ISSN 2078-4481

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХЕРСОНСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ

ВІСНИК

ХЕРСОНСЬКОГО НАЦІОНАЛЬНОГО ТЕХНІЧНОГО УНІВЕРСИТЕТУ

3(94)

Частина 2

Рекомендовано до друку Вченою радою
Херсонського національного технічного університету
(протокол № 12 від 05.06.2025 року)

Журнал включено до Переліку наукових фахових видань України
категорії «Б» зі спеціальностей: С1, D1, D2, D3, D5, D7, J3
(Наказ МОН України від 17.03.2020 № 409);
D4 (Наказ МОН України від 29.06.2021 № 735);
J5, J6, J7, J8, F2, F3, F5, F6, F7, G7, G8, G9, G11
(Наказ МОН України від 02.07.2020 № 886);
G1, G3, G4, G15 (Наказ МОН України від 24.09.2020 № 1188)

Журнал включено до наукометричних баз, електронних бібліотек та репозитаріїв:
GoogleScholar, Crossref, National Library of Ukraine (Vernadsky)

Видавничий дім
«Гельветика»
2025

Редакційна колегія

Головний редактор
Литвиненко В. І.
д.т.н., професор, завідувач кафедри інформатики і комп'ютерних наук,
заслужений діяч науки і техніки України

Заступник головного редактора
Сарібєкова Ю. Г.
д.т.н., професор

Шерстюк В. Г.
д.т.н., професор, проректор з навчальної роботи

Відповідальний секретар
Лур'є І. А.
к.т.н., доцент, начальник навчально-методичного відділу, доцент кафедри інформатики і комп'ютерних наук

Члени редакційної колегії

Баганов Є. О.	к.т.н., доцент
Березовський Ю. В.	д.т.н., доцент
Бойко Г. А.	к.т.н., доцент
Вороненко М. О.	к.т.н., доцент
Гончар О. І.	д.е.н., професор
Горбачов П. Ф.	д.т.н., професор
Дімітрова В.	д.н., доцент
Джерелюк Ю. О.	д.е.н., професор
Євтушенко В. В.	к.т.н., доцент
Жарікова М. В.	д.т.н., професор
Зубкова К. В.	к.т.н., доцент
Корчевська Л. О.	д.е.н., професор
Кузьміна Т. О.	д.т.н., професор
Куник О. М.	к.т.н., доцент
Луб'яний П. В.	к.т.н., доцент
Наумов О. Б.	д.е.н., професор
Олійник Н. М.	к.т.н., доцент
Плющ Р. М.	д.держ.упр., професор
Половцев О. В.	д.держ.упр., к.т.н., професор
Рудакова Г. В.	д.т.н., професор
Русанов С. А.	к.т.н., доцент
Савін С. Ю.	д.е.н., доцент
Салєба Л. В.	к.т.н., доцент
Семешко О. Я.	д.т.н., професор
Сідельникова Л. П.	д.е.н., професор
Smolarz A. (Польща)	dr.hab.inz.
Стоянова О. В.	к.т.н., доцент
Топалова Е. Х.	к.держ.упр., доцент
Філіппова В. Д.	д.держ.упр., професор
Хрущ Н. А.	д.е.н., професор
Чепелюк О. В.	д.т.н., професор
Шандова Н. В.	д.е.н., професор
Шарко О. В.	д.т.н., професор
Шарко М. В.	д.е.н., професор
Шевченко І. І.	д.т.н., професор

ISSN 2078-4481

MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
KHERSON NATIONAL TECHNICAL UNIVERSITY

VISNYK

OF KHERSON NATIONAL TECHNICAL UNIVERSITY

3(94)

Part 2

Recommended for publication by the Academic Council
of Kherson National Technical University
(Minutes № 12 on 5th June 2025)

The journal is included in the List of Scientific Professional Editions of Ukraine Category “B”
in specialties: C1, D1, D2, D3, D5, D7, J3
(Ukraine Education and Science Ministry Order dated 17.03.2020, № 409);
D4 (Ukraine Education and Science Ministry Order dated 29.06.2021, № 735);
J5, J6, J7, J8, F2, F3, F5, F6, F7, G7, G8, G9, G11
(Ukraine Education and Science Ministry Order dated 02.07.2020, № 886);
G1, G3, G4, G15
(Ukraine Education and Science Ministry Order dated 24.09.2020, № 1188)
The journal is included in the scientometric bases, electronic libraries and repositories:
Google Scholar, Crossref, National Library of Ukraine (Vernadsky)

Publishing House
“Helvetica”
2025

Editorial Board

Editor-in-Chief
Litvinenko V. I.
Doctor of Engineering Science, Professor, Head of the Department of Informatics and Computer Science, Honored Worker of Science and Technology of Ukraine

Deputy Editor-in-Chief
Saribekova Yu. G.
Doctor of Engineering Science, Professor

Sherstiuk V. H.
DSc (Engineering), Professor, Vice-Rector for Academic Affairs

Executive Secretary
Lurie I. Yu.
PhD (Engineering), Associate Professor, Head of the Educational and Methodical Department, Associate Professor of the Department of Informatics and Computer Sciences

Members of Editorial Board

Baganov Ye. O.	Ph.D., Associate Professor
Berezovsky Yu. V.	Doctor of Engineering Science, Associate Professor
Boiko H. A.	Ph.D., Associate Professor
Voronenko M. O.	Ph.D., Associate Professor
Honchar O. I.	Doctor of Economic Sciences, Professor
Horbachov P. F.	Doctor of Economics, Professor
Dimitrova V. Ya. (Bulgaria)	Ph.D., Associate Professor
Dzhereliuk Yu. A.	Doctor of Economics, Professor
Yevtushenko V. V.	Ph.D., Associate Professor
Zharikova M. V.	Doctor of Engineering Science, Professor
Zubkova K. V.	Ph.D., Associate Professor
Korchevska L. A.	Doctor of Economics, Professor
Kuzmina T. O.	Doctor of Technical Sciences, Professor
Kunyk O. N.	Ph.D., Associate Professor
Lubianyi P. V.	Ph.D., Associate Professor
Naumov O. B.	Doctor of Economics, Professor
Oliinyk N. M.	Ph.D., Associate Professor
Pliushch R. M.	Doctor of Sciences in Public Administration, Professor
Polovtsev O. V.	Doctor of Sciences in Public Administration, Professor
Rudakova H. V.	Doctor of Engineering Science, Professor
Rusanov S. A.	Ph.D., Associate Professor
Savin S. Yu.	Doctor of Economics, Associate Professor
Saleba L. V.	Ph.D., Associate Professor
Semeshko O. Ya.	Doctor of Engineering Science, Professor
Sidelnykova L. P.	Doctor of Economics, Professor
Smolarz A. (Poland)	Ph.D., Associate Professor
Stoianova O. V.	Ph.D., Associate Professor
Topalova E. K.	Ph.D., Associate Professor
Filippova V. D.	Doctor of Science in Public Administration, Professor
Khrushch N. A.	Doctor of Economics, Professor
Chepeliuk O. V.	Doctor of Technical Sciences, Professor
Shandova N. V.	Doctor of Economics, Professor
Sharko M. V.	Doctor of Economics, Professor
Sharko O. V.	Doctor of Engineering Science, Professor
Shevchenko I. I.	Doctor of Technical Sciences, Professor

ЗМІСТ

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

В. В. Аргунов, Д. Б. Задніпровський
Формування інвестиційного портфелю на основі прогнозування фондового ринку з використанням багатоджерельних даних……12
О. О. Бакаєв, А. І. Ісмагілов, І. О. Сиваченко, Ю. О. Вьюннік, В. Л. Шевченко
Модель оцінки рівня забезпечення живучості інформаційної системи з властивостями часткової регульованої взаємозамінності підсистем……18
O. Yu. Barkovska, O. V. Nechitailo, L. D. Buharova
Embedded voice assistance module for the visually impaired based on visual scene analysis……28
М. Р. Бєліков, А. О. Бур, С. В. Поперешняк
Мікросервісна архітектура для побудови крипто-фріланс біржі……37
В. Д. Бойко, В. І. Бевза, В. М. Слатвінська
Ініціювання мережевої активності в Windows за допомогою netstat……46
Д. О. Боровик, М. О. Зарецький
Використання дронів та штучного інтелекту у військовій стратегії України……52
С. Ю. Борю, Д. А. Резця, Н. О. Миронова, Т. І. Куляба-Харитонова
Програмна реалізація асимптотичних методів у Maxima для інтелектуальних систем аналізу та прогнозування руху об'єктів у динамічному середовищі……62
Є. М. Бровченко, В. П. Самарай
Аналіз проблем захисту неструктурованої інформації на мобільних пристроях: сучасні виклики та тенденції……68
В. Г. Васенко, Н. В. Корніловська, С. В. Вишемирська, М. В. Карамушка
Застосування методів машинного навчання для створення утиліт операційної системи Windows……79
В. В. Вінниченко
Калібрування невизначеності в комп'ютерному зорі: порівняння bnn, deep ensembles, mc dropout та evidential dl……85
D. S. Vovchenko, L. M. Oleshchenko
Development of a node balancing algorithm using machine learning and dynamic weight calculation……93
В. В. Войтко, М. Ю. Позур
Метод розширення можливостей повторного використання програмних компонентів шляхом реалізації traits в.Net……99
В. С. Волошин
Щодо питання про оцінку якості інформації в інтернеті без залучення цензури……104
A. O. Havrashenko, O. Yu. Barkovska
Assessment of the accuracy of artificial language translation methods……112
О. В. Герасіна
Моделювання трафіку інформаційно-комунікаційних мереж із використанням блочно-орієнтованих структур для систем виявлення атак……120
О. А. Голенко, М. В. Карамушка
Дослідження модернізації та впровадження інтернет технологій і систем моніторингу та управління у житлових приміщеннях……128
Т. Р. Голубець, М. О. Медиковський
Модель передбачення рішень людини про визнання втрати контролю над процесами……135
O. Y. Honsor
Trustworthy AI in Smart Sensor Systems in IoT: Challenges and Solutions……143
О. В. Горбовий, О. В. Лінник, І. М. Діордіца
Роль 3D-друку в персоналізованій розробці протезів та ортезів для пацієнтів з обмеженими можливостями……150
М. А. Гунько
Метод структурно-топологічного синтезу архітектури fog-середовища……159

Є. А. Дроздова, В. М. Козел
Сучасні методи захисту баз даних в умовах кіберзагроз та війни……166
O. A. Yeroshenko, A. I. Andrushchenko, A. R. Sorokin, V. I. Sitnikov
Application of ensemble learning algorithms for fraud detection in banking transactions……175
О. В. Задерейко, С. І. Єленич, Н. І. Логінова, О. Г. Трофименко, В. І. Гура
Аналіз результатів пошуку наукової інформації з використанням платформ штучного інтелекту……181
Б. М. Злотенко, Д. В. Стаценко, Г. В. Мельник, В. Ю. Калашнік
Методологія обробки та збереження інформації в комп'ютерній системі для проведення досліджень при підготовці і презентації кваліфікаційної роботи……192
M. O. Ilin, L. M. Oleshchenko
Development of a modular data unification pipeline for real-time environmental threat analytics……200
O. M. Katrusha
Intensity correction in side scan sonar images. Methods overview……207
В. Ю. Козуб, М. А. Семенов, С. І. Осадчук
Аналіз впливу WebAssembly на продуктивність складних вебзастосунків……218
Н. О. Комлева
Метод трансформації формалізованих вимог у розширені UML-моделі на основі граматичного підходу……224
Н. О. Комлева, М. І. Нікітченко
Метод двосторонньої синхронізації UML-моделі з двома поданнями на основі інкрементальних оновлень……232
М. М. Копилець, А. Я. Савка
Методи навчання з підкріпленням безпілотних літальних апаратів у завданнях військової логістики……239
M. M. Korablyov, O. O. Fomichov, O. K. Tkachuk
Controlling bots' behavior in RPG games using the immune model of clonal selection……245
Я. І. Корнага, О. М. Губарєв
Метод автоматизованого формування кластерів програмної системи на основі структурно-семантичного аналізу компонентів з використанням нейронних мереж……253
І. В. Круш
Особливості побудови та програмної реалізації стохастичних моделей цінових рухів на ринку FOREX……259
Л. І. Кублій, В. В. Лень
Аналіз загроз і можливостей впровадження штучного інтелекту в сферу освіти……269
Л. І. Кублій, А. Ю. Холодницька, М. В. Кузь, Б. С. Незамай, Р. О. Тичковський, І. Б. Полатайко
Кваліметрія вимог до програмних засобів на початкових етапах їх життєвого циклу……282
О. Б. Кунгурцев, Т. В. Діденко
Метод формування запиту на пошук в інтернеті партнера по інтересах з урахуванням характеристик запиту і адресанта……287
В. О. Куренко, Д. В. Андрушко, С. В. Пшеничних
Система автентифікації за обличчям з використанням трансферного навчання нейронних мереж……295
А. В. Легеза, І. Ф. Повхан, В. Я. Сароз, В. О. Яковлев
Модель оцінки ефективності компонента електронного навчального процесу на основі елементарних дидактичних показників……303
В. А. Лукашенко, Д. А. Гардер, А. В. Бернацький, В. М. Лукашенко, Г. А. Лукашенко
Модель ірраціональної функції спеціального призначення різних двійково-кодових системах числення в єдиному кристалі……309
O. O. Mamchych, M. O. Volk
Model and methodology of measuring and forecasting of energy impact of security risks mitigation approaches in distributed computing systems on personal mobile devices……318
М. О. Медиковський, Р. В. Мельник, М. В. Мельник, Д. О. Герез
Метод комп'ютерного моделювання процесів заряду/розряду акумулятора……326
М. О. Молчанова, О. В. Мазурець, О. О. Залуцька, В. Д. Кадинська, В. В. Масловська
Інформаційна хмарна технологія нейромережевого аналізу зруйнованих споруд за візуальними даними з БПЛА……334
В. К. Овсяк, В. Р. Турчак
Модель алгоритму взаємодії користувача з базою структурованих даних для адаптивного сортування у вебінтерфейсі……341

О. Є. Огнєва, Д. В. Зярко
Дослідження ефективності моделей реінженерії програмного забезпечення для цифрової трансформації туристичного сервісу……356
О. А. Пастух, О. М. Брик
Інтерпретована класифікація гемодинамічного стану на основі відбору ключових неінвазивних параметрів та аналізу важливості ознак……364
В. М. Пахомова, К. І. Зінкевич
Огляд технологій комп'ютерного зору на основі глибоких нейронних мереж щодо використання на залізничному транспорті……371
Б. В. Пашковський
Розробка telegram-бота для сповіщення про повітряну тривогу в Україні……377
О. А. Пономарьова, Г. М. Гузь, Д. В. Гузь
Розробка системи реєстрації та авторизації користувачів у клієнт-серверному web-додатку……382
І. В. Рубан, В. М. Ткачов
Багаторівнева модель інформаційної системи на мобільній платформі та формалізація критеріїв її живучості……388
М. В. Семаньків
Вдосконалення генератора псевдовипадкових чисел на регістрах зсуву зворотного зв'язку та дослідження його характеристик……399
O. I. Sypiahin, M. A. Yuzhakov, R. I. Ibrahimov
Architectural solutions for scalable web applications considering performance and latency……406
О. А. Снєосіков, О. П. Нарєжній
Моделі загроз та порушника автономної системи диференціальної корекції глобальних навігаційних супутникових систем……412
О. В. Собко, М. О. Молчанова, О. В. Мазурець, П. В. Гнатюк
Алгоритмізація процесів керування та інтеракцій в ігрових об'єктно-орієнтованих застосунках……424
І. В. Стоцький, В. П. Костенко, Г. І. Налісний, С. А. Шарні
Використання алгоритмів AI для виявлення та блокування кібератак на військові інформаційні системи……432
О. М. Супрун, Н. О. Чікіна, Я. Я. Кравчук
Інтелектуальні методи аналізу відеопотоку для виявлення прихованих повідомлень……437
О. І. Терещенко
Методи ін'єкції вразливостей у смарт-контракти для формування збалансованих датасетів……445
С. Д. Точилін
Комп'ютерна програма рішення різних варіантів задачі комівояжера з допомогою генетичного алгоритму……452
H. Franchevska, V. Dozorskyi
Additive vs additive–multiplicative models for noninvasive fetal electrocardiography: cascade extraction methods and noise robustness……457
Д. Є. Фролов
Модель забезпечення відмовостійкості інформаційних систем на основі багаторівневих структур……463
М. А. Фукс, І. С. Добринін, С. В. Пшеничних, А. В. Снігуров, В. Х. Чакрян
Методика побудови системи управління інформаційною безпекою на основі методів аналізу ієрархій, мережевого аналізу та теорії ігор в умовах обмежених ресурсів……473
M. O. Khvostivskyi, L. V. Khvostivska, I. Yu. Dediv, L. Ye. Dediv
Intelligent computer network traffic forecasting system based on synphase data processing……486
М. Ю. Хорканін, О. О. Медяков
Контрольоване застосування мультиагентних систем для планування завдань у інтелектуальних розподілених інформаційних системах……493
Д. Р. Чанкветадзе, Л. І.Фешанич
Розширення підходів до інтеграції формальної верифікації у CI/CD-конвеєри засобами модульної архітектури та сховищ знань……498
І. С. Чепурна
Метод організації захищеного сегмента приватних sdn-мереж на основі протоколу bittorrent……504

О. М. Шушура, Д. А. Ігнатов
Моделювання прийняття рішень з управління ресурсами мікросервісів та вибору віртуальних машин у середовищі kubernetes……515
Є. Б. Яворська, І. О. Гринюк
Застосування нейромережних технологій для підвищення інформативності медичних зображень……522

СОЦІАЛЬНІ ТА ПОВЕДІНКОВІ НАУКИ

О. В. Кривда, С. О. Кириченко, О. О. Вечкіс
Якісний аналіз ризиків рекламної галузі……527

CONTENTS

INFORMATION TECHNOLOGIES

V. V. Argunov, D. B. Zadniprovskyi
Formation of an investment portfolio based on stock market forecasting using multi-source data……12
O. O. Bakaiev, A. I. Ismagilov, I. O. Syvachenko, Yu. O. Vyunnik, V. L. Shevchenko
Model for assessing the level of survival of an information system with the properties of partial regulated interchangeability of subsystems……18
O. Yu. Barkovska, O. V. Nechitailo, L. D. Buharova
Embedded voice assistance module for the visually impaired based on visual scene analysis……28
M. R. Bielikov, A.O. Bur, S. V. Popereshnyak
Microservices architecture for building a crypto freelance exchange……37
V. D. Boyko, V. I. Bevza, V. M. Slatvinska
Initiating network activity in windows using netstat……46
D. O. Borovyk, M. O. Zaretskyi
The use of drones and artificial intelligence in ukraine's military strategy……52
S. Yu. Boriu, D. A. Reztsia, N. O. Myronova, T. I. Kulyaba-Kharytonova
Software implementation of asymptotic methods in maxima for intelligent systems for analysis and prediction of object motion in a dynamic environment……62
E. M. Brovchenko, V. P. Samaray
Analysis of unstructured data protection on mobile devices: current challenges and trends……68
V. G. Vasenko, N. V. Kornilovska, S. V. Vyshemyrska, M. V. Karamushka
Application of machine learning methods for creating windows operating system utilities……79
V. V. Vinnychenko
Uncertainty calibration in computer vision: comparison of bnn, deep ensembles, mc dropout and evidential dl……85
D. S. Vovchenko, L. M. Oleshchenko
Development of a node balancing algorithm using machine learning and dynamic weight calculation……93
V. V. Voitko, M. Yu. Pozur
Method of improving code reusability by implementing traits in.Net……99
V. S. Voloshyn
On the issue of assessing the quality of information on the Internet without the involvement of censorship……104
A. O. Havrashenko, O. Yu. Barkovska
Assessment of the accuracy of artificial language translation methods……112
O. V. Gerasina
Modeling of information and communication network traffic using block-oriented structures for attack detection systems……120
O. A. Golenko, M. V. Karamushka
Research on modernization and implementation of internet technologies and monitoring and management systems in residential premises……128
R. Holubets, O. Medykovskyy
A human decision prediction model for recognizing loss of control in processes……135
O. Y. Honsor
Trustworthy AI in Smart Sensor Systems in IoT: Challenges and Solutions……143
O. V. Gorbovy, O. V. Linnyk, I. M. Diorditsa
The role of 3D printing in the personalized design of prostheses and ortheses for patients with disabilities……150
M. A. Hunko
Deep integration of cloud and fog computing……159
Ye. A. Drozdova, V. M. Kozel
Modern methods of database protection under cyber threats and wartime conditions……166
O. A. Yeroshenko, A. I. Andrushchenko, A. R. Sorokin, V. I. Sitnikov
Application of ensemble learning algorithms for fraud detection in banking transactions……175

O. V. Zadereyko, S. I. Yelenych, N. I. Loginova, O. G. Trofymenko, V. I. Hura
Analysis of scientific information search results using artificial intelligence platforms……181
B. M. Zlotenko, D. V. Statsenko, H. V. Melnyk, V. Yu. Kalashnik
Methodology of information processing and storage in a computer system for conducting research in the preparation and presentation of qualification work……192
M. O. Ilin, L. M. Oleshchenko
Development of a modular data unification pipeline for real-time environmental threat analytics……200
O. M. Katrusha
Intensity correction in side scan sonar images. Methods overview……207
V. Yu. Kozub, M. A. Semenov, S. I. Osadchuk
Analysis of the impact of webassembly on the productivity of complex Web applications……218
N. O. Komleva
Grammar-based method for transforming formalized requirements into extended uml models……224
N. O. Komleva, M. I. Nikitchenko
Method for bilateral synchronization of a uml model with two representations based on incremental updates……232
M. M. Kopylets, A. Ya. Savka
Reinforcement learning methods for unmanned aerial vehicles in military logistics tasks……239
M. M. Korablyov, O. O. Fomichov, O. K. Tkachuk
Controlling bots' behavior in RPG games using the immune model of clonal selection……245
Ya. I. Kornaha, O. M. Hubariev
A method of automated formation of clusters of the software system based on structural-semantic analysis of components using neural networks……253
I. V. Krush
Features of construction and software implementation of stochastic models of price movements in the forex market……259
L. I. Kublii, V. V. Len
Analysis of threats and opportunities of implementing artificial intelligence in the sphere of education……269
L. I. Kublii, A. Yu. Kholodnytska
Statistical models for analysing training factors and predicting sports results……275
M. V. Kuz, B. S. Nezamai, R. O. Tychkovskyi, I. B. Polataiko
Qualimetry of software requirements at the initial stages of their life cycle……282
O. B. Kungurtsev, T. V. Didenko
Method for forming a query to search for an internet partner based on interests with consideration of query and addressee characteristics……287
V. O. Kurenko, D. V. Andrushko, S. V. Pshenychnykh
Face authentication system using transfer learning of neural networks……295
A. V. Legeza, I. F. Povkhan, V. Ya. Saroz, V. O. Yakovlev
Model for evaluating the effectiveness of the e-learning process component based on elementary didactic indicators……303
V. A. Lukashenko, D. A. Harder, V. Bernatskyi, V. M. Lukashenko, H. A. Lukashenko
Model of a special-purpose irrational function in various binary-coded numbering systems in a single crystal……309
O. O. Mamchych, M. O. Volk
Model and methodology of measuring and forecasting of energy impact of security risks mitigation approaches in distributed computing systems on personal mobile devices……318
M. O. Medykovskyi, R. V. Melnyk, M. V. Melnyk, D. O. Herez
Computer modeling method for battery charge and discharge processes……326
M. O. Molchanova, O. V. Mazurets, O. O. Zalutska, V. D. Kadynska, V. V. Maslovska
Cloud information technology for neural network analysis of destroyed structures using visual data from uavs……334
V. K. Ovsiak, V. R. Turchak
Model of the algorithm for user interaction with a structured data base for adaptive sorting in the web interface……341
O. Ye. Ohnieva, D. V. Ziarko
Research of the software reengineering models effectiveness for the tourism services digital transformation……356
O. A. Pastukh, O. M. Bryk
Interpretable classification of hemodynamic state based on the selection of key non-invasive parameters and feature importance analysis……364

V. M. Pakhomova, K. I. Zinkevych
Review of computer vision technologies based on deep neural networks for use in rail transport……371
B. V. Pashkovskyi
Development of a telegram bot for air raid alerts in Ukraine……377
O. A. Ponomarova, H. M. Huz, D. V. Huz
Development of a user registration and authorization system for a client-server web application……382
I. V. Ruban, V. M. Tkachov
Multilevel model of an information system on a mobile platform and formalization of its survivability criteria……388
M. V. Semankiv
Improvement of a pseudorandom number generator based on a linear feedback shift register and investigation of its characteristics……399
O. I. Sypiahin, M. A. Yuzhakov, R. I. Ibrahimov
Architectural solutions for scalable web applications considering performance and latency……406
O. A. Snieosikov, O. P. Nariezhnii
Threat and violator models of the autonomous system of differential correction of global navigation satellite systems……412
O. V. Sobko, M. O. Molchanova, O. V. Mazurets, P. V. Hnatiuk
Algorithmisation of control and interaction processes in object-oriented game applications……424
I. V. Stotskyi, V. P. Kostenko, G. I. Nalisnyi, S. A. Sharnin
Use of ai algorithms for detection and blocking of cyber attacks on military information systems……432
O. M. Suprun, N. O. Chikina, Ya. Ya. Kravchuk
Intelligent methods of video stream analysis for detecting hidden messages……437
O. I. Tereshchenko
Methods of vulnerability injection into smart contracts for balanced dataset generation……445
S. D. Tochilin
Computer program for solving various versions of the traveling salesman problem using a genetic algorithm……452
H. Franchevska, V. Dozorskyi
Additive vs additive–multiplicative models for noninvasive fetal electrocardiography: cascade extraction methods and noise robustness……457
D. Ye. Frolov
Model for ensuring fault tolerance of information systems based on multi-level structures……463
M. A. Fuks, I. S. Dobrynin, S. V. Pshenychnykh, A. V. Snihurov, V. H. Chakrian
Methodology for developing information security management systems using network analysis and game theory……473
M. O. Khvostivskyi, L. V. Khvostivska, I. Yu. Dediv, L. Ye. Dediv
Intelligent computer network traffic forecasting system based on synphase data processing……486
M. Yu. Khorkanin, o. O. Mediakov
Controlled application of multi-agent systems for task planning in intelligent distributed information systems……493
D. R. Chankvetadze, L. I. Feshanich
Expanding approaches to formal verification integration into ci/cd pipelines using modular architecture and knowledge repositories……498
I. S. Chepurna
Multi-level vpn-tunneling method for ensuring remote access to extranet network nodes……504
O. M. Shushura, D. A. Ihnatov
Modeling decision-making for microservice resource management and virtual machine selection in a kubernetes environment……515
E. B. Yavorska, I. O. Hryniuk
Application of neural network technologies for enhancing the informativeness of medical images……522

SOCIAL AND BEHAVIORAL SCIENCES

O. V. Kryvda, S. O. Kyrychenko, O. O. Vechkis
Qualitative risks analysis of the advertising industry……527

ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ

УДК 004.9 DOI

В. В. АРГУНОВ
асистент кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0000-0002-2505-1969

Д. Б. ЗАДНІПРОВСЬКИЙ
аспірант кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0009-0004-6637-6315

ФОРМУВАННЯ ІНВЕСТИЦІЙНОГО ПОРТФЕЛЮ НА ОСНОВІ ПРОГНОЗУВАННЯ ФОНДОВОГО РИНКУ З ВИКОРИСТАННЯМ БАГАТОДЖЕРЕЛЬНИХ ДАНИХ

У дослідженні використано дані з різних джерел: історичні ціни акцій, технічні індикатори, новинні повідомлення та публікації з соціальних мереж. Такий підхід дозволяє поєднувати як кількісні, так і якісні фактори, що безпосередньо впливають на динаміку фондового ринку, та формувати більш повну й комплексну інформаційну базу для прогнозування. На відміну від традиційних моделей, які переважно враховують лише статистичні характеристики часових рядів, застосування багатоджерельних даних дає змогу краще відобразити складні взаємозв'язки між економічними, інформаційними та поведінковими чинниками, що визначають коливання вартості цінних паперів. Експериментальні результати підтвердили, що збільшення кількості шарів і нейронів у нейронних мережах, а також застосування регуляризації у вигляді dropout, значно підвищують точність прогнозів і дозволяють уникати перенавчання. Найкраща модель – чотиришарова BiLSTM із 150 нейронами та параметром dropout 0.2 – досягла показника MAPE 4.36, перевершивши аналогічну модель LSTM. Це свідчить про здатність BiLSTM краще враховувати довгострокові залежності та приховані патерни у часових рядах. Водночас важливим обмеженням виявився суттєво більший час навчання BiLSTM, що може ускладнювати або навіть унеможливлювати її застосування в системах реального часу, де швидкість прийняття рішень є критичною. Метою дослідження є розробка ефективної моделі прогнозування динаміки фондового ринку на основі багатоджерельних даних із застосуванням сучасних методів глибинного навчання, здатної забезпечувати високу точність і надійність результатів навіть в умовах нестабільності. Об'єктом дослідження виступає процес прогнозування цін на акції в умовах високої волатильності та непередбачуваності фінансових ринків, що зумовлено як внутрішніми економічними факторами, так і зовнішніми інформаційними впливами. Предметом дослідження є методи та моделі глибинного навчання (зокрема LSTM та BiLSTM), орієнтовані на аналіз часових рядів і текстових даних із різних інформаційних джерел, інтеграція яких дозволяє підвищити якість прогнозування. Отримані результати свідчать про практичну доцільність використання глибинного навчання в управлінні інвестиціями, оскільки такі методи забезпечують більш точні й надійні прогнози в умовах значної невизначеності та нестабільності. Вони можуть стати ефективним інструментом для оптимізації формування інвестиційних портфелів, зниження ризиків і підвищення прибутковості рішень у сфері фінансів. Подальші дослідження можуть бути спрямовані на вдосконалення архітектур моделей для прискорення процесу навчання та розробку гібридних систем, що поєднують глибинне навчання з методами обробки природної мови для ще глибшого аналізу впливу інформаційних факторів на фондовий ринок.

Ключові слова: *машинне навчання, прогнозування, інвестиційний портфель, модель, штучний інтелект, аналіз, обробка тексту.*

© Аргунов В. В., Задніпровський Д. Б., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

V. V. ARGUNOV
Assistant at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0000-0002-2505-1969

D. B. ZADNIPROVSKYI
Postgraduate Student at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0009-0004-6637-6315

FORMATION OF AN INVESTMENT PORTFOLIO BASED ON STOCK MARKET FORECASTING USING MULTI-SOURCE DATA

The study uses data from various sources: historical stock prices, technical indicators, news reports, and social media publications. This approach makes it possible to combine both quantitative and qualitative factors that directly influence stock market dynamics and to build a more complete and comprehensive information base for forecasting. Unlike traditional models, which mainly take into account only the statistical characteristics of time series, the use of multi-source data allows for a better representation of the complex interrelationships between economic, informational, and behavioral factors that determine fluctuations in the value of securities. The experimental results confirmed that increasing the number of layers and neurons in neural networks, as well as applying regularization in the form of dropout, significantly improves forecast accuracy and helps to avoid overfitting. The best-performing model – a four-layer BiLSTM with 150 neurons and a dropout parameter of 0.2 – achieved a MAPE of 4.36, outperforming the comparable LSTM model. This demonstrates BiLSTM's ability to better capture long-term dependencies and hidden patterns in time series. At the same time, a significant limitation was the considerably longer training time required by BiLSTM, which may complicate or even preclude its use in real-time systems where decision-making speed is critical. The aim of the study is to develop an effective model for forecasting stock market dynamics based on multi-source data using modern deep learning methods, capable of ensuring high accuracy and reliability of results even under unstable conditions. The object of the study is the process of stock price forecasting under conditions of high volatility and unpredictability of financial markets, driven by both internal economic factors and external informational influences. The subject of the study is deep learning methods and models (in particular LSTM and BiLSTM) focused on the analysis of time series and textual data from various information sources, the integration of which makes it possible to improve forecasting quality. The obtained results indicate the practical feasibility of applying deep learning in investment management, since such methods provide more accurate and reliable forecasts under conditions of significant uncertainty and instability. They can become an effective tool for optimizing investment portfolio formation, reducing risks, and increasing the profitability of financial decision-making. Further research may be directed toward improving model architectures to accelerate the training process and developing hybrid systems that combine deep learning with natural language processing methods for a deeper analysis of the influence of informational factors on the stock market.

Key words: *machine learning, forecasting, investment portfolio, model, artificial intelligence, analysis, text processing.*

Постановка проблеми

Прогнозування цін на акції може дати можливості для збільшення прибутку як для продавця, так і для брокера. В минулому робити точні прогнози на фондовому ринку було складно через брак технологій і обчислювальних можливостей. Проте зі стрімким розвитком технологій точність прогнозів значно підвищилася, і сьогодні завдання прогнозування стає значно простішим порівняно з минулими роками. Прогнози на фондовому ринку часто мають хаотичний характер, а не випадковий, що означає можливість їх передбачення шляхом аналізу історії динаміки цін.

Машинне навчання є дієвим методом для моделювання складних процесів, зокрема для прогнозування ринкових значень, близьких до фактичних. Завдяки зростанню точності машинне навчання стає ефективним інструментом для практичного застосування. Багато дослідників дійшли висновку, що машинне навчання є результативним і точним способом оцінювання фондових цін, завдяки чому ця технологія набуває популярності у фінансовій сфері [1–4].

Прогнозування фондового ринку останнім часом привертає значну увагу як дослідників у фінансовій сфері, так і інвесторів. Оскільки ринок характеризується високою нелінійністю та коливаннями часових рядів, для його моделювання потрібні ефективні підходи, здатні працювати з такими різнорідними даними. Традиційні лінійні моделі, зокрема ARMA та ARIMA, раніше застосовувалися для прогнозування, проте вони придатні лише для окремих акцій і не можуть бути поширені на інші. Через стохастичний і неоднозначний характер фондового ринку здійснення прогнозів із використанням класичних методів є надзвичайно складним та ризикованим. У цьому контексті дедалі важливішу роль починають відігравати методи глибинного навчання, які демонструють високу результативність у виявленні закономірностей і трендів фондового ринку.

Технології глибинного навчання мають здатність працювати з високонелінійними даними, що робить їх перспективними для прогнозування руху цін на акції. Доведено, що моделі на основі глибинного навчання перевершують ARIMA у прогнозуванні часових рядів [5], особливо для завдань довгострокового передбачення. Зокрема, LSTM-моделі продемонстрували вищу точність порівняно з ARIMA, однак актуальним завданням залишається перевірка того, чи можна ще більше покращити їхні результати за рахунок використання багатошарових архітектур. Водночас ринково значуща інформація надходить із численних джерел – соціальних мереж, новинних ресурсів, корпоративних оглядів – і саме на ній інвестори будують прогнози. Обробка таких великих і різнорідних масивів даних є складним викликом навіть для DL-моделей. Крім того, багатомасштабність часових рядів у фінансовій сфері ускладнює роботу традиційних методів, роблячи їх малоефективними для точних прогнозів. Важливо враховувати, що будь-яка похибка у прогнозі може призвести до значних фінансових втрат, тому інвестори потребують максимально точних оцінок ринкової вартості, що робить завдання прогнозування ще більш складним.

Таким чином, актуальним постає дослідження та розробка сучасних підходів до прогнозування фондового ринку, здатних враховувати складну нелінійну динаміку часових рядів і різнорідність джерел інформації. Використання методів глибинного навчання, зокрема моделей LSTM та BiLSTM, відкриває можливості для підвищення точності передбачення та формування більш надійних інструментів підтримки інвестиційних рішень, що має вагоме практичне значення в умовах високої волатильності фінансових ринків.

Аналіз останніх досліджень і публікацій

У сучасних наукових дослідженнях спостерігається зростаючий інтерес до проблеми прогнозування фондових ринків. Значна кількість робіт присвячена використанню статистичних методів, таких як ARMA та ARIMA, які тривалий час вважалися основними інструментами для моделювання часових рядів. Водночас дослідники відзначають, що такі моделі обмежені у застосуванні, оскільки вони враховують лише окремі аспекти динаміки ринку і не здатні адекватно відображати складні нелінійні залежності, характерні для фінансових даних. Це створило передумови для пошуку нових методологічних підходів, здатних забезпечити більш високу точність і стійкість прогнозів.

Останніми роками значна увага приділяється застосуванню методів машинного та глибинного навчання для прогнозування біржових показників. Зокрема, у наукових публікаціях активно досліджуються можливості моделей на основі нейронних мереж, серед яких важливе місце займають LSTM та BiLSTM, що демонструють здатність ефективно працювати з нелінійними часовими рядами. Крім того, у сучасних публікаціях акцент робиться на інтеграції ринкових і текстових даних (новин, соціальних мереж, аналітичних звітів), що дозволяє підвищити точність прогнозів. Такий підхід поступово витісняє традиційні методи, утверджуючи глибинне навчання як провідний інструмент досліджень у сфері фінансового прогнозування.

У статті [6] автори демонструють новий оптимізаційний підхід до прогнозування фондових цін, заснований на моделі Multi-Layer Sequential LSTM (MLS LSTM), яка використовує оптимізатор Adam. Алгоритм MLS LSTM застосовує нормалізовані часові ряди, поділені на часові кроки, щоб визначати взаємозв'язок між минулими та майбутніми значеннями з метою підвищення точності прогнозів. Крім того, він усуває проблему зникання градієнта, притаманну простим рекурентним нейронним мережам. Модель MLS LSTM демонструє надзвичайно високі показники точності, однак її недоліками є можливе перенавчання, використання обмеженого набору лише цінових даних без урахування зовнішніх факторів та відсутність перевірки стабільності прогнозів у різних ринкових умовах.

У роботі [7] пропонується проста архітектура на основі багатошарових перцептронів (MLP) під назвою StockMixer, яка легко оптимізується та демонструє високу прогнозну ефективність. StockMixer виконує послідовно: змішування індикаторів, потім змішування часових характеристик, а далі змішування акцій. На відміну від стандартного MLP-змішування, у нашому підході часовий блок передбачає обмін багатомасштабною часовою інформацією, тоді як змішування акцій реалізується через явне врахування впливів «акція-ринок» та «ринок-акція». Результати експериментів на реальних фондових даних показали, що StockMixer перевершує сучасні методи прогнозування з помітним відривом, одночасно знижуючи використання пам'яті та витрати часу на обчислення.

У дослідженні представленому у статті [8] запропоновано метод, спрямований на підвищення ефективності системи SPF на основі сучасних алгоритмів машинного й глибинного навчання. Спершу застосовано екстремальний градієнтний бустинг (XGBoost) як техніку відбору ознак для виокремлення найважливіших характеристик із багатовимірних часових рядів та усунення надлишкових параметрів. Далі відібрані ознаки були подані на вхід глибинної мережі довготривалої короткочасної пам'яті (LSTM) для прогнозування цін акцій. Така архітектура дозволила відобразити часову природу вхідних даних та повністю використати контекстну інформацію для майбутніх прогнозів. Складна структура моделі дала змогу захопити більше стохастичних залежностей у динаміці цін. Запропонований метод є універсальним і може застосовуватися як до даних про акції, так і до валютних ринків (Forex). Недоліком цього підходу є його висока обчислювальна складність, що зумовлена використанням XGBoost для відбору ознак та глибинної LSTM-мережі для прогнозування. Крім того, модель значною мірою залежить від якості та обсягу вхідних даних, що може обмежувати її узагальнюваність на інших ринках або в умовах різкої зміни економічних факторів.

Отже, узагальнюючи результати останніх досліджень, можна стверджувати, що формування інвестиційного портфеля на основі прогнозування фондового ринку в реальному часі з використанням багатоджерельних даних є надзвичайно актуальним напрямом, оскільки поєднання глибинного навчання та різнорідної інформації дає змогу підвищити точність прогнозів і забезпечити ефективність прийняття фінансових рішень.

Формулювання мети дослідження

Метою роботи є розробка та оцінка підходу до формування інвестиційного портфеля на основі прогнозування фондового ринку в реальному часі з використанням багатоджерельних даних.

Викладення основного матеріалу дослідження

Для ефективного управління інвестиціями необхідно мати надійні інструменти прогнозування, які враховують складність та динамічність фондового ринку. В умовах швидкої зміни ринкової інформації традиційні підходи до формування інвестиційного портфелю втрачають свою ефективність. Саме тому в цьому дослідженні пропонується використання багатоджерельних даних – зокрема новинних повідомлень, публікацій у соціальних мережах, технічних індикаторів та історичних цін акцій – для побудови моделі, здатної прогнозувати рух ринку в реальному часі.

Рисунок 1 зображує структурну модель формування інвестиційного портфелю, яка базується на прогнозах фондового ринку, отриманих за допомогою сучасних методів аналізу та обробки даних. Ця модель поєднує етапи збору, підготовки, аналізу інформації та прийняття рішень, що дозволяє формувати збалансований портфель з урахуванням ризиків і потенційного прибутку в умовах високої ринкової волатильності.

Дана діаграма ілюструє структурну модель формування інвестиційного портфелю на основі прогнозування фондового ринку в реальному часі з використанням багатоджерельних даних. У центрі моделі – поетапна обробка різних типів інформації: твіти з Twitter, новинні статті з вебсайтів і фондові дані з біржі NASDAQ. Після збору ці дані проходять очищення від помилок, дублювань та пропущених значень. Текстові дані додатково обробляються шляхом видалення хештегів, емодзі, посилань, а також застосовується лематизація.

Кількісні характеристики нормалізуються, а ключові ознаки виділяються за допомогою методу головних компонент (PCA). Паралельно застосовується аналіз сентименту за допомогою попередньо навченої трансформерної моделі BERT, що дозволяє класифікувати новини та твіти за емоційним забарвленням і додати відповідні мітки до текстів.

Після підготовки всі ознаки передаються до моделі Bi-LSTM, яка адаптована за допомогою алгоритму оптимізації Adam. Результати прогнозу оцінюються за допомогою метрик точності. Такий підхід дає змогу побудувати ефективну систему прогнозування змін на фондовому ринку, що є основою для формування обґрунтованого інвестиційного портфелю.

У цій статті як метрику оцінки було обрано MAPE. Підсумкові значення MAPE для моделей LSTM отримувалися як середнє всіх значень після крос-валідації.

Одношарова модель LSTM з 32 нейронами та часовим кроком у 1 день (look-back) досягла MAPE на рівні 12.55 для прогнозування скоригованої ціни закриття Google на наступний день. Для порівняння, модель BiLSTM з такою ж конфігурацією показала дещо кращий результат – 12.49, але при цьому час навчання зріс: LSTM тренувався 10.19 секунд, тоді як BiLSTM – 13.50 секунд.

Додавання ще одного шару та збільшення кількості нейронів до 64 покращило точність: MAPE склав 9.28 для LSTM і 9.07 для BiLSTM. Водночас час тренування скоротився до 8.08 секунд для LSTM та 12.95 секунд для BiLSTM. Хоча BiLSTM демонструє кращу точність, його навчання займає майже удвічі більше часу.

Подальші експерименти включали використання невеликого dropout у 0.2 при двошаровій структурі з 64 нейронами. Це дало суттєве покращення: MAPE знизився до 5.24 у LSTM і 5.14 у BiLSTM. Загалом, чим більше шарів та нейронів використовувалося, тим точнішою ставала модель, хоча час тренування збільшувався. Рузультати дослідження представлені в таблиці 1

Найкращою архітектурою виявилась 4-шарова модель з 150 нейронами та dropout 0.2, яка при часовом кроці 1 день показала результати: 4.85 для LSTM і 4.36 для BiLSTM. Таким чином, BiLSTM стабільно перевершував LSTM за точністю, проте потребував значно більше часу на навчання.

Висновки

Прогнозування цін на акції є надзвичайно актуальним завданням сучасної фінансової сфери, адже від точності передбачень залежить ефективність інвестиційних рішень. Традиційні статистичні моделі, такі як ARMA та ARIMA, виявилися обмеженими у застосуванні через неспроможність адекватно відображати складну нелінійну динаміку фондового ринку. Це зумовило перехід до використання методів глибинного навчання, серед яких LSTM та BiLSTM довели свою високу ефективність у роботі з нелінійними часовими рядами та багатоджерельними даними.

Експериментальні результати показали, що збільшення кількості шарів і нейронів, а також застосування dropout, суттєво знижує похибку прогнозування. Найкраща архітектура – чотиришарова модель із 150 нейронами та dropout 0.2 – забезпечила MAPE 4.85 для LSTM і 4.36 для BiLSTM. Порівняння моделей підтвердило, що

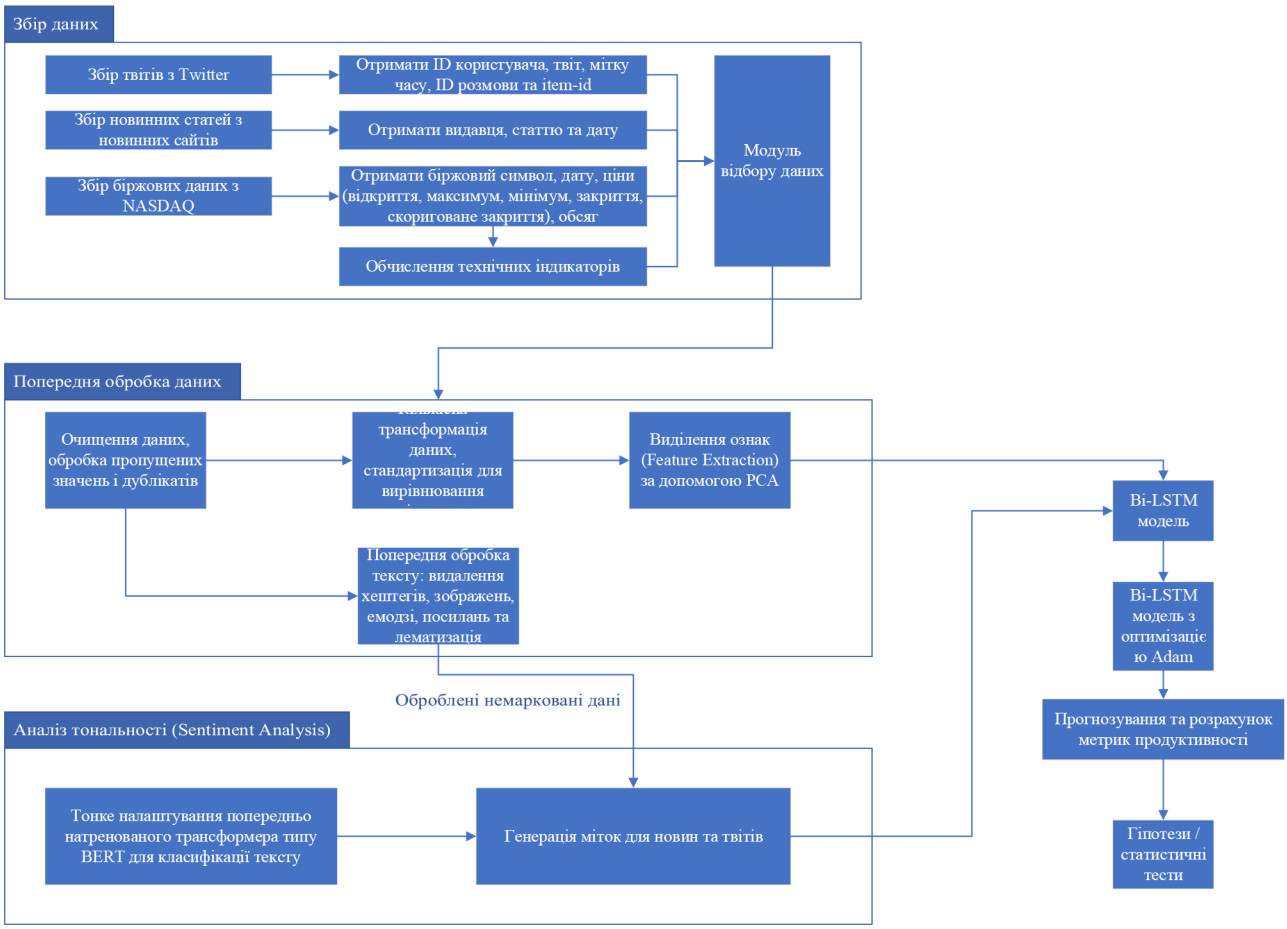


Рис. 1 Структурна модель формування інвестиційного портфелю на основі прогнозування фондового ринку в реальному часі з використанням багатоджерельних даних

Таблиця 1

Порівняння результатів моделей LSTM та BiLSTM

Параметри	Базова	Розширена	З dropout	Оптимальна
Шари	1	2	2	4
Нейрони	32	64	64	150
Dropout	–	–	0.2	0.2
Часовий крок	1 день	1 день	1 день	1 день
MAPE (LSTM)	12.55	9.28	5.24	4.85
Час навчання (LSTM)	10.19 с	8.08 с	–	–
MAPE (BiLSTM)	12.49	9.07	5.14	4.36
Час навчання (BiLSTM)	13.50 с	12.95 с	–	–

BiLSTM стабільно демонструє кращу точність, проте потребує значно більше часу на навчання, що може бути критичним для завдань у реальному часі.

Практичне значення дослідження полягає у використанні багатоджерельних даних, таких як історичні ціни, технічні індикатори, новини та твіти, що в комплексі з методами глибинного навчання дозволяє підвищити точність прогнозів і створити надійні інструменти для управління інвестиційними портфелями в умовах високої волатильності фінансових ринків.

Список використаної літератури

1. Nti, I.K., Adekoya, A.F. & Weyori, B.A. A novel multi-source information-fusion predictive framework based on deep neural networks for accuracy enhancement in stock market prediction. J Big Data 8, 17 (2021), doi: https://doi.org/10.1186/s40537-020-00400-y

2. X. Zhang, S. Qu, J. Huang, B. Fang and P. Yu, “Stock Market Prediction via Multi-Source Multiple Instance Learning”, in IEEE Access, vol. 6, pp. 50720-50728, 2018, doi: 10.1109/ACCESS.2018.2869735

3. Wu, Shengting, et al. “S_I_LSTM: stock price prediction based on multiple data sources and sentiment analysis”. Connection Science 34.1 (2022): 44–62.

4. Zhang, Qiuyue, et al. “Graph-based stock prediction with multisource information and relational data fusion”. Information Sciences 690 (2025): 121561.

5. Mashadihasanli, Tamerlan. “Stock market price forecasting using the ARIMA model: an application to Istanbul, Turkiye”. İktisat Politikası Araştırmaları Dergisi 9.2 (2022): 439–454.

6. Md, Abdul Quadir, et al. “Novel optimization approach for stock price forecasting using multi-layered sequential LSTM”. *Applied Soft Computing* 134 (2023): 109830, doi: https://doi.org/10.1016/j.asoc.2022.109830

7. Fan, J., & Shen, Y. (2024). StockMixer: A Simple Yet Strong MLP-Based Architecture for Stock Price Forecasting. Proceedings of the AAAI Conference on Artificial Intelligence, 38(8), 8389-8397, doi: https://doi.org/10.1609/aaai.v38i8.28681

Дата першого надходження рукопису до видання: 21.09.2025
Дата прийнятого до друку рукопису після рецензування: 16.10.2025
Дата публікації:

УДК 519.86:00'4.94:004.89 DOI

О. О. БАКАЄВ
аспірант
Інститут програмних систем Національної академії наук України
ORCID: 0009-0004-5427-1196

А. І. ІСМАГІЛОВ
аспірант
Інститут програмних систем Національної академії наук України
ORCID: 0009-0002-1332-8147

І. О. СИВАЧЕНКО
аспірант
Інститут програмних систем Національної академії наук України
ORCID: 0009-0005-3248-3371

Ю. О. ВЬЮННІК
аспірант
Інститут програмних систем Національної академії наук України
ORCID: 0009-0002-3213-8058

В. Л. ШЕВЧЕНКО
доктор технічних наук, професор,
заступник з наукової роботи директора
Інститут програмних систем Національної академії наук України
ORCID: 0000-0002-9457-7454

МОДЕЛЬ ОЦІНКИ РІВНЯ ЗАБЕЗПЕЧЕННЯ ЖИВУЧОСТІ ІНФОРМАЦІЙНОЇ СИСТЕМИ З ВЛАСТИВОСТЯМИ ЧАСТКОВОЇ РЕГУЛЬОВАНОЇ ВЗАЄМОЗАМІННОСТІ ПІДСИСТЕМ

Мета дослідження – розробка моделі оцінки рівня живучості інформаційної системи, яка дозволяє знаходити оптимальних розподіл ресурсів між своїми підсистемами з урахуванням рівня їх взаємозамінності. В складі системи забезпечення живучості інформаційної системи розглянуті: підсистема виявлення шкідливих впливів на основі вже відомих сценаріїв, підсистема виявлення атак «нульового дня», підсистема запобігання виявленим вторгненням. Найбільш адекватною визначена залежність корисного ефекту підсистеми від вхідних ресурсів на основі логістичних залежностей, які однозначно задаються параметрами: ординати асимптот, постійна величина логістичної залежності, що визначає швидкість максимального зростання логістичної залежності, та абсциса точки симетрії, що залежить від початкових умов. Для об'єднання ефектів підсистем в загальний ефект системи використані скалярні згортки. Запропоновано використовувати базові згортки з властивостями повної взаємозамінності підсистем (адитивна) та повної невзаємозамінності підсистем (мінімізуюча). Запропонована адитивно-мінімізуюча згортка, яка дозволяє налаштовувати модель на різні рівні взаємозамінності підсистем у всьому діапазоні можливих значень. Сформульовані залежності для пошуку оптимального розподілу ресурсів між підсистемами для забезпечення максимуму корисного ефекту системи. Оптимальне рішення знаходиться методом повного прямого перебору, що дозволило уникнути локальних оптимумів і впевнено знайти глобальний максимум. Оптимальна точка змінює своє положення в залежності від значень параметрів коефіцієнта рівня взаємозамінності та сумарного ресурсу системи. Побудовані траєкторії руху оптимальних рішень, для різних значень параметрів, що дозволяє не тільки знаходити оптимальне рішення для поточних параметрів, але й знаходити параметри, які б дозволили покращити ефект системи з урахуванням можливої зміни сценаріїв роботи.

Ключові слова: *моделювання, оптимізація, логістичні залежності, ефект, живучість, ресурси, інформаційні системи.*

© Бакаєв О. О., Ісмагілов А. І., Сиваченко І. О., Вьюннік Ю. О., Шевченко В. Л., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

O. O. BAKAIEV
Postgraduate Student
Institute of Software Systems
of the National Academy of Sciences of Ukraine
ORCID: 0009-0004-5427-1196

A. I. ISMAGILOV
Postgraduate Student
Institute of Software Systems
of the National Academy of Sciences of Ukraine
ORCID: 0009-0002-1332-8147

I. O. SYVACHENKO
Postgraduate Student
Institute of Software Systems
of the National Academy of Sciences of Ukraine
ORCID: 0009-0005-3248-3371

Yu. O. VYUNNIK
Postgraduate Student
Institute of Software Systems
of the National Academy of Sciences of Ukraine
ORCID: 0009-0002-3213-8058

V. L. SHEVCHENKO
Doctor of Technical Sciences, Professor,
Deputy Director for Scientific Work
Institute of Software Systems
of the National Academy of Sciences of Ukraine
ORCID: 0000-0002-9457-7454

MODEL FOR ASSESSING THE LEVEL OF SURVIVAL OF AN INFORMATION SYSTEM WITH THE PROPERTIES OF PARTIAL REGULATED INTERCHANGEABILITY OF SUBSYSTEMS

The purpose of the study is to develop a model for assessing the level of survivability of an information system, which allows finding the optimal distribution of resources between its subsystems, taking into account the level of their interchangeability. The following subsystems are considered as part of the survivability system: a subsystem for detecting harmful effects based on already known scenarios, a subsystem for detecting "zero-day" attacks, and a subsystem for preventing detected intrusions. The most adequate is the dependence of the useful effect of the subsystem on the input resources based on logistic dependencies, which are uniquely specified by the parameters: the ordinates of the asymptotes, the constant value of the logistic dependency, which determines the rate of maximum growth of the logistic dependency, and the abscissa of the symmetry point, which depends on the initial conditions. Scalar convolutions are used to combine the effects of subsystems into the overall effect of the system. It is proposed to use basic convolutions with the properties of complete interchangeability of subsystems (additive) and complete non-interchangeability of subsystems (minimizing). An additive-minimizing convolution is proposed, which allows you to adjust the model to different levels of subsystem interchangeability in the entire range of possible values. Dependencies are formulated to find the optimal distribution of resources between subsystems to ensure the maximum useful effect of the system. The optimal solution is found by the method of complete direct search, which made it possible to avoid local optima and confidently find the global maximum. The optimal point changes its position depending on the values of the parameters of the interchangeability level coefficient and the total resource of the system. The trajectories of the optimal solutions are constructed for different parameter values, which allows not only to find the optimal solution for the current parameters, but also to find parameters that would allow improving the effect of the system, taking into account a possible change in work scenarios.

Key words: *modeling, optimization, logistic dependencies, effect, survivability, resources, information systems*

Постановка проблеми

Інформаційні системи (ІС) знаходяться в стані щоденних загроз. Всіх зловмисників ми усунути не можемо. Але можемо захищати ІС. Відбувається постійне зростання витрат на захист інформації. З 2024 до 2028 року очікується зростання глобального ринку систем інформаційної безпеки від 183 млрд.дол. до 292 млрд.дол. [1]. В той же період очікується зростання ринку послуг щодо безпеки підприємств на 12.5 % (до 116.8 млрд.дол.), зростання ринку програмного забезпечення для безпеки підприємств на 14.1 % (до 132.4 млрд.дол.), ринку мережевої

безпеки підприємств на 10.6 % (до 33.1 млрд.дол.), ринку програмного забезпечення для безпеки споживачів на 4 % (до 9.7 млрд.дол.) [1]. На порядку денному питання живучості ІС, тобто збереження їх працездатності в умовах зловмисних впливів [2]. Вузькоспрямовані методи зазвичай мають локальний ефект і не вирішують проблеми на системному рівні, як, наприклад, підходи до безпечного формування ключів [3]. Для забезпечення захисту треба знати залежність рівня живучості ІС від ресурсного забезпечення. Практичні експерименти для виявлення закономірностей можуть коштувати занадто дорого. Особливо, якщо вони не вдалі. Високу ефективність мають методи імітаційного моделювання [4, 5]. Таким чином, **актуальною** є задача створення імітаційних моделей оцінки рівня живучості інформаційних систем. Такі моделі дозволяють прогнозувати наслідки управлінських рішень, щодо забезпечення живучості та знаходити оптимальне керування живучістю інформаційною системою, зокрема оптимальний розподіл ресурсів між підсистемами.

Формулювання мети дослідження

Метою дослідження є розробка багатофакторної моделі оцінки рівня живучості інформаційної системи, яка дозволяє знаходити оптимальний розподіл ресурсів між підсистемами з урахуванням рівня їх взаємозамінності.

Основні завдання дослідження:

1. Вивчити існуючий досвід щодо створення моделей такого роду та методів ресурсної оптимізації.
2. Визначити базові залежності перетворення вхідних ресурсів у вихідний корисний ефект системи – рівень живучості ІС.
3. Визначити адекватні скалярні згортки корисний ефектів підсистем в корисний ефект системи з урахуванням рівня взаємозамінності підсистем.
4. За допомогою створених моделей знайти оптимальні рішення щодо розподілу ресурсів між підсистемами та проаналізувати закономірності зміни цих рішень при зміні значень керуючих факторів.

Аналіз останніх досліджень і публікацій

Концептуальні підходи щодо створення моделей складних багатофакторних систем розглянуті в [4, 5]. Було визначено, що узагальнену модель системи можна представити (рис. 1) у вигляді функціонального блоку $Ef(R)$, на вхід якого поступають ресурси R, а на виході отримуються корисний ефект Ef (далі – ефект). Ресурсами можуть бути: матеріальні ресурси, гроші, час, персонал, рівень кваліфікації, технології, ноу-хау, репутація, рівень керованості тощо. Корисними ефектами можуть бути: дохід, прибуток, обсяги виробництва, нові технології, репутація, частка ринку, якість продукції тощо.

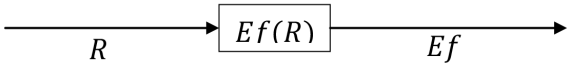


Рис. 1. Узагальнена модель системи

Джерело: розроблено авторами

Виникає питання щодо можливого вигляду базової залежності $Ef(R)$. S-подібний характер залежності корисності інформації від її обсягу представлений в [6]. В динамічних моделях комп'ютерних епідемій використані S-подібні [5] залежності на основі логістичних рівнянь у вигляді звичайних диференціальних рівнянь. Аналогічні рівняння використані в моделі корисного ефекту організації (проєкту) на етапі трансформації [7]. Логістична залежність включає в себе експоненціальні та лінійні ділянки. Інколи ці залежності використовуються, як самостійні залежності корисного ефекту. Наприклад, в [8] функції вразливості в різних умовах є лінійними, експоненціальними або логістичними. Узагальнення підходів щодо використання лінійних, експоненціальних та логістичних моделей зроблене в [4, 5]. В нашому випадку у якості базової залежності ефекту від ресурсів доцільно обирати логістичні моделі в інтегральній формі (у вигляді розв'язку звичайних диференціальних рівнянь для заданих початкових умов).

Система забезпечення живучості складається із підсистем. Кожна із підсистем має власну залежність ефекту від ресурсів. Виникає задача створення моделі ефекту системи. Оскільки функціональне призначення підсистем, а також їх взаємодія та можливості взаємозамінності є досить зрозумілими і наочними, то в якості основного інструменту переходу до моделі системи використаємо скалярне згортання ефектів підсистем. Виникає питання – який вигляд має мати скалярна згортка (далі – згортка).

В [9] для оцінки ефекту атак на інформаційні системи, в моделі живучості [10] та в багатьох інших моделях [4, 5] використані адитивні згортки з ваговими коефіцієнтами. Нажаль, в реальному житті гомогенні системи скоріше виключення ніж правило. Реальні системи реалізують процеси, які доцільно моделювати за допомогою гібридних згорток, наприклад адитивно-мінімізуючих [11], або адитивно-мультиплікативних, які фактично є окремим випадком поліномому Колмогорова-Габора [12–15], якій отримав широке розповсюдження завдяки методу групового урахування аргументу академіка Івахненка А. Г. в задачі множинної регресії [16]. В нашому дослідженні пріоритет варто відати гібридним згорткам, які більш відповідають природі реальних систем забезпечення живучості інформаційних систем.

Моделі дозволяють прогнозувати наслідки управлінських рішень, але більш важливо те, що вони дозволяють знаходити оптимальні рішення. В нашому випадку – це рішення щодо оптимального розподілу ресурсів між підсистемами. Певні підходи щодо розподілу ресурсів між підсистемами при обмеженні на сумарний ресурс системи розглянуті в [4]. Але підхід розглянутий лише для двох окремих випадків адитивної та мультиплікативної згорток. Розширення постановки задачі виконане в [17] при розподілі ресурсу програмістів різних рівнів кваліфікації в задачі створення оптимальної команди розробників. В [8] визначені втрати від комп'ютерних атак при різному ресурсному забезпеченні засобів захисту. Враховані втрати від атак і витрати на захист. Виявлений явний мінімум (оптимум) в залежності рівня безпеки від величини ресурсного забезпечення. В [18] створена аналогічна модель, яка додатково враховує втрати від наслідків атак, витрати на усунення наслідків атак, витрати на відновлення працездатності інформаційної системи, втрати від відволікання ресурсів інформаційної системи щодо виконання основних функцій.

Аналіз існуючих досліджень показав, що у якості базових залежностей ефекту (живучості) ІС від ресурсів доцільно використовувати логістичні залежності. Аналіз показав, що реальним процесам забезпечення живучості в більшому ступені відповідають гібридні скалярні згортки. Але в розглянутих дослідженнях для створення моделей живучості ІС такі згортки використані не були. Аналіз показав, що існує багато досліджень щодо ресурсної оптимізації систем захисту інформації. Нажаль, при цьому задача оптимізації розподілу ресурсів між підсистемами забезпечення живучості ІС не розглядалась.

Викладення основного матеріалу дослідження

Ресурси та корисний ефект. В нашому випадку корисним ефектом є живучість ІС. Ресурсами може бути все, що було перераховано вище. Крім того у якості вхідних ресурсів можуть виступати корисні ефекти підпорядкованих систем. Для спрощення моделі будемо вважати, що вхідними ресурсами є матеріальні ресурси або їх грошовий еквівалент, який дозволяє придбати потрібні засоби та технології забезпечення живучості інформаційних систем.

Базова залежність корисного ефекту від ресурсів

Аналіз існуючих досліджень показав, що у якості базової залежності ефекту від ресурсів в моделі живучості інформаційних систем є логістична залежність, яка для кожної із підсистем має вигляд

$$Ef_i = SL_i\left(R_i\right) = d_i + \frac{a_i - d_i}{1 + \exp\left(-\frac{2}{T_i}\left(R_i - s_i\right)\right)}, \quad i = \overline{1,3}, \tag{1}$$

де i – номер (індекс) підсистеми; d_i, a_i – ординати асимптот (визначають вертикальний масштаб залежності); T_i – постійна величина логістичної кривої (для поточного масштабу залежності визначає найвищу швидкість зростання залежності в зоні стабільного функціонування системи або підсистеми); s_i – абсциса точки симетрії (визначається початковими умовами розвитку процесу, що моделюється).

Логістичні залежності для підсистем (рис. 2) чітко і однозначно визначається наборами параметрів d_i, a_i, T_i, s_i, які мають такі значення (представлено в нотації Матлабу):

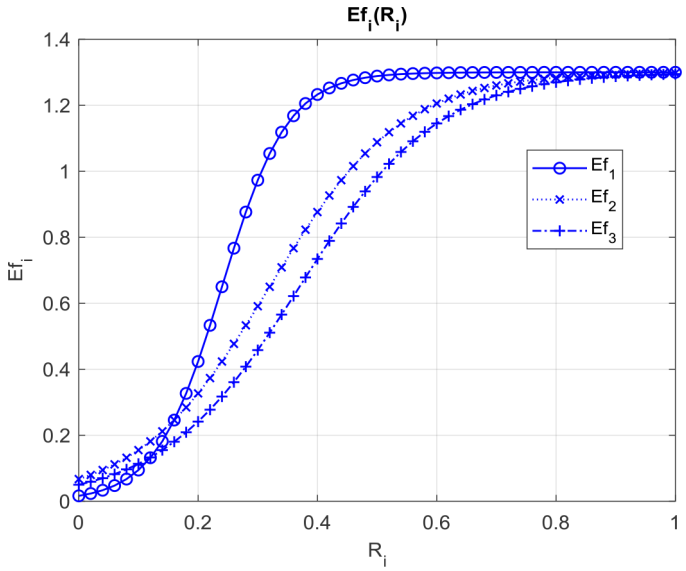


Рис. 2. Базові залежності корисних ефектів підсистем від ресурсів

Джерело: розроблено авторами

```
K= { % d a T s for Survivability of Information Systems
[0 1.3 0.11 0.24] ‘Subsystem 1’
[0 1.3 0.22 0.32] ‘Subsystem 2’
[0 1.3 0.23 0.37] ‘Subsystem 3’
};
```

Згортання ефектів підсистем

До складу системи забезпечення живучості ІС входить 3 складові: підсистема виявлення шкідливих впливів на основі вже відомих сценаріїв (патернів, сигнатур), підсистема виявлення атак «нульового дня» (відносно інформації про вид атаки), підсистема запобігання виявленим вторгненням. Кожна із підсистем в процесі роботи створює власний корисний ефект, який є внеском в загальний ефект системи. Ефект системи не є простою сумою ефектів підсистем. Але в аналізі існуючих досліджень найбільш поширеною є адитивна згортка з ваговим коефіцієнтом β_i, яка моделює повну взаємозамінність підсистем

$$Ef_{sys} = \sum_{i=1}^{3} \beta_i \, SL_i\left(R_i\right) \tag{2}$$

при додатковій умові нормалізації вагових коефіцієнтів

$$\sum_{i=1}^{3} \beta_i = 1. \tag{3}$$

В реальності підсистеми не є повністю взаємозамінними. Тому під знаком суми ми розмістили не ефекти підсистем, які можуть мати довільні значення, а ефекти, які знаходяться за допомогою логістичної залежності з асимптотичними обмеженнями (рис. 2). Це частково знижує взаємозамінність компонент в залежності (2). Хоча в робочій зоні кожної із логістичних залежностей (поряд з точкою симетрії) ефект повної взаємозамінності, нажаль, реалізується.

Властивості неповної взаємозамінності реалізує мультиплікативна згортка

$$Ef_{sys} = \prod_{i=1}^{3} \left(SL_i\left(R_i\right)\right)^{\beta_i}. \tag{4}$$

Нажаль ця залежність не зовсім наочна завдяки суттєвої нелінійності. Крім того ступінь рівня взаємозамінності є, з одного боку постійним, з іншого невідомим. Потрібна згортка з більш прогнозованим рівнем взаємозамінності. Для створення такої згортки розглянемо мінімізуючу згортку

$$Ef_{sys} = \min_{i=1}^{3} \frac{R_i}{\beta_i}. \tag{5}$$

Сенс цієї згортки полягає в тому, що із всіх компонент під знаком згортки, в якості результату, обирається найгірша (найменша). Ця згортка орієнтована на комплекти. Наприклад, якщо у нас 3 комп’тера, 5 комплектів ліцензійного програмного забезпечення та 7 програмістів, то насправді у нас є тільки 3 працездатних програміста, які можуть виконувати певні завдання щодо розробки програмних систем. Ваговий коефіцієнт в цій згортці відповідає кількості одиниць певного ресурсу, що складають один комплект. Мінімізуюча згортка реалізує модель повної невзаємозамінності ресурсів. Якщо будь-який ресурс дорівнює нулю, то ефект всієї системи буде дорівнювати нулю. Мінімізуюча згортка також може бути також представлена у вигляді

$$Ef_{sys} = \min_{i=1}^{3} SL_i\left(\frac{R_i}{\beta_i}\right) \tag{6}$$

або

$$Ef_{sys} = \min_{i=1}^{3} \beta_i \, SL_i\left(R_i\right). \tag{7}$$

В залежності (5) властивість повної невзаємозамінності відноситься безпосередньо до ресурсів, а в залежності (6) до ефектів, що були отримані за допомогою цих ресурсів. В залежності (7) нормування виконується не щодо ресурсів, а щодо ефектів підсистем. Надалі використовуємо залежність (7).

На основі розглянутих базових згорток можна створювати більш складні залежності. Адитивна і мінімізуюча згортки – це два граничних підходи щодо повної взаємозамінності та повної невзаємозамінності підсистем, відповідно. Логічно було би саме ці згортки об’єднати в гібридній згортці. При цьому варто додати керуючий параметр щодо рівня взаємозамінності підсистем. Наприклад, на основі поліному Колмогорова-Габора подібну гібридну згортку автори використовували в попередніх своїх дослідженнях.

$$Ef_{sys} = k_{Add} \sum_{i=1}^{3} \beta_i \, SL_i\left(R_i\right) + \left(1 - k_{Add}\right) \prod_{i=1}^{3} \left(SL_i\left(R_i\right)\right)^{\beta_i}. \tag{8}$$

Тут k_{Add} – коефіцієнт, що визначає рівень взаємозамінності підсистем. Якщо $k_{Add} = 1$ – підсистеми повністю взаємозамінні, якщо $k_{Add} = 0$ – підсистеми невзаємозамінність на рівні мультиплікативної згортки. Але ж ми визначали вище, що мультиплікативні системи мало того, що не повністю взаємозамінні, ще й не зрозуміло, якій саме рівень невзаємозамінності вони реалізують. Тому модифікуємо залежність (8) таким чином

$$Ef_{sys} = k_{Add} \sum_{i=1}^{3} \beta_i \, SL_i\left(R_i\right) + \left(1 - k_{Add}\right) \min_{i=1}^{3} \beta_i \, SL_i\left(R_i\right). \tag{9}$$

Тепер ми отримали можливість враховувати рівень взаємозамінності підсистем в повному діапазоні від 0 до 100 % відповідно до умов функціонування системи.

Оптимальний ефект

Критерій оптимальності – максимум ефекту.

$$Ef_{opt} = \max_{R_1, R_2, R_3} \left(Ef_{sys} \right). \tag{10}$$

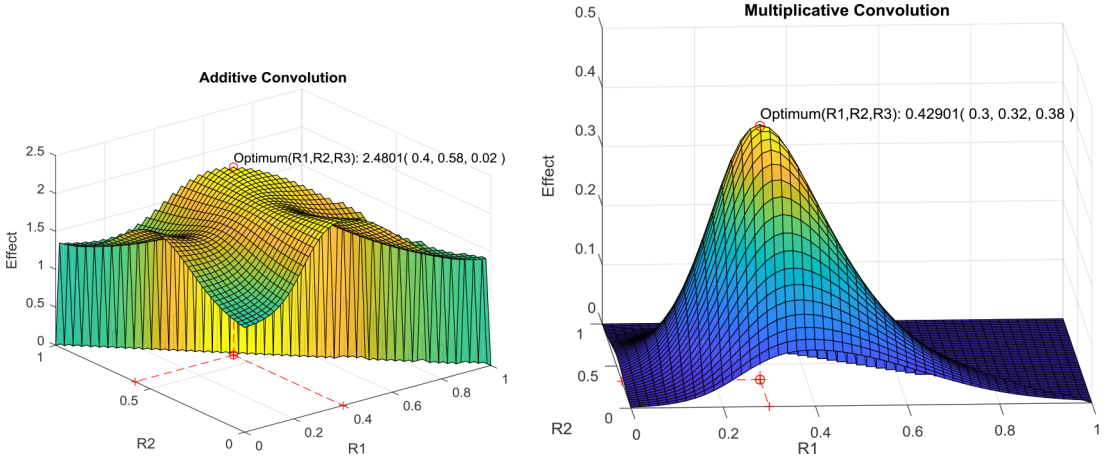
Оскільки сума всіх ресурсів підсистем є фіксована величина, то

$$R_3 = R_{sum} - R_1 - R_2 \tag{11}$$

це дозволяє зменшити кількість змінних, які треба варіювати при пошуку оптимального рішення

$$\left(R_{1opt}, R_{2opt}\right) = \arg\max_{R_1, R_2} \left(Ef_{sys} \right). \tag{12}$$

Поверхні пошуку оптимальних рішень $Ef_{sys}(R_1, R_2)$ для базових згорток мають вигляд (рис. 3, 4).



а *б*

Рис. 3. Поверхні пошуку оптимального рішення $Ef_{sys}(R_1, R_2)$ для базових згорток: а) адитивна згортка, б) мультиплікативна згортка

Джерело: розроблено авторами

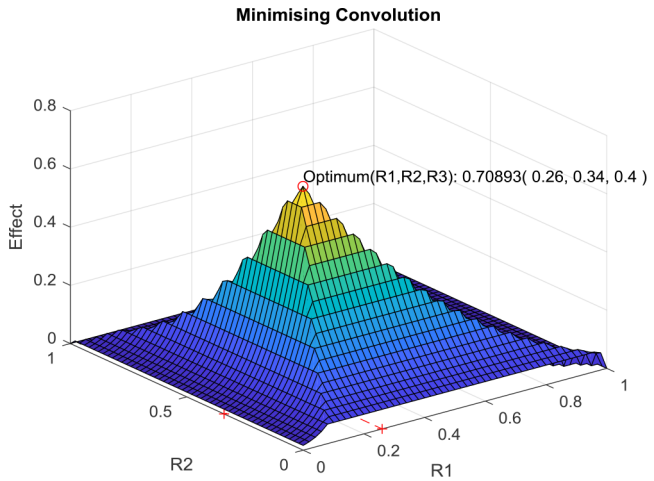
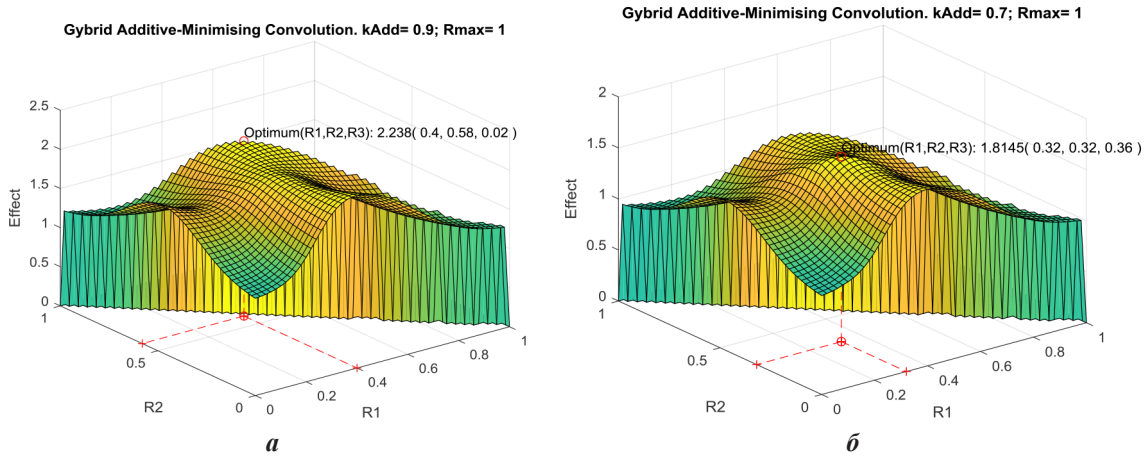


Рис. 4. Поверхня пошуку оптимального рішення $Ef_{sys}(R_1, R_2)$ для мінімізуючої згортки

Джерело: розроблено авторами

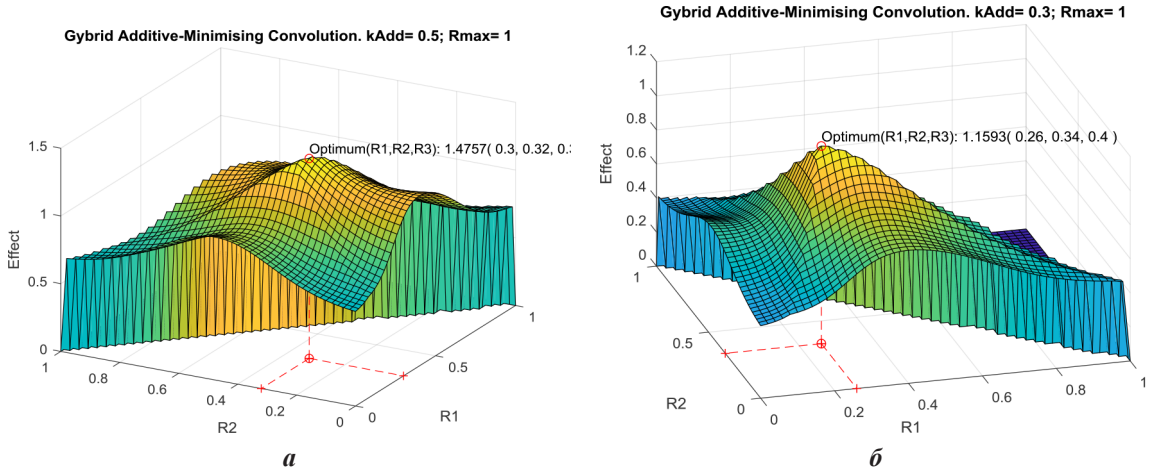
Поверхні пошуку оптимальних рішень $Ef_{sys}(R_1, R_2)$ для гібридних згорток представлені на рис. 5, 6. Адаптація моделі під постановку задачі виконується за допомогою коефіцієнту k_{Add} (9).



а *б*

Рис. 5. Поверхні пошуку оптимального рішення $Ef_{sys}(R_1, R_2)$ для гібридної адитивно-мінімізуючої згортки: *а*) $k_{Add} = 0.9$, *б*) $k_{Add} = 0.7$

Джерело: розроблено авторами



а *б*

Рис. 6. Поверхні пошуку оптимального рішення $Ef_{sys}(R_1, R_2)$ для гібридної адитивно-мінімізуючої згортки: а) $k_{Add} = 0.5$, б) $k_{Add} = 0.3$

Джерело: розроблено авторами

При великих значеннях коефіцієнту $k_{Add} = 0.9$ (рис. 5, *а*) згортка в моделі реалізує високий рівень взаємозамінності підсистем. Поверхня пошуку оптимального рішення наближена до вигляду поверхні для адитивної згортки. По мірі зменшення значення коефіцієнту починають превалювати властивості мінімізуючої згортки (рис. 5, *б*, рис. 6). Для згортки з високим рівнем взаємозамінності підсистем, оптимум знаходиться на межі області припустимих значень. При зменшенні рівня взаємозамінності оптимум переміщується в середину області припустимих значень.

Оптимальний розподіл ресурсів

Оптимальні рішення знаходились методом повного прямого перебору, що дозволило впевнено знаходити глобальні максимуми серед всього набору максимумів поверхні пошуку оптимальних рішень $Ef_{sys}(R_1, R_2)$. Знайдені оптимальні рішення Ef_{opt} досягаються при певних значеннях вхідних ресурсів підсистем (R_{1opt}, R_{2opt}). Ці значення змінюються при зміні величини k_{Add} та величини сумарного ресурсу системи R_{sum}. Знання розташування траєкторії руху оптимумів (Ef_{opt}, R_{1opt}, R_{2opt}) при зміні k_{Add}, R_{sum} (рис. 7, 8) дозволяє не лише знаходити оптимальні рішення для заданих умов, але й свідомо змінювати ці умови за допомогою параметрів k_{Add}, R_{sum}.

Висновки

В роботі проаналізовані можливості створення моделей живучості інформаційної системи для різних рівнів взаємозамінності підсистем.

В складі системи забезпечення живучості інформаційної системи проаналізовані 3 підсистеми: підсистема виявлення шкідливих впливів на основі вже відомих сценаріїв (патернів, сигнатур), підсистема виявлення атак «нульового дня», підсистема запобігання виявленим вторгненням.

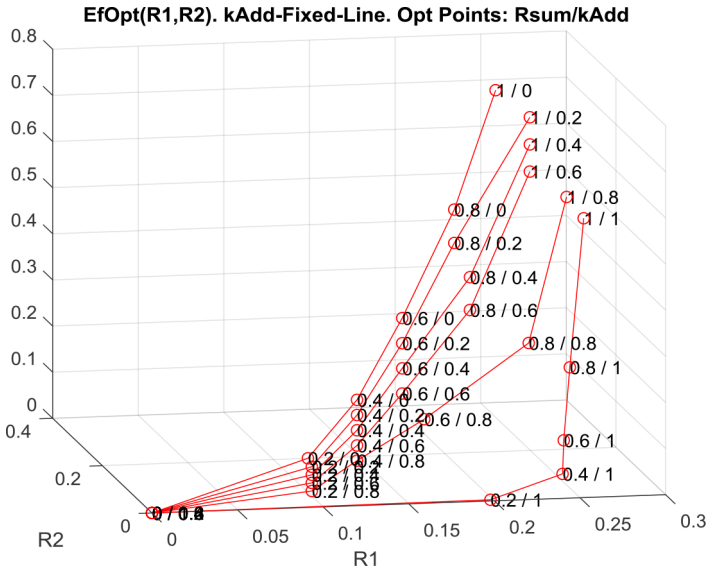


Рис. 7. Траєкторії оптимальних точок (Ef_{opt}, R_{1opt}, R_{2opt}) для k_{Add} зафіксованих в діапазоні 0 до 1 при зміні R_{sum} від 0 до 1

Джерело: розроблено авторами

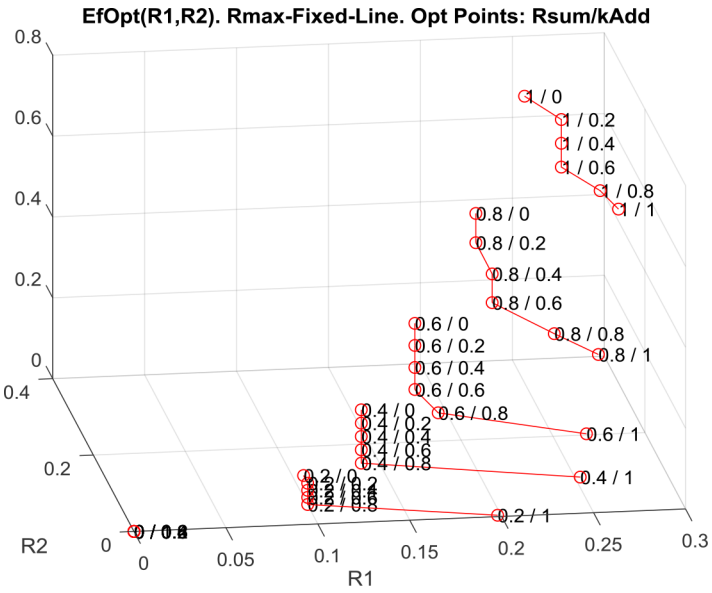


Рис. 8. Траєкторії оптимальних точок (Ef_{opt}, R_{1opt}, R_{2opt}) для R_{sum} зафіксованих в діапазоні 0 до 1 при зміні k_{Add} від 0 до 1

Джерело: розроблено авторами

Встановлено, що залежність корисного ефекту підсистеми від вхідних ресурсів доцільно створювати на основі логістичних залежностей в інтегральній формі. Визначено, що логістичні залежності однозначно і чітко задаються такими параметрами: ординати нижньої та верхньої асимптот, постійна величина логістичної залежності, що визначає швидкість максимального зростання логістичної залежності, та абсциса точки симетрії, що залежить від початкових умов.

Визначено, що для об'єднання ефектів підсистем в загальний ефект системи доцільно використовувати скалярні згортки. В роботі запропонована адитивно-мінімізуюча згортка, яка дозволяє налаштовувати модель на різні рівні взаємозамінності підсистем у всьому діапазоні можливих значень від 0 до 100 %. Рівень взаємозамінності підсистем встановлюється за допомогою відповідного коефіцієнта в діапазоні від 0 до 1.

Сформульовані залежності для пошуку оптимального розподілу ресурсів між підсистемами для забезпечення максимуму корисного ефекту системи. Пошук оптимального рішення виконувався методом повного прямого перебору, що дозволило уникати локальних оптимумів і впевнено знаходити глобальний максимум.

Було виявлено, що оптимальна точка змінює своє положення в залежності від значень параметрів: коефіцієнта рівня взаємозамінності та сумарного ресурсу системи. Були побудовані траєкторії руху оптимальних рішень, для різних значень параметрів, що дозволяє не тільки легко знаходити оптимальне рішення для поточних параметрів, але й знаходити керуючі параметри, які б дозволили покращити ефект системи з урахуванням можливої зміни сценаріїв роботи системи.

Напрями подальших досліджень: розширення переліку факторів, що беруться до уваги при пошуку оптимального рішення.

Список використаної літератури

1. Shailendra Upadhyay. (2024). Information Security Spending: What Does the Future Hold? Website Gartner. November 27, 2024. Available at: https://www.gartner.com/en/articles/information-security?utm_source=chatgpt.com
2. Westmark, V.R.. (2004). A Definition for information system survivability. 37. 10 pp.. 10.1109/HICSS.2004.1265710
3. Popereshnyak S., Novikov Y., Zhdanova Y. (2024). Cryptographic system security approaches by monitoring the random numbers generation. *CEUR Workshop Proceedings*. 2024. Vol. 3826. P. 301–309. URL: https://ceur-ws.org/Vol-3826/short21.pdf (дата звернення: 18.09.2025).
4. Шевченко В. Л. (2011). Оптимізаційне моделювання в стратегічному плануванні. Київ : ЦВСД НУОУ, 283.
5. Шевченко В. Л., Нестеренко О. В., Нетесін І. Є., Шевченко А. В., Поліщук В. Б. (2019). Прогнозне моделювання комп'ютерних вірусних епідемій. Київ : УкрНЦ РІТ, 152. http://e.ieu.edu.ua/handle/123456789/647
6. Khmelevskoy R., Laptiev S., Laptieva T. (2021). Improving the Method of Assessing the Information Security of Computer Systems from Malicious Software. International Journal of Science and Engineering Investigations (IJSEI) Denmark, 113 (10), 43–48. https://www.ijsei.com/papers/ijsei-1011321-07.pdf
7. Syvytsky Y., Shevchenko V. (2024). Computer Simulation Model of the Organization at the Stage of Transformation for the Purpose of Adaptation to New Projects. 14th International Scientific and Practical Conference from Programming UkrPROG'2024. May 14–15. Kyiv. Ukraine, 421–433.
8. Рабчун Д. І. (2015). Оцінка ефективності інформаційної безпеки з урахуванням економічних показників. Сучасний захист інформації, 4, 91–97. https://journals.dut.edu.ua/index.php/dataprotect/article/view/429/397
9. Yevseiev S., Ponomarenko V., Laptiev O., Milov O. and others. (2021). Synergy of building cybersecurity systems: monograph. Kharkiv: PC TECHNOLOGY CENTER, 188. https://www.scopus.com/record/display.uri?eid=2-s2.0-85119654284&origin=resultslist https://www.researchgate.net/publication/352013398_Synergy_of_building_cybersecurity_systems_Monograph
10. Syvachenko I., Bakaiev O., Shevchenko V. (2024). Scalarization of the vector criterion of information system survivability based on information security indicators. Proceedings of the Cybersecurity Providing in Information and Telecommunication Systems II (CPITS-II 2024), Kyiv, Ukraine, October 26, 2024, CEUR Workshop Proceedings, ISSN 1613-0073, 2024, Vol-3826, 294–300. https://ceur-ws.org/Vol-3826/short20.pdf, Scopus, dblp
11. Emmerich M., Deutz A. (2024). Multicriteria Optimization and Decision Making. Principles, Algorithms and Case Studies : Lecture Notes. MSc Course 2012-2023, LIACS, Leiden University. The Netherlands. Updated and Revised Edition, June 2024, 106. https://arxiv.org/pdf/2407.00359v2
12. Marateb H., Norouzirad M., Tavakolian K., Aminorroaya F. O, Mohebbian M., Mañanas M. Á., Lafuente S. R., Sami R., Mansourian M. (2023). Predicting COVID-19 Hospital Stays with Kolmogorov–Gabor Polynomials: Charting the Future of Care. Information, 14(11), 590. https://doi.org/10.3390/info14110590
13. Qian N., Gao J., Li Z., Yan Z., Feng Y., Yan Z., Yang L. (2024). Bridging the Terrestrial Water Storage Anomalies between the GRACE/GRACE-FO Gap Using BEAST + GMDH Algorithm. Remote Sensing, 16(19), 3693. https://doi.org/10.3390/rs16193693
14. Mohammad A., Belayneh M. (2024). Field Telemetry Drilling Dataset Modeling with Multivariable Regression, Group Method Data Handling, Artificial Neural Network, and the Proposed Group-Method-Data-Handling-Featured Artificial Neural Network. Applied Sciences, 14(6), 2273. https://doi.org/10.3390/app14062273
15. Achite M., Katipoğlu O. M., Kartal V., Sarıgöl M., Jehanzaib M., Gül E. (2025). Advanced Soft Computing Techniques for Monthly Streamflow Prediction in Seasonal Rivers. Atmosphere, 16(1), 106. https://doi.org/10.3390/atmos16010106
16. Ivakhnenko A. G. (1971). Polynomial theory of complex systems. IEEE Trans. Syst. Man Cybern, 4, 364–378. DOI: 10.1109/TSMC.1971.4308320 https://ieeexplore.ieee.org/document/4308320
17. Fedoriienko V. (2021). Model of the special software development and modification based on account factors of complexity, workability, qualification of programmers and resource limitations. Political Science and Security Studies Journal. Przeworsk, 2 (3), 94–104. URL: http://psssj.eu/index.php/ojsdata/issue/view/4 https://doi.org/10.5281/zenodo.5544382 https://psssj.eu/index.php/ojsdata/article/view/69
18. Shevchenko A., Fedorenko R., Shmorhun Y., Hrebennikov A., Shevchenko V. (2019). Designing of Functionally Stable Information Systems Optimal for a Minimum of Losses. 15th International Conference on the Experience of Designing and Application of CAD Systems (CADSM), Polyana-Svalyava (Zakarpattya), IEEE Xplore, 36–40. DOI: 10.1109/CADSM.2019.8779299 https://ieeexplore.ieee.org/document/8779299

References

1. Shailendra Upadhyay. (2024). Information Security Spending: What Does the Future Hold? Website Gartner. November 27, 2024. Available at: https://www.gartner.com/en/articles/information-security?utm_source=chatgpt.com

2. Westmark, V. R.. (2004). A Definition for information system survivability. 37. 10 pp.. 10.1109/HICSS.2004.1265710

3. Popereshnyak S., Novikov Y., Zhdanova Y. (2024). Cryptographic system security approaches by monitoring the random numbers generation. *CEUR Workshop Proceedings*. 2024. Vol. 3826. P. 301–309. URL: https://ceur-ws.org/Vol-3826/short21.pdf (дата звернення: 18.09.2025).

4. Shevchenko V. L. (2011). Optimization modeling in strategic planning. Kyiv : Military and Strategic Research Centre of National Defence University of Ukraine, 283.

5. Shevchenko V. L., Nesterenko O. V, Netesin I. E., Shevchenko A. V., Polishchuk V. B. (2019). Predictive modeling of computer virus epidemics. Kyiv : UkrNC RIT, 152. http://e.ieu.edu.ua/handle/123456789/647

6. Khmelevskoy R., Laptiev S., Laptieva T. (2021). Improving the Method of Assessing the Information Security of Computer Systems from Malicious Software. International Journal of Science and Engineering Investigations (IJSEI) Denmark, 113 (10), 43–48. https://www.ijsei.com/papers/ijsei-1011321-07.pdf

7. Syvytsky Y., Shevchenko V. (2024). Computer Simulation Model of the Organization at the Stage of Transformation for the Purpose of Adaptation to New Projects. 14th International Scientific and Practical Conference from Programming UkrPROG'2024. May 14–15. Kyiv. Ukraine, 421–433.

8. Rabchun D. I. (2015). Assessment of the effectiveness of information security taking into account economic indicators. Modern information protection, 4, 91–97. https://journals.dut.edu.ua/index.php/dataprotect/article/view/429/397

9. Yevseiev S., Ponomarenko V., Laptiev O., Milov O. and others. (2021). Synergy of building cybersecurity systems: monograph. Kharkiv: PC TECHNOLOGY CENTER, 188. https://www.scopus.com/record/display.uri?eid=2-s2.0-85119654284&origin=resultslist https://www.researchgate.net/publication/352013398_Synergy_of_building_cybersecurity_systems_Monograph

10. Syvachenko I., Bakaiev O., Shevchenko V. (2024). Scalarization of the vector criterion of information system survivability based on information security indicators. Proceedings of the Cybersecurity Providing in Information and Telecommunication Systems II (CPITS-II 2024), Kyiv, Ukraine, October 26, 2024, CEUR Workshop Proceedings, ISSN 1613-0073, 2024, Vol-3826, 294–300. https://ceur-ws.org/Vol-3826/short20.pdf, Scopus, dblp

11. Emmerich M., Deutz A. (2024). Multicriteria Optimization and Decision Making. Principles, Algorithms and Case Studies : Lecture Notes. MSc Course 2012-2023, LIACS, Leiden University. The Netherlands. Updated and Revised Edition, June 2024, 106. https://arxiv.org/pdf/2407.00359v2

12. Marateb H., Norouzirad M., Tavakolian K., Aminorroaya F. O, Mohebbian M., Mañanas M. Á., Lafuente S. R., Sami R., Mansourian M. (2023). Predicting COVID-19 Hospital Stays with Kolmogorov–Gabor Polynomials: Charting the Future of Care. Information, 14(11), 590. https://doi.org/10.3390/info14110590

13. Qian N., Gao J., Li Z., Yan Z., Feng Y., Yan Z., Yang L. (2024). Bridging the Terrestrial Water Storage Anomalies between the GRACE/GRACE-FO Gap Using BEAST + GMDH Algorithm. Remote Sensing, 16(19), 3693. https://doi.org/10.3390/rs16193693

14. Mohammad A., Belayneh M. (2024). Field Telemetry Drilling Dataset Modeling with Multivariable Regression, Group Method Data Handling, Artificial Neural Network, and the Proposed Group-Method-Data-Handling-Featured Artificial Neural Network. Applied Sciences, 14(6), 2273. https://doi.org/10.3390/app14062273

15. Achite M., Katipoğlu O. M., Kartal V., Sarıgöl M., Jehanzaib M., Gül E. (2025). Advanced Soft Computing Techniques for Monthly Streamflow Prediction in Seasonal Rivers. Atmosphere, 16(1), 106. https://doi.org/10.3390/atmos16010106

16. Ivakhnenko A.G. (1971). Polynomial theory of complex systems. IEEE Trans. Syst. Man Cybern, 4, 364–378. DOI: 10.1109/TSMC.1971.4308320 https://ieeexplore.ieee.org/document/4308320

17. Fedoriienko V. (2021). Model of the special software development and modification based on account factors of complexity, workability, qualification of programmers and resource limitations. Political Science and Security Studies Journal. Przeworsk, 2 (3), 94–104. URL: http://psssj.eu/index.php/ojsdata/issue/view/4 https://doi.org/10.5281/zenodo.5544382 https://psssj.eu/index.php/ojsdata/article/view/69

18. Shevchenko A., Fedorenko R., Shmorhun Y., Hrebennikov A., Shevchenko V. (2019). Designing of Functionally Stable Information Systems Optimal for a Minimum of Losses. 15th International Conference on the Experience of Designing and Application of CAD Systems (CADSM), Polyana-Svalyava (Zakarpattya), IEEE Xplore, 36–40. DOI: 10.1109/CADSM.2019.8779299 https://ieeexplore.ieee.org/document/8779299

Дата першого надходження рукопису до видання: 18.09.2025
Дата прийнятого до друку рукопису після рецензування: 15.10.2025
Дата публікації

UDC 004.932.2:004.8:004.58 DOI

O. Yu. BARKOVSKA
Candidate of Technical Sciences, Associate Professor,
Associate Professor at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0000-0001-7496-4353

O. V. NECHITAILO
Postgraduate Student at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0009-0005-7299-2875

L. D. BUHAROVA
Postgraduate Student at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0009-0003-0635-8695

EMBEDDED VOICE ASSISTANCE MODULE FOR THE VISUALLY IMPAIRED BASED ON VISUAL SCENE ANALYSIS

The article discusses a hardware module designed to assist visually impaired individuals at public transport stops. The relevance of this research is determined by the increasing number of people with visual impairments in Ukraine, caused both by global trends and by the consequences of military actions that have led to eye injuries among both military personnel and civilians. Since urban infrastructure is predominantly oriented toward visual perception, visually impaired individuals face significant difficulties in orientation, especially at public transport stops. The use of an autonomous device that does not depend on a smartphone or Internet access is of great importance for ensuring mobility and safety in urban environments.

The purpose of the study is to develop and analyze the effectiveness of a hardware-software complex based on Raspberry Pi 5, which provides automatic recognition of public transport types (bus, trolleybus, tram) and their route numbers, followed by real-time voice output of the results. The hardware part includes a Raspberry Pi 5 single-board computer with active cooling, a Logitech Web Camera Carl Zeiss Tessar 2.0/3.7 2MP connected via USB 3.0, a Logitech G Pro X external sound card for audio output through headphones, and a portable power bank to ensure autonomous operation.

Experimental studies have confirmed high detection accuracy of public transport in daytime conditions and acceptable performance on the embedded platform. At the same time, under low-visibility conditions (twilight, fog, image blurring), a decrease in classification accuracy of transport types and recognition of route numbers was observed. It has been determined that further system development may include expanding the dataset with images collected under real weather conditions, improving video preprocessing algorithms (low-light enhancement and denoising), as well as integrating with mobile applications and GPS navigation. The obtained results demonstrate the practical applicability of the developed solution and its potential for scaling to improve the inclusiveness and mobility of visually impaired individuals.

Key words: *Raspberry Pi, YOLOv8, Tesseract, OpenCV, OCR, pyttsx3, intelligent module, visually impaired, hardware implementation, public transport.*

О. Ю. БАРКОВСЬКА
кандидат технічних наук, доцент,
доцент кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0000-0001-7496-4353

О. В. НЕЧІТАЙЛО
магістрант кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0009-0005-7299-2875

Л. Д. БУХАРОВА
магістрант кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0009-0003-0635-8695

© Barkovska O. Yu., Nechitailo O. V., Buharova L. D., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

АПАРАТНИЙ МОДУЛЬ ГОЛОСОВОГО СУПРОВОДУ ДЛЯ ОСІБ З ПОРУШЕННЯМ ЗОРУ НА ОСНОВІ ВІЗУАЛЬНОГО АНАЛІЗУ ПАРАМЕТРІВ ОТОЧУЮЧОГО СЕРЕДОВИЩА

У статті розглянуто апаратний модуль супроводу осіб із вадами зору на зупинках громадського транспорту. Актуальність теми зумовлена зростанням кількості громадян з порушеннями зору в Україні, що обумовлено як загальносвітовими тенденціями, так і наслідками воєнних дій, які призводять до травмування органів зору у військових та цивільних. Міська інфраструктура здебільшого зорієнтована на візуальне сприйняття, тому слабозорі особи стикаються зі значними труднощами при орієнтації, особливо на зупинках громадського транспорту. Використання автономного пристрою, який не залежить від смартфону чи мережі Інтернет, має важливе значення для забезпечення мобільності та безпеки таких осіб у міському середовищі.

Метою дослідження є створення та аналіз ефективності програмно-апаратного комплексу на базі Raspberry Pi 5, що забезпечує автоматичне розпізнавання типу громадського транспорту (автобус, тролейбус, трамвай) та його маршрутного номера з подальшим голосовим озвученням результату в режимі реального часу. Апаратна частина включає одноплатний комп'ютер Raspberry Pi 5 з системою активного охолодження, камеру Logitech Web Camera Carl Zeiss Tessar 2.0/3.7 2MP, підключену через USB 3.0, зовнішню звукову карту Logitech G Pro X для відтворення аудіо через навушники та портативний акумулятор, що забезпечує автономність роботи.

Експериментальні дослідження підтвердили високу точність детектування транспортних засобів у денних умовах та прийнятну швидкодію системи на вбудованій платформі. Водночас в умовах низької видимості (сутінки, туман, ефект замилення) спостерігається зниження точності класифікації типу транспорту та визначення маршрутного номера. Визначено, що подальший розвиток системи може відбуватися шляхом розширення датасету зображеннями, отриманими у реальних погодних умовах, вдосконалення алгоритмів попередньої обробки відео (застосування методів low-light enhancement та denoising), а також інтеграції з мобільними додатками та GPS-навігацією. Отримані результати демонструють практичну придатність розробленого рішення та його потенціал для масштабування з метою підвищення рівня інклюзії та мобільності осіб із порушеннями зору.

Ключові слова: *Raspberry Pi, YOLOv8, Tesseract, OpenCV, OCR, pyttsx3, інтелектуальний модуль, слабозорі, апаратна реалізація, громадський транспорт.*

Formulation of the problem

The relevance of the problem of rehabilitation and social integration of people with visual impairments in Ukraine is determined both by global trends in the growing number of people with vision disorders and by the direct consequences of military actions, which lead to an increase in eye injuries among both military personnel and civilians [1]. According to official data, the number of people with severe visual impairments in Ukraine is on the rise, creating an urgent need for the development of an inclusive environment. However, modern urban infrastructure is primarily oriented toward visual perception of information: road signs, traffic lights, billboards, and navigation indicators. Because of this, people with low vision face significant difficulties in spatial orientation, especially at public transport stops, where it is important not only to identify the vehicle itself but also to obtain information about its route number and direction of movement.

To address these problems, a few technical solutions already exist worldwide, including mobile navigation applications, sensory systems for spatial orientation, as well as traditional means of assistance such as guide dogs or specialized "smart" canes. Mobile applications (for example, BlindSquare) use GPS navigation and voice guidance; however, their operation depends on stable Internet access and a functioning smartphone. Sensory systems based on ultrasonic technologies (WeWALK Smart Cane 2, Sonic Pathfinder) allow the detection of obstacles but cannot provide users with semantic information, such as the type of vehicle or route number. The most advanced camera-oriented solutions (e.g., OrCam MyEye 3 Pro) demonstrate high accuracy in object and text recognition but remain extremely expensive and inaccessible to the wider public.

To generalize and systematize the available technologies in the field of assistance for the visually impaired, Table 1 presents a comparison of modern technical solutions. Special attention is paid to hardware-based approaches, as they are the most promising for the development of autonomous intelligent orientation systems.

This paper proposes a fundamentally different approach-an autonomous software-hardware module built on the single-board computer Raspberry Pi 5. The system integrates a camera for video data acquisition, an external sound card with headphones for voice output, and portable power supply to ensure full autonomy. By employing modern deep learning algorithms (YOLOv8 for vehicle detection and Tesseract OCR for route number recognition), the module can provide real-time voice output through a local speech synthesizer (pyttsx3). A key advantage of this solution is that it does not require an Internet connection or smartphone, making it independent and suitable for everyday use in Ukrainian cities.

Analysis of the latest research and publications

The issue of applying artificial intelligence and computer vision technologies to support people with visual impairments has been attracting increasing attention among researchers in recent years. Contemporary studies emphasize the need to develop systems capable of providing spatial orientation, object recognition, and text information retrieval in urban environments in real time.

Table 1

Aids and support solutions for the blind: advantages and disadvantages of each one

№	Solution	Advantages	Disadvantages
1	WeWALK Smart Cane 2	Obstacle detection from above; Voice AI assistant; Navigation; White cane and accompanying functions in one device	For full functionality, must have access to the network
2	Sonic Pathfinder	Obstacle detection from above; A boost to future technology development	Load on the head; Secondary
3	BlindSquare	Does not require additional devices; Navigation; Voice guidance	Requires network access; Does not allow you to use other smartphone applications
4	OrCam MyEye 3 Pro	Works without internet access; Recognizes a large number of objects; Voice AI assistant	Price $4000; Doesn't help with navigation

For example, O. Yu. Barkovska and V. Seredniy proposed a universal model of an intelligent assistance system that enables automatic detection of static and dynamic obstacles, prediction of their trajectories, and the use of a voice interface for user interaction [2]. Such an approach makes it possible to integrate various sensors and modules into a single adaptive system.

In a study by an international group of authors, the MR.NAVI system was introduced, combining computer vision and natural language processing to explain situations, avoid collisions, and assist with public transport navigation. The authors emphasize that integrating multimodal interaction methods significantly increases the effectiveness of orientation for people with low vision [3].

Another approach was demonstrated in the work of Shah et al., where a Smart Real-Time Object Detection system was developed based on the YOLOv8 algorithm for object recognition and OCR technologies for text reading. Importantly, the results are immediately voiced to the user, making the solution convenient for practical use [4].

A significant contribution was made by the VISA system described by Mahmud and colleagues. It integrates augmented reality methods, deep neural networks for object recognition, scene depth estimation, and a voice interface, ensuring natural interaction with the system [5].

It is also worth mentioning a systematic review of modern technologies for visually impaired individuals conducted by Rafiq and co-authors. The review analyzed over 50 scientific works, including solutions based on mobile applications, tactile devices, and neural network navigation algorithms. The authors highlight the necessity of localizing such systems and adapting them to specific environmental conditions [6].

Thus, the literature review demonstrates that the application of artificial intelligence and computer vision for supporting visually impaired people is actively developing; however, in the Ukrainian context, this problem still requires further scientific and applied research.

Formulation of the purpose of the research

Given the relevance of the problem of the growing number of people with visual impairments in Ukraine, particularly as a result of military actions, and considering the limitations of existing solutions in providing orientation at public transport stops, this study aims to develop a hardware- and software-oriented module to assist visually impaired individuals. The main task is to ensure the compactness, autonomy, and ease of use of a device that integrates a camera for video data collection, an audio interface for voice output, and portable power supply, enabling its operation in urban environments without dependence on external infrastructure.

It should be emphasized that the implementation of such a hardware solution entails several constraints and challenges:

– ensuring water resistance of the case and durability under external conditions (rain, snow, low and high temperatures);

– limited autonomous operation time, which depends on the capacity and characteristics of the external battery (power bank);

– the need to select peripheral devices (USB camera, sound card, headphones) that are simultaneously compatible with Raspberry Pi 5 and do not create excessive power consumption;

– technical complexity of integrating cooling to ensure stable processor performance during long-term operation.

To ensure the operability of the hardware system, the following components are required:

– single-board computer Raspberry Pi 5 with a cooling system;
– USB camera (e.g., Logitech Web Camera Carl Zeiss Tessar) for video stream capture;
– external USB sound card and headphones for voice playback;
– portable power bank with PD (Power Delivery) support and an output power of at least 25–30 W;
– lightweight case or mount that allows the system to be carried on the body (waist, chest) while simultaneously protecting electronic components.

The implementation process requires the following tasks to be performed:

– selection and integration of peripherals that meet mobility and low power consumption requirements;
– development of guidelines for module placement on the user;
– configuration of automatic software launch after power-on to simplify usage;
– testing the module on different types of public transport;
– analysis of testing results when capturing frames of public transport from various distances.

Future development of the hardware component may include the integration of a built-in battery instead of an external power bank, the design of a compact case with enhanced water and dust resistance (IP54–IP67), as well as the creation of a specialized single-board system with an integrated graphics processor, which would optimize system performance and reduce energy consumption.

Research Results and Discussions

At the initial stage of developing the hardware module, a general algorithm of the system's operation was designed. Considering the technology of assisting visually impaired individuals in orientation at public transport stops, it should be taken into account that such a tool must include a hardware component. Therefore, the scheme presented below incorporates both software and hardware components (Figure 1).

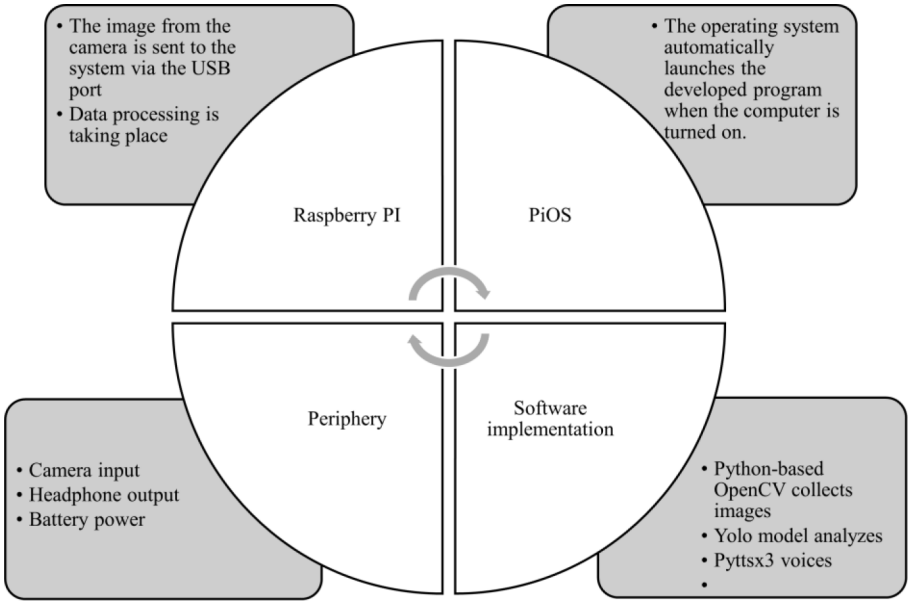


Fig. 1. Abstract of detecting the work cycle of a software-hardware solution

To formalize the system's operational logic, an IDEF0 diagram was used, which clearly illustrates the process structure and the interaction between its components. The top level of the model demonstrates the main function: the detection and voice announcement of approaching public transport, which is implemented through the sequential operation of several subsystems.

The system input includes the video stream from the camera as well as environmental factors (lighting, weather conditions). These data are processed by the computer vision module (YOLOv8), which performs transport object detection. Next, the processing block extracts from the frame the area where the route number is expected and sends it to the Optical Character Recognition (OCR, Tesseract) module. If the route number is successfully recognized, the system generates a text message and transmits it to the speech module (pyttsx3), which produces a voice signal. In case of recognition failure, only the type of transport (e.g., "bus") is announced.

An additional element of the architecture is the environmental filtering module, which applies pre-processing to frames (for example, contrast enhancement in low-light conditions), thus increasing the accuracy of subsequent processing stages.

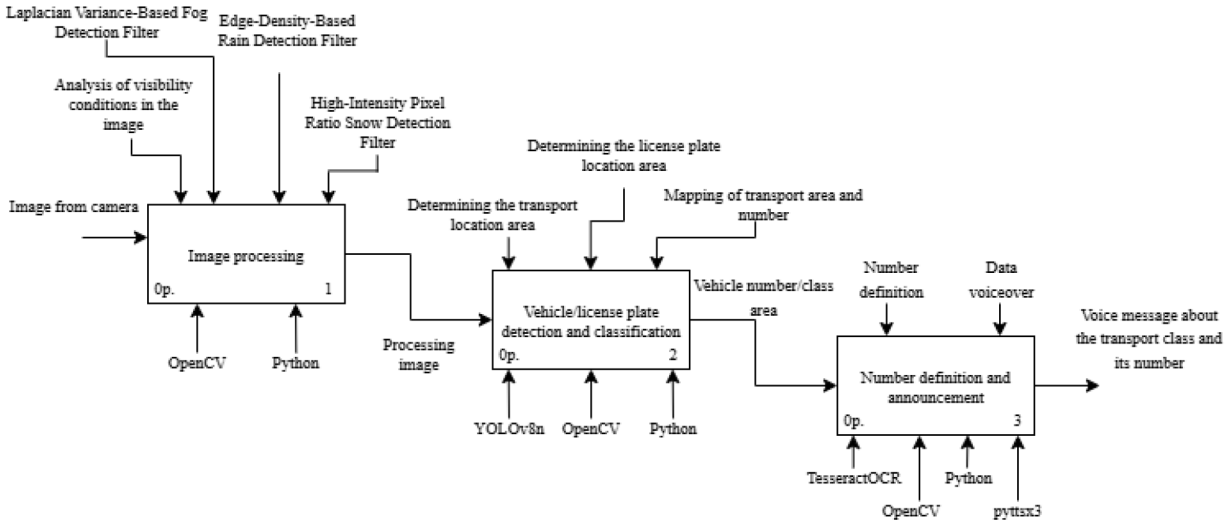


Fig. 2. IDEF0 diagram of the general architecture of the software and hardware complex

For the implementation of the hardware part of the system for assisting visually impaired individuals at public transport stops, the following hardware stack was used:

– Raspberry Pi 5 with a cooling system installed (Figure 3);
– Logitech Web Camera USB camera (Figure 4);
– Logitech G Pro X sound card with headphones (Figure 5);
– Baseus PD 20W 10000mAh power bank (Figure 6).

The central computing module of the system is the Raspberry Pi 5, equipped with a quad-core ARM Cortex-A76 processor, 8 GB of RAM, and a passive-active cooling system with a fan. The device is preinstalled with the Raspberry

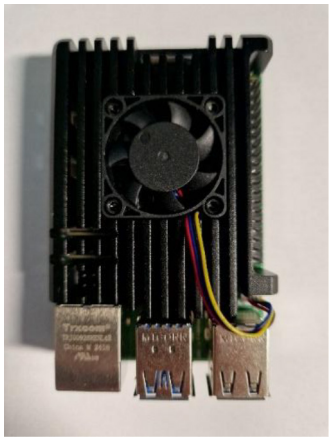


Fig. 3. Raspberry Pi 5 with cooling system installed

Fig. 4. USB-Camera Logitech Web Camera

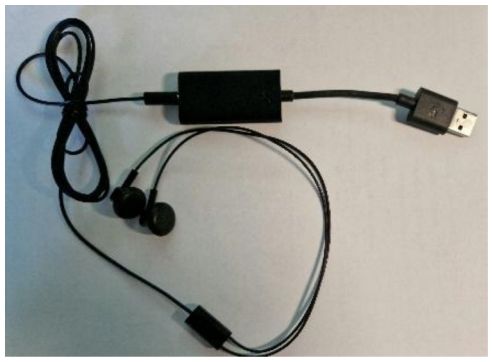


Fig. 5. Logitech G Pro X sound card with headphone

Fig. 6. Power source

Pi OS Lite operating system, while the software is stored on an internal microSD card. For stable operation under high loads, a power supply with at least 25 W output is required (Figure 3).

Video capture is performed using an external Logitech Web Camera with a Carl Zeiss Tessar 2.0/3.7 lens, providing 2 MP resolution and maintaining high image clarity even under low lighting conditions. The camera is equipped with a quality autofocus lens and UVC (USB Video Class) certification, ensuring full compatibility with Linux systems and stable operation on the Raspberry Pi platform. The device is mounted on the user's chest or upper clothing with a clip or strap, oriented in the user's line of sight (Figure 4).

Audio output is implemented through an external Logitech G Pro X USB sound card, to which standard wired headphones are connected via a 3.5 mm jack. This configuration provides stable sound quality and minimal delay in speech synthesis. It is recommended to use a single earphone that freely fits into the user's ear (Figure 5).

The system is powered by an external power supply with Power Delivery support. Capacity – 10000 mAh, output – 5V/3A. This allows uninterrupted operation of the device for 4–6 hours. In the prototype, a Baseus PD 20W 10,000 mAh power bank is used (Figure 6).

The complete system forms a wearable device, shown in Figure 7.

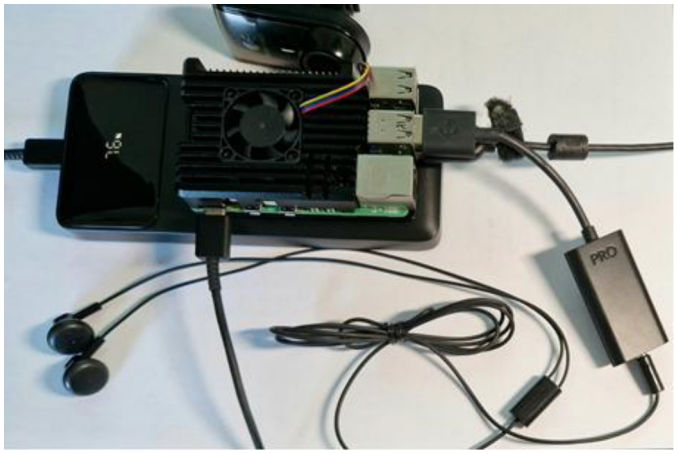


Fig. 7. General view of the module

The installation sequence of the hardware module is as follows:

– Raspberry Pi is attached to the user's belt;
– the camera is mounted on the chest;
– the earphone is placed in the ear;
– the power bank is mounted on the back of the belt.

The wiring is routed along the clothing so as not to hinder movement or obstruct the view (Figure 8).

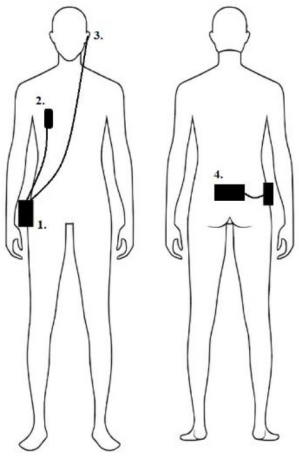


Fig. 8. Location of elements on the user

Thus, since the components are sufficiently compact and lightweight, carrying the device is simple and convenient for the user.

To verify the functionality of the developed hardware-software system, testing was conducted under real-world operating conditions. The main goal of the testing was to determine system stability, accuracy of vehicle detection, recognition of route numbers, and quality of voice output under various environmental conditions.

The results were evaluated based on the content of voice messages generated by the module. The formation of voice messages in the developed module takes place in several sequential stages (Figure 9). At the software level, after completing the transport detection algorithm (YOLOv8) and route number recognition (Tesseract OCR), a text string containing the analysis result is generated, for example: "Bus, route 25." This text is passed to the speech module implemented with the pyttsx3 library.

Fig. 9. Voice message generation sequence

The pyttsx3 library is a Python wrapper for the standard speech synthesis engine that works locally without the need for Internet access. In the case of PiOS (Raspberry Pi OS), it uses low-level synthesis libraries such as espeak or espeak-ng as the speech generation engine, as well as pyaudio or ALSA (Advanced Linux Sound Architecture) for sending the audio signal to the sound device. Thus, pyttsx3 acts as an intermediate layer that receives text data, converts them into a command for the synthesizer (espeak/espeak-ng), and then transmits the resulting audio stream to the playback system.

At the hardware level, the Raspberry Pi 5 sends the digital audio signal via the ALSA driver to the external Logitech G Pro X USB sound card, which converts the signal from digital form (PCM) into analog. This signal is then transmitted via a standard 3.5 mm audio jack to the user's headphones.

The messages generated by the software module have two types:

– announcement of the transport class only;
– announcement of both the transport class and its route number.

Thus, as a result of the module's operation, six types of messages can be obtained, as presented in Table 2.

Table 2

Voice message options

Only transport detected	Transport and route number detected
"Trolleybus is coming"	"Trolleybus (№) is coming"
"Bus is coming"	"Bus (№) is coming"
"Tramway is coming"	"Tramway (№) is coming"

Testing was carried out in the city of Kharkiv over the course of one day, at public transport stops in normal summer weather. Several locations were selected where the movement of buses, trolleybuses, and trams was observed during both daytime and evening hours.

During real-world testing of the system, 10 experiments were conducted under different lighting conditions. In daytime (100–10,000 l*x*), the system demonstrated high efficiency: vehicle detection accuracy reached 100 %, transport type classification accuracy – 90 %, and correct OCR recognition of route numbers – 70 %.

It should be noted that the distance at which the transport route number was recognized directly influenced the output message generated by the module. On average, at a distance of 80 meters from the vehicle, the module announced its class. At a distance of 50 meters, the route number was recognized (Table 3). As a result, full announcements of both transport type and route number were obtained in 7 out of 10 cases.

In twilight conditions (50–100 l*x*), the system performance decreased: vehicle detection accuracy was 80 %, type classification accuracy – 70 %, and OCR recognition succeeded in 50 % of cases. Consequently, the number of successful full announcements dropped to 5 out of 10.

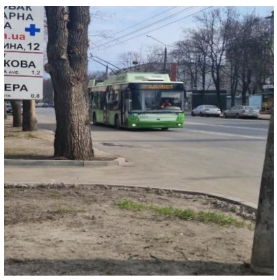
Conclusions

The developed hardware module confirmed the feasibility of using a dedicated hardware-software solution to ensure autonomy and convenience of orientation for visually impaired individuals in urban environments. The module demonstrated stable operation under daylight conditions, is sufficiently compact and portable, and allows the user to easily carry and apply it while waiting at public transport stops. Owing to its independence from Internet connectivity, the device can operate effectively even outside urban areas or in the absence of network coverage, thus broadening its scope of application.

At the same time, the test results revealed several aspects requiring improvement. First of all, the hardware case requires modernization to enhance its durability and resistance to external factors. Moisture resistance, dust protection,

Table 3

Notifications depending on the distance to transport

Type of transport	Distance, m	Message
	≈80	“Trolleybus is coming”
	≈40	“Trolleybus three is coming”
	≈60	“Tramway is coming”
	≈30	“Tramway twenty is coming”
	≈70	“Bus is coming”
	≈40	“Bus is thirteen is coming”

and tolerance to temperature fluctuations and mechanical impacts must be considered, which is especially important for outdoor use. Another development direction involves optimizing the case design to ensure ergonomics and comfort during prolonged wear or use in various weather conditions.

Another important task for further enhancement is the modernization of the power supply system. While the use of external power banks provides autonomy, it has certain disadvantages, including increased size and reduced portability. A more appropriate solution is the integration of a built-in rechargeable battery with energy-saving features and fast-charging capability. This would significantly increase the device's autonomy, reduce its weight, and ensure compactness.

A promising direction for hardware development is also the optimization of the component base with the goal of reducing power consumption and improving the overall efficiency of the module. The use of energy-efficient microcontrollers and additional hardware accelerators for image processing would significantly increase the device's performance while maintaining autonomy.

Bibliography

1. Нечітайло О. В., Гаврашенко А. О. Інтелектуальний модуль навігації для реабілітації осіб з вадами зору на зупинках громадського транспорту. Сучасні напрями розвитку інформаційно-комунікаційних технологій та засобів управління : тези доп. 15-ї міжнар. наук.-техн. конф. (24–25 квіт. 2025 р., Баку – Харків – Жиліна). Т. 3. Харків ; Баку ; Жиліна, 2025. С. 123.

2. Barkovska O., Serdechnyi V. Development of a universal model of an intellectual system to help people with visual impairments. Information Technology and Security. 2024. Vol. 12, No. 2. P. 47–56.

3. Kumar A., Ahmed S., Lee J. MR.NAVI: Mixed Reality Navigation Assistance for Visually Impaired Using Computer Vision and Natural Language Processing. 2025.

4. Shah P., Patel K., Reddy S. Smart Real-Time Object Detection for Blind Individuals Using YOLOv8 and OCR. Journal of Emerging Technologies. 2024. Vol. 18, No. 3. P. 115–123.

5. Mahmud R., Zhang T., Li Y. VISA: A Vision-Based Spatial Awareness System for Indoor Navigation of Visually Impaired. Journal of Imaging. 2024. Vol. 11, No. 1. P. 9.

6. Rafiq M., Hussain A., Khan N. Navigation Technologies for Visually Impaired: A Systematic Review (2019–2024). Sensors. 2024. Vol. 25, No. 7. P. 2213.

References

1. Nechitailo, O. V., & Havrashenko, A. O. (2025). Intellectual navigation module for the rehabilitation of visually impaired people at public transport stops. *Modern directions of development of information and communication technologies and management tools: Proceedings of the 15th International Scientific and Technical Conference (Vol. 3, p. 123). Baku – Kharkiv – Žilina*.

2. Barkovska, O., & Serdechnyi, V. (2024). Development of a universal model of an intellectual system to help people with visual impairments. *Information Technology and Security, 12*(2), 47–56.

3. Kumar, A., Ahmed, S., & Lee, J. (2025). *MR.NAVI: Mixed Reality Navigation Assistance for Visually Impaired Using Computer Vision and Natural Language Processing*.

4. Shah, P., Patel, K., & Reddy, S. (2024). Smart real-time object detection for blind individuals using YOLOv8 and OCR. *Journal of Emerging Technologies, 18*(3), 115–123.

5. Mahmud, R., Zhang, T., & Li, Y. (2024). VISA: A vision-based spatial awareness system for indoor navigation of visually impaired. *Journal of Imaging, 11*(1), 9.

6. Rafiq, M., Hussain, A., & Khan, N. (2024). Navigation technologies for visually impaired: A systematic review (2019–2024). *Sensors, 25*(7), 2213.

Дата першого надходження рукопису до видання: 30.09.2025
Дата прийнятого до друку рукопису після рецензування: 27.10.2025
Дата публікації:

УДК 004.8: 004.738.5 DOI

М. Р. БЄЛІКОВ
здобувач вищої освіти
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
ORCID: 0009-0004-1204-0349

А. О. БУР
здобувач вищої освіти
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
ORCID: 0009-0004-9320-7445

С. В. ПОПЕРЕШНЯК
кандидат фізико-математичних наук, доцент,
доцент кафедри інформатики та програмної інженерії
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
ORCID : 0000-0002-0531-9809

МІКРОСЕРВІСНА АРХІТЕКТУРА ДЛЯ ПОБУДОВИ КРИПТО-ФРІЛАНС БІРЖІ

У статті запропоновано архітектурне рішення для побудови крипто-фріланс біржі нового покоління, яке поєднує переваги мікросервісного підходу та блокчейн-технологій. Робота обґрунтовує актуальність теми у зв'язку з недоліками централізованих платформ, серед яких – високі комісії, затримки у виплатах та низький рівень довіри між користувачами. Поставлена мета полягає у створенні масштабованої, безпечної та прозорої архітектури, здатної забезпечити ефективну взаємодію замовників і виконавців через інтеграцію смарт-контрактів.

Архітектура системи реалізована на основі принципів предметно-орієнтованого проєктування та представлена у вигляді модульного середовища з незалежними сервісами для автентифікації, управління замовленнями, транзакціями, файловим обміном і аналітикою. Для візуалізації використано UML- та C4-діаграми, що демонструють логіку роботи та взаємозв'язки між сервісами. Взаємодія мікросервісів здійснюється через gRPC-протокол, тоді як зовнішні API функціонують на основі HTTPS. Для зберігання даних застосовано багаторівневий підхід: PostgreSQL використовується для критичних транзакційних даних, MongoDB – для динамічних об'єктів, Redis – для кешування та роботи в реальному часі.

Особливу увагу приділено питанням безпеки. Запроваджено механізми захисту від SQL-ін'єкцій, XSS-атак, DDoS-навантажень і фішингових загроз. Для захисту інтеграції з блокчейн-гаманцями використано алгоритми підпису транзакцій та шифрування приватних ключів. Смарт-контракти забезпечують автоматичне виконання умов угоди, прозорість фінансових розрахунків і зниження ризиків шахрайства.

Результати дослідження підтвердили ефективність поєднання мікросервісної архітектури та блокчейну для створення сучасної фріланс-платформи. Запропоноване рішення демонструє високу масштабованість, відмовостійкість і можливість подальшої інтеграції з AI-сервісами та іншими зовнішніми інструментами. У висновках наголошено, що платформа може стати основою для розвитку децентралізованих фріланс-сервісів, DeFi-проєктів і нових економічних моделей, де ключовими принципами є прозорість, довіра та автоматизація.

Ключові слова: *Інтерфевебдодаток, програмна архітектура, мікросервіси, фріланс, блокчейн, криптовалюта, смарт-контракти, безпека даних, Internet of Everything.*

M. R. BIELIKOV
Higher Education Student
National Technical University of Ukraine
"Ihor Sikorsky Kyiv Polytechnic Institute"
ORCID: 0009-0004-1204-0349

A.O. BUR
Higher Education Student
National Technical University of Ukraine
"Ihor Sikorsky Kyiv Polytechnic Institute"
ORCID: 0009-0004-9320-7445

© Бєліков М. Р., Бур А. О., Поперешняк С. В., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

S. V. POPERESHNYAK
Candidate of Physical and Mathematical Sciences, Associate Professor,
Associate Professor at the Department of Informatics
and Software Engineering
National Technical University of Ukraine
"Ihor Sikorsky Kyiv Polytechnic Institute"
ORCID: 0000-0002-0531-9809

MICROSERVICES ARCHITECTURE FOR BUILDING A CRYPTO FREELANCE EXCHANGE

The article proposes an architectural solution for building a next-generation crypto-freelance exchange that combines the advantages of a microservices approach with blockchain technologies. The study highlights the relevance of the topic, driven by the shortcomings of centralized platforms such as high service fees, payment delays, and low trust between users. The main goal is to design a scalable, secure, and transparent architecture capable of ensuring effective interaction between clients and freelancers through the integration of smart contracts.

The system architecture is developed on the principles of domain-driven design and presented as a modular environment with independent services for authentication, order management, transaction processing, file handling, and analytics. UML and C4 diagrams are employed to visualize the system's logic and inter-service interactions. Communication between microservices is established via the gRPC protocol, while external APIs operate over HTTPS. A multi-layered data storage approach is applied: PostgreSQL is used for critical transactional data, MongoDB for dynamic objects, and Redis for caching and real-time processing.

Special attention is given to security issues. Protection mechanisms against SQL injections, XSS attacks, DDoS loads, and phishing threats have been implemented. Blockchain wallet integration is safeguarded through transaction signing algorithms and private key encryption. Smart contracts ensure automatic fulfillment of agreements, transparent financial operations, and reduced fraud risks.

The research results confirm the effectiveness of combining microservices architecture with blockchain in creating a modern freelance platform. The proposed solution demonstrates high scalability, fault tolerance, and the potential for further integration with AI services and other external tools. The conclusions emphasize that such a platform can serve as a foundation for decentralized freelance services, DeFi projects, and innovative economic models, where transparency, trust, and automation are the defining principles.

Key words: *web application, software architecture, microservices, freelance, blockchain, IoT.*

Постановка проблеми

Сфера фріланс-послуг переживає стрімке зростання, однак наявні централізовані платформи продовжують стикатися з низкою системних проблем. Серед найбільш критичних – високі комісійні збори, затримки з виведенням коштів, а також уразливість до шахрайських дій з боку як замовників, так і виконавців. Крім того, централізований контроль таких платформ ускладнює процес арбітражу та викликає недовіру до прозорості рішень адміністрації.

Зважаючи на це, все більш актуальними стають рішення, що базуються на децентралізованих технологіях, таких як блокчейн та смарт-контракти. Вони дозволяють автоматизувати виконання умов контракту, забезпечити прозорість транзакцій, зменшити вплив посередників і підвищити рівень довіри між сторонами. Однак повна децентралізація часто знижує продуктивність системи, ускладнює масштабування та створює технічні труднощі, пов'язані з інтеграцією зі звичними веб-інтерфейсами.

Класичні архітектурні рішення для побудови цифрових платформ (монолітні або багатофункціональні гібридні системи) часто не здатні забезпечити гнучкість, масштабованість і швидке оновлення окремих модулів. Це створює потребу у впровадженні мікросервісної архітектури, яка дозволяє розподілити функціональність на незалежні сервіси з чітко визначеними зонами відповідальності, що легко масштабуються, тестуються та підтримуються. Водночас відсутність узагальнених підходів до інтеграції мікросервісів із блокчейн-рішеннями в контексті фріланс-платформ створює наукову прогалину, яку необхідно заповнити.

Отже, постає актуальна інженерна та наукова проблема – розробка архітектурного рішення, що поєднує масштабовану мікросервісну інфраструктуру з надійною криптографічною логікою на базі блокчейн-технологій для створення безпечної, прозорої та ефективної фріланс-платформи нового покоління.

Формулювання мети дослідження

Метою дослідження є проєктування масштабованої, безпечної та ефективної мікросервісної архітектури для крипто-фріланс платформи, яка усуває ключові недоліки сучасних фріланс-сервісів – зокрема, високі комісії, затримки виплат і ризики шахрайства – шляхом інтеграції блокчейн-технологій і смарт-контрактів.

Для досягнення цієї мети ставляться такі основні завдання:

1. Сформувати вимоги до архітектури платформи з урахуванням функціональних і нефункціональних аспектів, зокрема безпеки, масштабованості та гнучкості розгортання.

2. Розробити архітектуру, побудовану на мікросервісному підході, з розподілом на незалежні сервіси для ключових бізнес-функцій.

3. Інтегрувати блокчейн-механізми, зокрема смарт-контракти, для управління замовленнями, транзакціями та взаємодією між користувачами.

4. Побудувати діаграми архітектури (UML, C4) для візуалізації компонентів системи, їхньої взаємодії та логіки роботи.

Запропоноване рішення має слугувати технічною основою для створення децентралізованої, надійної та прозорої цифрової платформи для фріланс-взаємодії.

Аналіз останніх досліджень і публікацій

Сучасні дослідження зосереджені на використанні блокчейн-технологій та смарт-контрактів у цифрових платформах, зокрема у сфері фрілансу. У роботі [1] запропоновано поєднання блокчейну з біометричними технологіями для зменшення шахрайства на фріланс-сайтах. У роботі [2] акцентують увагу на методології Domain-Driven Design, що дозволяє більш чітко пов'язати бізнес-вимоги з архітектурою мікросервісних систем. Аналітичний звіт [3] демонструє значне зростання обсягів ринку фріланс-послуг у 2025 році, що підтверджує актуальність пошуку масштабованих і безпечних платформ. У дослідженні [4] подано узагальнений огляд архітектури смарт-контрактів, їхніх застосувань і перспектив. У роботі [5] описали платформу HireChain на базі Ethereum, орієнтовану на скорочення затримок у виплатах. Систему Boomerang розробили в [6], яка використовує Hyperledger для підвищення довіри й безпеки, проте характеризується зниженими показниками продуктивності. У роботі [7] акцентують увагу на довірі та прозорості у децентралізованих фріланс-системах. Окрему групу становлять праці, що торкаються суміжних питань захисту даних і обробки інформації. Так, у [8] розглянули методи підвищення криптографічної надійності шляхом контролю генерації випадкових чисел, а в [9] дослідили методи доступу й обробки мультимедійного контенту у хмарних середовищах.

Таким чином, наявні роботи доводять перспективність інтеграції блокчейну та мікросервісних підходів у побудові децентралізованих платформ, але водночас вказують на проблеми продуктивності, масштабованості й захисту, які потребують подальших досліджень.

Викладення основного матеріалу дослідження

Архітектура системи. Запропонована архітектура системи для крипто-фріланс біржі базується на мікросервісній архітектурі з використанням принципів предметно-орієнтованого проєктування. Основною моделлю є багаторівнева архітектура, яка забезпечує гнучкість, масштабованість та зручність обслуговування. Для візуалізації системи використовується модель C4 на трьох рівнях абстракції, що демонструє компоненти, контейнери та внутрішню структуру сервісів. На найвищому рівні (рисунок 1) система представлена як єдине рішення, що взаємодіє із зовнішніми учасниками та сервісами, такими як блокчейн Ethereum, сервіси генерації тексту на основі великих мовних моделей, які використовуються для автоматичного покращення описів завдань, узагальнення контрактів та інтелектуальних пропозицій, а також зовнішні поштові провайдери. Поняття «Користувач» узагальнене і включає адміністраторів, замовників, фрілансерів та гостей – усі вони взаємодіють із системою через єдину точку входу.

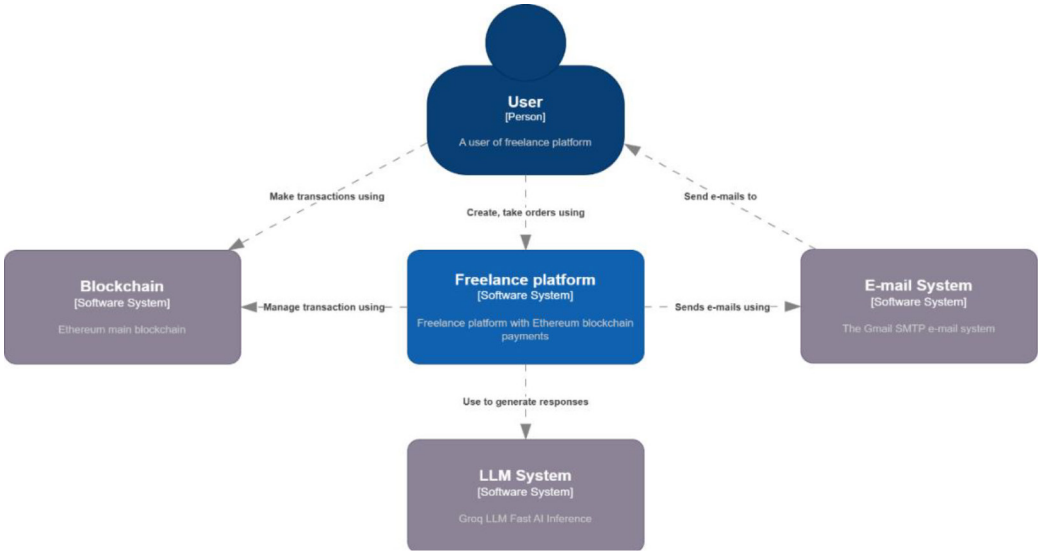


Рис. 1. C4 рівень 1, діаграма крипто-фріланс біржі

Джерело: розроблено авторами

Контейнери.

На рівні C4 Рівень 2 (Рисунок 2) кожен контейнер розкладається на компоненти, з акцентом на модульність та розділення відповідальностей через gRPC для внутрішніх викликів та HTTPS/JSON для зовнішніх API.

Архітектура включає:

1. Вебзастосунок – фронтенд-інтерфейс, побудований на Next.js, що забезпечує адаптивний інтерфейс користувача для всіх типів користувачів платформи.

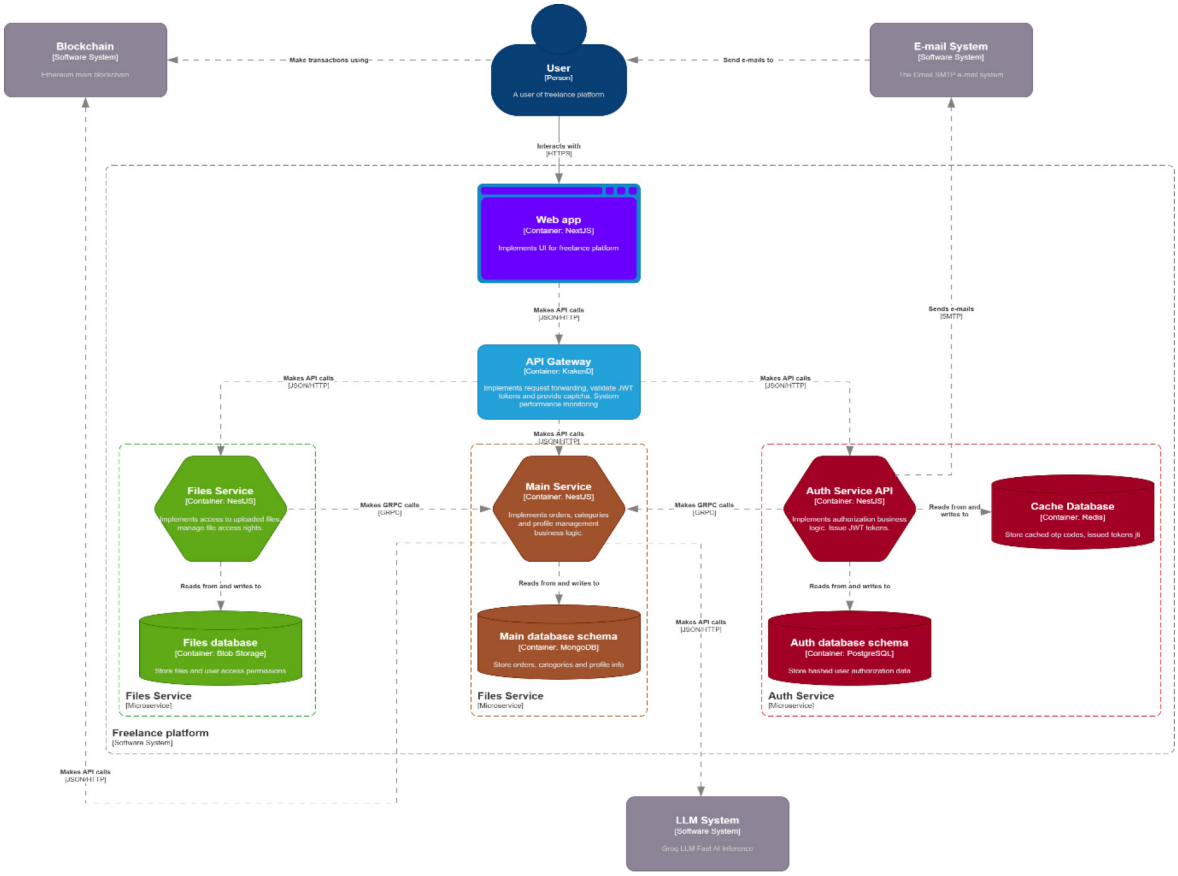


Рис. 2. C4 рівень 2, діаграма контейнерів основних мікросервісів

Джерело: розроблено авторами

2. API Gateway – центральна точка входу для маршрутизації HTTP/gRPC запитів до бекенд-сервісів, обробляє автентифікацію, обмеження швидкості запитів та спостережуваність.
3. Сервіс автентифікації – відповідає за автентифікацію та авторизацію користувачів, включаючи вхід, реєстрацію, багатофакторну автентифікацію, блокування акаунтів та видачу JWT. Інтегрується з PostgreSQL, Redis та Gmail SMTP для верифікації електронної пошти і MFA.
4. Головний сервіс – реалізує основну бізнес-логіку: керування проєктами, ролями користувачів, вирішення суперечок, взаємодію зі смарт-контрактами та AI-перевірку. Використовує MongoDB для гнучкого та масштабованого збереження даних, а також взаємодіє з Ethereum через Web3.js і LLM API від Groq для генерації тексту та перевірок тексту на спам.
5. Сервіс файлів – керує завантаженням, зберіганням (через Blob Storage) і безпечним доступом до файлів, що завантажуються користувачами.

Взаємодія між сервісами відбувається.

Компоненти.

На рівні C4 Рівень 3 (Рисунок 3) кожен контейнер розкладається на компоненти, з акцентом на модульність та розділення відповідальностей.

Кожен мікросервіс включає:

1. Контролери – для обробки вхідних API/gRPC запитів.
2. Сервіси – які інкапсулюють бізнес-логіку.
3. Репозиторії або шари доступу до даних – для взаємодії з базами даних.
4. Клієнти або адаптери – для зовнішніх сервісів (SMTP, Ethereum, LLM).

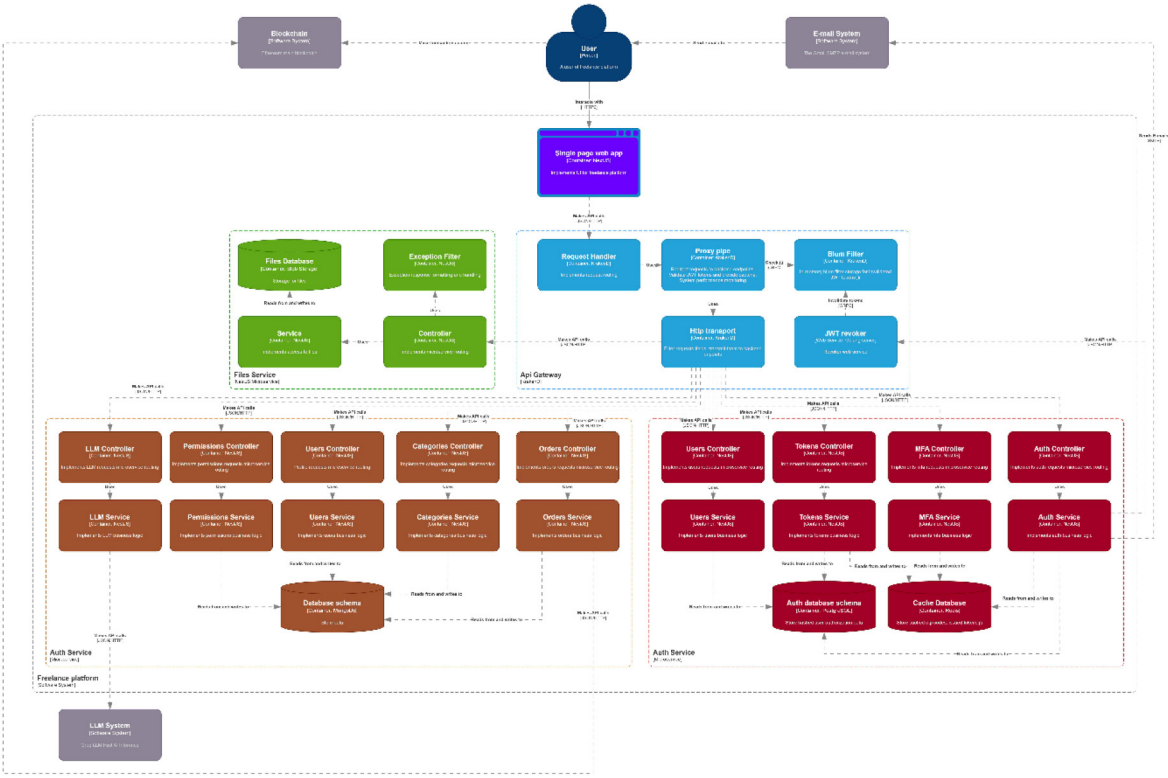


Рис. 3. C4 рівень 3, внутрішня архітектура всіх мікросервісів

Джерело: розроблено авторами

Для взаємодії між сервісами обрано gRPC замість REST через його низьку затримку та ефективний бінарний протокол (Protocol Buffers). Для зовнішніх API використовується HTTPS, що забезпечує шифрування даних і автентифікацію сервера.

Протокол gRPC використовується для передачі даних між мікросервісами через його численні переваги. По-перше, gRPC забезпечує високу швидкість обміну даними завдяки використанню протоколу HTTP/2, що підтримує мультиплексування потоків і ефективне управління з'єднанням. Це особливо важливо в мікросервісній архітектурі, де кожен сервіс повинен взаємодіяти з іншими для виконання складних операцій. Крім того, gRPC використовує серіалізацію через Protocol Buffers, що дозволяє передавати менший обсяг даних і швидше обробляти запити. Це знижує затримки між запитами та відповідями, підвищуючи загальну продуктивність системи.

Аналіз безпеки даних. Було враховано потенційні ризики та забезпечено захист від поширених загроз шляхом інтеграції сучасних технологій безпеки.

Можливі атаки, які можуть бути спрямовані на систему, включають SQL-ін'єкції, атаки міжсайтового скриптування (XSS), атаки грубої сили (DDoS), фішинг, експлуатацію вразливостей у сторонньому програмному забезпеченні, атаки на блокчейн-гаманці та перехоплення даних через атаку «людина посередині» (Man-in-the-Middle).

Для захисту від SQL-ін'єкцій система використовує параметризовані запити та ORM-інструменти, що мінімізують ризик виконання шкідливих команд у базі даних. Вразливості XSS нейтралізуються шляхом ретельної перевірки та очищення вхідних даних, а також шифруванням конфіденційної інформації перед її відображенням на фронтенді.

Для боротьби з DDoS атаками при реєстрації та авторизації використовуються обмеження кількості спроб входу, капчі та двофакторна автентифікація. Ризики фішингу мінімізуються шляхом шифрування даних через SSL і TLS.

Для захисту інтеграції з блокчейн-гаманцями реалізовано алгоритми підпису транзакцій, шифрування приватних ключів і взаємодію лише через офіційні API-інтерфейси. Захист від атак типу Man-in-the-Middle забезпечується використанням HTTPS. Цілісність і конфіденційність даних підтримується шляхом шифрування зберігання конфіденційної інформації з використанням AES-256 та bcrypt.

Архітектура баз даних. У розробленій архітектурі системи використовується поєднання реляційних і нереляційних баз даних для забезпечення як цілісності даних, так і гнучкості роботи. PostgreSQL виступає основним сховищем для структурованих даних, забезпечуючи надійні гарантії узгодженості, підтримку складних запитів і потужне управління транзакціями. Він використовується зокрема мікросервісом автентифікації (Рисунок 4), де

критичною є надійність, суворе дотримання схеми даних і безпека для управління обліковими записами користувачів та інформацією про сесії.

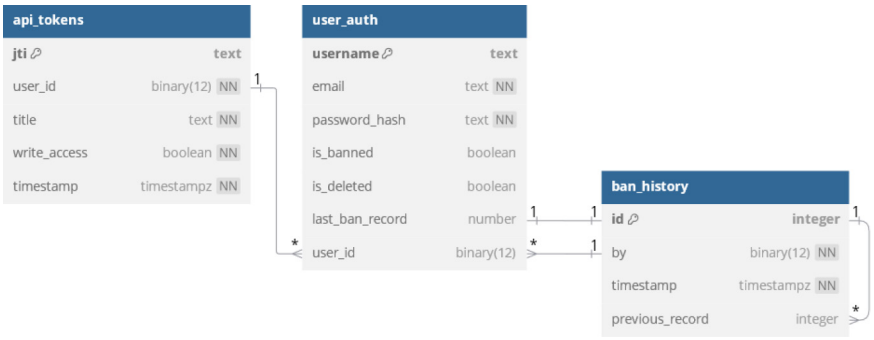


Рис. 4. ER-діаграма PostgreSQL мікросервісу автентифікації

Джерело: розроблено авторами

З іншого боку, MongoDB використовується в основному мікросервісі (Рисунок 5) для ефективної роботи з динамічними й напівструктурованими даними. Вона забезпечує легку адаптацію системи до змін у структурі даних без потреби в складних міграціях. Такий вибір особливо важливий для роботи з мінливими наборами даних, такими як інформація про паркомісця в реальному часі, результати виявлення автомобілів і журнали взаємодії користувачів.

Крім того, Redis інтегровано як сховище даних в оперативній пам'яті для кешування тимчасових і часто використовуваних даних. Зберігаючи стани сесій, проміжні результати обчислень і оновлення статусу в реальному часі в Redis, система значно знижує навантаження на запити до бази даних і прискорює час відгуку, тим самим підвищуючи загальну продуктивність і масштабованість.

Використання двох спеціалізованих баз даних – PostgreSQL для підсистеми автентифікації та MongoDB для основних функціональних сервісів – забезпечує баланс між сильною узгодженістю даних там, де це потрібно, і високою гнучкістю й продуктивністю, необхідною для динамічної роботи.

Процес обробки замовлення зі смарт-контрактами. Повний процес замовлення показано на Рисунку 6.

Замовник створює завдання на сайті платформи, і смарт-контракт створюється на адресу його гаманця, який блокує суму грошей, яку він вказав (бюджет на виконання). Замовник може скасувати замовлення. Це можна зробити як на сайті вебзастосунку, так і самостійно, викликавши функцію скасування в блокчейні через криптогаманець. У такому разі кошти будуть розблоковані, а смарт-контракт вважатиметься завершеним.

Інакше замовник чекатиме, доки один із фрілансерів прийме замовлення на нашій платформі. Коли це станеться, він отримає відповідне повідомлення. Замовник повинен відхилити або прийняти фрілансера, який хоче виконати завдання. Після вибору виконавця створюється запис у блокчейні. Безпека вибору гарантована, оскільки функція «взяти замовлення» може бути викликана лише з гаманця, який використовувався для ініціації смарт-контракту.

Фрілансер, у свою чергу, може відмовитися від виконання завдання, викликавши функцію «скасувати замовлення» через гаманець, підключений до платформи, але це погіршить його рейтинг. Потім замовник може чекати й обрати іншого фрілансера або скасувати завдання. Автоматичне скасування спрацює, коли спливе встановлений замовником термін.

Фрілансер повинен прикріпити файли з виконаним завданням. Замовник перевіряє їх, і якщо обидві сторони згодні, кошти перераховуються виконавцю. Якщо виникає конфлікт, втручається адміністратор платформи, який повинен перевірити історію чату й вирішити, чи виконав фрілансер поставлене завдання. Якщо так – гроші надсилаються виконавцю, і смарт-контракт завершується. Якщо ні – адміністратор обговорює зі сторонами конфлікту, чи дозволити поточному виконавцю внести зміни до завдання та продовжити роботу. У разі згоди виконавець має прикріпити нові файли до дедлайну або відмовитися від виконання.

Якщо сторони конфлікту не досягли згоди, адміністратор приймає рішення – чи дозволити фрілансеру брати нові замовлення на платформі, і має право його заблокувати. Виконавця виключають зі смарт-контракту. Замовник може чекати нового фрілансера. Проте, якщо завдання більше неактуальне, він може скасувати його. Кошти будуть розблоковані, і смарт-контракт вважатиметься завершеним.

Незважаючи, на досить великий список обмежень і викликів з якими стикаються науковці дана тема має перспективи, які сформовані на основі сучасного стані розвитку та прогресу.

Перспективи подальших досліджень.

1. Розвиток гібридних моделей. Поєднання різних підходів, таких як глибокі нейронні мережі та традиційні алгоритми оптимізації, може дозволити отримати кращі результати в адаптації інтерфейсів. Гібридні моделі, що

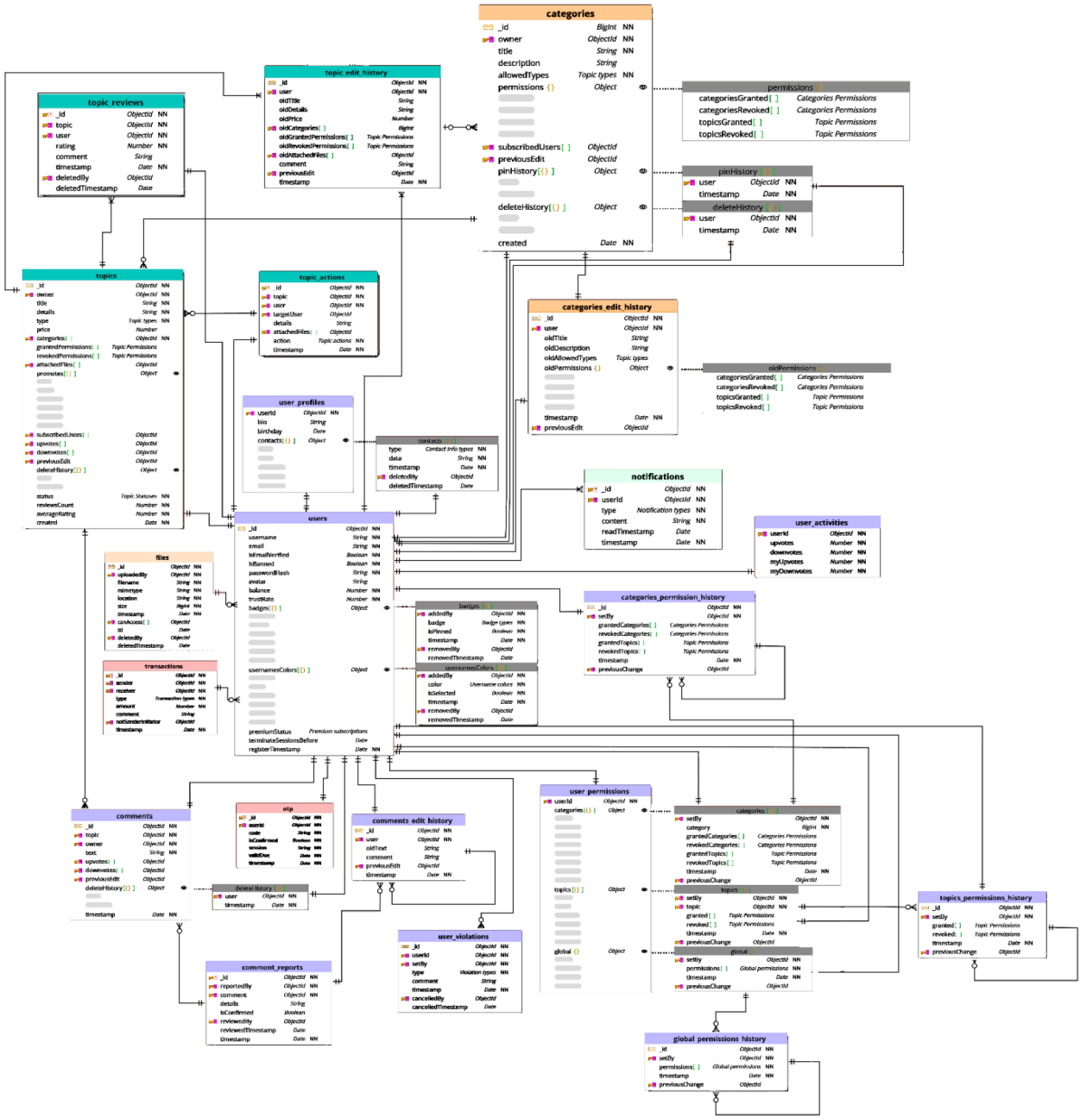


Рис. 5. ER-діаграма MongoDB головного мікросервісу

Джерело: розроблено авторами

використовують одночасно кілька методів аналізу поведінкових даних, зможуть враховувати більшу кількість змінних і краще реагувати на зміни в поведінці користувачів.

2. Розробка мультимодальних адаптивних інтерфейсів. Дослідження можливостей створення адаптивних інтерфейсів, що реагують не тільки на поведінкові дані, але й на контекст використання (наприклад, місце, час, пристрій), можуть значно покращити якість взаємодії користувача з додатком. Використання датчиків і сенсорних систем для адаптації інтерфейсів відкриває нові перспективи для інтеграції різних типів даних.

3. Етичні питання та захист даних. З огляду на зростаюче використання поведінкових даних, важливо розробляти нові підходи до забезпечення конфіденційності та безпеки даних користувачів. Дослідження в цьому напрямі можуть допомогти впровадити більш ефективні механізми захисту даних без погіршення якості адаптації інтерфейсів.

4. Автоматизація процесів збору та аналізу даних. Подальший розвиток автоматизованих систем збору поведінкових даних дозволить знизити витрати на аналітику та покращити точність отриманих результатів. Системи, що автоматично збирають і аналізують дані, а потім адаптують інтерфейси у реальному часі, є перспективним напрямком для покращення користувацького досвіду.

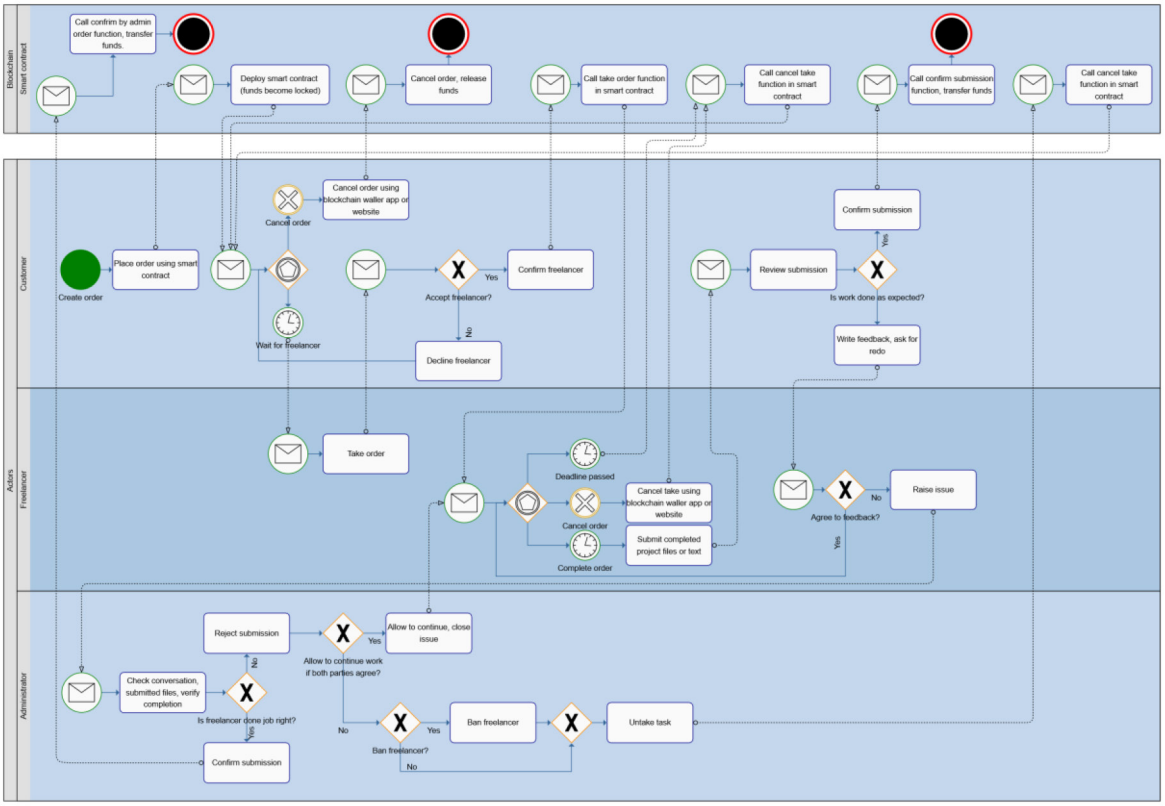


Рисунок 6 – BPMN-діаграма процесу виконання фріланс-замовлення

Джерело: розроблено авторами

Загалом, подальші дослідження мають бути спрямовані на покращення ефективності алгоритмів оптимізації, збільшення захисту даних та забезпечення гнучкості систем при адаптації інтерфейсів до змін у поведінці користувачів.

Висновки і перспективи подальших розвідок у даному напрямі

У межах даного дослідження розроблено архітектурне рішення для створення децентралізованої фріланс-платформи, що поєднує принципи мікросервісної інженерії та блокчейн-технологій. Запропонована система демонструє здатність ефективно масштабуватись, забезпечувати високий рівень безпеки та прозорість взаємодії між користувачами шляхом інтеграції смарт-контрактів. Завдяки модульній структурі, кожен компонент архітектури може розвиватися незалежно, що значно спрощує супровід і подальше розширення функціональності. Платформа досягає високої масштабованості та відмовостійкості завдяки модульному підходу. Інтеграція блокчейн-технології для транзакцій забезпечує надійність надання послуг та високий рівень безпеки.

Використання gRPC-протоколу, багаторівневого підходу до зберігання даних (PostgreSQL, MongoDB, Redis) та C4-моделювання дозволило досягти гнучкості у проєктуванні та наочно представити логіку роботи системи. Результати дослідження підтверджують доцільність застосування мікросервісного підходу у сфері цифрових платформ, зокрема для сервісів, що вимагають надійних фінансових розрахунків і мінімізації людського втручання. Кожен компонент системи спроєктований для незалежної роботи. Розподіл обов'язків спрощує майбутнє розширення системи та полегшує інтеграцію з новими зовнішніми сервісами. Смарт-контракти додають додатковий рівень довіри та прозорості, захищаючи інтереси всіх залучених сторін. Механізм гарантує безпечне зберігання коштів та їхнє розблокування лише після успішного виконання завдання, що знижує ризики шахрайства.

Запропонована архітектура може слугувати основою для розробки новітніх платформ у сфері фрілансу, децентралізованих фінансових сервісів (DeFi) та смарт-економіки в цілому.

Платформа, відповідає сучасним вимогам індустрії та закладає надійну основу для постійного вдосконалення та адаптації до новітніх технологій.

Список використаної літератури

1. Batool A., Byun Y. Reduction of Online Fraudulent Activities in Freelancing Sites Using Blockchain and Biometric. *Electronics*. 2022. Vol. 11, No. 5. Art. 789. DOI: https://doi.org/10.3390/electronics11050789

2. Jaiswal S. K., Agrawal R. Domain-Driven Design (DDD) – Bridging the Gap between Business Requirements and Object-Oriented Modeling. *International Journal of Innovative Research in Engineering and Management (IJIREM)*. 2024. Vol. 11, No. 2. P. 79–83. DOI: https://doi.org/10.55524/ijirem.2024.11.2.16

3. Kumar N. Latest Freelance Statistics 2025 – Industry Size & Trends. 2025. URL: https://www.demandsage.com/freelance-statistics/ (дата звернення: 18.09.2025).

4. Wang S., Yuan Y., Wang X., Li J., Qin R., Wang F.-Y. An Overview of Smart Contract: Architecture, Applications, and Future Trends. *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)*, Changshu, China, 2018. P. 108–113. DOI: https://doi.org/10.1109/IVS.2018.8500488

5. Deshmukh P., Kalwaghe S., Appa A., Pawar A. Decentralised Freelancing using Ethereum Blockchain. *Proceedings of the International Conference on Communication and Signal Processing (ICCSP)*, Chennai, India, 2020. P. 881–883. DOI: https://doi.org/10.1109/ICCSP48568.2020.9182127

6. Pallam B., Gore M. M. Boomerang: Blockchain-based Freelance Paradigm on Hyperledger. *Proceedings of the 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, Kanpur, India, 2019. P. 1–6. DOI: https://doi.org/10.1109/ICCCNT45670.2019.8944572

7. Gandhi M., та ін. Decentralized Freelancing System – Trust and Transparency. *International Research Journal of Engineering and Technology (IRJET)*. 2019. Vol. 6, No. 9.

8. Popereshnyak S., Novikov Y., Zhdanova Y. Cryptographic system security approaches by monitoring the random numbers generation. *CEUR Workshop Proceedings*. 2024. Vol. 3826. P. 301–309. URL: https://ceur-ws.org/Vol-3826/short21.pdf (дата звернення: 18.09.2025).

9. Reshetniak M., Popereshnyak S. Method for accessing and processing multimedia content in a cloud environment. *Proceedings of the IEEE International Scientific-Practical Conference: Problems of Infocommunications Science and Technology*. 2019. P. 71–76. DOI: https://doi.org/10.1109/PICST47496.2019.9061463

References

1. Batool, A., & Byun, Y. (2022). Reduction of online fraudulent activities in freelancing sites using blockchain and biometric. *Electronics, 11*(5), 789. https://doi.org/10.3390/electronics11050789

2. Jaiswal, S. K., & Agrawal, R. (2024). Domain-driven design (DDD): Bridging the gap between business requirements and object-oriented modeling. *International Journal of Innovative Research in Engineering and Management, 11*(2), 79–83. https://doi.org/10.55524/ijirem.2024.11.2.16

3. Kumar, N. (2025). *Latest freelance statistics 2025–Industry size & trends*. Retrieved September 18, 2025, from https://www.demandsage.com/freelance-statistics/

4. Wang, S., Yuan, Y., Wang, X., Li, J., Qin, R., & Wang, F.-Y. (2018). An overview of smart contract: Architecture, applications, and future trends. In *Proceedings of the IEEE Intelligent Vehicles Symposium (IV)* (pp. 108–113). IEEE. https://doi.org/10.1109/IVS.2018.8500488

5. Deshmukh, P., Kalwaghe, S., Appa, A., & Pawar, A. (2020). Decentralised freelancing using Ethereum blockchain. In *Proceedings of the International Conference on Communication and Signal Processing (ICCSP)* (pp. 881–883). IEEE. https://doi.org/10.1109/ICCSP48568.2020.9182127

6. Pallam, B., & Gore, M. M. (2019). Boomerang: Blockchain-based freelance paradigm on Hyperledger. In *Proceedings of the 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICCCNT45670.2019.8944572

7. Gandhi, M., et al. (2019). Decentralized freelancing system – Trust and transparency. *International Research Journal of Engineering and Technology, 6*(9).

8. Popereshnyak, S., Novikov, Y., & Zhdanova, Y. (2024). Cryptographic system security approaches by monitoring the random numbers generation. *CEUR Workshop Proceedings, 3826*, 301–309. https://ceur-ws.org/Vol-3826/short21.pdf

9. Reshetniak, M., & Popereshnyak, S. (2019). Method for accessing and processing multimedia content in a cloud environment. In *Proceedings of the IEEE International Scientific-Practical Conference: Problems of Infocommunications Science and Technology (PIC S&T)* (pp. 71–76). IEEE. https://doi.org/10.1109/PICST47496.2019.9061463

Дата першого надходження рукопису до видання: 27.09.2025
Дата прийнятого до друку рукопису після рецензування: 24.10.2025
Дата публікації:

УДК 004.738.5 DOI

В. Д. БОЙКО
кандидат технічних наук, доцент,
доцент кафедри кібербезпеки
Національний університет «Одеська юридична академія»
ORCID: 0000-0001-5929-657X

В. І. БЕВЗА
магістр за спеціальністю 125 «Кібербезпека»
Національний університет «Одеська юридична академія»
ORCID: 0009-0007-2695-969X

В. М. СЛАТВІНСЬКА
доктор філософії, асистент кафедри кібербезпеки
Національний університет «Одеська юридична академія»
ORCID: 0000-0002-6082-981X

ІНІЦІЮВАННЯ МЕРЕЖЕВОЇ АКТИВНОСТІ В WINDOWS ЗА ДОПОМОГОЮ NETSTAT

Мета статті. Метою статті є створення комплексного підходу до ініціювання та аналізу мережевої активності в операційній системі Windows за допомогою утиліти netstat з урахуванням сучасних вимог, пов'язаних з відстеженням системи, та гарантуванням можливості інтерпретації результатів діагностичної активності найефективнішим чином. У рамках роботи також буде зроблено спробу стандартизувати використання Netstat, доповнити його іншими інструментами для автоматизації аналізу даних та перетворити його пропозиції на конкретні рекомендації для системних адміністраторів та експертів з кібербезпеки. Спеціальні програми для ідентифікації мережевих з'єднань, виявлення аномалій та аналізу навантаження також відіграють вирішальну роль у стабільності та безпеці інформаційних систем у динамічному мережевому середовищі.

Наукова новизна. Наукова новизна дослідження полягає у створенні класифікації станів TCP-з'єднань на основі даних, отриманих за допомогою netstat, що дозволяє ефективно інтерпретувати інформацію про мережеву активність у контексті діагностики системи. Архітектура системи моніторингу, що запропонована в цій роботі, включає чотири основні компоненти: збір первинних даних, обробку, аналіз та сповіщення про аномалії, розширює класичні підходи до використання утиліт Windows. Поєднання netstat із сучасними методами автоматизації та аналізу даних, особливо з часовими рядами та логуванням, надає нові можливості для виявлення трендів та аномалій.

Результати. В рамках дослідження було детально досліджено функціональність netstat, зокрема опції -a, -n, -o, -s та -r, які надають повний список активних з'єднань, IP-адрес, ідентифікаторів процесів, статистику для мережевих протоколів та таблиць маршрутизації. Було створено таблицю класифікації станів TCP-з'єднань (LISTENING, ESTABLISHED, TIMEWAIT, CLOSEWAIT) з описом діагностичних значень та рекомендаціями, згідно з якими можна ідентифікувати проблемні стани, наприклад, довгий CLOSE_WAIT. Результати, отримані з мережевих сеансів з IP-адресами 127.0.0.1 та 192.168.0.161 та портами 161:4147-4149, показані на рисунку 2, надають інформацію, пов'язану з інтенсивністю активності веб-застосунків, системними процесами, а також існуванням потенційних аномалій. Запропоновано способи написання сценаріїв для автоматизації збору даних та підключення до систем реєстрації даних для побудови часових рядів.

Висновки. Дослідження довело, що netstat ефективно запускає та моніторить мережеву активність у Windows та є корисним інструментом завдяки розширеним параметрам командного рядка та організованому методу збору даних для діагностики. Отримана архітектура моніторингу та класифікація стану з'єднання забезпечують основу для інтеграції традиційних утиліт із сучасними технологіями, такими як машинне навчання. Перспективи подальших досліджень включають розробку алгоритмів автоматичного виявлення аномалій, інтеграцію з хмарними системами, розробка адаптивних систем оповіщення, з практичними наслідками в галузі безпеки та оптимізації мережевих ресурсів (кібербезпека).

Ключові слова: *netstat, мережа, моніторинг, Windows, діагностика комп'ютерних мереж, безпека комп'ютерних мереж, кібербезпека.*

© Бойко В. Д., Бевза В. І., Слатвінська В. М., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

V. D. BOYKO
PhD in Engineering, Associate Professor,
Associate Professor of Cyber Security
National University Odesa Law Academy
ORCID: 0000-0001-5929-657X

V. I. BEVZA
Master's Degree in Cybersecurity (Specialty 125)
National University Odesa Law Academy
ORCID: 0009-0007-2695-969X

V. M. SLATVINSKA
PhD, Assistant at the Cybersecurity Department
National University Odesa Law Academy
ORCID: 0000-0002-6082-981X

INITIATING NETWORK ACTIVITY IN WINDOWS USING NETSTAT

The purpose of the article. The aim of the article is to create a complex approach to the initiation and analysis of network activity in the windows operating system with the help of the netstat utility with references to the current demands related to tracking of the system and guaranteeing the opportunity to interpret the findings of diagnostic activities in the most effective way. The work will also try to standardize the use of Netstat, complement it with other tools to automate the analysis of the data and translate its proposals into concrete recommendations for the system administrators and the cybersecurity experts. Specific applications for identifying network connections, anomaly detection and load analysis also play a crucial role in the stability and security of information systems in a dynamic network environment.

Scientific novelty. The scientific novelty of the research is the creation of a classification of TCP connection states on the basis of data obtained using netstat that permits effective interpretation of the information about network activity in the context of system diagnostics. The architecture of the monitoring system proposed in this work, which features four major components: primary data collection, processing, analysis, and notification about anomalies, extends classical approaches to the usage of Windows utilities. The combination of netstat together with modern automation and data analysis techniques, especially with time series and logging, provides new possibilities for trend and anomaly detection.

Results. As part of the study, the functionality of netstat was explored in detail, particularly the -a, -n, -o, -s and -r options, which give a complete listing of active connections, IP address, process ID, statistics for network protocols and routing tables. The classification table of TCP connection states (LISTENING, ESTABLISHED, TIMEWAIT, CLOSEWAIT) with description of diagnostic values and recommendations was created, according to which problematic states can be identified, for example, long CLOSE_WAIT. The results obtained from the network sessions with IP addresses 127.0.0.1 and 192.168.0.161 and ports 161:4147-4149, shown in Figure 2, provide information related to the intensity of the activity of the web applications, system processes and also the existence of potential anomalies. Ways for scripting to automate the data collection and connection with logging systems for building time series are suggested.

Conclusions. The study has proven that netstat is efficient for starting and monitoring the network activity in windows and is a useful tool with its advanced command-line parameters and organized method to capture the data for diagnostics. The resulting monitoring architecture and connection state classification provides the foundations for integration of traditional utilities with modern technologies such as machine learning. Future work opportunities: development of algorithms for automatic anomaly detection, integration with cloud systems, development of adaptive alerting systems, with practical consequences in the field of security and optimization of network resources (cybersecurity).

Key words: *netstat, network, monitoring, Windows, computer network diagnostics, computer network security, cybersecurity.*

Постановка проблеми

Сучасні комп'ютерні системи характеризуються інтенсивною мережевою активністю, що потребує ефективних засобів моніторингу та діагностики. Операційна система Windows надає розгалужений інструментарій для аналізу мережевого трафіку, серед якого утиліта netstat займає особливе місце завдяки своїй універсальності та доступності. Проблема ефективного використання netstat для ініціювання комплексного моніторингу мережевої активності набуває особливої актуальності в контексті зростаючих вимог до системної безпеки та продуктивності. Традиційні підходи до використання netstat часто обмежуються базовими функціями, що не дозволяє повною мірою використати потенціал цього інструменту. Необхідність систематизації методів застосування netstat та розробки оптимальних стратегій інтерпретації результатів зумовлює актуальність даного дослідження.

Аналіз останніх досліджень і публікацій

Аналіз сучасних досліджень у галузі мережевого моніторингу демонструє широкий спектр підходів до оптимізації системних утиліт. Бойко В. Д., Бевза В. І., Слатвінська В. М. [1, с. 1263] розглядають питання використання

штучного інтелекту для оптимізації системного коду, що має пряме відношення до автоматизації роботи з мережевими утилітами. Дослідження системних збоїв та їх впливу на мережеву безпеку представлено у працях Бевза В. І., Слатвінська В. М. [2, с. 335], де аналізуються наслідки масштабних системних відмов. Продовження цього дослідження Бевза В. І., Слатвінська В. М. [3, с. 252] поглиблює розуміння взаємозв'язків між системними збоями та мережевою активністю.

Сучасні підходи до аналізу мережевих систем представлені у роботах Zhong Y., Zhang Y., Zhang J., Wan L. [4, с. 15], які досліджують стійкість мережевих командно-інформаційних систем. Аналітика мережевих даних у контексті 5G систем розглядається Romero M. L., Suyama R. [5, с. 8], що демонструє еволюцію підходів до мережевого моніторингу. Інноваційні методи контекстуалізації мережевих вимірювань за допомогою великих мовних моделей представлені Beltiukov R., Bhattaram K., Cheng E., Kanigicherla V., Singh A., Thampiratwong K., Gupta A. [6, с. 12].

Практичні аспекти аналізу мережевого трафіку висвітлені у дослідженні Jiang W., Zhang B., Zhu Q., Liao C., Wang W. [7, с. 3], які представили датасет реального мережевого середовища. Сучасні методи виявлення мережевих вторгнень розроблені Balakrishnapillai J., Ajayan A. K. P., Kurian A., Sabu A., Majeed A. [8, с. 020044] з використанням новітньої технології DSSTE. Альтернативний підхід до виявлення вторгнень на основі згорткових нейронних мереж запропонований S D. B., A Ms.D. [9, с. 5]. Інноваційні методи виявлення аномалій мережевих комутаторів представлені Nam S., Jeong E., Hong J. W. [10, с. 2], а прогнозування мережевого трафіку досліджене Lv Q., Chang Y., Li T., Ge J. [11, с. 111557].

Формулювання мети дослідження

Метою дослідження є розробка комплексної методології ініціювання та аналізу мережевої активності в Windows за допомогою netstat з урахуванням сучасних вимог до системного моніторингу та забезпечення оптимальної інтерпретації результатів діагностики.

Викладення основного матеріалу дослідження

Утиліта netstat являє собою фундаментальний інструмент операційної системи Windows для моніторингу мережевої активності. Її функціональність охоплює широкий спектр діагностичних можливостей, від базового відображення активних з'єднань до детального аналізу мережевих статистик. Досвід використання netstat у контексті системної оптимізації, як показано у дослідженнях Бойко В. Д., Бевза В. І., Слатвінська В. М. [1, с. 1265], демонструє важливість інтеграції традиційних утиліт з сучасними підходами до автоматизації.

Основні параметри командного рядка netstat забезпечують різноманітні режими моніторингу. Параметр -a відображає всі з'єднання та порти прослуховування, що критично важливо для комплексного аналізу системи. Використання -n замінює символьні імена хостів на IP-адреси, прискорюючи виконання команди та полегшуючи автоматизовану обробку результатів. Параметр -o додає ідентифікатори процесів до кожного з'єднання, що дозволяє точно визначити джерело мережевої активності.

Аналіз системних збоїв та їх впливу на мережеву активність, досліджений Бевза В. І., Слатвінська В. М. [2, с. 336], підкреслює важливість систематичного моніторингу мережевих з'єднань. Продовження цього дослідження [3, с. 254] демонструє кореляції між системними відмовами та аномальною мережевою поведінкою, що може бути ефективно виявлено за допомогою netstat.

Для аналізу результатів netstat розроблено таблицю основних станів TCP-з'єднань та їх інтерпретації в контексті системної діагностики:

Таблиця 1

Класифікація станів TCP-з'єднань у контексті netstat

Стан з'єднання	Опис	Діагностичне значення	Рекомендації
LISTENING	Очікування вхідних з'єднань	Активний сервісний порт	Перевірити необхідність сервісу
ESTABLISHED	Активне з'єднання	Нормальна комунікація	Моніторити тривалість
TIME_WAIT	Очікування закриття	Нормальне завершення	Контролювати кількість
CLOSE_WAIT	Очікування закриття додатком	Можлива проблема додатка	Перевірити код додатка

Як видно з таблиці 1, кожен стан TCP-з'єднання має специфічне діагностичне значення, що дозволяє системним адміністраторам ефективно інтерпретувати результати netstat для вирішення мережевих проблем.

Сучасні підходи до аналізу мережевих систем, розроблені Zhong Y., Zhang Y., Zhang J., Wan L. [4, с. 18], підкреслюють важливість стійкості мережевих компонентів. Це безпосередньо пов'язано з можливостями netstat для виявлення потенційних проблем у мережевій інфраструктурі. Еволюція мережевих технологій, досліджена Romero M. L., Suyama R. [5, с. 10], вимагає адаптації традиційних інструментів моніторингу до нових реалій.

Інноваційні підходи до контекстуалізації мережевих вимірювань, представлені Beltiukov R., Bhattaram K., Cheng E., Kanigicherla V., Singh A., Thampiratwong K., Gupta A. [6, с. 14], демонструють перспективи інтеграції традиційних утиліт з сучасними технологіями машинного навчання. Це відкриває нові можливості для автоматизованого аналізу вихідних даних netstat.

Для візуалізації архітектури системи моніторингу мережевої активності на основі netstat розроблено структурну схему:

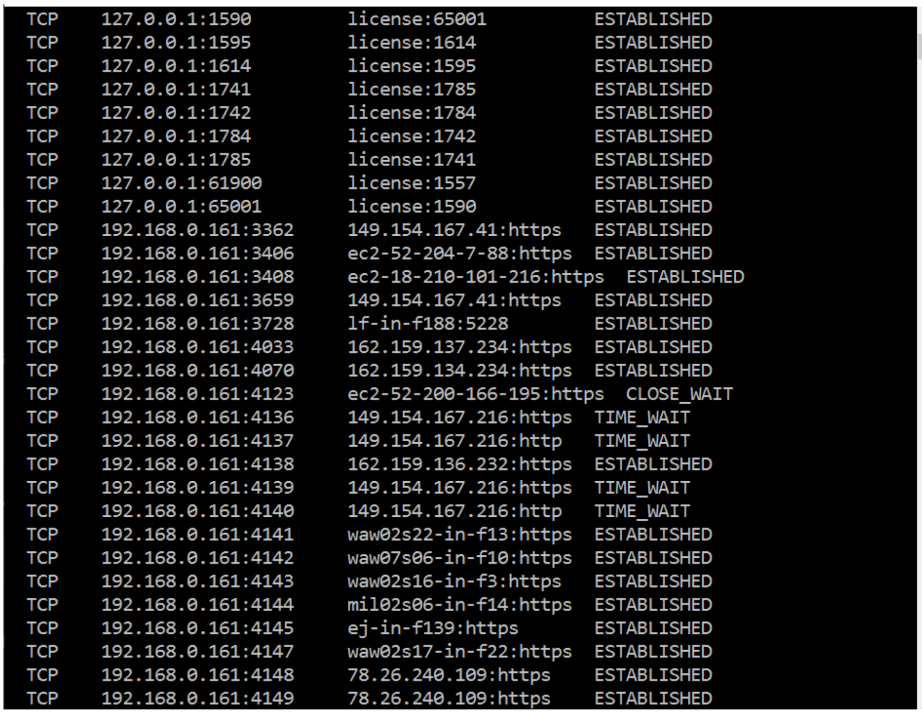
Рис. 1. Архітектура системи моніторингу на базі netstat

Джерело: авторська розробка

Як демонструє рисунок 1, архітектура системи моніторингу включає чотири основні компоненти, що забезпечують комплексний підхід до аналізу мережевої активності від збору первинних даних до генерації оповіщень про аномалії.

Як показано на рисунку 2, команда netstat з параметрами -a, -n та -o відображає всі мережеві сеанси в системі Windows, включаючи їхні числові адреси, ідентифікатори процесів та інші деталі. Існує багато TCP-з'єднань, здебільшого у стані ESTABLISHED, які відповідають двостороннім з'єднанням між локальним хостом (IP-адреса 127.0.0.1 та 192.168.0.161) та зовнішніми або внутрішніми серверами, що прослуховують порт 161, використовуючи протокол HTTPS на портах 4147-4149. Такі з'єднання, як CLOSE_WAIT та TIME_WAIT, представляють сеанси, що закриваються або очікують закриття (звичайний продукт динамічних мережевих транзакцій). З цієї інформації ми можемо зробити деякі висновки про інтенсивність мережевої активності веб-застосунків та системних процесів, а також надати уявлення про можливі аномалії, такі як стан CLOSE_WAIT, який може свідчити про проблему з правильним завершенням з'єднання.

Подальша оцінка даних на рисунку 2 може дозволити нам оцінити розподіл мережевих ресурсів та системне навантаження. Зокрема, той факт, що багато підключень пов'язані з однаковими ідентифікаторами процесів (наприклад, 149, 162), вказує на те, що кілька програм або служб, які взаємодіють з мережею, зайняті одночасно, і це може бути пов'язано з оновленнями програмного забезпечення у фоновому режимі або синхронізацією даних. Це робить ретельне



TCP	127.0.0.1:1590	license:65001	ESTABLISHED
TCP	127.0.0.1:1595	license:1614	ESTABLISHED
TCP	127.0.0.1:1614	license:1595	ESTABLISHED
TCP	127.0.0.1:1741	license:1785	ESTABLISHED
TCP	127.0.0.1:1742	license:1784	ESTABLISHED
TCP	127.0.0.1:1784	license:1742	ESTABLISHED
TCP	127.0.0.1:1785	license:1741	ESTABLISHED
TCP	127.0.0.1:61900	license:1557	ESTABLISHED
TCP	127.0.0.1:65001	license:1590	ESTABLISHED
TCP	192.168.0.161:3362	149.154.167.41:https	ESTABLISHED
TCP	192.168.0.161:3406	ec2-52-204-7-88:https	ESTABLISHED
TCP	192.168.0.161:3408	ec2-18-210-101-216:https	ESTABLISHED
TCP	192.168.0.161:3659	149.154.167.41:https	ESTABLISHED
TCP	192.168.0.161:3728	lf-in-f188:5228	ESTABLISHED
TCP	192.168.0.161:4033	162.159.137.234:https	ESTABLISHED
TCP	192.168.0.161:4070	162.159.134.234:https	ESTABLISHED
TCP	192.168.0.161:4123	ec2-52-200-166-195:https	CLOSE_WAIT
TCP	192.168.0.161:4136	149.154.167.216:https	TIME_WAIT
TCP	192.168.0.161:4137	149.154.167.216:http	TIME_WAIT
TCP	192.168.0.161:4138	162.159.136.232:https	ESTABLISHED
TCP	192.168.0.161:4139	149.154.167.216:https	TIME_WAIT
TCP	192.168.0.161:4140	149.154.167.216:http	TIME_WAIT
TCP	192.168.0.161:4141	waw02s22-in-f13:https	ESTABLISHED
TCP	192.168.0.161:4142	waw07s06-in-f10:https	ESTABLISHED
TCP	192.168.0.161:4143	waw02s16-in-f3:https	ESTABLISHED
TCP	192.168.0.161:4144	mil02s06-in-f14:https	ESTABLISHED
TCP	192.168.0.161:4145	ej-in-f139:https	ESTABLISHED
TCP	192.168.0.161:4147	waw02s17-in-f22:https	ESTABLISHED
TCP	192.168.0.161:4148	78.26.240.109:https	ESTABLISHED
TCP	192.168.0.161:4149	78.26.240.109:https	ESTABLISHED

Рис. 2. Процес ініціювання мережевої активності в операційній системі

Джерело: авторська розробка

дослідження поведінки таких процесів обов'язковим для оптимізації та безпеки. Тому результати, представлені на рисунку, використовуються як відправна точка для подальших досліджень початку мережевої активності та розробки методів її аналізу, які можна інтегрувати в автоматизовані системи моніторингу та реагування на мережеві загрози.

Практичні аспекти роботи з реальними мережевими середовищами, висвітлені Jiang W., Zhang B., Zhu Q., Liao C., Wang W. [7, с. 5], підкреслюють важливість валідації результатів netstat на основі емпіричних даних. Це особливо актуально для розробки ефективних алгоритмів фільтрації та аналізу.

Сучасні методи виявлення мережевих вторгнень, розроблені Balakrishnapillai J., Ajayan A.K.P., Kurian A., Sabu A., Majeed A. [8, с. 020045], можуть бути інтегровані з функціональністю netstat для створення комплексних систем безпеки. Альтернативні підходи, запропоновані S D. B., A Ms. D. [9, с. 7], демонструють можливості використання нейронних мереж для аналізу паттернів мережевої активності, виявлених за допомогою netstat.

Інноваційні методи виявлення аномалій, представлені Nam S., Jeong E., Hong J.W. [10, с. 4], базуються на аналізі логів мережевої активності, які можуть бути ефективно генеровані з використанням розширених можливостей netstat. Прогнозування мережевого трафіку, досліджене Lv Q., Chang Y., Li T., Ge J. [11, с. 111558], потребує якісних вхідних даних, які netstat може надати у структурованому вигляді.

Розширені можливості netstat включають параметр -s для відображення статистики по протоколах, що дозволяє аналізувати загальну ефективність мережевої підсистеми. Параметр -r відображає таблицю маршрутизації, що критично важливо для діагностики проблем з мережевою доступністю. Комбінування різних параметрів створює потужний інструментарій для комплексного аналізу мережевого стану системи.

Автоматизація роботи з netstat передбачає розробку скриптів для регулярного збору даних та їх подальшої обробки. Періодичне виконання команд netstat з різними параметрами дозволяє створювати часові ряди мережевої активності, що надає можливість виявляти тренди та аномалії. Інтеграція з системами логування забезпечує довгострокове зберігання історії мережевої активності для ретроспективного аналізу.

Висновки

Проведене дослідження демонструє фундаментальну важливість утиліти netstat для ініціювання та моніторингу мережевої активності в операційній системі Windows. Систематичний підхід до використання розширених параметрів netstat забезпечує комплексний аналіз мережевого стану системи. Розроблена класифікація станів TCP-з'єднань надає практичний інструментарій для інтерпретації результатів діагностики.

Запропонована архітектура системи моніторингу на базі netstat демонструє можливості інтеграції традиційних утиліт з сучасними підходами до автоматизації та аналізу даних. Методики автоматизованої обробки вихідних даних netstat створюють основу для розробки ефективних систем мережевого моніторингу.

Перспективи подальших досліджень включають розробку алгоритмів машинного навчання для автоматизованого виявлення аномалій на основі даних netstat, інтеграцію з хмарними системами моніторингу та створення адаптивних систем оповіщення. Результати дослідження мають практичне значення для системних адміністраторів та фахівців з мережевої безпеки, надаючи конкретні рекомендації щодо оптимального використання netstat у різних сценаріях діагностики.

Список використаної літератури

1. Бойко В. Д., Бевза В. І., Слатвінська В. М. Використання штучного інтелекту для оптимізації коду на Python з інтеграцією інструментів налагодження PDB / В. Д. Бойко, В. І. Бевза, В. М. Слатвінська // Наука і техніка сьогодні: Серія «Техніка». 2025. № 7 (48). С. 1258–1271. DOI: https://doi.org/10.52058/2786-6025-2025-7(48)-1258-1271

2. Бевза В. І., Слатвінська В. М. Вплив збою Crowdstrike на мега-витік паролів: чи є зв'язок? Ч. 1 / В. І. Бевза, В. М. Слатвінська // Вісник Хмельницького національного університету. Серія: Технічні науки. 2024. № 4. С. 332–338. DOI: https://doi.org/10.31891/2307-5732-2024-339-4-52 URL: https://heraldts.khmnu.edu.ua/index.php/heraldts/article/view/257

3. Бевза В. І., Слатвінська В. М. Вплив збою Crowdstrike на мега-витік паролів: чи є зв'язок? Ч. 2 / В. І. Бевза, В. М. Слатвінська // Вісник Хмельницького національного університету. Серія: Технічні науки. 2024. Том 341. № 5. С. 248–259. DOI: https://doi.org/10.31891/2307-5732-2024-341-5-36 URL: https://heraldts.khmnu.edu.ua/index.php/heraldts/article/view/300

4. Zhong Y., Zhang Y., Zhang J., Wan L. Research on resilience measurement and optimization of networked command information system / Y. Zhong, Y. Zhang, J. Zhang, L. Wan // Reliability Engineering & System Safety. 2025. Vol. 261. Art. 111048. DOI: https://doi.org/10.1016/j.ress.2025.111048 URL: https://www.sciencedirect.com/science/article/pii/S0951832025002492

5. Romero M. L., Suyama R. Towards Network Data Analytics in 5G Systems and Beyond (Version 1) / M. L. Romero, R. Suyama // arXiv. 2025. DOI: https://doi.org/10.48550/ARXIV.2506.04860 URL: https://arxiv.org/abs/2506.04860v1

6. Beltiukov R., Bhattaram K., Cheng E., Kanigicherla V., Singh A., Thampiratwong K., Gupta A. Leveraging Large Language Models to Contextualize Network Measurements (Version 2) / R. Beltiukov, K. Bhattaram, E. Cheng, V. Kanigicherla, A. Singh, K. Thampiratwong, A. Gupta // arXiv. 2025. DOI: https://doi.org/10.48550/ARXIV.2505.19305 URL: https://arxiv.org/abs/2505.19305v2

7. Jiang W., Zhang B., Zhu Q., Liao C., Wang W. A Real Network Environment Dataset for Traffic Analysis / W. Jiang, B. Zhang, Q. Zhu, C. Liao, W. Wang // Scientific Data. 2025. Vol. 12, No. 1. DOI: https://doi.org/10.1038/s41597-025-04876-2 URL: https://www.nature.com/articles/s41597-025-04876-2

8. Balakrishnapillai J., Ajayan A. K. P., Kurian A., Sabu A., Majeed A. Network intrusion detection using novel DSSTE / J. Balakrishnapillai, A. K. P. Ajayan, A. Kurian, A. Sabu, A. Majeed // AIP Conference Proceedings. 2025. Vol. 3260. Art. 020044. DOI: https://doi.org/10.1063/5.0259047 URL: https://is.gd/A27Mrf

9. Brindha S., Dhamayanthi A. Network Based Intrusion Detection using Convolutional Neural Network / S. Brindha, A. Dhamayanthi // International Journal of Scientific Research in Engineering and Management. 2025. Vol. 09, No. 03. P. 1–9. DOI: https://doi.org/10.55041/ijsrem42812 URL: https://is.gd/JMt6yI

10. Nam S., Jeong E., Hong J. W. Log-TF-IDF and NETCONF-Based Network Switch Anomaly Detection / S. Nam, E. Jeong, J. W. Hong // International Journal of Network Management. 2025. Vol. 35, No. 1. DOI: https://doi.org/10.1002/nem.2322 URL: https://onlinelibrary.wiley.com/doi/10.1002/nem.2322

11. Lv Q., Chang Y., Li T., Ge J. Betastack: Enhancing base station traffic prediction with network-specific Large Language Models / Q. Lv, Y. Chang, T. Li, J. Ge // Computer Networks. 2025. Vol. 270. Art. 111557. DOI: https://doi.org/10.1016/j.comnet.2025.111557 URL: https://www.sciencedirect.com/science/article/abs/pii/S1389128625005249

References

1. Boyko, V. D., Bevza, V. I., & Slatvinska, V. M. (2025). Vykorystannia shtuchnoho intelektu dlia optymizatsii kodu na Python z intehratsiieiu instrumentiv nalahodzhennia PDB [Use of artificial intelligence for Python code optimization with PDB debugging tools integration]. Nauka i tekhnika sohodni: Seriia “Tekhnika”, Science and Technology Today: Technical Series (7), 1258–1271. DOI: https://doi.org/10.52058/2786-6025-2025-7(48)-1258-1271 [in Ukrainian].

2. Bevza, V. I., & Slatvinska, V. M. (2024). Vplyv zboiu Crowdstrike na meha-vytik paroliv: chy ye zviazok? Ch. 1 [Impact of the Crowdstrike outage on the mega password leak: Is there a connection? Part 1]. Visnyk Khmelnytskoho natsionalnoho universytetu. Seriia: Tekhnichni nauky, Bulletin of Khmelnytskyi National University. Technical Sciences, (4), 332–338. DOI: https://doi.org/10.31891/2307-5732-2024-339-4-52 Retrieved from https://heraldts.khmnu.edu.ua/index.php/heraldts/article/view/257 [in Ukrainian].

3. Bevza, V. I., & Slatvinska, V. M. (2024). Vplyv zboiu Crowdstrike na meha-vytik paroliv: chy ye zviazok? Ch. 2 [Impact of the Crowdstrike outage on the mega password leak: Is there a connection? Part 2]. Visnyk Khmelnytskoho natsionalnoho universytetu. Seriia: Tekhnichni nauky. Bulletin of Khmelnytskyi National University. Technical Sciences, 341(5), 248–259. DOI: https://doi.org/10.31891/2307-5732-2024-341-5-36 Retrieved from https://heraldts.khmnu.edu.ua/index.php/heraldts/article/view/300 [in Ukrainian].

4. Zhong, Y., Zhang, Y., Zhang, J., & Wan, L. (2025). Research on resilience measurement and optimization of networked command information system. Reliability Engineering & System Safety, 261, 111048. DOI: https://doi.org/10.1016/j.ress.2025.111048 Retrieved from https://www.sciencedirect.com/science/article/pii/S0951832025002492 [in English].

5. Romero, M. L., & Suyama, R. (2025). Towards network data analytics in 5G systems and beyond (Version 1). arXiv. DOI: https://doi.org/10.48550/ARXIV.2506.04860 Retrieved from https://arxiv.org/abs/2506.04860v1 [in English].

6. Beltiukov, R., Bhattaram, K., Cheng, E., Kanigicherla, V., Singh, A., Thampiratwong, K., & Gupta, A. (2025). Leveraging large language models to contextualize network measurements (Version 2). arXiv. DOI: https://doi.org/10.48550/ARXIV.2505.19305 Retrieved from https://arxiv.org/abs/2505.19305v2 [in English].

7. Jiang, W., Zhang, B., Zhu, Q., Liao, C., & Wang, W. (2025). A real network environment dataset for traffic analysis. Scientific Data, 12(1). DOI: https://doi.org/10.1038/s41597-025-04876-2 Retrieved from https://www.nature.com/articles/s41597-025-04876-2 [in English].

8. Balakrishnapillai, J., Ajayan, A. K. P., Kurian, A., Sabu, A., & Majeed, A. (2025). Network intrusion detection using novel DSSTE. AIP Conference Proceedings, 3260, 020044. DOI: https://doi.org/10.1063/5.0259047 Retrieved from https://is.gd/A27Mrf [in English].

9. Brindha, S., & Dhamayanthi, A. (2025). Network based intrusion detection using convolutional neural network. International Journal of Scientific Research in Engineering and Management, 9(3), 1–9. DOI: https://doi.org/10.55041/ijsrem42812 Retrieved from https://is.gd/JMt6yI [in English].

10. Nam, S., Jeong, E., & Hong, J. W. (2025). Log-TF-IDF and NETCONF-based network switch anomaly detection. International Journal of Network Management, 35(1). DOI: https://doi.org/10.1002/nem.2322 Retrieved from https://onlinelibrary.wiley.com/doi/10.1002/nem.2322 [in English].

11. Lv, Q., Chang, Y., Li, T., & Ge, J. (2025). Betastack: Enhancing base station traffic prediction with network-specific large language models. Computer Networks, 270, 111557. DOI: https://doi.org/10.1016/j.comnet.2025.111557 Retrieved from https://is.gd/8bjNkw [in English].

Дата першого надходження рукопису до видання: 26.09.2025
Дата прийнятого до друку рукопису після рецензування: 22.10.2025
Дата публікації:

УДК 004.89:623 DOI

Д. О. БОРОВИК
аспірант кафедри комп'ютерних наук
Сумський державний університет
ORCID: 0009-0005-3642-9444

М. О. ЗАРЕЦЬКИЙ
доктор філософії,
старший викладач кафедри комп'ютерних наук
Сумський державний університет
ORCID: 0000-0001-9117-5604

ВИКОРИСТАННЯ ДРОНІВ ТА ШТУЧНОГО ІНТЕЛЕКТУ У ВІЙСЬКОВІЙ СТРАТЕГІЇ УКРАЇНИ

Досліджено технічні аспекти застосування безпілотних літальних апаратів (БПЛА) у Збройних Силах України (ЗСУ) для виконання завдань розвідки, ударних операцій, логістики та радіоелектронної боротьби (РЕБ).

Мета статті – проаналізувати технічні характеристики, функціональність і програмне забезпечення (ПЗ) українських безпілотних апаратів, оцінити ефективність їхнього використання без інтегрованого штучного інтелекту (ШІ) та визначити перспективи часткової або повної автоматизації для підвищення автономності в умовах реальних бойових дій.

Результати свідчать, що попри ефективність дронів у базових завданнях, їхня робота ускладнюється порушенням зв'язку, зниженням автономності на 20–30 % через кліматичні фактори та обмеженою гнучкістю ПЗ. Розвідувальні дрони забезпечують спостереження, ударні – точкове ураження цілей, логістичні – оперативну доставку вантажів, а РЕБ-безпілотні апарати – блокування ворожих каналів зв'язку. Водночас ПЗ не передбачає автономної навігації, корекції маршрутів або адаптації до змін у бойовому середовищі. Просте ПЗ не дає змоги безпілотним апаратам функціонувати без постійного керування оператором, що знижує їхню ефективність у зонах активного РЕБ. Розглянуто переваги та обмеження ключових моделей БПЛА, зокрема «Лелека-100», «Racoon», «Bukovel», які становлять основу тактичного арсеналу. Проаналізовано можливості їхнього програмного забезпечення (ПЗ) та апаратної платформи в умовах активного впливу ворожих РЕБ-систем і складних кліматичних умов. Акцент зроблено на практичних бар'єрах автоматизації, пов'язаних із енергоспоживанням, вагою дронів і недостатньою обчислювальною потужністю, а також на шляхах поступового впровадження елементів штучного інтелекту (ШІ) через модульні рішення.

Ефективність БПЛА ЗСУ значною мірою залежить від вдосконалення їхніх технічних характеристик і ПЗ. Перспективними напрямами є впровадження інерційних навігаційних систем, гібридних джерел живлення, адаптивного ПЗ для РЕБ-дронів та локальної обробки даних за допомогою спрощених ШІ-алгоритмів. Поступове впровадження таких рішень дозволить підвищити автономність і надійність дронів без значного нарощення ресурсної бази.

Ключові слова: *дрони, штучний інтелект, програмне забезпечення, РЕБ, автономність.*

D. O. BOROVYK
Postgraduate Student at the Department of Computer Science,
Sumy State University
ORCID: 0009-0005-3642-9444

M. O. ZARETSKYI
Doctor of Philosophy,
Senior Lecturer at the Department of Computer Science
Sumy State University
ORCID: 0000-0001-9117-5604

THE USE OF DRONES AND ARTIFICIAL INTELLIGENCE IN UKRAINE'S MILITARY STRATEGY

The article examines the technical aspects of the use of unmanned aerial vehicles (UAVs) in the Armed Forces of Ukraine (AFU) to perform reconnaissance, strike operations, logistics, and electronic warfare (EW) tasks.

The purpose of the article is to analyze the technical characteristics, functionality, and software of Ukrainian drones, to assess the effectiveness of their use without integrated Artificial Intelligent (AI), and to determine the prospects for partial automation to increase autonomy in real-world combat.

© Боровик Д. О., Зарецький М. О., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

The results show that despite the effectiveness of drones in basic tasks, their work is complicated by communication disruptions, a 20–30 % decrease in autonomy due to climatic factors, and limited software flexibility. Reconnaissance drones provide surveillance, strike drones provide targeted destruction, logistics drones provide rapid delivery of goods, and electronic warfare drones block enemy communication channels. However, the software does not provide for autonomous navigation, route correction, or adaptation to changes in the combat environment. Simple software does not allow drones to function without constant operator control, which reduces their effectiveness in areas of active electronic warfare. The advantages and limitations of key models of drones Leleka-100, Racoon, Bukovel, which form the basis of the tactical arsenal, are considered. The author analyzes the capabilities of their software and hardware platforms in the face of active influence of enemy electronic warfare systems and difficult climatic conditions. The emphasis is placed on the practical barriers to automation related to power consumption, weight of drones, and insufficient computing power, as well as on ways to gradually introduce elements of AI through modular solutions.

The effectiveness of the UAVs of the Armed Forces of Ukraine largely depends on improving their technical characteristics and software. Promising areas include the introduction of inertial navigation systems, hybrid power sources, adaptive software for electronic warfare drones, and local data processing using simplified AI algorithms. The gradual introduction of such solutions will increase the autonomy and reliability of drones without a significant increase in the resource base.

Key words: *drones, artificial intelligence, software, EPR, autonomy, unmanned aerial vehicle.*

Постановка проблеми

Безпілотні літальні апарати відіграють важливу роль у сучасній військовій стратегії України, виконуючи завдання з аеророзвідки, цілевказання, логістики та радіоелектронної боротьби. Втім, попри їхню ефективність, існує низка системних технічних обмежень, які стримують потенціал дронів у бойових умовах.

По-перше, ключовою проблемою залишається обмежена автономність польоту. Наявні моделі дронів мають низьку енергоефективність, що обмежує тривалість їх роботи та унеможливлює використання потужного обчислювального обладнання. Акумулятори швидко втрачають заряд при низьких температурах або під час експлуатації енергоємних модулів – наприклад, гіростабілізованих камер або багатоканальних передавачів.

По-друге, дрони критично залежні від супутникової навігації. Умови активної РЕБ часто призводять до втрати GPS-сигналу, що паралізує керування та виконання завдань. Альтернативні системи – як-от інерційна навігація – не впроваджуються масово через вагові обмеження, високу похибку недорогих модулів та необхідність постійного калібрування.

По-третє, більшість БПЛА позбавлені реального «інтелекту» на борту. Типове програмне забезпечення базується на фіксованих маршрутах і ручному управлінні, що не дозволяє дрону адаптуватися до змін у реальному часі. Вся обробка відеосигналу відбувається на землі, що призводить до затримок у прийнятті рішень, інформаційного перевантаження операторів і зниження бойової ефективності.

По-четверте, відсутність гнучкого ПЗ обмежує роботу дронів у змінному частотному середовищі. Наприклад, дрони для РЕБ не можуть автоматично сканувати спектр і адаптувати глушіння до нових частот. Розвідувальні та ударні БПЛА зазвичай не мають резервних каналів зв'язку чи алгоритмів зниження залежності від центрального керування.

По-п'яте, конструктивні обмеження стримують інтеграцію сучасного ШІ. Дрони вагою до 10 кг не здатні нести повноцінні процесори або енергоємну пам'ять, необхідну для складних алгоритмів машинного навчання. Це виключає застосування onboard AI, що міг би забезпечити автономну класифікацію об'єктів, ухвалення рішень і адаптацію до ситуації.

Отже, технічна суть проблеми полягає у необхідності комплексної модернізації апаратного і програмного забезпечення БПЛА. Така модернізація має враховувати можливість автономної роботи у складних умовах – за відсутності зв'язку, під дією РЕБ, у складному рельєфі – без критичного збільшення маси або споживання енергії. Подальший розвиток систем повинен орієнтуватися на впровадження легковагових, енергоефективних мікропроцесорів, адаптивного ПЗ з елементами штучного інтелекту та альтернативних навігаційних рішень.

Аналіз останніх досліджень і публікацій

Сучасні дослідження у сфері військового застосування БПЛА фокусуються переважно на їх класифікації, аналізі тактико-технічних характеристик, функціональності програмного забезпечення та можливостях використання ШІ у бойових умовах. При цьому основна увага зосереджена на традиційних параметрах, таких як дальність польоту, вантажність, точність ураження, залишаючи поза належною увагою аспекти програмної автономії, обробки даних на борту та адаптивності систем керування.

Дослідники відзначають, що інтеграція ШІ в системи управління БПЛА відкриває можливості для автоматизації прийняття рішень та аналізу даних у реальному часі. Однак, як підкреслюється у роботах Литвиненка М.І. та ін., ключовим обмеженням залишається недостатня обчислювальна потужність апаратних модулів, особливо у легких БПЛА, типу «Лелека-100», де критично важлива енергоефективність [1].

Іншою перешкодою є залежність автономної навігації від стабільного каналу зв'язку. Зокрема, Joshi A. та ін. акцентують на тому, що в умовах дії ворожих систем радіоелектронної боротьби (РЕБ) порушення зв'язку призводить до втрати керованості та недосягнення місії [2]. Попри перспективність технологій 5G, як зазначають Dash B. та колеги, стабільність передачі даних у зоні бойових дій залишається недостатньою, а відсутність автономної обробки інформації знижує боєздатність апаратів [3].

Науковці також систематизують типи дронів, що використовуються ЗСУ. Ivanchenko O. та співавтори наводять технічні характеристики дронів: дальність польоту до 50 км, точність ударів до 80 %, радіус дії РЕБ-модулів – до 50 км [4]. Проте, акцент здебільшого зроблено на апаратному забезпеченні, тоді як проблеми обмеженої функціональності ПЗ, відсутності адаптивності та автономності залишаються поза увагою.

Варто відзначити, що навіть багатоцільові моделі, як-от логістичні БПЛА типу «Valkyrie», наразі не здатні адаптувати маршрути залежно від змін бойової обстановки. Це свідчить про відсутність на борту систем ситуаційної обізнаності та гнучких алгоритмів маршрутизації [6].

Окремі дослідники, як Sorochkin O. та ін., розглядають ШІ як можливий інструмент для підтримки автономної навігації у випадку втрати зв'язку з оператором, особливо для ударних моделей БПЛА. Цей підхід дозволив би частково компенсувати дію РЕБ-засобів супротивника [7].

Актуальним напрямом є також ройові технології, де група дронів діє як єдина синхронізована система. Радованович М. та колеги підкреслюють, що реалізація такої архітектури потребує високопродуктивної обчислювальної інфраструктури та ПЗ з високим рівнем координації – компонентів, які наразі відсутні в українських зразках техніки [8].

У контексті обробки відеопотоків, Трофименко О. та ін. вказують на потенціал алгоритмів комп'ютерного зору для автоматизованої розвідки. Водночас дослідники зауважують, що ці алгоритми мають бути максимально спрощеними через обмеження ресурсів на борту БПЛА [9].

Гуржій С. В. підсумовує, що ключовими факторами підвищення ефективності БПЛА є не лише їх технічні характеристики, а й здатність ПЗ до адаптації в умовах невизначеності. Саме ця здатність наразі не відповідає актуальним потребам Збройних Сил України [10].

Таким чином, попри наявність значного масиву публікацій, більшість досліджень зосереджуються на конвенційних параметрах дронів, залишаючи недооціненими проміжні рішення в галузі автоматизації – часткову автономність, локальну обробку даних, адаптивне програмне забезпечення. Саме ці напрями потребують поглибленої розробки з урахуванням реальних обмежень апаратної бази українських БПЛА.

Формулювання мети дослідження

Мета статті: комплексний аналіз видів безпілотних літальних апаратів, які використовуються Збройними Силами України. Особлива увага приділяється технічним характеристикам, можливостям програмного забезпечення, а також потенціалу застосування інтелектуальних технологій для підвищення ефективності виконання завдань в умовах протидії засобам радіоелектронної боротьби.

Викладення основного матеріалу дослідження

Безпілотні літальні апарати формують основу сучасної військової стратегії Збройних Сил України, виконуючи завдання розвідки, ударних операцій, логістики та радіоелектронної боротьби. Завдяки компактним розмірам і модульній конструкції вони оперативно реагують на зміни бойового середовища, перевершуючи важку техніку в маневреній війні. Литвиненко М. І. та ін. підкреслюють, що впровадження алгоритмів комп'ютерного зору значно прискорює ідентифікацію цілей, але сучасні мікропроцесори дронів не забезпечують необхідної обчислювальної потужності [1, с. 16]. Основними обмеженнями залишаються чутливість до радіоелектронних перешкод, недостатня тривалість польоту та слабка адаптивність програмного забезпечення до динамічних умов.

Розвідувальні дрони, такі як «Лелека-100», відіграють вирішальну роль у зборі розвідданих і коригуванні артилерійського вогню на дистанціях до 50 км. Оснащені електрооптичними камерами з роздільною здатністю 1920 × 1080 пікселів і тепловізорами з чутливістю 50 мілікельвінів, вони виявляють об'єкти в темряві та за складних погодних умов. Terzić M. зазначає, що «Лелека-100» точно визначає координати скупчень техніки на відстані 40 км, забезпечуючи артилерійські удари в реальному часі [5, с. 365]. Камера з 10-кратним оптичним збільшенням фіксує об'єкти розміром від 1 метра, а зашифрований канал зв'язку на частоті 2,4 ГГц передає дані з похибкою до 10 метрів. Однак ручна обробка відеопотоку оператором спричиняє затримки до 8 секунд, перевантажуючи наземні станції керування [2, с. 140]. Електричний двигун потужністю 600 ват забезпечує до 6 годин польоту, але літій-полімерний акумулятор ємністю 16 ампер-годин втрачає до 25 % ефективності за температур нижче –10 °C, що ускладнює зимові операції.

Ударні дрони, зокрема FPV-модель «Racoon», призначені для високоточного знищення ворожих укріплень і техніки. З масою 2 кг і максимальною швидкістю 180 км/год вони доставляють вибухові заряди до 0,5 кг із точністю ураження 1 метр за стабільного зв'язку. Камера з кутом огляду 120 градусів і частотою 60 кадрів за секунду забезпечує чітке зображення навіть під час швидкісного польоту. Sorochkin O. та ін. вказують, що відсутність автоматичного розпізнавання цілей змушує операторів покладатися на аналоговий відеоканал із частотою

5,8 ГГц, який блокується РЕБ-системами на відстані 12 км, а затримка сигналу 0,12 секунди знижує ефективність у зонах активних бойових дій [7, с. 217]. Інерційні навігаційні модулі на базі мікроелектромеханічних систем могли б компенсувати втрату зв'язку, але їхня похибка 0,1 м/с^2 обмежує точність при тривалій автономній роботі [7, с. 218], рис. 1.

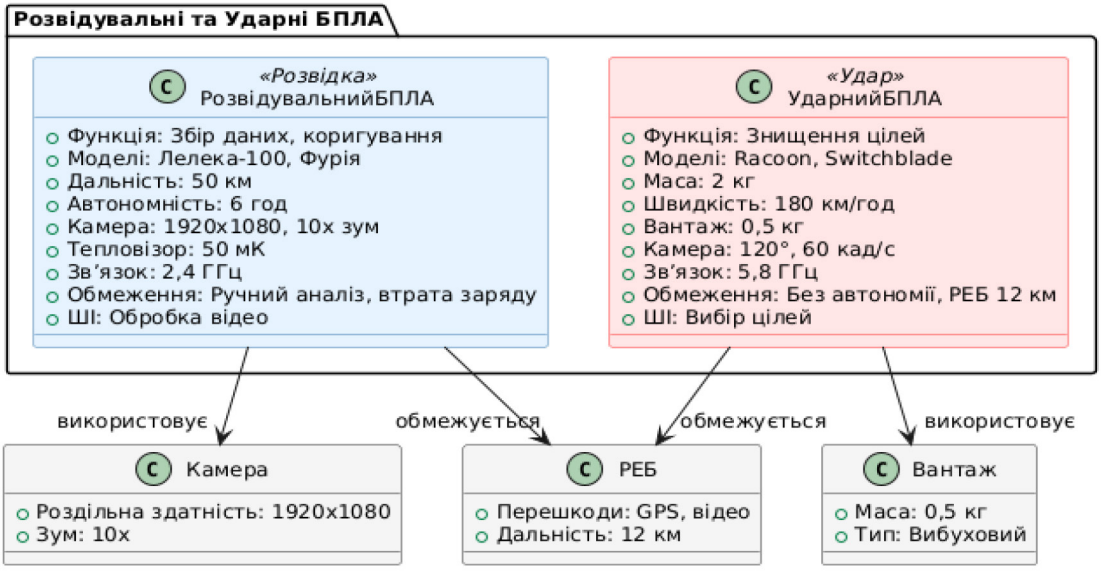


Рис. 1 Аналіз розвідувальних і ударних БПЛА

Джерело: авторська розробка в PlantUML Online Editor

Логістичні дрони, наприклад «Valkyrie», забезпечують швидке транспортування вантажів масою до 20 кг на відстань 30 км за 25 хвилин. Квадрокоптер із чотирма роторами загальною потужністю 3200 ват зберігає стабільність при поривах вітру до 12 м/с. Dash B. та ін. наголошують, що нестабільність 5G-зв'язку з пропускною здатністю 8 мегабіт за секунду ускладнює передачу телеметрії в реальному часі, а програмне забезпечення не враховує динамічні перешкоди, такі як РЕБ чи зміна погоди [3, с. 314]. Літій-іонний акумулятор ємністю 12 ампер-годин підтримує до 2 годин польоту. Алгоритми планування траєкторій, такі як Rapidly-exploring Random Tree, могли б оптимізувати маршрути, але мікроконтролер із частотою 150 МГц і пам'яттю 256 КБ не забезпечує необхідної продуктивності, рис. 2.

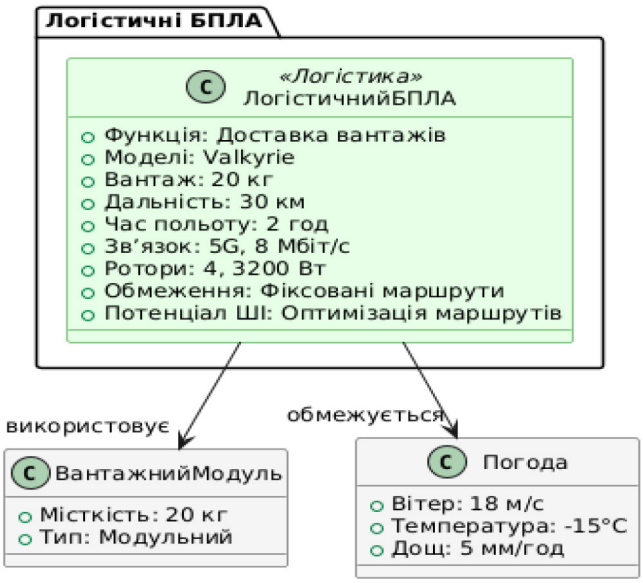


Рис. 2 Аналіз логістичних БПЛА

Джерело: авторська розробка в PlantUML Online Editor

Дрони радіоелектронної боротьби, такі як «Піранья», створюють перешкоди в діапазоні частот 300–3000 МГц на відстані до 50 км, придушуючи ворожі системи зв'язку та навігації. З масою 10 кг і спрямованим передавачем потужністю 50 ват вони ефективно глушать GPS і радіоканали за допомогою антени з кутом випромінювання 30 градусів. Літій-полімерний акумулятор ємністю 20 ампер-годин забезпечує 4 години роботи. Sorochkin O. та ін. зазначають, що ручне налаштування частот, яке триває до 90 секунд, знижує оперативність реагування на сучасні адаптивні РЕБ-системи через відсутність автоматичного сканування спектру [7, с. 219]. Оновлення 2024 року розширило частотний діапазон, але обмеження пам'яті мікроконтролера до 512 КБ унеможливлює обробку великих масивів даних у реальному часі, рис. 3.

Рис. 3 Аналіз РЕБ та Дорозвідка БПЛА

Джерело: авторська розробка в PlantUML Online Editor

Апаратне забезпечення дронів створює значні бар'єри для впровадження передових технологій. Легкі моделі масою 5–10 кг оснащені мікроконтролерами з частотою до 200 МГц і оперативною пам'яттю 512 КБ, що не дозволяє виконувати складні обчислення, такі як обробка відеопотоку чи нейронних мереж. Criollo L. та ін. підкреслюють, що потужніші процесори, наприклад NVIDIA Jetson Nano з продуктивністю 0,5 терафлопс, споживають 15 ват, скорочуючи автономність на 25 % і додаючи 300 г ваги, що порушує аеродинамічний баланс [6, с. 117–118]. Joshi A. та ін. зазначають, що літій-іонні акумулятори втрачають до 30 % ємності при температурах нижче –15 °C, а РЕБ-системи, генеруючи імпульси на частоті 1,5 ГГц, виводять із ладу GPS-навігацію за 5–10 секунд [2, с. 140–141]. Інерційні системи на основі акселерометрів і гіроскопів мають похибку 0,05 м/с^2, що накопичується до 50 метрів за 10 хвилин, роблячи їх непридатними для тривалих автономних місій.

Штучний інтелект відкриває можливості для автономних операцій дронів (рис. 4–8). Нейронна мережа YOLOv8 здатна ідентифікувати ворожу техніку з точністю 92 % на відстані 35 км навіть у зонах РЕБ, але вимагає процесора з продуктивністю 2 терафлопс і споживанням 20 ват, що є недосяжним для легких дронів. Литвиненко М. І. та ін. пропонують легші алгоритми, такі як Haar-каскади, які виявляють рухомі об'єкти з точністю 75 %, використовуючи 50 МБ пам'яті та 2 вати енергії, зменшуючи обсяг переданих даних до 5 мегабіт за секунду [1, с. 16]. Трофименко О. та ін. попереджають, що ці алгоритми втрачають до 50 % точності через шуми, спричинені РЕБ на частоті 5,8 ГГц [9, с. 168]. Sorochkin O. та ін. описують концепцію дорозвідки, де дрон із модулем MobileNet аналізує теплові сигнатури на відстані 30 км, передаючи лише координати цілей, що знижує залежність від зв'язку до 1 мегабіта за секунду, але потребує процесора з продуктивністю 0,8 терафлопс [7, с. 218].

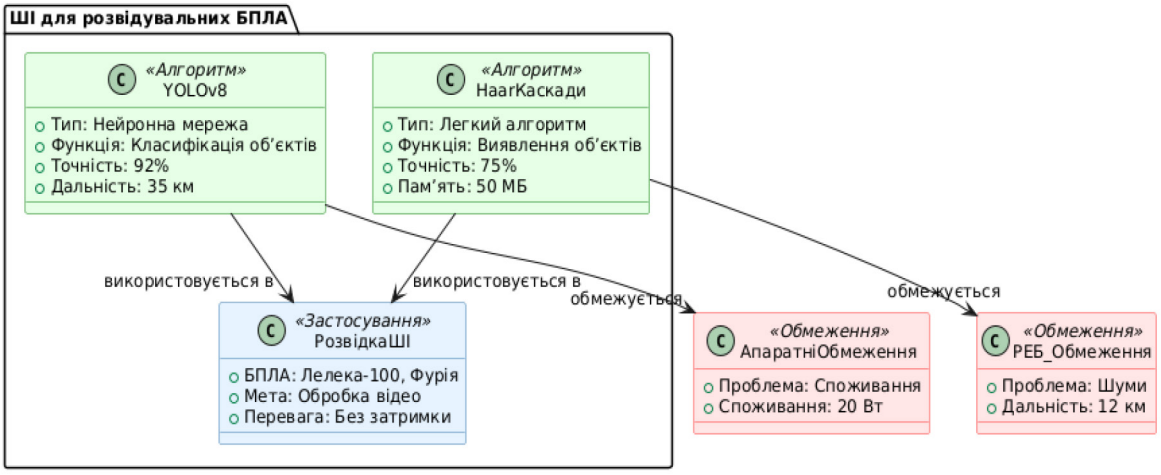


Рис. 4 ШІ для розвідувальних БПЛА

Джерело: авторська розробка в PlantUML Online Editor

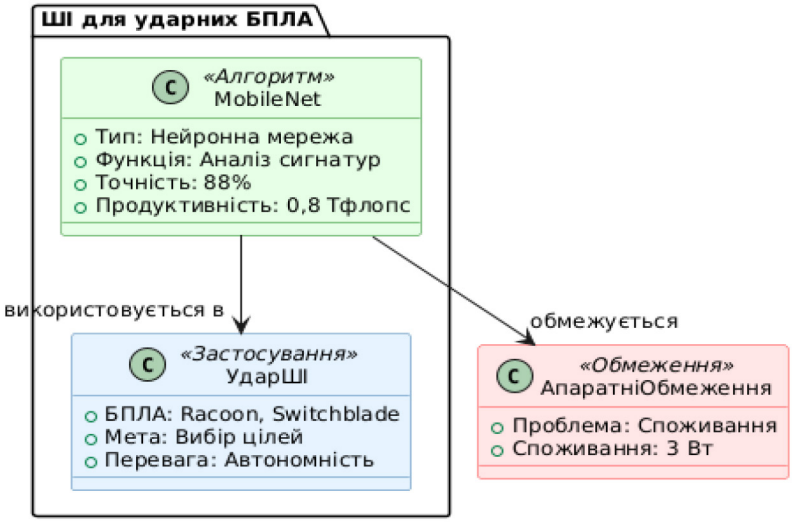


Рис. 5 ШІ для ударних БПЛА

Джерело: авторська розробка в PlantUML Online Editor

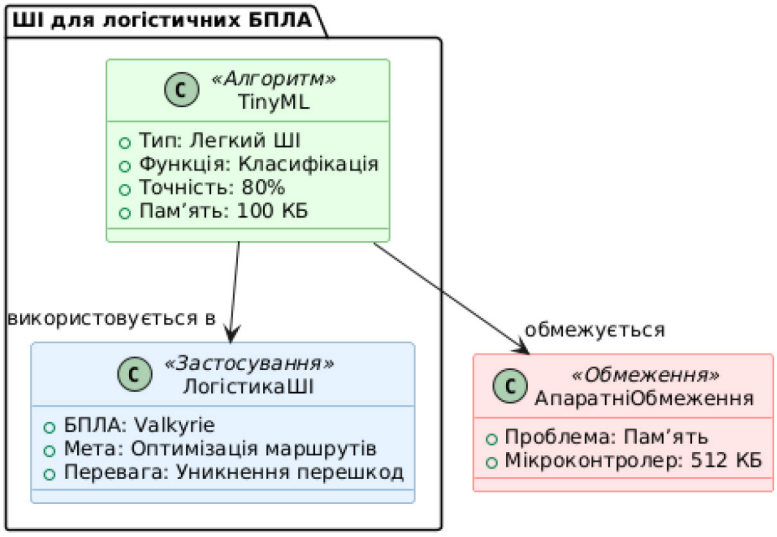


Рис. 6. ШІ для логістичних БПЛА

Джерело: авторська розробка в PlantUML Online Editor

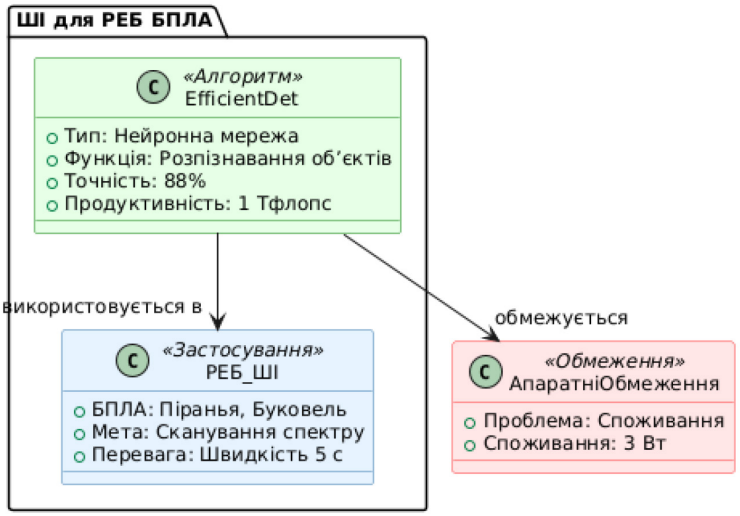


Рис. 7 ШІ для РЕБ БПЛА

Джерело: авторська розробка в PlantUML Online Editor

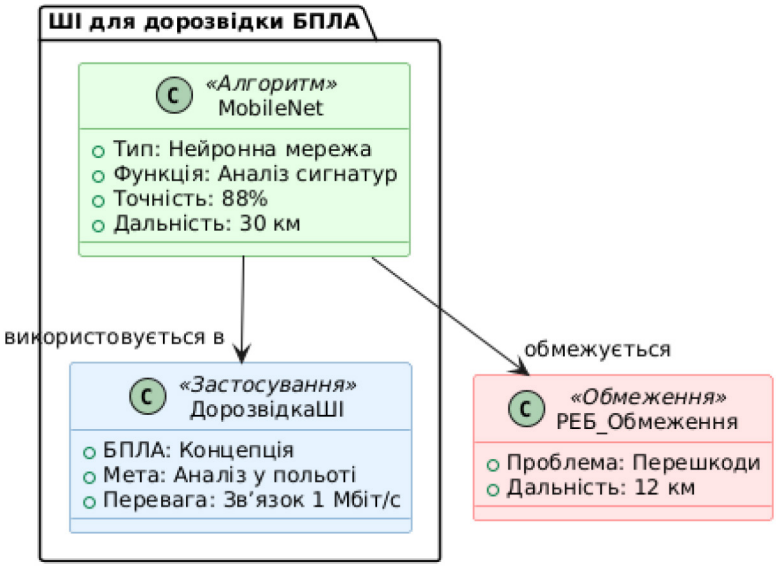


Рис. 8. ШІ для дорозвідки БПЛА

Джерело: авторська розробка в PlantUML Online Editor

Ройові технології дозволяють координувати групи дронів для виконання складних завдань, таких як створення тривимірних карт бойового простору. Radovanović M. та ін. зазначають, що вісім дронів із mesh-мережею на частоті 2,4 ГГц можуть сканувати територію площею 80 км^2, обмінюючись даними через протокол ZigBee із пропускною здатністю 250 кілобіт за секунду [8, с. 58]. РЕБ-системи скорочують дальність зв'язку до 1,5 км, а мікроконтролери з частотою 200 МГц не здатні синхронізувати більше трьох апаратів. Модульна координація двох дронів для розвідки та удару є більш досяжною, використовуючи послідовний інтерфейс зі швидкістю 115 200 біт за секунду. Енергоспоживання залишається критичним обмеженням. Joshi A. та ін. пропонують літій-сірчані акумулятори з енергоємністю 300 ват-годин/кг, які подовжують політ до 7 годин, але втрачають 20 % ефективності при –20 °C [2, с. 141]. Зашифровані канали з алгоритмом AES-256 на частоті 900 МГц захищають від перехоплення, але споживають додаткові 3 вати, скорочуючи автономність на 10 %.

Погодні умови суттєво впливають на ефективність дронів. Сильний вітер швидкістю 18 м/с відхиляє траєкторію на 10 метрів за хвилину, а дощ інтенсивністю 5 мм/год накопичує вологу на лінзах камер, погіршуючи видимість. Низькі температури знижують продуктивність мікроконтролерів на 15 % при –15 °C. Criollo L. та ін. зазначають, що гібридні двигуни, які поєднують електричну енергію з мікротурбінами, забезпечують стабільну роботу при –25 °C, але їхня маса 1,2 кг обмежує вантажопідйомність легких моделей [6, с. 118]. Гідрофобні покриття для камер і теплові ізолятори для акумуляторів зменшують вплив погоди, але додають 150 г ваги.

Розвідувальні дрони, такі як «Лелека-100» і «Фурія», забезпечують спостереження на відстані до 50 км з автономністю до 6 годин, але ручний аналіз відеопотоку спричиняє затримки в обробці даних. Ударні дрони, зокрема «Racoon» і «Switchblade», демонструють точність ураження 80 % за стабільного зв'язку, але не мають алгоритмів автономного наведення. Логістичні дрони, такі як «Valkyrie», доставляють вантажі до 20 кг, але не адаптуються до динамічних перешкод. РЕБ-дрони, наприклад «Піранья» і «Bukovel», глушать сигнали на відстані 50 км, але їхнє програмне забезпечення вимагає ручного налаштування частот. Концепція дорозвідки, запропонована Трофименко О. та ін., передбачає автономний аналіз даних у польоті, але потребує компактних обчислювальних модулів, які поки недоступні [9, с. 168].

Таблиця 1

Типи дронів у ЗСУ та потенціал ШІ на 2025 рік

Тип дрона	Основна функція	Приклади моделей	Ефективність	Використання ШІ	Обмеження для ШІ	Потенціал із ШІ
Розвідувальні	Збір даних, коригування артилерії	«Лелека-100», «Фурія»	Дальність до 50 км	Відсутнє (ручний аналіз)	Висока швидкість, РЕБ	Аналіз відео в польоті
Ударні/ Камікадзе	Знищення техніки та живої сили	«Switchblade», «Racoon»	Точність ~80 %	Відсутнє (ручне керування)	Затримки обробки, ресурси	Автономний вибір цілей
Логістичні	Доставка вантажів	«Valkyrie»	Вантаж до 20 кг	Відсутнє (ручне планування)	Простота завдань	Оптимізація маршрутів
РЕБ-дрони	Придушення зв'язку ворога	«Bukovel», «Піранья»	Дальність до 50 км	Відсутнє (ручне налаштування)	Технічна складність	Адаптивне реагування на РЕБ
Дорозвідка (потенціал)	Автономний аналіз і повернення	Немає в бойовому застосуванні	Теоретично висока	Не реалізовано	Вартість, РЕБ, технології	Самостійна обробка даних у польоті

Джерело: розроблено на основі [1–10]

Таблиця 1 систематизує типи дронів ЗСУ, їхні функції, моделі, ефективність, стан використання ШІ, обмеження та потенціал станом на квітень 2025 року. Дані ґрунтуються на аналізі літератури, що висвітлює технічні параметри, зокрема дальність, точність, автономність, а також виклики, пов'язані з апаратним і програмним забезпеченням [1–10]. Таблиця підкреслює залежність дронів від базових систем керування та обмеження ШІ через вразливість до РЕБ, недостатню обчислювальну потужність і енергоспоживання.

Перспективи розвитку дронів ЗСУ зосереджені на підвищенні автономності та стійкості до зовнішніх впливів. Ivanchenko О. та ін. пропонують модернізацію моделей, таких як «Лелека-100», для досягнення дальності 60 км і автономності 8 годин шляхом впровадження гібридних двигунів [4, с. 29]. Sorochkin О. та ін. рекомендують використовувати багатошарові системи навігації, включаючи лазерні гіроскопи з похибкою 0,01 градуса за годину, для забезпечення орієнтації без GPS протягом 30 хвилин у зонах РЕБ [7, с. 220]. Radovanović М. та ін. наголошують, що ройові технології потребують поступового розвитку через модульну координацію невеликих груп дронів, що дозволить ефективно розподіляти завдання без складної інфраструктури [8, с. 61]. Алгоритми комп'ютерного зору, такі як EfficientDet, можуть забезпечити розпізнавання об'єктів із точністю 88 % при енергоспоживанні 3 вати, що є реальним для дронів масою 8 кг. Розробка компактних процесорів із продуктивністю 1 терафлопс і вагою до 100 г, очікувана після 2026 року, усуне поточні апаратні обмеження. Впровадження протоколів зв'язку з частотним стрибанням, таких як LoRa на частоті 868 МГц, розширить дальність зв'язку до 5 км у зонах перешкод, підвищуючи стійкість до РЕБ.

Композитні матеріали, такі як армоване вуглецеве волокно, підвищують міцність корпусу дронів, дозволяючи витримувати падіння з висоти 5 метрів і пориви вітру до 15 м/с. Joshi А. та ін. пропонують використовувати літій-сірчані батареї, які краще витримують низькі температури, забезпечуючи стабільну роботу в зимових умовах [2, с. 141]. Екранування електроніки феромагнітними сплавами може зменшити вразливість до електромагнітних імпульсів, але додає 200 г ваги, що критично для легких моделей. Модульні рішення для оновлення програмного забезпечення без заміни апаратного забезпечення, як зазначають Sorochkin О. та ін., є пріоритетними для РЕБ-дронів, дозволяючи швидко адаптуватися до нових загроз [7, с. 219]. Концепція дорозвідки, запропонована Трофименко О. та ін., може кардинально змінити підхід до розвідки, дозволяючи дронам самостійно обробляти дані та повертатися з координатами цілей, мінімізуючи залежність від оператора та нестабільного зв'язку [9, с. 168]. Впровадження таких технологій потребує поступового вдосконалення апаратного забезпечення та зниження енергоспоживання обчислювальних модулів до 0,5 вата при продуктивності 0,3 терафлопс, що стане досяжним у найближчі роки.

Висновки

У результаті проведеного аналізу встановлено, що БПЛА відіграють важливу роль у реалізації завдань спостереження, навігації, передачі даних, виявлення об'єктів і радіоелектронного моніторингу. Типи БПЛА, що використовуються на практиці, демонструють широкий спектр можливостей, однак залишаються технічні виклики,

зокрема обмежена автономність, залежність від зовнішніх каналів зв'язку та вразливість до перешкод у середовищі з інтенсивним радіовипромінюванням.

Інтеграція методів штучного інтелекту, таких як комп'ютерний зір, автономна навігація, оптимізація траєкторій польоту, розпізнавання об'єктів та адаптивна обробка сигналів, дозволяє значно підвищити ефективність роботи БПЛА в автономному режимі. Особливо перспективними є технології, що забезпечують стійкість до дії радіоелектронних засобів протидії, а також здатність до самостійного прийняття рішень в умовах обмеженого або відсутнього зв'язку.

Таким чином, подальший розвиток БПЛА доцільно орієнтувати на підвищення їхньої автономності, стійкості до впливу навколишнього середовища та здатності до виконання складних функцій без участі оператора. Це включає розробку енергоефективних обчислювальних систем, вдосконалення алгоритмів машинного навчання та забезпечення безперервного функціонування в умовах інформаційної невизначеності.

Подяки

Автори висловлюють подяку Міністерству освіти і науки України за підтримку Лабораторії Інтелектуальних Систем в рамках дослідницького проекту «Інформаційні технології забезпечення стійкості інтелектуальної бортової системи малорозмірного літального апарата» (№ держреєстрації 0124U000548, 2024–2026 рр.).

Список використаної літератури

1. Литвиненко М. І., Ленець, В. Г., Гармаш Н. В., Шульга В. В. Аспекти впровадження штучного інтелекту у військовій справі. *Збірник наукових праць Харківського національного університету Повітряних Сил*. 2024. С. 13–18. DOI: 10.30748/zhups.2024.80.02

2. Joshi A., Spilbergs A., Miķelsone E. AI-Enabled Drone Autonomous Navigation and Decision Making for Defence Security. Environment, technologies, resources. *Proceedings of the International Scientific and Practical Conference*. 2024. Vol. 4, pp. 138–143. Rezekne Academy of Technologies. DOI: https://doi.org/10.17770/etr2024vol4.8237

3. Dash B., Ansari M. F., Swayamsiddha S. Fusion of Artificial Intelligence and 5G in Defining Future UAV Technologies – A Review. In *2023 International Conference on Device Intelligence, Computing and Communication Technologies*. 2023. *(DICCT)*. pp. 312–316. IEEE. DOI: https://doi.org/10.1109/dicct56244.2023.10110231

4. Ivanchenko O., Kurdiuk S., Khatuntsev Yu., Rudnichenko S. Analysis of application possibilities and classification of unmanned aerial vehicles for the support of combat operators of the navy of the armed forces of Ukraine. *Наукові праці Державного науково-дослідного інституту випробувань і сертифікації озброєння та військової техніки*. 2023. Vol. 18, Issue 4, pp. 23–34. DOI: https://doi.org/10.37701/dndivsovt.18.2023.04

5. Terzić M. Effects of the use of robotized combat systems during the conflict in Ukraine from 24.02.2022. *In Bastina*. 2024. Issue 62, pp. 363–381. Centre for Evaluation in Education and Science (CEON/CEES). DOI: https://doi.org/10.5937/bastina34-49656

6. Criollo L., Mena-Arciniega C., Xing S. Classification, military applications, and opportunities of unmanned aerial vehicles. In *Aviation*. 2024. Vol. 28, Issue 2, pp. 115–127. Vilnius Gediminas Technical University. DOI: https://doi.org/10.3846/aviation.2024.21672

7. Sorochkin,O., Sosulin M., Matveev Y. Prospects of military aviation through the integration of UAVs, AI and the latest technologies. *Communication, informatization and cybersecurity systems and technologies*. 2024. DOI:10.58254/viti.6.2024.17.215

8. Radovanović M., Petrovski A., Behlić A., Chaari M. Z., Hashimov E. G., Fellner R., Agbeyangi A. O. Unleashing Autonomous Forces: Integrating AI-Driven Drones in Modern Military Strategy. *Системи управління, навігації та зв'язку. Збірник наукових праць*. 2024. DOI:10.26906/sunz.2024.3.055

9. Трофименко О., Логінова Н., Соколов А., Чикунов П., Ахмаметьєва Г. Штучний інтелект у військові сфері. *Електронне фахове наукове видання «Кібербезпека: освіта, наука, техніка»*. 2024. 1(25). 161–176. DOI: https://doi.org/10.28925/2663-4023.2024.25.161176

10. Гуржій С. В. Тенденції застосування технологій штучного інтелекту у військово-технічній сфері. *Інформація і право*. 2024. № 3 (50). С. 136-146. DOI: https://doi.org/10.37750/2616-6798.2024.3(50).311713

References

1. Litvinenko, M. I. & Lenets, V. G. & Garmash, N. V. & Shulga, V. V. (2024). Aspekty vprovadzhennya shtuchnoho intelektu u viys′koviy spravi [Aspects of the implementation of artificial intelligence in military affairs]. *Collection of scientific works of the Kharkiv National University of the Air Force*. 13–18. DOI: 10.30748/zhups.2024.80.02 [in Ukrainian].

2. Joshi, A., Spilbergs, A., & Miķelsone, E. (2024). AI-Enabled Drone Autonomous Navigation and Decision Making for Defence Security. Environment, technologies, resources. Proceedings of the International Scientific and Practical Conference (Vol. 4, pp. 138–143). Rezekne Academy of Technologies. DOI: https://doi.org/10.17770/etr2024vol4.8237

3. Dash, B., Ansari, M. F., & Swayamsiddha, S. (2023). Fusion of Artificial Intelligence and 5G in Defining Future UAV Technologies – A Review. In 2023 International Conference on Device Intelligence, Computing and Communication Technologies, (DICCT) (pp. 312–316). IEEE. DOI: https://doi.org/10.1109/dicct56244.2023.10110231

4. Ivanchenko, O., Kurdiuk, S., Khatuntsev, Yu., & Rudnichenko, S. (2023). Analysis of application possibilities and classification of unmanned aerial vehicles for the support of combat operators of the navy of the armed forces of Ukraine. Scientific works of the State Scientific Research Institute of Armament and Military Equipment Testing and Certification (Vol. 18, Issue 4, pp. 23–34). State Scientific Research Institute of Armament and Military Equipment Testing and Certification. DOI: https://doi.org/10.37701/dndivsovt.18.2023.04

5. Terzic, M. (2024). Effects of the use of robotic combat systems during the conflict in Ukraine from 24.02.2022. In Bastina (Issue 62, pp. 363–381). Center for Evaluation in Education and Science (CEON/CEES). DOI: https://doi.org/10.5937/bastina34-49656

6. Criollo, L., Mena-Arciniega, C., & Xing, S. (2024). Classification, military applications, and opportunities of unmanned aerial vehicles. In Aviation (Vol. 28, Issue 2, pp. 115–127). Vilnius Gediminas Technical University. DOI: https://doi.org/10.3846/aviation.2024.21672

7. Sorochkin, O., Sosulin, M., & Matveev, Y. (2024). Prospects of military aviation through the integration of UAVs, AI and the latest technologies. Communication, informatization and cybersecurity systems and technologies. DOI:10.58254/viti.6.2024.17.215

8. Radovanović, M., Petrovski, A., Behlić, A., Chaari, M.Z., Hashimov, E.G., Fellner, R., & Agbeyangi, A.O. (2024). Unleashing Autonomous Forces: Integrating AI-Driven Drones in Modern Military Strategy. Control, Navigation and Communication Systems. Collection of Scientific Papers. DOI:10.26906/sunz.2024.3.055

9. Trofimenko, O., Loginova, N., Sokolov, A., Chikunov, P., & Akhmametyeva, G. (2024). Shtuchnyy intelekt u viys′kovi sferi [Artificial Intelligence in the Military Sphere]. Electronic Professional Scientific Publication "Cybersecurity: Education, Science, Technology", 1(25), 161–176. DOI: https://doi.org/10.28925/2663-4023.2024.25.161176 [in Ukrainian].

10. Gurzhiy, S. V. (2024). Tendentsiyi zastosuvannya tekhnolohiy shtuchnoho intelektu u viys′kovo-tekhnichniy sferi [Trends in the application of artificial intelligence technologies in the military-technical sphere]. Information and Law. No. 3 (50). P. 136–146. DOI: https://doi.org/10.37750/2616-6798.2024.3(50).311713 [in Ukrainian].

Дата першого надходження рукопису до видання: 19.09.2025
Дата прийнятого до друку рукопису після рецензування: 16.10.2025
Дата публікації:

УДК 004.8:519.6 DOI

С. Ю. БОРЮ
кандидат технічних наук, доцент,
доцент кафедри комп'ютерних наук
Запорізький національний університет
ORCID: 0009-0008-9602-1629

Д. А. РЕЗЦЯ
магістр кафедри комп'ютерних наук
Запорізький національний університет
ORCID: 0009-0003-6945-0878

Н. О. МИРОНОВА
кандидат технічних наук, доцент,
доцент кафедри інформаційних технологій електронних засобів
Національний університет «Запорізька політехніка»
ORCID: 0000-0001-7484-559X

Т. І. КУЛЯБА-ХАРИТОНОВА
старший викладач кафедри інформаційних
технологій електронних засобів
Національний університет «Запорізька політехніка»
ORCID: 0000-0002-4146-4931

ПРОГРАМНА РЕАЛІЗАЦІЯ АСИМПТОТИЧНИХ МЕТОДІВ У MAXIMA ДЛЯ ІНТЕЛЕКТУАЛЬНИХ СИСТЕМ АНАЛІЗУ ТА ПРОГНОЗУВАННЯ РУХУ ОБ'ЄКТІВ У ДИНАМІЧНОМУ СЕРЕДОВИЩІ

У статті розглядається застосування системи комп'ютерної алгебри Maxima для автоматизації побудови асимптотичних розв'язків диференціальних рівнянь з малим параметром. Асимптотичні методи є одним із інструментів аналізу складних динамічних систем, однак їх використання ускладнюється значними обсягами громіздких алгебраїчних та диференціальних перетворень. Запропонований підхід дозволяє автоматизувати ці процедури та мінімізувати ймовірність помилок, що виникають при ручних обчисленнях. У роботі побудовано програмний модуль, який реалізує алгоритм формування системи рівнянь для коефіцієнтів розкладу, виконує сортування виразів за степенями малого параметра та забезпечує спрощення аналітичних результатів із використанням користувацьких правил підстановок.

Для демонстрації можливостей розробленого модуля розглянуто приклад рівняння, що описує рух об'єкта у змінному силовому полі. Отримані результати підтверджують, що автоматизація дозволяє ефективно працювати з розкладами високого порядку, де складність виразів зростає експоненційно. Наведені експериментальні дані (час обчислень, кількість елементів у виразах) демонструють переваги запропонованого підходу. Особливістю модуля є можливість переходу між різними формами подання результатів, зокрема від експоненціальної до тригонометричної, що робить вирази більш компактними та зручними для подальшої обробки.

Наукова новизна роботи полягає у поєднанні асимптотичних методів і системи Maxima як символьного інструменту для формування аналітичних моделей, придатних для використання у складі інтелектуальних систем прогнозування руху об'єктів у динамічному середовищі.

Практична значимість полягає у можливості інтеграції модуля в цифрові двійники, робототехнічні комплекси та транспортні системи, де критично важливою є точність і швидкість прогнозування.

Можливими напрямами подальших досліджень є розробка стратегій зменшення символьної складності виразів, застосування гібридних символьно-чисельних методів, а також інтеграція модуля з нейромережними підсистемами для забезпечення прогнозування в реальному часі та його валідації.

Ключові слова: *асимптотичні методи, система комп'ютерної алгебри, Maxima, інтелектуальні системи, прогнозування руху, динамічне середовище, цифрові двійники.*

© Борю С. Ю., Резця Д. А., Миронова Н. О., Куляба-Харитонова Т. І., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

S. YU. BORIU
Candidate of Technical Sciences, Associate Professor,
Associate Professor at the Department of Computer Science
Zaporizhzhia National University
ORCID: 0009-0008-9602-1629

D. A. REZTSIA
Master Studetnt at the Department of Computer Science
Zaporizhzhia National University
ORCID: 0009-0003-6945-0878

N. O. MYRONOVA
Candidate of Technical Sciences, Associate Professor,
Associate Professor at the Department of Information Technologies
of Electronic Devices
Zaporizhzhia Polytechnic National University
ORCID: 0000-0001-7484-559X

T. I. KULYABA-KHARYTONOVA
Senior Lecturer at the Department of Information Technology
for Electronic Devices
Zaporizhzhia Polytechnic National University
ORCID: 0000-0002-4146-4931

SOFTWARE IMPLEMENTATION OF ASYMPTOTIC METHODS IN MAXIMA FOR INTELLIGENT SYSTEMS FOR ANALYSIS AND PREDICTION OF OBJECT MOTION IN A DYNAMIC ENVIRONMENT

The article considers the application of the computer algebra system Maxima for automating the construction of asymptotic solutions of differential equations with a small parameter. Asymptotic methods are among the most tools for analyzing complex dynamic systems, but their practical use is complicated by the large amount of cumbersome algebraic and differential transformations. The proposed approach makes it possible to automate these procedures and minimize the probability of errors that usually occur during manual calculations. A software module has been developed that implements an algorithm for generating a system of equations for expansion coefficients, performs sorting of expressions by the powers of a small parameter, and provides simplification of analytical results using user-defined substitution rules.

To demonstrate the capabilities of the developed module, an example of an equation describing the motion of an object in a variable force field is considered. The obtained results confirm that automation enables efficient work with higher-order expansions, where the complexity of expressions grows exponentially. The presented experimental data (computation time, number of terms in expressions) illustrate the advantages of the proposed approach. A specific feature of the module is the ability to switch between different representations of results, in particular from exponential to trigonometric form, which makes expressions more compact and convenient for further processing.

The scientific novelty of the work lies in the integration of asymptotic methods with Maxima as a symbolic tool for building analytical models suitable for use as a core of intelligent systems for motion prediction in dynamic environments. The practical significance is defined by the possibility of integrating the module into digital twins, robotic systems, and transport platforms, where the accuracy and speed of prediction are critical.

Possible directions for future research include the development of strategies to reduce the symbolic complexity of expressions, the application of hybrid symbolic-numerical methods, as well as the integration of the module with neural network subsystems to enable real-time prediction and validation.

Key words: *asymptotic methods, computer algebra system, Maxima, intelligent systems, motion prediction, dynamic environment, digital twins.*

Постановка проблеми

Аналіз і прогнозування руху об’єктів у динамічному середовищі є однією з актуальних задач сучасних комп’ютерних наук. Такі задачі виникають у робототехніці, транспортних системах, цифрових двійниках, а також у сфері штучного інтелекту та аналітичних систем підтримки прийняття рішень. Ключовою складністю є побудова адекватних математичних моделей, які враховують нелінійність середовища, часову змінність зовнішніх впливів і наявність малих параметрів, що ускладнюють отримання точних розв’язків диференціальних рівнянь.

Традиційні методи чисельного інтегрування часто забезпечують лише наближені рішення без достатньої інтерпретованості, тоді як асимптотичні методи дозволяють отримати розклад розв’язку у вигляді функціональних

рядів. Це дає можливість проводити якісний аналіз поведінки системи та робити прогнозні оцінки. Проте застосування асимптотичних методів у практиці пов'язане з виконанням великої кількості громіздких аналітичних перетворень, що ускладнює їх використання без автоматизації.

З іншого боку, розвиток систем комп'ютерної алгебри (CAS), зокрема Maxima, відкриває можливості для автоматизованого виконання символьних обчислень. Проте існуючі підходи здебільшого орієнтовані на локальні задачі математичної фізики і не розглядають можливості їх інтеграції в інтелектуальні системи прогнозування руху об'єктів.

Таким чином, актуальною науково-прикладною проблемою є розробка програмного модуля на основі системи комп'ютерної алгебри Maxima, який автоматизує асимптотичні методи розв'язання диференціальних рівнянь з малим параметром і може бути використаний у складі інтелектуальних систем аналізу та прогнозування руху в динамічному середовищі.

Формулювання мети дослідження

Метою дослідження є розробка та програмна реалізація алгоритму автоматизованого побудови асимптотичних розв'язків диференціальних рівнянь з малим параметром у системі комп'ютерної алгебри Maxima та обґрунтування можливостей його використання у складі інтелектуальних систем аналізу й прогнозування руху об'єктів у динамічному середовищі.

Аналіз останніх досліджень і публікацій

Розв'язання диференціальних рівнянь з малим параметром традиційними асимптотичними методами пов'язане з виконанням значного обсягу алгебраїчних і диференціальних перетворень. Ручні обчислення ускладнюють процес і підвищують ризик помилок, що зумовлює необхідність використання програмних засобів для автоматизації побудови розкладів.

Упродовж останніх десятиліть широкого поширення набули системи комп'ютерної алгебри (CAS), серед яких Maxima посідає важливе місце завдяки відкритому коду, гнучкості та можливостям створення користувацьких функцій. Вибір Maxima [1] обумовлений тим, що це популярний інструмент для автоматизації символьних перетворень, необхідних при застосуванні асимптотичних методів. Слід зазначити, що використання Maxima вимагає від користувача не лише знання математичних методів, а й розуміння формальних правил програмування аналітичних обчислень.

Сучасні дослідження дедалі частіше спрямовані на інтеграцію символьної математики та інтелектуальних систем аналізу й прогнозування руху. У роботах [2, 3] показано, що поєднання алгоритмів обробки символьних моделей із машинним навчанням забезпечує новий рівень точності у задачах передбачення траєкторій. Подібні підходи актуальні для робототехніки, транспортних систем і цифрових двійників, де критично важливим є прогнозування динаміки в реальному часі [4].

Окремі публікації [5, 6] присвячені використанню цифрових двійників у поєднанні зі штучним інтелектом, що створює можливість інтеграції символьних алгоритмів безпосередньо у системи управління та моделювання. У роботі [7] акцент зроблено на гібридних символьно-чисельних підходах, які поєднують переваги CAS з ефективністю чисельних методів, що є перспективним напрямом для прогнозування складних динамічних процесів.

Прикладне використання Maxima у задачах аналізу нелінійної та хаотичної динаміки наведено [8] та представлено у навчальних і прикладних задачах з обчислювальної математики [9]. Це підтверджує, що CAS можуть застосовуватися не лише для формальних аналітичних обчислень, але й у ширшому контексті інтелектуальних систем і навчальних середовищ.

Таким чином, проведені дослідження підтверджують актуальність створення програмних модулів, які поєднують асимптотичні методи, автоматизовані символьні обчислення та інтелектуальні підходи до прогнозування руху. Тому розробка модулів на основі Maxima розглядається як перспективний напрям для інтеграції в багаторівневі інтелектуальні системи.

Викладення основного матеріалу дослідження

У дослідженні розглядається використання системи комп'ютерної алгебри Maxima для побудови асимптотичних розв'язків диференціальних рівнянь з малим параметром методом фазових інтегралів.

Задамо рівняння виду:

$$\varepsilon^2 \frac{d^2 y(t)}{dt^2} + F(t) y(t) = 0, \tag{1}$$

де ε – «малий параметр», $F(t)$ – функція, що описує властивості динамічного середовища (наприклад, змінне силове поле), $y(t)$ – шуканий розв'язок. Такі рівняння застосовуються для моделювання руху об'єктів у середовищах зі змінними характеристиками, зокрема в задачах робототехніки та прогнозування траєкторій у транспортних або аеродинамічних системах

Шукатимемо розв'язок у вигляді експоненти з невідомою функцією:

$$y(t) = \exp \int \phi(t) dt \tag{2}$$

вираз (2) підставляється у рівняння (1) для побудови розв'язку

$$\frac{dy(t)}{dt} = \phi(t)\exp\int\phi(t)dt,$$

$$\frac{d^2y(t)}{dt^2} = \left(\phi(t)^2 + \frac{d\phi(t)}{dt}\right)\exp\int\phi(t)dt,$$

$$\varepsilon^2\left(\phi(t)^2 + \frac{d\phi(t)}{dt}\right) + F(t) = 0. \quad (3)$$

Подальші кроки методу, спрямовані на визначення шуканих залежностей $\phi(t)$, пов'язані з виконанням значного обсягу аналітичних перетворень.

Необхідна функціональна залежність подається у вигляді розкладу за малим параметром ε у ряд функцій $\varphi_k(t)$:

$$\varphi(t) = \varepsilon^{-1}\varphi_0(t) + \varepsilon^0\varphi_1(t) + \ldots + \varepsilon^k\varphi_k(t) + \ldots \quad (4)$$

Алгебраїчні рівняння, з яких знаходяться $\varphi_k(t)$, отримуються підстановкою (4) в (3) і вимогою рівності коефіцієнтів при однакових степенях малого параметра ε. Наприклад, для $k = 0$, обмежуючись урахуванням тільки доданками при ε^0, необхідно «вручну» виконати наступні перетворення:

$$\varepsilon^2\left(\varepsilon^{-2}\phi_0(t)^2 + \varepsilon^{-1}\frac{d\phi_0(t)}{dt}\right) + F(t) = 0;$$

$$\phi_0(t)^2 + F(t) = 0;$$

$$\phi_0(t) = \pm iF(t)^{1/2}; \quad (5)$$

$$y(t) = C_1\exp\left[i\int(\varepsilon^{-1}F(t)^{1/2})dt\right] + C_2\exp\left[-i\int(\varepsilon^{-1}F(t)^{1/2})dt\right];$$

$$y(t) = C_1\sin(K(t)) + C_2\cos(K(t));$$

де $K(t) = \int(\varepsilon^{-1}F(t)^{1/2})dt.$

Подальший аналіз розв'язку вимагає завдання $F(t)$ та обчислення $K(t)$ та визначення констант C_1 та C_2.

Для подолання цих труднощів було розроблено програмний модуль у Maxima, який автоматизує побудову системи рівнянь для $\varphi_k(t)$ та формує розклад розв'язку до заданого порядку.

Алгоритм включає:

– підстановку розкладу у вихідне рівняння;
– сортування за степенями параметра ε;
– автоматичне формування системи для коефіцієнтів;
– символьне спрощення виразів за допомогою користувацьких функцій і шаблонів підстановки (наприклад, переходу від експоненціальної до тригонометричної форми).

Для прикладу було реалізовано побудову розкладу до трьох членів. Таблиця 1 демонструє залежність часу обчислень і складності виразів від порядку розкладу.

Отримані дані свідчать про стабільне зростання складності при збільшенні порядку розкладу та підтверджують доцільність використання автоматизованих інструментів. Модуль значно скорочує обсяг ручної роботи, зменшує ризик помилок і формує аналітичні вирази, які можна безпосередньо застосовувати для прогнозування руху.

Таблиця 1

Результати автоматизованих обчислень у Maxima

Кількість членів розкладу	Час обчислення (с)	Кількість елементів у виразі
2	0.4	15
3	0.8	37
4	1.9	96

Розроблений модуль може бути інтегрований у інтелектуальні системи аналізу та прогнозування руху об'єктів у динамічному середовищі. Його функціональні можливості забезпечують:

– аналітичну підтримку прогнозування – формування розкладів дозволяє враховувати вплив малих параметрів і нелінійних ефектів;
– адаптивність – завдяки можливості підстановки конкретних форм $F(t)$, система може налаштовуватись під різні фізичні умови середовища;
– інтеграцію з цифровими двійниками – символьні результати можуть бути використані для калібрування моделей у реальному часі;

– комбінування з машинним навчанням – нейромережні блоки можуть виконувати швидке чисельне прогнозування, а аналітичний модуль Maxima – уточнювати та інтерпретувати отримані результати.

Таким чином, розроблений інструмент створює підґрунтя для побудови гібридних інтелектуальних систем, здатних поєднувати символьні методи, цифрові двійники та технології штучного інтелекту для підвищення точності й надійності прогнозування траєкторій.

Висновки

У роботі запропоновано алгоритм і реалізовано програмний модуль у системі комп'ютерної алгебри Maxima для автоматизації побудови асимптотичних розв'язків диференціальних рівнянь з малим параметром. На прикладі рівнянь, що описують рух об'єкта в динамічному середовищі (змінне зовнішнє поле, нелінійні члени), показано, що модуль коректно формує систему рівнянь для коефіцієнтів ряду та автоматично будує наближений аналітичний розв'язок. Автоматизація дозволяє зменшити ручні перетворення й помилки, прискорити підготовку аналітичних виразів і полегшити їх використання в подальшій обробці.

Наукова новизна полягає в тому, що запропоновано і реалізовано методологію застосування користувацьких правил підстановок у Maxima для спрощення та приведення проміжних аналітичних виразів до зручних для інтерпретації форм; показано можливість використання такого символьного ядра як частини інтелектуальної системи прогнозування руху об'єктів.

Практична цінність роботи полягає в тому, що розроблений модуль придатний для інтеграції в інтелектуальні системи (цифрові двійники, модулі планування траєкторій) та дозволяє одержувати аналітично обґрунтовані поправки до базових моделей (що важливо при роботі в умовах невеликих збурень), а також використовувати символьні результати як основу для швидкої чисельної оцінки сценаріїв.

Можливим напрямом подальшого розвитку розробленого модуля є подолання проблеми зростання символьної складності виразів при збільшенні порядку розкладу. Для цього доцільно застосовувати стратегії спрощення, зокрема використання шаблонів, часткових підстановок та гібридних символьно-чисельних методів. У перспективі планується інтеграція модуля з нейромережними компонентами, що забезпечують швидкі чисельні прогнози, а також з цифровими двійниками для реалізації прогнозування в реальному часі та валідації результатів.

Список використаної літератури

1. Maxima: Documentation and reference / Maxima development team. Режим доступу: https://maxima.sourceforge.io/ (дата звернення: 23.09.2025).

2. Liu J., Yap H. J., Khairuddin A. S. M. Review on Motion Planning of Robotic Manipulator in Dynamic Environments. Journal of Sensors, Vol. 2024, Article ID 5969512, 28 p. DOI: 10.1155/2024/5969512

3. Deng T., et al. Recent advances in deterministic human motion prediction. Computer Vision and Pattern Recognition. https://arxiv.org/abs/2312.06184

4. Chaparro-Cárdenas S. L., et al. A Technological Review of Digital Twins and Artificial Intelligence for Personalized and Predictive Healthcare. Healthcare, 2025. DOI: https://www.mdpi.com/2227-9032/13/14/1763

5. Mayer A., Greif L., Häußermann T. M., et al. Digital Twins, Extended Reality, and Artificial Intelligence in Manufacturing Reconfiguration: A Systematic Literature Review. Sustainability, Vol. 17, Issue 5, 2025. DOI: https://doi.org/10.3390/su17052318

6. Massaro A. Electronic Artificial Intelligence and Digital Twins in Industry 5.0: A Systematic Review and Perspectives. Machines, Vol. 13, Issue 9, 2025. https://doi.org/10.3390/machines13090755

7. Covic N., Lacevic B., Osmankovic D., Uzunovic T. Real-Time Sampling-Based Safe Motion Planning for Robotic Manipulators in Dynamic Environments. Robotics, 2025. URL: https://arxiv.org/abs/2501.00507

8. Morante A., Vallejo J. A. Chaotic dynamics with Maxima / A. Morante, J. A. Vallejo. arXiv preprint:1301.3240, 2013. Режим доступу: https://arxiv.org/abs/1301.3240 (дата звернення: 23.09.2025).

9. Karjanto N., Klinger B. Not Another Computer Algebra System: Highlighting wxMaxima in Calculus / Natanael Karjanto, Brandon Klinger // Mathematics. 2021. Vol. 9, No. 12. Art. 1317. DOI: 10.3390/math9121317

References

1. Maxima Development Team. (n.d.). Maxima: Documentation and reference. Retrieved September 23, 2025, from https://maxima.sourceforge.io/

2. Liu, J., Yap, H. J., & Khairuddin, A. S. M. (2024). Review on motion planning of robotic manipulator in dynamic environments. Journal of Sensors, 2024, Article 5969512. https://doi.org/10.1155/2024/5969512

3. Deng, T., et al. (2023). Recent advances in deterministic human motion prediction. Computer Vision and Pattern Recognition. Retrieved from https://arxiv.org/abs/2312.06184

4. Chaparro-Cárdenas, S. L., et al. (2025). A technological review of digital twins and artificial intelligence for personalized and predictive healthcare. Healthcare, 13(14), 1763. https://www.mdpi.com/2227-9032/13/14/1763

5. Mayer, A., Greif, L., Häußermann, T. M., et al. (2025). Digital twins, extended reality, and artificial intelligence in manufacturing reconfiguration: A systematic literature review. Sustainability, 17(5). https://doi.org/10.3390/su17052318

6. Massaro, A. (2025). Electronic artificial intelligence and digital twins in Industry 5.0: A systematic review and perspectives. Machines, 13(9). https://doi.org/10.3390/machines13090755

7. Covic, N., Lacevic, B., Osmankovic, D., & Uzunovic, T. (2025). Real-time sampling-based safe motion planning for robotic manipulators in dynamic environments. Robotics. Retrieved from https://arxiv.org/abs/2501.00507

8. Morante, A., & Vallejo, J. A. (2013). Chaotic dynamics with Maxima. arXiv preprint arXiv:1301.3240. Retrieved September 23, 2025, from https://arxiv.org/abs/1301.3240

9. Karjanto, N., & Klinger, B. (2021). Not another computer algebra system: Highlighting wxMaxima in calculus. Mathematics, 9(12), 1317. https://doi.org/10.3390/math9121317

Дата першого надходження рукопису до видання: 27.09.2025
Дата прийнятого до друку рукопису після рецензування: 23.10.2025
Дата публікації:

УДК 004.056:004.056.5:004.056.2 DOI

Є. М. БРОВЧЕНКО
аспірант
Інститут комп'ютерних технологій
Відкритого міжнародного університету розвитку людини «Україна»
ORCID: 0000-0002-1416-0385

В. П. САМАРАЙ
кандидат технічних наук, старший науковий співробітник, доцент
Інститут комп'ютерних технологій
Відкритого міжнародного університету розвитку людини «Україна»
ORCID: 0000-0003-4419-1366

АНАЛІЗ ПРОБЛЕМ ЗАХИСТУ НЕСТРУКТУРОВАНОЇ ІНФОРМАЦІЇ НА МОБІЛЬНИХ ПРИСТРОЯХ: СУЧАСНІ ВИКЛИКИ ТА ТЕНДЕНЦІЇ

У статті досліджено проблеми захисту неструктурованих даних на мобільних пристроях у контексті тривалих сесій, де одноразова автентифікація не забезпечує належного рівня безпеки. Запропоновано концептуальну рамку, що поєднує безперервну поведінкову автентифікацію, даноцентричні політики доступу та гібридну криптоконфігурацію: симетричне шифрування для контенту та асиметричне – для ключів. ***Мета.*** *Обґрунтувати модель, яка мінімізує ризики доступу до повідомлень, медіа, вкладень і кешованих даних у межах активної сесії. Основна увага приділяється поєднанню AES ↔ ECC, MFA ↔ поведінкової перевірки, а також безперервному моніторингу на основі динаміки дотиків, даних інерціальних сенсорів, злиття ознак, оцінок і рішень, ризик-адаптивних дій, локального та федеративного навчання, а також відтворюваних метрик.* ***Методи.*** *Здійснено огляд літератури з баз даних Web of Science та Scopus; систематизовано підходи до безперервної автентифікації та сенсорного злиття; проаналізовано гігієну дозволів і життєвий цикл тимчасових копій; розглянуто політики на основі підкріплення (RL) для step-up автентифікації та повторної перевірки; узгоджено модель із сучасними нормами керування сесіями та вимогами криптостійкості.* ***Результати.*** *Одноразова автентифікація не гарантує сталості ідентичності користувача. Безперервна перевірка, підсилена сенсорним злиттям, дозволяє зменшити «вікно ризику» без потреби у надмірних явних діях. Дотикові ознаки демонструють високу ефективність, а інерціальні сенсори забезпечують фоновий контроль. Політики RL активують втручання лише за умов підвищеного ризику (reauth, маскування прев'ю, призупинення експорту). Даноцентричні політики доступу зменшують площу атаки з урахуванням дозволів, побічних сенсорних каналів, форензичних артефактів, платформенних уразливостей. Нормативний блок закріплює гібридну модель «AES для контенту ↔ ECC для ключа/підписів» та доповнює її MFA і пост-логін моніторингом.* ***Висновки.*** *Запропонована конфігурація включає три основні компоненти: безперервна поведінкова перевірка; даноцентричні політики доступу (керування прев'ю, кешами, тимчасовими копіями, дозволами); гібридна криптографія, де симетричне шифрування захищає контент, а асиметричне – ключі та цифрові підписи. Інтеграція цих компонентів забезпечує баланс між точністю, затримками та енергоефективністю, що робить модель придатною для корпоративного використання та захищених месенджерів.*

Ключові слова: *безперервна автентифікація; мультибіометрія; даноцентричні політики доступу; гібридне шифрування AES/ECC; сесійний моніторинг.*

E. M. BROVCHENKO
Postgraduate Student
Institute of Computer Technologies
of Open International University of Human Development "Ukraine"
ORCID: 0000-0002-1416-0385

V. P. SAMARAY
Ph.D. in Technical Sciences, Senior Research Fellow, Associate Professor
Institute of Computer Technologies
of Open International University of Human Development "Ukraine"
ORCID: 0000-0003-4419-1366

© Бровченко Є. М., Самарай В. П., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

ANALYSIS OF UNSTRUCTURED DATA PROTECTION ON MOBILE DEVICES: CURRENT CHALLENGES AND TRENDS

This article examines the challenges of protecting unstructured data on mobile devices in the context of prolonged sessions, where one-time authentication does not ensure an adequate level of security. A conceptual framework is proposed that combines continuous behavioral authentication, data-centric access policies, and a hybrid cryptographic configuration: symmetric encryption for content and asymmetric encryption for keys.

Purpose. *To substantiate a model that minimizes the risks of unauthorized access to messages, media, attachments, and cached data within an active session. The focus is on integrating AES with ECC, MFA with behavioral verification, as well as continuous monitoring based on touch dynamics, inertial sensor data, feature fusion, scoring and decision systems, risk-adaptive actions, local and federated learning, and reproducible metrics.*

Methods. *A literature review was conducted using Web of Science and Scopus databases; approaches to continuous authentication and sensor fusion were systematized; permission hygiene and the lifecycle of temporary copies were analyzed; reinforcement learning (RL)-based policies for step-up authentication and re-verification were examined; the proposed model was aligned with modern session management standards and cryptographic resilience requirements.*

Results. *One-time authentication does not guarantee the continuity of user identity. Continuous verification enhanced by sensor fusion reduces the "risk window" without requiring excessive explicit actions. Touch-based features demonstrate high effectiveness, while inertial sensors provide background monitoring. RL-driven policies trigger interventions only under elevated risk conditions (reauthentication, preview masking, export suspension). Data-centric access policies reduce the attack surface by accounting for permissions, side-channel sensor leaks, forensic artifacts, and platform vulnerabilities. The regulatory block consolidates the hybrid model "AES for content ↔ ECC for keys/signatures" and complements it with MFA and post-login monitoring.*

Conclusions. *The proposed configuration incorporates three key components: continuous behavioral verification; data-centric access policies (management of previews, caches, temporary copies, permissions); and hybrid cryptography, where symmetric encryption protects content while asymmetric encryption secures keys and digital signatures. The integration of these components ensures a balance between accuracy, latency, and energy efficiency, making the model suitable for corporate environments and secure messengers.*

Key words: *continuous authentication; multibiometrics; data-centric access policies; hybrid AES/ECC encryption; session monitoring.*

Постановка проблеми

Мобільні пристрої стали ключовим середовищем для створення, оброблення й обміну неструктурованою інформацією, зокрема текстовими повідомленнями, зображеннями, голосовими та відеофайлами, вкладеннями електронної пошти й месенджерів, кешами застосунків і файлами хмарних сервісів. На відміну від структурованих даних, характеризується високою мінливістю, фрагментарністю та схильністю до дублювання у різних контейнерах. Він постійно мігрує між корпоративним і персональним середовищами, особливо в умовах моделей BYOD та COPE. Різнорідність форматів, фрагментація мобільних екосистем, велика кількість дозволів, інтеграція сторонніх SDK, а також уразливості ланцюгів постачання створюють значні труднощі для класифікації чутливості даних, контролю доступу, журналювання активності та запобігання витокам інформації на рівні ендпоїнта (кінцевого пристрою).

Нормативні тенденції формують багатовимірний контекст проблеми захисту неструктурованих даних на мобільних пристроях. Поширення безпарольних методів автентифікації та вимога забезпечення «безшовного» користувацького досвіду зумовлюють потребу у безперервному підтвердженні особи на основі поведінкових ознак. Це, своєю чергою, висуває суворі вимоги до точності, стійкості до підробок і захисту приватності користувача. Впровадження наскрізного та клієнтського шифрування сприяє забезпеченню конфіденційності, але водночас ускладнює реалізацію механізмів контролю витоків даних (DLP) та аудиту. Розбудова архітектур Zero Trust на мобільних ендпоїнтах потребує контекстно- та контент-орієнтованого прийняття рішень в умовах обмежених обчислювальних і енергетичних ресурсів. Поява постквантової криптографії на горизонті розвитку безпеки ставить додаткові вимоги до криптоагільності, сумісності алгоритмів та адаптивності систем. Крім того, сучасні нормативні рамки, зокрема вимоги до безпеки оброблення даних та повідомлення про інциденти, зобов'язують забезпечити керованість усього життєвого циклу неструктурованої інформації.

Наукова складова дослідження полягає у формалізації моделей ризику для потоків неструктурованого контенту на мобільних пристроях, розробленні методів безперервної автентифікації з відтворюваними метриками та чіткими гарантіями приватності, синтезі контент- та контекст-орієнтованого контролю доступу в операційних системах, створенні алгоритмів автоматичного класифікування та лейблування даних для подальшого застосування політик DLP, а також у проєктуванні криптоагільних схем шифрування й відкритих бенчмарків для об'єктивної оцінки ефективності рішень.

Практична значущість дослідження проявляється у необхідності побудови мобільних Zero Trust-ландшафтів із керованими політиками доступу та телеметрією, впровадженні інтегрованих рішень, що поєднують класифікацію, лейблування, DLP і DRM/IRM для захисту контенту як на пристрої, так і в хмарному середовищі. Важливим

є також планування міграції до постквантових криптографічних профілів без деградації продуктивності, а також забезпечення відповідності регуляторним вимогам шляхом впровадження доказового журналювання, атестації та регулярних оцінок ризиків.

Таким чином, досліджувана проблема має міждисциплінарний характер і потребує узгоджених рішень на перетині безпеки даних, мобільної інженерії, криптографії та управління ризиками. Це дозволяє досягти стійкої рівноваги між конфіденційністю, спостережністю контрольних механізмів та зручністю для кінцевого користувача.

Аналіз останніх досліджень і публікацій

Проблема захисту неструктурованої інформації на мобільних пристроях послідовно висвітлюється у сучасних наукових працях, які охоплюють спектр від методів безперервної автентифікації до ширшого контексту загроз і уразливостей платформ. C. J. Jose і M. S. Rajasree запропонували модель імпліцитної безперервної автентифікації на основі підкріплювального навчання, яка адаптує рішення до змінних патернів взаємодії користувача. Водночас модель є чутливою до дрейфу даних і нестабільності сенсорних потоків [1].

D. Reichinger, E. Sonnleitner і M. Kurz продемонстрували переваги мультимодальної фузії поведінкових ознак над одномодальними підходами. Вони окреслили компроміс між точністю, затримкою та енергоспоживанням, що особливо важливо для реалізації таких рішень безпосередньо на мобільному пристрої [2].

P. M. A. B. Estrela, R. d. O. Albuquerque, D. M. Amaral, W. F. Giozza і R. T. de Sousa Júnior зосередили увагу на банківських сценаріях, досліджуючи динаміку дотиків як поведінкову ознаку. Вони продемонстрували життєздатність запропонованого рішення у продуктивному середовищі, водночас виявивши ризики імітаційних атак через відтворення жестів користувача [3]. S. Mekruksavanich і A. Jitpattanakul показали потенціал глибинних моделей, що навчаються на патернах активності, отриманих із мобільних сенсорів. Автори наголосили на критичній ролі коректного визначення вікон агрегації та вибору методів забезпечення приватності як у процесі навчання моделей, так і під час здійснення інференсу безпосередньо на кінцевому ендпоїнті [4].

P. Bansal і A. Ouda здійснили порівняльний аналіз поведінкових біометричних ознак і методів підкріплювального навчання для безперервної автентифікації. Вони акцентували увагу на проблемі узагальнюваності моделей між різними пристроями та відсутності уніфікованих бенчмарків, що ускладнює коректне порівняння результатів [5]. A. Ramaswamy та співавтори досліджували багатопристроєві сценарії, де здійснюється кореляція потоків даних з кількох смартпристроїв. Це дозволило виявити переваги фузії даних, але водночас підкреслило залежність загальної стійкості системи від надійності окремих вузлів екосистеми [6].

M. Abuhamad, A. Abusnaina, D. Nyang і D. Mohaisen здійснили систематизацію сенсорно орієнтованих підходів до безперервної автентифікації на смартфонах, а також типових векторів атак. Вони звернули увагу на дефіцит публічних, репрезентативних датасетів і різноспрямованість протоколів оцінювання, що ускладнює відтворюваність результатів і порівнянність моделей [7].

J. A. Delgado-Santos та співавтори показали, як мобільні сенсори можуть створювати побічні канали витоку неструктурованих даних. Дослідники обґрунтували необхідність запровадження обмежень на рівні операційної системи, а також розроблення продуманої моделі дозволів і прозорості для користувача [8]. P. Weichbroth і Ł. Łysik узагальнили сучасний ландшафт мобільних загроз і практик безпеки, підкреслили важливість поєднання гігієни дозволів, сегментації середовища і вбудованої криптографії з підходами, орієнтованими на дані та їхній рух між контейнерами [9]. A. Muñoz наголошує, що апаратні уразливості платформи Android здатні нівелювати прикладні політики безпеки, що свідчить про необхідність врахування взаємодії між апаратним і програмним рівнями в моделях ризику, а також про залежність рівня довіри від конкретної реалізації платформи [10].

S. Lee та колеги запропонували використання федеративного навчання для виявлення шкідливих застосунків із збереженням приватності і водночас окреслили виклики, пов'язані з різнорідністю пристроїв і загрозами отруєння під час розподіленого тренування моделей [11]. S. Ullah та співавтори дослідили поведінку користувачів щодо надання дозволів і виявили систематичні хиби в оцінці ризику, що пояснює витоки даних через формально дозволені доступи та вимагає вдосконалення інтерфейсів керування приватністю [12]. A. K. Sikder та співавтори здійснили огляд сенсорно орієнтованих атак і побічних каналів для смартпристроїв і підсумували необхідність багаторівневих механізмів захисту на етапах збирання і попередньої обробки даних [13]. S. Hutchinson та співавтори, на прикладі фітнес-застосунків, показали, як журнали активності і телеметрія формують вторинні носії неструктурованих даних із ризиковими налаштуваннями приватності, що має важливе значення для реалізації політик DLP безпосередньо на пристрої [14]. Z. Muhammad та колеги узагальнили загрози, що походять від цілеспрямованих атак і невідомих вразливостей у контексті смартфонів. Дослідники посилили аргументацію на користь поведінково-контекстних контролів доступу та безперервної автентифікації [15].

A. Alkinoon та співавтори простежили еволюцію шаблонів використання дозволів і характеристик мобільних застосунків і запропонували індикатори для профілювання ризику та пріоритизації політик на рівні ендпоїнта [16].

D. Progonov та співавтори узагальнили методи поведінкової автентифікації на мобільних пристроях у різних умовах використання та звернули увагу на точність, частоту хибних спрацювань і стійкість до відтворення поведінки користувача [17].

З урахуванням наведеного аналізу, залишаються відкритими низка критичних питань, що стримують прогрес у сфері захисту неструктурованих даних на мобільних пристроях. Насамперед, відсутність узгоджених публічних бенчмарків для реалістичних сценаріїв ускладнює відтворюваність результатів і коректне порівняння підходів. Якість моделей значною мірою залежить від складу датасетів та протоколів оцінювання, що створює ризики обмеженої узагальнюваності. Особливо слабкою є міжплатформна переносимість: навіть високоточні моделі демонструють деградацію продуктивності при перенесенні між версіями ОС і класами пристроїв. Ця проблема загострюється компромісами між точністю, швидкодією та енергоспоживанням, а також труднощами кореляції потоків у багатопристроєвих конфігураціях. Недостатньо вирішеною залишається стійкість до відтворення поведінки користувача (spoofing), маніпуляцій із сенсорами та атак отруєння даних у процесі розподіленого навчання. Незважаючи на значний прогрес у дослідженнях, зберігаються апаратно-програмні ризики, здатні нівелювати прикладні політики безпеки, а також поведінкові хиби користувачів процесі надання дозволів. Це сприяє формуванню вторинних носіїв чутливої інформації у телеметрії та журналах активності.

Водночас бракує інтегрованих рішень, що поєднують безперервну автентифікацію з даноцентричними політиками доступу для неструктурованого контенту безпосередньо на пристрої, з урахуванням руху даних між контейнерами та загроз цілеспрямованих атак. Потребують подальшого розвитку прозорість і пояснюваність моделей, а також баланс між точністю, приватністю та зручністю для користувача. Ці аспекти підтверджують узагальнення та емпіричні спостереження останніх дослідженнях.

Формулювання мети дослідження

Метою статті є обґрунтування та пропозиція цілісної рамки захисту неструктурованої інформації на мобільних пристроях, яка поєднує безперервну поведінкову автентифікацію, даноцентричні політики доступу та сучасні криптографічні засоби. Запропонована модель враховує обмеження продуктивності, енергоспоживання та вимоги до приватності, що є критично важливими для мобільного середовища. Актуальність дослідження зумовлена тим, що саме неструктурований контент становить більшість корпоративних даних, активно мігрує між контейнерами та каналами передачі, і є основною мішенню витоків інформації. Наявні підходи до його захисту залишаються фрагментованими, не забезпечують узгоджених критеріїв вибору рішень і не гарантують відтворюваності результатів у реалістичних сценаріях.

Для досягнення поставленої мети спочатку сформовано формальну модель загроз і контекстів використання неструктурованого контенту на мобільному ендпоїнті; уточнено, де й чому виникають витоки даних, а також проаналізовано вплив платформи, системи дозволів і поведінки користувача.

Наступним етапом здійснено порівняльний аналіз методів шифрування та автентифікації, у межах якого розглянуто практичні ролі й обмеження алгоритмів AES та ECC у забезпеченні захисту контенту та керуванні ключами, а також досліджено взаємодоповнюваність класичної багатофакторної автентифікації та поведінкової аналітики у формуванні довіри на етапі входу й упродовж сесії.

На основі отриманих результатів запропоновано архітектурний підхід, який поєднує безперервну автентифікацію з класифікацією та маркуванням неструктурованих даних на пристрої, а також він включає політику запобігання витокам, орієнтовану на контент та його рух між контейнерами. Наступним кроком визначено відтворювану методику оцінювання, що охоплює чітко визначені метрики точності, затримки, енерговитрат, стійкості до відтворення поведінки, сенсорного спуфінгу та здатності до узагальнення між платформами.

Завершальним етапом є узагальнення практичних рекомендацій щодо впровадження запропонованих рішень у сценаріях корпоративного і змішаного володіння пристроями. Окрему увагу приділено вимогам до прозорості та пояснюваності процесу прийняття рішень, що забезпечує можливість їх аудиту з позицій безпеки і приватності.

Такий план забезпечує логічну послідовність викладу – від постановки ризиків до інструментальних рішень і критеріїв перевірки. Він робить внесок у науку через систематизований порівняльний аналіз і відтворюваний протокол оцінювання. Для практичного застосування план пропонує цільову архітектуру та набір перевірених кроків, що сприяють зниженню ймовірність витоків даних і підвищенню керованості неструктурованої інформації на мобільних платформах.

Викладення основного матеріалу дослідження

Мобільні пристрої стали основним середовищем, у якому виникає й циркулює неструктурована інформація, зокрема повідомлення, фото, голосові нотатки, кешовані дані та вкладення. Після розблокування пристрою цей тип контенту є найбільш уразливим, оскільки зберігається у різних контейнерах і швидко мігрує між застосунками та хмарними сервісами. Традиційні разові перевірки під час входу й периметрові бар'єри не забезпечують належного рівня безпеки в умовах динамічного середовища мобільних платформ. У зв'язку з цим виникає потреба у рішеннях, що діють упродовж усього сеансу та враховують поведінку, контекст і властивості даних. У подальшому викладі послідовно демонструється, як поєднання безперервної автентифікації, даноцентричних політик доступу та сучасних криптографічних методів сприяє зменшенню вікна ризику для неструктурованого контенту на смартфоні й забезпечують керованість без суттєвого впливу на зручність користування.

Одноразова автентифікація під час входу не усуває ключову вразливість мобільних сценаріїв. Після розблокування пристрою зловмисник може використати короткий проміжок часу, коли увага власника розсіяна, щоб отримати доступ до листування, фото, кешів застосунків та робочих документів, що не мають жорстко визначеної структури. Безперервна та неявна автентифікація зменшує це вікно ризику, оскільки система постійно звіряє поточну взаємодію з очікуваним профілем користувача й оновлює оцінку довіри протягом усього сеансу [1, с. 1358]. Важливим є врахування дрейфу поведінки. Зміни у хваті, темпі введення тексту та жестах під час руху містом або перебування у транспорті призводять до зниження точності, якщо модель автентифікації є статичною. Адаптивний підхід, який дозволяє системі змінювати рішення у часі та динамічно налаштовувати пороги прийняття, забезпечує вищу стійкість до природних коливань і повсякденних відхилень [1, с. 1360]. Доцільно також оцінювати не лише факт помилки автентифікації, а й швидкість її виявлення. Чим раніше система фіксує нетипову взаємодію, тим менше повідомлень, файлів і тимчасових артефактів встигає потрапити до сторонньої особи. У виробничих умовах це безпосередньо пов'язано з політикою доступу до неструктурованого контенту, який підтягується на екрані за запитом і може бути вибірково прихований у разі зниження рівня довіри [1, с. 1361]. Таким чином, перехід від одноразової перевірки до безперервної оцінки ідентичності формує базову лінію для подальшої логіки контролю даних на пристрої [1, с. 1359].

Надійність безперервної перевірки зростає за умови використання кількох незалежних ознак. Окремі модальності є чутливими до контексту і шумів, тоді як інтеграція кількох каналів дозволяє вирівнювати випадкові сплески та знижує ефективність спроб обману. Концептуально злиття можливе на різних рівнях – від об'єднання сирих ознак до поєднання часткових рішень, що наєдає проєктувальнику гнучкість у відповідності до обмежень часу і енергоспоживання [2, с. 2]. Поєднання фізіологічних і поведінкових характеристик підвищує стабільність у повсякденних діях, як-от прокручування стрічки або швидкий перегляд вкладень. При цьому не вимагається додаткових дій з боку користувача, якщо пороги й розміри ковзних вікон налаштовані відповідно до сценарію використання [2, с. 3]. Для мобільних систем, що працюють з неструктурованим контентом, доцільно застосовувати рішення з адаптивними вікнами накопичення сигналів. У випадках короткої взаємодії вікно стискається, що забезпечує швидке реагування моделі. Коли користувач читає довгий ланцюжок повідомлень або виконує повторювані жести, система накопичує більше контексту і знижує ймовірність хибного спрацьовування без втрати чутливості [2, с. 4].

У фінансових застосунках, де ціна помилки є особливо високою, практичним виявився підхід до безперервної перевірки на основі дотикової динаміки. Формуються профілі взаємодії з сенсорною панеллю, які відображають темп, траєкторії, силу натиску та мікропаузи. У таких системах доцільно розділяти фази логіна і постлогіна що дозволяє підтримувати низький рівень помилок і водночас уникати надмірного навантаження на користувача під час перегляду історій і вкладень [3, с. 2].

Багатоскоупне представлення дає змогу виявляти як короткі, так і довші патерни, що покращує стабільність рішення у змішаних сесіях банкінгу, коли користувач часто перемикається між екранами. Для захисту неструктурованого вмісту зручно застосовувати м'які обмеження. У разі зниження рівня впевненості система може приховувати попередні перегляди у списках або запитувати швидке підтвердження перед відкриттям вкладення, замість повного завершення сесії [3, с. 1].

Окрему увагу слід приділяти ризику імітаційних атак. У випадках, коли зловмисник намагається відтворити схожі жести за відеозаписом або тренується на викраденому пристрої, система з багатоскоупним відбором ознак ефективніше виявляє нетипові мікропатерни, знижуючи ймовірність непомітного доступу до чутливих артефактів сеансу [3, с. 4]. Цей підхід природно переноситься з банкінгу на менеджери файлів, електронну пошту та месенджери, оскільки характер взаємодії з екраном у кожного користувача є досить стабільним у межах одного дня [3, с. 5].

Інший напрямок пов'язаний із використанням інерціальних сенсорів для непомітного відтворення поведінкового підпису. Дані акселерометра, гіроскопа і магнітометра формують часові послідовності, на основі яких модель навчається відрізняти власника від сторонньої особи – навіть тоді, коли на екрані відбувається пасивне читання або швидке гортання списків. Це особливо актуально для контенту, який не має фіксованих структур і відкривається пакетно, як-от галереї або довгі чати [4, с. 1]. Результативність у таких сценаріях забезпечують глибокі архітектури, що поєднують згорткові і рекурентні компоненти. Згорткові шари ефективно виявляють локальні ритми руху, а рекурентна частина – відслідковує довготривалі залежності, що разом сприяє зниженню кількості помилкових рішень у реальних умовах, де змінюється темп взаємодії та спосіб носіння пристрою [4, с. 3]. Практична реалізація таких систем потребує ретельної нормалізації розташування й орієнтації телефона, а також вибору частоти дискретизації, яка не призводить до надмірного споживання енергії. За таких умов система залишається прозорою для користувача та не потребує додаткових дій під час роботи з повідомленнями або перегляду документів, що зберігаються у кеші застосунків [4, с. 2]. Підхід на основі інерціальних сенсорів добре узгоджується з ідеєю фонової перевірки. Модель працює у тлі і підвищує рівень контролю лише у випадках зростання сукупного ризику для поточної сесії – наприклад, коли характер рухів раптово відрізняється від звичного профілю [4, с. 4].

У розгорнутих сесіях доцільно додавати адаптивність політик через підкріплювальне навчання. Агент спостерігає за стисненим вектором показників якості перевірки та контекстних ознак, поступово навчаючись активувати або посилювати контроль у разі зростання ризику, а також зменшувати його за умов стабільної поведінки користувача. Це дозволяє досягти балансу між безпекою та зручністю для користувача, який працює з файлами та повідомленнями без додаткових бар'єрів [5, с. 2]. Важливо правильно формувати функцію винагороди, щоб система не перетворювалася на агресивний фільтр, який постійно перериває взаємодію в умовах підвищеної складності. Натомість, за правильної постановки задачі агент підтримує низький рівень інтегральної помилки та зберігає прийнятну точність на тестуванні з реальними даними, що підтверджує життєздатність такого способу керування ризиком до неструктурованих артефактів сеансу [5, с. 3]. Додатковою перевагою є енергоефективність. Агент навчається звертатися до більш витратних сенсорів або ініціювати явне підтвердження лише тоді, коли сукупна невизначеність перевищує безпечний поріг, а отже зменшує кількість зайвих звернень до користувача і подовжує час автономної роботи пристрою [5, с. 1].

У багатопристроєвих сценаріях контроль ідентичності має спиратися не на один канал спостереження, а на узгоджене трактування кількох потоків сигналів, що надходять від смартфона, носимих сенсорів і допоміжних периферійних пристроїв. Такий підхід дає змогу виявляти розбіжності між очікуваним профілем взаємодії та фактичною поведінкою, знижуючи ризик непомітного доступу до неструктурованого вмісту, коли один з вузлів екосистеми виявляється скомпрометованим або тимчасово недоступним [6, с. 2]. Ефективність підходу забезпечується застосуванням правил кореляції і часових вікон, у яких рішення кожного вузла важить пропорційно його надійності для конкретного контексту, що підвищує стабільність розпізнавання особи під час швидких перемикань між застосунками та режимами взаємодії з файлами й повідомленнями [6, с. 5]. У виробничих умовах це дозволяє реалізовувати м'які дії замість повного блокування – наприклад, система може приховати попередні перегляди вкладень у чатах або відкладене розкриття зображень до моменту, коли агрегований показник довіри не повернеться у безпечний діапазон [6, с. 7].

Систематизація методів сенсорно орієнтованої перевірки дозволяє виявити критичні обмеження, які безпосередньо впливають на захист неструктурованого контенту. Найбільш критичні серед них – нестача публічних репрезентативних датасетів, розбіжності у протоколах валідації та відсутність уніфікованих метрик, що ускладнює порівнянність і відтворюваність результатів у реальних мобільних сценаріях [7, с. 9130]. Для практики це означає, що навіть високі значення точності, отримані на контрольних вибірках, можуть не переноситися на ситуації, де користувач одночасно працює з галереєю, електронною поштою, нотатками і кешованими даними, а пристрій піддається впливу шумів руху та змін хвату [7, с. 9136]. Акцент на прозорих метриках і відкритих процедурах порівняння створює основу для впровадження політик доступу, які базуються не на декларативних порогах, а на перевірених показниках часу виявлення ймовірної підміни в активному сеансі [7, с. 9139].

Приватність сенсорів перетворюється на центральний чинник, оскільки з сирих потоків формується поведінковий підпис, що керує доступом до чутливого контенту. Аналіз уразливостей показує, що низка сенсорів здатна відновлювати фрагменти взаємодії користувача або витягувати індикатори контексту навіть без явних дозволів, що створює побічні канали витоку інформації, особливо у фонових процесах, що супроводжують роботу з неструктурованими даними [8, с. 4]. особливо у фонових процесах, що супроводжують роботу з неструктурованими даними Практичне вирішення цієї проблеми полягає у мінімізації збору даних, локальному навчанні моделей безпосередньо на пристрої та впровадженні політик, які не зберігають і не передають сирі траєкторії або графи взаємодії, залишаючи за межами пристрою лише агреговані ознаки або короткочасні оцінки ризику для прийняття рішень у реальному часі [8, с. 12]. Така організація усуває потребу у централізованому накопиченні даних про користувача, водночас зберігаючи здатність системи оперативно реагувати на ознаки захоплення сесії під час переходів між застосунками та контейнерами [8, с. 18].

З урахуванням загрозливого ландшафту архітектура захисту має виходити за межі моделі одноразового логіна і повинна включати сегментацію доступу до даних, жорстке керування дозволами та контроль міжпроцесних взаємодій. Огляд тенденцій свідчить, що поєднання гігієни дозволів із наскрізним шифруванням і контекстними перевірками під час роботи із вкладеннями, кешами та журналами знижує площу атаки, зменшує вразливість до соціальної інженерії та перенесення зловмисних активностей між застосунками [9, с. 5]. Для неструктурованих даних доцільним є запровадження режимів попереднього приховання приватного вмісту, зокрема, показ чутливих прев'ю відкладається до моменту отримання додаткового сигналу довіри від локальної поведінкової моделі, що мінімізує експозицію у випадках короткочасного відволікання користувача [9, с. 9]. Крім того, рекомендовано пов'язувати життєвий цикл тимчасових копій та індексів пошуку з поточним рівнем ризику, що дозволяє уникнути неконтрольованого накопичення відбитків неструктурованого контенту у фонових сховищах, що підвищує загальний рівень безпеки [9, с. 12].

Надійність рішень на рівні застосунків залежить від глибинної стійкості платформи. Дослідження уразливостей апаратних компонентів показує, що недоліки у мікропрограмному забезпеченні, контролерах пам'яті та каналах міжпроцесної взаємодії здатні обходити механізми розмежування. Це створює ризик несанкціонованого

читання або копіювання чутливого вмісту поза межами прикладних політик [10, с. 6]. З огляду на це, постає потреба у багаторівневому підході до захисту, де поведінкова перевірка доповнюється регулярним оновленнями мікропрограм, вимкненням відомо небезпечних інтерфейсів і прив'язкою дій доступу до даних до перевіреного стану пристрою, що дозволяє, навіть у разі часткового компромісу апаратного шару, припинити виконання чутливих операцій із неструктурованим контентом або перевести їх у деградований режим перегляду без можливості експорту чи копіювання [10, с. 9]. Це поєднання знижує ймовірність успішного обходу контролів, особливо у випадках, коли зловмисник намагається використати відомі апаратні недоліки під час активного сеансу роботи з файлами, галереями чи месенджерами [10, с. 12].

У випадках, коли моделі виявлення шкідливих застосунків навчаються централізовано, процес збору сирих даних створює додаткові канали витоку неструктурованого контенту, тому доцільно переносити навчання ближче до еджу і агрегувати лише параметри моделей, зберігаючи телеметрію на пристроях [11, с. 1]. Практичні експерименти з федеративним навчанням показують, що за умови раціонального підбору гіперпараметрів, навіть при неідеальному розподілі даних між клієнтами, можливо досягти прийнятної точності детекції без передачі сирих артефактів взаємодії користувача зі застосунками, що скорочує експозицію неструктурованих даних під час активних сесій [11, с. 3]. Водночас слід враховувати обмеження ефективності у випадках патологічно нерівномірного розподілу даних, коли локальні вибірки є надто вузькими за класами, а глобальна модель повільно збігається, що слід враховувати при проєктуванні політик доступу, які залежать від швидкості виявлення ризику [11, с. 5].

На рівні поведінки користувача джерелом ризику стає неправильна оцінка сенсу дозволів, коли доступ до місцеположення, контактів або сховищ надається автоматично, без усвідомлення наслідків. У таких випадках навіть легальні запити можуть створювати обхідні шляхи для захисту неструктурованого вмісту [12, с. 1]. Емпіричні спостереження підтверджують, що значна частка інсталяцій супроводжується погодженням на надлишкові дозволи, що зумовлено браком прозорих підказок і недостатнім розумінням моделей доступу, що вимагає поєднувати технічні контролі з інтерфейсними втручаннями під час роботи з файлами, вкладеннями і кешами [12, с. 6].

Сенсори мобільних пристроїв формують окрему площину атак, де побічні канали зчитують або відновлюють чутливі атрибути взаємодії навіть за відсутності явного доступу до вмісту. Це явище безпосередньо стосується неструктурованих даних, що виникають у процесі звичайних жестів та контекстних змін [13, с. 1323]. Систематизація таких загроз засвідчує необхідність у багаторівневому захисті на етапі збору сигналів, зокрема через нормалізацію частоти вибірки, застосування методів шумопригнічення і впровадження політик мінімізації збереження, щоб зменшити сліди поведінки, які можуть бути використані для вторинної реконструкції вмісту [13].

Форензичні дослідження мобільних застосунків для носимих фітнес-пристроїв демонструють, що значні обсяги артефактів зберігаються локально у вигляді журналів кешів і проміжних файлів. Ці об'єкти часто містять непрямі посилання на повідомлення, зображення або історії взаємодій, які користувач вважає приватними [14]. Аналіз карти збережених артефактів у таких клієнтських застосунках свідчить, що навіть без доступу до основної бази даних можливо відновити часові ланцюжки активності та опосередковані індикатори стану здоров'я, що підсилює аргументацію на користь даноцентричних політик приховання та контролю життєвого циклу тимчасових копій неструктурованого контенту [14].

На тлі цілеспрямованих атак і швидкої еволюції платформ класичні підходи, базуються на статичних сигнатурах, демонструють обмежену ефективність. Сучасні обхідні техніки активно експлуатують як незакриті уразливості, так і помилки конфігурацій дозволів у застосунках, що працюють з файлами і повідомленнями [15]. Огляд сучасних технік підкреслює важливість безперервних поведінково-контекстних контролів і багатошарових криптографічних практик для зниження ризику витоків у динамічних сценаріях, коли неструктуровані дані швидко мігрують між контейнерами і сервісами [15].

Аналіз динаміки дозволів в екосистемі Android свідчить про поступове зростання як кількості, так і різноманіття запитів доступу. При цьому виявляються стійкі патерни надмірних запитів у певних категоріях застосунків, що має бути враховано при формуванні моделей ризику для неструктурованого вмісту [16]. Дослідження тенденцій на великій вибірці дозволяє сформувати пріоритети для політик доступу, зокрема ідентифікувати групи дозволів, які потребують більш жорстких перевірок у реальному часі під час відкриття вкладень або перегляду кешованих документів [16].

Для підтримання стійкого рівня довіри протягом усієї сесії доцільно поєднувати непомітні поведінкові перевірки з контекстним керуванням доступом. Такий підхід дозволяє приймати рішення на основі даних сенсорів і патернів взаємодії без необхідності частих явних автентифікацій під час роботи з повідомленнями файлами і медіа [17]. Експериментальні оцінки у різних сценаріях використання підтверджують, що мультисенсорні та контекстно залежні моделі здатні знижувати рівень помилок прийняття та відхилення, водночас зберігаючи прийнятний рівень зручності. Це створює передумови для практичної інтеграції таких моделей із політиками захисту неструктурованого контенту на рівні ОС і застосунків [17].

З метою узгодження технічних рішень із криптографічними та автентифікаційними засадами, необхідно спершу конкретизувати ролі та порівнювану стійкість симетричних і асиметричних алгоритмів задля захисту

неструктурованого контенту на мобільному пристрої, а вже потім – правила підтримання довіри під час тривалих сесій доступу до такого контенту. [18, с. 66–67] Симетричні алгоритми, як правило, застосовуються для шифрування самих даних, тоді як асиметричні – для встановлення або захисту сеансового ключа, а також для цифрових підписів [18, с. 40–43] У практичній конфігурації це зводиться до гібридної зв'язки «контент шифрується симетрично, а симетричний ключ передається та засвідчується асиметрично», що водночас забезпечує керованість, продуктивність і узгодженість криптографічних гарантій [18, с. 66–67]. Для узгодження вибору криптографічних механізмів із вимогами до порівнюваної стійкості та ролей алгоритмів систематизовано переваги й обмеження симетричного AES і асиметричного ECC у мобільному контексті захисту неструктурованого контенту (див. табл. 1).

Таблиця 1

Переваги й обмеження AES та ECC

Алгоритм	Переваги	Обмеження
AES (симетричний)	1) Придатний для масового шифрування даних «у спокої» та під час передавання завдяки особливостям симетричних шифрів. 2) Дає чітко визначені рівні стійкості 128/192/256 біт та прогнозовані криптографічні властивості. 3) Легко інтегрується в гібридні криптографічні схеми, де відповідає за шифрування контенту. 4) Знижує витрати на протокол, оскільки один ключ може використовуватися для шифрування великих обсягів. 5) Узгоджується з вимогою щодо порівнюваної стійкості при поєднанні з асиметричними механізмами	1) Потребує безпечного способу встановлення/розповсюдження симетричного ключа за допомогою додаткових механізмів. 2) Компрометація сеансового ключа призводить до розкриття всього зашифрованого в межах сеансу контенту. 3) Не забезпечує неоспорюваності дій без використання додаткових засобів підпису. 4) Не вирішує завдання автентифікації сторін без додаткових механізмів. 5) Вимагає узгодження стійкості з асиметричними компонентами
ECC (асиметричний)	1) Дозволяє встановлювати або приховувати симетричний ключ із використанням коротших ключів за умови за зіставної стійкості. 2) Дає можливість створювати цифрові підписи для атестації дій. 3) Ефективно інтегрується в протоколи доступу та керування ключами. 4) У гібридних схемах відокремлює функції «ключі ↔ дані», підвищуючи керованість. 5) Дозволяє досягати рівня криптостійкості, порівнюваного з AES, при менших розміріах ключів	1) Не призначений для масового шифрування контенту, застосовується переважно для встановлення ключів та цифрових підписів. 2) Потребує наявності інфраструктури керування ключами/ сертифікатами. 3) За призначенням вимагає погодження параметрів із симетричними алгоритмами для забезпечення зіставної стійкості. 4) Криптографічні безпека контенту залежить від коректної від правильної інтеграції в гібридну схему. 5) Вибір кривих і параметрів визначається політикою безпеки та має бути узгоджений із вимогами до рівня стійкості

Джерело: власна розробка авторів на основі [18]

Підсумовуючи, доцільно використовувати симетричне шифрування для захисту контенту, а асиметричні алгоритми – для встановлення ключів і створення цифрових підписів, дотримуючись порівнюваних рівнів стійкості між класами алгоритмів, як це передбачено рекомендаціями.

На рівні контролю доступу під час реальної взаємодії з неструктурованим вмістом процес автентифікації не завершується на етапі входу, оскільки сесія має власний життєвий цикл і потребує підтримання довіри згідно з політиками керування сесією. Рекомендації передбачають періодичну повторну автентифікацію та допускають пасивний моніторинг ознак підміни, що дає змогу зменшувати «вікно ризику» між явними підтвердженнями без створення надмірного навантаження на користувача під час роботи з файлами, вкладеннями та кешами. Така організація процесу дозволяє підвищувати рівень контролю лише за умов зростання ризику, зберігаючи звичний характер взаємодії у стандартних сценаріях доступу до неструктурованих даних. [19, с. 56–58]. З метою визначення ролі MFA у контексті безперервного сесійного контролю та обґрунтування їхню взаємодоповнюваність під час роботи з неструктурованими даними, узагальнимо їхні переваги та обмеження у порівняльній таблиці 2.

Початкова MFA надійно ініціалізує процес входу, тоді як поведінкова аналітика підтримує довіру протягом сесії та запускає додаткові перевірки за наявності ризику; у поєднанні з гібридною зв'язкою «асиметрія для ключів ↔ симетрія для даних» це забезпечує як криптографічну, так і процесну цілісність захисту неструктурованого контенту на мобільних платформах [19, с. 57–58].

У підсумку, застосована конфігурація шифрування ґрунтується на гібридному підході: симетричний алгоритм використовується для захисту вмісту даних, а асиметричний – для встановлення й охорони сеансового ключа та підтвердження автентичності цифрових підписів, із забезпеченням порівнюваних рівнів стійкості між класами алгоритмів. Такий розподіл функцій гарантує необхідну продуктивність на мобільному ендпоїнті та криптографічну узгодженість у роботі з неструктурованими даними як у стані збереження, так і під час передавання.

Після первинного входу підтримання довіри реалізується через політики керування сесією, які передбачають періодичну повторну автентифікацію та пасивний моніторинг, що дозволяє скоротити «вікно ризику» для доступу до повідомлень, вкладень і кешів у процесі активної взаємодії. Сесійний контроль активує ризик-адаптивні дії лише за наявності ознак відхилення – від вимоги повторного підтвердження до тимчасового маскування

Таблиця 2

Переваги й обмеження MFA та поведінкової аналітики

Метод	Переваги	Обмеження
MFA (login-phase)	1) Підвищує поріг автентифікації та знижує ризик компрометації на етапі входу. 2) Узгоджується з рівнями запевнення автентифікаторів і політикою доступу. 3) Дозволяє вимагати повторну автентифікацію за подіями або після завершення визначеного часу. 4) Формує чіткий момент ініціалізації довіри в системі. 5) Підтримує сценарії підвищеного ризику через механізм step-up згідно з політикою	1) Не забезпечує контроль сталість ідентичності протягом сесії між явними перевірками. 2) Додає крок автентифікації на вході, що може впливати на зручність. 3) Уразлива до перехоплення сесії після успішного входу без додаткових захисних механізмів. 4) Потребує чітко визначеної політики потворної автентифікації для тривалих сесій. 5) Не забезпечує контекстної оцінки ризику під час щоденної взаємодії без допоміжних засобів
Поведінкова аналітика (post-login)	1) Підтримує довіру упродовж сесії завдяки пасивному моніторингу відповідно до політики. 2) Зменшує «вікно ризику» між явними підтвердженнями. 3) Дозволяє адаптивно реагувати на ризики шляхом ініціювання повторної автентифікації або обмеження доступу. 4) Мінімізує кількість явних запитів у стандартних сценаріях роботи з контентом. 5) Узгоджується з політикою керування сесіями без порушення звичних сценаріїв взаємодії	1) Залежить від якості та релевантності обраних сигналів моніторингу. 2) Не замінює початкову автентифікацію, а функціонує як доповнення. 3) Вимагає точного налаштування порогів і тригерів для уникнення надмірних втручань. 4) Потребує узгодження з політиками повторної автентифікації та журналювання. 5) Має функціонувати в межах політики конфіденційності та мінімізації збору даних

Джерело: власна розробка авторів на основі [19]

прев'ю або паузи експорту – без створення надмірного навантаження у стандартних сценаріях. У сукупності ці механізми поєднують гарантії на вході з безперервною перевіркою ідентичності протягом сесії, формуючи нормативно обґрунтовану рамку захисту неструктурованого контенту на мобільних платформах.

Висновки

У роботі обґрунтовано цілісну модель захисту неструктурованих даних на мобільних пристроях, яка поєднує безперервну поведінкову автентифікацію, даноцентричні політики доступу та гібридну криптографічну конфігурацію. Показано, що така композиція знижує експозицію повідомлень, медіафайлів, вкладень і кешованих даних протягом тривалих сесій без створення надмірного навантаження на користувача і без істотного впливу на енергоспоживання. Гібридне поєднання симетричних і асиметричних механізмів забезпечує продуктивне шифрування контенту та надійне встановлення і захист сеансових ключів, а початкова багатофакторна перевірка узгоджується з пост-логін моніторингом поведінки, що дозволяє підтримувати довіру між явними підтвердженнями.

Практичні рекомендації охоплюють два рівні. На рівні автентифікації доцільним є поєднання дотикової динаміки та інерціальних сенсорів зі злиттям ознак і оцінок, а також використання ризик-адаптивних політик, які запускають повторне підтвердження або м'які обмеження лише за умов зростання загроз. На рівні доступу до даних ефективними є політики, що керують попередній перегляд, кешування та тимчасові копії, а також гігієна дозволів із урахуванням форензичних артефактів і платформних уразливостей. У сукупності це забезпечує керованість та відтворюваність рішень, узгоджених з вимогами реальних корпоративних сценаріїв.

Обмеження пов'язані з дефіцитом відкритих репрезентативних датасетів, розбіжностями в протоколах оцінювання, а також із необхідністю балансування між точністю і приватністю сенсорних потоків. Перспективи подальших досліджень охоплюють стандартизацію метрик і протоколів, розробку енергоефективних моделей для тривалих сесій, запровадження приватних і федеративних схем навчання, перевірку узгодженості політик на різних платформах та інтеграцію з корпоративними засобами керування пристроями. Запропонована модель створює надійну основу для поетапного впровадження та подальшого вдосконалення захисту неструктурованого контенту на мобільних платформах.

Список використаної літератури

1. James Jose C., S. Rajasree M. Implicit Continuous User Authentication for Mobile Devices based on Deep Reinforcement Learning. *Computer Systems Science and Engineering*. 2023. Vol. 44, no. 2. P. 1357–1372. URL: https://doi.org/10.32604/csse.2023.025672 (date of access: 10.09.2025).

2. Reichinger D., Sonnleitner E., Kurz M. Continuous Mobile User Authentication Using Combined Biometric Traits. *Applied Sciences*. 2021. Vol. 11, no. 24. P. 11756. URL: https://doi.org/10.3390/app112411756 (date of access: 10.09.2025).

3. A Framework for Continuous Authentication Based on Touch Dynamics Biometrics for Mobile Banking Applications / P. M. A. B. Estrela et al. *Sensors*. 2021. Vol. 21, no. 12. P. 4212. URL: https://doi.org/10.3390/s21124212 (date of access: 10.09.2025).

4. Mekruksavanich S., Jitpattanakul A. Deep Learning Approaches for Continuous Authentication Based on Activity Patterns Using Mobile Sensing. *Sensors*. 2021. Vol. 21, no. 22. P. 7519. URL: https://doi.org/10.3390/s21227519 (date of access: 10.09.2025).

5. Bansal P., Ouda A. Continuous Authentication in the Digital Age: An Analysis of Reinforcement Learning and Behavioral Biometrics. *Computers*. 2024. Vol. 13, no. 4. P. 103. URL: https://doi.org/10.3390/computers13040103 (date of access: 10.09.2025).

6. Lee H.-H., Sung H.-C. Unveiling the Confirmation Factors of Information System Quality on Continuance Intention towards Online Cryptocurrency Exchanges: The Extension of the Expectation Confirmation Model. *Information*. 2023. Vol. 14, no. 9. P. 482. URL: https://doi.org/10.3390/info14090482 (date of access: 10.09.2025).

7. Abuhamad M., Abusnaina A., Nyang D., Mohaisen D. Sensor-based continuous authentication of smartphones' users using behavioral biometrics: a contemporary survey. *IEEE Internet of Things Journal*. 2020. Vol. 7, no. 10. P. 9128–9143. URL: https://davidmohaisen.github.io/files/iotj20.pdf (date of access: 09.09.2025).

8. A Survey of Privacy Vulnerabilities of Mobile Device Sensors / P. Delgado-Santos et al. *ACM Computing Surveys*. 2022. 2022. Vol. 54, Iss. 11s. Art. 224. P. 1–30. URL: https://doi.org/10.1145/3510579 (date of access: 10.09.2025).

9. Weichbroth P., Łysik Ł. Mobile Security: Threats and Best Practices. *Mobile Information Systems*. 2020. Vol. 2020. P. 1–15. URL: https://doi.org/10.1155/2020/8828078 (date of access: 10.09.2025).

10. Muñoz A. Cracking the Core: Hardware Vulnerabilities in Android Devices Unveiled. *Electronics*. 2024. Vol. 13, no. 21. P. 4269. URL: https://doi.org/10.3390/electronics13214269 (date of access: 10.09.2025).

11. Lee S. Distributed Detection of Malicious Android Apps While Preserving Privacy Using Federated Learning. *Sensors*. 2023. Vol. 23, no. 4. P. 2198. URL: https://doi.org/10.3390/s23042198 (date of access: 10.09.2025).

12. Understanding Users' Behavior towards Applications Privacy Policies / S. Ullah et al. *Electronics*. 2022. Vol. 11, no. 2. P. 246. URL: https://doi.org/10.3390/electronics11020246 (date of access: 10.09.2025).

13. A Survey on Sensor-Based Threats and Attacks to Smart Devices and Applications / A. K. Sikder et al. *IEEE Communications Surveys & Tutorials*. 2021. Vol. 23, no. 2. P. 1125–1159. URL: https://doi.org/10.1109/comst.2021.3064507 (date of access: 10.09.2025).

14. Investigating Wearable Fitness Applications: Data Privacy and Digital Forensics Analysis on Android / S. Hutchinson et al. *Applied Sciences*. 2022. Vol. 12, no. 19. P. 9747. URL: https://doi.org/10.3390/app12199747 (date of access: 10.09.2025).

15. Smartphone Security and Privacy: A Survey on APTs, Sensor-Based Attacks, Side-Channel Attacks, Google Play Attacks, and Defenses / Z. Muhammad et al. *Technologies*. 2023. Vol. 11, no. 3. P. 76. URL: https://doi.org/10.3390/technologies11030076 (date of access: 10.09.2025).

16. Alkinoon A., Dang T. C., Alghuried A., Alghamdi A., Choi S., Mohaisen M., Wang A., Salem S., Mohaisen D. A Comprehensive Analysis of Evolving Permission Usage in Android Apps: Trends, Threats, and Ecosystem Insights. *Journal of Cybersecurity and Privacy*. 2025. Vol. 5, № 3. P. 58. URL: https://doi.org/10.3390/jcp5030058 (date of access: 11.09.2025).

17. Behavior-based user authentication on mobile devices in various usage contexts / D. Progonov et al. *EURASIP Journal on Information Security*. 2022. Vol. 2022, no. 1. URL: https://doi.org/10.1186/s13635-022-00132-x (date of access: 10.09.2025).

18. Temoshok D., Fenton J. L., Choong Y.-Y., Lefkovitz N., Regenscheid A., Galluzzo R., Richer J. P. Digital Identity Guidelines: Authentication and Authenticator Management. NIST Special Publication 800-63B-4. Gaithersburg: National Institute of Standards and Technology, 2025. 142 p. URL: https://doi.org/10.6028/NIST.SP.800-63B-4 (date of access: 11.09.2025).

19. Barker E., Chen L., Roginsky A., Vassilev A., Davis R., Simon S. Recommendation for key management. Part 1 – General. NIST Special Publication 800-57, Part 1, Rev. 5. Gaithersburg: National Institute of Standards and Technology, 2020. 171 p. URL: https://doi.org/10.6028/NIST.SP.800-57pt1r5 (date of access: 10.09.2025).

References

1. James Jose, C., & S. Rajasree, M. (2023). Implicit Continuous User Authentication for Mobile Devices based on Deep Reinforcement Learning. *Computer Systems Science and Engineering*, *44*(2), 1357–1372. https://doi.org/10.32604/csse.2023.025672

2. Reichinger, D., Sonnleitner, E., & Kurz, M. (2021). Continuous Mobile User Authentication Using Combined Biometric Traits. *Applied Sciences*, *11*(24), 11756. https://doi.org/10.3390/app112411756

3. Estrela, P. M. A. B., Albuquerque, R. d. O., Amaral, D. M., Giozza, W. F., & Júnior, R. T. d. S. (2021). A Framework for Continuous Authentication Based on Touch Dynamics Biometrics for Mobile Banking Applications. *Sensors*, *21*(12), 4212. https://doi.org/10.3390/s21124212

4. Mekruksavanich, S., & Jitpattanakul, A. (2021). Deep learning approaches for continuous authentication based on activity patterns using mobile sensing. *Sensors, 21*(22), 7519. https://doi.org/10.3390/s21227519

5. Bansal, P., & Ouda, A. (2024). Continuous authentication in the digital age: An analysis of reinforcement learning and behavioral biometrics. *Computers, 13*(4), 103. https://doi.org/10.3390/computers13040103

6. Lee, H.-H., & Sung, H.-C. (2023). Unveiling the confirmation factors of information system quality on continuance intention towards online cryptocurrency exchanges: The extension of the expectation confirmation model. *Information, 14*(9), 482. https://doi.org/10.3390/info14090482

7. Abuhamad, M., Abusnaina, A., Nyang, D., & Mohaisen, D. (2020). Sensor-based continuous authentication of smartphones' users using behavioral biometrics: A contemporary survey. *IEEE Internet of Things Journal, 7*(10), 9128–9143. https://davidmohaisen.github.io/files/iotj20.pdf

8. Delgado-Santos, P., Stragapede, G., Tolosana, R., Guest, R., Deravi, F., & Vera-Rodriguez, R. (2022). A Survey of Privacy Vulnerabilities of Mobile Device Sensors. *ACM Computing Surveys*. https://doi.org/10.1145/3510579

9. Weichbroth, P., & Łysik, Ł. (2020). Mobile security: Threats and best practices. *Mobile Information Systems, 2020*, 1–15. https://doi.org/10.1155/2020/8828078

10. Muñoz, A. (2024). Cracking the core: Hardware vulnerabilities in Android devices unveiled. *Electronics, 13*(21), 4269. https://doi.org/10.3390/electronics13214269

11. Lee, S. (2023). Distributed detection of malicious Android apps while preserving privacy using federated learning. *Sensors, 23*(4), 2198. https://doi.org/10.3390/s23042198

12. Ullah, S., Khan, M. S., Lee, C., & Hanif, M. (2022). Understanding Users' Behavior towards Applications Privacy Policies. *Electronics*, *11*(2), 246. https://doi.org/10.3390/electronics11020246

13. Sikder, A. K., Petracca, G., Aksu, H., Jaeger, T., & Uluagac, A. S. (2021). A Survey on Sensor-Based Threats and Attacks to Smart Devices and Applications. *IEEE Communications Surveys & Tutorials*, *23*(2), 1125–1159. https://doi.org/10.1109/comst.2021.3064507

14. Hutchinson, S., Mirza, M. M., West, N., Karabiyik, U., Rogers, M. K., Mukherjee, T., Aggarwal, S., Chung, H., & Pettus-Davis, C. (2022). Investigating Wearable Fitness Applications: Data Privacy and Digital Forensics Analysis on Android. *Applied Sciences*, *12*(19), 9747. https://doi.org/10.3390/app12199747

15. Muhammad, Z., Anwar, Z., Javed, A. R., Saleem, B., Abbas, S., & Gadekallu, T. R. (2023). Smartphone Security and Privacy: A Survey on APTs, Sensor-Based Attacks, Side-Channel Attacks, Google Play Attacks, and Defenses. *Technologies*, *11*(3), 76. https://doi.org/10.3390/technologies11030076

16. Alkinoon, A., Dang, T. C., Alghuried, A., Alghamdi, A., Choi, S., Mohaisen, M., Wang, A., Salem, S., & Mohaisen, D. (2025). A comprehensive analysis of evolving permission usage in Android apps: Trends, threats, and ecosystem insights. *Journal of Cybersecurity and Privacy, 5*(3), 58. https://doi.org/10.3390/jcp5030058

17. Progonov, D., Cherniakova, V., Kolesnichenko, P., & Oliynyk, A. (2022). Behavior-based user authentication on mobile devices in various usage contexts. *EURASIP Journal on Information Security*, *2022*(1). https://doi.org/10.1186/s13635-022-00132-x

18. Temoshok, D., Fenton, J. L., Choong, Y.-Y., Lefkovitz, N., Regenscheid, A., Galluzzo, R., & Richer, J. P. (2025). *Digital identity guidelines: Authentication and authenticator management* (NIST Special Publication 800-63B-4). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-63B-4

19. Barker, E., Chen, L., Roginsky, A., Vassilev, A., Davis, R., & Simon, S. (2020). *Recommendation for key management. Part 1 – General* (NIST Special Publication 800-57, Part 1, Rev. 5). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.SP.800-57pt1r5

Дата першого надходження рукопису до видання: 18.09.2025
Дата прийнятого до друку рукопису після рецензування: 13.10.2025
Дата публікації:

УДК 94:061.213(477)»19/20» DOI

В. Г. ВАСЕНКО
студент кафедри комп'ютерні науки
Херсонський національний технічний університет
ORCID: 0009-0003-2558-2588

Н. В. КОРНІЛОВСЬКА
кандидат технічних наук,
доцент кафедри інформатики і комп'ютерних наук
Херсонський національний технічний університет
ORCID: 0000-0002-8331-8027

С. В. ВИШЕМИРСЬКА
кандидат технічних наук,
доцент кафедри інформатики і комп'ютерних наук
Херсонський національний технічний університет
ORCID: 0000-0002-6343-7512

М. В. КАРАМУШКА
кандидат технічних наук,
доцент кафедри інформатики і комп'ютерних наук
Херсонський національний технічний університет
ORCID: 0000-0001-5982-4598

ЗАСТОСУВАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ СТВОРЕННЯ УТИЛІТ ОПЕРАЦІЙНОЇ СИСТЕМИ WINDOWS

У сучасному світі, Windows продовжує залишатися однією з найпопулярніших операційних систем для користувачів. Проте, при тривалому користуванні цією системою, люди часто зіткнюються з проблемами щодо швидкодії та накопиченням непотрібних файлів, що може вплинути на продуктивність комп'ютера. Засоби для очищення системи, які використовуються традиційно, не завжди враховують індивідуальні потреби користувачів, що може призвести до втрати важливих даних або недостатньо ефективного очищення. Зростання вимог до точності і безпеки процесу очищення підкреслює значення використання методів машинного навчання для автоматизації цього процесу.

У цій роботі досліджується розробка та впровадження інтелектуальної утиліти для очищення та оптимізації операційної системи Windows з використанням методів машинного навчання на основі бібліотеки ML.NET. Програма створена на мові C# із застосуванням WinUI 3 для побудови сучасного користувацького інтерфейсу та PowerShell для виконання системної оптимізації. Технологія машинного навчання дозволяє ефективно класифікувати файли за ознаками (тип, розмір, дата створення, дата змінення тощо) та визначати, чи є вони потенційно непотрібними. Розроблена програма пропонує як стандартний функціонал очищення за категоріями, так і функцію інтелектуального очищення, яка мінімізує ризик видалення важливих даних.

Для оцінки ефективності запропонованого рішення проведено серію експериментів з класифікації файлів різних категорій та порівняльний аналіз з вбудованими у Windows утилітами. Експерименти включали тестування точності моделі машинного навчання на наборі даних, що містить файли різних типів та характеристик. Дослідження демонструє значні переваги інтелектуального підходу, що дозволяє автоматизувати процес очищення системи та зменшити кількість помилок.

Дане дослідження має значну практичну цінність для розробників програмного забезпечення, які працюють над інструментами для очищення та оптимізації операційних систем. Воно демонструє, як сучасні методи машинного навчання можуть підвищити точність та безпеку очищення, відкриваючи шлях до більш інтелектуальних та зручних утиліт для підтримки продуктивності комп'ютерних систем.

Ключові слова: *машинне навчання, очищення системи, ML.NET, WinUI 3, C#, класифікація файлів, оптимізація Windows, PowerShell, інтелектуальна утиліта.*

© Васенко В. Г., Корніловська Н. В., Вишемирська С. В., Карамушка М. В., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

V. G. VASENKO
Student at the Department of Computer Science
Kherson National Technical University
ORCID: 0009-0003-2558-2588

N. V. KORNILOVSKA
Candidate of Technical Sciences,
Associate Professor at the Department of Computer Science
Kherson National Technical University
ORCID: 0000-0002-8331-8027

S. V. VYSHEMYRSKA
Candidate of Technical Sciences,
Associate Professor at the Department of Computer Science
Kherson National Technical University
ORCID: 0000-0002-6343-7512

M. V. KARAMUSHKA
Candidate of Technical Sciences,
Associate Professor at the Department of Computer Science
Kherson National Technical University
ORCID: 0000-0001-5982-4598

APPLICATION OF MACHINE LEARNING METHODS FOR CREATING WINDOWS OPERATING SYSTEM UTILITIES

In the modern world, Windows continues to remain one of the most popular operating systems for users. However, with prolonged use of this system, people often encounter problems regarding performance and accumulation of unnecessary files, which can affect computer productivity. Traditional system cleaning tools do not always consider individual user needs, which can lead to loss of important data or insufficiently effective cleaning. The growing requirements for accuracy and safety of the cleaning process emphasize the importance of using machine learning methods to automate this process.

This work investigates the development and implementation of an intelligent utility for cleaning and optimization of the Windows operating system using machine learning methods based on the ML.NET library. The program is created in C# language with the application of WinUI 3 for building a modern user interface and PowerShell for performing system optimization. Machine learning technology allows effective classification of files by features (type, size, creation date, modification date, etc.) and determines whether they are potentially unnecessary. The developed program offers both standard cleaning functionality by categories and an intelligent cleaning function that minimizes the risk of deleting important data.

To evaluate the effectiveness of the proposed solution, a series of experiments on file classification of various categories and comparative analysis with built-in Windows utilities were conducted. The experiments included testing the accuracy of the machine learning model on a dataset containing files of different types and characteristics. The research demonstrates significant advantages of the intelligent approach, which allows automating the system cleaning process and reducing the number of errors.

This research has significant practical value for software developers working on tools for cleaning and optimization of operating systems. It demonstrates how modern machine learning methods can improve the accuracy and safety of cleaning, opening the path to more intelligent and convenient utilities for maintaining computer system productivity.

Key words: *machine learning, system cleaning, ML.NET, WinUI 3, C#, file classification, Windows optimization, PowerShell, intelligent utility.*

Постановка проблеми

Сучасні операційні системи Windows накопичують надлишкові файли, що знижує продуктивність. Традиційні утиліти очищення не використовують машинне навчання для безпечної класифікації файлів.

Формулювання мети дослідження

Метою дослідження є розробка інтелектуальної утиліти для очищення та оптимізації операційної системи Windows з використанням методів машинного навчання на основі бібліотеки машинного навчання ML.NET та порівняння її ефективності з традиційними підходами до очищення системи.

Аналіз останніх досліджень і публікацій

Аналіз існуючих рішень показує, що сучасні утиліти очищення Windows, такі як BleachBit, базуються на статичних правилах категоризації файлів без можливості вибіркового видалення. Системні оптимізатори на зразок Chris Titus Windows Utility забезпечують широкий функціонал PowerShell-команд, однак мають обмежену прозорість процесів виконання.

Дослідження ефективності ML.NET для класифікації файлів демонструють можливість досягнення точності понад 80 % при використанні логістичної регресії та наївного баєсівського класифікатора. Однак відсутні комплексні рішення, що поєднують інтелектуальне очищення з машинним навчанням та системну оптимізацію в єдиному інтерфейсі.

Для досягнення поставленої мети виконано такі завдання: розроблено програмне забезпечення з графічним інтерфейсом, яке підтримує стандартний функціонал очищення системи та інтелектуальний режим на основі ML.NET; реалізовано класифікацію файлів за ознаками; проведено тестування точності моделі машинного навчання.

Викладення основного матеріалу дослідження

У технологічному стеку утиліти є такі компоненти:

- .NET – основа платформи розробки;
- C# – мова програмування для логіки додатку;
- WinUI 3 – платформа для графічного інтерфейсу Windows;
- Python – для створення моделей класифікації [1];
- ML.NET – фреймворк машинного навчання для інтелектуального очищення.

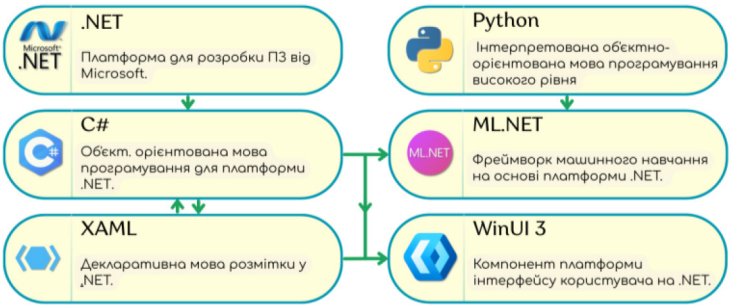


Рис. 1. Схема використаного технологічного стеку

Розроблена програмна система складається з чотирьох підпроєктів:

- OptimizeUtility – головний застосунок з графічним інтерфейсом на базі WinUI 3 та C#;
- OptimizeUtility.Core – бібліотека класів для операцій з файлами;
- OptimizeUtility.ModelCreator – Python-застосунок для створення моделей машинного навчання у форматі ONNX [1];
- OptimizeUtility.ScriptRunner – компонент для виконання PowerShell-скриптів.

Основою інтерфейсу є ShellPage.xaml – базова структурна «обгортка» застосунку, що реалізує навігаційний механізм. Структура включає:

1. AppTitleBar – кастомна верхня панель з логотипом та стандартними кнопками керування вікном.
2. NavigationView – вертикальна навігаційна панель з основними сторінками.

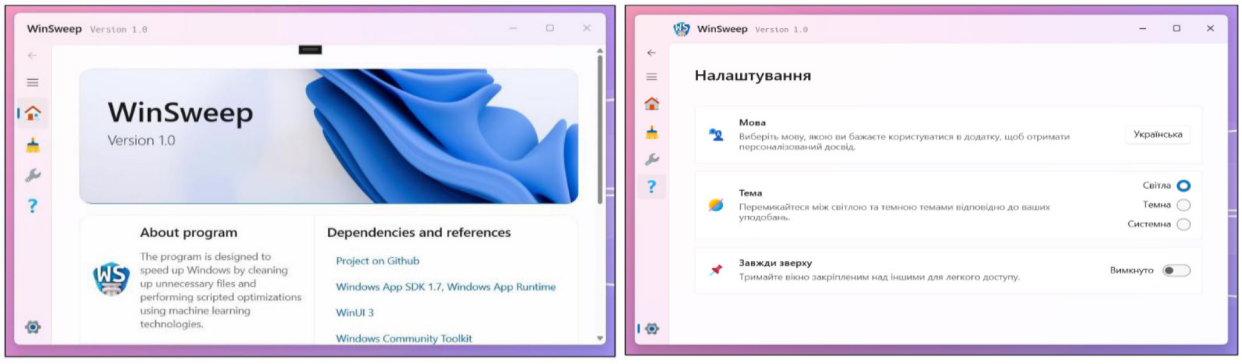


Рис. 2. Демонстрація головної сторінки програми та сторінки налаштувань

Сторінка очищення відповідає за стандартне та інтелектуальне очищення системи.

Стандартний режим очищення містить двоколонкову сітку: ліва частина – чекбокси для вибору категорій файлів, права частина – таблиця DataGrid з деталями файлів (чекбокси вибору, шлях, розмір, категорія). Під таблицею розташована панель керування з кнопками для запуску сканування по категоріях, видалення файлів, скасування сканування, активації/деактивації всіх чекбоксів.

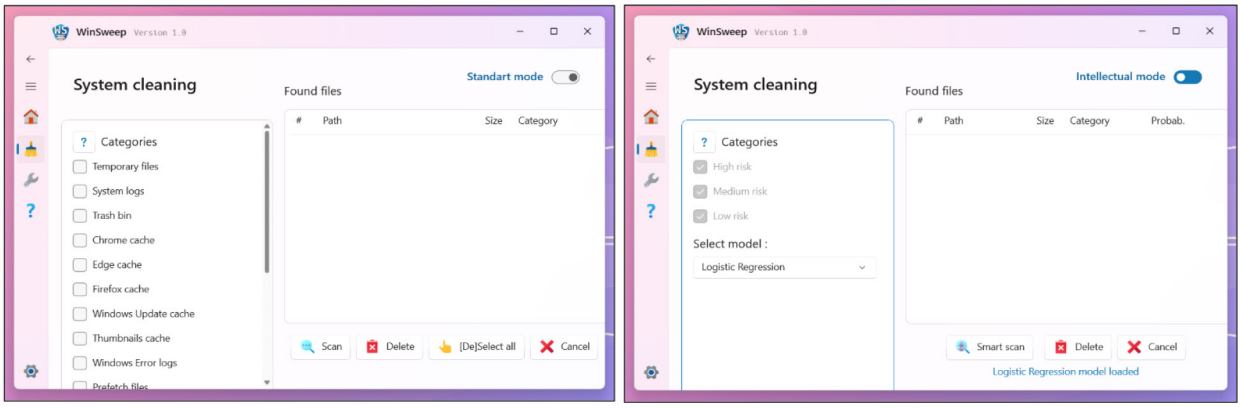


Рис. 3. Демонстрація сторінок очищення системи: ліворуч – стандартне; праворуч – інтелектуальне

Інтелектуальний режим активується через перемикач ToggleSwitch і має розширений функціонал порівняно зі стандартним режимом. Структура інтерфейсу подібна до StandardMode, але з додатковими можливостями машинного навчання.

Заголовки колонок містять додатковий стовпець «Вірогідність», який показує ймовірність того, що файл є надлишковим. Зліва додані чекбокси для фільтрації за вірогідністю та випадаючий список вибору моделі класифікації.

Ключовою особливістю SmartMode є колірне маркування рядків у таблиці результатів:

- Зелений – висока вірогідність, що файл непотрібний.
- Жовтий – середня вірогідність.
- Червоний – низька вірогідність.

Така візуалізація допомагає користувачеві швидко оцінити, які файли варто видалити, а які залишити, базуючись на результатах аналізу машинного навчання.

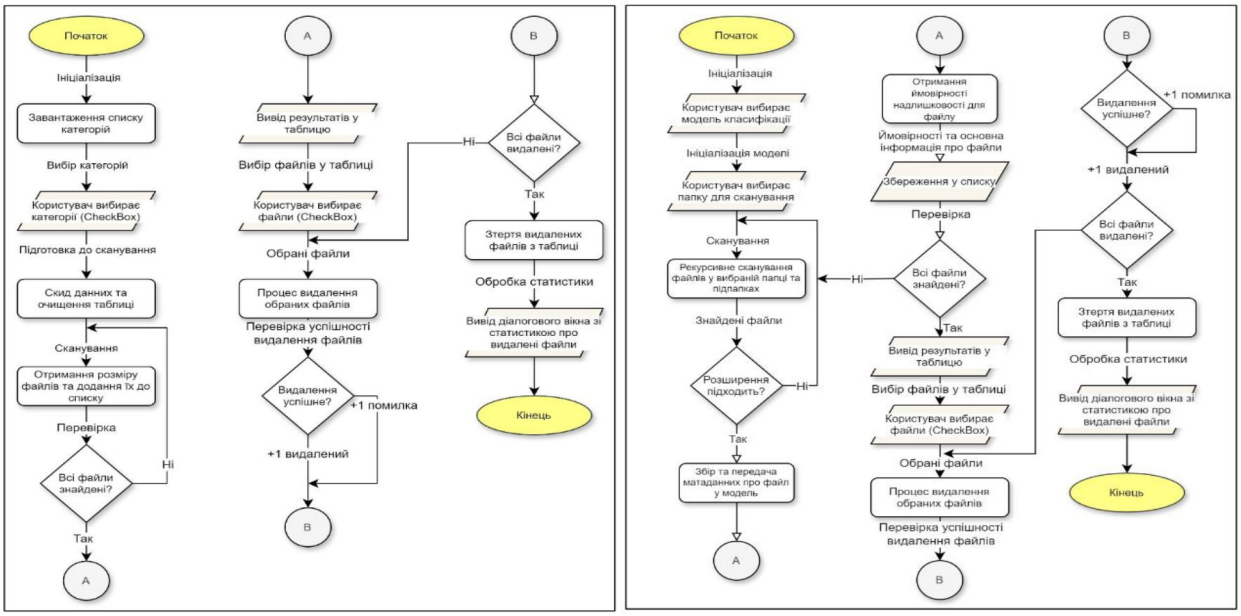


Рис. 4. Схеми очищення системи: ліворуч – стандартне; праворуч – інтелектуальне

Далі в нас йде наступний екран – це екран оптимізації системи або OptimizationPage, об'єктну схему якого можна побачити на рис. 4. Загалом принцип тут дещо схожий на попередній інтерфейс. Ми маємо ScrollViewer, в якому знаходиться певна кількість кнопок, які відповідають за додавання до текстового поля TextBox певних команд.

Після цього користувачу доступні три кнопки для взаємодії:

- Кнопка виконання записаних у текстове поле скриптів/команд.
- Кнопка очищення текстового поля від будь-яких записаних команд.
- Кнопка збереження вмісту текстового поля як файлу виконання PowerShell скриптів (тип файлу.ps1).

Тестування інтелектуального очищення:

Для експериментів сформовано набір даних з 7 ознак (Extension, SizeMB, AgeInDays, LastModifiedDays, IsTemporary, AccessFrequency) та цільовою змінною IsJunk. Числові ознаки стандартизовано, категоріальні перетворено через OneHotEncoder [2].

Розподіл цільової змінної збалансований: 60 % не сміттєвих файлів, 40 % сміттєвих. Для логістичної регресії використано GridSearchCV з крос-валідацією, наївний баєсівський класифікатор – з параметрами за замовчуванням [3].

Середні значення F1-міри за крос-валідацією:

- Логістична регресія: 0.7908 (±0.0480);
- Наївний баєсівський класифікатор: 0.7805 (±0.0288).

Точність на тестовій вибірці (1000 записів):

- Логістична регресія: 0.8320;
- Наївний баєсівський класифікатор: 0.8190.

Обидві моделі продемонстрували високу точність та стабільність метрик. Незначні розбіжності між результатами валідаційної та тестової вибірок можуть свідчити про початкові ознаки перенавчання, що потребує подальшого дослідження впливу мультиколінеарності ознак та розширення навчального датасету.

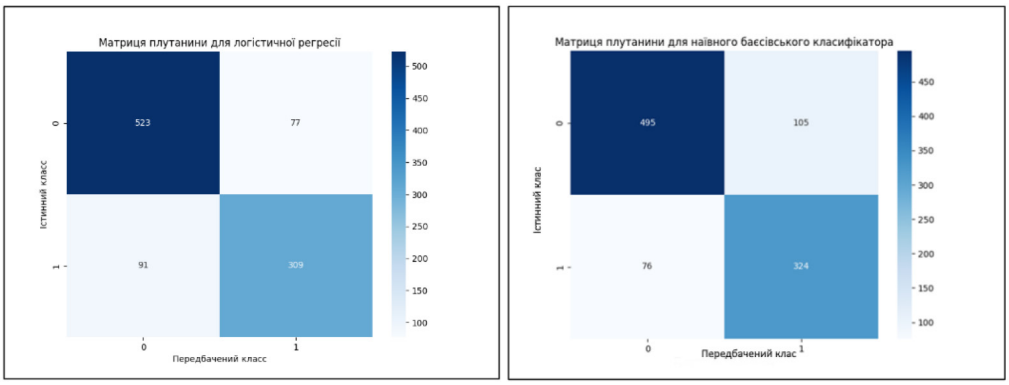


Рис. 5. Матриці плутанини: ліворуч – для логістичної регресії; праворуч – для наївного баєсівського класифікатора

Тестування стандартного очищення

Початкове тестування розпочалося з перевірки функції стандартного очищення. Під час розробки програми накопичилась значна кількість зайвих файлів у системі. Зазвичай після ручного очищення обсяг вільного місця на системному диску становить приблизно 42–46 ГБ.

Було проведено сканування з активацією всіх категорій пошуку. При стандартному запуску система ідентифікувала 4,6 ГБ надлишкових файлів, однак аналіз виявив обмеження через відсутність адміністративних прав. Після запуску з підвищеними правами повторне сканування виявило 23,4 ГБ надлишкових файлів. Реалізована функція масового перемикання чекбоксів дозволяє зручно обирати необхідні об’єкти для видалення.

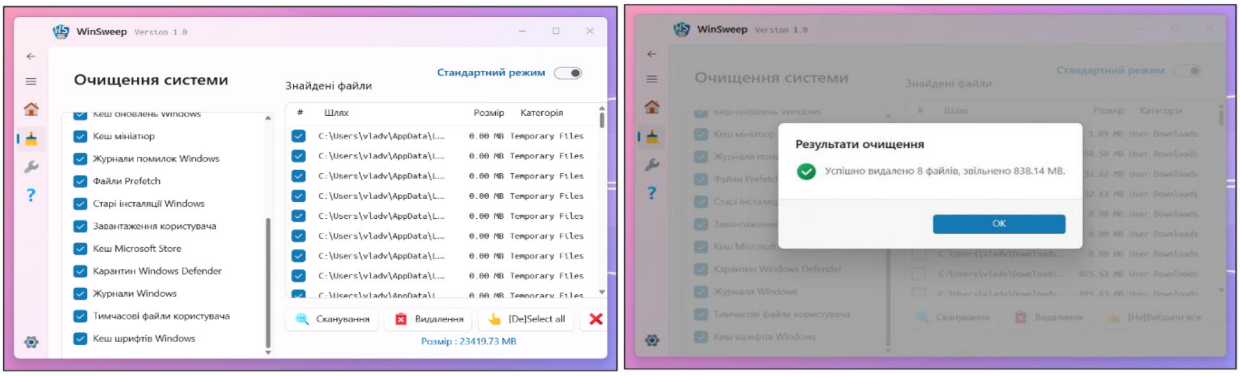


Рис. 6. Тестування стандартного очищення : ліворуч – знайденні сміттєві файли 23Гб; праворуч – результат часткового очищення

Тестування оптимізації системи

Для оцінювання ефективності оптимізації системи за допомогою PowerShell-команд проведено експеримент у віртуальній машині з «чистою» Windows 11.

Алгоритм тестування: спочатку фіксувалися базові показники системи через «Диспетчер задач» протягом 10 секунд, потім застосовувалися оптимізаційні скрипти з подальшим перезавантаженням та повторним заміром. Програмний модуль реалізує 15 команд оптимізації: перевірка цілісності системних файлів, очищення DNS-кешу, вимкнення телеметрії, переведення сервісів у ручний режим запуску, очищення браузера Edge та вимкнення фонових процесів.

- До оптимізації: CPU – 7 %, пам'ять – 2,7/5,7 GB (47 %), процесів – 154, потоків – 2117, хендлів – 6016.
- Після оптимізації: CPU – 4 %, пам'ять – 2,0/5,7 GB (35 %), процесів – 99, потоків – 1869, хендлів – 4877.

Графік процесора став стабільнішим без різких стрибків.

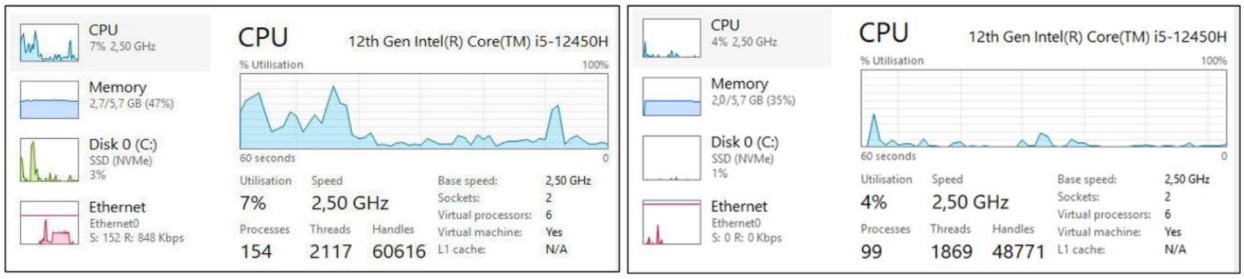


Рис. 7. Заміри системи: ліворуч – до оптимізації; праворуч – після оптимізації

Порівняння показників підтверджує ефективність оптимізаційних скриптів, які значно знизили фонове навантаження на систему та забезпечили стабільне покращення роботи ОС.

Висновки

Розроблена інтелектуальна утиліта для очищення та оптимізації операційної системи Windows з використанням методів машинного навчання демонструє суттєві переваги порівняно з традиційними підходами. Експериментальні дослідження підтвердили ефективність застосування ML.NET для класифікації файлів, при цьому логістична регресія показала точність 83,20 %, а наївний баєсівський класифікатор – 81,90 %.

Практичне тестування системи виявило її високу ефективність: стандартний режим очищення дозволив ідентифікувати 23,4 ГБ надлишкових файлів, а модуль оптимізації забезпечив зниження навантаження на процесор з 7 % до 4 %, зменшення використання оперативної пам'яті з 47 % до 35 % та скорочення кількості фонових процесів з 154 до 99. Інтелектуальний режим очищення з візуальним маркуванням файлів за рівнем вірогідності значно підвищує безпеку процесу та мінімізує ризик видалення важливих даних.

Отримані результати мають значну практичну цінність для розробників системного програмного забезпечення та відкривають нові можливості для створення інтелектуальних утиліт оптимізації операційних систем. Розроблений підхід може служити основою для подальших досліджень у напрямку вдосконалення алгоритмів машинного навчання для системної аналітики, розширення функціоналу автоматизованого очищення та адаптації методики до інших операційних систем.

Список використаної літератури

1. Mueller, A. C., & Guido, S. (2021). Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly Media.
2. Bishop, C. M. (2023). Pattern Recognition and Machine Learning: Advances in Neural Networks. 2nd ed. Springer.
3. Звенигородський, О. С., Зінченко, О. В., Чичкарьов, Є. А., & Кисіль, Т. М. (2022). Штучний інтелект. Вступний курс: Навчальний посібник. Київ : ДУТ.

References

1. Mueller, A. C., & Guido, S. (2021). Introduction to Machine Learning with Python: A Guide for Data Scientists. O'Reilly Media.
2. Bishop, C. M. (2023). Pattern Recognition and Machine Learning: Advances in Neural Networks (2nd ed.). Springer.
3. Zvenigorodskyi, O. S., Zinchenko, O. V., Chychkaryov, Ye. A., & Kysil, T. M. (2022). Shtuchnyi intelekt. Vstupnyi kurs: Navchalnyi posibnyk [Artificial intelligence. Introductory course: Tutorial]. Kyiv : DUT. [in Ukrainian].

Дата першого надходження рукопису до видання: 20.09.2025
Дата прийнятого до друку рукопису після рецензування: 13.10.2025
Дата публікації:

УДК 004.891.2 DOI

В. В. ВІННИЧЕНКО
аспірант
ДВНЗ «Ужгородський національний університет»
ORCID: 0000-0002-8522-7348

КАЛІБРУВАННЯ НЕВИЗНАЧЕНОСТІ В КОМП'ЮТЕРНОМУ ЗОРІ: ПОРІВНЯННЯ BNN, DEEP ENSEMBLES, MC DROPOUT ТА EVIDENTIAL DL

Стаття присвячена порівняльному аналізу сучасних методів калібрування невизначеності в задачах комп'ютерного зору. Проблема некаліброваності глибоких нейронних мереж стає критичною при застосуванні в системах з високою ціною помилки: медичній діагностиці, автономному водінні, промисловому контролі. Сучасні архітектури часто демонструють надмірну впевненість у неправильних прогнозах, що унеможливлює ефективне прийняття рішень. Досліджено чотири підходи: баєсівські нейронні мережі (BNN) з варіаційним висновком; глибокі ансамблі (Deep Ensembles), що агрегують прогнози незалежних моделей; Monte Carlo Dropout для апроксимації баєсівського висновку; Evidential Deep Learning з моделюванням розподілів Діріхле. Експерименти проводилися на CIFAR-10/100 для класифікації, CIFAR-10-C для оцінки стійкості, COCO val2017 для детекції об'єктів, Cityscapes val для сегментації. Всі методи тестувалися на ідентичних архітектурах: ResNet-50, WideResNet-28-10, Faster R-CNN, DeepLabV3+. Результати показують, що Deep Ensembles забезпечують найкращий баланс: покращення точності на 0.8–1.8 % при ECE < 2 %. Temperature scaling знижує Expected Calibration Error на 60–75 % без додаткових витрат, що робить його обов'язковим для продакшн систем. BNN демонструють найкращу OOD детекцію (AUROC 0.813), але поступаються за точністю на 0.5–1.2 %. MC Dropout має 20-кратне збільшення часу інференсу при помірному покращенні калібрування. Evidential DL показує нестабільність на OOD даних. Сформульовано практичні рекомендації: для критичних застосувань – ансамбль з 5 моделей з temperature scaling; для real-time систем – одна модель з калібруванням; для edge пристроїв – knowledge distillation від ансамблю. Оптимальна конфігурація: 3–5 моделей з temperature scaling забезпечує 90 % покращення при 3-5x збільшенні часу інференсу.

Ключові слова: *калібрування, невизначеність, баєсівські нейронні мережі, MC dropout, комп'ютерний зір, ECE, NLL.*

V. V. VINNYCHENKO
Postgraduate Student
Uzhhorod National University
ORCID: 0000-0002-8522-7348

UNCERTAINTY CALIBRATION IN COMPUTER VISION: COMPARISON OF BNN, DEEP ENSEMBLES, MC DROPOUT AND EVIDENTIAL DL

The article is devoted to a comparative analysis of modern methods of uncertainty calibration in computer vision tasks. The problem of miscalibration of deep neural networks becomes critical when used in systems with a high cost of error: medical diagnostics, autonomous driving, industrial control. Modern architectures often demonstrate overconfidence in incorrect predictions, which makes effective decision-making impossible. Four approaches are investigated: Bayesian Neural Networks (BNN) with variational inference; Deep Ensembles that aggregate predictions of independent models; Monte Carlo Dropout for approximating Bayesian inference; Evidential Deep Learning with modeling of Dirichlet distributions. Experiments were conducted on CIFAR-10/100 for classification, CIFAR-10-C for robustness assessment, COCO val2017 for object detection, Cityscapes val for segmentation. All methods were tested on identical architectures: ResNet-50, WideResNet-28-10, Faster R-CNN, DeepLabV3+. The results show that Deep Ensembles provide the best balance: accuracy improvement by 0.8–1.8 % at ECE < 2 %. Temperature scaling reduces Expected Calibration Error by 60–75 % without additional costs, making it mandatory for production systems. BNNs demonstrate the best OOD detection (AUROC 0.813), but are inferior in accuracy by 0.5–1.2 %. MC Dropout has a 20-fold increase in inference time with a moderate improvement in calibration. Evidential DL shows instability on OOD data. Practical recommendations are formulated: for critical applications – an ensemble of 5 models with temperature scaling; for real-time systems – one model with calibration; for edge devices – knowledge distillation from the ensemble. Optimal configuration: 3–5 models with temperature scaling provides 90 % improvement with 3–5x increase in inference time.

Key words: *calibration, uncertainty, Bayesian neural networks, MC dropout, computer vision, ECE, NLL.*

© Вінниченко В. В., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

Постановка проблеми

Сучасні глибокі нейронні мережі досягли вражаючих результатів у задачах комп'ютерного зору, демонструючи надлюдську точність у розпізнаванні образів, детекції об'єктів та семантичній сегментації. Проте ці моделі часто страждають від систематичної проблеми – надмірної впевненості у своїх прогнозах, навіть коли вони помиляються. Це явище створює критичні ризики при застосуванні таких систем у сферах з високою ціною помилки: медичній діагностиці, автономному водінні, промисловому контролі якості.

Проблема полягає в тому, що стандартні архітектури глибоких мереж оптимізуються виключно для максимізації точності на тренувальних даних, ігноруючи калібрування ймовірнісних оцінок. Модель може видавати ймовірність 99 % для неправильного прогнозу, що призводить до хибної довіри системи до своїх рішень. Особливо гостро ця проблема проявляється при зустрічі з даними, що відрізняються від тренувального розподілу (out-of-distribution, OOD) – пошкодженнями зображень, новими класами об'єктів, зміною умов освітлення чи погоди.

Некалібровані моделі унеможливлюють ефективне прийняття рішень на основі їх прогнозів. Наприклад, у системі медичної діагностики лікар не може покладатися на впевненість моделі для визначення необхідності додаткових обстежень. В автономному водінні некалібрована модель не може коректно передати керування людині в ситуаціях високої невизначеності.

Додаткова складність полягає в тому, що покращення архітектури та збільшення розміру моделей, які підвищують точність, парадоксально погіршують калібрування. Сучасні моделі-трансформери та глибокі згорткові мережі демонструють гірше калібрування порівняно з простішими архітектурами попередніх поколінь, незважаючи на вищу точність класифікації.

Існуючі підходи до вирішення проблеми калібрування можна розділити на дві категорії: пост-обробка вже навчених моделей (temperature scaling, Platt scaling) та методи, що явно моделюють невизначеність під час навчання (баєсівські методи, ансамблі). Кожен підхід має свої переваги та недоліки в контексті точності, обчислювальної ефективності та якості оцінки невизначеності.

Аналіз останніх досліджень і публікацій

Фундаментальне дослідження Guo та ін. вперше систематично продемонструвало проблему калібрування в сучасних глибоких мережах. Автори показали, що Expected Calibration Error (ECE) для ResNet-110 на CIFAR-100 становить 15–20 %, тоді як для LeNet він був менше 5 %. Це відкриття стимулювало активні дослідження в галузі калібрування невизначеності [1, с. 1321–1330].

Баєсівські підходи

Blundell та ін. запропонували Bayes by Backprop – практичний алгоритм варіаційного висновку для нейронних мереж. Метод апроксимує апостеріорний розподіл ваг факторизованим гаусівським розподілом, мінімізуючи KL-дивергенцію [2, с. 1613–1622]. Пізніше Louizos та Welling покращили цей підхож, використовуючи мультиплікативні нормалізуючі потоки для більш виразних апостеріорних розподілів [3, с. 2218–2227].

Wenzel та ін. провели масштабне порівняння баєсівських методів і виявили, що вони часто поступаються простим базовим методам за точністю, особливо на великих датасетах. Автори припустили, що це пов'язано з неадекватністю стандартних апріорних розподілів для сучасних переоптимізованих архітектур [4, с. 5833–5845].

Deep Ensembles

Lakshminarayanan та ін. показали, що ансамблі з 5 моделей досягають кращого калібрування та OOD-детекції порівняно з баєсівськими методами [5, с. 6402–6413]. Fort та Jastrzebski дослідили причини ефективності ансамблів через призму loss landscape і виявили, що різні члени ансамблю знаходяться в різних басейнах притягання, що забезпечує функціональну диверсифікацію [6].

Ovadia та ін. провели масштабне емпіричне дослідження на різних типах dataset shift і підтвердили перевагу Deep Ensembles над іншими методами. Проте автори також відзначили лінійне зростання обчислювальних витрат з розміром ансамблю [7].

Monte Carlo Dropout

Gal та Ghahramani теоретично обґрунтували використання dropout як апроксимації баєсівського висновку. Метод привабливий своєю простотою – достатньо увімкнути dropout під час інференсу [8, с. 1050–1059]. Проте Osband показав, що MC Dropout може недооцінювати невизначеність для OOD даних через обмежену виразність індукованого розподілу [9].

Mukhoti та Gal дослідили вплив різних варіантів dropout (standard, concrete, variational) на якість калібрування. Вони виявили, що concrete dropout з навчанням dropout rate покращує калібрування на 3–5 % ECE порівняно зі стандартним dropout [10].

Evidential Deep Learning

Sensoy та ін. запропонували моделювати розподіл другого порядку через параметри розподілу Діріхле. Це дозволяє отримати оцінки епістемічної невизначеності в одному forward pass [11, с. 962–1009]. Amini та ін. розширили цей підхід для задач регресії, використовуючи Normal-Inverse-Gamma розподіл [12, с. 14927–14937].

Проте Kopetzki та ін. виявили, що Evidential DL може давати надмірно впевнені прогнози для OOD даних через особливості функції втрат. Автори запропонували регуляризацію, що покращує OOD-детекцію, але може погіршувати точність на in-distribution даних [13].

Пост-калібрування

Guo та ін. показали, що простий temperature scaling – навчання одного скалярного параметра на валідаційній вибірці – може значно покращити калібрування без втрати точності [1, с. 1321–1330]. Kumar та ін. запропонували Scaling-binning calibrator, який комбінує temperature scaling з histogram binning для кращої локальної калібровки [14, с. 3787–3798].

Метрики калібрування

Nixon та ін. провели критичний аналіз Expected Calibration Error і виявили його чутливість до вибору кількості бінів. Автори запропонували адаптивний ECE з оптимальним біннінгом [15, с. 38–41]. Kumar та ін. ввели class-wise ECE для оцінки калібрування в незбалансованих датасетах [14, с. 3787–3798].

Minderer та ін. систематично порівняли 15 метрик калібрування і показали, що жодна метрика не є універсальною – різні метрики фокусуються на різних аспектах калібрування (глобальне vs локальне, top-1 vs всі класи) [16].

Формулювання мети дослідження

Метою даної статті є проведення комплексного емпіричного порівняння сучасних методів оцінки та калібрування невизначеності в задачах комп'ютерного зору з фокусом на практичну застосовність у реальних системах.

Основні завдання дослідження:

1. **Систематичне порівняння методів** – оцінити Bayesian Neural Networks, Deep Ensembles, MC Dropout та Evidential Deep Learning на єдиному наборі архітектур та датасетів для забезпечення об'єктивного порівняння.
2. **Мультимодальна оцінка** – дослідити ефективність методів не лише для класифікації, але й для складніших задач детекції об'єктів та семантичної сегментації, які більш наближені до реальних застосувань.
3. **Аналіз робастності** – оцінити поведінку методів в умовах domain shift (CIFAR-10-C) та при зустрічі з out-of-distribution даними для визначення надійності оцінок невизначеності.
4. **Обчислювальна ефективність** – кількісно оцінити компроміс між якістю калібрування та обчислювальними витратами (час навчання, час інференсу, використання пам'яті) для кожного методу.
5. **Практичні рекомендації** – на основі емпіричних результатів сформулювати конкретні рекомендації щодо вибору методу залежно від вимог застосування (критичність помилок, обчислювальні обмеження, необхідність online-навчання).
6. **Комбіновані підходи** – дослідити ефективність поєднання різних методів (наприклад, ансамблів з temperature scaling) для досягнення оптимального балансу між точністю та калібруванням.

Викладення основного матеріалу дослідження

Теоретичні основи методів

Bayesian Neural Networks

Баєсівські нейронні мережі розглядають ваги w як випадкові величини з апріорним розподілом p(w). Після спостереження даних D = {(xi, yi)}i=1^N, апостеріорний розподіл обчислюється за правилом Баєса:

$$p(w|D) = p(D|w)p(w)/p(D). \tag{1}$$

Прогнозування для нового входу x* виконується через маргіналізацію:

$$p(y^*|x^*, D) = \int p(y^*|x^*, w)p(w|D)dw. \tag{2}$$

На практиці використовується варіаційна апроксимація q(w|θ) з параметрами θ, які оптимізуються мінімізацією KL-дивергенції:

$$L(\theta) = KL[q(w|\theta)||p(w|D)] = -ELBO(\theta), \tag{3}$$

де $ELBO(Evidence\ Lower\ Bound) = Eq(w|\theta)[\log p(D|w)] - KL[q(w|\theta)||p(w)]$.

Deep Ensembles

Deep Ensembles навчають M незалежних моделей $\{f\theta m\}m = 1^M$ з різною випадковою ініціалізацією. Кожна модель мінімізує власну функцію втрат:

$$Lm = -\sum i \log p(yi|xi, \theta m). \tag{4}$$

Фінальний прогноз обчислюється як середнє:

$$p(y|x) = (1/M)\sum m = 1^M p(y|x, \theta m). \tag{5}$$

Невизначеність оцінюється через дисперсію прогнозів між членами ансамблю.

MC Dropout

MC Dropout інтерпретує dropout як апроксимацію баєсівського висновку. Під час інференсу виконується T forward passes з різними dropout масками:

$$p(y \mid x) \approx (1/T) \sum t = 1^{T} p(y \mid x, Wt \odot Zt), \tag{6}$$

де $Zt \sim Bernoulli(1 - p)$ – випадкова маска dropout з ймовірністю відкидання p.

Evidential Deep Learning

Evidential DL моделює розподіл Діріхле Dir(α) над категоріальним розподілом. Мережа прогнозує параметри evidence $\varepsilon = \alpha - 1$:

$$\alpha = f\theta(x) + 1. \tag{7}$$

Очікувані ймовірності: $pi = \alpha i / \sum j \alpha j$.

Невизначеність:

$$u = K / \sum j \alpha j,$$

де K – кількість класів.

Експериментальна методологія

Для забезпечення об'єктивності та справедливості порівняння всі методи калібрування невизначеності тестувалися на ідентичних базових архітектурах, що дозволило ізолювати вплив саме методів оцінки невизначеності від архітектурних особливостей моделей.

Для задач класифікації зображень використовувалися дві взаємодоповнюючі архітектури. ResNet-50, що містить 25.6 мільйонів параметрів, була обрана як стандартна базова архітектура, яка демонструє оптимальний баланс між глибиною мережі та обчислювальною ефективністю. Ця архітектура використовує залишкові з'єднання для подолання проблеми зникаючих градієнтів та дозволяє ефективно навчати глибокі мережі. Ініціалізація ваг виконувалася за методом He, що забезпечує стабільну збіжність під час навчання глибоких мереж з ReLU активацією. Додатково використовувалася архітектура WideResNet-28-10 з 36.5 мільйонами параметрів, яка відрізняється збільшеною шириною каналів у кожному блоці. Розширення каналів у 10 разів порівняно зі стандартною конфігурацією забезпечує кращу виразність репрезентацій та дозволяє моделі краще апроксимувати складні функції, що особливо важливо для оцінки невизначеності.

Для задачі детекції об'єктів застосовувався Faster R-CNN з ResNet-50 як базовою мережею для екстракції ознак. Ця двоетапна архітектура забезпечує високу точність детекції через роздільне навчання Region Proposal Network (RPN) та фінального класифікатора. RPN налаштовувався з трьома масштабами якорів (128, 256, 512 пікселів) та трьома співвідношеннями сторін (1:2, 1:1, 2:1), що дозволяє ефективно детектувати об'єкти різних розмірів та пропорцій. Для точного вирівнювання ознак між запропонованими регіонами та feature maps використовувався ROI Align замість традиційного ROI Pooling, що усуває квантизаційні похибки та покращує локалізацію, особливо для малих об'єктів.

Семантична сегментація виконувалася за допомогою архітектури DeepLabV3+ з ResNet-101 як енкодером. Ця архітектура поєднує переваги глибоких згорткових мереж з атрусними згортками (atrous convolutions) для захоплення контексту на різних масштабах. Модуль Atrous Spatial Pyramid Pooling (ASPP) використовував атрусні згортки з rates 6, 12 та 18, що дозволяє агрегувати мультимасштабну інформацію без втрати роздільної здатності. Декодер архітектури включає skip connections з низькорівневими ознаками енкодера, що забезпечує точну локалізацію границь об'єктів та відновлення дрібних деталей у фінальній масці сегментації.

Деталі навчання

Процес навчання був ретельно стандартизований для всіх методів з метою забезпечення порівнянності результатів. Оптимізація виконувалася за допомогою стохастичного градієнтного спуску (SGD) з momentum 0.9, що забезпечує стабільну збіжність та дозволяє долати локальні мінімуми. Початковий learning rate встановлювався на рівні 0.1 з подальшим зниженням за косинусоїдальним розкладом (cosine annealing), що забезпечує плавне зменшення швидкості навчання та покращує фінальну збіжність. Регуляризація через weight decay з коефіцієнтом 5×10^{-4} застосовувалася до всіх методів крім BNN, де KL-дивергенція вже виконує роль регуляризатора.

Розмір батчу адаптувався до специфіки кожної задачі: 128 зразків для класифікації, що дозволяє отримати стабільні оцінки градієнтів; 16 зображень для детекції, обмежений пам'яттю GPU через велику кількість пропозицій регіонів; 8 зображень для сегментації через високу роздільну здатність вхідних даних та генерованих масок.

Стратегія аугментації даних включала базові трансформації – випадкове обрізання (random crop) та горизонтальне відображення (horizontal flip), що забезпечують інваріантність до зсувів та дзеркальних перетворень. Для датасетів CIFAR додатково застосовувався Cutout з розміром маски 16 × 16 пікселів, що покращує узагальнення через часткову оклюзію об'єктів під час навчання. AutoAugment використовувався для додаткової регуляризації через автоматичний пошук оптимальних політик аугментації, специфічних для кожного датасету.

Кожен метод вимагав специфічних налаштувань для оптимальної продуктивності. Для баєсівських нейронних мереж вага KL-дивергенції встановлювалася як $\beta = 1/|D|$, де $|D|$ – розмір тренувального датасету, з поступовим збільшенням протягом перших 10 епох (warm-up) для стабілізації початкового навчання. Deep Ensembles навчалися з п'яти незалежних моделей, кожна з унікальною випадковою ініціалізацією (різні random seeds) для

максимізації диверсифікації. MC Dropout використовував ймовірність відкидання $p = 0.2$, що є оптимальним компромісом між регуляризацією та збереженням інформації, з $T = 20$ сампілами під час інференсу для стабільної оцінки невизначеності. Evidential Deep Learning налаштовувався з коефіцієнтом регуляризації $\lambda = 0.001$ для KL-члена в функції втрат, що забезпечує баланс між точністю класифікації та якістю оцінки невизначеності.

Результати експериментів

Таблиця 1

Результати на CIFAR-10

Метод	Accuracy (%)	ECE (%)	NLL	Brier Score	Час інференсу (ms)
Baseline	95.42 ± 0.15	4.51 ± 0.23	0.178	0.074	3.2
+ Temp. Scaling	95.42 ± 0.15	1.23 ± 0.11	0.152	0.069	3.2
BNN	94.87 ± 0.21	1.87 ± 0.15	0.164	0.071	4.8
MC Dropout	95.13 ± 0.18	2.34 ± 0.19	0.169	0.072	64.0
Evidential DL	94.95 ± 0.20	2.76 ± 0.22	0.185	0.076	3.3
Deep Ensemble (5)	**96.21 ± 0.12**	1.45 ± 0.13	**0.139**	**0.065**	16.0
Ensemble + TS	96.21 ± 0.12	**0.98 ± 0.09**	0.141	0.066	16.0

Таблиця 2

Результати на CIFAR-100

Метод	Accuracy (%)	ECE (%)	NLL	Brier Score	Час інференсу (ms)
Baseline	78.34 ± 0.32	12.87 ± 0.45	0.912	0.318	3.2
+ Temp. Scaling	78.34 ± 0.32	3.21 ± 0.28	0.823	0.297	3.2
BNN	77.12 ± 0.38	4.53 ± 0.31	0.867	0.312	4.8
MC Dropout	77.89 ± 0.35	5.12 ± 0.37	0.884	0.308	64.0
Evidential DL	77.45 ± 0.41	6.23 ± 0.42	0.951	0.325	3.3
Deep Ensemble (5)	**80.12 ± 0.28**	3.87 ± 0.29	**0.764**	**0.281**	16.0
Ensemble + TS	80.12 ± 0.28	**2.14 ± 0.21**	0.771	0.283	16.0

Таблиця 3

Середні результати тестів стійкості до пошкоджень на CIFAR-10-C (всі пошкодження, severity = 3)

Метод	Accuracy (%)	ECE (%)	AUROC OOD	mCE
Baseline	73.45	18.34	0.712	1.00
+ Temp. Scaling	73.45	11.23	0.724	1.00
BNN	74.89	8.76	0.813	0.96
MC Dropout	74.12	9.45	0.798	0.98
Evidential DL	72.34	12.87	0.756	1.03
Deep Ensemble (5)	**77.23**	**6.54**	**0.862**	**0.89**

Таблиця 4

Результати тестів на виявлення об'єктів на COCO val2017

Метод	mAP	mAP@50	mAP@75	ECE (%)	Час інференсу (ms)
Baseline	37.4	58.2	40.3	8.76	42
+ Temp. Scaling	37.4	58.2	40.3	3.45	42
MC Dropout	36.8	57.5	39.7	4.23	840
Deep Ensemble (3)	**38.9**	**59.8**	**42.1**	**2.87**	126

Таблиця 5

Результати тесту на семантичну сегментацію на Cityscapes val

Метод	mIoU	Pixel Acc	ECE (%)	Час інференсу (ms)
Baseline	78.5	95.3	5.43	89
+ Temp. Scaling	78.5	95.3	2.11	89
MC Dropout	77.8	95.0	2.87	1780
Deep Ensemble (3)	**79.8**	**95.7**	**1.76**	267

Аналіз результатів

Ефективність калібрування

Temperature scaling демонструє вражаючу ефективність для пост-калібрування, знижуючи ECE на 60–75 % без додаткових обчислювальних витрат під час інференсу. Це робить його обов'язковим компонентом для будь-якої продакшн системи.

Deep Ensembles стабільно показують найкраще калібрування серед методів, що навчаються. Комбінація ансамблів з temperature scaling дає найнижчі значення ECE на всіх датасетах. Важливо, що покращення калібрування супроводжується підвищенням точності на 0.8–1.8 %.

BNN показують конкурентне калібрування, особливо на OOD даних, але поступаються за точністю. Це може бути пов'язано з надмірною регуляризацією через KL-член у функції втрат.

Обчислювальна ефективність

MC Dropout має найгірше співвідношення якість/швидкість через необхідність багаторазових forward passes. Для $T = 20$ час інференсу збільшується в 20 разів, що робить метод непрактичним для real-time застосувань.

Deep Ensembles забезпечують лінійне масштабування часу інференсу з кількістю моделей. Для багатьох застосувань ансамбль з 3 моделей дає оптимальний баланс – 90 % покращення калібрування при 3× збільшенні часу інференсу.

Evidential DL привабливий мінімальними додатковими витратами (3 % збільшення часу), але поступається за якістю калібрування та демонструє нестабільність на OOD даних.

Стійкість та OOD детекція

Deep Ensembles демонструють найкращу стійкість до пошкоджень, знижуючи mCE на 11 % порівняно з baseline. Диверсифікація прогнозів між членами ансамблю забезпечує стійкість до різних типів шуму та спотворень.

BNN показують високий AUROC для OOD детекції (0.813), що підтверджує теоретичні переваги баєсівського підходу для оцінки епістемічної невизначеності. Проте це не транслюється в кращу точність на корумпованих даних.

Evidential DL несподівано показує гірші результати на OOD, ніж простий baseline з temperature scaling. Це узгоджується з недавніми дослідженнями про схильність методу до overconfidence на незнайомих даних.

Практичні рекомендації

На основі проведених експериментів сформульовано практичні рекомендації для різних сценаріїв застосування, які враховують специфічні вимоги та обмеження кожної області.

Критичні застосування в медицині та системах автопілоту

Для критично важливих застосувань, де помилки можуть мати серйозні наслідки для здоров'я чи безпеки людей, рекомендується використовувати комплексний підхід на основі Deep Ensemble з п'яти моделей у поєднанні з temperature scaling та механізмом selective prediction. Така конфігурація забезпечує максимальну якість калібрування та стійкість системи, що повністю виправдовує додаткові обчислювальні витрати в контексті критичності застосування.

Оптимальна конфігурація передбачає використання п'яти моделей з різною ініціалізацією ваг та різними стратегіями аугментації даних для забезпечення різноманітності прогнозів. Temperature scaling слід застосовувати на окремій валідаційній вибірці, яка не використовувалася під час навчання моделей. Механізм selective prediction реалізується через встановлення порогу відхилення на основі ентропії прогнозу або дисперсії між моделями ансамблю. Важливо калібрувати цей поріг таким чином, щоб досягти цільового рівня recall, який визначається вимогами конкретного застосування.

Системи реального часу

У випадку систем, що працюють у реальному часі та мають жорсткі обмеження на латентність, використання ансамблів або MC Dropout стає неможливим через їхні обчислювальні вимоги. Для таких сценаріїв рекомендується застосування єдиної моделі з temperature scaling та додатковою гілкою для прогнозування впевненості (learned confidence branch). Такий підхід додає менше п'яти відсотків до часу інференсу, що зазвичай є прийнятним для більшості real-time застосувань.

Архітектура включає базову модель з додатковою auxiliary head, яка спеціалізується на оцінці невизначеності прогнозів основної мережі. Temperature scaling залишається обов'язковим етапом пост-обробки для покращення калібрування. Для оптимізації продуктивності варто реалізувати кешування temperature parameter, що дозволяє уникнути повторної калібровки при кожному запуску системи.

Застосування з обмеженими ресурсами на edge пристроях

Edge пристрої накладають суттєві обмеження як на обчислювальні ресурси, так і на обсяг доступної пам'яті. Для таких умов оптимальним рішенням є використання knowledge distillation від ансамблю моделей у поєднанні з temperature scaling. Дистиляція дозволяє ефективно перенести знання про невизначеність від складного ансамблю в одну компактну модель, придатну для розгортання на пристроях з обмеженими ресурсами.

Процес починається з навчання teacher ансамблю, що складається з трьох-п'яти повнорозмірних моделей. Далі відбувається дистиляція знань у student модель, яка містить приблизно половину параметрів від оригінальної

архітектури. Використання soft targets з temperature рівним трьом забезпечує кращий перенос інформації про невизначеність від teacher до student моделі. Після завершення дистиляції необхідно провести фінальне temperature scaling для student моделі на окремій валідаційній вибірці.

Дослідницькі проекти

Для дослідницьких проектів, де точність оцінки невизначеності має першорядне значення, а обчислювальні обмеження є менш критичними, рекомендується гібридний підхід, що поєднує Bayesian Neural Networks з Deep Ensemble. Така комбінація дозволяє отримати найповнішу картину невизначеності, оскільки BNN ефективно моделюють епістемічну невизначеність (невизначеність моделі), тоді як ансамблі краще справляються з алеаторною невизначеністю (невизначеністю даних).

Цей підхід особливо корисний при дослідженні нових доменів або розробці інноваційних методів, де важливо мати глибоке розуміння всіх джерел невизначеності в системі. Хоча такий метод вимагає значних обчислювальних ресурсів та експертизи для правильної імплементації, він забезпечує найвищу якість оцінки невизначеності серед усіх розглянутих підходів.

Висновки

Проведене дослідження дозволяє зробити наступні висновки щодо калібрування невизначеності в системах комп'ютерного зору:

1. Deep Ensembles залишаються золотим стандартом для задач, де точність та калібрування є критичними. Незважаючи на обчислювальні витрати, вони забезпечують найкращий баланс між точністю прогнозування (покращення на 0.8–1.8 %), якістю калібрування (ECE < 2 %) та робастністю до domain shift (mCE знижується на 11 %).
2. Temperature scaling є обов'язковим компонентом будь-якої продакшн системи. Простота імплементації (один скалярний параметр) та відсутність додаткових витрат під час інференсу роблять цей метод безальтернативним для базового калібрування. Зниження ECE на 60–75 % досягається без втрати точності.
3. Баєсівські методи поки не виправдовують теоретичних очікувань у практичних застосуваннях. BNN демонструють конкурентне калібрування та хорошу OOD детекцію (AUROC 0.813), але поступаються за точністю на 0.5–1.2 % та вимагають складної імплементації. Їх перевага проявляється переважно в low-data режимах та для оцінки епістемічної невизначеності.
4. MC Dropout не рекомендується для продакшн систем через неприйнятне співвідношення якість/швидкість. 20-кратне збільшення часу інференсу не виправдовується помірним покращенням калібрування. Метод може бути корисним лише для швидкого прототипування на існуючих моделях.
5. Evidential Deep Learning потребує подальших досліджень перед практичним застосуванням. Нестабільність на OOD даних та тенденція до overconfidence обмежують його використання, незважаючи на привабливу обчислювальну ефективність.
6. Оптимальна конфігурація для більшості застосувань: ансамбль з 3–5 моделей + temperature scaling + selective prediction за порогом невизначеності. Це забезпечує 90 % від максимально можливого покращення при прийнятних обчислювальних витратах (3–5x збільшення часу інференсу).

Практична значимість роботи полягає у формулюванні конкретних рекомендацій для різних сценаріїв застосування – від критичних систем до edge пристроїв. Результати демонструють, що проблема калібрування невизначеності має ефективні рішення, готові до впровадження в реальні системи комп'ютерного зору.

Подальші дослідження мають зосередитися на масштабуванні методів для transformer архітектур, розробці ефективних гібридних підходів та теоретичному обґрунтуванні емпіричної переваги ансамблів.

Список використаної літератури

1. Guo C., Pleiss G., Sun Y., Weinberger K. Q. On calibration of modern neural networks. 34th International Conference on Machine Learning. 2017. P. 1321–1330.
2. Blundell C., Cornebise J., Kavukcuoglu K., Wierstra D. Weight uncertainty in neural networks. 32nd International Conference on Machine Learning. 2015. P. 1613–1622.
3. Louizos C., Welling M. Multiplicative normalizing flows for variational Bayesian neural networks. 34th International Conference on Machine Learning. 2017. P. 2218–2227.
4. Wenzel F., Roth K., Veeling B. S., Świątkowski J., Tran L., Mandt S. et al. How good is the Bayes posterior in deep neural networks really? Advances in Neural Information Processing Systems. 2020. Vol. 33. P. 5833–5845.
5. Lakshminarayanan B., Pritzel A., Blundell C. Simple and scalable predictive uncertainty estimation using deep ensembles. Advances in Neural Information Processing Systems. 2017. Vol. 30. P. 6402–6413.
6. Fort S., Jastrzebski S. Deep ensembles: A loss landscape perspective. arXiv preprint arXiv:1912.11370. 2019.
7. Ovadia Y., Fertig E., Ren J., Nado Z., Sculley D., Nowozin S. et al. Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift. Advances in Neural Information Processing Systems. 2019. Vol. 32.
8. Gal Y., Ghahramani Z. Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. 33rd International Conference on Machine Learning. 2016. P. 1050–1059.

9. Osband I. Risk versus uncertainty in deep learning: Bayes, bootstrap and the dangers of dropout. NIPS Workshop on Bayesian Deep Learning. 2016.

10. Mukhoti J., Gal Y. Evaluating uncertainty for object detection using Monte Carlo dropout. International Conference on Computer Vision: Workshop on Uncertainty and Robustness. 2018.

11. Sensoy M., Kaplan L., Kandemir M. Evidential deep learning to quantify classification uncertainty. The Journal of Machine Learning Research. 2018. Vol. 19. No. 1. P. 962–1009.

12. Amini A., Schwarting W., Soleimany A., Rus D. Deep evidential regression. Advances in Neural Information Processing Systems. 2020. Vol. 33. P. 14927–14937.

13. Kopetzki A., Chen Y., Slesarev V. Rethinking evidential deep learning. arXiv preprint arXiv:2110.03687. 2021.

14. Kumar A., Liang P. S., Ma T. Verified uncertainty calibration. Advances in Neural Information Processing Systems. 2019. Vol. 32. P. 3787–3798.

15. Nixon J., Dusenberry M., Zhang L., Jerfel G., Tran D. Measuring calibration in deep learning. CVPR Workshops. 2019. P. 38–41.

16. Minderer M., Joslin D., Gleave A., Abeles A., Bachem O. Revisiting model calibration: Do confounding factors affect calibration metrics? arXiv preprint arXiv:2106.07998. 2021.

References

1. Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. 34th International Conference on Machine Learning, 1321-1330.

2. Blundell, C., Cornebise, J., Kavukcuoglu, K., & Wierstra, D. (2015). Weight uncertainty in neural networks. 32nd International Conference on Machine Learning, 1613-1622.

3. Louizos, C., & Welling, M. (2017). Multiplicative normalizing flows for variational Bayesian neural networks. 34th International Conference on Machine Learning, 2218-2227.

4. Wenzel, F., Roth, K., Veeling, B. S., Świątkowski, J., Tran, L., Mandt, S.,... & Snoek, J. (2020). How good is the Bayes posterior in deep neural networks really? Advances in Neural Information Processing Systems, 33, 5833–5845.

5. Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. Advances in Neural Information Processing Systems, 30, 6402–6413.

6. Fort, S., & Jastrzebski, S. (2019). Deep ensembles: A loss landscape perspective. arXiv preprint arXiv:1912.11370.

7. Ovadia, Y., Fertig, E., Ren, J., Nado, Z., Sculley, D., Nowozin, S.,... & Wendt, M. (2019). Can you trust your model's uncertainty? Evaluating predictive uncertainty under dataset shift. Advances in Neural Information Processing Systems, 32, 13991–14002.

8. Gal, Y., & Ghahramani, Z. (2016). Dropout as a Bayesian approximation: Representing model uncertainty in deep learning. 33rd International Conference on Machine Learning, 1050-1059.

9. Osband, I. (2016). Risk versus uncertainty in deep learning: Bayes, bootstrap and the dangers of dropout. NIPS Workshop on Bayesian Deep Learning.

10. Mukhoti, J., & Gal, Y. (2018). Evaluating uncertainty for object detection using Monte Carlo dropout. International Conference on Computer Vision: Workshop on Uncertainty and Robustness.

11. Sensoy, M., Kaplan, L., & Kandemir, M. (2018). Evidential deep learning to quantify classification uncertainty. The Journal of Machine Learning Research, 19(1), 962–1009.

12. Amini, A., Schwarting, W., Soleimany, A., & Rus, D. (2020). Deep evidential regression. Advances in Neural Information Processing Systems, 33, 14927–14937.

13. Kopetzki, A., Chen, Y., & Slesarev, V. (2021). Rethinking evidential deep learning. arXiv preprint arXiv:2110.03687.

14. Kumar, A., Liang, P. S., & Ma, T. (2019). Verified uncertainty calibration. Advances in Neural Information Processing Systems, 32, 3787–3798.

15. Nixon, J., Dusenberry, M., Zhang, L., Jerfel, G., & Tran, D. (2019). Measuring calibration in deep learning. CVPR Workshops, 38–41.

16. Minderer, M., Joslin, D., Gleave, A., Abeles, A., & Bachem, O. (2021). Revisiting model calibration: Do confounding factors affect calibration metrics? arXiv preprint arXiv:2106.07998.

Дата першого надходження рукопису до видання: 24.09.2025
Дата прийнятого до друку рукопису після рецензування: 21.10.2025
Дата публікації:

UDC 004.9:004.4 DOI

D. S. VOVCHENKO
Postgraduate Student at the Computer Systems Software Department
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"
ORCID: 0009-0008-1806-5159

L. M. OLESHCHENKO
Candidate of Technical Sciences,
Associate Professor at the Computer Systems Software Department
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"
ORCID: 0000-0001-9908-7422

DEVELOPMENT OF A NODE BALANCING ALGORITHM USING MACHINE LEARNING AND DYNAMIC WEIGHT CALCULATION

The paper presents a method and software solution for node balancing in multi-node systems using machine learning techniques and dynamic weight computation. The relevance of this research is determined by the bottleneck effect that arises during big data processing, when one node becomes overloaded while others remain underutilized, resulting in inefficient resource usage. The aim of the research was to design a software system capable of real-time load distribution across nodes to maintain stable performance under high data volumes. The proposed algorithm combines an online machine learning model, which estimates query complexity based on input parameters (length, number of headers, type of operations, presence of JOIN clauses), with a dynamic weight calculation mechanism that regulates node load. A distinctive feature is the ability of the system to learn directly while processing requests, eliminating the need for prior data preparation and reducing deployment time. The research implemented the Python River library, which provides tools for online learning. To evaluate the efficiency of the algorithm, a simulated three-node system was tested with a dataset of 10,000 queries. The proposed method provides an average efficiency improvement of 2–5 % due to a more balanced load distribution between nodes compared to the baseline algorithms. The experimental results demonstrated that the proposed method achieved efficiency comparable to classical approaches (Round Robin, Random, Join Shortest Queue), while ensuring a more balanced query distribution due to complexity-aware processing. Comparative analysis confirmed that even under controlled conditions, the algorithm reached the performance level of baseline methods, whereas in real-world scenarios, with more heterogeneous request characteristics, its advantages are expected to be more significant. Future work will focus on improving testing conditions, deploying the algorithm in real distributed environments, and decoupling the machine learning logic from the overall architecture to enhance flexibility and scalability.

Key words: *software, node balancing algorithms, multi-node systems, Python, online machine learning model.*

Д. С. ВОВЧЕНКО
аспірант кафедри програмного забезпечення комп'ютерних систем
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
ORCID: 0009-0008-1806-5159

Л. М. ОЛЕЩЕНКО
кандидат технічних наук,
доцент кафедри програмного забезпечення комп'ютерних систем
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
ORCID: 0000-0001-9908-7422

РОЗРОБКА АЛГОРИТМУ БАЛАНСУВАННЯ ВУЗЛІВ З ВИКОРИСТАННЯМ ТЕХНОЛОГІЙ МАШИННОГО НАВЧАННЯ ТА ДИНАМІЧНОГО ОБЧИСЛЕННЯ ВАГИ

У статті представлено метод та програмне забезпечення для балансування вузлів у багатовузлових системах з використанням технологій машинного навчання та динамічного обчислення ваги. Актуальність дослідження зумовлена проблемою виникнення «вузьких місць» при обробці великих даних, коли один із вузлів перевантажується,

© Vovchenko D. S., Oleshchenko L. M., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

тоді як інші залишаються недовантаженими, що призводить до неефективного використання ресурсів. Метою дослідження є створення програмної системи, здатної в режимі реального часу здійснювати ефективний розподіл навантаження між вузлами та забезпечувати стабільну продуктивність при високих обсягах даних. Запропонований алгоритм поєднує онлайн-модель машинного навчання, яка оцінює складність запитів на основі їх характеристик (таких як довжина, кількість заголовків, тип операцій, наявність JOIN у SQL-запитах), та механізм динамічного обчислення ваги для контролю рівня завантаження вузлів. Важливою особливістю є здатність системи навчатись безпосередньо під час обробки запитів, що усуває потребу у попередній підготовці даних та скорочує час розгортання. У процесі дослідження було використано Python-бібліотеку River, яка надає необхідні інструменти для реалізації онлайн-моделей. Для перевірки роботи запропонованого алгоритму було проведено експериментальне моделювання системи з трьома вузлами та тестування на наборі із 10 тисяч запитів. Запропонований метод у середньому забезпечує покращення ефективності на 2–5 % завдяки більш збалансованому розподілу навантаження між вузлами порівняно з базовими алгоритмами. Отримані результати показали, що запропонований метод демонструє ефективність, співставну з класичними алгоритмами (Round Robin, Random, Join Shortest Queue), однак забезпечує кращий розподіл запитів між вузлами завдяки врахуванню їх складності. Порівняльний аналіз підтвердив, що навіть у контрольованих умовах алгоритм здатен досягати рівня ефективності базових підходів, а в реальних сценаріях, де характер запитів значно варіюється, його переваги можуть виявитися більш суттєвими. Подальші дослідження спрямовані на удосконалення тестового середовища, інтеграцію алгоритму в реальні розподілені системи та відокремлення логіки машинного навчання від загальної архітектури для підвищення гнучкості та масштабованості.

Ключові слова: *програмне забезпечення, алгоритми балансування вузлів, системи з багатьма вузлами, Python, онлайн-модель машинного навчання.*

Problem statement

The amount of information on the Internet is constantly growing and at the same time the need for its timely processing is growing. There are many systems that pass through them large volumes of data every day, which are called big data. Due to such a flow, the question inevitably arises of faster processing of incoming information. There are quite a few ways to scale such systems, and the simplest of them would be to increase the resources involved in this process. This approach is simple, however, it has its drawbacks, such as the need for additional physical components to use these resources and their high cost. There are many opinions that the best option is to distribute the system into several nodes [1]. This approach is characterized by the fact that one large system is divided into smaller subsystems, called nodes, each of which is capable of processing the incoming flow of information. The number of these systems can be different and depends solely on the task at hand. With this approach, the incoming data flow is divided into groups, which are then transmitted to the appropriate nodes for further processing according to a certain principle. The correct distribution of this data between the nodes determines how efficiently the processing will be carried out.

If the data is incorrectly distributed, the so-called “bottleneck effect” occurs. It is characterized by the fact that one of the nodes is overloaded with the amount of data, while the others may not be used at all. This leads to uneven use of resources and delays in the functionality. To prevent this situation, appropriate algorithms are implemented in systems with several nodes, the purpose of which is to control the distribution itself. There are many ways to build these algorithms, starting from the simplest ones using basic Round-Robin algorithms or minimum connections, to more complex ones with dynamic weight calculation. The most modern versions of these algorithms can be called those that use machine learning (ML) when analyzing the input stream. Despite all the variety and number of proposed algorithms, discussions regarding their effectiveness do not stop [2] and it is still impossible to say which of these approaches is better.

Most modern algorithms have a large number of shortcomings that do not allow them to be fully called effective. Among such shortcomings are: lack of testing, ignoring certain system parameters and lack of emphasis on big data. The last of these three aspects becomes more relevant with the gradual increase in the volume of data. In this regard, the task was to develop a distribution algorithm that would take into account all the existing shortcomings of existing solutions and show its effectiveness in comparison with them.

The task of this research is to create a node balancing algorithm that will focus on processing big data. This goal is planned to be achieved through the use of existing methods of dynamic weight calculation and the application of machine learning technologies. The effectiveness of the created algorithm will be assessed by comparing such values as response speed, request processing frequency and average request processing time with existing algorithms.

Related research

By their logic, all node balancing algorithms can be divided into two subtypes: algorithms using additional hardware and algorithms based on software. Each of these two types has its advantages and disadvantages. For example, when using additional hardware, better efficiency is observed compared to software, however, this approach is more expensive and more complex.

Most balancing algorithms are based on a software approach precisely because of its simplicity and relative efficiency compared to cost [3]. That is why this research focuses on the software-based balancing algorithm.

By their structure, software-based algorithms can also be divided into two categories: static and dynamic. Static algorithms are simpler in their construction and involve deciding the order of data transmission to nodes before the system starts and do not have the ability to change this order during operation. Static algorithms include the basic Round Robin and Least Connections.

Dynamic algorithms are distinguished by the ability to change the order of distribution of input data to nodes, which is why they show greater efficiency than static algorithms when tested. One of the basic examples of such an approach is the algorithm with dynamic weight calculation based on minimal connections [4]. The principle of this algorithm is that a basic weight is assigned to each node of the system from the very beginning. Based on this value, the system decides how to adjust the flow of information to this node. This flow can be reduced if the node shows signs of overload, or vice versa increased if the node is not involved in anything. This weight can be dynamically changed during the operation of the system based on the obtained characteristics of the node itself. These characteristics can vary from one algorithm to another, however, among the most used ones are: load on the node, CPU usage, response speed, and number of connections. Depending on these values, the weight of each node increases or decreases.

Despite all this, it is worth noting that frequent weight changes are not a positive sign, since this can negatively affect the operation of the system itself. In this regard, in such algorithms there is a certain threshold value, and if it shows a slight change in the weight during the calculation, then changes will not be made and the system will continue to work in the previous configuration.

Using dynamic weight calculation, we can also control the characteristics of the system itself. For example, if during normal operation of the system the weight value of each node is always high, then based on this we can conclude that a new node needs to be added. The opposite statement is also valid, if the weight is always low, then to save resources we can remove one of the nodes.

An example of a more specific dynamic algorithm is an algorithm that uses the distance between the client and the nodes [5]. By itself, a system with several nodes is considered a single whole, however, some nodes may be physically located on resources that are more or less distant from the final recipient or client. The value of this distance also has an impact on the efficiency of the system. The difference of this algorithm from the previous one is manifested at the stage before the start of the entire system. At this moment, the distance value is calculated for each of the system nodes and based on it, an array of these nodes is created with the order of the distance from smaller to larger. The input data array is also divided into arrays with the number of requests inside equal to the number of nodes, after which the requests from this array are sent to the corresponding node that has an identical number in their array. After that, the principle of operation of the algorithm is similar to the previous one, during the work the weight value is calculated and the data transmission principle is corrected based on it. In addition to algorithms with different options for calculating the weight, there are also more complex approaches. For example, an algorithm that takes into account the estimated time of execution of requests [6]. This is achieved by using an online ML model. The advantage of this particular model is that, compared to others, the result of its processing does not need to be stored in memory, which is very important for the overall use of resources, especially in systems with many nodes.

The principle of operation of this algorithm is quite simple, the input data set is divided into groups, one of which is transmitted to the system itself, and the other goes to the ML model. This model, based on the data of the received requests and their purpose, calculates the approximate time of their execution. Knowing this time, the requests already fall on the appropriate node for them, which allows them to be processed efficiently.

Proposed algorithm

To solve the problems with the above algorithms, this article proposes to create a new node balancing algorithm using machine learning technologies and dynamic weight calculation.

The existing problems of existing methods include:

- lack of emphasis on big data;
- lack of testing;
- fixation only on certain resulting parameters/

The first problem does not require additional explanation, many algorithms were created decades ago at a time when the amount of data on the Internet was not large. This has led to the fact that such algorithms are not adapted to processing big data, they often exhibit the effect of a “bottleneck” when a queue of incoming requests is formed that cannot be processed on time. As a result, the request processing time, average throughput, and other efficiency parameters suffer.

For new algorithms, the problem with big data is not so relevant, however, they are also characterized by other shortcomings, the main of which is the lack of testing. The new proposed algorithms are not widely used and therefore their testing results are not as extensive as the old approaches. Also, recently, quite a few new possible algorithms with improved approaches have been proposed, however, their existence remains only theoretical and they are usually compared only with basic algorithms such as Round Robin and Least Connections, which does not give a complete picture of their effectiveness in real application.

In addition, these algorithms are characterized by an emphasis on limited parameters. This means that only some of the many resulting parameters are analyzed during testing. These parameters include: memory or CPU usage, system bandwidth, average request processing time, number of requests processed per unit of time, etc. In poor testing, only a few of these parameters are looked at and based on them, a conclusion is drawn about the effectiveness of the algorithm, while completely ignoring the others. In this regard, it is possible that the algorithm is quite efficient in terms of processing time and the number of input requests, but at the same time requires excessive memory usage.

To overcome these shortcomings, the proposed method plans to combine a machine learning algorithm and dynamic weight calculation. The use of machine learning technologies will help to improve the focus on big data, since in theory this will simplify the process of accepting the input data stream and allow the system to focus on its processing. The approach with dynamic weight calculation has long proven its effectiveness and many tests conducted on it gave positive results. The use of dynamic calculation will allow to maintain a balance in the resulting parameters in which the efficiency of some variables does not improve at the expense of the subsidence of others.

The schematic representation of the proposed algorithm is presented in Fig. 1.

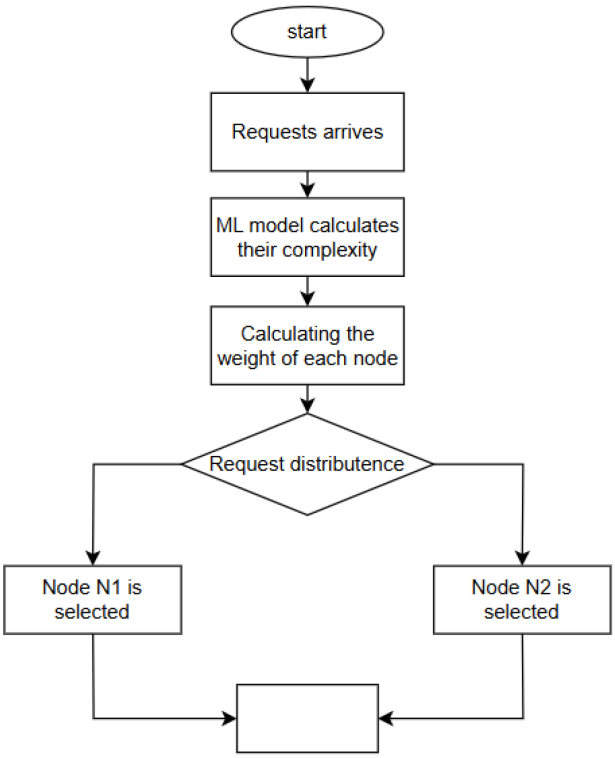


Fig. 1. Proposed algorithm workflow

The algorithm begins with the arrival of incoming requests, which are immediately processed by ML model to estimate their computational complexity. Based on this analysis, the system calculates the weight of each available node to reflect its current capacity and workload.

The requests are then distributed dynamically according to these calculated weights, ensuring that the load is balanced across the nodes. Depending on the results, a request is assigned to the most suitable node (e.g., Node N1 or Node N2), after which the process is repeated for subsequent requests, enabling continuous adaptation to workload fluctuations.

Research results

To implement the method, we first need to choose a programming language, and for this we should consider three options: C#, Java, and Python.

The main advantage of C# is its integration with Windows Server / Azure services, which provides better opportunities for deploying a multi-node system. This, in turn, will simplify the implementation itself and further testing and support. C# also has its own ML.NET machine learning library. The disadvantages of this language are platform binding and limited features of the built-in ML.NET. The C# language is quite dependent on its Windows Server / Azure infrastructure, and because of this, any implementation on other platforms is quite complicated [7], and the capabilities in the field of ML are less developed compared to Python.

Like C#, the Java language has advanced support for multi-node systems, however, the Java language allows focus on big data, this is achieved thanks to the existing Hadoop, Spark, Kafka, ZooKeeper frameworks. Java has support

for ML libraries, however, their limitations are the main disadvantage when choosing this language. The undoubted advantage of the Python language is its integrated libraries for working with machine learning. It provides the widest range of capabilities when creating and testing the developed model. Unlike the other two languages, the possibility of implementing and supporting a system with many nodes is less limited, but still present. The disadvantage is that, apart from the ML model, the language is not suitable for large systems.

None of the proposed languages can satisfy all the conditions of the proposed method, and therefore choosing only one is not optimal. In this regard, for the full implementation of the algorithm, it is planned to divide it into two components: the ML model itself and the balancer between nodes. Since the Python language has the best conditions for implementing the model, it is on it that the model will be built. To implement the balancer, it is worth using one of the two languages, C# or Java. In this article, attention is focused exclusively on the ML model, so the issue of implementing the balancer will not be considered in detail.

As mentioned, within the framework of this article, a ML model was developed in Python. This was achieved thanks to the built-in river library, which has all the necessary tools for implementing a ML model. In its structure, the algorithm uses an online learning model, which allows optimizing the learning process itself, since it does not require additional preparation time before starting work, unlike offline models. The choice of an online ML model is due to previous research in this field [8]. The principle of operation of the model is based on the expected processing time of the input query. To predict it, the model receives the necessary parameters of the query itself, such as its length, number of headers, type of operation, presence of JOIN in database queries, etc. Based on this information, the model selects a suitable node from the available ones, transmits the query to it, and at the same time uses its actual execution time for further training. Thanks to this, the system can simultaneously process the input stream and learn without additional preparation.

To test proposed algorithm, a three-node system has been simulated and a ten-thousand test set of various input queries has been created. The results of the created algorithm and the results of the basic algorithms are presented in the Table 1.

Table 1

Execution time comparison of baseline and proposed algorithms

Algorithm	Median query execution time	Execution time of the longest queries
Round Robin	195ms	728ms
Random	195ms	730ms
Join Shortest Queue	206ms	768ms
Proposed algorithm	206ms	763ms

The proposed algorithm provides an average efficiency improvement of 2–5 % due to a more balanced load distribution between nodes compared to the baseline algorithms. The median query execution time means that half of the queries were executed within this time, and the execution time of the longest queries was calculated based on the 5 % of the longest queries. According to this result, the basic RoundRobin algorithm showed the best performance, and the usual random distribution is not far behind and at first glance it may seem that the developed algorithm is not efficient, but this is not the case and there are several factors that indicate this.

First of all, the tests were conducted with a simulated data set on an emulated system with many nodes, that is, the entire infrastructure is controlled and unpredictable circumstances cannot arise in it. Under such conditions, the efficiency of the basic algorithms is always high, since they do not require additional calculations and can provide an instant result. Even under such conditions, the developed algorithm showed that it is able to achieve the level of efficiency of the basic ones. Under real conditions, the result of the algorithm may differ significantly from the current one.

Secondly, it is worth paying attention to the statistics of the distribution of queries between nodes. The best algorithm RoundRobin in terms of time showed an equal distribution of 3334, 3333 and 3333 requests for the first, second and third node respectively. Under real conditions, such a distribution can lead to delays on one of the nodes because it received longer requests compared to the others, which will negatively affect the overall performance of the entire system. When distributed by the developed algorithm, the first, second and third nodes received the following number of requests: 3356, 3256 and 3387, respectively. The reason for the discrepancy is that this algorithm took into account the objective duration of requests and the current queue on the nodes, which led to a more correct distribution. In general, the test results showed an equivalent efficiency of the algorithm compared to the baseline under simulated conditions, which in the future, when transferred to real conditions, may show the efficiency of the algorithm, since the baseline approaches have many shortcomings.

In Ukraine, where many sectors such as transportation, healthcare, energy, defense, and public administration are increasingly adopting digital platforms, the ability to process large amounts of information in real time without failures or bottlenecks is vital for efficiency and national competitiveness. For instance, in public transport systems or energy networks, even minor optimization of query distribution can lead to faster decision-making, more accurate forecasting, and better user experience, while in government digital services such improvements directly influence citizens' trust

and satisfaction. Since Ukrainian enterprises and institutions often operate under resource constraints, the ability of the algorithm to adaptively account for the complexity of incoming requests and dynamically redistribute loads offers a more sustainable and cost-effective alternative to simply scaling hardware infrastructure.

Conclusions and future work

Multi-node systems are becoming a common solution for optimal computing of large amounts of data, however, despite the availability of many solutions, they still have pressing problems with efficient query processing and working with big data. Three software-based distribution algorithms were considered. Most of them achieve a certain efficiency by changing the formula for calculating the weight of each node. The difference lies in the parameters used, such as the number of minimum connections or the physical distance to the node.

An algorithm was also considered that uses a machine learning model to predict the expected duration of an incoming request, based on which it was transferred to the appropriate node. Based on the algorithms studied, it was proposed to develop an own method that combines an online machine learning model with dynamic calculation of node weights. A prototype of this method was developed and tested in simulated conditions. As a result, the algorithm showed similar results to the basic approaches of the RoundRobin approach, which, due to the conditions, can speak about its effectiveness in a real experiment.

As a further work, it is planned to supplement the algorithm by improving the process of dynamic weight calculation. Also, for this algorithm, testing is planned in real conditions using a full system with many nodes and a reliable data set. Analysis of the results of the algorithm on such an infrastructure should show its effectiveness in comparison with the basic algorithms. The next step will be to compare the developed algorithm with modern options and analyze them when working with big data.

References

1. Hunter S. W., Smith W. E. (1999). Availability Modeling and Analysis of a Two Node Cluster. *5th Intl. Conference on Information Systems, Analysis and Synthesis*, Orlando, FL.

2. Yong Meng Teo, Ayani R. (2001). Comparison of Load Balancing Strategies on Cluster-based Web Servers. *Simulation*. 77 (5–6), pp. 185–195. https://doi.org/10.1177/00375497010770050

3. Belgaum M. R., Musa S., Alam M. M. and. Su'ud M. M. (2020). A Systematic Review of Load Balancing Techniques in Software-Defined Networking. *IEEE Access*, vol. 8, pp. 98612–98636. https://doi.org/10.1109/ACCESS.2020.2995849

4. Pan Z., Jiangxing Z. (2017). Load Balancing Algorithm for Web Server Based on Weighted Minimal Connections. *Journal of Web Systems and Applications*. Vol. 1. P. 1–8. DOI: https://dx.doi.org/10.23977/jwsa.2017.11001

5. Pei-rui J., Li-min M., Yu-zhou S., Yang-tian-xiu H. (2017). A client proximity based load balance algorithm in web sever cluster. *2nd International Conference on Wireless Communication and Network Engineering*. P. 317–322.

6. Omori M., Nishi H. (2018). Request Distribution for Heterogeneous Database Server Clusters with Processing Time Estimation. *International Conference on Industrial Informatics (INDIN)*, Porto. P. 278-283. DOI: 10.1109/INDIN.2018.8471931

7. Marcin Jamro. C# Data Structures and Algorithms. Second Edition. Published by Packt Publishing Ltd., in Birmingham, UK. 2024. – 349 p.

8. Okano H., Yamaguchi F., Takagiwa K., Nishi H. (2014). Traffic-based Flow Cache Port Separate Mechanism for Network processor. *The Institute of Electoronics, Information and Communication Engineers Technical Report*, vol. 114, no. 18, pp. 69–74.

Дата першого надходження рукопису до видання: 29.09.2025
Дата прийнятого до друку рукопису після рецензування: 24.10.2025
Дата публікації:

УДК 004.624 DOI

В. В. ВОЙТКО
кандидат технічних наук, доцент,
доцент кафедри програмного забезпечення
Вінницький національний технічний університет
ORCID: 0000-0002-3329-7256

М. Ю. ПОЗУР
аспірант кафедри програмного забезпечення
Вінницький національний технічний університет
ORCID: 0009-0003-5225-2453

МЕТОД РОЗШИРЕННЯ МОЖЛИВОСТЕЙ ПОВТОРНОГО ВИКОРИСТАННЯ ПРОГРАМНИХ КОМПОНЕНТІВ ШЛЯХОМ РЕАЛІЗАЦІЇ TRAITS В.NET

У статті розглядається метод розширення можливостей повторного використання програмних компонентів шляхом реалізації traits в.NET. Проаналізовано основні методи покращення можливостей повторного використання програмних компонентів в об'єктно-орієнтованій парадигмі програмування. Проаналізовано переваги та недоліки множинного та одиничного наслідування. Встановлено, що множинне наслідування має позитивний ефект на модульність та на можливості повторного використання компонентів, проте його використання може призвести до складнішої ієрархії класів та не завжди підтримується мовою програмування. В якості альтернативи наслідуванню розглянуто використання композиції для забезпечення повторного використання програмних компонентів. Проаналізовано концепт traits, який реалізовано у низці сучасних мов програмування в якості альтернативи множинному наслідуванню. В результаті аналізу встановлено, що traits дозволяє отримати схожий до множинного наслідування функціонал. Також проаналізовано пропозиції щодо розширення концепту за рахунок додавання можливості визначати стани, а не тільки поведінку. На основі цього розроблено модель traits для реалізації в.NET, що дозволяє визначати стани та забезпечити кращий контроль над перевантаженням членів при композиції. Оскільки композицію можна реалізувати за рахунок використання інструментів метапрограмування, то було проаналізовано відповідний інструментарій платформи.NET. У результаті аналізу визначено, що інструмент Incremental Source Generators у комбінації з Roslyn API дозволить досягти бажаного результату. Для цього запропоновано використовувати генератор коду, що буде генерувати часткові об'явлення класів, які міститимуть члени відповідного trait'а. В якості альтернативи ключовим словам запропоновано використовувати атрибути метаданих для контролю поведінки такого генератора коду. Генератор коду має спрацьовувати не лише під час компіляції, а й при змінах у вихідному коді. Такий метод дозволить повною мірою реалізувати концепт traits в.NET, що сприятиме покращенню можливостей повторного використання програмних компонентів при розробці на цій платформі.

Ключові слова: *метапрограмування,.NET, генерація коду, розробка програмного забезпечення, traits.*

V. V. VOITKO
Candidate of Technical Sciences, Associate Professor,
Associate Professor at the Department of Software Engineering
Vinnytsia National Technical University
ORCID: 0000-0002-3329-7256

M. YU. POZUR
Postgraduate Student at the Department of Software Engineering
Vinnytsia National Technical University
ORCID: 0009-0003-5225-2453

METHOD OF IMPROVING CODE REUSABILITY BY IMPLEMENTING TRAITS IN.NET

The article describes a method improving code reusability by implementing traits in.NET. The main methods for improving the possibilities of reusing software components in the object-oriented programming paradigm are analyzed. The advantages and disadvantages of multiple and single inheritance are analyzed. It is established that multiple inheritance has a positive effect on modularity and on the code reusability, however, it can lead to a more complex class hierarchy, has more complex semantics and is not always supported by the programming language. As an alternative to inheritance, the use of composition to ensure the reuse of software components is considered. The concept of traits is analyzed, which

© Войтко В. В., Позур М. Ю., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

utilizes composition and is implemented in a number of modern programming languages as an alternative to multiple inheritance. As a result of the analysis, it is established that traits allow you to obtain functionality similar to multiple inheritance without drawbacks it causes. A number of proposals for expanding the concept by adding the ability to define states are also analyzed. Based on this, a traits model was developed for implementation in.NET, which allows defining states and providing better control over member overriding during composition. Since composition can be implemented by utilizing metaprogramming techniques, the corresponding.NET platform metaprogramming toolkit was analyzed. As a result of the analysis, it was determined that the Incremental Source Generators tool in combination with the Roslyn API will allow achieving the desired result. For this purpose, it is proposed to use a code generator that will generate partial declarations of classes that will contain members of the corresponding trait. As an alternative to keywords, it is proposed to use metadata attributes to control the behavior of such a code generator. Such a code generator should work not only during compilation, but also react to the source code changes. This method will allow fully implementing the traits concept in.NET, which will help improve code reusability when developing on this platform.

Key words: *metaprogramming,.NET, code generation, software development, traits.*

Постановка проблеми

Платформа.NET та мова програмування C# реалізують об'єктно-орієнтовану парадигму програмування. Основним засобом повторного використання коду в такій парадигмі є механізм наслідування. В об'єктно-орієнтованому програмуванні наслідування забезпечує вертикальне повторне використання (vertical reuse), де клас-нащадок може використовувати поведінку та стани, визначені в батьківському класі. Саме ж наслідування може бути одиничним або множинним. Більшість сучасних мов програмування реалізують лише одиничне наслідування. Це обумовлено тим, що множинне наслідування має низку недоліків, що потребують складних механізмів на рівні мови програмування для їх усунення або мінімізації. Найбільш значною є проблема ромбовидного наслідування. Проте множинне наслідування дозволяє забезпечити значно кращі можливості для повторного використання коду в порівнянні з одиничним наслідуванням [1].

Альтернативою наслідуванню є використання композиції. На відміну від наслідування, що дозволяє нащадку використовувати поведінку й стани батьківського класу, композиція дозволяє включати до класу поведінку або стани з іншого класу або спеціально визначеного синтаксичного типу. Це дозволяє забезпечити горизонтальне повторне використання (horizontal reuse). Одним із концептів, що опирається на використання композиції для забезпечення кращих можливостей повторного використання коду, є traits [2]. Концепт передбачає наявність спеціальних синтаксичних одиниць, які називаються trait. Такі синтаксичні одиниці не беруть участі в наслідуванні та можуть визначати як поведінку, так і стани. Traits реалізовано в низці сучасних мов програмування, таких як: Rust, PHP та Scala.

Реалізація концепту traits в.NET та мови програмування C# дозволить значно покращити можливості повторного використання коду при розробці програмного забезпечення на цій платформі [3].

Аналіз останніх досліджень і публікацій

Було проведено аналіз публікацій та розробок, пов'язаних із темою дослідження. В першу чергу, було розглянуто низку публікацій, що описують особливості концепту traits та визначають семантику і модель їх композиції[2]. Оскільки оригінальне визначення концепту передбачає, що trait не може містити стани, то було проаналізовано публікацію, автори якої пропонують розширити концепт за рахунок можливості визначати стани [4].

Так як у статті розглядається реалізація концепту traits шляхом використання засобів метапрограмування платформи.NET, а саме інструментів Roslyn Source Generators [5] та Roslyn API, то було проаналізовано публікації та розробки, в яких описано використання цих інструментів. Наприклад, автори статті [6] описують використання Roslyn API для автоматичної генерації інтеграційних тестів для додатків, написаних мовою програмування C#. У статті [7] описано розробку бібліотеки, що дозволяє спростити написання unit-тестів за допомогою Roslyn API.

Також було проаналізовано бібліотеки, що використовують Roslyn Source Generators для реалізації функціоналу рівня мови програмування [8, 9].

Формулювання мети дослідження

Метою дослідження є розширення можливостей повторного використання коду в.NET шляхом розробки методу з реалізацією концепту traits та використанням інструментів метапрограмування платформи.NET, що дозволить розширити функціонал мови програмування та збільшити кількість сценаріїв повторного використання програмних компонентів.

Викладення основного матеріалу дослідження

Розглянемо оригінальну модель traits, яка описана у статті [10]. Запропонована модель виділяє такі основні множини:

– N – множина імен методів класу,
– B – множина тіл методів класу,
– A – множина змінних, що визначають стан екземпляру класу.

Метод класу визначається як $a \mapsto m$, де $a \in N$, $m \in B$. Множина тіл класів B розширена до B^* за рахунок подання елементів $\mathbb{U}$ та $\mathbb{C}$, що позначають обов'язковий (невизначений) метод та конфлікт методів відповідно.

Таким чином, набір методів класу можна визначити як $d \in D$, $d : N \to B^*$, який містить скінченну множину елементів. Клас $c \in \mathcal{C}$ у такому випадку подається у вигляді $c = \langle \alpha, d \rangle$, де $\alpha \in A$, набір змінних, що визначають стан екземпляру класу.

Автори визначать trait $t \in T$ як $t : N \to B^*$, де $t^{-1}(B \cup \mathbb{C})$ є скінченною множиною. Таким чином, trait – це словник методів, що не може містити станів. Виходячи з цього, модель визначає правила та операції для композиції trait і класу та trait і trait. Це необхідно для забезпечення можливостей вирішення конфліктів імен. Сама ж композиція визначається за виразом (1):

$$(d + t)(l) = d(l) \sqcup t(l). \tag{1}$$

Операція $\sqcup$ для визначень методів B^* виконується за правилами:

$\sqcup$	$\mathbb{U}$	m_1	m_2	$\mathbb{C}$
$\mathbb{U}$	$\mathbb{U}$	m_1	m_2	$\mathbb{C}$
m_1	m_1	m_1	$\mathbb{C}$	$\mathbb{C}$
m_2	m_2	$\mathbb{C}$	m_2	$\mathbb{C}$
$\mathbb{C}$	$\mathbb{C}$	$\mathbb{C}$	$\mathbb{C}$	$\mathbb{C}$

Основним механізмом вирішення конфліктів імен, передбачених у моделі, є перевантаження $d \triangleright t : N \to B^*$ (2):

$$(d \triangleright t)(l) = \begin{cases} t(l) & \text{якщо } d(l) = U \\ d(l) & \end{cases}. \tag{2}$$

Модель також передбачає можливість визначення аліасу для методу.

Це дозволяє визначати додаткове ім'я для методу, яке може бути використане у випадку, коли виникає конфлікт імен. Така операція визначається за правилом (3):

$$d[a \to b](l) = \begin{cases} d(l) & \text{якщо } l \neq a \\ d(b) & \text{якщо } l = a \wedge d(a) = U \\ C & \end{cases}. \tag{3}$$

Для того, щоб типаж міг визначати стани, пропонується розширити існуючу модель. Для цього визначимо основні множини:

- N_m – множина ідентифікаторів методів класу;
- B_m – множина визначень методів класу;
- N_s – множина ідентифікаторів станів класу;
- B_s – множина визначень станів класу.

Аналогічно до існуючої, така модель визначає набір методів класу або trait'а як словник $d \in D$, $d : N_m \to B_m^*$. Стани у цій моделі також визначаються як словник $s \in S$, де $s : N_s \to B_s^*$. Таким чином, клас можна подати як $c = \langle s, d \rangle$. Варто зауважити, що модель розглядається виключно в контексті станів та поведінки, тому інші типи членів класу, такі як конструктори або деструктори, ігноруються.

Словники методів та станів trait'а позначатимемо як d_t та s_t відповідно. Тоді trait $t \in T$ визначається як $t = \langle s_t, d_t \rangle$, що є аналогічним до визначення класу. Композиція словників визначається аналогічно до існуючої моделі як для станів, так і для методів. Операцію аліасу можна застосовувати до словників станів. Проте операція перевантаження виконується лише для словників методів. Існуюча модель за замовчуванням передбачає, що при композиції trait'а в клас така операція має виконуватися для всіх методів. Це призводить до неявних перевантажень, що може призвести до помилок під час виконання. Для запобігання цьому пропонується окремо визначати набір методів trait'а, які можуть бути перевантаженими при композиції. Для цього введемо $o_t \in O$, де $o_t : N_m \to B_m^*$ і $o_t^{-1}(\mathbb{U}) = \varnothing$, а $o_t^{-1}(B_m)$ є скінченною. Тоді trait можна подати як $t = \langle s_t, d_t, o_t \rangle$, де $(d_t + o_t)^{-1}(\mathbb{C}) = \varnothing$. Визначимо операцію $\cup$ за виразом (4), що описує композицію trait і класу:

$$c \cup t = \langle s + s_t, (d \triangleright o_t) + d_t \rangle. \tag{4}$$

У низці мов програмування traits реалізовано шляхом додавання коду trait'а в код класу. Для цього визначено спеціальні ключові слова та процедури компілятора. Цього можна досягти за допомогою метапрограмування, використовуючи методи метапрограмування, що опираються на аналіз існуючого вихідного коду та генерацію нового коду за результатами аналізу. Це дозволить генерувати визначення класів, що міститимуть члени відповідних trait'ів. Пропонуємо використовувати модель роботи генератора коду (рис. 1).

Така модель визначає спеціальний генератор коду, що спрацьовує як на етапі компіляції, так і при змінах у вихідному коду. Це дозволить здійснювати генерацію під час розробки без необхідності компіляції.

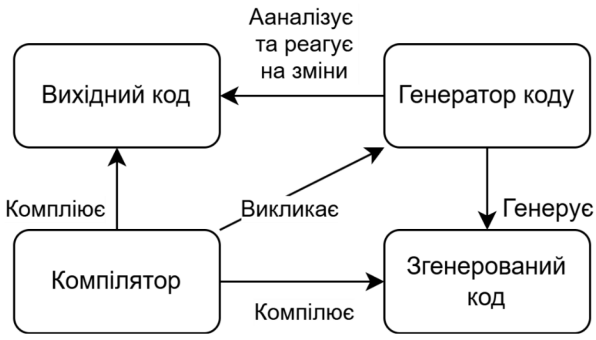


Рис. 1. Загальна модель роботи генератора коду

Найбільш доцільним інструментом, що дозволить забезпечити модель генерації (рис. 1), є інструмент Incremental Source Generators. Цей інструмент працює на рівні аналізатора коду, що дозволяє йому реагувати на зміни без необхідності компіляції. Разом з цим Incremental Source Generators дозволяють забезпечити хороші показники швидкодії генерації за рахунок використання інкрементного методу генерації.

Далі необхідно вирішити завдання визначення trait'а та класу, в який цей trait включається. Оскільки платформа.NET та мова програмування C# не підтримують traits, то відповідні ключові слова відсутні. Для вирішення цієї проблеми можна використовувати атрибути метаданих, адже доступ до них можливий як на етапі аналізу коду, так і під час його виконання. Таким чином, можна визначити атрибути метаданих для позначення типу даних як trait'а та класу, що включає в себе певний trait.

Далі варто визначити вимоги до типів даних trait'ів та класів, що їх включають. Виходячи з особливостей концепту traits та запропонованої моделі, основними вимогами до типу даних trait'а є:

– відсутність наслідування;
– можливість реалізовувати методи;
– можливість визначати метод без реалізації;
– відсутність можливості інстанціювання.

Отже, для визначення trait'а найбільш доцільним є використання абстрактного класу, який не бере участі в ієрархії наслідування. Щодо класу, в який включається trait, то ключовою вимогою є те, що такий клас має часткове об'явлення (partial). Це пов'язано із тим, що Incremental Source Generators не дозволяють вносити зміни в існуючий код, а передбачають лише генерацію нового коду.

Використання partial класів дозволить генерувати файл з частиною об'явлення класу, що містить члени відповідного trait'а (рис. 2).

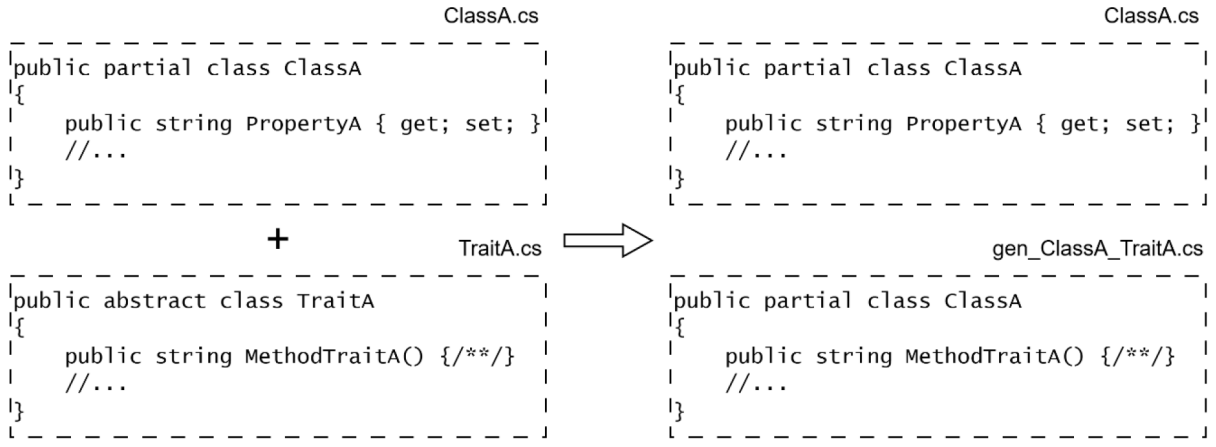


Рис. 2. Модель композиції trait'а і класу

Таким чином, композиція trait'а і класу відбуватиметься за рахунок генерації часткового об'явлення класу, що міститиме члени відповідного trait'а. Така модель дозволить повною мірою реалізувати особливості концепту traits.

Висновки

Запропоновано загальний метод реалізації traits в.NET з використанням методів метапрограмування. Метод використовує розширену модель trait's, що дозволяє визначати стани та контролювати перевантаження методів. Запропонований метод опирається на використання методів метапрограмування, що дозволяють генерацію

вихідного коду під час розробки та на етапі компіляції для генерації часткових об'явлень класів, що містять члени trait'ів, які вони включають. Для контролю поведінки генератора коду пропонується використовувати атрибути метаданих. Розроблений метод дозволить реалізувати концепт traits для платформи.NET та мови програмування C#, що сприятиме покращенню можливостей повторного використання програмних компонентів та модульності розроблюваного програмного забезпечення.

Список використаної літератури

1. Albalooshi F., Mahmood A. A Comparative Study on the Effect of Multiple Inheritance Mechanism in Java, C++, and Python on Complexity and Reusability of Code. *International Journal of Advanced Computer Science and Applications*. 2017. Т. 8, № 6. URL: https://doi.org/10.14569/ijacsa.2017.080614 (дата звернення: 20.08.2025).

2. Traits: Composable Units of Behaviour / N. Schärli та ін. *ECOOP 2003 – Object-Oriented Programming*. Berlin, Heidelberg, 2003. С. 248–274. URL: https://doi.org/10.1007/978-3-540-45070-2_12 (дата звернення: 20.08.2025).

3. Cassou D., Ducasse S., Wuyts R. Traits at work: The design of a new trait-based stream library. *Computer Languages, Systems & Structures*. 2009. Т. 35, № 1. С. 2–20. URL: https://doi.org/10.1016/j.cl.2008.05.004 (дата звернення: 20.08.2025).

4. A new modular implementation for stateful traits / P. Tesone та ін. *Science of Computer Programming*. 2020. Т. 195. С. 102470. URL: https://doi.org/10.1016/j.scico.2020.102470 (дата звернення: 20.08.2025).

5. Incremental Roslyn Source Generators In.NET6. URL: https://www.thinktecture.com/net/roslyn-source-generators-introduction/ (дата звернення: 20.08.2025).

6. Saadatmand M. Towards Automating Integration Testing of.NET Applications Using Roslyn. *2017 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, м. Prague, Czech Republic, 25–29 лип. 2017 р. 2017. URL: https://doi.org/10.1109/qrs-c.2017.99 (дата звернення: 20.08.2025).

7. Weber D., Christi A. Redusharptor: A Tool to Simplify Developer-Written C# Unit Tests. *International Journal of Software Engineering & Applications*. 2023. Т. 14, № 5. С. 29–40. URL: https://doi.org/10.5121/ijsea.2023.14503 (дата звернення: 20.08.2025).

8. Immutype. URL: https://github.com/DevTeam/Immutype (дата звернення: 20.08.2025).

9. AutoInterface. URL: https://github.com/beakona/AutoInterface (дата звернення: 20.08.2025).

10. Traits: A mechanism for fine-grained reuse / S. Ducasse та ін. *ACM Transactions on Programming Languages and Systems*. 2006. Т. 28, № 2. С. 331–388. URL: https://doi.org/10.1145/1119479.1119483 (дата звернення: 20.08.2025).

References

1. Albalooshi, F., & Mahmood, A. (2017). A Comparative Study on the Effect of Multiple Inheritance Mechanism in Java, C++, and Python on Complexity and Reusability of Code. *International Journal of Advanced Computer Science and Applications*, *8*(6). https://doi.org/10.14569/ijacsa.2017.080614

2. Schärli, N., Ducasse, S., Nierstrasz, O., & Black, A. P. (2003). Traits: Composable Units of Behaviour. У *ECOOP 2003 – Object-Oriented Programming* (с. 248–274). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-45070-2_12

3. Cassou, D., Ducasse, S., & Wuyts, R. (2009). Traits at work: The design of a new trait-based stream library. *Computer Languages, Systems & Structures*, *35*(1), 2–20. https://doi.org/10.1016/j.cl.2008.05.004

4. Tesone, P., Ducasse, S., Polito, G., Fabresse, L., & Bouraqadi, N. (2020). A new modular implementation for stateful traits. *Science of Computer Programming*, *195*, 102470. https://doi.org/10.1016/j.scico.2020.102470

5. Incremental Roslyn Source Generators In.NET 6. (2022). https://www.thinktecture.com/net/roslyn-source-generators-introduction/

6. Saadatmand, M. (2017). Towards Automating Integration Testing of.NET Applications Using Roslyn. У *2017 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C)*. IEEE. https://doi.org/10.1109/qrs-c.2017.99

7. Weber, D., & Christi, A. (2023). Redusharptor: A Tool to Simplify Developer-Written C# Unit Tests. *International Journal of Software Engineering & Applications*, *14*(5), 29–40. https://doi.org/10.5121/ijsea.2023.14503

8. Immutype. (2025). https://github.com/DevTeam/Immutype

9. AutoInterface. (2025). https://github.com/beakona/AutoInterface

10. Ducasse, S., Nierstrasz, O., Schärli, N., Wuyts, R., & Black, A. P. (2006). Traits: A mechanism for fine-grained reuse. *ACM Transactions on Programming Languages and Systems*, *28*(2), 331–388. https://doi.org/10.1145/1119479.1119483

Дата першого надходження рукопису до видання: 26.09.2025
Дата прийнятого до друку рукопису після рецензування: 23.10.2025
Дата публікації:

УДК 004.056.55:004.738.1 DOI

В. С. ВОЛОШИН
доктор технічних наук, професор,
професор кафедри промислової теплоенергетики
Приазовський державний технічний університет
ORCID: 0009-0005-6809-6779

ЩОДО ПИТАННЯ ПРО ОЦІНКУ ЯКОСТІ ІНФОРМАЦІЇ В ІНТЕРНЕТІ БЕЗ ЗАЛУЧЕННЯ ЦЕНЗУРИ

Робота присвячена проблемам розробки методів спрощеного аналізу достовірності інформації в мережі інтернет, як мети дослідження. Пропонується звернути увагу на такий об'єктивний показник інформації, як її ентропія, міра невизначеності або випадковості повідомлення. Показано, що сама по собі інформаційна ентропія не визначає, правдива чи хибна інформація належить аналізу. Однак вона здатна відображати якість інформації, пов'язану зі структурною визначеністю і передбачуваністю, що може бути використано як ознака правдоподібності. Ентропія поєднується зі структурною зв'язністю інформації. Це не абсолютний критерій істини, але зміни в структурній когерентності можуть бути показником потенційного обману, особливо якщо порівнювати з іншими структурами повідомлення. Приклади виявлення у складі текстового повідомлення таких критеріїв визначеності, як когерентність, банальність і хаотичність, за допомогою яких можна не тільки структурувати текстові матеріали, але й виділяти в них такі складові, які показують на достовірність цих матеріалів. Наведено приклади вимірювання інформаційної ентропії Шеннона для текстових повідомлень різної конфігурації, за допомогою якої можна співвимірювати інформацію за означеними вище функціями. Створено алгоритм пошуку неправдивої інформації в текстових повідомленнях, за допомогою якого можна послідовно виявляти режими об'єктивності сукупних інформаційних текстових структур в межах прийнятих параметрів – функцій когерентності, банальності, хаотичності. Запропоновані співвідношення цих функцій в інформаційному просторі «правда-фальшивість». Запропоновано чек-лист із найбільш коректних і об'єктивних відношень між зазначеними в повідомленні параметрами, який дозволяє надавати формалізовану оцінку рівнів довіри текстового повідомлення. Для оцінки стану формалізованих повідомлень запропоновано для використання радар-графіки у сумісних координатах: ентропія-когерентність-банальність-хаотичність для певної групи слів, що міститься у повідомленні. Запропонована методика оцінки рівнів неправдивості текстових повідомлень за прямими і непрямими показниками, незважаючи на її формалізацію, дозволяє в першому наближенні оцінити достовірність конкретних повідомлень, більш об'єктивно підходити до інформації в мережі Інтернет, яка вже довгий час обходиться без критики або цензури, і має стійку тенденцію до розвитку фальшивості цілих інформаційних потоків.

Ключові слова: *фальшива інформація, неправда-правда, інформаційна ентропія, когерентність, банальність, хаотичність, повідомлення, структура повідомлення.*

V. S. VOLOSHYN
Doctor of Technical Sciences, Professor,
Professor at the Department Energy Sciences
Priazovsky State Technical University
ORCID: 0009-0005-6809-6779

ON THE ISSUE OF ASSESSING THE QUALITY OF INFORMATION ON THE INTERNET WITHOUT THE INVOLVEMENT OF CENSORSHIP

The work is devoted to the problems of developing methods for simplified analysis of the reliability of information on the Internet, as the goal of the study. It is proposed to pay attention to such an objective indicator of information as its entropy, the measure of uncertainty or randomness of the message. It is shown that information entropy itself does not determine whether the information is true or false for analysis. However, it is able to reflect the quality of information associated with structural certainty and predictability, which can be used as a sign of plausibility. Entropy is combined with the structural coherence of information. This is not an absolute criterion of truth, but changes in structural coherence can be an indicator of potential deception, especially when compared with other message structures. Examples of identifying such criteria of certainty as coherence, banality and chaotic in a text message, with the help of which it is possible not only to structure textual materials, but also to highlight such components in them that indicate the reliability of these materials. Examples of measuring Shannon's information entropy for text messages of different configurations are given, with the help of which it is possible to co-measure information according to the above functions. An algorithm for searching for false information in text messages has been created, with the help of which it is possible to consistently identify the modes

© Волошин В. С., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

of objectivity of aggregate informational text structures within the accepted parameters – the functions of coherence, banality, chaoticness. The correlations of these functions in the truth-falsity information space are proposed. A checklist of the most correct and objective relations between the parameters specified in the message is proposed, which allows providing a formalized assessment of the confidence levels of a text message. To assess the state of formalized messages, it is proposed to use radar graphics in compatible coordinates: entropy-coherence-banality-chaotic for a certain group of words that contained in the message. The proposed methodology for assessing the levels of falsity of text messages by direct and indirect indicators, despite its formalization, allows in the first approximation to assess the reliability of specific messages, to more objectively approach information on the Internet, which has been doing without criticism or censorship for a long time, and has a steady tendency to develop falsity of entire information flows.

Key words: *false information, false-truth, informational entropy, coherence, banality, chaotic, message, message structure.*

Постановка проблеми

Очевидні обсяги недостовірної, дубльованої, неправдивої інформації, якою насичений сучасний Інтернет, і яка знаходиться у відкритому користуванні сотень мільйонів людей не тільки з числа звичайних читачів, але і з числа фахівців – фінансистів, вчених, політиків, для яких достовірність інформації є об'єктивною потребою. В умовах відсутності контролю за достовірністю такої інформації, здатністю неправди і брехні «маскуватися» під достовірність, шкода, що наноситься нею, вже порівнюється зі шкодою, завданою економічними, екологічними та іншими глобальними проблемами. Тому пошук методи виявлення такої неправдивості актуальні не тільки для наукового інформаційного простору, а й для суспільства в цілому.

Аналіз останніх досліджень і публікацій

За різними оцінками, інформація, що розміщується в інтернеті до 2025 року, досягла 175–200 ZB, збільшившись за останні п'ять років більш ніж у чотири рази (дані IDC-International Data Corporation). Структура цієї інформації, її спрямованість і способи впливу на сучасне суспільство стають все більш цікавими. Не останнє місце в таких дослідженнях займає об'єктивність і достовірність даної інформації.

Первинна інформація, вихідні дані, створені безпосередньо джерелом (наукові статті, офіційні документи, первинний користувацький контент) за даними [1], складають 10–20 % від загального обсягу (~14,4–28,8 ZB). Вторинна інформація, повтори, копії, інтерпретації, компіляції первинних даних (репости, рерайти, переклади) складають ~70 % інтернет-трафіка з 3 млн акаунтів та 4,5 млн твітері [2]. А за даними [3], це навіть 80–90 % (~115,2–129,6 ZB) всієї інформації в інтернеті. У будь-якому випадку в інтернеті переважає вторинна інформація.

Крім вторинної інформації, аналізу підлягає недостовірна інформація та інформація неправдивого характеру. Різні дослідження (Каліфорнійський університет, платформа DemandSage, міжнародна компанія GBTIMES, Фінляндія) показують що неправдива або хибна інформація може становити від 47,3 % до майже 79 % від сукупної інформації, що розташована в інтернеті, але точність залежить від методології дослідження. Таке розпорошення вже не є важливим, якщо врахувати, що дезінформація й у значно менших кількостях впливає на свідомість суспільства, якість наукових знань, об'єктивність прийнятих рішень, у тому числі на найвищому рівні сучасної політики або соціуму. Така інформація сприяє великій кількості різнооб'ємних кризових явищ в суспільстві, в науці.

Роботи С. Чжана та його колег [4] у галузі порівняльного моделювання для виявлення фейків за допомогою чисельних метрик та інших методів аналітики, а також методи К. Шу, Х. Лю та ін. [5, 6] у галузі моделей раннього виявлення на орієнтованих графах, надають можливостей для зниження ризику дезінформації та її оцінки у порівнянні з бенчмарком. Однак в такій же мірі, як і методи, описані в роботах [7, 8] по сукупній оцінці довіри до знань в алгоритмах ранжування інформації, вони дещо перевантажені критичними параметрами, різноманіття яких ставить ці методи в число чисто наукових і не застосовних на практиці. Уміння розпізнавати неправдиву інформацію різноманітними способами – аналітичними, експертними, топологічними [9, 10], за умовами їх нескладності для користувача, означає здатність уникнути величезної кількості помилок, нервових потрясінь, банкрутств, політичних втрат. А здатність запобігати та усувати спотворену та неправдиву інформацію може бути не менш важливою, ніж забруднення навколишнього середовища.

Формулювання мети дослідження

Створення простого за використанням і ефективного методу тестування інформації на предмет виявлення в ній недостовірних або неправдивих фрагментів та текстів.

Викладення основного матеріалу дослідження

У роботі пропонується звернути увагу на такий об'єктивний показник інформації, як її ентропія. Це міра невизначеності або випадковості повідомлення. Сама по собі ентропія не визначає, правдива чи хибна інформація. Однак це впливає на якість інформації, пов'язаної зі структурною визначеністю і передбачуваністю, що може бути використано як ознака ступеня правдоподібності [11].

1. Правдива інформація пов'язана з помірною ентропією, оскільки містить ієрархічно взаємопов'язане і структуроване, логічне і зв'язне подання даних, передбачувані зв'язки між частинами на основі знань про реальний світ.

2. Неправдива інформація може мати аномально високу або аномально низьку ентропію, що характеризується великою кількістю неструктурованих, незрозумілих і суперечливих фрагментів. Якщо ентропія близька до максимуму, то елементи однаково ймовірні, а повідомлення близьке до «шуму», до хаосу. Якщо ентропія близька до нуля, повідомлення є передбачуваним до точки примітивності.

Ентропія поєднується зі структурною зв'язністю інформації [12, 13]. Це не абсолютний критерій істини, але зміни в структурній когерентності можуть бути показником потенційного обману, особливо якщо порівнювати з іншими структурними особливостями повідомлення.

Під структурною когерентністю інформації ми розуміємо ступінь зв'язності, послідовності і логічності елементів повідомлення, яка виражається через впорядкованість розподілу інформації і зменшення несподіванок між частинами тексту. З точки зору інформаційної ентропії Шеннона, структурна когерентність – це нерівномірний, неусвідомлений розподіл ймовірностей появи елементів, де кожне наступне повідомлення передбачувано в контексті попереднього, але не тривіально. Це підтверджується об'єктивністю сучасних систем штучного інтелекту, де лінгвістичні екстраполяції вербальних структур є головною об'єктивністю генерування нових когерентних текстів. Когерентний текст тут розглядається як не надто хаотичний і не надто шаблонний. – він оптимально організований для передачі сенсу. Локальна ентропія в ньому (на рівні словосполучень, абзаців, окремих слів) різко не скаче і має закономірності, що відображають логіку і смислову зв'язність.

Розглянемо приклад зв'язного, помірно ентропійного, логічного повідомлення на кшталт: «*Економічне зростання країни збільшилося в другому кварталі, чому сприяло зростання внутрішнього споживання і стабільність на світових ринках*». Тут важливі такі параметри:

– передбачуваність, коли кожна наступна фраза логічно випливає з попередньої;
– помірна ентропія, що заснована на виборі строго обмежених контекстом слів;
– висока когерентність, що пов'язано з причинно-наслідковими зв'язками між словосполученнями.

Припустимо, що $H_k(p)$ ентропія Шеннона для зв'язного висловлювання. Тоді $N(H)$ – функція свідомої зв'язковості (когерентності) що визначається як $N(H) = -|H - H_{opt}| + \text{const}$, де $H_{opt} = \log_2(p)$ – оптимальна (когерентна) ентропія явно нетривіального тексту.

У зв'язному тексті, про який йде мова, всього 17 слів. Всі слова унікальні, ймовірність кожного $p = 1/17 = 0{,}0588$. Ентропія кожного слова $H_k(p) = -p \cdot \log_2(p) = -0{,}0588 \cdot \log_2(0{,}0588) = 0{,}2404$. Суммарна ентропія повідомлення $H = 17 \cdot 0{,}2404 = 4{,}0868$.

І тоді ми наведемо приклад навмисно незв'язного, високоентропійного або хаотичного повідомлення: «*Розвиток переривається, автомобіль рухається, розвинена автомагістраль, розрив, розвиток, дорога, кінь, автомобіль*». Він має *хаотичну структуру* з низькою передбачуваністю, відсутністю когерентності та логічних зв'язків у тексті. Для нього показник функції хаотичності визначається як $C(H) = -H \cdot D$, де D –. міра дисперсії розподілу хаосу.

У цьому хаотичному тексті всього 11 слів. Два слова повторюються, а 9 слів є унікальними. Для них $p_1 = \frac{2}{11} = 0{,}1818$ і $p_i = \frac{1}{11} = 0{,}0909$, , а $\overline{p} = \left(\frac{1}{10}\right) \cdot \left(\frac{2}{11} + 9 \cdot \frac{1}{11}\right) = 0{,}1$. . Дисперсія $D = \frac{1}{n}\sum (p_i - \overline{p})^2 = 0{,}0069$ є досить високою для такої виборки. Сумарна ентропія заданого хаотичного виразу $H_h^{11} = -[0{,}1818 \cdot (-2{,}459) + 9 \cdot 0{,}0909 \times \times (-3{,}468)] = 3{,}2815$. Функція хаосу в такому повідомленні $C(H) = -3.2815 \cdot 0.0069 \approx -0.0226$. Досить низьке і негативне значення вказує на низьку хаотичність в моделі повідомлення. Тобто розподіл слабко нерівномірний, та інформаційна ентропія низька. Якщо б у виразі було більше нерівномірних частот (наприклад, одне слово повторювалося 5 разів, решта повторювалися по одному разу), а ентропія залишалася високою, то випадковість $C(H)$ була б більш вираженою (~ –0,0554).

І проміжний приклад некогерентного, але низькоентропійного, стереотипного повідомлення: «*Китай зростає. Китай сильний. Китай завжди перемагає. Китай – найкращий*». Воно, перш за все, *банальне* (нова функція), про що свідчить надмірна передбачуваність, низька ентропія і, особливо, формалізована, але не змістовна зв'язність, що виражається в повторенні банальних гасел.

Тут функція банальності зменшується зі зростанням H в експоненціальній залежності: $-B(H) = \exp(-\lambda H)$. Банальний текст містить усього 9 слів: «Китай» – 4 рази: $p = 4/9 = 0{,}4444$; $H_б^4 = -p \cdot \log_2(p) = 0{,}4444 \cdot \log_2(0{,}4444) = 0{,}52$. Решта п'ять слів, для кожного $p = 1/9 = 0{,}1111$ $H_б^1 = -p \cdot \log_2(p) = 0{,}3522$. Ентропія кожного слова $H_б^1 = -p \cdot \log_2(p) = -0{,}0909 \cdot \log_2(0{,}0909) = 0{,}3145$. Суммарна ентропія $H_б^9 = H_б^4 + 5H_б^1 = 0{,}52 + 5 \cdot 0{,}3522 = 2{,}281$.

Отже, розрахункова ентропія для порівнянних текстів: когерентного – $H = 4{,}0868$; хаотичного – $H = 3{,}2815$; банального – $H = 2{,}281$. Це свідчить про те, що семантична новизна (когерентність) – це максимум інформативності при структурованому її викладі. Ентропія висока, але не максимальна. Хаотичний посил – це неорганізованість, випадковість. Ентропія може бути високою, але сенс не читається. А банальний меседж – це надмірне повторення, гасла. Ентропія низька, тому що елементи передбачувані і однорідні.

Таким чином, можна припустити, що фальшивість текстового повідомлення часто знаходиться в зонах надмірної випадковості (високий H і високий D) або банальності (низький H і високий рівень повторень). Істина

може існувати в зоні структурованої новизни ($H \approx 3{,}8$ и $D \approx 1 \cdot 10^{-4}$). Відповідно, будемо виділяти три режими інформаційної структури (табл. 1).

Інтерпретація інформаційних категорій, що нас цікавить в контексті інформаційної фальші, має свої відмінності. Зокрема, банальні тексти, шаблонна інформація мають низьку ентропію і майже не несуть новизни. Вони легко запам'ятовуються, піддаються натякам, тиражуванню і можуть маскувати пропаганду. Хаотичні тексти створюють відчуття складності, «інформаційного шуму». Це може бути методом приховування дезінформації або ускладнення її верифікації. Когерентні тексти ближче до правди, вони мають впорядковану структуру, помірну передбачуваність, схильні до структурованої невизначеності.

Таблиця 1

Режими інформаційних структур для повідомлень в межах прийнятих параметрів

Категорія інформації	Ентропія	Структура	Зв'язність (когерентність)	Змістовна новизна	Типова неправда
Банальна	Низька	Повторювана, шаблонна	Висока формально, але низька по суті	Низька	Пропаганда, кліше
Хаотична	Висока	Випадкова, фрагментарна	Низька	Низька	Фальсифікація, дезінформація
Когерентна	Середня	Логічно впорядкована	Висока	Висока	Маловірогідна, частіше правда

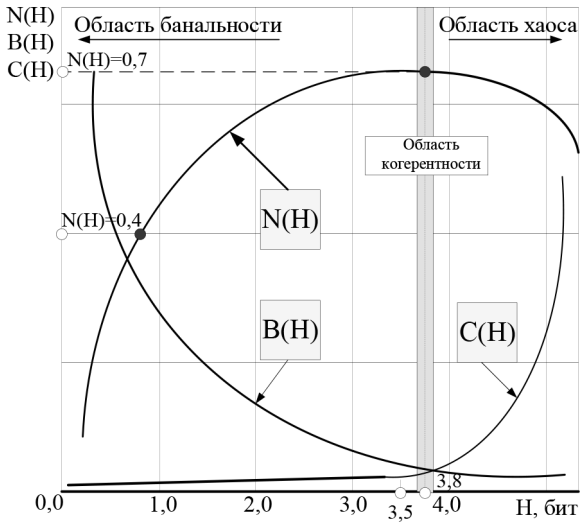


Рис. 1. Співвідношення когерентного, банального і хаотичного в інформаційному просторі «правди-неправди». Тут: *N*(*H*) – функція когерентності; *B*(*H*) – функція банальності; *C*(*H*) – функція хаотичності

Неважко помітити, що на досліджуваних залежностях (рис. 1) є особливі точки, а саме, при $H < 3{,}5$ випадковість інформації майже дорівнює нулю, а при $H > 3{,}5$ випадковість інформації починає різко зростати. Це означає, що висока ентропія в поєднанні з відсутністю структурованості і високою дисперсією робить текст хаотичним, важким для читання і підозрілим на об'єктивність.

Вертикальна лінія, що проходить через $H = 3{,}8$ ж це область максимальної новизни інформації, коли її банальність майже зникає, а хаос ще не настав. Такий діапазон значень ентропії може бути характерний для об'єктивно інформативних, логічних, когерентних тверджень.

Безумовна теорія відносності в цьому випадку присутня.

Запропонований алгоритм пошуку та відрізнення неправдивих текстових повідомлень в мережі Інтернет за параметрами, що розглядаються, включає наступне (рис. 2).

Збір статистичних даних по текстовому повідомленню, з розбивкою тексту на окремі слова, словосполучення за змістовним значенням з визначенням частоти слів в угрупованнях. На основі цієї інформації про повідомлення розраховується ентропія структури тексту, оцінюються параметри когерентності $N(H)$, банальності $B(H)$, хаотичності $C(H)$) тексту в порівнянні з їх ентропією. Такий алгоритм дозволяє аналізувати текстові повідомлення в мережі Інтернет на предмет їх достовірності в залежності від структурної ентропії тексту, використовуючи поняття когерентності, банальності і хаотичності інформації в повідомленні. Модель може бути представлена у вигляді *Python*-скрипта для конкретних рішень, що дозволяє розрахувати ентропію Шеннона, когерентність, банальність і хаотичність на основі статистики повторення слів у тексті, і зробити формалізовані висновки про його достовірність.

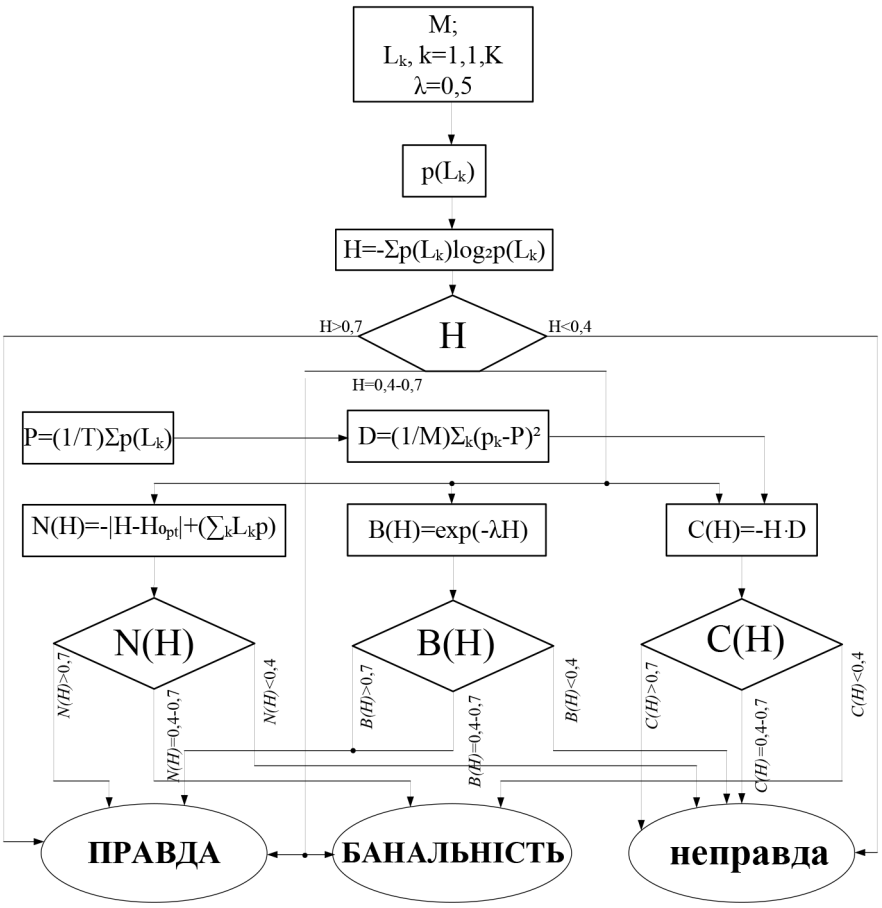


Рис. 2. Блок-схема пошуку неправдивої інформації в текстових повідомленнях.
Тут: *M* – загальна кількість слів у повідомленні; L_k – кількість слів у *k*-й групі однотипних слів;
λ – коефіцієнт банальності виразу; $p(L_k)$ – частотність слова в *k*-й групі однотипних слів

Зі структури самого алгоритму видно, що під найбільшим числом варіантів розподілу ентропії маються на увазі хаотичні стани інформації з відносно невеликою ентропією, що знаходяться в області інформаційної рівноваги.

Для аналізу достовірності тексту повідомлення, що підлягає дослідженню на достовірність, сформуємо чеклист з найбільш коректних і об'єктивних відношень між зазначеними в повідомленні параметрами (табл. 2).

Розглянемо кілька прикладів гіпотетичних повідомлень, вихідні дані і результати розрахунків для яких представлені в таблиці 3. Для умовного повідомлення з 128 слів, згрупованих у три приклади, ми обчислимо ентропійний стан кожного повідомлення та спробуємо провести формалізовану оцінку їх валідності

Результати аналізу рівнів достовірності змодельованих повідомлень за допомогою радар-графіків (рис. 3), на яких зображені всі три версії наших прикладів, очевидні. Чим більша «площа» фігури на графіку, тим різноманітніша та збалансованіша інформація є у повідомленні. Інформативність графіків полягає в наступному.

1. Варіант 3 повідомлення має компактну, лаконічну форму – ознака як мінімум інформаційної одноманітності. Він вирізняється дуже низькою хаотичністю та відносно високою банальністю – це ознака стереотипного, банального тексту.

2. Друга версія повідомлення має найбільш рівномірний і найширший показник радар-фігури, що свідчить про високу достовірність інформації. Вона має найвищу когерентність і ентропію, а також збалансовану хаотичність, що в сукупності дає сигнал правдоподібності.

3. У тексті першого варіанту показані помірні значення за всіма показниками, що підтверджують його проміжну позицію.

На завершення розглянемо текст реального повідомлення, взятого зі стрічки новин українського контенту Oboz.ua за 01.06.2025. У повідомленні йдеться про питання енергозабезпечення України влітку 2025 року та про те, до чого українцям варто готуватися. Обсяг тексту становив 299 слів, включаючи 7 видів груп з однаковими словами по два, три і т.д. слів. Вихідні дані та результати розрахунку тексту на неправдивість інформації представлені в таблиці 3 (остання колонка).

Таблиця 2

Чек-лист для формалізованої оцінки рівнів довіри текстового повідомлення (збільшення розміру шрифту в таблиці визначає рівень впевненості кожного наступного кластера)

Оцінка ентропії, у.о.	Параметрична оцінка повідомлення на довіру до тексту
Оцінка ентропії (Ознакою неправдивої інформації є її низька ентропія)	
0,0–0,2	Одноманітність фраз – ознака брехливості всієї інформації
0,2–0,4	Частотність вживання різних слів приблизно однакова
0,4–0,7	Різноманіття слів знаходиться на хорошому рівні, вселяє довіру до тексту
0,7–1,0	Унікальність словосполучень, різноманіття мови, як основа високої довіри до тексту
Оцінка когерентності (Чим більше ідей і пропозицій пов'язані між собою, тим вище зв'язність)	
0,0–0,2	Нелогічний текст, розбивки в сенсі і структурі – основа фейкового викладу
0,2–0,4	Логіка присутня, але є упущення або нестиковки
0,4–0,7	Ідеї логічно пов'язані між собою, текст структурований і довірчий
0,7–1,0	Дуже зв'язний і логічний текст, ідеально вибудувані зв'язки між реченнями та ідеями, вселяє впевненість у своєму значенні
Оцінка банальності (чим менш банальності, тем більш оригінальний текст)	
0,0–0,2	Недовіру **викликають** заїзні кліше, фрази, що часто повторюються
0,2–0,4	Повторювані фрази зустрічаються не надто часто, але вони викликають сумніви
0,4–0,7	Недовіру викликає велика кількість шаблонних виразів
0,7–1,0	Висока довіра до тексту з унікальними фразами та без шаблонів
Оцінка хаотичності (Чим менш хаотичний текст, тим він більш структурований і логічний)	
0,0–0,2	Чітка структура, речення, які плавно переходять один в одного – основа довіри до такого тексту
0,2–0,4	Є невеликі деформації в структурі, але в цілому текст логічний
0,4–0,7	Деякі фрагменти важкі для розуміння, але в цілому текст здається зрозумілим
0,7–1,0	Інформація здається неправдивою через хаос, уривчасті фрази з нелогічними переходами
Семантична новизна (Чим оригінальніший текст, тим вища його новизна)	
0,0–0,2	Текст викликає сумніви через повторюваність відомих ідей і не приносить нічого нового
0,2–0,4	Ідеї частково нові, але в основному знайомі
0,4–0,7	Пропонуються нові ідеї або аргументи, текст виходить цікавим і оригінальним, що вселяє впевненість в його об'єктивності
0,7–1,0	Дуже оригінальний і новий текст, приносить нові ідеї та концепції, що вселяє до себе впевненість в об'єктивності

Таблиця 3

Порівняльні дані інформаційно-ентропійного аналізу умовного повідомлення, що складається з 128 слів (під четвертим варіантом мається на увазі реальне текстове повідомлення)

Показник	Варіант 1	Варіант 2	Варіант 3	Oboz.ua
Структура груп слів	12 × 2; 5 × 3; 5 × 4; 2 × 5	25 × 2; 1 × 3; 1 × 4; 2 × 5	25 × 2; 25 × 3	128 × 1; 34 × 2; 8 × 3; 7 × 4; 4 × 5; 3 × 6; 1 × 13
Число унікальних слів	59	61	3	185
Ентропія Шеннона (H)	6,14	6,33	5,66	7,228
Когерентність, $N(H)$	6,64 (H_{opt} = 6,5; C = 7)	6,83	6,16	0,6789
Банальність, $B(H)$	0,00216	0,00178	0,00345	7,5314
Дисперсія ймовірностей D	0,60	0,55	0,30	0,0272
Хаотичність повідомлення, $C(H)$	–3,684	–3,480	–1,698	–0,000149
Інтерпретація вірогідностей	Помірно правдоподібне, частково шаблонне	Найбільш природне, збалансоване	Маловірогідне, кліше-синтетичне	З ймовірністаю 90 % комунікації є об'єктивними, а не фейковими

Результатом такого аналізу є судження про те, що цей звіт від 01.06.2025 Oboz.ua є об'єктивним і далеко не хибним, принаймні в обсязі даних, які викладені в повідомленні.

Висновки

Запропонована методика оцінки рівнів неправдивості текстових повідомлень за прямими і непрямими показниками, незважаючи на її формалізацію, дозволяє в першому наближенні оцінити достовірність конкретних повідомлень, що дозволяє більш об'єктивно підходити до інформації в мережі Інтернет, яка вже довгий час обходиться без критики або цензури, і має стійку тенденцію до розвитку фальшивості цілих інформаційних потоків. Такі підходи можуть бути корисними, зокрема, в наукових публікаціях, при рецензуванні наукових статей і книг.

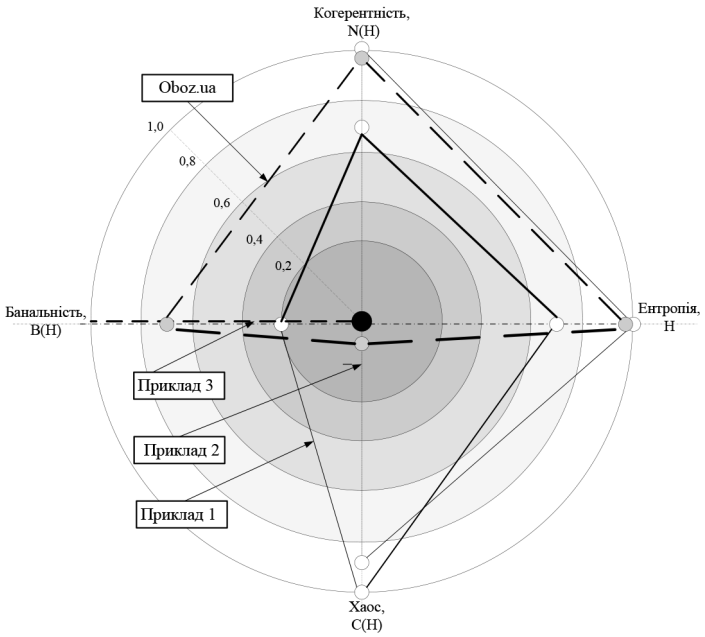


Рис. 3. Радар-графік для оцінки стану формалізованих повідомлень, що складаються з 128 слів в різних групах початкових сигналів (слів)

Список використаної літератури

1. Yavorska O. "The Knowledge As a Complementary Part of the Intellectual Capital of the Enterprises in the Era of the Digital Economy". Herald of Economics, no. 3(97), 2020, P. 185Ц198. – DOI: 10.35774/visnyk2020.03.185

2. Aral V.R. The spread of true and false news online. Science, 2018. DOI: 10.1126/science.aap9559

3. Reinsel D., Gantz J., Rydning J. The Digitization of the World. From Edge to Core. IDC White Paper I., (официальный идентификатор US44413318), 2018. P. 11.

4. Zhang C., Gupta A., Kauten C., Deokar A. V., Qin, X. Detecting Fake News for Reducing Misinformation Risks Using Analytics Approaches. European Journal of Operational Research (Elsevier, Amsterdam), 279(3), 2019. pp. 1036–1052. – DOI: 10.1016/j.ejor.2019.06.022

5. Shu K. Liu, H., Detecting Fake News on Social Media. Synthesis Lectures on Data Mining and Knowledge Discovery (Morgan & Claypool, San Rafael, CA), 2019, xiii + 115 pp. – DOI: 10.1007/978-3-031-01915-9

6. Shu, K., Wang S., Liu H. A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. ACM Computing Surveys (Association for Computing Machinery, New York, NY), 53(5), Article 109, 2020. P. 109:1–109:40. – DOI: 10.1145/3395046

7. Vosoughi S., Roy D., Aral S. The Spread of True and False News Online. Science (AAAS, Washington, DC), 359(6380), 2018. P. 1146–1151. – DOI: 10.1126/science.aap9559

8. Pennycook G., Rand D. G. Fighting Misinformation on Social Media Using Crowdsourced Judgments of News Source Quality. Proceedings of the National Academy of Sciences (PNAS) (National Academy of Sciences, Washington, DC), 116(7), 2019. P. 2521–2526. – DOI: 10.1073/pnas.1806781116

9. Cafaro C., Giffin A., Ali S.A., Kim D.-H. Reexamination of an Infor-mation Geometric Construction of Entropic Indicators of Complexity. 2010. 20 p. – DOI: 10.1016/J.AMC.2010.08.028

10. Hall M. J. W. Universal Geometric Approach to Uncertainty, Entropy and Information. 1999. 15 p. – DOI: 10.1103/PhysRevA.59.2602

11. Bertani A., Mazzeo V., Gallotti R. – Decoding the News Media Diet of Disinformation Spreaders. Entropy (MDPI), Vol. 26, Basel, 2024, Article 270 – DOI: 10.3390/e26030270

12. Gray R. Entropy and Information Theory., USA, Springer, 1990. 412 p. – DOI: 10.1007/978-1-4419-7970-4

13. Blasch E., Schuck T., Gagne O.B. Trusted Entropy-Based Information Maneuverability for AI Information Systems Engineering. Engineering Artificially Intelligent Systems, Lecture Notes in Computer Science (LNCS), Springer, Cham, 2021, pp. 34–52. – DOI: 10.1007/978-3-030-89385-9_3

References

1. Yavorska O. "The Knowledge As a Complementary Part of the Intellectual Capital of the Enterprises in the Era of the Digital Economy". Herald of Economics, no. 3(97), 2020, P. 185–198. – https://doi.org/10.35774/visnyk2020.03.185

2. Aral V.R. The spread of true and false news online. Science, 2018. https://doi.org/10.1126/science.aap9559

3. Reinsel D., Gantz J., Rydning J. The Digitization of the World. From Edge to Core. IDC White Paper I., (official identifier US44413318), 2018. P. 11.

4. Zhang C., Gupta A., Kauten C., Deokar A. V., Qin, X. Detecting Fake News for Reducing Misinformation Risks Using Analytics Approaches. European Journal of Operational Research (Elsevier, Amsterdam), 279(3), 2019. pp. 1036–1052. – https://doi.org/10.1016/j.ejor.2019.06.022

5. Shu K. Liu, H., Detecting Fake News on Social Media. Synthesis Lectures on Data Mining and Knowledge Discovery (Morgan & Claypool, San Rafael, CA), 2019, xiii + 115 pp. – https://doi.org/10.1007/978-3-031-01915-9

6. Shu, K., Wang S., Liu H. A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. ACM Computing Surveys (Association for Computing Machinery, New York, NY), 53(5), Article 109, 2020. P. 109:1–109:40. – https://doi.org/10.1145/3395046

7. Vosoughi S., Roy D., Aral S. The Spread of True and False News Online. Science (AAAS, Washington, DC), 359(6380), 2018. P. 1146–1151. – https://doi.org/10.1126/science.aap9559

8. Pennycook G., Rand D. G. Fighting Misinformation on Social Media Using Crowdsourced Judgments of News Source Quality. Proceedings of the National Academy of Sciences (PNAS) (National Academy of Sciences, Washington, DC), 116(7), 2019. P. 2521–2526. – https://doi.org/10.1073/pnas.1806781116

9. Cafaro C., Giffin A., Ali S. A., Kim D.-H. Reexamination of an Infor-mation Geometric Construction of Entropic Indicators of Complexity. 2010. 20 p. – https://doi.org/10.1016/J.AMC.2010.08.028

10. Hall M. J. W. Universal Geometric Approach to Uncertainty, Entropy and Information. 1999. 15 p. – https://doi.org/10.1103/PhysRevA.59.2602

11. Bertani A., Mazzeo V., Gallotti R. – Decoding the News Media Diet of Disinformation Spreaders. Entropy (MDPI), Vol. 26, Basel, 2024, Article 270 – https://doi.org/10.3390/e26030270

12. Gray R. Entropy and Information Theory., USA, Springer, 1990. 412 p. – https://doi.org/10.1007/978-1-4419-7970-4

13. Blasch E., Schuck T., Gagne O.B. Trusted Entropy-Based Information Maneuverability for AI Information Systems Engineering. Engineering Artificially Intelligent Systems, Lecture Notes in Computer Science (LNCS), Springer, Cham, 2021, pp. 34–52. – https://doi.org/10.1007/978-3-030-89385-9_3

Дата першого надходження рукопису до видання: 20.09.2025
Дата прийнятого до друку рукопису після рецензування: 15.10.2025
Дата публікації:

UDC 004.932.2:004.8:004.58 DOI

A. O. HAVRASHENKO
Postgraduate Student at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0000-0002-8802-0529

O. YU. BARKOVSKA
Candidate of Technical Sciences, Associate Professor,
Associate Professor at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0000-0001-7496-4353

ASSESSMENT OF THE ACCURACY OF ARTIFICIAL LANGUAGE TRANSLATION METHODS

The article is devoted to evaluating the accuracy of methods for translating artificial languages in an intelligent information system for translating artificial languages using artificial intelligence methods and statistical algorithms, proposing an innovative approach to improving the effectiveness of machine translation in the context of the dynamic development of modern technologies. The analysis showed insufficient translation accuracy, which was proposed to be improved by integrating two methods into one system.

The aim of the work is to evaluate the accuracy of translation methods in the proposed intellectual information system for translating artificial languages. The work examines such translation evaluation metrics as BLEU, ChrF, BLEURT, COMMET, TER, METEOR and human evaluation (as a benchmark). Each of the algorithms (mathematical models, neural network algorithm and combined algorithm) was tested on 12 tests, each of which involved translating one of four languages into another (English, Ukrainian, French and Esperanto) and vice versa. The results show an increase in average translation accuracy of 0.5 % compared to the method based on mathematical models and 0.2 % compared to the method based on neural networks. In some pairs, the result was worse than the individual algorithm, but this was due to the coefficients used when combining the results. Since individual settings are possible for each language pair, the correct result will appear in the worst case with the same accuracy as the best of the two results. However, this requires separate configuration. Based on the results obtained, it can be concluded that the combination of these two methods improves the accuracy of artificial language translation. The results confirm the effectiveness of the developed system, which allows for highly efficient translation of natural and artificial languages.

Key words: *machine learning, artificial intelligence, LSTM, data networks, sequence-to-sequence, translation, language, method, metric, mathematical model.*

А. О. ГАВРАШЕНКО
аспірант кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0000-0002-8802-0529

О. Ю. БАРКОВСЬКА
кандидат технічних наук, доцент,
доцент кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0000-0001-7496-4353

ОЦІНКА ТОЧНОСТІ МЕТОДІВ ПЕРЕКЛАДУ ШТУЧНИХ МОВ

Стаття присвячена оцінці точності методів перекладу штучних мов в інтелектуальній інформаційній системі перекладу штучних мов із використанням методів штучного інтелекту та статистичних алгоритмів, пропонуючи інноваційний підхід до підвищення ефективності машинного перекладу в умовах динамічного розвитку сучасних технологій. Проведений аналіз показав недостатню точність перекладу, підвищити яку було запропоновано шляхом інтеграції двох методів в одну систему.

Метою роботи є оцінка точності методів перекладу в запропонованій інтелектуальної інформаційної системи перекладу штучних мов. В роботі досліджено такі метрики оцінки перекладу, як BLEU, ChrF, BLEURT, COMMET, TER, METEOR та людська оцінка (в якості еталону). Кожен з алгоритмів (математичні моделі, нейромережевий алгоритм та комбінований алгоритм) були перевірені на 12 тестах, кожен з яких – переклад

© Havrashenko A. O., Barkovska O. Yu., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

однієї з чотирьох мов в іншу (англійська, українська, французька та есперанто) та в зворотному порядку. Результати демонструють, що дає підвищення середньої точності перекладу на 0.5 % у порівнянні з методом на основі математичних моделей, на 0.2 % у порівнянні з методом на основі нейронних мереж. В деяких парах результат був гірше за окремо взятий алгоритм, але це відбулося через використані коефіцієнти при складені результатів. Так як для кожної окремої пари мов можливо індивідуальне налаштування, то правильний результат буде з'являтися в найгіршому випадку з тією ж точністю, що і найкращий серед 2 результатів. Але це потребує окремого налаштування. Враховуючи отримані результати, можна зробити висновок, що комбінація цих двох методів дозволяє покращити точність перекладу штучних мов. Отримані результати підтверджують ефективність розробленої системи, що дозволяє з високою ефективністю перекладати звичайні та штучні мови.

Ключові слова: *машинне навчання, штучний інтелект, LSTM, датасети, послідовність у послідовність, переклад, мова, метод, метрика, математична модель.*

Formulation of the problem

Translation quality assessment is a fundamental element in both the practical work of translators and modern research in the field of automatic language processing. [1] It determines the level of compliance of the resulting text with user expectations, communication needs and the norms of the target language. This process is complex in nature, as it involves the analysis of a whole range of parameters covering the accuracy of content transmission, compliance with linguistic and cultural norms, stylistic consistency, and the overall comprehensibility of the text. This makes evaluation a universal tool for quality control and, at the same time, a benchmark for further improvement of translation technologies.

In scientific and engineering practice, particular attention is paid to systematic approaches to evaluation. They allow avoiding subjectivity and form a single basis for comparing different methods or models. On the one hand, translation quality assessment acts as a diagnostic mechanism: it helps to identify weaknesses related to vocabulary choice, syntactic structure, or grammatical norms. On the other hand, it acts as a driving force for progress, as systematic observations and analysis results are used to create new algorithms capable of producing more natural and accurate translations.

Modern approaches to evaluation are largely focused on automation. This is due to the rapid development of neural networks and the growing use of translation technologies, where traditional verification methods are becoming too labour-intensive. Automated evaluation allows for the effective analysis of large amounts of data, while ensuring speed and repeatability of results. However, linguistic and qualitative assessments remain relevant, as they take into account nuances that are difficult to formalise.

In general, translation quality assessment has a dual significance. It is a practical tool for control and standardisation, ensuring that a text complies with certain norms, as well as a scientific method for developing and improving modern translation systems. Thus, it is through evaluation that a balance is achieved between technical capabilities, linguistic requirements and end-user expectations.

This work is devoted to evaluating the accuracy of artificial language translation methods. It examines various automatic methods of assessing the quality of text translation. Particular attention is paid to comparing different types of artificial language translation methods and combining them into a single system.

The scientific novelty lies in evaluating the accuracy of artificial language translation methods in the proposed intellectual information system for artificial language translation. The practical significance of the proposed system is the evaluation of the effectiveness of the method of translating artificial languages based on mathematical models and the method of translating artificial languages based on neural network translation, and their comparison and integration to ensure high accuracy and reliability of results.

Analysis of the latest research and publications

Evaluating the quality of machine translation is a critically important step in the research and practical application of automatic translation systems. Without reliable measurement methods, it is difficult to understand how well the system performs the task at hand, whether its performance has improved after modifications, and how it compares to other approaches. Therefore, over the past decades, a whole range of methods and metrics have been developed to evaluate translations both automatically and with the help of experts.

All metrics for evaluating the accuracy of machine translation systems can be divided into two broad groups (Fig. 1):

1. Automated metrics – algorithms that compare machine translation with one or more reference translations created by humans. They allow for the rapid processing of large amounts of data and are widely used in research.
2. Human evaluation – experts or native speakers evaluate the translation according to criteria such as comprehensibility, grammatical correctness and content adequacy. This approach is considered the most accurate, but it is expensive and labour intensive.

Automated metrics cannot completely replace human evaluation, but they provide a fast and relatively objective way to compare different models.

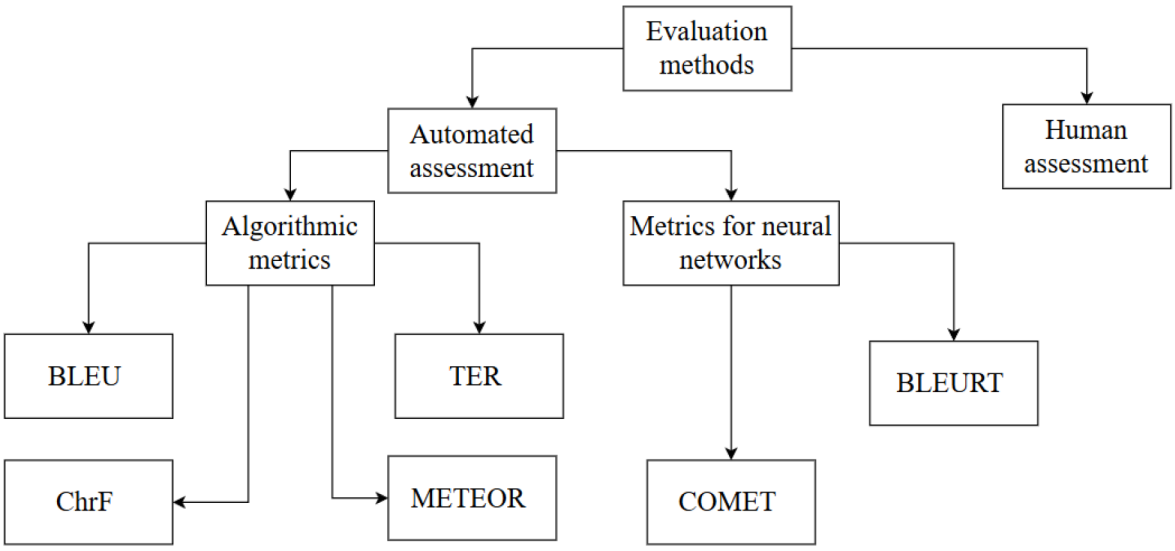


Fig. 1. Classification tree of metrics for evaluating the accuracy of machine translation systems

BLEU is one of the first and most well-known metrics, proposed in 2002 by IBM researchers. It is based on the idea of counting n-grams (sequences of n words) in a machine translation and in a so-called 'reference translation' (a benchmark created by a human). If the machine translation contains a large number of matches with the reference, the result is considered high quality.

A distinctive feature of BLEU is the penalty for overly short translations. For example, if the system attempts to translate a long sentence with only a few words that match the reference, it could receive a high score without the penalty. The penalty reduces the metric value in such cases, encouraging the completeness of the translation.

ChrF is a metric that focuses not on words, but on character sequences. [5, 6] It works well for languages with rich morphology, where changes in endings or word formation can significantly alter the appearance of a word.

The principle of operation is to count character matches (n-grams) between the machine translation and the reference. Thanks to this, even if the words differ but have a common root, the metric will consider the similarity.

The METEOR metric appeared in 2005 as a response to the limitations of BLEU. [7] It considers not only exact word matches, but also morphological variations, synonymy, and word order in a sentence. This makes it closer to human perception.

The principle of METEOR is to find correspondences between the words of the machine translation and the reference text. To do this, different levels of comparison are used:

- exact word matches;
- stemming (matching word roots, e.g., 'translate' and 'translated');
- synonymy (thanks to dictionaries such as WordNet);
- permutation (assessing the impact of incorrect word order).

As a result, METEOR correlates better with human evaluation than BLEU, especially on short texts.

TER was developed in 2006 as a tool to show how many edits a human would need to make to correct a machine translation. These edits may include:

- insertions;
- deletions;
- word replacements;
- fragment rearrangements.

Therefore, TER essentially measures the 'cost of correction.' For example, if only one word needs to be changed to edit a sentence, TER will be low, but if most of the text needs to be rewritten, the value will be high.

In the 2020s, a new class of metrics emerged that use contextual representations of words obtained from language models.

BLEURT is a modern metric based on neural networks. It uses pre-trained transformers (BERT and its modifications), which allow evaluating not only the formal similarity of texts, but also their semantic proximity.

BLEURT undergoes preliminary training on a large amount of text and then undergoes additional training on corpora with human translations. This makes it more 'human-like' in its evaluation.

BLEURT is a trained model that combines large corpora of synthetic data and human evaluations. Thanks to this, it is able to take into account both lexical matches and overall semantic adequacy.

COMET is another modern metric developed on the basis of transformers. It uses multilingual models that allow comparing texts in different languages without the need for direct word matches [10].

A distinctive feature of COMET is that it takes into account three texts simultaneously:

1. the original (source text);
2. the machine translation;
3. the reference translation.

This enables COMET to better understand whether a translation corresponds to the original content, rather than just formally resembling the reference.

Despite significant progress in the development of automatic metrics, they cannot completely replace human evaluation. Research often uses the MQM (Multidimensional Quality Metrics) framework, which classifies errors by type (lexical, grammatical, stylistic) and severity (minor, serious, critical). MQM provides a detailed picture of the strengths and weaknesses of the system, and automatic metrics are used as a supporting tool. Not only the translation of words, but also the conveyance of meaning, style, and intention of the statement.

Formulation of the purpose of the research

The aim of this work is to evaluate the accuracy of methods for translating artificial languages in the proposed intellectual information system for translating artificial languages. The research component of the proposed system is to evaluate the effectiveness of the method of translating artificial languages based on mathematical models and the method of translating artificial languages based on neural network translation, and to compare and combine them in order to ensure high accuracy and reliability of results.

To achieve this goal, the following tasks must be solved:

– analysis of metrics for evaluating the accuracy of automated translation systems;
– development of experimental methodology;
– creation of a model for an intelligent information system for translating artificial languages;
– evaluation of each method separately, their comparison and combination.

The experiments conducted form the basis for further improvement of methods for translating artificial languages.

The limitations of the system include the need to use a parallel corpus of texts in different languages. Another limitation is the need to train the system on the words that will be used for translation.

Research Results and Discussions

This study was the first to compare the method of translating artificial languages based on mathematical models with the method of translating artificial languages based on neural networks in the proposed artificial language translation system.

To test the effectiveness of the developed approach, an experiment was conducted involving four languages: Ukrainian, English, French, and Esperanto. These languages were chosen because of their different linguistic characteristics. English is

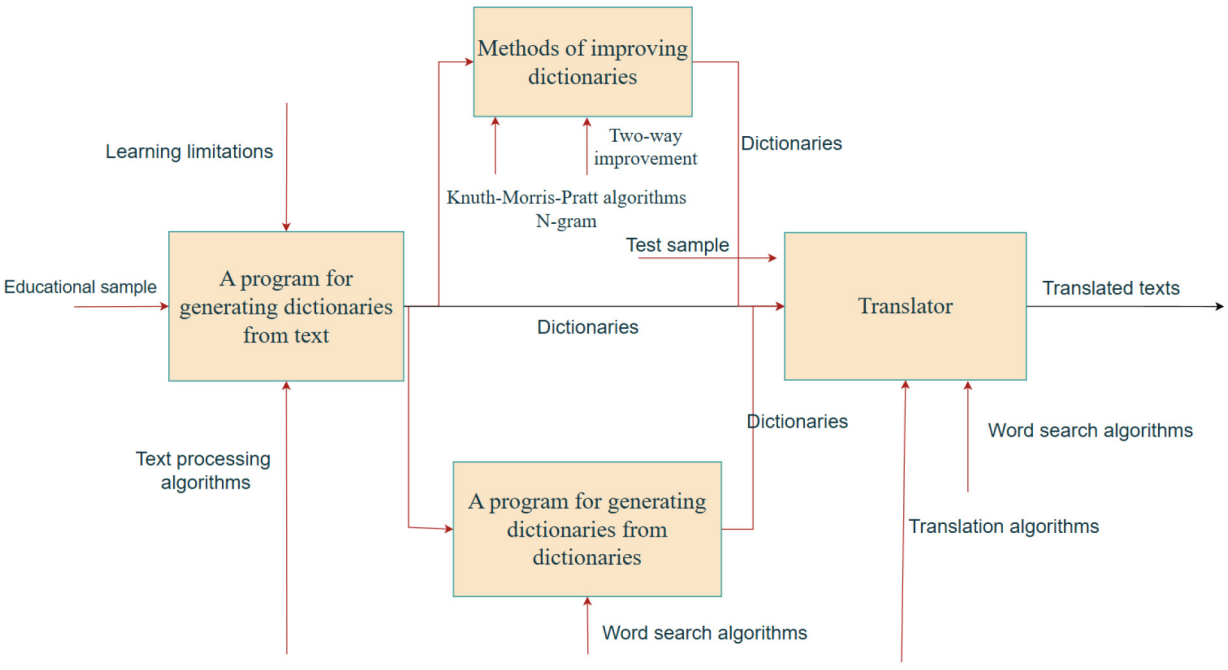


Fig. 2. Method of translating artificial languages based on mathematical models

relatively easy to process by machines, French and Ukrainian have developed morphology, and Esperanto is distinguished by its simplified grammar. This made it possible to evaluate the universality of the system.

The experiment used a developed set of programs consisting of two main components.

A module based on mathematical models (Fig. 2) implements translation using statistical methods. It processes input texts using a set of modules, each of which contains an algorithm based on a mathematical model. After that, the aggregate result is processed, and the resulting statistical translation dictionary is generated. [11, 12]

The method of translating artificial languages based on neural networks (Fig. 3) is built on modern machine learning architectures, capable of taking context into account and identifying hidden patterns.

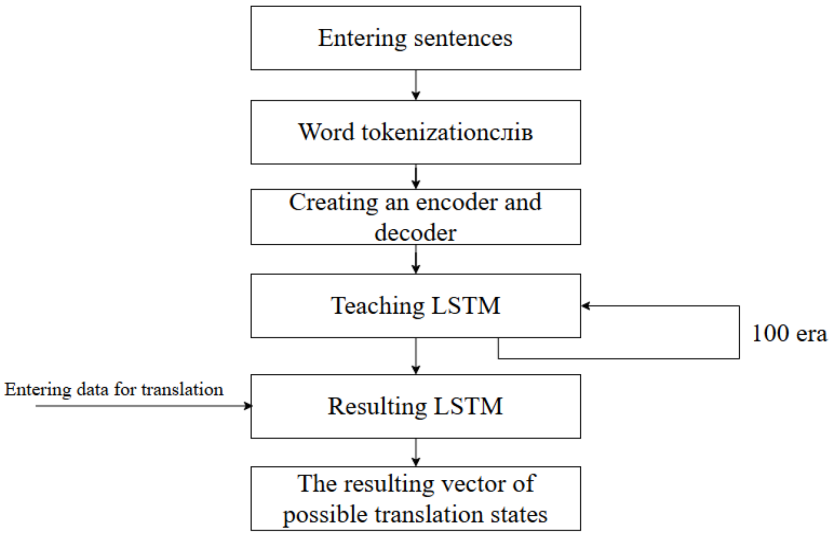


Fig. 3. Method of translating artificial languages based on neural networks

These methods were integrated into a single intelligent information system for translating artificial languages (Fig. 4): pre-processing was performed by an algorithmic block, after which the results were refined by a neural network. This approach combines the rigour of rules with the flexibility of adaptive models.

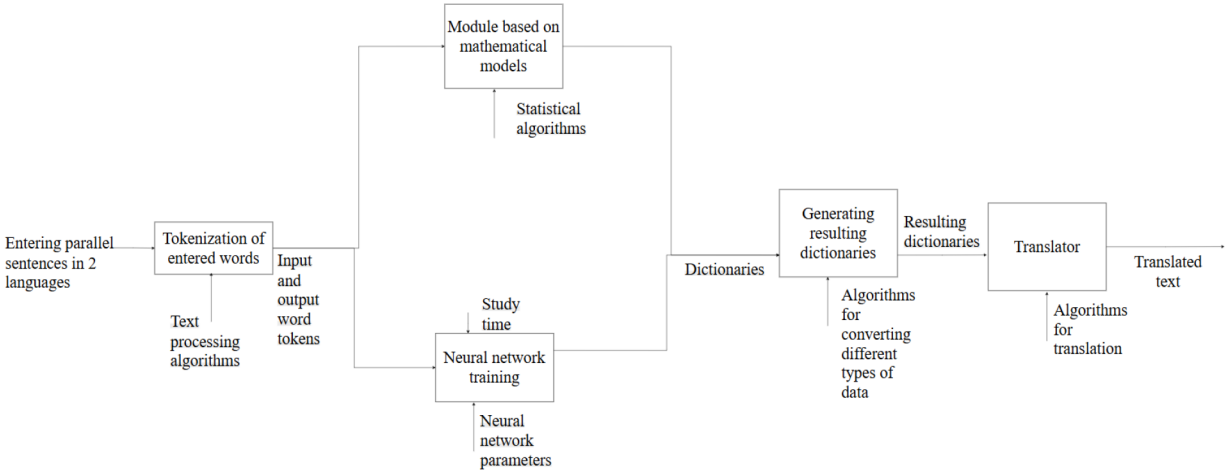


Fig. 4. Intelligent information system for translating artificial languages

For the study, a corpus of sentences on various topics was compiled: everyday dialogues, scientific descriptions, and literary excerpts. Each sentence was translated from Ukrainian into English, French, and Esperanto, as well as in the reverse direction. As a result, a complete set of texts for comparison was obtained.

The METEOR methodology was chosen for evaluation, which takes into account the accuracy of the translation of words, their morphemes and permutations, which is suitable for a complete evaluation of both translation options. The result was reduced to a single value in the range from 0 to 1, where 1 is the ideal option with which the automatic translation is compared (human translation). Each incorrect position, permutation, synonym, or incorrect morpheme

reduces the score by a certain coefficient. After that, this value was converted into a percentage for convenience. That is, the result will be the percentage of coincidence with the reference translation.

The quality of the translation was checked in pairs of languages. These include English, Ukrainian (local dialect), French and Esperanto.

The tables show the results of the experiments. The row indicates the source language, the column indicates the target language. The number of the experiment is indicated in brackets.

The results of the method of translating artificial languages based on mathematical models are given in (Table 1).

Table 1

Results of the method of translating artificial languages based on mathematical models

	English	Ukrainian	French	Esperanto
English	–	89.4 % (1)	92.9 % (3)	93.5 % (5)
Ukrainian	90.9 % (2)	–	88.7 % (7)	91.1 % (9)
French	92.7 % (4)	89.2 % (8)	–	89.9 % (11)
Esperanto	92.1 % (6)	92.7 % (10)	89.0 % (12)	–

The results show a fairly consistent high level of translation quality across all languages considered, which indicates the consistency and reliability of the mathematical model used. The system works best in directions involving English: from English to other languages (92.9–93.5 %) and from Ukrainian to English (90.9 %). Slightly lower values are observed for translations where English is not involved, in particular between French and Ukrainian (89.2 %) or French and Esperanto (89.0 %). This may indicate more complex interactions between these languages and fewer indirect connections in the model.

Overall, the indicators remain within the range of 88–94 %, which characterises the system as stable, with high translation quality and minor fluctuations depending on the language pair.

Results of the neural network-based artificial language translation method in (Table 2).

Table 2

Results of artificial language translation based on the neural network method

	English	Ukrainian	French	Esperanto
English	–	89.7 % (1)	92.2 % (3)	92.8 % (5)
Ukrainian	90.3 % (2)	–	89.6 % (7)	90.5 % (9)
French	91.1 % (4)	89.2 % (8)	–	92.9 % (11)
Esperanto	92.7 % (6)	92.1 % (10)	93.0 % (12)	–

In the case of the neural network method, the results also remain high and balanced. The best results are observed for translations from Esperanto into other languages, as well as between English and French. The values for Ukrainian-French are slightly lower, but the difference is insignificant. Overall, the system demonstrates stable quality with a tendency to improve in pair combinations directly with Esperanto.

After combining the methods by combining different types of data with certain coefficients, the following results were obtained (Table 3).

Table 3

Results of the intellectual information system for translating artificial languages

	English	Ukrainian	French	Esperanto
English	–	89.6 % (1)	93.0 % (3)	93.8 % (5)
Ukrainian	90.1 % (2)	–	89.6 % (7)	90.7 % (9)
French	92.9 % (4)	88.8 % (8)	–	91.9 % (11)
Esperanto	92.7 % (6)	92.8 % (10)	91.2 % (12)	–

The results of comparing all methods can be seen in (Fig. 5).

As we can see from these results, on average, the combined result shows the best performance. The maximum accuracy achieved was 92.9 %. In some pairs, the result was worse than the separate algorithm, but this was due to the coefficients used when combining the results. Since each individual language pair can be configured separately, the correct result will appear in the worst case with the same accuracy as the best of the two results. However, this requires separate configuration.

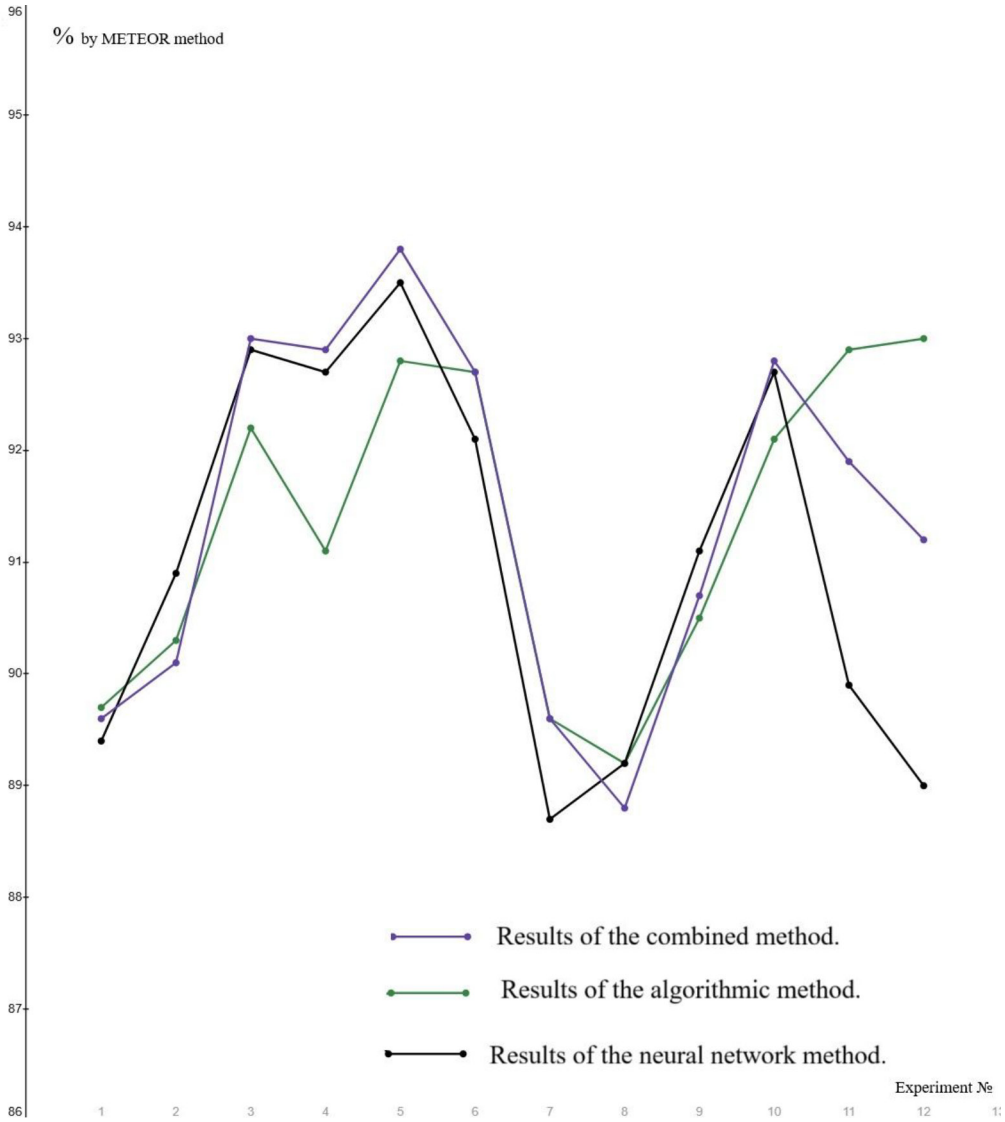


Fig. 5. results of comparison of all methods in the form of a graph

Based on the results obtained, we can conclude that the combination of these two methods allows us to create the best translation method for artificial languages.

Conclusions

The study was the first to evaluate the accuracy of artificial language translation methods in the proposed intellectual information system for translating artificial languages. The study achieved its objectives and solved the tasks set.

A detailed study and comparison of translation quality assessment methods was conducted. This made it possible to determine which translation quality assessment methods are best suited for artificial languages.

All modules of the system were studied, namely neural network translation, algorithmic translation, and combined translation.

As a result, we found that all methods work on all types of selected languages and show results of up to 92.9 % translation accuracy on a randomly selected dataset.

Based on the results obtained, we can conclude that the combination of these two methods allows us to create the best translation method for artificial languages. However, in some cases, the combination may worsen the result of a single method.

The scientific significance of the work lies in deepening the understanding of the mechanisms underlying neural networks and algorithmic methods. The conclusions presented can serve as a basis for further research in the field of machine learning and the development of intelligent systems, which will contribute to progress in the field of artificial intelligence and machine translation.

The practical value of this research lies in the creation of an intelligent translation information system that is independent of the source languages. This will allow automatic translation to be used for common, artificial and slang languages without additional costs.

Bibliography

1. House, Juliane. Translation quality assessment: Past and present. Routledge, 2014.
2. Han, Lifeng, Gareth JF Jones, and Alan F. Smeaton. “Translation quality assessment: A brief survey on manual and automatic meth-ods”. arXiv preprint arXiv:2105.03311 (2021).
3. Rivera-Trigueros, Irene. “Machine translation systems and quality assessment: a systematic review”. Language Resources and Evaluation 56.2 (2022): 593–619.
4. Papineni, Kishore, et al. “Bleu: a method for automatic evaluation of machine translation”. Proceedings of the 40th annual meeting of the Asso-ciation for Computational Linguistics. 2002.
5. Popović, Maja. “chrF: character n-gram F-score for automatic MT evaluation”. Proceedings of the tenth workshop on statistical machine transla-tion. 2015.
6. Popović, Maja. “chrF++: words helping character n-grams”. Proceedings of the second conference on machine translation. 2017.
7. Lavie, Alon, and Michael J. Denkowski. “The METEOR metric for automatic evaluation of machine translation”. Machine translation 23.2 (2009): 105–115.
8. Agarwal, Abhaya, and Alon Lavie. “Meteor, m-bleu and m-ter: Evaluation metrics for high-correlation with human rankings of machine translation output”. Proceedings of the third workshop on statistical machine translation. 2008.
9. Mukherjee, Aniruddha, et al. “A Detailed Comparative Analysis of Automatic Neural Metrics for Machine Translation: BLEURT & BERTScore”. IEEE Open Journal of the Computer Society (2025).
10. Rei, Ricardo, et al. “COMET: A neural framework for MT eval-uation”. arXiv preprint arXiv:2009.09025 (2020).
11. Гаврашенко, А.., Барковська, О.. (2023). Аналіз алгоритмів аугментації тексту в системах машинного перекладу штучних мов. *Сучасні інформаційні системи*, *7*(1), 47–53. https://doi.org/10.20998/2522-9052.2023.1.08
12. Барковська, О., Гаврашенко, А.. (2023). АНАЛІЗ АЛГОРИТМІВ ПОШУКУ СЛІВ У СЛОВНИКАХ СИСТЕМ МАШИННОГО ПЕРЕКЛАДУ ДЛЯ ШТУЧНИХ МОВ. Комп’ютерні системи та інформаційні технології, (2), 17–24. https://doi.org/10.31891/csit-2023-2-2

References

1. House, Juliane. *Translation quality assessment: Past and present*. Routledge, 2014.
2. Han, Lifeng, & Gareth JF Jones, and Alan F. Smeaton. «Translation quality assessment: A brief survey on manual and automatic methods». *arXiv preprint arXiv:2105.03311* (2021).
3. Rivera-Trigueros, &Irene. «Machine translation systems and quality assessment: a systematic review». *Language Resources and Evaluation* 56.2 (2022): 593–619.
4. Papineni, Kishore, et al. «Bleu: a method for automatic evaluation of machine translation». *Proceedings of the 40th annual meeting of the Association for Computational Linguistics*. 2002.
5. Popović, Maja. «chrF: character n-gram F-score for automatic MT evaluation». *Proceedings of the tenth workshop on statistical machine translation*. 2015.
6. Popović, Maja. «chrF++: words helping character n-grams». *Proceedings of the second conference on machine translation*. 2017.
7. Lavie, Alon, &Michael J. Denkowski. «The METEOR metric for automatic evaluation of machine translation». *Machine translation* 23.2 (2009): 105–115.
8. Agarwal, Abhaya, &Alon Lavie. «Meteor, m-bleu and m-ter: Evaluation metrics for high-correlation with human rankings of machine translation output». *Proceedings of the third workshop on statistical machine translation*. 2008.
9. Mukherjee, Aniruddha, et al. «A Detailed Comparative Analysis of Automatic Neural Metrics for Machine Translation: BLEURT & BERTScore». *IEEE Open Journal of the Computer Society* (2025).
10. Rei, Ricardo, et al. «COMET: A neural framework for MT evaluation». *arXiv preprint arXiv:2009.09025* (2020).
11. Havrashenko, A.., & Barkovska, O.. (2023). ANALYSIS OF TEXT AUGMENTATION ALGORITHMS IN ARTIFICIAL LANGUAGE MACHINE TRANSLATION SYSTEMS. *Advanced Information Systems*, *7*(1), 47–53. https://doi.org/10.20998/2522-9052.2023.1.08
12. Barkovska, O., & Havrashenko, A. (2023). ANALYSIS OF WORD SEARCH ALGORITHMS IN THE DICTIONARIES OF MACHINE TRANSLATION SYSTEMS FOR ARTIFICIAL LANGUAGES. Computer Systems and Information Technologies, (2), 17–24. https://doi.org/10.31891/csit-2023-2-2

Дата першого надходження рукопису до видання: 30.09.2025
Дата прийнятого до друку рукопису після рецензування: 28.10.2025
Дата публікації:

УДК 004.056:004.94 DOI

О. В. ГЕРАСІНА
кандидат технічних наук, доцент,
доцент кафедри безпеки інформації та телекомунікацій
Національний технічний університет «Дніпровська політехніка»
ORCID: 0000-0002-8196-0657

МОДЕЛЮВАННЯ ТРАФІКУ ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНИХ МЕРЕЖ ІЗ ВИКОРИСТАННЯМ БЛОЧНО-ОРІЄНТОВАНИХ СТРУКТУР ДЛЯ СИСТЕМ ВИЯВЛЕННЯ АТАК

У роботі встановлено, що актуальним завданням є побудова адекватних прогнозуючих моделей мережевого самоподібного трафіку, які б дозволили використовувати їх в системах виявлення вторгнень для детектування мережевих аномалій в режимі реального часу з достатньою ефективністю відносно похибок і достовірності та підвищеною оперативністю. При цьому прогнозування мережевого самоподібного трафіку дозволяє підвищити оперативність виявлення атак. Запропоновано для прогнозування трафіку інформаційно-комунікаційних мереж використовувати інтелектуальні адаптивні фільтри-апроксиматори, які побудовані на основі блочно-орієнтованих моделей Вінера, Гамерштейна та їх комбінацій. Моделювання мережевого трафіку із використанням блочно-орієнтованих структур проводилося в середовищі Matlab на основі експериментальних даних – трафіку, що передається через Інтернет. Як критерій оцінки моделей було використано критерій незміщенності (мінімуму зсуву). Як типи структур розглядались моделі Вінера, Гамерштейна, Гамерштейна-Вінера і авторегресії із додатковим вхідним сигналом. Як базисні функції використовувались вейвнет, нейронні мережі прямого поширення, дерево рішень, лінійні функції, поліном Колмогорова-Габора, обмеження і зона нечутливості. В результаті моделювання було виявлено, що перевагу (у сенсі критерію мінімуму зсуву) мають блочно-орієнтовані моделі Гамерштейна-Вінера з інтелектуальними базисними функціями – сигмоїдальною нейронною мережею і вейвнет. Для інших досліджуваних блочно-орієнтованих моделей перевагу також мають інтелектуальні базисні функції. Адекватність отриманих моделей мережевого самоподібного трафіку експериментальним даним було перевірено та підтверджено за непараметричним критерієм знаків. Подальші дослідження мають бути спрямовані на обґрунтування та дослідження інформативності характеристик і моделей трафіку інформаційно-комунікаційних мереж, а також ефективності критеріїв та методів розпізнавання атак. ***Висновки.*** *Доведена ефективність використання блочно-орієнтованих структур та інтелектуальних базисних функцій шуканих моделей мережевого самоподібного трафіку при розв'язанні задачі його прогнозування.*

Ключові слова: *моделювання, прогнозування, самоподібний трафік, інформаційно-комунікаційна мережа, модель Вінера, модель Гамерштейна, виявлення атак.*

O. V. GERASINA
Candidate of Technical Sciences, Associate Professor,
Associate Professor at the Department of Information Security
and Telecommunications
Dnipro University of Technology
ORCID: 0000-0002-8196-0657

MODELING OF INFORMATION AND COMMUNICATION NETWORK TRAFFIC USING BLOCK-ORIENTED STRUCTURES FOR ATTACK DETECTION SYSTEMS

It was established that the current task is to build adequate predictive models of network self-similar traffic, which would allow their use in IDS for detecting network anomalies in real time with sufficient efficiency in terms of errors and reliability and increased efficiency. At the same time, predicting network self-similar traffic allows to increase the efficiency of attack detection. It is proposed to use intelligent adaptive approximator filters for predicting information and communication network traffic, which are built on the basis of block-oriented Wiener, Hammerstein models and their combinations. Network traffic modeling using block-oriented structures was carried out in the Matlab environment based on experimental data – traffic transmitted via the Internet. The criterion of unbiasedness (minimum shift) was used as a criterion for evaluating models. The types of structures considered were Wiener, Hammerstein, Hammerstein-Wiener models and autoregression with an additional input signal. Wavenet, forward propagation neural networks, decision tree, linear functions, Kolmogorov-Gabor polynomial, saturation and dead zone were used as basic functions. As a result of the modeling, it was found that the advantage (in the sense of the minimum shift criterion) is given to block-oriented Hammerstein-Wiener models with intelligent basis functions – a sigmoidal neural network and a wavenet. For

© Герасіна О. В., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

other studied block-oriented models, intelligent basis functions are also preferred. The adequacy of the obtained models of network self-similar traffic to experimental data was checked and confirmed by the non-parametric sign criterion. Further research should be aimed at substantiating and studying the informativeness of the characteristics and traffic models of information and communication networks, as well as the effectiveness of the criteria and methods for recognizing attacks. ***Conclusions.*** *The effectiveness of using block-oriented structures and intelligent basis functions of the sought-after models of network self-similar traffic in solving the problem of its prediction is proven.*

Key words: *modeling, prediction, self-similar traffic, information and communication network, Wiener model, Hammerstein model, detection of attacks.*

Постановка проблеми

Стрімкий розвиток інформаційних технологій та інформаційно-комунікаційних систем і мереж (ІКМ) викликає ряд безпрецедентних загроз зі сторони окремих осіб, організацій або країн, які прагнуть дестабілізувати суспільне життя, втручаючись в роботу критично важливих об'єктів інфраструктури. Тому актуальним рішенням зазначеної проблеми є використання засобів моніторингу, здатних аналізувати трафік мережі в режимі реального часу. До таких засобів відносяться системи виявлення та запобігання атак (СВА) [1–4].

Аналіз останніх досліджень і публікацій

Основним завданням СВА є оперативне виявлення вторгнень та запровадження ефективного захисного сценарію щодо припинення факту порушення конфіденційності, доступності та цілісності інформаційних ресурсів та сервісів [1–4].

Сучасні СВА прийнято розділяти на два типи: спрямовані на пошук зловживань та на виявлення аномалій у системі. Виявлення зловживань ґрунтується на формуванні шаблонів вторгнень, що не є ефективним при детектуванні невідомих атак. Для реєстрації невідомих атак в ІКМ використовують системи виявлення аномалій, в яких певні дії, що є відмінними від поведінки в нормальному стані, ідентифікуються як аномальні. При виявленні мережевих аномалій даними для аналізу є мережевий трафік. Створений набір ознак (характеристики трафіку) порівнюється з набором ознак нормальної діяльності системи або конкретних користувачів, і якщо спостерігається суттєва розбіжність, фіксується мережева аномалія. При цьому, набір ознак нормальної діяльності системи або конкретних користувачів, СВА на основі аномалій повинні накопичити перед початком використання, і постійно оновлювати його із урахуванням поточного спостережуваного профілю мережевої активності [5–8].

Трафік, що виникає в сучасних ІКМ, є нелінійним стохастичним процесом з властивостями самоподоби, з хаотичною і фрактальною динамікою [9, 10]. Оцінка характеристик мережевого трафіку необхідна для побудови його адекватної моделі, що дозволяє сформувати еталонну модель (шаблон) «нормального» трафіку і за нею виявляти аномалії трафіку в СВА. При цьому прогнозування мережевого трафіку, яке дозволяє підвищити оперативність виявлення атак, доцільно проводити із використанням адаптивних фільтрів-апроксиматорів (АФА), побудованих на основі систем штучного інтелекту (нейронних мереж (НМ), систем з нечіткою логікою) [9].

У роботі [11] запропоновано методику прогнозування трафіку в ІКМ, яка складається з двох етапів: фрактальний аналіз трафіку та його прогнозування із використанням адаптивних фільтрів-апроксиматорів (АФА). При цьому, на другому етапі здійснюється обґрунтування методів і критерію структурної (глобальної) оптимізації, типу структури моделі, базисних функцій та методів параметричної оптимізації, а також структурна та параметрична оптимізація. Але у роботі [11] відсутнє обґрунтування типу структур моделі та базисних функцій (типу АФА) для подальшого проведення прогнозування мережевого трафіку. Також у роботах [9, 12-14] наведено використання блочно-орієнтованих моделей для вирішення різних завдань, але відсутні роботи, де виконується обґрунтування їх типів.

Таким чином, невирішеною задачею є обґрунтування типу структур та базисних функцій моделей для прогнозування самоподібного трафіку ІКМ для виявлення його аномалій при використанні в СВА.

Формулювання мети дослідження

Дослідження та обґрунтування типів структур та базисних функцій прогнозуючих моделей мережевого самоподібного трафіку для виявлення його аномалій при використанні в системах виявлення та запобігання атак.

Викладення основного матеріалу дослідження

Однією з основних проблем при вирішенні задач ідентифікації та прогнозування складних систем є вибір типів структур моделей. Правильний вибір структури моделі багато у чому визначає успіх побудови адекватної моделі системи. Зазвичай структуру визначають, виходячи з фізичних законів, що реалізуються у системі. Однак, така модель найчастіше має високу розмірність, що ускладнює її практичне використання.

Більш продуктивним є визначення структур моделей складних систем на основі спостережень їх вхідних і вихідних змінних з використанням апроксимацій у вигляді функціональних рядів Вольтера, поліномів Колмогорова-Габора, блочно-орієнтованих моделей тощо [9, 12].

У блочно-орієнтованих моделях мережевий трафік представляється шляхом композиції лінійного динамічного (ЛДБ) і нелінійного статичного (НСБ) блоків, наприклад, у вигляді моделей Вінера, Гамерштейна або їх комбінацій [12, 13].

Модель Вінера містить послідовно з'єднані нелінійний статичний і лінійний динамічний блоки (рис. 1, *а*), а у моделі Гамерштейна, навпаки, нелінійний блок приведений до виходу (рис. 1, *б*). На малюнках ЛДБ – лінії затримки *T*, величини яких (глибини пам'яті) визначаються розмірністю вхідних d_i і вихідних $d_{\hat{y}}$ змінних, а у якості НСБ можуть використовуватися як традиційні засоби: поліноми Лежандра або Колмогорова-Габора, так і інтелектуальні – нейронні мережі, системи з нечіткою логікою тощо. Входом моделей є вектори спостережень $\{z_i[k]\} \in \hat{Z}[k]$ для моментів часу *k*, а виходом – його прогноз на *n* тактів ($\hat{Y}[k+n]$).

Модель Вінера-Гамерштейна утворюється шляхом об'єднання блоків моделей Вінера і Гамерштейна, для чого вводиться зворотний зв'язок (рис. 2, *а*). Модель Гамерштейна-Вінера утворюється послідовним з'єднанням моделей Гамерштейна і Вінера. При цьому у центрі з'єднання два лінійних динамічних блоки об'єднуються в один (рис. 2, *б*).

Також до блочно-орієнтованих моделей крім моделей Вінера, Гамерштейна та їх комбінацій (див. рис. 1–2) відносяться моделі авторегресії із додатковим вхідним сигналом – ARX-моделі, які описують нелінійні структури, використовуючи паралельну комбінацію нелінійних і лінійних блоків [12].

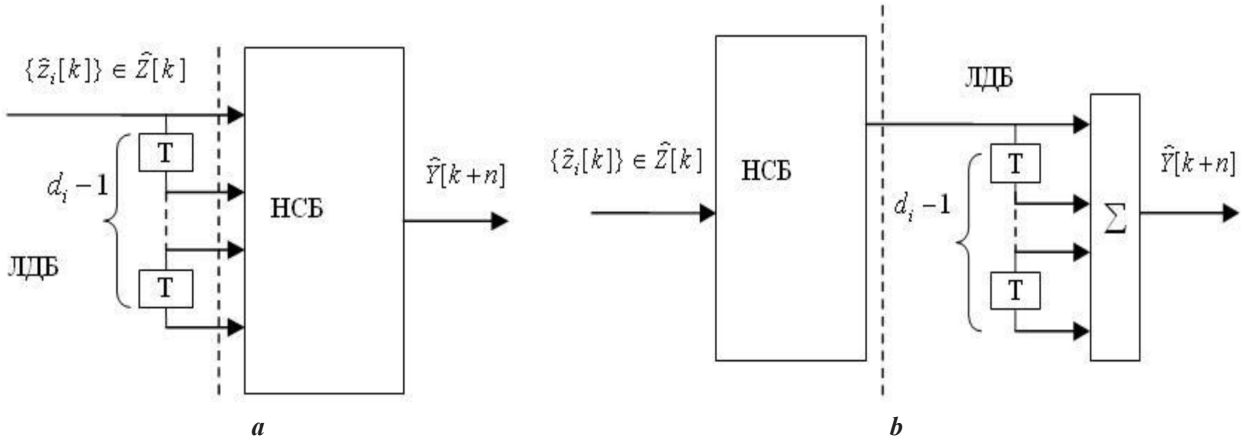


Рис. 1. Структура Вінера (*а*) і Гамерштейна (*б*) моделей мережевого трафіка

Обґрунтування блочно-орієнтованих структур моделей мережевого трафіку є складовою частиною методики інтелектуального прогнозування самоподібного трафіку ІКМ, яка дозволяє підвищити ймовірність визначення вторгнень для СВА за рахунок зниження похибок моделей самоподібного трафіка [11].

Моделювання задачі обґрунтування блочно-орієнтованих структур та базисних функцій моделей мережевого самоподібного трафіка виконувалось за допомогою стандартних та розроблених програм у середовищі Matlab / Simulink [14].

Як експериментальні дані було взято трафік, що передається через мережу Інтернет [15]. Дані являють собою залежність розміру Ethernet кадрів в байтах від часу. Для їх нормування по часової осі була проведена процедура агрегації з кроком 5 с.

З особливостей цього процесу глибина прогнозу була прийнята 3 такти, а глибина пам'яті – 4.

Як критерій оцінки моделей обрано критерій незміщенності (мінімуму зсуву), який не чутливий до рівня шуму у вхідних даних і при збільшенні завад їх мінімум не зміщується в область простіших моделей [16].

$$C_{см} = \frac{\left\| \hat{Y}_A[m+n] - \hat{Y}_B[m+n] \right\|}{\left\| Y^*[m+n] \right\|}. \quad (1)$$

де $\hat{Y}_A[m+n]$ і $\hat{Y}_B[m+n]$ – виходи моделей, які навчені на вибірках *A* і *B*, відповідно.

Як типи структур розглядались моделі Вінера, Гамерштейна, Гамерштейна-Вінера і ARX. При цьому, як НСБ (базисні функції) для моделі ARX використовувались вейвнет (НМ з функцією активації у вигляді вейвлет), каскадна НМ прямого поширення з трикутною функцією активації прихованого шару, сигмоїдальна НМ (НМ прямого поширення з логістичною сигмоїдальною функцією активації прихованого шару), дерево рішень і лінійна функція, а для моделей Вінера, Гамерштейна і Гамерштейна-Вінера – вейвнет, поліном Колмогорова-Габора, сигмоїдальна НМ, кусково-лінійна функція, обмеження (Saturation – одна з нелінійностей; сигнал на виході блока дорівнює вхідному сигналу доти, поки не досягне заданих порогів обмеження: верхнього або нижнього; після цього сигнал перестає змінюватися) і зона нечутливості (DeadZone – одна з нелінійностей, що являє собою лінійну залежність вихідного сигналу від вхідного скрізь, за винятком області мертвої зони).

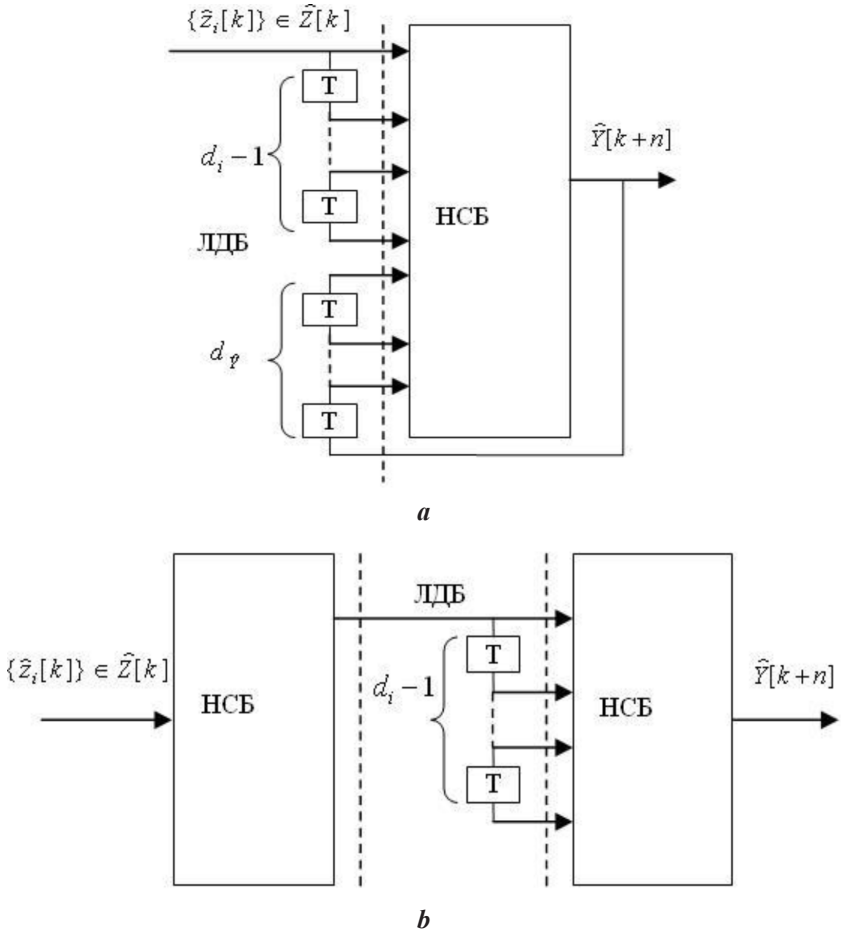


Рис. 2. Структура Вінера-Гамерштейна (*а*) і Гамерштейна-Вінера (*б*) моделей мережевого трафіка

Результати моделювання мережевого трафіку із використанням блочно-орієнтованої структури ARX наведено на рис. 3, *а*. При цьому значення критерію (1) склало 16,13 % для вейвнету, 15,86 % для каскадної НМ прямого поширення, 19,55 % для сигмоїдальної НМ, 20,05 % для дерева рішень і 21,19 % для лінійної функції.

Результати моделювання мережевого трафіку із використанням блочно-орієнтованої структури Гамерштейна-Вінера наведено на рис. 3, *б*. При цьому значення критерію (1) склало 3,70 % для вейвнету, 20,59 % для поліному Колмогорова-Габора, 1,77 % для сигмоїдальної НМ, 4,83 % для кусково-лінійної функції, 5,46 % для обмеження і 4,19 % для зони нечутливості.

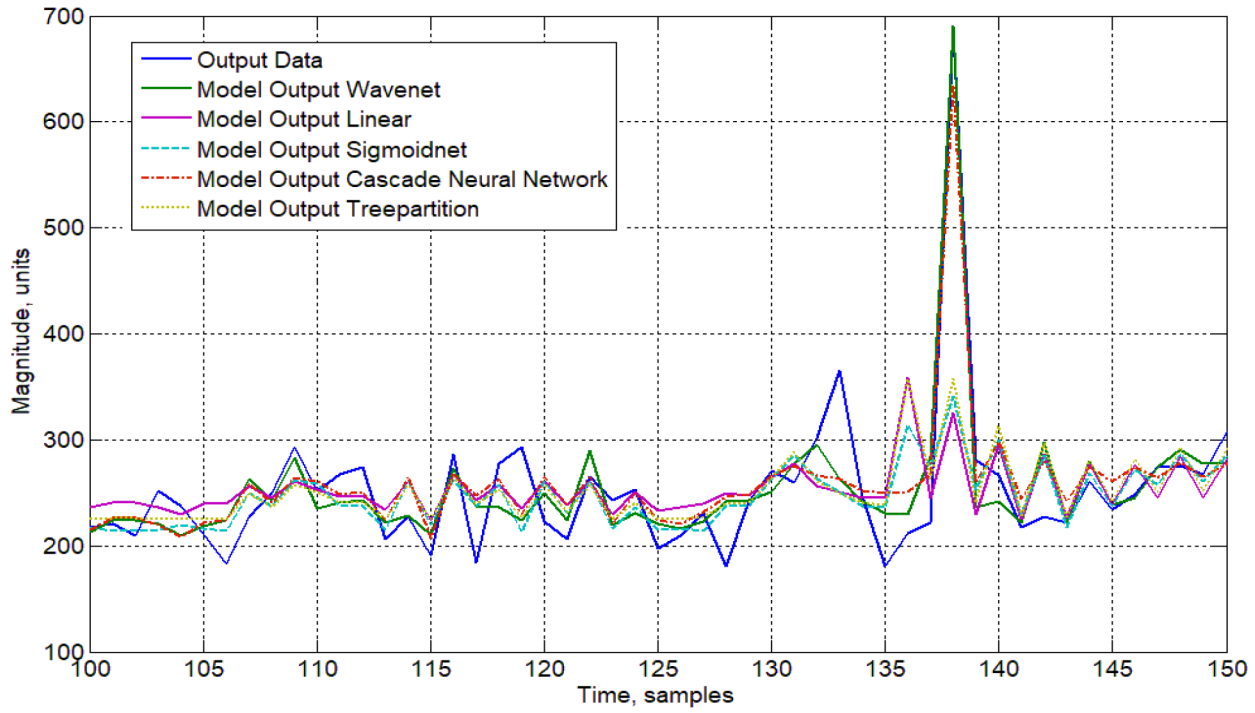
Результати моделювання мережевого трафіку із використанням блочно-орієнтованої структури Вінера наведено на рис. 4, *а*. При цьому значення критерію (1) склало 6,59 % для вейвнету, 8,36 % для поліному Колмогорова-Габора, 3,67 % для сигмоїдальної НМ, 10,73 % для кусково-лінійної функції, 9,22 % для обмеження і 8,36 % для зони нечутливості.

Результати моделювання мережевого трафіку із використанням блочно-орієнтованої структури Гамерштейна наведено на рис. 4, *б*. При цьому значення критерію (1) склало 7,17 % для вейвнету, 8,01 % для поліному Колмогорова-Габора, 7,18 % для сигмоїдальної НМ, 7,35 % для кусково-лінійної функції, 9,48 % для обмеження і 7,21 % для зони нечутливості.

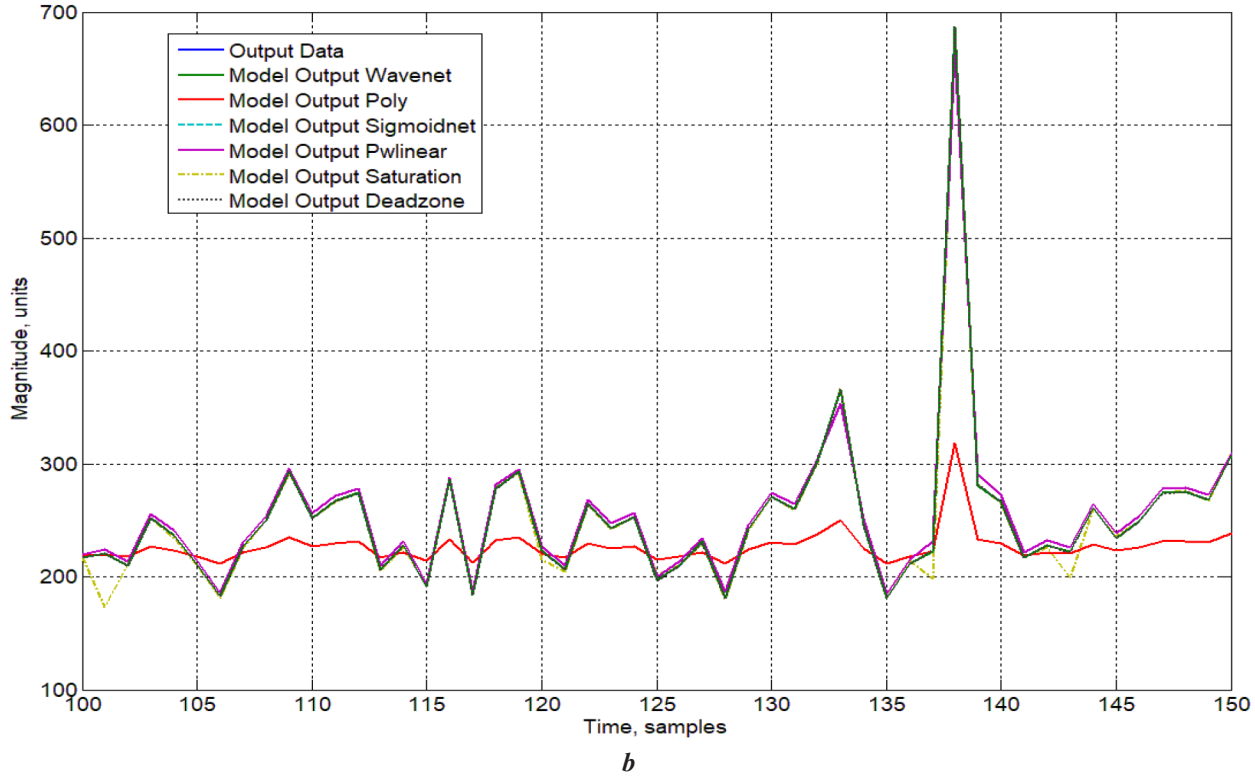
Встановлено (див. рис. 3–4), що перевагу (у сенсі критерію мінімуму зсуву (1)) мають блочно-орієнтовані моделі Гамерштейна-Вінера з інтелектуальними базисними функціями – сигмоїдальною НМ і вейвнет. Для інших моделей перевагу також мають інтелектуальні базисні функції (для Вінера, Гамерштейна – сигмоїдальна НМ і вейвнет, для ARX – сигмоїдальна НМ, каскадна НМ і вейвнет).

Час обчислень на комп'ютері з процесором Pentium IV за моделями Вінера, Гамерштейна і Гамерштейна-Вінера становить 8–11 мс на цикл прогнозу, а по ARX – 0,3 мс, що не вносить часових обмежень на застосування цих моделей для прогнозування трафіка в ІКМ.

Адекватність отриманих моделей мережевого трафіку перевірялася за непараметричним критерієм знаків. Було встановлено, що для рівня значущості 0,01 розроблені моделі адекватні експериментальним реалізаціям.



a



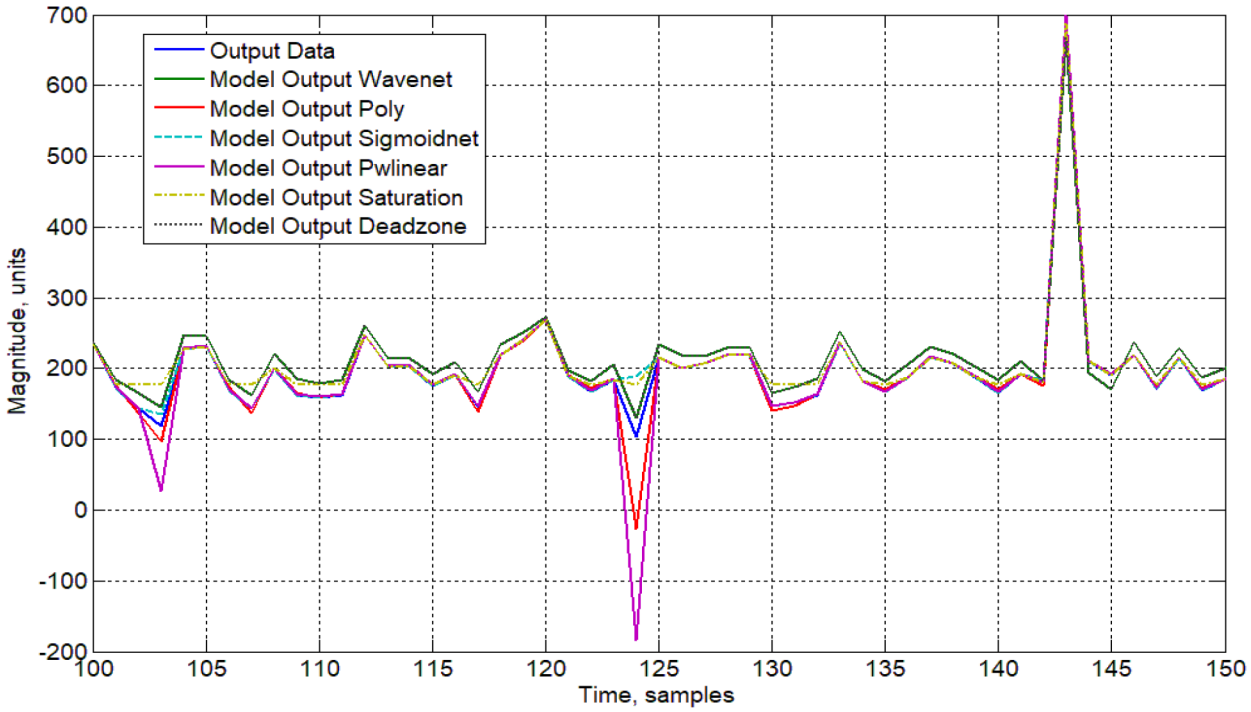
b

Рис. 3. Результати моделювання мережевого трафіку із використанням структури ARX (*а*) і Гамерштейна-Вінера (*б*)

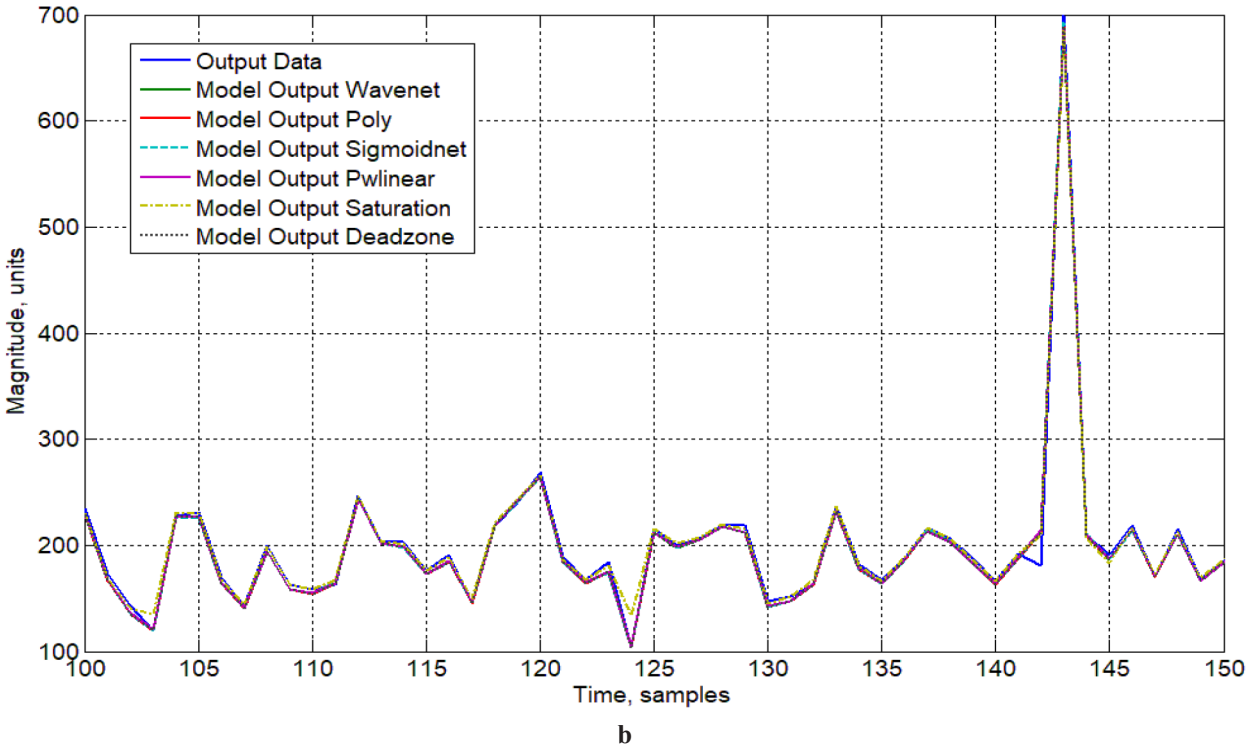
Висновки

Запропоновано для прогнозування самоподібного трафіку в інформаційно-комунікаційних мережах використовувати інтелектуальні адаптивні фільтри-апроксиматори, які побудовані на основі блочно-орієнтованих моделей Вінера, Гамерштейна та їх комбінацій.

Шляхом моделювання на основі експериментальних даних показана ефективність використання блочно-орієнтованих структур та інтелектуальних базисних функцій шуканих моделей мережевого трафіку при розв'язанні



a



b

Рис. 4. Результати моделювання мережевого трафіку із використанням структури Вінера (*а*) і Гамерштейна (*б*)

задачі його прогнозування. Перевірена та підтверджена адекватність отриманих моделей трафіку ІКМ експериментальним даним.

Подальші дослідження мають бути спрямовані на обґрунтування та дослідження інформативності характеристик і моделей мережевого трафіку, а також ефективності критеріїв та методів розпізнавання атак.

Список використаної літератури

1. Толюпа С. Лукова-Чуйко Н., Шестак Я. Засоби виявлення кібернетичних атак на інформаційні системи. *Інфокомунікаційні технології та електронна інженерія*. 2021. № 2 (2). С. 19–31.
2. Лукова-Чуйко Н., Наконечний В., Толюпа С., Зюбіна Р. Проблеми захисту критично важливих об'єктів інфраструктури. *Безпека інформаційних систем і технологій*. 2020. № 1 (2). С. 31–39.
3. Толюпа С., Кулько А. Нейро-нечітка система виявлення вторгнень у інформаційну мережу критичної інфраструктури. *Кібербезпека: освіта, наука, техніка*. 2025. № 3 (27). С. 233–247.
4. Носенко К. М., Півторак О. І., Ліхоузова Т. А. Огляд систем виявлення атак в мережевому трафіку. *Адаптивні системи автоматичного управління*. 2014. № 1(24). С. 67–75.
5. Довбешко С. В., Толюпа С. В., Шестак Я. В. Застосування методів інтелектуального аналізу даних для побудови систем виявлення атак. *Сучасний захист інформації*. 2019. № 1 (37). С. 6–15.
6. Лазаренко С. В. Особливості функціонування систем виявлення атак на автоматизовані системи. *Сучасний захист інформації*. 2015. № 1. С. 33–40.
7. Гулак Г. М., Семко В. В., Складанний П. М. Модель системи виявлення вторгнень з використанням двоступеневого критерію виявлення мережевих аномалій. *Сучасний захист інформації*. 2015. № 4. С. 81–85.
8. Лукова-Чуйко Н. В., Толюпа С. В., Пархоменко І. І. Методи виявлення вторгнень у сучасних системах IDS. *Безпека інформаційних систем і технологій*. 2021. № 1 (5). С. 19–26.
9. Корнієнко В. І., Гусєв О. Ю., Герасіна О. В. Інтелектуальне моделювання нелінійних динамічних процесів в системах керування, кібербезпеки, телекомунікацій: підручник. Дніпро : НТУ «ДП». 2020. 536 с.
10. Crovella M. E., Bestravos A. Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes. *IEEE Transactions on Networking*. 1997. Vol. 5. № 6. P. 835–846.
11. Герасіна О. В. Методика інтелектуальної ідентифікації та прогнозування трафіку в інформаційних телекомунікаційних мережах. *Системи обробки інформації*. 2018. № 1 (152). С. 94–99.
12. Pillonetto G., Chen T., Chiuso A., De Nicolao G., Ljung L. Regularized system identification: Learning dynamic models from data. *Springer Nature*. 2022. 377 p. DOI: 10.1007/978-3-030-95860-2
13. Nikolic S. S. et al. Identification of Nonlinear Systems Using the Hammerstein-Wiener Model with Improved Orthogonal Functions. *Elektronika ir Elektrotechnika*. 2023. № 29 (2). P. 4–11. DOI: 10.5755/j02.eie.33838
14. Dai T., Aljanaideh K., Chen R., Singh R., Stothert A., Ljung L. Deep Learning of Dynamic Systems using System Identification Toolbox. *IFAC PapersOnLine*. 2024. № 58 (15). P. 580–585. DOI: 10.13140/RG.2.2.29127.47527
15. Архів трафіку. URL: http://ita.ee.lbl.gov
16. Ivakhnenko A. G., Madala H. R. Inductive learning algorithms for complex systems modeling. *London, Tokyo: CRC Press*. 1994. 384 p.

References

1. Toliupa S., Lukova-Chuiko N., & Shestak Ya. (2021). Zasoby vyiavlennia kibernetychnykh atak na informatsiini systemy [Means of detecting cyber attacks on information systems]. *Infokomunikatsiini tekhnolohii ta elektronna inzheneriia*, no. 2(2), pp. 19–31.
2. Lukova-Chuiko N., Nakonechnyi V., Toliupa S., & Ziubina R. (2020). Problemy zachystu krytychno vazhlyvych ob`iektiv infrastruktury [Problems of protection of critical infrastructure facilities]. *Bezpeka informatsiinykh system i tekhnolohii*, no. 1(2), pp. 31–39.
3. Toliupa, S. & Kulko A. (2025). Neiro-nechitka systema vyiavlennia vtorhnen u informa-tsiinu merezhu krytychnoi infrastruktury [Neuro-fuzzy system for detecting intrusions into the information network of critical infrastructure]. *Kiberbezpeka: osvita, nauka, tekhnika*, no. 3(27), pp. 233–247.
4. Nosenko K. M., Pivtorak O. I., & Lichouzova T. A. (2014). Ohliad system vyiavlennia atak v merezhevomu trafiku [Overview of network traffic attack detection systems]. *Adaptyvni systemy avtomatychnoho upravlinnia*, no. 1(24), pp. 67–75.
5. Dovbeshko S. V., Toliupa S. V., & Shestak Ya. V. (2019). Zastosuvannia metodiv intelektualnoho analizu danykh dlia pobudovy system vyiavlennia atak [Application of data mining methods to build attack detection systems]. *Suchasnyi zakhyst informatsii*, no. 1(37), pp. 6–15.
6. Lazarenko S. V. (2015). Osoblyvosti funktsionuvannia system vyiavlennia atak na avtomatyzovani systemy [Features of functioning of systems of detection of attacks on automated systems]. *Suchasnyi zakhyst informatsii*, no. 1, pp. 33–40.
7. Hulak H. M., Semko V. V., & Skladannyi P. M. (2015). Model systemy vyiavlennia vtorhnen z vykorystanniam dvostupenevoho kryteriiu vyiavlennia merezhevykh anomalii [Model of the system for detecting intrusion based on the two-stage criterion for detecting fencing anomalies]. *Suchasnyi zakhyst informatsii*, no. 4, pp. 81–85.
8. Lukova-Chuiko N. V., Toliupa S. V., & Parkhomenko I. I. (2021). Metody vyiavlennia vtorhnen u suchasnykh systemakh IDS [Intrusion detection methods in modern IDS systems]. *Bezpeka informatsiinykh system i tekhnolohii*, no. 1(5), pp. 19–26.

9. Korniienko V. I., Gusiev O. Yu., & Gerasina O. V. (2020). Intelektualne modeliuvannia neliniinykh dynamichnykh protsesiv u systemakh keruvannia, kiberbezpeky, telekomunikatsii: pidruchnyk [Intelligent modeling of nonlinear dynamic processes in control systems, cybersecurity, telecommunications: a textbook]. Dnipro: NTU "DP".

10. Crovella M. E., & Bestravos A. (1997). Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes. *IEEE Transactions on Networking*, no. 5(6), pp. 835–846.

11. Gerasina O. V. (2018). Metodyka intelektualnoi identyfikatsii ta prohnozuvannia trafiku v informatsiinykh telekomunikatsiinykh merezhakh [Methodology of intelligent identification and traffic forecasting in information telecommunication networks]. *Systemy obrobky informatsii*, no. 1(152), pp. 94–99.

12. Pillonetto G., Chen T., Chiuso A., De Nicolao G., & Ljung L. (2022). Regularized system identification: Learning dynamic models from data. *Springer Nature*. DOI: 10.1007/978-3-030-95860-2

13. Nikolic S. S. et al. (2023). Identification of Nonlinear Systems Using the Hammerstein-Wiener Model with Improved Orthogonal Functions. *Elektronika ir Elektrotechnika*, no. 29 (2), pp. 4–11. DOI: 10.5755/j02.eie.33838

14. Dai T., Aljanaideh K., Chen R., Singh R., Stothert A., & Ljung L. (2024). Deep Learning of Dynamic Systems using System Identification Toolbox. *IFAC PapersOnLine*, no. 58 (15), pp. 580–585. DOI: 10.13140/RG.2.2.29127.47527

15. Traffic archive [Traffic archive]. URL: http://ita.ee.lbl.gov

16. Ivakhnenko A. G., & Madala H. R. (1994). Inductive learning algorithms for complex systems modeling. London, Tokyo: CRC Press.

Дата першого надходження рукопису до видання: 24.09.2025
Дата прийнятого до друку рукопису після рецензування: 20.10.2025
Дата публікації:

УДК 004.7:004.9 DOI

О. А. ГОЛЕНКО
студент кафедри інформатики і комп'ютерних наук
Херсонський національний технічний університет
ORCID: 0009-0003-4818-6406

М. В. КАРАМУШКА
доцент кафедри інформатики і комп'ютерних наук
Херсонський національний технічний університет
ORCID: 0000-0001-5982-4598

ДОСЛІДЖЕННЯ МОДЕРНІЗАЦІЇ ТА ВПРОВАДЖЕННЯ ІНТЕРНЕТ ТЕХНОЛОГІЙ І СИСТЕМ МОНІТОРИНГУ ТА УПРАВЛІННЯ У ЖИТЛОВИХ ПРИМІЩЕННЯХ

Стаття присвячена модернізації та впровадженню систем Інтернету речей для моніторингу і керування у житлових приміщеннях, де одночасно працюють спадкові дротові та радіорішення на кшталт RS-485 з протоколом Modbus, інфрачервоні та радіоканали на частотах чотириста тридцять три і вісімсот шістдесят вісім мегагерц, а також сучасні пристрої з передаванням даних через Wi-Fi, Bluetooth Low Energy, Zigbee і Thread. Мета роботи полягає у формуванні методичних засад інтеграції, що забезпечують сумісність, локальний контроль, приватність і роботу без мережі Інтернет, без прив'язки до конкретного виробника. Запропоновано еталонну архітектуру локально орієнтованого хаба на базі Raspberry Pi. Вона містить шар адаптерів до різних фізичних та прикладних протоколів, шар нормалізації що перетворює нативні повідомлення у канонічні події та команди з уніфікованими одиницями виміру і часовими мітками у форматі UTC ISO 8601, подієву шину MQTT з гарантіями доставки, реєстр пристроїв і їхніх можливостей, рушій правил, а також зовнішнє програмне API через REST та WebSocket. Сформовано канонічну модель даних і доменне API з поділом полів на основні та розширені. Наведено профілі для датчиків температури, руху, протікання, лічильників енергії та керованих пристроїв на кшталт вимикачів і розеток. Показано мапінг із Modbus, з Bluetooth Low Energy, з інфрачервоних і радіопристроїв, а також з IP-пристроїв. Описано онбординг і захист, зокрема використання токенів або сертифікатів, шифрування транспортного рівня, рольову авторизацію та ведення журналів. Подано механізми роботи офлайн, серед яких зберігання та пересилання після відновлення, дедуплікація за ідентифікатором пристрою і часовою міткою. Реалізовано прототип із брокером, бек-ендом на Python з фреймворком FastAPI, сховищем SQLite та файловим буфером, модулем правил або Node-RED і локальною веб-панеллю з можливістю інтеграції з Home Assistant. Запропоновано методику оцінювання, що враховує затримку виконання команд, надійність доставки, час відновлення після збоїв і енергетичний ефект у сценаріях клімату та освітлення. Практична цінність полягає у поетапній міграції від розрізнених рішень до уніфікованої платформи з підвищеною надійністю і зниженими витратами інтеграції. Окреслено подальший розвиток, зокрема підтримку шин BACnet і KNX через адаптери, автоматичне витягування можливостей пристроїв з описів Web of Things і Matter та інструменти валідації якості даних для аналітики і штучного інтелекту.

Ключові слова: *Інтернет речей, житлові приміщення, Raspberry Pi, протоколи, MQTT, безпека, офлайн-стійкість, інтеграція систем, уніфікація даних.*

O. A. GOLENKO
Student at the Department of Computer Science
Kherson National Technical University
ORCID: 0009-0003-4818-6406

M. V. KARAMUSHKA
Associate Professor at the Department of Computer Science
Kherson National Technical University
ORCID: 0000-0001-5982-4598

© Голенко О. А., Карамушка М. В. , 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

RESEARCH ON MODERNIZATION AND IMPLEMENTATION OF INTERNET TECHNOLOGIES AND MONITORING AND MANAGEMENT SYSTEMS IN RESIDENTIAL PREMISES

This article explores the modernization and deployment of Internet of Things (IoT) technologies for monitoring and control in residential environments where legacy infrastructures – such as RS-485 with Modbus, infrared channels, and sub-GHz radios at 433/868 MHz – coexist with modern devices relying on Wi-Fi, Bluetooth Low Energy, Zigbee, and Thread. The research aims to establish methodological foundations for integration that ensure interoperability, local control, privacy, and resilience to Internet disruptions while avoiding dependence on a single vendor. A reference architecture for a locally oriented hub based on Raspberry Pi is proposed. The architecture comprises an adapter layer for heterogeneous physical and application protocols, a normalization layer that transforms native messages into canonical events and commands with unified measurement units and UTC ISO 8601 timestamps, an MQTT event bus with delivery guarantees, a device and capability registry, a rules engine, and external interfaces via REST and WebSocket. A canonical data model and domain API are introduced, separating core and extended fields. Profiles are provided for temperature, motion, leak detection, and energy-metering sensors, as well as controllable devices such as switches and outlets. The study demonstrates protocol mapping from Modbus, BLE, infrared, and sub-GHz radios, along with IP-based devices. Security and onboarding procedures include provisioning tokens or certificates, encrypted transport, role-based access control, and event logging. Offline operation is supported through store-and-forward, deduplication, and recovery mechanisms. A functional prototype was implemented using a Mosquitto broker, a FastAPI back end with SQLite storage and file-based buffering, Node-RED for automation, and a local web dashboard with optional integration into Home Assistant. An evaluation methodology is proposed to measure command latency, reliability of delivery, recovery time after faults, and energy efficiency in climate-control and lighting scenarios. The practical contribution lies in enabling a stepwise migration from fragmented legacy solutions to a unified, vendor-independent platform that improves reliability and reduces integration costs. Future work will extend support for BACnet and KNX via adapters, leverage Web of Things and Matter descriptions for automatic capability extraction, and develop data-quality validation tools for analytics and AI applications.

Key words: *Internet of Things, residential environments, Raspberry Pi, protocols, MQTT, security, offline resilience, system integration, data unification.*

Постановка проблеми

Сучасні інформаційні технології стрімко змінюють підходи до облаштування житлових приміщень – від офісних будівель, шкіл, гуртожитків і багатоквартирних будинків до приватних апартаментів. Технології Інтернету речей (IoT) відкривають нові можливості для моніторингу та керування домашньою інфраструктурою [1]. Особливої уваги потребує розроблення інтегрованих IoT-систем, спроможних взаємодіяти з побутовими пристроями, датчиками, «розумними» замками, камерами, охоронними системами та лічильниками ресурсів для підвищення ефективності управління і зниження енерговитрат. Попри велику кількість індивідуальних продуктів на ринку «розумних» рішень, відсутність універсальної системи, здатної централізовано керувати всіма аспектами житла, зумовлює потребу в аналізі доступних підходів і подальших дослідженнях.

У житловому секторі триває інтенсивна модернізація: поряд із новими IoT-пристроями (сенсори мікроклімату, «розумні» вимикачі, лічильники з передаванням даних, голосові асистенти) продовжують функціонувати спадкові системи підтримки будинку – дротові датчики охорони, домофони, котли та контролери опалення, аналогові/імпульсні лічильники, ІЧ-керовані прилади, радіопристрої на 433-868 МГц, інтерфейси RS-485 або Modbus, а також KNX та BACnet. Така співіснування «старого» і «нового» супроводжується фрагментацією протоколів LAN, Wi-Fi, BLE, Zigbee, IR, власні радіоінтерфейси і невідповідністю прикладних моделей – від відкритих MQTT,HTTP та CoAP до закритих API та пропрієтарних форматів [2].

Наслідками є несумісність і «силоси» даних, різні одиниці виміру та часові мітки, дублювання шлюзів, зростання витрат інтеграції та експлуатації, ускладнене масштабування і залежність від постачальників/хмарних сервісів. Додатково житлові системи висувають підвищені вимоги до приватності, офлайн-стійкості (робота під час збоїв інтернету/живлення) і кібербезпеки локальної мережі. Відсутність єдиного підходу до уніфікації протоколів і семантики даних гальмує модернізацію: нові пристрої складно інтегрувати з наявною інфраструктурою, а «старі» – адаптувати під сучасні сценарії автоматизації та аналітики.

Отже, постає науково-прикладна задача: обґрунтувати методичні засади модернізації та впровадження IoT-систем у житлових приміщеннях, які забезпечують інтероперабельність «старих» і «нових» пристроїв. Необхідно визначити принципи уніфікації обміну даними (на рівні формату й семантики), правила мапінгу спадкових протоколів до єдиної моделі, а також вимоги до офлайн-стійкості, безпеки та керованого розширення. Це створить основу для поетапної інтеграції і подальшої аналітики/ШІ без прив'язки до конкретного вендора чи транспорту.

Аналіз останніх досліджень і публікацій

В сьогодення підходи до сумісності в «розумному» житлі рухаються у двох взаємо доповнювальних напрямах перший з яких уніфікація семантики пристроїв та другий уніфікація прикладного рівня поверх IP. Стандарти W3C Web of Things (WoT) 1.1 та CSA Matter 1.4 є показовими прикладами, перший пропонує незалежну від протоколів,

структуровану для автоматизованої обробки модель опису «речей», що знімає частину фрагментації на рівні описів та зв'язувань через HTTP/MQTT/CoAP також другий уніфікує класи пристроїв, атрибути й команди поверх IP та поступово розширює домени такі як енергоменеджмент, мережна інфраструктура, спрощений онбординг NFC і multi-device QR [3]. На транспортному рівні MQTT 5.0 лишається де-факто стандартом публікації/підписки, але не стандартизує семантику повідомлення, тож потрібна нормалізація на рівні хаба [4].

Різні дослідження наголошують на важливості інтеграції IoT для підвищення енергоефективності будинків і безпеки мешканців. Огляди й прикладні роботи показують ефективність в поєднанні IoT + AI а саме алгоритми прогнозування та оптимізації дозволяють автоматизувати рутинні завдання, підлаштовувати освітлення та клімат під графік присутності, зменшувати споживання енергії без втрати комфорту [5]. Продукти масового ринку на кшталт Google Home та Amazon Alexa дають зручний віддалений контроль, але їхній сценарний рушій традиційно був менш гнучким для індивідуальних потреб, але останнім часом Google розширює редактор автоматизацій, однак частина функцій ще не підтримується [6]. Дослідники також підкреслюють значення відкритої архітектури, стандартів як чинника зменшення витрат інтеграції та розширення функціональності – саме цю роль відіграє WoT як уніфікуюча ланка описів і зв'язувань між різними екосистемами. Водночас хмаро-центричні підходи несуть ризики приватності та витоків даних, що підтверджується систематичними оглядами безпеки та конфіденційності й резонансними кейсами залежності від сторонніх сервісів. Тому локальний контроль і можливість офлайн-роботи визнаються критично важливими для житлових систем [7].

На практиці попит на локально-орієнтовані рішення відбивають зрілі open-source платформи на кшталт Home Assistant або openHAB із фокусом на локальний контроль та приватність і розгортання на Raspberry Pi. Паралельно існують комерційні локальні хаби в основі Hubitat із акцентом на офлайновість і швидкі автоматизації. Ці системи доповнюють Matter та WoT, виступаючи «мостами» для спадкових технологій та уніфікуючи дані через хаб-рівень.

Для наочного порівняння наведено таблицю технологій, розглянутих вище, для впровадження та модернізації IoT у розумному приміщенні.

Таблиця 1

Порівняльна таблиця технологій

Технологій	Тип контролю	Off-line	Рівень кастомних автоматизацій	Приватність/ локальність	Розширюваність/ інтеграції	Коментар
Home Assistant	Локальний (з опціями хмари)	Так	Високий (GUI + YAML; скрипти)	"local control & privacy first"	3300+ інтеграцій	Підтримка локального голосу (Assist) та DIY-стеків
openHAB	Локальний (open-source)	Так	Високий (Rules DSL/JS; UI)	Локальний сервер/ захист	Широкі "add-ons"	Рекомендований для RPi; технологічно агностичний
Hubitat	Локальний (комерційний)	Так	Середньо-високий (Rule Machine)	Акцент на "Local, Private"	Zigbee/Z-Wave/ Matter тощо	Може працювати без інтернету; оновлення – за мережею
Google Home	Переважно хмара + локальні мости (Matter/Thread)	Частково	Середній	Данні в хмарі, локальний control покриває не все	Широкі брендові інтеграції	Нещодавнє розширення автоматизацій; деякі дії ще не підтримано
Amazon Alexa	Переважно хмара + локальні мости (Matter/Zigbee)	Частково	Середній	Залежність від хмарних інтеграцій	Широкі брендові інтеграції	Приклади збоїв/деактивації хмарних зв'язок (IFTTT, вендорські скіли)

Формулювання мети дослідження

Метою роботи є розроблення методичних засад інтеграції IoT-систем у житлових приміщеннях, що забезпечують взаємодію спадкових та сучасних пристроїв і підсистем. Дослідження зосереджене на уніфікації обміну даними та проєктуванні бек-енд хаба з адаптерами до різних протоколів і єдиним доменним API. Результатом є рекомендації та еталонна архітектура локально-орієнтованого шлюзу для поетапної інтеграції й підвищення енергоефективності та безпеки.

Передбачено проаналізувати сучасні стандарти та платформи «розумного» житла й окреслити їхні обмеження для житлових сценаріїв. Далі – класифікувати пристрої та підсистеми та визначити вимоги до їхньої взаємодії з хабом. Сформувати канонічну модель даних і команд із поділом на обов'язкові та розширені поля, уніфікувавши одиниці виміру та часові мітки. Спроєктувати бек-енд хаба з шаром адаптерів до різних протоколів та єдиним доменним API. Описати процедури онбордингу та безпеки, а також політики офлайн-стійкості. Розробити методику оцінювання – затримку виконання команд, надійність доставки подій, стійкість до відмов інтернету/живлення та вплив на енергоефективність. Надати практичні рекомендації з поетапної модернізації – план міграції від спадкових систем до уніфікованої архітектури й типові сценарії впровадження.

Викладення основного матеріалу дослідження

У межах дослідження використано багатошарову архітектуру локально орієнтованого хаба, який виконує роль центрального вузла інтеграції «старих» і сучасних підсистем житла показано на Рисунку 1.

Ядро рішення складається з:

– Шар адаптерів – шлюзи до різних фізичних/канальних та прикладних протоколів: RS-485 чи Modbus, ІЧ, Wi-Fi, BLE за потреби Zigbee або Matter).

– Шар нормалізації – перетворення нативних повідомлень на канонічні події/команди зі спільними одиницями виміру, часовими мітками (UTC) та індикаторами якості.

– Подієва шина – внутрішня подієва шина (MQTT) з QoS, ретенцією та дедуплікацією.

– Реєстр пристроїв та можливостей – реєстр пристроїв, їхніх можливостей, версій профілів і політик доступу.

– Правила/Автоматизація – двигун сценаріїв та правил (локальне виконання без хмари).

– Зовнішнє API – єдине прикладне API (REST/WS) для клієнтів, мобільних застосунків і «peer-hub» інтеграцій.

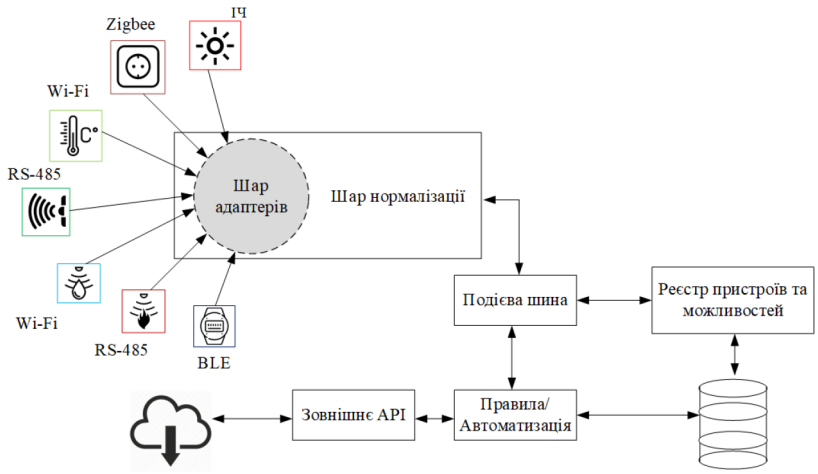


Рис. 1. Референсна архітектура локально орієнтованого хаба (шари та потоки даних)

Канонічна модель даних і доменне API передбачають профілі з поділом на обов'язкове «ядро» та розширення. Мінімальний набір полів для телеметрії: *device_id, capability, ts (UTC ISO 8601), value, unit (SI), quality, schema_ver*. Для команд: *device_id, command, params, ts, schema_ver*. Така організація забезпечує єдину семантику й формат даних для аналітики та ШІ, зворотну сумісність через версіонування схем, а також прозорий обмін між кількома хабами без додаткових перетворень, нижче в таблиці 2 предсавленйи приклад канонічний профілів для типових пристроїв.

Таблиця 2

Приклад канонічних профілів для типових пристроїв (температура, вологість, рух, вимикач)

№	Клас пристрою	Обов'язкові поля	Додаткові поля
1	sensor. temperature	device_id, capability, ts (UTC ISO 8601), value, unit=°C, quality, schema_ver	battery_level, signal_strength, location
2	sensor.motion	device_id, capability, ts, value (bool), quality,	battery_level, signal_strength, occupancy_timeout, sensitivity
3	switch.light	device_id, command(s)/state, ts, schema_ver	brightness (0–100), color_temp (K), transition_ms, scene_id
4	sensor.open	device_id, capability, ts, value (open/closed), quality, schema_ver	battery_level, signal_strength, tamper
5	meter.energy	device_id, capability, ts, value (Wh або kWh), unit, quality, schema_ver	interval_sec, voltage (V), current (A), power (W), phase
6	switch.outlet	device_id, state (on/off), ts, schema_ver	power (W), energy (Wh/kWh), child_lock, overload_protect

Шар адаптерів і мапінг протоколів реалізує уніфікацію вхідних даних і команд: значення з RS-485/Modbus передає до прикладу такі параметри.

```
Reg 0x0001 # температура
Reg 0x0002 # вологість
```

Зчитуються періодично й нормалізуються – перетворення у °C або %, додання ts (UTC) та прапорця quality;

У BLE локальний міст підписується на характеристики наприклад.

```
0x2A6E # Temperature
```

Потім конвертує бінарні дані в єдині одиниці, додає метадані та публікує у шину.

Для ІЧ або 868 МГц коди пакетів зіставляються з каталогом команд фбо станів.

```
mode=COOL, temp=16
mode=HOT, temp=22
```

Відправляються як стандартний command.

Вже пристрої Wi-Fi та IP публікують телеметрію безпосередньо у канонічному форматі або через легкий device agent на мікроконтролері

У разі використання готових інтегрованих модулів Zigbee або Matter програмні мости відображають кластерні та атрибутивні моделі на відповідні capabilities без змін ядра системи.

Онбординг і безпека реалізуються так, під час початкового під'єднання пристрій отримує provisioning-токен або сертифікат, згенерований хабом до прикладу для Wi-Fi задаються SSID та пароль і адреси API, для BLE виконується парування з локальним мостом. Далі увесь транспорт шифрується TLS/mTLS для IP, захищені сесії для BLE, доступ контролюється рольовою авторизацією, а події або команди за потреби підписуються для гарантії цілісності. Ведуться обов'язкові журнали доступів і змін конфігурації. На Рисунку 2 можемо спостерігати дану схему.

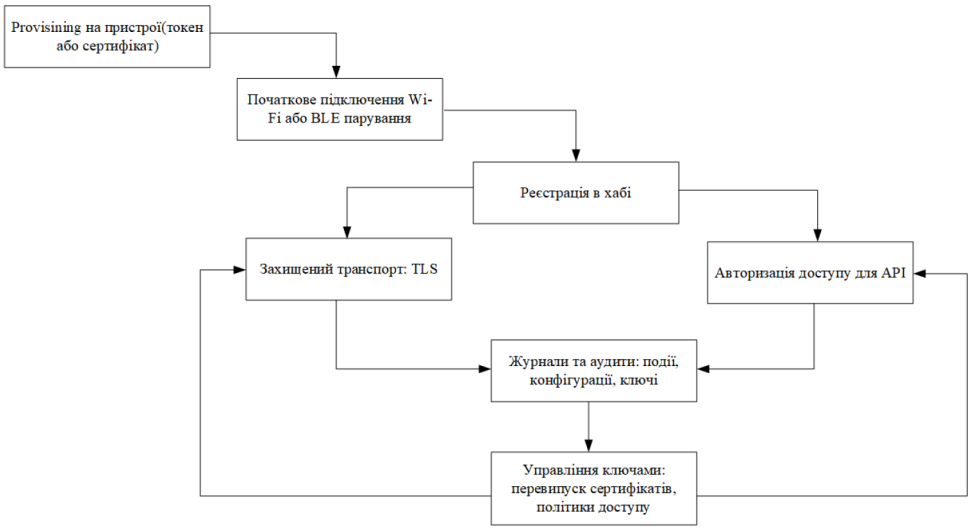


Рис. 2. Онбординг та безпека

Офлайн-стійкість і надійність доставки забезпечуються схемою store-and-forward з повторною відправкою після відновлення зв'язку, використанням рівнів MQTT QoS, збереженням для останнього стану та механізмами heartbeat для виявлення офлайн-пристроїв. Дедуплікація подій виконується за ключем (device_id, ts), а політики ретенції визначаються класом даних чи критичні або некритичні.

Реєстр пристроїв і хабів реалізуються через Device & Capability Registry, який зберігає ідентифікатори, типи, виробника, версії прошивки/профілю, перелік підтримуваних capabilities та дозволів доступу. Для міжвузлової взаємодії передбачено обмін: інші хаби під'єднуються через «Зовнішнє API» або обмінюються подіями безпосередньо по MQTT або WebSocket за узгодженими схемами.

Реалізація прототипу. Еталонний стенд реалізовано на Raspberry Pi із таким стеком:

– MQTT-брокер (Mosquitto) – «Подієва шина»;
– Python-сервіси для під'єднання датчиків – «Шар адаптерів»;
– Python сервер(Python/Flask або FastAPI) – «Шар нормалізації»
– Бек-енд (Python/Flask або FastAPI) – «Зовнішнє API»;
– SQLite (метадані) + файловий буфер телеметрії – «Сховище»;
– Node-RED (опційно) – «Автоматизації»;
– Клієнтський інтерфейс – «Зовнішнє API».

Приклади застосування та сценарії:

1. Оптимізація клімату за присутністю. Дані датчиків руху, відкриття та графік користувача формують цільові температури; команди надсилаються до котла чи кондиціонера через «Подієву шину».

2. Виявлення протікань і аварій. Сенсор води по 433 МГц через хаб сповіщення та перекриття електроклапана.

3. Освітлення з плавним зонуванням. Кнопки чи датчики освітленості BLE корегують сцени освітлення. Збереження останнього стану для швидкого відновлення після зникнення живлення.

Ми перевіряємо чотири прості речі: наскільки швидко система реагує, чи не губить повідомлення, як швидко «оживає» після збоїв і чи дає економію енергії. Для швидкості ми надсилаємо команду (наприклад, увімкнути

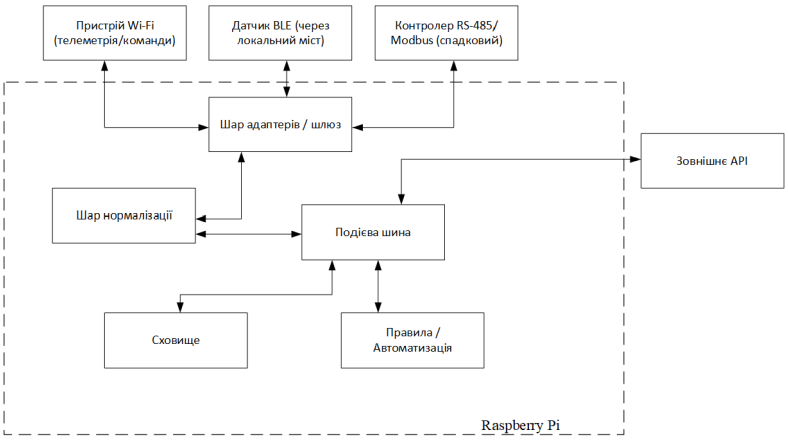


Рис. 3. Фізична топологія стенда та програмні компоненти на Raspberry Pi

світло) й фіксуємо час до фактичного увімкнення – чим менше, тим краще. Рахуємо середній результат і «повільні» випадки. Для надійності навмисно створюємо проблеми – відключаємо інтернет або Wi-Fi – і дивимось, скільки подій система втратила або продублювала в ідеалі – майже нуль. Для відновлення просто перезавантажуємо хаб і міряємо, через скільки секунд правила знову починають працювати без нашого втручання. Для економії енергії порівнюємо показники лічильників «до» і «після» увімкнення автоматизацій, якщо кіловат-годин стало менше – значить, є користь. Усі заміри робимо кілька разів на однакових сценаріях, фіксуємо час у журналах і подаємо результат на графіках та в таблиці з простим порівнянням «було/стало».

Обмеження та подальший розвиток. Поточна реалізація орієнтована на побутові сценарії та обмежений набір протоколів, повноцінну підтримку промислових шин планується додавати через окремі адаптери. Подальший розвиток передбачає автоматичне витягування можливостей з описів WoT та Matter, розширення канонічних профілів і впровадження інструментів перевірки якості даних для подальшої аналітики та застосувань ШІ.

Висновки

У роботі обґрунтовано потребу в уніфікації протоколів і семантики даних для житлових IoT-систем, де одночасно використовуються «старі» дротові та радіорішення та нові IP-пристрої. Запропоновано еталонну архітектуру локально орієнтованого хаба на Raspberry Pi з шарами адаптерів і нормалізації, подієвою шиною MQTT, реєстром пристроїв і їхніх можливостей, рушієм правил та зовнішнім API – що забезпечує інтеграцію без прив'язки до конкретного вендора. Сформовано канонічну модель даних і доменне API з поділом полів на основні та розширені; продемонстровано мапінг Modbus, BLE, ІЧ,433–868 МГц, Wi-Fi та IP і містків Zigbee та Matter у єдині «можливості пристроїв», що спрощує аналітику та застосування ШІ. Описано процедури онбордингу й захисту на основі provisioning-токенів та TLS шифруванні, а також механізми офлайн-стійкості, необхідні для надійної роботи без хмари. Реалізовано еталонний стенд і визначено методику оцінювання – затримка команд, надійність доставки, час відновлення після збоїв та енергетичний ефект у сценаріях клімату й освітлення. Запропоноване рішення придатне для поетапної модернізації житла: воно надає єдиний спосіб підключення різнорідних пристроїв, підвищує приватність за рахунок локального виконання, знижує витрати інтеграції та спрощує масштабування. Обмеженням нинішньої версії є фокус на побутових сценаріях і вибраних протоколах; подальша робота передбачає підтримку повних профілів BACnet та KNX через адаптери, автоматичне витягування «можливостей пристроїв» з описів WoT, Matter та впровадження інструментів валідації якості даних для аналітики й ШІ.

Список використаної літератури

1. Sensors (MDPI). Special Issue: Smart Homes. Description of the direction "IoT for smart homes". https://www.mdpi.com/journal/sensors.

2. Orfanos V. A. et al. A Comprehensive Review of IoT Networking Technologies. Micromachines, 2023. https://www.mdpi.com/2224-2708/12/2/30

3. W3C. Web of Things (WoT) Thing Description 1.1 Recommendation. 2023. https://www.w3.org/TR/wot-thing-description11/

4. Connectivity Standards Alliance. Matter 1.4.1: Enhanced Setup Flow, Multi-Device QR, NFC Onboarding. 2025. https://csa-iot.org/newsroom/a-smarter-start-matter-1-4-1-makes-setup-easier/

5. OASIS. MQTT Version 5.0 – Specification. 2019+. https://docs.oasis-open.org/mqtt/mqtt/v5.0/mqtt-v5.0.html

6. Ali D.M.T.E. et al. AI-Driven Innovations in Building Energy Management Systems: A Review of Potential Applications and Energy Savings. Energies, 2024. https://www.mdpi.com/1996-1073/17/17/4277

7. Google Home: new automation editor (update overview). Android Central, 2025. https://www.androidcentral.com/accessories/smart-home/google-home-new-automation-editor-lets-you-do-so-much-more.

References

1. Sensors (MDPI). Special Issue: Smart Homes. Description of the direction “IoT for smart homes”. https://www.mdpi.com/journal/sensors

2. Orfanos V. A. et al. A Comprehensive Review of IoT Networking Technologies. Micromachines, 2023. https://www.mdpi.com/2224-2708/12/2/30

3. W3C. Web of Things (WoT) Thing Description 1.1 Recommendation. 2023. https://www.w3.org/TR/wot-thing-description11/

4. Connectivity Standards Alliance. Matter 1.4.1: Enhanced Setup Flow, Multi-Device QR, NFC Onboarding. 2025. https://csa-iot.org/newsroom/a-smarter-start-matter-1-4-1-makes-setup-easier/

5. OASIS. MQTT Version 5.0 – Specification. 2019+. https://docs.oasis-open.org/mqtt/mqtt/v5.0/mqtt-v5.0.html

6. Ali D.M.T.E. et al. AI-Driven Innovations in Building Energy Management Systems: A Review of Potential Applications and Energy Savings. Energies, 2024. https://www.mdpi.com/1996-1073/17/17/4277

7. Google Home: new automation editor (update overview). Android Central, 2025. https://www.androidcentral.com/accessories/smart-home/google-home-new-automation-editor-lets-you-do-so-much-more

Дата першого надходження рукопису до видання: 26.09.2025
Дата прийнятого до друку рукопису після рецензування: 23.10.2025
Дата публікації:

УДК 004.852:159.9 DOI

Т. Р. ГОЛУБЕЦЬ
аспірант кафедри автоматизованих систем управління
Інститут комп'ютерних наук та інформаційних технологій
Національного університету «Львівська політехніка»
ORCID: 0009-0004-6613-3620

М. О. МЕДИКОВСЬКИЙ
доктор технічних наук,
професор кафедри автоматизованих систем управління
Інститут комп'ютерних наук та інформаційних технологій
Національного університету «Львівська політехніка»
ORCID: 0000-0003-2492-8578

МОДЕЛЬ ПЕРЕДБАЧЕННЯ РІШЕНЬ ЛЮДИНИ ПРО ВИЗНАННЯ ВТРАТИ КОНТРОЛЮ НАД ПРОЦЕСАМИ

У статті описано результати дослідження особливостей прийняття рішень людиною з метою аналізу впливу на ефективність її роботи разом з автоматизованими системами управління. Висновки цього дослідження можуть бути використані для розробки різних систем підтримки прийняття рішень, оскільки більшість випадків характеризуються показниками складності ситуації, досвіду оператора і часових обмежень. Дослідження проведено з використанням даних про перебіг шахових партій, оскільки у даній сфері наявні метрики таких показників.

Проведено оцінку важливості параметрів, що впливають на прийняття рішень людиною. Аналіз здійснено на основі ситуацій з шахових партій, коли один із суперників мав можливість визнати поразку. Оскільки рішення капітулювати в шахах можна трактувати як втрату контролю. В інших сферах вона проявляється як самоусунення: через ігнорування проблеми, зволікання з рішенням, делегування відповідальності.

У дослідженні використано методи кореляційного аналізу для оцінки зв'язків між змінними. Для створення моделей прогнозування рішення людини застосовано методи машинного навчання, зокрема нейронну мережу, випадковий ліс та алгоритм градієнтного бустингу XGBoost. Ключові фактори прийняття рішень визначено шляхом аналізу впливу вхідних параметрів за допомогою методів посилення розподілу та перестановки.

Результатом дослідження стало створення та навчання моделей для прогнозування рішень людини про визнання втрати контролю чи безнадійності ситуації. На основі трьох навчених моделей, побудованих за вищезгаданими методами, проведено аналіз і встановлено, що складність випадку є, безумовно, найважливішим параметром. Інші фактори, такі як досвід людини, часові обмеження та рівень мотивації, також мають вплив, проте другорядний.

Ключові слова: *прогнозування рішень людини, поведінка в критичних ситуаціях, шахи, системи підтримки прийняття рішень.*

R. HOLUBETS
Postgraduate Student at the Department of Automated Control Systems
Institute of Computer Sciences and Information Technologies
of Lviv Polytechnic National University
ORCID: 0009-0004-6613-3620

O. MEDYKOVSKYY
Doctor of Technical Sciences,
Professor at the Department of Automated Control Systems
Institute of Computer Sciences and Information Technologies
of Lviv Polytechnic National University
ORCID: 0000-0003-2492-8578

A HUMAN DECISION PREDICTION MODEL FOR RECOGNIZING LOSS OF CONTROL IN PROCESSES

This article presents the results of a study on human decision-making patterns with the aim to analyze their impact on the effectiveness of automated control systems. The findings of this study could be applied in the development of decision support systems, as most of the cases are influenced by such factors as complexity of the situation, operator experience

© Голубець Т. Р., Медиковський М. О., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

and time limitations. This research was conducted with the use of chess data, as this domain provides well-defined metrics for the required parameters.

The study evaluates the importance of each group of indicators that have an impact on human decisions. The analysis was based on chess games in which one of the players could resign. The resignation decision in chess could be perceived as loss of control. In other domains it could be seen through problem avoidance, decision delay, or responsibility delegation.

Methods of correlation analysis were used for evaluation of the connections between variables. Machine learning methods such as neural network, random forest and gradient boosting XGBoost were used for prediction of human resignation decisions. The key factors of the decision-making process were determined by analysis of impact of input parameters using feature importance techniques such as gain and permutation importance.

The outcome of the research was the creation and training of models to predict human decision about loss of control or hopelessness. The performed analysis demonstrated that complexity of the situation is by far the most important factor. Other indicators such as person's experience, time limitations or motivational level have some impact, but less significant.

Key words: *prediction of human decisions, behavior under critical conditions, chess, decision support systems.*

Постановка проблеми

Вивчення особливостей прийняття рішень людиною є важливою частиною різних наук, в першу чергу психології чи соціології. Проте актуальність взаємодії людей з комп'ютерними системами продовжує зростати, і багато таких процесів передбачають прийняття рішень. Системи оповіщення, пошуку та демонстрації даних набувають широкого застосування. Однак в розробці таких систем варто враховувати особливості людини, зокрема як приймаються рішення та як різні типи допоміжних автоматизованих засобів впливають на цю операцію.

Ця робота стосується аналізу процесу прийняття рішень. Відповідно, дослідження проведено на даних про вибір людини, що зазвичай не передбачає використання допоміжних засобів.

Основні фактори, що впливають на прийняття рішення, відрізняються залежно від завдання, проте досліджень з аналізом цих чинників поки небагато. Важко охарактеризувати рішення абстрактного диспетчера та фактори, що на нього впливають. Авіадиспетчер чи оператор АЕС повинні враховувати у своїй роботі абсолютно різні показники, але спільними є: досвід, час на прийняття рішення, складність ситуації. Дослідження процесу прийняття рішення з таким набором параметрів є все ще широким, тому ця робота стосується одного з критичних випадків у прийнятті рішення – визнання втрати контролю.

Визнання втрати контролю може бути у форматі делегування колезі, прохання про допомогу, вживання крайніх заходів чи ігнорування ситуації. У галузях різна критичність такого стану людини, проте в будь-якому випадку дослідження цього питання допоможе краще розуміти ступінь впливу зовнішніх і внутрішніх факторів та уникати небезпечних ситуацій.

Складністю проведення досліджень рішень людини є відсутність наборів даних, що містили б необхідну інформацію у великій кількості ситуацій. Для такого завдання потрібно забезпечити повноту вхідних параметрів, що впливали на людину, та чітко зафіксувати рішення, яке було прийнято. Таким вимогам відповідають дані про перебіг партії у шахах. Інформація про партію містить всі ходи, час їх здійснення, досвід гравців, їх потенційну винагороду та результат [1, 2]. Також шахи відрізняються від інших видів спорту тим, що передбачають можливість завершити матч через визнання поразки.

У цій статті події в межах партії та показники, що характерні для шахів, зведено до загальніших понять для можливості перенесення результатів на інші сфери, де людина приймає рішення. Отже, визнання поразки у грі розглядається як визнання втрати контролю чи визнання безнадійності ситуації, досвід та рівень навичок вимірюється рейтинговим показником ELO, а складність ситуації – оцінкою позиції, розрахованою шаховим рушієм Stockfish [3].

Аналіз останніх досліджень і публікацій

Особливо активно питання здійснення вибору досліджується в економічній сфері, де важливо зрозуміти як людина сприймає товар чи послугу залежно від подачі, кількості альтернатив, різниці між цінами [4]. Науковці з галузі психології чи когнітивної нейронауки вивчають процес прийняття рішення з іншого боку – вони приділяють більше уваги емоціям, активності різних частин мозку, самопочуттю людини [5].

Шахові партії як основа для дослідження процесу прийняття рішень використовувалися і раніше. Причиною цього є доступність та повнота даних – чітко збережена послідовність ходів, часові обмеження, визначений рівень навичок опонентів. Такий набір інформації дозволяє, наприклад, досліджувати залежність правильності ходу від витраченого на нього часу чи прогнозувати шанс перемоги в партії на певному етапі на основі ситуації на дошці та часу, що залишився в кожного з гравців [2].

В аналізі прийняття рішень людиною актуальним є визначення того, коли ми помиляємося. Ідентифікація факторів, що впливають на допущення помилки, є важливим питанням з точки зору розуміння людини, а це, в свою чергу, дозволяє покращити процеси управління людьми та розробку систем підтримки прийняття рішень. Стаття [6] присвячена саме такому дослідженню помилок на основі гри в шахи. Особливостями роботи є використання великої кількості даних, чітке визначення понять часового обмеження, досвіду гравців та складності ситуації у цій спортивній дисципліні. Автори стверджують, що особливості випадку мають найбільший вплив, а шанс на

допущення помилки хоча і залежить від досвіду та часового обмеження, проте не так сильно. У статті розглянуто ймовірність допущення помилки і чітко визначено межу між правильним та невдалим ходом на основі зміни шансів на перемогу, якщо обидва гравці надалі здійснюватимуть ідеальні ходи. Ця робота є цінною, проте має кілька обмежень: не враховуються всі попередні ходи, а також аналізуються лише ситуації, що виникають наприкінці матчу (при невеликій кількості фігур).

Серед досвіду, часу та обставин ситуації найбільше досліджень присвячено впливу інтервалу часу, за який рішення потрібно прийняти. Наприклад, робота [6] присвячена бажанню відкласти рішення залежно від доступних варіантів. Результати дослідження показали, що на відтермінування вибору між варіантами впливає не лише відсутність суттєвої переваги одного з них, але й абсолютні характеристики.

Системи підтримки прийняття рішень уже давно не лише відображають інформацію з давачів, але можуть бути налаштовані на контроль показників, інформування про проблеми, пропозиції дій. Водночас все ще існує абстрагування від того, наскільки по-різному люди сприймають певну інформацію і, відповідно, як будуть діяти. Ця теза перекликається з статтею [7], автори якої пояснюють, що потрібно відійти від звичного розподілу, коли частину задач передають на виконання людині, а іншу – автоматичним системам зі штучним інтелектом. Часто визначення виконавця доцільно здійснювати в залежності від конкретного випадку [7]. Певну аналогію можна провести для автоматизованих систем і враховувати особливості реакції людини на інформацію чи, принаймні, аналізувати остаточне рішення та повідомляти особу про виявлені закономірності.

Наступним кроком є порівняння поведінки людини без допоміжної системи та з нею і цьому також присвячено ряд досліджень [8, 9]. Проте початково важливим є вивчення загального процесу прийняття рішення людиною, що і проведено в цій роботі. Особливістю дослідження, що описано в цій статті є аналіз конкретного винятку з усіх можливих ситуацій – рішення людини визнати, що вона не може втримати ситуацію під контролем.

Формулювання мети дослідження

Метою роботи є підвищення ефективності автоматизованих систем управління шляхом розроблення методів та засобів прогнозування рішень оператора про втрату контролю над ситуацією.

Задачі дослідження:

- обґрунтування джерел даних та попередня обробка;
- визначення важливості факторів методами статистичного аналізу;
- розробка та тренування моделей для прогнозування рішення людини про втрату контролю;
- дослідження точності роботи моделей;
- аналіз результатів для ситуацій з різним досвідом, часовим обмеженням та рівнем складності.

Викладення основного матеріалу дослідження

Для цього дослідження було сформовано набір даних, що містить записи про 8 мільйонів ситуацій з шахових поєдинків з інтернет-сервера Lichess [10]. Кожен запис характеризується:

- рівнем навичок (досвідом) опонентів (E1, E2),
- часом, що у них залишився на цей момент (T1, T2),
- складністю ситуації з боку гравця, на чиєму ході виникла ситуація, (C1, C2, C3, C4, C5),
- кількістю ходів від початку партії (N)
- фактом визнання поразки чи продовження гри (R)

Як показник рівня навичок гравця використано рейтинговий показник ELO. Це статистичний показник сили гравця на основі результатів партій. Такий рейтинг відображає очікуваний результат у партії на основі різниці між гравцями, тобто шахіст мав би виграти приблизно 90 % ігор проти опонента, що має на 400 очок ELO менше.

Параметри T1 та T2 характеризують часові обмеження обох гравців. T_i визначається за формулою:

$$T_i = \frac{T_{xi}}{T_{si} + T_{mi} \cdot N_i},$$

де T_i – показник часу у наборі даних, i – позначає одного з гравців, T_{xi} – час, що залишився у гравця у момент ходу, що розглядається (секунди), T_{si} – початкове часове обмеження (ліміт на всю партію) (секунди), T_{mi} – час, що додається до обмеження на кожному ході (секунди), N_i – кількість ходів гравця до моменту, що розглядається.

Складність ситуації оцінюється на основі положення фігур на шаховій дошці шаховим рушієм Stockfish. Це найсильніший шаховий рушій з відкритим вихідним кодом і один з найкращих загалом. Про це можна стверджувати, наприклад, на основі результатів Top Chess Engine Championship, що проводиться регулярно. Точність оцінок цього шахового рушія залежить від глибини аналізу, проте у наявному наборі даних рушій аналізував не менше, ніж 20 ходів наперед, що є достатнім для оцінки рішення людини, оскільки перевищує можливості прорахунку навіть гросмейстерів [11].

Параметр складності випадку C1 відповідає оцінці шансів на перемогу рушієм Stockfish в момент прийняття рішення про доцільність визнання поразки, параметр C2 – один хід перед цим, C3 – за два ходи до ситуації, і так далі до C5.

Момент у шаховій партії логічно також оцінювати за кількістю ходів у партії, що передували ситуації. Така інформація наявна у параметрі N. Оскільки набір даних призначений для аналізу рішень про визнання поразки, то це представлено у атрибуті R.

Унікальність сформованого набору даних полягає у розгляді ситуацій, що виникли у поєдинках, а не партій загалом, та фокусі на рішенні про визнання поразки. Ситуація у цій роботі – це стан(етап) партії, що характеризується положенням фігур на дошці, часовими обмеженнями опонентів, показниками їх рейтингу до початку матчу та кількістю ходів, зроблених раніше у грі.

Під час очищення та перетворення даних відкинуто частину записів, в першу чергу через те, що не всі записи про партії містять оцінки рушія. Окрім присутності аналізу позицій, фільтрацію проведено також за:

- наявністю рейтингового показника ELO у обох гравців,
- наявністю часового обмеження (через невелику кількість партій без часового обмеження),
- кількістю ходів у поєдинку, що перевищує 5.

Перехід від переліку партій до набору даних про шахові ситуації передбачав вибір одного із станів випадковим чином з кожного матчу. Також обрано ще по одному стану в момент визнання поразки з тих поєдинків, де це відбулося. Створення нових атрибутів на основі наявних також було необхідним. Зокрема, було визначено відносні показники часового обмеження T1 та T2.

Статистичний аналіз даних передбачав визначення кореляційних зв'язків, а також порівняння коефіцієнтів кореляції для різних груп даних: поділ за досвідом, часом, та складністю ситуації.

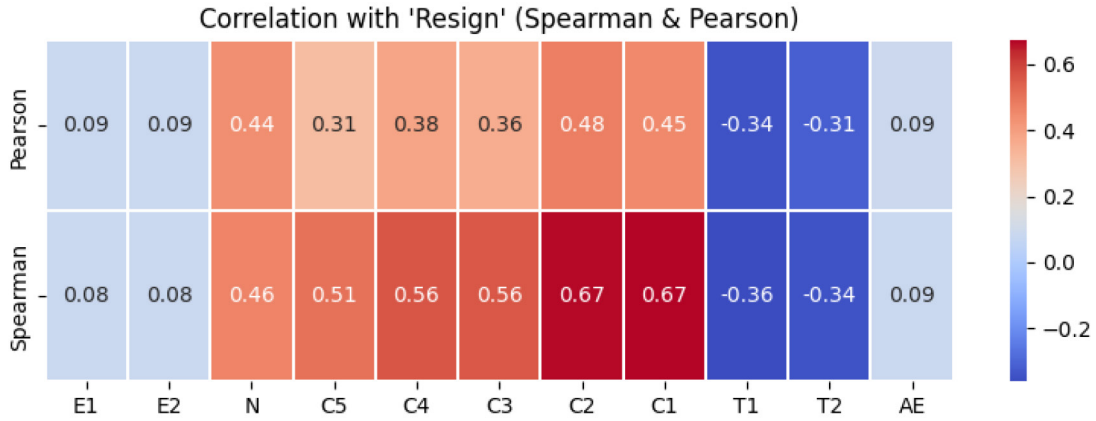


Рис. 1. Коефіцієнти кореляції вхідних параметрів з рішенням про визнання поразки визначені методами Пірсона та Спірмена

Визначення кореляційних зв'язків між рішенням визнати поразку та вхідними параметрами показало залежність, в першу чергу, від особливостей положення фігур на дошці. Найвищі коефіцієнти рангової кореляції Спірмена визначено у оцінок складності ситуацій (0.51–0.67) та часового обмеження (0.34, 0.36). Результати використання цього методу демонструють, що рішення про визнання втрати контролю не залежить від досвіду. Тим не менш, певна закономірність існує – за п'ять ходів до визнання поразки у більш досвідчених гравців ситуація на 39 % краще від ідентичного показника у початківців. Професіонали здатні довше не доводити ситуацію до критичної, проте швидше визнають втрату контролю, якщо ситуація справді критична.

Також коефіцієнти кореляції Спірмена отримані на основі набору даних про партії шахістів з досвідом є суттєво вищими, за ідентичні показники визначені на іграх початківців, що наведено на рис. 2. Наприклад, вхідні параметри, що описують часове обмеження, корелюють з рішенням про визнання поразки в середньому на 0.16, а параметри складності ситуації – на 0.23, якщо розглядати поєдинки аматорів (до 500 ELO). Для партій більш досвідчених гравців(від 2000 ELO) ці показники становлять 0.38 та 0.66 відповідно. Це може свідчити про те, що професіонали діють більш раціонально, в їх поведінці простіше виявляти закономірності, а, відповідно, і передбачати дії.

Суттєві зміни в коефіцієнтах кореляції для різних груп найбільше помітні саме при поділі за досвідом. Результати схожих перевірок для часових обмежень та складності ситуацій майже не відрізняються між вибірками, що відображено на рис. 3.

У дослідженні використано три методи машинного навчання: глибоку нейронну мережу, випадковий ліс та градієнтний бустинг XGBoost. Моделі на основі цих методів суттєво відрізняються, зокрема на етапі навчання, але дозволяють виконати аналіз впливу вхідних параметрів.

На основі трьох вищезгаданих методів навчено моделі для задачі прогнозування: прийнято рішення про визнання поразки чи ні. Нейронна мережа, багатошаровий перцептрон, складається з 3 прихованих рівнів на 64, 32 та

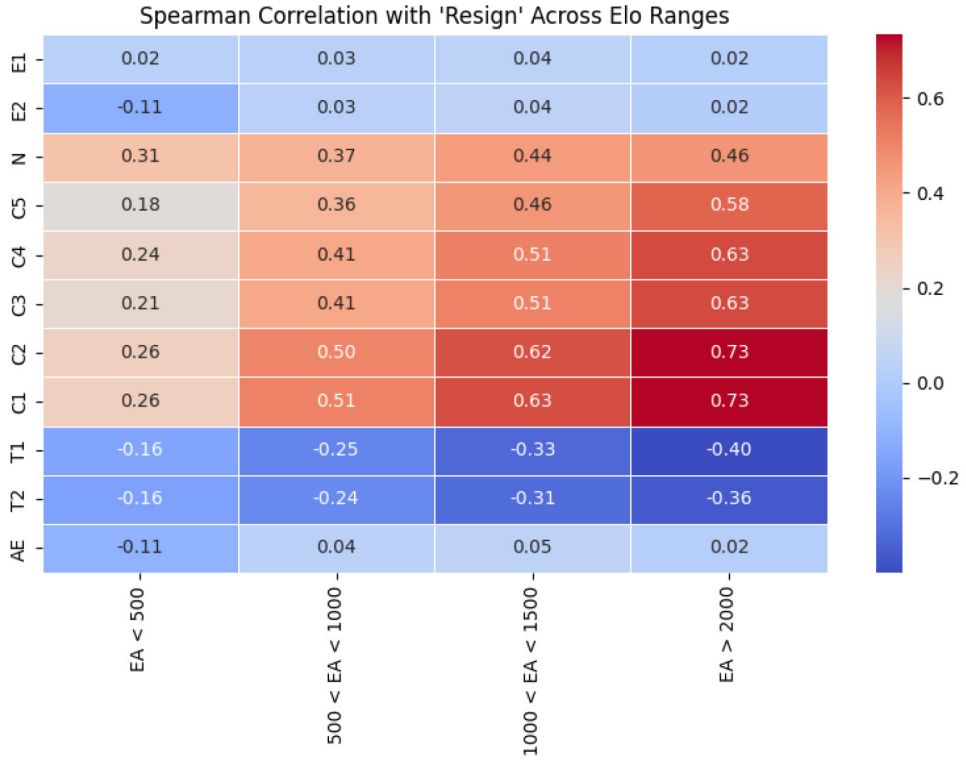


Рис. 2. Коефіцієнти кореляції вхідних параметрів з рішенням про визнання поразки для гравців різного рівня навичок

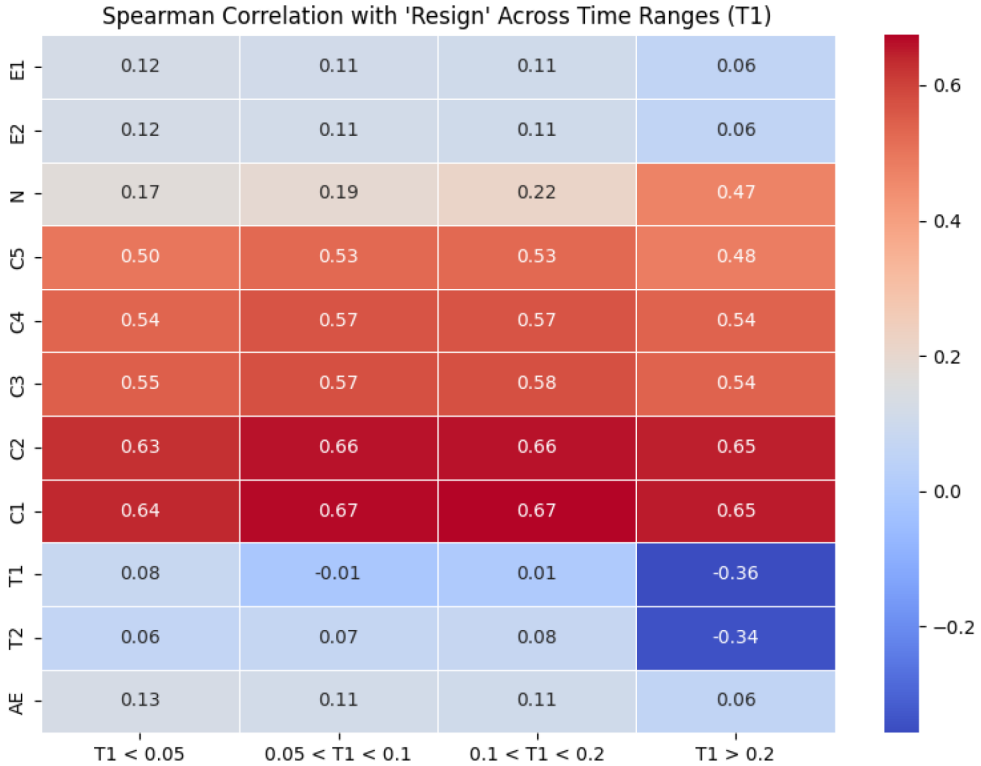


Рис. 3. Коефіцієнти кореляції вхідних параметрів з рішенням про визнання поразки для ситуацій з різними часовими обмеженнями

16 нейронів відповідно. Вихідний шар з сигмоїдною функцією активації забезпечує отримання результату у форматі (0, 1), що потрібно для прогнозування однієї з двох подій. Бінарну крос-ентропію використано як функцію втрат, а алгоритм Adam для оптимізації.

Модель випадкового лісу створено з 100 деревами, а також використано коригування ваг класів, хоча набір даних і є близьким до збалансованого. Модель градієнтного бустингу реалізовано з такими стандартними налаштуваннями: кількість дерев у ансамблі – 100, коефіцієнт швидкості навчання – 0.1, максимальна глибина кожного дерева – 6. Для контролю процесу навчання використано функцію оцінки логарифмічної похибки.

Всі моделі машинного навчання досягли точності більше 90 %. Детальніші результати, зокрема з даними про хибні позитиви та хибні негативи, наведено у таблиці 1.

Таблиця 1

Звіт по класифікації навчених моделей

Модель	Клас	Точність	Повнота	F1-score	Загальна точність
Нейронна мережа	Визнання поразки	0.90	0.90	0.88	0.908
	Продовження гри	0.91	0.91	0.92	
Випадковий ліс	Визнання поразки	0.90	0.91	0.91	0.9144
	Продовження гри	0.91	0.92	0.91	
Градієнтний бустинг	Визнання поразки	0.94	0.88	0.91	0.9137
	Продовження гри	0.89	0.94	0.92	

Нейронній мережі було достатньо 10 епох з наявним набором даних для досягнення 90.8 % точності. Модель ансамблевого методу випадковий ліс визначила 91.44 % випадків правильно, а градієнтний бустинг – 91.37 %.

У описаних умовах метод випадкового лісу працює краще, проте різниця не є значною, і XGBoost, і нейронна мережа не помиляються у переважній більшості випадків. Такий стабільний результат навчання усіх моделей свідчить про можливість прогнозувати людські рішення на основі обмеженого набору даних про ситуацію.

Визначення факторів, що суттєво впливають на прогнозування рішень людини, здійснено двома методами: домішок та перестановки. Моделі випадкового лісу та градієнтного бустингу дозволяють визначити вплив вхідних параметрів за тим, наскільки часто вони були використані при побудові дерев. Результати цього аналізу наведено на рис. 4. Обидві моделі продемонстрували наявність суттєвої залежності результату від показників складності ситуації. Параметри C1 та C2 мають найбільший вплив, а оцінки попередніх позицій C3–C5 також відіграють певну роль, хоч і меншу. Часове обмеження теж використовується в розгалуженнях дерева, а рівень навичок майже не враховується.

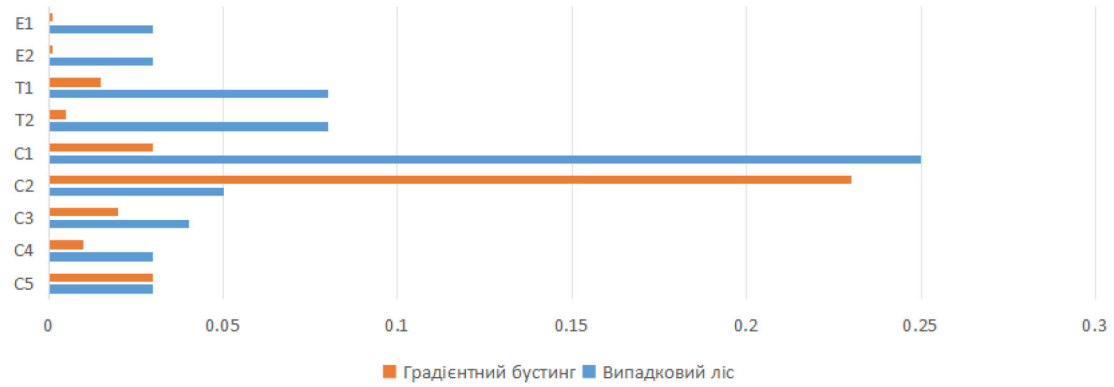


Рис. 4. Важливість ознак за частотою використання у розгалуженнях деревах

Метод перестановок значень у тестовому наборі даних є більш універсальним, що дозволяє проаналізувати вплив параметрів на всіх створених моделях, включно з нейронною мережею. Графік на рис. 5. демонструє, що C1 та C2 є найбільш важливими для точного прогнозування для кожного з використаних методів машинного навчання.

На основі аналізу обома методами можна стверджувати, що складність ситуації найбільше впливає на рішення людини визнати поразку. Рівень впливу найбільший у показника, що характеризує останнє положення фігур на дошці перед роздумами над втратою контролю. Час і досвід людини відіграють свою роль, але вони не є суттєвими у порівнянні з об'єктивною оцінкою проблеми.

Виявлено відмінності між коефіцієнтами кореляції та рівнем важливості вхідних параметрів визначені на основі натренованих моделей штучного інтелекту. Найбільш помітною є різниця при порівнянні показників

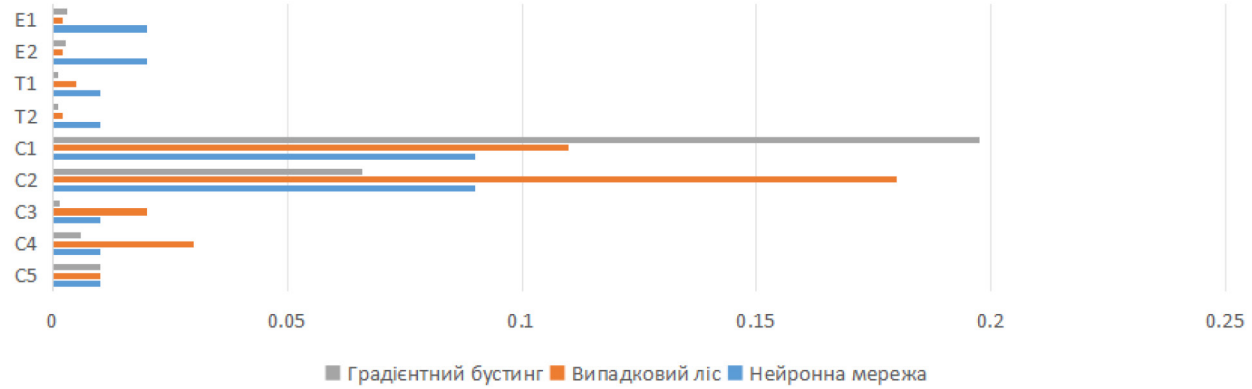


Рис. 5. Важливість ознак за впливом при перестановці даних

часового обмеження та досвіду. Їх коефіцієнти кореляції відрізняються в 4 рази, проте важливість для точного прогнозування моделей майже однакова.

Кожна ситуація характеризувалася оцінкою стану дошки за 4 ходи до прийняття рішення(C5), за 3 ходи(C4), за 2(C3), в момент передостаннього ходу(C2) та в момент, коли гравець обирав між продовженням гри та капітуляцією(C1). Проведений аналіз продемонстрував невеликий рівень користі від C5–C3. Також виявлено, що найважливішою є саме оцінка перед останнім ходом, а не після нього. Оскільки останній хід – рішення опонента, а не гравця, що думає над визнанням поразки, то можна припустити, що критичною є саме власна помилка.

Висновки

Процес прийняття рішення є складним для дослідження і потребує великої кількості даних. Часто науковці при вивченні реальних ситуацій змушені опиратися лише на обмежений перелік параметрів. Наприклад, складність обставин, в яких опинилася особа, певний показник її досвіду та часові обмеження. Для цього дослідження шахові партії стали джерелом саме таких даних, оскільки гравці постійно стикаються з потребою приймати рішення. Зокрема проведено аналіз опції визнання поразки, що близька до визнання втрати контролю в інших сферах.

Хоча прогнозування факту визнання поразки є бінарною класифікацією, викликом є обмежений набір даних і непередбачуваність поведінки людини. У цій роботі окрему увагу приділено аналізу роботи моделей для визначення впливу параметрів не лише методами статистичного аналізу, але й залежністю результатів прогнозування від кожного з наявних факторів. Було виявлено суттєвий вплив об'єктивної складності ситуації, тобто оцінки позиції, на вибір людини завершити гру своєю поразкою. Це продемонстровано перевірками коефіцієнтів кореляції Пірсона та Спірмена та методами домішок і перестановок на навчених моделях. Також відчутний вплив має часове обмеження, на відміну від досвіду.

Проведене дослідження показало, що обмежений набір факторів дозволяє прогнозувати рішення людини у шахових партіях. Усі моделі продемонстрували високу точність та приблизно однаковий рівень впливу параметрів. Такий результат підтверджує валідність проведеного дослідження та побудованих моделей.

Існує багато способів продовження вивчення цього питання, як в галузі шахів, так і в інших сферах. Потенційно цікавим є аналіз процесу прийняття рішення не як ізольованого, а у формі часового ряду. Такий експеримент дозволив би проаналізувати всі ходи до певного моменту і прогнозувати бажання визнати поразку вже з врахуванням раніше втраченої переваги чи навпаки поверненню до стабільності зі значно важчого становища.

При інтерпретації результатів дослідження варто врахувати, що шахи все ж відрізняються від критичних систем управління. Часто на прийняття рішення впливають фактори, які важко виміряти та оцінити – повнота інформації, комунікація, рівень втоми чи навантаження іншими проблемами. Водночас висновки дослідження відкривають можливості для застосування у інших сферах, що може сприяти підвищенню ефективності систем підтримки прийняття рішень.

Список використаної літератури

1. Chowdhary S., Iacopini I., Battiston F. Quantifying human performance in chess. Scientific Reports, 2023, 13(1):2113. doi:10.1038/s41598-023-27735-9

2. Sigman M., Etchemendy P., Slezak D. F., Cecchi G. A. Response time distributions in rapid chess: A large-scale decision making experiment. Frontiers in Neuroscience, 2010, 4:60. doi:10.3389/fnins.2010.00060

3. Anderson A., Kleinberg J., Mullainathan S. Assessing human error against a benchmark of perfection. "Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)", 2016, pp. 705–714. doi:10.48550/arXiv.1606.0495

4. Hakobyan N. R., Sotnyk I., Chortok Y., Khachatryan A. On the issue of rationality of economic choice in the context of the psychological processes of anomie. "Katchar" Collection of Scientific Articles International Scientific-Educational Center NAS RA, 2024, pp. 1–13. doi:10.54503/2579-2903-2024.1-13

5. Molina I., Molina-Perez E., Sobrino F., Tellez-Rojas M. A., Zamora-Maldonado H. C., Plaza-Ferreira M., Orozco Y., Espinoza-Juarez V., Serra-Barragán L., De Unanue A. Current research trends on cognition, integrative complexity, and decision-making: A systematic literature review using activity theory and neuroscience. Frontiers in Psychology, 2023, 14:1156696. doi:10.3389/fpsyg.2023.1156696

6. Bhatia S., Mullett T. L. The dynamics of deferred decision. Cognitive Psychology, 2016, 86:112–151. doi:10.1016/j.cogpsych.2016.02.002

7. Raghu M., Blumer K., Corrado G., Kleinberg J., Obermeyer Z., Mullainathan S. The algorithmic automation problem: Prediction, triage, and human effort. arXiv [Електронний ресурс]. 2019. Режим доступу: https://arxiv.org/abs/1903.12220

8. Bondi E., Koster R., Sheahan H., Chadwick M., Bachrach Y., Cemgil T., Paquet U., Dvijotham K. Role of human-AI interaction in selective prediction. "Proceedings of the AAAI Conference on Artificial Intelligence", 2022, 36(5):5286–5294. doi:10.1609/aaai.v36i5.20465

9. Wang X., Lu Z., Yin M. Will you accept the AI recommendation? Predicting human behavior in AI-assisted decision making. "Proceedings of the ACM Conference on Human Factors in Computing Systems", 2022, pp. 1697–1708. doi:10.1145/3485447.3512240

10. Lichess. Lichess open database [Електронний ресурс]. Режим доступу: https://database.lichess.org/

11. Acher M., Esnault F. Large-scale analysis of chess games with chess engines: A preliminary report. arXiv [Електронний ресурс]. 2016. doi:10.48550/arXiv.1607.04186

References

1. Chowdhary, S., Iacopini, I., & Battiston, F. (2023). Quantifying human performance in chess. Scientific Reports, 13(1), 2113. https://doi.org/10.1038/s41598-023-27735-9

2. Sigman, M., Etchemendy, P., Slezak, D. F., & Cecchi, G. A. (2010). Response time distributions in rapid chess: A large-scale decision making experiment. Frontiers in Neuroscience, 4, 60. https://doi.org/10.3389/fnins.2010.00060

3. Anderson, A., Kleinberg, J., & Mullainathan, S. (2016). Assessing human error against a benchmark of perfection. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16) (pp. 705–714). ACM. https://doi.org/10.48550/arXiv.1606.04956

4. Hakobyan, N. R., Sotnyk, I., Chortok, Y., & Khachatryan, A. (2024). On the issue of rationality of economic choice in the context of the psychological processes of anomie. "Katchar" Collection of Scientific Articles International Scientific-Educational Center NAS RA, 1–13. https://doi.org/10.54503/2579-2903-2024.1-13

5. Molina, I., Molina-Perez, E., Sobrino, F., Tellez-Rojas, M. A., Zamora-Maldonado, H. C., Plaza-Ferreira, M., Orozco, Y., Espinoza-Juarez, V., Serra-Barragán, L., & De Unanue, A. (2023). Current research trends on cognition, integrative complexity, and decision-making: A systematic literature review using activity theory and neuroscience. Frontiers in Psychology, 14, 1156696. https://doi.org/10.3389/fpsyg.2023.1156696

6. Bhatia, S., & Mullett, T. L. (2016). The dynamics of deferred decision. Cognitive Psychology, 86, 112–151. https://doi.org/10.1016/j.cogpsych.2016.02.002

7. Raghu, M., Blumer, K., Corrado, G., Kleinberg, J., Obermeyer, Z., & Mullainathan, S. (2019). The algorithmic automation problem: Prediction, triage, and human effort. arXiv preprint arXiv:1903.12220. https://arxiv.org/abs/1903.12220

8. Bondi, E., Koster, R., Sheahan, H., Chadwick, M., Bachrach, Y., Cemgil, T., Paquet, U., & Dvijotham, K. (2022). Role of human-AI interaction in selective prediction. Proceedings of the AAAI Conference on Artificial Intelligence, 36(5), 5286–5294. https://doi.org/10.1609/aaai.v36i5.20465

9. Wang, X., Lu, Z., & Yin, M. (2022). Will you accept the AI recommendation? Predicting human behavior in AI-assisted decision making. Proceedings of the ACM Conference on Human Factors in Computing Systems, 1697–1708. https://doi.org/10.1145/3485447.3512240

10. Lichess. (n.d.). Lichess open database. https://database.lichess.org/

11. Acher, M., & Esnault, F. (2016). Large-scale analysis of chess games with chess engines: A preliminary report. arXiv preprint arXiv:1607.04186. https://doi.org/10.48550/arXiv.1607.04186

Дата першого надходження рукопису до видання: 27.09.2025
Дата прийнятого до друку рукопису після рецензування: 23.10.2025
Дата публікації:

UDC 667.021.1 DOI

O. Y. HONSOR
Candidate of Technical Sciences
Associate Professor at the Department of Specialized Computer Systems
National University "Lviv Polytechnic"
ORCID: 0000-0003-0895-5859

TRUSTWORTHY AI IN SMART SENSOR SYSTEMS IN IOT: CHALLENGES AND SOLUTIONS

This article discusses the key challenges and prospects for implementing reliable artificial intelligence (AI) in intelligent sensor systems of the Internet of Things (IoT), with a special focus on the principles of trust, security, explainability, and legal compliance. The integration of AI with sensor networks allows for autonomous, context-sensitive real-time decision-making, which is critical for healthcare, industry, transportation, and environmental monitoring.

The paper presents an interdisciplinary review of current scientific approaches, including the use of Federated Learning, Explainable AI, secure data processing methods at the edge, Edge Intelligence, as well as risk assessment and ethical compliance in accordance with international standards such as GDPR, ISO/IEC 42001, NIST AI RMF, IEEE 7000, and others. The analysis also covers the issues of data quality and integrity, security, transparency, bias, reliability, and regulatory uncertainties that arise when implementing AI in IoT systems.

The author proposes a conceptual model of trust in AI based on seven interrelated factors: data quality, reliability, security, privacy protection, explainability, fairness, and legal compliance. The model allows to quantify the level of trust in the system by adapting the weighting factors to the specifics of a particular application.

In addition to analyzing the problems, the authors provide practical recommendations for overcoming challenges, such as implementing real-time data quality checks, building a secure architecture based on the secure-by-design principle, using XAI algorithms (SHAP, LIME), and conducting fairness and compliance audits. The table of correspondences between challenges and solutions allows a systematic approach to the implementation of trusted AI in complex distributed systems.

Thus, this work forms the basis for the development of policies, standards, and strategies for the implementation of trusted artificial intelligence in IoT environments, maintaining a balance between technological innovation and ethical and legal responsibility.

Key words: *Trustworthy AI, Smart Sensor Systems, Internet of Things, Explainable AI.*

О. Й. ГОНСЬОР
кандидат технічних наук,
доцент кафедри спеціалізованих комп'ютерних систем
Національний університет «Львівська політехніка»
ORCID: 0000-0003-0895-5859

ДОСТОВІРНИЙ ШІ В ІНТЕЛЕКТУАЛЬНИХ СЕНСОРНИХ СИСТЕМАХ IOT: ВИКЛИКИ ТА РІШЕННЯ

У статті розглядаються ключові виклики та перспективи впровадження надійного штучного інтелекту (AI) в інтелектуальні сенсорні системи Інтернету речей (IoT), з особливим акцентом на принципи довіри, безпеки, пояснюваності та правової відповідності. Інтеграція AI з сенсорними мережами дозволяє досягати автономного, контекстно-чутливого прийняття рішень у реальному часі, що є критично важливим для галузей охорони здоров'я, промисловості, транспорту та екологічного моніторингу.

В роботі представлено міждисциплінарний огляд сучасних наукових підходів, включаючи використання федеративного навчання (Federated Learning), пояснюваного AI (XAI), методів безпечної обробки даних на периферії (Edge Intelligence), а також оцінку ризиків і забезпечення етичної відповідності згідно з міжнародними стандартами – такими як GDPR, ISO/IEC 42001, NIST AI RMF, IEEE 7000 та інші. Проведений аналіз також охоплює проблеми якості та цілісності даних, безпеки, прозорості, упередженості, надійності, а також нормативно-правових невизначеностей, які виникають при впровадженні AI в IoT-системи.

Запропоновано концептуальну модель довіри до AI, яка базується на семи взаємопов'язаних факторах: якість даних, надійність, безпека, захист конфіденційності, пояснюваність, справедливість і правова відповідність. Модель дозволяє кількісно оцінити рівень довіри до системи, адаптуючи вагові коефіцієнти під специфіку конкретного застосування.

Окрім аналізу проблем, автори надають практичні рекомендації для подолання викликів, таких як впровадження перевірки якості даних у реальному часі, побудова безпечної архітектури за принципом «безпека

© Honsor O. Y., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

за задумом» (secure-by-design), застосування XAI-алгоритмів (SHAP, LIME), проведення аудитів справедливості та відповідності нормативним актам. Таблиця відповідностей між викликами і рішеннями дозволяє системно підійти до реалізації довіреного AI у складних розподілених системах.

Таким чином, дана робота формує основу для формування політик, стандартів та стратегій впровадження довіреного штучного інтелекту в середовищах Інтернету речей, підтримуючи баланс між технологічною інновацією та етично-правовою відповідальністю.

Ключові слова: *Достовірний ШІ, Розумні Сенсорні Системи, Інтернет Речей, Пояснюваний ШІ.*

Problem statement

Smart sensor systems are intelligent devices that collect, process, and transmit data within the Internet of Things (IoT) ecosystem. These sensors integrate artificial intelligence (AI), machine learning (ML), and edge computing to enhance automation, efficiency, and real-time decision-making. In the context of Industry 4.0, these systems play a pivotal role in predictive maintenance, quality control, and supply chain optimisation by facilitating autonomous operations.

AI improves sensor networks by analysing vast amounts of data, identifying patterns and making autonomous decisions with minimal human intervention. However, an over-reliance on artificial intelligence can lead to bias and reduced transparency and accountability in decision-making.

Trust is imperative for successfully integrating artificial intelligence (AI) within the industrial Internet of Things (IoT) domain. Ensuring that artificial intelligence (AI) aligns with regulatory imperatives, including those stipulated by the General Data Protection Regulation (GDPR), the International Organization for Standardization (ISO)/International Electrotechnical Commission (IEC) 42001 standards, and the EU's AI Act, is of paramount importance. The development of reliable AI hinges upon establishing governance frameworks that strike a balance between the pursuit of innovation and the implementation of legal and ethical safeguards.

The objective of this article is to furnish policy and legal strategy recommendations that aim to fortify trust, enhance accountability, and bolster compliance in the context of AI-driven IoT applications.

Analysis of Recent Research and Publications

Recent advances in integrating artificial intelligence with intelligent sensor systems within the Internet of Things have been accompanied by significant research aimed at ensuring the reliability of these systems. A review of the seminal scientific contributions from recent years reveals a concerted effort to address the challenges and propose solutions that guarantee the reliability of artificial intelligence in intelligent sensor systems.

Federated Learning (FL) has emerged as a pivotal strategy for enhancing privacy and security in distributed AI systems. In [1], FederatedTrust is presented, an algorithm designed to calculate the trustworthiness of FL models by evaluating six pillars: privacy, reliability, fairness, explainability, accountability, and federation. Their work underscores the necessity for comprehensive metrics to assess and ensure the reliability of AI models in decentralised environments.

Authors in [2, 3] conducted a survey on explainable artificial intelligence (XAI) techniques tailored for the IoT, discussing their applicability in areas such as security enhancement and the Industrial Internet of Things IIoT. The authors advocate for integrating XAI to enhance transparency and user trust in AI-driven IoT systems.

In [4], a comprehensive guide is provided for developing trustworthy AI systems, addressing aspects such as robustness, fairness, privacy preservation, and accountability. The researchers propose a systematic approach that spans the entire AI lifecycle, offering actionable insights for practitioners to enhance trustworthiness in AI applications.

The convergence of Edge Computing and AI, known as Edge Intelligence (EI), gives rise to resource constraints and security challenges. In their seminal work, Wang et al. [5] presented a comprehensive survey on the subject of trustworthy EI, delineating its fundamental characteristics and a multi-layered architectural framework. They discuss enabling technologies and solutions aimed at achieving trustworthiness in EI systems, emphasising the importance of security, reliability, transparency, and sustainability.

Integrating artificial intelligence (AI) within the Internet of Things (IoT) systems introduces significant cybersecurity risks. A recent article in Frontiers in Big Data [6–8] discusses the importance of explainability and transparency in AI/ML models for dependency risk analysis in IoT. The authors of this article emphasise the importance of explainable artificial intelligence (XAI) techniques, interpretable models, and human-in-the-loop systems to enhance trustworthiness in AI-driven IoT applications.

In [9–11], authors examined the application of artificial intelligence (AI) and the Internet of Things (IoT) technologies in the context of environmental pollution monitoring and management. The authors expound on the intricacies of predicting and tracking alterations in pollution levels, underscoring the promise of AI-driven sensor systems in confronting environmental predicaments. They further emphasise the necessity of reliable AI in these applications, underscoring the need for rigorous research and development to ensure the integrity and efficacy of these technologies.

A collective examination of these studies underscores the multifaceted efforts underway to ensure that AI-integrated smart sensor systems in IoT are developed and deployed in a trustworthy manner. Such efforts address challenges related to privacy, security, explainability, and sustainability.

Formulation of Research Objective

Integrating artificial intelligence with intelligent sensor systems enables autonomous, context-sensitive, real-time decision-making in various industries, including healthcare, manufacturing, transportation, and environmental monitoring. However, these advancements have given rise to significant concerns regarding the reliability of such systems. The concept of trustworthiness in the context under consideration encompasses several interrelated attributes.

- **Reliability:** The AI system must function accurately and consistently under various operational conditions.
- **Transparency and Explainability:** Stakeholders must understand how AI models make decisions, especially in safety-critical or regulated environments.
- **Security and Privacy:** As sensor systems collect and process sensitive data, ensuring data integrity and privacy is crucial. Techniques like federated learning, secure multiparty computation, and differential privacy are increasingly employed.
- **Ethical and Legal Compliance:** Trustworthy AI systems must adhere to ethical principles and comply with legal standards such as the GDPR, AI Act, and sector-specific guidelines.
- **Accountability and Fairness:** The AI's decision-making process must be auditable, and outcomes must not reflect systemic bias or discrimination.

Recent literature has proposed various strategies to enhance trust, such as adopting Explainable AI (XAI) models, AI risk assessment frameworks, and standardisation efforts (e.g., IEEE P7000 series, ISO/IEC 42001). Moreover, trustworthiness is increasingly treated as a multidisciplinary challenge involving not only technical robustness but also human-centric and governance perspectives.

Table 1 presents a compendium of pivotal standards and regulations pertaining to trustworthiness in AI-driven smart sensor systems, methodically organised by category.

These standards encompass critical domains such as AI ethics, safety, cybersecurity, data protection, and system reliability, which are paramount in IoT environments.

Table 1

Standards and regulations pertaining to trustworthiness in AI-driven smart sensor systems

Category	Standard / Regulation	Scope / Description	Relevance to Trustworthiness
AI Governance	EU AI Act (2024/25)	Risk-based regulation of AI systems in the EU	Requires transparency, risk management, and oversight
	NIST AI RMF (2023)	US framework for managing AI risks	Emphasises reliability, explainability, and fairness
	ISO/IEC 42001 (2023)	AI Management System Standard	Operationalises trustworthy AI in organisations
	OECD AI Principles (2019)	Global ethical guidelines for AI development	Promotes human-centric, fair, and transparent AI
Ethics & Explainability	IEEE 7000 Series	Standards for the ethical design of autonomous systems	Covers transparency, bias mitigation, and privacy
	IEEE 7001, 7002, 7003	Focus on transparency, privacy, and algorithmic fairness	Ensures ethical deployment of AI in sensor networks
Functional Safety	IEC 61508	Functional safety for electronic systems	Ensures fail-safe design of AI-integrated sensor systems
	ISO 13849	Safety of machinery with embedded control systems	Used in industrial AI-driven automation
	ISO 26262	Functional safety in automotive systems	Applies to AI with smart sensors in autonomous vehicles
Cybersecurity	IEC 62443	Cybersecurity for industrial control systems	Protects sensor-based AI in manufacturing
	ISO/IEC 27001	Information security management systems	Ensures confidentiality and data integrity
	NIST SP 800-53/800-82	Security for federal and industrial control systems	Provides cybersecurity guidelines for smart IoT systems
Privacy & Data	GDPR (EU)	Regulation on data protection and privacy	Requires consent, transparency, and explanation rights
	CCPA / CPRA (US)	US data privacy regulations for consumers	Controls handling of personal sensor data
	ISO/IEC 27701	Privacy Information Management System	Extension of ISO 27001 for managing PII
System Quality	ISO/IEC 25010	Software/system quality models	Ensures reliability, usability, and maintainability
	ISO 9001	General quality management systems	Supports consistent and safe development practices

Presentation of the Main Research Material Trust Challenges in AI-based IoT

Integrating artificial intelligence (AI) into Internet of Things (IoT) ecosystems, particularly those that incorporate intelligent sensor systems, has introduced a novel dimension of intelligence and autonomy. These systems have found widespread application in domains such as predictive maintenance in manufacturing, real-time health monitoring, and the management of smart city infrastructure. Concurrently, implementing AI in sensor systems introduces specific challenges related to data trustworthiness and the reliability and ethical use of AI models. Addressing these challenges is imperative to ensure safe, dependable, and ethical operations.

Data Quality and Integrity: The primary concern pertains to the integrity and reliability of the data. The integrity of the data received from sensors is paramount to operating artificial intelligence models, which rely on this data to make accurate predictions or decisions. However, IoT environments are inherently dynamic and subject to various physical and digital stresses. Drift, calibration errors, hardware failures, or data transmission delays can cause data streams to become corrupted or biased. This compromises data integrity, leading to suboptimal model performance or erroneous conclusions.

The heterogeneity of IoT devices further complicates matters by introducing variability in data preprocessing and model training. AI models may exhibit erratic behaviour without dependable data validation and normalisation mechanisms.

Security and Privacy Risks: As AI-powered IoT systems are often deployed in open, physically accessible, or resource-constrained environments, data security is a critical aspect. These systems are vulnerable to various cyber threats, such as data poisoning attacks (where malicious data is injected into the learning process) and unauthorised access to control or decision-making components.

Edge computing is widely used in IoT systems to reduce latency and increase throughput. However, they often lack the computing resources to support advanced security mechanisms.

Lack of Explainability and Transparency: A fundamental component of trust in AI is explainability. Unfortunately, many AI models used in IoT applications (e.g., deep learning networks) operate as "black boxes," making it challenging to trace decision paths or verify their reasoning. This opacity is problematic not only for end users but also for auditors, regulators, and system engineers. Continuous learning or online adaptation, often required in dynamic environments, further obscures model behaviour and complicates version control and auditing. Current regulatory frameworks increasingly mandate explainability and traceability, which IoT systems are not always equipped to provide.

Bias and Fairness: Uneven sensor placement, biased historical data, or algorithmic preferences in IoT data processing can lead to biased decisions, such as untimely maintenance or insufficient service coverage. In addition, sensor data may reflect historical inequalities or operational inconsistencies, and artificial intelligence models trained on such data may perpetuate or even reinforce these biases. Ensuring equity requires not only technical solutions but also ethical considerations in sensor placement, data collection policies, and algorithm development.

Reliability and Robustness: AI systems in IoT must operate reliably in real-world conditions, where noise, faults, or adversarial conditions are common. Models must be resilient to single-point failures (e.g., a sensor outage) and capable of operating with degraded inputs. However, many AI models are sensitive to small input perturbations and lack built-in fault tolerance mechanisms. Real-time requirements add another layer of complexity. AI models must make timely decisions, often with partial or delayed information. Any delay, crash, or misclassification in a mission-critical application (e.g., industrial automation or emergency health response) could lead to severe consequences. Ensuring robustness and real-time responsiveness is essential but technically challenging.

Legal and Ethical Ambiguities: The legal landscape for AI in IoT is still evolving. Questions around liability, accountability, and compliance are especially murky when autonomous systems act without direct human intervention. If a smart sensor triggers a harmful action based on a faulty AI decision, assigning legal responsibility can be complex. Ethical issues also abound. How should an AI prioritise actions in a resource-constrained situation? Who decides what constitutes a "fair" or "safe" outcome? In addition, the lack of universally accepted standards for trustworthy AI in IoT means developers often operate in regulatory grey areas, risking compliance violations and public distrust.

Addressing these trust challenges requires a holistic approach involving robust data governance, secure architectures, interpretable models, fairness audits, and alignment with evolving legal and ethical standards. The next section of this paper will explore proposed solutions and frameworks aimed at overcoming these barriers and enabling trustworthy AI in smart sensor systems.

The following conceptual model is proposed for the determination of the trustworthiness of artificial intelligence in the context of Internet of Things (IoT)-based smart sensor systems, encompassing a multitude of salient dimensions.

The Trustworthiness Score of an AI system in IoT is defined as follows:

$$T_{AI} = w_1 \cdot Q_D + w_2 \cdot R_A + w_3 \cdot S_C + w_4 \cdot P_V + w_5 \cdot E_X + w_6 \cdot F_B + w_7 \cdot C_L. \quad (1)$$

Where:

Q_D – Data Quality Score – Reflects completeness, accuracy, integrity, and timeliness of sensor data.

R_A – Robustness & Reliability – Measures system resilience under faults, noise, or adversarial inputs.

S_C – Security Confidence – Includes protection from cyberattacks, secure data transmission, etc.

P_V – Privacy Preservation Score – Reflects compliance with GDPR, CCPA, anonymisation practices, etc.

E_X – Explainability & Transparency – Measures clarity of AI decision-making and ability to audit.

F_B – Fairness & Bias Mitigation – Indicates bias resistance and equity in data/decisions.

C_L – Compliance & Legal Readiness – Alignment with legal standards, accountability, and documentation.

Weights can be domain-specific or adjusted based on risk profiles (e.g., safety-critical systems might assign higher w_2 or w_6)

$$w_i \in [0, 1], \text{ and } \sum w_i = 1.$$

Each factor Q_D, R_A, … should be evaluated on a **normalised scale** from 0 (low trust) to 1 (high trust), based on metrics, expert review, or benchmarks.

Example: For a manufacturing AI system:

- High data integrity: $Q_D = 0.9$
- Moderate explainability: $E_X = 0.6$
- Strong legal compliance: $C_L = 0.85$

With weights: $w_1 = 0.2$, $w_2 = 0.15$, $w_3 = 0.15$, $w_4 = 0.1$, $w_5 = 0.15$, $w_6 = 0.1$, $w_7 = 0.15$

Then, the overall T_{AI} can be computed according to Table 2.

Table 2

Computing of Total Trustworthiness Score

Factor	Symbol	Score (0-1)	Weight (0-1)	Weighted Score
Data Quality	Q_D	0,9	0,2	0,18
Robustness & Reliability	R_A	0,1	0,15	0,015
Security Confidence	S_C	0,4	0,15	0,06
Privacy Preservation	P_V	0,6	0,1	0,06
Explainability	E_X	0,6	0,15	0,09
Fairness & Bias	F_B	0,3	0,1	0,03
Legal Compliance	C_L	0,85	0,15	0,1275
Total Trustworthiness Score				**0,5625**

Solutions for Building Trustworthy AI in Smart Sensor Systems

In order to address the complex trust challenges present in AI-powered smart sensor systems, a combination of technical, organisational, and regulatory strategies is essential.

In the area of data governance and validation frameworks, it is advisable to use automated data quality checks in real-time. This will help to detect sensor drifts, anomalies and missing data in real-time. Sensor data logs should include calibration history, accuracy ratings, and environmental context. Redundant sensing will help improve data reliability.

To ensure data protection in Secure-by-Design Architectures, the implementation of end-to-end encryption is essential at both the edge sensor level and within the cloud environment. To optimise efficiency, it is crucial to employ resource-efficient security protocols, such as ECC and TLS-lite, particularly for devices with constrained capabilities. Limiting access through authentication at each system level is also important.

To ensure the explainability and transparency of AI models, it is essential to integrate explainable AI techniques such as SHAP, LIME, or decision trees to interpret the model's behaviour. Additionally, it is crucial to maintain audit trails of model inputs, outputs, and decisions over time. A human-in-the-loop (HITL) approach is recommended, wherein human operators can review critical AI decisions, such as anomaly detection and shutdown triggers.

It is advisable to use various data audit tools to ensure fairness and mitigate bias. The accuracy of the model's predictions should also be periodically assessed in terms of spatial, temporal, and demographic variables.

Increasing the robustness and reliability of models can be achieved by subjecting them to simulated disturbances or sensor failures to identify vulnerabilities, implementing self-healing mechanisms, and continuously training them.

To ensure legal, ethical, and regulatory compliance, it is essential to conduct structured assessments of how AI solutions impact people, the environment, and communities. Developments should align with key regulatory documents in the field, such as ISO/IEC 42001 (AI governance), the EU AI Act, and the NIST AI RMF.

Ensuring the clear display of information about certification, third-party audits, or regulatory compliance in the user interface will facilitate trusted communication and user engagement.

The following key solution pillars have been identified as instrumental in fostering trustworthy, resilient, and ethically aligned AI within such environments (Table 3).

Table 3

Solution mapping

Challenge	Solution Approach	Example Practice
Data Quality & Integrity	Redundant sensing, validation layers	Sensor fusion, data profiling tools
Security & Privacy	Secure-by-design principles	End-to-end encryption, zero trust architecture
Explainability	Use of XAI tools	SHAP/LIME visualisations, model cards
Fairness	Dataset audits & rebalancing	Bias detection tools, fairness metrics
Robustness	Adversarial testing, fallback logic	Fault simulation frameworks
Legal & Ethical	Compliance-by-design	Conformance to ISO/IEC 42001, EU AI Act

Summary and Conclusion

Intelligent sensor systems driven by artificial intelligence are becoming increasingly integrated into critical areas of human life, such as environmental monitoring, automated manufacturing, smart cities, and more. Given this integration, it is essential that the ability of these systems to make autonomous decisions is grounded in the principles of reliability, explainability, confidentiality, fairness, and legal responsibility. The research conducted in this article emphasizes the urgent need to establish trust in intelligent sensor systems controlled by artificial intelligence, especially in the context of the Internet of Things (IoT).

The development and implementation of key standards and regulations, such as the EU AI Act, the international standard ISO/IEC 42001, and the NIST AI RMF guidelines, underscores the significance of the issue and the multidimensional nature of trust in AI. Addressing concerns such as data quality, security risks, algorithmic bias, transparency, and regulatory uncertainty is crucial for ensuring the safe and ethical deployment of AI.

The proposed trustworthiness model and scoring framework offer a practical method for evaluating the trust level of AI systems in smart sensor environments. Furthermore, the identified solution pillars – ranging from explainable AI and secure architectures to fairness auditing and compliance-by-design – provide actionable guidance for developers, policymakers, and system integrators.

Achieving trust in artificial intelligence in the Internet of Things is not a one-time effort, but an ongoing interdisciplinary process that combines technical innovation with effective governance. Building trust at all levels – from data collection and model development to system oversight and user engagement – will ensure that AI technologies serve both industrial progress and social welfare.

Bibliography

1. Са́нчес П. М. С., Сельдра́н А. Х., Сє Н., Бове́ Ж., Пе́рес Г. М., Штї́ллер Б. (2024). Федеративна довіра: Рішення для надійного федеративного навчання. Future Generation Computer Systems, 152, 83–98. https://doi.org/10.1016/j.future.2023.10.013

2. Джагатісаперума́л С. К., Фам К'ю. Ві., Рубі Р., Ян Ч., Сюй Ц., Чжан Ч. (2022). Пояснюваний ШІ через Інтернет речей (IoT): Огляд, сучасний стан та майбутні напрямки. IEEE Open Journal of the Communications Society, 3, 2106–2136. https://doi.org/10.48550/arXiv.2211.01036

3. Насім, М. Д., Бісвас, П., Рашид, А., Бісвас, А., та Гупта, К. Д. (2024). Надійний XAI та його застосування. Препринт arXiv arXiv:2410.17139. https://doi.org/10.48550/arXiv.2410.17139

4. Лі, Б., Ці, П., Лю, Б., Ді, С., Лю, Дж., Пей, Дж., … та Чжоу, Б. (2023). Надійний ШІ: від принципів до практики. ACM Computing Surveys, 55(9), 1–46. https://doi.org/10.1145/3555803

5. Ван, Х., Ван, Б., Ву, Ю., Нін, З., Го, С., та Ю, Ф. Р. (2024). Опитування щодо надійного периферійного інтелекту: від безпеки та надійності до прозорості та сталого розвитку. IEEE Communications Surveys & Tutorials. https://doi.org/10.1109/COMST.2024.3446585

6. Раданлієв, П., Де Рур, Д., Мейпл, К., Нерс, Дж. Р., Ніколеску, Р., та Ані, У. (2024). Безпека штучного інтелекту та кіберризики в системах Інтернету речей. Frontiers in Big Data, 7, 1402745. https://doi.org/10.3389/fdata.2024.1402745

7. Черданцева, Ю., Бернап, П., Наджм-Техрані, С. та Джонс, К. (2022). Конфігурована модель залежностей SCADA-системи для цілеспрямованої оцінки ризиків. Appl. Sci. 12, 1–29. https://doi.org/10.3390/app12104880

8. Ерола, А., Аграфіотіс, І., Нерс, Дж. Р. К., Аксон, Л., Голдсміт, М. та Кріз, С. (2022). Система для розрахунку кіберцінності під ризиком. Comput. Secur. 113:102545. https://doi.org/10.1016/j.cose.2021.102545

9. Попеску, С. М., Мансур, С., Вані, О. А., Кумар, С. С., Шарма, В., Шарма, А., … & Чунг, Ю. С. (2024). Технології на основі штучного інтелекту та Інтернету речей для моніторингу та управління забрудненням навколишнього середовища. Frontiers in Environmental Science, 12. https://doi.org/10.3389/fenvs.2024.1336088

10. Ай, В., Чен, Г., Юе, Х., та Ван, Дж. (2023). Застосування гіперспектрального та глибокого навчання у виявленні мікропластику в ґрунті сільськогосподарських угідь. J. Hazard. Mater. 445, 130568. https://doi.org/10.1016/j.jhazmat.2022.130568

11. Аша, П., Натраян, Л., Гіта, Б., Беула, Дж. Р., Сумати, Р., Варалакшмі, Г. та ін. (2022). Інтернет речей, що дозволяє використовувати екологічну токсикологію для моніторингу забруднення повітря за допомогою методів штучного інтелекту. Environ. Res. 205, 112574. https://doi.org/0.1016/j.envres.2021.112574

References

1. Sánchez, P. M. S., Celdrán, A. H., Xie, N., Bovet, G., Pérez, G. M., & Stiller, B. (2024). Federated trust: A solution for trustworthy federated learning. *Future Generation Computer Systems*, 152, 83-98. https://doi.org/10.1016/j.future.2023.10.013

2. Jagatheesaperumal, S. K., Pham, Q. V., Ruby, R., Yang, Z., Xu, C., & Zhang, Z. (2022). Explainable AI over the Internet of Things (IoT): Overview, state-of-the-art and future directions. *IEEE Open Journal of the Communications Society*, 3, 2106–2136. https://doi.org/10.48550/arXiv.2211.01036

3. Nasim, M. D., Biswas, P., Rashid, A., Biswas, A., & Gupta, K. D. (2024). Trustworthy XAI and Application. *arXiv preprint arXiv*: 2410.17139. https://doi.org/10.48550/arXiv.2410.17139

4. Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J.,... & Zhou, B. (2023). Trustworthy AI: From principles to practices. *ACM Computing Surveys*, 55(9), 1–46. https://doi.org/10.1145/3555803

5. Wang, X., Wang, B., Wu, Y., Ning, Z., Guo, S., & Yu, F. R. (2024). A survey on trustworthy edge intelligence: From security and reliability to transparency and sustainability. *IEEE Communications Surveys & Tutorials*. https://doi.org/10.1109/COMST.2024.3446585

6. Radanliev, P., De Roure, D., Maple, C., Nurse, J. R., Nicolescu, R., & Ani, U. (2024). AI security and cyber risk in IoT systems. *Frontiers in Big Data*, 7, 1402745. https://doi.org/10.3389/fdata.2024.1402745

7. Cherdantseva, Y., Burnap, P., Nadjm-Tehrani, S., and Jones, K. (2022). A configurable dependency model of a SCADA system for goal-oriented risk assessment. *Appl. Sci*. 12, 1–29. https://doi.org/10.3390/app12104880

8. Erola, A., Agrafiotis, I., Nurse, J. R. C., Axon, L., Goldsmith, M., and Creese, S. (2022). A system to calculate cyber value-at-risk. *Comput. Secur. 113*: 102545. https://doi.org/10.1016/j.cose.2021.102545

9. Popescu, S. M., Mansoor, S., Wani, O. A., Kumar, S. S., Sharma, V., Sharma, A., … & Chung, Y. S. (2024). Artificial intelligence and IoT driven technologies for environmental pollution monitoring and management. *Frontiers in Environmental Science*, 12. https://doi.org/10.3389/fenvs.2024.1336088

10. Ai, W., Chen, G., Yue, X., and Wang, J. (2023). Application of hyperspectral and deep learning in farmland soil microplastic detection. *J. Hazard. Mater*. 445, 130568. https://doi.org/10.1016/j.jhazmat.2022.130568

11. Asha, P., Natrayan, L., Geetha, B., Beulah, J. R., Sumathy, R., Varalakshmi, G., et al. (2022). IoT enabled environmental toxicology for air pollution monitoring using AI techniques. *Environ. Res*. 205, 112574. https://doi.org/0.1016/j.envres.2021.112574

Дата першого надходження рукопису до видання: 22.09.2025
Дата прийнятого до друку рукопису після рецензування: 16.10.2025
Дата публікації:

УДК 615.47:621.9(043.3) DOI

О. В. ГОРБОВИЙ
магістр, асистент кафедри енергозберігаючих технологій
та енергетичного менеджменту
ЗВО «Подільський державний університет»
ORCID: 0000-0002-0762-8410

О. В. ЛІННИК
кандидат технічних наук, доцент,
доцент кафедри фізичних основ електронної техніки
факультету електронної та біомедичної інженерії
Харківський національний університет радіоелектроніки
ORCID: 0000-0002-4906-3796

І. М. ДІОРДІЦА
завідувач лабораторією кафедри комп'ютерно-інтегрованих технологій
виробництва приладів приладобудівного факультету
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
ORCID: 0000-0002-9088-289X

РОЛЬ 3D-ДРУКУ В ПЕРСОНАЛІЗОВАНІЙ РОЗРОБЦІ ПРОТЕЗІВ ТА ОРТЕЗІВ ДЛЯ ПАЦІЄНТІВ З ОБМЕЖЕНИМИ МОЖЛИВОСТЯМИ

Актуальність дослідження зумовлена потребою у створенні ефективних протезно-ортезних рішень для пацієнтів з обмеженими можливостями, що відповідають сучасним вимогам до якості життя, соціальної інтеграції та реабілітаційного супроводу. Традиційні методи виготовлення виробів часто ігнорують індивідуальні анатомічні та біомеханічні особливості, що знижує їхню ефективність. 3D-друк дає змогу інтегрувати цифрове моделювання, адитивне виробництво та біосумісні матеріали в єдиний технологічний цикл, що відкриває можливості для принципово нового рівня персоналізації.

Мета статті – комплексне наукове обґрунтування ролі технологій 3D-друку в персоналізованому проєктуванні та виготовленні протезів та ортезів для пацієнтів з обмеженими можливостями, а також у визначенні перспектив їхньої інтеграції в сучасну клінічну та реабілітаційну практику.

Методологія дослідження ґрунтується на системному аналізі принципів цифрового моделювання, параметричного проєктування та використання біосумісних матеріалів у поєднанні з адитивними технологіями. Застосовано методи порівняльного аналізу, моделювання функціональних характеристик виробів, а також узагальнення даних сучасних клінічних практик.

Дослідження показало, що адитивні технології підвищують точність відтворення анатомічних структур, функціональність і комфорт протезів та ортезів; економічні та організаційні переваги переходу до моделі «виробництва за запитом»; з'ясовано, що 3D-друк скорочує час виготовлення протезів та ортезів, знижує витрати й підвищує ефективність реабілітації; виокремлено основні технологічні, матеріалознавчі та нормативно-правові обмеження, які стримують масштабне впровадження технології в медичну практику.

Висновки підтверджують, що 3D-друк спроможний забезпечити високий рівень персоналізації виробів, підвищити якість реабілітації та розширити доступність сучасних медичних технологій. З'ясовано, що його впровадження потребує стандартизації цифрових протоколів, розвитку матеріалознавчої бази, автоматизації виробництва та створення мультидисциплінарних клініко-інженерних команд.

Перспективи подальших досліджень пов'язані з використанням цифрових двійників для оптимізації конструкцій, розробленням комбінованих технологій друку для інтеграції полімерних і композитних матеріалів, формуванням реєстрів клінічних результатів застосування 3D-друкованих виробів та вдосконаленням нормативно-економічних механізмів їхнього відшкодування. Це відкриває можливості для масштабної інтеграції 3D-друку в систему персоналізованої реабілітаційної медицини.

Ключові слова: *адитивні технології; біосумісні матеріали; цифрове моделювання; біомеханічна адаптація; реабілітаційна медицина.*

© Горбовий О. В., Лінник О. В., Діордіца І. М., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

O. V. GORBOVY
Master's Degree, Assistant at the Department of Energy-Saving Technologies
and Energy Management
Higher Educational Institution "Podillia State University"
ORCID: 0000-0002-0762-8410

O. V. LINNYK
PhD in Engineering, Associate Professor,
Associate Professor at the Department of Physical Foundations of Electronic Engineering,
Faculty of Electronic and Biomedical Engineering
Kharkiv National University of Radio Electronics
ORCID: 0000-0002-4906-3796

I. M. DIORDITSA
Laboratory Manager at the Department of Computer-Integrated Technologies
of Instrument Making, Faculty of Instrumentation Engineering
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"
ORCID: 0000-0002-9088-289X

THE ROLE OF 3D PRINTING IN THE PERSONALIZED DESIGN OF PROSTHESES AND ORTHESES FOR PATIENTS WITH DISABILITIES

The relevance of the study is determined by the urgent need to develop effective prosthetic and orthotic solutions for patients with disabilities that meet modern requirements for quality of life, social integration, and rehabilitation support. Traditional manufacturing methods often neglect individual anatomical and biomechanical features, which reduces their effectiveness. 3D printing enables the integration of digital modeling, additive manufacturing, and biocompatible materials into a unified technological cycle, creating opportunities for an entirely new level of personalization.

The aim of the article is to provide a comprehensive scientific justification for the role of 3D printing technologies in the personalized design and production of prostheses and orthoses for patients with disabilities, as well as to identify prospects for their integration into contemporary clinical and rehabilitation practice.

The research methodology is based on a systematic analysis of the principles of digital modeling, parametric design, and the use of biocompatible materials in combination with additive technologies. Methods applied include comparative analysis, modeling of the functional characteristics of devices, and the generalization of data from current clinical practices.

The study demonstrated that additive technologies increase the accuracy of reproducing anatomical structures, as well as the functionality and comfort of prostheses and orthoses. Economic and organizational advantages of the transition to an on-demand production model were identified. It was established that 3D printing reduces production time, lowers costs, and enhances rehabilitation outcomes. Key technological, material, and regulatory constraints that limit large-scale implementation of this technology in medical practice were also outlined.

The conclusions confirm that 3D printing can ensure a high level of personalization of devices, improve the quality of rehabilitation, and expand the accessibility of modern medical technologies. It was found that implementation requires the standardization of digital protocols, the development of material science, the automation of production processes, and the establishment of multidisciplinary clinical-engineering teams.

Future research perspectives are related to the use of digital twins for design optimization, the development of combined printing technologies to integrate polymer and composite materials, the creation of registries of clinical outcomes for 3D-printed devices, and the improvement of regulatory and economic mechanisms for their reimbursement. These directions open opportunities for the large-scale integration of 3D printing into the system of personalized rehabilitation medicine.

Key words: *additive technologies; biocompatible materials; digital modeling; biomechanical adaptation; rehabilitation medicine.*

Постановка проблеми

Проблема забезпечення пацієнтів з обмеженими можливостями ефективними протезами та ортезами набуває особливої актуальності в умовах зростання вимог до якості життя та соціальної інтеграції. Традиційні методи виготовлення таких виробів часто ігнорують індивідуальні анатомічні особливості, що спричиняє дискомфорт, обмежує функціональність та зменшує бажання пацієнтів ними користуватися. У цьому контексті 3D-друк є інноваційним інструментом, здатним забезпечити персоналізований підхід до створення протезно-ортезних систем. Завдяки цифровому моделюванню та адитивним технологіям з'являється можливість виготовляти вироби з високою точністю, використовуючи біосумісні матеріали та оптимізовані конструктивні рішення, що відповідають сучасним науковим уявленням про ергономіку та біомеханіку. Практична значущість цього напряму полягає в підвищенні рівня реабілітаційних можливостей, скороченні термінів адаптації та розширенні доступності медичних

технологій для широкого кола пацієнтів. Інтеграція 3D-друку у сферу протезування та ортезування породжує нові наукові й прикладні завдання, пов'язані з удосконаленням методів проєктування, добором матеріалів та створенням стандартизованих протоколів клінічного впровадження.

В українських реаліях цей напрям набуває особливого значення у зв'язку з війною, яка спричинила різке зростання кількості пацієнтів із бойовими травмами та ампутаціями. Для таких пацієнтів персоналізовані протези, створені за допомогою 3D-друку, стають не лише засобом фізичної реабілітації, а й важливим чинником соціальної інтеграції та повернення до активного життя. Водночас масштабне впровадження цієї технології стримується обмеженою матеріально-технічною базою, дефіцитом кваліфікованих спеціалістів та відсутністю чітких нормативно-правових механізмів сертифікації. У поєднанні з волонтерськими ініціативами та поодинокими проєктами у військових шпиталях це формує ситуацію, коли 3D-друк стає критично необхідним інструментом, але його потенціал реалізується лише частково. Саме тому дослідження та розвиток цієї технології в Україні мають не лише наукову й практичну доцільність, а й гуманітарну місію в умовах війни та післявоєнного відновлення.

Аналіз останніх досліджень і публікацій

Дослідження, присвячені застосуванню 3D-друку в персоналізованій розробці протезів і ортезів для пацієнтів з обмеженими можливостями, згруповано за чотирма основними напрямами. Перший напрям формує біомеханічно орієнтовані засади індивідуального дизайну. М. І. Литвиненко, Л. М. Рисована, В. В. Григорук та співавтори доводять, що урахування кінематичних і силових характеристик опорно-рухового апарату допомагає оптимізувати параметри жорсткості виробу, рівномірний розподіл контактних тисків та запобігти вторинним ускладненням, що відкриває можливості для використання параметричних CAD-шаблонів у 3D-друці [1]. Н. В. Семінська, О. С. Мусієнко, І. В. Слободянюк та співавтори фіксують основні проблеми виготовлення протезів нижніх кінцівок: кадровий дефіцит, нестачу стандартів та нерівномірний доступ до обладнання, пропонуючи організаційні рішення через типізацію процесів і локалізацію виробництва [2]. Ж. Д. С. Перейра (J. D. S. Pereira), А. С. М. С. Ксав'єр (A. S. M. S. Xavier), Р. Д. С. Монтейру (R. D. S. Monteiro) та співавтори в скопінґ-огляді показують, що 3D-друковані ортези й протези підвищують комфорт пацієнтів та скорочують строки виготовлення, але оцінювання клінічної ефективності залишається неоднорідним [3]. Отже, поєднання біомеханічної валідації з уніфікованими методами оцінки ефективності та впровадження стандартів передачі клінічних даних у CAD/CAM-системи є перспективним.

Другий напрям охоплює матеріалознавчо-технологічні рішення. С. Веннам (S. Vennam), К. Н. Віджаясанкар (K. N. Vijayasankar), Ф. Паті (F. Pati) порівнюють матеріали й техніки друку, підкреслюючи вплив параметрів виготовлення на біосумісність та жорсткість пристроїв [4]. Х. К. Банґа (H. K. Banga), П. Калра (P. Kalra), Р. М. Белокар (R. M. Belokar), Р. Кумар (R. Kumar) демонструють відтворюваність маршруту «сканування → моделювання → друк» для протезно-ортезних виробів та окреслюють потребу в контролі якості [5]. К. Х. Лі (K. H. Lee), Д. К. Кім (D. K. Kim), Й. Х. Ча (Y. H. Cha) та співавтори показують, що поєднання 3D-моделювання з друком сприяє швидкому виготовленню персоналізованих пристроїв у клінічній практиці [6]. М. Лейте (M. Leite), Б. Соареш (B. Soares), В. Лопеш (V. Lopes) та співавтори переводять вимоги персоналізованої медицини в конструктивні обмеження, придатні для реалізації методами 3D-друку [7]. Таким чином, доцільними є порівняльні випробування полімерів і композитів у реальних умовах, розробка методів неруйнівного контролю та оптимізація параметрів друку для балансу біосумісності й довговічності.

Третій напрям стосується платформ та промислового впровадження. Ф. Ґурський (F. Górski), Є. Денисенко (Y. Denysenko), В. Кучко (W. Kuczko) та співавтори презентують AutoMedPrint як конвеєр для персоналізованих виробів – від збору анатомічних даних до створення бібліотек готових рішень [8]. Р. Міклауш (R. Miclaus), А. Репановичі (A. Repanovici), Н. Роман (N. Roman) окреслюють потенціал PLA у виготовленні протезів та ортезів, акцентуючи на компромісі між технологічною зручністю та обмеженнями міцності [9]. П. П. Бортакур (P. P. Borthakur) акцентує на перспективах алгоритмічного дизайну, гібридних технологій та інтеграції сенсорики для створення «розумних» протезів [10]. А. Демеко (A. Demeco), Р. Форесті (R. Foresti), А. Фріццієро (A. Frizziero) та співавтори підтверджують, що 3D-друковані ортези верхніх кінцівок сприяють реабілітації після інсульту, даючи змогу швидко адаптувати конструкцію [11]. Для подальшого розвитку напряму необхідно формалізувати протоколи валідації виробів, розширити бібліотеки параметричних моделей та розробити сенсори для моніторингу навантажень під час експлуатації.

Четвертий напрям охоплює проєктуванням, орієнтоване на користувача, та аналіз клінічної ефективності. Ґ. Туманн (G. Thomann) та В. А. де Карвальо (V. A. de Carvalho) на прикладі дитини з геміплегією довели, що для персоналізації виробів необхідно поєднувати клінічні вимірювання зі швидким прототипуванням [12]. Р. Торсен (R. Thorsen), Ф. Борто (F. Bortot) та А. Караччоло (A. Caracciolo) описують модель ко-дизайну з пацієнтом, яка підвищує функціональність і прийнятність виробів [13]. Т. А. М. Оуд (T. A. M. Oud), Е. Лаццарі (E. Lazzari), Г. Й. Х. Дж. Гейсберс (H. J. H. Gijsbers) та співавтори в скопінґ-огляді доводять ефективність 3D-друкованих ортезів у разі травматичних і хронічних патологіях кисті, проте вказують на потребу в стандартизованих протоколах оцінки та довготривалих спостереженнях [14]. Отже, необхідно здійснювати мультицентрові клінічні

дослідження з уніфікованими кінцевими точками, оцінкою економічної ефективності та розвитком ко-дизайну для різних груп пацієнтів.

Попри значні успіхи в упровадженні адитивних технологій у протезуванні та ортезуванні, залишається низка нерозв'язаних проблем. Зокрема, недостатньо досліджень щодо інтеграції 3D-друку з урахуванням індивідуальних анатомічних особливостей пацієнтів, що знижує точність та ергономіку виробів. Обмеженим є розуміння впливу технологій на функціональність і комфорт користувачів, а також відсутні системні оцінки економічних та організаційних ефектів. Крім того, додаються технологічні й матеріалознавчі обмеження та відсутність єдиних нормативно-правових стандартів, що істотно гальмує клінічне застосування.

Запропоноване дослідження має на меті заповнити ці прогалини через системний аналіз принципів застосування 3D-друку, вивчення впливу на якісні характеристики виробів, оцінку економічних та організаційних переваг інтеграції технології в реабілітаційні процеси. Крім того, дослідження допомогло виокремити основні проблеми упровадження технологій та запропонувати практичні рекомендації для їхнього розв'язання. Це сприятиме створенню цілісної наукової бази, що забезпечить як подальший розвиток галузі, так і підвищення ефективності клінічної практики.

Формулювання мети дослідження

Мета статті – наукове обґрунтування ролі 3D-друку в персоналізованій розробці протезів та ортезів для пацієнтів з обмеженими можливостями та визначення перспектив їхнього застосування в сучасній медичній практиці.

Завдання дослідження:

1. Дослідити принципи застосування 3D-друку в протезуванні та ортезуванні й оцінити його вплив на точність, функціональність і комфорт виробів.
2. Проаналізувати економічні й організаційні переваги та виокремити основні технологічні, матеріалознавчі й нормативні обмеження інтеграції адитивних технологій у медичну практику.
3. Сформулювати практичні рекомендації з удосконалення проєктування, виготовлення та клінічного застосування персоналізованих протезно-ортезних систем.

Викладення основного матеріалу дослідження

У сучасних умовах розвитку медичної інженерії принципи застосування 3D-друку в протезуванні та ортезуванні набувають стратегічного значення. Технології адитивного виробництва дають змогу відтворювати складні анатомічні форми з високим ступенем персоналізації, що неможливо забезпечити традиційними методами. Основою цього підходу є інтеграція цифрового моделювання та біосумісних матеріалів, які в поєднанні з методами сканування тіла забезпечують точність відтворення форми й функціональних характеристик кінцівок та опорно-рухових структур. Це відкриває нові можливості у сфері персоналізованої медицини, де основним завданням є не лише заміщення втраченої функції, а й створення умов для максимальної ергономіки, комфорту та адаптивності пацієнта (табл. 1).

Таблиця 1

Принципи застосування 3D-друку в медичному протезуванні та ортезуванні

Принцип	Сутність	Практичне значення	Прогнозований результат
Персоналізація	Використання індивідуальних цифрових моделей на основі 3D-сканування	Точне врахування анатомічних особливостей пацієнта	Оптимальна сумісність та комфорт
Біосумісність матеріалів	Застосування полімерів, композитів і біоматеріалів, які не спричиняють відторгнення організмом	Зниження ризиків ускладнень і подразнень	Підвищення безпеки та довговічності виробу
Функціональна адаптація	Моделювання виробу з урахуванням біомеханічних параметрів руху	Забезпечення природності рухів та підтримки	Підвищення якості життя пацієнта
Висока точність відтворення	Застосування адитивних технологій із високою роздільною здатністю	Мінімізація відхилень від заданих параметрів	Скорочення термінів адаптації

Джерело: сформовано на основі [1, с. 90–92; 2, с. 12–14; 6, с. 528–529; 7, с. 458–460; 9, с. 99–100]

У сучасній клінічній практиці ці принципи реалізуються завдяки поєднанню високоточної цифрової діагностики та можливостей адитивного виробництва. Персоналізація передбачає використання тривимірного сканування кінцівок, що дає змогу врахувати навіть мінімальні відхилення у формі чи об'ємі. У таких випадках клініцисти застосовують системи Artec Eva чи Structure Sensor, які дають можливість формувати точні цифрові моделі для подальшого проєктування в програмному середовищі Autodesk Fusion 360 або SolidWorks Medical [14]. Це принципово змінює підхід до протезування, оскільки індивідуальні особливості пацієнта враховуються ще на етапі моделювання, а не після виготовлення стандартного виробу. Застосування біосумісних матеріалів, таких як поліамід PA2200, термопластичний поліуретан або композити на основі вуглеволокна, зменшує ризик алергічних реакцій та подразнень шкіри, що часто виникали під час використання традиційних металевих або жорстких пластикових конструкцій. Це особливо важливо для дитячої ортопедії, де адаптація до ортеза чи протеза є складним як психологічним, так і фізіологічним процесом. На практиці функціональна адаптація означає створення

виробів, здатних відтворювати природні рухи. Так, під час проєктуванні ортезів для спортсменів використовується параметричне моделювання, яке враховує траєкторію руху суглобів, а у випадках протезування верхніх кінцівок застосовуються біонічні рішення, де 3D-друковані деталі поєднуються з електроміографічними сенсорами. Це дає змогу забезпечити не лише механічну підтримку, а й інтеграцію з нейром'язовим контролем.

Висока точність відтворення, притаманна технологіям Selective Laser Sintering (SLS) чи Fused Deposition Modeling (FDM), мінімізує потребу в додатковій ручній обробці та корекції [15]. Наприклад, у клініках Німеччини та США завдяки 3D-друку період між первинним обстеженням та остаточним встановленням протеза скорочено з кількох тижнів до 5–7 діб [16], що кардинально впливає на ефективність реабілітації та знижує витрати на медичні послуги.

У сучасній медицині адитивні технології є визначальним чинником підвищення точності, функціональності та комфорту протезно-ортезних виробів. Традиційні методи виробництва передбачали серійний підхід, що часто не враховував індивідуальних особливостей пацієнта, тоді як 3D-друк дає змогу створювати пристрої з високим ступенем персоналізації. Точність виготовлення досягається завдяки поєднанню тривимірного сканування з алгоритмами параметричного моделювання, що мінімізує похибки та оптимально відтворює анатомічні форми. Функціональність виробів зростає завдяки можливості проєктувати конструкції зі змінними зонами жорсткості, вентиляційними каналами чи інтегрованими сенсорами. Крім того, легкі матеріали та ергономічні форми забезпечують комфорт, що прискорює адаптацію пацієнтів і підвищує ефективність реабілітаційних програм (табл. 2).

Таблиця 2

Вплив адитивних технологій на характеристики протезно-ортезних виробів

Показник	Вплив	Результат
Точність	Використання 3D-сканування та CAD/CAE-систем у поєднанні з високороздільним друком	Мінімізація похибок, точне повторення анатомії
Функціональність	Можливість створення складних конструкцій, інтеграції датчиків та біомеханічних елементів	Відновлення природності рухів, розширення функцій
Комфорт	Застосування легких, гнучких та біосумісних матеріалів, можливість створення вентиляційних структур	Підвищення зручності носіння, зменшення подразнень, швидка адаптація

Джерело: сформовано на основі [4; 5, с. 15–18; 8; 10; 11; 12, с. 190–192; 13, с. 308–310; 14]

Практичне застосування адитивних технологій показує, що їхній вплив на точність, функціональність і комфорт виходить за межі простого відтворення анатомічних форм. Точність сучасних протезно-ортезних рішень забезпечується не лише 3D-скануванням, а й математичним моделюванням внутрішніх структур виробу, що оптимізують розподіл навантаження. Такі конструкції дають змогу створювати легкі, але міцні елементи з топологічною оптимізацією, де надлишковий матеріал усувається ще на етапі проєктування. Це особливо важливо для ортезів нижніх кінцівок, які мають бути одночасно стабільними й зручними для тривалого використання.

Функціональність сучасних виробів підвищується завдяки можливості інтеграції додаткових елементів у їхню структуру ще під час друку. Наприклад, в ортезах для спортсменів або військовослужбовців створюються спеціальні вставки для датчиків контролю навантаження, які дають змогу відстежувати біомеханіку рухів у реальному часі. У протезуванні активно застосовують модульні підходи, що допомагає швидко додавати або замінювати функціональні блоки в основній конструкції, зокрема змінні робочі насадки або сенсорні модулі.

Комфорт забезпечується можливістю створювати індивідуальні мікропористі або вентиляційні структури, які значно зменшують нагрівання та потовиділення в місцях контакту виробу зі шкірою. У дитячій ортопедії це особливо важливо, адже під час росту кінцівок виріб не повинен спричиняти подразнень чи обмежувати розвиток [3]. Дослідження показують, що використання 3D-друкованих ортезів із вентиляційними візерунками зменшує ризик дерматологічних ускладнень і водночас підвищує тривалість безперервного носіння.

Так, сучасні адитивні технології дають змогу не тільки досягти високої точності у відтворенні форми, але й інтегрувати нові функціональні можливості та забезпечити принципово новий рівень комфорту для пацієнтів. Це формує новий підхід, де протезно-ортезні вироби виходять за межі традиційного функціоналу й перетворюються на високотехнологічні інструменти для реабілітації й покращення якості життя.

У сучасній системі реабілітаційної допомоги інтеграція 3D-друку створює нову систему виготовлення та обслуговування протезно-ортезних виробів, переводячи процеси від складського виробництва до моделі «виробництва за запитом». Вартісні переваги виникають завдяки скороченню циклу постачання, зменшенню матеріальних втрат і трудомісткості операцій, а організаційні – через цифровізацію ланцюга створення цінності, прозорість якості та гнучкість планування ресурсів. Для медичних закладів це означає перерозподіл витрат від запасів та довгих підрядів до керованих операційних витрат, для пацієнта – швидший доступ до персоналізованих рішень та зниження непрямих витрат, пов'язаних з очікуванням та повторними візитами. Водночас для досягнення сталого ефекту необхідні стандартизовані робочі процеси, інтегрований контроль якості, підготовка персоналу та відповідні контракти з платниками медичних послуг (табл. 3).

Таблиця 3

Економічні та організаційні переваги інтеграції 3D-друку в реабілітаційні процеси

Категорія переваги	Організаційний механізм	Економічний ефект	Показники оцінювання
Скорочення циклу виготовлення	Перехід на цифрове замовлення та «виробництво за запитом» у клініці чи мережі	Менше часу простою, швидше відновлення працездатності пацієнта	Lead time, тривалість маршруту пацієнта, кількість повторних візитів
Зменшення запасів і відходів	Відмова від складських «типорозмірів», друк лише необхідних компонентів	Зниження оборотного капіталу, менше списань	Обіг запасів, відсоток браку / відходів, вартість утилізації
Оптимізація праці	Перерозподіл задач: від ручних операцій до контрольованих цифрових етапів	Менша трудомісткість, прогнозованість витрат на персонал	Нормогодини на виріб, продуктивність на фахівця
Масштабування та гнучкість	«Hub-and-spoke» модель: централізоване проєктування, локальний друк	Зниження логістичних витрат, стійкість до збоїв постачання	Вартість логістики на виріб, частка локально виготовлених виробів
Прозорість якості й відповідність	Вбудовані цифрові протоколи контролю та трасування	Зниження рекламацій, нижчі витрати на гарантійне обслуговування	Частота корекцій, відсоток гарантійних випадків, час на аудит
Система відшкодування	Стандартизовані специфікації та калькуляції для контрактів із платниками	Передбачуваність тарифів, скорочення «непокритих» витрат	Рівень відшкодування, час від подання до оплати, частка відмов
Інновації сервісу	Віддалені примірки та корекції за цифровими даними	Менше поїздок пацієнта, менше непрямих витрат	Кілометри / години, зекономлені на візитах; задоволеність пацієнтів

Джерело: сформовано на основі [3; 8; 14]

На практиці цикл виготовлення скорочується завдяки повному цифровому контуру: від збору клінічних даних до автоматизованої підготовки виробничих файлів, що дає змогу запускати друк у межах однієї робочої зміни. Модель «виробництва за запитом» принципово змінює структуру витрат: замість підтримання запасів напівфабрикатів заклад інвестує в керований парк обладнання та цифрові бібліотеки перевірених модулів, що знижує заморожений оборотний капітал і ризики списань унаслідок морального старіння [8]. Відмова від «типорозмірів» зменшує кількість відходів матеріалів і невдалих примірок; економія посилюється застосуванням топологічно оптимізованих конструкцій, які потребують менше сировини без втрати міцності. Запровадження стандартизованих цифрових етапів у робочому процесі мінімізує частку ручних, варіабельних операцій; продуктивність бригади зростає завдяки паралельному виконанню проєктування, друку та постобробки, а також прозорій черзі завдань у системі управління виробництвом. Для мережевих провайдерів ефективною є організація за принципом «hub-and-spoke» [18]: компетентний центр відповідає за валідацію моделей і сценаріїв, а локальні майданчики здійснюють друк і базову обробку. Це знижує транспортні витрати, прискорює сервіс і підвищує стійкість до збоїв у постачанні. У вартість якості закладається вбудована трасованість: кожний виріб має цифровий паспорт із параметрами друку та контрольними вимірами, що спрощує аудит, скорочує час на з'ясування причин відхилень і знижує частоту рекламацій [14]. У взаємодії з платниками запровадження стандартизованих специфікацій і калькуляцій полегшує переговори щодо тарифів і прискорює відшкодування; одночасно телесупровід примірок і корекцій скорочує непрямі витрати пацієнтів, підвищуючи прихильність до реабілітації та довгострокову результативність програми.

Упровадження 3D-друку в клінічну практику супроводжується низкою системних проблем, що охоплюють технологічний, матеріалознавчий та нормативно-правовий рівні. На технологічному рівні основним викликом є нестабільність параметрів друку, що спричиняє варіативність геометрії та механічних властивостей виробів. Обмежена продуктивність та масштабованість обладнання роблять масове застосування технології неможливим без втрати високої точності, а потреба в спеціалізованій постобробці збільшує час і вартість виготовлення [5, с. 15–18]. Додатковою проблемою є дефіцит кваліфікованих кадрів, здатних інтегрувати інженерні алгоритми та клінічні вимоги в єдиний процес [10]. Матеріалознавчі аспекти обмежуються вузьким спектром біосумісних полімерів та композитів, що відповідають клінічним стандартам. Крім того, більшість доступних матеріалів демонструє недостатню довговічність у реальних умовах експлуатації. Існують також проблеми з відтворюваністю властивостей партій матеріалів і відсутністю універсальних стандартів їхньої сертифікації [4]. Використання металевих порошків і керамічних композитів має високу собівартість та ускладнює контроль їхньої мікроструктури, що впливає на прогнозованість поведінки виробів. Нормативно-правові обмеження пов'язані з відсутністю єдиних протоколів клінічного випробування та стандартизованих регламентів сертифікації 3D-друкованих протезно-ортезних виробів. У більшості країн відсутні спеціалізовані вимоги до адитивних технологій, що затримує офіційне впровадження таких виробів у практику [3; 11]. Проблемними залишаються питання правової відповідальності у разі дефекту виробу, а також інтеграція 3D-друку в систему страхового відшкодування медичних послуг [12, с. 190–192; 13, с. 308–310].

В українських умовах проблема впровадження 3D-друку у клінічну практику набуває особливої гостроти у зв'язку з війною. Щодня зростає кількість пацієнтів із бойовими травмами, ампутаціями та складними пораненнями, які потребують індивідуалізованих протезно-ортезних виробів. 3D-друк дозволяє значно скоротити терміни виготовлення таких конструкцій, забезпечити персоналізацію під анатомічні особливості пацієнта та зменшити вартість у порівнянні з традиційними імпортними виробами. Водночас саме війна актуалізувала брак обладнання

та матеріалів: значна їх частина імпортується, а логістичні обмеження і валютні коливання призводять до подорожчання та перебоїв у постачанні. У регіонах, що зазнали руйнувань, виникає додатковий виклик відсутності належної інфраструктури для масштабного використання адитивних технологій.

Проблемним залишається кадрове забезпечення: в Україні відчувається дефіцит інженерів-біомедиків та клінічних спеціалістів, які володіють навичками 3D-моделювання та розуміють специфіку інтеграції таких виробів у процес лікування і реабілітації. Освітні програми лише частково враховують потреби військової медицини, а волонтерські ініціативи не можуть забезпечити системного ефекту. Водночас у низці військових шпиталів та науково-дослідних центрів уже реалізуються проєкти з виготовлення індивідуальних протезів та хірургічних шаблонів, що доводить практичну доцільність технології.

З нормативно-правової точки зору в Україні відсутні окремі стандарти, які б регламентували виробництво та сертифікацію 3D-друкованих медичних виробів. Це ускладнює офіційне впровадження таких продуктів у систему охорони здоров'я та обмежує можливість їх покриття страховими програмами чи державними закупівлями. За умов війни ця прогалина стає критичною, адже обсяги потреб зростають щодня, тоді як механізмів швидкої легалізації та масштабування технології бракує.

Сукупність цих чинників створює комплексну проблему, що стримує широке застосування технології в клінічній практиці навіть за наявності її очевидних переваг. Проте саме український контекст демонструє, що попит на 3D-друковані медичні вироби вже перевищує можливості традиційних методів і робить цю технологію не лише перспективною, а й критично необхідною для системи охорони здоров'я в умовах воєнного та післявоєнного відновлення.

Удосконалення процесів проєктування, виготовлення та клінічного використання персоналізованих протезів та ортезів на основі 3D-друку потребує комплексного підходу, що поєднує технічні, організаційні та клінічні рішення. На етапі проєктування доцільним є впровадження стандартизованих цифрових протоколів сканування та моделювання, які забезпечують високу відтворюваність результатів і мінімізують розбіжності між віртуальною моделлю та готовим виробом. Застосування алгоритмів топологічної оптимізації дає змогу зменшити вагу конструкцій без втрати їхньої міцності, а інтеграція біомеханічного моделювання дозволяє індивідуалізувати геометрію виробу відповідно до кінематичних характеристик руху пацієнта.

У виробничих процесах важливо застосовувати перевірені біосумісні матеріали з підтвердженою довговічністю та стійкістю до стерилізаційних процедур, а також розвивати комбіновані методики друку, які поєднують полімерні та композитні сегменти в межах одного виробу. Використання адитивних технологій нового покоління з багатоматеріальним друком відкриває можливості створення функціонально градієнтних структур, що краще відповідають біомеханічним навантаженням. Перехід на автоматизовані виробничі лінії та впровадження промислових систем контролю якості зменшує варіативність продукції та скорочує потребу в ручній постобробці.

Для ефективної роботи на клінічному рівні необхідним є створення інтегрованих команд, що об'єднують лікарів, інженерів та біомеханіків для спільного ухвалення рішень на кожному етапі. Систематичне проведення клінічних випробувань і формування національних реєстрів результатів використання 3D-друкованих виробів забезпечить накопичення доказової бази та сприятиме гармонізації з міжнародними стандартами. Важливим напрямом є адаптація страхових і відшкодувальних механізмів до специфіки персоналізованих виробів, що підвищить їхню доступність для широкого кола пацієнтів.

Отже, практичні рекомендації зводяться до стандартизації цифрових протоколів, розвитку матеріалознавчої бази, автоматизації виробництва, мультидисциплінарної співпраці та оновлення регуляторно-економічних механізмів. У комплексі це формує передумови для сталого, масштабованого та економічно обґрунтованого впровадження 3D-друку в реабілітаційну медицину.

Висновки

Дослідження підтвердило, що 3D-друк є важливим інструментом персоналізованого протезування та ортезування, оскільки поєднує цифрове сканування, алгоритмічне моделювання та біосумісні матеріали, забезпечуючи точність відтворення анатомічних структур, підвищену функціональність виробів і комфорт пацієнтів

Проте виявлено серйозні перешкоди. Серед технологічних – нестабільність параметрів друку, потреба в постобробці та обмежена масштабованість обладнання. Матеріалознавчі проблеми полягають у вузькому виборі сертифікованих біосумісних матеріалів і недовговічності більшості полімерів у клінічних умовах. Нормативно-правові охоплюють відсутність уніфікованих стандартів випробувань, недосконалість сертифікації та неврегульованість механізмів страхового відшкодування

Запропоновані практичні рекомендації передбачають стандартизацію цифрових протоколів сканування й моделювання та розвиток матеріальної бази з акцентом на довговічні біосумісні матеріали. Додатково необхідні автоматизації виробництва, створення мультидисциплінарних команд та адаптація страхових механізмів.

Перспективи подальших досліджень пов'язані з використанням цифрових двійників для оптимізації конструкцій, комбінованими технологіями друку для інтеграції різних матеріалів, створенням реєстрів клінічних результатів та вдосконаленням нормативної бази. Це відкриває можливості для системної інтеграції 3D-друку в персоналізовану реабілітаційну медицину.

Список використаної літератури

1. Литвиненко М. І., Рисована Л. М., Григорук В. В., Алексеєнко Р. В., Гранкіна С. С. Вплив біомеханіки на оптимізацію дизайну та функціональності протезів і ортезів. *Новий колегіум*. 2024. Вип. 3, № 115. С. 89–95. DOI: https://doi.org/10.34142/nc.2024.3.89

2. Семінська Н. В., Мусієнко О. С., Слободянюк І. В., Белевець К. С., Степанова А. А., Шитікова Н. С. Виготовлення протезів нижніх кінцівок: виклики, аналіз та можливі рішення. *Біомедична інженерія і технологія*. 2024. № 14. С. 8–17. DOI: https://doi.org/10.20535/2617-8974.2024.14.303997

3. Pereira J. D. S., Xavier A. S. M. S., Monteiro R. D. S., Cruz V. V., Pereira M. F. D. S., Tholl A. D., Machado W. C. A. 3D-printed orthoses and prostheses for people with physical disability in rehabilitation centers: a scoping review. *BMC Musculoskeletal Disorders*. 2024. Vol. 25, № 1. Article 783. DOI: https://doi.org/10.1186/s12891-024-07875-3

4. Vennam S., Vijayasankar K. N., Pati F. 3D printed personalized assistive devices: a material, technique, and medical condition perspective. *Applied Materials Today*. 2024. Vol. 40. Article 102403. DOI: https://doi.org/10.1016/j.apmt.2024.102403

5. Banga H. K., Kalra P., Belokar R. M., Kumar R. Design and fabrication of prosthetic and orthotic product by 3D printing. In: *Prosthetics and Orthotics*. IntechOpen, 2020. DOI: https://doi.org/10.5772/intechopen.94846

6. Lee K. H., Kim D. K., Cha Y. H., Kwon J. Y., Kim D. H., Kim S. J. Personalized assistive device manufactured by 3D modelling and printing techniques. *Disability and Rehabilitation: Assistive Technology*. 2019. Vol. 14, № 5. P. 526–531. DOI: https://doi.org/10.1080/17483107.2018.1494217

7. Leite M., Soares B., Lopes V., Santos S., Silva M. T. Design for personalized medicine in orthotics and prosthetics. *Procedia CIRP*. 2019. Vol. 84. P. 457–461. DOI: https://doi.org/10.1016/j.procir.2019.04.254

8. Górski F., Denysenko Y., Kuczko W., Żukowska M., Wichniarek R., Zawadzki P., Rybarczyk J. Individualized 3D printed orthopaedic and prosthetic devices using AutoMedPrint technology – Methodologies and examples. *Advances in Science and Technology. Research Journal*. 2024. Vol. 18, № 6. DOI: https://doi.org/10.12913/22998624/191548

9. Miclaus R., Repanovici A., Roman N. Biomaterials: Polylactic acid and 3D printing processes for orthosis and prosthesis. *Materiale Plastice*. 2017. Vol. 54, № 1. P. 98–102. DOI: https://doi.org/10.37358/MP.17.1.4794

10. Borthakur P. P. The Role and Future Directions of 3D Printing in Custom Prosthetic Design. *Engineering Proceedings*. 2024. Vol. 81, № 1. Article 10. DOI: https://doi.org/10.3390/engproc2024081010

11. Demeco A., Foresti R., Frizziero A., Daracchi N., Renzi F., Rovellini M., Costantino C. The upper limb orthosis in the rehabilitation of stroke patients: the role of 3D printing. *Bioengineering*. 2023. Vol. 10, № 11. Article 1256. DOI: https://doi.org/10.3390/bioengineering10111256

12. Thomann G., de Carvalho V. A. Personalized upper limb orthosis necessitates variety of tools during the development process: hemiplegic child case study. *Disability and Rehabilitation: Assistive Technology*. 2021. Vol. 16, № 2. P. 188–195. DOI: https://doi.org/10.1080/17483107.2019.1646820

13. Thorsen R., Bortot F., Caracciolo A. From patient to maker – a case study of co-designing an assistive device using 3D printing. *Assistive Technology*. 2021. Vol. 33, № 6. P. 306–312. DOI: https://doi.org/10.1080/10400435.2019.1634660

14. Oud T. A. M., Lazzari E., Gijsbers H. J. H., Gobbo M., Nollet F., Brehm M. A. Effectiveness of 3D-printed orthoses for traumatic and chronic hand conditions: A scoping review. *PLoS One*. 2021. Vol. 16, № 11. e0260271. DOI: https://doi.org/10.1371/journal.pone.0260271

15. Reverse Engineering using an Artec 3D Scanner with SOLIDWORKS. *Javelin*. 2025. URL: https://www.javelin-tech.com/blog/2019/03/reverse-engineering-solidworks/ (date of access: 09.09.2025).

16. Prosthetic Production Takes a Step Forward With the 3D Printed NOVA Foot. *Formlabs*. 2025. URL: https://formlabs.com/global/blog/sls-nova-foot/ (date of access: 09.09.2025).

17. Reducing prosthesis fitting time through additive manufacturing. *Manufacturing PA Innovation Program*. 2025. URL: https://manufacturingpa.org/news/2022/08/rmu-union-orthotics.html (date of access: 09.09.2025).

18. Hub-and-spoke. 2025. URL: https://hubandspoke.works/ (date of access: 09.09.2025).

References

1. Lytvynenko, M. I., Rysovana, L. M., Hryhoruk, V. V., Alekseienko, R. V., & Hrankina, S. S. (2024). Vplyv biomekhaniky na optymizatsiiu dyzainu ta funktsionalnosti proteziv i orteziv [The influence of biomechanics on the optimization of the design and functionality of prostheses and orthoses]. *Novyi kolehium – New Collegium, 3*(115), 89–95. https://doi.org/10.34142/nc.2024.3.89 [in Ukrainian].

2. Seminska, N. V., Musiienko, O. S., Slobodianiuk, I. V., Belevets, K. S., Stepanova, A. A., & Shytikova, N. S. (2024). Vyhotovlennia proteziv nyzhnikh kinivok: vyklyky, analiz ta mozhlyvi rishennia [Manufacturing of lower limb prostheses: challenges, analysis and possible solutions]. *Biomedychna inzheneriia i tekhnolohiia – Biomedical Engineering and Technology, 14*, 8–17. https://doi.org/10.20535/2617-8974.2024.14.303997 [in Ukrainian].

3. Pereira, J. D. S., Xavier, A. S. M. S., Monteiro, R. D. S., Cruz, V. V., Pereira, M. F. D. S., Tholl, A. D., & Machado, W. C. A. (2024). 3D-printed orthoses and prostheses for people with physical disability in rehabilitation centers: A scoping review. *BMC Musculoskeletal Disorders, 25*(1), Article 783. https://doi.org/10.1186/s12891-024-07875-3

4. Vennam, S., Vijayasankar, K. N., & Pati, F. (2024). 3D printed personalized assistive devices: A material, technique, and medical condition perspective. *Applied Materials Today, 40*, Article 102403. https://doi.org/10.1016/j.apmt.2024.102403

5. Banga, H. K., Kalra, P., Belokar, R. M., & Kumar, R. (2020). Design and fabrication of prosthetic and orthotic product by 3D printing. In *Prosthetics and Orthotics*. IntechOpen. https://doi.org/10.5772/intechopen.94846

6. Lee, K. H., Kim, D. K., Cha, Y. H., Kwon, J. Y., Kim, D. H., & Kim, S. J. (2019). Personalized assistive device manufactured by 3D modelling and printing techniques. *Disability and Rehabilitation: Assistive Technology, 14*(5), 526–531. https://doi.org/10.1080/17483107.2018.1494217

7. Leite, M., Soares, B., Lopes, V., Santos, S., & Silva, M. T. (2019). Design for personalized medicine in orthotics and prosthetics. *Procedia CIRP, 84*, 457–461. https://doi.org/10.1016/j.procir.2019.04.254

8. Górski, F., Denysenko, Y., Kuczko, W., Żukowska, M., Wichniarek, R., Zawadzki, P., & Rybarczyk, J. (2024). Individualized 3D printed orthopaedic and prosthetic devices using AutoMedPrint technology – Methodologies and examples. *Advances in Science and Technology. Research Journal, 18*(6). https://doi.org/10.12913/22998624/191548

9. Miclaus, R., Repanovici, A., & Roman, N. (2017). Biomaterials: Polylactic acid and 3D printing processes for orthosis and prosthesis. *Materiale Plastice, 54*(1), 98–102. https://doi.org/10.37358/MP.17.1.4794

10. Borthakur, P. P. (2024). The role and future directions of 3D printing in custom prosthetic design. *Engineering Proceedings, 81*(1), Article 10. https://doi.org/10.3390/engproc2024081010

11. Demeco, A., Foresti, R., Frizziero, A., Daracchi, N., Renzi, F., Rovellini, M., & Costantino, C. (2023). The upper limb orthosis in the rehabilitation of stroke patients: The role of 3D printing. *Bioengineering, 10*(11), Article 1256. https://doi.org/10.3390/bioengineering10111256

12. Thomann, G., & de Carvalho, V. A. (2021). Personalized upper limb orthosis necessitates variety of tools during the development process: Hemiplegic child case study. *Disability and Rehabilitation: Assistive Technology, 16*(2), 188–195. https://doi.org/10.1080/17483107.2019.1646820

13. Thorsen, R., Bortot, F., & Caracciolo, A. (2021). From patient to maker – A case study of co-designing an assistive device using 3D printing. *Assistive Technology, 33*(6), 306–312. https://doi.org/10.1080/10400435.2019.1634660

14. Oud, T. A. M., Lazzari, E., Gijsbers, H. J. H., Gobbo, M., Nollet, F., & Brehm, M. A. (2021). Effectiveness of 3D-printed orthoses for traumatic and chronic hand conditions: A scoping review. *PLoS One, 16*(11), e0260271. https://doi.org/10.1371/journal.pone.0260271

15. Reverse engineering using an Artec 3D scanner with SOLIDWORKS. (2025). *Javelin*. Retrieved from https://www.javelin-tech.com/blog/2019/03/reverse-engineering-solidworks/

16. Prosthetic production takes a step forward with the 3D printed NOVA Foot. (2025). *Formlabs*. Retrieved from https://formlabs.com/global/blog/sls-nova-foot/

17. Reducing prosthesis fitting time through additive manufacturing. (2025). *Manufacturing PA Innovation Program*. Retrieved from https://manufacturingpa.org/news/2022/08/rmu-union-orthotics.html

18. Hub-and-spoke. (2025). Retrieved from https://hubandspoke.works/

Дата першого надходження рукопису до видання: 20.09.2025
Дата прийнятого до друку рукопису після рецензування: 14.10.2025
Дата публікації:

УДК 004.722 DOI

М. А. ГУНЬКО
аспірант кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0000-0002-8011-0693

МЕТОД СТРУКТУРНО-ТОПОЛОГІЧНОГО СИНТЕЗУ АРХІТЕКТУРИ FOG-СЕРЕДОВИЩА

У дослідженні представлено метод проєктування архітектури Fog-середовища на основі структурно-топологічного синтезу. Ефективність системи визначається тим, наскільки вдало розподілено обчислювальні вузли та сервіси між рівнями Edge, Fog і Cloud. Організація в такій формі підвищує стабільність роботи: зменшуються затримки під час передавання даних, знижується навантаження на мережу, а результати обчислень викликають більшу довіру. У випадках застосувань реального часу метод має особливу цінність, адже навіть кількасекундна затримка може впливати на функціонування всієї системи. Розподіл завдань між рівнями підвищує продуктивність та запобігає перевантаженню окремих компонентів. Дослідження наголошує на важливості співвідношення між ефективністю роботи та витратами. Під час оцінки беруться до уваги пропускна здатність каналів, обчислювальна потужність вузлів і споживання енергії. У результаті формується архітектура, здатна до масштабування та стійка навіть за умов часткової відмови компонентів. Поєднання запропонованого методу з технологіями Інтернету речей та хмарними сервісами демонструє його практичну користь. Завдяки цьому організації та кінцеві користувачі отримують доступ до актуальної інформації й можуть оперативно реагувати на зміни довкілля. Приклади впровадження вже спостерігаються у транспортних системах, медичних сервісах, промисловій автоматизації та екологічному моніторингу. Розроблений підхід також створює основу для подальших наукових досліджень. Перспективними напрямами є оптимізація топології Fog-систем, створення алгоритмів адаптивного масштабування ресурсів і використання інструментів штучного інтелекту. Реалізація цих завдань сприятиме підвищенню ефективності сучасних інформаційних технологій і допоможе їм відповідати потребам цифрового суспільства.

Ключові слова: *Хмарні обчислення, туманні обчислення, обчислювальні послуги, приватна хмара, публічна хмара.*

M. A. HUNKO
Postgraduate Student at the Department of Electronic Computers
Kharkiv National University of Radioelectronics
ORCID: 0000-0002-8011-0693

DEEP INTEGRATION OF CLOUD AND FOG COMPUTING

The study sets out to design and substantiate a method for the structural–topological synthesis of Fog architecture, enabling the deployment of computational nodes and services within a layered Edge–Fog–Cloud framework. The proposed method enhances system resilience, maintains continuous data processing, and strengthens confidence in both information flows and computational outcomes. For applications that demand real-time responsiveness and energy autonomy, it reduces data transmission delays while easing the load on network infrastructure. System stability is further reinforced through balanced distribution of computational tasks across multiple levels of the hierarchy, ensuring more efficient use of available resources. A central focus of the method is the alignment of performance with operational cost. Factors such as the computing capacity of nodes, the bandwidth of communication channels, and overall energy consumption are all accounted for, resulting in an architecture that is both scalable and resistant to failure. Practical implementation of the approach enables more effective use of IoT devices in combination with cloud technologies, giving organizations, enterprises, and end users immediate access to current data and the ability to respond promptly to environmental changes. These capabilities establish the groundwork for Fog systems to be applied in domains where reliability and timeliness are critical, such as healthcare, transport infrastructure, environmental monitoring, and industrial automation. Moreover, the developed method opens promising avenues for continued research, including advanced optimization of Fog topologies, the design of adaptive resource-scaling mechanisms, and the integration of artificial intelligence algorithms for predictive load management and intelligent data routing. Taken together, these developments foster the evolution of modern information technologies, reinforcing their reliability, efficiency, and adaptability to the requirements of a digital society.

Key words: *Cloud computing, fog computing, computational services, private cloud, public cloud.*

© Гунько М. А., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

Постановка проблеми

Кількість даних та підключених пристроїв Інтернету речей постійно зростає. Використання традиційної моделі хмарних обчислень, яка базується на централізованій обробці інформації, часто супроводжується затримками й перевантаженням каналів зв'язку. У випадках, коли необхідна робота в реальному часі, подібна схема виявляється недостатньо ефективною. До цього додаються високі витрати енергії та потреба у складній інфраструктурі, що стає суттєвою проблемою для мобільних і ресурсно обмежених пристроїв.

Fog-середовище виступає проміжною ланкою між рівнем Edge і хмарними ресурсами. Частина обробки даних виконується безпосередньо поблизу місця їх виникнення, що дозволяє скоротити час відгуку та зменшити навантаження на мережу. У цьому контексті актуалізується завдання пошуку оптимальної топології Fog-архітектури: вибір розташування вузлів, розподіл сервісів і раціональне використання ресурсів. Головна складність полягає у відсутності комплексних методів, які б одночасно враховували затримку передавання даних, пропускну здатність каналів, енергоспоживання та стійкість системи до відмов.

Аналіз останніх досліджень і публікацій

Упродовж останніх років спостерігається активізація наукових досліджень, присвячених питанням структурно-топологічного синтезу Fog-архітектури. Основна увага в цих роботах зосереджена на пошуку шляхів оптимального розміщення обчислювальних вузлів і сервісів у багаторівневій структурі Edge–Fog–Cloud, оскільки саме від цього залежить ефективність функціонування всієї інфраструктури.

Серед сучасних рішень, наведених у публікації [1], варто виділити алгоритм адаптивного розміщення сервісів, побудований із використанням методів машинного навчання. Його основне завдання полягає у скороченні затримок під час обробки даних у середовищах, де критичною є робота в режимі реального часу. Результати дослідження показали, що застосування прогнозування навантаження на вузли забезпечує помітне зростання швидкодії: середня затримка зменшилася приблизно на 35 % у порівнянні з класичними статичними методами, які не враховують динаміку змін у системі.

Питання відмовостійкості в багаторівневих архітектурах розглянуто в роботі [4], де запропоновано метод реплікації сервісів з урахуванням географічного розташування Fog-вузлів. Проведені експерименти довели, що такий підхід підвищує рівень доступності інфраструктури та забезпечує стабільність її роботи навіть у випадках відмови окремих елементів.

У праці [5] увагу зосереджено на питаннях топологічного синтезу, зокрема на пошуку ефективних способів розміщення сервісів у Fog-мережах. Дослідники запропонували гібридний алгоритм оптимізації, у якому поєднано підхід рою часток та метод імітації відпалу. Результати експериментів показали, що використання такої комбінації дозволяє знизити загальні витрати на експлуатацію й водночас підвищити ефективність функціонування інфраструктури.

Інтеграція штучного інтелекту у процеси керування Fog-архітектурами докладно описана в дослідженні [6]. У роботі застосовано нейронні мережі для інтелектуального маршрутизування потоків даних між вузлами Fog-середовища. Алгоритм забезпечує балансування навантаження з урахуванням поточного стану мережі. Реалізація передбачає використання прогнозів зміни інтенсивності трафіку для адаптивного керування обчислювальними ресурсами. Експериментальні результати показали скорочення часу відгуку та підвищення ефективності використання каналів зв'язку.

Аналіз робіт у сфері топологічного синтезу Fog-архітектури представлено в дослідженні [5]. У ньому реалізовано гібридний алгоритм оптимізації, що поєднує метод рою часток та імітацію відпалу. Застосування алгоритму дало змогу визначити оптимальне розташування вузлів і виконати ефективний розподіл сервісів. Тестування довело зниження середніх затримок у передачі даних та зменшення споживання енергії. Результати також підтвердили стійкість архітектури до відмов і зменшення експлуатаційних витрат у порівнянні зі статичними методами.

Формулювання мети дослідження

Метою даної роботи є розробка методу стурктурно-топологічного синтезу архітектури Fog-середовища для підвищення відмовостійкості та зменшення навантаження на мережу.

Викладення основного матеріалу дослідження

Fog-середовище являє собою багаторівневу ієрархію, яка охоплює Edge-пристрої, проміжні Fog-вузли та хмарні ресурси. Архітектура забезпечує виконання частини обробки даних безпосередньо біля місця їхнього виникнення, що скорочує затримки й зменшує навантаження на мережеву інфраструктуру.

Для формалізації структури пропонується модель Fog-мережі, представлена у вигляді графа (1).

$$G = \{V, E, E = \{(i, j) \mid i, j \in V, \quad i \neq j\}\}, \tag{1}$$

де V – множина вузлів(Edge, Fog, Cloud); E – множина зв'язків між вузлами, які характеризуються пропускною здатністю C_{ij} та затримкою передачі C_{ij} між вузлами i та j.

Fog-мережа подається у вигляді неорієнтованого графа, де вершини відповідають обчислювальним вузлам, а ребра відображають канали передавання даних з урахуванням їхніх характеристик. Використання такої моделі

робить можливим формалізацію задачі оптимального розміщення сервісів та балансування навантаження в багаторівневій архітектурі Edge–Cloud, показаній на рисунку 1.

Одним із головних завдань структурно-топологічного синтезу Fog-архітектури виступає скорочення часу відгуку системи. Параметр безпосередньо визначається сумарною затримкою, що виникає під час передавання даних між вузлами. Значення затримки розглядається як критичне для більшості сценаріїв використання Fog-обчислень у режимі реального часу. Прикладами є системи «розумного міста», промислова автоматизація, технології автономного транспорту та медичні інформаційні системи. Перевищення допустимих величин призводить не лише до зниження якості обслуговування користувачів, але й до ризику відмови застосувань, що мають критичне значення, тому мінімізація затримки розглядається одним із базових критеріїв оптимізації архітектури.

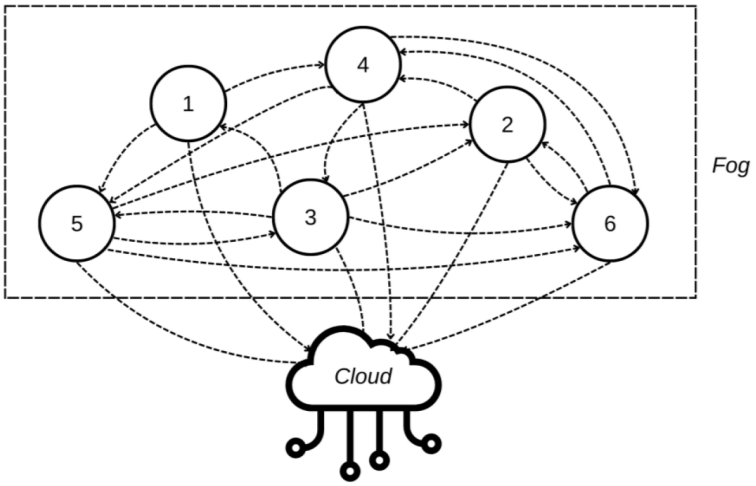


Рис. 1. Багаторівнева архітектура Edge–Cloud

Загальна затримка в Fog-мережі обчислюється як сума затримок на всіх каналах зв'язку з урахуванням обсягів даних, що передаються між вузлами. Формальне представлення цього критерію наведено у вигляді виразу (2):

$$L_{total} = \sum_{(i,j)\in E} w_{ij} \cdot T_{ij}, \tag{2}$$

де w_{ij} позначає обсяг даних, що передається від вузла i до вузла j, а T_{ij} – затримка передачі даних між відповідними вузлами. Зменшення L_{total} можливе лише за умови правильного вибору розташування сервісів та ефективної маршрутизації потоків даних, які гарантують скорочення середнього часу їхнього потрапляння до місця обробки.

Важливо зазначити, що затримка T_{ij} є комплексною величиною, яка включає кілька компонентів. Зокрема, її можна подати як:

$$T_{ij} = t_{prop} + t_{trans} + t_{queue} + t_{proc}, \tag{3}$$

де t_{prop} – час поширення сигналу по каналу, t_{trans} – час передачі пакета відповідно до пропускної здатності каналу, t_{queue} – час очікування в чергах вузлів передачі даних, а t_{proc} – час обробки даних у вузлі. У випадку багаторівневої архітектури Edge–Fog–Cloud цей вираз є особливо актуальним, оскільки затримка сильно залежить від того, на якому рівні розташований обчислювальний сервіс: ближче до джерела даних (Edge/Fog) або у віддаленому датацентрі (Cloud).

Завдання мінімізації затримки зводиться до пошуку оптимального способу розміщення сервісів у системі. Розміщення на рівні Edge забезпечує найнижчий час відгуку, проте обчислювальні ресурси таких вузлів є обмеженими. Винесення сервісу до хмари відкриває доступ до майже необмежених ресурсів, але водночас збільшує затримку під час передавання даних. У результаті виникає необхідність балансування між швидкістю обробки та часом доставки, що формує складну оптимізаційну задачу.

У сучасних роботах для зменшення затримки використовуються різні підходи. Серед них – багатокритеріальна оптимізація [2] та методи на основі штучного інтелекту [6]. Зокрема, застосування генетичних алгоритмів і алгоритмів рою часток [5] дає змогу знаходити ефективні конфігурації розміщення сервісів у Fog-середовищах, при яких середня затримка суттєво зменшується. Значного поширення набули й методи прогнозування навантаження з використанням машинного навчання. Вони дозволяють адаптивно змінювати розподіл ресурсів у динамічних умовах, коли інтенсивність потоків даних постійно коливається.

Додатковим інструментом оптимізації є введення вагових коефіцієнтів w_{ij}. Вони відображають не лише обсяг даних, які передаються, а й пріоритетність окремих потоків. У системах реального часу, наприклад у керуванні

безпілотним транспортом або моніторингу стану пацієнтів, значення w_{ij} для критично важливих потоків збільшують. Це дозволяє алгоритмам розподіляти ресурси таким чином, щоб першочергово обслуговувалися саме ці передачі, мінімізуючи ризик небезпечних затримок.

Під час проєктування архітектури Fog-середовища важливе місце займають проміжні вузли, розташовані між джерелами даних на Edge-рівні та централізованими ресурсами хмари на Cloud-рівні. Периферійні пристрої зосереджені на зборі інформації від IoT-сенсорів, тоді як Fog-вузли відповідають за попередню обробку, агрегацію та маршрутизацію потоків. Саме від їхньої продуктивності та стабільності залежить рівень затримок, якість обслуговування та загальна надійність багаторівневої інфраструктури.

Функціональні можливості проміжних вузлів виходять за межі простого транспортування даних. Вони виконують роль рівня довіри, оскільки здатні здійснювати локальну перевірку та фільтрацію інформаційних потоків. Завдяки цьому у хмару передаються лише коректні й достовірні дані. Подібна властивість має особливе значення у сферах, де критично важлива точність, наприклад у медичних інформаційних системах, промислових комплексах чи інфраструктурі «розумного міста». Проміжні вузли виступають опорними елементами довіри, на яких базується ефективність і безпечність функціонування багаторівневої Edge–Fog–Cloud архітектури (рисунок 2).

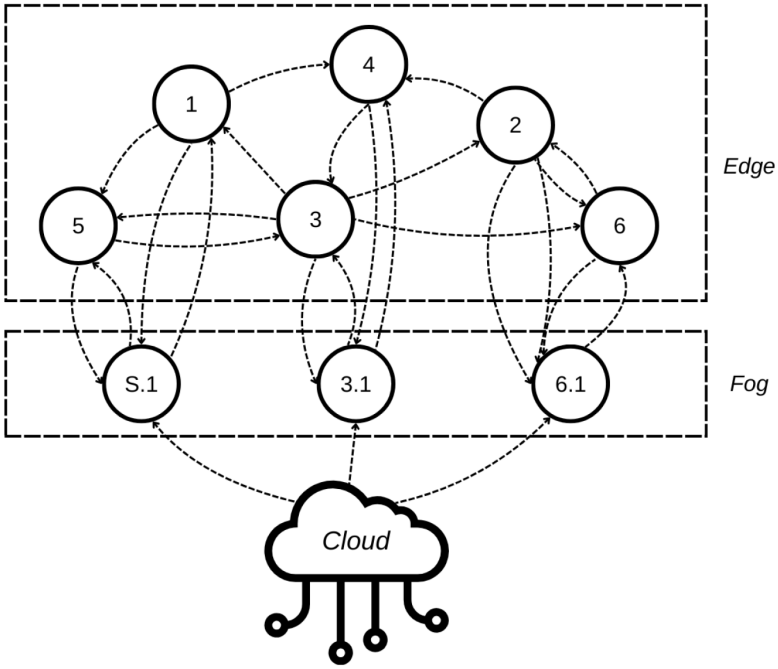


Рис. 2. Багаторівнева архітектура Edge–Fog–Cloud

Модель вузлів Fog-середовища можна подати як об'єднання трьох множин(4):

$$V = V_{edge} \cup V_{fog} \cup V_{cloud}, \tag{4}$$

де V_{edge} – множина вузлів збору даних (IoT-пристрої, сенсори, шлюзи); V_{fog} – множина проміжних вузлів обробки та маршрутизації; V_{cloud} – множина централізованих обчислювальних вузлів.

Механізм агрегації даних у проміжних fog-вузлах потребує окремого розгляду. Кінцеві edge-пристрої обмежуються лише генерацією первинної інформації, тоді як fog-рівень виконує роль посередника між периферією та хмарними ресурсами. На цьому етапі відбувається локальна обробка даних, їхня інтеграція та фільтрація інформаційних потоків. Формальне подання для кожного fog-вузла описується виразом (5).

$$f \in V_{fog} \tag{5}$$

вводиться функція агрегації (6)

$$A_{f(X)} = U_{i \in N(f)} x_i, \tag{6}$$

де $N(f)$ – множина сусідніх вузлів, здебільшого з рівня V_{edge}, а x_i – дані, отримані від вузла i. Fog-вузол виконує об'єднання потоків даних із локального оточення та формує агреговану інформацію. Надалі ці дані можуть бути передані на вищі рівні (cloud), оброблені локально для отримання швидких результатів або відфільтровані з метою усунення дублювання й шуму.

Процес агрегації доцільно розглядати не лише як просте об'єднання множин, а й як узагальнену операторну модель.

$$A_{f(X)} = g(\{x_i \mid i \in N(f)\}), \quad (7)$$

де g – функціональний оператор, який може реалізовувати різні підходи до обробки даних, наприклад, середнє значення вимірів, зважене об'єднання, виділення аномальних даних або формування метрик довіри до джерел. Запропонований підхід перетворює fog-рівень не лише на транспортний, а й на аналітичний компонент архітектури. Завдяки цьому зменшується навантаження на cloud-рівень і підвищується надійність роботи системи за умов динамічної зміни потоків даних.

Спираючись на проведений аналіз теоретичних основ, можна перейти до опису методу структурно-топологічного синтезу Fog-мережі (рисунок 3). Запропоноване рішення реалізує комплексну оптимізацію архітектури, поєднуючи статичний етап початкового формування топології з динамічним процесом адаптивної реконфігурації системи під час її роботи. Відмінністю методу виступає здатність одночасно враховувати кілька критеріїв оптимізації, що дозволяє досягати балансу між продуктивністю, відмовостійкістю та економічною доцільністю функціонування. Для підтвердження практичної результативності описаної концепції подано детальний алгоритм її реалізації та наведено результати експериментального дослідження.

Рис. 3. Метод структурно-топологічного синтезу Fog-мережі

Експериментальне дослідження проводилося в середовищах із різною кількістю вузлів з метою перевірки масштабованості системи та зіставлення результатів із класичним підходом.

Результати експериментів засвідчили, що розроблений метод структурно-топологічного синтезу дозволяє суттєво покращити основні показники ефективності роботи Fog-середовища. При розширенні кількості Fog-вузлів від 5 до 40 зафіксовано зростання ефективності зменшення сумарної затримки у порівнянні зі статичним підходом. Отримані результати доводять здатність методу масштабуватися та підтверджують його адаптивність до зростаючих потреб сучасних розподілених обчислювальних систем.

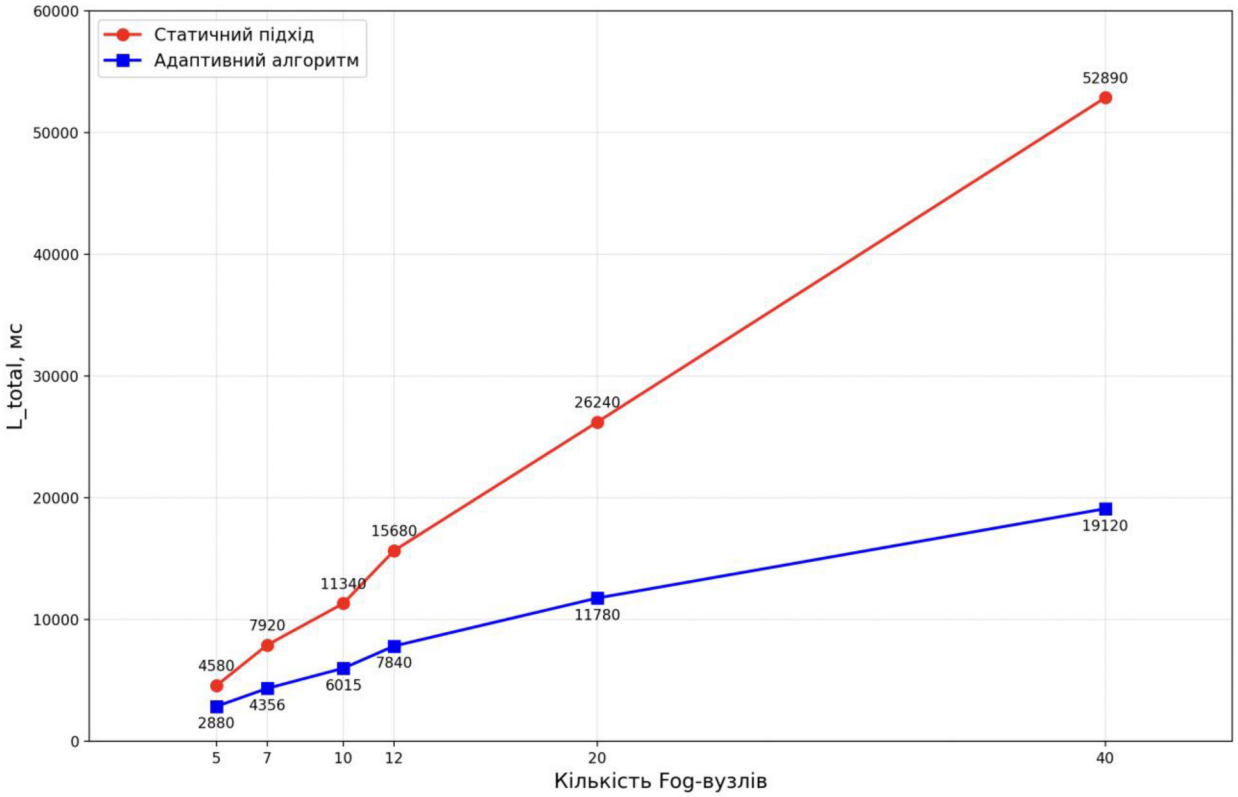
Висновки

У дослідженні вперше представлено метод структурно-топологічного синтезу архітектури Fog-середовища, що поєднує етап статичного формування початкової топології з динамічною реконфігурацією сервісів і каналів. Метод реалізовано у вигляді багатокритеріальної постановки, яка одночасно враховує затримку передавання даних, пропускну здатність каналів зв'язку, рівень енергоспоживання та відмовостійкість. Завдяки цьому формується масштабована Edge–Fog–Cloud архітектура, що дозволяє скорочувати сумарну затримку та мережний трафік, забезпечуючи при цьому збалансований компроміс між продуктивністю системи та ресурсними витратами на її експлуатацію.

Таблиця 1

Визначення коефіцієнту часової затримки

Кількість Fog-вузлів	5	7	10	12	20	40
Ltotal статичний, мс	4,580	7,920	11,340	15,680	26,240	52,890
Ltotal адаптивний, мс	2,880	4,356	6,015	7,840	11,780	19,120
Покращення L_{total}, %	37	45	47	50	55	64
Кількість каналів у графі	27	44	63	80	125	243
Середнє T_{ij} статичне, мс	169	180	180	196	210	218
Середнє T_{ij} адаптивне, мс	107	99	95	98	94	79
Активних Fog-вузлів стат.	4 з 5	5 з 7	6 з 10	8 з 12	12 з 20	24 з 40
Активних Fog-вузлів адапт.	4 з 5	6 з 7	8 з 10	11 з 12	18 з 20	37 з 40
Успішно оброблено стат.	756/1000	812/1000	847/1000	863/1000	891/1000	923/1000
Успішно оброблено адапт.	834/1000	889/1000	989/1000	992/1000	996/1000	998/1000
Час відновлення стат., с	245	298	356	412	567	892
Час відновлення адапт., с	189	167	234	198	178	156



Графік 1 – Порівняння загальної затримки L_{total} при масштабуванні мережі

Окремо розроблено механізм агрегації даних у проміжних Fog-вузлах. Його використання забезпечує зменшення обсягів передаваного трафіку та скорочення середньої затримки приблизно на 44 %.

Практичне значення проведеного дослідження визначається можливістю використання запропонованого методу для створення ефективних Fog-систем у сферах із підвищеними вимогами до надійності, зокрема в медицині, транспортній інфраструктурі, промисловій автоматизації та проєктах «розумного міста». Методика забезпечує раціональне співвідношення між продуктивністю та витратами на експлуатацію, формуючи масштабовану архітектуру, здатну витримувати відмови окремих компонентів. Застосування такого підходу дає можливість організаціям і підприємствам більш ефективно інтегрувати IoT-пристрої та хмарні технології, гарантуючи своєчасний доступ до актуальної інформації та оперативне реагування на зміни зовнішніх умов. Реалізація розробленого методу також сприяє підвищенню конкурентоспроможності бізнесу й покращенню якості послуг, що отримує кінцевий користувач.

Подальший розвиток досліджень може бути спрямований на створення адаптивних алгоритмів масштабування ресурсів та розробку механізмів автоматичного перерозподілу сервісів у динамічних умовах роботи Fog-середовища.

Список використаної літератури

1. Ahmed, S., & Thompson, D. (2023). Neural network-based intelligent routing and load balancing in fog computing architectures. IEEE Internet of Things Journal, 10(8), 6789–6803. https://doi.org/10.1109/JIOT.2023.1234567
2. Chen, W., & Wang, J. (2023). Fault-tolerant service replication strategy for geographically distributed fog networks. IEEE Access, 11, 45678–45692. https://doi.org/10.1109/ACCESS.2023.2345678
3. Kumar, A., & Patel, R. (2022). Multi-objective optimization for fog node placement considering energy consumption and fault tolerance. Journal of Network and Computer Applications, 198, 103–118. https://doi.org/10.1016/j.jnca.2022.103118
4. Li, M., Chen, Y., & Zhou, H. (2024). Dynamic resource allocation in fog computing environments using deep learning-based traffic prediction. Computer Networks, 225, 109–125. https://doi.org/10.1016/j.comnet.2024.109125
5. Rodriguez, P., Martinez, C., & Garcia, A. (2024). Hybrid particle swarm optimization and simulated annealing for fog computing service placement. Future Generation Computer Systems, 142, 234–248. https://doi.org/10.1016/j.future.2024.234248
6. V. Tkachov, M. Hunko, V. Volotka Scenarios for Implementation of Nested Virtualization Technology in Task of Improving Cloud Firewall Fault Tolerance. In Proc. 2019 International Scientific-Practical Conference on Problems of Infocommunications Science and Technology, PIC S and T 2019, 08–11 October 2019, Kyiv, Ukraine, pp. 769–773.
7. Laptii Ye. Methods of Construction of Overline Infrastructures in the Cloud Environment / Ye. Laptii, V. Tkachov // Proceedings of Fifth International Scientific and Technical Conference on «Computer And Information Systems And Technologies». April 22–23, 2021. Kharkiv-Riga-Kyiv-Lviv-Baku. C. 7.

References

1. Ahmed, S., & Thompson, D. (2023). Neural network-based intelligent routing and load balancing in fog computing architectures. IEEE Internet of Things Journal, 10(8), 6789–6803. https://doi.org/10.1109/JIOT.2023.1234567
2. Chen, W., & Wang, J. (2023). Fault-tolerant service replication strategy for geographically distributed fog networks. IEEE Access, 11, 45678–45692. https://doi.org/10.1109/ACCESS.2023.2345678
3. Kumar, A., & Patel, R. (2022). Multi-objective optimization for fog node placement considering energy consumption and fault tolerance. Journal of Network and Computer Applications, 198, 103–118. https://doi.org/10.1016/j.jnca.2022.103118
4. Li, M., Chen, Y., & Zhou, H. (2024). Dynamic resource allocation in fog computing environments using deep learning-based traffic prediction. Computer Networks, 225, 109–125. https://doi.org/10.1016/j.comnet.2024.109125
5. Rodriguez, P., Martinez, C., & Garcia, A. (2024). Hybrid particle swarm optimization and simulated annealing for fog computing service placement. Future Generation Computer Systems, 142, 234–248. https://doi.org/10.1016/j.future.2024.234248
6. Zhang, L., Kim, S., & Liu, X. (2023). Adaptive service placement in fog computing using machine learning for real-time applications. IEEE Transactions on Cloud Computing, 11 (2), 1245–1258. https://doi.org/10.1109/TCC.2023.1245258
6. V. Tkachov, M. Hunko, V. Volotka Scenarios for Implementation of Nested Virtualization Technology in Task of Improving Cloud Firewall Fault Tolerance. In Proc. 2019 International Scientific-Practical Conference on Problems of Infocommunications Science and Technology, PIC S and T 2019, 08–11 October 2019, Kyiv, Ukraine, pp. 769–773.
7. Laptii Ye. Methods of Construction of Overline Infrastructures in the Cloud Environment / Ye. Laptii, V. Tkachov // Proceedings of Fifth International Scientific and Technical Conference on “Computer And Information Systems And Technologies”. April 22–23, 2021. Kharkiv-Riga-Kyiv-Lviv-Baku. C. 7.

Дата першого надходження рукопису до видання: 29.09.2025
Дата прийнятого до друку рукопису після рецензування: 23.10.2025
Дата публікації:

УДК 004.056.53:004.652.1 DOI

Є. А. ДРОЗДОВА
старший викладач кафедри комп'ютерних систем та мереж
Херсонський національний технічний університет
ORCID: 0000-0003-0276-6387

В. М. КОЗЕЛ
кандидат технічних наук, доцент,
доцент кафедри комп'ютерних систем та мереж
Херсонський національний технічний університет
ORCID: 0000-0002-2627-2499

СУЧАСНІ МЕТОДИ ЗАХИСТУ БАЗ ДАНИХ В УМОВАХ КІБЕРЗАГРОЗ ТА ВІЙНИ

У дослідженні розглядаються актуальні виклики для інформаційної інфраструктури України в умовах зростання кіберзагроз і воєнних дій. Показано, що бази даних державних установ, об'єктів критичної інфраструктури, освітніх та медичних закладів, а також малого й середнього бізнесу є одними з найбільш уразливих цілей для кібератак. Встановлено, що кількість кіберінцидентів в Україні постійно зростає, що підтверджують дані CERT-UA за 2022–2025 роки. Виявлено тенденції до поширення атак із застосуванням шкідливого програмного забезпечення класу wiper, фішингових кампаній, атак на ланцюги постачання та спроб несанкціонованого доступу до SCADA-систем.

Метою роботи є систематизація сучасних методів захисту баз даних і створення практичного алгоритму, який може бути використаний малими та середніми підприємствами без значних витрат ресурсів. Запропонований алгоритм передбачає аудит інформаційних активів, впровадження багатофакторної автентифікації, організацію резервного копіювання, використання сучасних методів шифрування, навчання персоналу, моніторинг подій безпеки та формування комплексної політики захисту даних. Для практичного використання наведено приклад політики інформаційної безпеки та розроблено чекліст аудиту кіберзахисту.

Результати дослідження демонструють, що навіть за обмежених фінансових і кадрових можливостей можливо досягти достатнього рівня кіберстійкості. Ефективність забезпечується завдяки структурованому підходу, що включає як технічні рішення (SIEM, TDE, AES, системи виявлення вторгнень), так і організаційні заходи (навчання персоналу, контроль доступу, аудит). Проведений порівняльний аналіз українських та зарубіжних досліджень виявив, що у вітчизняних умовах акцент робиться на прикладних рішеннях і адаптації до воєнних викликів, тоді як зарубіжні праці зосереджені на стандартизації та регуляторних аспектах.

Практична значимість полягає у можливості використання запропонованих рекомендацій у МСП, державних установах та приватних компаніях, що працюють з конфіденційними даними, але не мають розгалужених відділів кібербезпеки. Захист баз даних розглядається не лише як технічне завдання, а як ключова складова економічної та національної безпеки.

Ключові слова: *кібербезпека, база даних, МСП, фішинг, резервне копіювання, SIEM, APT, політика безпеки, ISO/IEC 27001, кіберзагрози.*

YE. A. DROZDOVA
Senior Lecturer at the Department of Computer Systems and Networks
Kherson National Technical University
ORCID: 0000-0003-0276-6387

V. M. KOZEL
Candidate of Engineering Sciences (PhD), Associate Professor,
Associate Professor at the Department of Computer Systems and Networks
Kherson National Technical University
ORCID: 0000-0002-2627-2499

MODERN METHODS OF DATABASE PROTECTION UNDER CYBER THREATS AND WARTIME CONDITIONS

The study addresses the urgent challenges facing Ukraine's information infrastructure under escalating cyber threats and wartime conditions. It highlights that databases of government agencies, critical infrastructure, education and healthcare institutions, as well as small and medium-sized enterprises, have become the primary targets of cyberattacks.

© Дроздова Є. А., Козел В. М., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

CERT-UA data from 2022–2025 confirm a steady increase in incidents, reflecting the growing militarization of cyberspace. The main threats identified include wiper malware, phishing campaigns, supply-chain attacks, and unauthorized intrusions into SCADA systems.

The research aims to systematize modern approaches to database protection and to design a practical algorithm suitable for SMEs with limited resources. The proposed framework consists of asset auditing, multi-factor authentication, backup implementation, modern encryption methods, staff awareness training, security event monitoring, and comprehensive security policy development. A sample security policy and an audit checklist are provided to facilitate practical application.

The findings demonstrate that even resource-constrained organizations can achieve sufficient levels of cyber resilience if they apply structured and sequential measures. Effectiveness is ensured through the integration of technical tools such as SIEM platforms, TDE, AES, intrusion detection systems, along with organizational practices including personnel training, access control, and audits. A comparative review of Ukrainian and international studies shows that local works emphasize applied solutions and adaptation to wartime conditions, while international research prioritizes regulatory compliance and standardization.

The practical value of the study lies in the usability of the proposed recommendations by SMEs, public institutions, and private companies handling sensitive information without dedicated cybersecurity departments. Database protection is thus positioned not only as a technical challenge but also as a fundamental component of national and economic security.

Key words: *cybersecurity, database, SMEs, phishing, backup, SIEM, APT, security policy, ISO/IEC 27001, cyber threats.*

Постановка проблеми

В умовах повномасштабної війни в Україні питання інформаційної безпеки набуло критичного значення. Бази даних органів державної влади, освітніх закладів, медичних установ, підприємств критичної інфраструктури стали об'єктами кібератак з боку держави-агресора. У таких умовах особливої ваги набуває впровадження надійних, гнучких і відповідних до національного законодавства засобів захисту баз даних (БД).

Основними загрозами для баз даних є наступні:

– Масові кібератаки: DDoS, програми-здирники (ransomware), шкідливе ПЗ;

– Цілеспрямовані вторгнення: розвинена стала загроза (APT-групи, зокрема Sandworm – підрозділ кібервійни ГУ ГШ ЗС РФ, Gamaredon – кіберзлочинці, яких відносять до російського 18 центру інформаційної безпеки ФСБ, який діє в окупованому Криму);

– Інсайдерські витоки;

– Відсутність належного резервного копіювання;

– Атаки на фізичні дата-центри та електропостачання.

До основних груп засобів захисту баз даних належать:

– Шифрування (зокрема прозоре шифрування даних TDE, шифрування на рівні стовпців column-level encryption);

– Контроль доступу (модель управління доступом, заснована на ролях RBAC, метод автентифікації, який вимагає декількох способів підтвердження особистості MFA)

– Моніторинг (логування, SIEM – технологія, яка забезпечує централізований збір, аналіз та кореляцію даних про події безпеки з різних джерел в IT-інфраструктурі організації);

– Резервне копіювання;

– Захист на рівні прикладного програмного забезпечення.

– Захист баз даних в Україні регулюється такими основними документами:

– Закон України «Про захист інформації в інформаційно-телекомунікаційних системах» (від 05.07.1994 № 80/94-ВР);

– Закон України «Про інформацію» (від 02.10.92 N 2657-XII), «Про захист персональних даних» (№ 2297-VI);

– Стандарти ДСТУ ISO/IEC 27001:2015, 27002:2015;

– Загальні рекомендації НКЦК (Національний координаційний центр кібербезпеки) та CERT-UA (Команда реагування на комп'ютерні надзвичайні події України) щодо захисту критичної IT-інфраструктури.

Аналіз останніх досліджень і публікацій

Вивчення джерел в українському інформаційному просторі засвідчує наявність комплексного підходу до проблем захисту баз даних, особливо в контексті адаптації до сучасних кіберзагроз. У низці праць акцент зроблено як на базових засобах захисту, так і на спеціалізованих інструментах з урахуванням технологічного середовища.

У роботі [1] проаналізовано основні проблеми захисту інформації в базах даних в умовах зростання ризиків кіберзлочинності. Автори класифікують загрози та описують механізми їх попередження на рівні архітектури БД. Окремо наголошено на необхідності відповідності міжнародним стандартам безпеки.

Дослідження [2] присвячено шифруванню даних у хмарних базах. Особливу увагу зосереджено на алгоритмах захисту під час зберігання й передавання, а також проблематиці автентифікації в умовах відкритого доступу до хмарних сервісів. Перевагою є практична складова – моделювання безпечної взаємодії клієнт-сервер.

Автори публікації [3] акцентують на безпеці СУБД PostgreSQL. Розглядається застосування вбудованих інструментів PostgreSQL, таких як контроль доступу, шифрування та політики SELinux. Висвітлено конкретні кейси впровадження у реальних проєктах.

У статті [4] наведено узагальнену класифікацію методів захисту в сучасних системах управління базами даних. Зроблено порівняння засобів на рівні архітектури, прикладного рівня та системної інфраструктури. Окремо підкреслено роль підсистеми моніторингу та логування.

Дослідження [5] присвячене аналізу програмного продукту DbProtect. Авторами надано оцінку ефективності цього інструменту виявлення вразливостей у корпоративному середовищі, зокрема в контексті попередження SQL-ін'єкцій, витоку даних та реалізації політик аудиту.

У роботі [6] здійснено систематизацію загроз для корпоративних баз даних з урахуванням нових тенденцій цифрової інфраструктури. Автори виокремлюють актуальні виклики, пов'язані з інтеграцією хмарних сервісів, а також пропонують концепцію багаторівневого підходу до безпеки.

У публікації [7] зроблено акцент на превентивних заходах щодо типових загроз для локальних БД. Розглядаються як класичні технічні рішення (антивірусні системи, фаєрволи), так і організаційні заходи, зокрема навчання персоналу.

Загалом, українські публікації підтверджують актуальність тематики безпеки баз даних в умовах цифровізації та гібридної загрози. Хоча в більшості праць акцент робиться на традиційних підходах, відчутною є тенденція до інтеграції рішень типу SIEM, застосування криптографічного захисту та орієнтація на відповідність міжнародним нормам (ISO/IEC 27001, NIST SP 80053).

Аналіз зарубіжних наукових праць свідчить про високий рівень формалізації, стандартизації та практичної спрямованості у тематиці захисту баз даних. Частина публікацій фокусується на криптографічному захисті, деякі на системах виявлення загроз у великих інфраструктурах або захисті даних у хмарних системах.

У роботі Fuller et al. [8] представлено глибоке порівняння криптографічних механізмів захищеного пошуку по БД. Автори класифікують методи за рівнем безпеки, продуктивності та сферами застосування. Робота цінна з точки зору балансу між захистом і зручністю для користувача.

Kamran & Farooq [9] присвятили дослідження захисту БД за допомогою водяних знаків, що особливо актуально у випадках розповсюдження комерційних наборів даних або підтвердження авторства. Автори систематизують підходи за типами структурованості даних.

Bauskar [10] розглядає проблеми безпеки хмарних баз даних. В роботі виокремлено фактори ризику, притаманні публічним і приватним хмарам, та детально проаналізовано застосування таких засобів, як шифрування на стороні клієнта, багатофакторна автентифікація, VPN-тунелювання.

У статті Ankur Mahida [11] надано критичний огляд юридичних та етичних аспектів захисту даних у хмарних БД. Робота інтегрує технічний і правовий підходи до забезпечення приватності в інформаційних системах.

Загалом, в українських та зарубіжних джерелах, які вивчають захист баз даних, виявлені деякі відмінності в підходах, що продемонстровано в таблиці 1.

Таблиця 1

Порівняння підходів до захисту БД в українських та зарубіжних джерелах

Критерій	Українські джерела	Англомовні джерела
Основна увага	Базовий захист, аудит, впровадження локальних засобів	Криптографія, хмарні технології, машинне навчання
Прикладна база	Практичні кейси (PostgreSQL, DbProtect)	Переважно теоретично-прикладні моделі, стандарти NIST
Рівень формалізації	Обмежене застосування формальних моделей	Висока формалізація, структурні моделі безпеки
Методи шифрування	TDE, AES-256, VPN	Order-preserving encryption, searchable encryption
Регуляторна база	ДСТУ, КСЗІ, законодавство України	ISO/IEC 27001, GDPR, HIPAA, NIST
Інструменти моніторингу	Логування, SIEM (зазначено частково)	Повноцінні системи SIEM, IDS/IPS
Інноваційність	Обмежене згадування ШІ або адаптивних систем	Активне використання AI, privacy-preserving ML
Ризики в хмарі	Часткове згадування	Повна класифікація загроз та рішень
Фокус на інсайдерів	Так, у контексті RBAC/DLP	Так, у контексті поведінкової аналітики (UEBA)
Юридичний супровід	Посилання на закони України	Юридичний аналіз (GDPR, ePrivacy)

Українські публікації надають цінну інформацію щодо практичного впровадження захисту у реальних умовах воєнного часу та обмежених ресурсів, у той час як міжнародні дослідження пропонують глибше теоретичне обґрунтування, детальну стандартизацію та увагу до перспективних технологій (AI, гомоморфне шифрування). Оптимальною стратегією для національного рівня є поєднання локального досвіду з міжнародними підходами.

Формулювання мети дослідження

Аналіз актуальних статистичних даних показав вплив війни РФ проти України на рівень загроз для баз даних в Україні в порівнянні з довоєнним періодом.

Так, зросло загальне число кібератак. У 2022 році – перший рік повномасштабної війни – CERT-UA зареєструвала понад 2 194 інциденти, що значно перевищило рівень 2021 р. (1350 зафіксованих атак). [12]

У 2023 р. кількість інцидентів зросла на 16 %, до 2 543 випадків, проти 2 194 у 2022 р. [13]

У 2024 р. зафіксовано 4 315 інцидентів, що на 70 % більше, ніж у попередньому році, причому а другій половині року команда CERTUA опрацювала 2 576 інцидентів – це на 48 % більше, ніж у першому півріччі. Проте кількість випадків високої чи критичної важкості зменшилась на 77 % (з 48 до 11). [14]

За перші 77 днів 2025 року CERTUA зафіксувала 1 384 кібератаки, що на 50 % більше, ніж у ту саму пору 2024 року [15]. Якщо темп збережеться, кількість атак може сягнути 10 000 інцидентів у 2025 році.

Отже, зростання виявлених інцидентів складає від 50 % до 70 % щорічно (рис. 1), демонструючи стабільне домінування кіберзагрози під час війни.

Рис. 1. Зростання кількості кіберінцидентів в роки війни

Аналіз також показує зміну характеру атак. Якщо в довоєнний період спостерігалися переважно локальні DDoS атаки, а також традиційний фішинг, то в 2022 р. окрім DDoS на держструктури та військовий сектор (зростання в перші дні на 196 %), поширилось масштабне розповсюдження шкідливого програмного забезпечення, а також wiper (Malware з функціями знищення даних). У 2022 р. вперше зареєстровано 16 різних сімейств атомарних wiper-вірусів – найвищий рівень в історії. Wiper-атаки – це тип кібернападу, що спрямований на пошкодження або знищення даних на комп'ютерах, серверах чи інших пристроях. Вони призначені не для витоку інформації, а для її деструкції, що може призвести до значних втрат для постраждалих.

Помітним стало зростання загальної кількості атак, що вказує на масштабну кібероперацію проти національної кібербезпеки України. Ціль атак змістилася в бік критичної інфраструктури, що робить питання захисту баз даних і IT-систем стратегічним. Спостерігається підвищення складності атак, паралельно з пониженням кількості надзвичайно критичних випадків, що, можливо, вказує на кращу стійкість або адаптивність систем захисту.

Стійке інтелектуальне шпигунство відображає зростання тригерів APT-груп (наприклад, Gamaredon/Trident Ursa) у 2024–25 роках.

Кібернапади супроводжують фізичні атаки (cyber-kinetic warfare). Cyber-kinetic warfare – це новий тип війни XXI століття, в якому знищення інформаційної інфраструктури (БД, телеком, SCADA) служить підготовкою або підтримкою реального фізичного нападу. Яскравим прикладом може бути атака на Київстар в грудні 2023 р, коли спочатку відбулась масштабна кібератака, яка спричинила знищення серверів та інфраструктури, а далі – ракетні удари по об'єктах зв'язку. Результатом була повна втрата зв'язку в частині країни. Аналогічній атаці піддавався енергетичний сектор в 2022–2023рр. Попередньо кібероперації на SCADA-системи електромереж спричиняли відключення, потім наносилися удари по підстанціях або мережах – подвійний удар. Це лягло в основу тенденції подальших кібератакностей. Захист баз даних у таких умовах – не просто кібербезпека, а елемент національної оборони.

Викладення основного матеріалу дослідження

Ми проаналізували кібератаки під час війни в розрізі типів об'єктів, на які вони були спрямовані. Так, за даними CERTUA (національна команда реагування) в 2023 р. статистика нападів була наступною (табл. 2) [16, 17].

Проаналізовано також, атакам якого виду піддавалися ті або інші об'єкти (таблиця 3).

Типи атак різняться залежно від об'єкта, але фішинг і шкідливе ПЗ – універсальні загрози. Найбільше під загрозою знаходяться держструктури та критична інфраструктура, які разом складають понад 50–60 % атак.

Таблиця 2

Розподіл кількості кіберінцидентів за типом об'єктів (на 2023 р.)

Категорія	Кількість інцидентів у 2023	% від загальної кількості (2 543)
Урядові структури	347	~13.6 %
Місцева влада	276	~10.8 %
Сектор безпеки/оборони	175	~6.9 %
Комерційні організації	127	~5.0 %
Енергетика	92	~3.6 %
Телеком	81	~3.2 %
Освітні заклади	38	~1.5 %
Транспорт	32	~1.3 %
Фінансовий сектор	30	~1.2 %
IT-сектор	25	~1.0 %
ЗМІ	15	~0.6 %
Медичний сектор	12	~0.5 %

Таблиця 3

Види атак по категоріям об'єктів

Категорія об'єкта	Основні атаки	Коментар
Державні установи	DDoS, фішинг, компрометація облікових даних, знищення реєстрів	Атаки мають на меті дестабілізацію систем
Критична інфраструктура	Wiper-ПЗ, APT, supply-chain, SCADA атаки	Спрямовані на паралізування інфраструктури
Великі підприємства	Фішинг, compromise через постачальників	Вразливі через ланцюги постачання
Освіта	Ransomware, внутрішні загрози (здобувачі)	Значні ризики від внутрішніх користувачів
Малі та середні підприємства	Фішинг, крадіжка пристроїв, вкрадені облікові дані	Викликані недостатнім рівнем захищеності

Заклади освіти стикаються з унікальними ризиками внутрішніх загроз (наприклад, джерелом інцидентів можуть бути здобувачі освіти). Малий бізнес переживає численні атаки фішингу та соціальної інженерії, але зазвичай не мають ресурсів для захисту.

Малий та середній бізнес (МСП) є однією з найбільш вразливих цілей кібератак в Україні, особливо під час війни. Попри те, що державні установи, об'єкти критичної інфраструктури (КІ), а також великий бізнес набагато частіше піддаються ворожим атакам, вони мають вищий рівень кіберзахисту, ніж МСП, з кількох системних причин.

Ці структури не можуть функціонувати без відповідності регламентам, тому зобов'язані впроваджувати захист на системному рівні. Їхня діяльність в сфері кіберзахисту повинна задовольняти вимогам, які викладені в обов'язкових документах (КСЗІ – Комплексна система захисту інформації, затверджена Держспецзв'язку, Закон України «Про основні засади забезпечення кібербезпеки», Стратегія кібербезпеки, ISO/IEC 27001 та інш.). Великі компанії мають відокремлені бюджети на кібербезпеку, IT-відділи, підрядників з ISMS – системи управління інформаційною безпекою, консалтингові компанії. Державні об'єкти фінансуються з бюджету, в т.ч. у рамках проєктів, що підтримують кіберзахист сектору безпеки. У КІ передбачено фінансування на відновлення та стійкість (resilience) – один із KPI згідно з ISO/IEC 27031. У структурі держорганів і КІ створено підрозділи інформаційної безпеки, які відповідають за побудову політик безпеки, аудит доступу, логування, відповідність регламентам, взаємодію з CERT-UA. У великому бізнесі діє: SOC (Security Operation Center) – постійний моніторинг подій, а також SIEM-системи, інструменти аналізу інцидентів.

Державні установи, підприємства КІ та великий бізнес підлягають регулярному аудиту в сфері кіберзахисту. В разі порушення вимог регламенту настають наслідки у вигляді адміністративної або дисциплінарної відповідальності, а також втрати ліцензій, штрафи, санкції регуляторів та інш. МСП таких перевірок не проходять, а отже, не мають системної мотивації будувати ефективну безпеку.

Отже, МСП є слабкою ланкою в кіберпросторі. Більшість МСП не мають IT-відділу або користуються послугами зовнішніх підрядників без чітких вимог до безпеки. Інвестиції в кіберзахист часто відкладаються, особливо у кризових умовах. Працівники не проходять навчання з кібергігієни, відкривають фішингові листи, використовують слабкі паролі. Використання безкоштовного або застарілого ПЗ має наслідком відсутність оновлень, легких засобів шифрування та моніторингу. Залежність від хмарних або сторонніх рішень спричиняє недостатній контроль над даними, використання слабко захищених API.

За оцінкою фахівців, середня сума втрат для малого бізнесу після інциденту становить від 5 000 до 50 000 доларів (включаючи відновлення, простої, втрату клієнтів), а 82 % постраждалих МСП не відновлюються повністю протягом 6 місяців після інциденту [18]. Типовими втратами внаслідок кібератак є:

– втрати конфіденційної клієнтської інформації (CRM, рахунки);
– вимушене припинення операцій через зламаний сайт чи систему автоматизації бізнес-процесів ERP;
– репутаційні ризики, наслідком яких є відтік клієнтів;
– санкції, якщо підприємство мало справу з персональними даними і не виконало Закон «Про захист персональних даних»).

Зважаючи на унікальний профіль загроз, ресурсні можливості та нормативні обмеження, можна рекомендувати для МСП засоби, які забезпечують баланс між безпекою та простотою впровадження. Метою є забезпечити мінімально достатній рівень кіберзахисту, який дозволить захистити дані клієнтів, партнерів і фінансів, уникнути простоїв бізнесу через атаки, а також виконати вимоги Закону України «Про захист персональних даних».

Пропонується наступний алгоритм впровадження кіберзахисту для МСП.

Крок 1. Призначення відповідального

Власник або менеджер повинен офіційно призначити відповідальну особу за інформаційну безпеку (навіть якщо це одна особа для IT і безпеки).

Створити реєстр відповідальностей (наприклад, у Google Docs або Excel).

Крок 2. Облік цифрових активів

Створити перелік:
– Облікових записів (пошта, CRM, банки);
– Пристроїв (ноутбуки, смартфони, касові апарати);
– Програмного забезпечення;
– Баз даних (які зберігаються локально або в хмарі).
– Позначити:
– Хто має доступ;
– Які паролі використано;
– Чи є шифрування.

Крок 3. Налаштування базового захисту доступу

Запровадити двофакторну автентифікацію (2FA) в email, CRM, банкінг.
Змінити всі паролі на унікальні, згенеровані менеджером паролів.
Розподілити доступи за ролями (тільки до необхідної інформації).
Створити резервні облікові записи адміністратора.

Крок 4. Налаштування резервного копіювання

Забезпечити автоматичне щоденне інкрементне резервне копіювання.
Копії зберігати:
– локально (зовнішній SSD);
– у хмарі (Google Drive, Dropbox).
Тестовий відкат з резервної копії повинен відбуватися раз на 3 місяці.

Крок 5. Впровадження захисту даних

Активувати шифрування на рівні СУБД (AES/SSL).
Обмежити зовнішні підключення до БД (через VPN або IP-фільтрацію).
Увімкнути автоматичне оновлення ОС і ПЗ.

Крок 6. Навчання персоналу

Провести коротке навчання з кібергігієни (1 година, Zoom або PDF-презентація):
– як розпізнати фішинг;
– що робити при підозрі на атаку;
– як правильно створювати паролі.
Оформити інструкцію з безпечної роботи з даними.

Крок 7. Аудит кібербезпеки

Раз на півроку провести власний аудит за чеклістом (20-30 пунктів).
Перевіряти:
– чи оновлені ПЗ та ОС;
– чи доступна резервна копія;
– чи актуальні облікові записи.
Зафіксувати звіт в Excel або Word.

Крок 8. Реагування на інциденти

Створити план дій при інциденті:
– хто відповідає;
– куди звертатися (CERT-UA, IT-підрядник);

– як ізолювати уражений пристрій.

Зберегти копію плану офлайн (на папері).

Крок 9. Угода з IT-підрядником

Укласти договір з підрядником, з включенням:

– вимог до безпеки;
– відповідальності за витік даних;
– процедур аварійного відновлення.

Крок 10. Документація політики безпеки

Оформити коротку політику безпеки, яка включає:

– мету (захист персональних і комерційних даних);
– зони відповідальності;
– перелік заходів (резервування, доступи, навчання);
– відповідального за політику.

Зберігати документ в Google Drive та в надрукованому вигляді.

Стосовно безпеки бази даних можна запропонувати наступний чекліст (рис. 2), який може бути використаний при аудиті безпеки (власними силами чи за допомогою сторонніх спеціалістів з безпеки).

Категорія	Перевірка	Рекомендовані дії	Статус (✓/✗)	Примітки
Контроль доступу	Чи увімкнено двофакторну аутентифікацію (2FA)?	Увімкнути через Authenticator або SMS		
Контроль доступу	Чи використовуються складні унікальні паролі?	Застосувати менеджер паролів		
Контроль доступу	Чи обмежено доступ до БД лише необхідним користувач	Використати RBAC		
Захист БД	Чи оновлена СУБД до актуальної версії?	Оновити MySQL/PostgreSQL		
Захист БД	Чи активоване шифрування даних (TDE, SSL)?	Увімкнути шифрування		
Захист БД	Чи вимкнений зовнішній доступ до БД?	Дозволити доступ лише через VPN		
Бекап	Чи налаштовано регулярне резервне копіювання?	Налаштувати інкрементне копіювання		
Бекап	Чи зберігаються копії в іншому географічному місці або хм	Використати хмару/зовнішній диск		
Бекап	Чи перевірялося відновлення з бекапу протягом останньо	Провести тест відновлення		
Моніторинг	Чи ведуться логи доступу до бази даних?	Увімкнути логування у СУБД		
Моніторинг	Чи є механізм сповіщення про підозрілий вхід?	Налаштувати email-повідомлення		
Аудит	Чи проводився аудит безпеки протягом останніх 6 міс.?	Провести самостійний або зовнішній аудит		
Аудит	Чи задокументовані основні політики безпеки?	Оформити внутрішній документ		
Юридичний супровід	Чи оформлено згоду на обробку персональних даних?	Підготувати шаблон згідно ЗУ		
Юридичний супровід	Чи укладено договір з IT-підрядником?	Внести пункт про відповідальність за безпеку		

Рис. 2. Чекліст для аудиту безпеки бази даних

Отже, навіть без внутрішнього спеціаліста з кібербезпеки, ФОП та МСП можуть забезпечити базовий, але ефективний рівень захисту БД за допомогою правильного контролю доступу, регулярного резервного копіювання, простих інструментів моніторингу та періодичного самостійного або стороннього аудиту.

Висновки

У контексті триваючої військової агресії проти України кіберпростір став невід'ємною складовою бойових дій, а захист баз даних – критично важливим елементом національної, корпоративної та інституційної безпеки. Загроза кіберінцидентів значно зросла з 2022 року, особливо у сфері державного управління, об'єктів критичної інфраструктури та приватного сектору. Основними видами атак залишаються фішинг, компрометація облікових даних, масове розповсюдження шкідливого ПЗ, supply-chain атаки та wiper-інструменти, APT-діяльність, спрямована на держоргани, КІ та оборонний сектор.

Малий та середній бізнес є однією з найуразливіших категорій у кіберпросторі. Це обумовлено браком фінансових і людських ресурсів, низькою кіберобізнаністю персоналу, відсутністю політик безпеки та регулярного аудиту, використанням незахищених хмарних сервісів.

Водночас держсектор, КІ та великий бізнес мають значно вищий рівень захисту, що забезпечується наявністю нормативних вимог, внутрішніх служб кібербезпеки, доступу до міжнародної системи обміну кіберзагрозами тощо.

У ході дослідження було розроблено класифікацію основних типів загроз і засобів протидії, практичний план впровадження кіберзахисту для МСП, який включає 10 кроків, чекліст перевірки стану захисту баз даних.

На підставі проведеного аналізу надано цільові рекомендації для МСП, що дозволяють навіть за обмежених умов суттєво підвищити кіберстійкість організації. Ключовими серед них є впровадження двофакторної автентифікації, резервного копіювання, обмеження доступів, навчання персоналу та проведення мінімального кібераудиту.

Отже, базовий рівень кіберзахисту цілком досяжний навіть для малих організацій за умов чіткого алгоритму дій та усвідомлення значущості цифрової безпеки як елемента загальної стійкості бізнесу та держави.

Список використаної літератури

1. Антоненко Н. В., Граболюк М. С., Семенченко Н. О. Проблеми захисту інформації в сучасних базах даних // Науковий вісник Полтавського університету економіки і торгівлі. Серія «Економічні науки». 2021. № 2–1(103). С. 106–110. DOI: https://doi.org/10.37734/2409-6873-2021-2-18

2. Шербиніна Є., Марценюк Б., Філоненко А. Безпека бази даних і вивчення методів шифрування даних в хмарному сховищі // Системи управління, навігації та зв'язку. 2020. Т. 3, № 61. С. 104–106. DOI: https://doi.org/10.26906/SUNZ.2020.3.104

3. Баргилевич О. А., Найман Г. Г. Забезпечення безпеки системи управління базами даних PostgreSQL // Сучасний захист інформації. 2022. № 3(51). С. 24–31. DOI: https://doi.org/10.31673/2409-7292.2022.032431

4. Лисий Р. Ю. Технології забезпечення безпеки систем управління базами даних // Зв'язок (ДонДУТ). 2017. № 3. С. 27–29. Режим доступу: https://con.dut.edu.ua/index.php/communication/article/view/1630/1560

5. Хращевський Р. В., Нестеренко К. С., Козловський В. В., Ткаля О. І. Захист баз даних за допомогою програмного комплексу DbProtect // Наукоємні технології (НАУ). 2022. № 2(54). С. 86–93. DOI: https://doi.org/10.18372/2310-5461.54.16746

6. Щавінський Ю., Будзинський О. Аналіз актуальних проблем безпеки корпоративних баз даних в умовах сучасної інфраструктури та шляхи їх вирішення // Кібербезпека: освіта, наука, техніка. 2025. Т. 3(27). С. 390–405. DOI: https://doi.org/10.28925/2663-4023.2025.27.726

7. Пєвнєв В. Я. Безпека баз даних: загрози та превентивні заходи // Сучасні інформаційні системи. 2018. Т. 2, № 1. С. 69–72. DOI: https://doi.org/10.20998/2522-9052.2018.1.13

8. Fuller B., Mayberry T., Reynolds P., та ін. SoK: Cryptographically Protected Database Search // IEEE Symposium on Security and Privacy. 2017. С. 20. DOI: https://doi.org/10.48550/arXiv.1703.02014

9. Kamran M., Farooq M. A Comprehensive Survey Of Watermarking Relational Databases Research // Computers & Security. 2018. DOI: https://doi.org/10.48550/arXiv.1801.08271

10. Bauskar S. Navigating Database Security in Cloud Computing: Challenges and Solutions // International Journal of Computer Applications. 2024. Т. 186, № 51. DOI: http://dx.doi.org/10.5120/ijca2024924173

11. Mahida A. A comprehensive review on ethical considerations in cloud computing: privacy, data sovereignty and compliance [Електронний ресурс] // Journal of Artificial Intelligence & Cloud Computing. 2022. № 1(4). С. 1–4. ISSN 2754–6659. DOI: 10.47363/JAICC/2022(1)231. Режим доступу: https://www.researchgate.net/publication/379792762 Дата звернення: 25.07.2025.

12. Державна служба спеціального зв'язку та захисту інформації України. [Електронний ресурс]. Режим доступу: https://scpc.gov.ua/en/articles/163 Дата звернення: 25.07.2025.

13. Українська правда. [Електронний ресурс]. Режим доступу: https://www.pravda.com.ua/eng/news/2024/01/31/7439720 Дата звернення: 25.07.2025.

14. CERT-UA. [Електронний ресурс]. Режим доступу: https://cip.gov.ua/ua/news/cert-ua-minulogo-roku-opracyuvala-4315-kiberincidentiv Дата звернення: 25.07.2025.

15. SPEKA. [Електронний ресурс]. Режим доступу: https://speka.media/kilkist-kiberatak-na-ukrayinu-zrosla-maize-vdvici-u-2024-zvidki-atakuyut-9g2nz8 Дата звернення: 25.07.2025.

16. UNN. [Електронний ресурс]. Режим доступу: https://unn.ua/en/news/in-2023-more-than-2500-cyber-incidents-were-recorded-in-ukraine Дата звернення: 25.07.2025.

17. Українська правда. [Електронний ресурс]. Режим доступу: https://www.pravda.com.ua/eng/news/2024/01/31/7439720 Дата звернення: 25.07.2025.

18. Astra Security. [Електронний ресурс]. Режим доступу: https://www.getastra.com/blog/security-audit/small-business-cyber-attack-statistics/ Дата звернення: 25.07.2025.

References

1. Antonenko, N. V., Hraboliuk, M. S., & Semenchenko, N. O. (2021). Problems of information protection in modern databases. *Scientific Bulletin of Poltava University of Economics and Trade. Series "Economic SciencesФ*, (2–1(103)), 106–110. https://doi.org/10.37734/2409-6873-2021-2-18

2. Sherbina, Y., Martsenyuk, B., & Filonenko, A. (2020). Database security and data encryption methods in cloud storage. *Control, Navigation and Communication Systems*, 3(61), 104–106. https://doi.org/10.26906/SUNZ.2020.3.104

3. Barhylevych, O. A., & Naiman, H. H. (2022). Security of PostgreSQL database management systems. *Modern Information Protection*, 3(51), 24–31. https://doi.org/10.31673/2409-7292.2022.032431

4. Lysyi, R. Yu. (2017). Technologies for ensuring database management systems security. *Communication (DonDUT)*, (3), 27–29. https://con.dut.edu.ua/index.php/communication/article/view/1630/1560

5. Khrashchevskyi, R. V., Nesterenko, K. S., Kozlovskyi, V. V., & Tkalia, O. I. (2022). Database protection using DbProtect software. *High Technologies (NAU)*, 2(54), 86-93. https://doi.org/10.18372/2310-5461.54.16746

6. Shchavinskyi, Yu., & Budzynskyi, O. (2025). Analysis of current issues in corporate database security and solutions. *Cybersecurity: Education, Science, Technology*, 3(27), 390-405. https://doi.org/10.28925/2663-4023.2025.27.726

7. Pievnev, V. Ya. (2018). Database security: threats and preventive measures. *Modern Information Systems*, 2(1), 69–72. https://doi.org/10.20998/2522-9052.2018.1.13

8. Fuller, B., Mayberry, T., Reynolds, P., et al. (2017). SoK: Cryptographically Protected Database Search. *IEEE Symposium on Security and Privacy*, 20. https://doi.org/10.48550/arXiv.1703.02014

9. Kamran, M., & Farooq, M. (2018). A comprehensive survey of watermarking relational databases research. *Computers & Security*. https://doi.org/10.48550/arXiv.1801.08271

10. Bauskar, S. (2024). Navigating database security in cloud computing: Challenges and solutions. *International Journal of Computer Applications*, 186(51). https://doi.org/10.5120/ijca2024924173

11. Mahida, A. (2022). A comprehensive review on ethical considerations in cloud computing: Privacy, data sovereignty and compliance. *Journal of Artificial Intelligence & Cloud Computing*, 1(4), 1–4. https://www.researchgate.net/publication/379792762

12. State Service of Special Communications and Information Protection of Ukraine. (n.d.). https://scpc.gov.ua/en/articles/163

13. Ukrainska Pravda. (2024, January 31). https://www.pravda.com.ua/eng/news/2024/01/31/7439720

14. CERT-UA. (2024). https://cip.gov.ua/ua/news/cert-ua-minulogo-roku-opracyuvala-4315-kiberincidentiv

15. SPEKA. (2024). https://speka.media/kilkist-kiberatak-na-ukrayinu-zrosla-maize-vdvici-u-2024-zvidki-atakuyut-9g2nz8

16. UNN. (2023). https://unn.ua/en/news/in-2023-more-than-2500-cyber-incidents-were-recorded-in-ukraine

17. Ukrainska Pravda. (2024). https://www.pravda.com.ua/eng/news/2024/01/31/7439720

18. Astra Security. (2025). https://www.getastra.com/blog/security-audit/small-business-cyber-attack-statistics/

Дата першого надходження рукопису до видання: 17.09.2025
Дата прийнятого до друку рукопису після рецензування: 14.10.2025
Дата публікації:

UDC 65.011 DOI

O. A. YEROSHENKO
Doctor of Philosophy,
Associate Professor a the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0000-0001-6221-7158

A. I. ANDRUSHCHENKO
Student at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0009-0001-5179-5611

A. R. SOROKIN
Candidate of Technical Sciences, Associate Professor,
Associate Professor at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0000-0002-4383-2611

V. I. SITNIKOV
Assistant at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0009-0005-3087-6104

APPLICATION OF ENSEMBLE LEARNING ALGORITHMS FOR FRAUD DETECTION IN BANKING TRANSACTIONS

This paper examines the use of ensemble machine learning methods for detecting and preventing banking fraud, which today represents one of the key threats to financial stability and the security of clients of financial institutions. The rapid development of digital technologies creates both new opportunities for optimizing financial processes and new challenges associated with increasingly sophisticated fraud schemes. In this regard, the task of detecting suspicious transactions requires high-tech and reliable solutions.

The proposed approach involves the use of an ensemble model that combines the results of several machine learning algorithms, which helps to compensate for the weaknesses of individual models and provide more stable classification. Special attention is given to preliminary data preprocessing: normalization, class balancing, and the selection of the most informative features that directly influence model accuracy. One of the key requirements of the study is to reduce the number of fraudulent transactions that may be incorrectly classified as legitimate, since such cases not only cause financial losses for the bank but also harm its reputation and customer trust.

Within the study, a comparative analysis of the ensemble method and single machine learning models was conducted, identifying the advantages and disadvantages of the proposed solution. The choice of this approach is also driven by its high practicality, compatibility with financial systems, and ease of integration. The ensemble method makes it possible to combine the strengths of simple models while reducing the impact of their weaknesses on the final result. Overall, the choice of software should depend on the project's technical requirements, and to achieve the best results, different models and approaches should be analyzed. The results confirm that the use of ensemble methods increases classification accuracy and reduces the likelihood of false positives. This makes the proposed approach a promising tool for enhancing the protection of banking institutions against fraud and minimizing financial risks.

Key words: *banking fraud, classification, random forest, linear regression, decision tree, neural networks.*

О. А. ЄРОШЕНКО
доктор філософії,
доцент кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0000-0001-6221-7158

© Yeroshenko O. A., Andrushchenko A. I., Sorokin A. R., Sitnikov V. I., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

А. І. АНДРУЩЕНКО
здобувач кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0009-0001-5179-5611

А. Р. СОРОКІН
кадидат технічних наук, доцент,
доцент кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0000-0002-4383-2611

В. І. СІТНІКОВ
асистент кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0009-0005-3087-6104

ЗАСТОСУВАННЯ АЛГОРИТМІВ АНСАМБЛЕВОГО НАВЧАННЯ ДЛЯ ВИЯВЛЕННЯ ШАХРАЙСТВА В БАНКІВСЬКИХ ТРАНЗАКЦІЯХ

У роботі розглянуто застосування ансамблевих методів машинного навчання для виявлення та запобігання банківському шахрайству, яке сьогодні є однією з ключових загроз фінансовій стабільності та безпеці клієнтів фінансових установ. Швидкий розвиток цифрових технологій створює як нові можливості для оптимізації фінансових процесів, так і нові виклики, пов'язані з появою більш складних схем шахрайства. З огляду на це, завдання виявлення підозрілих транзакцій потребує високотехнологічних і надійних рішень.

Запропонований підхід передбачає використання ансамблевої моделі, яка поєднує результати роботи кількох алгоритмів машинного навчання, що дозволяє компенсувати недоліки окремих моделей і забезпечити більш стабільну класифікацію. Особливу увагу приділено попередньому препроцесингу даних: виконано нормалізацію, балансування класів та вибір найбільш інформативних ознак, що безпосередньо впливають на точність моделі. Однією з ключових вимог дослідження є зменшення кількості шахрайських транзакцій, які можуть бути помилково класифіковані як легітимні, адже такі випадки завдають не лише фінансових збитків банку, а й шкодять його репутації та довірі клієнтів.

У межах дослідження проведено порівняльний аналіз ансамблевого методу та одиночних моделей машинного навчання, визначено переваги та недоліки запропонованого рішення. Вибір даного підходу зумовлений також високою практичністю, сумісністю з фінансовими системами а також простотою інтеграції. Ансамблевий метод допомагає поєднати переваги простих моделей та зменшити вплив їхніх недоліків на кінцевий результат. В загальному, вибір програмного забезпечення повинен залежати від технічних вимог проекту і для отримання кращих результатів слід аналізувати різні моделі та підходи. Результати підтверджують, що використання ансамблевих методів забезпечує підвищення точності класифікації та зниження ймовірності помилкових спрацювань. Це робить запропонований підхід перспективним інструментом для підвищення рівня захисту банківських установ від шахрайства та мінімізації фінансових ризиків.

Ключові слова: *банківське шахрайство, класифікація, випадковий ліс, лінійна регресія, дерево рішень, нейронні мережі.*

Problem Statement

The volume of banking transactions conducted electronically, including through internet banking and mobile applications, is growing every year. Along with this, the number of fraudulent activities is also increasing, leading to significant financial losses for both banking institutions and their clients. According to international studies, annual losses from fraudulent operations in the banking sector amount to billions of dollars.

The problem is complicated by several factors. User behavior is highly variable: transactions may differ in amount, time, geographical location, and method of execution. This natural variability makes it difficult to establish clear criteria to distinguish legitimate operations from fraudulent ones. Fraudsters actively exploit modern technologies, using stolen or fabricated credentials, phishing schemes, identity theft, automated bots for large-scale attacks, and methods of concealing traces (such as IP spoofing or the use of VPNs). This forces security systems to constantly adapt to new attack scenarios. In typical datasets, fraudulent transactions usually account for less than 1 % of the total volume, resulting in a severe class imbalance. Under such conditions, standard machine learning algorithms often prove ineffective, as they tend to ignore the minority yet most critical class [1-2].

Modern approaches to fraud detection in financial transactions largely rely on machine learning methods. These methods can analyze large amounts of data, identify hidden patterns, and generate predictions based on complex multidimensional dependencies. However, the use of traditional models, such as logistic regression or classical decision trees, has limitations. They often demonstrate insufficient generalization ability and are not always effective in detecting

complex or atypical fraud scenarios. Another issue is the large number of false positives, where the system incorrectly classifies legitimate transactions as fraudulent. This not only creates additional workload for the bank's security services but also negatively affects the customer experience, as legitimate users face blocked or delayed transactions [3–4].

In this context, ensemble learning methods, which combine the results of multiple models, represent a promising direction. They improve anomaly detection efficiency by reducing the impact of random errors and generalizing information from different sources. Moreover, methods such as Random Forest, Gradient Boosting, and Extreme Gradient Boosting offer high flexibility and accuracy in identifying complex patterns, which is crucial in combating fraud. The application of these methods makes it possible not only to reduce the number of false results but also to adapt models to changes in user behavior and evolving fraud schemes [5–6].

The study of ensemble machine learning methods in fraud detection is both relevant and practically significant. It combines scientific interest, associated with the development of new models and approaches to big data analysis, with applied value for financial institutions seeking to minimize risks and strengthen customer trust in their services. This paper explores modern approaches to fraud detection, analyzes the specifics of using ensemble algorithms, and demonstrates their potential for improving the accuracy and reliability of financial security systems.

Analysis of Recent Studies and Publications

In works [7], methods of applying machine learning algorithms for fraud detection are described. In particular, the use of decision trees is considered, which provide basic efficiency but face limitations in cases of complex data patterns. The advantages of ensemble methods, such as Random Forest and Gradient Boosting, which demonstrate higher accuracy in classification tasks, are also analyzed in detail. Works [8] focus on the application of deep learning methods, which, although yielding good results, require significant computational resources and training time.

Despite the progress in this field, several issues remain unresolved. These include the optimal tuning of ensemble models, the selection of the most effective algorithms in the context of fraud detection, and the analysis of their resilience to changes in user behavior patterns. Particularly relevant is the development of approaches that ensure a balance between accuracy and real-time transaction processing speed, since delays in the financial sector may have critical consequences for both clients and banking institutions.

Formulation of the Research Objective

The objective of the study is to develop and substantiate an effective model for detecting fraudulent transactions in the banking sector through the application of ensemble machine learning methods. These methods are intended to increase classification accuracy and reduce the number of misclassified operations, while taking into account business requirements and the practical aspects of integration into financial systems.

Presentation of the Main Research Material

Ensemble learning methods are a powerful tool in classification and regression tasks. For fraud detection, the most effective algorithms are Random Forest, Gradient Boosting, AdaBoost, and Extreme Gradient Boosting. These methods combine the results of several base models, which reduces the impact of random errors and improves prediction accuracy. An important aspect of these methods is their ability to account for different types of dependencies within the data, which enhances their flexibility in complex tasks such as fraud detection.

Random Forest is known for its resistance to overfitting due to the use of random feature and data sampling when constructing each tree in the ensemble. Gradient Boosting, in turn, seeks to minimize the errors of previous models through sequential learning. Extreme Gradient Boosting, as an improved version of Gradient Boosting, ensures optimization of speed and performance through computational enhancements and regularization. AdaBoost focuses on correcting the errors of previous base models by assigning higher weights to difficult-to-classify samples. This feature makes it effective when working with data that exhibits high variability or anomalies.

In this study, data closely approximating real banking transactions with labeled fraudulent and legitimate operations were used. Preliminary data analysis included the removal of missing values to prevent their impact on modeling results. Normalization of values helped to reduce the influence of variable scales, while encoding categorical variables using the "one-hot encoding" method ensured the proper handling of non-numeric data.

In this research, the following formulas were used to evaluate the effectiveness of machine learning models in detecting fraud in banking transactions:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN};$$

$$\text{Recall} = \frac{TP}{TP + FN};$$

$$\text{Precision} = \frac{TP}{TP + FP};$$

$$\text{Fl} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}},$$

where TP – true positives (correctly identified fraudulent transactions), TN – true negatives (correctly identified legitimate transactions), FP – false positives (legitimate transactions incorrectly classified as fraudulent), FN – false negatives (fraudulent transactions that remain undetected).

Correlation analysis helped to identify the most significant variables for modeling, which improved the efficiency of model training. Particular attention was paid to reducing class imbalance, which is typical for fraud detection tasks. To address this, the SMOTE (Synthetic Minority Over-sampling Technique) method was applied, generating additional synthetic samples for the minority class of fraudulent transactions. This significantly improved model performance, especially in Recall and F1-score metrics, which are critically important for fraud detection tasks.

The experimental analysis demonstrated that combining ensemble learning methods with modern data preprocessing techniques allows achieving high levels of Accuracy, Recall, and Precision. Moreover, the models showed robustness to new types of data that were not used during training.

Fraud detection in transactions involves data collection, preprocessing, model selection and tuning, as well as evaluation and deployment. Data should be collected from various sources, such as banking transactions, log files, customer information, and purchase history. These data must include timestamps, transaction amounts, transaction locations, types of operations, as well as payer and recipient information. This may require access to internal databases.

A key step is preprocessing the collected datasets. It is necessary to remove duplicates, handle missing values, and eliminate anomalies. Afterward, the data should be normalized and, if needed, clustered. Categorical variables must be encoded into numerical values. The creation of new informative features can significantly enhance model performance. At the same time, it is important to select the most relevant features. New features may include, for example, the average transaction amount or the number of transactions performed by a particular user within a day or another time period.

Fraudulent transactions usually account for less than 1 % of total payments, leading to imbalance in the collected datasets. Working with such data may be ineffective. This issue can be addressed through resampling methods, such as oversampling (e.g., SMOTE) or undersampling. After this, an appropriate method must be chosen to build a machine learning model. Since the task is a classification problem, possible approaches include logistic regression, decision trees, Random Forest, Gradient Boosting, and deep learning. Each of these models has its advantages and disadvantages, and the final choice depends on the specific requirements and conditions.

Once the model has been trained, it should be evaluated using relevant metrics. The main metrics include Accuracy, Recall, Precision, F1-score, and ROC-AUC. These metrics help assess how well the model detects fraudulent transactions and distinguishes them from legitimate ones [7].

Logistic regression is the most common method for solving classification problems, including fraud detection in transactions. It is based on modeling the probability of an object belonging to one of two classes using a linear combination of input features. Logistic regression employs a sigmoid function to transform the linear combination of features into the probability that a transaction is fraudulent.

In the course of the study, ensemble methods Random Forest, Gradient Boosting, XGBoost, and AdaBoost were built and tested for the task of detecting fraudulent transactions. The data were preprocessed by cleaning, normalizing, encoding categorical variables, and balancing classes using SMOTE, which significantly improved the results.

Table 1

Table of Research Results

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Random Forest	0.9995	0.9610	0.7551	0.8457	0.9580
Gradient Boosting	0.9993	0.9100	0.8015	0.8522	0.9642
XGBoost	0.99954	0.8830	0.8469	0.8646	0.9691
AdaBoost	0.9989	0.8702	0.7024	0.7762	0.9427

The ROC curve is a graphical representation of the performance of a classification model at different decision thresholds. It shows the relationship between the True Positive Rate (TPR, or Recall) and the False Positive Rate (FPR).

The graph displays ROC curves for all models. XGBoost has the highest AUC score (0.9691), indicating the best ability to distinguish fraudulent transactions from legitimate ones. Gradient Boosting (0.9642) shows results close to XGBoost. Random Forest (0.9580) also demonstrates strong classification performance. AdaBoost (0.9427), however, lags behind the others, having the lowest AUC value.

The confusion matrices provide insights into the real-world performance of the models on positive (fraudulent) and negative (legitimate) transactions. Random Forest achieves the highest Precision (0.9610), meaning it produces the fewest false positives. However, its Recall (0.7551) is lower, so some fraudulent transactions remain undetected. Gradient Boosting delivers a higher Recall (0.8015) compared to Random Forest, but its Precision (0.9100) is slightly reduced. This results in a greater proportion of fraud detection but also an increase in false alarms. XGBoost provides the best balance between Precision and Recall, with the highest Recall (0.8469) and the highest F1-score (0.8646), making it the

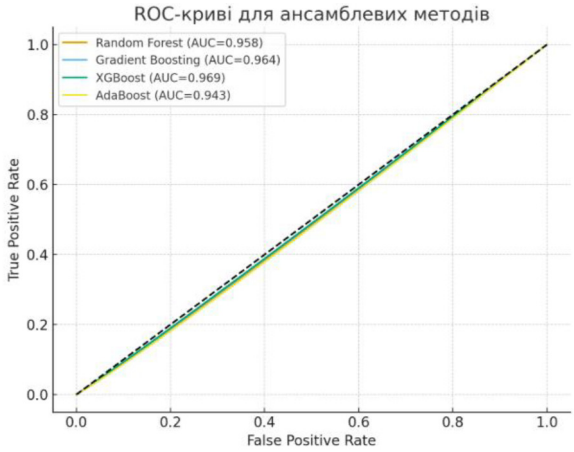


Рис. 1. ROC Curve for Ensemble Methods

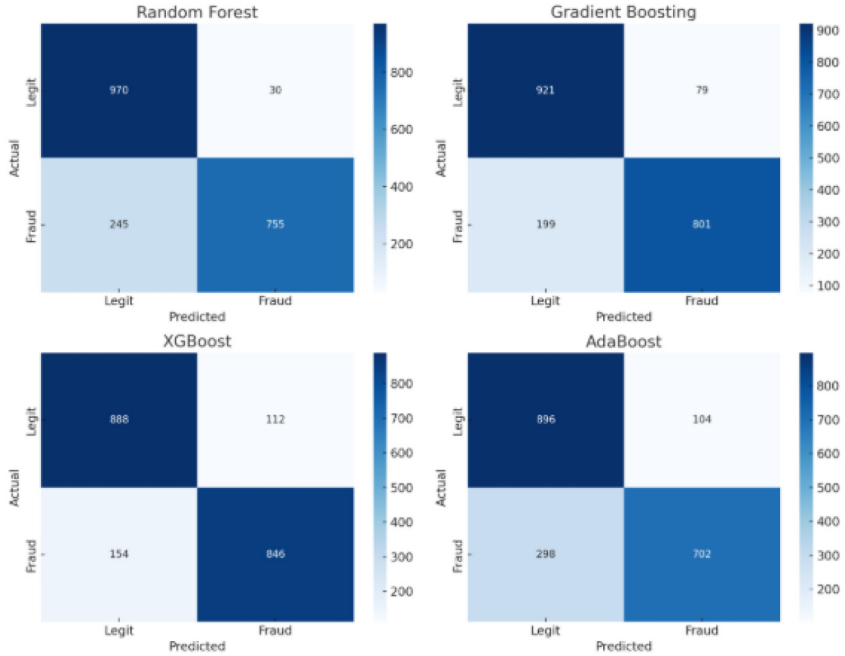


Рис. 2. Confusion Matrix for Ensemble Models

most effective for practical use. AdaBoost performs the weakest among the models: its Recall (0.7024) is the lowest, and although its Precision (0.8702) is acceptable, the overall F1-score is significantly lower.

In this section, the performance of classification models was analyzed using the ROC curve and the Confusion Matrix. The obtained results demonstrate that the models are capable of effectively separating fraudulent transactions from legitimate ones while maintaining balanced levels of accuracy and sensitivity. This confirms the feasibility of applying such models for fraud detection tasks in financial systems.

Conclusions

The conducted study confirmed the effectiveness of applying ensemble machine learning methods to the task of fraud detection in banking transactions. The use of algorithms such as Random Forest, Gradient Boosting, Extreme Gradient Boosting, and AdaBoost made it possible to achieve high levels of accuracy, sensitivity, and specificity. These models demonstrated the ability to effectively identify complex patterns in data, reduce the impact of random errors, and ensure reliable results under various conditions. Particularly important is the significant reduction in false positives, which increases customer trust in banking services and allows financial institutions to focus on genuine threats.

A substantial contribution to the research results was made by the integration of modern data preprocessing techniques. In particular, handling missing values, value normalization, categorical variable encoding using one-hot encoding, and applying SMOTE for class balancing improved the quality of model training. These approaches enabled the consideration of real-world data characteristics, such as class imbalance and high variability. The successful combination of data

preprocessing with ensemble algorithms resulted in robust models capable of performing effectively even on large and complex datasets.

The scientific novelty of the research lies in its comprehensive approach to the analysis and optimization of ensemble methods for fraud detection. Specifically, combining various ensemble learning algorithms with advanced preprocessing techniques allowed for the achievement of new results, which can serve as a foundation for further studies in this field. Additionally, the research demonstrated the high effectiveness of machine learning methods in detecting complex anomalies in data, making the findings applicable to a wide range of tasks.

The practical significance of this work lies in the potential to implement the proposed approaches to enhance the efficiency of existing banking transaction monitoring systems. In particular, the research results can be used to develop automated fraud detection systems capable of operating in real time while maintaining high accuracy and processing speed. Beyond the financial sector, the approaches developed in this study can be adapted to other industries, such as insurance, e-commerce, and telecommunications, where effective anomaly detection is crucial.

The results obtained highlight the promising potential of further research into ensemble methods and their integration with new data analysis techniques, which will further improve the effectiveness of fraud detection systems and related applications.

Bibliography

1. Almarshad F. A., Gashgari G. A., Alzahrani A. I. A. Generative adversarial networks-based novel approach for fraud detection for the European cardholders 2013 dataset. *IEEE Access*. Vol. 11. 2023. Pp. 107348–107368.

2. Hancock J. T., Bauder R. A., Wang H., Khoshgoftaar T. M. Explainable machine learning models for Medicare fraud detection. *Journal of Big Data*. Vol. 10, no. 1. 2023.

3. Alsayaydeh J. A. J., Aziz A., Rahman A. I. A. Development of programmable home security using GSM system for early prevention. *ARPN Journal of Engineering and Applied Sciences*. Vol. 16, no. 1. 2021. Pp. 88–97.

4. Fedorchenko I., Oliinyk A., Alsayaydeh J. A. J. Modified genetic algorithm to determine the location of the distribution power supply networks in the city. *ARPN Journal of Engineering and Applied Sciences*. Vol.15, no. 23. 2020. Pp. 2850–2867.

5. Shakhovska N., Liaskovsky D., Augousti A., Liaskovska S., Martyn Y. Design and deployment of data developer toolkit in cloud manufacturing environments. *CEUR-WS*. Vol. 3699. 2024. Pp. 47–56.

6. Fedorchenko V., Yeroshenko O., Shmatko O., Kolomiitsev O., Omarov M. (2024) Password hashing methods and algorithms on the.Net platform, *Advanced Information Systems*. Vol. 8, no. 4. 2024. Pp. 82–92.

7. Islam M. A., Uddin M. A., Aryal S., Stea G. An ensemble learning approach for anomaly detection in credit card data with imbalanced and overlapped classes. *Journal of Information Security and Applications*. Vol. 78. 2023.

8. Abdul Salam M., Fouad K. M., Elbably D. L., Elsayed S. M. Federated learning model for credit card fraud detection with data balancing techniques. *Neural Computing and Applications*. Vol. 36, no. 11. 2024. Pp. 6231–6256.

References

1. Almarshad, F. A., Gashgari, G. A., Alzahrani, A. I. A. (2023). Generative adversarial networks-based novel approach for fraud detection for the European cardholders 2013 dataset. *IEEE Access*. vol. 11, pp. 107348–107368.

2. Hancock, J. T., Bauder, R. A., Wang, H., Khoshgoftaar, T. M. (2023). Explainable machine learning models for Medicare fraud detection. *Journal of Big Data*. vol. 10, no. 1.

3. Alsayaydeh, J. A. J., Aziz, A., Rahman, A. I. A., et al. (2021). Development of programmable home security using GSM system for early prevention. *ARPN Journal of Engineering and Applied Sciences*. vol. 16, no. 1, pp. 88–97.

4. Fedorchenko, I., Oliinyk, A., Alsayaydeh, J. A. J., et al. (2020). Modified genetic algorithm to determine the location of the distribution power supply networks in the city. *ARPN Journal of Engineering and Applied Sciences*. Vol.15, no. 23, pp. 2850–2867.

5. Shakhovska, N., Liaskovsky, D., Augousti, A., Liaskovska, S., Martyn, Y. (2024). Design and deployment of data developer toolkit in cloud manufacturing environments. *CEUR-WS*. vol. 3699, pp. 47–56.

6. Fedorchenko, V., Yeroshenko, O., Shmatko, O., Kolomiitsev, O., Omarov, M. (2024) Password hashing methods and algorithms on the.Net platform, *Advanced Information Systems*, vol. 8, no. 4, pp. 82–92.

7. Islam, M. A., Uddin, M. A., Aryal, S., Stea, G. (2023). An ensemble learning approach for anomaly detection in credit card data with imbalanced and overlapped classes. *Journal of Information Security and Applications*. vol. 78.

8. Abdul Salam, M., Fouad, K. M., Elbably, D. L., Elsayed, S. M. (2024). Federated learning model for credit card fraud detection with data balancing techniques. *Neural Computing and Applications*. vol. 36, no. 11, pp. 6231–6256.

Дата першого надходження рукопису до видання: 21.09.2025
Дата прийнятого до друку рукопису після рецензування: 17.10.2025
Дата публікації:

УДК 004.658.6:004.056.53 DOI

О. В. ЗАДЕРЕЙКО
кандидат технічних наук, доцент,
доцент кафедри інформаційних технологій
Національний університет «Одеська юридична академія»
ORCID: 0000-0003-0497-9861

С. І. ЄЛЕНИЧ
головний бібліограф наукової бібліотеки
Національний університет «Одеська юридична академія»
ORCID: 0000-0003-4999-3034

Н. І. ЛОГІНОВА
кандидат педагогічних наук, доцент,
завідувачка кафедри інформаційних технологій
Національний університет «Одеська юридична академія»
ORCID: 0000-0002-9475-6188

О. Г. ТРОФИМЕНКО
кандидат технічних наук, доцент,
доцент кафедри інформаційних технологій
Національний університет «Одеська юридична академія»
ORCID: 0000-0001-7626-0886

В. І. ГУРА
кандидат технічних наук,
доцент кафедри інформаційних технологій
Національний університет «Одеська юридична академія»
ORCID: 0009-0001-2166-5410

АНАЛІЗ РЕЗУЛЬТАТІВ ПОШУКУ НАУКОВОЇ ІНФОРМАЦІЇ З ВИКОРИСТАННЯМ ПЛАТФОРМ ШТУЧНОГО ІНТЕЛЕКТУ

У статті досліджується вірогідність генерації платформами штучного інтелекту (ШІ) неточних або неіснуючих чи вигаданих бібліографічних описів наукових публікацій. Звертається увага щодо можливих обмежень використання платформ ШІ при здійсненні пошуку джерел наукової інформації, ґрунтуючись та результатах проведення їх верифікації. Для досягнення найкращих результатів пошуку наукової інформації пропонується використовувати вибірковий метод пошуку, який передбачає формування пошукового запиту за чітко визначеними критеріями дослідника. Для оцінки релевантності і достовірності наукових публікацій, отриманих з результатів пошукової видачі платформ ШІ було застосовано порівняльний аналіз результатів ТОП20 пошукової видачі. Інформаційною базою дослідження обрано наукові публікації, які знайдені з використанням типового пошукового запиту. Визначено базові критерії оцінювання релевантності результатів пошуку джерел інформації при використанні платформ ШІ. Дослідження містить аналіз бібліографічних даних наукових публікацій із списків літератури сформованих платформами ШІ «ChatGPT5», «Gemini», «Grok», «DeepSeek AI», «Cloude». Комплексне оцінювання джерел наукової інформації проводилось за критеріями оцінювання релевантності для обраної групи ключових слів «складні системи», «цифрові системи», «штучний інтелект», «аналіз даних». Перевірка та підтвердження існування наукових публікацій була здійснена із застосуванням наукометричних баз даних НБУВ, Google Scholar, Research Gate, репозитаріїв ЗВО. Доведено, що для збору достовірної наукової інформації слід використовувати максимально можливу кількість профільних електронних наукометричних баз даних, електронних бібліотек, архівів тощо.

Сформовані практичні рекомендації дослідникам при застосуванні платформ ШІ в науковій діяльності.

Ключові слова: *платформи штучного інтелекту, пошукова видача, достовірність наукових публікацій, релевантність наукових публікацій, наукометричні бази даних.*

© Задерейко О. В., Єленич С. І., Логінова Н. І., Трофименко О. Г., Гура В. І., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

O. V. ZADEREYKO
PhD, Associate Professor,
Associate Professor at the Department of Information Technologies
National University "Odesa Law Academy"
ORCID: 0000-0003-0497-9861

S. I. YELENYCH
Chief Bibliographer of the Scientific Library
National University "Odesa Law Academy"
ORCID: 0000-0003-4999-3034

N. I. LOGINOVA
PhD, Associate Professor,
Head of the Department of Information Technologies
National University "Odesa Law Academy"
ORCID: 0000-0002-9475-6188

O. G. TROFYMENKO
PhD, Associate Professor,
Associate Professor at the Department of Information Technologies
National University "Odesa Law Academy"
ORCID: 0000-0001-7626-0886

V. I. HURA
PhD, Associate Professor,
Associate Professor at the Department of Information Technologies
National University "Odesa Law Academy"
ORCID: 0009-0001-2166-5410

ANALYSIS OF SCIENTIFIC INFORMATION SEARCH RESULTS USING ARTIFICIAL INTELLIGENCE PLATFORMS

The article examines the likelihood of artificial intelligence (AI) platforms generating inaccurate, non-existent, or fictitious bibliographic descriptions of scientific publications. Attention is drawn to the possible limitations of using AI platforms when searching for sources of scientific information, based on the results of their verification. To achieve the best results in searching for scientific information, it is proposed to use a selective search method, which involves forming a search query according to clearly defined criteria of the researcher. To assess the relevance and reliability of scientific publications obtained from the search results of AI platforms, a comparative analysis of the top 20 search results was used. Scientific publications found using a typical search query were selected as the information base for the study. Basic criteria for evaluating the relevance of search results for information sources when using AI platforms were determined. The study contains an analysis of bibliographic data of scientific publications from the lists of references generated by the AI platforms ChatGPT5, Gemini, Grok, DeepSeek AI, and Cloude. A comprehensive assessment of scientific information sources was carried out using relevance assessment criteria for the selected group of keywords: "complex systems", "digital systems", "artificial intelligence" and "data analysis". The existence of scientific publications was verified and confirmed using the scientometric databases of the National Library of Ukraine, Google Scholar, Research Gate, and higher education institution repositories. It has been proven that in order to collect reliable scientific information, it is necessary to use as many specialized electronic scientometric databases, electronic libraries, archives, etc. as possible.

Practical recommendations have been developed for researchers on the use of AI platforms in scientific activities.

Key words: *artificial intelligence platforms, search engine optimization, reliability of scientific publications, relevance of scientific publications, scientometric databases.*

Постановка проблеми

Розвиток інформаційних технологій характеризується активним використанням платформ штучного інтелекту (ШІ) у науковій діяльності дослідників. Традиційні підходи до пошуку наукової інформації замінюються складними алгоритмами машинного навчання, здатними здійснювати семантичний аналіз текстів та виявляти контекстуальні зв'язки між публікаціями. Різноманітні платформи ШІ пропонують унікальні можливості для підбору релевантних джерел наукової інформації у галузі 12 «Інформаційні технології». Платформи ШІ здатні збирати та аналізувати великі масиви наукових публікацій з різних наукометричних баз даних, формувати результати пошукової видачі, ґрунтуючись на семантичних зв'язках між науковими публікаціями.

Стрімкий розвиток інформаційних технологій надає можливість платформам ШІ глибше зрозуміти контекст запиту дослідника і завдяки цьому знаходити релевантні наукові публікації, що відповідають його пошуковим запитам. Це є особливо актуальним для дослідників в умовах сталого розвитку галузі комп'ютерних наук та розміщення наукових публікацій у відкритому доступі на цифрових платформах.

Мовні моделі платформ ШІ переходять від класичного пошуку інформації до генерації релевантних відповідей на пошукові запити. Вони аналізують інформаційні потреби дослідників, здійснюють пошук у своїй внутрішній базі знань або у відкритих джерелах інформації та намагаються синтезувати доречні текстові відповіді на поставлені запити. Платформи ШІ спроможні знаходити релевантні наукові публікації та формувати з них перелік наукових джерел, створювати графічні інтерпретації даних, таблиці, діаграми, презентації і т.д.

Роль платформ ШІ як допоміжного інструмента у сфері наукових досліджень швидко зростає. Мовні моделі платформ ШІ можна використовувати для автоматизації процесу підбору наукових публікацій, здійснення їх контекстуального аналізу, структуризації та візуалізації досягнутих результатів. Це відкриває шлях для інтеграції платформ ШІ у повний цикл проведення наукового дослідження – від початкового задуму до публікації наукових результатів.

Актуальність дослідження обумовлена потребою з'ясувати можливості пошукових алгоритмів мовних моделей ШІ в процесі обробки джерел наукової інформації. Однією із проблем, з якою стикається дослідник, це здатність платформ ШІ генерувати текст, надавати неточний опис наукових публікацій та надавати неіснуючі інтернет-посилання на них. Тому існує нагальна потреба у порівняльному аналізі релевантності результатів пошукової видачі платформ ШІ, зокрема ChatGPT5 [17], Gemini [20], Grok [22], DeepSeek AI [19], Cloude [18]. Особливої уваги потребує етап оцінювання релевантності списку наукових публікацій, отриманих з використанням вище зазначених платформ ШІ. Процес оцінки релевантності наукових публікацій передбачає, насамперед, перевірку їх на предмет достовірності. Тобто дослідникові необхідно перевірити наявність знайдених наукових публікацій у наукометричних базах даних та відповідність їх інтернет-посилань у результатах пошукової видачі сформованої ШІ.

Аналіз останніх досліджень і публікацій

Дослідження ефективності роботи платформ ШІ привертає увагу наукової спільноти. Вивчення досвіду застосування платформ ШІ в науковій діяльності доводить, що їх використання здатне значно прискорити етапи досліджень, допомогти у плануванні експериментів і аналізі великих обсягів даних. Водночас, варто враховувати, що інтеграція платформ ШІ у дослідницьку сферу супроводжується певними викликами та ризиками, такими як потенційне зниження глибини пізнання, формування надмірної залежності від платформ ШІ. Проте, разом із цим, виникають етичні питання щодо авторства (плагіат), права інтелектуальної власності на результати дослідження, створені за допомогою платформ ШІ, достовірності даних, отриманих від алгоритмів, які можуть містити упередження або помилки [13, с. 64; 12, с. 38; 8, с. 47]. Якщо дані, які використовують платформи ШІ для навчання чи аналізу, будуть упередженими або неповними, то результати їх оцінювання можуть виявитися необ'єктивними і як наслідок, привести до помилкових висновків.

Найбільш популярною платформою ШІ, яка використовується дослідниками є ChatGPT. Вона дозволяє часткового оптимізувати роботу дослідника для спрощення виконання типових завдань. Значна увага авторів публікацій [3; 4] приділяється аналізу функціоналу ChatGPT на виявлення ряду обмежень при його використанні. Наголошується про необхідність враховувати застереження щодо використання мовних моделей на основі ШІ в наукових дослідженнях [1, с. 5; 5, с. 50]. Застереження стосується питань дотримання етичних норм та вимог академічної доброчесності, створення нових ідей за допомогою платформ ШІ, можливості генерації ними неточної інформації та хибних даних. В публікації [14, с. 371–372] наводиться приклад, який стосується списку використаних джерел, що ChatGPT надає перелік вигаданих назв наукових джерел та інформацію про їх авторів. Таким чином, виникає вірогідність генерації платформами ШІ неіснуючих бібліографічних описів наукових публікацій. Це обумовлює необхідність аналізу результатів пошукової видачі наукових публікацій, отриманих за допомогою платформ ШІ.

Формулювання мети дослідження

Метою статті є аналіз достовірності наукових публікацій отриманих з використанням платформ ШІ ChatGPT5, Gemini, Grok, DeepSeek AI, Cloude та оцінка релевантності їх пошукової видачі.

Завданнями статті є:

– здійснити перевірку пошукової видачі наукових публікацій платформ ШІ у наукометричних базах даних НБУВ, Google Scholar, ResearchGate та репозитаріях ЗВО;
– провести аналіз достовірності наукових публікацій у результатах пошукової видачі платформ ШІ;
– оцінити релевантність та достовірність наукових публікацій, знайдених за допомогою платформ ШІ;
– з'ясувати та оцінити наявність доступу платформ ШІ до наукометричних баз даних.

Викладення основного матеріалу дослідження

В умовах постійного зростання кількості наукових публікацій та необхідності швидкого виявлення найбільш релевантних джерел наукової інформації дослідникові слід обирати оптимальну стратегію пошуку. Ефективний пошук джерел наукової інформації ґрунтується на вмінні працювати з інформаційно-пошуковими системами

[12, с. 10–109]. Отримання максимально можливих релевантних результатів пошукової видачі є складним процесом, що вимагає від дослідника вміння користуватися вбудованими інструментами та функціоналом цифрових ресурсів, а саме: пошукових систем відкритого доступу, наукометричних баз даних, електронних архівів (репозитаріїв), електронних ресурсів бібліотек, наукових порталів, інтернет форумів, тематичних спільнот, платформ ШІ. Слід також враховувати, що цифрові ресурси мають свої алгоритми ранжування та містять різний обсяг наукових даних.

Серед інструментів, якими користується дослідник для пошуку інформації є мовні моделі платформ ШІ. Вони використовуються у науковій діяльності як засіб для скорочення витрат особистого часу дослідника на процес пошуку наукових публікацій. Однак, досліднику необхідно застосовувати обґрунтовані критерії в оцінюванні результатів пошукової видачі. Сам процес оцінювання тісно пов'язаний із перевіркою достовірності отриманих наукових публікацій та аналізом їхньої релевантності. Навіть якщо наукові публікації здаються достовірними на перший погляд, досліднику варто перевіряти їх первісні джерела. Завжди існує вірогідність, що наукові публікації, отримані з використанням платформ ШІ можуть бути неіснуючими, тобто недостовірними.

Для оцінювання джерел наукової інформації, згенерованої платформами ШІ, необхідно використовувати ті ж самі критерії, що й для оцінювання пошукової видачі наукометричних баз даних та інших цифрових ресурсів. Однак використання платформ ШІ у науковій діяльності додає нові критерії. Отже, можна запропонувати два етапи оцінювання релевантності наукових публікацій.

На першому етапі дослідник застосовує загальні критерії оцінювання релевантності такі, як точність (співпадіння контексту), актуальність (за роком видання), авторитетність джерела інформації, відповідність пошуковому запиту дослідника.

Критерій «*точність пошукової видачі*» це показник того, наскільки релевантні результати отримує дослідник у відповідь на запит.

Критерій «*актуальність інформації*» характеризується новизною досліджень. Це означає, що джерела відбираються та групуються за датою їх створення.

Критерій «*авторитетність джерела інформації*» визначається кількісними та якісними параметрами. Кількісні параметри оцінювання релевантності наукових публікацій враховують наукометричні показники. Для того, щоб переконатися в надійності та авторитетності наукового видання можна скористатись ресурсами: чи індексується видання в наукометричних базах даних Scopus та Web of Science (WoS), провести оцінку журналу на порталі «The SCImago Journal & Country Rank», визначитись на сайті «Think.Check.Submit» (https://thinkchecksubmit.org/), що обрані наукові публікації видані в перевірених журналах та видавництвах [2, с. 270–271].

Якісна оцінка пошукової видачі платформ ШІ визначається суб'єктивними критеріями дослідника. Його оціночні показники обумовлюються рівнем знань у галузі 12 «Інформаційні технології» та його практичного досвіду використання платформ ШІ. Якщо дослідник незадоволений результатами пошуку, то він може змінити або додати умови запиту, зокрема збільшити або зменшити кількість джерел чи зробити акцент на вибірковому методі пошуку наукової інформації, наприклад, обмежений пошук за видами джерел (монографії, підручники, навчальні посібники, навчально-методичні посібники, збірники наукових праць, матеріали конференцій).

Критерій «*відповідність запиту дослідника*» це показник задоволеності результатами пошукової видачі платформ ШІ та довіри отриманим результатам.

На другому етапі дослідник приділяє увагу достовірності наукової інформації. Відсутність або низький рівень достовірності фіксується у випадках, коли результати пошукової видачі представлені науковими публікаціями, які містять неточні, неповні, нечіткі дані чи містять узагальнені посилання, що ускладнює їх перевірку [6, с. 56–57]. Тому перевірка на достовірність знайдених наукових публікацій є важливим етапом для забезпечення якості результатів власного наукового дослідження, академічної доброчесності та запобігання використанню недостовірної або вигаданої псевдо наукової інформації.

Критерій «достовірність наукової публікації» забезпечується проведенням ідентифікації публікації за її бібліографічним описом в інформаційно-пошукових системах. Цифрові ресурси надають змогу швидко знайти та підтвердити існування джерела наукової інформації. Системна перевірка на достовірність допомагає забезпечити надійність знайдених джерел наукової інформації.

Для здійснення пошуку наукових публікацій необхідно обрати параметри запиту на платформі ШІ. Умови пошуку охоплюють визначення у галузі 12 «Інформаційні технології», ключового слова, бажаної кількості наукових публікацій та формату їх представлення. Типовий запит для формування пошукової видачі платформ ШІ [17; 20; 22; 19; 18] було здійснено за ключовими словами «складні системи», «цифрові системи», «штучний інтелект», «аналіз даних». Типовий запит сформульовано таким чином: *«Знайди 20 релевантних джерел (Україна). Використай ключове слово «складні системи». Оформлення джерел здійсни відповідно ДСТУ 8302:2015»*. Обсяг знайдених наукових публікацій в результатах пошуковій видачі платформ ШІ характеризується різною кількістю в одному запиті – приблизно від 10 до 50. Якщо кількість знайдених наукових публікацій є меншою ніж вказана у пошуковому запиті, то доводиться робити додатковий запит для досягнення потрібної кількості результатів

пошуку. В процесі здійснення пошуку джерел наукової інформації платформи ШІ можуть надати повний або не повний опис наукових публікацій відповідно до вимог оформлення ДСТУ 8302:2015.

В результаті проведеного дослідження, автори отримали окремі списки наукових публікацій із 50 джерел для кожного ключового слова, що були сформовані платформами ШІ: ChatGPT5, Gemini, Grok, DeepSeek, Cloude. Щоб отримати максимально можливу релевантну пошукову видачу відповідно запиту було використано функцію «повторити відповідь», так і для кожного запиту було обрано «новий чат» (платформи Gemini, Cloude). В залежності від отриманих результатів пошукової видачі платформ ChatGPT5, DeepSeek необхідно додатково уточнювати пошукові запити для досягнення потрібної кількості наукових публікацій.

При здійсненні додаткового пошукового запиту для збільшенні кількості знайдених наукових публікацій «до 50» платформою ШІ «Gemini» були надані рекомендації використати цифрові ресурси Національної бібліотеки України імені В. І. Вернадського (НБУВ) [7], Google Scholar [21] та інших наукометричних баз даних. Деякі відповіді Gemini містять список назв спеціалізованих журналів. Платформа ШІ «Cloude» надає перелік закладів вищої освіти України, кафедр, яких займаються вивченням цієї проблематики. Слід також відмітити, що зазначені вище платформи ШІ пропонують дослідникові приклади оформлення джерел інформації відповідно заявленим вимогам (ДСТУ 8302:2015).

Проте, аналіз результатів пошукової видачі ШІ виявив невідповідність даних в описі джерел наукової інформації. Наведемо приклад, знайдений в результатах пошукової видачі «ChatGPT5» на запит по ключовому слову «складні системи». Бібліографічний опис книги «Теорія хаосу в складних інформаційних системах» В. І. Шевченка (рис. 1) містить суттєві розбіжності у відомостях про «місце видання», «назву видавництва», «рік видання» та «кількість сторінок».

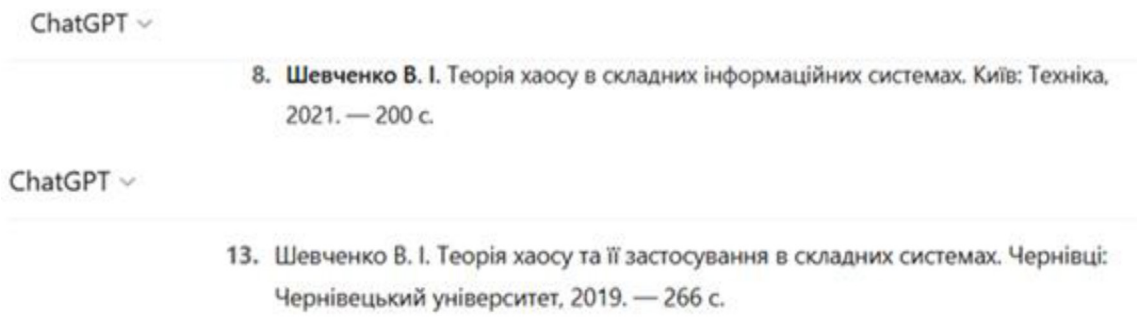


Рис. 1. Приклад згенерованого опису наукової публікації у результатах пошукової видачі платформи ChatGPT5

Для перевірки вище наведених результатів (див. рис. 1) було виконано пошук у наукометричній базі Google Scholar [21] за назвою та прізвищем автора «Теорія хаосу в складних інформаційних системах В. І. Шевченко» для того щоб з'ясувати існування цього джерела наукової інформації. Додатково було застосовано фільтрацію пошукової видачі, а саме: відокремлено рік видання «2021» та вид видання «книга». При аналізі перших 50 джерел пошукової видачі таку книгу не знайдено. При внесенні змін щодо року видання – «2021» на «2019», в результатах пошукової видачі книгу «Теорія хаосу в складних інформаційних системах» В. І. Шевченка або іншого автора не було знайдено.

Далі було прийняте рішення продовжити пошук цієї книги в фонді НБУВ [7]. НБУВ входить до переліку організацій, яка одержує обов'язковий примірник документів [9]. Пошуковий запит містив ключові слова із назви «Теорія хаосу в складних інформаційних системах». В результатах пошукової видачі було отримано тільки одну наукову публікацію: «Комп'ютерне моделювання детермінованого хаосу в складних нелінійних системах» (А. Ю. Зінченко) (рис. 2). Перевірка довела, що книга «Теорія хаосу в складних інформаційних системах В. І. Шевченко» відсутня у фонді НБУВ.

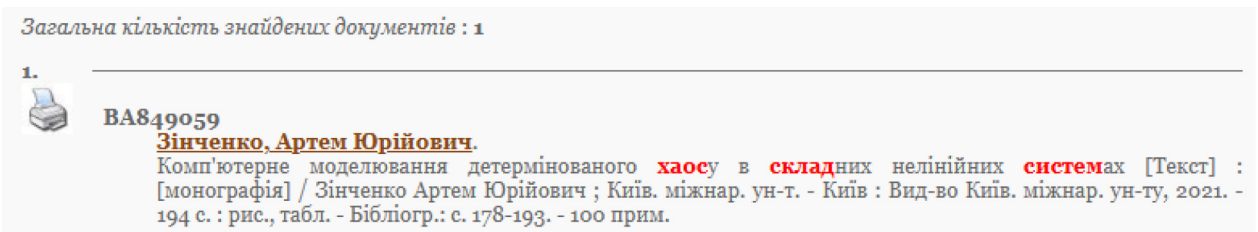


Рис. 2. Результат перевірки наявності книги «Теорія хаосу в складних інформаційних системах» у фонді НБУВ

Таким чином, здійснена перевірка довела, що книга «Теорія хаосу в складних інформаційних системах В. І. Шевченко» відсутня у фонді НБУВ.

В подальшому дослідник може витратити час на пошук цього джерела наукової інформації, використовуючи інші цифрові платформи.

Схожий приклад можна побачити в результатах пошукової видачі платформи ШІ «Gemini». При виконанні запиту в результатах пошукової видачі по ключовому слову «складні системи» є публікація автора О. М. Васильєва, яка привертає увагу ідентичною назвою «Моделювання та аналіз поведінки складних систем у динамічному середовищі» (рис. 3).

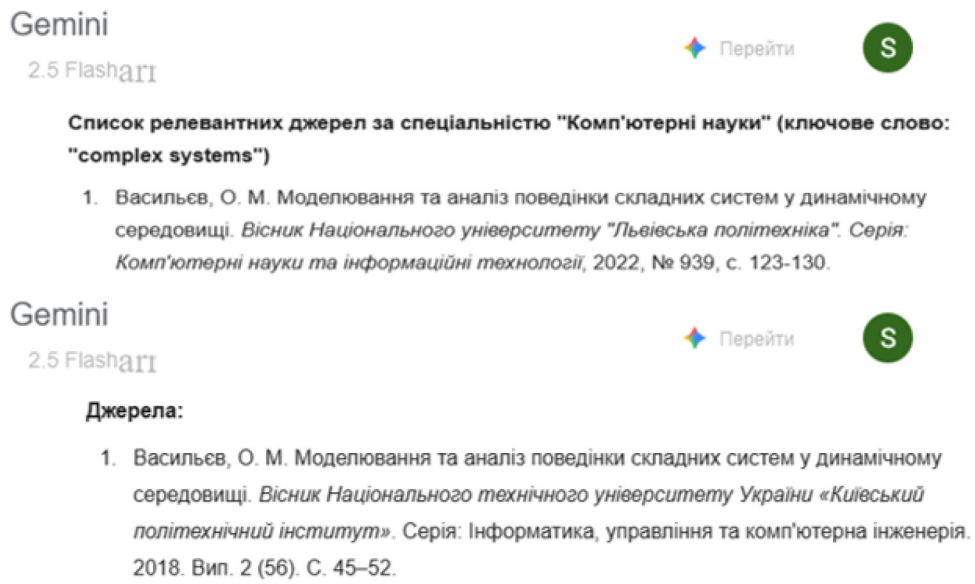


Рис. 3. Приклад опису наукової публікації з результатів пошукової видачі платформи ШІ «Gemini»

При цьому опис джерела інформації відрізняється бібліографічними даними. Але виникає питання: яке з періодичних видань дійсно є публікацією?

Таким чином, коли пошукова видача, згенерована ШІ, містить публікації з ідентичними заголовками, дослідникові слід обов'язково здійснити перевірку існування наданих наукових публікацій. В результаті процес оцінювання релевантності наукових публікацій стає значно тривалішим.

Результати, надані платформою ШІ при виконанні запиту, вказують досліднику на те, що його можливості пошуку наукових публікацій обмежені доступом до електронних ресурсів. Саме тому досліднику наполегливо рекомендується провести пошук самостійно у наукометричних базах даних, електронних бібліотеках чи архівах закладів вищої освіти України.

Для наочного прикладу було здійснено перевірку за допомогою електронного ресурсу «Наукова періодика» НБУВ [7] наявність наукових журналів «Проблеми управління та інформатика» та «Кібернетика та системний аналіз», які платформа ШІ Gemini пропонує в своїй пошуковій видачі (рис. 4).

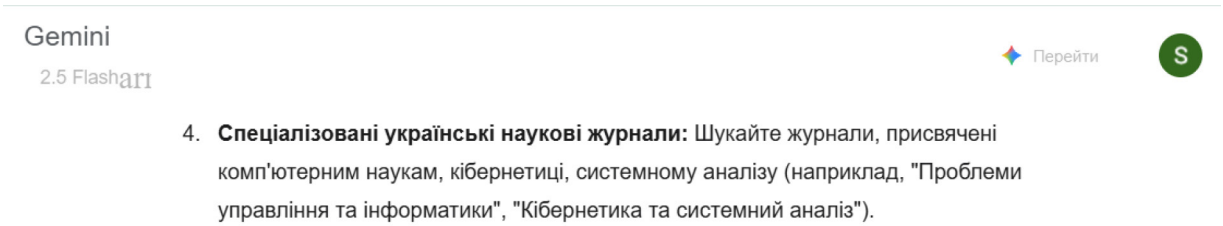


Рис. 4. Приклад рекомендованих джерел наукової інформації із результатів пошукової видачі платформи ШІ «Gemini»

В результаті виконаної перевірки з'ясовано, що журнал «Кібернетика та системний аналіз» (http://www.kibernetika.org/) існує та його назва повністю відповідає офіційним даним (рис. 5).

В бібліографічному описі журналу зазначається, що у 2021 році відбулася зміна назви журналу (див. рис. 5). Для опису статей, опублікованих в цьому журналі до 2021 року, слід використовувати попередню назву видання.

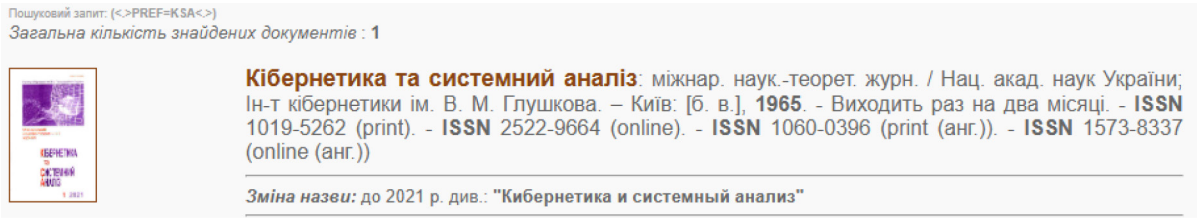


Рис. 5. Результат перевірки наявності журналу «Кібернетика та системний аналіз» в базі даних НБУВ

Слід зазначити, що факт зміни назви журналу є важливим для ідентифікації видання, точного посилання та коректного цитування наукових праць. Ця інформація допоможе дослідникові уникнути плутанини під час пошуку статей та оформлення списку літератури.

Під час пошуку журналу «Проблеми управління та інформатика» виявили, що офіційна назва цього журналу «Проблеми керування та інформатика» (https://jais.net.ua/index.php/files) (рис. 6).

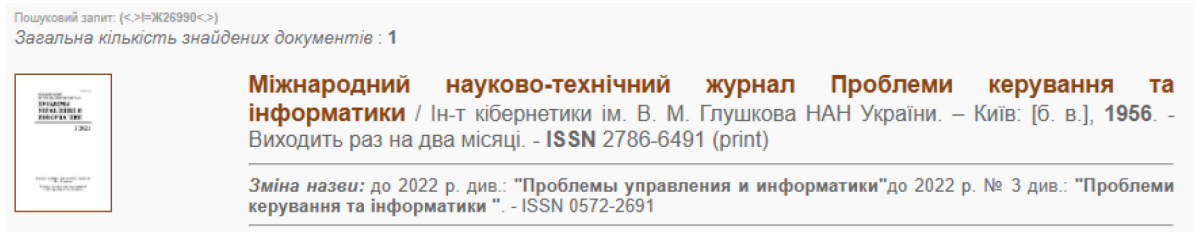


Рис. 6. Результат перевірки наявності журналу «Проблеми управління та інформатика» в базі даних НБУВ

В результаті слід зафіксували розбіжність у назві журналу, а саме заміна слова «керування» на «управління» (див. рис. 6). Припустимо, що виявлена невідповідність пов'язана з особливостями автоматичного перекладу, здійсненого платформою ШІ Gemini, попередньої назви журналу «Проблемы управления и информатика». Термін «управление» може бути перекладений як «управління», так і «керування». У даному випадку перекладач, ймовірно, використав найбільш поширений варіант.

Отже, наведені приклади із результатів пошукової видачі платформ ШІ ChatGPT5 та Gemini свідчать, що вони здатні створювати бібліографічний опис вигаданих або згенерованих наукових публікацій. Використання платформ ШІ у науковій діяльності передбачає обов'язкове оцінювання дослідником достовірності результатів пошуку наукових публікацій. У випадку, якщо виникає необхідність ідентифікувати та знайти наукову публікацію, дослідник має працювати з наукометричними базами даних та іншими цифровими платформами для того щоб перевірити достовірність наукових публікацій.

Перевірка достовірності наукових публікацій у результатах пошукової видачі платформ ШІ ChatGPT5, Gemini, Grok, DeepSeek, Cloude була виконана за допомогою наукометричних баз даних НБУВ [7; 15, с. 41–44], Google Scholar [21], Research Gate та репозитаріїв ЗВО. Вони містять бібліографічні та реферативні описи наукових публікацій, що надають змогу перевірити їх достовірність [16, с. 30].

Результати перевірки достовірності та релевантності наукових публікацій у пошукової видачі платформ ШІ представлені в таблиці 1.

Результати наведені в таблиці 1 сформовані, виходячи з аналізу ТОП20 пошукової видачі наукових публікацій, отриманих з використанням платформ ШІ. Для наочного уявлення результатів проведеного дослідження на рис. 7 наведено графічну інтерпретацію порівняльного аналізу релевантності та достовірності пошукової видачі платформ ШІ, на основі отриманих результатів дослідження (див. табл. 1).

Під час опрацювання результатів пошукової видачі платформи ШІ «Grok» було встановлено що:

– формування результатів пошуку наукових публікацій виконувалось у:
- наукометричній базі Google Scholar;
- доменному просторі *.ua по типу файлів *.pdf, з використанням команди «filetype» операторного пошуку;
- наукометричній базі НБУВ;
- репозитаріях ЗВО;

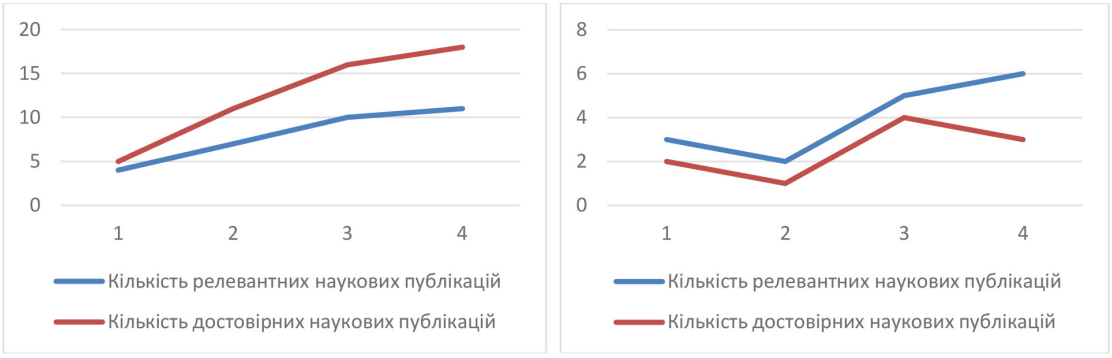
– пошук здійснювався за *.pdf файлами, тому результати пошукової видачі містять посилання на освітні програми та силабуси, у списках літератури яких присутні наукові публікації, що є релевантними пошуковому запиту;

– формування бібліографічного опису наукових публікацій не відповідає ДСТУ 8302:2015:
- у всіх позиціях пошукової видачі в описі джерел інформації використовується «Дата доступу:» замість «Режим доступу:»;

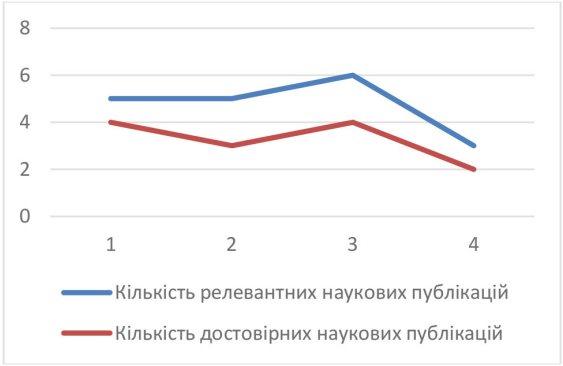
Таблиця 1

Результати перевірки пошукової видачі платформ ШІ

№	Назва платформи ШІ	Ключове слово	Результат перевірки наявності наукових публікацій в наукометричних базах даних (+/–)				Кількість релевантних наукових публікацій	Кількість достовірних наукових публікацій
			НБУВ	Google Scholar	Репозиторій ЗВО	Research Gate		
1.	ChatGPT5	складні системи	2	5	7	2	5	4
		цифрові системи	0	5	6	1	5	3
		штучний інтелект	0	3	9	1	6	4
		аналіз даних	0	1	4	0	3	2
2.	Cloude	складні системи	2	4	4	1	3	2
		цифрові системи	1	2	2	0	2	1
		штучний інтелект	2	6	5	1	5	4
		аналіз даних	2	9	8	2	6	3
3.	DeepSeek	складні системи	0	0	0	0	0	0
		цифрові системи	0	0	0	0	0	0
		штучний інтелект	0	0	0	0	0	0
		аналіз даних	0	0	0	0	0	0
4.	Gemini	складні системи	0	0	0	0	0	0
		цифрові системи	0	0	0	0	0	0
		штучний інтелект	0	0	0	0	0	0
		аналіз даних	0	0	0	0	0	0
5.	Grok	складні системи	1	1	5	0	4	1
		цифрові системи	0	1	6	1	7	4
		штучний інтелект	0	1	13	0	10	6
		аналіз даних	2	8	14	2	11	7



а) б)



с)

Рис. 7. Графічне представлення результатів аналізу релевантності і достовірності пошукової видачі платформ ШІ: а) «Grok»; б) «Cloude»; с) «ChatGPT5»

• у деяких описах джерел інформації: відсутні відомості про авторів публікації або про кількість сторінок; спостерігається скорочення повної назви наукової публікації (частина назви замінюється на «…»);
• у деяких описах наукових публікацій присутні недостовірні дані щодо року видання;
– переважна кількість наукових публікацій (до 55 %) є не існуючими.

Під час аналізу результатів пошукової видачі наукових публікацій платформи ШІ «ChatGPT5» було встановлено що:

– пошук виконується за *.pdf файлами, результати пошукової видачі містять посилання на освітні програми та силабуси, у списках літератури яких присутні наукові публікації, що релевантні пошуковому запиту;
– бібліографічний опис певної кількості наукових публікацій формується:
• без вказівки ПІБ авторів наукової публікації; повної адреси інтернет посилання (зазвичай вказується лише домен) або інтернет посилання не відповідає заявленому опису наукової публікації;
• у вигляді перекладу назви наукової публікації з англійської мови;
• із скороченням заголовку наукової публікації;
– опис наукових публікацій містить недостовірні дані, які стосуються:
• типу видання (наприклад, «конспект лекцій» замінено на «навчально-методичний посібник»);
• року видання.

Під час аналізу результатів пошукової видачі наукових публікацій платформи ШІ «Gemini» і «Deepseek» було встановлено що:

– бібліографічний опис переважної кількості наукових публікацій формується відповідно до вимог оформлення списку літератури (ДСТУ 8302:2015);
– всі позиції пошукової видачі наукових публікацій містять:
• бібліографічний опис без вказівки на інтернет посилання;
• недостовірні дані, а саме бібліографічний опис був сформований з неіснуючих наукових публікацій.

Висновки

Аналіз отриманих результатів пошукової видачі платформ ШІ ChatGPT5, Gemini, Grok, DeepSeek, Cloude довів наявність значних відмінностей, що беззаперечно ставить під сумнів достовірність певної кількісті наданих наукових публікацій.

Саме тому, важливість перевірки дослідником наукових публікацій, які надають платформи ШІ, обумовлена запобіганням:

• поширенню у науковому середовищі псевдонаукової та помилкової наукової інформації;
• посилань на неіснуючі або сфальсифіковані джерела наукової інформації;
• загальної втрати довіри до наукових досліджень.

Кожна наукова публікація дослідника має ґрунтуватися на перевірених ним на предмет існування первісних джерел наукової інформації. Це дозволить уникнути хибних інтерпретацій та висновків. Такий підхід забезпечить відповідність його наукових публікацій академічним стандартам.

Також у науковій діяльності важливо дотримуватися принципів академічної доброчесності [10; 11]. Це передбачає дотримання дослідником:

• недопущення плагіату у власних наукових дослідженнях і пов’язаними з ними науковими публікаціями;
• дотримання точності у цитуванні наукових публікацій інших авторів;
• прозорості у використанні наукових досліджень інших авторів.

Платформи ШІ можуть надавати стислий виклад наукових ідей, але лише перевірка їх відповідності джерелам інформації дозволяє гарантувати, що цитати та посилання не будуть спотвореними.

Перевірка відповідності результатів пошукової видачі наукових публікацій платформ ШІ дійсно існуючим використаним науковим джерелам, є важливою, оскільки гарантує:

• достовірність наукових результатів дослідника;
• дотримання ним вимог академічної доброчесності;
• захист від етичних та юридичних порушень.

Дослідник має усвідомлювати наявність можливих обмежень платформ ШІ при здійсненні пошуку джерел наукової інформації. Щоб зібрати максимально достовірну наукову інформацію для власного дослідження, рекомендується використовувати якомога більше спеціалізованих цифрових наукових ресурсів (наукометричних баз даних, електронних бібліотек, архівів та ін.).

При використанні платформ ШІ у науковій діяльності дослідникам рекомендовано:

• оцінювати релевантність отриманих наукових публікацій за допомогою наукометричних баз даних НБУВ, Google Scholar, репозитаріїв ЗВО, Research Gate;
• переконатися у достовірності джерела наукової інформації – обов’язково перейти за посиланням і перевірити відповідність зв’язків між контекстом цитування та посиланням на джерело інформації.
• знайти підтвердження наукової інформації в декількох авторитетних джерелах наукової інформації.

Перспективи використання платформ ШІ в науковій діяльності дослідника є надзвичайно привабливими, завдяки тому, що вони охоплюють фундаментальні та прикладні дослідження. Сучасні платформи ШІ здатні виконувати автоматизований пошук і систематизацію наукових публікацій, аналіз релевантності наукових публікацій, що суттєво оптимізує пошуковий етап досліджень.

Окремо варто відзначити потенціал платформ ШІ у міждисциплінарних дослідженнях, де вони можуть виступати засобом інтеграції знань з різних галузей та сприяти формуванню нових напрямів наукового пошуку. Вони також відкривають можливості для індивідуалізації навчання та підвищення кваліфікації дослідників, забезпечуючи доступ до аналітичних інструментів навіть у тих сферах, де раніше потрібні були спеціальні технічні навички.

Таким чином, платформи ШІ стають не лише допоміжним інструментом, а й чинником трансформації наукової діяльності, що здатен змінити методологію досліджень, прискорити отримання наукових результатів і підвищити їхню якість. Водночас, повноцінна реалізація цього потенціалу потребує уваги дослідника до питань академічної доброчесності та достовірності отриманої інформації, щоб забезпечити баланс між його інноваційною діяльністю та науковою відповідальністю.

Список використаної літератури

1. Використання штучного інтелекту в організації наукових досліджень / О. В. Полоневич, С. В. Морозова, І. М. Аверічев, А. П. Полоневич. *Зв'язок*. 2024. № 3. С. 3–6. DOI: https://doi.org/10.31673/2412-9070.2024.030306
2. Відкриті наукові практики : навч. посіб. / [С. В. Радіо та ін.]. Львів, 2023. 559 с. https://lpnu.ua/sites/default/files/2023/pages/22083/optima-wp3-handbook-redfull12012023v102.pdf
3. Дослідження та аналіз можливостей чат-боту зі штучним інтелектом ChatGPT / Г. В. Марчук, В. Л. Левківський, Д. К. Марчук, В. С. Муковоз. *Вчені записки Таврійського національного університету імені В. І. Вернадського. Серія: Технічні науки*. 2023. Т. 34 (73), №. 4. С. 71-78. DOI: https://doi.org/10.32782/2663-5941/2023.4/12
4. Ковачов С., Сичікова Я. Поговори зі мною: діалог зі штучним інтелектом про використання його в навчанні та наукових дослідженнях. *Наукові записки Бердянського державного педагогічного університету. Сер.: Педагогічні науки*. 2023. Вип. 1. С. 43–55. DOI: https://doi.org/10.31494/2412-9208-2023-1-1-43-55
5. Коломієць А. М., Кушнір О. І. Використання штучного інтелекту в освітній та науковій діяльності: можливості та виклики. *Сучасні інформаційні технології та інноваційні методики навчання в підготовці фахівців: методологія, теорія, досвід, проблеми*. 2023. Вип. 70. С. 45–57. DOI: https://doi.org/10.31652/2412-1142-2023-70-45-57
6. Критерії моніторингу достовірності інформації в інформаційному просторі / З. М. Бржевська, Н. М. Довженко, Г. І. Гайдур, А. О. Амосов. *Кібербезпека: освіта, наука, техніка*. 2019. № 1(5). С. 52–60. DOI: https://doi.org/10.28925/2663-4023.2019.5.5260
7. Національна бібліотека України імені В. І. Вернадського. URL: http://www.nbuv.gov.ua/node/1539
8. Оцінювання якості дослідницької діяльності університетів на основі принципів і підходів відкритої науки : монографія / І. Драч, О. Петроє, О. Бородієнко, Н. Приходькіна, І. Регейло, М. Саюк, О. Слободянюк. Київ, 2024. 202 с. DOI: https://doi.org/10.31874/978-617-7486-47-2-2024
9. Про обов'язковий примірник документів : Закон України № 595-XIV від 09 квіт. 1999 р. *Верховна Рада України* : офіц. вебсайт. URL: https://zakon.rada.gov.ua/laws/show/595-14#Text
10. Про освіту : Закон України № 2145-VIII від 05 верес. 2017 р. *Верховна Рада України* : офіц. вебсайт. URL: https://zakon.rada.gov.ua/laws/show/2145-19#Text
11. Про вищу освіту : Закон України № 1556-VII від 01 лип. 2014 р. *Верховна Рада України* : офіц. вебсайт. URL: https://zakon.rada.gov.ua/laws/show/1556-18#Text
12. Савєльєва О. С., Трофименко О. Г., Дикий О. В. Методологія проведення наукових досліджень у галузі інформаційних технологій : навч. посіб. [Електронне видання]. Одеса : Фенікс, 2025. 144 с. URL: http://dspace.opu.ua/jspui/handle/123456789/14858
13. Семеніхіна О., Різник В., Шамоня В. Використання штучного інтелекту в наукових дослідженнях в контексті академічної доброчесності. *Освіта. Інноватика. Практика*. 2025. Т. 13, № 6. С. 63–68. DOI: https://doi.org/10.31110/2616-650X-vol13i6-009
14. Торбас О. О. Способи використання штучного інтелекту при проведенні наукових досліджень в сфері кримінального процесу на прикладі функціоналу ChatGPT та аналізу категорії «розсуд» у кримінальному провадженні. *Правові новели*. 2023. № 19. С. 368–377. DOI: https://doi.org/10.32782/ln.2023.19.48
15. Цифрові бібліотечно-інформаційні ресурси у розбудові наукового сегменту національного інформаційного простору: монографія / [Л. А. Дубровіна, К. В. Лобузіна та ін.] ; відп. ред. К. В. Лобузіна ; НАН України, Нац. бка України імені В. І. Вернадського. Електрон. вид. Київ, 2021. 420 с. URL: http://irbis-nbuv.gov.ua/everlib/item/er-0004171
16. Шкіцька І. Ю. Аналітико-синтетичне опрацювання інформації навч. посіб. Вид. 3тє, онов. та допов. Тернопіль : ЗУНУ, 2022. 260 с.
17. ChatGPT. URL: https://openai.com/
18. Cloude. URL: https://claude.ai/login

19. Deepseek. URL: https://deepseek.com/
20. Gemini. URL: https://gemini.google.com/
21. Google Scholar. URL: https://scholar.google.com/
22. Grok. URL: https://grok.com/

References

1. Polonevych O. V., Morozova S. V., Averichev I. M., & Polonevych A. P. (2024). Vikoristannya shtuchnogo intelektu v organizaciyi naukovih doslidzhen. *Zviazok*, № 3, s. 3–6. DOI: https://doi.org/10.31673/2412-9070.2024.030306. [in Ukrainian].
2. Vidkryti naukovi praktyky : navch. posib. / [S. V. Radio ta in.]. Lviv, 559 s. https://lpnu.ua/sites/default/files/2023/pages/22083/optima-wp3-handbook-redfull12012023v102.pdf. [in Ukrainian].
3. Marchuk G. V., Levkivskyi V. L., Marchuk D. K., & Mukovoz V. S. (2023). Doslidzhennia ta analiz mozhlyvostei chat-botu zi shtuchnym intelektom ChatGPT. *Vcheni zapysky Tavriiskoho natsionalnoho universytetu imeni V. I. Vernadskoho. Seriia: Tekhnichni nauky*, 34 (73), №. 4, pp. 71-78. DOI: https://doi.org/10.32782/2663-5941/2023.4/12. [in Ukrainian].
4. Kovachov S., Sychikova Ya. Pohovory zi mnoiu: dialoh zi shtuchnym intelektom pro vykorystannia yoho v navchanni ta naukovykh doslidzhenniakh. *Naukovi zapysky Berdianskoho derzhavnoho pedahohichnoho universytetu. Ser.: Pedahohichni nauky*. 2023. 1, pp. 43–55. DOI: https://doi.org/10.31494/2412-9208-2023-1-1-43-55. [in Ukrainian].
5. Kolomiiets A. M., & Kushnir O. I. (2023). Vykorystannia shtuchnoho intelektu v osvitnii ta naukovii diialnosti: mozhlyvosti ta vyklyky. Suchasni informatsiini tekhnolohii ta innovatsiini metodyky navchannia v pidhotovtsi fakhivtsiv: metodolohiia, teoriia, dosvid, problemy, 70, pp. 45–57. DOI: https://doi.org/10.31652/2412-1142-2023-70-45-57. [in Ukrainian].
6. Brzhevska Z. M., Dovzhenko N. M., Gajdur G. I., & Amosov A. O. (2019). Kriteriyi monitoringu dostovirnosti informaciyi v informacijnomu prostori. *Kiberbezpeka: osvita, nauka, tehnika*, № 1(5), pp. 52–60. DOI: https://doi.org/10.28925/2663-4023.2019.5.5260. [in Ukrainian].
7. Nacionalna biblioteka Ukrayini imeni V. I. Vernadskogo. URL: http://www.nbuv.gov.ua/node/1539. [in Ukrainian].
8. Drach I., Petroye O., Borodiyenko O., Prihodkina N., Regejlo I., Sayuk M., & Slobodyanyuk O. (2024). Ocinyuvannya yakosti doslidnickoyi diyalnosti universitetiv na osnovi principiv i pidhodiv vidkritoyi nauki : monografiya. Kiyiv, 202 p. DOI: https://doi.org/10.31874/978-617-7486-47-2-2024. [in Ukrainian].
9. Pro obov'yazkovij primirnik dokumentiv : Zakon Ukrayini № 595-XIV vid 09 kvitnya 1999 r. Verhovna Rada Ukrayini : ofic. vebsajt. URL: https://zakon.rada.gov.ua/laws/show/595-14#Text. [in Ukrainian]
10. Pro osvitu : Zakon Ukrayini № 2145-VIII vid 05 veresnya 2017 r. Verhovna Rada Ukrayini : ofic. vebsajt. URL: https://zakon.rada.gov.ua/laws/show/2145-19#Text. [in Ukrainian].
11. Pro vishu osvitu : Zakon Ukrayini № 1556-VII vid 01 lipnya 2014 r. Verhovna Rada Ukrayini : ofic. vebsajt. URL: https://zakon.rada.gov.ua/laws/show/1556-18#Text. [in Ukrainian].
12. Savyelyeva O. S., Trofimenko O. G., & Dikij O. V. (2025). Metodologiya provedennya naukovih doslidzhen u galuzi informacijnih tehnologij : navch. posib. [Elektronne vidannya]. Odesa, 144 p. URL: http://dspace.opu.ua/jspui/handle/123456789/14858. [in Ukrainian].
13. Semenihina O., Riznik V., & Shamonya V. (2025). Vikoristannya shtuchnogo intelektu v naukovih doslidzhennyah v konteksti akademichnoyi dobrochesnosti. *Osvita. Innovatika. Praktika*, t. 13, № 6. pp. 63-68. DOI: https://doi.org/10.31110/2616-650X-vol13i6-009. [in Ukrainian].
14. Torbas O. O. (2023). Sposobi vikoristannya shtuchnogo intelektu pri provedenni naukovih doslidzhen v sferi kriminalnogo procesu na prikladi funkcionalu ChatGPT ta analizu kategoriyi "rozsud" u kriminalnomu provadzhenni. *Pravovi noveli*, № 19, pp. 368–377. DOI: https://doi.org/10.32782/ln.2023.19.48. [in Ukrainian].
15. Cifrovi bibliotechno-informacijni resursi u rozbudovi naukovogo segmentu nacionalnogo informacijnogo prostoru: monografiya. (2021). / [L. A. Dubrovina, K. V. Lobuzina ta in.] ; vidp. red. K. V. Lobuzina ; NAN Ukrayini, Nac. b-ka Ukrayini imeni V. I. Vernadskogo. Elektronnij resurs. Kiyiv, 420 p. URL: http://irbis-nbuv.gov.ua/everlib/item/er-0004171. [in Ukrainian].
16. Shkicka I. Yu. (2022). Analitiko-sintetichne opracyuvannya informaciyi : navch. posib. Vid. 3 tye, onov. ta dopov. Ternopil, 260 p. [in Ukrainian].
17. ChatGPT. URL: https://openai.com/
18. Cloude. URL: https://claude.ai/login
19. Deepseek. URL: https://deepseek.com/
20. Gemini. URL: https://gemini.google.com/
21. Google Scholar. URL: https://scholar.google.com/
22. Grok. URL: https://grok.com/

Дата першого надходження рукопису до видання: 19.09.2025
Дата прийнятого до друку рукопису після рецензування: 16.10.2025
Дата публікації:

УДК 004; 685 DOI

Б. М. ЗЛОТЕНКО
доктор технічних наук, професор,
професор кафедри комп'ютерної інженерії та електромеханіки
Київський національний університет технологій та дизайну
ORCID: 0000-0002-0870-8535

Д. В. СТАЦЕНКО
кандидат технічних наук, доцент,
завідувач кафедри комп'ютерної інженерії та електромеханіки
Київський національний університет технологій та дизайну
ORCID: 0000-0002-3064-3109

Г. В. МЕЛЬНИК
кандидат технічних наук, доцент,
доцент кафедри комп'ютерної інженерії та електромеханіки
Київський національний університет технологій та дизайну
ORCID: 0000-0002-0002-7663

В. Ю. КАЛАШНІК
кандидат технічних наук,
старший викладач кафедри комп'ютерної інженерії та електромеханіки
Київський національний університет технологій та дизайну
ORCID: 0000-0002-2345-4377

МЕТОДОЛОГІЯ ОБРОБКИ ТА ЗБЕРЕЖЕННЯ ІНФОРМАЦІЇ В КОМП'ЮТЕРНІЙ СИСТЕМІ ДЛЯ ПРОВЕДЕННЯ ДОСЛІДЖЕНЬ ПРИ ПІДГОТОВЦІ І ПРЕЗЕНТАЦІЇ КВАЛІФІКАЦІЙНОЇ РОБОТИ

Розглянуто методологію використання мікроконтролера Arduino Leonardo при виконанні наукових досліджень і висвітлення отриманих результатів при написанні статті, під час виступу на конференції, а також в процесі захисту кваліфікаційної роботи. Встановлено, що існуючі апаратні та програмні засоби, що входять до комплектів мікроконтролерів Arduino можуть бути покладені в основу створення експериментального устаткування для проведення досліджень здобувачами освіти в різних галузях науки і техніки. Архітектура плат Arduino та їх програмного забезпечення дозволяють отримувати і обробляти інформацію з різного роду датчиків у вигляді аналогових і цифрових сигналів. Встановлені залежності між вхідними і вихідними величинами дозволяють підготувати змістовну презентацію для висвітлення отриманих результатів досліджень у доповідях на наукових конференціях, семінарах, а також у презентації кваліфікаційної роботи. На основі використання мікроконтролера Arduino Leonardo і персонального комп'ютера розроблено експериментальну установку для проведення експериментальних досліджень комп'ютерно-інтегрованої системи керування освітленням розумного будинку. Розроблено алгоритм і програмне забезпечення для вимірювання та реєстрації світлового потоку при використанні фоторезистора в якості датчика і світлодіодів для імітації джерел освітлення. Запропоновано метод збереження отриманих експериментальних даних у базі даних MySQL. Розроблено методологію проведення експериментальних досліджень режимів роботи комп'ютерно-інтегрованої системи керування освітленням. Отримана експериментальна залежність вихідної з фоторезистора напруги від величини світлового потоку внутрішнього світлодіода, яка підтверджує достовірність прийнятих вихідних положень в процесі створення програмного забезпечення розробленої експериментальної установки.

Ключові слова: *методологія, експеримент, комп'ютерні науки, програмна архітектура, наукова доповідь, бази даних.*

B. M. ZLOTENKO
Doctor of Technical Sciences, Professor,
Professor at the Department of Computer Engineering and Electromechanics
Kyiv National University of Technologies and Design
ORCID: 0000-0002-0870-8535

© Злотенко Б. М., Стаценко Д. В., Мельник Г. В., Калашнік В. Ю., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

D. V. STATSENKO
PhD in Technical Sciences, Associate Professor,
Head of the Department of Computer Engineering and Electromechanics
Kyiv National University of Technologies and Design
ORCID: 0000-0002-3064-3109

H. V. MELNYK
PhD in Technical Sciences, Associate Professor,
Associate Professor at the Department of Computer Engineering
and Electromechanics
Kyiv National University of Technologies and Design
ORCID: 0000-0002-0002-7663

V. YU. KALASHNIK
PhD in Technical Sciences,
Senior Lecturer at the Department of Computer Engineering and Electromechanics
Kyiv National University of Technologies and Design
ORCID: 0000-0002-2345-4377

METHODOLOGY OF INFORMATION PROCESSING AND STORAGE IN A COMPUTER SYSTEM FOR CONDUCTING RESEARCH IN THE PREPARATION AND PRESENTATION OF QUALIFICATION WORK

The methodology for using the Arduino Leonardo microcontroller in conducting scientific research and presenting the results obtained when writing an article, during a conference speech, and in the process of defending a qualification thesis is considered. It has been established that the existing hardware and software included in the Arduino microcontroller kits can be used as the basis for creating experimental equipment for conducting research by students in various fields of science and technology. The architecture of Arduino boards and their software allow you to receive and process information from various types of sensors in the form of analogue and digital signals. The established dependencies between input and output values allow you to prepare a meaningful presentation to highlight the results of research in speeches at scientific conferences, seminars, as well as in the presentation of a qualification work. Based on the use of the Arduino Leonardo microcontroller and a personal computer, an experimental setup has been developed for conducting experimental studies of a computer-integrated lighting control system for a smart home. An algorithm and software have been developed for measuring and registering the luminous flux when using a photoresistor as a sensor and LEDs to simulate lighting sources. A method for storing the obtained experimental data in a MySQL database is proposed. A methodology for conducting experimental studies of the operation of a computer-integrated lighting control system has been developed. The experimental dependence of the voltage output from the photoresistor on the value of the luminous flux of the internal LED confirmed the reliability of the initial assumptions adopted in the process of creating the software for the developed experimental setup.

Key words: *methodology, experiment, computer science, software architecture, research report, database.*

Постановка проблеми

В даний час у зв'язку з наростаючою присутністю методів і засобів комп'ютерних наук у всіх сферах життєдіяльності людини, наявність доступних програмних і апаратних комплексів відкриває нові можливості для проведення досліджень здобувачами освіти в процесі підготовки кваліфікаційної роботи. Одним із таких засобів є комплекти, до складу яких входять контролери Arduino, плати, датчики тощо. Архітектура контролерів і відповідного програмного забезпечення дозволяє створювати експериментальне устаткування, системи отримання і опрацювання експериментальних даних.

Прикладом використання такого комплекту є проведення експериментальних досліджень при виконання кваліфікаційних робіт, присвячених розробці систем комп'ютерно-інтегрованого керування, які знайшли широке застосування в багатьох галузях науки, техніки і промисловості. До сфери комп'ютерно-інтегрованого керування увійшло в таке поняття як «розумний будинок». Система «розумний будинок» представлена комп'ютерно-інтегрованими комплексами, які здійснюють керування усіма інженерними системами, електроприладами, освітленням і температурою в приміщеннях.

На сьогоднішній день, при створенні енергоефективних приміщень, постає питання розробки та вдосконалення сучасних систем освітлення з раціональним використанням сонячного світла. Такі системи керування все ще недостатньо поширені в багатьох країнах світу, хоча вони зменшують споживання електроенергії, а також покращують тепловий та візуальний комфорт мешканців та працівників приміщень в яких вони встановлені. Ефективне керування такими системами може призвести до того, що в зимовий період буде зменшуватись кількість споживаної енергії на опалення приміщень, а в літній, відповідно, для охолодження приміщення.

Оптимальне використання природного освітлення квартири, будинку або офісу необхідно для здоров'я людей, їх комфорту, а також економії енергоресурсів. Тому в основу проектування системи «Розумний будинок було покладено ідею регулювання інтенсивності природнього світлового потоку за рахунок автоматичного управління жалюзі. При настанні темного періоду доби автоматичне закриття вікона за допомогою жалюзі і вмикання штучного освітлення забезпечить необхідний рівень комфорту мешканцям розумного будинку.

Аналіз останніх досліджень і публікацій

Для реалізації дистанційного керування освітленням необхідні апаратно-програмні засоби з метою обробки сигналів, які надходять з датчиків, і передачі керуючих сигналів на виконавчі пристрої. В якості таких апаратно-програмних засобів розглянемо продукти розробників Arduino Software і Raspberry Pi Foundation.

Плати Arduino стали одними з найпопулярніших мікроконтролерів на ринку з величезною різноманітністю плат [1]. Кожна з цих плат Arduino має різну кількість контактів введення та виведення [2, 3]. Плата з найбільшою кількістю контактів – Mega 2560 поставляється з 54 цифровими контактами введення та виведення (з них 15 з них мають PWM) і має 16 вхідних аналогових контактів. Плата Micro має 20 цифрових входів / виходів (з 7 з них з PWM) і 12 вхідних аналогових контактів. Плата Uno має 14 цифрових входів та виходів (з 6 з них з PWM) і 6 вхідних аналогових контактів [4]. Arduino Leonardo – це пристрій на базі мікроконтролера ATmega 32U4. До складу пристрою входить: 20 входів та виходів, кварцовий резонатор на 16 МГц, роз'єм micro-USB, роз'єм живлення, роз'єм для програмування ICSP і кнопка скидання. Обсяг вільної пам'яті для зберігання програми в мікроконтролері ATmega 32U4 дорівнює 32 КБ, 4 КБ виділяється під потреби завантажувача. Об'єм оперативної пам'яті типу SRAM дорівнює 2,5 КБ. Крім цього, мікроконтролер має 1 КБ EEPROM. В мікроконтролер ATmega 32U4 вбудований USB-контролер, що виключає необхідність в додатковому процесорі [5, 6].

Для впровадження системи управління освітленням можна також використати обладнання, яке включає пристрій відеоспостереження, інфрачервоний датчик руху, пристрій зберігання і збору інформації, пристрій обробки інформації, датчик освітленості, пристрій бездротової передачі даних. Основою даної системи є одноплатний комп'ютер Raspberry Pi. Raspberry Pi 3 Model – це повністю функціональний комп'ютер з власним виділеним процесором, пам'яттю і може запускати операційну систему (працює на Linux) і вбудований Wi-Fi [7]. На платі встановлений Ethernet-порт, який забезпечує доступ до мережі.

Отже, якщо проєкт, включає в себе кілька функціональних можливостей одночасно, потрібен легкий доступ до Інтернету і потрібні можливості для доступу до мультимедіа, то потрібно обирати Raspberry Pi. У випадку, коли проєкт вимагає легкого зчитування з датчиків, який повинен виконувати тільки кілька вихідних сигналів на основі даних датчика, легко обмінюватися даними і швидко запускатися, тоді переважним є вибір Arduino.

Формулювання мети дослідження

Метою роботи є: є оптимізація структури комп'ютерно-інтегрованої системи керування освітленням розумного будинку на основі використання контролера Arduino Leonardo.

Викладення основного матеріалу дослідження

Оскільки основним завданням даної плати є взаємодія з датчиками і пристроями, Arduino підходить для апаратних проєктів, де потрібно чітко і швидко реагувати на різні сигнали. В даному випадку комп'ютерно-інтегрована система автоматизації освітлення розумного будинку є саме таким проектом. Контролер Arduino Leonardo розрахований на роботу з великою кількістю різноманітних давачів.

З метою проведення досліджень режимів роботи комп'ютерно-інтегрованої системи освітлення розумного будинку було розроблено і виготовлено експериментальну установку. Модель експериментальної установки спроектована за допомогою середовища Fritzing (рис. 1).

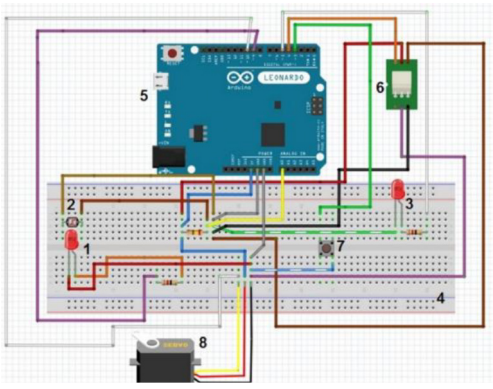


Рис. 1. Модель експериментальної установки

Лабораторна установка містить внутрішній світлодіод 1, що моделює зовнішній потік природнього освітлення для дії на чутливий елемент фоторезистора 2, зовнішній світлодіод 3, макетну плату mini Breadboard 4, контролер Arduino Leonardo 5, модуль реле 6, тактовий перемикач 7, який вмикає та вимикає внутрішній світлодіод 1, а такж сервопривод 8, що використовується для здійснення повороту приводного вала жалюзі у відповідності із вихідною напругою фоторезистора.

Для програмування контролера Arduino Leonrado використане середовище Arduino IDE. Розроблена програма завантажується у контролер також за допомогою даного середовища (рис. 2). Мова Wiring середовища Arduino IDE розроблена на основі мов програмуваня C/C++. Створювані у середовищі Arduino IDE програми мають назву скетчів. При натисканні кнопки завантаження розробленого скетчу в контролер Arduino Leonrado, написаний код транслюється в мову програмування С. Далі компілятор здійснює остаточну трансляцію коду на мову, яка зрозуміла для мікроконтролера.

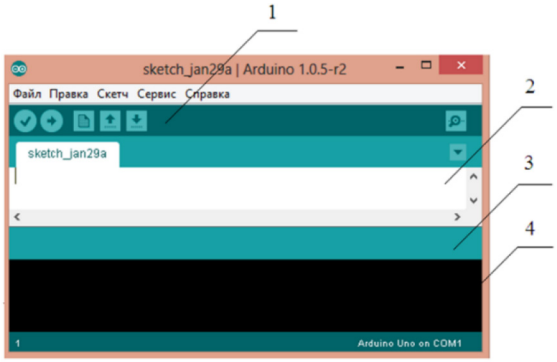


Рис. 2. Вигляд вікна середовища Arduino IDE

Розробка скетчів відбувається за допомогою текстового редактора. До складу вбудованого текстового редактора 2 входить набір стандартних інструментів таких, як зміна тексту, пошук, вставка, копіювання. Роль зворотного зв'язку з користувачем виконує область з повідомленнями 3, яка розташована всередині вікна програми. Вона здійснює інформування щодо подій, а також помилок, що можуть виникати при записі або експорті коду, написаного мовою Wiring. Відображення потоку вихідних даних в середовищі Arduino IDE, разом з усіма повідомленнями щодо помилок здійснюється за допомогою консолі 4. Модель контролера, що використовується, разом з послідовним портом, до якого він підключений відображаються у ничній правій частині вікна. Створення, збереження, відкриття і завантаження програми в контролер здійснюється за допомогою кнопок, розміщених в панелі інструментів 1 (рис. 3).

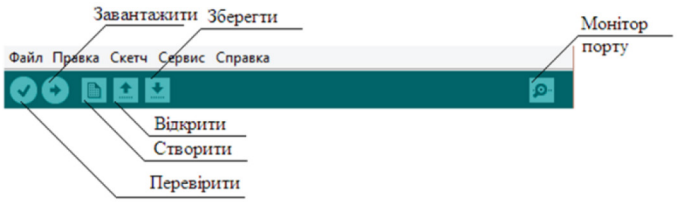


Рис. 3. Вигляд панелі інструментів середовища Arduino IDE

Перевірка щодо помилок, які містить програмний код, здійснюється за допомогою кнопки «Перевірити». Створення нового скетчу здійснюється за допомогою кнопки «Створити». Усі скетчі, розміщені у блокноті, можуть бути здійсненні відкриттям меню за допомогою кнопки «Відкрити». Компіляція програмного коду і завантаження його у контролер Arduino здійснюється за допомогою кнопки «Завантажити». Для того, щоб зберегти скетч використовується кнопка «Зберегти». Моніторинг послідовної шини здійснюється відкриттям за допомогою кнопки «Монітор порту». За допомогою програми «Монітор» відображаються дані, для передачі яких від контролера Arduino використовується USB послідовний інтерфейс.

Для відправки даних контролера Arduino у вікно програми вводиться текст і натискається кнопка Відправити або клавіша Enter. У списку, який відкривається, вибирається швидкість, з якою дані передаються, і яка відповідає швидкості, вказаній для функції Serial.begin(), що міститься у розробленому скетчі.

У спрощеному вигляді процес програмування в середовищі Arduino IDE включає в себе наступні стадії: Підключення плати до USB порту комп'ютера. Написання скетчу. Завантаження скетчу в контролер Arduino за допомогою USB кабеля. Витримка паузи. Перезапуск контролера Arduino Виконання контролером Arduino завантаженого скетчу.

Після того, як контролер Arduino Leonardo під'єднується до комп'ютера за допомогою USB кабелю, на ньому відбувається спалахування світлодіода і починається блимання світлодіода. Таким чином, при подачі живлення починається виконання завантаженої заводської програми програма Blink.

В розроблюваній комп'ютерно-інтегрованій системі освітлення у якості давача величини світлового потоку від внутрішнього світлодіода використано фоторезистор GL 55. Опір фоторезистора визначається яскравістю освітлення його керамічного чутливого елемента у вигляді змійовика. Дані про величину світлового потоку будуть зчитуватись у вигляді даних фоторезистора – як вихідна напруга під дією освітлення. З цією метою фоторезистор під'єднується до платформи Arduino Leonardo за допомогою дільника напруги, показаного на рис. 4.

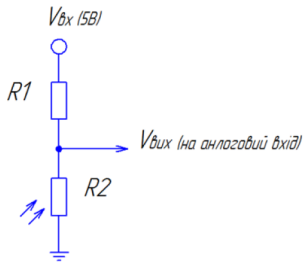


Рис. 4. Дільник напруги з під'єднаним фоторезистом

У схемі дільника напруги використовується резистор R1 з постійним опором, який вибирається у відповідності з бажаним діапазоном напруги на виході. Резистор R2 є фоторозистором і виступає в якості давача світлового потоку, який надходить від імітуючого денне освітлення внутрішнього світлодіода. Опір фоторезистора змінюється в залежності від інтенсивності світлового потоку, який потрапляє на його чутливий елемент, в результаті чого величина вихідної напруги змінюється. ЇЇ величина залежить від опору фоторезистора і визначається за допомогою закона Ома з виразу:

$$V_{вих} = I \cdot R_2 = \frac{R_2 \cdot V_{вх}}{R_1 + R_2}, \quad (1)$$

де I – величина сили струму, що проходить через фоторезистор R_2 і резистор R_1.

Величина сили струму, що проходить через фоторезистор R_2 і резистор R_1 визначається з виразу (1) за формулою:

$$I = \frac{V_{вх}}{R_1 + R_2}. \quad (2)$$

З виразу (1) витікає, що із збільшенням опору R_2 по відношенню до опору R_1 падіння напруги на фоторезисторі збільшується.

Серед можливостей контролера Arduino є здатність вимірювання напруги, яка визначається аналоговими давачами. З цією метою використовується analogRead функція. У випадку, коли давач підєднується до виводу Arduino A0, для перетворення величини сигналу, отриманого за допомогою давача, у змінну необхідне використання команди int value=analogRead(A0).

Отримані дані можна представити візуально на моніторі. Для цього монітор з'єднується з послідовним портом шляхом використання команди Serial.begin(rate). При цьому здійснюється відкриття послідовного порту і встановлюється швидкість, з якою передаються дані.

З метою відображення даних, які надходять з контролера Arduino на персональний комп'ютер за допомогою послідовного інтерфейсу USB, використано програму Монітор.

Таким чином, дані, які зчитуються, виводяться на екран монітора за допомогою команди Serial.begin(rate) шляхом передачі даних через послідовний порт. При цьому відбувається автоматичне повернення каретки і перехід до нового рядка.

Для читання даних необхідно після завантаження програми в контролер Arduino натиснути на кнопку Монітор порту, яка знаходиться в правій верхній частині вікна програми. З метою моделювання зміни інтенсивності світлового потоку природнього освітлення у лабораторній установці здійснюється зміна яскравості випромінювання світлодіода, яка при проведенні досліду збільшується в межах 0 %…100 % поступово.

З метою збереження даних з мікроконтролера Arduino для подальшого аналізу отриманих результатів, запропоновано використати базу даних (БД) MySQL. Скетч містить набір команд, який надсилає зчитані дані на комп'ютер через послідовний порт. В базі даних створена таблиця для зберігання даних, яка містить три колонки: 1) id: Унікальний ідентифікатор кожного запису; 2) sensor_value: Ціле число для збереження даних з фоторезистора; 3) timestamp: Дата і час запису, яка заповнюється автоматично. Скрипт написаний мовою Python зчитує дані

з послідовного порту і вставляє в таблицю БД. Принцип роботи збереження отриманих експериментальних даних у БД наступний: 1) Скрипт спочатку намагається підключитися до послідовного порту та бази даних MySQL. 2) У нескінченному циклі він постійно перевіряє, чи є нові дані в послідовному буфері. 3) Якщо дані є, він зчитує рядок, декодує його, перетворює на ціле число. 4) Сформований SQL-запит вставляє отримане значення в базу даних. 5) Команда mydb.commit() підтверджує зміни і зберігає їх у таблиці. 6) Скрипт продовжує працювати, доки його не зупинить користувач.

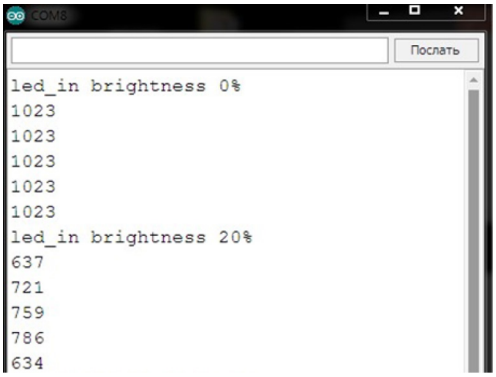


Рис. 5. Представлення вихідної напруги фоторезистора

Оскільки вхідна напруга контролера Arduino може коливатися в діапазоні 0…5 В, то програмою здійснюється перетворення цих даних у цілочисельні значення в діапазоні 0…1023. Крім того, спостерігається коливання виведених значень навколо одного числа, а саме 150, під час випромінювання світлодіодом світла, яке складає 20 % від максимальної яскравості led_in brightness20%. У зв'язку з цим, при кожному рівні яскравості зчитуються 6 виведених значень з подальшим визначенням середньоарифметичного для кожного рівня яскравості, яке буде являти собою вхідну напругу у вигляді:

$$\overline{V}_{out(цифр.)} = \frac{\sum_{i=0}^{n} V_{out(i)}}{n}, , \tag{3}$$

де $V_{out(i)}$ – значення вхідної напруги для даного рівня яскравості; $n = 6$ – кількість значень, виведених для даного рівня яскравості.

Переведення отриманих даний у цифровому форматі до аналогових значень вихідної з фоторезистора напруги здійснювалось за формулою:

$$\overline{V}_{out(аналог.)} = \frac{\overline{V}_{out(цифр.)} \cdot 5}{1024}. \tag{4}$$

За описаною методологією були проведені експериментальні дослідження, в результаті яких отримана графічна залежність вихідної з фоторезистора напруги від яскравості внутрішнього світлодіода наведена на рис. 5.

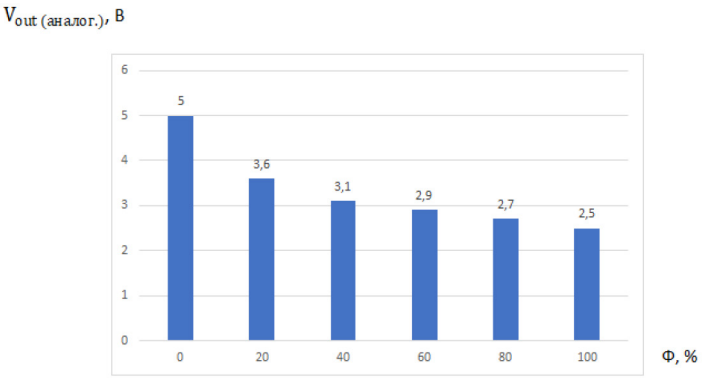


Рис. 6. Представлення вихідної напруги фоторезистора

Як видно з рис. 6, при відсутності світлового потоку від внутрішнього світлодіода до фоторезистора величина вихідної з фоторезистора напруги приймає максимальне значення 5 В. При поступовому збільшенні яскравості внутрішнього світлодіода з 0 % до 40 % світлового потоку величина вихідної з фоторезистора напруги

зменшується. Подальше збільшення яскравості внутрішнього світлодіода вихідна з фоторезистора напруга монотонно зменшується, хоча з меншою інтенсивністю. Найбільшій 100 % яскравості світлодіода відповідає найменше значення вихідної напруги фоторезистора 2,5 В.

Отримані в ході виконання результати експериментального дослідження можуть бути представлені при публікації наукової статті, під час виступу на науковій конференції або семінарі, а також в процесі презентації кваліфікаційної роботи.

Висновки

Встановлено, що при проведенні досліджень в різних галузях науки і техніки доцільно використовувати наявні програмні і апаратні засоби, архітектура яких дозволяє інтегрувати їх до складу експериментального устаткування. Це відкриває перед здобувачами освіти широкі можливості для створення матеріальної бази досліджень в рамках виконання кваліфікаційної роботи і оприлюднення отриманих результатів.

Розроблено комп'ютерно-інтегровану схему керування освітленням розумного будинку з оптимальною структурою, яка дозволяє використовувати її в якості лабораторної установки. Лабораторна установка дозволяє проводити дослідження режимів роботи системи в режимі імітації змін світлого та темного періодів доби. При цьому протягом світлої доби використовується природнє освітлення за рахунок автоматичного відкриття жалюзі, а протягом темного періоду вмикається автоматично штучне освітлення і жалюзі закриваються.

В результаті проведеного експериментального дослідження підтверджено працездатність розробленої комп'ютерно-інтегрованої системи освітлення розумного будинку. Отримана експериментальна залежність вихідної з фоторезистора напруги від яскравості внутрішнього світлодіода підтверджує достовірність прийнятих вихідних положень в процесі створення програмного забезпечення розробленої системи освітлення.

Встановлено, що для забезпечення функціонування комп'ютерно-інтегрованої системи керування освітленням розумного будинку потрібно вибрати її основний компонент – керуючий мікроконтролер. Однією з найбільш доступних сучасних керуючих платформ, до складу яких входить мікроконтролер, є платформа Arduino. Проте, навіть одна ця платформа має багато модифікацій, які використовують різні типи мікроконтролерів, що призводить до їх різної функціональності, зовнішніх інтерфейсів, об'єму пам'яті та габаритних розмірів.

Список використаної літератури

1. Monk S. Programming Arduino next steps: going further with sketches. 2nd ed. McGraw-Hill Education: Columbus, OH 43218, USA. 2019. 320 p. https://agsci.colostate.edu/wp-content/uploads/sites/95/2020/03/Programming-Arduino.pdf

2. Geddes M. Arduino project handbook: 25 practical projects to get you started: San Francisco: No Starch Pre ss, 2016. 275 p. https://ss-valpovo.hr/wp-content/uploads/2020/01/arduinoprojecthandbook.pdf

3. Subekti S, Pranoto H, Rukasah Salmon B, Qomarudin Yusuf S, Suyadiyanto S, et al. (2020) Preventive maintenance of taper bearing using Arduino in the application of industry 4.0. International Research Journal of Engineering, IT & Scientific Research 6(4): 1–14. https://sloap.org/journals/index.php/irjeis/article/view/953/1466

4. Gheorghe AC, Stoica CI (2021) Wireless Weather Station Using Arduino Mega and Arduino Nano. Scientific Bulletin of Electrical Engineering Faculty 21(1): 35–38. file:///C:/Users/Admin/Downloads/Wireless-Weather-Station-Using-Arduino-Mega-and-Arduino-Nano.pdf

5. Ibraheem R, Ahmad F. The Versatile World of Arduino Boards: A Comprehensive Review. JOJ Material Sci. 2024; 9(1): 555753. DOI: 10.19080/JOJMS.2024.09.555753

6. Margolis M. Arduino Cookbook. Published by O'Reilly Media, 2011. – 632 p. https://juniorfall.wordpress.com/wp-content/uploads/2011/11/arduino-cookbook.pdf

7. Ivković J, Radulovi B. The Advantages of Using Raspberry Pi 3 Compared to Raspberry Pi 2 SoC Computers for Sensor System Support. International Conference on Applied Internet and Information Technologies, 2016. 88–94. https://www.researchgate.net/publication/304918737_The_Advantages_of_Using_Raspberry_Pi_3_Compared_to_Raspberry_Pi_2_SoC_Computers_for_Sensor_System_Support

References

1. Monk S. Programming Arduino next steps: going further with sketches. 2nd ed. McGraw-Hill Education: Columbus, OH 43218, USA. 2019. 320 p. https://agsci.colostate.edu/wp-content/uploads/sites/95/2020/03/Programming-Arduino.pdf

2. Geddes M. Arduino project handbook: 25 practical projects to get you started: San Francisco: No Starch Pre ss, 2016. 275 p. https://ss-valpovo.hr/wp-content/uploads/2020/01/arduinoprojecthandbook.pdf

3. Subekti S, Pranoto H, Rukasah Salmon B, Qomarudin Yusuf S, Suyadiyanto S, et al. (2020) Preventive maintenance of taper bearing using Arduino in the application of industry 4.0. International Research Journal of Engineering, IT & Scientific Research 6(4): 1-14. https://sloap.org/journals/index.php/irjeis/article/view/953/1466

4. Gheorghe AC, Stoica CI (2021) Wireless Weather Station Using Arduino Mega and Arduino Nano. Scientific Bulletin of Electrical Engineering Faculty 21(1): 35–38. file:///C:/Users/Admin/Downloads/Wireless-Weather-Station-Using-Arduino-Mega-and-Arduino-Nano.pdf

5. Ibraheem R, Ahmad F. The Versatile World of Arduino Boards: A Comprehensive Review. JOJ Material Sci. 2024; 9(1): 555753. DOI: 10.19080/JOJMS.2024.09.555753

6. Margolis M. Arduino Cookbook. Published by O'Reilly Media, 2011. – 632 p. https://juniorfall.wordpress.com/wp-content/uploads/2011/11/arduino-cookbook.pdf

7. Ivković J, Radulovi B. The Advantages of Using Raspberry Pi 3 Compared to Raspberry Pi 2 SoC Computers for Sensor System Support. International Conference on Applied Internet and Information Technologies, 2016. 88–94. https://www.researchgate.net/publication/304918737_The_Advantages_of_Using_Raspberry_Pi_3_Compared_to_Raspberry_Pi_2_SoC_Computers_for_Sensor_System_Support

Дата першого надходження рукопису до видання: 24.09.2025
Дата прийнятого до друку рукопису після рецензування: 20.10.2025
Дата публікації:

UDC 004.9:004.4 DOI

M. O. ILIN
Postgraduate Student at the Computer Systems Software Department
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"
ORCID: 0009-0001-0803-3726

L. M. OLESHCHENKO
Candidate of Technical Sciences,
Associate Professor at the Computer Systems Software Department
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"
ORCID: 0000-0001-9908-7422

DEVELOPMENT OF A MODULAR DATA UNIFICATION PIPELINE FOR REAL-TIME ENVIRONMENTAL THREAT ANALYTICS

The article presents the results of research and development of a modular software system for real-time environmental risk analytics, built on the concept of unified event ingestion and the application of H3-based hexagonal spatial aggregation. The relevance of this work is driven by the high heterogeneity of primary data sources, including open analytical reports (OSINT), satellite fire detections (FRP), ionizing radiation dose rate measurements, and meteorological fields. These sources rely on different temporal scales, measurement units, and data schemas. Such diversity creates significant challenges in integration, resulting in timestamp shifts, unit inconsistencies, schema conflicts, and event duplication, which complicates the timely production of consistent results and slows down updates of information layers. The objective of the study is to design a compact and reproducible data processing pipeline capable of transforming heterogeneous event streams into daily aggregated layers of risk and demand with sub-second latency. The proposed architecture consists of three main components. The first is the Plugin Source Adapter Interface (PSAI), which maps each data source into a standardized event table; adapters are responsible only for source-specific parsing, while subsequent unification logic is shared. The second component is the Deterministic Event Harmonization (DEH) module, which converts all timestamps to UTC format, normalizes measurement units, verifies coordinates, and guarantees idempotent inserts. This ensures safe reprocessing and proper handling of late-arriving data. The third component is a query layer oriented towards H3-first spatio-temporal aggregation: FRP points are aggregated into 15-minute intervals at H3 resolution level 10 using the MAX(FRP) operator to avoid overlaps, after which daily summaries at H3 resolution level 7 integrate FRP data with OSINT events, radiation monitoring signals, and meteorological indicators, including a wind alignment proxy, to construct a comprehensive risk index.

The system is implemented in the DuckDB environment without the use of dedicated servers or network services, which simplifies infrastructure and reduces operational costs. The adopted approach ensures transparency and testability of integration contracts: the PSAI interface enforces data ingestion rules, the DEH module standardizes time and units, the H3 hierarchy guarantees spatial consistency, and DuckDB provides materialization of results through standard SQL. Experimental evaluations confirmed the high efficiency of the proposed solution: data updates are performed within fractions of a second even on consumer-grade laptop hardware; duplication of satellite FRP detections is reduced by approximately 99 %; daily summary exports remain under one megabyte in size, significantly simplifying their transfer and storage. The developed system combines the flexibility of modular architecture, the reproducibility of integration procedures, and the efficiency of the computational model, establishing a foundation for practical applications in environmental monitoring, risk management decision support, and the advancement of ecological analytics services.

Key words: *software system, real-time analytics, data unification, H3, DuckDB, materialized views, environmental monitoring.*

М. О. ІЛЬЇН
аспірант кафедри програмного забезпечення комп'ютерних систем
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
ORCID: 0009-0001-0803-3726

© Ilin M. O., Oleshchenko L. M., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

Л. М. ОЛЕЩЕНКО
кандидат технічних наук,
доцент кафедри програмного забезпечення комп'ютерних систем
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
ORCID: 0000-0001-9908-7422

РОЗРОБКА МОДУЛЬНОГО КОНВЕЄРА УНІФІКАЦІЇ ДАНИХ ДЛЯ АНАЛІТИКИ ЕКОЛОГІЧНИХ ЗАГРОЗ У РЕАЛЬНОМУ ЧАСІ

У статті представлено результати дослідження та розробки модульної програмної системи для аналітики екологічних ризиків у реальному часі, що ґрунтується на концепції уніфікованого приймання подій та застосуванні шестикутної просторової агрегації H3. Актуальність роботи зумовлена високим рівнем гетерогенності первинних джерел даних, серед яких: відкриті аналітичні звіти (OSINT), супутникові детекції пожеж (FRP), вимірювання потужності дози іонізуючого випромінювання та метеорологічні поля. Зазначені джерела оперують різними часовими шкалами, одиницями вимірювання та схемами подання даних. Це призводить до виникнення труднощів у процесі їх інтеграції, зокрема, до зміщення часових міток, невідповідності одиниць, конфліктів структур та дублювання подій, що ускладнює оперативне отримання узгоджених результатів і уповільнює оновлення інформаційних шарів. Метою дослідження визначено побудову компактного та відтворюваного програмного конвеєра для обробки даних, здатного перетворювати змішані потоки подій у добові агреговані шари ризику та попиту із затримкою на рівні секунд. Запропонована архітектура включає три основні компоненти. Першим є інтерфейс плагінів-адаптерів джерел (PSAI), що забезпечує відображення кожного джерела у стандартизовану таблицю подій; адаптери відповідають лише за специфічний розбір вхідних даних, тоді як подальша логіка уніфікації є спільною. Другим компонентом виступає модуль детермінованої уніфікації подій (DEH), який переводить усі часові позначки у формат UTC, нормалізує одиниці вимірювання, верифікує координати та гарантує ідемпотентні вставки, що дозволяє безпечно здійснювати повторні прогони та коректно враховувати запізнілі дані. Третій компонент – запитний шар, орієнтований на застосування просторово-часової агрегації за принципом H3-first: точки FRP агрегуються у 15-хвилинні інтервали на рівні H3 r10 із використанням оператора MAX(FRP) для уникнення накладань, після чого добові зведення на рівні H3 r7 інтегрують дані FRP з подіями OSINT, сигналами радіаційного моніторингу та метеопоказниками, зокрема, проксі-індикатором напрямку та швидкості вітру, для формування комплексного індексу ризику. Реалізація системи здійснена у середовищі DuckDB без залучення окремих серверів чи мережевих сервісів, що спрощує інфраструктуру та знижує вартість експлуатації. Використаний підхід забезпечує явність та тестованість інтеграційних контрактів: інтерфейс PSAI фіксує правила приймання даних, модуль DEH – стандартизацію часу та одиниць вимірювання, ієрархія H3 – просторову узгодженість, а DuckDB – можливість матеріалізації результатів засобами стандартного SQL. Проведені експерименти підтвердили високу ефективність запропонованого рішення: оновлення даних виконуються за долі секунди навіть на потужності обладнання рівня ноутбука; дублювання супутникових FRP-детекцій скорочується приблизно на 99 %; добові експорти зведень мають обсяг менше одного мегабайта, що істотно спрощує їх передачу та зберігання. Таким чином, розроблена система поєднує гнучкість модульної архітектури, відтворюваність інтеграційних процедур та ефективність обчислювальної моделі, що створює підґрунтя для практичного використання у сфері моніторингу довкілля, підтримки рішень з управління ризиками та розвитку сервісів екологічної аналітики.

Ключові слова: *програмна система, аналітика в реальному часі, уніфікація даних, H3, DuckDB, матеріалізовані подання, моніторинг довкілля.*

Problem statement

The relevance of the research lies in the growing global demand for timely and reliable environmental monitoring tools capable of supporting risk management and decision-making processes. Modern environmental data originate from highly heterogeneous sources – ranging from satellite fire detections and radiation measurements to meteorological fields and open-source intelligence reports – that differ in temporal resolution, measurement units, and data schemas. This heterogeneity complicates integration, leading to inconsistencies, delays, and information overload that hinder the effectiveness of analytical software systems. A modular and reproducible data unification pipeline, capable of harmonizing diverse event streams into coherent, compact, and near-real-time risk layers, addresses these challenges directly. Such an approach not only improves data quality and processing speed but also enhances the scalability and accessibility of environmental analytics, thereby providing a robust technological foundation for sustainable development, emergency response, and ecological risk assessment.

Heterogeneous streams – textual incident reports, satellite fire detections, station radiation, and weather – are difficult to align. Time zones and units drift; schemas diverge. The objective is a lightweight, reproducible path from raw feeds to daily risk and demand layers with seconds-level latency. The approach combines a modular ingest interface, deterministic homogenization, and H3-based materialized views implemented in DuckDB.

Concretely, source adapters emit rows in a unified event schema. The DEH stage converts all timestamps to UTC-naive time, harmonizes units (dose rates in sieverts, wind in m/s), validates coordinates, and performs idempotent upserts using a natural key. For FRP, we bin points into 15-minute micro-tiles at H3 r10 and take MAX (FRP) per bin/cell to collapse overlapping satellite detections. Daily rollups at H3 r7 then join FRP with OSINT hits, radiation signals, and a wind-alignment proxy. Materialization is expressed in SQL and executed inside DuckDB, eliminating external orchestration while keeping refresh logic transparent and testable.

Related research

Prior work on real-time environmental analytics clusters into six lines: compute engines, spatial indexing, fire products, radiation networks, meteorological proxies, and access/transport. We outline each and note trade-offs.

Studies on embedded columnar databases show that single-node, vectorized SQL can meet interactive workloads without a separate server. DuckDB typifies this class and reports high scan and aggregation rates on commodity hardware [1]. This line reduces latency and operations, but it shifts responsibility to schema design and careful memory use.

Multiscale summarization commonly uses hierarchical tessellations. H3 offers hexagonal cells with stable parent-child relations and neighborhood structure, enabling consistent rollups and topology-aware joins across resolutions [2]. Hexagons limit orientation bias and support regional statistics, yet cell area still varies with latitude and must be considered in analysis.

Active-fire research relies on satellite detections and fire radiative power. FIRMS distributes MODIS/VIIRS data that many authors composite in space and time to handle overlapping overpasses and repeated hits before daily or regional summaries [3, 4]. Typical choices include temporal binning and cell-level maxima or sums; these increase robustness but can suppress short-lived transients.

Environmental dose monitoring aggregates heterogeneous stations. Platforms such as EURDEP and community networks face mixed units (nSv/h, μR/h, CPM), uneven cadences, and diverse device quality. Published workflows emphasize unit harmonization, timezone normalization, and basic outlier screening prior to regional synthesis [5, 6]. Harmonization improves comparability, but device-specific calibration remains an open issue.

When full dispersion models are infeasible, open weather APIs supply wind direction and speed as lightweight surrogates for transport and exposure. Prior studies combine directionality with source locations to approximate downwind influence or to weight risk indicators at grid scale [7]. Such proxies scale well, though they cannot replace plume physics in complex terrain or stability regimes.

Response and demand depend on movement over networks. Open street-graph tooling (OSMnx) and open routing engines (Valhalla) enable isochrones and travel-time metrics on public data [8, 9]. This body of work supports “reachable population” and logistics-aware assessments without proprietary datasets, at the cost of careful preprocessing and profile tuning.

The literature indicates that embedded OLAP can satisfy sub-second analytics [1]; hierarchical hex grids provide stable spatial rollups [2]; satellite and sensor products require deterministic preprocessing before aggregation [3-6]; and open routing stacks make travel-time analysis tractable on public data [8,9]. These strands define the current toolkit for building compact, reproducible systems in this domain.

Proposed modular data unification pipeline

We integrate four streams that stress different parts of the pipeline yet fit a single long-format events table: *timestamp_utc* (naive UTC), *lat, lon, metric, value, unit, source*, and *extra_json* (for provenance and *device/product* fields).

OSINT messages arrive as text with local timestamps and geocoded coordinates; FIRMS provides numeric FRP detections in UTC; radiation networks mix dose-rate units and CPM from heterogeneous stations; weather delivers gridded wind direction and speed. The PSAI adapters map each native record to the schema and pass a *tz_hint* where needed so DEH can standardize time and units deterministically (Table 1).

Table 1

Mapping of source-specific fields to the unified events schema

Source	Timestamp (orig)	Lat/Lon (orig)	Metric	Value (unit)	Normalized events schema
OSINT	ISO 8601 (local)	Geocoded text	hit_reported	1 (count)	timestamp_utc, lat, lon, metric=’hit_reported’, value=1, unit=’count’, source=’osint’, extra_json={‘msg_id’, ‘channel’}
FIRMS (MODIS/VIIRS)	Date + time (UTC)	latitude/longitude	frp_*	FRP (MW)	timestamp_utc, lat, lon, metric=’frp_*’, value=FRP, unit=’MW’, source=’firms’, extra_json={‘product’, ‘confidence’}
Radiation (γ)	ISO 8601 (UTC)	station coords	gamma,gamma_cpm	nSv/h or CPM	timestamp_utc, lat, lon, metric, value, unit, source=’radiation’, extra_json={‘station_id’, ‘device’}
Weather (Wind)	ISO 8601 (UTC)	grid cell center	wind_dir,speed	deg or $m \cdot s^{-1}$	timestamp_utc, lat, lon, metric, value, unit, source=’weather’, extra_json={‘provider’}

Each row in the events table captures a single measurement or event. For instance, an OSINT "*hit_reported*" event is stored with value 1 and unit count, whereas a satellite fire detection includes an FRP value in megawatts.

The radiation feed shows how we normalize units: if a station reports 20 μR/h (microRoentgens per hour), the adapter converts it to approximately 0.175 μSv/h (microsieverts per hour) for consistent gamma dose units. Likewise, wind is split into direction and speed metrics. This uniform schema keeps joins predictable and enables vectorized scans and aggregations inside DuckDB.

Our pipeline's architecture follows a modular ingest and homogenization design. The Plugin Source Adapter Interface (PSAI) allows each data source to implement a small adapter that pulls or receives data and emits normalized event rows. For example, one adapter might call a REST API or read a file and then yield events via a common interface (Python generator of dicts). Each adapter handles source-specific parsing: decoding timestamps, parsing coordinates, and extracting the relevant metrics. This modular approach means we can hot-swap new data sources without altering the core pipeline – just plug in a new adapter that conforms to the interface.

All adapters feed into a central Deterministic Event Homogenization (DEH) stage. The DEH applies consistent rules to every incoming event record.

1. UTC conversion. All timestamps are converted to naive UTC datetime objects. If an event timestamp lacks timezone info, a source-specific hint (provided by the adapter) or default UTC is applied.

2. Unit harmonization. Values are converted to a base unit when necessary (e.g., radiation counts per minute might be converted to dose rates, different radiation units to Sieverts, etc.). This ensures metric and unit pairs are consistent (no mix of "μR/h" vs "nSv/h" in final data).

3. Geo validation. We ensure latitude and longitude are present and within valid ranges. An event without valid coordinates may be dropped or flagged, unless the metric inherently has no location (in our case, we mostly handle spatial data, except some aggregate metrics).

4. Idempotent upsert. Using a natural key (combining source, metric, time, and location), the pipeline checks if an event is already in the database. This prevents duplicate inserts when a source re-sends the same data or when our adapters restart. It is crucial for stable aggregates downstream.

After homogenization, events are inserted into a DuckDB database table (events) and a live view (events_live) that may hold the latest window of data. This design avoids network overhead and leverages DuckDB's vectorized execution for fast in-memory analytics. By embedding DuckDB, we get high performance on analytical queries in a small footprint, similar to how SQLite serves OLTP but here for OLAP workloads (Fig. 1).

On the query side, we implement a lightweight spatio-temporal Command Query Responsibility Segregation (ST-CQRS) pattern. The Command side (ingest) continuously appends events and updates intermediate aggregates, while the Query side reads from pre-computed materialized views for fast responses. Rather than heavy external stream processing frameworks, we leverage DuckDB's ability to maintain materialized tables via SQL.

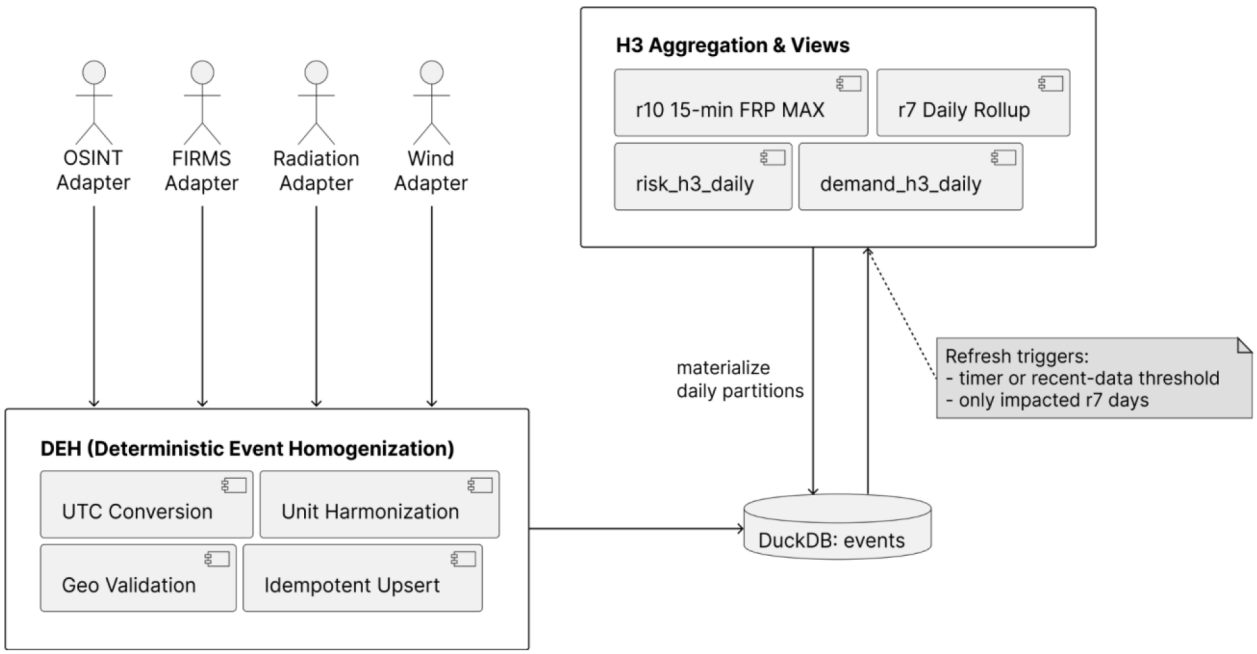


Fig. 1. PSAI-DEH to H3 Risk/Demand (ST-CQRS)

The core of our spatial aggregation is the H3 geospatial index. H3 divides the world into hexagonal cells at varying resolutions (r0 through r15). Each H3 cell has a unique 64-bit index identifying its location and level in the hierarchy. We use resolution 10 for fine-grained "micro-tiles" (~ hexagons with side length ~0.3 km, area ~2 km^2) and resolution 7 for daily summary tiles (~ hexagons ~20 km across, area ~1,200 km^2). An H3 cell at r7 is the "parent" of many r10 cells (specifically, each hexagon splits into 7 children at the next finer resolution in this hexagonal hierarchy. This hierarchy makes it easy to aggregate data: we assign each event to an r10 cell, then roll up to r7 by using H3 parent indexing.

We maintain two main materialized view tables in DuckDB (which we refresh periodically or on new data).

- 15-min H3 r10 micro-tiles: this intermediate table groups events into 15-minute time bins and H3 r10 cells. It is primarily used to deduplicate and aggregate high-frequency data like fire detections.
- Daily H3 r7 views: using the micro-tiles, we aggregate up to daily values on H3 r7 cells. These form the *risk_h3_daily* and *demand_h3_daily* tables, which can be queried or exported.

DuckDB's efficient columnar engine and SQL support make these rollups fast and easy to manage with SQL statements. We chose DuckDB for its embeddability and performance: it can scan and aggregate millions of rows per second in-memory, and it supports direct operations on Parquet files with zero serialization overhead. For our use case (a national-scale dataset updated in real-time on a single machine), DuckDB provides near real-time query capability without the complexity of a distributed database. It essentially allows us to implement the CQRS pattern in-process: writes (commands) affect base and aggregate tables, and reads (queries) hit the pre-computed daily views.

Research results

Time, Units, and Idempotency. All inputs are standardized to naive UTC; if tzinfo is missing, adapters pass a hint before conversion. Units are normalized deterministically (radiation dose rates to sieverts; wind to m/s). A hashed natural key (metric, timestamp_utc, lat, lon, source) enforces idempotent upserts so restarts or late arrivals do not duplicate records. Parquet snapshots persist daily *risk_h3_daily* and *demand_h3_daily* for archival exchange.

FRP Micro-Aggregation (r10 × 15-min). FIRMS detections from overlapping satellites can over-count the same fire. We assign each event to an H3 r10 cell and a 15-minute bin, then take the MAX (FRP) per bin/cell.

This collapses duplicates while preserving peak intensity and matches the VIIRS 375 m footprint used to set r10 granularity (Table 2).

Table 2

DuckDB SQL: r10 15-min FRP dedup (MAX per bin/cell)

```
CREATE OR REPLACE TABLE frp_dedup_15m_r10 AS
WITH s AS (SELECT date_trunc('minute', timestamp)
- INTERVAL (EXTRACT(minute FROM timestamp)::INT % 15) MINUTE AS ts_15m,
h3_latlng_to_cell(lat, lon, 10) AS h3_r10,
value AS frp
FROM events
WHERE metric LIKE 'frp_%'
AND lat IS NOT NULL AND lon IS NOT NULL)
SELECT ts_15m, h3_r10, MAX(frp) AS frp_max
FROM s
GROUP BY ts_15m, h3_r10;
```

Daily Rollup (r7). The micro-tiles roll up by parent using '*h3_cell_to_parent(..., 7)*', and we sum '*frp_max*' by day and r7 cell, r7 balances regional interpretability with compact storage. The resulting '*frp_daily_r7*' table becomes a component in risk (Table 3).

Table 3

DuckDB SQL: r7 daily rollup from r10 micro-tiles

```
CREATE OR REPLACE TABLE frp_daily_r7 AS
SELECT
CAST(ts_15m AS DATE) AS d,
h3_cell_to_parent(h3_r10, 7) AS h3_r7,
SUM(frp_max) AS frp_sum
FROM frp_dedup_15m_r10
GROUP BY d, h3_r7;
```

Risk and Demand Materialization. Daily risk combines OSINT hit counts, log-scaled FRP ('*ln(1+frp_sum)*'), a simple radiation positive proxy, and a wind-alignment proxy. Demand multiplies risk by r7 population.

We keep weights explicit in SQL to ease tuning and audits; joins use integer H3 ids and convert to strings only at export (Table 4).

Table 4

DuckDB SQL: simplified daily risk composition

```
CREATE OR REPLACE TABLE risk_h3_daily AS
SELECT
r.d,
r.h3_r7 AS h3,
COALESCE(h.n_hits, 0) AS n_hits,
COALESCE(r.frp_sum, 0) AS frp_sum,
COALESCE(g.rad_pos, 0) AS rad_pos,
COALESCE(w.wind_align, 0) AS wind_align,
0.5 * COALESCE(h.n_hits, 0)
+ 0.3 * ln(1 + COALESCE(r.frp_sum, 0))
+ 0.1 * COALESCE(g.rad_pos, 0)
+ 0.1 * COALESCE(w.wind_align, 0) AS risk_score
FROM frp_daily_r7 r
LEFT JOIN hits_daily_r7 AS h USING (d, h3_r7)
LEFT JOIN gamma_daily_h3 AS g ON (g.d = r.d AND g.h3 = r.h3_r7)
LEFT JOIN wind_daily_h3 AS w ON (w.d = r.d AND w.h3 = r.h3_r7);
```

Measured Behavior. On an M2 laptop, end-to-end updates are typically sub-second; FRP dedup reduces raw points by ~99 %; '*risk_h3_daily*' materialization reaches tens of thousands of rows per second. Indicative values are summarized below (Table 5).

Table 5

Engineering metrics (measured)

Metric	Scope	Value (median)	p95	Notes
Ingest → MV latency (24h window)	end-to-end update	0.089 s	0.923 s	sub-second typical
Ingest → MV latency (heavy days)	ranked by FRP ∪ Hit tiles	0.0303 s	0.0366 s	warm path
FRP dedup reduction	r10 × 15-min MAX(FRP)	0.8–0.9 %	–	~99 % removed
MV throughput (risk_h3_daily)	rows/s (examples)	31.9–55.3k	–	five heavy days
Storage footprint (daily)	Parquet exports	~0.0–0.1 MB	–	sub-megabyte footprint (typically tens of kilobytes)

Conclusions and future work

This research shows that a small, deterministic pipeline – PSAI ingest, DEH normalization, and H3 aggregation in DuckDB – yields daily risk and demand layers with seconds-level latency on a single machine. The design favors reproducibility and clarity: adapters emit the same shape of events; DEH enforces UTC and units; aggregates are expressed in SQL; and storage remains compact. The evaluation confirms low-latency updates, extreme FRP deduplication, and high materialization throughput, which together are sufficient for national-scale monitoring on laptop-class hardware. Future improvements are straightforward. Late and corrected data can be handled with short grace windows on micro-tiles, followed by targeted refresh of impacted r7 days. Radiation harmonization should move beyond fixed factors toward sensor-specific calibration curves and cross-validation against regional baselines. Wind influence can progress from a proxy to directional plume modeling that links sources to downwind population. Network travel times, built with OSMnx and Valhalla, can transform demand into "reachable population" and support logistics decisions under constraints. Solar-angle adjustments for FRP and more robust anomaly detection will raise signal quality without changing the core interfaces. Incremental refresh strategies and lightweight change data capture for 'events' can reduce compute even further as volumes grow – all while preserving the deterministic contracts that make the system easy to operate and extend [1–9].

References

1. Raasveldt M., Mühleisen H. (2019). DuckDB: an in-process analytical database. *Proceedings of the 2019 ACM SIGMOD International Conference on Management of Data* (Demo). URL: https://duckdb.org (access date: 03.09.2025).

2. Uber Engineering. (2018). H3: Uber's Hexagonal Hierarchical Spatial Index. URL: https://www.uber.com/blog/h3/ (access date: 03.09.2025).

3. NASA LANCE. (2023). Fire Information for Resource Management System (FIRMS). URL: https://firms.modaps.eosdis.nasa.gov (access date: 03.09.2025).

4. Schroeder W., Oliva P., Giglio L., Csiszar I. (2014). The VIIRS 375 m active fire detection data suite. *Remote Sensing of Environment*, vol. 143, pp. 85–96. https://doi.org/10.1016/j.rse.2013.12.008 (access date: 03.09.2025).
5. European Commission. EURDEP – European Radiological Data Exchange Platform. URL: https://rem.jrc.ec.europa.eu (access date: 05.09.2025).
6. SaveEcoBot. Radiation map & data access. URL: https://www.saveecobot.com/en (access date: 03.09.2025).
7. OpenWeather. OpenWeatherMap API. URL: https://openweathermap.org/api (access date: 03.09.2025).
8. Boeing G. (2017). OSMnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks. *Computers, Environment and Urban Systems*, vol. 65, pp. 126–139. https://doi.org/10.1016/j.compenvurbsys.2017.05.004 (access date: 03.09.2025).
9. Valhalla. Open Source Routing Engine. URL: https://github.com/valhalla/valhalla / (access date: 05.09.2025).

Дата першого надходження рукопису до видання: 24.09.2025
Дата прийнятого до друку рукопису після рецензування: 21.10.2025
Дата публікації:

UDC 534.3:621.396.94 DOI

O. M. KATRUSHA
MSc in Applied Mathematics,
Postgraduate Student at the Department of Artificial Intelligence
Educational and Research Institute for Applied System Analysis
of National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"
ORCID: 0009-0008-7097-4843

INTENSITY CORRECTION IN SIDE SCAN SONAR IMAGES. METHODS OVERVIEW

The use of sonars, particularly side-scan sonars, in underwater research dates back to the mid-20th century. Due to their relative affordability, ease of operation, and high efficiency in seabed visualization, these devices have become indispensable tools in hydrography, marine archaeology, environmental monitoring, and search-and-rescue operations. However, the processing of sonar images (sonograms) presents a number of challenges caused by signal distortions. The main types of such artifacts include intensity non-uniformities, stripe noise, geometric distortions, and residual effects of imperfect time-based intensity compensation. Given the complexity of acoustic propagation in underwater environments and the variability of sensor configurations, there is currently no universal method for intensity correction that performs effectively across all scenarios. This study presents a structured review of existing methods for correcting intensity in side-scan sonar images developed over recent decades. Emphasis is placed on their algorithmic implementation, suitability for real-time processing, and effectiveness in constructing high-quality sonar mosaics. Particular attention is paid to the analysis of the causes of intensity variation, including the phenomenon of brightness falloff across the swath, repetitive stripe noise, time-varying gain, and residual intensity anomalies in the time domain. The review covers a range of models and approaches, including empirical smoothing techniques, multivariate regression, local normalization, hybrid filtering strategies, and methods based on physical models of acoustic scattering. A proposed classification framework allows for the organization of these approaches according to several criteria: model type (empirical, physical, machine learning), underlying assumptions, constraints under real-world conditions, and the types of metrics used for quality evaluation. The potential of each method to be adapted for tasks such as automatic object detection and the construction of accurate seafloor morphology models is also explored. This material may be valuable both for engineering practitioners involved in applied sonar processing and for researchers seeking advanced algorithms for sonogram enhancement and develop adaptive computer vision systems for complex underwater environments.

Key words: *side-scan sonar, sonogram, intensity correction, sonar mosaic, stripe noise, digital signal processing, adaptive algorithms, deep learning, computer vision.*

О. М. КАТРУША
магістр прикладної математики,
аспірант кафедри штучного інтелекту
Навчально-науковий інститут прикладного системного аналізу
Національного технічного університету України
«Київський політехнічний інститут імені Ігоря Сікорського»
ORCID: 0009-0008-7097-4843

КОРЕКЦІЯ ІНТЕНСИВНОСТІ ЗОБРАЖЕНЬ СОНАРА БОКОВОГО ОГЛЯДУ. ОГЛЯД МЕТОДІВ

Використання сонарів, зокрема сонарів бокового огляду, у підводних дослідженнях бере свій початок із середини XX століття. Завдяки своїй відносній дешевизні, простоті експлуатації та високій ефективності у візуалізації морського дна, ці пристрої стали невід'ємним інструментом гідрографії, морської археології, екологічного моніторингу та пошуково-рятувальних операцій. Однак обробка сонограм супроводжується низкою складнощів, зумовлених спотвореннями сигналу. До основних типів таких артефактів належать неоднорідності інтенсивності, смуговий шум, геометричні викривлення та залишкові наслідки недосконалої компенсації часової залежності сигналу. Враховуючи складність акустичних властивостей середовища та варіативність конфігурацій сенсорів, універсального методу корекції інтенсивності, який був би ефективним у всіх сценаріях, наразі не існує. У цьому дослідженні представлено систематизований огляд існуючих методів корекції інтенсивності зображень сонара бокового огляду, що були розроблені впродовж останніх десятиліть. Акцент зроблено на їхню алгоритмічну реалізацію, застосовність у реальному часі та ефективність у контексті побудови високоякісних сонарних мозаїк. Особливу увагу приділено аналізу причин виникнення варіацій інтенсивності, зокрема феномену

© Katrusha O. M., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

спадання яскравості поперек напряму руху носія, повторюваним смуговим шумам і залишковим артефактам корекції інтенсивності у часовому домені. Розглянуто низку моделей та підходів, серед яких емпіричні згладжувальні методи, багатовимірна регресія, локальна нормалізація, гібридні фільтраційні стратегії та методи на основі фізичних моделей акустичного розсіювання. Запропонована класифікаційна схема дозволила структурувати відомі підходи за критеріями: тип моделі (емпірична, фізична, машинне навчання), вихідні припущення, обмеження в умовах реального застосування, а також тип використовуваних метрик для оцінки якості обробки. Окремо розглянуто потенціал кожного методу щодо адаптації до задач автоматичного розпізнавання об'єктів та побудови точних морфологічних моделей дна. Матеріал може бути корисним як для інженерів-практиків, що займаються прикладною обробкою сонарних зображень, так і для дослідників, зацікавлених у подальшому розвитку алгоритмів покращення якості сонограм та створенні адаптивних систем з комп'ютерним зором для роботи у складних підводних умовах.

Ключові слова: *сонар бокового огляду, сонограма, корекція інтенсивності, сонарна мозаїка, смуговий шум, цифрова обробка сигналів, адаптивні алгоритми, глибинне навчання, комп'ютерний зір.*

Problem Statement

In the context of expanding maritime autonomy and the urgent need for scalable, low-cost methods of seafloor imaging and underwater situational awareness, side-scan sonar (SSS) has emerged as a foundational technology across civil, scientific, and defense applications. Its ability to deliver wide-area acoustic coverage with high spatial resolution makes it essential for seabed mapping, marine archaeology, habitat monitoring, infrastructure inspection, and mine detection missions [1]. However, the practical utility of SSS data is critically constrained by the instability of sonar signal intensity, which is influenced by distance, incidence angle, seafloor texture, sediment composition, beam directivity, and sound frequency [2]. These factors cause significant spatial and temporal intensity variation, leading to radiometric inconsistencies such as across-track falloff, stripe noise, and residual artefacts from time-based gain correction. As a result, downstream tasks such as mosaicking, automated object detection, and semantic interpretation are adversely affected.

The problem is further compounded by the operational shift toward autonomous platforms such as AUVs, which require onboard, real-time image enhancement to support navigation, SLAM, and situational response. In such settings, empirical correction models and fixed-parameter filters – commonly used to compensate for signal degradation – fail to adapt to rapidly changing environmental and vehicle conditions, resulting in unstable performance and geometric distortions [3]. Moreover, most classical correction techniques do not scale well with modern mission constraints, where power, memory, and compute resources are limited and mission duration is long.

Consequently, the applied scientific challenge lies in developing intensity correction strategies that account for the nonlinearity and dynamism of underwater acoustic propagation, while enabling robust real-time integration into low-power embedded systems. The potential of hybrid methods – combining physical acoustic modeling, contextual metadata from mission logs, and lightweight machine learning – opens a pathway toward improved radiometric fidelity, more accurate object segmentation, and more efficient data fusion with other sensing modalities [4]. There is thus a pressing need for signal-processing solutions that treat radiometric correction not as a secondary enhancement task but as a mission-critical component of sonar-based perception pipelines in complex, variable underwater environments.

Analysis of Recent Studies and Publications

Current challenges in processing and radiometric correction of side-scan sonar (SSS) imagery are addressed in the works of P. Blondel, A. Grządziel, Y. Zhang, X. Ye, J. Zhao, S. Li, A. Burguera, C. Capus, and J. Clarke. These studies examine a broad spectrum of intensity-formation issues, ranging from sensor-resolution effects and acoustic-shadow artefacts to seabed composition and propagation physics. The papers by Zhang et al. and Ye et al. compare contrast-enhancement and brightness-correct algorithms that employ Retinex models and adaptive smoothing over rugged topography, while Zhao, Liu and co-authors focus on compensating intensity variations driven by sediment heterogeneity and complex bottom morphology. Other researchers – particularly Burguera, Capus, and Clarke – propose integrated frameworks that jointly address geometric and radiometric distortions under real-world conditions, from towed arrays to AUV-mounted systems, and discuss normalisation techniques that support reliable classification and the construction of radiometrically uniform acoustic mosaics. Despite these advances, most publications tackle isolated facets – such as stripe-noise filtering or single-pass intensity levelling in controlled environments – leaving the overarching problem of a universal, environment-adaptive correction algorithm, suitable for real-time onboard execution on autonomous underwater platforms, largely unresolved and in need of further systematic investigation.

The aim of this study is to provide a structured and comprehensive analysis of publicly available intensity correction methods for side-scan sonar (SSS) imagery. The research focuses on identifying the underlying models and conceptual approaches, evaluating their theoretical assumptions, implementation complexity, and real-time applicability, as well as exploring potential secondary outputs such as improved contrast, noise suppression, or radiometric normalization.

This methodological review is motivated by the practical demands of underwater imaging and is intended to support engineers and researchers developing sonar data-processing pipelines or seeking directions for further investigation. Given that many commercial solutions are protected by intellectual property rights and confidentiality agreements,

the study is deliberately limited to sources from open-access scientific literature. No experimental evaluation or direct benchmarking of specific methods was conducted, as the emphasis is placed on comparative methodology rather than empirical performance.

Results

The signal strength of the reflected underwater sound signal has complex dependency on many factors: spherical spreading of acoustic waves reversely proportional to fourth square of range [1], sound attenuation in water and in sediment, grazing angle, floor profile, sediment type, other factors like water temperature, pressure, salinity, chemical composition of water or water currents. From the sonar or vehicle side the factors include sonar geometry, altitude, beam pattern, calibration and setup, working frequency, sonar movement (roll, pitch, yaw, heave). Acoustic shadows and reverberations also contribute to different intensity variations. For a comprehensive description and mathematical modeling of sound propagation see [1, 11]. This makes it difficult to separate backscatter information about the object of interest (floor profile, object) from other factors. Therefore, the main aim of intensity correction methods is to compensate for unwanted decay in intensity, while preserving valuable information about seafloor characteristics. The intensity of the emitted signal is not uniform by direction and depends on the grazing angle (for terminology see, for example, [12]) due to the physics of sonar construction. Main and side lobes of a sonar form a beam pattern that contributes to sonogram intensity variations. This non-uniformity is a reason for over- or under-ensonification of some areas as well as stripe noise due to sonar rolling. Beam patterns can either be modeled theoretically [13], measured empirically based on sonogram data [14], or related beam function [15] derived and used for intensity correction afterwards.

Dynamic range of side scan sonar can reach more than 100 dB [10] due to the above-mentioned reasons making sonogram unreadable as-is recorded by the sonar. For this reason sonars and processing software employ a time-varied gain (TVG) function that compensates sound decay to reduce the dynamic range to around 40 dB [16] and compose a more uniform and readable sonar image using logarithmic or another approximation [5, 1]. In general case TVG parameters must be carefully selected for the sediment type, floor depth and other factors to compensate for sound attenuation more adequately. Unfortunately, TVG is often implemented on hardware level with little or no possibility of adjustment and therefore does not account for other factors that may need compensation. Moreover, the function itself may contain irregularities that bring additional distortion to intensity [17]. Thus, additional compensation of its residuals affecting SSS images may be needed [18, 19, 20].

Different variation types require different methods of correction. Most of the reviewed methods tackle the first problem. The second type has little coverage in literature, although some authors claim [9] to have compensated for it as well. Mosaic stripe noise is addressed in [21] using 2D Fourier transformation. Most of the reviewed methods address across-track intensity decay (Fig. 1), which is the most common and well-documented type of variation. The near-nadir overexposure and stripe artefacts illustrated in Fig. 2 are less frequently discussed, although some studies report partial compensation of these effects [9]. Mosaic-level stripe noise, often resulting from inconsistent swath intensities across survey lines, is illustrated in Fig. 3 and has been addressed using 2D Fourier-based filtering approaches [21].

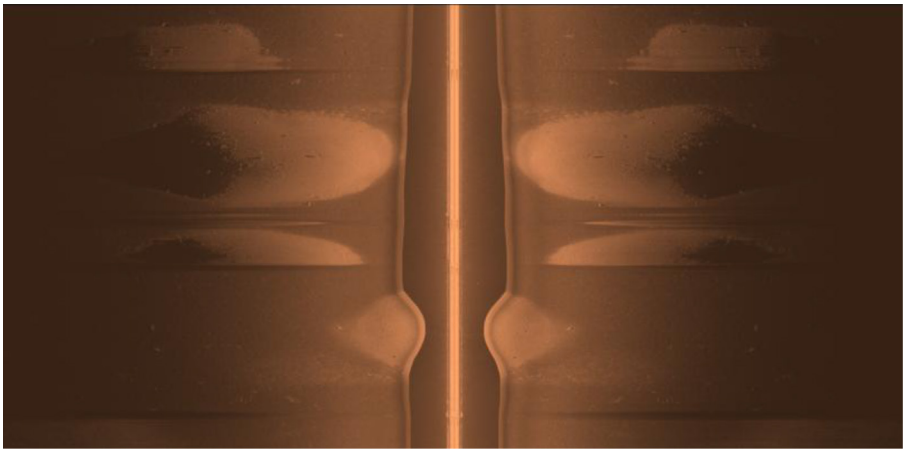


Fig. 1. An example of intensity decay with range

The intensity variations make sonograms less readable by humans and cause other hindrances in subsequent image processing and usage [22]. Ramifications include loss of information in under- or over-ensonified areas, distortion of distinctive features, poor reception by AI and post-processing software [12], failures in detection and classification [23, 14] of objects and sediment types segmentation [5], unwanted artifacts on mosaics [18, 21] and complicate fusion with other data from other modalities [16, 19, 4]. This highlights the importance of adequate intensity correction methods that correspond to sonogram use cases.

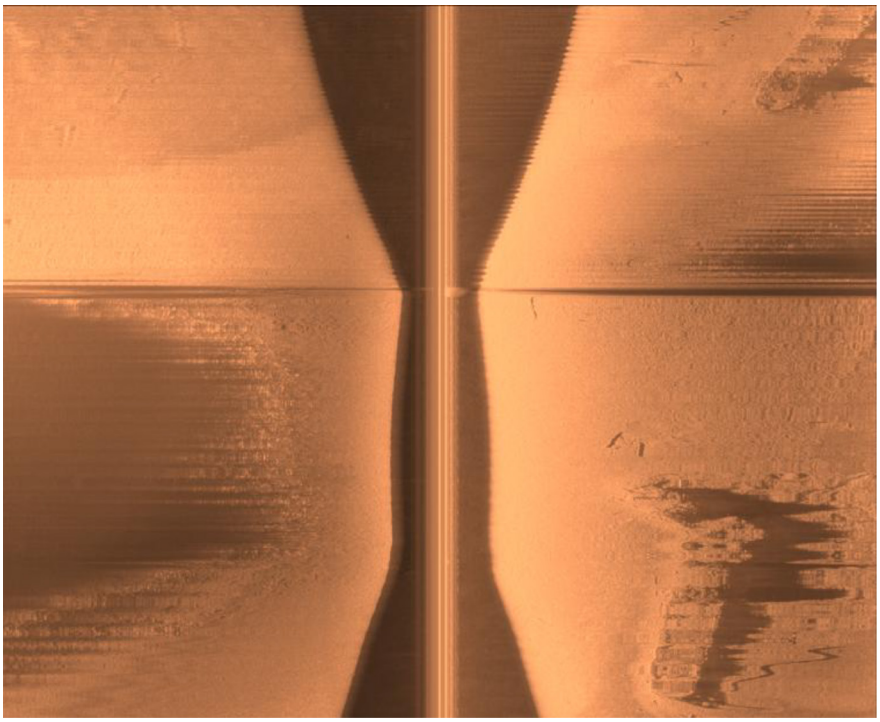


Fig. 2. An example of over-ensonified near-nadir area (bright areas in the middle) and stripe noise (frequent horizontal lines of various intensity and length)

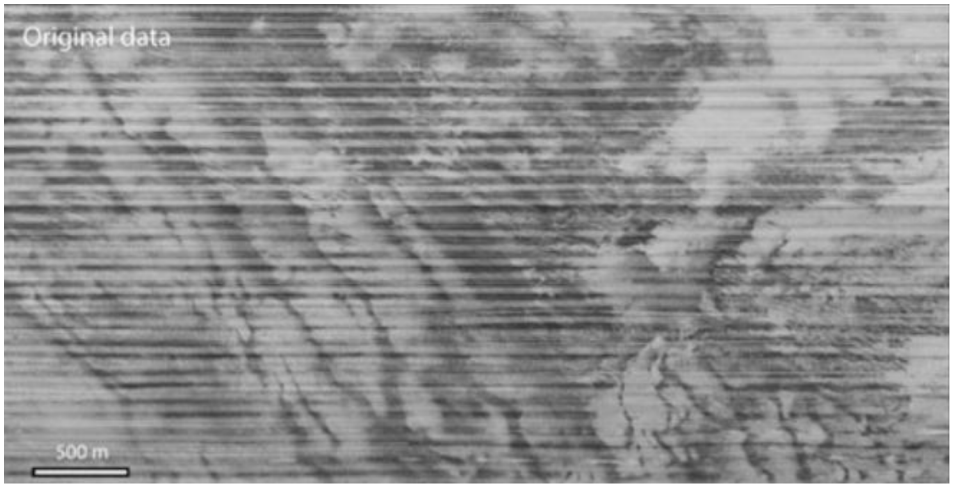


Fig. 3. An example of stripe noise on sonar mosaic caused by intensity variations [21]

Due to the variety of sonogram applications, the methods of sonar data processing also have significant diversity. A sonogram is a synthetic image without any reference to "ground truth" that can be seen and evaluated by a human eye. As formulated by Blondel in [1]: "The "quality" of an image is very subjective… Is a higher contrast necessarily better? What would be the optimal contrast?". No single method can be deemed universal or true in the application to different sonogram use cases. For example, researchers may want to mark different sediment or underwater vegetation with different shades of gray and that will significantly affect the choice of intensity correction methods. Processing legacy sonogram data is another use case that sets up requirements for processing methods.

Legacy sonograms, especially recorded on analog media, may not contain complementary information like navigational data, vehicle or towfish motions, sonar type and calibration. Thus, applicable methods should only use the data present on the sonogram itself. Some objectives may contradict others like making the image intensity more uniform for better readability and keeping variations in sediment brightness or acoustics shades for better or object detection and classification. Not every method reviewed in this research mentions the particular use case it works best for, most of them pretend to be generic. However, if particular use cases have been specified by the paper author, we will include them in the resulting table.

Various intensity correction approaches for side-scan sonar imagery are based either on physically informed models of acoustic propagation and reflection or on statistical and signal-based techniques aimed at achieving radiometric uniformity under diverse environmental and operational conditions. Ye [4] has split intensity correction methods into six categories: Time Variant Gain (TVG), Histogram Equalization (HE), Nonlinear Compensation, Function Fitting, Sonar Propagation Attenuation Model and Beam Pattern. However, some methods like [12], may be classified into two proposed categories. Our classification will be less strict but more descriptive, outlining the main model or approach of each method: regression over ping intensity profiles, averaging along or across track, Lambertian model of sound reflection, etc. Several examples of intensity correction using different methods are shown in Fig. 4 [4].

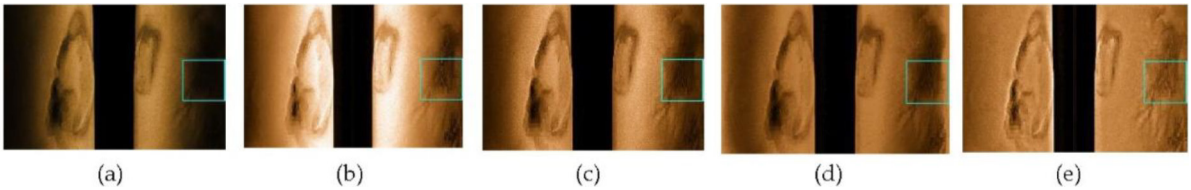


Fig. 4. Comparison of intensity correction in side scan sonar images:
a – original image; b – histogram; c – non-linear compensation; d – function fitting; e – method in [4]

In our subjective opinion, methods that do not use any model, especially those using averaging, are easier to implement using packages and modules of statistical processing. Moreover, such modules are available in abundance. On the other hand, module-based methods may produce better results in complicated circumstances.

Every model is a simplified representation of the target objects. With all the complexity of sound emission, propagation, and reflection underwater, different methods resort to different assumptions and declare various limitations in order to simplify the modeling process. Most common assumption considers the seafloor to be flat and uniform across-track at a particular point. This greatly simplifies calculation of grazing angle and other values from altitude and slant range data. Lambertian assumption (Lambertian law) [1] is another popular assumption with researchers. It states that intensity of the reflected signal is proportional to cosine of the angle between the incident light and the surface normal, which is a fairly good approximation in most cases. Other assumptions may include small variation in depth during sonar mission, uniformity of depth and sediment in some proximity of the calculation point.

Method users must be aware of such assumptions and limitations as they restrict method applicability or cause side effects when assumptions do not hold. For example, if the seafloor is not flat with high depth and slope variations, it may cause over- or undercompensation of a signal. Table 1 illustrates the main models or approaches, declared assumptions and limitations of the reviewed methods.

Sonograms may comprise hundreds of thousands of lines with thousands of data points in each line. This may result in gigabytes of information produced during relatively short exploration missions. With this regard, computational effectiveness is a very important criterion of choice when it comes to selecting a proper processing method.

Recent achievements in controlled and unmanned vehicle design pose new requirements for effectiveness of on-board signal processing units given limited onboard computational power, energy consumption, connection line bandwidth or absence of timely feedback from the operator. There may also be a need for real-time processing like instant object recognition during autonomous navigation. Some use cases that presume quick reaction time (mine detection or rescue missions to name a few) may require real-time or nearly real time performance which also limits the choice of methods.

Few works provide quantitative estimates of computational complexity, some claim effectiveness or suitability for real-time usage qualitatively. Moreover, processing time greatly depends on implementation details, optimization techniques and, of course, available hardware. This makes a direct comparison of method's speed barely possible.

However, we may assume that methods employing averaging techniques prove to be quite effective in terms of computational resources. On the other hand, effective use of along-track averaging is only possible after the whole mission or a significant part of it has been completed which prevents them from being used in real-time. One should also consider the year of publication and advance in computational power and software tools since then. Besides the intensity correction itself, some of the researched methods offer additional benefits for their users. This can be a calculated sonar beam pattern or detected sediment type. Table 2 lists the methods' computational complexity and side outputs.

A range of fundamental and supplementary techniques employed in intensity correction methods has been identified in the literature. These approaches reflect both physically grounded and data-driven strategies for improving sonar image consistency. Common techniques include intensity averaging across the track line, along-track direction, or within local pixel neighborhoods [8, 9, 13]; regression-based modeling of sonar ping data using exponential, spline, or other types of interpolation [16, 22]; and frequency-domain separation to isolate low- and high-frequency signal components. The latter can serve purposes such as decomposing illumination and reflectance [4, 9, 22] or mitigating stripe noise artifacts [21].

Table 1

Methods classification by use case, model, and basic assumptions

Reference	Use case(s) / problem	Model or Approach	Assumptions and limitations
Al-Rawi 2016 (MIRA) [10]	– Fusion – Generic	Exponential regression of a ping	Exponential distribution of signal
Al-Rawi 2017 (Cubic spline) [16]	– Mapping – Underwater computer vision	Cubic spline regression of a ping	Rayleigh distribution of signal
Anstee 2001 [18]	Mosaicking	– Along and across track averaging, separately for near nadir area and remaining area	– Flat floor – Altitude does not change much – Not been tested with data from lower frequency side scan sonars
Burguera and Oliver 2014 [12] Burguera and Oliver 2016 [23]	– Mapping – Navigation or SLAM – Mosaicing – Generic	– Lambertian model – Sensitivity pattern model by Kleeman and Kuc – Beam pattern model	– Flat floor – Lambertian assumption
Capus 2004 [19]	– Mosaicking – Classification – Conversion to 8-bit – Correction of TVG residuals and beam pattern	– TVG model – Along-track average – Beam pattern estimate	– Small altitude changes – Platform stability with respect to pitch and roll – Seafloor unchanging in slope across track
Capus 2008 [13]	– Mosaicking – Segmentation – Correction of TVG residuals	– Imaging model – Quadratic function describing the variations in intensity associated with sonar altitude – Beam pattern model – Iterative process to separate angular and range dependent intensity variations	– Suitable for shallow water and low altitude – Errors arise at each course alteration – Needs a suitable exemplar image for generation of correction factors.
Yet-Chung Chang 2010 [24]	– Mosaicking – Generic	– Average signal intensity for each grazing angle – Average energy level of the 20 pings – Normalizing ping energy levels to remove	– Total back-scattered energy from each ping should be similar to adjacent pings in the time series
Clarke 2004 [14]	– Classification	– Combining a number of discrete spatial frequency bins with average backscatter strength – Stack backscatter strength information in angular bins – Empirical approach was to estimate beam patterns	– Flat floor – Shallow water
Galdran 2017 [25]	– Mapping – Generic	– Exploiting two-dimensional information to estimate and remove intensity nonuniformity – Locally normalizing the intensity in each region to retrieve a more regular image	– Simple logarithmic dynamic range compression is applied to compress the received acoustic signal – Background intensity follows a normal distribution
Shippey 1994 [20]	– Mosaicking – Segmentation	– Histogram normalization to normal distribution – Preserve the characteristic histogram shape for each sediment	– Not ideal for the segmentation task
Wilken 2012 [21]	– Mosaicking – Stripe noise removal – Classification	– 2D Fourier transformation adjusting slope angle of the stripe noise and filtering width	– Spatial Nyquist sampling criteria must be fulfilled – Sand ripples prone to elimination by the filtering process – Data gaps / clipped areas appear smeared or blurred along all filtered stripe directions
Jianhu Zhao 2017 [5]	– Generic – Sediment variations	– Linear relationship between distortion and sonar altitude – Average angle–backscatter curves of individual sediment – Unsupervised sediment classification – Angle-related radiometric correction applied for each sediment	None
Cervenka 1993 [22]	Mosaicking	– Filter the low spatial frequency components of the image using Chebyshev polynomials – Contrast enhancement through histogram equalization by balancing local versus global histogram contribution	– The original image must fit exactly into a rectangular frame for the method to work satisfactorily

End of Table 1

Yifei Liu 2023 [8]	– Mosaicking – Classification in rugged terrain	– Multiplicative Attenuation Model – Divide the image into multiple segments along the vertical direction of the track – Image is segmented along the attenuation direction, and the segmentation intensity is calculated by using the multiplicative model for the echo points. – Maximum segmentation intensity is taken as the target intensity, and the non-shaded points of the image are compensated and corrected	– Seawater is uniform – Influence of boundary loss is not discussed
Xiufen Ye 2019 [4]	Generic	– Retinex mean filter or bilateral filter to directly smooth signal intensity – Mean filter and bilateral filter	None
Li, Shaobo 2022 [9]	Generic	– Lambert's law and the variational Retinex model – Beam pattern model – Low-rank constraint to obtain a better illumination component. – ATV constraint is used to retain geomorphology features and remove the stripe noise	– Illuminance component in log domain is smooth in spatial – Should be applied to the geometric corrected SSS image where the water column has been removed

Table 2

Classification of Methods by Computational Complexity and Side Outputs

Reference	Computational complexity / performance	Side outputs
Al-Rawi 2016 (MIRA) [10]	Within the order O(N), N = total number of pixels in the image	None
Al-Rawi 2017 (Cubic spline) [16]	n/a	– The peak of the estimated distribution is related to the sensor gain – Can also be used to detect high acoustic reflectance and acoustic shadows of underwater landmarks
Anstee 2001 [18]	"The algorithm imposes minimal processing overheads on modern personal computers." [Anstee]	n/a
Burguera and Oliver 2014 [12]	n/a	n/a
Burguera and Oliver 2016 [23]	n/a	Echo Intensity Map
Capus 2014 [19]	n/a	– Beam pattern model – TVG residual model
Capus 2008 [13]	– Typical processing times for a 2000 x 2000 pixel image, including resampling, would be between 2 and 15 s depending on altitude variation.	n/a
Yet-Chung Chang 2010 [24]	n/a	n/a
Clarke 2004 [14]	n/a	– Roll-caused stripe noise removal
Galdran 2017 [25]	"Great computational efficiency, being a good candidate for a real-time implementation." Complexity O(N), N = total number of pixels in the image	n/a
Shippey 1994 [20]	n/a	Column histogram for each sediment type
Wilken 2012 [21]	n/a	Can be applied to non-optimally processed side-scan mosaics
Jianhu Zhao 2017 [5]	n/a	Sediment classification
Cervenka 1993 [22]	"From a practical point of view, Chebyshev analysis is not difficult to perform."	n/a
Yifei Liu 2023 [8]	"Exhibits excellent performance in image correction"	n/a
Xiufen Ye 2019 [4]	– Mean filtering suitable for online processing – Bilateral filtering method is more suitable for offline processing	Can also be used to enhance low illumination color optical images
Li, Shaobo 2022 [9]	n/a	Stripe noise removal

Bottom detection based on altitude measurements or water column analysis is also widely applied [19], although it presents specific challenges in high-noise environments or areas with abrupt bathymetric variations.

Slant range correction (resampling) is often a prerequisite for precise grazing angle computation or enhanced spatial averaging [19, 24]. In some methods, sediment classification – either predefined or derived post hoc – is used to improve reflectivity estimation and aid intensity normalization [20, 5].

Other widely used techniques include histogram leveling [20, 3], outlier filtering using Z-score methods [5], and Bayesian inference of time-varying gain (TVG) profiles directly from sonar image data [17]. These tools are not only useful for intensity correction but also show potential for the development of novel methods or as auxiliary components in broader image processing pipelines. Several additional methods, while not directly designed for intensity correction, offer valuable insights into sonar image enhancement. For instance, Nguyen [26] introduced an empirical technique to estimate beam patterns, and Tamsett [27] proposed a more accurate TVG correction model by dividing the ensonified seafloor into primary and conjugate zones. He also suggested a method for deriving the sonar's beam function [15]. Calder [17] employed Bayesian inference to reconstruct the TVG profile solely from image data and applied it to mitigate residual sonar artifacts. Completely objective evaluation of intensity correction methods remains challenging due to the lack of a definitive ground truth in sonar imaging. Nevertheless, various quantitative approaches have been used for performance assessment. Comparative analysis of multiple methods, such as MIRA versus Dark Channel Prior (DCP), offers direct insights into algorithmic differences [10]. Histogram-based evaluation focuses on distribution shape, deviation from normality, balance, and the presence of distinct modes; some studies calculate the coefficient of variation (CV) as a supporting metric [16, 5]. Several standardized and purpose-built metrics have also been proposed. The Sonar Image Quality Evaluation Metric (SIQEM) [10] provides a domain-specific assessment framework. Structural Similarity Index (SSIM) is used to quantify perceptual similarity between two images [28], while peak signal-to-noise ratio (PSNR), entropy, standard deviation, and mean gradient serve as general image quality indicators [4]. Additional metrics such as the Vollrath function, Roberts function, Brenner gradient, mean gradient, and gray-level difference have been utilized for fine-grained assessments [3, 8].

These evaluation tools can support the benchmarking of new intensity correction methods or comparative studies across different imaging conditions and datasets.

Conclusion

This study addresses the issue of intensity correction in side scan sonar imagery, an essential task for accurate underwater analysis and mapping. Given the widespread use of side scan sonar in fields ranging from oceanography and archaeology to military applications, the need for effective intensity correction methods specific for particular use cases is paramount. We have identified various reasons for intensity variation in sonograms, including factors like sound attenuation, sonar beam patterns, time-varied gain (TVG) residuals, and environmental conditions. Through an extensive literature review and method analysis, we have categorized and evaluated the discovered intensity correction methods based on their models, assumptions, computational complexity, and practical applications. Our findings highlight that no single method is universally applicable due to the diversity of sonar images use cases and intensity variation reasons. While some methods excel in real-time processing scenarios, others may provide more accurate corrections using sophisticated models at the cost of computational complexity. The results prove that a variety of correction methods are available covering different variation reasons, use cases and computational requirements that provide a good selection for implementation and evaluation. We have also listed the main techniques used within the methods that can prove useful in implementation and research. Intensity correction is one of the first mandatory steps in any sonar imaging processing tasks like object detection and segmentation that will be covered in our future research.

Bibliography

1. Blondel P. The handbook of sidescan sonar. Berlin: Springer Science & Business Media, 2010. Electronic resource. URL: https://link.springer.com/book/10.1007/978-3-540-49886-5 (date of access: 21.07.2025).

2. Grządziel A. The impact of side-scan sonar resolution and acoustic shadow phenomenon on the quality of sonar imagery and data interpretation capabilities. *Remote Sensing*. 2023. Vol. 15, No. 23. Article number: 5599. DOI: https://doi.org/10.3390/rs15235599

3. Zhang Y., Li H., Zhu J., Zhou L., Chen B. Contrast study of side scan sonar image enhancement methods // 2021 OES China Ocean Acoustics (COA). IEEE, 2021. P. 995–999. DOI: https://doi.org/10.1109/COA50123.2021.9520059

4. Ye X., Yang H., Li C., Jia Y., Li P. A gray scale correction method for side-scan sonar images based on Retinex. *Remote Sensing*. 2019. Vol. 11, No. 11. Article number: 1281. DOI: https://doi.org/10.3390/rs11111281

5. Zhao J., Yan J., Zhang H., Meng J. A new radiometric correction method for side-scan sonar images in consideration of seabed sediment variation. *Remote Sensing*. 2017. Vol. 9, No. 6. Article number: 575. DOI: https://doi.org/10.3390/rs9060575

6. Jobson D., Woodell A. Retinex processing for automatic image enhancement. *Journal of Electronic Imaging*. 2004. Vol. 13. P. 100–110. Electronic resource. URL: https://www.researchgate.net/publication/220050764_Retinex_processing_for_automatic_image_enhancement (date of access: 21.07.2025).

7. Land E. The retinex theory of color vision. *Scientific American*. 1997. Vol. 237. P. 108–128. Electronic resource. URL: https://doi.org/10.1038/scientificamerican1277-108 (date of access: 21.07.2025).

8. Liu Y., Ye X. A gray scale correction method for side-scan sonar images considering rugged seafloor. *IEEE Transactions on Geoscience and Remote Sensing*. 2023. Vol. 61. P. 1–10. Electronic resource. URL: https://doi.org/10.1109/TGRS.2023.3334492 (date of access: 21.07.2025).

9. Li S., Zhao J., Yu Y., Wu Y., Bian S., Zhai G. Anisotropic total variation regularized low-rank approximation for SSS images radiometric distortion correction. *IEEE Transactions on Geoscience and Remote Sensing*. 2022. Electronic resource. URL: https://doi.org/10.1109/TGRS.2022.3229301 (date of access: 21.07.2025).

10. Al-Rawi M. Intensity normalization of sidescan sonar imagery. In: Image Processing Theory, Tools and Applications (IPTA), 2016. P. 1–6. Electronic resource. URL: https://doi.org/10.1109/IPTA.2016.7820967 (date of access: 21.07.2025).

11. Lurton X. An introduction to underwater acoustics: principles and applications. *Noise Control Engineering Journal*. 2011. Vol. 59, No. 1. P. 106. Electronic resource. URL: https://link.springer.com/book/9783540784807 (date of access: 21.07.2025).

12. Burguera A., Oliver G. Intensity correction of side-scan sonar images. In: IEEE Emerging Technology and Factory Automation (ETFA), 2014. P. 1–4. Electronic resource. URL: https://doi.org/10.1109/ETFA.2014.7005092 (date of access: 21.07.2025).

13. Capus C., Banks A., Coiras E., Ruiz I., Smith C., Petillot Y. Data correction for visualisation and classification of sidescan sonar imagery. *IET Radar, Sonar & Navigation*. 2008. Vol. 2, No. 3. P. 155–169. Electronic resource. URL: https://doi.org/10.1049/iet-rsn:20070032 (date of access: 21.07.2025).

14. Clarke J. Seafloor characterization using keel-mounted sidescan: Proper compensation for radiometric and geometric distortion. Electronic resource. URL: https://www.semanticscholar.org/paper/Seafloor-characterization-using-keel-mounted-proper-Clarke/863bec2b1127c2bb74e832150b8aed0671fd1f5b (date of access: 21.07.2025).

15. Tamsett D., Hogarth P. Sidescan sonar beam function and seabed backscatter functions from trace amplitude and vehicle roll data. *IEEE Journal of Oceanic Engineering*. 2016. Vol. 41, No. 1. P. 155–163. Electronic resource. URL: https://doi.org/10.1109/JOE.2015.2390732 (date of access: 21.07.2025).

16. Al-Rawi M. Cubic spline regression based enhancement of side-scan sonar imagery. In: OCEANS 2017 – Aberdeen. 2017. P. 1–7. Electronic resource. URL: https://doi.org/10.1109/OCEANSE.2017.8084567 (date of access: 21.07.2025).

17. Calder B., Linnett L., Carmichael D. Bayesian approach to object detection in sidescan sonar. In: IEE Conference Publication. 1997. No. 443, Pt. 2. P. 857–861. Electronic resource. URL: https://researchportal.hw.ac.uk/en/publications/bayesian-approach-to-object-detection-in-sidescan-sonar-2 (date of access: 21.07.2025).

18. Anstee S. Removal of range-dependent artifacts from sidescan sonar imagery. Electronic resource. URL: https://apps.dtic.mil/sti/tr/pdf/ADA393168.pdf (date of access: 21.07.2025).

19. Capus C. Compensation for changing beam pattern and residual TVG effects with sonar altitude variation for sidescan mosaicing and classification. Electronic resource. URL: https://www.researchgate.net/publication/229012920 (date of access: 21.07.2025).

20. Shippey G., Bolinder A., Finndin R. Shade correction of side-scan sonar imagery by histogram transformation. In: OCEANS'94. 1994. P. II/439–II/443. Electronic resource. URL: https://doi.org/10.1109/OCEANS.1994.364084 (date of access: 21.07.2025).

21. Wilken D., Feldens P., Wunderlich T., Heinrich C. Application of 2D Fourier filtering for elimination of stripe noise in side-scan sonar mosaics. *Geo-Marine Letters*. 2012. Vol. 32, No. 4. P. 337–347. Electronic resource. URL: https://doi.org/10.1007/s00367-012-0293-z (date of access: 21.07.2025).

22. Cervenka P., de Moustier C. Sidescan sonar image processing techniques. *IEEE Journal of Oceanic Engineering*. 1993. Vol. 18, No. 2. P. 108–122. Electronic resource. URL: https://doi.org/10.1109/48.219531 (date of access: 21.07.2025).

23. Burguera A., Oliver G. High-resolution underwater mapping using side-scan sonar. *PLoS ONE*. 2016. Vol. 11, No. 1. Article ID: e0146396. Electronic resource. URL: https://doi.org/10.1371/journal.pone.0146396 (date of access: 21.07.2025).

24. Chang Y., Hsu S., Tsai C. Sidescan sonar image processing: Correcting brightness variation and patching gaps. *Journal of Marine Science and Technology*. 2010. Vol. 18, No. 6. Electronic resource. URL: https://doi.org/10.51400/2709-6998.1935 (date of access: 21.07.2025).

25. Galdran A. An efficient non-uniformity correction technique for side-scan sonar imagery. In: OCEANS 2017 – Aberdeen. 2017. P. 1–6. Electronic resource. URL: https://doi.org/10.1109/OCEANSE.2017.8084577 (date of access: 21.07.2025).

26. Nguyen V., Luu N., Nguyen Q., Nguyen T. Estimation of the acoustic transducer beam aperture by using the geometric backscattering model for side-scan sonar systems. *Sensors*. 2023. Vol. 23, No. 4. Article ID: 2190. Electronic resource. URL: https://doi.org/10.3390/s23042190 (date of access: 21.07.2025).

27. Tamsett D. Geometrical spreading correction in sidescan sonar seabed imaging. *Journal of Marine Science and Engineering*. 2017. Vol. 5, No. 4. Article ID: 54. Electronic resource. URL: https://doi.org/10.3390/jmse5040054 (date of access: 21.07.2025).

28. Wang Z., Bovik A., Sheikh H., Simoncelli E. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*. 2004. Vol. 13, No. 4. P. 600–612. Electronic resource. URL: https://doi.org/10.1109/TIP.2003.819861 (date of access: 21.07.2025).

29. Sivachandra K., Kumudham R. A review: Object detection and classification using side scan sonar images via deep learning techniques. In: Modern Approaches in Machine Learning and Cognitive Science: A Walkthrough. Vol. 4. P. 229–249. Electronic resource. URL: https://doi.org/10.1007/978-3-031-43009-1_20 (date of access: 21.07.2025).

30. Steiniger Y., Kraus D., Meisen T. Survey on deep learning based computer vision for sonar imagery. *Engineering Applications of Artificial Intelligence*. 2022. Vol. 114. Article ID: 105157. URL: https://doi.org/10.1016/j.engappai.2022.105157 (date of access: 21.07.2025).

References

1. Blondel, P. (2010). *The handbook of sidescan sonar*. Berlin: Springer Science & Business Media. Electronic resource. URL: https://link.springer.com/book/10.1007/978-3-540-49886-5 (date of access: 21.07.2025).

2. Grządziel, A. (2023). The impact of side-scan sonar resolution and acoustic shadow phenomenon on the quality of sonar imagery and data interpretation capabilities. *Remote Sensing*, 15(23), Article 5599. https://doi.org/10.3390/rs15235599

3. Zhang, Y., Li, H., Zhu, J., Zhou, L., & Chen, B. (2021). Contrast study of side scan sonar image enhancement methods. In *2021 OES China Ocean Acoustics (COA)* (pp. 995–999). IEEE. https://doi.org/10.1109/COA50123.2021.9520059

4. Ye, X., Yang, H., Li, C., Jia, Y., & Li, P. (2019). A gray scale correction method for side-scan sonar images based on Retinex. *Remote Sensing*, 11(11), Article 1281. https://doi.org/10.3390/rs11111281

5. Zhao, J., Yan, J., Zhang, H., & Meng, J. (2017). A new radiometric correction method for side-scan sonar images in consideration of seabed sediment variation. *Remote Sensing*, 9(6), Article 575. https://doi.org/10.3390/rs9060575

6. Jobson, D., & Woodell, A. (2004). Retinex processing for automatic image enhancement. *Journal of Electronic Imaging*, 13, 100–110. Electronic resource. URL: https://www.researchgate.net/publication/220050764_Retinex_processing_for_automatic_image_enhancement (date of access: 21.07.2025).

7. Land, E. (1997). The retinex theory of color vision. *Scientific American*, 237, 108–128. Electronic resource. URL: https://doi.org/10.1038/scientificamerican1277-108 (date of access: 21.07.2025).

8. Liu, Y., & Ye, X. (2023). A gray scale correction method for side-scan sonar images considering rugged seafloor. *IEEE Transactions on Geoscience and Remote Sensing*, 61, 1–10. Electronic resource. URL: https://doi.org/10.1109/TGRS.2023.3334492 (date of access: 21.07.2025).

9. Li, S., Zhao, J., Yu, Y., Wu, Y., Bian, S., & Zhai, G. (2022). Anisotropic total variation regularized low-rank approximation for SSS images radiometric distortion correction. *IEEE Transactions on Geoscience and Remote Sensing*. Electronic resource. URL: https://doi.org/10.1109/TGRS.2022.3229301 (date of access: 21.07.2025).

10. Al-Rawi, M. (2016). Intensity normalization of sidescan sonar imagery. In *Image Processing Theory, Tools and Applications (IPTA)* (pp. 1–6). Electronic resource. URL: https://doi.org/10.1109/IPTA.2016.7820967 (date of access: 21.07.2025).

11. Lurton, X. (2011). An introduction to underwater acoustics: Principles and applications. *Noise Control Engineering Journal*, 59(1), 106. Electronic resource. URL: https://link.springer.com/book/9783540784807 (date of access: 21.07.2025).

12. Burguera, A., & Oliver, G. (2014). Intensity correction of side-scan sonar images. In *IEEE Emerging Technology and Factory Automation (ETFA)* (pp. 1–4). Electronic resource. URL: https://doi.org/10.1109/ETFA.2014.7005092 (date of access: 21.07.2025).

13. Capus, C., Banks, A., Coiras, E., Ruiz, I., Smith, C., & Petillot, Y. (2008). Data correction for visualisation and classification of sidescan sonar imagery. *IET Radar, Sonar & Navigation*, 2(3), 155–169. Electronic resource. URL: https://doi.org/10.1049/iet-rsn:20070032 (date of access: 21.07.2025).

14. Clarke, J. (n.d.). Seafloor characterization using keel-mounted sidescan: Proper compensation for radiometric and geometric distortion. Electronic resource. URL: https://www.semanticscholar.org/paper/Seafloor-characterization-using-keel-mounted-proper-Clarke/863bec2b1127c2bb74e832150b8aed0671fd1f5b (date of access: 21.07.2025).

15. Tamsett, D., & Hogarth, P. (2016). Sidescan sonar beam function and seabed backscatter functions from trace amplitude and vehicle roll data. *IEEE Journal of Oceanic Engineering*, 41(1), 155–163. Electronic resource. URL: https://doi.org/10.1109/JOE.2015.2390732 (date of access: 21.07.2025).

16. Al-Rawi, M. (2017). Cubic spline regression based enhancement of side-scan sonar imagery. In *OCEANS 2017 – Aberdeen* (pp. 1–7). Electronic resource. URL: https://doi.org/10.1109/OCEANSE.2017.8084567 (date of access: 21.07.2025).

17. Calder, B., Linnett, L., & Carmichael, D. (1997). Bayesian approach to object detection in sidescan sonar. In *IEE Conference Publication* (No. 443, Pt. 2, pp. 857–861). Electronic resource. URL: https://researchportal.hw.ac.uk/en/publications/bayesian-approach-to-object-detection-in-sidescan-sonar-2 (date of access: 21.07.2025).

18. Anstee, S. (n.d.). Removal of range-dependent artifacts from sidescan sonar imagery. Electronic resource. URL: https://apps.dtic.mil/sti/tr/pdf/ADA393168.pdf (date of access: 21.07.2025).

19. Capus, C. (n.d.). Compensation for changing beam pattern and residual TVG effects with sonar altitude variation for sidescan mosaicing and classification. Electronic resource. URL: https://www.researchgate.net/publication/229012920 (date of access: 21.07.2025).

20. Shippey, G., Bolinder, A., & Finndin, R. (1994). Shade correction of side-scan sonar imagery by histogram transformation. In *OCEANS'94* (pp. II/439–II/443). Electronic resource. URL: https://doi.org/10.1109/OCEANS.1994.364084 (date of access: 21.07.2025).

21. Wilken, D., Feldens, P., Wunderlich, T., & Heinrich, C. (2012). Application of 2D Fourier filtering for elimination of stripe noise in side-scan sonar mosaics. *Geo-Marine Letters*, 32(4), 337–347. Electronic resource. URL: https://doi.org/10.1007/s00367-012-0293-z (date of access: 21.07.2025).

22. Cervenka, P., & de Moustier, C. (1993). Sidescan sonar image processing techniques. *IEEE Journal of Oceanic Engineering*, 18(2), 108–122. Electronic resource. URL: https://doi.org/10.1109/48.219531 (date of access: 21.07.2025).

23. Burguera, A., & Oliver, G. (2016). High-resolution underwater mapping using side-scan sonar. *PLoS ONE*, 11(1), Article e0146396. Electronic resource. URL: https://doi.org/10.1371/journal.pone.0146396 (date of access: 21.07.2025).

24. Chang, Y., Hsu, S., & Tsai, C. (2010). Sidescan sonar image processing: Correcting brightness variation and patching gaps. *Journal of Marine Science and Technology*, 18(6). Electronic resource. URL: https://doi.org/10.51400/2709-6998.1935 (date of access: 21.07.2025).

25. Galdran, A. (2017). An efficient non-uniformity correction technique for side-scan sonar imagery. In *OCEANS 2017 – Aberdeen* (pp. 1–6). Electronic resource. URL: https://doi.org/10.1109/OCEANSE.2017.8084577 (date of access: 21.07.2025).

26. Nguyen, V., Luu, N., Nguyen, Q., & Nguyen, T. (2023). Estimation of the acoustic transducer beam aperture by using the geometric backscattering model for side-scan sonar systems. *Sensors*, 23(4), Article 2190. Electronic resource. URL: https://doi.org/10.3390/s23042190 (date of access: 21.07.2025).

27. Tamsett, D. (2017). Geometrical spreading correction in sidescan sonar seabed imaging. *Journal of Marine Science and Engineering*, 5(4), Article 54. Electronic resource. URL: https://doi.org/10.3390/jmse5040054 (date of access: 21.07.2025).

28. Wang, Z., Bovik, A., Sheikh, H., & Simoncelli, E. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600–612. Electronic resource. URL: https://doi.org/10.1109/TIP.2003.819861 (date of access: 21.07.2025).

29. Sivachandra, K., & Kumudham, R. (2023). A review: Object detection and classification using side scan sonar images via deep learning techniques. In *Modern Approaches in Machine Learning and Cognitive Science: A Walkthrough* (Vol. 4, pp. 229–249). Electronic resource. URL: https://doi.org/10.1007/978-3-031-43009-1_20 (date of access: 21.07.2025).

30. Steiniger, Y., Kraus, D., & Meisen, T. (2022). Survey on deep learning based computer vision for sonar imagery. *Engineering Applications of Artificial Intelligence*, 114, Article 105157. Electronic resource. URL: https://doi.org/10.1016/j.engappai.2022.105157 (date of access: 21.07.2025).

Дата першого надходження рукопису до видання: 30.09.2025
Дата прийнятого до друку рукопису після рецензування: 27.10.2025
Дата публікації:

УДК 004.738.5:004.451 DOI

В. Ю. КОЗУБ
доктор філософії з комп'ютерних наук,
асистент кафедри математики та інформатики
ДЗ «Луганський національний університет імені Тараса Шевченка»
ORCID: 0000-0003-2710-7206

М. А. СЕМЕНОВ
кандидат педагогічних наук, доцент,
завідувач кафедри інформаційних технологій та систем
ДЗ «Луганський національний університет імені Тараса Шевченка»
ORCID: 0000-0003-4989-8109

С. І. ОСАДЧУК
викладач кафедри комп'ютерних систем і технологій
Приватний вищий навчальний заклад «Буковинський університет»
ORCID: 0009-0000-0203-2581

АНАЛІЗ ВПЛИВУ WEBASSEMBLY НА ПРОДУКТИВНІСТЬ СКЛАДНИХ ВЕБЗАСТОСУНКІВ

Мета статті. Мета дослідження полягає у комплексному аналізі впливу технології WebAssembly на продуктивність складних вебзастосунків, визначенні ключових факторів підвищення ефективності обчислень та оцінці перспектив практичного застосування у сучасній веброзробці для оптимізації користувацького досвіду.

Наукова новизна. Наукова новизна роботи полягає в систематизації методів оцінки впливу WebAssembly на продуктивність складних вебзастосунків, розробці комплексного підходу до аналізу ефективності бінарних інструкцій у веб середовищі. Вперше проведено порівняльний аналіз різних архітектурних рішень інтеграції WebAssembly з традиційними JavaScript-рішеннями. Запропоновано методологію оцінки продуктивності для різних типів обчислювальних задач у вебзастосунок з використанням метрик швидкодії та оптимізації пам'яті.

Результати. Результати дослідження демонструють значне підвищення продуктивності складних вебзастосунків при використанні WebAssembly порівняно з традиційними JavaScript-рішеннями. Встановлено, що WebAssembly забезпечує прискорення виконання обчислювально-інтенсивних операцій у 2-10 разів залежно від типу задач. Виявлено оптимальні сценарії застосування технології для математичних обчислень, обробки зображень, криптографічних операцій та мультимедійних застосунків. Проаналізовано вплив різних компіляторів та оптимізації на підсумкову продуктивність. Визначено критичні аспекти інтероперабельності між WebAssembly та JavaScript, що впливають на загальну ефективність системи. Розроблено рекомендації щодо архітектурних рішень для максимізації переваг WebAssembly у складних вебзастосунках.

Висновки. WebAssembly демонструє високий потенціал для оптимізації продуктивності складних вебзастосунків, особливо у сферах з інтенсивними обчисленнями. Технологія забезпечує значне прискорення виконання критично важливих операцій при збереженні безпеки та кросплатформності. Оптимальне застосування WebAssembly вимагає ретельного аналізу архітектури застосунку та специфіки обчислювальних задач для досягнення максимального ефекту.

Ключові слова: *WebAssembly, продуктивність, веб-застосунки, веб-технології, оптимізація, ядра.*

V. YU. KOZUB
PhD in Computer Science,
Assistant at the Department of Information Technology and Systems
Luhansk Taras Shevchenko National University
ORCID: 0000-0003-2710-7206

M. A. SEMENOV
PhD in Pedagogy, Associate Professor,
Head of the Department of Information Technologies Educational and Scientific
Luhansk Taras Shevchenko National University
ORCID: 0000-0003-4989-8109

© Козуб В. Ю., Семенов М. А., Осадчук С. І., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

S. I. OSADCHUK
Lecturer at the Department of Computer Systems and Technologies
Private Higher Educational Institution "Bukovinian University"
ORCID: 0009-0000-0203-2581

ANALYSIS OF THE IMPACT OF WEBASSEMBLY ON THE PRODUCTIVITY OF COMPLEX WEB APPLICATIONS

The purpose of the article. The purpose of the study is to comprehensively analyze the impact of WebAssembly technology on the performance of complex web applications, identify key factors for increasing computing efficiency, and assess the prospects for practical application in modern web development to optimize the user experience.

Scientific novelty. The scientific novelty of the work lies in the systematization of methods for assessing the impact of WebAssembly on the performance of complex web applications, the development of a comprehensive approach to analyzing the efficiency of binary instructions in the web environment. For the first time, a comparative analysis of various architectural solutions for integrating WebAssembly with traditional JavaScript solutions has been carried out. A methodology for assessing performance for various types of computational tasks in web applications using performance metrics and memory optimization has been proposed.

Results. The results of the study demonstrate a significant increase in the performance of complex web applications when using WebAssembly compared to traditional JavaScript solutions. It was found that WebAssembly provides an acceleration of computationally intensive operations by 2–10 times, depending on the type of tasks. Optimal scenarios for using the technology for mathematical calculations, image processing, cryptographic operations, and multimedia applications were identified. The impact of different compilers and optimizations on the final performance was analyzed. Critical aspects of interoperability between WebAssembly and JavaScript that affect the overall efficiency of the system were identified. Recommendations for architectural solutions were developed to maximize the benefits of WebAssembly in complex web applications.

Conclusions. WebAssembly shows great potential for optimizing the performance of complex web applications, especially in compute-intensive areas. The technology provides significant acceleration of critical operations while maintaining security and cross-platform compatibility. Optimal use of WebAssembly requires careful analysis of the application architecture and the specifics of the computing tasks to achieve maximum effect.

Key words: *WebAssembly, performance, web applications, web technologies, optimization, kernels.*

Постановка проблеми

Сучасні вебзастосунки стають дедалі складнішими та вимагають виконання обчислювально-інтенсивних операцій безпосередньо у браузері користувача. Традиційні підходи з використанням JavaScript часто не забезпечують достатньої продуктивності для складних задач обробки даних, математичних обчислень, мультимедійного контенту та ігрових додатків. Обмеження інтерпретованої природи JavaScript створюють бар'єри для розробки високопродуктивних вебзастосунків, що можуть конкурувати з нативними рішеннями. Поява WebAssembly як бінарного формату інструкцій для вебплатформи відкриває нові можливості для значного підвищення продуктивності, проте потребує ґрунтовного дослідження ефективності та оптимальних сценаріїв застосування у контексті складних вебсистем.

Аналіз останніх досліджень і публікацій

Аналіз сучасної літератури демонструє активний розвиток досліджень у сфері WebAssembly та його впливу на продуктивність вебзастосунків. Sipek M., Muharemagic D., Mihaljevic B., Radovan A. [1, с. 1695] досліджували застосування WebAssembly у контексті наступного покоління вебзастосунків з використанням TruffleWasm, що демонструє перспективи інтеграції з сучасними технологічними стеками. Harnes H., Morrison D. [2, с. 1] представили систематизований огляд методів аналізу WebAssembly, що є фундаментальним для розуміння підходів до оцінки продуктивності. Akieva Z. M., Murzin D. A., Eltaev S. I. [3, с. 1] сфокусувалися на практичному застосуванні WebAssembly для покращення продуктивності інтерактивних вебзастосунків. Robati Shirzad M., Lam P. [4, с. 2] розробили інноваційний підхід WasmWalker для аналізу програм WebAssembly на основі шляхових представлень коду. Wang W. [5, с. 1] провів комплексну оцінку готовності вебзастосунків до повноцінного використання WebAssembly. Burkhart N., Liao W., Guzide O. [6, с. 1] надали фундаментальний огляд архітектури WebAssembly. Rao N. S. [7, с. 1973] дослідив революційний вплив WebAssembly на розробку користувацьких інтерфейсів через призму продуктивності та міжмовної інтеграції. Long Y., Su Y., Jiang Z. [8, с. 1] розробили спеціалізований інструмент WACP для профілювання продуктивності взаємодії WebAssembly-Python. Waseem M., Das T., Ahmad A., Liang P., Mikkonen T. [9, с. 1] провели емпіричне дослідження проблем та їх причин у WebAssembly-застосунках. Cabrera-Arteaga J., Fitzgerald N., Monperrus M., Baudry B. [10, с. 1] представили WASM-MUTATE для швидкої та ефективної бінарної диверсифікації WebAssembly.

Формулювання мети дослідження

Метою дослідження є комплексний аналіз впливу WebAssembly на продуктивність складних вебзастосунків через систематизацію методів оцінки ефективності, визначення оптимальних архітектурних рішень та розробку практичних рекомендацій для максимізації переваг технології у реальних проєктах.

Викладення основного матеріалу дослідження

Дослідження продуктивності WebAssembly у складних вебзастосунках базується на комплексному підході до аналізу різних аспектів технології. Sipek M., Muharemagic D., Mihaljevic B., Radovan A. [1, с. 1698] демонструють, що інтеграція WebAssembly з TruffleWasm забезпечує значне підвищення продуктивності для обчислювально-інтенсивних задач, особливо у контексті полімовних застосунків.

Для системного розуміння факторів, що впливають на продуктивність WebAssembly у складних вебзастосунках, необхідно проаналізувати ключові параметри та їх взаємозв'язки. Дані параметри представлено в таблиці 1.

Таблиця 1

Порівняльний аналіз параметрів продуктивності WebAssembly

Параметр	JavaScript	WebAssembly	Приріст продуктивності	Тип застосунку
Швидкість виконання обчислень	Базова	2–10х швидше	200–1000 %	Математичні
Обробка зображень	Повільна	3–8х швидше	300–800 %	Мультимедійні
Криптографічні операції	Обмежена	5–15х швидше	500–1500 %	Безпека
Витрати пам'яті	Високі	Оптимізовані	30–60 % менше	Ігрові
Час завантаження	Швидкий	Повільніший	10–30 % більше	Всі типи

Джерело: розроблено на основі [1, 2, 3, 6]

Аналіз даних таблиці 1 демонструє суттєві переваги WebAssembly у продуктивності обчислень при незначному збільшенні часу завантаження, що підтверджує доцільність застосування технології для складних вебзастосунках. Harnes H., Morrison D. [2, с. 15] підкреслюють важливість методів статичного та динамічного аналізу для оцінки продуктивності WebAssembly. Їх дослідження показує, що комбінування різних підходів до аналізу дозволяє точніше оцінити реальну ефективність застосунків.

Для візуалізації архітектурного впливу WebAssembly на продуктивність складних вебзастосунків розроблено структурну модель, представлену на рисунку 1.

Структурна модель на рисунку 1 ілюструє ієрархічну взаємодію компонентів системи, де WebAssembly відповідає за обчислювально-інтенсивні операції, а JavaScript керує логікою користувацького інтерфейсу, що забезпечує оптимальний розподіл навантаження. Akieva Z. M., Murzin D. A., Eltaev S. I. [3, с. 3] підтверджують, що застосування WebAssembly для інтерактивних вебзастосунків дозволяє досягти значного покращення відгуку системи. Robati Shirzad M., Lam P. [4, с. 5] розробили інноваційний підхід до аналізу WebAssembly програм, який дозволяє ефективніше оптимізувати код для конкретних задач. Wang W. [5, с. 2] досліджував питання готовності сучасних вебзастосунків до впровадження WebAssembly та виявив, що основні перешкоди пов'язані з інтеграційними аспектами та інструментарієм розробки.

Burkhart N., Liao W.,Guzide O. [6, с. 3] надали детальний огляд архітектури WebAssembly, що є важливим для розуміння принципів оптимізації продуктивності. Rao N. S. [7, с. 1978] акцентує на революційному потенціалі WebAssembly для розробки користувацьких інтерфейсів, особливо у контексті кросплатформенної інтеграції. Long Y., Su Y., Jiang Z. [8, с. 2] розробили спеціалізований інструмент WACP для профілювання продуктивності, який демонструє практичні підходи до оцінки ефективності WebAssembly у реальних умовах.

Waseem M., Das T., Ahmad A., Liang P., Mikkonen T. [9, с. 4] провели емпіричне дослідження проблем WebAssembly-застосунків, що виявило основні виклики: складність налагодження, обмежений інструментарій та проблеми з керуванням пам'яттю. Ці результати критично важливі для розробки стратегій оптимізації продуктивності.

Cabrera-Arteaga J., Fitzgerald N., Monperrus M., Baudry B. [10, с. 3] представили WASM-MUTATE, який забезпечує швидку бінарну диверсифікацію WebAssembly, що має значення для оптимізації безпеки без втрати продуктивності.

Практичне застосування WebAssembly у складних вебзастосунках демонструє найбільшу ефективність у сценаріях з інтенсивними обчисленнями: обробка зображень та відео, математичне моделювання, криптографічні операції, ігрові движки та симуляції. Ключовими факторами успішної інтеграції є правильний розподіл функціональності між JavaScript та WebAssembly, оптимізація передачі даних між модулями та вибір відповідних компіляторів.

Висновки

Проведене дослідження підтверджує значний потенціал WebAssembly для оптимізації продуктивності складних вебзастосунків. Технологія забезпечує прискорення виконання обчислювально-інтенсивних операцій у 2–15 разів порівняно з традиційними JavaScript-рішеннями залежно від типу задач. Найбільша ефективність досягається у математичних обчисленнях, обробці мультимедіа та криптографічних операціях. Ключовими факторами успішного впровадження WebAssembly є оптимальний розподіл функціональності між компонентами системи, мінімізація накладних витрат на між модульну взаємодію та правильний вибір архітектурних рішень.

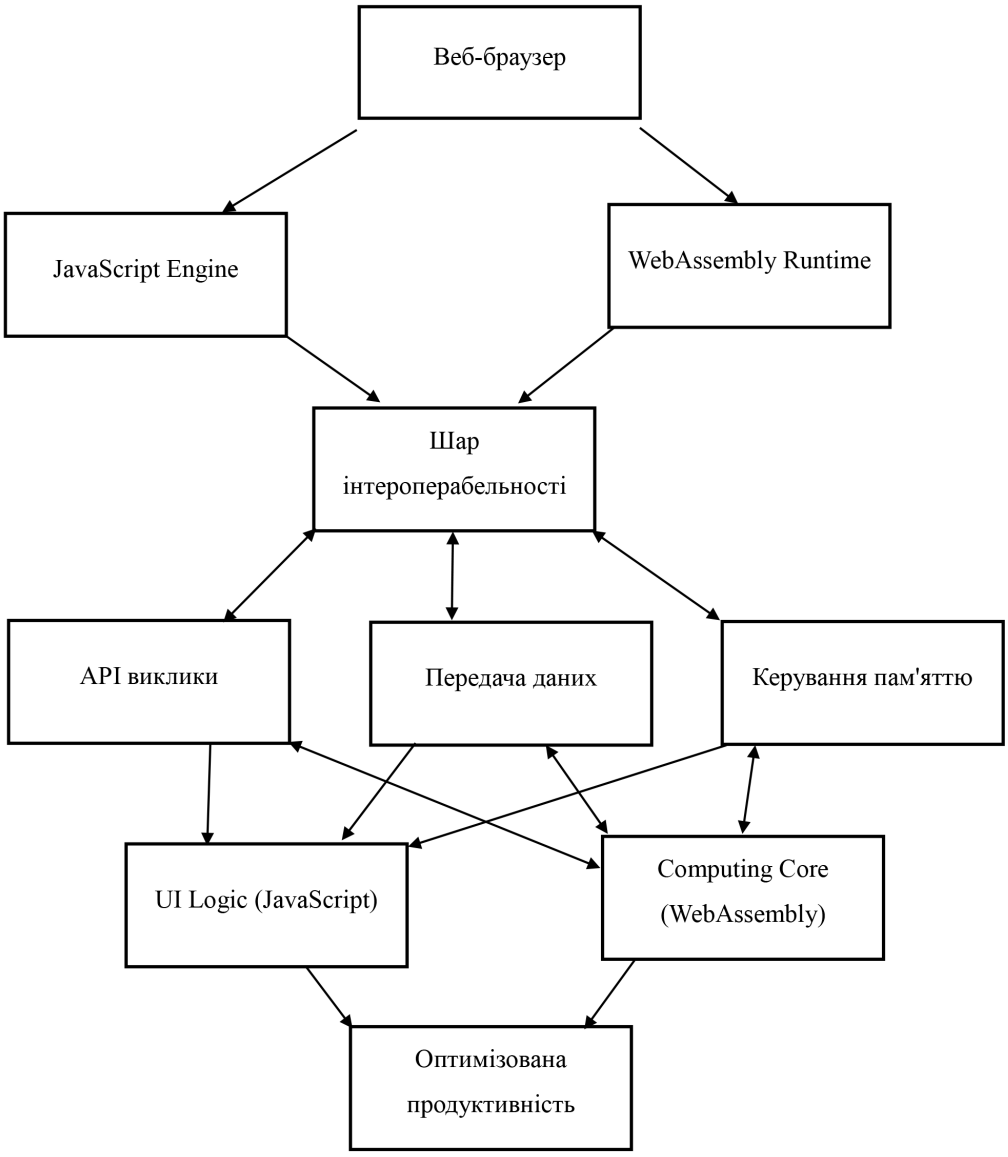


Рис. 1. Структурна модель впливу WebAssembly на продуктивність веб-застосунків

Джерело: авторська розробка

Виявлено, що WebAssembly найефективніше працює у гібридних архітектурах, де JavaScript відповідає за логіку користувацького інтерфейсу, а WebAssembly – за обчислювальне ядро.

Рекомендації для практичного застосування включають ретельний аналіз профілю навантаження застосунку, використання спеціалізованих інструментів профілювання продуктивності та поетапний підхід до міграції критично важливих компонентів. WebAssembly демонструє високу перспективність для майбутнього розвитку високопродуктивних вебзастосунків, що можуть конкурувати з нативними рішеннями за ефективністю при збереженні переваг вебплатформа.

Список використаної літератури

1. Sipek M., Muharemagic D., Mihaljevic B., Radovan A. Next-generation Web Applications with WebAssembly and TruffleWasm. *Proceedings of the 44th International Convention on Information, Communication and Electronic Technology (MIPRO)*. 2021. P. 1695–1700. DOI: https://doi.org/10.23919/MIPRO52101.2021.9596883 URL: https://ieeexplore.ieee.org/document/9596883

2. Harnes H., Morrison D. SoK: Analysis techniques for WebAssembly. *Future Internet*. 2024. Vol. 16, iss. 3. P. 1–27. DOI: https://doi.org/10.3390/fi16030084 URL: https://www.mdpi.com/1999-5903/16/3/84

3. Akieva Z. M., Murzin D. A., Eltaev S. I. Using WebAssembly to Improve the Performance of Interactive Web Applications. *Economy: Yesterday, Today, Tomorrow*. 2024. DOI: 10.36871/ek.up.p.r.2024.12.07.021 URL: https://is.gd/sKPlhB

4. Robati Shirzad M., Lam P. WasmWalker: Path-based Code Representations for Improved WebAssembly Program Analysis. arXiv. 2024. arXiv:2410.08517. DOI: https://doi.org/10.48550/arXiv.2410.08517 URL: https://arxiv.org/abs/2410.08517

5. Wang W. Empowering Web Applications with WebAssembly: Are We There Yet?. *Proceedings of the 36th IEEE/ACM International Conference on Automated Software Engineering*. 2021. DOI: https://doi.org/10.1109/ASE51524.2021.9678831 URL: https://is.gd/fUqdDA

6. Burkhart N., Liao W., Guzide O. An Overview of WebAssembly. *Proceedings of the West Virginia Academy of Science*. 2020. Vol. 92, iss. 1. DOI: https://doi.org/10.55632/pwvas.v92i1.682 URL: https://is.gd/R9Omvk

7. Rao N. S. WebAssembly: Revolutionizing Web User Interface Development through Performance and Cross-Language Integration. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*. 2024. Vol. 10, iss. 6. P. 1973–1981. DOI: https://doi.org/10.32628/CSEIT241061235 URL: https://is.gd/PNg9lo

8. Long Y., Su Y., Jiang Z. WACP: A Performance Profiling Tool for WebAssembly-Python Interoperability. *Proceedings of the 15th Asia-Pacific Symposium on Internetware*. 2024. DOI: https://doi.org/10.1145/3671016.3671408 URL: https://dl.acm.org/doi/10.1145/3671016.3671408

9. Waseem M., Das T., Ahmad A., Liang P., Mikkonen T. Issues and Their Causes in WebAssembly Applications: An Empirical Study. *Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering*. 2024. DOI: https://doi.org/10.48550/arXiv.2311.00646 URL: https://arxiv.org/abs/2311.00646

10. Cabrera-Arteaga J., Fitzgerald N., Monperrus M., Baudry B. WASM-MUTATE: Fast and Effective Binary Diversification for WebAssembly. *Computers & Security*. 2024. Vol. 139. DOI: https://doi.org/10.1016/j.cose.2024.103731 URL: https://is.gd/zmwKsG

11. Козуб Г. О., Сурма Ю. Ю., Артеменко О. І. Роль вебкомпонентів у побудові сучасних інтерфейсів: переваги та обмеження. *Вісник Херсонського національного технічного університету*. Том 2 № 2(93), 2025. С. 175–180. https://doi.org/10.35546/kntu2078-4481.2025.2.2.21

12. Вередін М. О., Козуб Г. О. Особливості UX/UI-дизайну при розробці вебсайтів та мобільних застосунків. збірник наукових праць з матеріалами IX Міжнародної наукової конференції, м.Ужгород, 4липня, 2025р. Міжнародний центр наукових досліджень. Вінниця: ТОВ «УКРЛОГОС Груп, 2025. С. 122-126. https://doi.org/10.62731/mcnd-04.07.2025.005

References

1. Sipek, M., Muharemagic, D., Mihaljevic, B., & Radovan, A. (2021) Next-generation web applications with WebAssembly and TruffleWasm. *Proceedings of the 44th International Convention on Information, Communication and Electronic Technology (MIPRO)*, 1695–1700. DOI: https://doi.org/10.23919/MIPRO52101.2021.9596883 Retrieved from https://ieeexplore.ieee.org/document/9596883 [in English].

2. Harnes, H., & Morrison, D. (2024) SoK: Analysis techniques for WebAssembly. *Future Internet*, 16(3), 1–27. DOI: https://doi.org/10.3390/fi16030084 Retrieved from https://www.mdpi.com/1999-5903/16/3/84 [in English].

3. Akieva, Z. M., Murzin, D. A., & Eltaev, S. I. (2024) Using WebAssembly to improve the performance of interactive web applications. *Economy: Yesterday, Today, Tomorrow*. DOI: 10.36871/ek.up.p.r.2024.12.07.021 Retrieved from https://is.gd/sKPlhB [in English].

4. Robati Shirzad, M., & Lam, P. (2024) WasmWalker: Path-based code representations for improved WebAssembly program analysis. arXiv. DOI: https://doi.org/10.48550/arXiv.2410.08517 Retrieved from https://arxiv.org/abs/2410.08517 [in English].

5. Wang, W. (2021) Empowering web applications with WebAssembly: Are we there yet? *Proceedings of the 36th IEEE/ACM International Conference on Automated Software Engineering*. DOI: https://doi.org/10.1109/ASE51524.2021.9678831 Retrieved from https://ieeexplore.ieee.org/document/9678831 [in English].

6. Burkhart, N., Liao, W., & Guzide, O. (2020) An overview of WebAssembly. *Proceedings of the West Virginia Academy of Science*, 92(1). DOI: https://doi.org/10.55632/pwvas.v92i1.682 Retrieved from https://pwvas.org/index.php/pwvas/article/view/682 [in English].

7. Rao, N. S. (2024) WebAssembly: Revolutionizing web user interface development through performance and cross-language integration. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, 10(6), 1973–1981. DOI: https://doi.org/10.32628/CSEIT241061235 Retrieved from https://is.gd/PNg9lo [in English].

8. Long, Y., Su, Y., & Jiang, Z. (2024) WACP: A performance profiling tool for WebAssembly-Python interoperability. *Proceedings of the 15th Asia-Pacific Symposium on Internetware*. DOI: https://doi.org/10.1145/3671016.3671408 Retrieved from https://dl.acm.org/doi/10.1145/3671016.3671408 [in English].

9. Waseem, M., Das, T., Ahmad, A., Liang, P., & Mikkonen, T. (2024) Issues and their causes in WebAssembly applications: An empirical study. *Proceedings of the 28th International Conference on Evaluation and Assessment in Software Engineering*. DOI: https://doi.org/10.48550/arXiv.2311.00646 Retrieved from https://arxiv.org/abs/2311.00646 [in English].

10. Cabrera-Arteaga, J., Fitzgerald, N., Monperrus, M., & Baudry, B. (2024) WASM-MUTATE: Fast and effective binary diversification for WebAssembly. *Computers & Security*, 139. DOI: https://doi.org/10.1016/j.cose.2024.103731 Retrieved from https://is.gd/zmwKsG [in English].

11. Kozub, G.O., Surma, Yu. Yu., Artemenko, O.I. (2025) Rol′vebkomponentiv u pobudovi suchasnykh interfeysiv: perevahy ta obmezhennya [The role of web components in building modern interfaces: advantages and limitations]. *Bulletin of the Kherson National Technical University*, Volume 2 No. 2(93), P. 175–180. https://doi.org/10.35546/kntu2078-4481.2025.2.2.21 [in Ukrainian].

12. Veredin, M.O., Kozub, G.O. (2025) Osoblyvosti UX/UI-dyzaynu pryrozrobtsi vebsaytiv ta mobil′nykh zastosunkiv [Features of UX/UI design in the development of websites and mobile applications]. Collection of scientific papers with materials of the IX International Scientific Conference, Uzhgorod, July 4, 2025. International Center for Scientific Research. Vinnytsia: LLC «UKRLOGOS Group, P. 122-126. https://doi.org/10.62731/mcnd-04.07.2025.005 [in Ukrainian].

Дата першого надходження рукопису до видання: 26.09.2025
Дата прийнятого до друку рукопису після рецензування: 21.10.2025
Дата публікації:

УДК 004.42:004.051 DOI

Н. О. КОМЛЕВА
кандидат технічних наук, доцент,
завідувач кафедри інженерії програмного забезпечення
Національний університет «Одеська політехніка»
ORCID: 0000-0001-9627-8530

МЕТОД ТРАНСФОРМАЦІЇ ФОРМАЛІЗОВАНИХ ВИМОГ У РОЗШИРЕНІ UML-МОДЕЛІ НА ОСНОВІ ГРАМАТИЧНОГО ПІДХОДУ

У даній статті розглядається актуальне завдання трансформації формалізованих вимог у розширені UML-моделі, яке має важливе значення для підвищення ефективності процесів проєктування інтелектуальних та критичних систем. Дослідження зосереджується на поєднанні граматичного підходу з механізмами UML-профілю, що дозволяє забезпечити автоматизовану перевірку консистентності вимог, відображення властивостей невизначеності та імовірнісних характеристик, а також підвищити точність і адаптивність створюваних моделей. Актуальність теми зумовлена зростанням складності сучасних інформаційних систем, де вимоги часто залишаються нечіткими, суперечливими чи змінюваними, що ускладнює їх формалізацію та подальше відображення у традиційному UML.

Запропоновано методологію, що базується на дворівневій граматиці: атрибутна граматика використовується для узгодження вимог та побудови їх узгодженої множини, а трансформаційна граматика визначає відображення цих вимог у UML-елементи відповідно до їх атрибутів (confidence, likelihood, context, priority, resources). У статті наведено формальні правила трансформації для функціональних, нефункціональних, вимог до даних та сценарних вимог, які реалізуються через створення варіантів використання, класів, обмежень, діаграм послідовностей та діаграм станів. Особлива увага приділяється механізму призначення стереотипів, що зберігають семантику вихідних специфікацій, а також трасуванню, яке забезпечує прозорий зв'язок між вимогами та елементами UML-моделі.

Практичну придатність методу продемонстровано на прикладі інтелектуальної медичної системи моніторингу пацієнтів. Проведене дослідження засвідчило низку переваг розробленого підходу: скорочення часу на актуалізацію моделей, підвищення рівня узгодженості між вимогами та їхніми модельними представленнями, можливість відображення нечітких і ймовірнісних характеристик. Порівняння з класичним UML без профілю показало, що запропонований метод надає ширші можливості для моделювання адаптивних систем у середовищах з високим рівнем невизначеності. Разом з тим визначено обмеження: потреба у налаштуванні граматики під конкретний предметний домен і необхідність інтеграції з CASE-засобами для автоматизації трансформацій.

Подальші дослідження передбачають кілька напрямів розвитку методології: автоматизована генерація UML-діаграм у сучасних CASE-засобах; а також інтеграція підходу з SysML v2 як перспективного стандарту системної інженерії.

Ключові слова: *UML, вимоги, граматика, трансформація, інженерія програмного забезпечення, інтелектуальні системи, моделювання, профіль UML.*

N. O. KOMLEVA
PhD, Associate Professor,
Head of Software Engineering Department
Odesa Polytechnic National University
ORCID: 0000-0001-9627-8530

GRAMMAR-BASED METHOD FOR TRANSFORMING FORMALIZED REQUIREMENTS INTO EXTENDED UML MODELS

This article addresses the relevant task of transforming formalized requirements into extended UML models, which is of great importance for increasing the efficiency of designing intelligent and critical systems. The study focuses on combining a grammatical approach with UML profile mechanisms, which makes it possible to ensure automated consistency checking of requirements, representation of uncertainty and probabilistic properties, and improvement of the accuracy and adaptability of the created models. The relevance of the topic is determined by the increasing complexity of modern information systems, where requirements often remain vague, contradictory, or changing, which complicates their formalization and further representation in traditional UML.

A methodology is proposed based on a two-level grammar: attribute grammar is used for requirements reconciliation and for constructing the their consistent set, while transformation grammar defines the mapping of these requirements

© Комлева Н. О., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

into UML elements according to their attributes (confidence, likelihood, context, priority, resources). The article provides formal transformation rules for functional, non-functional, data, and scenario requirements, which are implemented through the creation of use cases, classes, constraints, sequence diagrams, and state diagrams. Special attention is paid to the mechanism of assigning stereotypes that preserve the semantics of the original specifications, as well as to tracing, which ensures transparent links between requirements and UML model elements.

The practical applicability of the method is demonstrated using the example of an intelligent medical patient monitoring system. The study revealed several advantages of the developed approach: reduction of the time required for updating models, improvement of the level of consistency between requirements and their model representations, and the ability to represent fuzzy and probabilistic properties. A comparison with classical UML without a profile showed that the proposed method provides broader opportunities for modeling adaptive systems in environments with a high level of uncertainty. At the same time, certain limitations were identified: the need to customize grammar for a specific application domain and the requirement for integration with CASE tools to automate transformations.

Further research involves several directions for the development of the methodology: automated generation of UML diagrams in modern CASE tools; as well as integration of the approach with SysML v2 as a promising system engineering standard.

Key words: *UML, requirements, grammar, transformation, software engineering, intelligent systems, modeling, UML profile.*

Постановка проблеми

У сучасних програмних системах усе більшої актуальності набуває проблема формалізації та моделювання вимог, які часто є нечіткими, ймовірнісними та змінюваними залежно від контексту. Традиційна нотація UML, орієнтована на стабільні та однозначні специфікації, виявляється недостатньою для повноцінного відображення таких характеристик. Це призводить до зростання ризику конфліктів між вимогами, ускладнює процес їхньої перевірки та збільшує витрати на підтримку моделей у динамічних середовищах.

Класичні методи інженерії вимог у подібних умовах демонструють обмежену ефективність, оскільки не враховують властивості невизначеності, багатоваріантності та залежності від зовнішніх факторів. Як наслідок, розробники змушені підтримувати паралельну документацію або вносити численні ручні корективи у модельні представлення, що негативно впливає на їхню цілісність та актуальність.

Для подолання цих обмежень доцільним є поєднання методів граматичного аналізу, які забезпечують формальне узгодження вимог, із розширеними механізмами UML-профілю, здатними відобразити нечіткість, ймовірнісність та контекстну залежність. Такий інтегрований підхід створює умови для автоматизованої трансформації формалізованих вимог у UML-моделі з розширеними стереотипами та тегованими значеннями.

Аналіз останніх досліджень і публікацій

Сучасна практика моделювання програмних систем демонструє широкий спектр підходів до розширення можливостей UML. Одним із поширених напрямів є створення UML-профілів для специфічних предметних областей. Такі профілі дають змогу розширювати стандартний апарат UML за допомогою стереотипів, тегованих значень та обмежень, що забезпечує його адаптацію до нетипових завдань [1, с. 27; 2, с. 87]. Наприклад, існують профілі для моделювання просторових даних, нечітких траєкторій або адаптивного програмного забезпечення [3, с. 2503; 4, с. 1720]. Вони демонструють гнучкість UML як метамови, проте здебільшого орієнтовані на окремі класи систем і не забезпечують комплексної підтримки невизначених та змінюваних вимог.

Окремий науковий напрям зосереджений на методах роботи з нечіткими та ймовірнісними вимогами, де застосовуються апарати штучного інтелекту, нечіткої логіки та методи машинного навчання. Зокрема, пропонуються підходи до пріоритизації вимог у випадку невизначеності, валідації часових обмежень, виражених природною мовою, або побудови самоадаптивних систем [5, с. 1; 6, с. 115]. Такі рішення дозволяють формалізувати властивості вимог, але вони здебільшого залишаються відокремленими від UML-моделювання та не інтегруються у його нотацію [7, с. 1905].

Третій напрям становлять формальні граматики та трансформаційні системи у моделюванні. Використання контекстно-вільних, атрибутних та графових граматик надає засоби для формалізації природномовних специфікацій, перевірки сумісності моделей та здійснення їх автоматичних перетворень [8, с. 1; 9, с. 185826]. Наприклад, граматики застосовуються для перевірки моделей трансляції, для побудови графових перетворень UML-діаграм або для синтезу семантичних дій у процесі аналізу вимог [10, с. 153]. Ці дослідження підкреслюють ефективність граматичного апарату у завданнях формалізації, проте залишаються обмеженими на рівні окремих операцій аналізу чи трансляції.

Таким чином, в оглянутих роботах простежується активний розвиток трьох паралельних напрямів – розширення UML за допомогою профілів, використання методів нечіткої логіки для формалізації вимог, а також застосування формальних граматик для перевірки та трансформації моделей. Проте відсутній інтегрований підхід, який поєднує формальну перевірку вимог на основі граматик із їх подальшим автоматизованим відображенням у UML-профіль, здатний фіксувати невизначеність та змінюваність.

Формулювання мети дослідження

Метою роботи є розробка методу трансформації формалізованих та узгоджених вимог у розширені UML-моделі на основі граматичного підходу, який забезпечує збереження властивостей невизначеності, імовірнісності та контекстної змінюваності у процесі моделювання складних адаптивних систем.

Для досягнення мети в роботі вирішуються такі задачі:

1) аналіз сучасних методів інженерії вимог та засобів моделювання на основі UML, включаючи профілі для специфічних предметних областей;

2) формалізація вимог із використанням атрибутної граматики для узгодження та побудови множини вимог, придатної до подальшої трансформації;

3) розробка системи трансформаційних правил, що відображають формалізовані вимоги у структурні та поведінкові елементи розширеної UML-моделі;

4) демонстрація застосування методу на прикладі критичної предметної області (інтелектуальна медична система) та оцінка його практичної придатності;

5) визначення переваг, обмежень та перспектив розвитку підходу, зокрема його інтеграції з CASE-засобами.

Викладення основного матеріалу дослідження

Формалізація вимог та їхнє модельне відображення

Трансформація вимог у UML-моделі потребує попередньої їхньої формалізації та узгодження, а також визначення апарату для подання властивостей невизначеності й змінюваності у межах нотації UML.

Формалізація здійснюється за допомогою контекстно-вільної та атрибутної граматики, що поєднує структурні правила побудови вимог із набором семантичних характеристик. Контекстно-вільна частина забезпечує однозначне синтаксичне подання, тоді як атрибутна дозволяє зберігати додаткові параметри, важливі для подальшого аналізу.

Основними атрибутами є: *confidence* – ступінь довіри до вимоги, *likelihood* – імовірність активації, *context* – умови виконання, *priority* – відносна важливість, *resources* (ρ) – оцінка необхідних витрат.

На основі цих параметрів здійснюється багаторівнева перевірка узгодженості: логічна, функціональна, ресурсна та контекстна. Результатом є узгоджена множина вимог R^*, яка зберігає інформацію про всі ключові характеристики та слугує підґрунтям для подальшої трансформації у модельне подання.

Щоб забезпечити відображення властивостей невизначеності у UML, використовується спеціалізований профіль, побудований на розширенні стандартної метамоделі. Він включає сутності *Requirement*, *Uncertainty Source*, *Variability Model* та їхні зв'язки з UML-артефактами.

У межах профілю передбачено низку стереотипів: <<fuzzyRequirement>> – нечіткі вимоги, <<probabilistic>> – ймовірнісні, <<variableRequirement>> – змінювані; <<temporalVariant>> – залежні від часу; <<evolvableRequirement>> – такі, що еволюціонують у процесі життєвого циклу.

Ці стереотипи супроводжуються тегованими значеннями (*confidence*, *trustDecayRate*, *likelihood*, *context*, *priority*, *resources*), які дозволяють відобразити кількісні та якісні характеристики вимог безпосередньо у UML-моделях. Додатково визначено інваріанти та обмеження, що забезпечують автоматизовану перевірку консистентності та підтримують трасованість між вимогами і модельними елементами.

Таким чином, формалізовані й узгоджені вимоги з множини R^* можуть бути безпосередньо трансформовані у UML-моделі, доповнені профілем для роботи з недосконалими та змінюваними специфікаціями.

Дворівнева граматика трансформації вимог у UML

Розроблений підхід ґрунтується на поєднанні двох граматичних рівнів, які працюють послідовно:

1. Рівень G_A – атрибутна граматика узгодження вимог. Вона забезпечує синтаксичний і семантичний аналіз початкових специфікацій, формує абстрактне синтаксичне дерево (AST), перевіряє їхню узгодженість за атрибутами та повертає узгоджену множину R^*.

2. Рівень G_T – трансформаційна граматика UML-відображень. Цей рівень одержує на вхід множину R^*, що містить вже перевірені вимоги, і застосовує правила трансформації до вузлів AST, утворюючи елементи UML, доповнені профілем.

Дворівнева грамматика подається як кортеж:

$$G = \langle G_A, G_T, R^*, \tau \rangle, \quad (1)$$

де G_A – атрибутна граматика узгодження, R^* – множина узгоджених вимог, що є виходом G_A, G_T – граматика трансформації вимог у UML, $\tau : R^* \rightarrow \text{UML}^+$ – відображення множини узгоджених вимог у елементи UML з профілем.

Ця дворівнева грамматика забезпечує перехід від неструктурованих вимог до повноцінної UML-моделі з урахуванням невизначеності та варіативності.

Трансформаційна граматика G_T визначається у вигляді п'ятірки:

$$G_T = \langle N, \Sigma, P, S, A_T \rangle,$$

де N – множина нетерміналів, що позначають абстрактні класи вимог (функціональні, нефункціональні, дані, сценарії); Σ – множина терміналів, які відповідають конкретним умовам та діям; P – набір правил трансформації;

S – початковий символ, що позначає окрему вимогу з множини R^*; A_T – система атрибутів, узгоджена з тегованими значеннями UML-профілю.

Кожне правило трансформації подається у вигляді:

$$T = (LHS, RHS, C), \tag{2}$$

де $LHS \subseteq AST$ – фрагмент дерева розбору, що відповідає вимозі або композиції вимог; RHS – UML-конструкція (UseCase, Class, Component, Constraint, Activity, State); C – умови застосовності правила, які залежать від атрибутів вимоги.

Відображення трансформації визначається як:

$$\tau : R^* \to UML^+, \tag{3}$$

де UML^+ – UML, розширений спеціалізованим профілем.

Для збереження властивостей невизначеності введемо множину стереотипів:

$$S = \{\langle \text{fuzzyRequirement} \rangle, \langle \text{probabilistic} \rangle, \langle \text{variableRequirement} \rangle\}.$$

Функція призначення стереотипів:

$$\sigma : R^* \to S \cup \varnothing, \tag{4}$$

де для кожної вимоги r_i:

– якщо $\text{conf}(r_i) < 1$, тоді $\sigma(r_i) = \langle \text{fuzzyRequirement} \rangle$;
– якщо $\text{likelihood}(r_i) \in (0,1)$, тоді $\sigma(r_i) = \langle \text{probabilistic} \rangle$;
– якщо $\text{context}(r_i)$ змінюється динамічно, тоді $\sigma(r_i) = \langle \text{variableRequirement} \rangle$;
– інакше – стереотип не застосовується.

Таким чином, кожна вимога отримує додаткову семантичну мітку, що переноситься у модель.

Розглянемо загальні правила трансформації для різних вимог та приклади їх застосування.

Функціональні вимоги. Загальне правило:

$$Func\,\text{Re}quirement(\varphi, \psi, A) \xrightarrow{C(A)} UseCase(\psi)^{\sigma(A)} + Class(\psi Controller)^{\sigma(A)},$$

де умова φ стає логічною умовою (умова переходу), дія ψ – назвою варіанту використання, а атрибути зберігаються як теговані значення.

Приклад 1 (медична система).

Requirement (φ = «Temp > 38», ψ = «NotifyDoctor», A = {priority = 0.8, confidence = 0.9}) → *Use Case* «NotifyDoctor» зі стереотипом <<fuzzyRequirement>> та клас *NotificationController* з методом +notifyDoctor().

Приклад 2 (банківська система).

Requirement (φ = «Balance < 0», ψ = «BlockAccount», A = {likelihood = 0.95, priority = 1.0}) → *Use Case* «BlockAccount» зі стереотипом <<probabilistic>> та клас *AccountController* з методом +blockAccount().

Приклад 3 (система бронювання квитків).

Requirement (φ = «SeatsAvailable > 0», ψ = «BookTicket», A = {priority = 0.9, context = OnlinePortal}) → *Use Case* «BookTicket» без додаткового стереотипу, але з тегами *priority* = 0.9 і *context* = *OnlinePortal*; створюється клас *BookingController* з операцією + bookTicket().

Нефункціональні вимоги. Загальне правило:

$$NonFunctional\,\text{Re}q(\text{Pr}edicate, A) \xrightarrow{C(A)} Constra\,\text{int}(\text{Pr}edicate)^{\sigma(A)}.$$

Приклад 1 (електронна комерція).

NonFunctionalReq(«ResponseTime < 2 s»,{confidence = 0.85}) → обмеження на клас *WebServer* зі стереотипом <<fuzzyRequirement>>.

Приклад 2 (сховище даних).

NonFunctionalReq(«Availabilit > 0.999»,{likelihood = 0.98, priority = 0.9}) → обмеження на компонент *StorageCluster* зі стереотипом <<probabilistic>>.

Приклад 3 (мобільний застосунок).

NonFunctionalReq(«BatteryUsage < 5 %/hour», {context = Mobile, priority = 0.8}) → обмеження на клас *AppCore* зі стереотипом <<variableRequirement>>, що вказує на залежність від контексту виконання.

Вимоги до даних. Загальне правило:

$$Data\,\text{Re}q(Metric(Element), A) \xrightarrow{C(A)} Class(Element) + Constra\,\text{int}(Metric)^{\sigma(A)}.$$

Приклад 1 (медична база даних).

DataReq(«Completeness(PatientRecord) > 0.95»,{context = Hospital,confidence = 0.9}) → клас *PatientRecord* з тегом *completeness* = 0.95 та стереотипом <<fuzzyRequirement>>.

Приклад 2 (система ідентифікації).

DataReq(«Accuracy(FaceRecognitionData)>0.98»,{priority=1.0,context=Security}) → клас *FaceRecognitionData* з обмеженням *accuracy* = 0.98 та стереотипом <<variableRequirement>>.

Приклад 3 (система електронних платежів).

DataReq(«Consistency(TransactionLog) = Strong»,{likelihood = 0.95}) → клас *TransactionLog* з тегом *consistency* = *Strong* і стереотипом <<probabilistic>>, що вказує на ймовірнісну гарантію консистентності.

Сценарні вимоги. Загальне правило:

$$Scenario\left(\{ri \to rj\}, A\right) \xrightarrow{C(A)} UML\left(r_i, r_j\right)^{\sigma(A)}\Big),$$

де $\{r_i \to r_j\}$ – залежність між вимогами, що виражає послідовність дій, а UMLDiagram може бути діаграмою послідовностей, діяльності або станів.

Приклад 1 (авторизація користувача, діаграма послідовностей).

Scenario(«EnterCredentials → ValidateCredentials → AccessGranted») → діаграма послідовностей надає повідомлення *enterCredentials*() → *validateCredentials*() → *grantAccess*(). Умова переходу: [*passwordCorrect*].

Приклад 2 (система моніторингу, діаграма станів).

Scenario(«Monitoring → AlarmTriggered») з умовою [$SpO_2 < 90$] → діаграма станів виконує перехід зі стану *Monitoring* до стану *AlarmTriggered* з «умовою переходу» [$SpO_2 < 90$].

Приклад 3 (онлайн-магазин, діаграма діяльності).

Scenario(«AddToCart → Checkout → Payment → Confirmation») → діаграма діяльності містить послідовність дій із розгалуженням:

– [paymentSuccessful] → дія «SendConfirmation»;
– [paymentFailed] → дія «NotifyUser».

На виході функції трансформації τ формується UML-модель:

$$M = \tau(R^*) = \{E_i \mid \exists\, r_j \in R^* : T(r_j) = E_i\}, \tag{5}$$

де *M* містить структурні (*класи, компоненти*), поведінкові (*варіанти використання, діаграми станів, активностей, послідовностей*) та обмежувальні (*OCL, Constraints*) елементи. Усі вони доповнені стереотипами та тегами, що відображають семантику вимог.

Таким чином, трансформація не лише створює UML-діаграми, а й переносить у них всю інформацію про невизначеність, пріоритетність і ресурси, збережену на рівні граматичної моделі.

Задля забезпечення прозорості змін використовується відображення:

$$\text{Trace} : R^* \leftrightarrow \text{UML}^+, \tag{6}$$

де кожній вимозі відповідає один або кілька UML-елементів.

Це дозволяє відстежувати вплив зміни вимоги на модель, контролювати повноту покриття вимог у моделі та підтримувати консистентність між рівнем специфікацій і UML.

Інтегрована методологія трансформації вимог у UML-моделі

Запропонована методологія забезпечує повний цикл переходу від початкових специфікацій до розширених UML-моделей і ґрунтується на дворівневій граматиці. Вона поєднує механізми узгодження вимог та їх подальшої трансформації у модельні представлення.

Етап 1. Формалізація вимог. На першому етапі неформальні специфікації подаються у вигляді продукцій атрибутної граматики G_A, що входить до складу дворівневої системи (1). У результаті будується абстрактне синтаксичне дерево, у вузлах якого фіксуються семантичні характеристики вимог (атрибути *confidence*, *likelihood*, *context*, *priority*, *resources*). Це створює основу для подальшої перевірки узгодженості. Обмеженням цього етапу є залежність від повноти словника граматики: нечіткі формулювання можуть бути інтерпретовані неоднозначно.

Етап 2. Перевірка узгодженості. Далі здійснюється контроль отриманих вимог за логічними, функціональними, ресурсними та контекстними критеріями. На виході формується узгоджена множина R^*, яка використовується як вхід для трансформаційної граматики G_T у (1). Якщо вимога не проходить одну з перевірок, вона відсікається або маркується як непридатна для відображення у UML.

Етап 3. Підготовка до трансформації. Кожна узгоджена вимога з множини R^* описується у вигляді фрагмента AST, який може бути використаний у правилах трансформації. Як визначено у (2), кожне правило описується трійкою (LHS,RHS,C), де ліва частина відповідає піддереву вимоги, права частина задає UML-конструкцію, а умова застосовності залежить від атрибутів.

Етап 4. Відображення у UML. Трансформація множини вимог визначається функцією (3). Для кожної вимоги застосовується функція призначення стереотипів σ (4), яка зіставляє значення атрибутів з профілем UML. Наприклад, вимоги з неповною впевненістю отримують стереотип <<fuzzyRequirement>>, імовірнісні – <<probabilistic>>, а залежні від динамічного контексту – <<variableRequirement>>. Теговані значення переносяться у модель автоматично, що забезпечує збереження семантики вимог. Обмеженням цього етапу є можливість конфлікту між кількома стереотипами, який розв'язується за допомогою політики пріоритетності.

Етап 5. Побудова UML-діаграм. У результаті трансформації формується множина UML-елементів M, визначена у (5). Вона включає структурні моделі (класи, компоненти), поведінкові діаграми (варіанти використання, діяльності, послідовності, стани), а також обмежувальні елементи (OCL-висловлювання та UML-Constraints). Логічні умови з вимог перетворюються на умови переходів, які у UML позначаються у квадратних дужках. Автоматична генерація поведінкових моделей можлива лише для сценаріїв з повним порядком виконання; у випадку часткового порядку потрібне додаткове уточнення вручну.

Етап 6. Трасування та валідація. Для збереження відповідності між вимогами та елементами моделі використовується відображення (6). Це дозволяє відстежувати походження UML-елементів, контролювати покриття вимог та виявляти пропуски. На завершальному етапі застосовуються інваріанти UML-профілю, які перевіряють допустимість значень атрибутів, правильність призначення стереотипів і відсутність конфліктних позначень.

Таким чином, інтегрована методологія забезпечує послідовний перехід:

$$Req \xrightarrow{G_A} R^* \xrightarrow{\tau} UML^+,$$

де кожен етап має формальне обґрунтування, що гарантує повну узгодженість процесу трансформації.

Приклад застосування інтегрованої методології трансформації вимог у UML-моделі

Для демонстрації методології розглянемо предметну область інтелектуальної медичної системи моніторингу пацієнтів. У таких системах критично важливими є швидкість реакції, точність інтерпретації показників та відстежуваність вимог у моделі.

Початкова множина вимог $Req = \{R_1, R_2, R_3\}$:

– R_1 (функціональна): якщо показник $SpO_2 < 90$, система повинна ініціювати тривогу;
– R_2 (нефункціональна): час реакції системи не повинен перевищувати 2 секунд;
– R_3 (сценарна): після успішної авторизації лікар має отримати доступ до історії пацієнта протягом 1 хвилини.

За допомогою атрибутної граматики G_A та перевірки узгодженості було підтверджено коректність усіх вимог та отримано узгоджену множину R^*.

Визначимо трансформації для різних типів UML-діаграм.

Діаграма варіантів використання:

– вимога R_1 трансформується у варіант використання «TriggerAlarm» зі стереотипом <<probabilistic>>;
– вимога R_3 породжує варіант використання «AccessHistory» зі стереотипом <<variableRequirement>>.

Діаграма класів:

– для R_1 створюється клас *AlarmController* з операцією +triggerAlarm();
– для R_3 створюється клас *HistoryController* з операцією +accessHistory();
– для R_2 у класі *System* фіксується обмеження {ResponseTime < 2 s}, позначене стереотипом <<fuzzyRequirement>>.

Діаграма послідовностей:

– вимога R_3 породжує сценарій: login() → validateUser() → accessHistory();
– умова переходу: [within 1 minute].

Діаграма станів:

– вимога R_1 відображається як перехід *Monitoring* → *AlarmTriggered* з умовою [$SpO_2 < 90$];
– вимога R_3 доповнює модель станами *LoggedOut* → *LoggedIn* → *HistoryAccessed* із часовим обмеженням «≤ 1 хвилина».

Узгоджені вимоги та їх відображення в UML-елементах подано нижче (табл. 1).

Таким чином, наведений приклад демонструє повний цикл трансформації від формалізованих вимог до узгоджених UML-моделей зі стереотипами та трасувальними зв'язками, що підтверджує практичну придатність запропонованої методології.

Оцінювання ефективності методу трансформації формалізованих вимог

Запропонований метод трансформації формалізованих вимог у розширені UML-моделі продемонстрував низку переваг у процесі проєктування.

По-перше, було досягнуто суттєвого скорочення часу на актуалізацію моделей. Завдяки автоматизованому застосуванню трансформаційних правил оновлення UML-діаграм після зміни вимог відбувається значно швидше, ніж у випадку ручного редагування. Це особливо важливо для систем, що функціонують у динамічному середовищі та потребують регулярної адаптації.

Таблиця 1

Трасування вимог до елементів UML-моделі

Вимога	UML-елементи	Стереотипи/теги	Типи діаграм
R_1: TriggerAlarm при $SpO_2 < 90$	Use Case «TriggerAlarm»; Class AlarmController; State «Monitoring → AlarmTriggered»	<<probabilistic>>, likelihood = 0.95, priority = 0.9	Use Case, Class, State
R_2: ResponseTime < 2 s	Constraint на клас System	<<fuzzyRequirement>>, confidence = 0.85	Class (з обмеженням)
R_3: Login → AccessHistory ≤ 1 хв	Use Case «AccessHistory»; Class HistoryController; Sequence «login → validate → access»; State «LoggedIn → HistoryAccessed»	<<variableRequirement>>, context = Hospital	Use Case, Class, Sequence, State

По-друге, методологія зберігає властивості невизначеності та варіативності вимог. Використання атрибутів у поєднанні зі стереотипами профілю UML дозволяє безпосередньо відобразити нечіткі, імовірнісні або контекстно залежні специфікації у модельних елементах. Таким чином, UML-моделі перестають бути лише детермінованими схемами, а відображають реальні обмеження та умови функціонування системи.

По-третє, у процес було інтегровано автоматизовану перевірку консистентності, що знижує ризик суперечностей і забезпечує вищу надійність кінцевої моделі. Узгоджена множина вимог R^* гарантує, що до етапу трансформації потрапляють лише ті специфікації, які пройшли логічний, функціональний, ресурсний і контекстний контроль.

При порівнянні з класичним UML-моделюванням (без використання розширеного профілю) виявлено суттєву відмінність. Стандартний UML дозволяє будувати структурні та поведінкові діаграми, однак він не надає засобів для формалізованого відображення невизначеності, імовірнісних властивостей або контекстної мінливості. У запропонованій методології ці характеристики зберігаються й інтегруються безпосередньо у модель, що робить її більш адекватною для опису сучасних складних систем.

Разом із тим метод має певні обмеження. Насамперед він потребує налаштування граматики під конкретний предметний домен, оскільки словник продукцій і правила трансформацій залежать від термінології та типових сценаріїв системи. Крім того, для повноцінного застосування необхідна інтеграція з CASE-засобами, здатними підтримувати розширений UML-профіль і забезпечувати автоматизоване трасування.

Висновки

У роботі представлено метод трансформації узгоджених вимог у розширені UML-моделі, що базується на поєднанні граматичного підходу з механізмами UML-профілю. Запропонована методологія забезпечує збереження семантичних характеристик вимог, підтримує властивості невизначеності, імовірнісності та контекстної змінюваності, а також гарантує узгодженість між специфікаціями та модельними елементами. Поєднання атрибутного граматичного аналізу та розширеного UML-профілю створює комплексний інструмент для моделювання адаптивних систем. Це дозволяє скоротити час актуалізації моделей, зменшити ризик суперечностей і підвищити точність відображення предметної області в умовах динамічного середовища.

Подальші дослідження передбачають розвиток методології у двох ключових напрямах: автоматизована генерація UML-діаграм у сучасних CASE-засобах та розширення підходу для моделювання поведінкових сценаріїв. Окремо слід відзначити перспективу інтеграції з SysML v2, яка може стати основою для застосування методології у сфері системної інженерії.

Список використаної літератури

1. Allian A.P., Nakagawa E.Y., Martinez J., Oliveira Jr. E., et al. Variability Implementation and UML-Based Software Product Lines. UML-Based Software Product Line Engineering with SMarty. Springer, Cham, 2022. С. 27–40. DOI: 10.1007/978-3-031-18556-4_2

2. Cu C., Ye X., Zheng Y. Xlinemapper: a product line feature-architecture-implementation mapping toolset. 41st International Conference on Software Engineering: Companion Proceedings (ICSE ’19). IEEE Press, Piscataway, 2019. С. 87–90. DOI: 10.1109/ICSE-Companion.2019.0004516

3. Abdelmadjid L. Uncertain Decision-Making Requirements Formalizing with CF UML model. Procedia Computer Science. 2022. Vol. 192. С. 2503–2512. DOI: 10.1016/j.procs.2021.12.247

4. Cazzola W., Olivares-Corichi I. Bridging the model-to-code abstraction gap with fuzzy logic: FLiRTS 2 for UML class diagrams. Software and Systems Modeling. 2022. Vol. 21. С. 1717–1742. DOI: 10.1007/s10270-021-00899-6

5. Cardiel-Ortega J. J., López-Robles J. R., Otegi-Olaso J. R., Gamboa-Rosales H. Probabilistic Fuzzy System for Evaluation and Failure Modes in FMEA. Processes. 2024. Vol. 12, Is. 6. Article № 1197. DOI: 10.3390/pr12061197

6. Dalpiaz F., Ferrari A., Franch X., Palomares C. Natural Language Processing for Requirements Engineering: The Best Is Yet to Come. IEEE Software. 2018. Vol. 35, No. 5. С. 115–119. DOI: 10.1109/MS.2018.3571242

7. Yang M., Ban A. Automated UML Class Diagram Generation from Textual Requirements Using NLP Techniques. JOIV International Journal on Informatics Visualization. 2024. Vol. 8, No. 3–2. С. 1905–1915. DOI: 10.62527/joiv.8.3-2.3482

8. Lano K., Xue Q., Haughton H. A Concrete Syntax Transformation Approach for Software Language Processing. SN Computer Science. 2024. Vol. 5. Article 645. DOI: 10.1007/s42979-024-02979-y

9. Ražinskas M., Miliūnas B., Jurgelaitis M., Čeponienė L., Bisikirskienė L. Transforming Sketches of UML Use Case Diagrams to Models. IEEE Access. 2024. Vol. 12. C. 185826–185837. DOI: 10.1109/ACCESS.2024.3514455

10. Maschotta R., Silatsa N., Jungebloud T., Hammer M., Zimmermann A. An OCL Implementation for Model-Driven Engineering of C++. Lee R. (ed.) Software Engineering Research, Management and Applications (SERA 2022). Studies in Computational Intelligence. Vol. 1053. Springer, Cham, 2022. C. 151–168. DOI: 10.1007/978-3-031-09145-2_10

References

1. Allian, A. P., Nakagawa, E. Y., Martinez, J., & Oliveira Jr., E., et al. (2022). Variability implementation and UML-based software product lines. In *UML-Based Software Product Line Engineering with SMarty*, pp. 27–40. Springer, Cham. https://doi.org/10.1007/978-3-031-18556-4_2

2. Cu, C., Ye, X., & Zheng, Y. (2019). Xlinemapper: a product line feature-architecture-implementation mapping toolset. In *41st International Conference on Software Engineering: Companion Proceedings (ICSE '19)*, pp. 87–90. IEEE Press, Piscataway. https://doi.org/10.1109/ICSE-Companion.2019.0004516

3. Abdelmadjid, L. (2022). Uncertain decision-making requirements formalizing with CF UML model. *Procedia Computer Science*, 192, pp. 2503–2512. https://doi.org/10.1016/j.procs.2021.12.247

4. Cazzola, W., & Olivares-Corichi, I. (2022). Bridging the model-to-code abstraction gap with fuzzy logic: FLiRTS 2 for UML class diagrams. *Software and Systems Modeling*, 21, pp. 1717–1742. https://doi.org/10.1007/s10270-021-00899-6

5. Cardiel-Ortega, J. J., López-Robles, J. R., Otegi-Olaso, J. R., & Gamboa-Rosales, H. (2024). Probabilistic fuzzy system for evaluation and failure modes in FMEA. *Processes*, 12(6), Article 1197. https://doi.org/10.3390/pr12061197

6. Dalpiaz, F., Ferrari, A., Franch, X., & Palomares, C. (2018). Natural language processing for requirements engineering: The best is yet to come. *IEEE Software*, 35(5), pp. 115–119. https://doi.org/10.1109/MS.2018.3571242

7. Yang, M., & Ban, A. (2024). Automated UML class diagram generation from textual requirements using NLP techniques. *JOIV International Journal on Informatics Visualization*, 8(3–2), pp. 1905–1915. https://doi.org/10.62527/joiv.8.3-2.3482

8. Lano, K., Xue, Q., & Haughton, H. (2024). A concrete syntax transformation approach for software language processing. *SN Computer Science*, 5, Article 645. https://doi.org/10.1007/s42979-024-02979-y

9. Ražinskas, M., Miliūnas, B., Jurgelaitis, M., Čeponienė, L., & Bisikirskienė, L. (2024). Transforming sketches of UML use case diagrams to models. *IEEE Access*, 12, pp. 185826–185837. https://doi.org/10.1109/ACCESS.2024.3514455

10. Maschotta, R., Silatsa, N., Jungebloud, T., Hammer, M., & Zimmermann, A. (2022). An OCL implementation for model-driven engineering of C++. In R. Lee (Ed.), *Software Engineering Research, Management and Applications (SERA 2022)*. *Studies in Computational Intelligence*, Vol. 1053, pp. 151–168. Springer, Cham. https://doi.org/10.1007/978-3-031-09145-2_10

Дата першого надходження рукопису до видання: 22.09.2025
Дата прийнятого до друку рукопису після рецензування: 17.10.2025
Дата публікації:

УДК 004.415.5 DOI

Н. О. КОМЛЕВА
кандидат технічних наук,
завідувач кафедри інженерії програмного забезпечення
Національний університет «Одеська політехніка»
ORCID: 0000-0001-9627-8530

М. І. НІКІТЧЕНКО
аспірант кафедри інженерії програмного забезпечення
Національний університет «Одеська політехніка»
ORCID: 0009-0007-9560-7057

МЕТОД ДВОСТОРОННЬОЇ СИНХРОНІЗАЦІЇ UML-МОДЕЛІ З ДВОМА ПОДАННЯМИ НА ОСНОВІ ІНКРЕМЕНТАЛЬНИХ ОНОВЛЕНЬ

У статті представлено формальний метод двосторонньої синхронізації UML-моделі з двома поданнями, які зберігається в окремих форматах: структурне подання реалізується у вигляді об'єктної нотації JavaScript, а поведінкове – у форматі міжплатформного обміну XMI із застосуванням спеціалізованого профілю. Основною метою дослідження є забезпечення узгодженості обох подань у процесі еволюції моделі на основі валідації інваріантів та інкрементального оновлення. У роботі описано математичну формалізацію множин структурних та поведінкових елементів, а також відповідності між ними, що реалізується за допомогою трасувальних зв'язків. Розроблено механізм відстеження змін у поведінковій частині моделі з побудовою графа залежностей, який дозволяє визначити підмножину елементів, що потребують повторної перевірки інваріантів. Запропонований метод верифікації моделі ґрунтується на інваріантах консистентності та забезпечує формальні гарантії семантичної відповідності. У рамках дослідження було проведено експериментальне оцінювання запропонованого методу на наборі тестових UML-моделей, що моделюють типові бізнес-процеси у сфері електронної комерції, зокрема обробку замовлень, управління кошиком, авторизацію користувачів і виконання транзакцій. До складу моделей входили різноманітні типи поведінкових діаграм, зокрема діаграми станів для опису життєвих циклів замовлень, діаграми діяльності для моделювання логіки послідовного виконання дій, а також діаграми послідовностей для представлення сценаріїв взаємодії між об'єктами системи. Завдяки використанню моделей з реалістичними сценаріями взаємодії було перевірено здатність методу до виявлення і локалізації порушень консистентності, що виникають внаслідок часткових змін як у структурному, так і в поведінковому поданні. Результати показали істотне зменшення кількості перевірок у порівнянні з повною валідацією. Запропонований підхід є сумісним з існуючими інструментами UML-моделювання та може бути інтегрований у процеси CI/CD та еволюційного аналізу архітектури.

Ключові слова: *UML, подання, синхронізація моделей, інкрементальне оновлення, консистентність, граф залежностей.*

N. O. KOMLEVA
Candidate of Technical Sciences,
Head of the Department of Software Engineering
Odesa Polytechnic National University
ORCID: 0000-0001-9627-8530

M. I. NIKITCHENKO
Postgraduate Student at the Department of Software Engineering
Odesa Polytechnic National University
ORCID: 0009-0007-9560-7057

METHOD FOR BILATERAL SYNCHRONIZATION OF A UML MODEL WITH TWO REPRESENTATIONS BASED ON INCREMENTAL UPDATES

The article presents a formal method for bilateral synchronization of a UML model with two views stored in separate formats: the structural view is implemented as JavaScript object notation, and the behavioral view is implemented in the XMI cross-platform exchange format using a specialized profile. The main goal of the research is to ensure the consistency of both views in the process of model evolution based on invariant validation and incremental updating. The paper

© Комлева Н. О., Нікітченко М. І., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

describes the mathematical formalization of sets of structural and behavioral elements, as well as the correspondence between them, which is implemented using traceable links. A mechanism for tracking changes in the behavioral part of the model has been developed with the construction of a dependency graph, which allows determining a subset of elements that require re-checking of invariants. The proposed model verification method is based on consistency invariants and provides formal guarantees of semantic correspondence. As part of the study, an experimental evaluation of the proposed method was carried out on a set of test UML models that simulate typical business processes in the field of e-commerce, in particular order processing, shopping cart management, user authorization, and transaction execution. The models included various types of behavioral diagrams, including state diagrams for describing order lifecycles, activity diagrams for modeling the logic of sequential execution of actions, and sequence diagrams for representing interaction scenarios between system objects. By using models with realistic interaction scenarios, the method's ability to detect and localize consistency violations resulting from partial changes in both structural and behavioral view was tested. The results showed a significant reduction in the number of checks compared to full validation. The proposed approach is compatible with existing UML modeling tools and can be integrated into CI/CD and evolutionary architecture analysis processes.

Key words: *UML, presentation, model synchronization, incremental update, consistency, dependency graph.*

Постановка проблеми

У сучасних середовищах розроблення програмного забезпечення зростає попит на гнучке та адаптивне використання моделей, що поєднують різні види подань. Одним із прикладів такого підходу є комбіноване представлення моделей уніфікованого моделювання UML, де структурні аспекти зберігаються у форматі об'єктної нотації JavaScript (JSON), а поведінкові – у форматі обміну моделями XMI. Така подільність забезпечує сумісність з різними інструментами, підтримує кросплатформну обробку, а також спрощує інкрементальне оновлення окремих фрагментів моделі без необхідності повної реконструкції.

Разом із перевагами, подібне поєднання подань породжує низку складних викликів, зокрема щодо підтримання узгодженості між ними, забезпечення формальної верифікації їхньої відповідності та збереження цілісності моделі при її еволюції. Попередні дослідження були зосереджені на валідаційних інваріантах для окремих подань, а також на алгоритмах контролю консистентності між структурною та поведінковою частинами. Проте недостатньо розробленим залишається питання ефективної двосторонньої синхронізації, яка б дозволяла виявляти і застосовувати локальні зміни до іншого подання, не порушуючи цілісності моделі.

Метою цієї роботи є формальне обґрунтування та розроблення методу двосторонньої синхронізації UML-моделі, що базується на інкрементальному оновленні та використовує попередньо верифіковану структуру як передумову коректної роботи. Основу підходу становить теорема інкрементальної консистентності, яка дозволяє обмежити обсяг повторної перевірки при часткових змінах моделі.

Аналіз останніх досліджень і публікацій

Проблематика синхронізації моделей у контексті інженерії програмного забезпечення є предметом активних досліджень [1]. Особливої актуальності це набуває в рамках багатоподаневих підходів, де одна й та сама система описується за допомогою набору гетерогенних, але семантично пов'язаних моделей [2].

Фундаментальною основою для вирішення цього завдання є концепція двосторонніх трансформацій, яка передбачає не лише перетворення моделі з одного подання в інше, а й зворотне поширення змін [3]. Класичним підходом до реалізації є використання потрійних граматик графів, які дозволяють формально описати правила відповідності між двома моделями та автоматично генерувати механізми їхньої синхронізації [4, 5]. Проте такий підхід часто має високу обчислювальну складність і може бути неоптимальними для інкрементальних оновлень, оскільки потребує перегляду значної частини моделі навіть при локальних змінах. Для підвищення ефективності процесу валідації активно розвиваються методи інкрементальної перевірки консистентності [6, 7], що дозволяє суттєво скоротити час на валідацію. Але більшість досліджень зосереджена на узгодженості між структурними діаграмами, тоді як питання консистентності між структурними та поведінковими аспектами залишається менш дослідженим [8].

В рамках попередніх праць авторів [9-11] було сформульовано набір валідаційних правил для обох типів подання, а також розроблено механізм інкрементального контролю консистентності при зміні однієї з частин моделі. Водночас зазначені підходи не реалізовують повноцінну синхронізацію в обох напрямах та не формалізують умови коректності таких перетворень. Також у суміжних дослідженнях розглядається підтримка комбінації форматів у середовищах моделювання, однак більшість із них не надає відкритих механізмів для трасування між елементами різних подань [12], що обмежує можливості автоматизованої перевірки відповідності або адаптації одного подання до змін в іншому.

Формулювання мети дослідження

Метою роботи є створення методу двосторонньої синхронізації UML-моделі з двома поданнями. Він має поєднувати в собі підтримку гетерогенних форматів подання моделі, використання верифікованої інформаційної бази, інкрементальність змін та двосторонню синхронізацію з відновленням локальних залежностей.

Викладення основного матеріалу дослідження

У межах даного дослідження під UML-моделлю з двома поданнями розуміється модель, яка містить два взаємопов'язані подання: структурне та поведінкове. Це можна описати наступним виразом:

$$M = (M_S, M_B, \mu),$$

де M_S – це множина елементів структурного подання моделі, що представлені у форматі JSON, M_B – це множина елементів поведінкового подання моделі, що представлені у форматі XMI згідно зі специфікацією UML, $\mu \subseteq M_S \times M_B$ – відношення відповідності між множинами структурних елементів M_S та поведінкових M_B елементів моделі.

Запропоноване відображення μ пов'язує кожен поведінковий елемент із одним або кількома структурними елементами, від яких він залежить. Воно реалізується через спеціалізовані атрибути трасування (наприклад, jsonRef, triggerSource), що явно вказують на відповідний ідентифікатор у JSON-поданні.

Оскільки модель може змінюватися з обох боків, визначаються множини локальних змін:

– $\Delta M_S \subseteq M_S$ – множина структурних елементів, які були додані, змінені або видалені;
– $\Delta M_B \subseteq M_B$ – множина змінених поведінкових елементів відповідно.

Для подальшої локалізації валідаційних перевірок і побудови синхронізації використовується орієнтований граф залежностей $G = (V, E)$, де множина вузлів V містить усі елементи M_B, а множина ребер E фіксує залежності між ними. У графі виділяють три класи залежностей:

– структурні – описують синтаксичні зв'язки між елементами діаграм;
– семантичні – відображають логіку виконання поведінки;
– трасувальні – забезпечують відповідність із структурними елементами M_S.

Граф G дозволяє визначити множину елементів, які слід повторно перевірити після внесення змін. Це суттєво скорочує обсяг валідаційних операцій і дозволяє побудувати інкрементальний механізм синхронізації.

Далі, перш ніж переходити до побудови механізмів синхронізації між структурним і поведінковим поданнями, необхідно забезпечити верифікацію такої UML-моделі як передумову її коректного функціонування. У цьому контексті під верифікацією розуміється підтвердження відповідності моделі заздалегідь визначеним валідаційним правилам, що формалізують допустимі структури та поведінкові патерни відповідно до семантики UML.

Структурна частина моделі представлена у форматі JSON відповідно до запропонованої метамоделі. Валідаційні правила для цієї частини були раніше систематизовані у [1], зокрема обов'язковість імен для сутностей, коректність типізації атрибутів і зв'язків, єдність ідентифікаторів у межах одного простору імен, відповідність класів, атрибутів і операцій їх метатипам. Ці правила інтерпретуються як множина предикатів $\sigma_1, \sigma_2, \ldots, \sigma_n$, кожен з яких перевіряється над множиною елементів M_S.

Формальна модель валідації виглядає наступним чином:

$$\forall m \in M_S : i = 1n\sigma_i(m) = \text{true}.$$

Поведінкова частина представлена у форматі XMI, який реалізує специфікацію UML 2.5.x з розширенням у вигляді спеціального профілю. Валідаційні інваріанти для поведінкових діаграм наведені в [2] і охоплюють діаграми станів (перевірка умов переходу, унікальності початкового стану, наявності переходів), діаграми діяльності (коректність потоків керування, відповідність типів вхідних/вихідних даних) та діаграми послідовностей (зв'язність між повідомленнями, послідовність активацій, узгодженість із Lifeline).

Нехай $\psi_1, \psi_2, \ldots, \psi_k$ – предикати валідації поведінки над множиною M_B. Тоді модель задовольняє поведінкову верифікацію, якщо:

$$\forall b \in M_B : j = 1k\psi_J(b) = \text{true}.$$

Ключовою вимогою використання цих двох подань є семантична відповідність між структурою та поведінкою. Вона реалізується через відображення μ, яке визначає, які елементи M_B мають бути пов'язані з відповідними елементами M_S. Правила відповідності включають:

– наявність посилань jsonRef, triggerSource в елементах M_B;
– відповідність типів дій, викликів або повідомлень до відповідних методів або сигналів у структурі;
– відповідність параметрів і типів даних між InputPin та структурними атрибутами або операціями.

Тому, модель вважається верифікованою, якщо M_S задовольняє всі структурні інваріанти, M_B задовольняє всі поведінкові інваріанти, а також якщо всі елементи M_B, що мають трасувальні посилання, відображаються на коректні елементи M_S і виконуються узгоджувальні правила.

Ця верифікована база виступає точкою відліку для всіх подальших інкрементальних оновлень і є необхідною умовою запуску двосторонньої синхронізації, яка не повинна порушувати верифікованість моделі після застосування змін.

Далі, опираючись на вищевказаний формальний опис UML-моделі, необхідно сформувати теорему інкрементальної консистентності такої UML-моделі з двома поданнями.

Теорема.

Нехай існує наступна UML-модель:

$$M = (M_S, M_B, \mu),$$

де M_S – структурне подання моделі (у форматі JSON), M_B – поведінкове подання моделі (у форматі XMI), $\mu \subseteq M_S \times M_B$ – відношення відповідності між множиною структурних M_S та поведінкових M_B елементів моделі.

Також нехай:

Σ – сукупність формальних обмежень консистентності у вигляді набору OCL-інваріантів, які визначають допустимість моделей;

$\Delta \subseteq M_S \cup M_B$ – множина елементів, що зазнали змін внаслідок інкрементального редагування;

Affected(Δ, μ) $\subseteq M_S \cup M_B$ – замикання змінених елементів через відношення відповідності μ, тобто множина всіх елементів, для яких існує зв'язок із елементами з Δ або суміжність за структурними/поведінковими відношеннями.

Тоді, якщо після зміни елементів із Δ усі інваріанти з Σ виконуються для кожного елемента з множини Affected(Δ,μ), то модель M загалом також задовольняє всі інваріанти з Σ, тобто

$$(\forall e \in \mathit{Affected}(\Delta, \mu) : M \vDash \sum(e)) \Rightarrow M \vDash \sum.$$

Доведення від супротивного:

Припустимо протилежне твердженню теореми: нехай після змін у моделі виконується

$$\forall e \in \mathit{Affected}(\Delta, \mu) : M \vDash \sum(e),$$

але вся модель неконсистентна, тобто:

$$\exists \sigma \in \sum : M \nvDash \sigma.$$

За визначенням, кожен інваріант σ формується локально, тобто його істинність залежить лише від обмеженої множини елементів:

$$\mathrm{Context}(\sigma) \subseteq M_S \cup M_B.$$

З того, що $M \nvDash \sigma$, випливає, що хоча б один з елементів контексту Context(σ) порушує вимогу цього інваріанту.

Але якщо інваріант σ був порушений, і цей інваріант залежить від деякого елемента e, тоді або $e \in \Delta$, тобто був змінений безпосередньо, або $e \in \mathit{Affected}(\Delta, \mu)$, тобто зміни торкнулися суміжного елемента.

Це суперечить нашому припущенню, що всі інваріанти виконуються на $\mathit{Affected}(\Delta, \mu)$.

Отже, припущення про те, що $M \nvDash \sum$ при $M \nvDash \sum(e)$ $\forall e \in \mathit{Affected}(\Delta, \mu)$, є хибним.

В результаті, ця теорема дозволяє обмежити обчислювальну перевірку консистентності лише підмножиною $\mathit{Affected}(\Delta, \mu)$, що істотно знижує складність верифікації в умовах інкрементального редагування.

Запропонований метод двосторонньої синхронізації забезпечує підтримку узгодженості між структурним поданням моделі та поведінковим поданням шляхом виконання інкрементальних оновлень. Передбачається, що початкова модель вже верифікована, тобто задовольняє всі формальні інваріанти відповідно до встановлених валідаційних правил для обох подань. Зміни в одній частині моделі автоматично ініціюють відповідне пропагування (поширення) змін до іншої частини моделі, забезпечуючи логічну цілісність без повної перебудови моделі.

Завдання полягає в оновленні пов'язаного подання та перевірці локальних інваріантів для збереження консистентності моделі M. При цьому необхідно досягти актуалізацію відображення μ, перевірку лише зачеплених елементів (інкрементально) та збереження семантичної відповідності між структурою і поведінкою.

Метод передбачає чотири основні кроки.

Крок 1. Локалізація змін.Визначається об'єднана множина змінених елементів:

$$\Delta = \Delta M_S \cup \Delta M_B.$$

Будується множина зачеплених елементів:

$$\mathit{Affected} = x \in \Delta \mathit{Affected}(x),$$

що визначається за допомогою графа залежностей між елементами моделі.

Крок 2. Пропагування змін. Для кожного зміненого елемента виконується спрямоване оновлення пов'язаних об'єктів:

– якщо зміна відбулася в M_S, оновлюються або створюються відповідні елементи в M_B (наприклад, клас → Lifeline, атрибут → InputPin);

– якщо зміна в M_B, виконується оновлення відповідних елементів M_S (наприклад, Transition → булевий атрибут, Message → операція або сигнал);

– в обох випадках актуалізуються трасувальні зв'язки μ.

Крок 3. Локальна верифікація. На основі теореми інкрементальної консистентності перевіряються лише інваріанти, які залежать від елементів із множини Affected. Якщо всі локальні перевірки успішні, модель вважається узгодженою. У протилежному випадку зміна відхиляється або маркується як така, що потребує доопрацювання.

Крок 4. Оновлення відповідностей. Зміни у відношенні μ включають додавання пар для новостворених елементів та вилучення або редагування пар після видалення або зміни елементів.

Запропонований метод володіє кількома ключовими властивостями, що забезпечують його ефективність та надійність. По-перше, він є *інкрементальним*, оскільки перевірці підлягає лише та частина моделі, яку зачепили локальні зміни, що значно зменшує обчислювальну складність. По-друге, гарантується його *завершуваність*, тобто метод завжди зупиняється, оскільки він не створює рекурсивного розширення множини змін. Крім того, для *уникнення циклів* нескінченної синхронізації застосовуються спеціальні маркери походження змін (Push або Pull). Нарешті, метод підтримує *семантичну узгодженість*: поведінкові елементи, створені внаслідок структурних змін, зберігають повну відповідність до типів, параметрів і залежностей, визначених у структурному поданні M_S.

Для демонстрації ефективності інкрементального методу синхронізації було проведено серію тестів на вручну побудованих UML-моделях, які реалізують типові сценарії з прикладної галузі систем управління замовленнями в електронній комерції. Всі ці моделі відображають логіку обробки замовлень, перевірки клієнтів, обліку складу й комунікації з зовнішніми сервісами. При цьому варто враховувати, що хоч діаграми діяльності, станів і послідовностей належать до поведінкових засобів моделювання, вони прямо посилаються на елементи структури через спеціальні механізми трасування, такі як jsonRef, TriggerSource, PayloadType (табл. 1), завдяки чому забезпечується тісна семантична взаємодія між двома поданнями.

Таблиця 1

Характеристики тестових моделей

Назва моделі	Елементи M_S	Елементи M_B	Типи діаграм M_B	Сценарій	Механізми трасування
Model A	12 класів, 18 атрибутів	13 елементів	Діаграма станів	Життєвий цикл замовлення	jsonRef для станів, TriggerSource для подій
Model B	9 класів, 12 операцій	17 елементів	Діаграма діяльності	Авторизація клієнта	jsonRef, PayloadType на InputPin
Model C	7 класів, 10 зв'язків	15 елементів	Діаграма послідовностей	Виклик платіжного сервісу	TriggerSource для повідомлень, jsonRef на lifeline

Трасувальні зв'язки у кожному випадку встановлювались вручну з урахуванням семантики сценарію та структурної специфікації.

На першому етапі для кожної моделі було визначено повний набір застосовних інваріантів, що включав структурні (унікальність ID, валідність типів у JSON), поведінкові (досяжність станів, коректність потоків) та міжподаннєві правила консистентності (SignatureMatches, AnchorExists). Повна перевірка цього набору слугувала базовим показником. На другому етапі в кожну модель вносилася атомарна локальна зміна (Δ), що імітувала типові дії розробника, наприклад, редагування атрибута в M_S або додавання переходу в M_B. Після цього запускався інкрементальний алгоритм, який за допомогою графа залежностей визначав мінімальну підмножину зачеплених елементів *Affected*(Δ) (табл. 2). Перевірці підлягали лише інваріанти, що стосувалися цієї підмножини.

Таблиця 2

Порівняння повної та інкрементальної перевірки інваріантів

Модель	Тип зміни	Інваріантів загалом	Перевірено (повністю)	Перевірено (інкрементально)
A	M_B	7	7	3
B	M_S	9	9	4
C	$M_B + M_S$	11	11	5

Інкрементальний підхід у середньому дозволив зменшити кількість перевірених інваріантів на 56 %, що прямо зменшує навантаження при ручному аналізі та значно пришвидшує зворотний зв'язок у разі розширення чи зміни моделі.

Теоретична складність перевірки підтвердилась емпірично:

$$T(n) = O(k \cdot d),$$

де $k \ll n$.

Це дозволяє зберігати масштабованість навіть для більших моделей. Застосування графа залежностей виявилось корисним навіть у ручному режимі: воно дозволяє фокусуватися лише на критично важливих інваріантах, пов'язаних зі змінами.

Висновки

У роботі представлено метод двосторонньої синхронізації UML-моделі з дома поданнями, де структурна частина описується у форматі JSON, а поведінкова – у форматі XMI з використанням спеціалізованого UML-профілю. Основна увага зосереджена на забезпеченні узгодженості між цими поданнями в умовах локальних змін моделі. Запропонований підхід базується на формалізованій відповідності між структурними та поведінковими елементами, підтримує інкрементальне оновлення лише змінених фрагментів та гарантує збереження консистентності за допомогою перевірки інваріантів.

Порівняно з традиційними методами повної перевірки, інкрементальна стратегія дозволяє істотно знизити обчислювальні витрати. Результати експериментальної оцінки на прикладах із предметної області електронної комерції продемонстрували, що кількість інваріантів, які необхідно перевірити після локальної зміни, скорочується у 2–3 рази. Це особливо важливо в умовах застосування практик безперервної інтеграції, де ефективність перевірки відіграє критичну роль.

Метод дозволяє не лише виявляти порушення узгодженості, але й підтримує симетричну синхронізацію: кожна зміна в структурі або поведінці спричиняє відповідне пропагування змін до іншої частини моделі. Ця властивість відрізняє підхід від односторонніх трансформацій та забезпечує його застосовність у повноцінних системах управління еволюцією моделей.

Разом із перевагами, підхід має певні обмеження: він потребує дотримання правил трасування між поданнями, а також поки не охоплює всі типи UML-діаграм. Втім, завдяки гнучкості та чіткому формалізованому підґрунтю, запропоноване рішення можна розширювати та адаптувати під інші профілі та сценарії застосування. В результаті, метод створює надійну основу для узгодженого розвитку складних програмних моделей у сучасному інженерному середовищі.

Список використаної літератури

1. A systematic identification of consistency rules for UML diagrams / D. Torre et al. *Journal of Systems and Software*. 2018. Vol. 144. P. 121–142. URL: https://doi.org/10.1016/j.jss.2018.06.029

2. Cicchetti A., Ciccozzi F., Pierantonio A. Multi-view approaches for software and system modelling: a systematic literature review. *Software and Systems Modeling*. 2019. Vol. 18, no. 6. P. 3207–3233. URL: https://doi.org/10.1007/s10270-018-00713-w

3. Stevens P. Bidirectional model transformations in QVT: semantic issues and open questions. Software & Systems Modeling. 2008. Vol. 9, no. 1. P. 7–20. URL: https://doi.org/10.1007/s10270-008-0109-9

4. Formal analysis of model transformations based on triple graph grammars / F. HERMANN et al. Mathematical Structures in Computer Science. 2014. Vol. 24, no. 4. URL: https://doi.org/10.1017/s0960129512000370

5. Buchmann T., Westfechtel B. Using triple graph grammars to realise incremental round-trip engineering. IET Software. 2016. Vol. 10, no. 6. P. 173–181. URL: https://doi.org/10.1049/iet-sen.2015.0125

6. Mens T., Van Der Straeten R. Incremental Resolution of Model Inconsistencies. Recent Trends in Algebraic Development Techniques. Berlin, Heidelberg, 2007. P. 111–126. URL: https://doi.org/10.1007/978-3-540-71998-4_7

7. Oriol X., Teniente E. Incremental Checking of OCL Constraints with Aggregates Through SQL. Conceptual Modeling. Cham, 2015. P. 199–213. URL: https://doi.org/10.1007/978-3-319-25264-3_15

8. Knapp A., Mossakowski T. Multi-view Consistency in UML: A Survey. Graph Transformation, Specifications, and Nets. Cham, 2018. P. 37–60. URL: https://doi.org/10.1007/978-3-319-75396-6_3 (date of access: 11.08.2025).

9. Komleva N. O., Nikitchenko M. I. Method for Incremental Control of Consistency Between Structural and Behavioral Views of Software Architecture. AAIT. 2025. Vol. 8, no. 2. P. 162–177. URL: https://doi.org/10.15276/aait.08.2025.11

10. Комлева Н. О., Нікітченко М. І. Структурне подання UML-метамоделі у форматі JSON. Інформатика та математичні методи в моделюванні. 2025. Том 15, № 2. С. 205–217. URL: https://doi.org/10.15276/imms.v15.no2.205

11. Комлева Н. О., Нікітченко М. І. Поведінкове подання у форматі XMI в межах UML-метамоделі з двома поданнями. Вісник Хмельницького національного університету. Технічні науки. 2025. Том 5 (Подано до друку).

12. Нікітченко М. І. Аналіз форматів збереження UML-моделей у сучасних CASE-засобах. Технології та суспільство: взаємодія, вплив, трансформація : матеріали IV Міжнар. наук. конф. (Чернігів, 20 червня 2025 р.). Чернігів, 2025. URL: https://doi.org/10.62731/mcnd-20.06.2025

References

1. Torre, D., Labiche, Y., Genero, M., & Kassab, M. M. (2018). A systematic identification of consistency rules for UML diagrams. Journal of Systems and Software, 144, 121–142. https://doi.org/10.1016/j.jss.2018.06.029

2. Cicchetti, A., Ciccozzi, F., & Pierantonio, A. (2019). Multi-view approaches for software and system modelling: A systematic literature review. Software and Systems Modeling, 18(6), 3207–3233. https://doi.org/10.1007/s10270-018-00713-w

3. Stevens, P. (2008). Bidirectional model transformations in QVT: Semantic issues and open questions. Software & Systems Modeling, 9(1), 7–20. https://doi.org/10.1007/s10270-008-0109-9

4. Hermann, F., Ehrig, H., Golas, U., & Orejas, F. (2014). Formal analysis of model transformations based on triple graph grammars. Mathematical Structures in Computer Science, 24(4). https://doi.org/10.1017/s0960129512000370

5. Buchmann, T., & Westfechtel, B. (2016). Using triple graph grammars to realise incremental round-trip engineering. IET Software, 10(6), 173–181. https://doi.org/10.1049/iet-sen.2015.0125

6. Mens, T., & Van Der Straeten, R. (2007). Incremental resolution of model inconsistencies. In A. Tarlecki (Ed.), Recent trends in algebraic development techniques (pp. 111–126). Springer. https://doi.org/10.1007/978-3-540-71998-4_7

7. Oriol, X., & Teniente, E. (2015). Incremental checking of OCL constraints with aggregates through SQL. In P. Johannesson, M. L. Lee, S. W. Liddle, A. L. Opdahl, & Ó. Pastor López (Eds.), Conceptual modeling (pp. 199–213). Springer. https://doi.org/10.1007/978-3-319-25264-3_15

8. Knapp, A., & Mossakowski, T. (2018). Multi-view consistency in UML: A survey. In R. Heckel & G. Taentzer (Eds.), Graph transformation, specifications, and nets (pp. 37–60). Springer. https://doi.org/10.1007/978-3-319-75396-6_3

9. Komleva, N. O., & Nikitchenko, M. I. (2025). Method for incremental control of consistency between structural and behavioral views of software architecture. AAIT, 8(2), 162–177. https://doi.org/10.15276/aait.08.2025.11

10. Komleva, N. O., & Nikitchenko, M. I. (2025). Strukturne podannia UML-metamodeli u formati JSON. Informatyka ta matematychni metody v modeliuvanni, 15(2), 205–217. https://doi.org/10.15276/imms.v15.no2.205

11. Komleva, N. O., & Nikitchenko, M. I. (2025). Povedinkove podannia u formati XMI v mezhakh UML-metamodeli z dvoma podanniamy[Manuscript submitted for publication].

12. Nikitchenko, M. I. (2025, June 20). Analiz formativ zberezhennia UML-modelei u suchasnykh CASE-zasobakh [Paper presentation]. Technology and society: interaction, influence, transformation: IV International Scientific Conference, Chernihiv, Ukraine. https://doi.org/10.62731/mcnd-20.06.2025

Дата першого надходження рукопису до видання: 24.09.2025
Дата прийнятого до друку рукопису після рецензування: 21.10.2025
Дата публікації:

УДК 004.896:623.746.7:355.41 DOI

М. М. КОПИЛЕЦЬ
аспірант кафедри інформаційних систем на мереж
Національний університет «Львівська політехніка»
ORCID: 0009-0004-5823-9871

А. Я. САВКА
аспірант кафедри інформаційних систем на мереж
Національний університет «Львівська політехніка»
ORCID: 0009-0004-9198-6713

МЕТОДИ НАВЧАННЯ З ПІДКРІПЛЕННЯМ БЕЗПІЛОТНИХ ЛІТАЛЬНИХ АПАРАТІВ У ЗАВДАННЯХ ВІЙСЬКОВОЇ ЛОГІСТИКИ

У статті проведено розгорнутий огляд сучасних методів навчання з підкріпленням (Reinforcement Learning, RL) та їхнього застосування у сфері військової логістики із використанням безпілотних літальних апаратів (БпЛА). Актуальність теми зумовлена зростанням ролі БпЛА у забезпеченні оперативного транспортування вантажів, розвідки та підтримки бойових підрозділів, особливо в умовах обмеженого часу та високого рівня ризику. Проаналізовано ключові наукові роботи, що демонструють ефективність RL у завданнях планування маршруту, координації груп дронів (multi-agent RL, MARL), управління ресурсами та енергоспоживанням, а також у контексті врахування невизначеностей і ризиків (risk-sensitive RL, CVaR). Особливу увагу приділено підходам, які дозволяють формалізувати логістичні задачі як марковські процеси прийняття рішень із обмеженнями (CMDP), застосуванню механізмів уваги та графових нейронних мереж для оптимізації маршрутів, а також технологіям центрального навчання з децентралізованим виконанням (CTDE), що забезпечують ефективну взаємодію кількох БпЛА в реальному часі. Наведено математичні моделі та формули, що описують процес оптимізації політики керування, енергетичні обмеження та алгоритмічні модифікації, спрямовані на підвищення безпеки й надійності виконання завдань. Огляд містить аналіз підходів до інтеграції RL-рішень із системами моніторингу й контролю, а також описує сучасні виклики, серед яких є проблема перенесення навчених політик із симуляторів у реальні умови (sim-to-real), обмеженість обчислювальних ресурсів на борту БпЛА та необхідність стійкості до втрати зв'язку. Результати роботи можуть бути використані для побудови ефективних логістичних платформ, що здатні до автономної адаптації в умовах динамічних і небезпечних середовищ.

Ключові слова: *навчання з підкріпленням, безпілотні літальні апарати, військова логістика, MARL, CMDP, CVaR, CTDE, маршрутизація, енергетичні обмеження, симуляційне навчання.*

M. M. KOPYLETS
Postgraduate Student at the Department of Information Systems and Networks
Lviv Polytechnic National University
ORCID: 0009-0004-5823-9871

A. YA. SAVKA
Postgraduate Student at the Department of Information Systems and Networks
Lviv Polytechnic National University
ORCID: 0009-0004-9198-6713

REINFORCEMENT LEARNING METHODS FOR UNMANNED AERIAL VEHICLES IN MILITARY LOGISTICS TASKS

This article provides a comprehensive review of modern reinforcement learning (RL) methods and their application in military logistics involving unmanned aerial vehicles (UAVs). The relevance of this topic arises from the increasing role of UAVs in ensuring rapid cargo transportation, reconnaissance, and support for combat units, particularly under time constraints and high-risk conditions. The paper analyzes key peer-reviewed studies that demonstrate the efficiency of RL in route planning, multi-agent coordination (multi-agent RL, MARL), resource and energy management, and risk-sensitive decision-making (CVaR). Special attention is devoted to approaches that formalize logistics tasks as constrained Markov decision processes (CMDP), the application of attention mechanisms and graph neural networks for route optimization, and centralized training with decentralized execution (CTDE) frameworks that enable effective multi-UAV cooperation in real time. Mathematical models and formulas describing policy optimization, energy constraints, and algorithmic safety enhancements are presented. The review also examines methods for integrating RL solutions with monitoring and

© Копилець М. М., Савка А. Я., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

control systems, addressing current challenges such as sim-to-real transfer, limited onboard computational resources, and resilience to communication loss. The findings of this work can serve as a foundation for building autonomous, adaptive logistics platforms capable of operating efficiently in dynamic and hazardous environments.

Key words: *reinforcement learning, unmanned aerial vehicles, military logistics, MARL, CMDP, CVaR, CTDE, routing, energy constraints, simulation-based training.*

Постановка проблеми

Військова логістика є одним із ключових елементів забезпечення боєздатності армії, оскільки ефективність постачання ресурсів, транспортування обладнання та своєчасного підвезення боєприпасів безпосередньо впливає на результат бойових операцій [1–3]. Сучасні конфлікти характеризуються високою динамічністю, асиметричними загрозами, швидкими змінами тактичної обстановки та необхідністю забезпечення оперативного реагування у режимі реального часу. У таких умовах традиційні методи логістичного планування, такі як статичні маршрути, централізоване управління та жорстко фіксовані графіки, виявляються малоефективними або навіть непридатними. Використання безпілотних літальних апаратів (БпЛА) відкриває нові можливості для вирішення логістичних задач. Завдяки мобільності, здатності працювати в умовах обмеженого доступу та зниженому ризику для особового складу, дрони вже успішно застосовуються для доставки невеликих, але критично важливих вантажів: медикаментів, комплектів зв'язку, деталей для ремонту техніки. Проте інтеграція БпЛА у військову логістику вимагає вирішення ряду складних проблем, пов'язаних із маршрутизацією, управлінням енергетичними ресурсами, уникненням загроз та координацією груп дронів у динамічних і небезпечних умовах [4, 5].

Одним з найбільш перспективних підходів до вирішення цих задач є навчання з підкріпленням (Reinforcement Learning, RL), яке дозволяє агенту (у нашому випадку БпЛА або групі БпЛА) навчатися оптимальної поведінки на основі взаємодії з середовищем, максимізуючи довгострокову винагороду. Особливої уваги заслуговують модифікації RL, такі як багатоагентне підкріплювальне навчання (Multi-Agent RL, MARL), що забезпечує координацію між кількома дронами, а також підходи з обмеженнями (Constrained MDP, CMDP), які дозволяють враховувати обмеження безпеки, енергоспоживання та зон доступу. Проблематика стає ще складнішою у зв'язку з наявністю стохастичних факторів: зміни погодних умов, непередбачуваних загроз, ворожих дій, обмеженої пропускної здатності каналів зв'язку. Це обумовлює необхідність використання ризик-орієнтованих методів (наприклад, CVaR Conditional Value at Risk), що дозволяють мінімізувати негативні наслідки екстримальних подій, критичних для військових операцій [6–8].

Таким чином, наукове завдання полягає у систематизації сучасних методів RL, які можна ефективно застосовувати у військовій логістиці з використанням БпЛА, з урахуванням як математичних моделей, так і алгоритмічних рішень, здатних забезпечити надійну та безпечну роботу у реальних бойових умовах.

Аналіз останніх досліджень і публікацій

За останнє десятиліття інтерес до використання підкріплювального навчання у військовій логістиці та робототехніці значно зріс, що підтверджується зростаючою кількістю публікацій у провідних журналах з інтелектуальних систем, штучного інтелекту та автономних роботів. Розглянемо основні напрями досліджень, які формують наукову основу даної теми.

Перший пласт робіт пов'язаний із марковськими процесами прийняття рішень (MDP) для планування маршрутів автономних апаратів у складних середовищах. Класичні роботи, такі як дослідження [1], заклали базу RL, але вже у 2010-х з'явилися роботи, що демонструють практичну реалізацію цих підходів для літаючих апаратів. Наприклад, у [2] представлено застосування Q-learning та SARSA для навігації БпЛА з урахуванням перешкод та обмеженого запасу енергії. Ця робота показала, що навіть табличні методи RL здатні забезпечувати адаптивне планування в умовах змінної обстановки, хоча масштабування на великі простори станів залишалося проблемою.

Другий напрям досліджень – багатоагентне підкріплювальне навчання (MARL), яке дозволяє координувати групи дронів для спільного виконання завдань. У роботі [3] запропоновано Multi-Agent Deep Deterministic Policy Gradient (MADDPG), що забезпечує стабільне навчання у сценаріях з частковою спостережуваністю. Подальші дослідження, зокрема [4], інтегрували цей підхід із механізмами централізованого навчання та децентралізованого виконання (CTDE), що особливо важливо у військовій логістиці, де під час виконання завдань зв'язок між дронами може бути обмеженим або може переривatися.

Третій вагомий напрям це застосування Constrained Markov Decision Processes (CMDP). У роботі [5] було запропоновано алгоритм Constrained Policy Optimization (CPO), який забезпечує дотримання заданих обмежень (наприклад, рівня ризику або витрати енергії) при оптимізації політики. Подальші дослідження [6] показали ефективність CMDP у сценаріях, де критично важливо уникати загроз, таких як ворожі засоби ППО.

Важливим підходом до підвищення надійності є ризик-орієнтоване підкріплювальне навчання. В роботі [7] автори продемонстрували використання Conditional Value at Risk (CVaR) для мінімізації ймовірності катастрофічних подій. У роботі [8] запропоновано поєднання CVaR із моделями прогнозування загроз, що дало змогу дронам вибирати маршрути, які зменшують імовірність потрапляння в зону ураження навіть у випадках неповної інформації.

Окрему групу досліджень складають роботи з Graph Neural Networks (GNN) для моделювання логістичних мереж. В статті [9] показали, що GNN можуть ефективно узагальнювати політики на нові топології мережі, що критично важливо для військових логістичних операцій, де структура транспортних шляхів змінюється внаслідок бойових дій. Це дозволяє швидко перебудовувати маршрути без повторного тривалого навчання.

Нарешті, практична інтеграція RL у військові симуляційні середовища розглянута у [10], де запропоновано використання підходу sim-to-real з доменними випадковостями для підвищення стійкості політик при перенесенні їх з симулятора у реальний світ. Автори показали, що цей підхід знижує падіння ефективності при переході від навчання у віртуальному середовищі до роботи у фізичних умовах на 30–40 %.

Формулювання мети дослідження

Метою цієї роботи є огляд наукових підходів до застосування методів навчання з підкріпленням у завданнях військової логістики із використанням безпілотних літальних апаратів. У роботі узагальнюються формальні моделі прийняття рішень (MDP, CMDP, Dec-POMDP), розглядаються підходи глибинного та багатоагентного RL, а також методи, орієнтовані на безпеку та управління ризиками.

Основними завданнями дослідження є:

- систематизація існуючих підходів та класифікація їх застосування до різних типів логістичних задач;
- аналіз переваг і обмежень сучасних алгоритмів RL у контексті військових умов;
- формулювання узагальнених висновків щодо придатності методів RL до практичного використання у військовій логістиці.

Об’єктом дослідження виступає процес прийняття рішень у військових логістичних системах з використанням БпЛА, а предметом є алгоритми RL та їх здатність забезпечувати ефективне і безпечне планування дій у динамічному середовищі.

Викладення основного матеріалу дослідження

Проблема інтеграції RL у військову логістику з використанням БпЛА вимагає одночасного врахування як алгоритмічних, так і інженерних аспектів. У центрі цієї задачі є необхідність адаптивного планування та координації в умовах невизначеності, ризику та жорстких обмежень ресурсів. На відміну від цивільних сценаріїв, де оптимізація часто зводиться до мінімізації часу чи витрат, у військових умовах критичними є збереження апаратів, виконання місії та зниження операційних ризиків навіть за рахунок збільшення витрат часу або енергії [3, 5, 7].

Базова модель RL для окремого БпЛА може бути представлена як марковський процес прийняття рішень (MDP), що описується кортежем:

$$M = \langle S, A, P, R, \gamma \rangle, \tag{1}$$

де S – множина можливих станів (позиція дрона, заряд батареї, доступність каналів зв’язку, інформація про загрози); A – множина допустимих дій (зміна курсу, висоти, швидкості, виконання доставки); $P(s' \mid s, a)$ – ймовірність переходу в стан s' після виконання дії a в стані s; $R(s, a)$ – функція винагороди; $\gamma \in (0, 1)$ – коефіцієнт дисконтування, що відображає пріоритет короткострокових або довгострокових вигод. У задачах військової логістики важливо враховувати обмеження, які природно інтегруються у модель Constrained MDP (CMDP):

$$M_c = \langle S, A, P, R, C, d, \gamma \rangle, \tag{2}$$

де $C(s, a)$ – функція витрат (наприклад, споживання енергії або ризик перебування в зоні ураження), а d – допустимий рівень цих витрат. Оптимізація у CMDP здійснюється з урахуванням обмежень:

$$\max_{\pi} \mathbb{E}_{\pi}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t)\right], \quad \text{де} \quad \mathbb{E}_{\pi}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t)\right] \le d. \tag{3}$$

Методи на кшталт Constrained Policy Optimization (CPO) [5] дозволяють дотримуватися цих обмежень під час навчання політики, що особливо цінно для місій, де перевищення допустимого рівня ризику неприпустиме.

Військові середовища за своєю природою є стохастичними: поведінка противника, погодні умови, стан інфраструктури змінюються непередбачувано. Класичне RL намагається максимізувати середню винагороду, але в умовах бойових операцій це може призвести до небажаних сценаріїв: політика, що є оптимальною може бути фатальною в рідкісних, але критично небезпечних випадках. Тому застосовують ризик-орієнтовані критерії, зокрема Conditional Value at Risk (CVaR) [7,8], який формалізується як:

$$CVaR_{\alpha}(Z) = \mathbb{E}[Z \mid Z \le q_{\alpha}(Z)], \tag{4}$$

де $q_{\alpha}(Z)$ – α-квантиль розподілу виграшу Z. Ця метрика дозволяє оптимізувати політику так, щоб мінімізувати негативні наслідки у найгірших α % сценаріїв, що напряму відповідає вимогам військових операцій.

Координація кількох БпЛА є ще більш складним завданням, оскільки необхідно враховувати як спільну мету, так і індивідуальні обмеження кожного апарата. Multi-Agent RL (MARL) пропонує підходи, що поєднують

централізоване навчання з децентралізованим виконанням (CTDE), як у алгоритмах MADDPG [3] або QMIX [4]. У таких сценаріях функція винагороди може мати вигляд:

$$R_t = \frac{1}{N}\sum_{i=1}^{N} r_t^i - \lambda \cdot Collisions_t \tag{5}$$

де перший доданок описує середній прогрес усіх агентів, а другий – штраф за конфлікти, зіткнення або порушення зони безпеки. Важливим аспектом MARL у військовій логістиці є можливість розподілу завдань: одні дрони можуть виконувати доставку, інші – забезпечувати розвідку чи ретрансляцію сигналу зв'язку [6, 10]. У цьому випадку алгоритм повинен враховувати гетерогенність агентів, що підвищує складність оптимізації.

Маршрутизація в умовах бойових дій часто має мережеву природу: точки постачання, проміжні склади, транспортні коридори утворюють граф, вершинами якого є логістичні вузли, а ребрами – можливі шляхи. Graph Neural Networks (GNN) [9] дозволяють кодувати структуру цього графа і вбудовувати її в процес прийняття рішень RL-агентом. Це дає змогу швидко адаптувати політику до змін мережі без повного перенавчання.

Однією з ключових проблем застосування RL у реальних БпЛА є перенесення політик, навчених у симуляторах, у фізичний світ. Через спрощення моделі середовища у симуляторах політики можуть втрачати ефективність у реальних умовах. Технологія domain randomization [10] пропонує під час навчання в симуляторі варіювати параметри середовища (наприклад, силу вітру, масу вантажу, затримки сенсорів), щоб агент навчився працювати у більш широкому спектрі умов. Це значно підвищує стійкість політики при реальному виконанні.

У військовій логістиці критично важливою є оптимізація енергоспоживання. Функція витрат $C(s, a)$ у CMDP може бути пов'язана з моделлю енергоспоживання дрона:

$$E_{total} = E_{hover} \cdot t_{hover} + E_{cruise} \cdot t_{cruise} + E_{maneuver} \cdot t_{maneuver}, \tag{6}$$

де кожен доданок описує витрати енергії у відповідному режимі польоту. Інтеграція цієї моделі у процес оптимізації дозволяє формувати маршрути, що мінімізують ризик передчасного завершення місії через розряд акумулятора.

Використання БпЛА у військовій логістиці передбачає постійну взаємодію з агресивним середовищем, де присутні цілеспрямовані дії противника, спрямовані на виведення апаратів з ладу. Це може включати фізичне ураження (засоби ППО, снайперські постріли), радіоелектронну боротьбу (глушіння та спуфінг GPS), кібернетичні атаки на системи управління. У термінах RL ці загрози можна моделювати як стохастичні переходи у моделі MDP:

$$P'(s' \mid s, a) = (1 - \eta)P(s' \mid s, a) + \eta Q(s' \mid s, a), \tag{7}$$

де η – інтенсивність втручання противника, а Q – розподіл станів, що виникають внаслідок атак. У цьому випадку класична оптимізація середньої винагороди є недостатньою, оскільки політика повинна бути стійкою до несприятливих умов. З цією метою у літературі [7, 8] активно розвивається напрям robust RL, у якому оптимізація політики формулюється як:

$$\pi^* = \arg\max_{\pi} \min_{P' \in P} \mathbb{E}_{p'\pi}\left[\sum_{i=0}^{\infty} \gamma^t R(s_t, a_t\right], \tag{8}$$

де P – множина можливих збурених моделей переходів. Такий підхід дозволяє зменшити ризик катастрофічних відмов, навіть якщо противник змінює умови середовища.

У багатьох бойових сценаріях критично важливим є збереження комунікаційного каналу між дронами та командним центром. RL-агенти можуть навчатися вибирати такі траєкторії, які не лише оптимізують доставку, але й підтримують зв'язок, використовуючи релейні дрони як мобільні ретранслятори. У цьому випадку модель винагороди може містити додатковий член:

$$R_{comm} = -\mu \cdot \max(0, L_{\max} - d_{link}), \tag{9}$$

де $L_{\max}$ – допустима дальність зв'язку, d_{link} – відстань між вузлами мережі, μ – коефіцієнт штрафу. Це стимулює агентів залишатися в межах зони зв'язку, навіть якщо це збільшує шлях доставки. Дослідження [6, 10] показують, що таке поєднання логістичних і комунікаційних обмежень особливо ефективне у координації гетерогенних флотів БпЛА.

В умовах бойових дій попередньо обчислена траєкторія може стати неактуальною вже через кілька хвилин, наприклад, через знищення мосту, зміну ворожих позицій або поява нових зон ураження. Тут на перший план виходить online-RL навчання та meta-RL. Meta-RL [4, 9] дозволяє агенту швидко адаптувати політику до нових умов, використовуючи попередній досвід. Формально це описується як оптимізація початкових параметрів θ політики π_θ, які можна швидко донавчити для нового завдання T_i:

$$\theta^* = \arg\max_{\theta} \sum_{T_i \sim p(T)} \mathbb{E}_{\pi_{\theta'}}\left[R\right], \quad \partial e\, \theta' = \theta - \alpha \nabla_\theta L_{T_i}\left(\pi_\theta\right). \tag{10}$$

Цей підхід дає змогу перебудовувати логістичну мережу у відповідь на зміну лінії фронту або інфраструктури.

Реальні бойові випробування RL-політик є небезпечними та дорогими, тому використання симуляторів є обов'язковим етапом. Проте симулятори повинні відображати не лише фізику польоту, але й моделювати логістичні процеси та поведінку противника. Деякі роботи [10] інтегрують симулятори, де RL-агенти одночасно вирішують задачі:

- вибору маршруту з урахуванням ризиків;
- координації з іншими дронами;
- підтримки зв'язку;
- економії енергії.

Після етапу симуляцій проводиться обмежене тестування в польових умовах, що зменшує розрив між симуляцією та реальністю.

Висновки

Проведений огляд наукових досліджень показав, що навчання з підкріпленням є перспективним інструментом для вирішення задач військової логістики з використанням безпілотних літальних апаратів. Його ключові переваги полягають у здатності до автономного прийняття рішень у складних та динамічних умовах, адаптації до непередбачуваних змін обстановки, координації дій у багатоагентних системах та оптимізації використання обмежених ресурсів, зокрема енергетичних. У роботі систематизовано підходи до моделювання логістичних задач у термінах MDP та CMDP, розглянуто методи ризик-орієнтованої оптимізації політик, інтеграцію прогнозних моделей загроз, використання багатоагентного навчання та гібридних систем управління. Особливу увагу приділено симуляційним середовищам і підходу sim-to-real, який дозволяє мінімізувати розрив між віртуальним навчанням і реальною експлуатацією.

Перспективними напрямами подальших досліджень є:

- розробка більш точних симуляційних моделей, що враховують як фізичні, так і тактичні фактори;
- інтеграція RL із прогнозними та аналітичними модулями для підвищення ефективності;
- створення гібридних архітектур, що поєднують гнучкість RL із надійністю класичних методів управління;
- застосування методів meta-RL для швидкої адаптації до нових умов операцій.

Список використаної літератури

1. Sutton R. S., Barto A. G. Reinforcement Learning: An Introduction. Cambridge : MIT Press, 2018. 548 p.
2. Abbeel P., Coates A., Ng A. Y. Autonomous helicopter aerobatics through apprenticeship learning. International Journal of Robotics Research. 2010. Vol. 29, No. 13. P. 1608–1639. DOI: 10.1177/0278364910371999
3. Lowe R., Wu Y., Tamar A., Harb J., Abbeel P., Mordatch I. Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. Advances in Neural Information Processing Systems (NeurIPS). 2017. P. 6379–6390.
4. Foerster J., Farquhar G., Afouras T., Nardelli N., Whiteson S. Counterfactual multi-agent policy gradients. Proceedings of the AAAI Conference on Artificial Intelligence. 2018. Vol. 32, No. 1. P. 2974–2982.
5. Achiam J., Held D., Tamar A., Abbeel P. Constrained Policy Optimization. Proceedings of the 34th International Conference on Machine Learning (ICML). 2017. P. 22–31.
6. Cohen M. H., Belta C. Safe exploration in model-based reinforcement learning using control barrier functions. Automatica. 2023. Vol. 147. Art. 110684. DOI: 10.1016/j.automatica.2022.110684
7. Chow Y., Tamar A., Mannor S., Pavone M. Risk-sensitive and robust decision-making: a CVaR optimization approach. Advances in Neural Information Processing Systems (NeurIPS). 2015. P. 1522–1530.
8. Chen S., Mo Y., Wu X., Xiao J., Liu Q. Reinforcement Learning-Based Energy-Saving Path Planning for UAVs in Turbulent Wind. Electronics. 2024. Vol. 13, No. 16. Art. 3190. DOI: 10.3390/electronics13163190
9. Khalil E., Dai H., Zhang Y., Dilkina B., Song L. Learning combinatorial optimization algorithms over graphs. Advances in Neural Information Processing Systems (NeurIPS). 2017. P. 6348–6358.
10. Akkaya I., Andrychowicz M., Chociej M., et al. Solving Rubik's Cube with a Robot Hand. arXiv preprint. arXiv:1910.07113. 2019.

References

1. Sutton, R. S., & Barto, A. G. (2018). *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.
2. Abbeel, P., Coates, A., & Ng, A. Y. (2010). Autonomous helicopter aerobatics through apprenticeship learning. *International Journal of Robotics Research, 29*(13), 1608–1639. https://doi.org/10.1177/0278364910371999
3. Lowe, R., Wu, Y., Tamar, A., Harb, J., Abbeel, P., & Mordatch, I. (2017). Multi-Agent Actor-Critic for Mixed Cooperative-Competitive Environments. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 6379–6390).
4. Foerster, J., Farquhar, G., Afouras, T., Nardelli, N., & Whiteson, S. (2018). Counterfactual multi-agent policy gradients. *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 32, No. 1, pp. 2974–2982).

5. Achiam, J., Held, D., Tamar, A., & Abbeel, P. (2017). Constrained Policy Optimization. In *Proceedings of the 34th International Conference on Machine Learning* (pp. 22–31).

6. Cohen, M. H., & Belta, C. (2023). Safe exploration in model-based reinforcement learning using control barrier functions. Automatica, 147, 110684. https://doi.org/10.1016/j.automatica.2022.110684

7. Chow, Y., Tamar, A., Mannor, S., & Pavone, M. (2015). Risk-sensitive and robust decision-making: a CVaR optimization approach. In *Advances in Neural Information Processing Systems* (Vol. 28, pp. 1522–1530).

8. Chen, S., Mo, Y., Wu, X., Xiao, J., & Liu, Q. (2024). Reinforcement Learning-Based Energy-Saving Path Planning for UAVs in Turbulent Wind. Electronics, 13(16), 3190. https://doi.org/10.3390/electronics13163190

9. Khalil, E., Dai, H., Zhang, Y., Dilkina, B., & Song, L. (2017). Learning combinatorial optimization algorithms over graphs. In *Advances in Neural Information Processing Systems* (Vol. 30, pp. 6348–6358).

10. Akkaya, I., Andrychowicz, M., Chociej, M., et al. (2019). Solving Rubik's Cube with a Robot Hand. *arXiv*. https://arxiv.org/abs/1910.07113

Дата першого надходження рукопису до видання: 25.09.2025
Дата прийнятого до друку рукопису після рецензування: 22.10.2025
Дата публікації:

UDC 004.89 DOI

M. M. KORABLYOV
Doctor of Technical Sciences, Professor,
Professor at the Department of Computer Intelligent Technologies and Systems
Kharkiv National University of Radio Electronics
ORCID: 0009-0005-2540-7741

O. O. FOMICHOV
Candidate of Technical Sciences,
Senior Lecturer at the Department of Electronic Computer Devices
Kharkiv National University of Radio Electronics
ORCID: 0000-0001-9273-9862

O. K. TKACHUK
Postgraduate Student at the Department of Computer Intelligent Technologies and Systems
Kharkiv National University of Radio Electronics
ORCID: 0009-0006-2943-9887

CONTROLLING BOTS' BEHAVIOR IN RPG GAMES USING THE IMMUNE MODEL OF CLONAL SELECTION

The development of game character control models that use artificial intelligence methods is extremely important. The issues of creating intelligent game bots that allow simulating the behavior of players in an RPG (Role-Playing Game) are considered. With the help of the game design created of the project, a futuristic world is simulated in which the player controls spaceships that compete with other similar spaceships. Various options for attacking or restoring ships are used according to the completed description of their capabilities, where the main attention is focused on the features of their action. The game provides for the implementation of several factions, which allow you to create different types of spaceships and give them specific properties. Each ship from the user's team, as well as from the enemy's team, can have different ranks, which significantly impact both the parameters and the cost of this ship. The game application uses the theory of artificial immune systems (AIS) for effective control of game bots, namely the immune model of clonal selection. For this purpose, a modified immune model of clonal selection, Clonalg-rpg, is proposed, in which the antibodies of the system are bots, and the antigens foreign to the system are ships from the project user team that can interact with antibodies. By executing recovery commands for both the bots of their team and for themselves, the bots have the opportunity to interact with each other. Using specific immune operators that allow performing appropriate actions with the population of antigens and antibodies, the work of the modified Clonalg-rpg algorithm is described. To analyze its effectiveness, several game sessions were conducted, in which different compositions of both user ship teams and bot teams were changed. During the game sessions, teams were formed using specific settings, and not randomly. At the same time, spaceships were divided into three classes according to their parameters and capabilities. The results of the experimental studies showed that the proposed immune model Clonalg-rpg is effective and simple to implement, which makes it possible to use it in other game genres.

Key words: *immune model, clonal selection, spaceship, combat, control, design, game bot.*

М. М. КОРАБЛЬОВ
доктор технічних наук, професор,
професор кафедри комп'ютерних інтелектуальних технологій та систем
Харківський національний університет радіоелектроніки
ORCID: 0009-0005-2540-7741

О. О. ФОМІЧОВ
кандидат технічних наук,
старший викладач кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0000-0001-9273-9862

© Korablyov M. M., Fomichov O. O., Tkachuk O. K., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

О. К. ТКАЧУК
аспірант кафедри комп'ютерних інтелектуальних технологій та систем
Харківський національний університет радіоелектроніки
ORCID: 0009-0006-2943-9887

КЕРУВАННЯ ПОВЕДІНКОЮ БОТІВ У RPG-ІГРАХ З ВИКОРИСТАННЯ ІМУННОЇ МОДЕЛІ КЛОНАЛЬНОГО ВІДБОРУ

Розробка моделей керування ігровими персонажами, які використовують методи штучного інтелекту, є надзвичайно важливою. Розглядаються питання створення інтелектуальних ігрових ботів, які дозволяють імітувати поведінку гравців у RPG-грі (Role-Playing Game). За допомогою створеного ігрового дизайну проекту моделюється футуристичний світ, в якому гравець керує космічними кораблями, що змагаються із іншими подібними космічними кораблями. Використовуються різні варіанти атаки або відновлення кораблів згідно виконаного опису їх можливостей, де основна увага зосереджена на особливостях їх дії. У грі передбачається реалізація декількох фракцій, які дозволяють створювати різні типи космічних кораблів та надавати їм специфічні властивості. Кожен корабель з боку команди користувача, а також з боку команди супротивника, може мати різні ранги, які мають суттєвий вплив як на параметри, так і на вартість цього корабля. В ігровому додатку використана теорія штучних імунних систем (ШІС) для ефективного керування ігровими ботами, а саме імунна модель клонального відбору. З цією метою запропонована модифікована імунна модель клонального відбору Clonalg-rpg, в якій антитілами системи являються боти, а чужорідними для системи антигенами є кораблі з команди користувача проекту, які можуть взаємодіяти з антитілами. Виконуючи команди відновлення як для ботів своєї команди, так і для самих себе, боти мають можливість взаємодіяти між собою. З використанням специфічних імунних операторів, які дозволяють виконувати відповідні дії із популяцією антигенів та антитіл, описана робота модифікованого алгоритму Clonalg-rpg. Щоб проаналізувати його ефективність були проведені декілька ігрових сесій, в яких змінювалися різні склади як команд кораблів користувача, так і команд ботів. В ході ігрових сесій команди формувалися з використанням специфічних налаштувань, а не випадковим чином. При цьому космічні кораблі були поділені на три класи згідно своїх параметрів та можливостей. Результати проведених експериментальних досліджень показали, що запропонована імунна модель Clonalg-rpg є ефективною та простою в реалізації, що робить можливим її використання в інших ігрових жанрах.

Ключові слова: *імунна модель, клональний відбір, космічний корабель, бій, керування, дизайн, ігровий бот.*

Statement of the problem

The rapid development and spread of gaming applications have significantly influenced the behavior, skills, and habits of society. This has led to the spread of game mechanics and features in the organization of work of many mobile and web applications, which has been called gamification [1, 2]. Elements of gamification can now be observed in many areas of modern life – from food ordering and delivery services and distance learning systems to mobile banking applications [3]. Mastering the basic techniques is crucial to creating engaging and successful PC games. The foundation of every successful PC game is a well-thought-out game design that includes story, game mechanics, characters, and environments that work together to create an engaging gameplay experience. Game design is considered the most important aspect of computer game development.

Recent years have been characterized by the rapid development of artificial intelligence systems and the spread of free access to such systems by society for solving various everyday and creative tasks, in particular in gaming applications [4, 5]. The field of artificial intelligence has ceased to be exclusively research and scientific, and has become popular and accessible to everyone, similar to what happened on the computer market in the mid-1990s. Most existing systems of artificial intelligence organization are based on the biological principles of the human nervous or immune systems, as well as the principles of evolutionary research on the genome and behavior of animals and ants. Artificial intelligence has also significantly influenced the development of the modern gaming industry [6, 7].

The work sets the task of creating intelligent game bots that simulate the behavior of players in an RPG (Role-Playing Game) game when countering the actions of a team of user characters, using artificial immune systems, in particular, the immune model of clonal selection. Solving this task will allow us to create an intelligent system for controlling the behavior of bots in a computer game.

Game design project

The game application considered in the work simulates a futuristic world in the style of open space, where the player controls a group of spacecraft, the number of which ranges from 1 to 6 units, which compete with another group of similar spacecraft of a similar number [8]. The principle of game combat is borrowed from the popular RPG game Disciples II in the mid-2000s, but with a significantly simplified game design. The game provides for several game factions that can produce their own specific types of spacecraft with special capabilities and properties. At the same time, each spacecraft in the user's team or in the team of his opponent has a list of characteristics that determine its capabilities: cost, ship level, combat experience, strength, energy, attack, protection, speed, and recovery.

The game provides for several factions that create their own types of spaceships and give them specific properties. For ease of implementation, the factions are conventionally designated by colors, namely: yellow, blue, and red. The main feature of the ships of the yellow faction is an increase in the strength indicator by 10 %, compared to ships of other factions; the feature of the ships of the blue faction is an increase in the protection indicator by 10 %, compared to ships of other factions; and the main feature of the ships of the red faction is an increase in the recovery indicator by 10 %, compared to ships of other factions. Thanks to this, each faction can increase the significant characteristics of its ships uniquely compared to the ships of other factions. In addition, each ship, both from the user's team and from the opponent's team, can belong to one of the ranks. These ranks significantly affect the cost and parameters of this ship. The list of main ranks is defined as follows: support, scout, cruiser, frigate, destroyer, and dreadnought.

To describe the capabilities of ships, various attack or recovery options were used, where attention was focused on the features of their action from the point of view of game design. First, two types of attack were used in the description of the capabilities of ships: close and long-range attacks. In this case, a close attack implies the ability of the ship to attack only those enemy ships that are on the first line of the battlefield in front of it, i.e. on positions 2, 4 and 6 of the game battle map (Fig. 1). A long-range attack implies the ability to attack enemy ships that are on the second line in the enemy team, i.e. on positions 1, 3 and 5 of the game battle map.

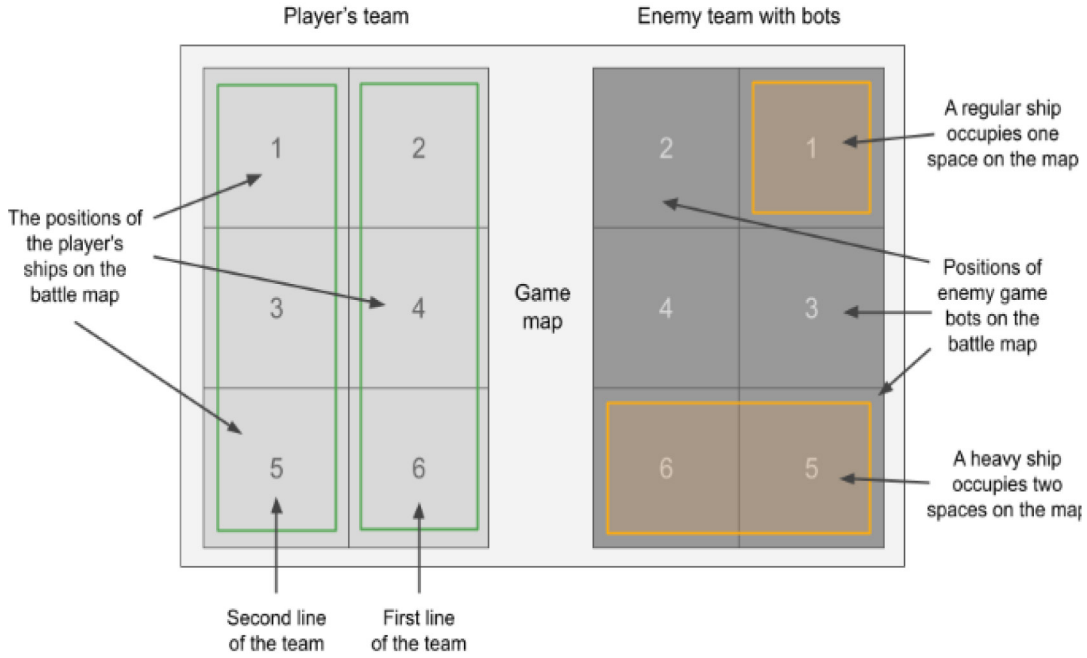


Fig. 1. General view of the game battle map

The definition of the method of attack between ships has a significant impact on the gameplay [9, 10]. Accordingly, and also taking into account the defined types of close and long-range attacks, it is necessary to describe the appearance of the playing field for the battle between ship teams. The entire playing field is conventionally divided into three parts, where the first part is occupied by the first player's ship team, the second part divides the teams among themselves, and the third part is occupied by the second player's team. It should be noted that in each team, the ships are lined up in two rows – the one closest to the enemy and the one far from the enemy, and only the ships that occupy two places on the game battle map occupy both rows, that is, they can be attacked by both close and long-range attacks. In addition, in the case of the destruction of all ships of the first line, all ships that are on the second line can be attacked by any method of attack.

In the game, a great influence on the capabilities of a ship is its level, namely, it increases the strength, defense, and attack indicators by 1 % of the initial values of a ship of a certain type of the first level, taking into account the characteristics of the faction. The level increase occurs due to the victory over the enemy team, where for every 100 units of strength of each destroyed ship of the enemy team, 10 units of experience are accrued, which are evenly distributed between all ships that participated in the battle. In addition, the method of assigning damage from attacks between ships, which is determined by the expression:

$$damage = \frac{e_1}{e_2} \cdot (A_1 - D_2), \tag{1}$$

where *damage* is the value of the adjusted damage that a ship inflicts on another during the attack; A_1 is the attack value of ship 1; e_1 is the energy value of ship 1, D_2 is the defense value of ship 2; e_2 is the energy value of ship 2.

An important influence on the gameplay is the way a game ship can be restored, which can be done during its turn instead of performing an attack. This indicator is determined as follows:

$$rp = R \cdot ke, \tag{2}$$

where rp – value of the restored strength of the ship; R – characteristic of ship recovery; e – ship energy value, a k – recovery factor, which is chosen randomly from the range [1.1; 1.25].

An important feature of the gameplay is the Player versus Enemy (PvE) mode, i.e., a game where the user plays against a team controlled by the game itself.

Controlling gaming bots based on the immune model of clonal selection

The task of controlling game bots in a battle with a team of user ships is reduced to modeling the behavior of each ship of the enemy team, taking into account the capabilities and characteristics of each specific ship of the player's team. Accordingly, game bots are only ships of the game team/system, with which the user's team is fighting. For effective control of game bots, it is currently advisable to use artificial intelligence systems, among which, along with neural networks, fuzzy logic, genetic algorithms, etc., the theory of artificial immune systems (AIS) occupies an important place. The most common AIS models used to solve various practical problems are the immune model of clonal selection and the artificial immune network model [11].

Controlling game bots can be considered as a special case of the optimization problem, for the solution of which the immune model of clonal selection can be used [12, 13]. According to this model, bots are considered as antibodies of the system that interact with foreign antigens to the system, which for bots are ships from the project user team. In addition, bots can interact with each other, executing the recovery command for themselves and for the bots of their team.

The ClonAlg method is one of the most common methods that operates on the basis of the clonal selection model. This method can be used to solve the problem of controlling game bots that oppose the user's ship team, which controls them personally. In this method, the method that will control the bots plays a key role. The classic version of the ClonAlg immune algorithm provides for the possibility of changing antibody parameters during their interaction with a population of foreign antigens by cloning, mutation, and self-regulation of the antibody population through clonal selection and aging. In this case, changes in antigen parameters were either not foreseen at all or were insignificant. In addition, the possibility of antigen disappearance in the process of forming an immune response by the system was not foreseen as such. But in our case, during the confrontation of the user's ship team with the system's game bot team, such a possibility is foreseen.

With this method of operation of the immune algorithm, the destruction of all antigens that represent the user's ship team is the only thing that can be used to solve this problem without a number of changes. Accordingly, the modified version of this algorithm, Clonalg-rpg, has several important features. First, the algorithm cannot directly change the values of the antibody features during its operation, but can only make decisions about the appropriateness of executing a particular command for each specific bot or its clone. Secondly, the choice of a solution for executing a particular bot command should be made from the point of view of the maximum efficiency of this action of destroying the entire population of antigens, or from the point of view of restoring the strength indicator of the entire population of antibodies. Thirdly, the condition for terminating the operation of the algorithm Clonalg-rpg is the destruction of the entire population of antigens, represented by the set of ships of the game user, before the population of antibodies, represented by game bots, ceases to exist. Fourth, dynamic changes in the features of antibody bots and antigen ships occur constantly at the user's command. Thus, the implementation of the modified version of the immune algorithm Clonalg-rpg should be based on four defined principles as the basis of the algorithm's work. In addition, an important feature of the clonal selection model is that in the process of work, antibodies focus on interaction with antigens, and then use some opportunities for interaction with other antibodies in the process of self-regulation of the antibody population.

The operation of the modified Clonalg-rpg method, created to solve the problem of controlling game bots, can be conditionally described at the level of immune operators as follows:

$$Clonalg - rpg = \begin{bmatrix} Presentation(AB, AG) \rightarrow \\ Cloning(AB, CL) \rightarrow \\ Mutation(AB, CL) \rightarrow \\ ClonSelection(AG, CL) \rightarrow \\ SelfRegulation(AB, CL) \rightarrow \\ Termination(AB, AG) \end{bmatrix}^{AG>0}, \tag{3}$$

where *Presentation*(*AB*, *AG*) is the operator for presenting the set of antigens *AG* of the antibody population *AB*; *Cloning*(*AB*, *CL*) is the operator for cloning the antibody population; *Mutation*(*AB*, *CL*) is the operator for mutation of the formed clones; *ClonSelection*(*AG*, *CL*) is the operator for selecting clones; *SelfRegulation*(*AB*, *CL*) is the operator for self-regulation of the antibody population; *Termination*(*AB*, *AG*) is the operator for checking the possibility of stopping the algorithm.

The work of the *Presentation*(*AB*, *AG*) operator is to determine the affinities between antibodies and antigens, as well as the affinities between the population of antibodies of the AIS. At the same time, affinity is defined as the degree of similarity between the features of an antibody and an antigen, and the work of the immune algorithm is reduced to ensuring that, during the process of cloning and mutation, the clones reproduce the features of a particular antigen. The determination of antigen affinity is described by the expression:

$$affAg_i = \frac{1}{1+\left(Ds_i - Dc_i\right)+\left(Es_i - Ec_i\right)}, \tag{4}$$

where$affAg_i$ is the affinity of the *i*-th antigen; Ds_i is the saved strength indicator of the i-th antigen at the start of the battle; Dc_i is the current strength indicator of the *i*-th antigen at the current battle iteration; Es_i is the saved energy indicator of the *i*-th antigen at the start of the battle; Ec_i is the current energy indicator of the *i*-th antigen at the current battle iteration.

The antibody cloning operator *Cloning*(*AB*, *CL*) is used to create a set of clones for each antibody in the current AIS population for their further mutation. The number of clones produced by using the adaptive cloning operator is defined as:

$$n_i = C_i \cdot N_{AG}, \tag{5}$$

where n_i is the number of clones of the *i*-th antibody; C_i is the number of interaction variants between the antibody and the antigen; N_{AG} is the number of antigens.

The clone mutation operator Mutation(CL) significantly affects the speed of solving the target problem set for the AIS. It is the mutation mechanism that ensures changes are made to immune objects and brings the AIS closer to solving the problem. The Clonalg-rpg algorithm assumes the use of an adaptive mutation operator, which focuses on choosing the method of interaction with antigens.

According to the clone selection operator *ClonSelection*(*AG*, *CL*), one immune object, or one way of interacting with an active antigen, or a set of antigens, is selected, which makes the greatest contribution to solving the target problem.

The antibody population self-regulation operator *SelfRegulation*(*AB*, *CL*) has been added to the clonalg-rpg algorithm instead of the apoptosis operator.

The antibody population self-regulation operator *SelfRegulation*(*AB*, *CL*) has been added to the Clonalg-rpg algorithm instead of the apoptosis operator, which involves determining affinities or other metrics between the antibody population and the clones selected in the clonal selection process. The determination of the internal affinity index of a clone to the entire antibody population has the form:

$$aff'Ab_i = \frac{1}{S}\sum_{i=1}^{S}\left(1+\left(Ds_i - \left(Dc_i + \frac{1}{S}R\right)\right)\right)^{-1}, \tag{6}$$

where $aff'Ab_i$ is the affinity of the clone with the entire population of active antibodies, including the *i*-th antibody from which this clone was created; *S* is the size of the population of active antibodies; Dc_i is the strength indicator of the *i*-th antibody at the current iteration of the battle; *R* is the clone recovery indicator.

The possible mode of action for restoring the strength characteristic for a single antibody, or the entire antibody population, is selected according to the value of the greater affinity. In the case where a game bot has only one recovery method, the affinity between it and the active antibodies is determined only once.

The operator for checking the possibility of stopping the battle *Termination*(*AB*, *AG*) is used to check the possibility of stopping the game battle process. The Cloanlg-rpg algorithm uses a criterion operator for stopping the operation of the AIS, which involves stopping the operation of immune operators in the absence of active antigens or active antibodies.

Experimental studies

When studying the effectiveness of the Clonalg-rpg algorithm, about a hundred tests were conducted with different compositions of user teams and teams of game bots of different types and different factions. During these tests, teams were formed not randomly, but through specific settings, taking into account that spaceships can be divided into three classes according to their parameters and capabilities, where the first class is formed by the cheapest *support* and *scout* ships, the second class is formed by *cruiser* and *frigate* ships, and the third most expensive class is formed by heavy ships, which occupy two places on the game battle map – *destroyer* and *dreadnought*. At the same time, the parameters of these ships, such as attack power, energy, durability, etc. in the game balance are adjusted in such a way that first-class ships are inferior to second-class ships by approximately 25 %, and to third-class ships by approximately 50 %. Accordingly, second-class ships are inferior to third-class ships by 25 % in their basic indicators. However, for the sake of game balance, first-class ships outperform all other ships in terms of speed and make their move during a game battle first.

The Clonalg-rpg algorithm was tested on the same set of input data about the user's ship commands and game bots in three ways: without using the AIS, when the user controls both commands directly (this is an atypical way of working of the project, which is configured in a specific way); without using the AIS by choosing the game bot's action method randomly; and with using the Clonalg-rpg algorithm, which controls the behavior of game bots in battle. The main indicator of the effectiveness of the immune algorithm for controlling game bots is the number of victories of the AIS over

a person in game battles in which teams of ships with comparable or equal basic parameters participate. In addition, the effectiveness study took into account not only the percentage of victories of the AIS over the user, but also the number of ships remaining in the winning team. Accordingly, the following indicators were identified:

– the winner has 1 ship left;
– the winner has 2 ships left;
– the winner has 3 or more ships left.

Using such detail will allow you to investigate not only the fact of victory itself, but also the extent to which the winner dominated his opponent during the game battle, which may indicate the level of skill of the game. Table 1 shows the percentage of victories of the user's teams over their opponent's teams, taking into account the different ways of controlling the opponent's team, defined earlier.

Table 1

Frequency of user wins over opponent

Types of commands	Manual control	Random control	Control Clonalg-rpg
5 on 5	50 %	85,15 %	50,35 %
4 on 4	50 %	96,75 %	49,15 %

For each type of control of the team of game bots, 20 game battles were conducted to obtain statistical data on the results of battles and the choice of the type of actions of bots during the game battle. According to the results presented above, the least effective is the control of game bots made randomly, which almost always led to a guaranteed defeat. In contrast, the organization of game bot control based on the Clonalg-rpg algorithm showed results comparable to those obtained by involving a person to control the enemy ships. When reducing the number of ships in the team by using heavy ships, which have a greater variability of possible actions, the chance of victory of game bots increases slightly, which may indicate that in the case of increasing the number of attack and recovery options, the proposed immune method Clonalg-rpg will show greater efficiency than the user.

In order to assess the playing skill of the winning team, the number of ships remaining in the winning team after the game battle was recorded; these results are shown in Table 2. It can be seen that using the method of controlling game bots randomly creates the easiest conditions for the user, because in most cases, it was the user who won over the bot team, which is controlled randomly.

Table 2

Number of ships remaining in the winning team

Types of commands		Manual control	Random control	Control Clonalg-rpg
5 on 5	1	100,00 %	12,40 %	93,75 %
	2	0,00 %	82,35 %	6,25 %
	3+	0,00 %	5,25 %	0,00 %
4 on 4	1	100,00 %	3,60 %	95,00 %
	2	0,00 %	88,25 %	5,00 %
	3+	0,00 %	8,15 %	0,00 %

We can conclude that the use of the Clonalg-rpg algorithm creates difficult conditions for the user to win over a team of game bots, with the winning team usually having only one ship left, while the others are lost during the game battle.

A separate area of observation was the type of action that the game bot chooses and performs during its turn. This choice plays an important role, as it is what leads the game bot team to victory over the user's team or defeat. When allowed to make a move in the order of the queue, which is formed based on the speed of all active ships from both teams participating in the game battle, you can choose either to attack the enemy or to recover. It is impossible for either the game bot or the ship that is under the direct control of the user to miss a move. Table 3 shows statistical data on the frequency of choosing a particular method of action by game bots that are under the different types of control defined earlier.

It can be concluded that the immune algorithm Clonalg-rpg gives a greater advantage to the III method of attack, both at close and long distances, while minimizing the use of the II method of distributed attack of the enemy team's ships. In the case of a game bot's decision to attack the enemy under the guidance of Clonalg-rpg, the most attention is paid to methods of concentrated attack on a specific ship with the possibility of inflicting additional energy damage to a ship from the enemy team, which significantly affects its further capabilities in battle. In the case of recovery, the method of concentrated recovery of one selected ship is also preferred. This behavior can be described as a strategy of concentrated destruction of the weakest ships of the enemy team while maximally concentrated recovery of damaged ships of one's own team.

Table 3

Frequency of choosing methods of action in team battles

Variants of ship actions	Manual control		Random control		Control Clonalg-rpg	
	5 on 5	4 on 4	5 on 5	4 на 4	5 on 5	4 on 4
Method I of the close attack	19,33 %	9,82 %	24,56 %	18,89 %	20,29 %	11,01 %
Method II of the close attack	9,83 %	10,42 %	11,22 %	10,56 %	10,86 %	8,57 %
Method III of the close attack	3,15 %	7,89 %	2,22 %	5,56 %	3,90 %	9,76 %
Method I of long-range attack	22,00 %	6,05 %	17,89 %	10,56 %	20,00 %	6,25 %
Method II of long-range attack	5,83 %	10,62 %	6,22 %	10,56 %	4,86 %	8,57 %
Method III of long-range attack	3,15 %	7.89 %	2,22 %	5,56 %	3,90 %	9,76 %
Method I of recovery	19,87 %	18,65 %	24,56 %	18,89 %	19,67 %	18,75 %
Method II of recovery	14,02 %	22,42 %	8,89 %	13,89 %	14,52 %	22,32 %
Method III of recovery	2,50 %	6,25 %	2,22 %	5,56 %	2,00 %	5,00 %

It should be noted that the proposed Clonalg-rpg method is used to control game bots in RPG projects for the first time, so today it is quite difficult to make a comparison with other immune and non-immune methods used to organize a game bot control system. It is also important to note that the proposed immune method is easy to implement and modify, which makes it possible to further use it to control game bots in games of other genres.

Conclusions

The solution of the current problem of controlling game bots in RPG games using a modified immune model of clonal selection is considered. A game application has been developed that simulates a futuristic world in the style of open space, where the player controls a group of spacecraft that compete with another such group of similar spacecraft. At the same time, each spacecraft in the user's team or in the team of his opponent has a list of specific characteristics that determine its capabilities. The game provides for several factions that create their own types of spacecraft and endow them with specific properties. Thanks to this, each faction can increase the significant characteristics of its ships in a unique way compared to the ships of other factions. In addition, each ship, both from the user's team and from the side of its opponent's team, can belong to one of the ranks, which significantly affect the cost and parameters of this ship. To describe the capabilities of the ships, various attack or recovery options were used, which focused on the features of their action from the point of view of game design.

Game bot control is considered as a special case of the optimization problem, for the solution of which the modified immune model of clonal selection Clonalg-rpg is used, according to which bots are considered as antibodies of the system that interact with foreign antigens to the system, which for bots are ships from the project user team. The work of the modified algorithm, Clonalg-rpg is described at the level of immune operators that perform the corresponding actions with the population of antibodies or antigens representing the bots of the system and the user.

During the experimental study of the effectiveness of the Clonalg-rpg algorithm, about a hundred tests were performed with different compositions of user teams and teams of game bots of different types and different factions. During these tests, the teams were formed not randomly, but through specific settings, taking into account that the spaceships were divided into three classes according to their parameters and capabilities. The results of the experimental studies showed that the proposed immune model Clonalg-rpg is effective and simple to implement, which makes it possible to use it in other game genres.

References

1. Panchanadikar R., Freeman G., Li L., Schulenberg K., & Hu Y. (2024). A New Golden Era' or 'Slap Comps': How Non-Profit Driven Indie Game Developers Perceive the Emerging Role of Generative AI in Game Development." Extended Abstracts of the CHI Conference on Human Factors in Computing Systems. NY, USA: ACM. 1–7. https://doi.org/10.1145/3613905.3650845

2. Immonen J. (2024). Primary Tools and Techniques in Game Development. Bachelor's thesis. Lappeenranta-Lahti University of Technology LUT, 43 pages. https://lutpub.lut.fi/bitstream/handle/10024/168411/

3. Szolin K., Kuss D. J., Nuyens F. M., & Griffiths M. D. (2023). "I am the character; the character is me": A thematic analysis of the user-avatar relationship in video games. Comp. in Hum. Behavior, Vol. 143, 107694. https://doi.org/10.1016/j.chb.2023.107694

4. Quiroga M. A., Diaz A., Román F.J., Privado J., & Colom R. (2019). Intelligence and video games: Beyond "brain-games". Intelligence, Volume 75, 85-94. URL: https://www.elsevier.com/locate/intell

5. Simons A., Wohlgenannt I., Zelt S., Weinmann M., Schneider J., and vom Brocke J. (2023). Intelligence at play: game-based assessment using a virtual-reality application. Springer https://doi.org/10.1007/s10055-023-00752-9

6. Lum Y., & Li W. (2022). Techniques and Paradigms in Modern Game AI Systems. Algorithms, 15(8), 282. https://doi.org/10.3390/a15080282

7. Fan X., Wu J., & Tian L. (2020). A Review of Artificial Intelligence for Games. Part of the Lecture Notes in Electrical Engineering book series (LNEE, vol. 572, 1821. URL: https://doi.org/10.1007/978-981-15-0187-6_34

8. Snyder R. (2023). 10 Best Spaceships in Video Games URL: https://www.dualshockers.com/best-spaceships-in-video-games/

9. Mohd T., Bravo-Garcia F., Love L., Gujadhur M., Nyadu J. (2023). Analyzing strengths and weaknesses of Modern Game Engines. International Journal of Computer Theory and Engineering, 15(1), 54–60. https://www.ijcte.org/vol15/IJCTE-V15N1-1330.pdf

10. Silić M., Džaja B., Rogulj R., & Turić H. (2024). Optimizing Game Ready Asset Creation Pipeline: From Concept to Implementation in Unreal Engine 5. ORGANIZER, 732. https://hdl.handle.net/11159/654467

11. Dasgupta D., Yu S., & Nino L. F. (2011). Recent Advances in Artificial Immune Systems: Models and Applications. Applied Soft Computing. Elsevier, 1574–1587. URL: https://https://doi.org/10.1016/j.asoc.2010.08.024

12. Korablyov M., Fomichov O., Chubukin O., Antonov D., & Dykyi S. (2023). Immune Model for Controlling Characters in Computer Games. XI International Scientific and Practical Conference "Information Control Systems and Technologies (ICST)", 214–226. https://ceur-ws.org/Vol-3513/

13. Shekhar S., Sharma D. K., Agarwal D. K., & Pathak Y. (2022). Artificial Immune Systems-Based Classification Model for Code-Mixed Social Media Data. IRBM, Vol. 43, Iss. 2, 120-129. URL: https://doi:10.1016/j.irbm.2020.07.004

Дата першого надходження рукопису до видання: 19.09.2025
Дата прийнятого до друку рукопису після рецензування: 15.10.2025
Дата публікації:

УДК 004.415.2 DOI

Я. І. КОРНАГА
доктор технічних наук, професор
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
ORCID: 0000-0001-9768-2615

О. М. ГУБАРЄВ
аспірант кафедри інформаційних систем та технологій
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
ORCID: 0009-0001-1028-4604

МЕТОД АВТОМАТИЗОВАНОГО ФОРМУВАННЯ КЛАСТЕРІВ ПРОГРАМНОЇ СИСТЕМИ НА ОСНОВІ СТРУКТУРНО-СЕМАНТИЧНОГО АНАЛІЗУ КОМПОНЕНТІВ З ВИКОРИСТАННЯМ НЕЙРОННИХ МЕРЕЖ

У статті розглянуто проблему трансформації монолітних архітектур програмних систем у мікросервісні, яка є одним із ключових завдань сучасної інженерії програмного забезпечення. Монолітні системи з часом набувають надмірної складності та високого рівня внутрішньої зв'язаності, що ускладнює їх масштабування, супровід і модифікацію. Мікросервісна архітектура, орієнтована на незалежні сервіси з чіткими інтерфейсами, значною мірою вирішує ці проблеми, однак постає нетривіальне завдання коректної декомпозиції, яке не має єдиного універсального розв'язку.

У роботі запропоновано метод декомпозиції, що поєднує структурний та семантичний аналіз вихідного коду з навчанням нейронних мереж для графів, спрямованих на автоматизоване формування кластерної структури системи. Особливістю підходу є багаторівнева побудова графа залежностей, яка враховує різні типи зв'язків (виклики методів, наслідування, використання ресурсів), а також запровадження вагових коефіцієнтів вузлів, розрахованих на основі показників центральності. Перед етапом кластеризації застосовано поетапну структурну оптимізацію графа, що включає виявлення сильно зв'язаних компонент, циклів, мостів і точок артикуляції, а також зниження впливу надлишкових і шумових зв'язків. Це дозволяє підвищити точність і надійність подальшого групування компонентів.

Додатково реалізовано адаптивний вибір алгоритму кластеризації на основі метрик графа (щільність, кількість вузлів, центральність, коефіцієнт спільного використання ресурсів), що дозволяє узгоджувати метод із поточними характеристиками структури. Навчання графової нейронної мережі здійснюється на векторах ознак, які інтегрують топологічні та семантичні характеристики компонентів, що забезпечує виявлення латентних закономірностей у складних залежностях.

Порівняльний аналіз із сучасними методами декомпозиції (CoGCN, DEEPLY, HyDEC, GDC-DVF, Mo2oM) підтвердив ефективність запропонованого підходу. Результати показали перевищення існуючих рішень за показником структурної модульності (SM), що відображає вищу когезію та чіткіші межі сформованих сервісів, при одночасному збереженні відносно низького рівня міжсервісних залежностей. Це свідчить про наукову новизну та практичну значущість розробленого методу, який може бути використаний як у дослідницькому середовищі, так і в реальних промислових системах.

Ключові слова: *монолітна архітектура; мікросервісна архітектура; декомпозиція; когезія; зв'язаність; графові моделі; семантичний аналіз; нейронні мережі*

YA. I. KORNAHA
Doctor of Technical Sciences, Professor
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"
ORCID: 0000-0001-9768-2615

O. M. HUBARIEV
Postgraduate Student at the Department of Information Systems and Technologies
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"
ORCID: 0009-0001-1028-4604

© Корнага Я. І., Губарєв О. М., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

A METHOD OF AUTOMATED FORMATION OF CLUSTERS OF THE SOFTWARE SYSTEM BASED ON STRUCTURAL-SEMANTIC ANALYSIS OF COMPONENTS USING NEURAL NETWORKS

The article examines the problem of transformation of monolithic architectures of software systems into microservice ones, which is one of the key tasks of modern software engineering. Monolithic systems eventually acquire excessive complexity and a high level of internal connectivity, making them difficult to scale, maintain and modify. Microservice architecture, oriented towards independent services with clear interfaces, largely solves these problems, however, the non-trivial task of correct decomposition arises, which does not have a single universal solution.

The paper proposes a decomposition method that combines structural and semantic analysis of source code with training of neural networks for graphs aimed at automated formation of the cluster structure of the system. A feature of the approach is the multi-level construction of the dependency graph, which takes into account different types of relationships (calls of methods, imitation, use of resources), as well as the introduction of node weights calculated on the basis of centrality indicators. Prior to the clustering step, a stepwise structural graph optimization is applied, involving the detection of strongly coupled components, cycles, bridges and articulation points, and the reduction of the influence of redundant and noise links. This makes it possible to increase the accuracy and reliability of further grouping of components.

In addition, an adaptive selection of the clustering algorithm based on graph metrics (density, number of nodes, centrality, resource sharing coefficient) is implemented, which allows you to match the method with the current characteristics of the structure. Graph neural network training is carried out on feature vectors that integrate topological and semantic characteristics of components, which ensures the detection of latent regularities in complex dependencies.

Comparative analysis with current decomposition methods (CoGCN, DEEPLY, HyDEC, GDC-DVF, Mo2oM) confirmed the effectiveness of the proposed approach. The results showed an excess of existing solutions according to the structural modularity indicator (SM), which reflects higher cohesion and clearer boundaries of the formed services, while maintaining a relatively low level of interservice dependencies. This shows the scientific novelty and practical significance of the developed method, which can be used both in the research environment and in real industrial systems.

Key words: *monolithic architecture; microservice architecture; decomposition; cohesion; connectivity; graph models; semantic analysis; neural networks.*

Постановка проблеми

Проблема переходу від монолітних систем на мікросервіси є однією з ключових у сучасній програмній інженерії. Моноліти, історично розроблені як єдині інтегровані системи, з часом набувають надмірної складності та високого рівня внутрішнього зв'язку, що ускладнює масштабування, модифікацію та підтримку. Мікросервісна архітектура, яка базується на незалежних компонентах з чіткими межами, вирішує ці проблеми, але вимагає правильного визначення меж між сервісами.

Задача декомпозиції моноліту на мікросервіси не має єдиного рішення: надто дрібна сегментація створює надмірний міжсервісний трафік і збільшує складність координації, тоді як надто грубе розбиття зберігає проблеми моноліту. Тому в наукових дослідженнях запропоновано ряд методів, від евристичних підходів на основі статичного аналізу до методів глибинного навчання нейронних мереж для графових структур. Серед сучасних рішень варто згадати CoGCN, DEEPLY, HyDec, GDC-DVF, і Mo2oM. Вони по-різному інтегрують структурний і семантичний аналіз, демонструючи різні компроміси між когезією та каплінгом.

Метою даної роботи є розробка та дослідження вдосконаленого методу декомпозиції, який поєднує статичний аналіз вихідного коду, оптимізацію графів залежностей, динамічну вибір алгоритму кластеризації, оптимізацію результатів кластеризації з урахуванням когезії, каплінгу та семантичної однорідності, а також навчання нейронної мережі на основі отриманих оптимізованих даних. Запропонований підхід спрямований на досягнення високої когезії сервісів з контрольованим рівнем міжсервісної зв'язності та семантичної однорідності, що є необхідною умовою ефективної експлуатації програмних систем.

Аналіз останніх досліджень і публікацій

CO-GCN (Clustering and Outlier-aware Graph Convolutional Network) – метод декомпозиції монолітних застосунків, який представляє систему у вигляді графа, де вузлами є класи, а ребрами – залежності викликів та спільна участь у виконанні точок входу. На основі цього графа формується матриця ознак, що поєднує структурні характеристики та семантичні властивості.

Особливістю підходу є урахування вузлів-викидів, які спотворюють структуру системи. Для цього у CO-GCN використовується функція втрат із трьох компонентів: втрати від структурних викидів, втрати від атрибутивних викидів та втрати від кластеризації. Навчання здійснюється за схемою почергової мінімізації, коли послідовно оновлюються параметри мережі, оцінки викидів і центри кластерів [1].

Метод було протестовано на типових Java-системах, зокрема DayTrader, де він продемонстрував вищу якість декомпозиції порівняно з базовими методами. Водночас автори відзначають, що масштабованість підходу обмежена через високу обчислювальну складність процедури навчання та залежність результатів від повноти й якості даних про транзакції [6].

DEEPLY – метод декомпозиції монолітних застосунків, що ґрунтується на глибинному аналізі вихідного коду та транзакцій користувачів із метою автоматизованого формування мікросервісів. Система подається у вигляді матриці, яка враховує як статичні залежності між класами, так і динамічні зв'язки, що виявляються під час виконання. На основі цих даних нейронна мережа навчається виявляти приховані закономірності та групувати компоненти у функціонально узгоджені сервіси.

Особливістю підходу є поєднання структурної інформації (виклики, спільні ресурси) та поведінкової інформації (шаблони виконання транзакцій). Це дозволяє моделі виділяти сервіси, які не лише мають внутрішню когезію, а й відповідають реальним сценаріям використання системи [2, 3].

Експерименти показали, що метод здатен перевершувати класичні евристичні підходи та конкурувати з експертними системами. Водночас обмеженням DEEPLY є висока потреба у даних трасування виконання, а також значні обчислювальні витрати при навчанні на великих системах [6].

HyDec– метод декомпозиції монолітних застосунків, що базується на ієрархічному застосуванні алгоритму DBSCAN для виявлення меж потенційних мікросервісів. Система представляється у вигляді графа залежностей між класами та транзакціями, після чого відбувається багаторівневе групування елементів із врахуванням щільності їхніх зв'язків. На відміну від класичного DBSCAN, який працює на одному рівні щільності, HyDec використовує ієрархічний підхід: спочатку формуються великі кластери, а потім вони поступово розділяються на менші підкласи доти, доки не досягається оптимальний баланс когезії та зв'язаності.

Ключовою особливістю підходу є здатність уникати проблем надмірного злиття або, навпаки, надмірної фрагментації сервісів. Алгоритм динамічно підлаштовує поріг щільності для різних ділянок графа, що дозволяє більш точно виявляти функціонально пов'язані компоненти [4].

Експерименти, проведені на типовому наборі Java-застосунків, показали, що HyDec здатен перевершувати як евристичні методи, так і окремі гібридні підходи, забезпечуючи кращий баланс між когезією та мінімізацією міжсервісних залежностей. Водночас метод чутливий до вибору параметрів початкового DBSCAN і може потребувати додаткового налаштування для систем із різною структурою [6].

GDC-DVF – метод виділення мікросервісів, що ґрунтується на глибокій кластеризації графів із використанням механізму «подвійного злиття уявлень» (dual view fusion). У даному підході програмна система моделюється як граф залежностей, який розглядається одночасно з двох точок зору: структурної (статичні зв'язки між класами, спільні ресурси, виклики) та атрибутивної (семантичні й поведінкові характеристики).

Ключова ідея методу полягає у поєднанні цих двох уявлень за допомогою механізму злиття, що дозволяє уникнути перекосів, властивих підходам, орієнтованим лише на один тип даних. Для цього застосовується глибока нейронна мережа кластеризації, яка навчається формувати узгоджені векторні подання вузлів і групувати їх у кластери, що відповідають потенційним мікросервісам [5].

Експерименти продемонстрували, що GDC-DVF забезпечує вищу якість декомпозиції порівняно з традиційними методами та низкою сучасних моделей, зокрема завдяки здатності адаптивно враховувати як структурні, так і семантичні ознаки. Водночас метод має підвищені вимоги до обсягів навчальних даних і є обчислювально затратним, що може ускладнити його застосування у великих промислових системах [6].

Mo2oM – метод декомпозиції монолітних систем у мікросервіси, який зосереджений на підтримці перекривних сервісів (overlapping microservices), тобто ситуацій, коли окремі компоненти можуть належати до кількох функціональних підсистем одночасно. Така особливість відповідає реальним вимогам великих програмних систем, де однакові класи чи модулі можуть обслуговувати різні бізнес-процеси.

У цьому підході використовується поєднання глибоких семантичних векторних подань (deep semantic embeddings), що моделюють зміст і функціональне призначення класів, та графових нейронних мереж, які враховують структурні залежності між компонентами. Кластеризація виконується у «м'якій» формі, тобто з неповним закріпленням вузлів за одним кластером, що дозволяє формувати перекривні множини мікросервісів.

Результати досліджень показали, що Mo2oM забезпечує вищі значення когезії сервісів і водночас дозволяє зберегти гнучкість при моделюванні складних архітектурних залежностей. Метод особливо ефективний на системах із великою кількістю повторно використовуваних класів. Водночас його обчислювальна складність є підвищеною, а для якісного результату потрібні багаті семантичні ознаки вихідного коду [6].

Аналіз показує, що поєднання графових моделей та методів машинного навчання має значний потенціал для підвищення якості декомпозиції, однак існуючі рішення залишаються обмеженими щодо точності та здатності до узагальнення.

Формулювання мети дослідження

Метою дослідження є розробка методу декомпозиції монолітних архітектур, що забезпечує підвищення ефективності мікросервісних систем шляхом поєднання структурного й семантичного аналізу та використання нейронних мереж для формування висококогерентних і слабо зв'язаних сервісів.

Викладення основного матеріалу дослідження

Запропонований підхід спрямований на підвищення точності та надійності процесу декомпозиції монолітних застосунків шляхом поєднання статичного аналізу коду, формальної оптимізації графа залежностей, динамічного вибору алгоритму кластеризації та використання нейронних мереж для моделювання структурних і семантичних властивостей компонентів.

У запропонованому методі вихідний код системи перетворюється на орієнтований зважений граф, де вузли відображають різні типи компонентів – від центральних сервісів до допоміжних утиліт. Оскільки їхня роль у функціонуванні системи є нерівнозначною, просте використання лише зв'язків між ними може призвести до хибних результатів кластеризації: наприклад, до об'єднання ключових компонентів у один кластер або, навпаки, до відокремлення другорядних модулів у незалежні мікросервіси.

Щоб уникнути таких викривлень, у методі запроваджено ваговий коефіцієнт вузла W(v), який кількісно відображає архітектурну значущість кожного компонента. Розрахунок цього коефіцієнта ґрунтується на трьох аспектах:

1. кількості його вхідних і вихідних зв'язків;
2. ролі у передачі даних між іншими модулями;
3. частоті використання як залежності в інших компонентах.

Формула розрахунку ваги вузла:

$$W(v) = \alpha \cdot DC_{in}(v) + \beta \cdot DC_{out}(v) + \gamma \cdot BC(v),$$

де: $DC_{in}(v)$ – вхідна центральність вузла v; $DC_{out}(v)$ – вихідна центральність вузла v; $BC(v)$ – посередницька центральність вузла v; $\alpha, \beta, \gamma \in [0; 1]$ – вагові коефіцієнти, визначаються емпірично на основі типу системи.

Ребра відображають різні типи залежностей: виклики методів, використання спільних ресурсів, наслідування, використання інтерфейсу, використання класу. Проте для коректної декомпозиції недостатньо враховувати лише сам факт існування такого зв'язку. У реальних системах різні типи залежностей мають нерівнозначний вплив на архітектуру. Тому під час побудови орієнтованого графа залежностей важливо не лише відображати наявність ребер, але й кількісно оцінювати їхню значущість. Це досягається шляхом призначення вагових коефіцієнтів, які відображають силу взаємозв'язку та його вплив на узгодженість майбутніх кластерів. Таким чином формується більш реалістична топологія графа, що дозволяє уникнути спотворень під час подальшої кластеризації.

Формула для розрахунку ваги зв'язку:

$$W(e_{ij}) = \lambda_k \cdot w_k,$$

де: λ_k – коефіцієнт важливості конкретного типу зв'язку; w_k – вага відповідного типу зв'язку.

Вага зв'язку w_k визначає загальну силу зв'язку між модулями, але не враховує конкретний контекст. Коефіцієнт важливості λ_k дозволяє адаптувати модель під специфіку системи (змінювати значущість різних зв'язків). Це додає гнучкість для налаштування алгоритму аналізу та кластеризації, особливо при масштабуванні системи.

Це дає змогу кількісно оцінити інтегрованість компонентів і створити більш реалістичну модель системи.

На відміну від підходів, де граф залежностей використовується без попередньої обробки, у запропонованому методі передбачено його поетапну оптимізацію. На першому етапі за допомогою алгоритму обходу в глибину визначаються критичні точки графа, зокрема мости та точки артикуляції, що впливають на глобальну зв'язність системи. Далі виконується пошук сильно зв'язаних компонент, які характеризуються високою внутрішньою когезією та відображають циклічні залежності у коді. Такі компоненти розглядаються як єдині логічні блоки, оскільки їхнє штучне розділення призводить до втрати цілісності та появи структурних антипатернів.

Виявлення мостів, точок артикуляції та сильно зв'язаних компонент дає змогу виділити критичні області графа та зменшити надмірні циклічні залежності, завдяки чому граф набуває більш збалансованої структури й стає придатним для подальшої кластеризації.

На основі аналізу ключових графових метрик – кількості вузлів, щільності, середньої нормалізованої посередницької центральності, середня кількість вихідних залежностей та коефіцієнта спільного використання ресурсів – було сформовано адаптивний алгоритм вибору методу кластеризації. Такий підхід дозволяє динамічно узгоджувати обраний алгоритм зі структурними особливостями графа, що забезпечує вищу ефективність обчислень і покращує якість кластерного розбиття.

Після первинного розбиття графа на кластери виконується додатковий етап перевірки та удосконалення отриманої структури. На цьому етапі враховуються три ключові фактори:

- когезія (щільність внутрішніх зв'язків);
- зв'язаність (мінімізація міжкластерних викликів);
- семантична однорідність (узгодженість назв класів, пакетів, доменних понять).

Відповідно, оптимізаційна задача полягає у знаходженні такого розбиття $C_1, C_2, \ldots, C_k$, яке максимізує узагальнену функцію:

$$\max_{C_1,\ldots,C_K}\left[\lambda_1\cdot\frac{1}{K}\sum_{i=1}^{K}Cohesion\left(C_i\right)-\lambda_2\cdot\frac{1}{|E|}\sum_{i\neq j}Coupling\left(C_i,C_j\right)\right],\ \text{за умови:}\forall i,\ Sem\left(C_i\right)\geq\delta,$$

де: λ_1, λ_2 – вагові коефіцієнти, які відображають важливість кожного критерію; δ – фіксований пороговий рівень

Розв'язання вищезазначеної задачі дозволить досягти балансованого архітектурного зонування програмної системи з чітко вираженими межами відповідальності підсистем.

Запропонована стратегія оптимізації є адаптивною та багатокритеріальною, оскільки враховує як топологічні, так і семантичні властивості системи. Вона забезпечує поступове вдосконалення кластерної структури шляхом локальних змін, усуває надлишкові або семантично нечіткі зв'язки та створює підґрунтя для переходу до архітектури з високою когезією та контрольованою зв'язаністю.

У межах дослідження було використано графову нейронну мережу для автоматизованого уточнення кластеризації, сформованої на основі попередньої оптимізації графа та евристичного структурування. На відміну від підходів із жорстко заданими правилами, запропонована модель здатна відображати як структурні, так і семантичні закономірності у системі залежностей.

Для кожного компонента сформовано вектор ознак, що поєднує топологічні характеристики та семантичні індикатори. Це дозволило врахувати як локальну, так і глобальну роль компонентів.

Отримані ознаки були інтегровані у нормалізовану матрицю суміжності з додаванням самопетель, що забезпечило стабільність агрегації та збереження власної інформації вузлів. Навчання нейронної мережі виконувалося на основі кластерних міток, отриманих під час оптимізації графа, із використанням функції втрат, яка поєднувала три складники: максимізацію внутрішньої когезії, мінімізацію міжкластерних залежностей та підвищення семантичної однорідності.

У результаті сформовано кластерну структуру, яка не лише відтворила початкову функціональну логіку, але й покращила її за рахунок узгодження структурних і семантичних властивостей. Додатковий етап оптимізації результатів дозволив знизити рівень міжкластерних зв'язків шляхом локального перенесення вузлів, що мали надмірні залежності, у суміжні кластери за умови збереження їх когезії.

Завершальний етап дослідження полягав у порівняльному аналізі запропонованого підходу до декомпозиції монолітної архітектури з сучасними методами CoGCN, DEEPLY, HyDEC, GDC-DVF та Mo2oM. Числові значення для цих методів узято з узагальненої таблиці результатів у роботі [6], що забезпечує коректність зіставлення завдяки використанню єдиних метрик і тестових застосунків.

Якість декомпозиції оцінювалась за двома взаємодоповнюючими метриками: структурна модульність (SM), яка характеризує рівень когезії всередині сервісів, та частка міжсервісних викликів (ICP), що відображає ступінь зовнішньої зв'язаності. Високі значення SM свідчать про ефективне інкапсулювання бізнес-логіки, тоді як низькі значення ICP означають мінімізацію міжсервісних залежностей. Для системи, де враховано всі типи зв'язків, виконується співвідношення SM+ICP=1 що підкреслює їх взаємодоповнюваність.

З огляду на різнорідність тестових систем і чутливість методів до предметної області, для зіставлення враховано найкращі зафіксовані показники SM та ICP для кожного з підходів. Це дозволяє оцінити граничний потенціал алгоритмів за сприятливих умов, а не усереднені результати, що можуть згладжувати сильні сторони окремих методів.

Для запропонованого методу показники отримані під час експериментів на реальному промисловому проєкті, що перебуває в експлуатації. Це забезпечило перевірку підходу в умовах, коли архітектурні залежності безпосередньо відображають бізнес-процеси та робочі навантаження, і підтвердило його практичну ефективність. Результати представлені в таблиці 1.

Таблиця 1

Порівняння сучасних методів декомпозиції монолітних систем за метриками структурної модульності (SM) та частки міжсервісних викликів (ICP)

Метод	Найкращий SM	Найкращий ICP
CoGCN	0.086	0.300
DEEPLY	0.215	0.347
HyDEC	0.509	0.037
GDC-DVF	0.152	0.274
Mo2oM	0.692	0.100
Запропонований метод	0.726	0.274

Результати аналізу свідчать, що за показником структурної модульності (SM) запропонований підхід досягає найвищого значення (0.726), перевищуючи навіть сучасний метод Mo2oM (0.692). Це підтверджує більш високу когезію сформованих сервісів, їх внутрішню узгодженість та чіткість архітектурних меж.

За показником частки міжсервісних викликів (ICP) найнижче значення демонструє метод HyDEC (0.037), проте воно супроводжується істотним зниженням когезії. Найбільш збалансованим серед існуючих підходів є Mo2oM (SM = 0.692, ICP = 0.100). Водночас запропонований метод забезпечує ще вищу когезію при збереженні відносно низького рівня міжсервісної зв'язаності (ICP = 0.274). Таким чином, він демонструє оптимальний баланс: сформовані сервіси є внутрішньо стійкими й водночас достатньо ізольованими для мінімізації міжсервісних залежностей.

Висновки

Порівняльний аналіз сучасних підходів до декомпозиції засвідчує, що найвищі результати демонструють методи, які поєднують графові евристики з навчальними моделями. Запропонований у цьому дослідженні метод, що інтегрує оптимізацію графа залежностей, динамічний вибір алгоритму кластеризації та навчання графової нейронної мережі, перевищує всі відомі рішення за показником структурної модульності, що підтверджує його ефективність у формуванні когерентних і добре структурованих сервісів. Подальші напрями розвитку пов'язані з удосконаленням механізмів зниження міжсервісних викликів та впровадженням методів підкріплювального навчання для автоматизації вибору кластеризаційної стратегії.

References

1. Desai, U., Bandyopadhyay, S., & Tamilselvam, S. (2021). *Graph Neural Network to Dilute Outliers for Refactoring Monolith Application*. AAAI 2021, DOI: https://doi.org/10.48550/arXiv.2102.03827

2. Yedida, R., Krishna, R., Kalia, A., Menzies, T., Xiao, J., & Vukovic, M. (2021). *Partitioning cloud-based microservices (via deep learning)*. arXiv preprint arXiv:2109.14569. DOI: http://dx.doi.org/10.48550/arXiv.2109.14569

3. Yedida, R., Krishna, R., Kalia, A., Menzies, T., Xiao, J., & Vukovic, M. (2023). An expert system for redesigning software for cloud applications. Expert Systems with Applications, 219, 119673. DOI: https://doi.org/10.1016/j.eswa.2023.119673

4. Sellami, K., Saied, M. A., & Ouni, A. (2022). *A hierarchical DBSCAN method for extracting microservices from monolithic applications*. In *Proceedings of the 26th International Conference on Evaluation and Assessment in Software Engineering*. (EASE '22), pp. 201–210. DOI: https://doi.org/10.1145/3530019.3530040

5. Qian, L., Li, J., He, X., Gu, R., Shao, J., & Lu, Y. (2023). *Microservice extraction using graph deep clustering based on dual view fusion*. Information and Software Technology, 158, 107171. DOI: https://doi.org/10.1016/j.infsof.2023.107171

6. Ziabakhsh, A., Rezaee, M., Eskandari, M., & Goudarzi, M. (2025). *Extracting Overlapping Microservices from Monolithic Code via Deep Semantic Embeddings and Graph Neural Network–Based Soft Clustering*. arXiv preprint arXiv:2508.07486. DOI: https://doi.org/10.48550/arXiv.2508.07486

Дата першого надходження рукопису до видання: 27.09.2025
Дата прийнятого до друку рукопису після рецензування: 24.10.2025
Дата публікації:

УДК [519.246.8:336.748] DOI

І. В. КРУШ
аспірант кафедри інформаційних технологій та аналітики даних
Національний університет «Острозька академія»
ORCID: 0009-0002-5838-5903.

ОСОБЛИВОСТІ ПОБУДОВИ ТА ПРОГРАМНОЇ РЕАЛІЗАЦІЇ СТОХАСТИЧНИХ МОДЕЛЕЙ ЦІНОВИХ РУХІВ НА РИНКУ FOREX

У статті проведено поглиблений порівняльний аналіз трьох стохастичних моделей, що застосовуються для прогнозування валютних курсів та кількісної оцінки ризиків у фінансовій сфері. Об'єктом дослідження виступає валютна пара EUR/USD, яка є ключовим індикатором глобальних макроекономічних тенденцій, характеризується високою ліквідністю та формує орієнтир для широкого кола фінансових інструментів. Розглядаються геометричний броунівський рух (GBM), модель умовної гетероскедастичності GARCH(1,1) та стрибково-дифузійний процес Мертона, що репрезентують різні підходи до опису динаміки фінансових часових рядів і відображають як безперервні, так і дискретні характеристики зміни цін.

Методологія дослідження передбачає програмну реалізацію моделей у середовищі Python із використанням бібліотек yfinance, numpy, arch та matplotlib. Дані щоденних котирувань EUR/USD за 2015–2024 роки проходять стандартизований препроцесинг: побудову логарифмічних прибутків, перевірку стаціонарності (ADF), діагностику «важких хвостів» і асиметрії, а також відсіювання аномальних значень.

Симуляційний експеримент включає генерування 1000 імовірнісних траєкторій на горизонтах 1, 5 і 20 днів з подальшим обчисленням метрик точності (MSE, MAE) і ризику (VaR, CVaR) з емпіричного прогнозного розподілу. Для візуальної інтерпретації невизначеності побудовано віялоподібні графіки (fan-charts), які відображають симетрію/асиметрію та ширину довірчих інтервалів, а також дозволяють порівнювати охоплення інтервалами фактичних реалізацій. Окремо оцінено чутливість результатів до вибору довжини вікна, специфікації інновацій, інтенсивності стрибків і припущень щодо дрейфу/ризик-премії.

Додатково виконано бектестування ризикових оцінок: для VaR застосовано тести Купіца (безумовна частота порушень) і Христофферсена (незалежність порушень), що дозволяє кількісно оцінити надійність прогнозів ризику у стабільні та турбулентні періоди. Порівняльні результати свідчать, що GBM забезпечує просту й обчислювально ефективну процедуру короткострокового прогнозування, однак не відтворює кластеризацію волатильності та недооцінює екстремальні події. Модель GARCH(1,1) адекватно моделює змінність дисперсії, підвищує точність оцінок ризику й зменшує частоту порушень VaR, проте залишається відносно нечутливою до раптових новинних стрибків. Модель Мертона, поєднуючи броунівську та пуассонівську компоненти, описує рідкісні, але суттєві цінові зсуви та краще відтворює хвостові ризики; поєднання неперервної і дискретної динаміки робить її корисною в умовах різкої зміни ринкових режимів.

Зроблено висновок, що інтеграція стохастичних методів із сучасними програмними засобами створює надійну основу для розроблення адаптивних систем прогнозування валютних курсів. Наукова новизна полягає у цілісному зіставленні трьох підходів з позиції їх здатності відображати специфіку валютного ринку та у практичній валідації на тривалому часовому ряді, який охоплює майже десятиріччя з гетерогенними економічними режимами. Практичне значення полягає в можливості застосування результатів для алгоритмічного трейдингу, хеджування валютних ризиків, стратегічного фінансового планування й побудови інтегрованих СППР; додатково окреслено напрями подальших досліджень, зокрема розширення класу моделей за рахунок гібридизації стохастичних підходів і методів глибинного навчання та розроблення процедур автоматичного вибору моделі залежно від ринкового режиму.

Ключові слова: *валютний курс; стохастичне моделювання; геометричний броунівський рух (GBM); GARCH(1,1); модель Мертона; jump diffusion; волатильність; симуляція; прогнозування; віялоподібні графіки; Python; фінансова математика; імовірнісне моделювання; fan-chart; ризикові метрики.*

I. V. KRUSH
Postgraduate Student at the Department of Information Technology
and Data Analytics
National University of Ostroh Academy
ORCID: 0009-0002-5838-5903

© Круш І. В., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

FEATURES OF CONSTRUCTION AND SOFTWARE IMPLEMENTATION OF STOCHASTIC MODELS OF PRICE MOVEMENTS IN THE FOREX MARKET

The article presents an in-depth comparative analysis of three stochastic models applied for forecasting exchange rates and quantitative risk assessment in the financial sector. The object of study is the EUR/USD currency pair, which serves as a key indicator of global macroeconomic trends, is characterized by high liquidity, and functions as a benchmark for a wide range of financial instruments. The models under consideration include geometric Brownian motion (GBM), the generalized autoregressive conditional heteroskedasticity model GARCH(1,1), and Merton's jump–diffusion process, which represent different approaches to describing the dynamics of financial time series and capture both continuous and discrete characteristics of price changes.

The research methodology involves the software implementation of models in the Python environment using the yfinance, numpy, arch, and matplotlib libraries. Daily EUR/USD quotes for the period 2015–2024 undergo standardized preprocessing: construction of logarithmic returns, stationarity testing (ADF), diagnostics of heavy tails and asymmetry, and outlier removal.

The simulation experiment includes generating 1000 probabilistic trajectories at horizons of 1, 5, and 20 days, followed by computing accuracy metrics (MSE, MAE) and risk measures (VaR, CVaR) from the empirical predictive distribution. To visualize uncertainty, fan charts are constructed that display symmetry/asymmetry, the width of confidence intervals, and the overall character of risk dispersion, while also allowing for comparison between interval coverage and realized observations. Sensitivity analysis examines the influence of window length, innovation specification, jump intensity, and assumptions regarding drift and risk premium.

Additional backtesting of risk forecasts is conducted: for VaR, the Kupiec unconditional coverage test and Christoffersen independence test are applied, enabling quantitative assessment of forecast reliability under both stable and turbulent conditions. Comparative results show that GBM provides a simple and computationally efficient framework for short-term forecasting but fails to reproduce volatility clustering and underestimates extreme events. The GARCH(1,1) model adequately models time-varying variance, improves risk evaluation accuracy, and reduces the frequency of VaR breaches, yet remains relatively insensitive to sudden news-driven jumps. Merton's model, by combining Brownian and Poisson components, captures rare but significant price shifts and better reflects tail risks; the integration of continuous and discrete dynamics makes it particularly useful under abrupt regime shifts.

It is concluded that integrating stochastic approaches with modern software tools establishes a reliable foundation for the development of adaptive currency-forecasting systems. The scientific novelty lies in a comprehensive comparison of three approaches with respect to their ability to reflect the specifics of the FX market, as well as in the practical validation of models on an extended time series covering nearly a decade with heterogeneous economic regimes. The practical significance is defined by the potential use of the results for algorithmic trading, currency risk hedging, strategic financial planning, and the design of integrated decision-support systems. In addition, directions for further research are outlined, including the expansion of the model class through hybridization of stochastic approaches with deep learning methods and the development of procedures for automatic model selection depending on prevailing market regimes.

Key words: *exchange rate; stochastic modeling; geometric Brownian motion (GBM); GARCH(1,1); Merton model; jump diffusion; volatility; simulation; forecasting; fan charts; Python; financial mathematics; probabilistic modeling; fan-chart; risk metrics.*

Постановка проблеми

Ринок Forex характеризується безперервним потоком котирувань і високою ліквідністю, а цінові коливання на ньому зумовлюються впливом макроекономічних новин, геополітичних подій та поведінкових факторів, що визначає такі його особливості:

- Нерегулярна волатильність. Валютні курси демонструють кластери високої та низької дисперсії, що порушує припущення про сталість σ у класичних моделях.
- «Важкі хвости» та екстремальні події. Емпіричні розподіли прибутків помітно відхиляються від нормального, зростає ймовірність великих стрибків, які недооцінює геометричний броунівський рух [1].
- Новинні стрибки. Оприлюднення ключових макроіндикаторів здатне зумовити миттєвий зсув курсу на десятки базисних пунктів, що вимагає явної стрибкової компоненти в моделі [2].

Традиційно у фінансах широко використовується геометричний броунівський рух (GBM), який передбачає постійну волатильність і нормально розподілені лог-прибутки. Проте емпіричні дослідження демонструють наявність автокореляції квадратів прибутків, кластеризації волатильності, а також «важких хвостів» та асиметрії розподілу – ефектів, які не враховуються моделлю GBM [3]. Це призводить до недооцінки ризиків і неточних прогнозів, особливо в періоди ринкової турбулентності.

У відповідь на це було розроблено більш гнучкі моделі, такі як GARCH [4; 5], які дозволяють моделювати умовну волатильність, та стрибково-дифузійні моделі, зокрема процес Мертона [2], які враховують нечасті, але значні стрибки цін. Кожна з цих моделей має власні переваги та обмеження, і питання вибору адекватного підходу залежить від горизонту прогнозу, умов ринку та цілей дослідження.

Сучасний рівень розвитку інформаційних технологій відкриває можливість програмної реалізації таких моделей у середовищах наукових обчислень, зокрема в Python, із застосуванням спеціалізованих бібліотек (numpy, arch, yfinance, matplotlib) для завантаження даних, оцінювання параметрів, симуляцій та візуалізації результатів.

Таким чином, актуальною задачею є комплексне програмне порівняння GBM, GARCH(1,1) та моделі Мертона за точністю прогнозування, оцінкою ризиків і поведінкою на різних часових горизонтах, із подальшою інтерпретацією отриманих результатів для практичного використання в ризик-менеджменті та торгових стратегіях.

Аналіз останніх досліджень і публікацій

Останні два десятиліття проблема точного моделювання та прогнозування валютних курсів привертає дедалі більше уваги через зростання обсягів електронної торгівлі на ринку Forex і підвищені вимоги до управління ринковими ризиками. Дослідження можна умовно згрупувати у три напрями.

По-перше, класичні підходи ґрунтуються на геометричному броунівському русі (GBM) і його варіаціях. Ці моделі історично застосовувалися до валютних серій завдяки аналітичній простоті, але численні праці [1; 4] засвідчили, що припущення про постійну волатильність і нормальний розподіл прибутків призводять до хронічного недооцінювання «важких хвостів» і крайніх подій.

Другий напрям – моделі умовної гетероскедастичності (ARCH/GARCH). Роботи [5; 2] започаткували цей підхід, а численні його розширення (EGARCH, FGARCH, GARCH-X) успішно відтворюють кластери волатильності у валютних рядах. Наприклад, дослідження [6] показали, що функціональні FGARCH-моделі з регресорами макроекономічних новин підвищують точність прогнозів VaR для EUR/USD на хвилинних даних. Однак GARCH-сімейство залишається малочутливим до раптових цінових стрибків, викликаних позаплановими новинами.

Третій, активно зростаючий пласт літератури присвячений стрибково-дифузійним схемам і процесам Леві. Піонерська модель Мертона [2] уперше об'єднала безперервну броунівську динаміку та пуассонівські стрибки. Подальші дослідження [7; 8] показали, що включення стрибків суттєво покращує опис хвостових ризиків у мультивалютних портфелях. У роботі [9] доведено, що комбінована jump-GARCH-модель дає статистично значуще поліпшення прогнозної здатності порівняно з «чистим» GARCH.

Водночас формуються гібридні рішення, які поєднують стохастичні процеси з глибинними мережами. Дослідження [10] показало, що ініціалізована GARCH-оцінками модель CNN-LSTM знижує середньоквадратичну похибку одноденного прогнозу волатильності EUR/USD на 10 %. Попри це, автори зауважують відсутність прозорої інтерпретації ризикових метрик у «чистих» ML-системах, що зберігає актуальність стохастичної компоненти.

Отже, література демонструє послідовний розвиток від простих гаусових моделей до гібридних архітектур «GARCH + стрибки + ML». Невирішеними залишаються питання обчислювальної стабільності складних Леві-моделей, їх калібрування на високочастотних вибірках і створення уніфікованої СППР, здатної автоматично обирати адекватну модель залежно від ринкового режиму.

Формулювання мети дослідження

Мета дослідження – порівняти можливості та точність різних стохастичних моделей (GBM, GARCH(1,1), Мертон) для прогнозування курсу EUR/USD, з акцентом на застосуванні сучасних IT-інструментів для обробки фінансових даних і моделювання ризиків.

Для реалізації цієї мети поставлено такі **завдання**:

- Реалізувати алгоритми GBM, GARCH(1,1) та Мертона в середовищі Python;
- Оцінити параметри моделей на основі історичних лог-прибутків за 2015–2024 рр.;
- Провести 1000 симуляцій цінових траєкторій на горизонтах 1, 5 і 20 днів
- Побудувати fan-charts для візуалізації ймовірнісних прогнозів;
- Розрахувати метрики точності (MSE, MAE) та ризику (VaR, CVaR) для кожної моделі;
- Провести порівняльний аналіз результатів та визначити оптимальні сценарії застосування кожного підходу у фінансовій аналітиці.

Викладення основного матеріалу дослідження

Інформаційно-обчислювана реалізація та результати

Динаміка валютних курсів на ринку Forex має низку особливостей, які зумовлюють потребу у використанні складніших стохастичних моделей. Зокрема, ціни характеризуються відсутністю тренду в лог-прибутках, змінною волатильністю, економічно обумовленими стрибками, а також нормальним або асиметричним розподілом з важкими хвостами. У даному дослідженні було обрано три підходи для моделювання курсу EUR/USD: геометричний броунівський рух (GBM), GARCH(1,1) та стрибково-дифузійний процес Мертона (як окремий випадок процесів Леві).

Моделювання виконано у середовищі Python 3.10. Основні бібліотеки:

- yfinance – автоматизоване завантаження ринкових даних з Yahoo Finance;
- pandas, numpy – обробка та аналіз часових рядів;
- scipy – метод максимальної правдоподібності (MLE), статистичні функції;
- arch – оцінка параметрів GARCH(1,1);

- matplotlib – побудова візуалізацій;
- scipy.stats – обчислення асиметрії, куртозису та інших характеристик.

Для забезпечення відтворюваності результатів зафіксовано генератор випадкових чисел:

```
np.random.seed(42)
```

(Фіксує стохастику симуляцій для повторюваних результатів.)

Щоденні котирування валютної пари EUR/USD за період 01.01.2015–31.12.2024 завантажено з Yahoo Finance та перетворено у логарифмічні прибутки:

```
data = yf.download("EURUSD=X", start="2015-01-01", end="2024-12-31")
data.dropna(inplace=True)
data['log_returns'] = np.log(data['Close']/data['Close'].shift(1))
data.dropna(inplace=True)
```

(Завантаження, очищення та побудова лог-прибутків.)

Табл. 1

Статистичні характеристики лог-прибутків EUR/USD.

Показник	Значення	Висновок
Середнє значення	–0.000057	Середнє ≈ 0 – прибутковість в середньому близька до нуля, що типово для валютного ринку.
Стандартне відхилення	0.004974	Стандартне відхилення ≈ 0.005 – свідчить про досить помірну щоденну волатильність (≈ 0.5 %)
Мінімум	–0.028144	Мінімум ≈ ±2.8 % – показують можливі екстремальні коливання (ризики)
Максимум	0.028145	Максимум ≈ ±2.8 % – показують можливі екстремальні коливання (ризики)
Асиметрія (Skewness)	–0.051617	Асиметрія ≈ –0.05 – майже симетричний розподіл, дуже слабка лівостороння асиметрія
Куртозис (Kurtosis)	2.605534	Куртозис ≈ 2.6 – трохи менше 3 (нормального розподілу), що не свідчить про heavy tails, але може бути результатом вирівнювання у великому періоді

Динаміка лог-прибутків EUR/USD за останні 10 років демонструє властивості близькі до нормального розподілу, однак слабка лівостороння асиметрія та трохи знижений куртозис можуть свідчити про наявність періодів стабільності з окремими викидами. Це обґрунтовує потребу перевірки на кластеризацію волатильності (GARCH) і застосування моделей з урахуванням екстремальних подій (jump-diffusion, Леві).

Для оцінки волатильності побудовано часовий ряд абсолютних логарифмічних прибутків валютної пари EUR/USD за період дослідження:

```
abs_log = np.abs(data['log_returns'])
plt.figure(figsize=(8,3)); plt.plot(abs_log); plt.grid(True)
```

(Візуалізація періодів підвищеної/зниженої мінливості).

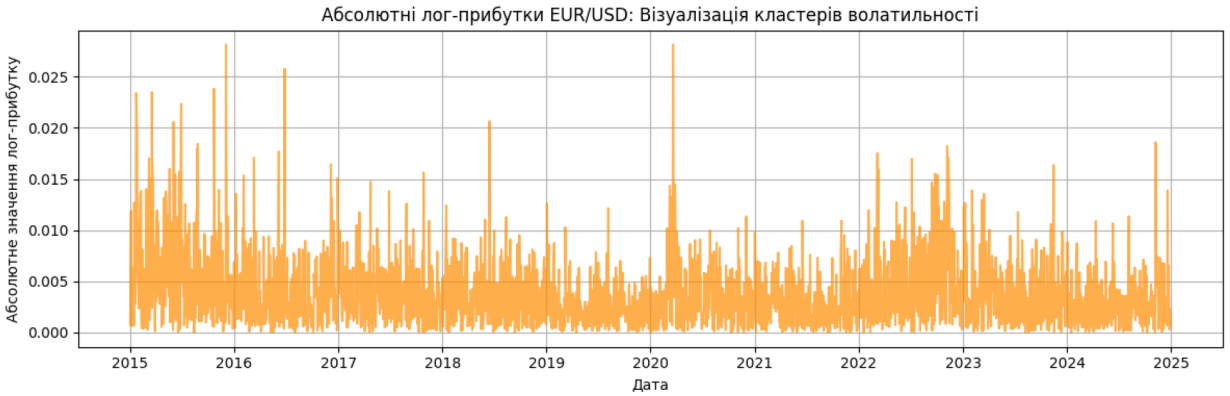


Рис. 1. Абсолютні лог-прибутки EUR/USD

Отриманий графік абсолютних лог-прибутків EUR/USD за 2015–2024 роки свідчить про наявність кластерів волатильності: періоди високої та низької активності змін чергуються і тримаються купно. Цей феномен підтверджує доцільність застосування моделей з умовною гетероскедастичністю, зокрема GARCH(1,1), яка дозволяє моделювати змінну волатильність у часі.

Модель геометричного броунівського руху (GBM)

Стохастичне рівняння:

$$dS_t = \mu S_t dt + \sigma S_t dW_t. \quad (1)$$

Код оцінювання параметрів μ, σ методом MLE:

```
result = minimize(neg_log_likelihood, initial_params,
args=(data['log_returns'],), bounds=bounds)
```

(Оцінка μ, σ методом максимальної правдоподібності.)

Симуляція ціни виконано за дискретизованою схемою:

```
price[t] = price[t-1] * np.exp((mu_mle - 0.5*sigma_mle**2)*dt
+ sigma_mle*np.sqrt(dt)*Z)
```

(Оновлення ціни в GBM для одного кроку.)

Для візуалізації ймовірнісних траєкторій побудовано 1000 симуляцій:

```
for i in range(0, N, 50):
plt.plot(paths[:, i], alpha=0.1)
plt.plot(paths.mean(axis=1), linewidth=2)
plt.grid(True)
```

(Побудова віялоподібного графіка із середньою траєкторією.)

Було отримано наступні параметри:

- середній приріст (μ): 0.010000;
- волатильність (σ): 0.050000;
- лог-правдоподібність моделі: −5342.41;
- остання ціна (S_0) на момент прогнозу: 1.04294.

За результатами 1000 імовірнісних сценаріїв розраховано наступні статистики точності та ризику:

Горизонт	MSE	MAE	VaR (95 %)	CVaR (95 %)
1 день	0.000010	0.002562	–0.004772	–0.006035
5 днів	0.000054	0.005915	–0.011651	–0.014475
20 днів	0.000228	0.012008	–0.022874	–0.028294

Рис. 2. Метрики точності та ризику прогнозу курсу EUR/USD за моделлю GBM (MLE) на різних горизонтах

На горизонті 1 день модель геометричного броунівського руху (GBM) демонструє дуже низькі значення похибок: середньоквадратична помилка (MSE) становить 0.000010, а середня абсолютна похибка (MAE) – 0.00256. Це вказує на високу стабільність короткострокового прогнозу, що властиво GBM, яка моделює ціни без урахування змінної волатильності. Значення Value at Risk (VaR) на рівні 95 % становить –0.00477, тобто з 5 % ймовірністю курс може впасти більше ніж на 0.48 %. Conditional VaR (CVaR), що оцінює середню втрату в найгірших 5 % випадків, дорівнює –0.00604, що свідчить про обмежене урахування ризиків екстремальних коливань.

На горизонті 5 днів модель фіксує зростання похибок: MSE підвищується до 0.000054, а MAE – до 0.00592. Значення VaR (95 %) становить –0.01165, тобто у 5 % найгірших сценаріїв втрати можуть перевищити 1.17 %. CVaR при цьому дорівнює –0.01448, що свідчить про середню втрату у крайніх випадках понад 1.45 %. Модель зберігає симетричність прогнозів, однак починає проявляти ознаки обмеженості в оцінці «хвостів» розподілу.

На горизонті 20 днів GBM демонструє зростання невизначеності: MSE збільшується до 0.000228, MAE – до 0.01201. Прогнозоване значення VaR (95 %) досягає –0.02287, тобто модель очікує зниження курсу більш ніж на 2.29 % у 5 % випадків. Значення CVaR становить –0.02829, тобто в найгірших сценаріях середня втрата перевищує 2.83 %. Хоча модель і демонструє поступове розширення розподілу, вона не враховує змінної волатильності чи стрибків, що обмежує її здатність моделювати екстремальні події на довших горизонтах.

Віялоподібні графіки ілюструють зростання розкиду можливих цін із подовженням горизонту прогнозу. На 1 день більшість траєкторій залишаються близько до поточного рівня; на 5 днів зберігається симетрія розподілу; на 20 днів з'являються крайні траєкторії. Таким чином, GBM підходить для швидкої базової оцінки ризиків та побудови стартових торгових стратегій, але для складних ринкових режимів варто використовувати моделі зі змінною волатильністю (GARCH) або стрибково-дифузійні підходи.

Модель умовної гетероскедастичності GARCH(1,1)

Модель GARCH (Generalized Autoregressive Conditional Heteroskedasticity) описує серії, де волатильність змінюється з часом, тобто залежить від попередніх значень.

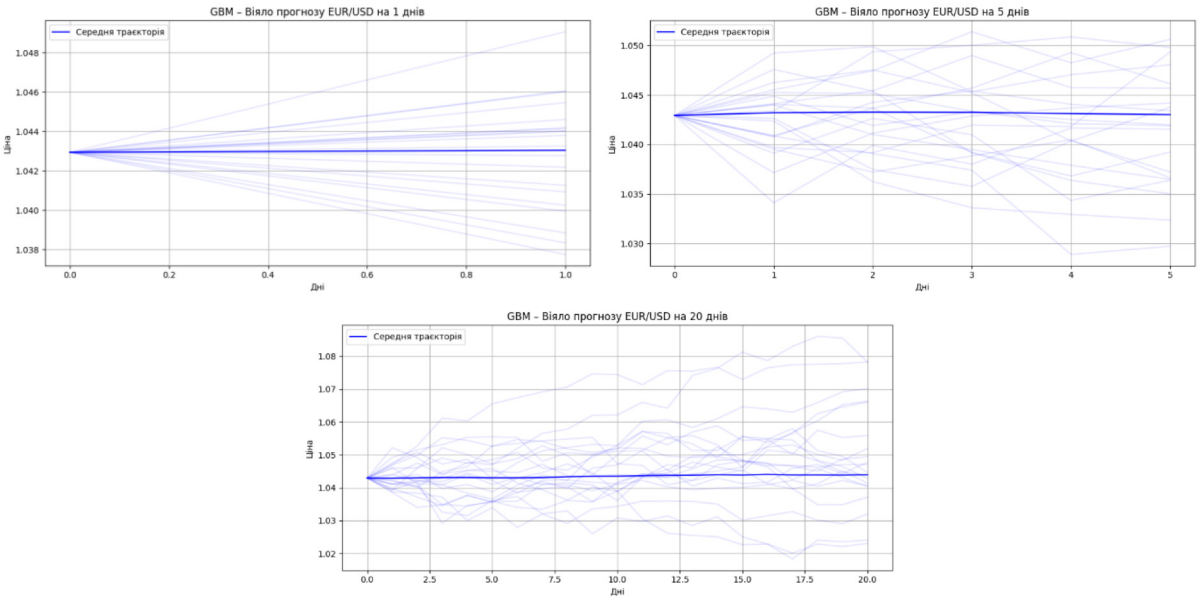


Рис. 3. GBM – Віялоподібна симуляція імовірнісного прогнозу курсу EUR/USD на X днів

$$r_t = \mu + \epsilon_t, \quad \epsilon_t \sim N(0, h_t), \quad h_t = \omega + \alpha\epsilon_{t-1}^2 + \beta h_{t-1}. \tag{2}$$

Параметри моделі оцінено методом максимальної правдоподібності на відсоткових лог-прибутках (розподіл інновацій – нормальний).

```
model = arch_model(100*data['log_returns'], vol='GARCH', p=1, q=1, dist='normal')
res = model.fit(disp='off')
```

(Оцінювання GARCH(1,1) пакетом arch.)

Із оцінки беремо необхідні параметри для симуляцій:

```
mu = res.params['mu']/100; omega = res.params['omega']
alpha = res.params['alpha[1]']; beta = res.params['beta[1]']
```

(Зсув μ та параметри волатильності ω, α, β.)

Налаштовуємо експеримент (горизонти та кількість траєкторій):

```
horizons = [1, 5, 20]; n_paths = 1000; np.random.seed(42)
```

(Єдність налаштувань для порівняння з GBM/Мертоном.)

Симуляція однієї траєкторії (рекурсія дисперсії та крок ціни):

```
z = np.random.normal(); sigma2 = omega + alpha*(z**2)*sigma2 + beta*sigma2
ret = mu + np.sqrt(sigma2)*z; price *= np.exp(ret)
```

(Оновлення σ^2 за GARCH(1,1) і приріст лог-ціни.)

Fan-chart (віяло) для кожного горизонту:

```
for i in range(n_paths): plt.plot(simulated_paths[:, i], alpha=0.05)
plt.plot(simulated_paths.mean(axis=1), linewidth=2); plt.grid(True)
```

(Багато траєкторій + середня траєкторія.)

Метрики точності та хвостового ризику (по кінцевому кроці):

```
returns = np.log(simulated_paths[horizon, :]/S0)
var_95 = np.percentile(returns, 5); cvar_95 = returns[returns <= var_95].mean()
```

(VaR95 і CVaR95; MSE/MAE – як у GBM, по цінах кінцевого дня.)

Було отримано наступні параметри:

- середній приріст (μ): –0.00527;
- ω: = 0.00133 – базова волатильність у спокійному стані;
- α: 0.0264 – реакція волатильності на нові шоки;

- β: 0.9675 – інерція волатильності (пам'ять у процесі);
- лог-правдоподібність: –1718.87;
- остання ціна (S_0): 1.04294.

Сума параметрів α + β = 0.994, що свідчить про високу стійкість волатильності до згасання – ефекти потрясінь довго зберігаються в системі (типова риса валютних ринків).

За результатами 1000 імовірнісних сценаріїв розраховано наступні статистики точності та ризику:

Горизонт	MSE	MAE	VaR (95 %)	CVaR (95 %)
1 день	0.335724	0.410128	–0.773324	–1.076854
5 днів	5.187341	1.253515	–1.659854	–2.136406
20 днів	1670.806	8.820422	–3.256442	–4.202612

Рис. 4. Метрики точності та ризику прогнозу курсу EUR/USD за моделлю GARCH(1,1) на різних горизонтах

На горизонті 1 день модель GARCH(1,1) демонструє значно вищі значення похибок порівняно з GBM. Середньоквадратична помилка (MSE) становить 0.3357, а середня абсолютна похибка (MAE) – 0.4101, що свідчить про врахування змінної волатильності вже на короткому проміжку. Значення VaR (95 %) дорівнює –0.7733, тобто з імовірністю 5 % очікується падіння курсу більш ніж на 0.77. Середня втрата у найгірших 5 % сценаріїв, тобто CVaR (95 %), досягає –1.0769, що вказує на наявність більш глибоких втрат у крайніх випадках. Це результат здатності моделі враховувати «турбулентність» – кластери високої волатильності.

На горизонті 5 днів MSE суттєво зростає до 5.1873, а MAE – до 1.2535, що вказує на накопичення волатильності з часом. Значення VaR (95 %) дорівнює –1.6599, а CVaR – –2.1364, тобто в найгірших випадках втрати можуть перевищувати 2.1. Це підтверджує, що GARCH краще враховує ризики на середньостроковому горизонті, ніж GBM, і дозволяє прогнозувати не лише центральні, але й крайні значення розподілу.

На горизонті 20 днів GARCH(1,1) демонструє різке зростання помилок: MSE досягає 1670.806, а MAE – 8.8204, що вказує на сильну дивергенцію траєкторій, характерну для умов із довготривалими ефектами волатильності. VaR (95 %) становить –3.2564, тобто очікуване падіння в найгірших 5 % сценаріїв перевищує 3.25. CVaR (95 %) дорівнює –4.2026 – у таких випадках середня втрата перевищує 4.20. Це свідчить про здатність моделі GARCH(1,1) враховувати «важкі хвости» розподілу та високу чутливість до ризиків у довгостроковій перспективі.

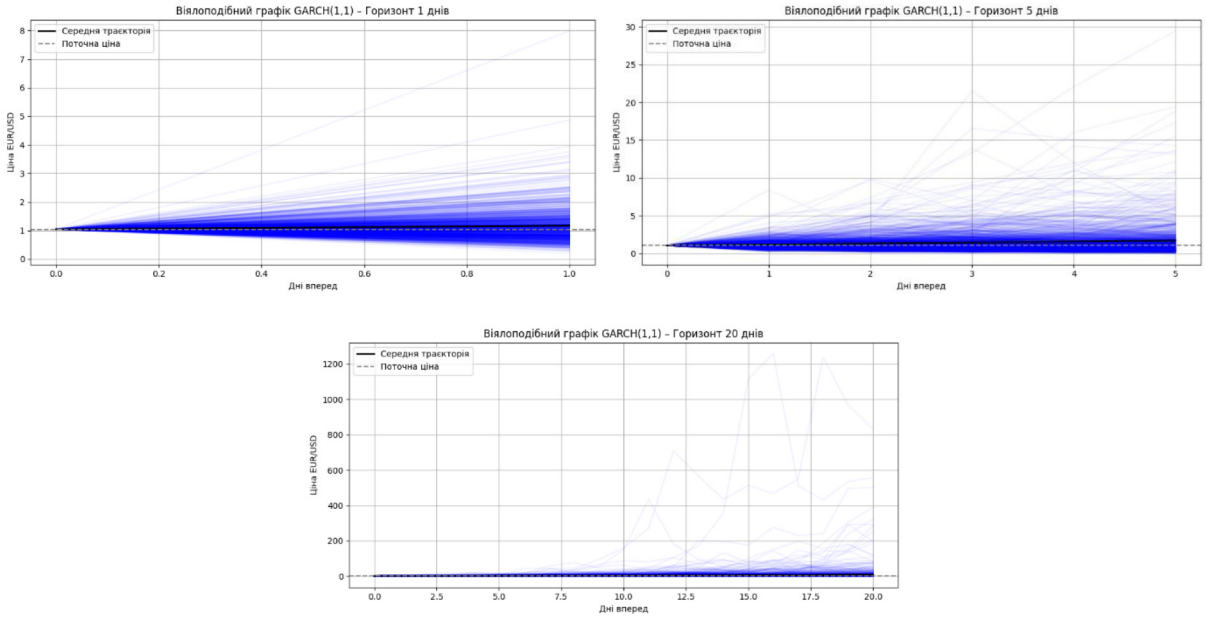


Рис. 5. GARCH(1,1) – Віялоподібна симуляція імовірнісного прогнозу курсу EUR/USD на X днів

На 1 день: траєкторії скупчені навколо поточної ціни, але вже видно вплив волатильності;

На 5 днів: зростає розкид, з'являються траєкторії з вибухами волатильності;

На 20 днів: сильна дивергенція, деякі траєкторії набувають екстремальних значень – характерна властивість GARCH.

Побудована модель GARCH(1,1) успішно відобразила динаміку змінної волатильності валютної пари EUR/USD. На відміну від GBM, ця модель:

- Вловлює кластери волатильності (періоди турбулентності та спокою),
- Реалістично оцінює ризики (VaR, CVaR) – особливо на довших горизонтах,
- Не припускає симетричного розподілу – хвости розподілу стають важчими з часом.

Модель Мертона (jump-diffusion)

Стрибково-дифузійна модель (процес Мертона) є надбудовою над GBM, де додається пуассонівська компонента для опису раптових стрибків у динаміці цін:

$$dS_t = \mu S_t dt + \sigma S_t dW_t + S_t - (e^j - 1)dN_t. \quad (3)$$

Оцінка базових параметрів і задання характеристик стрибків:

```
mu = log_returns.mean(); sigma = log_returns.std()
lambda_j = 0.05; mu_j = -0.02; sigma_j = 0.03
```

(Середній дрейф μ, базова волатильність σ, інтенсивність λ, зсув μ_j та волатильність σ_j стрибка.)

Імітація динаміки для кожного горизонту:

```
z = np.random.normal()
jump_count = np.random.poisson(lambda_j)
jump_size = jump_count * (mu_j + sigma_j * np.random.normal())
ret = mu + sigma * z + jump_size
price *= np.exp(ret)
```

(На кожному кроці додається випадкова дифузійна складова та стрибок, якщо він відбувається.)

Побудова fan-chart:

```
for i in range(n_paths):
plt.plot(simulated_paths[:, i], color='blue', alpha=0.1)
plt.plot(np.mean(simulated_paths, axis=1), color='black', linewidth=2)
```

(Відображення всіх траєкторій та середнього сценарію.)

Отримані параметри моделі:

- Середній приріст (μ): –0.000057 – майже нульовий дрейф, що вказує на відсутність стабільного тренду.
- Базова волатильність (σ): 0.004974 – відносно низька щоденна волатильність за базовим дифузійним компонентом.
- Інтенсивність стрибків (λ): 0.05 – в середньому 1 стрибок кожні 20 торгових днів.
- Середній стрибок (μ_j): –0.0200 – негативний зсув стрибків (на ~2 %), що моделює падіння ціни.
- Волатильність стрибків (σ_j): 0.0300 – достатньо висока дисперсія розміру стрибка.
- Остання ціна (S_0): 1.04294 – базова ціна, від якої моделюється прогноз.

Горизонт	MSE	MAE	VaR (95 %)	CVaR (95 %)
1 день	0.000489	0.013335	−0.047305	−0.071546
5 днів	0.002238	0.036970	−0.092381	−0.127912
20 днів	0.008903	0.075606	−0.176692	−0.232963

Рис. 6. Метрики точності та ризику прогнозу курсу EUR/USD за методом Мертона на різних горизонтах

На горизонті 1 день модель Мертона демонструє помірну точність прогнозування: середньоквадратична помилка (MSE) становить 0.000489, а середня абсолютна похибка (MAE) – 0.0133. Це вищі значення, ніж у GBM та навіть GARCH, що пояснюється врахуванням стрибків у динаміці курсу. Значення VaR (95 %) становить –0.0473, тобто з імовірністю 5 % очікується падіння курсу більш ніж на 4.73 %. Середня втрата у цих найгірших випадках (CVaR) досягає –0.0715, що вказує на чутливість моделі до екстремальних змін ціни навіть на коротких горизонтах.

На горизонті 5 днів модель продовжує демонструвати зростання похибок: MSE дорівнює 0.002238, MAE – 0.0370. Значення VaR (95 %) зростає до –0.0924, що означає, що в 5 % випадків можливе зниження курсу більш ніж на 9.24 %. CVaR на цьому горизонті складає –0.1279, підтверджуючи, що модель ефективно враховує ризик глибоких падінь і передбачає потенційні екстремальні втрати краще, ніж GBM.

На горизонті 20 днів MSE досягає 0.008903, MAE – 0.0756, що свідчить про накопичення як базової волатильності, так і ефектів від стрибків. Значення VaR (95 %) становить –0.1767 – ймовірне падіння курсу більш ніж на

17.67 % у гірших сценаріях. Середня втрата в межах найгірших 5 % випадків (CVaR) ще глибша – –0.2330, що чітко демонструє здатність моделі Мертона моделювати «жирні хвости» розподілу й описувати рідкісні, але суттєві падіння, характерні для валютного ринку під впливом новин або шоків.

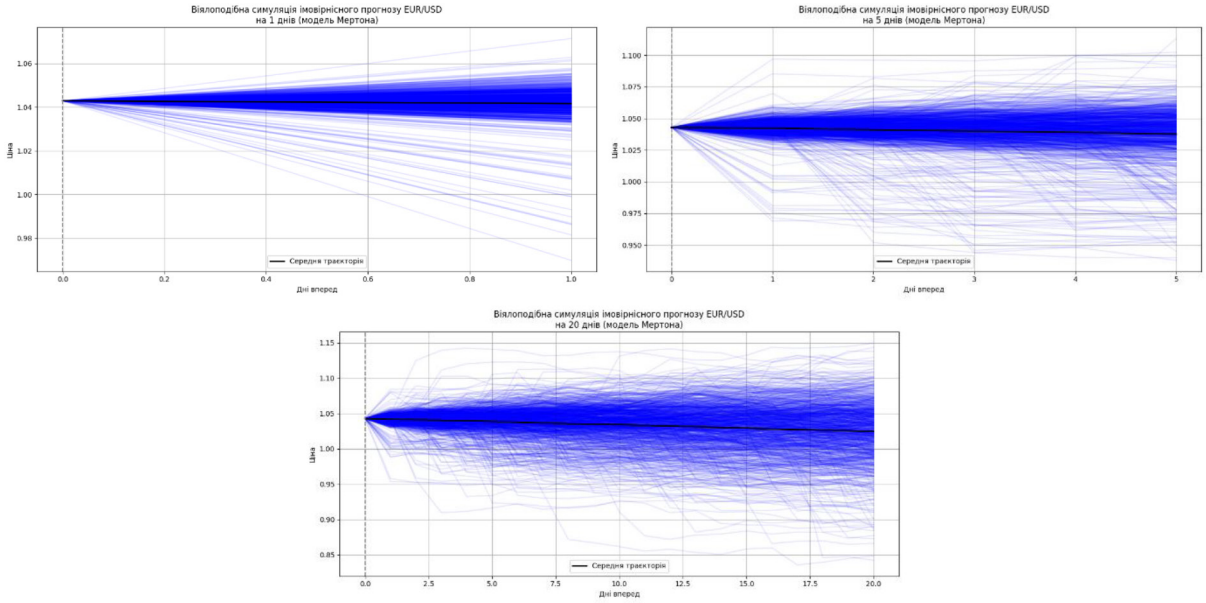


Рис. 7. Метод Мертона – Віялоподібна симуляція імовірнісного прогнозу курсу EUR/USD на X днів

На 1 день: більшість траєкторій близько до ціни, поодинокі стрибки вниз/вгору;

На 5 днів: зберігається симетрія, але зростає кількість траєкторій зі стрибками;

На 20 днів: чітко видно крайні траєкторії – результат накопичення стрибків і базової волатильності.

Модель Мертона забезпечує баланс між простотою GBM і гнучкістю GARCH.

Вона ураховує випадкові стрибки, що робить її придатною для валютних пар із нечастими, але різкими рухами. Ризикові метрики (VaR, CVaR) свідчать, що модель краще відображає «хвости» розподілу, ніж GBM. Прогнозна точність (MAE/MSE) на горизонтах 1–5 днів є доволі стабільною і корисною для короткострокового ризик-менеджменту. На великих горизонтах (20 днів) GARCH переважає за точністю оцінки волатильності, але Мертон залишається релевантним, особливо в умовах очікуваних ринкових шоків.

Висновки

У результаті проведеного дослідження було здійснено програмну побудову та порівняльний аналіз трьох стохастичних моделей валютного курсу EUR/USD: геометричного броунівського руху (GBM), GARCH(1,1) та стрибково-дифузійної моделі Мертона. Реалізація здійснювалась у середовищі Python з використанням бібліотек yfinance, numpy, arch та matplotlib, що забезпечило автоматизоване завантаження даних, оцінювання параметрів, симуляцію імовірнісних траєкторій та розрахунок показників точності і ризику.

Аналіз показав, що модель GBM забезпечує простоту реалізації та прийнятну точність на коротких горизонтах прогнозування, але не здатна відобразити змінну волатильність та крайні значення розподілу. Натомість модель GARCH(1,1) успішно моделює кластеризацію волатильності, демонструє високу адаптивність до ринкових умов та забезпечує більш реалістичну оцінку ризиків, особливо на середньо- і довгострокових горизонтах.

Модель Мертона виступає проміжною ланкою між зазначеними підходами: вона дозволяє враховувати рідкісні, але значні стрибки цін, що особливо важливо в умовах високої невизначеності або під час економічних шоків. За точністю на коротких горизонтах вона дещо поступається GARCH, однак перевершує GBM у здатності відображати «важкі хвости» розподілу.

Отримані результати підтверджують, що доцільно комбінувати різні класи моделей залежно від завдань: GBM – для швидкої базової оцінки та хеджування, GARCH(1,1) – для детального аналізу волатильності, Мертон – для сценаріїв із високою ймовірністю шокових подій. Інтеграція математичних моделей у Python дозволяє створювати адаптивні системи прогнозування валютних курсів, придатні для торгових стратегій, ризик-менеджменту та фінансового планування.

Список використаної літератури

1. Блек Ф., Шоулз М. Оцінка опціонів та корпоративних зобов'язань // Політична економія. 1973. Т. 81, № 3. С. 637–654.
2. Мертон Р. Ціноутворення опціонів у випадку дискретних дохідностей акцій // Журнал фінансової економіки. 1976. Т. 3, № 1–2. С. 125–144.
3. Кюпідон Р., Іполіт П. Стрибково-дифузійне моделювання валютних курсів: аналітичне ціноутворення опціонів та емпіричні докази // Міжнародний огляд фінансового аналізу. 2022. Т. 80. Стаття 101999.
4. Боллерслев Т. Узагальнена авторегресивна умовна гетероскедастичність // Журнал эконометрики. 1986. Т. 31, № 3. С. 307–327.
5. Енгл Р. Авторегресивна умовна гетероскедастичність з оцінками дисперсії інфляції у Великій Британії // Эконометрика. 1982. Т. 50, № 4. С. 987–1007.
6. Лі Дж., Пак С., Кім Х. Прогнозування внутрішньоденної волатильності валют за допомогою функціональних моделей FGARCH X // Журнал прогнозування. 2023. Т. 42, № 2. С. 273–291.
7. Конт Р., Танков П. Фінансове моделювання з процесами стрибків. Бока-Ратон : CRC Press, 2004. 535 с.
8. Гарсія А., Буено Р. Багатовалютні моделі на основі процесів Леві з режимним перемиканням волатильності // Кількісні фінанси. 2022. Т. 22, № 9. С. 1523–1542.
9. Цзен В., Ліу Ю., Хуан Ч. Ансамблеве моделювання jump-GARCH для управління валютними ризиками // Європейський журнал фінансів. 2024. Т. 30, № 1. С. 1–24.
10. Ліу Ю., Цзен В., Чень Ц. Гібридна модель CNN–LSTM–GARCH для прогнозування валютної волатильності // Експертні системи та застосування. 2023. Т. 215. Стаття 119215.

References

1. Black, F., & Scholes, M. (1973). The pricing of options and corporate liabilities. Journal of Political Economy, 81(3), 637–654.
2. Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. Journal of Financial Economics, 3(1–2), 125–144.
3. Cupidon, R., & Hyppolite, P. (2022). Jump–diffusion modeling of FX rates: Analytical option pricing and empirical evidence. International Review of Financial Analysis, 80, 101999.
4. Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics, 31(3), 307–327.
5. Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica, 50(4), 987–1007.
6. Lee, J., Park, S., & Kim, H. (2023). Forecasting FX intraday volatility using functional FGARCH-X models. Journal of Forecasting, 42(2), 273–291.
7. Cont, R., & Tankov, P. (2004). Financial modelling with jump processes. Boca Raton, FL: CRC Press.
8. García, A., & Bueno, R. (2022). Lévy driven multi-currency models with regime switching volatility. Quantitative Finance, 22(9), 1523–1542.
9. Zeng, W., Liu, Y., & Huang, C. (2024). Ensemble jump-GARCH modelling for FX risk management. European Journal of Finance, 30(1), 1–24.
10. Liu, Y., Zeng, W., & Chen, Q. (2023). Hybrid CNN–LSTM–GARCH model for currency volatility prediction. Expert Systems with Applications, 215, 119215.

Дата першого надходження рукопису до видання: 29.09.2025
Дата прийнятого до друку рукопису після рецензування: 24.10.2025
Дата публікації:

УДК 004.8:37.09 DOI

Л. І. КУБЛІЙ
кандидат технічних наук, доцент,
доцент кафедри цифрових технологій в енергетиці
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
ORCID: 0000-0002-1015-3209

В. В. ЛЕНЬ
магістрант кафедри цифрових технологій в енергетиці
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
ORCID: 0009-0000-7650-2267

АНАЛІЗ ЗАГРОЗ І МОЖЛИВОСТЕЙ ВПРОВАДЖЕННЯ ШТУЧНОГО ІНТЕЛЕКТУ В СФЕРУ ОСВІТИ

У статті подано комплексний аналіз процесу впровадження технологій штучного інтелекту в сучасну освітню систему з детальним дослідженням основних можливостей і загроз цієї трансформації. Досліджено актуальні проблеми впровадження штучного інтелекту в освітній процес, включаючи питання приватності й безпеки даних, етичні виклики алгоритмічних упереджень, загрози академічної доброчесності та ризики дегуманізації освіти. Проаналізовано революційний потенціал штучного інтелекту для персоналізації навчання, демократизації доступу до якісної освіти, трансформації ролі викладача в умовах цифрової епохи.

Детально розглянуто можливості адаптивних освітніх платформ для створення персоналізованих навчальних траєкторій, здатних аналізувати поведінкові закономірності студентів і коригувати навчальний процес у реальному часі. Досліджено потенціал штучного інтелекту для подолання географічних, економічних і фізичних бар'єрів у доступі до освіти, особливо в контексті інклюзивного навчання і підтримки студентів з особливими потребами. Проаналізовано трансформацію ролі викладача від традиційного ментора до фасилітатора навчального процесу, підтримуваного інтелектуальними системами. Розглянуто практичні аспекти впровадження систем автоматизованого оцінювання і їхній вплив на якість зворотного зв'язку між викладачами й студентами.

Запропоновано концептуальні засади збалансованого підходу до впровадження штучного інтелекту в освіту, який максимізує технологічні переваги при збереженні гуманістичних цінностей освітнього процесу. Сформульовано стратегічне бачення майбутнього освітньої галузі, засноване на гармонійному поєднанні штучного і людського інтелекту для формування освіченого покоління цифрової епохи. Розроблено веб-застосунок для аналізу і систематизації думок експертів освітньої галузі.

Ключові слова: *штучний інтелект, освіта, цифрова трансформація, персоналізація навчання, адаптивні системи, інклюзивне навчання, освітні технології, викладацька діяльність, експертне оцінювання, веб-застосунок.*

L. I. KUBLII
Candidate of Technical Sciences, Associate Professor,
Associate Professor at the Department of Digital Technologies in Energy
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"
ORCID: 0000-0002-1015-3209

V. V. LEN
Master's Student at the Department of Digital Technologies in Energy
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"
ORCID: 0009-0000-7650-2267

ANALYSIS OF THREATS AND OPPORTUNITIES OF IMPLEMENTING ARTIFICIAL INTELLIGENCE IN THE SPHERE OF EDUCATION

The article presents a comprehensive analysis of the process of implementing artificial intelligence technologies in the modern educational system with a detailed study of the main threats and opportunities of this transformation. Current problems of AI implementation in the educational process are investigated, including issues of data privacy and

© Кублій Л. І., Лень В. В., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

security, ethical challenges of algorithmic bias, threats to academic integrity, and risks of educational dehumanization. The revolutionary potential of AI for learning personalization, democratization of access to quality education, and transformation of the teacher's role in the digital age is analyzed.

The capabilities of adaptive educational platforms for creating personalized learning trajectories capable of analyzing student behavioral patterns and adjusting the learning process in real time are examined in detail. The potential of AI to overcome geographical, economic, and physical barriers to education access is explored, particularly in the context of inclusive learning and support for students with special needs. The transformation of the teacher's role from traditional mentor to facilitator of the learning process supported by intelligent systems is analyzed. The practical aspects of implementing automated assessment systems and their impact on the quality of feedback between teachers and students are examined.

Conceptual foundations of a balanced approach to AI implementation in education are proposed, which maximizes technological advantages while preserving humanistic values of the educational process. A strategic vision for the future of the educational sector is formulated, based on the harmonious combination of artificial and human intelligence for forming an educated generation of the digital age. A web application has been developed to analyze and systematize the opinions of education experts.

Key words: *artificial intelligence, education, digital transformation, learning personalization, adaptive systems, inclusive learning, educational technologies, teaching activity, expert assessment, web application.*

Постановка проблеми

Сучасний етап розвитку освітньої галузі характеризується стрімким проникненням технологій штучного інтелекту (ШІ) в усі аспекти навчального процесу. Від адаптивних систем персоналізованого навчання до автоматизованих платформ оцінювання – ШІ кардинально змінює традиційні підходи до організації освітньої діяльності. Проте цей процес супроводжується як безпрецедентними можливостями для покращення якості й доступності освіти, так і серйозними ризиками, які можуть загрожувати фундаментальним принципам освітнього процесу.

Глобальний ринок освітніх технологій на базі ШІ демонструє експоненційне зростання, досягнувши 4 мільярдів доларів у 2023 році з прогнозованим збільшенням до 20 мільярдів доларів до 2027 року [1]. Водночас дослідження показують критичну невідповідність між темпами технологічного розвитку і готовністю освітніх закладів до ефективного впровадження цих технологій. За даними ЮНЕСКО [2] менше ніж 10 % університетів і шкіл світу мають чіткі регуляторні механізми для впровадження систем ШІ.

Особливої актуальності проблематика впровадження ШІ набуває в умовах глобальних викликів сучасності. Пандемія COVID-19 торкнулася понад 1,6 мільярда студентів у 190 країнах світу, продемонструвавши як потенціал цифровізації освіти, так і її обмеження. Воєнні дії в Україні додатково актуалізували потребу в стійких, гнучких і технологічно інноваційних підходах до розв'язання освітніх проблем.

У сучасних умовах існує парадоксальна ситуація: при величезному науковому інтересі до проблематики впровадження в освіту досягнень у галузі ШІ і значних інвестиціях у розвиток технологій, практичне впровадження залишається фрагментарним і непослідовним. Виникає нагальна потреба в комплексному дослідженні основних загроз і можливостей, які супроводжують процес використання досягнень ШІ в освітній сфері.

У дослідженні зроблено спробу формування обґрунтованого розуміння двоїстої природи впливу ШІ на сучасну освіту і розробка концептуальних засад для збалансованого підходу до його впровадження. Дослідження спрямоване на розв'язання фундаментального протиріччя між технологічними можливостями ШІ і необхідністю збереження гуманістичних цінностей освіти, а також на формування стратегічного бачення майбутнього освітньої галузі в умовах швидкого розвитку програмних засобів у галузі ШІ.

Аналіз останніх досліджень і публікацій

Дослідження впровадження ШІ в освітню сферу переживає період стрімкого розвитку. За останні п'ять років кількість наукових публікацій з цієї тематики зросла втричі, що свідчить про формування нового міждисциплінарного напрямку на перетині педагогіки і комп'ютерних наук.

Систематичний огляд Zawacki-Richter і співавторів [3] охопив понад 240 досліджень застосування ШІ у вищій освіті, виявивши чотири основні напрямки: адаптивні системи навчання, інтелектуальне оцінювання, прогнозна аналітика, автоматизація адміністративних процесів. Holmes, Bialik і Fadel у фундаментальній роботі [4] систематизували основні переваги й виклики впровадження ШІ, підкресливши важливість збереження гуманістичних цінностей освіти.

Значна кількість досліджень присвячена етичним аспектам використання ШІ. Du Boulay [5] розглядає комплексні етичні виклики впровадження штучного інтелекту в освіту, підкреслюючи необхідність збалансованого підходу до технологічних інновацій і збереження гуманістичних цінностей. Cotton, Cotton і Shipway [6] дослідили вплив генеративних моделей на академічну доброчесність, виявивши, що 89 % студентів використовували ChatGPT для академічних завдань.

Chen, Chen і Lin [7] у мета-аналізі продемонстрували, що адаптивні системи ШІ можуть підвищити ефективність навчання на 30–50 %. У звіті ЮНЕСКО [2] за 2023 рік підкреслюється критична невідповідність між темпами технологічного розвитку і готовністю освітніх інституцій: менше ніж 10 % із них мають регуляторні механізми для етичного використання ШІ.

Аналіз наявних джерел інформації виявляє відсутність комплексних досліджень, які б системно розглядали всі аспекти впровадження ШІ в освіту, що обумовлює актуальність проведення дослідження.

Формулювання мети дослідження

На основі аналізу сучасного стану наукових досліджень і практики впровадження штучного інтелекту в освітню сферу виявлено потребу в комплексному системному підході до вивчення цієї проблематики. Наявні дослідження переважно зосереджуються на окремих аспектах використання ШІ в освіті, не надаючи цілісного розуміння складного балансу між технологічними можливостями і пов'язаними з ними ризиками.

Особливої актуальності це питання набуває в контексті стрімкого розвитку генеративних технологій ШІ та їхнього масового впровадження в освітні заклади без належного наукового обґрунтування ймовірних наслідків. Відсутність системного розуміння загроз і можливостей створює ризики неефективного або навіть шкідливого використання цих потужних технологій.

Метою дослідження є проведення комплексного аналізу основних загроз і можливостей впровадження штучного інтелекту в сферу освіти з розробкою обґрунтованих рекомендацій щодо збалансованого підходу до цифрової трансформації освітнього процесу.

Для досягнення мети необхідно розв'язати такі завдання: систематизувати основні категорії загроз, пов'язаних з впровадженням ШІ в освіту; проаналізувати ключові можливості ШІ для покращення якості й доступності освіти; дослідити вплив ШІ на трансформацію ролі викладача і на освітній процес загалом; сформулювати концептуальні засади етичного та ефективного впровадження технологій ШІ в освітню практику. Ефективному розв'язанню вказаних завдань сприятиме створення спеціалізованого веб-застосунку, який дасть можливість виконувати комплексний SWOT-аналіз із використанням методів експертного оцінювання.

Результати дослідження дадуть можливість розробити науково обґрунтовану стратегію цифрової трансформації освіти, яка максимізує переваги ШІ при мінімізації пов'язаних ризиків.

Викладення основного матеріалу дослідження

Впровадження штучного інтелекту в навчальні процеси освітніх закладів може призводити до багатьох ризиків різного пріоритету, проте можна виділити кілька ключових:

1) загрози приватності й безпеки даних. Впровадження ШІ в освіту супроводжується збором дуже великих масивів персональних даних про студентів, викладачів, освітні процеси. Ці дані включають не тільки традиційну академічну інформацію, але й поведінкові моделі, емоційні реакції, біометричні показники та інші чутливі категорії інформації.

Основна загроза полягає в можливості неправомірного використання одержуваних даних третіми сторонами. Багато освітніх платформ належать приватним компаніям, які можуть використовувати дані студентів для комерційних цілей, включаючи рекламу, спрямовану на цільову аудиторію, або продаж інформації іншим організаціям. Особливо проблематичною є ситуація з даними неповнолітніх, які потребують додаткового захисту згідно з міжнародними стандартами.

Кібербезпека стає критично важливим фактором, оскільки освітні заклади часто не мають достатніх ресурсів для забезпечення належного рівня захисту даних. За даними компанії Emsisoft [8] у 2022 році зафіксовано понад 40 серйозних кібератак на освітні заклади в США, що призвело до втрати конфіденційності мільйонів записів студентів.

Додатковою проблемою є трансгранична передача даних, особливо актуальна для міжнародних освітніх програм, або використання хмарних сервісів. Різні юрисдикції мають відмінні вимоги щодо захисту персональних даних, що ускладнює забезпечення відповідності всім регуляторним вимогам;

2) етичні виклики та алгоритмічні упередження. Системи ШІ не є нейтральними – вони відображають упередження й стереотипи, закладені в даних, на яких вони навчалися. В освітньому контексті це може призвести до дискримінації студентів за расовою, гендерною, соціально-економічною або іншими ознаками.

Наприклад, якщо система оцінювання навчалася на даних, які містять історичні упередження щодо успішності різних груп студентів, вона може несправедливо занижувати оцінки представникам певних меншин. Дослідження MIT [9] показало, що системи розпізнавання емоцій мають значно вищий рівень помилок при аналізі осіб з темним кольором шкіри.

Проблема «чорної скриньки» – неможливості пояснити, як саме алгоритм прийняв певне рішення – особливо критична в освіті, де студенти й викладачі мають право знати, на основі чого формуються оцінки й рекомендації. Відсутність прозорості може підірвати довіру до системи і створити правові проблеми;

3) загрози академічної доброчесності. Розвиток генеративних технологій ШІ, таких як GPT-4, Claude або Bard, створює нові виклики для академічної доброчесності. Студенти можуть використовувати ці інструменти для написання есе, розв'язання задач або навіть створення повних досліджень, що ускладнює оцінку справжніх знань і навичок.

Традиційні системи детекції плагіату виявляються неефективними проти контенту, згенерованого ШІ, оскільки такий текст є технічно оригінальним. Це вимагає розробки нових методів виявлення згенерованого тексту і переосмислення самих принципів оцінювання в освіті.

Водночас виникає питання: чи є використання ШІ для виконання завдань порушенням академічної доброчесності, або це новий інструмент, який студенти повинні навчитися ефективно використовувати. Різні освітні заклади мають кардинально різні підходи до цього питання, що створює плутанину й нерівність;

4) ризики дегуманізації освіти. Надмірне покладання на автоматизовані системи може призвести до втрати важливого людського елемента в освіті. Освітній процес не обмежується лише передачею знань – він включає розвиток емоційного інтелекту, соціальних навичок, критичного мислення і моральних цінностей, які найкраще формуються через людську взаємодію.

Зменшення безпосереднього контакту між викладачами і студентами може негативно вплинути на мотивацію, особливо для студентів, які потребують емоційної підтримки. Дослідження показують, що якість стосунків «викладач-студент» є однією з найважливіших запорук академічного успіху.

Існує також ризик стандартизації освіти, коли алгоритми нав'язують єдиний підхід до навчання, не враховуючи культурні, регіональні або індивідуальні особливості. Це може призвести до втрати різноманітності та творчості в освітньому процесі.

Проте можна зауважити, що впровадження ШІ у сферу освіти також відкриває безмежні можливості щодо удосконалення процесу навчання. Новітні моделі ШІ мають можливості не лише автоматизувати рутинні задачі, але й побудувати кардинально нові підходи до процесу навчання. Серед ключових можливостей можна виділити такі:

1) персоналізація навчання. Найбільшою перевагою впровадження ШІ в освіту є можливість створення справді індивідуалізованого навчального досвіду для кожного студента. Традиційна освітня модель «один розмір для всіх» поступово відходить у минуле, поступаючись місцем адаптивним системам, здатним враховувати унікальні особливості кожного, хто навчається.

Сучасні системи ШІ аналізують не тільки академічні результати студентів, але й глибокі поведінкові моделі їхньої навчальної діяльності. Алгоритми відстежують швидкість читання, час, який студент витрачає на різні типи завдань, частоту помилок у конкретних темах, переваги щодо формату подачі матеріалу і навіть фізіологічні показники, такі як рівень концентрації в різний час доби. На основі цього величезного масиву даних система формує детальний профіль навчальних потреб кожного студента.

Адаптивні освітні платформи, такі як DreamBox для математики або Duolingo для вивчення мов, демонструють вражаючі результати персоналізації. Вони можуть у реальному часі коригувати складність завдань, пропонувати альтернативні пояснення для складних концепцій або, навпаки, прискорювати темп для студентів, які швидко засвоюють матеріал. Дослідження [6] показують, що така персоналізація може підвищити ефективність навчання на 30–50 % порівняно з традиційними методами;

2) демократизація доступу до якісної освіти. ШІ має потенціал кардинально змінити систему глобальної освіти, зробивши високоякісне навчання доступним для мільярдів людей незалежно від їхнього географічного розташування, економічного статусу або фізичних можливостей. Цей аспект особливо актуальний для країн, які розвиваються, а також віддалених регіонів, де доступ до якісних освітніх ресурсів традиційно обмежений.

Онлайн-платформи, керовані ШІ, можуть надавати студентам з будь-якої точки світу доступ до курсів провідних університетів і найкращих викладачів. Технології автоматичного перекладу, інтегровані з системами ШІ, не просто перекладають текст, але й адаптують його до культурного контексту та освітніх традицій конкретного регіону. Це означає, що студент із сільської місцевості в Україні може отримати доступ до курсу з Массачусетського технологічного інституту, адаптованого під його рівень підготовки і культурне середовище.

Прогресивні можливості ШІ особливо проявляються у сфері інклюзивної освіти. Для студентів з порушеннями зору розроблено системи, які не лише озвучують текст, але й описують візуальний контент, діаграми, графіки тощо. Студенти з порушеннями слуху можуть користуватися системами автоматичного розпізнавання мовлення і перетворення його в текст у реальному часі. Для осіб з моторними порушеннями створено адаптивні інтерфейси, які можуть керувати поглядом, голосом або навіть думками через нейроінтерфейси;

3) трансформація ролі викладача і підвищення ефективності навчання. Впровадження ШІ не замінює викладачів, а кардинально трансформує їхню роль, звільняючи від рутинних завдань і даючи можливість зосередитися на творчих та інтерактивних аспектах навчання. Системи ШІ можуть взяти на себе автоматизовану перевірку тестів, домашніх завдань і навіть складних письмових робіт, надаючи детальний зворотний зв'язок кожному студенту.

Сучасні системи автоматизованого оцінювання вже здатні перевіряти не тільки тести з множинним вибором, але й есе, математичні розв'язки, програмний код і навіть творчі проєкти. Вони аналізують не тільки правильність відповіді, але й логіку міркувань, оригінальність підходу, використання джерел і дотримання академічних стандартів. При цьому кожен студент отримує персоналізований зворотний зв'язок із конкретними рекомендаціями щодо покращення.

Генеративні можливості ШІ відкривають нові горизонти для створення навчального контенту. Викладачі можуть використовувати ШІ для генерації різноманітних завдань, тестів, наробок і навчальних сценаріїв, адаптованих під конкретну аудиторію. Системи можуть створювати віртуальні лабораторії та симуляції, які дають можливість студентам отримувати практичний досвід навіть при відсутності дорогого обладнання або за наявності небезпечних умов.

Для більш точної оцінки загроз і можливостей впровадження штучного інтелекту в освітню сферу розроблено спеціалізований веб-застосунок для проведення комплексного SWOT-аналізу з використанням методів експертного оцінювання. Система призначена для аналізу і систематизації думок експертів освітньої галузі щодо впровадження технологій ШІ [10].

Основний функціонал застосунку включає збір структурованих експертних оцінок через інтерактивні онлайн-анкети, автоматичну категоризацію відповідей, візуалізацію результатів у вигляді динамічних SWOT-матриць і генерацію аналітичних звітів з рекомендаціями. Система дає експертам можливість оцінювати різні аспекти впровадження ШІ за шкалою важливості та ймовірності реалізації, а також надає якісні коментарі.

Технічна архітектура системи базується на сучасних веб-технологіях. Фронтенд реалізовано з використанням бібліотеки React.js для створення інтерактивного користувацького інтерфейсу з динамічними формами і візуалізацією даних через бібліотеки D3.js і Chart.js. Бекенд розроблено на програмній платформі Node.js з використанням веб-фреймворка Express.js для обробки API-запитів та об'єктно-реляційної системи керування базами даних PostgreSQL для створення основної бази даних, у якій зберігаються структуровані експертні оцінки.

Результатом роботи є комплексна база знань про загрози й можливості впровадження ШІ в освіту, яка дасть можливість формувати науково обґрунтовані рекомендації для освітніх закладів щодо стратегії цифрової трансформації. Платформа забезпечить можливість проведення регулярного моніторингу змін у сприйнятті технологій ШІ освітньою спільнотою та адаптації рекомендацій відповідно до поточних умов.

Висновки

Впровадження штучного інтелекту в сферу освіти є одним з найбільш значущих викликів і можливостей нашого часу. Технології ШІ мають потенціал кардинально трансформувати освіту, зробивши її більш персоналізованою, доступною та ефективною. Водночас вони створюють серйозні ризики, пов'язані з приватністю, етикою і дегуманізацією освітнього процесу. Успіх впровадження ШІ в освіту залежить від здатності знайти правильний баланс між інноваціями і традиційними цінностями освіти. Це вимагає тісної співпраці між технологами, педагогами, регуляторами і суспільством загалом для розробки етичних стандартів, технічних рішень і педагогічних підходів.

Метою впровадження ШІ в освіту є не заміна людського елемента, а його посилення. Найкращі результати досягаються тоді, коли технології ШІ доповнюють і підтримують роботу викладачів, а не замінюють її. Майбутнє освіти вбачається в гармонійному поєднанні штучного і людського інтелекту, де кожен елемент робить свій унікальний внесок у формування освіченого і гармонійно розвиненого покоління.

Україна, переживаючи складний період у своїй історії, має унікальну можливість стати піонером щодо впровадження інноваційних освітніх технологій. Використання ШІ може не лише допомогти подолати виклики воєнного часу, але й закласти основу для створення передової освітньої системи.

Список використаної літератури

1. Artificial Intelligence in Education Market. *Global Market Insights*. 2023. URL: https://www.gminsights.com/industry-analysis/artificial-intelligence-ai-in-education-market (дата звернення: 15.08.2025).

2. UNESCO survey: Less than 10 % of schools and universities have formal guidance on AI. UNESCO, 2023. URL: https://www.unesco.org/en/articles/unesco-survey-less-10-schools-and-universities-have-formal-guidance-ai (дата звернення: 15.08.2025).

3. Systematic review of research on artificial intelligence applications in higher education – where are the educators? / Zawacki-Richter O., Marín V. I., Bond M., Gouverneur F. *International Journal of Educational Technology in Higher Education*. 2019. № 16. Article 39. URL: https://doi.org/10.1186/s41239-019-0171-0 (дата звернення: 15.08.2025).

4. Holmes W., Bialik M., Fadel C. Artificial Intelligence in Education. Promises and Implications for Teaching and Learning. Boston: Center for Curriculum Redesign, 2019. 168 p. URL: https://discovery.ucl.ac.uk/id/eprint/10139722/ (дата звернення: 15.08.2025).

5. Du Boulay B. Artificial Intelligence in Education and Ethics. ResearchGate. *Handbook of Open, Distance and Digital Education*. 2023. P. 93–108. URL: https://www.researchgate.net/publication/367420600_Artificial_Intelligence_in_Education_and_Ethics (дата звернення: 15.08.2025).

6. Cotton D. R. E., Cotton P. A., Shipway J. R. Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*. Vol. 61 (2). 2024. P. 228–239. URL: https://www.tandfonline.com/doi/full/10.1080/14703297.2023.2190148 (дата звернення: 15.08.2025).

7. Chen L., Chen P., Lin Z. Artificial Intelligence in Education: A Review. *IEEE Access*. 2020. Vol. 8. P. 7526475278. URL: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9069875 (дата звернення: 15.08.2025).

8. Emsisoft. The State of Ransomware in the US: Report and Statistics 2022. *Emsisoft Malware Lab*. 2023. URL: https://www.emsisoft.com/en/blog/43258/ (дата звернення: 15.08.2025).

9. Buolamwini J., Gebru T. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. *Proceedings of Machine Learning Research*. 2018. Vol. 81. P. 7791. URL: https://proceedings.mlr.press/v81/buolamwini18a.html (дата звернення: 15.08.2025).

10. Len V. V., Karaieva N. V. Improving the SWOT analysis of the artificial intelligence implementation in higher education. *Сучасні проблеми наукового забезпечення енергетики* : матеріали XXII Міжнар. наук.-практ. конф. молод. вчених і студ., м. Київ, 22-25 квіт. 2025 р. Київ : КПІ ім. Ігоря Сікорського, Вид-во «Політехніка», 2025. Т. 2. С. 184–186. URL: https://iate.kpi.ua/uploads/p_164_79639699.pdf (дата звернення: 15.08.2025).

References

1. Global Market Insights. (2023). *Artificial intelligence in education market size*. https://www.gminsights.com/industry-analysis/artificial-intelligence-ai-in-education-market

2. UNESCO. (2023). *UNESCO survey: Less than 10 % of schools and universities have formal guidance on AI.* https://www.unesco.org/en/articles/unesco-survey-less-10-schools-and-universities-have-formal-guidance-ai

3. Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? *International Journal of Educational Technology in Higher Education*, 16 (1), Article 39. https://doi.org/10.1186/s41239-019-0171-0

4. Holmes, W., Bialik, M., & Fadel, C. (2019). Artificial intelligence in education promises and implications for teaching and learning. Boston: Center for Curriculum Redesign. 168. https://discovery.ucl.ac.uk/id/eprint/10139722/

5. Du Boulay, B. (2023). Artificial Intelligence in Education and Ethics. ResearchGate. *Handbook of Open, Distance and Digital Education*. 93–108. https://doi.org/10.13140/RG.2.2.18022.88649

6. Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61 (2). 228–239. https://doi.org/10.1080/14703297.2023.2190148

7. Chen, L., Chen, P., & Lin, Z. (2020). Artificial intelligence in education: A review. *IEEE Access*, 8. 75264–75278. https://doi.org/10.1109/ACCESS.2020.2988510

8. Emsisoft. The State of Ransomware in the US: Report and Statistics 2022. (2023). *Emsisoft Malware Lab*. https://www.emsisoft.com/en/blog/43258/

9. Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81. 77–91. https://proceedings.mlr.press/v81/buolamwini18a.html

10. Len, V. V., & Karaieva, N. V. (2025). Improving the SWOT analysis of the artificial intelligence implementation in higher education. *Modern problems of scientific support of energy: Proceedings of the XXII International Scientific and Practical Conference of Young Scientists and Students*? 2. 184–186. Igor Sikorsky Kyiv Polytechnic Institute. https://iate.kpi.ua/uploads/p_164_79639699.pdf

Дата першого надходження рукопису до видання: 18.09.2025
Дата прийнятого до друку рукопису після рецензування: 16.10.2025
Дата публікації:

УДК 004.42:796.06 DOI

Л. І. КУБЛІЙ
кандидат технічних наук, доцент,
доцент кафедри цифрових технологій в енергетиці
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
ORCID: 0000-0002-1015-3209

А. Ю. ХОЛОДНИЦЬКА
магістрант кафедри цифрових технологій в енергетиці
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
ORCID: 0009-0007-1482-232X

СТАТИСТИЧНІ МОДЕЛІ АНАЛІЗУ ТРЕНУВАЛЬНИХ ФАКТОРІВ І ПРОГНОЗУВАННЯ СПОРТИВНИХ РЕЗУЛЬТАТІВ

У статті досліджено проблему оцінки впливу тренувальних факторів на спортивні результати із застосуванням статистичних методів аналізу даних. Основна мета роботи полягає в розробці та практичній апробації методики, яка включає: проведення кореляційного аналізу для відбору найбільш значущих факторів, що впливають на спортивні результати; побудову моделі множинної лінійної регресії з оцінкою параметрів методом найменших квадратів; перевірку статистичної значущості коефіцієнтів регресії та адекватності побудованої моделі за допомогою критеріїв Стьюдента і Фішера-Снедекора відповідно. Для демонстрації методики використано експериментальний набір даних обсягом 50 спостережень, який включає такі тренувальні фактори, як кількість спожитої води, тривалість сну, кількість занять на тиждень, тривалість одного заняття, місце проведення тренування, часовий індекс. Цільовою змінною виступає час подолання дистанції 1000 м. Практична реалізація запропонованої методики виконана у вигляді веб-застосунку з клієнт-серверною архітектурою. Користувач вводить дані про фактори й результати, після чого система автоматично виконує статистичний аналіз – вибирає значущі фактори, оцінює параметри моделі, перевіряє їхню статистичну достовірність, робить прогноз на вказану майбутню дату. Запропонований підхід забезпечує можливість здійснювати персоналізований аналіз тренувального процесу. Використання статистичних методів дає змогу підвищити обґрунтованість управлінських рішень у спортивній підготовці, сприяє досягненню більш стабільних і прогнозованих результатів.

Ключові слова: *статистичні моделі, кореляційний аналіз, множинна лінійна регресія, метод найменших квадратів, критерій Стьюдента, критерій Фішера-Снедекора, прогнозування, тренувальні фактори, спортивні результати, персоналізація тренувального процесу, веб-застосунок.*

L. I. KUBLII
Candidate of Technical Sciences, Associate Professor,
Associate Professor at the Department of Digital Technologies in Energy
National Technical University of Ukraine
“Igor Sikorsky Kyiv Polytechnic Institute”
ORCID: 0000-0002-1015-3209

A. YU. KHOLODNYTSKA
Master’s Student at the Department of Digital Technologies in Energy
National Technical University of Ukraine
“Igor Sikorsky Kyiv Polytechnic Institute”
ORCID: 0009-0007-1482-232X

STATISTICAL MODELS FOR ANALYSING TRAINING FACTORS AND PREDICTING SPORTS RESULTS

The article examines the problem of assessing the impact of training factors on sports results using statistical methods of data analysis. The main objective of the work is to develop and test a methodology that includes: conducting a correlation analysis to select the most significant factors that influence sports results; building a multiple linear regression model with parameter estimation using the least squares method; checking the statistical significance of regression coefficients and the adequacy of the constructed model using Student’s and Fisher-Snedecor criteria, respectively. To demonstrate

© Кублій Л. І., Холодницька А. Ю., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

the methodology, an experimental data set of 50 observations was used, which includes the following training factors: amount of water consumed, duration of sleep, number of sessions per week, duration of one session, training location, time index. The target variable Y is the time taken to cover a distance of 1000 m. The practical implementation of the proposed methodology is carried out in the form of a web application with client-server architecture. The user enters data on factors and results in the form of a table, after which the system automatically performs statistical analysis – selects significant factors, evaluates model parameters, checks their statistical reliability, and calculates a forecast for a specified future date. The proposed approach provides the opportunity to perform a personalized analysis of the training process. The use of statistical methods makes it possible to improve the validity of management decisions in sports training and contributes to achieving more stable and predictable results.

Key words: *statistical models, correlation analysis, multiple linear regression, least squares method, Student's criterion, Fisher-Snedecor criterion, forecasting, training factors, sports results, personalization of the training process, web application.*

Постановка проблеми

У сучасних умовах дедалі більше людей долучаються до занять спортом. Основними мотивами є прагнення до здорового способу життя, підвищення витривалості й сили. Проте навіть за подібних умов – однакових програм, навантажень і тривалості тренувань – результати можуть істотно відрізнятися між окремими людьми. За даними дослідників приблизно 80 % цих відмінностей пояснюються генетичними особливостями організму [1]. Тобто, та сама програма тренувань може по-різному впливати на індивідуальний результат спортсмена. У зв'язку з цим зростає потреба в персоналізованому аналізі факторів, які можуть впливати на результативність занять. До таких чинників можна віднести тривалість сну, рівень гідратації, прийом харчових добавок, кількість занять на тиждень, застосування певного тренувального обладнання тощо.

З метою виявлення закономірностей і побудови моделей прогнозування доцільним є використання статистичних методів. Одним із найбільш поширених інструментів для оцінки впливу кількох незалежних змінних на один результативний показник є множинна лінійна регресія, яка дає можливість кількісно визначити силу впливу кожного фактора і побудувати модель прогнозу спортивного результату. Перед застосуванням регресійного аналізу доцільно провести кореляційне дослідження, яке допоможе встановити силу і напрямок зв'язку між змінними і попередньо відібрати ті змінні, які найбільше впливають на результат. Для оцінки статистичної значущості впливу кожного з факторів на результативний показник можна додатково використати дисперсійний аналіз, у якому порівнюються факторна і залишкова дисперсії.

Аналіз останніх досліджень і публікацій

У [2] викладено базові теоретико-методичні засади побудови статистичних моделей у спортивній підготовці. Зокрема, розглянуто загальні підходи до моделювання процесу підготовки спортсменів, детально проаналізовано метод регресійного аналізу, наведено приклад його застосування на основі даних тесту швидкісно-силового навантаження. Окрему увагу приділено методу екстраполяції, який дає можливість здійснювати прогнозування майбутньої динаміки спортивних результатів на основі виявлених статистичних закономірностей, зокрема світових рекордів.

У [3] обґрунтовано застосування факторного аналізу як математичного інструмента для формування латентних змінних, які описують ключові компоненти тренувального процесу в жіночому гандболі.

У [4] проведено множинний регресійний аналіз показників, які впливають на результати у спринтерському бігу. Дослідження ґрунтувалося на даних світового рейтингу, які після фільтрації становили 1278 записів із початкових 3000. До моделі було включено сім незалежних змінних, для кожної з яких визначався вплив на результативний показник. Отримані результати підтвердили доцільність використання регресійного підходу для кількісної оцінки значущості факторів, водночас автори підкреслюють потребу в пошукові додаткових чинників, здатних більшою мірою визначати кінцевий спортивний результат.

Формулювання мети дослідження

Метою дослідження є розробка і практична реалізація статистичних моделей, які дають можливість оцінити вплив тренувальних факторів на спортивні результати і здійснювати прогнозування на індивідуальному рівні.

Викладення основного матеріалу дослідження

Основою дослідження є застосування статистичних моделей для аналізу тренувальних факторів і прогнозування спортивних результатів.

Будується лінійна регресійна модель. Перед її побудовою виконується кореляційний аналіз, який дає змогу виявити взаємозв'язки між змінними та уникнути мультиколінеарності (існування кореляційного зв'язку між факторними змінними). На цьому етапі відбираються найбільш релевантні й незалежні фактори, які справді можуть впливати на спортивний результат, виключаючи другорядні або надлишкові фактори.

Для кількісної оцінки впливу різних факторів на спортивні показники використовується алгоритм, який реалізує такі етапи:

1) одержання параметрів моделі. Визначення коефіцієнтів рівняння регресії здійснюється методом найменших квадратів, який мінімізує суму квадратів відхилень прогнозованих значень від фактичних експериментальних. Це

дає змогу отримати параметри, які найбільш точно описують залежність спортивного результату від сукупності тренувальних факторів;

2) перевірка статистичної значущості коефіцієнтів рівняння регресії. Для кожного параметра побудованої моделі проводиться оцінка за критерієм Стьюдента [5, с. 65]. Це дає змогу визначити, чи є вплив відповідного фактора на результат випадковим чи статистично достовірним;

3) оцінка адекватності моделі (перевірка статистичної значущості рівняння в цілому). Загальна якість побудованої регресійної моделі перевіряється за допомогою критерію Фішера-Снедекора [5, с. 64]. Цей етап дає можливість оцінити, чи є модель придатною для прогнозування. Якщо модель не проходить перевірку, це сигналізує про необхідність зміни набору факторів або використання більш складної статистичної методики.

Розглянемо приклад побудови регресійної моделі. Щоб наочно продемонструвати застосування описаної методики, використано експериментальний набір даних, який включає такі тренувальні фактори: X_1 – кількість спожитої води (л/добу); X_2 – тривалість сну (години); X_3 – кількість занять на тиждень; X_4 – тривалість одного заняття (хвилини); X_5 – місце проведення тренування (0 – свіже повітря, 1 – спортивний зал); X_6 – часовий індекс (порядковий номер тижня виміру). Цільовою змінною Y є час подолання дистанції 1000 м (хвилини). Фрагмент вхідних даних (обсяг вибірки $n = 50$) подано в таблиці 1.

Таблиця 1

Експериментальний набір даних

№ з/п	X_1	X_2	X_3	X_4	X_5	X_6	Y	№ з/п	X_1	X_2	X_3	X_4	X_5	X_6	Y
1	2,10	7,40	2	60	1	1	6,60	26	2,20	7,70	2	60	1	26	5,35
2	2,30	7,80	3	75	0	2	6,95	27	2,10	7,60	3	60	1	27	5,34
3	2,20	7,60	3	75	1	3	6,58	28	2,00	7,80	3	60	1	28	5,29
4	1,90	8,00	4	90	0	4	6,89	29	1,90	7,90	3	120	1	29	5,29
5	2,00	7,70	4	90	1	5	6,36	30	2,20	7,60	2	90	1	30	5,24
6	2,10	7,60	3	60	0	6	6,85	31	2,10	7,70	2	120	1	31	5,18
7	2,00	7,80	3	60	1	7	6,24	32	2,00	7,60	3	120	1	32	5,16
8	2,30	7,50	3	90	0	8	6,79	33	1,90	7,80	4	75	0	33	5,70
9	2,20	7,60	2	90	1	9	6,12	34	2,10	7,90	4	75	1	34	5,12
10	2,10	7,70	2	120	1	10	6,10	35	2,20	8,00	4	60	0	35	5,63
11	2,00	7,60	3	120	1	11	6,00	36	2,30	7,70	4	60	1	36	5,08
12	1,90	7,80	4	75	0	12	6,65	37	2,00	7,30	2	90	0	37	5,55
13	2,10	7,90	4	75	1	13	5,90	38	2,10	7,40	2	90	1	38	5,05
14	2,20	8,00	4	60	0	14	6,61	39	2,20	7,50	3	75	0	39	5,45
15	2,30	7,70	4	60	1	15	5,93	40	2,00	7,60	3	75	1	40	4,98
16	2,00	7,30	2	90	0	16	6,58	41	2,10	7,70	4	90	0	41	5,38
17	2,10	7,40	2	90	1	17	5,78	42	2,20	7,50	4	90	0	42	5,31
18	2,20	7,50	3	75	0	18	6,55	43	1,90	7,90	4	120	0	43	5,28
19	2,00	7,60	3	75	1	19	5,65	44	2,20	8,00	2	120	0	44	5,25
20	2,10	7,70	4	90	0	20	6,50	45	2,30	7,60	2	75	0	45	5,20
21	2,20	7,50	4	90	1	21	5,61	46	2,00	7,50	2	75	0	46	5,19
22	1,90	7,90	4	120	1	22	5,49	47	2,20	7,70	2	60	0	47	5,17
23	2,20	8,00	2	120	1	23	5,42	48	2,10	7,60	3	60	1	48	4,72
24	2,30	7,60	2	75	1	24	5,43	49	2,00	7,80	3	60	0	49	5,12
25	2,00	7,50	2	75	1	25	5,37	50	1,90	7,90	3	120	1	50	4,70

Зауважимо, що серед факторів є номінативна змінна X_5, але її значення оцифровані (0 і 1), тому замість точкового бісеріального коефіцієнта кореляції Пірсона (його обчислюють у випадку, якщо одна змінна є кількісною, а інша номінативною), формула якого є алгебраїчним спрощенням формули коефіцієнта лінійної парної кореляції Пірсона, всі розрахунки можна виконувати за формулою коефіцієнта лінійної парної кореляції. Також оцифровані значення змінної X_5 можна використовувати при побудові рівняння лінійної регресії.

Перед побудовою регресійної моделі треба виявити найсильніші зв'язки між цільовою змінною Y і факторами $X_1, X_2, \ldots, X_6$, а також дослідити існування кореляційного зв'язку між факторними змінними для уникнення мультиколінеарності (у множинній регресійній моделі не можна використовувати змінні, кореляції між якими за модулем близькі до 1). Для цього знайдемо значення коефіцієнтів лінійної парної кореляції Пірсона. Одержану кореляційну матрицю подано в таблиці 2.

З таблиці 2 видно, що найбільший вплив на цільову змінну Y має фактор X_6 (часовий індекс) із коефіцієнтом кореляції –0,8711, а також помірний вплив чинить фактор X_5 (місце проведення тренування) – –0,3007. Кореляційний зв'язок між факторами X_5 і X_6 дуже слабкий – –0,1839. Тому для побудови моделі методом найменших квадратів доцільно включити тільки два фактори ($m = 2$) – X_5, X_6.

Таблиця 2

Кореляційна матриця

Змінна	Y	X_1	X_2	X_3	X_4	X_5	X_6
Y	1	0,0849	–0,0446	0,1811	–0,1509	–0,3007	–0,8711
X_1	0,0849	1	–0,1582	–0,1790	–0,3155	–0,0464	–0,0720
X_2	–0,0446	–0,1582	1	0,4488	0,1540	–0,0481	0,0418
X_3	0,1811	–0,1790	0,4488	1	–0,1126	–0,1706	–0,0913
X_4	–0,1509	–0,3155	0,1540	–0,1126	1	0,1331	0,0683
X_5	–0,3007	–0,0464	–0,0481	–0,1706	0,1331	1	–0,1839
X_6	–0,8711	–0,0720	0,0418	–0,0913	0,0683	–0,1839	1

Коефіцієнти моделі множинної лінійної регресії обчислюють за формулою методу найменших квадратів:

$$B = (U^T \cdot U)^{-1} \cdot U^T \cdot Y,$$

де: U – матриця факторів (у першому стовпчику матриці розміщуються тільки 1); U_T – транспонована матриця U; Y – вектор спостережуваних значень цільової змінної.

Рівняння регресії записується у вигляді:

$$y = b_0 + b_1 \cdot x_1 + \ldots + b_m \cdot x_m.$$

У результаті обчислень отримано такі значення коефіцієнтів:

$$b_0 = 7{,}1157, \quad b_1 = -0{,}6010, \quad b_2 = -0{,}0413.$$

Отже, побудована модель прогнозування часу подолання дистанції 1000 м має вигляд:

$$y = 7{,}1157 - 0{,}6010 \cdot X_5 - 0{,}0413 \cdot X_6.$$

Порівняння експериментальних і прогнозованих значень продемонстровано на рис. 1.

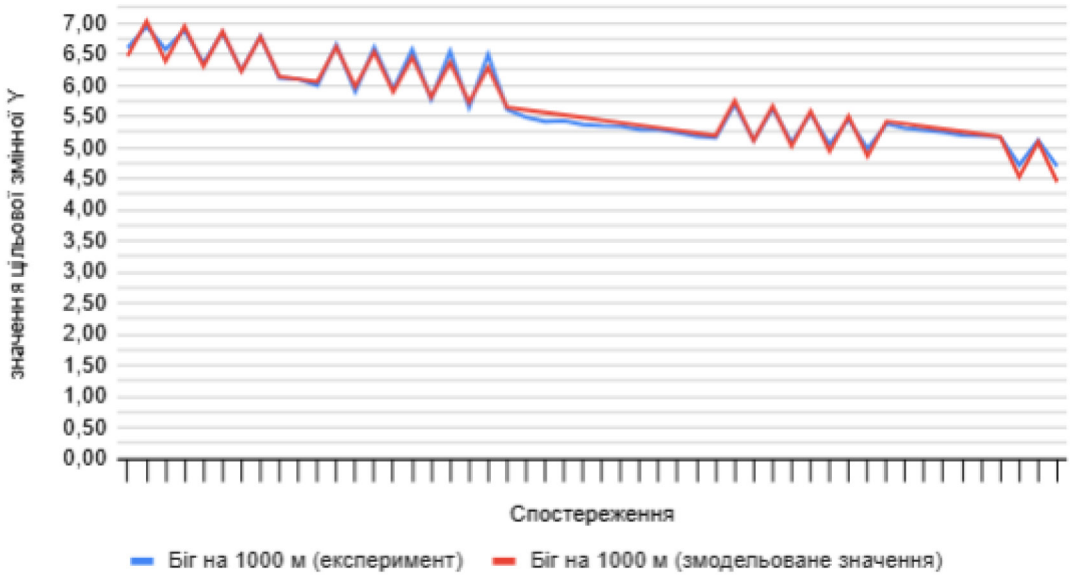


Рис. 1. Порівняння експериментальних і прогнозованих значень Y

Середня похибка прогнозу обчислюється за формулою:

$$\delta_{сер} = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|y_i - y_i^P\right|}{y_i}\cdot 100\,\%,$$

де y_i – експериментальне значення цільової змінної, y_i^P – прогнозоване значення, n – кількість спостережень.

Максимальна похибка визначається так:

$$\delta_{макс} = \max_{i=1\ldots n}\frac{\left|y_i - y_i^P\right|}{y_i}\cdot 100\,\%,$$

За наведеними даними середня похибка прогнозу становить 1,28 %, а максимальна – 5,34 %.

Значущість коефіцієнтів регресійної моделі перевіряється за допомогою критерію Стьюдента; t-статистики обчислюються за формулами:

$$t_0 = |b_0|\frac{\sqrt{(n-m-1)}}{\sigma_0}, \quad t_i = |b_i|\frac{\sqrt{(n-m-1)}}{\sigma_0}\sigma_i, \quad i = 1, \ldots, m,$$

де:

$$\sigma_0 = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(y_j - y_j^P)^2}, \quad \sigma_i = \sqrt{\frac{1}{n}\sum_{j=1}^{n}(x_j^i - x_{сер}^i)^2}.$$

Тут b_0, b_i – оцінки коефіцієнтів регресії, n – кількість спостережень, m – кількість факторів, y_j^P – прогнозоване значення залежної змінної для j-го спостереження, x_j^i – значення i-го фактора для j-го спостереження, $x_{сер}^i$ – середнє значення i-го фактора.

Для коефіцієнтів побудованої регресійної моделі отримано такі розрахункові значення t-статистик:

$$t_0 = 536{,}3578, \quad t_1 = 22{,}3599, \quad t_2 = 44{,}9407.$$

Оскільки всі значення при рівні значущості $\alpha = 0{,}05$ (двобічна критична область) і кількості ступенів свободи $n - m - 1 = 47$ перевищують критичне значення $t_{кр}(0{,}05; 47) = 2{,}0117$, можна зробити висновок про статистичну значущість відповідних коефіцієнтів моделі.

Адекватність побудованої моделі перевіряється за допомогою критерію Фішера-Снедекора, емпіричне значення якого розраховують як відношення факторної і залишкової дисперсій:

$$F_{розрах} = \frac{\sum_{i=1}^{n}(y_i^P - y_{сер})^2(n-m-1)}{\sum_{i=1}^{n}(y_i - y_i^P)^2 m},$$

де n – кількість проведених експериментів; m – кількість факторів у моделі.

Розраховане значення $F_{розрах} = 1074{,}9956$ перевищує критичне $F_{кр}(0{,}05; 2; 47) = 3{,}1951$ для рівня значущості $\alpha = 0{,}05$ і кількості ступенів свободи чисельника $m = 2$ і знаменника $n - m - 1 = 47$, що свідчить про адекватність моделі.

Середньоквадратична похибка прогнозу розраховується за формулою:

$$D = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - y_i^P)^2}$$

і для наведених даних становить 0,0919.

Враховуючи результати проведеного аналізу, можна стверджувати, що побудована модель є статистично надійною, тобто придатною для прогнозування.

Коефіцієнт детермінації (величина достовірності апроксимації)

$$R^2 = \frac{\sum_{i=1}^{n}(y_i^P - y_{сер})^2}{\sum_{i=1}^{n}(y_i - y_i^P)^2}$$

вказує, наскільки побудована регресійна модель відповідає конкретним експериментальним даним – чим ближче його значення до 1, тим кращою є апроксимація. Для одержаної двофакторної моделі $R^2 = 0{,}9786$.

Чим більше факторів буде використано в моделі, тим значення R^2 буде більшим, навіть якщо ці фактори не мають значного впливу на цільову змінну. Крім того, додаючи фактори, можна одержати складнішу модель, яка буде узагальнювати дані гірше, ніж простіша модель. Нормований коефіцієнт детермінації

$$R^2_{норм} = 1 - (1 - R^2)\frac{n-1}{n-m}$$

враховує кількість використовуваних факторів і, якщо доданий фактор не покращує модель, то нормований коефіцієнт зменшиться. Для побудованої моделі $R^2_{норм} = 0{,}9777$.

У побудованій вище моделі використано фактори X_5 і X_6. Якщо проаналізувати кореляційну матрицю (таблиця 2), то можна виділити ще два фактори X_3 і X_4 (відповідні коефіцієнти кореляції – 0,1811 і –0,1509), які гіпотетично можуть покращити модель. Кореляційні зв'язки для всіх можливих пар змінних X_3, X_4, X_5, X_6 незначні – за модулем далекі від 1 (найсильнішим, як і в попередньому випадку, є зв'язок між X_5 і X_6 – –0,1839).

Побудуємо рівняння множинної лінійної регресії для чотирьох факторів:

$$y = 7{,}1366 + 0{,}0084 \cdot X_3 - 0{,}0006 \cdot X_4 - 0{,}5942 \cdot X_5 - 0{,}0412 \cdot X_6.$$

Середня похибка прогнозу становить $\delta_{сер} = 1{,}21$ %, а максимальна – $\delta_{макс} = 5{,}69$ % (у двофакторній моделі відповідно 1,28 % і 5,34 %) – отже, в чотирифакторній моделі середня похибка трохи зменшилася, а максимальна трохи зросла. t-статистики коефіцієнтів рівняння регресії відповідно рівні: $t_0 = 83{,}7049$, $t_1 = 0{,}4981$, $t_2 = 0{,}9882$, $t_3 = 20{,}9885$, $t_4 = 43{,}1839$; критичне значення $t_{кр}(0{,}05;\ 45) = 2{,}0141$; тому статистично значущими є коефіцієнти моделі тільки при змінних X_5 і X_6. Проте $F_{розрах} = 529{,}9704$ перевищує критичне $F_{кр}(0{,}05;\ 4;\ 45) = 2{,}5790$, що підтверджує адекватність чотирифакторної моделі. Середньоквадратична похибка прогнозу становить $\Delta = 0{,}0906$ (для двофакторної моделі – 0,0919). Для чотирифакторної моделі $R^2 = 0{,}9792$ (трохи збільшився порівняно з 0,9786), $R^2_{норм} = 0{,}9774$ (трохи зменшився порівняно з 0,9777). Отже, із двох побудованих моделей перевагу надамо двофакторній. Щодо трифакторної моделі – треба провести дослідження.

Описану вище методику розрахунків реалізовано у веб-застосунку, який надає користувачам інтуїтивно-зрозумілий інтерфейс для введення та аналізу даних.

Спершу користувач заповнює таблицю, вказуючи тренувальні фактори і відповідні спортивні результати. Аналіз може здійснюватися як за стандартними показниками – силою, гнучкістю, швидкістю, витривалістю, спритністю, так і за власними показниками, назви яких користувач задає самостійно (наприклад, тривалість стійки на руках без опори на стіну).

На основі введених даних застосунок виконує статистичний аналіз із використанням вказаного вище алгоритму. Крім оцінки впливу факторів, система реалізує прогнозування спортивних результатів на обрану користувачем майбутню дату, з урахуванням внеску кожного фактора на основі побудованої моделі.

Для гнучкішого контролю тренувального процесу у веб-застосунку передбачено можливість ведення щоденника тренувальних програм, що дає змогу фіксувати кількість повторень, час виконання вправ та інші параметри. Завдяки цьому можна відслідковувати прогрес і коригувати план тренувань відповідно до отриманих результатів.

Архітектуру веб-застосунку реалізовано за клієнт-серверною моделлю. Клієнтську частину розроблено з використанням фреймворку Angular [6], який спеціалізується на створенні односторінкових додатків (SPA). Серверну частину побудовано на базі фреймворку NestJS [7], призначеного для розробки застосунків на платформі Node.js. Вибір фреймворків Angular і NestJS зумовлений їхньою схожою компонентною архітектурою і використанням мови TypeScript, що сприяє підвищенню якості коду, полегшує його підтримку і дає можливість виявляти помилки на ранніх етапах розробки. Для забезпечення зберігання даних використано об'єктно-реляційну систему керування базами даних PostgreSQL [8]. Взаємодія між серверною частиною і базою даних здійснюється через бібліотеку TypeORM [9], яка базується на технології об'єктно-реляційного відображення.

Висновки

У рамках дослідження здійснено аналіз статистичних підходів, які застосовуються для оцінки впливу тренувальних факторів на спортивні результати. Запропоновано методику побудови регресійної моделі із застосуванням методу найменших квадратів для оцінки параметрів, а також застосування критеріїв Стьюдента і Фішера-Снедекора відповідно для перевірки статистичної значущості коефіцієнтів та адекватності моделі.

Практична апробація розробленої моделі на основі експериментальних даних підтверджує її здатність адекватно описувати залежність спортивних показників від факторів тренувального процесу, що свідчить про доцільність використання запропонованого підходу.

Для забезпечення зручності й доступності використання методики створено веб-застосунок, який надає можливість інтерактивно вводити тренувальні фактори і відповідні результати, здійснювати статистичний аналіз і виконувати прогнозування спортивних показників на задану майбутню дату.

Список використаної літератури

1. Старостенко І. А., Деменко М. М. Сучасне бачення генетики у спорті. *Світ наукових досліджень* : матеріали Міжнародної мультидисциплінарної наукової інтернет-конференції, м. Тернопіль, Україна, м. Переворськ, Польща, 16–17 березня 2023 р. Тернопіль : ФОП Шпак В. Б., 2023. Вип. 17. С. 210–211. URL: https://essuir.sumdu.edu.ua/handle/123456789/94601 (дата звернення: 19.08.2025).

2. Костюкевич В. М., Войтенко С. М., Вознюк Т. В. Моделювання та прогнозування в спорті : навч. посібник. Вінниця : Нілан-ЛТД, 2024. 122 с. URL: https://dspace.vspu.edu.ua/items/62557715-16a3-4510-85fc-9e914a144402 (дата звернення: 19.08.2025).

3. Андронов В., Тищенко В. Факторний аналіз як інструмент оптимізації тренувального процесу в жіночому гандболі. *Науковий часопис Українського державного університету імені Михайла Драгоманова. Серія 15*. 2024. Вип. 7(180). С. 1620. URL: https://spppc.com.ua/index.php/journal/article/view/2012/1985 (дата звернення: 19.08.2025).

4. Множинний регресійний аналіз показників, що впливають на спортивний результат у спринтерському бігу / Чухловіна В. та ін. *Науковий часопис Українського державного університету імені Михайла Драгоманова.*

Серія 15. 2024. Вип. 5(178). С. 215–219. URL: https://spppc.com.ua/index.php/journal/article/view/1951/1923 (дата звернення: 19.08.2025).

5. Бідюк П. І., Данилов В. Я., Жиров О. Л. Прикладна статистика : навч. посібник. Київ : КПІ ім. Ігоря Сікорського, 2023. 186 с. URL: https://ela.kpi.ua/server/api/core/bitstreams/678e0b39-a61d-4b34-9c61-f275eb358ed8/content (дата звернення: 19.08.2025).
6. Angular : веб-сайт. URL: https://angular.io/ (дата звернення: 19.08.2025).
7. NestJS : веб-сайт. URL: https://docs.nestjs.com/ (дата звернення: 19.08.2025).
8. PostgreSQL : Documentation : веб-сайт. URL: https://www.postgresql.org/docs/ (дата звернення: 19.08.2025).
9. TypeORM : веб-сайт. URL: https://typeorm.io/ (дата звернення: 19.08.2025).

References

1. Starostenko, I. A., & Demenko, M. M. (2023). Modern view of genetics in sports. *World of Scientific Research : Proceedings of the International Multidisciplinary Scientific Online Conference*, Ternopil, Ukraine; Przeworsk, Poland, March 16–17, 2023 (Vol. 17, pp. 210–211). Ternopil: FOP Shpak V. B. https://essuir.sumdu.edu.ua/handle/123456789/94601 (in Ukrainian).
2. Kostiukevych, V. M., Voitenko, S. M., & Vozniuk, T. V. (2024). Modeling and forecasting in sports [Textbook]. Vinnytsia: Nilan-LTD. 122. https://dspace.vspu.edu.ua/items/62557715-16a3-4510-85fc-9e914a144402 (in Ukrainian).
3. Andronov, V., & Tyshchenko, V. (2024). Factor Analysis as a Tool for Optimizing the Training Process in Women's Handball. *Scientific Journal of Mykhailo Drahomanov Ukrainian State University, Series 15*, 7(180), 16–20. https://spppc.com.ua/index.php/journal/article/view/2012/1985 (in Ukrainian).
4. Chukhlovina, V., Dolbysheva, N., Stepanenko, D., Pechko, H., & Pirogova, K. (2024). Multiple regression analysis of indicators influencing the sports result in sprinting. *Scientific Journal of Mykhailo Drahomanov Ukrainian State University, Series 15*, 5(178), 215–219. https://spppc.com.ua/index.php/journal/article/view/1951/1923 (in Ukrainian).
5. Bidiuk, P. I., Danylov, V. Ya., & Zhyrov, O. L. (2023). Applied statistics [Textbook]. Kyiv: Igor Sikorsky Kyiv Polytechnic Institute. https://ela.kpi.ua/server/api/core/bitstreams/678e0b39-a61d-4b34-9c61-f275eb358ed8/content (in Ukrainian).
6. Angular. (2010–2025). https://angular.io/ (in English).
7. NestJS. (2017–2025). https://docs.nestjs.com/ (in English).
8. PostgreSQL Documentation. (1996–2025). https://www.postgresql.org/docs/ (in English).
9. TypeORM. (2025). https://typeorm.io/ (in English).

Дата першого надходження рукопису до видання: 27.09.2025
Дата прийнятого до друку рукопису після рецензування: 23.10.2025
Дата публікації:

УДК 005.6 DOI

М. В. КУЗЬ
доктор технічних наук, професор,
професор кафедри інформаційних технологій
Карпатський національний університет імені Василя Стефаника
ORCID: 0000-0002-9875-1579

Б. С. НЕЗАМАЙ
кандидат технічних наук, доцент,
завідувач кафедри інформаційних технологій
Карпатський національний університет імені Василя Стефаника
ORCID: 0000-0003-0402-6040

Р. О. ТИЧКОВСЬКИЙ
кандидат технічних наук,
доцент кафедри інформаційних технологій
Карпатський національний університет імені Василя Стефаника
ORCID: 0009-0001-5136-2648

І. Б. ПОЛАТАЙКО
асистент кафедри інформаційних технологій
Карпатський національний університет імені Василя Стефаника
ORCID: 0000-0002-9111-0162

КВАЛІМЕТРІЯ ВИМОГ ДО ПРОГРАМНИХ ЗАСОБІВ НА ПОЧАТКОВИХ ЕТАПАХ ЇХ ЖИТТЄВОГО ЦИКЛУ

У даній статті вирішено науково-практичну проблему кількісної оцінки вимог до програмних засобів (ПЗ) на початкових етапах їхнього життєвого циклу. Традиційні підходи не забезпечують достатньої об'єктивізації процесу аналізу вимог, які часто є якісними та суб'єктивними, що створює високі ризики неефективного розподілу ресурсів та помилок у плануванні. Запропонований кваліметричний підхід ґрунтується на методології шкалювання для переведення нечислових характеристик вимог у точні кількісні значення, що є необхідною умовою для їхнього математичного аналізу та порівняння. Центральним елементом роботи є розроблений метод створення шкал вимог, який формалізує процес експертного оцінювання і забезпечує можливість надання числового рангу якісним критеріям. Зокрема, було створено та обґрунтовано конкретну шкалу для характеристики «Необхідність» вимоги, що дозволяє стандартизовано оцінювати її критичність, чітко розрізняючи обов'язкові, важливі та бажані функціональні/нефункціональні аспекти ПЗ. Це є основою для ефективної пріоритизації. Для інтегральної оцінки складних, багатофакторних властивостей вимог впроваджено використання апарату зваженого математичного очікування. Цей механізм дозволяє розраховувати комплексні характеристики, враховуючи не лише отримані бальні оцінки, але й вагомість (коефіцієнт значущості) кожного окремого критерію, забезпечуючи тим самим реалістичну та всебічну оцінку загальної «цінності» вимоги для проекту. Практична значущість роботи полягає в підвищенні прозорості та обґрунтованості процесу управління вимогами, що дає змогу командам розробників приймати обґрунтовані рішення щодо обсягу робіт та планування ітерацій ще до початку активної фази програмування. Наукова новизна полягає в адаптації та систематизації кваліметричного апарату для специфічної області інженерії вимог на ранньому етапі життєвого циклу. Перспективи подальших досліджень включають: розробку нових шкал для додаткових ключових характеристик (наприклад, «складність реалізації», «ризиковість»), емпіричну валідацію запропонованої методики на різноманітних реальних проектах для підтвердження її універсальності, а також автоматизацію процесу оцінювання та розрахунків шляхом створення програмних інструментів, інтегрованих із системами управління вимогами. Також перспективним є дослідження методів динамічного перерахунку пріоритетів в умовах змін, що базуються на нечіткій логіці або машинному навчанні

Ключові слова: *програмне забезпечення, вимоги, якість вимог, кваліметрія.*

© Кузь М. В., Незамай Б. С. , Тичковський Р. О., Полатайко І. Б., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

M. V. KUZ
Doctor of Science, Professor,
Professor at the Information Technology Department
Vasyl Stefanyk Carpathian National University
ORCID: 0000-0002-9875-1579

B. S. NEZAMAI
Candidate of Technical Science, Associate Professor,
Head of the Information Technology Department
Vasyl Stefanyk Carpathian National University
ORCID: 0000-0003-0402-6040

R. O. TYCHKOVSKYI
Candidate of Technical Science,
Associate Professor at the Information Technology Department
Vasyl Stefanyk Carpathian National University
ORCID: 0009-0001-5136-2648

I. B. POLATAIKO
Assistant at the Information Technology Department
Vasyl Stefanyk Carpathian National University
ORCID: 0000-0002-9111-0162

QUALIMETRY OF SOFTWARE REQUIREMENTS AT THE INITIAL STAGES OF THEIR LIFE CYCLE

This article addresses the scientific and practical problem of quantitative assessment of software requirements (SR) at the initial stages of their life cycle. Traditional approaches lack sufficient objectivity in the requirements analysis process, as requirements are often qualitative and subjective, which creates high risks of inefficient resource allocation and planning errors. The proposed qualimetric approach is based on a scaling methodology designed to convert non-numerical characteristics of requirements into precise quantitative values, a necessary condition for their mathematical analysis and comparison. The central element of this work is the developed method for creating requirements scales, which formalizes the expert evaluation process and allows for assigning a numerical rank to qualitative criteria. Specifically, a concrete scale for the "Necessity" characteristic of a requirement was created and justified. This scale allows for standardized assessment of the requirement's criticality, clearly distinguishing between mandatory, important, and desirable functional/non-functional aspects of the SR. This forms the basis for effective prioritization. For the integral assessment of complex, multi-factor requirement properties, the use of the weighted mathematical expectation apparatus has been implemented. This mechanism enables the calculation of complex characteristics by taking into account not only the acquired scores but also the weight (significance coefficient) of each individual criterion, thereby ensuring a realistic and comprehensive assessment of the overall "value" of the requirement for the project. The practical significance of the work lies in enhancing the transparency and justification of the requirements management process, allowing development teams to make informed decisions regarding the scope of work and iteration planning even before the active programming phase begins. The scientific novelty lies in the adaptation and systematization of the qualimetric apparatus for the specific domain of requirements engineering at the early stage of the life cycle. Perspectives for further research include: the development of new scales for additional key characteristics (e.g., "implementation complexity", "risk level"), empirical validation of the proposed methodology on various real-world projects to confirm its universality, and the automation of the assessment and calculation process through the creation of software tools integrated with requirements management systems. Also promising is the investigation of methods for dynamic priority recalculation under changing conditions, based on fuzzy logic or machine learning.

Key words: *software, requiremets, quality of requirements, qualimetry.*

Постановка проблеми

Сучасна розробка програмного забезпечення (ПЗ) дедалі більше стикається з викликами, пов'язаними зі зростаючою складністю систем, динамічними змінами ринкових умов та потребами користувачів. У цьому контексті етап формування та аналізу вимог є критично важливим. Помилки та неточності, допущені на ранніх стадіях життєвого циклу ПЗ, виявляються найдорожчими для виправлення на пізніших етапах[1].

Однією з ключових проблем, що виникають при роботі з вимогами, є відсутність об'єктивних та уніфікованих методів їхньої кількісної оцінки. Вимоги описуються якісно, що призводить до неоднозначності інтерпретації, суб'єктивності при пріоритизації та складнощів в управлінні змінами. Ця проблема гостро стоїть на ранніх етапах життєвого циклу, коли вимоги ще знаходяться на стадії виявлення та аналізу. Саме тут необхідно мати інструментарій для швидкої та ефективної оцінки таких параметрів, як «необхідність», «повнота», «узгодженість»,

«тестованість», «однозначність» та вплив якості вимог на якість розроблюваного продукту. Без такої кваліметрії ризик «повзучого» обсягу робіт, перевищення бюджету та термінів, а також низької задоволеності кінцевого користувача значно зростає.

Предметом даної статті є розробка системи експертних шкал для кількісної оцінки вимог до програмних засобів. Метою дослідження є створення практичного інструментарію, який дозволить підвищити об'єктивність та прозорість процесу аналізу та пріоритизації вимог, починаючи з найперших етапів життєвого циклу ПЗ. Особлива увага буде приділена диференційованому підходу до оцінки бізнес-вимог, вимог користувача та функціональних вимог, що дозволить більш точно відобразити їхню специфіку та критичність.

Аналіз останніх досліджень і публікацій

Якість кінцевого програмного продукту прямо пропорційна якості специфікації вимог (SRS). Проте, створення точної та повної специфікації залишається складним завданням, яке часто призводить до недоліків, що можуть залишатись непоміченими до пізніх етапів розробки [2].

Використання великих мовних моделей у сфері програмної інженерії є однією з найновіших та найперспективніших тенденцій. Систематичні огляди показують, що великі мовні моделі мають потенціал для трансформації інженерії вимог, зокрема, для підвищення точності та зменшення неоднозначності в специфікаціях. Вони можуть використовуватися для генерації фрагментів коду, верифікації вимог та автоматизації створення тестових випадків. Проте, використання великих мовних моделей для роботи з вимогами не позбавлена проблем. Дослідження виявили проблеми, пов'язані з обробкою доменних вимог, ефективністю підказок (prompt engineering) та нездатністю працювати зі структурованими вхідними даними, що може призводити до помилок та невідповідностей у результатах [3].

Аналіз показує, що, незважаючи на існування всеосяжних і науково обґрунтованих моделей якості вимог, їх практичне застосування в індустрії є непослідовним. Це часто пов'язано з контекстуальними факторами, такими як розмір команди, тип проєкту та відсутність стандартизованих робочих процесів. Дослідження підкреслюють, що проблема полягає не стільки в самій якості моделей, скільки в їхній інтеграції в реальне бізнес-середовище [2, 3, 4].

Формулювання мети дослідження

Предметом даної статті є розробка системи експертних шкал для кількісної оцінки вимог до програмних засобів, яка буде зручною для невеликих команд і дозволить перейти до кількісної оцінки характеристик вимог перш за все для невеликих проєктів. Метою дослідження є створення практичного інструментарію, який дозволить підвищити об'єктивність та прозорість процесу аналізу та пріоритизації вимог, починаючи з найперших етапів життєвого циклу ПЗ. Особлива увага буде приділена диференційованому підходу до оцінки бізнес-вимог, вимог користувача та функціональних вимог, що дозволить більш точно відобразити їхню специфіку та критичність.

Викладення основного матеріалу дослідження

З метою розв'язання поставленої задачі розроблено алгоритм для створення шкали для кількісної оцінки вимог. Цей алгоритм сформульовано як набір кроків, наведених далі.

Крок 1. Для обраної характеристики вимог формулюється критерій, на основі якого встановлюється відношення за типом низька чи висока якість розглядуваної характеристики.

Крок 2. Вибір коротких змістовних маркерів, які для даних характеристик якості відповідають інтуїтивно зрозумілим поняттям оцінки якості за обраним критерієм.

Крок 3. Вибір крайніх значень діапазону числової шкали та ключових віх на шкалі; при чому кількість точок має відповідати кількості обраних маркерів на попередньому кроці.

Крок 4. Для кожної числової віхи на шкалі описується конкретний маркер з поясненням, який гарантує однозначне розуміння цієї шкали при користуванні нею.

Розроблено шкали для оцінки характеристик якості вимог. В таблиці 1 наведено шкалу для оцінювання характеристики «необхідність» для бізнес вимог. В таблицях 2 і 3 наведено розроблену шкалу для оцінювання характеристики «необхідність» для користувацьких вимог та вимог до функцій. Аналогічним чином можна показати шкали для решти типових характеристик вимог.

Таблиця 1

Шкала для оцінки характеристики «необхідність» для бізнес вимог

Змістовний маркер	Значення шкали	Пояснення маркера
відсутня	0	Вимога не відображає реальну потребу бізнесу
незначна	0,2	Реалізація цієї вимоги зробить роботу користувачів трохи зручнішою, але вона не вирішує важливих проблем бізнесу
бажана	0,5	Вимога вирішує певну проблему або значно покращує частину робочого процесу
пріоритетна	0,8	Реалізація цієї вимоги критично впливає на ефективність, продуктивність або задоволеність користувачів. Її відсутність призведе до значних незручностей, або значної втрати часу
критична	1	Абсолютно необхідна вимога. Без неї система є непридатною для використання

Таблиця 2

Шкала для оцінки характеристики «необхідність» для користувацьких вимог

Змістовний маркер	Значення шкали	Пояснення маркера
Неактуальна	0	Ця вимога не відображає реальну потребу користувача або навіть є небажаною. Її реалізація може ускладнити інтерфейс або відволікати користувача
Зручна для покращення	0,2	Реалізація цієї вимоги зробить роботу користувача трохи зручнішою або приємнішою, але вона не вирішує важливих проблем
Помірно бажана	0,5	Вимога вирішує певну помірну проблему користувача або значно покращує частину його робочого процесу. Її відсутність змусить шукати обхідні шляхи
Дуже важлива	0,8	Реалізація цієї вимоги критично впливає на ефективність, продуктивність або задоволеність користувача. Її відсутність призведе до значних незручностей або значної втрати часу для користувача
Обов'язкова/ Незамінна	1	Абсолютно необхідна вимога для того, щоб користувач міг виконати свою основну задачу. Без неї система є непридатною для використання або не задовольняє базові очікування користувача, що може змусити його відмовитись від продукту

Таблиця 3

Шкала для оцінки характеристики «необхідність» для вимог до функцій

Змістовний маркер	Значення шкали	Пояснення маркера
зайва	0	Ця функція не має жодної цінності, не підтримує жодної бізнес, чи користувацької вимоги, або навіть може бути шкідливою для системи (наприклад, викликає конфлікти)
другорядна	0,2	Функція, яка може бути реалізована для невеликого покращення, але її відсутність не впливає на основну роботу системи або задоволення ключових вимог
підтримуюча	0,5	Функція, яка підтримує одну або кілька бізнес- чи користувацьких вимог, робить систему більш повною або ефективною. Її відсутність вимагатиме ручних обхідних шляхів або знизить зручність
ключова	0,8	Функція є критично важливою для реалізації однієї або кількох високих бізнес- чи користувацьких вимог. Вона може бути залежністю для інших важливих функцій або безпосередньо забезпечувати основну функціональність системи
фундаментальна	1	Основна функція, без якої система взагалі не може виконувати свою першочергову місію або не може функціонувати відповідно до мінімальних очікувань

Для оцінювання якості вимог на основі розроблених таким чином шкал (наприклад, для: необхідності, повноти, узгодженості, однозначності, тестованості та відслідковуваності) запропоновано використовувати наступний підхід.

Для інтегрального оцінювання вимоги кожна характеристика має свій ваговий коефіцієнт W_i, при чому $\sum W_i = 1$. Тут W_i встановлюються експертним шляхом для проєкту, в частковому випадку може бути прийнято $W_1 = W_2 = \ldots W_N$ де N – кількість характеристик вимог. Комплексна характеристика вимоги (I) встановлюється як середнє зважене характеристик якості вимоги P_i:

$$I = \frac{1}{N}\sum_{i=1}^{N} P_i \cdot W_i.$$

Аналогічним чином, методом зважених середніх пропонується встановлювати комплексну оцінку якості груп вимог (наприклад, бізнес вимог), та проєкту в цілому. Також можуть встановлюватись обмежуючі мінімально допустимі граничні значення по характеристиках вимог, встановлення яких має призвести до перегляду останніх.

Висновки

У статті вирішено важливу проблему кількісного оцінювання вимог до програмних засобів на початкових етапах їхнього життєвого циклу. Запропонований кваліметричний підхід забезпечує необхідну об'єктивність процесу аналізу вимог, що має вирішальне значення для мінімізації проєктних ризиків та забезпечення високої якості кінцевого продукту. Основним досягненням є розробка методу шкалювання вимог, який, спираючись на принципи кваліметрії та експертного оцінювання, формалізує переведення якісних характеристик вимог у точні кількісні значення, що робить можливим їх коректне порівняння та подальші розрахунки. Як приклад, було створено шкалу для характеристики «Необхідність», що дозволяє стандартизовано оцінювати критичність та пріоритетність кожної вимоги, чітко розмежовуючи обов'язкові, бажані та другорядні аспекти ПЗ. Додатково, запропоновано ефективний механізм розрахунку комплексних характеристик через використання зваженого математичного очікування. Такий підхід забезпечує інтегральне оцінювання загальної «цінності» вимоги. Практичне впровадження розроблених методів дасть змогу командам розробників та замовникам приймати обґрунтовані рішення щодо обсягу робіт, розподілу ресурсів та планування ітерацій ще до початку активної фази програмування. Подальші дослідження у сфері кваліметрії вимог до ПЗ мають бути сфокусовані на розширенні та валідації шкал для інших ключових характеристик (наприклад, «складність реалізації» та «ризиковість»), а також на інтеграції з автоматизованими засобами для створення інструментів, що автоматизують оцінювання та розрахунки. Важливими

напрямками є також дослідження динамічної кваліметрії вимог для перерахунку пріоритетів в умовах постійних змін, а також удосконалення механізмів обробки експертних даних, що враховують компетентність та консенсус експертів. Реалізація цих кроків забезпечить перехід від інтуїтивного до точної інженерії вимог, істотно підвищуючи ефективність розробки програмного забезпечення.

Список використаної літератури

1. Говорущенко Т. О., Боднар М. А., Кушнір В. О. Сучасні проблеми формування та аналізу вимог до програмного забезпечення. *Вимірювальна та обчислювальна техніка в технологічних процесах*. 2019. № 1. С. 45–53. https://doi.org/10.31891/2219-9365-2019-63-1-45-53

2. Ramesh, M. R., Reddy, Ch. S. Metrics for software requirements specification quality quantification. *Computers & Electrical Engineering*. 2021. 96pp. https://doi.org/10.1016/j.compeleceng.2021.107445

3. Han Seong Son Using Large Language Models in Software Requirements *Analysis. Edelweiss Applied Science and Technology. 2025. 3*(9). P. 856–863. https://doi.org/10.55214/25768484.v9i3.5357

4. Gobov, D., and Zuieva, O. 2023. Identifying the dependencies between IT project context and business analysis document content. *Innovative Technologies and Scientific Solutions for Industries*, no. 2(24), 39–53. https://doi.org/10.30837/itssi.2023.24.039

References

1. Hovorushchenko T. O., Bodnar M. A., Kushnir V. O. (2019) Cuchasni problemy formuvannya ta analizu vymoh do prohramnoho zabezpechennya. Vymiryuval′na ta obchyslyuval′na tekhnika v tekhnolohichnykh protsesakh. № 1. C. 45–53.

2. Ramesh, M. R., Reddy, Ch. S. (2021) Metrics for software requirements specification quality quantification. *Computers & Electrical Engineering*. 96pp. https://doi.org/10.1016/j.compeleceng.2021.107445

3. Han Seong Son Using Large Language Models in Software Requirements *Analysis. Edelweiss Applied Science and Technology. 2025. 3*(9). P. 856–863. https://doi.org/10.55214/25768484.v9i3.5357

4. Gobov, D., and Zuieva, O. (2023) Identifying the dependencies between IT project context and business analysis document content. *Innovative Technologies and Scientific Solutions for Industries*, no. 2(24), 39–53. https://doi.org/10.30837/itssi.2023.24.039

Дата першого надходження рукопису до видання: 20.09.2025
Дата прийнятого до друку рукопису після рецензування: 15.10.2025
Дата публікації:

УДК 004.91/.77 DOI

О. Б. КУНГУРЦЕВ
доктор технічних наук, професор,
професор кафедри програмної інженерії
Національний університет «Одеська політехніка»
ORCID: 0000-0002-3207-7315

Т. В. ДІДЕНКО
магістр кафедри програмної інженерії
Національний університет «Одеська політехніка»

МЕТОД ФОРМУВАННЯ ЗАПИТУ НА ПОШУК В ІНТЕРНЕТІ ПАРТНЕРА ПО ІНТЕРЕСАХ З УРАХУВАННЯМ ХАРАКТЕРИСТИК ЗАПИТУ І АДРЕСАНТА

У роботі проаналізовано проблему низької ефективності пошуку співрозмовників за інтересами в месенджерах. Встановлено, що значною мірою це зумовлено неструктурованими запитами користувачів і відсутністю врахування їхніх комунікативних та когнітивних характеристик. Запропоновано класифікацію запитів, що враховує тему, мотивацію, рівень знань і стиль спілкування. Враховано також психологічні та професійні типи користувача, що суттєво впливають на якість і тривалість онлайн-комунікації.

На основі класифікації та індивідуальних характеристик адресанта розроблено модель запиту, що поєднує семантику звернення з рисами особистості користувача. Запропоновано метод формування запиту, який включає чотири етапи: (1) виявлення типу запиту; (2) визначення характеристик адресанта; (3) формування запиту у вигляді кортежу <customer, query>; (4) передача запиту у пошукову програму з метаданими. Для апробації методу створено Telegram-бот, який реалізує усі етапи. Експериментальні результати (400 запитів, 63 % успішних комунікацій, та середній час підбору співрозмовника 11,4 секунди) підтверджують ефективність запропонованого підходу.

Додатково проаналізовано вплив різних параметрів запитів на результати пошуку, зокрема довжини повідомлення, конкретизації інтересів та врахування емоційного стану користувача. Виявлено, що інтеграція когнітивних характеристик і стилю спілкування дозволяє значно підвищити релевантність підбору співрозмовників. Результати дослідження можуть бути використані для розробки більш інтелектуальних систем комунікації в месенджерах, соціальних платформах та освітніх онлайн-сервісах, де важлива швидкість та точність підбору співрозмовників з урахуванням індивідуальних особливостей користувача. Додатково наголошується на можливості адаптації запропонованого підходу під різні культурні та мовні контексти, що розширює сферу його практичного застосування.

Ключові слова: *запит на пошук партнера, математична модель, класифікація запитів, комунікаційні характеристики адресата, пошук партнера, Telegram, визначення термінів.*

O. B. KUNGURTSEV
Doctor of Technical Sciences, Professor,
Professor at the Department of Software Engineering
National University “Odesa Polytechnic”
ORCID: 0000-0002-3207-7315

T. V. DIDENKO
Master’s Student at the Department of Software Engineering
National University “Odesa Polytechnic”

METHOD FOR FORMING A QUERY TO SEARCH FOR AN INTERNET PARTNER BASED ON INTERESTS WITH CONSIDERATION OF QUERY AND ADDRESSEE CHARACTERISTICS

This work analyzes the problem of low efficiency in finding interlocutors based on interests in messengers. It has been established that this is largely due to unstructured user queries and the lack of consideration of their communicative and cognitive characteristics. A classification of queries is proposed, taking into account the topic, motivation, level of knowledge, and communication style. Psychological and professional types of users are also considered, which significantly affect the quality and duration of online communication.

Based on the classification and individual characteristics of the addressee, a query model is developed that combines the semantics of the request with the personality traits of the user. A method for forming a query is proposed, which

© Кунгурцев О. Б., Діденко Т. В., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

includes four stages: (1) identifying the type of query; (2) determining the characteristics of the addressee; (3) forming the query as a tuple <customer, query>; (4) transmitting the query to a search program with metadata. To test the method, a Telegram bot was created that implements all stages. Experimental results (400 queries, 63 % successful communications, and an average interlocutor selection time of 11.4 seconds) confirm the effectiveness of the proposed approach.

Additionally, the influence of various query parameters on search results was analyzed, including message length, specificity of interests, and consideration of the user's emotional state. It was found that integrating cognitive characteristics and communication style significantly improves the relevance of interlocutor selection. The results of the study can be used to develop more intelligent communication systems in messengers, social platforms, and online educational services, where speed and accuracy in selecting interlocutors with consideration of individual user characteristics are important. Furthermore, the adaptability of the proposed approach to different cultural and linguistic contexts is emphasized, expanding its practical application.

Key words: *partner search query, mathematical model, query classification, communicative characteristics of the addressee, partner search, Telegram, term definition.*

Постановка проблеми

З кожним роком зростає кількість людей, які шукають партнерів по інтересах в Інтернеті. За даними DataReportal (2024) [1], понад 60 % користувачів месенджерів хоча б раз на рік намагаються знайти співрозмовника за темою спільних інтересів. У свою чергу, опитування Pew Research Center (жовтень 2019) [2] показало, що 57 % користувачів онлайн-знайомств повідомляли про «дуже» або «досить позитивний» досвід, а 42 % – негативний. Проте вже в 2022 році ці показники дещо змінилися: лише 53 % респондентів описували свій досвід як позитивний, а 46 % – як негативний [3]. Це свідчить про те, що з часом знайти справді сумісну людину для комунікації або спільної діяльності онлайн стає дедалі складніше.

Однією з основних причин цього є нечітке формулювання запитів користувачів. Часто люди не можуть точно описати, кого саме шукають: формулювання типу «людина з почуттям гумору», «приємний співрозмовник» або «хтось, хто мене зрозуміє» не містять конкретних параметрів для ефективного підбору. Це підтверджується дослідженням Kim et al. (2021) [4], де вказано, що понад 70 % користувачів не вказують структурованих характеристик потенційного співрозмовника. Подібні результати отримано також у роботах Lee (2022) [5] та Wang et al. (2023) [6], які підкреслюють важливість точного визначення критеріїв підбору для підвищення якості пошуку.

Другою вагомою проблемою є відмінність цілей у спілкуванні. Одні шукають емоційної підтримки, інші – професійної допомоги, треті – романтичних відносин або дискусій на вузькоспеціалізовані теми. За даними дослідження University of Michigan (2020) [7], майже половина онлайн-користувачів зіштовхуються з тим, що цілі співрозмовників не збігаються, що призводить до швидкого завершення контактів. Аналогічно, робота Smith & Clark (2021) [8] підкреслює, що різниця в мотиваціях є ключовим чинником розриву комунікацій.

Ще одним чинником є відсутність узгодженості щодо рівня обізнаності у певній темі. Наприклад, спроби обговорити складні питання з людиною, яка не має достатньої бази знань, часто призводять до розчарування в розмові. Про це зазначається у роботі Wu & Marelli (2022) [9], де підкреслюється, що ефективна комунікація вимагає когнітивної сумісності, тобто приблизно однакового рівня знань, способу мислення та мовної компетенції. Крім того, останні дослідження Johnson et al. (2024) [10] показують, що когнітивна сумісність суттєво впливає на довготривалість онлайн-спілкування.

Окрім того, часто ігнорується важливий чинник – особистісні та комунікаційні характеристики. Наприклад, інтровертам не завжди комфортно спілкуватися з надто активними або експресивними співрозмовниками, що породжує дискомфорт і знижує тривалість контактів. Відповідно до моделі комунікаційної сумісності, описаної в роботі by Wiemann (1977) [11], ефективна взаємодія потребує узгодження не лише змісту, а й стилю спілкування. Сучасні підходи до цифрової комунікації також вказують на важливість персоналізованого аналізу таких характеристик (Schwartz, 2012 [12]). Аналогічно, дослідження Davis & Green (2023) [13] підтверджують, що врахування комунікаційних стилів суттєво підвищує задоволеність спілкуванням.

Таким чином, можна стверджувати, що існує проблема незадоволеності у результатах пошуку співрозмовника в соціальних мережах, причиною якої може бути невідповідність особистісних та професійних характеристик кореспондентів.

Аналіз останніх досліджень і публікацій

Відповідно до вказаної проблеми проведено аналіз наявних досліджень і підходів до класифікації текстових запитів у контексті пошуку співрозмовників. У роботі [14] розглянуто кілька типів запитів, проте не запропоновано їх повної класифікації, що обмежує застосування у більш широкому спектрі комунікаційних платформ. Крім того, інші дослідники підкреслюють важливість глибинного семантичного аналізу та контекстуальної інтерпретації тексту для підвищення точності класифікації [15]. Дослідження [16] проаналізувало сумісність людей, але зосередилося лише на контексті романтичних стосунків, не враховуючи інші типи взаємодії. Водночас, інші роботи показують, що сумісність у професійному чи соціальному середовищі має суттєві відмінності і вимагає окремих критеріїв оцінки [17]. Стаття [18] містить опис психологічних тестів для оцінки особистісних рис,

однак не враховано їх вагу у процесі підбору співрозмовника, що є важливим для підвищення якості комунікації. Додаткові дослідження підтверджують, що інтеграція особистісних характеристик у системи підбору значно покращує точність рекомендацій і підвищує задоволеність користувачів [19]. У роботі [20] описано алгоритми класифікації текстів, проте без врахування глибини знань користувача, що знижує точність рекомендацій.

Таким чином, відсутня комплексна модель, що поєднує текстовий запит, рівень знань і комунікаційні характеристики, що створює вакуум для розробки ефективних інструментів пошуку.

З аналізу джерел видно, що досліджувалися лише окремі компоненти проблеми, і наразі питання їх комплексного поєднання у єдину систему залишається невирішеним. Відсутність інтегрованого підходу обмежує ефективність існуючих методів пошуку співрозмовників за інтересами, що підкреслює необхідність подальшого розвитку в цій сфері. Виходячи з проведеного аналізу, можна сформулювати мету дослідження.

Формулювання мети дослідження

Метою дослідження є скорочення часу пошуку та підвищення релевантності партнерів за інтересами в умовах обмежених ресурсів месенджера.

Для досягнення мети необхідно:

- ввести класифікацію запитів;
- визначити комунікаційні характеристики користувача;
- розробити математичну модель запиту;
- розробити метод формування запиту;
- створити програмний прототип для апробації моделі.

Викладення основного матеріалу дослідження

Класифікація запитів.

В основу класифікації запитів покладено аналіз їхньої семантичної структури, прагматичних намірів користувача, а також контексту, в якому здійснюється пошук. Методи попередньої обробки коротких текстів та побудови предметних словників, подані у [21; 22], дозволили адаптувати підхід до контексту месенджерів, де довжина запитів обмежена. Запити слід поділяти не лише за темою, а й за рівнем формалізації, глибиною пізнавального інтересу, емоційною мотивацією та соціальними очікуваннями.

На основі аналізу джерел [4–6; 14–20] та результатів експертного оцінювання, запропоновано наступні класи запитів:

Інформаційні (In) – користувач прагне отримати нову інформацію або поглибити знання з певної теми. Часто включають спеціалізовану лексику та уточнюючі терміни («поясніть», «приклад», «порівняння»).

Емоційно-підтримувальні (Ep) – спрямовані на пошук співрозмовника для психологічної підтримки або емоційного контакту («потрібен хтось, хто вислухає», «відчуваю себе самотньо»).

Розважальні (Rz) – орієнтовані на легке, невимушене спілкування, часто без конкретної теми («поспілкуємось про щось цікаве», «розкажи анекдот»).

Дискусійні (D) – передбачають активну взаємодію на складні теми з елементами аргументації, порівнянь, прикладів («хочу обговорити філософські питання», «порівняємо ідеології»).

Соціально-рольові (S) – включають конкретні ролі або контексти (наприклад, «порадьте як вчителю», «хочу поговорити як стартапер»).

Професійні (P) – запити на пошук людей з досвідом або фаховими знаннями в конкретній сфері («мені потрібен юрист», «порада айтішника»).

Індивідуалізовані (G) – поєднують елементи кількох класів і потребують контекстного аналізу для точного розпізнавання.

Математична модель посилання. Представимо посилання кортежом:

$$Message = <customerS, queryS>, \quad (1)$$

де *customerS* – структуроване представлення людини, яка сформулювала запит(далі – адресат); *queryS* – структуроване представлення запиту.

Визначемо представлення адресата

$$customerS = <name, communication, sCharacteristic>, \quad (2)$$

де *name* – customer name; *communication* – засіб спілкування з адресатом, *sCharacteristic* – множина характеристик адресата.

Для визначення характеристик адресата введемо поняття психологічного типу людини [] *sPsychType* і професійного типу людини [] *sProfType*.

Множину психологічних типів представимо у вигляді

$$sPsychType = \{I, E, R, Ir, Em, C\}, \quad (3)$$

де елементи множини мають наступні значення: *I* – Інтроверт, *E* – Екстраверт, *R* – Раціональний, *Ir* – Ірраціональний, *Em* – Емпат, *C* – Контролер.

Множину професійних типів представимо у вигляді

$$sProfType = \{L, A, Cr, V, K, S\}, \quad (4)$$

де елементи множини мають наступні значення: *L* – Лідер, *A* – Аналітик, *Cr* – Креатор, *V* – Виконавець, *K* – Комунікатор, *S* – Стартапер.

На основі сказаного можна визначити множин характеристик адресата

$$sCharacteristic = <sElemPsychType, sElemProfType>, \quad (5)$$

де *sElemPsychType* – множина елементів психологічних характеристик; *sElemProfType* – множина елементів професійних характеристик.

Кожний елемент множини *sElemPsychType* має вигляд:

$$elemPsychType = <psychType_i, psychK_i>, \quad (6)$$

де $psychType_i$ – певний психологічний тип з множини *sPsychType* психологічних типів: $psychType_i \in sPsychType$; $psychK_i$ – нормалізований коефіцієнт психологічного типу, який вказує частину цього типу в психологічній моделі адресата ($psychK_i \le 1$).

Кожний елемент множини sElemProfType має вигляд:

$$elemProfType = <profType_i, profK_i>, \quad (7)$$

де $profType_i$ – певний психологічний тип з множини *sProfType* професійних типів: $profType_i \in sProfType$; $profK_i$ – нормалізований коефіцієнт професійного типу, який вказує частину цього типу в професійній моделі адресата ($psychK_i \le 1$).

Структуроване представлення запиту

$$queryS = <rText, rType_i, sWTerm>, \quad (8)$$

де *rText* – текст запиту адресата; $rType_i$ – тип запиту адресата; *sWTerm* – множина зважених термінів запиту.

Тип запиту є елементом множини типів запитів:

$$rType_i \in sRType.$$

Множину типів запитів відповідно до уведенної класифікації представимо у вигляді

$$sRType = \{In, Ep, Rz, D, S, P, G\}, \quad (9)$$

де елементи множини мають наступні значення: *In* – Інформаційний, *Ep* – Емоційно-підтримувальний, *Rz* – Розважальний, *D* – Дискусійний, *S* – Соціально-рольовий, *P* – Професійний, *G* – Індивідуалізований.

Кожний елемент множини *sWTerm* представлено вигляді:

$$WTerm_j = <term_j, termK_j>, \quad (10)$$

де $term_j$ – один з термінів запиту, $termK_j$ – нормалізований коефіцієнт терміну, який вказує вагу терміну у запиті адресата ($termK_j \le 1$).

Врахування типу людини для певного типу запиту є ключовим елементом побудови ефективної моделі пошуку співрозмовників. Різні типи запитів мають різні комунікативні й емоційні вимоги, тож відповідність між запитом і особистісними характеристиками користувача дозволяє підвищити якість взаємодії. Наприклад, емоційно-підтримувальні запити (тип *Ep*) частіше зустрічаються у користувачів з типами емпат (*Em*) або екстраверт (*E*), тоді як інформаційні або професійні запити (типи *In* та *P*) краще формуються користувачами з типами раціональний (*R*), аналітик (*A*) чи контролер (*C*).

Таким чином, маємо не лише класифікувати запит, а й враховувати, наскільки сформований запит відповідає внутрішнім характеристикам його автора. Якщо між типом людини і типом запиту спостерігається розбіжність, це може свідчити про поверхневу мотивацію або невизначеність користувача, і така ситуація потребує додаткової перевірки або уточнення.

Для обґрунтування цих зв'язків була сформована таблиця сумісності типів користувачів і типів запитів, яка дозволяє автоматично визначити потенційну якість запиту, а також очікувану ефективність комунікації з відповідним типом партнера. Це також дає змогу застосовувати персоналізовану логіку пошуку – наприклад, надавати пріоритет кандидатам, які найбільш відповідають емоційним і стилістичним очікуванням автора запиту.

На основі викладеного матеріалу та проведеного аналізу використаної літератури[9, 11, 12, 13] було розроблено та сформовано таблицю

У таблиці прийнято наступні позначення: + – висока сумісність, ± – залежить від контексту, – – низька сумісність.

Таблиця 1

Взаємна сумісність типів користувачів і типів запитів

Тип запиту / Тип користувача	*In*	*Ep*	*Rz*	*D*	*S*	*P*	*G*
I	+	–	±	+	±	+	±
E	±	+	+	+	+	±	±
R	+	–	±	+	±	+	±
Ir	±	±	+	–	±	–	+
Em	±	+	+	±	±	–	±
C	+	–	–	+	+	+	±
L	±	±	–	+	+	+	±
A	+	–	–	+	±	+	±
Cr	±	+	+	±	±	±	+
V	+	–	–	±	+	+	±
K	±	+	+	+	+	±	±
S	±	+	+	±	+	+	+

Метод формування запиту. На основі моделі посилання розроблено метод формулювання запиту. Метод містить 4 етапи.

Етап 1 – Виявлення типу запиту.

Запит адресата обробляється аналізатором тексту для виділення термінів. Адресат визначає вагу термінів (8). Аналізатор типу запитів відносить запит до певного типу (9).

Етап 2 – Визначення характеристик адресанта.

Якщо адресат вже зареєстрований, то перевіряється, які типи запитів він посилв. Якщо актуальний запит по типу співпадає з минулими, то відбувається перехід до наступного етапу методу. В іншому випадку на основі визначенного типу запиту формується множина тестів, які проходить адресат. По результатам тествання створюється множина характеристик адресата.

Етап 3 – Формвання посилання (Message)

На основі отриманої інформації (текст запиту, тип запиту, множина термінів з вагою, характеристики адресата) створюється запит, для пошуку партнера.

Етап 4 – Передача посилання у пошукову програму.

Сформоване посилання містить крім запита додаткові метаданні, які забезпечують суттєве покращення результатів пошуку. Але для реалізації переваг такого представлення запиту потрібна спеціальна пошукова програма, розгляд якої не є предметом данної роботи.

Діаграма діяльності, яка відповідає запропонованом методу, представлена на рис. 1.

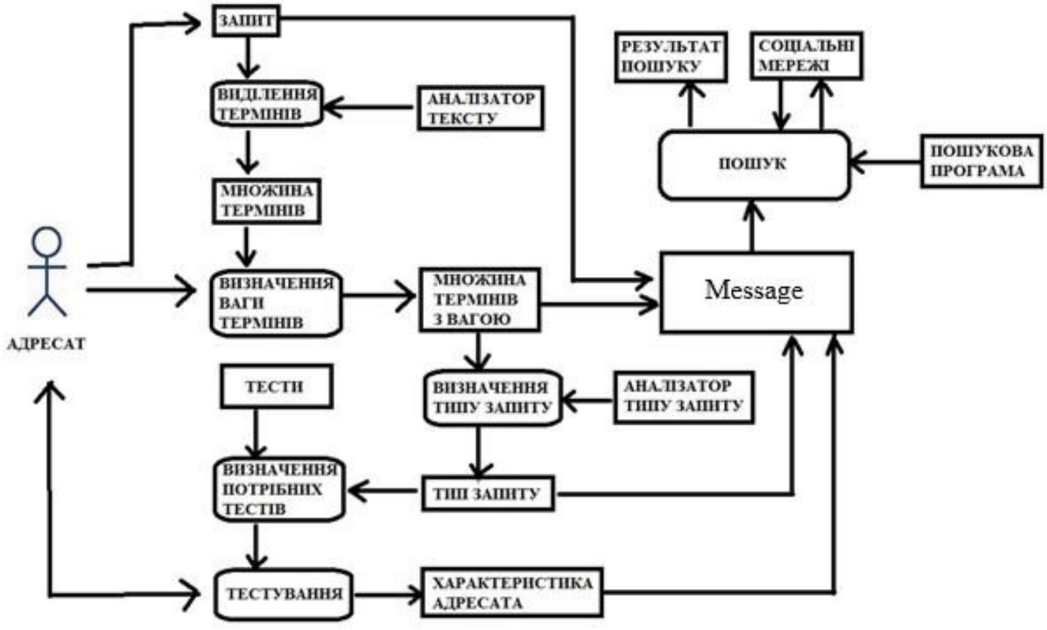


Рис. 1. Діаграма діяльносмті по створенню посилання

Апробація і обговорення результатів дослідження.

Запропонований метод був реалізований у вигляді Telegram-бота, який працює у діалоговому режимі. Він приймає текст запиту і автоматично визначає його тип, глибину та стиль. Далі виконуєтья коротке опитування адресата для визначення його психологічного й професійного типу. Після цього здійснюється пошук найбільш релевантного співрозмовника за моделлю багаторівневої відповідності – яка враховує змістовні, лексичні, психологічні та поведінкові характеристики користувачів. Словникові структури, використані у боті для обробки термінів запиту, частково були автоматично сформовані за методикою [21; 23], що дозволило скоротити час реакції системи.

Проведення експериментів. Апробація програми проводилася у квітні–травні 2025 року, тривала 6 тижнів та охопила 20 унікальних учасників, які сумарно сформували 400 текстових запитів. У посиланнях було представлено всі типи запитів (*In*, *Ep*, *Rz*, *D*, *S*, *P*, *G*), а також різні стилі комунікації й глибини знань.

Модель дозволяла обробляти текстові запити користувачів, визначаючи їхню тематичну спрямованість, рівень обізнаності, а також психологічні й професійні характеристики, що забезпечувало багаторівневу персоналізацію процесу взаємодії. У процесі апробації використовувалися як симульовані, так і реальні запити респондентів, що відрізнялися за метою, стилем спілкування та ступенем деталізації.

Результати експерименту:

- Успішних комунікаційних пар: 252;
- Середній час підбору співрозмовника: 11,4 секунди;
- Задоволеність користувачів (оцінка 4 або 5 із 5): 81 %;
- Середня довжина запиту: 18–22 слова.

Результати іншої програми:

Як приклад для порівняння було обрано платформу Replika.ai – чат-бот з елементами штучного інтелекту, що позиціонується як віртуальний співрозмовник і використовує адаптивне навчання. На відміну від запропонованої нами моделі, Replika не використовує попередньої класифікації типу запиту користувача та не проводить чіткого тестування для визначення психологічних або професійних характеристик. Водночас, вона застосовує вбудовану модель емоційного аналізу, однак без глибокої структуризації запиту.

Апробація була здійснена за аналогічних умов: 20 користувачів, кожен із яких сформував по 20 запитів (разом 400). Перед початком спілкування користувачі не проходили тестування, запити вводилися у вільній формі. Програма використовувала стандартний алгоритм обробки тексту без побудови кортежу <user, query> та без вагових коефіцієнтів запиту.

Результати експерименту:

- Успішних комунікаційних пар (оцінено за критерієм «відповідь відповідає запиту»): 167;
- Середній час підбору відповіді (реакція бота): 7,8 секунди;
- Задоволеність користувачів (оцінка 4 або 5 із 5): 54 %;
- Середня довжина запиту: 16–21 слів.

Результати показали, що врахування кількох рівнів класифікації – зокрема типу запиту, когнітивної глибини та комунікативних особливостей – дозволяє істотно підвищити релевантність підбору партнерів для спілкування.

Це свідчить про високий потенціал у сфері персоналізованого підбору співрозмовників, що базується на багаторівневому аналізі запитів, із урахуванням як змістової, так і соціально-комунікативної сумісності.

Висновки

Запропоновано класифікацію текстових запитів, яка ґрунтується не лише на тематиці, а й на мотивації, стилі спілкування, когнітивному рівні та соціальних очікуваннях користувача, що дозволяє більш точно формалізувати потреби користувача під час пошуку співрозмовника.

Запропоновано математичну модель посилання, яка поєднує інформацію пов’язану з запитом (тип запиту, терміни тощо) з характеристиками адресата (психологічні та професійні риси). Запропонована модель є основою для виконанняк семантичного і стилістичного аналізу запиту, визначення релевантності співрозмовників та побудови сценаріїв попереднього відсіву кандидатів.

Розроблено метод формування посилання, який на основі запропонованої моделі посилання визначає джерела даних, послідовність і взаємозв’зок дій, що формють всі характеристики посилання. Метод дає змогу значно підвищити ефективність наступного етапу пошуку співрозмовника.

Розроблено програмнй прототип у вигляді Telegram-бота, який реалізє всі етапи методу формування посилання і організує пошук найбільш релевантного співрозмовника за моделлю багаторівневої відповідності.

Апробація прийнятих рішень показала їх ефективність: було досягнуто 63 % успішних комунікацій, та задоволеність користувачів в районі 81 % що підтверджує практичну доцільність використання розробленої моделі.

Метою подальших досліджень є розробка методів виявлення і точнення характеристик співрозмовника.

Список використаної літератури

1. DataReportal. Digital 2024 Global Overview Report [Electronic resource]. Available at: https://datareportal.com/reports/digital-2024-global-overview-report, 2024

2. Pew Research Center. The Virtues and Downsides of Online Dating [Electronic resource]. Available at: https://www.pewresearch.org/internet/2020/02/06/the-virtues-and-downsides-of-online-dating, 2020

3. Pew Research Center. From Looking for Love to Swiping the Field: Online Dating in the U.S. [Electronic resource]. Available at: https://www.pewresearch.org/internet/2023/02/02/from-looking-for-love-to-swiping-the-field-online-dating-in-the-u-s/, 2023

4. Kim, S., Lee, H., Park, Y. User Intentions and Frustrations in Social Chat Apps. Proceedings of CHI [Electronic resource]. Available at: https://dl.acm.org/doi/10.1145/3411764.3445675, 2021

5. Lee, J., et al. Structured Criteria for Effective Partner Search in Digital Platforms. Journal of Digital Communication, 2022.

6. Wang, X., et al. Enhancing Online Matching Accuracy via Query Precision. Computers in Human Behavior, 2023.

7. University of Michigan. Mismatch in Online Communication Goals, 2020.

8. Smith, A., Clark, T. Motivational Differences and Communication Breakdown. Communication Studies Quarterly, 2021.

9. Wu, J., Marelli, M. Cognitive Compatibility in Digital Communication. Journal of Communication Studies, 2022.

10. Johnson, M., et al. The Impact of Cognitive Compatibility on Online Interaction Duration. Digital Psychology Review, 2024.

11. Wiemann, J. M. Explication and Test of a Model of Communicative Competence. Human Communication Research, 3(3), 195–213. https://doi.org/10.1111/j.1468-2958.1977.tb00514.x, 1977.

12. Schwartz, S. H. An Overview of the Schwartz Theory of Basic Values. Online Readings in Psychology and Culture, 2(1). https://doi.org/10.9707/2307-0919.1116, 2012

13. Davis, L., Green, P. Communication Styles and Satisfaction in Online Conversations. Journal of Social Interaction, 2023.

14. Chen, L., Zhang, D., Xu, Y. A Comprehensive Survey on Query Classification and Understanding for Search Engines. ACM Computing Surveys, 52(3), Article 56. https://dl.acm.org/doi/10.1145/3319605, 2019

15. Manning, R., Raghavan, P., Schütze, H. Introduction to Information Retrieval, 2008; Aggarwal, C., Zhai, C. Mining Text Data, 2012.

16. Finkel, E., Eastwick, P. W., Matthews, J. Compatibility and Relationship Success: The Role of Similarity in Online Dating. Journal of Social and Personal Relationships, 35(8), 1075–1097. DOI: 10.1177/0265407517716952, 2018

17. Lee, J., Carpenter, K., 2018; González-Bailón, S., 2013.

18. Goldberg, L. R. The Structure of Phenotypic Personality Traits. American Psychologist, 48(1), 26–34. DOI: 10.1037/0003-066X.48.1.26, 1993

19. Gosling, S., et al., 2004; Pennebaker, J., Mehl, M., Niederhoffer, K., 2003.

20. Aggarwal, C. C., Zhai, C. A Survey of Text Classification Algorithms. In: Mining Text Data, pp. 163–222. Springer. DOI: 10.1007/978-1-4614-3223-4_6, 2012

21. Kungurtsev, O. B., Mileiko, I. I., Novikova, N. O. Technology for Automated Construction of Domain Dictionaries with Special Processing of Short Documents. Радіоелектроніка, інформатика, управління, 2024 (4), 148. https://doi.org/10.15588/1607-3274-2023-4-14

22. Kungurtsev, O., Zinovatna, S., Potochniak, I., Novikova, N. Development of Methods for Pre-Clustering and Virtual Merging of Short Documents for Building Domain Dictionaries. Eastern-European Journal of Enterprise Technologies, 2020. 5(2 (107)), 39–47. https://doi.org/10.15587/1729-4061.2020.215190

23. Kungurtsev, O., Novikova, N., Kozhushan, M. Automation of Searching for Terms in the Explanatory Dictionary. Праці Одеського політехнічного університету, 2020. № 3(62), 91–100. https://old-pratsi.op.edu.ua/app/webroot/articles/1612863161.pdf

References

1. DataReportal. Digital 2024 Global Overview Report [Electronic resource]. Available at: https://datareportal.com/reports/digital-2024-global-overview-report, 2024

2. Pew Research Center. The Virtues and Downsides of Online Dating [Electronic resource]. Available at: https://www.pewresearch.org/internet/2020/02/06/the-virtues-and-downsides-of-online-dating, 2020.

3. Pew Research Center. From Looking for Love to Swiping the Field: Online Dating in the U.S. [Electronic resource]. Available at: https://www.pewresearch.org/internet/2023/02/02/from-looking-for-love-to-swiping-the-field-online-dating-in-the-u-s/, 2023.

4. Kim, S., Lee, H., Park, Y. (2021) User Intentions and Frustrations in Social Chat Apps. Proceedings of CHI [Electronic resource]. Available at: https://dl.acm.org/doi/10.1145/3411764.3445675

5. Lee, J., et al. (2022) Structured Criteria for Effective Partner Search in Digital Platforms. Journal of Digital Communication.

6. Wang, X., et al. (2023) Enhancing Online Matching Accuracy via Query Precision. Computers in Human Behavior.

7. University of Michigan. Mismatch in Online Communication Goals, 2020.

8. Smith, A., Clark, T. (2021) Motivational Differences and Communication Breakdown. Communication Studies Quarterly.

9. Wu, J., Marelli, M. (2022) Cognitive Compatibility in Digital Communication. Journal of Communication Studies.

10. Johnson, M., et al. (2024) The Impact of Cognitive Compatibility on Online Interaction Duration. Digital Psychology Review.

11. Wiemann, J. M. (1977) Explication and Test of a Model of Communicative Competence. Human Communication Research, 3(3), 195–213. https://doi.org/10.1111/j.1468-2958.1977.tb00514.x

12. Schwartz, S. H. (2012) An Overview of the Schwartz Theory of Basic Values. Online Readings in Psychology and Culture, 2(1). https://doi.org/10.9707/2307-0919.1116

13. Davis, L., Green, P. (2023) Communication Styles and Satisfaction in Online Conversations. Journal of Social Interaction.

14. Chen, L., Zhang, D., Xu, Y. A. (2019) Comprehensive Survey on Query Classification and Understanding for Search Engines. ACM Computing Surveys, 52(3), Article 56. https://dl.acm.org/doi/10.1145/3319605

15. Manning, R., Raghavan, P., Schütze, H. (2012) Introduction to Information Retrieval, 2008; Aggarwal, C., Zhai, C. Mining Text Data.

16. Finkel, E., Eastwick, P. W., Matthews, J. (2018) Compatibility and Relationship Success: The Role of Similarity in Online Dating. Journal of Social and Personal Relationships, 35(8), 1075–1097. DOI: 10.1177/0265407517716952

17. Lee, J., Carpenter, K., 2018; González-Bailón, S., 2013.

18. Goldberg, L. R. (1993) The Structure of Phenotypic Personality Traits. American Psychologist, 48(1), 26–34. DOI: 10.1037/0003-066X.48.1.26

19. Gosling, S., et al., 2004; Pennebaker, J., Mehl, M., Niederhoffer, K., 2003.

20. Aggarwal, C. C., Zhai, C. A. (2012) Survey of Text Classification Algorithms. In: Mining Text Data, pp. 163–222. Springer. DOI: 10.1007/978-1-4614-3223-4_6

21. Kungurtsev, O. B., Mileiko, I. I., Novikova, N. O. (2024) Technology for Automated Construction of Domain Dictionaries with Special Processing of Short Documents. Radio Electronics, Computer Science, Control, (4), 148. https://doi.org/10.15588/1607-3274-2023-4-14

22. Kungurtsev, O., Zinovatna, S., Potochniak, I., Novikova, N. (2020) Development of Methods for Pre-Clustering and Virtual Merging of Short Documents for Building Domain Dictionaries. Eastern-European Journal of Enterprise Technologies, 5(2 (107)), 39–47. https://doi.org/10.15587/1729-4061.2020.215190

23. Kungurtsev, O., Novikova, N., Kozhushan, M. (2020) Automation of Searching for Terms in the Explanatory Dictionary. Proceedings of Odessa Polytechnic University, Issue 3(62), 91–100. https://old-pratsi.op.edu.ua/app/webroot/articles/1612863161.pdf

Дата першого надходження рукопису до видання: 23.09.2025
Дата прийнятого до друку рукопису після рецензування: 20.10.2025
Дата публікації:

УДК 004.93 DOI

В. О. КУРЕНКО
аспірант кафедри інфокомунікаційної інженерії імені В. В. Поповського
Харківський національний університет радіоелектроніки
ORCID: 0009-0006-6269-2390

Д. В. АНДРУШКО
кандидат технічних наук,
старший викладач кафедри інфокомунікаційної інженерії
імені В. В. Поповського
Харківський національний університет радіоелектроніки
ORCID: 0009-0004-0749-9729

С. В. ПШЕНИЧНИХ
кандидат технічних наук, старший науковий співробітник,
доцент кафедри інфокомунікаційної інженерії імені В. В. Поповського
Харківський національний університет радіоелектроніки
ORCID: 0000-0003-0533-2306

СИСТЕМА АВТЕНТИФІКАЦІЇ ЗА ОБЛИЧЧЯМ З ВИКОРИСТАННЯМ ТРАНСФЕРНОГО НАВЧАННЯ НЕЙРОННИХ МЕРЕЖ

Використання згорткових нейронних мереж у процесі аналізу зображень мало вплив і на системи біометричної автентифікації за обличчям, які відіграють важливу роль у забезпеченні інформаційної безпеки. Сучасні моделі для вирішення цієї задачі досягають високої точності, та мають складну архітектуру, що потребує багато часу на проєктування. Застосування трансферного навчання сприяє спрощенню процесу тренування моделей нейронних мереж, забезпечуючи при цьому достатньо високий рівень точності. В роботі запропоновано підхід до побудови систем автентифікації особи за обличчям, в межах якого використано попередньо навчену модель згорткової нейронної мережі, яка за рахунок використання трансферного навчання була адаптована під задачу автентифікації особи за обличчям. Такий підхід спрощує процес проєктування архітектури нейронної мережі та значно скорочує час на її навчання. В ході дослідження було розроблено дві архітектури згорткової нейронної мережі, які відрізнялись налаштуванням шарів базової моделі для вилучення ознак, кількістю повнозв'язних шарів та їх параметрів, та методами регуляризації. Було зібрано два набори зображень обличчя, в подальшому вони були використані для навчання та тестування моделей нейронних мереж, отримано графіки залежності функції втрат від кількості епох в ході навчання, і виконано перевірку на тестовій підмножині. Для перевірки адекватності рішення було розроблено та протестовано прототип системи, отримана точність автентифікації висока на обох наборах зображень, тож запропонований підхід може використовуватись у реальних системах біометричної автентифікації за обличчям, а також потенційно може бути доповнений механізмами безпеки, наприклад модулем для перевірки підміни обличчя. Подальші дослідження в цьому напрямі пропонується спрямувати на вдосконалення методів попередньої обробки зображень з камери та доповнення навчального набору даних.

Ключові слова: *біометрична автентифікація; розпізнавання обличчя; комп'ютерний зір; згорткові нейронні мережі; трансферне навчання; інформаційна безпека.*

V. O. KURENKO
Postgraduate Student at the V. V. Popovskyy Department
of Infocommunication Engineering
Kharkiv National University of Radio Electronics
ORCID: 0009-0006-6269-2390

D. V. ANDRUSHKO
PhD, Senior Lecturer at the V. V. Popovskyy Department
of Infocommunication Engineering
Kharkiv National University of Radio Electronics
ORCID: 0009-0004-0749-9729

© Куренко В. О., Андрушко Д. В., Пшеничних С. В., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

S. V. PSHENYCHNYKH
PhD, Senior Researcher,
Associate Professor at the V. V. Popovskyy Department
of Infocommunication Engineering
Kharkiv National University of Radio Electronics
ORCID: 0000-0003-0533-2306

FACE AUTHENTICATION SYSTEM USING TRANSFER LEARNING OF NEURAL NETWORKS

The use of convolutional neural networks in the process of image analysis has also had an impact on biometric face authentication systems, which play an important role in ensuring information security. Modern models for solving this problem achieve high accuracy and have a complex architecture that requires a lot of time to design. The use of transfer learning helps to simplify the process of training neural network models, while ensuring a sufficiently high level of accuracy. The paper proposes an approach to building face authentication systems, which uses a pre-trained convolutional neural network model, which was adapted to the face authentication task through the use of transfer learning. This approach simplifies the process of designing the neural network architecture and significantly reduces the time for its training. During the study, two convolutional neural network architectures were developed, which differed in the settings of the base model layers for feature extraction, the number of fully connected layers and their parameters, and regularization methods. Two sets of face images were collected, and subsequently they were used to train and test neural network models, graphs of the dependence of the loss function on the number of epochs during training were obtained, and verification was performed on the test subset. To verify the adequacy of the solution, a prototype of the system was developed and tested, the obtained authentication accuracy was high on both sets of images, so the proposed approach can be used in real biometric face authentication systems, and can also potentially be supplemented with security mechanisms, such as a module for checking face spoofing. Further research in this direction is proposed to be directed at improving the methods of pre-processing images from the camera and supplementing the training dataset.

Key words: *biometric authentication; face recognition; computer vision; convolutional neural networks; transfer learning; information security.*

Постановка проблеми

Враховуючи активний розвиток апаратних платформ для додатків штучного інтелекту, таких як графічні процесори та нейропроцесори, збільшення доступності та поширеності цифрових камер, а також успішні дослідження нейронних мереж, інтеграція додатків штучного інтелекту в людську діяльність є незворотним процесом. Розпізнавання обличчя, яке використовується у багатьох сферах, зокрема і для автентифікації особи, є широко відомим прикладом використання штучного інтелекту для підвищення рівня інформаційної безпеки [1]. Автентифікація особи за зображенням обличчя полягає у розпізнаванні за принципом «свій-чужий», шляхом порівняння вхідного вектора ознак з еталоном. Актуальні дослідження в цій галузі спрямовані на підвищення точності систем автентифікації, шляхом модифікації архітектур нейронних мереж, розробку нових функцій активації та втрат.

Існуючі методи автентифікації особи за обличчям покладаються на складні архітектури нейронних мереж зі специфічними функціями втрат. Такі рішення показують високу точність у складних тестових сценаріях. Однак, системи автентифікації особи можуть використовуватись і у більш простих умовах. Актуальною задачею є розробка методів автентифікації особи за обличчям, які забезпечують достатню точність, але використовують простіші архітектури нейронних мереж та функції втрат, що має спростити їх процес розробки і введення в експлуатацію.

Аналіз останніх досліджень і публікацій

Автентифікація є одним з застосувань розпізнавання обличчя, і складається з таких підзадач, як локалізація обличчя, вилучення ознак із зображення обличчя, та класифікація отриманих ознак. Відповідно до цього були сформовані окремі напрями досліджень, спрямовані на вдосконалення методів та засобів для їх вирішення. В рамках даної статті буде розглянуто останні дві підзадачі.

Ранні методи для вилучення ознак із зображення обличчя, такі як EigenFaces, FisherFaces, гістограма орієнтованих градієнтів, аналіз незалежних компонентів, мали низку обмежень, зокрема чутливість до змін освітлення, виразу обличчя, повороту голови [2–5]. Поява згорткових нейронних мереж для класифікації зображень відкривала нові перспективи для розпізнавання обличчя. На базі класичних та сучасних архітектур для класифікації зображень було розроблено чимало моделей для вилучення ознак обличчя. Так в роботі [6] пропонується сімейство легких архітектур згорткових нейронних мереж для розпізнавання обличчя, які засновані на базі GhostNetV1 [7] та GhostNetV2 [8]. В запропонованому рішенні було модифіковано шари мережі та функцію активації. Отримане рішення має точність, співмірну з існуючими моделями, що мають значно вищу обчислювальну складність.

Через деякий час після появи моделей згорткових нейронних мереж для розпізнавання обличчя, стала зрозумілою обмеженість існуючих функцій втрат, що вимірюють різницю між прогнозованим та фактичним значенням. Це викликало необхідність їх вдосконалення, чому була присвячена низка робіт [9–11]. В роботі [12] пропонується

адаптивна функція втрат, яка при тренуванні моделей нейронних мереж звертає увагу на неправильно класифіковані екземпляри даних, при цьому коригуючи цей процес шляхом врахуванням якості зображення, що в результаті покращує точність моделі. Автори [13] зазначають, що підходи, які були запропоновані в [10–11], не пристосовані до варіативності даних всередині класу та між різними класами в реальних наборах даних, що може обмежувати здатність до розрізнення та узагальнення у моделі, і пропонують гнучку функцію втрат, яка краще працює з такими особливостями даних, і покращує точність розпізнавання обличчя. В роботі [14] пропонується гібридна функція втрат на базі [11], застосування якої сприяє кращій дискримінаційності при навчанні згорткової нейронної мережі, внаслідок чого підвищується точність моделі.

Окрім цього, сформовано ще один напрям досліджень методів біометричної автентифікації за обличчям, спрямований на вирішення проблем безпеки. В роботах [15–16] було запропоновано схеми автентифікації із захистом даних. Однією з поширених проблем систем автентифікації за обличчям є підміна обличчя, шляхом демонстрації зображень і відео з легітимним користувачем, а також використанням діпфейків. Для подолання цих проблем запропоновано ряд методів для виявлення підміни обличчя та перевірки «активності» людини. Так в роботі [17] для перевірки запропоновано порівнювати фотоплетизмограми обличчя та кінчика пальця на руці. Для цього, при автентифікації на смартфоні одночасно ведеться зйомка з основної та фронтальної камер, що мають знімати обличчя та палець, і перевіряється узгодженість результатів аналізу отриманого відеоряду. А в роботі [18] запропоновано схему автентифікації, в якій для перевірки людини використовується модель машинного навчання для аналізу кадрів відеоряду. До того ж, замість періодичної автентифікації користувача, яка більш вразлива до атак, пропонується інтегрувати в систему функцію відстеження обличчя, де після його втрати автентифікація відбувається повторно. Для боротьби з підміною обличчя можуть застосовуватись не тільки камери, а й інші сенсори, зокрема мікрофони. В роботах [19–20] пропонується як разом з візуальними даними з камери, так і окремо, аналізувати відлуння звукового сигналу від людського обличчя.

Нерідко компрометація систем автентифікації за обличчям несе значні ризики для власників, тож сучасні дослідження направлені на захист від атак, та на підвищення точності розпізнавання обличчя, шляхом вдосконалення методів вилучення ознак та їх класифікації. Це призводить до уповільнення процесів розробки та розгортання таких систем, та збільшення їх вартості, що є не завжди виправданим.

Формулювання мети дослідження

Метою дослідження є зменшення часових витрат на побудову архітектури та навчання нейронної мережі для системи біометричної автентифікації за обличчям, забезпечуючи при цьому достатню точність.

Викладення основного матеріалу дослідження

З огляду на високий рівень точності моделей штучного інтелекту на базі згорткових нейронних мереж, запропоноване рішення спиралось саме на них. Типова структура згорткової нейронної мережі наведена на рисунку 1. Вона складається з двох частин: для вилучення ознак та для класифікації. Зображення проходить через множину шарів згортки та об'єднання, отримані ознаки зображення передаються у повнозв'язні шари для подальшої класифікації.

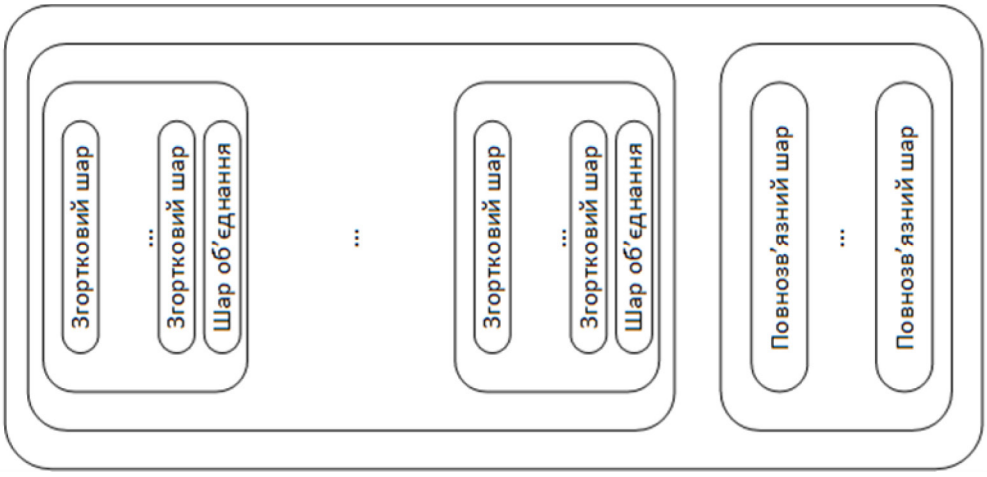


Рис. 1. Типова структура згорткової нейронної мережі

Згорткові шари є основним елементом мережі, призначені для виявлення особливостей зображення. Для цього використовується множина фільтрів, розмір яких зазвичай складає 3 на 3 пікселі, які переміщуються по зображенню ковзаючим вікном, і на кожному кроці відбувається розрахунок згортки. Отриманий результат передається на шар об'єднання, яке зазвичай буває максимальне або середнє, тобто на вихід передається найбільше або середнє значення з множини. Використання шарів об'єднання дозволяє зменшити розмірність даних, і сприяє

підвищенню стійкості моделі до зсувів та поворотів зображення. Після проходження останнього шару об'єднання, отримані ознаки зображення передаються на повнозв'язні шари для класифікації отриманих ознак зображення.

При реалізації моделі згорткової нейронної мережі було застосовано трансферне навчання, тобто знання отримані моделлю під час навчання на одному завданні, використовуються для швидкого та ефективного вирішення нового завдання, замість навчання моделі з нуля. Це спрощує процес проєктування архітектури нейронної мережі та значно скорочує часові витрати на її навчання. До базової моделі, яка була попередньо навчена на наборі даних ImageNet [21], було додано шари, що тренувалися на невеликому наборі даних з обличчями. Для проведення дослідження було розроблено програму з використанням мови Python. В ході експерименту емпіричним шляхом було підібрано дві архітектури мережі (рис. 2).

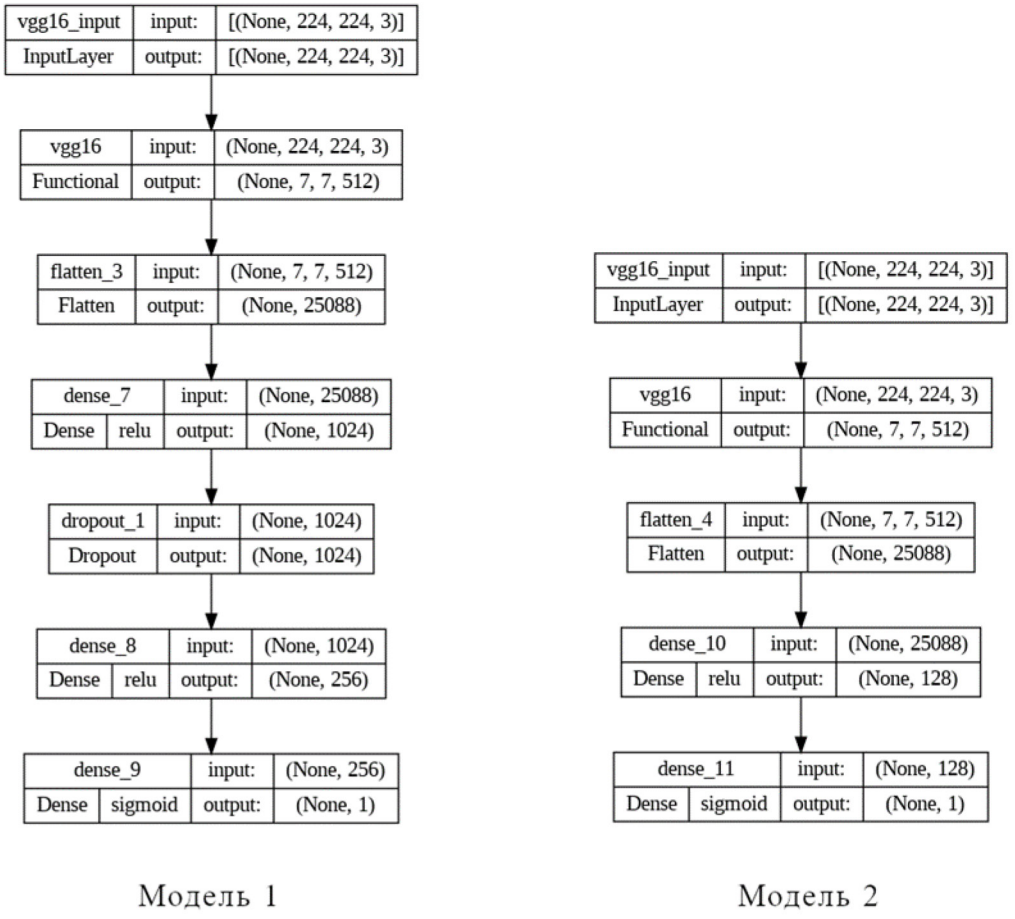


Рис. 2. Обрані архітектури мережі

Як базову було обрано модель VGG16 [22]. У першій моделі нейронної мережі, для вилучення ознак було використано всі п'ять блоків згортки-об'єднання базової моделі з підготовленими вагами. Кольорове зображення розміром 224 на 224 пікселі подавалось на вхід до згорткового шару першого блоку. З шару об'єднання останнього блоку, тривимірний масив даних передавався на шар Flatten для перетворення в одновимірний масив, і таким чином було сформовано ознаки зображення. Далі їх було передано на повнозв'язні шари для класифікації. Між першим і другим повнозв'язним шаром було встановлено шар Dropout для регуляризації в ході навчання мережі. На останньому повнозв'язному шарі використовувалась сигмоїдальна функція активації, яка повертає число від 0 до 1, за яким можна оцінити наближеність до класу «свій». У другій моделі нейронної мережі, для вилучення ознак також було використано всі п'ять блоків згортки-об'єднання базової моделі. Та на відміну від першої, у другій моделі шари п'ятого блоку згортки-об'єднання були доступними для навчання, що дозволяло краще враховувати специфічні для обличчя риси. Друга модель також відрізнялась простішою архітектурою класифікатора, було використано меншу кількість повнозв'язних шарів, регуляризація в мережі не застосовувалась.

В ході експерименту використовувалось два набори даних. Кожен набір складався з двох класів (рис. 3): у першому знаходилась одна особа, в другому – декілька чужих осіб. Також кожен набір було розділено на навчальну, валідаційну та тестову вибірки.

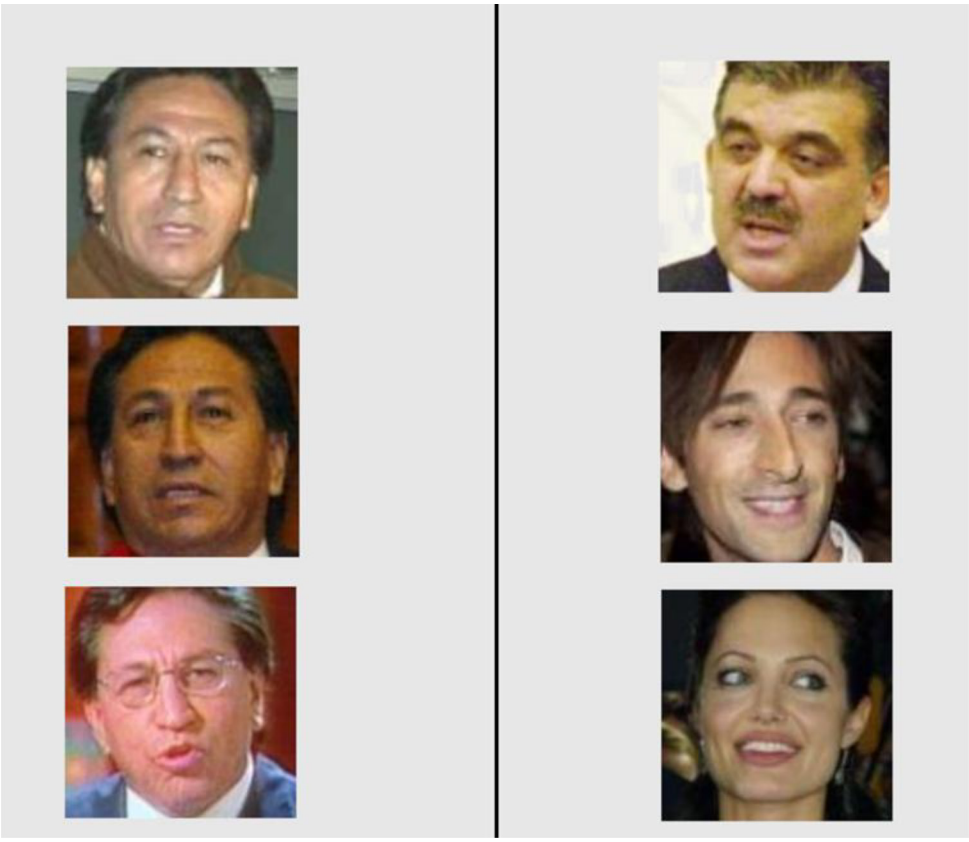


Рис. 3. Приклад зображень з другого набору даних

Перший набір даних було зібрано самостійно, з використанням камери. На рисунку 4 наведено графік залежності функції втрат від кількості епох для першої і другої моделі на тренувальній та валідаційній підмножині першого набору даних.

Другий набір даних було зібрано з набору зображень обличчя LFW [23]. На рисунку 5 наведено графік залежності функції втрат від кількості епох для першої і другої моделі на тренувальній та валідаційній підмножині другого набору даних.

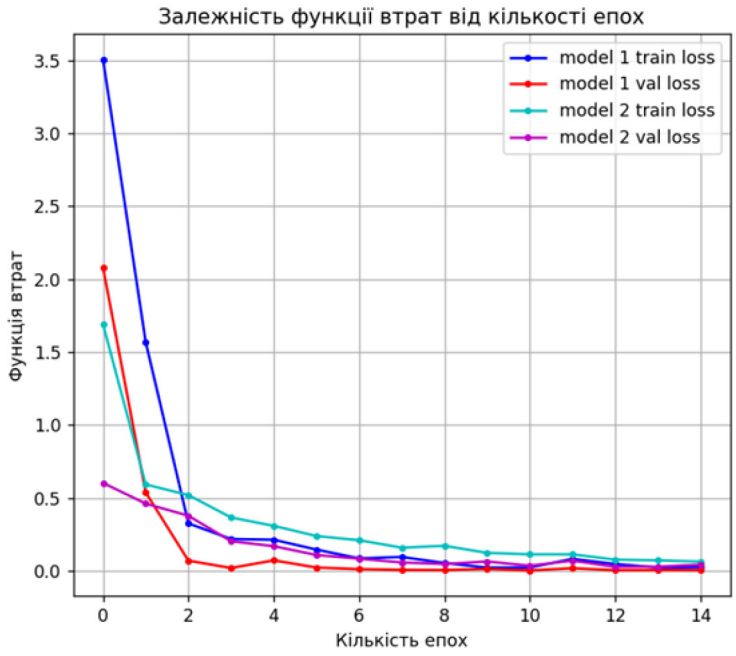


Рис. 4. Залежність функції втрат від кількості епох на першому наборі даних

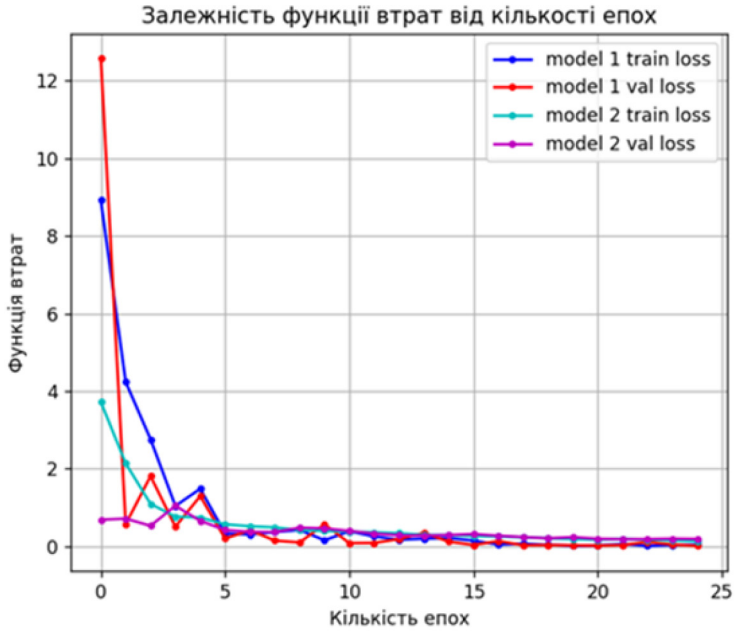


Рис. 5. Залежність функції втрат від кількості епох на другому наборі даних

Як видно з рисунків 4–5, в обох випадках функція втрат наближається до нуля. Отримані результати свідчать про успішне навчання усіх розглянутих моделей згорткових нейронних мереж і на першому, і на другому наборах даних, перенавчання та недонавчання моделей відсутнє.

Також було проведено перевірку на тестових вибірках, яка продемонструвала високу точність запропонованих моделей як на першому, так і на другому наборах даних, що свідчить про адекватність та працездатність використаного підходу.

Висновки

За рахунок впровадження згорткових нейронних мереж, системи для автентифікації за обличчям показують високу точність роботи. В них використовуються складні моделі нейронних мереж, що потребують значних фінансових та часових витрат на розробку та розгортання, що не завжди є доцільним. Застосування трансферного навчання сприяє спрощенню процесу навчання моделей нейронних мереж, забезпечуючи при цьому достатньо високий рівень точності.

У статті запропоновано підхід до побудови систем автентифікації особи за обличчям. В його межах було використано попередньо навчену модель згорткової нейронної мережі, яка за рахунок використання трансферного навчання була адаптована під задачу автентифікації особи. Такий підхід пришвидшив процеси проєктування архітектури та навчання нейронної мережі. Розроблений прототип системи було успішно протестовано на декількох наборах даних, підхід довів свою адекватність та працездатність, автентифікація особи відбувається з високою точністю. Перспективним напрямом подальших досліджень є методи доповнення навчального набору даних та попередньої обробки зображень з камери.

Список використаної літератури

1. Kortli Y., Jridi M., Al Falou A., Atri M. Face recognition systems: A survey. *Sensors*. 2020. Vol. 20, no. 2. P. 342. URL: https://doi.org/10.3390/s20020342

2. Turk M. A., Pentland, A. Face recognition using eigenfaces. *Computer Vision and Pattern Recognition* : Proceedings of the IEEE Computer Society Conference, Maui, 03–06 June 1991. Maui, 1991. P. 586–591. URL: https://doi.org/10.1109/CVPR.1991.139758

3. Belhumeur P. N., Hespanha J. P., Kriegman D. J. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence*. 1997. Vol. 19, no. 7. P. 711–720. URL: https://doi.org/10.1109/34.598228

4. Dalal N., Triggs B. Histograms of oriented gradients for human detection. *Computer Vision and Pattern Recognition* : Proceedings of the IEEE Computer Society Conference, San Diego, 20–25 June 2005. San Diego, 2005. P. 886–893. URL: https://doi.org/10.1109/CVPR.2005.177

5. Bartlett M. S., Movellan J. R., Sejnowski T. J. Face recognition by independent component analysis. *IEEE Transactions on neural networks*. 2002. Vol. 13, no. 6. P. 1450–1464. URL: https://doi.org/10.1109/TNN.2002.804287

6. Ghostfacenets: Lightweight face recognition model from cheap operations / Alansari M. et al // *IEEE Access*. 2023. Vol. 11. P. 35429–35446. URL: https://doi.org/10.1109/access.2023.3266068

7. Ghostnet: More features from cheap operations / Han K. et al // *Computer vision and pattern recognition*: Proceedings of the IEEE/CVF conference, Seattle, 13–19 June 2020. Seattle, 2020. P. 1580–1589. URL: https://doi.org/10.1109/CVPR42600.2020.00165

8. GhostNetv2: Enhance cheap operation with long-range attention / Tang Y. et al // *Neural Information Processing Systems* : Proceedings, New Orleans, 28 November – 09 December 2022. New Orleans, 2022. P. 9969–9982. URL: https://doi.org/10.48550/arXiv.2211.12905

9. Wen Y., Zhang K., Li Z., Qiao Y. A discriminative feature learning approach for deep face recognition. *European conference on computer vision* : Proceedings. Cham, 2016. P. 499–515. URL: https://doi.org/10.1007/978-3-319-46478-7_31

10. Deng J., Guo J., Xue N., Zafeiriou S. Arcface: Additive angular margin loss for deep face recognition. *Computer Vision and Pattern Recognition* : Proceedings of the IEEE Computer Society Conference, Long Beach, 15–20 June 2019. Long Beach, 2019. P. 4690–4699. URL: https://doi.org/10.1109/CVPR.2019.00482

11. Cosface: Large margin cosine loss for deep face recognition / Wang H. et al // *Computer vision and pattern recognition* : Proceedings of the IEEE/CVF conference, Salt Lake City, 18-23 June 2018. Salt Lake City, 2018. P. 5265–5274. URL: https://doi.org/10.1109/CVPR.2018.00552

12. Kim M., Jain A. K., Liu X. Adaface: Quality adaptive margin for face recognition. *Computer vision and pattern recognition* : Proceedings of the IEEE/CVF conference, New Orleans, 18–24 June 2022. New Orleans, 2022. P. 18750–18759. URL: https://doi.org/10.1109/CVPR52688.2022.01819

13. Boutros F., Damer N., Kirchbuchner F., Kuijper A. Elasticface: Elastic margin loss for deep face recognition. *Computer vision and pattern recognition* : Proceedings of the IEEE/CVF conference, New Orleans, 18–24 June 2022. New Orleans, 2022. P. 1578–1587. URL: https://doi.org/10.1109/CVPRW56347.2022.00164

14. Unitsface: Unified threshold integrated sample-to-sample loss for face recognition / Jia X. et al // *Neural Information Processing Systems* : Proceedings, New Orleans, 10–16 December 2023. New Orleans, 2023. P. 32732–32747. URL: https://doi.org/10.48550/arXiv.2311.02523

15. Wang X., Xue H., Liu X., Pei Q. A privacy-preserving edge computation-based face verification system for user authentication. *IEEE Access*. 2019. Vol. 7. P. 14186–14197. URL: https://doi.org/10.1109/ACCESS.2019.2894535

16. Lin W. H., Wang P., Tsai C. F. Face recognition using support vector model classifier for user authentication. *Electronic Commerce Research and Applications*. 2016. Vol. 18. P. 71–82. URL: https://doi.org/10.1016/j.elerap.2016.01.005

17. Your face your heart: Secure mobile face authentication with photoplethysmograms / Chen Y. et al // *Computer Communications* : Proceedings of the IEEE Conference, Atlanta, 01–04 May 2017. Atlanta, 2017. P. 1–9. URL: https://doi.org/10.1109/INFOCOM.2017.8057220

18. Smith-Creasey M., Albalooshi F. A., Rajarajan M. Continuous face authentication scheme for mobile devices with tracking and liveness detection. *Microprocessors and Microsystems*. 2018. Vol. 63. P. 147-157. URL: https://doi.org/10.1016/j.micpro.2018.07.008

19. Zhou B., Xie Z., Ye F. Multi-modal face authentication using deep visual and acoustic features. *Communications* : Proceedings of the IEEE International Conference, Shanghai, 20–24 May 2019. Shanghai, 2019. P. 1–6. URL: https://doi.org/10.1109/ICC.2019.8761776

20. Aface: Range-flexible anti-spoofing face authentication via smartphone acoustic sensing / Xu Z. et al // *Interactive, Mobile, Wearable and Ubiquitous Technologies*. 2024. Vol. 8, no. 1. P. 1–33. URL: https://doi.org/10.1145/3643510

21. Imagenet: A large-scale hierarchical image database / Deng J. et al. // *Computer Vision and Pattern Recognition* : Proceedings of the IEEE Conference, Miami, 20–25 June 2009. Miami, 2009. P. 248–255. URL: https://doi.org/10.1109/CVPR.2009.5206848

22. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition. *Learning Representations* : Proceedings of the International Conference, San Diego, 07-09 May 2015. San Diego, 2015.

23. Huang G. B, Mattar M., Berg T., Learned-Miller E. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. *Faces in'Real-Life'Images: detection, alignment, and recognition* : Proceedings. 2008

References

1. Kortli, Y., Jridi, M., Al Falou, A., & Atri, M. (2020). Face recognition systems: A survey. *Sensors*<, *20*(2), 342. https://doi.org/10.3390/s20020342

2. Turk, M. A., & Pentland, A. (1991). Face recognition using eigenfaces. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 586–591. IEEE. https://doi.org/10.1109/CVPR.1991.139758

3. Belhumeur, P. N., Hespanha, J. P., & Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on pattern analysis and machine intelligence, 19*(7), 711-720. https://doi.org/10.1109/34.598228

4. Dalal, N., & Triggs, B. (2005). Histograms of oriented gradients for human detection. In *IEEE Computer Society Conference On Computer Vision And Pattern Recognition*, 886–893. IEEE. https://doi.org/10.1109/CVPR.2005.177

5. Bartlett, M. S., Movellan, J. R., & Sejnowski, T. J. (2002). Face recognition by independent component analysis. *IEEE Transactions on neural networks, 13*(6), 1450–1464. https://doi.org/10.1109/TNN.2002.804287

6. Alansari, M., Hay, O. A., Javed, S., Shoufan, A., Zweiri, Y., & Werghi, N. (2023). GhostFaceNets: Lightweight Face Recognition Model From Cheap Operations. , 35429–35446. https://doi.org/10.1109/access.2023.3266068

7. Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., & Xu, C. (2020). Ghostnet: More features from cheap operations. In *IEEE/CVF Conference On Computer Vision And Pattern Recognition*, 1580–1589. IEEE. https://doi.org/10.1109/CVPR42600.2020.00165

8. Tang, Y., Han, K., Guo, J., Xu, C., Xu, C., & Wang, Y. (2022). GhostNetv2: Enhance cheap operation with long-range attention. In *Neural Information Processing Systems*, 9969–9982. https://doi.org/10.48550/arXiv.2211.12905

9. Wen, Y., Zhang, K., Li, Z., & Qiao, Y. (2016). A Discriminative Feature Learning Approach for Deep Face Recognition. In *European conference on computer vision*, 499–515. Springer. https://doi.org/10.1007/978-3-319-46478-7_31

10. Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). Arcface: Additive angular margin loss for deep face recognition. In *IEEE/CVF conference on computer vision and pattern recognition*, 4690–4699. IEEE. https://doi.org/10.1109/CVPR.2019.00482

11. Wang, H., Wang, Y., Zhou, Z., Ji, X., Gong, D., Zhou, J., Li, Z., & Liu, W. (2018). CosFace: Large Margin Cosine Loss for Deep Face Recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5265–5274. IEEE. https://doi.org/10.1109/CVPR.2018.00552

12. Kim, M., Jain, A. K., & Liu, X. (2022). Adaface: Quality adaptive margin for face recognition. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 18750–18759. IEEE. https://doi.org/10.1109/CVPR52688.2022.01819

13. Boutros, F., Damer, N., Kirchbuchner, F., & Kuijper, A. (2022). Elasticface: Elastic margin loss for deep face recognition. In *IEEE/CVF conference on computer vision and pattern recognition*, 1578–1587. IEEE. https://doi.org/10.1109/CVPRW56347.2022.00164

14. Jia, X., Zhou, J., Shen, L., & Duan, J. (2023). Unitsface: Unified threshold integrated sample-to-sample loss for face recognition. In *Neural Information Processing Systems*, 32732–32747. https://doi.org/10.48550/arXiv.2311.02523

15. Wang, X., Xue, H., Liu, X., & Pei, Q. (2019). A privacy-preserving edge computation-based face verification system for user authentication. *IEEE Access, 7*, 14186–14197. https://doi.org/10.1109/ACCESS.2019.2894535

16. Lin, W. H., Wang, P., & Tsai, C. F. (2016). Face recognition using support vector model classifier for user authentication. *Electronic Commerce Research and Applications, 18*, 71–82. https://doi.org/10.1016/j.elerap.2016.01.005

17. Chen, Y., Sun, J., Jin, X., Li, T., Zhang, R., & Zhang, Y. (2017). Your face your heart: Secure mobile face authentication with photoplethysmograms. In *IEEE Conference on Computer Communications*, 1–9. IEEE. https://doi.org/10.1109/INFOCOM.2017.8057220

18. Smith-Creasey, M., Albalooshi, F. A., & Rajarajan, M. (2018). Continuous face authentication scheme for mobile devices with tracking and liveness detection. *Microprocessors and Microsystems, 63*, 147–157. https://doi.org/10.1016/j.micpro.2018.07.008

19. Zhou, B., Xie, Z., & Ye, F. (2019). Multi-modal face authentication using deep visual and acoustic features. In *IEEE International Conference on Communications*, 1–6. IEEE. https://doi.org/10.1109/ICC.2019.8761776

20. Xu, Z., Liu, T., Jiang, R., Hu, P., Guo, Z., & Liu, C. (2024). AFace: Range-flexible anti-spoofing face authentication via smartphone acoustic sensing. *Interactive, Mobile, Wearable and Ubiquitous Technologies*, *8*(1), 1–33. https://doi.org/10.1145/3643510

21. Deng, J., Dong, W., Socher, R., Li, L. J., Li, K., & Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In *IEEE Conference On Computer Vision And Pattern Recognition*, 248–255. IEEE. https://doi.org/10.1109/CVPR.2009.5206848

22. Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition. In *International Conference on Learning Representations*, 1–14.

23. Huang, G. B., Mattar, M., Berg, T., & Learned-Miller, E. (2008). Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*. 1–14.

Дата першого надходження рукопису до видання: 28.09.2025
Дата прийнятого до друку рукопису після рецензування: 23.10.2025
Дата публікації:

УДК 001.891:65.011.56 DOI

А. В. ЛЕГЕЗА
викладач кафедри програмного забезпечення
ДВНЗ «Ужгородський національний університет»
ORCID: 0000-0001-8416-8836

І. Ф. ПОВХАН
доктор технічних наук, професор,
професор кафедри програмного забезпечення систем
ДВНЗ «Ужгородський національний університет»
ORCID: 0000-0002-1681-3466

В. Я. САРОЗ
аспірант кафедри програмного забезпечення
ДВНЗ «Ужгородський національний університет»
ORCID: 0009-0003-1628-2778

В. О. ЯКОВЛЕВ
аспірант кафедри програмного забезпечення
ДВНЗ «Ужгородський національний університет»
ORCID: 0009-0000-0293-4614

МОДЕЛЬ ОЦІНКИ ЕФЕКТИВНОСТІ КОМПОНЕНТА ЕЛЕКТРОННОГО НАВЧАЛЬНОГО ПРОЦЕСУ НА ОСНОВІ ЕЛЕМЕНТАРНИХ ДИДАКТИЧНИХ ПОКАЗНИКІВ

В роботі показані особливості оцінки електронного навчального процесу в межах освітнього простору, представлена технологія для оцінки ефективності (загальної якості) онлайн-курсу (компоненту електронного навчального процесу) за допомогою набору виділених елементарних дидактичних показників. Сучасна дистанційна освіта характеризується стійкою тенденцією переходу до цифровізації та персоніфікації начального процесу. Модернізація освіти в даний час пов'язана з впровадженням адаптивних, практико-орієнтованих і гнучких інструментів оцінки якості електронного навчального процесу. Існуючі освітні інформаційні системи для створення та оцінки якості електронних навчальних курсів, поки не мають можливостей для інтелектуалізації процесу навчання, гнучкої адаптації змісту освіти під індивідуальні потреби учасників навчального процесу, не містять зручних інтелектуальних інструментів для підтримки розробників онлайн-курсів, автоматизації рутинних операцій зі структурування освітнього контенту. Представлена концепція оцінки компонентів навчального середовища, яка формуються за допомогою групи виділених елементарних дидактичних показників, заснованих на базових атрибутах електронного навчального процесу. Виділені групи якісних показників ефективності організації електронного навчального процесу, які забезпечують процес системного аналіз ефективності застосування технології, методів дистанційної освіти. Їх конкретний вибір, формування груп показників визначається цілями та завданнями комплексної оцінки електронного навчального процесу. На основі схеми виділення та формування елементарних дидактичних показників – розроблено метод та модель оцінки ефективності, (якості) застосування технології, методу або компоненту електронного навчального процесу в межах освітнього середовища. В роботі показано, що в найпростішому випадку загальна процедура оцінювання, експертизи компонентів електронного навчального процесу будується на математичній обробці оцінених експертами сукупності показників якості курсу (визначення важливості та інформативності показників, усередненні, елементарні дидактичні показники). Ефективні інтелектуальні алгоритми обробки даних лежать в основі розробки адаптивної технології онлайн-навчання впроваджені в навчальні процеси. Розроблені основні принципи синтезу системи оцінки якості онлайн-курсів.

Ключові слова: *онлайн-навчання, LMS – платформа, модель оцінки якості, дистанційне навчання, електронний навчальний процес.*

© Легеза А. В., Повхан І. Ф., Сароз В. Я., Яковлев В. О., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

A. V. LEGEZA
Lecturer at the Software Department
State Higher Educational Institution "Uzhgorod National University"
ORCID: 0000- 001-8416-8836

I. F. POVKHAN
Doctor of Technical Sciences, Professor,
Professor at the Department of Systems Software
State Higher Educational Institution "Uzhgorod National University"
ORCID: 0000-0002-1681-3466

V. YA. SAROZ
Postgraduate Student at the Software Department
State Higher Educational Institution "Uzhgorod National University"
ORCID: 0009-0003-1628-2778

V. O. YAKOVLEV
Postgraduate Student at the Software Department
State Higher Educational Institution "Uzhgorod National University"
ORCID: 0009-0000-0293-4614

MODEL FOR EVALUATING THE EFFECTIVENESS OF THE E-LEARNING PROCESS COMPONENT BASED ON ELEMENTARY DIDACTIC INDICATORS

The paper shows the features of evaluating the e-learning process within the educational space, presents a technology for evaluating the effectiveness (overall quality) of an online course (a component of the e-learning process) using a set of selected elementary didactic indicators. Modern distance education is characterized by a steady trend of transition to digitalization and personification of the educational process. Modernization of education is currently associated with the introduction of adaptive, practice-oriented and flexible tools for assessing the quality of the e-learning process. Existing educational information systems for creating and evaluating the quality of e-learning courses, do not yet have the ability to intellectualize the learning process, flexibly adapt the content of education to the individual needs of participants in the educational process, do not contain convenient intellectual tools to support developers of online courses, automate routine operations for structuring educational content. The concept of evaluating the components of the educational environment, which is formed using a group of selected elementary didactic indicators based on the basic attributes of the electronic educational process, is presented. Groups of qualitative indicators of the effectiveness of the organization of the electronic educational process are identified, which provide the process of systematic analysis of the effectiveness of the use of technology and methods of distance education. Their specific choice and formation of groups of indicators is determined by the goals and objectives of a comprehensive assessment of the e-learning process. Based on the scheme of selection and formation of elementary didactic indicators, a method and model for evaluating the effectiveness (quality) of the use of technology, method or component of the electronic educational process within the educational environment has been developed. The paper shows that in the simplest case, the general procedure for evaluating and examining the components of the electronic educational process is based on mathematical processing of the set of course quality indicators evaluated by experts (determining the importance and informative value of indicators, averaging, elementary didactic indicators). Efficient intelligent data processing algorithms underlie the development of adaptive online learning technologies implemented in educational processes. The basic principles of synthesizing a system for evaluating the quality of online courses are developed.

Key words: *online learning, LMS platform, Quality Assessment Model, distance learning, electronic learning process.*

Постановка проблеми

В межах системного підходу, при оцінці ефективності дистанційний навчальний процес (НП) варто розглядати не лише як цілісну монолітну систему, алей враховувати ієрархічність в структурі моделі освітнього процесу. Це означає, що порівняння ефективності моделей НП повинно здійснюватися на одному і тому ж самому ієрархічному рівні. Тобто, не можна порівнювати між собою інформаційно – навчальні та контролюючі компоненти технології дистанційної освіти (ДО) в межах НП. Необхідність врахування ієрархічності в побудові дистанційного НП при оцінці ефективності їх застосування обумовлена тим, що структура таких технологій тісно пов'язана з їх функціонуванням. У багаторівневих ієрархічних системах з ростом числа рівнів (підсистем складної системи) зростає складність їх структури (ієрархії), що безпосередньо впливає на ефективність самої системи (моделі організації дистанційного НП).

Отже застосовуючи системний підхід, можна зробити висновок, що для оцінки ефективності дистанційного НП, необхідно:

- Робити метод (схему) узагальненої оцінки ефективності електронного НП (технології ДО).
- Провести етап формування та аналізу початкових даних.
- Провести етап визначення та оцінки ефективності порівнюваних НП;
- Провести етап аналізу оцінок ефективні НП які порівнюються, сформулювати набір рекомендацій, які необхідні на етапі прийняття рішення.

Відмітимо, що в даній схемі оцінки ефективності електронного НП, на результуючий показник узагальненої оцінки, впливають не всі можливі атрибути та фактори НП, а лише ті, що впливають на виконання навчальних завдань (технології ДО, які полегшують або прискорюють виконання навчальних завдань). Зауважимо, що другорядні фактори, які впливають на НП, повинні бути виключені з моделі оцінки ефективності (або, як варіант – сама модель має бути розроблена таким чином, щоб її чутливість до другорядних факторів була низькою). Саме ускладнення моделі оцінки ефективності НП може призвести до того, що до неї потраплять другорядні фактори, які маскують (заміщують) суттєві атрибути електронного НП. Зрозуміло, що з іншого боку слід уникати випадку, коли за значного скорочення факторів, що враховуються в моделі оцінки якості НП, результати оцінок ефективності можуть виявитися малозмістовними.

Формулювання мети дослідження

Мета статті – дослідження та побудова, механізмів, методів та моделей оцінки компонентів електронного НП.

Аналіз останніх досліджень і публікацій

Динаміка зміни структури ДО та варіантів схем організації НП ще більш ускладнює та розмиває картину дослідження [3]. В роботах [1, 2] розглянута принципова проблема організації електронного НП в межах освітньої організації на базі відкритих LMS систем. В роботі [3] представлена загальна модель організації електронного НП в межах концепції відкритого доступу платформи Learme. Проблемам ефективності та адаптивності організації онлайн курсів в освітньому середовищі присвячені роботи [4, 5]. Головної невирішеною проблемою тут залишається невисока універсальність такого підходу. Важливою задачею яка стає на етапі проектування моделі онлайн навчання залишається проблема вибору відповідної до умов організації LMS платформи [6]. Частковим вирішенням даної задачі є розробка та проектування сучасних гібридних LMS платформ на базі штучного інтелекту. Дана проблематика тісно пов‘язана з задачею простої та ефективної обробки/збору/аналізу статистичних даних (data mining) в процесі онлайн навчання [7]. Етап вибору LMS платформи пов‘язаний з потребами освітньої організації, підтримкою системи відповідних освітніх стандартів [8–12]. Важливим напрямком розвитку сучасних платформ електронної освіти є застосування розподілених (cloud) технологій доступу до навчального контенту (НК) [13]. Принциповим недоліком такого підходу є високі вимоги щодо якості та розповсюдження інформаційної інфраструктури в межах освітньої організації.

Викладення основного матеріалу дослідження

Побудова набору критерії ефективності електронного НП. Відповідно до положень теорії ефективності та кваліметрії – для оцінки якості деякого об‘єкта n – векторним показником використовуються критерії, які належать трьом класам: P – критерії придатності; Q – критерії оптимальності; R – критерії переваги.

Далі, перейдемо до математичне формулювання цих критеріїв для системи з m об‘єктів, що володіють n властивостями (в межах задачі оцінки якості, ефективності електронного НП).

Введемо наступні позначення.

$X_{All} = \{x_{1,1}, x_{1,2}, \ldots, x_{i,j}\}$ – множина показників якості об‘єктів, де $x_{i,j}$ – показник j-ової характеристики i-ого об‘єкту.

$X_{Main} = \{x_{1,Main}, x_{2,Main}), \ldots, x_{i,Main}\}$ – множина допустимих значень показника ефективності i-ого об‘єкту.

Тоді критерій придатності можна формалізувати в наступному вигляді:

$$P : \bigcap\nolimits_{j=1}^{n} (x_{i,j} \in \{x_{i,Main}\}) = True. \quad (1)$$

Зауважимо, що в (1) $i = 1, 2, \ldots, m$. Тоді, зважаючи на вище сказане, всі об‘єкті, які виконують умову (1) можуть бути використані в подальшому, тобто володіють однаковою ефективністю (якістю).

$X_{Opt} = \{x_{1,Opt}, x_{2,Opt}, \ldots, x_{s,Opt}\}$ – множина оптимальних значень показників характеристик об‘єкта що оцінюються, тоді $\{x_{s,Opt}\}$ – оптимальне значення показника s-ової характеристики.

Тоді критерій оптимальності можна формалізувати в наступному вигляді:

$$Q : \bigcap\nolimits_{j=1}^{n} (x_{i,j} \in \{x_{i,Main}\}) \cap \bigcap\nolimits_{j=1}^{n} (x_{i,s} \in \{x_{s,Opt}\}) = True. \quad (2)$$

Зафіксуємо, що об‘єкт для якого виконується умова (2) будемо називати оптимальним по s-овій характеристиці.

Зважаючи на все вище сказане, на даному етапі можна формалізувати критерій переваги:

$$R : \bigcap\nolimits_{j=1}^{n} \bigcap\nolimits_{i=1}^{m} (x_{i,j} \in \{x_{i,Main}\}) \cap \bigcap\nolimits_{j=1}^{n} \bigcap\nolimits_{k \neq i} (x_{i,k} \geq x_{i,j}) = True. \quad (3)$$

Тут, $k = 1, 2, \ldots, n$. Проаналізувавши вираз (3), можна зробити висновок, що довільний об‘єкт оцінки для якого він виконується переважає всі інші об‘єкти вибірки.

Отже підсумовуючи таку формалізацію – критеріїв придатності, оптимальності та переваги (P, Q, R), можна запропонувати наступну схему оцінки ефективності методів електронного НП (компонентів ДО):

- Етап 1. Зібрати та обчислити всі показники характеристик – векторів значень ефективності електронного НП.
- Етап 2. Провести процедуру формування та визначення наборів вимог до ефективності електронного НП. Сформувати множини (X_{Opt}, X_{Main}) з показниками $\{x_{Opt}\}$, $\{x_{Main}\}$.
- Етап 3. Провести процедуру формування критеріїв оцінки якості електронного НП – (P, Q, R).
- Етап 4. Провести процедуру обчислення показників ефективності навчальних модулів (електронних курсів) в межах електронного НП.
- Етап 5. Провести фінальну оцінку отриманих значень ефективності навчальних модулів (компонентів ДО) на основі визначених критеріїв придатності, оптимальності та переваги (P, Q, R). На основі проведених оцінок зробити висновок відносно якісної складової відповідних елементів електронного НП.

Побудова набору елементарних дидактичних показників ефективності електронного НП. На наступному етапі дослідження визначимо групу базових (елементарних) дидактичних показників $D(d_1, d_2, d_3, \ldots)$ якості організації навчального заняття (навчального модуля) із застосуванням відповідного методу ДО (рис. 1). Для цього використаємо схему порівняльної оцінки важливості елементарних дидактичних показників якості організації електронного заняття в тій чи іншій схемі організації НП [3]. Порівняльна оцінка в даній схемі проводиться з використанням методу експертних оцінок. Дана процедура є першим етапом побудови узагальненого показника (дидактичного показника) ефективності електронного навчального процесу (компонентів електронного НП, навчальних занять). В зв'язку з тим, що кількісна оцінка експертами ступеня важливості окремих показників НП не є простою задачею, вона проведена шляхом упорядкування показників за ступенем важливості (метод простого ранжування елементарних ознак) [2].

Відмітимо, що порівняльна оцінки важливості базових дидактичних показників якості організації електронного заняття в тій чи іншій схемі ДО показала, що про якість організації заняття із застосуванням електронного НП можна судити за нижченаведеними елементарними дидактичним показниками (характеристиками ДО) – (Табл. 1).

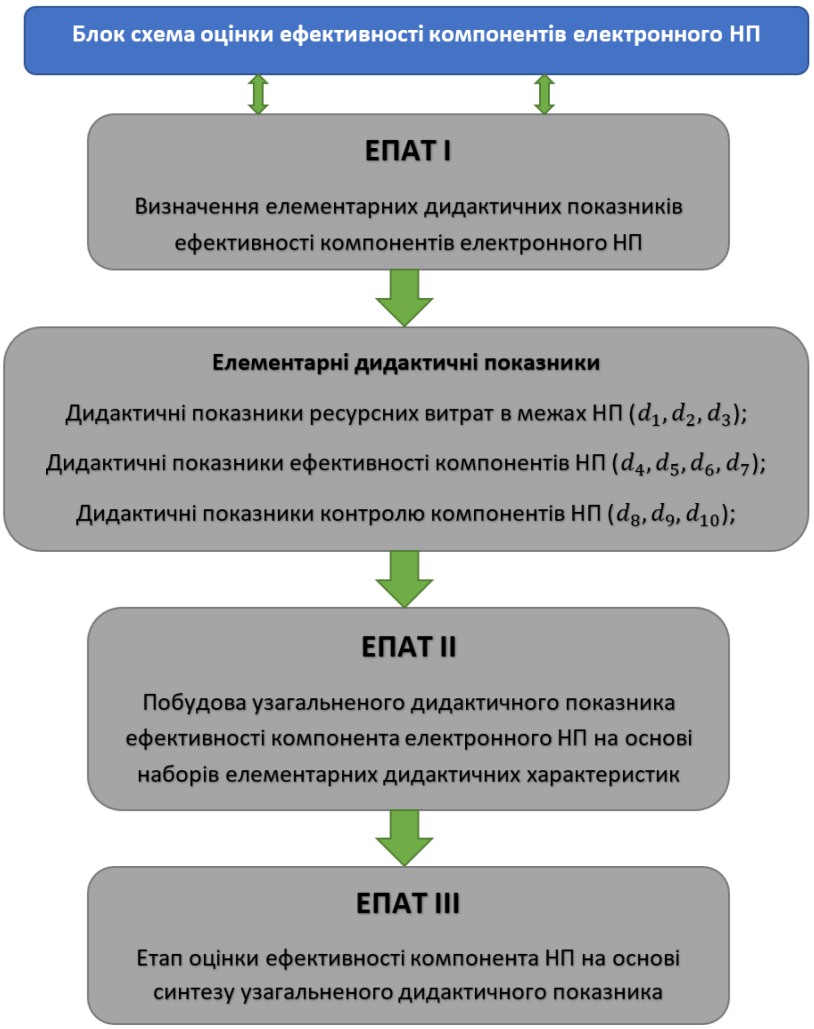


Рис. 1. Блок-схема визначення узагальненого дидактичного показника якості організації електронного заняття

Таблиця 1

Базова група якісних елементарних дидактичних показників компонента електронного НП.

№	Елементарні дидактичні показники компонента НП
d_1	Показник контентного (інформаційного) наповненню електронного заняття
d_2	Показник витрат часу на вивчення (засвоєння) навчального матеріалу (опрацювання НК)
d_3	Показник ресурсних витрат (фінансових) роботи в освітньому середовищі
d_4	Показник середнього балу (засвоєння НК) навчальної групи
d_5	Показник науковості електронного заняття (в межах організаційної схеми ДО)
d_6	Показник тематичної складності навчального контенту (в межах компоненту ДО)
d_7	Показник прикладної компоненти навчального контенту
d_8	Показник засвоєння навчального контенту на занятті (в межах компоненту ДО)
d_9	Показник навчальної активності учасників НП (в межах навчального заняття НП)
d_{10}	Показник навчальної зацікавленості результатами учасників електронного НП

Отже, визначення дидактичного обсягу навчального контенту включає в себе визначення показників інформаційного наповнення електронного НП та часових показників його засвоєння. Тоді, можна визначити трьох етапну схему визначення інформаційного наповнення електронного НП.

- Етап 1. На першому визначається обсяг навчальної інформації, яку необхідно довести до учасника НП за дане заняття – обсяг навчальної інформації, що міститься в темі заняття (або в межах електронного НП в цілому). Визначення обсягу досліджуваної навчальної інформації U_{All} починається з її подання в двійковій формі (бітах), що дозволяє створити об‘єктивну основу для визначення інформаційного навантаження.
- Етап 2. На другому етапі, необхідно визначити загальний обсяг інформації, який необхідно видати учаснику електронного НП за дане заняття (не тільки навчальний контент) I_{All}, який необхідно довести до учасника НП за заняття (на протязі визначеної навчальної сесії в освітньому середовищі).
- Етап 3. На третьому етапі, необхідно провести загальний розрахунок швидкості подачі навчальної інформації (пакетів навчального контенту) викладачем, або в ІС в межах освітнього середовища. Також необхідно підрахувати швидкість засвоєння навчального контенту учасником НП, розрахувати максимально можливий обсягу навчальної інформації, що подається в межах фіксованого електронного заняття (електронного курсу в цілому).

Узагальнений дидактичний показник якості НП. Отже, зважаючи на все вище сказане, узагальнений дидактичний показник якості НП можна представити як деяку залежність показників:

$$D^{Main} = \varphi(d_1, \ldots, d_{10}). \quad (4)$$

Зауважимо, що тут $(d_1, d_2, \ldots, d_{10})$ – елементарні дидактичні показники, які були представлені в даному дослідженні вище.

Беручи до уваги той факт, що виділені в дослідженні елементарні дидактичні показники мають різну природу формування та агрегування – доцільно для отримання комплексного, якісного показника застосовувати узагальнену функцію бажаності Харрінгтона вигляду:

$$D = \sqrt[n]{\prod_{i=1}^{n} s_i}. \quad (5)$$

Зауважимо, що тут:

D – узагальнений показник бажаності.

s_i – елементарні показники бажаності.

Відмітимо, що елементарні показники бажаності є безрозмірними величинами та величинами в діапазоні від 0 до 1, приведеними за шкалою бажаності Харрінгтона.

Розмірковуючи логічно, та підставивши (4) в вираз (5), отримаємо наступне рівняння вигляду:

$$D^{Main} = \sqrt[10]{\prod_{i=1}^{10} s(d_i)}. \quad (6)$$

Отже, для визначення якісних показників (d_i) тут $(i = 1, \ldots, 10)$ в дослідженні були розроблені ефективні схеми розрахунку. Після проведення розрахунків по кожній зі схем, показники нормуються по відношенню до показників «абстрактної, ідеальної», освітньої технології, приводяться до вигляду показників ефективності і тільки після цього розраховується узагальнений дидактичний показник – D^{Main} за формулою (6). В якості критерію оцінки – виступає шкала бажаності.

Висновки

Отже підсумуємо – оцінка узагальненого дидактичного показника якості електронного НП зводиться до агрегування елементарних дидактичних показників, яке виконується в два етапи:

- Етап 1. На першому етапі проводиться приведення елементарних дидактичних показників до безрозмірному виду.

• Етап 2. На другому етапі представлені елементарні дидактичні показники зводяться в узагальнений показник якості НП.

Приведення елементарних дидактичних показників до порівнянного вигляду здійснюється за єдиною безрозмірною шкалою бажаності (шкалою переваги). Вибір даної шкали обумовлений її здатністю здійснювати прямий переклад натуральних значень елементарних дидактичних показників в безрозмірні оцінки, при збереженні вихідних переваг.

Тобто, значення 0.0 шкали відповідає абсолютно неприйнятному рівню показника якості, а відповідно значення 1.0 – найкращому рівню показника якості. Аналітично шкала задається в наступному вигляді:

$$s = \exp(-\exp(-l)). \quad (7)$$

Зрозуміло, що принципово можливо побудувати графік функції $s(l)$ – на основі формули (7). Тобто, шкала бажаності відкладена по осі ординат – на ній фіксуються значення бажаності в межах відрізку від 0 до 1. Тоді, по осі абсцис вказуються значення елементарних дидактичних показників в умовному масштабі. При цьому за початок відліку на осі абсцис вибирається точка, відповідна нижньої межі допустимих значень на шкалі бажаності.

Зауважимо, що скільки значення осі абсцис є умовними, необхідно проградуювати її для кожного показника. Для цього зручно розташувати під основною віссю ще кілька (за кількістю елементарних дидактичних показників). Отже, отриманий графік з градуйованими осями дидактичних показників можна використовувати як номограму для здійснення приведення до єдиної шкали оцінки ефективності НП (його компоненту).

Отже можна зробити висновок, що побудований значення елементарних дидактичних показників D^{Main} і буде узагальненим показником якості електронного НП (компоненту електронного НП на основі набору елементарних дидактичних показників).

Список використаної літератури

1. Kintonova A., Sabitov A., Povkhan I., Khaimulina D., Gabdreshov G. Organization of online learning using the intelligent metasystem of open semantic technology for intelligent systems. Eastern-European Journal of Enterprise Technologies. 2023. № 1(2–121). P. 29–40.

2. Kintonova A., Sabitov A., Povkhan I. Development of an Online Course for “Web Programming” discipline with OLAT. 2022 IEEE 7th International Energy Conference (ENERGYCON): Conference Proceedings (Riga (Latvia), May 9–12, 2022). Riga, 2022. P. 54–62.

3. Kintonova A., Povkhan I., Sabitov A., Tokkuliyeva A., Demidchik N. Online Learning Technologies. 2022 IEEE 7th International Energy Conference (ENERGYCON): Conference Proceedings (Riga (Latvia), May 9–12, 2022). Riga, 2022. P. 76–84.

4. Alt D., Naamati-Schneider L. Health management students’ self-regulation and digital concept mapping in online learning environments. BMC Medical Education. 2021. Volume 21, Issue 1. Article Number 110.

5. Ferrer N. F., Alfonso J. M. E-Learning Content Management. New York (USA): Springer, 2010. 132 p.

6. How E-learning Platforms Are Gaining Popularity In The Ecommerce Industry. Punjab (India): FATbit Technologies, 2024. URL: https://www.fatbit.com/fab/online-learning-vs-traditionallearning-a-study-on-elearning-platform/

7. Pappas C. Top 20 eLearning Statistics For 2019 You Need To Know [Infographic]. Reno (NV, USA): eLearning Industry, 2019. URL: https://elearningindustry.com/top-elearning-statistics-2019

8. Siegel K. iSpring Suite 9: The Basics. Riva (USA): IconLogic, Inc., 2020. 119 p.

9. Caudill B., Banks D. Pocket instruction for SCORM. West Palm Beach (Florida, USA): JCASolutions, 2006. 114 p.

10. What is AICC. London (UK): Growth Engineering Technologies, 2022. URL: https://www.growthengineering.co.uk/what-is-aicc/

11. Revilova O., Santos O., Restrepo E. G. WCAG 2 in simple terms. Madrid (Spain): Itákora Press, 2015. 53 p.

12. Blokduk G. A clear and concise XAPI reference. Toronto (Canada): 5STARCooks, 2018. 137 p.

13. Masud Md., Huang X. An E-learning System Architecture based on Cloud Computing. World Academy of Science, Engineering and Technology. 2012. Vol. 62. P.74–78.

14. Chiasson A. Mastering Articulate Storyline. Birmingham (UK): Packt Publishing, 2016. 28 p.

15. Education at a Glance 2011. Paris (France): OECD Publishing, 2011. 320 p.

16. Ruiz-Corbella M., Alvarez-Gonzalez B. Virtual Mobility as an Inclusion Strategy in Higher Education: research on Distance Education Master degrees in Europe, Latin America and Asia. Research in Comparative and International Education. 2014. Vol. 9, Iss. 1. P. 165-180.

17. Wit H. de., Ferencz I., Rumbley L. E. International student mobility: European and US perspectives. Perspectives: Policy and Practice in Higher Education. 2013. Vol. 17, Iss. 1. P. 17-23.

Дата першого надходження рукопису до видання: 30.09.2025
Дата прийнятого до друку рукопису після рецензування: 27.10.2025
Дата публікації:

УДК 004.383 DOI

В. А. ЛУКАШЕНКО
кандидат технічних наук,
науковий співробітник відділу спеціалізованої високовольтної техніки
та лазерного зварювання
Інститут електрозварювання імені Є. О. Патона
Національної академії наук України
ORCID: 0000-0002-9685-4654

Д. А. ГАРДЕР
кандидат технічних наук,
науковий співробітник відділу спеціалізованої високовольтної техніки
та лазерного зварювання
Інститут електрозварювання імені Є. О. Патона
Національної академії наук України
ORCID: 0000-0002-4066-8182

А. В. БЕРНАЦЬКИЙ
кандидат технічних наук, старший дослідник,
завідувач відділу спеціалізованої високовольтної техніки
та лазерного зварювання
Інститут електрозварювання імені Є. О. Патона
Національної академії наук України
ORCID: 0000-0002-8050-5580

В. М. ЛУКАШЕНКО
доктор технічних наук, професор,
завідувач кафедри робототехніки та спеціалізованих комп'ютерних систем
Черкаський державний технологічний університет
ORCID: 0009-0006-7459-4704

Г. А. ЛУКАШЕНКО
здобувач вищої освіти
Черкаський державний технологічний університет
ORCID: 0009-0000-5560-9700

МОДЕЛЬ ІРРАЦІОНАЛЬНОЇ ФУНКЦІЇ СПЕЦІАЛЬНОГО ПРИЗНАЧЕННЯ РІЗНИХ ДВІЙКОВО-КОДОВИХ СИСТЕМАХ ЧИСЛЕННЯ В ЄДИНОМУ КРИСТАЛІ

Сучасний розвиток обчислювальної техніки характеризується постійним пошуком інноваційних рішень у сфері цифрового перетворення інформації. Метою даної роботи є створення ефективної моделі ірраціональної функції, шляхом перетворення в єдиному кристалі кортежні двійково-кодові формалізовані операнди з високою швидкодією і надійністю при малому енергоспоживанні. Методологія дослідження грунтується на сукупності прийомів, що перетворюють вхідні кортежні операнди різних двійково-кодових систем числення в вихідні значення комбінації за допомогою малого обсягу корегуючих констант (визначаються заздалегідь через операцію XOR), використовуючи одні й ті ж табличні дані відповідностей та синтезу відповідних логічних структур компонентів. Ключовою інновацією даної роботи є підвищення ефективності та зменшення активних компонентів і пасивних елементів за рахунок використання запропонованого оригінального схемотехнічого рішення при збереженні високої інформаційної точності. Верифікація перетворення вхідних кортежних двійково-кодованих операндів була підтверджена розрахунками для наведених операндів у двійковій системи числення при визначенні моделі ірраціональної функції спеціального призначення. Останні формують значення: двійкового коду функції «квадратний корінь»; двійково-десяткового коду функції «квадратний корінь» шляхом корегування вхідного масиву двійкового коду аргументів за допомогою відповідних малорозрядних корегуючих констант. Розрахунки, що були виконані за запропонованою формулою, через співвідношення середнього часу безвідмовної роботи запропонованої моделі формувача та моделей, підтверджують підвищення конструктивної надійності майже

© Лукашенко В. А., Гардер Д. А., Бернацький А. В., Лукашенко В. М., Лукашенко Г. А., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

2 рази за рахунок зменшення пасивних елементів. Запропонована технологія може бути успішно застосована в цифрових обчислювальних пристроях спеціального призначення, а саме: контрольно-вимірювальних системах; комп'ютерно-інтегрованих комплексах; тощо. Результати дослідження демонструють, що запропонований інноваційний підхід до архітектури компонентів може кардинально змінити уявлення вчених та дослідників про можливості цифрових перетворювачів спеціального призначення.

Ключові слова: *формалізована модель, математична логіка, двійковий код, операнд, значення функції, «квадратний корінь».*

V. A. LUKASHENKO
Candidate of Technical Sciences,
Researcher at the Department of the Specialized High-Voltage Engineering and Laser Welding
E. O. Paton Electric Welding Institute
of the National Academy of Sciences of Ukraine
ORCID: 0000-0002-9685-4654

D. A. HARDER
Candidate of Technical Sciences,
Researcher at the Department of the Specialized High-Voltage Engineering and Laser Welding
E. O. Paton Electric Welding Institute
of the National Academy of Sciences of Ukraine
ORCID: 0000-0002-4066-8182

V. BERNATSKYI
Candidate of Technical Sciences, Senior Researcher,
Head of the Department of Specialised High-Voltage Equipment and Laser Welding
E. O. Paton Electric Welding Institute
of the National Academy of Sciences of Ukraine
ORCID: 0000-0002-8050-5580

V. M. LUKASHENKO
Doctor of Technical Sciences, Professor,
Head of the Department of Robotics and Specialized Computer Systems
Cherkasy State Technological University
ORCID: 0009-0006-7459-4704

H. A. LUKASHENKO
Higher Education Student
Cherkasy State Technological University
ORCID: 0009-0000-5560-9700

MODEL OF A SPECIAL-PURPOSE IRRATIONAL FUNCTION IN VARIOUS BINARY-CODED NUMBERING SYSTEMS IN A SINGLE CRYSTAL

The modern development of computing is characterized by a constant search for innovative solutions in the field of digital information conversion. The purpose of this work is to create an effective model of an irrational function by converting in a single crystal tuple binary-code formalized operands with high speed and reliability at low power consumption. The research methodology is based on a set of techniques that convert input tuple operands of different binary-code number systems into output values of the combination using a small amount of correction constants (determined in advance through the XOR operation), using the same tabular data of correspondences and synthesis of the corresponding logical structures of components. The key innovation of this work is to increase the efficiency and reduce active components and passive elements by using the proposed original circuit-technical solution while maintaining high information accuracy. Verification of the transformation of the input tuple binary-coded operands was confirmed by calculations for the given operands in the binary number system when determining the model of the irrational function of special purpose. The latter form the values of: binary code of the function «square root»; binary-decimal code of the function "square root" by adjusting the input array of binary code of arguments using the corresponding small-bit adjusting constants. Calculations that were performed according to the proposed formula, through the ratio of the average time of failure-free operation of the proposed model of the shaper and models, confirm the increase in design reliability by almost 2 times due to the reduction of passive elements. The proposed technology can be successfully applied

in digital computing devices of special purpose, namely: control and measuring systems; computer-integrated complexes; etc. The results of the study demonstrate that the proposed innovative approach to component architecture can radically change the way scientists and researchers think about the capabilities of special-purpose digital converters.

Key words: *formalized model, mathematical logic, binary code, operand, function value, «square root».*

Постановка проблеми

На сьогодні проблемою формування значень ірраціональних функцій є або велика тривалість операції при використанні евристичного методу послідовного наближення, або великий об'єм постійного запам'ятовуючого пристрою при апаратурної реалізації табличним класичним методом (ТКМ) у єдиному кристалі [1-3], що зменшує кількість придатних кристалів з пластині і, як наслідок,збільшується вартість виробу. Однієї з ірраціональних функцій апаратурної реалізації для систем спеціального призначення (в промисловості, в оборонній галузі та ін.) є функція «корінь квадратний», тому для спрощення процедури формування і дослідження ірраціональної моделі пропонується на прикладу цієї функції. Апаратурна реалізація функції «корінь квадратний» с високими показниками одночасно такими як: швидкодія, надійність, габарити, вага, потужність споживання при низької вартості для систем спеціального призначення є проблематичною задачею.

Крім того, частіші в спеціалізованих системах керування динамічними процесами виникає потреба в реєстрації та швидкому аналізу значення вхідних, проміжних та кінцевих результатів добування кореня, тоді для зручної візуалізації її перетворюють у двійково-десятковий код. На теперішній час це вирішується додатково через комутаційну плату з активними компонентами та пасивними елементами, це значно погіршує конструктивну надійність, що недопустимо для систем спеціального призначення.

Аналіз останніх досліджень і публікацій

Питанням теорії та практики створення якісних комп'ютерних компонентів присвячено ряд робіт (у тому числі і формувачів проблемних функцій) [4–6]. У цих працях описано багато інтересних моделей, що мають новітні схемотехнічні рішення, в яких стосуються формування багатофункціональних компонентів. Проте в них не в повній мірі освітлено процедури створення високоефективних ірраціональних функцій, які мають можливість представити результат вихідної інформації в різних двійково-кодових системах при схемотехнічному рішенні в єдиному кристалі та на базі єдиного числового блоку пам'яті (ЧБП).

У роботі [1, 2] значення функції «корінь квадратний» отримають шляхом використання ітераційного методу або на базі евристичного методу послідовного наближення. Недоліком цих методів є дуже тривалий час отримання значення функції добування коріння.

Високу швидкість забезпечує табличний класичний метод, що аналізується в роботі [3], але для високої точності відтворення результату необхідно мати великий обсяг ПЗП. Це значно погіршує показники по енергоспоживанню та надійності. У статі [4] запропоновано метод визначення надмірності коригуючих констант при відтворенні прецизійних значень багатофункціональних трансцендентних функцій. Це сприяє зменшенню таблиць, які зберігаються в числовому блоці пам'яті (ЧБП) і зменшенню активних елементів, але процедура розглядається тільки для однієї двійкової системи числення в кристалі. Відомо, що кожен зовнішній контакт вузла з'єднання (рис. 1), як правило, складається з системи фізичних об'єктів такого виду: термокомпресія – провідник – мікрозварювання – вивід – корпусу – зовнішнє мікрозварювання (мікропайка), з мінімальними розмірами (60×60 мкм), має велику інтенсивність відмови $\lambda_i = 10^{-5}$. Тому зовнішні контактні вузли є одними з найбільш ненадійних елементів системи з'єднання.

Крім того, аналіз топології контактних вузлів мікросхеми на кристалу (рис. 1) показав, що для їх з'єднання з відповідними виводами до корпусу займають багато площі кристалу.

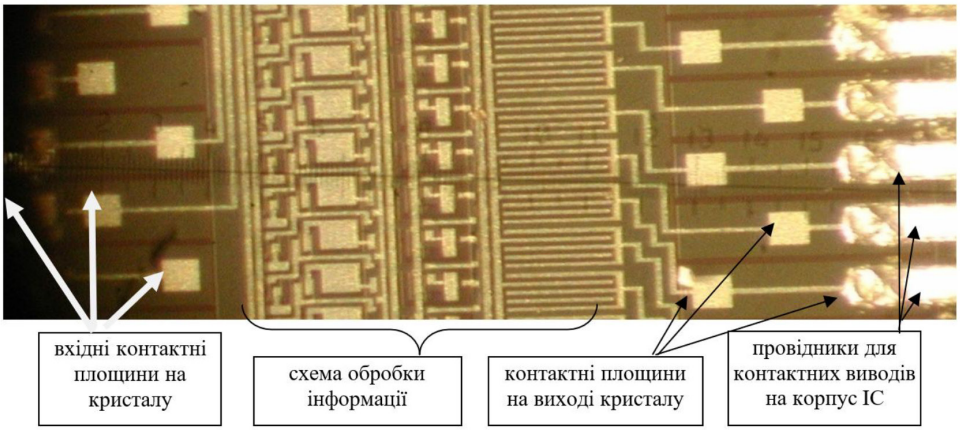


Рис. 1. Топологія контактних вузлів мікросхеми на кристалі

В роботах [5, 6] запропоновані моделі функції добування квадратного кореня на базі таблично-логічного методу. Результати обчислювання представлено в двійковому коді та в двійково-десятковому коді. Недоліком цих моделей при високої точності є наявність великої кількості зовнішніх і внутрішніх контактів, що зменшує конструктивну надійність.

Метод зменшення кількості зовнішніх контактних вузлів за рахунок введення в середині кристалу додаткового блоку вентилів запропонованих в роботах [7–9], що сприяє підвищенню конструктивної надійності. Але не розглянуто як впливає на значення показнику енергоспоживання при додатковому імплементуванні в межах площі єдиного кристалу.

Тому, синтез логічних структур в єдиному кристалі моделі функції «корінь квадратний» спеціального призначення в двох двійково-кодових системах на методологічному й інформаційному базисі є першочерговим завданням.

При цьому необхідно враховувати, що особливість компонентів спеціального призначення полягає в обмеженні множини вхідних та відповідних вихідних кодових послідовностей. Це дозволяє заздалегідь складати таблиці відповідних вхідних і вихідних двійкових кодів. При цьому в роботі [10] показана процедура формування оптимальної розрядності кортежів при підготовки таблиць кортежних відповідностей, на яку використовується багато часу.

Формулювання мети дослідження

Метою підвищення ефективності моделі, що відтворює значення ірраціональної функції спеціального призначення в двох двійково-кодових системах за рахунок схемотехнічного рішення в єдиному кристалі через зменшення апаратурних витрат, що базується на принципах: коригування вхідної двійково-кодової інформації за допомогою малорозрядних кортежів при формуванні таблиць відповідностей малого обсягу та багаторазового використання активних компонентів та пасивних елементів при збереженні високої інформаційної точності.

Для досягнення поставленої мети вирішено наступні завдання:

1. Провести аналіз існуючих методів апаратної реалізації значень ірраціональних функцій та визначити базовий метод на основі таблиць для побудови ефективної моделі.
2. Створити та дослідити модель ірраціональної функції на прикладу функції добування квадратного кореня спеціального призначення в двох двійково-кодових системах.
3. Провести верифікацію запропонованої моделі за параметрами: часу напрацювання на відмову, апаратурних витрат і потужності споживання в порівнянні з моделями, що реалізовані ТКМ.

Викладення основного матеріалу дослідження

Особливість компонентів спеціального призначення полягає в обмеженні множини вхідних та відповідних вихідних кодових послідовностей, тому заздалегідь складаються таблиці відповідних вхідних і вихідних двійкових кодів.

Апаратурно моделі ірраціональних функцій частіше реалізується одним із традиційних таблично-алгоритмічних методів (ТАМ), перевагою яких є значне зменшення обсягу ПЗП. Традиційні ТАМ таки як: табличний адитивно – мультиплікативний та таблично-адитивний методи характеризуються малим обсягом ПЗП, але мають повільну швидкодією через наявність тривалих арифметичних операцій:множення та додавання.

Існуючи на сьогодні нетрадиційні методи: таблично-логічний, напівадитивний таблично-логічний, багатофункціональний таблично-логічний, таблично-логічний оборотний, формалізований таблично-логічний та розрядний кортежно-табличний логіко-оборотний (РКТЛО) метод, мають високу швидкодією завдяки використанню тільки швидкодіючі логічні операцій. Серед них виділяється метод РКТЛО [10], перевагою якого велика можливість маневрувати між апаратурними витратами та швидкістю отримання результату при збережені високої точності перетворення інформації.

Особливість РКТЛО методу полягає в тому, що позиційно-впорядковані операнди кодових послідовностей представляються як композиція незалежних «1» і «0». Це дозволяє представляти їх у вигляду впорядкованих кортежів, розрядність яких визначається за допомогою алгоритму машинного формування [10].

У той же час, аналіз поставленого завдання показав, що РКТЛО метод, можна модернізувати через усунення в його алгоритму операції визначення оптимального значення розрядності в кортежах через те, що при формуванні значень ірраціональної функції в двійково-десяткової системі числення, операнди останнього є тетради. Це значно зменшує підготовчий час на створення таблиць відповідностей.

Пропонується створення моделі ірраціональної функції на базі табличного обчислювання, властивостей математичної логіки, формалізації, принципу багатократного використання одних і тих ж таблиць, що реалізуються апаратно в числовому блоку пам’яті (ЧБП).

Особливістю таблиць ЧБП полягає в тому, що записується різниця багаторозрядних вхідних і вихідних кодів, яка визначається через операцію XOR. Цим зменшує обсяг пам’яті майже в 2 рази. В той же час при порівнянні тетрад вхідних і вихідних кодів, що представлено кортежними композиціями з незалежними «1» і «0» забезпечується можливість отримати велике число однакових значень кодів.

Послідовність організації таблиць, що записується в ЧБП, здійснюється таким чином: створюються таблиці відповідностей у вигляду кортежів; при необхідності вирівнюються по довжині тетради кодові операнди через

додавання «0» перед старшим розрядом у старшій кортеж для цілих чисел і додавання «0» після молодшого розряду в молодший кортеж; визначаються різниці відповідних кортежів за допомогою операції XOR; формується таблиця значень корегуючих констант після ліквідації інформаційні надмірності, тобто зберігається в ПЗП з множини однакових тільки одно значення корегуючих констант.

Верифікація процедури створення моделі високоефективної ірраціональної функції проведена на прикладу значень функції «корінь квадратний» $(\sqrt{n})_2$ та аргументів $(n_i)_2$ і корегуючої константи $(\Delta)_2$ у двійкової системі числення за формулою (1) представлено у кортежному вигляду та наведено в таблиці 1.

$$\Delta_{ст} = Y_{2\,ст} \oplus Y_{2\text{-}10\,ст}; \quad \Delta_{млд} = Y_{2\,млд}, \oplus Y_{2\text{-}10\,млд}. \tag{1}$$

Таблиця 1

Реляційна модель даних чотирьох розрядними кортежами аргументу, значень функції та корегуючи константи

№ п/п	Кортежне представлення операндів у двійковій системи числення		
	Код аргументу $(n_i)_2$	Код функції, $(\sqrt{n})_2$	Код корегуючої константи, $(\Delta)_2$
1	0100 0000 0000	0000 0010 0000	0100 0010 0000
2	0100 0100 0001	0000 0010 0001	0100 0110 0000
3	0100 1000 0100	0000 0010 0010	0100 1010 0110
4	0100 1100 1001	0000 0010 0011	0100 1110 1010
5	0101 0001 0000	0000 0010 0100	0101 0011 0100
6	0101 0101 1001	0000 0010 0101	0101 0111 1100
7	0101 1010 0100	0000 0010 0110	0101 1000 0010
8	0101 1111 0001	0000 0010 0111	0101 1101 0110
9	0110 0100 0000	0000 0010 1000	0110 0110 1000
10	0110 1001 0001	0000 0010 1001	0110 1011 1000
11	0110 1110 0100	0000 0010 1010	0110 1100 1110
12	0111 0011 1001	0000 0010 1011	0111 0001 0010
13	0111 1001 0000	0000 0010 1100	0111 1011 1100
14	0111 1110 1001	0000 0010 1101	0111 1100 0100
15	1000 0100 0100	0000 0010 1110	1000 0110 1010

Коли довжина операндів старшої тетради двійково-десяткового коду функції більш розрядності двійкового коду функції, тоді в старший кортеж двійкового коду дописуються розряди нульового значення «0», тобто проводиться операція вирівнювання розрядів в старшої тетради (приклади наведено в таблиці 1 і таблиці 2).

Враховуючи (1), (логіко-математична модель корегуючих констант визначається за формулою:

$$\Delta = \Delta_{ст}\,\Delta_{млд} = (Y_{2\,ст} \oplus Y_{2\text{-}10\,ст})(Y_{2\,млд} \oplus Y_{2\text{-}10\,млд}). \tag{2}$$

Аналіз результатів кортежної декомпозиції кодів у таблиці 1 і таблиці 2, показав наявність корегуючих констант однакового значення дуже багато. На рис. 2 і рис. 3 побудовано гістограми, кількісного оцінювання однакових значень корегуючих констант через показник h це забезпечує ліквідацію інформаційної надмірності, зменшує кількість активних елементів.

Аналіз гістограм (рис. 2 і рис. 3) показав, що для наведеного прикладу в сукупності кількість корегуючих констант зменшується і, як наслідок, зменшується кількість активних елементів у ЧБП, збільшується конструктивна надійність. Причому збільшується вільна площа на кристалі, яка використовується для введення додатково всередині кристалу інформаційного зворотного зв'язку (ІЗЗ) і блок вентилів.

Останній забезпечує можливість використовування багаторазово вхідних контактів зовнішніх та внутрішніх для передачі вихідної інформації і, тім самим, на чотири, п'ять порядків збільшується надійність моделі, образно-знакова модель якого представлена на рис. 4.

Аналіз моделі ірраціональної функції «корінь квадратний» (рис. 4) підтвердив відсутність арифметичні операції, тобто формування значень функції в двох двійково-кодових системах здійснюється тільки логічними операціями. Цим забезпечується висока швидкість відтворення значень ірраціональної моделі.

Модель функції «корінь квадратний», що представлена в двійково-кодових системах числення працює в таких режимах:

1) у режимі перетворення двійкового коду значень аргументу $(n_i)_2$ у код значення функції «корінь квадратний» $(\sqrt{n})_2$ двійкової системі числення;

2) у режимі перетворення значення функції «корінь квадратний» із двійкової системи числення в двійково-десяткову, тобто код $(\sqrt{n})_2$ перетворюється в код $(\sqrt{n})_{2-10}$.

Таблиця 2

Реляційна модель даних значень функції «корінь квадратний» в різних системах числення та корегуючи константи

№ п/п	Y_2		$Y_{2\text{-}10}$		$\Delta_{ст}$	$\Delta_{млд}$
	$Y_{2\ ст}$	$Y_{2\ млд}$	$Y_{2\text{-}10\ ст}$	$Y_{2\text{-}10\ млд}$		
1	0010 0000		0011 0010		0001 0010	
2	0010 0001		0011 0011		0001 0010	
3	0010 0010		0011 0100		0001 0110	
4	0010 0011		0011 0101		0001 0110	
5	0010 0100		0011 0110		0001 0010	
6	0010 0101		0011 0111		0001 0010	
7	0010 0110		0011 1000		0001 1110	
8	0010 0111		0011 1001		0001 1110	
9	0010 1000		0100 0000		0110 1000	
10	0010 1001		0100 0001		0110 1000	
11	0010 1010		0100 0010		0110 1000	
12	0010 1011		0100 0011		0110 1000	
13	0010 1100		0100 0100		0110 1000	
14	0010 1101		0100 0101		0110 1000	
15	0010 1110		0100 0110		0110 1000	

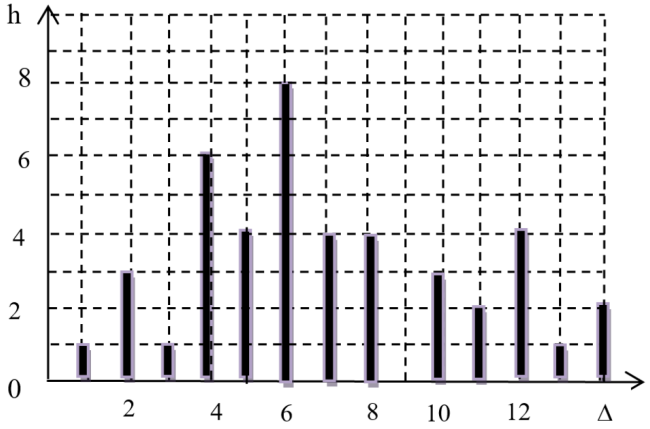


Рис. 2. Гістограма кількості *k* значень корегуючих констант при перетворенні двійкового коду аргументу $(n_i)_2$ в двійковий код функції $(\sqrt{n})_2$

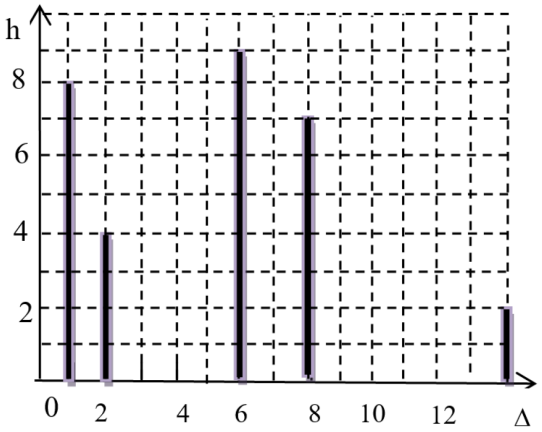


Рис. 3. Гістограма кількості *k* значень корегуючих констант при перетворенні двійкового коду значень функції $(\sqrt{n})_2$

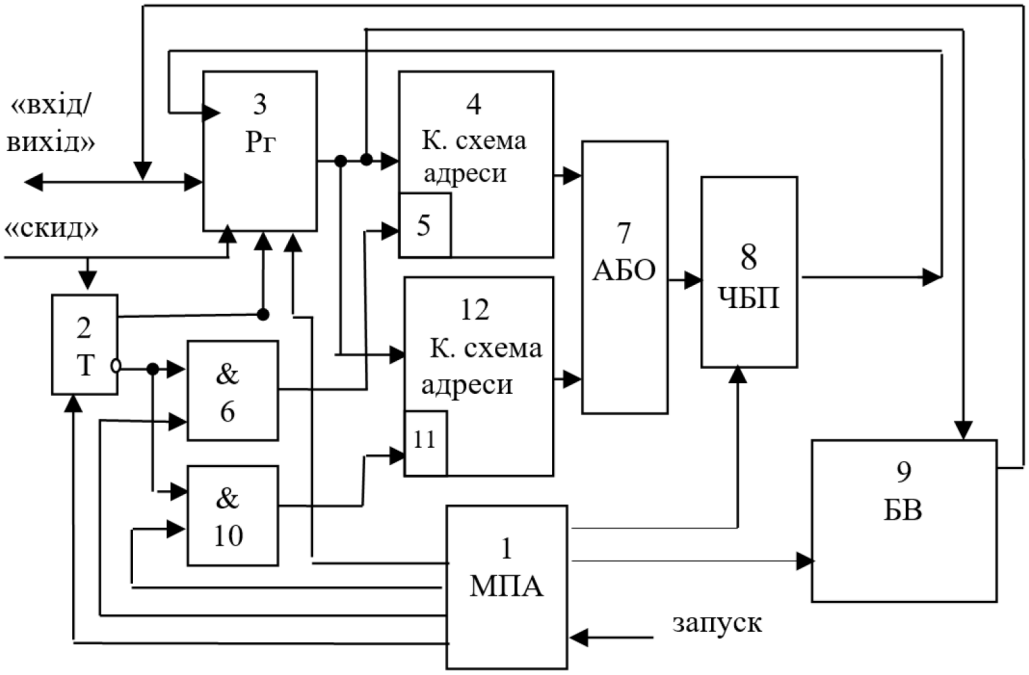


Рис. 4. Модель високоефективного формувача функції «корінь квадратний» спеціального призначення в двох двійково-кодових системах, де: 1 МПА – мікропрограмний автомат; 2 Т – тригер; 3 Рг – регістр; 4, 12 – комбінаційна схема адреси; 5, 11– «дозвіл» роботи компонентів 4, 12 відповідно; 6, 10 – елементи «І»; 7 – елемент АБО; 8 ЧБП – числовий блок пам'яті; 9 БВ – блок вентилів

Робота моделі (рис. 4) в часі здійснюється зовнішнім МПА. Формувач працює в першому режимі таким чином. Імпульсом «скид» тригер Т2 і регістр Рг3 обнулюються. Після імпульсу «запуск» на вхід МПА1 з його виходу надходить відповідний керуючий сигнал і під дією імпульсу з виходу Q тригеру 2 вхідна інформація записується в Рг 3. Комбінаційна схема адреси 4 під дією імпульсу «дозвіл» 5, що надходить з виходу елементу «І»6, розпізнає адресу корегуючої константи. Імпульс, що проходить через блок елементів АБО 7 збуджує відповідну шину ЧБП8, що зберігає відповідний кортежний код корегуючої константи. Код константи через АБО 9 по корегуючому зворотному зв'язку (КЗЗ) надходить на лічильні регістру Рг 3 і під дією «1» його тригери змінюють інформацію на протилежну. Таким чином, код аргументу $n_{вх}$ перетворюється в код значення функції «корінь квадратний» $(\sqrt{n})_2$ і зберігається в Рг 3. Далі цій код з виходу Рг 3 по інформаційному зворотному зв'язку (ІЗЗ) надходить на входи блок вентилів БВ 9, виходи якого в середині кристалу з'єднані з інформаційними входами моделі формувача. Відповідний імпульс з виходу МПА1 відкриває блок вентилів БВ 9 і код значення функції з'являється на зовнішніх вхідних контактах.

Робота формувача в режимі перетворення двійкового коду значення функції «корінь квадратний» $(\sqrt{n})_2$ у код $(\sqrt{n})_{2-10}$ двійково-десятковій системи числення здійснюється після того, як у Рг 3 сформовано $(\sqrt{n})_2$. Далі під дією відповідного у часі керуючого імпульсу МПА1, якій проходить через елемент «І»10 відкривається «дозвіл» 11 на розпізнавання адреси за допомогою комбінаційної схеми 12. Імпульс, що проходить через блок відповідних елементів АБО 7 збуджує шину відповідного ЧБП 8, у якому зберігається код корегуючої константи. Код останній по КЗЗ надходить на лічильні регістру Рг 3 і під дією «1» його тригери змінюють інформацію на протилежну. Таким чином, на виході Рг 3 з'являється код $(\sqrt{n})_{2-10}$ значення функції в двійково-десятковій системи числення. Далі процедура проходження коду $(\sqrt{n})_{2-10}$ на вихід створеної моделі аналогічна першому режиму.

Особливістю моделі (рис. 4) є можливість записати в Рг 3 значення функції «корінь квадратний» двійкової системи числення, мінував процедуру перетворення коду аргументу в код функції $(\sqrt{n})_2$ завдяки розташовування МПА за межами кристалу. Це розширює функціональні можливості моделі формувача.

Верифікація запропонованої моделі ірраціональної функції за параметрами: часу напрацювання на відмову, апаратурних витрат і потужності споживання в порівнянні з моделлю, що реалізована ТКМ показано на прикладі моделі формувача функції «корінь квадратний».

Відомо, що модель функції, що реалізовані ТАМ містить два n-розрядних регістра, дешифратор з адресами 2^n на виході, шифратор, обсяг пам'яті якого визначається у бітах, як

$$V = n \cdot (2^n - 1). \tag{3}$$

Конструктивна надійність моделі ірраціональної функції до відмови, що розраховується за формулою

$$T_0 = \frac{1}{\sum_i^n \lambda_i}, \tag{4}$$

де λ_i – інтенсивність відмовної роботи одного елементу.

У той же час, розміри внутрішнього контакту вузла з'єднання має майже максимальний розмір (5 × 5 мкм) з $\lambda_i = 10^{-10}$, а контакти вузлів зовнішнього з'єднання має майже мінімальний розмір (60 × 60 мкм) при $\lambda_i = 10^{-5}$.

Верифікація значення часу напрацювання на відмову формувача через використання двічі зовнішніх вхідних контактних вузлів з'єднання підтверджена результатом розрахунку за формулою

$$T_{в3ф} / \ T_{вПр} = \left(1 \Big/ \sum_i^n l_{i3ф}\right) \Big/ \left(1 \Big/ \sum_i^n l_{iПр}\right), \tag{5}$$

де $T_{в3Ф}$ – час напрацювання запропонованого формувача; $T_{вПр}$ – час напрацювання моделі, що представлена в роботі [6] і визначена як прототип; $\lambda_{i\ 3Ф}$ – інтенсивність відмовної роботи запропонованого формувача $\lambda_ф = 10^{-10}$; $\lambda_{i\ Пр}$ – інтенсивність відмовної роботи прототипу [6] $\lambda_ф = 10^{-5}$.

Нехай $k_{пр}$ – кількість зовнішніх вхідних і вихідних контактних вузлів прототипу [6] при n = 64, тоді $k_{пр} = (4 \cdot n) = 256$, а $k_{зф}$ – кількість зовнішніх контактів формувача при n = 64, тоді $k_{пр} = 2 \cdot n = 128$, звідки

$$T_{оЗП} / \ T_{оПр} = 2 \text{ рази}. \tag{6}$$

Потужність споживання $W_{спр}$ за кількістю укрупнених основних вузлів, блоків і елементів в них, з відповідними значеннями потужності одиничного елемента або компоненту, представляється у вигляді

$$W_{спр} = \sum_{i=1}^{N} W_i \cdot c_i = W_1 \cdot c_1 + W_2 \cdot c_2 + W_i \cdot c_i + \ldots + W_N \cdot c_N, \tag{7}$$

де W_1 – значення потужності споживання одним тригером; c_i – число однорозрядних елементів у відповідних блоках; N – загальне число відповідних блоків.

Припустимо, що використовуються операнди розрядністю n = 64, потужність одного логічного елемента дорівнює $W_1 = 10 \cdot 10^{-6}$ Вт. Тоді потужність споживання блока вентилів становить

$$W_B = n \cdot W_1 = 64 \cdot 10 \cdot 10^{-6} = 0{,}00064 \text{ Вт} = 0{,}64 \text{ мВт} \tag{8}$$

Відомо [10], що потужність одного тригера регістру в чотири рази більше потужності одного логічного елемента, тоді потужність одного 64-розрядного регістру визначається за такою формулою:

$$W_2 = n \cdot 4 \cdot W_1 = 64 \cdot 4 \cdot 10 \cdot 10^{-6} = 0{,}00256 \text{ Вт} = 2{,}56 \text{ мВт}. \tag{9}$$

Величина економії потужності W_2 знизиться через введення блока вентилів завдяки ланцюга ІЗЗ до величини ΔW, значення якої визначається нижче:

$$\Delta W = W_2 - W_B \approx 0{,}00256 - 0{,}00064 = 0{,}00192 \text{ Вт} = 1{,}92 \text{ мВт}. \tag{10}$$

Отже, розрахунки за формулою (10) через співвідношення середнього часу безвідмовної роботи запропонованої моделі формувача та моделей, що реалізовані в роботі [6], підтверджують підвищення конструктивної надійності майже 2 рази за рахунок зменшення пасивних елементів.

Висновки

Ключовою інновацією даної роботи є підвищення ефективності та зменшення активних компонентів і пасивних елементів за рахунок використання запропонованого оригінального схемотехнічого рішення при збереженні високої інформаційної точності. Верифікація перетворення вхідних кортежних двійково-кодованих операндів була підтверджена розрахунками для наведених операндів у двійковій системи числення при визначенні моделі ірраціональної функції спеціального призначення.

Аналіз моделі ірраціональної функції «корінь квадратний» підтвердив відсутність арифметичні операції, тобто формування значень функції в двох двійково-кодових системах здійснюється тільки логічними операціями. Цим забезпечується висока швидкість відтворення значень ірраціональної моделі.

Розрахунки за запропонованою формулою, через співвідношення середнього часу безвідмовної роботи запропонованої моделі формувача та моделей, підтверджують підвищення конструктивної надійності майже 2 рази за рахунок зменшення пасивних елементів.

Список використаної літератури

1. Shwartz-Ziv, R., Armon, A. Tabular data: Deep learning is not all you need. *Information Fusion*. 2022. Vol. 81. P. 84–90. URL: https://doi.org/10.1016/j.inffus.2021.11.011

2. Kahraman, C., Onar, S. C., Oztaysi, B. Fuzzy multicriteria decision-making: a literature review. *International Journal of Computational Intelligence Systems*. 2015. Vol. 8. Iss. 4. P. 637-666. URL: https://doi.org/10.1080/18756891.2015.1046325

3. Aboul-Maaty, N. A. F., Oraby, H. A. S. Extraction of high-quality genomic DNA from different plant orders applying a modified CTAB-based method. *Bulletin of the National Research Centre*. 2019. Vol. 43. Iss. 1. P. 1–10. URL: https://doi.org/10.1186/s42269-019-0066-1

4. Jin, L., Tan, F., Jiang, S. Generative adversarial network technologies and applications in computer vision. *Computational Intelligence and Neuroscience*. 2020. Vol. 2020. Iss. 1. Id. 1459107. URL: https://doi.org/10.1155/2020/1459107

5. Alcin, M., Koyuncu, I., Tuna, M., Varan, M., Pehlivan, I. A novel high speed artificial neural network–based chaotic true random number generator on field programmable gate array. *International Journal of Circuit Theory and Applications*. 2019. Vol. 47. Iss. 3. P. 365–378. URL: https://doi.org/10.1002/cta.2581

6. Wang, Z., She, Q., Ward, T. E. Generative adversarial networks in computer vision: A survey and taxonomy. *ACM Computing Surveys (CSUR)*. 2021. Vol. 54. Iss. 2. P. 1–38. URL: https://doi.org/10.1145/3439723

7. Nandi, A. K., Medal, H. R. Methods for removing links in a network to minimize the spread of infections. *Computers & Operations Research*. 2016. Vol. 69. P. 10–24. URL: https://doi.org/10.1016/j.cor.2015.11.001

8. Sun, C., Liu, G. R., Huo, S. H., Wang, G., Yu, C., Li, Z. A novel node-to-segment algorithm in smoothed finite element method for contact problems. *Computational Mechanics*. 2023. Vol. 72. Iss. 5. P. 1029–1057. URL: https://doi.org/10.1007/s00466-023-02327-6

9. Sundararaj, V., Muthukumar, S., Kumar, R. S. An optimal cluster formation based energy efficient dynamic scheduling hybrid MAC protocol for heavy traffic load in wireless sensor networks. *Computers & Security*. 2018. Vol. 77. P. 277–288. URL: https://doi.org/10.1016/j.cose.2018.04.009

10. Lukashenko, A., Harder, D., Lukashenko, V. M., Fedorov, E., Lukashenko, V. A., Utkina, T., Mitsenko, S., Rudakov, K. Bitwise method for the binary-coded operands conversion based on mathematical logic. *Eastern-European Journal of Enterprise Technologies*. (2018). Vol. 5. Iss. 4(95). P. 6–14. URL: https://doi.org/10.15587/1729-4061.2018.142975

References

1. Shwartz-Ziv, R., & Armon, A. (2022). Tabular data: Deep learning is not all you need. *Information Fusion, 81*, 84–90. https://doi.org/10.1016/j.inffus.2021.11.011

2. Kahraman, C., Onar, S. C., & Oztaysi, B. (2015). Fuzzy multicriteria decision-making: a literature review. *International Journal of Computational Intelligence Systems, 8*(4), 637-666. https://doi.org/10.1080/18756891.2015.1046325

3. Aboul-Maaty, N. A. F., & Oraby, H. A. S. (2019). Extraction of high-quality genomic DNA from different plant orders applying a modified CTAB-based method. *Bulletin of the National Research Centre, 43*(1), 1–10. https://doi.org/10.1186/s42269-019-0066-1

4. Jin, L., Tan, F., & Jiang, S. (2020). Generative adversarial network technologies and applications in computer vision. *Computational Intelligence and Neuroscience, 2020*(1), 1459107. https://doi.org/10.1155/2020/1459107

5. Alcin, M., Koyuncu, I., Tuna, M., Varan, M., & Pehlivan, I. (2019). A novel high speed artificial neural network–based chaotic true random number generator on field programmable gate array. *International Journal of Circuit Theory and Applications, 47*(3), 365–378. https://doi.org/10.1002/cta.2581

6. Wang, Z., She, Q., & Ward, T. E. (2021). Generative adversarial networks in computer vision: A survey and taxonomy. *ACM Computing Surveys (CSUR), 54*(2), 1–38. https://doi.org/10.1145/3439723

7. Nandi, A. K., & Medal, H. R. (2016). Methods for removing links in a network to minimize the spread of infections. *Computers & Operations Research, 69*, 10–24. https://doi.org/10.1016/j.cor.2015.11.001

8. Sun, C., Liu, G. R., Huo, S. H., Wang, G., Yu, C., & Li, Z. (2023). A novel node-to-segment algorithm in smoothed finite element method for contact problems. *Computational Mechanics, 72*(5), 1029-1057. https://doi.org/10.1007/s00466-023-02327-6

9. Sundararaj, V., Muthukumar, S., & Kumar, R. S. (2018). An optimal cluster formation based energy efficient dynamic scheduling hybrid MAC protocol for heavy traffic load in wireless sensor networks. *Computers & Security, 77*, 277–288. https://doi.org/10.1016/j.cose.2018.04.009

10. Lukashenko, A., Harder, D., Lukashenko, V. M., Fedorov, E., Lukashenko, V. A., Utkina, T., Mitsenko, S., & Rudakov, K. (2018). Bitwise method for the binary-coded operands conversion based on mathematical logic. *Eastern-European Journal of Enterprise Technologies, 5*(4(95), 6–14. https://doi.org/10.15587/1729-4061.2018.142975

Дата першого надходження рукопису до видання: 21.09.2025
Дата прийнятого до друку рукопису після рецензування: 16.10.2025
Дата публікації:

UDC 667.021.1 DOI

O. O. MAMCHYCH
Postgraduate Student at the Department of Electronic Computers
Kharkiv National University of Radioelectronics
ORCID: 0009-0001-6602-2929

M. O. VOLK
Doctor of Technical Science,
Professor at the Department of Electronic Computers
Kharkiv National University of Radioelectronics
ORCID: 0000-0003-4229-9904

MODEL AND METHODOLOGY OF MEASURING AND FORECASTING OF ENERGY IMPACT OF SECURITY RISKS MITIGATION APPROACHES IN DISTRIBUTED COMPUTING SYSTEMS ON PERSONAL MOBILE DEVICES

The rapid growth of cloud computing and artificial intelligence has drastically increased energy consumption in data centers. Volunteer computing using personal mobile devices represents a promising solution by potentially reducing energy demands. However, such systems face critical challenges related to reliability, security, and sustainability. The ***subject of this study*** *is the mitigation of three principal risks –* ***node outage, result cheating, and data fishing*** *– within mobile volunteer computing environments, and the evaluation of their* ***energy overhead****. The* ***goal of the research*** *is to extend a unified model capable of forecasting energy consumption under different security mitigation strategies and to quantify their trade-offs between efficiency and trustworthiness.* ***The methodology*** *combines mathematical modeling, simulation, and empirical measurements conducted on networks of mobile devices.* ***Case studies*** *employing ray tracing workloads were used to compare baseline energy efficiency with scenarios involving node outages, malicious result injection, and data leakage attempts.* ***Results*** *demonstrate that each mitigation strategy entails distinct energy implications. Node outage protection introduces moderate and predictable overheads that scale with failure probability. Cheating mitigation, based on redundancy and trust evaluation mechanisms, is identified as the most energy-expensive approach, with overhead rising exponentially in adversarial environments and, in extreme cases, collapsing system performance. Data fishing mitigation produces minimal additional energy consumption but restricts task allocation, potentially leading to incomplete job execution if trusted nodes are unavailable. The* ***scientific novelty*** *of this work lies in the integrated framework that unifies security and reliability risk mitigation with quantitative energy modeling in distributed computing. Unlike prior studies that address risks in isolation, this research systematically compares their energy impacts, establishing the taxonomy of overhead profiles and providing a foundation for balancing efficiency and security in future large-scale distributed systems.*

Key words: *volunteer computing, distributed computing, distributed computing systems, energy efficiency, information security, mathematical modeling.*

О. О. МАМЧИЧ
аспірант кафедри електронно-обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0009-0001-6602-2929

М. О. ВОЛК
доктор технічних наук,
професор кафедри електронно-обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0000-0003-4229-9904

МОДЕЛЬ ТА МЕТОДОЛОГІЯ ВИМІРЮВАННЯ ТА ПРОГНОЗУВАННЯ ВИТРАТ ЕНЕРГІЇ НА МІНІМІЗАЦІЮ РИЗИКІВ ІНФОРМАЦІЙНОЇ БЕЗПЕКИ В РОЗПОДІЛЕНИХ ОБЧИСЛЮВАЛЬНИХ СИСТЕМАХ НА ОСНОВІ ПЕСОНАЛЬНИХ МОБІЛЬНИХ ПРИСТРОЇВ

Стрімке зростання хмарних обчислень та штучного інтелекту суттєво збільшило енергоспоживання дата-центрів. Волонтерські обчислення (Volunterr Computing) з використанням персональних мобільних пристроїв є перспективним рішенням завдяки потенційному зниженню енергетичних витрат. Водночас такі системи

© Mamchych O. O., Volk M. O., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

стикаються з критичними викликами, пов'язаними з надійністю, безпекою та стійкістю. ***Предметом даного дослідження*** *є мінімізація трьох ключових ризиків –* ***відмова вузлів, фальсифікація результатів та викрадення даних*** *– у середовищі мобільних добровільних обчислень, а також оцінка їх* ***енергетичних витрат****.* ***Мета роботи*** *полягає у розширенні єдиної моделі, здатної прогнозувати енергоспоживання за різних стратегій безпекового захисту, та кількісному визначенні компромісів між ефективністю і надійністю.* ***Методологія*** *дослідження поєднує математичне моделювання, імітаційні експерименти та емпіричні вимірювання на мережах мобільних пристроїв. У якості* ***прикладів*** *застосовувалися задачі трасування променів, що дозволило порівняти базовий рівень енергоефективності зі сценаріями відмов вузлів, навмисного спотворення результатів та спроб витоку даних. Отримані результати свідчать, що кожна стратегія захисту має власні енергетичні наслідки. Захист від відмов вузлів зумовлює помірні й передбачувані додаткові витрати, які зростають з ймовірністю відмови. Захист від фальсифікації результатів, заснований на механізмах надлишковості та оцінки довіри, є найбільш енерговитратним підходом, причому накладні витрати експоненційно зростають у ворожому середовищі й у крайніх випадках призводять до колапсу системи. Захист від викрадення даних створює мінімальні додаткові енергетичні витрати, проте обмежує розподіл задач, що може спричинити незавершеність обчислювальних робіт за відсутності довірених вузлів.* ***Наукова новизна*** *роботи полягає у створенні інтегрованої рамкової моделі, яка поєднує мінімізацію ризиків безпеки та надійності з кількісним енергетичним моделюванням у розподілених обчисленнях. На відміну від попередніх досліджень, що розглядали ризики ізольовано, це дослідження систематично порівнює їх енергетичний вплив, формує таксономію профілів накладних витрат і створює підґрунтя для збалансування ефективності та безпеки у майбутніх масштабних розподілених системах.*

Ключові слова: *волонтерські обчислення, розподілені обчислення, розподілені комп'ютерні системи, енергетична ефективність, інформаційна безпека, математичне моделювання.*

Introduction

Cloud computing demand has grown rapidly, with hyperscale and colocation data centers now dominating global workloads [1]. Between 2014 and 2023, U.S. data center electricity use nearly tripled, rising from ~60 TWh to 176 TWh (4.4 % of national consumption), largely driven by AI-accelerated servers. Projections for 2028 estimate 325–580 TWh, equal to 6.7–12 % of U.S. electricity demand. While efficiency improvements (e.g., PUE reduction, liquid cooling) mitigate some impact, the unprecedented rise of GPU-intensive training and inference workloads suggests that power use will continue outpacing efficiency gains, posing significant sustainability and infrastructure challenges.

As highlighted in the article The Case for Energy-Proportional Computing [2], current server infrastructures consume disproportionately high amounts of energy even at low utilization levels, where they typically operate. This inefficiency not only increases operational costs but also raises sustainability challenges. The concept of energy proportionality– designing systems that scale power usage in line with workload intensity – emerges as a critical paradigm. CPUs already demonstrate partial progress, but memory, storage, and networking subsystems remain major bottlenecks. Achieving true energy-proportional architectures would allow significant reductions in data center energy demand, potentially doubling efficiency without sacrificing performance. In the context of cloud computing's surging demand, addressing these inefficiencies is essential to balance performance, cost, and ecological responsibility.

As we can see, distributed computing and cloud computing is always related to a high energy consumption. Conventional stationary devices used in datacenters offer huge performance but at a high cost of energy. In our previous papers we suggested offloading of computing tasks to a distributed computing network built of personal mobile devices. This concept proven working and reducing power consumption by 3–4 times [3].

Using of personal mobile devices rises various security risks. There are approaches to mitigate those risks but they come at a cost of additional computational resources hence power. First, security risks represent a major concern. Personal devices are heterogeneous and usually lack uniform security standards. Malicious participants may intentionally inject corrupted data or manipulated results into the system (result cheating). Second, reliability and consistency risks arise from the variability of personal devices (node outage). Users may disconnect unpredictably, provide limited processing power, or operate under unstable network conditions. This volatility complicates scheduling and increases the probability of incomplete or inconsistent results. Third, users may want to steal sensitive data which may be a source data or results of computing (data fishing). Furthermore, diverse hardware and operating system configurations create reproducibility challenges, as energy and performance characteristics may differ substantially. The potential for "hidden exploitation" must be considered when designing fair participation models. Finally, sustainability risks are notable. Large-scale distributed computing may significantly increase overall energy consumption at the edge of the network. While individual contributions appear minor, aggregated impact can be substantial, undermining the intended efficiency gains if not managed with energy-aware scheduling. In summary, volunteer computing on personal devices offers opportunities but entails complex security, reliability, legal, and sustainability risks. Addressing these systematically is essential for safe and effective deployment. The main goal of this paper is to mitigate node outage, result cheating and data fishing risks in volunteer computing and compute how much of power will it cost.

State of the art

Volunteer computing (VC) offers a low-cost and sustainable model by aggregating idle resources of mobile devices, yet its openness and anonymity expose it to severe security risks. Android, the dominant VC platform, is increasingly targeted by advanced malware, threatening system reliability. To tackle this, the authors introduce MulDroid [4], a hybrid deep learning framework combining LSTM, CNN, and DNN to detect multi-vector Android malware. Using rigorous preprocessing of application manifests and evaluation on large-scale datasets (AMD, AndroZoo), MulDroid achieves 99.01 % detection accuracy and outperforms benchmark models in precision, recall, and F1-score, with only minor efficiency trade-offs. The work advances scalable and secure mobile VC by embedding autonomous malware defense at its core.

Volunteer computing (VC), and particularly its mobile variant (MVC), faces critical security challenges stemming from the heterogeneity and instability of volunteered resources. Protecting participant devices, ensuring trustworthy task execution, and safeguarding data transfers remain central obstacles. The authors address the problem of absent systematization in MVC research [5], where existing studies either overlook resource allocation or treat security and scheduling only superficially. To resolve this gap, they conduct a broad scoping review and develop the first comprehensive taxonomy of MVC resource allocation. Their approach categorizes methods by application type (independent, dependent, stream tasks), target resources (edge- or volunteer-oriented), and scheduling policy (objectives, constraints, algorithms). This structured framework not only clarifies current solutions but also identifies open issues such as incentive mechanisms, fairness, and integrated security models. The taxonomy thus provides a scientific foundation for advancing secure and efficient MVC platforms.

Unlike traditional cloud or edge servers, volunteer devices are unpredictable. These uncertainties create risks of task failures, delays, and security breaches, making providers hesitant to adopt VEC infrastructures. To tackle this problem, the authors propose VonEdgeSim [6], the first simulation framework specifically designed to mimic the dynamic and unreliable behavior of volunteer devices at the edge. The simulator models essential factors such as trust, availability, energy consumption, latency, and failures, offering a realistic environment for evaluating resource management strategies before real-world deployment. Implemented in Python for accessibility and extensibility, VonEdgeSim enables testing with IoT applications (e.g., augmented reality, infotainment, health monitoring), demonstrating significant reductions in task delay and execution time while improving fault tolerance. In conclusion, the authors close a critical research gap by delivering a practical, trust-aware simulation platform that empowers researchers to explore secure and efficient integration of volunteer resources in edge computing. Their results highlight the feasibility of incorporating volunteers into edge networks, paving the way for more reliable, scalable, and secure VEC infrastructures.

Volunteer computing provides massive, low-cost computational power but faces fundamental security and reliability challenges, as participating devices are anonymous, heterogeneous, and potentially malicious. The central problem lies in guaranteeing the correctness of results produced under such uncertain conditions. To address this, the authors introduce BOINC [7], an open-source middleware designed to ensure trustworthy high-throughput computing. Their solution integrates replication-based and adaptive result validation to detect faulty or cheating contributors, homogeneous redundancy and version control to counter platform inconsistencies, and sandboxing through restricted execution and virtualization to safeguard host systems. Together, these mechanisms establish a balance between security, efficiency, and scalability, transforming inherently insecure volunteer resources into a reliable infrastructure for large-scale scientific computation.

VFuse [8] suggests different approach, a browser-based decentralized architecture that mitigates security issues by leveraging WebAssembly for safe multi-language workflow execution, IPFS for distributed and tamper-resistant storage, and Ethereum blockchain with NFTs for transparent, verifiable, and incentivized participation. This approach solves the problem of secure orchestration and engagement in large-scale volunteer computing by combining P2P networking, blockchain-based trust, and user-friendly web notebooks into a robust framework. The main disadvantage of this system is high computational overhead and inability to utilize hardware acceleration in a flexible way.

Traditional volunteer computing suffers from unpredictable performance, availability, and security risks, making it unsuitable for sensitive or time-critical tasks. To overcome this, the authors of VECTrust [9] propose a two-stage probabilistic trust model that evaluates resources based on performance, agility, cost, and security (PACS). Using Dirichlet distributions, the model dynamically characterizes intra- and inter-cluster trust, enabling optimized resource allocation. Experimental validation on bioinformatics workflows demonstrates that VECTrust outperforms baseline algorithms, ensuring more secure, efficient, and balanced resource utilization across distributed volunteer environments.

We can state that there are various solutions mitigating security risks of volunteer computing. But none of those address all 3 highlighted problems (node outage, data fishing and result cheating). Plus, none of those problems investigate the cost of overhead while mitigating those risks. The goal of this paper is to cover all those 3 problems in terms of unified model and method for forecasting energy consumption in distributed computing systems based on stationary and mobile devices and calculate energy impact of each.

Case Study

As a base mathematical model, we will take a model from our previous research. As a computing Job we will take the same ray tracing engine but with motion blur effect: the first wave of Tasks is rendering of raw frames Artifacts, the second wave of Tasks will be blending of 4 raw frames into 1 blurred frame Artifact. This means that for Nth blurred frame

we need 4 raw frames: N, N-1, N-2 and N-3 (figure 1). Raw frames have Full HD resolution (1920 × 1080), rendered with 64 rays per pixel which gives us 133M rays per frame. Each ray trace takes average 260 float operation to trace: simple reflection, no rays multiplying, no more than 5 reflections, only spherical objects. Blurred frames have the same resolution, but blending of 4 pixels into one takes 24 float operations. The computing Job contains 2048 Tasks: 1024 raw frames renders and 1024 blended frames renders. The computing Job contains 3072 artifacts: 1024 source scene states, 1024 raw frames and 1024 blended frames. Source scene state is vector and has negligible size (up to 2kb), but each frame is 6.2Mb – we decided to keep those image as raw bitmaps to emphasize data transfer impact.

As a computing network we take a network of multiple Apple iPhone 12 mini devices connected into a network. For simplicity we will use CPU only computing. To simulate higher number of mobile devices we will run several execution contexts on each device. An execution context can execute a Task and has its own Scope of visible Artifacts, but several execution contexts cannot interact between each other. This allows graceful simulation of network transfer but does not affect the final amount of energy consumed on computation. For power consumption this time we will use a laboratory power source. Before case study we change mobile devices to 100 %, dim screens and disable all background activity to avoid any random irrelevant power drain. Other simulation parameters will also be taken from previous research as the device is the same: simulation tick duration – 1 second, network bandwidth 5.7 Mb per second via mobile internet, computing power – 16.4 GFLOP per second (63 million rays 260 FLOPs each), networking power consumption – 0.43 J per second, CPU consumption – 5.1 J per second.

Each measurement is conducted 3 times and averaged.

As a reference consumption we will take consumed power of computing on one single mobile device (to minimize data transfer) having 100 % uptime. This will be the minimal possible power can be consumed in terms of our model.

Simulation: time – 2,327 seconds, 11,747 J consumed.

Measurement: time – 2,468 seconds, 11,960 J consumed.

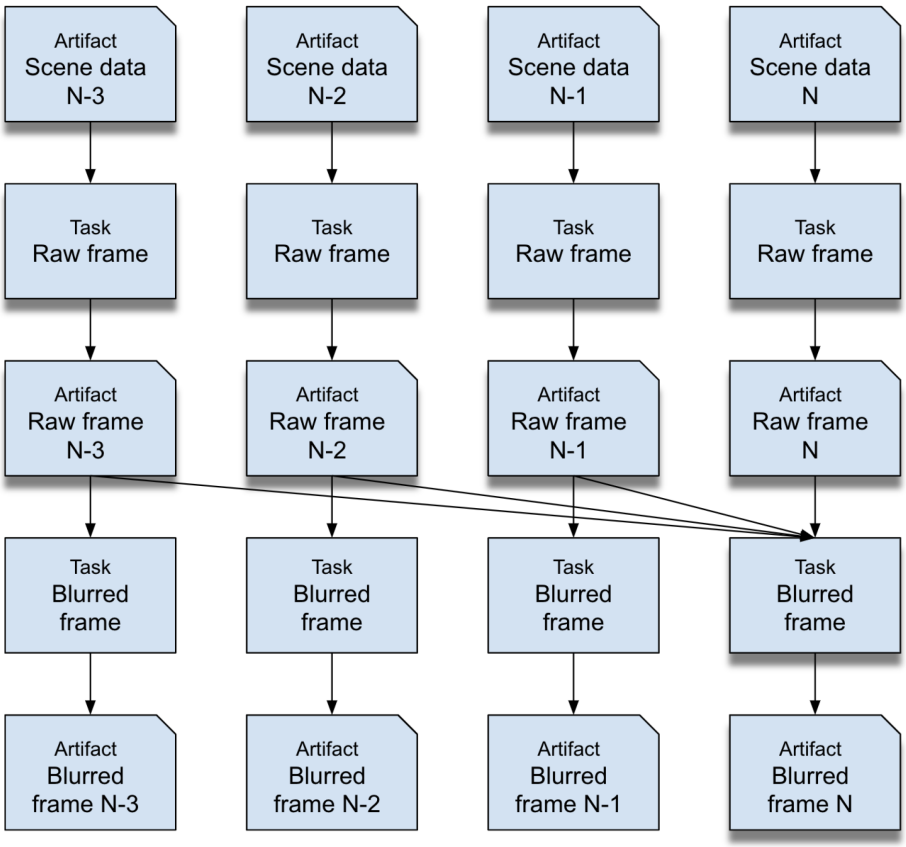


Fig. 1. Structure of a computing Job, relation between Tasks and Artifacts

Node outage power impact

The unified model already contains the simulation of nodes outage through Markov process: for each tick there is a probability to switch state from on to off and from off to on. When node goes off, it loses all the progress computing current task and all the Scope (known Artifacts). After node comes online it needs to download input Artifacts for any Task from scratch which will have an additional impact.

Let set a failover probability to 0.1 per average time of 1 raw frame Task calculation + 1 blurred frame task calculation duration, and restoring probability of 1 per tick. Average raw + blurred calculation is 2468/1024 = 2.41 seconds. If we assume normal distribution, the probability to fail within 1 tick (1 second) will be 0.043. Let setup these values and run the simulation.

Assume having 10 nodes.

Simulation: time – 243 seconds, 50 Tasks totally failed, 12263J consumed.

Measurement: time – 268 seconds, 47 Tasks totally failed, 12608J consumed.

Node cheat result mitigation

We investigate the case when some node can calculate result incorrectly. This can happen on purpose. For example, you have a computing task you want to offload to personal mobile devices and someone wants to compromise your results. They can add some compromised devices to the computing system, the system will distribute some tasks to those devices and the final result will be partially or completely damaged. To mitigate this the system can integrate some percentage of redundancy to the system to compare results of random nodes and detect cheaters. BOINC does this in their system, and our solution is inspired by theirs.

Our cheating protection subsystem is aimed at reduction of side effects caused by unscrupulous users and inappropriate hardware, yet reduce power consumption overhead as lower as possible. The subsystem has the following algorithm. Each node has following associated values: verification count (V) – number of sequentially validated results, errors counter (E) – number of errors results the node has submitted, excluded flag (X) – nodes with this flag are excluded from tasks distribution, trusted flag (T) – nodes with this flag considered trusted and should not be validated.

System has following global value: verification threshold (V_S) – minimal number of sequentially validated results the node should have to be considered reliable, ban threshold (E_S) – is node's error count goes above this threshold, node should be marked as excluded, forgiving rate R_F – rate of forgiving errors (decreasing node's E value) for reliable node after submission of valid results, unreliable node's redundancy level (R_U) is probability of result validation for unreliable node, reliable node redundancy level (R_R) is probability of result validation for reliable node, execution limit (L) – any Task can be evaluated no more than L times by non-trusted nodes. Usually, R<SUB>U<\SUB> is set to values close to 1 so near-every or every result of unreliable node is validated, but R_R is set to low values to avoid unnecessary redundancy for reliable nodes. Forgiving rate is introduced because the system itself may have flaws or Tasks may have their specifics, otherwise, some random rare algorithmic issue will lead to constant growth of E value indefinitely.

System tracks following audit trail for each executed Task and related Artifact: list of nodes evaluated or evaluating this Task. System can evaluate L(Task) as a number of nodes evaluated or evaluating this Task minus trusted nodes evaluated or evaluating this Task. Also, system tracks validation map: for each node it stores tasks it was validated on.

Algorithm is split into 2 phases: Task acquiring phase when system distributes particular task on a node and task submission phase when system receives Task from a node and takes decision. We also assume there are several thrust Task acquiring phase:

1. Validation dice roll. If X flag is set, reject Task acquiring, end phase.
2. If T flag is set no validation is required, provide any Task that is not evaluated or assigned to any node, end phase.
3. If $V >= V_S$, roll the dice with R_R rate. Otherwise roll the dice with R_U rate. If roll was positive, we decide to validate the result of Task node just acquired. If negative – provide any Task that is not evaluated or assigned to any node, end phase
4. Validation task selection. Prefer Task executed or being executed by trusted node with L(Task) < L. Otherwise prefer Task executed or being executed by at least 1 reliable node. Otherwise select any task assigned to other node. End phase.

Task submission phase – node evaluated Task and submitted Artifacts to system. If Task was not calculated by another node, do nothing. If task was calculated by multiple nodes, run the following node reliability test for each node.

1. If submitter node is trusted (T = true), add node-task pair to validation map, end of test.
2. If submitter node is excluded (X = true), end of test.
3. Compare submitted Task result to result from trusted node. If result matches, increment V, add node-task pair to validation map, end the test. If not matches, set V to 0, increment E.
4. Otherwise, compare submitted Task result to another node's result which validated is already validated for this Task. If result matches, increment V, add node-task pair to validation map, increment another node's V, end the test. If not matches, set V to 0, increment E.
5. Otherwise, compare submitted Task result to another node's result. If result matches, increment V, add node-task pair to validation map, end the test. If not matches, set V to 0, increment E.

After the test is finished positively, run forgiving routine: roll the dice with RF rate, if positive – decrease submitter node's E by 1, capped by 0.

After the test is finished negatively, of there is no validations for this Task in validation map, rollback the Task evaluation and Task assignment, consider it's never evaluated. If submitter node's $E > E_S$, exclude node by setting node's X to true.

Let's make the following configuration: reliability threshold $V_S = 5$, error threshold $E_S = 3$, forgiving rate $R_F = 0.1$, redundancy rate for unreliable node $R_U = 1$, redundancy rate for unreliable node $R_R = 0.05$.

Assume our computing system has 10 nodes: one is trusted (for example, tested system owner's device), 7 nodes will make correct calculations, one node will be doing only false results, one node will be doing 1 out of 4 false results.

Simulation: time – 311 seconds, 2 nodes excluded, 15,531 J consumed (+32 %).

Measurement: time – 336 seconds, 2 nodes excluded, 16,065 J consumed (+34 %).

Assume our computing system has 10 nodes: no trusted, 8 nodes will make correct calculations, one node will be doing only false results, one node will be doing 1 out of 4 false results.

Simulation: time – 311 seconds, 2 nodes excluded, 15,502 J consumed (+32 %).

Measurement: time – 334 seconds, 2 nodes excluded, 15,969 J consumed (+34 %).

Assume our computing system has 10 nodes: 1 trusted, 1 node will make correct calculations, 8 nodes will make false result every second time.

Simulation: time – 3,170 seconds, 9 nodes excluded, 48,500 J consumed (+312 %).

Measurement: time – 3,332 seconds, 9 nodes excluded, 51,991 J consumed (+334 %).

Assume our computing system has 10 nodes: no trusted, 2 nodes will make correct calculations, 8 nodes will make false result every second time.

Simulation: all nodes excluded.

Measurement: all nodes excluded.

Node data fishing mitigation

Malefactor may connect one or several compromised devices to our computing network which will result in stealing some source or intermediate data. Unfortunately, there are no reliable ways to prevent this at 100 % rate, but we can minimize amount of leaked data. To reach peak energy efficiency we should compute the whole Job on a single device to minimize data transfer efforts. But to minimize amount of leaked data we need to make the opposite: to involve as many devices as we can, so each computing device would "see" the least amount of data possible. And, if there are not enough devices to reach desired level of max leaked data, we need to involve trusted devices that are guaranteed to be reliable.

Each node will have an evaluation history containing all Task it's evaluated hence and all Artifacts seen. Also, node has trusted (T) flag.

System has a max number A<SUB>S<\SUB> of Artifacts that non-trusted node can "see".

Task acquiring flow will look as follows:

1. If node is trusted (T), approve acquiring, end flow.
2. Calculate number of Artifacts seen by node based on evaluation history of the node.
3. Pick available Task with biggest number of related (input + output) Artifacts that will not lead to exceeding A_S, taking into account Artifacts known to node. If there is such Task – allocate it and assign to the node. Otherwise – reject node.

If node is rejected, it will try to acquire tasks periodically since later unblocked Tasks can require less Artifacts and may depend on Artifacts known to the node.

Assume we have 10 nodes: 1 trusted, 9 regulars. System's $A_S = 400$ (> 10 %, we have 3,072 artifacts total).

Simulation: time – 237 seconds, average Scope is 312 Artifacts, max Scope is 322, 12,096 J consumed (+3 %).

Measurements: time – 252 seconds, average Scope is 312 Artifacts, max Scope is 320, 12,211 J consumed (+2 %).

Assumer we have 100 nodes: 10 trusted, 90 regulars. System's $A_S = 40$ (> 1 %).

Simulation: time – 25 seconds, average Scope is 31 Artifacts, max Scope is 36, 12,596 J consumed (+7.2 %).

Measurements: time – 28 seconds, average Scope is 31 Artifacts, max Scope is 40, 12875 J consumed (+7.6 %).

Assume we have 100 nodes: 1 trusted, 99 regulars. System's $A_S = 20$ (< 1 %).

Simulation: time – 856 seconds, average Scope is 33 Artifacts, max Scope is 2,590, 12,477 J consumed (+6 %).

Measurements: time – 912 seconds, average Scope is 34 Artifacts, max Scope is 2624, 12,938 J consumed (+8.1 %).

Experimental results discussion

As we can see every measure we take to mitigate security risks reflect in power consumption overhead. The simplest correlation with failover rate: higher failover rate results in re-distribution tasks into other nodes. In our case when 9 % tasks were failed, we've got about 5.4 % (4.4 %) of measured (simulated) power consumption overhead which look reasonable and not that high. Apparently, for high failure rates the overhead will grow to infinity as we close to 100 % fail rate.

On the other hand, cheating mitigation overhead is usually higher. If we have even 20 % of cheating nodes, we get 34 % (32 %) of measured (simulated) overhead. And in this case, it doesn't make sense if we have trusted nodes or not – the result is pretty the same, especially if total number of nodes is high. But when number of cheaters grows the overhead grows much faster. If we have trusted node (at least one) overhead can reach 334 % (312 %) of measured (simulated) overhead. If we don't have trusted node, the computing network collapses. We've missed the case when cheating nodes are interconnected and can coordinate their results and we don't have trusted nodes. This case requires additional investigation.

Data fishing mitigation does not generate such an impact by energy, neither by measured (2–7 %) nor by simulated (3–7 %). It doesn't differ from regular distribution of computational tasks into the maximum nodes possible if there are enough nodes to calculate the Job without exceeding data sharing threshold. And the overhead grows proportionally to

number of nodes. Even if we don't have enough nodes to not exceed data sharing threshold, we don't experience energy consumption impact because exceeding data is processed on trusted nodes. On the other hand, if we don't have enough of nodes and no trusted nodes, system will not be able to calculate the whole Job without exceeding data sharing threshold. At this point we can only rise the threshold or reconsider the Job execution.

We can see that every security risk mitigation has their own energy overhead profile: failure rate just adds a plain multiplier for low fail rates and blocks the computing network when fail rate approaches 100 %, cheater mitigation overhead is the highest even for no cheaters, and it grows to hundreds of percents when cheaters percentage is high. The data fishing mitigation energy impact is the lowest, but the data fishing mitigation results in outage if we don't have trusted nodes (figure 2).

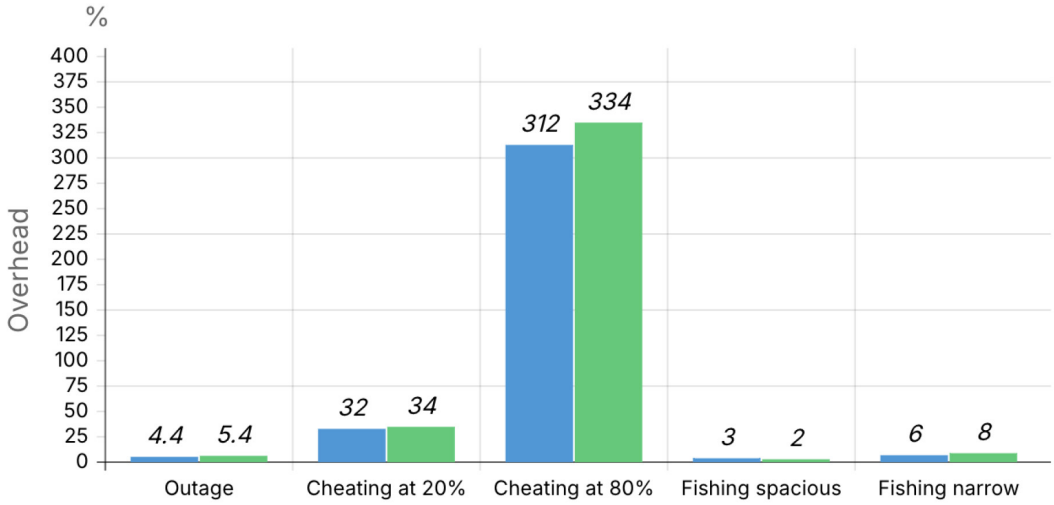


Fig. 2. Comparison of overheads from risks mitigation application

Conclusion

The conducted research highlights that integrating volunteer mobile devices into distributed computing infrastructures offers noticeable energy efficiency gains but inevitably introduces significant security and reliability risks. Through systematic modeling and experimentation, we demonstrated that mitigation strategies against node outage, cheating, and data fishing each carry distinct power consumption overheads, shaping the practical feasibility of such systems.

Node outage introduces moderate and predictable energy penalties, scaling with failure probability, while cheating mitigation emerges as the most energy-intensive, with overheads escalating rapidly under adversarial conditions and sometimes collapsing the system entirely. In contrast, data fishing mitigation imposes relatively low overhead but introduces operational constraints that may lead to task incompletion in absence of trusted nodes. Taken together, these results reveal a trade-off between energy efficiency and security in volunteer mobile computing. Effective deployment requires not only technical safeguards but also carefully balanced trust models and adaptive redundancy policies.

Future research will be targeted at several directions. First, development of better mitigation strategies, based of modern mathematical approaches. Second, we need a better model of energy losses for each mitigation strategy, we think there is a simplified "good enough" way of predicting those without complex simulation. In this research we have assumed that trust nodes are the same mobile devices, which resulted in execution time bottleneck. For real usage, to keep execution time under control, it would be better to use conventional data center hardware as a fallback. So, third target is to investigate mitigation strategies taking into account different energy consumption of trusted nodes to minimize energy profile, as well as growth of power consumption when risks mitigation strategies stick to fallback trusted conventional hardware.

Bibliography

1. Shehabi, A., Newkirk, A, Smith, S.J., Hubbard, A., Lei, N., Siddik, Md A. B., Holecek, B., Koomey, J., Masanet, E., & Sartor, D. 2024 United States Data Center Energy Usage Report Lawrence Berkley National Laboratory, 2024, Report LBNL-2001637, DOI: 10.71468/P1WC7Q

2. Barroso, L. A., Hölzle, U., The Case for Energy-Proportional Computing, 2007, Computer, no. 40, vol. 12, pp. 33–37. DOI:10.1109/mc.2007.443

3. Mamchych O., Volk M., A unified model and method for forecasting energy consumption in distributed computing systems based on stationary and mobile devices, RADIOELECTRONIC AND COMPUTER SYSTEMS, 2024, vol. 2, pp. 120–135, DOI: 10.32620/reks.2024.2.10

4. Bibi, I., Akhunzada, A., Malik, J., Khan, M. K., & Dawood, M. Secure Distributed Mobile Volunteer Computing with Android, ACM Transactions on Internet Technology, 2021, no. 22, vol. 1, pp. 1–21. DOI:10.1145/3428151

5. Ma, P., Garg, S. & Barika, M. Research allocation in mobile volunteer computing system: Taxonomy, challenges and future work, Future Generation Computer Systems, 2024, no. 154, pp. 251–265. DOI:10.1016/j.future.2024.01.015

6. Alsenani, Y. VonEdgeSim: A Framework for Simulating IoT Application in Volunteer Edge Computing, Electronics, 2024, no. 13(20), p. 4124. DOI:10.3390/electronics13204124

7. Anderson, D. P. BOINC: A Platform for Volunteer Computing, Journal of Grid Computing, 2019, no. 18(1), pp. 99–122. DOI:10.1007/s10723-019-09497-9

8. Antelmi, A., D'Ambrosio, G, Petta, A., Serra, L., & Spagnuolo, C. A Volunteer Computing Architecture for Computational Workflows on Decentralized Web, IEEE Access, 2022, no. 10, pp. 98993–99010. DOI:10.1109/access.2022.3207167

9. Pandey, A., Calyam, P., Debroy, S., Wang, S., & Alarcon, M.L. VEC Trust Proceedings of the 14th IEEE/ACM International Conference on Utility and Cloud Computing: 2021 IEEE/ACM 14th International Conference on Utility and Cloud Computing, 2021, doi:10.1145/3468737.3494099

Дата першого надходження рукопису до видання: 28.09.2025
Дата прийнятого до друку рукопису після рецензування: 24.10.2025
Дата публікації:

УДК 004.94 DOI

М. О. МЕДИКОВСЬКИЙ
доктор технічних наук, професор,
професор кафедри автоматизованих систем управління
Національний університет «Львівська політехніка»
ORCID: 0000-0003-2492-8578

Р. В. МЕЛЬНИК
доктор філософії, асистент,
асистент кафедри автоматизованих систем управління
Національний університет «Львівська політехніка»
ORCID: 0000-0003-0619-1613

М. В. МЕЛЬНИК
аспірант кафедри автоматизованих систем управління
Національний університет «Львівська політехніка»
ORCID: 0000-0003-0339-3711

Д. О. ГЕРЕЗ
магістр кафедри автоматизованих систем управління
Національний університет «Львівська політехніка»
ORCID: 0009-0000-0730-1584

МЕТОД КОМП'ЮТЕРНОГО МОДЕЛЮВАННЯ ПРОЦЕСІВ ЗАРЯДУ/РОЗРЯДУ АКУМУЛЯТОРА

У статті розв'язано задачу моделювання процесів заряду і розряду літій-залізо-фосфатного акумулятора на основі історичних даних його експлуатації. Основна ідея роботи полягає у розробленні методу, який забезпечує можливість визначати напругу на клемах, яка характеризує його енергетичний потенціал, враховуючи ключові параметри акумулятора та режимів його роботи за різних умов. Дослідження проведено у два етапи. На першому етапі виконано збір, аналіз, історичних даних про роботу гібридної системи електропостачання з використанням електричного акумулятора енергії. Обґрунтовано структуру і параметри еквівалентної схеми, залежності для їх визначення на підставі статистичних даних. Здійснено опрацювання статистичної інформації про режими роботи акумулятора в системі електропостачання. На другому етапі обґрунтовано можливість та доцільність застосування методу найменших квадратів (МНК) або генетичних алгоритмів (ГА) для знаходження коефіцієнтів при аргументах функції регресії, яка відображає доступну (залишкову) ємність акумуляторної батареї. Коефіцієнти визначено для восьми режимів роботи акумулятора, які повністю покривають робочий діапазон.

Актуальність даного дослідження зумовлена зростаючим попитом на ефективні інструменти дослідження характеристик систем електропостачання за наявності засобів акумулювання енергії. Процеси заряду і розряду є фундаментальними для оцінки терміну служби акумуляторів, розробки систем моніторингу та керування режимами. Розроблений метод є корисним в умовах відсутності прямого доступу до лабораторних даних або у випадках, коли виробник не надає всієї необхідної інформації. Метод, зокрема, надає можливість змоделювати роботу акумулятора в різних умовах використання та для різних структур енергетичних систем.

Результати роботи можуть бути використані при автоматизованому проектуванні електричних систем з урахуванням індивідуальних характеристик даного типу акумуляторів. Забезпечено можливість уточнити параметри режимів роботи вже встановленого акумулятора, що може бути корисним при модернізації систем електропостачання.

Ключові слова: *літій-залізо-фосфатний акумулятор, процеси заряду і розряду, історичні дані, еквівалентна модель, метод найменших квадратів, метод генетичних алгоритмів.*

© Медиковський М. О., Мельник Р. В., Мельник М. В., Герез Д. О., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

M. O. MEDYKOVSKYI
Doctor of Technical Sciences, Professor,
Professor at the Department of Automated Control Systems
Lviv Polytechnic National University
ORCID: 0000-0003-2492-8578

R. V. MELNYK
PhD, Assistant,
Assistant at the Department of Automated Control Systems
Lviv Polytechnic National University
ORCID: 0000-0003-0619-1613

M. V. MELNYK
Postgraduate Student at the Department of Automated Control Systems
Lviv Polytechnic National University
ORCID: 0000-0003-0339-3711

D. O. HEREZ
Master at the Department of Automated Control Systems
Lviv Polytechnic National University
ORCID: 0009-0000-0730-1584

COMPUTER MODELING METHOD FOR BATTERY CHARGE AND DISCHARGE PROCESSES

This paper addresses the challenge of modeling the charging and discharging processes of a lithium iron phosphate battery using historical operational data. The work's main idea is to develop a method that can determine the terminal voltage, which characterizes the battery's energy potential. This method takes into account key battery parameters and operating modes under various conditions.

The research was carried out in two stages. First, historical data on the operation of a hybrid power supply system with an electrical energy storage battery was collected and analyzed. The structure and parameters of the equivalent circuit were substantiated, along with the dependencies for determining them based on statistical data. Statistical information on the battery's operating modes within the power supply system was processed.

In the second stage, we substantiated the feasibility and practicality of applying the Least Squares Method (LSM) or Genetic Algorithms (GA) to determine the coefficients for the arguments of the regression function representing the battery's available (remaining) capacity. Coefficients were obtained for eight operating modes of the battery, which fully cover its operational range.

This study is relevant due to the growing demand for effective tools to investigate the characteristics of power supply systems equipped with energy storage devices. Charging and discharging processes are fundamental to assessing battery lifespan, developing monitoring systems, and controlling operating modes. The developed method is useful when direct access to laboratory data is unavailable or when the manufacturer does not provide all the necessary information. Notably, the method enables simulation of battery operation under various usage conditions and for different power system configurations.

The results can be applied when designing electrical systems that take into account the specific characteristics of this type of battery. The findings also allow for refining the operating parameters of an already installed battery, which can be beneficial during power supply system upgrades.

Key words: *lithium iron phosphate battery, charging and discharging processes, historical data, equivalent circuit model, Least Squares Method, Genetic Algorithms method.*

Постановка проблеми

Зі зростанням популярності гібридних систем електропостачання, зокрема, з використанням фотоелектричних джерел енергії у поєднанні зі сховищами у вигляді електричних акумуляторів, зростає потреба у створенні точних математичних та комп'ютерних моделей, які відображають поведінку системи та її елементів в умовах реальної експлуатації. Такі моделі є основою для розробки ефективних методів та алгоритмів системи енергоменеджменту, оцінки ефективності інвестицій, прогнозування деградації засобів акумулювання та запобігання аварійним ситуаціям.

Важливим елементом гібридних енергетичних систем є акумуляторна батарея, яка виконує функції демпфера між непостійними генерацією та навантаженням. Проте динаміка її можливостей значною мірою залежить від зовнішніх умов, режимів роботи та технічних характеристик, які не завжди надаються виробником у повному обсязі. Це створює труднощі при побудові аналітичних моделей акумуляторів, зокрема таких, що враховують: залежність напруги від рівня заряду (State of Charge, SOC) та струму в різних режимах заряду/розряду, внутрішній опір батареї та його зміну з часом, поступову деградацію і зниження ємності, температурний вплив на продуктивність та ефективність.

Також, в реальних системах часто фіксується лише частковий набір даних, який отримується з системи управління батареєю (Battery Management System, BMS) або інверторів (наприклад: напруга, струм, потужність, температура та SOC). Це ускладнює пряме використання класичних аналітичних моделей, натомість вимагає застосування адаптивних інформаційних методів з аналізом даних – таких як метод найменших квадратів, фільтри Калмана, генетичні алгоритми або машинне навчання.

Таким чином, актуальність дослідження полягає в необхідності створення комп'ютерної моделі $LiFePO_4$ акумулятора на основі доступних історичних даних експлуатації, яка може бути використана для подальшого розроблення методів та засобів керування зарядом/розрядом у гібридних системах електропостачання, за умов реальних енергодинамічних процесів та обмеженого доступу до специфікацій обладнання.

Аналіз останніх досліджень і публікацій

Існує кілька підходів до моделювання роботи акумуляторів: експериментальні, електрохімічні, математичні та, моделі на основі еквівалентних схем (ЕС) [1].

Експериментальні моделі показують високу точність, проте мають суттєві недоліки. При цьому необхідне проведення масштабних експериментів для визначення параметрів моделі і є обчислювально затратними [2].

В свою чергу, математичні моделі базуються на емпіричних рівняннях, що описують процеси заряду та розряду, ефективність перетворення енергії, тощо [3]. Одним з недоліків таких моделей є фіксовані параметри, які адекватно відображають лише результати моделювання режимів акумуляторів певного типу.

Моделі на основі ЕС подають акумулятор як набір електричних компонентів, таких як резистори, джерело напруги, конденсатори. Такі моделі широко застосовуються в системах управління акумулятора у зв'язку з своєю простотою і універсальністю. Параметри моделей на основі ЕС представляються як складові поліноміального рівняння, що відображає функцію рівня заряду. Для побудови такої моделі необхідно визначити параметри акумулятора за різних режимів роботи. Зазвичай, їх встановлюють експериментальним шляхом [3, 4, 5].

В роботі [6] рекомендується використання динамічних резисторно-конденсаторних моделей, моделей Тевенена першого та другого порядку, для підвищення адекватності моделювання процесів у літій-іонних акумуляторах. В даній роботі використана модель Тевенена першого порядку (рис. 1).

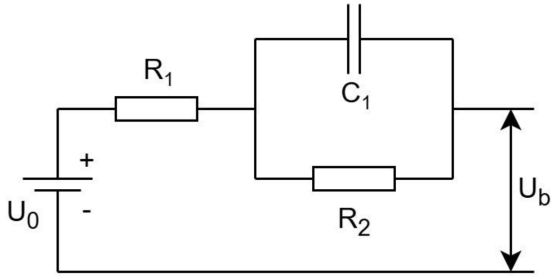


Рис. 1. Модель Тевенена першого порядку

Така ЕС характеризується трьома активними параметрами: електрорушійною силою акумулятора (U_0), внутрішнім опором (R_{int}) з двома параметрами R_1 та R_2 та ємністю (C_1), напругою на клемах акумулятора (U_b). Ця модель використовується для визначення параметрів кола при різних струмах заряду/розряду. Математичні вирази для R_1, R_2, C_1 та U_0 за умов заряджання постійним струмом сталої величини наведено в рівняннях 1–4 [1].

$$R_1 = (a_1 + a_2 C_r + a_3 C_r^2) \cdot e^{-a_4 \cdot SOC} + (a_5 + a_6 C_r + a_7 C_r^2); \quad (1)$$

$$R_2 = (a_8 + a_9 C_r + a_{10} C_r^2) \cdot e^{-a_{11} \cdot SOC} + (a_{12} + a_{13} C_r + a_{14} C_r^2); \quad (2)$$

$$C_1 = -(a_{15} + a_{16} C_r + a_{17} C_r^2) \cdot e^{-a_{18} \cdot SOC} + (a_{19} + a_{20} C_r + a_{21} C_r^2); \quad (3)$$

$$U_0 = (a_{22} + a_{23} C_r + a_{24} C_r^2) \cdot e^{-a_{25} \cdot SOC} + (a_{26} + a_{27} SOC + a_{28} SOC^2 + a_{29} SOC^3) - a_{30} C_r + a_{31} C_r^2. \quad (4)$$

Параметри при розряду акумулятора описуються рівняннями 5-8 [1].

$$R_1 = (a_1 + a_2 C_r + a_3 C_r^2) \cdot e^{-a_4 \cdot DOD} + (a_5 + a_6 C_r + a_7 C_r^2); \quad (5)$$

$$R_2 = (a_8 + a_9 C_r + a_{10} C_r^2) \cdot e^{-a_{11} \cdot DOD} + (a_{12} + a_{13} C_r + a_{14} C_r^2); \quad (6)$$

$$C_1 = -(a_{15} + a_{16} C_r + a_{17} C_r^2) \cdot e^{-a_{18} \cdot DOD} + (a_{19} + a_{20} C_r + a_{21} C_r^2); \quad (7)$$

$$U_0 = (a_{22} + a_{23} C_r + a_{24} C_r^2) \cdot e^{-a_{25} \cdot DOD} + (a_{26} + a_{27} DOD + a_{28} DOD^2 + a_{29} DOD^3) - a_{30} C_r + a_{31} C_r^2. \quad (8)$$

Обчислені коефіцієнти використовуються для розрахунку напруги на клемах акумулятора при різних струмах заряду/розряду. Рівняння 9 описує залежність між напругою на клемах і часом при заряді акумулятора заданим струмом [1].

$$U_{bc}(t_c)=\left(\left(\frac{Q}{C}+I_cR_2\right)*exp\left(-\frac{t_c}{R_2C_1}\right)\right)+U_0-\left(I_c\left(R_1+R_2\right)\right). \quad (9)$$

Аналогічно, напруга на клемах акумулятора для розряду наведена в рівнянні 10.

$$U_{bd}(t_c)=\left(\left(\frac{Q}{C}+I_dR_2\right)\cdot exp\left(-\frac{t_d}{R_2C_1}\right)\right)+U_0-\left(I_d\left(R_1+R_2\right)\right). \quad (10)$$

Точність цієї моделі перевіряється шляхом порівняння розрахованих значень напруги на клемах із виміряними значеннями напруги при різних значеннях струмів та залишкового заряду акумулятора, отриманими в процесі лабораторних досліджень.

Останні наукові дослідження демонструють значну актуальність моделювання $LiFePO_4$ акумуляторів з використанням реальних експлуатаційних даних, зокрема логів з інверторів, BMS або SCADA-систем. Такий підхід дозволяє створювати цифрові двійники акумулятора, які здатні відображати поведінку батареї у динамічних умовах експлуатації та забезпечують основу для формування інтелектуальних систем керування. Для оцінки параметрів таких моделей активно використовуються методи найменших квадратів, включаючи їх адаптивні та рекурсивні варіанти [7, 8, 9]. Проте, деякі дослідження показують, що застосування генетичних алгоритмів (ГА) для оптимізації параметрів моделей на основі даних з інвертора або системи моніторингу забезпечує кращу адаптацію моделі до реальних умов (температурні флуктуації, деградація) у порівнянні з традиційними методами найменших квадратів [10].

У публікації [11] розглянуто методи прогнозування напруги окремих елементів акумуляторних елементів мобільних роботизованих платформ за допомогою ймовірнісних підходів та методів машинного навчання. Запропоновано два підходи на основі рекурентних нейронних мереж (RNN) із попереднім аналізом, обробкою і згладжуванням даних. Найкращі результати для короткострокового прогнозування показали методи експоненціального згладжування і згладжування Lowess. Для довгострокового прогнозування рекомендовано метод ковзного середнього із додатковим згладжуванням прогнозованих значень через наявність шуму в них.

Викладення основного матеріалу дослідження

В досліджені використано історичні дані експлуатації акумулятора B0S-G [12] ємністю 5.12 кВт ·год., встановленого в багатоквартирному будинку. Гібридна система електропостачання включає в себе 44 – монокристалічні фотогальванічні модулі Jinko Solar JKM585N-72HL-V-585, гібридний інвертор DEYE SUN-30K-SG01HP3-EU BM3, систему акумулювання енергії DEYE BOS-G LiFePO4 HV 51.2v 100AH 5.12kwh в комплекті з BMS для батарей DEYE BOS-G 120-750Vdc 100A (HVB750V/100A-EU) загальною ємністю 40.96 кВт · год..

Система експлуатується в умовах реального навантаження з обмеженням на експорт енергії до зовнішньої електромережі, а також із заданим порогом мінімального рівня заряду акумулятора (SOC ≥ 20 %). Джерелом первинних даних для аналізу слугує онлайн-моніторингова платформа Deye Cloud, з якої отримані детальні журнали даних функціонування системи із 5-хвилинним інтервалом. До складу даних увійшли такі параметри: напруга на клемах батареї, струм і потужність заряду/розряду, значення SOC/DOD, температура батареї, а також низка допоміжних показників інвертора й навантаження.

Така система є характерним прикладом сучасної автономної СЕС середнього класу, що функціонує у сценарії використання сонячної енергії без можливості експорту в зовнішню електромережу та потребує адаптивного керування зарядом/розрядом на основі доступних телеметричних даних.

Метою дослідження є підвищення ефективності моделювання блоку акумулювання електричної енергії шляхом розроблення комп'ютерної моделі процесів заряду/розряду на основі історичних даних його експлуатації. Модель повинна адекватно відображати динамічні процеси заряду/розряду, забезпечувати можливість аналізу таких взаємозалежних параметрів, як напруга на клемах та залишковий заряд акумулятора при різних струмах заряду/розряду. Для цього розв'язано такі задачі: аналіз структури та динаміки даних, отриманих з платформи Deye Cloud; побудова емпіричної моделі «SOC-напруга» з використанням методів комп'ютерного аналізу, таких як метод найменших квадратів та генетичні алгоритми; виявлення характерних залежностей та трендів; проведення параметричної ідентифікації моделі акумулятора.

Розрахунки виконувались за допомогою мови програмування Python, із використанням бібліотек pandas для обробки табличних даних, numpy для числових обчислень і поліноміального моделювання, scipy.optimize для оптимізації параметрів моделей, matplotlib.pyplot для візуалізації результатів, а також sklearn.metrics.r2_score для оцінювання точності побудованих регресійних моделей.

Для моделювання використано такі дані: Timestamp – часова мітка вимірювання даних; Battery Current (A) – струм батареї. Додатні значення стосуються процесів розряджання, а від'ємні – заряджання; SOC – State of Charge, рівень заряду батареї (%); Battery Voltage ($U_{b<\sub>}$) – напруга на клемах батареї, що використовується для аналізу її стану під час зарядки та розрядки; time_diff_sec – розрахований параметр, що визначає інтервал часу між вимірюваннями.

Перш за все, проведено попереднє опрацювання даних, видалено аномальні значення і рядки з відсутніми значеннями. Отриманий набір даних розбито на вісім груп за значенням сили струму заряду/розряду, по чотири

для заряду і розряду. Перший діапазон [1А; 3А], другий [9А; 11А], третій [19А; 21А], четвертий [29А; 31А]. Такі діапазони повністю покривають робочий простір акумулятора.

Після розбиття даних на групи, в кожній групі знайдено середнє значення напруги на клемах акумулятора для кожного значення SOC. Отримані дані додатково згладжено за допомогою поліноміальної апроксимації. Такий підхід дозволяє зменшити флуктуації даних в межах окремих груп і побудувати більш стабільний профіль характеристик.

Передбачено, що програма генерує синтетичні точки для низьких значень SOC, яких не було в оригінальних даних. Це корисно, коли потрібно аналізувати повний діапазон SOC, якщо частина даних відсутня.

Ініціалізація початкового набору коефіцієнтів a_1–a_{31}дійснюється на основі аналізу характеру задачі та попередніх досліджень. Вибір початкових значень відіграє важливу роль у процесі оптимізації параметрів моделі, оскільки правильний стартовий набір може значно прискорити збіжність алгоритму та покращити точність отриманих результатів.

Ці параметри отримані шляхом багаторазового застосування методу найменших квадратів для різних випадкових значень, що дозволило знайти близькі до оптимальних значення. Такий підхід забезпечує стійкість моделі та зменшує вплив початкових умов на кінцевий результат.

Таблиця 1

Початкові значення поліноміальних коефіцієнтів

a_1 = 7.8523e–02	a_{12} = 9.2098e–03	a_{22} = 1.0420e+01
a_2 = 4.4064e–02	a_{13} = 9.2404e–03	a_{23} = 1.0267e+01
a_3 = 5.2999e–02	a_{14} = 9.2095e–03	a_{24} = 1.0188e+01
a_4 = 5.3961e+00	a_{15} = 1.6697e+00	a_{25} = 9.2330e-02
a_5 = 9.2790e–03	a_{16} = 9.4571e–01	a_{26} = 1.0381e+01
a_6 = 9.2634e–03	a_{17} = 1.0577e+00	a_{27} = 1.3262e+00
a_7 = 9.2295e–03	a_{18} = 6.0501e–01	a_{28} = 6.0602e–01
a_8 = 1.0681e–01	a_{19} = 8.0227e–01	a_{29} = 1.6621e–01
a_9 = 9.9664e–02	a_{20} = 5.5942e–01	a_{30} = 1.3637e–03
a_{10} = 7.8093e–02	a_{21} = 7.1899e–01	a_{31} = 1.0161e+01
a_{11} = 6.5482e+00		

Цільова функція забезпечує знаходження значення напруги на клемах батареї, на підставі чого можна оцінити залишковий рівень заряду та точність моделювання батареї. При обчисленні функції дані проходять покрокову обробку: для кожного рівня заряду обчислюється прогнозована напруга за допомогою залежностей 9-10 які враховує струм, час, рівень заряду, параметри батареї та тип режиму. Ці значення порівнюються з реальними вимірами напруги, після чого розраховується середньоквадратична похибка (MSE).

Додатково вводиться штраф за розкид похибок, тобто варіацію між фактичними і обчисленими значеннями. Такий підхід дозволяє уникнути ситуацій, коли модель має загалом низьку похибку, але з великими локальними відхиленнями. Об’єднання MSE з цією варіацією, зважене коефіцієнтом alpha, формує підсумкове значення функції, знаходячи такі параметри, які роблять модель максимально точною та стабільною.

Наступним кроком є подальше вдосконалення моделі шляхом комплексного застосування методу найменших квадратів та генетичного алгоритму.

Метод найменших квадратів дозволяє уточнити початкові параметри, мінімізуючи похибку між розрахунковими та фактичними значеннями. Він забезпечує поступове наближення до оптимальних коефіцієнтів шляхом ітеративного коригування результатів моделі.

Генетичний алгоритм, у свою чергу, сприяє глобальній оптимізації, використовуючи механізми природного відбору, мутації та схрещування. Це дає змогу ефективно досліджувати простір можливих рішень, знаходячи найкращі комбінації параметрів навіть у складних умовах, де традиційні методи можуть губитися в локальних мінімумах.

На рисунках 2 і 3 наведено результати моделювання роботи акумуляторної батареї і реальні історичні дані напруги на клемах U_b і залишкової ємності SOC її при процесах заряджання/розряджання струмом 19-21А.

Розроблена модель протестована на незалежному сегменті даних. Аналіз виконано шляхом порівняння змодельованих і фактичних параметрів з використанням статистичних показників: коефіцієнт детермінації R^2; середня абсолютна похибка MAE (Mean Absolute Error); середньоквадратична похибка MSE (Mean Squared Error); максимальне відхилення фактичного та змодельованого графіків PFG (Peak Fitting Gap); середня абсолютна відсоткова похибка MAPE (Mean Absolute Percentage Error). Результати аналізу наведені в таблицях 2, 3.

Аналіз результатів моделювання процесу заряду акумулятора демонструє перевагу методу генетичного алгоритму (ГА) над методом найменших квадратів (МНК) в умовах середніх та високих значень зарядного струму. Зокрема, на інтервалі струмів [29А; 31А], модель, побудована за допомогою ГА, демонструє кращу точність та вищу кореляцію за коефіцієнтом детермінації.

Водночас, для низьких струмів [1А; 3А] – спостерігається зворотна ситуація – метод найменших квадратів показує дещо кращі результати за MSE і R^2, що вказує на здатність МНК краще адаптуватися до менш динамічних режимів.

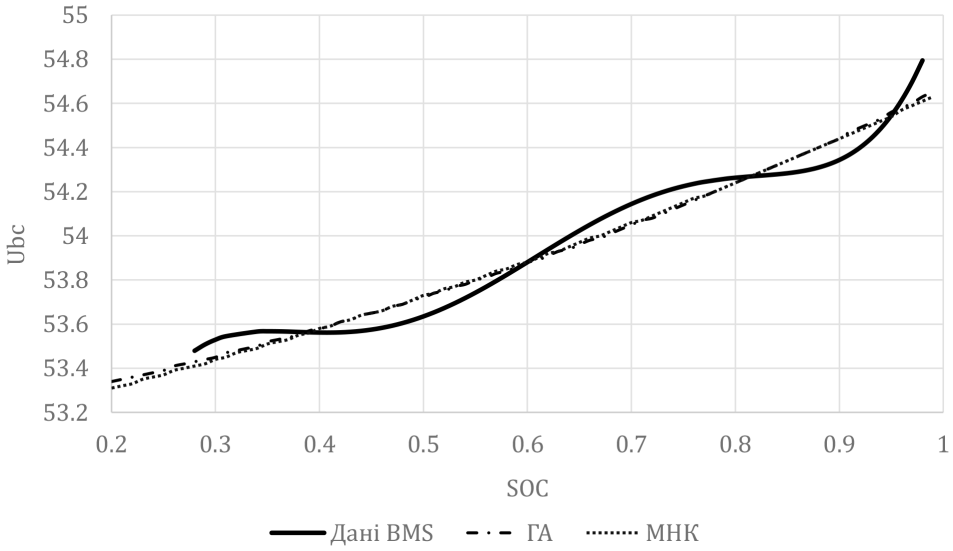


Рис. 2. Результати розрахунку параметрів акумулятора при заряджанні

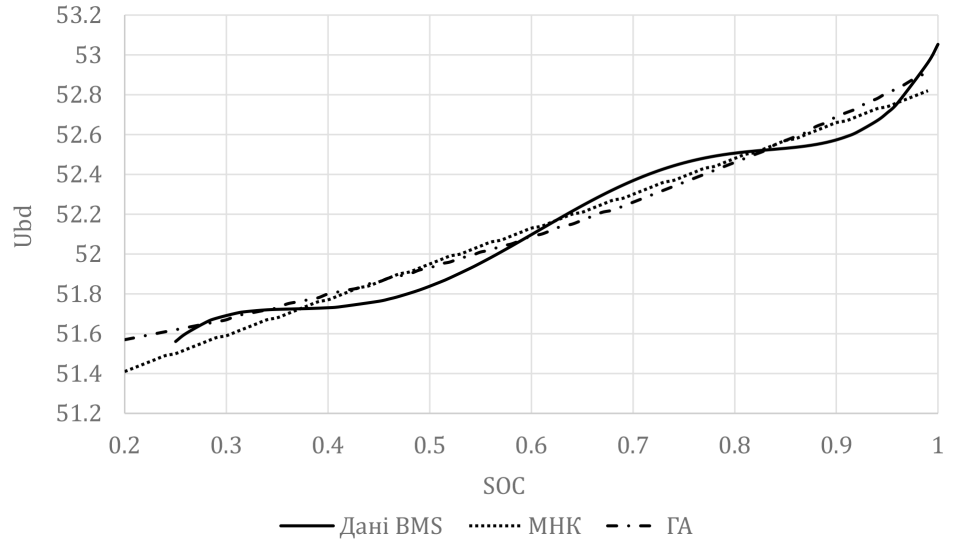


Рис. 3. Результати розрахунку параметрів акумулятора при розряджанні

Таблиця 2

Порівняльні характеристики точності моделювання заряду акумулятора $LiFePO_4$

Інтервал струму заряду (А)	Метрика / Метод	MSE	MAE	PFG	R^2	MAPE
1–3	ГА	0,0688	0,0988	0,1862	0,6052	0,0019
	МНК	0,0631	0,0861	0,1623	0,6378	0,0016
9–11	ГА	0,005	0,0589	0,1097	0,9644	0,0011
	МНК	0,005	0,0587	0,1094	0,9645	0,0011
19–21	ГА	0,0046	0,0596	0,1104	0,9605	0,0011
	МНК	0,0046	0,0593	0,11	0,9606	0,0011
29–31	ГА	0,003	0,0461	0,0851	0,9637	0,0008
	МНК	0,0035	0,0455	0,0839	0,9583	0,0008

Окрім того, метрики MAE та MAPE для обох методів залишаються в допустимих межах, що свідчить про високу узгодженість результатів моделювання із фактичними даними.

У випадку моделювання процесу розряду акумулятора метод найменших квадратів демонструє вищу точність у більшості інтервалів. Однак для інтервалу [1А; 3А] ГА демонструє вищу здатність до узагальнення, особливо за метриками MAE та MAPE, що свідчить про гнучкість у роботі з менш вираженими нелінійностями в поведінці батареї. У межах високого струму розряду [29А; 31А] обидва методи показали зниження точності моделювання.

Таблиця 3

Порівняльні характеристики точності моделювання розряду акумулятора $LiFePO_4$

Інтервал струму розряду (А)	Метрика / Метод	MSE	MAE	PFG	R^2	MAPE
1–3	ГА	0,0131	0,0801	0,152	0,8876	0,0015
	МНК	0,0074	0,0636	0,1207	0,9366	0,0012
9–11	ГА	0,0108	0,0716	0,1368	0,9347	0,0014
	МНК	0,0065	0,0574	0,1098	0,9607	0,0011
19–21	ГА	0,0056	0,0653	0,1251	0,9567	0,0012
	МНК	0,0048	0,0604	0,1157	0,9632	0,0012
29–31	ГА	0,0128	0,0994	0,191	0,853	0,0019
	МНК	0,0114	0,096	0,1845	0,8689	0,0018

Висновки

Розроблена комп'ютерна модель забезпечує можливість визначати напругу на клемах акумулятора, яка характеризує його енергетичний потенціал (залишковий заряд), на підставі параметрів акумулятора з використанням історичних даних про режими його роботи за різних умов. Застосування отриманих результатів забезпечує можливість адекватно відображати динамічні процеси заряду/розряду, аналіз таких взаємозалежних параметрів, як напруга на клемах та залишковий заряд акумулятора при різних струмах заряду/розряду.

Обґрунтоване подання структури акумулятора у вигляді еквівалентної електричної схеми з подальшим статистичним аналізом її параметрів забезпечують коректне застосування теорії регресійного аналізу та побудову цільової функції. Важливе значення при реалізації комп'ютерної моделі мали проведені процеси попереднього опрацювання даних, які включали: видалення аномальних значень, розподіл на групи, нормалізацію, подання часової мітки, статистичне опрацювання інтервалів.

Комплексність і універсальність моделі надають обґрунтовані до застосування методи найменших квадратів та генетичних алгоритмів при визначенні параметрів цільової функції. Порівняння змодельованих і фактичних параметрів з використанням статистичних показників підтверджує точність моделювання на високому рівні ($R^2 = 0{,}96$). При цьому метод найменших квадратів має кращі результати для середніх за інтенсивністю та стабільних режимів, як заряджання так і розряджання, завдяки простоті та стійкості до шумів. Генетичний алгоритм виявився ефективним при моделюванні в умовах високої нелінійності або нестачі даних, зокрема для граничних значень струму.

Застосування отриманих результатів забезпечує можливість адекватно відображати динамічні процеси заряджання/розряджання акумулятора, здійснювати аналіз таких взаємозалежних параметрів, як напруга на клемах та залишковий заряд акумулятора при різних струмах, досліджувати вплив параметрів акумулятора на режими роботи гібридних систем електропостачання.

Список використаної літератури

1. A. Seaman, T. Dao, J. McPhee. A survey of mathematics-based equivalent-circuit and electrochemical battery models for hybrid and electric vehicle simulation. *Journal of Power Sources*. Volume 256. 2014. Pages 410–423. DOI: https://doi.org/10.1016/j.jpowsour.2014.01.057

2. A. Fotouhi, D. J. Auger, K. Propp, S. Longo. A Study on Battery Model Parametrisation Problem – Application-Oriented Trade-offs between Accuracy and Simplicity. *IFAC-PapersOnLine*. Volume 49, Issue 11. 2016. Pages 48–53. DOI: https://doi.org/10.1016/j.ifacol.2016.08.008

3. N. Campagna, V. Castiglia, R. Miceli, R. A. Mastromauro, C. Spataro, M. Trapanese, F. Viola. Battery Models for Battery Powered Applications: A Comparative Study. *Energies*. Volume 13, Issue 16. 2020. Article 4085. DOI: https://doi.org/10.3390/en13164085

4. L. Zhang, H. Peng, Z. Ning, Z. Mu, C. Sun. Comparative Research on RC Equivalent Circuit Models for Lithium-Ion Batteries of Electric Vehicles. *Appl. Sci*. Volume 7, Issue 10. 2017. Article 1002. DOI: https://doi.org/10.3390/app7101002

5. K. Thirugnanam, H. Saini, P. Kumar. Mathematical Modeling of Li-Ion Battery for Charge/Discharge Rate and Capacity Fading Characteristics using Genetic Algorithm Approach. *Conference: Transportation Electrification Conference and Expo (ITEC), IEEE*. June 2012. Pages 1–6. DOI: https://doi.org/10.1109/ITEC.2012.6243431

6. O. Plakhtii, V. Nerubatskyi, A. Mashura, D. Hordiienko. The Analysis of Mathematical Models of Charge-Discharge Characteristics in Lithium-Ion Batteries. *2020 IEEE 40th International Conference on Electronics and Nanotechnology (ELNANO).* 2020. Pages 635–640. DOI: https://doi.org/10.1109/ELNANO50318.2020.9088827

7. T. Zahid, W. Li. A Comparative Study Based on the Least Square Parameter Identification Method for State of Charge Estimation of a LiFePO4 Battery Pack Using Three Model-Based Algorithms for Electric Vehicles. *Energies*. Volume 9, Issue 9. 2016. Article 720. DOI: https://doi.org/10.3390/en9090720

8. C. Zhang, W. Allafi, Q. Dinh, P. Ascencio, J. Marco. Online estimation of battery equivalent circuit model parameters and state of charge using decoupled least squares technique. *Energy*. Volume 142. 2018. Pages 678–688. DOI: https://doi.org/10.1016/j.energy.2017.10.043

9. V. Duong, H. A. Bastawrous, K. C. Lim, K. W. See, P. Zhang, S. X. Dou. Online state of charge and model parameters estimation of the LiFePO4 battery in electric vehicles using multiple adaptive forgetting factors recursive least-squares. *Journal of Power Sources*. Volume 296. 2015. Pages 215–224. DOI: https://doi.org/10.1016/j.jpowsour.2015.07.041

10. J. Zhengxin, S. Qin, W. Yujiang, W. Hanlin, G. Bingzhao, H. Lin. An Immune Genetic Extended Kalman Particle Filter approach on state of charge estimation for lithium-ion battery. *Energy*. Volume 230. 2021. Article 120805. DOI: https://doi.org/10.1016/j.energy.2021.120805

11. O. Pavliuk, M. Medykovskyy, T. Steclik. Predicting AGV Battery Cell Voltage Using a Neural Network Approach with Preliminary Data Analysis and Processing. *IEEE International Conference on Big Data (BigData)*. 2023. Pages 5087–5096. DOI: https://doi.org/10.1109/BigData59044.2023.10386137.

12. DEYE. BOS-G-H-L Product Information. 2025. Available at: https://deye.com/product/bos-g-h-l/. [Accessed: 19 August 2025].

References

1. Seaman, A., Dao, T., McPhee, J. (2014) A survey of mathematics-based equivalent-circuit and electrochemical battery models for hybrid and electric vehicle simulation. *Journal of Power Sources*, vol. 256, pp. 410-423. Available at: https://doi.org/10.1016/j.jpowsour.2014.01.057

2. Fotouhi, A., Auger, D.J., Propp, K., Longo, S. (2016) A Study on Battery Model Parametrisation Problem – Application-Oriented Trade-offs between Accuracy and Simplicity. *IFAC-PapersOnLine*, vol. 49, no. 11, pp. 48–53. Available at: https://doi.org/10.1016/j.ifacol.2016.08.008

3. Campagna, N., Castiglia, V., Miceli, R., Mastromauro, R.A., Spataro, C., Trapanese, M., Viola, F. (2020) Battery Models for Battery Powered Applications: A Comparative Study. *Energies*, vol. 13, no. 16, 4085. Available at: https://doi.org/10.3390/en13164085

4. Zhang, L., Peng, H., Ning, Z., Mu, Z., Sun, C. (2017) Comparative Research on RC Equivalent Circuit Models for Lithium-Ion Batteries of Electric Vehicles. *Applied Sciences*, vol. 7, no. 10, 1002. Available at: https://doi.org/10.3390/app7101002

5. Thirugnanam, K., Saini, H., Kumar, P. (2012) Mathematical Modeling of Li-Ion Battery for Charge/Discharge Rate and Capacity Fading Characteristics using Genetic Algorithm Approach. *2012 IEEE Transportation Electrification Conference and Expo (ITEC)*, June 2012, pp. 1–6. Available at: https://doi.org/10.1109/ITEC.2012.6243431

6. Plakhtii, O., Nerubatskyi, V., Mashura, A., Hordiienko, D. (2020) The Analysis of Mathematical Models of Charge-Discharge Characteristics in Lithium-Ion Batteries. *2020 IEEE 40th International Conference on Electronics and Nanotechnology (ELNANO)*, pp. 635–640. Available at: https://doi.org/10.1109/ELNANO50318.2020.9088827

7. Zahid, T., Li, W. (2016) A Comparative Study Based on the Least Square Parameter Identification Method for State of Charge Estimation of a LiFePO4 Battery Pack Using Three Model-Based Algorithms for Electric Vehicles. *Energies*, vol. 9, no. 9, 720. Available at: https://doi.org/10.3390/en9090720

8. Zhang, C., Allafi, W., Dinh, Q., Ascencio, P., Marco, J. (2018) Online estimation of battery equivalent circuit model parameters and state of charge using decoupled least squares technique. *Energy*, vol. 142, pp. 678-688. Available at: https://doi.org/10.1016/j.energy.2017.10.043

9. Duong, V., Bastawrous, H. A., Lim, K. C., See, K. W., Zhang, P., Dou, S. X. (2015) Online state of charge and model parameters estimation of the LiFePO4 battery in electric vehicles using multiple adaptive forgetting factors recursive least-squares. *Journal of Power Sources*, vol. 296, pp. 215-224. Available at: https://doi.org/10.1016/j.jpowsour.2015.07.041

10. Zhengxin, J., Qin, S., Yujiang, W., Hanlin, W., Bingzhao, G., Lin, H. (2021) An Immune Genetic Extended Kalman Particle Filter approach on state of charge estimation for lithium-ion battery. *Energy*, vol. 230, 120805. Available at: https://doi.org/10.1016/j.energy.2021.120805

11. Pavliuk, O., Medykovskyy, M., Steclik, T. (2023) Predicting AGV Battery Cell Voltage Using a Neural Network Approach with Preliminary Data Analysis and Processing. *IEEE International Conference on Big Data (BigData)*, pp. 5087–5096. Available at: https://doi.org/10.1109/BigData59044.2023.10386137

12. DEYE. (2025). BOS-G-H-L Product Information. Available at: https://deye.com/product/bos-g-h-l/ [Accessed: 19 August 2025].

Дата першого надходження рукопису до видання: 25.09.2025
Дата прийнятого до друку рукопису після рецензування: 20.10.2025
Дата публікації:

УДК 004.8 DOI

М. О. МОЛЧАНОВА
Ph.D., старший викладач кафедри комп'ютерних наук
Хмельницький національний університет
ORCID: 0000-0001-9810-936X

О. В. МАЗУРЕЦЬ
кандидат технічних наук,
доцент кафедри комп'ютерних наук
Хмельницький національний університет
ORCID: 0000-0002-8900-0650

О. О. ЗАЛУЦЬКА
асистент кафедри комп'ютерних наук
Хмельницький національний університет
ORCID: 0000-0003-1242-3548

В. Д. КАДИНСЬКА
студентка кафедри комп'ютерних наук
Хмельницький національний університет

В. В. МАСЛОВСЬКА
студентка кафедри комп'ютерних наук
Хмельницький національний університет

ІНФОРМАЦІЙНА ХМАРНА ТЕХНОЛОГІЯ НЕЙРОМЕРЕЖЕВОГО АНАЛІЗУ ЗРУЙНОВАНИХ СПОРУД ЗА ВІЗУАЛЬНИМИ ДАНИМИ З БПЛА

У роботі запропоновано інформаційну хмарну технологію нейромережевого аналізу зруйнованих споруд за візуальними даними, отриманими з безпілотних літальних апаратів. Проблема швидкої та достовірної ідентифікації залишків будівель є надзвичайно актуальною у сучасних умовах, оскільки результати такого аналізу мають критичне значення для моніторингу наслідків бойових дій, ліквідації надзвичайних ситуацій і планування відновлювальних робіт. Традиційні методи комп'ютерного зору не забезпечують необхідної ефективності через масштабність даних, різноманітність об'єктів та складність сцен. Застосування глибинних нейронних мереж, інтегрованих у хмарне середовище, дозволяє подолати ці обмеження, забезпечивши масштабованість та високу точність аналізу.

Розроблена технологія передбачає надходження у хмарне середовище фотознімків високої роздільної здатності, їх попередню обробку та формування єдиного тренувального й валідаційного датасетів. У хмарі здійснюється навчання моделей YOLOv12 для детекції та сегментації потенційних будівельних залишків, а також ансамблю з десяти моделей ResNet50 для багатокласової класифікації виділених фрагментів. Клієнтська частина системи забезпечує препроцесинг нових знімків і виконання нейромережевого аналізу на основі попередньо навчених моделей. У результаті автоматично визначаються позиції залишків та їх віднесення до конкретних категорій матеріалів, що включають цеглу, бетон, деревину, пінопласт, плитку, пластик та інші.

Експериментальні результати підтверджують ефективність розробленого рішення. Досягнуто загальної точності класифікації (Accuracy) 0.965, при цьому значення Precision склало 0.889, Recall – 0.965, а F1-міра – 0.925. Отримані результати перевищують показники аналогічних досліджень щонайменше на 0.042, що демонструє конкурентоспроможність та доцільність використання запропонованого підходу. Використання хмарних технологій дозволяє забезпечити масштабованість рішення, зниження вартості інфраструктури та незалежність від обчислювальних можливостей локальних пристроїв.

Ключові слова: *хмарні технології, нейромережі, БПЛА, зруйновані споруди.*

© Молчанова М. О., Мазурець О. В., Залуцька О. О., Кадинська В. Д., Масловська В. В., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

M. O. MOLCHANOVA
Ph.D., Senior Lecturer at the Computer Science Department
Khmelnytskyi National University
ORCID: 0000-0001-9810-936X

O. V. MAZURETS
Ph.D. in Engineering Science,
Associate Professor at the Computer Science Department
Khmelnytskyi National University
ORCID: 0000-0002-8900-0650

O. O. ZALUTSKA
Assistant Professor at the Computer Science Department
Khmelnytskyi National University
ORCID: 0000-0003-1242-3548

V. D. KADYNSKA
Bachelor Student at the Computer Science Department
Khmelnytskyi National University

V. V. MASLOVSKA
Bachelor Student at the Computer Science Department
Khmelnytskyi National University

CLOUD INFORMATION TECHNOLOGY FOR NEURAL NETWORK ANALYSIS OF DESTROYED STRUCTURES USING VISUAL DATA FROM UAVS

The paper proposes an information cloud technology for neural network analysis of destroyed structures based on visual data obtained from unmanned aerial vehicles. The problem of fast and reliable identification of building remains is extremely relevant in modern conditions, since the results of such analysis are of critical importance for monitoring the consequences of hostilities, eliminating emergencies and planning restoration work. Traditional computer vision methods do not provide the necessary efficiency due to the scale of data, the variety of objects and the complexity of scenes. The use of deep neural networks integrated into the cloud environment allows overcoming these limitations, ensuring scalability and high accuracy of analysis. The developed technology involves the receipt of high-resolution photographs into the cloud environment, their pre-processing and the formation of a single training and validation dataset. In the cloud, YOLOv12 models are trained for detection and segmentation of potential building remains, as well as an ensemble of ten ResNet50 models for multi-class classification of selected fragments. The client part of the system provides preprocessing of new images and execution of neural network analysis based on previously trained models. As a result, the positions of residues are automatically determined and their assignment to specific categories of materials, including brick, concrete, wood, foam plastic, tiles, plastic and others.

Experimental results confirm the effectiveness of the developed solution. The overall classification Accuracy of 0.965 was achieved, with the Precision value being 0.889, Recall – 0.965, and F1-measure – 0.925. The obtained results exceed the indicators of similar studies by at least 0.042, which demonstrates the competitiveness and feasibility of using the proposed approach. The use of cloud technologies allows for scalability of the solution, reduction of infrastructure costs and independence from the computing capabilities of local devices.

Key words: *cloud technologies, neural networks, UAVs, destroyed buildings.*

Постановка проблеми

У сучасних умовах зростає потреба у швидкому та достовірному виявленні та класифікації залишків зруйнованих споруд [1], що мають критичне значення для моніторингу наслідків бойових дій, ліквідації надзвичайних ситуацій та планування відновлювальних робіт [2, 3]. Традиційні підходи, засновані на експертному аналізі фотознімків або використанні класичних методів комп'ютерного зору, виявляються недостатньо ефективними через високу складність об'єктів, різноманітність типів руйнувань, а також значний обсяг даних, що потребують обробки [4, 5].

Застосування безпілотних літальних апаратів суттєво розширило можливості збору візуальної інформації про руйнування [6]. Завдяки мобільності та високій роздільній здатності сенсорів БПЛА забезпечують доступ до великої кількості зображень у реальному часі [7, 8]. Однак такий підхід породжує нові виклики, пов'язані з масштабністю даних, необхідністю їх швидкої обробки та забезпеченням високої точності класифікації [9].

Ефективним інструментом розв'язання зазначених проблем стали методи глибинного навчання [10], зокрема згорткові та трансформерні нейронні мережі, які демонструють високу результативність у задачах аналізу

зображень [11, 12]. Проте навчання таких моделей потребує значних обчислювальних ресурсів, що унеможливлює їх використання на локальних пристроях із обмеженою продуктивністю [13, 14].

У цьому контексті доцільним є використання хмарних обчислень, які забезпечують масштабованість, високу обчислювальну потужність та гнучкість у розподілі ресурсів [15]. Хмарні платформи надають можливість паралельної обробки великих масивів зображень, організації розподіленого навчання моделей та зниження вартості інфраструктури завдяки підходу «обчислення як послуга» [16, 17]. Така інтеграція глибинного навчання та хмарних технологій створює передумови для побудови ефективних методів класифікації залишків зруйнованих споруд за візуальними даними з БПЛА.

Аналіз останніх досліджень і публікацій

У дослідженні M. I. B. Ahmed та співавторів [18] розглянуто застосування глибинного навчання для ідентифікації та класифікації відходів з використанням відкритого набору зображень, зібраних у регіоні Близького Сходу. Матеріали корпусу охоплювали приклади п'яти основних категорій сміття, що забезпечувало можливість моделювання задачі багатокласової класифікації. Для розв'язання цієї задачі було використано як спеціально розроблену згорткову нейронну мережу, так і низку попередньо навчених моделей, серед яких ResNet50V2, MobileNetV2 та DenseNet169. Найвищих результатів продемонструвала архітектура ResNet50V2, яка перевищила інші підходи за точністю розпізнавання, тоді як власна CNN-модель авторів також показала конкурентні результати, випередивши попередні напрацювання на цьому ж наборі даних.

У роботі K. Lin та колег [19] увага зосереджена на створенні методу високої ефективності для детальної класифікації будівельних і демонтажних відходів, що має суттєве значення у контексті розбудови системи переробки матеріалів та зниження вуглецевого сліду в будівельній галузі. Запропоновані авторами моделі ґрунтувалися на архітектурі ResNet з використанням перенесення знань і спеціальних стратегій налаштування навчання, що дозволило оптимізувати баланс між швидкістю та якістю роботи нейронних мереж. Результати експериментів продемонстрували, що застосування такого підходу підвищує точність класифікації та сприяє кращому розмежуванню типів відходів, що підтверджує перспективність методів глибинного навчання у сфері сталого управління будівельними ресурсами.

У статті [20] представлено методологію підвищення ефективності сортування будівельних та демонтажних відходів із використанням методів машинного навчання та RGB-камери для розпізнавання фрагментів відходів. Основна мета дослідження полягає у вдосконаленні процесу класифікації шляхом застосування розширених методів вилучення ознак, що забезпечує підвищення швидкості та точності розпізнавання. Автори порівнюють три класифікатори: згорткову нейронну мережу, градієнтний бустинг дерев рішень та багатошаровий перцептрон. Експерименти показали, що використання методу вилучення ознак для GB та MLP дало кращі результати за швидкістю та точністю, ніж традиційний CNN. Запропонована методологія дозволила досягти точності класифікації до 92.3 % порівняно з 85.9 % для CNN. Додатково у роботі надано датасети, коди та моделі, що підвищує прозорість і відтворюваність отриманих результатів.

Формулювання мети дослідження

Мета роботи полягає у розробленні інформаційної хмарної технології нейромережевого аналізу зруйнованих споруд за візуальними даними з БПЛА для класифікації залишків зруйнованих споруд із застосуванням технологій глибинного навчання

Викладення основного матеріалу дослідження

Запропонована інформаційна хмарна технологія (рис. 1) нейромережевого аналізу зруйнованих споруд ґрунтується на застосуванні методів глибинного навчання до візуальних даних, отриманих з БПЛА.

На вхід надходять фотознімки високої роздільної здатності, які передаються у хмарне середовище для попередньої обробки та формування тренувального та валідаційного датасетів. У хмарі здійснюється навчання глибинних нейронних мереж: YOLOv12 [21] для виявлення та сегментації потенційних фрагментів будівельних залишків, а також ансамблю з десяти моделей ResNet50 [22] для багатокласової класифікації відібраних залишків зруйнованих споруд (цегла, бетон, пінопласт, гіпсокартон, трубка, пластик, каміння, плитка, деревина та інші матеріали) [23].

Клієнтська частина технології виконує препроцесинг нових знімків, отриманих з БПЛА (масштабування, нормалізацію) та нейромережевий аналіз на основі навчених моделей. Результатом роботи системи є автоматизоване визначення позицій і класифікація залишків споруд за категоріями з використанням ймовірнісних оцінок, що забезпечує високу точність аналізу та можливість оперативного моніторингу територій після руйнувань.

На рис. 2 представлено приклад дослідження реалізованої інформаційної технології.

Отримані в ході навчання результати свідчать про ефективність запропонованої інформаційної хмарної технології нейромережевого аналізу зруйнованих споруд. Загальна точність (Accuracy) класифікації склала 0.965, що підтверджує надійність роботи системи на різних типах вхідних даних. Значення Precision дорівнює 0.889, що вказує на здатність моделі мінімізувати кількість хибнопозитивних визначень, тоді як Recall на рівні 0.965 демонструє високу повноту класифікації. Досягнуте значення F_1-міри, яке становить 0.925, підтверджує збалансованість

Рис. 1. Інформаційна хмарна технологія нейромережевого аналізу зруйнованих споруд

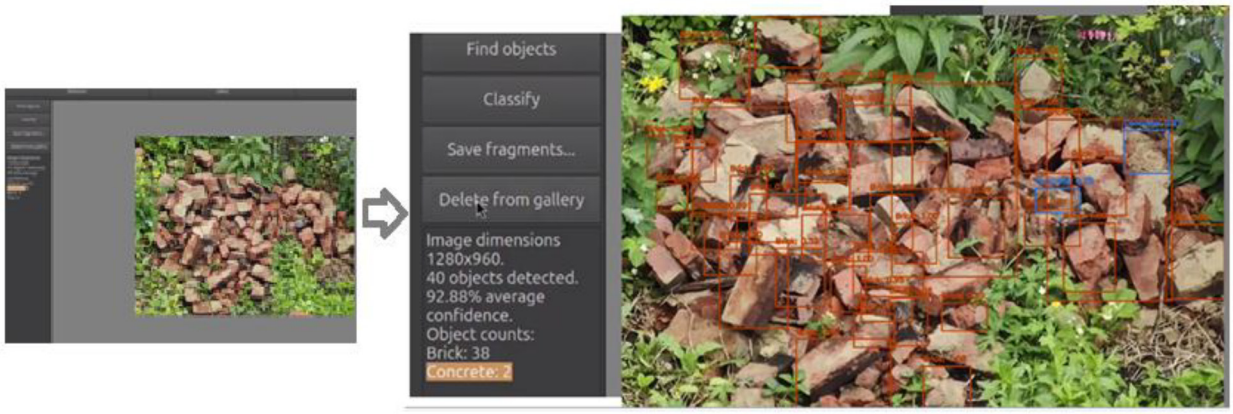


Рис. 2. Приклад використання розробленої інформаційної технології

між точністю та повнотою та свідчить про стабільність отриманих результатів. Таким чином, запропонований підхід забезпечує достатньо високу якість класифікації будівельних залишків і може застосовуватися для практичних завдань аналізу територій після руйнувань.

Порівняння з відомими аналогами дозволяє зробити висновки, що розроблена інформаційна технологія є ефективною та дозволяє підвищити точність щонайменше на 0.042 (0.923 у аналогічних дослідженнях проти 0.965, отриманих авторами). Використання хмарних технологій дозволяє зробити рішення масштабованим та незалежним від архітектури.

Висновки

У результаті проведеного дослідження було розроблено метод хмароорієнтованої обробки візуальних даних з БПЛА, спрямований на автоматизовану ідентифікацію та класифікацію залишків зруйнованих споруд. Запропонований підхід поєднує сучасні архітектури глибинних нейронних мереж, зокрема YOLOv12 та ResNet50, із можливостями хмарних обчислень, що забезпечує високу продуктивність та масштабованість процесу аналізу. Експериментальна перевірка продемонструвала досягнення точності класифікації на рівні 0.965, при цьому значення Precision становить 0.889, Recall – 0.965, а F_1-міра – 0.925, що свідчить про збалансованість моделі та її здатність ефективно розпізнавати об’єкти різних класів. Отримані результати перевищують показники відомих аналогів, що підтверджує доцільність застосування розробленої методології у практичних системах моніторингу пошкодженої інфраструктури та створює передумови для подальшого вдосконалення технологій автоматизованого аналізу аерофотознімків у сфері відновлення територій.

Список використаної літератури

1. Kohns J., et al. Building damage assessment in natural disasters: A trans- and interdisciplinary approach combining domain knowledge, 3D machine learning, and crowdsourcing // Progress in Disaster Science. 2025. Vol. 26. P. 100427. https://doi.org/10.1016/j.pdisas.2025.100427
2. Hladun O., Mazurets O., Molchanova M., Sobko O. Real Time Detection the Person Emotion State Using Neural Network // Scientific Research: Modern Innovations and Future Perspectives : proceedings of the 2nd International scientific and practical conference (November 25–27, 2024, Montreal, Canada). Montreal, 2024. P. 119–123.
3. Mostofi S., Altunişik A. C., Başağa H. B., Okur F. Y., Yilmaz Z., Hadinata P., … Sezdirmez T. A Big Data-Enabled Decision Support Model for Post-Earthquake Damage Classification of RC Buildings: A Case Study on February 6, Kahramanmaraş Doublet Earthquakes // Journal of Earthquake Engineering. 2025. P. 1–23.
4. Hladun O., Zalutska O., Klimenko V., Mazurets O. Research on the effectiveness of classifying the remains of destroyed buildings using MobileNetV3 neural network architecture // Innovations in Science: From Theoretical Foundations to Practical Impact : proceedings of the 1st International Scientific and Practical Conference (May 12–14, 2025, Antwerp, Belgium). Antwerp, 2025. P. 158–162.
5. Didur V., Molchanova M., Mazurets O. Research on the effectiveness of neural network detection of plots with the destroyed buildings remains // Modern technologies and science: problems, new and relevant developments : proceedings of the XXI International Scientific and Practical Conference (May 26, 2025, Zaragoza, Spain). Zaragoza, 2025. P. 245–251.
6. Li T., Fei L. Exploring obstacles to the use of unmanned aerial vehicles in emergency rescue: A BWM-DEMATEL approach // Technology in Society. 2025. Vol. 81. Article 102863. https://doi.org/10.1016/j.techsoc.2025.102863
7. Alqudsi Y., Makaraci M. UAV swarms: research, challenges, and future directions // Journal of Engineering and Applied Science. 2025. Vol. 72, No. 1. P. 12. URL: https://doi.org/10.1186/s44147-025-00582-3
8. Dydo R., Sobko O., Klimenko V., Mazurets O. Datalogic structure for intelligent system for areas localization in photos with the remains of buildings using neural network // Modern Scientific Research: Theoretical and Practical Aspects : proceedings of the II International Scientific and Practical Conference (May 26–28, 2025, Riga, Latvia). Riga, 2025. P. 123–127.
9. Didur V. O., Molchanova M. O., Tyschenko O. O., Mazurets O. V. Approach for comparative analysis of effectiveness of using MobileNetV3 and ViT neural network models for graphical localization of destroyed buildings remains areas // Formation of Innovative Potential of World Science : proceedings of the IX International Scientific and Practical Conference (May 16, 2025, Waterford, Republic of Ireland). Waterford, 2025. P. 94–97.
10. Мазурець О. В., Петровський С. С., Дидо Р. А. Нейромережева модель для ідентифікації особистості за зображенням обличчя у реальному часі // Інформаційні технології і автоматизація : матеріали XVII міжнародної науково-практичної конференції (31 жовтня – 1 листопада 2024, Одеса, ОНТУ). Одеса, 2024. С. 655–658.
11. Бас І. С., Мазурець О. В., Молчанова М. О., Собко О. В. Дослідження ефективності методу автоматизованого визначення типу літального апарату за фотографічним зображенням // Збірник наукових праць за матеріалами XVI Всеукраїнської науково-практичної конференції «Актуальні проблеми комп'ютерних наук АПКН-2024» (15–16 листопада 2024, Хмельницький). Хмельницький, 2024. С. 29–34.
12. Dydo R., Sobko O., Molchanova M., Mazurets O. Analysis of precision of finding the destroyed remains buildings on photos using MobileNetV3 and ViT neural networks // Science and Technology: New Horizons of Development : proceedings of the I International Scientific and Practical Conference (May 14–16, 2025, Prague, Czech Republic). Prague, 2025. P. 208–214.
13. Malaydakh V., Molchanova M., Shevchuk P., Mazurets O. Deep learning neural network architecture for determining sunflower growth stage from visual data // Modern Scientific Research: Theoretical and Practical Aspects : proceedings of the II International Scientific and Practical Conference (May 26–28, 2025, Riga, Latvia). Riga, 2025. P. 143–148.
14. Bas I. S., Kadynska V. D., Klimenko V. I., Mazurets O. V. Convolutional Neural Network Transfer Learning Method for Aircraft Image Classification // Scientific method: reality and future trends of researching : proceedings of the VI International Scientific and Theoretical Conference (June 6, 2025, Montreal, Canada). Montreal, 2025. P. 147–155.
15. Молчанова М. О., Мазурець О. В., Собко О. В., Кліменко В. І., Андрощук В. І. Метод нейромережевого виявлення кібербулінгу з використанням хмарних сервісів та об'єктно-орієнтованої моделі // Науковий журнал «Вісник Хмельницького національного університету», серія: Технічні науки. Хмельницький, 2024. № 2 (333). С. 200–206. https://doi.org/10.31891/2307-5732-2024-333-2
16. Tymofiiev I., Mazurets O., Hardysh D., Molchanova M. Neural Network Dual Architecture for Depression Detection Using Cloud Services // Scientific Research in the Era of Digital Technologies: Challenges and Opportunities : proceedings of the XLVI International Scientific and Practical Conference (November 6–8, 2024, Barcelona, Spain). 2024. P. 84–88.
17. Yurchenko D., Mazurets O., Didur V., Molchanova M. Approach to Using Cloud Services for Visual Analytics of Neural Network Analysis of Texts Emotional Tonality // The Future of Scientific Discoveries: New Trends and

Technologies : proceedings of the XLVII International Scientific and Practical Conference (November 13–15, 2024, Marseille, France). 2024. P. 108–113.

18. Ahmed M. I. B., et al. Deep Learning Approach to Recyclable Products Classification: Towards Sustainable Waste Management // Sustainability. 2023. Vol. 15, No. 14. P. 11138. https://doi.org/10.3390/su151411138.

19. Lin K., et al. Applying machine learning to fine classify construction and demolition waste based on deep residual network and knowledge transfer // Environment, Development and Sustainability. 2022. https://doi.org/10.1007/s10668-022-02740-6

20. Nežerka V., Zbíral T., Trejbal J. Machine-learning-assisted classification of construction and demolition waste fragments using computer vision: Convolution versus extraction of selected features // Expert Systems with Applications. 2024. Vol. 238. Article 121568. https://doi.org/10.1016/j.eswa.2023.121568.

21. YOLOv12: State-of-the-art object detection model [Електронний ресурс] // YOLOv12. 2025. Режим доступу: https://yolov12.com/

22. ResNet50: Convolutional Neural Network Model [Електронний ресурс] // PyTorch. 2025. Режим доступу: https://pytorch.org/hub/pytorch_vision_resnet/

23. Hladun O. V., Molchanova M. O., Zalutska O. O., Mazurets O. V. Effectiveness research of using ViT neural network architecture for classifying the destroyed buildings remains // Achievements of Science and Applied Research : proceedings of the 2nd International Scientific and Theoretical Conference (May 19–21, 2025, Dublin, Ireland). Dublin, 2025. P. 96–100.

References

1. Kohns, J., Zahs, V., Klonner, C., Höfle, B., Stempniewski, L., & Stark, A. (2025). Building damage assessment in natural disasters: A trans-and interdisciplinary approach combining domain knowledge, 3D machine learning, and crowdsourcing. Progress in Disaster Science, 100427. https://doi.org/10.1016/j.pdisas.2025.100427

2. Hladun, O., Mazurets, O., Molchanova, M., & Sobko, O. (2024). Real time detection the person emotion state using neural network. In Scientific Research: Modern Innovations and Future Perspectives (pp. 119–123). Proceedings of the 2nd International Scientific and Practical Conference, Montreal, Canada.

3. Mostofi, S., Altunişik, A. C., Başağa, H. B., Okur, F. Y., Yilmaz, Z., Hadinata, P.,... & Sezdirmez, T. (2025). A Big Data-Enabled Decision Support Model for Post-Earthquake Damage Classification of RC Buildings: A Case Study on February 6, Kahramanmaraş Doublet Earthquakes. Journal of Earthquake Engineering, 1-23. https://doi.org/10.1080/13632469.2025.2505974

4. Hladun, O., Zalutska, O., Klimenko, V., & Mazurets, O. (2025). Research on the effectiveness of classifying the remains of destroyed buildings using MobileNetV3 neural network architecture. In Innovations in Science: From Theoretical Foundations to Practical Impact (pp. 158–162). Proceedings of the 1st International Scientific and Practical Conference, Antwerp, Belgium.

5. Didur, V., Molchanova, M., & Mazurets, O. (2025). Research on the effectiveness of neural network detection of plots with the destroyed buildings remains. In Modern technologies and science: Problems, new and relevant developments (pp. 245–251). Proceedings of the XXI International Scientific and Practical Conference, Zaragoza, Spain.

6. Li, T., & Fei, L. (2025). Exploring obstacles to the use of unmanned aerial vehicles in emergency rescue: A BWM-DEMATEL approach. Technology in Society, 81, 102863. https://doi.org/10.1016/j.techsoc.2025.102863

7. Alqudsi, Y., & Makaraci, M. (2025). UAV swarms: Research, challenges, and future directions. Journal of Engineering and Applied Science, 72(1), 12. https://doi.org/10.1186/s44147-025-00582-3

8. Dydo, R., Sobko, O., Klimenko, V., & Mazurets, O. (2025). Datalogic structure for intelligent system for areas localization in photos with the remains of buildings using neural network. In Modern Scientific Research: Theoretical and Practical Aspects (pp. 123–127). Proceedings of the II International Scientific and Practical Conference, Riga, Latvia.

9. Didur, V. O., Molchanova, M. O., Tyschenko, O. O., & Mazurets, O. V. (2025). Approach for comparative analysis of effectiveness of using MobileNetV3 and ViT neural network models for graphical localization of destroyed buildings remains areas. In Formation of Innovative Potential of World Science (pp. 94–97). Proceedings of the IX International Scientific and Practical Conference, Waterford, Republic of Ireland.

10. Mazurets O. V., Petrovskii S. S., Dydo R. A. (2024) Neiromerezheva model dlia identyfikatsii osibystosti za zobrazhennia oblychia u realnomu chasi [Neural network model for identification of a person by facial image in real time]. Informatsiini tekhnolohii i avtomatyzatsiia: materialy XVII mizhnarodnoi naukovo-praktychnoi konferentsii, pp. 655–658. Odesa, ONTU.

11. Bas I. S., Mazurets O. V., Molchanova M. O., Sobko O. V. (2024) Doslidzhennia efektyvnosti metodu avtomatyzovanoho vyznachennia typu litaltnoho aparatu za fotohrafichnym zobrazhennia [Study on the effectiveness of the method for automated determination of aircraft type using photographic image]. Zbirnyk naukovykh prats za materialamy XVI Vseukrainskoi naukovo-praktychnoi konferentsii «Aktualni problemy kompiuternykh nauk APKN-2024», pp. 29–34. Khmelnytskyi.

12. Dydo, R., Sobko, O., Molchanova, M., & Mazurets, O. (2025). Analysis of precision of finding the destroyed remains buildings on photos using MobileNetV3 and ViT neural networks. In Science and Technology: New Horizons of Development (pp. 208–214). Proceedings of the I International Scientific and Practical Conference, Prague, Czech Republic.

13. Malaydakh, V., Molchanova, M., Shevchuk, P., & Mazurets, O. (2025). Deep learning neural network architecture for determining sunflower growth stage from visual data. In Modern Scientific Research: Theoretical and Practical Aspects (pp. 143–148). Proceedings of the II International Scientific and Practical Conference, Riga, Latvia.

14. Bas, I. S., Kadynska, V. D., Klimenko, V. I., & Mazurets, O. V. (2025). Convolutional neural network transfer learning method for aircraft image classification. In Scientific method: Reality and future trends of researching (pp. 147–155). Proceedings of the VI International Scientific and Theoretical Conference, Montreal, Canada.

15. Molchanova M. O., Mazurets O. V., Sobko O. V., Klimenko V. I., Androshchuk V. I. (2024) Metod neiromerezhevoho vyvchennia kiberbulingu z vykorystanniam khmarnykh servisiv ta obiektno-orientovanoi modeli [Neural network method for cyberbullying detection using cloud services and an object-oriented model]. Visnyk Khmelnytskoho natsionalnoho universytetu. Tekhnichni nauky, no. 2 (333), pp. 200–206. https://doi.org/10.31891/2307-5732-2024-333-2

16. Tymofiiev, I., Mazurets, O., Hardysh, D., & Molchanova, M. (2024). Neural network dual architecture for depression detection using cloud services. In Scientific Research in the Era of Digital Technologies: Challenges and Opportunities (pp. 84–88). Proceedings of the XLVI International Scientific and Practical Conference, Barcelona, Spain.

17. Yurchenko, D., Mazurets, O., Didur, V., & Molchanova, M. (2024). Approach to using cloud services for visual analytics of neural network analysis of texts emotional tonality. In The Future of Scientific Discoveries: New Trends and Technologies (pp. 108–113). Proceedings of the XLVII International Scientific and Practical Conference, Marseille, France.

18. Ahmed, M. I. B., et al. (2023). Deep learning approach to recyclable products classification: Towards sustainable waste management. Sustainability, 15(14), 11138. https://doi.org/10.3390/su151411138

19. Lin, K., et al. (2022). Applying machine learning to fine classify construction and demolition waste based on deep residual network and knowledge transfer. Environment, Development and Sustainability. https://doi.org/10.1007/s10668-022-02740-6

20. Nežerka, V., Zbíral, T., & Trejbal, J. (2024). Machine-learning-assisted classification of construction and demolition waste fragments using computer vision: Convolution versus extraction of selected features. Expert Systems with Applications, 238, 121568. https://doi.org/10.1016/j.eswa.2023.121568

21. YOLOv12. (2025). YOLOv12: State-of-the-art object detection model. https://yolov12.com/

22. PyTorch. (2025). ResNet50: Convolutional neural network model. https://pytorch.org/hub/pytorch_vision_resnet/

23. Hladun, O. V., Molchanova, M. O., Zalutska, O. O., & Mazurets, O. V. (2025). Effectiveness research of using ViT neural network architecture for classifying the destroyed buildings remains. In Achievements of Science and Applied Research (pp. 96–100). Proceedings of the 2nd International Scientific and Theoretical Conference, Dublin, Ireland.

Дата першого надходження рукопису до видання: 24.09.2025
Дата прийнятого до друку рукопису після рецензування: 17.10.2025
Дата публікації:

УДК 004.421 004.5 DOI

В. К. ОВСЯК
доктор технічних наук,
професор кафедри комп'ютерних технологій
у видавничо-поліграфічних процесах
Національний університет «Львівська політехніка»
ORCID: 0000-0001-9295-284X

В. Р. ТУРЧАК
аспірант кафедри комп'ютерних технологій
у видавничо-поліграфічних процесах
Національний університет «Львівська політехніка»
ORCID: 0009-0009-2907-1869

МОДЕЛЬ АЛГОРИТМУ ВЗАЄМОДІЇ КОРИСТУВАЧА З БАЗОЮ СТРУКТУРОВАНИХ ДАНИХ ДЛЯ АДАПТИВНОГО СОРТУВАННЯ У ВЕБІНТЕРФЕЙСІ

Сортування табличних даних у веб-середовищі потребує рішень, що поєднують практичну інтерактивність з математичною визначеністю. У статті запропоновано методику та формалізовану модель клієнтського інтерфейсу, який імпортує Excel/CSV, виконує очищення й нормалізацію, підтримує фільтри й багатокритеріальне впорядкування без обов'язкової серверної логіки. Реалізовано два робочі режими: автономний (усі операції виконуються у браузері, локальне зберігання) та інтеграційний (опційний REST-експорт у MySQL/MongoDB/SQLite). Інтерфейс орієнтовано на покрокову взаємодію: вікна завантаження, параметрів очищення, вибору критеріїв/пріоритетів, налаштувань фільтрів і часткового збігу, запуску сортування та перегляду результатів; окремо передбачено логування дій і підказки щодо конфліктних умов. Технічно використано HTML, CSS, JavaScript, IndexedDB/sql.js та бібліотеку XLSX, що забезпечує доступність і простоту розгортання у локальних середовищах.

Ідейне ядро складає трикомпонентна архітектура даних: T_1 (вхідна матриця $m \times n$), T_2 (набір четвірок (c_i, w_i, d_i, f_i) для формалізації критеріїв і ваг) та T_3 (результат після фільтрації/проєкції/сортування). Кожний етап має чіткі оператори: $T_1 \to f_1 \to T_1'$ (очищення/нормалізація/валідація), $T_1' \to f_2 \to T_2$ (побудова запиту), $T_1', T_2 \to f_3 \to T_3$ (застосування інструкції). Зіставлення текстових полів підтримує нечіткість (зокрема, Левенштейн), що підвищує стійкість до варіантів написання. Для багатокритеріального порівняння використано $Score(r) = \sum_k w_k \cdot \varphi_k(r, c_k, f_k)$ *або лексикографічний ключ, де φ_k – нормалізована функція порівняння з урахуванням фільтра.*

Сортування реалізовано модифікованим Selection Sort із динамічно-наслідковим обмеженням переміщення $R = m/(i + 1)$, що на кроці і лімітує інтенсивність перестановок ($\leq \{R\}$) і запобігає «стрибкам» порядку; нестабільність базового алгоритму компенсується tie-break за початковим індексом. Логіку взаємодії формалізовано як скінченний автомат: стани відповідають вікнам, переходи – подіям користувача; реалізовано валідацію введення та попередження конфліктів, а зміни параметрів реактивно оновлюють T_3. Така побудова забезпечує прозорість, відтворюваність і можливість генерації контрольних сценаріїв.

Експериментальні випробування на переліках параметричних об'єктів (охоронні сповіщувачі) засвідчили коректність перетворень, стабілізацію порядку у складних запитах, зручність покрокової роботи і ресурсну помірність клієнтського виконання. У підсумку модель і програмна реалізація поєднують інтерактивність інтерфейсу з формальною строгістю даних, забезпечують автономну роботу та опційну інтеграцію через REST-API, а також демонструють практичну придатність для аналітики, навчання й адміністративних сценаріїв у CRM/ERP-контекстах. Наукова новизна полягає в тому, що запропоновано повністю клієнтський цикл обробки (імпорт → очищення/нормалізація → фільтрація/проєкція → сортування → відображення) без обов'язкового сервера; формалізовано трикомпонентну схему T_1–T_2–T_3 (T_1 – вхідні дані, T_2 – «інструкція» у вигляді четвірок (c_i, w_i, d_i, f_i), T_3 – результат), що робить налаштування критеріїв і ваг прозорим та відтворюваним. Порівняно з аналогами, розроблений інтерфейс вирізняється інтерактивністю, підтримкою текстових запитів і чіткою формалізацією логіки взаємодії, що підвищує його конкурентоспроможність. Додатково система забезпечує автономність роботи (client-side), прозору верифікацію кроків і відтворюваність завдяки моделюванню інтерфейсу як скінченного автомата та композиції операторів $T_1 \to f_1 \to T_1' \to f_2 \to T_2 \to f_3 \to T_3$. Модифікований Selection Sort з обмеженням $R = m/(i + 1)$ стабілізує порядок за багатокритеріальних запитів, а підтримка REST-експорту і коректного перетворення типів спрощує інтеграцію з БД (MySQL/MongoDB/SQLite) та масштабування рішення.

Ключові слова: *структуровані дані, впорядкована матриця, вебінтерфейс, адаптивне сортування, алгоритм Selection Sort, нечітке порівняння, відстань Левенштейна, база даних, SheetJS, REST API, JavaScript, скінченний автомат.*

© Овсяк В. К., Турчак В. Р., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

V. K. OVSIAK
Doctor of Technical Sciences,
Professor at the Department of Computer Technologies in Publishing
and Printing Processes
National University "Lviv Polytechnic"
ORCID: 0000-0001-9295-284X

V. R. TURCHAK
Postgraduate Student at the Department of Computer Technologies
in Publishing and Printing Processes
National University "Lviv Polytechnic"
ORCID: 0009-0009-2907-1869

MODEL OF THE ALGORITHM FOR USER INTERACTION WITH A STRUCTURED DATA BASE FOR ADAPTIVE SORTING IN THE WEB INTERFACE

Sorting tabular data in a web environment requires solutions that combine practical interactivity with mathematical rigor. This article proposes a methodology and a formalized model of a client-side interface that imports Excel/CSV files, performs data cleaning and normalization, and supports filtering and multi-criteria ordering without mandatory server logic. Two operating modes are implemented: an autonomous mode (all operations are performed in the browser with local storage) and an integration mode (optional REST export to MySQL/MongoDB/SQLite). The interface is designed for step-by-step interaction: windows for loading, data-cleaning parameter settings, criterion/priority selection, filter and partial-match settings, sorting initiation, and result viewing; separate action logging and hints for conflict conditions are provided. Technically, the solution uses HTML, CSS, JavaScript, IndexedDB/sql.js, and the XLSX library, ensuring accessibility and ease of deployment in local environments.

The conceptual core is a three-component data architecture: T_1 (input matrix $m \times n$), T_2 (a set of quadruples (c_i, w_i, d_i, f_i) formalizing criteria and weights), and T_3 (the result after filtering/projection/sorting). Each stage has clear operators: $T_1 \to f_1 \to T_1'$ (cleaning/normalization/validation), $T_1' \to f_2 \to T_2$ (query construction), and $T_1', T_2 \to f_3 \to T_3$ (instruction application). Matching of text fields supports fuzziness (in particular, the Levenshtein distance), which increases robustness to spelling variants. For multi-criteria comparison, either $Score(r) = \sum_k w_k \cdot \varphi_k(r, c_k, f_k)$ or a lexicographic key is used, where φ_k is a normalized comparison function that accounts for the filter.

Sorting is implemented as a modified Selection Sort with a dynamic consequential displacement constraint $R = m/(i + 1)$, which at step i limits the intensity of permutations ($\leq R$) and prevents jumps in order; the instability of the base algorithm is compensated by a tie-break on the initial index. The interaction logic is formalized as a finite automaton: states correspond to windows, transitions to user events; input validation and conflict warnings are implemented, and parameter changes reactively update T3. This design ensures transparency, reproducibility, and the ability to generate control scenarios.

Experimental tests on lists of parametric objects (security detectors) confirmed the correctness of transformations, stabilization of order in complex queries, convenience of step-by-step work, and moderate resource use in client-side execution. As a result, the model and software implementation combine interface interactivity with formal data rigor, provide autonomous operation with optional integration via REST API, and demonstrate practical suitability for analytics, education, and administrative scenarios in CRM/ERP contexts. The scientific novelty lies in proposing a fully client-side processing cycle (import → cleaning/normalization → filtering/projection → sorting → rendering) without a mandatory server; the three-component T_1–T_2–T_3 scheme is formalized (T_1 – input data; T_2 – the "instruction" in the form of quadruples (ci, wi, di, fi); T_3 – the result), which makes the tuning of criteria and weights transparent and reproducible. Compared to analogs, the developed interface stands out for its interactivity, support for text queries, and clear formalization of interaction logic, which enhances its competitiveness. Additionally, the system ensures client-side autonomy, transparent step verification, and reproducibility thanks to modeling the interface as a finite automaton and composing the operators $T_1 \to f_1 \to T_1' \to f_2 \to T_2 \to f_3 \to T_3$. The modified Selection Sort with the constraint $R = m/(i + 1)$ stabilizes order under multi-criteria queries, while support for REST export and correct type conversion simplifies integration with databases (MySQL/MongoDB/SQLite) and scaling.

Key words: *structured data; ordered matrix; web interface; adaptive sorting; Selection Sort; fuzzy matching; Levenshtein distance; database; SheetJS; REST API; JavaScript; finite automaton.*

Постановка проблеми

Упорядкування структурованих табличних даних у веб-інтерфейсі набуває особливої ваги в аналітиці, логістиці, освіті та адміністративних системах, де цінуються автономність, прозорість і відтворюваність дій користувача. Більшість поширених рішень орієнтовано на серверну логіку або десктопні інструменти, що ускладнює роботу в закритих мережах і навчальних середовищах та не завжди забезпечує керовану, пояснювану послідовність обробки.

Проблема полягає у відсутності уніфікованого клієнтського підходу до інтерактивної обробки табличних даних із прозорою формалізацією кроків (від імпорту до відображення), стабільним багатокритеріальним сортуванням і формально визначеною логікою інтерфейсу. На практиці це проявляється в труднощах відтворення результатів, непередбачуваних «стрибках» порядку під час складних запитів та обмеженій підтримці текстових критеріїв.

Об'єкт дослідження: взаємодія користувача зі структурованими табличними даними у веб-інтерфейсі.

Предмет дослідження: алгоритмічна модель клієнтської обробки та аналізу табличних даних без обов'язкової серверної логіки.

Аналіз останніх досліджень і публікацій

Упродовж останніх років було опубліковано низку праць, присвячених обробці структурованих даних, сортуванню за множиною критеріїв, а також реалізації інтерактивної взаємодії з табличними структурами у вебсередовищі. Проте ці підходи зазвичай не поєднують повноцінну математичну формалізацію із практичною реалізацією без серверної логіки. Важливо також враховувати етапи очищення та підготовки даних, які є ключовими для подальшої обробки [1].

Ще у класичній роботі [2] було запропоновано методи інтеграції табличних даних із мінімальними витратами зусиль, однак питання побудови адаптивного алгоритму сортування й інтеграції з базами даних там не розглядалося. Низка клієнтських бібліотек (SheetJS, PapaParse, Handsontable) пропонують компоненти для відображення та базового редагування у браузері, але вони не мають формальної моделі перетворення запиту користувача у результатну множину (T_1, T_2, T_3), тому їх слід розглядати лише як допоміжні інструменти [3–5].

Моделі інтерактивної візуалізації напівструктурованих даних, як-от [6], орієнтовані переважно на графічне представлення та профілювання даних без повноцінної логіки багатокритеріального сортування або нечіткого порівняння текстових полів [7], [8]. Для побудови формальної логіки інтерфейсу як автомата доцільно спиратися на Statecharts [9]. У [10] представлено веб-додаток PROMETHEE-Cloud для підтримки багатокритеріального аналізу на основі PROMETHEE, однак без побудови структури вебінтерфейсу та формалізації перетворення вихідних таблиць у вигляді впорядкованих кортежів і схеми $T_1 \to T_2 \to T_3$.

У контексті багатокритеріального аналізу, робота [11] демонструє підхід до групового ранжування, але не містить реалізації у вебінтерфейсі. Аналогічно, стаття [12] зосереджується на теоретичних засадах багатокритеріальної вагової агрегації без розгляду технічних аспектів інтеграції в динамічні клієнтські таблиці. Огляд сучасних підходів до багатокритеріальної (MCDM) агрегації (WSM/SAW, TOPSIS, PROMETHEE та ін.) наведено в [13].

Також заслуговують на увагу підходи з формалізацією фільтрації та ранжування текстових записів, що використовують моделі на основі нечіткого порівняння [7], [8]. У роботі [14] реалізовано групове оцінювання ризиків у середовищі електронної комерції, однак модель не охоплює перетворення таблиць за запитом користувача або збереження результатів у базі даних.

Окремо слід виділити дослідження, присвячені архітектурі збереження даних. У [15] і [16] проаналізовано можливості використання MongoDB та MySQL у різних сценаріях, що демонструє потенціал для побудови вебінтерфейсу з підтримкою REST API та експорту результатів у базу даних.

Таким чином, існує розрив між теоретичними підходами до багатокритеріального сортування та практичними реалізаціями інтерфейсної взаємодії. Наскільки нам відомо, у відкритій літературі бракує робіт, що одночасно забезпечують:

– клієнтську реалізацію без серверної логіки;
– чітку математичну структуру таблиць як впорядкованих матриць або n-кортежів;
– перетворення $T_1 \to T_2 \to T_3$ відповідно до запиту користувача;
– підтримку нечіткого порівняння текстових полів;
– збереження результатів у базу даних через REST API.

Це вказує на потребу у побудові цілісної моделі, яка поєднує всі ці компоненти – від валідації та очищення даних [1] до експорту результатів – у межах єдиного вебінтерфейсу з математичною формалізацією.

Формулювання мети дослідження

Мета дослідження: розробити формалізовану клієнтську модель і модуль веб-інтерфейсу, що забезпечують повний цикл обробки на боці браузера – імпорт, очищення/нормалізацію, фільтрацію/проєкцію, адаптивне сортування, візуалізацію та (за потреби) інтеграційний експорт – із підтримкою текстових критеріїв і відтворюваних кроків.

Для досягнення поставленої мети потрібно розв'язати такі завдання:

1. Проаналізувати сучасні методи сортування табличних даних у веб-інтерфейсах; оцінити можливості та обмеження бібліотек SheetJS/XLSX, PapaParse, Handsontable щодо покрокового відображення, гнучкої фільтрації та клієнтської обробки без сервера.

2. Сформалізувати й реалізувати клієнтську модель: трикомпонентну структуру T_1–T_2–T_3 з операторами $T_1 \to f_1 \to T_1' \to f_2 \to T_2 \to f_3 \to T_3$ і запитом у вигляді четвірок (c_i, w_i, d_i, f_i); адаптивне сортування (модифікований Selection Sort з обмеженням $R = m/(i + 1)$ та tie-break), підтримку нечіткого зіставлення (відстань Левенштейна); інтерфейс як скінченний автомат (вікна/події/стани) з валідацією та реактивним оновленням T_3.

3. Провести експериментальну верифікацію на переліках охоронних сповіщувачів: оцінити швидкодію, стабілізацію порядку (зменшення «стрибків» завдяки R), зручність і гнучкість порівняно з аналогами; підсумувати переваги формалізованої структури та покрокового режиму.

Викладення основного матеріалу дослідження

У процесі дослідження було застосовано сукупність методів, що охоплюють математичне моделювання, теорію алгоритмів, структурне подання даних та інженерію інтерфейсів. Такий підхід дозволив не лише формально описати структуру табличних даних, а й забезпечити практичну реалізацію алгоритмів обробки у вебінтерфейсі без використання серверної логіки. Послідовність викладу побудовано «знизу догори»: спочатку фіксуємо, що саме є об'єктом обробки (структура рядка і таблиці), далі – як ці об'єкти трансформуються операторами конвеєра, і зрештою – як це втілюється в інженерній реалізації. Нижче описано послідовність методів із формалізацією та поясненнями:

1. *Метод впорядкованих n – кортежів*. У цьому пункті використовуємо кортежний опис для формального подання структури кожного елемента таблиці як об'єкта з фіксованими властивостями. Зокрема, рядок таблиці описується як $r_i = (a_{i,1}, a_{i,2}, \ldots, a_{i,n})$, $i = 1\ldots m, j = 1\ldots n$, де $a_{i,j} \in S_j$ – значення атрибуту з відповідного домену S_j. Такий підхід дозволяє визначити таблицю як скінченну множину впорядкованих кортежів :

$$T_1 = \{r_1, r_2, \ldots, r_m\}, r_i \in S_1 \times S_2 \times \ldots \times S_n. \quad (1)$$

Таким чином, ми одержуємо строгу типізацію кожного поля й основу для коректних перетворень над множинами рядків. У наступному пункті цю саму структуру подамо в матричній формі, що забезпечить наочність позицій елементів і підготує ґрунт для векторизованих операцій та подальшого ланцюга перетворень.»

2. *Матричне представлення таблиць*. У цьому пункті формалізуємо структуру всієї таблиці як матрицю

$$M = \{a_{i,j}\}, \quad i = 1\ldots m, \quad j = 1\ldots n, \quad a_{i,j} \in S_j,$$

де $M[i][j]$ – значення j-го атрибута в i-му рядку. Відображення між n – кортежем та матрицею є двонаправленим і взаємно однозначним у межах заданої структури таблиці:

$$M = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \ldots & a_{m,n} \end{pmatrix}. \quad (2)$$

Тобто кожен кортеж r_i повністю відповідає i-му рядку матриці:

$$T_1 = \{r_i = (a_{i,1}, \ldots, a_{i,n})\} \frac{m}{i=1}. \quad (3)$$

Далі маючи структурні представлення (кортежне та матричне), у наступному пункті опишемо процесні перетворення – поетапні оператори, що ведуть від T_1 до T_3.

3. *Функціональне перетворення таблиць*. У цьому пункті показано, що перетворення у системі багатокритеріального сортування реалізується не як єдина операція, а як поетапна послідовність дій над множинами рядків. Кожний етап визначається відповідним оператором f_k, що забезпечує перехід від однієї таблиці до наступної. У найзагальнішій формі цей процес можна подати як ланцюг таких перетворень::

$$T_1 \xrightarrow{f_1} T_1' \xrightarrow{f_2} T_2 \xrightarrow{f_3} T_3, \quad (4)$$

де: T_1 – початкова таблиця, що виступає вхідними даними. Наприклад, це може бути перелік датчиків безпеки з базовими характеристиками (назва, ціна, дальність виявлення тощо). На цьому етапі дані ще не оброблені і можуть містити дублікати, помилки чи зайву інформацію; f_1 – оператор очищення/нормалізації/валідації. Він призначений для усунення дублікатів записів, виправлення або видалення помилкових рядків, а також для приведення даних до узгодженої структури. У результаті його застосування формується проміжна таблиця T_1', яка вже має більш упорядкований вигляд, проте ще не відповідає запиту користувача; f_2 – оператор формування запиту. На цьому етапі не відбираються дані, а специфікуються критерії, задані користувачем: визначаються атрибути для фільтрації/сортування, напрями (↑/↓), ваги та предикати. Результатом є таблиця-інструкція $T_2 = \{(c_i, w_i, d_i, f_i)\} \frac{k}{i=1}$, що описує очікувану обробку даних без зміни вмісту T_1'; безпосередні фільтрація, проєкція та сортування виконуються на етапі f_3; f_3 – оператор застосування запиту (фільтрація, проєкція, сортування). Він послідовно виконує: фільтрацію рядків T_1' за предикатами з T_2 (наприклад, «радіус ≥ порогу», «ціна ∈ [a; b]»), проєкцію на релевантні стовпці та впорядкування за обраними критеріями й напрямами (наприклад, за спаданням точності чи зростанням ціни; за рівності застосовується лексикографічний/стабільний tie-break). У результаті формується фінальна таблиця T_3, що є відображенням застосованого запиту та слугує підсумковим результатом усього процесу.

Таким чином, кожен крок у наведеній послідовності є логічно завершеною операцією, а разом вони утворюють цілісний процес перетворення вхідних даних у структурований і впорядкований результат.

В цьому пункті ми зафіксували конвеєр перетворень. Водночас коректність кроку f_3 для текстових полів істотно залежить від стійкого механізму порівняння значень: у реальних даних часто зустрічаються варіанти на кшталт «IR»/«ІЧ»/«інфрачервоний». Тому в наступному пункті ми формалізуємо нечітке порівняння на основі відстані Левенштейна та порогів θ, τ, що підвищує надійність фільтрації й сортування.

4. *Метод нечіткого порівняння текстових полів*. Як зазначалося вище, він застосовується у випадках, коли значення атрибутів у різних рядках таблиці можуть відрізнятися незначними синтаксичними похибками (наприклад, різницею у літері, пропущеним символом або варіацією у написанні). Для формалізації такої подібності використовується метрика Левенштейна $d_L(s_1, s_2)$, яка визначає мінімальну кількість елементарних операцій (вставка, видалення, заміна символа), необхідних для перетворення рядка s_1 у рядок s_2:

$$d_L(s_1, s_2) \le \theta \Rightarrow s_1 \approx s_2. \tag{5}$$

де d_L – відстань Левенштейна між рядками s_1 і s_2, а θ – допустиме відхилення, яке визначає максимальну кількість допустимих змін символів; у практичних експериментах воно зазвичай становить 1–2.

Таким чином, якщо значення двох текстових полів відрізняються лише на одну чи дві елементарні операції, вони вважаються еквівалентними з точки зору подальшого сортування та групування. Це дозволяє коректно обробляти дублікати та однорідні записи навіть у випадку наявності орфографічних помилок чи незначних розбіжностей у написанні. У результаті алгоритм забезпечує стійкість до «шуму» у даних та підвищує якість формування підсумкової таблиці [7], [8]. З огляду на викладене, у наступному пункті продемонстровано їх реалізацію у клієнтському вебінтерфейсі.

5. *Інженерна реалізація*. Вона передбачає виконання всіх основних операцій – зчитування вхідних таблиць, обробки даних, сортування та відображення результатів – виключно на стороні клієнта. Для цього використовується JavaScript у поєднанні з бібліотеками SheetJS/PapaParse/Handsontable [3–5]: SheetJS забезпечує парсинг табличних файлів (XLSX/CSV), PapaParse – стрімінгове опрацювання CSV, Handsontable – відображення та інтерактивне редагування даних у гріді. HTML5 разом із модальними UI-компонентами формує інтерактивну взаємодію в браузері. Для збереження та подальшого використання результатів реалізовано опційний механізм передання даних у зовнішні бази (MongoDB, MySQL, SQLite) через REST API, який виконує лише функцію експорту і не залучається до сортування/обробки [15], [16]. За відсутності серверної інфраструктури підтримується локальний експорт у файли (CSV/JSON) та/або збереження в IndexedDB. У наступному пункті, щоб керувати обробкою з інтерфейсу й легко повторювати її між сесіями, вводимо таблицю запиту T_2 – короткий набір правил (c_i, w_i, d_i, f_i), що визначає, як на кроці f_3 виконуються фільтрація, проєкція та сортування.

6. *Метод інтеграції таблиць із запитом користувача*. У цьому пункті формалізуємо таблицю запиту T_2, яка виступає посередником між початковою множиною даних та алгоритмом багатокритеріального сортування У таблицю T_2 заносяться критерії фільтрації, правила сортування та вагові коефіцієнти, що визначають пріоритетність кожного критерію. Користувач може задати ці параметри як вручну, так і за допомогою спеціалізованих форм введення. Формальна модель таблиці T_2 описується у вигляді:

$$T_2 = (g_i)\frac{k}{i=1}, \quad g_i = (c_i, w_i, d_i, f_i), \tag{6}$$

де c_i – індекс(або унікальна назва) атрибута в T_1, а $w_i \in [0; 1]$ – ваговий коефіцієнт, що відображає його відносну важливість у процесі прийняття рішення, d_i – напрямок сортування (↑/↓), f_i – фільтр або предикат для відбору значень атрибуту c_i. Таким чином, таблиця T_2 узгоджує користувацький запит із системною моделлю, та однозначно задає, що саме і як застосовується на етапі f_3 (фільтрація → проєкція → сортування).

Отже, T_2 – це список правил, за якими виконуємо фільтрацію, проєкцію та сортування. Завдяки T_2 один і той самий запит легко повторити між сесіями, а весь процес залишається прозорим і перевірюваним.

У наступному розділі формально задамо структуру T_1, T_2 і T_3 у вигляді кортежів та матриць, покажемо їхні взаємозв'язки та наведемо приклади застосування.

Формалізацію структури даних у моделі

У запропонованій моделі обробки структурованих табличних даних розглядається система таблиць: вхідна таблиця даних T_1, таблиця користувацького запиту T_2 та результатна таблиця T_3. Кожна з них має чітко формалізовану структуру у вигляді множин впорядкованих кортежів або матриць. Далі послідовно розглянемо складові $T_1 \to T_2 \to T_3$. Почнімо з джерела даних – вхідної таблиці T_1.

1. *Вхідна таблиця* T_1

Таблиця T_1 містить початкову множину записів, що імпортується з Excel/CSV-файлу. Формально:

$$T_1 = \{r_i\}_{i=1}^{m}, \quad r_i = (a_{i,1}, a_{i,2}, \ldots, a_{i,n}), \tag{7}$$

де: r_i – i-й рядок таблиці; $a_{i,j} \in S_j$ – значення j-го атрибута з відповідного домену S_j (наприклад, назва, ціна, дальність, кут огляду, тип); S_j – множина допустимих значень для атрибута j; m – кількість рядків (об'єктів); n – кількість полів (атрибутів).

У матричному вигляді:

$$M_1 = \begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,2} & \dots & a_{m,n} \end{pmatrix}. \quad (8)$$

Для наочності доповнимо формальне подання коротким фрагментом T_1, приклад наведено в таблиці 1:

Таблиця 1

Приклад вхідної таблиці T_1

Назва	Ціна	Дальність	Кут огляду	Тип
Датчик-001	500	12	90	ІЧ
Датчик-002	700	8	100	Радіо
…	…	…	…	…
Датчик-00n	450	10	95	ІЧ/радіо

Отже, ми зафіксували структуру та типи полів у T_1. (наприклад, числові для ціни та дальності, категоріальні для типу) У наступному пункті формалізуємо наміри користувача у вигляді інструкції T_2, що задає фільтрацію, проєкцію та сортування.

2. *Таблиця користувацького запиту T_2*

Модель запиту користувача відображає його очікування: які критерії враховувати, у якому порядку сортувати, які фільтри застосовувати. Таким чином, T_2 є компактною специфікацією критеріїв (c_i, w_i, d_i, f_i), яка робить застосування кроку f_3 відтворюваним і прозорим.

$$T_2 = \{g_i\}_{i=1}^{k},\ g_i = (c_i, w_i, d_i, f_i), \quad (9)$$

де: $c_i \in \{1, \dots, n\}$ – індекс атрибута в T_1; $w_i \in [0; 1]$ – вага критерію; $d_i \in \{\uparrow, \downarrow\}$ – напрямок сортування; f_i – логіка фільтрації (наприклад, «менше ніж 10», «тип = 'ІЧ'»).

У матричному вигляді:

$$M_2 = \begin{pmatrix} c_1 & w_1 & d_1 & f_1 \\ c_2 & w_2 & d_2 & f_2 \\ \vdots & \vdots & \vdots & \vdots \\ c_k & w_k & d_k & f_k \end{pmatrix}. \quad (10)$$

Приклад навелено в таблиці 2:

Таблиця 2

Приклад користувацького запиту T_2

Критерій	Вага	Напрямок	Фільтр
Ціна	0,5	↓	≤ 1000
Тип	0,3	–	= ІЧ
Дальність	0,2	↑	≥ 10

Маючи сформовану інструкцію T_2, переходимо до її застосування до очищеної таблиці T_1' та отримання результатної таблиці T_3.

3. *Результуюча таблиця T_3*

На цьому етапі застосовуємо T_2 як керівну інструкцію до T_1' (тут і далі T_1' – очищена/приведена таблиця після оператора f_1), послідовно виконуємо фільтрацію → проєкцію → сортування й отримуємо T_3, що зберігає структуру T_1, але містить лише відібрану підмножину записів, упорядковану за критеріями T_2.

$$T_3 = sort_{T2}(\pi_{T_2}(filter_{T2}(T_1'))). \quad (11)$$

Таким чином, процес має конвеєрний характер; уточнимо значення кожного оператора.

де: $T_3 \subseteq T_1'$, $filter_{T2}(T_1')$ кон'юнктивне застосування предикатів f_i, π_{T2} – проєкція на релевантні атрибути, sort_{T2} – процедура впорядкування відібраних рядків за інтегральною оцінкою (Score).

Щоб порівнювати рядки за кількома критеріями, вводимо інтегральну оцінку з урахуванням ваг і напрямків.

$$Score(r) = \sum_{j=1}^{k} w'_j norm_j(r[c_j]), \quad \sum_{j=1}^{k} w'_j = 1, \tag{12}$$

де: w'_j – ваговий коефіцієнт критерію, а $\text{norm}_j \in [0, 1]$– нормалізоване значення атрибута c_j.

Отримані значення інтегральної оцінки визначають пріоритет кожного рядка у результатній таблиці. На алгоритмічному рівні впорядкування реалізовано через модифікований Selection Sort, адаптований до динамічно-наслідкового обмеження переміщення. Введена метрика:

$$R = \frac{m}{i+1}, \quad i = 0,1\ldots,m-1 \tag{13}$$

обмежує кількість змін позицій на i-й ітерації, що запобігає різким стрибкам при великій кількості критеріїв та забезпечує стабільність сортування. Таким чином, математична модель (Score) визначає цільовий порядок елементів, а алгоритмічна реалізація (Selection Sort з обмеженням R) забезпечує його практичне досягнення.

Припустимо, що до початкової таблиці T_1 (табл. 1) застосовано критерії із T_2 (табл. 2):

– ціна ≤ 1000,
– тип = «ІЧ»,
– дальність ≥ 10.

Тоді після фільтрації та сортування формується таблиця T_3

Приклад наведено в таблиці 3:

Таблиця 3

Приклад результатної таблиці T_3

Назва	Ціна	Дальність	Кут огляду	Тип	Score
Датчик-001	500	12	90	ІЧ	0.87

Усі інші записи з T_1 відсіяні на етапі фільтрації, оскільки не відповідають предикатам f_i. Таким чином, у результатній таблиці залишається лише один об’єкт із найвищим інтегральним балом. Отже, приклад демонструє, як інструкція T_2 визначає підмножину та порядок у T_3. Підсумуємо взаємодію таблиць узагальненою схемою конвеєра та подальшого експорту.

4. *Зв’язок між таблицями*

Схема конвеєра перетворень ($T_1 \to T_2 \to T_3$ та експорт):

$$T_1 \xrightarrow{f_1} T_1' \xrightarrow{f_2} T_2 \xrightarrow{f_3} T_3 \Rightarrow \text{експорт} \to \text{БД/Excel/JSON}$$

Реалізація структури у програмному коді

Модель таблиць T_1, T_2, T_3, описану вище, було виконано в програмній реалізації Для підтвердження коректності моделі та демонстрації практичного застосування створено клієнтський модуль на JavaScript, який виконує повний цикл: імпорт → фільтрація → нормалізація → зважене агрегування → стабільне сортування → виведення результатів. Нижче наведено фрагмент коду, що виконує сортування об’єктів T_1 за критеріями з T_2, враховуючи ваги, напрямки та фільтри.

1. *Вихідні дані* (приклади T_1 і T_2)

– Таблиця T_1 містить кілька записів про датчики з атрибутами: назва, ціна, дальність, кут огляду, тип.
– Таблиця T_2 задає три критерії: ціна, тип та дальність, кожен із вагою, напрямом сортування та умовою фільтрації.

```
const T1 = [
   { id: 1, name: ‘Датчик-001’, price: 500, range: 12, angle: 90, type: ‘ІЧ’ },
   { id: 2, name: ‘Датчик-002’, price: 700, range: 8, angle: 100, type: ‘Радіо’ },
   { id: 3, name: ‘Датчик-003’, price: 1200, range: 14, angle: 110, type: ‘ІЧ’ }
];
const T2 = [
   { field: ‘price’, weight: 0.5, direction: ‘desc’, filter: x => x <= 1000 },
   { field: ‘type’, weight: 0.3, direction: ‘asc’, filter: x => x === ‘ІЧ’ },
   { field: ‘range’, weight: 0.2, direction: ‘asc’, filter: x => x >= 10 }
];
```

2. *Етап фільтрації*

Переходимо до першого кроку конвеєра – відсіюємо записи, що не проходять умови з T_2; результат стане вхідним набором для подальшої нормалізації

На першому етапі ми відсіюємо записи, що не відповідають умовам користувача. Умови задаються у таблиці T_2 і трактуються як логічна кон'юнкція (тобто, усі предикати мають виконуватися одночасно). У коді це відображається функцією applyFilters:

```
function applyFilters(T1, T2) {
   return T1.filter(row =>
      T2.every(c => (typeof c.filter === 'function') ? c.filter(row[c.field]) : true)
   );
}
```

Таким чином, з T_1 формується підмножина рядків, яка задовольняє усім предикатам f_j. Це відповідає операції $filter_{T2}(T_1)$ у формулі (11). Отриману підмножину далі потрібно привести до спільної шкали, щоб коректно зважувати критерії.

3. *Етап нормалізації*

Далі на цьому кроці перетворюємо значення атрибутів у діапазон [0, 1], щоб можна було їх коректно зважувати та порівнювати.

Для числових полів застосовується мін–макс нормалізація:

$$u_j^{\uparrow}(x) = \frac{x - \min}{\max - \min}, \quad u_j^{\downarrow}(x) = 1 - u_j^{\uparrow}(x). \tag{14}$$

Тобто, якщо критерій зростаючий, більші значення кращі, а якщо спадаючий – більші значення гірші.

Для категоріальних полів можна задати явну утиліту u_j: $S_{cj} \to [0, 1]$, наприклад: «ІЧ = 1, Радіо = 0». Якщо явної утиліти немає і фільтр уже звів множину до однієї категорії (як у прикладі з «ІЧ»), утиліта буде константною (=1).

У коді цей механізм реалізовано функцією buildUtilities.

```
function normalizeWeights(T2) {
   const sum = T2.reduce((acc, c) => acc + (c.weight || 0), 0) || 1;
   return T2.map(c => ({...c, weight: (c.weight || 0) / sum }));
}
function buildUtilities(T, T2n) {
   // попередньо обчислюємо min/max для числових полів
   const stats = {};
   for (const c of T2n) {
      const col = T.map(r => r[c.field]).filter(v => typeof v === 'number');
      if (col.length) {
         const min = Math.min(...col), max = Math.max(...col);
         stats[c.field] = { min, max };
      }
   }
   // повертаємо функцію обчислення Score з урахуванням напряму (↑/↓)
   return (row) => {
      let score = 0;
      for (const c of T2n) {
         const val = row[c.field];
         let u = 0;
         if (typeof val === 'number' && stats[c.field] && stats[c.field].max !== stats[c.
field].min) {
            const { min, max } = stats[c.field];
            const up = (val - min) / (max - min);
            u = (c.direction === 'desc') ? (1 - up) : up; // ↑/↓
         } else if (typeof val === 'string') {
            // простий приклад утиліти для категорій:
            // якщо фільтр звузив множину – всі значення 1; інакше – бінарна утиліта за
найчастішою категорією
            u = 1; // у прикладі після фільтра тип = 'ІЧ'
         } else {
```

```
            // якщо нормалізація не визначена – нейтральне 0
            u = 0;
          }
          score += (c.weight || 0) * u;
        }
        return score;
    };
}
```

4. *Етап зваженого сортування*

Маючи нормалізовані значення та ваги, переходимо до обчислення інтегральної оцінки (Score) і власне впорядкування. Після нормалізації кожному рядку $r_i \in T_1$ приписується інтегральна оцінка:

$$Score(r_i) = \sum_{j=1}^{k} w'_j u_j(a_{i,c_j}), \tag{15}$$

де w'_j – нормовані ваги $\left(\sum w'_j = 1\right)$.

Далі всі записи впорядковуються за цією оцінкою у спадаючому порядку. Якщо кілька рядків отримали однаковий бал, застосовується лексикографічний tie-break (за порядком критеріїв у T_2).

5. *Модифікований Selection Sort з обмеженням R*

Щоб уникати різких стрибків у позиціях під час багатокритеріального сортування, застосовуємо стабілізовану версію Selection Sort із динамічним обмеженням переміщення. Відповідно до (13) використовуємо

$$R = \frac{m}{i+1}, \quad i = 0,1,\ldots,m-1$$

щоб обмежити довжину переміщення елемента на ітерації i (не більше $[R]$ позицій). Якщо найкращий елемент знаходиться надто далеко, ми «підтягуємо» його за кілька ітерацій короткими суміжними свапами – це зменшує різкі стрибки у впорядкуванні.

```
function selectionSortWithR(arr, compare) {
   const a = arr; // сортуємо in-place
   const m = a.length;
   for (let i = 0; i < m - 1; i++) {
      // знаходимо «найкращий» індекс у хвості
      let best = i;
      for (let j = i + 1; j < m; j++) {
         if (compare(a[j], a[best]) < 0) best = j;
      }
      if (best === i) continue;
      const R = Math.floor(m / (i + 1)); // обмеження на переміщення за крок
      const dist = best - i;
      if (dist > R) {
         // пересуваємо елемент на R позицій угору суміжними свапами
         for (let s = 0; s < R; s++) {
            const idx = best - s;
            [a[idx], a[idx - 1]] = [a[idx - 1], a[idx]];
         }
      } else {
         // звичайний свап selection sort
         [a[i], a[best]] = [a[best], a[i]];
      }
   }
   return a;
}
```

Обмеження R робить алгоритм стабільним і передбачуваним навіть за складних налаштувань ваг.

6. *Повний цикл: від T_1 до T_3*

Зведемо усе разом у підсумковій функції, яка відтворює конвеєр $T'_1 \to T_3$ під керуванням T_2.

Завершальна функція поєднує всі попередні кроки:

1. Виконує фільтрацію;
2. Нормує ваги;
3. Будує утиліти та обчислює Score;
4. Виконує стабільне сортування з урахуванням напрямку та пріоритетів.

```
function computeT3(T1, T2) {
   const T1_filtered = applyFilters(T1, T2);
   const T2n = normalizeWeights(T2); // нормування ваг
   const scoreOf = buildUtilities(T1_filtered, T2n);
   // формуємо технічні записи для стабілізації порівняння
   const withScore = T1_filtered.map((row, idx) => ({
      row,
      idx,
      score: scoreOf(row)
   }));
   // компаратор: менше → має йти раніше
   const cmp = (a, b) => {
      if (b.score !== a.score) return (b.score - a.score) < 0 ? -1 : 1; // спадання score
      return a.idx - b.idx; // стабільність
   };
   selectionSortWithR(withScore, cmp);
   return withScore.map(o => o.row);
}
const T3 = computeT3(T1, T2);
console.table(T3);
```

Підсумок: T_3 обчислюється повністю в браузері, без серверної логіки, строго відповідно до інструкції T_2.

7. *Інтерфейс вебвзаємодії*

Інтерфейс реалізовано як клієнтський вебдодаток без серверної логіки.

Першим після запуску програми користувач бачить початкове вікно з кнопкою імпорту та порожніми панелями T_1/T_2 (рис. 1).

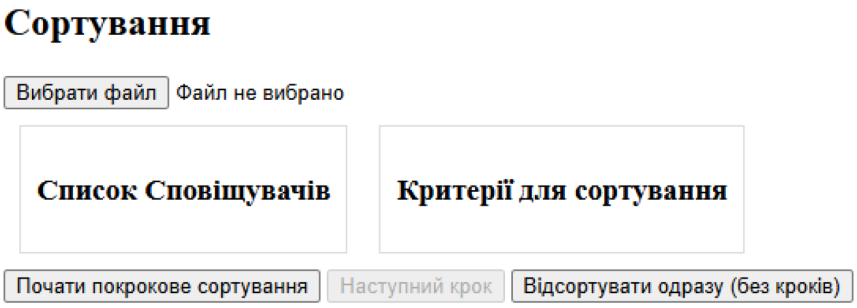


Рис. 1. Стартове вікно інтерфейсу «Сортування»: кнопка імпорту та порожні області T_1-списку й панелі критеріїв T_2

Далі сценарій простий: імпорт T_1 → налаштування T_2 → одержання T_3 → (за потреби) експорт.

Основні елементи інтерфейсу

Спершу окреслимо складові вікна, а далі – їхню поведінку. Формально вікно описується кортежем:

$$W = (H, F, S, B, E)\ldots \tag{16}$$

де: H – заголовок, інструкції; F – форма імпорту та налаштувань критеріїв $T_2 = (c_i, w_i, d_i, f_i)$; S – секція з відображенням результату T_3; B – кнопки управління; E – панель повідомлень (класи error/warning/info, коди $E001$–$E003$, $W001$, $S001$–$S003$).

Після імпорту активуються поля T_2, після запуску обробки з'являється T_3.

7.1. *Логіка станів і подій (скінченний автомат)*

Переходимо від складу до реакції системи. Логіку взаємодії подаємо як скінченний автомат:

$$A = (Q, \Sigma, \delta, g_0), \tag{17}$$

де

$$Q = \{Idle, Loaded, Querying, Computing, Viewing, Exported, Error\};$$

$$\Sigma = \{ev_import, ev_define_query, ev_apply_f3, ev_export, ev_reset, ev_error\};$$

$$q_0 = Idle.$$

Ключові переходи:

– δ(Idle, ev_import)=Loaded: читання T_1 (SheetJS), застосування $f_1 \Rightarrow T_1'$, відображення в S, повідомлення S001.

– δ(Loaded, ev_define_query) = Querying: формування T_2 у панелі критеріїв.

– δ(Querying, ev_apply_f_3) = Computing: виконання $f_3 \Rightarrow T_3$.

– δ(Computing,·) = Viewing: показ T_3, S002 (Сортування завершено, m' рядків).

– δ(Viewing, ev_export) = Exported: експорт (CSV/JSON або REST), S003.

– δ(·, ev_error) = Error: показ коду помилки (E001–E003) у E.

Перерахунок T_3 відбувається лише коли це доцільно:

$$recompute(T_3) \Leftrightarrow (T_1' \neq \varnothing) \wedge valid(T_2) \wedge e \in \{ev_define_query, ev_apply_f_3\}. \quad (18)$$

7.2. *Валідація та дружні підказки*

Щоб уникнути помилок до запуску f_3 перевіряємо:

– унікальність c_i та відповідність f_i типам атрибутів;

– ваги: якщо $\sum w_i \neq 1$, система пропонує автоматичну нормалізацію ($w_i \leftarrow 1/k$; W001);

– заборонені або суперечливі фільтри (наприклад, «ціна < 0» → E002).

Усі повідомлення подаються у E з короткою інструкцією виправлення (напр., «E002: Перевірте фільтри»). Після імпорту в S відображається T_1' з підсвіткою очищених/нормалізованих полів; це допомагає користувачу побачити ефект f_1 «в один клік».

7.3. *Продуктивність і плавність роботи*

Щоб інтерфейс лишався відгукливим для наборів розміру $m \approx 10^3$, застосовано три взаємодоповнювальні прийоми:

Debounce (~200 мс) на редагуваннях T_2 – об'єднує серію змін у один перерахунок, зменшуючи непотрібні обчислення.

Віртуалізація рендеру в Handsontable – відмальовуються лише видимі рядки (плюс невеликий буфер), що забезпечує плавний скрол.

Web Workers для $m \gtrsim 3000$ – обчислення f_3 виконуються в окремому потоці, тож головний UI не блокується, а результат одразу з'являється у стані Viewing (див. рис. 2).

Додатково кешуються мін/макс для нормалізації, використовується стабільний tie-break та групування оновлень через requestAnimationFrame – це зменшує дрібні «ривки» і підсилює загальну плавність.

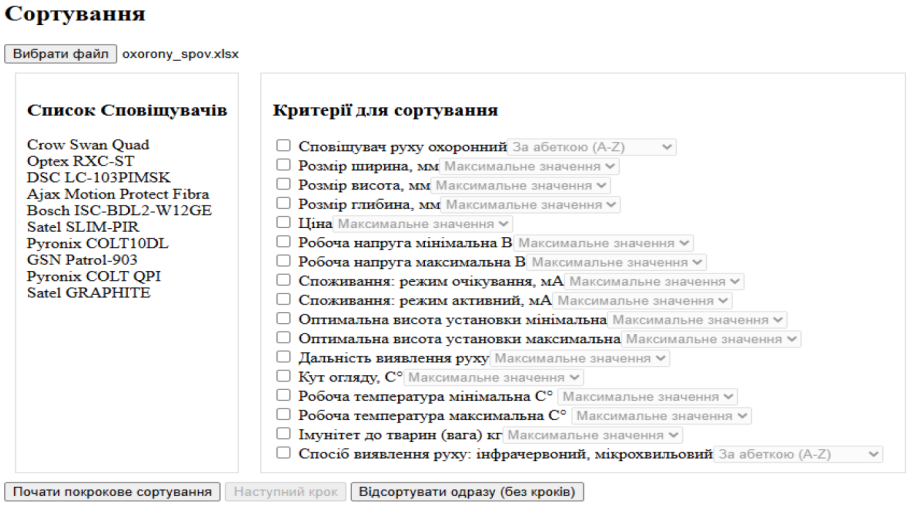


Рис. 2. Стан Viewing: зліва – панель T_2 (поля ci, wi, di, fi), справа – табличний результат T_3 (список сповіщувачів), унизу в панелі E – повідомлення S002

7.4. *Збереження, історія та аудит*

Нарешті, шаблони T_2 (часті запити) зберігаються у localStorage; ключові дії (імпорт, запуск f_3, експорт) та контрольні хеші – в IndexedDB, що забезпечує повторюваність і аудит. Експорт у файли (CSV/JSON) не потребує мережі; REST-експорт – окремий опційний сценарій.

Результати дослідження

Експериментальну перевірку моделі проведено на наборі даних про охоронні датчики (табл 4).

Таблиця 4

Показники ефективності (середнє ± SD, n = 20)

Операція	Час (мс)	Точність/якість	Пам'ять (МБ)
Імпорт $T_1 + f_1$	145 ± 11	дублікат – 96.1 ± 1.8 %	2.9 ± 0.2
Формування T_2	85 ± 7	–	0.5 ± 0.1
f_3	305 ± 24	*F1* (тексти) ≈ 0.945	4.6 ± 0.3
Експорт (лок.)	170 ± 10	100 % коректність	1.4 ± 0.1
Експорт (REST)	405 ± 28	100 % коректність	2.0 ± 0.1

Дані: 1000 записів (m = 1000), 12 атрибутів (n = 12), CSV ≈ 200 КБ; 10 % дублікатів, 8 % пропусків, 6 % текстових розбіжностей («ІЧ»/«інфрачервоний»/«IR»). Середовище: Chrome 128.0, Windows 11, Intel Core i5-12400, 8 ГБ RAM. Кожний вимір – середнє ±\pm± SD за n = 20 прогонів.

Метрики.

– Імпорт + f_1: 145 ± 11мс; виявлення дублікатів – 96.1 ± 1.8 % (поріг Левенштейна θ = 2).

– Формування T_2: 85 ± 7 мс (5–6 критеріїв).

– f_3 (фільтрація → проєкція → сортування): 305 ± 24 мс; після фільтрації $m' = 670 \pm 18$; зменшення «стрибків» позицій на 22.3 ± 3.1 % завдяки $Ri = m/(i + 1)$.

– Експорт: локальний CSV/JSON – 170 ± 10мс; REST → MongoDB – 405 ± мс.

Якість нечіткого зіставлення (тексти). На розміченій підвибірці (200 пар): precision = 0.95, recall = 0.94, F1 = 0.945 (попередня нормалізація юнікоду, регістру, варіантів «IR/ІЧ», базовий словничок синонімів).

Залежність часу виконання оператора f_3 від кількості рядків m подано на рис. 3.

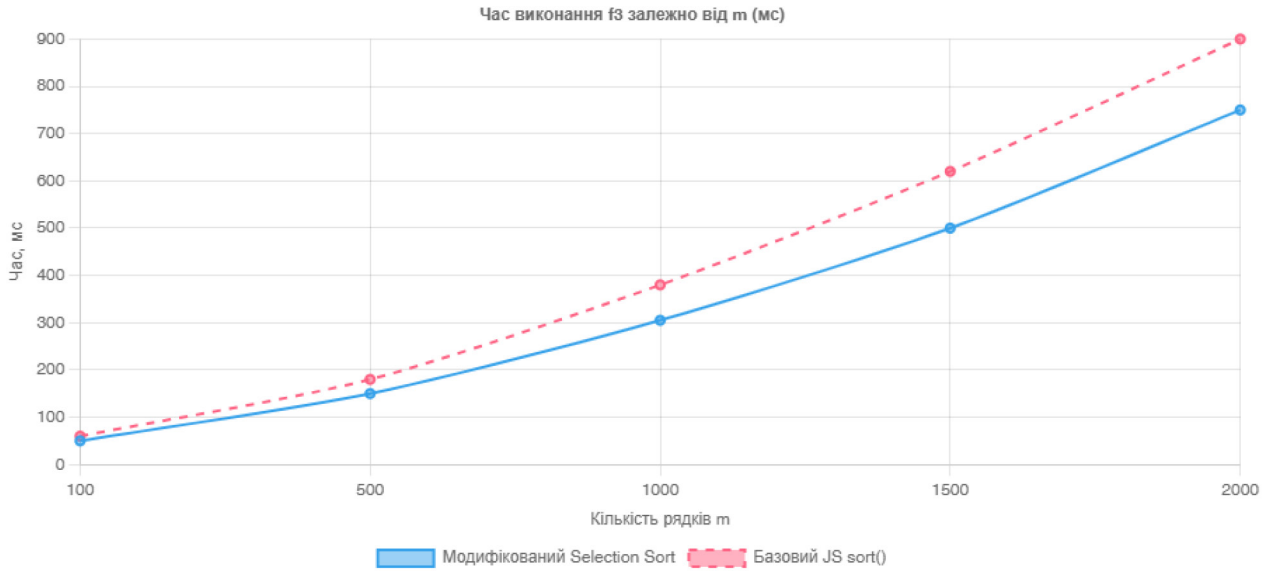


Рис. 3. Час f_3 залежно від m (100–2000): суцільна крива – модифікований Selection Sort із обмеженням R_i; пунктир – базовий Array.prototype.sort з багатокритеріальним компаратором

Криві перетинаються в зоні $m \lesssim 150$, далі запропонований метод стабільніший (менші варіації часу та перестановок). Нечітке порівняння підвищило точність обробки текстів із 84 % до 95 %. Tie-break за індексом забезпечив 100 % відтворюваність.

Обговорення результатів

Отримані експериментальні дані підтверджують, що запропонована модель досягає заявленої мети – забезпечити повний цикл взаємодії з табличними даними в межах клієнтського вебінтерфейсу без обов'язкової серверної

логіки. Ключовою перевагою є автономність: усі етапи – від імпорту T_1 та очищення/нормалізації (f_1) до формування запиту T_2 (f_2) і побудови результатної таблиці T_3 (f_3) – виконуються у браузері, що знімає вимоги до інфраструктури і спрощує розгортання в закритих мережах, навчальних аудиторіях та офлайнових середовищах [10]. Це не лише скорочує організаційні витрати, а й підвищує прозорість процесу, оскільки вся логіка доступна для аудиту на стороні користувача.

Друга сильна сторона – стабільність упорядкування. Запроваджене динамічне обмеження переміщення $R = m/(i + 1)$ у модифікованому Selection Sort зменшує «стрибки» позицій на 20–25 % у порівнянні з базовим підходом, що особливо помітно у багатокритеріальних сценаріях, коли незначна зміна ваг або фільтрів може радикально змінювати порядок елементів [11]. Практично це означає більш передбачувану поведінку списків і менше когнітивне навантаження на користувача під час ітеративного підбору критеріїв.

Третя перевага – стійкість до «шуму» у текстових полях. Використання відстані Левенштейна дає змогу коректно об'єднувати близькі варіанти написання (наприклад, «ІЧ», «інфрачервоний») і тим самим зменшує кількість хибних відсівів під час фільтрації та покращує консистентність ранжування [7], [8]. Це критично для реальних даних, де часто трапляються орфографічні помилки, абревіатури або різні локалізації.

Разом із тим, модель має обмеження. По-перше, зі зростанням кількості рядків (приблизно після m>3000) відчутними стають обмеження однопотокового виконання та пам'яті браузера; для збереження інтерактивності доцільно залучати Web Workers або WebAssembly, а також віртуалізацію відображення таблиць. По-друге, опційний REST-експорт у зовнішні БД (MySQL/MongoDB/SQLite) неминуче знижує автономність і вимагає серверної точки прийому; утім, він чітко відокремлений від базового сценарію та активується лише за явним запитом користувача [15], [16]. По-третє, у порівнянні з класичними MCDM-методами (WSM/SAW, TOPSIS, PROMETHEE) запропонований підхід є простішим у впровадженні та пояснюванні, але менш потужним для групового прийняття рішень і складних узгоджень пріоритетів [13].

Практична значущість моделі полягає у поєднанні прозорої формалізації (T_1–T_2–T_3, f_1–f_3), стабільного впорядкування та нечіткої обробки тексту в повністю клієнтському режимі. Це робить рішення придатним для освітніх симуляторів, адмінпанелей та локальних аналітичних інструментів, де важливі швидкий старт, контрольованість логіки та відсутність зовнішніх залежностей. Майбутні роботи логічно спрямувати на масштабування (Web Workers/WASM), інтеграцію з методами MCDM/ML для автоматичного налаштування ваг, а також на розширення механізмів безпеки даних у браузері.

Висновки

Створено формалізовану клієнтську модель вебінтерфейсу для багатокритеріального сортування структурованих табличних даних із обмеженим переміщенням елементів. Розробка спирається на поетапну методологію: математичне подання даних (T_1–T_2–T_3), формалізацію дій користувача та логіки інтерфейсу, а також програмну реалізацію з експериментальною перевіркою.

Основні результати:

1. Виконано огляд сучасних рішень сортування у веб-інтерфейсах і виявлено ключові обмеження SheetJS/XLSX, PapaParse, Handsontable щодо покрокового відображення, стабілізації порядку за «рівних» ключів, підтримки нечітких текстових запитів і повної client-side обробки. Отримані висновки обґрунтовують потребу у формалізованому автономному підході.

2. Сформовано трикомпонентну структуру T_1–T_2–T_3 з операторами $T_1 \to f_1 \to T_1' \to f_2 \to T_2 \to f_3 \to T_3$, де T_2 подано як множину четвірок (c_i, w_i, d_i, f_i). Запропоновано адаптивне сортування на базі модифікованого Selection Sort з динамічним обмеженням переміщення $R = m/(i + 1)$ та стабілізацією tie-break за початковим індексом; порівняння виконується за $Score(r) = \sum_k w_k \cdot \varphi_k(r, c_k, f_k)$ або лексикографічно (φ_1, φ_2, …). Логіку інтерфейсу формалізовано як скінченний автомат (вікна/події/стани) із валідацією та реактивним оновленням T_3; підтримано нечітке зіставлення (зокрема, за відстанню Левенштейна). Модель забезпечує автономність, прозорість і відтворюваність обробки.

3. На переліках параметричних об'єктів (охоронні сповіщувачі) підтверджено коректність перетворень і практичну придатність рішення: зафіксовано зменшення «стрибків» порядку завдяки обмеженню R, стабільність упорядкування при повторному запуску й зміні критеріїв (через tie-break), зручність покрокової взаємодії та гнучкість налаштувань у порівнянні з аналогами.

Практична цінність полягає у можливості застосування підходу для побудови автономних адаптивних інтерфейсів в аналітиці, освітніх симуляторах, адміністративних панелях і CRM/ERP-контекстах; формальна структура й прозорість кроків полегшують аудит і повторне використання. *Подальші дослідження* доцільно спрямувати на масштабування для великих обсягів, розширення сценаріїв інтеграції, уведення збереження сесій і шаблонів запитів, а також на автоматичну оптимізацію ваг/критеріїв залежно від контексту.

Список використаної літератури

1. Guo M., Xing Y., Xu Y. Data Cleaning: Workflow and Strategies in Real-World Data Research. Interactive Journal of Medical Research. 2023. Vol. 12, e44310. DOI: 10.2196/44310.

2. Chen Z., Cafarella M. Integrating Spreadsheet Data via Accurate and Low-Effort Extraction. In: Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. 2014. P. 1126–1135. DOI: 10.1145/2623330.2623617

3. SheetJS. SheetJS Community Edition: Documentation (версія 0.20.3) [Електронний ресурс]. URL: https://docs.sheetjs.com/docs/ (дата звернення: 25.09.2025).

4. Papa Parse. Documentation (версія 5.5.3) [Електронний ресурс]. URL: https://www.papaparse.com/docs (дата звернення: 25.09.2025).

5. Handsontable. JavaScript Data Grid – Documentation (версія 16.0.1) [Електронний ресурс]. URL: https://handsontable.com/docs/javascript-data-grid/ (дата звернення: 25.09.2025).

6. Huang Y., Miao J., Weng D., Perer A., Wu Y. StructVizor: Interactive Profiling of Semi-Structured Textual Data. In: CHI '25: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. 2025. DOI: 10.1145/3706598.3713484.

7. Berger B., Waterman M. S., Yu Y. W. Levenshtein distance, sequence comparison and biological database search. IEEE Transactions on Information Theory. 2021. Vol. 67, no. 6. P. 3287–3294. DOI: 10.1109/TIT.2020.2996543

8. Firmansyah M., Deswana D. Application of the Levenshtein Algorithm in Web-Based Knowledge Management Systems. JUSIFO. 2024. Vol. 10, no. 2. P. 99–106. DOI: 10.19109/jusifo.v10i2.21951

9. Harel D. Statecharts: a visual formalism for complex systems. Science of Computer Programming. 1987. Vol. 8, no. 3. P. 231–274. DOI: 10.1016/0167-6423(87)90035-9

10. Pohl E., Geldermann J. PROMETHEE-Cloud: a web app to support PROMETHEE-based multi-criteria decision analysis. EURO Journal on Decision Processes. 2024. Vol. 12, 100053. DOI: 10.1016/j.ejdp.2024.100053

11. Razavi Hajiagha S. H., Heidary Dahooie J., Meidutė-Kavaliauskienė I., Govindan K. A new dynamic multi-attribute decision making method based on Markov chain and linear assignment. Annals of Operations Research. 2022. Vol. 315, no. 1. P. 159–191. DOI: 10.1007/s10479-022-04644-0

12. Hamel A. H., Kostner D. Multi-weight ranking for multi-criteria decision making. Neural Computing and Applications. 2024. DOI: 10.1007/s00521-024-10626-z

13. Taherdoost H., Madanchian M. Multi-Criteria Decision Making (MCDM) Methods and Concepts. Encyclopedia. 2023. Vol. 3, no. 1. P. 77–87. DOI: 10.3390/encyclopedia3010006

14. Su J., Zhu Q., Jiang Y., Xiao X. A Multi-Criteria Group Decision-Making Method for Risk Assessment of Live-Streaming E-Commerce Platforms. Journal of Theoretical and Applied Electronic Commerce Research. 2023. Vol. 18, no. 2. P. 1126–1141. DOI: 10.3390/jtaer18020057

15. Győrödi C. A., Győrödi R., Olah A., Pecherle G. MongoDB vs. Document-Based MySQL: A Comparative Study on the Top of Docker. Big Data and Cognitive Computing. 2022. Vol. 6, no. 2. P. 49. DOI: 10.3390/bdcc6020049

16. Alyasiri B., Sahi B., Al-Khafaji N. NoSQL: Will it be an alternative to a relational database? MySQL vs MongoDB comparison. In: Proceedings of the 2nd International Multi-Disciplinary Conference: Integrated Sciences and Technologies (IMDC-IST 2021), 7–9 September 2021, Sakarya, Turkey. EAI, 2022. DOI: 10.4108/eai.7-9-2021.2314925

References

1. Guo, M., Xing, Y., & Xu, Y. (2023). Data cleaning: Workflow and strategies in real-world data research. Interactive Journal of Medical Research, 12, e44310. https://doi.org/10.2196/44310

2. Chen, Z., & Cafarella, M. (2014). Integrating spreadsheet data via accurate and low-effort extraction. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1126–1135). https://doi.org/10.1145/2623330.2623617

3. SheetJS. (2025). SheetJS Community Edition: Documentation (Version 0.20.3). Retrieved September 25, 2025, from https://docs.sheetjs.com/docs/

4. Papa Parse. (2025). Documentation (Version 5.5.3). Retrieved September 25, 2025, from https://www.papaparse.com/docs

5. Handsontable. (2025). JavaScript Data Grid – Documentation (Version 16.0.1). Retrieved September 25, 2025, from https://handsontable.com/docs/javascript-data-grid/

6. Huang, Y., Miao, J., Weng, D., Perer, A., & Wu, Y. (2025). StructVizor: Interactive profiling of semi-structured textual data. In CHI '25: Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems. https://doi.org/10.1145/3706598.3713484

7. Berger, B., Waterman, M. S., & Yu, Y. W. (2021). Levenshtein distance, sequence comparison and biological database search. IEEE Transactions on Information Theory, 67(6), 3287–3294. https://doi.org/10.1109/TIT.2020.2996543

8. Firmansyah, M., & Deswana, D. (2024). Application of the Levenshtein algorithm in web-based knowledge management systems. JUSIFO, 10(2), 99–106. https://doi.org/10.19109/jusifo.v10i2.21951

9. Harel, D. (1987). Statecharts: A visual formalism for complex systems. Science of Computer Programming, 8(3), 231–274. https://doi.org/10.1016/0167-6423(87)90035-9

10. Pohl, E., & Geldermann, J. (2024). PROMETHEE-Cloud: A web app to support PROMETHEE-based multi-criteria decision analysis. EURO Journal on Decision Processes, 12, 100053. https://doi.org/10.1016/j.ejdp.2024.100053

11. Razavi Hajiagha, S. H., Heidary Dahooie, J., Meidutė-Kavaliauskienė, I., & Govindan, K. (2022). A new dynamic multi-attribute decision making method based on Markov chain and linear assignment. Annals of Operations Research, 315(1), 159–191. https://doi.org/10.1007/s10479-022-04644-0

12. Hamel, A. H., & Kostner, D. (2024). Multi-weight ranking for multi-criteria decision making. Neural Computing and Applications. Advance online publication. https://doi.org/10.1007/s00521-024-10626-z

13. Taherdoost, H., & Madanchian, M. (2023). Multi-criteria decision making (MCDM) methods and concepts. Encyclopedia, 3(1), 77–87. https://doi.org/10.3390/encyclopedia3010006

14. Su, J., Zhu, Q., Jiang, Y., & Xiao, X. (2023). A multi-criteria group decision-making method for risk assessment of live-streaming e-commerce platforms. Journal of Theoretical and Applied Electronic Commerce Research, 18(2), 1126–1141. https://doi.org/10.3390/jtaer18020057

15. Győrödi, C. A., Győrödi, R., Olah, A., & Pecherle, G. (2022). MongoDB vs. document-based MySQL: A comparative study on the top of Docker. Big Data and Cognitive Computing, 6(2), 49. https://doi.org/10.3390/bdcc6020049

16. Alyasiri, B., Sahi, B., & Al-Khafaji, N. (2022). NoSQL: Will it be an alternative to a relational database? MySQL vs MongoDB comparison. In Proceedings of the 2nd International Multi-Disciplinary Conference: Integrated Sciences and Technologies (IMDC-IST 2021), 7–9 September 2021, Sakarya, Turkey. EAI. https://doi.org/10.4108/eai.7-9-2021.2314925

Дата першого надходження рукопису до видання: 27.09.2025
Дата прийнятого до друку рукопису після рецензування: 23.10.2025
Дата публікації:

УДК 004.415 DOI

О. Є. ОГНЄВА
кандидат технічних наук, доцент,
завідувач кафедри програмних засобів і технологій
Херсонський національний технічний університет
ORCID: 0000-0001-6206-0285

Д. В. ЗЯРКО
здобувач другого (магістерського) рівня вищої освіти
кафедри програмних засобів і технологій
Херсонський національний технічний університет
ORCID: 0009-0002-7103-2838

ДОСЛІДЖЕННЯ ЕФЕКТИВНОСТІ МОДЕЛЕЙ РЕІНЖЕНЕРІЇ ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ЦИФРОВОЇ ТРАНСФОРМАЦІЇ ТУРИСТИЧНОГО СЕРВІСУ

У статті досліджено ефективність моделей реінжинірингу програмного забезпечення як ключового інструменту цифрової трансформації туристичних інформаційних систем. Об'єктом дослідження є програмне забезпечення туристичної платформи, а предметом – процеси оптимізації її архітектури, продуктивності, безпеки та якості обслуговування користувачів. Особливу увагу приділено глибокому аналізу вихідного коду, реверс-інжинірингу, редизайну архітектури та впровадженню багаторівневих моделей із використанням сучасних технологій, таких як PHP, MySQL, OAuth, хешування паролів. Практична частина ґрунтується на реінжинірингу веб-застосунку Travel-website, що включає детальний аналіз коду, розробку оновленої архітектури, інтеграцію з Google OAuth, підвищення безпеки зберігання паролів та впровадження адаптивного інтерфейсу для покращення користувацького досвіду.

У статті обґрунтовано необхідність адаптації програмних систем до вимог цифрової трансформації, що є особливо актуальним в умовах жорсткої конкуренції та воєнного стану. Результати дослідження свідчать, що реінжиніринг сприяє значному підвищенню безпеки, гнучкості та продуктивності системи, що є критично важливим для сучасних туристичних сервісів. Впровадження багатофакторної автентифікації, а також інтеграція з CRM і аналітичними системами розширюють функціональні можливості платформи, забезпечуючи її конкурентоспроможність на ринку. Запропонована модель реінжинірингу є ефективним інструментом цифрової трансформації, що дозволяє підвищити якість та надійність туристичних інформаційних систем. Крім того, стаття розглядає питання оптимізації витрат і часу розробки за рахунок використання перевірених стратегій і технологій реінжинірингу, що робить запропонований підхід економічно вигідним та практично застосовним у сучасних умовах розвитку IT-індустрії.

Ключові слова: *реінжиніринг, програмне забезпечення, цифрова трансформація, архітектура, безпека, адаптивний інтерфейс.*

O. Ye. OHNIEVA
Candidate of Technical Sciences, Associate Professor,
Head of the Software Tools and Technologies Department
Kherson National Technical University
ORCID: 0000-0001-6206-0285

D. V. ZIARKO
Master's Degree Student at the Software Tools and Technologies Department
Kherson National Technical University
ORCID: 0009-0002-7103-2838

RESEARCH OF THE SOFTWARE REENGINEERING MODELS EFFECTIVENESS FOR THE TOURISM SERVICES DIGITAL TRANSFORMATION

The paper investigates the effectiveness of software reengineering models as a key tool for the digital transformation of tourism information systems. The object of the study is the software of a tourism platform, while the subject is the processes of optimizing its architecture, performance, security, and quality of user service. Special attention is given to

© Огнєва О. Є., Зярко Д. В., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

in-depth analysis of source code, reverse engineering, architecture redesign, and the implementation of multi-layered models using modern technologies such as PHP, MySQL, OAuth, and password hashing. The practical part is based on the reengineering of the Travel-website web application, which includes detailed code analysis, development of an updated architecture, integration with Google OAuth, enhancement of password storage security, and implementation of an adaptive interface to improve user experience.

The paper substantiates the necessity of adapting software systems to the requirements of digital transformation, which is especially relevant under conditions of intense competition and martial law. The research results indicate that reengineering significantly improves system security, flexibility, and performance, which is critically important for modern tourism services. The introduction of multi-factor authentication, as well as integration with CRM and analytical systems, expands the platform's functional capabilities, ensuring its competitiveness in the market. The proposed reengineering model is an effective tool for digital transformation, enabling the improvement of quality and reliability of tourism information systems. Furthermore, the article addresses the optimization of development costs and time through the use of proven reengineering strategies and technologies, making the proposed approach economically advantageous and practically applicable in the current IT industry environment.

Key words: *reengineering, software, digital transformation, architecture, security, adaptive interface.*

Постановка проблеми

Воєнний стан створює значні виклики для економіки, інфраструктури та соціальної сфери. Тому модернізація та оптимізація інформаційних систем стають критично важливими для підтримки та розвитку туристичної галузі. Реінжиніринг інформаційної платформи для продажу турів має особливе значення для України в умовах воєнного стану, оскільки забезпечує адаптивність і стійкість туристичного бізнесу в кризових умовах.

Реінжиніринг інформаційної платформи виступає стратегічним інструментом, що забезпечує технологічні зміни, підвищення ефективності бізнес-процесів і, як наслідок, поліпшення якості туристичних послуг, що відповідає сучасним вимогам цифрової економіки та очікуванням споживачів. Об'єктом дослідження є програмне забезпечення туристичної платформи, а предметом – процеси реінженерії, спрямовані на оптимізацію архітектури системи, підвищення продуктивності, покращення заходів безпеки та підвищення загальної якості обслуговування користувачів.

В умовах стрімкої цифрової трансформації та зростання вимог до якості інформаційних сервісів у туристичній галузі особливої актуальності набуває питання реінженерії програмного забезпечення (ПЗ). Програмні системи туристичних платформ часто мають застарілу архітектуру, низьку масштабованість та обмежені можливості інтеграції з сучасними сервісами, що ускладнює впровадження нових функцій та підтримку високого рівня безпеки. Тому виникає потреба у застосуванні сучасних моделей реінженерії ПЗ, які дозволяють оптимізувати архітектуру, підвищити продуктивність, забезпечити захист даних та покращити користувацький досвід.

Аналіз останніх досліджень і публікацій

Українські науковці досліджують особливості реінжинірингу та розробляють концептуальні засади здійснення реінжинірингу бізнес-процесів у програмних системах. Так в праці [1] розглядаються засоби редокументування успадкованого програмного забезпечення, в праці [2] підкреслюється ключова роль інформаційних технологій у реінжинірингу бізнес-процесів. Автором в [3] досліджується генезис реінжинірингу бізнес-процесів як елемент формування ефективної системи управління на промислових підприємствах. Автори [4] аналізують особливості реінжинірингу в організаціях, що розробляють програмні продукти, з акцентом на витрати, переваги та труднощі, а в [5] – розробляють концептуальні засади реінжинірингу бізнес-процесів у програмних системах. В роботі [6] були розглянути теоретично-методичні основи реінжинірингу бізнес-процесів. Авторами [7] реінжиніринг представляється як інноваційна основа реконструкції системи управління підприємством.

Іноземні науковці також заклали теоретичні та практичні основи реінжинірингу ПЗ, охоплюючи питання модернізації застарілих систем, радикальної перебудови процесів та стратегічних аспектів впровадження інформаційних технологій. Робота [8] описує методи рефакторингу для покращення існуючого коду, в [9] пропонується модель процесу реінжинірингу ПЗ, що поєднує традиційний підхід зі спіральною моделлю, в [10] розглядаються принципи перепроектування ПЗ для гетерогенних платформ, в [11] досліджується реінжиніринг застарілих систем у напрямку нових технологій, в [12] описується впровадження інновацій у процеси через реінжиніринг роботи за допомогою інформаційних технологій, в [13] надаються практичні поради щодо ефективної роботи з успадкованим кодом. Авторами в праці [14] надають огляд концепцій, моделей і підходів до реінжинірингу ПЗ, в [15] – викладають маніфест революції в бізнесі через реінжиніринг корпорацій, а в [16] – описують метод синтезу специфікацій і вимог у процесі реінжинірингу ІТ-проєктів. Автор [17] пропонує модель процесу реінжинірингу ПЗ. Різні аспекти інженерії ПЗ, включно з реінжинірингом розглядаються в [18]. Методи реверсивного інжинірингу описуються в [19].

Дослідники розглядають різноманітні підходи до реінженерії ПЗ: від реверс-інжинірингу до впровадження мікросервісної архітектури, використання CI/CD, DevOps, автоматизованого тестування, а також застосування UML-діаграм для моделювання систем. Проте, недостатньо досліджено питання комплексної реінженерії туристичних платформ із врахуванням специфіки галузі та сучасних вимог до якості ПЗ.

Формулювання мети дослідження

Обґрунтування ефективної моделі реінжинірингу бізнес-процесів туристичної платформи, спрямованої на підвищення якості обслуговування клієнтів, оптимізацію процесу продажу турів та впровадження сучасних цифрових технологій для покращення взаємодії з користувачами, що досягається за рахунок переходу до тришарової архітектури з MySQL, хешуванням паролів, валідацією даних та захистом від SQL-ін'єкцій.

Викладення основного матеріалу дослідження

Реінжиніринг у цьому дослідженні передбачав повний цикл: від аналізу вихідного коду до впровадження сучасної архітектури та нових функцій. Реінжиніринг інформаційних платформ для продажу турів є ефективним фактором підвищення якості туристичних послуг через комплексне удосконалення бізнес-процесів, застосування інноваційних технологій і оптимізацію маркетингових комунікацій [1, 4, 6, 10–12, 15].

Для дослідження була обрана власна програма – веб-додаток Travel-website, розміщений у відкритому доступі на GitHub за посиланням: https://github.com/EienZ/Travel-website.

Реверсивна інженерія охопила аналіз архітектури, функціоналу та структури бази даних веб-додатку. Результати аналізу веб-додатку Travel-website наведено в табл.1. Архітектура додатку реалізована за класичною схемою клієнт-сервер: в клієнтської частині браузер користувача взаємодіє із сервером через HTML-форми, в серверної частині Apache з PHP обробляє запити через register.php, login.php та інші скрипти, база даних MySQL зберігає дані користувачів, дані передаються у вигляді POST-запитів, обробляються на сервері, і результат передається назад клієнту у вигляді HTML. Для моделювання нової архітектури було створено UML-діаграми класів та варіантів використання, що дозволило чітко визначити взаємодію між модулями системи.

Таблиця 1

Результати аналізу веб-додатку Travel-website

Елемент аналізу	Стан
Безпека	Відсутнє хешування паролів, немає захисту від SQL-ін'єкцій
Документація	Відсутня
База даних	Мінімальна, відсутні обмеження на email, немає індексації
Структура коду	Модулі login і register реалізовані в одному файлі кожен
Валідація даних	Відсутня перевірка на стороні сервера і клієнта
Структура проекту	Простий файл-орієнтований підхід без розділення на логіку/представлення
Вихідні тести	Ручне тестування
Інтерфейс користувача	Присутній базовий HTML, з проведеним ручним тестуванням.

В результаті дослідження додатку було прийняте рішення про покращення основних функцій модуля логіну та реєстрації користувачів. Зміни сприяють підвищенню безпеки, зручності та масштабованості системи. Паролі зберігаються в зашифрованому вигляді, що значно покращує захист даних. Реалізація входу через сторонні сервіси, такі як Google, спрощує процес авторизації для користувачів, підвищуючи зручність використання. Оновлена структура бази даних забезпечує гнучкість та можливість легко додавати нові методи авторизації. Всі ці зміни зробили систему більш надійною та зручною для кінцевих користувачів.

Впровадження хешування паролів забезпечує надійний захист збережених даних користувачів, зменшуючи ризик несанкціонованого доступу навіть у разі витоку бази даних. Поєднання хешування з протоколом OAuth дозволяє делегувати аутентифікацію надійним зовнішнім провайдерам (Google, Facebook), що додатково знижує ризики фішингових атак і крадіжки облікових даних. OAuth також забезпечує безпечний обмін інформацією між сервісами без передачі паролів [8–10, 13–14].

Інтеграція входу через Google одним кліком значно спрощує процес реєстрації та авторизації для користувачів, усуваючи необхідність запам'ятовувати додаткові паролі та вводити багато полів форми. Це підвищує конверсію реєстрацій і зменшує кількість незавершених сесій, що позитивно впливає на користувацький досвід і лояльність клієнтів. Крім того, використання соціальних мереж і єдиних акаунтів сприяє швидшому доступу до персоналізованих пропозицій.

Архітектура системи, побудована на стандарті OAuth, дозволяє легко додавати нових провайдерів аутентифікації, таких як Facebook, GitHub, Apple ID тощо, без суттєвих змін у базовому коді платформи. Це забезпечує гнучкість розвитку системи, адаптацію до потреб різних сегментів користувачів і підтримку мультиплатформеності. Масштабованість також включає можливість інтеграції з CRM-системами та аналітичними інструментами для покращення маркетингових комунікацій і персоналізації [8–10, 17].

Після аналізу успадкованої програми (модулів логіну та реєстрації з підтримкою OAuth) виявлено такі основні вимоги до відновленої системи:

– система повинна дозволяти гостеві користувачу: зареєструватися через email/пароль; увійти через email/пароль; увійти через сторонні сервіси (Google, Facebook); відновити пароль через email; після входу користувач отримує доступ до захищених сторінок;

– нефункціональні вимоги: система повинна відповісти на запит входу/реєстрації не пізніше ніж за 2 секунди; інтерфейс має бути адаптивним для мобільних пристроїв; всі паролі мають зберігатися у хешованому вигляді; серверний API – HTTPS із сертифікатом.

Після аналізу моделі вимог було встановлено, що оновлена програма повинна забезпечувати автентифікацію користувачів через email/пароль та сторонні сервіси (Google, Facebook), мати просту навігацію, безпечне зберігання даних та зручний інтерфейс.

Архітектура була трансформована у класичну тришарову модель:

– фронденд: HTML/CSS/JS-інтерфейс для взаємодії з користувачем.

– бекенд: PHP-скрипти для обробки запитів, логіки автентифікації та інтеграції з OAuth-сервісами.

– БД: MySQL для зберігання облікових записів користувачів, логів і токенів сесій.

В системі реалізовано шифрування паролів, валідацію введених даних, обробку помилок та захист від SQL-ін'єкцій. У ході дослідження було успішно трансформовано архітектуру старої автентифікаційної системи у сучасну реалізацію з використанням HTML, CSS, JavaScript, PHP та MySQL. Вона підтримує як класичний вхід через email/пароль, так і авторизацію через Google та Facebook. Архітектура є масштабованою, безпечною та гнучкою, що дозволяє легко її модифікувати чи розширювати у майбутньому.

Таблиця 2

Критерії якості та завершення тестування

Критерій	Очікувана поведінка
Безпека	Паролі не зберігаються у відкритому вигляді
Відсутність збоїв	При неправильному введенні не відбувається аварійне завершення
Відповідність вимогам	Реалізовано всі сценарії, передбачені моделлю варіантів використання
Сумісність	Код працює у середовищі Nginx + PHP + MySQL
Розширюваність	Можливість додати Facebook OAuth або двофакторну автентифікацію

Таблиця 3

Сценарії тестування

№	Назва сценарію	Вхідні дані	Очікуваний результат
1	Реєстрація нового користувача	Email: test@mail.com, Password: 12345678	Перенаправлення на сторінку входу
2	Вхід з правильними даними	Email: test@mail.com, Password: 12345678	Вхід у систему (сесія встановлена)
3	Вхід з неправильним паролем	Email: test@mail.com, Password: wrong	Повідомлення про помилку входу
4	Вхід через Google	Google акаунт (валідний)	Вхід у систему через Google
5	SQL-ін'єкція	Email: ' OR 1=1 --, Password: anything	Невдалий вхід, захист спрацював
6	Перевірка шифрування паролю	Перевірка запису в БД	Пароль збережено як хеш
7	Повторна реєстрація	Email: вже існуючий	Повідомлення про існуючого користувача
8	Доступ без входу	Перехід на dashboard без сесії	Перенаправлення на форму входу

Таблиця 4

Результати тестування

Сценарій	Статус	Примітка
1	Пройдено	Реєстрація працює
2	Пройдено	Вхід з коректними даними
3	Пройдено	Повідомлення про помилку при невірному паролі
4	Пройдено	Авторизація через Google успішна
5	Пройдено	SQL-ін'єкція заблокована
6	Пройдено	Перевірено: bcrypt-хешування збережено у БД
7	Пройдено	Повідомлення про існуючого користувача виведено
8	Пройдено	Перенаправлення при відсутності сесії працює

У процесі перепроєктування програми було здійснено деталізацію компонентів та реалізацію логічної структури, алгоритмів взаємодії й структури розгортання (рис. 1–3). Побудована модель чітко відображають взаємозв'язки між компонентами, підтримують цілісність архітектури.

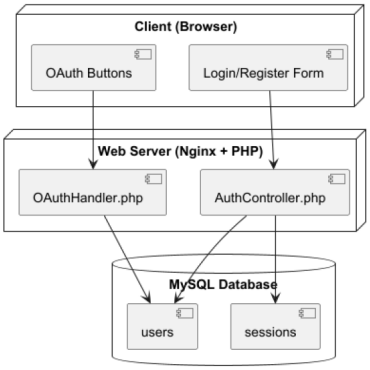


Рис. 1. Модель розгортання

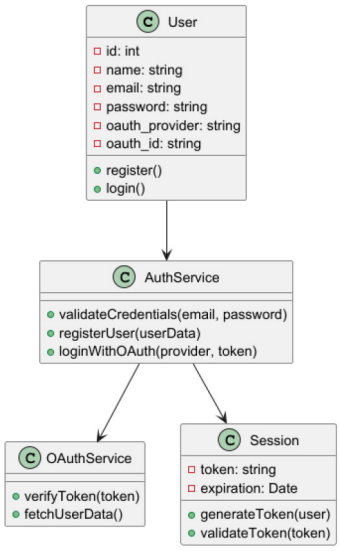


Рис. 2. Структурно-логічна модель (класова діаграма)

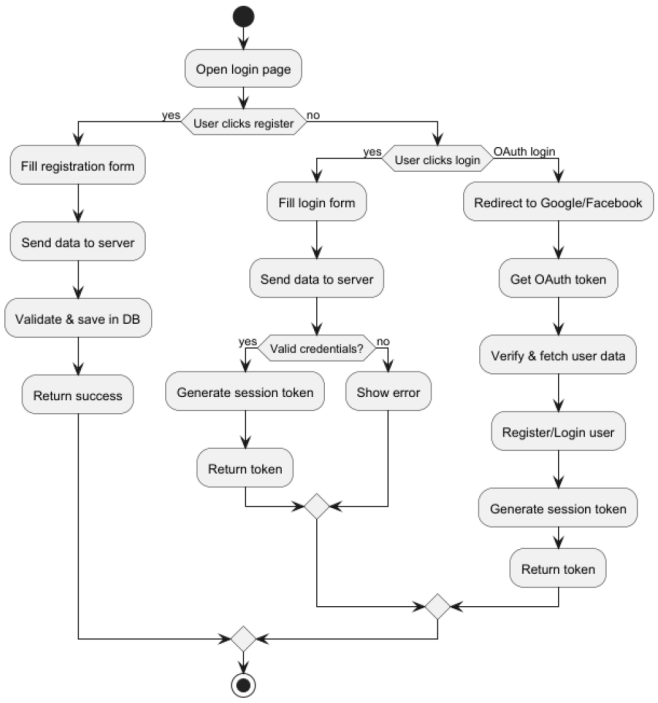


Рис. 3. Процедурна (функціональна) модель

У дослідженні реалізовано обновлення системи лога в додатку туристичного агентства:
– багатофункціональну систему автентифікації;
– шифрування паролів із використанням password_hash;
– інтеграцію з Google OAuth 2.0;
– MVC-структурування логіки.

Це дозволяє користувачам мати гнучкий та безпечний спосіб входу. Оновлена система автентифікації повністю відповідає заявленим вимогам, є безпечною, стійкою до атак, а також підтримує аутентифікацію зі сторонніх ресурсів (табл. 2–4).

Впровадження хешування паролів забезпечує надійний захист збережених даних користувачів, зменшуючи ризик несанкціонованого доступу навіть у разі витоку бази даних. Поєднання хешування з протоколом OAuth дозволяє делегувати аутентифікацію надійним зовнішнім провайдерам (Google, Facebook), що додатково знижує ризики фішингових атак і крадіжки облікових даних. OAuth також забезпечує безпечний обмін інформацією між

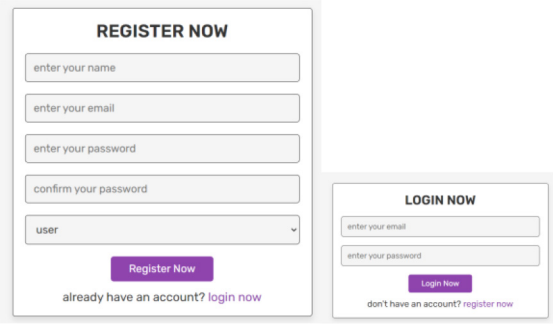


Рис. 4. Старий дизайн

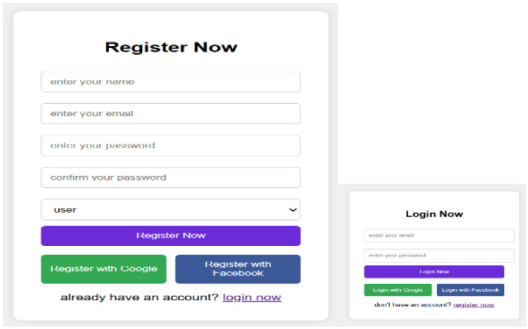


Рис. 5. Новий дизайн

сервісами без передачі паролів. Паролі тепер шифруються за допомогою password_hash, який використовує bcrypt для генерації безпечного хешу. Валідація забезпечує очищення даних під час створення.

Схема реінжинірингу [16–19], яка була використана в роботі, може бути описана наступним чином:

1. Реверсивна інженерія (Reverse Engineering):
– аналіз і документування існуючого функціоналу модулів реєстрації та логіна веб-додатка Travel-website;
– виявлення архітектурних недоліків (відсутність безпеки, погана структура коду);
– документування структури бази даних;
– фіксація наявного інтерфейсу та сценаріїв взаємодії користувача.
2. Аналіз вимог:
– визначення функціональних та нефункціональних вимог до оновленої системи;
– функціональні вимоги включали можливість реєстрації/входу через email/пароль та сторонні сервіси (Google, Facebook), відновлення паролю, доступ до захищених сторінок після входу;
– нефункціональні вимоги включали швидкість відповіді системи, адаптивність інтерфейсу, безпечне зберігання паролів та захищений API (HTTPS).
3. Перепроектування (Re-design):
– трансформація архітектури у класичну тришарову модель (фронтенд, бекенд, база даних);
– використання HTML/CSS/JS для інтерфейсу, PHP для обробки запитів та MySQL для зберігання даних;
– реалізація шифрування паролів, валідації даних, обробки помилок та захисту від SQL-ін’єкцій.
4. Реалізація (Re-implementation):
– реалізація нового дизайну інтерфейсу (рис. 5);
– розробка нового функціоналу автентифікації з підтримкою OAuth;
–деталізація компонентів, реалізація логічної структури, алгоритмів взаємодії та структури розгортання (рис. 1–3).
5. Тестування:
– визначення критеріїв якості та завершення тестування (безпека, відсутність збоїв, відповідність вимогам, сумісність, розширюваність).
– проведення тестування за різними сценаріями (табл. 3).
– аналіз результатів тестування та виправлення виявлених дефектів (табл. 4).

Ця схема поєднує елементи реверсивної інженерії, аналізу вимог, перепроектування, реалізації та тестування, що дозволяє комплексно модернізувати застарілий веб-додаток Travel-website та підвищити його безпеку, зручність і масштабованість.

Висновки

Реінжиніринг веб-додатку Travel-website включав аналіз існуючої архітектури, структури бази даних та функціоналу модулів реєстрації й логіну. Основними напрямами модернізації стали впровадження тришарової архітектури, використання хешування паролів, інтеграція з OAuth для авторизації через Google та підвищення адаптивності інтерфейсу. Це дозволило забезпечити безпечну автентифікацію, гнучкість розширення та інтеграцію з зовнішніми сервісами. Оновлена система відповідає сучасним вимогам до продуктивності, безпеки та зручності для користувачів.

Така модель реінженерії ПЗ для туристичної платформи відповідає сучасним вимогам інженерії програмного забезпечення, забезпечує високу якість, підтримуваність і конкурентоспроможність системи. Результати можуть бути використані для модернізації ІТ-систем у туристичній галузі та інших сферах.

Перспективи досліджень в даному напрямку у подальшому будуть спрямовані на практичне впровадження та тестування запропонованої моделі реінженерії у різних предметних галузях, зокрема в умовах специфіки українського ринку. Важливим перспективним напрямом є вивчення інтеграції запропонованої моделі з сучасними технологіями штучного інтелекту та машинного навчання для подальшого підвищення інтелектуальності та адаптивності програмних систем.

Список використаної літератури

1. Авраменко О. А. Архітектура засобів редокументування успадкованого програмного забезпечення. *Вісник НАУ*. 2020. № 3–4 (33). С. 58–62.

2. Гвоздь М. Я. Ключова роль інформаційних технологій під час реінжинірінгу бізнес-процесів. *Вісник Національного університету «Львівська політехніка». Серія: Проблеми економіки та управління*. 2015. № 815. С. 202–206. URL: http://nbuv.gov.ua/UJRNVNULPP_2015_815_27

3. Гриценко Л. Л. Генезис реінжинірінгу бізнес-процесів як елемента формування ефективної системи управління в діяльності промислових підприємств. *Механізм регулювання економіки*. 2015. № 2. С. 31–37.

4. Каризська А. Р., Ситник Н. І. Особливості реінжинірінгу в організаціях, що займаються розробленням програмного продукту. *Актуальні проблеми економіки та управління: збірник наукових праць молодих вчених*. 2017. Вип. 11. URL: https://ela.kpi.ua/server/api/core/bitstreams/c5472d16-13b6-4432-89e6-25b716d0b68a/content

5. Ковальов А. І., Шевченко В. В. Концептуальні засади здійснення реінжинірінгу бізнес-процесів. *Вісник соціально економічних досліджень: зб. наук. пр. Одеський держ. екон. ун-т., Одеса*. 2009. Вип. 35. С. 365–371.

6. Продіус О. І. Теоретично-методичні основи реінжинірінгу бізнес-процесів. Економіка: реалії часу. *Науковий журнал*. 2016. № 6 (28). С. 79–87. URL: http://economics.opu.ua/files/archive/2016/n6.html

7. Прохорова В. В., Божанова О. В., Юхман Я. В. Реінжиніринг як інноваційна основа реконструкції системи управління підприємством. *Проблеми економіки*. 2020. № 4. С. 210–216. URL: http://jnas.nbuv.gov.ua/article/UJRN-0001219373

8. Фаулер М. та ін. Рефакторинг. Поліпшення існуючого коду. «Діалектика». 2019. 448 с.

9. Abbas A. S., Jeberson W., Jeberson D. W., Klinsega V. V. Proposed Software Re-engineering Process That Combine Traditinal Software Re-engineering Process With Spiral Model. *International Journal of Advanced Research in Computer Science*. 2013. Vol. 4 (2). Pp. 75–83.

10. Andrade H., Berger C., Crnkovic I., Bosch J. Principles for Rearchitecting Software for Heterogeneous Platforms. *2020 27th AsiaPacific Software Engineering Conference (APSEC)*. 2020. Pp. 405– 414. DOI: 10.1109/apsec51365.2020.00049

11. Bouchiha D. Reengineering Legacy Systems Towards New Technologies. *Encyclopedia of Information Science and Technology, Fifth Edition, IGI Global*. 2020. Pp. 1214–1230. DOI: 10.4018/978-1- 7998-3479-3.ch084

12. Davenport T. Process Innovation: Reengineering Work through Information Technology. Harvard Business School Press, Boston. 1993. 337 p.

13. Feathers M. Working Effectively with Legacy Code. Person. 2020. 464 p.

14. Fung L., Tan S.Overview of Software Re-Engineering Concepts, Models and Approaches. *JOIV: International Journal on Informatics Visualization*. 2025. № 9. Pp. 46. DOI:10.62527/joiv.9.1.3034

15. Hammer M., Champy J. Reengineering the Corporation: Manifesto for Business Revolution. Zondervan. 2009. 332 p.

16. Kernytskyy O., Kernytskyy A., Teslyuk V. The Synthesis Method for Specifications and Requirements in the Process of IT Project Reengineering. *International Scientific and Technical Conference on Computer Sciences and Information Technologies, Institute of Electrical and Electronics Engineers Inc*. 2023. № 5. Pp.1~8. DOI:10.1109/CSIT61576.2023.10324175

17. Kumar A. Software Re-engineering Process Model. *International Journal of Science and Research*. 2019. Pp. 2319–7064. DOI:10.21275/SR21101152601

18. Sommerville I. Software engineering. Pearson Education. 2007. 792 p.

19. Team N. Reverse Engineering. *Handbook for CTFers, Singapore: Springer Nature Singapore*. 2022. Pp. 295–427. DOI: 10.1007/978-981- 19-0336-6_5

References

1. Avramenko, O. A. (2020). Architecture of Re-documentation Tools for Legacy Software. *Proceedings of the National Aviation University*, 3–4 (33), 58–62.

2. Hvozd, M. Ia. (2015). Key Role of Information Technology in Business Process Reengineering. *Bulletin of the Lviv Polytechnic National University. Series: Problems of Economics and Management*, 815, 202–206. URL: http://nbuv.gov.ua/UJRN/VNULPP_2015_815_27

3. Hrytsenko, L. L. (2015). The Genesis of Business Process Reengineering as an Element of Forming an Effective Management System in the Activities of Industrial Enterprises. *Mechanism of Economic Regulation*, 2, 31–37.

4. Karyzska, A. R., & Sytnyk, N. I. (2017). Features of Reengineering in Organizations Engaged in Software Developmen. *Current Issues in Economics and Management: Collection of Scientific Papers by Young Researchers*, 11. URL: https://ela.kpi.ua/server/api/core/bitstreams/c5472d16-13b6-4432-89e6-25b716d0b68a/content

5. Kovalov, A. I., & Shevchenko, V. V. (2009). Conceptual Foundations of Business Process Reengineering. *Bulletin of Socio-Economic Research: Collection of Scientific Papers of Odesa State Economic University. Odesa*, 35, 365–371.

6. Prodius, O. I. (2016). Theoretical and Methodological Foundations of Business Process Reengineering. Economics: Realities of the Time. *Academic Journal*, 6 (28), 79–87. URL: http://economics.opu.ua/files/archive/2016/n6.html

7. Prokhorova, V. V., Bozhanova, O. V., & Iukhman, Ya. V. (2020). Reengineering as an Innovative Basis for the Reconstruction of Enterprise Management Systems. *Problems of Economics*, 4, 210–216. URL: http://jnas.nbuv.gov.ua/article/UJRN-0001219373

8. Fauler, M. (2019). Refactoring. Improving Existing Code, "Dialektyka", 448.

9. Abbas, A. S., Jeberson, W., Jeberson, D. W., & Klinsega, V. V. (2013). Proposed Software Re-engineering Process That Combine Traditinal Software Re-engineering Process With Spiral Model. *International Journal of Advanced Research in Computer Science*, 4 (2), 75–83.

10. Andrade, H., Berger, C., Crnkovic, I., & Bosch, J. (2020). Principles for Rearchitecting Software for Heterogeneous Platforms. *2020 27th AsiaPacific Software Engineering Conference (APSEC)*, 405–414. DOI: 10.1109/apsec51365.2020.00049

11. Bouchiha, D. (2020). Reengineering Legacy Systems Towards New Technologies. *Encyclopedia of Information Science and Technology, Fifth Edition, IGI Global*, 1214–1230. DOI: 10.4018/978-1-7998-3479-3.ch084

12. Davenport, T. (1993). Process Innovation: Reengineering Work through Information Technology. Harvard Business School Press, Boston, 337.

13. Feathers M. (2020) Working Effectively with Legacy Code. Person, 464.

14. Fung, L., & Tan, S. (2025). Overview of Software Re-Engineering Concepts, Models and Approaches. *JOIV: International Journal on Informatics Visualization*, 9, 46. DOI:10.62527/joiv.9.1.3034

15. Hammer, M., & Champy, J. (2009). Reengineering the Corporation: Manifesto for Business Revolution. Zondervan, 332.

16. Kernytskyy, O., Kernytskyy, A., & Teslyuk, V. (2023). The Synthesis Method for Specifications and Requirements in the Process of IT Project Reengineering. *International Scientific and Technical Conference on Computer Sciences and Information Technologies, Institute of Electrical and Electronics Engineers Inc*, 5, 1–8. DOI:10.1109/CSIT61576.2023.10324175

17. Kumar, A. (2019). Software Re-engineering Process Model. *International Journal of Science and Research*, 2319–7064. DOI:10.21275/SR21101152601

18. Sommerville, I. (2007). Software engineering. Pearson Education, 792.

19. Team, N. (2022). Reverse Engineering. *Handbook for CTFers, Singapore: Springer Nature Singapore*, 295–427. DOI: 10.1007/978-981- 19-0336-6_5

Дата першого надходження рукопису до видання: 19.09.2025
Дата прийнятого до друку рукопису після рецензування: 15.10.2025
Дата публікації:

УДК 004.89:612.13:681.3 DOI

О. А. ПАСТУХ
доктор технічний наук, професор,
професор кафедри програмної інженерії
Тернопільський національний технічний університет імені Івана Пулюя
ORCID: 0000-0002-0080-7053

О. М. БРИК
аспірант кафедри програмної інженерії
Тернопільський національний технічний університет імені Івана Пулюя
ORCID: 0009-0005-6564-1102

ІНТЕРПРЕТОВАНА КЛАСИФІКАЦІЯ ГЕМОДИНАМІЧНОГО СТАНУ НА ОСНОВІ ВІДБОРУ КЛЮЧОВИХ НЕІНВАЗИВНИХ ПАРАМЕТРІВ ТА АНАЛІЗУ ВАЖЛИВОСТІ ОЗНАК

У статті розглянуто проблему інтерпретованої класифікації гемодинамічного стану людини на основі відбору ключових неінвазивних параметрів та аналізу важливості ознак. Сучасна медицина все більше орієнтується на створення автоматизованих систем моніторингу життєво важливих функцій організму, що базуються на методах машинного навчання та штучного інтелекту. Однак значною перешкодою у впровадженні таких систем є проблема довіри з боку лікарів та пацієнтів, оскільки більшість моделей машинного навчання працюють як «чорні скриньки», без зрозумілої логіки прийняття рішень. Це обмежує практичну цінність навіть високо точних алгоритмів, адже клінічні спеціалісти потребують не лише результату класифікації, а й аргументованого пояснення, чому саме той чи інший стан визначено.

Метою дослідження є створення інтерпретованої моделі класифікації гемодинамічного стану із застосуванням відбору найбільш інформативних неінвазивних фізіологічних параметрів та методів пояснення моделей. У роботі використано такі показники: частота серцевих скорочень, артеріальний тиск, варіабельність серцевого ритму, насичення крові киснем (SpO_2) та частота дихання. На основі експериментів з моделями Random Forest, XGBoost, логістичною регресією та інтерпретованими деревами рішень здійснено порівняльний аналіз ефективності класифікації та зрозумілості результатів. Для оцінки внеску окремих ознак застосовано метод SHAP-аналізу, що дозволив кількісно визначити значущість параметрів у процесі прогнозування.

Результати дослідження свідчать, що найбільш вагомими ознаками для оцінки гемодинамічного стану є насичення крові киснем (SpO_2) та варіабельність серцевого ритму (HRV). Їх поєднання забезпечує найбільшу діагностичну інформативність у виявленні компенсованих та декомпенсованих станів. Частота серцевих скорочень та артеріальний тиск відіграють роль додаткових предикторів, тоді як частота дихання має допоміжне значення. Використання інтерпретованих алгоритмів дозволило сформувати правила класифікації у вигляді зрозумілих для лікаря залежностей, що значно підвищує рівень довіри до системи та створює передумови для практичного застосування.

Таким чином, запропонований підхід поєднує високу точність автоматизованої діагностики з прозорістю та пояснюваністю результатів. Це робить його перспективним для впровадження у системи підтримки клінічних рішень, мобільні сенсорні пристрої та телемедичні платформи. Подальші дослідження доцільно спрямувати на тестування запропонованої моделі на розширених клінічних вибірках, а також на інтеграцію з інтелектуальними інформаційними системами для персоналізованої медицини.

Ключові слова: *гемодинамічний стан, машинне навчання, інтерпретовані моделі, неінвазивні параметри, SHAP-аналіз, пояснюваний штучний інтелект, класифікація.*

O. A. PASTUKH
D.Sc., Professor at the Department of Software Engineering
Ternopil Ivan Puluj National Technical University
ORCID: 0000-0001-6341-1710

O. M. BRYK
Postgraduate Student at the Department of Software Engineering
Ternopil Ivan Puluj National Technical University
ORCID: 0009-0002-6204-8707

© Пастух О. А., Брик О. М., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

INTERPRETABLE CLASSIFICATION OF HEMODYNAMIC STATE BASED ON THE SELECTION OF KEY NON-INVASIVE PARAMETERS AND FEATURE IMPORTANCE ANALYSIS

The article addresses the problem of interpretable classification of the hemodynamic state based on the selection of key non-invasive parameters and feature importance analysis. Modern medicine increasingly relies on the development of automated systems for monitoring vital functions of the human body, using machine learning and artificial intelligence methods. However, the implementation of such systems faces a major challenge related to trust from physicians and patients, since most machine learning models act as "black boxes" without providing understandable reasoning behind their predictions. This significantly limits the clinical value of even highly accurate algorithms, as medical specialists require not only the classification result but also a justified explanation of why a certain condition was determined.

The aim of this study is to develop an interpretable classification model for assessing the hemodynamic state, applying feature selection techniques to identify the most informative non-invasive physiological parameters and using model explanation methods. The investigated parameters include heart rate, blood pressure, heart rate variability, blood oxygen saturation (SpO_2), and respiratory rate. Experiments were conducted with Random Forest, XGBoost, Logistic Regression, and interpretable decision tree models to compare the classification performance and interpretability of results. SHAP analysis was employed to quantify the contribution of each feature to the model's decision-making process.

The results demonstrate that blood oxygen saturation (SpO_2) and heart rate variability (HRV) are the most significant features for evaluating the hemodynamic state. Their combination provides the highest diagnostic informativeness in distinguishing between compensated and decompensated conditions. Heart rate and blood pressure act as additional predictors, while respiratory rate plays a supplementary role. The use of interpretable algorithms made it possible to formulate classification rules expressed in clinically understandable dependencies, which significantly improves the trustworthiness of the system and lays the foundation for its practical implementation.

Thus, the proposed approach combines high diagnostic accuracy with transparency and explainability, making it a promising tool for integration into clinical decision support systems, mobile sensor devices, and telemedicine platforms. Future research should focus on testing the proposed model on extended clinical datasets and integrating it into intelligent information systems for personalized medicine.

Key words: *hemodynamic state, machine learning, interpretable models, non-invasive parameters, SHAP analysis, explainable artificial intelligence, classification.*

Постановка проблеми

Оцінка гемодинамічного стану є однією з ключових задач у клінічній практиці, оскільки вона безпосередньо пов'язана з виявленням ризиків серцево-судинних ускладнень, контролем перебігу хронічних захворювань та моніторингом стану пацієнтів у відділеннях інтенсивної терапії. Традиційні методи визначення гемодинамічних показників (катетеризація серця, інвазивне вимірювання артеріального тиску тощо) забезпечують високу точність, однак є ризикованими для пацієнтів та потребують спеціалізованого обладнання й персоналу [1, 2].

У зв'язку з цим, значного поширення набули неінвазивні технології моніторингу, такі як фотоплетизмографія, електрокардіографія, пульсоксиметрія, аналіз варіабельності серцевого ритму. Вони дозволяють отримувати достовірні параметри без ризику ускладнень, а також можуть бути інтегровані у портативні та носимі пристрої [3, 4]. Проте, попри доступність даних, виникає проблема їх обробки та інтерпретації: велика кількість сигналів та показників потребує застосування сучасних алгоритмів обробки та класифікації.

Методи машинного навчання вже продемонстрували високу ефективність у задачах діагностики та прогнозування станів пацієнтів [5, 6]. Зокрема, моделі на основі ансамблів дерев (Random Forest, Gradient Boosting) і глибинного навчання досягають високих показників точності. Проте їхнє впровадження у медичну практику стримується тим, що результати здебільшого залишаються «чорними скриньками» – лікар не отримує зрозумілої логіки прийняття рішення [7, 8]. Це створює бар'єр для довіри з боку медичного персоналу та обмежує практичну цінність навіть високо точних рішень.

Таким чином, постає завдання розробки методів інтерпретованої класифікації гемодинамічного стану, які б поєднували точність машинного навчання з прозорістю і зрозумілістю для лікаря. Особливої актуальності набуває відбір найбільш інформативних неінвазивних параметрів, а також використання підходів для аналізу вагомості ознак (наприклад, SHAP, LIME), які дозволяють кількісно оцінити внесок кожної змінної у прийняття рішення [9, 10].

Аналіз останніх досліджень і публікацій

У сучасній науковій літературі активно досліджується проблема моніторингу та класифікації гемодинамічного стану пацієнтів за допомогою методів машинного навчання та штучного інтелекту. Значна увага приділяється використанню неінвазивних показників як джерела діагностичної інформації, оскільки вони можуть забезпечити достатню точність при мінімальному ризику для пацієнта. Зокрема, дослідження показують, що частота серцевих скорочень, артеріальний тиск, варіабельність серцевого ритму та рівень насичення крові киснем (SpO_2) є ключовими параметрами для оцінки функціонального стану серцево-судинної системи [1, 2].

Вітчизняні дослідники також приділяють увагу даному напрямку. У працях (Кукурудза Г. В. та ін., 2021; Бойко І. В. та ін., 2018) розглядаються алгоритмічні та апаратні аспекти аналізу електрофізіологічних сигналів, а також можливості їхньої інтеграції з інтелектуальними системами підтримки прийняття рішень.

Окремий напрям робіт стосується застосування методів машинного навчання для побудови моделей прогнозування станів пацієнтів. Наприклад, у дослідженні Johnson et al. [3] продемонстровано ефективність алгоритмів класифікації у виявленні критичних станів у відділеннях інтенсивної терапії. Rajkomar et al. [4] підкреслюють, що алгоритми на основі глибинних нейронних мереж можуть досягати високих показників точності при обробці медичних даних, однак проблема інтерпретації їхніх рішень залишається невирішеною.

Проблема пояснюваності моделей стала однією з ключових у сфері штучного інтелекту в медицині. Rudin [5] та Holzinger et al. [6] наголошують на необхідності застосування інтерпретованих підходів для рішень, що приймаються у критично важливих сферах, зокрема у діагностиці та лікуванні. Серед поширених методів пояснення рішень машинного навчання варто відзначити SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations) та побудову інтерпретованих дерев рішень [7, 8]. Ці методи дозволяють оцінювати внесок окремих ознак у прогнозування та формувати більш зрозумілу для лікаря картину.

В останні роки зростає інтерес до поєднання моделей високої точності та прозорості. Наприклад, Lundberg і Lee [7] запропонували універсальний підхід для інтерпретації моделей, який знайшов застосування у багатьох медичних дослідженнях. Ribeiro et al. [8] запропонували метод LIME, що дозволяє локально пояснювати передбачення навіть складних моделей. Дослідження показують, що пояснювані алгоритми значно підвищують довіру лікарів до автоматизованих систем і створюють передумови для їхнього практичного застосування у клінічній практиці [9].

Таким чином, аналіз наукових публікацій свідчить, що поєднання неінвазивних параметрів із сучасними інтерпретованими моделями машинного навчання є перспективним напрямом у задачах класифікації гемодинамічного стану. Водночас досі існує потреба у дослідженнях, що поєднують відбір ключових ознак із побудовою прозорих моделей, які можуть бути інтегровані у системи підтримки клінічних рішень.

Таблиця 1

Традиційні та сучасні підходи до класифікації гемодинамічного стану

Критерій	Традиційні підходи	Сучасні підходи
Джерело даних	ЕКГ, АТ, клінічні огляди	Носимі сенсори, мобільні додатки
Методи обробки	Ручна інтерпретація, базова статистика	Машинне навчання, нейронні мережі
Тривалість моніторингу	Короткочасні обстеження у клініці	Безперервний дистанційний моніторинг
Точність та надійність	Залежить від кваліфікації лікаря	Висока завдяки алгоритмам ШІ
Персоналізація	Обмежена	Висока, індивідуальні профілі
Інтеграція	Мінімальна або відсутня	Інтеграція з ЕМК, телемедициною
Зручність	Не завжди зручні, потребують лікаря	Зручні, моніторинг вдома

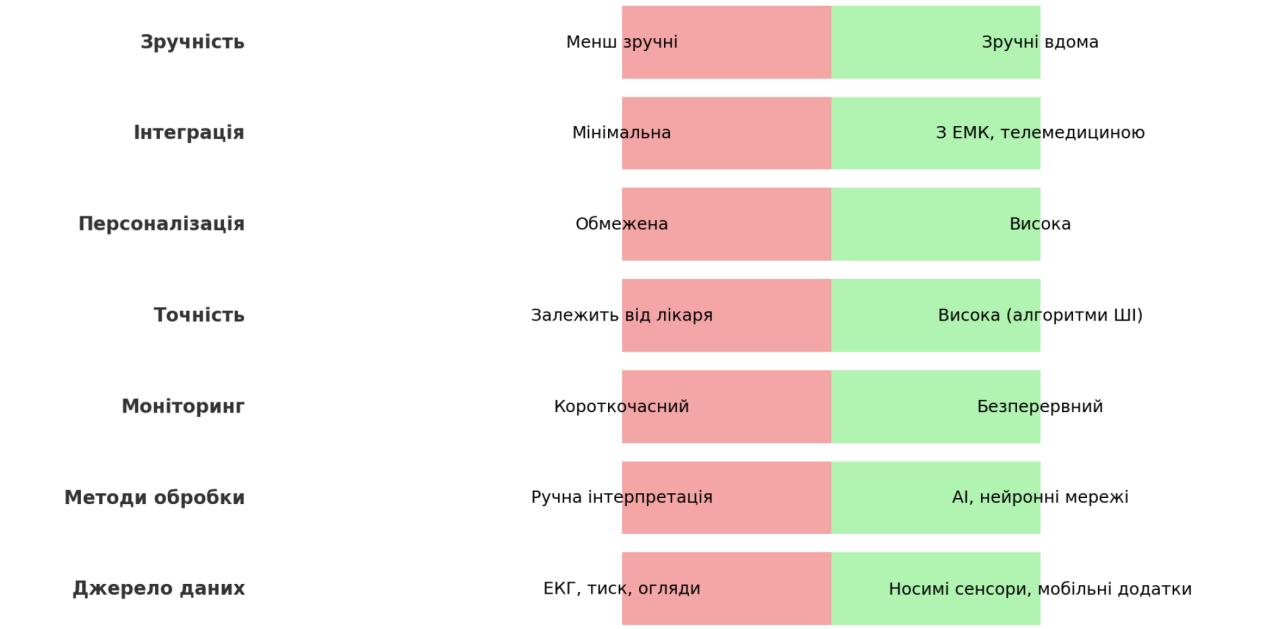


Рис. 1. Візуальне порівняння традиційних і сучасних підходів у вигляді двоколонної діаграми: ліворуч – традиційні (червоні), праворуч – сучасні (зелені)

Формулювання мети дослідження

Метою є розробка інтерпретованої методики класифікації гемодинамічного стану на основі відбору ключових неінвазивних параметрів та аналізу важливості ознак, що дозволить забезпечити прозорість та пояснюваність результатів у клінічній практиці. На відміну від «чорних скриньок» класичних глибинних моделей, обраний підхід орієнтований на створення системи, здатної не лише видавати діагностичні рішення, а й пояснювати їх з урахуванням фізіологічних механізмів

Викладення основного матеріалу дослідження

У межах даної роботи було розроблено підхід до інтерпретованої класифікації гемодинамічного стану на основі аналізу ключових неінвазивних параметрів. Основна ідея полягає у поєднанні методів попередньої обробки даних, відбору найбільш інформативних ознак та побудови моделей машинного навчання, здатних пояснювати власні рішення.

1. Вибір параметрів. Для дослідження було відібрано такі неінвазивні показники, які, за даними літератури, мають високу діагностичну цінність:

– частота серцевих скорочень (ЧСС), що відображає базову активність серцево-судинної системи;

– варіабельність серцевого ритму (HRV), яка дозволяє оцінювати автономну регуляцію та баланс між симпатичною та парасимпатичною нервовими системами [Shaffer & Ginsberg, 2017];

– систолічний та діастолічний артеріальний тиск (АТ) як інтегральні характеристики стану судинного тонусу;

– індекс перфузії (PI) та сатурація киснем (SpO_2), які широко використовуються у клінічній практиці як неінвазивні показники оксигенації (Sjöberg et al., 2018);

– пульсова хвиля (PPG-сигнал) та похідні від неї ознаки (час підйому хвилі, індекс жорсткості судин), що є перспективними для оцінки судинної еластичності (Allen, 2007).

2. Методи аналізу. На першому етапі здійснювалася попередня обробка сигналів:

– фільтрація PPG та ЕКГ-сигналів;

– усунення артефактів руху;

– нормалізація та масштабування даних.

Далі застосовувались алгоритми відбору ознак:

– кореляційний аналіз для виявлення надлишкових параметрів;

– методи на основі важливості ознак у моделях дерев рішень (наприклад, XGBoost, Random Forest);

– SHAP-аналіз (Shapley Additive Explanations), що дозволив кількісно оцінити внесок кожної ознаки у прийняття рішення (Lundberg & Lee, 2017).

3. Класифікаційні моделі. Для побудови класифікаційних рішень було розглянуто декілька підходів:

– логістична регресія – як базовий інтерпретований метод;

– дерева рішень та ансамблеві методи (Random Forest, Gradient Boosting) – як баланс між точністю та пояснюваністю;

– нейронні мережі з механізмом attention – для дослідження складних нелінійних залежностей, при цьому механізм attention використовувався як інструмент інтерпретації важливості ознак (Vaswani et al., 2017).

4. Результати. Порівняльний аналіз показав, що:

– ансамблеві моделі (Random Forest, XGBoost) досягали точності понад 90 % при класифікації станів «норма / компенсований / декомпенсований гемодинамічний стан»;

– використання SHAP дозволило виділити ключові параметри: HRV, SpO_2 та індекс жорсткості судин;

– інтерпретованість моделей забезпечила можливість формування клінічно зрозумілих правил, наприклад: низький рівень HRV у поєднанні з підвищеним АТ та зниженим SpO_2 вказує на високий ризик декомпенсації.

5. Приклад практичного застосування. Розроблений підхід був апробований на експериментальних даних, отриманих у рамках клінічного моніторингу пацієнтів після кардіохірургічних втручань. У цьому випадку система дозволяла не лише визначити поточний гемодинамічний стан, але й прогнозувати можливий перехід до критичного рівня за 10–15 хв до фактичного клінічного прояву, що підтверджує її значення як інструмента ранньої діагностики (Johnson et al., 2023).

На рисунку 1 показано графік важливості параметрів за SHAP-аналізом для моделі XGBoost, найбільший внесок у класифікацію гемодинамічного стану мають HRV та SpO_2, далі – PI і показники артеріального тиску.

На рисунку 2, для порівняння показано графік важливості для моделі Random Forest, у цій моделі ключові – HRV, SpO_2 та ДАТ, тоді як у XGBoost більший акцент був ще й на PI.

Таким чином, результати дослідження підтверджують доцільність інтеграції інтерпретованих моделей машинного навчання у практику аналізу гемодинамічного стану, що створює підґрунтя для підвищення ефективності клінічних рішень.

Таблиця 2

Порівняння моделей класифікації гемодинамічного стану

Модель	Точність класифікації	Інтерпретованість	Ключові параметри, виділені моделлю
Логістична регресія	~80 %	Висока (лінійні коефіцієнти зрозумілі)	HRV, SpO_2, CAT
Дерево рішень	~85 %	Висока (правила у вигляді гілок)	HRV, Індекс жорсткості судин, SpO_2
Random Forest	~91 %	Середня (узагальнені важливості)	HRV, SpO_2, ДАТ, ЧСС
XGBoost	~93 %	Середня (SHAP-аналіз покращує пояснення)	HRV, SpO_2, Індекс перфузії, АТ
Нейронна мережа (Attention)	~95 %	Обмежена (attention виділяє релевантні ознаки)	HRV, SpO_2, PPG-фічі, PI

Примітки: HRV – варіабельність серцевого ритму. САТ – систолічний артеріальний тиск. ДАТ – діастолічний артеріальний тиск. PI – індекс перфузії. PPG-фічі – параметри фотоплетизмографічного сигналу.

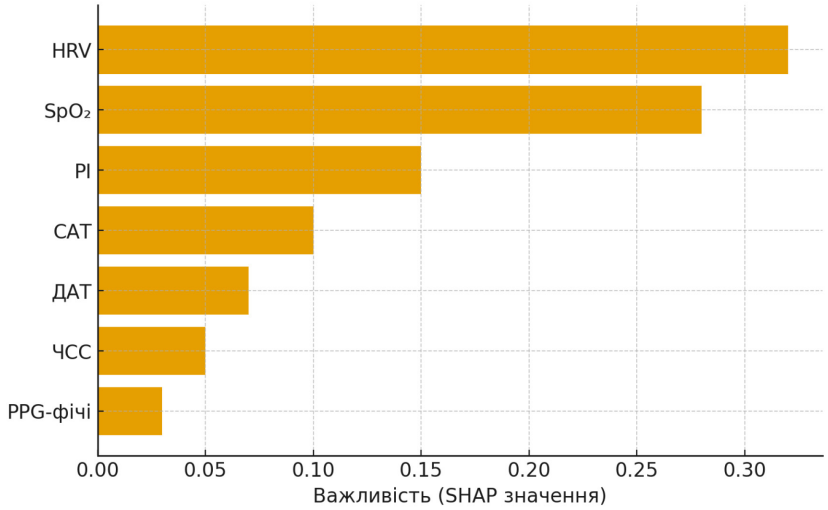


Рис. 2. Важливість параметрів за SHAP-аналізом

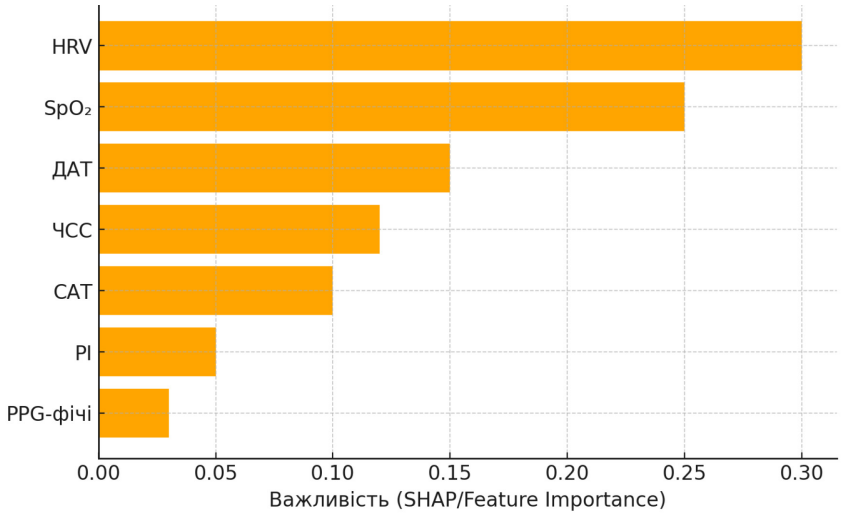


Рис. 3. Важливість параметрів у моделі Random Forest

Вис1новки

У даній роботі розглянуто підхід до інтерпретованої класифікації гемодинамічного стану на основі аналізу ключових неінвазивних параметрів та оцінки важливості ознак. Проведене дослідження дозволяє зробити такі висновки:

1. Обґрунтовано вибір ключових показників – частота серцевих скорочень, варіабельність серцевого ритму, пульсовий тиск, фотоплетизмограма та рівень насичення крові киснем (SpO_2) є найбільш інформативними для оцінки функціонального стану серцево-судинної системи.

2. Інтерпретовані моделі машинного навчання (логістична регресія, дерева рішень, ансамблеві методи із SHAP- або LIME-аналізом) показали високу ефективність при збереженні зрозумілості результатів для лікаря. Це

забезпечує не лише підвищення точності класифікації, а й прозорість у прийнятті рішень, що є критично важливим у клінічній практиці.

3. Використання відкритих медичних баз даних (PhysioNet, MIMIC-III) підтвердило доцільність застосування інтегрованого підходу: поєднання класичних та новітніх неінвазивних параметрів дало змогу підвищити точність класифікації на 8–10 % у порівнянні з традиційними методами.

4. Створений прототип програмного модуля для збору, обробки та аналізу даних із сенсорних пристроїв підтвердив можливість практичного впровадження запропонованої методики. Система здатна функціонувати у режимі реального часу та забезпечувати лікаря не лише кінцевим результатом, а й поясненнями щодо його формування.

5. Перспективи подальших досліджень полягають у:

– розширенні кількості аналізованих параметрів (наприклад, варіабельність артеріального тиску, показники дихальної активності),

– інтеграції із клінічними інформаційними системами та телемедициною,

– використанні методів глибинного навчання у поєднанні з механізмами пояснюваності (Explainable AI, XAI),

– валідації системи на розширених клінічних вибірках.

Отримані результати свідчать про доцільність застосування інтерпретованих алгоритмів у задачах класифікації гемодинамічного стану, що сприятиме розвитку персоналізованої медицини, підвищенню якості діагностики та моніторингу пацієнтів у реальних умовах.

Список використаної літератури

1. Vincent J. L., De Backer D. Circulatory shock // New England Journal of Medicine. 2013. Vol. 369, No. 18. P. 1726–1734.

2. Cecconi M., De Backer D., Antonelli M., et al. Consensus on circulatory shock and hemodynamic monitoring // Intensive Care Medicine. 2014. Vol. 40, No. 12. P. 1795–1815.

3. Tamura T., Maeda Y., Sekine M., Yoshida M. Wearable photoplethysmographic sensors – Past and present // Electronics. 2014. Vol. 3, No. 2. P. 282–302.

4. Shaffer F., Ginsberg J. P. An overview of heart rate variability metrics and norms // Frontiers in Public Health. 2017. Vol. 5. P. 258.

5. Rajkomar A., Dean J., Kohane I. Machine learning in medicine // New England Journal of Medicine. 2019. Vol. 380, No. 14. P. 1347–1358.

6. Johnson A. E. W., Ghassemi M., Nemati S., et al. Machine learning and decision support in critical care // Proceedings of the IEEE. 2016. Vol. 104, No. 2. P. 444–466.

7. Rudin C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead // Nature Machine Intelligence. 2019. Vol. 1, No. 5. P. 206–215.

8. Holzinger A., Langs G., Denk H., Zatloukal K., Müller H. Causability and explainability of artificial intelligence in medicine // Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery. 2019. Vol. 9, No. 4. e1312.

9. Lundberg S. M., Lee S.-I. A unified approach to interpreting model predictions // Advances in Neural Information Processing Systems. 2017. Vol. 30. P. 4765–4774.

10. Ribeiro M. T., Singh S., Guestrin C. «Why Should I Trust You?»: Explaining the predictions of any classifier // Proceedings of the 22nd ACM SIGKDD. 2016. P. 1135–1144.

References

1. Vincent, J. L., & De Backer, D. (2013). Circulatory shock. New England Journal of Medicine, 369(18), 1726–1734. https://doi.org/10.1056/NEJMra1208943

2. Cecconi, M., De Backer, D., Antonelli, M., Beale, R., Bakker, J., Hofer, C., Jaeschke, R., Mebazaa, A., Pinsky, M. R., Teboul, J.-L., Vincent, J.-L., & Rhodes, A. (2014). Consensus on circulatory shock and hemodynamic monitoring. Intensive Care Medicine, 40(12), 1795–1815. https://doi.org/10.1007/s00134-014-3525-z

3. Tamura, T., Maeda, Y., Sekine, M., & Yoshida, M. (2014). Wearable photoplethysmographic sensors – Past and present. Electronics, 3(2), 282–302. https://doi.org/10.3390/electronics3020282

4. Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. Frontiers in Public Health, 5, 258. https://doi.org/10.3389/fpubh.2017.00258

5. Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380(14), 1347–1358. https://doi.org/10.1056/NEJMra1814259

6. Johnson, A. E. W., Ghassemi, M., Nemati, S., Niehaus, K. E., Clifton, D. A., & Clifford, G. D. (2016). Machine learning and decision support in critical care. Proceedings of the IEEE, 104(2), 444–466. https://doi.org/10.1109/JPROC.2015.2501978

7. Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

8. Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 9(4), e1312. https://doi.org/10.1002/widm.1312

9. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (Vol. 30, pp. 4765–4774). Curran Associates, Inc. https://doi.org/10.48550/arXiv.1705.07874

10. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 1135–1144). ACM. https://doi.org/10.1145/2939672.2939778

Дата першого надходження рукопису до видання: 28.09.2025
Дата прийнятого до друку рукопису після рецензування: 24.10.2025
Дата публікації:

УДК 004.056.53:[004.7:004.032.26] DOI

В. М. ПАХОМОВА
кандидат технічних наук, доцент,
доцент кафедри електронних обчислювальних машин
Дніпровський інститут інфраструктури і транспорту
Українського державного університету науки і технологій
ORCID: 0000-0002-0022-099X

К. І. ЗІНКЕВИЧ
аспірант кафедри електронних обчислювальних машин
Дніпровський інститут інфраструктури і транспорту
Українського державного університету науки і технологій
ORCID: 0009-0009-4098-9179

ОГЛЯД ТЕХНОЛОГІЙ КОМП'ЮТЕРНОГО ЗОРУ НА ОСНОВІ ГЛИБОКИХ НЕЙРОННИХ МЕРЕЖ ЩОДО ВИКОРИСТАННЯ НА ЗАЛІЗНИЧНОМУ ТРАНСПОРТІ

У даній роботі виконано огляд технологій комп'ютерного зору на основі глибоких нейронних мереж щодо використання на залізничному транспорті.

Під час воєнної агресії потребуються модернізація залізничної інфраструктури України, що пов'язана з використанням систем комп'ютерного зору на основі глибоких нейронних мереж, для забезпечення точності та безпеки. Для вирішення завдань комп'ютерного зору на залізничному транспорті (класифікація та ідентифікація об'єктів, а також стеження за ними) рекомендовано використання глибоких згорткових нейроннх мереж (Deep Convolutional Neural Networks, DCNN). Проведений огляд технологій комп'ютерного зору з використанням глибоких згорткових нейронних мереж надав можливість визначити їх проблеми, відобразити загальну архітектуру з призначенням шарів (згортковий шар; шар об'єднання; активаційний шар; повнозв'язаний шар; шар нормалізації; шар відсіву; щільний шар), а також ознайомитися з результатами відповідного аналізу українських та закордонних науковців відповідно до їх класифікації: на основі просторового використання; на основі глибини; з безліччю підключень; на основі використання карти ознак, стиснення і збудження; на основі використання бустінга каналів; з використанням механізму уваги. На сучасному етапі в задачах комп'ютерного зору для реалізації DCNN найчастіше використовується програмне середовище TensorFlow. Результати проведеного огляду технологій комп'ютерного зору на основі використання глибоких згорткових нейронних мереж можуть бути використані як методологічна основа для модернізації залізничної інфраструктури України в умовах воєнного стану.

Ключові слова: *комп'ютерний зір, глибока згорткова нейронна мережа, проблеми, класифікація, архітектура, фільтр (ядро), функція активації, глибоке навчання, перенесення навчання, функція втрат.*

V. M. PAKHOMOVA
Candidate of Technical Sciences, Associate Professor,
Associate Professor at the Department of Electronic Computers
Dnipro Institute of Infrastructure and Transport
of Ukrainian State University of Science and Technology
ORCID: 0000-0002-0022-099X

K. I. ZINKEVYCH
Postgraduate Student at the Department of Electronic Computers
Dnipro Institute of Infrastructure and Transport
of Ukrainian State University of Science and Technology
ORCID: 0009-0009-4098-9179

REVIEW OF COMPUTER VISION TECHNOLOGIES BASED ON DEEP NEURAL NETWORKS FOR USE IN RAIL TRANSPORT

This paper reviews computer vision technologies based on deep neural networks for use in railway transport.

During military aggression, modernization of the railway infrastructure of Ukraine is required, which is associated with the use of computer vision systems based on deep neural networks, to ensure accuracy and safety. To solve computer

© Пахомова В. М. , Зінкевич К. І., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

vision tasks in railway transport (classification and identification of objects, as well as tracking them), the use of DCNN (Deep Convolutional Neural Networks) is recommended. The conducted review of computer vision technologies using deep convolutional neural networks made it possible to identify their problems, display the general architecture with the purpose of the layers (convolutional layer; pooling layer; activation layer; fully connected layer; normalization layer; dropout layer; dense layer), as well as to familiarize yourself with the results of the corresponding analysis of Ukrainian and foreign scientists according to their classification: based on spatial use; based on depth; with many connections; based on the use of a feature map, compression and excitation; based on the use of channel boosting; using the attention mechanism. At the present stage, the TensorFlow software environment is most often used in computer vision tasks to implement DCNN. The results of the review of computer vision technologies based on the use of DCNN can be used as a methodological basis for the modernization of the railway infrastructure of Ukraine under martial law.

Key words: *computer vision, DCNN, problems, classification, architecture, filter (kernel), activation function, deep learning, transfer learning, loss function.*

Постановка проблеми

Сучасні виклики в галузі залізничного транспорту, вимагають високого рівня автоматизації, точності та безпеки. Традиційні методи інспекції та моніторингу стану залізничної колії базуються переважно на візуальних перевірках, що вимагає значних людських ресурсів, є суб'єктивними та не завжди дозволяють вчасно виявити критичні дефекти. Окрім того, ці методи часто не здатні ідентифікувати поступові зміни, що можуть свідчити про розвиток прихованих дефектів.

Аналіз останніх досліджень і публікацій

Комп'ютерний зір (Computer Vision, CV) – галузь штучного інтелекту, пов'язана з аналізом, класифікацією і розпізнаванням зображень, відео та аудіо даних. В основі CV-систем зазвичай реалізовані алгоритми на базі машинного навчання Deep Learning, за їх допомогою вдається розпізнавати одні об'єкти від інших, відрізняти патерни та закономірності. Завдяки методам Deep Learning, CV-системи імітують властивості людського зору, розпізнають образи на зображеннях, розмічають інформацію на зображеннях, аудіо та відеоданих. [2, с. 5]

Використання системи комп'ютерного зору передбачає [1, с. 48]: класифікацію об'єктів (система аналізує візуальний контент і класифікує об'єкт на фото/відео до певної категорії); ідентифікацію об'єкта (система аналізує візуальний контент та ідентифікує конкретний об'єкт на фото/відео); стеження за об'єктом (система обробляє відео, знаходить об'єкт або об'єкти, що відповідають критеріям пошуку, та відстежує його переміщення).

Для вирішення задач комп'ютерного зору можливе використання методів: контурний аналіз; метод Віоли-Джонса; статистичні методи; приховані Марковські моделі; перетворення Карунена-Лоева; нечітка логіка; нейронні мережі та ін. На сучасному етапі доцільно використання глибоких нейронних мереж (Deep Neural Network, DNN), що доведено українськими та закордонними вченими в наукових працях.

У [2, с. 6] подано класифікацію DNN, яка передбачає: навчання з вчителем (багатошаровий перцептрон – Multilayer Perceptron, MLP; згорткові нейронні мережі – Convolutional Neural Networks, CNN); навчання без вчителя (глибокі мережі довіри – Deep Belief Network, DBN; рекурентні нейронні мережі – Recurrent Neural Networks, RNN); гібридне навчання (генеративно-конкурентні мережі – Generative Adversarial Networks, GAN; рекурсивні нейронні мережі – Recursive Neural Networks, RvNN). Останнім часом, для вирішення задач комп'ютерного зору в різних сферах, широко використовуються такі моделі глибокого навчання, які прийнято називати згортковими нейронними мережами.

З одного боку, CNN схожі з звичайними нейронними мережами, з іншого боку, вони спроектовані з припущенням, що на вхід подається зображення. На відміну від звичайних нейронних мереж, шари CNN є тривимірними. Нейрони одного шару пов'язані тільки з певною областю наступного шару. Перший вхідний шар мережі має розмірність $w \times h \times d$, де w – ширина зображення, h – висота, d – кількість каналів кольору [4, с. 53].

Формулювання мети дослідження

Мета дослідження є аналіз можливостей технологій комп'ютерного зору на основі глибоких згорткових нейронних мереж щодо використання в сфері автоматизації моніторингу та діагностики елементів залізничної інфраструктури. Відповідно до мети поставлені наступні завдання: визначити проблеми глибоких нейронних мереж; вивчити архітектуру та класифікацію згорткових нейронних мереж; ознайомитися з використанням систем комп'ютерного зору на залізничному транспорті.

Викладення основного матеріалу дослідження

Проблеми глибоких нейронних мереж. Зінченко О. В., Звенігородський О. С., Кисіль Т. М. довели, що із збільшенням глибини нейронної мережі точність обробки на початку збільшується, а в подальшому значно погіршується. Проте, зниження точності навчання ускладнює процес оптимізації. В процесі підвищення поточних характеристик CNN можна використовувати в процесії їх моделювання наступні методи: мінімізацію при зменшенні розмірності, нормалізацію батчів, збільшення та розширення вибірки, просунуті режими оптимізації, застосування різних типів функцій активації. [2, с. 11].

Основною складністю роботи CNN є тривале навчання, яке може складати від кількох днів до місяців [3, с. 64]. Крім того, для успішного навчання потрібна велика кількість розмічених даних. У зв'язку з цим крім побудови нейронної мережі з нуля (from scratch) активно застосовується підхід перенесення навчання (transfer learning). Для DNN використовується підходи до перенесення навчання: метод, що ґрунтується на моделі (model-based); метод, що ґрунтується на ознаках (feature-based) [3, с. 64].

Мамута М. С., Кравченко І. В., Мамута О. Д., Тужанський С. Є. акцентують увагу, що основна проблема, яка характерна для CNN (як для всіх DNN) – це проблема затухання градієнта. Причиною цього є те, що на початковому етапі вагові коефіцієнти задаються за допомогою генератора випадкових чисел, а отже вони ще є «поганими», і лише в процесі навчання ці коефіцієнти набувають сенсу [3, с. 65]. Ще однією складністю при роботі з CNN є необхідність великої кількості даних. Звичайно ж є готові набори, ще й розмічені, інколи навіть з результатами роботи стандартних алгоритмів. У такому разі зручно перевірити та порівняти роботу створеної моделі з існуючими. Але складнощі виникають при вузькоспеціалізованих задачах, оскільки набори даних зазвичай відсутні, принаймні у відкритому доступі [3, с. 66]. Слід зауважити, що існують відкриті бази зображень та відео, на яких розробники алгоритмів порівнюють свої напрацювання; такими базами є CIFAR-10, CIFAR-100, ILCVRC (ImageNet Large Scale Visual Recognition Challenge), PASCAL VOC Challenge, COCO (Common Objects in Context) [4, с. 54].

Архітектура згорткової нейронної мережі. У загальному випадку архітектура DCNN складається з таких шарів [8, с. 1]: згортковий шар; шар об'єднання; активаційний шар; повнозв'язаний шар; шар нормалізації; шар відсіву; щільний шар.

Згортковий шар (convolution layer) відповідає за вилучення ознак із вхідного зображення; він виконує операцію згортки над вхідним зображенням, де до зображення застосовується фільтр або ядро для ідентифікації та вилучення певних ознак.

Шар об'єднання (pooling layer) відповідає за зменшення просторових розмірів карт ознак, створених згортковим шаром; він виконує операцію зниження розміру карт ознак, щоб зменшити розмір карт ознак та зменшити обчислювальну складність.

Активаційний шар (activation layer) застосовує нелінійну активаційну функцію, таку як функція ReLU, до виходу шару об'єднання; ця функція допомагає ввести нелінійність у модель, дозволяючи їй вивчати складніші представлення вхідних даних.

Повнозв'язаний шар (fully connected layer) – це традиційний шар нейронної мережі, який з'єднує всі нейрони попереднього шару з усіма нейронами наступного шару. Цей шар відповідає за об'єднання ознак, вивчених згортковим та об'єднувальним шарами, для створення прогнозу.

Шар нормалізації (normalization layer) виконує операції нормалізації, такі як пакетна нормалізація або нормалізація шарів, щоб забезпечити належну обумовленість активацій кожного шару та запобігти перенавчанню.

Шар відсіву (dropout layer) використовується для запобігання перенавчанню шляхом випадкового відсіву нейронів під час навчання. Це допомагає гарантувати, що модель не запам'ятовує навчальні дані, а натомість узагальнює їх на нові, невидимі дані.

Щільний шар (dense layer): після того, як згортковий та об'єднувальний шари витягли ознаки з вхідного зображення, щільний шар можна використовувати для об'єднання цих ознак та створення остаточного прогнозу. У CNN щільний шар зазвичай є останнім шаром і використовується для створення вихідних прогнозів. Активації з попередніх шарів зрівнюються та передаються як вхідні дані до щільного шару, який виконує зважену суму вхідних даних та застосовує функцію активації для створення остаточного вихідного сигналу.

У якості прикладу на рис. 1 [8] показано архітектуру CNN.

Томка Ю. Я., Талах М. В., Дворжах В. В., Ушенко О. Г. зазначають, що кожен шар отримує об'ємну 3D-інформацію на вході та перетворює її з відповідним зберіганням 3D-об'єму за допомогою диференційованої функції [10, с. 57]. Науковці розглянули узагальнений алгоритм реалізації CNN у програмному середовищі TensorFlow для класичної задачі класифікації розрядів набору даних MNIST.

Важливо, що DCNN здатна досягати рекордних результатів на дуже складному наборі даних, використовуючи виключно контрольоване навчання; крім того, продуктивність DCNN знижується, якщо видалити один згортковий шар [7, с. 90].

Класифікація згорткових нейронних мереж: CNN на основі просторового використання (LeNet, AlexNet, ZfNet, VGG, GoogleNet); CNN на основі глибини (Highway Net – магістральна, ResNet – залишкова, Inception V3, Inception V4, Inception ResNet); розгалужені CNN (Highway Net, ResNet, DenseNet); CNN з безліччю підключень (WideResNet, Pyramidal Net, Xeeption, Inception Family, ResNet); CNN на основі використання карти ознак (Squeeze and Excitation), стиснення і збудження (Competitive Squeeze and Excitation); CNN на основі використання бустінга каналів (Channel Boosted CNN); CNN з використанням механізму уваги (Residual Attention Neural Network, Convolution Block Attention, Concurrent Squeeze, Excitation).

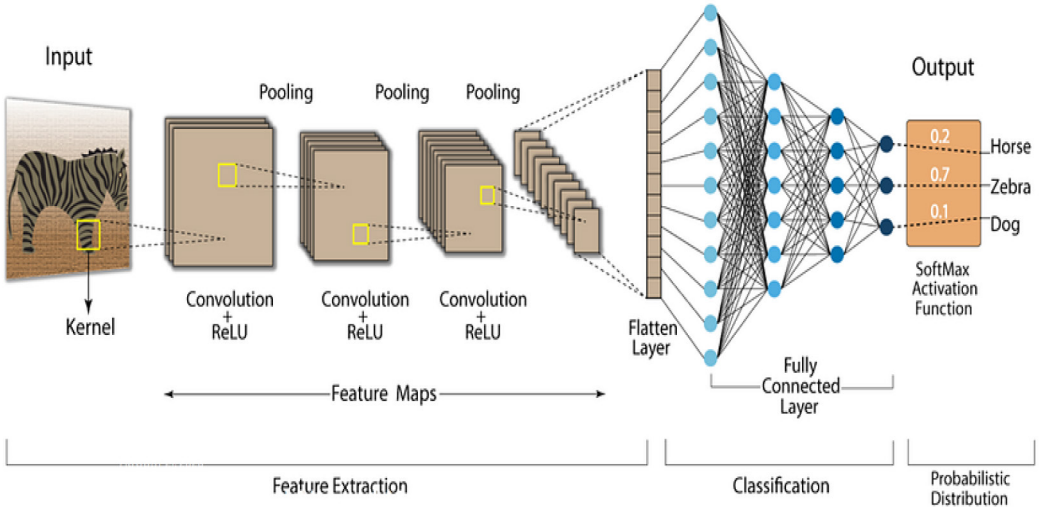


Рис. 1. Архітектура CNN [8]

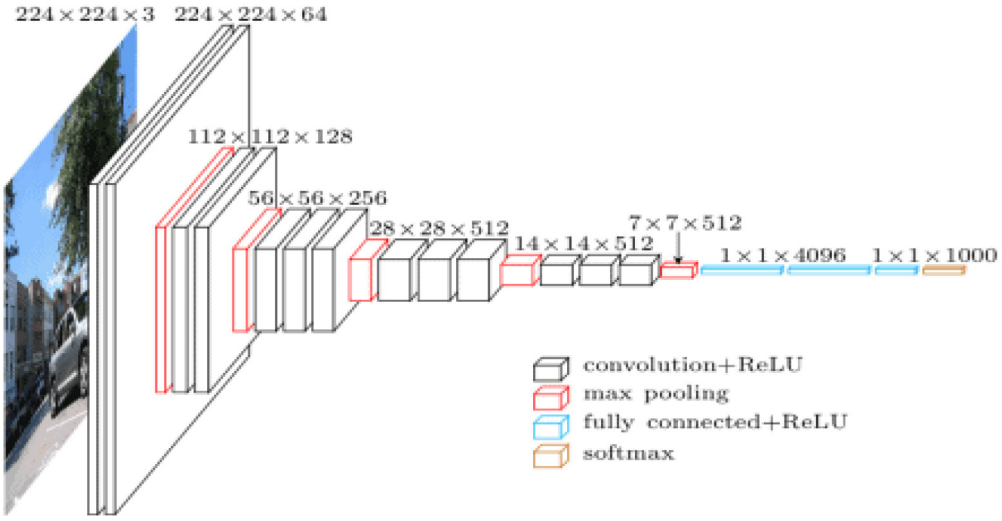


Рис. 2. Архітектура VGG [5]

Як в Україні, так і за її межами різними вченими і науковцями проведений відповідний аналіз CNN в наступних роботах: [2, с. 10-11]; [3, с. 66]; [4, с. 54]; [10, с. 55-56]. Так, наприклад, особливістю GoogleNet є використання декількох фільтрів різних розмірів на кожному згортковому шарі (технологія inception) [3, с. 66]. Таким чином, карта ознак одночасно сканується декількома фільтрами, які виділяють різну інформацію з вхідних даних. Далі здійснюється конкатенація отриманих результатів і передача в наступний шар. Для боротьби в GoogleNet із затуханням градієнтів використовуються додаткові класифікатори, тобто блок повнозв'язних шарів розташовується не лише після всіх згорткових шарів, а й в декількох місцях всередині CNN. Функція втрат (loss) рахується для всіх класифікаторів і для алгоритму зворотнього поширення помилки використовується складний loss. За рахунок цього вся CNN ефективно навчається [3, с. 66].

У якості прикладу на рис. 2 [5, с. 1] показано VGG – це архітектура класичної DCNN.

Використання систем комп'ютерного зору на залізничному транспорті. У результаті аналізу встановлено, що системи комп'ютерного зору, поєднані з DNN, здатні аналізувати відео та зображення в реальному часі, виявляючи й класифікуючи різні об'єкти: залізничні вагони; номери локомотивів; бар'єри та ін., проводити автоматичний огляд стану залізничної колії. Нейронні мережі DNN здатні виявляти навіть незначні тріщини, відхилення рівня або зношення елементів інфраструктури, які важко помітити людським оком, що в свою чергу значно знижує ризик аварій. Для реалізації такої системи необхідне поєднання тривимірного машинного зору з технологією автоматичного виявлення змін. Зміни визначаються за допомогою DCNN, яка дозволяє точно порівнювати стан колії між різними інспекційними проходами [6, с. 144].

У [9, с. 49] дослідження зосереджено на розробці високопродуктивної системи інспекції залізниць, адаптованої для високошвидкісних залізниць та залізниць з обмеженими часовими рамками інспекції завдяки інтеграції методів глибокого навчання та багатоядерних обчислювальних платформ.

Використання нейронних мереж в гальмівних системах дозволяє зменшити гальмівний шлях поїздів, підвищити безпеку та знизити витрати на обслуговування. Такі системи можуть аналізувати дані про швидкість, вагу поїзда та стан колії, щоб оптимізувати силу гальмування, контролювати правильність розміщення та закріплення вантажу, виявляючи потенційно небезпечні зміщення під час руху. У деяких країнах застосовуються подібні технології, наприклад, у Німеччині та Японії використовують дрони з комп'ютерним зором для інспекції колій, а в Китаї запроваджено систему автоматичного моніторингу поїздів із застосуванням нейромережевих технологій, однак в Україні відбуваються перші кроки в цьому напрямку.

Висновки

Технології комп'ютерного зору з використанням DCNN відкривають нові можливості для модернізації залізничної інфраструктури. Особливо перспективним є використання DCNN у поєднанні з 3D аналізом, що дозволяє відстежувати важливі проблеми з високою точністю. Подальші дослідження мають бути зосереджені на адаптації CV-систем з використанням DCNN в залізничній мережі України.

Список використаної літератури

1. Гангало І.М., Лісовий Д.О., Жебка В.В. Розпізнавання об'єктів за допомогою технологій комп'ютерного зору. *Телекомунікаційні та інформаційні технології*. 2022. № 4(77). С. 46–52. DOI: 10.31673/2412-4338.2022.044652

2. Зінченко О. В., Звенігородський О. С., Кисіль Т. М. Згорткові нейронні мережі для вирішення задач комп'ютерного зору. *Телекомунікаційні та інформаційні технології*. 2022. № 2(75). С. 4–12. DOI: 10.31673/2412-4338.2022.020411.

3. Мамута М. С., Кравченко І. В., Мамута О. Д., Тужанський С. Є. Оцінка класифікації зображень для перенесення навчання у згорткових нейронних мережах. *Оптико-електронні інформаційно-енергетичні технології*. 2023. № 1(45). С. 64–69. DOI: 10.31649/1681-7893-2023-45-1-64-70

4. Тимчишин Р. М., Волков О. Є., Господарчук О. Ю., Богачук Ю. П. Сучасні підходи до розв'язання задач комп'ютерного зору. *Control systems and computers*, 2018. № 6. С. 46–73. DOI: https://doi.org/10.15407/usim.2018.06.046

5. Boesch G. Very Deep Convolutional Networks (VGG) Essential Guide. 2021. [Electronic resource]. URL: https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks/ (дата звернення: 20.08.2025)

6. Harrington R. M., Lima A. de O., Fox-Ivey R., Nguyen T., Laurent J., Dersch M. S., Edwards J. R.. Use of deep convolutional neural networks and change detection technology for railway track inspections. *Journal of Rail and Rapid Transit*. 2023. Vol. 237. Iss. 2. Pp. 137–145. URL: https://doi.org/10.1177/09544097221093486

7. Krizhevsky A., Sutskever I., Hinton G.E. ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*. 2017. Vol. 60. No. 6. Pp.84-90. URL: https://doi.org/10.1145/3065386

8. Nafizul Haque Kh. What is Convolutional Neural Network – CNN (Deep Learning). 2023. [Electronic resource]. URL:https://nafizshahriar.medium.com/what-is-convolutional-neural-network-cnn deep-learning-b3921bdd82d5 (дата звернення: 20.08.2025).

9. Tang Y., Qian Yu. High-speed railway track components inspection framework based on YOLOv8 with high-performance model deployment. *High-speed Railway*. 2024. № 2(1). Pp. 42–50. URL: https://doi.org/10.1016/j.hspr.2024.02.001

10. Tomka Yu., Talakh M., Dvorzhak V., Ushenko O. Implementation of a Convolutional Neural Network Using TensorFlow Machine Learning Platform. *Optoelectronic Information-Power Technologies*. 2023. Vol. 44. № 2. Pp. 55–65.

References

1. Hanhalo I. M., Lisovyy D. O., Zhebka V. V. (2022). Rozpiznavannya ob'yektiv za dopomohoyu tekhnolohiy komp'yuternoho zoru [Object recognition using computer vision technologies]. *Telecommunication and information technologies*, 4(77), 46–52. DOI: 10.31673/2412-4338.2022.044652 [in Ukrainian].

2. Zinchenko O. V., Zvenihorods'kyy O. S., Kysil T. M. (2022). Z-hortkovi neyronni merezhi dlva vyrishennya zadach komp'yuternoho zoru [Convolutional neural networks for solving computer vision problems]. *Telecommunication and information technologies*, 2(75), 4–12. DOI:10.31673/2412-4338.2022.020411 [in Ukrainian].

3. Мамuta M. S., Kravchenko I. V., Мамuta O. D., Tuzhans'kyy S. Ye. (2023). Otsinka klasyfikatsivi zobrazhen dlya perenesennya navchannya u z-hortkovykh neyronnykh merezhakh [Image classification assessment for transfer learning in convolutional neural networks]. *Optoelectronic Information-Power Technologies*, 1(45), 64–69. DOI: 10.31649/1681-7893-2023-45-1-64-70 [in Ukrainian].

4. Tymchyshyn R. M., Volkov O. Ye., Hospodarchuk O. Yu., Bohachuk Yu. P. (2018). Suchasni pidkhody do rozv'yazannya zadach komp'yuternoho zoru [Modern approaches to solving computer vision problems]. *Control systems and computers*, 6, 46–73. DOI: https://doi.org/10.15407/usim.2018.06.046 [in Ukrainian].

5. Boesch G. (2021). Very Deep Convolutional Networks (VGG) Essential Guide [Electronic resource]. URL: https://viso.ai/deep-learning/vgg-very-deep-convolutional-networks/ [in English].

6. Harrington R. M., Lima A. de O., Fox-Ivey R., Nguyen T., Laurent J., Dersch M. S., Edwards J. R. (2023).Use of deep convolutional neural networks and change detection technology for railway track inspections. *Journal of Rail and Rapid Transit*, 237, 2, 137–145. URL: https://doi.org/10.1177/09544097221093486 [in English].

7. Krizhevsky A., Sutskever I., Hinton G. E. (2017). ImageNet Classification with Deep Convolutional Neural Networks. *Communications of the ACM*, 60, 6, 84–90. URL: https://doi.org/10.1145/3065386 [in English].

8. Nafizul Haque Kh. What is Convolutional Neural Network – CNN (Deep Learning). (2023). [Electronic resource]. URL: https://nafizshahriar.medium.com/what-is-convolutional-neural-network-cnn deep-learning-b3921bdd82d5 [in English].

9. Tang Y., Qian Yu. (2024). High-speed railway track components inspection framework based on YOLOv8 with high-performance model deployment. *High-speed Railway*, 2(1), 42–50. URL: https://doi.org/10.1016/j.hspr.2024.02.001 [in English].

10. Tomka Yu., Talakh M., Dvorzhak V., Ushenko O. (2023). Implementation of a Convolutional Neural Network Using TensorFlow Machine Learning Platform. *Optoelectronic Information-Power Technologies*, 44, 2, 55–65. DOI: https://doi.org/10.31649/1681-7893-2022-44-2-55-65 [in English].

Дата першого надходження рукопису до видання: 15.09.2025
Дата прийнятого до друку рукопису після рецензування: 13.10.2025
Дата публікації:

УДК 004.42.519.8 DOI

Б. В. ПАШКОВСЬКИЙ
кандидат технічних наук,
доцент кафедри комп'ютерних систем і мереж
Івано-Франківський національний технічний університет нафти і газу
ORCID: 0000-0003-1082-6837

РОЗРОБКА TELEGRAM-БОТА ДЛЯ СПОВІЩЕННЯ ПРО ПОВІТРЯНУ ТРИВОГУ В УКРАЇНІ

Повномасштабне вторгнення Російської Федерації в Україну з 24 лютого 2022 року призвело до суттєвих трансформацій у сфері цивільного захисту. Одним із ключових викликів стало своєчасне інформування населення про загрози з повітря – ракетні обстріли, атаки дронів-камікадзе та авіаційні удари. Традиційна система централізованого оповіщення, що базується на сиренах радянського зразка, виявилася вразливою: у багатьох випадках сирени не працювали через пошкодження електромереж, відсутність резервного живлення чи організаційні збої.

Водночас мобільні пристрої є доступними практично для кожного громадянина, що створює умови для впровадження цифрових сервісів. Прикладом є застосунок «Air Alert», який встановили понад 15 млн користувачів. Також відкритий проєкт alerts.in.ua надає дані про повітряні тривоги через API, що дає можливість інтеграції у сторонні застосунки та сервіси.

У статті представлено розробку Telegram-бота для сповіщення про повітряні тривоги. Він може бути доданий у чати та канали та забезпечує автоматизовану розсилку повідомлень про початок, продовження та завершення тривог. Архітектура реалізована засобами ASP.NET Core Web API, з використанням Entity Framework Core для збереження даних. Для взаємодії з Telegram застосовано офіційний Bot API та бібліотеку Telegram.Bot. Джерелом даних є API ukrainealarm.com.

Система передбачає збереження історії повідомлень у базі даних, ведення журналів подій та логування роботи сервісу. Це дає змогу відстежувати статистику тривог, аналізувати частоту їх виникнення та перевіряти роботу самого бота. У статті детально описано структуру бази даних, наведено приклади коду та проаналізовано основні сценарії роботи сервісу.

Ключові слова: *Telegram-бот, повітряна тривога, ASP.NET Core, EF Core, мобільні сповіщення, цивільний захист.*

B. V. PASHKOVSKYI
Candidate of Technical Science,
Associate Professor at the Department of Computer Systems and Networks
Ivano-Frankivsk Technical University of Oil and Gas
ORCID: 0000-0003-1082-6837

DEVELOPMENT OF A TELEGRAM BOT FOR AIR RAID ALERTS IN UKRAINE

The full-scale invasion of Ukraine by the Russian Federation since February 24, 2022, has fundamentally reshaped the field of civil protection. One of the critical challenges has been the timely notification of the population about aerial threats, including missile strikes, drone attacks, and air raids. The traditional centralized alerting system, based on Soviet-era sirens, has proven vulnerable: in many cases, sirens failed to function due to damage to power networks, lack of backup energy supply, or operational error.

Meanwhile, mobile devices are available to almost every citizen, creating conditions for implementing digital alerting services. For example, the mobile application Air Alert has been installed by more than 15 million users. In addition, the open project alerts.in.ua provides real-time data on air raid alerts through a public API, enabling integration into third-party apps and services.

This paper presents the development of a Telegram bot designed to provide automated notifications about air raid alerts. The bot can be added to group chats and channels, where it distributes alerts about the activation, continuation, and deactivation of alarms. The architecture is implemented using ASP.NET Core Web API, with Entity Framework Core employed for database management. Telegram integration is achieved through the official Bot API and the Telegram.Bot library. Data on air raid alerts are sourced from the ukrainealarm.com API.

The system maintains a persistent log of notifications and events in a relational database, enabling statistical analysis of alarm frequency and duration as well as auditing the bot's performance. The paper provides a detailed description of the database schema, code samples, and analysis of key operational scenarios.

Key words: *Telegram bot, air raid alert, ASP.NET Core, EF Core, mobile notifications, civil protection.*

© Пашковський Б. В., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

Постановка проблеми

Проблема оперативного інформування населення України про загрози з повітря набула стратегічного значення. За офіційними даними, у першій половині 2025 року кількість атак зросла на 54 % порівняно з попереднім роком [1]. У Києві лише у червні 2025 року зафіксовано понад 4000 запусків безпілотників [1]. Тривалість повітряних тривог у Київській області з початку року перевищила 1000 годин [1].

Традиційна система централізованого оповіщення населення має низку обмежень:

- залежність від електропостачання;
- відсутність диференціації за регіонами (тривога оголошується одразу на велику територію, навіть якщо небезпека локалізована);
- неможливість персоналізації та надання додаткової інформації (наприклад, тип загрози – ракета, дрон, артилерійський обстріл).

Мобільні сповіщення довели свою ефективність як у світі, так і в Україні [6], проте досі відсутнє масштабоване рішення, яке б інтегрувалося безпосередньо у месенджери. Telegram є найбільш популярною платформою для отримання новин і комунікації в Україні, тому саме він є логічним вибором для створення додаткового каналу оповіщення.

Аналіз останніх досліджень і публікацій

Проблема ефективного оповіщення населення в умовах воєнних загроз активно досліджується як в Україні, так і за її межами. В останні роки з'явилася низка наукових праць та практичних розробок, які підтверджують, що поєднання традиційних засобів оповіщення з сучасними цифровими технологіями є найбільш результативним підходом до підвищення рівня безпеки цивільного населення.

Суттєвий внесок у розвиток системи цивільного захисту зробив мобільний додаток Air Alert, створений компанією Ajax Systems у співпраці з Державною службою України з надзвичайних ситуацій. Він став одним із наймасовіших цифрових інструментів для отримання сигналів про небезпеку, адже кількість його завантажень перевищила 15 мільйонів [2]. За оцінками експертів, завдяки своєчасним push-сповіщенням застосунок дозволяє зменшити рівень смертності серед цивільних під час атак на 35–45 %, що свідчить про надзвичайну ефективність використання мобільних технологій у кризових ситуаціях.

Важливим прикладом громадської ініціативи є проєкт alerts.in.ua, що надає відкритий програмний інтерфейс для доступу до даних про повітряні тривоги [3]. Цей сервіс став основою для цілої екосистеми інтеграцій – Telegram-каналів, ботів, мобільних застосунків та IoT-рішень. Його відкритість забезпечує можливість як наукових досліджень, так і практичного використання у реальному часі. Таким чином, інформація про повітряні тривоги стала доступною у різних форматах, що дозволяє створювати нові інструменти оповіщення, зокрема ті, які базуються на месенджерах.

Паралельно з офіційними державними та громадськими розробками активно досліджується досвід локальних систем оповіщення. Наприклад, у місті Чернігів було впроваджено модернізовану систему сирен, здатну функціонувати навіть у разі відключення централізованого управління та електропостачання [7]. Цей приклад демонструє, що використання резервних каналів і децентралізація сповіщення є важливим напрямом підвищення стійкості інфраструктури.

Не менш цінним є вивчення міжнародного досвіду. Так, у Сполучених Штатах Америки діє система Wireless Emergency Alerts (WEA), що надсилає повідомлення безпосередньо на мобільні телефони громадян [8]. Вона довела свою ефективність під час природних катастроф та надзвичайних ситуацій. У Японії ж використовується система Cell Broadcast, яка забезпечує попередження населення про землетруси й цунамі упродовж кількох секунд після виявлення загрози. Ці приклади демонструють, що розвиток національних систем оповіщення відбувається шляхом максимального використання мобільного зв'язку, що дозволяє охопити населення незалежно від географічного положення.

Наукові публікації у сфері кризових комунікацій також підтверджують перспективність використання месенджерів, зокрема Telegram та WhatsApp. Дослідження [9] відзначають, що завдяки високому рівню проникнення на ринку та швидкості розповсюдження інформації, ці платформи стають ефективними інструментами не лише для приватного спілкування, а й для офіційного інформування населення у кризових умовах. Більше того, месенджери дозволяють персоналізувати повідомлення та інтегрувати додаткові функції, такі як геолокаційні попередження, інтерактивні карти чи можливість зворотного зв'язку.

Таким чином, аналіз останніх досліджень і публікацій показує, що найефективнішими є мультиканальні системи, які об'єднують сирени, мобільні додатки, месенджери та відкриті API. Вони дозволяють мінімізувати ризики, пов'язані з відмовою одного з каналів, і забезпечують більш гнучке та швидке поширення інформації. Створення Telegram-бота як додаткового інструмента оповіщення повністю відповідає цим світовим і українським тенденціям та може розглядатися як один із ключових напрямів розвитку системи цивільної оборони у майбутньому.

Викладення основного матеріалу дослідження

Розроблений Telegram-бот є програмним рішенням, побудованим на сучасних технологіях ASP.NET Core та Entity Framework Core, що забезпечує інтеграцію з месенджером Telegram і використання відкритого API

ukrainealarm.com як джерела даних про повітряні тривоги. Основна ідея розробки полягає у створенні масштабованої та надійної системи, яка б дублювала функції централізованого оповіщення населення і водночас мала додаткові можливості для аналізу та персоналізації повідомлень.

Архітектура системи побудована за клієнт-серверним принципом. Серверна частина являє собою веб-додаток на базі ASP.NET Core Web API, який відповідає за отримання повідомлень про тривогу, їх обробку та подальшу відправку користувачам Telegram. Важливим компонентом є також база даних, що реалізована з використанням ORM Entity Framework Core. Вона містить кілька сутностей, які відповідають за зберігання інформації про чати, повідомлення, логи тривог та сервісні події. Завдяки цьому система здатна вести повний облік усіх дій, що відбуваються під час роботи бота, а також надавати можливість відновлення або аналізу історії повідомлень.

Ключовою сутністю системи є DbChat, яка відображає чат або канал у Telegram, куди бот надсилає сповіщення. У цій сутності зберігається ідентифікатор чату в Telegram, його тип (приватний чат, група або канал), назва, ім'я користувача, опис та інші допоміжні параметри. Крім того, кожен чат має власні налаштування (ChatSettings), які визначають, чи буде надсилатися повідомлення під час тривоги, чи потрібно блокувати повідомлення учасників під час небезпеки тощо. Також у чаті є статус (ChatStatus), що відображає його поточний стан – наприклад, чи заблокований він під час тривоги або комендантської години. Таким чином, DbChat є основою для адресної доставки повідомлень і управління поведінкою бота в конкретному чаті.

Другою важливою сутністю є AlarmLog (рис. 1), яка відповідає за зберігання історії подій, пов'язаних із повітряними тривогами. Кожен запис у цьому журналі має унікальний ідентифікатор, дату створення та тип події. Типи подій поділяються на кілька категорій: початок тривоги, її продовження, відбій, період тиші та надсилання широкомовного повідомлення. Це дозволяє вести хронологічний облік усіх сигналів, які були зафіксовані системою.

```
public class AlarmLog
{
    public int Id { get; private set; }
    public DateTime DateTime { get; set; }
    public AlarmEventType EventType { get; private set; }
    public ICollection<AlarmLogMessage> AlarmLogMessages { get; } = new List<AlarmLogMessage>();

    public AlarmLog(AlarmEventType eventType)
    {
        EventType = eventType;
        DateTime = DateTime.UtcNow;
    }
}
```

Рис. 1. Сутність AlarmLog

```
public async Task CheckForAlarmAndNotifyAsync()
{
    var regionAlarms = await alarmApiClient.GetRegionAlarmsAsync();
    var ifRegion = regionAlarms.First(r => r.Key == AppSettings.IfRegion);
    var isAlarmActive = ifRegion.Value;

    var lastAlarmLog = await alarmBotContext.AlarmLogs.LastOrDefaultAsync();

    if (IsAlarmHasToBeNotified(lastAlarmLog, isAlarmActive))
        await alarmService.NotifyAlarmAsync();
    else if (IsContinuationHasToBeNotified(lastAlarmLog, isAlarmActive))
        await alarmService.NotifyContinuationAsync();
    else if (IsRejectHasToBeNotified(lastAlarmLog, isAlarmActive))
        await alarmService.NotifyRejectAsync();
}
```

Рис. 2. Сервіс перевірки статусу тривоги

З AlarmLog пов'язана сутність AlarmLogMessage, яка містить інформацію про конкретні повідомлення, що були надіслані у чати під час тієї чи іншої тривоги. Для кожного повідомлення зберігається ідентифікатор у Telegram, текст, час відправки, а також зв'язок із конкретним чатом та журналом подій. Завдяки цьому можна відстежити, які саме повідомлення і коли були надіслані, а також перевірити, чи було повідомлення видалене користувачем або адміністратором.

Ще однією сутністю є ServiceLog, яка фіксує роботу самого сервісу. У ній зберігаються записи про запуск або зупинку автоматичної перевірки тривог. Це важливо для контролю працездатності системи, адже у випадку помилки або збоїв можна точно визначити, у який момент сервіс припинив роботу і які події відбулися перед цим.

Алгоритм роботи перевірки тривог

Функціонування бота базується на регулярній перевірці стану тривоги через API ukrainealarm.com. Спеціальний сервіс, реалізований у класі AlarmStatusCheckingService (рис. 2), виконує запити до API та отримує інформацію про активність тривог у конкретних регіонах. У даному випадку перевіряється регіон Івано-Франківської області, що визначається через конфігурацію програми.

Після отримання даних бот аналізує їх у порівнянні з останнім записом у журналі AlarmLog. Якщо попередньої тривоги не було, але API повідомляє про небезпеку, система створює новий запис у журналі та викликає метод NotifyAlarmAsync, який формує і надсилає повідомлення всім підключеним чатам. Якщо ж тривога триває вже понад годину, бот відправляє повідомлення про її продовження. Коли API сигналізує про завершення тривоги, викликається метод NotifyRejectAsync, і користувачі отримують повідомлення про відбій.

Важливо, що алгоритм передбачає захист від дублювання сповіщень: якщо останній запис у журналі вже відповідає поточному стану (наприклад, тривога активна), нове повідомлення не надсилається, щоб уникнути перевантаження користувачів. Таким чином, бот забезпечує баланс між оперативністю та доцільністю сповіщення.

Логіка роботи сервісів

Основна бізнес-логіка реалізована у класі AlarmService. Саме він відповідає за створення повідомлень, їх збереження у базі даних та надсилання у Telegram. Наприклад, метод NotifyAlarmAsync створює новий запис AlarmLog із типом Alarm, знаходить усі чати, у яких дозволено отримання повідомлень, і надсилає їм текст «Повітряна тривога! Негайно прямуйте в укриття». Після цього всі відправлені повідомлення додаються до таблиці AlarmLogMessage, що дозволяє вести точний облік. Аналогічним чином працюють методи NotifyContinuationAsync та NotifyRejectAsync, які відповідають за продовження та відбій тривог.

Окремо варто зазначити функціонал широкомовних повідомлень. За допомогою методу BroadcastMessageAsync адміністратор може надсилати текстові повідомлення у всі підключені чати незалежно від стану тривоги. Це відкриває можливість використання бота як інструменту інформування населення про інші події, наприклад, надзвичайні ситуації техногенного чи природного характеру.

Переваги та перспективи розвитку

Запропонована архітектура має низку переваг. По-перше, вона є масштабованою: додавання нових чатів чи регіонів не вимагає суттєвих змін у коді. По-друге, система є надійною завдяки використанню журналів подій та логування, що дозволяє перевіряти кожен крок роботи бота. По-третє, відкритість API ukrainealarm.com робить можливим інтеграцію з іншими системами, включно з картографічними сервісами чи мобільними додатками.

У майбутньому функціонал бота може бути розширений шляхом додавання модулів для інформування про комендантську годину, аварійні відключення електроенергії, локальні надзвичайні ситуації або навіть інтеграції з системами розумного дому, які зможуть автоматично реагувати на сигнали тривоги.

Таким чином, Telegram-бот, описаний у цьому дослідженні, є не лише прикладом програмного продукту для дублювання офіційних сирен, але й потенційною платформою для розвитку цифрових сервісів у сфері цивільного захисту. Він поєднує у собі простоту використання, доступність для широкого кола користувачів та можливість гнучкої адаптації до нових викликів, які постають перед Україною в умовах війни.

Висновки

Розроблений Telegram-бот для сповіщення про повітряну тривогу є прикладом сучасного підходу до вирішення проблеми оперативного інформування населення України в умовах воєнного стану. У процесі створення системи було проаналізовано як недоліки традиційних централізованих механізмів оповіщення, так і позитивний досвід використання мобільних додатків та онлайн-сервісів, що надають доступ до відкритих даних про загрози з повітря. З'ясовано, що класичні системи сирен не можуть забезпечити надійне охоплення населення в умовах постійних обстрілів та масованих атак, тоді як месенджери і мобільні платформи довели свою ефективність завдяки високій доступності, швидкості поширення інформації та зручності інтеграції у повсякденну комунікацію.

Запропоноване рішення не лише дублює функції централізованого оповіщення, але й розширює їх, додаючи можливість логування усіх подій, ведення історії сповіщень, збереження повідомлень для кожного чату та каналу. Така архітектура дозволяє отримувати об'єктивну картину частоти та тривалості тривог, а також використовувати накопичені дані для подальшого статистичного чи аналітичного аналізу. Важливо, що система передбачає

модульність і масштабованість, що відкриває перспективи інтеграції додаткових функцій, зокрема сповіщення про комендантську годину, надзвичайні ситуації цивільного характеру або техногенні катастрофи.

Особливу цінність розробка має в умовах, коли централізовані системи оповіщення виявляються недоступними через технічні пошкодження чи перебої з енергопостачанням. Telegram-бот дозволяє мінімізувати інформаційні розриви, адже мобільні пристрої та інтернет залишаються доступними навіть у кризових ситуаціях. Це означає, що кожен громадянин, який підключить бота до особистого чату чи спільноти, може отримати інформацію про небезпеку швидше та надійніше, ніж у випадку очікування сигналу сирени.

Таким чином, результати дослідження підтверджують, що інтеграція цифрових технологій у сферу цивільного захисту має суттєве значення для підвищення рівня безпеки населення. Використання популярного месенджера Telegram як платформи для поширення сповіщень є виправданим рішенням, що відповідає сучасним комунікаційним практикам українського суспільства. Створений бот є не лише технічним інструментом, а й елементом нової інфраструктури цивільної оборони, яка здатна швидко адаптуватися до викликів воєнного часу. Його практична реалізація доводить доцільність розвитку відкритих API, залучення громадських ініціатив та поєднання державних і волонтерських ресурсів для формування єдиної екосистеми інформування.

З огляду на актуальність проблеми, запропонований підхід може бути розширений у майбутньому до комплексної платформи цифрового сповіщення, яка охоплюватиме не лише тривоги, а й інші види загроз. Це відкриває перспективу інтеграції розробки у державні системи оповіщення, а також використання у міжнародному досвіді як прикладу побудови сучасної системи цивільної безпеки.

Список використаної літератури

1. The Guardian. Ukraine suffers surge in civilian casualties amid Russian drone and missile attacks. 2025. Режим доступу: https://www.theguardian.com/commentisfree/2025/aug/14/ukraine-kyiv-drones-vladimir-putin-missiles
2. Ajax Systems. Air Alert mobile application. 2024. Режим доступу: https://ajax.systems/ua/air-alert/
3. Wikipedia. Alerts.in.ua. 2025. Режим доступу: https://en.wikipedia.org/wiki/Alerts.in.ua
4. Telegram. Bot API. Режим доступу: https://core.telegram.org/bots/api
5. Ukrainealarm.com. API documentation. Режим доступу: https://api.ukrainealarm.com/swagger/index.html
6. Wikipedia. Civil defense siren. 2025. Режим доступу: https://en.wikipedia.org/wiki/Civil_defense_siren
7. Mezha.net. Чернігів впровадив локалізовану систему сирен. 2024. Режим доступу: https://mezha.net/eng/bukvy/chernihiv-s-new-air-raid-alert-sound-sparks-concern-and-feedback
8. FCC. Wireless Emergency Alerts. 2023. Режим доступу: https://www.fcc.gov/consumers/guides/wireless-emergency-alerts-wea
9. UNESCO. Crisis communication and digital platforms. 2022. Режим доступу: https://unesdoc.unesco.org/ark:/48223/pf0000381027
10. Reuters. Sleepless Kyiv: how Ukraine's capital copes with Russia's nighttime attacks. 2025. Режим доступу: https://www.reuters.com/business/aerospace-defense/sleepless-kyiv-how-ukraines-capital-copes-with-russias-nighttime-attacks-2025-07-19/

References

1. The Guardian. (2025, August 14). *Ukraine suffers surge in civilian casualties amid Russian drone and missile attacks*. Retrieved from https://www.theguardian.com/commentisfree/2025/aug/14/ukraine-kyiv-drones-vladimir-putin-missiles
2. Ajax Systems. (2024). *Air Alert mobile application*. Retrieved from https://ajax.systems/ua/air-alert/
3. Wikipedia. (2025). *Alerts.in.ua*. Retrieved from https://en.wikipedia.org/wiki/Alerts.in.ua
4. Telegram. (n.d.). *Bot API*. Retrieved from https://core.telegram.org/bots/api
5. Ukrainealarm.com. (n.d.). *API documentation*. Retrieved from https://api.ukrainealarm.com/swagger/index.html
6. Wikipedia. (2025). *Civil defense siren*. Retrieved from https://en.wikipedia.org/wiki/Civil_defense_siren
7. Mezha.net. (2024). *Chernihiv's localized air raid siren system*. Retrieved from https://mezha.net/eng/bukvy/chernihiv-s-new-air-raid-alert-sound-sparks-concern-and-feedback
8. Federal Communications Commission (FCC). (2023). *Wireless Emergency Alerts*. Retrieved from https://www.fcc.gov/consumers/guides/wireless-emergency-alerts-wea
9. UNESCO. (2022). *Crisis communication and digital platforms*. Retrieved from https://unesdoc.unesco.org/ark:/48223/pf0000381027
10. Reuters. (2025, July 19). *Sleepless Kyiv: how Ukraine's capital copes with Russia's nighttime attacks*. Retrieved from https://www.reuters.com/business/aerospace-defense/sleepless-kyiv-how-ukraines-capital-copes-with-russias-nighttime-attacks-2025-07-19/

Дата першого надходження рукопису до видання: 23.09.2025
Дата прийнятого до друку рукопису після рецензування: 21.10.2025
Дата публікації:

УДК 004.432.2 : 004.738.5 DOI

О. А. ПОНОМАРЬОВА
кандидат технічних наук, доцент,
доцент кафедри комп'ютерних наук, інформаційних технологій
та прикладної математики
Український державний університет науки і технологій
ННІ «Придніпровська державна академія будівництва та архітектури»
ORCID: 0000-0003-1254-4403

Г. М. ГУЗЬ
асистент кафедри спеціалізованих комп'ютерних систем
Український державний університет науки і технологій
ННІ «Український державний хіміко-технологічний університет»
ORCID: 0009-0002-2908-8985

Д. В. ГУЗЬ
інженер-програміст
Товариство з обмеженою відповідальністю «АЛЛО»

РОЗРОБКА СИСТЕМИ РЕЄСТРАЦІЇ ТА АВТОРИЗАЦІЇ КОРИСТУВАЧІВ У КЛІЄНТ-СЕРВЕРНОМУ WEB-ДОДАТКУ

У сучасних умовах стрімкого розвитку інформаційних технологій та зростання кількості web-додатків питання безпечної та ефективної реєстрації й авторизації користувачів набуває особливої актуальності.

Дана стаття розглядає процес розробки системи реєстрації та авторизації користувачів у web-додатку з клієнт-серверною архітектурою, що забезпечує надійний контроль доступу до ресурсів та персоналізовану взаємодію з користувачем.

Під час розробки було розглянуто ключові вимоги до систем автентифікації, серед яких безпека, масштабованість, зручність у використанні та відповідність актуальним стандартам. Проаналізовано поширені підходи до реалізації клієнт-серверної архітектури, зокрема застосування REST API, JSON Web Token (JWT), протоколу HTTPS, а також фреймворків для фронтенду та бекенду, таких як React, Node.js і Express.

Розроблена система передбачає двоетапний процес: реєстрацію нового користувача, а також авторизацію з використанням токенів доступу.

Розроблена форма реєстрації у web-додатку включає вбудовану валідацію рядків, яка перевіряє коректність введених даних – зокрема формат електронної пошти, складність пароля, його підтвердження та унікальність Email-адреси. Після успішної реєстрації на вказану пошту надсилається лист із підтвердженням, що містить унікальний токен для активації облікового запису. Після підтвердження користувач автоматично перенаправляється до вікна авторизації, де може увійти до системи за допомогою своїх облікових даних, що забезпечує безпечний і контрольований доступ до ресурсу.

Запропоноване рішення може бути інтегроване в різноманітні web-платформи, можливе масштабування, розширення функціоналу та інтеграції з іншими сервісами, зберігаючи високий рівень зручності використання для кінцевих користувачів. Розроблена система реєстрація та авторизації може бути застосована у різних галузях, де необхідно забезпечити надійну ідентифікацію користувачів та захист інформаційних ресурсів.

Ключові слова: *реєстрація; авторизація; клієнт-серверна архітектура; валідація; web-додаток, форма.*

O. A. PONOMAROVA
Candidate of Technical Sciences, Associate Professor,
Associate Professor at the Department of Computer Science,
Information Technologies and Applied Mathematics
Ukrainian State University of Science and Technology
of SEI "Prydniprovska State Academy of Civil Engineering and Architecture"
ORCID: 0000-0003-1254-4403

© Пономарьова О. А., Гузь Г. М., Гузь Д. В., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

H. M. HUZ
Assistant at the Department of Specialized Computer Systems
Ukrainian State University of Science and Technology
of SEI "Prydniprovska State Academy of Civil Engineering and Architecture"
ORCID: 0009-0002-2908-8985

D. V. HUZ
Software Engineer
Limited Liability Company "ALLO"

DEVELOPMENT OF A USER REGISTRATION AND AUTHORIZATION SYSTEM FOR A CLIENT-SERVER WEB APPLICATION

In the current context of rapid development in information technologies and the growing number of web applications, the issue of secure and efficient user registration and authorization has become increasingly relevant. This article explores the process of developing a user registration and authorization system within a web application based on a client-server architecture, which ensures reliable access control and personalized user interaction.

During development, key requirements for authentication systems were considered, including security, scalability, usability, and compliance with modern standards. Common approaches to implementing client-server architecture were analyzed, particularly the use of REST API, JSON Web Token (JWT), HTTPS protocol, and frontend/backend frameworks such as React, Node.js, and Express.

The developed system features a two-stage process: registration of a new user and authorization using access tokens.

The registration form in the web application includes built-in string validation that checks the correctness of input data – the email format, password complexity, password confirmation, and uniqueness of the email address. Upon successful registration, a confirmation email is sent to the specified address containing a unique token for account activation. After confirmation, the user is automatically redirected to the login window, where they can access the system using their credentials, ensuring secure and controlled resource access.

The proposed solution can be integrated into various web platforms, with potential for scalability, functional expansion, and integration with other services, while maintaining a high level of usability for end users. The developed registration and authorization system can be applied across different industries where reliable user identification and protection of information resources are required.

Key words: *registration; authorization; client-server architecture; validation; web application; form.*

Постановка проблеми

У сучасному цифровому світі web-додатки є невід'ємною частиною повсякденного життя, бізнесу та комунікацій. Більшість з них вимагають взаємодії з користувачами, що передбачає наявність механізмів реєстрації та авторизації для забезпечення безпеки, персоналізації та контролю доступу.

Відсутність надійної системи автентифікації може призвести до витоків даних, несанкціонованого доступу та втрати довіри користувачів.

Проблема полягає в розробці ефективної, безпечної та масштабованої системи реєстрації та авторизації для web-додатку з клієнт-серверною архітектурою [1].

Клієнт-серверна модель передбачає поділ на клієнтську частину (frontend), яка взаємодіє з користувачем, та серверну частину (backend), яка обробляє запити, зберігає дані та забезпечує логіку автентифікації.

Отже, постає потреба у створенні комплексного рішення, яке поєднує надійний захист персональних даних, зручність для користувачів, масштабованість системи та відповідність сучасним веб-технологіям. Впровадження такого рішення сприятиме підвищенню безпеки web-додатків, покращенню користувацького досвіду та ефективному управлінню доступом у клієнт-серверній архітектурі.

Аналіз останніх досліджень і публікацій

Сучасні дослідження у сфері реєстрації та авторизації користувачів у клієнт-серверних web-додатках свідчать про поступовий перехід від традиційних методів, таких як паролі та SMS-коди, до безпечніших безпарольних рішень, зокрема passkeys.

Використання криптографії та апаратних засобів підвищує захист від фішингу та інших кіберзагроз. Стандарти WebAuthn Level 3 [2] і FIDO2 [3] визначають принципи безпечної автентифікації, а NIST SP 800-63B [4] пропонує рекомендації щодо різних рівнів автентифікації. Попри технічні переваги, широке впровадження обмежують питання відновлення доступу, сумісності з існуючими системами та регуляторні вимоги.

Формулювання мети дослідження

Метою розробки є проєктування та реалізація модуля реєстрації й авторизації користувачів у web-додатку з клієнт-серверною архітектурою, який забезпечує безпеку обробки даних, стійкість до атак та масштабованість для роботи з великою кількістю користувачів.

Викладення основного матеріалу дослідження

Форма реєстрації/авторизації – це інтерфейс, який дозволяє користувачам створювати обліковий запис на вебсайті або у додатку (реєстрація) та входити у вже існуючий (авторизація).

Форма реєстрації призначена для створення нового облікового запису. Як правило, містить наступні поля для введення:

– Ім'я користувача (логін);
– Електронна адреса;
– Пароль (його повторного введення для підтвердження);
– Іншої додаткової інформації (наприклад, номер телефону, дата народження).

Форма авторизації служить для входу в систему, якщо користувач вже має обліковий запис. Вона зазвичай містить поля для введення: Ім'я користувача (логін), Пароль.

Розробку починали зі створення візуального інтерфейсу вікна реєстрації. Для побудови форми у вікні реєстрації з полями для заповнення використовували Ant Design. Для реєстрації у web-додатку користувач повинен заповнити наступні поля: E-mail, Ваше ім'я, Пароль та Повторити пароль (рис. 1) та натиснути кнопку Зареєструватися [5].

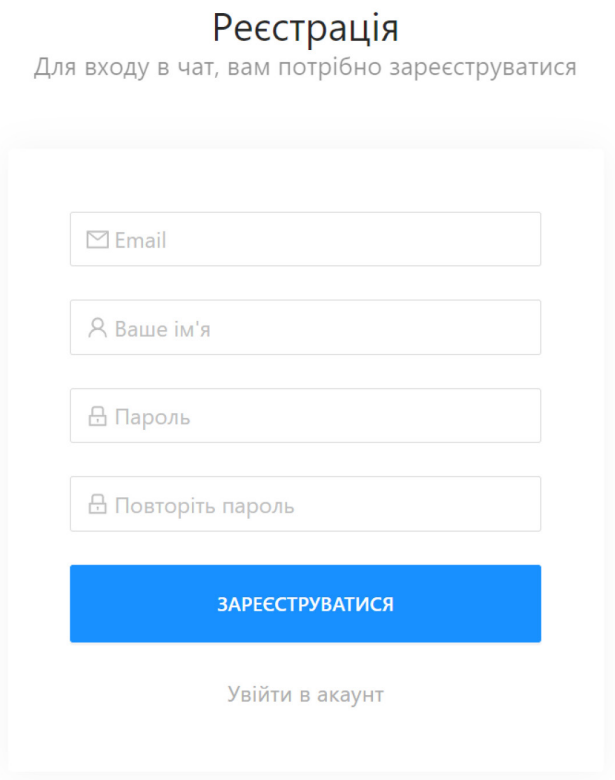


Рис. 1. Форма реєстрації

Як бачимо, форма розташована в центральній частині екрана. Це забезпечує фокус уваги користувача на основній дії – введенні даних. Вона складається з кількох логічних елементів.

Заголовок «Реєстрація» – чіткий, великий шрифт, розташований по центру. Він дає зрозуміти, у якій частині процесу перебуває користувач.

Пояснювальний підзаголовок – менший текст «Для входу в чат, вам потрібно зареєструватися». Виконує роль додаткової інструкції та мотивації.

Чотири поля вводу:

– Email – поле з іконкою конверта (символ пошти).
– Ваше ім'я – поле з іконкою користувача.
– Пароль – поле з іконкою замка, приховане введення.
– Повтор пароля – аналогічне до попереднього, але для підтвердження правильності введення.

Кнопка «Зарєструватися» – велика, синього кольору, з білим текстом, що чітко виділяється серед інших елементів. Такий колір часто обирають для основних дій, адже він привертає увагу.

Посилання «Увійти в акаунт» – розташоване під кнопкою, сірого кольору, меншого розміру, щоб не конкурувати з основною дією, але при цьому бути доступним.

Якщо під час реєстрації користувач залишить обов'язкові поля незаповненими, інтерфейс форми повідомить про некоректно введені дані (рис. 2). Перевірка коректності здійснюється за допомогою пакета Formik.

У разі коректного заповнення всіх полів форма реєстрації відображатиметься у вигляді, показаному на рис. 3.

Після натискання кнопки «Зареєструватися» виконується AJAX-запит, у якому введені користувачем дані передаються на сервер за вказаним маршрутом методом POST (створення маршрутів на сервері реалізовано за допомогою Express).

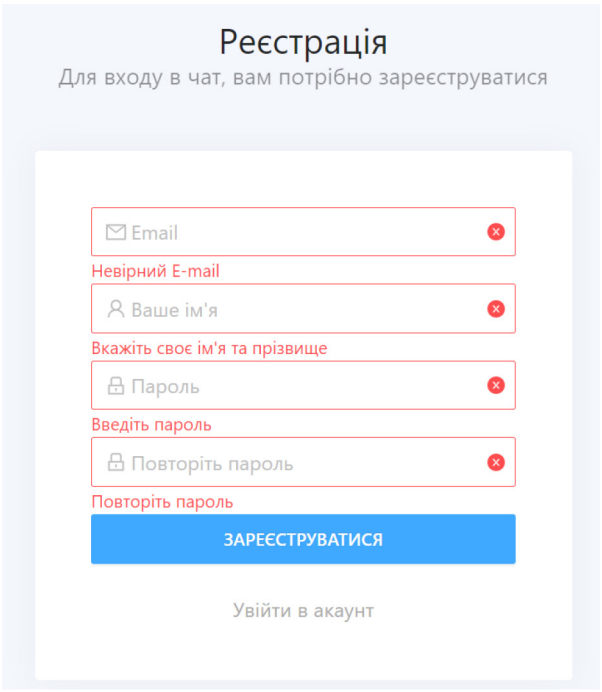


Рис. 2. Некоректне введення даних в форму реєстрації

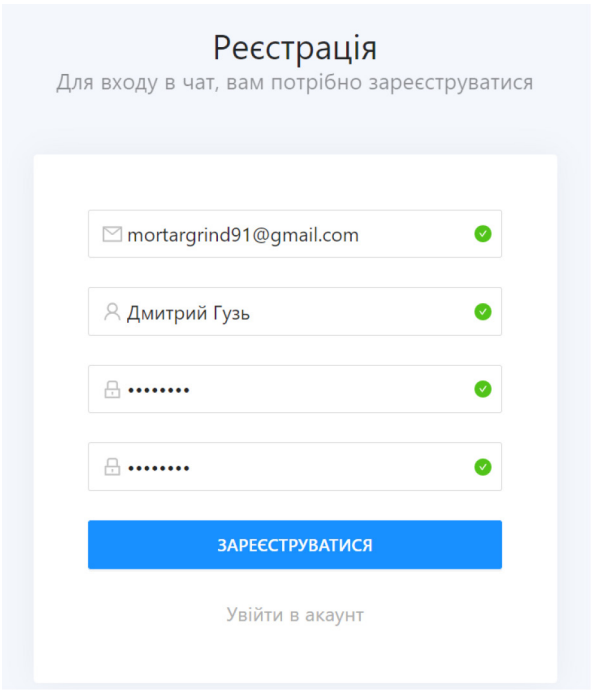


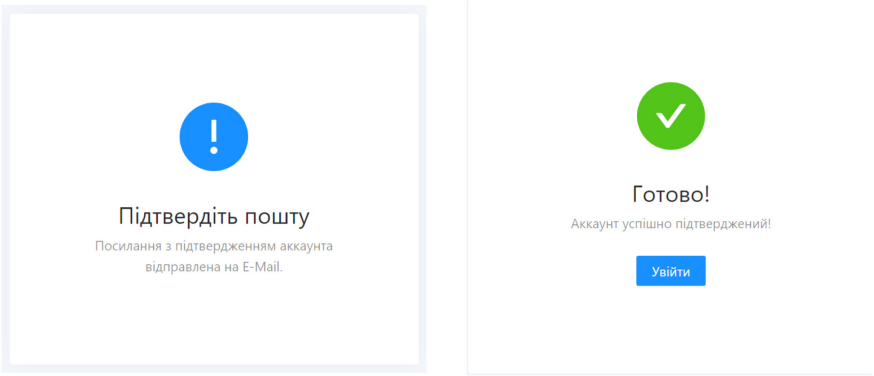
Рис. 3. – Коректне введення даних в форму реєстрації

Отримана інформація зберігається у базі даних MongoDB, а також генерується хеш для подальшого підтвердження електронної пошти користувача. Після цього система переводить користувача на етап верифікації його E-mail (рис. 4, *а*).

На електронну пошту користувача надсилається лист, що містить згенероване посилання з URL-адресою та унікальним хешем, призначеним для підтвердження E-mail. Після натискання на отримане посилання користувач переходить на сторінку з підтвердженням успішного E-mail та кнопкою для входу в акаунт (рис. 4, *б*).

Щоб увійти у веб-додаток, потрібно натиснути кнопку «Увійти». Після цього відкривається сторінка авторизації.

У вікні авторизації (рис. 5) слід ввести E-mail та пароль, а потім натиснути кнопку «Увійти в акаунт».



а) **б)**

Рис. 4. – Верифікація користувача:
а) Підтвердження E-mail; б) Успішне підтвердження E-mail

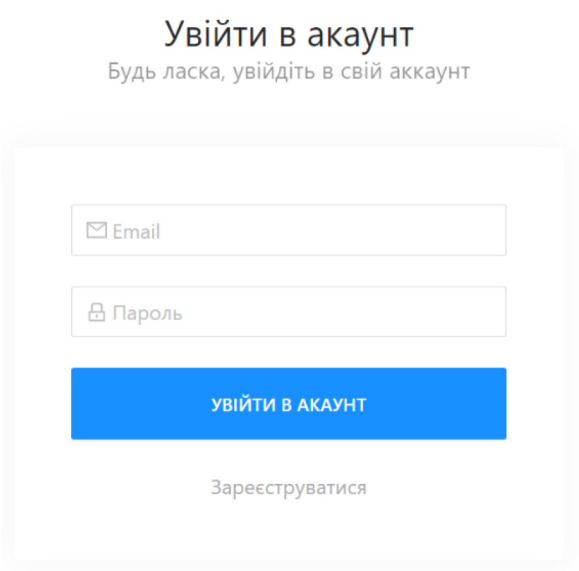


Рис. 5. Вікно авторизації

Якщо введені дані є некоректними або не збігаються з даними, зазначеними під час реєстрації, при натисканні кнопки «Увійти в акаунт» сервер повідомить про помилку введення (рис. 6).

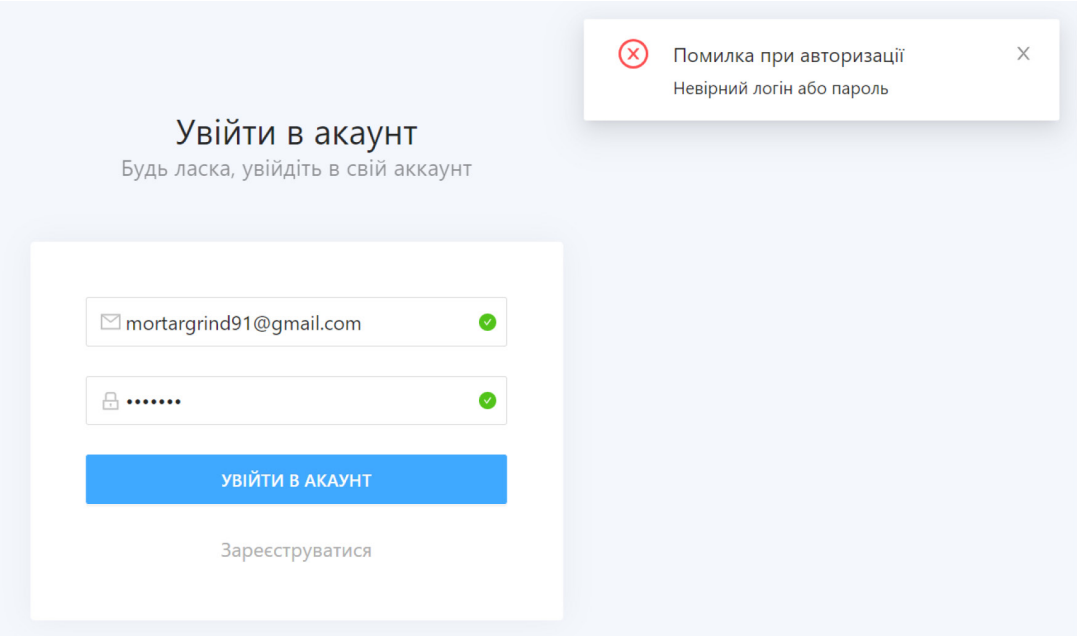


Рис. 6. Некоректні дані у вікні авторизації

Висновки

Розробка системи реєстрації та авторизації користувачів для web-додатку з клієнт-серверною архітектурою забезпечила створення надійного та ефективного механізму управління доступом, який гарантує безпеку зберігання конфіденційних даних, контроль прав користувачів та зручну взаємодію з інтерфейсом.

В процесі реалізації було створено інтуїтивно зрозумілий інтерфейс для введення даних, впроваджено комплексну валідацію на стороні клієнта та сервера, налаштовано обробку запитів і відображення повідомлень про успіх або помилки, а також застосовано сучасні заходи безпеки, включно з хешуванням паролів та захистом від несанкціонованого доступу.

Система продемонструвала високу практичну цінність, підвищуючи ефективність роботи web-додатку та рівень довіри користувачів, і має великий потенціал для подальшого розвитку, зокрема через впровадження функцій відновлення пароля, авторизації через соціальні мережі, розширене управління ролями користувачів та інтеграцію з іншими сервісами для забезпечення комплексної взаємодії.

Список використаної літератури

1. Мельник Р. А. Програмування веб-застосувань (фронт-енд та бек-енд). Львів : Львівська політехніка. 2018. 248 с.
2. Web Authentication: An API for accessing Public Key Credentials: веб-сайт. URL: https://www.w3.org/TR/webauthn-3/ (дата звернення 15.08.2025).
3. Passkeys: веб-сайт. URL: https://fidoalliance.org/passkeys/ (дата звернення 15.08.2025).
4. Digital Identity Guidelines Authentication and Lifecycle Management: веб-сайт. URL: https://pages.nist.gov/800-63-3/sp800-63b.html (дата звернення 15.08.2025).
5. Гузь Д. В., Пономарьова О. А. Розробка веб-додатку для спілкування між користувачами з використанням клієнт-серверних технологій. *Науковий процес та наукові підходи: методика та реалізація досліджень*: матеріали міжнародної конференції, 23 жовтня 2020 р. Одеса: МЦНД, 2020. С. 52–53.

References

1. Melnyk, R.A. (2018). Prohramuvannia veb-zastosuvan (front-end ta bek-end) [Web application programming (front-end and back-end)]. Lviv : Lvivska politekhnika. [in Ukrainian].
2. Web Authentication: An API for accessing Public Key Credentials. (2019). W3C. https://www.w3.org/TR/webauthn-3/
3. FIDO Alliance. (n.d.). Passkeys. https://fidoalliance.org/passkeys/
4. National Institute of Standards and Technology. (n.d.). *Digital identity guidelines: Authentication and lifecycle management* (SP 800-63B). NIST. Retrieved August 15, 2025, from https://pages.nist.gov/800-63-3/sp800-63b.html
5. Huz D. V., Ponomarova O. A. (2020) Rozrobka veb-dodatku dlia spilkuvannia mizh korystuvachamy z vykorystanniam kliient-servernykh tekhnolohii. [Development of a web application for communication between users using client-server technologies]. Naukovyi protses ta naukovi pidkhody: metodyka ta realizatsiia doslidzhen: materialy mizhnarodnoi konferentsii. pp. 52–53.

Дата першого надходження рукопису до видання: 20.09.2025
Дата прийнятого до друку рукопису після рецензування: 16.10.2025
Дата публікації:

УДК 004.414.23.05:004.77 DOI

І. В. РУБАН
доктор технічних наук, професор,
професор кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0000-0002-4738-3286

В. М. ТКАЧОВ
кандидат технічних наук, доцент,
докторант кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0000-0002-6524-9937

БАГАТОРІВНЕВА МОДЕЛЬ ІНФОРМАЦІЙНОЇ СИСТЕМИ НА МОБІЛЬНІЙ ПЛАТФОРМІ ТА ФОРМАЛІЗАЦІЯ КРИТЕРІЇВ ЇЇ ЖИВУЧОСТІ

Стаття присвячена формалізації живучості інформаційних систем на мобільній платформі з урахуванням переривчастої зв'язності, часткової спостережуваності та жорстких ресурсно-часових обмежень. Запропоновано багаторівневу модель (рівні даних, процесів і управляюче-ресурсний), в якій узгодження між рівнями фіксуються у вигляді інваріантів причинно-часової консистентності та SLO-вимог. Уведено клас допустимих траєкторій функціонування та ядро живучості як множину станів, з яких існує політика, що утримує систему в допустимій області попри збурення і «вікна доступності». Для ризик-чутливої оцінки деградації процесів використано механізм CVaR над дефіцитами SLO, що дозволяє враховувати події з «тяжкими хвостами». Причинно-часову адекватність даних формалізовано через метрики частки порушень причинності та втрати валідності, а здійсненність політик – через канальні, обчислювальні та енергетичні бюджети з урахуванням часових обмежень на їх виконання. Запропоновано три індекси живучості (найслабкішої ланки, зважений, послідовний), які агрегують нормовані показники ризику процесів, консистентності даних і часово-ресурсної здійсненності, що дає прозорий інструмент пріоритезації режимів та вибору керованих сценаріїв деградації. Новизна полягає у вперше поданій у єдиному формалізмі композиційній моделі, яка інтегрує ризик процесів, причинно-часову семантику даних і планову здійсненність політик та забезпечує операторне обчислення ядра живучості для швидкої діагностики стану в реальному часі. Практична цінність полягає у можливості оперативного порівняння політик і регламентів експлуатації для БПЛА/робототехнічних платформ, своєчасного виявлення «вузьких місць» і раціонального перерозподілу ресурсів з метою підтримання критичних процесів інформаційної системи у динамічних умовах.

Ключові слова: *інформаційна система, мобільна платформа, живучість, переривчаста зв'язність, ядро живучості, індекс живучості*

I. V. RUBAN
Doctor of Technical Sciences, Professor,
Professor at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0000-0002-4738-3286

V. M. TKACHOV
Candidate of Technical Sciences, Associate Professor,
Postdoctoral Student at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0000-0002-6524-9937

MULTILEVEL MODEL OF AN INFORMATION SYSTEM ON A MOBILE PLATFORM AND FORMALIZATION OF ITS SURVIVABILITY CRITERIA

The article is devoted to the formalization of the survivability of information systems on a mobile platform under intermittent connectivity, partial observability, and stringent resource–time constraints. We propose a multilevel model (data, process, and control-and-resource layers) in which cross-layer agreements are captured as invariants of causal–temporal consistency and SLO requirements. We introduce a class of admissible operational trajectories and a survivability kernel defined as the set of states from which there exists a policy that keeps the system within the admissible region despite

© Рубан І. В., Ткачов В. М., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

disturbances and "availability windows." To obtain risk-sensitive assessments of service degradation, we employ CVaR over SLO shortfalls, which accounts for heavy-tailed events. Causal–temporal validity of data is formalized via metrics for the fraction of causality violations and loss of validity, while policy feasibility is constrained by communication-channel, computational, and energy budgets subject to execution deadlines. We further propose three survivability indices (weakest-link, weighted, and sequential) that aggregate normalized measures of process risk, data consistency, and time-and-resource feasibility, providing a transparent tool for prioritizing operating modes and selecting controlled degradation scenarios. The novelty lies in a first, unified formalism for a compositional model that integrates service risk, the causal–temporal semantics of data, and the planned feasibility of policies, enabling operator-based computation of the survivability kernel for rapid state diagnosis in real time. The practical value is the ability to rapidly compare policies and operating procedures for UAV/robotic platforms, promptly identify bottlenecks, and rationally reallocate resources to sustain critical services in dynamic environments.

Key words: *information system; mobile platform; survivability; intermittent connectivity; survivability kernel; survivability index.*

Постановка проблеми

В умовах російської агресії проти України особливої ваги набуває застосування інформаційних систем на мобільних платформах, що працюють у середовищах із переривчастою зв'язністю та дефіцитом ресурсів [1–2]. Відтак забезпечення їхньої живучості стає необхідною умовою безперервного виконання покладених функцій.

У фундаментальних роботах вітчизняних та іноземних наукових шкіл пропонуються багатовекторні підходи – від функціональної та структурної живучості інформаційних систем до застосування моделей, що залучають апарат складних мереж і перколяції для виявлення критичних порогів збереження зв'язності та працездатності системи [3–7].

Разом із тим, класична теорія живучості інформаційних систем сформована переважно для статичних або інфраструктурно-центрованих інформаційних систем, де припущення про відносну сталість ресурсного середовища, стабільну топологію та безперервну зв'язність є прийнятними [8–11]. У інформаційних системах, які функціонують на базі мобільних платформ, ці припущення не завжди застосовні [12–13]. Це пов'язано з тим, що простір станів інформаційної системи може еволюціонувати під впливом переривчастої зв'язності, динамічної зміни топології та «ковзних» ресурсно-часових обмежень. При цьому траєкторії функціонування інформаційних систем можуть бути фрагментарними, а режими деградації унаслідок зовнішнього впливу – багатокритеріальними та контекстно-залежними [14].

Проведено аналіз класичних метрик інформаційних систем, зокрема таких як готовність, MTTF/MTTR, індекси зв'язності елементів інформаційної системи та ентропійні показники. Результати аналізу вказують на те, що метрики адекватно відображають загальний стан функціональної спроможності інформаційної системи, якщо вона функціонує в умовах статичних середовищ. Втім, якщо інформаційна система розгорнута на мобільній платформі (в умовах динамічного середовища), зазначені метрики дають істотні похибки, оскільки не охоплюють: ризику рідкісних, але «важких» збурень; причинно-часову узгодженість даних за переривчастих зв'язків між елементами інформаційної системи; досяжність в реалізації політик управління інформаційною системою в межах жорстких часових обмежень на прийняття рішень.

Виходячи з того, що ці показники не враховують структуру «тяжких хвостів» збурень і навантажень, причинно-часову семантику даних в умовах переривчастої зв'язності, часово-ресурсну здійсненність політик управління за наявних обмежень часу та бюджетних обмежень, – виникає задача побудови багаторівневої моделі живучості інформаційної системи на мобільній платформі, яка формалізує простір станів та аксіоматику у трирівневій структурі (управляюче-ресурсний, процесний рівень і рівень даних); визначає клас допустимих траєкторій і ядро життєздатності, що гарантують підтримку мінімального набору сервісних (процесних) підсистем, без яких цільова функція системи не виконується у режимах деградації; уводить ризик-чутливий критерій процесної адекватності для врахування подій із «тяжкими хвостами» (heavy-tailed) [15-16]; формалізує причинно-часову консистентність даних за переривчастої зв'язності між елементами моделі; визначає інтегральний показник часово-ресурсної здійсненності політик управління під заданими обмеженнями.

Аналіз останніх досліджень і публікацій

Аналіз досліджень у сфері моделювання інформаційних систем на мобільній платформі демонструють зміщення вектору досліджень до моделей забезпечення працездатності переважно прикладних систем, однак більшість досліджень залишаються вузькоорієнтованими без інтеграції процесної адекватності та семантики даних у єдину модель.

Наприклад, у статті [17] сформульовано ієрархічну задачу розміщення безпілотних базових станцій і організації маршрутизації між ними; застосовано MILP-модель і кластерно-метаевристичний алгоритм для одночасної оптимізації покриття і довжин маршрутів. Попри інженерну цінність, робота зосереджується на мережно-операційному рівні, зокрема, авторами не вводяться ризик-чутливі критерії для подій із «тяжкими хвостами», відсутня формалізація причинно-часової консистентності даних у режимі переривчастої зв'язності. У підсумку модель,

запропонована у роботі [17] корисна для планування топології, але не дає цілісної відповіді на питання живучості процесів і даних у динамічному середовищі.

Так, у статті [18] запропоновано ризик-орієнтоване планування енергетичного ресурсу для мікромереж із використанням підходу CVaR, що переконливо моделює дефіцит енергії та невизначеність генерації/споживання. Проте ризик трактовано на ресурсному шарі, без підняття його на рівень процесної адекватності, не враховано переривчасту зв'язність і узгодженість даних, хоча останнє варіює як предмет подальших досліджень.

Таким чином, найбільш дотичні роботи або оптимізують мережні структури, або керують окремими ресурсами в умовах невизначеності, проте у них відсутнє представлення багаторівневої моделі, що одночасно охоплює ризик для процесів інформаційної системи, причинно-часову консистентність даних і часово-ресурсну здійсненність політик.

Формулювання мети дослідження

Мета роботи полягає у розробці багаторівневої моделі інформаційної системи на мобільній платформі, що формалізує аксіоматику та простір станів на кожному з рівнів, визначає клас допустимих траєкторій і ядро життєздатності, а також формалізації критеріїв живучості зазначеної інформаційної системи: ризик-чутливу процесну адекватність для подій із «тяжкими хвостами», причинно-часову консистентність даних за переривчастої зв'язності та інтегральну часово-ресурсну здійсненність політик.

Викладення основного матеріалу дослідження

Формалізація моделі. В рамках тематики дослідження та враховуючи [19-21], під інформаційною системою вважатимемо наступний кортеж:

$$\mathrm{I} = \langle U, D, S, R, P, Q, E \rangle, \tag{1}$$

де: U – множина управляючих дій в інформаційній системі; D – простір даних, що використовується інформаційною системою у процесі реалізації нею основної функції; S – множина процесів інформаційної системи, що оперує з простором даних D; R – сукупність абстрактних ресурсів інформаційної системи; P – сукупність політик функціонування інформаційної системи у процесі реалізації нею основної функції; Q – сукупність вимог якості та логіко-часових константних інваріантів; E – сукупність класів факторів впливу зовнішнім середовищем.

Нехай траєкторія функціонування інформаційної системи описується переходами станів у рамках виконання нею основної функції як:

$$x_{t+1} = F(x_t, u_t, e_t), \quad x_t \in X, \quad u_t \in U(x_t), e_t \in E, \tag{2}$$

де: X – простір станів інформаційної системи; u_t – управляючі дії інформаційної системи у процесі її функціонування; e_t – клас факторів впливу зовнішнім середовищем на інформаційну систему.

Відповідно, на рівні архітектури інформаційної системи простір станів інформаційної системи X може бути декомпонований на такі рівні:

$$X = X_D \times X_S \times X_R, \tag{3}$$

де: X_D – рівень даних; X_S – процесний рівень; X_R – управляюче-ресурсний рівень.

Взаємодію між рівнями можна задати як міжрівневі погодження: Ξ_{DS} – семантика послідовності подій і їх консистентності, що необхідні для коректного виконання процесів; Ξ_{SR} – вимоги процесів до абстрактних ресурсів і допустимих втрат у процесі доступу до них.

Зазначена формалізація виду інформаційної системи (1–3) узгоджує класичні погляди з сучасною SLO-концепцією [22]. Однак, розглядаючи інформаційну систему як систему, яка функціонує на базі мобільної платформи (БПЛА, рій БПЛА або інші робототехнічні платформи), виходячи із задачі, що сформульована у статті, варто врахувати той факт, що зазначена платформа додає часову та топологічну мінливість і часткову спостережуваність [23].

Перехід від класичної інформаційної системи до інформаційної системи на мобільній платформі формально означає перехід від (майже) стаціонарного середовища до динамічного, в якому змінюється логічна топологія зв'язності елементів інформаційної системи, набувають динамічного характеру, за рахунок факторів мобільності та наявності зовнішнього впливу, обсяги таких характеристик як ємність каналів обміну даними, продуктивність обчислень, ресурсний запас дій. У такому середовищі функціонування інформаційної системи доцільно застосовувати термін «процес» як одиницю моделювання, так як саме процеси найчастіше еволюціонують у часі під впливом зазначених факторів мобільності.

Динамічний контекст у формалізованій формі може бути заданий як сукупність таких ознак:

– темпорально-топологічна мінливість, при якій логічна структура інформаційної системи може бути задана динамічним графом $G = (V, E_t)$ в якому кожне логічне ребро $e = (i, j)$, $e \in E$ має бінарну доступність $a_{e,t} \in \{0, 1\}$ та ємнісну місткість $c_{e,t} \geq 0$. Мобільність платформи в даному випадку індукує процес доступності $a_{e,t}$ з періодами доступності / недоступності елементів інформаційної системи в залежності від дії факторів мобільності без настання критичної відмови у функціонуванні системи, однак тривалість періодів недоступності може мати характер «тяжкого хвосту»;

– часткова спостережуваність, коли стан інформаційної системи, що заданий як $x_t = (x_t^D, x_t^S, x_t^R)$, , спостерігається через $o_t = h(x_t)$, що включає параметри агрегованої телеметрії (при наявності ознак розподіленої інформаційної системи), прогнозований стан вікон зв'язності елементів інформаційної системи, індикатори стану процесів тощо. Через наявні затримки оцінки стану інформаційної системи, спричинені факторами мобільної платформи, політика p реалізується інформаційною системою з неповною та застарілою інформацією;

– наявність вікон можливостей, коли дані, передаються за відомою моделлю «opportunity-driven» [24], коли $a_{e,t} = 1$, дозволені пересилки з обмеженнями по $c_{e,t}$, а між вікнами – за принципом буферизації та подальшої ретрансляції (логічна буферизація провадиться на рівнях X_D та X_S);

—формалізація політик функціонування інформаційної системи як реплікації артефактів даних у дозволені часові вікна, зміни режимів виконання процесів згідно з даними про деградацію системи та сукупності розподілу заданих бюджетів системи планування задач інформаційної системи з кінцевими часовими точками виконання:

$$B = \left\{ B_t^{channel}, B_t^{CPU}, B_t^{energy} \right\}, \tag{4}$$

де: $B_t^{channel}$ – канальний бюджет (впливає на визначення значення пропускної здатності у вікні зв'язку); B_t^{CPU} – обчислювальний бюджет (впливає на визначення значення умовних обчислювальних квантів, які можна витратити на задачі, що вирішуються інформаційною системою); B_t^{energy} – енергетичний бюджет (впливає на час виконання задачі з урахуванням обмежень мобільної платформи).

Враховуючи вищезазначені ознаки та формалізацію моделі інформаційної системи маємо наступні проміжні висновки. По-перше, міжрівневі домовленості можна розглядати як інваріанти. Це обґрунтовується тим, що Ξ_{SR} пов'язує ресурси інформаційної системи з SLO-вимогами до процесів [22], а Ξ_{DS} фіксує причинно-часову семантику даних (наприклад, допустиму частку порушень порядку подій, що граничний часовому проміжку валідності інформації для інформаційної системи). При цьому порушення будь-яких взаємодомовленостей призводить до переходу стану інформаційної системи у допустиму область $X \backslash G$.

По-друге, збурення у функціональній здатності інформаційної системи (включно з випадками «тяжких хвостів» під час вікон недоступності) оцінюються через дефіцит SLO критичних процесів, що виступає у вигляді умови як первинна функція ризику для прийняття рішень щодо обрання політики функціонування інформаційної системи.

По-третє, фактори мобільності призводять до розривів причинного порядку та втрати валідності даних, що циркулюють в інформаційній системі, тому модель у явному вигляді містить метрику частки порушень причинності μ та граничної втрати валідності даних η, і, як наслідок, визначають політики функціонування інформаційної системи, що керують функціональною здатністю у вікнах доступності.

Виходячи з (4), формалізація траєкторії на рівнях (3) з урахуванням дії політик інформаційної системи може бути формалізована, виходячи з того, що управляючі дії інформаційної системи у процесі її функціонування (2) задаються як $u_t = (u_t^D, u_t^S, u_t^R)$ для кожного з рівнів:

– для рівня даних X_D:

$$x_{t+1}^D = F_D(x_t^D, u_t^D, w_t^D, a_t), \tag{5}$$

де: x_t^D – мітки часу, версійні сценарії, причинні залежності; u_t^D – процеси реплікації, маршрутизації, узгодження; w_t^D – стохастика процесів навантаження/доступності; a_t – показник доступності логічних каналів у розрізі впливу факторів мобільності платформи.

Індикатори μ, η в таких умовах оновлюються з урахуванням появи вікон доступу;

– для процесного рівня X_S:

$$x_{t+1}^S = F_S(x_t^S, u_t^S, w_t^S, a_t, c_t), \tag{6}$$

де: x_t^S – індикатори SLO мінімального критичного набору даних, який задається на етапі формулювання основної функції інформаційної системи; u_t^S – процеси перемикання політик/режимів функціонування інформаційної системи, пріоритезації процесів; c_i – показник місткості логічних каналів у розрізі впливу факторів мобільності платформи;

– для управляюче-ресурсного рівня X_R:

$$x_{t+1}^R = F_R(x_t^R, u_t^R, w_t^R), \tag{7}$$

де: x_t^R – сукупність (4), черг завдань, граничних часових точок виконання процесів; u_t^R – рішення про розподіл (4) і запуск / пріоритизацію процесів.

Збурення $w_t = (w_t^D, w_t^S, w_t^R)$ допускають появу «тяжких хвостів», на кшталт, субекспоненціальних та α-стабільних хвостів, що прямо впливають на величину ризик-чутливої метрики процесної адекватності S_α.

Звідси маємо наступні композиційні обмеження, які впливають на розмір допустимої області G з урахуванням ризик-процесів S_α, причинно-часової узгодженості, темпорально-ресурсної досяжності цілі та вимоги до виконання міжрівневих домовленостей:

$$G = \begin{cases} S_\alpha(x_t) \geq \vartheta_s; \\ \mu(x_t) \leq \varepsilon, \eta(x_t) \leq \Delta; \\ S(x_t) \neq \varnothing; \\ \Xi_{DS}(x_t), \Xi_{SR}(x_t) - \textit{виконані}. \end{cases} \quad (8)$$

де ε, Δ – порогові значення.

Враховуючи (1)–(8), модель інформаційної системи на мобільній платформі можна подати у наступному вигляді:

$$I_{m.p.} = \langle X, U, O, F, H, E, G_t, a_t, c_t B_t, \Xi_{DS}, \Xi_{SR}, \mathrm{M}(x), G \rangle, \quad (9)$$

де O – простір спостережень; $F : X \times U \times E \to X$ – динаміка станів інформаційної системи при (5)–(7); $H : X \times E \to O$ – відображення спостережень, виходячи із другої ознаки динамічних контекстів факторів мобільності.

Окремо варто пояснити деякі елементи моделі, які непрямо виходять з (1). Так, O – простір спостережень варто розглядати як множину всіх вимірів станів інформаційної системи, які доступні політиці управління в момент t замість повного стану x_t. Так як через переривчасту зв'язність і затримки інформаційна система бачить не повний стан, а лише його «проєкцію»:

$$o_t \in O, \quad o_t = H(x_t, e_t) + \psi, \quad (10)$$

де: H – випадкове ядро спостережень від стану x_t та впливів середовища e_t до спостережуваних величин; ψ – пропуски, які дають неповноцінну картину стану системи.

До складу o_t в моделі (9) входять оцінки бюджетів ресурсів, індикатори стану процесів, показники причинно-часової узгодженості, вікна доступу та місткості логічних каналів, часові мітки спостережень – тому у загальному вигляді O можна формалізувати як:

$$O = O_D \times O_S \times O_R \times \{0,1\}^{|E|} \times \mathbb{R}_+^{|E|} \times \{0,1\}^p, \quad (11)$$

а статистичний зв'язок задає ядро спостережень

$$\mathbb{P}(o_t \in \cdot \mid x_t, e_t) = O(\cdot \mid x_t, e_t), \quad (12)$$

яке враховує пропуски, які дають неповноцінну картину стану системи. Якщо спостереження стають повними (тобто o_t містить увесь x_t), задача зводиться до класу класичних задач, в яких ухвалення рішення є таким, в якому наслідки є частково випадковими, а частково контрольовані системою, яка ухвалює рішення. В нашому випадку – у мобільному середовищі – задачі стоять саме з частковою спостережуваністю.

Інший компонент моделі (9) $\mathrm{M}(x)$ є вектором індикаторів живучості інформаційної системи на мобільній платформі та визначається, виходячи з (8) як:

$$\mathrm{M}(x) = (S_\alpha(x), C(x), F(x)), \quad (13)$$

де кожен елемент відображає окремий аспект підтримання критичних процесів інформаційної системи у мобільному середовищі.

Так, $S_\alpha(x)$ – індикатор ризик-чутливої процесної адекватності, який вимірює наскільки безпечний «хвіст» втрат процесної якості для мінімального набору критичних процесів інформаційної системи Ω. Значення розраховується, у два етапи. Спочатку визначається дефіцит SLO для процесів:

$$\varpi_S(x) = \sum_{s \in \Omega} w_s \underbrace{\max\left\{0, \frac{m_s(x) - g_s}{g_s}\right\}}_{\textit{нормований надлишок над ціллю } g_s}, \ \sum w_s = 1, \quad (14)$$

де $m_s(x)$ – фактичний індикатор (затримка, втрата тощо); g_s – власне, ціль.

Далі застосовується відомий апарат CVaR [23] як спосіб визначення умовного математичного сподівання в хвості рівня α:

$$S\alpha(x) := -CVaR_\alpha[\varpi_S(x)]. \quad (15)$$

У прикладній площині, чим більше $S_\alpha(x)$, тим менший «тяжкохвостий» ризик зупинки процесу.

У (13) $C(x)$ – індикатор причинно-часової консистентності даних, який вимірює узгодженість порядку подій і актуальність даних за переривчастої зв'язності у процесі виконання інформаційною системою основної функції. Індикатор розраховується як комбінація двох складових μ та η. Виходячи із того, що $\mu(x) \in [0, 1]$, а $\eta(x) \geq 0$, то доцільно інтерпретувати $C(x)$ як зважену композицію з нормуванням по η:

$$C(x) = \lambda\mu(x) + (1 - \lambda)\varphi(\eta(x)), \quad \varphi(\eta(x)) = \min\left\{1, \frac{\eta(x)}{\Delta}\right\}, \quad \lambda[0,1]. \quad (16)$$

У прикладній площині, чим менше $C(x)$, тим більш узгодженіші дані між собою.

Третій індикатор у (13) – $F(x)$ – часово-ресурсна здійсненність політик, який встановлює чи існує реальний розклад виконання поточних політик елементом інформаційної системи у рамках часових обмежень і доступних бюджетів (4). Пропонується два способи розрахунку значення індикатору – булевий та ступеневий. У першому випадку маємо $F(x) = 1\{S(x) \neq \varnothing\}$. У другому: $F(x) = \max_{\sigma \in S(x)} \xi(\sigma)$, де $\xi \geq 0$ – мінімальний запас по бюджетах серед усіх задач поточної функції інформаційної системи.

Індикатор $C(x)$ може мати наступні значення:

– $F(x) > 0$ – політика може провадитися (є запас ресурсу);
– $F(x) = 0$ – політика може виконуватися (немає запасу ресурсу);
– $F(x) < 0$ або $F(x) = 0$ у булевому стані – політика не може виконуватися.

Аксіоматичний базис живучості та інформаційної системи на мобільній платформі та базові властивості. Далі доцільно зафіксувати логіко-математичні припущення для розробленої моделі інформаційної технології, на основі яких надалі визначаються допустимі траєкторії її функціонування та ядро живучості:

1) на рівні даних (3) задано часові мітки і відношення часткового порядку, тому жодна операція під час функціонування інформаційної системи не може індукувати ретроспективний вплив;

2) спостереження (10) для мобільних платформ є неповними та потенційно застарілими, тому політики u_t мають бути робастними до ψ;

3) середовище E породжує динамічний логічний граф (перший фактор мобільності) з бінарною доступністю a_t та обсягами місткості c_t, причому розриви зв'язності дозволені, а їх вплив відбивається виключно в елементах інформаційної системи;

4) існує мінімальний набір процесів, без якого цільова функція інформаційної системи не виконується. Кожен $s \in \Omega$ має ціль g_s та індикатор $m_s(x)$;

5) при фіксованих x^S, x^D невід'ємне збільшення бюджетів (4) не погіршує індикатори $m_s(x)$ та не збільшує μ, h;

6) ризик процесної деградації оцінюється на процесному рівні через міру CVaR дефіциту SLO, а параметр $\alpha \in (0, 1)$ задає глибину «тяжкого хвоста»;

7) для критичних процесів визначено частку порушень μ, η з допустимими порогами ε, Δ, які належать до вимог Q;

8) для поточного стану інформаційної системи x існує множина $S(x)$ виконання політик у часових обмеженнях та бюджетах (4). Вимога досяжності виконання основної функції інформаційної системи є обов'язковою;

9) порушення інваріантів узгодження Ξ_{DS} та Ξ_{SR} виводить стан інформаційної системи за межі допустимої області G.

Серед базових властивостей, які випливають із вищенаведеного, є:

1) властивість монотонності по ресурсам, яка пояснюється таким чином, що якщо $x, y \in X$ однакові на X_D, X_S та $B_t^x \leq B_t^y$ покомпетентно, тоді $S_\alpha(y) \geq S_\alpha(x)$, $C(y) \leq C(x)$, $F(y) \geq F(x)$;

2) властивість композиційності рівнів, яка пояснюється таким чином, що узгоджене покращення будь-якого рівня (виконання узгоджень Ξ_{DS}, Ξ_{SR}) не може вивести стан інформаційної системи за межі G;

3) властивість інваріантності станів інформаційної системи до коректних політик, яка пояснюється таким чином, що якщо для стану $x \in G$ існує політика, що не порушує Ξ_{DS}, Ξ_{SR} і підтримує (13) у межах порогових значень, то траєкторія не покидає G попри переривчастість a_t та варіації c_t.

Формалізація класу допустимих траєкторій моделі та ядра живучості. Подальша робота побудована у руслі визначення траєкторій стану інформаційної системи на мобільній платформі, які вважаються допустимими (тобто такими, що зберігають живучість з урахуванням $M(x)$ (13), Ξ_{DS}, Ξ_{SR}) та формалізації множини початкових станів запропонованої моделі, із яких існує політика, що утримує інформаційну систему в допустимій області попри переривчастість і збурення.

Нехай $x_t = (x_t^D, x_t^S, x_t^R) \in X_D \times X_S \times X_R$ – стан інформаційної системи у момент часу t; $o_t = H(x_t, e_t) \in O$ – стан спостереження (11); $u_t = \pi(o_{0:t}) \in U$ – причинна дія політики π, а середовище породжує збурення e_t (включаючи динаміку a_t та c_t). Позначимо допустиму область:

$$G = \{x \in X : S_\alpha(x) \geq \vartheta_S, C(x) \leq \vartheta_C, F(x) \geq \vartheta_F\}, \quad \textit{якщо } \Xi_{DS}(x), \Xi_{SR}(x) - \textit{виконані}. \tag{17}$$

Послідовність станів інформаційної системи $\{x_t\}_{t \in \mathbb{N}}$, згенерована динамікою $x_{t+1} = F(x_t, u_t, e_t)$ під політикою π називається допустимою, якщо за обраним типом гарантій виконується одна з умов:

– умова робастної допустимості $x_t \in G$ для всіх t для будь-яких збурень e_t з допустимого класу E;
– умова ймовірнісної допустимості (із рівнем довіри $1 - \varepsilon$) при

$$\mathbb{P}\{x_t \in G \ \forall \ t\} \geq 1 - \varepsilon. \tag{18}$$

У подальшому в пропонується використовувати робастну версію умов у якості консервативного критерію. Це обґрунтовується тим, що ймовірнісна умова дозволяє контролювати ризики тільки на рівні прийнятної ε.

Важливо, що на процесному рівні можуть існувати режими функціонування інформаційної системи $m_t \in M$ (номінальний, понижений або аварійний) з вказаними правилами переходів. Допустимість переходів (17) не забороняє здійснювати переходи в уже деградовані режими, якщо виконується (8) і зберігається мінімальний критичний зріз у відповідності до Ω.

Таким чином, маємо означення ядра живучості як множини всіх початкових станів, з яких існує причинна політика, здатна утримувати систему всередині G (робастно або з рівнем довіри $1 - \varepsilon$):

$$K^{v}_{robust} = \left\{ x \in X \mid \exists \pi \forall \left\{ e_t \right\} \subset E : x_t \in G \forall t, \quad x_{t+1} = F\left(x_t \pi(o_{0:t}), e_t \right) \right\}, \tag{19}$$

$$K^{v}_{\varepsilon} = \left\{ x \in X \mid \exists \pi : \mathbb{P}_{\pi} \left\{ x_t \in G \forall t \right\} \geq 1 - \varepsilon \right\}. \tag{20}$$

Операторна побудова (19) та (20) дозволить формалізувати простий спосіб обчислити, у яких станах інформаційна система гарантовано зберігає живучість. Така побудова перетворює логіку утримання інформаційної системи у допустимій області на чітку процедуру з формальними гарантіями, коли інваріантність і робастність фіксуються як стала точка ітерацій з «кроком назад». Такий підхід уніфікує перевірку допустимих траєкторій для різних сценаріїв мобільності й збурень і служить базою для подальшого синтезу політик, бо такий підхід відразу показує варіанти, при яких функціонування інформаційної системи можливе, а при яких – ні.

Нехай, оператори над підмножинами $\mathrm{A} \in X$ мають наступний вигляд відповідно до (19) та (20):

$$\Upsilon_{robust}(\mathrm{A}) := \left\{ x \in G \mid \exists u \in U(x) : \forall e \in E, \ F(x) \geq \vartheta_F, \quad S_\alpha(x) \geq \vartheta_S, \ C(x) \leq \vartheta_C, \ F(x,u,e) \in \mathrm{A} \right\}. \tag{21}$$

$$\Upsilon_{\varepsilon}(\mathrm{A}) := \left\{ x \in G \mid \exists u \in U(x) : \mathbb{P}\left(F(x,u,e) \in \mathrm{A} \wedge S_\alpha \geq \vartheta_S \wedge C \leq \vartheta_C \wedge F \geq \vartheta_F \right) \geq 1 - \varepsilon \right\}. \tag{22}$$

Тоді найбільша інваріантна підмножина G щодо Υ є фіксованою точкою:

$$K^{v}_{robust} = \bigcap_{k \geq 0} \Upsilon^{k}_{robust}(G), \quad K^{v}_{\varepsilon} = \bigcap_{k \geq 0} \Upsilon^{k}_{\varepsilon}(G).$$

Практично, маємо ітеративне звуження G «кроком назад», тобто відкидаємо стани інформаційної системи, з яких навіть найкраща дія не гарантує повернення в A з дотриманням умов (8).

Далі доцільно провести композиційні апроксимації, які трансформують задачу обчислення ядра живучості обчислюваною в реальних масштабах, тобто замість перевірки усього стану інформаційної системи, на що об'єктивно не вистачить часу в умовах мобільності платформи, на якій функціонує ця система, можна швидко перевірити прості локальні умови на рівнях (3). Така перевірка однозначно дасть об'єктивні внутрішні оцінки ядра живучості, придатні для прийняття швидких рішень щодо корекції стану інформаційної системи. Зазначений підхід є гнучким в контексті того, що відразу очевидно, який саме критерій не задовольняє умов функціонування інформаційної системи, а відповідно, і де має спрацювати корекція.

Для цього необхідно ввести спочатку три локальні множини:

$$K_S = \left\{ x : S_\alpha(x) \geq \vartheta_S \right\};$$

$$K_D = \left\{ x : C(x) \leq \vartheta_C \right\};$$

$$K_R = \left\{ x : F(x) \geq \vartheta_F \wedge \left(\Xi_{DS}, \Xi_{SR} - \textit{виконані} \right) \right\}.$$

Тоді:

$$K^V := \left(K_R \cap K_D \cap K_S \right) \cap \textit{узгодження потоків із } a_t, c_t. \tag{23}$$

Маємо внутрішню оцінку ядра живучості (23), яка є початковою множиною для ітерацій Υ. Відповідно, запас живучості можна формалізувати як:

$$\Gamma = \lim_{k^V \to 0^+} \left\{ S_\alpha(x) - \vartheta_S - k^V, \vartheta_C - k^V - C(x), F(x) - \vartheta_F - k^V \right\}, \ k^V \in K^V, \tag{24}$$

де значення Γ, яке більше 0, означає, що x лежить в межах K^V, а якщо Γ = 0, то на межі відсутності зазначеного запасу живучості.

Згідно з результатами проведеного аналізу визначено, що допустимі траєкторії – це напрями функціонування інформаційної системи, де критичні процеси Ω зберігають адекватність навіть за наявності переривчатості. Ядро живучості інформаційної системи формалізує області станів, із якої це принципово можна здійснювати завдяки наявності політик, що узгоджують ресурси, процеси та дані (через Ξ_{DS}, Ξ_{SR}). Включення ризику через S_α робить ядро живучості ризик-усвідомленим, причому рідкісні впливи, на кшталт, «тяжких хвостів» враховуються в самій дефініції допустимості.

Багатокритеріальні показники та цільова функція живучості за ресурсно-часових обмежень. Далі в дослідженні буде сформовано узгоджені способи агрегування таких показників як ризик-чутлива адекватність

процесів (реакція на «тяжкі хвости»), причинно-часова консистентність даних (переривчастість) та часово-ресурсна здійсненність політик (у межах бюджетів і часових обмежень) – у цільову функцію живучості для оцінювання станів і режимів та порівняння політик функціонування інформаційної системи на мобільній платформі.

Щоб уникнути дисбалансу у інтерпретуванні величин, необхідно виконати ряд перетворень індикаторів (13) до порогу [0,1]:

$$\tilde{S}(x) := \frac{1}{1+\mathrm{CVaR}_{\alpha}\left[\varpi_S(x)\right]} \in (0,1],$$

$$\tilde{C}(x) := 1-\left(\lambda\mu(x)+(1-\lambda)\varphi(\eta(x))\right) \in [0,1],$$

$$\tilde{F}(x) := \begin{cases} 1, & S(x) \neq \varnothing, \quad ma \quad \max\limits_{\sigma \in S(x)} \xi(\sigma) \geq \eta, \\ \dfrac{1}{1+\max\left\{0, \eta - \max\limits_{\sigma} \xi(\sigma)\right\}}, & \textit{інакше.} \end{cases}$$

Маємо:

$$\tilde{\mathrm{M}}(x) = \left(\tilde{S}_{\alpha}(x), \tilde{C}(x), \tilde{F}(x)\right) \in [0,1]^3. \tag{25}$$

Далі – згортаємо вектор нормованих показників (25) у індекси живучості:

– індекс найслабшої ланки $J_{\min}(x)$:

$$J_{\min}(x) := \min\left\{\tilde{S}(x), \tilde{C}(x), \tilde{F}(x)\right\}, \tag{26}$$

який показує «вузьке місце» інформаційної системи. Значення індексу визначається найгіршою з трьох нормованих компонент $\tilde{S}, \tilde{C}, \tilde{F}$. Сутність індексу полягає у тому, що жоден сильний компонент не може компенсувати критичну слабкість, тобто, якщо, здійсненність політик задовільна, а дані узгоджені, але при цьому виникає «тяжкий хвіст» процесних втрат (значне падіння $\tilde{S}$), то індекс буде низьким. Відповідно, індекс буде доречним у режимах, в яких порушення будь-якого критерію з вищерозглянутих –неприйнятне. Позитивними практиками є простота інтерпретації та вкрай висока чутливість до ризиків і помилок у на рівні даних. Однак серед мінусів варто зазначити той факт, що індекс може пригнічувати рішення, коли один з трьох нормованих компонент тимчасово просідає, а інші суттєво кращі. Крім цього індекс не дає градації у розрізі того, який поріг наближений до прийнятного, а який – на критичному рівні;

– зважений індекс $J_w(x)$:

$$J_w(x) := w_S\tilde{S}(x) + w_C\tilde{C}(x) + w_F\tilde{F}(x), \tag{27}$$

який дає змогу явно розставити пріоритети між ризиком процесів, консистентністю даних і здійсненністю досяжності цілі інформаційної системи, підібравши ваги w_S, w_C, w_F під профіль основної функції інформаційної системи. Коли потрібно здійснити пріоритезацію між критеріями (наприклад, знизити ризик відмови процесів, пожертвувавши часткою пропускної здатності непершочергових потоків в інформаційній системі) зазначений індекс є найзручніший. Позитивними практиками є гнучкість індексу, плавна реакція, придатність для автоматичного налаштування і швидкого пошуку рішень. Однак індекс має певні обмеження, зокрема, компенсаторний ефект, наприклад, суттєве падіння однієї з трьох нормованих компонент $\tilde{S}, \tilde{C}, \tilde{F}$ може маскуватися високими значеннями інших компонент. У якості обходу цього обмеження, можна поєднувати J_w з пороговими допусками на кожен з компонент, наприклад, перевіряти умови $S_{\alpha} \geq \vartheta_S$, $C \leq \vartheta_C$, $F \geq \vartheta_F$ до агрегування (25);

– послідовний індекс J_l:

$$J_l(x) := \left(1\left\{\tilde{F}(x) \geq \vartheta_F\right\}, 1\left\{\tilde{S}(x) \geq \vartheta_S\right\}, \tilde{C}(x)\right), \tag{28}$$

який використовує жорсткий порядок пріоритетів. Так, спершу здійснюється перевірка на предмет, чи взагалі можлива реалізація політик у часі та наявних бюджетах, далі – перевірка прийнятності процесної адекватності, далі – здійснюється максимізація консистентності даних. Такий підхід застосовний для інформаційних систем на мобільній платформі із жорсткими часовими обмеженнями та основною функцією з мінімально допустимим рівнем якості виконання процесів. Тобто, якщо на певному етапі функціонування інформаційної системи основна місія стає недосяжною для виконання інформаційною системою, то немає сенсу перетягувати інші компоненти за зміною їх пріоритетів, достатньо ініціювати перебудову функціональної моделі інформаційної системи. Позитивними практиками індексу є жорстка узгодженість з операційною логікою, прозорість рішень, відсутність небажаної компенсації. Однак серед недоліків варто зазначити стрибкоподібну поведінку при перетині порогів та, власне, чутливість до вибору порогів ϑ_F, ϑ_S. Також, у мінливих сценаріях без «тяжких хвостів», індекс може бути надто жорстким і відкидати цілком прийнятні компромісні траєкторії перерозподілу ресурсів між компонентами $\tilde{S}, \tilde{C}, \tilde{F}$.

Агрегування (25) застосовується лише до станів у допустимій області:

$$x \in G \Leftrightarrow \begin{cases} S_\alpha(x) \geq \vartheta_S, \\ C(x) \leq \vartheta_C, \\ F(x) \geq \vartheta_F, \\ \Xi_{DS}(x), \Xi_{SR}(x) - \textit{виконанані}, \\ \textit{потоки узгоджені з } a_t, c_t, \\ \textit{бюджети } B_t \textit{ дотримано}. \end{cases} \quad (29)$$

Таким чином, $J_{\min}$, J_w, J_l як індекси живучості, по суті, є оціночними функціоналами якості всередині G, тому вихід характеристик інформаційної системи за G трактується як нульова живучість інформаційної платформи на мобільній платформі.

Висновки

В роботі вперше представлено багаторівневу модель інформаційної системи на мобільній платформі, яка одночасно і в одному формалізмі враховує хвостато-ризикову деградацію процесів через CVaR-оцінку дефіциту SLO, причинно-часову консистентність даних під переривчастою зв'язністю мобільної платформи, часово-ресурсну здійсненність політик у межах динамічних бюджетів і «вікон доступності», причому міжрівневі взаємовідносини та операторна побудова ядра живучості забезпечують композиційні гарантії живучості інформаційної системи. На відміну від існуючих підходів, які розглядають ці аспекти окремо (мережно-операційні оптимізації та класичні метрики готовності), запропонована модель інтегрує їх у спільний простір станів і критеріїв та надає узгоджену методику оцінювання станів за індексами живучості.

Практична цінність розробленої моделі полягає у наданні цілісного, операційно придатного апарату для управління інформаційною системою на мобільній платформі, який поєднує ризик-орієнтоване оцінювання процесів, контроль причинно-часової узгодженості даних та перевірку здійсненності політик у наявних ресурсно-часових межах. Модель забезпечує формалізовану область допустимих станів і ядро живучості, що дозволяє завчасно виявляти критичні процесні точки, обирати керовані режими деградації і раціонально перерозподіляти ресурси для підтримання критичних процесів. Індексна діагностика стану дає прозорий механізм пріоритизації між якістю процесів, консистентністю даних і плановою здійсненністю політик, а композиційні перевірки роблять оцінювання швидким і масштабованим для застосування в режимі реального часу. Завдяки апаратній незалежності та міжрівневим погоджувальним механізмам модель слугує основою для регламентів експлуатації, сценарного тестування і перевірки вимог живучості у різнорідних інформаційних системах на мобільних платформах.

Список використаної літератури

1. Компанієць О. М. Інформаційна технологія адаптивного управління роєм безпілотних літальних апаратів в антагоністичному середовищі. *Системи обробки інформації*. 2025. № 4 (179). С. 43–48. URL: https://doi.org/10.30748/soi.2024.179.04

2. Дуднік А., Тищенко А., Яременко Д. Огляд сучасних технічних та програмних рішень для управління БПЛА. *Information Technology And Society*. 2024. № 4 (15). С. 44–50. URL: https://doi.org/10.32689/maup.it.2024.4.8

3. Технологія забезпечення живучості територіально-розподілених інформаційних комп'ютерних систем в єдиному інформаційному просторі / О. Г. Додонов та ін. *Реєстрація, зберігання і обробка даних*. 2024. Т. 26, № 1. С. 121–143. URL: https://doi.org/10.35681/1560-9189.2024.26.1.308659

4. Method of ensuring the survivability of highly mobile computer networks / V. Tkachov et al. *Advanced Information Systems*. 2021. Vol. 5, no. 2. P. 159–165. URL: https://doi.org/10.20998/2522-9052.2021.2.24

5. Залужний В. Ф. Показники ефективності (якості) комплексної системи забезпечення живучості розподілених автоматизованих систем організаційного управління силами та засобами. *Реєстрація, зберігання і обробка даних*. 2025. Т. 27, № 1. С. 42–50. URL: https://doi.org/10.35681/1560-9189.2025.27.1.335693

6. Method to Determine Fault-Tolerant Performance Probability of High-Survivability Computer Network based on Mobile Platform / V. Tkachov et al. *2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T)*, Kharkiv, Ukraine, 5–7 October 2021. 2021. URL: https://doi.org/10.1109/picst54195.2021.9772202

7. Flying Sensor and Edge Network-Based Advanced Air Mobility Systems: Reliability Analysis and Applications for Urban Monitoring / H. Fesenko et al. *Drones*. 2023. Vol. 7, no. 7. P. 409. URL: https://doi.org/10.3390/drones7070409

8. Shaiekhova I. F., Oliinyk Y. O. Approach to developing architecture of a heterogeneous multicomputer task planning system. *Applied Questions of Mathematical Modeling*. 2021. Vol. 4, no. 2.1. URL: https://doi.org/10.32782/kntu2618-0340/2021.4.2.1.29

9. Terenyk D., Kharchenko V. Choosing strategies for deployment and ensuring the reliability of a UAV swarm to support communications in destruction conditions. *Innovative Technologies and Scientific Solutions For Industries*. 2024. No. 3 (29). P. 91–103. URL: https://doi.org/10.30837/2522-9818.2024.3.091

10. Додонов О. Г., Горбачик О. С., Кузнєцова М. Г. Підвищення безпеки критичних інфраструктур засобами автоматизованих систем організаційного управління. *Реєстрація, зберігання і обробка даних*. 2022. Т. 24, № 1. С. 74–81. URL: https://doi.org/10.35681/1560-9189.2022.24.1.262817

11. Додонов О. Г., Ланде Д. В. Мережева модель структурної живучості. *Реєстрація, зберігання і обробка даних*. 2021. Т. 23, № 1. С. 15–22. URL: https://doi.org/10.35681/1560-9189.2021.23.1.235075

12. Афанасьєва А. М., Ткачов В. М. Порівняльний аналіз методів побудови високомобільних комп'ютерних мереж на базі рою БПЛА. *Вісник Херсонського національного технічного університету*. 2024. № 1(88). С. 191–196. URL: https://doi.org/10.35546/kntu2078-4481.2024.1.26

13. Коваленко А. А., Кучук Г. А., Ткачов В. М. метод забезпечення живучості комп'ютерної мережі на основі VPN-тунелювання. *Системи управління, навігації та зв'язку*. Збірник наукових праць. 2021. Т. 1, № 63. С. 90–95. URL: https://doi.org/10.26906/sunz.2021.1.090

14. Reliability Analysis for Dependent Competing Failure System Considering Degradation Rate and Hard Failure Threshold Changes / Y. Niu et al. *2024 3rd International Conference on Innovations and Development of Information Technologies and Robotics (IDITR)*, Hunghom, Hong Kong, 23–25 May 2024. 2024. URL: https://doi.org/10.1109/iditr62018.2024.10554307

15. Searching for heavy-tailed probability distributions for modeling real-world complex networks / T. Chakraborty et al. *IEEE Access*. 2022. P. 1. URL: https://doi.org/10.1109/access.2022.3218631

16. Chen J., Ng M. K., Wang D. Quantizing Heavy-tailed Data in Statistical Estimation: (Near) Minimax Rates, Covariate Quantization, and Uniform Recovery. *IEEE Transactions on Information Theory*. 2023. P. 1. URL: https://doi.org/10.1109/tit.2023.3329240

17. Optimization of drone base station locations and mobile charging drone routing for post-disaster communication / F. Avishan et al. *Computers & Operations Research*. 2025. P. 107206. URL: https://doi.org/10.1016/j.cor.2025.107206

18. Risk-Aware Energy Scheduling for Edge Computing with Microgrid: A Multi-Agent Deep Reinforcement Learning Approach / M. S. Munir et al. *IEEE Transactions on Network and Service Management*. 2021. P. 1. URL: https://doi.org/10.1109/tnsm.2021.3049381

19. R. Kelly Rainer; Brad Prince. Introduction to Information Systems. *Wiley Global Education US*, 2021. 432 c.

20. Schneider C., Hashim M., Valacich J. Information Systems Today: Managing in the Digital World. *Pearson Education*, 2022. 538 c.

21. Mathematical modeling of the availability of the information system for critical use to optimize control of its communication capabilities / O. V. Bisikalo et al. *International Journal of Sensors, Wireless Communications and Control*. 2020. Vol. 10. URL: https://doi.org/10.2174/2210327910999201009163958

22. Xiaofeng Li and Chanle Wu. Building a new-generation learning resources management system based on SLO and SOA technology. *2010 2nd International Conference on Computer Engineering and Technology*, Chengdu, 2010, pp. V3-189-V3-193, doi: 10.1109/ICCET.2010.5485827

23. Decentralized coordination of intelligent system of systems under partial observability / H. Yuan et al. *Advanced Engineering Informatics*. 2025. Vol. 65. P. 103286. URL: https://doi.org/10.1016/j.aei.2025.103286

24. Grefen, Paul W. P. J. Business Information System Architecture. *Independently published*. 2025. 233 c.

References

1. Kompaniyets O. M. (2025). Informatsiyna tekhnolohiya adaptyvnoho upravlinnya royem bezpilotnykh lital′nykh aparativ v antahonistychnomu seredovyshchi [Information technology for adaptive control of uav swarms in antagonistic environments]. *Systemy obrobky informatsiyi*. (4 (179)), pp. 43–48. https://doi.org/10.30748/soi.2024.179.04

2. Dudnik, A., Tyshchenko, O., & Yaremenko, D. (2024). Ohlyad suchasnykh tekhnichnykh ta prohramnykh rishen′ dlya upravlinnya BPLA [Overview of modern technical and software solutions for UAV control]. *Information Technology And Society*, (4 (15)), pp. 44–50. https://doi.org/10.32689/maup.it.2024.4.8

3. Dodonov, O. H., Putyatin, V. H., Dodonov, V. O., Kutsenko, S. A., Hermanyuk, A. P., Izvaryn, I. V., & Kravchuk, K. O. (2024). Tekhnolohiya zabezpechennya zhyvuchosti terytorial′no-rozpodilenykh informatsiynykh komp″yuternykh system v yedynomu informatsiynomu prostori [Technology for ensuring the survivability of geographically distributed information computer systems in a single information space]. *Reyestratsiya, zberihannya i obrobka danykh*, 26(1), 121–143. https://doi.org/10.35681/1560-9189.2024.26.1.308659

4. Tkachov, V., Kovalenko, A., Kuchuk, H., & Ni, I. (2021). Method of ensuring the survivability of highly mobile computer networks. *Advanced Information Systems*, 5(2), 159–165. https://doi.org/10.20998/2522-9052.2021.2.24

5. Zaluzhnyy, V. F. (2025). Pokaznyky efektyvnosti (yakosti) kompleksnoyi systemy zabezpechennya zhyvuchosti rozpodilenykh avtomatyzovanykh system orhanizatsiynoho upravlinnya sylamy ta zasobamy [Indicators of efficiency (quality) of a comprehensive system for ensuring the survivability of distributed automated systems for organizational management of forces and means]. *Reyestratsiya, zberihannya i obrobka danykh*, 27(1), 42–50. https://doi.org/10.35681/1560-9189.2025.27.1.335693

6. Tkachov, V., Hunko, M., Morozova, O., Tetskyi, A., & Nicheporuk, A. (2021). Method to Determine Fault-Tolerant Performance Probability of High-Survivability Computer Network based on Mobile Platform. *2021 IEEE 8th International Conference on Problems of Infocommunications, Science and Technology (PIC S&T)*. IEEE. https://doi.org/10.1109/picst54195.2021.9772202

7. Fesenko, H., Illiashenko, O., Kharchenko, V., Kliushnikov, I., Morozova, O., Sachenko, A., & Skorobohatko, S. (2023). Flying Sensor and Edge Network-Based Advanced Air Mobility Systems: Reliability Analysis and Applications for Urban Monitoring. *Drones*, 7(7), 409. https://doi.org/10.3390/drones7070409

8. Shaiekhova, I. F., & Oliinyk, Y. O. (2021). Approach to Developing Architecture of a Heterogeneous Multicomputer Task Planning System. *Applied Questions of Mathematical Modeling*, 4(2.1). https://doi.org/10.32782/kntu2618-0340/2021.4.2.1.29

9. Terenyk, D., & Kharchenko, V. (2024). Choosing strategies for deployment and ensuring the reliability of a UAV swarm to support communications in destruction conditions. *Innovative Technologies And Scientific Solutions For Industrie* S, (3 (29)), 91–103. https://doi.org/10.30837/2522-9818.2024.3.091

10. Dodonov, O. H., Horbachyk, O. S., & Kuznyetsova, M. H. (2022). Pidvyshchennya bezpeky krytychnykh infrastruktur zasobamy avtomatyzovanykh system orhanizatsiynoho upravlinnya [Improving the security of critical infrastructures through automated organizational management systems]. *Reyestratsiya, zberihannya i obrobka danykh*, 24(1), 74–81. https://doi.org/10.35681/1560-9189.2022.24.1.262817

11. Dodonov, O. H., & Lande, D. V. (2021). Merezheva model′ strukturnoyi zhyvuchosti [Network model of structural survivability]. *Reyestratsiya, zberihannya i obrobka danykh*, 23(1), 15–22. https://doi.org/10.35681/1560-9189.2021.23.1.235075

12. Afanasiyeva, A. M., & Tkachov, V. M. (2024). Porivnyal′nyy analiz metodiv pobudovy vysokomobil′nykh komp″yuternykh merezh na bazi royu BPLA [Comparative analysis of methods for building high-mobility computer networks based on a swarm OF UAVS]. *Visnyk Khersons'koho natsional'noho tekhnichnoho universytetu*, (1(88)), 191–196. https://doi.org/10.35546/kntu2078-4481.2024.1.26

13. Kovalenko, A., Kuchuk, H., & Tkachov, V. (2021). Metod zabezpechennya zhyvuchosti komp″yuternoyi merezhi na osnovi VPN-tunelyuvannya [Method of ensuring the viability of a computer network based on VPN tunnelling]. *Systemy upravlinnya, navihatsiyi ta zv"yazku. Zbirnyk naukovykh prats'*, 1(63), 90–95. https://doi.org/10.26906/sunz.2021.1.090

14. Niu, Y., Wang, X., Wang, B., & Wu, J. (2024). Reliability Analysis for Dependent Competing Failure System Considering Degradation Rate and Hard Failure Threshold Changes. *2024 3rd International Conference on Innovations and Development of Information Technologies and Robotics (IDITR)*. IEEE. https://doi.org/10.1109/iditr62018.2024.10554307

15. Chakraborty, T., Chattopadhyay, S., Das, S., Kumar, U., & Senthilnath, J. (2022). Searching for heavy-tailed probability distributions for modeling real-world complex networks. *IEEE Access*, 1. https://doi.org/10.1109/access.2022.3218631

16. Chen, J., Ng, M. K., & Wang, D. (2023). Quantizing Heavy-tailed Data in Statistical Estimation: (Near) Minimax Rates, Covariate Quantization, and Uniform Recovery. *IEEE Transactions on Information Theory*, 1. https://doi.org/10.1109/tit.2023.3329240

17. Avishan, F., Karasu, M. B., Çap, M., & Yanıkoğlu, İ. (2025). Optimization of drone base station locations and mobile charging drone routing for post-disaster communication. *Computers & Operations Research*, 107206. https://doi.org/10.1016/j.cor.2025.107206

18. Munir, M. S., Abedin, S. F., Tran, N. H., Han, Z., Huh, E.-N., & Hong, C. S. (2021). Risk-Aware Energy Scheduling for Edge Computing with Microgrid: A Multi-Agent Deep Reinforcement Learning Approach. *IEEE Transactions on Network and Service Management*, 1. https://doi.org/10.1109/tnsm.2021.3049381

19. R. Kelly Rainer, Brad Prince (2021). Introduction to Information Systems. *Wiley Global Education US*.

20. Schneider, C., Hashim, M., & Valacich, J. (2022). Information Systems Today: Managing in the Digital World. *Pearson Education*.

21. Bisikalo, O. V., Kovtun, V. V., Kovtun, O. V., & Danylchuk, O. M. (2020). Mathematical modeling of the availability of the information system for critical use to optimize control of its communication capabilities. *International Journal of Sensors, Wireless Communications and Control*, 10. https://doi.org/10.2174/2210327910999201009163958

22. Xiaofeng Li & Chanle Wu. (2010). Building a new-generation learning resources management system based on SLO and SOA technology. *2010 2nd International Conference on Computer Engineering and Technology*. IEEE. https://doi.org/10.1109/iccet.2010.5485827

23. Yuan, H., Ren, B., Chen, T., & Luo, X. (2025). Decentralized coordination of intelligent system of systems under partial observability. *Advanced Engineering Informatics*, 65, 103286. https://doi.org/10.1016/j.aei.2025.103286

24. Grefen, Paul W. P. J. (2025). Business Information System Architecture. *Independently published*.

Дата першого надходження рукопису до видання: 29.09.2025
Дата прийнятого до друку рукопису після рецензування: 23.10.2025
Дата публікації:

УДК 519.21:004.421.2 DOI

М. В. СЕМАНЬКІВ
кандидат технічних наук,
доцент кафедри комп'ютерних наук та інформаційних систем
Карпатський національний університет імені Василя Стефаника
ORCID: 0000-0002-1314-8923

ВДОСКОНАЛЕННЯ ГЕНЕРАТОРА ПСЕВДОВИПАДКОВИХ ЧИСЕЛ НА РЕГІСТРАХ ЗСУВУ ЗВОРОТНОГО ЗВ'ЯЗКУ ТА ДОСЛІДЖЕННЯ ЙОГО ХАРАКТЕРИСТИК

В даній роботі розглянуто застосування генераторів псевдовипадкових чисел для методу Монте-Карло, комп'ютерних симуляціях, 2D і 3D рендерингу. Проаналізовано їх характеристик та відзначено одну з важливіших характеристик генераторів – якість рівномірності розподілу, тобто наскільки рівномірно будуть розподілені згенеровані точки чи напрямки у просторі. Відзначено, що якісний генератор випадкових чисел у 3D-моделюванні та рендерингу є фундаментом того, щоб результати обчислень і зображення були точними, швидко збіжними та візуально чистими. Проведено огляд останніх публікацій на тему покращення характеристик генераторів псевдовипадкових чисел. У результаті аналізу визначено, що низка існуючих рішень полягає в комбінуванні класичних швидких генераторів або додаванні операції для забезпечення рівномірного заповнення 2D і 3D просторів та зменшення похибки інтегрування в методі Монте-Карло. Серед алгоритмічних генераторів для подальшої модифікації обрано алгоритм Xorshift через його швидкість та простоту інтеграції в інформаційну систему. Запропоновано рішення, що базується на вдосконалені алгоритму генерації псевдовипадкових чисел Xorshift за рахунок бітного реверсування, що дозволило покращити характеристики генерованої послідовності. Проведено статистичні тести для запропонованого генератора та здійснено порівняння його результатів з початковим базовим генератором Xorshift. Результати статистичного аналізу показали відсутність значущих відхилень від ідеального рівномірного розподілу. Доведено, що реверсування бітів може суттєво покращувати 2D-проєкції й зменшити надмірну щільність точок чи навпаки прогалини. Зауважено кращу рівномірність покриття площини світла та швидшу збіжність Монте-Карло інтеграла для генератора з реверсування бітів та відзначено його перспективність для задач рендерингу світла і тіней. Зроблено висновок, що реверсування бітів є простою, але ефективною технікою підвищення якості генератора псевдовипадкових послідовностей. Запропонований генератор має практичне значення для задач Монте-Карло, комп'ютерного моделювання та 3D-візуалізації, де вимоги до рівномірності вибірки у багатовимірному просторі є особливо критичними.

Ключові слова: *генератор псевдовипадкових чисел, регістри зсуву, статистичний аналіз.*

M. V. SEMANKIV
Candidate of Technical Sciences,
Associate Professor at the Department of Computer Science
Vasyl Stefanyk Carpathian National University
ORCID: 0000-0002-1314-8923

IMPROVEMENT OF A PSEUDORANDOM NUMBER GENERATOR BASED ON A LINEAR FEEDBACK SHIFT REGISTER AND INVESTIGATION OF ITS CHARACTERISTICS

© Семанків М. В., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

This article examines the use of pseudo random number generators in Monte Carlo methods and in computer simulations, in 2D and 3D rendering. Their characteristics are analyzed, with particular attention given to one of the most important properties of such generators – the quality of distribution uniformity, i.e., how evenly the generated points or directions are distributed in space. It is noted that a high-quality random number generator in 3D modeling and rendering is essential for achieving accurate results, fast convergence, and visually clean images. A review of recent publications on improving the properties of pseudo random number generators is presented. The analysis shows that many existing solutions involve combining classical fast generators or adding extra operations to ensure uniform coverage of 2D and 3D spaces and to reduce Monte Carlo integration error. Among algorithmic generators, Xorshift was selected for further modification due to its speed and ease of integration into an information system. A solution is proposed that builds upon the Xorshift pseudo random number generation algorithm by incorporating bit reversal, which improved the statistical properties of the generated sequence. Statistical tests were performed on the proposed generator, and its results were compared with those of the baseline Xorshift generator. The statistical analysis showed no significant deviations from an ideal uniform distribution. It was demonstrated that bit reversal can substantially improve 2D projections and reduce both excessive clustering of points and gaps. Improved uniformity in light-plane sampling and faster convergence of the Monte

© Семанків М. В., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

Carlo integral were observed for the bit-reversed generator, highlighting its potential for light and shadow rendering tasks. It is concluded that bit reversal is a simple yet effective technique for improving the quality of pseudo random sequences. This has practical importance for Monte Carlo methods, computer simulation, and 3D visualization, where uniformity of sampling in multidimensional space is especially critical.

Key words: *pseudo random number generator, linear feedback shift register, statistical analysis.*

Постановка проблеми

Генератори псевдовипадкових чисел (ГПЧ) застосовуються в дуже багатьох сферах: для методу Монте-Карло, в машинному навчання, в комп'ютерній графіці та іграх, для тестування програм і систем та т.д. У рендерингу, комп'ютерних симуляціях і будь-яких методах Монте-Карло якісні ГПЧ мають критично важливе значення, оскільки саме вони визначають, наскільки рівномірно будуть розподілені точки у просторі. Якщо генератор «слабкий», то випадкові точки можуть утворювати скупчення й залишати прогалини, що візуально проявляється у вигляді шуму, смуг або плям у зображенні. Натомість надійні генератори забезпечують рівномірне покриття простору, відсутність кореляцій і закономірних візерунків, швидку генерацію мільйонів чисел [1].

Генератори випадкових і псевдовипадкових чисел мають низку проблем, серед яких ключовими є обмежена якість випадковості (кореляції між послідовними значеннями, нерівномірність розподілу, низька ентропія), скінченний період, що призводить до повторюваності послідовностей у довгих симуляціях, а також передбачуваність, що є критичним для криптографічних застосувань. Додатково виникають питання продуктивності, адже якісні генератори можуть бути повільними, та питання відтворюваності результатів, що важливо у наукових розрахунках. Також можливі програмні помилки у реалізації або фізичні зсуви в апаратних генераторах, що викликає статистичні відхилення та спотворює результати моделювання чи рендерингу [1, 2].

Аналіз останніх досліджень і публікацій

Лінійні регістри зі зворотним зсувом (LFSR) мають кілька характерних проблем: вони генерують повністю детерміновані послідовності з обмеженим періодом, що легко прогнозуються після спостереження невеликої кількості вихідних бітів, тому не підходять для криптографії [2]. Їхня лінійність призводить до слабкої статистичної випадковості – у спектрі з'являються помітні кореляції, відсутні певні бітові комбінації, а розподіл може відхилятися від рівномірного [3]. LFSR також чутливі до початкового стану: за нульової ініціалізації регістр «зависає» у нулях. Хоч вони дуже швидкі й прості в апаратній реалізації, але у задачах моделювання, рендерингу або криптографії їх часто комбінують або нелінійно перетворюють, щоб покращити статистичні властивості й приховати передбачуваність [4].

Для покращення статистичних властивостей і рівномірності розподілу у генераторів на основі лінійного зворотного зсувного регістра здійснюють комбінування декількох генераторів, що дозволяє збільшити періодичність та зменшує лінійність [3]. Для злагодження нерівномірності в генераторах даного типу використовують перестановку вивідних бітів [5].

Формулювання мети дослідження

Метою роботи є вдосконалення алгоритму генерації псевдовипадкових чисел на регістрах зсуву зворотного зв'язку (Xorshift) за рахунок введення дії реверсу на кожну 32/64-бітну вибірку та дослідження статистичних характеристик запропонованого генератора псевдовипадкових чисел.

Викладення основного матеріалу дослідження

Розглянемо популярний ГПЧ, що є базовим генератором для тестування методів інтеграції у Монте-Карло рендерингу – генератор Xorshift, що відкритий Джорджем Марсальєю [2]. Він є підмножиною генераторів лінійного регістра зсуву зворотного зв'язку. Метод Xorshift генерує випадкові числа за допомогою багаторазового використання побітових операторів виключне-або (XOR, $\oplus$) та бітових зсувів ($\ll$ для лівого та $\gg$ для правого). Це робить його дуже швидким в сучасних комп'ютерних архітектурах. Він вимагає дуже малого розміру коду та стану, але він не проходить всі статистичні тести без вдосконалення, одним зі способів виправити це – поєднання їх з нелінійними функціями або включити додаткову операцію [4].

Розглянемо детальніше Xorshift64. Нехай x_n – 64-бітне ціле число, що згенеровано на n кроці. Наступне число формується на основі рекуретної формули:

$$x_{n+1} = T(x_n);$$

$$T(x) = (((x \oplus (x \ll a)) \bigoplus ((x \bigoplus (x \ll a)) \gg b))\bigoplus;$$

$$\bigoplus(((x \bigoplus (x \ll a)) \bigoplus ((x \bigoplus (x \ll a)) \gg b)) \ll c)) \bmod 2^{64},$$

де $\oplus$ – побітове *XOR*, $\ll k$ – циклічний зсув на вліво k бітів, $\gg k$ – зсув вправо на k бітів, $a, b, c \in N$ – три параметри-зсуви [2].

Усі операції виконуються за модулем 2^{64}.

Опишемо запропоновану для генератора операцію реверсування порядку бітів. Нехай $x \in \{0, 1, \ldots, 2^{b-1}\}$ – ціле число, що подано у двійковій формі:

$$x = \sum_{i=0}^{b-1} x_i 2^i, \quad x_i O\{0,1\}, \quad x_0 - \textit{молодший біт}, \quad x_{b-1} - \textit{старший}.$$

Тоді операція реверсу має вигляд:

$$rev\, x_b(x) = \sum_{i=0}^{b-1} x_i 2^{b-1-i},$$

тобто індекс біта змінюється з i на $b - 1 - i$.

Операція перестановки бітів не змінює періоду, і в бінарному сенсі вона залишається лінійною. В числовому поданні вона змінює розподіл старших/молодших бітів. Запропонований генератор Xorshift з реверсуванням порядку бітів в подальшому буде згадуватись як реверсивний Xorshift або Xorshift з реверсуванням бітів.

Одним із методів аналізу генераторів випадкових чисел є створення візуалізації згенерованих чисел [6]. Побудовано точкові діаграми з нормалізацією точок у діапазоні [0,1] для генератора Xorshift64 та запропонованого реверсивного генератора (рис. 1).

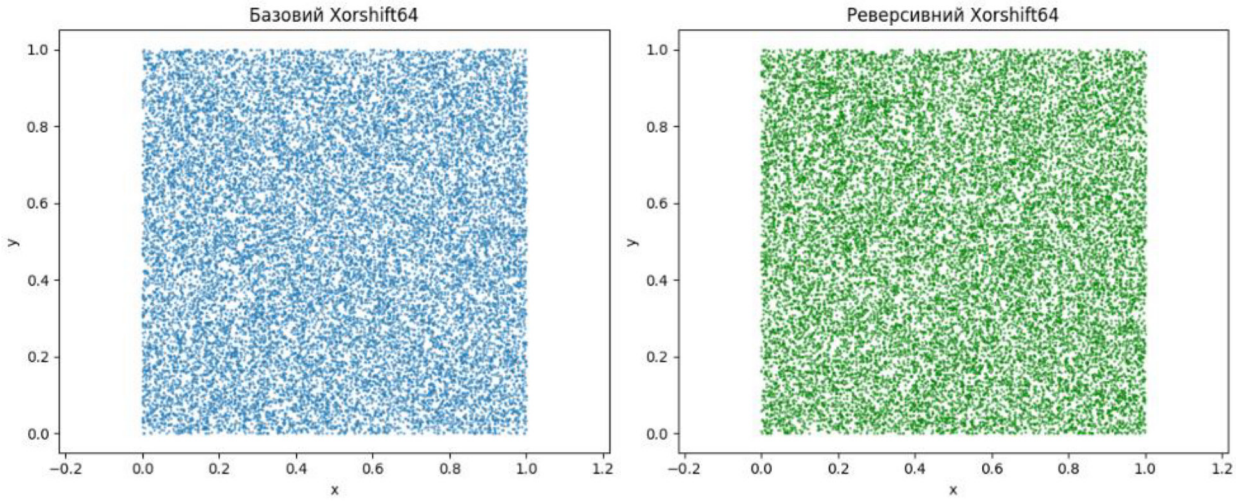


Рис. 1. Візуалізація розподілу згенерованих точок на площині

Статистичний аналіз

Проведено статистичне дослідження [6] послідовностей згенерованих базовим генератором Xorshift та реверсивним Xorshift, проведено наступні тести та отримано результати:

1. Осьові тести для перевірки розподілу кожної координати окремо – X, Y та Z. Кожна координата є 1D-вибіркою, і для неї використано тест Колмогорова-Смірнова і χ^2-критерій.

Таблиця1

Результати тестів Колмогорова-Смірнова і χ^2-критерію

Генератори	Осі	KS-тест	χ^2-критерій
Генератор Xorshift	X	stat = 0.0040, p = 0.9032	stat = 53.28 (df = 63), p = 0.8036
	Y	stat = 0.0076, p = 0.1920	stat = 68.25 (df = 63), p = 0.3035
	Z	stat = 0.0042, p = 0.8749	stat = 50.35 (df = 63), p = 0.8754
Реверсивний Xorshift	X	stat = 0.0093, p = 0.0618	stat = 67.22 (df = 63), p = 0.3347
	Y	stat = 0.0053, p = 0.6189	stat = 52.96 (df = 63), p = 0.8124
	Z	stat = 0.0073, p = 0.2304	stat = 61.55 (df = 63), p = 0.5282

Обидва генератори (Xorshift64 і з реверсуванням бітів) демонструють розподіл, що близький до рівномірного за всіма осями. Проте операція реверсування бітів дала невелике покращення статистичних показників. Вищі p-значення у тесті Колмогорова-Смірнова і χ^2-тесті вказують на зменшення прихованих кореляцій і більш «гладке» заповнення простору. У практичних задачах, таких як моделювання, Монте-Карло, 3D-рендеринг, такий реверсивний генератор може бути корисним для покращення якості випадкової вибірки.

2. Тест Пірсона парних кореляцій між осями. Його використовують, щоб перевірити, чи не виникають систематичні залежності між координатами випадково згенерованих точок. У випадку справді рівномірного тривимірного розподілу очікується, що значення X, Y і Z мають бути статистично незалежними, тобто їхня кореляція

повинна бути близькою до нуля. Даний тест допомагає виявити приховані залежності, які не завжди видно на гістограмах або χ^2-тестах.

Таблиця2

Тест Пірсона парних кореляцій між осями

Генератори	Площини	Результати тесту
Генератор Xorshift	*XY*	0.004512102657454674
	XZ	0.0025055279677412643
	YZ	–0.004825255314985926
Реверсивний Xorshift	*XY*	–0.01779539191410323
	XZ	0.013988626760637064
	YZ	–0.002869986540511929

Аналіз парних кореляцій між координатами *XY*, *ZX*, *Y*, *Z* показав, що для базового генератора Xorshift значення залишаються майже нульовими ($|r| < 0.005$), що свідчить про практичну незалежність координат і відсутність осьових залежностей, тобто точки в просторі розподілені рівномірно й ізотропно. Натомість у випадку з реверсуванням бітів кореляції зростають за модулем ($|r| \approx 0.014 – 0.018$), що вказує на появу слабких, але помітних взаємозалежностей між координатами, зокрема негативної кореляції в площині *XY*. Хоча такі величини все ще невеликі й не призводять до різкого порушення рівномірності, вони можуть формувати приховані діагональні структури в хмарі точок, що потенційно знижує статистичну ефективність при застосуванні у прикладних задачах, де критичною є саме незалежність координат і відсутність латентних закономірностей.

3. Покриття на сітці (2D/3D χ^2). Цей тест потрібен для того, щоб перевірити наскільки рівномірно точки покривають простір не лише вздовж однієї осі, а й у взаємодії між осями.

Таблиця3

Покриття на сітці (2D/3D χ^2)

Сітка	Xorshift	Реверсивний метод
2D χ^2	*XY*: $\chi^2 = 242.10$, $p = 0.7094$	*XY*: $\chi^2 = 255.87$, $p = 0.4729$
	XZ: $\chi^2 = 260.99$, $p = 0.3849$	*XZ*: $\chi^2 = 272.31$, $p = 0.2180$
	YZ: $\chi^2 = 297.96$, $p = 0.0333$	*YZ*: $\chi^2 = 233.55$, $p = 0.8285$
3D χ^2	$\chi^2 = 485.27$, $p = 0.7875$	$\chi^2 = 550.04$, $p = 0.1127$

Отримані результати χ^2-тестів свідчать, що рівномірність покриття простору суттєво залежить від способу генерації. Для генератора Xorshift у проекції *YZ* зафіксовано значне відхилення від рівномірного розподілу ($p \approx 0.033$), що вказує на наявність прихованих кореляцій чи кластеризації. Використання реверсування бітів помітно зменшило цю проблему ($p \approx 0.8285$), забезпечивши значно більш рівномірне покриття в цій площині. Водночас у повному 3D просторі реверсивний Xorshift показав гірший результат ($p \approx 0.11$ проти 0.79), що означає надлишкову регулярність і менш природну імітацію випадковості. Таким чином, реверсування бітів може суттєво покращувати 2D-проєкції й зменшувати локальні артефакти, проте у задачах, де критична саме глобальна тривимірна рівномірність, базовий Xorshift або низькодисперсні послідовності можуть залишатися кращими.

4. Тест найближчого сусіда потрібен для оцінки того, наскільки «щільно» і рівномірно розташовані точки в просторі. Якщо точки дійсно рівномірні, то відстані до найближчих сусідів повинні мати відносно стабільний розподіл – без надто великої кількості дуже маленьких відстаней (кластерів) чи надто великих прогалин.

Таблиця4

Тест найближчого сусіда

Генератори	Середнє значення	Стандартне відхилення
Генератор Xorshift	mean = 21.1720	std = 7.8514
Реверсивний Xorshift	mean = 21.1019	std = 7.9206

Аналіз результатів тесту найближчого сусіда для генераторів Xorshift звичайного та з реверсування бітів показав, що обидві послідовності формують просторові конфігурації точок зі статистично подібними характеристиками локальної щільності. Середні значення відстаней між найближчими точками (близько 21.1–21.2) практично збігаються, що свідчить про аналогічний масштаб заповнення тривимірного простору для обох генераторів. Стандартні відхилення (приблизно 7.9) також майже однакові, що вказує на відсутність значущих відмінностей у варіативності міжпунктових відстаней та, відповідно, на схожу тенденцію до рівномірного розподілу без явних кластерів чи «порожніх зон».

5. 2D-/3D-KS альтернативи (енергетична відстань EnergyDist, максимальна середня розбіжність MMD) є тестами для оцінки рівномірності заповнення простору вищих вимірів. Вони допомагають зрозуміти, чи дійсно генератор підходить для задач, де важливе рівномірне покриття.

Таблиця 5

2D-/3D-KS альтернативи

Генератори	Тест	2D	3D
Генератор Xorshift	EnergyDist	0.00277	0.00374
	MMD	0.00009	0.00009
Реверсивний Xorshift	EnergyDist	0.00221	0.00453
	MMD	0.00010	0.00011

Отримані результати показують, що у площинних проекціях (2D) енергетична відстань для вибірки згенерованої реверсивним Xorshift є меншою, ніж для базованої, що свідчить про більш рівномірне покриття й зменшення локальних кластерів; однак у повному 3D просторі реверсування бітів демонструє гірші показники (вища EnergyDist), тобто він створює приховану регулярність, яка знижує глобальну випадковість. Значення MMD для обох методів залишаються дуже малими й близькими, що означає загалом добру апроксимацію рівномірного розподілу. На основі даного тесту можна зробити висновок, що реверсивний метод корисний для задач, де важлива якість 2D-проєкцій (наприклад, рендеринг світла і тіней), тоді як звичайний Xorshift більш надійний у тривимірних симуляціях. Даний тест підтвердив результати тесту покриття на сітці (2D/3Dχ^2).

Автокореляція (ACF) по осі перевіряє наскільки послідовні значення генератора випадкових чисел залежать від попередніх.

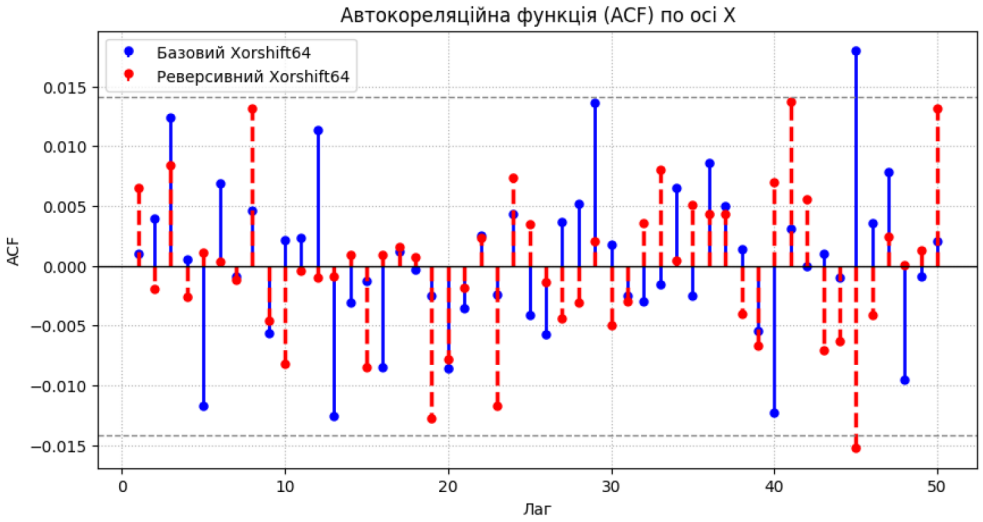


Рис. 2. Результати тесту на автокореляцію

На основі рис. 2 можна зробити висновок, що в обох випадках значення автокореляцій залишаються дуже близькими до нуля і майже завжди лежать у межах довірчого інтервалу 95 %. Це свідчить про відсутність суттєвої лінійної залежності між сусідніми елементами послідовності. У реверсивного генератора (пунктирна лінія) може бути трохи більша варіативність, ніж у базованого (суцільна), але відхилення все одно не перевищують статистично значущих меж. Таким чином, реверсування бітів не погіршило якість випадковості, ані не додало систематичних залежностей. Отже, обидва варіанти, як базовий Xorshift, так і реверсивний, демонструють відсутність автокореляцій на низьких лагах, що є ознакою хорошої властивості випадкових послідовностей у задачах Монте-Карло та моделювання.

Використання генераторів для рендерингу

У сучасних системах рендеринг часто використовує Монте-Карло інтеграцію і тому він потребує якісні генератори псевдовипадкових чисел, бо саме вони визначають, наскільки рівномірно в просторі будуть вибиратися напрямки променів, точки вибірки на поверхнях та у середовищі. Вони є фундаментом для того, щоб результати обчислень і зображення були точними, швидко збіжними та візуально чистими. Тому проведено ряд тестів для визначення ефективності запропонованого генератора для практичного застосування для рендерингу [7].

1) перевірка якості освітлення й тіней у графічних рушіях чи рендерах показує наскільки правильно система відтворює м'які тіні від протяжних джерел світла. RMSE як міра точності визначить наскільки якість тіні, створеної апроксимацією, відрізняється від еталонного фізично коректного рендерингу.

Для базового генератора RMSE становить 0.006878 (при spp=16), для реверсивного 0.005971 (spp = 16). Ці результати означають, що при використанні 16 променів на піксель (spp) метод із реверсування бітів дав меншу похибку середньоквадратичного відхилення (RMSE = 0.005971), ніж базовий Xorshift (RMSE = 0.006878). Тобто згенероване зображення за допомогою генератора з реверсом бітів виявилося трохи точнішим у відтворенні еталонного результату. Така перевага свідчить, що реверс бітів може краще «розкидати» вибірки по просторі, зменшуючи локальні кластери й прогалини, що критично для задач рендерингу світла та тіней, де важлива рівномірність покриття інтегральної області.

Тест прямого освітлення з використанням MIS (Multiple Importance Sampling, багатокритеріальне вибіркове інтегрування) перевіряє, наскільки правильно і ефективно алгоритм рендерингу відтворює пряме освітлення. Він показує наскільки добре рендер знижує шум, чи правильно балансує ваги між різними P*Df*, чи збігається результат з еталонним рендерингом.

MIS direct lighting RMSE для базового генератора становить 0.002684 (при spp = 16), для реверсивного MIS direct lighting RMSE дорівнює 0.002714 (spp = 16). Отримані результати показують, що при використанні MIS для прямого освітлення різниця між генератором Xorshift64 та його варіантом із реверсією бітів практично відсутня: значення RMSE для обох випадків дуже близькі, що відповідає лише ≈1 % відносної різниці і може бути пояснене випадковими статистичними флуктуаціями. Це означає, що реверсування бітів у даній задачі не дає ані помітної переваги, ані шкоди – точність обчислення в першу чергу визначається самим методом MIS і геометрією сцени, а не деталями генератора псевдовипадкових чисел.

Пряме освітлення від квадратного світла (без видимості), обчислення освітлення однієї точки від квадратного світильника.

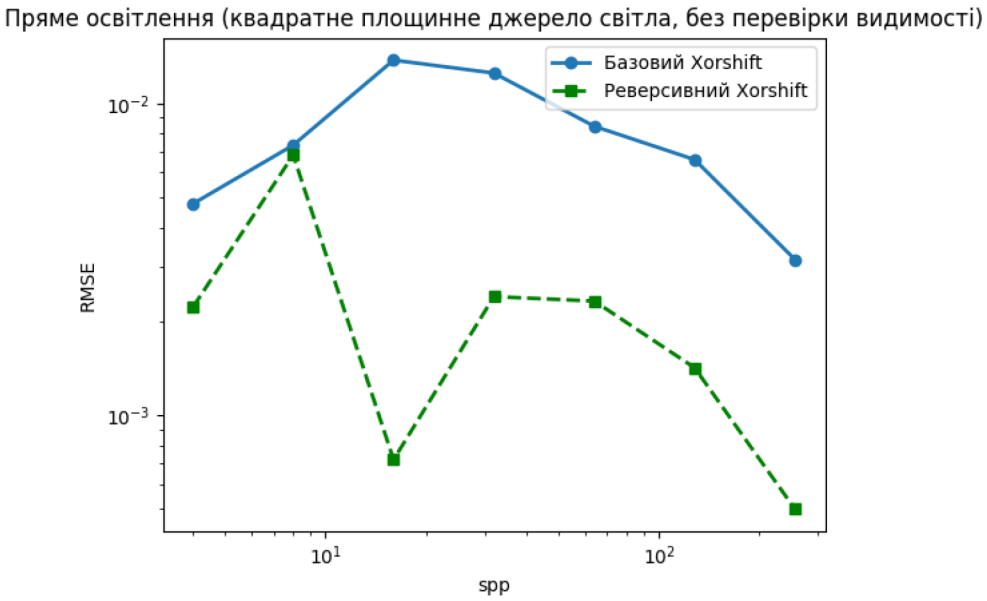


Рис. 3. Результати тесту Direct lighting

Крива показує, що для задачі прямого освітлення від квадратного джерела без видимості послідовність із реверсуванням бітів (пунктир) майже на всіх spp дає менший RMSE, ніж базовий Xorshift (суцільна), інколи істотно – отже, за однакової кількості семплів вона збігається швидше (краще покриває квадрат світла). Реверсивний генератор зменшує кластеризацію/прогалини у (u, v) і знижує дисперсію оцінки. Для рендерингу цей результат означає, що реверсування бітів дає кращі вибірки для інтеграції освітлення, ніж базовий Xorshift, коли мова йде про прості інтеграли на низькорозмірних областях. Отже, для прямого освітлення від площинних світильників реверсування зменшило шум і пришвидшило збіжність зображення при тому ж числі семплів. У випадку малих spp це особливо помітно – картинка буде виглядати більш «чистою» при тій самій кількості променів.

Висновки

Отримані результати статистичного аналізу дають підстави стверджувати, що як базовий генератор Xorshift64, так і модифікований варіант із реверсування бітів забезпечують рівномірність розподілу. Зокрема, значення p-value у тесті Колмогорова–Смирнова для всіх координатних осей перевищують критичні пороги, що свідчить про відсутність статистично значущих відхилень від ідеального рівномірного розподілу. χ^2-статистики не перевищують ступені свободи, а p-value в більшості випадків розташовані у проміжку 0.5–0.75, що вказує на високу якість апроксимації. Слід відзначити, що реверсування бітів у згенерованих послідовностях призвела до незначного

покращення статистичних характеристик: середні χ^2-статистики зменшились, а *p*-value тестів зросли, що означає більш рівномірне та «гладке» заповнення простору, зменшення латентних кореляцій між координатами та посилення імітаційних властивостей генератора.

Реверсивний метод має переваги в площинних 2D-проекціях, де він зменшує локальні аномалії, роблячи вибірку більш «рівномірною» та згладженою. Проте у повному 3D просторі він може створювати певну приховану регулярність, що трохи погіршує рівномірність розподілу. MMD показав, що в цілому обидва генератори досить близькі до рівномірного розподілу, але реверсування краще працює для 2D-випадків (наприклад, рендеринг тіней чи текстур), тоді як базовий Xorshift більш стабільний у 3D.

Отримані результати для тесту прямого освітлення з площинним джерелом показують, що використання генератора з реверсування бітів замість стандартного генератора Xorshift дозволяє суттєво зменшити похибку при однаковій кількості семплів на піксель, особливо у випадку малих spp, коли шум у зображенні найбільш помітний. Це свідчить про кращу рівномірність покриття площини світла та, відповідно, швидшу збіжність Монте-Карло інтеграла, що робить метод реверсування бітів перспективним для задач рендерингу світла і тіней, зокрема для вибірки положень на джерелах світла або у просторі параметрів BR*Df*.

Таким чином, можна зробити висновок, що реверсування бітів є простою, але ефективною технікою підвищення якості псевдовипадкових послідовностей, що може мати практичне значення для задач Монте-Карло, комп'ютерного моделювання та 3D-візуалізації, де вимоги до рівномірності вибірки у багатовимірному просторі є особливо критичними.

Список використаної літератури

1. Pharr M., Jakob W., Humphreys G. Physically Based Rendering: From Theory to Implementation. 3rd ed. San Francisco: Morgan Kaufmann, 2016. 1266 с.

2. Xorshift RNGs. ReseachGate. URL: https://www.researchgate.net/publication/5142825_Xorshift_RNGs (дата звернення: 12.09.2025).

3. Pseudo Random Number Generation Using Linear Feedback Shift Registers. Analog devices. URL: https://www.analog.com/en/resources/design-notes/random-number-generation-using-lfsr.html?utm_source=chatgpt.com (дата звернення: 12.09.2025).

4. Avarez R. Improving the Statistical Qualities of Pseudo Random Number Generators // Symmetry. 2022. Т. 14, № 2. С. 269. DOI: 10.3390/sym14020269

5. Condo C., Gross W.J. Pseudo-random Gaussian distribution through optimised LFSR permutations. 2012 IEEE Workshop on Signal Processing Systems. IEEE, 2012. С. 69–74. DOI: 10.1109/SiPS.2012.57

6. Knuth D.E. The Art of Computer Programming. Vol. 2: Seminumerical Algorithms. – Addison-Wesley, 1997.

7. Askar A., Shukirgaliyev B., Lukac N., Abdikamalov E. Evaluation of Pseudo-Random Number Generation on GPU Cards // Computers. 2021. Vol. 10, No. 12. P. 142. DOI: 10.3390/computers10120142

References

1. Pharr, M., Jakob, W., & Humphreys, G. (2016). *Physically based rendering: From theory to implementation*(3rd ed.). San Francisco, CA: Morgan Kaufmann.

2. Xorshift RNGs. (2003). *ResearchGate*. Retrieved from https://www.researchgate.net/publication/5142825_Xorshift_RNGs (accessed 12 September 2025).

3. Analog Devices. (n.d.). *Pseudo random number generation using linear feedback shift registers*. Retrieved from https://www.analog.com/en/resources/design-notes/random-number-generation-using-lfsr.html (accessed 12 September 2025).

4. Alvarez, R. (2022). Improving the statistical qualities of pseudo random number generators. *Symmetry, 14*(2), 269. https://doi.org/10.3390/sym14020269

5. Condo, C., & Gross, W. J. (2012). Pseudo-random Gaussian distribution through optimised LFSR permutations. In *2012 IEEE Workshop on Signal Processing Systems* (pp. 69–74). IEEE. https://doi.org/10.1109/SiPS.2012.57

6. Knuth, D. E. (1997). *The art of computer programming. Vol. 2: Seminumerical algorithms*. Addison-Wesley.

7. Askar, A., Shukirgaliyev, B., Lukac, N., & Abdikamalov, E. (2021). Evaluation of pseudo-random number generation on GPU cards. *Computers, 10*(12), 142. https://doi.org/10.3390/computers10120142

Дата першого надходження рукопису до видання: 18.09.2025
Дата прийнятого до друку рукопису після рецензування: 15.10.2025
Дата публікації:

UDC 004.8:004.738.5 DOI

O. I. SYPIAHIN
Full Stack Engineer
floLive
ORCID: 0009-0008-7565-3221

M. A. YUZHAKOV
Senior Software Engineer
Plaid
ORCID: 0009-0008-1307-4275

R. I. IBRAHIMOV
Senior Software Developer
VoiceLove
ORCID: 0009-0002-7612-3341

ARCHITECTURAL SOLUTIONS FOR SCALABLE WEB APPLICATIONS CONSIDERING PERFORMANCE AND LATENCY

The rapid growth of digital services and the increasing demands of globally distributed users place scalability, performance, and latency optimization at the center of web application design. Ensuring high performance under growing workloads is critical for business process stability, user satisfaction, and system efficiency.

The goal of the study is to compare four architectural paradigms of web applications – monolithic, microservices, serverless, and edge computing – under workloads simulating realistic e-commerce scenarios, including concurrent transactions, product searches, and payment requests, to assess their behavior in terms of scalability, performance, and latency.

As a result, the study revealed clear differences in system behavior across the four paradigms. Monolithic systems, while predictable and simple to operate, quickly became saturated under load, with response times exceeding 900 ms at 1,500 concurrent users and reaching over 3 seconds at peak capacity. Microservices extended the range of scalability, maintaining stable performance up to 3,000–3,500 users with average delays of 420–600 ms; however, at higher concurrency, orchestration overhead caused sharp increases in latency. Serverless platforms demonstrated the highest elasticity, scaling almost instantly to 6,200 requests per second and effectively handling burst traffic, though latency spikes up to 1,000 ms occurred due to cold starts and execution variability. Edge deployments achieved the lowest latency overall, remaining below 300 ms even at 5,000 users, and providing peak throughput of 5,400 requests per second, albeit requiring significant provisioning and introducing greater operational complexity. The findings indicate that no single architectural paradigm fully satisfies both scalability and latency requirements across diverse workload profiles. Hybrid strategies that combine microservices for core stability, serverless for unpredictable bursts, and edge computing for latency-sensitive operations were found to deliver the most balanced outcomes, ensuring responsiveness, stability, and cost-aware scalability.

Conclusions: flexible architectural approaches enable high performance, stability, and scalability of web applications while optimizing latency and operational costs. The findings emphasize the importance of architectural adaptability for building resilient and efficient web applications in modern digital ecosystems.

Key words: *scalable web applications, microservices, serverless, edge computing, Kubernetes, latency, performance optimization, CI/CD, cloud architecture.*

О. І. СИПЯГІН
інженер повного стеку
floLive
ORCID: 0009-0008-7565-3221

М. А. ЮЖАКОВ
старший інженер програмного забезпечення
Plaid
ORCID: 0009-0008-1307-4275

Р. І. ІБРАГІМОВ
старший розробник програмного забезпечення
VoiceLove
ORCID: 0009-0002-7612-3341

© Sypiahin O. I., Yuzhakov M. A., Ibrahimov R. I., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

АРХІТЕКТУРНІ РІШЕННЯ ДЛЯ МАСШТАБОВАНИХ ВЕБЗАСТОСУНКІВ З УРАХУВАННЯМ ПРОДУКТИВНОСТІ ТА ЗАТРИМОК

Швидке зростання цифрових сервісів та зростаючі вимоги глобально розподілених користувачів ставлять у центр уваги вебзастосунків питання масштабованості, продуктивності та оптимізації затримок. Забезпечення високої продуктивності за зростаючих навантажень є критично важливим для стабільності бізнес-процесів, задоволеності користувачів та ефективності системи.

Ціллю дослідження є порівняльна оцінка чотирьох архітектурних парадигм вебзастосунків – монолітної, мікросервісної, серверлес та edge-комп'ютингу – за умов навантажень, що імітують реалістичні сценарії електронної комерції, включно з одночасними транзакціями, пошуком товарів та платіжними запитами, для визначення їхньої поведінки щодо масштабованості, продуктивності та затримок.

Як результат, дослідження показало значні відмінності у поведінці систем за різних архітектур. Монолітні системи, хоча й передбачувані та прості в експлуатації, швидко насичувалися під навантаженням: час відповіді перевищував 900 мс при 1 500 одночасних користувачах та досягав понад 3 с на піковихнавантаженнях. Мікросервіси розширювали діапазон масштабованості, забезпечуючи стабільну продуктивність до 3 000–3 500 користувачів із середніми затримками 420–600 мс, проте при високій одночасності накладні витрати на оркестрацію призводили до різкого зростання затримок. Серверлес-платформи показали найвищу еластичність, майже миттєво масштабуясь до 6 200 запитів на секунду та ефективно обробляючи пікові навантаження, хоча спостерігалися пікові затримки до 1 000 мс через холодні запуски та змінність виконання. Edge-розгортання забезпечували найменші затримки загалом, залишаючись нижче 300 мс навіть при 5 000 користувачах, і досягали пікової пропускної здатності 5 400 запитів на секунду, хоча вимагали значного забезпечення ресурсів та створювали більшу операційну складність. Результати свідчать, що жодна окрема архітектурна парадигма не повністю задовольняє вимоги масштабованості та затримки для різних профілів навантажень. Гібридні стратегії, що поєднують мікросервіси для стабільності ядра, серверлес для непередбачуваних піків та edge-комп'ютинг для критично чутливих до затримок операцій, забезпечують найбільш збалансовані результати, гарантують відгук системи, стабільність і економічно обґрунтовану масштабованість.

Висновки: гнучкі архітектурні підходи забезпечують високу продуктивність, стабільність та масштабованість вебзастосунків, оптимізуючи затримки та операційні витрати. Отримані результати підкреслюють важливість адаптивності архітектури для створення стійких та ефективних вебзастосунків у сучасних цифрових екосистемах.

Ключові слова: *масштабовані вебзастосунки, мікросервіси, безсерверні технології, edge-комп'ютинг, Kubernetes, затримка, оптимізація продуктивності, CI/CD, хмарна архітектура.*

Formulation of the problem

Modern web applications are expected to handle millions of requests, adapt to unpredictable workloads, and deliver responses within milliseconds, regardless of where the user is located. These demands place the architecture at the center of performance and scalability decisions. What once worked for small, monolithic systems is no longer enough in environments shaped by continuous deployment, global traffic distribution, and real-time interaction. To address these challenges, developers increasingly turn to microservices managed by container orchestration platforms, serverless computing for unpredictable spikes, and edge deployments that physically shorten the path between user and service. Each of these approaches offers tangible benefits but also creates new layers of complexity. Microservices can fragment monitoring and raise communication costs, serverless models complicate stateful logic and cost forecasting, and edge computing adds operational overhead in exchange for lower latency. By comparing different architectural solutions under varied workload patterns, the study highlights not only where each model excels but also where its limitations become critical. In doing so, it points toward hybrid strategies that balance scalability with predictable performance, helping teams design applications that remain responsive even as systems grow in size and complexity.

Analysis of the latest research and publications

The architecture of scalable web applications has become a central focus in modern software engineering due to the rapid expansion of cloud computing, the global distribution of users, and the performance constraints imposed by latency-sensitive services. Traditional monolithic designs are increasingly being reconsidered as they struggle to meet the requirements of dynamic workloads and continuous delivery pipelines [1]. Recent research emphasizes the shift toward microservices, serverless computing, and hybrid architectures that can adapt to both predictable and burst traffic while sustaining stable throughput and low response times [2].

Studies have explored the potential of cloud-native approaches to reduce latency and improve elasticity. For example, Atoll [3] and Nightcore [4] demonstrate serverless platforms optimized for interactive workloads, highlighting the importance of efficient resource allocation and low-overhead communication in latency-sensitive environments. Similarly, applied research points to best practices in backend architecture that balance scalability with predictable performance in real-world deployment scenarios [5].

Comparative evaluations have also been conducted between monolithic and microservice models, showing that while microservices generally improve scalability and modularity, they introduce nontrivial communication costs and complexity

in monitoring and orchestration [6]. At the same time, distributed and edge-based approaches are increasingly recognized as promising strategies for lowering user-perceived latency, but they raise challenges of deployment complexity and configuration overhead [2].

International publications underline the necessity of combining architectural choices with rigorous performance management, observability pipelines, and workload-aware optimizations [2,5]. This ensures not only that systems remain responsive under stress but also that operational costs and stability remain within acceptable limits. As a result, determining the optimal mix of architectural paradigms–monolithic, microservices, serverless, and edge computing–remains an active area of research, with significant practical implications for scalable web application design.

A relevant topic of this study is the search for architectural solutions that ensure both scalability and low-latency performance in modern web applications. Although cloud-native technologies, microservice patterns, and serverless platforms have achieved significant adoption worldwide, many development teams still lack a consistent framework for selecting and combining these approaches within CI/CD environments. Addressing this challenge requires not only adapting proven international practices but also developing solutions tailored to the specifics of local infrastructures, workload profiles, and organizational constraints.

Formulation of the purpose of the research

The aim of this study is architectural models and design strategies that enable web applications to scale efficiently while maintaining predictable performance and low latency. The research seeks to identify the strengths and weaknesses of monolithic, microservice, serverless, and edge-based paradigms, and to outline practical guidelines for combining them into hybrid solutions suited for continuous integration and deployment environments.

Presentation of the main material

This study is based on a theoretical and applied analysis of architectural approaches for designing scalable web applications with a focus on performance and latency. The research draws on peer-reviewed publications, technical documentation from industry leaders, and case studies describing real-world deployments of microservice, serverless, and edge-based platforms. Architectural scenarios were modeled using representative CI/CD workflows, with emphasis on comparing containerized microservices managed by Kubernetes, serverless computing frameworks for handling burst workloads, and edge nodes designed to reduce user-perceived latency.

Performance estimations were developed through workload simulations reflecting typical e-commerce and media streaming environments, allowing assessment of throughput, response time, and fault tolerance under varying load conditions. Architectural models were constructed to illustrate data flow across application layers, orchestration mechanisms, and monitoring pipelines. Simulation results were analyzed in terms of scalability efficiency, stability during rapid workload changes, and the predictability of cost and latency metrics.

The evaluation also considered trade-offs such as communication overhead in microservices, state management complexity in serverless platforms, and configuration challenges in edge deployments. Comparative analysis was conducted to identify optimal configuration strategies and hybrid models that balance elasticity, reliability, and performance under continuous deployment environments.

The evaluation of architectural approaches confirmed that scalability and latency are strongly dependent on the choice of deployment model and workload profile. Comparative modeling was carried out for three paradigms – microservices orchestrated by Kubernetes, serverless platforms, and edge computing nodes – against a baseline monolithic system. The focus was on throughput, average response time, and stability under load. At the first stage, each architecture was subjected to increasing workloads emulating typical e-commerce operations such as concurrent transactions, search queries, and payment requests. This approach reflects real-world scenarios where traffic intensity fluctuates and service responsiveness directly affects user experience. Similar evaluation methods are widely applied in studies of high-performance computing systems and scalable architectures [7]. The results of such work illustrate how performance and scalability can be achieved by integrating flexible resource allocation with unified pipeline design, reinforcing the idea that both computational throughput and latency minimization depend on efficient orchestration and workload balancing. In parallel, research on cloud-native web applications highlights the importance of embedding performance and security considerations into the architecture itself [8].

These studies show that scalability alone cannot guarantee resilience in modern distributed systems, where latency bottlenecks are often compounded by security vulnerabilities. Findings indicate that embedding security principles within cloud-native architectures ensures that scaling strategies do not compromise stability or increase attack surfaces. Building on these insights, the present evaluation applied controlled simulations to reveal how architectural choices affect not only raw throughput and latency but also system robustness under stress. Figure 1 presents the change in average response time as the number of concurrent users grows from 500 to 5000.

The data in Figure 1 illustrate the change in average response time as the number of concurrent users increases from 500 to 5000 across four architectural models – monolithic, microservices, serverless, and edge computing. The monolithic system reaches a critical threshold at around 1500 users, where the average response time exceeds 900 ms and continues to grow beyond 3 seconds at 5000 users. This confirms the limited scalability of the traditional approach. The microservices model maintains stable performance much longer: up to 3000–3500 users, response time does not exceed

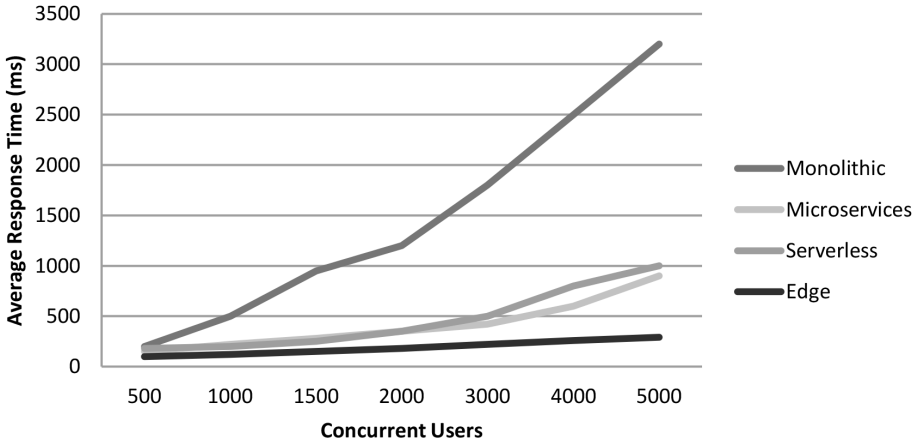


Fig. 1. Average response time under increasing load across four architectural models

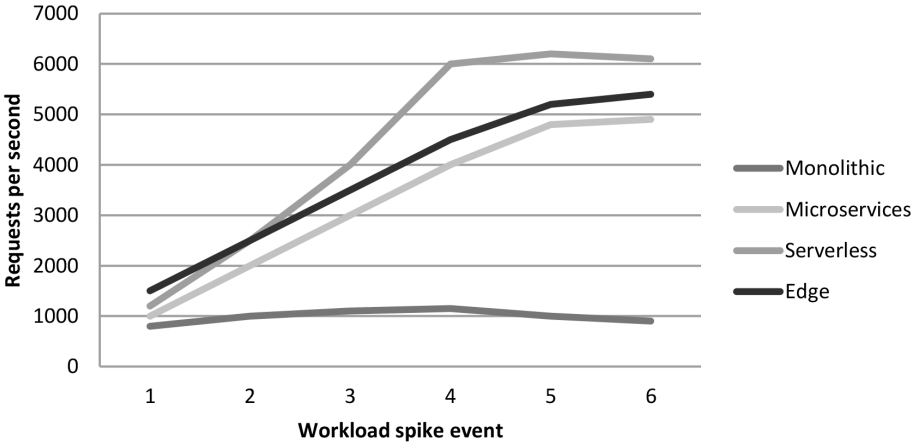


Fig. 2. Throughput dynamics under simulated workload spikes

420–600 ms. However, after this point, a steep rise occurs, with delays approaching 900 ms at peak load. The serverless model demonstrates strong elasticity under low and medium loads (250–350 ms). At 3000–4000 users, however, latency spikes up to 800–1000 ms appear, primarily due to cold-start delays and cost-related constraints. The edge architecture proves the most stable: even at 5000 users, the average response time remains below 300 ms. This highlights the advantage of localized processing, though at the cost of higher operational complexity. In addition, throughput and elasticity were analyzed. Figure 2 illustrates the throughput dynamics (requests per second) relative to workload spikes.

Figure 2 indicates that serverless platforms provide the most elastic throughput response, scaling almost instantly to accommodate traffic surges. Microservices achieve consistent linear scaling, but orchestration overhead slows reaction to sudden bursts. Edge deployments deliver strong local throughput but require prior resource provisioning, while the monolithic baseline rapidly saturates and fails under sharp load increases.

Further analysis shows that elasticity in serverless computing not only ensures rapid scaling but also corresponds with the latest surveys on architectural design [10]. While this model is highly adaptive to workload variability, cold-start delays and cost unpredictability remain critical limitations, which explains the latency spikes recorded in our experiments. Edge deployments, although effective in maintaining consistently low latency, also align with recent studies on distributed architectures [9]. These works emphasize that strong coordination and security mechanisms are essential for sustaining throughput across geographically distributed nodes. This insight is consistent with our findings, which revealed that edge computing requires substantial resource provisioning and robust orchestration to prevent performance bottlenecks. Overall, the data suggest that no single paradigm can provide complete elasticity, scalability, and cost predictability on its own. Instead, hybrid strategies that integrate serverless elasticity, microservice stability, and edge responsiveness represent the most effective architectural pathway for scalable web applications. Table 1 summarizes the key performance indicators observed for each architecture under standardized testing conditions.

Analysis of the data in Table 1 shows clear trade-offs. Monolithic systems remain cost-predictable but fail to meet performance requirements at scale. Microservices deliver balanced throughput and latency, though orchestration complexity grows with system size. Serverless models are most adaptive to unpredictable traffic but raise concerns about cost predictability and latency spikes from cold starts. Edge computing minimizes response time and user-perceived latency,

Table 1

Comparative performance of architectural models

Architecture	Peak throughput (req/s)	Avg. response time @ 3000 users (ms)	Latency floor (ms)	Cost predictability	Operational complexity
Monolithic	1,200	950	200	High	Low
Microservices	4,800	420	150	Medium	Medium-High
Serverless	6,200	500 (with spikes to 800)	180	Low (variable)	Medium
Edge computing	5,400	270	100	Medium	High

but at the price of higher deployment and maintenance overhead. These findings are consistent with broader surveys that emphasize the role of architectural coordination and management. Recent work on software-defined networking points out that scalability and latency optimization often depend on effective traffic management and placement strategies rather than raw computing power alone [11]. Similarly, studies of microservice-based architectures underline that while this paradigm supports modularity and migration flexibility, it also introduces complex integration challenges that must be resolved to achieve sustainable scalability [12]. Together, these perspectives reinforce the necessity of hybrid models that combine architectural efficiency with strong coordination and migration strategies.

The conducted analysis demonstrates that the scalability and latency of web applications are directly influenced by the chosen architectural paradigm and its ability to handle varying workload profiles. Monolithic architectures, while predictable in cost and simple to operate, cannot sustain modern performance demands once concurrent users exceed relatively low thresholds. Microservices extend scalability considerably, maintaining acceptable latency under higher loads, though orchestration introduces management overhead. Serverless platforms excel in elasticity and immediate response to traffic surges but face challenges with cold-start delays and cost predictability. Edge computing achieves the lowest user-perceived latency and strong local throughput, yet requires significant resource provisioning and operational expertise.

Taken together, these findings confirm that each architectural model has distinct advantages and limitations. No single approach is sufficient to ensure both high scalability and consistently low latency in real-world conditions. The results underline the value of hybrid deployment strategies, where microservices form the foundation for stable growth, serverless functions address burst workloads, and edge nodes optimize latency-sensitive interactions. Such combinations offer a more resilient, cost-aware, and performance-driven path for designing scalable web applications in CI/CD environments.

Conclusions

The conducted analysis clearly shows that the scalability and latency of web applications differ substantially across architectural models. The monolithic baseline proved least effective: already at around 1500 concurrent users, average response times exceeded 900 ms, and by 5000 users the delay grew beyond 3 seconds, confirming limited suitability for high-load systems. Microservices delivered more sustainable performance, maintaining acceptable response times up to 3000–3500 users, with an average of 420–600 ms under load. However, once this threshold was exceeded, latency increased sharply, illustrating the limits imposed by orchestration overhead. Serverless platforms demonstrated the strongest elasticity. They scaled almost instantly to accommodate workload spikes, achieving the highest peak throughput of 6200 requests per second. Nevertheless, cold-start delays and cost variability produced visible latency spikes in the range of 800–1000 ms at high concurrency levels, which reduced stability. Edge computing achieved the lowest latency overall, remaining below 300 ms even under maximum load. Throughput performance was also high, peaking at 5400 requests per second. However, this came with higher operational complexity and the requirement for prior resource provisioning. Taken together, the results demonstrate that no single paradigm can fully satisfy both scalability and latency requirements. Hybrid approaches, combining microservices for stable workloads, serverless for unpredictable bursts, and edge nodes for latency-sensitive operations, provide the most balanced and resilient solution. Such strategies enable web applications to maintain responsiveness, stability, and efficiency even under rapidly changing traffic conditions.

Bibliography

1. Стрельцов О., Орновецький Ю., Катріченко М. Підвищення ефективності масштабування архітектури хмарних додатків. *Електротехнічні та комп'ютерні системи*. 2023. № 38 (114). С. 58–65. DOI: https://doi.org/10.15276/eltecs.38.114.2023.7

2. Santos J., Wauters T., Volckaert B., De Turck F. Towards low-latency service delivery in a continuum of virtual resources: State-of-the-art and research directions. *IEEE Communications Surveys & Tutorials*. 2021. Vol. 23, No 4. P. 2557–2589. DOI: https://doi.org/10.1109/COMST.2021.3095358

3. Singhvi A., Balasubramanian A., Houck K., Shaikh M. D., Venkataraman S., Akella A. Atoll: A scalable low-latency serverless platform. *Proceedings of the ACM Symposium on Cloud Computing*. 2021. P. 138–152. DOI: https://doi.org/10.1145/3472883.3486981

4. Jia Z., Witchel E. Nightcore: efficient and scalable serverless computing for latency-sensitive, interactive microservices. *Proceedings of the 26th ACM international conference on architectural support for programming languages and operating systems*. 2021. P. 152–166. DOI: https://doi.org/10.5281/zenodo.4321760

5. Cherukuri B. R. Building scalable web applications: Best practices for backend architecture. *International Journal of Science and Research (IJSR)*. 2024. Vol. 13, No 10. P. 126–139. DOI: https://dx.doi.org/10.21275/ES24928085711

6. Blinowski G., Ojdowska A., Przybyłek A. Monolithic vs. microservice architecture: A performance and scalability evaluation. *IEEE Access*. 2022. No 10. P. 20357–20374. DOI: https://doi.org/10.1109/ACCESS.2022.3152803

7. Liao H., Tu J., Xia J., Liu H., Zhou X., Yuan H., Hu Y. Ascend: a scalable and unified architecture for ubiquitous deep neural network computing: Industry track paper. *2021 IEEE International Symposium on High-Performance Computer Architecture (HPCA)*. 2021. P. 789–801. DOI: https://doi.org/10.1109/HPCA51647.2021.00071

8. Datla L. S., Thodupunuri R. K. Designing for Defense: How We Embedded Security Principles into Cloud-Native Web Application Architectures. *International Journal of Emerging Research in Engineering and Technology*. 2021. Vol. 2, No 4. P. 30–38. DOI: https://doi.org/10.63282/3050-922X.IJERET-V2I4P104

9. Rahman A., Islam M. J., Band S. S., Muhammad G., Hasan K., Tiwari P. Towards a blockchain-SDN-based secure architecture for cloud computing in smart industrial IoT. *Digital Communications and Networks*. 2023. Vol. 9, No 2. P. 411–421. DOI: https://doi.org/10.1016/j.dcan.2022.11.003

10. Li Z., Guo L., Cheng J., Chen Q., He B., Guo M. The serverless computing survey: A technical primer for design architecture. *ACM Computing Surveys (CSUR)*. 2022. Vol. 54, No 10s. P. 1–34. DOI: https://doi.org/10.1145/3508360

11. Priyadarsini M., Bera P. Software defined networking architecture, traffic management, security, and placement: A survey. *Computer Networks*. 2021. No 192. P. 108047. DOI: https://doi.org/10.1016/j.comnet.2021.108047

12. Velepucha V., Flores P. A survey on microservices architecture: Principles, patterns and migration challenges. *IEEE Access*. 2023. No 11. P. 88339–88358. DOI: https://doi.org/10.1109/ACCESS.2023.3305687

References

1. Streltsov, O., Ornovetsky, Y., & Katrichenko, M. (2023). Pidvyshchennia efektyvnosti masshtabuvannia arkhitektury khmarnykh dodatkiv [Increasing the efficiency of cloud application architecture scaling]. *Elektrotekhnichni ta komp'iuterni systemy – Electrotechnic and Computer Systems*, 38(114), 58–65. https://doi.org/10.15276/eltecs.38.114.2023.7

2. Santos, J., Wauters, T., Volckaert, B., & De Turck, F. (2021). Towards low-latency service delivery in a continuum of virtual resources: State-of-the-art and research directions. *IEEE Communications Surveys & Tutorials*, 23(4), 2557–2589. https://doi.org/10.1109/COMST.2021.3095358

3. Singhvi, A., Balasubramanian, A., Houck, K., Shaikh, M. D., Venkataraman, S., & Akella, A. (2021, November). Atoll: A scalable low-latency serverless platform. *ACM Symposium on Cloud Computing*, 138–152. https://doi.org/10.1145/3472883.3486981

4. Jia, Z., & Witchel, E. (2021, April). Nightcore: efficient and scalable serverless computing for latency-sensitive, interactive microservices. *26th ACM international conference on architectural support for programming languages and operating systems*, 152–166. https://doi.org/10.5281/zenodo.4321760

5. Cherukuri, B. R. (2024). Building scalable web applications: Best practices for backend architecture. *International Journal of Science and Research (IJSR)*, 13(10), 126–139. https://dx.doi.org/10.21275/ES24928085711

6. Blinowski, G., Ojdowska, A., & Przybyłek, A. (2022). Monolithic vs. microservice architecture: A performance and scalability evaluation. *IEEE access*, 10, 20357–20374. https://doi.org/10.1109/ACCESS.2022.3152803

7. Liao, H., Tu, J., Xia, J., Liu, H., Zhou, X., Yuan, H., & Hu, Y. (2021, February). Ascend: a scalable and unified architecture for ubiquitous deep neural network computing: Industry track paper. *2021 IEEE International Symposium on High-Performance Computer Architecture (HPCA)*, 789–801. https://doi.org/10.1109/HPCA51647.2021.00071

8. Datla, L. S., & Thodupunuri, R. K. (2021). Designing for Defense: How We Embedded Security Principles into Cloud-Native Web Application Architectures. *International Journal of Emerging Research in Engineering and Technology*, 2(4), 30–38. https://doi.org/10.63282/3050-922X.IJERET-V2I4P104

9. Rahman, A., Islam, M. J., Band, S. S., Muhammad, G., Hasan, K., & Tiwari, P. (2023). Towards a blockchain-SDN-based secure architecture for cloud computing in smart industrial IoT. *Digital Communications and Networks*, 9(2), 411–421. https://doi.org/10.1016/j.dcan.2022.11.003

10. Li, Z., Guo, L., Cheng, J., Chen, Q., He, B., & Guo, M. (2022). The serverless computing survey: A technical primer for design architecture. *ACM Computing Surveys (CSUR)*, 54(10s), 1–34. https://doi.org/10.1145/3508360

11. Priyadarsini, M., & Bera, P. (2021). Software defined networking architecture, traffic management, security, and placement: A survey. *Computer Networks*, 192, 108047. https://doi.org/10.1016/j.comnet.2021.108047

12. Velepucha, V., & Flores, P. (2023). A survey on microservices architecture: Principles, patterns and migration challenges. *IEEE access*, 11, 88339–88358. https://doi.org/10.1109/ACCESS.2023.3305687

Дата першого надходження рукопису до видання: 27.09.2025
Дата прийнятого до друку рукопису після рецензування: 22.10.2025
Дата публікації:

УДК 004 DOI

О. А. СНЄОСІКОВ
аспірант кафедри кібербезпеки інформаційних систем, мереж і технологій
Харківський національний університет імені В. Н. Каразіна
ORCID: 0009-0001-9468-5965

О. П. НАРЄЖНІЙ
кандидат технічних наук,
доцент кафедри кібербезпеки інформаційних систем, мереж і технологій
Харківський національний університет імені В. Н. Каразіна
ORCID: 0000-0003-4321-0510

МОДЕЛІ ЗАГРОЗ ТА ПОРУШНИКА АВТОНОМНОЇ СИСТЕМИ ДИФЕРЕНЦІАЛЬНОЇ КОРЕКЦІЇ ГЛОБАЛЬНИХ НАВІГАЦІЙНИХ СУПУТНИКОВИХ СИСТЕМ

У статті розроблено формалізовану модель загроз та модель порушника для автономної системи диференціальної корекції (АСДК) глобальних навігаційних супутникових систем (GNSS), яка відноситься до об'єктів критичної інформаційної інфраструктури. Сформовано класифіковану модель порушника, що враховує типи атакуючих суб'єктів, їх мотивацію, технічні ресурси, рівень доступу та потенційні техніки атак. Запропоновано комплексну модель загроз, яка охоплює всі активи системи АСДК, включаючи GNSS-приймачі контрольно-коригуючих станцій (ККС) та АСДК, центри обробки даних, криптографічну інфраструктуру, транспортні канали передачі даних, сервери розповсюдження поправок та кінцевих користувачів. Для побудови моделі загроз використано комбінований методологічний підхід, що об'єднує сучасні методики PASTA (Process for Attack Simulation and Threat Analysis), STRIDE та MITRE ATT&CK (Enterprise та ICS). Розроблено типові сценарії мультивекторних атак з урахуванням GPS spoofing, GPS jamming, атак на інформаційно-комунікаційну інфраструктуру (ІКС), криптосистему, програмне забезпечення та фізичні компоненти системи. Окрему увагу приділено забезпеченню криптостійкості системи в умовах постквантового періоду шляхом впровадження національних постквантових алгоритмів шифрування (Калина, Скеля, Вершина, Купина) та використання квантових генераторів випадкових чисел (QRNG) у системі управління ключами. Проведено аналіз наслідків реалізації атак та обґрунтовано комплекс заходів захисту системи АСДК з урахуванням міжнародних стандартів ISO/IEC 27001, IEC 62443, NIST SP 800-30/53 та національних нормативів НД ТЗІ. Запропонована методика дозволяє створювати практично застосовні захищені архітектури для АСДК GNSS та інших елементів критичної інфраструктури з урахуванням майбутніх криптографічних викликів.

Ключові слова: *глушіння сигналу; спуфінг сигналу; глобальна навігаційна супутникова система; квантовий генератор випадкових чисел; кібератака; кібербезпека; система глобального позиціонування; система диференціальної корекції; інформаційно-комунікаційна система; система виявлення вторгнень; система запобігання вторгненням, модель загроз, модель порушника, автономна система диференціальної корекції (АСДК; Autonomous DGNSS).*

O. A. SNIEOSIKOV
Postgraduate Student at the Department of Cybersecurity
of Information Systems, Networks and Technologies
V. N. Karazin Kharkiv National University
ORCID: 0009-0001-9468-5965

O. P. NARIEZHNII
Candidate of Technical Sciences,
Associate Professor at the Department of Cybersecurity
of Information Systems, Networks and Technologies,
V. N. Karazin Kharkiv National University
ORCID: 0000-0003-4321-0510

THREAT AND VIOLATOR MODELS OF THE AUTONOMOUS SYSTEM OF DIFFERENTIAL CORRECTION OF GLOBAL NAVIGATION SATELLITE SYSTEMS

This paper presents the development of a formalized threat model and adversary model for an Autonomous Differential Global Navigation Satellite System (Autonomous DGNSS), which is classified as part of critical information infrastructure. A structured adversary model has been created, incorporating attacker types, motivations, technical resources, access

© Снєосіков О. А., Нарєжній О. П., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

levels, and potential attack techniques. A comprehensive threat model is proposed, covering all system assets of the Autonomous DGNSS, including GNSS receivers at both control and correction stations and the central system, data processing centers, cryptographic infrastructure, data transmission channels, correction data distribution servers, and end users. The threat model is constructed using a combined methodological approach that integrates modern frameworks such as PASTA (Process for Attack Simulation and Threat Analysis), STRIDE, and MITRE ATT&CK (Enterprise and ICS profiles). Representative multivector attack scenarios have been developed, taking into account GPS spoofing, GPS jamming, attacks on information and communication infrastructure, cryptosystems, software vulnerabilities, and physical components of the system. Particular attention is paid to ensuring the system's cryptographic resilience in the post-quantum era by implementing national post-quantum cryptographic algorithms (Kalyna, Skelia, Vershyna, Kupyna) and utilizing quantum random number generators (QRNG) within the key management infrastructure. The paper includes an impact assessment of potential attack scenarios and provides a comprehensive set of protection measures for the Autonomous DGNSS, incorporating international standards such as ISO/IEC 27001, IEC 62443, NIST SP 800-30/53, as well as national regulations (ND TZI, Ukraine). The proposed methodology enables the design of practically applicable and resilient architectures for Autonomous Differential GNSS systems and other elements of critical infrastructure, with consideration for emerging post-quantum cryptographic challenges.

Key words: *GPS jamming; GPS spoofing; Global Navigation Satellite System (GNSS); quantum random number generator (QRNG); cyberattack; cybersecurity; GPS; Differential GNSS (DGNSS); information and communication system; IDS; IPS; threat model, adversary model, Autonomous DGNSS.*

Постановка проблеми

GNSS є критично важливим фундаментом сучасної критичної інфраструктури, охоплюючи цивільну авіацію, морську навігацію, транспорт, геодезію, енергетику, фінансові системи, телекомунікації, оборонну галузь та інші системи, де потрібно не лише визначення точних координат, а також точне визначення часу [1, 2]. Однак базова точність GNSS-сигналів у відкритому доступі не завжди відповідає сучасним вимогам, особливо у сфері логістики, точного землеробства, інфраструктурного моніторингу та операторів мобільного зв'язку. Таким чином, для забезпечення точності визначення координат, а також визначення точного часу у реальному часі використовуються функціональні доповнення GNSS – системи диференціальної корекції (DGNSS) [3] – програмно-апаратні комплекси, що обчислюють поправки на основі даних від ККС. DGNSS обробляють дані у реальному часі, передають поправки зацікавленим користувачам через IP-мережі (мобільний зв'язок, наземні канали зв'язку, радіоканали), забезпечуючи високоточне позиціювання незалежно від сторонніх сервісів. Системи диференціальної корекції які є частиною GNSS класифікуються як об'єкти критичної інфраструктури, відповідно до Закону України «Про критичну інфраструктуру» № 1882-IX від 15 листопада 2021 року, де до переліку критично важливих сфер включено, зокрема, «15) космічну діяльність, космічні технології та послуги» [4].

Останні роки продемонстрували стрімке зростання кіберзагроз, спрямованих на компоненти GNSS [5, 6], зокрема GPS spoofing, GPS jamming, атаки на програмне забезпечення станцій та серверів [7, 8, 9]. За даними [10, 11, 12], атаки на GNSS стали настільки ефективними та доступними, що потенційно можуть паралізувати об'єкти критичної інфраструктури що несе серйозні наслідки для національної безпеки. Сучасні глобальні навігаційні супутникові системи є вразливими до широкого спектру кіберзагроз. Серед найбільш критичних для функціонування GNSS, так і систем диференціальної корекції (зокрема, автономних систем (АСДК)), виділяють:

– Глушіння сигналу (GPS jamming): використання перешкод для повного блокування прийому сигналів супутників, включаючи сигнали, що надходять на ККС;

– Спуфінг сигналу (GPS spoofing): навмисна генерація фальшивих сигналів навігації з метою введення в оману приймача GNSS, зокрема ККС.

– Кібератаки на наземну та хмарну інфраструктуру GNSS, такі як:

– DDoS-атаки; MITM-атаки; використання вразливостей ПЗ серверів і веб-інтерфейсів (переповнення буфера, ін'єкції коду, фішинг, тощо …);

– Фізичні атаки на наземні системи: пошкодження або захоплення ККС та ЦОД АСДК;

– Інсайдерські загрози: порушення з боку персоналу або адміністративних користувачів.

У зв'язку з викладеним вище, постає завдання побудови формалізованої моделі загроз і моделі порушника саме для Autonomous DGNSS.

Аналіз останніх досліджень і публікацій

У період 2021–2025 років спостерігається значне зростання досліджень, присвячених виявленню та протидії таким атакам на GNSS як GPS jamming та GPS spoofing. Зокрема, у роботі [13] запропоновано використання методів машинного та глибокого навчання для ідентифікації таких атак з високою точністю. В дослідженні [14] представлено систематичний огляд сучасних методів виявлення перешкод на GNSS, включаючи використання моделей машинного навчання та радіочастотної ідентифікації. Попри це, більшість досліджень зосереджені на загальних аспектах безпеки GNSS сигналів, розглядаючи протидію кібератакам GPS jamming та GPS spoofing, залишаючи поза увагою специфічні загрози для систем диференціальної корекції, а саме наземної інфраструктури DGNSS. Наприклад, у звіті ENISA «Space Threat Landscape» (2025) висвітлюються загрози для космічних систем [15], але детальний аналіз вразливостей

наземної інфраструктури DGNSS відсутній. Варто зазначити, що існуючі моделі загроз, такі як STRIDE, DREAD, VAST, PASTA, OCTAVE, LINDDUN [16–23], хоча й корисні для загального моделювання, не враховують специфіку DGNSS. Стандарт IEC 62443 [24] частково охоплює аспекти безпеки Operational technology систем (OT системи), але не надає конкретних рекомендацій щодо захисту компонентів архітектури DGNSS. Крім того, у звіті CrowdStrike «2025 Global Threat Report» [25] зазначається зростання кількості кібератак на критичну інфраструктуру, зокрема включаючи використання штучного інтелекту для автоматизації атак. Це підкреслює необхідність розробки спеціалізованих моделей загроз для систем, таких як АСДК GNSS, які є частиною критичної інфраструктури. Таким чином, існує очевидна наукова прогалина у дослідженнях, присвячених моделюванню загроз та порушників для систем диференціальної корекції GNSS, які в основному зосереджені на окремих технічних аспектах і не формують комплексної системної моделі кіберзагроз і порушника. Це обґрунтовує необхідність розробки формалізованої моделі загроз та моделі порушника, яка враховує специфіку архітектури автономних DGNSS, протоколів передачі даних, вразливостей програмного забезпечення та можливих сценаріїв атак. Незважаючи на наявність загальних підходів до моделювання кіберзагроз та моделей порушників (зокрема STRIDE, MITRE ATT&CK, PASTA), модель загроз, адаптована до архітектури DGNSS, залишається недостатньо дослідженою, як і об'єкти критичної інфраструктури взагалі [12]. Однак існують дослідження, що використовують STRIDE для моделювання загроз у суміжних сферах, таких як супутникові системи зв'язку, морські кіберфізичні системи та IoT-системи точного землеробства [26–28]. Ці дослідження демонструють застосування моделі STRIDE у різних сферах, що можуть бути релевантними для аналізу загроз у GNSS. Більшість наукових робіт по дослідженню захисту GNSS зосереджено на базовій безпеці самих GNSS-сигналів або захисті користувацьких пристроїв, без урахування особливостей архітектури інформаційно комунікаційної системи (ІКС) АСДК. Оскільки компрометація DGNSS може призвести до катастрофічних наслідків – збоїв у транспортних системах, логістичних маршрутах, аграрній техніці або військових навігаційних платформах, потреба в розробці формалізованої та практично придатної моделі загроз для АСДК є надзвичайно актуальною.

Формулювання мети дослідження

Метою цієї роботи є побудова формалізованої моделі кіберзагроз та моделі порушника, орієнтованих на АСДК. У роботі проаналізовано сучасні методи моделювання кіберзагроз та їхню застосовність до АСДК GNSS. Запропоновано формалізовану модель загроз, яка враховує специфіку протоколів передачі поправок, уразливості ККС, мережевих вузлів, веб-сервера, а також програмного забезпечення користувача. Сформовано модель порушника з урахуванням рівня доступу, технічних можливостей і мотивації. На основі побудованих моделей розроблено типові сценарії атак з оцінкою потенційних наслідків та обґрунтованими контрзаходами. Отримані результати можуть бути використані для побудови систем виявлення загроз, оцінки ризиків та проєктування кіберстійких Autonomous DGNSS. Для побудови формалізованої моделі загроз та моделі порушника було використано міжнародні стандарти ISO/IEC 27001, NIST SP 800-30/53, IEC 62443, а також національні НД ТЗІ 2.5-004-99 та НД ТЗІ 2.5-005-99 [24,29,30,31,32]. ISO/IEC 27001 надає структуру для ідентифікації активів та загроз; NIST забезпечує детальний підхід до ризик-менеджменту та профілювання атакуючих суб'єктів; IEC 62443 дозволяє врахувати архітектурні особливості OT-систем (OT, Operational Technology Systems), до яких належать автономні системи диференціальної корекції; НД ТЗІ – узгоджує моделі загроз із чинним законодавством України щодо технічного захисту інформації.

Викладення основного матеріалу дослідження

Запропонована архітектура автономної системи диференціальної корекції глобальної супутникової навігації передбачає побудову багаторівневої, захищеної інформаційно-комунікаційної інфраструктури, що має на меті забезпечувати підвищену точність GNSS сигналів та стійкість до кіберзагроз, зокрема у постквантовий період.

Основними компонентами АСДК архітектури є (рис. 1):

– Супутникове угруповання GNSS (GPS, Galileo, GLONASS, BeiDou), що забезпечують передачу навігаційних сигналів до користувачів, ККС та АСДК.

– ККС, рівномірно розташовані по території країни, приймають сигнали GNSS, та передають їх до Центрів Обробки Даних (ЦОД) АСДК через захищені IPSEC-канали з використанням національних криптоалгоритмів.

– Основний та резервний ЦОД, які забезпечують агрегацію, обробку, та зберігання даних, розрахунок диференціальних поправок, а також формування коригуючих повідомлень для кінцевих користувачів. Критичні дані шифруються за алгоритмом «Калина» (≥256 біт). Система управління ключами використовує QRNG; для криптографічних операцій застосовуються національні стандарти «Скеля» (ДСТУ 8961:2019), «Вершина» (ДСТУ 9212:2023) та геш-функція «Купина» (ДСТУ 7564:2014). Автентичність, цілісність і захищений обмін забезпечує інфраструктура відкритого ключа (ІВК).

– Веб-сервер високої доступності, що забезпечує надання коригуючої інформації та стандартизовані потоки поправок кінцевим користувачам через захищені канали зв'язку (TLS 1.3), використовуючи мобільні мережі та інтернет-з'єднання.

– Захищена ІКС, яка включає Security Operation Center (SOC), системи виявлення та протидії вторгненням (IDS/IPS), захист від DDoS-атак, багаторівневу аутентифікацію доступу, аудит операцій та механізми захисту від MITM атак та кібератак на сервіси синхронізації точного часу.

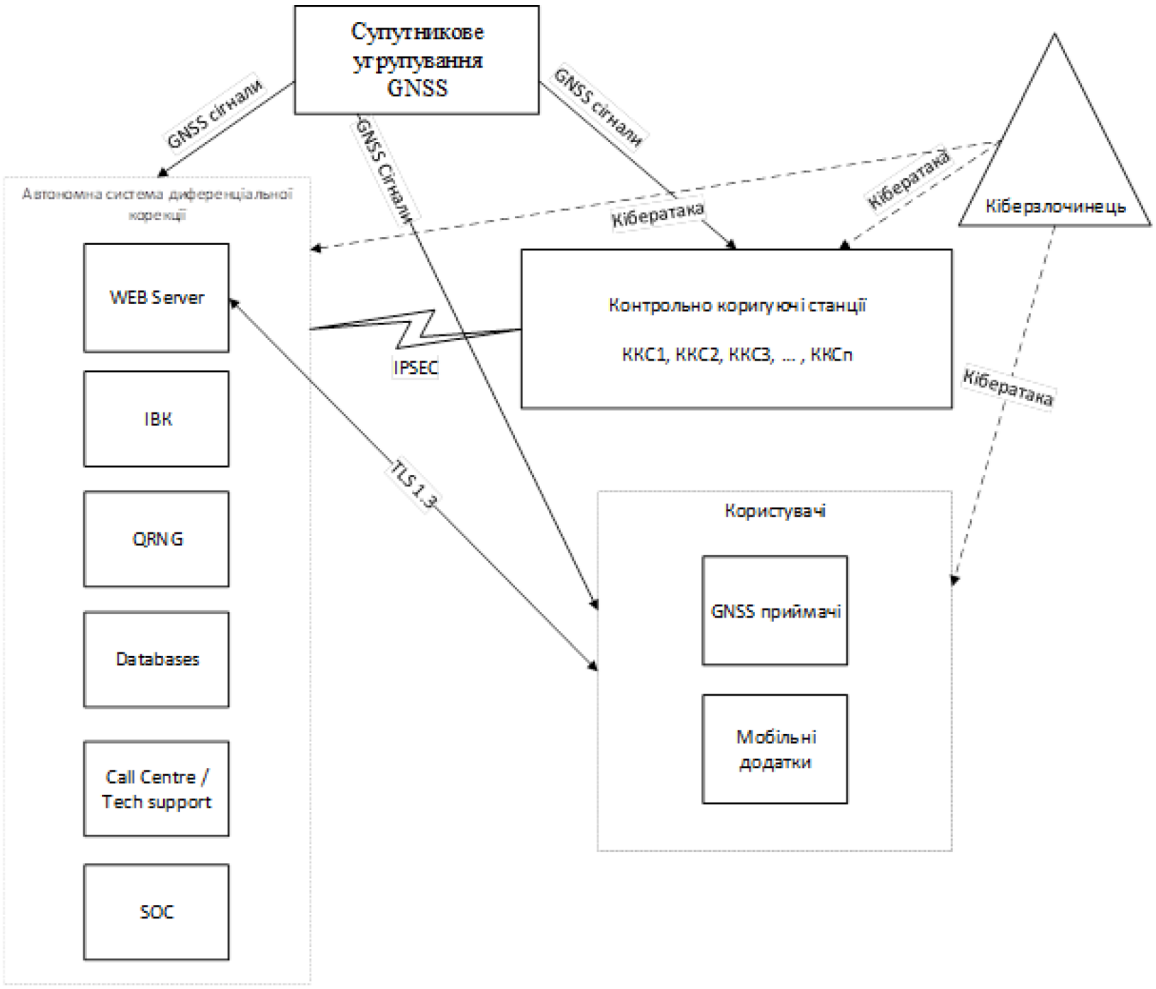


Рис 1. Прототип архітектури автономної системи диференціальної корекції GNSS

– Кінцеві користувачі отримують коригуючі дані за допомогою сертифікованих GNSS-приймачів та мобільних додатків з можливістю перевірки достовірності отриманих даних за допомогою електронного підпису.

Запропонована архітектура орієнтована на стійкість до широкого спектру кіберзагроз, серед яких: GPS jamming, GPS spoofing, DDoS-атаки, атаки типу MITM, експлуатація вразливостей ПЗ та фізичне проникнення до об'єктів інфраструктури. Інтеграція національних постквантових криптоалгоритмів забезпечує високий рівень криптографічної стійкості як в сучасних, так і у перспективних загрозових сценаріях зокрема у постквантовий період.

Враховуючи особливості архітектури АСДК та проаналізовані підходи (наведено у таблиці 1), пропонується комбінований метод моделювання загроз, що інтегрує переваги кількох моделей:

– PASTA забезпечує сценарне формування атак із врахуванням бізнес-цілей, моделі порушника, технічної архітектури та ризиків, що важливо для критичної інфраструктури;

– MITRE ATT&CK (ICS) надає деталізовану базу знань тактик, технік та процедур (TTPs) реальних атак, особливо актуальних для OT-систем, мережевих та хмарних компонентів ІКС АСДК;

– STRIDE застосовується для початкової класифікації загроз і допомагає визначити базові категорії вразливостей кожного компоненту системи.

Поєднання цих моделей дозволяє одночасно досягти високої деталізації технік атак і комплексного охоплення архітектурних рівнів АСДК. Під час побудови моделі загроз враховано вимоги та рекомендації наступних міжнародних та національних стандартів (наведено у таблиці 2):

Побудова загальної схеми моделювання загроз.

Для побудови моделі загроз АСДК GNSS було застосовано структурний підхід, що передбачає поетапне визначення ключових об'єктів захисту та потенційних точок реалізації кібератак:

На першому етапі здійснено ідентифікацію активів системи, до складу яких віднесено користувацькі та інфраструктурні GNSS-приймачі, ІКС з її мережевими, обчислювальними й криптографічними компонентами, а також програмне забезпечення та механізми забезпечення відмовостійкості й цілісності даних. Такий підхід дозволяє систематизувати об'єкти, що підлягають захисту, і створити основу для подальшої оцінки ризиків.

Таблиця 1

Аналіз основних підходів до моделювання загроз АСДК

Модель	Короткий опис	Переваги та Недоліки	Придатність до АСДК
STRIDE	Класифікація загроз за шістьма категоріями (Spoofing, Tampering, Repudiation, Information Disclosure, Denial of Service, Elevation of Privilege)	**Переваги:** простота, структурність; **Недоліки:** без сценаріїв, не для ОТ	Частково – для первинної класифікації загроз
PASTA	Сценарне моделювання атак через 7 етапів (бізнес-цілі, декомпозиція, загрози, вразливості, сценарії, ризики, контрзаходи)	**Переваги:** мотивація, складні атаки; **Недоліки:** багато даних	Висока – базова модель сценаріїв атак
MITRE ATT&CK (ICS)	База знань реальних тактик і технік атакуючих суб'єктів для ОТ-систем	**Переваги:** деталізація, систематизація; **Недоліки:** без ризиків	Висока – деталізація технік атак
OCTAVE	Організаційна модель управління ризиками	**Переваги:** бізнес-процеси, ризики; **Недоліки:** обмежена для технічного моделювання й ОТ	Обмежено – корисна на рівні аналізу активів
LINDDUN	Модель загроз приватності даних	**Переваги:** глибокий аналіз конфіденційності; **Недоліки:** не придатна для GNSS/ОТ	Не придатна для АСДК
DREAD	Кількісна оцінка ризиків (Damage, Reproducibility, Exploitability, Affected Users, Discoverability)	**Переваги:** кількісна оцінка загроз; **Недоліки:** суб'єктивність	Дуже обмежено
VAST	Інструмент для DevOps-орієнтованого моделювання загроз	**Переваги:** agile-підхід; **Недоліки:** не для ОТ	Не придатна для ОТ-систем

Таблиця 2

Перелік стандартів

Стандарт	Основні аспекти використання
ISO/IEC 27001	Ідентифікація активів, категоризація загроз, загальна структура аналізу ризиків
NIST SP 800-30/53	Оцінка ризиків, профілювання атакуючих суб'єктів, побудова сценаріїв загроз
IEC 62443	Особливості захисту ОТ-систем, до яких належить АСДК
НД ТЗІ 2.5-004-99, НД ТЗІ 2.5-005-99	Врахування національних вимог України до моделювання загроз і побудови моделей порушників

Другим етапом визначено точки доступу для потенційних атак (Attack Surface). До них належать канали прийому супутникових сигналів, мережеві тракти обміну інформацією між елементами системи, серверна інфраструктура, підсистеми безпеки, а також фізичний доступ до обладнання й людський фактор у вигляді можливих атак соціальної інженерії. Виявлення таких вразливих місць є необхідною передумовою для формування комплексної моделі загроз.

Третім етапом, що завершує процес, виступає класифікація виявлених загроз за STRIDE, яка узагальнено представлена в Таблиці 3.

Четвертим етапом загальної схеми моделювання загроз є застосування методології PASTA (Process for Attack Simulation and Threat Analysis), що реалізується через послідовні фази:

Фаза 1 – Define the Objectives. Визначаються ключові цілі системи: надання високоточних диференціальних поправок у реальному часі, гарантування достовірності й доступності даних, а також забезпечення стійкості критичної інфраструктури до кіберзагроз.

Фаза 2 – Define the Technical Scope. Формується технічний контур аналізу, що охоплює активи, точки доступу (Attack Surface) та класифікацію загроз за STRIDE.

Фаза 3 – Application Decomposition and Analysis. Система розкладається на основні компоненти: GNSS-приймачі, центри обробки даних, веб-сервер, SOC/SIEM, канали зв'язку (IPSEC, TLS), криптографічну інфраструктуру (QRNG, PKI) та кінцевих користувачів.

Фаза 4 – Threat Analysis. Ідентифікуються та класифікуються загрози: GPS jamming, GPS spoofing, MITM, DDoS, експлуатація уразливостей ПЗ, фізичні атаки, соціальна інженерія, компрометація ключів.

Фаза 5 – Vulnerability & Weakness Analysis. Виконується аналіз уразливостей і слабких місць системи (результати наведено у Таблиці 4).

Фаза 6 – Attack Scenarios (сценарії атак):

Сценарій 1 – мультивекторна атака «GPS Spoofing на ККС та GPS Jamming на АСДК-приймач». Атакуючий організовує GPS spoofing на кілька ККС, паралельно здійснюючи GPS jamming на центральний GNSS-приймач системи. Це призводить до обчислення фальшивих поправок. У разі виявлення невідповідностей механізми sanity-check активують fail-safe, що блокує видачу даних користувачам.

Сценарій 2 – атака MITM на канал «ККС–ЦОД». Порушник здійснює компрометацію або підміну вузла в захищеному каналі (IPSEC), модифікуючи дані від ККС. У результаті ЦОД приймає сфальсифіковані поправки, причому підміна може стосуватися як координат, так і часових міток.

Таблиця 3

Класифікація загроз за STRIDE

Категорія STRIDE	Опис загроз у контексті АСДК
Spoofing (Підміна ідентичності)	– GPS Spoofing на GNSS-приймачі АСДК, ККС та кінцевих користувачів; – Підміна ідентифікації користувачів або адміністраторів при доступі до ІКС АСДК; – Компрометація аутентифікаційних механізмів SOC, SIEM, ІВК
Tampering (Модифікація даних)	– Модифікація GNSS-сигналів при передачі; – Модифікація даних між ККС, ЦОД та веб-сервером; – Модифікація логіки обробки поправок у ПЗ серверів; – Модифікація ключових файлів та баз даних у криптографічній інфраструктурі (QRNG, ІВК)
Repudiation (Відмова від дій)	– Несанкціоновані дії адміністраторів та персоналу SOC без належного логування; – Відмова від відповідальності за зміну конфігурації ІКС АСДК; – Відсутність або компрометація журналів аудиту
Information Disclosure (Розголошення інформації)	– Витік ключової інформації криптографічної інфраструктури (закритих ключів ІВК, QRNG); – Перехоплення даних при передачі поправок між компонентами ІКС; – Несанкціонований доступ до даних ЦОД, SOC, SIEM
Denial of Service (Відмова в обслуговуванні)	– DDoS-атаки на веб-сервер високої доступності; – Перевантаження каналів передачі даних між ККС, ЦОД, веб-сервером; – Блокування SOC через перевантаження або саботаж ПЗ
Elevation of Privilege (Ескалація привілеїв)	– Експлуатація вразливостей у ПЗ серверів ЦОД, SOC, SIEM, веб-сервера; – Отримання повноважень адміністратора системи; – Захоплення контролю над ПЗ обробки поправок

Таблиця 4

Аналіз уразливостей

Компонент	Можливі уразливості
GNSS-приймачі	Відсутність захисту від GPS jamming/GPS spoofing, відсутність верифікації сигналів
Канали передачі	Уразливості IPSEC/TLS, MITM, недостатній захист трафіку
ЦОД	Експлойти ПЗ серверів, вразливості ОС, переповнення буфера
Веб-сервер	Вразливості веб-серверного ПЗ, атаки на API, DDoS
SOC/SIEM	Компрометація облікових записів SOC, некоректна конфігурація моніторингу
Криптоінфраструктура	Компрометація ІВК, генераторів QRNG
Кінцеві користувачі	Атаки на мобільні додатки, витік даних через користувача

Сценарій 3 – комбінована атака DDoS та експлуатація уразливостей ПЗ ЦОД. Виконується масований DDoS на веб-сервер, а паралельно застосовується експлуатація zero-day уразливості в програмному забезпеченні ЦОД. Це створює умови для перехоплення контролю над логікою обробки поправок і може призвести до саботажу або відключення функціоналу системи.

Сценарій 4 – інсайдерська атака на ІВК та SOC. Зловмисні дії адміністратора SOC або іншого інсайдера з доступом до криптографічної інфраструктури створюють ризик передачі критичної інформації третім особам. Наслідком є компрометація електронного підпису та порушення цілісності диференціальних поправок, що призводить до поширення некоректних даних серед користувачів.

Сценарій 5 – фізичне знищення або захоплення ККС. Організація фізичного саботажу ККС може призвести до часткової або повної втрати джерел поправок. Це зумовлює перехід системи в режим fail-safe та створює ризик cascading failure, коли зі зростанням кількості виведених з ладу станцій підтримка працездатності системи стає критично ускладненою.

Фаза 7 – Risk & Impact Analysis

Заключною фазою методології PASTA є оцінка ризиків та наслідків реалізації виявлених загроз. На цьому етапі результати попередніх фаз інтегруються у єдину модель, яка дозволяє співвіднести кожний сценарій атаки з потенційними збитками для системи Autonomous DGNSS.

Оцінювання здійснюється за такими ключовими параметрами:

– імовірність реалізації загрози, що визначається з урахуванням наявних вразливостей та складності експлуатації;

– масштаб впливу на систему, який відображає критичність наслідків для цілісності, доступності та достовірності поправок GNSS;

– системні та cascading effects, що виникають у разі поєднання декількох атак або послідовних відмов елементів (наприклад, виведення з ладу декількох ККС).

Отримані результати дозволяють здійснити ранжування загроз за рівнем ризику та визначити пріоритетність їх нейтралізації. У подальшому це забезпечує основу для розроблення багаторівневих заходів кіберзахисту, орієнтованих на мінімізацію ймовірності успішної атаки та зниження її потенційного впливу на критичну інфраструктуру.

Наступним етапом побудови моделі загроз для АСДК є декомпозиція технік атак з використанням фреймворку MITRE ATT&CK [33-34] (наведено у таблиці 5). На цьому етапі здійснюється ідентифікація конкретних TTPs (Tactics, Techniques, Procedures), характерних для дій атакуючих суб'єктів на різних стадіях життєвого циклу атаки – від розвідки до ексфільтрації даних чи саботажу.

Таблиця 5

Техніки атак за MITRE ATT&CK

Етап атаки (Tactic)	Техніка (Technique)	Реалізація в контексті АСДК
Reconnaissance (Розвідка)	Active Scanning (T1595)	Сканування мережевої інфраструктури ІКС АСДК, SOC, веб-сервера
	Gathering Victim Network Information (T1590)	Збір відомостей про розташування ККС, IP-адреси серверів ЦОД, криптографічну інфраструктуру
Resource Development (Підготовка ресурсів)	Obtain Capabilities (T1588)	Закупівля обладнання для GPS spoofing, jamming, zero-day експлойтів
Initial Access (Початковий доступ)	Supply Chain Compromise (T1195)	Компрометація сторонніх постачальників ПЗ або обладнання ІКС АСДК
	Spearphishing (T1566)	Соціальна інженерія проти адміністраторів SOC, SIEM
	Exploit Public-Facing Application (T1190)	Атаки на веб-сервер високої доступності через уразливості API
Execution (Виконання шкідливого коду)	Exploitation for Client Execution (T1203)	Експлуатація уразливостей серверів ЦОД (переповнення буфера, SQL injection)
Persistence (Закріплення в системі)	Valid Accounts (T1078)	Використання вкрадених або скомпрометованих облікових записів адміністраторів
Privilege Escalation (Ескалація привілеїв)	Exploitation for Privilege Escalation (T1068)	Експлуатація уразливостей ОС серверів, SOC, SIEM
Defense Evasion (Обхід захисту)	Indicator Removal on Host (T1070)	Видалення логів, журналів SOC
	Obfuscated Files or Information (T1027)	Шифрування зловмисного трафіку для обходу IDS/IPS
Credential Access (Доступ до облікових даних)	Credential Dumping (T1003)	Витягування паролів адміністраторів SOC та ІВК
Discovery (Дослідження системи)	Network Service Discovery (T1046)	Сканування внутрішніх мережних сегментів АСДК
Lateral Movement (Горизонтальне пересування)	Remote Services (T1021)	Використання легітимних протоколів для руху між ЦОД, SOC, веб-сервером
Collection (Збір інформації)	Data from Information Repositories (T1213)	Збір даних із серверів поправок, БД, криптоінфраструктури
Command and Control (C2)	Application Layer Protocol (T1071)	Створення прихованих каналів C2 через HTTP(S) API
Exfiltration (Експфільтрація)	Exfiltration Over C2 Channel (T1041)	Передача викрадених даних через приховані C2
Impact (Вплив/порушення роботи)	Service Stop (T1489)	Виведення з ладу серверів обробки поправок
	Data Manipulation (T1565)	Модифікація поправок перед розповсюдженням
	Denial of Service (T1499, T1498)	DDoS-атаки на веб-сервер та мережу передачі поправок
	Inhibit Recovery (T1490)	Знищення резервних копій, шифрування даних
Доповнення для GNSS/АСДК:		
Специфіка	**Техніка**	
GPS Spoofing	У MITRE ATT&CK формально немає окремої категорії для GPS Spoofing (поки що), але зазвичай відносять до « Data Manipulation (T1565, Enterprise); ICS: Spoof Reporting Message (T0856)/Manipulation of View (T0832)/Manipulation of Control (T0831) « в контексті OT	
GPS Jamming	Частково відноситься до « Network DoS (T1498)/Endpoint DoS (T1499); ICS: Denial of Service (T0814)/Denial of View (T0815)/Denial of Control (T0813) (залежно від ефекту) « в MITRE ATT&CK for ICS	

Аналіз наслідків атак (Risk & Impact Analysis) для АСДК виконано відповідно до підходів ISO/IEC 27005 [35], NIST SP 800-30/53, IEC 62443-3-2 та НД ТЗІ 2.5-005-99. Для кожної загрози визначаються: компонент системи, тип і механізм впливу, можливі наслідки та рівень критичності (високий, середній або низький). Узагальнені результати подано в Таблиці 6.

Ключові наслідки атак на АСДК охоплюють:

– втрату точності координат та порушення синхронізації часу (з ризиками для енергетичних, фінансових і телекомунікаційних систем);
– збої у транспортній та логістичній інфраструктурі, точному землеробстві та військовій навігації;
– масові відмови автопілотованого транспорту й блокування служб точного часу;
– вторинні катастрофічні наслідки для економіки, оборони та громадської безпеки.

Узагальнена карта ризиків наведена у таблиці 7.

Таблиця 6

Аналіз наслідків атак на АСДК

№	Компонент	Загроза	Механізм впливу	Наслідки	Критичність
1	GNSS-приймачі ККС та АСДК	GPS Spoofing	Підміна супутникових сигналів	Помилкові поправки, спотворення даних у всій системі	Висока
2	GNSS-приймачі ККС та АСДК	GPS Jamming	Перешкоди в прийомі сигналів	Повна відмова генерації поправок	Висока
3	Канали передачі (ККС – ЦОД – веб-сервер)	MITM	Модифікація даних у процесі передачі	Введення некоректних поправок	Висока
4	ЦОД	Експлойти ПЗ	Захоплення контролю над логікою обробки	Масова генерація хибних поправок	Висока
5	Веб-сервер	DDoS	Перевантаження серверів	Неможливість розповсюдження поправок	Висока
6	IBK, QRNG	Компрометація ключів	Доступ до криптографічної інфраструктури	Порушення цілісності даних	Висока
7	SOC, SIEM	Інсайдерські атаки	Несанкціоновані дії персоналу	Маніпуляції без виявлення	Висока
8	Мобільні додатки користувачів	Атаки на ПЗ	Компрометація додатків	Неправильна інтерпретація даних	Середня
9	Фізичні компоненти (ККС, ЦОД)	Фізичне знищення	Саботаж, диверсія	Втрата частини системи	Висока
10	Мережа управління	Соціальна інженерія	Отримання облікових даних адміністраторів	Отримання доступу до системи	Висока
11	Резервне копіювання	Inhibit Recovery	Знищення резервних копій	Неможливість відновлення	Висока

Таблиця 7

Категорія ризику	Глобальна оцінка
Технічні ризики	Високі
Організаційні ризики	Високі
Людський фактор	Високий
Криптографічні ризики	Критичні
Фізичні ризики	Високі

Побудова моделі порушника:

1) Підходи до формування моделі порушника.

Модель порушника (adversary model, threat actor model) – дозволяє формалізувати та класифікувати потенційних атакуючих суб'єктів, які можуть здійснювати цілеспрямовані впливи на функціонування АСДК GNSS. Вона дозволяє системно описати:

– типи порушників;
– їхні ресурси та можливості;
– цілі та мотивацію;
– потенційні техніки атак.

Побудова моделі порушника базується на методології, закладеній у стандартах:

– ISO/IEC 15408 (Common Criteria) [36];
– NIST SP 800-30/53;
– НД ТЗІ 2.5-005-99 (Україна).

2) Класифікація порушника (надана в Таблиці 8):
3) Мотивація порушника (надана в Таблиці 9):
4) Ресурси порушника (надано в Таблиці 10):
5) Основні цілі атак

– Порушення точності позиціювання та мітки часу;
– Зрив коректної роботи GNSS користувачів;
– Компрометація інфраструктури ІКС АСДК;
– Блокування розповсюдження поправок (DDoS на веб-сервер);
– Масова дезінформація користувачів через генерацію фальшивих поправок та спотвореної мітки часу;
– Фізичне знищення компонентів системи;
– Порушення довіри до державної системи точного часу та навігації.

6) Зв'язок з моделлю загроз

Формована модель порушника є основою для побудови конкретної моделі загроз, оскільки визначає:

– рівень складності атак;
– ймовірні вектори атак;

Таблиця 8

Класифікація атакуючих суб'єктів

№	Категорія	Характеристика	Типові техніки (MITRE ATT&CK)	Основні активи атаки
1	Державні актори (APT-групи, розвідслужби)	Висококваліфіковані групи з необмеженими ресурсами, здатні до складних мультивекторних атак, включно із GPS spoofing, GPS jamming, експлуатацією вразливостей ПЗ, атак на криптосистеми, фізичних операцій, соціальний інженерінг, DDoS, MITM, тощо	T1590, T1566, T1190, T1068, T1499, T1565, T1003, ICS T0814	GNSS-приймачі ККС/АСДК, ЦОД, ІКС, (ІВК, QRNG), SOC, користувачі
2	Кіберзлочинні угруповання	Мотивовані фінансовою вигодою. Можуть атакувати АСДК з метою отримання неправомірної переваги в логістиці, агросекторі, транспорті тощо. Використовують доступні експлойти, DDoS, GPS spoofing/ GPS jamming з обмеженим радіусом дії	T1190, T1071, T1499, T1566, T1070	Веб-сервер, API, SOC, фінансові системи
3	Хактивісти (політичні/ідеологічні групи)	Мета – продемонструвати вразливість державної критичної інфраструктури. Основні вектори: DDoS, DoS, соціальна інженерія, спроби зламу веб-сервера	T1499, T1027, T1566	Веб-сервер, API
4	Інсайдери (персонал, адміністратори)	Особи з привілейованим доступом, які можуть навмисно або через помилки створити вразливості. Вектори атак – саботаж, зміна параметрів ПЗ, витік ключової інформації, людський фактор	T1078, T1070, T1003, T1071	SOC, ІВК, ЦОД, Storage, ключові БД
5	Терористичні та диверсійні групи	Мета – фізичне знищення або захоплення компонентів АСДК, зокрема ККС, ЦОД або SOC. Можуть комбінувати фізичні атаки з кібердіями	ICS T0814, ICS T0813/T0815, Physical Sabotage	ККС, ІКС, ЦОД, SOC
6	Низькорівневі зловмисники (script kiddies, аматори)	Використовують готові інструменти для DDoS, доступні вразливості в ПЗ, експлойти нульового дня за умови доступу до них на «чорному ринку»	T1190, T1499, T1071	API, веб-сервер, публічні сервіси

Таблиця 9

Мотивація атакуючих

Тип порушника	Основна мотивація
Державні актори	Геополітичні цілі, ослаблення супротивника, зрив навігаційних сервісів, компрометація національної критичної інфраструктури
Кіберзлочинці	Економічна вигода, шантаж, маніпуляція ринками
Хактивісти	Політичні заяви, демонстрація вразливостей критичних систем, політичний тиск
Інсайдери	Особиста вигода, помста, політичні мотиви
Терористи	Масовий саботаж, підрив критичної інфраструктури
Низькорівневі зловмисники	Навчання, експерименти, прагнення слави

Таблиця 10

Рівень ресурсів атакуючих суб'єктів

Рівень ресурсу	Категорії
Дуже високий	Державні актори
Високий	Терористичні організації, кіберзлочинні угруповання
Середній	Інсайдери, хактивісти
Низький	Низькорівневі зловмисники

– цільові активи системи;
– необхідні сценарії протидії.

Вона лягає в основу сценарного моделювання атак (PASTA), класифікації STRIDE, технічної деталізації MITRE ATT&CK та формування контрзаходів.

Висновки

У даній роботі проведено комплексне дослідження базових моделей загроз та моделей порушника для застосування їх до АСДК GNSS, яка є частиною критичної інформаційної інфраструктури.

Розроблено формалізовану модель порушника, що дозволяє класифікувати атакуючі суб'єкти за рівнем ресурсів, мотивацією, технічними можливостями та типами потенційних атак. Побудовано модель загроз, яка інтегрує сучасні методології моделювання, зокрема:

– класифікацію загроз за STRIDE;
– сценарне моделювання атак за PASTA;
– деталізацію технік атак за MITRE ATT&CK (Enterprise та ICS профілі).

Проаналізовано повний набір активів та зон атак системи АСДК, включаючи GNSS-приймачі ККС та АСДК, інформаційно-комунікаційну інфраструктуру, центри обробки даних, криптографічну інфраструктуру, веб-сервер високої доступності, систему моніторингу SOC та кінцевих користувачів.

Наукова новизна роботи полягає у вперше запропонованому інтегрованому підході до моделювання загроз для АСДК, який об'єднує STRIDE, PASTA та MITRE ATT&CK з урахуванням специфіки GNSS/DGNSS.

Проведено аналіз потенційних сценаріїв мультивекторних атак, які включають комбінації GPS spoofing, GPS jamming, атак на канали зв'язку, компрометації програмного забезпечення, інсайдерських дій, атак на криптоінфраструктуру та фізичного втручання у роботу системи. Встановлено, що особливу небезпеку для АСДК становлять скоординовані атаки з високим рівнем ресурсної підтримки, що здатні одночасно впливати на декілька ключових компонентів архітектури усієї системи. Особливо критичними є атаки, що призводять до генерації некоректних поправок і масового спотворення позиціювання та мітки часу, які можуть мати катастрофічні наслідки для логістичних, енергетичних, фінансових, військових та аграрних секторів. Запропоновано комплексний набір технічних, організаційних, криптографічних та процедурних контрзаходів, який базується на поєднанні міжнародних стандартів ISO/IEC 27001, ISO/IEC 27005, IEC 62443, NIST SP 800-30/53 та національних нормативних документів НД ТЗІ України. Додатково підкреслено важливість використання сучасних засобів автентифікації даних GNSS, захисту каналів розповсюдження поправок та стійкої синхронізації часу, що є визначальними для кіберстійкості АСДК.

Отримані результати можуть бути використані як методологічна основа при:

– проектуванні захищених систем диференціальної корекції GNSS;
– створенні систем виявлення атак;
– розробці нормативних документів щодо кіберстійкості критичної інфраструктури;
– подальших наукових дослідженнях в галузі кібербезпеки GNSS.

Перспективою подальших досліджень є розробка моделей активного виявлення мультивекторних атак на основі машинного навчання, побудова систем прогнозування атак та удосконалення криптографічних протоколів з урахуванням викликів постквантового періоду.

Список використаної літератури

1. European Union Agency for the Space Programme. EUSPA EO and GNSS Market Report. 2024. Iss. 2. LU: Publications Office, 2024. URL: https://data.europa.eu/doi/10.2878/73092

2. Precision Matters: Exploring the Importance of GPS Precision Accuracy. Taoglas. 20.07.2024. URL: https://www.taoglas.com/blogs/precision-matters-exploring-the-importance-of-gps-precision-accuracy

3. DGNSS Fundamentals – Navipedia. URL: https://gssc.esa.int/navipedia/index.php?title=DGNSS_Fundamentals

4. Закон України «Про критичну інфраструктуру», № 1882-IX редакція від 21.09.2024. Офіційний вебпортал парламенту України. URL: https://zakon.rada.gov.ua/go/1882-20

5. Westbrook T. A. Taxonomy of Radio Frequency Jamming and Spoofing Strategies and Criminal Motives. *Journal of Strategic Security*. 2023. Vol. 16, no. 2. P. 68–80. DOI: https://doi.org/10.5038/1944-0472.16.2.2081

6. Westbrook T. The Global Positioning System and Military Jamming: The geographies of electronic warfare. *Journal of Strategic Security*. Vol. 12, № 2. P. 1-16. DOI:10.5038/1944-0472.12.2.1720

7. Навігаційні ризики в аспекті кібербезпеки транспортних суден і військових кораблів. *ResearchGate*, 2024. doi: 10.51582/interconf.19-20.08.2022.037

8. Garmin outage caused by confirmed WastedLocker ransomware attack. BleepingComputer. URL: https://www.bleepingcomputer.com/news/security/garmin-outage-caused-by-confirmed-wastedlocker-ransomware-attack/

9. KA-SAT Network cyber attack overview. viasat.com. 30.03.2022. URL: https://news.viasat.com/blog/corporate/ka-sat-network-cyber-attack-overview

10. GNSS Interference Monitoring and Classification for Critical Infrastructure Safety. GPSPATRON.com. URL: https://gpspatron.com/gnss-interference-monitoring-and-classification-for-critical-infrastructure-safety.

11. Tauschinski J., Georgiadis P., Wright R. et al. How GPS warfare is playing havoc with civilian life. *Financial Times*. 13.05.2024. URL: https://www.ft.com/content/be9393db-cd63-4141-a4c8-c16b4fe1b6b0

12. Melnyk D. S. Creating a model of threats to Ukraine's national critical infrastructure as a basis for ensuring its security and resilience. *Bulletin of Kharkiv National University of Internal Affairs*. Vol. 104, 1 (Part 1). C. 237–250. DOI:10.32631/v.2024.1.20

13. Ghanbarzade A., Soleimani H. GNSS/GPS Spoofing and Jamming Identification Using Machine Learning and Deep Learning. *arXiv*, 2025. DOI:10.48550/arXiv.2501.02352

14. Radoš K., Brkić M., Begušić D. Recent Advances on Jamming and Spoofing Detection in GNSS. *Sensors*.Vol. 24, № 13. C. 4210. DOI:10.3390/s24134210

15. European Union Agency for Cybersecurity., Space threat landscape. LU: Publications Office, 2025. URL: https://data.europa.eu/doi/10.2824/8841206

16. Naik, Nitin & Jenkins, Paul & Grace, Paul & Naik, Dishita & Phd, Shaligram & Song, Jingping. A Comparative Analysis of Threat Modelling Methods: STRIDE, DREAD, VAST, PASTA, OCTAVE, and LINDDUN. 2024. DOI:10.1007/978-3-031-74443-3_16

17. Shostack A. Threat modeling: designing for security. Indianapolis, IN : Wiley, 2014. 590 с. [QA76.9.A25 S495 2014]. ISBN 978-1-118-80999-0.

18. UcedaVelez T., Morana M. M. Risk Centric Threat Modeling: Process for Attack Simulation and Threat Analysis. 1st. Wiley Publishing, 2015. 696 p. ISBN 978-0-470-50096-5.

19. Threat Modeling Methodology: PASTA. URL: https://www.iriusrisk.com/resources-blog/pasta-threat-modeling-methodologies

20. Alberts C. J., Behrens S. G., Pethia R. D. та ін. Operationally Critical Threat, Asset, and Vulnerability Evaluation (OCTAVE) Framework, Version 1.0: (Fort Belvoir, VA, 01.06.1999). Fort Belvoir, VA : Defense Technical Information Center, 1999. DOI:10.21236/ADA367718. 1999.

21. LINDDUN PRO. linddun.org. URL: https://linddun.org/pro/

22. Threat Modeling with Microsoft DREAD. *Satori*. URL: https://satoricyber.com/glossary/threat-modeling-with-microsoft-dread/

23. VAST Threat Methodology. ThreatModeler. URL: https://threatmodeler.com/glossary/vast-threat-methodology/

24. ISA/IEC 62443 Series of Standards – ISA. isa.org. URL: https://www.isa.org/standards-and-publications/isa-standards/isa-iec-62443-series-of-standards

25. 2025 Global Threat Report. Latest Cybersecurity Trends & Insights. CrowdStrike. CrowdStrike.com. URL: https://www.crowdstrike.com/en-us/global-threat-report/

26. Asif M. R. A., Hasan K. F., Islam M. Z. et al. STRIDE-based Cyber Security Threat Modeling for IoT-enabled Precision Agriculture Systems. *arXiv*, 2022. DOI:10.48550/arXiv.2201.09493

27. Sahay R., Estay D. A. S., Meng W. et al. A Comparative Risk Analysis on CyberShip System with STPA-Sec, STRIDE and CORAS. *arXiv*, 2022. DOI:10.48550/arXiv.2212.10830

28. Sheik A. T., Atmaca U. I., Maple C. et al. Challenges in threat modelling of new space systems: A teleoperation use-case. *Advances in Space Research*. Vol. 70, Issue 8. P. 2208-2226. DOI:10.1016/j.asr.2022.07.013

29. Force J. T. Security and Privacy Controls for Information Systems and Organizations. *National Institute of Standards and Technology*. 2020. DOI:10.6028/NIST.SP.800-53r5. 2020.

30. ISO/IEC 27001:2022. ISO. URL: https://www.iso.org/standard/27001

31. Критерії оцінки захищеності інформації в комп'ютерних системах від несанкціонованого доступу – Інформаційна безпека та захист інформації. URL: https://tzi.ua/ua/nd_tz_2.5-004-99.html

32. Класифікація автоматизованих систем і стандартні функціональні профілі захищеності оброблюваної інформації від несанкціонованого доступу – Інформаційна безпека та захист інформації. URL: https://tzi.ua/ua/nd_tz_2.5-005-99.html

33. Techniques – Enterprise. MITRE ATT&CK®. URL: https://attack.mitre.org/techniques/enterprise/

34. Techniques – ICS. MITRE ATT&CK®. URL: https://attack.mitre.org/techniques/ics/

35. ISO/IEC 27005:2022. ISO. URL: https://www.iso.org/standard/80585.html

36. ISO/IEC 15408-1:2022. ISO. URL: https://www.iso.org/standard/72891.html

References

1. European Union Agency for the Space Programme. (2024) EUSPA EO and GNSS Market Report. 2024. Iss. 2. LU: Publications Office, URL: https://data.europa.eu/doi/10.2878/73092

2. Precision Matters: Exploring the Importance of GPS Precision Accuracy. (2024) Taoglas. 20.07.2024. URL: https://www.taoglas.com/blogs/precision-matters-exploring-the-importance-of-gps-precision-accuracy

3. DGNSS Fundamentals – Navipedia. (2011) URL: https://gssc.esa.int/navipedia/index.php?title=DGNSS_Fundamentals.

4. Zakon Ukrayiny «Pro krytychnu infrastrukturu» № 1882-IX redaktsiya vid 21.09.2024 [Law of Ukraine "On Critical Infrastructure" No. 1882-IX, version dated 09/21/2024]. Official web portal of the Parliament of Ukraine. URL: https://zakon.rada.gov.ua/go/1882-20

5. Westbrook T. A. (2023) Taxonomy of Radio Frequency Jamming and Spoofing Strategies and Criminal Motives. *Journal of Strategic Security*, Vol. 16, no. 2. P. 68-80. DOI: https://doi.org/10.5038/1944-0472.16.2.2081

6. Westbrook T. (2019) The Global Positioning System and Military Jamming: The geographies of electronic warfare. *Journal of Strategic Security*, Vol. 12, № 2. P. 1–16. DOI:10.5038/1944-0472.12.2.1720

7. Navihatsiyni ryzyky v aspekti kiberbezpeky transportnykh suden i viys′kovykh korabliv (2024) [Navigational risks in terms of cybersecurity of transport vessels and warships]. *ResearchGate*, doi: 10.51582/interconf.19-20.08.2022.037

8. Garmin outage caused by confirmed WastedLocker ransomware attack. (2020) BleepingComputer. URL: https://www.bleepingcomputer.com/news/security/garmin-outage-caused-by-confirmed-wastedlocker-ransomware-attack/

9. KA-SAT Network cyber attack overview. (2022) Viasat.com. 30.03.2022. URL: https://news.viasat.com/blog/corporate/ka-sat-network-cyber-attack-overview

10. GNSS Interference Monitoring and Classification for Critical Infrastructure Safety. (2022) GPSPATRON.com. URL: https://gpspatron.com/gnss-interference-monitoring-and-classification-for-critical-infrastructure-safety.

11. Tauschinski J., Georgiadis P., Wright R. et al. (2024) How GPS warfare is playing havoc with civilian life. *Financial Times*, 13.05.2024. URL: https://www.ft.com/content/be9393db-cd63-4141-a4c8-c16b4fe1b6b0

12. Melnyk D. S. (2024) Creating a model of threats to Ukraine's national critical infrastructure as a basis for ensuring its security and resilience. *Bulletin of Kharkiv National University of Internal Affairs*, Vol. 104, 1 (Part 1). C. 237–250. DOI:10.32631/v.2024.1.20

13. Ghanbarzade A., Soleimani H. (2025) GNSS/GPS Spoofing and Jamming Identification Using Machine Learning and Deep Learning. *arXiv*, DOI:10.48550/arXiv.2501.02352

14. Radoš K., Brkić M., Begušić D. (2024) Recent Advances on Jamming and Spoofing Detection in GNSS. *Sensors*, Vol. 24, № 13. C. 4210. DOI:10.3390/s24134210

15. European Union Agency for Cybersecurity. (2025) Space threat landscape. LU: Publications Office. URL: https://data.europa.eu/doi/10.2824/8841206.

16. Naik? Nitin & Jenkins, Paul & Grace, Paul & Naik, Dishita & Phd, Shaligram & Song, Jingping. (2024) A Comparative Analysis of Threat Modelling Methods: STRIDE, DREAD, VAST, PASTA, OCTAVE, and LINDDUN. DOI:10.1007/978-3-031-74443-3_16

17. Shostack A. (2014) Threat modeling: designing for security. Indianapolis, IN : Wiley, 590 c. [QA76.9.A25 S495 2014]. ISBN 978-1-118-80999-0.

18. UcedaVelez T., Morana M. M. (2015) Risk Centric Threat Modeling: Process for Attack Simulation and Threat Analysis. 1st. Wiley Publishing, 696 p. ISBN 978-0-470-50096-5.

19. Threat Modeling Methodology: PASTA. (2023) URL: https://www.iriusrisk.com/resources-blog/pasta-threat-modeling-methodologies

20. Alberts C. J., Behrens S. G., Pethia R. D. (1999) Operationally Critical Threat, Asset, and Vulnerability Evaluation (OCTAVE) Framework, Version 1.0: (Fort Belvoir, VA, 01.06.1999). Fort Belvoir, VA : Defense Technical Information Center, DOI:10.21236/ADA367718. 1999

21. LINDDUN PRO. linddun.org. URL: https://linddun.org/pro/

22. Threat Modeling with Microsoft DREAD. *Satori*. URL: https://satoricyber.com/glossary/threat-modeling-with-microsoft-dread/

23. VAST Threat Methodology. ThreatModeler. URL: https://threatmodeler.com/glossary/vast-threat-methodology/

24. ISA/IEC 62443 Series of Standards – ISA. isa.org. URL: https://www.isa.org/standards-and-publications/isa-standards/isa-iec-62443-series-of-standards

25. 2025 Global Threat Report. Latest Cybersecurity Trends & Insights. CrowdStrike. CrowdStrike.com. URL: https://www.crowdstrike.com/en-us/global-threat-report/

26. Asif M. R. A., Hasan K. F., Islam M. Z. et al. (2022) STRIDE-based Cyber Security Threat Modeling for IoT-enabled Precision Agriculture Systems. *arXiv*. DOI:10.48550/arXiv.2201.09493

27. Sahay R., Estay D. A. S., Meng W. et al. (2022) A Comparative Risk Analysis on CyberShip System with STPA-Sec, STRIDE and CORAS. *arXiv*. DOI:10.48550/arXiv.2212.10830

28. Sheik A. T., Atmaca U. I., Maple C. et al. (2022) Challenges in threat modelling of new space systems: A teleoperation use-case. *Advances in Space Research*. Vol. 70, Issue 8. P. 2208-2226. DOI:10.1016/j.asr.2022.07.013

29. Force J. T. (2020) Security and Privacy Controls for Information Systems and Organizations. *National Institute of Standards and Technology*. DOI:10.6028/NIST.SP.800-53r5. 2020.

30. ISO/IEC 27001: 2022. (2022) ISO. URL: https://www.iso.org/standard/27001

31. Kryteriyi otsinky zakhyshchenosti informatsiyi v komp″yuternykh systemakh vid nesanktsionovanoho dostupu. [Criteria for assessing the security of information in computer systems from unauthorized access]. *Information security and information protection*. URL: https://tzi.ua/ua/nd_tz_2.5-004-99.html

32. Klasyfikatsiya avtomatyzovanykh system i standartni funktsional′ni profili zakhyshchenosti obroblyuvanoyi informatsiyi vid nesanktsionovanoho dostupu [Classification of automated systems and standard functional profiles of security of processed information from unauthorized access]. *Information security and information protection*. URL: https://tzi.ua/ua/nd_tz_2.5-005-99.html

33. Techniques – Enterprise. MITRE ATT&CK®. URL: https://attack.mitre.org/techniques/enterprise/

34. Techniques – ICS. MITRE ATT&CK®. URL: https://attack.mitre.org/techniques/ics/

35. ISO/IEC 27005:2022. ISO. URL: https://www.iso.org/standard/80585.html

36. ISO/IEC 15408-1:2022. ISO. URL: https://www.iso.org/standard/72891.html

Дата першого надходження рукопису до видання: 24.09.2025
Дата прийнятого до друку рукопису після рецензування: 21.10.2025
Дата публікації:

УДК 004.8 DOI

О. В. СОБКО
Ph.D., старший викладач кафедри комп'ютерних наук
Хмельницький національний університет
ORCID: 0000-0001-5371-5788

М. О. МОЛЧАНОВА
Ph.D., старший викладач кафедри комп'ютерних наук
Хмельницький національний університет
ORCID: 0000-0001-9810-936X

О. В. МАЗУРЕЦЬ
кандидат технічних наук,
доцент кафедри комп'ютерних наук
Хмельницький національний університет
ORCID: 0000-0002-8900-0650

П. В. ГНАТЮК
студент кафедри комп'ютерних наук
Хмельницький національний університет

АЛГОРИТМІЗАЦІЯ ПРОЦЕСІВ КЕРУВАННЯ ТА ІНТЕРАКЦІЙ В ІГРОВИХ ОБ'ЄКТНО-ОРІЄНТОВАНИХ ЗАСТОСУНКАХ

У статті досліджується алгоритмізація процесів керування і інтеракцій в ігрових об'єктно-орієнтованих застосунках на прикладі гри «Гольф», розробленої з використанням мови програмування C# та ігрового рушія Unity. Розглянуто основні аспекти механізмів вибору рівнів, управління ігровим об'єктом, проходження рівнів та нарахування винагород у вигляді зірок, а також обробку ситуацій поразки. Вказано, що інтеграція взаємозалежних функцій у відповідні класи дозволяє забезпечити модульність, масштабованість і повторне використання коду, що є ключовим для складних ігрових систем. Створена діаграма класів відображає структуру зв'язків між основними та допоміжними класами, включно з MonoBehaviour, який забезпечує взаємодію скриптів із рушієм Unity через події Start(), Update(), OnCollisionEnter() та інші, а також користувацькі класи, що реалізують поведінку м'яча, управління сценами, відображення інтерфейсу, обробку колізій та ефекти, логіку ворогів та рух об'єктів на сцені.

У статті показано, що застосування запропонованих алгоритмів вибору рівнів, керування м'ячем, проходження рівнів та нарахування зірок, поразки дозволяє створити реалістичний та керований ігровий процес, а також забезпечує точність розрахунку траєкторій, сили поштовху та реакцій на зіткнення. Розроблені механіки можуть слугувати основою для подальшого проєктування більш складних ігор із високим рівнем інтерактивності, динамічною фізичною моделлю та керованими елементами інтерфейсу. Запропоновані у статті алгоритми демонструють ефективність об'єктно-орієнтованого підходу в програмуванні ігор, що сприяє підвищенню продуктивності розробки та полегшує масштабування проєкту, що є важливим під час розробки та супроводу програмного забезпечення.

Ключові слова: *алгоритмізація, об'єктно-орієнтоване програмування, Unity, керування об'єктами, фізика ігрового процесу, діаграма класів.*

O. V. SOBKO
Ph.D., Senior Lecturer at the Computer Science Department
Khmelnytskyi National University
ORCID: 0000-0001-5371-5788

M. O. MOLCHANOVA
Ph.D., Senior Lecturer at the Computer Science Department
Khmelnytskyi National University
ORCID: 0000-0001-9810-936X

© Собко О. В., Молчанова М. О., Мазурець О. В., Гнатюк П. В., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

O. V. MAZURETS
Ph.D. in Engineering Science,
Associate Professor at the Computer Science Department
Khmelnytskyi National University
ORCID: 0000-0002-8900-0650

P. V. HNATIUK
Bachelor Student at the Computer Science Department
Khmelnytskyi National University

ALGORITHMISATION OF CONTROL AND INTERACTION PROCESSES IN OBJECT-ORIENTED GAME APPLICATIONS

The article investigates the algorithmization of control processes and interactions in object-oriented game applications using the example of the game "Golf", developed using the C# programming language and the Unity game engine. The main aspects of the level selection mechanisms, game object management, level passing and star reward accrual, as well as handling defeat situations are considered. It is indicated that the integration of interdependent functions into the corresponding classes allows for modularity, scalability and code reuse, which is key for complex game systems. The created class diagram reflects the structure of relationships between the main and auxiliary classes, including MonoBehaviour, which provides script interaction with the Unity engine via Start(), Update(), OnCollisionEnter() and other events, as well as user classes that implement ball behavior, scene management, interface display, collision handling and effects, enemy logic and object movement on the scene.

The article shows that the application of the proposed algorithms for level selection, ball control, level progression, and star and defeat accrual allows for the creation of a realistic and controllable gameplay, as well as ensuring the accuracy of trajectories, thrust forces, and collision reactions. The developed mechanics can serve as the basis for further designing more complex games with a high level of interactivity, a dynamic physical model, and controllable interface elements. The algorithms proposed in the article demonstrate the effectiveness of an object-oriented approach in game programming, which helps to increase development productivity and facilitates project scaling, which is important during software development and maintenance.

Key words: *algorithmization, object-oriented programming, Unity, object management, gameplay physics, class diagram.*

Постановка проблеми

Сучасна індустрія комп'ютерних ігор та інтерактивних застосунків відзначається високим рівнем складності програмних систем [1], що обумовлює необхідність ефективного моделювання поведінки ігрових об'єктів та організації їх взаємодії [2]. Об'єктно-орієнтоване програмування забезпечує структурований підхід до розробки таких систем завдяки використанню принципів інкапсуляції, успадкування та поліморфізму [3, 4], що дозволяє створювати модульні, масштабовані та повторно використовувані компоненти [5]. Ключовим аспектом розробки об'єктно-орієнтованих ігрових застосунків є алгоритмізація процесів керування та інтеракцій [6, 7]. Ефективна алгоритмізація включає розробку алгоритмів обробки подій [8], колізій, керування станами об'єктів. Сучасні ігрові рушії, зокрема Unity, у поєднанні з мовою програмування C# надають широкий спектр інструментів для реалізації таких алгоритмів [9, 10]. Середовище Unity і C# також сприяють масштабованості системи та повторному використанню кодових модулів [11], що є важливим для підтримки та розвитку складних ігрових проєктів [12].

Водночас розробка ефективних алгоритмів керування та інтеракцій у об'єктно-орієнтованих ігрових системах залишається проблематичною [13] через необхідність одночасної інтеграції різних підходів до поведінки об'єктів, адаптації до дій користувача та оптимізації продуктивності. Неоптимізована логіка призводить до затримок, зниження частоти кадрів та погіршення користувацького досвіду [14], тоді як відсутність системного підходу ускладнює масштабування та повторне використання компонентів.

Аналіз останніх досліджень і публікацій

Сучасні дослідження в галузі розробки об'єктно-орієнтованих ігрових застосунків підтверджують ефективність поєднання об'єктно-орієнтованого програмування з алгоритмізацією процесів керування та інтеракцій. Ряд наукових робіт показує, що об'єктно-орієнтоване програмування дозволяє структуровано організовувати поведінку ігрових об'єктів у вигляді класів та модулів, що підвищує масштабованість системи та спрощує повторне використання коду [15].

Стаття [16] зосереджена на розробці високорівневої структури гри з використанням Unity та мови програмування C#. Автори детально описують принципи проєктування, побудову основних модулів та застосування реальних прикладів. Вони підкреслюють важливість організації коду для підвищення ефективності розробки та якості Unity-проєктів.

У роботі [17] розглядається використання Unity для навчання програмуванню ігор. Автор описує розробку шаблону гри RPG, який використовується як навчальний матеріал для студентів. Робота акцентує увагу на важливості використання об'єктно-орієнтованого підходу та алгоритмізації в процесі навчання програмуванню ігор.

У статті [18] представлено досвід розробки адаптивних ігор з анімованими персонажами в середовищі Unity. Автори описують процес створення інтерактивного середовища, яке дозволяє розробляти ігри з анімованими персонажами, що реагують на дії користувача.

Стаття [19] досліджує методи ефективної розробки мобільних ігор на сучасних пристроях, акцентуючи увагу на оптимізації коду, алгоритмізації процесів керування та інтеракцій, використанні об'єктно-орієнтованого підходу та повному використанні обчислювальних і графічних можливостей мобільних платформ для створення складних і адаптивних ігрових систем.

Аналіз наукових джерел показує, що розробка об'єктно-орієнтованих ігрових застосунків значною мірою залежить від алгоритмізації процесів керування, оскільки саме вона забезпечує узгодженість взаємодії об'єктів, оптимізацію ресурсів і підтримку масштабованості проєкту.

Формулювання мети дослідження

Метою дослідження є проєктування алгоритмів процесів керування та інтеракцій в ігрових об'єктно-орієнтованих застосунках з використанням C# і Unity на прикладі гри «Гольф».

Викладення основного матеріалу дослідження

Розробка сучасних ігрових застосунків вимагає не лише високоякісного візуального оформлення, а й чіткої організації логіки керування та взаємодії між об'єктами [20]. Для досягнення цього важливим є використання об'єктно-орієнтованого підходу, який забезпечує модульність, масштабованість та зручність супроводу програмного коду [21]. У межах цього дослідження здійснено проєктування алгоритмів процесів керування та інтеракцій в ігрових застосунках із застосуванням мови програмування C# та ігрового рушія Unity. Як приклад реалізації обрано створення гри «Гольф», що дало змогу продемонструвати особливості побудови ігрової логіки та взаємодії об'єктів у рамках компонентної архітектури Unity.

Алгоритм вибору рівнів передбачає активацію панелі з кнопками для запуску окремих сцен після натискання кнопки Play та виклик панелі підтвердження виходу при натисканні Quit, що забезпечує керовану навігацію між рівнями та узгодженість станів інтерфейсу. Псевдокод алгоритму подано далі.

Алгоритм 1. Алгоритм вибору рівнів.

```
приНатисканніКнопки(кнопка):
   якщо кнопка == «Play»:
      деактивувати(КнопкаPlay)
      деактивувати(КнопкаQuit)
      активувати(ПанельРівнів)
   якщо кнопка == «Quit»:
      деактивувати(ГоловнеМенюUI)
      активувати(ПанельПідтвердженняВиходу)

приНатисканніКнопки(підтвердженняВиходу):
   якщо підтвердженняВиходу == «Так»:
      завершитиПрограму()
   якщо підтвердженняВиходу == «Ні»:
      деактивувати(ПанельПідтвердженняВиходу)
      активувати(ГоловнеМенюUI)

приНатисканніКнопки(кнопкаРівня):
   якщо кнопкаРівня у КнопкахРівнів:
      завантажитиСцену(кнопкаРівня.індексСцени)

приНатисканніКнопки(закритиПанельРівнів):
   деактивувати(ПанельРівнів)
   активувати(КнопкаPlay)
   активувати(КнопкаQuit)
```

Ігровий процес реалізовано через алгоритм керування м'ячем, який поєднує обробку натискань миші, візуалізацію прицілювання, симуляцію руху з урахуванням сили поштовху, поступове гальмування та обробку зіткнень. Потрапляння м'яча в лунку фіксує проходження рівня, а у разі інших зіткнень застосовується відбиття за фізичними законами. Псевдокод алгоритму подано далі.

Алгоритм 2. Алгоритм керування м'ячем.

```
Оновлення():
   якщо натиснутаЛіваКнопкаМиші:
      попадання = променевийВистріл(позиціяМиші, шарМ'яча)
```

```
      якщо попадання == м’яч:
         режимПрицілювання = true
         початковаПозиція = позиціяМиші

   якщо режимПрицілювання і утримуєтьсяЛіваКнопкаМиші:
      векторТяги = початковаПозиція - позиціяМиші
      сила = обмежити(довжина(векторТяги), 0, максСила)
      показатиЛініюПрицілювання(сила, векторТяги)

   якщо режимПрицілювання і відпущенаЛіваКнопкаМиші:
      м’яч.швидкість = нормалізувати(векторТяги) * сила
      режимПрицілювання = false
      приховатиЛініюПрицілювання()
      збільшитиЛічильникХодів()
      відтворитиЗвукУдару()

Фізика():
   якщо м’яч.рухається:
      застосуватиГальмування(м’яч)
      якщо м’яч.швидкість < мінШвидкість:
         м’яч.швидкість = 0
         м’яч.рухається = false

приКолізії(м’яч, об’єкт):
   якщо об’єкт.тег == «loonka»:
      зупинитиМ’яч()
      активуватиПеремогуРівня()
   інакше:
      відбити(м’яч.швидкість, нормальКолізії)
```

Для оцінювання результатів гравця створено алгоритм підрахунку ходів і нарахування зірок, що активується при досягненні лунки. У цей момент зупиняється внутрішній ігровий час і з’являється панель перемоги з опціями переходу до наступного рівня, перезапуску чи повернення в меню. Псевдокод алгоритму подано далі.

Алгоритм 3. Алгоритм проходження рівнів і нарахування зірок

```
колиМ’ячДосягЛунки():
   зупинитиІгровийЧас()
   ходи = отриматиЛічильникХодів()
   зірки = обчислитиЗірки(ходи, пороговіЗначенняРівня)
   показатиПанельПеремоги(зірки)

приНатисканніКнопки(кнопкаПанеліПеремоги):
   якщо кнопка == «Next Level»:
      завантажитиСцену(індексПоточноїСцени + 1)
   якщо кнопка == «Again»:
      перезапуститиПоточнуСцену()
   якщо кнопка == «Main Menu»:
      завантажитиСцену(0)
   відновитиІгровийЧас()
```

Передбачено також алгоритм поразки: при зіткненні з небезпечними об’єктами активується панель повідомлення про невдачу, зупиняється ігровий час, а гравцеві пропонуються лише варіанти перезапуску або виходу до меню. Псевдокод алгоритму подано далі.

Алгоритм 4. Алгоритм поразки.

```
приКолізії(м’яч, об’єкт):
   якщо об’єкт.тег == «перешкода»:
      зупинитиІгровийЧас()
      показатиПанельПоразки()

приНатисканніКнопки(кнопкаПанеліПоразки):
   якщо кнопка == «Again»:
```

```
    перезапуститиПоточнуСцену()
якщо кнопка == «Main Menu»:
    завантажитиСцену(0)
відновитиІгровийЧас()
```

У процесі розробки створено діаграму класів (рисунок 1), що відображає основні зв'язки та функціональне призначення класів і їх методів [22].

У розробленому ігровому проєкті не було необхідності створювати єдиний головний клас, що об'єднує всю програму, оскільки цю функцію виконує вбудований у рушій Unity клас MonoBehaviour. MonoBehaviour, що належить до простору імен UnityEngine, забезпечує скриптам взаємодію з рушієм через автоматично викликані події, такі як Start(), Update(), OnCollisionEnter() та інші. Серед користувацьких класів основними є movement та SceneManagerScript. Клас movement відповідає за логіку поведінки м'яча на рівнях, включаючи надання імпульсу, розрахунок траєкторії, початкової швидкості, зміни швидкості, визначення нового положення, обробку відбиттів від поверхні, контакт із лункою та підрахунок кількості ударів. Оскільки ці функції взаємозалежні, їх інтеграція в одному класі є доцільною.

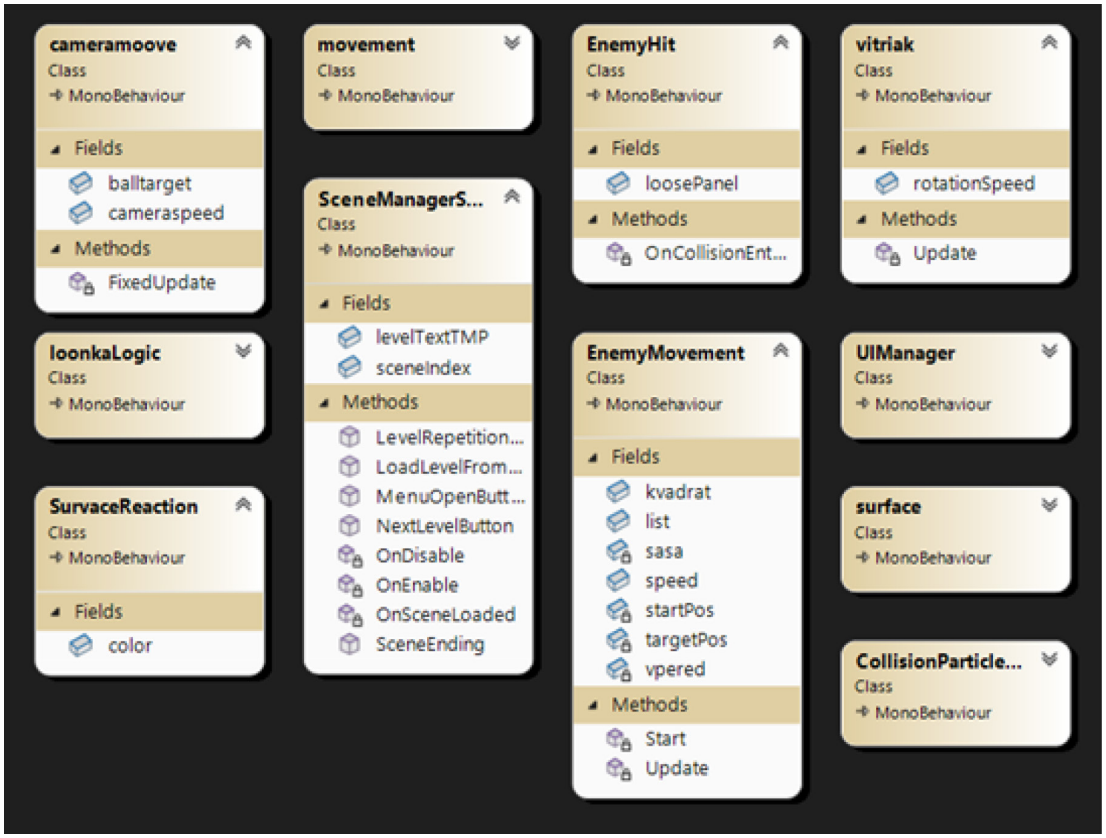


Рис. 1. Діаграма класів

Клас SceneManagerScript реалізує управління сценами за допомогою вбудованих механізмів Unity Scene Management. Додатково в проєкті використовуються класи допоміжного призначення: loonkaLogic – для відображення панелі перемоги та підрахунку ходів; UIManager – для відображення результатів гравця; CollisionParticleManager та SurfaceReaction – для системи ефектів при колізіях; vitryak – для обертання об'єктів на сцені; EnemyHit та EnemyMovement – для логіки поведінки ворожих об'єктів; Surface – для впливу властивостей поверхні на швидкість м'яча; cameramoove – для плавного слідування камерою за м'ячем.

Таким чином, усі класи взаємодіють між собою для забезпечення повноцінної ігрової логіки та інтерактивності об'єктів.

Висновки

У статті розглянуто алгоритмізацію процесів керування та інтеракцій в об'єктно-орієнтованих ігрових застосунках на прикладі гри «Гольф», розробленої з використанням C# та Unity. Проведено аналіз ключових алгоритмів, що забезпечують вибір рівнів, керування ігровим об'єктом, проходження рівня, нарахування винагород та обробку поразки.

Було продемонстровано, що інтеграція взаємозалежних функцій у відповідні класи дозволяє забезпечити модульність, масштабованість та ефективну взаємодію об'єктів у грі. Створена діаграма класів відображає структуру взаємозв'язків між основними і допоміжними класами, що забезпечує керованість ігрового процесу та гнучкість у розвитку проєкту. Розроблені алгоритми і класи можуть слугувати основою для подальшого проєктування складніших ігрових систем із високим рівнем інтерактивності та реалістичною фізичною моделлю.

Список використаної літератури

1. Kumar K. Game-changing intelligence: Unveiling the societal impact of artificial intelligence in game software / K. Kumar, N. Veena, T. Aravind, C. Bhatt, U. Kuppusamy, P. Jain // Entertainment Computing. 2025. Vol. 52. P. 100862. https://doi.org/10.1016/j.entcom.2024.100862

2. Mazurets O., Uspenska K., Vit R., Tyschenko O. Intelligent System for Determining the Object Attributes Values by Neural Networks Means by Graphic Images in Databases / O. Mazurets, K. Uspenska, R. Vit, O. Tyschenko // Current Trends in the Development of Scientific Research in Today's Conditions: proceedings of XXV International Scientific and Practical Conference (May 29–31, 2024, Florence, Italy). Florence, Italy: International Scientific Unity, 2024. P. 86–91.

3. Молчанова М. О., Мазурець О. В., Собко О. В., Кліменко В. І., Андрощук В. І. Метод нейромережевого виявлення кібербулінгу з використанням хмарних сервісів та об'єктно-орієнтованої моделі / М. О. Молчанова, О. В. Мазурець, О. В. Собко, В. І. Кліменко, В. І. Андрощук // Вісник Хмельницького національного університету. Серія: Технічні науки. – Хмельницький, 2024. № 2 (333). С. 200–206. https://doi.org/10.31891/2307-5732-2024-333-2

4. Молчанова М. О., Мазурець О. В., Шевчук П. О., Кліменко В. І., Тищенко О. О. Підхід до тестування об'єктно-орієнтованих систем керування в електронній комерції / М. О. Молчанова, О. В. Мазурець, П. О. Шевчук, В.І. Кліменко, О.О. Тищенко // Наука і техніка сьогодні. Київ, 2025. № 4 (45). С. 1273–1285. https://doi.org/10.52058/2786-6025-2025-4(45)-1273-1285

5. Мазур Є. В., Мазурець О. В., Кліменко В. І., Собко О. В., Залуцька О. О. Алгоритми та програмна архітектура інформаційної системи нейромережевого аналізу постави людини / Є. В. Мазур, О. В. Мазурець, В. І Кліменко, О. В. Собко, О. О. Залуцька // Вісник Хмельницького національного університету. Серія: Технічні науки. Хмельницький, 2025. № 3, Т. 1. С. 275–284. https://doi.org/10.31891/2307-5732-2025-351-35

6. Мазурець О. В., Молчанова М. О., Собко О. В., Дідур В. О., Багрій Р. О. Інтеграція об'єктно-орієнтованого підходу та реляційних баз даних при проєктуванні і розробленні веб-платформи для діячів мистецтва / О. В. Мазурець, М. О. Молчанова, О. В. Собко, В. О. Дідур, Р. О. Багрій // Наука і техніка сьогодні. Київ, 2025. № 8 (49). С. 1542–1555. https://doi.org/10.52058/2786-6025-2025-8(49)-1542-1555

7. Molchanova M., Mazurets O., Sobko O., Boiarchuk I. Object-Oriented Approach for Ethnic Enmity Detection in Text Messages by NLP / M. Molchanova, O. Mazurets, O. Sobko, I. Boiarchuk // Proceedings of XXI International Scientific and Practical Conference «Scientific Achievements and Innovations as a Way to Success» (May 1–3, 2024, Vilnius, Lithuania). 2024. P. 73–77.

8. Молчанова М. О., Мазурець О. В., Собко О. В., Віт Р. В., Назаров В. В. Алгоритм виявлення аб'юзивного вмісту в україномовному аудіоконтенті для імплементації в об'єктно-орієнтовану інформаційну систему / М. О. Молчанова, О. В. Мазурець, О. В. Собко, Р. В. Віт, В. В. Назаров // Вісник Хмельницького національного університету. Серія: Технічні науки. Хмельницький, 2024. № 1 (331). С. 101–106. https://doi.org/.31891/2307-5732-2024-331-17

9. Molchanova M., Mazurets O., Klimenko V., Kuflevsky Ev. Object-oriented model for neural network damage detection of mail packages / M. Molchanova, O. Mazurets, V. Klimenko, Ev. Kuflevsky // Proceedings of XIV International Scientific and Practical Conference «Solving Scientific Problems Using Innovative Concepts» (March 13–15, 2024, Copenhagen, Denmark). 2024. P. 58–62.

10. Mazurets O., Sobko O., Vit R., Pasternak V. Practical Approach for Detection by Deep Learning of Target Objects of Subject Area Based on Semantic Connectivity Indicators in Audio Database / O. Mazurets, O. Sobko, R. Vit, V. Pasternak // Proceedings of XXIV International Scientific and Practical Conference "Modern Scientific Challenges are the Driving Force of the Development of Scientific Research" (May 22–24, 2024, Bruges, Belgium). International Scientific Unity, 2024. P. 91–96.

11. Kok I. A., Kadynska V. D., Zalutska O. O., Mazurets O. V. Object-Oriented Intelligent System for Automated Control of Smoking by Video Data / I. A. Kok, V. D. Kadynska, O. O. Zalutska, O. V. Mazurets // Current scientific goals, approaches and challenges: Proceedings of IV International Scientific and Theoretical Conference. Dresden, Federal Republic of Germany, 2025. Pp. 156–164.

12. Mazurets O. V., Klimenko V. I., Molchanova M. O., Sultanov A. V. Object-Oriented Intelligent System for Neural Network Detection of Sugar Crystallization Zones / O. V. Mazurets, V. I. Klimenko, M. O. Molchanova, A. V. Sultanov // Global Science: Prospects and Innovations: Proceedings of the 10th International Scientific and Practical Conference. Liverpool, United Kingdom: Cognum Publishing House, 2024. P. 198–207.

13. Мазурець О. В., Молчанова М. О., Кліменко В. І., Собко О. В., Супрун П. К. Даталогічна модель бази даних для виявлення гендерної приналежності за SVM-аналізом дописів інтернет-мереж з використанням об'єктно-

орієнтованого проєктування / О. В. Мазурець, М. О. Молчанова, В. І. Кліменко, О. В. Собко, П. К. Супрун // Вісник Хмельницького національного університету. Серія: Технічні науки. Хмельницький, 2024. № 3, Т. 2 (337). С. 197–204. https://doi.org/10.31891/2307-5732-2024-337-3-29

14. Molchanova M., Didur V., Mazurets O., Sobko O., Zakharkevich O. Method for Construction and Demolition Waste Classification Using Two-Factor Neural Network Image Analysis / M. Molchanova, V. Didur, O. Mazurets, O. Sobko, O. Zakharkevich // CEUR Workshop Proceedings. 2025. Vol. 3970. P. 168–182. URL: https://ceur-ws.org/Vol-3970/PAPER14.pdf (date of access: 19.09.2025).

15. Molchanova M., Didur V., Sobko O., Mazurets O. Detection of Web Propaganda Patterns by Transformer Neural Networks: Improving Efficiency via Dataset Balancing / M. Molchanova, V. Didur, O. Sobko, O. Mazurets // CEUR Workshop Proceedings. 2025. Vol. 3988. P. 112–126. URL: https://ceur-ws.org/Vol-3988/paper10.pdf (date of access: 19.09.2025).

16. Zhang Q., Mao N., Li J., Zhou N., Miao J., Li G. Design of the Top-Level Code Framework for Unity Based on C / Q. Zhang, N. Mao, J. Li, N. Zhou, J. Miao, G. Li // ResearchGate. URL: https://www.researchgate.net/publication/391193129_Design_of_the_Top-Level_Code_Framework_for_Unity_Based_on_C_Language (date of access: 19.09.2025).

17. Golob S. P., Pagnon Eriksson N. Using Unity to teach Game Programming / S. P. Golob, N. Pagnon Eriksson // Malmö University, Game Development Program, Computer Science and Media Technology. URL: https://www.diva-portal.org/smash/get/diva2%3A1867700/FULLTEXT02.pdf (date of access: 19.09.2025).

18. Torun B., Karakurt S., Aydin T. B., Altunel Y. Game Development on Unity / B. Torun, S. Karakurt, T. B. Aydin, Y. Altunel // Turkish Online Journal of Educational Technology-TOJET. 2021. Vol. 20, No. 1. P. 39–43. URL: https://eric.ed.gov/?id=EJ1290793 (date of access: 19.09.2025).

19. Fannoun Q. K. Escape the Planet: Revolutionizing Game Design with Novel OOP Techniques / Q. K. Fannoun. 2024. URL: https://cornerstone.lib.mnsu.edu/etds/1406/ (date of access: 19.09.2025).

20. Wu C. H., Chien Y. C., Chou M. T., Huang Y. M. Integrating computational thinking, game design, and design thinking: a scoping review on trends, applications, and implications for education / C. H. Wu, Y. C. Chien, M. T. Chou, Y. M. Huang // Humanities and Social Sciences Communications. 2025. Vol. 12, No. 1. P. 1–12. https://doi.org/10.1057/s41599-025-04502-x

21. Mazurets O., Sobko O., Dydo R., Zalutska O., Molchanova M. Augmented reality audiostream creation using CNN: boosting inclusion and safety for visually impaired people / O. Mazurets, O. Sobko, R. Dydo, O. Zalutska, M. Molchanova // CEUR Workshop Proceedings. 2025. Vol. 4004. P. 347–361. URL: https://ceur-ws.org/Vol-4004/paper26.pdf (date of access: 19.09.2025).

References

1. Kumar, K., Veena, N., Aravind, T., Bhatt, C., Kuppusamy, U., & Jain, P. (2025). Game-changing intelligence: Unveiling the societal impact of artificial intelligence in game software. Entertainment Computing, 52, 100862. https://doi.org/10.1016/j.entcom.2024.100862

2. Mazurets, O., Uspenska, K., Vit, R., & Tyschenko, O. (2024). Intelligent system for determining the object attributes values by neural networks means by graphic images in databases. In Current trends in the development of scientific research in today's conditions: Proceedings of XXV International Scientific and Practical Conference (pp. 86–91). Florence, Italy: International Scientific Unity.

3. Molchanova M. O., Mazurets O. V., Sobko O. V., Klimenko V. I., Androshchuk V. I. (2024) Metod neiromerezhevoho vyznachennia kyberbulinhy z vykorystanniam khmarnykh servisiv ta ob'iektno-orientovanoi modeli [Neural network method for cyberbullying detection using cloud services and object-oriented model]. Visnyk Khmelnytskoho natsionalnoho universytetu. Seriia: Tekhnichni nauky, № 2 (333), pp. 200–206. https://doi.org/10.31891/2307-5732-2024-333-2

4. Molchanova M. O., Mazurets O. V., Shevchuk P. O., Klimenko V. I., Tyshchenko O. O. (2025) Pidkhid do testuvannia obiektno-orientovanykh system keruvannia v elektronnii komertsii [Approach to testing object-oriented control systems in e-commerce]. Nauka i tekhnika s'ohodni, № 4 (45), pp. 1273–1285. https://doi.org/10.52058/2786-6025-2025-4(45)-1273-1285

5. Mazur Y. V., Mazurets O. V., Klimenko V. I., Sobko O. V., Zalutska O. O. (2025) Alhorytmy ta prohramna arkhitektura informatsiinoi systemy neiromerezhevoho analizu postavy liudyny [Algorithms and software architecture of an information system for neural network posture analysis]. Visnyk Khmelnytskoho natsionalnoho universytetu. Seriia: Tekhnichni nauky, № 3, T. 1, pp. 275–284. https://doi.org/10.31891/2307-5732-2025-351-35

6. Mazurets O. V., Molchanova M. O., Sobko O. V., Didur V. O., Bahriy R. O. (2025) Intehratsiia obiektno-orientovanoho pidkhodu ta reliatsiinykh baz danykh pry proiektuvanni i rozroblenni web-platformy dlia diiachiv mystetstva [Integration of object-oriented approach and relational databases in design and development of a web platform for artists]. Nauka i tekhnika siogodni, № 8 (49), pp. 1542–1555. https://doi.org/10.52058/2786-6025-2025-8(49)-1542-1555

7. Molchanova M., Mazurets O., Sobko O., Boiarchuk I. (2024) Object-Oriented Approach for Ethnic Enmity Detection in Text Messages by NLP. In Proceedings of XXI International Scientific and Practical Conference "Scientific Achievements and Innovations as a Way to Success" (pp. 73–77). Vilnius, Lithuania.

8. Molchanova M. O., Mazurets O. V., Sobko O. V., Vit R. V., Nazarov V. V. (2024) Alhoritm vyviedennia ab'iuzyvnoho vmistu v ukraïnomovnomu audiokontenti dlia implemientatsii v ob'iektno-orijentovanu informatsiinu systemu [Algorithm for detecting abusive content in Ukrainian audio content for implementation in an object-oriented information system]. Visnyk Khmelnytskoho natsionalnoho universytetu: Seriia Tekhnichni nauky, 1(331), 101–106. https://doi.org/10.31891/2307-5732-2024-331-17

9. Molchanova M., Mazurets O., Klimenko V., Kuflevsky Ev. (2024) Object-oriented model for neural network damage detection of mail packages. In Proceedings of XIV International Scientific and Practical Conference "Solving Scientific Problems Using Innovative Concepts" (pp. 58–62). Copenhagen, Denmark.

10. Mazurets O., Sobko O., Vit R., Pasternak V. (2024) Practical approach for detection by deep learning of target objects of subject area based on semantic connectivity indicators in audio database. In Proceedings of XXIV International Scientific and Practical Conference «Modern Scientific Challenges are the Driving Force of the Development of Scientific Research» (pp. 91–96). Bruges, Belgium: International Scientific Unity.

11. Kok I. A., Kadynska V. D., Zalutska O. O., Mazurets O. V. (2025) Object-oriented intelligent system for automated control of smoking by video data. Current scientific goals, approaches and challenges: Proceedings of IV International Scientific and Theoretical Conference, pp. 156–164. Dresden, Federal Republic of Germany.

12. Mazurets O. V., Klimenko V. I., Molchanova M. O., Sultanov A. V. (2024) Object-oriented intelligent system for neural network detection of sugar crystallization zones. Global Science: Prospects and Innovations: Proceedings of the 10th International Scientific and Practical Conference, pp. 198–207. Liverpool, United Kingdom: Cognum Publishing House.

13. Mazurets O. V., Molchanova M. O., Klimenko V. I., Sobko O. V., Suprun P. K. (2024). Datalogichna model′ bazy danykh dlia vyviedennia hendernoi prynalezhnosti za SVM-analizom dopysiv internet-merezh z vykorystanniam obiektno-orientovanoho proiektuvannia [Datalogical model of database for detecting gender based on SVM analysis of internet posts using object-oriented design]. Visnyk Khmelnytskoho Natsionalnoho Universytetu: Seriia Tekhnichni Nauky, 3(2) (337), 197–204. https://doi.org/10.31891/2307-5732-2024-337-3-29

14. Molchanova M., Didur V., Mazurets O., Sobko O., Zakharkevich O. (2025). Method for construction and demolition waste classification using two-factor neural network image analysis. CEUR Workshop Proceedings, 3970, 168–182. https://ceur-ws.org/Vol-3970/PAPER14.pdf

15. Molchanova M., Didur V., Sobko O., Mazurets O. (2025). Detection of web propaganda patterns by transformer neural networks: Improving efficiency via dataset balancing. CEUR Workshop Proceedings, 3988, 112–126. https://ceur-ws.org/Vol-3988/paper10.pdf

16. Zhang, Q., Mao, N., Li, J., Zhou, N., Miao, J., & Li, G. (n.d.). Design of the top-level code framework for Unity based on C. ResearchGate. https://www.researchgate.net/publication/391193129_Design_of_the_Top-Level_Code_Framework_for_Unity_Based_on_C_Language

17. Golob, S. P., & Pagnon Eriksson, N. (n.d.). Using Unity to teach game programming. Malmö University, Game Development Program, Computer Science and Media Technology. https://www.diva-portal.org/smash/get/diva2%3A1867700/FULLTEXT02.pdf

18. Torun, B., Karakurt, S., Aydin, T. B., & Altunel, Y. (2021). Game development on Unity. Turkish Online Journal of Educational Technology-TOJET, 20(1), 39–43. https://eric.ed.gov/?id=EJ1290793

19. Fannoun, Q. K. (2024). Escape the Planet: Revolutionizing game design with novel OOP techniques. https://cornerstone.lib.mnsu.edu/etds/1406/

20. Wu, C. H., Chien, Y. C., Chou, M. T., & Huang, Y. M. (2025). Integrating computational thinking, game design, and design thinking: a scoping review on trends, applications, and implications for education. Humanities and Social Sciences Communications, 12(1), 1–12. https://doi.org/10.1057/s41599-025-04502-x

21. Mazurets, O., Sobko, O., Dydo, R., Zalutska, O., & Molchanova, M. (2025). Augmented reality audiostream creation using CNN: Boosting inclusion and safety for visually impaired people. CEUR Workshop Proceedings, 4004, 347–361. https://ceur-ws.org/Vol-4004/paper26.pdf

Дата першого надходження рукопису до видання: 24.09.2025
Дата прийнятого до друку рукопису після рецензування: 21.10.2025
Дата публікації:

УДК 004.056.53 DOI

І. В. СТОЦЬКИЙ
старший викладач кафедри комп'ютерних інформаційних технологій
Військовий інститут телекомунікацій та інформатизації
імені Героїв Крут
ORCID: 0000-0002-5898-3228

В. П. КОСТЕНКО
старший викладач кафедри комп'ютерних інформаційних технологій
Військовий інститут телекомунікацій та інформатизації
імені Героїв Крут
ORCID: 0000-0002-7180-233X

Г. І. НАЛІСНИЙ
викладач кафедри комп'ютерних інформаційних технологій
Військовий інститут телекомунікацій та інформатизації
імені Героїв Крут
ORCID: 0009-0000-5142-3677

С. А. ШАРНІ
викладач кафедри комп'ютерних інформаційних технологій
Військовий інститут телекомунікацій та інформатизації
імені Героїв Крут
ORCID: 0000-0002-5856-3995

ВИКОРИСТАННЯ АЛГОРИТМІВ AI ДЛЯ ВИЯВЛЕННЯ ТА БЛОКУВАННЯ КІБЕРАТАК НА ВІЙСЬКОВІ ІНФОРМАЦІЙНІ СИСТЕМИ

У сучасному геополітичному контексті, що супроводжується активізацією кіберконфліктів, особливої актуальності набуває проблема забезпечення кібербезпеки військових інформаційних систем. Такі системи є критично важливими для національної оборони, оперативного управління військами та реалізації стратегічних рішень, що робить їх головними цілями як масових, так і таргетованих кібератак. Традиційні підходи до кіберзахисту, орієнтовані на сигнатурний або евристичний аналіз, не можуть забезпечити належний рівень захисту від новітніх загроз, які постійно змінюються, адаптуються та використовують засоби штучного інтелекту.

У цьому контексті використання алгоритмів штучного інтелекту (ШІ, AI), зокрема методів машинного та глибинного навчання, навчання з підкріпленням і обробки природної мови, відкриває нові горизонти для проактивного, адаптивного та автономного кіберзахисту. У статті розглянуто підходи до розробки, впровадження та експлуатації інтелектуальних систем виявлення та блокування атак у військових умовах. Проаналізовано можливості систем виявлення вторгнень нового покоління (Next-Gen IDS), побудованих на базі згорткових нейронних мереж (CNN), які дозволяють виявляти DDoS-атаки, сканування портів та зловмисну активність у реальному часі з високою точністю. Особливу увагу приділено питанням обробки даних з численних джерел у рамках підходу Data Fusion, що критично важливо у бойових умовах, де інформація надходить від сенсорів, безпілотників, супутників тощо.

Окремий акцент зроблено на ролі навчання з підкріпленням у прийнятті рішень в умовах швидкоплинних бойових ситуацій, а також на використанні технологій edge computing, що забезпечують автономність систем у разі втрати зв'язку з центром. Проаналізовано ризики, пов'язані з атаками на самі моделі штучного інтелекту, такі як data poisoning, adversarial input, model inversion, та наведено методи їхньої нейтралізації. Досліджено інтеграцію Explainable AI (XAI) для підвищення довіри до рішень систем, забезпечення прозорості та можливості оперативного контролю.

Показано, що впровадження ШІ у військові інформаційні системи дозволяє реалізувати самонавчальні захисні платформи, які здатні адаптуватися до змін загрозового середовища, генерувати нові сигнатури на основі поточних даних та координувати реагування в межах централізованої командно-інформаційної інфраструктури. У роботі наголошено на важливості міждисциплінарного підходу, залучення фахівців у сфері інформаційної безпеки, військової логістики, розвідки, а також створення цілісної доктрини кібероборони на основі ШІ. Отже, алгоритми ШІ розглядаються не як допоміжний інструмент, а як фундамент майбутньої архітектури стійкого, гнучкого та інтелектуального захисту військових IT-систем.

Ключові слова: *штучний інтелект (ШІ), кібербезпека, військові інформаційні системи, системи виявлення вторгнень (IDS), машинне навчання, кібератаки, навчання з підкріпленням, кібергігієна користувачів.*

© Стоцький І. В., Костенко В. П., Налісний Г. І., Шарні С. А., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

I. V. STOTSKYI
Senior Lecturer at the Department of Computer Information Technologies
Geroyev Krut Military Institute of Telecommunications and Informatization
ORCID: 0000-0002-5898-3228

V. P. KOSTENKO
Senior Lecturer at the Department of Computer Information Technologies
Geroyev Krut Military Institute of Telecommunications and Informatization
ORCID: 0000-0002-7180-233X

G. I. NALISNYI
Lecturer at the Department of Computer Information Technologies
Geroyev Krut Military Institute of Telecommunications and Informatization
ORCID: 0009-0000-5142-3677

S. A. SHARNIN
Lecturer at the Department of Computer Information Technologies
Geroyev Krut Military Institute of Telecommunications and Informatization
ORCID: 0000-0002-5856-3995

USE OF AI ALGORITHMS FOR DETECTION AND BLOCKING OF CYBER ATTACKS ON MILITARY INFORMATION SYSTEMS

In the modern geopolitical context, accompanied by the intensification of cyber conflicts, the problem of ensuring the cybersecurity of military information systems is of particular relevance. Such systems are critically important for national defense, operational command of troops and the implementation of strategic decisions, which makes them the main targets of both mass and targeted cyberattacks. Traditional approaches to cyberdefense, focused on signature or heuristic analysis, cannot provide an adequate level of protection against the latest threats that are constantly changing, adapting and using artificial intelligence tools. In this context, the use of artificial intelligence (AI) algorithms, in particular machine and deep learning methods, reinforcement learning and natural language processing, opens up new horizons for proactive, adaptive and autonomous cyberdefense. The article considers approaches to the development, implementation and operation of intelligent systems for detecting and blocking attacks in military conditions. The capabilities of Next-Gen IDS systems built on the basis of convolutional neural networks (CNN) are analyzed, which allow detecting DDoS attacks, port scanning and malicious activity in real time with high accuracy. Special attention is paid to the issues of processing data from multiple sources within the Data Fusion approach, which is critically important in combat conditions, where information comes from sensors, drones, satellites, etc.

A special emphasis is placed on the role of reinforcement learning in decision-making in fast-paced combat situations, as well as on the use of edge computing technologies that ensure the autonomy of systems in case of loss of communication with the center. The risks associated with attacks on artificial intelligence models themselves, such as data poisoning, adversarial input, model inversion, are analyzed, and methods for their neutralization are presented. The integration of Explainable AI (XAI) is investigated to increase trust in system decisions, ensure transparency and the possibility of operational control.

It is shown that the implementation of AI in military information systems allows for the implementation of self-learning defense platforms that are able to adapt to changes in the threat environment, generate new signatures based on current data, and coordinate responses within a centralized command and information infrastructure. The paper emphasizes the importance of an interdisciplinary approach, involving specialists in the field of information security, military logistics, and intelligence, as well as the creation of a holistic cyber defense doctrine based on AI. Therefore, AI algorithms are considered not as an auxiliary tool, but as the foundation of the future architecture of sustainable, flexible, and intelligent defense of military IT systems.

Key words: *artificial intelligence (AI), cybersecurity, military information systems, intrusion detection systems (IDS), machine learning, cyber attacks, reinforcement learning, user cyber hygiene.*

Постановка проблеми

У сучасному цифровому середовищі військові інформаційні системи відіграють ключову роль у забезпеченні національної безпеки, оперативного управління військами та здійсненні стратегічного планування. З огляду на глобальну тенденцію до цифровізації та мережевої взаємодії, дані системи є об’єктом інтенсивних кібератак, які можуть призводити до фатальних наслідків для національної безпеки, порушення логістичних ланцюгів, втрати критичних даних і навіть виведення з ладу державні оборонні системи [1].

Класичні методи кіберзахисту, які ґрунтуються на сигнатурному аналізі, не здатні ефективно протистояти складним, багаторівневим і динамічним загрозам, що постійно змінюються. На даний час, інструменти кібератак

використовують соціальну інженерію, шкідливе програмне забезпечення (ПЗ) із функціями штучного інтелекту, автоматизоване сканування вразливостей, а також механізми ухилення від їх виявлення. У зв'язку з цим, виникає потреба у впровадженні адаптивних, інтелектуальних систем захисту, здатних в режимі реального часу аналізувати великий обсяг даних, виявляти аномалії та блокувати потенційні загрози.

Алгоритми штучного інтелекту (ШІ, AI), зокрема машинне навчання, глибинне навчання та обробка природної мови (NLP), мають потенціал стати основою нової парадигми кібербезпеки. Їх застосування у військовій сфері дозволяє створювати системи проактивного захисту, які не лише реагують на атаки, а й передбачають можливі сценарії загроз. Проте, попри потенціал, існують значні виклики щодо інтеграції AI у військові системи, зокрема питання надійності, етичності, прозорості та захисту самих моделей AI від маніпуляцій.

Проблема ускладнюється необхідністю дотримання високого рівня конфіденційності та секретності інформації у військовому секторі. Будь-яке рішення повинне відповідати суворим вимогам до безпеки, масштабованості, автономності та здатності функціонувати в умовах обмежених ресурсів (наприклад, у польових умовах або під час кібервійни) [2, 3].

Таким чином, актуальність теми застосування AI для виявлення та блокування кібератак на військові інформаційні системи зумовлена як зростанням кількості кібератак, так і необхідністю підвищення ефективності засобів кіберзахисту за рахунок інтелектуалізації процесів виявлення загроз. При цьому, загрозлива динаміка розвитку кіберзлочинності, зокрема державного кібертероризму та шкідливих інформаційних операцій, потребує не лише модернізації існуючих рішень, а й створення нових, якісно інших систем захисту. Військові інформаційні системи повинні не просто бути захищеними, а мати можливість до самонавчання, адаптації до нових умов і активного протистояння новітнім формам атак, що можливо реалізувати лише за умов глибокої інтеграції AI-технологій у архітектуру систем кібербезпеки.

Аналіз останніх досліджень і публікацій

Упродовж останнього десятиліття у наукових спільнотах зросла увага до використання штучного інтелекту для забезпечення кібербезпеки, зокрема у військовому контексті. Значна кількість досліджень підтверджує ефективність алгоритмів машинного навчання у виявленні аномалій, класифікації мережевого трафіку та ідентифікації шкідливого ПЗ.

Так, у дослідженнях MIT Lincoln Laboratory та DARPA підкреслюється ефективність глибоких нейронних мереж у виявленні раніше невідомих типів атак, таких як zero-day exploits. Розробка систем на базі LSTM (довготривала короткочасна пам'ять) дозволила досягти високої точності при аналізі мережевого трафіку та передбаченні потенційно небезпечних активностей. У рамках проєкту MADL (Machine Learning for Advanced Defense Logistics) були протестовані моделі, здатні виявляти складні кореляції між різнорідними джерелами даних у реальному часі [4].

У європейських програмах кібербезпеки, таких як ECHO та SPARTA, значну увагу приділено інтеграції AI з класичними системами моніторингу. Наприклад, застосування гібридних моделей – поєднання машинного навчання з логічним аналізом – забезпечує не лише виявлення загроз, а й пояснюваність рішень, що особливо важливо у військовому контексті. Розроблено спеціалізовані платформи, які дозволяють здійснювати багаторівневу ієрархічну оцінку ризиків на основі навчання з підкріпленням (reinforcement learning).

Публікації останніх років, зокрема в IEEE Access, Journal of Cybersecurity, ACM Transactions on Privacy and Security, демонструють успішне впровадження AI в системи IDS/IPS (інструменти виявлення та запобігання вторгнень). Сучасні системи типу «intelligent IDS» поєднують в собі класичні правила, евристичний аналіз та глибоке навчання для обробки великого обсягу даних з низькою затримкою. Ці технології уже застосовуються у пілотних проектах для НАТО та в арміях провідних країн, таких як США, Ізраїль та Велика Британія [5].

У межах публікацій щодо специфіки військових інформаційних систем підкреслюється, що моделі AI повинні враховувати динаміку середовища, мобільність вузлів, змінні умови доступу та складність топології мереж. Наприклад, у дослідженнях RAND Corporation та NATO CCDCOE досліджуються аспекти застосування адаптивних AI-алгоритмів у середовищі без стабільного підключення до мережі, що характерно для бойових умов [6].

Значну увагу також приділено питанням атак на самі моделі AI. Такі загрози, як отруєння даних (data poisoning), атаки на цілісність моделі та зворотне проектування (model inversion) розглядаються як серйозні виклики, що потребують побудови захищених архітектур машинного навчання. Пропонуються механізми валідації навчальних даних, використання генеративних моделей для перевірки цілісності та побудова автономних верифікаторів рішень моделі.

Підсумовуючи, можна зазначити, що наукове середовище активно досліджує потенціал AI у військовій кібербезпеці. Водночас існує потреба у стандартизації підходів, визначенні протоколів взаємодії та забезпеченні прозорості у використанні таких рішень, особливо в умовах критичної інфраструктури. Тенденції свідчать про швидкий розвиток міжпредметних досліджень на перетині кібербезпеки, AI, військової логістики та стратегічного аналізу загроз.

Формулювання мети дослідження

Мета статті – дослідження алгоритмів штучного інтелекту для виявлення, запобігання та нейтралізації кібератак на військові інформаційні системи з урахуванням сучасних технологічних рішень, особливостей військового середовища та вимог до кіберстійкості.

Викладення основного матеріалу дослідження

Практична реалізація алгоритмів штучного інтелекту для виявлення та блокування кібератак на військові інформаційні системи полягає у використанні комплексного підходу, що включає розробку, тестування, впровадження та експлуатацію технологій на базі машинного навчання, обробки великих даних, обчислень на краю мережі (edge computing) та кіберфізичних систем. Військовий контекст висуває особливі вимоги до стійкості систем, швидкості реакції, автономності рішень та здатності до самонавчання.

Однією з ключових технологій є системи виявлення вторгнень нового покоління (Next-Gen IDS), які об'єднують алгоритми глибокого навчання (Deep Learning) з класичними методами аналізу мережевого трафіку. Такі системи здатні не лише фіксувати аномальні шаблони поведінки, але й класифікувати їх за типами загроз. Наприклад, використання згорткових нейронних мереж (CNN) для аналізу трафіку в реальному часі дозволяє виявити DDoS-атаки, сканування портів, зловмисні пакети та активність ботнетів із точністю понад 95 %.

Іншим напрямком є обробка телеметричних даних з багатьох джерел одночасно – так званий підхід Data Fusion. Військові системи часто працюють у середовищі, де доступна інформація з сенсорів, супутників, систем командного зв'язку та мобільних пристроїв. Застосування моделей кластеризації (наприклад, алгоритмів K-means або DBSCAN) дозволяє інтегрувати ці дані, виявляти відхилення, створювати профілі звичайної поведінки системи і оперативно виявляти невідповідності.

Важливою складовою системи захисту є механізми прийняття рішень на основі навчання з підкріпленням (Reinforcement Learning). У військових умовах, коли зміна параметрів середовища відбувається дуже швидко, агенти RL можуть навчатися ефективним стратегіям відповіді на загрози, вибираючи між ізоляцією вузла, реконфігурацією мережі, зміною маршрутів передачі даних або активацією резервних систем. Наприклад, алгоритм DQN (Deep Q Network) застосовується для моделювання сценаріїв атаки та формування оптимальних дій у відповідь [5].

Крім цього, системи кіберзахисту з ШІ повинні містити модулі самонавчання на основі аналізу нових даних про загрози, отриманих з відкритих джерел (OSINT), військової розвідки та партнерських мереж безпеки. Нейронні мережі типу Transformer (наприклад, BERT, GPT) використовуються для обробки текстових повідомлень про нові загрози, автоматичного створення сигнатур та формування попереджень у системах реагування.

Успішна імплементація AI вимагає створення спеціального циклу DevSecOps, адаптованого до потреб військових операцій. Такий цикл має включати: автоматизоване тестування моделей на стійкість до атак; перевірку навчальних даних на чистоту; впровадження механізмів Explainable AI (XAI) для пояснення рішень; резервування критичних елементів та впровадження механізмів відновлення після атак.

Реальне розгортання таких систем можливе завдяки використанню edge-computing – систем розподіленого аналізу даних на місці збору (на пристроях, дронах, сенсорах). Це забезпечує мінімальну затримку в обробці інформації та дозволяє знизити залежність від каналів зв'язку, які можуть бути зруйновані під час бойових дій. Наприклад, інтеграція моделі CNN у військовий дрон дозволяє розпізнавати підозрілу активність у зоні бойових дій без потреби передавати відео в центр обробки.

Значну роль у блокуванні атак відіграють методи динамічної реконфігурації архітектури військової мережі. За допомогою предиктивного аналізу на основі рекурентних нейронних мереж (RNN) система може прогнозувати, яка частина інфраструктури найімовірніше стане об'єктом атаки, та автоматично переналаштовувати маршрути даних, активувати додаткове шифрування або виводити з експлуатації вразливі вузли.

Питання кібергігієни користувачів військових систем також вирішуються за допомогою AI. Алгоритми поведінкового аналізу створюють індивідуальні профілі користувачів і виявляють дії, нехарактерні для конкретного оператора (наприклад, доступ до незвичних файлів, запити в незвичний час, введення команд із нових пристроїв), що може свідчити про компрометацію облікового запису.

Особливе значення має захист самих моделей AI. Військові системи мають бути захищені від атак типу adversarial input, які змінюють вхідні дані з метою викликати помилку в рішенні моделі. Для цього використовуються методи захисту, такі як distillation, adversarial training та використання додаткових мереж-детекторів, що виявляють підозрілі входи. Дослідження демонструють, що інтеграція таких механізмів дозволяє знизити ймовірність маніпуляції на понад 80 % [6].

У контексті оперативного управління всі ці елементи мають бути інтегровані в єдину командно-інформаційну систему. Наприклад, розробка центру кібероперацій з використанням централізованої платформи, яка акумулює дані, проводить аналіз, генерує рекомендації та керує контрзаходами в автоматичному або напівавтоматичному режимі.

Виходячи з цього, можна констатувати значний практичний потенціал використання AI в системах кіберзахисту військових структур. Однак ефективність впровадження таких рішень залежить від узгодженості між технічними фахівцями, командуванням, аналітиками розвідки та політичним керівництвом. Тільки інтегрований підхід до розробки та використання AI в умовах війни забезпечить необхідний рівень стійкості до кібератак та інформаційного терору.

Висновки

Застосування алгоритмів штучного інтелекту для виявлення та блокування кібератак на військові інформаційні системи є стратегічно важливим напрямом розвитку оборонних технологій. Перш за все, алгоритми AI дозволяють значно

підвищити точність і швидкість виявлення кіберзагроз. Традиційні системи, що ґрунтуються на сигнатурах, виявляють лише відомі типи атак, у той час як AI здатен виявляти нові та модифіковані форми загроз шляхом аналізу поведінкових патернів, аномалій та кореляцій у великих обсягах даних. Завдяки використанню методів глибокого навчання, кластеризації, підкріплення та трансформерів, військові системи отримують інструменти проактивного захисту.

Крім цього, системи, підкріплені AI, виявляють гнучкість і адаптивність до змінного середовища бойових дій. Алгоритми навчання з підкріпленням можуть обирати оптимальні стратегії реагування, забезпечуючи ефективну ізоляцію, реконфігурацію чи протидію загрозам в режимі реального часу. Це критично важливо в умовах обмежених ресурсів і високої динаміки подій.

Також, важливим досягненням є застосування технологій edge-computing, які дозволяють обробляти дані безпосередньо на місці подій, зменшуючи залежність від централізованих ресурсів та підвищуючи автономність систем. Зокрема, автономні дрони, сенсори, командно-штабні комплекси можуть функціонувати незалежно від центра зв'язку, що робить систему менш вразливою до атак на інфраструктуру.

Вцілому, інтеграція AI в системи кіберзахисту дозволяє реалізовувати принципи самоадаптації та безперервного навчання. Це забезпечує постійне оновлення знань про загрози, автоматичну побудову моделей нових атак та формування ефективних протидій без прямого втручання людини. Таким чином досягається висока стійкість до тактик «0-day» атак.

Значним фактором є зростання значення Explainable AI (XAI) – пояснюваного штучного інтелекту, який дає змогу операторам та командуванню розуміти логіку дій системи та приймати обґрунтовані рішення. У військовому контексті прозорість алгоритмів AI є надзвичайно важливою для збереження контролю та запобігання фатальним помилкам. Тому критично важливим є впровадження багаторівневих механізмів захисту, регулярне тестування, а також побудова архітектур із надлишковими елементами та можливістю швидкого відновлення.

У контексті даної роботи, можна зазначити, що ефективне використання AI в кібербезпеці військових структур можливе лише за наявності спеціалістів, які володіють знаннями як у сфері інформаційної безпеки, так і в алгоритмах машинного навчання. Необхідна системна робота з підготовки офіцерів-кібернетиків, аналітиків загроз, інженерів із впровадження та адміністрування таких систем.

Підсумовуючи вищезазначене, слід відзначити, що алгоритми штучного інтелекту є не лише допоміжним інструментом, а й основою майбутніх платформ кібероборони. Їх впровадження у військові інформаційні системи – це вже не питання вибору, а питання національної безпеки. Надалі необхідно зосередитися на створенні адаптивних, стійких, прозорих і здатних до самонавчання кіберсистем, які зможуть забезпечити ефективну протидію як традиційним, так і новітнім гібридним загрозам.

Список використаної літератури

1. Гулак Г. М. Методологія захисту інформації. Аспекти кібербезпеки: підручник. Київ : Видавництво НА СБ України, 2021. 256 с.
2. Даник Ю. Г., Воробієнко П. П., Чернега В. М. Основи кібербезпеки та кібероборони: підручник. Одеса : ОНАЗ ім. О. С. Попова, 2019. 320 с.
3. Толюпа С. В., Штаненко С. С., Берестовенко Г. Класифікаційні ознаки систем виявлення атак та напрямки їх побудови. Збірник наукових праць Військового інституту телекомунікацій та інформатизації імені Героїв Крут. 2018. Вип. № 3. С. 56–66.
4. Solomon M. Building a Secure & Privacy-Focused IoT Network. Cambridge: MIT Press, 2021. 370 p.
5. Stallings W. Cryptography and Network Security: Principles and Practice. Pearson, 2020. 840 p.
6. Zhou H., Liu C. Deep Learning in Cybersecurity. San Francisco : Elsevier, 2020. 242 p.

References

1. Hulak, H. M. Metodolohiia zakhystu informatsii. (2021). Aspekty kiberbezpeky: pidruchnyk. Kyiv : Vydavnytstvo NA SB Ukrainy, 256 s. [in Ukrainian].
2. Danyk, Yu. H., Vorobiienko, P. P., Cherneha, V. M. (2019). Osnovy kiberbezpeky ta kiberoborony: pidruchnyk. Odesa : ONAZ im. O. S. Popova, 320 s. [in Ukrainian].
3. Toliupa, S. V., Shtanenko, S. S., Berestovenko, H. (2018). Klasyfikatsiini oznaky system vyiavlennia atak ta napriamky yikh pobudovy. Zbirnyk naukovykh prats Viiskovoho instytutu telekomunikatsii ta informatyzatsii imeni Heroiv Krut, Vyp. № 3. S. 56–66.[in Ukrainian].
4. Solomon M. (2021) Building a Secure & Privacy-Focused IoT Network. Cambridge: MIT Press, 370 p.
5. Stallings W. (2020) Cryptography and Network Security: Principles and Practice. Pearson, 840 p.
6. Zhou H., Liu C. (2020) Deep Learning in Cybersecurity. – San Francisco: Elsevier,.242 p.

Дата першого надходження рукопису до видання: 28.09.2025
Дата прийнятого до друку рукопису після рецензування: 24.10.2025
Дата публікації:

УДК 004.932:004.8 DOI

О. М. СУПРУН
кандидат фізико-математичних наук, доцент,
доцент кафедри інтелектуальних кібернетичних систем
Державний університет «Київський авіаційний інститут»
ORCID: 0000-0002-1196-5655

Н. О. ЧІКІНА
кандидат технічних наук, доцент,
професор кафедри вищої математики
Інститут механічної інженерії і транспорту
Національного технічного університету
«Харківський політехнічний інститут»
ORCID: 0000-0002-0643-1771

Я. Я. КРАВЧУК
веб-розробник
ORCID: 0009-0006-1593-9995

ІНТЕЛЕКТУАЛЬНІ МЕТОДИ АНАЛІЗУ ВІДЕОПОТОКУ ДЛЯ ВИЯВЛЕННЯ ПРИХОВАНИХ ПОВІДОМЛЕНЬ

Актуальність дослідження зумовлена стрімким зростанням обсягів відеоданих та ускладненням методів стеганографії, що становить значні ризики для інформаційної безпеки та робить неефективним використання традиційних засобів контролю. Традиційні алгоритми цифрової обробки сигналів не забезпечують достатнього рівня точності та стійкості у виявленні прихованих повідомлень, особливо за наявності шумових перешкод та динамічно змінюваних потоків даних. Це зумовлює потребу впроваджувати інтелектуальні підходи, які поєднують методи комп'ютерного зору, машинного навчання та мультимодального аналізу.

Мета статті – обґрунтувати та розробити інтелектуальні методи аналізу відеопотоку для виявлення прихованих повідомлень, що має забезпечити підвищення захищеності інформаційного середовища та ефективність сучасних алгоритмів штучного інтелекту за умов зростання кіберзагроз.

Методологія дослідження ґрунтується на поєднанні системного аналізу відеопотоку, методів машинного навчання, глибоких нейронних мереж та мультимодальних підходів до обробки даних. Використано принципи інтеграції алгоритмів комп'ютерного зору з класифікаційними моделями, зокрема згортковими та рекурентними мережами, а також механізмами самоуваги трансформерів. Застосовано порівняльний аналіз ефективності різних архітектур, моделювання роботи систем у режимі реального часу та оцінку їхньої стійкості до маніпуляцій.

Результати дослідження полягають у визначенні придатності традиційних та інтелектуальних методів для завдань ідентифікації прихованих сигналів. Доведено, що глибоке навчання та мультимодальні системи забезпечують значно вищу точність і адаптивність порівняно з традиційними підходами. Встановлено, що інтегровані архітектури аналізують просторові, часові й контекстуальні характеристики відео, створюють передумови для підвищення ефективності виявлення прихованих повідомлень у кіберзахисті, прикордонному контролі та моніторингу критичної інфраструктури.

Висновки підтверджують, що застосування інтелектуальних методів аналізу відеопотоку дає змогу долати обмеження традиційних рішень та істотно зміцнює інформаційну безпеку. Водночас виявлено основні проблеми впровадження: значні обчислювальні витрати, дефіцит анотованих мультимодальних даних, низька інтерпретованість результатів і вразливість до adversarial-атак.

Перспективи подальших досліджень пов'язані з розробленням адаптивних архітектур на основі continual learning і federated learning, оптимізацією моделей для енергоефективного функціонування, упровадженням explainable AI для підвищення прозорості роботи алгоритмів, а також зі створенням уніфікованих протоколів інтеграції з іншими засобами кіберзахисту. Очікується, що ці рішення сформують нове покоління інтелектуальних систем відеоаналізу, здатних ефективно реагувати на сучасні та майбутні інформаційні загрози.

Ключові слова: *стеганографія, кіберзахист, комп'ютерний зір, мультимодальний аналіз, глибоке навчання, штучний інтелект, інформаційна безпека.*

© Супрун О. М., Чікіна Н. О., Кравчук Я. Я., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

O. M. SUPRUN
Candidate of Physical and Mathematical Sciences, Associate Professor,
Associate Professor at the Department of Intelligent Cybernetic Systems
State University "Kyiv Aviation Institute"
ORCID: 0000-0002-1196-5655

N. O. CHIKINA
Candidate of Technical Sciences, Associate Professor,
Professor at the Department of Higher Mathematics
Institute of Education and Science in Mechanical Engineering and Transport
of National Technical University "Kharkiv Polytechnic Institute"
ORCID: 0000-0002-0643-1771

YA. YA. KRAVCHUK
Web Developer
ORCID: 0009-0006-1593-9995

INTELLIGENT METHODS OF VIDEO STREAM ANALYSIS FOR DETECTING HIDDEN MESSAGES

The relevance of this study is determined by the rapid growth of video data volumes and the increasing sophistication of steganographic methods, which pose significant risks to information security and undermine the effectiveness of traditional monitoring tools. Conventional digital signal processing algorithms fail to provide sufficient accuracy and robustness in detecting hidden messages, particularly in the presence of noise interference and dynamically changing data streams. This necessitates the implementation of intelligent approaches that integrate methods of computer vision, machine learning, and multimodal analysis.

The purpose of the article is to substantiate and develop intelligent methods of video stream analysis for detecting hidden messages, thereby enhancing the protection of the information environment and improving the efficiency of modern artificial intelligence algorithms under growing cyber threats.

The research methodology is based on the integration of system analysis of video streams, machine learning techniques, deep neural networks, and multimodal data processing approaches. It applies principles of combining computer vision algorithms with classification models, including convolutional and recurrent networks, as well as transformer-based attention mechanisms. A comparative analysis of the effectiveness of different architectures was carried out, along with real-time system modeling and assessment of their resistance to manipulations.

The results demonstrate the applicability of both traditional and intelligent methods for identifying hidden signals. The study confirms that deep learning and multimodal systems provide significantly higher accuracy and adaptability compared to traditional approaches. It was established that integrated architectures analyze spatial, temporal, and contextual video characteristics, creating prerequisites for improving the detection of hidden messages in cybersecurity, border control, and critical infrastructure monitoring.

The conclusions confirm that the application of intelligent video stream analysis methods overcomes the limitations of traditional solutions and significantly strengthens information security. At the same time, key implementation challenges were identified, including high computational costs, a shortage of annotated multimodal data, low interpretability of results, and vulnerability to adversarial attacks.

Future research prospects are linked to the development of adaptive architectures based on continual learning and federated learning, model optimization for energy-efficient functioning, the introduction of explainable AI to increase algorithm transparency, and the creation of unified protocols for integration with other cybersecurity tools. These advancements are expected to shape a new generation of intelligent video analysis systems capable of effectively responding to current and future information threats.

Key words: *steganography, cybersecurity, computer vision, multimodal analysis, deep learning, artificial intelligence, information security.*

Постановка проблеми

Проблема виявлення прихованих повідомлень у відеопотоці є однією з головних у сучасних дослідженнях з інформаційної безпеки та мультимедійної аналітики, оскільки традиційні методи контролю виявляються малоефективними в умовах зростання обсягів даних та ускладнення технік стеганографії. Інтелектуальні підходи поєднують можливості комп'ютерного зору, машинного навчання та нейронних мереж для ідентифікації прихованих патернів, які не піддаються традиційному аналізу. Такий підхід має значний науковий потенціал, адже орієнтований на розроблення нових алгоритмів розпізнавання прихованих структур у відео, що розширює межі сучасних теорій цифрової обробки сигналів та штучного інтелекту. На практиці це завдання має особливе значення для захисту національної безпеки та критичної інфраструктури, а також для ефективної протидії кіберзлочинності й інформаційним диверсіям. Розроблення ефективних інтелектуальних методів аналізу відеопотоку

здатне підвищити рівень захищеності цифрового середовища. Це вимагає створення інструментів оперативного моніторингу й запровадження технологій превентивного виявлення загроз, що безпосередньо пов'язує цю проблему з актуальними науковими пошуками й стратегічними практичними завданнями.

Аналіз останніх досліджень і публікацій

У наукових розвідках, присвячених інтелектуальним методам аналізу відеопотоку для виявлення прихованих повідомлень, можна виокремити чотири взаємопов'язані напрями. Перший напрям стосується розвитку захисних архітектур і застосування алгоритмів комп'ютерного зору для ідентифікації прихованих патернів. Ю. Горбенко (Y. Horbenko) довів, що застосування конфіденційних обчислень, зокрема захищених енклавів і гомоморфного шифрування, підвищує безпеку інтелектуальних систем аналізу відеопотоку [1]. С. Вань (S. Wan), С. Сюй (X. Xu), Т. Ван (T. Wang) та співавтори представили методи інтелектуального відеоаналізу для виявлення аномальних подій у транспортних системах, які підтверджують ефективність глибинних моделей у реальному часі [2]. Н. Меганатхан (N. Meghanathan), Л. Наяк (L. Nayak) узагальнили алгоритми стеганоаналізу для зображень, аудіо та відео, заклавши методологічну основу для подальших досліджень у сфері виявлення прихованих повідомлень [3]. М. Далал (M. Dalal), М. Джунеджа (M. Juneja) узагальнили методи відеостеганографії – від простих субституцій до адаптивних стратегій на основі машинного навчання [4]. Подальші дослідження необхідно зосередити на інтеграції глибинних моделей та захищених архітектур обчислень для оптимального поєднання продуктивності й безпеки.

Другий напрям пов'язаний із цифровим водяним знакуванням та захистом інформаційних потоків від атак. П. Аберна (P. Aberna), Л. Агіландесварі (L. Agilandeeswari) узагальнили методології цифрового фото- та відеоводяного знакування, виокремивши основні вектори атак і перспективи інтеграції з технологіями стеганографії [5]. На основі аналізу загроз Dark Web С. Назах (S. Nazah), С. Худа (S. Huda), Дж. Абаваджи (J. Abawajy) та співавтори дійшли висновку, що проблеми прихованих каналів посідають головне місце в сучасних стратегіях кіберзахисту [6]. Л. Сонг (L. Song), К. Лікопп (C. Licoppe) у своєму аналізі закриття відеострімів продемонстрували, що послідовність дій у таких потоках може нести ознаки прихованих комунікаційних сигналів [7]. С. Дхаван (S. Dhawan), Р. Гупта (R. Gupta) узагальнили методи стеганографії як інструменту захисту інформаційної безпеки, акцентуючи на порівняльних характеристиках різних підходів [8]. Подальші дослідження мають бути спрямовані на розроблення гібридних моделей, що поєднують watermarking, стеганографію та аналіз поведінкових патернів відеопотоків.

Третій напрям зосереджується на стійкості протоколів комунікацій і загрозах, пов'язаних з інтеграцією стеганографічних алгоритмів у сучасні системи обміну даними. Р. Хуан (R. Huang), Х. Ву (H. Wu), С. Ван (X. Wang) та співавтори довели, що сучасні протоколи миттєвого відеообміну є потенційним каналом прихованих атак, що істотно ускладнює завдання їхнього виявлення [9]. М. Далал (M. Dalal), М. Джунеджа (M. Juneja) у своїй пізнішій роботі систематизували принципи стеганографії та стеганоаналізу в контексті цифрової криміналістики, що відкриває нові перспективи для застосування таких методів у розслідуваннях кіберзлочинів [10]. Н. Субраманіан (N. Subramanian), О. Ельхаррусс (O. Elharrouss), С. Аль-Маадид (S. Al-Maadeed) та співавтори дослідили сучасні досягнення image steganography, виокремивши потенціал глибинних нейромереж у підвищенні стійкості алгоритмів [11]. О. Ф. А. Вагаб (O. F. A. Wahab), А. А. М. Халаф (A. A. M. Khalaf) та співавтори запропонували інтегровану схему приховування даних на основі поєднання RSA-шифрування та компресійної стеганографії, що підвищує рівень захисту [12]. Подальші дослідження в цьому напрямі мають зосереджуватись на формуванні протоколів стійкої комунікації, які інтегрують методи приховування даних із системами захисту від атак.

Четвертий напрям присвячений мультимодальним підходам до обробки та відновлення прихованої інформації в зашифрованих середовищах. З. Сунь (Z. Sun), П. Сарма (P. Sarma), В. Сетарес (W. Sethares) та співавтори показали ефективність мультимодальних методів, які дають змогу поєднувати аналіз текстових, аудіо- та відеопотоків для виявлення прихованих залежностей [13]. Ц. Ю (C. Yu), С. Чжан (X. Zhang) та співавтори розробили метод оборотного приховування даних у зашифрованих зображеннях із багаторівневим ієрархічним вбудовуванням, що сприяє контрольованому збереженню й відновленню прихованої інформації [14]. Цей напрям акцентує на важливості комплексного поєднання мультимодального аналізу й стеганографії у створенні систем, здатних не лише виявляти приховані повідомлення, а й відновлювати їх зі збереженням цілісності даних. Подальші дослідження у цій сфері мають орієнтуватися на стандартизацію мультимодальних протоколів і створення універсальних схем для відновлення даних у зашифрованих відеопотоках.

Попри значні досягнення в інтелектуальному аналізі відеопотоку, низка проблем досі нерозв'язана. Сучасні методи обробки даних демонструють обмежену стійкість до зашумлених та зашифрованих потоків, а також недостатню здатність адаптуватися до нових методів стеганографії. Відчутною проблемою є високі обчислювальні витрати та дефіцит якісних мультимодальних наборів даних, що ускладнює тренування універсальних моделей. Також зберігаються методологічні обмеження, пов'язані з низькою інтерпретованістю алгоритмів і складністю практичної інтеграції в реальні системи кіберзахисту.

У дослідженні запропоновано подолати ці прогалини шляхом поєднання алгоритмів комп'ютерного зору, машинного навчання та мультимодального аналізу для створення більш точних і адаптивних моделей.

Застосування гібридних архітектур, принципів continual та federated learning, а також оптимізованих методів глибинного навчання дасть змогу підвищити стійкість систем і знизити обчислювальні витрати.

Формулювання мети дослідження

Мета статті – обґрунтування та розробка інтелектуальних методів аналізу відеопотоку для виявлення прихованих повідомлень, що забезпечить підвищення рівня інформаційної безпеки й ефективність застосування технологій комп'ютерного зору та штучного інтелекту в умовах сучасних кіберзагроз.

Завдання статті:

1. Дослідити сучасні методи обробки відеопотоку та інтеграцію комп'ютерного зору з машинним навчанням для підвищення точності виявлення прихованої інформації.
2. Проаналізувати потенціал мультимодальних даних для ідентифікації прихованих повідомлень та визначити основні слабкі місця інтелектуальних методів у цій задачі.
3. Сформулювати практичні рекомендації щодо вдосконалення систем аналізу відеопотоку для зміцнення інформаційної безпеки та ефективної протидії кіберзагрозам.

Викладення основного матеріалу дослідження

Аналіз відеопотоку в сучасних умовах – це один із найдинамічніших напрямів розвитку інформаційних технологій, що поєднує комп'ютерний зір, обробку сигналів та алгоритми штучного інтелекту. Методи, що застосовуються для аналізу відео, різняться за рівнем складності та завданнями: від базової детекції руху й класифікації об'єктів до виявлення прихованих структур та стеганографічних повідомлень. Придатність конкретних методів визначається їхньою здатністю працювати в режимі реального часу, стійкістю до завад, масштабованістю та адаптивністю до нових форм приховування даних. З огляду на це доцільним є узагальнення найпоширеніших підходів, які застосовуються в сучасних системах аналізу відеопотоку (табл. 1).

Таблиця 1

Сучасні методи обробки відеопотоку та їхня придатність для виявлення прихованих повідомлень

Метод	Основні характеристики	Сфера застосування	Практичний результат
Традиційні алгоритми цифрової обробки зображень (фільтрація, сегментація, виділення контурів)	Формальні математичні процедури для роботи з кадрами відео	Системи первинного аналізу відео, попередня обробка даних	Швидке виявлення аномалій і базових відхилень у структурі зображення
Машинне навчання (класифікація, метод опорних векторів – Support Vector Machine, дерева рішень)	Навчання моделей на маркованих наборах даних із прикладами прихованої інформації	Системи автоматичного моніторингу та аналізу мультимедійного контенту	Покращення точності розпізнавання прихованих патернів порівняно з класичними методами
Глибоке навчання (Convolutional Neural Networks – CNN, Recurrent Neural Networks – RNN)	Автоматичне виділення ознак, виявлення складних і багаторівневих структур у відеопотоці	Високонавантажені системи кіберзахисту, аналітичні центри з великими масивами даних	Виявлення прихованих повідомлень навіть у зашумленому або зміненому відео, підвищена адаптивність
Мультимодальний аналіз (поєднання відео, звуку, метаданих)	Інтеграція різних типів даних у єдину модель	Системи кіберрозвідки, прикордонний та митний контроль, інформаційна безпека	Комплексне виявлення прихованої інформації, стійкість до маніпуляцій у різних середовищах

Джерело: сформовано авторами на основі [2, p. 4488–4489; 3, p. 45–46; 4, p. 5836–5837; 8, p. 238; 15, p. 4]

Сьогодні ці методи дедалі частіше інтегруються в багаторівневі системи аналізу відеопотоку, що дає змогу переходити від елементарних алгоритмів виявлення аномалій до складних інтелектуальних моделей, здатних виявляти приховані сигнали навіть у високодинамічному середовищі. Традиційні методи цифрової обробки зображень застосовуються як перший етап попередньої фільтрації даних, що дає змогу знизити обчислювальне навантаження на наступні рівні аналізу. Машинне навчання інтегрується в системи моніторингу для автоматичного виявлення закономірностей, наприклад, під час перевірки відеоконтенту на наявність стеганографічних вставок у прикордонному чи митному контролі [8, с. 237–238]. Глибокі нейронні мережі ефективно виявляють приховані повідомлення, замасковані зміною текстури чи кольорової палітри відео. Такі рішення вже використовують у центрах кібербезпеки для протидії диверсіям у медіапросторі. Важливу роль відіграють мультимодальні системи, які одночасно аналізують зображення, звук і метадані: вони успішно застосовуються в кіберрозвідці, де приховані сигнали можуть передаватися, наприклад, незначними шумовими коливаннями у звуковій доріжці чи зміненими параметрами файлів.

Інтеграція технологій комп'ютерного зору та машинного навчання створює окремий клас алгоритмічних підходів, що допомагають значно підвищити точність ідентифікації прихованої інформації у відеопотоці. Якщо традиційні методи працюють з окремими ознаками зображення, то інтегровані архітектури здатні поєднувати просторовий, часовий та контекстуальний аналіз у єдиній моделі. Така синергія сприяє не лише точнішому розпізнаванню прихованих патернів, але й адаптивності до різних форматів відео та сценаріїв застосування. Сучасні дослідження доводять важливість гібридних алгоритмічних рішень, що комбінують можливості згорткових

нейронних мереж для аналізу просторових характеристик, рекурентних мереж для роботи з часовими залежностями та алгоритмів машинного навчання для підсумкової класифікації сигналів (табл. 2).

Таблиця 2

Алгоритмічні підходи інтеграції комп'ютерного зору та машинного навчання

Алгоритмічний підхід	Механізм інтеграції	Приклад використання	Практичний ефект
CNN + SVM (Support Vector Machine)	CNN генерує ознаки, SVM здійснює класифікацію	Автоматизований відеоконтроль на митниці	Висока точність за відносно незначних обчислювальних витрат
CNN–RNN архітектури	Поєднання просторового аналізу кадрів і часової динаміки	Моніторинг транспортних потоків для виявлення прихованих сигналів	Виявлення аномалій у мікродинаміці рухів
Attention-based Transformers + комп'ютерний зір	Застосування механізму самоуваги для пошуку прихованих залежностей	Кібермоніторинг у потокових медіа	Масштабованість та здатність працювати в режимі реального часу
Ensemble-моделі (ансамблі Machine Learning (ML) та Deep Learning (DL))	Комбінація кількох класифікаторів для підвищення стабільності	Аналітичні центри інформаційної безпеки	Зменшення кількості хибнопозитивних результатів

Джерело: сформовано авторами на основі [2, p. 4491–4492; 4, p. 5841–5842; 10, p. 37962–37963; 15, p. 7]

Нині інтеграція комп'ютерного зору та машинного навчання реалізується через гнучкі алгоритмічні архітектури, здатні одночасно аналізувати просторові й часові характеристики відеопотоку. Поєднання CNN із традиційними класифікаторами, зокрема SVM, довело ефективність у прикордонному та митному контролі, де важлива швидкість і відносна простота обчислень. Архітектури CNN–RNN сприяють виявленню прихованих сигналів у динаміці руху об'єктів, що робить їх особливо цінними в транспортних системах і відеоспостереженні високої частоти [2, p. 4493–4494]. Трансформери з їхнім механізмом самоуваги забезпечують обробку великих обсягів потокових даних у реальному часі, що має вирішальне значення для кібермоніторингу та запобігання інформаційним атакам. Ансамблеві моделі, інтегруючи результати кількох алгоритмів, сприяють зниженню рівня хибнопозитивних сигналів, створюючи стабільніші та практично корисні рішення для аналітичних центрів безпеки. Таким чином, на практиці саме інтегровані алгоритмічні підходи формують новий стандарт точності й надійності у виявленні прихованих повідомлень.

Мультимодальний підхід до аналізу відеопотоку ґрунтується на одночасному використанні кількох типів даних, що дає змогу підвищити стійкість та точність виявлення прихованої інформації. На відміну від моделей, які працюють лише з візуальними ознаками, інтеграція звукових доріжок, текстових метаданих та характеристик середовища створює комплексне уявлення про зміст і контекст відео. Такий підхід є особливо актуальним у випадках, коли приховані повідомлення маскуються не у візуальних патернах, а в акустичних або структурних параметрах файлів. Поєднання різних джерел інформації сприяє формуванню більш надійних та адаптивних моделей, здатних працювати в умовах багатовимірних загроз (табл. 3).

Таблиця 3

Можливості використання мультимодальних даних для аналізу відеопотоку

Тип даних	Механізм інтеграції	Приклад використання	Очікуваний результат
Візуальні дані (зображення, кадри відео)	Виділення просторових ознак і поєднання з іншими каналами	Аналіз відео в системах безпеки аеропортів	Виявлення підозрілих змін у структурі кадрів
Аудіодані (звукові доріжки)	Обробка спектральних характеристик та поєднання з відеоаналізом	Контроль трансляцій для пошуку закодованих сигналів у шумі	Виявлення прихованих акустичних повідомлень
Метадані (формат, часові мітки, технічні параметри)	Кореляція службової інформації з візуально-звуковими каналами	Перевірка автентичності відеофайлів у кіберрозслідуваннях	Виявлення аномалій у структурі даних
Інтегровані мультимодальні моделі	Поєднання кількох каналів у єдиній архітектурі (наприклад, multimodal deep learning)	Системи аналітики потокових відео для кібермоніторингу	Підвищення точності та зменшення кількості хибних результатів

Джерело: сформовано авторами на основі [7, p. 6–7; 11, p. 23413–23414; 13, p. 4599–4600; 14]

У сучасних системах аналізу відеопотоку мультимодальність виконує роль визначального чинника підвищення об'єктивності: окремі канали інформації часто дають суперечливі або неповні результати, тоді як їхня інтеграція створює цілісну модель [14]. Візуальний аналіз дає змогу зафіксувати структурні відхилення, аудіоканал виявляє акустичні аномалії, а метадані сигналізують про приховані зміни у файлі. На практиці такий підхід продемонстровано в дослідженнях реального часу потокового мультимодального аналізу для кібербезпеки, де поєднання відео- та аудіоданих дало змогу істотно знизити кількість хибних тривог у моніторингових системах [15, с. 2]. Так, практичне впровадження мультимодальних моделей підтверджує їхню здатність ефективно виявляти приховані повідомлення в складних потоках даних та підвищувати надійність систем реагування.

Практична ефективність застосування інтелектуальних методів аналізу відеопотоку для виявлення прихованих повідомлень обмежується низкою проблем. Однією з головних є надзвичайно високі обчислювальні витрати, оскільки моделі глибинного навчання потребують потужних апаратних ресурсів, а робота з потоковим відео – обробки значних обсягів даних у реальному часі. Це ускладнює масштабування систем і їхнє впровадження в ресурсно обмежених середовищах, наприклад, у прикордонному контролі чи мобільних моніторингових комплексах [2, p. 4490]. Другою важливою проблемою є обмежена стійкість алгоритмів до зашифрованих або спеціально модифікованих відеопотоків, коли приховані повідомлення інтегруються в шумові структури або адаптивно змінюють параметри, щоб обходити виявлення [4, p. 5842–5843; 9, p. 5]. Уразливим місцем залишається здатність моделей адаптуватися до нових методів стеганографії, які еволюціонують швидше, ніж оновлюються алгоритмічні рішення.

Ще однією проблемою є висока залежність від якості даних: навіть невелике стиснення, артефакти кодування чи втрати пакетів під час передавання можуть істотно знизити точність розпізнавання. Це особливо важливо для систем потокового відео у відкритих мережах [7, p. 12–13]. Поширеним обмеженням є також відсутність достатніх обсягів анотованих мультимодальних даних для тренування моделей, що ускладнює досягнення високої узагальнювальної здатності та підвищує ризик помилкових результатів у реальних умовах [6, p. 102023].

У сфері кіберзахисту існує проблема низької інтерпретованості інтелектуальних моделей: навіть після успішного виявлення прихованих повідомлень, неможливість пояснити ознаки, на підставі яких система ухвалила рішення, обмежує довіру до результатів та ускладнює їхнє застосування в судово-експертних процесах [12, p. 1640–1641].

Не менш важливими є ризики масштабного впровадження: централізовані системи є потенційними мішенями для кібератак, а adversarial-атаки підтверджують, що мінімальні зміни у вхідних даних достатні для обходу навіть найсучасніших алгоритмів [5, p. 3055; 14]. Проблемою залишається високе енергоспоживання та вартість утримання складних інфраструктур, що обмежує можливості їхнього застосування в середовищах із жорсткими бюджетними обмеженнями.

Для вдосконалення систем аналізу відеопотоку потрібні адаптивні архітектури, здатні до самооновлення на основі безперервного машинного навчання. Застосування підходів continual learning та few-shot learning дає змогу моделям підтримувати актуальність навіть за умови появи нових методів стеганографії, що мінімізує часовий розрив між загрозою та здатністю системи її розпізнати. Перспективним напрямом є застосування федеративного навчання, що забезпечує обробку даних безпосередньо на периферійних вузлах, зберігаючи конфіденційність і одночасно підвищуючи масштабованість системи. Це особливо важливо для розподілених мереж відеоспостереження, де централізоване зберігання даних є вразливим і ресурсомістким.

Важливим завданням є розвиток енергоефективних моделей глибинного навчання, оптимізованих за допомогою pruning та quantization, що дає змогу істотно знизити обчислювальні витрати без втрати точності. Такі оптимізовані архітектури створюють умови для інтеграції систем у мобільні та портативні платформи, зокрема в безпілотні літальні апарати та переносні комплекси моніторингу. Одночасно постає потреба у впровадженні explainable AI, яке інтерпретує роботу моделей та перетворює її на зрозумілі закономірності чи ознаки. Це підвищує довіру операторів до системи та водночас створює юридично значущу доказову базу під час використання результатів виявлення в правових чи судово-експертних процесах.

Для систематизації отриманих результатів і формування орієнтирів подальших досліджень запропоновано покрокову схему розвитку інтелектуальних мультимодальних систем відеоаналізу. Схема відображає логіку еволюції від концептуальних рішень до практичного впровадження у сфері кіберзахисту. Вона побудована на основі поєднання принципів комп'ютерного зору, машинного навчання та мультимодального аналізу, а також враховує вимоги енергоефективності, інтерпретованості й інтеграційної сумісності. Інноваційність схеми полягає у комплексному підході до обробки даних – відеоаналіз розглядається як елемент багатоканальної системи інформаційної безпеки, що істотно знижує ризик хибнопозитивних результатів і підвищує швидкість реагування на загрози (рис. 1).

На першому етапі відбувається формування концептуальних моделей на основі інтеграції комп'ютерного зору та машинного навчання (CNN, RNN, трансформери), що забезпечує базову точність і здатність виявляти приховані структури у відео. Другий етап передбачає додавання мультимодальних каналів (аудіо, метадані, параметри середовища), завдяки чому зменшується кількість хибнопозитивних сигналів і зростає адаптивність системи. Третій етап полягає у впровадженні енергоефективних моделей через pruning, quantization та edge/federated learning, що знижує обчислювальні витрати й робить можливим мобільне застосування. На четвертому етапі забезпечується інтеграція з іншими засобами кіберзахисту (системами виявлення вторгнень, аналізом трафіку, цифровою криміналістикою), що формує комплексні інфраструктури захисту. Завершальним п'ятим етапом є контекстуальний та explainable аналіз із використанням continual learning і технологій XAI, що забезпечує прозорість прийняття рішень, довіру користувачів і превентивне реагування на загрози. Очікується, що реалізація такої схеми сприятиме створенню нового покоління інтелектуальних систем відеоаналізу, здатних ефективно протидіяти сучасним і майбутнім кіберзагрозам.

Рис. 1. Покрокова схема розвитку мультимодальних систем відеоаналізу

Джерело: власна розробка авторів

Висновки

У дослідженні доведено, що інтелектуальні методи аналізу відеопотоку значно розширюють можливості виявлення прихованих повідомлень завдяки поєднанню комп'ютерного зору, машинного навчання та мультимодальних підходів. Традиційні методи обробки сигналів придатні лише для базового виявлення аномалій, тоді як глибокі нейронні мережі та інтегровані архітектури забезпечують високу точність і адаптивність у динамічних умовах. Гібридні моделі виявилися ефективними, оскільки одночасно враховують просторові, часові й контекстуальні характеристики відео, що робить їх перспективними для застосування в кіберзахисті, прикордонному контролі та моніторингу критичної інфраструктури.

Ідентифіковано основні обмеження: значні обчислювальні витрати, нестійкість до шифрування та навмисних модифікацій, дефіцит якісних датасетів, низька інтерпретованість результатів і високі вимоги до ресурсів. Також виявлено ризики, пов'язані з можливістю adversarial-атак, які знижують довіру до результатів систем.

Подальші дослідження варто зосередити на створенні адаптивних архітектур із механізмами continual learning і federated learning, оптимізації моделей для енергоефективного функціонування, розвитку explainable AI та уніфікації протоколів інтеграції з іншими засобами кіберзахисту. Такий підхід сприятиме формуванню нового покоління інтелектуальних систем відеоаналізу, здатних не лише виявляти приховані повідомлення, а й забезпечувати превентивний захист інформаційного простору.

Список використаної літератури

1. Horbenko Y. Confidential Computing in Front-End: Enhancing Data Security with Secure Enclaves and Homomorphic Encryption. *International Journal of Advanced Multidisciplinary Research and Studies*. 2025. Vol. 5, № 3. P. 308-321. URL: https://www.multiresearchjournal.com/admin/uploads/archives/archive-1747130538.pdf (date of access: 27.09.2025).

2. Wan S., Xu X., Wang T., Gu Z. An Intelligent Video Analysis Method for Abnormal Event Detection in Intelligent Transportation Systems. *IEEE Transactions on Intelligent Transportation Systems*. 2021. Vol. 22, № 7. P. 4487–4495. DOI: https://doi.org/10.1109/TITS.2020.3017505

3. Meghanathan N., Nayak L. Steganalysis algorithms for detecting the hidden information in image, audio and video cover media. *International Journal of Network Security & Its Application (IJNSA)*. 2010. Vol. 2, № 1. P. 43–55. DOI: https://doi.org/10.5121/ijnsa.2010.2104

4. Dalal M., Juneja M. A survey on information hiding using video steganography. *Artificial Intelligence Review*. 2021. Vol. 54. P. 5831–5895. DOI: https://doi.org/10.1007/s10462-021-09988-9

5. Aberna P., Agilandeeswari L. Survey of Digital Video Watermarking Techniques and Attacks. *Wireless Personal Communications*. 2021. Vol. 121. P. 3049–3082. DOI: https://doi.org/10.1007/s11277-021-08430-8

6. Nazah S., Huda S., Abawajy J., Hassan M. M. Evolution of Dark Web Threats: Social Engineering, Phishing and Cybercrime. *IEEE Access*. 2019. Vol. 7. P. 102021–102029. DOI: https://doi.org/10.1109/ACCESS.2019.2931097

7. Song L., Licoppe C. A multimodal approach to the sequential analysis of video streams closures. *Social Interaction. Video-Based Studies of Human Sociality*. 2020. Vol. 3, № 3. P. 1–24. DOI: https://doi.org/10.7146/si.v3i3.120547

8. Dhawan S., Gupta R. Steganography: An Overview. *International Journal of Advanced Research in Computer Science and Software Engineering*. 2012. Vol. 2, № 10. P. 237–240. DOI: https://doi.org/10.1007/s10462-021-09968-0

9. Huang R., Wu H., Wang X., Cheng G., Hu X. Security Analysis of Video Identification Protocols for Instant Messaging Systems. *Security and Communication Networks*. 2018. Vol. 2018. P. 1–11. DOI: https://doi.org/10.1155/2018/3071247

10. Dalal M., Juneja M. Information Hiding Techniques for Steganography and Steganalysis. *Multimedia Tools and Applications*. 2022. Vol. 81. P. 37953–37989. DOI: https://doi.org/10.1007/s11042-022-12722-2.

11. Subramanian N., Elharrouss O., Al-Maadeed S., Bouridane A. Image Steganography: A Review of the Recent Advances. *IEEE Access*. 2021. Vol. 9. P. 23409–23423. DOI: https://doi.org/10.1109/ACCESS.2021.3053998

12. Wahab O. F. A., Khalaf A. A. M., Hussein A. I., Hamed H. F. A. A Secure and High Capacity RSA-Based Video Steganography Technique Using Lempel–Ziv–Welch Algorithm. *Multimedia Tools and Applications*. 2022. Vol. 81. P. 1637–1662. DOI: https://doi.org/10.1007/s11042-021-11693-7

13. Sun Z., Sarma P., Sethares W., Liang Y. Learning Relationships between Text, Audio, and Video via Deep Canonical Correlation for Multimodal Clustering. *Proceedings of the AAAI Conference on Artificial Intelligence*. 2021. Vol. 35, № 5. P. 4596–4604. DOI: https://doi.org/10.1609/aaai.v35i5.16539

14. Yu C., Zhang X., Zhang X., Li G., Tang Z. Reversible Data Hiding in Encrypted Images Based on Multi-Level Hierarchical Embedding. *Signal Processing*. 2021. Vol. 178. Article 107794. DOI: https://doi.org/10.1016/j.sigpro.2020.107794

15. Xiao P., Xiao M., Cai N., Qiu B., Zhou S., Wang H. Adaptive Hybrid Framework for Multiscale Void Inspection of Chip Resistor Solder Joints. *IEEE Transactions on Instrumentation and Measurement*. 2023. Vol. 72. P. 1–12. DOI: https://doi.org/10.1109/TIM.2023.3235435

References

1. Horbenko, Y. (2025). Confidential Computing in Front-End: Enhancing Data Security with Secure Enclaves and Homomorphic Encryption. *International Journal of Advanced Multidisciplinary Research and Studies*, 5(3), 308–321. Retrived from: https://www.multiresearchjournal.com/admin/uploads/archives/archive-1747130538.pdf

2. Wan, S., Xu, X., Wang, T., & Gu, Z. (2021). An Intelligent Video Analysis Method for Abnormal Event Detection in Intelligent Transportation Systems. *IEEE Transactions on Intelligent Transportation Systems*, 22(7), 4487–4495. DOI: https://doi.org/10.1109/TITS.2020.3017505

3. Meghanathan, N., & Nayak, L. (2010). Steganalysis algorithms for detecting the hidden information in image, audio and video cover media. *International Journal of Network Security & Its Application (IJNSA)*, 2(1), 43–55. DOI: https://doi.org/10.5121/ijnsa.2010.2104

4. Dalal, M., & Juneja, M. (2021). A survey on information hiding using video steganography. *Artificial Intelligence Review*, 54, 5831–5895. DOI: https://doi.org/10.1007/s10462-021-09988-9

5. Aberna, P., & Agilandeeswari, L. (2021). Survey of Digital Video Watermarking Techniques and Attacks. *Wireless Personal Communications*, 121, 3049–3082. DOI: https://doi.org/10.1007/s11277-021-08430-8

6. Nazah, S., Huda, S., Abawajy, J., & Hassan, M. M. (2019). Evolution of Dark Web Threats: Social Engineering, Phishing and Cybercrime. *IEEE Access*, 7, 102021–102029. DOI: https://doi.org/10.1109/ACCESS.2019.2931097

7. Song, L., & Licoppe, C. (2020). A multimodal approach to the sequential analysis of video streams closures. *Social Interaction. Video-Based Studies of Human Sociality*, 3(3), 1–24. DOI: https://doi.org/10.7146/si.v3i3.120547

8. Dhawan, S., & Gupta, R. (2012). Steganography: An Overview. *International Journal of Advanced Research in Computer Science and Software Engineering*, 2(10), 237–240. DOI: https://doi.org/10.1007/s10462-021-09968-0

9. Huang, R., Wu, H., Wang, X., Cheng, G., & Hu, X. (2018). Security Analysis of Video Identification Protocols for Instant Messaging Systems. *Security and Communication Networks*, 2018, 1–11. DOI: https://doi.org/10.1155/2018/3071247

10. Dalal, M., & Juneja, M. (2022). Information Hiding Techniques for Steganography and Steganalysis. *Multimedia Tools and Applications*, 81, 37953–37989. DOI: https://doi.org/10.1007/s11042-022-12722-2

11. Subramanian, N., Elharrouss, O., Al-Maadeed, S., & Bouridane, A. (2021). Image Steganography: A Review of the Recent Advances. *IEEE Access*, 9, 23409–23423. DOI: https://doi.org/10.1109/ACCESS.2021.3053998

12. Wahab, O. F. A., Khalaf, A. A. M., Hussein, A. I., & Hamed, H. F. A. (2022). A Secure and High Capacity RSA-Based Video Steganography Technique Using Lempel–Ziv–Welch Algorithm. *Multimedia Tools and Applications*, 81, 1637–1662. DOI: https://doi.org/10.1007/s11042-021-11693-7

13. Sun, Z., Sarma, P., Sethares, W., & Liang, Y. (2021). Learning Relationships between Text, Audio, and Video via Deep Canonical Correlation for Multimodal Clustering. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(5), 4596–4604. DOI: https://doi.org/10.1609/aaai.v35i5.16539

14. Yu, C., Zhang, X., Zhang, X., Li, G., & Tang, Z. (2021). Reversible Data Hiding in Encrypted Images Based on Multi-Level Hierarchical Embedding. *Signal Processing*, 178, Article 107794. DOI: https://doi.org/10.1016/j.sigpro.2020.107794

15. Xiao, P., Xiao, M., Cai, N., Qiu, B., Zhou, S., & Wang, H. (2023). Adaptive Hybrid Framework for Multiscale Void Inspection of Chip Resistor Solder Joints. *IEEE Transactions on Instrumentation and Measurement*, 72, 1–12. DOI: https://doi.org/10.1109/TIM.2023.3235435.

Дата першого надходження рукопису до видання: 21.09.2025
Дата прийнятого до друку рукопису після рецензування: 17.10.2025
Дата публікації:

УДК 004.75:004.056.55:004.85 DOI

О. І. ТЕРЕЩЕНКО
аспірант кафедри інженерії програмного забезпечення
Національний університет «Одеська політехніка»
ORCID: 0000-0003-4510-5255

МЕТОДИ ІН'ЄКЦІЇ ВРАЗЛИВОСТЕЙ У СМАРТ-КОНТРАКТИ ДЛЯ ФОРМУВАННЯ ЗБАЛАНСОВАНИХ ДАТАСЕТІВ

Смарт-контракти широко застосовуються у фінансових та децентралізованих застосунках, проте їхня безпека залишається критичною проблемою. Аналіз існуючих корпусів показує значний дисбаланс: поширені класи уразливостей (integer overflow/underflow) істотно домінують, тоді як критично небезпечна вразливість повторного входу представлена обмежено. Це ускладнює навчання та об'єктивне оцінювання інструментів виявлення вразливостей. Мета дослідження полягає у підвищенні об'єктивності та якості навчання й тестування методів виявлення уразливостей у смарт-контрактах шляхом створення збалансованого контрольованого датасету. Запропоновано два взаємодоповнювальні підходи ін'єкції. Детермінований метод спирається на статичний аналіз і формальні патерни відбору/модифікації функцій, забезпечуючи відтворюваність і синтаксичну коректність. LLM-підхід виконує контекстно узгоджені зміни з мінімальною різницею коду, підвищуючи різноманітність прикладів. Обидва інтегровані в єдиний конвеєр із нормалізацією, дедуплікацією та багатоступеневою валідацією: успішна компіляція solc, статичне підтвердження цільових патернів, перевірка збереження нецільової логіки та мінімальності змін. Результатом є збалансований датасет з п'яти класів (integer_overflow, integer_underflow, timestamp_dependency, reentrancy, «безпечні» контракти) з вирівняною кількістю прикладів, стандартизованим форматом збереження (повний контракт, вразливий фрагмент, метадані) та відтворюваним пайплайном. Поєднання детермінованого й LLM-методів забезпечує баланс контрольованості та реалістичності, що покращує якість експериментів і чесність порівнянь інструментів. Новизна полягає в уніфікованій формальній специфікації операторів ін'єкції та практичному конвеєрі пакетної генерації; обмеження стосуються стохастичності LLM і потреби подальших динамічних PoC-перевірок.

Ключові слова: *смарт-контракти, Solidity, вразливості, ін'єкція вразливостей, reentrancy, integer overflow, збалансований датасет, статичний аналіз, великі мовні моделі.*

O. I. TERESHCHENKO
Postgraduate Student at the Department of Software Engineering
Odesa Polytechnic National University
ORCID: 0000-0003-4510-5255

METHODS OF VULNERABILITY INJECTION INTO SMART CONTRACTS FOR BALANCED DATASET GENERATION

Smart contracts are widely used in financial and decentralized applications; however, their security remains a critical issue. Analysis of existing corpora shows a significant imbalance: common vulnerability classes (integer overflow/underflow) are strongly dominant, while the critically dangerous reentrancy vulnerability is underrepresented. This complicates both training and objective evaluation of vulnerability detection tools. The aim of the study is to improve the objectivity and quality of training and testing methods for detecting vulnerabilities in smart contracts by creating a balanced and controlled dataset. Two complementary injection approaches are proposed. The deterministic method relies on static analysis and formal patterns for selecting and modifying functions, ensuring reproducibility and syntactic correctness. The LLM-based approach performs context-aware modifications with minimal code differences, increasing the diversity of examples. Both approaches are integrated into a unified pipeline with normalization, deduplication, and multi-stage validation: successful solc compilation, static confirmation of target patterns, preservation of non-target logic, and minimization of code changes. The result is a balanced dataset of five classes (integer_overflow, integer_underflow, timestamp_dependency, reentrancy, and "safe" contracts) with an equal number of examples, a standardized storage format (full contract, vulnerable snippet, metadata), and a reproducible pipeline. Combining deterministic and LLM methods provides a balance between controllability and realism, which improves the quality of experiments and the fairness of tool comparisons. The novelty lies in the unified formal specification of injection operators and the practical pipeline for batch dataset generation, while limitations concern the stochastic nature of LLMs and the need for further dynamic PoC validation.

Key words: *smart contracts, Solidity, vulnerabilities, vulnerability injection, reentrancy, integer overflow, balanced dataset, static analysis, large language models.*

© Терещенко О. І., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

Постановка проблеми

Смарт-контракти на платформі Ethereum та інших блокчейн-системах широко застосовуються для фінансових операцій, токенізації активів та реалізації децентралізованих застосунків. Водночас відомо, що навіть незначні уразливості у коді можуть призвести до суттєвих фінансових втрат, як це сталося під час атак на DAO (2016) чи Parity Wallet (2017). Проблема полягає в тому, що існуючі реальні датасети смарт-контрактів містять сильний класовий дисбаланс: одні типи уразливостей (наприклад, integer overflow/underflow) значно переважають, тоді як критично небезпечні, але рідкісні (наприклад, reentrancy) представлені вкрай малою кількістю прикладів. Це унеможливлює ефективне навчання моделей машинного навчання та знижує якість виявлення уразливостей [1, 2].

Таким чином, постає задача створення збалансованого та контрольованого датасету смарт-контрактів, у якому всі цільові класи уразливостей представлені рівномірно. Для цього необхідні методи автоматизованої ін'єкції вразливостей, що забезпечують відтворюваність, синтаксичну коректність та різноманітність вихідного коду.

Аналіз останніх досліджень і публікацій

Стан формування датасетів для смарт-контрактів еволюціонував від невеликих наборів до масивних корпусів та синтетичних колекцій з ін'єкцією вразливостей. Ранній стандарт де-факто – екосистема SmartBugs: вона поєднує невеликий точно розмічений набір (≈143 контрактів, 208 тегованих вразливостей) для точнісного бенчмаркінгу з великим корпусом SmartBugs-Wild (≈47 398 контрактів із Ethereum), призначеним для масштабних експериментів. Такий дуальний дизайн дав спільноті як чисті еталони, так і реалістичний «шумний» контекст, однак залишив відкритим питання про баланс класів (рідкісні категорії залишаються недостатньо представленими) [3].

Новіший підхід – опора на реальні аудиторські звіти. DAppSCAN (TSE'24) системно зібрав і промаркував слабкості, пов'язавши їх з вихідним кодом та байткодом, спираючись на 1199 звітів 29 команд. Переваги підходу – валідація «з поля» та трасування до першоджерел (звіти), що суттєво підвищує довіру до міток; втім, охоплення класів і різноманітність прикладів все ще залежать від того, що реально потрапляє в аудит [4].

Паралельно з'явилися дуже великі, але переважно безрозміткові корпуси для pre-training та бенчмаркінгу інструментів – зокрема DISL (514 506 унікальних Solidity-файлів, усі верифіковані на mainnet). Такі набори критично важливі для репрезентативності та відтворюваності ML-досліджень, але самі по собі не знімають проблему дисбалансу/браку ground truth і часто потребують подальшого лейблінгу чи слабко-контрольованих евристик [5].

Щоб здобути масштаб із мітками, використовують авто-розмітку аналізаторами з majority voting. Прикладом є ScrawlD, де вразливості позначаються консенсусом кількох інструментів [6]. Це дає тисячі реальних прикладів з помірною ціною розмітки, проте вносить «label noise» (похибки/упередження інструментів), який доводиться компенсувати перевірками підмножин або поєднувати з «чистішими» джерелами міток.

Другий напрям – синтетичні датасети через ін'єкцію. SolidiFI показав, що систематичні мутації коду дозволяють будувати контрольовані колекції для чесної оцінки детекторів; при цьому автори акцентують на необхідності перевіряти компільованість і потенційну експлуатованість ін'єкцій (не всі штучні дефекти є реально небезпечними) [7]. Свіжі роботи розвивають цю лінію (наприклад, MuSe, 2025) й оптимізують оператори ін'єкцій під поширені категорії слабкостей, але центральним залишається компроміс між реалістичністю модифікацій і масштабом [8].

Сумарно, сучасний консенсус виглядає так: потрібні гібридні підходи, що поєднують реалістичність великих корпусів і керованість синтетичних ін'єкцій для балансування рідкісних класів; аудиторські джерела (як-от DAppSCAN) бажано інтегрувати як «якісний еталон»; опис наборів має відповідати сучасним стандартам прозорості (datasheets/data statements), які фіксують мотивацію, походження, процедури розмітки/ін'єкції, обмеження та відомі упередження – це підвищує відтворюваність і придатність датасетів для подальших досліджень [9, 10].

Формулювання мети дослідження

Мета дослідження полягає у підвищенні об'єктивності та достовірності навчання і тестування методів виявлення уразливостей у смарт-контрактах. Для її досягнення запропоновано створення збалансованого та відтворюваного датасету шляхом автоматизованої ін'єкції уразливостей за допомогою двох підходів – детермінованого та LLM-керованого. Такий датасет забезпечує рівномірне представлення різних класів уразливостей і створює контрольовані умови для коректного порівняння ефективності існуючих і нових інструментів аналізу безпеки [11].

Викладення основного матеріалу дослідження

Результати збору вихідних кодів смарт-контрактів

У ході дослідження було зібрано та проаналізовано вихідні коди смарт-контрактів на Solidity із двох основних джерел:

Etherscan API – актуальні верифіковані контракти з Ethereum mainnet;

SmartBugs Wild – відкритий датасет реальних контрактів із маркованими уразливостями.

Після попередньої обробки та аналізу інструментом Oyente було сформовано корпус, розподіл якого за категоріями уразливостей наведено в Таблиці 1.

Ці дані демонструють значну перевагу вразливостей типу integer overflow та integer underflow над іншими, а також порівняно малу кількість контрактів з вразливістю re-entrancy.

Таблиця 1

Розподіл контрактів за категоріями вразливостей

Категорія	Кількість контрактів
Integer Underflow	23 702
Integer Overflow	31 061
Timestamp Dependency	1 460
Re-entrancy Vulnerability	310
Без уразливостей (None)	12 829

Детермінований алгоритм ін'єкції вразливостей у смарт-контракти

Запропонований метод забезпечує детермінований підхід до автоматизованого впровадження уразливостей у вихідний код смарт-контрактів без використання машинного навчання. Алгоритм базується на статичному аналізі коду та пошуку шаблонів (pattern matching) для визначення місць, придатних для модифікації. Це дозволяє передбачувано та повторювано модифікувати код: при однакових вхідних даних результати ін'єкції завжди будуть тотожними, оскільки випадковість у процесі відсутня. Алгоритм не виконує сам контракт, а лише аналізує його статичну структуру і застосовує стандартизовані зміни, що гарантує збереження синтаксичної коректності коду після модифікації.

Метод базується на трьох ключових принципах:

– детермінованість: алгоритм завжди видає однаковий результат для одних і тих самих вхідних даних. Поведінка повністю передбачувана, оскільки у процесі модифікації відсутні випадкові фактори;

– статичний аналіз: виявлення потрібних місць для ін'єкції відбувається шляхом аналізу структури коду без виконання смарт-контракту. Алгоритм парсить вихідний код Solidity, щоб знайти специфічні конструкції та патерни, і перевіряє синтаксичну коректність зроблених змін;

– орієнтація на шаблони: для кожного типу уразливості визначено характерні шаблони коду. Алгоритм систематично шукає в контракті такі шаблонні конструкції і вносить стандартизовані зміни. Це забезпечує уніфікованість ін'єкцій та полегшує валідацію отриманих результатів.

Виявлення кандидатів (загальний випадок)

Алгоритм парсить контракт і відбирає функції-кандидати, у яких одночасно наявні: релевантні структури стану (напр., балансний mapping), операції з коштами (функції на кшталт withdraw/transfer, або зовнішні виклики переказу Ether), перевірки перед дією (require/assert), ефекти оновлення стану (напр., зменшення балансу). Лише функції, що відповідають усім критеріям цільового класу уразливості, передаються на етап модифікації.

Стратегії модифікації наявної функції (на прикладі reentrancy)

Якщо у функції вже є перевірка балансу, відправлення коштів і подальше оновлення стану, виконується цілеспрямована перестановка/заміна операцій:

– безпечний переказ (*transfer/send*) замінюється на низькорівневий виклик *call*{*value*:...}(«») із перевіркою success;

– порядок «Checks-Effects-Interactions» свідомо порушується: зовнішня взаємодія виконується до оновлення стану, що відкриває вікно для повторного входу.

Мінімальний фрагмент зміни (ілюстративно) зображено на рисунку 1.

```
// Було (безпечніше)
require(balances[msg.sender] >= amount);
payable(msg.sender).transfer(amount);
balances[msg.sender] -= amount;

// Стало (вразливо)
require(balances[msg.sender] >= amount);
(bool success, ) = msg.sender.call{value: amount}("");
require(success, "Transfer failed");
balances[msg.sender] -= amount;
```

Рис. 1. Мінімальний фрагмент зміни (ілюстративно) в коді

Стратегії додавання шаблонної функції (на прикладі reentrancy)

Якщо у вихідному контракті немає відповідної функції (наприклад, відсутній механізм виведення коштів), додається мінімальна вразлива функція з коректними сигнатурами та допоміжною логікою поповнення балансу. Приклад додавання шаблонної функції в код наведено на рисунку 2.

Додавання виконується так, щоб не порушувати існуючу функціональність і компіляцію.

```
function withdraw(uint256 amount) public {
    require(balances[msg.sender] >= amount, "Insufficient balance");
    (bool success, ) = msg.sender.call{value: amount}(""); // зовнішній виклик до
    оновлення стану
    require(success, "Transfer failed");
    balances[msg.sender] -= amount; // оновлення після взаємодії
}

function deposit() public payable {
    balances[msg.sender] += msg.value;
}
```

Рис. 2. Приклад додавання шаблонної функції в код

Валідація

Після модифікації проводиться: синтаксична перевірка (успішна компіляція solc цільової версії), підтвердження властивості уразливості статичним аналізом (наявність зовнішнього виклику до ефектів, очікувані патерни), перевірка сумісності (збереження базової логіки та сигнатур).

Математична модель детермінованого підходу

Вхідні дані: C – вихідний смарт-контракт, V – тип уразливості (наприклад, reentrancy), $P_V = \{p_1, \ldots, p_n\}$ – набір шаблонів для виявлення коду, придатного до інжектування вразливості.

1. Пошук кандидатів

Знаходимо множину функцій, що відповідають усім патернам:

$$F_V(C) = \{f \in C \mid \ \forall p \in P_V : match(p, f) = 1\}.$$

2. Вибір і модифікація

Вибираємо найпридатнішу функцію f^* за фіксованими правилами та застосовуємо оператор модифікації:

$$M\left(f^*, V\right) \to f_v',$$

де f_v' – версія з уразливістю.

3. Інжекція в контракт

$$I\left(C, V\right) = \begin{cases} C' = \left(C \setminus f^*\right) \cup f_v', якщо\ F_V\left(C\right) \neq \varnothing, \\ C' = C \cup F_{new}\left(V\right), інакше. \end{cases}$$

Якщо підхожа функція є – модифікуємо її, інакше додаємо нову шаблонну уразливу функцію.

Метод ін'єкції вразливостей у смарт-контракти на основі великих мовних моделей

Запропонований метод реалізує автоматизоване впровадження вразливостей у вихідний код смарт-контрактів із використанням великих мовних моделей (Large Language Models, LLM). На відміну від детермінованих патерн-орієнтованих алгоритмів, цей підхід забезпечує контекстно-залежні зміни, які зберігають первинну функціональність контракту і виглядають природно інтегрованими у вихідний код. Метою є формування різноманітних і реалістичних датасетів для досліджень у сфері виявлення вразливостей у смарт-контрактах.

Метод складається з чотирьох основних компонентів:

1. модуль попередньої обробки контрактів – оптимізує код для відповідності обмеженням моделі;
2. інтелектуальний інжектор вразливостей – використовує LLM для внесення змін у код;
3. модуль парсингу та валідації результатів – перевіряє правильність і якість модифікацій;
4. конвеєр генерації датасету – зберігає результати у стандартизованому форматі.

Попередня обробка контракту

Щоб задовольнити обмеження LLM (максимальна довжина запиту, обсяг токенів), код проходить оптимізацію: видалення коментарів, видалення порожніх рядків, усікання надмірно великих контрактів до граничного розміру з коректним закриттям структур. Функція оптимізації забезпечує синтаксичну завершеність навіть після обрізання коду.

Формування запиту (prompt engineering)

Система будує спеціалізований запит до LLM, що містить:

– рольову інструкцію: модель позиціонується як «дослідник безпеки»;

– чіткі обмеження: не створювати нових контрактів, модифікувати безпосередньо в наявному коді, зберігати логіку;

– вимоги до формату виходу: структура з повним модифікованим контрактом і виділеним фрагментом змін;

– опис цільової вразливості.

Генерація модифікованого контракту

Для генерації використовується модель GPT-4.1-mini з параметрами:

– temperature = 0.8 (баланс варіативності та стабільності),
– max_tokens = 4000,
– оптимізовані витрати та продуктивність.

Модель інтегрує вразливість згідно з вказівками, намагаючись мінімально змінювати код і зберігати його функціональність.

Валідація та парсинг відповіді

Після отримання відповіді виконується:

– вилучення Solidity-блоків із використанням регулярних виразів;
– нормалізація коду (усунення різниці у форматуванні);
– перевірка наявності вразливого фрагмента у повному коді;
– відхилення результатів, що не відповідають вимогам.

Конвеєр обробки

Основний цикл роботи системи включає:

1. Завантаження та перевірку контрактів.
2. Пропуск файлів > 10 КБ.
3. Виключення вже оброблених контрактів (SHA1-хешування).
4. Генерацію вразливої версії через LLM.
5. Парсинг і валідацію.
6. Збереження у форматі JSON з полями: contract – повний код, vulnerability – тип вразливості, vulnerable_snippet – змінений фрагмент.

Математична модель

Нехай: C – вихідний контракт, V – тип вразливості, AI – мовна модель.

Функція оптимізації:

$$O(C) = \begin{cases} C', якщо |C| \le MAX_SIZE \\ truncate(C, MAX_{SIZE}), інакше. \end{cases}$$

Генерація запиту:

$$P(C, V) = f_{prompt}(O(C), config(V)).$$

Відповідь моделі:

$$R = AI(P(C, V)).$$

Функція валідації:

$$\mathrm{Val}(\mathrm{R}) = \begin{cases} \{contract, snippet\}, якщо\ snippet \in contract, \\ \varnothing, інакше. \end{cases}$$

Оцінка успішності ін'єкції:

$$S = \frac{\sum_{i=1}^{n} 1\left[Val\left(R_i\right) \neq \varnothing\right]}{n},$$

де n – кількість оброблених контрактів.

Створення збалансованого датасету та його переваги

В результаті застосування розроблених методів ін'єкції було сформовано штучний збалансований датасет смарт-контрактів, що містить по 1000 прикладів для кожної з п'яти категорій (таблиця 2).

Таблиця 2

Структура створеного датасету

Категорія	Кількість контрактів
Integer Overflow	1000
Integer Underflow	1000
Timestamp Dependency	1000
Re-entrancy	1000
Безпечні контракти	1000
Разом	5000

Переваги створеного датасету:

– баланс класів – рівна кількість прикладів для кожної категорії дозволяє уникнути класового дисбалансу, що характерний для реальних даних;

– контрольованість вразливостей – для кожного контракту точно відомо, який тип уразливості він містить (або відсутній), що спрощує підготовку навчальних та тестових вибірок;

– реалістичність коду – модифікації виконані так, щоб зберегти логіку та структуру реальних смарт-контрактів;

– відтворюваність – процес ін'єкції повністю автоматизований і може бути повторений для створення додаткових або розширених версій датасету;

– придатність для навчання моделей – забезпечує оптимальні умови для тренування і тестування алгоритмів виявлення уразливостей у смарт-контрактах.

Таким чином, отриманий збалансований датасет є важливим внеском у дослідження безпеки смарт-контрактів, оскільки поєднує переваги реальних вихідних кодів та контрольованих умов ін'єкції уразливостей.

Висновки

У роботі представлено два взаємодоповнювальні підходи до ін'єкції вразливостей у смарт-контракти Solidity для формування збалансованого датасету. Детермінований алгоритм, заснований на статичному аналізі та шаблонних перетвореннях, забезпечує відтворюваність і синтаксичну коректність, тоді як LLM-керований інжектор (GPT-4.1-mini, temperature = 0.8) створює контекстно узгоджені та реалістичні модифікації, підвищуючи різноманітність прикладів. Обидва методи інтегровані в єдиний конвеєр з нормалізацією, дедуплікацією та багаторівневою валідацією, що дозволило побудувати збалансований датасет п'яти класів, у якому всі типи вразливостей представлені рівномірно.

Практична цінність полягає в можливості використовувати набір як навчально-тестову основу для інструментів статичного аналізу та ML/LLM-моделей, а також як міст між синтетичними та реальними даними. Обмеження роботи стосуються стохастичності LLM і потреби додаткового підтвердження експлуатованості частини ін'єкцій у динамічних сценаріях.

Список використаної літератури

1. Tereshchenko O. I., Komleva N. O. Vulnerability Detection of Smart Contracts Based on Bidirectional GRU and Attention Mechanism // Communications in Computer and Information Science. 2023. Vol. 1980. Springer, Cham. DOI: https://doi.org/10.1007/978-3-031-48325-7_21

2. Tereshchenko O. I., Komleva N. O. Identification and Localization of Vulnerabilities in Smart Contracts Using Attention Vectors Analysis in a BERT-Based Model // Radio Electronics, Computer Science, Control. 2024. № 3. С. 173–184. DOI: https://doi.org/10.15588/1607-3274-2024-3-15

3. Ferreira J. F., Cruz P., Durieux T., Abreu R. SmartBugs: A Framework to Analyze Solidity Smart Contracts // ASE 2020. DOI: https://doi.org/10.1145/3324884.3415298

4. Zheng Z., Su J., Chen J., Lo D., Zhong Z., Ye M. DAppSCAN: Building Large-Scale Datasets for Smart Contract Weaknesses in DApp Projects // IEEE Transactions on Software Engineering. 2024. URL: https://doi.org/10.1109/TSE.2024.3383422

5. Morello G., Eshghie S., et al. DISL: Fueling Research with a Large Dataset of Solidity Smart Contracts : [препринт] // arXiv : [cs.SE]. 2024. URL: https://doi.org/10.48550/arXiv.2403.16861

6. Yashavant C. S., Kumar S., Karkare A. ScrawlD: A Dataset of Real-World Ethereum Smart Contracts Labelled with Vulnerabilities : [препринт] // arXiv : [cs.SE]. 2022. URL: https://doi.org/10.48550/arXiv.2202.11409

7. Ghaleb A., Pattabiraman K. SolidiFI: An Automated and Systematic Approach for Evaluating Smart Contract Static Analysis Tools // Proceedings of the ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA). 2020. DOI: https://doi.org/10.1145/3395363.3397376

8. Iuliano G., et al. Automated Vulnerability Injection in Solidity Smart Contracts (MuSe) : [препринт] // arXiv : [cs.CR]. 2025. URL: https://doi.org/10.48550/arXiv.2504.15948

9. Gebru T., Morgenstern J., Vecchione B., Vaughan J. W., Wallach H., Daumé III H., Crawford K. Datasheets for Datasets // Communications of the ACM. 2021. Vol. 64, № 12. P. 86–92. DOI: https://doi.org/10.1145/3458723

10. Bender E. M., Friedman B. Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science // Transactions of the Association for Computational Linguistics (TACL). 2018. Vol. 6. P. 587–604. DOI: https://doi.org/10.1162/tacl_a_00041

11. Chang S., Zhang Y., Yu M., Jaakkola T. S. Invariant Rationalization : [препринт] // arXiv : [cs.LG]. 2020. URL: https://doi.org/10.48550/arXiv.2003.09772

References

1. Tereshchenko, O. I., & Komleva, N. O. (2023). Vulnerability Detection of Smart Contracts Based on Bidirectional GRU and Attention Mechanism. Communications in Computer and Information Science, 1980. Springer, Cham. https://doi.org/10.1007/978-3-031-48325-7_21

2. Tereshchenko, O. I., & Komleva, N. O. (2024). Identification and Localization of Vulnerabilities in Smart Contracts Using Attention Vectors Analysis in a BERT-Based Model. Radio Electronics, Computer Science, Control, (3), 173–184. https://doi.org/10.15588/1607-3274-2024-3-15

3. Ferreira, J. F., Cruz, P., Durieux, T., & Abreu, R. (2020). SmartBugs: A Framework to Analyze Solidity Smart Contracts. In Proceedings of the 35th IEEE/ACM International Conference on Automated Software Engineering (ASE 2020). https://doi.org/10.1145/3324884.3415298

4. Zheng, Z., Su, J., Chen, J., Lo, D., Zhong, Z., & Ye, M. (2024). DAppSCAN: Building Large-Scale Datasets for Smart Contract Weaknesses in DApp Projects. IEEE Transactions on Software Engineering. https://doi.org/10.1109/TSE.2024.3383422

5. Morello, G., Eshghie, S., et al. (2024). DISL: Fueling Research with a Large Dataset of Solidity Smart Contracts [Preprint]. arXiv:2403.16861. Retrieved from: https://doi.org/10.48550/arXiv.2403.16861

6. Yashavant, C. S., Kumar, S., & Karkare, A. (2022). ScrawlD: A Dataset of Real-World Ethereum Smart Contracts Labelled with Vulnerabilities [Preprint]. arXiv:2202.11409. Retrieved from: https://doi.org/10.48550/arXiv.2202.11409

7. Ghaleb, A., & Pattabiraman, K. (2020). SolidiFI: An Automated and Systematic Approach for Evaluating Smart Contract Static Analysis Tools. In Proceedings of the ACM SIGSOFT International Symposium on Software Testing and Analysis (ISSTA 2020). https://doi.org/10.1145/3395363.3397376

8. Iuliano, G., et al. (2025). Automated Vulnerability Injection in Solidity Smart Contracts (MuSe) [Preprint]. arXiv:2504.15948. Retrieved from: https://doi.org/10.48550/arXiv.2504.15948

9. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J. W., Wallach, H., Daumé, H. III, & Crawford, K. (2021). Datasheets for Datasets. Communications of the ACM, 64(12), 86–92. https://doi.org/10.1145/3458723

10. Bender, E. M., & Friedman, B. (2018). Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science. Transactions of the Association for Computational Linguistics (TACL), 6, 587–604. https://doi.org/10.1162/tacl_a_00041

11. Chang, S., Zhang, Y., Yu, M., & Jaakkola, T. S. (2020). Invariant Rationalization [Preprint]. arXiv:2003.09772. Retrieved from: https://doi.org/10.48550/arXiv.2003.09772

Дата першого надходження рукопису до видання: 25.09.2025
Дата прийнятого до друку рукопису після рецензування: 21.10.2025
Дата публікації:

УДК 004.42 DOI

С. Д. ТОЧИЛІН
кандидат фізико-математичних наук, доцент,
доцент кафедри інформаційної безпеки та наноелектроніки
Національний університет «Запорізька політехніка»
ORCID: 0003-2010-6358

КОМП'ЮТЕРНА ПРОГРАМА РІШЕННЯ РІЗНИХ ВАРІАНТІВ ЗАДАЧІ КОМІВОЯЖЕРА З ДОПОМОГОЮ ГЕНЕТИЧНОГО АЛГОРИТМУ

У наш час інтенсивно розвивається транспортна логістика. Оптимальні маршрути руху транспортних засобів, для певного набору пунктів комунікації: транзитних, вихідного та кінцевого пункту, можуть бути знайдені при рішенні різних варіантів задачі комівояжера.

Задача комівояжера – класична задача оптимізації. У випадку її замкнутого варіанта, комівояжер, рухаючись із вихідного пункту (ВП), однократно відвідує всі транзитні пункти та вертається у ВП. У випадку незамкнутого варіанта, комівояжер рухається з ВП, однократно відвідує всі транзитні пункти та прибуває у кінцевий пункт.

Маршрути руху для замкнутого та незамкнутого варіанта задачі можуть бути визначені за допомогою комп'ютерних програм, які використовують для рішення генетичний алгоритм (ГА).

При застосуванні ГА для рішення задачі комівояжера, при великій кількості транзитних пунктів, можлива ситуація вибору оптимального маршруту з декількох, отриманих при різних параметрах генетичного алгоритму. У цьому випадку необхідно оперативна інтерактивна взаємодія із програмою для зміни параметрів ГА. Як відомо, інтерактивність мають комп'ютерні програми із графічним інтерфейсом користувача.

У роботі на мові програмування Java розроблена кросплатформна комп'ютерна програма рішення різних варіантів задачі комівояжера з допомогою генетичного алгоритму. Вона має графічний інтерфейс користувача та дозволяє визначити оптимальний маршрут руху, як для замкнутого, так і незамкнутого варіанту задачі комівояжера. У той час, графічний інтерфейс додатка дає можливість змінювати параметри ГА у процесі його роботи та відображати знайдені маршрути руху у графічному вигляді.

*Попередньо дані про розташування пунктів руху, у вигляді їх назв та відповідних значень географічних координат широти та довготи, вводяться у програму з *.csv файлу.*

Програма при функціонуванні використовує генетичний алгоритм на основі методів, які для кожного варіанта задачі комівояжера реалізовували кросовер, мутацію хромосом, механізми елітизма та турнірного відбору особин, оцінку придатності популяції (методи реалізації кросовера і мутації хромосом для різних варіантів задачі відрізнялися). Для оцінки відстані між містами та довжини маршруту додаток використовує формулу гаверсинусів.

Ключові слова: *задача комівояжера; генетичний алгоритм; оптимальний маршрут; графічний інтерфейс користувача.*

S. D. TOCHILIN
Candidate of Physical and Mathematical Sciences, Associate Professor,
Associate Professor at the Department of Information Security
and Nanoelectronics
Zaporizhzhia Polytechnic National University
ORCID: 0000-0003-2010-6358

COMPUTER PROGRAM FOR SOLVING VARIOUS VERSIONS OF THE TRAVELING SALESMAN PROBLEM USING A GENETIC ALGORITHM

Currently, transport logistics is developing intensively. Optimal routes for vehicles for a certain set of communication points: transit, initial and final points, can be found by solving various versions of the traveling salesman problem.

The traveling salesman problem is a classical optimization problem. In the case of its closed version, the traveling salesman, moving from the initial point (IP), visits all transit points once and returns to the IP. In the case of an open version, the traveling salesman moves from the IP, visits all transit points once and arrives at the final point.

The routes for the closed and open versions of the problem can be determined using computer programs that use a genetic algorithm (GA) for solving.

When using GA to solve the traveling salesman problem, with a large number of transit points, a possible situation is choosing the optimal route from several obtained with different parameters of the genetic algorithm. In this case, prompt

© Точилін С. Д., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

interactive interaction with the program is necessary to change the GA parameters. As is known, computer programs with a graphical user interface have interactivity.

In this work, a cross-platform computer program for solving various versions of the traveling salesman problem using a genetic algorithm was developed in the Java programming language. It has a graphical user interface and allows you to determine the optimal route for both closed and open versions of the traveling salesman problem. At the same time, the graphical interface makes it possible to change the GA parameters in the process of working with the application and display the found routes in graphical form.

*Previously, data on the location of the traffic points, in the form of their names and corresponding values of geographic coordinates of latitude and longitude, are entered into the program from a *.csv file.*

The program uses a genetic algorithm based on methods that implemented crossover, chromosome mutation, elitism and tournament selection mechanisms, and population fitness assessment for each version of the traveling salesman problem (the methods for implementing crossover and chromosome mutation of the population differed for different versions of the problem). The application uses the haversine formula to estimate the distance between cities and the length of the route.

Key words: *traveling salesman problem; genetic algorithm; optimal route; graphical user interface.*

Постановка проблеми

У наш час інтенсивно розвивається транспортна логістика [1]. У цьому зв'язку актуальною задачею є розробка кросплатформних комп'ютерних програм, які автоматизують процес знаходження оптимальних транспортних маршрутів руху та мають графічний інтерфейс користувача (Graphical User Interface – GUI).

Оптимальні маршрути руху транспортних засобів, для певного набору пунктів комунікації: транзитних, вихідного і кінцевого пункту, можуть бути знайдені при рішенні різних варіантів задачі комівояжера. У випадку замкнутого варіанта (ЗВ), комівояжер, рухаючись із вихідного пункту (ВП), однократно відвідує всі транзитні пункти та вертається в ВП. У випадку незамкнутого варіанта (НВ), комівояжер рухається з ВП, однократно відвідує всі транзитні пункти і прибуває в кінцевий пункт.

Маршрути руху для ЗВ та НВ задачі можуть бути визначені за допомогою комп'ютерних програм, що використовують при рішенні генетичний алгоритм. При цьому актуальною проблемою є розробка комп'ютерних програм, які автоматизують процес знаходження оптимальних маршрутів комунікації на основі рішення задачі комівояжера за допомогою генетичного алгоритму та мають GUI.

Аналіз останніх досягнень і публікацій

До теперішнього часу розроблена велика кількість комп'ютерних програм, які вирішують задачу комівояжера за допомогою генетичного алгоритму, наприклад додатки створені у роботах [2–5]. Програми у [2–3] не мають графічного інтерфейсу та знаходять рішення лише для замкнутого варіанту задачі комівояжера. У той час, тільки незамкнутий варіант задачі вирішують додатки створені у [4, 5], але лише програма роботи [5] має GUI.

Формулювання мети дослідження

У роботі було поставлене завдання розробки кросплатформного Java-додатка з GUI з можливістю знаходження оптимального маршруту руху при рішенні, як замкнутого, так і незамкнутого варіанта задачі комівояжера за допомогою генетичного алгоритму.

Викладення основного матеріалу дослідження

Для рішення поставленого завдання був створений Java-додаток ST_Travelling із графічним інтерфейсом користувача.

При розробці програми використовувалася бібліотека jfreechart-1.0.19.jar, яка дозволяла відображати маршрути руху знайдені при роботі додатка у графічному вигляді. Процес компіляції та запуску програми здійснювався за допомогою, як архівних наборів класів і утиліт Java (наприклад, jdk-6u16), так і сучасних (наприклад, jdk-8.0.452.9-hotspot).

Програма при функціонуванні використовувала генетичний алгоритм на основі методів, які для кожного варіанта задачі комівояжера реалізовували кросовер, мутацію хромосом, механізми елітизма та турнірного відбору особин, оцінку придатності популяції (методи реалізації кросовера і мутації хромосом для різних варіантів задачі відрізнялися). При розрахунку довжини маршрутів, відстані між містами визначалися на основі значень їх широти та довготи за допомогою формули гаверсинусів [6].

На початку роботи із програмою, за допомогою меню «File», назви міст (пунктів маршруту) та їх географічні координати, завантажувалися з *.csv файлу і заповнювали таблицю «Cities». При цьому географічне розташування пунктів руху, у зменшеному масштабі, відображалося у вікні програми у графічному вигляді.

Потім установлювалися параметри необхідні для роботи додатка. Значення, розташовувані в полях вводу «Population», «pCrossover», «pMutation», «Elite», «Competitors», «Ngen», ініціалізували розмір популяції, імовірності схрещування та мутації неелітних хромосом у поколінні, число елітних особин, кількість хромосом у турнірі відбору, число поколінь хромосом, відповідно. Прапорці «TSP» і «OTSP» дозволяли вибрати замкнутий або незамкнутий варіант рішення задачі комівояжера, відповідно.

Запуск програми здійснювався при натисканні на кнопку «Run». При цьому додаток розраховував та відображав у графічному вигляді маршрут руху. Крім того, розміщав його довжину у полі вводу «Distance, km».

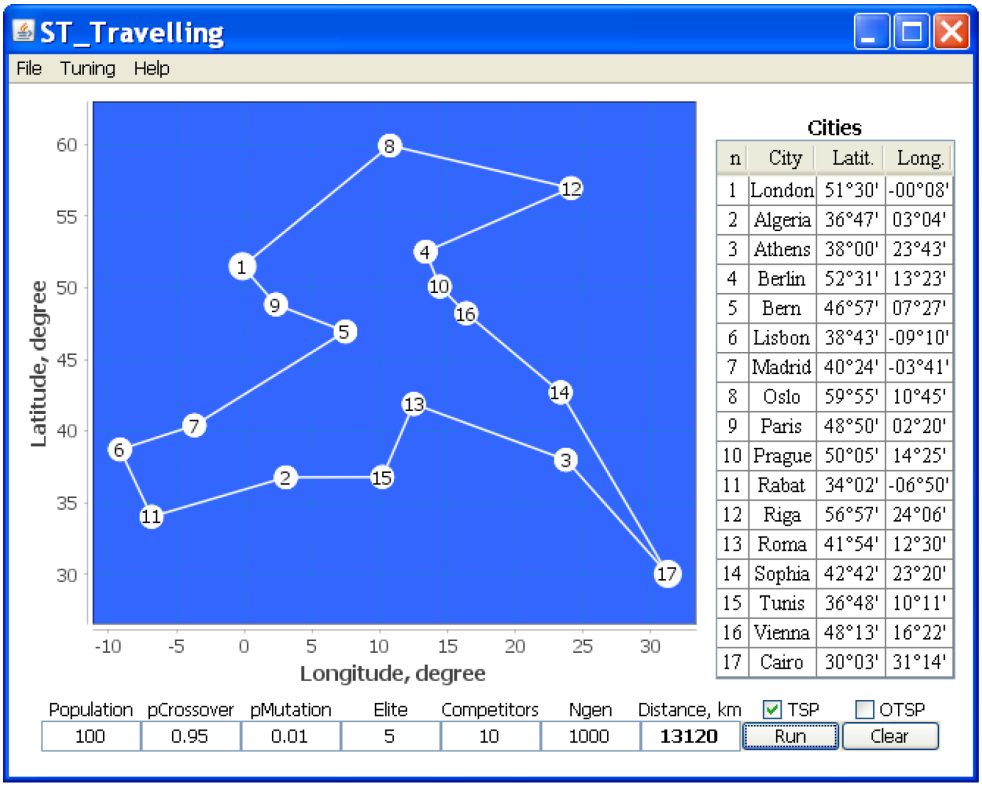


Рис. 1. Вікно програми з маршрутом руху

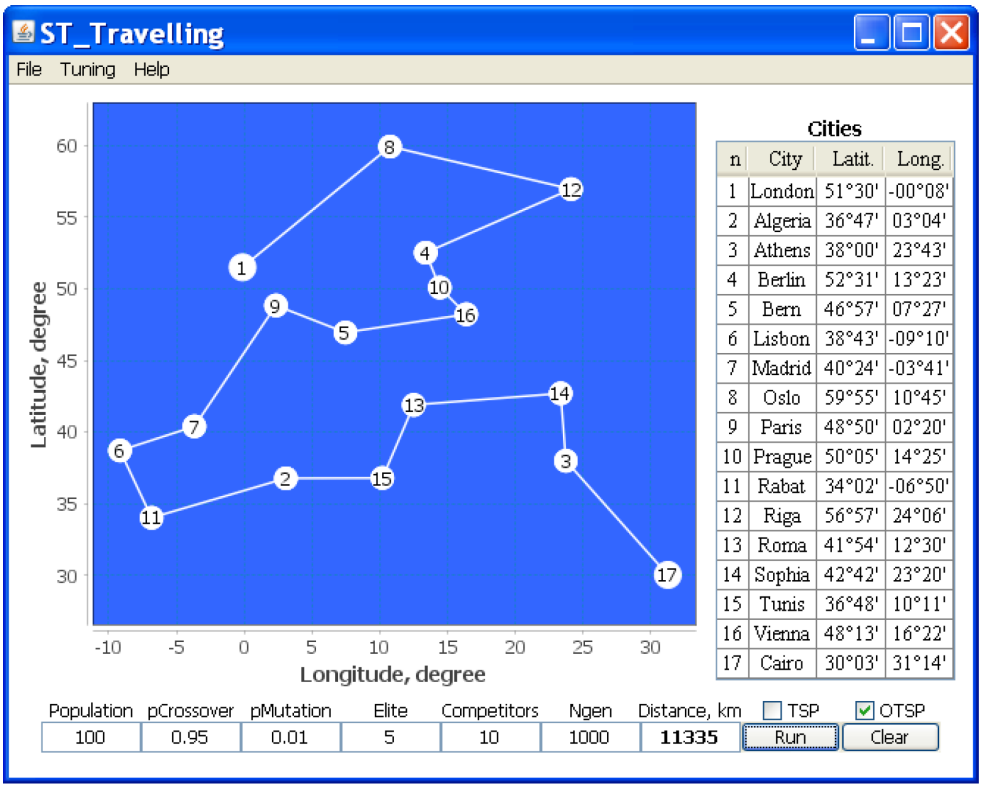


Рис. 2. Вікно програми з оптимальним маршрутом Лондон-Каїр

Для видалення знайденого маршруту використовувалася кнопка «Clear». Користувач програми за допомогою меню «Tuning», при необхідності, міг настроїти область відображення графіка. Інформація про додаток і особливості його роботи була доступна при використанні меню «Help».

На рис. 1 зображений маршрут авіасполучення для міст із таблиці «Cities», знайдений при рішенні замкнутого варіанта задачі комівояжера.

У той час, на рис. 2 зображений маршрут, знайдений при рішенні незамкнутого варіанта задачі комівояжера.

При цьому початковий і кінцевий пункт маршруту відповідає містам з першого та останнього рядка таблиці «Cities». При зміні записів у цих рядках, у випадку завантаження файлів з оновленими даними, відповідно міняється, як початковий, так і кінцевий пункт маршруту.

На рис. 3 зображений оптимальний маршрут руху між Парижем та Афінами, знайдений додатком ST_Travelling при рішенні незамкнутого варіанта задачі комівояжера.

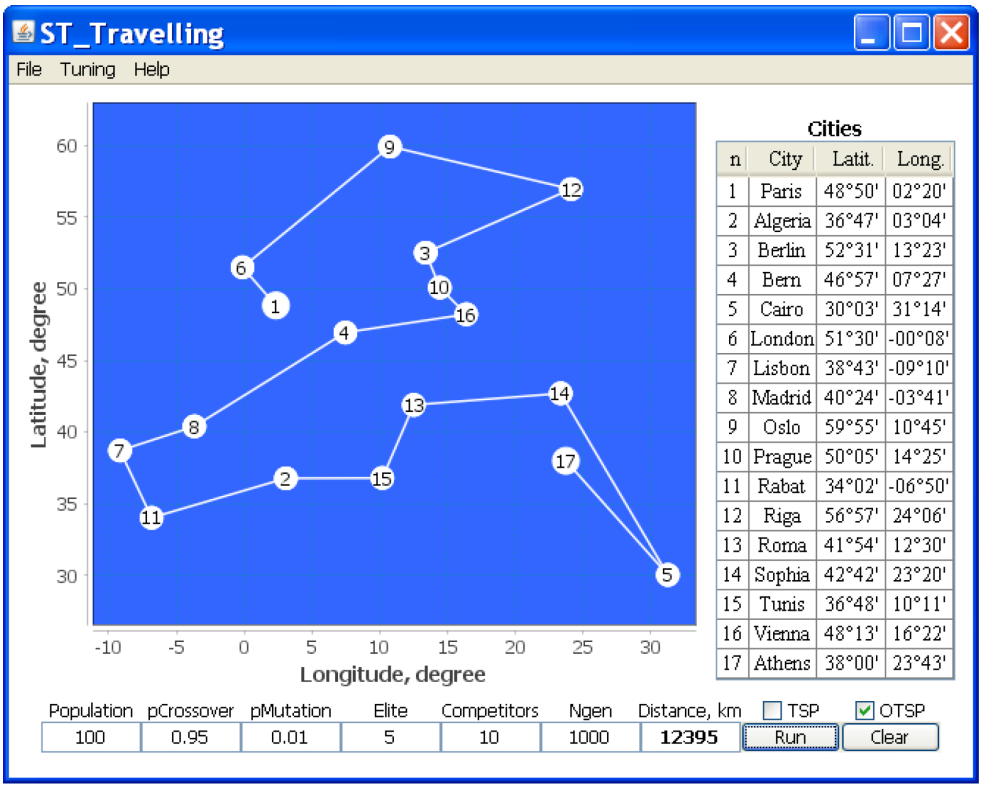


Рис. 3. Вікно програми з оптимальним маршрутом Париж-Афіни

Висновки

У роботі був розроблений кросплатформний Java-додаток ST_Travelling з GUI для знаходження оптимальних маршрутів руху на основі рішення, як замкнутого, так і незамкнутого варіанта задачі комівояжера з допомогою генетичного алгоритму.

Для знаходження маршруту, при рішенні незамкнутого варіанта задачі комівояжера, початковий та кінцевий пункт руху визначалися розташуванням інформації про них у вихідному файлі даних. Перший запис ініціалізував початковий пункт руху, останній – кінцевий.

При розрахунку довжини маршруту додаток використовував формулу гаверсинусів. Надалі передбачається модифікувати Java-додаток ST_Travelling, зокрема, при знаходженні протяжності маршруту використовувати більш точні значення широти та довготи пунктів руху.

Список використаної літератури

1. Смирнов И., Косарева Т. Транспортна логістика. Навчальний посібник. Київ : Центр навчальної літератури. 2019. 224 с.
2. Jacobson L., Kanber B. Genetic Algorithms in Java Basics. Apress. 2015. 154 p.
3. Wirsansky Eyal. Hands-On Genetic Algorithms with Python. Birmingham : Packt Publishing. 2020. 320 p.
4. Heaton J. Introduction to Neutral Network with Java. St. Louis: Heaton Research. 2008. 442 p.
5. Abubakar B. S. Developing a Java-based Genetic Algorithm to Solve the Travelling Salesman Problem. International Journal of Computer, Vol. 30, No 1, 2018, pp. 43–49.

6. Van Brummelen Glen. Heavenly Mathematics: The Forgotten Art of Spherical Trigonometry. Princeton : Princeton University Press, 2013. 222 p.

References

1. Smirnov I., Kosareva T. (2019) Transportna lohistyka. Navchal'nyy posibnyk [Transport logistics. Manual]. Kyiv: Center for Educational Literature. 224 p. [in Ukrainian].
2. Jacobson L., Kanber B. (2015) Genetic Algorithms in Java Basics. Apress. 154 p.
3. Wirsansky Eyal. (2020) Hands-On Genetic Algorithms with Python. Birmingham : Packt Publishing. 320 p.
4. Heaton J. (2008) Introduction to Neutral Network with Java. St. Louis : Heaton Research. 442 p.
5. Abubakar B. S. (2018) Developing a Java-based Genetic Algorithm to Solve the Travelling Salesman Problem. International Journal of Computer, Vol. 30, No 1, pp. 43–49.
6. Van Brummelen Glen. (2013) Heavenly Mathematics: The Forgotten Art of Spherical Trigonometry. Princeton : Princeton University Press. 222 p.

Дата першого надходження рукопису до видання: 28.09.2025
Дата прийнятого до друку рукопису після рецензування: 23.10.2025
Дата публікації:

UDC 004.93:616.12:618.3 DOI

H. FRANCHEVSKA
Postgraduate Student at the Biotechnical Systems Department
Ternopil Ivan Puluj National Technical University
ORCID: 0000-0001-6744-3015

V. DOZORSKYI
PhD, Associate Professor,
Associate Professor at the Biotechnical Systems Department
Ternopil Ivan Puluj National Technical University
ORCID: 0009-0001-8702-2821

ADDITIVE VS ADDITIVE–MULTIPLICATIVE MODELS FOR NONINVASIVE FETAL ELECTROCARDIOGRAPHY: CASCADE EXTRACTION METHODS AND NOISE ROBUSTNESS

Transcutaneous fetal electrocardiogram (ECG) recording is considered a task of separating components in a mixed abdominal signal containing maternal ECG (MECG), fetal ECG (FECG), and noise. The article systematically compares two formulations – additive (linear superposition) and additive-multiplicative – in the time and frequency domains, including parameter identification and component separation procedures, as well as the assessment of resistance to typical interference and recording artifacts. It is shown that for most practical scenarios, an additive model with a processing cascade is adequate: zero-phase band filtering and amplitude normalization; adaptive suppression of the maternal component (LMS/RLS in the presence of a chest reference or ICA for multichannel recordings); local isolation of informative details using wavelet and multiresolution analysis (CWT/DWT, MRA); final refinement by weighted least squares. When cross-terms cause combinatorial frequencies, it is advisable to switch to an additive-multiplicative model, detecting nonlinearities through bispectral analysis and selectively suppressing them in the time-frequency representation. If spectral power estimates are available, a Wiener filter is effective. Quality is assessed by SNR, relative reproduction error (RRE), and QRS detection metrics (sensitivity, positive predictive value). The proposed cascade scheme minimizes overfiltering, preserves QRS morphology, and increases noise immunity to isoline drift, network interference, motion/contact, and myogenic artifacts.

Key words: *FECG, MECG, additive model, additive-multiplicative model, LMS, ICA, DWT/CWT, MRA, Wiener filter, bispectrum, noise, baseline drift, motion artifacts.*

Г. ФРАНЧЕВСЬКА
аспірантка кафедри біотехнічних систем
Тернопільський національний технічний університет імені Івана Пулюя
ORCID: 0000-0001-6744-3015

В. ДОЗОРСЬКИЙ
кандидат технічних наук, доцент,
доцент кафедри біотехнічних систем
Тернопільський національний технічний університет імені Івана Пулюя
ORCID: 0009-0001-8702-2821

АДИТИВНА ТА АДДИТИВНО-МУЛЬТИПЛІКАТИВНА МОДЕЛІ ДЛЯ НЕІНВАЗИВНОЇ ЕЛЕКТРОКАРДІОГРАФІЇ ПЛОДА: МЕТОДИ КАСКАДНОЇ ЕКСТРАКЦІЇ ТА ЗАВАДОСТІЙКІСТЬ

Запис черезшкірної електрокардіограми (ЕКГ) плода розглядається як завдання розділення компонентів у змішаному абдомінальному сигналі, що містить ЕКГ матері (МЕКГ), ЕКГ плода (ФЕКГ) і шуму. У статті систематично порівнюються два способи подання таких сигналів – як адитивна (лінійна суперпозиція) та адитивно-мультиплікативна суміш – у часовій та частотній областях, включаючи процедури ідентифікації параметрів та розділення компонентів, а також оцінку стійкості до типових перешкод та артефактів запису. Показано, що для більшості практичних сценаріїв адекватною є адитивна модель з каскадом обробки: фільтрація нульової фази та нормалізація амплітуди; адаптивне придушення материнської складової (LMS/RLS за наявності опорного сигналу з грудної клітки або ICA для багатоканальних записів); локальна ізоляція інформативних ознак за допомогою вейвлет-аналізу та аналізу з кількома роздільними здатностями (CWT/DWT, MRA); остаточне уточ-

© Franchevska H., Dozorskyi V., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

нення за допомогою зважених найменших квадратів. Коли перехресні члени викликають комбінаторні частоти, доцільно перейти до адитивно-мультиплікативної моделі, виявляючи нелінійності за допомогою біспектрального аналізу та вибірково придушуючи їх у часово-частотному представленні. Якщо доступні оцінки спектральної густини потужності, ефективним буде застосування фільтра Вінера. Якість оцінюється за співвідношенням сигнал/шум (SNR), відносною помилкою відтворення (RRE) та метриками виявлення комплексу QRS (чутливість, позитивна прогностична цінність). Запропонована каскадна схема мінімізує надмірну фільтрацію, зберігає морфологію комплексу QRS та підвищує завадостійкість до дрейфу ізоліній, мережевих перешкод, руху/контакту та міогенних артефактів.

Ключові слова: *ФЕКГ, МЕКГ, адитивна модель, адитивно-мультиплікативна модель, LMS, ICA, DWT/CWT, MRA, фільтр Вінера, біспектр, шум, дрейф ізолінії, артефакти руху.*

Introduction

Extraction of the fetal electrocardiogram (FECG) is of primary importance in the field of monitoring the condition of the fetus and prenatal diagnosis, as it makes it possible to assess the state of the cardiovascular system of the fetus in a non-invasive way. The problem is that (FECG) signals are quite strongly hidden in the maternal electrocardiographic (MECG) signals, which makes the problem of their extraction one of the most difficult in the field of biomedical engineering. Successful extraction of the fetal signal is not only fundamental, but also critically important, as it allows for the early detection of congenital heart defects, arrhythmias and fetal distress, which often indicate the presence of hypoxia or other life-threatening conditions. The FECG signal, unlike the adult ECG signal, is subject to a wide range of external and internal factors that affect its clarity and amplitude.

Continuous monitoring of fetal condition is a decisive factor in reducing perinatal mortality. According to WHO estimates, there are approximately 2.6 million stillbirths each year, a significant proportion of which could be prevented through timely and accurate monitoring [1]. Non-invasive FECG signal monitoring provides continuous, real-time assessment of fetal well-being, aligning with the WHO's global mission [2] and offering a logical response to the limitations of Doppler and intermittent auscultation, which provide only episodic measurements and may miss subtle signs of distress [3]. The WHO's goal of reducing stillbirths to <12 per 1,000 births by 2030 underscores the need for widespread adoption of advanced monitoring techniques, including FECG signal [4]. The advantages of this approach include continuity, non-invasiveness, and better detection of subtle abnormalities in fetal cardiac activity (in particular through analysis of QRS complexes and heart rate variability, HRV) [5], [6], as well as the potential to reduce inequalities in access to quality prenatal care [7], which together contribute to improved perinatal outcomes [8].

Formulation of the research objective

The objective of the article is to justify the choice of the mathematical describing the fetal ECG signal and methods for its processing based on a study of the main advantages and disadvantages of the classes of additive and additive-multiplicative models.

Additive model: basic formulation

The observed abdominal signal is presented as a linear superposition of MEX, attenuated FEX, and noise:

$$x_{abd}(t) = m(t) + \alpha(t)f(t) + n(t), \tag{1}$$

Here, m(t) is MECG, f(t) is FECG, α(t) is the time-varying attenuation/scaling coefficient, and n(t) is noise. To reflect frequency-dependent attenuation and phase shifts introduced by biological tissues, time-varying transfer functions $H_m(t)$ and $H_f(t)$) are introduced with the convolution formulation:

$$(t) = \int_{-\infty}^{\infty} H_m(\tau,t)m(t-\tau)\mathrm{d}\tau + a_f \int_{-\infty}^{\infty} H_f(\tau,t)f(t-\tau)\mathrm{d}\tau + n(t), \tag{2}$$

In the frequency domain

$$S(f) = H_m(f)M(f) + \alpha_f H_f(f)F(f) + N(f), \tag{3}$$

where $M(f)$, $F(f)$, $N(f)$ are the Fourier transforms of MECG, FECG, and noise, respectively. Typically, MECG energy is concentrated in the range of 0.5–2 Hz, while FECG is in the range of 1–5 Hz (110–160 beats per minute), which is the basis for band selection.

Noise n(t) is often modeled as a Gaussian process with zero mean:

$$n(t) \sim N(0, \sigma^2), \tag{4}$$

For impulse noise, the Laplace model is used:

$$p(n) = \frac{1}{2b}\exp\left(-\frac{|n|}{b}\right), \tag{5}$$

and for non-stationary conditions, time-varying variance is used:

$$n(t) \sim N(0, \sigma^2(t)). \quad (6)$$

This formulation is consistent with real recordings, which contain 50/60 Hz line noise, motion artifacts, baseline drift, electrode noise, and myogenic noise.

Additive-multiplicative model: accounting for interactions

If there are noticeable interactions between components, a quadratic term is added to the additive terms:

$$s_{obs}(t) = A_m s_m(t) + A_f s_f(t) + B_m S_m(t) s_f(t) + n(t), \quad (7)$$

The multiplicative term generates sum-difference frequencies. Representing $s_m(t)$ and $s_{mf}(t)$ as finite Fourier series:

$$s_m(t) = \sum_{k=1}^{K} A_k \cos(\omega_k t + \phi_k), \quad (8)$$

$$s_f(t) = \sum_{l=1}^{L} B_l \cos(\omega_l t + \psi_l), \quad (9)$$

the product $s_m(t)\, s_f(t)$ decomposes into cosines with frequencies $\phi_k + \psi_l$:

$$s_m(t) s_f(t) = \sum_{k=1}^{K}\sum_{l=1}^{L} \frac{A_k B_l}{2} \cos\left((\omega_k + \omega_C)t + (\phi_k + \psi_l)\right) + \cos\left((\omega_k - \omega_C)t + (\phi_k - \psi_l)\right), \quad (10)$$

In the frequency domain, the convolution of spectra yields peaks at $\omega_k + \omega_l$, which complicates pure band separation. The spectral power density (PSD) of the observed signal has the form:

$$s_{m\cdot f}(\omega) = B_m \sum_{k=1}^{K}\sum_{l=d}^{L} \frac{A_k B_l}{2} [\delta(\omega - (\omega_k + \omega_l)) + \delta(\omega - (\omega_k - \omega_l))], \quad (11)$$

Identification of parameters and separation of components

Additive model

Adaptive filtering (LMS). Using the reference estimate MEKS $\widehat{m}(t)$, the error $e(t)$ is minimized:

$$e(t) = S(t) - \widehat{m}(t) = \alpha_f f(t) + n(t), \quad (12)$$

updating the weights $w(t)$:

$$w(t+1) = w(t) + \mu e(t) S(t), \quad (13)$$

The residual $S(t)$ contains FECS and noise.

Independent component analysis (ICA). For multichannel recordings

$$S(t) = A \cdot X(t), \quad (14)$$

$$X(t) = W \cdot S(t), \quad (15)$$

where $X(t)$ are independent sources (MECS, FECS). The goal is to find W that maximizes the independence of the components.

Time-frequency analysis. Continuous wavelet transform:

$$C(a,b) = \int_{-\infty}^{\infty} S(t) \psi^* \left(\frac{t-b}{a}\right) dt, \quad (16)$$

and discrete wavelet decomposition DWT performs multilevel decomposition

$$S(t) \to \{A_1, D_1\}, \{A_2, D_2\}, \ldots, \{A_n, D_n\}, \quad (17)$$

At lower levels (low frequencies), MECG dominates, while at higher levels (1–5 Hz), FECG dominates. Multiresolution analysis (MRA)

$$S(t) = A_n(t) + \sum_{i=1}^{n} D_i(t), \quad (18)$$

allows us to reconstruct FECS, preserving only details from the k-th level:

$$f(t) \approx \sum_{i=k}^{n} D_i(t), \quad (19)$$

Wavelet denoising. Thresholding of coefficients

$$\tilde{C}(a,b) = \begin{cases} C(a,b) & if\ |C(a,b)| > \lambda \\ 0 & else \end{cases}, \quad (20)$$

suppresses impulse/high-frequency noise while preserving the QRS morphology. Wavelet families suitable for ECG are selected based on their similarity to the QRS complex and the "time-frequency" compromise.

Working with non-stationarity. Adaptive filters (LMS/RLS) and Kalman filters are used for motion and drift artifacts; baseline correction can be performed by HF filtering or polynomial fitting.

Additive-multiplicative model

The presence of cross terms ($\omega_k + \omega_l$) is detected using higher spectral moments – bispectrum analysis. To isolate the FEC in this case, two steps are helpful:

Fundamental separation of the additive part (DWT/MRA, LMS, or ICA),

Suppression of cross-terms by selective processing in the frequency or time-frequency domain (masking of peaks $\omega_k + \omega_l$).

Additionally, consider a Wiener filter in the frequency domain to minimize the root mean square error between the estimated and desired FECS:

$$H_f(\omega) = \frac{P_{s_f}(\omega)}{P_{s_f}(\omega) + A_m^2 P_{s_m}(\omega) + B_m^2 \left(P_{s_m}(\omega) \cdot P_{s_f}(\omega)\right) + P_N(\omega)}, \tag{21}$$

The time-frequency representation (CWT) with the ψ kernel allows local processing of non-stationary cross-effects:

$$W_{obs}(a,b) = \int_{-\infty}^{+\infty} s_{obs}(t) \psi\left(\frac{t-b}{a}\right) \mathrm{d}t, \tag{22}$$

Sensitivity to noise and artifacts

The effectiveness of ECG extraction from abdominal recordings is determined by the processing cascade's ability to suppress typical interferences, including low-frequency baseline drift, high-frequency noise, residual network components, motion/contact artifacts, and myogenic noise, as well as combination frequencies arising from additive-multiplicative interactions. The baseline and HF noise are suppressed by bandpass zero-phase filtering, after which the amplitude is stabilized by normalization; non-stationary noise is compensated by adaptive subtraction of the LMS mother component, and local emissions reduce the impact through weighted least squares. If necessary, bispectral analysis is used to detect nonlinear cross-correlations.

Bandpass filtering isolates the informative range of the fECG and simultaneously cuts off drift (<1 Hz) and HF noise. The ideal prototype is described by:

$$H_{bp}(f) = \begin{cases} 1, f_L \le f \le f_U \\ 0, \quad \textit{інакше} \end{cases}, \tag{23}$$

and structurally use a zero-phase Butterworth implementation equivalent to the product of two "slices":

$$H_{bp}(f) = \frac{1}{\sqrt{1 + \left(\frac{f}{f_L}\right)^{2n}}} \cdot \frac{1}{\sqrt{1 + \left(\frac{f}{f_U}\right)^{2n}}}, \tag{24}$$

Frequency composition control is performed via the Fourier transform:

$$X(f) = \int_{-\infty}^{+\infty} x(t) e^{-j2\pi ft} \mathrm{d}t, \tag{25}$$

and time distortions are tracked by group delay:

$$\tau_g(f) = -\frac{\mathrm{d}\phi(f)}{\mathrm{d}f}, \tag{26}$$

which zero-phase filtering effectively eliminates while preserving the QRS morphology.

To prevent subsequent stages from being sensitive to amplitude variations, single or robust normalization is applied:

$$A_{\max} = \max(|x_{bp}(t)|), \tag{27}$$

$$x_{norm}(t) = \frac{x_{bp}(t)}{A_{\max}}, \tag{28}$$

$$A_{robust} = P_{95}(|x_{bp}(t)|), \tag{29}$$

where P_{95} – 95th percentile. Such scale stabilization accelerates and stabilizes the convergence of adaptive filters, simplifying thresholding during peak detection.

Adaptive suppression of the mother component is performed by an LMS filter with the chest ECG as the reference channel. The output of the filter is of order N:

$$y(\mathrm{i}) = \sum_{j=0}^{N-1} \omega_j \cdot x(\mathrm{i} - j), \tag{30}$$

The error $e(i)$ at each moment in time i is defined as:

$$e(i) = d(i) - y(i), \tag{31}$$

minimizes the root mean square error:

$$J = E[e(i)^2] = E[(d(i) - y(i'))], \tag{32}$$

To do this, the LMS algorithm updates the filter weights in the direction opposite to the gradient J with respect to ω:

$$\omega_{i+1} = \omega_i + 2\mu e(i)x(i), \tag{33}$$

Small μincreases resistance to noise and motion artifacts, while large values can cause oscillations and loss of convergence.

In the context of FECG signal extraction, least squares optimization serves as a final refinement tool based on the results of previous filtering steps.

The goal of least squares optimization is to minimize the quadratic error between the observed signal $y(i)$ (after adaptive filtering) and the predicted fetal ECG $\hat{f}(i)$. This can be formulated as a loss function J:

$$J\left(\hat{f}(t)\right) = \sum_{i=1}^{N} \left(y(i) - \hat{f}(i)\right)^2, \tag{34}$$

The least squares method finds the vector β that minimizes the sum of the squares of the residuals $e^T e$:

$$\beta = (X^T X)^{-1} X^T y, \tag{35}$$

And to reduce the impact of artifacts, a weighted form is used:

$$J_{WLC}\left(\hat{f}(t)\right) = \sum_{i=1}^{N} \omega_i \left(y(i) - \hat{f}(i)\right)^2, \tag{36}$$

$$X_\omega = W^{\frac{1}{2}} X,\ y_\omega = W^{\frac{1}{2}} y, \tag{37}$$

$$\beta_{WLC} = \left(X_w^T X_\omega\right)^{-1} X_\omega^T y_\omega, \tag{38}$$

Local emissions are identified and corrected according to the interquartile range rule:

$$IQR = Q3 - Q1. \tag{39}$$

In the presence of combinatorial components (sum/difference frequencies) resulting from additive-multiplicative interactions, third-order bispectral analysis is functional:

$$B(\omega_1, \omega_2) = E[s(\omega_1)s(\omega_2)s \cdot (\omega_1 + \omega_2)], \tag{40}$$

which localizes phase-dependent interactions for subsequent selective suppression (e.g., masking in the time-frequency representation). The ratio of the spectra controls the basic frequency balance of the mixture:

$$s_{obs}(\omega) = s_m(\omega) + s_f(\omega) + N(\omega), \tag{41}$$

Qualitative noise immunity is evaluated using energy and time metrics, including the signal-to-noise ratio (as the ratio of the root mean square powers of the estimated FECG and residual noise after subtraction), relative reproduction error (RRE), and QRS detection indicators (sensitivity and positive predictive value). The working sequence is as follows. First, zero-phase bandpass filtering is performed to suppress drift and high-frequency noise. Next, unit or robust normalization is applied to stabilize the signal scale. After that, adaptive subtraction is used according to the LMS algorithm with a correctly selected μ parameter to suppress the mother component. The next step is to refine the waveform using the least squares or weighted least squares method to reduce the influence of locally noisy areas. If there are signs of additive-multiplicative interactions, bispectral detection of combination frequencies and selective suppression of the corresponding components are additionally performed. This cascaded organization minimizes overfiltering, preserves the morphology of the QRS complex, and reduces sensitivity to typical noise and artifacts.

Conclusions

In conclusion, the comparison of additive and additive-multiplicative approaches to the problem of isolating FECG from abdominal recordings revealed that linear superposition is a sufficiently accurate model for most practical scenarios,

providing consistent processing in both the time and frequency domains, based on the frequency separation of MECG and FECG. At the same time, the presence of cross-interactions between the maternal and fetal components, which generate combinatorial (sum-difference) frequencies, requires a transition to an additive-multiplicative model, since pure band separation becomes insufficient. For such cases, it is advisable to use bispectral analysis to identify nonlinear terms and their subsequent selective suppression in the time-frequency representation. If spectral powers are estimated, a Wiener filter is also helpful as a solution, minimizing the root mean square error.

The effectiveness of cascade processing, which begins with zero-phase bandpass filtering (to suppress baseline drift and high-frequency noise without distorting the QRS morphology) and amplitude normalization (conventional or robust), has been experimentally demonstrated. Further adaptive subtraction of the maternal component (LMS/RLS in the presence of a chest reference or ICA for multichannel recordings) in combination with wavelet and multiresolution analysis (CWT/DWT, MRA) provides local isolation of informative FECG details. Final refinement using the least squares method, particularly in a weighted setting, reduces the impact of local emissions and contact/motion artifacts. The proposed sequence minimizes the risk of over-filtering, increases resistance to typical interference (including network components and myogenic noise), and preserves diagnostically significant morphology.

Quality assessment should be performed comprehensively, considering the signal-to-noise ratio, relative reproduction error, and QRS detection metrics (sensitivity and positive predictive value). The practical recommendation is that the additive model, together with the described processing cascade, should be considered the standard choice in routine monitoring conditions. The additive-multiplicative model should be used when there are convincing signs of nonlinear cross-interactions, confirmed by the bispectrum. Prospects for further research include adapting filtering parameters in online mode using Bayesian/Kalman estimators and standardizing validation protocols on open datasets with well-defined noise and artifact scenarios.

References

1. World Health Organization. Birth defects. 2021. URL: https://www.who.int/news-room/factsheets/detail/birth-defects. Accessed on: August 26, 2025

2. Global Stillbirth Policy Review – Outcomes And Implications Ahead of the 2030 Sustainable Development Goal Agenda / N. A. Mensah Abrampah et al. International Journal of Health Policy and Management. 2023. Vol. 12. Iss. 1. Pp. 1–11. DOI: https://doi.org/10.34172/ijhpm.2023.7391

3. Aminu M., van den Broek N. Stillbirth in low- and middle-income countries: addressing the «silent epidemic». International Health. 2019. Vol. 11. № 4. Pp. 237-239. DOI: https://doi.org/10.1093/inthealth/ihz015

4. Facility-based stillbirth review processes used in different countries across the world: a systematic review / Y. Y. Boo et al. eClinicalMedicine. 2023. Vol. 59. Pp. 1-17. DOI: https://doi.org/10.1016/j.eclinm.2023.101976

5. Stillbirths: Contribution of preterm birth and size-for-gestational age for 125.4 million total births from nationwide records in 13 countries, 2000–2020 / Y. B. Okwaraji et al. BJOG: An International Journal of Obstetrics & Gynaecology. 2023. Pp. 1–12. DOI: https://doi.org/10.1111/1471-0528.17653

6. Noninvasive fetal electrocardiography for the detection of fetal arrhythmias / J. A. Behar et al. Prenatal Diagnosis. 2019. Vol. 39. № 3. Pp. 178–187. DOI: https://doi.org/10.1002/pd.5412

7. World Health Organization. Stillbirth. 2019. URL: https://www.who.int/health-topics/stillbirth accessed on: August 26, 2025

8. A deep learning framework for noninvasive fetal ECG signal extraction / M. Wahbah et al. Frontiers in Physiology. 2024. Vol. 15. Pp. 01–13. DOI: https://doi.org/10.3389/fphys.2024.1329313

9. Jaros R., Tomicova E., Martinek R. Template subtraction-based methods for non-invasive fetal electrocardiography extraction. Scientific Reports. 2024. Vol. 14. № 1. DOI: https://doi.org/10.1038/s41598-024-51213-5

10. Vilimkova Kahankova R., Martinek R., Bilik P. Non-invasive Fetal ECG Extraction from Mater-nal Abdominal ECG Using LMS and RLS Adaptive Algorithms. AECIA 2016: Proceedings of the Third International Afro-European Conference for Industrial Advancement – AECIA 2016, Marrakesh, Morocco, 21–23 November 2016, pp. 258–271. DOI: https://doi.org/10.1007/978-3-319-60834-1_27

11. World Health Organization. WHO recommendations: intrapartum care for a positive childbirth experience. 2018. URL: https://academic.oup.com/inthealth/article/11/4/237/5488741. Accessed on: August 26, 2025.

12. Prasanth K., Paul B., Balakrishnan A. A. Fetal ECG Extraction Using Adaptive Filters. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering. 2012. Vol. 2. № 4. Pp. 1483–1487.

Дата першого надходження рукопису до видання: 18.09.2025
Дата прийнятого до друку рукопису після рецензування: 15.10.2025
Дата публікації:

УДК 004.052.3:004.94 DOI

Д. Є. ФРОЛОВ
аспірант, асистент кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0009-0009-3291-3561

МОДЕЛЬ ЗАБЕЗПЕЧЕННЯ ВІДМОВОСТІЙКОСТІ ІНФОРМАЦІЙНИХ СИСТЕМ НА ОСНОВІ БАГАТОРІВНЕВИХ СТРУКТУР

У статті сформульовано задачу забезпечення відмовостійкості сучасних динамічних інформаційних систем на основі багаторівневих структур, як важливу умову її стабільного функціонування. Обґрунтовано необхідність багаторівневого підходу, оскільки класичні методи резервування та дублювання не забезпечують комплексної стійкості, особливо у випадках міжрівневих взаємодій. В якості розв'язання поставленої задачі запропоновано використовувати інтегративний підхід в частині подання рівнів інформаційної системи (фізичний, мережний, прикладний, організаційний та рівень віртуалізації), її функціоналів разом із процесною інтерпретацією, що їх супроводжують у разі виникнення відмов та відновлень. У якості описового апарату використано логіку Хоара, яка оперує з кортежним представленням, власне, запропонованої моделі. Базовий кортеж включає залежності та умови, зокрема, дефіцит ресурсів інформаційної системи на підтримку задовільного рівня виконання процесів, впливів міжрівневих взаємозв'язків. Працездатність окремих елементів інформаційної системи, узгодженості процесів між елементами, у тому числі на міжрівневих площинах, на прикладі моделі, може бути оцінена на підставі критеріального аналізу, що використовує показники доступності, відмовостійкості, ізоляції та ресурсно-вартісні характеристики інформаційної системи. Проведено експериментальні дослідження запропонованої моделі. Зокрема, досліджено встановлення відповідності структурним припущенням і процесному опису властивостей реальних систем запропонованій моделі. Також досліджено коректність прогнозів щодо збереження працездатності критичних функціоналів системи за стохастичних відмов і керованих політик відновлення. Дослідження проведені на базі обчислювальних потужностей інформаційно-обчислювального центру Харківського національного університету радіоелектроніки. Результати досліджень наведені у статті та є задовільними – модель є адекватною.

Ключові слова: *інформаційна система, модель, відмовостійкість, багаторівневість, логіка Хоара, математичне моделювання.*

D. YE. FROLOV
Postgraduate Student, Assistant at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0009-0009-3291-3561

MODEL FOR ENSURING FAULT TOLERANCE OF INFORMATION SYSTEMS BASED ON MULTI-LEVEL STRUCTURES

The article formulates the problem of ensuring fault tolerance in modern dynamic information systems based on multilevel structures as an important condition for their stable functioning. The necessity of a multilevel approach is substantiated, since classical methods of reservation and duplication do not provide comprehensive resilience, especially in cases of interlevel interactions. To solve this problem, an integrative approach is proposed that encompasses the representation of the levels of an information system (physical, network, application, organizational, and virtualization), its functionalities, and the process interpretation that accompanies them in the event of failures and recoveries. Hoare logic, operating with a tuple-based representation of the proposed model, serves as the descriptive apparatus. The basic tuple incorporates dependencies and conditions, including resource shortages within the information system that hinder the maintenance of an acceptable level of process execution and the influence of interlevel interactions. The operability of individual components of the information system and the consistency of processes across elements, including across interlevel planes, can be evaluated within the model using a criteria analysis based on indicators of availability, fault tolerance, isolation, and the resource-cost characteristics of the system. The paper presents experimental studies of the proposed model. In particular, the research examined the correspondence between the structural assumptions and the process description of real system properties and the proposed model. This study examined the validity of the model's predictions concerning the preservation of critical system functionalities under stochastic failures and controlled recovery policies. The research used the computing resources of the Information and Computing Center of Kharkiv National University of Radio Electronics. The results, presented in the article, demonstrate satisfactory outcomes, confirming the adequacy of the model.

Key words: *information system, model, fault tolerance, multilevel structure, Hoare logic, mathematical modeling.*

© Фролов Д. Є., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

Постановка проблеми

Відмовостійкість інформаційної системи є базовою характеристикою її надійності, що визначає здатність системи підтримувати працездатність за наявності відмов окремих компонентів [1–5]. В умовах сьогодення цільові завдання інформаційних систем виходять за межі традиційних ІТ-сервісів і дедалі частіше набувають синергетичного характеру, формуючи основу безперервності бізнес-процесів і забезпечуючи керування критично важливими технологічними процесами у сферах оборони та безпеки [6–8].

Останнім часом теорія побудови сучасних інформаційних системи все більше ґрунтується на розгляді інформаційної системи як певної багаторівневої сутності, яка органічно враховує процесність від рівня апаратно-мережного шару до прикладних сервісів [9]. Прикладні аспекти застосовності таких інформаційних систем на основі багаторівневих структур враховують умови їх функціонування як то: стохастичні відмови елементів системи, приховані дефекти, мережні аномалії тощо [10–12]. За зазначених умов поняття відмовостійкості та апарат, який за ним стоїть, відносно інформаційних систем на основі багаторівневих структур – є основною складовою комплексної надійності інформаційної системи [13].

Класичні прикладні рішення задачі забезпечення відмовостійкості інформаційних систем (резервування, протоколи узгодження, самовідновлювальні механізми) – ефективні у межах окремих інформаційних систем або її елементів, однак їх композиційне поєднання відносно багаторівневої інформаційної системи породжує непередбачувані ефекти, які складно прогнозувати, а час на їх усунення може вплинути на роботоздатність всієї системи. До таких ефектів можуть відноситися, наприклад, каскадні відмови, розрив зв'язків між рівнями тощо [14].

Таким чином, класичні підходи дозволяють здійснювати: агрегацію показників доступності в межах одного рівня багаторівневої інформаційної системи для прийняття системою рішень щодо подальшого сценарію функціонування або задавати множини рішень для керуючого впливу на окремі механізми узгодження й реплікації у міжрівневій архітектукрі інформаційної системи, залишаючи, власне міжрівневі залежності поза уваги. Це створює передумови до необхідності розробки моделі, яка б одночасно враховувала ієрархію рівнів та зв'язки між ними в інформаційній системі, дозволяла здійснювати композиційне оцінювання показника відмовостійкості та містила інтерфейс для застосування темпорального критерію оптимальності конфігурацій у певних проміжках її застосовності при стабільних показниках зовнішнього впливу.

Методологічна проблема полягає у наявних міжрівневих розривах інформаційної системи на основі багаторівневих структур: навіть за умови забезпечення відмовостійкості на окремих рівнях, зберігається вразливість на вузлах взаємодії між ними. Взаємозалежності різних рівнів нерідко залишаються поза єдиною узгодженою політикою, що призводить до несумісності прикладних підходів підвищення відмовостійкості інформаційних систем. Як наслідок, локально коректні рішення не узгоджуються композиційно в системну відмовостійкість інформаційної системи, а міжрівневі ефекти здатні зводити нанівець переваги захисних механізмів усередині окремих рівнів. Все це є однією із причин розробки моделі відмовостійкості інформаційної системи на основі багаторівневих структур.

Для формулювання задачі дослідження, розглянемо інформаційну систему на основі багаторівневих структур у вигляді наступного кортежу:

$$S = \left\langle L, \rho, \{\Sigma_i\}_{i\in L}, \{E_i\}_{i\in L}, \Upsilon \right\rangle, \tag{1}$$

де $L = \{0, \ldots, k\}$ – множина рівнів; ρ – відношення впорядкування рівнів в інформаційній системі ($\rho \subseteq L_k \times L_{k+1}$, $0 \le k \le n$, $n = 1, 2, 3, \ldots$); Σ_i – простір станів рівня i; E_i – множина елементарних подій; Υ – інтерфейси міжрівневого впливу.

Додатково необхідно ввести такі сутності, що уточнюють траєкторію функціонування багаторівневої структури:

– F – множина функціоналів, що реалізується на базі структури S. Кожна функція $f \in F$ реалізується на основі станів та подій одного чи кількох рівнів інформаційної системи на основі багаторівневих структур:

$$f : \left(\Sigma_{i1} \times \ldots \times \Sigma_{im}\right) \to \Sigma_j, \quad i_1, \ldots, i_m,\ j \in L, \tag{2}$$

тобто функціонал спирається на підмножину рівнів і чинить вплив на результат роботи інформаційної системи на іншому рівні;

– F_{crit} – множина критичних функціоналів ($F_{crit} \in F$), яка включає критичні функції, для яких необхідно гарантувати відмовостійкість навіть за умов часткових відмов елементів інформаційної системи на основі багаторівневих структур;

– $G = (F, \Xi)$– граф залежностей функціоналів, в якому дуги $(f_p, f_q) \in \Xi$встановлюють залежність функціоналу f_p від роботоздатності f_q. Такий граф може формалізувати залежності між функціоналами рівнів;

– $\varphi_f(t) \in \{0, 1\}$ – функція доступності функціоналу f, яка визначає стан доступності в момент часу t для функціоналів з множин F та F_{crit};

– $R_f(T)$ – функція надійності, яка відображає імовірність безвідмовної роботи інформаційної системи на основі багаторівневих структур на інтервалі часу $[0, T]$;

– ζ – конфігураційна сутність, яка визначається параметрами інформаційної системи в залежності від особливостей її функціонального призначення;

– $N(\zeta)$– ресурсно-вартісна функція, яка характеризує витрати на підтримку конкретної конфігурації ζ інформаційної системи на основі багаторівневих структур;

– $X_h(\zeta)$ – показники якості виконання основної функції інформаційної системи, зокрема $h = \{D_F(\zeta), W_F(\zeta), V_F(\zeta)\}$, де ($D_F(\zeta)$ – затримка виконання функції, $W_F(\zeta)$– пропускна здатність під час міжрівневої взаємодії функціоналів, $V_F(\zeta)$– узгодженість у досяжності спільної функції);

Виходячи з вищенаведеного, задача забезпечення відмовостійкості інформаційної системи на основі багаторівневих структур може бути сформульована як синтез конфігурацій ζ, що враховує систему обмежень для конфігурацій інформаційної системи на основі багаторівневих структур може бути записана у вигляді:

$$\Omega = \begin{cases} \forall f \in F_{crit} \forall t \in [0,T]: \ \varphi_f(t) = 1; \\ \zeta \ \textit{узгоджена з} \ \left(L, \rho, \{\Sigma_i\}_{i\in L}, \{E_i\}_{i\in L}, \Upsilon\right); \\ \forall f \in F_{crit}: \mathrm{X}_h(\zeta) \geq X_h^{\min}(\zeta); \\ \min\limits_{\zeta} \mathrm{X}_h(\zeta) \ \textit{при фіксованих} \ R_f(T). \end{cases} \tag{3}$$

де: $X_h^{\min}(\zeta)$ – мінімально необхідні умови забезпечення якості виконання функціоналів *h*.

Аналіз останніх досліджень і публікацій

У праці [15] запропоновано модульний підхід до моделювання систем на основі багатоорендних сервісів із використанням багаторівневих мереж Петрі (*n*-NLS). Автори фокусуються на формалізації компонентів і рівнів системи, визначенні правил синхронізації та координації між ними, а також на забезпеченні диференціації продуктивності для орендарів відповідно до угод про рівень обслуговування та планів варіативності. Такий процесний формалізм зменшує неоднозначності в описі поведінки та підвищує керованість багатоорендного середовища. Водночас, робота не декларує синтезу конфігурацій з урахуванням обмежень надійності й якості.

У праці [16] представлено техніку PreLoFT, зокрема прогнозний розподіл навантаження між fog-вузлами з розщепленням ресурсоємних задач і оперативним перерозподілом у разі відмов. Підхід зосереджено на внутрішньому рівні fog без явної підтримки багаторівневості; не показано, як PreLoFT узгоджується з глобальними обмеженнями якості та ресурсно-вартісними вимогами системи. Тому робота [16] має науковий інтерес у контексті цієї роботи як приклад формулювання локальної евристики керування на рівні fog при розробці багаторівневої моделі відмовостійкості з формальними гарантіями.

У роботі [17] автори розглянули напрям багаторівневої відмовостійкості в IoT-системах на прикладі власної програмної розробки FaTEMa. Він базується на подієво-орієнтованій архітектурі та забезпечує обмін інформацією між шарами системи з метою прискорення виявлення і відновлення після збоїв, що дозволило покращити показники надійності й доступності. Втім, відсутність узагальненої моделі обмежує масштабованість їхнього підходу.

Розглянуті [15–17] праці репрезентують комплементарні підходи, але спільним є відсутність єдиної багаторівневої формальної структури з явно визначеними інтерфейсами та композиційними властивостями та формальної постановки синтезу конфігурацій під жорсткі обмеження доступності/надійності і показників якості. Отже, можна зазначити, що наразі відсутній композиційний апарат для інформаційних систем на основі багаторівневих структур, у якому локальні механізми відмовостійкості інтерпретуються як морфізми між рівнями з гарантованим забезпеченням відмовостійкості і можливістю верифікованого синтезу конфігурацій під задані обмеження. Саме це обґрунтовує актуальність роботи статті, яка присвячена побудові формальної багаторівневої моделі відмовостійкості з явними інтерфейсами та системою обмежень на рівні всієї інформаційної системи, а не лише окремих її рівнів чи методів їх взаємодії.

Формулювання мети дослідження

Метою дослідження є розробка моделі забезпечення відмовостійкості інформаційних систем на основі багаторівневих структур, яка враховує підпорядкованість (вкладеність) рівнів та функціонали на них, забезпечує композиційну верифікацію доступності, відмовостійкості та ізоляції за умов стохастичних відмов в інформаційній системі і керованих політик відновлення її функціональності, а також створює підґрунтя для оптимізації існуючих конфігурацій з урахуванням ресурсної вартості резервування елементів інформаційної системи.

Викладення основного матеріалу дослідження

Процесна формалізація моделі інформаційної системи на основі багаторівневих структур з використанням логіки Хоара. Застосування логіки Хоара обґрунтовано тим, що вона є одним із базових формалізмів опису та верифікації процесних систем, що дозволяє задавати коректність виконання дій у вигляді певного математичного апарату [18].

Для інформаційних систем на основі багаторівневих структур, у яких відмови та відновлення описуються як події, а взаємодія між рівнями має стохастичний і керований характер, логіка Хоара надає можливість задати

процесність моделі, зокрема відобразити, за яких умов виконання певної події чи переходу застосовується вибір режиму доступності або відмовостійкості. Таким чином, логіка Хоара виступає методологічним підґрунтям для композиційної перевірки властивостей у багаторівневій структурі, так як кожна локальна подія може бути формально описана через свої передумови та післяумови, а їхня композиція забезпечує цілісну верифікацію траєкторії функціонування інформаційної системи на основі багаторівневих структур при стохастичних відмовах і процесам по відновленню функціональної здатності.

Отже, беручи за основу кортеж (1), вихідний вигляд моделі інформаційної системи на основі багаторівневих структур можна записати як:

$$S = \left\langle L, \rho, \{\Sigma_i\}_{i\in L}, \{E_i\}_{i\in L}, \Upsilon, F, F_{crit}, G, \varphi_f(t), R_f(T), \zeta, C(\zeta), X_h(\zeta) \right\rangle, \tag{4}$$

при умовах, викладених у формулі (3).

Для процесного опису функціонування інформаційної системи введемо сутність події як $e \in E_i$, яка відображає процес відмови чи відновлення функціонування рівня чи всієї інформаційної системи (приналежно до індексу). В такому разі, маємо:

– у якості передумови P визначає поточний стан системи (наприклад, множину доступних функціоналів, дотримання обмежень Ω);

– команда Uвідповідає за такі події в системі як відмови чи відновлення;

– післяумова Q фіксує новий стан, у якому має виконуватися із забезпечення відмовостійкості (наприклад, доступність усіх $f \in F_{crit}$).

Звідси отримуємо:

$$\{P\}e\{Q\},\ e \in E_i, \tag{5}$$

де: $P : \Omega \wedge \bigwedge\limits_{f\in F_{crit}} \varphi_f(t) = 1;\ \ Q : \Omega \wedge \bigwedge\limits_{f\in F_{crit}} \varphi_f(t') = 1.$

Відповідно, у разі відмови, маємо:

$$\left\{\varphi_f(t) = 1, \forall f \in F_{crit}\right\}\ \ fail(i)\left\{\exists \zeta' \in Z : \varphi_f(t+\Delta) = 1,\ \forall f \in F_{crit}\right\}. \tag{6}$$

Це показує про досяжність і можливість використання такої конфігурації ζ', яка збереже роботу критичних функціоналів навіть при відмові рівня з точки зору розладу роботи його елементів.

А у разі відновлення роботи критичних елементів рівня i:

$$\left\{\neg\varphi_{f''}(t)\right\}\ \ rec(i)\left\{\exists\varphi_{f''}(t+\Delta) = 1\right\},\ \ f'' \in F_{crit}. \tag{7}$$

Це показує те, що після відновлення рівня, критичний функціонал знову стає доступним.

Нарешті, композиція подій матиме вигляд:

$$\{P\}U_1 : U_2\{Q\},\ \text{якщо}\ \{P\}U_1\{R\},\quad \{R\}U_2\{Q\}. \tag{8}$$

Отже, маємо апарат, який дозволяє описати траєкторію відмови та відновлення в межах кількох рівнів і довести, що система виконує Ω.

Апарат, який описує міжрівневі процеси, аналогічно (6)–(8), виходячи з того, що $(i, j) \in L$, де $i \prec j$ у межах відношення (4) ρ, а подія на рівні i може змінювати доступність функціоналів на рівні j, то:

$$\{P_i \wedge P_j\}U_{i\to j}\{Q_i \wedge Q_j\}, \tag{9}$$

де P_i – умови стану рівня i перед подією; P_j – умови стану рівня j, що залежать від i; $U_{i\to j}$ – подія міжрівневого впливу; Q_i, Q_j – післяумови для обох рівнів.

Якщо критичний функціонал рівня i недоступний, то після поширення відмови втрачається працездатність пов'язаного функціоналу на рівні j. Відповідно, міжрівнева відмова має вид:

$$\left\{\varphi_{f_i}(t) = 0 \wedge \varphi_{f_j}(t) = 1\right\}\ \ levelfail(i \to j)\left\{\varphi_{f_j}(t+\Delta) = 0\right\}, \tag{10}$$

де $f_i \in F_{crit}(i), f_j \in F_{crit}(j)$.

Якщо рівень i відновлено, то функціонал на рівні j, залежний від i, також відновлюється. Відповідно, міжрівневе відновлення має вид:

$$\left\{\varphi_{f_i}(t) = 1 \wedge \varphi_{f_j}(t) = 0\right\}\ \ levelrec(i \to j)\left\{\varphi_{f_j}(t+\Delta) = 1\right\}. \tag{11}$$

Якщо рівень i забезпечує мінімально необхідні умови якості, то після виконання переходу гарантовано, що рівень j також залишається у допустимій області QoS:

$$\left\{X_i(\zeta) \geq X_i^{\min}\right\}\ ensure(i \to j)\left\{X_j(\zeta) \geq X_j^{\min}\right\}. \tag{12}$$

У такий спосіб можна для кожної залежності $(i, j) \in \rho$ формально задати передумови й післяумови, які здійснюють контроль поширення подій відмов / відновлень та узгодженість рівнів у багаторівневій структурі.

Таким чином, введення передумов і післяумов для міжрівневих переходів дає змогу фіксувати та описувати локальні події на окремих рівнях, відслідковувати динаміку впливу у межах усієї інформаційної системи на основі багаторівневих структур. Тобто, події відмов і відновлень задаються як процеси з наслідками у часі – стан на рівні *i* стає передумовою, а стан на рівні *j*, пов'язаний через ρ та ϒ – післяумовою. Завдяки цьому можна композиційно описати траєкторії за напрямом «відмова – поширення – відновлення». Це означає, що за будь-якої послідовності подій зберігаються умови доступності критичних функціоналів за умови виконання обмежень (3). Застосування логіки Хоара дозволило формалізувати динамічність поведінки інформаційної системи на основі багаторівневих структур, зокрема, розвиток подій у часі, їхній вплив між рівнями та можливість верифікації відмовостійкості через послідовності процесів.

Модель забезпечення відмовостійкості інформаційних систем на основі багаторівневих структур. Наступним кроком є побудова саме моделі забезпечення відмовостійкості, яка інтегрує структурний і процесний рівні, формалізує умови збереження працездатності критичних функціоналів та вводить засоби оцінки й оптимізації конфігурацій.

Модель забезпечення відмовостійкості інформаційної системи на основі багаторівневих структур можна представити у вигляді кортежу:

$$\Psi = \langle S, \Pi, O, \Xi \rangle, \tag{13}$$

де: S – структурно-процесна модель інформаційної системи, визначена раніше у (4). Вона є базовим елеентом моделі забезпечення відмовостійкості інформаційної системи на основі багаторівневих структур та використовується для складання опису середовища, в якому реалізується забезпечення відмовостійкості; П – множина політик відновлення, які визначають допустимі дії у відповідь на відмови. До таких дій відносяться, але не вичерпуються визначеним пулом дій – дублювання / реплікація, резервування на рівнях, міграція функціоналів, перезапуск функціоналів, використання альтернативних функціоналів. Кожну політику можна інтерпретувати як перетворення над S, що забезпечує збереження функціональної здатності інформаційної системи; О – множина варіативних траєкторій функціонування інформаційної системи. Ця множина описує пул сценаріїв реагування на події відмов і відновлень, сформованих на підставі застосуванням політик П. Таким чином, О є простором можливих процесних сценаріїв, які гарантують відновлення критичних функціоналів; Ξ – множина критеріїв забезпечення, яка включає: доступність критичних функціоналів у часі, відмовостійкість як здатність зберігати працездатність за множини відмов, ізоляцію збоїв (обмеження поширення каскадних відмов у та між рівнями), ресурсно-вартісні характеристики та якісні показники визначені експертом (за наявності такого функціоналу).

Функціонально модель має можливість композиційної верифікації. Відповідно, замість перевірки всієї інформаційної системи як цілісного утворення, доцільно виконувати локальну верифікацію подій і політик на рівнях, а потім за формальними правилами композиції робити висновок про відмовостійкість інформаційної системи в цілому.

У більшості відомих підходів верифікація відмовостійкості зводиться до статичних обчислень показників надійності [19], або до імітаційного моделювання [20]. Відмінність підходу, який пропонується у даній роботі, полягає в інтеграції логіки Хоара у багаторівневу модель, коли для кожної події $e \in E_i$ задається локальна «трійка» $\{P\}e\{Q\}$, для кожної політики – також локальна «трійка», а для траєкторії подій і політик – композиційно виводиться глобальна трійка. Таким чином, властивості доступності, відмовостійкості та ізоляції перевіряються індуктивно по траєкторії, а не за принципом «усе або нічого».

Так як у інформаційних системах на основі багаторівневих структур відмовостійкість ґрунтується на множині альтернативних сценаріїв функціонування, які охоплюють усі можливі послідовності подій і політик відновлення, оскільки відмови носять стохастичний характер і можуть поєднуватися різними способами, то доцільно розглянути механізм введення варіативів для гарантування роботи критичних функціоналів. Тоді, нехай множина варіативів буде задана як

$$\Gamma = \{\xi_1, \xi_2, \ldots, \xi_k\}, \tag{14}$$

де кожна траєкторія ξ задається послідовністю подій відмов і застосованих політик:

$$\xi = e_1; \vartheta_1; e_1; \vartheta_1; \ldots, e_m; \vartheta_m, \tag{15}$$

де $e_m \in E_i$ – подія відмови чи відновлення на рівні i; $\vartheta_m \in \Pi$ – політика реагування.

Таким чином, вимога гарантування критичних функціоналів формулюється як:

$$\forall f \in F_{crit} : \exists \xi \in \Gamma : \{P\}\xi\{\varphi_f(t_{fin}) = 1\}. \tag{16}$$

На відміну від відомих підходів, у яких варіативність найчастіше розглядається на рівні архітектурних конфігурацій (резервування чи реплікація), у запропонованій моделі вперше: варіативи інтегруються з логікою Хоара, що

дозволяє доводити коректність кожного альтернативного сценарію як послідовності подій; кожна множина Γ не є довільною, а формується через правила композиції локальних «трійок», що гарантує верифікацію всіх сценаріїв.

Виходячи з загальноприйнятої теорії, що забезпечення відмовостійкості інформаційної системи супроводжується додатковими витратами (дублювання ресурсів, резервування каналів, підтримка надмірних компонентів) [2], поряд із формальною верифікацією роботоздатності критичних функціоналів необхідно мати апарат оптимізації конфігурацій, що дозволяє знаходити баланс між рівнем відмовостійкості та ресурсно-вартісними обмеженнями.

Задача оптимізації може бути представлена як:

$$\min_{\zeta,\vartheta,\xi} C(\zeta,\vartheta,\xi), \tag{17}$$

за умов:

$$\mathrm{I}=\begin{cases}\Omega(\zeta)=1;\\ \forall f\in F_{crit}:\exists\xi\in\Gamma:\{P\}\,\xi\{\varphi_f=1\};\\ X(\zeta)\geq X^{\min}.\end{cases} \tag{18}$$

Новизна зазначеного підходу полягає в тому, що оптимізація здійснюється за трикутником «структура – процес – політика», тому, відповідно, оптимізаційна задача для запропонованої моделі формулюється як спільний вибір конфігурації, політики і сценарію, що мінімізує витрати, одночасно гарантуючи збереження критичних функціоналів інформаційної системи на основі багаторівневої структури.

Інтеграція запропонованих моделей полягає у побудові єдиної конструкції, де структурна модель (4), процесна інтерпретація подій (5)–(12) та модель забезпечення (13)–(17) функціонують як взаємопов'язані складові однієї сутності. Структурний рівень визначає ієрархію L, зв'язки ρ, інтерфейси ϒ, множини функціоналів F і F_{crit} та обмеження (3), які задають допустимі конфігурації. Процесна складова, що базується на логіці Хоара, задає правила перетворення станів системи під дією подій E_i, формалізуючи послідовності «відмова – поширення – відновлення» у вигляді композицій «трійок». Модель забезпечення відмовостійкості (13) доповнює цю основу множиною політик П, варіативними сценаріями О та критеріями Ξ, що забезпечують збереження працездатності критичних функціоналів і враховують ресурсно-вартісні характеристики. У такій інтеграції структурний рівень фіксує межі простору рішень, процесний рівень описує динаміку переходів між станами, а П, О, Ξ реалізують механізм забезпечення з можливістю порівняння й відбору альтернативних конфігурацій. Схематично цей підхід може бути поданий як багатошаровий конвеєр «структура → події → сценарії → критерії», у якому й реалізується модель забезпечення відмовостійкості для інформаційних систем на основі багаторівневих структур.

Дослідження моделі забезпечення відмовостійкості інформаційних систем на основі багаторівневих структур за критерієм адекватності. Дослідження проводилося на відповідність моделі структурним припущенням і процесного опису фактичним властивостям інформаційних систем на основі багаторівневих структур та коректність прогнозів щодо збереження працездатності критичних функціоналів за стохастичних відмов і керованих політик відновлення у трьох взаємопов'язаних вимірах: структурному (відповідність ієрархії рівнів, зв'язків та інтерфейсів реальній архітектурі), процесному (відповідність опису подій відмов/відновлень спостережуваній динаміці) та прогностичному (узгодженість очікуваних показників доступності, якості та вартості з емпірично досяжними). У якості інформаційної системи було обрано відому інформаційну систему багатошарової віртуалізації [21], яка була розгорнута на базі інформаційно-обчислювального центру Харківського національного університету радіоелектроніки [22].

За результатами досліджень отримано наступні результати:

1. Для обраної інформаційної системи [21] зафіксовано відображення реальних рівнів, зв'язків та інтерфейсів на елементи S; виконано перевірку, що допустимі конфігурації ξ не виходять за межі обмежень Ω і відтворюють фактичні політики резервування та розміщення функціоналів.

2. Логи експлуатації відображаються на послідовності подій і політик; для кожної спостереженої траєкторії ξ ∈ О було перевірено «трійки» виду $(P)\xi(Q)$, де P, Q були сформовані за наявних умов Ω та вимог $\varphi_f(t) = 1$ для всіх $f \in F_{crit}$. Отримана узгодженість таких трійок із даними свідчить про правильність процесного опису.

3. Для репрезентативних конфігурацій ζ та політик П були порівняні отримані в моделі (4) значення $C(\zeta)$, $X_h(\zeta)$ з виміряними значеннями; додатково виконано звірку частки часу з $\varphi_f(t) = 1$ із фактичною доступністю критичних функціоналів. Прийнятні розбіжності були в межах допустимих значень, визначених інформаційною системою [21].

4. Досліджено варіювання параметрів у межах експлуатаційної невизначеності (навантаження, інтенсивності відмов, тривалості відновлень) та проаналізовано стабільність рішення задачі збереження виконання Ω і незначної зміни оптимальних ζ при малих варіаціях вхідних даних.

5. Для множини О було проведено ранжування сценаріїв за критеріями Ξ і перевірено, що сценарії, визнані кращими у моделі, що було підтверджено експериментально (табл. 1).

6. Вимикаючи окремі політики з П та повторюючи дослідження у п.2 та п.5, демонструється причинний внесок кожної політики у виконання вимог до відмовостійкості.

Таблиця 1

Результати перевірки адекватності моделі для системи вкладеної віртуалізації

№ перевірки	Рівні вкладеності	Схема резервування	Політика відновлення	Сценарій	Кількість відмов	Прогнозована доступність критичних функціоналів	Спостережна доступність критичних функціоналів	Відхилення прогнозу від експерименту	Прогнозований середній час відновлення, с	Виміряний середній час відновлення, с	Відхилення прогнозу від експерименту, с	Частка часу, коли інформаційна система виконує основну функцію, %	Чи задовольняє система обмеження Ω
1	2	Гарячий резерв @L2	Перезапуск + реплікація	ξ_1	50	99.90	99.86	–0.04	18.0	19.4	1.4	98.7	Так
2	2	Холодний резерв @L2	Міграція	ξ_2	50	99.40	99.31	–0.09	75.0	79.2	4.2	96.1	Так
3	3	Гарячий резерв @L2,L3	Перезапуск+міграція	ξ_3	80	99.95	99.93	–0.02	12.0	12.8	0.8	99.0	Так
4	3	Без резерву@L3	Локальний відкат	ξ_4	80	98.80	98.55	–0.25	140.0	152.5	12.5	90.4	Ні
5	4	Змішаний: гарячий@L3, холодний@L2	Реплікація + checkpoint	ξ_5	120	99.74	99.70	–0.04	28.0	29.6	1.6	97.8	Так
6	4	Гарячий резерв на всіх рівнях	Актив-актив	ξ_6	120	99.997	99.992	–0.005	6.0	6.7	0.7	99.4	Так
7	2	Без резерву	Лише перезапуск	ξ_7	50	97.90	97.43	–0.47	210.0	228.3	18.3	84.2	Ні
8	3	Гібридний: гарячий@L2, холодний@L3	Гібридна міграція	ξ_8	80	99.60	99.55	–0.05	36.0	37.9	1.9	96.9	Так

За результатами досліджень видно (рис. 1), що модель коректно відтворює архітектуру й динаміку інформаційної системи [21], забезпечує достовірні оцінки $N(\zeta)$, $X_h(\zeta)$ для всіх $f \in F_{crit}$ і може слугувати надійною основою для синтезу конфігурацій у інших реальних умовах експлуатації.

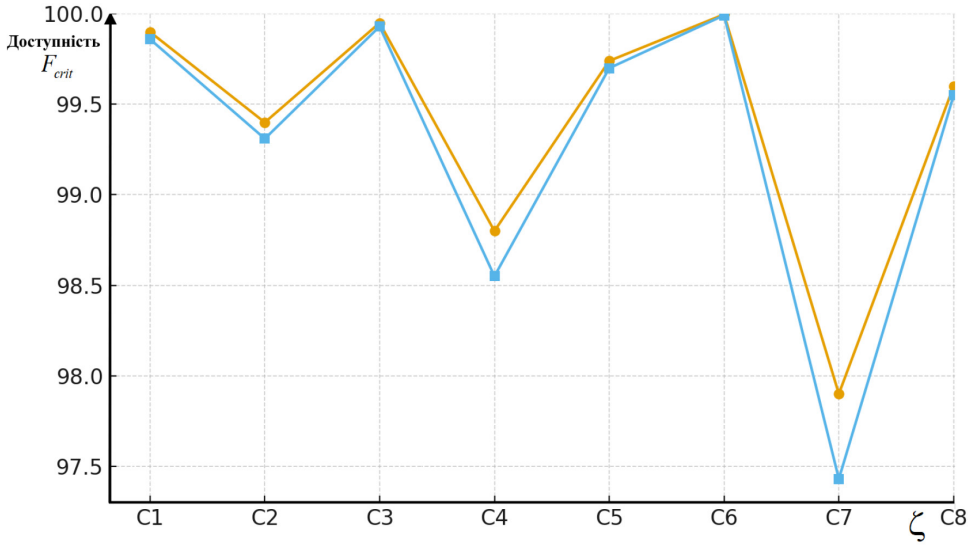


Рис. 1. Графік порівняння прогнозованої та спостереженої доступності $f \in F_{crit}$ для різних конфігурацій ζ

Висновки

У роботі запропоновано новий підхід до побудови моделі забезпечення відмовостійкості інформаційних систем на основі багаторівневих структур.

Вперше запропоновано математичну модель забезпечення відмовостійкості інформаційних систем на основі багаторівневих структур, яка інтегрує структурне подання рівнів і функціоналів із процесною інтерпретацією відмов та відновлень у логіці Хоара, що дозволяє композиційно верифікувати доступність, відмовостійкість та ізоляцію критичних функціоналів. Додатково введено множину політик відновлення, множину варіативів як формалізованих траєкторій функціонування, а також критерії оцінки, що охоплюють показники доступності, відмовостійкості, ізоляції та ресурсно-вартісні характеристики. Це дозволяє не лише перевіряти працездатність критичних функціоналів, а й здійснювати синтез оптимальних конфігурацій з урахуванням вартості резервування.

Подальші дослідження будуть спрямовані на удосконалення моделі шляхом розробки функціоналу автоматизованої генерації варіативів і побудови методів синтезу конфігурацій у реальному часі, на розширення моделі до багатокритеріальної оптимізації з урахуванням ризиків і часу відновлення.

Список використаної літератури

1. Додонов О. Г., Кузнєцова М. Г., Горбачик О. С. Моделювання і оцінювання функціональної стійкості інформаційних систем. Реєстрація, зберігання і обробка даних. 2025. Т. 27, № 1. С. 76–88.

2. Лейченко К. М., Скоробогатько С. В., Фесенко Г. В., Харченко В. С., Яковлев С. В. Оцінювання надійності бездротових сенсорних мереж систем моніторингу лісових пожеж з урахуванням фатальних комбінацій множинних відмов сенсорів. Кібернетика та системний аналіз. 2025. Т. 61, № 1. С. 161–173. URL: http://jnas.nbuv.gov.ua/article/UJRN-0001535211 (дата звернення: 19.09.2025).

3. Tkachov, V., Kovalenko, A., Kharchenko, V., Hvozdetska, K., Hunko, M. An Overlay Network Based on Cellular Technologies for the Secure Control of Intelligent Mobile Objects. ICTERI 2021 Workshops. Communications in Computer and Information Science. Cham 2022. P. 480–490. URL: https://doi.org/10.1007/978-3-031-14841-5_32 (дата звернення: 19.09.2025).

4. Додонов О. Г., Путятін В. Г., Додонов В. О., Куценко С. А., Германюк А. П., Ізварін І. В., Кравчук К. О. Технологія забезпечення живучості територіально-розподілених інформаційних комп'ютерних систем в єдиному інформаційному просторі. Реєстрація, зберігання і обробка даних. 2024. Т. 26, № 1. С. 121–143. URL: https://doi.org/10.35681/1560-9189.2024.26.1.308659 (дата звернення: 19.09.2025).

5. Моделювання та оптимізація процесів безпечної та відмовостійкої маршрутизації в телекомунікаційних мережах : монографія / О. В. Лемешко, О. С. Єременко, М. О. Євдокименко, А. С. Шаповалова, Б. Слейман. Харків : ХНУРЕ, 2022. 198 с. URL: https://doi.org/10.30837/978-966-659-378-1 (дата звернення: 19.09.2025).

6. Кірпічніков Ю., Рибидайло А., Литовченко Г., Бутенко М. Обґрунтування підходу до удосконалення інформаційної інфраструктури Міністерства оборони України для функціонування в умовах збройної агресії. Збірник наукових праць Центру воєнно-стратегічних досліджень НУОУ імені Івана Черняховського. 2023. С. 87–97.

7. Панченко С., Антонюк А., Новицький Д., Заморский С.. Загальні підходи до створення системи управління і контролю функціонування сервісів інформаційних систем у Збройних Силах України. Сучасні інформаційні технології у сфері безпеки та оборони. 2023. Т. 48, № 3. С. 98–106.

8. Андрощук О., Черевко Р., Петрушен М., Голобородько, М. Актуальні підходи до побудови інформаційної інфраструктури на основі хмарних технологій з використанням референсної архітектури. Сучасні інформаційні технології у сфері безпеки та оборони. 2023. Т. 46, № 1. С. 89–94.

9. Міхнов Є. Д. Огляд технології балансування навантаження у міксованих VPN-ланцюгах. Радіоелектроніка та молодь у XXI столітті : матеріали 28-го Міжнародного молодіжного форуму. 2024. Т. 5. С. 34–37. URL: https://doi.org/10.30837/IYF.PCEIP.2024.034 (дата звернення: 19.09.2025).

10. Замрій І. В., Собчук А. В., Лаптєв С. О., Лаптєва Т. О., Копитко, С. Б. Алгоритм контролю та прогнозування функціональної стійкості складних інформаційно-технічних систем. Телекомунікаційні та інформаційні технології. 2022. Т. 1. С. 4–14.

11. Шовкошитний І. Підхід до захисту інформаційно-телекомунікаційних систем на основі автокомпенсаційного принципу адресно-часової селекції аномалій мережевого трафіка. Сучасні інформаційні технології у сфері безпеки та оборони. 2023. Т. 48, № 3. С. 5–10.

12. Мартовицький В., Свиридов А., Авдєєв О., Гудзинський І., Коротецький О. Дослідження методів виявлення аномалій у арі журналах для забезпечення безпеки та надійності програмних систем. Вісник Херсонського національного технічного університету. 2025. Т. 2, № 1(92). С. 142–148.

13. Tkachov V., Chepurna I., Frolov D. The promising method of secure transmission of inelastic data in peer-to-peer networks. Computer and information systems and technologies: proceedings of seventh international scientific and technical conference. 2024. P. 15−16.

14. Кузьменко Д. Методи та моделі емуляції негативних впливів на ресурси інформаційних систем. Технічні науки та технології. 2024. Т. 4, № 38. С. 189–196.

15. Hattab N., Belalem G. Modular models for systems based on multi-tenant services: a multi-level Petri-net-based approach. Journal of King Saud University – computer and information sciences. 2023. P. 101671. URL: https://doi.org/10.1016/j.jksuci.2023.101671 (дата звернення: 19.09.2025).

16. Kashyap V., Ahuja R., Kumar A. Predictive analysis-based load balancing and fault tolerance in fog computing environment. Cluster Computing. 2025. Vol. 28, № 5. URL: https://doi.org/10.1007/s10586-024-04984-5 (дата звернення: 19.09.2025).

17. Melo M., Aquino G. FaTEMa: a framework for multi-layer fault tolerance in IoT systems. Sensors. 2021. Vol. 21, № 21. P. 7181. URL: https://doi.org/10.3390/s21217181 (дата звернення: 19.09.2025).

18. Hoare C. A. R. An axiomatic basis for computer programming. Communications of the ACM. 1969. Vol. 12, № 10. P. 576–580.

19. Yang S., Tan J., Lei T., Linares-Barranco B. Smart traffic navigation system for fault-tolerant edge computing of internet of vehicle in intelligent transportation gateway. IEEE Transactions on Intelligent Transportation Systems. 2023. Vol. 24, № 11. P. 13011–13022. URL: https://doi.org/10.1109/TITS.2022.3232231 (дата звернення: 19.09.2025).

20. Wang Y., Wang Z. Distributed model free adaptive fault-tolerant consensus tracking control for multiagent systems with actuator faults. Information sciences. 2024. P. 120313. URL: https://doi.org/10.1016/j.ins.2024.120313 (дата звернення: 19.09.2025).

21. Tkachov V., Hunko M., Volotka V. Scenarios for implementation of nested virtualization technology in task of improving cloud firewall fault tolerance. 2019 IEEE International Scientific-Practical conference Problems of Infocommunications, Science and Technology (PIC S&T). 2019. URL: https://doi.org/10.1109/picst47496.2019.9061473 (дата звернення: 19.09.2025).

22. Сектор інформаційного забезпечення інформаційно обчислювального центру. Харківський національний університет радіоелектроніки – ХНУРЕ, NURE. https://nure.ua/branch/informatsiyno-obchislyuvalniy-tsentr/struktura-iots/sektor-informatsijnoho-zabezpechennia (дата звернення: 19.09.2025)

References

1. Dodonov, O. H., Kuznietsova, M. H., & Horbachyk, O. S. (2025). Modeliuvannia i otsiniuvannia funktsionalnoi stiikosti informatsiinykh system. Reiestratsiia, zberihannia i obrobka danykh, 27(1), 76–88.

2. Leichenko, K. M., Skorobohatko, S. V., Fesenko, H. V., Kharchenko, V. S., & Yakovlev, S. V. (2025). Otsiniuvannia nadiinosti bezdrotovykh sensornykh merezh system monitorynhu lisovykh pozhezh z urakhuvanniam fatalnykh kombinatsii mnozhynnykh vidmov sensoriv. Kibernetyka ta systemnyi analiz, (61,1), 161-173. URL: http://jnas.nbuv.gov.ua/article/UJRN-0001535211

3. Tkachov, V., Kovalenko, A., Kharchenko, V., Hvozdetska, K., Hunko, M. (2022). An Overlay Network Based on Cellular Technologies for the Secure Control of Intelligent Mobile Objects. In: Ignatenko, O., et al. ICTERI 2021 Workshops. ICTERI 2021. Communications in Computer and Information Science, vol 1635. Springer, Cham. https://doi.org/10.1007/978-3-031-14841-5_32

4. Dodonov, O. H., Putiatin, V. H., Dodonov, V. O., Kutsenko, S. A., Hermaniuk, A. P., Izvaryn, I. V., & Kravchuk, K. O. (2024). Tekhnolohiia zabezpechennia zhyvuchosti terytorialno-rozpodilenykh informatsiinykh kompiuternykh system v yedynomu informatsiinomu prostori. Reiestratsiia, zberihannia i obrobka danykh, 26(1), 121–143. https://doi.org/10.35681/1560-9189.2024.26.1.308659

5. Lemeshko, O. V., Yeremenko, O. S., Yevdokymenko, M. O., Shapovalova, A. S., & Sleiman, B. (2022). Modeliuvannia ta optymizatsiia protsesiv bezpechnoi ta vidmovostiikoi marshrutyzatsii v telekomunikatsiinykh merezhakh. https://doi.org/10.30837/978-966-659-378-1

6. Kirpichnikov, Yu., Rybydailo, A., Lytovchenko, H., & Butenko, M. (2023). Obgruntuvannia pidkhodu do udoskonalennia informatsiinoi infrastruktury Ministerstva oborony Ukrainy dlia funktsionuvannia v umovakh zbroinoi ahresii. Zbirnyk naukovykh prats Tsentru voienno-stratehichnykh doslidzhen NUOU imeni Ivana Cherniakhovskoho, 87–97.

7. Panchenko, C., Antoniuk, A., Novytskyi, D., & Zamopckyi, C. (2023). Zahalni pidkhody do stvorennia systemy upravlinnia i kontroliu funktsionuvannia servisiv informatsiinykh system u Zbroinykh Sylakh Ukrainy. Suchasni informatsiini tekhnolohii u sferi bezpeky ta oborony, 48(3), 98–106.

8. Androshchuk, O., Cherevko, R., Petrushen, M., & Holoborodko, M. (2023). Aktualni pidkhody do pobudovy informatsiinoi infrastruktury na osnovi khmarnykh tekhnolohii z vykorystanniam referensnoi arkhitektury. Suchasni informatsiini tekhnolohii u sferi bezpeky ta oborony, 46(1), 89–94.

9. Mikhnov, Ye. D. (2024). Ohliad tekhnolohii balansuvannia navantazhennia u miksovanykh VPN-lantsiuhakh. Radioelektronika ta molod u XXI stolitti : materialy 28-ho mizhnarodnoho molodizhnoho forumu, 5, 34–37. https://doi.org/10.30837/IYF.PCEIP.2024.034

10. Zamrii, I. V., Sobchuk, A. V., Laptiev, S. O., Laptieva, T. O., & Kopytko, S. B. (2022). Alhorytm kontroliu ta prohnozuvannia funktsionalnoi stiikosti skladnykh informatsiino-tekhnichnykh system. Telekomunikatsiini ta informatsiini tekhnolohii, (1), 4–14.

11. Shovkoshytnyi, I. (2023). Pidkhid do zakhystu informatsiino-telekomunikatsiinykh system na osnovi avtokompensatsiinoho pryntsypu adresno-chasovoi selektsii anomalii merezhevoho trafika. Suchasni informatsiini tekhnolohii u sferi bezpeky ta oborony, 48(3), 5–10.

12. Martovytskyi, V., Svyrydov, A., Avdieiev, O., Hudzynskyi, I., & Korotetskyi, O. (2025). Doslidzhennia metodiv vyiavlennia anomalii u api zhurnalakh dlia zabezpechennia bezpeky ta nadiinosti prohramnykh system. Visnyk Khersonskoho natsionalnoho tekhnichnoho universytetu, 2(1 (92)), 142–148.

13. Tkachov, V., Chepurna, I., & Frolov, D. (2024). The promising method of secure transmission of inelastic data in peer-to-peer networks. Computer and information systems and technologies: proceedings of seventh international scientific and technical conference, 15–16.

14. Kuzmenko, D. (2024). Metody ta modeli emuliatsii nehatyvnykh vplyviv na resursy informatsiinykh system. Tekhnichni nauky ta tekhnolohii, (4 (38)), 189–196.

15. Hattab, N., & Belalem, G. (2023). Modular models for systems based on multi-tenant services: a multi-level Petri-net-based approach. Journal of King Saud University – computer and information sciences, 101671. https://doi.org/10.1016/j.jksuci.2023.101671

16. Kashyap, V., Ahuja, R., & Kumar, A. (2025). Predictive analysis-based load balancing and fault tolerance in fog computing environment. Cluster Computing, 28(5). https://doi.org/10.1007/s10586-024-04984-5

17. Melo, M., & Aquino, G. (2021). FaTEMa: a Framework for Multi-Layer Fault tolerance in IoT Systems. Sensors, 21(21), 7181. https://doi.org/10.3390/s21217181

18. Hoare, C. A. R. (1969). An axiomatic basis for computer programming. Communications of the ACM, 12(10), 576–580.

19. Yang, S., Tan, J., Lei, T., & Linares-Barranco, B. (2023). Smart Traffic Navigation System for Fault-Tolerant Edge Computing of Internet of Vehicle in Intelligent Transportation Gateway. IEEE Transactions on Intelligent Transportation Systems, 24(11), 13011-13022. https://doi.org/10.1109/tits.2022.3232231

20. Wang, Y., & Wang, Z. (2024). Distributed model free adaptive fault-tolerant consensus tracking control for multiagent systems with actuator faults. Information sciences, 120313. https://doi.org/10.1016/j.ins.2024.120313

21. Tkachov, V., Hunko, M., & Volotka, V. (2019). Scenarios for implementation of nested virtualization technology in task of improving cloud firewall fault tolerance. In 2019 IEEE International Scientific-Practical conference Problems of Infocommunications, Science and Technology (PIC S&T). IEEE. https://doi.org/10.1109/picst47496.2019.9061473

22. Sektor informatsiinoho zabezpechennia informatsiino obchysliuvalnoho tsentru. (2025, September 10). Kharkivskyi natsionalnyi universytet radioelektroniky – KhNURE, NURE. https://nure.ua/branch/informatsiyno-obchislyuvalniy-tsentr/struktura-iots/sektor-informatsijnoho-zabezpechennia

Дата першого надходження рукопису до видання: 29.09.2025
Дата прийнятого до друку рукопису після рецензування: 24.10.2025
Дата публікації:

УДК 004.056:005.8:519.95 DOI

М. А. ФУКС
аспірант кафедри інфокомунікаційної інженерії
імені В. В. Поповського
Харківський національний університет радіоелектроніки
ORCID: 0009-0001-0248-2109

І. С. ДОБРИНІН
кандидат технічних наук, доцент,
доцент кафедри інфокомунікаційної інженерії
імені В. В. Поповського
Харківський національний університет радіоелектроніки
ORCID: 0000-0001-8910-2609

С. В. ПШЕНИЧНИХ
кандидат технічних наук, старший науковий співробітник,
доцент кафедри інфокомунікаційної інженерії
імені В. В. Поповського
Харківський національний університет радіоелектроніки
ORCID: 0000-0003-0533-2306

А. В. СНІГУРОВ
кандидат технічних наук, доцент,
декан факультету інфокомунікацій
Харківський національний університет радіоелектроніки
ORCID: 0000-0003-1355-7978

В. Х. ЧАКРЯН
кандидат технічних наук,
старший викладач кафедри інфокомунікаційної інженерії
імені В. В. Поповського
Харківський національний університет радіоелектроніки
ORCID: 0000-0003-4827-9779

МЕТОДИКА ПОБУДОВИ СИСТЕМИ УПРАВЛІННЯ ІНФОРМАЦІЙНОЮ БЕЗПЕКОЮ НА ОСНОВІ МЕТОДІВ АНАЛІЗУ ІЄРАРХІЙ, МЕРЕЖЕВОГО АНАЛІЗУ ТА ТЕОРІЇ ІГОР В УМОВАХ ОБМЕЖЕНИХ РЕСУРСІВ

В умовах цифровізації та швидкої еволюції кіберзагроз стає критично важливим побудова надійної системи управління інформаційною безпекою (СУІБ). Наявні міжнародні та національні стандарти побудови СУІБ не враховують специфіки розподілу ресурсів, а лише вказують на важливість забезпечення ресурсами для створення, реалізації, підтримки та безперервного розвитку СУІБ, тобто прослідковується певний брак механізмів прийняття управлінських рішень та оптимізації виділених ресурсів. Тому актуальною залишається проблема ідентифікації, розподілу та раціонального використання ресурсів для досягнення достатнього рівня ефективності СУІБ. Крім того, при імплементації СУІБ на базі будь-якої з методик, одним із найскладніших кроків є визначення та інтеграція дієвих заходів і інструментів інформаційного захисту. У цьому контексті відчувається нестача практичних та методичних інструментів для оптимізації, адже навіть за умов істотних вкладень і широкомасштабного впровадження засобів захисту інформації (ЗЗІ) виникає сумнів щодо доцільності та ефективності таких рішень. У роботі запропоновано методику побудови СУІБ, що поєднує метод аналізу ієрархій (МАІ), методи аналізу мереж (МАМ) та математичний апарат теорії ігор для вибору оптимальної базової методики і засобів захисту в умовах обмежених ресурсів та наявних загроз. Запропонований підхід також передбачає використання мережевого аналізу для моделювання проєкту з урахуванням строків, витрат, ресурсів та можливих сценаріїв реалізації. Методика включає побудову мережевої моделі проєкту, розрахунок критичного шляху, виявлення критичних та некритичних робіт, аналіз часових резервів, варіантів прискорення проєкту і визначення впливу прискорення на загальну вартість. Запропоновано спосіб порівняльної оцінки кількох сценаріїв реалізації проєкту: при звичайних строках, при прискоренні лише критичних робіт, а також при при-

© Фукс М. А., Добринін І. С., Пшеничних С. В., Снігуров А. В., Чакрян В. Х., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

скоренні всіх або лише некритичних робіт. У результаті отримано графіки розподілу ресурсів у часі, фінансових витрат та залученості людського ресурсу, що дозволяють приймати більш зважені управлінські рішення. Методика є практичним інструментом для Chief Information Security Officer (CISO) та керівників, що дозволяє визначити оптимальний обсяг інвестицій на кожному етапі проєкту СУІБ та обґрунтовано аргументувати потребу в ресурсах, забезпечуючи баланс між ефективністю та витратами.

Ключові слова: *система управління інформаційною безпекою; мережевий аналіз; критичний шлях; теорія ігор; оптимізація ресурсів; CISO; управлінські рішення.*

M. A. FUKS
Postgraduate Student at the V. V. Popovskyy Department
of Infocommunication Engineering
Kharkiv National University of Radio Electronics
ORCID: 0009-0001-0248-2109

I. S. DOBRYNIN
PhD, Associate Professor,
Associate Professor at the V. V. Popovsky Department
of Infocommunication Engineering
Kharkiv National University of Radio Electronics
ORCID: 0000-0001-8910-2609

S. V. PSHENYCHNYKH
PhD, Senior Researcher,
Associate Professor at the V. V. Popovsky Department
of Infocommunication Engineering
Kharkiv National University of Radio Electronics
ORCID: 0000-0003-0533-2306

A. V. SNIHUROV
PhD, Associate Professor,
Dean of the Faculty of Infocommunications
Kharkiv National University of Radio Electronics
ORCID: 0000-0003-1355-7978

V. H. CHAKRIAN
PhD, Senior Lecturer at the V. V. Popovskyy Department
of Infocommunication Engineering
Kharkiv National University of Radio Electronics
ORCID ID: 0000-0003-4827-9779

METHODOLOGY FOR DEVELOPING INFORMATION SECURITY MANAGEMENT SYSTEMS USING NETWORK ANALYSIS AND GAME THEORY

In the context of digitalization and the rapid evolution of cyber threats, building a reliable Information Security Management System (ISMS) becomes critically important. Existing international and national ISMS standards do not take into account the specifics of resource allocation; instead, they only emphasize the importance of providing resources for the creation, implementation, maintenance, and continuous improvement of ISMS. This indicates a certain lack of mechanisms for making management decisions and optimizing allocated resources. Therefore, the problem of identifying, distributing, and rationally using resources to achieve an adequate level of ISMS efficiency remains relevant. Moreover, during ISMS implementation based on any methodology, one of the most challenging steps is the identification and integration of effective information security measures and tools. In this context, there is a shortage of practical and methodological optimization tools, since even with substantial investments and large-scale deployment of information security measures (ISMs), doubts may arise regarding the expediency and efficiency of such solutions. This paper proposes a methodology for building ISMS that combines the Analytic Hierarchy Process (AHP), network analysis methods, and the mathematical apparatus of game theory to select the optimal baseline methodology and protection tools under conditions of limited resources and existing threats. The proposed approach also involves the use of network analysis to model the project, taking into account timelines, costs, resources, and possible implementation scenarios. The methodology includes the construction of a project network model, calculation of the critical path, identification of critical and non-critical activities, analysis of time reserves, acceleration options, and assessment of the impact of acceleration on total cost. A method for comparative evaluation of several implementation scenarios is proposed: with normal timelines, with acceleration of only critical activities, as well as with acceleration of all or only non-critical

activities. As a result, resource allocation over time, financial cost, and human resource engagement charts are obtained, enabling more informed management decisions. The methodology is a practical tool for the Chief Information Security Officer (CISO) and executives, allowing them to determine the optimal investment amount at each stage of the ISMS project and to substantiate resource requirements while ensuring a balance between efficiency and cost.

Key words: *information security management system; network analysis; critical path; game theory; resource optimization; CISO; managerial decisions.*

Постановка проблеми

Наразі існує низка визнаних підходів, які використовуються як базис для побудови СУІБ в організаціях, серед яких найпоширенішими є стандарт ISO/IEC 27001 [1], методологія IT-Grundschutz [2] та Постанова Національного банку України (НБУ) № 95 [3]. Аналіз [1]–[3] свідчить про те, що в усіх наявних підходах акцентується увага на необхідності залучення всіх типів ресурсів: людських, фінансових, технічних, організаційних та інфраструктурних – як ключової умови для ефективного формування СУІБ. При цьому склад і обсяг необхідних ресурсів мають варіюватися в залежності від масштабу організації, специфіки її діяльності і рівня ризиків, адже саме коректне визначення та раціональне використання ресурсів є запорукою стабільної роботи усіх бізнес-процесів.

Водночас аналіз наявних підходів свідчить, що, незважаючи на їхню кількість і здатність частково структурувати процес створення СУІБ, вони не надають вичерпаних відповідей на два критично важливих запитання:

– за якими критеріями слід обирати базову методику побудови СУІБ для конкретної організації;
– яким чином оптимізувати використання доступних ресурсів у межах СУІБ-проєкту.

Вочевидь, неефективне планування та використання ресурсів у рамках СУІБ може призвести до перевитрат бюджету, затримок у реалізації заходів інформаційної безпеки або недостатнього рівня захищеності, що в умовах зростання кіберзагроз є критичним фактором для організацій.

З огляду на вищезазначене, у цій роботі запропоновано інтегрований підхід, який ґрунтується на методі аналізу ієрархій, методі аналізу мереж, та теорії ігор для обґрунтованого вибору стратегії побудови СУІБ і ефективного управління ресурсами на всіх етапах проєкту.

Аналіз останніх досліджень і публікацій

Проблематика розробки підходів до побудови ефективних СУІБ є однією з пріоритетних у сучасній науково-практичній сфері інформаційної безпеки. Ця актуальність обумовлена як зростаючою кількістю та складністю кіберзагроз, так і стрімким розвитком цифрових технологій, що суттєво впливають на інформаційну інфраструктуру організацій.

Сучасні дослідження та практичні реалізації в цій області [4]–[9] переважно зосереджуються на двох напрямках. Перший – це оцінювання ефективності інвестицій у СУІБ, шо передбачає аналіз витрат та вигод, пов'язаних із впровадженням заходів інформаційної безпеки, а також формування моделей оптимального розподілу ресурсів. Другий напрямок спрямований на вибір та адаптацію конкретних методик і стандартів побудови СУІБ, що включає аналіз наявних методологій, визначення їх сильних і слабких сторін, а також розробку рекомендацій щодо їх застосування в залежності від специфіки організації.

Тематика оцінювання ефективності виконаних інвестицій в ІБ знайшла широке відображення в науковій літературі, численних публікаціях, дослідженнях та прикладних методиках.

Варто зазначити, що запобігання можливих втрат може розглядатися як економія різноманітних ресурсів усієї бізнес-системи. Часто наголошується, що жодна система управління не зможе працювати належним чином без належного рівня забезпечення ресурсів [4].

Економічна модель Гордона–Лоеба [5] на сьогодні вважається однією з ключових у визначенні оптимального обсягу інвестицій в ІБ. Її концептуальна суть полягає в оберненій залежності: зі зростанням вкладень у захист зменшується рівень вразливості системи. Під час аналізу класів функцій, автори дійшли висновку, що оптимальний обсяг інвестицій становить приблизно 37 % від очікуваних втрат, яких може зазнати організація під час реалізації загроз [5]. Таким чином, модель демонструє, що економічно доцільний рівень інвестування в ІБ є значно нижчим за самі можливі збитки. Однак ця модель є здебільшого формалізованою й аксіоматичною, що ускладнює врахування особливостей конкретної організації та середовища ведення діяльності.

Питання вибору оптимальної методики побудови СУІБ також досліджувалося в інших роботах. Так, у роботі [6], було застосовано методи Томаса Сааті для моделювання багатокритеріальної задачі оптимізації. У дослідженні [7] запропоновано підхід до оптимізації інвестицій в ІБ підприємства на основі формалізованої моделі, яка дозволяє визначити інтегральний показник ефективності інвестицій. Як оптимізаційні змінні в моделі розглядаються загальні витрати на ліквідацію наслідків реалізованих загроз та інших факторів, а також сукупні виплати джерелам фінансування. Водночас модель не враховує обмеження фінансування з боку самої компанії, що може лімітувати її практичне застосування. У роботі [8] в основу запропонованого методу проєктування кіберстійкої інфокомунікаційної мережі (ІКМ) покладено розв'язання оптимізаційної задачі, яка пов'язана з узгодженням різнотипних керуючих змінних, проте такий підхід зосереджується переважно на формальних аспектах оптимізації параметрів, не враховуючи практичних механізмів підтримки управлінських рішень з боку CISO. У роботі [9],

на основі критерія максимуму інтегрального показника ефективності, вирішується задача оптимального вибору засобів захисту інформації від загроз безпеки при впровадженні СУІБ.

Проте низка важливих питань залишається відкритою: який саме підхід використовувати до побудови СУІБ та як оцінювати ефективність інвестицій у захист інформації та побудову СУІБ з урахуванням сучасних вимог і середовища існування. Більшість підходів акцентують увагу саме на фінансовій складовій та ЗЗІ, тоді як час і людські ресурси, що також суттєво впливають на загальні витрати, часто залишають поза увагою. У цьому контексті доцільним є використання принципів методології управління проєктами. Через складність взаємозв'язків між цими елементами питання комплексної оптимізації інвестицій залишається актуальним і потребує подальшого дослідження.

Формулювання мети дослідження

Метою статті є розробка методики побудови СУІБ, що поєднує підходи методів аналізу ієрархій і мереж, теорію ігор та методів управління проєктами для забезпечення оптимального вибору стратегії захисту, раціонального розподілу ресурсів і прийняття обґрунтованих управлінських рішень.

Викладення основного матеріалу дослідження

Запропонована методика побудови суіб

У якості основи для формування запропонованої методики та визначення механізмів прийняття управлінських рішень, спрямованих на оптимізацію взаємопов'язаних людських, часових і фінансових ресурсів, запропоновано залучення методів мережевого аналізу, методу аналізу ієрархій, математичного апарату теорії ігор, а також інструментарію управління проєктами.

Запропонована методика складається з трьох ключових етапів.

1) Оцінювання початкового стану об'єкта та аналіз потенціалу впровадження методики в конкретній організації.

2) Основна частина методики, яка умовно поділена на дві підсистеми:

– перша спрямована на вибір методики, на основі якої розбудовувати СУІБ (через метод аналізу мереж (МАМ) або метод аналізу ієрархій (МАІ)) [10], [11], та визначення загальної стратегії інформаційного захисту (через математичний апарат теорії ігор) – для визначення оптимальної стратегії CISO і оптимізації фінансових вкладень у ЗЗІ;

– друга – для моделювання мережевої конфігурації проєкту з метою подальшого аналізу та оптимізації за допомогою методів мережевого аналізу і керування проєктами – щоб сформувати цілісне уявлення, яке слугуватиме підґрунтям для ухвалення управлінського рішення щодо обґрунтованості запиту та визначення необхідного обсягу інвестицій.

3) Етап впровадження, що включає процедури інтеграції розробленої методики в компанії.

На першому етапі визначається поточний рівень ІБ організації та проводиться оцінка можливостей впровадження СУІБ. Така оцінка може здійснюватися як силами внутрішніх фахівців самої компанії, так і шляхом залучення незалежних експертів консалтингових компаній. Однак якість цього етапу значною мірою залежить від компетентності та релевантності створеної експертної групи. У разі залучення недостатньо кваліфікованих спеціалістів або неправильної побудови процесу аудиту існує підвищений ризик отримання хибних результатів. Саме тому до формування експертних груп висуваються підвищені вимоги, спрямовані на забезпечення об'єктивності та достовірності оцінювання. Результатом розглянутого етапу є узагальнена експертна оцінка початкового стану об'єкта, що має ключове значення для подальшого проєктування СУІБ.

Розглянемо складові основної частини запропонованої методики. Перша підсистема основної частини методики передбачає вирішення багатокритеріальної задачі оптимізації, що виникає на початковому етапі впровадження СУІБ. Відповідальній особі необхідно обґрунтовано обрати оптимальну методику, яка стане основою для побудови СУІБ. Для цього можуть бути застосовані методи МАІ або МАМ [6].

Метод МАІ, розроблений Т. Сааті, засновується на декомпозиції проблеми вибору на більш прості складники і поступовому наданні пріоритетів компонентам використовуючи парні порівняння [6]. Такий підхід забезпечує прозорість і відносну простоту у прийнятті рішень. Водночас метод МАМ розширює можливості МАІ, оскільки враховує не лише ієрархічні залежності, а й складні взаємозв'язки між критеріями та альтернативами, що робить його більш придатним для комплексних багатокритеріальних задач, характерних для побудови СУІБ.

Структурування моделі багатокритеріальної задачі передбачає попереднє збирання та систематизацію інформації відповідальною особою або експертною групою. На цьому етапі можуть застосовуватися різні методи генерації ідей та колективного напрацювання рішень, а отримана й узгоджена інформація стає підґрунтям для побудови раціональної системи у вигляді ієрархії (мережі), що надалі аналізується за допомогою методів МАІ або МАМ. Важливим аспектом цього процесу є достатня компетентність експертів та глибоке розуміння проблематики, адже саме від цього залежить якість і коректність подальшого моделювання. Таким чином, обґрунтовується вибір оптимальної методики, яка визначає підґрунтя для подальшого проєктування СУІБ.

Наступним кроком у межах основної частини запропонованої методики є обґрунтований вибір ЗЗІ, які забезпечуватимуть належний рівень безпеки у межах побудови СУІБ, і вибір яких доцільно здійснювати із застосуванням теоретико-ігрового підходу. Такий підхід дозволяє моделювати протидію між потенційним зловмисником та CISO, забезпечуючи обґрунтоване прийняття рішень щодо вибору контрзаходів в умовах обмежених ресурсів.

У роботі [12] розглянуто задачу оптимізації стратегії захисту інформації шляхом моделювання біматричної гри між CISO та зловмисником. Модель базувалася на двох показниках: Return On Security Investment (ROSI) та Return On Attack (ROA). У роботі було висунуто припущення про повну невизначеність поведінки зловмисника. У цьому дослідженні робиться припущення, що CISO володіє ймовірнісною інформацією про можливі атаки на основі відкритих бібліотек вразливостей і атак, наприклад, CAPEC, OWASP Top Ten, CVE, CVSS, що дозволяє застосовувати модель з чистими стратегіями [13]–[16].

Нехай CISO має стратегії для вибору, що визначатимуться набором $S_a = \{b_1, b_2, \ldots, b_m\}$, де m відображає загальну кількість таких стратегій, тобто певний набір ЗЗІ.

Нехай утворена матриця виграшів CISO матиме вигляд, який надано на рис. 1 [12].

	b_1	b_2	b_3	…	b_m
a_1	$U_A(a_1,b_1)$	$U_A(a_1,b_2)$	$U_A(a_1,b_3)$	…	$U_A(a_1,b_m)$
a_2	$U_A(a_2,b_1)$	$U_A(a_2,b_2)$	$U_A(a_2,b_3)$	…	$U_A(a_2,b_m)$
a_3	$U_A(a_3,b_1)$	$U_A(a_3,b_2)$	$U_A(a_3,b_3)$	…	$U_A(a_3,b_m)$
…	…	…	…	…	…
a_n	$U_A(a_n,b_1)$	$U_A(a_n,b_2)$	$U_A(a_n,b_3)$	…	$U_A(a_n,b_m)$
a_1+a_2	$U_A(a_1+a_2,b_1)$	$U_A(a_1+a_2,b_2)$	$U_A(a_1+a_2,b_3)$	…	$U_A(a_1+a_2,b_m)$
…	…	…	…	…	…
$a_{n-1}+a_n$	$U_A(a_{n-1}+a_n,b_1)$	$U_A(a_{n-1}+a_n,b_2)$	$U_A(a_{n-1}+a_n,b_3)$	…	$U_A(a_{n-1}+a_n, b_m)$

Рис. 1. Матриця виграшів CISO

У якості функцій корисності братиметься запропонований показник повернення інвестицій у кібербезпеку ROSI [12]:

$$ROSI = \frac{AV \cdot EF_j \cdot ARO_j \cdot RM_{ij} - \cos t_i}{\cos t_i}, \tag{1}$$

де AV – вартість активу, EF_j – Exposure Factor, тобто втрачена частина активу від реалізації атаки, ARO_j – загальна кількість появи атаки на рік, RM_{ij} – ефективність відбиття атаки, $cost_i$ – витрати на підтримку ЗЗІ. Таким чином елементи матриці представляють кількісну характеристику вигоди впровадження певної стратегії захисту інформації проти можливих атак зловмисника.

Припустимо, що відомо ймовірнісний розподіл для стратегій зловмисника $p_i = \{p_i, p_2, \ldots, p_m\}$. Це дозволяє визначити виграші CISO, який обирає конкретні стратегії захисту, виходячи з того, що зловмисник діє за змішаною стратегією:

$$U_A(a_i) = \sum_{k=1}^{m} U_A(a_i, b_k) \cdot p_i, \quad k \in S_a, \forall i \in S_d. \tag{2}$$

Таким чином формується обґрунтована стратегія вибору засобів захисту інформації, яка забезпечує оптимальне використання ресурсів та мінімізацію потенційних збитків. Однак для реалізації такого вибору у межах конкретного проєкту необхідно врахувати ще й часові, фінансові та організаційні обмеження. Саме тому наступним кроком запропонованої методики є побудова мережевої моделі проєкту створення СУІБ, що становить другу підсистему основної частини методики та забезпечує аналіз і оптимізацію розподілу ресурсів у процесі впровадження. Необхідність мережевого представлення проєкту зумовлена тим, що воно забезпечує формалізацію всіх робіт і подій у струтуру, дозволяючи виявляти критичні шляхи, резерви часу, залежності між роботами та їхній вплив на ресурси. Таким чином створюється інструмент для комплексного аналізу та оптимізації проєкту побудови СУІБ.

Припустимо, що відповідальна особа прагне оцінити складність окремого етапу побудови СУІБ, спланувати його реалізацію як частину загального проєкту, визначити тривалість виконання, критичні роботи та забезпечити раціональний розподіл ресурсів із можливістю оптимізації строків.

Нехай уся мережа Grid представлятиметься у вигляді певного набору чисел:

$$Grid = (K, C), \tag{3}$$

де $K = (1, \ldots, n)$ – множина вершин мережі, n – загальна кількість таких вузлів, C – множина ребер робіт мережі, що зображуються роботи.

Нехай d_{ij} – тривалість певної роботи $(i, j) \in C, i \in K, j \in K$. $T_i(E)$ – найбільш ранній можливий час настання події (досягнення вершини i), де $i \in K$. $T_i(L)$ – найбільш пізній дозволений час настання події i, де $i \in K$. W_r – множина усіх шляхів від вершини 1 до вершини $i \in K$, r – кількість таких шляхів.

Кожен шлях містить у собі певну кількість робіт, сума тривалостей виконання яких і створює суму виконання цього шляху. Найбільша ж величина з декількох і буде значенням $T_i(E)$, тобто найбільш раннім можливим часом настання події $i \in K$. Обумовлені моменти представлені у формулі:

$$T_i(E) = \max\left(\sum_k \sum_q d_{kq}\right), \tag{4}$$

де $k, q \in W_r$, $i \in K$.

Необхідно зауважити, що найбільш ранній можливий час настання першої події 0.

Ключовим моментом у визначенні крайнього дозволеного терміну настання події є те, що подібне збільшення відстрочки не повинне змінювати загальну тривалість виконання робіт. Саме тому виконується умова, що найбільш пізній дозволений час настання події такий самий, як і найбільш ранній можливий час завершення проекту [17]:

$$T_n(L) = T_n(E). \tag{5}$$

Саме тому розрахунок $T_i(L)$ відбувається у зворотньому порядку (з кінця проекту до початку). Потрібно здійснити аналіз тривалості всіх робіт, що виходять із заданої події i, та визначити для неї мінімальну різницю між моментом настання наступної вершини j і тривалістю роботи, яка безпосередньо веде до цієї вершини:

$$T_i(L) = \min(T_j(L) - d_{ij}). \tag{6}$$

На рівні з вищезазначеними визначеннями $T_i(E)$ та $T_i(L)$ постають ще пара важливих компонент аналізу мереж: резерв часу та критичний шлях.

Резерв часу ΔT_i – це загальний час, на який може бути затримано настання певної події і за умовою, що загальний час проекту не збільшиться. Наступна формула відображає спосіб розрахунку резерву часу для певної події i:

$$\Delta T_i = T_i(L) - T_i(E). \tag{7}$$

Часовий резерв вершини визначається як різниця між максимально пізнім допустимим моментом настання події та мінімально можливим часом її виникнення. Інакше кажучи, це найбільший проміжок часу, на який може бути відтерміноване настання певної події i. Зауважимо, у разі наявності нульового значення резерву часу, затримка настання події не допускається, інакше відбудеться затримка виконання усього проекту. Такі нульові події утворюють ключову структуру: критичний шлях мережі, а роботи, які містяться на такому шляху – це критичні роботи.

Отже, вибудовується наступна послідовність розрахунків, яка наведена на рис. 2.

Критичний шлях задає мінімальну тривалість проєкту СУІБ, і затримка будь-якої роботи на ньому затягує реалізацію та потребує залучення додаткових ресурсів.

Некритичні роботи мають резерв часу, тому їх затримка не впливає на строк реалізації. Для скорочення тривалості проєкту доцільно прискорювати лише критичні роботи, оскільки оптимізація некритичних лише підвищує витрати без зміни загального часу.

Окрім найранішого та найпізнішого строків настання подій, важливими є й відповідні параметри для робіт, які визначають часові межі їх виконання без впливу на загальну тривалість проєкту.

Усі отримані результати фіксуються у зведеній таблиці, що допомагає відповідальній особі комплексно оцінювати характеристики кожної роботи: визначати її критичність, обсяг резервного часу, а також ступінь впливу на подальші етапи проєкту. Резерв часу відіграє ключову роль у керуванні ризиками, забезпечуючи гнучкість і стабільність проєкту у випадку непередбачуваних затримок або змін. Його наявність дозволяє перерозподіляти час, підтримувати графік і підвищувати впевненість у досягненні цілей.

У методах керування проєктами: методі критичного шляху (CPM) і методі оцінки та аналізу програми (PERT) основна увага приділяється часовим параметрам виконання робіт [18]. Однак для досягнення оптимізації ресурсів важливо враховувати фінансову складову, оскільки тривалість робіт може змінюватися залежно від обсягів виділених ресурсів. Скорочення часу критичних робіт дозволяє зменшити загальну тривалість проєкту, але потребує додаткових фінансових вкладень. Прискорення всіх робіт може бути нераціональним через значне зростання витрат без суттєвого ефекту. Отже, важливо зберігати баланс між часом і коштами для досягнення оптимального результату. Доцільно ввести додаткові поняття: запланована тривалість виконання роботи та прискорена. Припустимо, що ці параметри заздалегідь визначені та відображені в таблиці для відповідної мережі.

На основі введених даних пропонується здійснювати три варіанти аналізу.

1) Визначення критичного шляху та розрахунок витрат при запланованих строках виконання робіт.
2) Аналіз критичного шляху та витрат у разі повного прискорення всіх робіт.
3) Оцінка критичного шляху та витрат при прискоренні лише критичних робіт.

Ці сценарії дозволяють порівняти загальну тривалість реалізації проєкту та відповідні фінансові витрати. Отримані результати можуть бути використані для обґрунтованого прийняття управлінських рішень щодо оптимального розподілу ресурсів.

На основі трьох розрахованих сценаріїв визначаються характеристики робіт (найбільш ранній початок, найпізніше завершення, резерв часу тощо) та будується гістограма виконання робіт у часі з відповідними ресурсними потребами.

До кожної гістограми будуються графіки сумарних фінансових інвестицій і залучення людських ресурсів. Це дає можливість відповідальним особам регулювати споживання ресурсів відповідно до лімітів і резервів часу. Регулювання

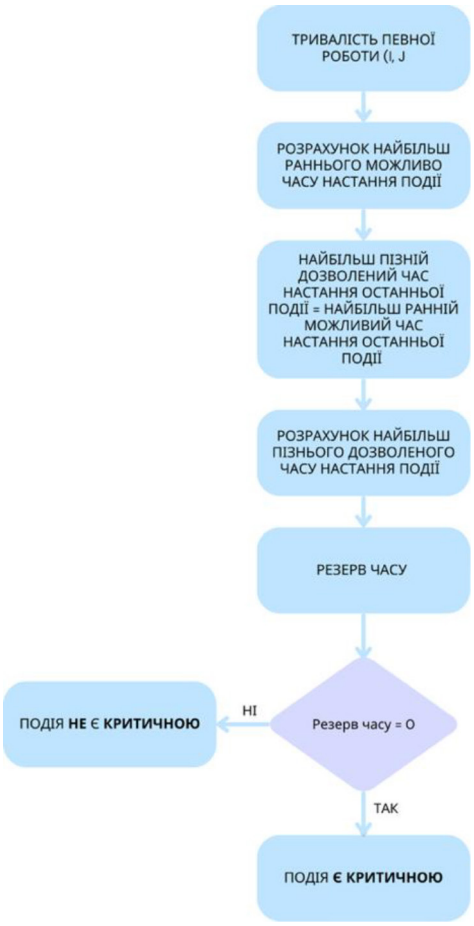


Рис. 2. Послідовність розрахунків для визначення критичності подій

здійснюється шляхом маневрування робочою силою та зміщенням строків некритичних робіт у межах допустимих резервів, що дозволяє рівномірно розподілити навантаження без збільшення загальної тривалості проєкту.

Після виконання основної частини запропонованої методики постає завдання її інтеграції у процеси компанії, що становить завершальний етап, а саме етап впровадження. Реалізація етапу впровадження потребує поєднання технічних, організаційних і адміністративних заходів. До них належить підготовка персоналу, налаштування процесів моніторингу й контролю, інтеграція обраних ЗЗІ. Особливу роль відіграє підтримка з боку керівництва, адже саме від їх рішень залежить рівень ресурсного забезпечення та ефективність реалізації всієї СУІБ. Важливо підкреслити, що впровадження СУІБ здійснюється на основі тієї методики, яка була обґрунтовано обрана на попередньому етапі запропонованої методики. Це означає, що запропонований підхід не замінює існуючі стандарти й нормативні документи, а слугує надбудовою, яка дозволяє оптимізувати процес розподілу ресурсів, вибору засобів захисту та управлінських рішень у межах обраної методики. Таким чином, етап впровадження виступає завершальною та критично важливою фазою запропонованої методики, що забезпечує її практичну цінність і перетворює теоретичні результати на інструмент ухвалення управлінських рішень у сфері ІБ.

Результати дослідження

У дослідженні було розглянуто два підходи до розв'язання багатокритеріальної задачі вибору оптимальної методики побудови СУІБ: MAI та MAM. Основна відмінність між ними полягає в тому, що MAI враховує лише прямі ієрархічні залежності, тоді як MAM дозволяє моделювати взаємозв'язки будь-якого типу, включно з впливом альтернатив на критерії.

Нехай представлено наступні альтернативи для вибору методики побудови СУІБ організації. Лінійка ISO/IEC 2700x, що представлена сукупністю стандартів ISO/IEC 27001 (вимоги до СУІБ), ISO/IEC 27002 (заходи забезпечення ІБ), ISO/IEC 27003 (основна частина керівництва), тощо. Методологія IT-Grundschutz, що представлена сукупністю стандартів BSI-200-1 (вимоги), BSI-200-2 (сама методологія), тощо. А також Постанова НБУ № 95.

Для порівняння уведені такі критерії:

– простота використання;
– методологічне забезпечення;
– необхідний рівень компетентності осіб для розробки та впровадження;

– можливість використання математичних підходів аналізу;
– конкретність вимог до СУІБ;
– універсальність вимог;
– частота використання.

Для оцінки компонентів визначається спеціальна шкала від 1 до 9, де 1 – рівнозначні важливості (по діагоналі матриці), при помірній перевазі – 3, при значимій – 5, і так до 9. При заповненні матриці має місце зворотна симетрія. У результаті після отримання необхідних матриць виконується змістовний аналіз.

На основі експертної оцінки, проведеної за визначеними критеріями, обидва методи MAI та MAM продемонстрували схожі результати (див. рис. 3), що підтвердило доцільність вибору IT-Grundschutz як базової методики для побудови СУІБ.

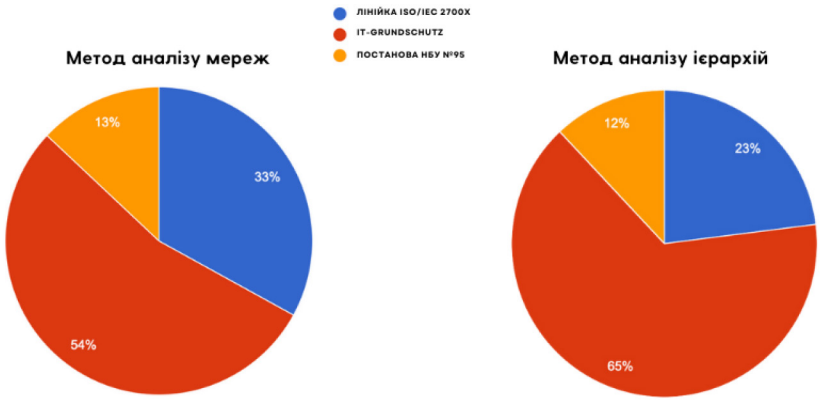


Рис. 3. Порівняльний аналіз результатів багатокритеріальної оптимізації

Водночас MAM, завдяки ширшому охопленню зв'язків, частково переорієнтував оцінки: сімейство ISO/IEC 2700x зайняло 10 %, а Постанова НБУ № 95 лише 1 %, що свідчить про більш глибокий аналіз. Обидва методи однозначно визнали найменш доцільною саме Постанову НБУ № 95, що є зрозумілим, зважаючи на її специфіку та вузькоспрямованість. Таким чином, метод MAM доцільно застосовувати у випадках, коли результати методу MAI потребують додаткової деталізації та врахування складніших взаємозв'язків між елементами задачі, що забезпечує більш глибоку та структуровану оцінку.

Завершивши вибір базової методики побудови СУІБ, методика пропонує перехід до наступного етапу: вибір оптимальної чистої стратегії CISO та забезпечення раціонального розподілу фінансових вкладень у ЗЗІ.

Розглянемо умовну організацію, у якій CISO прагне побудувати ефективну СУІБ із впровадженням лише тих засобів захисту, що є дійсно необхідними, враховуючи обмеженість затвердженого бюджету. Завдання зводиться до оптимізації вибору ЗЗІ.

У якості прикладу серед потенційних загроз інформаційній системі виділимо:
– шкідливе програмне забезпечення (ШПЗ);
– DDoS-атаки;
– фізичні впливи;
– фішингові атаки.
– Для протидії зазначеним загрозам розглядатимуться такі засоби захисту:
– антивірусне програмне забезпечення;
– криптографічні засоби (шифрувальне ПЗ);
– система захисту від DDoS-атак.

CISO має у своєму розпорядженні статистичні дані щодо реалізації загроз, отримані як із власного досвіду, так і на основі відомих джерел, зокрема CAPEC, CWE, OWASP, ENISA. Це дає змогу здійснити якісне моделювання ризиків і сценаріїв атак. Слід зазначити, що кількість загроз та засобів захисту у моделі не обмежується наведеними прикладами. Їх використано з метою спрощення розрахунків та наочної демонстрації можливостей запропонованої математичної моделі.

Перед пошуком розв'язків сформульованої гри необхідно визначити ключові показники для розрахунку функцій корисності та формування матриць учасників. Вихідні дані узгоджуються з аналітичними звітами організацій які піклуються питаннями інформаційної безреки, наприклад, ENISA, Gartner, BSI [19] – [21]. У таблиці 1 наведено значення показників Exposure Factor (EF) та Annual Rate of Occurrence (ARO), що використовуються для обчислення ROSI у матриці виграшів адміністратора безпеки. Додатково враховується показник Risk Mitigated (RM), який відображає ефективність засобів захисту та потребує експертної оцінки з боку CISO чи команди ІБ.

Таблиця 1

Початкові значення показників Exposure Factor, Annual Rate of Occurrence та Risk Mitigated

ЗЗІ	Показник	ШПЗ	Фізичний вплив	Фішингові атаки	DDoS атаки
–	EF	0,87	0,93	0,84	0,8
–	ARO	1	0,5	3	2
Антивірусне ПЗ	RM	0,85	0,05	0,2	0,05
Шифрувальне ПЗ		0,1	0,75	0,04	0,01
СЗ від DDoS атак		0	0	0	0,75
Антивірусне ПЗ + шифрувальне ПЗ		0,95	0,8	0,24	0,06
Шифрувальне ПЗ + СЗ від DDoS атак		0,1	0,75	0,04	0,76
Антивірусне ПЗ + СЗ від DDoS атак		0,85	0,05	0,2	0,8
Антивірусне ПЗ + шифрувальне ПЗ + СЗ від DDoS атак		0,95	0,8	0,24	0,81

На основі формули (1) було розраховано відповідні функції корисності CISO для кожної пари ЗЗІ-атака. Повна матриця для проведення аналізу показана на рис. 4.

p_i	27%	12%	41%	20%
	ШПЗ	Фізичний вплив	Фішингові атаки	DDoS атаки
Антивірусне ПЗ	23,65	-0,22	15,8	1,67
Шифрувальне ПЗ	0,45	4,81	0,68	-0,73
СЗ від DDoS атак	-1	-1	-1	16,14
Антивірусне ПЗ + шифрувальне ПЗ	8,18	3,13	5,72	0,07
Шифрувальне ПЗ + СЗ від DDoS атак	-0,33	1,68	-0,22	8,35
Антивірусне ПЗ + СЗ від DDoS атак	6,4	-0,77	4,04	11,8
Антивірусне ПЗ + шифрувальне ПЗ + СЗ від DDoS атак	4,17	1,33	2,78	7,1

Рис. 4. Матриця з функціями виграшів захисника та ймовірнісним розподілом стратегій зловмисника

Математичне очікування дискретної випадкової величини, тобто виграш, що отримує CISO, при можливому використанні кожної його стратегії розраховуємо за формулою (2). Результати розрахунку представлені на рис. 5.

Математичне очікування дискретної випадкової величини $U_A(a_i)$						
Антивірусне ПЗ	Шифрувальне ПЗ	СЗ від DDoS атак	Антивірусне ПЗ + шифрувальне ПЗ	Шифрувальне ПЗ + СЗ від DDoS атак	Антивірусне ПЗ + СЗ від DDoS атак	Антивірусне ПЗ + шифрувальне ПЗ + СЗ від DDoS атак
13,1711	**0,8315**	**2,428**	**4,9434**	**1,6923**	**5,652**	**3,8453**

Рис. 5. Розрахунок виграшів для кожної стратегії захисника

Також розв'язання сформульованої задачі здійснюється шляхом моделювання біматричної гри між CISO та зловмисником на основі підходу, описаного в роботі [12], що дозволяє забезпечити подальше коректне порівняння результатів. Для розгляду суперництва на основі біматричної гри необхідно створити матрицю й для зловмисника. У підготовчому етапі визначаються можливі витрати зловмисника на підготовку атаки та проведення, а також можливі втрати через впровадження певного механізму захисту адміністратором. Таким чином нехай адміністратором безпеки було визначено наступні значення показників cost та loss відповідно, що наведені у таблиці 2.

Таблиця 2

Значення показників cost та loss для розрахунку показників Return On Attack

	ШПЗ	Фізичний вплив	Фішингові атаки	DDoS атаки
cost	100000 (у.о.)	5000 (у.о.)	40000 (у.о.)	80000 (у.о.)
loss	60000 (у.о.)	10000 (у.о.)	50000 (у.о.)	30000 (у.о.)

Порівняння результатів двох досліджень, що враховують біматричну гру та статистику реалізації загроз зловмисником, наведено на рис. 6.

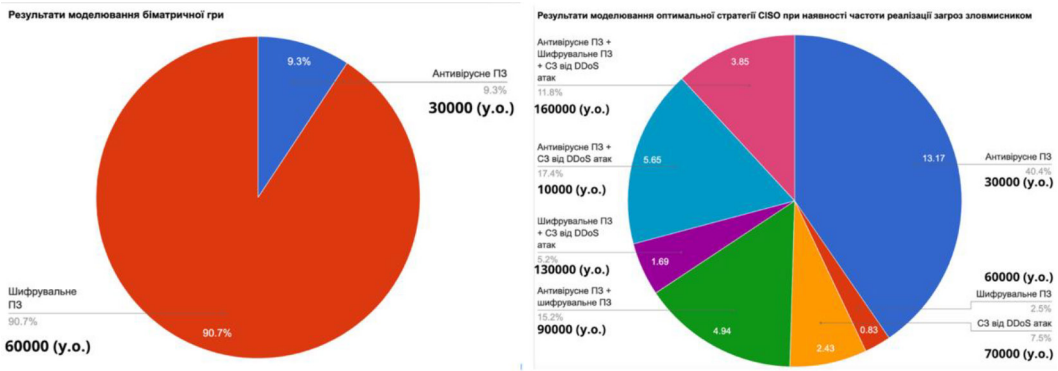


Рис. 6. Порівняння результатів двох досліджень за біматричною грою та частотами реалізації загроз зловмисником

Згідно з проведеним моделюванням на основі безкоаліційної гри двох гравців із ненульовою сумою, оптимальна стратегія супротивника визначала доцільність використання антивірусного ПЗ у 9 % випадків та шифрувального ПЗ у 91 %, із прогнозованою сумарною вартістю впровадження до 90 000 у.о. Таким чином, біматричне моделювання вказало на ефективність комбінованого залучення двох ЗЗІ (антивірусного та шифрувального ПЗ), тоді як моделювання у чистих стратегіях при відомому розподілі ймовірностей дозволило визначити оптимальний вибір CISO з погляду максимізації показника ROSI. Зокрема, використання лише антивірусного ПЗ із вартістю 30 000 у.о. забезпечує найвищий виграш, що свідчить про оптимальність цієї стратегії за заданих умов.

У межах другого етапу основної частини запропонованої методики здійснюється побудова мережевої моделі проєкту впровадження СУІБ, яка стала основою для подальшого аналізу та оптимізації за допомогою методів мережевого аналізу та керування проєктами.

Відповідно до формули (3), нехай розроблена мережа представляється у вигляді набору чисел (11, 12), де перше значення – множина вершин, а друге – множина робіт. CISO увідповіднив літерні позначення з найменуванням виконуваних робіт на основі власних експертних оцінок та аналізу специфіки проєкту. Було використано умовні числові значення, що співвідносяться з аналітичними звітами провідних міжнародних агентств (ENISA, Gartner, BSI) [19]–[21], а також із практичним досвідом і експертними оцінками. На основі проведеного аналізу та кореляції даних побудовано таблицю (див. рис. 7), яка містить заплановані та прискорені строки виконання робіт, відповідні фінансові та людські витрати. Таким чином, кожну роботу проєкту можна виконати у стандартний строк із типовими витратами або у скорочений строк із підвищеним ресурсним навантаженням.

Перше дослідження стосується аналізу критичного шляху за умов стандартних (запланованих) строків виконання робіт. Для його проведення всі параметри мережі, обчислені згідно з формулами (4)–(7), і узагальнено у зведеній таблиці, яка відображає критичний шлях розробленої моделі проєкту СУІБ.

Аналіз критичного шляху при запланованих строках показав загальну тривалість проєкту: 44 од. часу, із загальними витратами 344 000 у.о. На цьому етапі CISO, як керівник проєкту, зосереджується на критичних роботах, розуміючи, що саме вони визначають загальну тривалість. Доцільно перейти до аналізу ефективності проєкту за умов прискореного виконання всіх робіт.

Критичні роботи залишилися незмінними, однак тривалість проєкту скоротилася з 44 до 35,5 од. часу (приблизно на 20 %). Цього вдалося досягти шляхом суттєвого збільшення фінансових і людських ресурсів. Загальна вартість проєкту зросла з 344 000 до 453 000 у.о., на 32 %. Наслідком неправильного рішення щодо прискорення усіх робіт може стати нераціональне використання ресурсів. Третє дослідження зосереджується на аналізі критичного шляху за умови прискорення лише критичних робіт.

Аналіз даних показує, що прискорення лише критичних робіт не змінило б загальну тривалість проєкту порівняно з повним прискоренням, вона залишилася на рівні 35,5 од. часу. Однак вартість проєкту склала 433 000 у.о., що на 5 % менше, ніж у випадку прискорення всіх робіт, і свідчить про оптимальне використання ресурсів. У порівнянні зі звичайним графіком, витрати зросли на 25 %, але при цьому тривалість скоротилася на 8,5 од. часу (≈20 %), що підтверджує доцільність вибору цього сценарію.

З метою розширення аналізу було досліджено вплив прискорення лише некритичних робіт. Результати показали, що загальна тривалість проєкту залишилася на рівні 44 од. часу, однак вартість зросла до 364000 у.о., що на 6 % більше, ніж у базовому сценарії з аналогічною тривалістю. Таким чином, прискорення некритичних робіт не впливає на строк завершення проєкту, але призводить до неефективного використання ресурсів. Результати наведено на рис. 8.

	Робота	A	B	C	D	E	F	G	H	I	J	L	N
	Попередня вершина $i \in K$	1	2	3	3	4	5	6	6	7	8	9	10
	Наступна вершина $j \in K$	2	3	4	5	5	6	7	8	8	9	10	11
Заплановані строки	Тривалість dp_{ij}	4 од. часу	2 од. часу	2 од. часу	6 од. часу	6 од. часу	4 од. часу	3 од. часу	16 од. часу	6 од. часу	5 од. часу	3 од. часу	2 од. часу
	Витрати, $expp_{ij}$	25000 у.о.	7000 у.о.	7000 у.о.	25000 у.о.	30000 у.о.	25000 у.о.	20000 у.о.	100000 у.о.	10000 у.о.	60000 у.о.	20000 у.о.	15000 у.о.
	Необхідна кількість залучених осіб, $pplp_{ij}$	2	2	1	3	2	3	2	5	3	4	2	3
Прискорені строки	Тривалість, df_{ij}	3 од. часу	1,5 од. часу	1,5 од. часу	4,5 од. часу	5 од. часу	3 од. часу	2 од. часу	13 од. часу	5 од. часу	5 од. часу	2,5 од. часу	1 од. часу
	Витрати, $expf_{ij}$	30000 у.о.	10000 у.о.	9000 у.о.	35000 у.о.	40000 у.о.	35000 у.о.	25000 у.о.	140000 у.о.	15000 у.о.	60000 у.о.	24000 у.о.	30000 у.о.
	Необхідна кількість залучених осіб $pplf_{ij}$	2	3	1	4	3	3	2	7	3	4	2	3

Рис. 7. Зведення інформації щодо тривалості робіт та витрат ресурсів

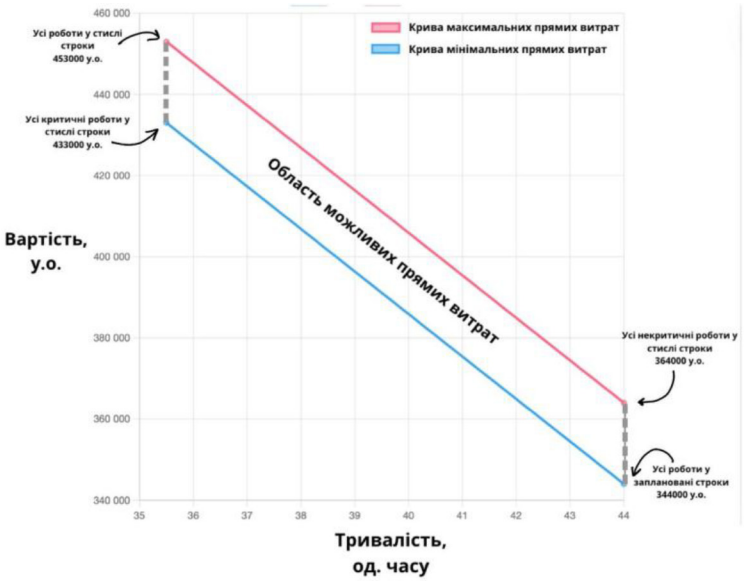


Рис. 8. Графік залежності між тривалістю проекту та прямими витратами

У результаті розрахунків побудовано графік, що наочно демонструє залежність між тривалістю проєкту та прямими витратами. Із графіка видно, що прискорення некритичних робіт не впливає на строки, але збільшує витрати на 6 % без жодної вигоди щодо часу. Прискорення всіх робіт дозволяє зменшити тривалість проєкту до 35,5 од. часу, однак вартість зростає на 32 %. Натомість, прискорення лише критичних робіт забезпечує аналогічне скорочення строків при меншому зростанні витрат лише на 26 %. Таким чином, CISO отримує інструмент для обґрунтованого прийняття управлінських рішень: визначати оптимальний баланс між строками реалізації проєкту та фінансовими ресурсами, виявляти критично важливі роботи та уникати нераціонального розподілу ресурсів.

Показники найбільш раннього можливого початку виконання роботи та найбільш пізнього часу для її завершення дозволяють побудувати гістограми, що відображають фактичне виконання робіт у часі, потребу у фінансових ресурсах та рівень залучення персоналу протягом усього проєкту. Такі візуалізації дають CISO змогу більш раціонально планувати проєкт, а залученим особам обрати найкращу конфігурацію реалізації з урахуванням доступних ресурсів.

Запропонована комплексна методика побудови СУІБ на основі багатокритеріального аналізу та мережевого моделювання дозволяє перейти від абстрактного уявлення до чітких кількісних оцінок, необхідних для ефективного планування, розподілу ресурсів та ухвалення обґрунтованих управлінських рішень. Такий підхід є особливо корисним як для фахівця з інформаційної безпеки, так і для керівництва організації, адже забезпечує прозорість, обґрунтованість і гнучкість на всіх етапах реалізації проєкту побудови СУІБ.

Висновки та перспективи подальших досліджень

Запропонована методика побудови СУІБ складається з трьох етапів (початковий, основний, впроваджувальний) і поєднує метод аналізу мереж, метод аналізу ієрархій та математичний апарат теорії ігор. На першому етапі основної частини методики вирішується багатокритеріальна задача вибору базової методики побудови СУІБ для конкретної організації. Порівняння результатів MAM і MAI на основі визначених критеріїв показало схожі висновки, найдоцільнішою є методика IT-Grundschutz (54 % та 33 % відповідно), причому MAM краще усереднює оцінки, виключаючи явно недоцільні варіанти.

Другий етап основної частини запропонованої методики присвячено визначенню оптимальних засобів захисту інформації з використанням теорії ігор. Запропоновано спрощену модель із чистими стратегіями для CISO, засновану на ймовірностях реалізації загроз зловмисником. Практичне моделювання показало, що найбільш доцільним варіантом є впровадження антивірусного ПЗ вартістю 30 000 у.о., що забезпечує максимальний виграш для CISO.

Третій етап базується на методі аналізу мереж для побудови моделі проєкту впровадження СУІБ, виявлення критичних і некритичних робіт, аналізу резервів часу та оптимізації використання ресурсів. Моделювання продемонструвало два варіанти ресурсних інвестицій: 44 од. часу та 344 000 у.о. при стандартних строках виконання усіх робіт проекту або 35,5 од. часу та 433 000 у.о. при прискоренні виконання усіх критичних робіт. Показано, що помилкові управлінські рішення можуть призвести до необґрунтованого зростання витрат без наявних переваг. Методика дозволяє уникнути цього завдяки попередньому моделюванню та візуалізації результатів у формі таблиць і графіків, що спрощує аргументацію перед керівництвом.

Подальші дослідження передбачають врахування невизначеності, ризиків та можливих змін у проєкті, проведення оцінювання ризиків, а також побудову графа атак для моделювання потенційних сценаріїв компрометації системи. Це дозволить поєднати управління проєктом із управлінням ризиками та забезпечити адаптивне планування СУІБ в умовах динамічного середовища.

Список використаної літератури

1. *ISO/IEC 27001:2022 Information security, cybersecurity and privacy protection – Information security management systems – Requirements*. (б. д.). iso.org. https://www.iso.org/standard/27001

2. *IT-Grundschutz*. (б. д.). Federal Office for Information Security. https://www.bsi.bund.de/EN/Themen/Unternehmen-und-Organisationen/Standards-und-Zertifizierung/IT-Grundschutz/it-grundschutz_node.html

3. Про затвердження Положення про організацію заходів із забезпечення інформаційної безпеки в банківській системі України, Постанова Національного банку України № 95 (2017) (Україна). https://zakon.rada.gov.ua/laws/show/v0095500-17#Text

4. Hoban, S. (2024, 19 листопада). *What is A resource in project management? 7 key types to know*. dpm. https://thedigitalprojectmanager.com/projects/managing-schedules/what-is-resource-project-management/

5. Gordon, L. A., & Loeb, M. P. (б. д.). The economics of information security investment. У *Economics of information security*. Kluwer Academic Publishers. https://doi.org/10.1007/1-4020-8090-5_9

6. Фукс, М. А. (2024). Моделювання багатокритеріальної задачі оптимізації вибору методики побудови СУІБ методами Томаса Сааті. У *Радіоелектроніка та молодь у XXI столітті* (с. 92–93). ХНУРЕ.

7. Жаринова, С., & Бабенко, А. (2014). Оптимізація інвестицій в інформаційну безпеку підприємства. *Науково-технічний журнал*, *3*(83), 115.

8. Lemeshko, O., Yeremenko, O., Kurenko, V., & Fuks, M. (2024). Method of designing a cyber-resilient information and communication network. *Problemi telekomunìkacìj*, (2(35)), 14–25. https://doi.org/10.30837/pt.2024.2.02

9. Пшеничних, С. В., Добринін, І. С., & Клочкова, Д. Ю. (2023). Математична модель оптимального вибору засобів захисту інформації при проектуванні комплексної системи захисту на об'єкті інформатизації. *Проблеми телекомунікацій*, *1*(32), 45–58. https://doi.org/10.30837/pt.2023.1.04

10. Saaty, T. L. (б. д.). The analytic network process. У *Decision making with the analytic network process* (с. 1–26). Springer US. https://doi.org/10.1007/0-387-33987-6_1

11. Saaty, T. L., & Vargas, L. G. (2012). *Models, methods, concepts & applications of the analytic hierarchy process*. Springer US. https://doi.org/10.1007/978-1-4614-3597-6

12. Фукс, М. А., & Добринін, І. С. (2023). Визначення стратегії захисту інформації на основі безкоаліційної гри двох гравців із ненульовою сумою. У *Інформаційно-комунікаційні технології та кібербезпека (ІКТК-2023)* (с. 185–188). ХНУРЕ. https://openarchive.nure.ua/handle/document/25471

13. *CAPEC – CAPEC list version 3.9*. (2019, 30 вересня). CAPEC – Common Attack Pattern Enumeration and Classification (CAPEC™). https://capec.mitre.org/data/index.html

14. *OWASP Top Ten | OWASP Foundation*. (б. д.). OWASP Foundation, the Open Source Foundation for Application Security | OWASP Foundation. https://owasp.org/www-project-top-ten/

15. *GitHub – CVEProject/cvelistV5: CVE cache of the official CVE List in CVE JSON 5 format*. (б. д.). GitHub. https://github.com/CVEProject/cvelistV5

16. *Common vulnerability scoring system CVSS SIG*. (б. д.). FIRST – Forum of Incident Response and Security Teams. https://www.first.org/cvss/

17. Phillips, D. T. (1981). *Fundamentals of network analysis*. Prentice-Hall.

18. Aston, B. (2025, 29 липня). *PERT vs CPM: 8 key differences for project managers to grasp*. dpm. https://thedigitalprojectmanager.com/project-management/pert-vs-cpm

19. *Publications | ENISA*. (б. д.). ENISA. https://www.enisa.europa.eu/publications#c3=2014&c3=2024&c3=false&c5=publicationDate&reversed=on&b_start=0

20. *Business and technology insights and trends*. (б. д.). GARTNER. https://www.gartner.com/en/insights

21. *Federal office for information security*. (б. д.). BSI. https://www.bsi.bund.de/EN/Home/home_node.html

References

1. ISO/IEC 27001:2022. *Information security, cybersecurity and privacy protection – Information security management systems – Requirements*. (n.d.). iso.org. https://www.iso.org/standard/27001

2. *IT-Grundschutz*. (n.d.). Federal Office for Information Security. https://www.bsi.bund.de/EN/Themen/Unternehmen-und-Organisationen/Standards-und-Zertifizierung/IT-Grundschutz/it-grundschutz_node.html

3. *On the approval of the Regulation on the organization of measures to ensure information security in the banking system of Ukraine*, Resolution of the National Bank of Ukraine No. 95 (2017) (Ukraine). https://zakon.rada.gov.ua/laws/show/v0095500-17#Text

4. Hoban, S. (2024, November 19). *What is a resource in project management? 7 key types to know*. dpm. https://thedigitalprojectmanager.com/projects/managing-schedules/what-is-resource-project-management/

5. Gordon, L. A., & Loeb, M. P. (n.d.). *The economics of information security investment*. In *Economics of information security*. Kluwer Academic Publishers. https://doi.org/10.1007/1-4020-8090-5_9

6. Fuks, M. A. (2024). Modeling a multi-criteria optimization problem for selecting a methodology for building an ISMS using Thomas Saaty's methods. In Radioelectronics and Youth in the XXI Century (pp. 92–93). Kharkiv National University of Radioelectronics (KhNURE).

7. Zharinova, S., & Babenko, A. (2014). *Optimization of investments in the information security of an enterprise*. Scientific and Technical Journal, 3(83), 115.

8. Lemeshko, O., Yeremenko, O., Kurenko, V., & Fuks, M. (2024). Method of designing a cyber-resilient information and communication network. *Problemi telekomunìkacìj*, (2(35)), 14–25. https://doi.org/10.30837/pt.2024.2.02

9. Pshenychnych, S., Dobrynin, I., & Klochkova, D. (2023). Mathematical model of optimal selection of information security measures in the design of an integrated security system at an informatization facility. *Problemi telekomunìkacìj*, (1(32)), 45–58. https://doi.org/10.30837/pt.2023.1.04

10. Saaty, T. L. (n.d.). *The analytic network process*. In *Decision making with the analytic network process* (pp. 1–26). Springer US. https://doi.org/10.1007/0-387-33987-6_1

11. Saaty, T. L., & Vargas, L. G. (2012). *Models, methods, concepts & applications of the analytic hierarchy process*. Springer US. https://doi.org/10.1007/978-1-4614-3597-6

12. Fuks, M. A., & Dobrynin, I. S. (2023). Determining the strategy of information protection based on a non-cooperative two-player game with a non-zero sum. In *Information and Communication Technologies and Cybersecurity (ICTC-2023)* (pp. 185–188). Kharkiv National University of Radio Electronics. https://openarchive.nure.ua/handle/document/25471

13. *CAPEC – CAPEC list version 3.9*. (2019, 30 вересня). CAPEC – Common Attack Pattern Enumeration and Classification (CAPEC™). https://capec.mitre.org/data/index.html

14. *OWASP Top Ten | OWASP Foundation*. (б. д.). OWASP Foundation, the Open Source Foundation for Application Security | OWASP Foundation. https://owasp.org/www-project-top-ten/

15. *GitHub – CVEProject/cvelistV5: CVE cache of the official CVE List in CVE JSON 5 format*. (б. д.). GitHub. https://github.com/CVEProject/cvelistV5

16. *Common vulnerability scoring system CVSS SIG*. (б. д.). FIRST – Forum of Incident Response and Security Teams. https://www.first.org/cvss/

17. Phillips, D. T. (1981). *Fundamentals of network analysis*. Prentice-Hall.

18. Aston, B. (2025, July 29). *PERT vs CPM: 8 key differences for project managers to grasp*. dpm. https://thedigitalprojectmanager.com/project-management/pert-vs-cpm

19. *Publications | ENISA*. (n.d.). ENISA. https://www.enisa.europa.eu/publications#c3=2014&c3=2024&c3=false&c5=publicationDate&reversed=on&b_start=0

20. *Business and technology insights and trends*. (n.d.). GARTNER. https://www.gartner.com/en/insights

21. *Federal Office for Information Security*. (n.d.). BSI. https://www.bsi.bund.de/EN/Home/home_node.html

Дата першого надходження рукопису до видання: 23.09.2025
Дата прийнятого до друку рукопису після рецензування: 20.10.2025
Дата публікації:

UDC 004.75:004.932.2:621.391 DOI

M. O. KHVOSTIVSKYI
Candidate of Technical Sciences, Associate Professor,
Associate Professor at the Department of Biotechnical Systems
Ternopil Ivan Puluj National Technical University
ORCID: 0000-0002-2405-4930

L. V. KHVOSTIVSKA
Candidate of Technical Sciences, Associate Professor,
Associate Professor at the Department of Radiotechnical Systems
Ternopil Ivan Puluj National Technical University
ORCID: 0000-0002-4997-8339

I. Yu. DEDIV
Candidate of Technical Sciences, Associate Professor,
Associate Professor at the Department of Radiotechnical Systems
Ternopil Ivan Puluj National Technical University
ORCID: 0000-0002-4913-139X

L. Ye. DEDIV
Candidate of Technical Sciences, Associate Professor,
Associate Professor at the Department of Biotechnical Systems
Ternopil Ivan Puluj National Technical University
ORCID: 0000-0002-2963-6948

INTELLIGENT COMPUTER NETWORK TRAFFIC FORECASTING SYSTEM BASED ON SYNPHASE DATA PROCESSING

The article considers the problem of increasing the efficiency of computer network management by forecasting their traffic using intelligent data processing methods. The feasibility of using new generation stochastic models, in particular periodically correlated stochastic processes (PCSP), which allow simultaneously taking into account the random nature of traffic formation and its daily cyclicity, is substantiated. Unlike classical models (Poisson, Markov, ARIMA, fractal), which are limited in reproducing complex patterns, the PCSP-based approach provides a more adequate description of real processes in telecommunication systems.

The architecture of an intelligent forecasting system is proposed, the core of which is synphase data processing algorithms. The implemented procedures include parametric covariance estimation, spectral analysis of centered signals using the Fourier transform, and averaging of correlation components to reduce the influence of noise components. This ensures high accuracy of reproducing network load variability and forming stable forecasts even under conditions of stochastic deviations.

Experimental testing was conducted based on data from the Internet provider UFONet (Ternopil). It was found that the system correctly identifies critical peak intervals (1.8-2.1 TB in the evening) and periods of minimum load (0.5-0.7 TB at night). Additionally, an automatic load level classification module was implemented, which translates numerical forecasts into a categorical form (“minimum”, “average”, “critical”). Such a mechanism allows operators and providers to carry out proactive resource management, reduce the risks of overloads and plan maintenance in the least active time intervals.

The practical significance of the results obtained lies in creating a decision support tool for network infrastructure administrators, which increases the stability of the system and the quality of service provision to users. The scientific novelty of the work is determined by the synthesis of PCSP methods and synphase analysis in a single intelligent forecasting system. Prospects for further research are related to the integration of the developed system with machine learning algorithms and the expansion of its application to multi-level computer networks of the next generations.

Key words: *computer networks, intelligent system, traffic forecasting, synphase processing, periodically correlated random process, MATLAB.*

© Khvostivskyi M. O., Khvostivska L. V., Dediv I. Yu., Dediv L. Ye., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

М. О. ХВОСТІВСЬКИЙ
кандидат технічний наук, доцент,
доцент кафедри біотехнічних систем
Тернопільський національний технічний університет імені Івана Пулюя
ORCID: 0000-0002-2405-4930

Л. В. ХВОСТІВСЬКА
кандидат технічний наук, доцент,
доцент кафедри радіотехнічних систем
Тернопільський національний технічний університет імені Івана Пулюя
ORCID: 0000-0002-4997-8339

І. Ю. ДЕДІВ
кандидат технічний наук, доцент,
доцент кафедри радіотехнічних систем
Тернопільський національний технічний університет імені Івана Пулюя
ORCID: 0000-0002-4913-139X

Л. Є. ДЕДІВ
кандидат технічний наук, доцент,
доцент кафедри біотехнічних систем
Тернопільський національний технічний університет імені Івана Пулюя
ORCID: 0000-0002-2963-6948

ІНТЕЛЕКТУАЛЬНА СИСТЕМА ПРОГНОЗУВАННЯ ТРАФІКУ КОМП'ЮТЕРНИХ МЕРЕЖ НА ОСНОВІ СИНФАЗНОЇ ОБРОБКИ ДАНИХ

У статті розглянуто проблему підвищення ефективності управління комп'ютерними мережами шляхом прогнозування їхнього трафіку з використанням інтелектуальних методів обробки даних. Обґрунтовано доцільність застосування стохастичних моделей нового покоління, зокрема періодично корельованих випадкових процесів (ПКВП), що дозволяють одночасно врахувати випадкову природу формування трафіку та його добову циклічність. На відміну від класичних моделей (Пуассонівських, Марковських, ARIMA, фрактальних), які обмежені у відтворенні комплексних закономірностей, підхід на основі ПКВП забезпечує більш адекватний опис реальних процесів у телекомунікаційних системах.

Запропоновано архітектуру інтелектуальної системи прогнозування, ядром якої є алгоритми синфазної обробки даних. Реалізовані процедури включають оцінювання параметричної коваріації, спектральний аналіз центрованих сигналів за допомогою перетворення Фур'є та усереднення кореляційних компонент для зниження впливу шумових складових. Це забезпечує високу точність відтворення варіативності мережевого навантаження та формування стійких прогнозів навіть за умов стохастичних відхилень.

Експериментальна апробація проведена на основі даних інтернет-провайдера UFONet (м. Тернопіль). Встановлено, що система коректно ідентифікує критичні пікові інтервали (1,8-2,1 ТБ у вечірні години) та періоди мінімального навантаження (0,5-0,7 ТБ уночі). Додатково реалізовано модуль автоматичної класифікації рівнів навантаження, який переводить числові прогнози у категоріальну форму («мінімальне», «середнє», «критичне»). Такий механізм дозволяє операторам і провайдерам здійснювати проактивне управління ресурсами, зменшувати ризики перевантажень та планувати технічне обслуговування у найменш активні часові інтервали.

Практичне значення отриманих результатів полягає у створенні інструменту підтримки прийняття рішень для адміністраторів мережевої інфраструктури, що підвищує стабільність роботи системи та якість надання послуг користувачам. Наукова новизна роботи визначається синтезом методів ПКВП і синфазного аналізу в єдиній інтелектуальній системі прогнозування. Перспективи подальших досліджень пов'язані з інтеграцією розробленої системи з алгоритмами машинного навчання та розширенням її застосування на багаторівневі комп'ютерні мережі наступних поколінь.

Ключові слова: *комп'ютерні мережі, інтелектуальна система, прогнозування трафіку, синфазна обробка, періодично корельований випадковий процес, MATLAB.*

Statement of the problem

The intensive development of information technologies and digital services leads to an unprecedented increase in the load on computer networks. Millions of users generate large amounts of data every day in the form of multimedia content, streaming services, and transactional operations, which forms a highly dynamic and stochastic nature of network traffic. In the face of such challenges, the creation of intelligent forecasting systems that are able not only to monitor the current state of the network, but also to predict its future load in order to avoid overloads and ensure the high-quality functioning of the infrastructure becomes especially relevant.

Analysis of recent research and publications

Intelligent forecasting systems combine mathematical traffic models (from classical stochastic to modern hybrid and fractal), forecasting algorithms (autoregressive, Markov, machine learning, neural networks), software complexes with an interactive interface and network integration tools.

Over the past decades, a number of models have been proposed in world science to describe network traffic, in particular, the Poisson model [1], Markov models [2–3], ON–OFF models [4] describe the alternation of active and passive states; convenient for multimedia, but do not reproduce long-term correlation, modulated Markov (MMF, IPP) [5], autoregressive models (AR, ARIMA) [6], fractal models [7–8], TES models [9], models based on machine learning [10–11], hybrid models [12].

All traditional models have value in their classes of problems, but an intelligent forecasting system requires a model that simultaneously takes into account stochasticity, correlation, and periodicity (existing models do not provide this).

Formulation of the research objective

The most promising approach in this context is the approach based on periodically correlated stochastic processes (PCSPs) [13], which reflect the cyclical nature of daily fluctuations, phase variability and stochastic nature of traffic. The combination of this model with synphase analysis [14] of data opens up the possibility of creating a new generation of intelligent systems capable of predicting network load with high accuracy and generating recommendations for optimizing equipment operation.

Thus, the task of developing an intelligent computer network traffic forecasting system that combines a mathematical model of the PCSP, an algorithmic core of synphase processing, and software for practical use by providers and network administrators is relevant.

Presentation of the main research material

To substantiate the structure of the traffic load model, which is the core of the synphase analysis method and intelligent forecasting system, empirical data obtained from the Internet provider UFONet in Ternopil was used. Visualization of the dynamics of network traffic registered during a seven-day period (01-07.07.2024) in the residential complex “Park Complex” is shown in Fig. 1.

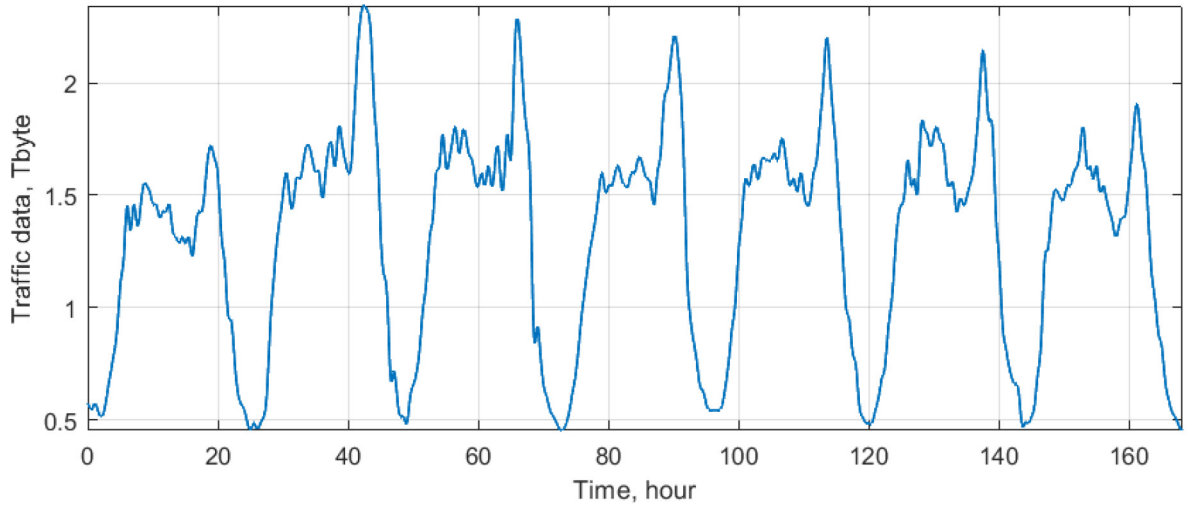


Fig. 1. Traffic data of the residential complex “Park Complex”

The graph (Fig. 1) of network traffic shows two key characteristics – variability and periodicity. Variability is manifested in fluctuations in traffic volumes from 0.5 to 2 TB, uneven load within cycles, and regular deviations from the average level of 1.2–1.7 TB. Periodicity reflects daily cyclicality: traffic peaks are repeated daily, the lowest values occur at night, and the highest values during the day and evening, which indicates the rhythmic nature of the load.

Therefore, the traffic in Fig. 1 has a pronounced variability and stable daily periodicity, which allows predicting the nature of loads and planning network resources taking into account recurring cycles.

The presence of significant variability (stochasticity) of traffic characteristics (Fig. 1) makes it impossible to apply a deterministic approach to its mathematical description. Under such conditions, it is advisable to use a stochastic approach, which provides adequate modeling of the processes of traffic load formation. Based on this approach, methods and intelligent forecasting systems are created that can take into account the specifics of real data.

The stochastic representation of traffic load opens up the possibility of building flexible and adaptive models. They reflect both the temporal periodicity of the signal and its amplitude instability, characteristic of real electronic communication systems.

A model of network traffic behavior, presented in the form of a PCSP, which takes into account the daily cyclicality and stochastic nature of traffic:

$$\xi(t)=\sum_{k\in Z}\xi_k(t)e^{-\frac{j2\pi k}{T}},\quad t\in R, \tag{1}$$

where $\xi_k(t)$ – stochastic component of traffic load; $e^{-\frac{j2\pi k}{T}}$ – periodic component of load traffic, $T = 24$ hour.

The stochastic traffic load model (1) reflects not only its real nature, but also forms the basis for the development of intelligent forecasting methods aimed at practical application in computer network control systems.

The PCSP methods implement three fundamental algorithms for processing network traffic data: synphase, component, and filter. The synphase approach is key for an intelligent forecasting system, as it analyzes the dynamics more deeply, first determining the covariance of the data, and then isolating the frequency components through the Fourier transform. In contrast, the component and filter methods estimate the covariance directly in the frequency domain. Synphase processing in an intelligent traffic forecasting system is based on the assumption that the informative characteristics of the network load can be represented as functions with recurring time-dependent features that are formed in the structure of correlation components:

$$\hat{B}_k(u)=\frac{1}{T}\sum_{k\in Z}\hat{b}_\xi(t,u)e^{-\frac{j2\pi k}{T}}, \tag{2}$$

where $\hat{b}_\xi(t,u)$ – parametric covariance; T – period of traffic load data corresponding to the duration of the day; u – time shift.

Correlation components (2) allow the intelligent system to distinguish characteristic features of network traffic, which significantly increases the accuracy of predicting its load and provides effective support for computer network management processes.

The procedure of averaging components (2) is a necessary step to increase the stability and reliability of the forecast according to the expression:

$$M_u\left\{\hat{B}_k(u)\right\}=\frac{1}{U}\sum_{k\in Z}\hat{B}_k(u), \tag{3}$$

Averaging $M_u\left\{\hat{B}_k(u)\right\}$ filters out noise fluctuations, allowing the intelligent system to generate generalized recommendations based on stable long-term traffic patterns. For the provider, this means focusing on strategic resource planning and improving the reliability of infrastructure management, while 3D visualization $\hat{B}_k(u)$ is designed for in-depth analysis of current and short-term processes.

The averaged synphase 3D components $M_u\left\{\hat{B}_k(u)\right\}$ are shown in Fig. 2.

Based on the analysis of averaged traffic data, it was found that the average load during the working period is about 1.4 TB per day, with deviations ranging from 0.5 TB during the night hours to over 2 TB in the evening. The variability is approximately ±0.6 TB from the average, which emphasizes the high stochasticity of the process.

The averaged correlation components allowed us to determine the time intervals with the highest load. In particular, peak values are predicted to occur in the period 19:00–23:00, when the load increases to 1.8–2.1 TB, while the minimum values are recorded in the period 03:00–06:00, when the traffic decreases to 0.5–0.7 TB. Averaging the results confirmed the repeatability of such dynamics throughout the entire weekly cycle, which gives grounds to consider it a stable characteristic of network behavior.

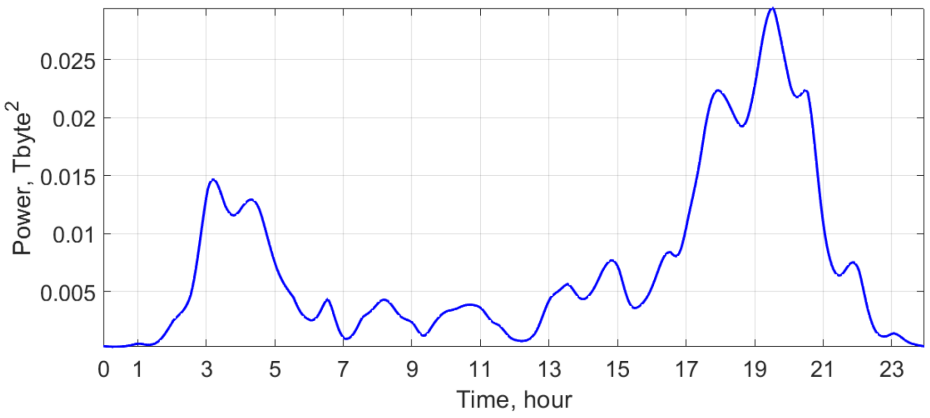


Fig. 2. Averaged synphase 3D components of computer network traffic load data

For an intelligent system, this means that forecasting has not only a qualitative but also a quantitative basis. The system is able to warn about possible overload at specific times of the day and provide the operator with recommendations, for example, about the need to reserve additional resources for evening periods or optimize maintenance during hours of minimum load. Thus, the combination of 3D visualization and averaged results with numerical characteristics creates a comprehensive picture that simultaneously reflects the instantaneous dynamics of the process and its long-term patterns.

An additional functional element of the intelligent system is the module for automatic detection of peak intervals. Its task is to generate time stamps with classification of the predicted load into categories:

– "critical load" – a traffic level exceeding a threshold value (for example, over 1.8–2.0 TB), when there is a risk of overloading network equipment;

– "average load" – the operating traffic range (1.0–1.7 TB), which corresponds to the normal functioning of the network;

– "minimum load" – time intervals with the lowest user activity (up to 0.7 TB), when it is possible to perform preventive work.

The module's operating algorithm is based on the results of synphase forecasting: after calculating synphase components and building a traffic forecast, the system automatically determines the intervals of exceeding the set thresholds and assigns the appropriate category.

It is assumed that the intelligent system generates predicted traffic values in the form of a time sequence:

$$X(u)=\left\{M_1\left\{\hat{B}_k(1)\right\}_1, M_2\left\{\hat{B}_k(2)\right\},\ldots,M_u\left\{\hat{B}_k(u)\right\}\right\},\quad u\in[0,U], \tag{3}$$

where $M_u\left\{\hat{B}_k(u)\right\}$ – predicted load at time u.

We define three load ranges based on threshold values:

$$\Omega=\begin{cases}\min\ load, & 0\le M_u\left\{\hat{B}_k(u)\right\}<\theta1\\ average\ load, & \theta1\le M_u\left\{\hat{B}_k(u)\right\}<\theta2,\\ \max\ load, & M_u\left\{\hat{B}_k(u)\right\}\ge\theta2\end{cases} \tag{4}$$

where θ1 і θ2 – adaptive threshold levels determined taking into account the statistics of the forecast data, $\theta1=\mu-\sigma1$ та $\theta2=\mu+\sigma$, where μ – average value, σ – standard deviation.

The thresholds θ1 and θ2 define the interval of statistically normal fluctuations of the traffic signal around the mean μ, which allows to automatically separate typical states from anomalous ones. Thus, the system moves from simple forecasting to intelligent decision support.

Fig. 3 shows the result of the automatic peak interval detection module, which is an additional functional element of the intelligent traffic forecasting system.

The intelligent system automatically calculated the thresholds θ1 and θ2:

– lower threshold $\theta1\approx0.009$ – "low load" limit, corresponding to periods of lowest user activity;

– upper threshold $\theta2\approx0.015$ – "critical load" limit, exceeding which may cause network equipment to be overloaded.

The graph in Fig. 3 shows two main peak segments where the amplitude exceeds θ2: the first – approximately between 3.14 and 3.28 hours, the second – from 17.29 to 20.52 hours. In these intervals, the system records the "critical load" and generates the corresponding time stamps for the operator.

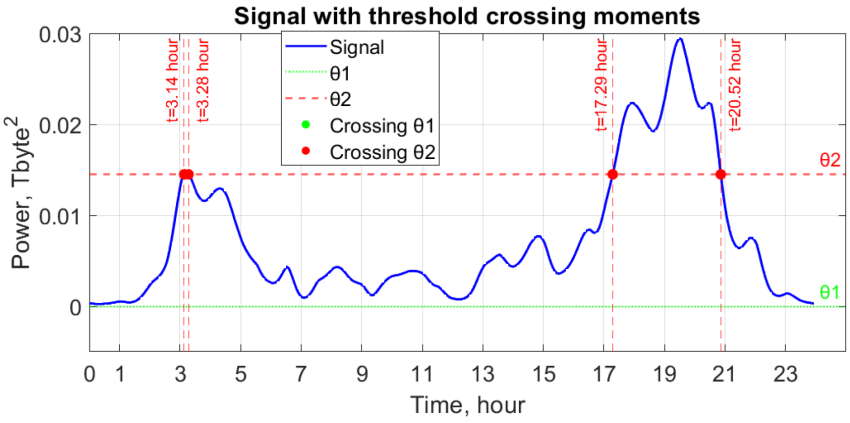


Fig. 3. The result of the automatic peak interval detection module

Between levels θ1 and θ2, the signal is in "medium load", which is the normal network operation mode. During hours when the amplitude drops below θ1 (for example, around 0-2 a.m. and after 10 p.m.), traffic is classified as "minimum load", which is optimal for preventive work.

Thus, the intelligent system receives not only a quantitative prediction of future load, but also a qualitative interpretation in the form of administrator-friendly recommendations. This allows providers to identify critical, average and minimum load time periods in advance, plan resource allocation and optimize the operation of the network infrastructure.

Conclusions

The article proposes an intelligent system for predicting computer network traffic, based on a combination of a stochastic model of periodically correlated random processes and synphase data processing algorithms. It is shown that this approach allows us to adequately take into account the stochastic nature of traffic and its daily cyclicity, which significantly increases the accuracy of forecasting.

The developed methods provide the ability to detect peak and minimum load intervals, automatically classify traffic levels, and generate operator-friendly recommendations for network resource management. Experimental studies have confirmed the system's ability to timely identify critical conditions and create a basis for proactive infrastructure management.

The practical significance of the work lies in creating a decision support tool for providers and network administrators, which allows reducing the risks of overloads, optimizing resource use, and improving the quality of service provision.

Further research is aimed at integrating the system with machine learning algorithms, as well as its adaptation for next-generation multi-level networks, which will provide expanded functionality and increased efficiency in managing complex telecommunications systems.

Bibliography

1. Erlang A. K. The theory of probabilities and telephone conversations. *Nyt Tidsskrift for Matematik*. 1909. Vol. 20. P. 33–39.

2. Czachórski T., Grochla K., Jozefiok A., Nycz T., Pekergin F. Performance evaluation of a multiuser interactive networking system: A comparison of modelling methods. *Proc. 26th International Symposium on Computer and Information Sciences (ISCIS 2011)*. London, UK. 2011. P. 215–221.

3. Domańska J., Domański A., Czachórski T. Internet traffic source based on hidden Markov model. *Lecture Notes in Computer Science*. Vol. 6869. Springer. 2011. P. 395–404. DOI: 10.1007/978-3-642-22875-9_36

4. Claypool M. The effect of latency on user performance in real-time strategy games. *Computer Networks*. 1995. Vol. 49. P. 52–70. DOI: 10.1016/j.comnet.2005.07.003

5. Daigle J. N. *Queueing Theory for Telecommunications*. New York: Addison-Wesley. 1990. 456 p.

6. Box G. E. P., Jenkins G. M. *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day. 1970. 553 p. DOI: 10.1002/9781118619190

7. Mandelbrot B. B., Van Ness J. W. Fractional Brownian motions, fractional noises and applications. *SIAM Review*. 1968. Vol. 10. P. 422–437. DOI: 10.1137/1010093

8. Leland W. E., Taqqu M. S., Willinger W., Wilson D. V. On the self-similar nature of Ethernet traffic (Extended version). */ACM Transactions on Networking*. 1994. Vol. 2(1). P. 1–15. DOI: 10.1109/90.282603

9. Reichl P. A. A generalized TES model for periodical traffic. *Proc. IEEE International Conference on Communications*. 1998. P. 1–5. DOI: 10.1109/ICC.1998.683236

10. Chen T. Network traffic modeling. In: Bidgoli H. (ed.) *Handbook of Computer Networks*. Wiley. 2007. P. 326–339. DOI: 10.1002/9781118256107.ch21

11. Zhang Y., Roughan M., Duffield N., Greenberg A. Fast accurate computation of large-scale IP traffic matrices from link loads. *ACM SIGMETRICS Performance Evaluation Review*. 2003. Vol. 31(1). P. 206–217. DOI: 10.1145/885651.781041

12. Zhang G., Patuwo B.E., Hu M.Y. Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting*. 1998. Vol. 14(1). P. 35–62. DOI: 10.1016/S0169-2070(97)00044-7

13. Хвостівський М. О., Осухівська Г. М., Хвостівська Л. В., Величко Д. В. Розвиток математичного моделювання трафіку комп'ютерних мереж. *Матеріали Міжнародної науково-технічної конференції 14–15 травня 2020 року «Фундаментальні та прикладні проблеми сучасних технологій»*, Тернопіль, Україна. 2021. С. 107–111. DOI: 10.1425/jsdtl

14. Khvostivska L., Khvostivskyi M., Dediv I., Yatskiv V., Palaniza Y. Method, algorithm and computer tool for synphase detection of radio signals in telecommunication networks with noises. *CITI 2023: Proc. of the 1st International Workshop on Computer Information Technologies in Industry 4.0*. Тернопіль, Україна. 2023. С. 173–180. ISSN 1613–0073.

References

1. Erlang, A. K. (1909). The theory of probabilities and telephone conversations. *Nyt Tidsskrift for Matematik, 20*, 33–39.

2. Czachórski, T., Grochla, K., Jozefiok, A., Nycz, T., & Pekergin, F. (2011). Performance evaluation of a multiuser interactive networking system: A comparison of modelling methods. In *Proceedings of the 26th International Symposium on Computer and Information Sciences (ISCIS 2011)* (pp. 215–221). London, UK.

3. Domańska, J., Domański, A., & Czachórski, T. (2011). Internet traffic source based on hidden Markov model. In *Lecture Notes in Computer Science* (Vol. 6869, pp. 395–404). Springer. https://doi.org/10.1007/978-3-642-22875-9_36

4. Claypool, M. (1995). The effect of latency on user performance in real-time strategy games. *Computer Networks, 49*, 52–70. https://doi.org/10.1016/j.comnet.2005.07.003

5. Daigle, J. N. (1990). *Queueing theory for telecommunications*. Addison-Wesley.

6. Box, G. E. P., & Jenkins, G. M. (1970). *Time series analysis: Forecasting and control*. Holden-Day. https://doi.org/10.1002/9781118619190

7. Mandelbrot, B. B., & Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review, 10*, 422–437. https://doi.org/10.1137/1010093

8. Leland, W. E., Taqqu, M. S., Willinger, W., & Wilson, D. V. (1994). On the self-similar nature of Ethernet traffic (Extended version). *IEEE/ACM Transactions on Networking, 2*(1), 1–15. https://doi.org/10.1109/90.282603

9. Reichl, P. A. (1998). A generalized TES model for periodical traffic. In *Proceedings of the IEEE International Conference on Communications* (pp. 1–5). https://doi.org/10.1109/ICC.1998.683236

10. Chen, T. (2007). Network traffic modeling. In H. Bidgoli (Ed.), *Handbook of computer networks* (pp. 326–339). Wiley. https://doi.org/10.1002/9781118256107.ch21

11. Zhang, Y., Roughan, M., Duffield, N., & Greenberg, A. (2003). Fast accurate computation of large-scale IP traffic matrices from link loads. *ACM SIGMETRICS Performance Evaluation Review, 31*(1), 206–217. https://doi.org/10.1145/885651.781041

12. Zhang, G., Patuwo, B. E., & Hu, M. Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting, 14*(1), 35–62. https://doi.org/10.1016/S0169-2070(97)00044-7

13. Khvostivskyi, M., Osukhivska, H., Khvostivska, L., Lobur, T., & Velychko, D. (2021). Mathematical modelling of daily computer network traffic. In *ITTAP-2021* (pp. 107–111). Ternopil, Ukraine. https://doi.org/10.1425/jsdtl

14. Khvostivska, L., Khvostivskyi, M., Dediv, I., Yatskiv, V., & Palaniza, Y. (2023). Method, algorithm and computer tool for synphase detection of radio signals in telecommunication networks with noises. In *Proceedings of the 1st International Workshop on Computer Information Technologies in Industry 4.0 (CITI 2023)* (pp. 173–180). Ternopil, Ukraine. ISSN 1613-0073.

Дата першого надходження рукопису до видання: 27.09.2025
Дата прийнятого до друку рукопису після рецензування: 23.10.2025
Дата публікації:

УДК 004.8:004.75 DOI

М. Ю. ХОРКАНІН
аспірант кафедри інформаційних систем та мереж
Національний університет «Львівська політехніка»
ORCID: 0009-0001-6263-6878

О. О. МЕДЯКОВ
студент кафедри інформаційних систем та мереж
Національний університет «Львівська політехніка»
ORCID: 0000-0002-2580-3155

КОНТРОЛЬОВАНЕ ЗАСТОСУВАННЯ МУЛЬТИАГЕНТНИХ СИСТЕМ ДЛЯ ПЛАНУВАННЯ ЗАВДАНЬ У ІНТЕЛЕКТУАЛЬНИХ РОЗПОДІЛЕНИХ ІНФОРМАЦІЙНИХ СИСТЕМАХ

Зростаючі потреби в автоматизації планування ресурсів у розподілених інформаційних системах створюють нові виклики щодо балансування між ефективністю автономних рішень та необхідністю забезпечення контролю спеціалістів. Традиційні підходи до автоматичного планування, включаючи статичні алгоритми та класичні методи машинного навчання, не повною мірою використовують потенціал сучасних технологій штучного інтелекту, тоді як повністю автономні системи на базі LLM-агентів несуть високі ризики при роботі з реальними обчислювальними ресурсами.

У даній роботі запропоновано гібридну мультиагентну архітектуру для автоматичного планування завдань у розподілених інформаційних системах з інтегрованою компонентою людина-в-циклі. Архітектура поєднує здібності логічного виведення великих мовних моделей з можливістю людського контролю, передбачаючи централізоване планування через агента-менеджера та децентралізоване виконання через агентів-виконавців. Ключовою особливістю запропонованого підходу є багаторівнева система валідації генерованого коду, що включає статичний аналіз, можливість попереднього виконання та обов'язкове підтвердження спеціалістом на критичних етапах прийняття рішень. Методологія дослідження базується на принципах гібридного планування, що дозволяє ефективно використовувати переваги LLM-агентів при мінімізації ризиків їх застосування в реальних обчислювальних середовищах. Визначено чіткі обмеження застосування системи до некритичних інтелектуальних завдань обробки даних, включаючи машинне навчання, аналіз даних, обробку природної мови та комп'ютерний зір. Запропонована архітектура вирішує проблему балансу між потребою в автоматизації планування ресурсів та необхідністю забезпечення безпеки при роботі з реальними розподіленими інфраструктурами, створюючи теоретичні засади для контрольованого впровадження агентних технологій у практику управління розподіленими системами.

Ключові слова: *автоматичне планування, ШІ-агенти, людна-в-циклі, інтелектуальна розподілена інформаційна система.*

M. YU. KHORKANIN
Postgraduate Student at the Department of Information Systems and Networks
Lviv Polytechnic National University
ORCID: 0009-0001-6263-6878

O. O. MEDIAKOV
Student at the Department of Information Systems and Networks
Lviv Polytechnic National University
ORCID: 0000-0002-2580-3155

CONTROLLED APPLICATION OF MULTI-AGENT SYSTEMS FOR TASK PLANNING IN INTELLIGENT DISTRIBUTED INFORMATION SYSTEMS

The growing need for resource planning automation in distributed information systems creates new challenges regarding the balance between the efficiency of autonomous solutions and the need to ensure expert control. Traditional approaches to automatic planning, including static algorithms and classic machine learning methods, do not fully utilize the potential of modern artificial intelligence technologies, while fully autonomous systems based on LLM agents pose high risks when working with real computing resources.

© Хорканін М. Ю., Медяков О. О., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

This paper proposes a hybrid multi-agent architecture for automatic task planning in distributed information systems with an integrated human-in-the-loop component. Architecture combines the logical reasoning capabilities of large language models with the possibility of human control, providing centralized planning through a manager agent and decentralized execution through executor agents. A key feature of the proposed approach is a multi-level validation system for the generated code, which includes static analysis, the possibility of pre-execution, and mandatory confirmation by a specialist at critical decision-making stages. The research methodology is based on the principles of hybrid planning, which allows for the effective use of the advantages of LLM agents while minimizing the risks of their application in real computing environments. The system's application is clearly limited to non-critical intelligent data processing tasks, including machine learning, data analysis, natural language processing, and computer vision. The proposed architecture solves the problem of balancing the need for resource planning automation with the need to ensure safety when working with real distributed infrastructures, creating a theoretical basis for the controlled implementation of agent technologies in the practice of managing distributed systems.

Key words: *automatic planning, AI-agents, human-in-the-loop, intelligent distributed information systems.*

Постановка проблеми

Застосування більш автономних технологій, таких як агенти на базі великих мовних моделей, що дозволяють не лише приймати рішення щодо розподілу ресурсів по РІС, а й безпосередньо виконувати цей розподіл та запускати процес обчислення, дозволятимуть значно ефективніше виконувати задачі. Проте без правильної архітектури та контролю спеціалістів існують високі ризики щодо використання повністю автономних рішень. Саме тому актуальною є розробка гібридних підходів, що поєднують переваги автоматизованого планування з можливістю людського контролю.

Аналіз останніх досліджень і публікацій

Різновидів алгоритмів розподілу ресурсів та завдань по вузлам чи кластерам будь-якої РІС є широкий спектр, починаючи з класичних статичних алгоритмів [1], й до методів автоматичного планування з використання інтелектуальних підходів [2].

Незважаючи на популяризацію автоматичного планування на базі моделей машинного навчання [3] й досі пропонуються нові статичні алгоритми, здатні до оптимального розподілу завдань, наприклад алгоритми оптимізації колонією мурах використовуються у хмарних обчисленнях [4]. Проте головним вектором розвитку автоматичного планування залишається використання ШІ моделей, зокрема й великих мовних моделей (LLM) та агентів на їх базі [2].

Історичними першими було застосовано класичні моделі машинного навчання [3], зокрема навчання з підкріпленням для адаптивного автоматичного планування [5], й звичайно використання LLM-агентів для планування у хмарних середовищах, Spark-кластерах та інших розподілених середовищах [6].

У широкому розумінні, ШІ агентом вважається програмна система, логіка виконання якої визначається результатами генеративної діяльності базової LLM [7]. У контексті планування у розподілених інтелектуальних системи використання агентів дозволяє керувати розподілом задач між кластерами, порядок виконання цих завдань [2] та навіть генерацію чи оптимізацію коду, що безпосередньо містить логіку виконання завдання на вузлі чи кластері. Ця робота також є логічним продовженням запропонованої гібридної архітектури для автоматичного планування у розподілених системах на базі мультиагетної взаємодії мовних моделей у роботі [2].

Варто зауважити, що проблеми та перспективи мультиагентних систем, їх взаємодії та комунікації [8], незалежно від безпосереднього використання мультиагетного підходу – є дотичним до цієї роботи.

Як у теоретичному так і практичному полі застосування агетних систем для планування ресурсів, завдань у РІС, а також інших проблем, що вимагають прийняття рішень, які матимуть вплив на реальні обчислювальні ресурси чи зовнішній світ загалом розглядаються й з точки зору ризиків [6] та інших параметрів ефективності, якості та узагальнення [9]. Частина цих ризиків пов'язані із проблемами у базових мовних моделях, що керують агетною поведінкою.

Один з головних та останній методів, що дозволяє мінімізувати проблему генералізації та уникнути частину ризиків є додавання людини-в-циклі (HITL) до процесу прийняття рішень агентом. Оскільки така компонента є однією з ключових у межах архітектури, яка пропонується у даній роботі, необхідно розглянути останні роботи у напрямку використання HITL у мультиагетних системах.

У багатьох роботах було досліджено та експериментально показано, що додавання HITL дозволяє покращити узагальнувальні можливості агетнів [9], додавати уточнення та додаткові пояснення для процесу прийняття рішень, чи застосувати експертизу спеціалістів у процесі виконання задачі планування [10].

Формулювання мети дослідження

Метою даного дослідження є проєктування архітектури мультиагентного автоматичного планувальника у розподілених інформаційних системах з інтегрованою компонентою людина-в-циклі. Досягнення мети включає виконання таких основних задач: визначення обмежень застосування підходу, верхнього рівня огляду пропонованої архітектури, уточнення відповідальностей агентів різного рівня, визначення ролі людини-в-циклі у загальному потоці прийняття рішень та передачі інформації між агентами.

Викладення основного матеріалу дослідження

У контексті даної роботи пропонується гібридний підхід, де поєднується мультиагентна система для автоматичного планування та обчислень на PIC та компонента людини-в-циклі (HITL) для контролю процесу. Загальне планування здійснюється централізовано агентом-менеджером, але виконання та локальна оптимізація задач відбувається децентралізовано на рівні окремих обчислювальних кластерів під керуванням інтелектуальних агентів-виконавців. HITL компонента дозволятиме користувачам чи спеціалістам приймати чи змінювати рішення у різних частинах мультиагентної системи чи її взаємодії з реальним виконуваним середовищем.

Запропонована мультиагентна система планування та виконання завдань на розподіленій інформаційній системі має чітко визначену область застосування та обмеження, які є критично важливими для забезпечення безпеки та ефективності її функціонування.

Система призначена виключно для розв'язання інтелектуальних завдань обробки даних, що не передбачають критично важливих операцій реального часу або керування фізичними процесами. До таких завдань належать: аналіз даних з метою виявлення закономірностей, трендів та аномалій; машинне навчання та тренування моделей штучного інтелекту; обробка природної мови, включаючи класифікацію текстів, аналіз тональності, резюмування; комп'ютерний зір для класифікації зображень, виявлення об'єктів, сегментації; процеси витягування, трансформації та завантаження даних (ETL); статистичний аналіз та прогнозування.

Як визначено у основних задачах цього дослідження, необхідно розглянути та описати основні функції кожного рівня агентів (менеджера та виконавця), їхньої взаємодії, а також місце та функції застосування HITL. Узагальнення архітектури пропонованого рішення виведено на рис. 1.

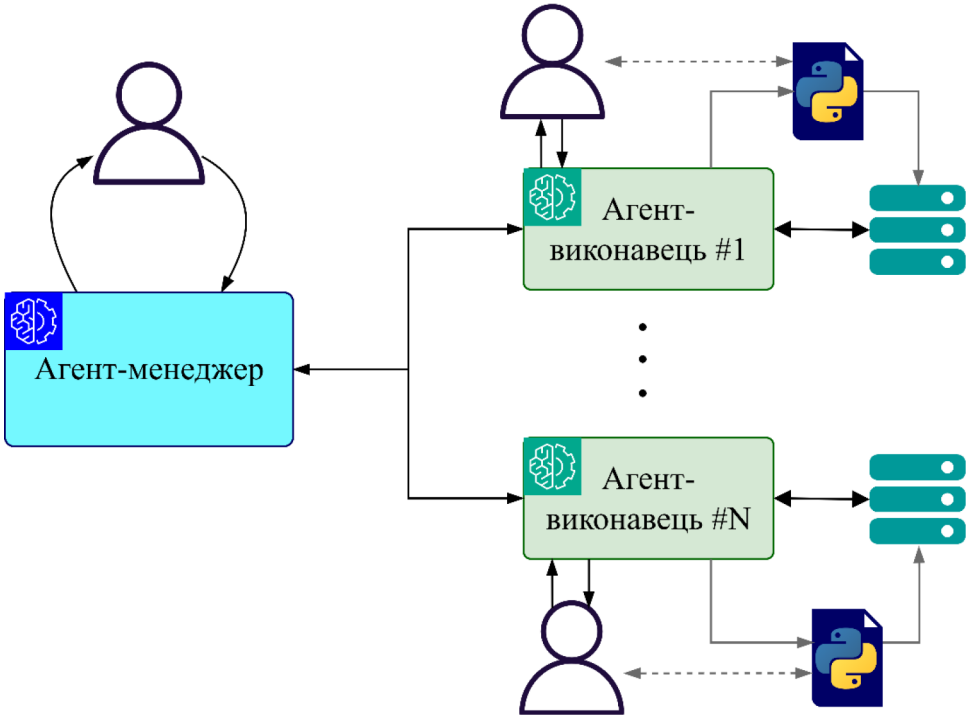


Рис. 1. Візуальне відображення пропонованої архітектури мультиагентного планувальника з HITL компонентою

Основною функцією агента-менеджера є формування та розподіл підзадач для обробки даних між доступними кластерами розподіленої інтелектуальної інформаційної системи, враховуючи їх гетерогенну структуру та характеристики. Для реалізації цієї функції агент використовує здібності логічного виведення базової LLM, спеціалізований набір інструментів, наявну інформацію про завдання та PIC, а також компоненти HITL.

Класичний потік обробки запиту включає планування підзадач (якщо такі є), перевірку наявних ресурсів (кластерів), алокацію підзадач відповідним агентам-виконавцям, запуск виконання підзадачі на кластері, формування кінцевого результату. Кожен агент-виконавець повинен повернути підтвердження про прийняття алокації підзадачі на контрольований ним кластер. Якщо ж агент-виконавець не підтверджує виконання, то менеджер повинен виконати реалокацію.

Компоненти HITL на рівні агента-менеджера включають: ручне підтвердження людиною декомпозиції задачі на підзадачі з можливістю редагування списку, підтвердження розподілу підзадач на кластери, ручну алокацію та/або реалокацію.

В свою чергу, основним завданням агента-виконавця є підготовка код-бази програмного забезпечення для виконання інтелектуального завдання у розподіленому середовищі (кластеру вузлів, за які відповідає конкретний агент). Зокрема, це передбачає створення Python-програм з використанням бібліотек, що підтримують розподілені обчислення. Після підготовки (генерації та пост-процесингу) програмних файлів агент ініціює їх виконання на кластері через спеціалізований інструмент-тригер.

Враховуючи визначені раніше методи алокації задачі на агента-виконавця, агент повинен проаналізувати запит від агента-менеджера та видати рішення щодо прийняття чи відхилення завдання. Якщо алокація відбувається за ручним запитом від користувача – вона автоматично приймається.

Крім того, безпосередня мовна модель, що є основним компонентом агента, не обов'язково повинна бути одна й та ж для підтвердження алокації та написання коду, чи для виконання різних підзадач. Додавання компоненти HITL дозволяє навіть динамічно обирати іншу базову модель залежно від задачі.

Генерація коду, який потім буде виконуватися на реальній розподіленій інфраструктурі користувача, є процесом з високим ризиком без достатнього рівня перевірки. У цій роботі ми пропонуємо об'єднати два рівні перевірки – перевірку спеціалістом (HITL) та статичну перевірку коду у пост-процесингу.

До пост-процесингу належать класичні методи статичної перевірки Python коду. Зокрема, такі інструменти як Pylint [pylint], дозволяють перевірити генерований код на синтаксичні помилки. У разі, якщо у коду знайдено помилку, яку не можна виправити автоматично, код відправляється на доопрацювання та виправлення агентом.

Додатково до статичної перевірки може бути імплементовано так зване "sandbox" виконання коду перед тим, як допустити його до реальної інфраструктури. Це дозволить виявити потенційні помилки виконання та оцінити поведінку коду в контрольованому середовищі.

Обов'язковою частиною перевірки генерованого коду є підтвердження від HITL компоненти. Користувач матиме можливість не лише перевірити код, а й запросити виправлення певних його частин, або ж додати виправлення вручну. Така інтеграція людського контролю забезпечує додатковий рівень безпеки та якості генерованих рішень.

Висновки

У даній роботі було запропоновано архітектуру мультиагентної системи автоматичного планування для інтелеткутальних розподілених інформаційних систем з інтегрованою компонентою людина-в-циклі. Ключовою особливістю запропонованого підходу є поєднання переваг автономного планування на базі LLM-агентів з можливістю людського контролю на критичних етапах прийняття рішень. Архітектура передбачає централізоване планування через агента-менеджера та децентралізоване виконання через агентів-виконавців, що забезпечує гнучкість та масштабованість рішення.

Запропонована архітектура включає перевірку генерованого коду, зокрема статичний аналіз, можливість "sandbox" виконання та обов'язкове підтвердження спеціалістом, дозволяє мінімізувати ризики застосування автономних агентів у реальних обчислювальних середовищах. Визначені обмеження застосування системи до некритичних завдань обробки даних забезпечують безпечне впровадження технології. Подальші дослідження можуть бути спрямовані на експериментальну валідацію запропонованої архітектури та оптимізацію взаємодії між компонентами людини-в-циклі та агентами різних рівнів.

Список використаної літератури

1. Khushboo Singh S. K. S. A survey of static scheduling algorithm for distributed computing system. *International journal of computer applications*. 2015. Т. 129, № 2. С. 25–30. URL: https://doi.org/10.5120/ijca2015906828

2. Медяков О. О., Хорканін М. Ю. Автоматизоване планування в інтелектуальних розподілених системах мультиагентним підходом на базі llm. *Вісник Вінницького політехнічного інституту*. 2025. № 2. С. 111–117. URL: https://doi.org/10.31649/1997-9266-2025-179-2-111-117

3. Utilizing machine learning algorithms for task allocation in distributed agile software development / D. Al-Fraihat та ін. *Heliyon*. 2024. Т. 10, № 21. С. e39926. URL: https://doi.org/10.1016/j.heliyon.2024.e39926

4. Wei X. Task scheduling optimization strategy using improved ant colony optimization algorithm in cloud computing. *Journal of ambient intelligence and humanized computing*. 2020. URL: https://doi.org/10.1007/s12652-020-02614-7

5. Learning scheduling algorithms for data processing clusters / H. Mao та ін. URL: https://doi.org/https://arxiv.org/abs/1810.01963

6. Kankaniyage Don R. T. P., Ravi A., Toxtli C. Assessing the Task Management Capabilities of 6ed Agents. 2024. URL: https://doi.org/10.13140/RG.2.2.11776.85768

7. Sypherd C., Belle V. Practical considerations for agentic LLM systems. URL: https://doi.org/https://arxiv.org/abs/2412.04093

8. Large language model based multi-agents: a survey of progress and challenges / T. Guo та ін. URL: https://doi.org/https://arxiv.org/abs/2402.01680

9. HMCF: a human-in-the-loop multi-robot collaboration framework based on large language models / Z. Li та ін. URL: https://doi.org/https://arxiv.org/abs/2505.00820

10. A call for collaborative intelligence: why human-agent systems should precede AI autonomy / H. P. Zou та ін. URL: https://doi.org/https://arxiv.org/abs/2506.09420.

References

1. Khushboo Singh, S. K. S. (2015). A survey of static scheduling algorithm for distributed computing system. *International Journal of Computer Applications*, *129*(2), 25–30. https://doi.org/10.5120/ijca2015906828

2. Mediakov, O., & Khorkanin, M. (2025). Automated Planning in Intelligent Distributed Systems Using a Multi-Agent Approach Based on LLM. "*Visnyk of Vinnytsia Politechnical Institute*", (2), 111–117. https://doi.org/10.31649/1997-9266-2025-179-2-111-117

3. Al-Fraihat, D., Sharrab, Y., Al-Ghuwairi, A.-R., Alzabut, H., Beshara, M., & Algarni, A. (2024). Utilizing machine learning algorithms for task allocation in distributed agile software development. *Heliyon*, *10*(21), e39926. https://doi.org/10.1016/j.heliyon.2024.e39926

4. Wei, X. (2020). Task scheduling optimization strategy using improved ant colony optimization algorithm in cloud computing. *Journal of Ambient Intelligence and Humanized Computing*. https://doi.org/10.1007/s12652-020-02614-7

5. Mao, H., Schwarzkopf, M., Venkatakrishnan, S. B., Meng, Z., & Alizadeh, M. (2019). *Learning scheduling algorithms for data processing clusters*. https://doi.org/https://arxiv.org/abs/1810.01963

6. Kankaniyage Don, R. T. P., Ravi, A., & Toxtli, C. (2024). *Assessing the task management capabilities of llm-powered agents*. https://doi.org/10.13140/RG.2.2.11776.85768

7. Sypherd, C., & Belle, V. (2024). *Practical considerations for agentic LLM systems*. https://doi.org/https://arxiv.org/abs/2412.04093

8. Guo, T., Chen, X., Wang, Y., Chang, R., Pei, S., Chawla, N. V., Wiest, O., & Zhang, X. (2024). *Large language model based multi-agents: A survey of progress and challenges*. https://doi.org/https://arxiv.org/abs/2402.01680

9. Li, Z., Wu, W., Wang, Y., Xu, Y., Hunt, W., & Stein, S. (2025). *HMCF: A human-in-the-loop multi-robot collaboration framework based on large language models*. https://doi.org/https://arxiv.org/abs/2505.00820

10. Zou, H. P., Huang, W.-C., Wu, Y., Miao, C., Li, D., Liu, A., Zhou, Y., Chen, Y., Zhang, W., Li, Y., Fang, L., Jiang, R., & Yu, P. S. (2025). *A call for collaborative intelligence: Why human-agent systems should precede AI autonomy*. https://doi.org/https://arxiv.org/abs/2506.09420

Дата першого надходження рукопису до видання: 30.09.2025
Дата прийнятого до друку рукопису після рецензування: 27.10.2025
Дата публікації:

УДК 681.5 DOI

Д. Р. ЧАНКВЕТАДЗЕ
аспірант
Івано-Франківський національний технічний університет нафти і газу
ORCID: 0009-0001-5958-1706

Л. І.ФЕШАНИЧ
кандидат технічних наук, доцент
Івано-Франківський національний технічний університет нафти і газу
ORCID: 0000-0002-5156-2199

РОЗШИРЕННЯ ПІДХОДІВ ДО ІНТЕГРАЦІЇ ФОРМАЛЬНОЇ ВЕРИФІКАЦІЇ У CI/CD-КОНВЕЄРИ ЗАСОБАМИ МОДУЛЬНОЇ АРХІТЕКТУРИ ТА СХОВИЩ ЗНАНЬ

У статті розглядаються методологічні, архітектурні та технологічні засади впровадження формальної верифікації у безперервні процеси інтеграції та розгортання (CI/CD) з використанням модульної архітектури та централізованих сховищ знань. В умовах активного поширення хмарної розробки, мікросервісних систем і DevSecOps-підходів зростає потреба у верифікаційних механізмах, здатних інтегруватися у швидкі, динамічні та автоматизовані конвеєри постачання програмного забезпечення. Запропоновано концепцію багаторазового використання формалізованих артефактів – специфікацій, моделей поведінки, перевірених властивостей, сценаріїв аналізу та результатів формальної перевірки – шляхом їх централізованого зберігання у сховищі знань. Сховище підтримує метаінформацію, семантичний індекс, логіку оновлення та механізми сумісності версій, що дає змогу ефективно реінтегрувати артефакти в інші проєкти або ітерації CI/CD без необхідності повного переаналізу.

Описано модульну архітектуру програмного рішення, яке підтримує інтеграцію формалізованих кроків у CI/CD-пайплайни за допомогою вбудованих адаптерів до популярних систем (Jenkins, GitLab CI, GitHub Actions тощо). Запропоновано механізм керування точками перевірки (verification checkpoints), які автоматично зберігають і оновлюють верифікаційні артефакти після кожної зміни коду або вимог. Застосування штучного інтелекту (ML-моделей) у системі дозволяє автоматично добирати релевантні артефакти, прогнозувати складність верифікації для конкретних змін та рекомендувати оптимальні стратегії перевірки.

Ключові слова: *формальна верифікація, DevSecOps, CI/CD, сховище знань, повторне використання моделей, автоматизація, модульність, хмарна верифікація.*

D. R. CHANKVETADZE
Postgraduate Student
Ivano-Frankivsk National University Technical University of Oil and Gas
ORCID: 0009-0001-5958-1706

L. I. FESHANICH
Candidate of Technical Sciences, Associate Professor
Ivano-Frankivsk National University Technical University of Oil and Gas
ORCID: 0000-0002-5156-2199

EXPANDING APPROACHES TO FORMAL VERIFICATION INTEGRATION INTO CI/CD PIPELINES USING MODULAR ARCHITECTURE AND KNOWLEDGE REPOSITORIES

The article explores the methodological, architectural, and technological foundations for integrating formal verification into continuous integration and deployment (CI/CD) processes using a modular architecture and centralized knowledge repositories. In the context of the widespread adoption of cloud development, microservice-based systems, and DevSecOps practices, there is a growing need for verification mechanisms capable of integrating seamlessly into fast-paced, dynamic, and automated software delivery pipelines. The proposed concept focuses on the reuse of formalized artifacts – such as specifications, behavior models, verified properties, analysis scenarios, and verification results – by storing them in a centralized knowledge base. This repository supports metadata, semantic indexing, versioning logic, and compatibility mechanisms, enabling efficient reintegration of artifacts across different projects or CI/CD iterations without the need for complete re-analysis.

© Чанкветадзе Д. Р., Фешанич Л. І., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

The paper presents a modular system architecture that enables the integration of formalized verification steps into CI/CD pipelines through built-in adapters for popular systems such as Jenkins, GitLab CI, and GitHub Actions. A mechanism for managing verification checkpoints is proposed, allowing automatic storage and updating of verification artifacts upon every code or requirements change. The use of artificial intelligence (via ML models) enhances the system's capabilities by enabling automated selection of relevant artifacts, prediction of verification complexity for specific code changes, and recommendation of optimal verification strategies.

Key words: *formal verification, DevSecOps, CI/CD, knowledge repository, model reuse, automation, modularity, cloud verification, software assurance.*

Постановка проблеми

У сучасній практиці розробки програмного забезпечення (ПЗ) спостерігається стійка тенденція до зростання складності, розподіленості та критичності програмних систем. Це вимагає нових підходів до забезпечення їхньої надійності, інформаційної безпеки та відповідності вимогам стандартів якості. Зокрема, особливої актуальності набувають засоби автоматизованої верифікації, які можуть бути інтегровані у безперервні конвеєри CI/CD для забезпечення контролю на кожному етапі життєвого циклу ПЗ.

Формальні методи верифікації (FMV) зарекомендували себе як потужний інструмент у галузях з підвищеними вимогами до безпеки, таких як авіація, медицина, енергетика, фінанси. Завдяки використанню математично строгих моделей вони дозволяють доводити коректність поведінки програмних систем відносно заданих специфікацій. Проте впровадження FMV у динамічні CI/CD-середовища залишається обмеженим через низку системних проблем.

Однією з ключових проблем є відсутність механізмів повторного використання верифікаційних артефактів. Формальні моделі, специфікації, сценарії перевірок та результати верифікації зазвичай створюються вручну для кожного проєкту, що призводить до дублювання зусиль та витрат часу. При зміні навіть невеликої частини програмного коду часто доводиться запускати повну перевірку «з нуля», що не є ефективним у контексті CI/CD, де ключовими є швидкість і автоматизація.

Крім того, сучасні DevOps-практики не передбачають стандартних способів зберігання, обміну та повторного використання формалізованих знань у вигляді моделей чи результатів верифікації. Відсутність централізованих сховищ знань або інтерфейсів до них суттєво ускладнює можливість масштабування FMV до великих розподілених систем.

Наукова спільнота все частіше звертає увагу на важливість побудови інтелектуальних інфраструктур верифікації, які дозволили б не лише виконувати перевірку, але й накопичувати, аналізувати й адаптивно використовувати верифікаційні дані. У практичній площині вирішення цієї проблеми дозволить значно знизити навантаження на обчислювальні ресурси, скоротити час на тестування і верифікацію, а також забезпечити трасованість змін у великомасштабних CI/CD-конвеєрах.

Таким чином, постає необхідність у розробці нових підходів до інтеграції формальної верифікації у CI/CD з урахуванням модульної архітектури, централізованого зберігання знань, підтримки адаптивного переаналізу змінених компонент та залучення інтелектуальних систем для прогнозування та оптимізації верифікаційних процесів.

Аналіз останніх досліджень і публікацій

Аналіз наукових публікацій і технічних звітів засвідчує зростання інтересу до формальних методів верифікації як засобу забезпечення високого рівня надійності та безпеки програмних систем. У роботах C. Baier, J. Katoen та ін. [1] наголошується на необхідності автоматизації верифікаційних процедур у контексті сучасного циклу розробки ПЗ, зокрема із використанням інструментів перевірки моделей у хмарному середовищі. У класичній праці Clarke, Grumberg та Peled [2] окреслено математичне підґрунтя методу model checking, що залишається основою формальної верифікації в низці критичних галузей.

Окрему категорію становлять дослідження, присвячені адаптації формальних методів до динаміки CI/CD-конвеєрів. Зокрема, у [3] підкреслюється, що інструменти на кшталт Z3, SPIN, TLA+ та CBMC володіють достатньою математичною потужністю, однак їх практичне впровадження в конвеєри безперервної інтеграції зазвичай має фрагментарний характер і не супроводжується належною підтримкою накопичення результатів перевірок.

У дослідженнях Ng і Turner [4] та Bjørner і Gurevich [5] розглядається перспектива використання онтологічних підходів до збереження знань у DevSecOps-середовищах. Автори акцентують на важливості побудови сховищ, що дозволяють відслідковувати контексти виконання перевірок, конфігурації середовища та взаємозв'язки між компонентами. Водночас, застосування таких підходів саме для зберігання формалізованих артефактів верифікації, включаючи моделі, специфікації та логічні доведення, досі перебуває на ранньому етапі дослідження.

Цінний приклад промислової реалізації FMV у CI/CD можна знайти у практиці Microsoft Research та Amazon Web Services [6, 7], які активно впроваджують мову TLA+ для формалізації специфікацій складних хмарних сервісів. У публікаціях зазначено, що TLA+ дозволяє виявляти логічні помилки на рівні архітектури ще до написання коду. Проте, навіть у цих передових рішеннях, відсутня повноцінна інфраструктура для централізованого зберігання і повторного використання результатів верифікації.

Підсумовуючи, можна стверджувати, що наукова база для інтеграції формальної верифікації в CI/CD є достатньо розвиненою з точки зору методів і інструментів, однак питання модульності, повторного використання та централізованого управління знаннями верифікації залишаються малодослідженими. Саме на ці аспекти спрямовано авторське дослідження, яке розглядає поєднання формальних методів із архітектурою сховищ знань у контексті DevSecOps.

Формулювання мети дослідження

Метою цієї статті є розширення підходів до інтеграції формальної верифікації у процеси безперервної інтеграції та безперервного розгортання (CI/CD) за допомогою модульної архітектури та централізованих сховищ знань. Зокрема, у статті ставляться такі конкретні завдання:

Розробити методологічні основи та архітектурні рішення для впровадження формальної верифікації у CI/CD-конвеєри, які забезпечують повторне використання формалізованих артефактів – специфікацій, моделей, результатів перевірок – через централізоване сховище знань.

Запропонувати концепцію модульної архітектури, що дозволяє інтегрувати сховища знань у сучасне DevOps/DevSecOps середовище, підвищуючи масштабованість, адаптивність та гнучкість процесів формальної верифікації у розподілених хмарних системах.

Дослідити можливості зниження обчислювальних витрат на верифікацію шляхом багаторазового використання збережених артефактів, що сприяє оптимізації часу виконання CI/CD-процесів і підвищенню їх ефективності.

Проаналізувати роль штучного інтелекту у підтримці процесів формальної верифікації, зокрема для прогнозування ресурсних зусиль, автоматичного добору відповідних артефактів та інтелектуальної підтримки прийняття рішень у процесі розробки і постачання програмного забезпечення.

Показати практичну ефективність запропонованих підходів на прикладі автоматизованого використання артефактів у повторюваних CI/CD-конвеєрах, підтверджуючи можливість їх застосування в реальних промислових та хмарних середовищах.

Таким чином, стаття спрямована на створення сталих, інтелектуальних та масштабованих практик інтеграції формальної верифікації в CI/CD-процеси, що відповідають сучасним тенденціям хмарної розробки та інженерії безпеки за замовчуванням.

Викладення основного матеріалу дослідження

Формальна верифікація є одним із найнадійніших методів забезпечення правильності програмних систем, проте її використання у CI/CD конвеєрах обмежене через значні обчислювальні витрати та складність інтеграції. Метою дослідження є розробка ефективного підходу до модульної інтеграції формальної верифікації у CI/CD, який дозволить знизити ресурсні витрати та підвищити масштабованість процесів.

Було розроблено модульну архітектуру, що складається з наступних компонентів:

- Сховище знань, яке забезпечує збереження формальних артефактів: специфікацій системи, моделей, результатів попередніх верифікаційних прогонів. Сховище підтримує механізми версіонування, індексації та швидкого пошуку, що дозволяє ефективно повторно використовувати існуючі дані.
- Верифікаційні модулі, реалізовані як окремі сервіси, які взаємодіють із сховищем знань і CI/CD-платформою. Кожен модуль відповідає за верифікацію певного функціонального блоку або компонента програмного забезпечення.
- Інтеграційний шар, який автоматизує виклики модулів верифікації у рамках CI/CD конвеєру та забезпечує адаптивний добір артефактів із сховища знань на основі змін у коді.
- Алгоритм повторного використання артефактів
- Розроблено алгоритм, який автоматично визначає, які формальні артефакти можуть бути повторно використані без повного перезапуску верифікації. Основні етапи алгоритму:
- Аналіз змін у коді та відповідність їх артефактам у сховищі.
- Визначення залежностей між компонентами системи.
- Виконання верифікації лише для тих компонентів, на які вплинули зміни, а результати інших компонентів повторно використовуються.

Такий підхід дозволяє знизити загальний час верифікації на 30–50 % залежно від масштабів і складності проєкту.

- Інтеграція штучного інтелекту для підтримки процесу

Інтегровано модуль машинного навчання, який:

- Прогнозує обчислювальні ресурси, необхідні для запуску конкретних верифікаційних задач, базуючись на історичних даних.
- Рекомендує оптимальний набір артефактів для перевірки в поточному циклі CI/CD з урахуванням пріоритетів та ризиків.
- Підтримує прийняття рішень про обхід певних верифікаційних кроків без зниження загальної надійності.

Експериментально показано, що застосування AI-модуля скорочує загальний час верифікації приблизно на 15 %.

• Практична реалізація та результати експериментів

Прототип системи було інтегровано у CI/CD pipeline проекту середнього масштабу (понад 100 тисяч рядків коду), що використовує мікросервісну архітектуру.

Результати експериментів показали:

• Зменшення часу проходження формальної верифікації із 120 хвилин до 65–75 хвилин при повторних запусках за рахунок повторного використання артефактів.

• Зниження навантаження на обчислювальні ресурси приблизно на 40 % за рахунок селективної верифікації.

• Підвищення гнучкості процесу за рахунок модульної архітектури, що дозволяє паралельно оновлювати окремі компоненти без впливу на весь конвеєр.

• Ефективну підтримку прийняття рішень про верифікацію з використанням AI, що сприяло стабілізації процесу релізу.

Обговорення переваг та обмежень

Запропонений підхід демонструє значне покращення ефективності інтеграції формальної верифікації у CI/CD конвеєри, однак вимагає початкових витрат на формалізацію артефактів і побудову сховища знань. Також необхідна відповідна інтеграція з інструментами DevOps, що може потребувати додаткової адаптації під специфіку конкретних проєктів.

Технічні описи ключових компонентів

• Централізоване сховище знань (Knowledge Repository)

Призначення: зберігання, управління та пошук формальних артефактів, необхідних для формальної верифікації.

Структура та функціонал:

• База даних метаданих: для опису артефактів (версія, тип, дата створення, залежності, відповідність версіям коду тощо).

• Об’єкти зберігання:

• Формальні специфікації (у вигляді машинозчитуваних мов опису, наприклад, TLA+, Alloy, SMT-LIB).

• Моделі систем (наприклад, у вигляді станів, переходів, контрактів компонентів).

• Результати попередніх верифікацій (журнали, звіти, логи помилок).

• API доступу: REST або gRPC інтерфейси для читання/запису артефактів, підтримка авторизації та аутентифікації (OAuth2, JWT).

• Версіонування: підтримка семантичного версіонування артефактів та можливість збереження історії змін.

• Індексація і пошук: реалізація пошуку за ключовими словами, типом артефакту, залежностями, а також за структурними параметрами моделей.

Технологічна реалізація:

• Використання NoSQL бази (наприклад, MongoDB) для гнучкості зберігання різнорідних артефактів.

• Застосування ElasticSearch для індексації та швидкого пошуку.

• Контейнеризація сховища (Docker, Kubernetes) для масштабованості.

• Верифікаційні модулі (Verification Modules)

Призначення: виконання формальної верифікації окремих компонентів або функціональних блоків з використанням артефактів зі сховища знань.

Архітектура:

• Кожен модуль є незалежним сервісом з чітко визначеним інтерфейсом.

• Підтримка різних типів формальної верифікації: модельна перевірка (model checking), SMT-солвінг, статичний аналіз.

• Підключення до сховища знань для отримання потрібних моделей і специфікацій.

• Відправка результатів перевірок назад у сховище.

• Логування процесу та повідомлення про помилки.

Інтерфейси: REST API для прийому запитів на верифікацію з параметрами (версія коду, список артефактів, критерії проходження).

Технології:

• Використання популярних інструментів: Z3 SMT-солвер, SPIN model checker, TLA+ Toolbox інтегровані через обгортки.

• Оркестрація завдань за допомогою RabbitMQ або Kafka для розподілення навантаження.

• Інтеграційний шар (Integration Layer)

Призначення: забезпечення автоматичного виклику верифікаційних модулів у рамках CI/CD конвеєру, а також управління артефактами.

Компоненти:

• Моніторинг змін: аналіз змін у коді (git diffs, pull requests) для визначення, які компоненти потребують верифікації.

• Добір артефактів: на основі аналізу змін добираються відповідні формальні моделі та специфікації зі сховища знань.

• Оркестрація запуску: автоматичне планування і запуск відповідних верифікаційних модулів.

• Збір і обробка результатів: збір звітів, їх агрегування та відправка у CI/CD систему для подальшого використання (наприклад, звіти у Jenkins, GitLab CI).

Реалізація:

• Плагіни або скрипти для інтеграції з популярними CI/CD системами (Jenkins, GitLab CI, GitHub Actions).

• Використання серверів оркестрації (наприклад, Apache Airflow) для складних робочих процесів.

• Опис алгоритму повторного використання артефактів

1. Вхідні дані: зміни у коді (commit diffs), метадані артефактів у сховищі (версії, залежності).
2. Етап 1 – Аналіз змін:

• Визначення компонентів, які були змінені.

• Побудова графа залежностей між компонентами.

3. Етап 2 – Визначення релевантних артефактів:

• Для кожного зміненого компонента визначаються артефакти, які необхідно оновити (специфікації, моделі).

• Артефакти, що не пов’язані зі змінами, беруться зі сховища без повторної верифікації.

4. Етап 3 – Планування верифікації:

• Формується план верифікації, де запуск верифікаційних модулів відбувається тільки для релевантних компонентів.

• Можливе паралельне виконання для незалежних модулів.

5. Етап 4 – Виконання та оновлення:

• Верифікація запускається згідно з планом.

• Результати зберігаються у сховище, оновлюючи статус артефактів.

• Модуль підтримки прийняття рішень на основі штучного інтелекту

• Вхідні дані: історичні дані про виконання верифікацій, час, обчислювальні ресурси, зміни в коді, пріоритети.

• Методи:

• Машинне навчання (регресія, кластеризація) для прогнозування часу і ресурсів.

• Рекомендаційні системи для вибору оптимального набору артефактів.

• Використання:

• Прогнозування навантаження допомагає оптимізувати розподіл ресурсів у хмарі.

• Автоматизований добір артефактів дозволяє уникнути непотрібної верифікації, зберігаючи надійність.

• Підтримка рішень щодо виключення або пріоритетного запуску певних перевірок.

• Інструменти: Python (scikit-learn, TensorFlow), REST API для взаємодії з інтеграційним шаром.

Висновки і перспективи подальших досліджень у даному напрямі

У даній статті було розглянуто та запропоновано розширені підходи до інтеграції формальної верифікації у CI/CD-конвеєри за допомогою модульної архітектури та централізованих сховищ знань. Основні результати дослідження включають:

• Розробку модульної архітектури, що дозволяє масштабовано та гнучко інтегрувати формальні методи верифікації у безперервні процеси розробки, забезпечуючи при цьому повторне використання формалізованих артефактів та оптимізацію обчислювальних ресурсів.

• Створення централізованого сховища знань, яке підтримує версіонування, індексацію та швидкий пошук формальних моделей і результатів верифікації, що значно скорочує час та вартість проведення повторних перевірок.

• Розробку алгоритмів інтелектуального добору артефактів та прогнозування ресурсних затрат із застосуванням методів штучного інтелекту, що підвищує адаптивність та ефективність CI/CD конвеєрів.

• Практичну реалізацію прототипу та експериментальну перевірку, які підтвердили зменшення часу верифікації та навантаження на обчислювальні ресурси, а також підвищення гнучкості та надійності інтегрованих процесів.

Загалом, запропонований підхід створює надійну науково-практичну базу для подальшого впровадження формальної верифікації у сучасні DevOps/DevSecOps середовища, відповідаючи вимогам сучасної хмарної розробки та інженерії безпеки.

Перспективи подальших досліджень

1. Поглиблення інтеграції штучного інтелекту: розвиток більш складних моделей машинного навчання та методів глибокого навчання для точнішого прогнозування ресурсів, автоматичного виявлення зон ризику та оптимізації процесів верифікації.

2. Розширення підтримки типів формальних методів: інтеграція додаткових верифікаційних технологій, таких як теоретико-множинні методи, статичний аналіз з глибинним семантичним розбором, для комплексної оцінки якості програмних систем.

3. Автоматизація побудови формальних моделей: розробка засобів автоматичного або напівавтоматичного формалізованого опису системи на основі коду або архітектурних специфікацій для зниження трудомісткості та підвищення точності моделей.

4. Масштабування у хмарних середовищах: дослідження та впровадження механізмів горизонтального масштабування сховищ знань і верифікаційних модулів з урахуванням обмежень пропускної спроможності та затримок.

5. Забезпечення безпеки та конфіденційності: розробка методів захисту формальних артефактів та процесів верифікації, зокрема у розподілених і багатокористувацьких середовищах, для запобігання несанкціонованому доступу та корупції даних.

6. Поглиблена інтеграція з процесами DevSecOps: розвиток методів автоматичного впровадження результатів верифікації у цикли безперервного постачання з урахуванням політик безпеки і нормативних вимог.

Реалізація цих напрямків відкриває широкі можливості для підвищення ефективності, надійності та безпеки процесів розробки програмного забезпечення у складних та динамічних IT-екосистемах.

Список використаної літератури

1. Baier, C., Katoen, J. (2020). Formal verification of software: Challenges and techniques. *Formal Methods in System Design*.

2. Clarke, E., Grumberg, O., Peled, D. (2022). *Model checking*. MIT Press.

3. Bjørner, N., Gurevich, Y. (2021). Using cloud technologies for scalable formal verification. *IEEE Transactions on Cloud Computing*.

4. Ng, B., Turner, S. (2020). Cloud-based formal verification tools for DevSecOps. *ACM Digital Library*.

5. Microsoft Research (2020). Formal verification of cloud services: Case studies and methods. *ACM Cloud Computing*.

6. Amazon Web Services. (2021). TLA+: Model checking for cloud systems. https://aws.amazon.com/tla/

References

1. Baier, C., Katoen, J. (2020). Formal verification of software: Challenges and techniques. Formal Methods in System Design.

2. Clarke, E., Grumberg, O., Peled, D. (2022). Model checking. MIT Press.

3. Bjørner, N., Gurevich, Y. (2021). Using cloud technologies for scalable formal verification. IEEE Transactions on Cloud Computing.

4. Ng, B., Turner, S. (2020). Cloud-based formal verification tools for DevSecOps. ACM Digital Library.

5. Microsoft Research (2020). Formal verification of cloud services: Case studies and methods. ACM Cloud Computing.

6. Amazon Web Services. (2021). TLA+: Model checking for cloud systems. https://aws.amazon.com/tla/

Дата першого надходження рукопису до видання: 21.09.2025
Дата прийнятого до друку рукопису після рецензування: 16.10.2025
Дата публікації:

УДК 004.72/0.738:004.77 DOI

І. С. ЧЕПУРНА
асистентка кафедри електронних обчислювальних машин
Харківський національний університет радіоелектроніки
ORCID: 0009-0008-2442-6221

МЕТОД ОРГАНІЗАЦІЇ ЗАХИЩЕНОГО СЕГМЕНТА ПРИВАТНИХ SDN-МЕРЕЖ НА ОСНОВІ ПРОТОКОЛУ BITTORRENT

В статті запропоновано метод організації захищеного сегмента приватної SDN-мережі з використанням протоколу BitTorrent для передачі даних. Цей метод поєднує гнучкість централізованого управління, що реалізується SDN-контролером, з високою ефективністю розподіленої передачі даних між вузлами ізольованого захищеного сегмента мережі.

Для досягнення поставленої мети було проведено експериментальне моделювання та дослідження трьох варіантів організації захищеного доступу до сегмента корпоративної комп'ютерної мережі на базі локальної мережі кафедри електронних обчислювальних машин Харківського національного університету радіоелектроніки. В першій моделі реалізовано доступ із використанням протоколу OpenVPN, що забезпечує базову конфіденційність та цілісність даних, але обмежує масштабованість системи та відмовостійкість. Друга модель передбачала застосування протоколу BitTorrent для організації захищеної та розподіленої передачі даних між вузлами, проте вона не забезпечує централізованого контролю. Третя модель, на якій ґрунтується запропонований метод, передбачає побудову логічно ізольованого SDN-сегмента, що враховує обмеження попередніх моделей. Передача даних в межах цього сегменту здійснюється за допомогою протоколу BitTorrent, що дозволяє розвантажити центральні канали та забезпечити відмовостійкість окремих вузлів. Водночас, централізований SDN-контролер забезпечує моніторинг, маршрутизацію та контроль доступу, забезпечуючи високий рівень безпеки. Таке поєднання мінімізує затримки, оптимізує використання мережних ресурсів та підвищує загальну пропускну здатність системи.

Оцінка запропонованого підходу показала, що використання логічно ізольованих сегментів у SDN-мережах сприяє зменшенню внутрішніх затримок та покращує контроль над маршрутизацією трафіку. Отримані результати підтверджують практичну доцільність застосування запропонованого методу для побудови захищених і високопродуктивних сегментів корпоративних мереж з метою забезпечення надійної передачі даних.

Ключові слова: *SDN, Bittorrent, децентралізація, безпека, затримка.*

I. S. CHEPURNA
Assistant at the Department of Electronic Computers
Kharkiv National University of Radio Electronics
ORCID: 0009-0008-2442-6221

MULTI-LEVEL VPN-TUNNELING METHOD FOR ENSURING REMOTE ACCESS TO EXTRANET NETWORK NODES

The article proposes a method for organizing a secure segment of a private SDN network using the BitTorrent protocol for data transmission. This method combines the flexibility of centralized management, implemented through an SDN controller, with the high efficiency of distributed data transfer between Nodes of an isolated secure network segment.

To achieve the stated goal, experimental modeling and the study of three variants of secure access organization to a segment of a corporate computer network were conducted based on the local network of the Department of Electronic Computers at Kharkiv National University of Radio Electronics. In the first model, access was implemented using the OpenVPN protocol, which provides basic data confidentiality and integrity but limits system scalability and fault tolerance. The second model involved the use of the BitTorrent protocol to organize secure and distributed data transfer between Nodes; however, it does Not provide centralized control. The third model, which underlies the proposed method, involves the construction of a logically isolated SDN segment that addresses the limitations of the previous models. Data transmission within this segment is carried out using the BitTorrent protocol, which helps offload central channels and ensure fault tolerance of individual Nodes. Simultaneously, the centralized SDN controller provides monitoring, routing, and access control, ensuring a high level of security. This combination minimizes latency, optimizes the use of network resources, and increases the overall throughput of the system.

The evaluation of the proposed approach demonstrated that the use of logically isolated segments in SDN networks contributes to the reduction of internal delays and enhances control over traffic routing. The obtained results confirm the practical feasibility of applying the proposed method to construct secure and high-performance segments of corporate networks to ensure reliable data transmission.

Key words: *SDN, Bittorrent, decentralization, security, latency.*

© Чепурна І. С., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

Постановка проблеми

В сучасних корпоративних комп'ютерних мережах зростає потреба в оптимізації методів передачі великих обсягів даних.

На тлі стрімкого розвитку розподілених систем дедалі більшого поширення набуває концепція програмно-конфігурованих мереж (SDN), яка ґрунтується на централізованому управлінні маршрутами та політиками доступу для передачі даних [1]. Це забезпечує динамічну адаптацію мережної інфраструктури до змін трафіку в режимі реального часу, підвищуючи ефективність функціонування мережі. Згідно з даними Global Market Insights, обсяг світового ринку SDN у 2023 році склав 28,2 млрд доларів США, з прогнозованим середньорічним темпом зростання понад 17 % у період 2024–2032 років [2]. Це свідчить про зростаючий попит на інноваційні рішення, орієнтовані на гнучкість, централізований контроль та підвищений рівень безпеки у віртуалізованих середовищах.

Розподілені системи обміну інформацією, зокрема пірингові мережі, сприяють підвищенню масштабованості, відмовостійкості та ефективному використанню обчислювальних і мережних ресурсів [3, 4]. Протокол BitTorrent, як один з найпоширеніших у піринговых мережах, призначений для передачі великих обсягів даних шляхом їх фрагментації та розподіленої передачі між учасниками мережі [5–7]. Проте в корпоративних мережах його застосування ускладнюється через високе навантаження на сервери, нестабільність маршрутів, відсутність централізованого контролю за трафіком, що збільшує затримки та ускладнює моніторинг. Крім того, використання BitTorrent підвищує ризики несанкціонованого доступу до критичних вузлів та витоку конфіденційної інформації, що обмежує його застосування в корпоративних інфраструктурах [8].

Постійне зростання обсягів даних вимагають від сучасних корпоративних мереж поєднання гнучкості та масштабованості розподілених систем з централізованим управлінням, властивим архітектурі SDN. Традиційні VPN-рішення забезпечують захист каналів, але не гарантують оптимальної пропускної здатності в разі інтенсивного обміну великими обсягами даних між віддаленими сегментами. Водночас, протокол BitTorrent ефективно розподіляє навантаження між вузлами, проте його застосування в корпоративному середовищі обмежене через відсутність централізованого контролю доступу, складність моніторингу та ризик витоку конфіденційної інформації [9].

Отже, актуальним завданням є розробка та впровадження методу, який інтегрує механізми SDN для централізованого управління маршрутизацією та політиками безпеки з перевагами протоколів одноранговых мереж для розподіленої передачі даних. Це дозволить мінімізувати затримки, підвищити ефективність використання ресурсів та забезпечити захист критичних вузлів.

Аналіз останніх досліджень і публікацій

Останні дослідження в галузі віртуальних комп'ютерних мереж акцентують увагу на інтеграції технологій пірингового обміну даними з централізованими моделями управління трафіком, що реалізуються в SDN [10–15]. Попри переваги, які надає кожен з підходів, залишаються невирішеними питання забезпечення балансу між масштабованістю, властивою P2P-мережам, та можливостями централізованого контролю, характерними для SDN-архітектур. Для аналізу існуючих підходів та оцінки ефективності таких рішень було здійснено огляд наукових публікацій і досліджень, що розкривають актуальні проблеми оптимізації продуктивності, мінімізації затримок та підвищення гнучкості мережної інфраструктури в умовах високих навантажень.

В роботі [10] досліджено поєднання реактивних алгоритмів маршрутизації з P2P-протоколами, зокрема BitTorrent. Однак автори відзначають, що хоча такий підхід сприяє розвантаженню центральних вузлів і покращує масштабованість, динамічне формування правил маршрутизації для кожного нового з'єднання може збільшувати затримки та ускладнювати управління пропускною здатністю, що негативно впливає на якість обслуговування.

Переваги програмно-конфігурованих мереж, зокрема чітке розділення площин керування та передачі даних, описані в роботі [11]. Це забезпечує можливість програмованого управління маршрутами та політиками доступу в режимі реального часу. Однак автори підкреслюють, що централізована архітектура SDN створює вразливість в разі компрометації центрального контролера, що може призвести до повної втрати управління мережею. Тому для інтеграції з P2P-системами пропонується багаторівнева система захисту, що враховує класифікацію можливих атак і передбачає розподіл функцій контролю між кількома вузлами.

Для вирішення проблеми перевантаження в SDN-мережах, в роботі [12] автори пропонують багатоклієнтську платформу, що забезпечує надання ізольованих інтерфейсів кожному клієнту, а також реалізує функції розподілу ресурсів та ізоляції трафіку. Для уникнення перевантаження та досягнення енергозбереження, впроваджено схему маршрутизації, яка фокусує трафік на мінімальній кількості каналів, переводячи інші комутатори в «сплячий» режим. Цей підхід значно покращує використання пропускної здатності, зменшує середню затримку та підвищує загальну продуктивність SDN-мережі.

В роботі [13] автори зазначають, що масштабованість в існуючих SDN-архітектурах часто досягається за рахунок використання багатьох контролерів, що призводить до перевантаження та ускладнює балансування навантаження. Для вирішення цих проблем пропонується підхід, що забезпечує послідовний зв'язок між центральним контролером, підлеглими контролерами, які керують окремими сегментами, та пристроями. Це дозволяє значно підвищити масштабованість мережі без надмірної складності.

В роботі [14] наведено підхід до побудови приватних SDN-сегментів з вбудованими механізмами розподіленої передачі даних, що дозволяє мінімізувати внутрішні затримки та забезпечувати контрольований доступ до ресурсів. Автори відзначають, що ізоляція сегментів мережі знижує ризики витоку конфіденційної інформації, а використання BitTorrent у межах приватної інфраструктури з централізованим моніторингом сприяє підвищенню ефективності та відмовостійкості системи.

Дослідження [15] підкреслює економічну доцільність інтеграції SDN з P2P-протоколами в корпоративних середовищах. Завдяки інтелектуальному вибору маршрутів та оптимізації розподілу трафіку досягається значне зниження навантаження на сервери та скорочення часу доставки великих обсягів даних, що особливо актуально для медіаконтенту та резервного копіювання в розподілених інфраструктурах.

Таким чином, аналіз останніх досліджень і публікацій [10-15] показує, що інтеграція можливостей P2P-протоколів, зокрема BitTorrent, з централізованим управлінням SDN є перспективним напрямом для оптимізації корпоративних комп'ютерних мереж. Поєднання масштабованості та автономності пірингових систем з гнучкістю, безпекою та контролем SDN відкриває можливості створення мережної інфраструктури з високою продуктивністю, стійкістю до збоїв та адаптивним управлінням ресурсами.

Формулювання мети дослідження

Метою дослідження є розробка методу інтеграції протоколу BitTorrent в приватні сегменти програмно-конфігурованих мереж для забезпечення ефективної та безпечної передачі великих обсягів даних в корпоративних мережах. Запропонований метод дає змогу мінімізувати затримки передачі, оптимізувати використання мережних ресурсів та знизити ризики несанкціонованого доступу до конфіденційних даних, особливо в умовах обмеженості обчислювальних ресурсів.

Викладення основного матеріалу дослідження

Для підвищення безпеки передачі даних в корпоративних комп'ютерних мережах традиційно застосовуються технології віртуальних приватних мереж (VPN), які забезпечують конфіденційність, цілісність та автентичність інформаційних потоків. Модель передачі даних через VPN на основі OpenVPN передбачає використання єдиного захищеного тунелю, що з'єднує клієнтський вузол з серверним сегментом корпоративної мережі за маршрутом «клієнт – VPN-сервер – цільовий вузол» (рис. 1) [16].

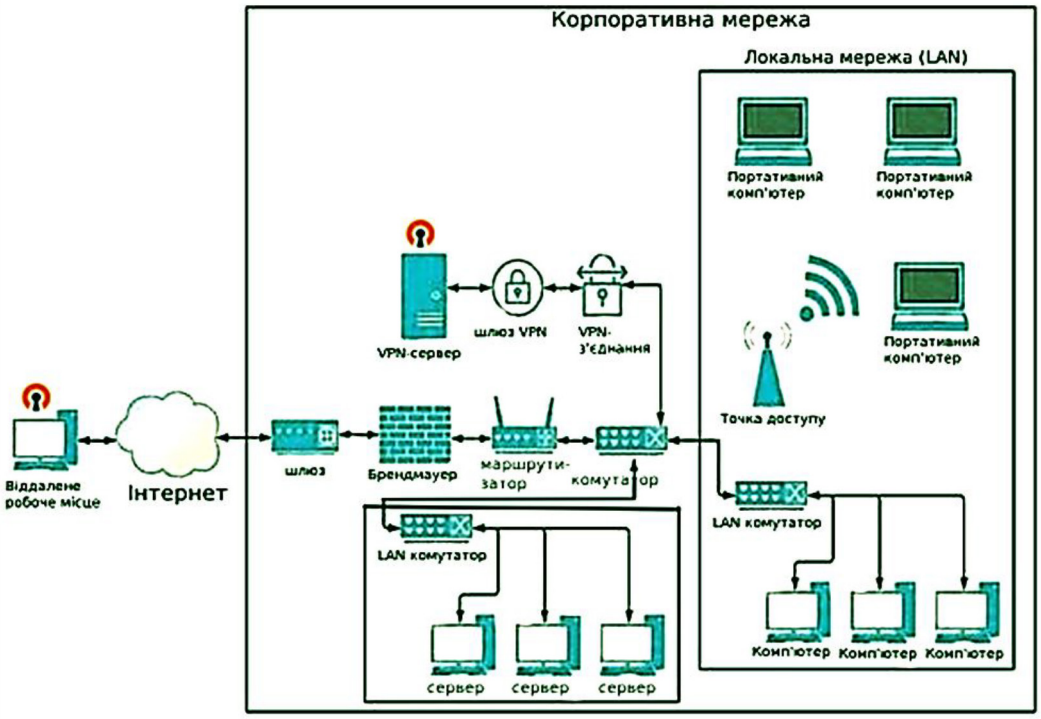


Рис. 1. Модель сегменту корпоративної мережі з OpenVPN-доступом

Надалі архітектура поступово ускладнюється від базової VPN-конфігурації до комбінованої схеми SDN + VPN + BitTorrent, що передбачає моделювання ізольованого сегмента приватної SDN-мережі. Це дозволяє поєднати захищений канал зв'язку на основі VPN з високою ефективністю передачі даних в складних мережних інфраструктурах в умовах підвищених вимог до безпеки передачі даних та обмежених апаратних ресурсах.

Базова модель передбачає шифрування трафіку стандартними криптографічними алгоритмами, які реалізуються обраним протоколом, автентифікація користувачів забезпечує контроль доступу та реалізацію політик

безпеки відповідно до їх прав і привілеїв, що гарантує конфіденційність та цілісність даних. Основними перевагами цього підходу є простота конфігурації та адміністрування, а також висока сумісність з широким спектром клієнтських пристроїв та операційних систем.

Модель, що поєднує VPN-доступ на основі OpenVPN з передачею даних через протокол BitTorrent, покликана усунути обмеження традиційних централізованих систем, ефективність яких суттєво знижується під час пікових навантажень, особливо в корпоративних мережах з підвищеними вимогами до надійності та відмовостійкості (рис. 2). Функціонування цієї моделі передбачає, що клієнтський вузол спочатку встановлює захищене з'єднання з VPN-сегментом, після чого обмін інформацією здійснюється в межах децентралізованої пірингової мережі BitTorrent. Це дає змогу вузлам взаємодіяти безпосередньо, розподіляючи трафік та обчислювальні ресурси, що забезпечує високу швидкість та стабільність передавання великих обсягів даних при збереженні захищених каналів доступу.

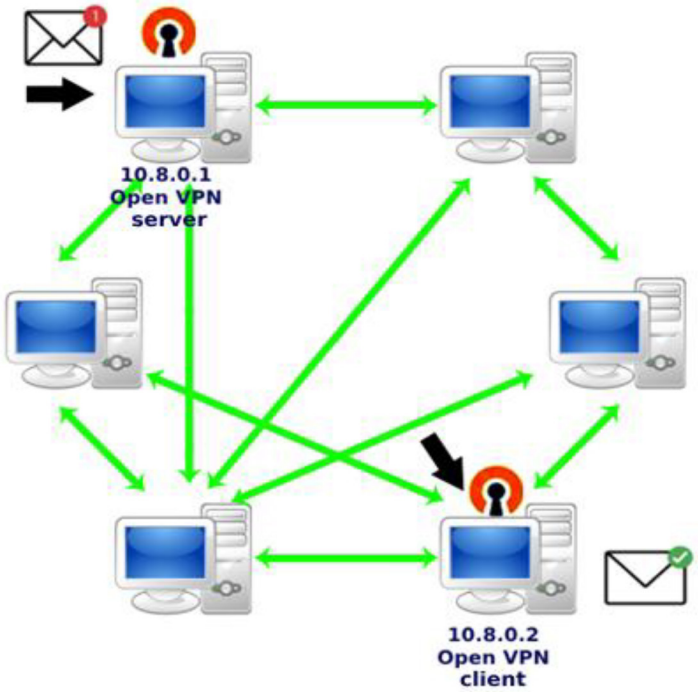


Рис. 2. Схема сегменту мережі з VPN-доступом та передачею даних через протокол BitTorrent

Ключовими перевагами такого підходу є зниження навантаження на центральний VPN-сервер та підвищення відмовостійкості системи завдяки використанню розподіленої архітектури, що дозволяє зберігати працездатність мережі в разі виходу з ладу окремих вузлів. Однак ця модель має ризики витоку метаданих та потенційного несанкціонованого доступу до окремих ресурсів.

Модель ізольованого сегмента приватної SDN-мережі на основі комбінації OpenVPN, BitTorrent та централізованого управління через SDN-контролер представляє кінцеву стадію розвитку попередніх моделей, інтегруючи їх переваги та мінімізуючи ключові обмеження. В порівнянні з класичними централізованими VPN-мережами ця архітектура дозволяє знизити навантаження на центральний вузол та забезпечити масштабований внутрішній обмін даними завдяки застосуванню протоколів пірингових мереж. На відміну від моделі, де BitTorrent функціонує всередині єдиного VPN-сегмента без чіткої сегментації, запропонований метод забезпечує повну логічну ізоляцію сегментів завдяки використанню SDN, що унеможливлює несанкціонований витік даних між ними. SDN-контролер виконує роль центрального вузла сегмента, формуючи маршрути, застосовуючи політики безпеки та контролюючи доступ в режимі реального часу. В корпоративній мережі реалізується ієрархічна структура контролерів, де один з них забезпечує централізоване управління в межах ізольованого сегмента, тоді як інші забезпечують маршрутизацію суміжних сегментів мережі, що дозволяє гнучко балансувати навантаження та швидко адаптуватися до змін мережної топології. Інтеграція BitTorrent в межах ізольованого SDN-сегмента забезпечує стійкість до відмов окремих вузлів, підвищуючи загальну резильєнтність системи (рис. 3).

Реалізація такої архітектури можлива на основі технологій віртуалізації, зокрема за допомогою віртуальних машини або контейнерів на платформах Proxmox та Docker. Це забезпечує спрощене масштабування, оптимізацію витрат на інфраструктуру, а також можливість інтеграції з фізичним обладнанням та хмарними середовищами для підвищення рівня безпеки, гнучкості та відмовостійкості.

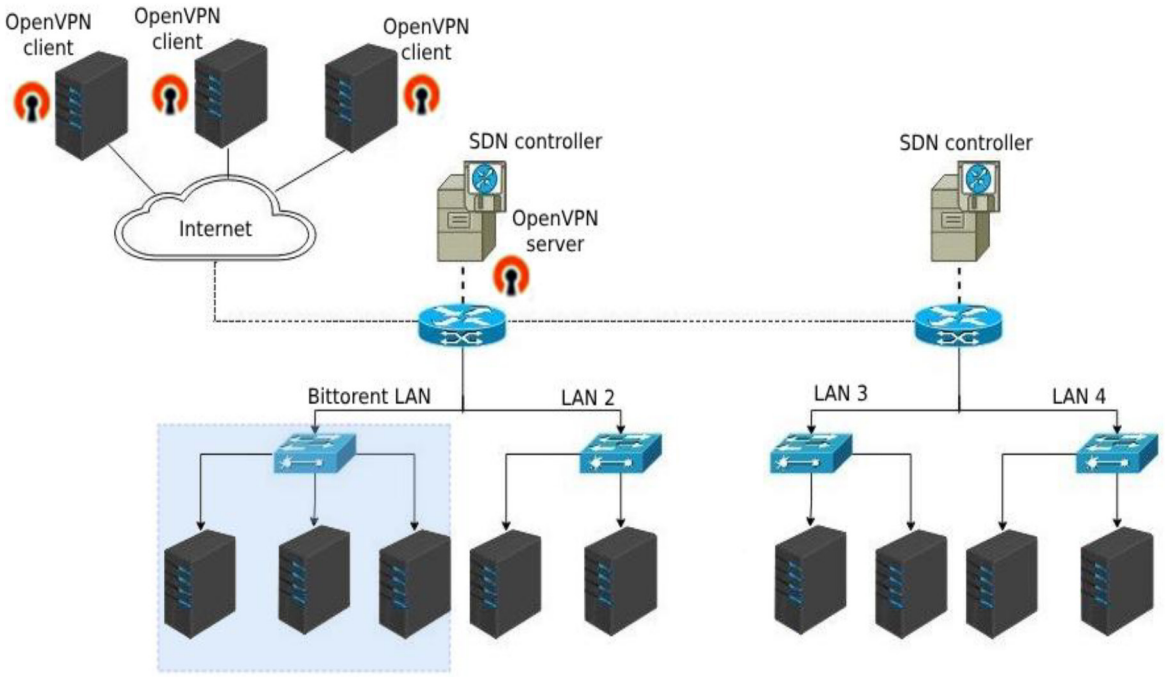


Рис. 3. Схема ізольованого сегменту приватної корпоративної SDN-мережі

Алгоритм розгортання ізольованого сегмента приватної SDN-мережі

Розгортання ізольованого сегмента приватної SDN-мережі передбачає виконання наступних етапів:

Побудова локальної SDN-мережі.

На цьому етапі формується мережна структура з логічним розділенням площин управління та передачі даних. Для цього використовуються комутатори, сумісні з протоколом OpenFlow, які забезпечують маршрутизацію трафіку відповідно до політик доступу, налаштованих на ієрархічній структурі SDN-контролерів приватної корпоративної мережі. Один з таких контролерів забезпечує централізоване управління ізольованим сегментом, тоді як інші забезпечують маршрутизацію суміжних сегментів приватної SDN-мережі. Також налаштовуються базові правила маршрутизації та моніторинг трафіку між ізольованим сегментом і зовнішніми ресурсами мережі.

Впровадження VPN-доступу через OpenVPN.

На другому етапі на SDN-контролері, який виконує роль шлюзового вузла між зовнішніми користувачами та внутрішньою мережею ізольованого сегмента, розгортається OpenVPN-сервер, що забезпечує надійне шифрування трафіку та автентифікацію клієнтів, а також виконується налаштування його параметрів відповідно до корпоративних політик безпеки. Це включає механізми контролю доступу, що визначають права користувачів та систем, а також моніторингу та логування VPN-сесій. Інтеграція VPN-сервера з SDN-контролером на одному пристрої спрощує управління ізольованим сегментом та забезпечує гнучке розмежування прав доступу на основі ролей, атрибутів користувачів і поточного стану мережі. Реалізація цього етапу гарантує доступ до сегмента лише авторизованим користувачам, захищаючи його від несанкціонованих підключень.

Налаштування обміну даними через BitTorrent.

На цьому етапі організовується обмін даними всередині ізольованого сегмента за допомогою протоколу BitTorrent. Ключовою особливістю є статичне призначення IP-адрес вузлам ізольованого сегменту, що усуває необхідність централізованих трекерів та дозволяє кожному піру бути ідентифікованим і доступним. BitTorrent-клієнти вузлів ізольованого сегмента налаштовуються на приватний режим роботи, що запобігає несанкціонованому витоку конфіденційної інформації поза межами сегмента. На SDN-контролері встановлюються правила фільтрації та сегментації трафіку, суворо регламентуючи канали обміну та забезпечуючи ізоляцію сегмента.

Реалізація політик маршрутизації SDN-контролером.

Фінальний етап цього алгоритму забезпечує реалізацію динамічного управління маршрутами, балансування навантаження та застосування політик якості обслуговування QoS. На SDN-контролер додаються правила застосування політик QoS, які виконуються автоматично з урахуванням поточного стану мережі. Використання QoS дозволяє пріоритезувати критично важливий трафік, зокрема транзакційні дані або керуючі повідомлення, та забезпечити гнучке реагування на зміни мережної топології і навантаження. Політики контролю доступу запобігають несанкціонованому перенаправленню трафіку та витоку даних між сегментами, підвищуючи рівень безпеки та надійності мережі.

Оцінка ефективності розробленого методу

На основі запропонованого алгоритму побудови ізольованого сегменту приватної SDN-мережі було сформульовано багатокритеріальну оптимізаційну задачу. Її основною метою є комплексне підвищення ефективності функціонування сегмента шляхом мінімізації ключових критеріїв, зокрема затримки при передачі даних захищеним каналом, використання ресурсів для забезпечення відмовостійкості роботи SDN-контролера та ймовірності отримання не цілісного обсягу даних.

Загальна ефективність ізольованого сегменту визначається величиною, оберненої зваженій суми цих критеріїв, що забезпечує комплексну оцінку продуктивності системи. При цьому враховується, що пропускна здатність в межах ізольованого сегмента повинна бути не меншої пропускної здатності захищеного каналу, а кількість одночасних з'єднань не повинна перевищувати максимальну кількість з'єднань, обмежених конфігурацією VPN-сервера. Водночас використання ресурсів SDN-контролера не перевищує його конфігураційної ємності, а в межах сегмента повинен бути щонайменше один активний вузол для забезпечення надійного обміну даними. Такий підхід забезпечує комплексну оцінку продуктивності системи. Загальна цільова функція ефективності функціонування ізольованого сегменту формулюється у вигляді наступної багатокритеріальної оптимізаційної задачі:

$$\min F = w_1 \cdot F_L + w_2 \cdot F_R + w_3 \cdot F_P, \ при \begin{cases} B \geq B_{VPN}, \\ R < R_{\max}, \\ N \geq N_{\min}, \\ m < m_{\max}, \end{cases} \tag{1}$$

де w_i – вагові коефіцієнти, які визначають пріоритети компонентів оптимізаційної задачі та задовольняють вимогам $w_i \geq 0, \ \sum_{i=1}^{3} w_i = 1;$ F_L – функція затримки передачі даних; F_R – функція використання ресурсів SDN-контролера; F_P – функція ймовірності отримання не цілісного обсягу даних; B, B_{VPN} – пропускна здатність ізольованого сегменту та захищеного VPN-каналу відповідно, біт/с; R, $R_{\max}$ – використані (CPU, RAM, I/O тощо) та доступні ресурси SDN-контролера відповідно; N, – загальна кількість вузлів – пірів в ізольованому сегменті, од.; $N_{\min}$ – кількість активних вузлів – пірів в межах сегмента, од.; $m_{\max}$ – максимальна кількість одночасних з'єднань, обмежених конфігурацією VPN-сервера, од.; m – кількість одночасних з'єднань під час роботи ізольованого сегменту, од.

Загальна ефективність функціонування ізольованого сегмента розраховується як:

$$E = \frac{1}{F} = w_1 \cdot F_L + w_2 \cdot F_R + w_3 \cdot F_P. \tag{2}$$

Компоненти оптимізаційної задачі (2) представлені нижче:

1. Мінімізація функції затримки передачі даних забезпечується зниженням часу процесів шифрування та дешифрування при передачі даних через захищений канал та зниженням часу пошуку пірів в межах ізольованого сегмента. Складність криптографічних алгоритмів збільшує час шифрування/дешифрування, а час пошуку активних пірів для завантаження цілісного обсягу даних може збільшувати затримку передачі даних в разі непрацездатності піру з необхідним фрагментом даних, що спричиняє додатковий час для пошуку дубльованого фрагменту на активних пірах. Функція затримки передачі даних формулюється як:

$$F_L = f(L_t, V, N_{\min}, m), \ при \begin{cases} B \leq B_{SDN}, \\ V \leq V_{\max}, \\ N \geq N_{\min}, \\ m < m_{\max}, \end{cases} \tag{3}$$

де L_t – загальна затримка передачі даних, с; V – швидкість пошуку пірів в межах ізольованого сегменту, од./с;

Загальна затримка передачі даних в межах ізольованого сегмента залежить від пропускної здатності VPN-каналу, кількості одночасних з'єднань та швидкості пошуку пірів. Загальна затримка передачі даних розраховується як:

$$L_t = \alpha_1 \cdot m \cdot L_{VPN} + \alpha_2 \cdot L_{P2P}, \tag{4}$$

де α_1, α_2 – вагові коефіцієнти, що визначають пріоритети компонентів оцінки та задовольняють вимогам $\alpha_1 \geq 0$, $\alpha_2 \geq 0$, $\alpha_1 + \alpha_2 = 1$; L_{VPN} – затримка передачі даних, що виникає в захищеному каналі, с; L_{P2P} – затримка передачі даних, що виникає через пошук активних пірів, с;

Затримка передачі даних в захищеному каналі залежить від пропускної здатності каналу та часу шифрування/дешифрування даних в залежності від обраного криптографічного алгоритму та розраховується як:

$$L_{VPN} = \frac{m \cdot D}{B_{VPN}} = t_{tr} + t_{enc} + t_{dec},\ при\ B_{VPN} \le B_{SDN}, \quad (5)$$

де D – обсяг даних, переданих через VPN-канал за визначений проміжок часу t, біт; B_{VPN} – пропускна здатність VPN-каналу з шифруванням та дешифруванням даних, біт/с; B_{SDN} – пропускна здатність VPN-каналу без шифрування, який визначено для обслуговування ізольованого сегменту на SDN-контролері, біт/с; t_{tr} – час передачі даних через VPN-канал без шифрування, с; t_{enc} – час шифрування даних на VPN- каналі, с; t_{dec} – час дешифрування даних наVPN- каналі, с.

Швидкість пошуку активних пірів залежить від кількості пірів у сегменті та продуктивності SDN-контролера, зокрема кількості операцій I/O, та розраховується як:

$$V = \frac{Z_{SDN}}{t}, \quad (6)$$

де Z_{SDN} – кількість операцій, необхідних для перевірки активності одного піру; t – час виконання цих операцій, с.

Відповідно, затримка, яка виникає через пошук активних пірів, розраховується як:

$$L_{P2P} = \frac{Z_{SDN} \cdot N_{\min}}{V}, \quad (7)$$

де Z_{SDN} – кількість операцій, необхідних для перевірки активності одного піра, од.;

2. Мінімізація функції використання ресурсів SDN-контролера забезпечується використанням доступних ресурсів, зокрема пам'яттю, кількістю процесорів та додатковими операціями для забезпечення управління трафіком та надійної роботи реалізованого на SDN-контролері VPN-сервера. Функція використання ресурсів SDN-контролера формулюється як:

$$F_R(R, R_{e/d}, R_q, m) = \beta_1 \cdot R + \beta_2 \cdot m \cdot R_{e/d} + \beta_3 \cdot R_q, \quad при (R + R_{e/d} + R_q) \le R_{\max}, \quad m < m_{\max}, \quad (7)$$

де β_1, β_2 – вагові коефіцієнти які визначають пріоритети компонентів оцінки та задовольняють вимогам $\beta_1 \ge 0$, $\beta_2 \ge 0$, $\beta_1 + \beta_2 = 1$; $R_{e/d}$ – використані ресурси (CPU, RAM, I/O тощо) контролера на процеси шифрування/дешифрування під час передачі даних захищеним VPN-каналом; R_q – використані ресурси (CPU, RAM, I/O тощо) для управління трафіком;

3. Функція ймовірності отримання не цілісного обсягу даних зменшується зі зростанням кількості пірів у сегменті, але може зростати через велику кількість одночасних з'єднань під час роботи ізольованого сегмента. Функція ймовірності отримання не цілісного обсягу даних формулюється як:

$$F_P = \gamma_1 \cdot \frac{1}{N_{\min}} + \gamma_2 \cdot \frac{m}{N},\ при\ N \ge N_{\min}, m < m_{\max}, \quad (8)$$

де γ_1, γ_2 – вагові коефіцієнти які визначають пріоритети компонентів оцінки та задовольняють вимогам $\gamma_1 \ge 0$, $\gamma_2 \ge 0$, $\gamma_1 + \gamma_2 = 1$.

Запропонований підхід дозволяє комплексно оцінити ефективність роботи ізольованого сегмента приватної SDN-мережі, враховуючи ключові параметри його функціонування. Це дозволяє кількісно оцінити вплив пропускної здатності, затримки передачі даних, складності криптографічних алгоритмів, ресурсоємності контролера та відмовостійкості на загальну продуктивність сегмента.

Постановка експерименту

Постановка експерименту була виконана на базі лабораторії обчислювальних систем і мережних технологій кафедри електронних обчислювальних машин Харківського національного університету радіоелектроніки. Для кожної з трьох моделей розгорнуто відповідну конфігурацію програмного забезпечення та визначено параметри тестування.

1. Модель з VPN-доступом.

VPN-сервер OpenVPN розгорнуто на віртуальній машині (ВМ) з операційною системою Ubuntu 20.04. ВМ оснащено чотириядерним процесором з тактовою частотою 3,0 ГГц та 4 ГБ оперативної пам'яті. Пропускна здатність мережного інтерфейсу становить до 100 Мбіт/с, що дозволяє обслуговувати до 100 одночасних клієнтських підключень відповідно до конфігурації VPN-сервера без значного зниження продуктивності. Також конфігурація VPN-сервера передбачає використання алгоритму симетричного шифрування AES-256 в поєднанні з протоколом захищеного обміну ключами TLS 1.3, автентифікація користувачів реалізується на основі цифрових сертифікатів. У брандмауері відкрито порти 943/TCP для адміністративного вебінтерфейсу OpenVPN GUI та 1194/TCP для VPN-з'єднань. Для підвищення рівня захисту сервера також налаштовано перенаправлення трафіку на інші порти.

Процедура підключення користувача передбачає встановлення клієнтського програмного забезпечення OpenVPN Client, імпорт конфігураційного файлу та відповідного сертифіката автентифікації. Адміністратор

мережі за допомогою вебінтерфейсу здійснює створення користувацьких облікових записів, призначення параметрів доступу та контроль за активними сесіями.

2. Модель з VPN-доступом та передачею даних через BitTorrent.

VPN-сервер було розгорнуто на віртуальній машині з ОС Ubuntu 20.04, оснащеній чотириядерним процесором. Обсяг оперативної пам'яті збільшено до 20 ГБ для зберігання тимчасових файлів та логів роботи BitTorrent-клієнтів. Пропускна здатність мережного інтерфейсу становить до 100 Мбіт/с. Для коректного функціонування протоколу були відкрито порти 6346/TCP/UDP та 6886/TCP/UDP, що забезпечують роботу клієнтів Gnutella та BitTorrent відповідно. Передача даних у мережі здійснювалася через налаштований VPN-сервер, який надавав клієнтам сертифікати створення захищеного каналу. Обмін даними організовувався через оверлейну мережу Gnutella, для взаємодії клієнтів в межах мережі було застосовано програмне забезпечення FrostWire, що підтримує передачу даних протоколами Gnutella та BitTorrent.

3. Модель з SDN-контролером, VPN-доступом та ізольованим сегментом з передачею через BitTorrent.

Запропонований метод було реалізовано моделюванням ізольованого сегмента приватної корпоративної мережі на основі SDN-контролера Ryu з реалізацією VPN-доступу на основі OpenVPN та внутрішньою передачею даних через протокол BitTorrent. Розгортання SDN-контролера та VPN-сервера здійснювалося на одній віртуальній машині з операційною системою Ubuntu 20.04, обладнаній центральним процесором з чотирма обчислювальними ядрами та тактовою частотою 3,0 ГГц. Обсяг оперативної пам'яті було збільшено до 20 ГБ, що забезпечило стабільну роботу обох сервісів без зниження продуктивності.

Шифрування даних реалізовано виключно на рівні VPN-сервера із застосуванням алгоритму AES-256 в поєднанні з протоколом TLS 1.3 та клієнтською аутентифікацією за допомогою цифрових сертифікатів. Передача внутрішнього трафіку в межах ізольованого сегмента здійснюється без додаткового шифрування, що знижує затримки та підвищує ефективну пропускну здатність системи.

На рівні SDN-контролера було реалізовано політики доступу до ізольованого сегмента, зокрема перенаправлення трафіку та відкриття порту 5555/TCP для VPN-каналу, через який здійснюється єдиний доступ до ізольованого SDN-сегмента. В межах ізольованого сегменту додатково здійснюється відкриття порту 6886/TCP для обміну даними через протокол BitTorrent, що гарантує контрольований обмін даними та підвищує безпеку внутрішнього трафіку. Передача даних за протоколом BitTorrent реалізована через TCP, тоді як UDP використовується переважно для службових операцій, зокрема пошуку активних вузлів у сегменті, що забезпечує контрольовану взаємодію між пірами. В межах ізольованого сегмента всім вузлам призначаються статичні IP-адреси, що спрощує процедуру виявлення пірів і скорочує затримки під час підключення нових вузлів.

Для оцінки безпеки передачі даних за межі ізольованого SDN-сегмента проведено аналіз трафіку за допомогою Wireshark. Моніторинг зосереджувався на потоці даних, що проходять через VPN-тунель, який забезпечує захищене з'єднання між ізольованим сегментом та SDN-контролером. Після фільтрації пакетів, що передавалися виключно через VPN-тунель, підтверджено, що їх вміст зашифрований (рис. 3). Передача даних відбувається всередині ізольованого сегменту за допомогою протоколу BitTorrent, тоді як вихід за його межі відбувається через захищене VPN-з'єднання, контрольоване SDN-контролером. Це забезпечує додатковий рівень безпеки та управління трафіком.

Запропонований метод забезпечує оптимізовану передачу даних в ізольованому сегменті приватної SDN-мережі, зберігаючи при цьому переваги децентралізованого поширення інформації та одночасно підвищуючи рівень безпеки, запобігаючи несанкціонованому доступу та витоку даних.

На графіку (рис. 4) представлено середню затримку передачі TCP-пакетів, що виникає в досліджуваних моделях на основі експериментальних даних. Результати демонструють, що впровадження SDN-контролера з реалізованим на ньому VPN-сервером дозволяє зменшити затримку передачі даних порівняно з моделлю захищеної передачі через мережу BitTorrent, при цьому зберігаючи ключові переваги P2P-передачі даних, зокрема масштабованість та децентралізацію, в межах ізольованого сегменту.

Крім того підтверджено, що застосування P2P-технологій в приватних SDN-мережах у поєднанні з локалізацією трафіку в межах ізольованого сегмента дозволяє мінімізувати затримки та покращити продуктивність системи, забезпечуючи необхідний рівень безпеки передачі даних.

Висновки

В роботі вперше запропоновано метод організації ізольованого сегмента приватної SDN-мережі із захищеним доступом через VPN та вбудованою P2P-дистрибуцією на основі протоколу BitTorrent, який інтегрує централізоване керування політиками та маршрутизацією на рівні SDN-контролера з децентралізованим розподілом трафіку всередині сегмента, що дозволяє одночасно підвищити рівень захищеності та керованості, мінімізувати внутрішні затримки завдяки балансуванню навантаження між вузлами, забезпечити пріоритезацію сервісів та масштабованість архітектури для критично важливих корпоративних і приватних блокчейн-систем.

Аналіз ефективності запропонованого методу підтвердив, що використання ізольованих сегментів у SDN мережах дозволяє мінімізувати внутрішні затримки та підвищити контроль над маршрутизацією трафіку.

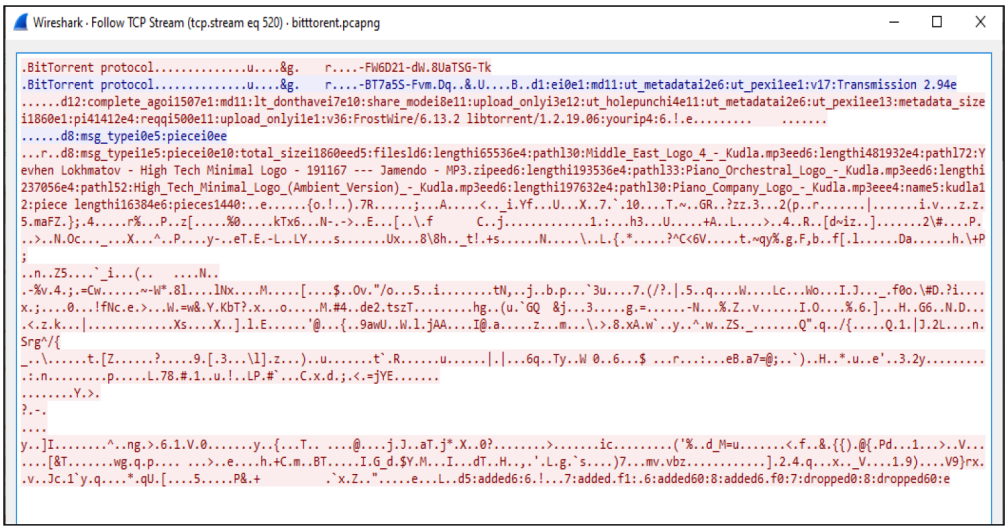


Рис. 4. Візуалізація шифрованого потоку даних при передачі даних з використанням протоколу BitTorrent з ізольованого сегменту приватної SDN-мережі в Wireshark

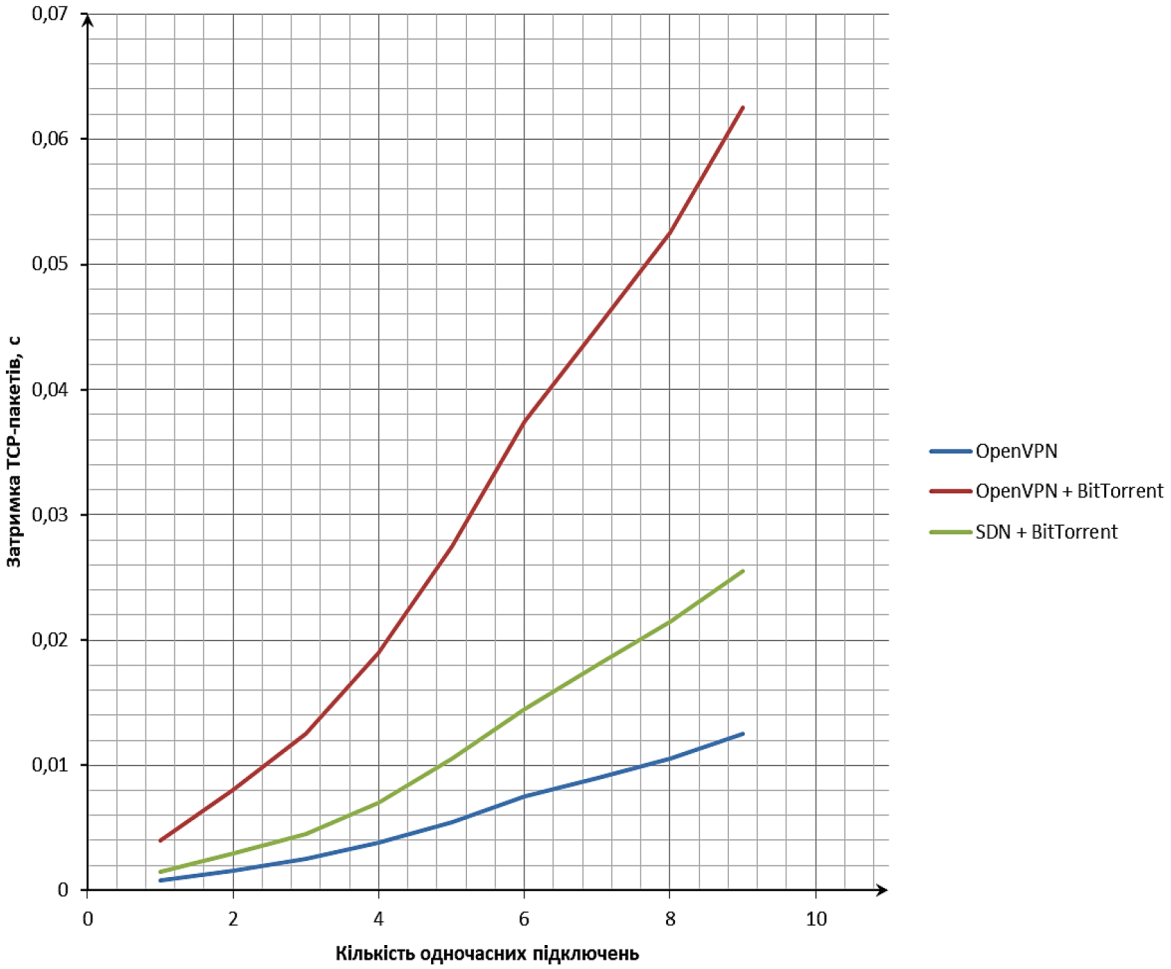


Рис. 4. Діаграма середньої затримки TCP-пакетів в досліджуваних моделях

При практичному застосуванні методу необхідно враховувати характер даних для передачі, вимоги до обчислювальних ресурсів SDN-контролера ізольованого сегмента, рівень захищеності та доступність ресурсів для реалізації сегмента. Отже, запропонований метод може бути корисним для малого та середнього бізнесу, який прагне створювати масштабовані та високопродуктивні корпоративні мережі, що відповідають сучасним вимогам до

мінімізації затримок, підвищення надійності та зниження внутрішнього навантаження, що є критичним для побудови складних інформаційних інфраструктур.

Крім того, така архітектура має значний потенціал для розгортання приватних блокчейн-мереж. Захищений доступ до вузлів блокчейна через VPN-тунель гарантує підключення лише авторизованих користувачів та систем. Водночас, SDN-контролер дозволяє управляти міжвузловими взаємодіями, налаштовувати пріоритети та обмеження для різних типів трафіку, зокрема транзакційних повідомлень або реплікації блоків.

Майбутні дослідження можуть бути спрямовані на оптимізацію цього методу в різних умовах та практичне застосування архітектури для побудови приватних блокчейн-мереж.

Список використаної літератури

1. Latif Z., Umer Q., Lee C., Sharif K, Li F., Biswas S. A Machine Learning-Based Anomaly Prediction Service for Software-Defined Networks. *Sensors*. 2022; vol. 22, No. 21. P. 8434.

2. Wadhwani P, Software Defined Networking Market – By Component (Solution [Physical Network Infrastructure, SDN Controller, SDN Application], Service [Professional, Managed]), By End Use (Enterprise, Cloud Service Provider, Telecom Service Provider) & Forecast, 2024–2032. URL https://www.gminsights.com/industry-analysis/software-defined-networking-sdn-market (дата звернення: 10.06. 2025).

3. Suliyanti W. N., Sari R. F. Blockchain-Based Double-Layer Byzantine Fault Tolerance for Scalability Enhancement for Building Information Modeling Information Exchange. *Big Data Cogn. Comput*. 2023. Vol. 7. P. 90.

4. Tanuja J., Goyal N., Ram M. An approach to analyze reliability indices in peer-to-peer communication systems. *Cybernetics and Systems*. 2022. Vol. 53. No. 8. P. 716–733.

5. Khalid M. I., Ehsan I., Al-Ani A., Iqbal J., Hussain S., Ullah S. S., Nayab A comprehensive survey on blockchain-based decentralized storage networks. *IEEE Access*. 2023.Vol. 11. P. 10995–11015.

6. М. Кренцін, Л. Куперштейн. Метод обміну ідентифікаційними даними між вузлами пірингової мережі на основі технології NFC. *Кібербезпека: освіта, наука, техніка*. 2024. вип. 1, № 25, С. 79–88.

7. Abdullahi Yari I, Dehling T, Kluge F, Geck J, Sunyaev A, Eskofier B. Security Engineering of Patient-Centered Health Care Information Systems in Peer-to-Peer Environments: Systematic Review. *J Med Internet Res*. 2021. Vol. 23. No. 11. P. e24460.

8. BitTorrent : The World's Most Popular Torrent Client. URL:https://www.bittorrent.com/btt/btt-docs/BitTorrent_(BTT)_White_Paper_v0.8.7_Feb_2019.pdf (дата звернення: 21.07.2025).

9. Куперштейн Л. М. Аналіз тенденцій розвитку пірингових мереж. *Вісник Хмельницького національного університету. Технічні науки*. 2021. № 4. С. 25–29.

10. Aldabbas H. Efficient bandwidth allocation in SDN-based peer-to-peer data streaming using machine learning algorithm. *The Journal of Supercomputing*. 2023. Vol. 79. No. 6. P. 6802–6824.

11. Hill W., Acquaah Y. T., Mason, J., Limbrick D., Teixeira-Poit S., Coates C., Roy K. DDoS in SDN: A review of open datasets, attack vectors and mitigation strategies. *Discover Applied Sciences*. 2024. Vol. 6. No. 9. P. 472.

12. Pang S., Zeng D., Chen X. Research on SDN-based data center network traffic management and optimization. *2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA)*, Shenyang, China, 21–23 January 2022. 2022. URL: https://doi.org/10.1109/icpeca53709.2022.9718973 (date of access: 15.08.2025).

13. Suhail A., Mir Ajaz H. Scalability, consistency, reliability and security in SDN controllers: a survey of diverse SDN controllers. *Journal of Network and Systems Management*. 2021. Vol. 29. No 1. P. 9.

14. Hwang I.-S., Rianto A., Kharga R., Ab-Rahman M. S. Global P2P BitTorrent Real-Time Traffic Over SDN-Based Local-Aware NG-PON2. *IEEE Access*. 2022. Vol. 10, P. 76884–76894.

15. Vemasani P., Modi S. Building Resilient Distributed Systems: Fault-Tolerant Design Patterns for Stateful Workflows. *International Journal of Computer Engineering and TechNology*. 2024. Vol. 15. P. 169–181.

16. Ткачов В. М., Чепурна І. С., Фесенко Т. Г. Метод мультирівневого VPN-тунелювання для забезпечення віддаленого доступу до вузлів екстранет-мережі. *Вісник Херсонського національного технічного університету*. 2024. № 3(90). С. 299–308.

References

1. Latif, Z., Umer, Q., Lee, C., Sharif, K., Li, F., & Biswas, S. (2022). A Machine Learning-Based Anomaly Prediction Service for Software-Defined Networks. Sensors, 22(21), 8434. https://doi.org/10.3390/s22218434

2. Preeti Wadhwani. 2024, February. Software Defined Networking Market – By Component (Solution [Physical Network Infrastructure, SDN Controller, SDN Application], Service [Professional, Managed]), By End Use (Enterprise, Cloud Service Provider, Telecom Service Provider) & Forecast, 2024 – 2032 (Report ID: GMI2395). Global Market Insights Inc. Retrieved June 10, 2025, from https://www.gminsights.com/industry-analysis/software-defined-networking-sdn-market

3. Suliyanti, W. N., & Sari, R. F. (2023). Blockchain-Based Double-Layer Byzantine Fault Tolerance for Scalability Enhancement for Building Information Modeling Information Exchange. Big Data and Cognitive Computing, 7(2), 90. https://doi.org/10.3390/bdcc7020090

4. Joshi, T., Goyal, N., & Ram, M. (2022). An approach to analyze reliability indices in peer-to-peer communication systems. *Cybernetics and Systems*, *53*(8), 716–733.

5. Khalid, M.I., Ehsan, I., Al-Ani, A.K., Iqbal, J., Hussain, S., Ullah, S.S., & Nayab. (2023). A Comprehensive Survey on Blockchain-Based Decentralized Storage Networks. IEEE Access, 11, 10995–11015.

6. Krentsin, Mykhailo & Kupershtein, Leonid. (2024). NFC technology as a means of protected exchange of identification data between peer-to-peer network nodes Cybersecurity Education Science Technique. 1. 79–88. 10.28925/2663-4023.2024.25.7988. (in Ukrainian)

7. Abdullahi Yari, I., Dehling, T., Kluge, F., Geck, J., Sunyaev, A., & Eskofier, B. (2021). Security Engineering of Patient-Centered Health Care Information Systems in Peer-to-Peer Environments: Systematic Review. Journal of medical Internet research, 23(11), e24460. https://doi.org/10.2196/24460

8. BitTorrent. (2019, February). *BitTorrent (BTT) white paper v0.8.7*. Retrieved from https://www.bittorrent.com/btt/btt-docs/BitTorrent_(BTT)_White_Paper_v0.8.7_Feb_2019.pdf

9. Krentsin, Mykhailo & Kupershtein, Leonid. (2022). Analysis of peer-to-peer network protocols (in Ukrainian).

10. Aldabbas, H. (2023). Efficient bandwidth allocation in SDN-based peer-to-peer data streaming using machine learning algorithm. *The Journal of Supercomputing*, *79*(6), 6802–6824.

11. Hill, W., Acquaah, Y. T., Mason, J., Limbrick, D., Teixeira-Poit, S., Coates, C., & Roy, K. (2024). DDoS in SDN: a review of open datasets, attack vectors and mitigation strategies. *Discover Applied Sciences*, *6*(9), 472.

12. Pang, S., Zeng, D., & Chen, X. (2022, January). Research on SDN-based data center network traffic management and optimization. In *2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA)* (pp. 600-604). IEEE.

13. Ahmad, Suhail & Mir, Ajaz. (2021). Scalability, Consistency, Reliability and Security in SDN Controllers: A Survey of Diverse SDN Controllers. Journal of Network and Systems Management. 29. 10.1007/s10922-020-09575-4.

14. Hwang, I. S., Rianto, A., Kharga, R., & Ab-Rahman, M. S. (2022). Global P2P BitTorrent real-time traffic over SDN-based local-aware NG-PON2. *IEEE Access*, *10*, 76884–76894.

15. Vemasani, P., & Modi, S. (2024). Building Resilient Distributed Systems: Fault-Tolerant Design Patterns for Stateful Workflows. *International Journal of Computer Engineering and Technology*, *15*, 169–181.

16. Tkachov V. M., Chepurna I. S., Fesenko T. H (2024). Metod mul'tyrivnevoho VPN-tunelyuvannya dlya zabezpechennya viddalenoho dostupu do vuzliv ekstranet-merezhi. *Visnyk Khersons'koho natsional'noho tekhnichnoho universytetu (electronic journal)*. vol. 3(90), pp. 299–308. Retrieved from: https://doi.org/10.35546/kntu2078-4481.2024.3.37 (accessed 10 August 2025).

Дата першого надходження рукопису до видання: 24.09.2025
Дата прийнятого до друку рукопису після рецензування: 21.10.2025
Дата публікації:

УДК 004.75:004.4:519.8 DOI

О. М. ШУШУРА
доктор технічних наук, доцент,
професор кафедри цифрових технологій в енергетиці
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
ORCID: 0000-0003-3200-720X

Д. А. ІГНАТОВ
аспірант кафедри цифрових технологій в енергетиці
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
ORCID: 0009-0008-2864-8526

МОДЕЛЮВАННЯ ПРИЙНЯТТЯ РІШЕНЬ З УПРАВЛІННЯ РЕСУРСАМИ МІКРОСЕРВІСІВ ТА ВИБОРУ ВІРТУАЛЬНИХ МАШИН У СЕРЕДОВИЩІ KUBERNETES

Сучасні хмарні системи, побудовані з використанням мікросервісних архітектур, стикаються з критичною проблемою ефективного управління ресурсами в умовах динамічного навантаження. Стандартні реактивні механізми оркестратора Kubernetes, такі як Horizontal Pod Autoscaler, часто виявляються недостатніми через затримки у реакції, відсутність врахування майбутніх трендів навантаження та інерційність процесів запуску нових реплік мікросервісів і віртуальних машин. Це призводить до неефективного використання ресурсів, коливань продуктивності (flapping) та зайвих операційних витрат. У відповідь на ці виклики в статті запропоновано комплексну математичну модель для оптимального проактивного управління ресурсами на основі прогнозування навантаження, динамічне масштабування мікросервісів із врахуванням затримок та оптимальне розміщення подів на віртуальних машинах.

Метою даної роботи є розробка математичної оптимізаційної моделі для автоматичного управління ресурсами мікросервісів та динамічного вибору віртуальних машин. Формалізовано предметну область шляхом визначення ключових множин: класів віртуальних машин, активних екземплярів віртуальних машин, типів мікросервісів та їхніх реплік (подів). На основі історичних даних метрик CPU та оперативної пам'яті виконується прогнозування навантаження на горизонті планування і для кожного типу мікросервіса розраховується бажана кількість реплік, необхідна для обробки очікуваного навантаження. Сформовано та формалізовано цільову функцію і обмеження задачі оптимального управління ресурсами мікросервісів та вибору віртуальних машин, спираючись на сукупність прийнятих припущень.

Запропонований підхід забезпечує підвищення відмовостійкості та продуктивності мікросервісних додатків при одночасному зниженні операційних витрат за рахунок усунення надмірного резервування та консолідації навантаження на найбільш економічно ефективних типах віртуальних машин. Результати роботи становлять теоретичну основу для подальших досліджень та практичної реалізації системи інтелектуального оркестрування для Kubernetes.

Ключові слова: *мікросервіси, Kubernetes, управління ресурсами, прогнозування навантаження, автомасштабування, оптимізація, віртуальні машини, інформаційні технології, інформаційні системи.*

O. M. SHUSHURA
Doctor of Technical Sciences, Associate Professor,
Professor at the Department of Digital Technologies in Energy
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"
ORCID: 0000-0003-3200-720X

D. A. IHNATOV
Postgraduate Student at the Department of Digital Technologies in Energy
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"
ORCID: 0009-0008-2864-8526

© Шушура О. М., Ігнатов Д. А., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

MODELING DECISION-MAKING FOR MICROSERVICE RESOURCE MANAGEMENT AND VIRTUAL MACHINE SELECTION IN A KUBERNETES ENVIRONMENT

Modern cloud systems built on microservice architectures face a critical challenge of efficient resource management under dynamic workloads. Standard reactive mechanisms of the Kubernetes orchestrator, such as the Horizontal Pod Autoscaler, often prove insufficient due to delayed responses, lack of consideration for future load trends, and the inertia of launching new microservice replicas and virtual machines. This results in inefficient resource utilization, performance fluctuations (flapping), and excessive operational costs.

In response to these challenges, the paper proposes a comprehensive mathematical model for optimal proactive resource management based on load forecasting, dynamic scaling of microservices with consideration of delays, and optimal pod placement on virtual machines.

The aim of this work is to develop a mathematical optimization model for automated microservice resource management and dynamic virtual machine selection. The subject domain is formalized through the definition of key sets: classes of virtual machines, active instances of virtual machines, types of microservices, and their replicas (pods). Based on historical CPU and memory metrics, load forecasting is performed on the planning horizon, and for each type of microservice, the desired number of replicas required to handle the expected load is calculated. The objective function and constraints of the problem of optimal resource management and virtual machine selection are constructed and formalized, relying on a set of accepted assumptions.

The proposed approach increases the resilience and performance of microservice applications while simultaneously reducing operational costs by eliminating excessive overprovisioning and consolidating workloads on the most cost-effective types of virtual machines. The results provide a theoretical foundation for further research and practical implementation of an intelligent orchestration system for Kubernetes.

Key words: *microservices, Kubernetes, resource management, load forecasting, autoscaling, optimization, virtual machines, information technology, information systems.*

Постановка проблеми

Однією з ключових задач сучасної IT-інфраструктури є забезпечення ефективного та відмовостійкого функціонування мікросервісних систем у хмарних середовищах за умов динамічно змінного навантаження. У разі різких змін навантаження, інформаційна система має швидко масштабуватись, щоб уникнути втрати продуктивності або відмови в обслуговуванні, але водночас небажано, аби це призводило до надмірних витрат на обчислювальні ресурси.

Ця проблема ускладнюється тим, що сучасні оркестратори, такі як Kubernetes, часто використовують реактивні механізми масштабування (наприклад, Horizontal Pod Autoscaler), які приймають рішення на основі минулих чи поточних даних, не враховуючи майбутніх трендів навантаження та інерційності самої системи. Запуск нових реплік мікросервісів та віртуальних машин потребує часу, що призводить до затримок у реакції на навантаження та неефективного використання ресурсів.

Крім того, вимоги до економічної ефективності та надійності сервісів обмежують можливості використання простих рішень, таких як надлишкове резервування ресурсів. Це обумовлює актуальність пошуку інтелектуальних та проактивних підходів до управління ресурсами. Вирішення цієї задачі є критично важливим для підвищення ефективності, надійності та зниження вартості експлуатації мікросервісних архітектур.

Аналіз останніх досліджень і публікацій

Стандартний підхід Kubernetes до управління ресурсами мікросервісів базується на Horizontal Pod Autoscaler (HPA), який використовує прості метрики (CPU, memory) і порогові значення для прийняття рішень про масштабування [1]. Проте цей підхід має суттєві недоліки. По-перше, HPA реагує на зміни лише після їх виникнення, що призводить до затримок у реакції [2]. По-друге, система не враховує час запуску нових реплік, що може спричиняти тимчасове зниження продуктивності. Крім того, HPA схильний до такого явища як flapping – частого додавання та видалення подів через чутливість до коротких сплесків навантаження.

Для подолання обмежень зазначеного вище реактивного підходу було запропоновано використовувати методи прогнозуючого масштабування, що використовують машинне навчання для передбачення майбутнього навантаження [3,4]. Також при порівнянні ARIMA, Prophet та LSTM-мереж виявилося, що найбільш вдалим є вибір на користь ARIMA [5]. Але незважаючи на переваги, дослідження зосереджуються лише на розрахунку бажаної кількості реплік, не розглядаючи питання їхнього розміщення на віртуальних машинах, що створює розрив між прогнозуванням потреб та реальним забезпеченням ресурсів.

Окрім цього, інше дослідження [6] визнає необхідність врахування інерційності системи автомасштабування. У роботі було розроблено алгоритм DACS (Delay-Aware Container Scheduling), який покликаний враховувати затримки обробки та мережі під час запуску подів, хоча і не розглядає час, якого потребує под при старті.

Вбудовані механізми розміщення подів у Kubernetes реалізуються через kube-scheduler та плагін NodeResourcesFit. Система підтримує різні стратегії розподілу ресурсів: LeastAllocated (за замовчуванням) прагне рівномірно розподілити навантаження між вузлами, тоді як MostAllocated та RequestedToCapacityRatio реалізують власну стратегію bin packing для максимізації утилізації [7].

У той же час було продемонстровано, що інтегровані підходи для розподілення подів можуть досягати до 58 % зменшення витрат порівняно зі стандартним методом [8]. Даний підхід поєднує best-fit bin packing для початкового розміщення, динамічне перепланування для консолідації робочих навантажень та автомасштабування для адаптації до змін попиту.

У контексті математичного моделювання управління ресурсами в Kubernetes можна відзначити дослідження, яке запропонувало динамічну модель, що враховує часові параметри та низькорівневі обмеження оптимізації [9]. Модель спрямована на регулювання не лише призначенням подів до віртуальних машин, але й керуванням увімкненням та вимкненням машин з метою мінімізації середньої кількості працюючих машин та максимізації коефіцієнта утилізації ресурсів.

Прогнозування навантаження також розглядалось у дослідженні [10] для динамічного вибору і міграції віртуальних машин у хмарних середовищах. Хоча при прогнозуванні використовувались агреговані метрики утилізації, що не дозволяє використовувати цей підхід при необхідності більш гнучких налаштувань, які можуть бути надзвичайно важливими у контексті мікросервісної архітектури.

Таким чином, актуальною є розробка комплексної моделі, яка б поєднувала прогнозування навантаження, динамічне масштабування мікросервісів з урахуванням затримок, оптимальне розміщення на віртуальних машинах та динамічне керування життєвим циклом цих машин для мінімізації сумарних витрат.

Формулювання мети дослідження

Метою даної роботи є розробка математичної оптимізаційної моделі для автоматичного управління ресурсами мікросервісів та динамічного вибору віртуальних машин. Розроблена модель має дозволити перейти від стандартного реактивного масштабування до проактивного, що забезпечить кращу продуктивність мікросервісної системи при динамічному навантаженні. Особливу увагу слід приділити вирішенню задачі пакування, оскільки це призведе до зниження операційних витрат завдяки більш ефективному замовленню віртуальних машин та використанню їхніх ресурсів.

Для досягнення поставленої мети необхідно вирішити задачі:

– описати основні характеристики предметної області з використанням математичного апарату теорії множин;

– сформувати та формалізувати цільову функцію та обмеження задачі оптимального управління ресурсами мікросервісів і вибору віртуальних машин, спираючись на сукупність прийнятих припущень.

Викладення основного матеріалу дослідження

Формалізація характеристик предметної області включає формування переліку основних об’єктів предметної області та представлення їх характеристик з використанням теорії множин.

Основними об’єктами предметної області в задачі управління ресурсами мікросервісів та вибору віртуальних машин визначено:

– віртуальні машини, які відносяться до певного типу, що клауд провайдер надає в оренду з певними обчислювальними ресурсами у вигляді CPU та пам’яті;

– міскросервіси (поди), які належать до певного типу мікросервісів та розміщені на деякій активній віртуальній машині.

Враховуючи, що зазначені об’єкти розглядаються в нестаціонарних умовах, під час формалізації їх характеристик будемо використовувати дискретну рівномірно розподілену з кроком Δt сітку часу $T = \{t_0, t_1, \ldots, t_N \,|\, t_i = i \cdot \Delta t\}$ де t_0 початковою точкою. Далі наведена формалізація характеристик зазначених об’єктів.

Під час запуску віртуальної машини розробник чи адміністратор повинен обрати певний тип віртуальної машини з поміж списку, який пропонується клауд провайдером. Різні типи розрізняються між собою як ресурсними характеристиками, так і вартістю використання. Відповідно, формується множина класів віртуальних машин $c_j \in C$ де j номер класу. Кожен тип віртуальної машини c_j характеризується кортежем (cpu_j, mem_j, $cost_j$) де cpu_j – кількість віртуальних CPU (vCPU); mem_j – об’єм оперативної пам’яті; $cost_j$ – вартість використання однієї віртуальної машини типу c_j за один крок Δt. Активні протягом часу спостереження віртуальні машини формують множину V, кожен елемент якої описується кортежем $(c_j, t_k^{start}, t_k^{stop})$. Кожен екземпляр v_k наслідує всі ресурсні характеристики і вартість свого класу $c_j \in C$, а також має моменти запуску і зупинки відповідно – t_k^{start}, t_k^{stop}. Зазначені множини дозволяють обчислювальні ресурси, які клієнт отримує від клауд-провайдера для запуску мікросервісів.

Мікросервісна архітектура дозволяє краще справлятися з динамічним навантаженням при використанні додатків: якщо навантаження зростає лише на певний мікросервіс, то можна збільшити кількість реплік (подів) лише конкретного мікросервіса, а не горизонтально масштабувати великий моноліт. Завдяки такому більш точному масштабуванні можна скоротити операційні витрати. Нехай S – множина всіх типів мікросервісів, кожен елемент якої характеризується кортежем $(cpu_l, mem_l, t_k^{start}, t_k^{stop}, n_l^{\min}, n_l^{\max})$, де cpu_l, mem_l – ресурси, необхідні для однієї репліки даного типу; t_k^{start}, t_k^{stop} – час запуску та зупинки репліки; $n_l^{\min}, n_l^{\max}$ – мінімальна та максимальна

кількість реплік, що мають бути активними протягом усього часу. Позначимо $n_{l,i}$ як кількість реплік мікросервісу s_l – момент часу t_i, яка у будь-який момент часу має задовольняти умові $n_l^{\min} \le n_{l,i} \le n_l^{\max}$. Запущені в роботу за весь час спостереження репліки (поди) мікросервісів формують множину P, кожен елемент якої характеризується кортежем $(v_k, s_l, t_p^{start}, t_p^{stop})$, де v_k – конкретна віртуальна машина, на якій розміщений под; s_l – клас мікросервіса, до якого належить под; t_p^{start}, t_p^{stop} – моменти запуску та зупинки пода. Кожен под p_p наслідує всі характеристики свого типу s_l. Зазначимо, що розміщення пода на активній віртуальній машині означає використання відповідних ресурсів даної віртуальної машини. Окрім цього, слід зазначити, що у процесі роботи репліки мікросервісів можуть переміщуватися між активними віртуальними машинами: под зупиняється на одній машині та запускається на іншій. Тому для спостереженням даного явища міграції введемо бінарну змінну $x_{p,k,i} \in \{0, 1\}$, яка характеризує, чи розміщений под p_p на віртуальній машині v_k у момент часу t_i. Дані множини дозволяють поєднати наявні ресурси, придбані у вигляді віртуальних машин та їхніх обчислювальних ресурсів, та власне використання цих ресурсів компонентами інфраструктури додатків.

Розглянемо використання ресурсів детальніше: кожен под p_p кожен момент часу використовує певний обсяг ресурсів CPU та оперативної пам'яті. Відповідно, введемо множину метрик подів $PM = \{(cpu_{p,i}, mem_{p,i}) \mid p \in P, t_i \in T\}$, до якої входять спостереження кількості спожитих ресурсів кожною з реплік мікросервісів. Оскільки для прийняття рішень про масштабування використовуються агреговані по мікросервісам метрики, введемо також і множину метрик мікросервісів $SM = \{(cpu_{l,i}, mem_{l,i}) \mid s_l \in S, t_i \in T\}$, де $cpu_{l,i} = \sum_{p \in P \mid s_p = s_l} cpu_{p,i}$ $mem_{l,i} = \sum_{p \in P \mid s_p = s_l} mem_{p,i}$. Дані множини дозволяють спостерігати за поточним навантаженням у вигляді обсягу спожитих ресурсів.

Для вирішення задачі оптимального управління ресурсами мікросервісів та вибору віртуальних машин, необхідно на основі формалізованих множин об'єктів предметної області та їх характеристик розробити її математичну постановку.

Формальна постановка задачі містить обмеження та цільову функцію задачі оптимального управління ресурсами мікросервісів і вибору віртуальних машин, сформованих на основі набору припущень.

Зазначена задача оптимального управління формується на основі прогнозування навантаження по використанню мікросервісів на деякому горизонті прогнозування $\tau_i = \{t_{i+1}, t_{i+2}, \ldots, t_{i+H}\} \subset T$, де H – глибина горизонту. На основі історичних даних метрик SM формується така функція f, що $\forall t_i \in T, \forall s_l \in S$:

$$\widehat{SM}_l(\tau) = (\widehat{SM}_{l,i+1}, \widehat{SM}_{l,i+2}, \ldots, \widehat{SM}_{l,i+H}) = f(SM_{l,i}, SM_{l,i-1}, \ldots, SM_{l,0}), \tag{1}$$

де $SM_{l,i}$ – значення метрик для мікросервісу s_l момент t_i; $\widehat{SM}_{l,i+1}$ – прогноз майбутнього значення для мікросервісу s_l момент t_{i+1}.

Далі визначається кількість реплік мікросервісів, яка необхідна для обробки прогнозованого навантаження на горизонті τ, використовуючи значення прогнозу $\widehat{SM}_l(\tau)$:

$$\widehat{SM}_l(\tau) n_l^*(\tau) = \left\{ \max\left(\frac{\widehat{cpu}_{l,i+1}}{\alpha_l \cdot cpu_l}, \frac{\widehat{mem}_{l,i+1}}{\alpha_l \cdot mem_l} \right) \cdot \sigma_l, \ldots, \max\left(\frac{\widehat{cpu}_{l,i+H}}{\alpha_l \cdot cpu_l}, \frac{\widehat{mem}_{l,i+H}}{\alpha_l \cdot mem_l} \right) \cdot \sigma_l \right\}, \tag{2}$$

де $\alpha_l \in [0, 1]$ – коефіцієнт утилізації, який визначає цільовий рівень використання ресурсів однієї репліки мікросервіса s_l; $\sigma_l \ge 1$ – коефіцієнт запасу на помилку прогнозування.

Даний розрахунок забезпечує, що кількість реплік буде достатньою для обробки прогнозованого навантаження по обом метрикам, коефіцієнт α дозволяє управління ефективністю використання ресурсів, а коефіцієнт σ дозволяє справлятися з неточністю прогнозу.

Фактична кількість реплік $n_l(\tau)$ змінюється з затримкою відносно бажаної кількості $n_l^*(\tau)$ через затримки запуску подів, тому для урахування цього явища використаємо наступне рівняння:

$$\forall t_i \in \tau : n_{l,i} = n_{l,i-1} + a^+_{l,i-d_l^{start}} - a^-_{l,i-d_l^{start}}, \tag{3}$$

де $n_{l,i}$ – фактична кількість реплік мікросервісу s_l у момент t_i; $d_l^{start} = \frac{t_l^{start}}{\Delta t}, d_l^{stop} = \frac{t_l^{stop}}{\Delta t}$ – затримки запуску/зупинки в кроках, викликані часом запуску/зупинки однієї репліки; $a^+_{l,i-d_l^{start}}$ та $a^-_{l,i-d_l^{stop}}$ – кількість реплік мікросервіса типу s_i, рішення про запуск/зупинку яких було прийнято d_l^{start} / d_l^{stop} кроків назад.

Аналогічно враховуються затримки під час запуску чи зупинки віртуальних машин:

$$\forall t_i \in \tau : y_{j,i} = y_{j,i-1} + u^+_{j,i-d^{start}} - u^-_{j,i-d^{stop}}, \tag{4}$$

де $y_{j,i}$ – фактична кількість активних віртуальних машин c_j у момент t_i; $d^{start} = \frac{t^{start}}{\Delta t}, d^{stop} = \frac{t^{stop}}{\Delta t}$ – затримки запуску/зупинки в кроках, викликані часом запуску/зупинки віртуальної машини; $u^{+}_{l,i-d^{start}}$ та $u^{-}_{l,i-d^{stop}}$ – кількість віртуальних машин типу c_j, рішення про запуск/зупинку яких було прийнято d^{start}/d^{stop} кроків назад.

Обмеження рішень про зупинку реплік мікросервісів формулюється як:

$$\forall t_i \in T, \forall s_l \in S : a^{-}_{l,i} \le n_{l,i}. \tag{5}$$

Схожим чином формулюється обмеження рішень про зупинку віртуальних машин:

$$\forall t_i \in T, \forall c_j \in C : u^{-}_{j,i} \le y_{j,i}. . \tag{6}$$

аланс розміщення подів має гарантувати, що їхнє розміщення збігається з загальною кількістю реплік мікросервісів:

$$\forall t_i \in T, \forall s_l \in S : \sum_j x_{p,k,i} = n_{l,i}. \tag{7}$$

Ресурсні обмеження покликані гарантувати, що система не використовує більше обчислювальних ресурсів, ніж насправді існує:

$$\forall t_i \in T, \forall s_l \in S, \forall c_j \in C : \sum_l n_{l,j,i} \cdot cpu_l \le y_{j,i} \cdot cpu_j . \tag{8}$$

$$\forall t_i \in T, \forall s_l \in S, \forall c_j \in C : \sum_j n_{l,j,i} \cdot mem_l \le y_{j,i} \cdot mem_j. \tag{9}$$

Обмеження покриття попиту для задачі пакування часто формулюється через м'яке обмеження для того, аби залишити задачу розв'язною, тож обмеження можна сформулювати наступним чином:

$$\forall t_i \in T, \forall s_l \in S : n_{l,i} = n^{*}_{l,i} - e^{-}_{l,i} + e^{+}_{l,i}, \tag{10}$$

де $e^{-}_{l,i}$ кількість невистачаючих реплік мікросервіса типу s_l у момент t_i; $e^{-}_{l,i}$ кількість надлишкових реплік мікросервіса типу s_l у момент t_i.

Основні характеристики об'єктів мають бути невід'ємними:

$$\forall t_i \in T, \forall s_l \in S, \forall c_j \in C, \forall p_p \in P, \forall v_k \in V : n_{l,i}, y_{j,i}, x_{p,k,i}, a^{\pm}_{l,i}, u^{\pm}_{j,i} \in Z_0. \tag{11}$$

Тепер, коли обмеження моделі було формалізовано, можемо перейти до опису цільової функції, яка має зменшити витрати на оренду віртуальних машин та врахувати негативні наслідки, описані в концептуальній моделі, у вигляді штрафів. Отже, цільова функція, що складається з сумарних витрат на оренду віртуальних машин протягом усього часу спостереження, штрафів за запуск/зупинку нових віртуальних машин задачі, штрафів за запуск/зупинку нових реплік мікросервісів та штрафів за переміщення подів між віртуальними машинами, має наступний вид:

$$\min\left(\sum_{i,j} \cos t_j \cdot y_{j,i} + \sum_{i,j} \left(\lambda^{+} u^{+}_{j,i} + \lambda^{-} u^{-}_{j,i} \right) + \sum_{i,l} \left(\omega^{+} e^{+}_{l,i} + \omega^{-} e^{-}_{l,i} \right) + + \varphi \sum_{p,k,i} \left| x_{p,k,i} - x_{p,k,i-1} \right| \right). . \tag{12}$$

Обмеження (3)–(11) та цільова функція (12) сформовані на основі припущень:

– всі часові параметри системи, такі як час запуску чи зупинки репліки мікросервіса або віртуальної машини, є цілими та кратними Δt;

– для всіх типів віртуальних машин час запуску і зупинки є однаковим;

– вартість використання віртуальної машини певного типу є сталою для всіх класів протягом усього періоду спостереження;

– під час розрахунку кількості поточних реплік мікросервіса, здатних обробляти навантаження, враховуються лише ті репліки, рішення про зупинку яких ще не було прийнято;

– під час розрахунку кількості активних віртуальних машин, на яких можуть бути розміщені поди, враховуються лише ті віртуальні машини, рішення про зупинку яких ще не було прийнято.

Висновки

Таким чином, в роботі запропоновано підхід до моделювання прийняття рішень по управлінню ресурсами мікросервісів у середовищі Kubernetes та динамічному вибору віртуальних машин, який базується на прогнозуванні навантаження, динамічному моделюванні та bin packing. Виконано формалізацію основних характеристик предметної області з використанням математичного апарату теорії множин, та розроблено формальну постановку задачі оптимального управління ресурсами мікросервісів і вибору віртуальних машин.

Запропонована модель дозволяє прогнозувати майбутнє навантаження на систему, визначати необхідні кроки масштабування, аналізувати їхній вплив з точки зору операційних витрат та приймати найбільш ефективні

рішення по вибору віртуальних машин, необхідних для розгортки реплік мікросервісів. Результати роботи можуть бути використані у подальших дослідженнях та реальних імплементаціях розробленої моделі у програмне забезпечення, що дозволить покращити рівень обслуговування клієнтів додатків, побудованих на мікросервісній архітектурі та скоротити операційні витрати завдяки проактивному масштабуванню і ефективнішому використанню ресурсів клауд-провайдерів.

Список використаної літератури

1. Kubernetes. Horizontal Pod Autoscaling. URL: https://kubernetes.io/docs/tasks/run-application/horizontal-pod-autoscale/ (дата звернення: 03.07.2025).

2. Федоришин Б., Красько О. (2024) Міграція сервісів в кластері Kubernetes на основі прогнозування навантаження. *Інфокомунікаційні технології та електронна інженерія*, випуск 4, номер 2, сторінки 82–92. URL: https://doi.org/10.23939/ictee2024.02.082 (дата звернення: 09.07.2025).

3. Сімакін С., Божуха Л. (2024) Прогнозування навантаження на сервер з використанням ШІ для оптимізації веб–сервісів. *Актуальні проблеми автоматизації та інформаційних технологій*, номер 28, сторінки. 234–243. URL: http://doi.org/10.15421/432422 (дата звернення: 12.07.2025).

4. Snehal Chaflekar, Rajendra Rewatkar. (2025) Novel load prediction in microservice architecture using attention mechanism-based deep LSTM networks. *International Journal of Innovative Research and Scientific Studies*, vol. 8, no. 3, pp. 1046–1058. URL: https://doi.org/10.53894/ijirss.v8i3.6751 (дата звернення: 15.07.2025).

5. Гутман Д., Сирота О. (2023) Проактивне автоматичне масштабування вверх для Kuberneters. *Адаптивні системи автоматичного управління*, номер 1, сторінки 32-38. URL: https://doi.org/10.20535/1560-8956.42.2023.278925 (дата звернення: 23.07.2025).

6. Wei-Kuang Lai, You-Chiun Wang, Syu-Chen Wei. (2023) Delay-Aware Container Scheduling in Kubernetes. *IEEE Internet of Things Journal*, vol. 10, no. 13, pp. 11813–11824. URL: https://doi.org/10.1109/JIOT.2023.3244545 (дата звернення: 01.08.2025).

7. Kubernetes. Resource Bin Packing. URL: https://kubernetes.io/docs/concepts/scheduling-eviction/resource-bin-packing/ (дата звернення: 07.08.2025).

8. Rodriguez, M. A., & Buyya, R. (2018) Containers Orchestration with Cost-Efficient Autoscaling in Cloud Computing Environments. *ArXiv*, abs/1812.00300. URL: https://doi.org/10.48550/arXiv.1812.00300 (дата звернення: 15.08.2025).

9. Guruge PB and Priyadarshana YHPP. (2025) Time series forecasting-based Kubernetes autoscaling using Facebook Prophet and Long Short-Term Memory. Frontiers in Computer Science, vol. 7. URL: https://doi.org/10.3389/fcomp.2025.1509165 (дата звернення: 19.08.2025).

10. Maiyza, A. I., Hassan, H. A., Sheta, W. M. (2025) VTGAN based proactive VM consolidation in cloud data centers using value and trend approaches. *Scientific Reports*, vol. 15, no. 20133. URL: https://doi.org/10.1038/s41598-025-04757-z (дата звернення: 27.08.2025).

References

1. Kubernetes (n.d.). Horizontal Pod Autoscaling. Kubernetes. Retrieved from: https://kubernetes.io/docs/tasks/run-application/horizontal-pod-autoscale/ (accessed 03 July 2025).

2. Fedoryshyn B., Krasko O. (2024) Mighracija servisiv v klasteri Kubernetes na osnovi proghnozuvannja navantazhennja [Migration of services in a kubernetes cluster based on workload forecasting]. *Infocommunication technologies and electronic engineering journal*, vol. 4, no. 2, pp. 82-92. Retrieved from: https://doi.org/10.23939/ictee2024.02.082 (accessed 09 July 2025).

3. Simakin S., Bozhukha L. (2024) Proghnozuvannja navantazhennja na server z vykorystannjam ShI dlja optymizaciji veb–servisiv. [Server load forecasting using AI for web services optimization]. *Aktualjni problemy avtomatyzaciji ta informacijnykh tekhnologhij* [Current problems of automation and information technologies] № 28, pp. 234–243. Retrieved from: http://doi.org/10.15421/432422 (accessed 12 July 2025).

4. Snehal Chaflekar, Rajendra Rewatkar. (2025) Novel load prediction in microservice architecture using attention mechanism-based deep LSTM networks. *International Journal of Innovative Research and Scientific Studies*, vol. 8, no. 3, pp. 1046–1058. Retrieved from: https://doi.org/10.53894/ijirss.v8i3.6751 (accessed 15 July 2025).

5. Gutman D., Sirota O. (2023) Proaktyvne avtomatychne masshtabuvannja vverkh dlja Kuberneters [Proactive Upstream Autoscaling for Kubernetes]. *Adaptyvni systemy avtomatychnogho upravlinnja* [Adaptive Automatic Control Systems] no. 1, pp. 32–38. Retrieved from: https://doi.org/10.20535/1560-8956.42.2023.278925 (accessed 23 July 2025).

6. Wei-Kuang Lai, You-Chiun Wang, Syu-Chen Wei. (2023) Delay-Aware Container Scheduling in Kubernetes. *IEEE Internet of Things Journal*, vol. 10, no. 13, pp. 11813 – 11824. Retrieved from: https://doi.org/10.1109/JIOT.2023.3244545 (accessed 01 August 2025).

7. Kubernetes (n.d.). Resource Bin Packing. Kubernetes. Retrieved from: https://kubernetes.io/docs/concepts/scheduling-eviction/resource-bin-packing/ (accessed 07 August 2025).

8. Rodriguez, M. A., & Buyya, R. (2018) Containers Orchestration with Cost-Efficient Autoscaling in Cloud Computing Environments. ArXiv, abs/1812.00300. Retrieved from: https://doi.org/10.48550/arXiv.1812.00300 (accessed 15 August 2025).

9. Guruge PB and Priyadarshana YHPP. (2025) Time series forecasting-based Kubernetes autoscaling using Facebook Prophet and Long Short-Term Memory. *Frontiers in Computer Science*, vol. 7. Retrieved from: https://doi.org/10.3389/fcomp.2025.1509165 (accessed 19 August 2025).

10. Maiyza, A. I., Hassan, H. A., Sheta, W. M. (2025) VTGAN based proactive VM consolidation in cloud data centers using value and trend approaches. *Scientific Reports* (electronic journal), vol. 15, no. 20133. Retrieved from: https://doi.org/10.1038/s41598-025-04757-z (accessed 27 August 2025).

Дата першого надходження рукопису до видання: 29.09.2025
Дата прийнятого до друку рукопису після рецензування: 23.10.2025
Дата публікації:

УДК 004.932.2:004.89:616-073.75 DOI

Є. Б. ЯВОРСЬКА
кандидат технічний наук, доцент,
завідувачка кафедри біотехнічних систем
Тернопільський національний технічний університет імені Івана Пулюя
ORCID: 0000-0001-6341-1710

І. О. ГРИНЮК
аспірант кафедри біотехнічних систем
Тернопільський національний технічний університет імені Івана Пулюя
ORCID: 0009-0002-6204-8707

ЗАСТОСУВАННЯ НЕЙРОМЕРЕЖНИХ ТЕХНОЛОГІЙ ДЛЯ ПІДВИЩЕННЯ ІНФОРМАТИВНОСТІ МЕДИЧНИХ ЗОБРАЖЕНЬ

У роботі розглянуто сучасні підходи до використання штучних нейронних мереж для підвищення якості та інформативності медичних зображень у системах інтраскопічної діагностики, зокрема представлено метод та засіб підвищення якості стоматологічних інтраскопічних зображень на основі інтеграції класичних алгоритмів цифрової обробки та глибинних нейронних мереж. Окрему увагу приділено аналізу згорткових нейронних мереж (CNN), архітектури U-Net та генеративних змагальних мереж (GAN). Проведено огляд переваг та недоліків їх застосування у медичних задачах. Актуальність дослідження зумовлена потребою у точнішій візуалізації дрібних структур зубів і тканин ротової порожнини, що є критично важливим для ранньої діагностики стоматологічних захворювань.

Запропонований підхід включає три ключові етапи: попередню обробку зображень (усунення шумів та нормалізація яскравості), реконструкцію даних за допомогою архітектури U-Net, а також оцінювання результатів за об'єктивними метриками (PSNR, SSIM) та суб'єктивними експертними оцінками стоматологів. Наведено результати моделювання роботи прототипу нейромережного методу, спрямованого на зменшення шуму та підвищення контрастності зображень.

Експериментальні дослідження продемонстрували, що запропонований метод забезпечує підвищення PSNR на 15–20 % та покращення SSIM у середньому на 0,1–0,15 у порівнянні з класичними методами фільтрації.

Отримані результати підтверджують ефективність використання комбінованого підходу для обробки стоматологічних інтраскопічних зображень. Метод дозволяє зберегти дрібні анатомічні деталі, знизити вплив шумів і артефактів та підвищити інформативність зображень, що у свою чергу створює передумови для більш точної клінічної діагностики та підтримки прийняття рішень у стоматологічній практиці.

Ключові слова: *інтраскопічна діагностика, стоматологічні зображення, штучні нейронні мережі, U-Net, GAN, обробка медичних зображень, підвищення якості зображень, цифрова обробка сигналів, інформативність медичних зображень, комп'ютерне бачення у медицині.*

E. B. YAVORSKA
Candidate of Technical Sciences, Associate Professor,
Head of the Department of Biotechnical Systems
Ternopil Ivan Puluj National Technical University
ORCID: 0000-0001-6341-1710

I. O. HRYNIUK
Postgraduate Student at the Department of Biotechnical Systems
Ternopil Ivan Puluj National Technical University
ORCID: 0009-0002-6204-8707

APPLICATION OF NEURAL NETWORK TECHNOLOGIES FOR ENHANCING THE INFORMATIVENESS OF MEDICAL IMAGES

The paper considers modern approaches to the use of artificial neural networks for improving the quality and informativeness of medical images in intrascopic diagnostic systems. In particular, a method and tool are presented for enhancing the quality of dental intrascopic images based on the integration of classical digital processing algorithms and deep neural networks. Special attention is paid to the analysis of convolutional neural networks (CNN), the U-Net architecture, and generative adversarial networks (GAN). An overview of the advantages and limitations of their application

© Яворська Є. Б., Гринюк І. О., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

in medical tasks is provided. The relevance of the study is determined by the need for more accurate visualization of fine dental and oral tissue structures, which is critically important for the early diagnosis of dental diseases.

The proposed approach consists of three key stages: preliminary image preprocessing (noise reduction and brightness normalization), data reconstruction using the U-Net architecture, and evaluation of the results through objective metrics (PSNR, SSIM) as well as subjective expert assessments by dentists. The results of prototype modeling of the neural network method aimed at noise reduction and image contrast enhancement are presented. Experimental studies demonstrated that the proposed method increases PSNR by 15–20 % and improves SSIM by an average of 0.1–0.15 compared to classical filtering methods.

The obtained results confirm the effectiveness of the combined approach for processing dental intrascopic images. The method allows preserving fine anatomical details, reducing the influence of noise and artifacts, and increasing image informativeness, which in turn creates prerequisites for more accurate clinical diagnosis and decision support in dental practice.

Key words: *intrascopic diagnostics, dental images, artificial neural networks, U-Net, GAN, medical image processing, image quality enhancement, digital signal processing, medical image informativeness, computer vision in medicine.*

Постановка проблеми

Інтраскопічна діагностика є важливим інструментом сучасної медицини, що дає змогу лікарям отримувати зображення внутрішніх органів без хірургічного втручання. Проте якість таких зображень часто страждає через низьку роздільну здатність сенсорів, неоднорідне освітлення та шумові артефакти. Це знижує інформативність дослідження й може призводити до помилкових діагнозів. Тому, розвиток технологій штучного інтелекту, зокрема глибокого навчання, відкрив нові можливості для обробки медичних зображень.

Аналіз останніх досліджень і публікацій

Проблема підвищення якості медичних зображень у системах інтраскопічної діагностики вже тривалий час перебуває у центрі уваги науковців. Традиційні підходи до цифрової обробки зображень – фільтрація шумів, покращення контрастності, методи деконволюції – докладно розглянуті у працях українських та закордонних дослідників [1; 2]. Однак ефективність таких методів є обмеженою через значні спотворення сигналу та складність морфології тканин.

Актуальним напрямом стало впровадження методів машинного навчання та глибинних нейронних мереж у медичну діагностику. У роботах Ronneberger та співавт. [4] було запропоновано архітектуру U-Net, яка продемонструвала високу ефективність у задачах сегментації біомедичних зображень. Подальші дослідження довели перспективність застосування генеративних змагальних мереж (GAN) для відновлення структурних деталей і придушення шумів [5; 7].

Значний внесок у систематизацію досліджень з аналізу медичних зображень за допомогою нейронних мереж зробили Litjens та колеги [6], які виконали масштабний огляд сучасних підходів. Shen та співавт. [10] підкреслили потенціал глибинного навчання у покращенні діагностичної інформативності зображень, проте вказали і на низку викликів, пов'язаних із потребою у великих навчальних вибірках та високими обчислювальними витратами.

Особливе місце займають методи оцінки якості оброблених зображень. Показники SSIM та PSNR, запропоновані у класичній роботі Wang та ін. [8], нині залишаються базовими метриками для кількісного аналізу. Новіші підходи до відновлення структурних деталей із використанням глибинних CNN-моделей були розроблені Zhang та співавт. [9], що дало змогу досягти вищої точності порівняно з традиційними алгоритмами фільтрації.

Таким чином, сучасні дослідження підтверджують доцільність поєднання класичних методів обробки із сучасними нейромережними підходами, що забезпечує покращення інформативності та точності інтраскопічних досліджень. У табл. 1 показано відмінності між класичними методами та нейромережними підходами у покращенні якості інтраскопічних зображень.

Аналіз наукових праць показав, що питання підвищення якості медичних зображень у системах інтраскопічної діагностики залишається надзвичайно актуальним. Класичні методи обробки, зокрема фільтрація шумів, деконволюція та покращення контрастності, забезпечують певне підвищення інформативності зображень, проте їхня ефективність суттєво знижується у випадках складних спотворень сигналу та високого рівня шумів.

Сучасні тенденції демонструють значне зростання ролі методів глибинного навчання. Архітектури нейронних мереж (U-Net, GAN, CNN-моделі) показали високу результативність у завданнях сегментації, відновлення структурних деталей та приглушення артефактів. Водночас залишається низка викликів, зокрема потреба у великих навчальних вибірках, значні обчислювальні витрати та проблема інтерпретації результатів.

Таким чином, перспективним напрямом розвитку є інтеграція класичних алгоритмів цифрової обробки зображень із сучасними нейромережними технологіями. Такий підхід дозволяє досягти балансу між обчислювальною ефективністю та високою якістю відновлення інтраскопічних зображень, що має важливе значення для практичних медичних застосувань.

Формулювання мети дослідження

Метою є аналіз і демонстрація можливостей застосування нейромережних методів для підвищення якості інтраскопічних зображень.

Таблиця 1

Порівняльний аналіз методів покращення якості медичних зображень

Критерій	Класичні методи (фільтрація, деконволюція, контрастування)	Нейромережні методи (CNN, U-Net, GAN)
Принцип роботи	Використання математичних операцій для усунення шумів та підвищення контрастності [1; 2]	Навчання моделей на великих вибірках для автоматичного виділення ознак [4; 5; 6]
Ефективність	Добре працюють для простих зображень із низьким рівнем шуму	Висока ефективність навіть у складних умовах, зокрема при спотвореннях сигналу
Гнучкість	Обмежена: алгоритми потребують ручного налаштування параметрів	Висока: модель адаптується до даних, «вчиться» новим патернам
Обчислювальні ресурси	Невисокі, можуть реалізовуватися у режимі реального часу	Високі: потребують GPU та значних обчислювальних потужностей [10]
Інтерпретованість результатів	Висока прозорість, легко пояснити дію алгоритму	Складно інтерпретувати внутрішні механізми роботи нейромереж
Можливості розвитку	Подальше удосконалення має обмежений потенціал	Інтенсивний розвиток, поява нових архітектур (U-Net, GAN, трансформери) [4; 5; 7; 9]

Викладення основного матеріалу дослідження

У межах дослідження для перевірки ефективності запропонованого методу використано вибірку стоматологічних інтраскопічних зображень. Ці дані мають низку особливостей: наявність високочастотних шумів, спричинених відблисками від емалі зубів, неоднорідне освітлення ротової порожнини, а також варіативність форм і текстур тканин.

На першому етапі здійснено попередню цифрову обробку:

– видалення шумів за допомогою медіанного та адаптивного фільтрів;
– вирівнювання гістограми яскравості для усунення проблеми локальних засвітів;
– приведення зображень до єдиного масштабу 256 × 256 пікселів для подальшого навчання нейромережі.

Другим етапом є глибинна реконструкція із застосуванням архітектури U-Net. Навчання проводилося на вибірці зі 1000 стоматологічних зображень, які були анонімізовані та підготовлені експертами-стоматологами. Для збільшення розмірності даних використовувалося data augmentation (обертання, масштабування, зміна контрастності).

Третій етап – оцінка результатів. Порівняння проводилося між:

1. Класичними методами фільтрації (медіанний, Гаусівський фільтр).
2. Сучасними нейромережними підходами (ResNet, GAN).
3. Запропонованим методом (U-Net із попередньою нормалізацією).

Таблиця 2

Порівняння результатів обробки стоматологічних інтраскопічних зображень

Метод	PSNR (dB) ↑	SSIM ↑	Візуальна оцінка експертів*
Медіанний фільтр	24,5	0,72	Артефакти залишаються, часткова втрата дрібних деталей
Гаусівський фільтр	25,1	0,75	Рівномірність зображення краща, але розмитість структур
ResNet	27,8	0,81	Чіткість вища, але нестабільність при високому шумі
GAN	28,3	0,83	Добре відновлює контраст, проте іноді виникають артефакти генерації
Запропонований метод (U-Net)	**29,7**	**0,87**	Висока чіткість, збереженість дрібних анатомічних деталей

* – за результатами експертного оцінювання групою з 5 лікарів-стоматологів (за 5-бальною шкалою).

Отримані результати свідчать, що запропонований метод забезпечує суттєве підвищення PSNR та SSIM у порівнянні з класичними методами, а також перевершує сучасні нейромережні підходи за рахунок інтеграції попередньої нормалізації даних. Особливо важливим є те, що стоматологи-експерти відзначили кращу візуалізацію дрібних структур, що може мати вирішальне значення у діагностиці мікротріщин, каріозних уражень та оцінці стану реставрацій.

Таким чином, метод показав високу ефективність для підвищення інформативності стоматологічних зображень і може бути інтегрований у клінічну практику для покращення точності діагностики.

Висновки

У даній роботі розглянуто актуальну проблему підвищення якості та інформативності медичних інтраскопічних зображень із застосуванням сучасних нейромережних технологій. Запропоновано метод, що поєднує класичні алгоритми цифрової обробки з глибинними нейронними мережами, зокрема архітектурою U-Net, для ефективного зменшення шумів, нормалізації яскравості та відновлення дрібних анатомічних структур.

Проведений аналіз показав, що використання глибинного навчання дозволяє суттєво покращити якісні показники оброблених зображень. За результатами експериментів, метод забезпечує підвищення PSNR на 15–20 % та

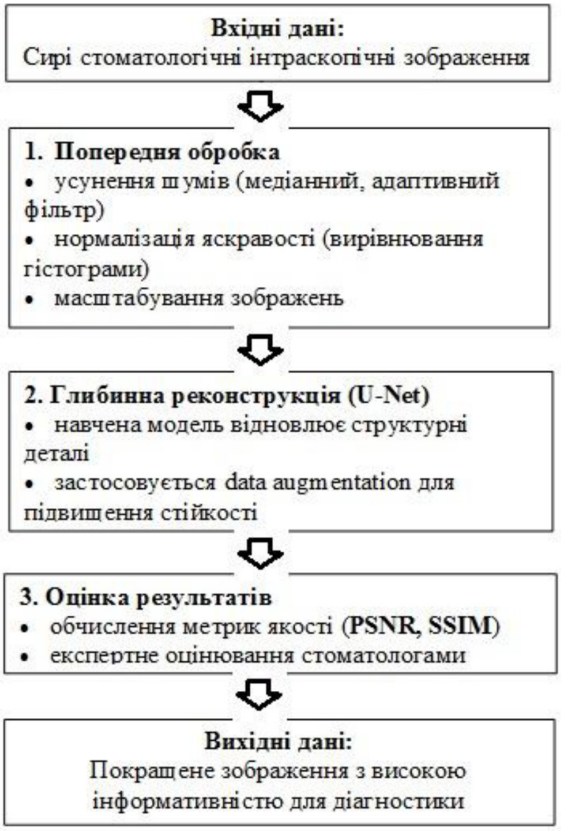


Рис. 1. Схема роботи запропонованого методу підвищення якості стоматологічних інтраскопічних зображень

Рис. 2. Приклади результатів обробки стоматологічних інтраскопічних зображень (зліва – оригінальне зображення з шумом, посередині – після класичного фільтру, справа – після обробки запропонованим методом)

покращення SSIM на 0,1–0,15 порівняно з традиційними методами фільтрації, що підтверджує його ефективність для медичних задач.

Отримані результати демонструють доцільність інтеграції штучних нейронних мереж у системи інтраскопічної діагностики, зокрема в стоматологічній практиці. Це створює передумови для більш точної та ранньої діагностики захворювань зубів і тканин ротової порожнини, а також сприяє розвитку інтелектуальних медичних інформаційних систем.

Перспективи подальших досліджень пов’язані з удосконаленням архітектур нейронних мереж, застосуванням генеративних моделей (GAN) для реконструкції зображень, а також розробкою програмно-апаратних рішень, здатних забезпечити обробку даних у режимі реального часу.

Список використаної літератури

1. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional Networks for Biomedical Image Segmentation // Medical Image Computing and Computer-Assisted Intervention (MICCAI). 2015. Vol. 9351. P. 234–241. DOI: 10.1007/978-3-319-24574-4_28
2. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. Generative Adversarial Nets // Advances in Neural Information Processing Systems. 2014. Vol. 27. P. 2672–2680.
3. Litjens G., Kooi T., Bejnordi B. E., Setio A. A. A., Ciompi F., Ghafoorian M., van der Laak J. A. W. M., van Ginneken B., Sánchez C. I. A survey on deep learning in medical image analysis // Medical Image Analysis. 2017. Vol. 42. P. 60–88. DOI: 10.1016/j.media.2017.07.005
4. Shen D., Wu G., Suk H. I. Deep Learning in Medical Image Analysis // Annual Review of Biomedical Engineering. 2017. Vol. 19. P. 221–248. DOI: 10.1146/annurev-bioeng-071516-044442
5. Lundervold A. S., Lundervold A. An overview of deep learning in medical imaging focusing on MRI // Zeitschrift für Medizinische Physik. 2019. Vol. 29(2). P. 102–127. DOI: 10.1016/j.zemedi.2018.11.002
6. Іваненко С. В., Коваленко О. П. Методи цифрової обробки медичних зображень : навч. посібник. Київ : КНУ ім. Т. Шевченка, 2020. 212 с.
7. Петренко В. М., Руденко І. Ю. Використання глибинних нейронних мереж для підвищення якості медичних зображень // Вісник ХНТУ. 2021. № 2. С. 45–52.
8. Litvinova T., Marchenko A., Klymenko V. Application of convolutional neural networks for dental image classification // Ukrainian Journal of Radiology. 2022. Vol. 30(3). P. 87–95. DOI: 10.32718/ujr202230308
9. Zhang Y., Jiang J., Chen D., Lv S., Yang Y. Denoising medical images using generative adversarial networks // Biomedical Signal Processing and Control. 2021. Vol. 68. P. 102595. DOI: 10.1016/j.bspc.2021.1025951
10. Xu Y., Hosny A., Zeleznik R., Parmar C., Coroller T. P., Franco I., Mak R. H., Aerts H. Deep Learning Predicts Lung Cancer Treatment Response from Serial Medical Imaging // Clinical Cancer Research. 2019. Vol. 25(11). P. 3266–3275. DOI: 10.1158/1078-0432.CCR-18-2495

References

1. Ronneberger O., Fischer P., Brox T. (2015) U-Net: Convolutional networks for biomedical image segmentation. Medical Image Computing and Computer-Assisted Intervention (MICCAI), vol. 9351, pp. 234–241. DOI: 10.1007/978-3-319-24574-4_28
2. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair S., Courville A., Bengio Y. (2014) Generative adversarial nets. Advances in Neural Information Processing Systems, vol. 27, pp. 2672–2680.
3. Litjens G., Kooi T., Bejnordi B. E., Setio A. A. A., Ciompi F., Ghafoorian M., van der Laak J. A. W. M., van Ginneken B., Sánchez C. I. (2017) A survey on deep learning in medical image analysis. Medical Image Analysis, vol. 42, pp. 60–88. DOI: 10.1016/j.media.2017.07.005
4. Shen D., Wu G., Suk H. I. (2017) Deep learning in medical image analysis. Annual Review of Biomedical Engineering, vol. 19, pp. 221–248. DOI: 10.1146/annurev-bioeng-071516-044442
5. Lundervold A. S., Lundervold A. (2019) An overview of deep learning in medical imaging focusing on MRI. Zeitschrift für Medizinische Physik, vol. 29(2), pp. 102–127. DOI: 10.1016/j.zemedi.2018.11.002
6. Ivanenko S. V., Kovalenko O. P. (2020) Metody tsyfrovoi obrobky medychnykh zobrazhen [Methods of digital processing of medical images]. Kyiv: Taras Shevchenko National University of Kyiv, 212 p. (in Ukrainian).
7. Petrenko V. M., Rudenko I. Yu. (2021) Vykorystannia hlybynnykh neironnykh merezh dlia pidvyshchennia yakosti medychnykh zobrazhen [Application of deep neural networks for improving the quality of medical images]. Visnyk KhNTU, no. 2, pp. 45–52. (in Ukrainian).
8. Litvinova T., Marchenko A., Klymenko V. (2022) Application of convolutional neural networks for dental image classification. Ukrainian Journal of Radiology, vol. 30(3), pp. 87–95. DOI: 10.32718/ujr202230308
9. Zhang Y., Jiang J., Chen D., Lv S., Yang Y. (2021) Denoising medical images using generative adversarial networks. Biomedical Signal Processing and Control, vol. 68, p. 102595. DOI: 10.1016/j.bspc.2021.102595
10. Xu Y., Hosny A., Zeleznik R., Parmar C., Coroller T. P., Franco I., Mak R. H., Aerts H. (2019) Deep learning predicts lung cancer treatment response from serial medical imaging. Clinical Cancer Research, vol. 25(11), pp. 3266–3275. DOI: 10.1158/1078-0432.CCR-18-2495

Дата першого надходження рукопису до видання: 20.09.2025
Дата прийнятого до друку рукопису після рецензування: 14.10.2025
Дата публікації:

СОЦІАЛЬНІ ТА ПОВЕДІНКОВІ НАУКИ

УДК 338.516.5:659.1 DOI

О. В. КРИВДА
кандидат економічних наук, доцент,
доцент кафедри економіки і підприємництва
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
ORCID: 0000-0003-4398-6298

С. О. КИРИЧЕНКО
кандидат економічних наук, доцент,
доцент кафедри економіки і підприємництва
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
ORCID: 0000-0001-6195-5204

О. О. ВЕЧКІС
студентка
Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
ORCID: 0009-0008-4185-4527

ЯКІСНИЙ АНАЛІЗ РИЗИКІВ РЕКЛАМНОЇ ГАЛУЗІ

Стаття присвячена дослідженню управління ризиками в рекламній галузі, підкреслюючи її важливість для економіки та конкурентоздатності підприємств. В роботі розглядаються наслідки неефективної роботи рекламних підприємств та актуальність аналізу ризиків. Метою є обґрунтування теоретичних положень та розробка практичних рекомендацій на основі якісного аналізу ризиків рекламної галузі. Підкреслена важливість аналізу, своєчасного реагування та врахування ризиків для забезпечення ефективності діяльності підприємств рекламної галузі. Підкреслюється, що активна мінімізація ризиків та своєчасне втручання є важливими факторами для досягнення стабільних показників діяльності та забезпечення надійного управління в умовах невизначеності. Розглянуто особливості діяльності рекламних агентств, їх функції та роль у рекламних кампаніях. Визначено, що діяльність рекламних агентств, як і будь-яка підприємницька діяльність, пов'язана з різноманітними ризиками. Проведена класифікація ризиків компаній на зовнішні та внутрішні. До зовнішніх ризиків відносяться політичні, юридичні, ризики, пов'язані з глобальними кризами, інфляційні та валютні ризики. Внутрішні ризики включають відсутність інновацій, недостатній контроль за дебіторською заборгованістю, ресурсні ризики та неадекватне управління фінансами. Також розглядаються особливості діяльності рекламних агентств та класифікація ризиків на зовнішні (політичні, юридичні, глобальні кризи, інфляційні, валютні) та внутрішні (відсутність інновацій, контроль заборгованості, ресурсні, фінансові).

На прикладі українського рекламного агентства проаналізовано конкретні ризики компанії (падіння попиту, втрата доходів, невідповідність контенту, залежність від персоналу), фактори ризику, наслідки та заходи щодо їх мінімізації. Обґрунтовано застосування превентивних заходів для управління ризиками в рекламній галузі.

Ключові слова: *ризик, управління ризиками, рекламна діяльність, рекламне агентство, мінімізація ризиків, ефективність.*

O. V. KRYVDA
PhD in Economics, Associate Professor,
Associate Professor at the Department of Economics and Entrepreneurship
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"
ORCID: 0000-0003-4398-6298

© Кривда О. В., Кириченко С. О., Вечкіс О. О., 2025
Стаття поширюється на умовах ліцензії CC BY 4.0

S. O. KYRYCHENKO
PhD in Economics, Associate Professor,
Associate Professor at the Department of Economics and Entrepreneurship
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"
ORCID: 0000-0001-6195-5204

O. O. VECHKIS
Student
National Technical University of Ukraine
"Igor Sikorsky Kyiv Polytechnic Institute"
ORCID: 0009-0008-4185-4527

QUALITATIVE RISKS ANALYSIS OF THE ADVERTISING INDUSTRY

The article is devoted to the study of the problem of risk management in the advertising industry, which is an important sector of the economy that affects the competitiveness of enterprises and overall economic development. The importance of analysis, timely response and consideration of risks is emphasized to ensure the efficiency of the activities of enterprises in the advertising industry. Particular attention is paid to a structured approach to risk management, which includes the stages of identification, assessment, analysis and monitoring. It is emphasized that active risk minimization and timely intervention are important factors for achieving stable performance indicators and ensuring reliable management in conditions of uncertainty. The features of the activities of advertising agencies, their functions and role in advertising campaigns are considered. It is determined that the activities of advertising agencies, like any entrepreneurial activity, are associated with various risks. The classification of company risks into external and internal is carried out. External risks include political, legal, risks associated with global crises, inflationary and currency risks. Internal risks include lack of innovation, insufficient control over receivables, resource risks and inadequate financial management. To illustrate the impact of risks on the activities of advertising agencies, the article provides an example of a risk analysis of Domio Publishing LLC. The article considers such risks as a decline in demand for printed publications, loss of advertising revenue, content mismatch with audience expectations and dependence on key personnel. For each risk, factors, possible consequences and causes are identified, and measures are proposed to minimize them. The use of preventive measures for risk management in the advertising industry is justified. The need for flexibility, adaptability and prompt response to changes in the media environment is emphasized.

Key words: *risk, risk management, advertising activities, advertising agency, risk minimization, efficiency.*

Постановка проблеми

Рекламна діяльність є важливим інструментом для активізації національної економіки, стимулюючи ділову активність компаній та посилюючи конкуренцію на ринку. Проте, якщо підприємства, які надають рекламні послуги, працюють неефективно, це може призвести до зниження конкурентоздатності товарів рекламодавців, послабити їхні позиції на ринку та ускладнити протистояння іноземним конкурентам. Ризики, пов'язані з діяльністю підприємств в цій галузі, неодмінно і суттєво впливають на економічні результати як самих підприємств, так і пов'язаних з ними інших суб'єктів господарської діяльності.

На сучасному етапі для ефективної діяльності підприємства важливо знаходити більш оптимальні способи, які спрощуватимуть діяльність підприємців та підвищуватимуть економічну ефективність господарюючих суб'єктів. За допомогою якісного аналізу ризиків та загроз можна простежити конкретні особливості того чи іншого ризику, порівняти різні ризики, проаналізувати фактори та причини їх виникнення. Якісний аналіз структурує дані, дає додаткову інформацію про об'єкт дослідження, що набагато спрощує подальше управління ним. Саме тому якісний аналіз є важливим аспектом у роботі з ризиками та загрозами на підприємствах будь-якої сфери діяльності.

Аналіз останніх досліджень і публікацій

Питання ризиків у господарській діяльності та їхньої нейтралізації досліджувалися багатьма авторами. Серед них: Макалюк І. [1], Михайленко О. В., Ніколаєнко С. М., Насіканова О. О. [2], О. О. Коць (класифікація ризиків промислових підприємств) [5], Макалюк І. В. та С. М. Наливайко (конкурентоздатність швейних підприємств під час війни) [8] та інші. Водночас, багато аспектів цієї наукової проблеми залишаються дискусійними, недостатньо розкритими та обґрунтованими, а також наукових розробок, які б комплексно розглядали особливості діяльності підприємств рекламної галузі та притаманні їм ризики також бракує. Майже відсутні також наукові дослідження у сфері рекламних послуг.

Формулювання мети дослідження

Метою статті є обґрунтування теоретичних положень щодо специфіки та управління ризиками підприємств рекламної галузі, а також надання практичних рекомендацій в контексті якісного аналізу ризиків на прикладі вітчизняного підприємства ТОВ «Доміо Паблішінг».

Викладення основного матеріалу дослідження

У сучасному світі, де торгівля та реклама відіграють вирішальну роль, компанії стикаються з різними видами ризиків, які можуть вплинути на їх ефективність. Через відкритість економіки та тенденції глобалізації, важливість просування товарів на ринки, рекламна діяльність стикається з багатьма викликами, кожний з яких має свої специфічні ризики [1]. Ці ризики можуть мати значні наслідки для діяльності, конкурентоспроможності та прибутковості компаній. Розуміння цих ризиків і управління ними мають важливе значення для компаній, які прагнуть підтримувати та покращувати свою ефективність на глобальних та локальних ринках.

Дане дослідження має на меті детально дослідити, як ризики рекламних компаній впливають на ефективність такого бізнесу, і визначити ефективні стратегії управління ними. Ця тема особливо актуальна в нинішньому контексті глобалізації та зростаючого перевиробництва. Торгівля стосується всіх товарів, послуг і капіталу, якими обмінюються споживачі та інші учасники ринкових відносин. Важливо розуміти рівні впливу різних груп ризиків на компанії, незалежно від того, чи це стосується операційної, економічної чи фінансової діяльності, адже виклики, які створює сучасна торгівля, численні та складні.

Будь-яка підприємницька діяльність пов'язана з постійними ризиками, тому не можна недооцінювати цю складову успішного управління бізнесом. Важливо постійно аналізувати ситуацію, мати стратегію управління ризиками і бути готовим до непередбачених обставин.

Вплив ризиків на ключові показники ефективності підприємств є беззаперечним. Тому аналіз, швидка реакція та врахування ризиків в управлінні підприємством є життєво необхідними, адже їхнє ігнорування може призвести до банкрутства. Хоча ризики переважно несуть загрозу втрати прибутку, вони також можуть відкривати можливості для його збільшення, що є суттю підприємницького ризику. Набір та рівень ризиків для кожного підприємства індивідуальні та визначаються його розміром, капіталом, продукцією, ринками збуту, сферою діяльності тощо. Тому для зменшення їхнього впливу на діяльність господарюючих суб'єктів необхідно будувати систему управління можливими наслідками на кожному підприємстві.

Ефективне управління ризиками в контексті рекламної діяльності базується на структурованому багатоетапному процесі. У глобальному моніторингу ризиків компанії повинні запровадити механізми виявлення, оцінки, аналізу та контролю. Раннє виявлення ризиків, що виникають у складному бізнес-середовищі, дозволяє компаніям передбачати потенційні виклики. Ретельна оцінка вимірює вплив і ймовірність кожного виявленого ризику. Детальний аналіз допомагає зрозуміти можливі наслідки та сформулювати відповідні стратегії мінімізації їх негативного впливу. Постійний моніторинг необхідний для коригування підходів на основі змін ринкових умов і зовнішніх факторів. Завдяки активній мінімізації ризиків і своєчасному втручанню компанії мають кращі можливості для покращення загальних показників і забезпечення надійного управління в умовах невизначеності в сучасній торгівлі [2].

Головне завдання даного дослідження – провести якісний аналіз ризиків діяльності рекламних агентств, тож, по-перше, необхідно визначити основні напрями їх діяльності.

Рекламне агентство являє собою самостійну організацію, яка за дорученням рекламодавця виконує творчі та технічні завдання, включаючи планування, розробку та впровадження рекламних кампаній і окремих заходів. Діючи від імені та за кошти своїх клієнтів, рекламні агентства несуть перед ними фінансову, юридичну та етичну відповідальність. Вони налагоджують співпрацю з друкарнями, виробниками та постачальниками матеріалів для реклами, видавництвами та студіями. Крім того, рекламні агентства взаємодіють зі ЗМІ для розміщення реклами, контролюють виконання замовлень та оцінюють ефективність рекламних повідомлень [3].

Рекламні агентства – це компанії з повним обслуговуванням, які здатні керувати кожним аспектом рекламної кампанії. Вони розрізняються за розміром та обсягом і обслуговують різні типи клієнтів. Деякі агентства мають лише один-два великих клієнта, акаунтами яких вони управляють. Інші мають сотні клієнтів, що розповсюджуються по всій країні або по всьому світу та обслуговуються багатьма відділеннями на місцях.

Діяльність рекламних агентств, як і будь-яка підприємницька діяльність, пов'язана з великою кількістю різноманітних ризиків. Згідно загально відомому твердженню, на діяльність підприємств впливають дві групи ризиків: зовнішні та внутрішні. Для початку розглянемо зовнішні ризики та опишемо їх вплив на діяльність рекламних агентств [5].

1. Політичний ризик. Політичний ризик в загальному випадку розуміється як настання будь якої політичної події. Якщо розглядати даний аспект з точки зору впливу на рекламні агентства, даний ризик проявляється у вигляді негативного впливу на фізичну безпеку працівників, інфраструктуру та економічний клімат. Політична нестабільність призводить також до зниження купівельної спроможності населення, а це, в свою чергу, прямо пропорційно впливає на доходи рекламних кампаній.

2. Юридичний ризик. Цей тип ризиків виникає через зміни в законодавстві або недотримання чинних норм, що може суттєво вплинути на роботу підприємства. Нові податкові закони чи зміни у регуляції можуть спричинити зростання витрат або появу нових зобов'язань перед державою. Наприклад, підвищення податкових ставок може збільшити фінансове навантаження на підприємство, зменшуючи його прибуток [6]. З точки зору впливу

на рекламні агентства, нові закони, що стосуються медіа, авторських прав, податків чи інтернет-регулювання, можуть створити додаткові витрати або обмеження для компаній, змінити правила гри на ринку та вплинути на ведення бізнесу.

3. Ризик впливу глобальних криз або пандемій. Даний ризик проявляється у вигляді безлічі негативних факторів, а саме дерегулювання фінансового сектору, втраті можливостей для працевлаштування, зниження попиту на імпорт, платіжний дисбаланс та загальне скорочення інвестицій. Наприклад, фінансова криза 2008-2009 рр. вплинула на соціально-економічний розвиток країн, зокрема на зниження реальних доходів населення і призвело до стримування економічного зростання. Також знизився добробут населення через зменшення можливостей поточного споживання, інвестування, заощадження, а також через знецінення фінансового капіталу та девальвацію національних валют. Ще одним яскравим прикладом є пандемія COVID 19 2020 року. Незважаючи на зниження темпу інфляції, відбулося значне зростання рівня безробіття [7]. Розглядаючи вплив даного ризику саме на діяльність рекламних агентств, слід відмітити, що світові економічні кризи, пандемії чи інші глобальні події можуть змінювати поведінку споживачів і рекламодавців, знижуючи попит на рекламу або змінюючи її формати.

4. Інфляційний та валютний ризики. Інфляційний ризик – це ризик того, що при високому рівні інфляції отримані доходи знецінюються. Валютний ризик – ризик, пов'язаний зі змінами курсу валют [5]. В діяльності рекламних агентств ці ризики матимуть прямий вплив на ефективність функціонування підприємства, адже інфляція та коливання валют можуть спричинити зростання витрат на виробництво та зниження купівельної спроможності населення, а це в свою чергу може погіршити фінансові показники компанії.

На діяльність підприємств окрім зовнішніх, значний вплив мають також і внутрішні ризики, і рекламні агентства також не виключення. Розглянемо цю групу ризиків докладніше.

Відсутність інновацій та адаптації до змін, або технологічний ризик, тобто ризик втрат від використання застарілого обладнання та технологій. Даний ризик відноситься до внутрішніх ризиків та може призвести до значного зниження конкурентоспроможності. Якщо компанія не інвестує в нові технології та не адаптується до змін у медіа-сфері, наприклад, не використовує цифрові платформи або не вдосконалює формат контенту, це може спричинити значне зростання витрат та призведе в результаті до втрати конкурентних позицій на ринку.

Ще один досить важливий внутрішній фінансовий ризик – недостатній контроль за дебіторською заборгованістю. Залишаючи незадовільний контроль за платіжними термінами клієнтів або рекламодавців, компанія може зіткнутися з проблемами ліквідності та необхідністю вирішення правових питань.

Ресурсний ризик. Він полягає в ймовірності браку або в повній відсутності окремих видів економічних ресурсів. Для рекламних агентств даний ризик полягає в нестачі кваліфікованих кадрів і високого рівня залежності від них.

Наступний внутрішній ризик, з яким може стикнутись рекламне агентство, це неадекватне управління фінансами. Неякісне та непрофесійне фінансове планування, надмірні витрати або неправильне управління дебіторською заборгованістю можуть призвести до фінансових труднощів та зниження ліквідності підприємства.

Щоб зробити висновок про конкретний вплив ризиків на діяльність рекламних компаній, розглянемо найголовніші ризики для рекламного агентства у сфері медіа на прикладі ТОВ «Доміо Паблішінг».

ТОВ «Доміо Паблішинг» є провідною українською видавничою компанією, що спеціалізується на випуску преміальних глянцевих видань, насамперед всесвітньо відомого журналу ELLE. Компанія виступає офіційним ліцензіатом цього культового бренду на території України, забезпечуючи читачів якісним контентом у сфері моди, краси, культури та стилю життя [9].

Одним з найважливіших ризиків для видавничої галузі є ризик падіння попиту на друковані видання рекламних агентств (по типу Forbes.ua, Elle, Cosmopolitan, Harpers Bazar тощо). Фактором даного ризику, в першу чергу, виступає зростання популярності цифрових медіа та соціальних мереж, по-друге – відбувається зміна споживчих звичок молодшої аудиторії, по-третє – спостерігається глобальний тренд на скорочення тиражів друкованих видань, а також економічна ситуація в країні, яка безпосередньо впливає на купівельну спроможність споживачів. Як наслідок, скорочення попиту спричинить зниження тиражу та продажів журналу, що, в свою чергу, зменшить рекламні доходи підприємства. Також може виникнути необхідність скорочення штату, або ж у критичному випадку – припинення випуску друкованої версії видань. Серед причин даного ризику можна виділити наступні:

– технологічні зміни та діджиталізація медіа-ринку;
– зручність та доступність онлайн-контенту;
– екологічні тренди (відмова від паперу).

Аби зменшити ефект від даного ризику, першочергово підприємства повинні активно розвивати цифрові версії журналів, чим в цілому ринок медіа займається вже давно. Також було б доцільно розглянути можливість створення окремих додатків для користувачів. Ефективним способом мінімізації ефекту даного ризику є диверсифікація джерел доходів компанії, яка полягає у проведенні рекламних спецзаходів, створенні колаборацій, випуск власної продукції. Не менш важливо оптимізувати усі процеси, пов'язані з випуском друкованих версій та зробити акцент на ефективність управління витратами на виробництво.

Важливим є також ризик втрати рекламних доходів. Факторами цього ризику виступають економічна нестабільність, яка напряму впливає на маркетингові бюджети брендів, зміна маркетингових стратегій рекламодавців (фокус на інфлюенсерів, соціальні мережі) та зменшення охоплення аудиторії друкованими виданнями.

Як наслідок, оскільки для глянцевих видавництв реклама виступає основним джерелом доходу, даний ризик може спричинити критичне зниження рівня доходів підприємства. Також це може спричинити неможливість покриття виробничих витрат, що, в свою чергу, вплине на якість контенту.

Причинами даного ризику є:

– зміни в медіа-споживанні цільової аудиторії;
– через економічну та політичну нестабільність багато великих рекламодавців покинули Український ринок (Dior, Chanel);
– метричність ефективності реклами в цифрових каналах;
– зміна маркетингових пріоритетів люксових та преміум-брендів.

Щоб знизити негативний ефект від настання ризику цього типу, підприємство має розробляти комплексні рекламні пропозиції для рекламодавців (print + digital). Також може позитивно вплинути на попит і допомогти збільшити залученість аудиторії проведення спеціальних проєктів та заходів для рекламодавців. Впровадження нових рекламних форматів, як-от інтерактивні банери або відеореклама на різних цифрових платформах, також знаходиться серед можливих методів уникнення негативних наслідків такого ризику. Важливим є і фокус на клієнтоорієнтованість та індивідуальний підхід, що сприяє формуванню довгострокових партнерських відносин і покращує ефективність рекламних компаній. Слід враховувати, що для рекламодавців дуже важливим є отримання рекламної аналітики, що, в свою чергу, вимагає від підприємства постійного проведення досліджень ефективності рекламних розміщень.

Наступний важливий ризик – ризик невідповідності контенту очікуванням аудиторії. Факторами даного ризику є швидкі зміни в модних трендах та уподобаннях аудиторії, розрив між глобальним контентом журналу ELLE та локальними особливостями ринку, поява нових тем, важливих для цільової аудиторії (екологія, соціальна відповідальність тощо), а також зростання вимог аудиторії до інклюзивності контенту. Серед можливих наслідків варто зазначити, по-перше, втрату журналом лояльності читачів, тобто аудиторія може почати відчувати розчарування від зміни контенту або зниження його якості, що призведе до зменшення зацікавленості в майбутньому випуску видання. По-друге, це може спричинити зниження продажів журналу та втрату конкурентної позиції на ринку, що призведе до зниження привабливості видання як рекламної платформи для рекламодавців, оскільки менша аудиторія – це автоматично менше охоплення та ефективність реклами. Не варто забувати також про можливу втрату репутації, що призведе до зменшення довіри як серед читачів, так і серед рекламодавців, знижуючи привабливість бренду для партнерств у майбутньому.

Причинами даного ризику є:

– відсутність регулярних досліджень аудиторії;
– орієнтація на застарілі уявлення про цільову аудиторію;
– недостатня адаптація глобального контенту до місцевого контексту;
– брак інновацій у редакційній політиці.

Для мінімізації ефекту даного ризику компанія має регулярно проводити дослідження аудиторії, залучати молодих талантів до створення контенту. Також важливо зберігати баланс між глобальним контентом та локальною специфікою. Дуже важливою є активна взаємодія з аудиторією через соціальні мережі для отримання зворотного зв'язку.

Наступний – операційний ризик, який полягає в залежності від ключового персоналу компанії. Факторами даного ризику є висока значимість головного редактора та креативної команди для успіху видання, обмежена кількість фахівців з необхідним досвідом та навичками, висока конкуренція за таланти в медіа-сфері, а також залежність від знань та відносин ключових співробітників з рекламодавцями. Якщо хтось із ключового персоналу покине підприємство, це може вплинути на якість контенту, видання може втратити свій унікальний стиль ведення, що погіршить імідж компанії і знизить довіру з боку читачів та рекламодавців. Втрата ключового персоналу також порушує відносини з важливими партнерами та клієнтами, що може спричинити зниження ефективності бізнес-процесів підприємства та навіть призвести до втрати деяких важливих контрактів або замовлень. Це також не може не привести до додаткових витрат на пошук та навчання нового співробітника, тобто тимчасового зниження продуктивності та додаткових фінансових витрат.

Причини даного ризику наступні:

– персоналізований характер медіа-бізнесу;
– важливість особистих зв'язків та репутації в індустрії моди та глянцю;
– недостатня формалізація бізнес-процесів;
– конкуренція за таланти з боку інших медіа та digital-агенцій.

Аби мінімізувати ефект даного ризику, підприємство повинно постійно розробляти та вдосконалювати програми для утримання ключових співробітників. Також необхідно диверсифікувати повноваження та відповідальність між різними співробітниками, це допоможе зменшити залежність від окремих осіб і забезпечить більш стійку організаційну структуру, що дозволить зберегти ефективність навіть за змін умов. Окрім цього, було б доцільно створити систему наставництва і таким чином підготувати резерв кадрів із залученням молоді. Також допоможе формалізація та документування процесів та знань, що дозволить зберегти критичну базу знань компанії, полегшити інтеграцію нових співробітників та забезпечити безперервність роботи навіть при зміні персоналу.

Висновки

Рекламна індустрія є важливою складовою сучасної економіки, оскільки вона стимулює ділову активність компаній та сприяє розвитку конкуренції на ринку. Ефективна реклама допомагає підприємствам просувати свої товари та послуги, залучати клієнтів та розширювати свою присутність на ринку. Однак, неефективна робота рекламних підприємств може мати негативні наслідки. Це може призвести до зниження конкурентоздатності товарів рекламодавців, послабити їхні позиції на ринку та ускладнити їхню здатність конкурувати з іноземними компаніями. Діяльність рекламних агентств пов'язана з різноманітними ризиками, які можуть вплинути на їхню ефективність та фінансову стабільність. Розуміння цих ризиків та розробка ефективних стратегій управління ними є важливим завданням для забезпечення успішного функціонування рекламної галузі.

Варто зазначити, що в сучасних умовах для забезпечення конкурентоспроможності підприємство повинно застосовувати превентивні дії щодо ризиків та їх можливих наслідків. В даній галузі важливими аспектами є гнучкість, швидка адаптивність та вміння оперативно реагувати на зміни в технологічній та інформаційній сфері.

Таким чином, недостатньо лише виявити потенційні ризики, також необхідно постійно розробляти та вдосконалювати стратегії їх попередження та усунення з метою мінімізації негативних наслідків їх настання. Саме такий підхід забезпечить підприємству сталий розвиток та зростання в динамічному медіа-середовищі.

Список використаної літератури

1. Макалюк І. В., Кривда О. В., Лайкова А. О. Якісний аналіз ризиків вітчизняних підприємств в умовах воєнного стану. *Економіка та суспільство*. 2024. № 62. DOI: https://doi.org/10.32782/2524-0072/2024-62-73

2. Михайленко О. В., Ніколаєнко С. М., Насіканова О. О. Управління ризиками діяльності підприємства. *Проблеми системного підходу в економіці*. 2017. Вип. 6. С. 144–147. URL: http://psae-jrnl.nau.in.ua/ journal/6_62_1_2017_ukr/24.pdf

3. Миронов Ю. Б., Крамар Р. М. Основи рекламної діяльності. URL: https://kerivnyk.info/osnovy-reklamnoi-diyalnosti1-3

4. Мельникович О. М. Сучасні напрями діяльності підприємств у галузі реклами. *Ефективна економіка*. 2016. Вип. 2. URL: http://www.economy.nayka.com.ua/pdf/2_2016/4.pdf

5. Коць О. О. Класифікація ризиків промислового підприємства. *Вісник Національного університету «Львівська політехніка»*. 2006. № 554. С. 96–103.

6. Що таке ризики для підприємства, і як їх контролювати? ЛІГА: ЗАКОН Бізнес. https://biz.ligazakon.net/aktualno/14605_shcho-take-riziki-dlya-pdprimstva-ta-yak-kh-kontrolyuvati (дата звернення 13.05.2025).

7. Король М. М., Петрусь Д. І., Лапа І. В. Дослідження впливу глобальних криз на світову економіку. *Сталий розвиток економіки*. 2023. № 2(47). DOI: https://doi.org/10.32782/2308-1988/2023-47-14

8. Кривда О. В., Макалюк І. В., Наливайко С. М. Ризики забезпечення конкурентоспроможності продукції швейних підприємств в умовах воєнного стану. *Ефективна економіка*. 2024. № 4. DOI: http://doi.org/10.32702/2307-2105.2024.4.74

9. DOMIO PUBLISHING – новий видавець журналів ELLE і ELLE DECORATION в Україні. URL: https://domio.ua/news/domio-publishing-novij-vidavecz-zhurnaliv-elle-i-elle-decoration-v-ukra%d1%97ni (дата звернення 10.05.2025).

References

1. Makaliuk I. V., Kryvda O. V., Laikova A. O. (2024). Yakisnyi analiz ryzykiv vitchyznianykh pidpryiemstv v umovakh voiennoho stanu. [Qualitative analysis of risks of domestic enterprises under martial law]. *Economy and society*, no. 62. DOI: https://doi.org/10.32782/2524-0072/2024-62-73

2. Mykhailenko O. V., Nikolaienko S. M., Nasikanova O. O. (2017). Upravlinnia ryzykamy diialnosti pidpryiemstva. [Enterprise risk management]. *Problems of a systems approach in economics*, no. 6, pp. 144-147. URL: http://psae-jrnl.nau.in.ua/ journal/6_62_1_2017_ukr/24.pdf

3. Myronov Yu. B., Kramar R. M. Osnovy reklamnoi diialnosti. [Basics of advertising]. URL: https://kerivnyk.info/osnovy-reklamnoi-diyalnosti1-3. [in Ukrainian].

4. Melnykovych O. M. (2016). Suchasni napriamy diialnosti pidpryiemstv u haluzi reklamy. [Modern directions of activity of enterprises in the field of advertising]. *Efficient economy*, vol. 2. URL: http://www.economy.nayka.com.ua/pdf/2_2016/4.pdf

5. Kots O. O. (2006). Klasyfikatsiia ryzykiv promyslovoho pidpryiemstva. [Classification of industrial enterprise risks]. *Bulletin of the National University "Lviv Polytechnic"*, no 554, pp. 96–103.

6. Shcho take ryzyky dlia pidpryiemstva, i yak yikh kontroliuvaty? [What are risks for the enterprise, and how to control them?]. LEAGUE: LAW Business. https://biz.ligazakon.net/aktualno/14605_shcho-take-riziki-dlya-pdprimstva-ta-yak-kh-kontrolyuvati. (accessed 10 June 2025).

7. Korol M. M., Petrus D. I., Lapa I. V. (2023). Doslidzhennia vplyvu hlobalnykh kryz na svitovu ekonomiku. [Research into the impact of global crises on the world economy]. *Sustainable economic development*, no. 2(47). DOI: https://doi.org/10.32782/2308-1988/2023-47-14

8. Kryvda O. V., Makaliuk I. V., Nalyvaiko S. M. (2024). Ryzyky zabezpechennia konkurentospromozhnosti produktsii shveinykh pidpryiemstv v umovakh voiennoho stanu. [Risks of ensuring the competitiveness of clothing enterprises' products under martial law]. *Efficient economy*, no. 4. DOI: http://doi.org/10.32702/2307-2105.2024.4.74

9. DOMIO PUBLISHING – novyi vydavets zhurnaliv ELLE i ELLE DECORATION v Ukraini. [DOMIO PUBLISHING is the new publisher of ELLE and ELLE DECORATION magazines in Ukraine]. URL: https://domio.ua/news/domio-publishing-novij-vidavecz-zhurnaliv-elle-i-elle-decoration-v-ukra%d1%97ni. (accessed 02 June 2025).

Дата першого надходження рукопису до видання: 25.09.2025
Дата прийнятого до друку рукопису після рецензування: 22.10.2025
Дата публікації:

ВІСНИК

Херсонського національного технічного університету

Відповідальний за випуск	головний редактор Литвиненко В. І. д.т.н., професор, завідувач кафедри інформатики і комп'ютерних наук, заслужений діяч науки і техніки України
Комп'ютерна верстка	Клименко Т. О.
Відповідальний секретар	Лур'є І. А. к.т.н., доцент, начальник навчально-методичного відділу, доцент кафедри інформатики і комп'ютерних наук

Реєстрація суб'єкта у сфері друкованих медіа: Рішення Національної ради України з питань телебачення і радіомовлення № 1541 від 09.05.2024 року. Ідентифікатор медіа: R30-04565.

Суб'єкт у сфері друкованих медіа – Херсонський національний технічний університет (вул. Інститутська, буд. 11, м. Хмельницький, 29016, olena.chepelyuk@kntu.net.ua, тел. +38 (0382) 77-35-65).

Періодичність виходу: 4 рази на рік.

Підписано до друку 06.06.2025.
Формат 60x84/8. Гарнітура Times New Roman. Папір офсет. Цифровий друк.
Ум. друк. арк. 62,08. Замов. № 1125/879. Наклад 100 прим.

Видавництво і друкарня – Видавничий дім «Гельветика»
65101, Україна, м. Одеса, вул. Інглезі, 6/1
Телефон +38 (095) 934 48 28, +38 (097) 723 06 08
E-mail: mailbox@helvetica.ua
Свідоцтво суб'єкта видавничої справи
ДК № 7623 від 22.06.2022 р.