

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет навчально-науковий центр заочної форми навчання
(повна назва)

Кафедра електронних обчислювальних машин
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

Рівень вищої освіти другий (магістерський)

Модель машинного навчання в піринговій мережі
медичних установ

(тема)

Виконав:

студент II курсу, групи СПЗМ-22-1
Свірщевський К.О.
(прізвище, ініціали)

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

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

Освітня програма Системне програмування
(повна назва освітньої програми)

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

Допускається до захисту

Зав. кафедри ЕОМ

(підпис)

Коваленко А.А.

(прізвище, ініціали)

2024 р.

Харківський національний університет радіоелектроніки

Факультет навчально-науковий центр заочної форми навчання

Кафедра електронних обчислювальних машин

Рівень вищої освіти другий (магістерський)

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

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

Освітня програма Системне програмування
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри _____
(підпис)

“ _____ ” _____ 20__ р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

студенту Свірщевському Кирилу Олександровичу
(прізвище, ім'я, по батькові)

1. Тема роботи Модель машинного навчання в піринговій мережі медичних установ

затверджена наказом по університету від “ 01 ” квітня 2024 р. № 45 Стз

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

3. Вхідні дані до роботи 1) медичні установи, об'єднані в однорангову мережу;

2) порівняння з класифікаторами НММ, CNN, RBFN, тощо;

3) цільові функції: швидкість і точність формування прогнозів.

4. Перелік питань, що потрібно опрацювати у роботі _____

1) аналіз проблеми побудови сучасних медичних інформаційних систем;

2) вибір та обґрунтування методів навчання;

3) розробка моделі для пірингової мережі медичних закладів;

4) удосконалення запропонованої системи;

5) проведення експериментальних досліджень;

б) висновки.

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	Аналіз проблеми	02.04.24-08.04.24	
2	Вибір та обґрунтування методів навчання	09.04.24-16.04.24	
3	Розробка базової моделі системи	17.04.24-22.04.24	
4	Удосконалення запропонованої системи	23.04.24-06.05.24	
5	Проведення експериментів	07.05.24-23.05.24	
6	Оформлення матеріалів кваліфікаційної роботи	24.05.24-03.06.24	
7	Подання кваліфікаційної роботи керівникові та її попередній захист	04.06.24-07.06.24	
8	Подання кваліфікаційної роботи на рецензування	08.06.24-12.06.24	

Дата видачі завдання 01 квітня 2024 р.

Студент _____
(підпис)

Керівник роботи _____
(підпис)

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

РЕФЕРАТ

Пояснювальна записка кваліфікаційної роботи: 81 с., 18 рис., 10 табл., 1 дод., 47 джерел.

ДІАГНОСТИКА, МАШИННЕ НАВЧАННЯ, МЕДИЧНА ІНФОРМАЦІЙНА СИСТЕМА, ПОКАЗНИКИ ФІЗІОЛОГІЧНОЇ ДІЯЛЬНОСТІ, ПРОГНОЗУВАННЯ, СИСТЕМА ПРИЙНЯТТЯ РІШЕНЬ, ЧАСОВЕ ВІКНО.

Метою кваліфікаційної роботи є розробка моделі машинного навчання в пiринговiй мережi медичних установ.

В роботі досліджено технології, пов'язаними з комп'ютерними мережами, машинним навчанням, великими даними та розподіленими обчисленнями. Запропоновано систему R2P-навчання, яка сприяє підвищенню ефективності навчання та точності діагностики з великою кількістю даних пацієнтів. Оскільки потоки даних мають значний вплив на систему, замість безперервного надсилання часових рядів даних пацієнтів, дані організовуються та передаються в певному часовому вікні на основі терміновості пацієнтів. Запропонований алгоритм базується на концепції NEWS.

ABSTRACT

Master's thesis: 81 pages, 18 figures, 10 tables, 1 appendix, 47 sources.

DECISION MAKING SYSTEM, DIAGNOSTICS, MACHINE LEARNING, MEDICAL INFORMATION SYSTEM, PREDICTION, TIME WINDOW, VITAL INDICATORS.

The purpose of the qualification work is to develop a machine learning model for the P2P network of medical institutions.

Technologies related to computer networks, machine learning, big data and distributed computing are explored in the work. A P2P learning system is proposed, which contributes to the improvement of learning efficiency and diagnostic accuracy with a large amount of patient data. Because data streams have a significant impact on the system, instead of continuously sending a time series of patient data, the data is organized and sent in a specific time window based on the urgency of the patients. The proposed algorithm is based on the NEWS concept.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ	8
ВСТУП	10
1 АНАЛІЗ ПРОБЛЕМИ	12
1.1 Загальні відомості	12
1.2 Проблеми дослідження.....	17
1.3 Задачі дослідження	20
2 ВИБІР МЕТОДУ НАВЧАННЯ	22
2.1 Метод ELM	22
2.2 Інкрементне навчання за допомогою ELM	25
2.3 Попередня обробка даних	27
3 ШВИДКЕ НАВЧАННЯ В ПРИНГОВІЙ МЕРЕЖІ	30
3.1 Загальні відомості	30
3.2 Узагальнена модель системи	34
3.2.1 Збір даних і прогнозування в одному вузлі.....	36
3.2.2 Обмін даними між вузлами.....	36
3.1 ELM-SM	37
3.1.1 Генерація часткової моделі	38
3.1.2 Часткова модель з використанням MD5.....	40
3.1.3 Процес оновлення ELM-SM на одноранговому пристрої	41
3.2 Висновок за розділом.....	43
4 ПРІОРИТЕЗАЦІЯ ДАНИХ	44
4.1 Загальні відомості	44
4.2 Огляд запропонованого методу.....	45
4.3 Життєво важливі ПФД, клінічні події та критерії оцінки терміновості.....	48
4.4 Процесор на основі часових вікон.....	51

4.4.1 Блок даних у часовому вікні і його пріоритет	52
4.4.2 Збирання даних за допомогою фільтра.....	53
4.5 Процесор пріоритетів.....	54
4.6 Експериментальні оцінки та результати.....	55
4.6.1 Попередня обробка даних	56
4.6.2 Класифікація за пріоритетністю	57
4.6.3 Ефективність збирання даних.....	58
4.6.4 Середній час очікування.....	60
4.7 Висновки за розділом	63
ВИСНОВКИ.....	64
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ	65
ДОДАТОК А ГРАФІЧНИЙ МАТЕРІАЛ КВАЛІФІКАЦІЙНОЇ РОБОТИ	71

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ
І ТЕРМІНІВ

- АТ – артеріальний тиск
- ДАТ – діастолічний артеріальний тиск
- ЕКГ – електрокардіограма
- ІКТ – інформаційно-комунікаційні технології
- ПФД – показники фізіологічної діяльності
- САТ – систолічний артеріальний тиск
- СПР – система прийняття рішень
- ЧДР – частота дихальних рухів
- ЧСС – частота серцевих скорочень
- CDSS – система підтримки клінічних рішень (англ., Clinical Decision Support System)
- CNN – згорточна нейронна мережа (англ., Convolution Neural Network)
- ELM – нейронна мережа типу «машина екстремального навчання» (англ., Extreme Learning Machine)
- ELM-SM – удосконалена модель ELM з використанням часткових моделей (англ., Extreme Learning Machine with Semi-Model)
- FIFO – дисципліна обслуговування «першим прийшов – першим вийшов» (англ., First In – First Out)
- HMM – прихована модель Маркова (англ., Hidden Markov Model)
- IoT – Інтернет речей (англ., Internet of Things)
- ML – машинне навчання (англ., Machine Learning)
- NEWS – Національний рейтинг раннього попередження у Великобританії (англ., National Early Warning Score)
- NN – нейронні мережі (англ., Neural Networks)
- OS – онлайн-послідовне навчання (англ., Online-Sequential Learning)
- SPO₂ – насичення крові киснем

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

TWBP – процесор на основі часових вікон (англ., Time-Window Based Processor)

TWDC – дані часового вікна (англ., Time-Window Data Chunk).

ВСТУП

Інвестиції в глобальний ринок медичних інформаційних технологій значно зростають, що робить системи охорони здоров'я все більш популярними в багатьох різних медичних закладах по всьому світу [1]. Таке значне збільшення чітко відображає те, як громадськість та державні установи зосереджуються на вдосконаленні систем охорони здоров'я, що сприяє подовженню тривалості життя та зменшенню витрат на охорону здоров'я.

В умовах швидкої зміни інформаційних технологій в сфері охорони здоров'я та у зв'язку із застосуванням технологій машинного навчання (ML) розробляються і досліджуються різні типи систем автоматизації. Ці системи збирають різноманітну інформацію про стан здоров'я пацієнтів, і лікар використовує отримані дані для прийняття ефективних рішень [2]. З іншого боку, дані використовуються для навчання «розумних» систем охорони здоров'я. В якості однієї з поширених систем охорони здоров'я можна зазначити систему підтримки клінічних рішень (CDSS) [3], яка використовується для автоматичного встановлення надійних діагнозів, що забезпечує якість і безпеку лікування в різних медичних закладах і сприяє зменшенню навантаження на лікарів із великою кількістю пацієнтів.

Досягнення в Інтернеті речей (IoT) прискорюють розвиток розумних систем охорони здоров'я, постачаючи до них дані ширшого типу [4]. На основі величезної кількості (більше 90 мільйонів) переносних пристроїв у різних сферах, включаючи охорону здоров'я, у 2014 році [5] з'явився новий сегмент IoT під назвою «Wearable IoT» (WIoT). Системи охорони здоров'я з WIoT призводять до поступового переходу від систем, орієнтованих на лікарню, до середовища, орієнтованого на людину, яка може віддалено контролювати пацієнтів.

Отже, враховуючи значний прогрес у створенні натільних датчиків,

пристроїв Інтернету речей і розвиток технологій ML, різні дослідження орієнтуються на системи дистанційного моніторингу охорони здоров'я, вбудовані в CDSS. Багато рішень Інтернету речей для сфери охорони здоров'я пропонуються різними компаніями (Microsoft, Alter Calsoft Labs, КАА та ін.) з метою допомогти пацієнтам і лікарям. Ці інтелектуальні системи забезпечують надійні рішення для віддаленого спостереження за людьми похилого віку з хронічними захворюваннями, які є основною причиною багатьох смертей у більшості західних країн. Ці захворювання зазвичай викликані нездоровим способом життя, неправильним харчуванням і вродженими генетичними проблемами [6]. У цих медичних системах дані пацієнтів безперервно збираються з пристроїв IoT та передаються до медичних установ для подальшого аналізу. Після аналізу даних надсилається відповідне клінічне рішення, яке в режимі реального часу відображає стан здоров'я пацієнту.

Використовуючи різноманітні передові технології, розумні системи охорони здоров'я відіграють важливу роль у покращенні медичних послуг. Проте такі системи стикаються з проблемою збільшення обсягу даних пацієнтів, нових захворювань і змін існуючих моделей захворювань, які вимагають передової інтелектуальної системи для забезпечення високої швидкості обробки з високою діагностичною точністю.

1 АНАЛІЗ ПРОБЛЕМИ

1.1 Загальні відомості

Різноманітні роботи дослідників у сфері розумних систем охорони здоров'я пропонують багатообіцяючі рішення. Однією з поширених інтелектуальних систем охорони здоров'я є CDSS. Вона може діагностувати різні захворювання, використовуючи медичні записи пацієнтів за різними класифікаторами ML. Наприклад, можуть використовуватися дані ЕКГ для виявлення серцевих проблем пацієнтів [7], певна кількість показників фізіологічної діяльності – для прогнозування декількох клінічних подій [8]. Можна також зазначити рішення [9] для діагностики раку печінки за допомогою магнітно-резонансної томографії.

Портативні смарт-пристрої збирають численні сигнали фізіологічної діяльності пацієнта за допомогою датчиків на тілі та відправляють їх на сервер. Клінічний заклад може мати локальний сервер або розміщений у хмарі. CDSS використовується для інтелектуального прогнозування, яке підтримує діагноз лікарів. Якщо було виявлено аномальну подію, система може повідомити пацієнта належним чином.

Крім CDSS, існують різні види розумних систем, які обслуговують пацієнтів. У багатьох дослідженнях пропонуються системи моніторингу стану здоров'я, що надають допомогу користувачам, яким важко отримати доступ до медичних послуг. В якості прикладу можна навести різноманітні дослідження, присвячені Ambient Assisted Living (AAL) [10]. AAL – це технологічний підхід із використанням інформаційно-комунікаційних технологій (ІКТ), який можна використовувати для підтримки пацієнтів або людей похилого віку вдома. AAL дозволяє захистити здоров'я різних користувачів, дозволяючи їм жити незалежно від інших у власному домі. Він спрощує повсякденну діяльність за допомогою домашньої автоматизації та

забезпечує надійні інтелектуальні рішення для моніторингу та догляду за користувачами.

На рисунку 1.1 показана типова архітектура систем медичного моніторингу, яка включає два основних компоненти: модуль збору даних і платформу обробки даних. Наприклад, інтелектуальний пристрій IoT постійно збирає дані сигналів фізіологічної діяльності пацієнтів за допомогою їхніх переносних датчиків, а потім надсилає ці дані CDSS. CDSS може бути розміщено на хмарній платформі або на локальному сервері медичного закладу. Система не тільки забезпечує «розумні» рішення (класифікація або прогнозування), які допомагають клініцистам (наприклад, лікарям або медсестрам), але також надсилає належне сповіщення пацієнту, коли виявлено аномальну клінічну подію. Медичні дані пацієнтів також використовуються в якості навчальних наборів для оновлення CDSS. Оновлений CDSS може мати кращу точність діагностики та виявляти більше захворювань у разі майбутніх подій.

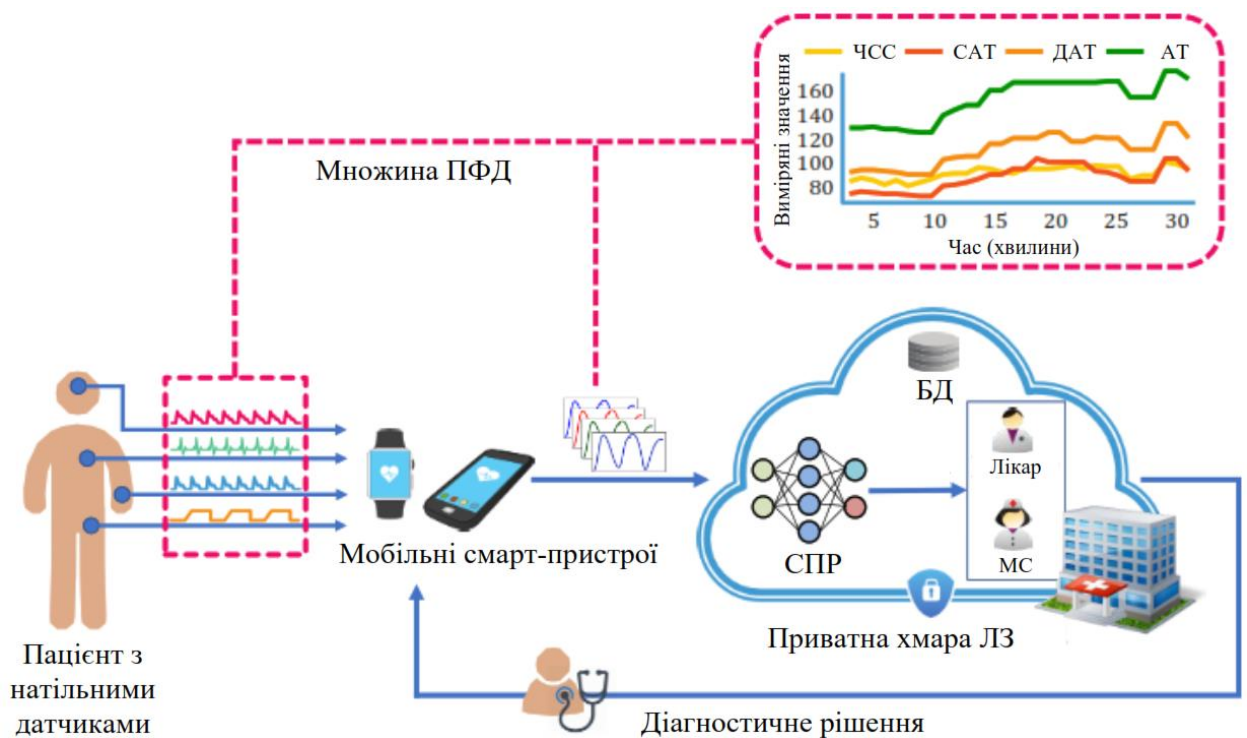


Рисунок 1.1 – Типова архітектура системи медичного моніторингу

Оскільки щодня генеруються величезні обсяги необроблених медичних даних від пацієнтів [11], системи охорони здоров'я відчувають значний тиск на три етапи обробки цих медичних даних, а саме збір даних, вивчення даних і прогнозування даних. Дані, створені датчиками, мають три основні ознаки великих даних [12]: обсяг; швидкість; різноманітність. Ситуація стає більш критичною, оскільки люди старшого віку вимагають постійного моніторингу. Ефективний процес навчання з великими даними покращується завдяки застосуванню хмарних платформ, які використовуються в багатьох дослідженнях [8, 13]. Слід також зазначити як інші проблеми, наприклад, конфіденційність [14].

Незважаючи на те, що дослідники запропонували різні медичні інтелектуальні системи для вирішення різноманітних завдань, існують два суттєвих обмеження, які впливають на продуктивність системи:

- ізолюваність системи – система не поділяє знання, отримані з даних пацієнтів в системах, що використовуються різними медичними установами. Обмін знаннями може підвищити точність та ефективність діагностики, але система має розглянути дієвий та надійний метод вивчення нових невідомих знань, які надсилаються іншими системами;

- безпріоритетна обробка даних – система обробляє дані однаково, не враховуючи терміновість пацієнтів, що призводить до потенційних затримок у лікуванні важких станів пацієнтів. Обробка даних включає передачу даних від пристроїв IoT до медичних установ (а саме збір даних) і прогнозування даних у медичних установах.

Задля вирішення проблеми ізолюваності, інформаційні системи розподіленого медичного закладу мають бути об'єднані між собою різними мережевими методами. У результаті виникає медична спільнота для обміну професійними знаннями між різними медичними установами. Така спільнота (рисунок 1.2) може бути реалізована на основі певної топології мережі за одним із двох варіантів:

- централізована мережа;

- розподілена/однорангова/пірингова (P2P) мережа.



Рисунок 1.2 – Співтовариство у сфері охорони здоров'я

Типові підходи розподіленого аналізу даних [15] на основі централізованої мережі можна продемонструвати за допомогою моделі обміну даними, показаної на рисунку 1.3(a). Цей підхід є вигідним з точки зору спільного використання та вивчення даних у централізований спосіб. Спочатку кожен клієнт збирає дані та надсилає їх на центральний сервер. Далі центральний сервер оновлює власний класифікатор на основі цих даних, після чого останній оновлений класифікатор синхронізується з усіма клієнтами. Однак централізована модель стикається з деякими обмеженнями:

- висока центральна залежність. Усі клієнти залежать від центрального сервера та ізольовані від інших клієнтів. Під час технічного обслуговування або періоду збою центрального сервера клієнти не можуть отримувати оновлення від центрального сервера.

- незбалансованість процесів обміну та навчання. Центральний сервер інтенсивно вимагатиме обчислювальних і комунікаційних ресурсів зі

збільшенням розміру вибірок даних і кількості клієнтів значно збільшуються.

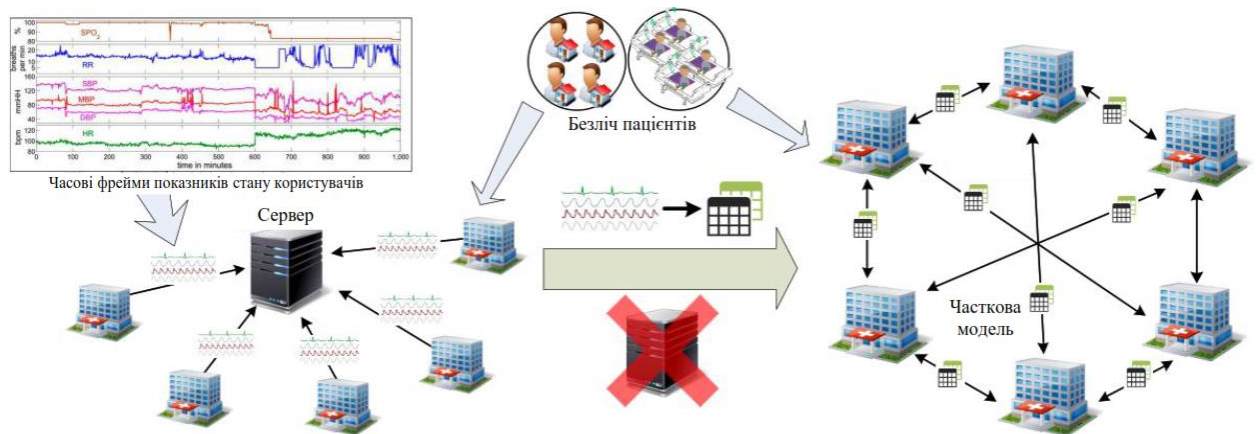


Рисунок 1.3 – Порівняння моделі централізованого навчання та моделі P2P-навчання

Отже, у моделі централізованого навчання кожна медична установа надсилає дані про показники фізіологічної діяльності (ПФД) з пристроїв IoT пацієнтів на центральний сервер для обробки. У запропонованій моделі навчання на основі пірингової мережі медичні установи підключаються без центрального сервера. Медичні знання, отримані з даних ПФД, розподіляються між медичними установами. Таким чином, пірингова мережа, яка показана на рисунку 1.3(б), використовується для покращення медичної інформаційної системи, за рахунок усунення залежності від центрального сервера та балансування навантаження в кожному медичному закладі.

В цій роботі процес обміну та навчання в піринговій мережі позначається як P2P-навчання, яке також деталізується як інкрементний процес навчання [16]. Сценарій для ілюстрації системи P2P-навчання показано на рисунку 1.4. Припустимо, що є три лікарні (Н1, Н2, Н3), які збирають дані від пацієнтів і роблять прогноз. Спочатку Н1 має лише набір даних D_A , за допомогою якого може виявляти хворобу «А»; Н2 має набір даних D_B , за допомогою якого може виявляти хворобу «В»; Н3 має набір даних D_C , за допомогою якого може виявляти хворобу «С». Після обміну та

вивчення знань між Н1, Н2 та Н3, усі вони мають здатність виявляти хвороби «А», «В» і «С».

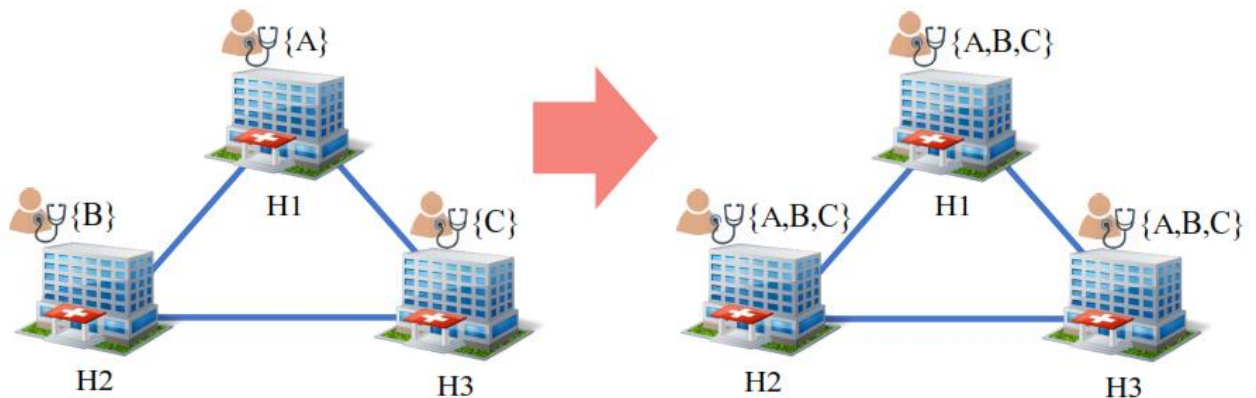


Рисунок 1.4 – Приклад системи навчання P2P

Головна мета цієї роботи – дослідження покращеної розумної системи охорони здоров'я для подолання певних обмежень шляхом підтримки P2P-навчання та обробки даних з урахуванням терміновості пацієнтів. Система повинна забезпечувати хорошу точність діагностики і ефективну обробку даних з точки зору їхнього збирання даних, аналізу і прогнозування.

1.2 Проблеми дослідження

Дослідження аналізує проблеми, пов'язані з розробкою швидкої системи P2P-навчання, за допомогою якої кожна медична установа (вузол, «пір» – від англ. Peer) може підвищити точність своєї діагностики, використовуючи знання, отримані від інших установ, і може оптимізувати потік обробки даних пацієнтів на основі різної терміновості. Така система повинна діагностувати нові захворювання, не торкаючись необроблених наборів навчальних даних із новими захворюваннями. Таким чином, необхідно розглянути п'ять дослідницьких проблем, які перелічені нижче.

Проблема 1. Існуючі класифікатори ML витрачають багато часу на навчання та не підтримують P2P-навчання. Багатьма дослідженнями для

діагностики у сфері охорони здоров'я запропоновані різні класифікатори ML: наприклад, машина опорних векторів (SVM) [17], прихована модель Маркова (НММ) [18], нейронні мережі (NN) [19]. Більшість із цих існуючих класифікаторів мають повільний час навчання, що обмежує їх застосування в охороні здоров'я, оскільки багато характеристик існуючих захворювань змінюються, а нові захворювання з'являються зі зміною способу життя та навколишнього середовища. Наприклад, навчання НММ на 600 зразках займає 30 хвилин у віртуальній машині Amazon x3.large [20]. Незважаючи на те, що ELM [21] забезпечує швидку та високу точність прогнозування для аналізу медичних даних, цей метод не підтримує R2P-навчання.

Проблема 2. Великі масиви необроблених медичних даних для навчання. Для прийняття рішень клініцистами і навчання розумних системи охорони здоров'я використовується різноманітна інформація про стан здоров'я, зібрана від пацієнтів [2]. Наприклад, передові технології датчиків, що їх можна носити, дозволили контролювати численні життєво важливі показники пацієнта в будь-який час і в будь-якому місці. У міру накопичення багатьох життєво важливих показників великої кількості пацієнтів виникає проблема великих даних. Наприклад, життєво важливі показники, такі як частота серцевих скорочень (ЧСС), артеріальний тиск (АТ), частота дихальних рухів (ЧДР) і насичення крові киснем (SpO_2), є важливою частиною великих медичних даних [22]. Якщо числове значення кожного життєво важливого показника містить 4 байти, а період збирання даних становить 1 хвилину, то для 6 життєво важливих показників загалом збирається 24 байти даних за хвилину, що еквівалентно 33,75 кілобайтам (КБ) на день, або 12 мегабайтам (МБ) на рік. Якщо такі дані зібрати від 5 мільйонів пацієнтів, то обсяг даних становитиме 57,3 петабайта (ПБ) на рік. Обмін такими великими обсягами даних між різними медичними установами є неможливим з точки зору часу.

Проблема 3. Кожен вузол має різні порядок отримання та вивчення знань, якими поділяються інші вузли.

У пропонованій системі P2P-навчання кожен вузол ділиться знаннями та отримує знання від інших вузлів, що значною мірою залежить від пропускної здатності мережі та продуктивності комп'ютера. Наприклад, затримка мережі враховує час, необхідний для того, щоб пакет даних дістався від однієї визначеної точки до іншої, а пропускна здатність – це здатність дротової або бездротової мережі передавати максимальну кількість даних від однієї точки до іншої (через окрему комп'ютерну мережу або використовуючи підключення до Інтернету) за певний проміжок часу. Затримка мережі та пропускна здатність суттєво впливають на час отримання пакету в одноранговому пристрої. Очевидно, що менша затримка та більша пропускна здатність можуть скоротити час передачі. Крім того, загальній продуктивності комп'ютера сприяє швидкість центрального процесора – швидкість, з якою процесор може завершити цикл обробки, і яка зазвичай вимірюється в мегагерцах (МГц) або гігагерцах (ГГц). Це означає, що процесор 1,8 ГГц має приблизно вдвічі більшу швидкість, ніж процесор 900 МГц. Таким чином, система P2P-навчання стикається з труднощами, пов'язаними з невизначеними замовленнями отримання та навчання серед вузлів, спричинених різною продуктивністю обробки мереж і комп'ютерів.

Проблема 4. Кожен вузол має динамічні мітки даних як у процесі навчання, так і в процесі прогнозування. Традиційне навчання та прогнозування вимагає попереднього визначення та фіксації міток даних, що в реальних випадках є ненадійним. З розвитком технологій і зміною способу життя в різних медичних установах будуть виявлятися нові захворювання. Різні місця також впливають на оцінку захворювань. Наприклад, у розвинених країнах спостерігається більше випадків діабету, тоді як у країнах, що розвиваються, більше хворих на малярію. Це означає, що знання, якими користуються різні медичні установи, містять різні захворювання. Наприклад, на рисунку 1.4, вузол Н1 має вивчити знання, що містять невідомі мітки даних «В» і «С», а мітки даних «А» і «В» невидимі для Н3. Крім того, медичний заклад може передбачити нові зразки, які не належать

до жодних вивчених захворювань. Наприклад, вузол Н1 виконує знищення зразків із захворюванням С, маючи лише відомості про захворювання А та В. Ці проблеми, пов'язані з мітками динамічних даних, необхідно вирішувати як у процесах навчання, так і в процесах прогнозування системи P2P-навчання.

Проблема 5. Терміновість пацієнта динамічна з урахуванням послідовних у часі даних. В медичних системах дані про ПФД зазвичай збираються від пацієнтів і використовуються для прогнозування. Як показано на рисунку 1.1, ці дані є послідовними в часі і зазвичай збираються з певною періодичністю (наприклад, кожен хвилину або 10 хвилин). На основі налаштувань клінічних пріоритетів цим даним можуть бути надані різні пріоритети, які використовуються для сортування потоку пацієнтів, щоб пацієнти з більш невідкладними станами мали діагностуватися або лікуватися раніше пацієнтів з менш невідкладними захворюваннями. Але ці послідовні в часі дані мають різну кількість зростаючих і спадних тенденцій, як буде показано далі, через що терміновість пацієнта змінюється з часом. Потрібно враховувати точність і ефективність оцінки терміновості пацієнтів за допомогою послідовних у часі даних про ПФД.

1.3 Задачі дослідження

Для того щоб розв'язати проблемні питання дослідження, слід визначити ряд задач.

Задача 1. Розробити швидку систему P2P-навчання для медичної сфери, яка здатна прогнозувати клінічні події за допомогою життєво важливих ПФД. Слід розглянути проблеми (проблема 1 і проблема 2), пов'язані з ефективністю навчання та обміну інформацією в системі P2P-навчання. Важливість роботи з цими проблемами полягає насамперед у забезпеченні ефективної архітектури P2P для подальших досліджень. Оскільки різні медичні дані можуть використовуватися для діагностики охорони здоров'я,

життєво важливі ПФД (наприклад, частота серцевих скорочень і артеріальний тиск) спрямовані на впровадження та аналіз пропонованої системи P2P-навчання.

Задача 2. Покращити запропоновану систему охорони здоров'я за рахунок розподіленого онлайн-послідовного навчання (OS-навчання) [23] і прогнозування. Це дослідницьке питання призначене для вирішення проблем, пов'язаних із невизначеним порядком навчання серед вузлів і динамічних міток даних (проблеми 3 і 4). Зокрема, мета полягає в тому, щоб підвищити надійність системи P2P шляхом вбудовування OS-навчання, що дозволить класифікаторам ML вивчати навчальні дані, доступні в послідовному порядку.

Задача 3. Оптимізувати потік обробки в пропонованій системі з точки зору збору даних і прогнозування. Отже, потрібно створити надійну систему P2P, ефективну для навчання та спільного використання. Однак базова система не враховує терміновість даних пацієнтів під час збору та прогнозування даних, що впливає на медичні послуги, які надаються пацієнтам, з точки зору часу очікування та обсягу зібраних даних. Це питання розглядає проблему 5, пов'язану з оцінкою терміновості пацієнта. Виникає задача оптимізувати систему P2P і зробити її більш практичною та ефективною, ввівши терміновість пацієнтів.

2 ВИБІР МЕТОДУ НАВЧАННЯ

В цьому розділі розглянуто базовий метод ELM, інкрементний ELM і однокласовий ELM, які забезпечують основу для розробки системи P2P-навчання. Також описано попередньо оброблений набір даних, який використовується для оцінки ефективності запропонованих методів у всіх розділах.

2.1 Метод ELM

ELM [21] – це одношарова нейронна мережа (SLFN) без налаштування прихованого рівня. Основна перевага ELM полягає в тому, що він долає обмеження алгоритмів зворотного поширення, які зазвичай використовуються в штучних нейронних мережах, шляхом випадкового генерування вагових коефіцієнтів вхідних даних і аналітичного обчислення вагових коефіцієнтів вихідних даних. Основні обмеження зворотного поширення включають надмірну підгонку, високу вартість обчислення процесу навчання та локальні мінімуми. Крім того, швидкість навчання та продуктивність ELM також значно кращі, ніж інші звичайні алгоритми навчання.

Архітектуру ELM показано на рисунку 2.1. В ELM є три різних шари: вхідний, прихований і вихідний. Зразки з мітками та функціями подаються на вхідний шар ELM, а потім обробляються випадковими ваговими значеннями та функцією активації у прихованому шарі. Результат моделі розраховується на основі вихідних матриць із прихованого шару.

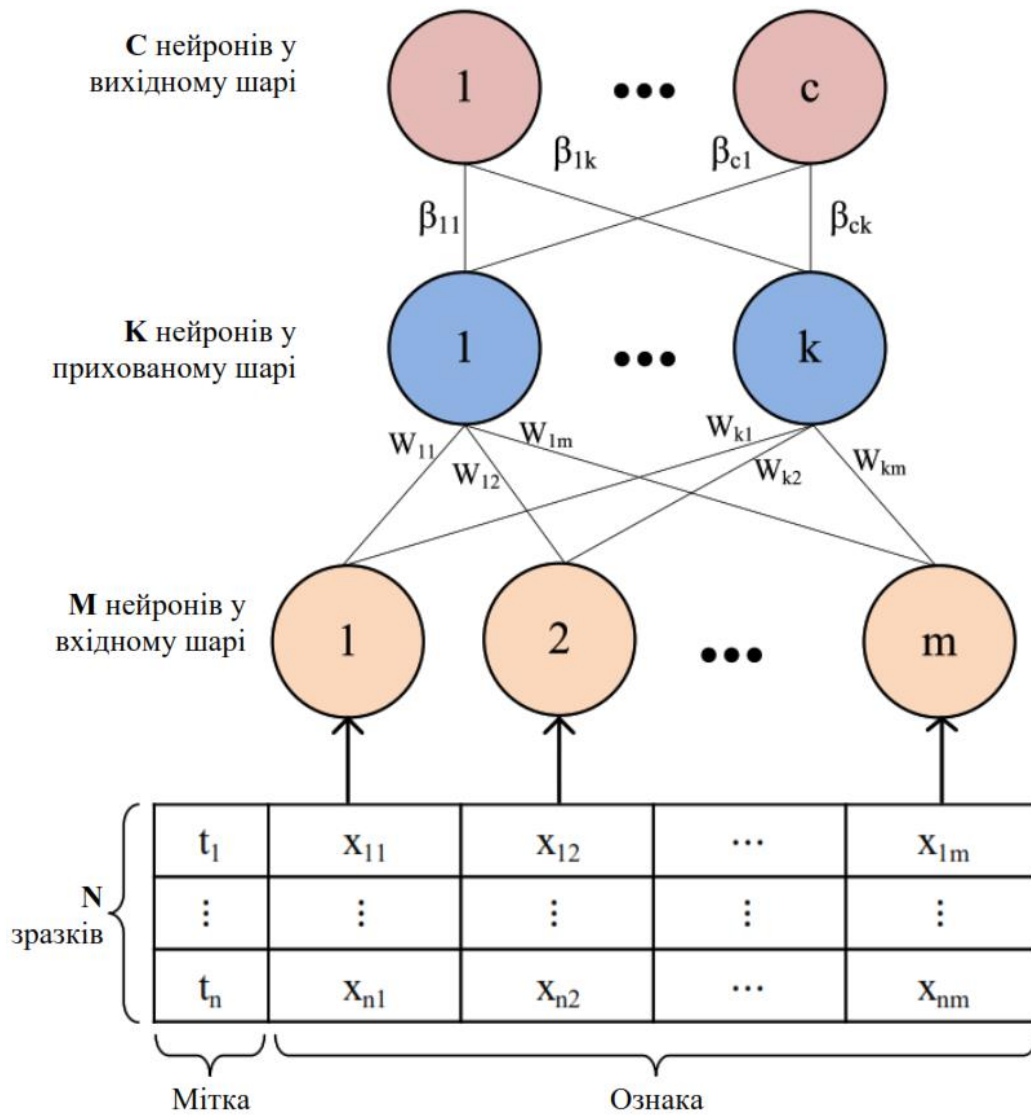


Рисунок 2.1 – Архітектура ELM

Припустимо, що ELM має M нейронів у вхідному шарі, K нейронів у прихованому шарі та C нейронів у вихідному шарі для N довільних окремих вибірок (\vec{x}_i, \vec{t}_i) , де $\vec{x}_i = [x_{i1}, x_{i2}, x_{i3}, \dots, x_{im}]^T \in \mathbb{R}^M$ і $\vec{t}_i = [t_{i1}, t_{i2}, t_{i3}, \dots, t_{ic}]^T \in \mathbb{R}^C$. Елементи m , i та c представляють індекс ознак, зразків і нейронів відповідно. ELM із K прихованих нейронів математично моделюється як

$$\vec{o}_i = \sum_{k=1}^K [\vec{\beta}_k \cdot G(\vec{\omega}_k, b_k, \vec{x}_i)] = \sum_{k=1}^K [\vec{\beta}_k \cdot G(\vec{\omega}_k \cdot \vec{x}_i + b_k)] \quad (2.1)$$

$$(\vec{o}_i, \vec{\beta}_k \in \mathbb{R}^C, \vec{\omega}_k \in \mathbb{R}^M, b_k \in \mathbb{R}, i = 1, 2, \dots, N),$$

де $\vec{\omega}_k = [\omega_{k1}, \omega_{k2}, \omega_{k3}, \dots, \omega_{km}]^T$ і b_k – випадкові вхідні ваги в k -му прихованому вузлі, $\vec{\beta}_k = [\beta_{k1}, \beta_{k2}, \beta_{k3}, \dots, \beta_{kc}]^T$ – ваговий вектор, що з'єднує k -й прихований вузол і вихідні вузли, $\vec{o}_i = [o_{i1}, o_{i2}, o_{i3}, \dots, o_{ic}]$ є i -м вихідним вектором ELM, і, нарешті, $G(*)$ відповідає виходу функції активації, яка використовується в нейронах прихованого шару. Зокрема, значення елементів у \vec{t}_i дорівнює 1, коли вихід нейрона належить до класу вибірки, а решта є -1. ELM може оцінити ці N вибірок з нульовою помилкою, що є основним принципом алгоритму найменших квадратів.

Оцінка

$$\sum_{n=1}^N \|\vec{o}_i - \vec{t}_i\| = 0 \quad (2.2)$$

може бути означена як

$$H \cdot \beta = T, \quad (2.3)$$

де

$$H = \begin{bmatrix} G(\vec{\omega}_1, b_1, \vec{x}_1) & \cdots & G(\vec{\omega}_k, b_k, \vec{x}_1) \\ \vdots & \ddots & \vdots \\ G(\vec{\omega}_1, b_1, \vec{x}_n) & \cdots & G(\vec{\omega}_k, b_k, \vec{x}_n) \end{bmatrix}_{N \times K} = \begin{bmatrix} g_{11} & \cdots & g_{1k} \\ \vdots & \ddots & \vdots \\ g_{n1} & \cdots & g_{nk} \end{bmatrix}_{N \times K} \quad (2.4)$$

$$\beta = \begin{bmatrix} \vec{\beta}_1 \\ \vdots \\ \vec{\beta}_k \end{bmatrix}_{K \times C} \quad \text{та} \quad T = \begin{bmatrix} \vec{t}_1 \\ \vdots \\ \vec{t}_n \end{bmatrix}_{N \times C} \quad (2.5)$$

H – вихідна матриця прихованого шару ELM із певним вхідним

набором даних $X = [\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_n]$. Розв'язок мінімальної норми вищевказаної лінійної системи за методом найменших квадратів можна виразити як

$$\hat{\beta} = H^\dagger \cdot T = (H^T H)^{-1} H^T T, \quad (2.6)$$

$$H^T H = \begin{bmatrix} u_{11} & \cdots & u_{1k} \\ \vdots & \ddots & \vdots \\ u_{k1} & \cdots & u_{kk} \end{bmatrix}_{K \times K} \quad (2.7)$$

та

$$H^T T = \begin{bmatrix} v_{11} & \cdots & v_{1c} \\ \vdots & \ddots & \vdots \\ v_{k1} & \cdots & v_{kc} \end{bmatrix}_{K \times C}, \quad (2.8)$$

де H^\dagger є псевдооберненим, який екстраполює обернену матрицю H у (2.3).

2.2 Інкрементне навчання за допомогою ELM

Задля пришвидшення процесу навчання з великими даними, пропонується інкрементний підхід до навчання на основі ELM [24].

Нехай початковий навчальний набір даних буде $D = (X, T)$, де $|D = \{(\bar{x}_j, \vec{t}_j) \mid \bar{x}_j \in \mathbb{R}^M, j = 1, 2, \dots, N\}$, N – кількість вхідних зразків. Щойно отриманий навчальний набір даних представлений $\Delta D = (\Delta X, \Delta T)$, де $\Delta D = \{(\bar{x}_j, \vec{t}_j) \mid \bar{x}_j \in \mathbb{R}^M, j = N + 1, N + 2, \dots, N + \Delta N\}$, ΔN – кількість нових зразків.

Коли новий навчальний набір даних об'єднується з оригінальним, маємо:

$$D' = \left(\begin{bmatrix} X \\ \Delta X \end{bmatrix}, \begin{bmatrix} T \\ \Delta T \end{bmatrix} \right) = \left\{ (\bar{x}_j, \bar{t}_j) \mid \bar{x}_j \in \mathbf{R}^M, j=1,2,\dots,N, N+1, N+2, \dots, N+\Delta N \right\}.$$

Нова вихідна матриця прихованого шару H' може бути отримана з H і ΔH . Тепер маємо H' і T' , де

$$H' = \begin{bmatrix} H \\ \Delta H \end{bmatrix} \quad \text{та} \quad T' = \begin{bmatrix} T \\ \Delta T \end{bmatrix}. \quad (2.9)$$

Перетворимо (2.7) і (2.8) відповідно до оператора множення матриць:

$$H'^T H' = \begin{bmatrix} H \\ \Delta H \end{bmatrix}^T \begin{bmatrix} H \\ \Delta H \end{bmatrix} = \begin{bmatrix} H^T & \Delta H^T \end{bmatrix} \begin{bmatrix} H \\ \Delta H \end{bmatrix} = H^T H + \Delta H^T \Delta H, \quad (2.10)$$

$$H'^T T' = \begin{bmatrix} H \\ \Delta H \end{bmatrix}^T \begin{bmatrix} T \\ \Delta T \end{bmatrix} = \begin{bmatrix} H^T & \Delta H^T \end{bmatrix} \begin{bmatrix} T \\ \Delta T \end{bmatrix} = H^T T + \Delta H^T \Delta T. \quad (2.11)$$

Оскільки $U = H^T H$ та $V = H^T T$, представимо (2.10) і (2.11) у такому вигляді:

$$U' = U + \Delta U \quad \text{та} \quad V' = V + \Delta V. \quad (2.12)$$

Тоді (2.6) стає

$$\hat{\beta} = U^{-1} V. \quad (2.13)$$

Зазвичай ELM має справу з даними двох або більше класів, проте зазначено [25], що також є можливість однокласової класифікації.

Аналізуючи (2.1) і (2.3), коли цільовий клас T є лише одним класом, а

саме із всіма 1 у T , β стає лінійно-апроксимованим відображенням H на T , що геометрично є наближенням у гіперплощині [9]. Тоді відстань будь-якої точки (вибірки) до гіперплощини, побудованої позаядерним ELM, визначається як [26]

$$d = |H \cdot \beta - T|. \quad (2.14)$$

Оцінка відстані доводить, що ELM здатний виявляти зразки аномалій на основі навчання з одним класом.

2.3 Попередня обробка даних

З метою оцінки продуктивності запропонованих методів у цій роботі, використовується набір даних з архіву MIT Physiobank MIMIC-II [27], який містить різні типи життєво важливих даних різних пацієнтів. Набір даних попередньо обробляється перед валідацією запропонованого методу для створення міток і ознак з різних діапазонів життєво важливих ПФД, а також для фільтрації шуму із вхідного набору даних. Для попередньої обробки медичних даних застосовуються відповідні методи [20] і програмна система MATLAB версії R2021b.

В застосованому обробленому наборі даних є 180 ознак у вибірці, яка класифікується за п'ятьма мітками. Перед навчанням запропонованого ELM-SM всі значення ознак відображаються у діапазон $[-1, 1]$ за масштабуванням min-max. Характеристики зразка та мітки описані нижче. Кожен зразок обробленого набору даних містить 6 сигналів ПФД, як показано на рисунку 2.2:

- частота серцевих скорочень (ЧСС);
- систолічний артеріальний тиск (САТ);
- діастолічний артеріальний тиск (ДАТ);
- середній артеріальний тиск (АТ);

- частота дихальних рухів (ЧДР);
- насичення крові киснем (SPO_2).

Тривалість кожного сигналу ПФД становить 60 хвилин, яка розбивається на 10-хвилинні вікна спостереження. Тому маємо шість підвбірок від кожного ПФД. З кожної підвбірки виділяють п'ять ознак. Ознаками (характеристиками) є середнє значення, стандартне відхилення, медіана, кількість тенденцій зростання та зменшення. Таким чином, одна вибірка містить 180 ознак (6 ПФД x 6 підвбірок x 5 ознак).

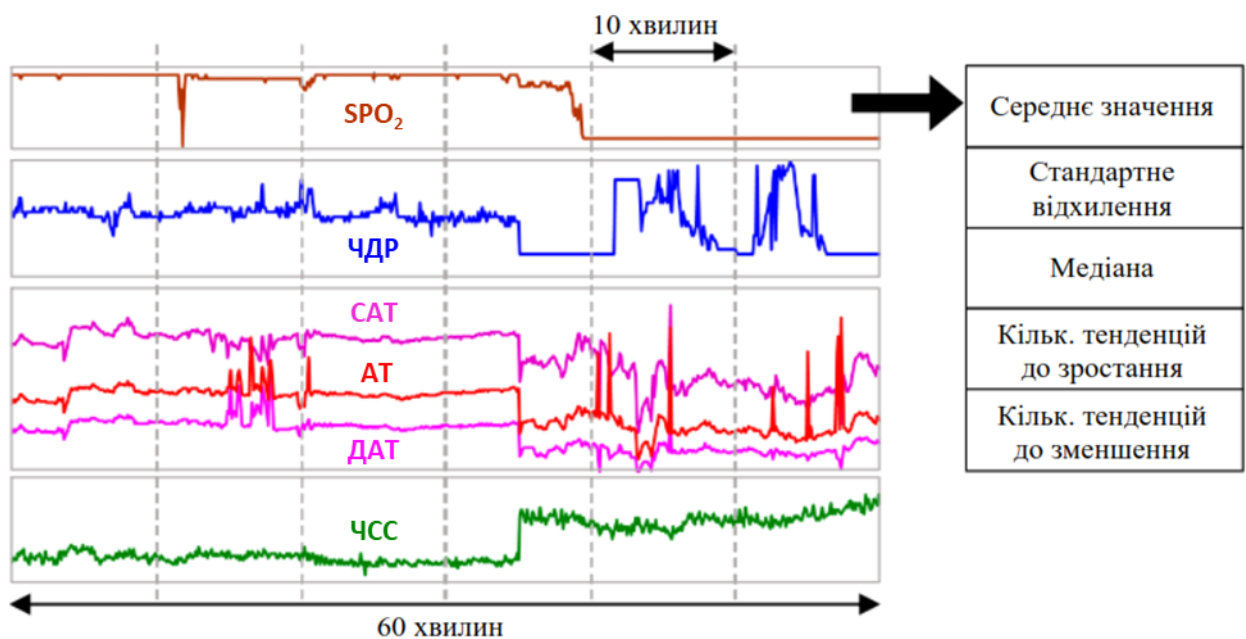


Рисунок 2.2 – Приклад вибірки за 60 хвилин

За допомогою запропонованої системи здійснюється передбачення п'яти різних подій, які перераховані в таблиці 2.1. Якщо життєво важливі ПФД виходять за межі норми (очікуваних значень), ідентифікується клінічна подія. Наприклад, NNNN означає, що всі життєво важливі сигнали знаходяться в межах норми, тоді як інші чотири клінічні події (ТНТН, ВНТН, ТТТН, ТНВН) представляють ситуації виходу життєво важливих ПФД за межі очікуваних діапазонів [28], як показано в таблиці 2.1.

Таблиця 2.1 – Цільові події для оцінки, їхні позначки та узагальнене нормальне значення життєво важливих ПФД

Подія	Мітка	ПФД			
		ЧСС	АТ	ЧДР	SPO ₂
Всі значення в межах норми	NNNN				
Одночасна тахікардія, гіпотензія, тахіпноє та гіпоксія більше 30 хв.	THTH	В >100	Н (САТ<90 і ДАТ<60) або АТ<70	В >17	Н <93%
Одночасна брадикардія, гіпотензія, тахіпноє та гіпоксія більше 30 хв.	BHTH	Н <60	Н (САТ<90 і ДАТ<60) або АТ<70	В >17	Н <93%
Одночасна тахікардія, гіпертензія, тахіпноє та гіпоксія більше 30 хв.	TTTX	В >100	В (САТ >120 і ДАТ>80) або АТ>105	В >17	Н <93%
Одночасна тахікардія, гіпотензія, брадипноє та гіпоксія більше 30 хв.	THTB	В >100	Н (САТ<90 і ДАТ<60) або АТ<70	Н <12	Н <93%

3 ШВИДКЕ НАВЧАННЯ В ПРІНГОВІЙ МЕРЕЖІ

3.1 Загальні відомості

Як було зазначено у розділі 1, вигода від використання технологій ML в сучасних медичних інформаційних системах полягає в можливості виявляти симптоми різних захворювань відповідно до життєво важливих ПФД, які збираються з різних переносних пристроїв, у визначений часовий період. Отримавши ці дискретні зразки даних, система навчається (наприклад, годину або півгодини) підбирати нові дані й таким чином може діагностувати майбутні медичні записи з більшою точністю. Однак, навіть використовуючи обчислювальну потужність хмари, для більшості алгоритмів машинного навчання все одно потрібно багато часу на навчання, що створює нові проблеми для ефективного процесу навчання в реальному часі.

Крім того, медичні установи зазвичай використовують приватні хмарні сервіси для зберігання даних і прискорення процесів навчання [29], але при цьому рідко розглядають можливість обміну даними з іншими установами. Інформаційні системи в різних установах можуть підвищити свою діагностичну точність шляхом ефективного вивчення спільних даних (чим більше даних доступно, тим точніший діагноз може поставити система).

Різноманітні дослідницькі роботи пропонують рішення щодо розумних медичних систем; однак вони страждають від деяких обмежень. Дослідники [14] аналізують різні виклики в медичних інформаційних системах, коли проблема великих даних вирішується шляхом впровадження хмарних платформ, що використовується в багатьох дослідженнях. Є пропозиція [8] домашньої системи моніторингу із використанням прихованої моделі Маркова (НММ). Слід зазначити систему для діагностики станів пацієнтів за допомогою відео- та аудіосигналів [30]. Деякі дослідники [31, 32] пропонують перспективну повсюдну систему з різними ПФД.

Передбачається [31] використання хмари, біосенсору і смартфона для спостереження за пацієнтами з хронічними захворюваннями легенів; автори розробляють застосунок для операційної системи iOS, щоб мати змогу отримувати результати аналізу з хмари. Схожа робота [32] пропонує медичну систему з програмою для смартфона та сервером, який керує інформацією про здоров'я. Застосунок для смартфона збирає дані та відображає стан здоров'я пацієнта, визначений сервером. Однак ці системи є ізольованими і не передбачають можливість обміну своїми знаннями на користь інших медичних установ.

Що стосується питань конфіденційності, то дослідники пропонують застосувати методи на основі шифрування для розробки безпечного протоколу передачі [14], використовувати блокчейн для обміну інформацією про пацієнтів у розподілений спосіб [33]. Ці методи зосереджені на домені збереження конфіденційності та не враховують особливостей навчання P2P. Крім того, методи на основі шифрування споживають багато часу для шифрування та дешифрування записів пацієнтів, обмін зашифрованими необробленими даними є неефективним, а також виникає потреба враховувати, чи мають різні медичні установи однакові права на використання необроблених даних із записів пацієнтів.

В клінічній діагностиці, системах підтримки рішень широко використовуються для виявлення аномальних явищ і симптомів захворювання різні методів машинного навчання, наприклад НММ [18, 34], машина опорних векторів (SVM) [17] і нейронні мережі (NN) [19]. Однак більшість із них є «дорогими» з точки зору обчислень. Таблиця 3.1 надає загальне порівняння з точки зору обчислювальної вартості між трьома спорідненими роботами. Незважаючи на те, що ці дослідження використовували різні набори даних і експериментальні параметри, можна помітити, що класифікатори, які використовуються в цих системах (НММ, згорточна нейронна мережа (CNN) і мережа радіальних базисних функцій (RBFN)) займають більше 15 хвилин для навчання, що стає серйозною та

критичною проблемою зі збільшенням вибірки та складними класифікаторами для навчання у великих даних.

Таблиця 3.1 – Витрати часу в практичних реалізаціях схожих робіт

Робота	Класифікатор	Кількість зразків	машина	Час навчання
[20]	HMM	600	m3.xlarge на Amazon (4 ЦП і 8 ГБ ОП)	30 хв.
[35]	CNN	606	2-ядерний ЦП і 8 ГБ ОП	понад 25 хв.
[7]	RBFN	70 000	4-ядерний ЦП Intel 2,4 ГГц і 2 ГБ ОП	понад 15 хв.

Для підвищення продуктивності узагальнення в раніше запропонованих медичних діагностичних системах [7, 36] розглянуто питання ефективності і запропоновано класифікатор на основі ELM [21]. Порівняно з іншими алгоритмами машинного навчання, ELM є швидшим у плані навчання, оскільки йому не потрібно ітераційно обчислювати параметри прихованих шарів. Таблиця 3.2 показує інше порівняння між різними дослідницькими роботами з точки зору їх точності у виявленні клінічних подій за допомогою різних класифікаторів у діагностичних медичних системах. Можна помітити, що ELM є найефективнішим з точки зору класифікації клінічних подій у наборах даних ПФД.

Щоб задовольнити вимоги обробки великих даних, різні хмарні постачальники, такі як Microsoft і Amazon, пропонують служби даних, які зберігають і шукають записи пацієнтів [29]. Однак ці рішення зосереджені на використанні переваг хмарного зберігання даних, а не на обчислювальних можливостях. У різноманітних дослідженнях для хмарних обчислень зазвичай використовується централізований алгоритм, наприклад MapReduce. Разом з тим, були запропоновані методи розподіленого навчання з ELM [37, 38] і HMM [20] з використанням MapReduce для прискорення

процесу навчання за допомогою обчислювальної потужності Cloud. Крім того, можна виділити комбіновані моделі на основі MapReduce, які дозволяють уникнути проблеми перекриття та підвищити точність класифікаторів.

Таблиця 3.2 – Дослідження алгоритмів машинного навчання для застосування в медичних інформаційних системах

Робота	Класифікатор	Набір даних	Показники	Клінічні події	Точність
[20]	HMM	MIMIC-II	ЧСС, ЧДР, SPO ₂ , САТ	5	97,7%
[17]	SVM	набір даних Оксфордської онкологічної лікарні	ЧСС, ЧДР, SPO ₂ , САТ	2 (норма або ні)	96%
[19]	нейромережа	MIMIC-II	АТ	2 (гострий епізод гіпотензії або ні)	94%
[7]	ELM	MIT-BIH аритмія	ЕКГ	6	98,72%
[39]	kNN	клівлендський набір даних про серцеві захворювання	13 ознак	7	97,4%
[35]	CNN	набір даних із центрального Китаю за 2013-2015 рр.	79 ознак	2 (інфаркт мозку або ні)	94,8%
[40]	дерево рішень	набір даних про діабет індіанців піма	8 ознак	2 (діабет 2 типу або ні)	80%
[40]	мережа Байєса	набір даних про діабет індіанців піма	8 ознак	2 (діабет 2 типу або ні)	74%

Слід зауважити, що описані вище методи сильно залежать від основних центральних серверів і мало враховують вартість передачі даних. Фреймворк MapReduce фокусується на підвищенні швидкості обробки для одного користувача (ведучого) за рахунок використання переваг розподілених обчислень або парадигми хмарних обчислень (ведені). Це означає, що кожен медичний заклад може найняти будь-яку хмарну службу з MapReduce, щоб ефективно працювати з багатьма записами пацієнтів. Але за цією структурою лише ведучий має повні знання для завдання класифікації. Крім того, він не розглядає, як ділитися знаннями між різними ведучими для підвищення точності прогнозування, а це означає, що заклади навряд чи будуть ділитися своїми власними знаннями, щоб допомогти іншим. Обмін необробленими даними між ведучими також не є ефективним методом через великі обсяги нових даних, які збираються щодня в медичному закладі.

3.2 Узагальнена модель системи

З метою подолання недоліків, описаних в роботах різних дослідників, пропонується нова система P2P-навчання з використанням метода ELM-SM. Запропонована система має здатна до повністю розподіленого навчання, що може бути корисним для інших систем, підключених до пірингової мережі. Запропонований метод на основі часткових моделей дозволяє зменшити розмір необроблених даних і прискорити процес обміну та навчання.

Отже, побудова розумної медичної систему на основі архітектурі P2P надає швидкість та ефективність з точки зору навчання та обміну інформацією. Реалізація системи відбувається шляхом генерації напівмоделей із необроблених даних з подальшим обміном ними з хмарними медичними установами. Кожен медичний заклад розглядається як рівноправний (вузол, пір) в запропонованій моделі.

Модель запропонованої системи наведено на рисунку 3.1. За основу моделі взяті існуючі реалізації [31, 32]. Розглядається сценарій, коли літній

пацієнт живе сам вдома, а його фізіологічний стан постійно контролюється віддаленою інформаційною системою медичного закладу. Різні бездротові датчики прикріплюються до тіла пацієнта та надсилають дані ПФД на смарт-пристрій, який можна носити (наприклад, смарт-годинник). Пристрій передає дані в приватну хмару, яку медичний заклад наймає для обробки. У хмарі виконуються різні алгоритми для очищення даних, навчання ELM-SM і виявлення знань. Результат навчання далі використовується для оцінювання майбутніх клінічних станів на основі нещодавно отриманих даних від пацієнтів. Клініцисти (наприклад, лікарі та медсестри) у медичному закладі можуть приймати діагностичні рішення відповідно до клінічних станів пацієнта та належним чином повідомляти пацієнта.

Систему можна розуміти як таку, що складається з двох частин:

- збір даних і прогнозування в одному вузлі;
- обмін даними між вузлами.

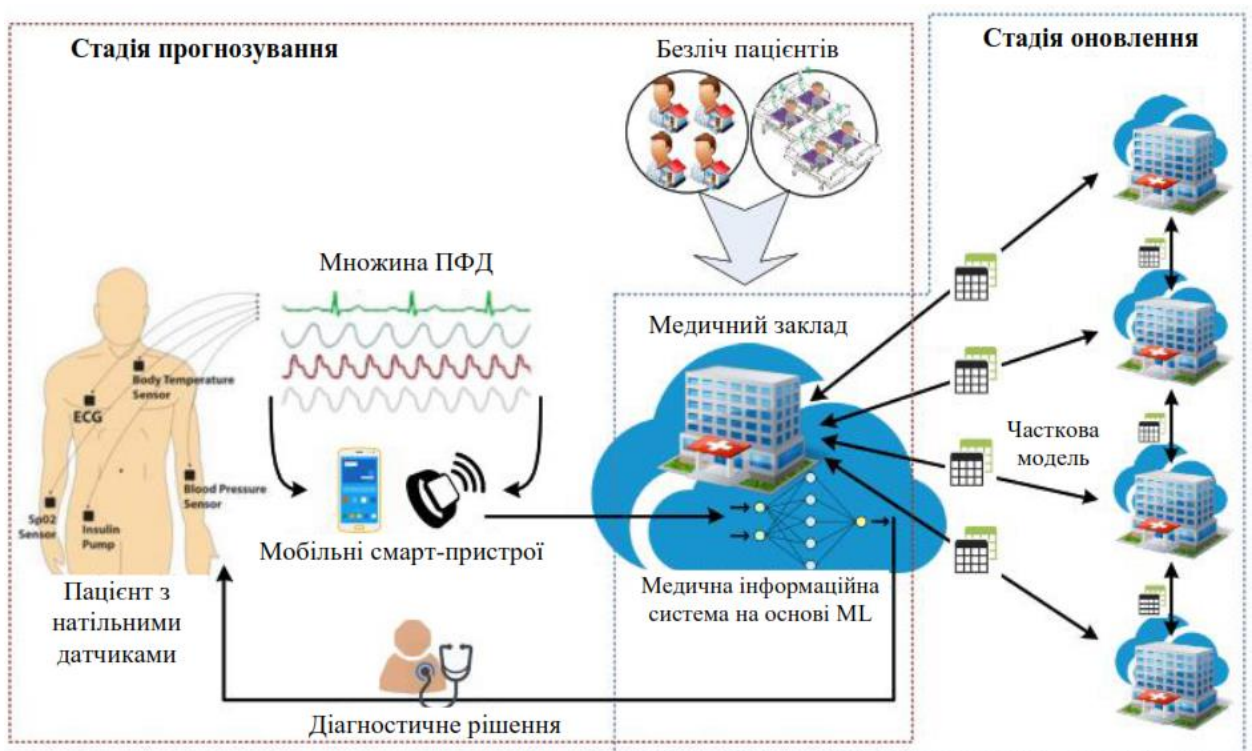


Рисунок 3.1 – Модель пропонованої системи

3.2.1 Збір даних і прогнозування в одному вузлі

У системі, зображеній на рисунку 3.1, послідовні в часі дані від сенсорів на тілі пацієнта безперервно надсилаються на його портативні пристрої IoT. Потім пристрої передають ці дані ПФД певному одноранговому пристрою для подальшої обробки. Дані зберігаються в хмарі та використовуються для навчання розумної медичної системи. Навчальний набір даних містить записи життєво важливих ПФД пацієнтів та їхні мітки (клінічні події). Крім того, ELM-SM, встановлений на кожному одноранговому пристрої, має здатність виявляти аномальні клінічні події нещодавно отриманих даних від пристроїв IoT, а потім одноранговий пристрій може повідомляти пацієнтів про будь-які подальші дії. Наприклад, медзаклад може автоматично викликати пацієнтів для запису до лікарів, якщо виявить відхилення в записах.

3.2.2 Обмін даними між вузлами

У запропонованій системі кожен одноранговий вузол має певну кількість сусідніх однорангових вузлів, які з'єднуються з ним безпосередньо, що утворює мережу P2P із певним протоколом. Рисунок 3.1 показує, що є чотири сусідні однорангові вузли, які підключаються до однорангового вузла в середині зображення. Одноранговий вузол витягує часткову модель із навчального набору даних і ділиться нею зі своїми сусідами. Приклад процесу спільного використання між сусідніми вузлами показано на рисунку 3.2. Вузол А ділиться новою частковою моделлю з сусіднім одноранговим вузлом В; вузол В ділиться нею з сусіднім вузлом С, а С – з сусіднім D.

Часткова модель визначається як легкий вихід під час процесу навчання ELM, який буде використовуватися для обчислення повної моделі. Кожен вузол відфільтровує вивчені часткові моделі від інших і припиняє ділитися ними з сусідніми вузлами. Як правило, часткова модель менша за

необроблений навчальний набір даних, що робить запропоновану систему більш ефективною в процесі обміну даними.

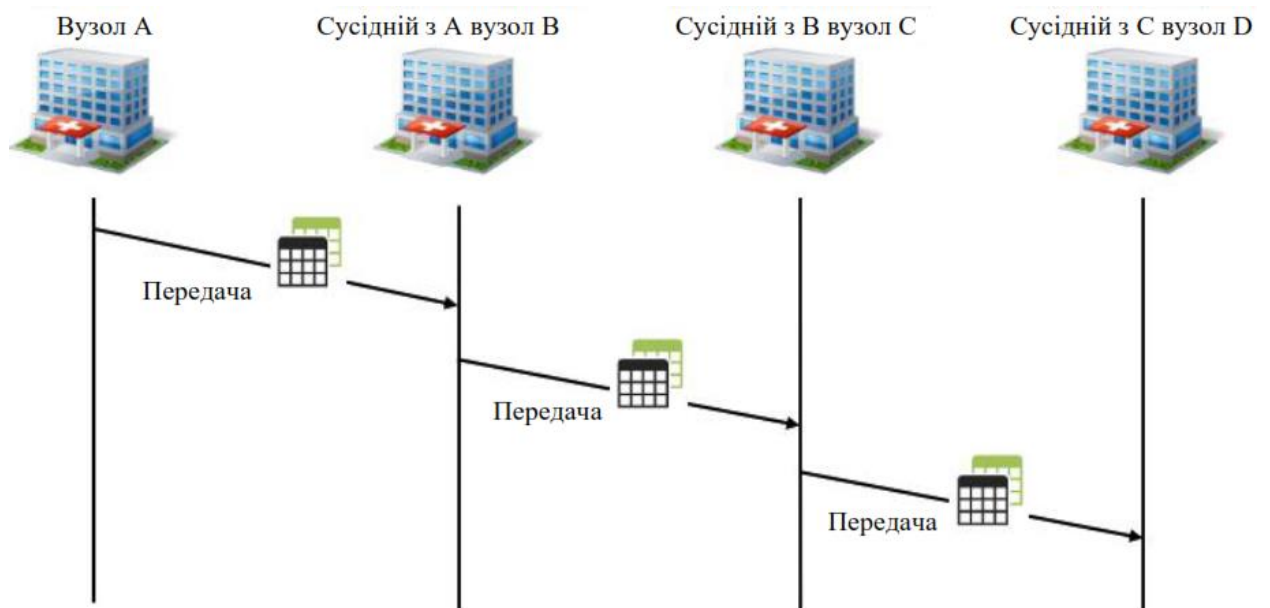


Рисунок 3.2 – Приклад обміну даними

3.1 ELM-SM

Розглянемо процес генерації часткової моделі, оновлення ELM-SM за допомогою новоствореної часткової моделі від сусіднього однорангового вузла та фільтрації різних версій часткових моделей для запобігання повторного навчання в одноранговому вузлі.

Навчальний необроблений набір даних передається у вхідний шар. Далі, у прихованому шарі виконуються певні обчислення з випадковими вхідними вагами; вихідні дані прихованого шару містять два компоненти ($C1$ і $C2$), які утворюють часткову модель. Часткова модель використовується для розрахунку повної моделі, а саме, з симетричної матриці ($H^T H$) генерується $C1$.

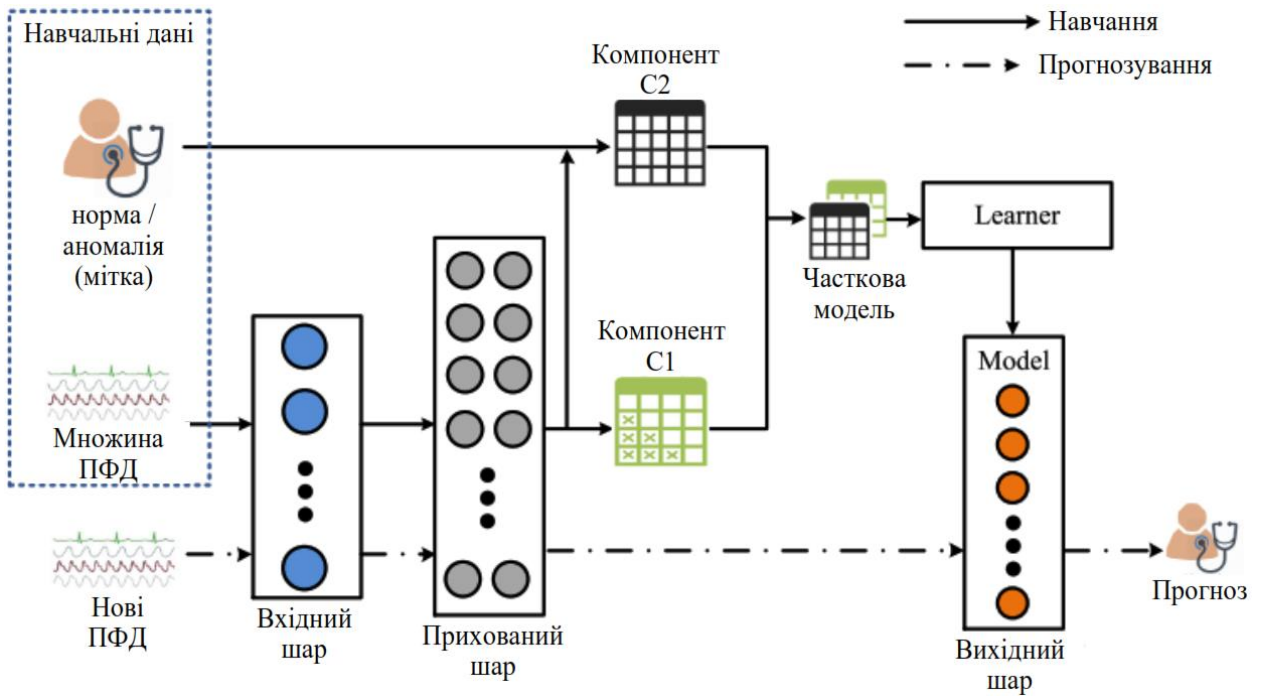


Рисунок 3.3 – Процес створення часткової моделі на основі ELM

3.1.1 Генерація часткової моделі

На рисунку 3.3 показано процес генерації часткової моделі, яка складається із двох компонентів – C1 і C2.

Компонент C1 складається з елементів результуючої матриці ($H^T H$), створеної за допомогою (2.7). Відповідно до визначення симетричної матриці маємо $u_{ij} = u_{ji}$ в

$$H^T H = \begin{bmatrix} u_{11} & \cdots & u_{1k} \\ \vdots & \ddots & \vdots \\ u_{k1} & \cdots & u_{kk} \end{bmatrix}_{K \times K}. \quad (3.1)$$

Отже, C1 визначається як $[u_{ij}]$, показано в рівнянні нижче, де $i \leq j$:

$$C1 = [u_{11} \ u_{12} \ \dots \ u_{1k} \ u_{22} \ u_{23} \ \dots \ u_{2k} \ \dots \ u_{kk}]. \quad (3.2)$$

Можна легко перетворити $C1$ в $H^T H$ з кількістю прихованих нейронів K . При цьому слід зазначити, що розмір $H^T H$, як і $C1$, залежить від кількості нейронів у прихованому шарі.

Компонент $C2$ розглядається як $H^T T$, створений на основі (2.8) з вихідною матрицею прихованого шару H і векторами міток даних T . Відповідно до (2.8), розмір $C2$ залежить не тільки від кількості прихованих нейронів, але й від кількості вихідних нейронів.

Нарешті, розраховується модель на основі часткової моделі за допомогою (2.6). Рисунок 3.4 показує приклад двох компонентів $C1$ і $C2$ у частковій моделі. Часткова модель генерується ELM із 3 прихованими нейронами та 5 цільовими виходами.

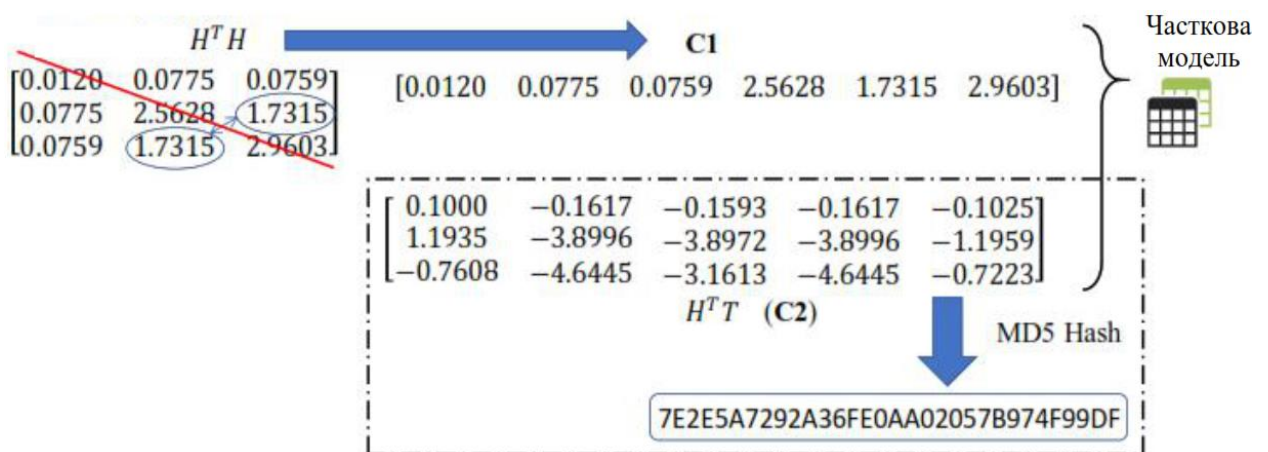


Рисунок 3.4 – Приклад часткової моделі та її значення MD5

Пропонований процес генерації часткової моделі показаний в лістингу 3.1. Кожен одноранговий вузол повинен мати однакові випадкові вхідні ваги (ω і b у (2.1)) на кроці 6. На початку матриця навчальної вибірки X та їх матриця міток T передаються у функцію GENERATE на кроці 5. Матриця міток містить вектори міток пов'язаних вибірок. Значення елементів у векторі міток дорівнює 1, коли вихідний нейрон належить до класу вибірки, а решта належать -1. Наприклад, вектор мітки (-1,-1, 1,-1,-1) означає, що існує 5

цільових класів і цей вектор належить до третього класу. Прихована вихідна матриця (H) обчислюється на основі (2.1) на кроці 6. На кроці 7 і кроці 8 обчислюються $M = H^T H$ та $C_2 = H^T T$, які показано у (2.7) і (2.8), відповідно. Нарешті, генерується часткова модель C_1 і C_2 , якою можна ділитися з іншими вузлами.

Лістинг 3.1 – Процедура формування часткової моделі

```

1: Input:  $X$  and  $T$ 
2: Output:  $C_1$  and  $C_2$ 
3: Initialize:  $w, b, H, M, C_1, C_2$ 
4:
5: procedure GENERATE( $X, T$ )
6:    $H \leftarrow X * w + b$                                 ▷ Equation (2.1)
7:    $M \leftarrow H^T * H$                                   ▷ Equation (2.7)
8:    $C_2 \leftarrow H^T * T$                                 ▷ Equation (2.8)
9:    $C_1 \leftarrow$  Elements of  $M$  where  $i \leq j$           ▷ Equation (3.2)
10:  return  $C_1, C_2$ 
11: end procedure

```

3.1.2 Часткова модель з використанням MD5

Для прискорення процесу порівняння моделі використовується значення MD5, яке розраховується з часткової моделі і застосовується у фільтрі даних. Незважаючи на те, що кожен компонент часткової моделі менший, ніж великомасштабний необроблений набір навчальних даних, порівняння двох різних компонент пов'язане з великими обчислювальними витратами. З метою пришвидшення процесу пропонується застосувати хеш-алгоритм MD5, який зазвичай використовується як контрольна сума для перевірки цілісності даних. Зміна навіть біта в даних може призвести до різниці 128-бітного хеш-коду. Оскільки кількість вихідних нейронів зазвичай набагато менша за кількість прихованих нейронів, це означає, що C_2 також стає набагато меншим за C_1 . Таким чином, C_2 використовується для обчислення хеш-коду MD5. Приклад версії часткової моделі, згенерованої алгоритмом хешування MD5, показаний на рисунку 3.4. Слід зазначити, що

для обчислення хеш-коду використовуються в кожному елементі матриці лише перші шість цифр після коми через точність числа з рухомою комою, яка прийнята в комп'ютерних засобах.

3.1.3 Процес оновлення ELM-SM на одноранговому пристрої

Рисунок 3.5 демонструє процес оновлення часткової моделі у вузлі. У пропонованій системі P2P кожен одноранговий вузол отримує часткові моделі від сусідніх вузлів у невизначений час і в невизначеному порядку. Але у кожному процесі оновлення вузол використовує лише нову часткову модель.

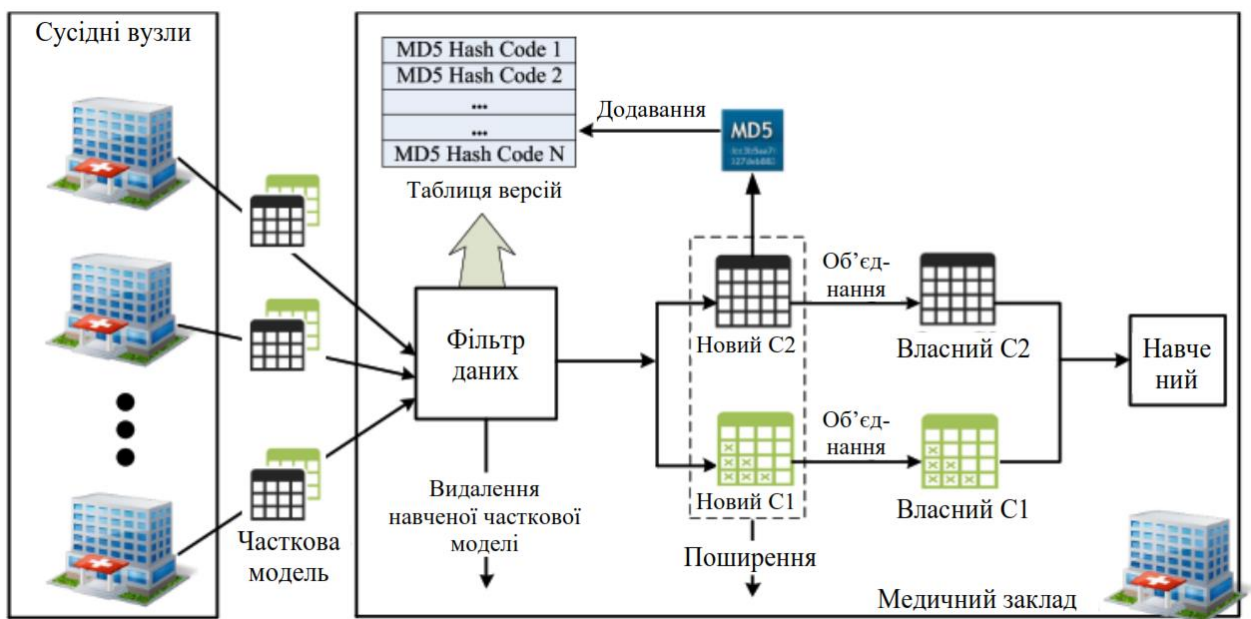


Рисунок 3.5 – Процес оновлення часткової моделі у вузлі

Перед обробкою часткової моделі вузол фільтрує вивчену часткову модель, а також припиняє надавати спільний доступ до неї своїм сусідам. Вивчені часткові моделі виявляються фільтром даних із таблицею версій. Таблиця містить версії вивчених часткових моделей, які генеруються на основі хеш-алгоритму MD5. Потім кожен компонент цієї нової часткової

моделі об'єднується в оригінальну часткову модель в ELM-SM. Нарешті, вже оновлена часткова модель використовується для обчислення нового класифікатора на основі (2.6), що демонструє рисунок 3.3.

Процес оновлення представлено в лістингу 3.2. Функція оновлення приймає на вхід нові C_1 і C_2 , згенеровані процедурою 3.1 від інших вузлів. Локальні компоненти часткової моделі представлені M_1 і M_2 , відповідно. Спочатку розраховується версія MD5 (на кроці 12). Потім до цієї версії застосовується фільтр даних, використовуючи свою таблицю, щоб з'ясувати, чи був вже навчений ELM-SM. Якщо версії не існує, ELM-SM у вузлі оновлюється новою частковою моделлю. Спочатку додаються M_1 і M_2 до нових C_1 і C_2 , відповідно (кроки 15 і 16). Потім M_1 перетворюється на матрицю M , яка використовується для обчислення нового класифікатора на основі (2.6) (крок 19), а в таблицю версій записується нова версія (крок 18).

Лістинг 3.2 – Процедура оновлення часткової моделі

```

1: Input:  $C_1$  and  $C_2$ 
2: Output:  $\beta$  of ELM
3: Initialize:  $M_1, M_2, VersionTable$ 
4:
5: procedure DATA_FILTER(newVersion)
6:   if VersionTable has newVersion then
7:     return true
8:   else
9:     return false
10:  end if
11: end procedure
12:
13: procedure UPDATE( $C_1, C_2$ )
14:   newVersion  $\leftarrow$  Calculate MD5 value of  $C_2$ 
15:   if DATA_FILTER(newVersion) then
16:     return
17:   end if
18:    $M_1 += C_1$ 
19:    $M_2 += C_2$  ▷ Equation (2.11)
20:   Transform  $M_1$  into matrix  $M$ 
21:   Add newVersion into VersionTable
22:   return  $M^{-1} * M_2$  ▷ Equation (2.6)
23: end procedure

```

3.2 Висновок за розділом

У цьому запропоновано нову систему P2P-навчання для обміну медичними знаннями, з метою створення спільноти медичних закладів у хмарному середовищі. Завдяки обраному підходу, усунуто залежність від центрального сервера. Крім того, будь-який медичний заклад може поділитися знаннями, які він уже отримав із необроблених даних, із сусідніми закладами. В результаті сусідні заклади вивчають набори даних із спільних знань, не витрачаючи багато часу на вивчення необроблених наборів даних. Таким чином, замість передачі необроблених даних ПФД в традиційних централізованих моделях навчання, запропонована система передає часткові моделі своїм сусідам, що робить її більш ефективною з точки зору навчання та обміну знаннями.

До того ж, пропонована система має явну перевагу щодо конфіденційності, оскільки перетворює необроблений набір даних у новий випадковий простір, а необроблену інформацію неможливо отримати з часткової моделі. Однак потрібно окремо проаналізувати характеристики безпеки і конфіденційності в подальшій роботі. Також доцільно провести дослідження складніших мережевих топологій, що, як очікується, дозволить покращити продуктивність запропонованої системи P2P-навчання.

4 ПРИОРИТЕЗАЦІЯ ДАНИХ

4.1 Загальні відомості

Як було зазначено вище, існуючі медичні інформаційні системи часто не враховують терміновість пацієнтів при обробці їхніх медичних даних, що призводить до затримки надання допомоги невідкладним пацієнтам, оскільки обсяг даних, зібраних від пацієнтів, є величезним, і аномальні та нормальні дані збираються разом. Для вирішення цієї проблеми доцільно визначити оптимальні схеми збору даних і прогнозування з урахуванням пріоритету даних, який може представляти терміновість пацієнтів.

Різні дослідження [20, 41, 42] для прогнозування аномалій враховують численні ПФД (наприклад, ЕКГ, артеріальний тиск, частоту серцевих скорочень, частоту дихальних рухів та насичення киснем). Але більшість з них знаходяться на теоретичному рівні і ще не придатні для широкого використання.

Запропоновані фреймворки [13, 20] прикладного рівня потребують певного удосконалення. Наприклад, один з них [13] в основному фокусується на оцінці особистого стану на основі прихованої марковської моделі (НММ). Для прийняття рішення, які дані потрібно передати, використовуються спеціальні правила. Другий досліджує хмарну структуру, щоб справлятися з тиском зберігання та обробки величезної кількості даних. Обидва зазначених приклади використовують мобільний пристрій для постійного збору та передачі необроблених даних ПФД на сервер. НММ також використовується для прогнозування клінічних подій. Ще один практичний приклад – апарат BioSign [43], який мінімізує час визначення критичної клінічної ситуації, але не підтримує прогнозування.

Наведені вище системи демонструють хороші практичні рішення медичних систем, але всі вони обробляють дані пацієнтів за принципом

«перший прийшов – перший вийшов». Однак у реальних клінічних випадках зазвичай спочатку розглядається невідкладність пацієнта, щоб визначити порядок надання медичних послуг. Прикладом може бути ефективна система сортування [4], яка використовується у відділеннях невідкладної допомоги (ED) під час прогнозування госпіталізації у відділення інтенсивної терапії або внутрішньолікарняної смертності. Прикладом хорошого засобу прогнозування результатів для пацієнтів є Національний рейтинг раннього попередження (англ., National Early Warning Score – NEWS) [44], який може надати додаткову цінність для моніторингу пацієнтів у відділенні невідкладної допомоги та в стаціонарі. Відповідно до існуючих клінічних установок, терміновість пацієнта відіграє значну роль у спостереженні пацієнтів у стаціонарі. Тому під час обробки медичних даних пацієнтів розумна система охорони здоров'я повинна враховувати цей чинник.

4.2 Огляд запропонованого методу

Розширена медична інформаційна система, в якій використовуються життєво важливі ПФД, може бути продемонстрована рисунком 4.1. У цій системі є дві частини: клієнт для збирання даних і сервер для аналізу даних.

Клієнт використовує портативні пристрої (наприклад, смартфон або смарт-годинник) для постійного збирання даних про стан пацієнта із датчиків на тілі та для оцінки пріоритету цих медичних даних на основі певного часового вікна (наприклад, 10 хвилин). Пріоритет даних представляє терміновість пацієнта, що означає те, що більш важкий пацієнт має вищий пріоритет даних. Крім того, ці медичні дані часового вікна та їхні пріоритети організовані у фрагменти даних часового вікна (англ., Time-Window Data Chunk – TWDC). Клієнт вибирає, які фрагменти даних потрібно передати на сервер, виходячи з їх пріоритетів.

Сервер використовується медичним закладом для надання точного прогнозу стану здоров'я пацієнтів на основі їхніх медичних даних, зібраних

від клієнта. Порядок передбачення даних пацієнтів оптимізується за допомогою пріоритету даних, що гарантує, що рішення щодо діагнозу надаються у відповідний час залежно від терміновості пацієнтів. Крім того, завдяки дедалі більшій кількості даних про пацієнтів, сервер може надати більш точне рішення для оцінки даних пацієнтів і прогнозування різних медичних станів, запровадивши R2P-навчання, описане вище.

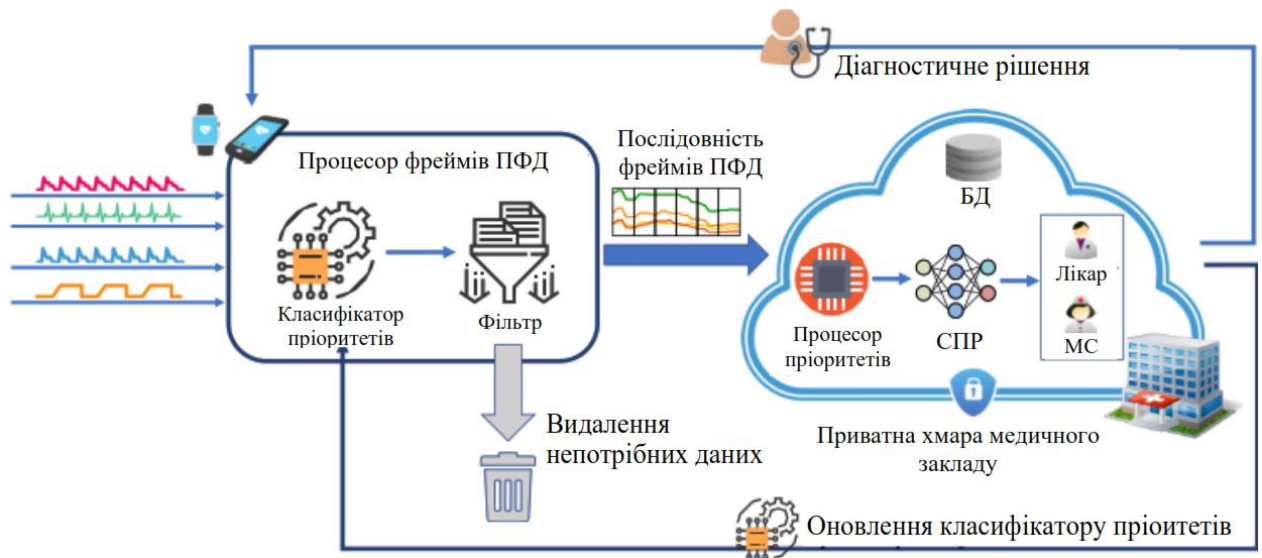


Рисунок 4.1 – Запропонована медична система з дистанційним моніторингом

Основним компонентом клієнта є процесор на основі часових вікон (англ., Time-Window Based Processor – TWBP), який складається з двох ключових частин: класифікатора пріоритетів для оцінки даних і фільтра для відбору даних.

Класифікатор пріоритетів використовує метод машинного навчання (ML) для автоматичної класифікації TWDC за різними пріоритетами. Пріоритет даних відіграє важливу роль в оптимізації системи. Класифікатор ML навчається за допомогою медичних зразків, що містять 4 життєво важливі ПФД, а мітки навчальних даних ідентифікуються алгоритмом оцінки даних на основі реальних клінічних умов.

Фільтр забезпечує ефективне збирання медичних даних пацієнтів з

урахуванням різних пріоритетів даних. У пропонованій системі аномальні дані є більш цінними та значущими, ніж «нормальні» дані, оскільки система повинна надавати відповідні діагностичні рішення під час виявлення аномальних даних. Отже, фільтр націлений на аномальні дані, які мають вищий пріоритет. Слід враховувати, що потрібно забезпечити повну медичну інформацію, яка передаватиметься на сервер для діагностики клініцистами.

На сервері основними є три компоненти: база даних для зберігання даних, пріоритетний процесор для керування чергою прогнозування та CDSS для прогнозування даних.

База даних використовується для зберігання даних для запису TWDC від різних пацієнтів. Ці записи можна використовувати для встановлення надійних довгострокових діагнозів для пацієнтів і для виявлення корисних медичних знань, які можуть покращити розумну систему охорони здоров'я.

Пріоритетний процесор допомагає організувати порядок передбачення медичних даних пацієнтів на основі пріоритету даних, що може скоротити час очікування даних термінових пацієнтів і сприяти наданню допомоги пацієнтам у відповідний час залежно від їхньої терміновості. В цій роботі пропонується алгоритм для оцінки довгострокового пріоритету даних на основі пріоритетів часових рядів TWDC. Наприклад, якщо потрібно обчислити пріоритет даних за годину, враховуються пріоритети 6 часових рядів TWDC із використанням 10-хвилинних вікон.

CDSS є машинним класифікатором, який може точно прогнозувати стан здоров'я пацієнтів, використовуючи медичні дані, зібрані від пацієнтів у режимі реального часу. Для того, щоб відповідати змінам у клінічному середовищі (наприклад, появі нових захворювань), CDSS здатний до P2P-навчання, яке сприяє підвищенню точності та ефективності діагностики.

Збір даних клієнтом реалізований у три етапи.

Етап 1. Портативний інтелектуальний пристрій постійно збирає дані про ПФД пацієнта з датчиків на тілі. Як було зазначено раніше, розглядаються 6 ПФД (ЧСС, САТ, ДАТ, АТ, ЧДР і SPO₂), які

використовуються для виявлення 5 клінічних подій.

Етап 2. Процесор (у пристрої) обробляє дані на основі певного вікна (наприклад, 10 хвилин). Дані у межах вікна визначаються як блок даних у часовому вікні (TWDC), а процесор – як процесор на основі часових вікон (TWBP). У процесорі реалізуються дві основні функції: класифікація пріоритету TWDC і фільтрація TWDC.

Етап 3. Після фільтрації процесор передає часові ряди TWDC з їх пріоритетом на сервер медичного закладу.

Сервер обробляє дані в три етапи.

Етап 1. Після отримання достатньої кількості TWDC протягом певного періода спостереження (наприклад, 60 хвилин) сервер групує ці TWDC разом і обчислює загальний пріоритет, враховуючи всі TWDC за період спостереження. Далі сервер розміщує ці згруповані дані в списку очікування на основі пріоритету; вищий пріоритет має більший індекс, що означає менший час на обробку.

CDSS прогнозує медичні стани за допомогою згрупованих TWDC і після виявлення аномальної клінічної події надсилає дані клініцистам. Потім пацієнту надаються коректні діагностичні рішення.

TWDC зберігаються в базі даних для резервного копіювання та подальшого видобутку знань. Дані також використовуються для оновлення класифікатора пріоритетів і створення медичних знань, якими можна ділитися та які покращують прогнозування діагнозів.

4.3 Життєво важливі ПФД, клінічні події та критерії оцінки терміновості

Пропонована система враховує числові дані тенденцій шести життєво важливих ПФД, наведених у таблиці 4.1, для виявлення ранніх ознак клінічного погіршення та оцінки ефекту лікування. Значення життєво важливих ПФД є різними, оскільки на ці значення впливають різні стани

(наприклад, вік і стать) пацієнтів. Для постановки базового діагнозу медична наука визначає загальний діапазон норми для кожного ПФД, наведеного в таблиці 4.1.

Таблиця 4.1 – Життєво важливі ПФД та їх нормальний діапазон

ПФД	Нормальний діапазон	ПФД	Нормальний діапазон
ЧДР	12-18 вдихів за хв.	САТ	90-120 мм рт. ст.
SPO ₂	95-100%	ДАТ	60-90 мм рт. ст.
ЧСС	60-100 ударів за хв.	АТ	60-110 мм рт. ст.

Тахікардія та брадикардія визначаються як підвищення та зниження ЧСС, відповідно. Підвищення артеріального тиску відоме як гіпертонія, а падіння – гіпотонія. Зростання та зниження ЧДР називаються тахіпное та брадипное відповідно. Дефіцит SPO₂ називається гіпоксією. Пропонована система виявляє, що ці клінічні стани виникають одночасно та тривають протягом певного періоду часу, наведеного в таблиці 2.1. Вважають ефективним підходом використовувати дані за 1 годину для прогнозування майбутньої клінічної події [20].

В системі реалізовано алгоритм контрольованого навчання для оцінки терміновості пацієнтів і досягнення належного пріоритету даних. Для присвоєння міток зразкам система використовує метод, схожий на NEWS [45]. Оскільки важко автоматично отримати рівень свідомості (LOC) пацієнтів за допомогою розумних пристроїв, система видаляє LOC з метою спрощення методу оцінки. Крім того, оскільки через слабкий внесок параметрів діастолічного артеріального тиску (ДАТ) і температури в оцінку NEWS пропонує видалити температуру, для оцінки терміновості даних пропонований метод розглядає 4 із 6 життєво важливих ПФД.

Як показано в таблиці 4.2, показниками, які беруться до уваги, є ЧДР, SPO₂, ЧСС і САТ. Відповідні бали надаються на основі їх порогових значень. Наприклад, якщо значення ЧДР становить від 9 до 11, надається оцінка 1.

Слід враховувати, що всі значення є цілими, а пороги САТ відрізняються від NEWS. Результат NEWS становить 0, коли значення САТ становить від 110 до 219, але нормальне значення САТ пацієнта завжди має бути в межах 80 і 120. Оцінка 0 для 3 інших життєво важливих ПФД означає, що значення знаходиться в межах норми. Отже, для узгодженості система змінює порогові значення САТ, якщо його значення перевищує 110.

Таблиця 4.2 – Життєво важливі ПФД, що визначають сортування

ПФД	Оцінка						
	-3	-2	-1	0	+1	+2	+3
ЧДР	≤ 8		9-11	12-20		21-24	≥ 25
SPO ₂	≤ 91	91-92	94-95	≥ 96			
ЧСС	≤ 40		41-50	51-90	91-110	111-130	≥ 131
САТ	≤ 90	91-100	101-110	110-120	121-170	171-219	≥ 220

Таблиця 4.3 – Визначення пріоритету за оцінкою життєво важливих ПФД

Пріоритет	Мітка	Час очікування	Діапазон	Загальний рахунок
1	норма	-	0	0
2	увага	≤ 1 год.	(0,1]	1-3
3	невідкладна	≤ 0,5 год.	(1,2]	4-6
4	екстрена	0	(2, 3]	≥ 7 або оцінка +3/-3

За аналогією NEWS [46], пропонується метод оцінки [47] розраховує загальний бал усіх життєво важливих ПФД і класифікує його за 4 пріоритетами. Критерії класифікації базуються на часі очікування, який визначено в NEWS. Детальне визначення всіх пріоритетів наведено в таблиці 4.3. Зокрема, якщо загальна оцінка дорівнює 0, система розглядає можливість відхилення даних, оскільки всі життєво важливі ПФД знаходяться в межах

норми. Діапазон кожного пріоритету використовується для оцінки послідовної групи пріоритетів, що пояснюється далі.

4.4 Процесор на основі часових вікон

На відміну від того, щоб постійно передавати дані від клієнта до сервера, система працює з TWDC у TWBP і вирішує, який TWDC потрібен для діагностики. Рисунок 4.2 пояснює робочий процес TWBP з чотирма даними про ПФД в часі (ЧСС, САТ, ЧДР і SPO_2). TWBP реалізує 3 етапи:

- виділяються ознаки в TWDC усіх життєво важливих ПФД;
- класифікатор пріоритетів визначає пріоритет TWDC;
- відповідно до пріоритету, фільтр вирішує, який TWDC є значущим для діагностики.

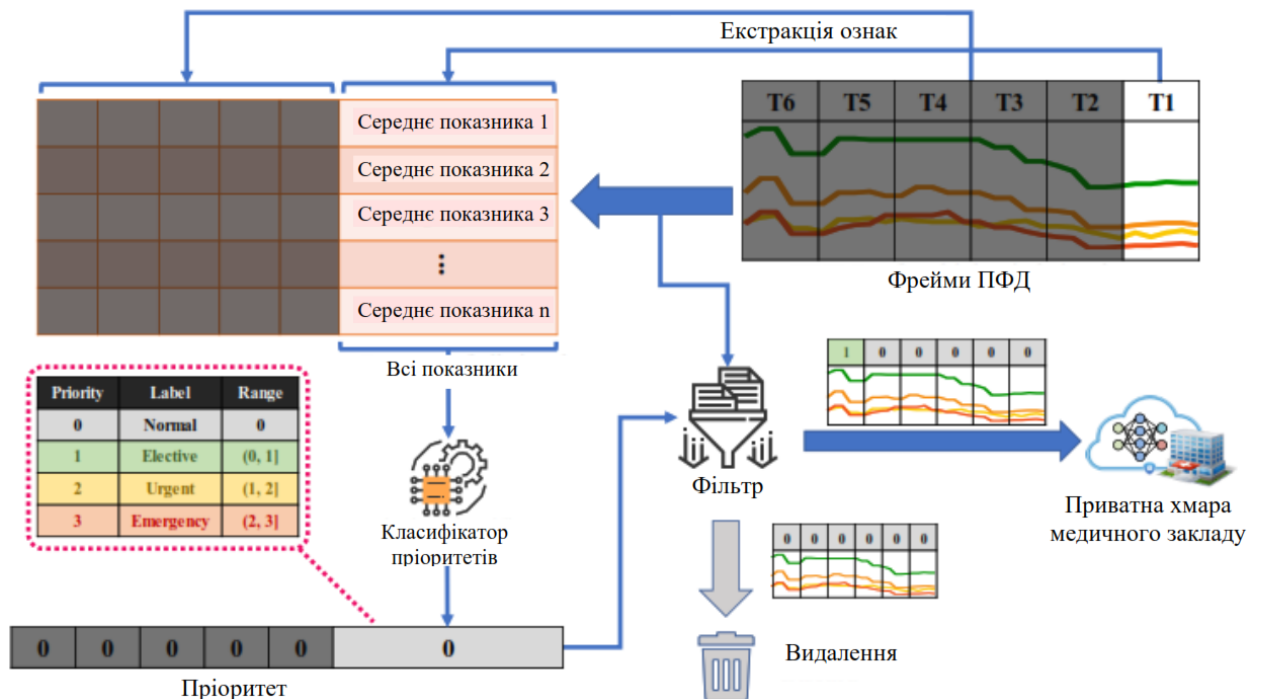


Рисунок 4.2 – Робочий процес процесора на основі часових вікон

4.4.1 Блок даних у часовому вікні і його пріоритет

Основна роль класифікатора пріоритетів у TWBP полягає в позначенні TWDC відповідним рівнем пріоритету. Нижче введено TWDC і його пріоритет, який є основою для запропонованої системи обробки даних пацієнтів.

Припустимо, що набір дискретних часових рядів даних (X) тривалості часу (T) розбивається на K вікон (W_s) однакового розміру. Є N зразків у кожному вікні (W), де N дорівнює T/K . Отже, дані часового ряду можна розглядати як послідовність TWDC – $X_1(t), X_1(t), X_1(t), \dots, X_K(t)$. Потім зразки в кожному вікні використовуються для побудови ознак. Наприклад, якщо в системі використовується 10-хвилинне вікно, 60-хвилинні дані можна розділити на послідовність із 6 TWDC (від T_1 до T_6 , показаних на рисунку 4.2). Щоб визначити пріоритет TWDC, середні значення кожного ПФД отримують із зразків у вікні:

$$f = \frac{\sum_{k=0}^{K=N} (X_k(t))}{N}. \quad (4.1)$$

Тоді всі середні значення з усіх життєво важливих ПФД групуються як вектор ознак $(f_1, f_2, f_3, \dots, f_s)$ де s – загальна кількість ПФД. Оскільки описаний вище алгоритм оцінки працює лише з цілими числами, усі компоненти вектора ознак потрібно змінити на значення числа, округленого до найближчого цілого (f'):

$$f' = \lfloor f + 0.5 \rfloor. \quad (4.2)$$

За аналогією з існуючим дослідженням [20], у запропонованій системі прийнято 10-хвилинне вікно. Після виділення ознак навчений класифікатор

пріоритетів може визначити пріоритет TWDC. Для класифікації пріоритетів можна використовувати будь-які види алгоритмів машинного навчання.

4.4.2 Збирання даних за допомогою фільтра

Фільтр у TWBP визначає суттєві медичні дані на основі свого пріоритету, що допоможе оптимізувати збір даних шляхом зменшення їхнього обсягу.

Враховуючи пріоритет 0, значення всіх ПФД у TWDC знаходяться в нормальному діапазоні. Якщо аномальний TWDC існує в буфері, потрібно передати усі послідовні TWDC для подальшого прогнозування. Рисунок 4.3 показує приклад збирання даних. Коли новий TWDC надходить у TWBP кожний період часу (t), синє буферне вікно зсувається ліворуч із збільшенням t . У кожному t відображається статус TWDC.

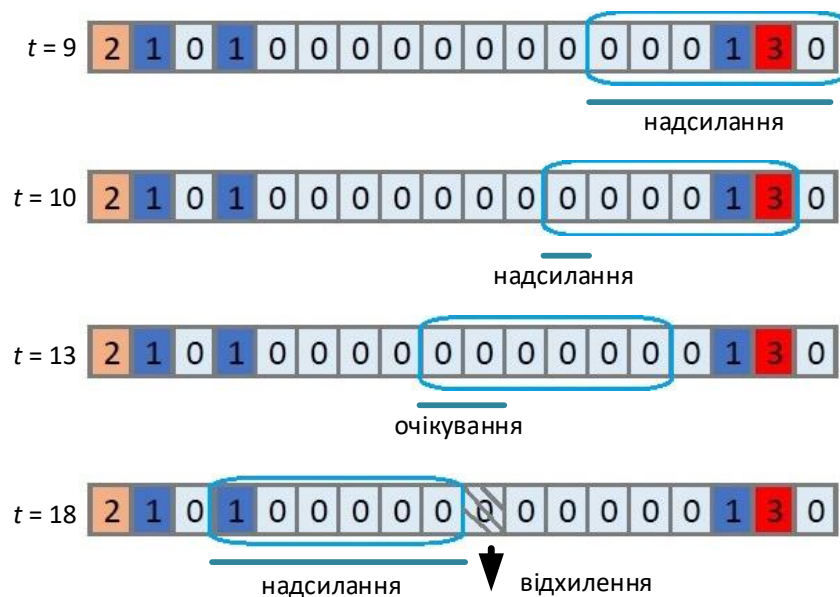


Рисунок 4.3 – Приклад передачі даних

Існує три статуси TWDC: надсилання, очікування та відхилення. Якщо пріоритет TWDC у буфері перевищує 0, ці 6 TWDC надсилаються на сервер.

Якщо пріоритети нового TWDC та решти дорівнюють 0, новий TWDC позначається як очікуваний. Зокрема, очікування TWDC означає, що системі потрібно отримати більше TWDC, щоб вирішити, надсилати чи відхилити TWDC. Коли буфер виходить із незавершеного TWDC, незавершений TWDC відкидається. Очевидно, що замість того, щоб просто надсилати TWDC один за одним, алгоритм коригує процес збирання даних на основі аномальних TWDC.

4.5 Процесор пріоритетів

За аналогією з іншим дослідженням [20], для здійснення діагностики система враховуватиме послідовність TWDC замість одного окремого TWDC. Це відбувається у процесорі пріоритетів, показаному на рисунку 4.4.

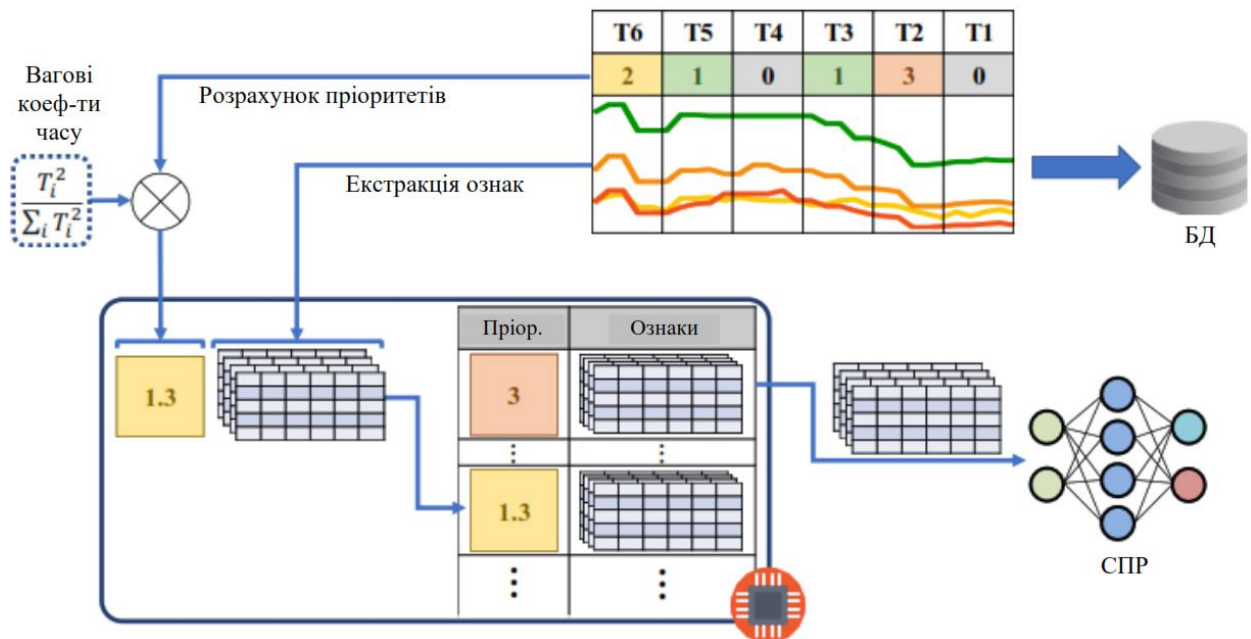


Рисунок 4.4 – Робочий процес пріоритетного процесора

Як було зазначено раніше, у вікні W використовуються всього K пріоритетів часових рядів для визначення загального пріоритету цих TWDC. Для того, щоб зберегти значущість часу, замість усереднення всіх пріоритетів

визначені вагові коефіцієнти часу

$$\omega_i = \frac{T_i^2}{\sum_{i=1}^{i=K} T_i^2} \text{ та } T_i = i * W. \quad (4.3)$$

Позначимо P як множину всіх пріоритетів, тоді $P = (p_1, p_2, \dots, p_K)$. Загальний пріоритет p' безперервних TWDC розраховується таким чином:

$$p' = P \cdot W = \sum_{i=1}^{i=K} p_i * w_i. \quad (4.4)$$

Значення p' в межах певного діапазону наведено в таблиці 4.3. Тоді система поміщає всі витягнуті функції послідовних TWDC у список очікування на основі їхнього пріоритету, як показано на рисунку 4.4. Для прогнозування вибираються дані з найвищим пріоритетом. Процес вилучення функцій описано в розділі 2. У TWDC кожного ПФД виділяється 5 ознак, якими є середнє значення, стандартне відхилення, медіана, кількість тенденцій до зростання, кількість тенденцій до зменшення. Для прогнозування майбутньої клінічної події використовується матриця ознак, яка включає всі витягнуті ознаки послідовних TWDC.

4.6 Експериментальні оцінки та результати

Для оцінки ефективності запропонованих алгоритмів використані реальні медичні дані з набору MIMIC-II [27]. Із попередньо оброблених даних було отримано набори даних для навчання та тестування. Використовуються чотири життєво важливі ПФД (ЧСС, САТ, ЧДР і SpO_2) для оцінки чотирьох пріоритетів даних, що представляють терміновість пацієнтів. Продуктивність запропонованої системи аналізується за

допомогою трьох вимірювань:

- точність класифікації пріоритетів: показує продуктивність трьох різних класифікаторів, які класифікують дані за чотирма пріоритетами;
- ефективність збирання даних: використовується для оцінки продуктивності збирання даних у запропонованій системі з точки зору обсягу даних і частоти їх надсилання. Експерименти проводяться з 10 різними співвідношеннями (від 0,1 до 1) аномальних даних у системі;
- середній час очікування: використовується для вимірювання продуктивності в черзі очікування. Пріоритети даних часових рядів генеруються з використанням двох різних дискретних розподілів: рівномірного розподілу та нормального розподілу.

4.6.1 Попередня обробка даних

Щоб оцінити точність призначення пріоритетів даних, розглянемо 6 життєво важливих ПФД із числового набору даних MIMIC-II фізіобанку МІТ, публічного набору даних, що містить велику кількість даних пацієнтів. Приймаються лише записи, що містять принаймні 24-годинні дані чисельного тренду цих 6 ПФД. Більшість ПФД знімаються протягом однієї хвилини. Дані, що збираються за секунду, перетворюються на вибірку за хвилину шляхом усереднення всіх значень за хвилину. Дані, у яких відсутні значення протягом тривалого періоду, і дані з шумом також фільтруються. Нарешті, для експериментів отримано 1023 записи.

Як зазначалося в попередньому розділі, для визначення пріоритету використовуються 4 із 6 життєво важливих ПФД – ЧСС, САТ, ЧДР і SpO_2 . У кожному ПФД усереднюються всі значення протягом 10-хвилинного вікна, тому кожен зразок має 4 середні значення. Потім застосовується алгоритм описаний вище для надання міток всім зразкам. Щоб збалансувати набір даних, випадковим чином вибираються 1500 зразків кожного пріоритету. Враховуючи, що загалом є 4 пріоритети, у наборі даних є 6000 зразків. Набір

даних нормалізується за допомогою лінійного перетворення z-показника. 70% вибірок становлять набір даних навчання, а решта – набір даних тестування, які наведені в таблиці 4.4.

Таблиця 4.4 – Попередньо оброблені дані з 4 пріоритетами даних

Пріоритет даних	0	1	2	3
Тип	нормальний	аномальний		
Кількість навчальних проб	1050	1050	1050	1050
Кількість досліджуваних зразків	450	450	450	450
Всього	1500	1500	1500	1500

4.6.2 Класифікація за пріоритетністю

Точність різних класифікаторів нейронної мережі наведено в таблиці 4.5. Зокрема, класифікатори, засновані на дереві рішень, не розглядаються в експериментах, оскільки вони добре класифікують мітки даних, створені з правил. У цьому тесті вони можуть забезпечити понад 99% точності. В системі MATLAB реалізована ELM [21], до складу якої входить 500 прихованих нейронів із сигмоїдною функцією активації. Класифікатори результатів запускаються в Weka 3.8 із налаштуваннями за замовчуванням. Використано ядро RBF в SMO.

Таблиця 4.5 – Порівняння точності різних алгоритмів навчання

	Багатошаровий песептрон	SMO	ELM
Точність, %	71,89	66,83	80,6
Час навчання, с	2,82	13,53	1,09

ELM демонструє найкращу точність 80,6% серед усіх потенційних

класифікаторів нейронних мереж. Матриця плутанини класифікації показана в таблиці 4.6. Порівняно з базовим рівнем (25%), ELM демонструє значне покращення точності класифікації.

Таблиця 4.6 – Матриця плутанини після виконання класифікації

	норма	увага	невідкладна	екстрена
норма	445	136	10	1
увага	4	248	18	8
невідкладна	1	66	380	63
екстрена	0	0	42	378

Крім загальної точності та матриці плутанини, для кожного пріоритету застосовуються різні показники точності (достовірність, чутливість і специфічність), показані в таблиці 4.7. З цього спостереження видно, що класифікація не чутлива до даних із вибіркоvim пріоритетом.

Таблиця 4.7 – Вимірювання ефективності кожного пріоритету за допомогою ELM

	норма	увага	невідкладна	екстрена
Точність, %	75,169	88,889	74,510	90,000
Чутливість, %	98,670	55,111	84,444	84,000
Специфічність, %	89,111	97,705	90,377	96,891

4.6.3 Ефективність збирання даних

У цьому експерименті використано Python з метою симулювати процес передачі даних. Для спрощення симуляції, вважаємо пріоритет від 1 до 3 ненормальним пріоритетом 1. На основі 10 різних аномальних коефіцієнтів даних від 0,1 до 1 створюється список із 6000 двійкових значень.

Співвідношення аномальних даних розраховується як

$$R_{\text{abnormal}} = \frac{\text{Кількість "1" у списку}}{\text{Довжина списку}}. \quad (4.5)$$

Коефіцієнт передачі даних розраховується як

$$P_{\text{data}} = \frac{\text{Кількість надісланих даних}}{\text{Довжина списку}}. \quad (4.6)$$

Коефіцієнт запитів на надсилання даних розраховується як

$$F_{\text{sr}} = \frac{\text{Кількість надісланих запитів}}{\text{Довжина списку}}. \quad (4.7)$$

Рисунок 4.5 пояснюється у такий спосіб. Якщо аномальними є 10% TWDC, то можна відкинути понад 20%. Коли аномальне співвідношення становить 40%, передаються 99,8% TWDC. Якщо в списку є понад 40% аномальних TWDC, необхідно передати всі TWDC.

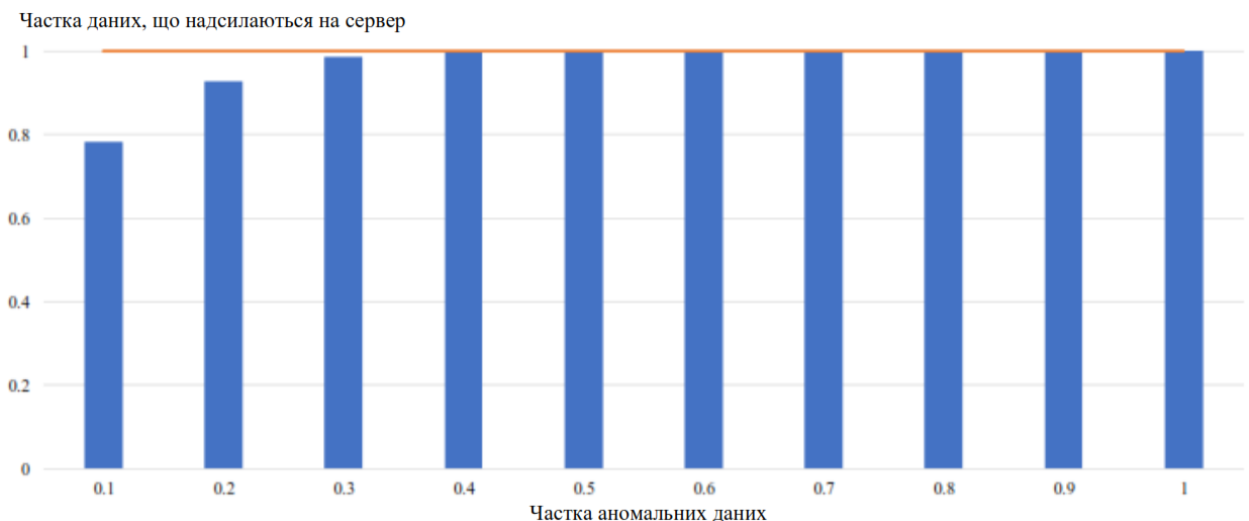


Рисунок 4.5 – Частка даних, надісланих на сервер, з різними аномальними співвідношеннями

Подібно до відсотка передачі даних, коли список містить 10% аномальних TWDC, більше половини запитів на надсилання даних можна зберегти, оскільки система групує послідовні TWDC і надсилає один раз. І 0,2% запитів зберігаються з 60% аномальним співвідношенням даних. Якщо в списку понад 60% аномальних TWDC, кожен TWDC надсилається на сервер один за одним, що означає, що співвідношення надсилання запитів дорівнює 1.

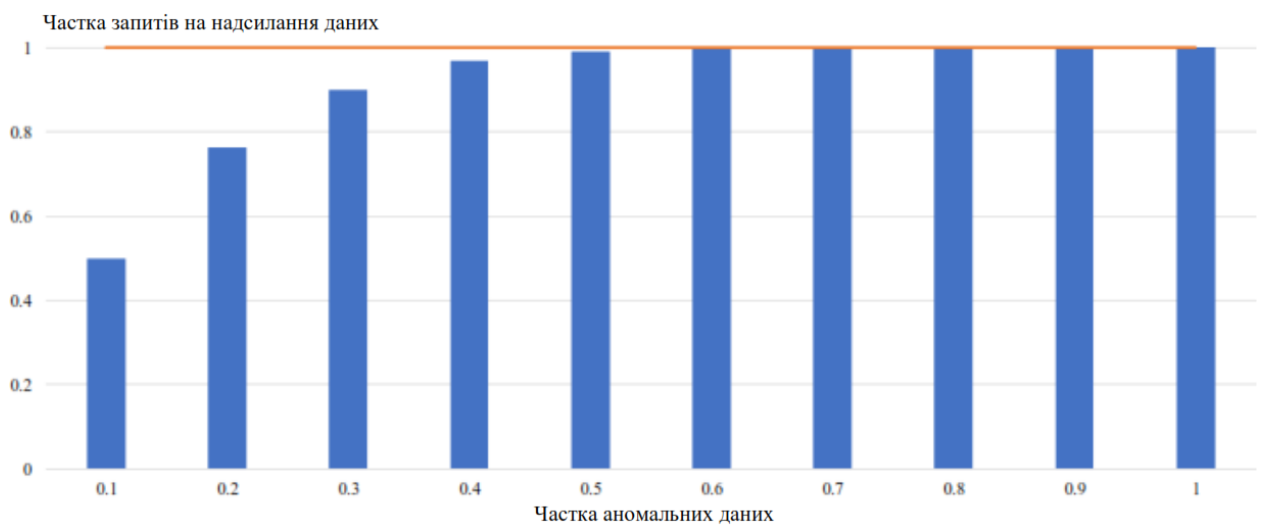


Рисунок 4.6 – Частка запитів на надсилання даних з різною часткою аномалій

4.6.4 Середній час очікування

Реалізовано експеримент для оцінки часу очікування записів пацієнтів перед прогнозуванням. Припускаємо, що список очікування може містити Q записів для подальшого прогнозування. Після завершення прогнозування всіх Q -записів нові Q -записи надходять до списку очікування. CDSS щоразу споживає лише 1 запис і витрачає t на прогнозування запису. Симуляція виконана на Python. Генеруються 3000 записів із 3 аномальними пріоритетами з рівномірного розподілу та стандартного нормального розподілу відповідно. Таким чином, є $3000/Q$ пакетів для прогнозування, а

час очікування обчислюється шляхом усереднення часу очікування записів із різними пріоритетами в усіх пакетах. У разі відсутності пріоритету система обробляє всі дані за принципом «першим прийшов – першим вийшов» (FIFO). В іншому випадку система впорядковує дані в списку очікування відповідно до їх пріоритетів, а потім обробляє їх один за іншим.

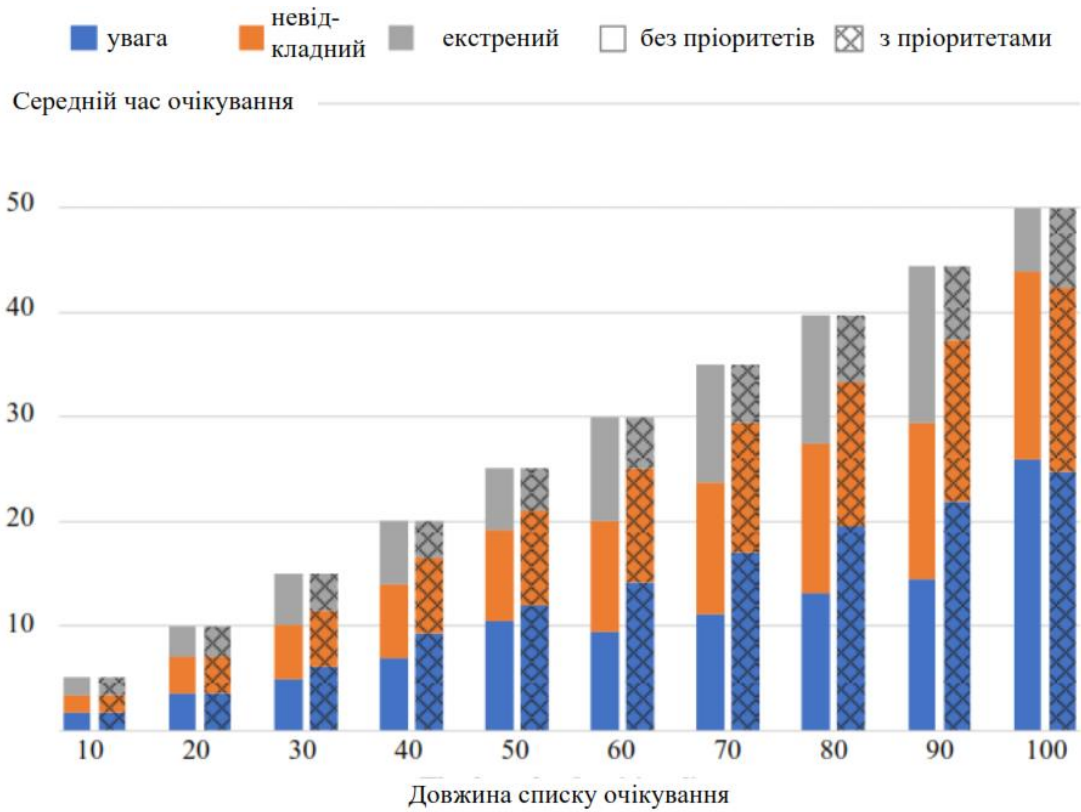
Результати з різними розподілами даних показані на рисунку 4.7. З рисунків можна помітити, що коли список очікування стає довшим, система збільшує час очікування даних із нижчим пріоритетом (Elective).

На рисунку 4.7(а), час очікування без урахування пріоритету коливається зі збільшенням довжини списку очікування, тоді як результати з пріоритетом є стабільними незалежно від довжини списку очікування. На рисунку 4.7(б), порівняно з результатами без урахування пріоритету даних, запропонований метод може значно скоротити час очікування термінових даних.

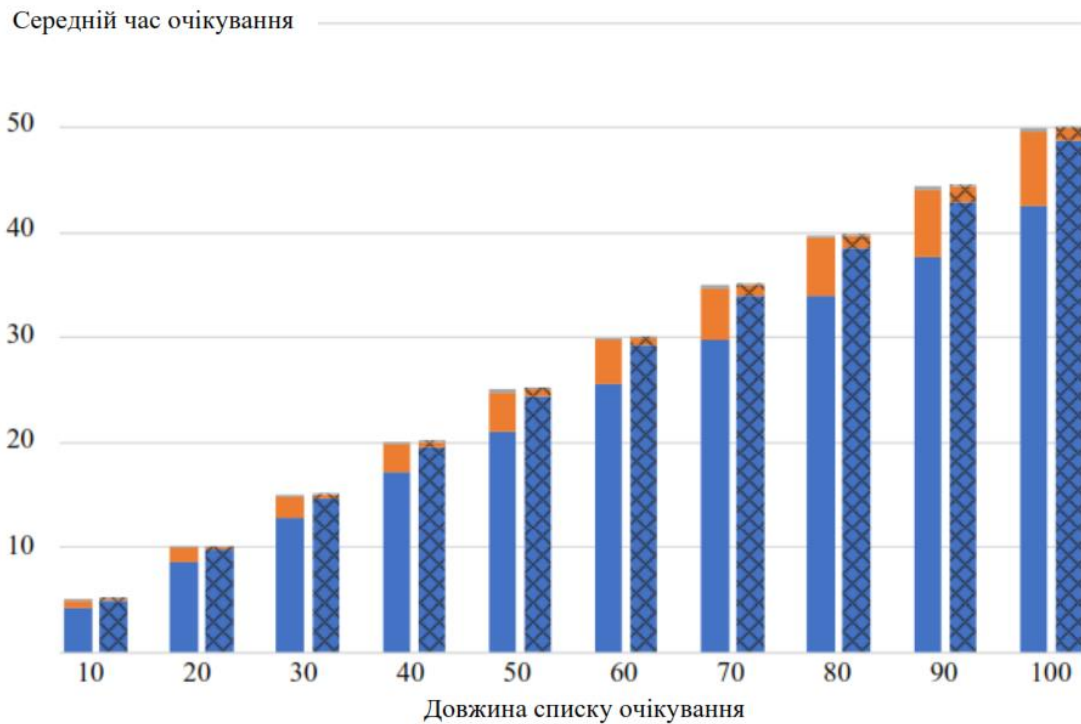
Таблиця 4.8 демонструє порівняння часу очікування системи з пріоритетом і без нього, коли довжина списку очікування становить 10 і всі пріоритети даних відповідають нормальному розподілу. Припускаємо, що система та клініцисти витрачають 10 хвилин на діагностику кожного з даних. Можна помітити, що час очікування всіх даних з різним пріоритетом менший, ніж максимальний час очікування з клінічного визначення NEWS. Однак система може оптимізувати потік обробки даних на основі пріоритету. Завдяки запропонованому методу, можна скоротити час очікування термінових даних на 75%.

Таблиця 4.8 – Порівняння часу очікування

	норма	увага	невідкладна	екстрена
NEWS	-	< 1 год.	< 0,5 год.	0
CDSS з пріоритетом	-	10 хв.	1 хв.	0
CDSS без пріоритету	-	26 хв.	4 хв.	0,2 хв.



(а)



(б)

Рисунок 4.7 – Час очікування записів із пріоритетом і без нього: (а) рівномірний розподіл пріоритетів; (б) нормальний розподіл пріоритетів

4.7 Висновки за розділом

В цьому розділі представлено медичну інформаційну систему з оцінкою пріоритету даних, що дозволяє оптимізувати збирання даних і прогнозування щодо розміру даних і часу очікування. На основі аналізу реальних клінічних умов розроблено нові алгоритми для оцінювання пріоритетів даних, які можуть представляти терміновість пацієнтів. Завдяки цьому не тільки збирається менше даних від пацієнтів, але також надається повна медична інформація, необхідна для довгострокового встановлення точного діагнозу. Розглядаючи різні розподіли пріоритетів даних у реальному випадку, експерименти показують, що запропонований метод здатен підвищити ефективність збирання даних і оптимізувати час очікування відповідно до терміновості пацієнтів. У майбутньому можна розробити персоналізовану оцінку даних із усвідомленням контексту, яка може забезпечити точнішу оцінку особистої інформації [47].

ВИСНОВКИ

В роботі досліджено новітні системи та алгоритми, натхненні різними техніками, пов'язаними з комп'ютерними мережами, машинним навчанням, великими даними та розподіленими обчисленнями.

Багатьма медичними установами використовуються інформаційні системи на основі ML, які постійно збирають дані показників фізіологічної діяльності від пацієнтів і виявляють їхній клінічний статус (наприклад, захворювання) за допомогою різних класифікаторів на основі машинного навчання. Щоб підвищити точність діагностики, класифікатори потрібно часто навчати. Пропонована система P2P-навчання сприяє підвищенню ефективності навчання та точності діагностики з великою кількістю даних пацієнтів.

Надійна, ефективна для навчання та обміну система дозволяє кожному медичному закладу ділитися своїми знаннями з іншими. За винятком процесу навчання та обміну інформацією, потоки обробки збору даних і прогнозування мають значний вплив на систему. Замість безперервного надсилання часових рядів даних пацієнтів, дані організуються та передаються в певному часовому вікні на основі терміновості пацієнтів, що називається блоком даних часового вікна (TWDC). Кожному TWDC надається пріоритет, який оцінюється запропонованим алгоритмом та представляє терміновість пацієнта. Дані з вищим пріоритетом обробляються перед тими з нижчим пріоритетом, що оптимізує очікування даних перед прогнозуванням. Запропонований алгоритм натхненний системою NEWS. У пропонованій системі лише цінні TWDC для аналізу, які визначаються пріоритетом даних, надсилаються до медичного закладу, що може зменшити обсяг переданих даних.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Свірщевський К.О. Інтелектуальна система формування діагнозу стану здоров'я користувача // Проблеми інформатизації. Тези доповідей десятої міжнародної науково-технічної конференції, Черкаси: ЧДТУ, – 2022. – С. 86
2. V. Midha, M. N. Ngafeeson, K. Ghosh, M. Ngafeeson, J. Sun, M. Merhi, K. Koong, L. Lui, and M. Yunis. Healthcare it: Current research and future trends. *Int. J. Electronic Healthcare*, 9(2/3):89, 2023.
3. K. Dunn Lopez, S. M. Gephart, R. Raszewski, V. Sousa, L. E. Shehorn, and J. Abraham. Integrative review of clinical decision support for registered nurses in acute care settings. *Journal of the American Medical Informatics Association*, 24(2):441-450, 2017.
4. L. Catarinucci, D. de Donno, L. Mainetti, L. Palano, L. Patrono, M. L. Stefanizzi, and L. Tarricone. An iot-aware architecture for smart healthcare systems. *IEEE Internet of Things Journal*, 2(6):515-526, Dec 2021.
5. M. S. Report. Wearable computing devices, like apple's iwatch, will exceed 485 million annual shipments by 2018. ABI Research, Feb 2013.
6. W. C. Willett, J. P. Koplan, R. Nugent, C. Dusenbury, P. Puska, and T. A. Gaziano. Prevention of chronic disease by means of diet and lifestyle changes. In D. T. Jamison, J. G. Breman, A. R. Measham, G. Alleyne, M. Claeson, D. B. Evans, P. Jha, A. Mills, and P. Musgrove, editors, *Disease Control Priorities in Developing Countries*, chapter 44, pages 833-850. The International Bank for Reconstruction and Development/The World Bank, 2006.
7. J. Kim, H. S. Shin, K. Shin, and M. Lee. Robust algorithm for arrhythmia classification in ecg using extreme learning machine. *Biomedical engineering online*, 8(1):31, 2009.
8. A. R. M. Forkan and I. Khalil. A probabilistic model for early prediction of abnormal clinical events using vital sign correlations in home-based monitoring.

In Pervasive Computing and Communications (PerCom), 2016 IEEE International Conference on, pages 1-9. IEEE, 2016.

9. W. Huang, N. Li, Z. Lin, G. Huang, W. Zong, J. Zhou, and Y. Duan. Liver tumor detection and segmentation using kernel-based extreme learning machine. In 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBC 2013, Osaka, Japan, July 3-7, 2013, pages 3662-3665, 2013.

10. W. Kurschl, S. Mitsch, and J. Schoenboeck. Modeling situation-aware ambient assisted living systems for eldercare. In Information Technology: New Generations, 2009. ITNG'09. Sixth International Conference on, pages 1214-1219. IEEE, 2009.

11. A. Pantelopoulos and N. G. Bourbakis. A survey on wearable sensor-based systems for health monitoring and prognosis. *IEEE Trans. Systems, Man, and Cybernetics, Part C*, 40(1):1-12, 2020.

12. . A. Beyer and D. Laney. The importance of big data: a definition. Stamford, CT: Gartner, pages 2014-2018, 2012.

13. P. Jiang, J. Winkley, C. Zhao, R. Munnoch, G. Min, and L. T. Yang. An intelligent information forwarder for healthcare big data systems with distributed wearable sensors. *IEEE Systems Journal*, 10(3):1147-1159, 2016.

14. F. Touati and R. Tabish. U-healthcare system: State-of-the-art review and challenges. *Journal of Medical Systems*, 37(3):9949, May 2013.

15. J. Erickson. Database technologies: Concepts, methodologies, tools, and applications: Concepts, methodologies, tools, and applications. volume 1, pages 157-164. IGI Global, 2009.

16. A. Gepperth and B. Hammer. Incremental learning algorithms and applications. In European Symposium on Artificial Neural Networks (ESANN), Bruges, Belgium, 2016.

17. L. Clifton, D. A. Clifton, P. J. Watkinson, and L. Tarassenko. Identification of patient deterioration in vital-sign data using one-class support vector machines. In Computer Science and Information Systems (FedCSIS), 2011

Federated Conference on, pages 125-131. IEEE, 2011.

18. F. J. Ordonez, P. de Toledo, and A. Sanchis. Activity recognition using hybrid generative/discriminative models on home environments using binary sensors. *Sensors*, 13(5):5460-5477, 2013.

19. J. Henriques and T. Rocha. Prediction of acute hypotensive episodes using neural network multi-models. In *Computers in Cardiology*, 2009, pages 549-552. IEEE, 2009.

20. A. R. M. Forkan and I. Khalil. Peace-home: Probabilistic estimation of abnormal clinical events using vital sign correlations for reliable home-based monitoring. *Pervasive and Mobile Computing*, 38:296-311, 2017.

21. G.-B. Huang, H. Zhou, X. Ding, and R. Zhang. Extreme learning machine for regression and multiclass classification. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, 42(2):513-529, 2012.

22. Y. L. Zheng, X. R. Ding, C. C. Y. Poon, B. P. L. Lo, H. Zhang, X. L. Zhou, G. Z. Yang, N. Zhao, and Y. T. Zhang. Unobtrusive sensing and wearable devices for health informatics. *IEEE Transactions on Biomedical Engineering*, 61(5):1538-1554, May 2014.

23. N. Liang, G. Huang, P. Saratchandran, and N. Sundararajan. A fast and accurate online sequential learning algorithm for feedforward networks. *IEEE Trans. Neural Networks*, 17(6):1411-1423, 2006.

24. J. Xin, Z. Wang, L. Qu, and G. Wang. Elastic extreme learning machine for big data classification. *Neurocomputing*, 149:464-471, 2015.

25. W. Yan. One-class extreme learning machines for gas turbine combustor anomaly detection. In *Neural Networks (IJCNN)*, 2016 International Joint Conference on, pages 2909-2914. IEEE, 2016.

26. Q. Leng, H. Qi, J. Miao, W. Zhu, and G. Su. One-class classification with extreme learning machine. *Mathematical problems in engineering*, 2015, 2015.

27. M. Saeed, M. Villarroel, A. T. Reisner, G. Clifford, L.-W. Lehman, G. Moody, T. Heldt, T. H. Kyaw, B. Moody, and R. G. Mark. Multiparameter

intelligent monitoring in intensive care ii (mimic-ii): a public-access intensive care unit database. *Critical care medicine*, 39(5):952, 2011.

28. A. Budiman, M. I. Fanany, and C. Basaruddin. Constructive, robust and adaptive os-elm in human action recognition. In *Industrial Automation, Information and Communications Technology (IAICT), 2014 International Conference on*, pages 39-45. IEEE, 2014.

29. S. P. Ahuja, S. Mani, and J. Zambrano. A survey of the state of cloud computing in healthcare. *Network and Communication Technologies*, 1(2):12, 2012.

30. M. S. Hossain. Cloud-supported cyberphysical localization framework for patients monitoring. *IEEE Systems Journal*, 11(1):118-127, March 2017.

31. N. Ravikumar, N. H. Metcalfe, J. Ravikumar, and R. Prasad. Smartphone applications for providing ubiquitous healthcare over cloud with the advent of embeddable implants. *Wireless Personal Communications*, 86(3):1439-1446, Feb 2016.

32. S. Kim, S. Yeom, O. Kwon, D. Shin, and D. Shin. Ubiquitous healthcare system for analysis of chronic patients' biological and lifelog data. *IEEE Access*, 6:8909-8915, 2018.

33. C. McFarlane, M. Beer, J. Brown, and N. Prendergast. *Patientory: A healthcare peer-to- peer emr storage network v1.1*, 2017.

34. I. Stanculescu, C. K. Williams, and Y. Freer. Autoregressive hidden markov models for the early detection of neonatal sepsis. *IEEE journal of biomedical and health informatics*, 18(5):1560-1570, 2014.

35. M. Chen, Y. Hao, K. Hwang, L. Wang, and L. Wang. Disease prediction by machine learning over big data from healthcare communities. *IEEE Access*, 5:8869-8879, 2017.

36. L. Shang, J.-G. Wang, W.-S. Yao, and S.-F. Chen. Classification approach based on evolutionary neural networks. *Ruan Jian Xue Bao(Journal of Software)*, 16(9):1577-1583, 2005.

37. J. Xin, Z. Wang, L. Qu, G. Yu, and Y. Kang. A-elm: Adaptive

distributed extreme learning machine with mapreduce. *Neurocomputing*, 174:368-374, 2016.

38. X. Bi, X. Zhao, G. Wang, P. Zhang, and C. Wang. Distributed extreme learning machine with kernels based on mapreduce. *Neurocomputing*, 149:456-463, 2015.

39. M. Shouman, T. Turner, and R. Stocker. Applying k-nearest neighbour in diagnosing heart disease patients. *International Journal of Information and Education Technology*, 2(3):220, 2012.

40. N. Sarma, S. Kumar, and A. K. Saini. A comparative study on decision tree and bayes net classifier for predicting diabetes type 2. In *ICRTIET-2014*, 2014.

41. H. Cao, L. Eshelman, N. Chbat, L. Nielsen, B. Gross, and M. Saeed. Predicting icu hemodynamic instability using continuous multiparameter trends. In *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, pages 3803-3806. IEEE, 2008.

42. O. Salem, Y. Liu, A. Mehaoua, and R. Boutaba. Online anomaly detection in wireless body area networks for reliable healthcare monitoring. *IEEE J. Biomedical and Health Informatics*, 18(5):1541-1551, 2014.

43. L. Tarassenko, A. Hann, A. Patterson, E. Braithwaite, K. Davidson, V. Barber, and D. Young. Biosign : multi-parameter monitoring for early warning of patient deterioration. pages 71 - 76, 12 2005.

44. C. Barfod, M. M. P. Lauritzen, J. K. Danker, G. Soletormos, J. L. Forberg, P. A. Berlac, F. Lippert, L. H. Lundstr0m, K. Antonsen, and K. H. W. Lange. Abnormal vital signs are strong predictors for intensive care unit admission and in-hospital mortality in adults triaged in the emergency department-a prospective cohort study. *Scandinavian journal of trauma, resuscitation and emergency medicine*, 20(1):28, 2012.

45. N. Alam, I. Vegting, E. Houben, B. van Berkel, L. Vaughan, M. Kramer, and P. Nanayakkara. Exploring the performance of the national early warning score (news) in a european emergency department. *Resuscitation*, 90:111-115,

2015.

46. N. C. E. Committee et al. National early warning score national clinical guideline no. 1. 2013.

47. Росінський Д.М., Свірщевський К.О. Оптимізація збору оперативних даних в піринговій мережі медичної установи // «Системи управління навігації та зв'язку», – Випуск 2 (72), – Полтава, Національний університет “Полтавська політехніка імені Юрія Кондратюка”, – 2023. – С. 142-145.