

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії
(повна назва факультету)

Кафедра комп'ютерних наук
(повна назва кафедри)

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня

магістр

(назва освітнього ступеня)

на тему: Дослідження алгоритмів машинного навчання для потреб систем
підтримки прийняття рішень в умовах пандемії

Виконав: студент VI курсу, групи СНм-61
спеціальності 122 Комп'ютерні науки
(шифр і назва спеціальності)

(підпис)

Фіголь В.В.

(прізвище та ініціали)

Керівник

(підпис)

Гром'як Р.С.

(прізвище та ініціали)

Нормоконтроль

(підпис)

Мацюк О.В.

(прізвище та ініціали)

Завідувач кафедри

(підпис)

Боднарчук І.О.

(прізвище та ініціали)

Рецензент

(підпис)

Стухляк П.Д.

(прізвище та ініціали)

Тернопіль
2022

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет комп'ютерно-інформаційних систем і програмної інженерії
(повна назва факультету)

Кафедра комп'ютерних наук
(повна назва кафедри)

ЗАТВЕРДЖУЮ

Завідувач кафедри

Боднарчук І.О.
(підпис) (прізвище та ініціали)

« 26 » травня 2022 р.

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ**

на здобуття освітнього ступеня Магістр
(назва освітнього ступеня)

за спеціальністю 122 Комп'ютерні науки
(шифр і назва спеціальності)

Студенту Фіголь Віталію Володимировичу
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження алгоритмів машинного навчання для потреб систем підтримки прийняття рішень в умовах пандемії

Керівник роботи Гром'як Роман Сильвестрович, к.ф.-м.н., доцент кафедри КН
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом ректора від « 28 » жовтня 2021 року № 4/7-909

2. Термін подання студентом завершеної роботи 23 травня 2022р.

3. Вихідні дані до роботи Наукові публікації про алгоритми машинного навчання, системи підтримки прийняття рішень, особливості їх функціонування та використання в умовах пандемії

4. Зміст роботи (перелік питань, які потрібно розробити)

Вступ. 1 Стан та перспективи досліджень в галузі алгоритмів машинного навчання та їх використання в умовах пандемії. 2 Обчислювальний експеримент для дослідження алгоритмів машинного навчання в умовах пандемії. 3. Аналіз результатів обчислювального експерименту дослідження алгоритмів машинного навчання в умовах пандемії. 4 Охорона праці та безпека в надзвичайних ситуаціях. Висновки. Додатки

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, слайдів)

1 Титульна сторінка. 2 Тема, Мета, Об'єкт, Предмет дослідження. 3 Завдання дослідження.

4 Актуальність дослідження. 5 COVID-19 Поширення від людини до людини. 6 Блок-схема відбору публікацій щодо досліджень із застосуванням запитів і критеріїв включення.

7 Розподіл виявлених публікацій по наукометричних БД. 8 Класифікація супутніх наукових публікацій. 9 Блок-схема узагальненого конвеєра дослідження алгоритмів машинного

навчання для потреб систем підтримки прийняття рішень в умовах пандемії. 10 Архітектура нейронної мережі для вибору ознак при дослідженні алгоритмів машинного навчання в

умовах пандемії. 11 Відносна важливість початкових розглянутих ознак. 12 Точність досліджуваних алгоритмів машинного навчання. 13 F1-розподіл та AUC досліджуваних

алгоритмів. 14. Результати навчання та функціонування шести різних алгоритмів машинного навчання. 15 Висновки. 16 Висновки. 17 Завершальний слайд.

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Охорона праці	Дмитроца Л.П., доцент		
Безпека в надзвичайних ситуаціях	Клепчик В.М., старший викладач		

7. Дата видачі завдання 28 жовтня 2021 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1.	Ознайомлення з завданням до кваліфікаційної роботи	28.10.2021-31.10.2021	<i>Виконано</i>
2.	Підбір наукових джерел про алгоритми машинного навчання, системи підтримки прийняття рішень, особливості їх функціонування та використання в умовах пандемії	1.11.2021-14.11.2021	<i>Виконано</i>
3.	Переклад та опрацювання наукових джерел про алгоритми машинного навчання, системи підтримки прийняття рішень, особливості їх функціонування та використання в умовах пандемії	15.11.2021-19.12.2021	<i>Виконано</i>
4.	Виконання дослідження алгоритмів машинного навчання для потреб систем підтримки прийняття рішень в умовах пандемії	10.01.2022-30.01.2022	<i>Виконано</i>
5.	Оформлення розділу «Стан та перспективи досліджень в галузі алгоритмів машинного навчання та їх використання в умовах пандемії»	31.01.2022-21.02.2022	<i>Виконано</i>
6.	Оформлення розділу «Обчислювальний експеримент для дослідження алгоритмів машинного навчання в умовах пандемії»	14.03.2022-27.03.2022	<i>Виконано</i>
7.	Оформлення розділу «Аналіз результатів обчислювального експерименту дослідження алгоритмів машинного навчання в умовах пандемії»	28.03.2022-03.04.2022	<i>Виконано</i>
8.	Виконання завдання до підрозділу «Охорона праці»	04.04.2022-10.04.2022	<i>Виконано</i>
9.	Виконання завдання до підрозділу «Безпека в надзвичайних ситуаціях»	11.04.2022-17.04.2022	<i>Виконано</i>
10.	Оформлення кваліфікаційної роботи	18.04.2022-23.04.2022	<i>Виконано</i>
11.	Нормоконтроль	27.04.2022-30.04.2022	<i>Виконано</i>
12.	Перевірка на плагіат	10.05.2022	<i>Виконано</i>
13.	Попередній захист кваліфікаційної роботи	12.05.2022	<i>Виконано</i>
14.	Захист кваліфікаційної роботи	27.05.2022	

Студент

(підпис)

Фіголь В.В.

(прізвище та ініціали)

Керівник роботи

(підпис)

Гром'як Р.С.

(прізвище та ініціали)

АНОТАЦІЯ

Дослідження алгоритмів машинного навчання для потреб систем підтримки прийняття рішень в умовах пандемії // Кваліфікаційна робота освітньо-наукового рівня «Магістр» // Фіголь Віталій Володимирович // Тернопільський національний технічний університет імені Івана Пулюя, факультет комп'ютерно-інформаційних систем і програмної інженерії, кафедра комп'ютерних наук, група СНм-61 // Тернопіль, 2022 // С. 71, рис. – 34, табл. – 1, кресл. – 17, додат. – 1, бібліогр. – 31.

Ключові слова: COVID-19, конвеєр, машинне навчання, нейромережа, пандемія, прийняття рішень, прогнозування.

Кваліфікаційна робота присвячена дослідженню алгоритмів машинного навчання для потреб систем підтримки прийняття рішень в умовах пандемії. В першому розділі кваліфікаційної роботи описано процес пошуку наукових публікацій щодо використання алгоритмів машинного навчання в умовах пандемії. Розглянуто машинне навчання та глибоке навчання для COVID-19. В другому розділі кваліфікаційної роботи проаналізовано показники оцінки використання алгоритмів машинного навчання в умовах пандемії. Подано опис процесу тестування NN в процесі дослідження алгоритмів машинного навчання в умовах пандемії. В третьому розділі кваліфікаційної роботи описано результати ідентифікації ключових біомаркерів. Виконано порівняння прогностичних характеристик алгоритмів машинного навчання для потреб систем підтримки прийняття рішень в умовах пандемії.

Об'єкт дослідження: процеси аналітичного опрацювання з використанням алгоритмів машинного навчання. Предмет дослідження: методи збирання та аналітичного опрацювання біометричних сутностей для потреб систем підтримки рішень.

ANNOTATION

Study of machine learning algorithms for the needs of decision support systems in a pandemic // Qualification work of the educational level "Master" // Fihol Vitalii Volodymyrovych // Ternopil National Technical University named after Ivan Pulyuy, Faculty of Computer Information Systems and Software Engineering, Department of Computer Science, SNnm-61 group // Ternopil, 2021 // P. 71, fig. - 34, tables - 1, chair. - 17, annexes - 1, references. - 31.

Key words: COVID-19, conveyor, machine learning, neural network, pandemic, decision making, forecasting.

Qualification work is devoted to machine learning algorithms study for decision support systems needs in a pandemic. The qualification work first section describes the searching process for scientific publications on the machine learning algorithms use in a pandemic. Machine learning and deep learning for COVID-19 are considered. The qualification work second section analyzes the indicators for assessing the machine learning algorithms use in a pandemic. A NN testing process description in the studying machine learning algorithms process in a pandemic is given. The qualification work third section describes identification results of key biomarkers. The prognostic characteristics comparison of machine learning algorithms for decision support systems needs in a pandemic is performed.

Object of research: analytical processing using machine learning algorithms processes. Research subject: collection and analytical processing methods of biometric entities for decision support systems needs.

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

DL (англ. Deep Learning) – глибоке навчання.

DSS (англ. Decision Support System) – системи підтримки прийняття рішень.

ML (англ. Machine Learning) – машинне навчання.

NN (англ. Neural Network) – нейронна мережа.

RF (англ. Random Forest) – випадкові ліси.

SVM (англ. Support Vector Machine) – машина опорних векторів.

ВООЗ – Всесвітня організація охорони здоров'я.

КТ – комп'ютерна томографія.

ЗМІСТ

ВСТУП	8
1 СТАН ТА ПЕРСПЕКТИВИ ДОСЛІДЖЕНЬ В ГАЛУЗІ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ ТА ЇХ ВИКОРИСТАННЯ В УМОВАХ ПАНДЕМІЇ.....	10
1.1 Пандемія COVID-19	10
1.2 Опис процесу пошуку наукових публікацій щодо використання алгоритмів машинного навчання в умовах пандемії	12
1.3 Аналіз стану досліджень щодо використання алгоритмів машинного навчання в умовах пандемії	16
1.4 Машинне навчання та глибоке навчання для COVID-19.....	20
1.5 Висновок до першого розділу	22
2 ОБЧИСЛЮВАЛЬНИЙ ЕКСПЕРИМЕНТ ДЛЯ ДОСЛІДЖЕННЯ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ В УМОВАХ ПАНДЕМІЇ.....	23
2.1 Формування та попередня обробка пандемічних наборів даних	23
2.2 Розподіл даних для унікальної сегрегації пацієнтів	24
2.3 Конвеєр машинного навчання для дослідження використання алгоритмів машинного навчання в умовах пандемії	27
2.4 Показники оцінки використання алгоритмів машинного навчання в умовах пандемії	29
2.5 Важливість функції XGBoost для оцінки біомакерів	30
2.6 Вибір NN-функції для дослідження алгоритмів машинного навчання в умовах пандемії	31
2.7 Навчання NN в процесі дослідження алгоритмів машинного навчання в умовах пандемії	33
2.8 Тестування NN в процесі дослідження алгоритмів машинного навчання в умовах пандемії	34

	7
2.9 Висновок до другого розділу	36
3 АНАЛІЗ РЕЗУЛЬТАТІВ ОБЧИСЛЮВАЛЬНОГО ЕКСПЕРИМЕНТУ ДОСЛІДЖЕННЯ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ В УМОВАХ ПАНДЕМІЇ.....	37
3.1 Результати ідентифікації ключових біомаркерів	37
3.2 Порівняння прогностичних характеристик алгоритмів машинного навчання для потреб систем підтримки прийняття рішень в умовах пандемії.....	42
3.3 Ефективність алгоритмів машинного навчання.....	53
3.4 Висновок до третього розділу	55
4 ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ	56
4.1 Організація та перевірка стану охорони праці на підприємстві під час пандемії	56
4.2 Основні принципи і способи забезпечення життєдіяльності.....	61
ВИСНОВКИ.....	66
ПЕРЕЛІК ДЖЕРЕЛ	68
ДОДАТКИ	

ВСТУП

Актуальність теми. Коронавірусна хвороба 2019 року (COVID-19), спричинена вірусом SARS-CoV-2, є гострим респіраторним захворюванням, яке ВООЗ класифікувала як пандемію. Раптовий сплеск кількості інфікованих громадян і високий рівень смертності спричинили зростання навантаження на державні та світові організації, установи та заклади в галузі охорони здоров'я. Отже, дуже важливо визначити ключові фактори перебігу пандемії, щоб оптимізувати стратегію лікування пацієнтів, використання та постачання обмежених матеріальних та людських ресурсів. В сучасній медичній практиці управлінські рішення приймаються із використанням систем підтримки прийняття рішень (DSS). Тому актуальним є дослідження алгоритмів машинного навчання (ML) для потреб систем підтримки прийняття рішень (DSS) в умовах пандемії. При цьому потрібно дослідити різні моделі ML, зокрема, нейронні мережі, логістичну регресію, XGBoost, випадкові ліси, SVM, дерева рішень, щоб визначити модель, яка досягає стабільно високої точності прогнозування часу та складності перебігу хвороби. Проведення дослідження спрямоване на пошук програмно-алгоритмічних рішень, які зможуть допомогти прискорити процеси прийняття рішень у різнотипових системах в галузі охорони здоров'я та підвищити точність діагностування та якість цілеспрямованого лікування громадян.

Мета і задачі дослідження. Метою даної кваліфікаційної роботи освітньо-наукового рівня «Магістр» є підвищення рівня повноти подання інформації щодо перебігу пандемічних процесів та полегшення процесів прийняття рішень в умовах пандемії. Для досягнення поставленої мети було потрібно виконати наступні завдання:

– Проаналізувати стан досліджень в предметній області алгоритмів машинного навчання в умовах пандемії.

– Проаналізувати формування та попередню обробку пандемічних наборів даних.

– Дослідити існуючі на даний час методи сформовані на основі алгоритмів машинного навчання.

Об’єкт дослідження процеси аналітичного опрацювання з використанням алгоритмів машинного навчання.

Предмет дослідження методи збирання та аналітичного опрацювання біометричних сутностей для потреб систем підтримки рішень.

Наукова новизна одержаних результатів кваліфікаційної роботи полягає у тому, що отримав подальший розвиток метод сформований на основі конвеєра машинного навчання.

Практичне значення одержаних результатів. На основі багатокритеріального аналізу алгоритмів машинного навчання обґрунтовано переваги найромережних методів для потреб систем підтримки прийняття рішень в умовах пандемії.

Апробація результатів магістерської роботи. Основні результати проведених досліджень обговорювались на ІХ науково-технічній конференції «ІНФОРМАЦІЙНІ МОДЕЛІ, СИСТЕМИ ТА ТЕХНОЛОГІЇ» Тернопільського національного технічного університету імені Івана Пулюя (м. Тернопіль, 2021 р.).

Публікації. Основні результати кваліфікаційної роботи опубліковано у двох працях конференції (Див. додатки А).

Структура й обсяг кваліфікаційної роботи. Кваліфікаційна робота складається зі вступу, чотирьох розділів, висновків, списку літератури з 31 найменування та одного додатка. Загальний обсяг кваліфікаційної роботи складає 71 сторінку, з них 46 сторінок основного тексту, який містить 34 рисунки та одну таблицю.

1 СТАН ТА ПЕРСПЕКТИВИ ДОСЛІДЖЕНЬ В ГАЛУЗІ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ ТА ЇХ ВИКОРИСТАННЯ В УМОВАХ ПАНДЕМІЇ

1.1 Пандемія COVID-19

Спалах нової коронавірусної хвороби 2019 року (COVID-19), спричиненої важким гострим респіраторним синдромом коронавірус два (SARS-CoV-2), почався в Ухані, Китай, у грудні 2019 року. Відтоді він швидко поширився по всьому світу. Раптове зростання кількості випадків спричинило підвищене навантаження на світові та національні системи охорони здоров'я, що супроводжувалось гострою нестачею обширного переліку ресурсів. Дослідження алгоритмів машинного навчання для потреб систем підтримки прийняття рішень в умовах пандемії має важливе значення, оскільки воно допомагає визначити фактори що впливають на ризику смерті пацієнтів та керувати рішеннями щодо оперативного розподілу дефіцитних медичних та людських ресурсів під час динамічного перебігу триваючої пандемії COVID-19.

Основними ознаками інфекції COVID-19 є лихоманка, сухий кашель і утруднення дихання. Деякі пацієнти можуть мати болі в м'язах, відчувати втому та втрату смаку чи нюху (аносмія), а до 10% мають симптоми, пов'язані з розладами шлунково-кишкового тракту, зокрема діарею. Як вважалося спочатку, одним із потенційних шляхів поширення вірусу між людьми є прямий контакт. Передача захворювання може відбуватися на відстані до 6 футів (1,81 м). Біологічні матеріали інфікованої людини поширюються під час розмови або чхання, тому повітряно-крапельний шлях можна вважати однією з основних причин поширення хвороби. Крім того, симптоми COVID-19 у деяких випадках не помітні. Таким чином, соціальне дистанціювання може зменшити ймовірність зараження.

На рисунку 1.1 показано, як COVID-19 поширюється від людини до іншої людини, коли немає соціального дистанціювання.

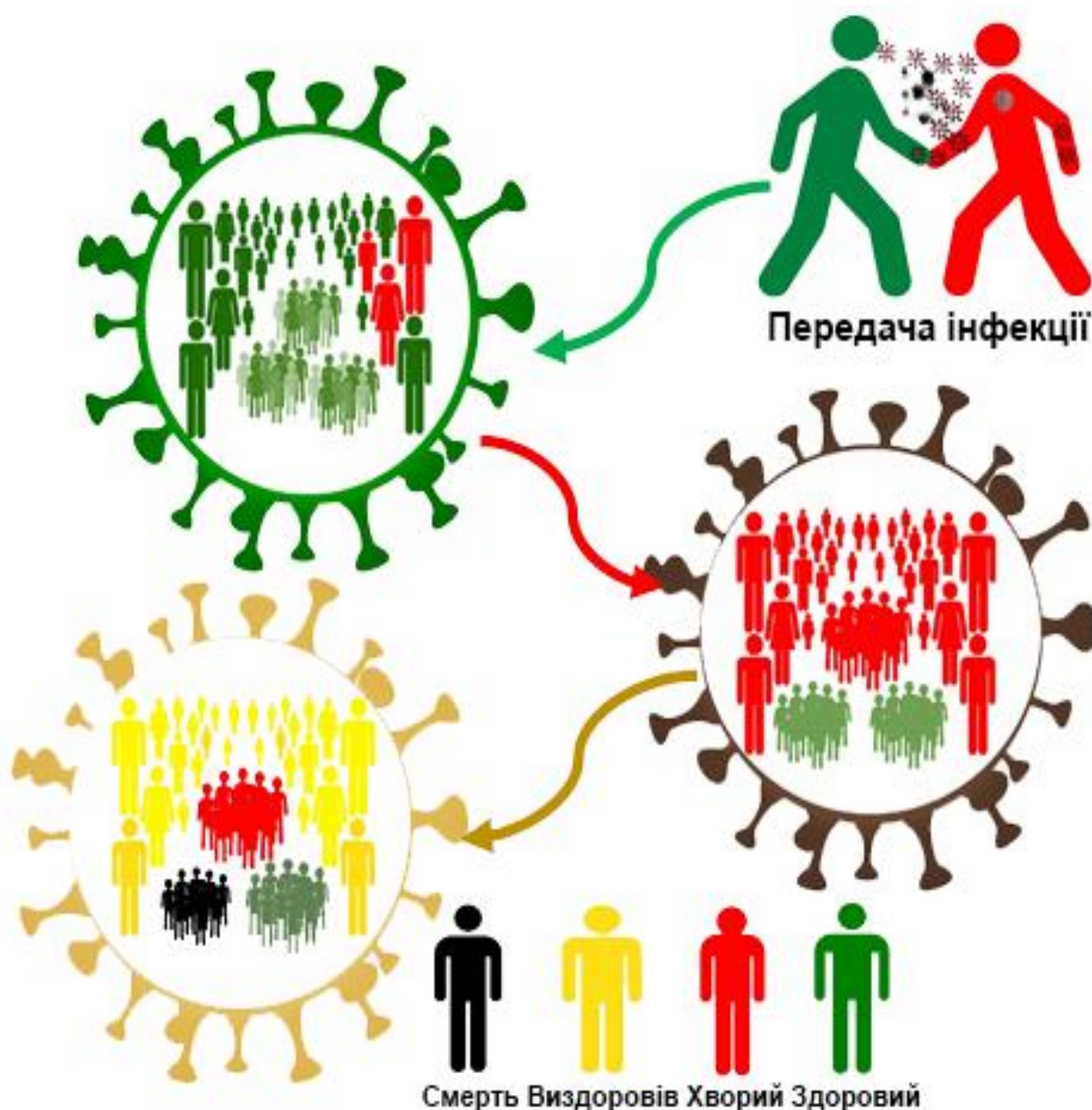


Рисунок 1.1 – COVID-19 Поширення від людини до людини [1]

У групі найбільшого ризику є люди похилого віку та люди з супутніми захворюваннями, зокрема серцево-судинними захворюваннями та діабетом [2]. Оскільки хвороба поширюється по всьому світу то відбувається виявлення зростаючого переліку симптомів і особливостей, які впливають на

смертність пацієнтів. Наявність великого та динамічно зростаючого набору ознак, на які впливає коронавірусне захворювання, ускладнює розуміння ступеня їх впливу на смертність пацієнтів. Дослідження алгоритмів машинного навчання для потреб систем підтримки прийняття рішень в умовах пандемії може допомогти при аналізі великі наборів даних для пришвидшення процесів виявлення закономірностей та надання моделей з високою точністю оцінювання факторів ризику.

1.2 Опис процесу пошуку наукових публікацій щодо використання алгоритмів машинного навчання в умовах пандемії

Автори [3] опублікували результати дослідження щодо відбору публікацій з використанням найбільш релевантних ключових слів. Зокрема, «COVID-19», «машинне навчання» та «глибоке навчання».

Зібрані лише дослідження, описані англійською мовою. При цьому використані цифрові бази даних «Elsevier», «IEEE Xplore», «MDPI», «Springer», «IOP Press» та «Wiley».

Описана авторами процедура пошуку включала п'ять кроків. На першому етапі проведено пошук опублікованих результатів досліджень щодо пандемії COVID-19. Всього було знайдено двісті п'ятдесят п'ять робіт.

При пошуку були проаналізовані дослідження щодо використання методів ML та DL для пандемії COVID-19. На цьому етапі було використано три ключові слова – «COVID-19», «машинне навчання» та «глибоке навчання» в різних комбінаціях у поєднанні з операторами «AND» та «OR».

На другому етапі були виключені дублікати документів. Всього було виключено п'ятдесят п'ять публікацій.

На третьому кроці відібрані публікації були відфільтровані на основі назв та анотацій. Роботи з темами, що не відносяться до досліджуваної

предметної області, були виключені. Понад сто публікацій було виключено і лише сто публікацій було відібрано та передано на наступний етап.

Блок-схема процедури відбору публікацій подана на рисунку 1.2.

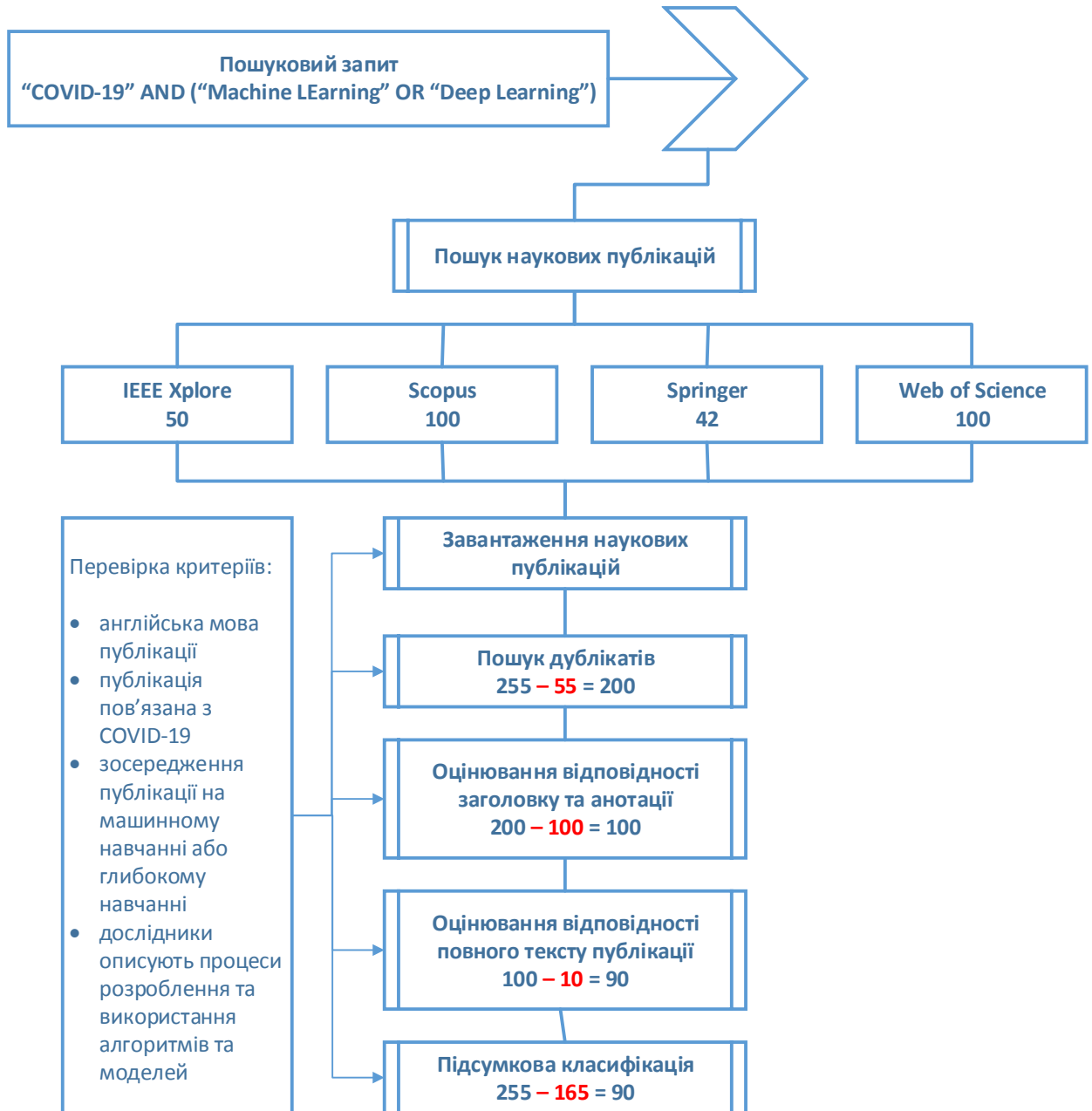


Рисунок 1.2 – Блок-схема відбору публікацій щодо досліджень із застосуванням запитів і критеріїв включення

Розподіл знайдених публікацій по наукометричних БД подано на рисунку 1.3 [3].

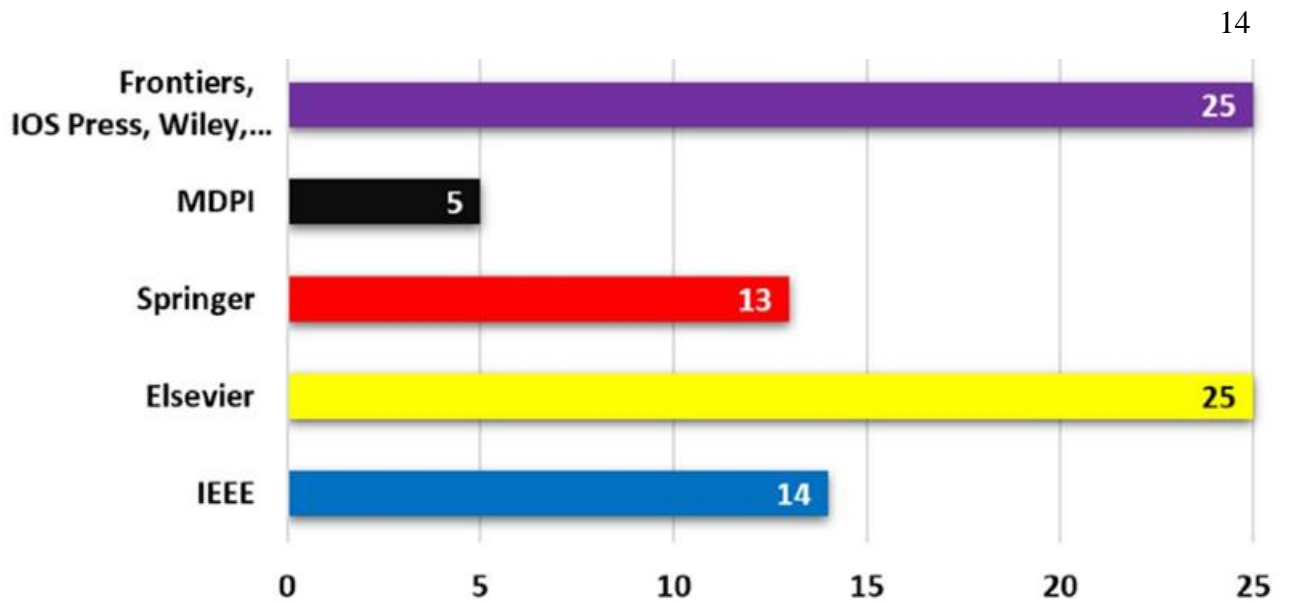


Рисунок 1.3 – Розподіл виявлених публікацій по наукометричних БД

Саме більше відповідних публікацій – двадцять п’ять було виявлено в наукометричній базі даних Elsevier.

Серед проаналізованих публікацій сто сімдесят сім – у відкритому доступі а двадцять три в закритому доступі (див. рисунок 1.4).

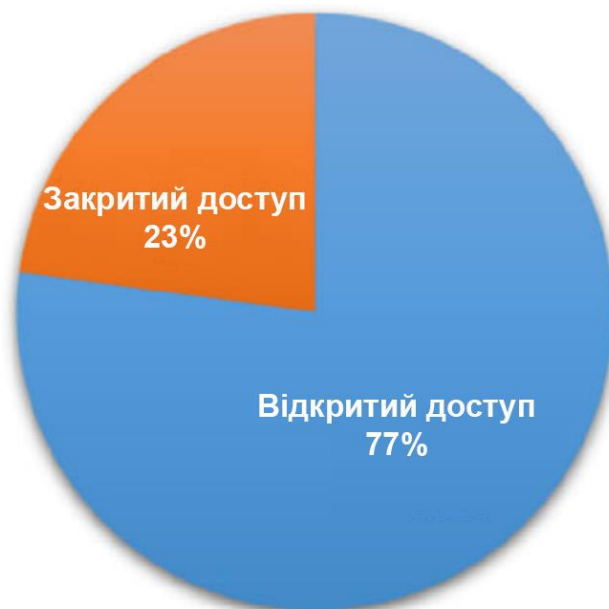


Рисунок 1.4 – Доступність виявлених публікацій [3]

Серед відібраних публікацій було виявлено одні тези конференцій та дев’яносто дев’ять статей (див. рисунок 1.5).

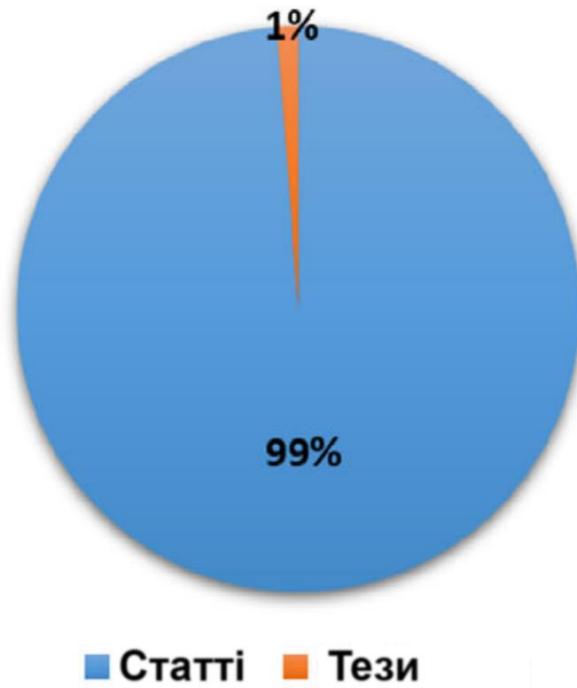


Рисунок 1.5 – Розподіл статей та матеріалів конференцій серед виявлених публікацій [3]

Розподіл відібраних публікацій по категоріях підходів до навчання подано на рисунку 1.6 [3].



Рисунок 1.6 – Розподіл відібраних публікацій по категоріях підходів до навчання

На четвертому етапі було розглянуто кожну публікацію, щоб переконатися, що її тема знаходиться в межах визначеного напрямку досліджень. При цьому було виключено десять статей. До підсумкової вибірки було включено дев'яносто робіт, пов'язаних із методами ML та DL та їх використанням для процесів прийняття рішень під час пандемії COVID-19. На ключових етапах були використані ті самі критерії прийнятності.

1.3 Аналіз стану досліджень щодо використання алгоритмів машинного навчання в умовах пандемії

У ряді досліджень [4], [5], [6] застосовано алгоритми ML для виявлення та діагностування інфекції COVID-19 у громадян. Вони забезпечують можливість для швидшого скринінгу пацієнтів під час пандемії та допомагають подолати проблеми з виконанням полімеразної ланцюгової реакції зворотної транскрипції (RT-PCR) у масштабах популяцій.

Більшість досліджень COVID-19 на основі ML базуються на основі опрацювання рентгенівських знімків грудної клітки або КТ [4]. Такі дані важко отримати в умовах низького ресурсу [7], тому виникає необхідність формування моделей на основі альтернативних даних, отриманих за допомогою легкодоступних та недорогих тестів.

Виявлення COVID-19 на основі аудіозаписів зразків кашлю виявилось перспективним [6]. Однак для прогнозування ступеня тяжкості перебігу COVID-19 потрібні масштабовані програмно-алгоритмічні рішення. Звичайні аналізи крові показали перспективний рівень прогнозування тяжкості перебігу інфекційних захворювань.

Для прогнозування ризику виникнення важких ускладнень і смертності було запропоновано використання різних моделей ML [8], [9], [10]. Використання ML є важливим при прийнятті рішень в умовах обмеженості матеріальних та людських ресурсів при збільшенні кількості інфікованих

COVID-19 пацієнтів. При цьому важливими є питанням щодо розподілу ресурсів серед пацієнтів в залежності від прогнозів перебігу захворювання.

Ванг [3] запропонував дві різні моделі, сформовані на основі аналітичного опрацювання клінічних і лабораторних характеристик. Клінічна модель, розроблена з урахуванням віку, гіпертонії та ішемічної хвороби серця в анамнезі, показала AUC 0,83 для когорти підтвердження. Лабораторна модель, корельована з віком, використовує високочутливий С-реактивний білок (hs-CRP), насичення периферичних капілярів киснем (SpO_2), враховувала кількість нейтрофілів та лімфоцитів, D-димер, аспартатамінотрансферазу (AST) і швидкість клубочкової фільтрації (ШКФ).

При цьому краща дискримінаційна потужність з AUC 0,88 була продемонстрована для когорти валідації. Когорта перевірки була сформована на основі даних сорока п'яти пацієнтів інфікованих COVID-19. Серед них п'ятнадцять осіб померли, а тридцять вижили. XGBoost та зворотній понижуючий вибір були використані для вибору ознак, з подальшою багатоваріантною логістичною регресією для класифікації. Клінічна модель може виявитися корисною завдяки простоті збору даних щодо усіх трьох ознак.

Шанг [11] запропонував систему оцінювання COVID-19 з метою класифікації пацієнтів на групи низького та високого ризику. Пацієнти групи високого ризику матимуть значно вищі шанси на смерть, ніж пацієнти з групи низького ризику. Для аналізу ознак і формування моделі прогнозування використовувалися багатофакторний аналіз і коефіцієнти двійкової логістичної регресії ласо. Створена модель показала хорошу дискримінаційну здатність з AUC 0,938 для незалежної когорти підтвердження. Було використано вісім різних змінних, зокрема, вік та параметри крові.

Хіє [12] описує визначення SpO_2 , кількості лімфоцитів, віку та лактатдегідрогеназу (ЛДГ) в якості набору важливих ознак для створення

моделі прогнозування перебігу захворювання. Він використовував багатофакторну логістичну регресію для завдання класифікації, яка продемонструвала AUC 0,98 для незалежного набору перевірки. Автори створили номограму для визначення ймовірності смертності.

Хіменес-Содем [13] використано дані про COVID-19 з Данії та Великобританії для побудови моделей прогнозування смертності. Зазначені моделі продемонстрували значення AUC 0,906 при діагностиці, 0,818 при госпіталізації та 0,721 при надходженні пацієнтів у відділення реанімації. Виділено загальні фактори ризику – вік, індекс маси тіла та гіпертонія.

Болурані [14] розроблено моделі ML для прогнозування дихальної недостатності протягом двох діб годин після госпіталізації інфікованого COVID-19 пацієнта. Модель XGBoost продемонструвала найвищу середню точність 0,919 та AUC 0,77. Найбільш впливові змінні – тип доставки кисню, який використовується у відділенні невідкладної допомоги, вік, рівень індексу тяжкості невідкладної допомоги, частота дихання, рівень лактату в сироватці та демографічні характеристики.

Більшість опублікованих досліджень демонструють узгодженість результатів з різними періодами перебігу захворювання. Тому аналіз алгоритмів машинного навчання для потреб систем підтримки прийняття рішень в умовах пандемії необхідний для оцінки ефективності моделей та надійного застосування в реальних сценаріях, коли пацієнт може перебувати на будь-якій стадії захворювання.

Група авторів, очолена Ян [15] проаналізували зразки крові чотирьохсот восьмидесяти п'яти пацієнтів з Уханя (Китай) і розробили програмно-алгоритмічне рішення на основі XGBoost. Запропонована клінічно-діюча модель XGBoost з одним деревом використовувала три значимо-важливі характеристики ЛДГ, лімфоцити та hs-CRP.

Розроблені рекурсивні правила прийняття рішень із трьома ознаками разом з порогом. Це забезпечило інтерпретаційне рішення машинного

навчання з точністю щонайменше 90% протягом усіх днів перебігу захворювання. Автори проаналізували узгодженість запропонованої моделі з точністю дев'яносто чотири відсотки за сім днів.

Однак, можливо, їхні результати є необ'єктивними [16]. Використаний для визначення консистенції метод, був відхилений через зростання кількості зразків з наближенням дня результату. При цьому показник F1 на перший день – 0,97 та зменшується до 0,68 на сімнадцятий день від результату. Це свідчить про невідповідність. Незбалансований набір тестів призводить до ненадійності результатів. Бажано точно прогнозувати протягом перших днів після зараження, що допомагає розробити стратегію лікування на ранніх стадіях.

Тому потрібно дослідити конвеєр машинного навчання, щоб подолати недоліки існуючих моделей та покращити продуктивність. Слід проаналізувати набори даних, надані у відкритих джерелах, які містять біомаркери, отримані з аналізів крові, для розробки моделей. Доцільно створити рішення, використовуючи важливість функцій XGBoost і класифікацію нейронної мережі (NN).

При цьому потрібно виконати вибір ознак, щоб мінімізувати кількість ознак, які можна використовувати для розробки моделей стратифікації ризиків. Велика кількість функцій для невеликої когорти іноді може призвести до перенавчання. Доцільно обрати ключові біомаркери для прогнозування, наприклад комбінацію з п'яти ознак.

Нейронна мережа може забезпечити високу продуктивність прогнозування з точністю до 96% [17], а графік важливості функції XGBoost додає моделі інтерпретації. Аналіз графіків функцій, які показують чіткі тенденції розвитку, також дає додаткове уявлення про функції.

Різні алгоритми та надійне тестування були реалізовані в [17] з високим рівнем довіри до запропонованої моделі. Запропонована авторами модель виявилась точною та послідовною протягом періоду захворювання пацієнта.

Вона може бути використана при швидшій діагностиці з меншою кількістю функцій і більшою достовірністю. Системи підтримки прийняття рішень в галузі охорони здоров'я можуть використовувати модель для оптимізації стратегії лікування за рахунок цілеспрямованого використання ресурсів.

1.4 Машинне навчання та глибоке навчання для COVID-19

Враховуючи швидке та географічно широке розповсюдження коронавірусу, алгоритми ML та DL доволі часто використовуються для підвищення ефективності традиційних методів виявлення або прогнозування COVID-19. Для діагностування та виявлення COVID-19 використовуються алгоритми ML, наприклад, контрольоване навчання та неконтрольоване навчання. Крім того, в проаналізованих наукових публікаціях було використано для виявлення COVID-19 декілька алгоритмів DL. Також використовуються гібридні підходи, які поєднують ML та DL для діаностування та виявлення випадків інфікування COVID-19. На рисунку 1.7 подано класифікацію методів ШІ сформованих на основі алгоритмів ML та DL [1].

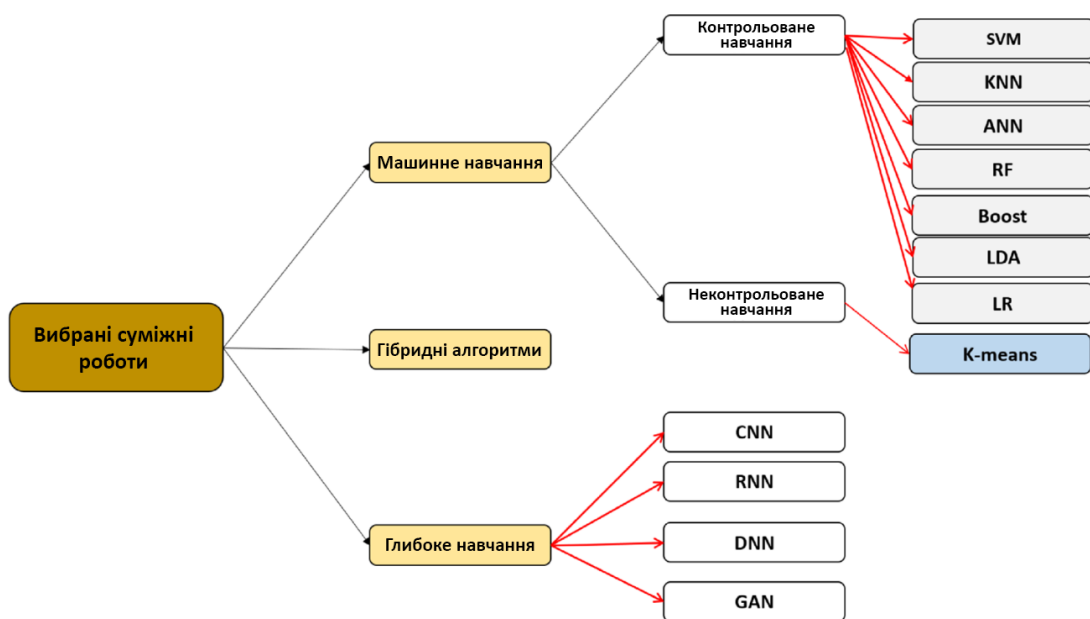


Рисунок 1.6 – Класифікація супутніх наукових публікацій

Алгоритми ML використовуються для подання інструкцій комп'ютерам, щодо роботи в процесі самостійного навчання. Вони сформовані на основі методів аналітичного опрацювання даних, які передбачають формування та підгонку моделей, які дозволяють машинам «навчатися» на практиці та робити прогнози.

На рисунку 1.8 подано розподіл по країнах публікацій щодо дослідження ML для COVID-19.

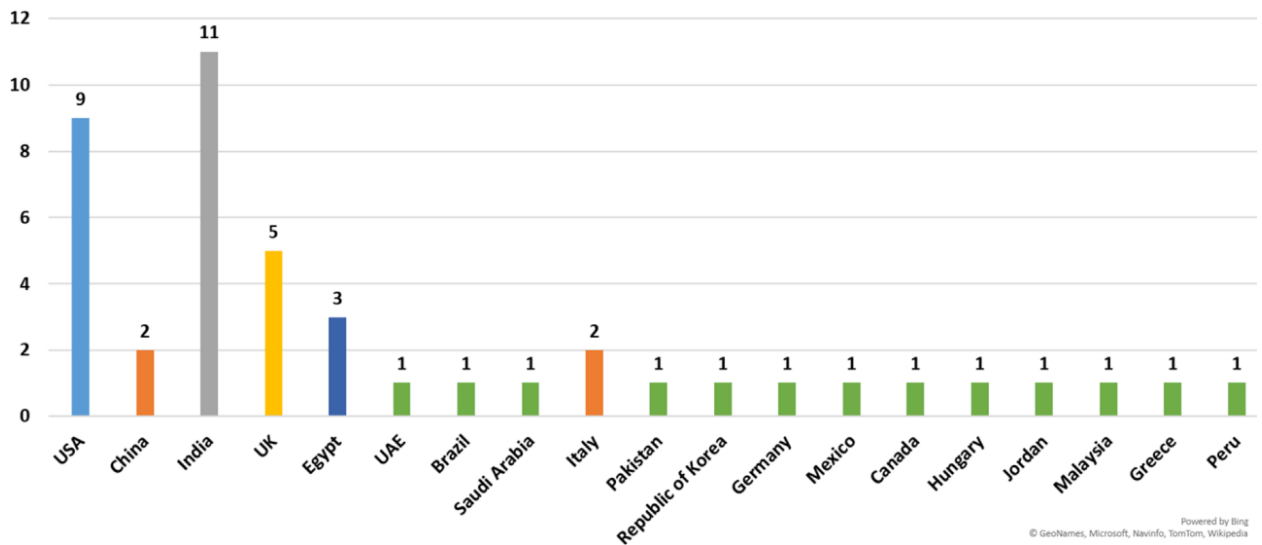


Рисунок 1.8 – Розподіл по країнах публікацій щодо дослідження ML для COVID-19 [1]

Загалом, алгоритми ML, які використовуються для діагностики COVID-19, можна розділити на два підходи – контрольоване навчання та неконтрольоване навчання.

Алгоритми ML використовується для виявлення COVID-19 шляхом аналізу вмісту входних рентгенівських або КТ-зображень та видобування з них унікальних ознак. Відповідно до зазначених функцій відбувається прогнозування класифікації входних зображень звичайний випадок або як випадок інфікування.

На рисунку 1.9 подано географічний розподіл публікацій щодо дослідження ML для COVID-19.

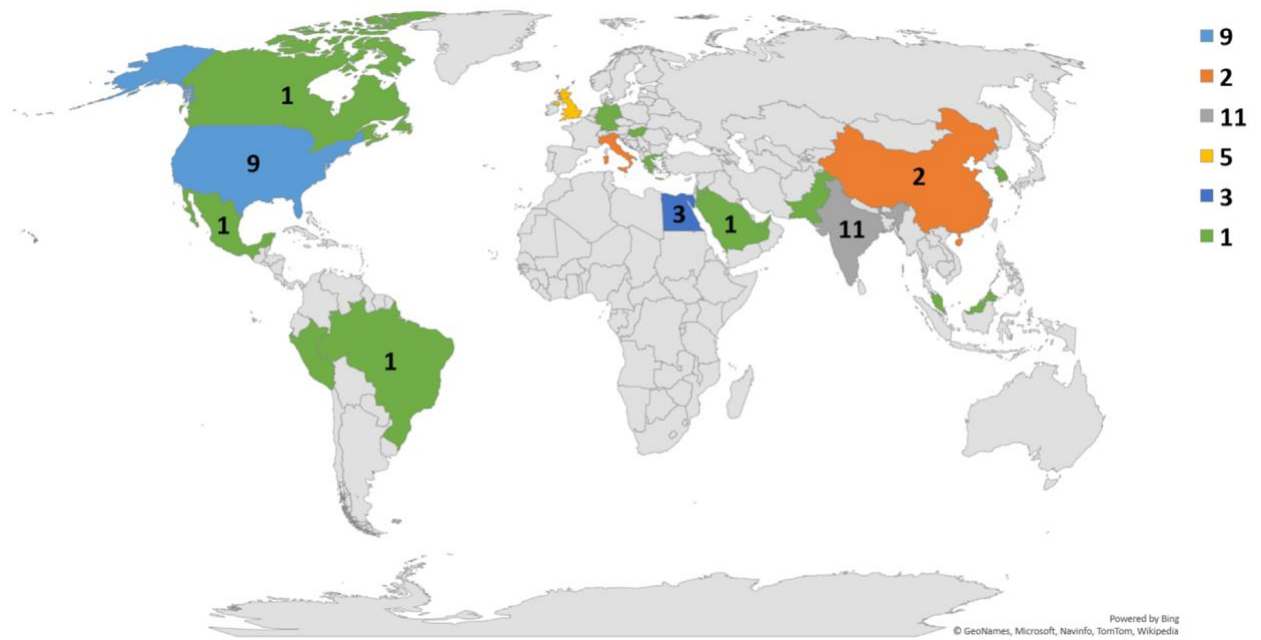


Рисунок 1.9 – Географічний розподіл публікацій щодо дослідження ML для COVID-19 [1]

1.5 Висновок до першого розділу

В першому розділі кваліфікаційної роботи освітнього-наукового рівня «Магістр» описано пандемію COVID-19. Наведено опис процесу пошуку наукових публікацій щодо використання алгоритмів машинного навчання в умовах пандемії. Проаналізовано стан досліджень щодо використання алгоритмів машинного навчання в умовах пандемії. Розглянуто машинне навчання та глибоке навчання для COVID-19.

2 ОБЧИСЛЮВАЛЬНИЙ ЕКСПЕРИМЕНТ ДЛЯ ДОСЛІДЖЕННЯ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ В УМОВАХ ПАНДЕМІЇ

2.1 Формування та попередня обробка пандемічних наборів даних

В процесі дослідження алгоритмів машинного навчання використано дані для аналізу ознак, навчання та тестування [15]. Цей набір даних містить дві тисячі сімсот сімдесят дев'ять електронних записів підтверджених або підозрюваних на COVID-19 пацієнтів з лікарні Тунцзи в м. Ухань (Китай). Початковий набір даних сформовано з часового ряду трьохсот семидесяти п'яти інфікованих COVID-19 пацієнтів з сімдесят чотирма біомаркерами, часом вибірки даних, часом госпіталізації, часом виписки та бінарним показником летальності для осіб що вижили чи померли.

Ян та ін. [15] використовували лише дані остаточних зразків кожного окремого пацієнта для навчання та тестування.

В процесі дослідження алгоритмів машинного навчання розглянуто зразки з усіх днів періоду спостереження кожного пацієнта для навчання та тестування. Для кожного пацієнта обрано декілька рядків даних, що представляли покази, зроблені в різні дні спостереження. Для деяких днів також проводилося декілька операцій читання даних у різний час. Всі рядки даних, що представляють пацієнта в один день, були об'єднані разом, щоб створити єдиний вузол даних для кожного унікального дня пацієнта.

Коли існували об'єкти спостереження з декількома записами, зробленими протягом одного дня в різний час, для генерування об'єднаної єдиного вузла даних враховуються показники, які були зняті найраніше цього конкретного дня. Оскільки доцільно, щоб модель якомога раніше вивчала особливості, які передбачають ризики ускладнень та летальності для пацієнта. Характеристики, значення яких були відсутні в більш ніж семидесяти відсотків випадків, були вилучені та не використовувалися для

подальшого аналізу. В процесі первинного опрацювання даних слід додати новий стовпець «Кількість днів до підсумкового результату», щоб позначити, скільки днів залишилося вибірці даних до дня результату. Значення обчислюється відніманням дати дня читання даних від дати дня виписки чи смерті. Після генерації та обробки даних досліджуваний набір даних містить відомості щодо понад двісті пацієнтів, що вижили, та сто шістьдесят дев'ять померлих пацієнтів. Проаналізовано особливості щодо двох класів.

Відсутні значення в навчальному наборі були обчислені за допомогою алгоритму «KNearest Neighbor». Значення для імпутації було розраховано шляхом усереднення значень десяти найближчих сусідів, зважених по відношенню до евклідової відстані. Для нормалізації навчальних даних було використано «Min-Max» масштабування. Його використано, оскільки більшість функцій не відповідають нормальному розподілу та забезпечують збереження структури вузлів даних у кожній функції.

2.2 Розподіл даних для унікальної сегрегації пацієнтів

Набір даних після попередньої обробки складається сформовано з понад тисячу сімсот п'ятидесяти вузлів даних, що відповідають трьохсот семидесяти пацієнтам, серед яких п'ятдесят чотири проценти одужало, а сорок шість процентів померли від COVID-19. Кількість вузлів даних для кожного з пацієнтів коливається від одного до дванадцяти. Дані були зібрані протягом різних днів настання одного з двох результатів.

Щоб забезпечити ексклюзивність відомостей щодо пацієнтів у навчальному та тестовому наборах, вісімдесят відсоткі пацієнтів були випадковим чином обрані для навчального набору, а решта двадцять відсотків пацієнтів були обрані для тестового набору.

Розподіл тренувальних даних по днях до результату показано на рисунку 2.1.

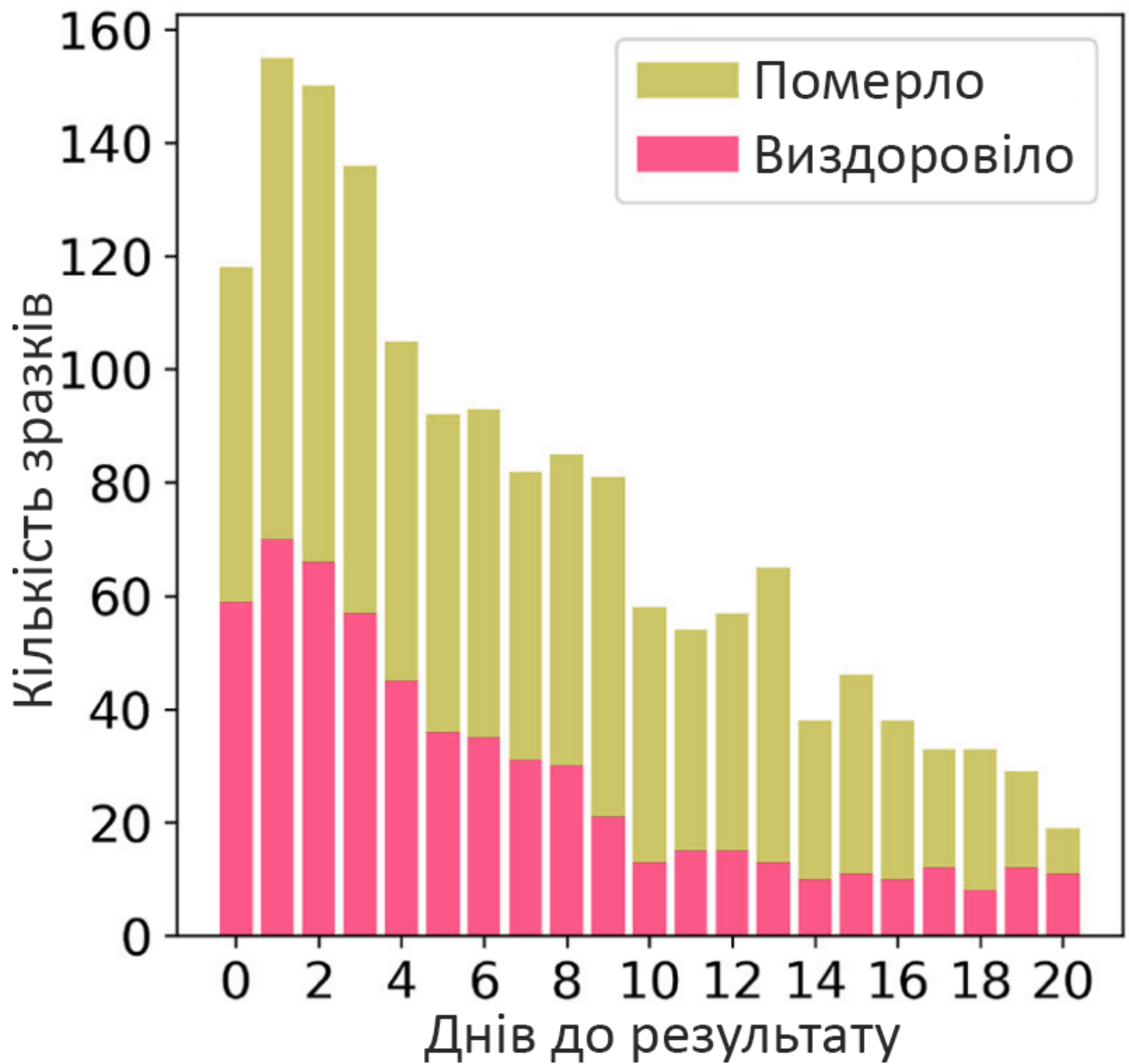


Рисунок 2.1 – Розподіл тренувальних даних

Унікальна сегрегація пацієнтів важлива, оскільки включення вузлів даних від одного пацієнта до навчальних і тестових наборів може призвести до упередженості.

Поділ «80/20» забезпечив унікальний навчальний набір, що складається з понад тисячу чотириста вузлів даних, та тестовий набір, що сформовано на основі більш як як трьохсот сорока вузлів даних.

Розподіл тестових даних по днях до результату показано на рисунку 2.2.

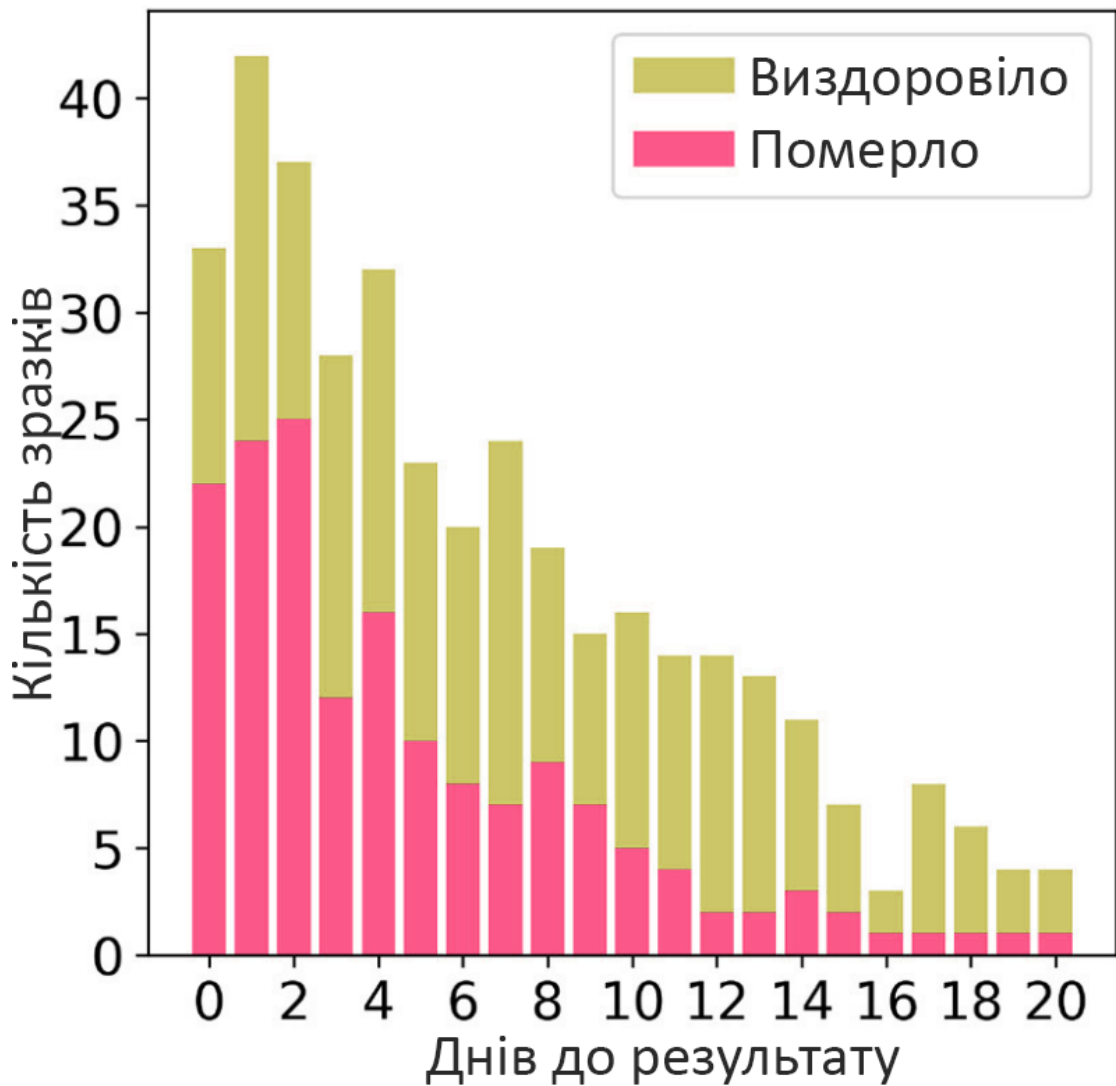


Рисунок 2.2 – Розподіл тренувальних даних

Заняття розподіляються на всі дні в оптимальних співвідношеннях та є співмірними. Оскільки метою кваліфікаційної роботи є дослідження алгоритмів машинного навчання для потреб систем підтримки прийняття рішень в умовах пандемії, то не повинно виникати залежності від кількості днів до результату. Тому всі отримані від пацієнтів показники розглядаються як унікальні вузли даних для подальшого дослідження.

При цьому доцільно порівнювати результати з та без використання методу К-найближчого сусіда (KNN). Алгоритм KNN корисний для узгодження вузлів даних з їх найближчими k сусідами в багатовимірних просторах.

Відповідно згенеруємо призначений тестовий набір так, що кожен вузол даних у тестовому наборі буде обчислено з використанням значень, що відповідають вибраним п'яти об'єктам із десяти їх найближчих сусідів у наборі поїздів.

В процесі дослідження отримано тестовий набір, який містив понад двісті десять зразків, що належать сорока п'яти пацієнтам. Співвідношення померлих до загальних пацієнтів становить 0,57 та свідчить, що в тестовому наборі добре представлено обидва класи пацієнтів. Це дозволяє гарантувати надійність результатів для обох класів. Незараховані тестові набори призначені для оцінювання продуктивності алгоритмів, коли тестовий набір не має синтезованих значень. Для цього було вилучено рядки даних, у яких були відсутні значення будь-якої з п'яти функцій.

В результаті отримано тестовий набір, який містить сто п'ятнадцять зразків, що належать сорока п'яти пацієнтам із співвідношенням померлих до загальної кількості пацієнтів 0,5.

2.3 Конвеєр машинного навчання для дослідження використання алгоритмів машинного навчання в умовах пандемії

Усі досліджувані алгоритми повинні пройти процедуру навчання на вибірках за всі дні з набору даних для оцінювання здатності прогнозування, незалежно від кількості днів до результату. Після попереднього опрацювання даних для визначення ступеня важливості ознак доцільно використати класифікатор XGBoost. Для вибору ознак доцільно використати нейронну мережу. Отримана таким чином оптимальна комбінація функцій потім повинна бути використана для навчання різних класифікаційних моделей сформованих на основі алгоритмів машинного навчання.

На рисунку 2.3 подано блок-схему узагальненого конвеєра дослідження алгоритмів машинного навчання [17] для потреб систем підтримки прийняття рішень в умовах пандемії.

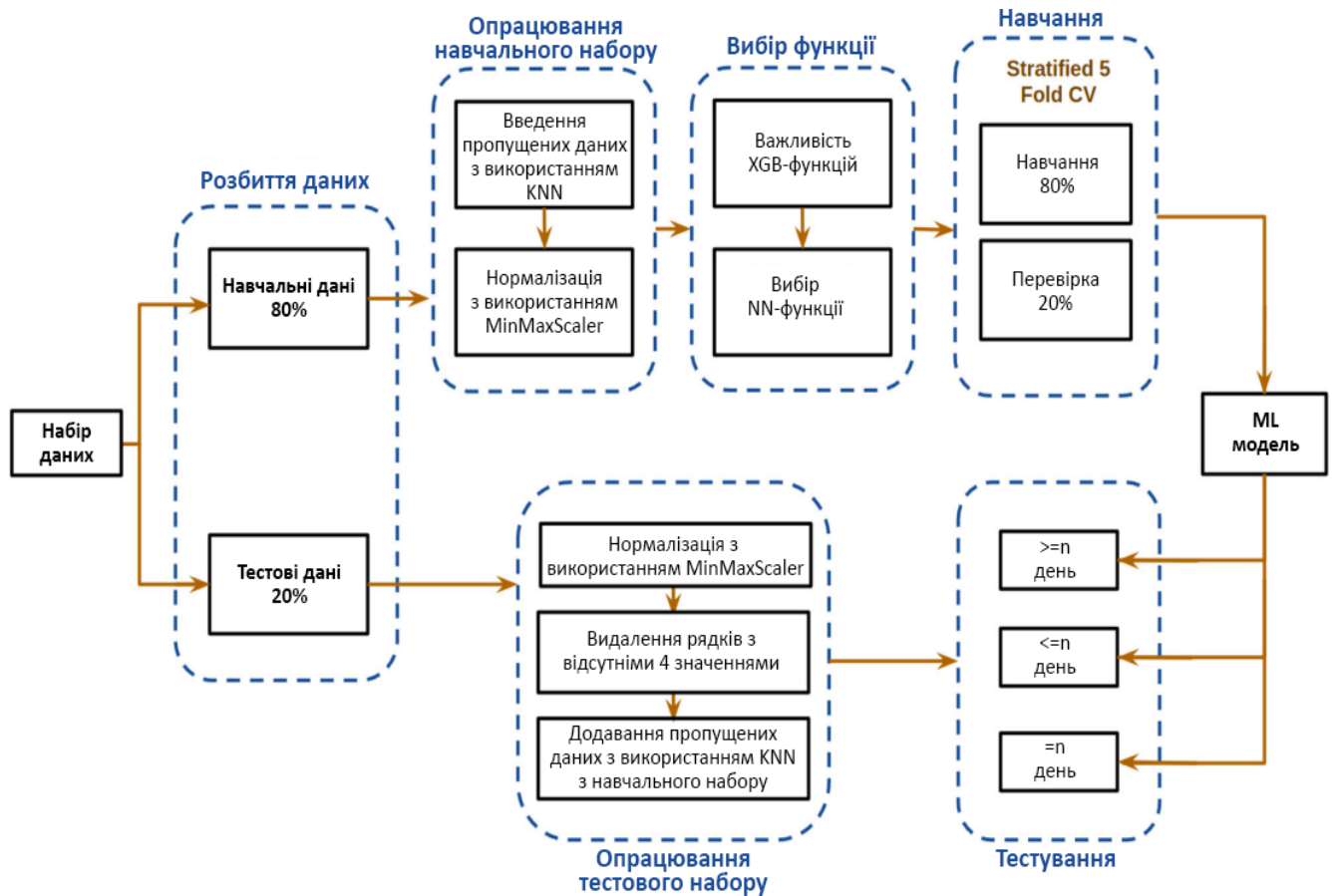


Рисунок 2.3 – Блок-схема узагальненого конвеєра дослідження алгоритмів машинного навчання для потреб систем підтримки прийняття рішень в умовах пандемії

Потім навчені моделі повинні були перевірені трьома різними способами, кожен з яких має свої переваги. Повинна бути здійснена п'ятикратна перехресна перевірка для оцінки передбачуваної здатності та статистичної значущості досліджуваних моделей, що сформовані на основі ML-алгоритмів. Слід провести оцінку моделей на основі різних показників, середнього значення та стандартного відхилення.

2.4 Показники оцінки використання алгоритмів машинного навчання в умовах пандемії

Прогнозну продуктивність контрольованих моделей потрібно оцінювати за допомогою показників: істинно-позитивний (TP), істинно-негативний (TN), хибнопозитивний (FP) і хибнонегативний (FN).

ROC-AUC чи AUC – це «Площа під кривою робочої характеристики приймача», або скорочено крива ROC. Вона забезпечує сукупне вимірювання ефективності за всіма можливими порогоми класифікації. Крива ROC відображає два параметри. Істинно-позитивна оцінка:

$$TPR = \frac{TP}{TP + FN} .$$

Та хибно-позитивна оцінка:

$$FPR = \frac{FP}{FP + TN} .$$

При цьому AUC вимірює всю двовимірну площу під кривою ROC, починаючи від (0,0) та закінчуючи (1,1).

Точність є важливою метрикою для всіх моделей класифікації. При проведенні дослідження алгоритмів машинного навчання тестовий набір даних не є незбалансованим, тому можна сформулювати добре уявлення про прогнозну продуктивність моделі:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Оцінка F1 – це середнє гармонійне значення точності та запам'ятовування:

$$F1score = \frac{2 \times Precision \times Recall}{Precision + Recall},$$

де,

$$Precision = \frac{TP}{TP + FP},$$

та,

$$Recall = \frac{TP}{TP + FN}$$

2.5 Важливість функції XGBoost для оцінки біомакерів

Щоб оцінити біомаркери, які найбільше впливатимуть на результат доцільно використати XGBoost для отримання відносної важливості. XGBoost – це потужний алгоритм машинного навчання, який використовується для оцінювання найбільш дискримінаційних функцій з результатів дослідження моделей та алгоритмів [18]. Остаточна важливість досліджуваного алгоритму обчислюється з використанням середнього значення значень для всіх дерев.

Важливість ознаки над деревом визначається кількістю разів, коли функція використовувалась для поділу дерева. Вона зважується за допомогою квадратичного покращення, зробленого розбиттям у продуктивності моделі чи алгоритмі [19]. Доцільно визначити середню

важливість ознак за допомогою сотні ітерацій випадкового вибору 80% стрічок з навчальному наборі.

При цьому слід прийняти параметр максимальної глибини XGBoost рівним трьом. Швидкість навчання слід встановити 0,2, а параметр регуляризації α потрібно встановити п'яти та прийняти метою логістичну регресію. Всі інші параметри доцільно прийняти за замовчуванням.

2.6 Вибір NN-функції для дослідження алгоритмів машинного навчання в умовах пандемії

На рисунку 2.4 подано архітектуру нейронної мережі для дослідження алгоритмів машинного навчання для потреб систем підтримки прийняття рішень в умовах пандемії [17]. Де n – кількість аналізованих ознак.

Після визначення порядку важливості потрібно використати нейронну мережу (NN) для пошуку оптимальної кількості ознак, необхідних для дослідження алгоритмів машинного навчання для потреб систем підтримки прийняття рішень в умовах пандемії. NN можуть використовуватись для вивчення складних відношень між ознаками.

Вхідний шар повинен мати таку ж кількість нейронів, що й кількість аналізованих ознак. Архітектура NN дослідження алгоритмів машинного навчання для потреб систем підтримки прийняття рішень в умовах пандемії включає два прихованих шари з активацією ReLU. Перший прихований шар повинен мати в двічі більшу кількість нейронів, ніж вхідний шар.

В процесі дослідження алгоритмів машинного навчання для потреб систем підтримки прийняття рішень в умовах пандемії використовується оптимізатор Адама з швидкістю навчання 0,001 та планувальник «Reduce On Plateau» з порогом терпіння, оскільки значення «п'ять» було використано для оновлення ваг та швидкості навчання.

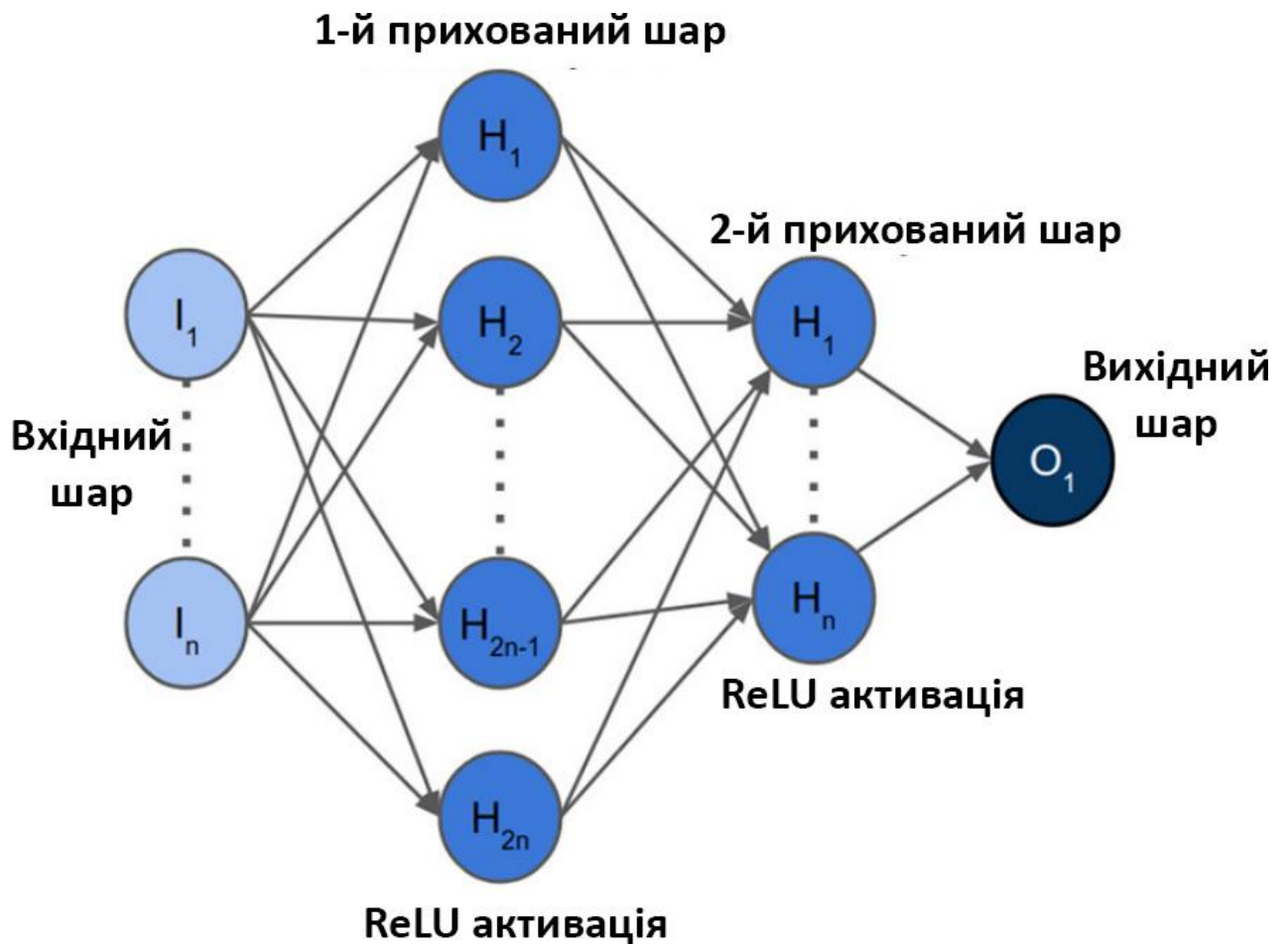


Рисунок 2.4 – Архітектура нейронної мережі для вибору ознак при дослідженні алгоритмів машинного навчання в умовах пандемії

Другий прихований шар повинен мати таку ж кількість нейронів, як вхідний шар. Бінарна перехресна ентропія з логітами буде функцією втрат.

Теоретично неглибока NN здатна працювати так само добре або навіть краще, ніж логістична регресія, при умові, що NN не перенавчиться на навчальному наборі [20]. Для запобігання перенавчанню доцільно використати ранню зупинку та зниження швидкості навчання.

Середню AUC кожного набору під час вибору функції потрібно розрахувати за допомогою стратифікованої 5-кратної перехресної перевірки. Потім потрібно порівняти продуктивність кожного набору функцій для вибору оптимального.

2.7 Навчання NN в процесі дослідження алгоритмів машинного навчання в умовах пандемії

Щоб порівняти продуктивність різних алгоритмів машинного навчання, досліджено шість різних моделей. Для вхідних даних моделей машинного навчання було використано найоптимальніший набір функцій. GridCV із стратифікованою 5-кратною перехресною валідацією використано для обширного налаштування гіперпараметрів логістичної регресії, випадкових лісів, XGBoost, машини опорних векторів та дерев рішень.

1. Для навчання була розроблена нейронна мережа, подібна до тієї, що використовується в процесі вибору функцій прогнозування. Кількість нейронів у вхідному шарі дорівнювала оптимальній кількості ознак, визначеній шляхом вибору ознак. При цьому перший і другий приховані шари мають подвійну та рівну кількість нейронів як вхідний шар, відповідно. Задана швидкість навчання 0,00001. Щоб забезпечити належну відповідність навчальним даним терпіння планувальника задано рівним п'ятидесяти. Граничний показник класифікації, який оцінювався щодо набору перевірки та F1-Score, використовувався як показник для порівняння продуктивності. У випадку, якщо декілька відсічок давали однаковий F1-Score – вибиралось граничне значення, найближче до 0,5.

2. Логістична регресія – це модель, яка інтерпретується та добре працює на лінійно розділених простих даних [21]. Процес навчання відбувся за допомогою «лінійного» вирішувача через невеликий розмір набору даних, штраф L1, допуск критеріїв зупинки – заданий рівними 0,0001, обернену силу регуляризації C, задану рівній десяти та масштабування перехоплення задане рівним одиниці.

3. Випадкові ліси (RF) – це надійна техніка на основі дерев, корисна для обробки відсутніх даних та викидів [22]. Метод зберігає свою точність, незважаючи на малі розміри даних [23]. Його навчали за критерієм Джіні з

максимальною глибиною рівною дев'яти, мінімальною кількістю зразків у вузлі листа рівною один та кількістю дерев рівною дев'яносто.

4. XGBoost навчався з цільним набором як логістична регресія, гамою рівною нулю, швидкість навчання 0,2, максимальним кроком дельти рівним нулю, максимальною глибиною рівною чотирьом, мінімальною сумою ваги екземпляра дитини рівною нулю, лямбда-параметром регуляризації L2 – сім та підвибіркою рівною одиниці.

5. Машина опорних векторів (SVM) – це техніка машинного навчання, яка має високу здатність узагальнення та корисна для невеликого розміру даних [24]. При порівнянні чотирьох типів ядер, зокрема, лінійного, «поліноміального», радіальної базисної функції (RBF) і сигмовидного помічено, що «поліноміальне» ядро працює найкраще. Модель SVM була навчена з «полі» ядром з ступенем рівним трьом, гамою масштаб, параметром регуляризації C рівним п'яти та максимальною кількістю ітерацій рівною п'ятиста тисячам.

6. Дерева рішень передбачають значення залежної змінної, вивчаючи прості правила прийняття рішень, виведені з даних [25]. Їх навчали з критерієм «ентропія», максимальною глибиною рівною дев'яти, мінімальною кількістю вибірок, необхідною для розміщення на листковому вузлі, рівною дев'яти, мінімальною кількістю вибірок, необхідною для поділу внутрішнього вузла, рівною двом та випадковим роздільником.

2.8 Тестування NN в процесі дослідження алгоритмів машинного навчання в умовах пандемії

Стратифікована п'ятикратна перехресна перевірка використана для отримання шести навчених та перевірених на різних складках моделей машинного навчання в умовах пандемії [16]. Через невеликий розмір набору даних було обрано п'ятикратну перехресну перевірку для достатньої варіації,

щоб представити основний розподіл. Потім ці моделі були перевірені на тестовому наборі а результати були усереднені для визначення ефективності прогнозування моделі.

Спочатку нормалізували тестовий набір за допомогою «MinMax» масштабування, яке вписувалося в комплект поїзда. Деякі з рядків мали відсутні значення, тому їх було оброблено двома способами:

1. Рядки з відсутніми значеннями для будь-яких чотирьох або більше з п'яти вибраних функцій були усунені. Відсутні значення кожного зразка в отриманому тестовому наборі були розраховані із середнім значенням його відповідних десяти найближчих сусідів у наборі проїздів за допомогою KNN. При цьому найближчі сусіди були знайдені лише щодо вибраних п'яти ознак. Найближчі сусіди були визначені через обернену евклідову відстань між точками даних.

2. Було вилучено всі рядки, для яких відсутнє будь-яке з п'яти значень. Набір не має імпутації та включає лише рядки в яких були всі значення. Це створює тестовий набір, який може моделювати стовідсотковий реальний сценарій тестування.

Потім відбувалось тестування на трьох випадках. Для реалізації справжньої прогнозної продуктивності та її узгодженості моделі були перевірені за допомогою трьох випадків:

1. У випадку $\leq n$, якщо до результату залишилося лише n або менше днів, тестові зразки в яких значення «Кількість днів до результату» було більше n -го дня, були вилучені. Потім було проведено тестування кумуляції решти зразків.

2. У випадку $\geq n$, на n або більше днів наперед, випробувальні зразки для яких значення «Кількість днів до результату» було менше ніж n -й день, були вилучені. Потім було проведено тестування кумуляції решти зразків.

3. У випадку $= n$, рівно за n днів до результату, були обрані тестові зразки, які мали значення «Кількість днів до результату» рівну n -му дню.

2.9 Висновок до другого розділу

В другому розділі кваліфікаційної роботи освітнього-наукового рівня «Магістр» описано формування та попередня обробка пандемічних наборів даних. Розглянуто розподіл даних для унікальної сегрегації пацієнтів. Подано опис конвеєра машинного навчання для дослідження використання алгоритмів в умовах пандемії. Проаналізовано показники оцінки використання алгоритмів машинного навчання в умовах пандемії. Висвітлено важливість функції XGBoost для оцінки біомаркерів. Проведено вибір NN-функції для дослідження алгоритмів машинного навчання в умовах пандемії. Виконано навчання NN в процесі дослідження алгоритмів машинного навчання в умовах пандемії. Подано опис процесу тестування NN в процесі дослідження алгоритмів машинного навчання в умовах пандемії.

3 АНАЛІЗ РЕЗУЛЬТАТІВ ОБЧИСЛЮВАЛЬНОГО ЕКСПЕРИМЕНТУ ДОСЛІДЖЕННЯ АЛГОРИТМІВ МАШИННОГО НАВЧАННЯ В УМОВАХ ПАНДЕМІЇ

3.1 Результати ідентифікації ключових біомаркерів

Враховуючи, серйозність та швидкість поширення COVID-19, значну кількість смертельних результатів, необхідна велика кількість лабораторних тестів для оцінювання стану пацієнтів.

Вибір ознак для визначення найважливіших біомаркерів для оцінки ризиків, є важливою задачею, оскільки отримання меншої кількості лабораторних тестів означає швидший та ефективніший процес прийняття рішень в умовах пандемії. На рисунку 3.1 показано порядок відносної важливості всіх початкових розглянутих ознак [17].

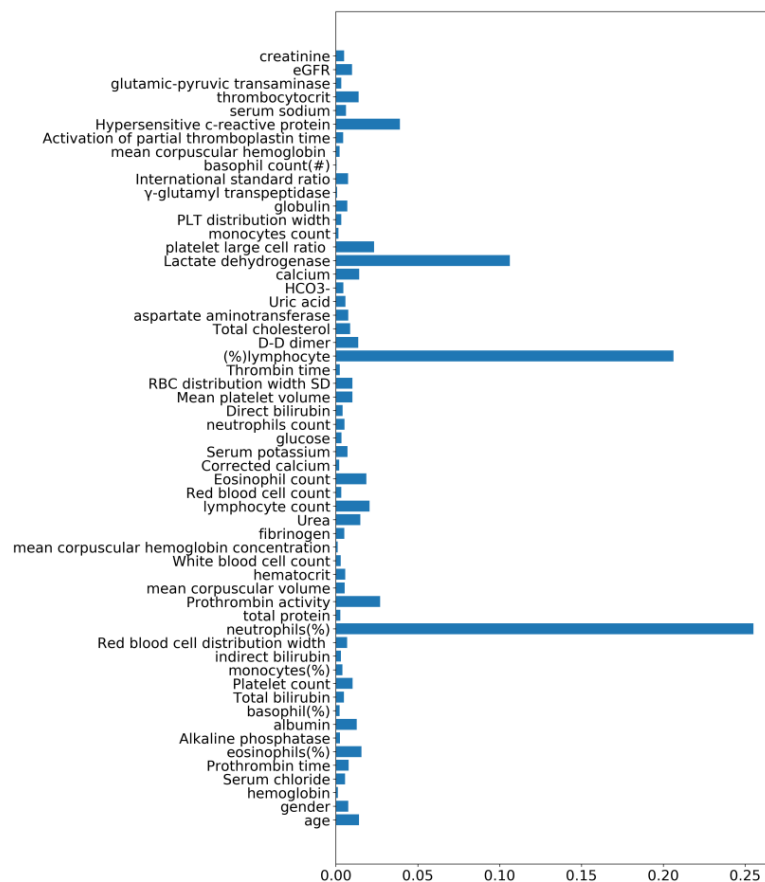


Рисунок 3.1 – Відносна важливість початкових розглянутих ознак

Відносну важливість доступних функцій функції XGBoost визначено за допомогою алгоритму підвищення градієнта XGBoost. При цьому виявлено чотири основні ознаки – «відсоток нейтрофілів», «відсоток лімфоцитів», «ЛДГ» та «hs-CRP». Список функцій відсортовано в порядку спадання важливості. Потім у верхню частину списку важливості ознак додано поле «Вік» (див. рисунок 3.2) завдяки простоті збору даних та дослідженням [2], які демонструють важливість фактора у визначенні прогресування перебігу COVID-19.

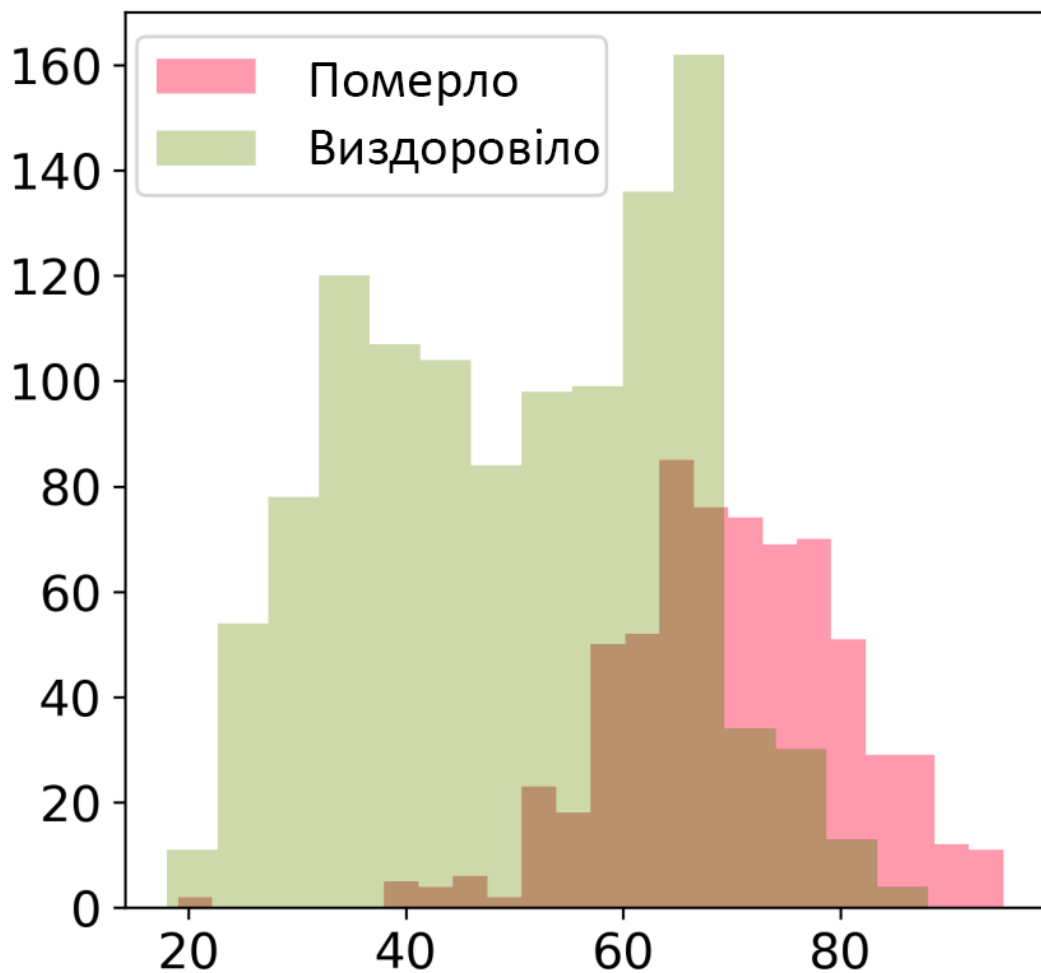


Рисунок 3.2 – Розподіл досліджуваної множини інформації про пацієнтів по віку

Розподіл ЛДГ щодо класів пацієнтів, які вижили та померли подано на додатковому рисунку 3.3.

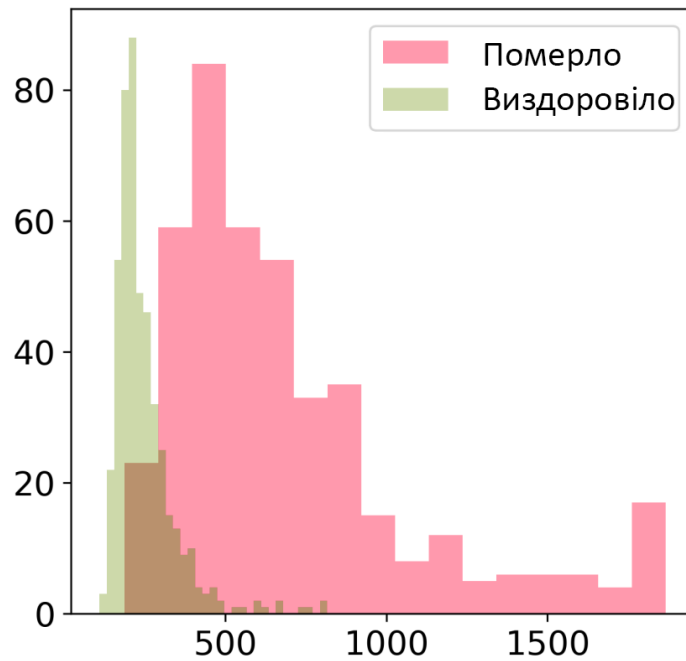


Рисунок 3.3 – Розподіл ЛДГ пацієнтів по класах

Розподіл нейтрофілів щодо класів пацієнтів, які вижили та померли подано на додатковому рисунку 3.4.

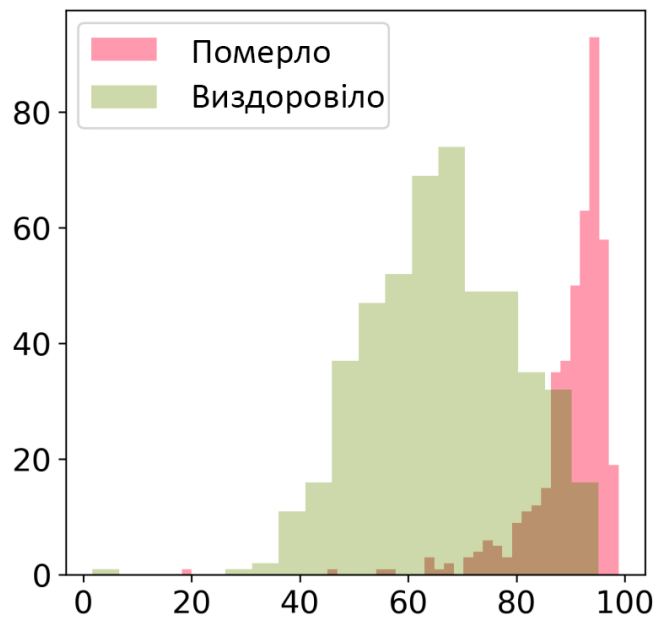


Рисунок 3.4 – Розподіл нейтрофілів пацієнтів по класах

Розподіл лімфоцитів щодо класів пацієнтів, які вижили та померли подано на додатковому рисунку 3.5.

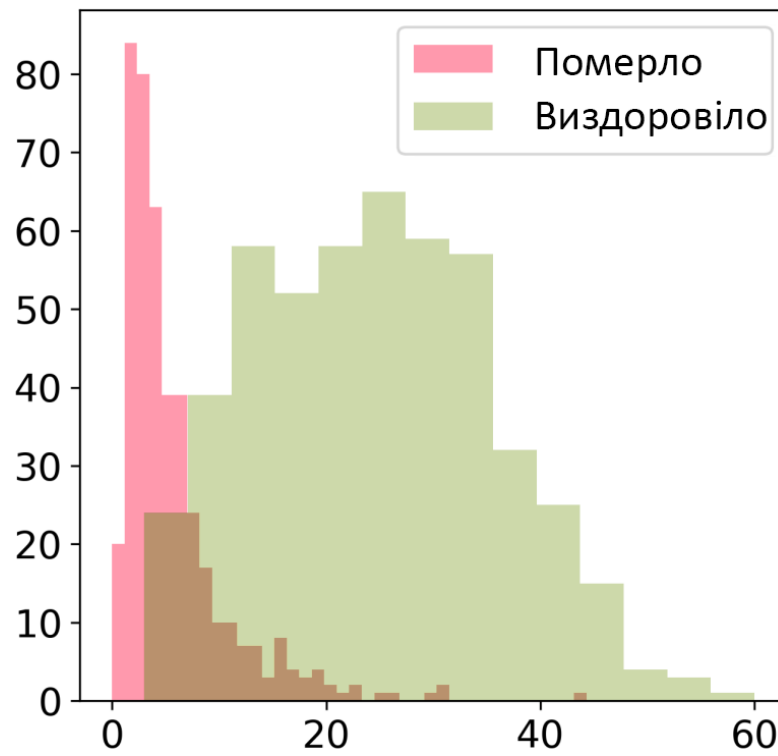


Рисунок 3.5 – Розподіл лімфоцитів піцієнтів по класах

Розподіл hs-CRP щодо класів пацієнтів, які вижили та померли подано на додатковому рисунку 3.6.

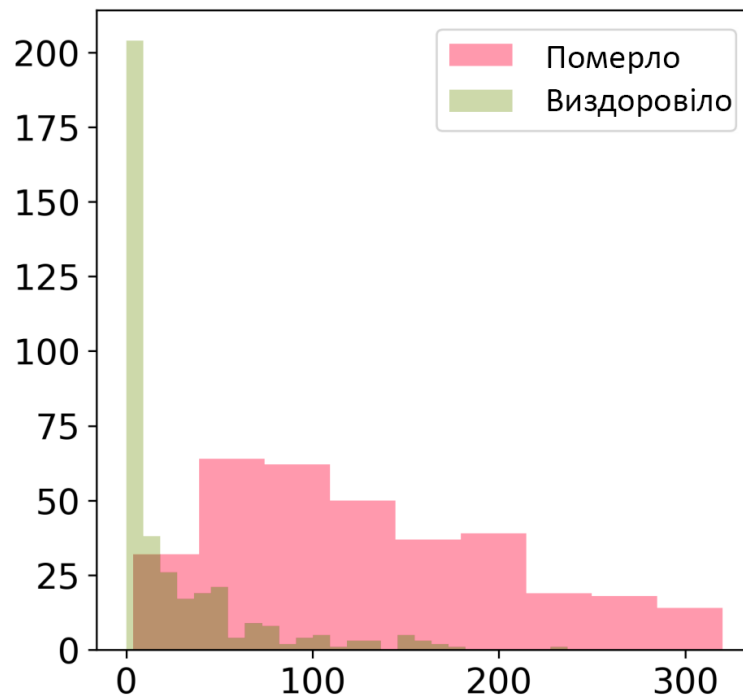


Рисунок 3.6 – Розподіл лімфоцитів піцієнтів по класах

Вибір функцій NN відбувався для оцінки AUC кожного набору функцій під час прямого вибору за допомогою нейронної мережі були нанесені на графік. Метою цієї процедури було максимізувати показник AUC та мінімізувати кількість функцій, вибраних для розробки моделей. Слід відмітити, що середній показник AUC за п'ятьма ознаками становить 0,95.

Кількість вибраних функцій у порівнянні із середнім показником AUC [2], отриманим за допомогою нейронної мережі для вибору ознак подана на рисунку 3.7.

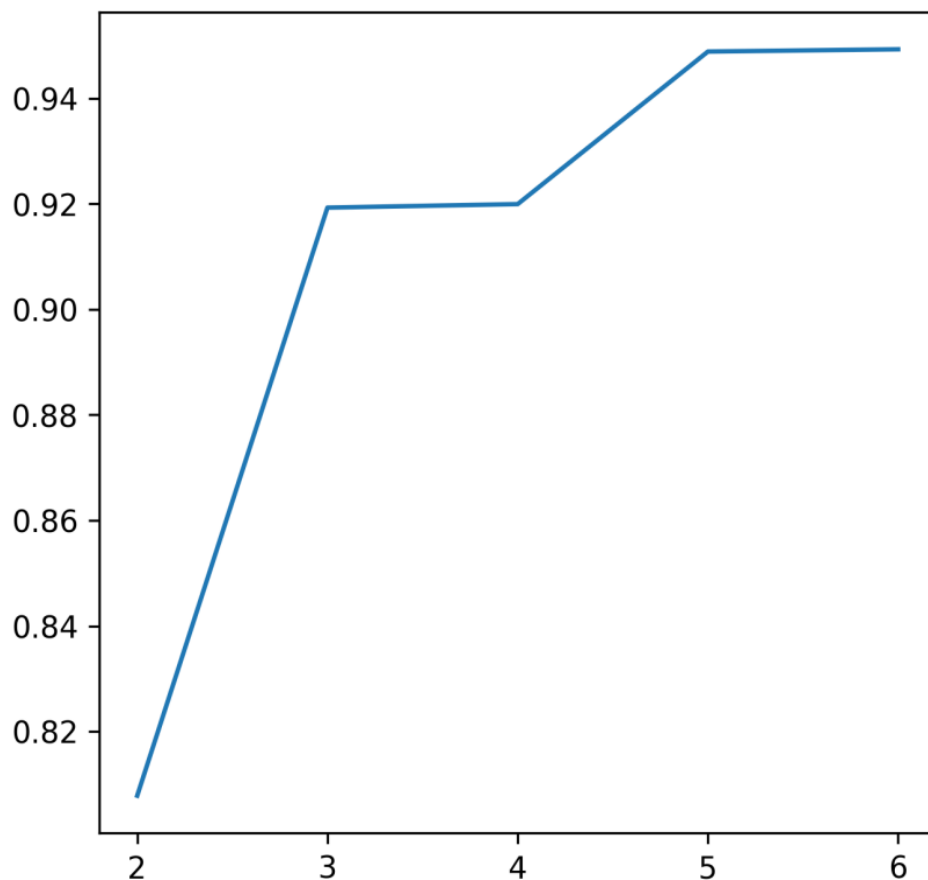


Рисунок 3.7 – Кількість вибраних функцій у порівнянні із середнім показником AUC

Тому, після впорядкування функцій у порядку спадання їх відносної важливості, доцільно використовувати набір з перших п'яти функцій. Зокрема, «вік», «відсоток нейтрофілів», «відсоток лімфоцитів», «ЛДГ» і «hs-CRP». Додавання шостої функції не привело до значного збільшення AUC.

Також слід відзначити, що відмова від функції «вік» заважає продуктивності моделі, водночас додавання функції «стать» не покращує продуктивність моделі.

3.2 Порівняння прогностичних характеристик алгоритмів машинного навчання для потреб систем підтримки прийняття рішень в умовах пандемії

В кваліфікаційній роботі освітньо-наукового рівня «Магістр» проводилось дослідження шести різних алгоритмів машинного навчання для потреб систем підтримки прийняття рішень в умовах пандемії, а саме «нейронна мережа», «SVM», «логістична регресія», «випадкові ліси», «XGBoost» та «дерева рішень». На рисунку 3.8 показано точність [2] досліджуваних для потреб систем підтримки прийняття рішень в умовах пандемії шести різних алгоритмів машинного навчання.

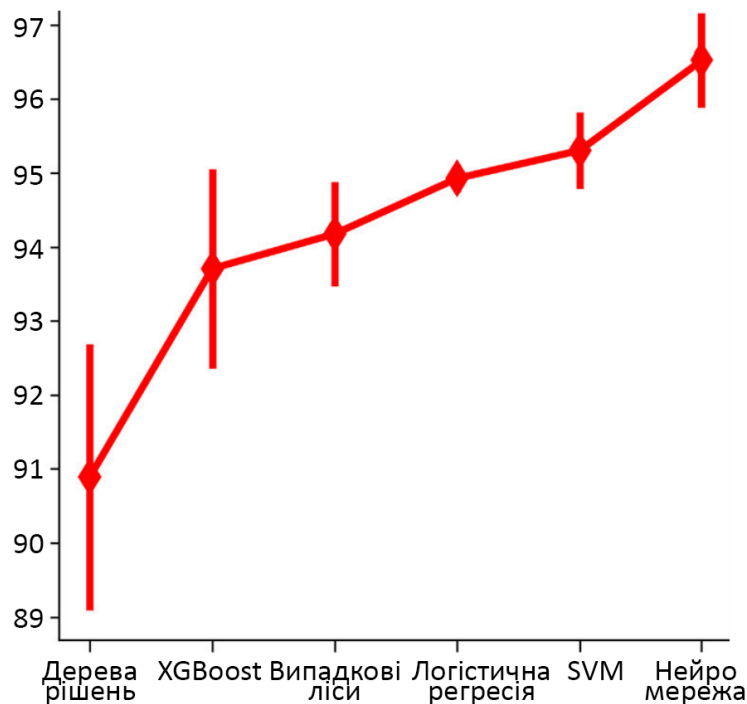


Рисунок 3.8 – Точність досліджуваних для потреб систем підтримки прийняття рішень в умовах пандемії алгоритмів машинного навчання

Спочатку кожен алгоритм після стратифікованої п'ятикратної перехресної перевірки був протестований на тестовому наборі, який містив зразки з усього періоду спостереження пацієнтів.

F1-розподіл та AUC розроблених на основі досліджуваних алгоритмів машинного навчання моделей подано на рисунку 3.9 [2].

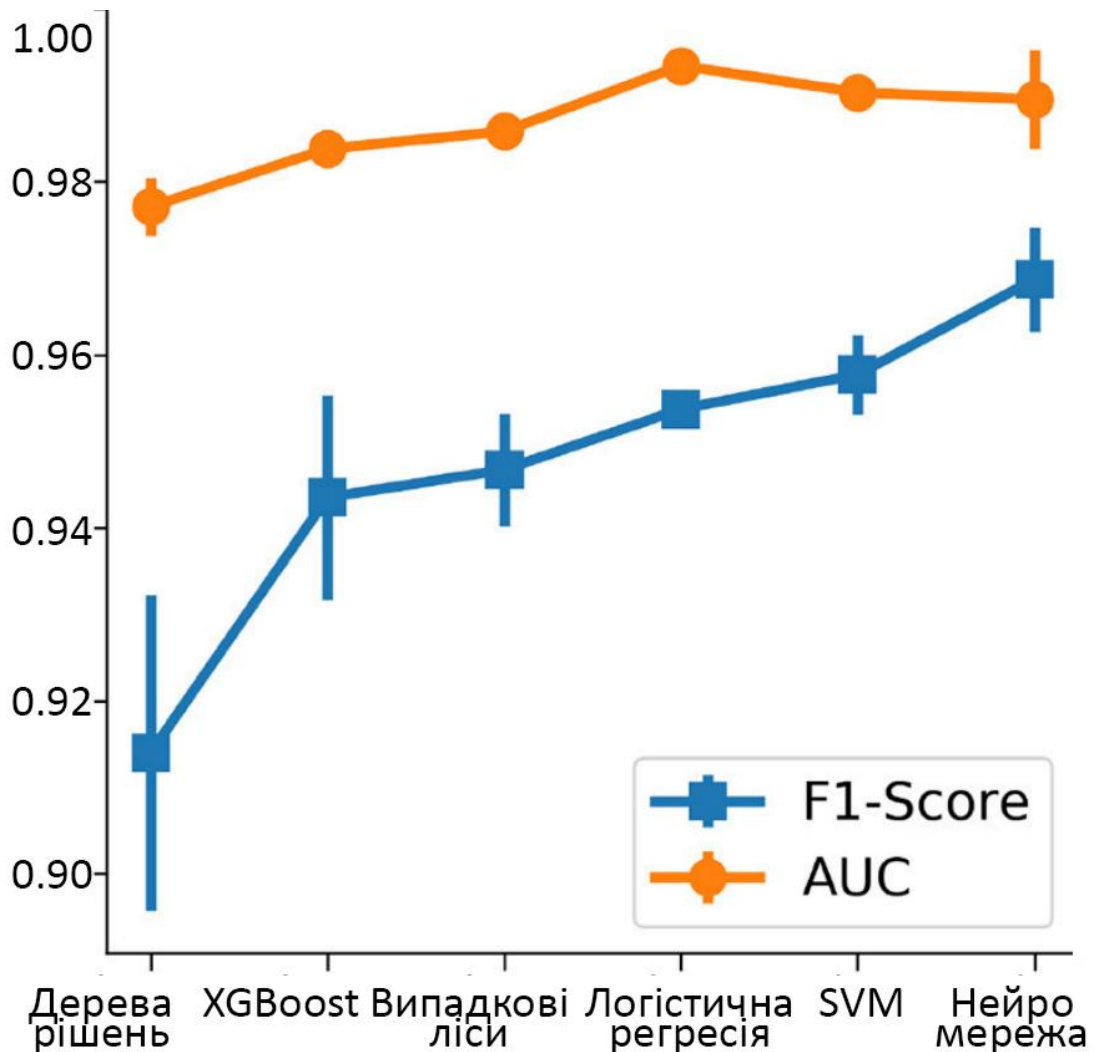


Рисунок 3.9 – F1-розподіл та AUC досліджуваних для потреб систем підтримки прийняття рішень в умовах пандемії алгоритмів машинного навчання

Слід відмітити, що найкращою моделлю є нейронна мережа, оскільки вона продемонструвала [2] кращі результати навчання та функціонування (див. таблицю 3.1).

Таблиця 3.1 – Результати навчання та функціонування досліджуваних для потреб систем підтримки прийняття рішень в умовах пандемії шести різних алгоритмів машинного навчання

Алгоритм (Модель)	Середня точність %	Середній F1-розподіл %	Середня AUC
Нейромережа	96.526 (0.637)	0.9687 (0.006)	0.9895 (0.0057)
SVM	95.305 (0.514)	0.9577 (0.0046)	0.9903 (0.0014)
Логістична регресія	94.929 (0.188)	0.9537 (0.0018)	0.9934 (0.00015)
Випадкові ліси	94.178 (0.703)	0.9467 (0.0065)	0.9858 (0.0020)
XGBoost	93.709 (1.3477)	0.9435 (0.0118)	0.9838 (0.0022)
Дерева рішень	90.892 (1.7963)	0.914 (0.0182)	0.9771 (0.0033)

На рисунку 3.10 подано криві втрат навчання та перевірки нейронної мережі. Оскільки дані не були поділені на пакети то кількість ітерацій відображає епохи [2].

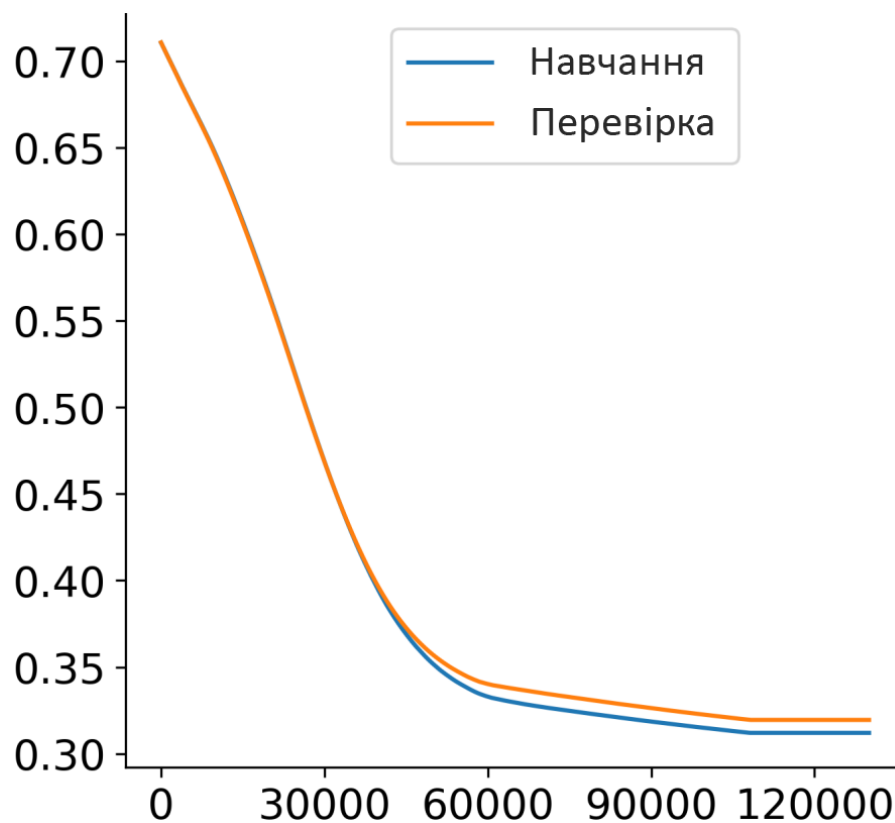


Рисунок 3.10 – Криві втрат навчання та перевірки нейронної мережі

Крива втрат демонструє високий рівень прилягання та має високу AUC-оцінку, що демонструє високу здатність розрізняти ознаки бінарної класифікації.

Надійність і застосовність моделей, сформованих з використанням досліджуваних алгоритмів машинного навчання, в різних налаштуваннях доцільно додатково перевірити трьома способами. Зокрема, в першому випадку досліджується продуктивність моделі нейронної мережі [2], коли тестовий набір складається з даних, лише тоді якщо до дня результату залишилося n або менше днів (див. рисунок 3.11).

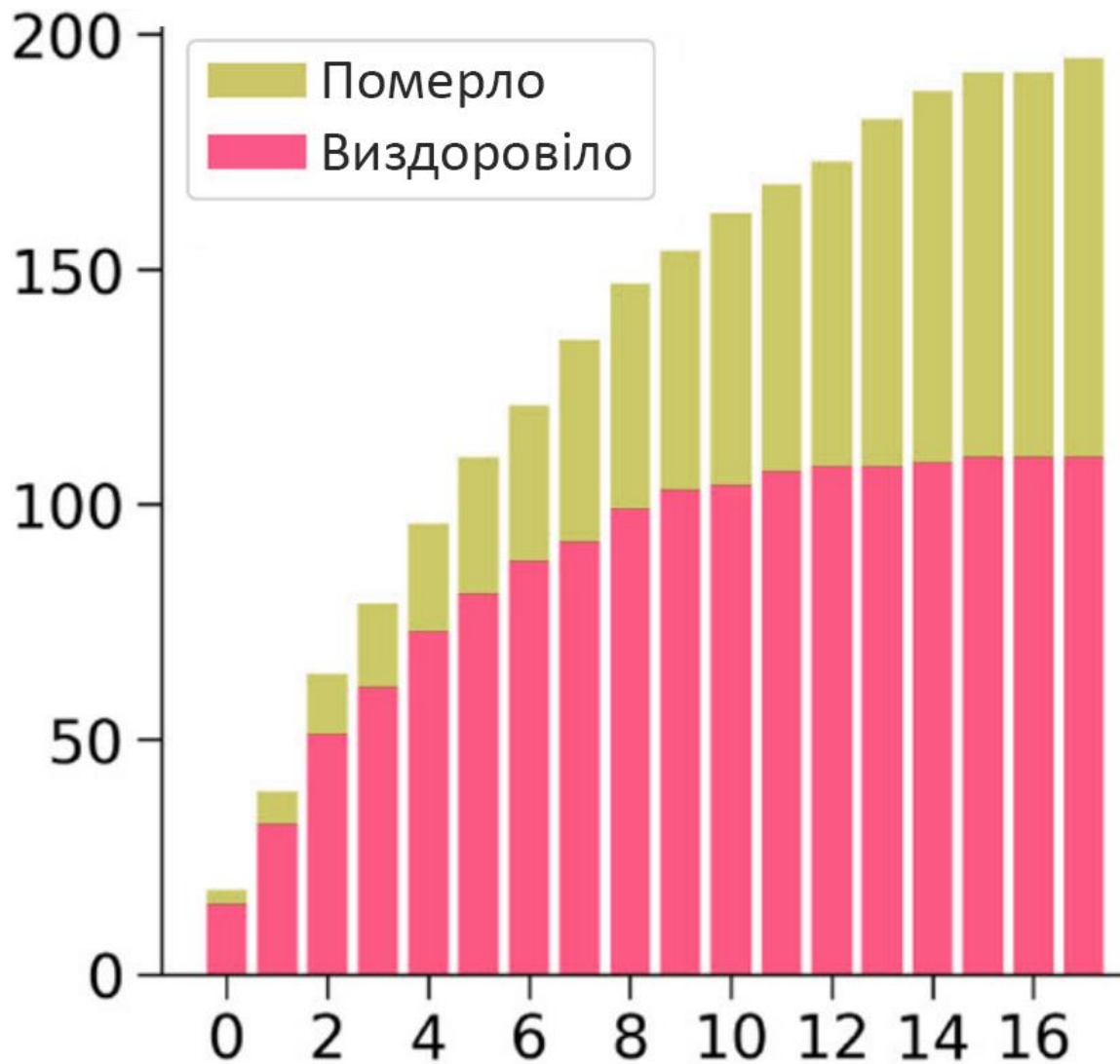


Рисунок 3.11 – Розподіл за класами кумульованих точок даних станом на $\leq n$ -й день для всіх зразків у введеному наборі тестів

Точність моделі оцінюється за різні дні до результату (див. рисунок 3.12). Тестування проводилося щодо різних значень n [2] та помітно, що точність стабільно висока протягом від n до сімнадцяти днів, якщо до результату залишилося лише n або менше днів. Для значення діапазону n від нуля до сімнадцяти днів значення точності були в близькому діапазоні від дев'яносто семи до дев'яносто дев'яти відсотків, що вказує на високі прогностні характеристики моделі. Кількість зразків після сімнадцятого дня є відносно невеликою, щоб значимо вплинути на точність впродовж наступних днів, тому доцільно проводити всі аналізи лише до сімнадцяти днів.

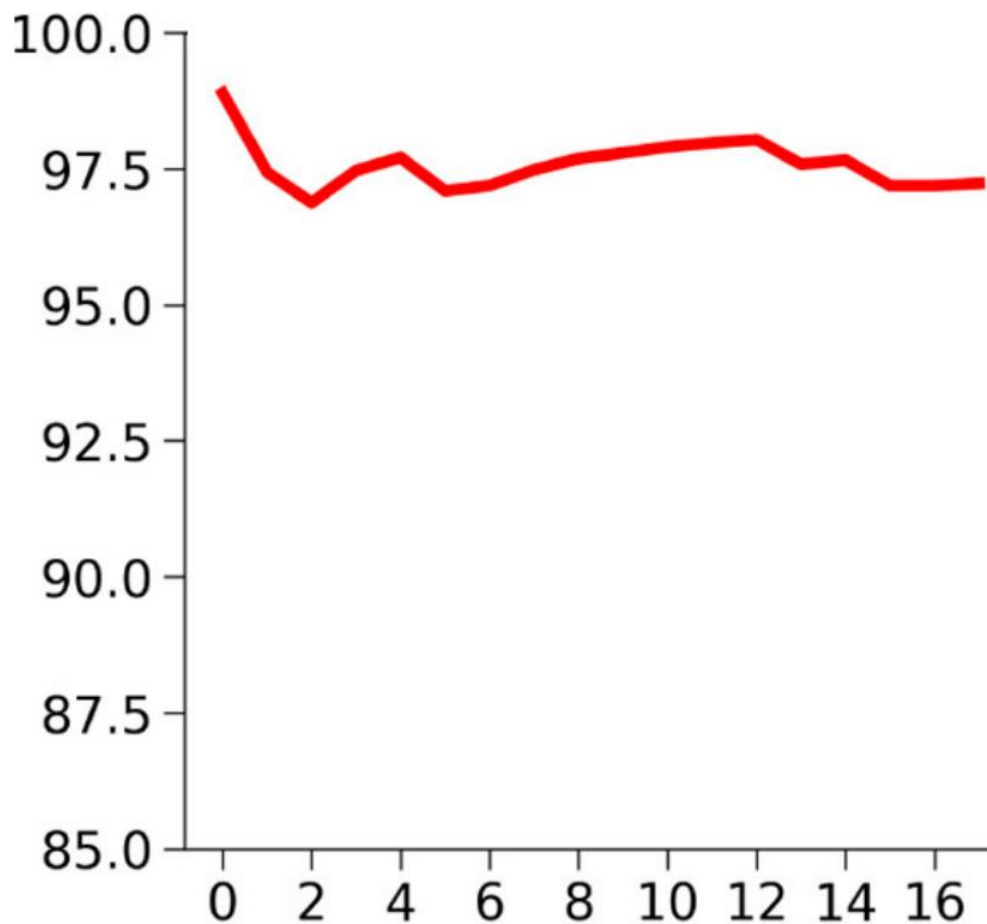


Рисунок 3.12 – Точність моделі нейронної мережі за різні дні до результату

F1-оцінка та AUC моделі оцінювалися за різні дні до результату. На рисунку 3.13 спостерігаються високі та стабільні показники AUC [2], зокрема, найкращий дорівнює одиниці а найгірший нуль цілих дев'яносто

дев'ять сотих. При цьому найкращий розподіл F1 дорівнює 0,99 а найгірший рівний 0,97 у різні дні. Це підтверджує стабільну точність нейромережних алгоритмів щонайменше дев'яносто сім процентів, якщо до результату залишилося будь-яку кількість днів.

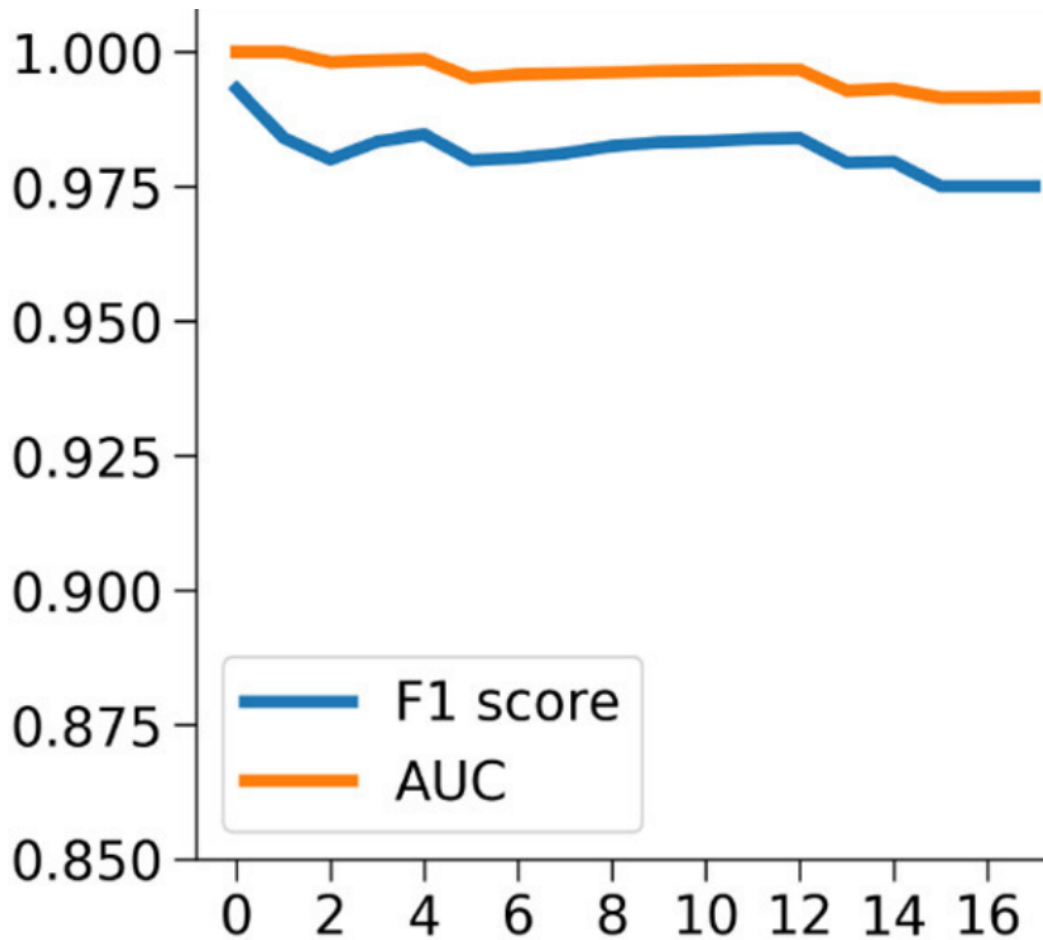


Рисунок 3.13 – F1-оцінка та AUC моделі оцінювалися за різні дні до результату

Недоліком першого випадку те, що вибірки, ближчі до результату, можуть домінувати над результатами, оскільки для них наявно більше ближчих до результату зразків.

В другому випадку при якому досліджується продуктивність моделі на n або більше днів наперед, розглядається для усунення цього недоліку. Невідомо, коли настане день результату, тому важливо проаналізувати ефективність алгоритмів щодо будь-якого дня до дня результату. Це також

допоможе оцінити ефективність алгоритмів протягом усього періоду перебігу захворювання. Оскільки кількість зразків зменшується по відношенню кількості днів до результату, то в кожній сукупності можуть переважати лише зразки найближчого до результату дня. На рисунку 3.14 подано продуктивність нейронної мережі на даних тесту з використанням другого випадку, коли кількість днів до результату більше або дорівнює n [2]. Зокрема подано розподіл за класами кумульованих точок даних в $\leq n$ -й день для всіх зразків у введеному тестовому наборі.

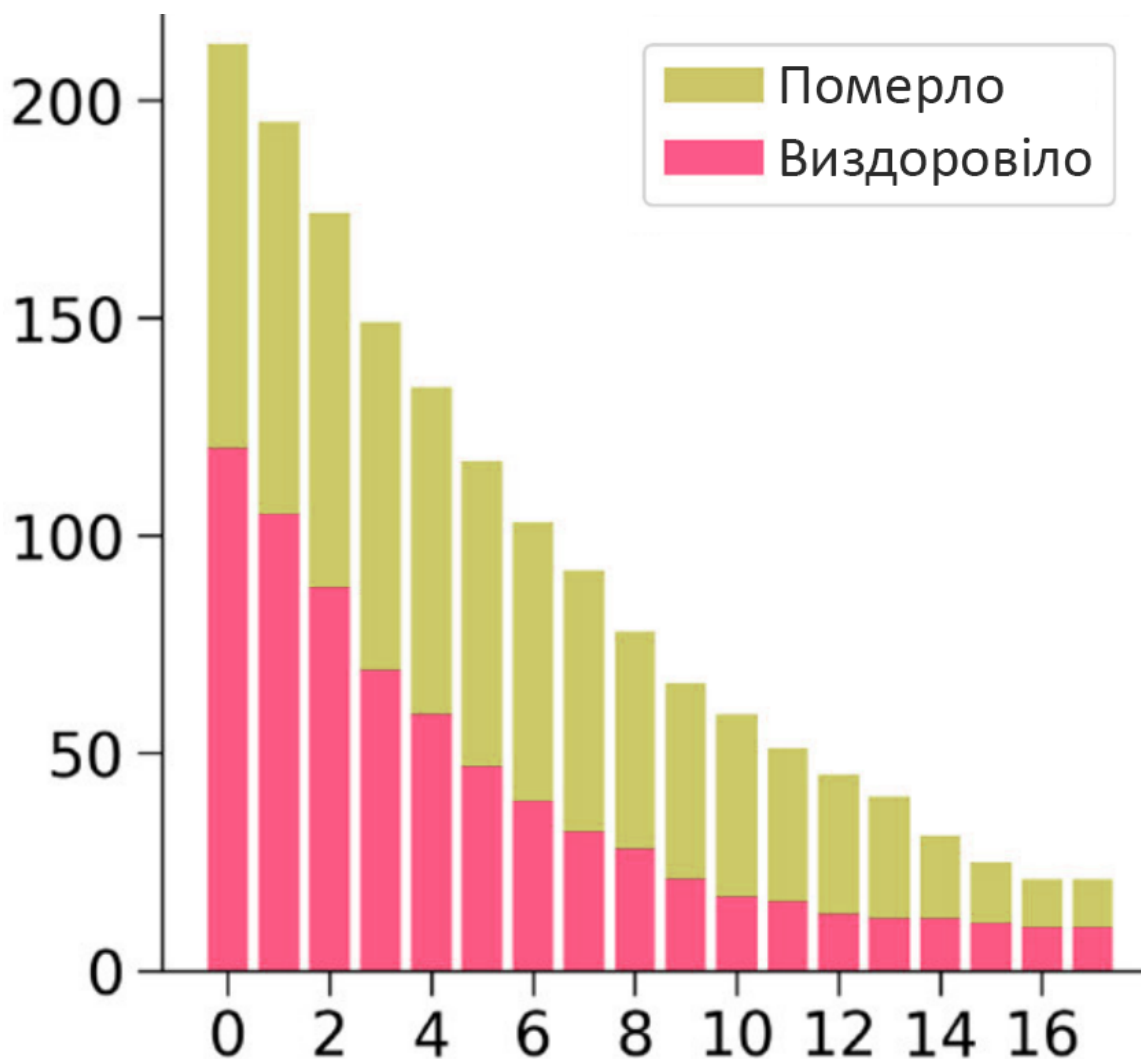


Рисунок 3.14 – Продуктивність нейронної мережі на даних тесту з використанням другого випадку, коли кількість днів до результату більше або дорівнює n

Точність сформованої в другому випадку моделі, в свою чергу, оцінюється за різні дні до результату (див. рисунок 3.15).

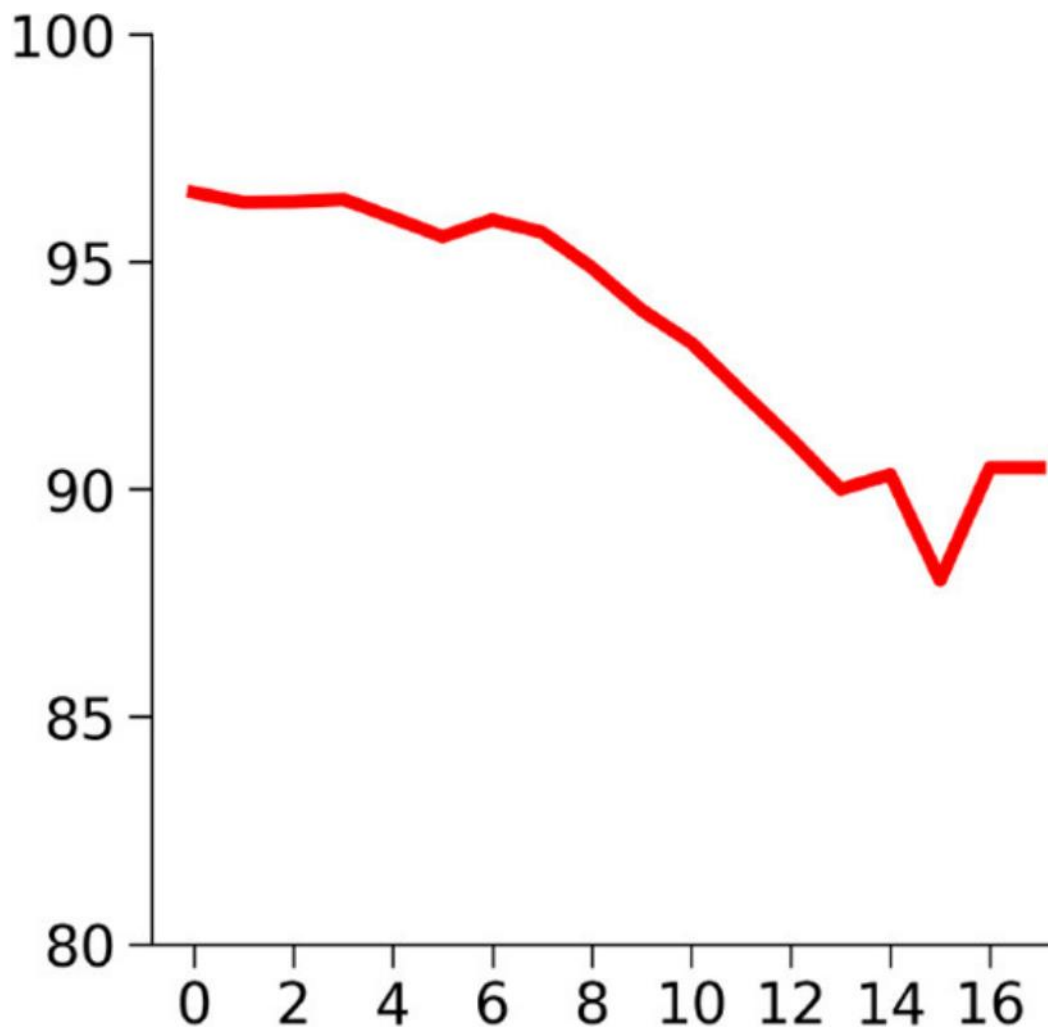


Рисунок 3.15 – Точність моделі нейронної мережі в другому випадку

F1-оцінка і AUC моделі оцінювалися за різні дні до результату в другому випадку подані на рисунку 3.16 [2]. Отже, це дає точніший огляд. Зазначений рис. 3.15 узгоджується з інтуїтивним рівнем, як наслідок стає важче передбачити результат, коли відбувається віддалення від дня результату. Тим не менш, модель починається з високою точністю 96,5 відсотків на день результату та залишається досить стабільною. Найнижча точність вісімдесят вісім відсотків спостерігалася, коли модель прогнозує п'ятнадцять та більше днів наперед.

На рисунку 3.16 подібну консистенцію спостерігали з показниками AUC та F1 [2]. AUC починається від 0,99 в день результату та досягає найнижчої точки на п'ятнадцятий день із значенням 0,96. F1-оцінка стартує з 0,97 та досягає найнижчої точки 0,84 на тринадцятий день. Однак модель здійснює прогнозування досить точно – 95,7 відсотків, при $n \geq 1$ тиждень, що свідчить про високу продуктивність.

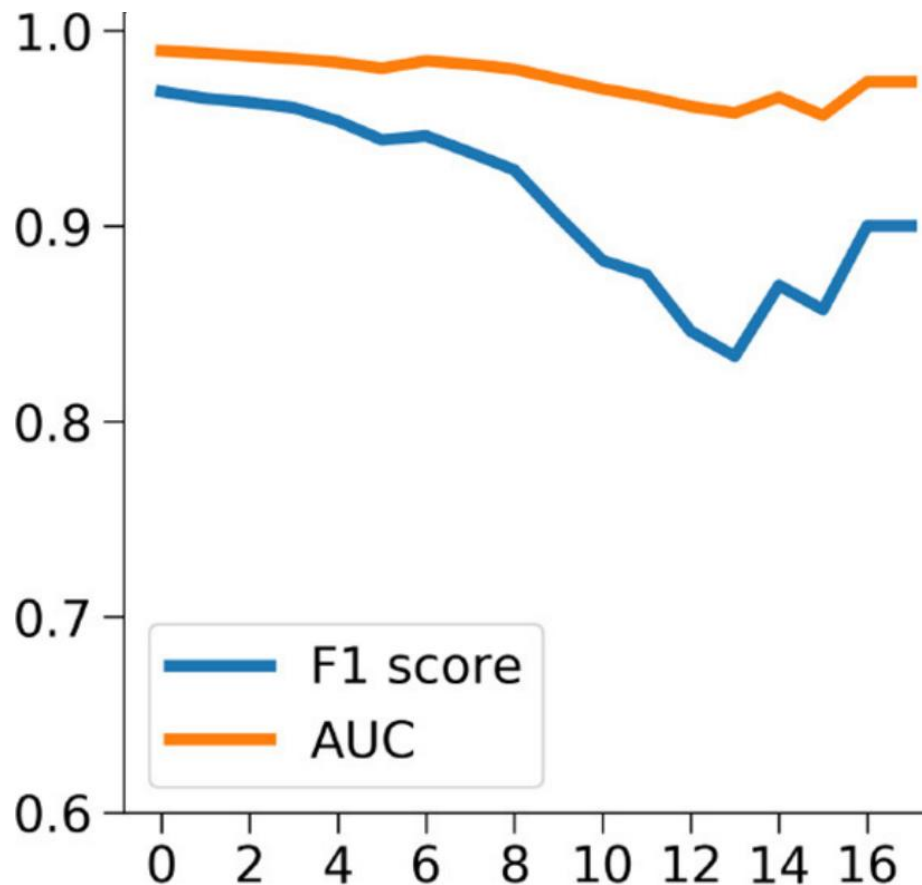


Рисунок 3.16 – F1-оцінка і AUC алгоритму в другому випадку

В третьому випадку, для ще суворішого тестування оцінимо ефективність моделі рівно на n -й день до дня результату. Продуктивність нейронної мережі на даних тесту з використанням третього випадку [2], коли кількість днів до результату дорівнює n подана на рисунку 3.17. Тут жоден зразок з будь-якого іншого дня не може вплинути на результати. Подано розподіл за класами кумульованих точок даних $\leq n$ -й день для всіх зразків у введеному наборі тестів.

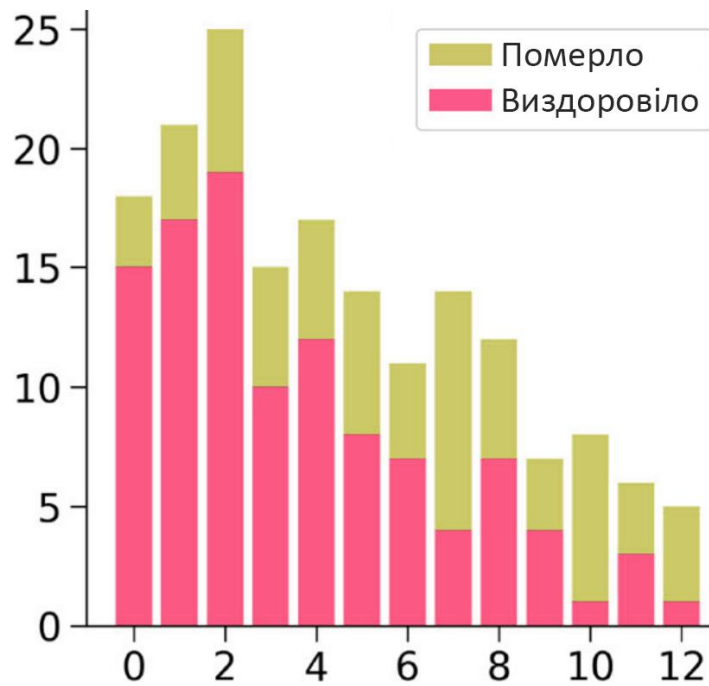


Рисунок 3.17 – Розподіл за класами кумульованих точок даних ($\leq n$ -й день) для всіх зразків у введеному наборі тестів

Незважаючи на те, що точність коливається в залежності від днів, на рисунку 3.18 показано, що з використанням нейромережних алгоритмів прогнозує результат з точністю 98,9% в день остаточного результату та досягає найнижчого значення 92,85% на п'ятий день [2].

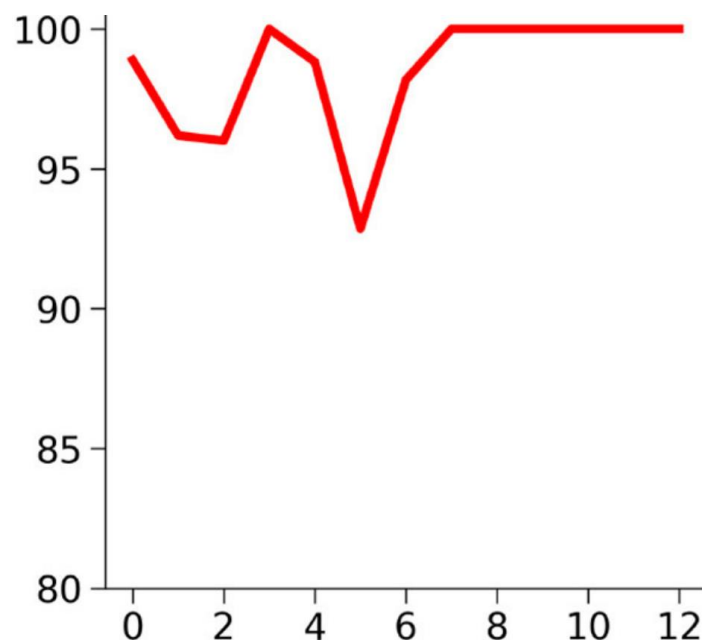


Рисунок 3.18 – Точність нейронної мережі в т ретьому випадку

Оцінки AUC та F1 мають подібну тенденцію. починаючи з одиниці та 0,99, відповідно, у день результату та досягнувши найнижчого показника на п'ятий день із 0,95 та 0,94 відповідно (див. рисунок 3.19 [2]).

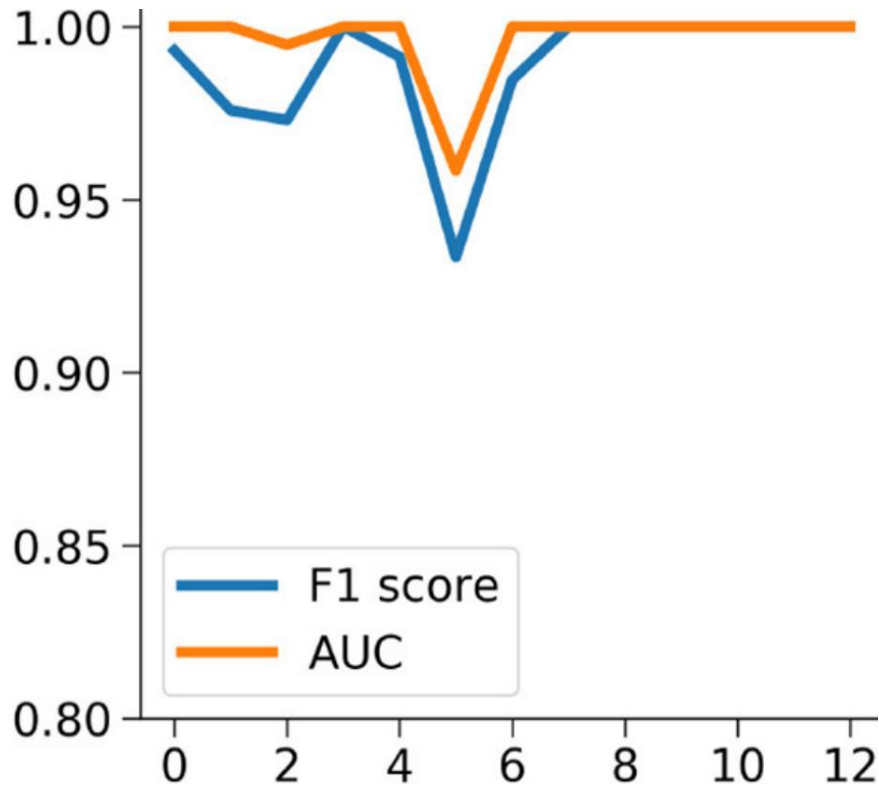


Рисунок 3.19 – F1-оцінка і AUC алгоритму в третьому випадку

Цілком можливо, що стовідсоткова точність з сьомого по дванадцятий дні може бути пов'язана із меншою кількістю аналізованих зразків. Проте слід відзначити, що були присутні обидва класи пацієнтів. Сформована на основі нейромережних алгоритмів модель послідовно передбачала усі випадки коректно з сьомого по дванадцятий день з показниками «AUC» та «F1» на одиницю.

Сформовані на основі алгоритмів машинного навчання моделі були навчені на основі даних навчального набору. Моделі навчалися з використанням шести різних алгоритмів із використанням даних тесту без імпутації, тобто без використання будь-яких синтетичних даних. При цьому вузли даних із відсутніми записами просто видалялися.

3.3 Ефективність алгоритмів машинного навчання

Ефективність методів сформованих на основі алгоритмів машинного навчання для потреб систем підтримки прийняття рішень в умовах пандемії, зазвичай оцінюється на основі декількох вимірювань. Найпопулярнішим показником є точність класифікації. В [1] подано використання інших критеріїв та описані їх математичні формулювання. Зокрема представлені оціночні заходи, які використовуються при оцінці результатів використання алгоритмів машинного навчання для класифікації COVID-19. Кожен набір даних використовує один або декілька комплексних показників оцінювання.

На рисунку 3.20 зображено критерії, які використовуються для оцінки моделей діагностики COVID-19 сформованих на основі алгоритмів машинного навчання. Більше 18 заходів оцінки використовуються найсучаснішими. Очевидно, що точність є основним критерієм оцінки, яким користується більшість дослідників, за яким слідує чутливість.

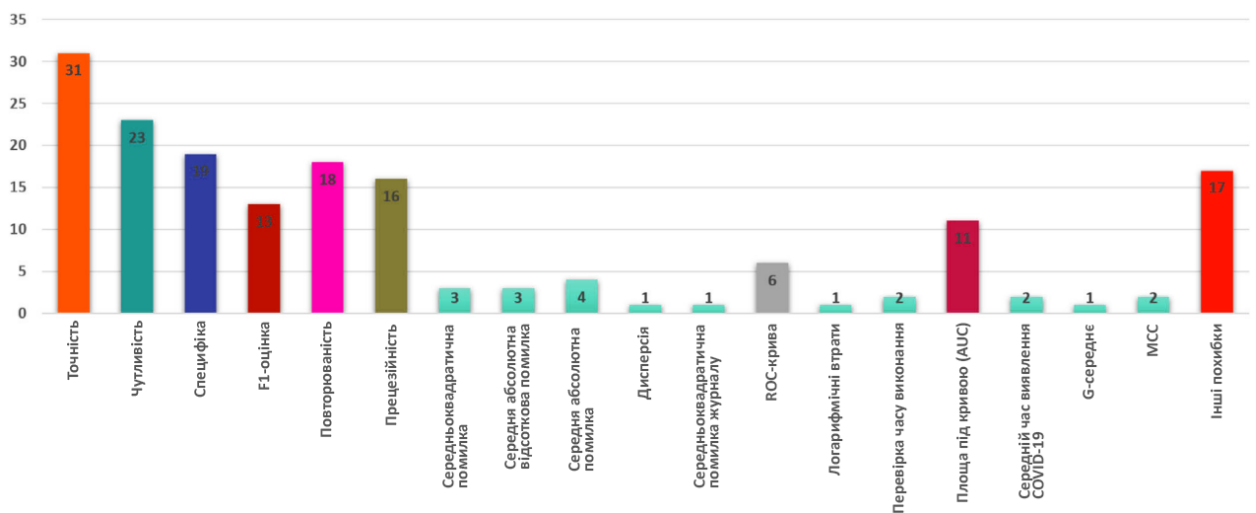


Рисунок 3.20 – Критерії оцінки оцінки моделей діагностики COVID-19 сформованих на основі алгоритмів машинного навчання [1]

На рис. 3.20 показано класифікатори, які використовуються для оцінки методів на сонові алгоритмів машинного навчання потреб систем підтримки

рішень в умовах пандемії COVID-19. Як показано на рисунку 3.21, класифікатор на основі нейромереж є основною технікою, яку використовує більшість дослідників у сфері COVID-19 [1]. Підвищення, K-середнє та логістична регресія займають друге місце за відсотком використання цих класифікаторів у наукових роботах.

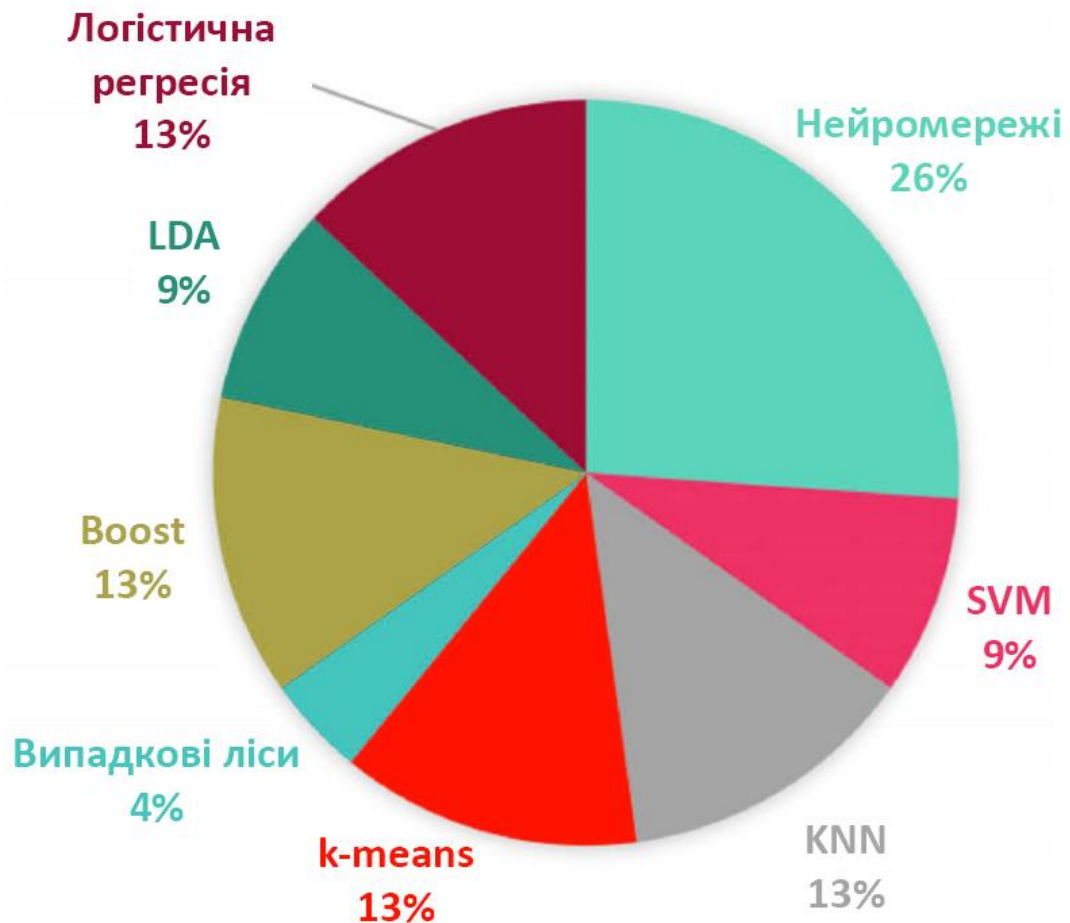


Рисунок 3.21 – Відсоток використання підходів машинного навчання для COVID-19

Як зазначалося методи сформовані на основі алгоритмів машинного навчання інтенсивно використовуються для боротьби з COVID-19. Хоча методи ML та DL показують успішні результати для їх перевірених випадків COVID19, є численні проблеми, які можна розглянути для підвищення якості досліджень у такому напрямку.

3.4 Висновок до третього розділу

В третьому розділі кваліфікаційної роботи описано результати ідентифікації ключових біомаркерів. Виконано порівняння прогностичних характеристик алгоритмів машинного навчання для потреб систем підтримки прийняття рішень в умовах пандемії. Нейронна мережа прогнозує перебіг інфекційних захворювань в умовах пандемії з точністю дев'яносто шість відсотків для всіх днів протягом періоду захворювання і з точністю дев'яносто відсотків для більш ніж шіснадцяти днів наперед. Важливість розглянутої в попередньому розділі функції XGBoost забезпечує інтерпретацію моделі, яка може бути актуальною в клінічних умовах. Надійність досліджених алгоритмів машинного навчання була ретельно перевірена за допомогою трьох різних сценаріїв. Досліджено ефективність алгоритмів машинного навчання в умовах пандемії.

4 ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ

4.1 Організація та перевірка стану охорони праці на підприємстві під час пандемії

Кваліфікаційна робота освітньо-наукового рівня «магістр» присв'ячена дослідженню алгоритмів машинного навчання для потреб систем підтримки прийняття рішень в умовах пандемії. Машинне навчання – це інноваційна підмножина систем штучного інтелекту в галузі інформатики. Вона доволі часто використовує статистичні методи, підходи та прийоми для реалізації алгоритмічних засобів «навчання» штучних інтелектуальних систем з метою поступового покращення їх продуктивності. Зокрема, систем підтримки прийняття різнотипових рішень в умовах пандемії, запит на розвиток яких сформовано суспільством внаслідок швидкого поширення коронавірусних інфекційних захворювань. Розробкою та інформаційно-технологічним супроводом систем підтримки прийняття рішень переважно займаються ІТ-компанії різного розміру та складу.

На кожному підприємстві, установі, організації чи ІТ-компанії, відповідно до статті 10 Закону України «Про захист населення від інфекційних хвороб» від 6 квітня 2000 р. № 1645-III, потрібно та необхідно неухильно дотримуватись санітарно-гігієнічних та санітарно-протиепідемічних норм та правил при виконанні будь-яких видів діяльності.

Згідно статті 7 Закону України «Про забезпечення санітарного та епідемічного благополуччя населення» від 24 лютого 1994 року № 4004-XII, установи, організації та підприємства (в тому числі ІТ-компанії), зобов'язані розробляти, запроваджувати та неухильно здійснювати санітарні та протиепідемічні заходи.

Згідно вимог «Директиви 89/391/ЄЕС Ради про введення заходів, що сприяють поліпшенню безпеки і гігієни праці працівників на виробництві» та

ДСТУ OHSAS 18001:2010 «Системи управління гігієною та безпекою праці», розроблені та запроваджені заходи повинні містити належним чином реалізовані процедури оцінювання ризику на робочих місцях, які охоплюють весь перелік визначених вітчизняним та міжнародним законодавством біологічних загроз та небезпек [26].

Відповідно до рекомендацій Всесвітньої організації охорони здоров'я (ВООЗ) на кожному підприємстві, організації, установі та в кожній ІТ-компанії необхідно сформувавши план щодо запобігання захворювання. Зазначений план повинен містити [27]:

- Заходи для забезпечення працівників ІТ-компанії спеціальними дезінфекційними засобами. Засоби повинні розміщуватись поблизу робочих місць. Працівників ІТ-підприємства потрібно регулярно інформувати щодо потреби часто мити руки.

- Забезпечення співробітників ІТ-підприємства засобами індивідуального захисту (ЗІЗ). Для утилізації ЗІЗ, в разі потреби, слід встановити спеціальні контейнери.

- Обмеження особистого спілкування, без дискримінації. Обмеження перебування персоналу та працівників ІТ-компанії, що належать до груп підвищеного ризику, в місцях скупчення персоналу. Зокрема, громадяни, для яких захворювання спричиняє найбільший ризик – це особи з легеневиими або серцевими захворюваннями, цукровим діабетом, люди похилого віку.

- Організація систематичних перевірок працівників перед початком робочої зміни з метою виявлення співробітників з ознаками або симптомами захворювання. Слід звертати увагу на легкий кашель, підвищену температуру тіла. При виявленні даних симптомів громадянина слід негайно скерувати на прийом до сімейного лікаря, або додому.

- Повинно бути здійснено розробку та виконано розміщення інформаційних плакатів, які повідомляють про небезпеку інфекційного захворювання, шляхи його розповсюдження, методи та способи захисту. На

плакатах слід розмістити інформацію про особливості використання ЗІЗ, змінного одягу, заходи для дотримання гігієни рук, а також заохочення працівників ІТ-компанії залишатися вдома та звертатися до сімейного лікаря у випадку появи симптомів інфекційного захворювання.

- Розроблення гнучкого графіка роботи персоналу та працівників ІТ-компанії з метою мінімізації скупчення громадян в одній локації.

- Розташування одноразових гігієнічних серветок поряд з часто відвідуваними місцями та предметами, які часто використовують. Зокрема, дверними ручками, клавіатурами, пультами дистанційного керування. Це повинно бути зроблено, щоб працівники могли регулярно протирати ці предмети перед кожним використанням.

- Систематичне очищення та дезінфекція території ІТ-компанії, умтанови, організації чи підприємства. Особливо в місцях масового скупчення громадян. Організація провітрювання приміщень.

Якщо в ІТ-компанії, організації, установі чи на підприємстві виявлено громадян інфікованих вірусом SARS-Cov-2 або хворих на COVID-19 слід послугоуватись рекомендованим планом дій:

- Провести дезінфекція та очищення. При цьому необхідно організувати обов'язкове очищення та дезінфекцію приміщень, в яких існує підозра або виявлені підтверджені випадки інфікування. Всі поверхні, з якими контактувала інфікована або хвора особа, слід ретельно очистити та дезинфікувати. Зокрема, дверні ручки, клавіатури, посуд, санвузли, телефони та загальнодоступні зони, де інфікований або хворий громадянин міг переміщатися чи перебувати. Це також стосується вбиралень та коридорів, де він міг бути навіть впродовж короткого періоду часу.

- Слід ізолювати громадянина. При цьому для зменшення ризику розповсюдження інфекції доцільно виділити окреме ізольоване приміщення для розташування хворого до того моменту прибуття на виклик бригади центру екстреної медицини та медицини катастроф – «швидка допомога».

– У разі підозри щодо інфікування одного з працівників ІТ-компанії, підприємства, установи чи організації необхідно організувати окреме робоче місце, розташоване на відстані не менше ніж 2 метри від інших робочих місць. Якщо є можливість перемістити його в окреме приміщення або обмежену зону з можливістю періодичного провітрювання приміщення.

При підтвердженні підозри:

– необхідно обов'язково звернутися до відповідних структур для організації дезінфекції приміщень.

– Громадяни, які контактували з інфікованою особою, повинні бути відправлені на самоізоляцію.

– Слід здійснити інформування. Роботодавці повинні обов'язково інформувати працівників про виявлені випадки захворювання на COVID-19 або SARS-Cov-2 на робочих місцях. Але при цьому слід зберігати конфіденційність. Поширення інформації про випадки інфікування може викликати занепокоєння або спричинити виникнення стресу в колег інфікованих громадян. Тому потрібно організувати інформування з надійних джерел. При цьому потрібно постійно підтримувати з контакт зі співробітниками та при можливості дозволити працювати згідно гнучкого графіку або дистанційно.

– Слід провести утилізацію одягу. Всі одноразові ЗІЗ та одяг хворої особи слід запакувати в поліетиленові мішки. Щільно запакувати, а потім помістити їх в ще один поліетиленовий мішок. Його слід щільно запакувати, зробити відповідну позначку та зберігати в безпечному місці до отримання остаточного результату досліджень. У випадку отримання негативних результатів тестування можна утилізувати разом зі звичайними відходами. При отриманні позитивних результатів тестів слід звернутися у відповідні організації для отримання алгоритму подальших дій.

Держпраці наголошує, що у зв'язку з пандемією потребує змін інструкція з охорони праці [28]. Порядок та особливості внесення змін до

актів з безпеки праці та здоров'я працівників ІТ-компаній, установ, організацій та підприємства регламентується Порядком опрацювання і затвердження роботодавцем нормативних актів з охорони праці, що діють на підприємстві, затвердженим наказом Державного комітету України по нагляду за охороною праці 21.12.1993 № 132 та зареєстрованим в Міністерстві юстиції України 07.02.1994 за № 20/229.

До опрацювання проєкту нормативного акту з безпеки праці та здоров'я працівників залучаються кваліфіковані фахівці в галузі охорони праці та компетентні щодо правових питань, представники профспілок та громадських об'єднань, уповноважені найманими працівниками особи з питань охорони праці.

При необхідності роботодавці можуть залучити на договірній основі до розроблення та опрацювання проєктів нормативних актів фахівців сторонніх організацій. Зокрема, щодо змін до актів підприємств, що містять обов'язкові для дотримання працівниками специфічних вимог з охорони праці при виконанні ними робіт певного виду або за певною професією на всіх робочих місцях та інструкцій з охорони праці. Держпраці нагадує, що перегляд інструкцій з охорони праці проводиться відповідно до передбачених нормативно-правовими актами термінів, на основі яких їх розроблено. Проте не рідше ніж один раз на п'ять років. Для робіт з підвищеною небезпекою або там, де є потреба у професійному доборі чи підготовці, це повинно відбуватись не рідше ніж один раз на три роки. Проте пандемія та карантинні обмеження зумовили необхідність внесення змін до інструкцій з охорони праці з метою доповнення їх поданими вище в цьому параграфі правилами поведінки.

Забезпечення, організація та своєчасна перевірка стану охорони праці в ІТ-компаніях під час пандемії дозволить підвищити ефективність запровадження розглянутих в кваліфікаційній роботі алгоритмів машинного навчання для потреб систем підтримки прийняття рішень.

4.2 Основні принципи і способи забезпечення життєдіяльності

Кваліфікаційна робота освітньо-наукового рівня «магістр» присв'ячена дослідженню алгоритмів машинного навчання для потреб систем підтримки прийняття рішень в умовах пандемії. Пандемія COVID-19 спричинила обширний перелік суттєвих змін в усіх сферах людського життя та діяльності. Окрім того, ситуацію ускладнила війна в Україні. Тому актуальним питанням розділу «Безпека в надзвичайних ситуаціях» є дослідження основні принципів і способів забезпечення життєдіяльності.

Можна виділити ряд важливих проблем, які постійно перебувають у полі зору людства для забезпечення нормальних умов життя і праці [29].

1. Дотримання параметрів середовища перебування людини в необхідних для життєдіяльності межах. Це одна із складних проблем, які стоять перед світовим співтовариством, пов'язана з тим, що трудова діяльність людей з року в рік активізується, ускладнюється, спричиняє введення новіших знарядь праці та технологій.

Як наслідок загострюється проблема технологічної безпеки та збільшується навантаження на всі структурні частини навколишнього середовища. Є очевидною небезпека виснаження природних ресурсів, незворотних забруднень і зміни безпечних параметрів середовища, за якими створюються реальні умови для виникнення різного роду небезпек.

Отже, кожна держава повинна мати професійно придатні структури і системи захисту від наслідків імовірних небезпек. Головною метою таких систем є захист населення та зниження рівня ризику при виникненні певних небезпек шляхом запобігання, реагування і ліквідації їх наслідків.

2. Забезпечення населення всіма видами енергоресурсів (електроенергією, газом, нафтопродуктами, кам'яним вугіллям, водою тощо). Енергетична криза, що існує на даний час, суттєво впливає на

життєдіяльність людей. Це одна з найбільш актуальних проблем забезпечення безпеки будь-якої країни світу.

3. Забезпечення населення всіма необхідними параметрами і нормами матеріального середовища життя. Гострою проблемою для багатьох людей у різних країнах є недостатня кількість житла, комунального транспорту, суспільних закладів, спортивних комплексів, медичних закладів та інших елементів системи життєзабезпечення.

4. Забезпечення продуктами харчування. Продукти харчування є фізіологічною основою життєдіяльності. Із збільшенням чисельності населення ця проблема стає особливо гострою. Якщо людство не розробить нові перспективні технології вирощування продуктів харчування і своєчасно не адаптується до них, може виникнути небезпечна ситуація глобального масштабу.

5. Наявність і раціональне використання питної (прісної) води. Йдеться про охорону прісної води від забруднення, що може призвести до непридатності її використання для потреб населення. Звідси впливає важливість очищення води, боротьба з промисловим і побутовим забрудненням, виснаженням водою.

6. Ліквідація, переробка або використання відходів виробництва. Особливо небезпечними є відходи атомних, хімічних, біологічних виробництв, кількість яких щорічно зростає і, відповідно, збільшується кількість відходів.

Тому основні проблеми життєдіяльності людини зводяться до комплексу заходів щодо захисту, раціонального використання природних ресурсів і відновлення живої природи. Зокрема збереження та відновлення рослинного та тваринного світу, його різноманіття. Комплекс заходів щодо захисту, раціонального використання природних ресурсів також включає збереження та відновлення неживої природи. Зокрема, ґрунтів, водних ресурсів, атмосфери, надр, кліматичних умов тощо.

Розглянемо основні принципів забезпечення життєдіяльності [30]:

1. Безперервне забезпечення фізіологічних процесів організму людини, що залежить від таких факторів: повітря, питна вода, продукти харчування, тепло, світло, предмети споживання житло, одяг, взуття.

2. Принципи взаємозв'язку і взаємозалежності з навколишнім середовищем. Життєдіяльність забезпечується такими факторами навколишнього середовища, як параметри споживання, енергоресурси, корисні копалини, продукти харчування, елементи штучного середовища.

Водночас, життєдіяльність впливає на середовище життя, змінює та регулює параметри споживання, виснажує енергоресурси, корисні копалини, видозмінює клімат, рослинний і тваринний світ, забруднює навколишнє середовище.

Внаслідок того, що науково-технічний прогрес ще не досяг такого розвитку, щоб усі технологічні процеси були безпечними, безвідходними і безаварійними, вірогідність виникнення техногенних і технологічних криз не виключається.

3. Принцип раціональної організації праці за метою, часом, місцем і нормами. Належна організація праці включає управління, принципи організації, цілі і завдання, засоби праці, виробничу діяльність і результати праці. Порушення норм праці, технологічних процесів, фізичне зношення засобів виробництва, нанесення моральної шкоди, як правило, призводять до аварійних ситуацій.

4. Принцип матеріального заохочення при організації життєдіяльності, що безпосередньо пов'язаний з продуктивністю праці, яка визначається:

- людським фактором (способом матеріального заохочення);
- працездатністю виробничого персоналу;
- ступенем підготовленості до праці (професійним, фізіологічним, психологічним).

– певний вплив на продуктивність праці мають також індивідуальні особливості працівників, їхні фізіологічні і психологічні можливості, параметри навколишнього середовища, технічні та організаційні умови.

Порушення уваги, послідовності виконання технологічних операцій, норм і вимог до технічної документації, низький рівень професійної підготовки також можуть призвести до виникнення надзвичайних ситуацій.

5. Принцип захисту здоров'я, меж і умов життєдіяльності. Для реалізації цього принципу людство створило спеціальні інститути: медичного забезпечення, оборони, екологічного захисту, моралі. Окремі інститути як структурні частини життєдіяльності можуть створюватись для захисту людей і народного господарства в особливих (надзвичайних) ситуаціях. До них можна віднести: цивільну оборону, міністерство з надзвичайних ситуацій, комісії з питань техногенно-екологічної безпеки і надзвичайних ситуацій, штаби цивільної оборони.

6. Ліквідація негативних наслідків життєдіяльності. У вищеназваних принципах зазначені випадки, при яких життєдіяльність окремих груп людей і населення Землі загалом супроводжується небезпечними ситуаціями техногенного, природного, соціально-екологічного характеру. Тому кожна держава повинна мати професійно придатні структури для ліквідації наслідків небезпечних ситуацій.

Проаналізуємо способи забезпечення життєдіяльності. Кожен принцип має певну кількість способів його реалізації [31]. Найбільш типовими з них є:

1. Усі види трудової діяльності, практики.
2. Способи розробки і створення засобів праці.
3. Забезпечення життєдіяльності людини параметрами штучного середовища.
4. Здоровий спосіб життя, профілактика здоров'я, відпочинок.
5. Оптимальне обмеження параметрів споживання ресурсів, забезпечення та їх поповнення.

6. Функціонування захисних систем у звичайних умовах і в умовах виникнення різних небезпек.

Основні функції БЖД повинні забезпечити безпеку життєдіяльності людини, охорону навколишнього середовища через:

- опис життєвого простору, його зонування за значеннями негативних чинників на основі експертизи джерел негативних впливів, їх взаємного розташування та режиму дії, а також із урахуванням кліматичних, географічних та інших особливостей регіону або зони діяльності;

- формування вимог безпеки до джерел негативних чинників, їх гранично допустимих викидів, енергетичних впливів, допустимого ризику тощо;

- організацію моніторингу стану навколишнього середовища та контролю джерел негативних впливів;

- розробку та використання систем захисту;

- реалізацію заходів щодо запобігання та ліквідації наслідків надзвичайних ситуацій;

- навчання населення основам БЖД, підготовку спеціалістів усіх рівнів і видів діяльності.

Отже, можна зробити висновок, що прийнятні умови життєдіяльності людини великою мірою ґрунтуються на забезпеченні належних умов безпеки перебування людини у навколишньому середовищі. А поява наукової галузі знань про безпеку життєдіяльності людини – це закономірний процес розвитку суспільства, який обумовлюється, з одного боку, бурхливим розвитком науково-технічного прогресу, процесами глобалізації і створенням комфортних умов життя для людини, а з іншого – зростанням різного роду небезпек. В тому числі спричинених пандемією COVID-19 та бойовими діями на території України.

ВИСНОВКИ

В першому розділі кваліфікаційної роботи освітнього рівня «Магістр»:

- Описано пандемію COVID-19.

- Наведено опис процесу пошуку наукових публікацій щодо використання алгоритмів машинного навчання в умовах пандемії.

- Проаналізовано стан досліджень щодо використання алгоритмів машинного навчання в умовах пандемії.

- Розглянуто машинне навчання та глибоке навчання для COVID-19.

В другому розділі кваліфікаційної роботи:

- Описано формування та попередня обробка пандемічних наборів даних.

- Розглянуто розподіл даних для унікальної сегрегації пацієнтів.

- Подано опис конвеєра машинного навчання для дослідження використання алгоритмів в умовах пандемії.

- Проаналізовано показники оцінки використання алгоритмів машинного навчання в умовах пандемії.

- Висвітлено важливість функції XGBoost для оцінки біомаркерів.

- Проведено вибір NN-функції для дослідження алгоритмів машинного навчання в умовах пандемії.

- Виконано навчання NN в процесі дослідження алгоритмів машинного навчання в умовах пандемії.

- Подано опис процесу тестування NN в процесі дослідження алгоритмів машинного навчання в умовах пандемії.

В третьому розділі кваліфікаційної роботи:

- Описано результати ідентифікації ключових біомаркерів.

- Виконано порівняння прогностичних характеристик алгоритмів машинного навчання для потреб систем підтримки прийняття рішень в

умовах пандемії. Нейронна мережа прогнозує перебіг інфекційних захворювань в умовах пандемії з найбільш прийнятною точністю.

– Досліджено ефективність алгоритмів машинного навчання в умовах пандемії.

У розділі «Охорона праці та безпека в надзвичайних ситуаціях» розглянуто організацію та перевірку стану охорони праці на підприємстві під час пандемії. Описано основні принципи і способи забезпечення життєдіяльності.

ПЕРЕЛІК ДЖЕРЕЛ

- 1 Alyasseri, Zaid Abdi Alkareem, et al. "Review on COVID-19 diagnosis models based on machine learning and deep learning approaches." *Expert systems* 39.3 (2022): e12759.
- 2 Liu Y, Mao B, Liang S, Yang JW, Lu HW, Chai YH, et al. Association between age and clinical characteristics and outcomes of COVID-19. *Eur Respir J.* (2020) 55:2001112. doi: 10.1183/13993003.01112-2020.
- 3 Wang K, Zuo P, Liu Y, Zhang M, Zhao X, Xie S, et al. Clinical and laboratory predictors of in-hospital mortality in patients with COVID-19: a cohort study in Wuhan, China. *Clin Infect Dis.* (2020) 71:2079–88. doi: 10.1093/cid/ciaa538.
- 4 Arpaci I, Huang S, Al-Emran M, Al-Kabi MN, Peng M. Predicting the COVID-19 infection with fourteen clinical features using machine learning classification algorithms. *Multim Tools Appl.* (2021) 80:11943–57. doi: 10.1007/s11042-020-10340-7.
- 5 Wan Y, Zhou H, Zhang X. An interpretation architecture for deep learning models with the application of COVID-19 diagnosis. *Entropy.* (2021) 23:204. doi: 10.3390/e23020204.
- 6 Imran A, Posokhova I, Qureshi HN, Masood U, Riaz MS, Ali K, et al. AI4COVID-19: AI enabled preliminary diagnosis for COVID-19 from cough samples via an app. *Inform Med Unlock.* (2020) 20:100378. doi: 10.1016/j.imu.2020.100378.
- 7 Silverstein J. *Most of the World Doesn't Have Access to X-Rays.* The Atlantic. (2018). Available online at: <https://www.theatlantic.com/health/archive/2016/09/radiology-gap/501803/>.
- 8 Lu J, Hu S, Fan R, Liu Z, Yin X, Wang Q, et al. ACP risk grade: a simple mortality index for patients with confirmed or suspected severe acute respiratory

syndrome coronavirus 2 disease (COVID-19) during the early stage of outbreak in Wuhan, China. *MedRxiv [Preprint]*. (2020). doi: 10.1101/2020.02.20.20025510.

9 Gemmar P. An interpretable mortality prediction model for COVID19 patients-alternative approach. *MedRxiv [Preprint]*. (2020). doi: 10.1101/2020.06.14.20130732.

10 Chowdhury MEH, Rahman T, Khandakar A, Al-Madeed S, Zughaier SM, Doi SM, et al. An early warning tool for predicting mortality risk of covid-19 patients using machine learning. *arXiv preprint arXiv:2007.15559*. (2020).

11 Shang Y, Liu T, Wei Y, Li J, Shao L, Liu M, et al. Scoring systems for predicting mortality for severe patients with COVID-19. *EClinicalMedicine*. (2020) 24:100426. doi: 10.1016/j.eclinm.2020.100426.

12 Xie J, Hungerford D, Chen H, Abrams ST, Li S, Wang G, et al. Development and external validation of a prognostic multivariable model on admission for hospitalized patients with COVID-19. *medRxiv [Preprint]*. (2020). doi: 10.1101/2020.03.28.20045997.

13 Jimenez-Solem E, Petersen TS, Hansen C, Hansen C, Lioma C, Igel C, et al. Developing and validating COVID-19 adverse outcome risk prediction models from a bi-national european cohort of 5594 patients. *Sci Rep*. (2021) 11:3246. doi: 10.1038/s41598-021-81844-x.

14 Bolourani S, Brenner M, Wang P, McGinn T, Hirsch JS, Barnaby D, et al. A machine learning prediction model of respiratory failure within 48 hours of patient admission for COVID-19: model development and validation. *J Med Intern Res*. (2021) 23:e24246. doi: 10.2196/24246.

15 Yan L, Zhang HT, Yuan Y. An interpretable mortality prediction model for COVID-19 patients. *Nat Mach Intell*. (2020) 2:283–8. doi: 10.1038/s42256-020-0180-7.

16 Hansen LK, Salamon P. Neural network ensembles. *IEEE Trans Pattern Anal Mach Intell*. (1990) 12:993–1001.

17Karthikeyan, Akshaya, et al. "Machine learning based clinical decision support system for early COVID-19 mortality prediction." *Frontiers in public health* 9 (2021).

18Chen T, Guestrin C. Xgboost: a scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016). p. 785–94.

19Elith J, Leathwick JR, Hastie T. A working guide to boosted regression trees. *J Anim Ecol.* (2008) 77:802–13. doi: 10.1111/j.1365-2656.2008.01390.x.

20Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review. *J Biomed Inform.* (2002) 35:352–9. doi: 10.1016/S1532-0464(03)00034-0.

21Hosmer DW Jr, Lemeshow S, Sturdivant RX. *Applied Logistic Regression*. Vol. 398 (Hoboken, NJ: John Wiley & Sons 2013).

22Rehman A, Khan MA, Mehmood Z, Saba T, Sardaraz M, Rashid M, et al. Microscopic melanoma detection and classification: a framework of pixel-based fusion and multilevel features reduction. *Microsc Res Techn.* (2020) 83:410–23. doi: 10.1002/jemt.23429.

23Breiman L. Random forests. *Mach Learn.* (2001) 45:5–32. doi: 10.1023/A:1010933404324.

24Khan SA, Nazir M, Khan MA, Saba T, Javed K, Rehman A, et al. Lung nodule detection framework from computed tomography images using support vector machine. *Microsc Res Techn.* (2019) 82:1256–66. doi: 10.1002/jemt.23275.

25Saba T. Computer vision for microscopic skin cancer diagnosis using handcrafted and non-handcrafted features. *Microsc Res Techn.* (2021). doi: 10.1002/jemt.23686.

26Чеберячко С., Дерюгін О., Охорона праці на підприємствах в умовах пандемії COVID-19. Охорона праці і пожежна безпека. №9. 2020р.

27Опара Н.М., Сіромаха М.В. Охорона праці на підприємствах в умовах пандемії COVID-19. Інноваційні аспекти системи безпеки праці, захисту інтелектуальної власності: матеріали VI Всеукраїнської науково-практичної інтернет-конференції. (м. Полтава, 25-26 березня 2021 р.). Полтава: ПДАА, 2021. С. 95-98.

28Інструкція з охорони праці потребує змін у зв'язку з пандемією, – наголошує Держпраці. URL: <https://www.sop.com.ua/news/2580-nstruktsya-z-ohoroni-prats-potrebu-zmn-u-zvyazku-z-pandemyu-nagoloshu-derjsprats>.

29Проблеми життєдіяльності. URL: https://pidru4niki.com/11800912/bzhd/problemi_zhittyediyalnosti.

30Основні принципи забезпечення безпечної життєдіяльності. URL: https://studopedia.com.ua/1_58559_osnovni-printsipi-zabezpechennya-bezpechnoi-zhittiediyalnosti.html.

31Горденко С.І. Основні принципи та методи забезпечення безпеки життєдіяльності людей. «Молодий вчений» • № 9.1 (36.1), вересень, 2016 р. С.54-58.

ДОДАТКИ

Тези конференції

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ТЕРНОПІЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ ІВАНА ПУЛЮЯ**

МАТЕРІАЛИ

ІХ НАУКОВО-ТЕХНІЧНОЇ КОНФЕРЕНЦІЇ

**«ІНФОРМАЦІЙНІ МОДЕЛІ,
СИСТЕМИ ТА ТЕХНОЛОГІЇ»**



8–9 грудня 2021 року

**ТЕРНОПІЛЬ
2021**

УДК 001
М34

ПРОГРАМНИЙ КОМІТЕТ

Голова: Лупенко Сергій Анатолійович – докт. техн. наук, професор.

Співголови: Марущак Павло Орестовиц – проректор з наукової роботи, докт. техн. наук, професор.

Баран Ігор Олегович – канд. техн. наук, доцент, декан факультету ФІС.

Науковий секретар: Семенишин Галина Мирославівна – старший викладач.

Члени: докт. фіз.-мат. наук, професор В. Кривень; професор кафедри комп'ютерних наук, докт. тех. наук Я. Литвиненко; канд. техн. наук, доцент, Г. Осухівська; докт. техн. наук, професор М. Карпінський; канд. пед. наук, доцент Ж. Баб'як; докт. фіз.-мат. наук, професор М. Петрик; канд. техн. наук, доцент Н. Загородна.

ОРГАНІЗАЦІЙНИЙ КОМІТЕТ

Голова: Скоренький Юрій Любомирович – канд. техн. наук, доцент.

Члени: доцент кафедри комп'ютерних наук, к.т.н. В. Никитюк; доцент кафедри програмної інженерії, к.т.н. Д. Михалик; асистент М. Стадник; асистент Н. Шаблій; ст. викладач Л. Джиджора.

Матеріали ІХ науково-технічної конфіції «Інформаційні моделі, системи та М34 технології» Тернопільського національного технічного університету імені Івана Пулюя, (Тернопіль, 8 – 9 грудня 2021 р.). – Тернопіль: Тернопільський національний технічний університет імені Івана Пулюя, 2021. – 210 с.

Адреса оргкомітету: ТНТУ ім. І. Пулюя, м. Тернопіль, вул. Руська, 56, 46001, тел. (0352) 52-41-33, факс (0352) 254983.

E-mail: conffis2021@gmail.com

Редагування, оформлення, верстка: Семенишин Г.М.

СЕКЦІЇ КОНФЕРЕНЦІЇ, ЯКІ ПРЕДСТВЛЕНІ В ЗБІРНИКУ

- Математичне моделювання;
- Інформаційні системи та технології;
- Комп'ютерні системи та мережі;
- Програмна інженерія та моделювання складних розподілених систем;
- Новітні фізико-технічні та освітні технології.

В збірнику надруковано тези доповідей ІХ науково-технічної конференції «Інформаційні моделі, системи та технології» (Тернопіль, 8–9 грудня 2021 р.) за такими науковими напрямками: математичне моделювання; інформаційні системи та технології; комп'ютерні системи та мережі; програмна інженерія та моделювання складних розподілених систем; новітні фізико-технічні та освітні технології.

Розрахований на науковців, викладачів та студентів вузів.

За зміст тез та дотримання норм академічної доброчесності відповідальність несе автор.

© Тернопільський національний технічний університет імені Івана Пулюя, 2021

УДК 004.4

Д.В. Палкова, В.В. Дацик, В.В. Фіголь, Т.О. Маєвський(Тернопільський національний технічний університет імені Івана Пулюя, Україна)
(Технічний коледж ТНТУ імені Івана Пулюя, Україна)**ІНФОРМАЦІЙНІ ТЕХНОЛОГІЇ В УМОВАХ ПАНДЕМІЇ**

UDC 004.4

D.V. Palkova, V.V. Datsyk, V.V. Fihol, T.O. Maievskyi**INFORMATION TECHNOLOGIES IN A PANDEMIC CONDITION*****Ключові слова:* ІНФОРМАЦІЙНА СИСТЕМА, ТЕХНОЛОГІЯ, ПАНДЕМІЯ.*****Ke ywords:* INFORMATION SYSTEM, TECHNOLOGY, PANDEMIC.**

Пандемія COVID-19 спричинила суттєвий вплив на лікарняні медичні системи, підприємництво, освіту та економіку. Телемедицина, дистанційна робота та онлайн-освіта набувають важливості для допомоги людському суспільству з метою уповільнення поширення коронавірусу [1]. Пандемія спричинила зростання попиту та, як наслідок, зусиль в галузі інноваційних інформаційних технологій для боротьби з спричиненими COVID-19 збитками [2].

Пандемія розширила можливості впровадження інформаційно-технологічних рішень та надала можливості проведення наукових досліджень і запровадження прикладних застосунків для управління інформацією, практичної роботи та проєктування та використання інноваційних технологій [3]. Пришвидшений перехід до телемедицини, дистанційної роботи та онлайн-освіти в результаті загрози коронавірусу свідчить про те, що цифрові технології надають багато переваг. Вони відіграють важливу роль в процесах управління та зменшення ризиків, спричинених карантинними заходами під час пандемії та після неї [4]. Інформаційні системи та технології також відіграють визначальну роль в галузі охорони здоров'я, супроводі процесів підтримки прийняття клінічних рішень, оперативного реагування на кризові і надзвичайні ситуації та управлінні ризиками [5]. Обширне коло фахівців у сфері інформаційних систем та технологій проводять дослідження та роботи для боротьби з пандемією. З цією метою розробляються програмно-алгоритмічні комплекси та застосунки для боротьби з поширенням вірусу, відстеження та прогнозування його поширення, захисту медичних закладів та установ від несанкціонованих спроб доступу та кібератак [6]. Дослідники в галузі інформаційних систем та технологій вносять вагомий вклад у глобальні зусилля, спрямовані на боротьбу з COVID-19 та майбутніми пандеміями [7]. Вони використовують здобутий досвід та знання щодо реагування на кризи, процесів прийняття рішень, дистанційної роботи, управління віртуальними мережами, командної роботи, аналізу великих за обсягом наборів даних тощо. На даний час присутні запити щодо досліджень у сфері інформаційних систем для боротьби з COVID-19.

Пандемія спричиняє вагомий вплив на процеси проєктування, розроблення, практичного впровадження та використання інформаційних систем та технологій [8]. Дослідники та розробники в царині інформаційних систем та технологій допоможуть провести аналіз даних щодо пандемії COVID-19. Вони зможуть долучитись до потенційно нових напрямків досліджень. Зокрема, полегшити роботи щодо соціального дистанціювання, онлайн-комерції, розпізнавання обличчя в масках та інших варіантах використання COVID-19 застосунків. Напрямки проведення досліджень – конфіденційність, краудсорсинг, передача даних, відстеження процесів та подій, робототехніка, ланцюги постачання, автономні системи, інструменти віртуального спілкування, дистанційна освіта, розділення робіт та задач, приватність [9].

В роботі [10] Рей окреслив множину напрямків досліджень в галузі інформаційної безпеки для підвищення стійкості до пандемій та екстремальних подій, що включає переосмислення системної політики в галузі охорони здоров'я. Автор роботи [11] опублікував оцінювання спричинених COVID-19 ключових викликів до інформаційних систем та технологічних перспектив. Він запропонував рекомендації щодо управління інформацією у процесах цифрової трансформації охорони здоров'я, освіти, праці та життя.

Література.

1. Chavez, N. & Kounang, N. (2020). A man diagnosed with Wuhan coronavirus near Seattle is being treated largely by a robot. Available at <https://www.cnn.com/2020/01/23/health/ushttps://www.cnn.com/2020/01/23/health/us-wuhan-coronavirus-doctor-interview/index.html>
2. O'Leary, D. E. (2020). Evolving Information Systems and Technology Research Issues for COVID-19 and Other Pandemics. *Journal of Organizational Computing and Electronic Commerce*. Available at <https://www.tandfonline.com/doi/full/10.1080/10919392.2020.1755790>.
3. Sein, M. K. (2020). The Serendipitous Impact of COVID-19 Pandemic: A Rare Opportunity for Research and Practice. *International Journal of Information Management*, Volume 55, 102164.
4. Richter, A. (2020). Locked-down digital work. *International Journal of Information Management*, Volume 55, 102157.
5. Chen, R., Sharman, R., Chakravarti, N., Rao, H. R., & Upadhyaya, S. J. (2008). Emergency response information system interoperability: Development of chemical incident response data model. *Journal of the Association for Information Systems*, 9(3), pp. 200-230.
6. Mingis, K. (2020). Tech pitches in to fight COVID-19 pandemic. Available at <https://www.computerworld.com/article/3534478/tech-pitches-in-to-fight-covid-19https://www.computerworld.com/article/3534478/tech-pitches-in-to-fight-covid-19-pandemic.html>
7. Agerfalk, P., Conboy, K., & Myers, M. (2020). The European Journal of Information Systems Call for Papers: Special Communications on Information Systems in the Age of Pandemics. Available at <https://techjournals.wixsite.com/techjournals/ejis-is-pandemics>.
8. Sein, M. K. (2020). The Serendipitous Impact of COVID-19 Pandemic: A Rare Opportunity for Research and Practice. *International Journal of Information Management*, Volume 55, 102164.
9. O'Leary, D. E. (2020). Evolving Information Systems and Technology Research Issues for COVID-19 and Other Pandemics. *Journal of Organizational Computing and Electronic Commerce*. Available at <https://www.tandfonline.com/doi/full/10.1080/10919392.2020.1755790>.
10. Rai, A. (2020). The COVID-19 Pandemic: Building Resilience with IS Research. *MIS Quarterly*, 44(2), 02.
11. Dwivedi, Y. K., Hughes, D. L., Coombs, C., Constantiou, I., Duan, Y., Edwards, J. S., Raman, R. (2020). Impact of COVID-19 pandemic on information management research and practice: Transforming education, work and life. *International Journal of Information Management*, Volume 55, 102211.

УДК 004.4

Д.В. Палкова, В.В. Дацук, В.В. Фіголь, О.В. Яскілка

(Тернопільський національний технічний університет імені Івана Пулюя, Україна)

(Технічний коледж ТНТУ імені Івана Пулюя, Україна)

ВПЛИВ ПАНДЕМІЇ НА РОЗВИТОК ІНФОРМАЦІЙНИХ ТЕХНОЛОГІЙ

UDC 004.4

D.V. Palkova, V.V. Datsyk, V.V. Fihol, O.V. Yaskilka

THE PANDEMIC IMPACT ON THE INFORMATION TECHNOLOGY DEVELOPMENT

Ключові слова: ІНФОРМАЦІЙНА ТЕХНОЛОГІЯ, ПАНДЕМІЯ, РОЗВИТОК.

Key words: INFORMATION TECHNOLOGY, PANDEMIC, DEVELOPMENT.

COVID-19 – це безпрецедентна пандемія у всьому світі. З точки зору впливу на поведінку людей її можна порівняти з Другою світовою війною, Великою депресією та іспанським грипом 1918 року. У світі продовжує зростати кількість інфікованих COVID-19 пацієнтів та спричинених ним смертей. Дослідники, лікарі та епідеміологи у всьому світі намагалися вивчити нові аспекти COVID-19. Швидкий розвиток та експоненційне зростання пандемії виявило та ускладнило обширний перелік проблем в системі охорони здоров'я та системах доставки по всьому світу [1]. Зокрема, обмеження лікарняних ресурсів, ліжок, персоналу, запасів та системах охорони здоров'я в цілому стало очевидним через зростання кількості випадків інфікування COVID-19. Медичним закладам та системам охорони здоров'я довелося швидко розширювати потужності [2].

Для боротьби з пандемією COVID-19 регулярно вводяться карантинні заходи, що включають запровадження фізичного дистанціювання. Які практично вплинули на всі сфери людської діяльності. Зокрема, торгівля, освіта, робота, соціальна взаємодія та розваги перейшли з офлайн-категорії в онлайн. Це призвело до прискореного розвитку та поширення нових інформаційних технологій серед обширного кола звичайних громадян. Водночас для них ще більше зростає цифровий розрив з громадянами без доступу до технологій. Подана в роботі [3] структура системи «дані-людина» демонструє багатопрофільну дорожню карту контролю та боротьби з інфекційними захворюваннями. Вони зосереджені на трьох компонентах:

- 1) Відбір, інтеграція та передавання даних про здоров'я громадян.
- 2) Сумісність систем.
- 3) Рекомендаційні системи та інтерфейси для управління поведінкою людей.

Для боротьби з пандемією нещодавно було розроблено ряд нових інформаційно-технологічних програм та застосунків. Зокрема мобільні програмні засоби відстеження контактів інфікованих COVID-19 громадян та цільові чат-боти. Використання зазначених технологій може допомогти зменшити вплив пандемії на окремих громадян, організації та суспільство в цілому. Ефективне та інноваційне використання інформаційних технологій зможе допомогти в процесах виявлення коронавірусу в громадах, відстеження стану інфікованих пацієнтів, покращення процедур лікування інфікованих COVID-19 громадян. Допоможе розробити медикаментозне лікування та вакцини [4].

Інформаційні технології на основі штучного інтелекту, зокрема, сформовані на основі алгоритмів машинного навчання, розпізнавання зображень та глибокого навчання, можуть бути використані для раннього виявлення та діагностування COVID-19, швидшого відкриття ліків та розробки нових методів лікування [5]. На даний час ряд розробників перепрофілювали існуючі інформаційні системи на основі штучного інтелекту для сприяння реалізації заходів соціального дистанціювання та відстеження соціальних контрактів громадян [6]. Інформаційна технологія 3D-друку допомагає у виготовленні масок для обличчя та засобів індивідуального захисту

медичних працівників. Компанія «Markforged» співпрацює з корпорацією «Neurophotometrics» з метою виробництва тампонів в носоглотку, виготовлених за допомогою 3D-друку. Вони використовуються для оперативного обстеження на COVID-19. На виготовлення окремого екземпляра потрібно менше трьох хвилин [7].

Аналіз великих за обсягом наборів даних ефективно використовується для ідентифікації осіб, що потребують карантину на основі історії подорожей, прогнозування процесів поширення COVID-19, прискорення розробки противірусних препаратів та вакцин, покращення розуміння процесів поширення COVID-19. На Тайвані було успішно застосовано аналітику великих даних для виявлення випадків інфікування COVID-19 та оповіщення громадян в режимі реального часу [8].

Високопродуктивні обчислювальні інфраструктури та суперкомп'ютери для вирішення складних наукових проблем та обробки великих за обсягом наборів даних використовуються для розробки нових ліків та вакцин у коротші терміни. Високопродуктивний обчислювальний консорціум COVID-19 був створений з метою використання обчислювальних ресурсів та суперкомп'ютерів у США. До консорціуму входять шістнадцять державних та приватних організацій [9].

Література.

1. Melvin SC, Wiggins C, Burse N, Thompson E, Monger M. The Role of Public Health in COVID-19 Emergency Response Efforts from a Rural Health Perspective. *Prev Chronic Dis* 2020;17:200256.
2. World Health Organization. WHO Releases Guidelines to Help Countries Maintain Essential Health Services During the COVID-19 Pandemic. Geneva, Switzerland: World Health Organization; 2020. <https://www.who.int/news-room/detail/30-03-2020-who-releasesguidelines-to-help-countries-maintain-essentialhealth-services-during-the-covid-19-pandemic>.
3. Bardhan, I., Chen, H., & Karahanna, E. (2020). Connecting Systems, Data, and People: A Multidisciplinary Research Roadmap for Chronic Disease Management. *Management Information Systems Quarterly*, 44(1), pp. 185-200.
4. Johnstone, S. (2020). A viral warning for change. COVID-19 Versus the Red Cross: Better Solutions Via Blockchain and Artificial Intelligence (February 3, 2020). University of Hong Kong Faculty of Law Research Paper, (2020/005).
5. Brohi, S. N., Jhanjhi, N. Z., Brohi, N. N., & Brohi, M. N. (2020). Key Applications of State-of-the-Art Technologies to Mitigate and Eliminate COVID-19. Available at https://www.techrxiv.org/articles/Key_Applications_of_State.
6. Sipior, J. C. (2020). Considerations for Development and Use of AI in Response to COVID-19. *International Journal of Information Management*, Volume 55, 102170.
7. Markforged. (2020). Fiberflex: 3D Printed Nasal Swabs For Covid-19 Testing. Available at <https://markforged.com/covid-19/#swabs>.
8. Van Bavel, J. J., Baicker, K., Boggio, P. S., Capraro, V., Cichocka, A., Cikara, M., Drury, J. (2020). Using social and behavioural science to support COVID-19 pandemic response. *Nature Human Behaviour*, pp. 1-12.
9. Woo, T. (2020). Cloud Players and Research Groups Join The Fight Against COVID-19 With High-Performance Computing. Forrester. Available at <https://go.forrester.com/blogs/cloudhttps://go.forrester.com/blogs/cloud-players-and-research-groups-join-the-fight-against-covid-19-with-high-performance-computing>.