

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет комп'ютерної інженерії та управління
(повна назва)

Кафедра електронних обчислювальних машин
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

Рівень вищої освіти другий (магістерський)

Метод прогнозування серцево-судинних захворювань
за основі методів машинного навчання

(тема)

Виконав:

студент II курсу, групи КСМм-22-2
Вракіна К.П.
(прізвище, ініціали)

Спеціальність 123 «Комп'ютерна інженерія»
(код і повна назва спеціальності)

Тип програми освітньо-професійна
(освітньо-професійна або освітньо-наукова)

Освітня програма Комп'ютерні системи та мережі
(повна назва освітньої програми)

Керівник: доц. Мартовицький В.О.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри ЕОМ

(підпис)

Коваленко А.А.

(прізвище, ініціали)

2024 р.

Харківський національний університет радіоелектроніки

Факультет _____ комп'ютерної інженерії та управління _____

Кафедра _____ електронних обчислювальних машин _____

Рівень вищої освіти _____ другий (магістерський) _____

Спеціальність _____ 123 «Комп'ютерна інженерія» _____
(код і повна назва)

Тип програми _____ освітньо-професійна _____
(освітньо-професійна або освітньо-наукова)

Освітня програма _____ Комп'ютерні системи та мережі _____
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

“ _____ ” _____ 20__ р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

студенту _____ Вракіній Каріні Павлівні _____
(прізвище, ім'я, по батькові)

1. Тема роботи Метод прогнозування серцево-судинних захворювань на основі методів машинного навчання

затверджена наказом по університету від “ 06 ” листопада 2023 р. № 1298 Ст

2. Термін подання студентом роботи до екзаменаційної комісії _____ 15 січня 2024 р.

3. Вхідні дані до роботи Спостереження лікарів, відомі спроби подібних експериментів

4. Перелік питань, що потрібно опрацювати у роботі Ознайомлення з медичною літературою, дослідження відомих експериментів у цій галузі, вивчення технології машинного навчання, створення архітектури програми, програмування, тестування.

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (слайдів) _____

Слайд-презентація – 12 слайдів _____

6. Консультанти розділів роботи (заповнюється за наявності консультантів згідно з наказом, зазначеним у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Опрацювання теоретичного матеріалу	07.11.2023	
2	Аналіз та вибір методів рішення задачі	20.11.2023	
3	Розробка алгоритмів вирішення, вибір засобів	05.12.2023	
4	Проектування програмної реалізації завдання	18.12.2023	
5	Тестування та усунення проблем	25.12.2023	
6	Оформлення пояснювальної записки	14.01.2024	

Дата видачі завдання 06 листопада 2023 р.

Студент _____
(підпис)

Керівник роботи _____
(підпис)

доц. Мартовицький В.О.
(посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка кваліфікаційної роботи: 91 с., 26 рис., 2 дод., 16 джерел.

MACHINE LEARNING, ССЗ, PYTORCH, ПРОГРАМА, КОД.

Метою кваліфікаційної роботи є дослідження і окреслення методів машинного навчання, які можуть бути ефективно застосовані для прогнозування серцевих захворювань та корекції поведінкових факторів ризику в профілактиці серцево-судинних захворювань (ССЗ).

У ході виконання кваліфікаційної роботи було поставлено задачу спрогнозувати ймовірність виникнення проблем із серцем у людей на основі інформації про їхнє здоров'я і точно визначити проблему, завдяки алгоритмам машинного навчання. Результати цього дослідження можуть бути використані медиками для встановлення точного діагнозу серця пацієнтів на основі проаналізованих даних, що передбачить можливість розвитку хвороби серця.

ABSTRACT

Master's thesis: 91 pages, 26 figures, 2 appendices, 16 sources.

MACHINE LEARNING, CVD, PYTORCH, PROGRAM, CODE.

The major goal of this thesis is research and delineation of machine learning methods that can be effectively applied to the prediction of heart diseases and correction of behavioral risk factors in the prevention of cardiovascular diseases (CVD).

In the course of this thesis, the main task was to predict the probability of heart problems among people based on information about their health and accurately identify the problem, thanks to machine learning algorithms. The results of this study can be used by doctors to establish an accurate diagnosis of the heart of patients based on the analyzed data, which will predict the possibility of heart disease progression.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ.....	7
ВСТУП	8
1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ.....	10
1.1 Обґрунтування актуальності обраної теми	10
1.2 Сфери застосування	12
1.3 Обґрунтування доцільності вдосконалення існуючих рішень.....	14
1.4 Постановка задачі.....	16
2 ПРИНЦИПИ РОБОТИ МАШИННОГО НАВЧАННЯ ТА АЛГОРИТМИ ОБРОБКИ ДАННИХ.....	18
2.1 Історія виникнення машинного навчання	18
2.2 Сфери застосування та фреймворки	23
2.3 Принципи роботи та методи	26
3 РОЗРОБКА СТРУКТУРИ І АЛГОРИТМІВ ПРОГРАМИ.....	36
3.1 Модель.....	36
3.2. Матеріали та методи	40
3.3 Результати роботи.....	48
4 РОЗРОБКА ТА ТЕСТУВАННЯ.....	56
4.1 Програмне забезпечення	56
4.2 Допоміжні прилади	57
4.3 Схема роботи	68
ВИСНОВКИ.....	71
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ	72
ДОДАТОК А Графічний матеріал кваліфікаційної роботи.....	74
ДОДАТОК Б Обробка програми	81
Б.1 Код обробки програми	81

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ
І ТЕРМІНІВ

ООП – об'єктно-орієнтоване програмування

ССЗ – серцево-судинні захворювання

ШІ – штучний інтелект

AL – активне навчання (англ., Active Learning)

EHR – електронні медичні записи (англ., Electronic Medical Records)

KDD – виявлення знань у базі даних (англ., Knowledge Discovery in Database)

ML – машинне навчання (англ., Machine Learning)

SVM – опорні векторні машини (англ., Support Vector Machines)

ВСТУП

Серцево-судинні захворювання (ССЗ) представляють собою глобальну кризу охорони здоров'я, створюючи значний тягар для систем охорони здоров'я та спричиняючи значну частку смертності в усьому світі. Поширеність серцево-судинних захворювань та їх тяжкі наслідки зробили цю проблему винятково актуальною в наш час. Надзвичайно важливо вирішити цю проблему комплексно, маючи на меті не лише пом'якшити руйнівний вплив серцево-судинних захворювань, але й зміцнити загальний рівень громадського здоров'я та знизити глобальні рівні смертності. Для досягнення цієї мети машинне навчання стало ключовим інструментом для прогнозного аналізу та втручання, пропонуючи нові шляхи виявлення, запобігання та лікування ССЗ. Це комплексне дослідження заглиблюється в формулювання проблеми, що розглядається, і ключову роль машинного навчання в прогнозуванні та профілактиці серцево-судинних захворювань.

Глобальна поширеність серцево-судинних захворювань, включаючи такі стани, як хвороби серця та інсульт, досягла тривожного рівня, а їх наслідки виходять за межі географічних кордонів і соціально-економічних відмінностей. Ця пандемія серцево-судинних захворювань започаткувала нову еру викликів у сфері охорони здоров'я, що вимагає інноваційних підходів для полегшення страждань і зниження приголомшливих рівнів смертності. Саме на цьому фоні формулювання проблеми постає центральною проблемою[4].

Проблема є багатовимірною та багатогранною, вона торкається життя мільйонів людей і впливає на якість їхнього життя. Терміновість вирішення цієї проблеми зумовлена гострою потребою у розробці інструментів прогнозування, які могли б ідентифікувати осіб із високим ризиком і забезпечити своєчасне втручання для запобігання або ефективного лікування серцево-судинних захворювань. Таким чином, роль машинного навчання

розглядається у значному світлі, пропонуючи надію та потенційні рішення в пошуках покращення громадського здоров'я та продовження життя людей.

Це комплексне дослідження охопить багатогранний ландшафт прогнозування та профілактики серцево-судинних захворювань через призму машинного навчання. Основна передумова полягає в тому, що використання потужності машинного навчання може розкрити важливу інформацію з даних пацієнтів, революціонізуючи ландшафт охорони здоров'я та врятувавши незліченну кількість життів. У наступних розділах надано детальний аналіз формулювання проблеми, значення прогнозування ССЗ та методи, за допомогою яких алгоритми машинного навчання застосовуються для вирішення цієї глобальної проблеми охорони здоров'я[5].

1 АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ

1.1 Обґрунтування актуальності обраної теми

Актуальність теми використання машинного навчання як основної технології прогнозування серцево-судинних захворювань на основі даних пацієнтів є величезною та багатогранною. Ця актуальність впливає з гострих глобальних проблем охорони здоров'я, пов'язаних із серцево-судинними захворюваннями, і трансформаційного потенціалу, який пропонує машинне навчання для вирішення цих проблем. Надалі буде перелічено декілька ключових моментів, які підкреслюють важливість і актуальність цієї теми.

Глобальна криза охорони здоров'я: серцево-судинні захворювання представляють собою глобальну кризу охорони здоров'я, яка має значний вплив на рівень смертності, системи охорони здоров'я та економіку в усьому світі. Нагальність вирішення цієї кризи підкреслюється високою поширеністю серцево-судинних захворювань, що робить цю тему надзвичайно актуальною в галузі охорони здоров'я та громадського здоров'я.

Зростання захворюваності: захворюваність на серцево-судинні захворювання зростає, вражаючи людей різних вікових груп і демографічних груп. Ця висхідна тенденція вимагає інноваційних і масштабованих рішень для виявлення та управління людьми, які знаходяться в групі ризику. Машинне навчання забезпечує керований даними підхід для вирішення цієї проблеми, що розвивається.

Велика кількість даних: завдяки поширенню електронних медичних записів, переносних пристроїв і додатків, пов'язаних зі здоров'ям, існує велика кількість даних про пацієнтів, доступних для аналізу. Машинне навчання унікально підходить для обробки та отримання цінної інформації з цих великих наборів даних, що робить його дуже актуальною технологією для прогнозування серцево-судинних захворювань[4].

Прецизійна медицина: машинне навчання дозволяє налаштовувати втручання в галузі охорони здоров'я, концепція, відома як прецизійна медицина. Пристосовуючи прогнози та втручання до окремих пацієнтів на основі їхніх унікальних профілів даних, машинне навчання сприяє більш ефективному та персоналізованому підходу до профілактики та лікування ссз[6].

Економічно ефективне охорона здоров'я. Ефективні прогностичні моделі, створені за допомогою машинного навчання, можуть допомогти більш ефективно розподіляти ресурси охорони здоров'я, зменшуючи непотрібні витрати та оптимізуючи використання ресурсів. Це особливо актуально в контексті обмежених бюджетів охорони здоров'я та необхідності надавати якісну допомогу зростаючому населенню[5].

Вплив на громадське здоров'я: точне прогнозування серцево-судинних захворювань за допомогою машинного навчання має глибокий вплив на громадське здоров'я. Це дає змогу розробляти цільові стратегії охорони здоров'я та втручання, зменшуючи загальний тягар серцево-судинних захворювань для громад і країн. Ця актуальність поширюється на державну політику та планування охорони здоров'я.

Покращення якості життя: прогнозування та профілактика серцево-судинних захворювань зосереджені не лише на подовженні життя, але й на покращенні якості життя людей із групи ризику. Виявляючи людей на ранній стадії, машинне навчання може допомогти їм приймати обґрунтовані рішення та змінювати спосіб життя, сприяючи більш здоровому та повноцінному життю[5]. Виклики масштабу: глобальний масштаб проблеми серцево-судинних захворювань вимагає рішень, які можна застосовувати у великому масштабі. Машинне навчання з його можливостями автоматизованого аналізу та масштабованістю є актуальною технологією для ефективного вирішення цієї проблеми.

Досягнення в дослідженні: постійний прогрес у дослідженнях і технологіях машинного навчання постійно розширює можливості

прогнозування ссз. Актуальність цієї теми ще більше підкреслюється потенціалом постійних інновацій та вдосконалення.

Міждисциплінарний підхід: машинне навчання для прогнозування серцево-судинних захворювань передбачає співпрацю між медичними працівниками, дослідниками даних і дослідниками. Цей міждисциплінарний підхід має важливе значення для розробки комплексних рішень, підкреслюючи актуальність цієї теми для сприяння співпраці в багатьох сферах[4].

Таким чином, актуальність використання машинного навчання як основної технології для прогнозування серцево-судинних захворювань на основі даних пацієнтів глибоко вкорінена в нагальності вирішення глобальної кризи охорони здоров'я. Трансформаційний потенціал машинного навчання, його здатність аналізувати велику кількість даних пацієнтів і його здатність надавати індивідуальні та керовані даними рішення роблять його центральною та надзвичайно актуальною технологією в прагненні зміцнити громадське здоров'я, знизити смертність і покращити якість лікування. життя людей із ризиком серцево-судинних захворювань.

1.2 Сфери застосування

Використання машинного навчання для прогнозування серцево-судинних захворювань на основі даних пацієнтів стає все більш поширеним у сфері охорони здоров'я та дослідженнях. Ось кілька прикладів застосування машинного навчання в цьому контексті:

Оцінка ризиків і раннє виявлення. Алгоритми машинного навчання використовуються для розробки моделей оцінки ризиків, які можуть передбачити ймовірність розвитку серцево-судинних захворювань у людини. Ці моделі враховують такі фактори, як вік, стать, сімейний анамнез, вибір способу життя та біомаркери для оцінки ризику[2].

Медична візуалізація: Машинне навчання використовується для аналізу медичних зображень, таких як ехокардіограми, МРТ і КТ, щоб виявити структурні аномалії, оцінити роботу серця та виявити ранні ознаки серцевих захворювань.

Носимі пристрої – Smart Wearables: Носимі пристрої, оснащені датчиками, постійно збирають дані про частоту серцевих скорочень, артеріальний тиск і фізичну активність. Алгоритми машинного навчання обробляють ці дані, щоб відстежувати зміни з часом, виявляти аномалії та надавати користувачам і постачальникам медичних послуг зворотний зв'язок у реальному часі[3].

Електронні медичні записи (EHR): Інтелектуальний аналіз даних. Машинне навчання використовується для аналізу електронних медичних записів для отримання цінної інформації. Він може ідентифікувати моделі та кореляції в даних пацієнтів, допомагаючи клініцистам приймати більш обґрунтовані рішення щодо лікування та профілактики.

Геномні дані – прогноз геномного ризику: методи машинного навчання застосовуються до геномних даних для ідентифікації генетичних маркерів, пов'язаних із серцево-судинними захворюваннями. Це дозволяє розробляти точніші прогностичні моделі на основі генетичного профілю людини.

Стратифікація пацієнтів: Алгоритми машинного навчання можуть класифікувати пацієнтів у різні групи ризику на основі даних про їх здоров'я. Це допомагає виявити осіб із групи високого ризику, які потребують більш інтенсивного моніторингу та втручання.

Прихильність до лікування: Моделі машинного навчання можуть передбачати прихильність до лікування на основі історичних даних пацієнта та моделей поведінки. Це має вирішальне значення для того, щоб пацієнти дотримувалися призначеного плану лікування.

Телемедицина та віддалений моніторинг – віддалений моніторинг здоров'я: Машинне навчання використовується в телемедичних рішеннях для дистанційного моніторингу життєво важливих показників пацієнтів і

виявлення аномалій, забезпечуючи своєчасне втручання та зменшуючи потребу в особистих візитах.

Навчання пацієнтів і поведінкові втручання: Машинне навчання може аналізувати дані пацієнтів, щоб надавати персоналізовані рекомендації щодо зміни способу життя, наприклад дієти та фізичних вправ, щоб зменшити ризик серцево-судинних захворювань[7].

Підтримка клінічних рішень - інструменти прогнозування ризиків: Системи підтримки клінічних рішень на основі машинного навчання допомагають постачальникам медичних послуг приймати рішення щодо лікування та втручань, надаючи оцінку ризиків і рекомендації на основі доказів.

Дослідження та розробка ліків: Машинне навчання використовується для аналізу великих наборів даних для виявлення потенційних препаратів-кандидатів для лікування серцево-судинних захворювань. Він також може передбачити реакцію пацієнта на певні ліки, створюючи більш персоналізовані плани лікування[11].

Управління здоров'ям населення: Машинне навчання застосовується до даних про стан здоров'я на рівні населення, щоб визначити тенденції, кластери та регіони з високим ризиком, що дозволяє органам охорони здоров'я впроваджувати цільові заходи та політику.

Здатність технології обробляти й аналізувати великі набори даних, виявляти тонкі закономірності та надавати персоналізоване розуміння має великі надії на покращення профілактики, лікування та загальних результатів для людей із ризиком серцево-судинних захворювань або тих, хто живе з ними.

1.3 Обґрунтування доцільності вдосконалення існуючих рішень

Значення прогнозування серцево-судинних захворювань неможливо переоцінити. Своєчасне та точне прогнозування серцево-судинних захворювань пропонує безліч переваг як на індивідуальному, так і на

суспільному рівнях. Деякі з ключових факторів, які підкреслюють значення прогнозування ССЗ, включають:

Раннє втручання – раннє втручання осіб із ризиком серцево-судинних захворювань дозволяє випадково втручатись, наприклад, спосіб змінити життя або лікуватися. Це може значно зменшити тяжкість захворювання та покращити результати[4].

Профілактична медична допомога прогностичні моделі дозволяють впроваджувати профілактичні заходи охорони здоров'я, адаптовані до конкретної людини, зосереджуючись на факторах ризику та поведінки, які можна змінити, щоб зменшити ймовірність ССЗ.

Розподіл ресурсів – ресурси охорони здоров'я часто обмежені. Інструменти прогнозування можуть допомогти більш ефективно розподілити ці ресурси, визначаючи осіб із найвищим ризиком, гарантуючи, що вони отримають увагу та погляд, які вони потребують.

Планування громадської охорони здоров'я. Установлення громадської охорони здоров'я може отримати прибуток від прогностичних моделей для розробки та впровадження цільового здоров'я у сферу охорони здоров'я та політики, спрямованої на зменшення загального тягаря серцево-судинних захворювань.

Зниження витрат на охорону здоров'я – тимчасова профілактика та втручання можуть призвести до зменшення витрат на охорону здоров'я, пов'язаних із лікуванням ССЗ. Це, у своїй системі, може зменшити фінансовий тиск на окремих осіб і охорони здоров'я.

Покращена якість життя – прогнозування та запобігання серцево-судинним захворюванням не тільки рятує життя, але й покращує якість життя людей із групою ризику. Люди можуть прийняти обґрунтовані рішення щодо свого здоров'я, що призводить до кращого самопочуття.

Довгострокове управління здоров'ям – прогнозування ССЗ не обмежується початковою оцінкою ризику; це також добре в довгостроковому управлінні здоров'ям. Постійний моніторинг і коригування на основі

прогнозних моделей можуть допомогти людям підтримувати здоров'я серцево-судинної системи з часом.

У світі цих факторів очевидно, що прогнозування серцево-судинних захворювань є трансформаційним підходом до управління цією глобальною проблемою охорони здоров'я. Машинне навчання, здатне аналізувати величезну кількість даних про допомогу та виявляти складні закономірності, може революціонізувати прогнозування серцево-судинних захворювань і, як наслідок, покращити громадське здоров'я та добробут у глобальній масштабі[7].

1.4 Постановка задачі

В основі проблеми лежить нагальна потреба вирішити проблему високого рівня смертності, пов'язаної з серцево-судинними захворюваннями. Цей імператив впливає з самого масштабу проблеми та її далекосяжних наслідків, що впливають на окремих людей, громади та нації. Серцево-судинні захворювання є основною причиною смерті в усьому світі, завдаючи значних збитків системам охорони здоров'я та економіці. Грандіозність проблеми неможливо переоцінити, і вона потребує проактивної та стратегічної реакції.

Постановка проблеми охоплює кілька критичних аспектів:

Високий рівень смертності: серцево-судинні захворювання спричиняють значну частку глобальної смертності, що робить їх першочерговою проблемою громадського здоров'я. Вони забирають життя із загрозливою швидкістю, і їхній вплив виходить за рамки просто статистичних даних, залишаючи родини втраченими, а суспільства припиненими[4].

Поширеність: серцево-судинні захворювання широко поширені, вражаючи людей різного віку та походження. Цей широкомасштабний вплив підкреслює необхідність системного та цілеспрямованого підходу до вирішення проблеми.

Якість життя: крім смертності, серцево-судинні захворювання значно

погіршують якість життя хворих. Це може призвести до погіршення здоров'я, зниження продуктивності.

Економічні наслідки: серцево-судинні захворювання лягають важким економічним тягарем на системи охорони здоров'я, уряди та окремих людей. Вартість лікування, втрата продуктивності та потреба в довгостроковому догляді ще більше ускладнюють проблему.

Глобальний вплив: проблема не обмежується певним регіоном чи країною, вона завжди виходить за рамки кордонів і впливає на різні групи населення. Отже, це вимагає скоординованих глобальних зусиль для ефективного вирішення проблеми[5].

У світлі цих багатогранних викликів формулювання проблеми стає критично важливою вправою для визначення найбільш ефективних стратегій прогнозування та запобігання. МЛобіцяючий шлях вирішення цих проблем і внесення трансформаційних змін у підхід до ССЗ. Використовуючи потужність даних і складних алгоритмів, машинне навчання може надати інформацію, яка дозволить медичним працівникам прогнозувати та зменшувати ризик серцево-судинних захворювань із безпрецедентною точністю та ефективністю.

Основна задача цієї стратегії полягає в тому, щоб спрогнозувати ймовірність виникнення проблем із серцем у людей на основі інформації про їхнє здоров'я (холестерин, артеріальний тиск, вік тощо), що дозволяє швидко й точно визначити проблему. У цьому дослідженні було використано десять алгоритмів машинного навчання, включаючи SVM, Random Forest, Gradient Boost, Decision Tree, eXtreme Gradient Boost, KNN, LR і MLP. Запропоноване дослідження аналізує дані про продуктивність і оцінює кілька методів класифікації, щоб передбачити захворювання серця. За результатами обстеження буде поставлений точний діагноз стану серця пацієнта. Звіт про стан здоров'я пацієнта використовується лікарем для введення даних. Модель будується з використанням даних як вхідних і передбачає можливість розвитку хвороби серця.

2 ПРИНЦИПИ РОБОТИ МАШИННОГО НАВЧАННЯ ТА АЛГОРИТМИ ОБРОБКИ ДАНИХ

2.1 Історія виникнення машинного навчання

Термін «машинне навчання» був введений у 1959 році Артуром Самуелем, співробітником ІВМ і піонером у галузі комп'ютерних ігор і штучного інтелекту. Синонім самонавчання комп'ютерів також використовувався в цей період[8].

На початку 1960-х років компанія Raytheon розробила експериментальну «навчальну машину» з пам'яттю на перфострічці під назвою «Кібертрон» для аналізу сигналів гідролокатора, електрокардіограм і мовленнєвих моделей за допомогою елементарного навчання з підкріпленням. Його постійно «навчав» людина-оператор/вчитель розпізнавати закономірності та оснащував кнопкою «дурень», щоб змушувати його повторно оцінювати неправильні рішення. Репрезентативною книгою про дослідження машинного навчання в 1960-х роках була книга Нільссона про Learning Machines, присвячена переважно машинному навчанню для класифікації шаблонів. Інтерес, пов'язаний з розпізнаванням образів, тривав у 1970-х роках, як описано Дудою та Хартом у 1973 році. У 1981 році була зроблена доповідь про використання стратегій навчання, щоб нейронна мережа навчилася розпізнавати 40 символів (26 літер, 10 цифр і 4 спеціальні символи) з комп'ютерного терміналу.

Том М. Мітчелл надав широко цитоване, більш формальне визначення алгоритмів, які вивчаються в області машинного навчання: «Кажуть, що комп'ютерна програма вчиться на досвіді E щодо деякого класу завдань T і мірою продуктивності P , якщо її продуктивність у завданнях у T , як вимірюється P , покращується з досвідом E ." Це визначення завдань, у яких йдеться про машинне навчання, пропонує принципово операційне визначення,

а не визначення поля в когнітивних термінах. Це слідує за пропозицією Алана Тюрінга в його статті "Обчислювальна техніка та інтелект", в якій питання "Чи можуть машини мислити?" замінюється питанням «Чи можуть машини робити те, що ми (як мислячі сутності) можемо робити?».

Сучасне машинне навчання має дві мети: одна – класифікувати дані на основі розроблених моделей, інша – спрогнозувати майбутні результати на основі цих моделей. Гіпотетичний алгоритм для класифікації даних може використовувати комп'ютерний зір родимок у поєднанні з навчанням під наглядом, щоб навчити його класифікувати ракові родимки. Алгоритм машинного навчання для біржової торгівлі може інформувати трейдера про майбутні потенційні прогнози.

Як науковий напрям, машинне навчання виникло в результаті пошуків штучного інтелекту (ШІ). На початку ШІ як академічної дисципліни деякі дослідники були зацікавлені в тому, щоб машини навчалися на даних. Вони намагалися підійти до проблеми різними символічними методами, а також тим, що тоді називали «нейронними мережами»; це були в основному перцептрон та інші моделі, які пізніше виявились перевинаходами узагальнених лінійних моделей статистики. Також використовували ймовірнісні міркування, особливо в автоматизованій медичній діагностиці[6].

Однак посилення акценту на логічному, заснованому на знаннях підході спричинило розрив між ШІ та машинним навчанням. Імовірнісні системи страждали від теоретичних і практичних проблем збору та представлення даних. До 1980 року експертні системи стали домінувати над штучним інтелектом, а статистика була в опіці. Робота над символічним навчанням/навчанням, заснованим на знаннях, продовжувалася в рамках штучного інтелекту, що призвело до індуктивного логічного програмування, але більш статистична лінія досліджень тепер була поза областю власне штучного інтелекту, у розпізнаванні образів та пошуку інформації. Дослідження нейронних мереж були залишені ШІ та інформатикою приблизно в той же час. Цю лінію також було продовжено за межами області AI/CS, як

«коннекціонізм» дослідниками з інших дисциплін, включаючи Хопфілда, Румельхарта та Гінтона. Їхній головний успіх прийшов у середині 1980-х років із повторним винаходом зворотного поширення.

Машинне навчання (ML), реорганізоване та визнане окремою сферою, почало процвітати в 1990-х роках. Сфера змінила свою мету з досягнення штучного інтелекту на вирішення вирішуваних проблем практичного характеру. Він змістив фокус із символічних підходів, успадкованих від ШІ, на методи та моделі, запозичені зі статистики, нечіткої логіки та теорії ймовірностей[6].

Машинне навчання та інтелектуальний аналіз даних часто використовують однакові методи та значною мірою збігаються, але в той час як машинне навчання зосереджується на передбаченні на основі відомих властивостей, отриманих із навчальних даних, інтелектуальний аналіз даних фокусується на виявленні (раніше) невідомих властивостей у даних (це етап аналізу виявлення знань у базах даних). Інтелектуальний аналіз даних використовує багато методів машинного навчання, але з різними цілями; з іншого боку, машинне навчання також використовує методи інтелектуального аналізу даних як «навчання без нагляду» або як етап попередньої обробки для підвищення точності учня. Значна частина плутанини між цими двома дослідницькими спільнотами (які часто мають окремі конференції та окремі журнали, основним винятком є ECML PKDD) виходить із основних припущень, з якими вони працюють: у машинному навчанні ефективність зазвичай оцінюється з огляду на здатність відтворювати відомі знання, тоді як у виявленні знань і інтелектуальний аналіз даних (KDD) ключовим завданням є виявлення раніше невідомих знань. Оцінений з огляду на відомі знання, неінформований (неконтрольований) метод буде легко перевершувати інші контрольовані методи, тоді як у типовому завданні KDD контрольовані методи не можуть бути використані через відсутність навчальних даних[8].

Машинне навчання також має тісний зв'язок з оптимізацією: багато проблем навчання формулюються як мінімізація деякої функції втрат на

навчальному наборі прикладів. Функції втрат виражають розбіжність між прогнозами моделі, що навчається, і фактичними екземплярами проблеми (наприклад, у класифікації потрібно призначити мітку екземплярам, а моделі навчаються правильно прогнозувати попередньо призначені мітки набору приклади)[13].

Різниця між оптимізацією та машинним навчанням виникає через мету узагальнення: хоча алгоритми оптимізації можуть мінімізувати втрати на навчальному наборі, машинне навчання займається мінімізацією втрат на невидимих зразках. Характеристика узагальнення різних алгоритмів навчання є активною темою поточних досліджень, особливо для алгоритмів глибокого навчання.

Машинне навчання та статистика є тісно пов'язаними областями з точки зору методів, але відрізняються своєю основною метою: статистика робить висновки про популяцію на основі вибірки, тоді як машинне навчання знаходить узагальнювані прогностичні моделі. За словами Майкла І. Джордана, ідеї машинного навчання, від методологічних принципів до теоретичних інструментів, мають довгу передісторію в статистиці. Він також запропонував термін наука про дані як заповнювач, щоб назвати загальну область.

Традиційний статистичний аналіз вимагає апріорного вибору моделі, яка найбільше підходить для набору даних дослідження. Крім того, для аналізу включено лише суттєві або теоретично релевантні змінні, засновані на попередньому досвіді. Навпаки, машинне навчання не побудоване на попередньо структурованій моделі; скоріше дані формують модель, виявляючи основні закономірності. Чим більше змінних (вхідних даних) використовується для навчання моделі, тим точнішою буде остаточна модель.

Лео Брейман розрізняв дві парадигми статистичного моделювання: модель даних і алгоритмічну модель, де «алгоритмічна модель» означає більш-менш алгоритми машинного навчання, такі як Random Forest.

Деякі статистики перейняли методи машинного навчання, що призвело до комбінованої галузі, яку вони називають статистичним навчанням.

Аналітичні та обчислювальні методи, отримані з глибоко вкоріненої фізики неупорядкованих систем, можуть бути поширені на масштабні проблеми, включаючи машинне навчання, наприклад, для аналізу вагового простору глибоких нейронних мереж. Таким чином, статистична фізика знаходить застосування в області медичної діагностики.

Обчислювальний аналіз алгоритмів машинного навчання та їх продуктивності – це розділ теоретичної інформатики, відомий як теорія обчислювального навчання за допомогою моделі ймовірно приблизно правильного навчання (РАС). Оскільки навчальні набори скінченні, а майбутнє невизначене, теорія навчання зазвичай не дає гарантій продуктивності алгоритмів. Натомість досить поширеними є ймовірнісні межі продуктивності. Декомпозиція відхилення–дисперсія є одним із способів кількісної оцінки помилки узагальнення.

Для найкращої продуктивності в контексті узагальнення складність гіпотези повинна відповідати складності функції, що лежить в основі даних. Якщо гіпотеза є менш складною, ніж функція, то модель недостатньо підігнала дані. Якщо у відповідь ускладнити модель, то похибка навчання зменшується. Але якщо гіпотеза занадто складна, то модель підлягає переобладнанню, і узагальнення буде гіршим[6].

На додаток до меж продуктивності, теоретики навчання вивчають часову складність і доцільність навчання. У теорії обчислювального навчання обчислення вважається можливим, якщо його можна виконати за поліноміальний час. Існує два типи результатів складності часу: Позитивні результати показують, що певний клас функцій можна вивчити за поліноміальний час. Негативні результати показують, що певні класи не можна вивчити за поліноміальний час.

2.2 Сфери застосування та фреймворки

Машинне навчання спирається на алгоритми. Ці алгоритми дуже складні для розуміння та роботи, якщо ви не фахівець з даних або експерт з машинного навчання.

Таким чином, структура машинного навчання спрощує алгоритми машинного навчання. Фреймворк ML – це будь-який інструмент, інтерфейс або бібліотека, які дозволяють легко розробляти моделі ML, не розуміючи базових алгоритмів[8].

Існує безліч фреймворків машинного навчання, призначених для різних цілей. Можливо, TensorFlow, PyTorch і scikit-learn є найпопулярнішими фреймворками ML. Однак вибір фреймворка для використання залежатиме від роботи, яку ви намагаєтеся виконати. Ці фреймворки орієнтовані на математику та статистичне моделювання (машинне навчання) на відміну від навчання нейронних мереж (глибоке навчання).

TensorFlow і PyTorch є прямими конкурентами через свою схожість. Вони обидва надають багатий набір інструментів лінійної алгебри та можуть виконувати регресійний аналіз.

Scikit-learn існує вже давно і найбільше знайомий програмістам R, але він має велике застереження: він не створений для роботи в кластері.

Spark ML створено для роботи в кластері, оскільки це суть Apache Spark.

А тепер ближче розглянемо деякі конкретні фреймворки.

TensorFlow був розроблений у Google Brain, а потім перетворений у проект з відкритим кодом. TensorFlow може: виконувати регресію, класифікацію, нейронні мережі тощо; працює як на процесорах, так і на графічних процесорах.

TensorFlow є одним із фреймворків де-факто машинного навчання, які використовуються сьогодні, і він безкоштовний.

TensorFlow – це повноцінний дослідницький і виробничий інструмент для машинного навчання. Це може бути дуже складним, але це не обов'язково.

Як і електронну таблицю Excel, TensorFlow можна використовувати просто або більш досвідчено:

TF є досить простим для звичайного користувача, який хоче повернути передбачення на заданому наборі даних. TF також може працювати для досвідчених користувачів, які бажають налаштувати кілька конвеєрів даних, трансформувати дані відповідно до своєї моделі, налаштувати всі шари та параметри своєї моделі та навчатися на кількох машинах, зберігаючи конфіденційність користувача[6].

TF вимагає глибокого розуміння масивів NumPy. TensorFlow побудований з тензорів. Це спосіб обробки тензорів; отже, інструмент NumPy від Python. NumPy – це платформа Python для роботи з n-вимірними масивами (1-вимірний масив – це вектор, 2-вимірний масив – це матриця тощо). Замість того, щоб автоматично перетворювати масиви на одноразові вектори (істинно-хибне представлення), очікується, що це завдання буде виконувати спеціаліст з даних.

Але TensorFlow має багатий набір інструментів. Наприклад, функції активації для нейронних мереж можуть виконувати всю важку статистичну роботу. Якщо ми визначаємо глибоке навчання як здатність створювати нейронні мережі, то TensorFlow це робить. Але він також може вирішувати більш повсякденні проблеми, як-от регрес.

PyTorch був розроблений FAIR, Facebook AI Research. На початку 2018 року команда FAIR об'єднала Caffe2, іншу структуру машинного навчання, у PyTorch. Це провідний конкурент TensorFlow. Коли інженери вирішують використовувати платформу ML, їхній вибір зазвичай зводиться до питання: «Ми використовуємо TensorFlow чи PyTorch?» Кожен з них служить своїм цілям, але є досить взаємозамінними.

Як TensorFlow, PyTorch: займається регресією, класифікацією, нейронними мережами тощо; працює як на процесорах, так і на графічних процесорах; PyTorch вважається більш пітонічним. Там, де TensorFlow може запустити модель швидше та з деякими налаштуваннями, PyTorch вважається

більш настроюваним, дотримуючись більш традиційного підходу до об'єктно-орієнтованого програмування через створення класів.

Досліджено, що PyTorch має швидший час навчання. Ця швидкість є незначною для багатьох користувачів, але може мати значення для великих проектів. PyTorch і TensorFlow знаходяться в активній розробці, тому порівняння швидкості, ймовірно, не буде різним.

Іноді для вимірювання вірогідного успіху гіпотези потрібен лише швидкий тест. Scikit-learn – це старий стандарт у світі науки про дані, і може бути корисним запустити швидкі ескізи моделі ML, щоб побачити, чи може модель мати певну інтерпретацію.

Scikit – ще один пакет Python, який може виконувати багато корисних завдань машинного навчання: лінійна регресія, регресії дерева рішень, регресії випадкового лісу, K-nearest neighbour, SVM, стохастичні моделі градієнтного спуску.

Scikit надає такі інструменти аналізу моделі, як матриця плутанини, щоб оцінити, наскільки добре працює модель. Багато разів ви можете розпочати роботу з ML у scikit-learn, а потім перейти до іншого фреймворку. Наприклад, scikit-learn має чудові інструменти попередньої обробки даних для швидкого кодування категоріальних даних. Після попередньої обробки даних за допомогою Scikit ви можете перемістити їх у TensorFlow або PyTorch.

Spark ML – може працювати в кластерах. Іншими словами, він може обробляти дійсно велике множення матриці, беручи фрагменти матриці та запускаючи це обчислення на різних серверах. (Матричне множення є однією з найважливіших операцій машинного навчання.) Для цього потрібна розподілена архітектура, щоб у вашому комп'ютері не вичерпувалося пам'яті та він не працював занадто довго під час роботи з великими обсягами даних.

Spark ML є складним, але замість того, щоб працювати з масивами NumPy, він дозволяє працювати зі структурами даних Spark RDD, що зрозуміє кожен, хто використовує Spark у ролі великих даних. І ви можете використовувати Spark ML для роботи з кадрами даних Spark SQL, які знають

більшість програмістів Python. Таким чином, він створює для вас щільні та іскристі вектори міток ознак, усуваючи певну складність підготовки даних для введення в алгоритми машинного навчання.

Torch стверджує, що це найпростіший фреймворк ML. Це стара бібліотека машинного навчання, вперше випущена в 2002 році. Раніше, у PyTorch, Python був обраним методом доступу до основних таблиць, у яких Torch виконує свої обчислення. Сам Torch можна використовувати за допомогою Lua з менеджером пакетів LuaRocks. Відносна простота Torch пояснюється його інтерфейсом мови програмування Lua. Lua справді проста. Тут немає чисел з плаваючою речовиною чи цілих чисел, лише числа. І всі об'єкти в Lua є таблицями. Таким чином, легко створювати структури даних. І він надає багатий набір простих для розуміння функцій для розділення таблиць і додавання до них.

Як і TensorFlow, основним елементом даних у Torch є тензор. Ви створюєте його, написавши `torch.Tensor`. CLI (інтерфейс командного рядка) надає вбудовану довідку та допомагає з відступами. Люди, які користувалися Python, відчують полегшення, оскільки це означає, що ви можете вводити функції на місці, не починаючи спочатку, коли ви зробили помилку. А для тих, хто любить складність і розріджений код, Torch підтримує функціональне програмування. Іншими новими фреймворками майбутнього є huggingface.co, що є однією з найкращих бібліотек машинного навчання, яка створює хороші базові моделі для дослідників, створені на основі TensorFlow і PyTorch та Keras – це бібліотека нейронних мереж, створена на основі TensorFlow, щоб зробити моделювання машинного навчання простим[8].

2.3 Принципи роботи та методи

Основна функція машинного навчання включає в себе два основних етапи: вивчення даних, визначення шаблонів і створення моделі на їх основі та прийняття рішень і прогнозування на основі моделей.

Це не передбачає явного програмування; натомість алгоритми машинного навчання (ML) навчають програмне забезпечення, щоб воно стало здатним до навчання та вдосконалення. Ці алгоритми потребують величезних обсягів інформації, а їх ретельне вивчення, синтаксичний аналіз і аналіз дозволяють їм ідентифікувати закономірності та розробляти інструкції для прийняття рішень і прогнозів.

Досить часто ML вважають частиною штучного інтелекту; інші кажуть, що лише його «інтелектуальна» частина є підполем ШІ. У «Книзі чому» Джудея Перл пояснює різницю між машинним навчанням і штучним інтелектом: машинне навчання покладається на пасивне спостереження, тоді як ШІ активно взаємодіє з навколишнім середовищем. У будь-якому випадку, ML має багато спільних підходів і принципів роботи зі ШІ.

Розробка програмного забезпечення на основі ML значною мірою залежить від підготовки моделі. Серед ключових етапів процесу виділяють наступне.

Вибір і підготовку навчального набору даних. Навчальні дані – це набір даних, який представляє інші дані, які модель ML оброблятиме та перероблятиме під час вирішення свого завдання[13].

Залежно від випадку навчальні дані можуть бути помічені, що означає, що вони мають певну функцію або належать до певного класу, який модель повинна ідентифікувати. Якщо дані не позначені, модель позначає функції або призначає класи самостійно.

Обов'язкова вимога до навчальних даних полягає в тому, що вони повинні бути підготовлені таким чином, щоб їх було легко обробляти моделлю ML. Отже, його слід проаналізувати, нормалізувати та розділити на кілька підмножин, які використовуватимуться для навчання, тестування та оцінювання.

Підготовка даних дуже важлива. Кажуть, що робота спеціаліста з даних полягає на 90% у підготовці даних і лише на 10% у навчанні моделі на підготовлених даних.

Вибір методу машинного навчання. Специфіка проектів розробки програмного забезпечення на основі машинного навчання залежить від методу машинного навчання, який обирає команда: контрольоване, неконтрольоване, підкріплення та напівконтрольоване навчання. Метод визначається відповідно до способу навчання алгоритмів[10].

Контрольоване навчання – як впливає з назви, цей тип навчання виконується під контролем спеціаліста з даних. Спеціаліст з даних маркує вхідні дані, встановлює необхідні змінні, визначає бажані результати та визначає правильність оцінки. Цей тип навчання добре працює в таких випадках:

- регресійне моделювання;
- ансамблювання;
- бінарна класифікація;
- багатокласова класифікація;
- виявлення та сегментація об'єктів;
- навчання без контролю.

З назви також видно, що цей тип розробки програмного забезпечення на основі ML не потребує спеціаліста для моніторингу того, як навчаються алгоритми. Неконтрольоване навчання спирається на заздалегідь визначені набори даних і прогнози.

У цьому випадку алгоритми навчаються, використовуючи навчальні дані без міток, і перевіряють набори даних на наявність шаблонів, які можуть допомогти вибрати значущі зв'язки та підмножини даних. Цей тип навчання використовується для:

- виявлення аномалії;
- кластеризації даних;
- зменшення ваги;
- об'єднання видобутку.

Навчання без нагляду є основою глибокого навчання та нейронних мереж.

Напівконтрольоване навчання – це поєднання контрольованого та неконтрольованого навчання, що означає, що дослідники даних надають алгоритмам дані з мітками, але алгоритми можуть досліджувати дані та розвивати власне розуміння без втручання людини. На основі позначених даних, наданих людьми, алгоритми визначають особливості даних і застосовують їх до непозначених даних.

Чому обирають цей підхід? Це легше керувати, ніж неконтрольований метод, але є більш економічно ефективним і займає менше часу, ніж контрольоване навчання. Найбільш яскравими прикладами його застосування в проектах розробки програмного забезпечення є системи виявлення шахрайства та машинного перекладу.

Навчання з підкріпленням – цей тип ML використовується, коли процес, який машина повинна навчитися виконувати, складається з багатьох кроків, і кожен з них має чітко визначені правила. У процесі навчання дослідники даних забезпечують алгоритм зворотним зв'язком і підказками щодо якості його роботи. Це заохочуватиме алгоритм ML шукати позитивні винагороди (підкріплення).

Гарними прикладами використання навчання з підкріпленням є відеоігри, програмне забезпечення для автоматизації на основі робототехніки та управління ресурсами.

Вибір алгоритму для виконання з навчальними даними. Алгоритм машинного навчання – це комбінація процедур обробки даних, вбудованих у програмний код.

Факторами, які визначають тип алгоритмів ML, які будуть використовуватися, є бізнес-проблема, яку необхідно вирішити, доступні ресурси та специфіка даних (незалежно від того, мічені вони чи не мічені, і їх кількість).

Як правило, алгоритми поділяються на такі категорії залежно від проблеми, яку вони використовують для вирішення:

Класифікація: вони вибирають між кількома класами та надають для них імовірності: логістична регресія, наївний байєсівський метод, опорна векторна машина, K-найближчі сусіди тощо.

Регресія: вони дозволяють моделі передбачити кількість змінної: Naive Bayes, LARS Lasso, AdaBoost, XGBoost, Elastic Net, Random Forest тощо.

Кластеризація: вони дозволяють групувати подібні дані та маркувати їх відповідно до групи, до якої вони належать: K-Means, Kohonen, TwoStep тощо.

Зменшення розмірності: вони дозволяють поєднувати або відкидати незначущі дані: аналіз головних компонентів (PCA), лінійний дискримінантний аналіз (LDA), прямий вибір ознак, факторний аналіз тощо.

Існує дві групи алгоритмів ML залежно від типу даних: алгоритми ML, що використовуються з позначеними даними (регресія, алгоритми на основі екземплярів і дерева рішень), і ті, що використовуються з даними без міток (алгоритми кластеризації та асоціації).

Навчання алгоритму для отримання бажаних результатів. Навчання алгоритму є ітераційним, і кожен цикл включає:

- пряме поширення;
- порівняння результатів з бажаним результатом;
- коригування алгоритму для отримання найкращих результатів і повторення цих кроків, доки не буде отримано результат, необхідний для заданої норми ймовірності.

Отримане рішення є моделлю машинного навчання. Після підготовки моделі ML її можна використовувати з новими даними для вирішення нових бізнес-завдань і поступово вдосконалювати з точки зору ефективності та точності.

Машинне навчання актуальне в багатьох галузях і галузях і має здатність розвиватися з часом. Ось шість реальних прикладів використання машинного навчання.

Розпізнавання зображень є добре відомим і широко поширеним прикладом машинного навчання в реальному світі. Він може ідентифікувати

об'єкт як цифрове зображення на основі інтенсивності пікселів у чорно-білих або кольорових зображеннях[12].

Реальними прикладами є позначення рентгенівських знімків як ракових чи ні, призначення імені сфотографованому обличчю (він же «тегування» в соціальних мережах), розпізнання почерку, розділяючи одну літеру на менші зображення.

Машинне навчання також часто використовується для розпізнавання обличчя на зображенні. Використовуючи базу даних людей, система може визначити спільні риси та зіставити їх із обличчями. Це часто використовується в правоохоронних органах.

Також розпізнавання мовлення, машинне навчання може переводити мову в текст. Деякі програмні програми можуть перетворювати живий голос і записане мовлення в текстовий файл. Мова також може бути сегментована за інтенсивністю в частотно-часових діапазонах.

Реальні приклади розпізнавання мовлення: голосовий пошук, голосовий набір, контроль приладу.

Програмне забезпечення для розпізнавання мовлення найчастіше використовують на таких пристроях, як Google Home або Amazon Alexa.

Машинне навчання може допомогти в діагностиці захворювань. Багато лікарів використовують чат-боти з можливостями розпізнавання мовлення, щоб розрізнити закономірності в симптомах.

Реальні приклади медичної діагностики: постановка діагнозу або рекомендація варіанту лікування, онкологія та патологія використовують машинне навчання для розпізнавання ракової тканини, аналіз рідини організму.

У випадку рідкісних захворювань спільне використання програмного забезпечення для розпізнавання обличчя та машинного навчання допомагає сканувати фотографії пацієнтів і ідентифікувати фенотипи, які корелюють із рідкісними генетичними захворюваннями.

Статистичний арбітраж є ще одним прикладом використання. Арбітраж – це автоматизована торгова стратегія, яка використовується у фінансах для управління великою кількістю цінних паперів. Стратегія використовує торговий алгоритм для аналізу набору цінних паперів з використанням економічних змінних і кореляцій.

Реальні приклади статистичного арбітражу: алгоритмічний трейдинг, який аналізує мікроструктуру ринку, аналіз великих наборів даних, визначення можливостей арбітражу в реальному часі, оптимізація стратегії арбітражу для покращення результатів.

Прогностична аналітика – машинне навчання може класифікувати доступні дані на групи, які потім визначаються правилами, встановленими аналітиками. Коли класифікація завершена, аналітики можуть розрахувати ймовірність помилки.

Реальні приклади прогнозової аналітики: передбачення, чи є транзакція шахрайською чи законною, удосконалення системи прогнозування для розрахунку ймовірності несправності.

Прогностична аналітика є одним із найперспективніших прикладів машинного навчання, застосованого для всього; від розробки продукту до ціноутворення на нерухомість.

Екстракція – машинне навчання може витягувати структуровану інформацію з неструктурованих даних. Організації накопичують величезні обсяги даних від клієнтів. Алгоритм машинного навчання автоматизує процес анування наборів даних для інструментів прогнозової аналітики.

Реальні приклади видобутку: створення моделі для прогнозування розладів голосових зв'язок, розробка методів профілактики, діагностики та лікування захворювань, допомагайте лікарям швидко діагностувати та лікувати проблеми

Як правило, ці процеси виснажливі. Але машинне навчання може відстежувати та витягувати інформацію для отримання мільярдів зразків даних.

Машинне навчання пропонує широкий спектр методів і прийомів боротьби з серцево-судинними захворюваннями (ССЗ). Ці методи використовують потужність аналізу даних і прогнозного моделювання, щоб допомогти в ранньому виявленні, оцінці ризику, профілактиці та лікуванні ссз. Ось кілька ключових методів і технік, за допомогою яких машинне навчання сприяє цій важливій справі:

Вибір функцій і розробка: алгоритми машинного навчання можуть ідентифікувати найбільш релевантні характеристики або змінні в даних пацієнтів, які вказують на ризик серцево-судинних захворювань. Вибір функцій і розробка допомагають зменшити розмірність даних і підвищити ефективність прогнозних моделей.

Алгоритми класифікації: алгоритми класифікації машинного навчання, такі як логістична регресія, дерева рішень, випадкові ліси та машини опорних векторів, використовуються для класифікації людей у різні групи ризику. Ці моделі можуть передбачити, чи має пацієнт низький, помірний або високий ризик розвитку серцево-судинних захворювань.

Регресійні моделі: методи регресії, такі як лінійна регресія та гребенева регресія, використовуються для прогнозування постійних результатів, таких як артеріальний тиск або рівень холестерину. Ці моделі можуть допомогти оцінити серйозність факторів ризику ССЗ.

Глибоке навчання: глибокі нейронні мережі, підмножина машинного навчання, використовуються для таких завдань, як аналіз зображень у медичній візуалізації, що дозволяє ідентифікувати структурні аномалії та точно оцінювати ризик серцево-судинних захворювань.

Методи ансамблю: такі комплексні методи, як пакетування та посилення, поєднують кілька моделей машинного навчання для підвищення точності прогнозування та зменшення надмірного оснащення. Вони використовуються для створення надійних моделей прогнозування ризику ссз.

Обробка природної мови (nlp): використовується для отримання цінної інформації з неструктурованих клінічних нотаток, звітів і розповідей пацієнтів

в електронних медичних записах. Це допомагає отримувати анамнез пацієнта та симптоми, пов'язані з ССЗ.

Алгоритми кластеризації: алгоритми кластеризації, такі як k-середні та ієрархічна кластеризація, застосовуються для сегментації пацієнтів у групи на основі подібних характеристик, допомагаючи у стратифікації пацієнтів та індивідуальних втручаннях.

Виявлення аномалії: алгоритми виявлення аномалій виявляють незвичайні моделі або викиди в даних пацієнтів. Ці методи є вирішальними для раннього виявлення порушень життєво важливих ознак або рівня біомаркерів[15].

Аналіз часових рядів: моделі часових рядів використовуються для аналізу тенденцій і змін у даних пацієнтів з часом, що особливо важливо для моніторингу та лікування хронічних ССЗ.

Незбалансована обробка даних: такі методи, як надмірна вибірка, недостатня вибірка та генерація синтетичних даних, використовуються для усунення незбалансованих наборів даних, гарантуючи, що моделі машинного навчання надають точні прогнози як для осіб із високим, так і з низьким ризиком.

Зменшення розмірності: аналіз головних компонентів і t-розподілене стохастичне вбудовування сусідів (t-sne) використовуються для зменшення розмірності даних пацієнта великої розмірності, зберігаючи важливу інформацію.

Зрозумілість моделі: такі методи, як shap (додаткові пояснення shapley) і lime (локальні інтерпретовані модельно-незалежні пояснення) допомагають пояснити рішення, прийняті моделями машинного навчання, підвищуючи їх прозорість і надійність у клінічних умовах.

Передача навчання: передача навчання використовує попередньо підготовлені моделі на великих наборах даних і точно налаштовує їх на конкретні завдання прогнозування серцево-судинних захворювань, зменшуючи потребу у великих позначених даних.

Системи постійного моніторингу та оповіщення: моделі машинного навчання можна інтегрувати в системи безперервного моніторингу, щоб виявляти раптові зміни в даних пацієнтів і ініціювати сповіщення для своєчасного втручання.

Геномний аналіз: машинне навчання використовується для аналізу геномних даних для визначення генетичних маркерів, пов'язаних із ризиком серцево-судинних захворювань, пропонуючи персоналізовану оцінку ризику на основі генетичної інформації.

Прогностичне моделювання прихильності до лікування: моделі машинного навчання передбачають прихильність пацієнтів до призначених ліків, забезпечуючи ефективне дотримання пацієнтами своїх планів лікування.

Обробка даних у реальному часі: машинне навчання дозволяє обробляти дані з переносних пристроїв у режимі реального часу, надаючи негайну інформацію та втручання пацієнтам і постачальникам медичних послуг.

Міждисциплінарна співпраця: машинне навчання сприяє співпраці між медичними працівниками, дослідниками даних і дослідниками, заохочуючи міждисциплінарний підхід до вирішення складних проблем серцево-судинних захворювань.

Ці методи та техніки демонструють універсальність машинного навчання у лікуванні серцево-судинних захворювань. Використовуючи інформацію, що базується на даних, і прогнозне моделювання, машинне навчання сприяє ранньому виявленню, оцінці ризиків, профілактичним стратегіям і індивідуальним втручанням, що зрештою покращує лікування серцево-судинних захворювань і зменшує їх вплив на глобальну громадськість.

3 РОЗРОБКА СТРУКТУРИ І АЛГОРИТМІВ ПРОГРАМИ

3.1 Модель

Прогнозування відіграє вирішальну роль у клінічній практиці. Зокрема, точність відіграє життєво важливу роль у прогнозуванні захворювань серця. Останні внески кількісно оцінюють можливості алгоритмів машинного навчання для досягнення того ж. Алгоритми та методи машинного навчання застосовувалися до різних наборів медичних даних для автоматизації аналізу великих і складних даних. У цьому контексті експериментальне дослідження вказує на запропоновану модель, що визначає підхід машинного навчання, наприклад класифікацію, для виявлення наявності серцевих захворювань.

Хвороба серця вважається однією зі складних і найсмертоносніших хвороб людини у світі. Діагностика та лікування захворювань серця є дуже складними, особливо в країнах, що розвиваються, через рідкісну доступність діагностичного обладнання та брак лікарів та інших ресурсів, що впливає на правильне прогнозування та лікування серцевих пацієнтів. Точна та правильна діагностика ризику серцево-судинних захворювань у пацієнтів необхідна для зниження пов'язаних із ними ризиків серйозних проблем із серцем та покращення безпеки серця. Інвазивні методи діагностики серцевих захворювань базуються на аналізі історії хвороби пацієнта, звіті про фізичне обстеження та аналізі відповідних симптомів медичними експертами. Усі ці методи здебільшого спричиняють неточну діагностику та часто затримують результати діагностики через людські помилки.

Для вирішення цих складнощів у діагностиці захворювань серця розроблені прогностичні моделі з використанням штучної нейронної мережі та машинного навчання. Ці системи навчаються та випробовуються, щоб отримати кращу точність прогнозування захворювань [12].

Сфера машинного навчання дозволяє ідентифікувати приховані шаблони та створювати аналітичні структури, включаючи кластеризацію, класифікацію, регресію та кореляцію, шляхом інтеграції та застосування різних методів, таких як моделі машинного навчання, нейронні мережі та пошук інформації. Отже, методи машинного навчання продемонстрували великий потенціал для підтримки прийняття клінічних рішень, допомоги в розробці клінічних настанов і алгоритмів управління, а також сприяння створенню заснованих на доказах клінічних практик для лікування серцево-судинних захворювань (ССЗ). Крім того, раннє виявлення серцево-судинних захворювань за допомогою методів машинного навчання може зменшити потребу в обширних і дорогих клінічних і лабораторних дослідженнях, що призведе до зменшення фінансового навантаження як на систему охорони здоров'я, так і на окремих осіб.

Задля яскравого прикладу можна навести модель підсилення градієнта для прогнозування наявності серцево-судинних захворювань і визначення найбільш прогностичного значення на основі значень грубих наборів. Після цього для аналізу серцево-судинних захворювань використовується низка методів машинного та глибокого навчання. Нижче наведено основні внески цього дослідження.

Використовуючи перехресну перевірку та розділену перевірку, відкрийте для себе алгоритм машинного навчання з покращеною продуктивністю, який застосовуватиметься для виявлення серцево-судинних захворювань.

Застосування відповідної техніки вибору ознак може оптимізувати точність передбачення. Використання надійного алгоритму машинного навчання може покращити раннє прогнозування розвитку серцево-судинних захворювань (ССЗ) на ранніх стадіях, сприяючи ранньому втручання та вибору ключових функцій для підтримки алгоритмів відновлення.

Прогнозування серцево-судинних захворювань за допомогою найсучаснішого набору даних про серцево-судинні захворювання.

Надання достовірних консультацій медичним та медичним фахівцям щодо суттєвих змін у сфері охорони здоров'я.

Багато дослідників вивчають низку рамок очікування серцевих захворювань, використовуючи різні методи аналізу даних. Вони використовують набори даних і різноманітні обчислення на додаток до результатів тестування та майбутньої роботи, яка буде можлива над структурою, і досягають більш продуктивних результатів. Дослідники завершили численні дослідницькі спроби досягти ефективних методів і високої точності в розпізнаванні розладів, пов'язаних з серцем [3].

Дослідження Tran побудувало інтелектуальну систему з використанням методу моделювання Naive Bayes data mining. Це веб-додаток, у якому користувач відповідає на заздалегідь запрограмовані запитання. Він намагається знайти приховану інформацію в базі даних і порівнює значення користувача з навченим набором даних. Він може надати відповіді на складні запитання щодо діагностики серцевих захворювань, дозволяючи медичним працівникам приймати більш обґрунтовані клінічні рішення, ніж традиційні системи підтримки прийняття рішень. Це також знижує витрати на лікування, надаючи ефективний догляд.

Gnaneswar демонструє важливість моніторингу частоти серцевих скорочень під час їзди на велосипеді. Велосипедисти можуть впоратися з велосипедними зустрічами, такими як велосипедний ритм, щоб визначити рівень активності, відстежуючи свій пульс під час прискорення. Контролюючи навантаження на педалі, велосипедисти можуть уникнути перетренованості та серцевої недостатності. Пульс велосипедиста можна використовувати для визначення інтенсивності вправи. Пульс можна виміряти за допомогою датчика, який можна носити. На жаль, датчик не фіксує всю інформацію через рівні проміжки часу, наприклад одну секунду, дві секунди тощо. Отже, нам знадобиться модель очікування пульсу, щоб заповнити прогалини.

Робота Gnaneswar спрямована на використання Feedforward Brain Organization для побудови прогностичної моделі для пульсу з урахуванням

циклічного ритму. На другому джерелами даних є пульс і ритм. Результатом є прогнозований пульс на наступну секунду. Використовуючи структуру мозку з прямою передачею, зв'язок між пульсом і велосипедним ритмом представляється статистично. Мутіярса розширює адміністрування медичної допомоги на основі цих аргументів. Численні прориви в дистанційному зв'язку були зроблені в очікуванні серцевих захворювань. Використання методів інтелектуального аналізу даних (DM) для виявлення та локалізації коронарної хвороби є дуже корисним. Під час їхньої оцінки проводиться порівняльний аналіз кількох розрахунків видобутку інформації для однієї та змішаної породи, щоб визначити, яке обчислення найбільш точно прогнозує коронарну хворобу [15].

Єшвендра стверджує, що використання обчислень ШІ для прогнозування різних захворювань зростає. Це поняття настільки важливе та різноманітне через здатність обчислень штучного інтелекту мати порівнянну перспективу з людиною для підвищення точності прогнозу коронарної хвороби. Patil зазначає, що правильна діагностика серцевих захворювань є однією з найважливіших біомедичних проблем, які необхідно вирішити. Три методи видобутку інформації: опорна векторна машина, наївна байєва та дерево рішень. Ці методи були використані для створення мережі емоційної підтримки для їхнього бажаного варіанту. Тріполіті стверджує, що ідентифікація захворювань із високим рівнем поширеності, таких як хвороба Альцгеймера, Паркінсона, діабет, рак молочної залози та коронарна хвороба, є одним із найважливіших біомедичних тестів, які потребують негайної уваги. Gonsalves спробував спрогнозувати коронарну ССЗ за допомогою машинного навчання та історичних медичних даних. Oikonomou надає огляд різноманітної інформації, яка зустрічається в умовах хронічних захворювань. Використовуючи численні методи машинного навчання, вони роз'яснили теорію екстремальних цінностей, щоб краще вимірювати тяжкість хронічних захворювань і ризику.

Згідно з Ібрагімом, системи, засновані на машинному навчанні, можна використовувати для прогнозування та діагностики захворювань серця. Методи активного навчання (AL) підвищують точність класифікації шляхом інтеграції зворотного зв'язку між користувачем і експертною системою з розрідженими даними. Крім того, Pratiyush досліджували роль ансамблевих класифікаторів у системі ХАІ у прогнозуванні захворювань серця на основі наборів даних про ССЗ. У запропонованій роботі використовувався набір даних, що включає 303 екземпляри та 14 атрибутів, з характеристиками атрибутів категорійного, цілого та дійсного типу, а завдання класифікації базувалося на таких методах класифікації, як KNN, SVM, наивний метод Байєса, AdaBoost, пакетування та LR [9].

У літературі зроблені спроби створити стратегії для прогнозування діагностики серцевих захворювань. Через високу розмірність текстового введення багато традиційних алгоритмів машинного навчання не можуть включити його в процес прогнозування одночасно. Як наслідок, у цій статті досліджується та розробляється набір надійних алгоритмів машинного навчання для покращення раннього прогнозування розвитку серцево-судинних захворювань, що дозволяє швидко втручатися та одужати.

3.2. Матеріали та методи

Дані завжди відіграють експериментах дослідження роль у процесі прогнозування. Для розробки цієї прогностичної моделі були враховані дані про захворювання серця, доступні в репозиторії машинного навчання UCI. Дані складаються з наведених нижче атрибутів.

S No	Attribute Name	Type of Attribute
1	Age	Real
2	Sex	Binary
3	Chest Pain Type	Nominal
4	Resting Blood Pressure	Real
5	Cholesterol	Real
6	Fasting Blood Sugar	Binary
7	Resting Electrocardiographic Results	Nominal
8	Maximum Heart Rate Achieved	Real
9	Exercise Induced Angina	Binary
10	Old peak	Real
11	Slope Of The Peak Exercise ST Segment	Ordered
12	Number Of Major Vessels	Real
13	Thal	Nominal
14	Num	Real

Рисунок 3.1 –Кроки, необхідні для створення системи прогнозування захворювань серця

Розподіл даних можна побачити на рисунках нижче.

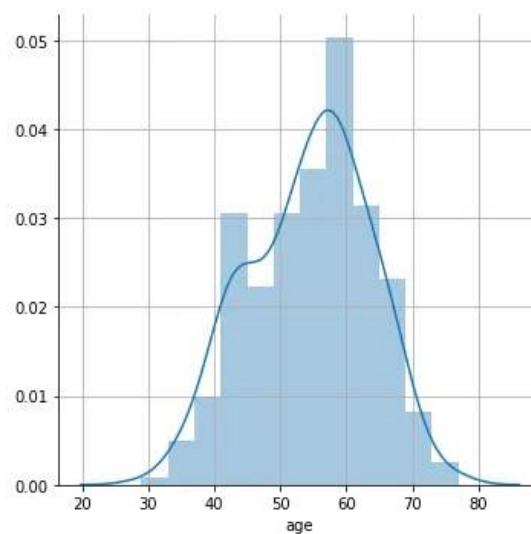


Рисунок 3.2 – Розподіл за віком

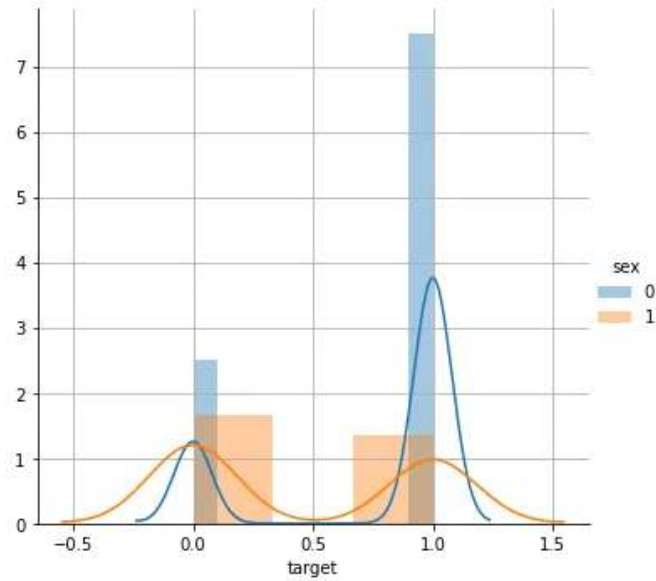


Рисунок 3.3 – Розподіл, який показує, що і чоловіки, і жінки однаково схильні до впливу

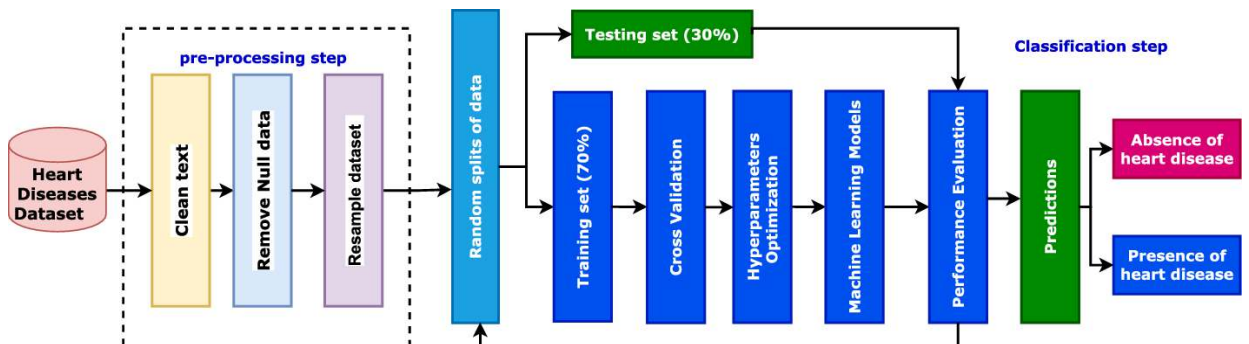


Рисунок 3.4 – Основні етапи запропонованої методики

У запропонованій схемі класифікації випадків захворювань серця, спочатку проводиться пошуковий аналіз. Проводиться комплексний аналіз як цілі, так і функцій, а змінні категорії перетворюються на числові значення. Для порівняння моделей, що розглядаються, використовуються різні критерії. Результати кожної моделі аналізуються, і вибирається оптимальна модель для розглянутої проблеми. Запропонована модель ретельно досліджується, а бібліотека Optuna використовується для налаштування гіперпараметрів моделі, щоб побачити, наскільки вони покращені. Запропонована модель

розділена на три фази: попередня обробка, навчання та класифікація (рисунок 3.4).

Перед навчанням вибраних моделей важливо звернути увагу на відсутні значення холестерину, які спочатку були введені як 0. Для цього дані розділені на групи на основі наявності підтвердженого захворювання серця, і використовується середнє значення кожної групи щоб заповнити відсутні значення. Щоб оцінити, чи впливають ці змінні на прогнозування серцевих захворювань на основі їхніх значень Шеплі, до моделей було включено умови взаємодії, щоб зафіксувати будь-які можливі кореляції між елементами даних. SHAP (SHapley Additive ExPlanations) використовує теорію ігор для визначення значущості кожної характеристики та може використовуватися для пояснення як окремих передбачень моделі, так і сукупних результатів моделі. SHAP визначає величину внеску кожного предиктора в результат моделі шляхом усереднення граничних внесків кожної функції за всіма можливими комбінаціями ознак.

Алгоритм машинного навчання буде правильно навчено після попередньої обробки та нормалізації наборів даних. Після модифікації даних їх довільно класифікують на навчальний набір і тестовий набір, при цьому 70% рядків призначають навчальному набору, а 30% – тестовому. К-кратність – це поширений метод перехресної перевірки, який передбачає виконання великої кількості відповідних тестів для визначення типового показника точності моделі. Ця методика існує досить давно. Для перевірки запропонованої стратегії розгортаються такі процедури ШІ, як SVC, MultinomialNB, K-Neighbor, BernoulliNB, SGD, Random forest і Decision Tree для найкращих умов результату.

XGBoost (Extreme Gradient Boosting) – це контрольований метод навчання для підвищення точності прогнозування шляхом поєднання кількох дерев рішень. XGBoost ітеративно додає дерева рішень за допомогою посилення градієнта, при цьому кожне наступне дерево намагається

виправити помилки попередніх дерев. Остаточний прогноз є зваженою сумою всіх окремих прогнозів дерева.

Наївний алгоритм Байєса призначає рівну вагу всім характеристикам або якостям. Алгоритм стає більш ефективним, оскільки одна властивість не впливає на іншу. Відповідно до Ясіна 2020, наївний класифікатор Байєса (NBC) є простим, ефективним і добре відомим алгоритмом категоризації тексту. NBC використовує теорему Байєса для класифікації документів з 1950-х років, і вона теоретично обґрунтована. Апостеріорна оцінка використовується для визначення класу за допомогою наївного класифікатора Байєса. Характеристики, наприклад, класифікуються на основі їх найвищого умовного потенціалу [15].

Наївний Байєс Бернуллі – це статистичний метод, який дає логічні результати на основі наявності або відсутності необхідного тексту. Дискретний розподіл Бернуллі вводиться в цей класифікатор. Під час ідентифікації небажаного ключового слова або позначення певного типу слова в тексті цей тип простого байєсівського класифікатора є корисним. Він також відрізняється від багаточленного підходу тим, що генерує двійковий вихід, такий як 1–0, True-False або Yes–No. Стохастична система або процедура – це така, яка має як частину рішення випадкової відповідності. Стохастичний градієнтний спуск (SGD) рандомізує кілька зразків даних, а не весь набір даних у кожній ітерації. Як наслідок, замість обчислення суми градієнтів для всіх екземплярів кожна ітерація обчислює градієнт функції вартості для окремого прикладу. SGD – це метод визначення оптимальних властивостей гладкості диференційованої або субдиференційованої цільової функції, який є ітераційним.

Дерево рішень – це широко відомий метод машинного навчання, у якому дані неодноразово розподіляються на основі певних параметрів. Дерево має дві обхідні сутності: вузли та листя. Листя представляють рішення або результати, тоді як вузли рішень поділяють дані. Дерева рішень можна використовувати в комбінації для розв’язування проблем (ансамблеве

навчання). Алгоритм випадкового лісу вирішує проблеми переобладнання, пов'язані з алгоритмами дерева рішень. Алгоритм здатний вирішувати проблеми регресії та класифікації, а також оцінювати велику кількість атрибутів, щоб визначити, які з них є найважливішими. Випадкові дані можуть вивчатися без добре спланованих змін даних.

Алгоритм K-Nearest Neighbor (K-NN) класифікує нові спостереження на основі їх відстані від відомих прикладів. Базуючись на більшості голосів його сусідів і функції відстані як інструменті вимірювання, випадок позначається в класі з найвищою частотою серед його k-найближчих сусідів. У задачах класифікації k-NN повертає приналежність до класу. Тоді як у задачах регресії він повертає значення властивості об'єкта. Чи використовується k-NN для класифікації чи регресії, впливає на результат. Оскільки цей метод покладається на відстань для класифікації, нормалізація може значно покращити навчальні дані. Якщо ознаки відповідають різним фізичним одиницям або масштабам, стандартизація може значно підвищити точність навчальних даних.

В експериментах дослідження було використувано Google Colab як платформу впровадження для моделей машинного навчання. Платформа включає в себе віртуальну машину, яка працює на серверах Google і надає користувачам доступ до середовища Python, яке включає популярні наукові бібліотеки даних, такі як TensorFlow, PyTorch і Scikit-Learn. Google Colab – це хмарне середовище для ноутбуків Jupyter, яке пропонує безкоштовний доступ до обчислювальних ресурсів, таких як віртуальна машина з 12 ГБ оперативної пам'яті та до 100 ГБ місця на жорсткому диску. Розмір пам'яті, виділеної віртуальній машині, становить до 25 ГБ, також можна ввімкнути опції з великим обсягом оперативної пам'яті до 52 ГБ для великих моделей або даних [11]. Віртуальна машина працює на серверах Google і оснащена графічним процесором NVIDIA Tesla K80, що дозволяє нам ефективно навчати моделі глибокого навчання. Крім того, Google Colab надає широкий спектр попередньо встановлених бібліотек і інструментів, що полегшує встановлення

та використання необхідних залежностей. Віртуальна машина працює на базі операційної системи Linux, що забезпечує стабільність і надійність середовища реалізації. Крім того, операційною системою, яка використовується віртуальною машиною, є Linux Ubuntu, яка поставляється з попередньо встановленими різними системними бібліотеками та інструментами.

Дані про стан серця, використані в цьому дослідженні, є синтезом наборів даних зі сховища машинного навчання UCI та містять одинадцять характеристик, які можна використовувати для прогнозування наявності серцевої недостатності, поширеного серцево-судинного захворювання, яке значно підвищує ймовірність серцево-судинних захворювань. Цільова змінна є двійковим атрибутом, який вказує на діагноз серцевої недостатності, якщо HeartDisease = 1 (рисунок 3.5). Крім того, також представлено список змінних і опис характеристик у наборі даних про захворювання серця (рисунок 3.6).

Age	Sex	Type chest pain	BP resting	Cholesterol	BS fasting	ECG resting	HR max	Angina exercise	Old peak	ST slope	Disease of heart
41	M	ATA	142	287	0	Nor1	173	N	0.0	Upper	0
48	F	NAP	162	182	0	Nor1	157	N	1.0	Flat1	1
38	M	ATA	132	273	0	ST	98	N	0.0	Upper	0
49	F	ASY	136	224	0	Nor1	109	Y	1.5	Flat1	1
53	M	NAP	152	185	0	Nor1	123	N	0.0	Upper	0

Рисунок 3.5 – Зразок набору даних про серцеву недостатність

Variable	Interpretation
Age	Patient's Age/year
Gender	Patient's Gender, Male/Female
Type of chest pain	Type of chest pain: i. TA: Typical Angina ii. ATA: Atypical Angina iii. NAP: Non-Anginal Pain iv. ASY: Asymptomatic
Resting blood pressure	Patient's Blood Pressure/mmHg.
Total Cholesterol	Patient's Cholesterol (mg/dl).
Blood Glucose level (Fasting)	Patient's fasting blood glucose level. i. glucose > 120 mg/dL = 1 ii. glucose below 120 mg/dL = 0
ECG at rest	Electrocardiography (at rest): i. Normal ii. ST: ST segment and/or T wave abnormality iii. LVH: Probable or Definite Left Ventricular Hypertrophy
Heart Rate at Maximum	Maximum Heart Rate, heart beats per minute.
Angina on Exercising	Exercise-associated Angina, present /absent.
Old peak	Measure of ST Depression.
ST_Slope	Slope of Peak Exercise. i. Up: up sloping ii. Flat iii. Down: down sloping

Рисунок 3.6 – Симптоми, ознаки та лабораторні дослідження сукупності захворювань серця

Datasets	#Observations
Cleveland	303
Hungarian	294
Stalog (Heart) Data Set	270
Long Beach VA	200
Switzerland	123
Total	1190
Duplicated	272
Final dataset	918

Рисунок 3.7 – Різні набори даних, які використовуються для створення набору даних про захворювання серця

	Age	RestingBP	Cholesterol	FastingBS	MaxHR	Oldpeak	HeartDisease
Count	918	918	918	918	918	918	918
Max	77	200	603	1	202	6.20	1
Min	28	0	0	0	60	-2.6	0
Mean	53.51	132.39	198.79	0.23	136.81	0.89	0.55
Std	9.43	18.51	109.38	0.42	25.46	1.06	0.49
25%	47	120	173.25	0	120	0	0
50%	54	130	223	0	138	0.60	1
75%	60	140	267	0	156	1.50	1

Рисунок 3.8 – Зведена статистика числових змінних

	Sex	TypeChestPain	ECGResting	AnginaExercise	ST_Slope
Count	920	920	920	920	920
Unique	2	3	4	2	4
Top	M	ASY	Normals	N	Flat1
Freq	735	486	562	557	470

Рисунок 3.9 – Зведена статистика категоріальних змінних

Variable	Value	Total patients	Proportion of heart disease
Sex	M	725	90.2%
	F	193	9.8%
ChestPainType	ASY	496	77.2%
	NAP	203	14.2%
	ATA	173	4.7%
	TA	46	3.9%
RestingECG	Normal	552	56.1%
	ST	178	23.0%
	LVH	188	20.9%
ExerciseAngina	Y	371	62.2%
	N	547	37.8%
ST_Slope	Flat	460	75.0%
	Up	395	15.4%
	Down	63	9.6%

Рисунок 3.10 – Частка захворювань серця

Набір даних було створено шляхом об'єднання різноманітних наборів даних, які раніше були доступні незалежно та не об'єднувалися раніше. У цьому наборі даних п'ять наборів даних про серце об'єднані в 11 загальних характеристиках, що робить його найбільшим набором даних про серцеві захворювання, доступним для дослідницьких цілей. Конкретні набори даних, використані для курації цього складеного набору даних (рисунок 3.7).

Набір даних Heart Disease містить 918 спостережень і 12 стовпців. Можна побачити підсумовані основні статистичні дані для числових характеристик (рисунок 3.8). Зрозуміло, що середнє значення віку становить 53 роки, а максимальне – 77 років (рисунок 3.8). Подібним чином представлена статистика категоріальних атрибутів (рисунок 3.9), де зображені унікальні значення атрибута ChestPainType дорівнюють 4, а верхнє – «ASY».

Також можна побачити підсумовані основні деталі числових характеристик (рисунок 3.10). Зрозуміло, що змінна Стать має два основні значення: чоловіки (М) і жінки (Ж), так що частка захворювань серця для М становить 90,2%, а для Ж – 9,8%. Подібним чином представлено статистику атрибута ChestPainType (рисунок 3.10), є 4 значення (ASY, NAP, ATA та TA), і найчастішим є ASY 77,2%.

3.3 Результати роботи

Примітно, що класифікації в значенні атрибута хвороби серця досить добре збалансовані. У 508 із 918 пацієнтів, які брали участь у дослідженні, була діагностована серцева недостатність, а у 410 – ні. Пацієнти із захворюваннями серця мають середній вік 57 років, тоді як пацієнти без захворювань серця мають типовий вік 51 рік. Як показано (рисунок 3.11), близько 63% чоловіків мають серцеві захворювання, тоді як приблизно 25% жінок мають серцеві захворювання [7]. Жінка має ймовірність захворювання серця 25,91%. Чоловік має ймовірність захворювання серця 63,17%.

Heart Disease Statistics

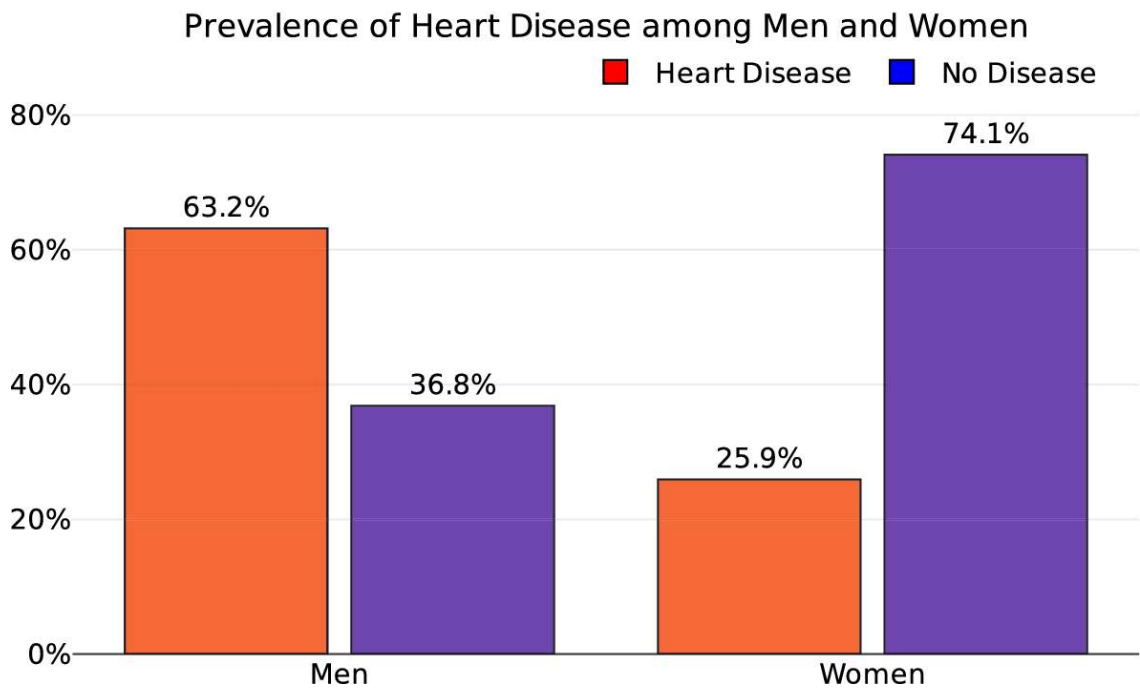
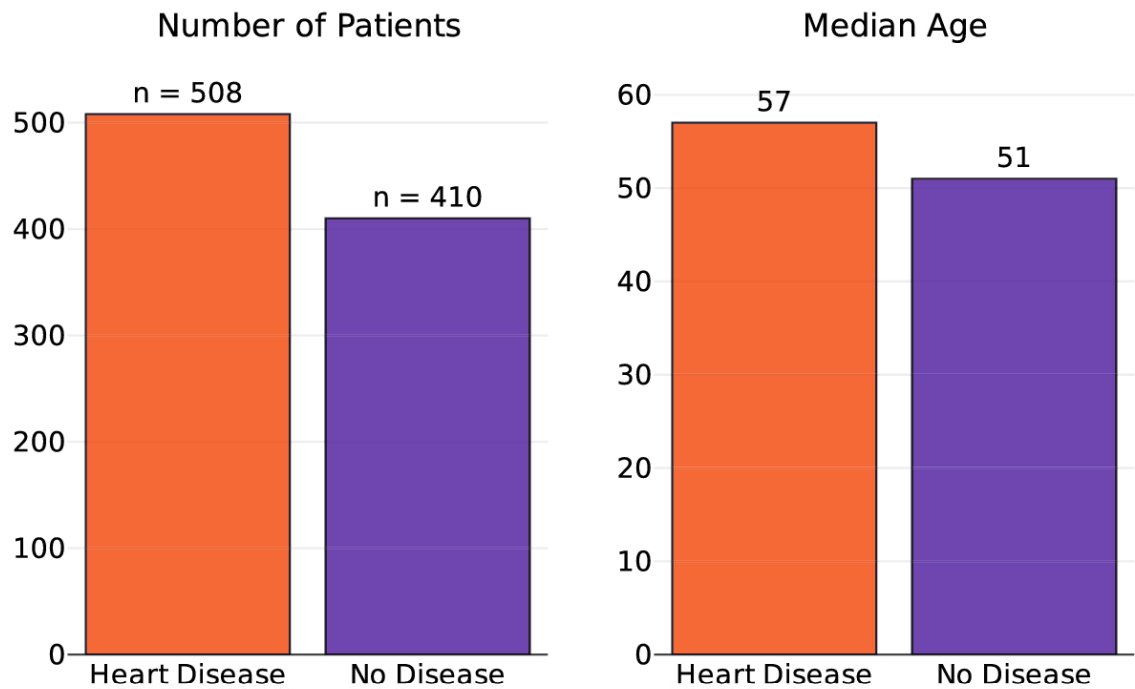


Рисунок 3.11 – Статистика серцевих захворювань

Надалі буде продемонстровано діапазони захворювань серця для віку, систолічного артеріального тиску, холестерину, частоти серцевих скорочень і депресії сегмента ST (рисунок 3.12). Вікова діаграма пацієнтів із серцево-

судинними захворюваннями припадає на вік від 51 до 62 років, як це зображено на віковій діаграмі. У цій категорії також є кілька молодших людей, що виходять за нижню межу. Люди, які не мають серцево-судинних захворювань, мають віковий діапазон, який є дещо більш варіабельним, але більш рівномірним, і немає викидів. Переважна більшість пацієнтів, які належать до цієї категорії, досить молоді, віком від 43 до 57 років.

Крім того, коробкові діаграми між групами для змінної Pulse Pressure Pressure надзвичайно схожі. Обидва мають верхні та нижні викиди, при цьому у переважної більшості пацієнтів артеріальний тиск падає від 120 до 145 мм рт. Як показано (рисунок 3.12), середній артеріальний тиск в обох групах становить приблизно 130 мм рт. Крім того, для змінної холестерину розподіл холестерину, здається, зміщений вправо, особливо серед осіб із серцевими захворюваннями, де значна кількість спостережень була зареєстрована зі значеннями холестерину 0. Як показано (рисунок 3.12), середня частота серцевих скорочень у людей із захворюваннями становить 150 ударів на хвилину, а у людей із захворюваннями серця – 126 ударів на хвилину.

У випадку змінної депресії сегмента ST (OldPeak) існує дисперсія між розподілом груп депресії сегмента ST. Депресія ST більш варіабельна у пацієнтів із захворюваннями серця з численними більшими викидами. У більшості цих пацієнтів спостерігаються депресії ST від 0 до 2 мм із середнім значенням 1,2 мм. У пацієнтів без захворювань серця діапазон вузкий, між 0 і 0,6 мм, з медіаною депресії ST 0 мм, однак розподіл цієї групи в цілому більш спотворений (рисунок 3.12).

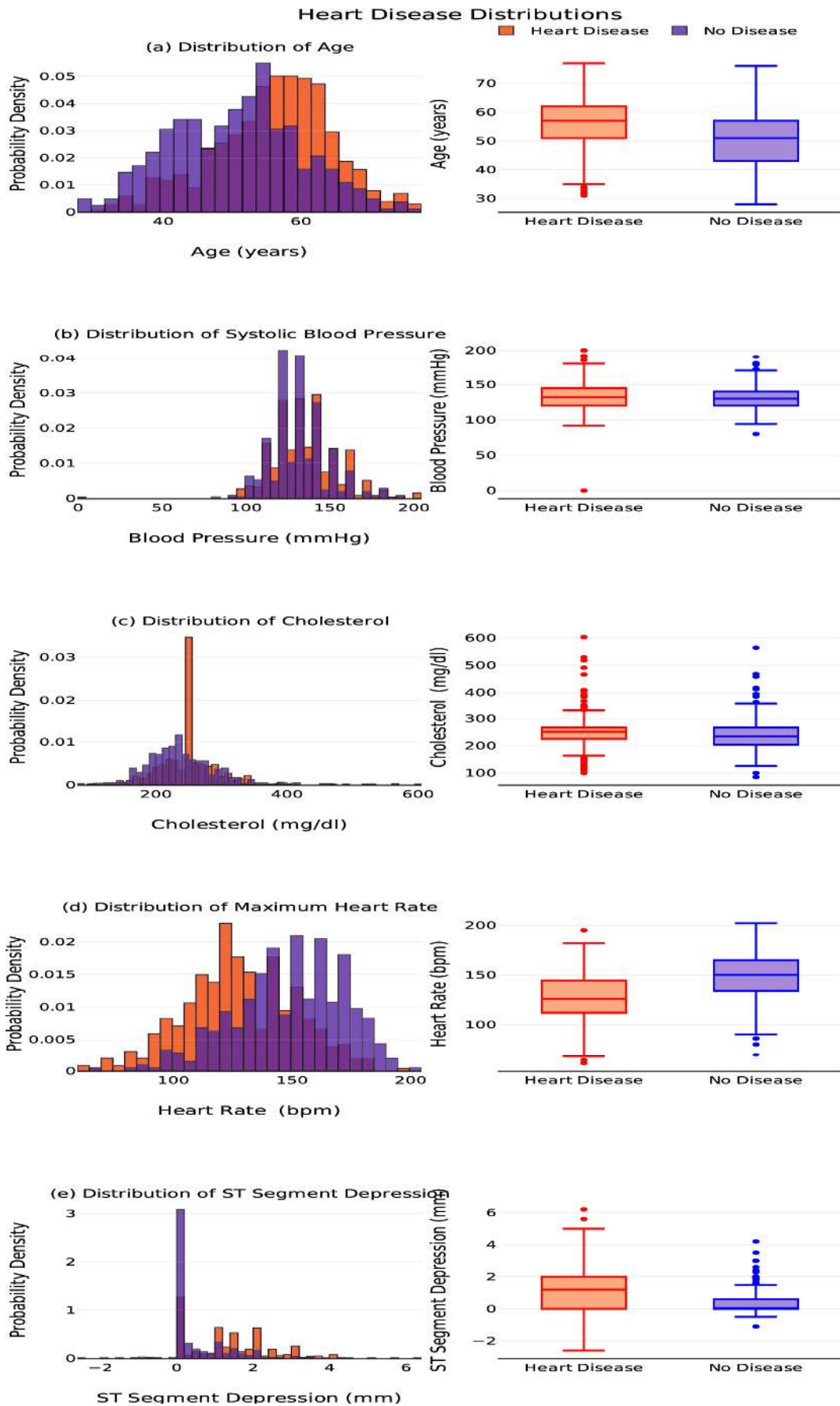


Рисунок 3.12 – Поширення захворювань серця

Розглянемо кореляційну матрицю (рисунок 3.13), пов'язану з набором даних про захворювання серця. Серцева хвороба має найсильніший позитивний зв'язок із OldPeak (кореляція = 0,4) і найсильніший негативний зв'язок із MaxHR (кореляція = -0,4), згідно з кореляційною матрицею. Вік і MaxHR також мають досить високий зв'язок із кореляцією 0,38. Як видно (рисунок 3.13), частота серцевих скорочень має тенденцію до зниження з віком. Результати спостерігають слабку кореляцію між числовими ознаками та цільовою змінною на основі матриці. Олдпек (число, пов'язане з депресією) позитивно корелює із захворюваннями серця. Хвороба серця негативно корелює з максимальною частотою серцевих скорочень. Холестерин має цікавий негативний зв'язок із захворюваннями серця.

	Age	RestingBP	Cholesterol	FastingBS	MaxHR	Oldpeak	HeartDisease
HeartDisease	0.28	0.11	-0.23	0.27	-0.4	0.4	1.0
Oldpeak	0.26	0.16	0.05	0.05	-0.16	1.0	0.4
MaxHR	-0.38	-0.11	0.24	-0.13	1.0	-0.16	-0.4
FastingBS	0.2	0.07	-0.26	1.0	-0.13	0.05	0.27
Cholesterol	-0.1	0.1	1.0	-0.26	0.24	0.05	-0.23
RestingBP	0.25	1.0	0.1	0.07	-0.11	0.16	0.11
Age	1.0	0.25	-0.1	0.2	-0.38	0.26	0.28

Рисунок 3.13 – Кореляційна матриця для набору даних Heart Disease

Нижче проілюстровано кореляцію між хворобою серця та змінними категорії (рисунок 3.14). Близько 80% діабетиків мають проблеми з серцем. Пацієнти зі стенокардією фізичного навантаження мають навіть більшу

частоту серцево-судинних захворювань, понад 85%. Понад 65% пацієнтів із діагностованою хворобою серця мали аномалії зубця ST-T на ЕКГ у стані спокою, що є найбільшим відсотком серед категорій. Пацієнти з плоским або спадаючим нахилом ST під час фізичних навантажень мають найвищу частоту серцево-судинних захворювань – 82,8% і 77,8% відповідно.

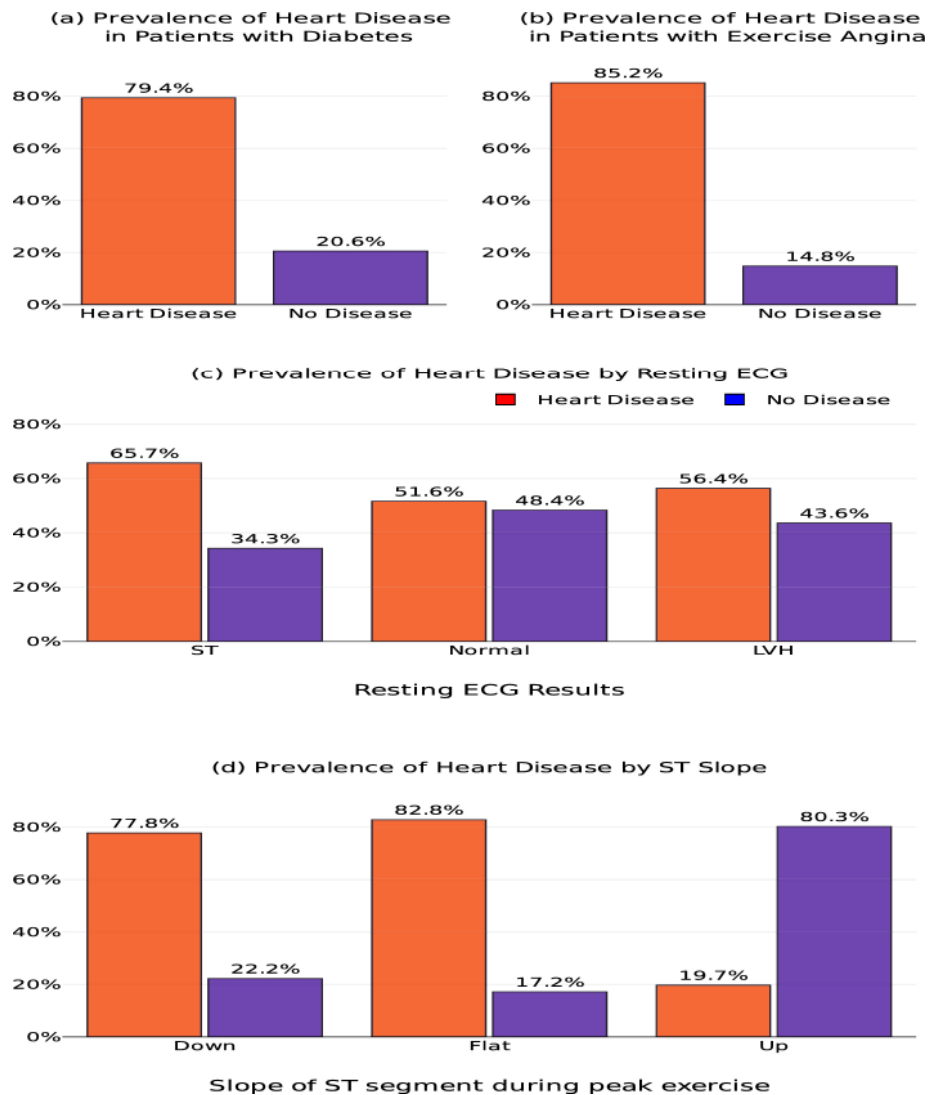


Рисунок 3.14 – Поширеність захворювань серця за ЕКГ результатами

Безсимптомний біль у грудях при серцевих захворюваннях у майже 77%, відсутність болю в грудях (безсимптомний) є найпоширенішим симптомом у пацієнтів із серцевими захворюваннями (рисунок 3.15). Крім того, захворювання серця приблизно в дев'ять разів частіше зустрічаються у

чоловіків, ніж у жінок, серед пацієнтів із серцево-судинним діагнозом. Пацієнт із безсимптомним болем у грудях (ASY) приблизно в шість разів частіше страждає серцевими захворюваннями, ніж пацієнт із атиповою стенокардією (АТА).

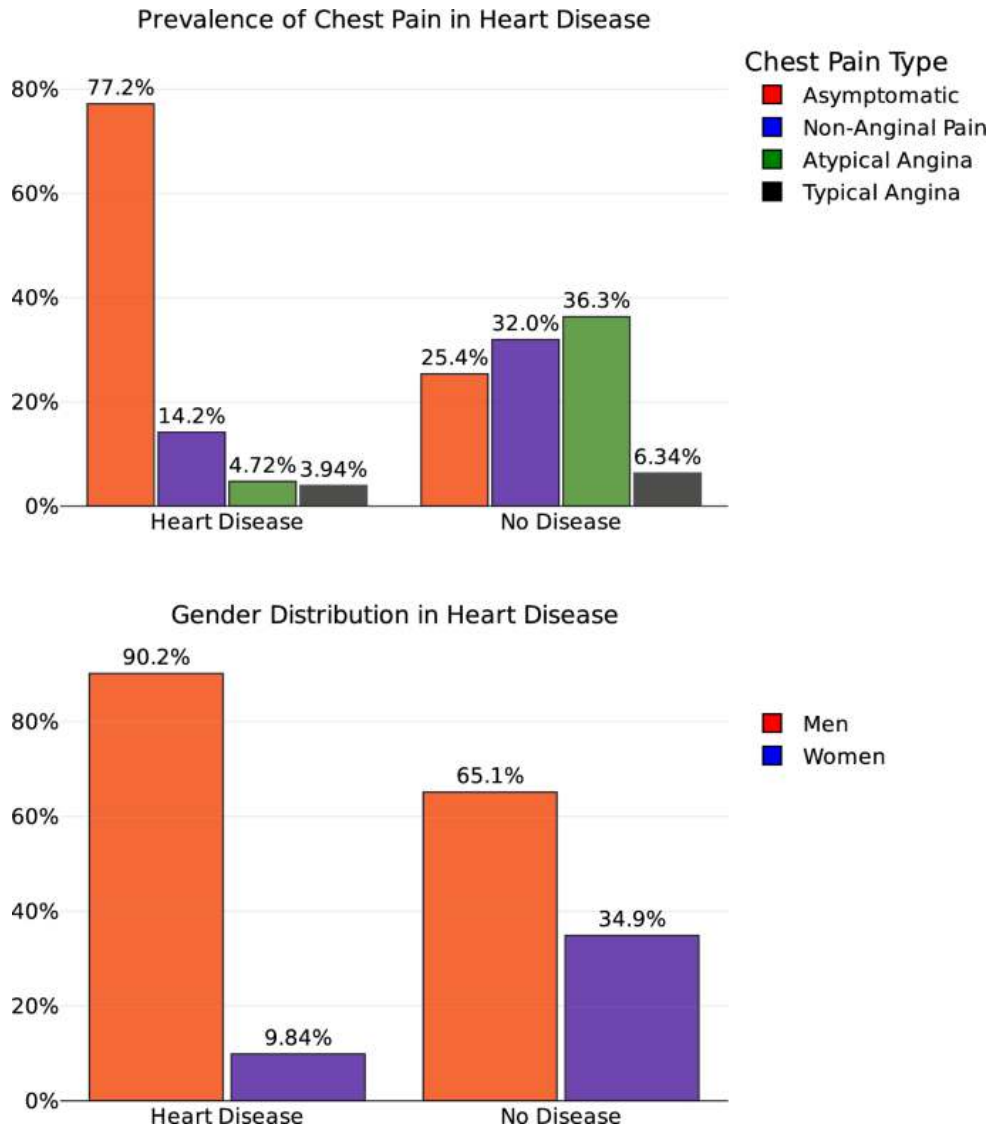


Рисунок 3.15 – Поширеність болю в грудній клітині серед даних про захворювання серця

Загальна інформація, отримана в результаті дослідницького аналізу даних, дані для цільової змінної майже збалансовані. Зв'язок між числовими характеристиками та цільовою змінною слабкий. Олдпик (число, пов'язане з

депресією) позитивно корелює із захворюваннями серця. Серцева хвороба негативно корелює з максимальною частотою серцевих скорочень. Цікаво, що існує негативний зв'язок між холестеринем і хворобами серця. Чоловіки приблизно в 2,44 рази частіше страждають серцевими захворюваннями, ніж жінки. Існують чіткі відмінності між типами болю в грудях. Пацієнти з безсимптомним болем у грудях (ASY) приблизно в шість разів частіше страждають від серцевих захворювань, ніж у пацієнтів з атиповою стенокардією (ATA). ЕКГ у спокої: значення електрокардіограми у спокої співставні. Пацієнти з аномаліями пульсу ST-T мають вищий ризик розвитку серцевих захворювань, ніж ті, у кого їх немає [16]. Стенокардія фізичного навантаження: люди, які мають стенокардію, спричинену фізичним навантаженням, мають майже в 2,4 рази більше шансів мати серцеві захворювання, ніж люди, які не страждають. Нахил сегмента ST при максимальному навантаженні змінюється. ST Slope Up має значно нижчий ризик серцево-судинних захворювань, ніж інші два сегменти. Стенокардія, спричинена фізичним навантаженням, з оцінкою «Так» майже в 2,4 рази частіше призводить до захворювання серця, ніж стенокардія, спричинена фізичним навантаженням, з оцінкою «Ні».

4 РОЗРОБКА ТА ТЕСТУВАННЯ

4.1 Програмне забезпечення

PyTorch – це платформа машинного навчання на основі бібліотеки Torch, яка використовується для таких програм, як комп'ютерне бачення та обробка природної мови, спочатку розроблена Meta AI, а тепер є частиною парасольки Linux Foundation. Це безкоштовне програмне забезпечення з відкритим кодом, випущене за модифікованою ліцензією BSD. Хоча інтерфейс Python є більш відшліфованим і основним напрямком розробки, PyTorch також має інтерфейс C++.

На основі PyTorch створено ряд програм для глибокого навчання, зокрема Tesla Autopilot, Pyro від Uber, Transformers від Hugging Face, PyTorch Lightning і Catalyst.

PyTorch надає дві функції високого рівня: тензорні обчислення (наприклад, NumPy) із потужним прискоренням за допомогою графічних процесорів (GPU) і глибокі нейронні мережі, побудовані на основі стрічкової системи автоматичного диференціювання.

Meta (раніше відома як Facebook) працює як з PyTorch, так і згортковою архітектурою для швидкого вбудовування функцій (Caffe2), але моделі, визначені двома фреймворками, були взаємно несумісними. Проект Open Neural Network Exchange (ONNX) був створений Meta та Microsoft у вересні 2017 року для перетворення моделей між фреймворками. Caffe2 було об'єднано з PyTorch наприкінці березня 2018 року [10]. У вересні 2022 року Meta оголосила, що PyTorch буде управлятися PyTorch Foundation, новоствореною незалежною організацією – дочірньою компанією Linux Foundation.

PyTorch визначає клас під назвою Тензор (`torch.Tensor`) для зберігання та роботи з однорідними багатовимірними прямокутними масивами чисел.

Тензори PyTorch схожі на масиви NumPy, але також можуть працювати на графічному процесорі NVIDIA, що підтримує CUDA. PyTorch також розробляє підтримку для інших платформ GPU, наприклад, ROCm від AMD і Metal Framework від Apple.

Тензор у фізиці схожий на тензор PyTorch тим, що це здебільшого багатовимірний масив. Єдина додаткова особливість тензора фізика, якої немає в тензорі PyTorch, полягає в тому, що під час індексування його записів деякі індекси записуються під або над індексом.

Тензор у фізиків підтримує чотири основні операції:

- доповнення;
- тензорний добуток;
- скорочення;
- зміна основи.

З них розрізнення ко/контраваріантності впливає лише на операцію зміни основи. На інші три операції не впливає ко/контраваріантність, і тому їх легко реалізувати для тензорів PyTorch.

4.2 Допоміжні прилади

Ішемічна хвороба серця або ішемічна хвороба серця (ІХС) розвивається, коли основні кровоносні судини, що живлять ваше серце, пошкоджуються або захворюють. Коронарні артерії - це кровоносні судини, які постачають кисень і кров до серця. Часто ІХС може стати причиною інфаркту. Це найпоширеніша форма хвороби серця в США, де вона є причиною понад 370 000 смертей на рік (Hage та ін., 2009).

За останнє десятиліття бездротовий безперервний моніторинг здоров'я був значним застосуванням нової технології охорони здоров'я. Загалом, ідея системи безперервного моніторингу здоров'я та розумних медичних пристроїв має величезний потенціал для бізнесу та добробуту людей. Ця технологія забезпечує пристрій у режимі реального часу, який може завчасно

попереджати людей про стан їх здоров'я. Перш ніж пояснювати основні моменти, було б краще обговорити деякі фізіологічні датчики, які можна використовувати для побудови безперервних систем моніторингу здоров'я. Ці датчики можна використовувати для відстеження людей у реальному часі. Крім того, їх можна використовувати як для діагностики, так і для моніторингу. Безперервний віддалений моніторинг працездатності використовує різні бездротові протоколи та мережі, включаючи Bluetooth, Wi-Fi, Zigbee та інші (LAN, WAN). Фізіологічні датчики, шлюз (з'єднання доступу) і хмара є основними будівельними блоками постійного моніторингу здоров'я (зберігання даних). Частота серцевих скорочень показує здоров'я нашого серця і допомагає визначити стан серцево-судинної системи. У клінічних умовах частота серцевих скорочень контролюється в контрольованих умовах, таких як вимірювання крові, вимірювання тону серця та електрокардіограма (ЕКГ). Однак його можна контролювати і вдома. Наше серце б'ється, щоб перекачувати збагачену киснем кров до наших тканин і виводити з них утворені клітинні відходи [8]. Чим більше наше серце працює для досягнення цих цілей, чим більше ми використовуємо свої м'язи, тим швидше наше серце буде битися, щоб перекачувати більше крові (Ornish, 2010). Монітор серцевого ритму – це, по суті, носна система, яка робить знімок серцевих скорочень і вимірює кількість ударів за хвилину (ударів на хвилину), щоб можна було точно контролювати серцевий ритм за допомогою деталей. Існує два типи методів моніторингу серця: електричні та оптичні. Електричний підхід має середню похибку 1 відсоток і середню вартість 150,00 доларів США. Оптичний метод має 15-відсотковий рейтинг точності та середню вартість 20 доларів. Середня частота серцевих скорочень людини в стані спокою для дорослих чоловіків становить приблизно 70 ударів за хвилину, а для дорослих жінок – близько 75 ударів за хвилину. Частота серцевих скорочень сильно відрізняється від однієї людини до іншої в залежності від фізичної підготовки, віку та генетичних особливостей. Існує декілька методів вимірювання частоти серцевих скорочень, таких як

фонокардіограма (ФКГ), ЕКГ, вимірювання форми хвилі артеріального тиску та вимірювання пульсу, але такі методи є клінічними та дорогими (Kaisti et al., 2019).

За останні два десятиліття алгоритми машинного навчання стають все більш популярними в різних сферах. В останні роки алгоритми машинного навчання використовувалися в галузі охорони здоров'я для аналізу великих обсягів даних з метою покращення розуміння захворювань. У результаті алгоритми машинного навчання можуть бути корисними для виявлення прихованих шаблонів у наборах медичних даних. Глибоке та машинне навчання досягли величезного прогресу. У Masethe & Masethe (2014) було проведено експеримент для прогнозування серцевих нападів і проведено порівняння, щоб знайти найкращий метод прогнозування. Це може бути корисним інструментом для лікарів, щоб передбачати критичні випадки у своїй практиці та надавати відповідні поради. Точність прогнозування, визначена алгоритмами J48, REPTREE та SIMPLE CART, свідчить про те, що використовувані параметри є надійними індикаторами для прогнозування наявності серцевих захворювань, але це дослідження має той недолік, що не може передбачити серцеві захворювання в режимі реального часу. У Chaurasia & Pal (2013) для побудови моделі використовуються три класифікатори: ID3, CART і DT, причому CART є найточнішим із 83,49% і 0,23 с. Важливими характеристиками серцевих захворювань є ср (біль у грудях), нахил (нахил пікового сегмента фізичного навантаження), Exang (стенокардія, спричинена фізичним навантаженням) і Restecg (електрокардіографічний показник у спокої). Однак це дослідження має той недолік, що не може передбачити захворювання серця в реальному часі. Sudhakar & Manimekalai (2014) описують різні методи, які використовувалися останніми роками для розрахунку коефіцієнта прогнозування серцевих захворювань. ANN, BN, дерева рішень і алгоритми класифікації – це деякі з методів, які використовуються, але це дослідження має той недолік, що не може передбачити захворювання серця в реальному часі. У дослідженні діагностики

захворювань серця за допомогою арбітражу нейронних мереж Оланійі, Ойедотун і Аднан (2015) використовуються два методи: SVM і ANN. Машина опорних векторів є найкращим алгоритмом для діагностики серцевих захворювань із рівнем точності 87,5 відсотка та високими значеннями чутливості та специфічності, але прогнозування серцевих захворювань є статичним і без реального часу. У Oloom et al. (2015), для моніторингу він використовується датчик пульсу.

Для створення моделі використовуються три класифікатори: BayesNet, SVM і FT, причому SVM є найточнішим з точністю 88,3%. Joshi & Nair (2015) підтверджують, що три методи прогнозування серцево-судинних захворювань, такі як «Дерево рішень», «Найв-Байес» і «Найближчий сусід», були використані для розробки моделі з використанням методів інтелектуального аналізу даних на основі класифікації, причому дерево рішень було найточнішим із точністю 92,3% [16]. У цій роботі не було прогнозу в реальному часі. Sultana, Haider & Uddin (2016) стверджують, що в аналізі методів інтелектуального аналізу даних для прогнозування захворювань серця використовуються п'ять класифікаторів: KStar, J48, SMO, Bayes Net і багатошаровий перцептрон, причому SMO показує найвищу продуктивність із 89-відсотковою точністю. У цій роботі не було зроблено прогнозів у реальному часі. Для пацієнтів із проблемами серця використовуються датчики для забезпечення постійного моніторингу. Для прикладу розглянемо роботу Алі (2017). Використовуючи датчик серцевого ритму та пороговий алгоритм для виявлення аномалій, було запропоновано систему моніторингу серцевого ритму та виявлення серцевого нападу на основі Arduino. Через неідеальний характер компонентів система має обмеження, пов'язані з ненаданням точних результатів системи моніторингу. У Majumder, Mondal & Deen (2017) представлено сучасне дослідження щодо фізіологічних параметрів і систем моніторингу активності, розроблених на платформі, що носить. Методи глибокого навчання викликали значний інтерес у медичних дослідників. Недавнє опитування показує, що близько половини медичних організацій

планують використовувати глибоке навчання (Kamilaris & Prenafeta-Boldú, 2018). Алгоритми глибокого навчання пропонують чудову можливість для раннього виявлення захворювань серця. Таким чином, використання глибоких нейронних мереж в аналізі набору даних дозволяє значно швидше та надійніше діагностувати серцеві захворювання, ніж за допомогою традиційних методів. Методи, звані глибинним навчанням, навчаються представляти об'єкти від низького рівня (необроблений вхід) до високого та більш абстрактного (Makantasis та ін., 2015). Він складається з кількох рівнів, які представляють різні рівні представлення; звідси і назва глибоке навчання (Gulli & Pal, 2017). Глибоке навчання визначається як автоматичний структурований алгоритм або ієрархічна структура, яка емулює навчання людини з метою отримання певних знань. Він виділяється тим, що для нього не потрібні попередньо запрограмовані правила, але сама система здатна самостійно навчитися виконувати завдання на етапі попереднього навчання. Він також характеризується тим, що складається з переплетених штучних нейронних мереж для обробки інформації. В основному використовується для автоматизації прогнозного аналізу. Алгоритми, які складають систему глибокого навчання, знаходяться в різних нейронних шарах, що складаються з ваг (чисел). Система в основному поділяється на три рівні: вхідний рівень (він складається з нейронів, які асимілюють вхідні дані, такі як зображення або база даних), рівень (це мережа, яка виконує обробку інформації та виконує проміжні обчислення. Чим більше нейронів у цьому шарі, тим складніше виконуються обчислення), вихідний рівень (це остання ланка в ланцюгу, і саме мережа приймає рішення або робить висновок, надаючи вихідні дані).

Як працює глибоке навчання? Комп'ютерні програми, які використовують глибоке навчання, проходять через той самий процес, що й маленька дитина, яка вчиться ідентифікувати об'єкт? Кожен алгоритм в ієрархії застосовує нелінійне перетворення до своїх вхідних даних і використовує швидкість навчання для створення статистичної моделі як

виходу. Ітерації тривають, доки результат не досягне прийнятного рівня точності. Це називається глибоким навчанням, тому що всі дані повинні пройти через усі приховані шари для етапу обробки. Типовим і суттєвим прикладом методу типу глибокого навчання є багатошарові перцептрони, які в основному є математичною функцією, яка відображає вхідні дані на вихідні дані. Це композиція (або мережа) кількох простіших нелінійних функцій (Karlik & Olgac, 2011), які називаються нейронами або перцептронами (рисунок 4.1), де вагові коефіцієнти, пов'язані з кожною з композиції, навчаються (зазвичай шляхом зворотного розповсюдження), щоб вивчити функцію, яка пов'язує вхідні дані з виходами контрольованої проблеми навчання [12].

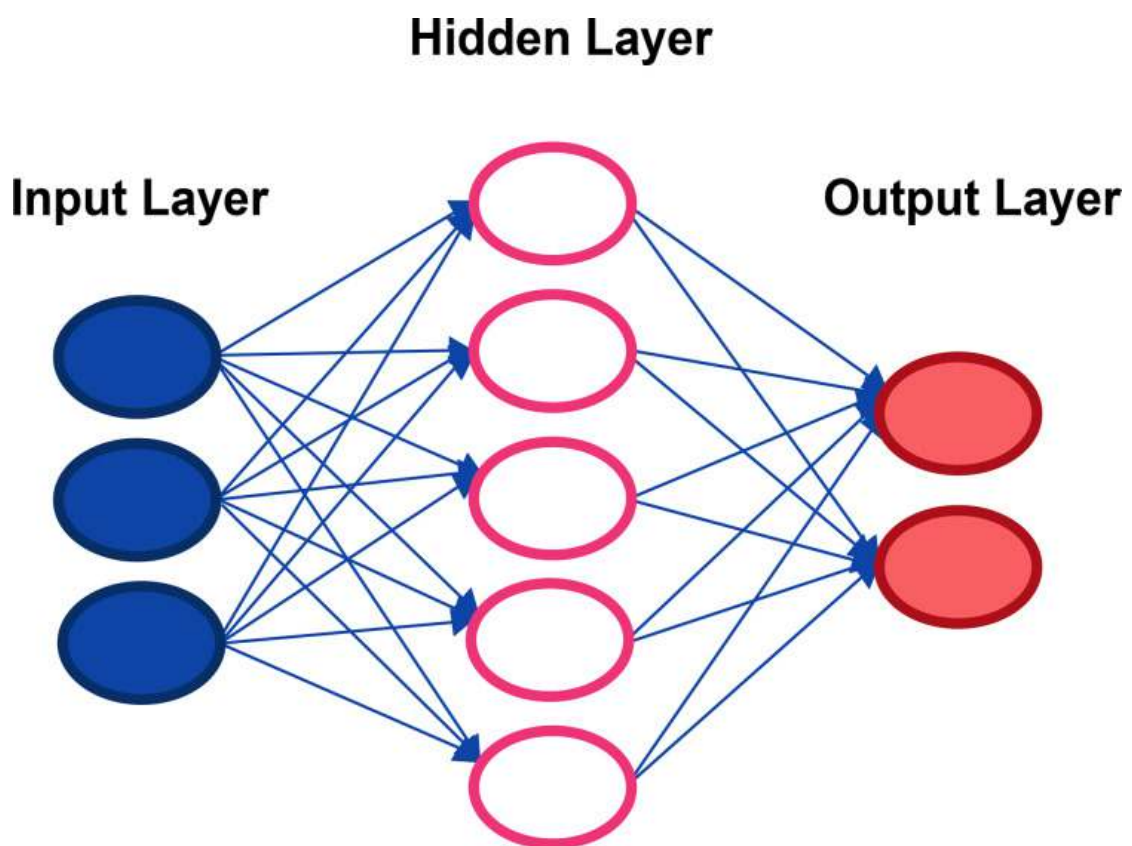


Рисунок 4.1 – Представлення класичної нейронної мережі

У фреймворку SHDML використовується багато методологій, наприклад: носяча вбудована система, Android, робочий стіл, ML, DL.

В останні роки розумні годинники почали випускати великі електронні компанії, а також нові стартапи. Однією з перших пропозицій був Samsung Galaxy Gear у вересні 2013 року. Приблизно через рік у квітні 2015 року було випущено Apple Watch. Більшість пропонованих інтерфейсів для моніторингу здоров'я використовують три архітектури: Бездротова мережа на тілі (WBAN), що складається з носимих пристроїв, датчики як блок збору даних, зв'язок і мережа, а також сервісний рівень. Було запропоновано вбудовану систему, яка містить мікроконтролер, пов'язаний із датчиком пульсу, який вимірює частоту серцевих скорочень людини. Мікроконтролер передає зібрані медичні дані в програму Android через бездротовий зв'язок Bluetooth. Додаток для Android зберігає частоту серцевих скорочень на хмарному сервері, і користувач може отримати доступ до вимірюваної частоти серцевих скорочень; кроки враховуються в додатку для Android. Користувач підтверджує, чи страждає він на ішемічну хворобу серця чи ні, використовуючи модуль глибокого навчання через додаток для Android/настільного комп'ютера.

Вбудована функція використовується для створення переносного комп'ютера. Вбудована система – це «комп'ютерна система», яка складається з комбінації мікропроцесорів, пам'яті та периферійних пристроїв введення/виводу зі спеціальним призначенням із сучасними вбудованими системами, що базуються головним чином на інтегрованому мікроконтролері.

Носима система – це реалізований на основі мікроконтролера atmega32 із датчиком пульсу та іншими модулями, датчик пульсу на основі технології PPG, датчики PPG (фотоплетизмографії) використовують технологію на основі світла для вимірювання швидкості кровотоку, що контролюється насосною дією серця. Було додано модуль Bluetooth для взаємодії з додатком Android для надсилання ударів BPM за хвилину та РК-монітор для перегляду BPM на носіїв. Запропонована структура реалізована шляхом розробки «бібліотеки датчиків пульсу», яка обробляє інтерфейси аналого-цифрового перетворювача АЦП і ТАЙМЕР з мікроконтролерами (рисунок 4.2). Світлодіод випромінює світло, яке потрапляє безпосередньо на вену. У венах

буде більше кровотоку всередині них лише тоді, коли серце качає (систоличний), у венах буде невеликий кровотік у них під час (діастолічний), тому, якщо ми відстежуємо потік крові, ми також можемо стежити за серцевими скороченнями. Датчик світла (рисунок 4.3), буде вловлювати більше світла, оскільки воно відбиватиметься потоком крові. Порогове значення «Т», яке визначається як початкова точка для кожного биття (25% або 50%) сигналу, аналогово-цифровий перетворювач (АЦП) є початком зчитування аналогового сигналу в точці Т, яка представляє отриманий світло кровотоку. Сигнал перетворюється на цифрове значення, із застосуванням шумопоглинання (рисунок 4.4) (дикротична виїмка, яка відображає закриття аортального клапана). Кожні 2 мс АЦП зупиняє мікропроцесор для зчитування з датчика, ця операція була досягнута за допомогою механізму переривання.

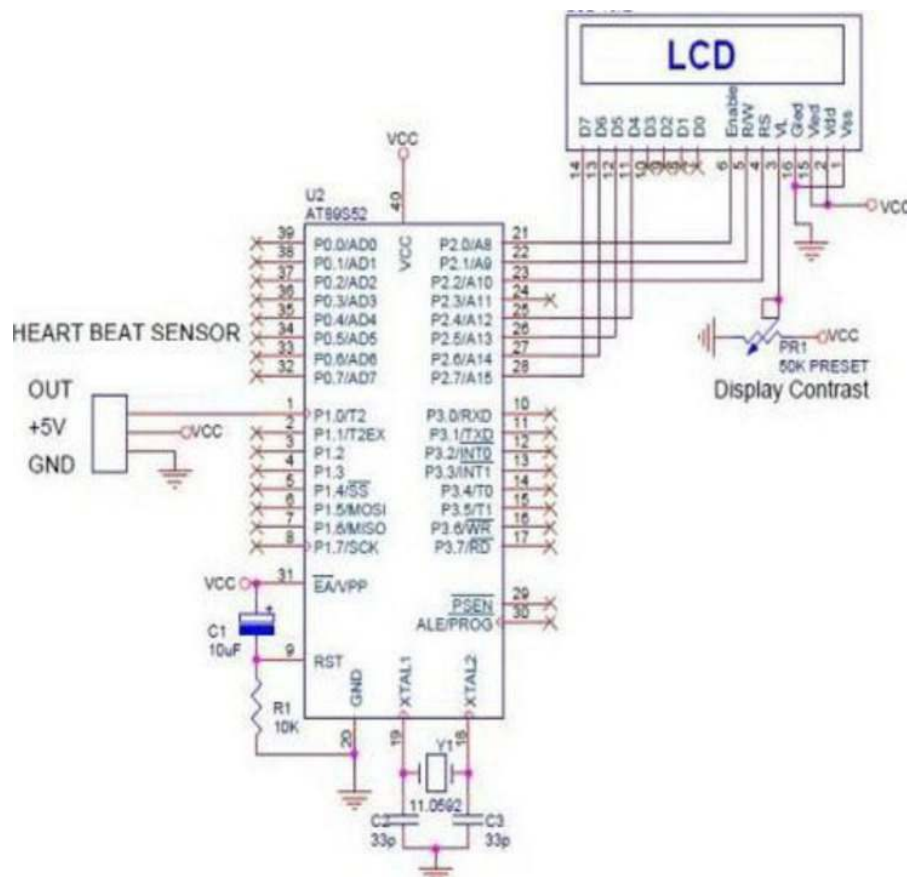


Рисунок 4.2 – Інтерфейси аналого-цифрового перетворювача АЦП і таймер з мікроконтролерами

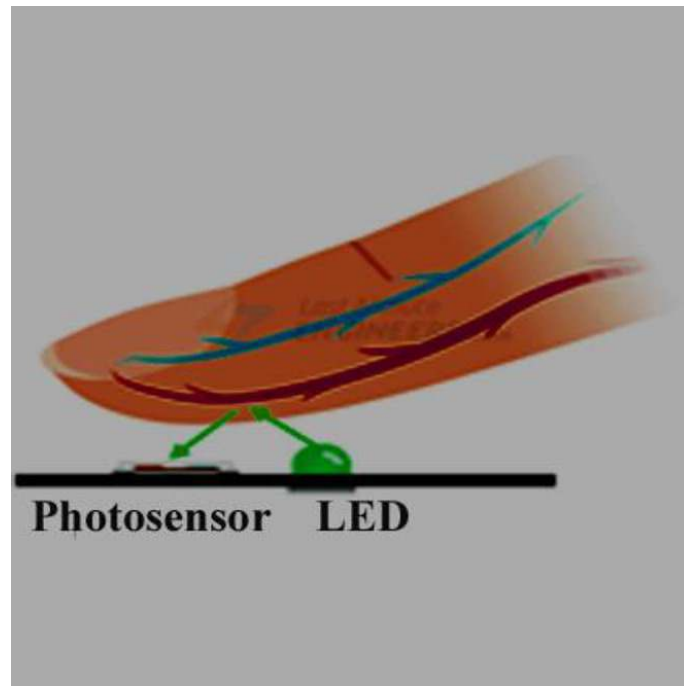


Рисунок 4.3 – Датчик світла

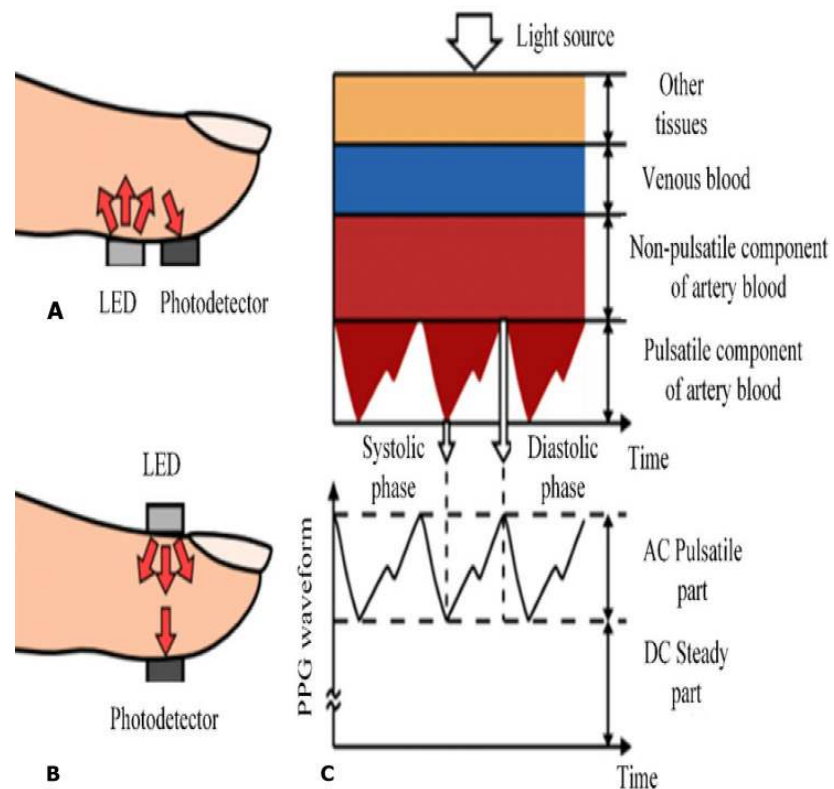


Рисунок 4.4 – Представлення електрокардіограми пацієнта з тренувальних даних

Структура SHDML складається з двох етапів. Перший ступінь ШДМЛ здатний контролювати стан ЧСС пацієнта. Частота серцевих скорочень визначається датчиком фотоплетизмографії (PPG) [9]. Сигнал обробляється за допомогою мікроконтролера ATmega32 для визначення частоти серцевих скорочень за хвилину. Потім він надсилає частоту серцевих скорочень, представлену як удари в хвилину, до програми через з'єднання Bluetooth.

Другий етап, SHDML, використовувався в системах підтримки медичних рішень для прогнозування та діагностики захворювань серця. Завдяки тренуванню техніки глибокого та машинного навчання, яка аналізує дані користувача для виявлення захворювань серця (рисунок 4.5). Оскільки наші дані, на яких навчатиметься машина, включають бажаний результат, тому використовувані типи DL та ML є машинним навчанням під наглядом.

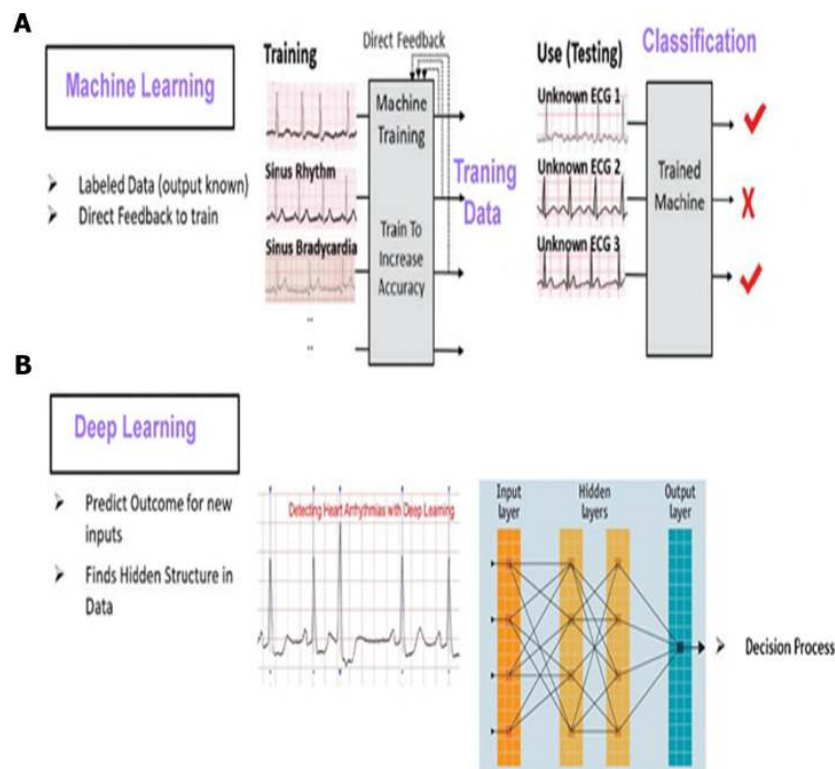


Рисунок 4.5 – Структура SHDML складається з ML і DL

Вхідні дані: збираються дані, на основі яких буде тренуватися модель, і вони мають бути точними, надійними та збалансованими. Використовується

база даних у реальному часі Firebase Cloud для збору та зберігання автентичних даних на відміну від традиційної реляційної бази даних (рисунок 4.7). Firebase: це серверна база даних у реальному часі як послуга (BaaS), яка дозволяє зберігати список об'єктів у форматі дерева [14]. Дані можна синхронізувати дані між пристроями (рисунок 4.7).

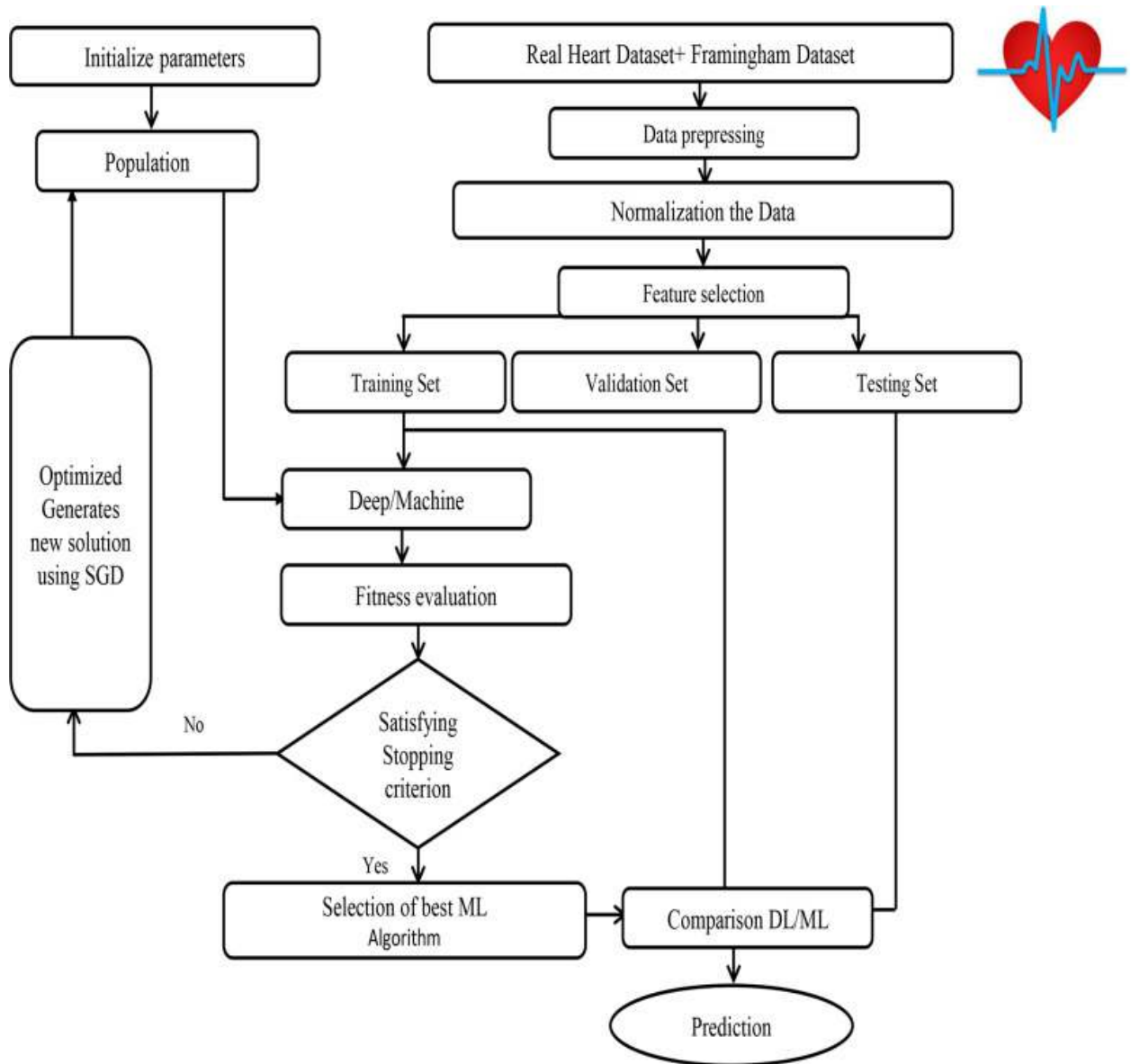


Рисунок 4.6 – Блок-схема інфраструктури SHDML

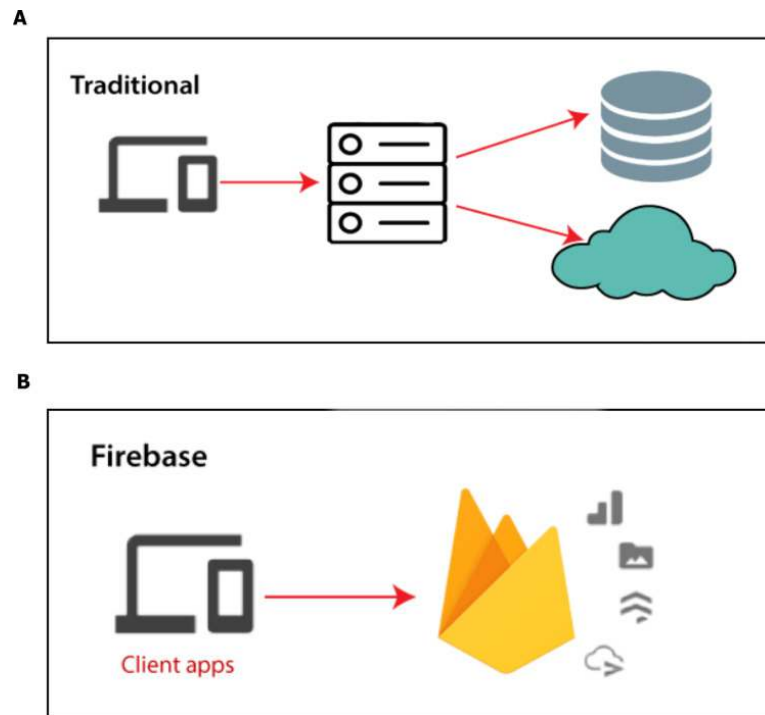


Рисунок 4.7 – База даних Firebase у реальному часі

4.3 Схема роботи

Глибоке навчання – це підполе машинного навчання, яке є важливим у нашому проєкті, оскільки допомагає вбудованій системі думати, а додатку для Android – передбачати, що користувач матиме серцевий напад у найближчі десять років. Ми використовували саме бібліотеку «Pytorch», тому що це бібліотека, яку легко вивчити та побудувати модель, а також з нею можна боротися з додатками Android у підключенні, тому це найкращий вибір серед бібліотек глибокого навчання. Ми використовували «CNN», щоб допомогти нам у покращенні нашої продуктивності завдяки вилученню зв’язків і ознак, що збільшує характеристики наших даних шляхом пошуку нових зв’язків між усіма ними, щоб отримати найкращий шаблон і комбінацію наших даних для прогнозування результатів [13].

Біологічно створені штучні нейронні мережі, також відомі як багатошарові перцептрони (MLP), здатні моделювати дуже складні нелінійні функції. Одним із найважливіших методів машинного навчання є штучні

нейронні мережі (ШНМ). Це орієнтовані на мозок системи, створені для імітації людського мислення, як це означає слово «нейронні». Вхідний, вихідний і прихований шари нейронних мереж складаються з трьох шарів. Прихований шар, як правило, складається з блоків, які перетворюють вхідні дані на шаблон, яким може керувати вихідний рівень. ШНМ є чудовими інструментами для вилучення та навчання комп'ютера ідентифікації шаблонів, які є надто складними або розпливчастими для розпізнавання програмістом-людиною [16]. З 1940-х років нейронні мережі використовувалися, а в останні десятиліття вони стали важливою частиною штучного інтелекту завдяки появі нової техніки, відомої як «зворотне поширення», яка дозволяє мережам знати, як змінювати свої приховані шари нейронів у випадках, коли результати не відповідають очікуванням творця (Nashif et al., 2018).

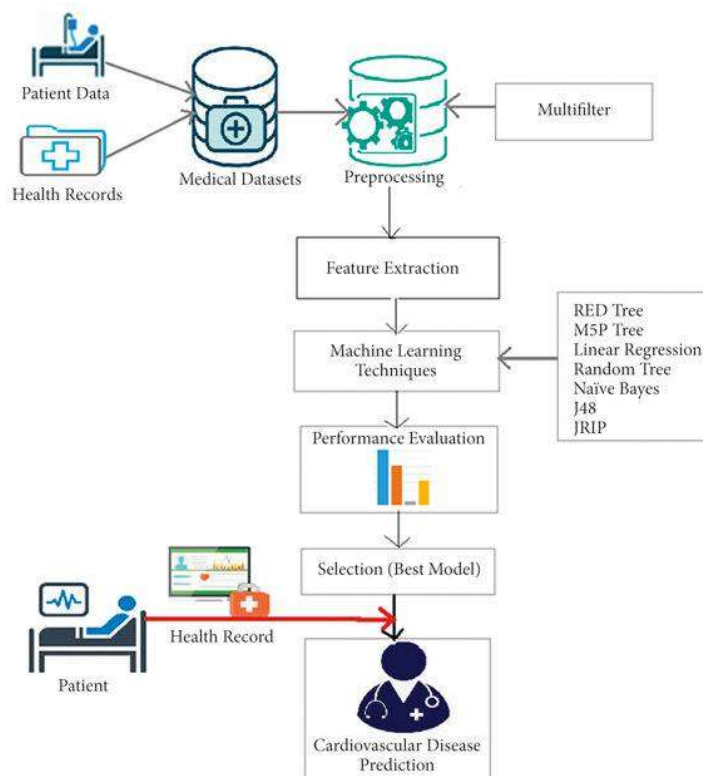


Рисунок 4.8 – Фреймворк запропонованої системи прогнозування серцево-судинних захворювань

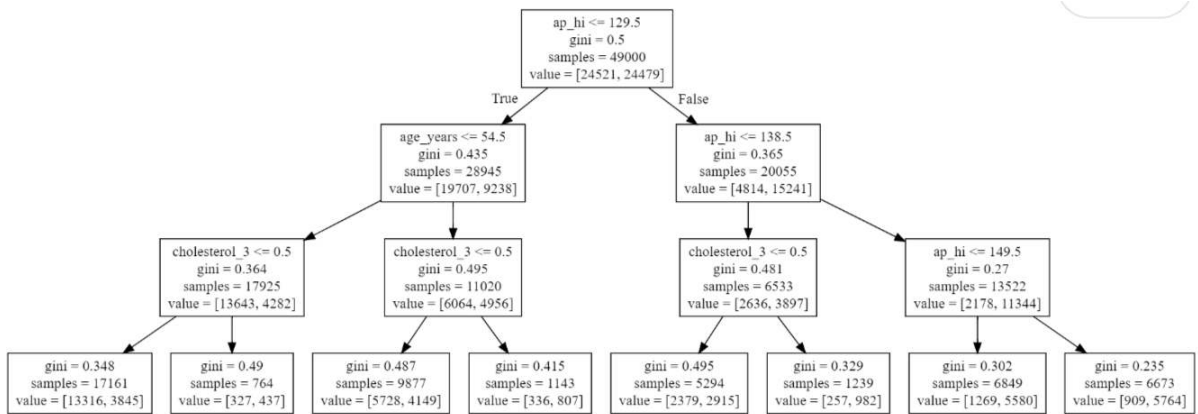


Рисунок 4.9 – Результати збору даних пацієнтів і їх візуалізація

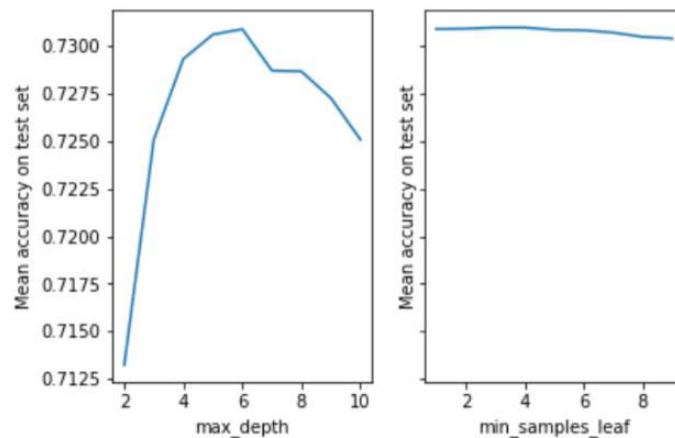


Рисунок 4.10 – результати обробки даних у вигляді графіку залежності параметрів `max_depth` і `min_samples_leaf`

Виходи з попередніх станів подаються як вхідні дані для поточного стану в рекурентних нейронних мережах (RNN). Приховані шари RNN мають здатність запам'ятовувати деталі. Вихідні дані, згенеровані в попередньому стані, використовуються для оновлення прихованого стану. RNN можна використовувати для прогнозування часових рядів, оскільки він має довго-короткочасну пам'ять (Ali et al., 2021), яка дозволяє відновлювати попередні вхідні дані.

ВИСНОВКИ

Захворювання серця є важкими і щороку спричиняють багато смертей. Якщо пацієнт ігнорує попереджувальні ознаки серцевої хвороби, жахливі наслідки можуть наступити швидко. Основною метою було окреслити кілька методів аналізу даних, які можуть бути ефективно застосовані для прогнозування серцевих захворювань. Тобто, щоб робити прогнози, які є точними та ефективними, використовуючи при цьому менше характеристик і тестів. У цьому дослідженні використовуються різні методи попередньої обробки та алгоритми машинного навчання для проведення поглибленого аналізу та отримання результатів. Було використано різноманітні класифікації навчання під наглядом, щоб класифікувати діагноз серцевих проблем, при цьому Успіх у класифікаціях діагностики захворювань серця має бути очевидним, якщо характеристики лікуються належним чином. Якщо характеристики зберігаються належним чином, категоризація серцево-судинних захворювань може працювати належним чином.

Цю роботу можна продовжити для діагностики серцевих захворювань за допомогою зрозумілої техніки машинного навчання, щоб встановити точність, справедливість, прозорість і результати моделі. Крім того, використання самостійно створеного набору даних дозволить розширити це дослідження.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. G. Mancia, S. Oparil, P.K. Whelton, M. McKee, A. Dominiczak, F.C. Luft, K. AlHabib, F. Lanas, A. Damasceno, D. Prabhakaran, G. La Torre The technical report on sodium intake and cardiovascular disease in low-and middle-income countries by the joint working group of the World Heart Federation, the European Society of Hypertension and the European Public Health Association Eur Heart J, 38 [Текст], (10) (2017). 712-719 с.
2. A. Coronary Heart disease [Текст], (2020), 468 с.
3. A. D'Souza To predict cardiac disease, data mining techniques are applied 74-77 Int. J.Res.Eng.Sci., 3 (3) [Текст], (2015) 75 с.
4. Sanjiv, J.S. (2015), [Phenomapping for Novel Classification of Heart Failure With Preserved Ejection Fraction], Circulation, №131, [Текст], 269–279 с.
5. Choi E. (2017) [Using recurrent neural network models for early detection of heart failure onset], Journal of the American Medical Informatics Association, №24, [Текст], 361–370 с.
6. Koelio L.P., Richard V. (2016) [Construction of machine learning systems in Python], DMK Publishing House, [Текст], 303 с.
7. W.J. Loesche Periodontal disease as a risk factor for heart disease Compendium, 15 (8) [Текст], (1994), 976-978 с.
8. S.K. Sen Using machine learning methods in heart disease detection and prediction Int. J.Eng.Comput. Sci., 6 (6) [Текст], (2017), 21623-21631 с.
9. R. Hertel, R. Benlamri A deep learning segmentation-classification pipeline for x-ray-based covid-19 diagnosis” Biomedical Engineering Advances (2022), Article 100041.
10. Chattopadhyay, M. Maitra MRI-Based brain tumor image detection using CNN based deep learning method Neurosci.Inf. (2022), Article 100060.
11. S.K. Mamatha, H.K. Krishnappa, N. Shalini Graph theory based segmentation of magnetic resonance images for brain tumor detection Pattern

Recogn Image Anal, 32 (1) [Текст], (2022), 153-161 с.

12. G.N. Ahmad, H. Fatima, A.S. Saidi Efficient medical diagnosis of human heart diseases using machine learning techniques with and without GridSearchCV IEEE Access [Текст], (2022), 576-596 с.

13. M.M. Rahman, M.R. Rana, Nur-A-Alam, M.S.I. Khan, K.M.M. Uddin A web-based heart disease prediction system using machine learning algorithms Netw. Biol., 12 (2) [Текст], (2022), 64-81 с.

14. S.K. Dey, M.M. Rahman, A. Howlader, U.R. Siddiqi, K.M.M. Uddin, R. Borhan, E.U. Rahman Prediction of dengue incidents using hospitalized patients, metrological and socio-economic data in Bangladesh: a machine learning approach PLoS One, 17 (7) (2022), Article e0270933.

15. D.V.B. Oliveira, J.F. da Silva, T.A. de Sousa Araújo, U.P. Albuquerque Influence of religiosity and spirituality on the adoption of behaviors of epidemiological relevance in emerging and Re-emerging diseases: the case of dengue fever J Relig Health, 61 (1) [Текст], (2022), 564-585 с.

16. Врaкiнa К. П. Proceedings of XII International Scientific and Practical Conference. 2022. С. 161-164. URL: EURASIAN-SCIENTIFIC-DISCUSSIONS-18-20.12.2022.pdf (sci-conf.com.ua)