

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

бакалавр

(назва освітнього ступеня)

на тему: Дослідження контрольованого машинного навчання для
прогнозування зараження COVID-19 на основі епідеміологічних наборів даних

Виконав: студент IV курсу, групи СТ-41

спеціальності 126 Інформаційні системи та

(шифр і назва спеціальності)

технології

(підпис)

Квашнін Д.О.

(прізвище та ініціали)

Керівник

(підпис)

Липак Г.І.

(прізвище та ініціали)

Нормоконтроль

(підпис)

Шимчук Г.В.

(прізвище та ініціали)

Завідувач кафедри

(підпис)

Боднарчук І.О.

(прізвище та ініціали)

Рецензент

(підпис)

Скоренький Ю.Л.

(прізвище та ініціали)

Тернопіль
2021

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Безпека життєдіяльності, основи хорони праці	Гурик О.Я., доцент кафедри МТ		

7. Дата видачі завдання 25 січня 2021 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1.	Ознайомлення з завданням до кваліфікаційної роботи	25.01.2021	Виконано
2.	Підбір джерел про контрольоване машинне навчання для прогнозування зараження COVID-19 та епідеміологічні набори даних.	26.01.2021-01.02.2021	Виконано
3.	Переклад та опрацювання джерел про контрольоване машинне навчання для прогнозування зараження COVID-19 та епідеміологічні набори даних.	02.02.2021-08.02.2021	Виконано
4.	Виконання дослідження щодо контрольованого машинного навчання для прогнозування зараження COVID-19 на основі епідеміологічних наборів даних.	09.02.2021-13.02.2021	Виконано
5.	Оформлення розділу «Предметна область, літературні джерела та набори даних».	14.02.2021-19.02.2021	Виконано
6.	Оформлення розділу «Дослідження контрольованого машинного навчання для прогнозування зараження COVID-19 на основі епідеміологічних наборів даних»	20.02.2021-24.02.2021	Виконано
7.	Виконання завдання до підрозділу «Безпека життєдіяльності»	07.06.2021-08.06.2021	Виконано
8.	Виконання завдання до підрозділу «Основи хорони праці»	07.06.2021-08.06.2021	Виконано
9.	Оформлення кваліфікаційної роботи	07.06.2021-08.06.2021	Виконано
10.	Нормоконтроль	07.06.2021-08.06.2021	Виконано
11.	Перевірка на плагіат	08.06.2021	Виконано
12.	Попередній захист кваліфікаційної роботи	09.06.2021	Виконано
13.	Захист кваліфікаційної роботи	25.06.2021	

Студент

_____ (підпис)

Квашнін Д.О.

_____ (прізвище та ініціали)

Керівник роботи

_____ (підпис)

Липак Г.І.

_____ (прізвище та ініціали)

АНОТАЦІЯ

Дослідження контрольованого машинного навчання для прогнозування зараження COVID-19 на основі епідеміологічних наборів даних // Кваліфікаційна робота освітнього рівня «Бакалавр» // Квашнін Дмитро Олександрович // Тернопільський національний технічний університет імені Івана Пулюя, факультет комп'ютерно-інформаційних систем і програмної інженерії, кафедра комп'ютерних наук, група СТ-41 // Тернопіль, 2021 // С. 44, рис. – 11, табл. – 5, кресл. – 12, додат. – 0, бібліогр. – 44.

Ключові слова: COVID-19, аналітичне опрацювання, дані, дерево рішень, набір, машинне навчання, пандемія.

Кваліфікаційна робота присвячена дослідженню методів контрольованого машинного навчання котрі використовуються для прогнозування зараження COVID-19 на основі епідеміологічних наборів даних. Метою даної кваліфікаційної роботи освітнього рівня «Бакалавр» є підвищення рівня поінформованості громадян щодо інфікування COVID-19 шляхом аналітичного опрацювання епідеміологічних наборів даних з використанням методів контрольованого машинного навчання.

В першому розділі кваліфікаційної роботи освітнього рівня «Бакалавр» проаналізовано предметну область засобів та методів машинного навчання в контексті їх використання для опрацювання відомостей щодо COVID-19. Виконано аналіз літературних джерел.

В другому розділі кваліфікаційної роботи проаналізовано матеріали та методи дослідження. Описано методи контрольованого машинного навчання. Подано аналіз коефіцієнта кореляції. Розглянуто прогнозні моделі щодо зараження COVID-19. Виконано аналіз дерева рішень. Та подано оцінювання результатів.

ANNOTATION

Study of controlled machine learning for COVID-19 infecting forecast based on epidemiological datasets // Qualification work of educational level "Bachelor" // Kvashnin Dmytro Oleksandrovych // Ternopil National Technical University named after Ivan Pulyuy, Faculty of Computer Information Systems and Software Engineering, Department of Computer Science, group ST-41 // Ternopil, 2021 // P. 44, fig. – 11, table. – 5, chair. – 12, added. – 0, bibliogr. – 44.

Key words: COVID-19, analytical processing, data, decision tree, set, machine learning, pandemic.

The qualification work is devoted to the study of controlled machine learning methods used to predict COVID-19 infection based on epidemiological data sets. The purpose of this qualification work of the educational level "Bachelor" is to increase the level of awareness of citizens about the infection of COVID-19 by analytical processing of epidemiological data sets using the methods of controlled machine learning.

The first section of the qualification work of the educational level "Bachelor" analyzes the subject area of tools and methods of machine learning in the context of their use for processing information on COVID-19. The analysis of literary sources is executed.

In the second section of the qualification work the materials and research methods are analyzed. Methods of controlled machine learning are described. The analysis of the correlation coefficient is given. Predictive models for COVID-19 infection are considered. The analysis of the decision tree is performed. And an evaluation of the results is given.

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

ANN (англ. Artificial Neural Network) – Штучна нейронна мережа.

CNN (англ. Convolutional Neural Network) – Згорткова нейронна мережа.

DL (англ. Deep Learning) – Глибинне навчання.

ELM (англ. Extreme Learning Machine) – Екстремальне машинне навчання.

GAN (англ. Generative Adversarial Network) – Генеративна мережа на основі змагання.

MIT (англ. Massachusetts Institute of Technology) – Массачусетський технологічний інститут.

ML (англ. Machine Learning) – Машинне навчання.

RMSE (англ. Root Mean Square Error) – Середньоквадратична помилка.

SL (англ. Self-Controlled Learning) – Самоконтрольоване навчання.

ШІ – Штучний інтелект.

ЗМІСТ

ВСТУП	7
1 ПРЕДМЕТНА ОБЛАСТЬ, ЛІТЕРАТУРНІ ДЖЕРЕЛА ТА НАБОРИ ДАНИХ	9
1.1 Аналіз предметної області.....	9
1.2 Аналіз літературних джерел.....	13
1.3 Підготовка та попередній аналіз структури наборів даних	16
1.4 Висновок до першого розділу	20
2 ДОСЛІДЖЕННЯ КОНТРОЛЬОВАНОГО МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ЗАРАЖЕННЯ COVID-19 НА ОСНОВІ ЕПІДЕМІОЛОГІЧНИХ НАБОРІВ ДАНИХ.....	21
2.1 Матеріали і методи.....	21
2.2 Методи контрольованого машинного навчання	22
2.3 Аналіз коефіцієнта кореляції.....	25
2.4 Прогнозні моделі щодо зараження COVID-19	28
2.5 Аналіз дерева рішень	30
2.6 Оцінювання результатів.....	31
2.7 Висновок до другого розділу	34
3 БЕЗПЕКА ЖИТТЄДІЯЛЬНОСТІ, ОСНОВИ ХОРОНИ ПРАЦІ	35
3.1 Долікарська допомога при кровотечах	35
3.2 Допомога при теплових і сонячних ударах	36
3.3 Висновок до третього розділу	38
ВИСНОВКИ.....	39
ПЕРЕЛІК ДЖЕРЕЛ	40

ВСТУП

Актуальність теми. COVID-19 або 2019-nCoV вже не є пандемією, а радше ендеміком, оскільки понад шістьсот тисяч людей у всьому світі втратили життя після зараження хворобою. На даний час не існує спеціального лікування проти COVID-19, тому життя з хворобою та її симптомами неминуче. Ця реальність спричинила значне навантаження на національні системи охорони здоров'я у всьому світі, особливо в країнах, що розвиваються. Для знищення пандемії COVID-19 не існує ані ефективної, клінічно перевіреної стратегії протівірусних препаратів, ані затвердженої вакцини. При цьому існують альтернативи, які можуть зменшити навантаження не лише на обмежені щодо ресурсів системи охорони здоров'я, але й на економічний сектор; найперспективніші включають використання неклінічних методів, таких як машинне навчання, видобування даних, глибинне навчання та інші елементи на основі ШІ. Ці альтернативи полегшать діагностику та прогноз для пацієнтів з пандемією COVID-19. Тому розроблення контрольованих моделей машинного навчання для аналітичного опрацювання відомостей щодо COVID-19 є актуальним напрямком сучасних досліджень.

Мета і задачі дослідження. Метою даної кваліфікаційної роботи освітнього рівня «Бакалавр» є підвищення рівня поінформованості громадян щодо інфікування COVID-19 шляхом аналітичного опрацювання епідеміологічних наборів даних з використанням методів контрольованого машинного навчання. Для досягнення поставленої мети потребують вирішення ряд наступних завдань:

- Проаналізувати предметну область досліджень щодо методів аналітичного опрацювання епідеміологічних наборів даних по COVID-19.
- Проаналізувати процеси формування моделей класифікації на основі методів машинного навчання для виявлення випадків зараження COVID-19.

– Провести порівняльне оцінювання методів контрольованого машинного навчання при аналізі епідеміологічних наборів даних по COVID-19.

Практичне значення одержаних результатів. Сформовані результати оцінки методів ML можуть бути практично використані при виборі оптимального методу для аналітичного опрацювання епидемиологічних відомостей щодо COVID-19.

1 ПРЕДМЕТНА ОБЛАСТЬ, ЛІТЕРАТУРНІ ДЖЕРЕЛА ТА НАБОРИ ДАНИХ

1.1 Аналіз предметної області

Останній патогенний спалах гострого респіраторного синдрому коронавірусу (SARS-CoV-2) відповідає за глобальну пандемію 2019-nCoV або COVID-19 [1]. Вірус виник у штаті Ухань, Китай, наприкінці грудня 2019 року. Вірус вперше був виявлений у кажанів і потрапив до людини через посередницьких господарів [2]. Коронавіруси (CoV), які включають Близькосхідний респіраторний коронавірусний синдром (MERS-CoV) Тяжкий гострий респіраторний коронавірусний синдром (SARS-CoV) та новий коронавірус 2019 року (2019-nCoV), викликають різні захворювання ссавців та птахів [3]. Під час першого виявлення вірусу китайський Протокол діагностики та лікування нової коронавірусної пневмонії зазначав, що COVID-19 можна виявити без позитивних результатів тестів на SARS-CoV-2 за допомогою трьох методів:

1. Позитивна КТ грудної клітки.
2. Значні клінічні прояви включають кашель, задишку та інші ознаки інфекції нижніх дихальних шляхів.
3. Лабораторні результати, що показують лімфопенію та за потреби лейкопенію [4].

Значно поширюються симптоми лихоманки COVID-19 (98%), кашлю (76%) та діареї (3%), які часто важчі у дорослих людей із хронічними захворюваннями [5]. Багато пацієнтів відзначають задишку, яка у багато випадків, як видається, є симптомом грипу [6]. COVID-19 поширюється експоненціально по всьому світу, оскільки він був вперше визнаний наприкінці 2019 року [7]. Пандемія вплинула на всі міжнародні перевезення та понад 209 країн та територій у всьому світі [8]. Будучи оголошеною

надзвичайною ситуацією в галузі охорони громадського здоров'я та міжнародною проблемою, пандемія передається шляхом безпосереднього контакту з тілесною рідиною інфікованої особи, чханням та кашлем [9]. На даний час від COVID19 не має ні ліків ні затверджених вакцин Хоча зараз значні зусилля докладаються вченими у всьому світі для їх пошуку [2]. Антивірусні засоби, зокрема арбідол, рибавірин, «hINFa-2b», хлорохін, фавіпіравір, ремдезивір, лопінавір, ритонавір тощо, на даний час використовуються в клінічних випробуваннях щодо COVID-19 [10] та лікарських рослин, які, як вважають, мають захисний ефект на захворювання дихальних шляхів. використовуються для управління пандемією COVID-19 [11].

Безсимптомні випадки та відсутність наборів діагнозів призводять до затримки або навіть пропуску діагностики, піддаючи пацієнтів, відвідувачів та медичних працівників впливу інфекції COVID-19. Це становить велику загрозу для секторів охорони здоров'я та економіки. Тому неклінічні методи, зокрема машинне навчання, видобування даних, експертні системи та інші методи штучного інтелекту, повинні відігравати важливу роль у діагностиці та стримуванні пандемії COVID-19. Використання нетерапевтичних підходів може зменшити навантаження на системи охорони здоров'я, одночасно забезпечуючи найкращі діагностичні та передбачувані методи для COVID-19.

Машинне навчання (ML) є однією з найдосконаліших концепцій штучного інтелекту (ШІ) і забезпечує стратегічний підхід до розробки автоматизованих, складних та об'єктивних алгоритмічних прийомів для мультимодального та розмірного біомедичного або математичного аналізу даних [12]. Алгоритми ML здатні читати та модифікувати свою структуру на основі набору спостережуваних даних з адаптацією, здійсненою шляхом оптимізації функції витрат або мети [13]. ML вже продемонструвало свій потенціал для діагностики, виявлення, терапевтичного спостереження та

лікування багатьох захворювань [14]. Методи ML можна класифікувати за чотирма способами:

1. Методи навчання під контролем – це методи навчання або алгоритми навчання ML, які пов'язують попередній та поточний набір даних за допомогою маркованих даних для прогнозування майбутніх подій [15]. Процес навчання починається з навчального процесу з набору даних та розвиває цілеспрямовану діяльність для прогнозування вихідних значень [16]. Методи здатні забезпечити пошук результатів у наборах вхідних даних за допомогою адекватного навчального процесу та порівнювати результати з фактичними результатами та очікуваннями для виявлення помилок та модифікації моделі відповідно до результатів [17].

2. Неконтрольоване навчання – це техніки ML, які використовуються, коли набір навчальних даних не класифікується або не маркується [8]. Методи навчання виводять функцію для видобування прихованих знань або шаблонів з немаркованого набору даних [18]. Методика не визначає належний результат, а скоріше видобуває спостереження та закономірності з набору даних [19].

3. Навчальні методи напівконтролю – це методи навчання, які лежать між контрольованими та неконтрольованими методами навчання, де в процесі навчання використовуються марковані та немарковані набори даних [18]. Як правило, методи навчання розглядають менший набір даних із маркуванням та більший набір даних без міток [20]. Методи навчання можуть бути адаптовані для досягнення вищої точності. Зазначені методи є кращими, коли маркований набір даних потребує компетентних та відповідних ресурсів для навчання в ньому [21].

4. Методи навчання з підкріпленням взаємодіють із навчальним середовищем шляхом виявлення помилок [18]. Затримка винагород та пошук методом спроб та помилок – це загальні риси прийомів навчання з підкріпленням. Ці прийоми використовуються для виявлення ідеальної

поведінки в конкретному контексті з метою підвищення ефективності моделі [22].

Процес машинного навчання починається зі збору даних з різних ресурсів [23]. Наступним кроком є виправлення попередньо опрацьованих даних, щоб виправити проблеми, пов'язані з їх структурою та вмістом, і зменшити розмір простору, видаливши недійсні дані з файлів для відбору цікавих даних [24]. Іноді значення набору даних може бути важливим для прийняття системою рішень, але алгоритми машинного навчання розробляються з використанням інших концепцій. Зокрема, для аналізу даних, видобування корисних та нових знань, прихованих моделей або інформації з минулого досвіду використовуються підходи з статистики, теорії контролю та теорії ймовірності тощо [23]. Наступним кроком є оцінка продуктивності моделей та оптимізація моделі. Це вдосконалює модель за допомогою нових наборів даних та правил [7]. Методи ML використовуються в різних сферах, зокрема медицині, інженерії, освіті, виробництві, прогнозуванні, управлінні дорожнім рухом та роботизації [23]. На рисунку 1.1 подано процес навчання для розробки прогнозних моделей.



Рисунок 1.1 – Процес навчання для розробки прогнозних моделей

Останнім часом методи ML використовуються для аналізу великих за обсягами структурованих та неструктурованих наборів даних. Діагностика малярії, класифікація тифу та судинних захворювань, оцінка ризику діабету,

аналіз геномних та генетичних даних є прикладами біомедичного використання методів ML [23].

У цій роботі контрольовані методи ML використовуються для розробки прогнозних моделей зараження COVID-19, на основі маркування епідеміологічних даних для виявлення позитивних та негативних випадків COVID-19 з контрольованими алгоритмами навчання, зокрема дерева рішень, логістичні регресії та наївний Байєса, підтримку векторних машин та штучні нейронні мережі.

1.2 Аналіз літературних джерел

На даний час проведено багато досліджень із використанням різних методів та засобів на основі ШІ для діагностики, прогнозування зараження та реабілітації COVID-19. У роботі [8] подана прогнозна модель інтелектуального аналізу даних для реабілітації пацієнтів з COVID-19, яка була розроблена з використанням чотирьох алгоритмів інтелектуального аналізу даних. Серед них модель, зроблена на основі дерева рішень, має найвищу точність 99,85%. У роботі [25] розроблені згорткові нейронні мережі, які прогнозують новий коронавірус на основі рентгенівських зображеннями. Техніка глибокого навчання, яка є однією з підгалузей ML, натхненна структурою людського мозку, використовується для автоматичного прогнозування пацієнтів з COVID-19. Було використано набір даних з рентгенівськими зображеннями грудної клітини. Були використані попередньо навчені моделі «InceptionV3», «ResNet50» та «Inception ResNetV2», котрі пройшли навчання та тестування на наборі даних. Результат дослідження показав, що попередньо навчена модель «ResNet» дала найвищу точність серед трьох моделей: 98%. Це показує придатність моделі для допомоги медичним працівникам у прийнятті високоефективних рішень в клінічній практиці, що допоможе виявити COVID-19 на ранніх

стадіях зараження. У роботі [26] модифікована модель видобування відомостей щодо інфекцій «SEIR» та модель ML для прогнозування тенденції пандемії COVID-19, котрі були розроблені у Китаї. Моделі ефективно прогнозували піки та розміри пандемії. Досліджувались відомості щодо міграції населення та оновлені епідеміологічні дані COVID-19.

ML-підхід був використаний для аналітичного опрацювання даних щодо ГРВІ 2003 року для прогнозування процесів перебігу пандемії. У роботі [27] було здійснено видобування даних для прогнозування захворюваності на COVID-19 на основі даних про тенденції Google на центральному сході. Для оцінки кількості позитивних випадків COVID-19 використовувались моделі довготривалої короткострокової пам'яті та лінійної регресії. Моделі оцінювались із використанням метрики середньоквадратичної помилки RMSE та десяти методів перехресної перевірки. RMSE довготривалої короткочасної пам'яті та моделей лінійної регресії становили 27,273 та 7,367, відповідно. Дослідження передбачало тенденцію спалаху COVID-19. Подібні прогнози можуть допомогти менеджерам в галузі охорони здоров'я та розробникам політики ефективно планувати, розподіляти та використовувати ресурси в галузі охорони здоров'я. Довідник [28] ідентифікував внутрішню геномну сигнатуру COVID-19, використовуючи підхід без вирівнювання на основі ML. Цей підхід включав контрольований ML-аналіз цифрового сигналу аналізу геному, доповнений процес прийняття рішень та аналіз коефіцієнта рангової кореляції Спірмена для результатів перевірки.

Результат дослідження підтверджує гіпотезу дослідження кажана як джерело походження COVID-19. При цьому дослідження додатково класифікує пандемію як сарбековірус у складі бетакоронавірусу. Понад п'ять тисяч унікальних геномних послідовностей із набору даних, загальним обсягом у вісім мільйонів, були проаналізовані з точністю понад 90%.

У роботі [18] підхід, сформований на основі машинного навчання, був розроблений для прогнозування в реальному режимі часу спалаху COVID-19

з використанням сповіщень про новини, про які повідомляє Media Cloud, та офіційного звіту про охорону здоров'я Китайського центру з контролю та профілактики, пошукової діяльності в Інтернеті та щоденний прогноз від GLEAM, котрий сформовано на основі механістичної моделі, побудованої на агентах. Підхід використовує кластеризацію, яка дозволяє дослідити геопросторову синхронність перебігу COVID-19 у китайських провінціях. Цей підхід може дати точний прогноз на два дні раніше часу. Підхід, керований ML, також застосовувався для прогнозування тяжкості перебігу інфекції у пацієнтів. У дослідженні було запропоновано клінічний набір даних з Уханю. Модель прогнозування на основі алгоритму «ML Xgboost» була розроблена та протестована на двадцяти дев'яти пацієнтах. Модель змогла передбачити ризик смертності пацієнтів COVID-19 та клінічний шлях до розпізнавання критичних важких. Модель допомогла лікарям з ідентифікацією пацієнтів COVID-19. Завдяки своєчасним медичним втручанням, модель потенційно могла зменшити ризик смертності.

У дослідженні [29] для прогнозування кандидатів на вакцину COVID-19 використовували засоби зворотної вакцинології ML та «Vaxign». Інструмент зворотної вакцинології «Vaxign» передбачав, що S білок є ймовірною адгезією. Водночас «Vaxign ML» передбачав, що S білок має високий показник захисної антигенності. Прогнозування вакцинації у дослідженні пропонує нові стратегії для ефективної та безпечної розробки вакцини COVID-19.

Довідник [30] представив підхід, керований даними, для аналізу пандемії COVID-19 з урахуванням її ранньої динаміки зараження та часового майнера знань на основі даних з США. Видобуваються відомості щодо інфекційної сили, ступеня важкості інфекції, що є важливим для оцінки безсимптомних інфекцій та прогнозування нових інфекцій з часом. Підхід виявив дуже значну кількість випадків безсимптомних інфекцій пандемії COVID-19, з відставанням близько десяти днів. Кількісно було підтверджено,

що інфекційна сила вірусу сильна, з приблизно 0,14% переходу від легкої до серйозної інфекції.

Супутні роботи, вказують що методи ML та інші методи ШІ відігравали важливу роль у прогнозуванні, діагностиці та стримуванні пандемії COVID-19. Це може допомогти зменшити навантаження на державні системи охорони здоров'я. На основі проведеного аналізу та опрацювання наукових публікацій можемо зробити висновок, що на даний час опубліковано доволі мало відомостей щодо маркування епідеміологічних наборів даних для позитивних та негативних випадків COVID-19 та для розробки контрольованих моделей ML для прогнозування інфекції COVID-19. Тому, дане дослідження має на меті дослідити ці прогалини.

1.3 Підготовка та попередній аналіз структури наборів даних

Для дослідження потрібно використовувати епідеміологічні набори даних про позитивні та негативні випадки COVID19 розміщені на офіційних веб-сайтах. Зокрема набір даних [31] був отриманий на основі інформації Епідеміологічного нагляду за вірусними респіраторними хворобами, про яку повідомляли чотириста сімдесят п'ять підрозділів моніторингу вірусних респіраторних захворювань (USMER). Набір даних містить результати лабораторних досліджень зворотної транскрипції полімеразної ланцюгової реакції «RT-PCR» для випадків COVID-19 (див. таблицю 1.1).

Таблиця 1.1 – Відомості про структуру набору даних [31]

№	Особливість	Значення	Тип даних
1	2	3	4
1	Вік	більше за 0	int64
2	Стать	0 = жінка, 1 = мужчина	int64
3	Пневмонія	0 = негативний, 1 = позитивний	int64

Графік формування профілю набору даних для критерію [31] результат подано на рисунку 2.1.

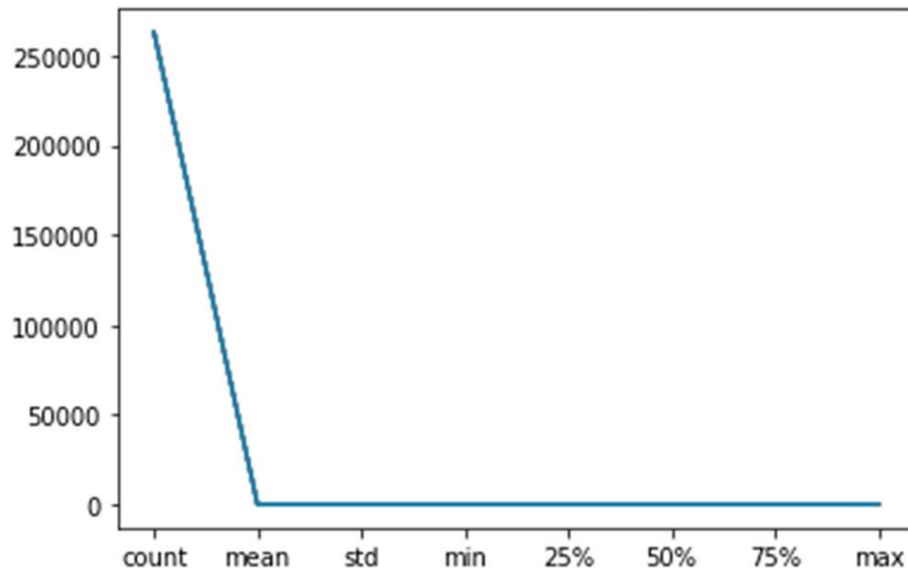


Рисунок 2.2 – Діаграма профілю набору даних

В наборах даних присутні відомості щодо віку пацієнтів, інфікованих COVID-19. Графік частоти віку пацієнтів, інфікованих COVID-19 подано на рисунку 2.3.

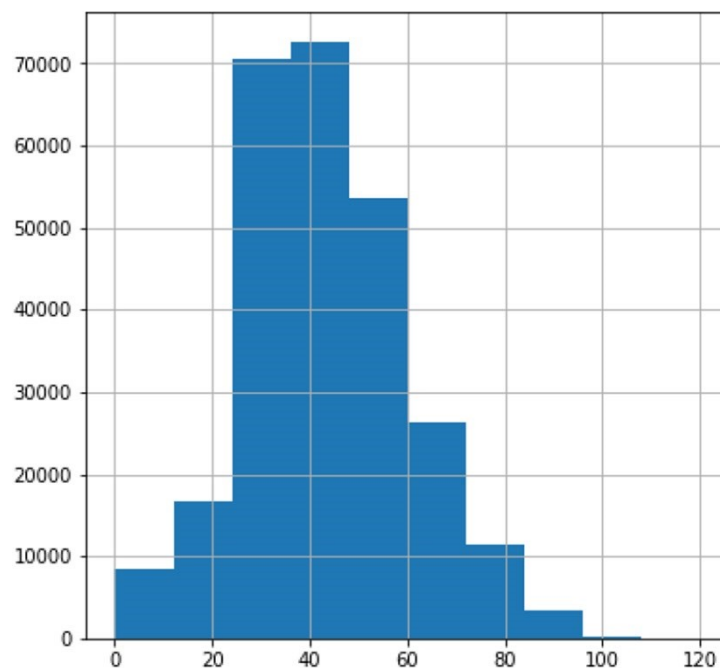


Рисунок 2.3 – Графік частоти віку пацієнтів, інфікованих COVID-19

В наборах данных присутствуют сведения о поле пациентов, инфицированных COVID-19. График гендерного распределения пациентов, инфицированных COVID-19, представлен на рисунке 2.4.

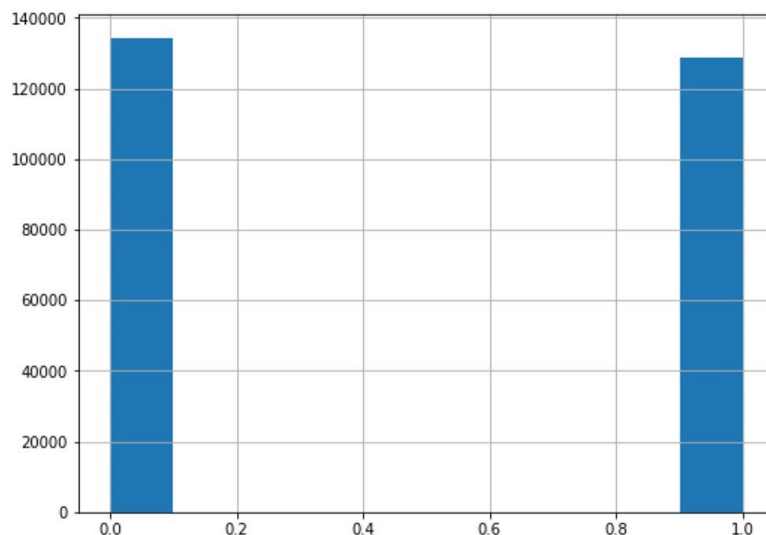


Рисунок 2.4 – График гендерного распределения пациентов, инфицированных COVID-19

В наборах данных присутствуют сведения о результатах тестирования пациентов, инфицированных COVID-19. График распределения результатов тестирования пациентов, инфицированных COVID-19, представлен на рисунке 2.5.

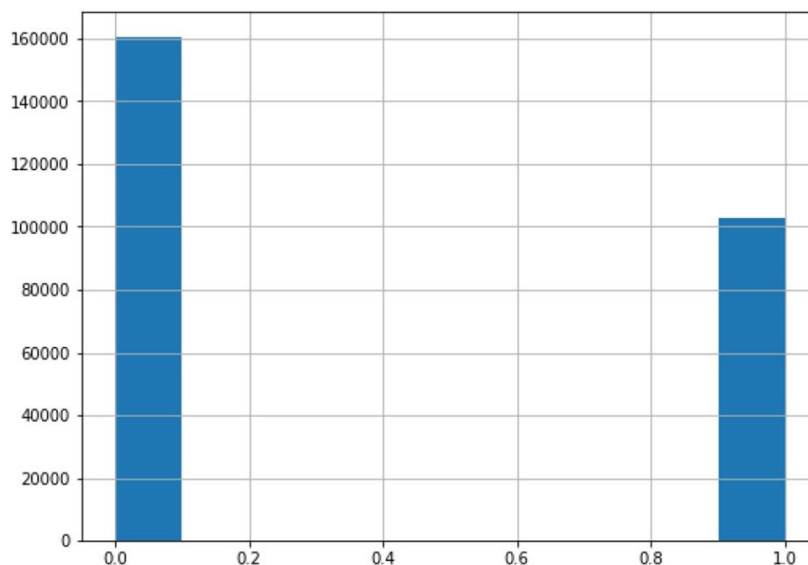


Рисунок 2.5 – Частота результатов тестирования пациентов, инфицированных COVID-19

В таблиці 1.3 подано відомості про профіль набору даних.

Таблиця 1.3 – Інформація про профіль набору даних [31]

S. No.	Властивість	Мінімум	Максимум	Середній	Відхилення
1	Вік	0	120	42.59	16.90
2	Стать	0	1	0.49	0.50
3	Пневмонія	0	99	0.17	0.81
4	Діабет	0	98	0.51	6.07
5	Астма	0	98	0.38	5.80
6	Гіпертонія	0	98	0.52	5.84
8	Ожиріння	0	98	0.53	5.92
10	Куріння	0	98	0.46	5.98
11	Результат	0	1	0.39	0.49

Епідеміологічний набір даних про позитивні та негативні випадки COVID-19 має понад сорок ознак – стовпців. При цьому до набору даних [31] включено лише дві демографічні ознаки, зокрема це вік та стать і вісім клінічних ознак, які включають пневмонію, діабет, астму, гіпертонію, ожиріння та один фактор високого ризику, який є результатом куріння. Окрема колонка містить відомості щодо «RT-PCR» COVID-19 у наборі даних пацієнтів.

1.4 Висновок до першого розділу

В першому розділі кваліфікаційної роботи освітнього рівня «Бакалавр» проаналізовано предметну область засобів та методів машинного навчання в контексті їх використання для опрацювання відомостей щодо COVID-19. Виконано аналіз літературних джерел. Описано процес підготовки та попереднього аналізу структури наборів даних.

2 ДОСЛІДЖЕННЯ КОНТРОЛЬОВАНОГО МАШИННОГО НАВЧАННЯ ДЛЯ ПРОГНОЗУВАННЯ ЗАРАЖЕННЯ COVID-19 НА ОСНОВІ ЕПІДЕМІОЛОГІЧНИХ НАБОРІВ ДАНИХ

2.1 Матеріали і методи

Методологія формування керованих моделей машинного навчання для прогнозування зараження COVID-19 за допомогою набору епідеміологічних даних подана в роботі [31], подана на рисунку 2.1.

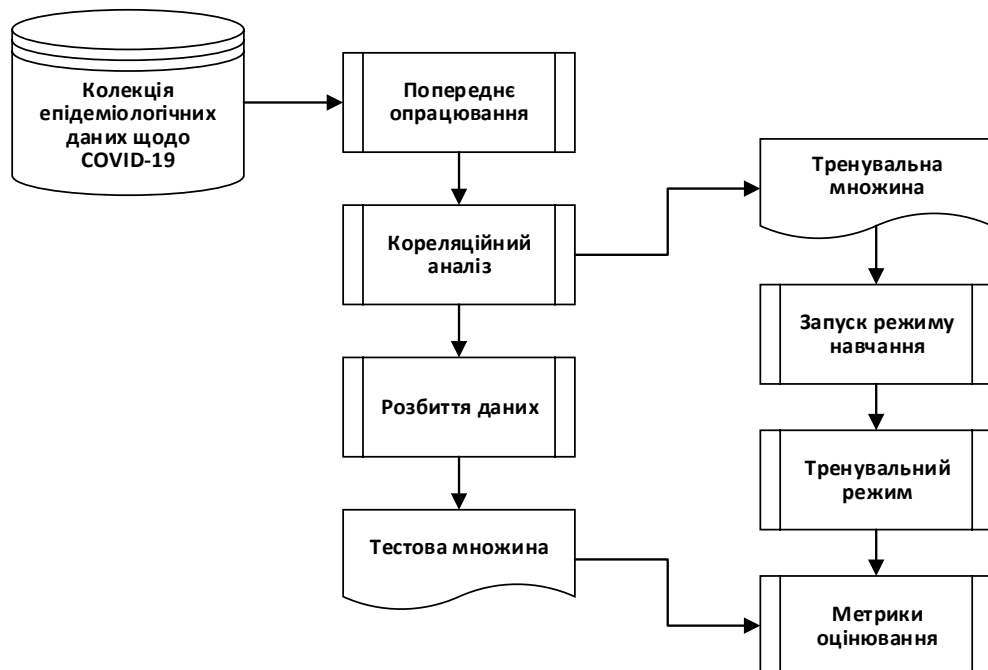


Рисунок 2.1 – Формування моделей класифікації на основі методів машинного навчання для виявлення випадків зараження COVID-19 [31]

На початку набір епідеміологічних даних проходить процедуру попереднього опрацювання. Потім відбувається кореляційний аналіз, в процесі виконання якого формується тренувальна множина. Після процедури розбиття даних формується тестова множина, яка використовується для перевірки метрик оцінювання.

2.2 Методи контрольованого машинного навчання

Контрольовані методи навчання потребують, щоб люди забезпечували вхідні та необхідні результати відповідно до додавання зворотного зв'язку щодо точності прогнозування в процесі навчання [24]. Для цього використовуються алгоритми Наївний Байес, логістична регресія та дерево рішень, які застосовуються для формування моделі прогнозування зараження COVID-19 із використанням епідеміологічних наборів даних, в яких позначено позитивні та негативні випадки COVID-19.

Алгоритм ML Наївного Байєса використовується для класифікаційних навчальних завдань, в яких екземпляри набору даних дискримінуються на основі зазначеної ознаки [21]. Алгоритм має імовірнісний характер і водночас базується на теоремі Байєса [32]. Теорема Байєса:

$$P(A|C) = \frac{P(C|B)P(B)}{P(C)}. \quad (2.1)$$

Алгоритм логістичної регресії ML використовується для класифікаційних навчальних завдань, в яких визначається асоціація проти категоріально залежних ознак від незалежних ознак [33]. Алгоритм навчання використовується, коли залежні ознаки мають двійкові значення, зокрема 0 та 1 – «істина» чи «хибне», «негативне» чи «позитивне», або «Так» чи «Ні» [29]. Нижче наведено математичне рівняння алгоритму логістичної регресії (2.2), яке використовується для обчислення зв'язку між залежними ознаками та незалежними атрибутами або ознаками набору даних:

$$i = \text{Logistic regression}(p) = \ln\left(\frac{p}{1-p}\right). \quad (2.2)$$

Алгоритм дерева рішень ML використовується для розподілу навчальних занять. Де дерево будується шляхом розділення набору даних на менші набори, доки кожен розділ не стане чистим і класифікація даних залежить від типу даних [34]. Розділ атрибута набору даних числового типу даних $(B) \leq z$, де z – значення домену B для всього категоріального атрибута розділу типу даних C , утворюють значення (C) , $D \in E$, коли E дорівнює невеликому набору (B) . Для усунення шуму з набору даних використовується метод обрізки остаточної побудови дерева при повному вирощуванні [35]. Алгоритм дерева рішень використовується як один з найефективніших алгоритмів навчання завдяки своїй здатності обробляти всі типи даних, зокрема безперервні типи та деталізовані дані, розуміння та судження [36].

Підтримка векторних машин (SVM) – це алгоритм навчання, який використовується для навчальних завдань регресії та класифікації. Точки набору даних представлені у SVM-просторі поділяються на точки та групи з подібними структурами, що потрапляють в ті самі групи [37]. Площини поділяють набір меж та простір даних між групами даних для завдання регресії або класифікації [20], як показано на рисунку 2.2.

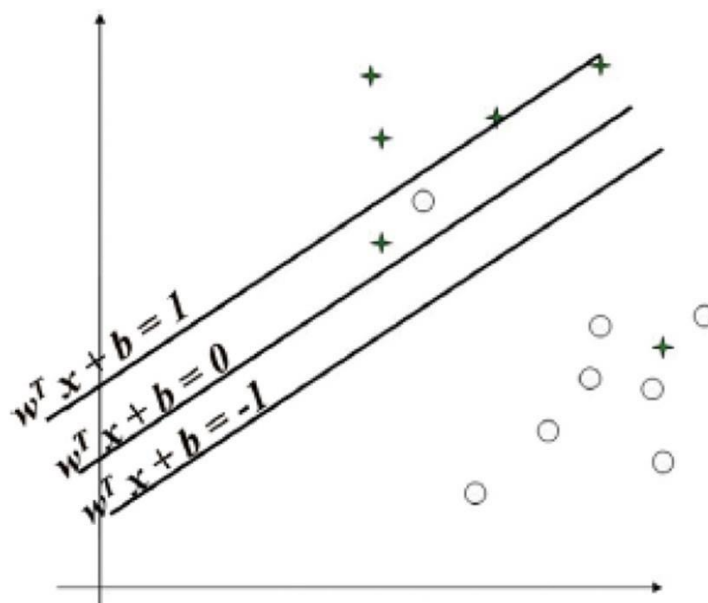


Рисунок 2.2 – Репрезентація SVM

Дані вважаються p -вимірними для лінійних SVM, які можна розділити на розмір площин $p-1$, відомих як гіперплощини [38]. Найкращий варіант вибирається на основі відстані між розділеними класами [29].

Представлення SVM між зазначеними двома категоріями називається максимальною маржею. Точка даних p визначається як:

$$(\vec{x}_1, y_1), \dots, (\vec{x}_n, y_n), \quad (2.3)$$

Реальний вектор представлений x_l , який може бути або -1 , або 1 , що представляє x_l . Гіперплощина побудована таким чином, щоб зменшити відстань між класами $y = -1$ та $y = 1$ відповідно, які визначені в рівнянні (2.4) нижче:

$$\vec{w} \cdot \vec{x} - b = 0. \quad (2.4)$$

Штучна нейронна мережа (ANN) імітує функції та діяльність мозку людини, який ідентифікується як вузли, який технічно відомий як штучні нейрони [38]. Нейрони обмінюються даними та інформацією між собою у формі 0 та 1 або їх комбінацій. Функція передачі додала всі зважені вхідні дані як:

$$z = \sum_{n=1}^n w_n x_n + w_b b, \quad (2.5)$$

Кожен нейрон має певну вагу, яка йому надається, що вказує на його функції та ролі в системі [38]. Структура ANN розділена на рівні, від рівня прийому даних, рівня введення, середнього або прихованого шару до вихідного шару, який називається шаром вилучення або класифікації. Кожен шар має певну функцію виконувати та перетворювати дані у відповідну

інформацію для отримання кінцевого та оптимального результату [23]. Функція активації та передачі відіграє вирішальну роль у діяльності, яку здійснюють нейрони. Таким чином, b представляє значення упередженості, яке часто дорівнює 1.

2.3 Аналіз коефіцієнта кореляції

Аналіз коефіцієнта кореляції використовується для визначення сильної взаємозв'язку між двома наборами ознак набору даних, які можуть бути як залежними, так і незалежними ознаками або змінними [39]. Отже, значення r є кінцевим числом від -1 до $+1$, що показує сильний взаємозв'язок між двома наборами залежних та незалежних ознак або змінних [40]. Відносини можуть бути позитивними, якщо число позитивне; так само відносини можуть бути від'ємними, якщо число від'ємне. Ідея підходу до аналізу коефіцієнта кореляції полягає в тому, що важливість відповідного набору ознак у наборі даних може бути визначена шляхом оцінки зв'язку між залежними та незалежними ознаками [41]. Набір ознак вважається хорошим для моделі ML, якщо залежні ознаки співвідносяться із незалежними ознаками [42]. Характеристику можна оцінити за рівнянням. (2.6):

$$\text{Importance} = \frac{\overline{\text{kavg}(\text{corr}_{fc})}}{\sqrt{k + k(k-1)\overline{\text{avg}(\text{corr}_{ff})}}}, \quad (2.6)$$

Це коефіцієнт кореляції між залежним набором ознак та незалежною ознакою, який є критерієм ранжування для оцінки набору ознак, $\text{avg}(\text{corr}_{fc})$ – середнє значення кореляції між залежним ознакою та незалежним ознакою, $\text{avg}(\text{corr}_{ff})$ – середнє значення кореляції між набором ознак, а k – кількість ознак.

На рисунку 2.3 показаний коефіцієнт кореляційної діаграми різних залежних ознак.

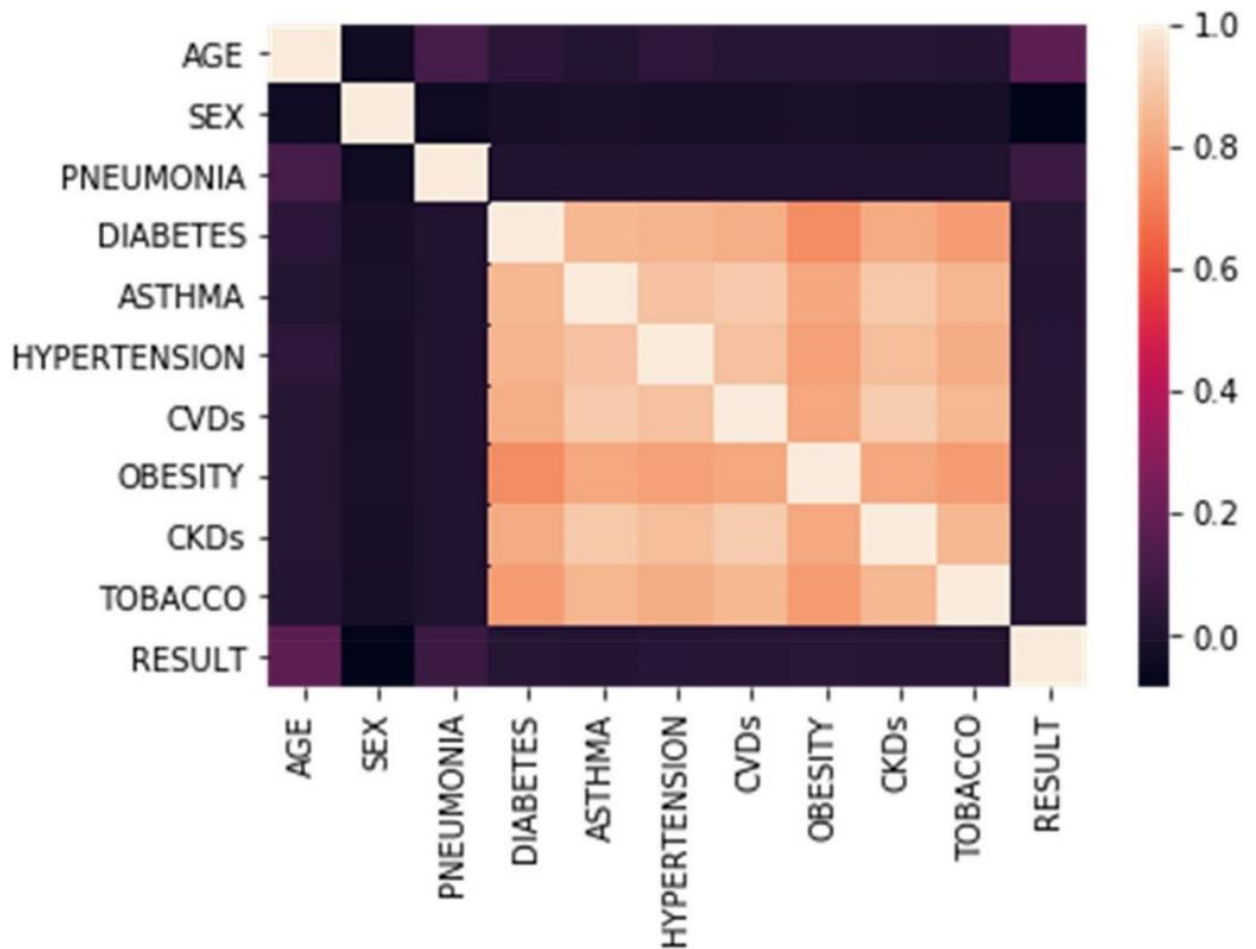


Рисунок 2.3 – Коефіцієнт кореляції Scatterplot для характеристики набору даних [31]

Зокрема:

- «AGE» – вік;
- «SEX» – стать;
- «PNEUMONIA» – пневмонію;
- «DIABETES» – діабет;
- «ASTHMA» – астму;
- «HYPERTENSION» – гіпертонію;

- «CVDs» – серцево-судинні захворювання;
- «OBESITY» – ожиріння;
- «SKDs» – ХХН;
- «TOBACCO» – куріння

Відомості подано у відношеннях до «RESULT» – результату, який є незалежною властивістю набору даних.

На рисунку 2.4 показана відповідна матриця кореляції незалежних ознак набору даних щодо залежних ознак.

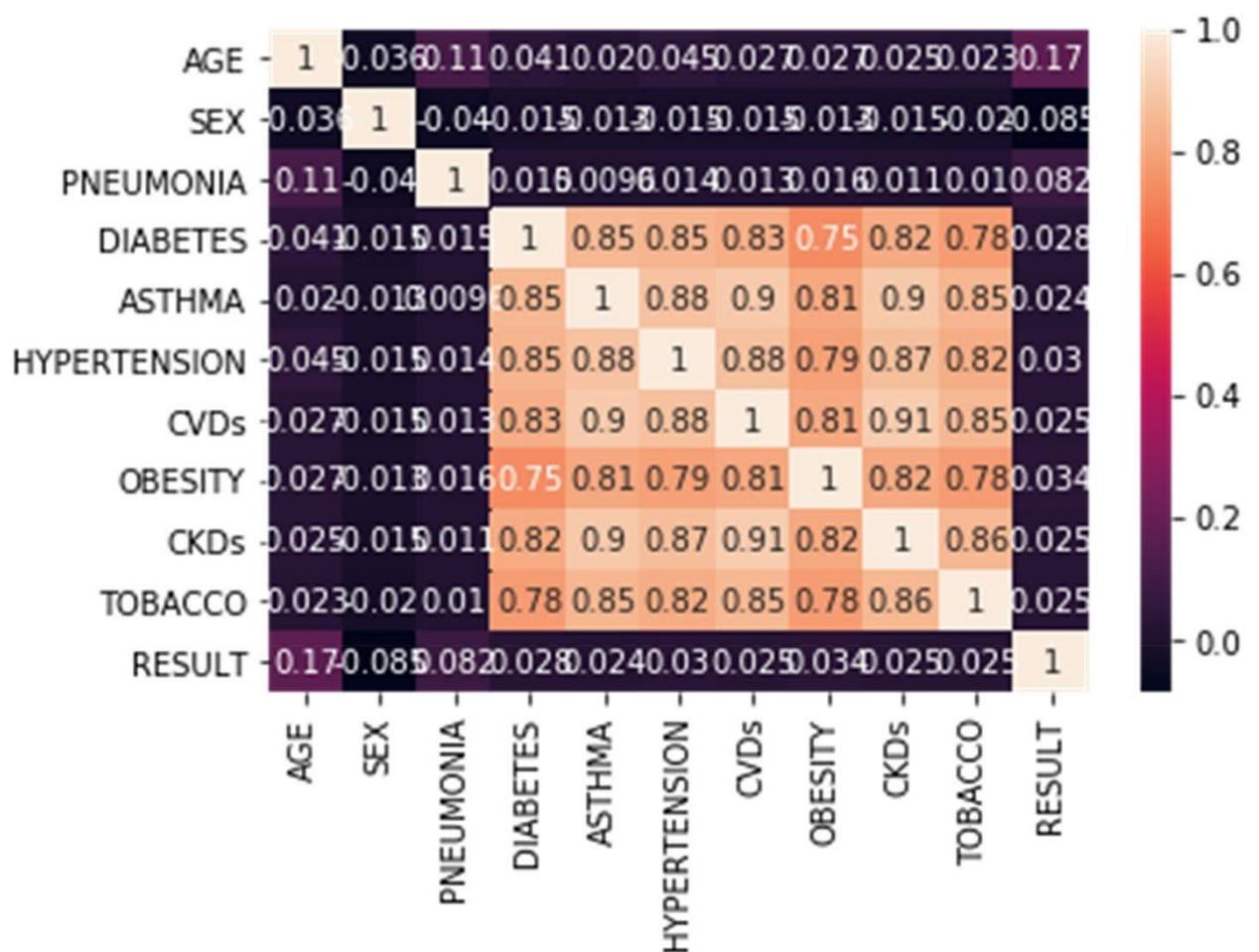


Рисунок 2.4 – Матриця кореляції характеристик набору даних [31]

В таблиці 2.1 подано відомості щодо коефіцієнта кореляції, обчисленого на основі набору даних [31].

Таблиця 2.1 – Значення r – коефіцієнта кореляції [31]

№	Атрибут		r значення	Співвідношення коефіцієнтів кореляції
	Залежний	Незалежний		
1	Вік	Результат тесту на RT-PCR COVID-19	0.17	Слабка позитивна залежність коефіцієнта кореляції
2	Стать		0.085	
3	Пневмонія		0.082	
4	Діабет		0.028	
5	Астма		0.024	
6	Гіпертонія		0.03	
7	ССЗ		0.025	
8	Ожиріння		0.034	
9	ССЗ		0.025	
10	Тютюн		0.025	

Для всіх відомостей поданих в наборі даних [31] щодо кожної залежної ознаки спостерігається слабка позитивна залежність коефіцієнта кореляції по відношенню до результату тесту «RT-PCR» COVID-19.

2.4 Прогнозні моделі щодо зараження COVID-19

Моделі ML, що використовуються для контролю зараження COVID-19 розроблені з використанням дерева рішень, логістичної регресії та наївного Байеса. Окремо використовувались алгоритми машинного навчання SVM та ANN із набором даних епідеміології для позитивних та негативних випадків COVID-19. До розробки моделі було проведено аналіз коефіцієнта кореляції між різними залежними та незалежними ознаками, щоб визначити сильний взаємозв'язок між кожною залежною та незалежною ознакою набору даних. Усі залежні ознаки мають позитивний коефіцієнт кореляції зв'язок із незалежною ознакою набору даних. Однак це залежність слабого

позитивного коефіцієнта кореляції, яку всі залежні ознаки мають від незалежної ознаки. Набір даних епідеміології для позитивних та негативних випадків COVID-19 був розділений на навчальні та тестові набори. Таким чином, моделі були навчені моделям із 80% навчальних даних та протестовані з рештою 20% набору даних. В [31] автори описують п'ять різних моделей, що були розроблені з використанням алгоритмів класифікації машинного навчання, щоб передбачити, чи інфікований пацієнт COVID-19, використовуючи алгоритми класифікації ML, зокрема дерево рішень, логістичну регресію, наївного Баєса, SVM та ANN.

Моделі оцінювались за допомогою метрики оцінки ефективності, чутливості та специфічності для визначення їх ефективності та якості.

Метрика оцінки точності показує відсоток випадків набору даних, правильно передбачених моделлю, розробленою алгоритмом машинного навчання [7]. Точність визначається за формулою:

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fn + fp}. \quad (2.7)$$

Хоча метрика оцінки чутливості показує відсоток позитивних пацієнтів на COVID-19 за моделями, як це виражається у рівнянні нижче (8).

$$\text{Sensitivity} = \frac{tp}{tp + fn}. \quad (2.8)$$

Показник оцінки специфічності віддзеркалює відсоток пацієнтів з негативним тестом COVID-19 за моделями, як це виражається у рівнянні:

$$\text{Specificity} = \frac{tn}{tn + fp}. \quad (2.9)$$

На рисунку 2.5 подана модель дерева рішень для прогнозування зараження COVID-19.

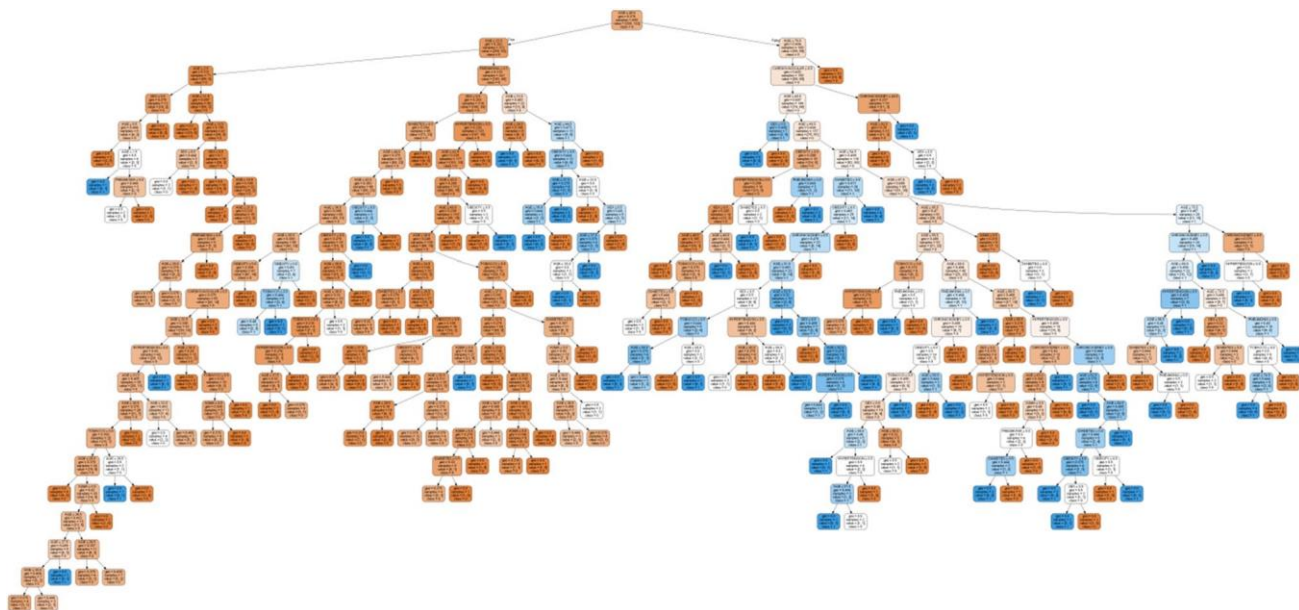


Рисунок 2.5 – Модель дерева рішень для прогнозування зараження COVID-19

2.5 Аналіз дерева рішень

Рішення логістичний алгоритм наївного Байєса забезпечує 92.40%, водночас модель прийняття рішень – 89.20%. Модель логістичної регресії – 86.34% за якою слідує модель наївного Беєса, яка має найменший показник – 83.76%. Для специфічності, яка показує відсоток негативно заражених COVID19 за моделями, модель Наїва Байєса виявилася найкращою моделлю серед усіх моделей із 94.30%, за якою слідує модель дерева рішень, яка має 93.22%, а модель логістичної регресії, яка має 87.34 %, то модель ANN, яка має 83,30%, а потім модель SVM, яка має не менше 76,50%.

Модель дерева рішень показала, що вікова характеристика є найважливішою характеристикою серед усіх залежних особливостей набору даних, включаючи клінічні особливості. Модель вказує на те, що більшість людей старше 45 років схильні до зараження COVID-19 у порівнянні з людьми молодшого віку. Подібним чином люди, які страждають на

пневмонію, ХХН, ССЗ, діабет, астму, ожиріння та гіпертонію, частіше заражаються COVID-19. Що стосується статі, чоловіки більш схильні до зараження COVID-19, ніж жінки. А курці частіше заражаються, ніж некурці. Модель допоможе медичним працівникам щодо діагностики підозрюваних хворих на COVID-19, і це доповнить тестування на RT-PCR COVID-19, тим самим зменшуючи навантаження на системи охорони здоров'я.

Моделі ML, що контролюються, можуть бути використані як методи ретроспективної оцінки або інструменти для перевірки випадків зараження COVID-19. Це дослідження показує, як прогностичні моделі зараження COVID-19 ML можуть бути розроблені, перевірені та використані як інструменти для швидкої діагностики випадків зараження COVID-19. Дослідження також показує важливі ролі та місце керованих алгоритмів ML у прогнозуванні, діагностиці та стримуванні пандемії COVID-19, що може допомогти зменшити навантаження на обмежені ресурси системи охорони здоров'я у більшості країн світу, особливо країн, що розвиваються.

2.6 Оцінювання результатів

Раннє прогнозування COVID-19 може бути корисним для зменшення навантаження на системи охорони здоров'я, допомагаючи діагностувати пацієнтів з COVID-19. У роботі [31] було розроблено дерево рішень, логістичну регресію та наївний Байєс. При цьому алгоритми SVM та ANN контролювали навчальні класифікаційні моделі для прогнозування зараження COVID-19 з використанням набору даних епідеміології для позитивних та негативних випадків COVID-19. Ефективність усіх моделей оцінювали на основі параметрів точності. Результат роботи моделей подано на рисунку 2.6 та у таблиці 2.2 відповідно.

Модель, розроблена з використанням дерева рішень, виявилася найкращою з усіх моделей, розроблених з точки зору точності, 94,99% порівняно з іншими моделями, розробленими з логістичною регресією, наївними Байесом, SVM та ANN, які мають 94,41%, 94,36%, 92,40% та Точність 89,20%. Хоча для чутливості, яка правильно показує відсоток позитивних пацієнтів на COVID-19 за моделями, SVM виявився найкращою моделлю серед усіх моделей з 93,34%, а потім ANN Model

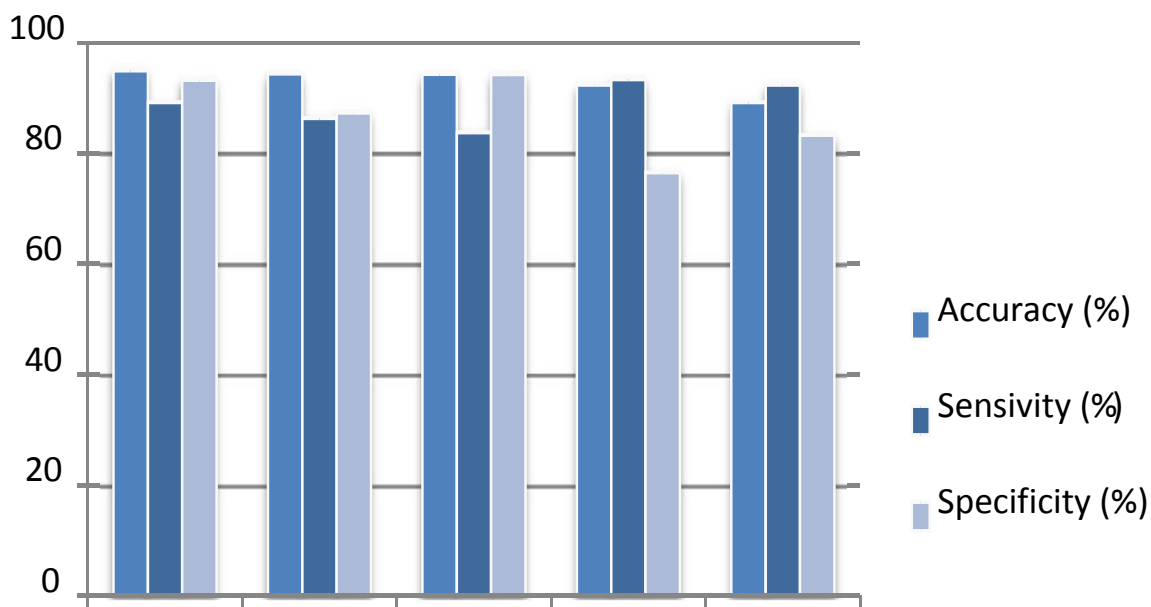


Рисунок 2.6 – Результат оцінки методів ML для аналітичного опрацювання відомостей щодо COVID-19

Таблиця 2.2 – Результат оцінки методів ML для аналітичного опрацювання відомостей щодо COVID-19 [31]

№	Модель	Точність (%)	Чутливість, %	Специфічність, %
1	2	3	4	5
1	Дерево рішень	94.99	89.2	93.22
2	Логістична регресія	94.41	86.34	87.34
3	Наївний Баес	94.36	83.76	94.3

Продовження таблиці 2.2

1	2	3	4	5
4	Підтримка векторної машини	92.4	93.34	76.5
5	Штучна нейронна мережа	89.2	92.4	83.3

На даний час пандемія COVID-19 виявляється ендемічною, як і інші інфекційні захворювання, включаючи ВІЛ/СНІД, туберкульоз, кір та гепатит. Це зачепило майже двісті тринадцять країн і територій у всьому світі та два глобальні шляхи міжнародних перевезень, тим самим змусивши ВООЗ оголосити хворобу COVID-19 надзвичайною ситуацією в галузі охорони здоров'я. Що викликає міжнародне занепокоєння. COVID-19 передається при безпосередньому контакті з інфікованою людиною через чхання та кашель і не має вакцини та ліків, дозволених з медичної точки зору. Доклінічні методи, такі як методи ML, використовуються як альтернативний засіб діагностики та прогнозування пацієнтів з інфекцією COVID-19 з метою доповнення та зменшення навантаження на обмежені системи охорони здоров'я майже у всіх країнах світу. Та з метою відродження сильно постраждалого економічного сектору. У цій роботі були досліджені контрольовані моделі ML для виявлення зараження COVID-19 сформовані на основі дерев рішень, логістичної регресії та наївними алгоритмами навчання Байєса, використовуючи епідеміологічний набір даних про позитивні та негативні випадки COVID-19. За відомостями [31] моделі були навчені з 80% даних про навчання та протестовані з рештою 20% даних. Модель, розроблена з використанням дерева рішень, виявилася найкращою з усіх моделей, розроблених з точки зору точності – 94,99%.

2.7 Висновок до другого розділу

В другому розділі кваліфікаційної роботи проаналізовано матеріали та методи дослідження. Описано методи контрольованого машинного навчання. Подано аналіз коефіцієнта кореляції. Розглянуто прогностні моделі щодо зараження COVID-19. Виконано аналіз дерева рішень. Та подано оцінювання результатів.

3 БЕЗПЕКА ЖИТТЄДІЯЛЬНОСТІ, ОСНОВИ ХОРОНИ ПРАЦІ

3.1 Долікарська допомога при кровотечах

Ознаки артеріальної кровотечі: кров яскраво-червоного кольору «б'є фонтаном», пульсує, призводить до значної і швидкої крововтрати.

Ознаки венозної кровотечі: кров безперервно витікає з рани, темно-червоного кольору; залежно від діаметру пошкодженої вени кровотеча може бути від незначної до інтенсивної.

Згідно Правил надання домедичної допомоги [43] необхідно:

1. Переконатися у відсутності небезпеки.
2. Провести огляд постраждалого, визначити наявність свідомості та дихання.
3. Викликати бригаду екстреної (швидкої) медичної допомоги.
4. Одягнути рукавички.
5. Якщо кровотечі немає:
 - підняти кінцівку;
 - накласти на рану чисту, стерильну серветку та бинтову пов'язку.
6. Якщо кровотеча артеріальна:
 - накласти на рану чисту, стерильну серветку та натиснути безпосередньо на рану;
 - підняти кінцівку;
 - якщо кровотеча не зупинена, накласти на рану пов'язку, що тисне, та при можливості одночасно притиснути артерію на відстані;
 - якщо кровотеча не зупинена, накласти джгут.
7. Якщо кровотеча венозна:
 - накласти на рану чисту, стерильну серветку та здійснити тиск безпосередньо на рану;
 - підняти кінцівку;

– якщо кровотеча не зупинена, накласти на рану пов'язку.

8. Покласти постраждалого на спину з ногами, піднятими трохи вище рівня серця.

9. Вкрити постраждалого покривалом.

Якщо дихання немає, розпочати серцево-легеневу реанімацію:

– виконати 30 натискань на грудну клітку глибиною не менше 5 см (не більше 6 см), з частотою 100 натискань (не більше 120) за хвилину;

– виконати 2 вдихи з використанням маски-клапану, дихальної маски, тощо. При відсутності захисних засобів можна не виконувати штучне дихання, а проводити тільки натискання на грудну клітку. Виконання двох вдихів повинно тривати не більше 5 секунд;

– поперемінно повторювати попередні два пункти до приїзду швидкої.

Важливо змінювати особу, що проводить натиснення на грудну клітку, кожні 2 хвилини.

3.2 Допомога при теплових і сонячних ударах

Тепловий удар – стан викликаний загальним перегрівом організму під впливом надмірного підвищення навколишньої температури [44]. Головною причиною перегрівання є порушення терморегуляції (при фізичній перенарузі, зневодненні організму, порушенні функції потових залоз). Перші ознаки теплового удару: млявість, головний біль, запаморочення, почервоніння обличчя, підвищення температури тіла, сонливість, блювота, марення. Якщо ж причини перегріву не були вчасно усунені, може бути тепловий удар, що супроводжується блідістю, посинінням шкірних покривів, шкіра при тепловому ударі холодна та вкрита потом, пульс нитковидний, спостерігається втрата свідомості.

Основними симптоми та ознаками теплового удару є:

– висока температура тіла (40 С і вище);

- з'являється спрага;
- відсутність потовиділення. При тепловому ударі, викликаному спекотною погодою, шкіра стає гарячішою і сухою на дотик. А при тепловому ударі, викликаному напруженою фізичною роботою, шкіра зазвичай волога, липка;
- почервоніння шкіри;
- прискорення дихання, підвищення частоти серцевих скорочень;
- може розвинутися пульсуючий головний біль, виникнуть такі симптоми, як: судоми, галюцинації, втрата свідомості.

Перша допомога:

- якщо ви відчули перші симптоми теплового удару, негайно викличте швидку допомогу. Якщо у вас немає можливості викликати невідкладну медичну допомогу, слід зробити наступне:
 - якщо ви перебуваєте на вулиці, негайно зайдіть в прохолодне кондиціоноване приміщення;
 - зніміть тісний одяг, розв'яжіть краватку, зніміть взуття;
 - якщо є можливість, оберніться вологим простирадлом або ввімкніть вентилятор, прийміть прохолодний душ або ванну. Так само, для зниження температури тіла можна прикласти до області шиї, спини, пахв і паху мішечки з льодом.
 - рекомендується випити 1 літр води з додаванням 2 чайних ложок солі;

У жодному випадку не пийте алкогольні напої і напої з високим вмістом кофеїну (чай, кава, капучино), оскільки вони порушують терморегуляцію організму.

Сонячний удар – різновид теплового. Сонячний удар – стан, що виникає внаслідок сильного перегріву голови прямими сонячними променями, в результаті чого розширюються мозкові кровоносні судини, відбувається приплив крові до голови. Частіше сонячний удар відбувається в

безвітряну погоду при високій вологості повітря. Перші ознаки удару: почервоніння обличчя та сильний головний біль. Далі можуть приєднатися: нудота, запаморочення, шум у вухах, потемніння в очах, блювота, задишка, втрата свідомості.

Перша допомога:

– швидко перенести постраждалого в прохолодне місце, покласти на спину, піднявши дещо ноги, зняти або розстебнути одяг. Змочити голову холодною водою або покласти на неї змочений холодною водою рушник, холодні примочки на лоб, тім'яну ділянку, потилицю, на пахові, підключичні, підколінні, пахові ділянки, де зосереджено багато кровоносних судин. Можна зробити вологе обгортання, або протерти тіло постраждалого шматочком льоду, облити його прохолодною водою, але обережно і не довго;

– у важких випадках слід одразу зважити на характер дихання постраждалого, перевірити, чи не порушена у нього прохідність дихальних шляхів. Виявивши, що язик запав, а в роті є блювотні маси, повернути голову постраждалого на бік і очистити порожнину рота бинтом або носовою хустинкою, накрученою на палець;

– якщо дихання слабке, або його немає, терміново почати робити штучне дихання методом “рот у рот” чи “рот у ніс” до появи самостійного глибокого дихання. Якщо відсутній пульс, а зіниці розширені і не реагують на світло, слід провести весь комплекс реанімації – штучне дихання і непрямий масаж серця.

3.3 Висновок до третього розділу

В третьому розділі кваліфікаційної роботи описана долікарська допомога при кровотечах. Розглянута Допомога при теплових і сонячних ударах.

ВИСНОВКИ

В першому розділі кваліфікаційної роботи освітнього рівня «Бакалавр»:

– Проаналізовано предметну область засобів та методів машинного навчання в контексті їх використання для опрацювання відомостей щодо COVID-19.

– Виконано аналіз літературних джерел.

– Описано процес підготовки та попереднього аналізу структури наборів даних.

В другому розділі кваліфікаційної роботи:

– Проаналізовано матеріали та методи дослідження.

– Описано методи контрольованого машинного навчання.

– Подано аналіз коефіцієнта кореляції.

– Розглянуто прогностні моделі щодо зараження COVID-19.

– Виконано аналіз дерева рішень.

– Та подано оцінювання результатів.

У повному обсязі виконано дослідження в розділі «Безпека життєдіяльності, основи хорони праці».

ПЕРЕЛІК ДЖЕРЕЛ

- 1 Duda, O., Pasichnyk, V., Kunanets, N., Antonii, R., & Matsiuk, O. (2020, September). Multidimensional Representation of COVID-19 Data Using OLAP Information Technology. In 2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT) (Vol. 2, pp. 277-280). IEEE.
- 2 Oleksii Duda, Liliana Dzhydzhora, Oleksandr Matsiuk, Andrii Stanko, Nataliia Kunanets, Volodymyr Pasichnyk, Oksana Kunanets. Mobile Information System for Monitoring the Spread of Viruses in Smart Cities. SISN. 2020; Volume 8: pp. 65 - 70.
- 3 Duda O., Kunanets N., Matsiuk O., Pasichnyk V., Rzhеuskyi A. (2021) Aggregation, Storing, Multidimensional Representation and Processing of COVID-19 Data. In: Shakhovska N., Medykovskyy M.O. (eds) Advances in Intelligent Systems and Computing V. CSIT 2020. Advances in Intelligent Systems and Computing, vol 1293, pp 875-889. Springer, Cham. ISBN978-3-030-63270-0.
- 4 Chinese diagnosis and treatment plan of COVID-19 patients (The fifth edition). <http://www.nhc.gov.cn/yzygj/s7653p/202002/3b09b894ac9b4204a79db5b8912d4440.shtml>. 2020.
- 5 Huang C, et al. (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet; ahead of print.
- 6 Lisa EG, Vineet DM. Return of the coronavirus: 2019-nCoV. Viruses. 2020;12(2):135. <https://doi.org/10.3390/v12020135>.
- 7 Muhammad LJ, Algehyne EA, Usman SS. Predictive supervised machine learning models for diabetes mellitus. Sn Comput Sci. 2020;1:240. <https://doi.org/10.1007/s42979-020-00250-8>.

8 Muhammad LJ, Islam MM, Usman SS, et al. Predictive data mining models for novel coronavirus (COVID-19) infected patients' recovery. Springer Nat Comput Sci. 2020. <https://doi.org/10.1007/s42979-020-00216-w>.

9 WHO, Key messages and actions for COVID-19 prevention and control in schools, March 2020.

10 Dong L, Hu S, Gao J. Discovering drugs to treat coronavirus disease 2019 (COVID-19). *Drug Discov Ther.* 2020;14:58–60.

11 <https://www.sciencemag.org/news/2020/05/unproven-herbal-remedy-again-st-covid-19-could-fuel-drug-resistant-malariascientists>.

12 Duda, O., Kunanets, N., Martsenko, S., Matsiuk, O., Pasichnyk, V., Building secure Urban information systems based on IoT technologies. *CEUR Workshop Proceedings 2623*, pp. 317-328. 2020.

13 Duda, O., et al, Selection of Effective Methods of Big Data Analytical Processing in Information Systems of Smart Cities. *CEUR Workshop Proceedings 2631*, pp. 68-78. 2020.

14 Haruna AA,. Muhammad LJ, Yahaya BZ, et al. (2019) An improved C4.5 data mining driven algorithm for the diagnosis of coronary artery disease. In: *International Conference on Digitization (ICD)*, Sharjah, United Arab Emirates, pp 48–52.

15 Mathkunti NM, Rangaswamy S. Machine learning techniques to identify dementia. *SN Comput Sci.* 2020;1:118. <https://doi.org/10.1007/s42979-020-0099-4>.

16 Rustam F, et al. COVID-19 future forecasting using supervised machine learning models. *IEEE Access.* 2020. <https://doi.org/10.1109/access.2020.2997311>.

17 Rasheed OA, Mohammed E, Iris S et al. (2020) Comparison of supervised machine learning classification techniques in prediction of locoregional recurrences in early oral tongue cancer. *Int J Med Inform* 136.

18 Dianbo L, Leonardo C, Canelle P et al. (2020) A machine learning methodology for real-time forecasting of the 2019–2020 COVID19 outbreak using Internet searches, news alerts, and estimates from mechanistic models, <https://arxiv.org/abs/2004.04019>.

19 Muhammad LJ, et al. Using decision tree data mining algorithm to predict causes of road traffic accidents, its prone locations and time along Kano–Wudil Highway. *Int J Database Theory Appl*. 2017;10(11):197–208.

20 Daniel R, Schridder A, Kern D. Supervised machine learning for population genetics: a new paradigm. *Trend Genet*. 2018;34–4:301–12.

21 Adrien PG, Yunpeng G, Gregory M, et al. A comparison of supervised machine learning algorithms for mosquito identification from backscattered optical signals. *Ecol Inform*. 2020;58:101090.

22 Sadiq H, Muhammad LJ, Yakubu A. Mining social media and DBpedia data using Gephi and R. *J Appl Comput Sci Math*. 2018;12(1):14–20.

23 Huang GB, Zhu QY, Siew CK. Extreme learning machine: theory and applications. *Neurocomputing*. 2006;70(1):489–501.

24 Li Y, Hai-Tao Z, Jorge G et al. (2020) A machine learning-based model for survival prediction in patients with severe COVID19 infection medRxiv 2020.02.27.20028027; doi: <https://doi.org/https://doi.org/10.1101/2020.02.27.20028027>.

25 Narin A, Kaya C, Pamuk Z (2020) Automatic detection of coronavirus disease (COVID-19) using X-ray images and deep convolutional neural networks. arXiv 2003.10849

26 Yang Z, Zeng Z, Wang K, et al. Modified SEIR and AI prediction of the epidemics trend of COVID-19 in China under public health interventions. *J Thorac Dis*. 2020;12(3):165–74. <https://doi.org/10.21037/jtd.2020.02.64>.

27 Ayyoubzadeh SM, Ayyoubzadeh SM, Zahedi H, et al. Predicting COVID-19 incidence through analysis of Google trends data in Iran: data mining and deep learning pilot study. *JMIR Public Health Surveill*. 2020;6(2):e18828.

28 Gurjit SR, Soltysiak MPM, El Roz H, de Souza CPE, Hill KA, Kari L. Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study, bioRxiv 2020.02.03.932350; doi: <https://doi.org/https://doi.org/10.1101/2020.02.03.932350>.

29 Edison O, Mei UW, Anthony H et al. (2020) COVID-19 coronavirus vaccine design using reverse vaccinology and machine learning. bioRxiv preprint doi: <https://doi.org/https://doi.org/10.1101/2020.03.20.000141>.

30 Malik M (2020) Machine learning the phenomenology of COVID-19 from early infection dynamics, <https://arxiv.org/abs/2003.07602>.

31 <https://www.kaggle.com/marianarfranklin/mexico-covid19-clinical-data/metadata>.

32 Ahmad IS, Bakar AA, Yaakub MR, et al. A survey on machine learning techniques in movie revenue prediction. SN Comput Sci. 2020;1:235. <https://doi.org/10.1007/s42979-020-00249-1>.

33 Asadi H, Dowling R, Yan B, Mitchell P, et al. Machine learning for outcome prediction of acute ischemic stroke post intra-arterial therapy. PLoS ONE. 2014;9:2.

34 Muhammad LJ et al. (2019) Performance evaluation of classification data mining algorithms on coronary artery disease dataset. In: IEEE 9th International Conference on Computer and Knowledge Engineering (ICCCKE 2019), Ferdowsi University of Mashhad. 2019.

35 Hussain S et al. (2019) Performance evaluation of various data mining algorithms on road traffic accident dataset. In: Satapathy S, Joshi A. (eds) Information and communication technology for intelligent systems. Smart innovation, systems and technologies, 106.

36 Muhammad LJ, Yahaya BZ, Garba A, et al. Multi query optimization algorithm using semantic and heuristic approaches. Int J Database Theory Appl. 2016;6(9):219.

37 Hearst MA, Dumais ST, Osuna E, Platt J, Scholkopf B. Support vector machines. *IEEE Intell Syst Appl.* 1998;13(4):18–28.

38 Kaur H, Kumari V. Predictive modeling and analytics for diabetes using a machine learning approach. *Appl Comput Inform.* 2018. <https://doi.org/10.1016/j.aci.2018.12.004>.

39 Deborah JR. How to Interpret a Correlation Coefficient r . *Dummies.* <https://www.dummies.com/education/math/statistics/howto-interpret-a-correlation-coefficient-r/>.

40 Muhammad LJ, Garba A, Abba G. Security challenges for building knowledge based economy in Nigeria. *Int J Secur Appl.* 2015;9(1):119.

41 Muhammad LJ, Garba A, Abba G. Security challenges for building knowledge based economy in Nigeria. *Int J Secur Appl.* 2015;9(1):119.

42 Muhammad LJ, Garba EJ, Oye ND, et al. On the problems of knowledge acquisition and representation of expert system for diagnosis of coronary artery Disease (CAD). *Int J u- and e- Serv Sci Technol.* 2018;11(3):50–9.

43 Порядок надання домедичної допомоги постраждалим при рані кінцівки, в тому числі ускладненій кровотечею. <https://zakon.rada.gov.ua/laws/show/z0759-14/paran2#n2>.

44 Перша допомога при сонячному і тепловому ударах. <https://bozhedarivska-selrada.gov.ua/news/1576496951/>.