

Міністерство освіти і науки України
Чорноморський національний університет імені Петра Могили

Кваліфікаційна наукова
праця на правах рукопису

Гончаров Денис Сергійович

УДК 004.67:004.738.5;616-01/09

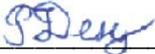
ДИСЕРТАЦІЯ

**«МОДЕЛІ ТА МЕТОДИ ПОБУДОВИ СПЕЦІАЛІЗОВАНИХ СИСТЕМ
МОНІТОРИНГУ СТАНУ ЗДОРОВ'Я ЛЮДИНИ»**

Спеціальність 123 – Комп'ютерна інженерія
Галузь знань 12 Інформаційні технології

Подається на здобуття наукового ступеня доктора філософії

Дисертація містить результати власних досліджень. Використання ідей, результатів і текстів інших авторів мають посилання на відповідне джерело.

 Д. С. Гончаров

Науковий керівник Чуйко Геннадій Петрович, д-р фіз.-мат. наук, професор

Миколаїв – 2026

АНОТАЦІЯ

Гончаров Д. С. Моделі та методи побудови спеціалізованих систем моніторингу стану здоров'я людини. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії (PhD) за спеціальністю 123 Комп'ютерна інженерія. – Чорноморський національний університет імені Петра Могили, Миколаїв, 2026.

Дисертація присвячена вирішенню нагальної науково-практичної задачі створення та впровадження моделей і методів побудови спеціалізованих систем моніторингу стану здоров'я, які забезпечують повний цикл роботи з гетерогенними біомедичними даними – від їх реєстрації та підготовки до контролю якості, зменшення розмірності, стиснення, передавання та інтелектуального аналізу – у розподіленому контурі «периферія – мережа – сервер/аналітика». Суть задачі полягає в тому, щоб за обмежених ресурсів периферійних вузлів і мережевої пропускної здатності отримувати швидкі, відтворювані та інтерпретовані діагностичні результати без втрати діагностично значущої інформації, поєднуючи легковагові моделі і правила, робастні процедури стабілізації даних та безвтратне стиснення медичних зображень із можливістю апаратного прискорення.

Метою дослідження є вдосконалення спеціалізованих комп'ютерних систем моніторингу стану здоров'я людини шляхом розроблення та впровадження методів відбору й ранжування інформативних ознак та зменшення розмірності вхідних даних.

Об'єкт дослідження

Спеціалізовані комп'ютерні та кіберфізичні системи інтернету речей та пов'язані з ними ІТ-інфраструктури, у яких реалізуються процеси збору, передавання, зберігання та опрацювання масивів даних.

Предмет дослідження

Моделі та методи організації інформаційних процесів попередньої обробки й контролю якості, відбору та редукції інформативних ознак,

передавання гетерогенних біомедичних даних та їх інтелектуального аналізу у складі багаторівневих розподілених комп'ютерних систем моніторингу стану здоров'я людини.

Наукова новизна отриманих результатів:

– **вперше** запропоновано концептуальну модель багаторівневої системи моніторингу стану здоров'я людини, яка, на відміну від монолітних, базорієнтованих, та інших, поєднує рівні подання, прикладної логіки та даних, забезпечує балансування навантаження і паралельне опрацювання запитів від ІоМТ-пристроїв; це дало змогу зменшити частку відмов, що підвищує надійність централізованого збору;

– **вперше** запропоновано метод попередньої оцінки цінності набору WBCD для первинного скринінгу, який на відміну від відомих поєднує ранжування 9 біомаркерів за 8 оцінювачами. Атрибути статистично поділено на 3 групи релевантності: високі, середні і низькі при значущості 95 %, що дало можливість збільшення виявлення кількості викидів до 10,34 %, для класу «patients» до 14,06 %;

– **удосконалено** модель інтерпретованого діагностичного рішення на основі дерева J48 для набору WBCD, яка, на відміну від повного дерева, будується на скороченому наборі релевантних атрибутів, та забезпечує зменшення структурної складності дерева на 43,5 % за вузлами і 41,7 % за листами, що підвищує зручність інтерпретації діагностичних рішень;

– **набули подальшого розвитку** метод скорочення вимірності для WBCD, який на відміну від відомих передбачає послідовне застосування: стандартизації, відбору ознак та PCA-перетворення, що дало можливість зменшити розмірність на 86,7 % зі збереженням 91 % варіативності даних.

Практичне значення отриманих результатів:

Практичне значення результатів дисертаційного дослідження полягає в тому, що запропоновані рішення дають інженерно відтворювану основу для побудови та налаштування діагностичних модулів у системах моніторингу здоров'я, де дані є гетерогенними, зашумленими й часто містять аномальні

спостереження. Формалізована попередня оцінка цінності біомедичних наборів даних і ранжування біомаркерів дозволяє ще до етапу навчання моделей обґрунтовано визначати, які ознаки є найбільш інформативними, а також налаштовувати робастні процедури контролю якості даних і обробки викидів, зменшуючи ризик отримання нестабільних або «випадкових» діагностичних висновків у практичній експлуатації. Удосконалена схема скорочення вимірності, що поєднує стандартизацію, відбір ознак і подальше перетворення простору ознак, забезпечує зниження обчислювального навантаження та вимог до пам'яті при збереженні діагностично значущої інформації, що безпосередньо підвищує придатність аналітики до використання в обмежених ресурсами вузлах і в розподілених контурах обробки. Розвинений метод порівняльного аналізу в середовищі WEKA забезпечує системний, прозорий і повторюваний вибір діагностичних рішень за комплексом показників якості та дозволяє отримувати інтерпретовані моделі у вигляді спрощених дерев або (і) правил, що полегшує валідацію, супровід і інтеграцію таких рішень у прикладні програмні компоненти системи. У сукупності ці результати скорочують час і трудомісткість розроблення діагностичної аналітики, підвищують її надійність на «реальних» даних та підтримують практичне впровадження у багаторівневному контурі збору, підготовки й аналізу медичних даних, що в дисертації пов'язано з використанням у прикладних реалізаціях та освітньому процесі.

Основні результати дисертаційної роботи впроваджено:

- у НДР (№ ДР 012U109898, 2021–2022рр.) «Розробка модулів автоматизації бездротових приладів відновлення пост-інфарктних, пост-інсультних пацієнтів в індивідуальних умовах віддаленої реабілітації» (Акт впровадження від 26 грудня 2022 р.);
- у НДР (№ ДР 0125U000904, 2025-2029рр.) «Інженерія даних як процеси виявлення, збору, реєстрації та подальшої обробки даних в рамках побудови апаратно-програмної інфраструктури даних» (Акт впровадження від 21 січня 2026 р.);

– у навчальний процес за першим (бакалаврським) рівнем вищої освіти на кафедрі комп'ютерної інженерії Чорноморського національного університету імені Петра Могили при проведенні лекційних занять та лабораторних робіт з дисциплін «Сенсори, перетворювачі» українською мовою та «Machine Learning and Pattern Recognition» англійською мовою (Акт впровадження від 10 вересня 2025 р.).

Апробація результатів дисертації

Матеріали дисертаційної праці доповідалися, обговорювалися та здобули схвалення на науково-технічних конференціях та семінарах:

- Всеукраїнська науково-практична конференція «Могилянські читання» (Миколаїв, 2022–2025 рр.);
- Всеукраїнська науково-практична конференція молодих вчених «Інформаційні технології та інженерія» (Миколаїв, 2023–2024 рр.);
- Міжнародна науково-практична конференція Free and open source software, Харків, 14–16 лютого 2023;
- Всеукраїнська науковопрактична конференція «Актуальні завдання медичної, біологічної фізики та інформатики» Вінниця: 7 квітня 2023;
- 2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Dortmund, Germany, 7–9 September 2023;
- 2023 IEEE 13th International Conference on Electronics and Information Technologies (ELIT), Lviv, Ukraine, 26–28 September 2023;
- Міжнародна науково-практична конференція «Ольвійський форум» (Миколаїв, 2022–2024 рр.).

У вступі обґрунтовано тему, актуальність дослідження моделей та методів побудови спеціалізованих систем моніторингу стану здоров'я людини, зумовлену зростанням обсягів медичних даних і потребою передавати та обробляти їх швидко й надійно в телемедичних та ІоМТ-сценаріях. Визначено мету роботи, сформульовано основні завдання, об'єкт і предмет дослідження. Показано, що для практичного використання таких систем важливими є

зменшення трафіку та обчислювального навантаження, а також збереження діагностично важливої інформації під час попередньої обробки, стиснення та подальшої аналітики. Обґрунтовано вибір підходу, у якому частина обробки виконується на периферійному вузлі (edge), а інтелектуальний аналіз здійснюється у середовищі WEKA після підготовки даних у форматі ARFF.

Перший розділ присвячено теоретико-методичним засадам побудови систем моніторингу стану здоров'я людини. У ньому проаналізовано монолітну, database-centric, client-server та distributed-архітектури, а також обґрунтовано доцільність переходу до багаторівневих IoT-рішень. Розглянуто моделі й методи інтеграції інструментарію WEKA до інформаційних систем моніторингу, зокрема через Django ORM, формат ARFF, Python та віддалене завантаження підготовлених наборів даних. Узагальнено особливості наборів даних для машинного навчання і діагностики раку молочної залози, а також сформульовано вимоги до спеціалізованих систем моніторингу, включно з безвтратним стисненням медичних зображень. За результатами аналізу обґрунтовано вибір JPEG-LS і показано перспективність застосування ПЛІС як периферійного обчислювального вузла в таких системах.

Другий розділ дисертації присвячено концептуальним моделям та методам інтелектуального аналізу даних у медичному моніторингу. Розроблено та систематизовано математичний апарат побудови й застосування моделей фізіологічних процесів, зокрема просторово-часові моделі динаміки та дифузійних явищ, термодинамічну модель на основі рівняння біотепла, а також модель ангиогенного «перемикання» як важливого механізму прогресування патологічних станів. Запропоновано метод підготовки інформативних ознак для машинного навчання: від процедур попередньої обробки, виділення ROI та нормалізації – до радіомічної екстракції ознак. Додатково розглянуто синтез алгоритмів класифікації та кластеризації для виявлення патологій і ризик-стратифікації, включно з ансамблевими підходами, глибокими архітектурами з механізмами уваги, методами виявлення аномалій, пояснюваним ШІ та

федеративним навчанням як засобом збереження конфіденційності в розподілених сценаріях.

Третій розділ присвячено практичній реалізації моделей і методів інтелектуального аналізу даних із використанням WEKA в контурі багаторівневої системи моніторингу стану здоров'я людини. У розділі запропоновано модель інтеграції WEKA через Django ORM і формат ARFF, метод формування ARFF-наборів даних, метод попередньої оцінки цінності набору Coimbra Breast Cancer Dataset для первинного скринінгу, а також метод скорочення вимірності набору WBCD із формуванням наборів ds1–ds3. Реалізовано метод порівняльного аналізу ефективності діагностичних рішень у WEKA та побудовано інтерпретовану діагностичну модель на основі дерева рішень J48. Отримані результати підтвердили доцільність використання контрольованого формування вибірок, зменшення розмірності та інтерпретованих моделей для задач медичної діагностики.

Четвертий розділ дисертації присвячено впровадженню та експериментальній апробації розробленої системи моніторингу. У розділі наведено результати впровадження в науково-дослідні роботи та навчальний процес, описано практичну реалізацію багаторівневого контуру, організацію рівня даних на базі MySQL, модулі збору, попередньої обробки й серверної аналітики ІоМТ-даних. Також розглянуто програмно-апаратну реалізацію безвратного стиснення медичних зображень, інтеграцію стиснених даних у контур інтелектуального аналізу, реалізацію периферійного вимірювального модуля на основі AD8232 та оцінку метрологічних і експлуатаційних характеристик системи на реальних наборах даних.

У висновках узагальнено, що запропонований підхід до стиснення і скорочення ознак зменшує обсяг даних та вимоги до ресурсів, зберігаючи високу якість діагностики. Показано, що для набору вісконсинського набору даних про рак молочної залози можна суттєво зменшити кількість ознак і при цьому отримувати стабільно високі результати класифікації, а для набору Коїмбри можливо будувати прості та зрозумілі правила на обмеженій кількості

показників. Запропонована інтеграція «периферія – сервер – WEKA» разом із використанням FPGA підсилює швидкодію та енергоефективність, що важливо для IoMT-систем.

***Ключові слова:** хмарні технології, віддалена реабілітація, мобільні застосунки, ПЛІС, штучний інтелект, моніторинг енергоспоживання, Інтернет медичних речей, моніторинг, безвратне стиснення, сенсор, інтелектуальні системи, IoT-пристрої, віддалений доступ, JPEG-LS, глибоке навчання.*

ABSTRACT

Honcharov D. S. Models and methods for building specialized human health monitoring systems. Qualifying scientific work on the rights of the manuscript.

Dissertation for obtaining the Doctor of Philosophy scientific degree in the specialty 123 Computer engineering (branch of knowledge 12 – Information Technologies). – Petro Mohyla Black Sea National University, Mykolaiv, 2025.

The dissertation is devoted to solving the urgent scientific and practical problem of creating and implementing models and methods for building specialized health monitoring systems that provide a full cycle of work with heterogeneous biomedical data in a distributed circuit «periphery – network – server/analytics». Moreover, this cycle includes such stages as registration and preparation for quality control, dimensionality reduction, compression, transmission and intelligent data analysis. The essence of the problem is to obtain fast, reproducible and interpretable diagnostic results without losing diagnostically significant information, with limited resources of peripheral nodes and network bandwidth, by combining lightweight models and rules, robust data stabilization procedures and lossless compression of medical images with the possibility of hardware acceleration.

The research goal and objectives

The goal of the study is to improve specialized computer systems for monitoring human health by developing and implementing methods for selecting and ranking informative features and reducing the dimensionality of input data.

Object of research

Specialized computer and cyber-physical systems of the Internet of Things and related IT infrastructures, which implement the processes of collecting, transmitting, storing and processing data arrays.

Subject of research

Models and methods for organizing information processes of pre-processing and quality control, selection and reduction of informative features, transmission of heterogeneous biomedical data and their intellectual analysis as part of multi-level distributed computer systems for monitoring human health.

Scientific novelty of the obtained results

For the first time:

- the conceptual model of a multi-level human health monitoring system has been proposed, which, unlike monolithic, base-oriented, and others, combines the presentation, application logic, and data levels, provides load balancing and parallel processing of requests from IoMT devices; this makes it possible to reduce the failure rate, which increases the reliability of centralized collection;
- the method for preliminary assessment of the value of the CBCD set for primary screening has been proposed, which, unlike the known ones, combines the ranking of 9 biomarkers by 8 evaluators. The attributes are statistically divided into 3 relevance groups: high, medium and low at a significance of 95 %, which makes it possible to increase the detection of the number of outliers to 10.34 %, for the «patient» class to 14.06 %.

Improved:

- the model of interpreted diagnostic solution based on the J48 tree for the WBCD set has been improved, which, unlike the full tree, is built on a reduced set of relevant attributes and provides a reduction in the structural complexity of the tree by 43.5 % for nodes and 41.7 % for leaves, which increases the convenience of interpreting diagnostic solutions.

Have been developed further:

- a dimensionality reduction method for WBCD, which, unlike the known ones, involves the sequential application of standardization, feature selection, and PCA transformation, which makes it possible to reduce the dimensionality by 86.7 % while preserving 91 % of the data variability.

Practical significance of the obtained results:

The practical significance of the dissertation research results is that the proposed solutions provide an engineering reproducible basis for building and tuning diagnostic modules in health monitoring systems, where the data is heterogeneous, noisy, and often contain anomalous observations. Formalized preliminary assessment of the value of biomedical datasets and biomarker ranking allow, even before the model training

stage, to reasonably determine which features are the most informative, as well as to configure robust data quality control and outlier processing procedures, reducing the risk of obtaining unstable or «random» diagnostic conclusions in practical operation. The improved dimensionality reduction scheme that combines standardization, feature selection, and subsequent feature space transformation provides a reduction in computational load and memory requirements while preserving diagnostically significant information, which directly increases the suitability of analytics for use in resource-constrained nodes and in distributed processing circuits. The developed method of comparative analysis in the WEKA environment provides a systematic, transparent and repeatable selection of diagnostic solutions according to a set of quality indicators and allows obtaining interpreted models in the form of simplified trees and rules, which simplifies the validation, maintenance and integration of such solutions into the applied software components of the system. Taken together, these results reduce the time and complexity of developing diagnostic analytics, increase its reliability on «real» data, and support practical implementation in a multi-level circuit of collecting, preparing, and analyzing medical data; which in the dissertation is associated with use in applied implementations and the educational process.

The main results of the dissertation research have been implemented:

- the research work (№ 012U109898, 2021–2022) «Development of automation modules for wireless devices for the recovery of post-infarction, post-stroke patients in individual conditions of remote rehabilitation» (Implementation Act dated December 26, 2022);
- the research work (№ 0125U000904, 2022–2029) «Data engineering as the processes of detecting, collecting, registering and further processing data within the framework of building a hardware and software data infrastructure» (Implementation Act of January 21, 2026);
- the educational process at the first (bachelor's) level of higher education at the Department of Computer Engineering of the Petro Mohyla Black Sea National University during lectures and laboratory work in the disciplines «Sensors,

Converters» in Ukrainian and «Machine Learning and Pattern Recognition» in English (Implementation Act dated September 10, 2025).

Approbation of the dissertation results

The materials of the dissertation were presented, discussed, and approved at the following scientific and technical conferences and seminars:

- All-Ukrainian Scientific and Practical Conference «Mohyla Readings» (Mykolaiv, 2022–2025);
- All-Ukrainian Scientific and Practical Conference of Young Scientists «Information Technologies and Engineering» (Mykolaiv, 2023–2024);
- International Scientific and Practical Conference «Free and Open Source Software» (Kharkiv, February 14–16, 2023);
- All-Ukrainian Scientific and Practical Conference «Current Issues of Medical and Biological Physics and Informatics» (Vinnytsia, April 7, 2023);
- 2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS) (Dortmund, Germany, September 7–9, 2023);
- 2023 IEEE 13th International Conference on Electronics and Information Technologies (ELIT) (Lviv, Ukraine, September 26–28, 2023);
- International Scientific and Practical Conference «Olvian Forum» (Mykolaiv, 2022–2024).

In the introduction, the topic and the relevance of the study of models and methods for building specialized human health monitoring systems are substantiated, due to the growth in the volume of medical data and the need to transmit and process them quickly and reliably in telemedicine and IoMT scenarios. The purpose of the work is determined, the main tasks, object and subject of the study are formulated. It is shown that for the practical use of such systems, it is important to reduce traffic and computational load, as well as preserve diagnostically important information during pre-processing, compression and subsequent analytics. The choice of an approach is justified, in which part of the processing is performed on the peripheral node (edge),

and the intelligent analysis is performed in the WEKA environment after data preparation in the ARFF format.

The first chapter is devoted to the theoretical and methodological foundations for building human health monitoring systems. It analyzes monolithic, database-centric, client-server, and distributed architectures, and substantiates the feasibility of transitioning to multi-level IoMT solutions. Models and methods for integrating WEKA into monitoring information systems are considered, in particular through Django ORM, the ARFF format, Python, and remote loading of prepared datasets. The features of datasets for machine learning and breast cancer diagnosis are generalized, and the requirements for specialized monitoring systems, including lossless compression of medical images, are formulated. Based on the analysis, the choice of JPEG-LS is substantiated, and the perspective of using FPGA as a peripheral computing node in such systems is shown.

The second chapter of the dissertation is devoted to conceptual models and methods of intelligent data analysis in medical monitoring. The mathematical framework for modeling physiological processes has been developed and systematized, including spatiotemporal dynamics and diffusion processes, thermodynamic approaches based on the bioheat equation, as well as models of the angiogenic «switch» as an important mechanism of pathological progression. A methodology for preparing informative features for machine learning is proposed: from preprocessing procedures, ROI selection, and normalization to radiomic feature extraction. In addition, the synthesis of classification and clustering algorithms for pathology detection and risk stratification is considered, including ensemble approaches, deep architectures with attention mechanisms, anomaly detection methods, explainable AI, and federated learning as a means of preserving confidentiality in distributed scenarios.

The third chapter is devoted to the practical implementation of models and methods for intelligent data analysis using WEKA within the framework of a multi-level human health monitoring system. The chapter proposes a model for WEKA integration through Django ORM and the ARFF format, a method for generating ARFF

datasets, a method for the preliminary assessment of the value of the Coimbra Breast Cancer Dataset for primary screening, as well as a method for dimensionality reduction of the WBCD dataset with the formation of ds1–ds3 subsets. A method for the comparative analysis of the effectiveness of diagnostic decisions in WEKA was implemented, and an interpretable diagnostic model based on the J48 decision tree was developed. The obtained results confirmed the feasibility of using controlled dataset formation, dimensionality reduction, and interpretable models for medical diagnostic tasks.

The fourth chapter of the dissertation is devoted to the implementation and experimental validation of the developed monitoring system. The chapter presents the results of implementation in research projects and the educational process, and describes the practical implementation of the multi-level framework, the organization of the data layer based on MySQL, and the modules for collection, preprocessing, and server-side analytics of IoMT data. It also considers the hardware-software implementation of medical images lossless compression, the compressed data integration into the intelligent analysis framework, the implementation of a peripheral measurement module based on AD8232, and the evaluation of the metrological and operational characteristics of the system using real datasets.

In the conclusions, summarize that the proposed approach to compression and feature reduction decreases the data volume and resource requirements while maintaining high diagnostic quality. It is shown that for the WBCD dataset the number of features can be significantly reduced while still achieving consistently high classification results, and for the Coimbra dataset it is possible to build simple and clear rules using a limited set of indicators. The proposed «edge – server – WEKA» integration, together with the use of FPGA, improves performance and energy efficiency, which is important for IoMT systems.

Keywords: *cloud technologies, remote rehabilitation, mobile applications, FPGA, artificial intelligence, power consumption monitoring, Internet of medical things, monitoring, lossless compression, sensor, intelligent systems, IoT devices, remote access, JPEG-LS, deep learning.*

СПИСОК ОПУБЛІКОВАНИХ НАУКОВИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

**Наукові праці, в яких опубліковані основні результати дисертації та
відповідають п. 8 Постанови КМУ від 12.01.2022 № 44**

1. Chuiko G., Honcharov D. Dimensionality cutback and deep learning algorithms efficacy as to the breast cancer diagnostic dataset. *Radioelectronic and Computer Systems*. 2024. Vol. 2024, no. 4. P. 91–98. DOI: 10.32620/reks.2024.4.08. URL: <https://nti.khai.edu/ojs/index.php/reks/article/view/2652> (Last accessed: 12.01.2026). **ISSN: 2663-2012 Scopus(Q3)**

2. Chuiko G., Honcharov D. Breast cancer dataset from Coimbra: Pre-ratings of its value to machine learning and diagnosis. *Digital Technologies Research and Applications*. 2025. Vol. 4, no. 2. P. 182–193. DOI: 10.54963/dtra.v4i2.1348. URL: <https://ojs.ukscip.com/index.php/dtra/article/view/1348> (Last accessed: 12.01.2026). **ISSN: 2754-5687 Scopus**

3. Chuiko G., Dvornik O., Darnapuk Y., Honcharov D., Yaremchuk O. Asleep adults' breathing patterns via data mining of electromyograms. *Proceedings of the IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, Dortmund, Germany, 2023. P. 550–554. DOI: 10.1109/IDAACS58523.2023.10348674. URL: <https://ieeexplore.ieee.org/document/10348674> **ISSN: 2770-4254 Scopus**

4. Кандиба І. О., Гончаров Д. С., Гончарова Н. В. Інтеграція інструментарію Weka до багаторівневої системи моніторингу стану здоров'я людини. *Наука і техніка сьогодні*. 2025. № 5 (46). DOI: 10.52058/2786-6025-2025-5(46)-1571-1584. URL: <https://perspectives.pp.ua/index.php/nts/article/view/24521> (дата звернення: 12.01.2026). **кат. Б**

5. Кандиба І. О., Гончаров Д. С., Гончарова Н. В. Архітектура багаторівневої системи моніторингу стану здоров'я людини. *Вчені записки Таврійського національного університету імені В. І. Вернадського. Серія*

«Технічні науки». 2024. № 2. С. 78–84. DOI: 10.32782/2663-5941/2024.2/11. URL: https://tech.vernadskyjournals.in.ua/journals/2024/2_2024/13.pdf кат. Б

6. Гончаров Д. С. Інтелектуальні моделі та методи моніторингу стану здоров'я людини на основі даних сенсорних мереж. *Вчені записки Таврійського національного університету імені В. І. Вернадського. Серія: Технічні науки*. 2025. № 4. С. 80–86. DOI: 10.32782/2663-5941/2025.4.2/11. URL: https://www.tech.vernadskyjournals.in.ua/journals/2025/4_2025/part_2/13.pdf кат Б

7. Гончаров Д. С. Діагностика раку грудей: ІОТ-система з FPGA та WEKA-аналітикою. *Технічні науки та технології*. 2025. № 3 (41). С. 185–192. DOI: 10.25140/2411-5363-2025-3(41)-185-192. URL: <http://tst.stu.cn.ua/article/view/345197> кат. Б

8. Кандиба І.О., Фісун М. Т., Горбань Г. В., Антипова К. О., Гончаров Д. С. Аналіз даних в системі моніторингу стану здоров'я людини засобами Python. *Наука і техніка сьогодні*. 2024. № 5(33). С. 1162–1176. DOI: 10.52058/2786-6025-2024-5(33)-1162-1175. ISSN: 2786-6025 кат. Б

Список публікацій, які засвідчують апробацію матеріалів дисертації

– матеріали конференцій, індексовані в наукометричній базі Scopus:

9. Chuiko G., Dvornik O., Darnapuk Y., Honcharov D., Krainyk Y., Yaremchuk O. Attribute selection, outliers impact study and visualization within breast cancer detection. *2023 IEEE 13th International Conference on Electronics and Information Technologies, Lviv, Ukraine, 26–28 September 2023*. P. 1–5. DOI: 10.1109/elit61488.2023.10310922.

– інші матеріали конференцій:

10. Гончаров Д. С. Стиснення інформації у системах моніторингу стану здоров'я: методи сімейства LZx. *Інформаційні технології та інженерія* : тези доповідей Всеукр. наук.-практ. конф. молодих вчених, аспірантів і студентів, Миколаїв, 7–10 лютого 2023 р. С. 65–67. URL: <https://dspace.chmnu.edu.ua/jspui/handle/123456789/606> (дата звернення: 12.01.2026).

11. Гончаров Д. С., Гончарова Н. В. Огляд та застосування оптичного датчика MAX30105 у медицині. *Free and Open Source Software (FOSS-2023)*: тези доп. XI Міжнар. наук.-практ. конф. / Харків. нац. економ. ун-т ім. Семена Кузнеця, Харків, 14–16 лютого 2023 р. Харків: ХНЕУ ім. Семена Кузнеця, 2023. С. 51–52. URL: <http://repository.hneu.edu.ua/bitstream/123456789/29041/1/foss-2023-theses.pdf> (дата звернення: 12.01.2026).

12. Гончаров Д. С., Чуйко Г. П. ЕКГ-модуль ad8232: будова, характеристики і застосування. *Актуальні завдання медичної, біологічної фізики та інформатики*: тези доп. II Всеукр. наук.-практ. конф., Вінниця, 7 квітня 2023 р. С. 113–114. URL: <https://dspace.vnmu.edu.ua/handle/123456789/6161> (дата звернення: 12.01.2026).

13. Гончаров Д. С. Обробка даних з приладів за допомогою weka на прикладі раку молочної залози. *Ольвійський форум – 2023: стратегії країн Причорноморського регіону в геополітичному просторі: тези доповідей: Технічні науки*/ Миколаїв: ЧНУ, 15-18 червня 2023. С. 13–17. URL: <https://dspace.chmnu.edu.ua/jspui/handle/123456789/1194> (дата звернення: 12.01.2026).

14. Гончаров Д. С. Емпірична оцінка якості вимірювань датчика пульсу шляхом статистичної обробки даних. *Ольвійський форум – 2024: стратегії країн Причорноморського регіону в геополітичному просторі: тези доповідей: Технічні науки*/ Миколаїв: ЧНУ, 20-23 червня 2024. С. 146–151. URL: <https://dspace.chmnu.edu.ua/jspui/handle/123456789/1194> (дата звернення: 12.01.2026).

15. Гончаров Д. С. Застосування камер в медичних роботах-маніпуляторах. *Інформаційні технології та інженерія*: тези доповідей: Всеукраїнська наук.-практ. конф. молодих вчених, аспірантів і студентів. Миколаїв: ЧНУ, 29–02 лютого, 2024. С. 74–76 URL: <https://files.znu.edu.ua/files/Bibliobooks/Inshi79/0058873.pdf#page=74> (дата звернення: 12.01.2026).

16. Чуйко Г. П., Гончаров Д. С., Бущук М. С. Відбір атрибутів та скорочення розмірності даних щодо раку молочної залози з Knowledge flowe. *Могілянські читання – 2024*: тези доповідей Всеукр. наук.-практ. конф. Комп'ютерні науки. Технічні науки, Миколаїв, 6–10 листоп. 2024. С. 152–155. URL: <https://dspace.chmnu.edu.ua/jspui/handle/123456789/2507> (дата звернення: 12.01.2026).

ЗМІСТ

| | |
|---|-----------|
| ПЕРЕЛІК СКОРОЧЕНЬ | 21 |
| ВСТУП | 22 |
| РОЗДІЛ 1 ТЕОРЕТИКО-МЕТОДИЧНІ ЗАСАДИ ПОБУДОВИ СИСТЕМ МОНІТОРИНГУ СТАНУ ЗДОРОВ'Я ЛЮДИНИ..... | 29 |
| 1.1 Монолітна, database-centric, front-end/back-end та distributed-архітектури в системах моніторингу стану здоров'я людини | 29 |
| 1.2 Моделі та методи інтеграції інструментарію WEKA до інформаційних систем моніторингу..... | 46 |
| 1.3 Огляд наборів даних для машинного навчання та діагностики раку молочної залози..... | 58 |
| 1.4 Формалізація вимог до спеціалізованих систем моніторингу та діагностики..... | 61 |
| 1.5 Можливості застосування ПЛІС (FPGA) в системі моніторингу..... | 62 |
| Висновки до розділу 1..... | 63 |
| РОЗДІЛ 2 КОНЦЕПТУАЛЬНІ МОДЕЛІ ТА МЕТОДИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ..... | 65 |
| 2.1 Математичні моделі опису фізіологічних процесів: моделі Гомперца, просторової гетерогенності, Пеннеса та ін. | 65 |
| 2.2 Методи підготовки даних та формування інформативних ознак для машинного навчання на прикладі діагностики раку молочної залози..... | 70 |
| 2.3 Синтез алгоритмів класифікації та кластеризації для виявлення загрозливих та патологічних станів | 76 |
| Висновки до розділу 2..... | 81 |
| РОЗДІЛ 3 ПРАКТИЧНА РЕАЛІЗАЦІЯ МОДЕЛЕЙ І МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ З ВИКОРИСТАННЯМ WEKA | 83 |

| | |
|--|------------|
| | 20 |
| 3.1 Модель інтеграції інструментарію WEKA до багаторівневої системи моніторингу стану здоров'я людини..... | 83 |
| 3.2 Метод формування ARFF-наборів даних для WEKA..... | 86 |
| 3.3 Метод попередньої оцінки цінності набору CBCD для первинного скринінгу..... | 88 |
| 3.4 Метод скорочення вимірності WBCD із формуванням наборів ds1–ds3... | 95 |
| 3.5 Метод порівняльного аналізу ефективності діагностичних рішень у WEKA | 101 |
| 3.6 Модель інтерпретованого діагностичного рішення на основі дерева J48106 | |
| 3.7 Модель індукції діагностичних правил на основі JRip..... | 111 |
| 3.8 Метод валідації кластерної структури даних за метрикою FMI..... | 116 |
| Висновки до розділу 3..... | 121 |
| РОЗДІЛ 4 ВПРОВАДЖЕННЯ ТА ЕКСПЕРИМЕНТАЛЬНА АПРОБАЦІЯ РОЗРОБЛЕНОЇ СИСТЕМИ МОНІТОРИНГУ | 123 |
| 4.1 Впровадження результатів дослідження | 123 |
| 4.2 Програмно-апаратна реалізація методів стиснення зображень без втрат для медичної візуалізації..... | 144 |
| 4.3 Оцінка метрологічних та експлуатаційних характеристик системи на реальних наборах даних..... | 145 |
| 4.4 Експериментальна перевірка багаторівневої системи моніторингу..... | 154 |
| Висновки до розділу 4..... | 157 |
| ВИСНОВКИ..... | 159 |
| СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ..... | 163 |
| ДОДАТОК А Акти впровадження..... | 177 |
| ДОДАТОК Б Список опублікованих наукових праць за темою дисертації | 181 |

ПЕРЕЛІК СКОРОЧЕНЬ

| | |
|-------|--|
| ССЗ | – Серцево-судинні захворювання |
| РМЗ | – Рак молочної залози |
| ШІ | – Штучний інтелект |
| МІС | – Медична інформаційна система |
| | |
| ARFF | – Attribute-Relation File Format |
| CBCD | – Coimbra Breast Cancer Dataset |
| FPGA | – Field-Programmable Gate Array |
| GLRLM | – Gray Level Run Length Matrix |
| IoMT | – Internet of Medical Things |
| IoT | – Internet of Things |
| ROI | – Region of Interest |
| RPM | – Remote patient monitoring |
| WBCD | – Wisconsin Breast Cancer Dataset |
| WEKA | – Waikato Environment for Knowledge Analysis |

ВСТУП

Сучасна медицина дедалі активніше переходить від епізодичних вимірювань до безперервного дистанційного моніторингу, що спирається на кіберфізичні системи та Інтернет медичних речей (ІоМТ). Такі системи об'єднують сенсорні вузли, канали зв'язку, серверні або хмарні компоненти, де безперервно відбуваються процеси отримання, передавання, зберігання, опрацювання та захисту медичних даних. Зростання обсягів і різноманітності цих даних посилює вимоги до продуктивності, надійності та відтворюваності обчислень, а також до стабільності роботи в умовах змінних мережевих параметрів. У телемедичних сценаріях критичними стають затримки, стійкість до збоїв і захищеність інформаційних потоків, оскільки від цього залежить своєчасність та коректність рішень.

Практична реалізація ІоМТ-платформ стикається з інженерним протиріччям: для якісної діагностичної підтримки потрібні точні алгоритми інтелектуального аналізу, але обчислювальні ресурси периферійних вузлів часто обмежені, як і їхній енергобюджет. Передавання великих масивів даних у мережі підсилює потребу в ефективному зменшенні потоку або стисканні без втрати діагностично значущої інформації, інакше система стає надто повільною або дорогою в експлуатації. Навіть за достатньої точності моделі залишається проблема відтворюваності результатів у різних умовах збору даних і на різних апаратно-програмних конфігураціях, що ускладнює масштабування. Крім точності, важливою є інтерпретованість, адже результати мають бути прозорими для фахівця та придатними до перевірки на рівні ознак, правил і критеріїв. Отже, актуальність підсилюється необхідністю одночасно узгодити обчислювальну ефективність, пропускну здатність, надійність і вимоги до якості аналітики.

Тому виникає потреба в обґрунтованих моделях і методах побудови спеціалізованих систем моніторингу, які забезпечують узгоджену роботу «периферія – мережа – сервер/аналітика» без втрати керованості та контролю якості. Ключовим стає таке перетворення даних і ознак, яке зменшує обчислювальне навантаження та обсяг передавання, але зберігає діагностичну

інформативність і дозволяє відтворювати результати на різних наборах і в різних умовах. Не менш важливою є побудова легковагових діагностичних моделей, здатних працювати швидко та стабільно, водночас залишаючись достатньо інтерпретованими для практичного використання. Особливої уваги потребують сценарії, де поряд із табличними біомаркерами обробляються зображення, оскільки саме вони створюють пікові навантаження на обчислювальний контур і канали зв'язку. Перспективним є поєднання програмних процедур підготовки даних і навчання моделей із можливістю прискорення окремих етапів попередньої обробки та стискання на спеціалізованих обчислювальних засобах, коли цього вимагає режим наближення до реального часу. Водночас система має зберігати властивості надійності та безпеки, адже обробляє чутливу інформацію і працює в розподіленому середовищі, де збої та атаки є практично неминучими. Саме поєднання цих вимог формує актуальність теми та визначає необхідність подальшого розвитку підходів до проектування, реалізації й перевірки спеціалізованих систем моніторингу стану здоров'я людини.

Мета та завдання дослідження

Метою роботи є вдосконалення спеціалізованих комп'ютерних систем моніторингу стану здоров'я людини шляхом розроблення та впровадження методів відбору й ранжування інформативних ознак та зменшення розмірності вхідних даних. Запропонований підхід спрямований на мінімізацію обчислювальних витрат при збереженні високих показників прогностичної точності, що є критичним для систем, які працюють в умовах обмежених ресурсів.

Для досягнення поставленої мети розв'язано такі основні завдання:

- провести критичний аналіз існуючих підходів до проектування спеціалізованих систем моніторингу та сформулювати вимоги до алгоритмів інтелектуальної обробки даних у контексті архітектур ІоМТ;
- обґрунтувати методи та технологічні засади структурування біомедичних показників для їхньої подальшої обробки та аналізу засобами середовища WEKA;

- розробити та дослідити методику скорочення інформативного простору ознак (відбір, фільтрація, редукція) для підвищення ефективності класифікації;
- здійснити порівняльну оцінку ефективності методів глибинного навчання та класичних алгоритмів машинного навчання в середовищі WEKA, аналізуючи їхню роботу як на вихідних масивах даних, так і в умовах зниженої розмірності ознакового простору;
- дослідити можливості інтерпретації моделей і сформулювати «легковагові» діагностичні правила для задач первинного скринінгу;
- запропонувати апаратно-програмний підхід, у якому попередня обробка (стиснення) даних може бути реалізована на FPGA з подальшим імпортом підготовлених даних у WEKA для навчання та тестування моделей.

Об'єктом дослідження є спеціалізовані комп'ютерні та кіберфізичні системи інтернету речей та пов'язані з ними IT-інфраструктури, у яких реалізуються процеси збору, передавання, зберігання та опрацювання масивів даних.

Предмет дослідження – моделі та методи організації інформаційних процесів попередньої обробки й контролю якості, відбору та редукції інформативних ознак, передавання гетерогенних біомедичних даних та їх інтелектуального аналізу у складі багаторівневих розподілених комп'ютерних систем моніторингу стану здоров'я людини.

Методи дослідження

Для розв'язання поставлених завдань використано комплекс методів дослідження комп'ютерних та кіберфізичних ІоМТ-систем і штучного інтелекту, методи системного аналізу та побудови багаторівневих моделей ІоМТ-систем, методи підготовки та структуризації даних, методи відбору ознак і зниження вимірності, методи машинного та глибокого навчання й кластеризації в середовищі WEKA, а також статистичні методи оцінювання якості (крос-валідація, PRC та ROC-аналіз, коефіцієнт кореляції Метьюса, показники узгодженості кластерів тощо). Для апаратної реалізації та експериментальної

перевірки застосовано методи проектування цифрових систем на FPGA (опис логіки, верифікація та реалізація у відповідних САПР), а також методи експериментальної оцінки коефіцієнта стиснення та характеристик інтеграції даних у контур аналітики.

Практичне значення отриманих результатів: Практичне значення результатів дисертаційного дослідження полягає в тому, що запропоновані рішення дають інженерно відтворювану основу для побудови та налаштування діагностичних модулів у системах моніторингу здоров'я, де дані є гетерогенними, зашумленими й часто містять аномальні спостереження. Формалізована попередня оцінка цінності біомедичних наборів даних і ранжування біомаркерів дозволяє ще до етапу навчання моделей обґрунтовано визначати, які ознаки є найбільш інформативними, а також налаштувати робастні процедури контролю якості даних і обробки викидів, зменшуючи ризик отримання нестабільних або «випадкових» діагностичних висновків у практичній експлуатації. Удосконалена схема скорочення вимірності, що поєднує стандартизацію, відбір ознак і подальше перетворення простору ознак, забезпечує зниження обчислювального навантаження та вимог до пам'яті при збереженні діагностично значущої інформації, що безпосередньо підвищує придатність аналітики до використання в обмежених ресурсах вузлах і в розподілених контурах обробки. Розвинений метод порівняльного аналізу в середовищі WEKA забезпечує системний, прозорий і повторюваний вибір діагностичних рішень за комплексом показників якості та дозволяє отримувати інтерпретовані моделі у вигляді спрощених дерев або (і) правил, що полегшує валідацію, супровід і інтеграцію таких рішень у прикладні програмні компоненти системи. У сукупності ці результати скорочують час і трудомісткість розроблення діагностичної аналітики, підвищують її надійність на «реальних» даних та підтримують практичне впровадження у багаторівневному контурі збору, підготовки й аналізу медичних даних, що в дисертації пов'язано з використанням у прикладних реалізаціях та освітньому процесі.

Результати дисертаційної роботи мають суттєве практичне значення в галузі ІоМТ-систем моніторингу стану здоров'я, телемедичних технологій передавання та зберігання медичних даних, інтелектуального аналізу (WEKA-аналітики). Запропоновані підходи можуть бути використані у клінічній діагностиці та скринінгу, у прикладних НДР, при розробці спеціалізованих кіберфізичних медичних систем, а також у підготовці фахівців зі спеціальності 123 «Комп'ютерна інженерія».

Запропоновані підходи дозволяють:

- зменшити трафік і вимоги до пам'яті в телемедичних ІоМТ-системах за рахунок структурного та безвратного стиснення даних без спотворення діагностично важливих ознак;
- підвищити стабільність діагностичних рішень на зашумлених даних і за наявності викидів завдяки робастним процедурам контролю якості даних.

Основні результати дисертаційної роботи впроваджено:

- у НДР (№ ДР 012U109898, 2021–2022рр.) «Розробка модулів автоматизації бездротових приладів відновлення пост-інфарктних, пост-інсультних пацієнтів в індивідуальних умовах віддаленої реабілітації» (Акт впровадження від 26 грудня 2022р.);
- у НДР (№ ДР 0125U000904, 2025–2029рр.) «Інженерія даних як процеси виявлення, збору, реєстрації та подальшої обробки даних в рамках побудови апаратно-програмної інфраструктури даних» (Акт впровадження від 21 січня 2026 р.);
- у навчальний процес за першим (бакалаврським) рівнем вищої освіти на кафедрі комп'ютерної інженерії Чорноморського національного університету імені Петра Могили при проведенні лекційних занять та лабораторних робіт з дисциплін «Сенсори, перетворювачі» українською мовою та «Machine Learning and Pattern Recognition» англійською мовою (Акт впровадження від 10 вересня 2025 р.).

Особистий внесок здобувача

Основний зміст роботи, всі теоретичні та практичні результати, висновки і дослідження, що представлено до захисту, одержані автором самостійно. В опублікованих роботах (додаток Б) особисто здобувачеві належать наступні наукові результати:

– в статті [1] проведено комп'ютерні експерименти в середовищі WEKA для Wisconsin Breast Cancer dataset з метою суттєвого скорочення розмірності без втрати прогностичної здатності: виконано послідовну попередню обробку та редукцію ознак (стандартизація – відбір ознак CfsSubsetEval з формуванням ds2 на 11 атрибутів – перетворення PrincipalComponents із параметром $R = 0,91$, що зменшує ds2 до 4 головних компонент і охоплює 91 % дисперсії, формуючи ds3);

– у статті [2] здійснено оцінювання набору Coimbra Breast Cancer Dataset, сформованого для задачі первинної діагностики раку молочної залози, та поставлено мету розробити візуальний класифікатор для підтримки прогнозування й моніторингу на основі рутинних показників крові та фізичних даних у середовищі WEKA;

– у роботі [3] автор виконав підготовку та попередню обробку ЕМГ-даних, провів інтелектуальний аналіз для виділення характерних патернів дихання під час сну;

– у роботі [5] запропонував і описав багаторівневу архітектуру системи моніторингу стану здоров'я, визначив функції рівнів (периферійний вузол, передавання, серверна частина та аналітика), а також обґрунтував місце обробки даних і прийняття діагностичних рішень у загальному контурі системи.

Апробація результатів дисертації:

Матеріали дисертаційної праці доповідалися, обговорювалися та здобули схвалення на науково-технічних конференціях та семінарах:

– Всеукраїнська науково-практична конференція «Могилянські читання» (Миколаїв, 2022–2025 рр.);

- Всеукраїнська науково-практична конференція молодих вчених «Інформаційні технології та інженерія» (Миколаїв, 2023–2024 рр.);
- Міжнародна науково-практична конференція Free and open source software, Харків, 14–16 лютого 2023;
- Всеукраїнська науково-практична конференція «Актуальні завдання медичної, біологічної фізики та інформатики» Вінниця: 7 квітня 2023;
- 2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS), Dortmund, Germany, 7–9 September 2023;
- 2023 IEEE 13th International Conference on Electronics and Information Technologies (ELIT), Lviv, Ukraine, 26–28 September 2023;
- Міжнародна науково-практична конференція «Ольвійський форум» (Миколаїв, 2022–2024 рр.).

Публікації

Згідно до теми дисертаційного дослідження опубліковано 16 наукових праць, з них 3 праці – у наукових фахових виданнях України кат. А, які також проіндексовані у наукометричній базі Scopus; 7 праць – у збірниках матеріалів міжнародних науково-практичних конференцій, проіндексованих у наукометричній базі Scopus.

Структура та обсяг дисертації

Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел до кожного розділу та додатків. Основний зміст викладено на 176 сторінках друкованого тексту, містить 26 рисунків та 23 таблиці. Список використаних джерел містить 107 найменування. Загальний обсяг роботи становить 184 сторінки.

РОЗДІЛ 1

ТЕОРЕТИКО-МЕТОДИЧНІ ЗАСАДИ ПОБУДОВИ СИСТЕМ МОНІТОРИНГУ СТАНУ ЗДОРОВ'Я ЛЮДИНИ

1.1 Монолітна, database-centric, front-end/back-end та distributed-архітектури в системах моніторингу стану здоров'я людини

Побудова сучасних систем моніторингу стану здоров'я людини потребує не лише вибору сенсорів (sensors), каналів передавання даних і засобів зберігання інформації, а й обґрунтованого вибору архітектури всієї системи. Саме архітектурна організація визначає, яким чином будуть взаємодіяти між собою пристрої збору даних, прикладна логіка, модулі візуалізації, аналітичні сервіси та база даних. Для систем медичного моніторингу це особливо важливо, оскільки такі системи мають працювати з потоками неоднорідних біомедичних показників, підтримувати безперервне спостереження, забезпечувати збереження історії вимірювань і водночас залишатися достатньо гнучкими для розширення функціоналу [1, 2].

У роботі [3] показано, що для побудови систем моніторингу стану здоров'я людини можуть застосовуватися різні відомі архітектурні моделі, зокрема монолітна архітектура застосунку, архітектура, орієнтована на базу даних, архітектура з поділом на клієнтську та серверну частини, а також підхід розподілених обчислень. Кожна з них має власні переваги, але окремо не забезпечує повною мірою вимоги, характерні для телемедичних та ІоМТ-систем, де необхідно одночасно підтримувати збір даних від пристроїв, централізоване зберігання, масштабування, балансування навантаження та подальшу аналітичну обробку. Тому на першому етапі доцільно розглянути особливості цих відомих архітектурних моделей, а також їх придатність до використання в системах моніторингу стану здоров'я людини [3, 4].

Серед суміжних досліджень, релевантних побудові сучасних систем моніторингу стану здоров'я людини, доцільно виділити праці, присвячені автоматизації реабілітаційних пристроїв, ІоТ-пристроїв, нейромережевому

аналізу даних та ефективності бездротових мереж. У роботі [5] розглянуто *methods and tools of formation of general indexes for automation of devices in rehabilitative medicine for post-stroke patients*, що підтверджує актуальність розроблення інтелектуальних систем для підтримки реабілітаційних процедур. У статті [6] показано приклад створення IoT-пристрою на основі лічильника Гейгера, що демонструє практичні підходи до побудови сенсорних вузлів, віддаленого доступу до даних та інтеграції периферійних засобів у мережеві системи спостереження. Робота [7], присвячена *neural network for emotions recognition*, ілюструє можливості *artificial intelligence* та *deep learning* для обробки складних сигналів і прийняття рішень в умовах невизначеності. У [8] розглянуто методи підвищення ефективності бездротових мереж передавання даних для систем IoT, що є важливим для забезпечення надійної роботи IoT-пристроїв у системах моніторингу. Додатково робота [9] щодо попереднього моделювання навантаження центру віддаленого керування підтверджує значення керування ресурсами розподілених платформ, у яких стабільність функціонування залежить від узгодженої роботи каналів зв'язку, обчислювальних вузлів і сервісів віддаленого доступу.

1.1.1 Монолітна архітектура застосунку (Monolithic application)

Монолітна архітектура застосунку належить до найвідоміших базових моделей побудови програмних систем. Її характерною ознакою є те, що основні функціональні частини системи реалізуються в межах одного програмного цілого, тобто як єдиний застосунок, у якому поєднані інтерфейс, прикладна логіка та механізми доступу до даних. У роботі [3] архітектури програмних систем у медичній інформатиці поділено, зокрема, на монолітні та розподілені, причому монолітна архітектура розглядається як варіант, що складається з єдиного пакета або цілісного програмного модуля. У роботі [11] монолітний підхід також описано як такий, за якого один застосунок охоплює всі сервіси й функції проєкту.

Для початкових етапів розроблення монолітна архітектура має очевидні переваги. Вона є зрозумілою з точки зору проєктування, відносно простою у розгортанні й зручною для невеликих систем, де кількість користувачів, джерел даних і сценаріїв взаємодії ще не є значною. Саме тому в окремих прикладних рішеннях та ранніх версіях інформаційних систем такий підхід тривалий час залишався поширеним. У роботі [4] показано, що в системах віддаленого моніторингу пацієнтів традиційно серверні застосунки часто будувалися саме за монолітною схемою. Робота також містить класифікацію прикладних рішень і окреслює критичні проблеми галузі, серед яких пріоритетними названо захист приватності пацієнтів та оптимізацію енергоспоживання пристроїв (рисунок 1.1).

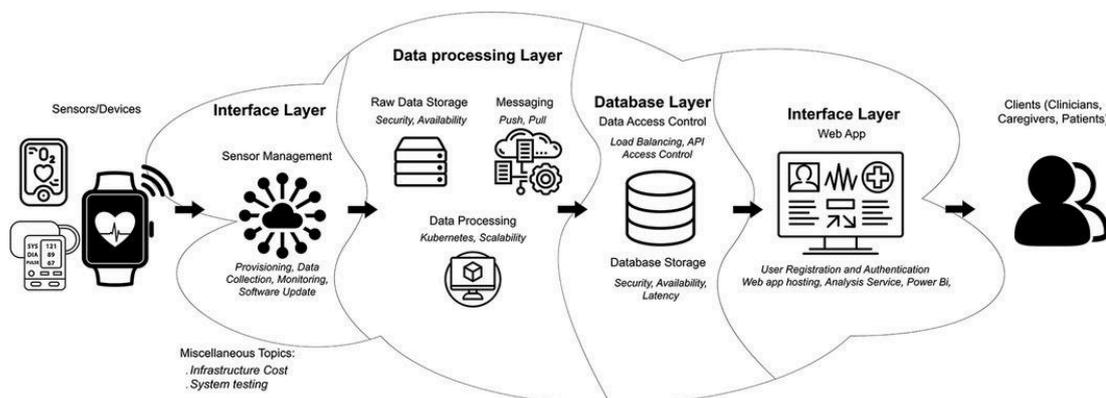


Рисунок 1.1 – Автори виділяють такі архітектурні шари: сенсорний, комунікаційний, обробка даних, прикладний рівень

У роботі [10], присвяченій архітектурам wearable-систем моніторингу, також простежується логіка побудови рішень від простих централізованих схем до складніших багатокомпонентних структур, що підтверджує практичну придатність простих архітектур лише для обмежених за масштабом сценаріїв.

Разом із тим для систем моніторингу стану здоров'я людини монолітна архітектура має низку суттєвих обмежень. Якщо один застосунок одночасно відповідає за приймання даних від сенсорів (sensor), їх перевірку, обробку, збереження в базі даних, візуалізацію та керування доступом, то зі зростанням кількості підключених пристроїв і користувачів така структура ускладнюється. У роботі [4] прямо зазначено, що в монолітній схемі один вебсерверний

застосунок виконує отримання даних, логіку обробки та оновлення бази даних; при цьому будь-яка зміна вимагає повторного тестування всієї системи, а сама структура стає складною для супроводу. Подібна проблематика відображена й у роботі [3], де систематичний огляд архітектури медичних інформаційних систем (англ. health information systems, далі МІС) показує зміщення сучасних рішень від монолітних схем до більш модульних і сервісно-орієнтованих архітектур, оскільки саме вони краще підтримують масштабованість, інтероперабельність і стійкість системи.

Для медичних і телемедичних систем це має особливе значення. У таких застосунках архітектура повинна не лише забезпечувати збереження даних, а й підтримувати безперервне надходження показників, безпечний доступ, інтеграцію аналітичних модулів і поступове нарощування функціональності. У роботі [4] підкреслено, що для віддаленого моніторингу пацієнтів масштабованість і захист даних є критичними вимогами, а тому для повного використання можливостей хмарної інфраструктури доцільніше переходити до мікросервісної організації. Робота [2] також показує, що для медичних інформаційних систем більш придатними є архітектури з розділенням рівнів подання, прикладної логіки та керування даними, оскільки це полегшує підтримку, зміну окремих компонентів і подальший розвиток системи.

Отже, монолітна архітектура застосунку є важливою відомою моделлю, яку доцільно розглядати як вихідний етап еволюції програмних систем моніторингу. Вона може бути придатною для невеликих або прототипних рішень, однак у задачах безперервного моніторингу стану здоров'я людини, особливо в ІоМТ-сценаріях і сценаріях віддаленого спостереження за пацієнтами (англ. remote patient monitoring), її можливості є обмеженими. Саме це пояснює подальший перехід до більш гнучких архітектур, у яких окремі функції системи розділяються між кількома рівнями або сервісами.

1.1.2 Архітектура, орієнтована на базу даних (Database-centric)

Архітектура, орієнтована на базу даних (БД), належить до відомих моделей побудови інформаційних систем, у яких база даних виконує не лише функцію зберігання, а й значною мірою визначає спосіб організації прикладної логіки та доступу до інформації. У роботі [12] підхід, орієнтований на БД, подано як архітектуру, у якій можливості сучасної системи керування базами даних використовуються ширше, ніж просте збереження записів, зокрема для забезпечення безпеки, надійності та розширюваності застосунків. Близьку за змістом ідею викладено у роботі [13], де запропоновано архітектуру, яка заснована на використанні тільки систем керування реляційною базою даних (англ. Relational Database Management System only Architecture, RDBMS-only architecture) вебзастосунків, у межах якої основна логіка системи максимально зосереджується на рівні реляційної бази даних.

Для медичних інформаційних систем та систем моніторингу стану здоров'я людини такий підхід є природним, оскільки саме БД забезпечує накопичення результатів вимірювань, ведення історії спостережень, формування вибірок та централізований доступ до медичної інформації. У роботі [12] прямо запропоновано орієнтовану на БД архітектуру для домашнього моніторингу здоров'я, причому наголошено, що механізми систем керування базами даних (СКБД) можуть бути використані не лише для керування даними, а й для підтримки прикладних функцій інфраструктури домашнього медичного обслуговування (англ. home-based healthcare). Подібну ідею розвинуто і в роботі [14], де активна обробка всередині бази даних розглядається як засіб підтримки систем допомоги в побуті (англ. Ambient Assisted Living Systems), тобто база даних фактично стає активним компонентом архітектури, а не пасивним сховищем.

Перевага такого підходу полягає в тому, що він дозволяє централізувати роботу з даними, зменшити дублювання механізмів доступу та спростити забезпечення цілісності інформації. Для систем моніторингу це важливо, оскільки біомедичні дані надходять послідовно, накопичуються протягом

тривалого часу і повинні бути доступними для подальшого аналізу. Разом із тим орієнтація переважно на базу даних не розв'язує повністю задачі, пов'язані з розподілом навантаження, координацією сервісів, гнучким масштабуванням та відокремленням прикладної логіки від рівня зберігання. Саме тому в сучасних медичних інформаційних системах модель орієнтації на БД частіше розглядається як важлива складова архітектури, але не як вичерпне універсальне рішення. Це добре узгоджується з роботою [2], у якій для МІС обґрунтовано багаторівневу референтну архітектуру з окремими рівнями подання, прикладної логіки та керування даними.

Ще одне обмеження підходу, пов'язаного з орієнтацією на БД, проявляється тоді, коли система переходить від простого накопичення даних до безперервного дистанційного моніторингу. У таких сценаріях важливими стають не лише запити до бази даних, а й обмін повідомленнями між пристроями, обробка потоків вимірювань, інтеграція аналітичних модулів, масштабування серверної частини та захист персональних даних. У роботі [4], присвяченій інфраструктурі віддаленого спостереження за пацієнтом на хмарних платформах, підкреслено, що зростання вимог до масштабованості та підтримки сервісів призводить до відходу від простіших централізованих схем на користь гнучкіших архітектур. Подібний висновок міститься і в систематичному огляді [3], де показано, що сучасні МІС дедалі частіше реалізуються на основі багаторівневого, сервісно-орієнтованого, гібридного та інших більш модульних підходів.

Отже, архітектура, орієнтована на базу даних, є важливою відомою моделлю, що добре підходить для централізованого накопичення, впорядкування та початкового опрацювання медичної інформації. Водночас у системах безперервного моніторингу стану здоров'я людини її можливості є недостатніми, якщо потрібно забезпечити масштабування, балансування навантаження, інтеграцію аналітики та розмежування функцій між різними рівнями системи. Саме тому підхід, орієнтований на БД, доцільно розглядати як

одну з базових складових складнішої багаторівневої архітектури, а не як завершене універсальне рішення.

1.1.3 Архітектура з поділом на клієнтську та серверну частини

Архітектура з поділом на клієнтську та серверну частини є однією з найвідоміших моделей побудови інформаційних систем, у якій функції надання сервісів і функції їх використання логічно розмежовуються між двома сторонами взаємодії. У роботі [15] клієнт-серверний підхід визначено як організацію системи у вигляді набору інформаційних сервісів, які надаються серверами, та набору клієнтів, що звертаються до цих сервісів. У роботі [16] запропоновано клієнт-серверну архітектуру для МІС нового покоління, орієнтовану на довготривале зберігання та використання клінічних даних.

Практична цінність такої моделі полягає в тому, що вона дозволяє відокремити інтерфейс користувача від централізованого зберігання даних і прикладної логіки. Для систем моніторингу стану здоров'я людини це є важливим, оскільки користувачеві потрібен зручний доступ до результатів вимірювань, тоді як серверна частина повинна забезпечувати приймання, зберігання, оновлення та передачу даних іншим компонентам системи. У роботі [2] під час побудови референтної архітектури МІС окремо виділено рівні подання, прикладної логіки та керування даними, що фактично розвиває і деталізує ідею клієнт-серверного розмежування функцій.

Для медичних застосунків клієнт-серверна модель виявилася особливо корисною в тих випадках, коли необхідно забезпечити віддалений доступ до клінічної інформації або результатів обстежень. У роботі [17] запропоновано систему телерадіології на основі клієнт-серверної архітектури, яка забезпечує швидкий доступ до інтерактивної тривимірної візуалізації для первинної діагностики гострого інсульту за допомогою смартфона. У роботі [10] показано, що архітектури носимих систем моніторингу здоров'я (англ. *wearable health-monitoring systems*) також природно тяжіють до поділу на сторону збору та

передавання даних, проміжний вузол або шлюз і серверну частину, де відбуваються накопичення, обробка та подальше використання інформації.

Як приклад практичного застосування архітектури з поділом на клієнтську та серверну частини можна навести роботу [18], у якій запропоновано систему дистанційного моніторингу на основі пульсоксиметра. У цій системі клієнтську частину утворює мережевий пульсоксиметр, який виконує вимірювання та передає дані, тоді як серверна частина представлена персональним сервером моніторингу, що забезпечує приймання, зберігання, відображення та керування даними в реальному часі. Така організація підтверджує, що клієнт-серверний підхід є практично придатним для систем дистанційного біомедичного спостереження, де необхідно розмежувати функції збору показників і централізованого моніторингу (monitoring).

Разом із тим для сучасних систем безперервного медичного моніторингу простого поділу на клієнтську і серверну частини часто виявляється недостатньо. Якщо кількість джерел даних зростає, а система має підтримувати не лише відображення інформації, а й аналітику, інтеграцію з іншими сервісами, масштабування та підвищені вимоги до безпеки, то двокомпонентна схема поступово ускладнюється. У роботі [2] це фактично відображено через перехід від загальної логіки клієнт-серверної взаємодії до більш деталізованої багаторівневої референтної архітектури. У систематичному огляді [3] також показано, що для МІС сьогодні широко використовуються не лише клієнт-серверні, а й багаторівневі, сервісно-орієнтовані, гібридні та інші більш модульні архітектурні шаблони.

Ще одне обмеження клієнт-серверного підходу полягає в тому, що він задає загальну логіку взаємодії, але сам по собі не визначає, яким чином мають бути організовані серверні сервіси, розподіл навантаження, керування потоками даних від великої кількості пристроїв і взаємодія між кількома підсистемами. Для задач моніторингу стану здоров'я людини цього вже недостатньо, оскільки такі системи одночасно поєднують носимі сенсори (sensors), мобільні пристрої, бази даних, засоби візуалізації та аналітичні модулі. Саме тому клієнт-серверну

архітектуру доцільно розглядати як важливу відому модель, яка створила основу для подальшого переходу до більш гнучких багаторівневих рішень.

Отже, архітектура з поділом на клієнтську та серверну частини є важливим етапом розвитку інформаційних систем медичного призначення. Її перевагами є логічне розмежування функцій, підтримка віддаленого доступу до інформації та зручність інтеграції користувацьких інтерфейсів із централізованими сервісами. Водночас для сучасних систем моніторингу стану здоров'я людини вона не завжди є достатньою в ізольованому вигляді, тому в подальшому доповнюється багаторівневою організацією, окремими сервісами обробки та складнішими механізмами взаємодії між компонентами.

1.1.4 Архітектура розподілених обчислень (Distributed computing)

Архітектура розподілених обчислень належить до відомих моделей побудови інформаційних систем, у яких обробка даних і виконання функцій системи не зосереджуються в одному програмному модулі або на одному сервері, а розподіляються між кількома взаємопов'язаними підсистемами чи вузлами. У роботі [3] зазначено, що програмні архітектури в медичній інформатиці доцільно поділяти, зокрема, на монолітні та розподілені, причому розподілена архітектура передбачає наявність відокремлюваних підсистем, між якими розмежовано виконання окремих функцій. У роботі [19] запропоновано розподілену обчислювальну модель для середовищ Інтернету речей, орієнтовану на спільне використання обчислювальних ресурсів розумних пристроїв у задачах моніторингу стану здоров'я.

Практичне значення такого підходу для систем моніторингу стану здоров'я людини полягає в тому, що він дозволяє розподілити процеси збору, передавання, попередньої обробки, зберігання та аналітичного опрацювання даних між кількома рівнями або вузлами системи. Це особливо важливо в умовах, коли медичні показники надходять від великої кількості сенсорів і мають оброблятися без істотних затримок. У роботі [19] запропоновано розподілену сенсорну та моніторингову інфраструктуру для IoT-середовищ, метою якої є

оптимізація використання біомедичних сенсорів та обчислювальних ресурсів для підтримки розширеного застосування моніторингу. Розвитком цього підходу є робота [20], у якій запропоновано тришарову ієрархію edge, fog та cloud (периферійний, туманний та хмарний рівні архітектури відповідно) для дистанційного моніторингу життєвих показників, де первинні рішення приймаються ближче до джерела даних, а складніша обробка переноситься на вищі рівні.

У медичних застосунках архітектура розподілених обчислень набула особливого значення у зв'язку з розвитком туманних обчислень (туманні обчислення (англ. Fog Computing) – це проміжна, розподілена архітектура між кінцевими пристроями (IoT) та хмарою (cloud), яка обробляє дані ближче до їхнього джерела) та IoMT. У роботі [21] запропоновано архітектуру, орієнтовану на туманні обчислення (або fog-орієнтовану архітектуру), моніторингу для керування сповіщеннями в реальному часі, де проміжний рівень між пристроями та хмарою забезпечує зменшення затримки, локальний аналіз і швидке передавання тривожних повідомлень лікарям або доглядальникам. У роботі [22] розроблено туманно-хмарну (fog-cloud) архітектуру для IoMT, у межах якої швидкий і простіший аналіз виконується на рівні туману (fog-рівні), а більш ресурсомістке машинне навчання – у хмарі (cloud-рівень). Така організація є показовою для медичного моніторингу, оскільки дозволяє поєднати оперативність реагування з можливістю складного аналітичного опрацювання великих масивів даних.

Разом із тим сама по собі ідея розподілу обчислень не усуває всіх труднощів побудови систем медичного моніторингу. Зі збільшенням кількості вузлів, каналів обміну та сервісів зростають вимоги до координації між компонентами, захисту даних, забезпечення якості обслуговування, енергоефективності та надійного керування ресурсами. У роботі [23] підкреслено, що для e-Health-застосунків (e-Health – українська електронна система охорони здоров'я) під час використання туманних обчислень (fog computing) особливо важливими залишаються питання затримки, безпеки,

приватності, енергоефективності та керування ресурсами. У цій же роботі наголошено, що перехід від традиційної хмарної моделі до архітектури, орієнтованої на туманні обчислення, (fog-архітектури) зумовлений саме потребою обробляти великі потоки даних ближче до джерела та краще задовольняти вимоги систем реального часу.

Крім того, у практичних системах дистанційного моніторингу розподілені обчислення зазвичай реалізуються не у вигляді абстрактної мережі рівноправних вузлів, а як ієрархічна або багаторівнева організація. Це добре видно з робіт [20, 21, 22], де розподілена модель поступово уточнюється через використання периферійний, туманного та хмарного рівнів (edge-, fog- та cloud-рівнів), кожен із яких має власну роль у загальному контурі опрацювання медичних даних. Тобто, в сучасних системах моніторингу стану здоров'я людини архітектура розподілених обчислень є не кінцевою формою побудови системи, а радше теоретичною та технологічною основою для подальшого переходу до багаторівневих архітектурних рішень.

Отже, архітектура розподілених обчислень є важливою відомою моделлю, що створює підґрунтя для побудови масштабованих систем медичного моніторингу з рознесенням функцій між кількома вузлами. Її перевагами є можливість розподілу навантаження, наближення частини обробки до джерел даних і підтримка складніших сервісів аналітики. Водночас для практичного використання в системах моніторингу стану здоров'я людини вона потребує подальшої структуризації у вигляді багаторівневих архітектур, у яких чітко визначено функції рівнів збору, проміжної обробки, серверної логіки та хмарного зберігання й аналізу.

1.1.5 Багаторівневі ІоМТ-архітектури як напрям розвитку систем моніторингу

Розвиток систем моніторингу стану здоров'я людини засвідчив, що окреме використання монолітної архітектури, архітектури, орієнтованої на базу даних, клієнт-серверної моделі або розподілених обчислень не повністю задовольняє

вимоги сучасних телемедичних застосувань. У міру поширення Інтернету речей (Internet of Medical Things) увага дослідників змістилася до багаторівневих архітектур, у яких окремо виділяються рівні збору даних, передавання, проміжної або крайової (периферійної) обробки, хмарного зберігання та прикладного використання результатів. У роботі [24] Biomedical IoT (біомедичний інтернет речей) розглянуто як сукупність технологій, архітектурних елементів і сервісних підходів, що забезпечують інтеграцію біомедичних пристроїв, комунікаційних протоколів і аналітичних платформ у єдиному контурі моніторингу. У роботі [25] додатково підкреслено, що поєднання штучного інтелекту та ІоМТ створює основу для побудови інтелектуальних біомедичних систем, у яких сенсорні пристрої, мережеве передавання даних, хмарна інфраструктура та аналітичні модулі функціонують як взаємопов'язані складові єдиної системи охорони здоров'я.

У роботі [26] наведено порівняльний аналіз різних архітектур ІоМТ і підкреслено, що для розумних систем охорони здоров'я (англ. smart healthcare systems) типовою є організація, у якій нижній рівень відповідає за отримання даних від медичних пристроїв, середні рівні – за мережеву взаємодію та обчислювальне розвантаження, а верхні – за аналітику й надання сервісів. Подібний висновок зроблено і в роботі [27], де розглянуто використання мобільні периферійні обчислення (англ. mobile edge computing) в Інтернету медичних речей (англ. Internet of Medical Things) і показано багат шарову побудову архітектури із рівнями збору даних датчиками (пристроями), комунікації, периферійних та хмарних обчислень (англ. sensing, communication, edge та cloud computing відповідно). Це свідчить про те, що багаторівнева організація вже стала не окремим випадком, а одним із провідних напрямів розвитку ІоМТ-архітектур.

Практична цінність багаторівневих ІоМТ-архітектур полягає в тому, що вони дозволяють розміщувати обчислення там, де це є найбільш доцільним. Первинний збір і попереднє опрацювання даних можуть виконуватися безпосередньо поблизу медичного пристрою, тоді як ресурсомістка аналітика,

довготривале зберігання та інтеграція з іншими сервісами переносяться на вищі рівні. У роботі [20] для системи NB-ІоТ-моніторингу (NB-ІоТ-моніторинг – це технологія віддаленого контролю пристроїв у реальному часі, що використовує енергоефективний вузькосмуговий стандарт стільникового зв'язку LPWAN) здоров'я запропоновано ієрархію edge, fog і cloud (периферійний, туманний та хмарний рівні архітектури відповідно), яка дозволяє поліпшити продуктивність і безпеку порівняно з менш гнучкими схемами. У роботі [22] також показано, що поєднання периферійного, туманного та хмарного рівнів у архітектурі віддаленого моніторингу за станом здоров'я створює кращі умови для зменшення затримок, підтримки сервісів реального часу та розмежування функцій між рівнями системи.

Додатковим підтвердженням доцільності багаторівневого підходу є робота [28], у якій розглянуто оптимізацію програмної архітектури мобільних біомедичних сенсорних засобів, що надають аналітичні сервіси для діагностики та підтримки прийняття рішень. У цій роботі запропоновано архітектурні концепції для мобільних і веборієнтованих компонентів, призначених для збору та опрацювання біомедичних даних у великомасштабних системах моніторингу пацієнтів. Це підтверджує, що сучасні системи медичного моніторингу дедалі частіше проєктуються не як ізольовані пристрої, а як багаторівневі програмно-апаратні комплекси, у яких сенсорні вузли, мобільні компоненти та аналітичні сервіси функціонують у межах єдиного архітектурного контуру.

Як приклад прикладної реалізації багаторівневого кіберфізичного підходу можна навести роботу [29], у якій описано біомедичну кіберфізичну систему цілодобового моніторингу функцій легень у пацієнтів із COVID-19. У цій системі нижній рівень утворюють засоби вимірювання та збору фізіологічних показників, проміжний рівень пов'язаний із передаванням даних через смартфон і мережу Інтернет, а верхній рівень реалізує серверну обробку, доступ лікаря до поточних показників, історії спостережень і сповіщень про критичний стан пацієнта. Така організація підтверджує, що багаторівневий розподіл функцій є

практично доцільним не лише для загальних ІоМТ-систем, а й для спеціалізованих кіберфізичних систем моніторингу стану людини.

Як ще один приклад прикладної реалізації багаторівневого кіберфізичного підходу можна навести роботу [30], у якій описано прототип системи моніторингу фізичного стану оператора літального апарата. У цій системі нижній рівень утворюють сенсорні вузли збору фізіологічних показників, проміжний рівень пов'язаний із мікроконтролерною обробкою та передаванням даних, а верхній рівень реалізує серверні функції, вебсервіси та відображення результатів у вебінтерфейсі. Така організація підтверджує, що багаторівневий розподіл функцій є практично доцільним не лише для загальних ІоМТ-систем, а й для спеціалізованих кіберфізичних систем моніторингу стану людини (рисунок 1.2).

Водночас перехід до багаторівневих ІоМТ-архітектур зумовлений не лише прагненням підвищити продуктивність, а й необхідністю забезпечити масштабованість, захист медичних даних, інтероперабельність і стійкість до зростання кількості підключених пристроїв. У роботі [24] наголошено, що сучасні ІоМТ-системи повинні підтримувати обробку даних у реальному часі, надійний обмін між різними платформами та безпечно використання сенсорної інфраструктури в домашніх і доклінічних умовах. У новіших оглядових працях також підкреслюється, що багаторівнева побудова ІоМТ-систем є природною відповіддю на складність медичного середовища, оскільки дозволяє краще розподіляти функції між пристроями, мережевими вузлами та хмарними сервісами.

Отже, багаторівневі ІоМТ-архітектури слід розглядати як закономірний напрям розвитку систем моніторингу стану здоров'я людини. Вони поєднують ідеї розподілених обчислень, клієнт-серверної взаємодії, централізованого зберігання даних та локальної обробки поблизу джерел вимірювань (туманні обчислення). Саме така логіка розвитку архітектурних рішень створює теоретичне підґрунтя для подальшого обґрунтування концептуальної моделі багаторівневої архітектури системи моніторингу стану здоров'я людини.

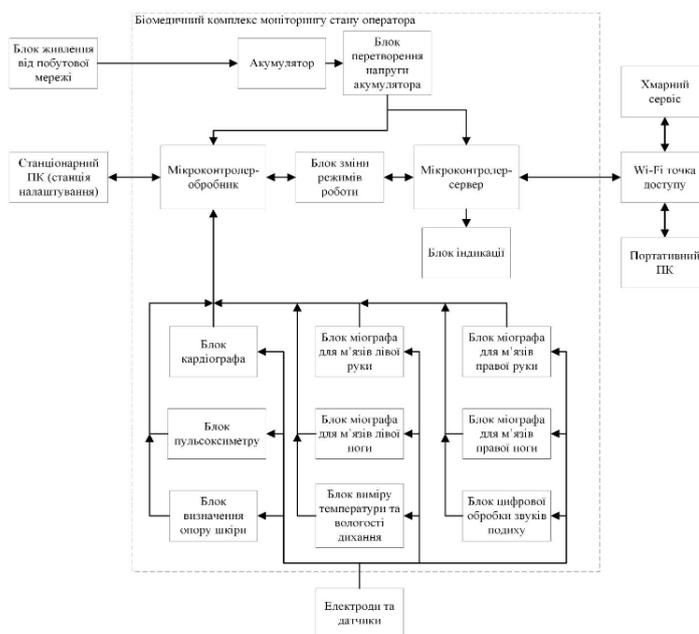


Рисунок 1.2 – Структурна схема комплексу моніторингу здоров'я

Таким чином, сучасна архітектура систем моніторингу еволюціонує у бік створення «невидимих» для користувача, але високотехнологічних рішень. Об'єднання методів інтелектуальної обробки [25, 31], принципів побудови складних біомедичних систем та технологій великих даних формує фундамент для нової якості медичного обслуговування, де процес збору даних є повністю автоматизованим та інтегрованим у щоденний побут пацієнта.

1.1.6 Програмні аспекти та захист даних

Невід'ємною складовою будь-якої архітектури моніторингу, поряд з апаратним забезпеченням, є програмна екосистема, яка відповідає за логіку обробки та збереження накопиченої інформації. Цей аспект детально розглядається у дослідженні [32], присвяченому програмній реалізації систем моніторингу біомедичної інформації. Автори [32] наголошують, що ефективність роботи комплексу залежить не лише від точності датчиків, а й від алгоритмів, закладених у програмне забезпечення. Дослідження підкреслює необхідність створення гнучких програмних оболонок, здатних адаптуватися під різні типи медичних даних, забезпечувати стабільну роботу системи навіть під

час значних навантажень на обчислювальні ресурси та підтримувати моніторинг енергоспоживання ключових вузлів системи.

Однак, в умовах цифровізації медицини на перший план виходить проблема конфіденційності, оскільки біомедичні показники належать до чутливих персональних даних. Це питання ґрунтовно аналізується у праці О. М. Ключко [33], де розглядаються медичні інформаційні системи крізь призму захисту інформації. У статті [33] стверджується, що архітектура сучасної системи моніторингу повинна базуватися на принципах «security by design» (безпека через дизайн). Це означає інтеграцію механізмів шифрування, розмежування прав доступу та захищених протоколів передачі ще на етапі проєктування системи, а не як додаткову опцію.

Синтезуючи підходи, викладені в [32] та [33] можна зробити висновок, що сучасна програмна архітектура біомедичного моніторингу є дворівневою задачею. З одного боку, вона вимагає оптимізованого коду для швидкої обробки сигналів, як це показано в [32], а з іншого – повинна гарантувати відповідність суворим стандартам кібербезпеки, описаним у [33], що є критичним для довіри користувачів до таких технологій.

1.1.7 Приклади реалізації на базі конкретних сенсорів та гаджетів

Теоретичні моделі архітектури знаходять своє практичне втілення у конкретних апаратних розробках, спектр яких варіюється від масової споживчої електроніки до вузькоспеціалізованих вимірювальних комплексів. Потенціал використання загальнодоступних гаджетів розкрито у роботі [34], де описується автоматизована система моніторингу на базі смартгодинника. Дослідники доводять, що інтеграція носимих пристроїв у контур спостереження дозволяє забезпечити безперервність контролю життєвих показників без створення дискомфорту для користувача, фактично роблячи моніторинг фоновим процесом.

Разом з тим, для вирішення складніших клінічних завдань розробляються біотехнічні системи з функцією зворотного зв'язку, які виходять за рамки

пасивного спостереження. Так, І. М. Огородник [35] пропонує комплексне рішення для неінвазивного діагностування та, що важливо, корекції функціонального стану людини. Цю лінію активного управління здоров'ям продовжує дослідження [36], які розробили електронний комплекс для контролю вмісту інсуліну. У праці [36] показано приклад реалізації замкненої системи управління, де моніторинг є проміжною ланкою для прийняття рішення про введення препарату, що критично важливо для пацієнтів з ендокринними порушеннями.

З іншої сторони, фундаментом для надійності всіх вищезгаданих систем виступають інновації на рівні первинних перетворювачів. Зокрема, [37] акцентує увагу на важливості прецизійних вимірювань, пропонуючи використання брегівських сенсорів температури для біомедичного застосування. Робота [37] демонструє, що вдосконалення бази на мікрорівні дозволяє досягти принципово нової якості первинних даних, яка є необхідною умовою для коректної роботи алгоритмів верхнього рівня в будь-якій сучасній архітектурі моніторингу.

1.1.8 Огляд суміжних прикладних рішень у хмарних, мобільних та реабілітаційних медичних системах

Додатково до розглянутих архітектурних і програмно-технічних рішень доцільно врахувати суміжні прикладні дослідження, у яких розкриваються окремі аспекти функціонування сучасних медичних інформаційних систем, пов'язані з використанням хмарних технологій, мобільних застосунків та засобів віддаленої реабілітації. Такі роботи є важливими для розуміння ширшого контексту розвитку систем моніторингу стану здоров'я людини, оскільки демонструють інтеграцію сервісної, комунікаційної та користувацької складових у єдиному цифровому середовищі.

У роботі [38] розглянуто підхід до об'єднання GNSS-даних у хмарному середовищі на основі багатоагентного аналізу протоколів автентифікації в корпоративних логістичних системах. Хоча дослідження безпосередньо не належить до телемедичної галузі, воно демонструє значущість cloud technologies

(хмарні технології) для побудови розподілених інформаційних платформ, де критично важливими є захищений доступ, багатоджерельне злиття даних та координація взаємодії між окремими агентами системи. Для систем моніторингу стану здоров'я такий підхід є показовим з погляду організації хмарної інфраструктури, у межах якої можуть акумулюватися, синхронізуватися та аналізуватися потоки медичних даних від різних пристроїв і користувачів.

У статті [39] досліджено можливості remote rehabilitation (віддалена реабілітація) постінсультних пацієнтів із ушкодженнями нервів кисті та пальців за допомогою бездротових пристроїв в індивідуальних умовах. Практична цінність такого підходу полягає в тому, що реабілітаційні процедури можуть бути перенесені з клінічного середовища в домашні умови без втрати керованості та спостережуваності процесу. Для тематики дисертації це важливо, оскільки підтверджує актуальність побудови спеціалізованих систем, у яких збір фізіологічних даних, передавання результатів і контроль стану пацієнта здійснюються дистанційно з використанням сенсорних і бездротових вузлів.

У роботі [40] запропоновано систему автоматизованого формування черги викликів лікаря пацієнтами в лікарні з візуалізацією через mobile applications (мобільні застосунки). Це дослідження показує, що мобільні застосунки в медичних системах виконують не лише допоміжну, а й функціонально значущу роль, забезпечуючи інтерфейс взаємодії між пацієнтом, медичним персоналом і сервісною логікою системи. У ширшому сенсі такі рішення підтверджують доцільність включення мобільного рівня до архітектури систем моніторингу, де користувач отримує доступ до результатів спостереження, повідомлень, рекомендацій або сервісів зворотного зв'язку.

1.2 Моделі та методи інтеграції інструментарію WEKA до інформаційних систем моніторингу

Інтеграція засобів інтелектуального аналізу даних до систем моніторингу стану здоров'я людини є окремою архітектурною та програмною задачею. У таких системах недостатньо лише накопичувати й передавати результати

вимірювань, а необхідно також забезпечити їх подальше перетворення у придатне для аналізу подання, запуск алгоритмів класифікації, кластеризації, регресії або пошуку асоціативних правил, а також повернення результатів до прикладного контуру системи. У зв'язку з цим особливого значення набуває вибір способу, яким саме середовище аналізу взаємодіє з базою даних, серверною логікою та внутрішнім поданням інформації в системі моніторингу.

У роботі [41] показано, що інструментарій WEKA підтримує роботу не лише з файлами наборів даних, а й із базами даних через JDBC-з'єднання. Разом із тим у роботах [42], [44] наголошено, що для медичних інформаційних систем і систем дистанційного моніторингу важливими є не лише функціональні можливості обробки даних, а й розмежування контурів доступу, безпека роботи з персональною інформацією, контрольованість джерел даних і відтворюваність аналітичних процедур. Саме тому в сучасних інформаційних системах доцільно розглядати не один, а кілька відомих способів інтеграції аналітичного інструментарію, зокрема пряме підключення до бази даних, програмну інтеграцію через серверний код, а також формування стандартизованих наборів даних для їх подальшого окремого аналізу.

1.2.1 Пряме підключення інструментарію WEKA до бази даних через JDBC

Пряме підключення інструментарію WEKA до бази даних через JDBC є одним із найвідоміших способів інтеграції цього середовища з інформаційними системами. Його сутність полягає в тому, що WEKA отримує доступ до реляційної бази даних за допомогою стандартного механізму Java Database Connectivity (підключення до бази даних Java), після чого може виконувати SQL-запити, отримувати вибірки та безпосередньо використовувати їх для подальшого аналізу. У роботі [41] описано, що для такого підключення необхідно налаштувати JDBC-драйвер і параметри з'єднання, а також показано, що в WEKA для цього використовується механізм DatabaseUtils.props, який визначає драйвер, адресу бази даних та інші параметри доступу.

Перевагою цього підходу є його безпосередність. За наявності коректно налаштованого з'єднання дослідник або розробник може працювати з актуальними даними без попереднього ручного експорту у проміжні формати. Для задач експериментального аналізу це зручно, оскільки дозволяє швидко виконувати SQL-вибірки, змінювати склад атрибутів і без додаткових кроків передавати результати до алгоритмів WEKA. У роботі [41] та супровідній документації до WEKA такий спосіб подано як стандартну підтримувану можливість середовища, що робить його природною вихідною моделлю інтеграції аналітичного модуля з підсистемою зберігання даних.

Разом із тим для систем моніторингу стану здоров'я людини пряме підключення аналітичного середовища до робочої бази даних має низку обмежень. По-перше, у такому випадку контур аналітики отримує безпосередній доступ до сховища, у якому містяться персональні та медичні дані, що ускладнює розмежування прав доступу й підвищує вимоги до захисту системи. У роботах підкреслено, що для МІС та середовищ телемедицини питання безпеки, контрольованого доступу та захисту медичних даних є принциповими, а архітектурні рішення мають мінімізувати ризик несанкціонованого доступу до чутливої інформації.

По-друге, пряме підключення WEKA до бази даних тісно пов'язує аналітичний модуль із внутрішньою структурою сховища. Це означає, що зміна схеми таблиць, назв полів або логіки зберігання може вимагати додаткового переналаштування SQL-запитів і параметрів інтеграції. Для багаторівневої системи моніторингу, у якій серверна логіка, база даних, вебінтерфейс і модулі аналізу повинні розвиватися відносно незалежно, така жорстка залежність є небажаною. У роботі [44] показано, що для медичних інформаційних систем (англ. health information systems) більш придатними є архітектури з чітким розмежуванням рівня даних, прикладної логіки та подання, оскільки це підвищує зручність супроводу і гнучкість системи.

По-третє, у практичних системах моніторингу важливо забезпечити повторюваність аналітичних процедур. Якщо аналіз виконується безпосередньо

над поточним станом бази даних, то склад вибірки може змінюватися залежно від моменту виконання запиту, наявності нових записів або змін у даних. Це створює додаткові труднощі для валідації результатів, повторного запуску експериментів та порівняння моделей. Саме тому пряме JDBC-підключення доцільно розглядати як відомий базовий спосіб інтеграції WEKA, який є функціонально можливим і зручним для окремих задач, але не завжди є найкращим рішенням для постійного використання в контурі системи моніторингу стану здоров'я людини.

Отже, пряме підключення інструментарію WEKA до бази даних через JDBC є відомим і технічно доступним способом інтеграції аналітичного середовища з інформаційною системою. Його перевагами є простота безпосереднього доступу до актуальних даних та відсутність обов'язкового проміжного етапу експорту. Водночас для медичних систем такий підхід має обмеження, пов'язані з безпекою, жорсткою залежністю від структури сховища та недостатнім розмежуванням процесів генерації даних і їх окремого аналізу. Саме це обґрунтовує необхідність розгляду інших моделей і методів інтеграції WEKA до багаторівневих систем моніторингу стану здоров'я людини.

1.2.2 Об'єктно-реляційне відображення Django ORM як метод формування вибірок даних

Об'єктно-реляційне відображення є відомим підходом до організації доступу до даних, за якого прикладна програма працює не безпосередньо з SQL-запитами, а з об'єктами прикладної мови програмування, які пов'язані з таблицями та полями бази даних. У роботі [43] зазначено, що в Django модель є основним джерелом опису даних, причому кожна модель зазвичай відповідає одній таблиці бази даних, а кожен атрибут моделі – окремому полю. У роботі [43] також показано, що доступ до даних у Django організується через Manager і QuerySet, які дозволяють будувати вибірки, фільтрувати записи та поетапно уточнювати запити без переходу до низькорівневого SQL.

Для систем моніторингу стану здоров'я людини такий підхід є зручним насамперед тому, що він дозволяє формувати робочі вибірки на рівні прикладної логіки системи, не відкриваючи аналітичному контуру прямий доступ до всієї структури бази даних. У роботі при побудові референтної архітектури медичної інформаційної системи (англ. health information systems; МІС) окремо виділено рівні подання, прикладної логіки та керування даними, що добре узгоджується з використанням ORM як проміжного механізму між прикладним кодом і сховищем. У роботі [44] також підкреслено, що в сучасних лікарняних цифрових платформах шар даних має бути відокремлений від алгоритмічного та прикладного рівнів, щоб забезпечити керованість, масштабування та безпечну інтеграцію аналітики.

Практична цінність Django ORM полягає в тому, що він дозволяє описати структуру даних на рівні моделей, а далі отримувати потрібні записи через фільтрацію, зв'язки між сутностями, агрегацію та інші стандартні механізми. Для задач медичного моніторингу це означає, що з наявної бази можна сформувати не довільну початкову вибірку, а саме той набір атрибутів, який потрібен для подальшого аналізу, наприклад за пацієнтом, періодом, типом вимірювання або групою фізіологічних показників. Офіційна документація Django прямо вказує, що QuerySet є базовим механізмом отримання об'єктів із бази даних, а саме ORM (Object-Relational Mapping; об'єктно-реляційне відображення) надає достатньо засобів для побудови складних запитів без використання raw SQL (сирій SQL – безпосередній SQL-запит до бази даних, написаний вручну). Це робить його придатним як метод формування відтворюваних вибірок даних у серверному контурі системи моніторингу.

Для медичних інформаційних систем це особливо важливо, оскільки робота з даними електронних медичних записів часто ускладнюється великою кількістю таблиць, неоднорідністю полів та необхідністю додаткової валідації. У систематичному огляді [45] показано, що вилучення та підготовка даних з EMR (electronic medical record; електронна медична картка пацієнта) є складним і трудомістким процесом, який часто потребує ручної участі фахівців, а також

стандартизації процесів і інструментів. Використання ORM у такому контексті не усуває всіх труднощів, але дозволяє перенести формування вибірок із рівня ручних SQL-операцій на рівень контрольованого прикладного коду, що є більш зручним для супроводу та повторного використання.

Разом із тим об'єктно-реляційне відображення не слід розглядати як універсальне розв'язання всіх задач інтеграції даних. У роботі [43] показано, що в медичних інформаційних системах застосування ORM може супроводжуватися ускладненням запитів, надлишковими JOIN-операціями та зниженням продуктивності в складних схемах даних. Офіційна документація Django також прямо рекомендує спочатку використовувати ORM, але в окремих випадках визнає потребу в raw SQL і окремо виносить питання оптимізації доступу до бази даних. Це означає, що ORM є сильним і зручним методом формування вибірок, однак для великих або складно нормалізованих медичних сховищ він потребує уважного проєктування моделей, запитів та механізмів оптимізації.

Отже, об'єктно-реляційне відображення Django ORM доцільно розглядати як відомий метод формування вибірок даних у багаторівневих системах моніторингу стану здоров'я людини. Його перевагами є зрозуміле моделювання сутностей предметної області, робота з даними через об'єкти прикладної мови, зменшення залежності від ручного SQL та зручність побудови повторюваних вибірок. Водночас у системах медичного призначення його застосування має супроводжуватися врахуванням продуктивності, складності запитів і необхідності подальшого стандартизованого подання сформованих даних для аналітичних засобів.

1.2.3 Формат Attribute-Relation File як модель подання даних для WEKA

Attribute-Relation File Format є відомою моделлю подання даних, яка історично пов'язана з середовищем WEKA і використовується як один із базових способів опису навчальних вибірок для задач машинного навчання. У документації WEKA зазначено, що ARFF є ASCII-текстовим файлом, який описує множину екземплярів із фіксованим набором атрибутів. Такий файл має

дві основні частини: заголовок, у якому задаються назва відношення, перелік атрибутів і їхні типи, та секцію даних, де подано самі записи. Окремо вказано, що ARFF підтримує числові, номінальні, рядкові, датовані та реляційні атрибути, а також має розширення для розріджених подань і ваг екземплярів. У роботах [41,47] підкреслено, що саме ARFF було створено разом із WEKA як власний формат системи, а сам інструментарій може завантажувати дані з файлів, URL-адрес і баз даних, причому серед підтримуваних форматів ARFF названо першим.

Практична цінність ARFF полягає в тому, що він поєднує в одному текстовому поданні і схему даних, і самі значення атрибутів. Для систем моніторингу стану здоров'я людини це є важливим, оскільки дозволяє сформувати не просто вибірку записів, а формально визначений набір ознак, готовий до подальшого використання засобами WEKA. У роботі [46] описано регіональну систему медичного аналізу даних, у якій дані з лікарняних баз після централізації перетворюються у формат ARFF і далі завантажуються до WEKA для попередньої обробки та аналізу. У роботі [48] для онтологічно-керованої системи побудови діагностичних моделей зазначено, що дані, вилучені з корпоративного клінічного сховища, подаються у формі ARFF-файлу, сумісного з WEKA та іншими інструментами аналізу даних. Це показує, що ARFF використовується не лише як внутрішній формат середовища WEKA, а і як проміжний стандартизований контракт між медичним сховищем і модулем аналізу.

Для задач медичної аналітики ARFF є зручним ще й тому, що дозволяє явно зафіксувати склад ознак, їхній порядок і типи, а отже, забезпечити відтворюваність експериментів. У низці робіт, присвячених медичному машинному навчанню, дослідники спеціально переводять підготовлені дані в ARFF перед запуском моделей у WEKA. Зокрема, в дослідженні аналізу вірусних послідовностей для побудови випадкового лісу дані було конвертовано в ARFF для використання у WEKA Explorer, а самі ARFF-набори подано як супровідні матеріали. В іншій відкритій роботі з аналізу PubMed-запитів для визначення

асоціативних правил наголошено, що базовим форматом вводу в WEKA є саме ARFF, а для аналізу було сформовано ARFF-файл на десятки тисяч записів. Це підтверджує, що ARFF добре підходить для завдань, де важливо чітко описати атрибути, клас і самі екземпляри до початку аналітичної обробки.

Разом із тим формат ARFF не слід розглядати як універсальний засіб представлення будь-яких медичних даних без попередньої підготовки. Його сильна сторона полягає саме у роботі зі структурованими наборами ознак, але для складних клінічних даних, текстових записів, багатотабличних сховищ або динамічних потоків вимірювань спочатку потрібні етапи видалення, агрегування, нормалізації та приведення даних до фіксованої ознакової схеми. Це добре видно з робіт [46, 48], де ARFF не виникає безпосередньо з первинного джерела, а формується після попереднього етапу перетворення даних з XML, клінічного сховища, лог-файлу або іншого представлення. Отже, ARFF доцільно трактувати не як заміну бази даних чи сховищу медичних записів, а як стандартизовану модель подання вже підготовленої вибірки для подальшого аналізу в середовищі WEKA.

Отже, формат Attribute-Relation File Format (ARFF) доцільно розглядати як відому модель подання даних для інструментарію WEKA. Його перевагами є простота текстового представлення, поєднання опису схеми та даних в одному файлі, підтримка різних типів атрибутів і зручність перенесення підготовлених наборів між системою формування даних і середовищем аналізу [41, 47]. Водночас для систем моніторингу стану здоров'я людини використання ARFF передбачає попереднє формування контрольованої вибірки ознак на рівні прикладної логіки системи, що логічно підводить до розгляду способів програмного створення та завантаження таких наборів у контурі інтеграції WEKA.

1.2.4 Програмна інтеграція WEKA з Python через бібліотеку `python-weka-wrapper3`

Програмна інтеграція інструментарію WEKA з Python є відомим способом поєднання засобів машинного навчання WEKA з серверною логікою прикладної системи. У документації [49] зазначено, що бібліотека `python-weka-wrapper3` дозволяє використовувати WEKA з Python 3, запускаючи й керуючи Java Virtual Machine (JVM) через `jpure`. У джерелі [49] також підкреслено, що ця бібліотека дає змогу запускати алгоритми й фільтри WEKA з Python та надає доступ до API WEKA через тонкі обгортки над JNI-викликами. Таким чином, `python-weka-wrapper3` можна розглядати як відомий метод програмної інтеграції WEKA до серверного або дослідницького Python-контексту.

Практична цінність такого підходу полягає в тому, що аналітичні процедури перестають бути ізольованими від прикладного коду системи. У прикладах [49] показано, що через `python-weka-wrapper3` можна не лише завантажувати набори даних і запускати JVM, а й будувати класифікатори, виконувати крос-валідацію з оцінюванням якості, візуалізувати ROC-криві, застосовувати кластеризацію, добування асоціативних правил, атрибутний відбір і навіть працювати з доступом до баз даних через JDBC. Це означає, що WEKA може використовуватися не тільки як окремий інтерактивний інструмент, а і як викликаний із Python аналітичний модуль, вбудований у серверний сценарій обробки даних.

Для систем моніторингу стану здоров'я людини така інтеграція є особливо корисною, коли потрібно поєднати контур збору й підготовки даних із контуром аналітики без ручного втручання користувача. У роботі [44] запропоновано п'ятирівневу архітектуру платформ штучного інтелекту для лікарняних систем, де окремо виділено інфраструктурний, рівень даних, алгоритмічний, прикладний і безпековий рівні. У роботі [42] запропоновано архітектуру дистанційного моніторингу пацієнтів на комерційних хмарних платформах, у якій поєднано збір сигналів, обробку, зберігання та вебдоступ до результатів. За такої організації програмний виклик аналітики з Python є зручним, оскільки дозволяє

вбудовувати запуск моделей WEKA у прикладну логіку системи, не відокремлюючи аналіз від серверного контуру обробки даних.

Ще одна важлива перевага `python-weka-wrapper3` полягає у підтримці відтворюваності аналітичного середовища. У [49] наведено механізми фіксація, встановлення та ініціалізація системи (англ. `freeze`, `install` та `bootstrap` відповідно), які дозволяють зафіксувати набір установлених пакетів WEKA та відтворити відповідне середовище в іншому віртуальному оточенні або на іншій машині. Для систем моніторингу стану здоров'я людини це має практичне значення, оскільки спрощує перенесення аналітичного контуру між тестовим і робочим середовищами та робить результати запуску алгоритмів більш відтворюваними. У роботі [44] також показано, що для МІС відокремлення рівнів і повторне використання архітектурних рішень є важливою умовою супроводжуваності системи.

Разом із тим програмна інтеграція WEKA з Python має і свої обмеження. У документації [49] прямо вказано, що `python-weka-wrapper3` охоплює переважно базову неграфічну функціональність WEKA, а також потребує Python 3, OpenJDK 11 або новішої версії та низки додаткових бібліотек, серед яких обов'язковою є `jpure`. У повідомленні команди WEKA зазначено, що ранні версії обгортки спиралися на `javabridge`, який певний час підтримувався недостатньо належним чином і ускладнював установлення, зокрема для ARM-пристроїв Mac; перехід на `JPure` зробив інсталяцію простішою без істотних змін інтерфейсу. Це означає, що `python-weka-wrapper3` є корисним і практичним методом програмної інтеграції, однак він залишається залежним від Java-середовища, конфігурації JVM і загальної узгодженості програмного середовища.

Отже, програмну інтеграцію WEKA з Python через бібліотеку `python-weka-wrapper3` доцільно розглядати як відомий метод вбудовування інструментарію WEKA у серверний або дослідницький контур обробки даних. Його перевагами є можливість автоматизованого запуску алгоритмів WEKA з Python, поєднання аналітики з прикладною логікою системи та підтримка відтворюваного середовища виконання. Водночас у системах моніторингу стану здоров'я

людини такий підхід потребує належного розмежування рівнів доступу до даних, узгодженого середовища Java і Python та стандартизованого подання сформованих вибірок для подальшого аналізу. Це логічно підводить до розгляду способу, за якого дані спочатку формуються у прикладному контурі, а потім окремо завантажуються в WEKA для обробки.

1.2.5 Завантаження віддалених наборів даних у WEKA як спосіб розділення генерації та аналізу даних

Завантаження підготовлених наборів даних у WEKA з файлу, URL-адреси або іншого зовнішнього джерела є відомим способом використання цього середовища, за якого формування даних і їх подальший аналіз виконуються не в одному й тому самому контурі. У документації до WEKA зазначено, що всі алгоритми середовища працюють із вхідними даними у формі єдиної реляційної таблиці, яка може бути прочитана з файлу або сформована на основі запиту до бази даних. У довіднику до використання інструменту WEKA Workbench також прямо вказано, що файли на локальному комп'ютері не є єдиним джерелом даних для WEKA: користувач може відкрити URL, після чого WEKA завантажує ARFF-файл через HTTP, або відкрити базу даних і отримати вибірку через SQL-запит, яку WEKA читає як ARFF-подання [50].

Такий підхід має практичне значення для систем моніторингу стану здоров'я людини, оскільки дозволяє спочатку сформувати керовану вибірку в межах прикладної системи, а вже потім передати її до аналітичного середовища у стандартизованому вигляді. Офіційна документація ARFF підкреслює, що цей формат призначений для опису набору екземплярів із фіксованим набором атрибутів, а документація до використання інструменту ArffLoader, запропонованого у WEKA, показує, що відповідний завантажувач у WEKA підтримує як пакетний, так і інкрементний режим читання і водночас реалізує інтерфейс URLSourcedLoader [51]. Отже, ARFF можна розглядати як проміжне стандартизоване подання, через яке дані, підготовлені в одній системі, передаються до WEKA для окремого аналізу.

Для медичних інформаційних систем така схема є особливо доцільною, коли потрібно розмежувати рівень формування даних і рівень алгоритмічної обробки. У референтній архітектурі МІС окремо виділяються рівні подання, прикладної логіки та керування даними, а в сучасній п'ятирівневій архітектурі платформ штучного інтелекту (англ. artificial intelligence, AI) для лікарняних систем розділено інфраструктурний, рівень даних, алгоритмічний, прикладний і безпековий рівні. Це означає, що для систем медичного моніторингу архітектурно виправданим є підхід, за якого дані спочатку проходять контрольоване формування на рівні прикладної логіки та рівня даних, а аналітичний модуль отримує вже підготовлений набір ознак.

Перевага завантаження віддалених наборів даних у WEKA полягає в тому, що воно дозволяє розділити два різні процеси: генерацію даних і їх окремий аналіз. На практиці це означає, що система моніторингу може сама визначати склад ознак, часовий інтервал, критерії відбору записів та правила попередньої підготовки, після чого передавати до WEKA вже сформований набір. Така організація зменшує залежність аналітичного модуля від внутрішньої структури робочої бази даних і робить експерименти більш відтворюваними, оскільки аналіз виконується не над змінним поточним станом сховища, а над зафіксованим поданням вибірки. Це є висновком із поєднання можливостей Workbench щодо завантаження даних із URL або бази даних та архітектурного принципу розмежування рівнів у медичних інформаційних системах.

Разом із тим такий підхід має і свої обмеження. Він передбачає наявність проміжного етапу підготовки набору даних, а тому якість подальшого аналізу безпосередньо залежить від коректності сформованого ARFF-подання, повноти атрибутів і актуальності вибірки. Крім того, завантаження підготовленого набору в WEKA є зручним для пакетної або повторюваної аналітики, але саме по собі не замінює механізмів потокової обробки чи постійної синхронізації з робочою базою даних. Це узгоджується з тим, що документація WEKA розглядає ARFF та URL-завантаження як підтримувані джерела даних, а не як повноцінну заміну всіх інших способів інтеграції.

Отже, завантаження віддалених наборів даних у WEKA доцільно розглядати як спосіб інтеграції аналітичного інструментарію до систем моніторингу стану здоров'я людини, за якого формування вибірки й її подальший аналіз виконуються в різних, логічно розмежованих контурах. Його перевагами є стандартизоване подання даних, менша залежність від структури робочої бази, краща відтворюваність експериментів і природне узгодження з багаторівневою архітектурою медичних інформаційних систем. Саме цей підхід є найближчим до моделі інтеграції WEKA, у якій набір даних спочатку формується в системі моніторингу, а потім окремо завантажується та обробляється засобами WEKA.

1.3 Огляд наборів даних для машинного навчання та діагностики раку молочної залози

Ефективність моделей машинного навчання в задачах виявлення й класифікації раку молочної залози значною мірою залежить від характеристик доступних навчальних наборів даних: способі подачі (табличні дані, мамографія, гістопатологія, WSI), розміру й репрезентативності вибірки, якості анотацій та доступності метаданих. Аналіз сучасних публікацій показує, що дослідницьке поле опирається на суміш історичних табличних наборів даних та нових зображувальних колекцій, проте для переходу до клінічної цінності моделям бракує узгоджених стандартів і широких зовнішніх валідацій [52].

Для перевірки класичних алгоритмів (SVM, k-NN, decision trees, ensemble) широко використовуються компактні табличні набори даних, що містять числові морфометричні ознаки та невелику кількість прикладів [52]. Такі набори зручні для швидкого прототипування та реплікації експериментів, але їхня обмежена варіативність і мала клінічна складність не дозволяють отримувати висновки щодо продуктивності моделей у реальних клінічних потоках пацієнтів [53]. Крім того, у деяких роботах досліджується оптимізація відбору ознак і стабілізація результатів за умови мультиколінеарності – питання, важливе для табличних наборів даних з великою кількістю взаємопов'язаних параметрів [54].

Сучасні огляди підкреслюють істотну неоднорідність мамографічних наборів: частина баз складається з цифрових знімків сучасних апаратів, інші – з оцифрованих плівкових зображень; формати, рівні розмітки (BIRADS, bounding boxes, маски) і умови ліцензування суттєво різняться між колекціями. Ця розбіжність у форматах і якості знижує інтероперабельність і ускладнює використання моделей на зовнішніх по відношенню до них даних, що підтверджено аналізом відповідності принципам FAIR (від англ. findable, accessible, interoperable, reusable) – згідно з ним більшість існуючих наборів не задовольняє всім вимогам. У роботах також відзначається перспективність застосування GAN-методів для синтезу додаткових прикладів, проте синтетичні дані мають проходити ретельну перевірку, щоб не вводити систематичні похибки.

Перехід до цифрової патології сприяв появі низки публічних наборів H&E-зображень і WSI (від англ. whole-slide images), які стали базою для розвитку згорткових нейронних мереж у задачах класифікації та сегментації. Огляд відкритих WSI-колекцій фіксує обмежену кількість повноцінно задокументованих наборів і відсутність уніфікованих метаданих, що обмежує можливості різностороннього оцінювання моделей; водночас TCGA-BRCA і ряд інших колекцій залишаються найчастіше вживаними джерелами для досліджень у галузі комп'ютерної патології [55]. При цьому робота з WSI вимагає специфічних технік – патчинг, агрегація локальних прогнозів, управління балансом на рівні патчів та зразків, а також контролю якості анотацій, оскільки часті помилки розмітки призводять до серйозних похибок у поведінці глибоких моделей.

У ряді сучасних праць підкреслюється перевага мультимодальних підходів, які поєднують мамографію, гістологію та клінічні маркери (анамнез, біопсійні висновки, молекулярні дані) [56]. Такі колекції дозволяють створювати більш інформативні моделі для прогнозу ризику, стратифікації пацієнтів і персоналізованих рекомендацій, але проблеми інтеграції, анонімізації і різні формати зберігання обмежують їхнє широке поширення. Окремі дослідження

також демонструють, що зовнішня валідація на незалежних мультимодальних вибірках значно знижує ризик перенавчання і дає адекватнішу оцінку переносимості моделі.

Часті критичні зауваження стосуються дефіциту докладних метаданих і нестачі стандартних форматів розмітки: відсутність уніфікованого опису апаратного забезпечення, параметрів сканування, схем анотації й їхньої верифікації ускладнює повторне використання і порівняння результатів різних робіт. Огляди та емпіричні дослідження натякають на потребу у формалізації вимог до опису наборів даних (включно з версіями анотацій, якістю зображень та демографічною інформацією) для підвищення інтероперабельності і відтворюваності.

Для компенсації обмежень наборів дослідники широко застосовують трансферне навчання із попередньо навчених архітектур, агресивну аугментацію зображень, стратегії боротьби з дисбалансом класів (oversampling, focal loss) та синтетичні підходи (GAN) для розширення вибірки. Водночас кілька робіт наголошують на тому, що ці техніки не замінюють потреби у багатосайтних, стандартизованих і якісно анотованих даних.

З огляду на узагальнення проаналізованих матеріалів, можна запропонувати наступні настанови:

- пріоритет – документувати повні метадані (формат файлів, обладнання, схема анотацій, демографія) і дотримуватися принципів FAIR;
- проводити зовнішню валідацію на незалежних колекціях і публікувати результати з розбиттям за центрами збору даних;
- використовувати мульти-модальні дані, де це можливо, але ретельно документувати процес інтеграції та процедури анотації;
- при синтезі даних (GAN) перевіряти статистичну сумісність з реальними прикладами і оцінювати вплив синтетики на клінічні метрики.

Таким чином, існуючі відкриті та напіввідкриті набори даних створюють міцну базу для розробки алгоритмів підтримки діагностики раку молочної залози, але їхня неоднорідність в форматах, якості анотацій та метаданих

обмежує загальну переносимість моделей. Подальший прогрес потребує системної роботи над стандартизацією, забезпеченням докладної документації й регулярною зовнішньою валідацією – тоді інструменти штучного інтелекту матимуть реальні шанси перейти від демонстраційних проєктів до безпечного застосування в клінічній практиці.

1.4 Формалізація вимог до спеціалізованих систем моніторингу та діагностики

Для оцінки ефективності було проведено порівняння різних варіацій алгоритмів JPEG для стиснення даних на тестових зображеннях мамографії. Результати експерименту наведено в таблиці 1.1.

Таблиця 1.1 – Оцінка ефективності стиснення

| Алгоритм | Тип стиснення | Коефіцієнт стиснення (CR) | Особливості |
|----------------------------|----------------------|----------------------------------|---|
| JPEG | З втратами | 1,26 | Непридатний через втрату деталей |
| JPEG 2000 | Без втрат | 1,69 | Висока обчислювальна складність |
| JPEG-LS | Без втрат | 1,97 | Оптимальний баланс швидкості та CR |
| Deep Learning Lossy | Адаптивні втрати | 1,97 | Високий потенціал, але ще не є стандартом |

Отримані у таблиці 1.1 результати, де JPEG-LS демонструє найкращий компроміс для безвратного підходу, доцільно зіставити з даними незалежних тестів та оглядів, виконаних для потокових сценаріїв. Зокрема, в оглядовій статті наведено, що досяжний рівень стискання JPEG-LS суттєво залежить від складності зображення й може змінюватися в широких межах (від близько 1,2 до понад 7) [93]. Для програмної реалізації автори також показують швидкодію порядку сотень Мбіт/с, що важливо як нижня оцінка для реального часу у потоковій передачі. Для програмної реалізації автори також показують

швидкодію порядку сотень Мбіт/с, що важливо як нижня оцінка для реального часу у потоковій передачі.

Окремо перевірено, що застосування JPEG-LS у запропонованому контурі не спотворює інформативність даних для подальшої діагностики: під час валідації інтегрованого контуру на повному наборі ознак (31–32 атрибути) отримано коефіцієнт кореляції 0,9159.

У підсумку, JPEG-LS забезпечує безвтратне стискання з коефіцієнтом близько 1,97 і зберігає високу узгодженість результатів подальшого інтелектуального аналізу в WEKA, що підтверджує доцільність вибору алгоритму для телемедичних сценаріїв передавання медичних зображень.

Також важливим є аспект стиснення медичних зображень без втрат. Для задач, де неприпустима втрата дрібних діагностично значущих деталей, застосування lossy-алгоритмів обмежене. У проведених експериментах JPEG-LS продемонстрував найкраще співвідношення коефіцієнта стискання та обчислювальної складності для задач безвтратної обробки медичних зображень. Теоретично JPEG-LS базується на низькоскладному прогнозуванні та ентропійному кодуванні в межах сімейства LOCO-I, яке розроблялося як спрощений підхід до використання контекстних моделей, забезпечуючи компроміс між ступенем стиснення та обчислювальною складністю. [55].

1.5 Можливості застосування ПЛІС (FPGA) в системі моніторингу

У задачах спеціалізованих систем моніторингу стану здоров'я людини ключовими є три взаємопов'язані вимоги: обробка даних у реальному часі, обмежені енергоресурси носимих (периферійних) пристроїв та раціональне використання каналу зв'язку під час передачі сигналів і медичних зображень у хмарні сервіси чи аналітичні підсистеми. У таких умовах апаратна платформа має забезпечувати детерміновані часові характеристики, високу пропускну здатність для потоків даних і водночас бути придатною до інтеграції у компактні пристрої.

Одним із практично придатних шляхів є використання програмованих логічних інтегральних схем (ПЛІС, FPGA) як периферійного обчислювального вузла (edge-вузла) поруч із джерелом даних. Сучасні оглядові роботи з тематики FPGA у медицині підкреслюють, що FPGA поєднують енергоефективність, обробку в реальному часі та можливість апаратно-програмного співпроектування (англ. hardware/software co-design) для задач діагностики, медичної візуалізації та моніторингу пацієнтів [56].

Для носимих медичних пристроїв (Wearable Medical Devices) додатково критичними є габарити та автономність. Огляди, присвячені енергоефективним wearable-системам, зазначають, що FPGA можуть бути компромісом між мікроконтролерами (низьке споживання, але обмежена продуктивність) та ASIC (висока енергоефективність, але низька гнучкість), особливо коли потрібна багатоканальна обробка сигналів і вбудовані елементи «легкої» аналітики на пристрої [57].

Щодо реалізованості JPEG-LS на ПЛІС, у відкритих публікаціях та оглядах FPGA-реалізацій підкреслюється, що JPEG-LS є відносно зручним для апаратного втілення завдяки модульній структурі та можливості паралелізації окремих етапів (обробка контекстів, обчислення похибки прогнозу, Golomb-кодування, режими regular/run) [58]. Окрім того, вітчизняні публікації також розглядають потенціал перенесення JPEG-LS на ПЛІС із використанням паралельної обробки та оптимізованих структур даних, що безпосередньо корелює з завданням зменшення витрат часу на стискання в потоці.

Висновки до розділу 1

У першому розділі проаналізовано архітектурні, програмні й технологічні підходи до побудови систем моніторингу стану здоров'я людини. Показано, що монолітна, орієнтована на базу даних, клієнт-серверна та розподілена архітектури мають окремі переваги, але не повною мірою відповідають вимогам сучасних телемедичних та ІоМТ-систем. Обґрунтовано доцільність

багаторівневої архітектурної організації з розмежуванням функцій збору, передавання, обробки, зберігання та аналітики даних.

Проаналізовано відомі моделі та методи інтеграції інструментарію WEKA до інформаційних систем моніторингу. Встановлено, що пряме підключення до бази даних через JDBC є технічно можливим, але має обмеження з погляду безпеки, керованості доступом і відтворюваності експериментів. Показано, що для багаторівневих систем доцільнішим є підхід, за якого вибірки формуються в прикладному контурі засобами Django ORM, перетворюються у формат ARFF і далі окремо завантажуються до WEKA або запускаються через програмну інтеграцію з Python.

У розділі також узагальнено особливості наборів даних для задач машинного навчання та діагностики раку молочної залози. Показано, що наявні відкриті табличні та зображувальні набори даних є корисною основою для розроблення моделей, однак відрізняються за форматом, повнотою метаданих, якістю анотацій і можливістю зовнішньої валідації.

На основі аналізу вимог до спеціалізованих систем моніторингу та діагностики обґрунтовано доцільність безвтратного стискання медичних зображень. Встановлено, що алгоритм JPEG-LS забезпечує найкращий компроміс між коефіцієнтом стискання, обчислювальною складністю та збереженням інформативності даних. Також показано перспективність застосування ПЛІС для попередньої обробки і стискання даних у реальному часі.

Отже, у першому розділі сформовано теоретичне та технологічне підґрунтя для подальшого розроблення концептуальної моделі багаторівневої системи моніторингу стану здоров'я людини.

РОЗДІЛ 2

КОНЦЕПТУАЛЬНІ МОДЕЛІ ТА МЕТОДИ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ

2.1 Математичні моделі опису фізіологічних процесів: моделі Гомперца, просторової гетерогенності, Пеннеса та ін.

Створення високоефективних спеціалізованих систем для моніторингу стану здоров'я, зокрема в контексті ранньої діагностики та супроводу пацієнтів з раком молочної залози (РМЗ), неможливе без ґрунтового математичного базису. Математичні моделі в цьому випадку виступають інструментом, що дозволяє формалізувати складні біофізичні процеси, переходячи від емпіричних спостережень до предиктивного аналізу. Для побудови адекватної моделі моніторингу необхідно синтезувати підходи, що описують три взаємопов'язані рівні патогенезу: кінетику клітинної проліферації, процеси неоваскуляризації та термодинамічні аномалії біологічної тканини.

Ключовим компонентом алгоритмічного забезпечення системи є моделі, здатні прогнозувати динаміку розвитку новоутворення у часовому просторі. Аналіз профільної літератури вказує на те, що динаміка солідних пухлин, до яких відносяться карциноми молочної залози, найбільш коректно описується класом звичайних диференціальних рівнянь.

Хоча класична експоненціальна модель закону Мальтуса є базовою для опису клітинного поділу, її застосування обмежене виключно ранніми, субклінічними стадіями розвитку. Як аргументовано в роботах Панунзі та співавт. [59], після досягнення пухлиною певного критичного об'єму спостерігається ефект насичення, зумовлений вичерпанням локальних ресурсів (гіпоксія, дефіцит нутрієнтів) та просторовими обмеженнями.

2.1.1 Модель опису росту пухлини: феноменологічна модель Гомперца

У сучасній біоматематиці загальноновизнаним стандартом для опису росту РМЗ є феноменологічна модель Гомперца. Її фундаментальна відмінність від

логістичної кривої Ферхюльста полягає в асиметричному характері гальмування росту: питома швидкість проліферації спадає за експоненціальним законом. Математичний запис моделі має вигляд:

$$\frac{dN(t)}{dt} = rN(t) \ln \ln \left(\frac{K}{N(t)} \right), \quad (2.1)$$

де змінні інтерпретуються таким чином:

- $N(t)$ – поточний об'єм біомаси пухлини або кількість клітин у момент часу t ;
- r – коефіцієнт проліферативної активності, що характеризує агресивність пухлини;
- K – гранична ємність середовища (англ. carrying capacity), що визначається рівнем васкуляризації тканини.

Емпіричні дослідження Алгефарі [60] підтверджують, що крива Гомперца забезпечує вищу точність апроксимації експериментальних даних *in vivo* порівняно з альтернативними сигмоїдальними функціями. Це пояснюється тим, що дана модель враховує нелінійне зниження фракції клітин, що діляться, в міру збільшення пухлини, що корелює з формуванням некротичного ядра [61].

Для адаптації моделі до завдань моніторингу терапевтичної відповіді, пропонується використання розширеного рівняння, яке включає члени природної елімінації та зовнішнього впливу:

$$\frac{dN}{dt} = \alpha N^\beta - \delta N - \Phi(t, N), \quad (2.2)$$

де перший доданок αN^β відображає метаболічно-залежний ріст (зазвичай через поверхневий обмін, де $\beta \approx 2/3$), δ – константа апоптозу, а функціонал $\Phi(t, N)$ описує цитотоксичний ефект лікування або імунний контроль [62].

2.1.2 Просторово-часова динаміка та дифузійні процеси: модель просторової гетерогенності

Для задач медичної візуалізації та точної локалізації вогнища патології одновимірних моделей росту недостатньо. Необхідним є перехід до моделі просторової гетерогенності, що реалізується через апарат диференціальних рівнянь у частинних похідних типу «реакція-дифузія».

Розподіл щільності злоякісних клітин $c(x, t)$ у тривимірному просторі молочної залози описується рівнянням балансу маси [63]:

$$\partial c / \partial t = \nabla \cdot (D(x) \nabla c) + S(c) - \lambda c. \quad (2.3)$$

У цьому рівнянні ключову роль відіграє тензор дифузії $D(x)$, який в умовах реальної анатомії є анізотропним. Згідно з сучасними дослідженнями [64], інвазія клітин відбувається переважно вздовж протокових структур та фасцій, зустрічаючи опір у жировій та сполучній тканинах. Член $S(c)$ описує джерело (проліферацію), часто в логістичній формі $\rho c(1 - c/K)$.

Важливим аспектом моделі є врахування взаємодії пухлини з позаклітинним матриксом. Інвазивний ріст супроводжується секрецією протеолітичних ферментів (наприклад, металопротеїназ), що деградують навколишні тканини. Це явище описується системою зв'язаних рівнянь для концентрації ферменту $m(x, t)$ та щільності матриксу $v(x, t)$ [65]:

$$\frac{\partial v}{\partial t} = kmv, \quad (2.4)$$

де k – коефіцієнт деградації. Такий підхід дозволяє відтворити складну морфологію пухлини (спікули, інфільтративні краї), що є критичною діагностичною ознакою під час аналізу медичних зображень в автоматизованих системах [66].

2.1.3 Термодинамічна модель: рівняння біотепла Пеннеса

Оскільки розроблювана система орієнтована на використання неінвазивних методів, зокрема термометрії, математичний апарат повинен встановлювати чіткий причинно-наслідковий зв'язок між внутрішніми патологічними процесами та температурним полем на поверхні тіла.

Фундаментом для вирішення прямої та оберненої задач теплопереносу в біологічних середовищах є рівняння Пеннеса (англ. Pennes Bioheat Transfer Equation) [67]. Для тканини молочної залози воно записується як:

$$\rho c_p \frac{\partial T}{\partial t} = \nabla \cdot (k \nabla T) + \omega_b c_b (T_a - T) + q_{met} + q_{ext}. \quad (2.5)$$

Фізичний зміст рівняння полягає в балансі теплових потоків:

- $\nabla \cdot (k \nabla T)$ – кондуктивна теплопередача через тканину;
- $\omega_b c_b (T_a - T)$ – конвективне тепловідведення (або нагрів) за рахунок мікроциркуляції крові (перфузія), де ω_b – об'ємна швидкість кровотоку;
- q_{met} – ендогенне тепло, що виділяється в результаті метаболізму.

Специфікою злоякісних утворень РМЗ є суттєве підвищення локальної теплопродукції та перфузії. Дослідження [68] демонструють, що термогенез пухлини може перевищувати фонові значення в десятки разів внаслідок прискореного гліколізу та запальних процесів.

В рамках цієї моделі пропонується задавати функцію метаболічного джерела тепла $q_{met}(x)$ залежно від просторового розподілу пухлинних клітин $c(x, t)$, отриманого в попередньому пункті:

$$q_{met}(x) = q_{norm} + \alpha \cdot c(x, t). \quad (2.6)$$

Така інтеграція дозволяє розв'язувати обернену задачу теплопровідності (ІНТР): реконструкцію параметрів пухлини (розмір, глибина, метаболічна активність) на основі аналізу поверхневих термограм [69]. Це є основою функціональної діагностики в розроблюваній системі.

2.1.4 Модель ангіогенного перемикання

Для підвищення прогностичної цінності системи моніторингу необхідно враховувати процес неоангіогенезу – формування нової судинної мережі, що забезпечує перехід пухлини в фазу швидкого росту та метастазування. Математичний опис цього процесу базується на моделі хемотаксису ендотеліальних клітин у відповідь на градієнт ангіогенних факторів (англ. Tumour Angiogenic Factors, TAF), таких як судинний ендотеліальний фактор зростання (англ. Vascular Endothelial Growth Factor, VEGF).

Використовуючи підхід [70], динаміку щільності ендотеліальних клітин $e(x, t)$ можна описати рівнянням:

$$\frac{\partial e}{\partial t} = D_e \nabla^2 e - \nabla \cdot (\chi e \nabla c_{TAF}) + \Gamma(e). \quad (2.7)$$

де другий доданок в правій частині відповідає за хемотаксис клітин до пухлини (проти градієнта TAF), а $\Gamma(e)$ описує їх проліферацію. Включення ангіогенного модуля в загальну модель є критичним, оскільки поява нових капілярів змінює параметр перфузії ω_b у рівнянні Пеннеса, що безпосередньо впливає на теплову картину, яку реєструє система моніторингу.

2.1.5 Гібридний підхід: стохастика та нейромережева апроксимація

Зважаючи на високу варіабельність фізіологічних параметрів між різними пацієнтами, суто детерміновані моделі можуть мати обмежену точність. Тому пропонується застосування гібридної архітектури, де коефіцієнти диференціальних рівнянь (наприклад, r , K , q_{met}) розглядаються як стохастичні змінні з відповідними розподілами ймовірностей.

Для персоналізації моделі або, іншими словами, налаштування під конкретного пацієнта доцільно використовувати Байєсівський висновок, що дозволяє оновлювати апріорні розподіли параметрів під час надходження нових даних моніторингу:

$$P(\text{Дані}) \propto P(\text{Дані}|\theta)P(\theta). \quad (2.8)$$

З точки зору обчислювальної реалізації, перспективним напрямом є використання методів глибокого навчання, [71], а саме фізично-обґрунтованих нейронних мереж (англ. Physics-Informed Neural Networks, PINNs) [72]. Нейронна мережа в такій схемі виступає апроксиматором розв'язку системи диференціальних рівнянь, що дозволяє суттєво прискорити обчислення та реалізувати алгоритми прогнозування безпосередньо вбудованими засобами системи моніторингу, де важливим додатковим аспектом є моніторинг енергоспоживання периферійних обчислювальних вузлів. Такий підхід є важливим і для сценаріїв, у яких потрібен віддалений доступ до результатів моніторингу та прогнозування з боку лікаря або пацієнта без перенесення всієї обчислювальної логіки на клієнтський пристрій.

2.2 Методи підготовки даних та формування інформативних ознак для машинного навчання на прикладі діагностики раку молочної залози

Формування вектору інформативних ознак для інтелектуального аналізу стану пацієнток із РМЗ потребує врахування високої коморбідності та специфічних кардіотоксичних ризиків. Згідно з фундаментальною науковою заявою «Американської асоціації серця» [73], серцево-судинні захворювання (ССЗ) є провідною причиною отримання інвалідності та смертності серед осіб, що пройшли терапію РМЗ. Спільність етіологічних чинників обох патологій, таких як вік, генетична схильність та спосіб життя, детально ілюструється на рисунку 2.1 [73], де відображено перетин факторів ризику, що формують підґрунтя для подальшого моніторингу.



Рисунок 2.1 – Фактори ризику ССЗ та РМЗ [73]

Додаткові складності вносять специфічні механізми впливу, наведені на рисунку 2.2 [73], які вказують на пряму залежність між розвитком кардіоваскулярних ускладнень та агресивністю протипухлинного лікування.

| | Ризик ССЗ | Ризик РМЗ |
|--------------------------------------|-----------|-----------|
| Здорове харчування | ↓ | ↓ |
| Західна дієта | ↑ | ↑ |
| Легке та помірне споживання алкоголю | ↓ | ↑ |
| Червоне/оброблене м'ясо | ↑ | ↑ |
| Фізична активність | ↓ | ↓ |
| Сидячий спосіб життя | ↑ | ↑ |
| Пременопаузне ожиріння | ↑ | ↓ |
| Куріння | ↑ | ↑ |
| Рання менструація | ↑ | ↑ |
| Рання менопауза | ↑ | ↓ |
| Гормонозамісна терапія | ↑ | ↑ |

Рисунок 2.2 – Фактори пов'язані з розвитком ССЗ та РМЗ [73]

Систематизація цих впливів представлена у таблиці 2.1 [73], де класифіковано типи терапії (антрацикліни, таргетні препарати, радіація) та відповідні їм несприятливі ефекти, включаючи аритмії, ішемію та серцеву недостатність.

Таблиця 2.1 – Лікування раку та побічні ефекти з боку серцево-судинної системи [73]

| Тип терапії | Серцево-судинні побічні ефекти |
|--------------------|--|
| Антрацикліни | Дисфункція лівого шлуночка, серцева недостатність, міокардит, перикардит, фібриляція передсердь, шлуночкова тахікардія, фібриляція шлуночків |
| Алкілюючі агенти | Дисфункція лівого шлуночка, серцева недостатність, міокардит, перикардит, артеріальний тромбоз, брадикардія, фібриляція передсердь, надшлуночкова тахікардія |
| Таксани | Брадикардія, блокада серця, шлуночкова ектопія |
| Антиметаболіти | Коронарний тромбоз, спазм коронарних артерій, фібриляція передсердь, шлуночкова тахікардія, фібриляція шлуночків |
| Ендокринна терапія | Венозний тромбоз, тромбоемболія, периферичний атеросклероз, порушення ритму, дисфункція клапанів, перикардит, серцева недостатність |
| Радіотерапія | Ішемічна хвороба серця, кардіоміопатія, захворювання клапанів серця, аритмії |

Проте аналіз літератури свідчить про неоднорідність таких впливів. Як зазначають [74], певні схеми лікування (наприклад, застосування тамоксифену або бевацизумабу у специфічних популяціях) можуть не демонструвати очікуваного зростання артеріальних подій, а подекуди виявляти потенційно кардіопротекторні властивості за рахунок корекції ліпідного профілю. Така варіативність результатів лише підкреслює складність задачі моніторингу та необхідність персоналізованого підходу.

Для вирішення проблеми пропонується впровадження систем раннього виявлення дисфункцій, концептуальна модель яких базується на спрощеній схемі детекції та лікування систолічної дисфункції лівого шлуночка, наведеній на рисунку 2.3 [73]. Використання ЕКГ-параметрів (зокрема HRV та QT-інтервалу) як вхідних ознак дозволяє реалізувати методи предиктивної діагностики, що є критично важливим для запобігання незворотним змінам у міокарді пацієнта.



Рисунок 2.3 – Спрощене зображення виявлення, профілактики та лікування систолічної дисфункції лівого шлуночка при РМЗ [73]

Особливу наукову цінність для розробки методів інтелектуального аналізу становить адаптація клінічних протоколів у предиктивні математичні моделі. Зокрема, на рисунку 2.3 представлено детерміновану схему виявлення та лікування систолічної дисфункції лівого шлуночка, яка може бути формалізована як динамічний граф станів пацієнта. У контексті побудови спеціалізованої системи моніторингу, цей підхід трансформується з реактивного

(лікування за фактом виявлення зниження фракції викиду) у проактивний – заснований на безперервному аналізі біосигналів.

Запропонований метод підготовки ознак базується на використанні ЕКГ-даних, отриманих за допомогою сенсорного модуля AD8232 [75], для ідентифікації латентних маркерів кардіотоксичності. Згідно з таблицею 2.1, такі препарати як трастузумаб та антрацикліни викликають специфічні порушення провідності та варіабельності ритму, що дозволяє виділити наступні ключові ознаки для навчання нейромережових моделей:

- спектральні характеристики HRV: як індикатори порушення вегетативної регуляції серця під дією цитостатиків;
- динаміка інтервалу QT: як маркер ризику виникнення фатальних аритмій, описаних у структурі факторів ризику (рисунок 2.2);
- морфологічні деформації зубця T: що можуть корелювати з ішемічними процесами, згаданими у переліку несприятливих ефектів.

Інтеграція цих ознак у систему виявлення аномалій дозволяє реалізувати механізм раннього попередження. На відміну від традиційного підходу, що базується на спільних факторах ризику РМЗ та серцево-судинних хвороб (рисунок 2.1), інтелектуальна система здатна вловлювати мікрозміни в сигналі ще до того, як пацієнт потрапить у категорію високого ризику згідно зі стандартними протоколами.

Водночас, враховуючи дані [74] щодо гетерогенності відповідей на різні типи терапії, алгоритмічна база повинна підтримувати механізми валідації ознак, щоб диференціювати кардіопротекторні ефекти (наприклад, від тамоксифену) та патологічні зміни. Це обґрунтовує необхідність використання ансамблевих методів класифікації, здатних обробляти нелінійні взаємозв'язки між онкологічним лікуванням та кардіоваскулярною відповіддю організму.

Ефективність сучасних систем моніторингу та діагностики онкологічних захворювань, зокрема раку молочної залози, визначається не лише архітектурою класифікатора, але й якістю та релевантністю вхідного простору ознак. Останні дослідження 2021–2024 років свідчать про зміну парадигми від використання

простих геометричних дескрипторів до комплексного аналізу – радіоміки та автоматичної генерації ознак за допомогою глибоких нейронних мереж [76]. У межах дисертаційної роботи запропоновано підхід до формування простору ознак на основі гібридної стратегії, що поєднує інтерпретовані клінічні ознаки з високорівневими представленнями, отриманими методами трансферного навчання.

2.2.1 Сучасні підходи до попередньої обробки та сегментації

Якість екстракції ознак напряму залежить від точності виділення зони інтересу (англ. Region of Interest, ROI). Традиційні методи порогової сегментації в сучасних роботах поступаються місцем семантичній сегментації на базі глибокого навчання, що дозволяє працювати з зашумленими даними (наприклад, термограмами або УЗД низької роздільної здатності).

Згідно з дослідженнями [77], для препроцесингу медичних зображень найбільш ефективним є використання адаптивних фільтрів згладжування зі збереженням меж (англ. edge-preserving smoothing), у поєднанні з методами нормалізації гістограми (CLAHE). Це дозволяє нівелювати варіабельність умов зйомки, що є критичним для систем моніторингу.

Для автоматичної сегментації пухлини в роботі застосовано архітектуру Attention U-Net. На відміну від класичного U-Net, ця модифікація інтегрує механізми уваги (англ. attention gates), які дозволяють моделі фокусуватися на локальних патологічних структурах, ігноруючи нерелевантні фонові тканини. Як зазначають [78], такий підхід підвищує коефіцієнт Дайса (англ. Dice coefficient) під час сегментації РМЗ до 0,92–0,95, що значно перевищує показники класичних морфологічних методів.

Математично процес сегментації можна представити як відображення вхідного тензора зображення X у бінарну маску Y функцією f_{θ} , параметри якої θ оптимізуються шляхом мінімізації комбінованої функції втрат (Dice Loss + Binary Cross Entropy):

$$L_{total} = L_{Dice} + \lambda L_{BCE} = \left(1 - \frac{2\sum Y_{true}Y_{pred}}{\sum Y_{true} + \sum Y_{pred}}\right) - \lambda \sum [y_{true} \ln y_{pred} + (1 - y_{true}) \ln(1 - y_{pred})]. \quad (2.9)$$

2.2.2 Радіоміка: високопродуктивна екстракція ознак

Сучасним стандартом в описі фенотипу пухлини є радіоміка. Цей метод передбачає вилучення великої кількості кількісних показників з медичних зображень, які неможливо оцінити візуально. У роботі [79] підкреслюється, що радіомні ознаки (англ. radiomic features) мають високу кореляцію з генотипом пухлини та її відповіддю на терапію.

Методика включає розрахунок трьох груп параметрів:

1) ознаки першого порядку (гістограмні): описують розподіл інтенсивності вокселів без врахування просторових відношень. Ключовими є ентропія, ексцес (англ. kurtosis) та асиметрія (англ. skewness). Високий ексцес на термограмах може свідчити про наявність «гарячої точки» з аномальним метаболізмом;

2) текстурні ознаки другого порядку: сучасні протоколи рекомендують використовувати не лише матриці суміжності (англ. Gray-Level Co-occurrence Matrix, GLCM), а й матриці довжин серій рівнів сірого (англ. Gray Level Run Length Matrix, GLRLM) та матриці розмірних зон (англ. Gray level size zone matrix, GLSZM). У своєму дослідженні [80] довели, що ознаки GLRLM, зокрема акцент на довгостроковій перспективі (англ. Long Run Emphasis, LRE), є надійними маркерами гетерогенності злоякісних утворень:

$$LRE = \frac{\sum_{i=1}^{Ng} \sum_{j=1}^{Nr} j^2 p(i,j)}{2a}. \quad (2.10)$$

2.3 Синтез алгоритмів класифікації та кластеризації для виявлення загрозливих та патологічних станів

Розробка інтелектуального ядра системи моніторингу раку молочної залози вимагає переходу від простих детермінованих правил до адаптивних

імовірнісних моделей. Специфіка задачі полягає у необхідності вирішення двох дуальних проблем: класифікації (віднесення стану пацієнта до одного з відомих класів: «норма», «доброякісна патологія», «злоякісне утворення») та кластеризації (виявлення прихованих груп ризику або нових підтипів патологій без попередньої розмітки). У прикладному аспекті такі моделі мають бути придатними до інтеграції в системи, де первинні дані надходять від IoT-пристроїв, що забезпечують безперервне або періодичне збирання медичних показників. У такому контексті системи моніторингу доцільно розглядати як інтелектуальні системи, здатні адаптувати процедури виявлення патологічних станів до структури вхідних даних і умов спостереження.

2.3.1 Ансамблеві методи класифікації для гетерогенних даних

У системах моніторингу вхідний вектор ознак часто є гетерогенним: він містить кількісні показники (радіомні ознаки зображень), категоріальні дані (фактори ризику, генетичні маркери) та часові ряди (динаміка температури). Для таких задач, як демонструють дослідження [81], монолітні моделі поступаються ансамблевим методам на основі градієнтного бустингу.

В рамках дисертаційної роботи пропонується використання архітектури стекового узагальнення (англ. Stacked Generalization, Stacking). На першому рівні (base learners) навчаються різноманітні моделі, що компенсують слабкі сторони одна одної, а саме:

1) LightGBM: використовується для обробки числових радіомних ознак завдяки алгоритму односторонньої вибірки на основі градієнта (англ. Gradient-based One-Side Sampling, GOSS), який забезпечує високу швидкість навчання на великих вибірках;

2) CatBoost: оптимізований для роботи з категоріальними ознаками анамнезу без необхідності попереднього One-Hot кодування, що запобігає розростанню розмірності простору ознак [82].

На другому рівні (meta-learner) використовується логістична регресія або мілководна нейромережа, яка агрегує передбачення базових моделей. Математично це виражається як пошук оптимальної функції композиції Φ :

$$\hat{y} = \Phi(f_{LGBM}(x), f_{CatBoost}(x), \dots, f_n(x)). \quad (2.11)$$

Такий підхід, згідно з роботою [83], дозволяє підвищити метрику AUC-ROC на 4–6 % порівняно з поодинокими класифікаторами, зменшуючи дисперсію помилки прогнозування.

2.3.2 Глибокі архітектури: Vision Transformers та механізми уваги

Для класифікації візуальних даних (термограм, мамограм) революційним кроком останніх років стало впровадження архітектури зорових трансформерів (англ. Vision Transformers, ViT). На відміну від згорткових мереж, які мають обмежене рецептивне поле (локальне вікно згортки), трансформери аналізують глобальні залежності між усіма ділянками зображення.

У системі пропонується імплементація гібридної моделі свін трансформера (англ. Swin Transformer або Shifted Window Transformer), описаної в дослідженні [84] та адаптованої для медичних задач у 2024 році. Ця архітектура використовує механізм «зсувного вікна» (англ. shifted windows) для обчислення функції уваги (англ. self-attention), що дозволяє будувати моделі взаємозв'язків між пухлиною та віддаленими лімфатичними вузлами або судинною сіткою.

Функція уваги обчислюється за формулою:

$$Attention(Q, K, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}} 1 + B\right) V, \quad (2.12)$$

де Q , K , V – матриці запитів, ключів та значень відповідно, а B – відносне позиційне кодування. Включення позиційного кодування є критичним для медичної діагностики, оскільки локалізація патології (наприклад, верхній зовнішній квадрант) має прогностичне значення.

Для боротьби з проблемою дисбалансу класів (кількість випадків патології значно менша за норму), яка є типовою для медичних наборів даних, стандартна функція втрат Cross-Entropy замінюється на Focal Loss [85]:

$$FL(\rho_t) = -\alpha_t (1 - \rho_t)^\gamma \log \log (p_t). \quad (2.13)$$

Параметр фокусування $\gamma > 0$ зменшує вагу «легких» прикладів (фону), змушуючи модель концентруватися на складних для класифікації випадках пухлин на ранніх стадіях.

2.3.3 Кластеризація для стратифікації ризиків

У задачах моніторингу часто виникає ситуація, коли розмічені дані відсутні або необхідно виявити нові закономірності перебігу хвороби. Для цього застосовуються методи кластеризації.

Замість класичного методу k-середніх, який погано працює з кластерами складної форми, пропонується використання алгоритму глибокої вбудованої кластеризації (англ. Deep Embedded Clustering, DEC). Як показано в роботі [86], DEC одночасно навчає автоенкодер для отримання стиснутого латентного представлення даних z та оптимізує центри кластерів у цьому просторі.

Цільова функція DEC визначається як дивергенція Кульбака-Лейблера (KL) між розподілом м'яких присвоєнь Q і допоміжним цільовим розподілом P :

$$L = KL(P \parallel Q) = \sum_i \sum_j \rho_{ij} \log \log \frac{\rho_{ij}}{q_{ij}}. \quad (2.14)$$

Цей підхід дозволяє автоматично групувати пацієнтів за фенотипами (наприклад, «швидкоростучі», «повільні», «запальні») на основі багатовимірних даних, не вимагаючи втручання лікаря на етапі навчання.

2.3.4 Виявлення аномалій для раннього попередження

Специфіка системи моніторингу полягає в необхідності виявлення відхилень від індивідуальної норми пацієнта ще до того, як вони сформується у чітку клінічну картину. Це задача One-Class Classification.

Ефективним рішенням, згідно з дослідженнями [87], є використання варіаційних автоенкодерів (англ. Variational Autoencoder, VAE) або GANomaly (Generative Adversarial Networks for Anomaly Detection). Модель навчається виключно на даних здорового стану конкретного пацієнта (або популяційній нормі), намагаючись реконструювати вхідні дані.

За умови появи патологічних змін (наприклад, локальна гіпертермія або зміна ехогенності тканини) помилка реконструкції (англ. Reconstruction Error) різко зростає:

$$E_{rec} = \| x - G(E(x)) \|_2. \quad (2.15)$$

Якщо E_{rec} перевищує адаптивний поріг, система генерує сигнал тривоги про «загрозливий стан». Перевага цього методу полягає в його здатності детектувати будь-яку аномалію, навіть ту, яка не була присутня в навчальній вибірці (zero-shot learning).

2.3.5 Пояснюваність рішень (Explainable AI, XAI)

Критичною вимогою до медичних алгоритмів є інтерпретованість. Лікар не може покладатися на рішення «чорної скриньки». Тому синтезована алгоритмічна схема включає модуль XAI.

Для візуальних моделей використовуються методи Score-CAM (Class Activation Mapping), які підсвічують на зображенні зони, що найбільше вплинули на рішення нейромережі (теплові карти уваги). Для табличних даних та ансамблів застосовується метод SHAP (SHapley Additive exPlanations) заснований на теорії кооперативних ігор. Згідно з [88], значення Шеплі дозволяють оцінити локальний внесок кожної ознаки (наприклад, вік,

температура, розмір ущільнення) у фінальну ймовірність малігнізації для конкретного пацієнта. Це перетворює систему з «оракула» на прозорого асистента.

2.3.6 Федеративне навчання

З урахуванням вимог конфіденційності медичних даних у роботі застосовано підхід розподіленого навчання класифікаторів. Для цього використано алгоритм федеративного навчання FedAvg, за яким локальні моделі навчаються на пристроях користувачів або в локальних базах клінік, а на центральний сервер передаються лише оновлення параметрів (градієнти ваг), без передавання сирих даних пацієнтів.

Дослідження [89] підтверджують, що федеративне навчання досягає точності, порівнянної з централізованим, при цьому забезпечуючи повну відповідність стандартам GDPR/HIPAA. Рівняння оновлення глобальних ваг має вигляд:

$$w_{t+1} = w_t - \eta \sum_{k=1}^K \frac{\eta_k}{\eta} g_k, \quad (2.16)$$

де η_k – обсяг даних k -го клієнта, g_k – локальний градієнт.

Висновки до розділу 2

У другому розділі дисертаційної роботи розглянуто та систематизовано підходи до формування математичного й алгоритмічного забезпечення спеціалізованої системи моніторингу стану здоров'я людини на прикладі раку молочної залози. Проведене дослідження спрямоване на поєднання класичних методів побудови математичних моделей із сучасними інструментами аналізу біомедичних даних та машинного навчання.

У підрозділі 2.1 сформовано комплекс моделей, що відображають основні фізіологічні та біофізичні процеси росту й поширення пухлин. Показано, що експоненціальні моделі придатні лише для ранніх стадій, тоді як модель Гомперца точніше описує нелінійну динаміку росту за умов ресурсних

обмежень. Урахування просторової неоднорідності, дифузії та взаємодії з позаклітинним матриксом дало змогу детальніше описати інвазивний характер злоякісних утворень. Включення рівняння біотепла Пеннеса дозволило встановити зв'язок між метаболічною активністю пухлини й температурними аномаліями на поверхні тіла, що важливо для неінвазивної термометрії. Окремо обґрунтовано доцільність урахування ангиогенезу та стохастичної природи фізіологічних параметрів для персоналізації моделей.

У підрозділі 2.2 розглянуто підходи та методи формування інформативного простору ознак для задач інтелектуального аналізу медичних даних. Показано, що точність подальшої класифікації значною мірою визначається етапами попередньої обробки даних і сегментації зони інтересу. Застосування сучасних архітектур глибокого навчання для сегментації зменшує вплив шумів і варіабельності умов отримання зображень. Запропонований метод екстракції радіомних ознак забезпечує кількісний опис структурної та функціональної неоднорідності пухлин, що підвищує діагностичну інформативність вхідних даних.

У підрозділі 2.3 узагальнено методи побудови алгоритмів класифікації, кластеризації та виявлення аномалій для гетерогенних біомедичних даних. Обґрунтовано доцільність ансамблевих підходів для зменшення похибок і підвищення стійкості моделей. Показано перспективність трансформерних архітектур для аналізу медичних зображень і некерованого навчання для стратифікації пацієнтів за рівнем ризику. Зазначено, що механізми пояснюваності підвищують довіру до автоматизованого аналізу, а федеративне навчання сприяє конфіденційності даних у розподілених системах.

У цілому результати, отримані в другому розділі, формують логічно узгоджену наукову основу для подальшої реалізації програмно-алгоритмічних компонентів системи моніторингу. Запропоновані моделі та методи створюють передумови для підвищення точності раннього виявлення патологічних змін і можуть бути використані як базис для подальших експериментальних досліджень та практичної апробації розроблюваної системи.

РОЗДІЛ 3

ПРАКТИЧНА РЕАЛІЗАЦІЯ МОДЕЛЕЙ І МЕТОДІВ ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ З ВИКОРИСТАННЯМ WEKA

3.1 Модель інтеграції інструментарію WEKA до багаторівневої системи моніторингу стану здоров'я людини

Для включення засобів інтелектуального аналізу даних до багаторівневої системи моніторингу стану здоров'я людини у роботі запропоновано модель інтеграції інструментарію WEKA, яка узгоджується з архітектурою системи на базі Python/Django та хмарних технологій і забезпечує відокремлення процесів збирання, зберігання і забезпечує відокремлення процесів збирання медичних даних, у тому числі з апаратних вузлів на основі ПЛІС, їх зберігання і підготовки від етапу подальшої аналітичної обробки.. Необхідність такої інтеграції зумовлена тим, що сучасні системи моніторингу на основі концепції ІоМТ, зокрема в задачах віддаленої реабілітації, забезпечують накопичення великої кількості показників стану здоров'я, однак для підтримки діагностичних рішень ці дані мають бути перетворені у придатний вигляд для алгоритмів машинного навчання та методів штучного інтелекту. Основні положення запропонованої моделі апробовано в авторській публікації [90, 91].

Під час формування моделі було розглянуто три можливі способи інтеграції інструментарію WEKA до системи моніторингу. Перший спосіб передбачає пряме підключення WEKA до бази даних засобами JDBC. Такий підхід є технічно можливим, однак для медичної системи він пов'язаний із необхідністю надання прямого доступу до СКБД, що ускладнює ізоляцію аналітичного контуру від робочого контуру зберігання персоналізованих даних. Другий спосіб полягає у виклику WEKA безпосередньо з Python-коду через бібліотеку `python-weka-wrapper3`. Цей варіант розширює можливості автоматизації аналізу, але потребує додаткового серверного оточення, пов'язаного з підтримкою JVM та суміжних бібліотек, що збільшує навантаження на серверну частину системи. Третій спосіб, який покладено в

основу запропонованої моделі, передбачає формування вибірки в системі моніторингу та її подальше окреме подання для обробки засобами WEKA. Саме цей варіант у роботі прийнято як основний [90].

Обраний у роботі варіант інтеграції інструментарію WEKA, заснований на формуванні вибірки в системі моніторингу та її подальшому поданні у форматі ARFF для окремої аналітичної обробки, схематично наведено на рисунку 3.1.

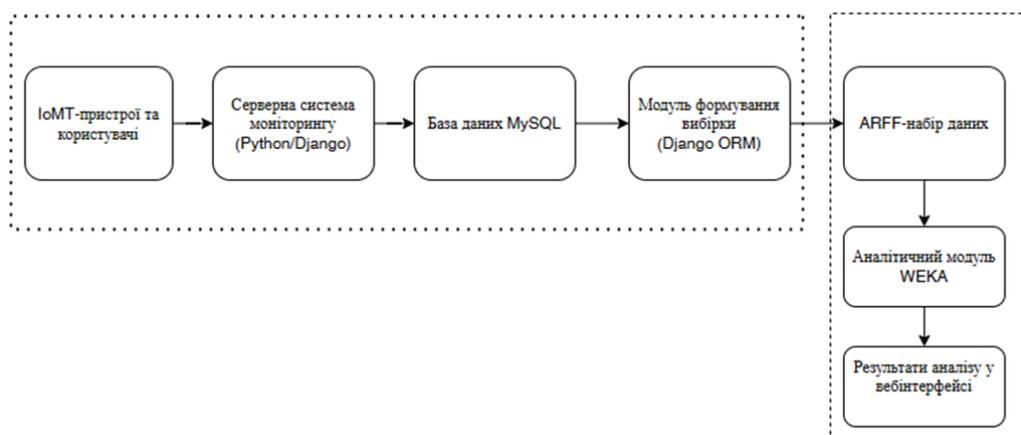


Рисунок 3.1 – Модель інтеграції інструментарію WEKA до багаторівневої системи моніторингу стану здоров'я людини

Як показано на рисунку 3.1, запропонована модель передбачає формування цільової вибірки в основному контурі системи моніторингу засобами Django ORM, її подання у форматі ARFF та подальшу обробку в аналітичному модулі WEKA з передаванням результатів у вебінтерфейс або мобільні застосунки [92].

Сутність запропонованої моделі полягає в тому, що первинні дані від IoT-пристроїв, інших пристроїв моніторингу, мобільних застосунків, зокрема від вузлів на основі ПЛІС, та користувацьких підсистем накопичуються в базі даних багаторівневої системи, після чого на рівні прикладної логіки формується контрольована вибірка ознак для подальшого аналізу. Формування такої вибірки виконується засобами Django ORM, що дає змогу працювати з даними через об'єктно-орієнтовані моделі, виконувати фільтрацію, агрегацію та поєднання пов'язаних сутностей без прямого втручання в структуру запитів до бази даних. У результаті з робочої БД виділяється не вся множина записів, а лише той набір атрибутів, який потрібний для конкретної аналітичної задачі [92, 93].

Наступним етапом моделі є подання сформованої вибірки у форматі ARFF, який використовується як проміжна форма обміну між системою моніторингу та середовищем WEKA. Таке подання дає змогу відокремити робочу базу даних від аналітичного контуру, стандартизувати структуру набору ознак і забезпечити відтворюваність подальшого аналізу. Після формування ARFF-набору він може бути переданий до WEKA для класифікації, кластеризації, регресії або добування правил. Таким чином, інструментарій WEKA у запропонованій моделі функціонує не як компонент прямого доступу до БД, а як окремий аналітичний модуль, що працює з уже підготовленими наборами даних [90, 94].

Перевагою такої моделі є розмежування функцій між компонентами системи. Багаторівнева система моніторингу відповідає за збирання, зберігання та попередню організацію даних, тоді як WEKA використовується для виконання інтелектуального аналізу над сформованими наборами ознак. Це зменшує залежність аналітичного модуля від внутрішньої структури робочої бази даних, спрощує підготовку вибірок для різних задач та узгоджується з принципом розподілу рівнів у багаторівневій архітектурі. Водночас запропонована модель не замінює потокову обробку даних у режимі реального часу, а орієнтована насамперед на контрольоване формування й подальший аналіз вибірок у межах діагностичних і дослідницьких сценаріїв. За такої організації багаторівнева система моніторингу набуває властивостей інтелектуальної системи, у якій функції збирання, підготовки та аналітичної інтерпретації медичних даних розподілено між взаємопов'язаними програмними модулями [95].

Отже, у роботі запропоновано модель інтеграції інструментарію WEKA до багаторівневої системи моніторингу стану здоров'я людини, у якій формування вибірки здійснюється засобами Django ORM у межах основної системи, подання даних реалізується у форматі ARFF, а подальший аналіз виконується засобами WEKA в окремому аналітичному контурі. Така організація інтеграції є доцільнішою за пряме підключення WEKA до БД та менш навантажує серверний або хмарний контур системи, ніж повне вбудовування інструментарію WEKA у серверну логіку застосунку.

3.2 Метод формування ARFF-наборів даних для WEKA

Для забезпечення коректної передачі даних із багаторівневої системи моніторингу до аналітичного модуля WEKA у роботі [90] використано метод формування робочих наборів даних у форматі ARFF. Необхідність цього методу зумовлена тим, що WEKA використовує ARFF як один із базових форматів вводу/виводу, а сам формат призначений для подання множини екземплярів із фіксованим набором атрибутів у текстовому вигляді. У структурі ARFF-файлу задаються назва відношення, перелік атрибутів та секція даних, що робить цей формат придатним для контрольованого перенесення вибірок із серверної системи до середовища інтелектуального аналізу [92, 94].

Запропонований метод орієнтований на використання в межах серверного контуру системи моніторингу, побудованої на базі Python/Django. Його вхідними даними є записи, що зберігаються в робочій базі даних MySQL та відбираються відповідно до поставленої аналітичної задачі. На відміну від прямого передавання всієї структури бази даних до WEKA, метод передбачає попереднє формування цільової вибірки засобами Django ORM. Це дає змогу виділити лише ті атрибути, які є необхідними для подальшої класифікації, кластеризації або побудови правил, та відокремити аналітичний контур від робочого контуру зберігання персоналізованих даних.

Метод формування ARFF-набору доцільно подати як послідовність із чотирьох основних етапів. На першому етапі виконується визначення складу атрибутів відповідно до прикладної задачі аналізу. Для цього з БД відбираються лише релевантні поля, наприклад демографічні характеристики, фізіологічні показники та службові ознаки, що мають діагностичне значення. На другому етапі здійснюється формування вибірки засобами Django ORM, що забезпечує об'єктно-орієнтоване отримання записів із пов'язаних таблиць без прямого втручання в SQL-структуру запитів. На третьому етапі виконується опис структури ARFF-файлу, тобто задаються відношення (relations), атрибути (attributes) та їхні типи. На четвертому етапі підготовлені записи переносяться до

секції даних (data), унаслідок чого формується завершений ARFF-набір, придатний для завантаження в WEKA.

З погляду структури результату метод забезпечує побудову ARFF-файлу, що містить дві основні частини: заголовок і дані. У заголовку визначається назва набору даних і перелік атрибутів із зазначенням їхніх типів. У секції даних розміщуються самі екземпляри спостережень. ARFF підтримує числові, номінальні, рядкові та датовані атрибути, що дозволяє використовувати його для подання різнорідних медичних даних у межах однієї структури. Саме ця властивість робить формат придатним для задач моніторингу стану здоров'я людини, де в одному наборі можуть поєднуватися кількісні вимірювання, категоріальні ознаки та службова інформація.

Як практичний засіб формування ARFF-подання розглянуто використання бібліотеки `liac-arff`, де структура набору задається через словник Python із ключами `description` (опис), `relation` (відношення), `attributes` (атрибут) і `data` (дані, значення). У такому разі сформований ARFF-набір виступає стандартизованим проміжним поданням між модулем збору й зберігання даних та аналітичним модулем WEKA [90, 92].

Перевагою запропонованого методу є те, що він забезпечує відтворюваність аналітичних експериментів, спрощує повторне використання вибірок і зменшує залежність модуля аналізу від внутрішньої структури робочої бази даних. Після формування ARFF-набір може бути безпосередньо завантажений у WEKA Explorer або використаний в інших інструментах обробки даних, якщо потрібне подальше порівняння моделей чи додаткова аналітика. Отже, метод формування ARFF-наборів є необхідною процедурною складовою інтеграції інструментарію WEKA до багаторівневої системи моніторингу стану здоров'я людини, зокрема для сценаріїв віддаленої реабілітації.

3.3 Метод попередньої оцінки цінності набору CBCD для первинного скринінгу

Для оцінювання придатності набору Coimbra Breast Cancer Dataset [95] до задач первинного скринінгу у роботі запропоновано метод попередньої оцінки його цінності, який поєднує ранжування біомаркерів за групою оцінювачів WEKA, статистичне виділення підгруп релевантності ознак та аналіз зашумленості й викидів у даних. Такий підхід дає змогу ще до побудови підсумкового класифікатора встановити, які атрибути є найбільш інформативними, які мають допоміжне значення, а також наскільки надійними є самі дані для подальшого діагностичного використання. Основні положення цього підходу апробовано в авторській публікації [96, 97].

Набір CBCD [98] є компактним двокласовим набором даних, сформованим для задачі первинної діагностики раку молочної залози. Він містить 116 спостережень, із яких 52 належать до класу healthy (здоровий), а 64 – до класу patients (пацієнти). Ознаковий опис набору включає дев'ять числових атрибутів, отриманих із рутинних аналізів крові та антропометричних вимірювань: Age (вік), BMI (індекс маси тіла), Glucose (рівень глюкози в крові), Insulin (рівень інсуліну), HOMA, Leptin (рівень лептину), Adiponectin (рівень адипонектину), Resistin (рівень резистину), MCP1 (показник, який вказує на наявність в крові запального білка цитокіну). Саме поєднання невеликої кількості відносно доступних показників робить цей набір привабливим для задач попереднього скринінгу, однак його невеликий обсяг і потенційна зашумленість вимагають попередньої перевірки перед використанням у діагностичних моделях [96].

Запропонований метод реалізується послідовно. На першому етапі виконується оцінювання релевантності дев'яти атрибутів за допомогою кількох незалежних оцінювачів WEKA. На другому етапі результати окремих оцінювачів узгоджуються через побудову рангових векторів для кожної ознаки та визначення її медіанного рангу. На третьому етапі на основі відстаней між ранговими векторами виявляються групи ознак із близькою релевантністю. На четвертому етапі перевіряється статистична значущість такого групування.

Нарешті, на п'ятому етапі оцінюється зашумленість набору за допомогою процедур виявлення викидів і локальних аномалій. Саме сукупність цих процедур і становить метод попередньої оцінки цінності набору СВCD.

Для початкового ранжування атрибутів було використано вісім оцінювачів WEKA: Correlation Ranking Filter (фільтр ранжування за кореляцією), Gain Ratio Feature Evaluator (оцінювач ознак за коефіцієнтом приросту інформації), Information Gain Ranking Filter (фільтр ранжування за інформаційним приростом), OneR Feature Evaluator (оцінювач ознак за методом OneR за якістю найпростішого правила класифікації, побудованого на її основі), Relief Ranking Filter (фільтр ранжування за методом визначення важливості ознак за їхньою здатністю розрізняти близькі об'єкти різних класів), SVM Feature Evaluator (оцінювач ознак на основі методу опорних векторів), Symmetrical Uncertainty Ranking Filter (оцінює силу зв'язку між ознакою та класом на основі симетризованої міри інформаційної залежності) та J48 Feature Evaluator (оцінювач ознак на основі дерева рішень J48). Кожний із них формує власний порядок атрибутів за спаданням релевантності, тому для одержання більш стійкого результату доцільно аналізувати не окремий рейтинг, а їх узгоджену сукупність.

Результати ранжування дев'яти атрибутів набору СВCD за вісьмома оцінювачами WEKA наведено на рисунку 3.2.

| # | WEKA Evaluator | Attribute Ranking (Best First) |
|---|--|---|
| 1 | Correlation Ranking Filter | Glucose, HOMA, Insulin, Resistin, BMI, MCP1, Age, Adiponectin, Leptin |
| 2 | Gain Ratio feature evaluator | Age, Glucose, HOMA, Leptin, Resistin, BMI, Adiponectin, Insulin, MCP1 |
| 3 | Information Gain Ranking Filter | Glucose, HOMA, Age, Leptin, Resistin, BMI, Adiponectin, Insulin, MCP1 |
| 4 | OneR feature evaluator | Glucose, Age, Resistin, HOMA, Insulin, BMI, Adiponectin, Leptin, MCP1 |
| 5 | RELIEF Ranking Filter | Age, Glucose, Resistin, Insulin, HOMA, BMI, Adiponectin, Leptin, MCP1 |
| 6 | SVM feature evaluator | Glucose, BMI, Resistin, Insulin, HOMA, Age, Leptin, Adiponectin, MCP1 |
| 7 | Symmetrical Uncertainty Ranking Filter | Glucose, Age, HOMA, Leptin, Resistin, BMI, Adiponectin, Insulin, MCP1 |
| 8 | J48 feature evaluator | Glucose, Age, Resistin, HOMA, MCP1, BMI, Adiponectin, Leptin, Insulin |

Рисунок 3.2 – Ранжування атрибутів набору СВCD за оцінювачами WEKA

Після отримання восьми незалежних порядків для кожної ознаки було сформовано ранговий вектор із восьми компонент. Наприклад, для атрибута Glucose відповідний вектор має вигляд (1,2,1,1,2,1,1,1), тобто в більшості оцінювачів саме цей показник потрапляє на найвищі позиції. Подальше

визначення медіанного рангу для кожної ознаки дало змогу впорядкувати весь набір атрибутів у такій послідовності: Glucose, Age, HOMA, Resistin, BMI, Insulin, Leptin, Adiponectin, MCP1. Уже на цьому етапі видно, що перші чотири атрибути мають стійко вищу релевантність, тоді як три останні займають найнижчі позиції майже в усіх використаних оцінювачах.

Однак для прийняття рішення про подальше скорочення ознакового простору недостатньо лише лінійного впорядкування. Важливо встановити, чи утворюють ознаки природні підгрупи близької релевантності. З цією метою для рангових векторів було побудовано матрицю мангеттенських відстаней. Вибір саме цієї метрики зумовлений тим, що компоненти ранг-векторів є цілими числами, а сама метрика добре відображає сумарну різницю між двома порядками. Для виявлення близьких за релевантністю ознак побудовано матрицю мангеттенських відстаней між їх ранговими векторами; її візуалізацію наведено на рисунку 3.3.

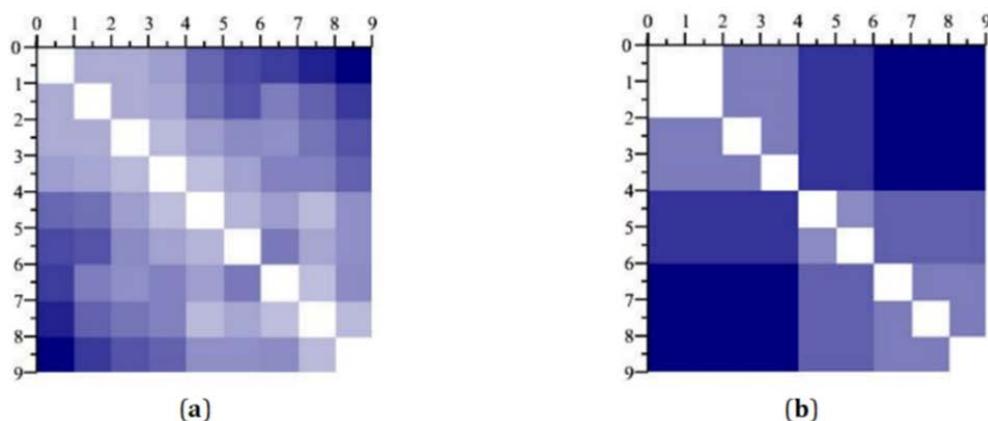


Рисунок 3.3 – Теплова карта матриці мангеттенських відстаней між ранговими векторами атрибутів СВСД: (а) початкове подання; (б) подання після групування ознак за релевантністю

Аналіз теплової карти показує, що атрибути не лише впорядковуються за спаданням значущості, а й природно групуються у три підмножини. До першої, найбільш релевантної підгрупи, належать Glucose, Age, HOMA, Resistin. Другу, середню за релевантністю, утворюють BMI та Insulin. До третьої, найменш релевантної підгрупи, входять Leptin, Adiponectin, MCP1. Такий результат є

важливим, оскільки для подальшого аналізу доцільно орієнтуватися не стільки на жорстке місце кожної окремої ознаки в загальному рейтингу, скільки на її належність до певного рівня інформативності.

Для перевірки того, що виявлене групування не є випадковим, було використано парний двовибірковий t-тест для множини пар ранг-векторів. Отримані результати показали, що відмінності між ознаками всередині кожної підгрупи переважно є статистично незначущими, тоді як відмінності між самими підгрупами є статистично значущими на рівні довіри 95 %. У термінах статистичної перевірки це відповідає рівню значущості $\alpha = 0,05$, тобто довірчому рівню 95 %. Тут α – наперед заданий поріг статистичної значущості, що визначає допустиму ймовірність помилки першого роду, а ρ – р-значення, тобто розрахована за результатами тесту ймовірність отримати спостережувану різницю за умови відсутності реальної різниці між порівнюваними підгрупами.

$$\alpha = 1 - 0,95 = 0,05 \quad (3.1)$$

Тому за умови $\rho < 0,05$ різницю між порівнюваними підгрупами вважали статистично значущою, а за $\rho \geq 0,05$ – статистично незначущою.

Це означає, що поділ атрибутів на три групи має не лише візуальне, а й статистичне підґрунтя. Статистичне розділення ознак на три підгрупи релевантності додатково ілюструється діаграмою box-plot, наведеною на рисунку 3.4.

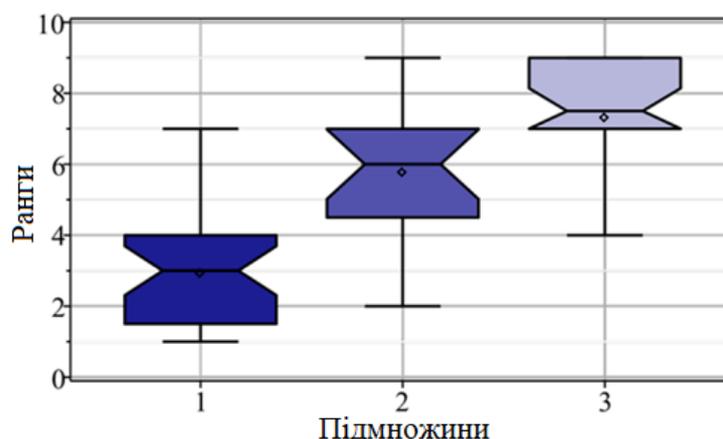


Рисунок 3.4 – Воx-plot для трьох підгруп релевантності атрибутів СВCD

Діаграма на рисунку 3.4 наочно підтверджує, що підгрупи мають різні діапазони рангових значень. Хоча межі між ними не є абсолютно різкими, загальна картина узгоджується з попереднім висновком про існування трьох рівнів релевантності. Практично це означає, що у задачах первинного скринінгу можна насамперед спиратися на першу підгрупу ознак, другу розглядати як допоміжну, а третю – як таку, що потенційно підлягає вилученню або окремому обґрунтуванню перед включенням у модель.

Разом із тим релевантність ознак не вичерпує питання цінності набору даних. Навіть інформативний набір може виявитися недостатньо надійним через велику кількість шумових спостережень, локальних аномалій або неправильно класифікованих записів. Тому наступною складовою методу стало оцінювання зашумленості набору СВCD.

На першому кроці цього етапу застосовано фільтр на основі міжквартильного розмаху (IQR), який виявляє екстремальні значення за класичними статистичними правилами. У результаті застосування IQR-фільтра було ідентифіковано 12 спостережень із 116, які можуть розглядатися як викиди. У відсотковому вираженні це становить:

$$\frac{12}{116} \cdot 100 \% = 10,34 \% \quad (3.2)$$

Із цих 12 спостережень 3 належать до класу healthy (здоровий), а 9 – до класу patients (пацієнти). Для класу patients частка таких об'єктів становить:

$$\frac{9}{64} \cdot 100 \% = 14,06 \% \quad (3.3)$$

тоді як для класу healthy вона дорівнює:

$$\frac{3}{52} \cdot 100 \% = 5,77 \% \quad (3.4)$$

Це свідчить, що зашумленість у класі patients є вищою, ніж у класі healthy. Додатково було використано фільтр CAIRAD, який дає змогу виявляти підозрілі значення атрибутів та оцінювати частку «шумових» значень для кожної ознаки

окремо. Крім того, цей фільтр виявив чотири випадки потенційної помилкової класифікації, що для малого набору даних є особливо чутливим чинником, оскільки безпосередньо впливає на позадіагональні елементи матриці помилок.

Для уточнення практичної придатності набору до первинного скринінгу виконано оцінювання зашумленості його атрибутів; відповідні результати наведено в таблиці 3.1.

Таблиця 3.1 – Частка шумових значень атрибутів CBCD за фільтром CAIRAD

| Властивість | Глюкоза | Вік | НОМА | Резистин | МСР1 |
|--------------------|---------|---------|--------|-------------|------|
| Noise, % | 6 | 14 | 6 | 6 | 0 |
| Властивість | ВМІ | Інсулін | Лептин | Адипонектин | |
| Noise, % | 9 | 12 | 8 | 16 | |

Показник Noise для кожного атрибута були визначені як відношення кількості значень, позначених фільтром CAIRAD як шумові або підозрілі, до загальної кількості спостережень у наборі CBCD з подальшим переведенням у відсотки:

$$Noise\% = \frac{N_{noisy}}{116} \cdot 100\% \quad (3.5)$$

Оскільки у таблиці 3.1 наведено округлені відсоткові значення, їм відповідають приблизні абсолютні кількості: для 6 % – близько 7 значень, для 8 % – близько 9, для 9 % – близько 10, для 12 % – близько 14, для 14 % – близько 16, для 16 % – близько 19 значень.

Дані таблиці 3.1 підтверджують, що рівень шуму розподілений між атрибутами нерівномірно. Найвищі значення зафіксовано для адипонектину у розмірі 16 %, віку – 14 % та інсуліну – 12 %. Для індексу маси тіла шумова частка становить 9 %, для лептин – 8 %, для глюкози, НОМА і резистину – по 6 %, тоді як для МСР1 вона дорівнює 0 %. Таким чином, навіть серед найменш

релевантних і найбільш релевантних ознак ситуація із зашумленістю не є однаковою, що ще раз підкреслює доцільність окремого аналізу як релевантності, так і якості самих даних.

Ще один погляд на проблему аномалій дає алгоритм Local Outlier Factor (LOF), який оцінює локальну щільність оточення кожного об'єкта. Для точок, що належать до «ядра» вибірки, значення LOF зазвичай близьке до 1, тоді як для ізольованіших спостережень воно зростає. Застосування цього підходу до СВCD дало змогу не лише виявити підозрілі об'єкти, а й оцінити загальну структуру аномалій у наборі. Для оцінювання локальних аномалій у наборі СВCD побудовано розподіл оцінок Local Outlier Factor, наведений на рисунок 3.5.

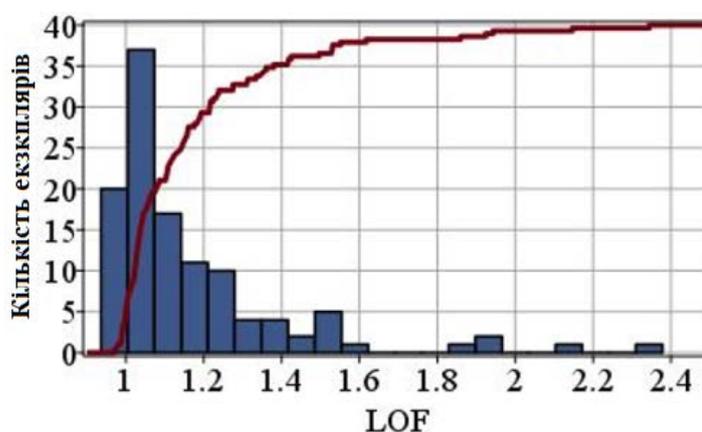


Рисунок 3.5 – Гістограма значень LOF для набору СВCD та відповідна кумулятивна функція розподілу

Як видно з рисунку 3.5, розподіл LOF є асиметричним і має виражений «довгий хвіст» аномальних спостережень. Межу між «ядром» вибірки та зоною викидів доцільно пов'язувати з рівнем $LOF \approx 1,3$. За такої межі «ядро» набору становить 95 об'єктів, а «хвіст» – 21 об'єкт, що у відсотковому вираженні становить відповідно

$$\frac{95}{116} \cdot 100 \% = 81,90 \% \quad (3.6)$$

та

$$\frac{21}{116} \cdot 100 \% = 18,10 \% \quad (3.7)$$

Ця оцінка добре узгоджується з попередніми результатами, отриманими за IQR і CAIRAD, і ще раз свідчить, що набір містить помітну частку проблемних спостережень. У прикладному сенсі це означає, що набір може бути корисним для попереднього скринінгу, але не повинен розглядатися як безумовно надійна основа для відповідальних клінічних висновків без додаткової перевірки.

Отже, запропонований метод попередньої оцінки цінності набору CBCD для первинного скринінгу поєднує три взаємодоповнювальні аспекти аналізу: ранжування ознак за релевантністю, статистичне виділення підгруп інформативності та оцінювання зашумленості набору. За результатами застосування методу встановлено, що найбільш релевантними є Glucose, Age, HOMA, Resistin, середню релевантність мають BMI та Insulin, а найменшу – Leptin, Adiponectin, MCP1. Одночасно показано, що набір містить помітний рівень шуму та викидів: за IQR-оцінкою виявлено 12 із 116 спостережень, тобто 10,34 % набору, а для класу patients ця частка становить 14,06 %. За LOF-аналізом до аномального «хвоста» може бути віднесено 21 із 116 об'єктів, тобто 18,10 % вибірки. Це дає підстави вважати набір CBCD перспективним для задач первинного скринінгу, але таким, що потребує обережного використання, урахування зашумленості та доцільного скорочення ознакового простору перед побудовою діагностичних моделей.

3.4 Метод скорочення вимірності WBCD із формуванням наборів ds1–ds3

Для підвищення практичної придатності Вісконсинського набору даних про рак молочної залози (англ. Wisconsin Breast Cancer Dataset, WBCD) до задач діагностики та одночасного зменшення надлишковості ознакового опису у роботі використано метод послідовного скорочення вимірності з формуванням трьох робочих наборів даних: ds1, ds2 та ds3. Необхідність такого підходу зумовлена тим, що WBCD, попри високу інформативність, містить 30 числових ознак, а отже створює доволі об'ємний простір ознак для подальшого аналізу. За таких умов постає типова для машинного навчання проблема надмірної

розмірності, коли збільшення кількості атрибутів не обов'язково приводить до кращого результату, а в окремих випадках, навпаки, ускладнює побудову моделі та її інтерпретацію. Основні положення цього підходу апробовано в авторській публікації [99, 100].

Набір WBCD [101] містить 569 спостережень і описує два класи пухлин – malignant (злаякісна) та benign (доброякісна). У його структурі використано 30 числових ознак, які отримані на основі десяти базових характеристик клітинного ядра, визначених із мамографічних зображень. Для кожної з цих характеристик у наборі задано три представлення: середнє значення, стандартне відхилення та, так зване, «найгірше» значення. У результаті формується досить детальний, але водночас і громіздкий ознаковий опис, що й робить задачу скорочення вимірності актуальною.

Запропонований метод передбачає не одноразове, а послідовне триетапне перетворення початкового набору. Спочатку виконується стандартизація числових атрибутів, далі – відбір найбільш доцільних ознак, а після цього – додаткове зниження розмірності методом головних компонент. У такий спосіб утворюються три набори даних, які відрізняються між собою розмірністю ознакового простору, але зберігають єдину цільову ознаку класу. Послідовність формування наборів ds1, ds2 та ds3 у межах запропонованого методу скорочення вимірності WBCD наведено в таблиці 3.2.

На першому етапі всі числові атрибути початкового набору було стандартизовано. Для цього використано стандартний фільтр WEKA Standardize, у результаті чого кожна числова ознака була приведена до нульового середнього значення та одиничного стандартного відхилення. Така процедура є доцільною, оскільки початкові атрибути мають різні масштаби вимірювання, а їх безпосереднє використання без вирівнювання масштабу могло б спотворювати подальший аналіз. Після цього було сформовано перший робочий набір ds1, який містить усі 30 стандартизованих числових ознак і номінальний атрибут класу [99].

Таблиця 3.2 – Послідовність формування наборів ds1, ds2 та ds3 у методі скорочення вимірності WBCD

| Позначення набору | Вхідні дані | Перетворення | Вихідний простір ознак | Призначення |
|-------------------|----------------|--|--|--|
| ds1 | (повний набір) | Standardize (стандартизація ознак) | 30 стандартизованих ознак + клас | Базове представлення для порівняння; вирівнює масштаби ознак |
| ds2 | ds1 | AttributeSelection (<i>CfsSubsetEval</i> + <i>BestFirst</i>) | 11 відібраних ознак + клас | Скорочення розмірності без різкого зниження інформативності |
| ds3 | ds2 | PrincipalComponents (PCA) з параметром $R = 0,91$ | 4 головні компоненти (91 % дисперсії) + клас | Максимально компактне представлення для швидкої аналітики |

Другий етап полягав у відборі підмножини найбільш змістовних атрибутів із набору ds1. Для цього у середовищі WEKA було використано фільтр відбору ознак AttributeSelection з оцінювачем CfsSubsetEval, який оцінює підмножину ознак на основі кореляційного критерію, у поєднанні з процедурою пошуку BestFirst (першого кращого). Така комбінація дає змогу не просто ранжувати ознаки поодиночки, а знайти підмножину атрибутів, яка є інформативною з погляду зв'язку з цільовим класом і водночас не містить надмірно дубльованих характеристик. У результаті цього кроку з 30 вихідних числових ознак було залишено 11. Саме вони утворили другий робочий набір ds2. Таким чином, на

етапі переходу від ds1 до ds2 кількість числових ознак було зменшено з 30 до 11, тобто на:

$$\frac{30-11}{30} \cdot 100 \% = 63,3 \%. \quad (3.8)$$

Це означає, що після відбору ознак у наборі ds2 збережено лише 36,7 % початкового ознакового опису. Важливо, що після цього перетворення набір уже стає помітно компактнішим, але ще не настільки скороченим, щоб вважати його оптимальним для «легковагових» моделей.

Третій етап було виконано над набором ds2. На цьому кроці використано фільтр PrincipalComponents на основі методу головних компонент, який дає змогу перейти від початкових 11 ознак до нового простору головних компонент. У роботі було обрано таку конфігурацію перетворення, за якої зберігається близько 91 % загальної дисперсії. У результаті набір ds2 було зведено до 4 головних компонент, які й утворили третій робочий набір ds3. Саме цей набір є найбільш скороченим і водночас зберігає основну частину інформації, накопиченої в попередньому ознаковому просторі. Перехід від початкових 30 числових ознак до 4 головних компонент означає, що розмірність ознакового простору було зменшено на

$$\frac{30-4}{30} \cdot 100 \% = 86,7 \%. \quad (3.9)$$

Іншими словами, у наборі ds3 збережено лише 13,3 % від початкової кількості числових ознак, а саме скорочення становить приблизно 7,5 раза:

$$\frac{30}{4} = 7,5. \quad (3.10)$$

Показник 91 % у цьому випадку відповідає частці загальної дисперсії даних, яку збережено після перетворення методом головних компонент, тобто 4 компоненти набору ds3 пояснюють 91 % варіативності початкового 11-вимірного простору ds2.

Основні характеристики сформованих наборів ds1, ds2 та ds3 наведено в таблиці 3.3.

Таблиця 3.3 – Характеристики наборів даних ds1, ds2 та ds3, сформованих у процесі скорочення вимірності WBCD

| Набір даних | Характер перетворення | Кількість числових ознак | Додаткові характеристики |
|--------------------|---------------------------------|---------------------------------|--|
| ds1 | Стандартизація початкового WBCD | 30 | Усі числові атрибути стандартизовано |
| ds2 | Відбір ознак із ds1 | 11 | Підмножина атрибутів, відібрана засобами CfsSubsetEval + BestFirst |
| ds3 | РСА-перетворення набору ds2 | 4 | 4 головні компоненти, що охоплюють близько 91 % дисперсії |

Таким чином, повне скорочення ознакового простору від 30 до 4 вимірів означає зменшення розмірності на 86,7 %, або приблизно у 7,5 раза. Це і є основний практичний результат запропонованого методу. При цьому важливо підкреслити, що мета роботи полягала не в механічному зменшенні кількості атрибутів за будь-яку ціну, а в пошуку такого рівня редукції, який ще не призводить до істотної втрати корисної інформації. Саме тому перехід до ще

меншої кількості компонент, наприклад до двох, у роботі не приймався, оскільки це вже вимагало б згоди на помітніше зниження частки охопленої дисперсії.

Запропонований метод є послідовним і логічно завершеним. Набір ds1 використовується як базове стандартизоване подання повного WBCD. Набір ds2 відображає результат структурного відбору найбільш придатних атрибутів. Набір ds3 є найкомпактнішим представленням, у якому збережено основну частину мінливості даних. Така трирівнева побудова дає змогу не лише скоротити простір ознак, а й надалі порівнювати поведінку моделей на повному, частково скороченому та максимально зменшеному поданнях одного й того самого набору даних.

Щоб перевірити, чи не призводить така редукція до втрати прогностичної здатності, сформовані набори ds1–ds3 було далі використано в серії порівняльних експериментів із класифікаторами глибокого навчання у WEKA. Одержані результати показали, що навіть після послідовного скорочення ознакового простору якість класифікації залишається високою. Зокрема, частка правильно класифікованих екземплярів для досліджених моделей перебувала в межах приблизно 95,87–97,68 %, а статистично значущого погіршення якості в результаті переходу від ds1 до ds3 виявлено не було. Це означає, що запропонований метод не лише зменшує розмірність, а й зберігає практичну цінність набору для подальшої діагностичної обробки. Показники частки правильно класифікованих екземплярів для наборів ds1–ds3 наведено в таблиці 3.4.

Таблиця 3.4 – Частка правильно класифікованих екземплярів для наборів ds1, ds2 та ds3 при використанні класифікаторів глибокого навчання у WEKA

| Набір даних | D14jMlp | MLP | Voted Perceptron |
|-------------|--------------|--------------|------------------|
| ds1 | 97,68 (1,80) | 96,72 (2,20) | 96,45 (2,11) |
| ds2 | 97,58 (1,79) | 96,64 (1,90) | 96,78 (2,14) |

Продовження таблиці 3.4

| Набір даних | DI4jMlp | MLP | Voted Perceptron |
|-------------|--------------|--------------|------------------|
| ds3 | 96,40 (2,26) | 95,87 (2,47) | 95,96 (2,33) |

Наведені в таблиці 3.4 значення свідчать, що перехід від повного набору до значно скороченого не супроводжується різким падінням точності. Невелике зниження окремих показників є очікуваним, однак у межах проведених експериментів воно не мало критичного характеру. Отже, скорочення простору ознак до чотирьох головних компонент можна вважати обґрунтованим компромісом між компактністю подання та збереженням діагностичної інформативності [99].

Таким чином, у роботі використано метод скорочення вимірності WBCD, який передбачає послідовне застосування трьох перетворень – стандартизації, відбору ознак і подальшого переходу до простору головних компонент – із формуванням наборів ds1, ds2 та ds3. Запропонований метод скорочення вимірності забезпечив послідовний перехід від 30 початкових числових ознак до 11 відібраних атрибутів, а далі – до 4 головних компонент, що відповідає зменшенню розмірності на 86,7 % при збереженні 91 % загальної дисперсії даних. Це створює підґрунтя для подальшого використання скорочених наборів у порівняльному аналізі діагностичних рішень і побудові більш компактних та інтерпретованих моделей [99, 100].

3.5 Метод порівняльного аналізу ефективності діагностичних рішень у WEKA

Після формування робочих наборів даних і скорочення ознакового простору виникає практичне питання: яку саме модель доцільно вважати найпридатнішою для діагностичного використання. Для відповіді на це питання у роботі застосовано метод порівняльного аналізу ефективності діагностичних рішень у середовищі WEKA. Його призначення полягає не лише у формальному

зіставленні кількох алгоритмів, а у виборі такого рішення, яке забезпечує належну якість класифікації, залишається стабільним під час повторних перевірок і, за можливості, не втрачає інтерпретованості. Саме тому порівняння виконується не за однією метрикою, а за їх сукупністю [96, 99].

У загальному вигляді запропонований метод включає кілька послідовних етапів. Спочатку формується набір алгоритмів кандидатів для розв'язання конкретної діагностичної задачі. Далі для кожного алгоритму виконується навчання і перевірка за єдиною схемою валідації. Після цього обчислюються показники якості, які дають змогу оцінити не лише загальну частку правильних відповідей, а й здатність моделі коректно виявляти цільовий клас. На завершальному етапі результати зіставляються між собою, а відмінності між алгоритмами оцінюються з урахуванням статистичної стійкості одержаних показників. Така процедура дає змогу уникнути ситуації, коли вибір моделі робиться лише за однією «найкращою» цифрою без урахування її варіації та практичного змісту.

У роботі оцінювання виконувалося за схемою 10-fold cross-validation (десятиблокової перехресної перевірки) з повтореннями. За такої схеми вибірка багаторазово ділиться на навчальні та тестові частини, причому кожний фрагмент даних по черзі використовується як контрольний. Це дозволяє зменшити залежність висновків від одного випадкового розбиття вибірки та отримати більш надійну оцінку якості. Для перевірки того, чи є спостережувані відмінності між алгоритмами реальними, а не випадковими, доцільно використовувати статистичне порівняння на рівні довіри 0,95. У середовищі WEKA для цього зручно застосовувати скоригований парний t-критерій, що особливо важливо під час багаторазових експериментів.

Базовим показником у методі є Accuracy (правильність), тобто частка правильно класифікованих екземплярів. Проте для задач медичної діагностики цього недостатньо. Висока загальна точність (правильність) не гарантує, що модель добре виявляє саме пацієнтів, а не «виграє» лише за рахунок коректного розпізнавання простішого класу. Тому додатково аналізуються Precision

(точність), Recall (повнота) і F1 (міра Ван Різбергена; середнє гармонічне точності Precision та повноти Recall), причому основний акцент робиться на цільовому класі «пацієнти». Такий підхід дозволяє окремо оцінити, наскільки модель здатна знаходити випадки захворювання, не створюючи надмірної кількості хибних спрацювань. Для більш повного зіставлення також використовується PRC area – площа під кривою «точність-повнота», яка є особливо інформативною, коли важливо контролювати баланс між пропусками і хибними спрацюваннями.

Першим прикладом застосування методу є набір Coimbra Breast Cancer Dataset, для якого було виконано порівняння п'яти базових класифікаторів WEKA: NaiveBayes (наївний байєсівський аналіз), Ibk (метод k-найближчих сусідів), J48 (дерево рішень), MultilayerPerceptron (багатошаровий перцептрон), SMO (метод опорних векторів). У цьому випадку порівняння проводиться не лише за Accuracy, а й за показниками Precision, Recall та F1 саме для класу «пацієнти», оскільки саме цей клас має найбільше прикладне значення для сценарію первинного скринінгу. Результати порівняння алгоритмів класифікації на наборі Коїмбри наведено в таблиці 3.5.

Таблиця 3.5– Порівняння алгоритмів класифікації на наборі Коїмбри

| Алгоритм (WEKA) | Accuracy, % | Precision (patients = 2) | Recall (patients = 2) | F1(patients = 2) |
|----------------------|----------------|-----------------------------|--------------------------|------------------|
| NaiveBayes | 60,3448 | 0,765 | 0,406 | 0,531 |
| IBk (k = 3) | 67,2414 | 0,724 | 0,656 | 0,689 |
| J48 | 68,9655 | 0,726 | 0,703 | 0,714 |
| MultilayerPerceptron | 65,5172 | 0,688 | 0,688 | 0,688 |
| SMO (SVM) | 66,3793 | 0,705 | 0,672 | 0,688 |

Як видно з таблиці 3.5, найвищу Accuracy у цій серії експериментів показав алгоритм J48. Водночас його перевага полягає не лише у відсотку правильно класифікованих випадків. Для класу «пацієнти» він також демонструє найбільш збалансоване співвідношення між Precision, Recall і F1. Це означає, що модель не просто дає прийнятний загальний результат, а й досить впевнено виявляє саме ті спостереження, які мають найбільшу клінічну цінність у задачі первинного скринінгу. З огляду на це J48 доцільно розглядати не тільки як один з алгоритмів у переліку, а як базову інтерпретовану модель для подальшого аналізу [96].

Другим прикладом застосування методу є набір Wisconsin Breast Cancer Dataset, для якого порівняння виконувалося вже не між різними типами класифікаторів, а між трьома нейромережевими алгоритмами D14jMpr (багатошаровий перцептрон на основі бібліотеки DeepLearning4J), Multilayer Perceptron (класичний багатошаровий перцептрон) та Voted Perceptron (голосуючий перцептрон) на трьох представленнях даних: ds1, ds2 і ds3 [3]. Тут мета порівняльного аналізу була дещо іншою: не лише виявити кращий алгоритм, а й перевірити, чи не призводить скорочення вимірності до прихованої втрати якості. Саме тому результати зіставлялися одночасно за рядками таблиць між наборами ds1–ds3 і за стовпцями між самими алгоритмами.

Відсоток правильно класифікованих екземплярів для наборів ds1–ds3 наведено в таблиці 3.4. Наведені значення показують, що після переходу від повного набору до скорочених представлень ds2 і ds3 відсоток правильно класифікованих екземплярів залишається високим. Формально спостерігається невелике зниження окремих показників, однак воно не має різкого характеру. Більш того, розкид результатів, відображений у стандартних відхиленнях, залишається в межах, типовий для стабільних моделей за перехресної перевірки. Отже, перехід до компактнішого простору ознак не супроводжується критичною втратою прогностичної здатності [99].

Проте для діагностичних задач важливо оцінити не лише Accuracy, а й поведінку моделей у координатах «точність-повнота». Саме тому додатково було виконано зіставлення за PRC area. Цей показник є особливо корисним тоді,

коли потрібно переконатися, що навіть за скорочення простору ознак модель зберігає здатність надійно виявляти цільовий клас. Площа під кривими PRC для наборів ds1–ds3 наведена в таблиці 3.6.

Таблиця 3.6 – Площа під кривими PRC для наборів ds1–ds3

| Набір даних | Алгоритм | | |
|-------------|-------------|-----------------------|------------------|
| | Dl4jMlp | Multilayer Perceptron | Voted Perceptron |
| ds1 | 1,00 (0,01) | 0,99 (0,01) | 0,97 (0,02) * |
| ds2 | 0,99 (0,01) | 0,99 (0,01) | 0,97 (0,02) * |
| ds3 | 0,99 (0,01) | 0,99 (0,01) | 0,96 (0,02) * |

Із таблиці 3.6 видно, що всі моделі демонструють дуже високі значення PRC area. Для Dl4jMlp та Multilayer Perceptron ці значення практично наближаються до одиниці на всіх трьох представленнях даних. Для Voted Perceptron вони також залишаються високими, хоча виявляються статистично гіршими, що й позначено символом «*». Водночас принципово важливо, що між самими наборами ds1, ds2, ds3 не спостерігається різкого погіршення PRC area. Це означає, що скорочення ознак не руйнує здатність моделей відокремлювати цільовий клас навіть у координатах, чутливих до співвідношення між точністю та повнотою.

Отже, застосований метод порівняльного аналізу ефективності діагностичних рішень у WEKA дає змогу оцінювати моделі не ізольовано, а в єдиній процедурі зіставлення. Для набору Коїмбри це дозволило аргументовано виділити J48 як компроміс між якістю та інтерпретованістю. Для WBCD метод показав, що перехід від ds1 до ds3 забезпечує істотне скорочення ознакового простору без суттєвої деградації показників Accuracy і PRC area. Таким чином,

метод створює обґрунтовану основу для подальшого вибору як інтерпретованих моделей, так і більш компактних представлень даних у задачах діагностики.

3.6 Модель інтерпретованого діагностичного рішення на основі дерева J48

У задачах медичної діагностики важливо не лише отримати правильний результат класифікації, а й мати можливість пояснити, чому система віднесла об'єкт саме до певного класу. Це особливо актуально для сценаріїв первинного скринінгу, де лікареві або дослідникові потрібен не тільки формальний висновок, а й зрозуміла логіка його отримання. З огляду на це у роботі окремо розглянуто модель інтерпретованого діагностичного рішення на основі дерева J48, яка дає змогу подати процедуру класифікації як послідовність прозорих умов на значеннях ознак. Основні результати, покладені в основу цього підходу, висвітлено в авторській публікації [97, 100].

Модель J48 у роботі використано не як універсальний класифікатор «на всі випадки», а як інтерпретований інструмент для побудови пояснюваного діагностичного рішення. Її застосування є доцільним після попереднього скорочення ознакового простору, оскільки саме на компактному наборі найбільш змістовних атрибутів дерево рішень набуває практичної цінності. Якщо дерево будується на надмірно великій кількості слабоінформативних ознак, воно швидко втрачає наочність, ускладнює ручний аналіз і перестає виконувати головну функцію – пояснювати логіку класифікації. Саме тому побудова J48 у роботі спирається на попередні етапи відбору ознак і порівняльного аналізу моделей.

У межах запропонованого підходу дерево J48 формувалося для Вісконсинського набору даних про рак молочної залози (Wisconsin Breast Cancer Dataset, WBCD) на скороченому ознаковому описі. Попередній етап порівняльного аналізу дозволив виокремити компактну підмножину найбільш інформативних ознак, які забезпечують достатню якість класифікації та водночас зберігають змістовну інтерпретованість. Найважливішими для

побудови дерева виявилися атрибути `perimeter_worst` (максимальне значення периметра ядра клітини), `area_worst` (максимальне значення площі ядра клітини), `concave_points_worst` (максимальна кількість увігнутих точок контуру ядра), `texture_mean` (середнє значення текстурної характеристики ядра), `concave_points_mean` (середня кількість увігнутих точок контуру ядра). Саме вони утворюють основу спрощеної моделі рішення, яка дає змогу перейти від числового ознакового простору до системи логічних умов.

Обраний підхід дозволяє розглядати дерево рішень не просто як технічний вихід середовища WEKA, а як формалізовану модель прийняття діагностичного рішення. Кореневий вузол дерева відповідає найбільш роздільній ознаці, а подальші вузли уточнюють класифікацію через граничні значення інших атрибутів. Таким чином, кожний шлях від кореня до листка можна інтерпретувати як окремий сценарій діагностичного висновку. Така форма подання є принципово важливою, оскільки робить результат зрозумілим не лише з позиції машинного навчання, а й з погляду прикладного використання. Структуру побудованої моделі інтерпретованого рішення на основі дерева J48 наведено на рисунок 3.6.

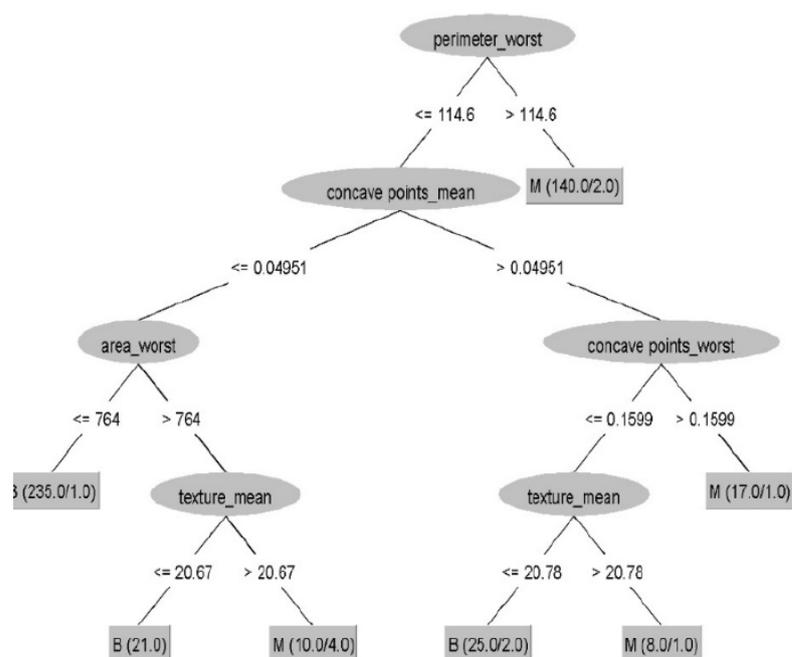


Рисунок 3.6 – Спрощене дерево рішень J48 для діагностичної класифікації об'єктів набору WBCD

Як показано на рисунку 3.6, для налаштованої в роботі моделі J48 отримано компактну структуру дерева, що містить 13 вузлів і 7 листків [4]. Важливо підкреслити, що ці значення характеризують не алгоритм J48 загалом, а саме ту конфігурацію дерева, яку було отримано в даному дослідженні після скорочення ознакового простору та налаштування параметрів. Такий розмір дерева є достатньо малим, щоб його можна було аналізувати вручну, і водночас достатнім для збереження діагностично значущої інформації. У практичному сенсі це означає, що модель не перевантажена зайвими розгалуженнями й може бути використана як основа для пояснюваної підтримки прийняття рішень [97].

Для кількісної оцінки якості побудованої моделі було використано не одну, а кілька взаємодоповнювальних метрик. Це дає змогу уникнути ситуації, коли висока загальна точність приховує слабкі місця моделі в розпізнаванні окремих класів. Для налаштованого дерева J48 отримано такі результати: $\text{precision} = 0,967$, $\text{MCC} = 0,929$, $\text{ROC area} = 0,978$. Сукупно ці значення свідчать про те, що модель забезпечує не лише високу частку правильних класифікацій, а й добру узгодженість прогнозів із реальними класами та стійке розділення класів у просторі рішень. Основні показники якості налаштованої моделі J48 наведено в таблиці 3.7.

Таблиця 3.7 – Показники якості моделі інтерпретованого діагностичного рішення на основі дерева J48

| Показник | Precision | MCC | ROC area | Кількість вузлів дерева | Кількість листків |
|-----------------|------------------|------------|-----------------|--------------------------------|--------------------------|
| Значення | 0,967 | 0,929 | 0,978 | 13 | 7 |

Показник $\text{precision} = 0,967$ означає, що серед об'єктів, які модель віднесла до відповідного класу, переважна більшість справді належить до нього. Коефіцієнт кореляції Метьюса $\text{MCC} = 0,929$ є особливо важливим для задач двокласової діагностики, оскільки враховує всі елементи матриці помилок і дозволяє оцінити узгодженість між прогнозованими та істинними класами

значно повніше, ніж звичайна точність. Значення ROC area = 0,978 свідчить про високу здатність моделі розрізняти класи за різних порогів прийняття рішення. Разом ці показники підтверджують, що компактність дерева досягнута без суттєвої втрати якості.

Однак справжня цінність моделі J48 у даній роботі полягає не лише в числових метриках, а в можливості безпосереднього перетворення дерева на набір правил типу «IF – THEN». Саме в такій формі модель стає зручною для практичного прочитання. Кожен маршрут від кореня до листка відповідає окремому правилу, яке містить послідовність умов на діагностично значущих атрибутах і завершується віднесенням об'єкта до одного з класів. Це дає змогу пояснити, які саме ознаки виявилися вирішальними в конкретній класифікації. Приклади правил «якщо – то», отриманих зі спрощеного дерева J48, наведено в таблиці 3.8.

Таблиця 3.8 – Приклади правил «якщо – то», отриманих зі спрощеного дерева рішень J48

| № | Правило типу «якщо – то» (IF – THEN) | Клас |
|---|--|------|
| 1 | IF perimeter_worst ≤ 114.6 AND concave_points_mean ≤ 0.04951 AND area_worst ≤ 764 THEN ... | B |
| 2 | IF perimeter_worst ≤ 114.6 AND concave_points_mean ≤ 0.04951 AND area_worst > 764 AND texture_mean ≤ 20.67 THEN ... | B |
| 3 | IF perimeter_worst ≤ 114.6 AND concave_points_mean ≤ 0.04951 AND area_worst > 764 AND texture_mean > 20.67 THEN ... | M |
| 4 | IF perimeter_worst ≤ 114.6 AND concave_points_mean > 0.04951 AND concave_points_worst ≤ 0.1599 AND texture_mean ≤ 20.78 THEN ... | B |
| 5 | IF perimeter_worst ≤ 114.6 AND concave_points_mean > 0.04951 AND concave_points_worst ≤ 0.1599 AND texture_mean > 20.78 THEN ... | M |

Продовження таблиці 3.8

| № | Правило типу «якщо – то» (IF – THEN) | Клас |
|---|---|------|
| 6 | IF perimeter_worst \leq 114.6 AND concave_points_mean $>$ 0.04951 AND concave_points_worst $>$ 0.1599 THEN ... | М |

Наведені правила показують, що дерево J48 у цьому випадку опирається на невелику кількість добре інтерпретованих ознак, пов'язаних із геометричними та текстурними характеристиками утворення. З погляду практики це важливо, оскільки модель не створює враження «чорної скриньки»: користувач може простежити, за якими умовами об'єкт віднесено до доброякісного або злроякісного класу. Більше того, сама структура правил вказує на пріоритетні для моделі предиктори та порогові значення, за якими відбувається розділення випадків [97].

Інтерпретованість у цьому підході не слід зводити лише до факту наявності дерева. Справжня перевага моделі полягає в тому, що дерево залишається достатньо компактним, спирається на обмежений перелік змістовних ознак і водночас забезпечує високу якість класифікації. Саме поєднання цих властивостей робить J48 придатним для сценаріїв, де важлива пояснюваність результату, можливість ручної перевірки логіки та потенційна інтеграція моделі в прикладні програмні модулі системи моніторингу.

Отже, у роботі побудовано модель інтерпретованого діагностичного рішення на основі дерева J48, яка функціонує на скороченому ознаковому описі набору WBCD і забезпечує поєднання двох важливих властивостей: високої якості класифікації та зрозумілої логічної структури рішення. Для налаштованої моделі отримано 13 вузлів і 7 листків, а також показники precision = 0,967, MCC = 0,929, ROC area = 0,978. Це дає підстави розглядати дерево J48 як придатну основу для пояснюваних діагностичних рішень у задачах первинного скринінгу та підтримки аналізу біомедичних даних.

3.7 Модель індукції діагностичних правил на основі JRip

У задачах первинного скринінгу важливим є не лише отримання прийнятної точності класифікації, а й можливість подати результат у формі, придатній для безпосереднього прочитання та практичного використання. Саме тому в роботі окремо розглянуто модель індукції діагностичних правил на основі алгоритму JRip, який реалізує підхід Repeated Incremental Pruning to Produce Error Reduction (повторне інкрементне обрізання для зменшення помилок). На відміну від моделей типу «чорної скриньки», JRip формує набір правил виду «IF – THEN», що дає змогу безпосередньо простежити, які саме поєднання ознак приводять до віднесення об'єкта до певного класу. Основні положення цього підходу та результати його застосування до набору Coimbra Breast Cancer Dataset (набір про рак молочної залози Коїмбри; CBCD) висвітлено в авторській публікації [96].

Доцільність використання JRip у цій роботі зумовлена специфікою набору CBCD, який містить невелику кількість спостережень, порівняно компактний ознаковий опис і водночас помітний рівень шуму. За таких умов надто складні моделі не завжди є найкращим вибором, оскільки їх важче інтерпретувати, а виграш у якості може бути нестійким. Натомість модель індукції правил дозволяє перейти від абстрактного багатовимірного простору ознак до системи логічних умов, які можуть бути прочитані та перевірені людиною. Для задачі попереднього скринінгу це є важливою перевагою, оскільки дає можливість сформулювати прозорий діагностичний сценарій на основі відносно невеликої кількості показників.

У межах запропонованого підходу JRip використовується не ізольовано, а після попередніх етапів аналізу набору CBCD. Спочатку було оцінено релевантність ознак і встановлено, що найбільш змістовними є Glucose (глюкоза), Age (вік), HOMA (відношення глюкози та інсуліну), Resistin (резистин), тоді як BMI (індекс маси тіла) та Insulin (рівень інсуліну) мають середню релевантність, а Leptin (лептин), Adiponectin (адипонектин), MCP1 (рівень цитокину) – низьку. Окремо було показано, що набір містить помітну

частку шуму та викидів. Це створило підстави для перевірки кількох варіантів подання даних: повного набору, очищеного від частини аномальних спостережень, редукованого шестиатрибутного набору та найбільш компактного чотириатрибутного набору. Саме на такому тлі модель JRip розглядається як засіб не лише класифікації, а й побудови зрозумілої системи діагностичних правил.

Щоб оцінити стійкість моделі, у роботі було виконано багаторазове тестування JRip у режимі 5-fold cross-validation (п'ятивибіркової кросвалідації). Такий підхід дозволяє зменшити вплив випадкового розбиття вибірки та перейти від одиничного результату до середніх значень показників і їхні варіації. Для кожного варіанта набору даних проводилася серія з великої кількості повторів, що дало змогу порівняти не лише середню якість класифікації, а й стабільність отриманих оцінок. Важливо, що при цьому аналізувалися не лише випадки покращення, а й статистична значущість виявлених змін.

У результаті було встановлено, що перехід від повного набору SVCD до очищеного та далі до скорочених подань супроводжується певним зростанням середніх показників якості. Якщо розглядати Accuracy (правильність) як найбільш наочний індикатор, то її середні значення змінювалися послідовно: 70,2 % для повного набору, 79,7 % для очищеного, 81,1 % для редукованого шестиатрибутного подання та 81,6 % для найбільш компактного чотириатрибутного набору. Проте разом із цим стандартні відхилення залишалися достатньо великими, тому формально зафіксоване покращення не можна трактувати як безумовно статистично значуще. Це важливий результат сам по собі: він показує, що редукція та часткове очищення даних не гарантують різкого підвищення якості, але водночас і не руйнують її [96].

Після аналізу узагальнених показників особливу увагу було приділено питанню інтерпретованості. Саме тут JRip виявився найбільш корисним. Алгоритм формує не дерево, а компактний набір правил, який у готовому вигляді можна подати як логічну схему прийняття рішення. Для лікаря або дослідника така форма є значно ближчою до звичної практики, ніж абстрактні вагові

коефіцієнти чи приховані шари нейромережі. Кожне правило задає конкретне поєднання порогових значень біомаркерів, при якому спостереження відноситься до класу здорових або пацієнтів.

Таблиця 3.9 – Порівняльні результати застосування JRip до різних варіантів подання набору SVCD

| Варіант набору SVCD | Повний набір | Очищений набір | Редукований шестиатрибутний набір | Редукований чотириатрибутний набір |
|----------------------------|---|---|---|---|
| Характеристика подання | Початковий набір SVCD | Після вилучення помилково класифікованих спостережень | Без Leptin, Adiponectin, MCP1 | Додатково без BMI та Insulin |
| Кількість атрибутів | 9 | 9 | 6 | 4 |
| Кількість екземплярів | 116 | 100 | 100 | 100 |
| Ассурація, % (M ± SD) | 70,2 ± 8,7 | 79,7 ± 9,1 | 81,1 ± 8,6 | 81,6 ± 8,6 |
| Коментар | Базовий варіант без очищення і редукції | Підвищення середньої точності після фільтрації | Скорочення ознакового простору без помітної втрати якості | На ньому отримано правила JRip |

Найбільш показові правила було отримано на найкомпактнішому чотириатрибутному поданні набору. Саме цей результат є принципово важливим, оскільки демонструє: навіть після істотного скорочення ознакового опису модель зберігає не лише прийнятну якість, а й практичну прозорість. У

підсумку сформовано три явні правила для класу healthy (здоровий) і одне правило за замовчуванням для класу patients (пацієнти). Приклади правил IF–THEN, індукованих алгоритмом JRip для скороченого набору CBCD, наведено в таблиці 3.10.

Таблиця 3.10 – Діагностичні правила «IF – THEN», отримані за допомогою JRip для скороченого подання набору CBCD

| № | Правило (IF – THEN) | Клас | Покриття (кількість випадків) | Ймовірність |
|---|--|---------|-------------------------------------|-------------|
| 1 | IF Glucose \leq 90 AND Resistin \leq 12,9361 AND HOMA \geq 0,827271 THEN | healthy | 17 | 100 % |
| 2 | IF Age \leq 36 AND Glucose \leq 90 THEN | healthy | 7 | 100 % |
| 3 | IF Age \geq 66 AND Glucose \leq 102 AND Resistin \leq 12,766 THEN | healthy | 9 | 100 % |
| 4 | ELSE | patient | 67 | 93 % |

Наведені в таблиці 3.10 правила є змістовно важливими з кількох причин. По-перше, вони показують, що для прийняття попереднього рішення виявилось достатньо дуже обмеженого набору показників, причому ключову роль у правилах відіграють Glucose, Age, Resistin та частково HOMA. По-друге, усі три явні правила належать до класу healthy, тоді як решта випадків автоматично відноситься до класу patients. Така структура не випадкова: вона відображає прагнення моделі спочатку виокремити відносно чіткі сценарії «нормального» профілю, а всі інші, менш типові комбінації ознак, трактувати як потенційно проблемні [96].

Окрему увагу варто звернути на практичну простоту одержаних правил. Перше правило поєднує поріг Glucose з додатковими обмеженнями на Resistin та

НОМА і покриває 17 випадків із повною правильністю в межах цього правила. Друге правило ще компактніше: воно базується лише на двох умовах Age та Glucose і охоплює 7 випадків без помилок. Третє правило вводить іншу вікову зону та дещо м'якший поріг для Glucose, але знову використовує обмеження на Resistin; воно покриває 9 випадків і також не дає помилок у межах цього локального правила. Правило за замовчуванням відносить усі інші випадки до класу patient, забезпечуючи покриття решти спостережень із імовірністю близько 93 %.

Саме така побудова робить модель JRip особливо привабливою для задач первинного скринінгу. Вона не вимагає аналізу великої кількості параметрів, не приховує логіку рішення і в найпростішому випадку може бути реалізована навіть у вигляді набору послідовних перевірок у програмному модулі. У прикладному сенсі це означає, що для автоматизованого попереднього відбору потенційно ризикових випадків достатньо контролювати лише кілька ключових показників. Така економність особливо цінна для сценаріїв, де необхідно мінімізувати витрати часу, зусиль і вартості обстеження.

Разом із тим результати моделі слід інтерпретувати обережно. Набір SVCD є невеликим і зашумленим, а тому будь-який класифікатор, навіть інтерпретований, не може розглядатися як остаточне клінічне рішення. У цьому сенсі JRip слугує насамперед інструментом пояснюваного попереднього аналізу. Його правила корисні як логічний шаблон для первинного скринінгу та як засіб виявлення найбільш інформативних комбінацій ознак, але не замінюють повноцінної медичної діагностики.

Для кращого розуміння меж застосовності моделі доцільно також враховувати узагальнені оцінки якості, отримані для набору SVCD. Вони показують, що для поточного стану цього набору реалістичні значення показників перебувають у відносно широких межах, а отже будь-який висновок про «найкращий» результат має бути стриманим. Водночас навіть за таких обмежень JRip демонструє важливу перевагу: він перетворює набір біомаркерів

на чіткі, легко програмовані правила. Орієнтовні межі основних показників якості для моделі JRip на наборі CBCD наведено в таблиці 3.11.

Таблиця 3.11 – Орієнтовні межі показників якості моделі JRip для набору CBCD

| Показник | Accuracy, % | FP Ratio | Кappa | F1-Measure | ROC Area |
|----------|----------------|-----------|-----------|------------|-----------|
| Діапазон | 72–90 | 0,00–0,20 | 0,43–0,79 | 0,72–0,89 | 0,72–0,86 |

Наведені в таблиці 3.11 межі не слід трактувати як характеристики одного-єдиного запуску алгоритму. Вони радше відображають робочий інтервал, у межах якого може коливатися якість моделі залежно від схеми тестування, варіанта набору даних і параметрів налаштування.

Отже, побудовано модель індукції діагностичних правил на основі JRip, орієнтовану на задачу первинного скринінгу за даними набору CBCD. Модель базується на попередньому відборі інформативних ознак, допускає використання скорочених подань набору та формує компактний набір правил виду «IF – THEN», придатний для безпосереднього прочитання і програмної реалізації. Найбільш практично значущий результат полягає в тому, що навіть на чотириатрибутному поданні набору вдалося отримати прості правила, засновані переважно на атрибутах Glucose, Age, Resistin і НОМА, які забезпечують прийнятну якість класифікації та високий рівень пояснюваності. Саме це дозволяє розглядати JRip як основу для побудови прозорих діагностичних рішень у межах попереднього скринінгу.

3.8 Метод валідації кластерної структури даних за метрикою FMI

Після скорочення розмірності набору даних постає окреме питання: чи не втрачається при цьому внутрішня структура даних, тобто чи зберігається природне групування об'єктів у просторі ознак. Для перевірки цього у роботі застосовано метод валідації кластерної структури даних, який ґрунтується на

зіставленні результатів некерованого групування з еталонними класами за метрикою Fowlkes-Mallows Index (FMI). Такий підхід дає змогу оцінити не тільки поведінку класифікаторів у задачі з учителем, а й те, наскільки самі дані після редукції зберігають змістовний поділ на класи. Основні положення цього підходу та результати його застосування до набору WBCD висвітлено в авторській публікації [99].

На відміну від попередніх підпунктів, де основна увага приділялася контрольованому навчанню, у цьому випадку використано некеровані нейромережеві методи кластеризації. У роботі для цього було обрано два алгоритми, реалізовані в середовищі WEKA: Self-Organizing Map (самоорганізаційна карта Кохонена, SOM) та Learning Vector Quantization (квантування навчальних векторів, LVQ). Обидва підходи належать до нейромережевих методів групування, але відрізняються логікою формування кластерів. SOM відображає багатовимірний простір у низьковимірне впорядковане представлення та дає змогу виявляти приховану топологію даних. LVQ працює через набір репрезентативних векторів і фактично виконує розподіл об'єктів за принципом близькості до прототипів. Використання саме цих двох алгоритмів є доцільним, оскільки вони дозволяють перевірити, чи зберігається структура даних не лише в задачі класифікації, а й на рівні природного кластерного поділу.

У межах запропонованого підходу кластеризацію виконано для трьох представлень набору Wisconsin Breast Cancer Dataset: ds1, ds2 та ds3. Набір ds1 відповідає повному стандартизованому ознаковому опису, ds2 – скороченому поданню після відбору ознак, а ds3 – найбільш компактному представленню після переходу до чотирьох головних компонент. Саме таке поетапне зіставлення є принципово важливим: воно дає можливість перевірити, чи не руйнується структура даних у процесі скорочення розмірності.

Запропонований метод включає кілька послідовних кроків. На першому етапі для кожного з наборів ds1–ds3 виконують кластеризацію засобами SOM та LVQ. На другому етапі в середовищі WEKA застосовують режим classes-to-

clusters evaluation (оцінка відповідності еталонних класів та кластерів), який дозволяє зіставити сформовані кластери з еталонними класами malignant (злаякісна) та benign (доброякісна). На третьому етапі за елементами одержаної матриці відповідності обчислюють індекс FMI (індекс Фоулкса-Меллоу), що кількісно характеризує ступінь узгодженості між кластерною структурою і реальним поділом об'єктів на класи. На завершальному етапі одержані значення порівнюють між собою для різних алгоритмів і різних представлень даних.

Метрика FMI є зручною саме для цього завдання, оскільки вона відображає, наскільки добре збігаються парні відношення типу «ці два об'єкти потрапили в один кластер» та «ці два об'єкти належать до одного класу». На практиці значення FMI, близькі до одиниці, свідчать про високу узгодженість між кластерами та еталонними класами, тоді як нижчі значення вказують на перекриття класів, шум або нестійкість групування [99].

Формально індекс Фоулкса-Меллоу для двокласового випадку можна подати так:

$$FMI_{positive} = \frac{TP}{\sqrt{(TP+FP)(TP+FN)}}, FMI_{negative} = \frac{TN}{\sqrt{(TN+FP)(TN+FN)}} \quad (3.11)$$

де TP, FP, FN, TN – відповідні елементи матриці відповідності класів і кластерів. Такий запис показує, що індекс одночасно враховує і правильні парні збіги, і помилки віднесення, а тому є більш змістовним, ніж проста візуальна оцінка кластерного поділу.

Результати оцінювання узгодженості між кластерами та еталонними класами за індексом Фоулкса-Меллоу наведено на рисунку 3.7.

Як видно з рисунку 3.7, для обох алгоритмів і для всіх трьох представлень даних значення FMI перебувають у межах 0,81–0,99. Це дає підстави говорити про достатньо високу узгодженість між кластерами, сформованими в процесі некерованого групування, та реальними класами набору WBCD. Інакше кажучи, навіть після редукції ознак об'єкти не втрачають своєї природної схильності до групування відповідно до діагностичного поділу.

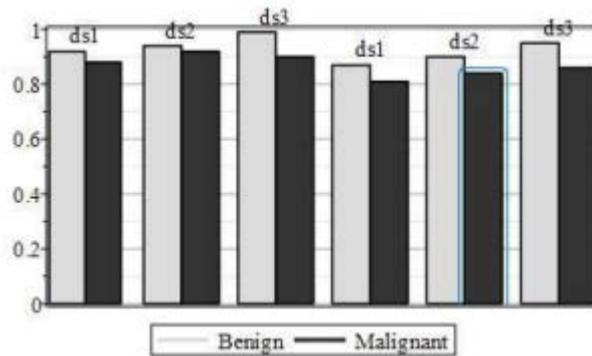


Рисунок 3.7 – Індеси Фуулкса-Меллоу (FMI) для оцінювання узгодженості кластерів і класів у SOM та LVQ на наборах ds1–ds3

Разом із тим одержані результати не слід трактувати як ознаку ідеального розділення класів. Для реальних медичних даних це взагалі не є типовою ситуацією. Навіть за наявності доброякісних і злоякісних випадків частина спостережень закономірно перекривається в просторі ознак, а також можуть бути присутні шумові значення та нетипові випадки. Саме тому в цьому підході важливо не вимагати від кластеризації абсолютної сепарації, а оцінювати, чи зберігається загальна кластерна закономірність після скорочення розмірності.

Окремо слід зазначити, що в роботі аналізувалася не лише узгодженість класів і кластерів, а й ступінь відокремленості самих кластерів. Для цього використовували внутрішній критерій Variance Ratio Criterion (VRC). Одержані значення 1,78 для LVQ та 1,87 для SOM є відносно низькими, що вказує на неповне розділення кластерів. Проте цей результат не суперечить оцінкам за FMI. Він лише показує, що кластери в даних WBCD не є ідеально відокремленими, однак попри це вони достатньо добре відображають реальний класовий поділ. У прикладному сенсі це навіть важливіше за формальну «красу» кластерів, оскільки підтверджує збереження змістовної структури даних. Для зручності інтерпретації основні висновки щодо кластерної структури наборів ds1–ds3 можна звести до таблиця 3.12.

Наведені результати підтверджують, що перехід до скорочених представлень ds2 і ds3 не руйнує внутрішню організацію даних. Це особливо важливо в контексті всієї роботи, оскільки дає додаткове, незалежне від

класифікаційних метрик підтвердження коректності редукції ознак. Якщо б після скорочення розмірності кластерна структура руйнувалася, то навіть за прийнятних значень Accuracy можна було б сумніватися в змістовній придатності редукованих даних. Однак оцінки за FMI показують, що цього не відбувається.

Таблиця 3.12 – Узагальнена інтерпретація результатів валідації кластерної структури наборів ds1–ds3

| Алгоритм | Представлення даних | Діапазон значень FMI | Інтерпретація |
|----------|---------------------|----------------------|--|
| SOM | ds1–ds3 | 0,81–0,99 | Висока узгодженість кластерів з еталонними класами |
| LVQ | ds1–ds3 | 0,81–0,99 | Висока узгодженість кластерів з еталонними класами |
| SOM, LVQ | ds1–ds3 | VRC: 1,78–1,87 | Кластери не є ідеально розділеними, але відображають класову структуру |

Отже, у роботі запропоновано метод валідації кластерної структури даних за метрикою FMI, який базується на зіставленні результатів SOM та LVQ кластеризації з еталонними класами набору WBCD. Метод дозволив показати, що для представлень ds1–ds3 значення індексу Фоулкса-Меллоу перебувають у межах 0,81–0,99, тобто відповідають достатньо високому рівню узгодженості між кластерами і реальними класами. Це свідчить про те, що навіть після скорочення ознакового простору дані зберігають змістовну внутрішню

структуру, а отже редуковані представлення можуть бути використані не лише для класифікації, а й для подальшого аналізу групування об'єктів.

Висновки до розділу 3

У третьому розділі здійснено практичну реалізацію моделей і методів інтелектуального аналізу даних із використанням WEKA у складі багаторівневої системи моніторингу стану здоров'я людини. Запропоновано модель інтеграції інструментарію WEKA до серверного контуру системи на базі Python/Django, у якій формування вибірок виконується засобами Django ORM, а передавання даних до аналітичного модуля реалізується через формат ARFF, що забезпечує відокремлення робочого та аналітичного контурів, відтворюваність експериментів і зменшення залежності від структури бази даних.

Для набору Coimbra Breast Cancer Dataset розроблено метод попередньої оцінки його цінності для задач первинного скринінгу, який поєднує ранжування ознак, статистичне виділення підгруп релевантності та оцінювання зашумленості. Установлено, що найбільш релевантними є ознаки Glucose, Age, НОМА і Resistin, тоді як набір містить помітну частку шуму та аномальних спостережень, що обґрунтовує необхідність обережного використання цього набору і доцільність скорочення ознакового простору перед побудовою діагностичних моделей.

Для набору Wisconsin Breast Cancer Dataset реалізовано метод послідовного скорочення вимірності з формуванням наборів ds1, ds2 і ds3, що забезпечив перехід від 30 початкових ознак до 11 відібраних атрибутів і далі до 4 головних компонент, тобто зменшення розмірності на 86,7 % при збереженні 91 % загальної дисперсії даних. Показано, що таке скорочення не призводить до критичної втрати прогностичної здатності, оскільки частка правильно класифікованих екземплярів для досліджених моделей залишалася в межах 95,87–97,68 %.

На основі порівняльного аналізу ефективності діагностичних рішень у WEKA встановлено, що для набору Коїмбри доцільно використовувати дерево

рішень J48 як компроміс між якістю та інтерпретованістю, а для наборів ds1–ds3 підтверджено збереження високих значень Accuracy і PRC area навіть після скорочення ознакового простору. Додатково побудовано інтерпретовану модель діагностичного рішення на основі дерева J48 для WBCD, яка має 13 вузлів і 7 листків та характеризується показниками precision = 0,967, MCC = 0,929, ROC area = 0,978. Також побудовано модель індукції діагностичних правил на основі JRip для CBCD, яка на скороченому чотириатрибутному поданні формує компактні правила типу «IF – THEN» і забезпечує середню точність $81,6 \pm 8,6\%$. У цілому результати розділу підтверджують практичну придатність запропонованих моделей і методів для побудови компактних, інтерпретованих і придатних до програмної реалізації діагностичних рішень у багаторівневих системах моніторингу стану здоров'я людини.

РОЗДІЛ 4

ВПРОВАДЖЕННЯ ТА ЕКСПЕРИМЕНТАЛЬНА АПРОБАЦІЯ РОЗРОБЛЕНОЇ СИСТЕМИ МОНІТОРИНГУ

4.1 Впровадження результатів дослідження

4.1.1 Впровадження результатів дослідження в НДР № 0121U109898

Результати дослідження методів стиснення інформації були впроваджені під час виконання державної бюджетної НДР «Розробка модулів автоматизації бездротових приладів відновлення пост-інфарктних, пост-інсультних пацієнтів в індивідуальних умовах віддаленої реабілітації» (№ держ. реєстрації 0121U109898), що виконувалася у Чорноморському національному університеті імені Петра Могили у 2021–2022 рр.

У межах зазначеної НДР було використано результати порівняльного аналізу алгоритмів стиснення LZ77, LZ78, LZW та RLE для різних типів даних з метою обґрунтування доцільності їх застосування у медичній галузі. У матеріалах дослідження показано, що ефективність стискання істотно залежить від типу вхідних даних: для розрізнених текстових послідовностей алгоритми RLE є малоефективними, тоді як для даних із довгими серіями однакових символів їх застосування є доцільним; алгоритми сімейства LZ забезпечують кращі результати для текстових послідовностей за рахунок використання словникового підходу.

Практичне значення для НДР мали одержані кількісні оцінки. Зокрема, в роботі наведено, що кодування текстових даних засобами української мови характеризується в середньому на 20–25 % більшою надлишковістю порівняно з аналогічними даними англійською мовою, що враховувалося при аналізі потенціалу стискання текстової інформації. Крім того, при порівнянні базових алгоритмів LZ77 і LZ78 для однакових умов кодування показано, що LZ78 є ефективнішим: для окремого символу виграш становить приблизно 1,5 раза, а для кодованої послідовності – до 2 разів. Також у табличному порівняльному аналізі для тестових текстових даних отримано коефіцієнти стиснення 1,20 для

LZ77 і 1,42 для LZ78, а для рядка з повторюваними символами – відповідно 1,10 і 1,43.

Отримані результати були використані у НДР як аналітична основа для вибору підходів до зменшення надлишковості даних у бездротових приладах відновлення та віддаленої реабілітації, де обсяг передавання інформації безпосередньо впливає на швидкодію обміну та вимоги до зберігання даних. Факт впровадження зазначених результатів підтверджується актом впровадження результатів дисертаційної роботи.

4.1.2 Впровадження результатів дослідження в НДР № 0125U000904

Практичне впровадження результатів дисертаційної роботи підтверджено актом впровадження при виконанні НДР № 0125U000904 «Методи та засоби обробки медичних сигналів для підвищення ефективності діагностичних систем» у Чорноморському національному університеті імені Петра Могили. У межах впровадження використано модель контуру інтеграції «edge – сервер – аналітика WEKA», у якій дані, отримані від периферійного рівня, після надходження на серверний рівень зберігаються у базі даних MySQL, далі формуються у вибірки для аналізу та конвертуються у формат ARFF для подальшої обробки засобами WEKA.

Технічні засади цього рішення висвітлено в публікаціях автора. У статті «Архітектура багаторівневої системи моніторингу стану здоров'я людини» наведено концептуальну модель багаторівневої архітектури на основі Python та Django, описано використання MySQL, кількох паралельних вебсерверів і балансувача навантаження, що забезпечує функціонування системи на основі хмарних технологій. У статті «Аналіз даних в системі моніторингу стану здоров'я людини засобами Python» обґрунтовано використання Django ORM, MySQL та формату ARFF для підготовки даних до аналітичної обробки. У статті «Інтеграція інструментарію WEKA до багаторівневої системи моніторингу стану здоров'я людини» безпосередньо описано підключення WEKA до MySQL, формування ARFF-наборів із використанням Django ORM та виконання

класифікації засобами python-weka-wrapper3, зокрема із застосуванням алгоритму J48.

Отже, впровадження в межах НДР полягає у практичному використанні інтегрованого контуру «edge – сервер – аналітика WEKA», який забезпечує послідовний ланцюг збирання, зберігання, підготовки та інтелектуального аналізу медичних даних у телемедичних сценаріях.

4.1.3 Впровадження результатів дисертаційної роботи у навчальний процес

Результати дисертаційної роботи впроваджено у навчальний процес кафедри комп'ютерної інженерії Чорноморського національного університету імені Петра Могили під час викладання дисциплін «Сенсори, перетворювачі» та «Machine Learning and Pattern Recognition» у вигляді лекційних матеріалів і методичних матеріалів для виконання практичних робіт.

Зокрема, до навчальних матеріалів включено тему «Використання технологій ІоМТ (інтернет медичних речей) для збору та попередньої обробки даних серцевого ритму», сформовану на основі результатів, висвітлених у публікаціях «Архітектура багаторівневої системи моніторингу стану здоров'я людини» та «Аналіз даних в системі моніторингу стану здоров'я людини засобами Python», у яких розглянуто концепцію Internet of Medical Things, особливості збору біометричних показників, зокрема серцевого ритму, багаторівневу організацію системи моніторингу, а також засоби первинного опрацювання та підготовки даних до подальшого аналізу.

Також до навчально-методичного забезпечення включено тему «Використання алгоритмів зниження розмірності та глибокого навчання для обробки датасетів біомедичних даних», підготовлену на основі публікацій «Dimensionality cutback and deep learning algorithms efficacy as to the breast cancer diagnostic dataset», «Breast Cancer Dataset from Coimbra Pre-Ratings of Its Value to Machine Learning and Diagnosis» та «Attribute Selection, Outliers Impact Study, and Visualization within Breast Cancer Detection», у яких висвітлено питання відбору

ознак, скорочення розмірності, візуалізації даних і застосування алгоритмів глибокого навчання до діагностичних біомедичних наборів даних.

Отже, впровадження в навчальний процес полягає у використанні результатів дисертаційної роботи як основи для підготовки лекційних і практичних матеріалів на кафедрі комп'ютерної інженерії ЧНУ імені Петра Могили.

4.1.4 Організація рівня даних та структура БД (MySQL)

Рівень даних у реалізованій багаторівневій системі моніторингу стану здоров'я виконує функції централізованого накопичення, зберігання та впорядкування медичної й службової інформації, необхідної для подальшої аналітичної обробки. У запропонованому програмному контурі цей рівень реалізовано на базі MySQL, що забезпечує зберігання результатів вимірювань, даних пацієнтів, відомостей про пристрої, а також службових сутностей вебзастосунку [91, 92].

Структуру бази даних спроектовано з урахуванням потреб моніторингової системи та особливостей реалізації серверної частини засобами Django. До її складу входять таблиці, що охоплюють: профілі користувачів, медичні показники та часові ряди вимірювань, характеристики пристроїв, а також дані автентифікації, авторизації, журналювання й інші сервісні об'єкти. Загалом у реалізованій системі структура БД містить 20 таблиць, взаємозв'язки між якими забезпечують цілісність даних і підтримують роботу прикладного та аналітичного контурів системи.

Як видно з рисунок 4.1, організація рівня даних орієнтована на довготривале накопичення результатів спостереження, що є характерним для задач дистанційного моніторингу стану здоров'я. Такий підхід дає змогу зберігати не лише окремі вимірювання, а й контекст їх отримання: часові мітки, джерело надходження, тип пристрою та службові параметри обробки. Це створює основу для подальшого формування узгоджених вибірок, їх експорту у формат ARFF та використання в аналітичному модулі WEKA [91].

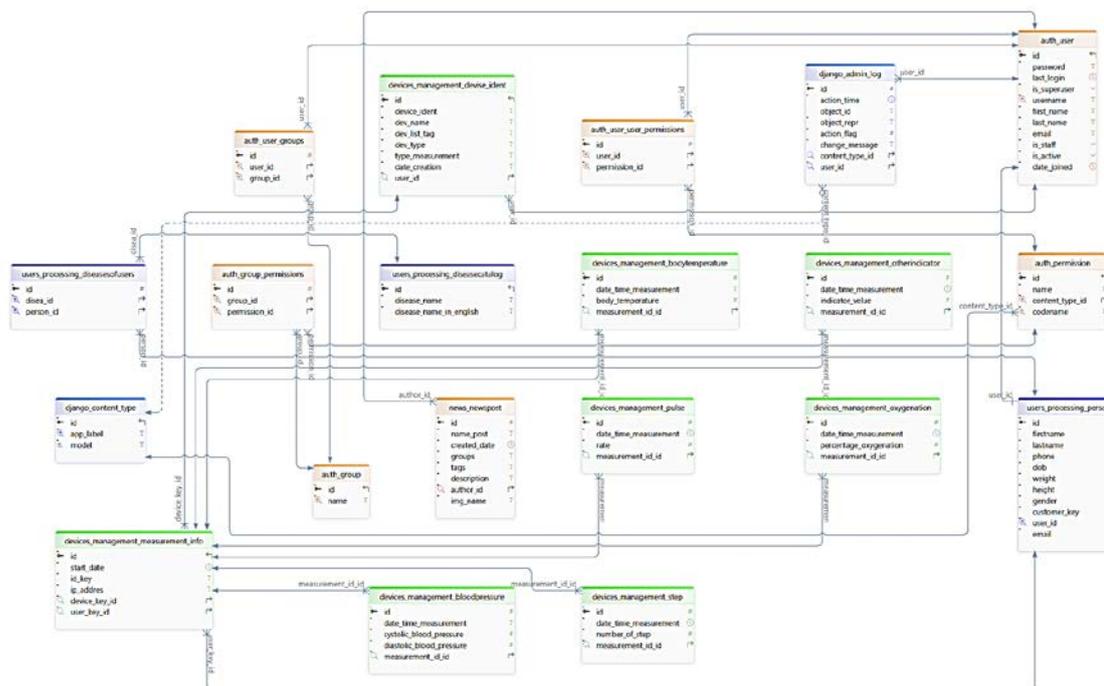


Рисунок 4.1 – Структура БД наведеної системи моніторингу стану здоров'я людини

Взаємодія серверного рівня з базою даних здійснюється через ORM-механізм Django, що забезпечує зручну роботу з даними на рівні прикладної логіки, спрощує супровід програмного коду та зменшує кількість ручних SQL-операцій. У результаті реалізований рівень даних виступає не лише сховищем медичної інформації, а й проміжною ланкою між підсистемою збору даних і контуром інтелектуального аналізу.

4.1.5 Модуль збору, попередньої обробки та серверної аналітики ІоМТ-даних

Розроблена програмна система моніторингу стану здоров'я орієнтована на роботу з потоком вимірювань, що надходять від розподілених джерел у парадигмі ІоМТ. У такому сценарії ключову роль відіграє модуль, який з одного боку забезпечує приймання показників (пульс, тиск, температура, оксигенація тощо), а з іншого – виконує їх первинну перевірку, нормалізацію, агрегування та підготовку до подальшого аналізу в модулях штучного інтелекту. Змістовно цей модуль є «містком» між рівнем збору даних (things) та серверно-аналітичним

контуром (fog/cloud), оскільки саме на ньому вирішуються типові проблеми реальних медичних вимірювань: нерівномірність у часі, пропуски, шум, різна точність сенсорів і залежність показників від характеристик пристрою та умов зняття даних. Відповідні принципи проектування узгоджуються з підходами, викладеними у публікаціях автора щодо багаторівневої архітектури та аналізу даних у системах моніторингу стану здоров'я. У сукупності така організація створює основу для побудови інтелектуальних систем моніторингу стану здоров'я, у яких первинне збирання даних поєднується з їх автоматизованою попередньою обробкою та подальшою серверною аналітикою. Подібні підходи до попередньої обробки та інтелектуального аналізу часових біомедичних сигналів були апробовані також на ЕМГ-даних під час аналізу патернів дихання уві сні [92, 95, 102].

На етапі збору даних система приймає вимірювання від мобільного застосунку або інтегрованих сенсорних IoT-пристроїв у вигляді повідомлень, що містять: ідентифікатор користувача (пацієнта), тип показника, значення, часову мітку та (за можливості) ідентифікатор пристрою. Така організація обміну даними забезпечує віддалений доступ до сервісів моніторингу, за якого користувач або периферійний пристрій може передавати показники до серверного контуру без безпосередньої локальної взаємодії з ним. Наявність інформації про пристрій є важливою, оскільки точність і стабільність вимірювань можуть суттєво відрізнятися для різних класів обладнання; це впливає як на якість аналітики, так і на коректність подальших діагностичних висновків. Тому в базі даних поряд із медичними параметрами зберігаються метадані про пристрій та службові сутності застосунку (структура БД наведена в пункті 4.1.4, рисунок 4.1). Сама система може на пряму збирати дані через GET/POST запити або імпортувати зібрані дані у форматі CSV, що дозволяє використовувати дані зібрані сторонніми сервісами, наприклад Google Fit (рисунок 4.2).

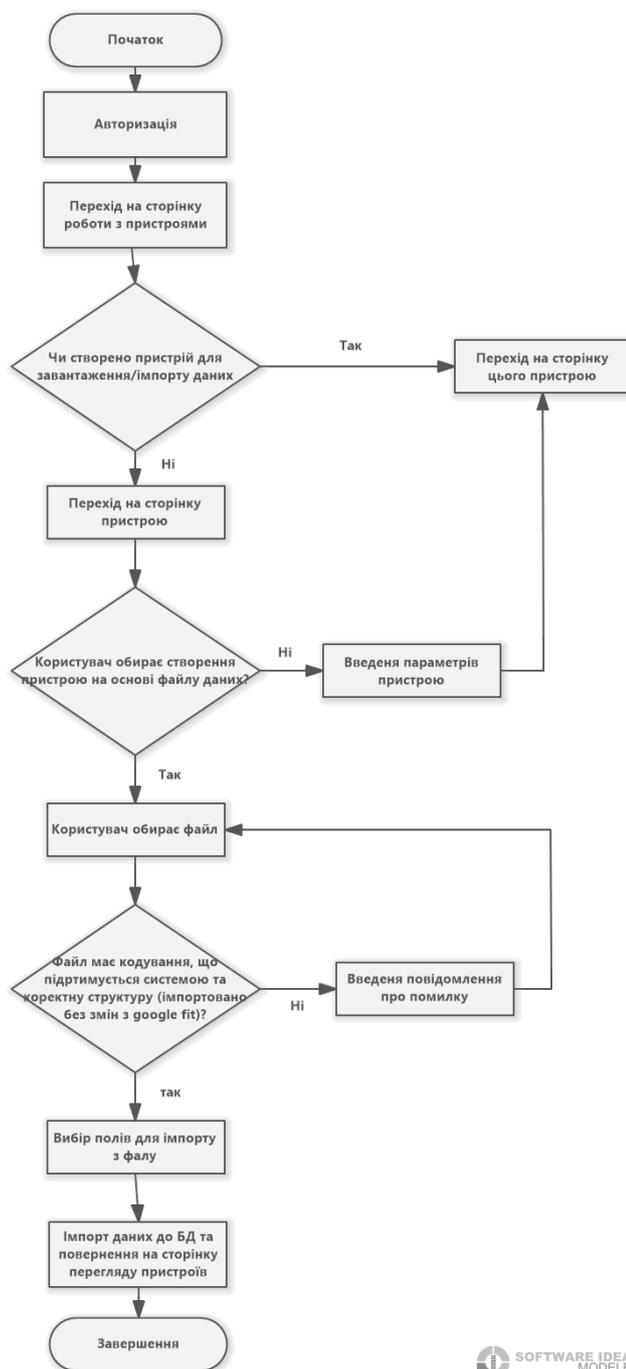


Рисунок 4.2 – Блок-схема процесу імпорту даних

Попередня обробка даних на серверному рівні передбачає кілька послідовних кроків [90].

Перший крок – валідація вхідних даних. Перевіряється коректність типів, діапазонів значень, наявність обов’язкових полів, узгодженість одиниць вимірювання. На практиці це захищає систему від «помилкових пакетів» і некоректних значень, що виникають під час збоїв сенсорів, розриві зв’язку або помилці клієнтського ПЗ.

Другий крок – нормалізація часових міток. Вхідні дані можуть надходити нерівномірно (наприклад, частіше під час активності користувача і рідше вночі), а також із часовими зсувами. Тому на сервері доцільно приводити час до єдиного формату та зберігати як «час вимірювання» (з боку пристрою), так і «час отримання» (з боку сервера). Це спрощує аудит і відтворюваність аналізу.

Третій рівень – обробка пропусків і дублювань. Для сенсорних даних типовими є прогалини та повтори (повторне надсилання пакета після відновлення мережі). На цьому кроці застосовується дедуплікація та логіка «останнє значення – найсвіжіша мітка» для повторів. Для пропусків важливо відокремлювати «відсутність вимірювання» від «нульового значення», оскільки це різні ситуації з точки зору інтерпретації.

Четвертий рівень – фільтрація шуму та згладжування. Для окремих показників (наприклад, пульс) можливі короткі сплески через артефакти датчика. Залежно від мети аналізу може застосовуватись просте ковзне усереднення або медіанна фільтрація на короткому вікні, щоб зменшити вплив випадкових піків без спотворення загального тренду. Важливо, щоб параметри фільтрації були фіксовані й описані в роботі, оскільки це впливає на результати порівняння моделей [103, 104].

П'ятий рівень – уніфікація формату зберігання. Після перевірок і нормалізації запис зберігається у БД у стандартизованому вигляді. За рахунок цього подальші процедури аналізу можуть працювати за узгодженою схемою даних і повторно використовувати однакові запити.

Таблиця 4.1 – Етапи попередньої обробки даних

| Етап | Мета | Результат |
|-------------------|--------------------------|-------------------------------|
| Валідація | Контроль формату | Відсів некоректних пакетів |
| Нормалізація часу | Узгодження часових міток | Коректна часово-рядна історія |

Продовження таблиці 4.1

| Етап | Мета | Результат |
|------------------------|-----------------------------|--------------------------|
| Пропуски або дублікати | Відновлення послідовності | Очищена серія вимірювань |
| Фільтрація шуму | Зниження артефактів сенсора | Стабільні тренди |
| Уніфікація зберігання | Підготовка до аналітики | Узгоджені записи в БД |

Окремо слід врахувати питання безпеки, оскільки система обробляє медичні дані. На рівні вебзастосунку доцільно використовувати стандартні механізми захисту, притаманні Django (рольова модель доступу, захист від підміни запитів), а під час передавання даних – захищений канал (HTTPS). У роботі щодо аналізу даних у системі моніторингу також розглядається підхід до посилення захисту через застосування криптографічних бібліотек для шифрування даних під час пересилання, що є релевантним для IoT-сценаріїв.

Після накопичення даних у БД наступним завданням є їхній оперативний і зрозумілий аналіз для користувача (лікаря або пацієнта). На практиці аналітичний модуль виконує дві функції: формує «аналітичні зрізи» даних із БД (вибірki за пацієнтом, періодом, типом показника), реалізує статистичний і візуальний аналіз, який потрібний як для первинної інтерпретації стану, так і для подальшої підготовки даних до інтелектуальних моделей. У статті [92] про застосування Python для аналізу даних у системі моніторингу зазначено доцільність використання спеціалізованих бібліотек Pandas, Matplotlib та Seaborn, які є стандартом де-факто для такого класу задач.

У реалізації модуль може працювати у двох режимах. Перший – інтегрований, коли аналіз виконується всередині Django-застосунку на основі вибірок, сформованих ORM-запитами. Другий – сервісний, коли аналітика винесена в окремий компонент (мікрсервіс), що самостійно звертається до БД і повертає результати у вигляді агрегованих даних. Обидва підходи є сумісними з

загальною багаторівневою архітектурою: інтегрований режим простіший для впровадження, а сервісний підвищує масштабованість і дозволяє ізолювати важкі обчислення від вебчастини [90, 92].

Типовий цикл аналітики виглядає наступним чином:

- формування DataFrame на основі вибірки з БД (ключові поля: час, значення, тип показника, ідентифікатор пацієнта або пристрою);
- описова статистика: кількість спостережень, середнє, стандартне відхилення, кuartилі, мін/макс. Це корисно як «контроль якості» даних і як швидка діагностична характеристика;

На рисунку 4.3 наведено приклад первинного статистичного опису вибірки засобами Pandas, що дозволяє швидко оцінити межі, варіативність та наявність викидів.

```
import numpy as np
import pandas as pd
import mysql.connector as sql

db_connection = sql.connect(host='127.0.0.1', database='customer_health',
                             user='****', password='*****')

pd.set_option("display.precision",2)
df = pd.read_sql('SELECT * FROM devices_management_pulse where measurement_id=2',
                 con=db_connection)
print(df['rate'].describe())
```

✓ 0.3s

| | |
|----------------------------|--------|
| count | 500.00 |
| mean | 74.31 |
| std | 6.45 |
| min | 63.00 |
| 25% | 69.00 |
| 50% | 75.00 |
| 75% | 80.00 |
| max | 85.00 |
| Name: rate, dtype: float64 | |

Рисунок 4.3 – Приклад первинного статистичного аналізу даних

- аналіз розподілів і трендів: гістограми, лінійні графіки, порівняння сегментів (наприклад, «до/після» лікувального втручання). Приклад аналізу розподілу показника за різні часові зрізи подано на рисунку 4.4, що дає змогу виявляти зміщення розподілу та потенційні аномалії;

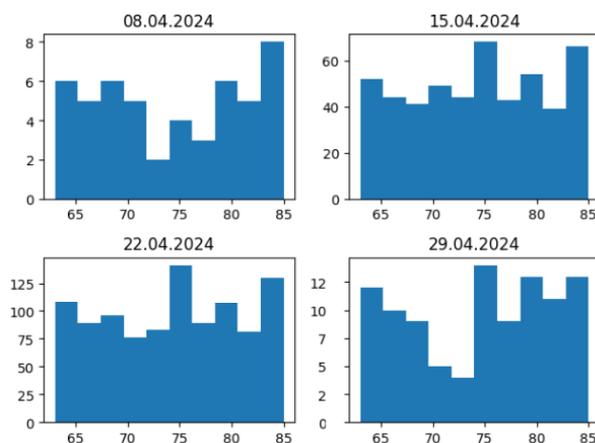


Рисунок 4.4 – Гістограма розподілу біомедичного показника за різними часовими зрізами

– кореляційний аналіз (за потреби): оцінка взаємозв'язків між показниками, що може бути корисним для відбору ознак або пошуку вторинних залежностей.

На рисунку 4.5 показано приклад візуального аналізу взаємозв'язку двох показників засобами Seaborn, що використовується для попередньої оцінки залежностей між ознаками.

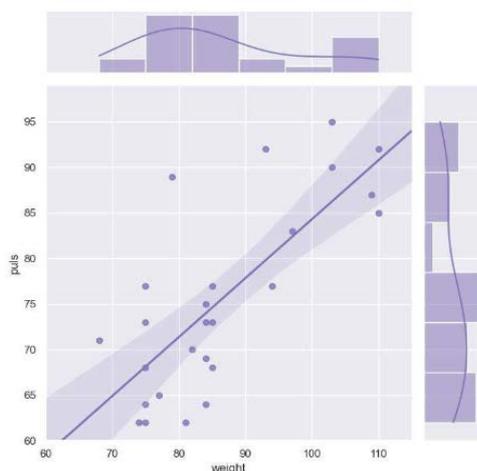


Рисунок 4.5 – Діаграма розсіювання з лінією регресії та маргінальними розподілами Seaborn для оцінювання взаємозв'язку показників

У підсумку аналітичний модуль виконує не лише роль «візуалізатора», а й забезпечує підготовку даних до застосування інтелектуальних моделей. Зокрема, результати попередньої обробки та структуризації можуть експортуватися у формат, придатний для подальшого використання в WEKA (наприклад, у вигляді

ARFF), або передаватися в модуль машинного навчання як підготовлена вибірка. Це узгоджується з підходами інтеграції інструментарію WEKA до багаторівневих систем моніторингу, де аналітичний компонент може працювати як окрема підсистема, що запускає алгоритми навчання та повертає результати у вебінтерфейс [90, 92, 102].

4.1.6 Інтеграція WEKA в програмний контур Django/Python

Інтеграція інструментарію WEKA у програмний контур вебсистеми є ключовою умовою практичного використання експериментально перевірених моделей (п. 3.4) у режимі реального часу. Основна складність полягає в різниці технологічних стеків: серверна частина системи побудована на Python/Django, тоді як WEKA реалізована на Java. Тому в роботі розглянуто кілька взаємодоповняльних підходів інтеграції – від «файлового» обміну через ARFF до прямого виклику Java-API з Python через спеціалізовані bridge-бібліотеки.

Для обміну даними між серверною частиною та WEKA використовується спеціалізований формат Attribute-Relation File Format (ARFF), який є базовим для завантаження даних у WEKA та містить опис відношення (relation), атрибутів (імена і типи) і блоку даних. З погляду програмного контуру це дозволяє зафіксувати «контракт даних»: порядок ознак, їхні типи, домени номінальних значень і назву класової змінної – тобто все, що критично для відтворюваності прогнозів та уникнення помилок сумісності.

Формування наборів даних виконується на основі даних, отриманих із БД через Django ORM (запити з фільтрацією, зв'язками, відбором полів). Далі ORM-дані можуть бути перетворені в ARFF за допомогою бібліотеки liac-arff, яка дозволяє описати ARFF у вигляді структури Python dictionary (description, relation, attributes, data). Такий спосіб зручний тим, що сервер одразу генерує ARFF з «живих» даних системи моніторингу й не залежить від ручних операцій експорту.

У межах системи доцільно розглядати два режими інтеграції.

Перший – це вбудований режим WEKA у середовищі Python. У цьому випадку виклик алгоритмів WEKA виконується безпосередньо з Python через bridge-бібліотеку. На практиці для інтеграції Java-WEKA з Python застосовують, зокрема, бібліотеки на кшталт weka3.pyweka3 та подібні рішення, які дають змогу завантажувати дані, запускати алгоритми й інтерпретувати результати з Python-коду. У вже сформованому програмному контурі (Django) подібний механізм зручно реалізується як окремий сервісний модуль, який:

- ініціалізує JVM (Java Virtual Machine);
- завантажує попередньо навчений класифікатор;
- приймає на вхід дані пацієнта у форматі JSON.dictionary;
- конвертує їх у ARFF.Instances;
- повертає клас і, за потреби, оцінки впевненості.

Другий – винесений режим. Практична особливість полягає в тому, що інтеграція WEKA через python-bridge потребує додаткового інструментарію на сервері (а саме: JDK/JVM та інші залежності), що збільшує навантаження на вузол, який паралельно приймає дані від ІоМТ-пристроїв, і підсилює потребу в моніторингу енергоспоживання обчислювального контуру. Тому для масштабованих сценаріїв доцільно відокремлювати «вузол аналітики» (WEKA-контур) від «вузла збору/вебінтерфейсу» (Django-контур). Це дозволяє:

- ізолювати важкі обчислення;
- спростити експлуатацію (оновлення JDK/WEKA не зачіпає вебчастину);
- збалансувати ресурси під реальні навантаження.

Для регулярної роботи з наборами, які постійно оновлюються, а дані постійно доповнюються новими вимірюваннями, зручним є механізм завантаження даних у WEKA безпосередньо через HTTP у межах використання хмарних технологій. WEKA підтримує роботу з віддаленими файлами у форматі ARFF або CSV за URL-адресою без попереднього локального збереження, у поєднанні з хмарними технологіями це спрощує автоматизацію тестування моделей і доступ до актуальних даних «майже в реальному часі».

У такій схемі Django може виступати постачальником набору даних: формувати в формат ARFF динамічно (за періодом, пацієнтом, типом вимірювання), а WEKA – споживачем, який завантажує цей ARFF по URL. Подібний підхід також зручний для відділення контурів і узгоджується з використанням хмарних технологій: WEKA працює на окремому пристрої, але має контрольований доступ до актуальних наборів даних.

Сценарії використання інтеграції: онлайн-прогноз і batch-перенавчання. На рівні прикладної логіки доцільно розвести два сценарії.

Онлайн-сценарій (оперативний прогноз).

Вебінтерфейс або API ініціює діагностичний запит; сервер формує вхідний вектор ознак (з останніх вимірювань чи агрегатів) і передає його в WEKA-контур. У вбудованому режимі це відбувається як прямий виклик класифікатора з Python після конвертації JSON в ARFF/Instances; результат повертається у Django-представлення для відображення. Критичний момент є сталість схеми ознак: атрибути мають подаватися в тому ж порядку і з тими ж типами, що й під час навчання моделі. Batch-сценарій (періодичне навчання або переоцінка).

Система (за розкладом або після накопичення достатнього обсягу нових даних) формує оновлений набір даних через ORM і експортує його в ARFF, після чого на вузлі WEKA виконується перенавчання та валідація. Перевага в тому, що такий режим можна реалізувати або на окремому пристрої з прямим підключенням до БД, або через HTTP-завантаження актуального ARFF. Результатом batch-процесу є нова версія моделі (наприклад, серіалізований файл), яка далі підвантажується в онлайн-контур для прогнозування.

Описані механізми інтеграції узгоджуються з багаторівневою архітектурою (presentation – logic – data):

- Data layer надає дані через ORM/SQL;
- Logic layer формує ARFF і керує викликом моделей;
- Presentation layer відображає прогноз і візуальні пояснення.
- З практичного погляду, під час впровадження WEKA в продакшн-контур слід враховувати:

контур слід враховувати:

- збільшення вимог до серверного середовища у вбудованому режимі (JDK/JVM та залежності), через що доцільно виділяти окремий вузол аналітики;
- необхідність контрольованого доступу до наборів даних під час HTTP-завантаження (автентифікація, токени), оскільки медичні дані є чутливими;
- версіонування моделей і фіксацію схеми ознак (ARFF-контракт) для забезпечення відтворюваності прогнозів.

У підсумку інтеграція WEKA в програмний контур Django може бути реалізована як в ролі вбудованого механізму із прямим викликом Java-моделей, так і як сервісна підсистема на окремому вузлі. Обидва підходи підтримують автоматизоване формування наборів даних через ORM, стандартизацію обміну через ARFF та подальше використання моделей WEKA в задачах медичного моніторингу й діагностики. Наведений приклад розглядається як демонстрація механізму автоматизованого відбору ознак у «WEKA KnowledgeFlow», у практичному контурі системи аналогічний підхід застосовується до сформованих з БД ARFF-наборів, що використовуються для навчання моделей.

4.1.7 Структурно-семантичний аналіз інформаційних потоків та стратегія попередньої обробки даних

На етапі системного аналізу предметної області було виявлено, що вхідні дані мають суттєві відмінності у своїй природі, ентропії та статистичному розподілі. Це зумовило необхідність розділення попередньої обробки на два паралельні конвеєри [95].

Аналіз метаболічного профілю та антропометрії (вектор біохімічних ознак)

Перший інформаційний потік базується на концепції «рідкої біопсії» та скринінгових гематологічних тестах. У якості референсної моделі для цього потоку в роботі використано набір даних про рак молочної залози Коїмбри (CBCD). Важливість цього набору полягає в тому, що він відображає системні зміни в організмі, які часто передують візуалізації пухлини на мамографії.

Цей набір даних формує вектор ознак, що включає дев'ять ключових параметрів (таблиця 4.2).

Таблиця 4.2 – Структура вектору ознак набору даних CBCD

| Група показників | Атрибути | Обґрунтування та особливості |
|---------------------------------------|---------------------------|--|
| Антропометричні показники | Вік (Age) | Базові фізичні параметри. Відомо, що надлишкова вага корелює з підвищеним рівнем естрогенів, що є визнаним фактором ризику розвитку патологій. |
| | Індекс маси тіла (BMI) | |
| Показники вуглеводного обміну | Глюкоза (Glucose) | Вимірюються натщесерце. Індекс HOMA-IR дозволяє оцінити наявність інсулінорезистентності, що часто супроводжує метаболічні порушення. |
| | Інсулін (Insulin) | |
| | HOMA-IR | |
| Адіпокіни та маркери запалення | Лепин | Специфічні білки, що виділяються жировою тканиною або імунними клітинами. Вони виступають біомаркерами запальних процесів та енергетичного гомеостазу. |
| | Адопонектин (Adiponectin) | |
| | Резистин (Resistin) | |
| | MCP-1 | |

З точки зору Data Science, цей потік даних характеризується низькою розмірністю, але високою щільністю інформації. Тут відсутня просторова кореляція, проте спостерігаються складні нелінійні залежності (наприклад,

співвідношення лептину до адипонектину часто є більш інформативним, ніж їхні абсолютні значення). Тому для цього типу даних головним викликом є не зменшення розмірності, а пошук прихованих закономірностей класифікації.

Другий, значно об'ємніший потік даних, формується в результаті цифрової обробки зображень тонкоголкової аспіраційної біопсії (англ. Fine Needle Aspiration, FNA). Цей етап спирається на вісконсинський набір даних про рак молочної залози (WBCD). На відміну від біохімії, тут увага зосереджена на геометричних примітивах, що описують форму клітинних ядер. Аналогічна логіка роботи з візуальними медичними потоками даних характерна і для камерних підсистем у медичних роботизованих комплексах [93, 105].

Комп'ютерний зір трансформує аналогове зображення у числовий масив для отримання різних метрик для кожного ядра, наведених у таблиці 4.3.

Таблиця 4.3 – Параметри аналогових зображень

| Атрибут | Опис |
|----------------------------|---|
| Radius (повнота) | Середня відстань від центру ядра до точок на його периметрі. |
| Texture (текстура) | Стандартне відхилення значень шкали сірого. Дозволяє оцінити неоднорідність внутрішньої структури ядра. |
| Perimeter (периметр) | Розмірна характеристика (довжина контуру). |
| Area (площа) | Розмірна характеристика (площа внутрішньої області). |
| Smoothness (гладкість) | Локальні варіації довжини радіуса (наскільки рівним є контур). |
| Compactness (компактність) | Похідна величина, що розраховується за формулою: $\frac{perimeter^2}{area} - 1$ |
| Concavity (увігнутість) | Виразеність увігнутих ділянок контуру (западин). |

Продовження таблиці 4.3

| Атрибут | Опис |
|--|---|
| Symmetry (симетрія) | Параметр симетрії форми ядра. |
| Fractal Dimension (фрактальна розмірність) | Оцінка складності «берегової лінії» периметра (наближення контуру). |

Для кожного з цих параметрів обчислюються середнє значення, стандартна помилка (SE) та «найгірше» (максимальне) значення, що загалом формує 30-вимірний простір ознак.

Аналіз вектору візуальних ознак (WBCD) виявив критичну для машинного навчання проблему – високу мультиколінеарність. Це явище, коли дві або більше незалежних змінних сильно корелюють між собою, створюючи інформаційний шум. Наприклад, геометрично очевидно, що зі збільшенням радіуса кола неминуче зростають його периметр і площа. Збереження всіх трьох параметрів у вхідному векторі нейромережі не додає нових знань про об'єкт, а лише збільшує обчислювальну складність і ризик перенавчання.

Для верифікації припущення про наявність мультиколінеарності та оцінки взаємозв'язків між ознаками виконано кореляційний аналіз, після чого застосовано процедури редукції (відбору) ознак для підвищення стійкості та швидкодії класифікаторів.

На рисунку 4.6 наведено приклад архітектури багат шарового перцептрона (MLP) у WEKA для скороченого набору ds3 (4 головні компоненти), що ілюструє перехід від первинних ознак до компактного представлення. Як видно, перцептрон містить чотири нейрони (pc1, pc2, pc3 та pc4, що відповідають кількості головних компонент) у вхідному шарі, три нейрони у прихованому шарі та два вихідні нейрони. Для алгоритмічного вирішення цього завдання було використано інструментарій середовища WEKA. Застосовано метод CorrelationAttributeEval у комбінації з алгоритмом пошуку Ranker. Суть методу полягає в оцінці прогностичної здатності кожного окремого

атрибута відносно цільового класу (діагнозу), при цьому штрафується надлишкова кореляція з вже відібраними атрибутами.

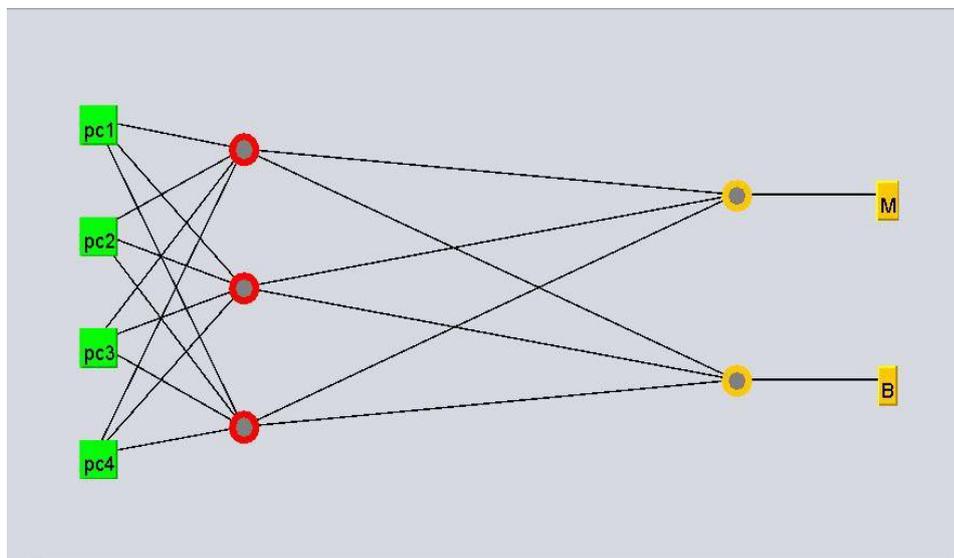


Рисунок 4.6 – Знімок екрана WEKA багатошарового перцептрона (MLP) для скороченого набору даних ds3

Для реалізації діагностичного контуру в середовищі WEKA налаштовано класифікатори MLP, Voted Perceptron та D14jMlp. Параметри їх конфігурації наведено в таблиці 4.4.

Таблиця 4.4 – Конфігурації класифікаторів

| Алгоритм | Конфігурація |
|------------------|--|
| MLP | <code>weka.classifiers.functions.MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a</code> |
| Voted Perceptron | <code>weka.classifiers.functions.VotedPerceptron -I 1 -E 1.0 -S 1 -M 10000</code> |
| D14jMlp | <code>weka.classifiers.functions.D14jMlpClassifier -S 1 -cache-mode MEMORY -early-stopping "weka.dl4j.earlystopping .Early Stopping -maxEpochsNoImprovement 0 -valPercentage 0.0" -normalization "Standardize training data" -iterator "weka.dl4j.iterators.instance.DefaultInstanceIterator -bs 1" -</code> |

Продовження таблиці 4.4

| Алгоритм | Конфігурація |
|--------------------------|---|
| Dl4jMlp (продовження) | <pre> iteration-listener "weka.dl4j.listener.EpochListener -eval true -n 5" -layer "weka.dl4j.layers.OutputLayer -lossFn \"weka.dl4j.lossfunctions.LossMCXENT\" -nOut 2 -activation \"weka.dl4j.activations.ActivationSoftmax \" -name \"Output layer\"" -logConfig "weka.core.LogConfiguration -append true -dl4jLogLevel WARN -logFile C:\\Users\\master\\wekafiles\\wekaDeeplearning4j.log - nd4jLogLevel INFO -wekaDl4jLogLevel INFO" -config "weka.dl4j.NeuralNetConfiguration -biasInit 0.0 -biasUpdater \"weka.dl4j.updater.Sgd -lr 0.001 -lrSchedule \\\"weka.dl4j.schedules.ConstantSchedule -scheduleType EPOCH\\\" \" -dist \"weka.dl4j.distribution.Disabled \" - dropout \"weka.dl4j.dropout.Disabled \" - gradientNormalization None -gradNormThreshold 1.0 -l1 NaN -l2 NaN -minimize -algorithm STOCHASTIC_GRADIENT_DESCENT-updater \"weka.dl4j.updater.Adam -beta1MeanDecay 0.9 - beta2VarDecay 0.999 -epsilon 1.0E-8 -lr 0.001 -lrSchedule \\\"weka.dl4j.schedules.ConstantSchedule -scheduleType EPOCH\\\" \" -weightInit XAVIER -weightNoise \"weka.dl4j.weightnoise.Disabled \" -numEpochs 10 - numGPUs 1 -averagingFrequency 10 -prefetchSize 24 - queueSize 0 -zooModel "weka.dl4j.zoo.CustomNet - channelsLast false -pretrained NONE\" \"weka.dl4j.weightnoise.Disabled \" -numEpochs 10 - numGPUs 1 -averagingFrequency 10 -prefetchSize 24 - queueSize 0 -zooModel "weka.dl4j.zoo.CustomNet - channelsLast false -pretrained NONE" </pre> |

Таким чином, розроблена підсистема попередньої обробки забезпечує адаптивну підготовку даних: для біохімічного потоку (CBCD) зберігається повний вектор через складність зв'язків, тоді як для морфологічного потоку (WBCD) виконується агресивна редукція розмірності для підвищення ефективності класифікаторів.

4.1.8 Реалізація периферійного вимірювального модуля (AD8232) та первинна обробка сигналу

Для практичного впровадження описаного методу підготовки інформативних ознак було спроектовано вимірювальний модуль на основі інтегрованої схеми AD8232. AD8232 є низькопотужною інтегрованою схемою, призначеною для реєстрації серцевого ритму та електричної активності серця у смузі 1–100 Гц, із двома вхідними каналами для зчитування сигналів від електродів. Реєстрований сигнал проходить через підсилювально-фільтрувальний тракт, що забезпечує придушення перешкод і підготовку даних до подальшої цифрової обробки. Додатково модуль дозволяє оцінювати якість контакту електродів за рахунок вимірювання імпедансу шкіри, а також має вбудований захист від електростатичного розряду та перенапруги, що підвищує стійкість роботи в умовах експлуатації. Цей компонент обрано як базовий фронтенд для збору біопотенціалів серця, оскільки він дозволяє ефективно реалізувати ланцюг підсилення та фільтрації, необхідний для подальшої математичної обробки. Апаратна конфігурація системи (рисунок 4.7) забезпечує консолідацію мультимодальних даних шляхом поєднання сигналів електричної активності серця з іншими фізіологічними показниками, що були визначені як релевантні в теоретичній частині дослідження. Як альтернативні периферійні засоби реєстрації фізіологічних параметрів у системах моніторингу можуть використовуватися й оптичні сенсори типу MAX30105, що розширює спектр придатних апаратних рішень [106].

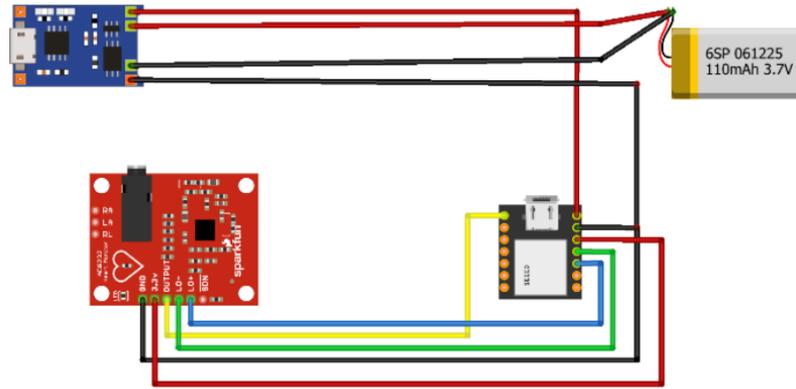


Рисунок 4.7 – Схема пристрою «ЕКГ-МОДУЛЬ AD8232»

Програмна реалізація методів обробки на цьому етапі фокусується на детекції та аналізі інтервалів PR, QT та комплексу QRS. Це дозволяє перетворити «сирий» аналоговий сигнал у структурований набір даних, придатний для інтелектуального аналізу, що відповідає стратегії проактивного моніторингу.

4.2 Програмно-апаратна реалізація методів стиснення зображень без втрат для медичної візуалізації

Сучасні системи моніторингу стану здоров'я людини, що базуються на концепції ІоМТ, генерують величезні масиви даних, серед яких найбільш об'ємними є результати візуальної діагностики (мамографія, термографія, УЗД). Характерною особливістю графічних медичних даних є їхня висока надлишковість, яка безпосередньо впливає на швидкість передачі сигналів від сенсорних вузлів до центрів прийняття рішень та на вартість зберігання інформації [93, 105].

У контексті систем реального часу затримка в передачі даних може бути критичною для своєчасного реагування медичного персоналу. У медицині пріоритетним є стиснення без втрат, оскільки будь-які спотворення сигналу, характерні для методів із втратами (наприклад, стандартного JPEG), можуть призвести до некоректного аналізу дрібних патологічних структур і діагностичних помилок [93, 107].

В рамках дослідження проаналізовано кілька сімейств алгоритмів, придатних для реалізації в ІоМТ контурі: словникові методи сімейства LZ (LZ77

та LZ78), методи RLE, диференціальне кодування, а також JPEG-LS [105] як спеціалізований стандарт безвтратного стиснення. Для структурованих потоків встановлено перевагу LZ78 над LZ77 (орієнтовно у 1,5 раза за коефіцієнтом стиснення), а для медичних зображень як базовий безвтратний варіант обрано JPEG-LS через баланс ступеня стиснення та обчислювальної складності [107].

Таким чином, це дає можливість зменшити навантаження на мережу під час пересилання даних на сервер шляхом зменшення об'єму даних після стиснення, при цьому не втрачаючи важливих елементів даних, які можуть мати високу діагностичну важливість.

На стороні сервера виконується декомпресія даних, після чого вони надходять до БД (MySQL) та імпортуються в середовище WEKA для подальшого аналізу. Процес включає: трансформацію даних і формування ARFF-набору; запуск алгоритмів класифікації (RandomForest або J48) чи інших задач аналізу. Таким чином, стиснення розглядається як частина наскрізного контуру «архівування – декомпресія – аналітика», а не як ізольована процедура.

4.3 Оцінка метрологічних та експлуатаційних характеристик системи на реальних наборах даних

Для підтвердження надійності розробленої системи моніторингу було проведено емпіричну оцінку якості вимірювань датчика серцевого ритму AD8232. Основною метою цього етапу було визначення точності реєстрації електричної активності серця в умовах, наближених до реальної експлуатації.

Далі наведені три емпіричні моделі для датчика AD8232, розраховані на основі методу апроксимації функції. Таким чином, за допомогою датчика по п'ять разів було зроблене вимірювання серцевого ритму у десятих пацієнтів. Отримана множина даних стала експериментальною вибіркою для побудови моделі [104].

На основі цих даних було вираховане математичне очікування Y_m та стандартна Δ і відносна σ похибки для оцінки відхилення фактичних вимірів від очікуваних (таблиця 4.5).

Отримані точки даних були нанесені на графік з метою визначення, який тип функції описуватиме закономірність найкраще.

Таблиця 4.5 – Експериментальна вибірка для побудови моделі

| X | Y_1 | Y_2 | Y_3 | Y_4 | Y_5 | Y_m | σ | Δ |
|----|-------|-------|-------|-------|-------|-------|----------|----------|
| 1 | 112 | 116 | 98 | 106 | 71 | 100,6 | 11,92197 | 6,44 |
| 2 | 77 | 61 | 65 | 90 | 90 | 76,6 | 9,05048 | 5,44 |
| 3 | 75 | 84 | 96 | 117 | 79 | 90,2 | 11,28815 | 6,52 |
| 4 | 100 | 75 | 100 | 80 | 105 | 92,0 | 9,00617 | 5,8 |
| 5 | 91 | 94 | 80 | 74 | 108 | 89,4 | 8,78888 | 4,96 |
| 6 | 119 | 87 | 91 | 98 | 114 | 101,8 | 9,39749 | 5,88 |
| 7 | 109 | 113 | 100 | 80 | 116 | 103,6 | 9,66782 | 5,44 |
| 8 | 109 | 68 | 86 | 116 | 118 | 99,4 | 14,44991 | 8,96 |
| 9 | 107 | 81 | 110 | 92 | 78 | 93,6 | 9,73653 | 5,96 |
| 10 | 66 | 92 | 106 | 87 | 77 | 85,6 | 10,09510 | 5,64 |

Спираючись на розміщення точок даних на графіку, було визначено, що найкраще їх описуватимуть ступенева функція або поліноми різних ступенів (рисунок 4.8).

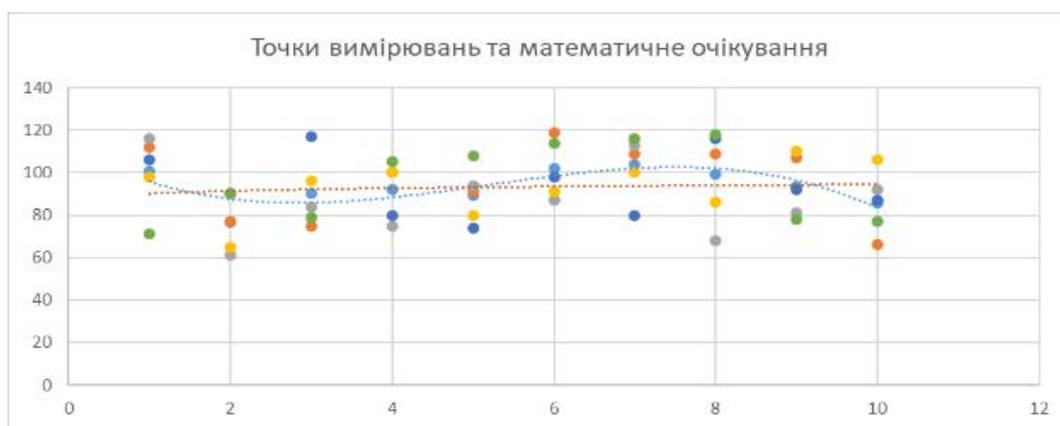


Рисунок 4.8 – Точки експериментальних даних на графіку

Таким чином, були обрані такі моделі, як поліноми третього та п'ятого ступенів та ступенева функція, також були визначені числові коефіцієнти для цих функцій. Отримані евристичні моделі для цих даних мають такий вигляд:

$$y_1 = -0,3299x^3 + 5,1269x^2 - 21,155x + 112,05; \quad (4.1)$$

$$y_2 = -0,015x^5 + 0,4719x^4 - 5,7946x^3 + 33,331x^2 - 82,891x + 154,08; \quad (4.2)$$

$$y_3 = 90,098x^{0,0205}. \quad (4.3)$$

Таблиця 4.6 – Отримані очікувані значення за трьома емпіричними моделями

| <i>X</i> | <i>Y_m</i> | <i>Ye₁</i> | <i>Ye₂</i> | <i>Ye₃</i> |
|-----------|----------------------|-----------------------|-----------------------|-----------------------|
| 1 | 100,6 | 95,6920 | 99,1823 | 90,09800 |
| 2 | 76,6 | 87,6084 | 82,3356 | 91,38739 |
| 3 | 90,2 | 85,8198 | 83,5107 | 92,15017 |
| 4 | 92,0 | 88,3468 | 90,4040 | 92,69523 |
| 5 | 89,4 | 93,2100 | 96,6375 | 93,12023 |
| 6 | 101,8 | 98,4300 | 99,9588 | 93,46893 |
| 7 | 103,6 | 102,0274 | 100,4411 | 93,76476 |
| 8 | 99,4 | 102,0228 | 98,6832 | 94,02179 |
| 9 | 93,6 | 96,4368 | 94,0095 | 94,24908 |
| 10 | 85,6 | 83,2900 | 82,6700 | 94,45287 |

Також, були визначені стандартна Δ і відносна ϵ похибки для трьох емпіричних моделей (таблиця 4.7).

Таблиця 4.7 – Стандартні та відносні похибки для трьох моделей

| <i>X</i> | Δ_1 | Δ_2 | Δ_3 | ϵ_1 | ϵ_2 | ϵ_3 |
|----------|------------|------------|------------|--------------|--------------|--------------|
| 1 | 4,9080 | 1,4177 | 10,5020 | 0,048105 | 0,014115 | 0,11119 |
| 2 | -11,0084 | -5,7356 | -14,7874 | -0,107900 | -0,057100 | -0,15656 |
| 3 | 4,3802 | 6,6893 | -1,95017 | 0,042932 | 0,066599 | -0,02065 |
| 4 | 3,6532 | 1,5960 | -0,69523 | 0,035806 | 0,015890 | -0,00736 |
| 5 | -3,8100 | -7,2375 | -3,72023 | -0,037340 | -0,072060 | -0,03939 |
| 6 | 3,3700 | 1,8412 | 8,33107 | 0,033030 | 0,018331 | 0,08820 |

Продовження таблиці 4.7

| X | Δ_1 | Δ_2 | Δ_3 | ϵ_1 | ϵ_2 | ϵ_3 |
|----|------------|------------|------------|--------------|--------------|--------------|
| 7 | 1,5726 | 3,1589 | 9,835237 | 0,015414 | 0,031450 | 0,10413 |
| 8 | -2,6228 | 0,7168 | 5,378214 | -0,025710 | 0,007137 | 0,05694 |
| 9 | -2,8368 | -0,4095 | -0,649080 | -0,027800 | -0,004080 | -0,00687 |
| 10 | 2,3100 | 2,9300 | -8,852870 | 0,022641 | 0,029171 | -0,09373 |

Таким чином, найкращі показники абсолютних і відносних похибок показує друга емпірична модель (поліном 5-го ступеня; рисунок 4.9), але цього ще замало. Для остаточної оцінки адекватності отриманих моделей була визначена ступінь їхньої придатності [104].

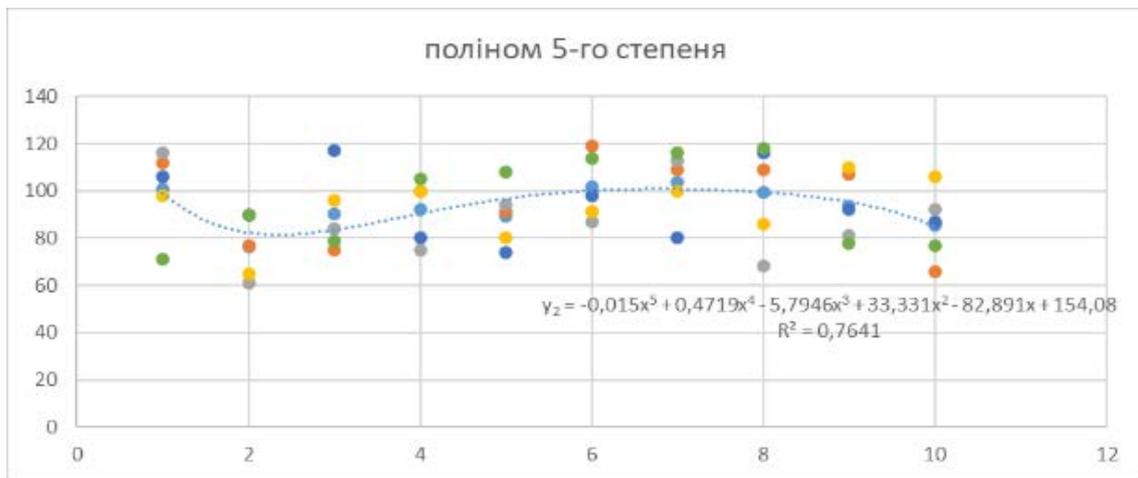


Рисунок 4.9 – Друга модель – поліном 5-го ступеня

Адекватність (придатність) емпіричної моделі була визначена з допомогою нерівності:

$$\|\epsilon_s(i)\| < \frac{3K|y_i|_{max}}{100|F_s|_{max}} - \frac{\|F_s[\Phi(i)]\| - \|y(i)\|}{|F_s|_{max}} \quad (4.4)$$

Таким чином, якщо всі значення емпіричної моделі відповідають цьому критерію придатності, то її можна вважати адекватною.

Таблиця 4.8 – Ступінь придатності трьох емпіричних функцій

| X | ε_1 | E_1 | ε_2 | E_2 | ε_3 | E_3 |
|-----|-----------------|-----------------|-----------------|-----------------|-----------------|---------------|
| 1 | 0,048105 | 0,127307 | 0,048105 | 0,094568 | 0,048105 | 0,196742 |
| 2 | -0,1079 | -0,02869 | -0,1079 | 0,023349 | -0,1079 | -0,071 |
| 3 | 0,042932 | 0,122134 | 0,042932 | 0,147052 | 0,042932 | 0,064907 |
| 4 | 0,035806 | 0,115008 | 0,035806 | 0,096343 | 0,035806 | 0,078193 |
| 5 | -0,03734 | 0,041859 | -0,03734 | 0,008396 | -0,03734 | 0,046167 |
| 6 | 0,03303 | 0,112233 | 0,03303 | 0,098784 | 0,03303 | 0,173757 |
| 7 | 0,015414 | 0,094616 | 0,015414 | 0,111903 | 0,015414 | 0,189682 |
| 8 | -0,02571 | 0,053495 | -0,02571 | 0,08759 | -0,02571 | 0,142495 |
| 9 | -0,0278 | 0,051398 | -0,0278 | 0,076376 | -0,0278 | 0,078682 |
| 10 | 0,022641 | 0,101843 | 0,022641 | 0,109624 | 0,022641 | -0,00817 |

Тепер добре видно, що по критерію адекватності перша модель (поліном 3-го ступеня; рисунок 4.10) є найкращою, бо в неї критерію адекватності не відповідає лише одне значення з десяти, а в двох інших – по два.

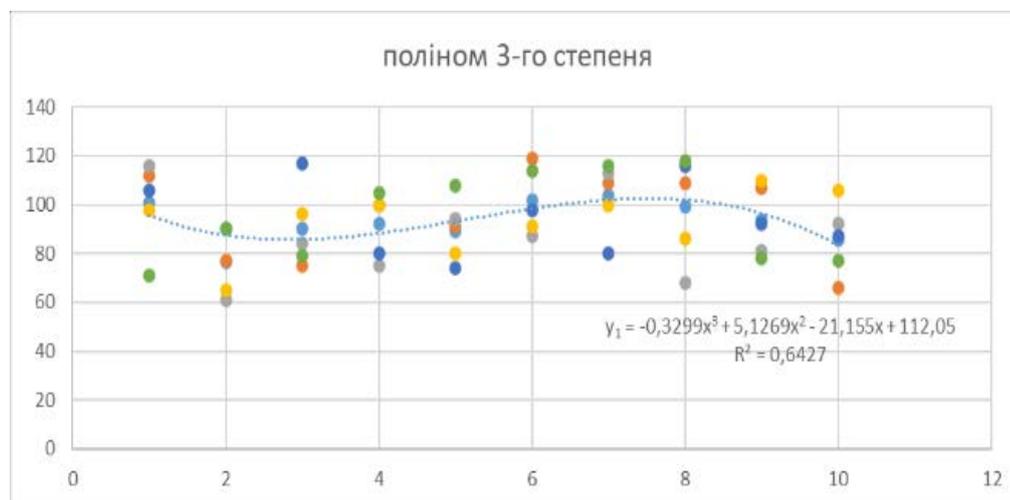


Рисунок 4.10 – Перша модель – поліном 3-го ступеня

Отже, по критерію адекватності перша модель поліном 3-го ступеня є найкращою, а найкращі показники абсолютних і відносних похибок показує друга емпірична модель поліном 5-го ступеня.

Окрім того, статистична обробка даних вимірювального каналу використана для оцінювання стійкості сигналу до артефактів руху та електростатичних розрядів, захист від яких інтегровано в архітектуру мікросхеми. Запропонований метод оцінювання передбачає такі кроки:

1) аналіз стабільності сигналу: шляхом статистичної обробки вибірок ЕКГ-сигналів було встановлено, що використання вбудованого підсилювача та фільтрів AD8232 (рисунок 4.11) забезпечує високу амплітудну точність, необхідну для подальшої цифрової обробки мікроконтролером [104, 106];

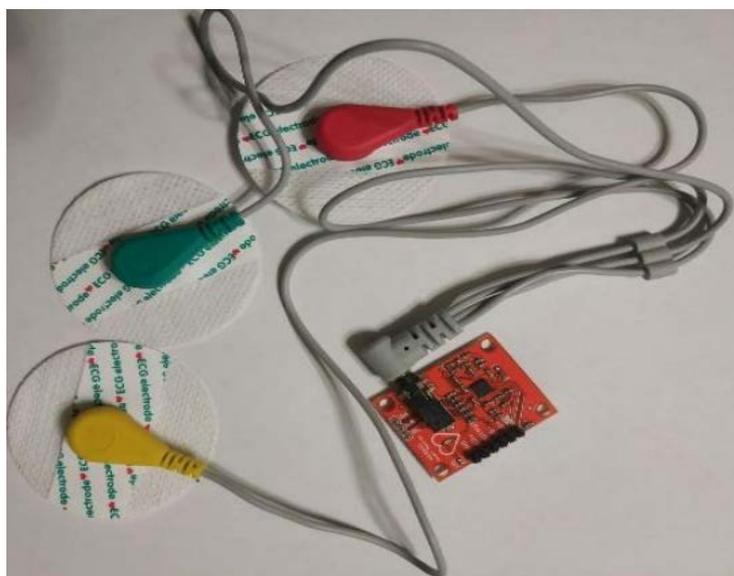


Рисунок 4.11 – Модуль серцевого ритму AD8232

2) верифікація часових параметрів: на основі отриманих відліків було проведено розрахунок інтервалів, критичних для діагностики патологій, що виникають під час терапії раку молочної залози:

- інтервал PR: тривалість у діапазоні 0,12–0,20 с підтвердила коректність вилучення сигналу з електродів;
- комплекс QRS: стабільна детекція зубців R дозволила забезпечити високу точність вимірювання частоти серцевих скорочень (ЧСС);
- інтервал QT: статистична оцінка нормальних значень у межах 0,36–0,44 с за умови серцевого ритму 60–100 ударів на хвилину підтвердила можливість використання системи для моніторингу кардіотоксичності.

4.3.1 Метрологічна оцінка точності та діагностичної потужності моделей

Експериментальна перевірка далі виконується для вузлів системи, яка спрямована на оцінку їх метрологічних та експлуатаційних характеристик. Для оцінки метрологічної спроможності системи було проведено серію експериментів з використанням алгоритмів машинного навчання у середовищі WEKA версії 3.9.6.

До експлуатаційних характеристик таких вузлів доцільно також віднести моніторинг енергоспоживання, особливо для периферійних модулів тривалого телемедичного спостереження.

По-перше, було виконано аналіз класифікації на повному наборі даних вісконсинського набору даних (WBCD). Таким чином, застосування класифікаторів SVM та LMT на повному наборі WBCD (30 атрибутів) продемонструвало високу точність: показники Precision та Recall сягнули 0,98, а коефіцієнт кореляції Метьюса (MCC) – 0,95–0,96. Використання глибокого навчання (Dl4jMlp) дозволило досягти точності в діапазоні 95,9–97,7 % зі стандартним відхиленням менше 2,5 %. [99]

Таблиця 4.9 – Середньозважені оцінки продуктивності для двох різних класифікаторів повного набору даних

| Класифікатор | Точність | Повність | Каппа-статистика | MCC | ROC |
|---------------------|-----------------|-----------------|-------------------------|------------|------------|
| SVM | 0,98 | 0,98 | 0,95 | 0,96 | 0,97 |
| LMT | 0,98 | 0,98 | 0,95 | 0,95 | 1,00 |

Далі була виконана оцінка на базі біомаркерів з набору даних Коїмбри (CBCD). Набір CBCD виявився більш складним для аналізу через високий рівень статистичного шуму. Оцінка точності (Accuracy) для CBCD варіюється в межах 72–90 %. Зокрема, було встановлено, що рівень глюкози є найбільш релевантним

метрологічним показником (перший ранг), що прямо корелює з ризиком патології.

На останньому етапі була виконана валідація через крос-валідацію. Для виключення перенавчання використовувалася 10-кратна крос-валідація. Було зафіксовано, що за умови переходу від навчання на всьому наборі до крос-валідації кількість хибнопозитивних результатів (FP) у наборі WBCD зростає з 0 до 12, що свідчить про чутливість системи до малих вибірок.

4.3.2 Вплив зниження вимірності на експлуатаційну ефективність

Мінімізація кількості ознак є критичною для швидкодії ІоМТ-пристроїв. Тому для прискорення обробки результатів була виконана спроба зменшення розмірності контрольних наборів даних [95, 100].

Для зменшення розмірності набору WBCD була виконана його потрійна фільтрація. Таким чином, завдяки використанню фільтрів CfsSubsetEval та методу головних компонент (PCA), розмірність набору WBCD було скорочено з 30 атрибутів до чотирьох ключових компонентів, що зберігають 91 % дисперсії. Це не призвело до статистично значущого падіння точності, що підтверджує можливість використання спрощених моделей у носимих пристроях [99].

Для формування скороченого представлення ds2 на морфологічному потоці WBCD виконано відбір ознак в середовищі WEKA із використанням оцінювача CorrelationAttributeEval у поєднанні з алгоритмом пошуку Ranker. Процедура передбачала оцінювання прогностичної здатності кожного атрибута щодо цільового класу та штрафування надлишкової кореляції між відібраними ознаками, що зменшує мультиколінеарність. Додатково відсіяно атрибути зі слабким зв'язком із діагнозом (умовно $< 0,2$) та ознаки-дублікати, які повторюють одну й ту саму інформацію (наприклад, `area_mean` і `perimeter_mean`).

У результаті експерименту було сформовано скорочений, але інформаційно насичений набір ознак. Атрибути з низькою кореляцією до діагнозу ($< 0,2$), а також ті, що дублювали інформацію (наприклад, `area_mean`, `perimeter_mean`), були відсіяні.

Також, було виконане ранжування в CBCD. За допомогою алгоритмів InfoGain та RELIEF було виділено підмножину з 4 ознак (Glucose, Age, HOMA, Resistin). Використання лише цих параметрів дозволяє системі працювати швидше, зберігаючи прийнятний рівень Карра-статистики (близько 0,65).

Для перевірки впливу ступеня редукції ознак на якість класифікації сформовано три представлення набору даних: ds1 (повний набір ознак), ds2 (після відбору ознак), ds3 (після додаткової редукції до мінімального представлення). Оцінювання виконано в середовищі WEKA за схемою 10-fold cross-validation із 10 повтореннями, рівнем довіри 0,95 та статистичною перевіркою відмінностей (скоригований парний t-тест). Отримані результати (таблиця 4.10) демонструють, що навіть для найбільш редукovanого набору ds3 точність залишається високою, а значущих відмінностей між ds1–ds3 за показником Accuracy не спостерігається, що підтверджує практичну доцільність використання скорочених представлень у периферійних IoT-пристроях.

Таблиця 4.10 – Відсоток правильно класифікованих екземплярів Accuracy

| Набір даних | D14jMlp | MLP | Voted Perceptron |
|-------------|--------------|--------------|------------------|
| ds1 | 97,68 (1,80) | 96,72 (2,20) | 96,45 (2,11) |
| ds2 | 97,58 (1,79) | 96,64 (1,90) | 96,78 (2,14) |
| ds3 | 96,40 (2,26) | 95,87 (2,47) | 95,96 (2,33) |

4.3.3 Аналіз стійкості до зашумлених даних та викидів

Метрологічна надійність системи залежить від її здатності ігнорувати артефакти сенсорних мереж і статистично нетипові спостереження. Таким чином, була виконана низка перевірок обраних наборів даних, в тому числі виявлення викидів та їхнього впливу на достовірність рішень.

За допомогою фільтра Interquartile Range у наборі WBCD було виявлено 55 випадків (близько 10 %), які класифікуються як викиди; з них 17 належать класу Benign і 38 – класу Malignant [97].

З іншої сторони, використання Local Outlier Factor з пороговим значенням 1,3 дозволило чітко розмежувати «ядро» даних та аномальні відхилення, що підвищує стабільність діагностичних правил. У роботі [96] підкреслено, що частка таких спостережень є суттєвою і не може розглядатися як незначний фактор, оскільки вона потенційно впливає на структуру помилок класифікації. Таким чином, було встановлено, що у наборі CBCD рівень шуму також є високим: IQR-фільтр ідентифікує 12 «аномальних» інстансів із 116 (приблизно 10 % у середньому), причому для класу «patients» шум досягає близько 14 % [96].

Для малих вибірок (зокрема CBCD) навіть поодинокі «некоректні» записи та шумові значення можуть істотно збільшувати позадіагональні елементи матриці невідповідностей, тобто прямо формувати частки FP/FN. Додатково, застосування CAIRAD дозволяє не лише позначати «noisy data» (шуми) як пропуски, але й виявляти сутності, що ведуть до систематичних помилок класифікації (зокрема, виявлено 4 потенційно проблемні сутності). Оцінювання частки «шумових значень» за атрибутами (CAIRAD) показує, що зашумлення є нерівномірним: наприклад, для атрибуту Age досягає 14 %, для атрибуту Adiponectin – 16 %, для Insulin – 12 %. Це обґрунтовує необхідність окремого контролю якості даних і використання робастних підходів до моделей.

У випадку CBCD показано наявність «довгого хвоста» розподілу LOF, який зливається з «ядром» приблизно на рівні $LOF = 1,3$. За обраним порогом оцінено, що «core» складає 95 сутностей, тоді як «tail» (хвіст) містить 21 викид, у роботі [96] автори пов'язують викиди з позадіагональними елементами матриці помилок і зазначають, що прогнозна точність для такого розділення може бути порядку 82 %, що є прийнятним для попереднього скринінгу, але недостатнім для відповідальних клінічних рішень без додаткової валідації.

4.4 Експериментальна перевірка багаторівневої системи моніторингу

Експериментальну перевірку працездатності та продуктивності багаторівневої системи моніторингу проведено в умовах імітації інтенсивного користувацького навантаження. Основною метою експерименту було

порівняння монолітної та багаторівневої реалізацій системи за кількістю і часткою помилкових запитів в умовах однакового навантаження, оскільки саме цей показник найбільш повно відображає стійкість системи до перевантаження. Серверні компоненти системи були розгорнуті в контейнеризованому середовищі Docker, що забезпечило ізоляцію сервісів, відтворюваність умов експерименту та контроль використання обчислювальних ресурсів. Для серверного середовища було виділено 8 ГБ оперативної пам'яті.

Генерування навантаження здійснювалося з окремого клієнтського пристрою – ноутбука з процесором AMD Ryzen 5 5500U та 16 ГБ оперативної пам'яті. Для надсилання запитів використовувався інструмент Locust, за допомогою якого моделювалася одночасна робота великої кількості користувачів. У межах експерименту було задано навантаження 1000 віртуальних користувачів із темпом нарощування 10 користувачів за секунду, а тривалість одного тестового запуску становила 5 хвилин. Відповідні параметри навантажувального тестування наведено на рисунку 4.12.

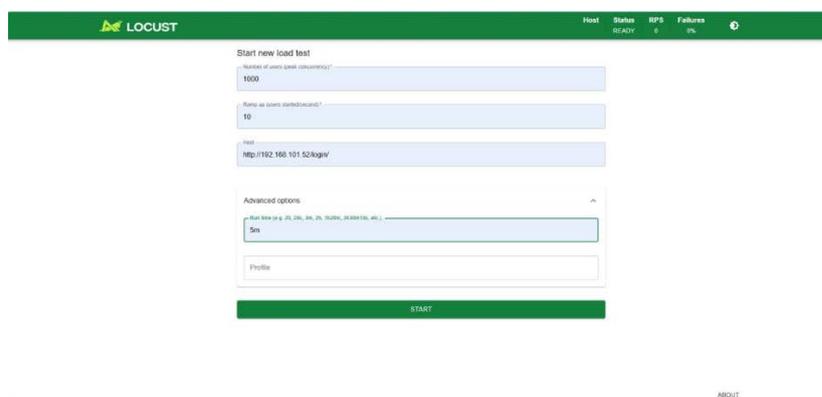


Рисунок 4.12 – Налаштування навантажувального тестування в середовищі Locust для експериментальної перевірки системи моніторингу

Під час експериментальної перевірки оцінювалися такі показники, як кількість запитів за секунду, час відповіді системи та частка запитів, що завершилися з помилкою. Для обґрунтування доцільності запропонованої архітектури було проведено порівняння результатів навантажувального тестування монолітної та багаторівневої реалізацій системи.

Результати тестування монолітної реалізації наведено на рисунку 4.13. Із наведених графіків видно, що після виходу системи на робочий режим кількість оброблених запитів стабілізувалася на рівні близько 296,9 запитів/с, водночас частка помилкових запитів досягала 39 %. Це свідчить про суттєве зниження надійності монолітної реалізації в умовах високого навантаження. Крім того, графік часу відповіді демонструє зростання затримок для найповільніших запитів, що вказує на погіршення стабільності обслуговування при збільшенні кількості одночасних користувачів.



Рисунок 4.13 – Результати навантажувального тестування монолітної реалізації системи моніторингу в середовищі Locust

Результати тестування багаторівневої реалізації наведено на рисунок 4.14. У цьому випадку система забезпечила вищу пропускну здатність близько 316,2 запитів/с, а частка помилкових запитів зменшилася до 18 %. Отже, багаторівнева архітектура продемонструвала кращу стійкість до інтенсивного потоку звернень та вищу ефективність обробки запитів порівняно з монолітним підходом. Незважаючи на наявність зростання часу відповіді для частини запитів при досягненні пікового навантаження, загальна робота багаторівневої системи виявилася більш стабільною.



Рисунок 4.14 – Результати навантажувального тестування багаторівневої реалізації системи моніторингу в середовищі Locust

Із наведених графіків видно, що після виходу системи на робочий режим кількість оброблених запитів стабілізувалася на рівні близько 296,9 запитів/с, водночас розподілена багаторівнева система стабілізувалась на рівні 316,2 запитів/с. Ці дані свідчать про більшу на 8 % пропускну здатність багаторівневої архітектури.

Порівняльний аналіз отриманих результатів показав, що використання багаторівневої архітектури є доцільним для систем дистанційного моніторингу. У порівнянні з монолітною реалізацією вона забезпечує вищу пропускну здатність і меншу частку помилок при однакових умовах тестування. Це підтверджує практичну придатність запропонованого підходу для побудови масштабованих систем медичного моніторингу, орієнтованих на роботу в умовах значного користувацького навантаження.

Висновки до розділу 4

У четвертому розділі показано, що запропоновані в роботі підходи не залишилися лише на рівні теоретичних моделей, а були доведені до практичної реалізації та перевірені в реальних умовах. Результати дослідження впроваджено в науково-дослідні роботи та в навчальний процес, а також реалізовано програмно-апаратний контур системи моніторингу, який об'єднує рівень збору даних, серверну обробку, базу даних MySQL і аналітичний модуль WEKA. У

розділі обґрунтовано, що для медичних зображень доцільно використовувати безвратне стиснення, зокрема JPEG-LS, а для біомедичних даних – попередню обробку, уніфікацію та подальший інтелектуальний аналіз у межах єдиного контуру. Окремо підтверджено працездатність периферійного вимірювального модуля на основі AD8232, який придатний для реєстрації серцевого ритму та виділення діагностично важливих часових параметрів сигналу. Експериментальна перевірка показала, що навіть після скорочення розмірності даних точність класифікації залишається високою, а використання скорочених наборів ознак є практично доцільним для ІоМТ-систем, де важливі швидкодія та економія ресурсів. Також встановлено, що якість діагностичних рішень суттєво залежить від рівня шуму, викидів і повноти вхідних даних, тому контроль якості даних є обов'язковою умовою надійної роботи системи.

ВИСНОВКИ

Дисертаційна робота на тему «Моделі та методи побудови спеціалізованих систем моніторингу стану здоров'я людини» присвячена розв'язанню актуальної науково-прикладної задачі підвищення ефективності спеціалізованих систем моніторингу стану здоров'я людини шляхом удосконалення методів формування інформативного простору ознак, побудови інтерпретованих діагностичних моделей та оптимізації обробки мультимодальних медичних даних у програмно-апаратних контурах. У межах дослідження сформовано підхід, який поєднує інтелектуальний аналіз даних у середовищі WEKA, механізми редукції вимірності, валідацію якості моделей та інженерні рішення для інтеграції даних у реальні інформаційні потоки.

Отримані результати підтверджують, що цілеспрямоване «стиснення» інформативного простору (як у сенсі відбору редукції ознак для класифікації, так і у сенсі безвтратного стиснення медичних зображень) є практично значущим інструментом, який дозволяє зменшувати обчислювальні витрати й вимоги до передавання даних без втрати діагностично важливої інформації.

Поставлену мету дисертаційного дослідження, а саме: вдосконалення спеціалізованих комп'ютерних систем моніторингу стану здоров'я людини шляхом розроблення та впровадження методів відбору й ранжування інформативних ознак та зменшення розмірності вхідних даних було досягнуто.

За результатами проведеного дослідження отримано такі основні наукові результати:

1) розроблено модель спеціалізованої системи моніторингу стану здоров'я людини з контуром «edge – сервер – WEKA», яка інтегрує зменшення розмірності (редукцію ознак 30 до 4 та безвтратне стиснення зображень) із уніфікованим ORM – ARFF-імпортом і подальшим навчанням моделей у WEKA; це забезпечило надійність класифікації за PRC area у діапазоні 0,96–1,00 та скорочення обсягу передавання зображень із коефіцієнтом стиснення $k = 1,97$;

2) розроблено метод структурного стиснення табличних біомедичних даних шляхом потрійної послідовної фільтрації та редукції ознак у WEKA;

показано, що для WBCD редукція з 30 до 4 не призводить до істотного погіршення якості: відсоток правильно класифікованих екземплярів становить 95,9–97,7 % при стандартному відхиленні $< 2,5$ %, а PRC area перебуває в межах 0,96–1,00;

3) розроблено метод побудови «легковагових» інтерпретованих діагностичних моделей для первинного скринінгу на основі рутинних аналізів крові набору даних про рак молочної залози Коїмбри із пріоритезацією біомаркерів (глюкоза, вік – ранг 1–2) та індукцією правил/дерев (JRip, J48) на редукованому просторі 4–6 атрибутів; досягнуто точність класифікації 68,97 %, повнота для класу пацієнтів $\text{Recall}(\text{patients}) = 0,703$ і $\text{F1}(\text{patients}) = 0,714$ для кращого базового класифікатора (J48) і сформовано прозорі правила типу «IF – THEN»;

4) розроблено метод валідації внутрішньої структури даних через зіставлення результатів некерованого групування з еталонними класами за метрикою FMI; показано, що для SOM (самоорганізовані карти) і LVQ (квантування векторів з навчанням) на наборах ds1–ds3 значення FMI перебувають в межах 0,81–0,99, що підтверджує узгодженість кластерної структури з класами після редукції ознак;

5) обґрунтовано та впроваджено безвтратне стиснення медичних зображень як критично важливий етап збереження діагностичних деталей у телемедичних системах; експериментально підтверджено доцільність JPEG-LS як оптимального компромісу між коефіцієнтом стиснення та обчислювальною складністю ($\text{CR} = 1,97$), а також показано збереження високої якості подальшого аналізу в WEKA (коефіцієнт кореляції 0,9159 на повному наборі атрибутів);

6) розроблено та апробовано практичні механізми інтеграції даних у контур інтелектуального аналізу «БД (MySQL) – ORM – ARFF – WEKA»: забезпечено автоматизоване формування наборів в формат ARFF із БД, контроль коректності структури атрибутів і записів та відтворюваний сценарій оцінювання (10-кратна крос-валідація) для трьох подань даних ds1–ds3;

7) спроектовано вимірювальний модуль реєстрації ЕКГ на базі інтегрованої схеми AD8232 і виконано експериментальну оцінку похибок вимірювання ЧСС; за даними таблиці 4.8 середня абсолютна похибка кращої емпіричної моделі становить 3,17 уд./хв (максимальна 7,24 уд./хв), а середня відносна – 3,16 % (максимальна 7,21 %). Побудовано емпіричні моделі апроксимації, що можуть використовуватися для попереднього калібрування та підвищення надійності вимірювань у системі моніторингу.

Кількісні показники, що характеризують отримані результати, узагальнено нижче:

- вісконсинський набір даних про рак молочної залози (повний ds1 та скорочені до одинадцяти атрибутів ds2 і до чотирьох атрибутів ds3 набори даних): після редукції з 30 до 4, точність залишилась високою: для класифікатора D14jMPr 97,68 % ($\sigma = 1,80$) на ds1 і 96,40 % ($\sigma = 2,26$) на ds3; для класифікатора MLP 96,72 % ($\sigma = 2,20$) і 95,87 % ($\sigma = 2,47$) відповідно; область під PR-кривою становить 0,96–1,00;

- вісконсинський набір даних про рак молочної залози (повний набір, 30 ознак): класифікатори SVM (метод опорних векторів) та LMT (логістичне модельне дерево) забезпечують точність Precision = 0,98, повноту Recall = 0,98, коефіцієнт кореляції Метьюса MCC = 0,95–0,96 та площу під ROC-кривою ROC-area = 0,97–1,00;

- набір даних про рак молочної залози Коїмбри: найкращий базовий класифікатор J48 – точність Precision = 68,97 %, повнота для класу пацієнтів Recall(patients) = 0,703 та F1(patients) = 0,714 (діапазон точності для розглянутих моделей 60,34–68,97 %);

- валідація кластерної структури (SOM – самоорганізовані карти; та LVQ – квантування векторів з навчанням) індекс Фоулкса-Меллоу (FMI) перебуває в межах 0,81–0,99 для ds1–ds3.

Безвтратне стиснення зображень: JPEG-LS дає коефіцієнт стиснення $k = 1,97$ (порівняно з JPEG2000 коефіцієнт стиснення $k = 1,69$); після імпорту в WEKA збережено високу узгодженість аналізу (коефіцієнт кореляції 0,9159).

– стійкість до шуму та викидів: у «WBCD» виявлено 55 викидів (10 %); у наборі даних про рак молочної залози Коїмбри – 12/116 (10 %), при цьому для класу пацієнти частка шумових значень сягає 14 %; за LOF (локальний фактор випадіння) порогом 1,3 оцінено «ядро» core = 95 і «хвіст» tail = 21 екземпляру.

Таким чином, у дисертації сформовано та апробовано комплексний підхід до побудови спеціалізованих систем моніторингу стану здоров'я, який поєднує:

– методи формування інформативного простору та інтерпретовані моделі для первинної діагностики,

– механізми перевірки надійності результатів через валідаційні метрики та зіставлення кластерів із класами;

– інженерні рішення інтеграції даних (ARFF/HTTP) та зменшення обсягів передавання (JPEG-LS, FPGA) у телемедичних контурах. Отримані результати можуть бути використані при створенні і модернізації ІоМТ-систем, телемедичних сервісів і програмно-апаратних комплексів для попереднього скринінгу, дистанційного спостереження та аналітичної підтримки прийняття рішень у медицині.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Hauschild, A. C., Martin, R., Holst, S. C., Wienbeck, J., Heider, D. Guideline for software life cycle in health informatics. *iScience*. 2022. Iss. 9. P. 105534. DOI: 10.1016/j.isci.2022.105534.
2. Tummers J., Tobi H., Catal C., Tekinerdogan B. Designing a reference architecture for health information systems. *BMC Med Inform Decis Mak*. 2021. Iss. 8. P. 210. DOI: 10.1186/s12911-021-01570-2.
3. Casanova R., Villa-Garzon F. A., Branch-Bedoya J. W. Architectural patterns for health information systems: a systematic review. *Front Digit Health*. 2025. Iss. 17. P. 1694839. DOI: 10.3389/fdgth.2025.1694839.
4. Boikanyo K., Zungeru A. M., Sigweni B., Yahya A., Lebekwe C. Remote patient monitoring systems: applications, architecture, and challenges. *Scientific African*. 2023. Vol. 20. P. e01638. DOI: 10.1016/j.sciaf.2023.e01638.
5. Trunov A., Beglytsia V., Gryshchenko G., Ziuzin V., Koshovyi V. Methods and tools of formation of general indexes for automation of devices in rehabilitative medicine for post-stroke patients. *Eastern-European Journal of Enterprise Technologies*. 2021. Vol. 4, iss. 2 (112). P. 35–46. DOI: 10.15587/1729-4061.2021.239288.
6. Ovchar S., Mohyla V., Saltovskiy B., Puzyrov S. Development of an IoT device based on a Geiger counter. *Automation of Technological and Business Processes*. 2023. Vol. 15, iss. 3. P. 51–57. DOI: 10.15673/atbp.v15i3.2625.
7. Tiahunova M., Tronkina O., Kirichek G., Skrupsky S. The neural network for emotions recognition under special conditions. *CEUR Workshop Proceedings*. 2021. Vol. 2864. P. 121–134. DOI: 10.32782/cm15/2864-11.
8. Карнаух Д., Тягунова М., Киричек Г. Методи та можливості покращення ефективності бездротових мереж передачі даних для систем «ІоТ». *Вчені записки ТНУ імені В. І. Вернадського. Серія: Технічні науки*. 2025. Т. 36 (75), № 2. С. 84–90. DOI <https://doi.org/10.32782/2663-5941/2025.2.1/13>.
9. Obukhova K., Savchuk T., Zhuravska I., Boiko A., Nikolskyi V., Puzyrov S. Prevention of unmanned vessels collisions due to pre-modeling the remote control

center CPU load. *CSIT Proceedings: 16th International Conference on Computer Science and Information Technologies*. Int. conf., Lviv, Ukraine, Sep. 22–25, 2021. Vol. 2. P. 202–205. DOI: 10.1109/CSIT52700.2021.9648800.

10. Custodio V., Herrera F.J., López G., Moreno J.I. A Review on Architectures and Communications Technologies for Wearable Health-Monitoring Systems. *Sensors*. 2012. Vol. 12, iss. 10. P. 13907–13946. DOI: 10.3390/s121013907.

11. Ren Z., Wang W., Wu G., Gao C., Chen W., Wei J., Huang T. Migrating web applications from monolithic structure to microservices architecture. *Internetware 18: Proc. of The Tenth Asia-Pacific Symposium on Internetware*, Beijing, China, Sep. 16, 2018. P. 1–10. DOI: 10.1145/3275219.327523.

12. De Morais W. O., Lundström J., Wickström N. A. Database-centric architecture for home-based health monitoring. *Ambient Assisted Living and Active Aging: Proc. of the 5th International Work-Conference, IWAAL*. Int. conf., Carrillo, Costa Rica, Dec. 2–6, 2013. P. 26–34. DOI: 10.1007/978-3-319-03092-0_4.

13. Vicente A., Etcheverry L., Sabiguero A. An RDBMS-only architecture for web applications. *Proceedings of the 2021 XLVII Latin American Computing Conference (CLEI)*. Int. conf., Cartago, Costa Rica, Oct. 25 – 29, 2021. DOI: 10.1109/CLEI53233.2021.9640017.

14. deMorais W. O., Lundström J., Wickström N. Active in-database processing to support ambient assisted living systems. *Sensors*. 2014. Vol. 14, iss. 8. P. 14765–14785. DOI: 10.3390/s140814765.

15. Hripcsak G. IAIMS architecture. *Journal of the American Medical Informatics Association*. 1997. Vol. 4, no. 2, suppl. 1. P. 20–30. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC61487/> (дата звернення: 11.03.2026).

16. Huff S. M., Haug P. J., Stevens L. E., Dupont R. C., Pryor T. A. HELP the next generation: a new client-server architecture. *Proceedings of the Annual Symposium on Computer Application in Medical Care*. 1994. P. 271–275. URL: <https://pmc.ncbi.nlm.nih.gov/articles/PMC2247764/> (дата звернення: 11.03.2026).

17. Mitchell J.R., Sharma P., Modi J., Simpson M., Thomas M., Hill M.D., Goyal M. A smartphone client-server teleradiology system for primary diagnosis of

acute stroke. *Journal of medical internet research*. 2011. Vol. 13, iss. 2 P. 1–12. DOI: 10.2196/jmir.1732.

18. Pak J. G., Park K. H. Advanced Pulse Oximetry System for Remote Monitoring and Management. *Journal of Biomedicine and Biotechnology*. 2012. Vol. 2012. P. 930582. DOI: 10.1155/2012/930582.

19. Mora H., Gil D., Terol R.M., Azorín J., Szymanski J. An IoT-based computational framework for healthcare monitoring in mobile environments. *Sensors*. 2017. Iss. 17. P. 2302. DOI: 10.3390/s17102302.

20. Daraghmi Y.A., Daraghmi E.Y., Daraghma R., Fouchal H., Ayaida M. Edge–Fog–Cloud computing hierarchy for improving performance and security of nb-iot-based health monitoring systems. *Sensors*. 2022. Iss. 22. P. 8646. DOI: 10.3390/s22228646.

21. Elhadad A., Alanazi F., Taloba A. I., Abozeid, A. Fog computing service in the healthcare monitoring system for managing the real-time notification. *Journal of healthcare engineering*. 2022. Vol. 2022. P. 5337733. DOI: 10.1155/2022/5337733.

22. Yıldırım E., Cicioğlu M., Çalhan A. Fog-cloud architecture-driven Internet of Medical Things framework for healthcare monitoring. *Medical and biological engineering and computing*. Iss. 61. P. 1133–1147. DOI: 10.1007/s11517-023-02776-4.

23. Vilela P. H., Rodrigues J. J. P. C., Righi R. D. R., Kozlov S., Rodrigues V. F. Looking at fog computing for e-health through the lens of deployment challenges and applications. *Sensors*. Iss. 20. P. 2553. DOI: 10.3390/s20092553.

24. Aledhari M., Razzak R., Qolomany B., Al-Fuqaha A., Saeed F. Biomedical IoT: Enabling Technologies, Architectural Elements, Challenges, and Future Directions. *IEEE Access*. 2022. Vol. 10. P. 31306–31339. DOI: 10.1109/ACCESS.2022.3159235.

25. Manickam P., Mariappan S. A., Murugesan S. M., Hansda S., Kaushik A., Shinde R., Thipperudraswamy S. P. Artificial Intelligence (AI) and Internet of Medical Things (IoMT) Assisted Biomedical Systems for Intelligent Healthcare. *Biosensors*. 2022. Vol. 12, iss. 8. P. 562. DOI: 10.3390/bios12080562.

26. He P., Huang D., Wu D., He H., Wei Y., Cui Y., Wang R., Peng L. A Survey of Internet of Medical Things: Technology, Application and Future Directions. *Digital Communications and Networks*. 2024. Article in press. DOI: 10.1016/j.dcan.2024.11.013.

27. Awad A. I., Fouda M. M., Khashaba M. M., Mohamed E. R., Hosny K. M. Utilization of Mobile Edge Computing on the Internet of Medical Things: A Survey. *ICT Express*. 2023. Vol. 9, iss. 3. P. 473–485. DOI: 10.1016/j.icte.2022.05.006.

28. Chmielewski M., Pieczonka P., Kukielka M., Gutowski T. Software architecture optimization of mobile biomedical sensor-based tools providing analytical services for disease diagnostics and assistance. *Procedia Manufacturing*. 2020. Vol. 44. P. 575–582. DOI: 10.1016/j.promfg.2020.02.253.

29. Мельник А. О., Морозов Ю. В., Гаваньо Б. І., Гупало П. А. Біомедична кіберфізична система цілодобового моніторингу функцій легень у пацієнтів із Covid-19. *Комп'ютерні системи та мережі*. 2020. Вип. 2, № 1. DOI: 10.23939/csn2020.01.001.

30. Вадурін К. О., Перекрест А. Л., Гученко М. І. Прототип кіберфізичної системи моніторингу фізичного стану оператора літального апарата. *Системи управління, навігації та зв'язку*. 2022. № 4 (70). С. 57–65. DOI: 10.26906/SUNZ.2022.4.057.

31. Khan Z. F., Alotaibi S. R. Applications of artificial intelligence and big data analytics in m-health: a healthcare system perspective. *Journal of Healthcare Engineering*. 2020. Vol. 2020. P. 8894694. DOI: 10.1155/2020/8894694.

32. Lodewyk K., Wiebe M., Dennett L., Larsson J., Greenshaw A., Hayward J. Wearables research for continuous monitoring of patient outcomes: A scoping review. *PLOS Digital Health*. 2025. Vol. 4, № 5. P. e0000860. DOI: 10.1371/journal.pdig.0000860

33. Ключко О. М. Медична інформаційна система моніторингу стану здоров'я населення із захистом персональних даних. *Medical Informatics and Engineering*. 2020. № 1. С. 17–28. DOI: 10.11603/mie.1996-1960.2020.1.11126

34. Bin K. J., Pretto L. R. De, Sanchez F. B., Battistella L. R. Digital platform to continuously monitor patients using a smartwatch: preliminary report. *JMIR Formative Research*. 2022. Vol. 6, № 9. P. e40468. DOI: 10.1016/j.media.2024.103320

35. Огородник І. М. Біотехнічна система для неінвазивного діагностування та корекції функціонального стану людини: дис. канд. техн. наук (доктора філософії): 05.11.17. Харків, 2021. 217 с. URL: <https://repo.btu.kharkov.ua//handle/123456789/7813>

36. Gopalakrishnan K., Balakrishnan A., Govardhanan K., Selvarasu S. Propositional Inference for IoT Based Dosage Calibration System Using Private Patient-Specific Prescription against Fatal Dosages. *Sensors*. 2023. Vol. 23, № 1. P.336 DOI: 10.3390/s23010336

37. Rohan R., Venkadeshwaran K., Ranjan P. Recent advancements of fiber Bragg grating sensors in biomedical application: a review. *Journal of Optics*. 2024. Vol. 53, № 1. P. 282–293. DOI: 10.1007/s12596-023-01134-9

38. Burlachenko I. S., Savinov V. Y., Tohoiev O. R., Zhuravska I. M. The Cloud GNSS Data Fusion Approach Based on the Multi-Agent Authentication Protocols' Analysis in the Corporate Logistics Management Systems. *Radio Electronics, Computer Science, Control*. 2022. No. 4. P. 95–105. DOI: 10.15588/1607-3274-2021-4-9.

39. Zhuravska I. Remote rehabilitation of the post-stroke patients with the hand and finger nerves damaged through wireless devices in individual conditions. *Journal of Medical Pharmaceutical and Allied Sciences*. 2021. Vol. 10, Iss. 5. P. 3691–3697. DOI: 10.22270/jmpas.V10I5.1594.

40. Zhuravska I., Dvoretzkyi M., Kulakovska I., Boiko A., Dvoretzka S. The Queue's Automated Creation of Doctor's Calls by Patients in the Hospital with Visualization via the Mobile Application. *Journal of Mobile Multimedia*. 2023. Vol. 19, No. 3. P. 897–934. DOI:10.13052/jmm1550-4646.19311

41. Kirkby R., Frank E. WEKA Explorer User Guide for Version 3-4-3. 2004. P. 3–5. URL: <https://weka.sourceforge.io/manuals/ExplorerGuide.pdf> (дата звернення: 11.01.2026).
42. Cao M., Ramezani R., Katakwar V. K., Zhang W., Boda D., Wani M., Naeim A. Developing Remote Patient Monitoring Infrastructure Using Commercially Available Cloud Platforms. *Frontiers in Digital Health*. 2024. Vol. 6. P. 1399461. DOI: 10.3389/fdgth.2024.1399461.
43. Django Software Foundation. Models. Django Documentation: вебсайт. URL: <https://docs.djangoproject.com/en/6.0/topics/db/models/> (дата звернення: 11.03.2026).
44. Maimaitiaili M., Jiamaliding Y., Dai G., Xiao H., Kuerbanjiang W., Yi Y. Artificial Intelligence Platform Architecture for Hospital Systems: Systematic Review. *Journal of Medical Internet Research*. 2025. Vol. 27. P. e79788. DOI: 10.2196/79788.
45. Lim H. C., Wong H., Philip R., Van Der Vegt A., Choo K.-K. R., Pole J. D., Sullivan C. Streamlining Electronic Medical Record Data Extraction and Validation in Digital Hospitals: A Systematic Review to Identify Optimal Approaches and Methods. *Learning Health Systems*. 2025. Vol. 9, iss. 4. P. e70024. DOI: 10.1002/lrh2.70024.
46. Robu R., Stoicu-Tivadar V. Regional Medical Data Mining System. *Studies in Health Technology and Informatics*. 2011. Vol. 165. P. 117–122. DOI: 10.3233/978-1-60750-735-2-117.
47. Hall M., Frank E., Holmes G., Pfahringer B., Reutemann P., Witten I. H. The WEKA Data Mining Software: An Update. *SIGKDD Explorations*. 2009. Vol. 11, no. 1. P. 10–18. DOI: 10.1145/1656274.1656278.
48. Haug P. J., Ferraro J. P., Holmen J., Wu X., Mynam K., Ebert M., Dean N., Jones J. An ontology-driven, diagnostic modeling system. *Journal of the American Medical Informatics Association*. 2013. Vol. 20, iss. e1. P. e102–e110. DOI: 10.1136/amiajnl-2012-001376.
49. Reutemann P. python-weka-wrapper3 documentation. URL: <https://fracpete.github.io/python-weka-wrapper3/> (дата звернення: 11.01.2026).

50. WEKA. Class ConverterUtils.DataSource. *WEKA API Documentation*: вебсайт. URL: <https://weka.sourceforge.io/doc.stable/weka/core/converters/ConverterUtils.DataSource.html> (дата звернення: 11.03.2026).

51. WEKA. ArffLoader. *WEKA API Documentation*: вебсайт. URL: <https://weka.sourceforge.io/doc.dev/weka/core/converters/ArffLoader.html> (дата звернення: 11.01.2026).

52. Zahoor S., Lali I. U., Khan M. A., Javed K., Mehmood W. Breast Cancer Detection and Classification using Traditional Computer Vision Techniques: A Comprehensive Review. *Current medical imaging*. 2020. Iss. 16. P. 1187–1200. DOI: 10.2174/1573405616666200406110547

53. Yang Z., Wei T., Liang Y., Yuan X., Gao R., Xia Y., Zhou J., Zhang Y., Yu Z. A foundation model for generalizable cancer diagnosis and survival prediction from histopathological images. *Nature Communications*. 2025. Vol. 16, Iss 1. P. 2366. DOI: 10.1038/s41467-025-57587-y

54. Boddu A. S., Jan A. A systematic review of machine learning algorithms for breast cancer detection. *Tissue and Cell*. 2025. Vol. 95. P. 102929. DOI: 10.1016/j.tice.2025.102929

55. Weinberger M.J., Seroussi G. From LOCO-I to the JPEG-LS Standard. *HP Labs Technical Report*. 1999. 20 p. URL: <https://shiftright.com/mirrors/www.hpl.hp.com/techreports/1999/HPL-1999-3.pdf> (дата звернення: 11.03.2026).

56. Aridhi E., Laabidi K., Mami A. FPGA technology in healthcare: A comprehensive review of hardware and software solutions for diagnostics, imaging, and patient care. *Array*. 2025. Vol. 28. P. 100622. DOI: 10.1016/j.array.2025.100622

57. Khan M. I., Silva B. da. Harnessing FPGA Technology for Energy-Efficient Wearable Medical Devices. *Electronics*. 2024. Vol. 13, iss. 20. P. 4094. DOI: 10.3390/electronics13204094

58. Wang H. FPGA Implementation of JPEG-LS Remote Sensing Image Coding Algorithm. Proceedings of the 2017 4th International Conference on Machinery, Materials and Computer (MACMC 2017). Int. conf., Xi'an, China, November 27-29,

2017. URL: <https://www.atlantis-press.com/proceedings/macmc-17> (дата звернення: 11.01.2026).

59. Pompa M., Urso G., Panunzi S., Drexler D. A., Gombos B., Gaetano A. De. A Mathematical Model of Breast Cancer Growth and Drug Resistance Evolution Under Chemotherapy. *Mathematics*. 2025. Vol. 13, iss. 7. P. 1115. DOI: 10.3390/math13071115

60. Algefary A. A characterization of common Lyapunov diagonal stability using Khatri-Rao products. *AIMS Mathematics*. 2024. Vol. 9, Iss. 8. P. 20612–20626. DOI: 10.3934/math.20241001

61. Vaghi C., Rodallec A., Fanciullino R., Ciccolini J., Mochel J. P., Matri M., Poignard C., Ebos J. M. L., Benzekry S. Population modeling of tumor growth curves and the reduced Gompertz model improve prediction of the age of experimental tumors. *PLoS computational biology*. 2020. Vol. 16, iss. 2. P. e1007178. DOI: 10.1371/journal.pcbi.1007178

62. Alqahtani R. T., Ajbar A., Aljebli E. H. Dynamics of a model of tumor–immune cell interactions under chemotherapy. *Mathematics*. 2025. Vol. 13, iss. 13. P. 2200. DOI: 10.3390/math13132200

63. Do N.-T., Le P. B., Le T. S., Tran T. T., Pham Q. H. Vibration characteristics of h-FGS plates resting on an elastic foundation under double blast loading. *Alexandria Engineering Journal*. 2024. Vol. 86. P. 131–143. DOI: 10.1016/j.aej.2023.11.050

64. Metzcar J., Wang Y., Heiland R., Macklin P. A review of cell-based computational modeling in cancer biology. *JCO clinical cancer informatics*. 2019. Vol. 2. P. 1–13. DOI: 10.1200/CCI.18.00069

65. Singh H., Srivastava H. M. *Mathematical Methods in Medical and Biological Sciences*. MA: Elsevier, 2024. 250 p. DOI: 10.1016/C2023-0-01386-3.

66. Agur Z., Elishmereni M., Kheifetz Y. Personalizing oncology treatments by predicting drug efficacy, side-effects, and improved therapy: mathematics, statistics, and their integration. *Wiley Interdisciplinary Reviews: Systems Biology and Medicine*. 2014. Vol. 6, iss. 3. P. 239–253. DOI: 10.1002/wsbm.1263

67. Баталія Б. О., Терещенко М. Ф. Моделювання розподілу температурних градієнтів при взаємодії лазерного випромінювання з біологічною тканиною. *Вісник Київського політехнічного інституту. Серія Приладобудування*. 2023. № 66 (2). С. 93–99. DOI: 10.20535/1970.66(2).2023.295047
68. Gutierrez C., Owens A., Medeiros L., Dabydeen D., Sritharan N., Phatak P., Kandlikar S. G. Breast cancer detection using enhanced IRI-numerical engine and inverse heat transfer modeling: model description and clinical validation. *Scientific Reports*. 2024. Vol. 14, iss. 1. P. 3316. DOI: 10.1038/s41598-024-53856-w
69. Horvat I. D., Iljaž J. Numerical solving of dynamic thermography inverse problem for skin cancer diagnosis based on non-Fourier bioheat model. *Strojniški vestnik-Journal of Mechanical Engineering*. 2025. Vol. 71, iss. 9–10. P. 271–283. DOI: 10.5545/sv-jme.2025.1368
70. Travasso R. D. M., Corvera Poiré E., Castro M., Rodriguez-Manzanique J. C., Hernández-Machado A. Tumor angiogenesis and vascular patterning: a mathematical model. *PloS one*. 2011. Vol. 6, iss. 5. P. e19989. DOI: 10.1371/journal.pone.0019989
71. Raissi M., Perdikaris P., Karniadakis G. E. Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational physics*. 2019. Vol. 378. P. 686–707. DOI: 10.1016/j.jcp.2018.10.045
72. Yang L., Meng X., Karniadakis G. E. B-PINNs: Bayesian physics-informed neural networks for forward and inverse PDE problems with noisy data. *Journal of Computational Physics*. 2021. Vol. 425. P. 109913. DOI: 10.1016/j.jcp.2020.109913
73. Mehta L. S., Watson K. E., Barac A., Beckie T. M., Bittner V., Cruz-Flores S., Dent S., Kondapalli L., Ky B., Okwuosa T., Piña I. L., Volgman A. S. Cardiovascular Disease and Breast Cancer: Where These Entities Intersect: A Scientific Statement From the American Heart Association. *Circulation*. 2018. Vol. 137, iss. 8. P. e30–e66. DOI: 10.1161/CIR.0000000000000556
74. Leeuwen M. T. van, Luu S., Gurney H., Brown M. R., Webber K., Pearson S.-A., Hunt L., Vajdic C. M. Cardiovascular toxicity of targeted therapies for cancer:

a protocol for an overview of systematic reviews. *BMJ open*. 2018. Vol. 8, iss. 6. P. e021064. DOI: 10.1136/bmjopen-2017-021064

75. Гончаров Д. С., Чуйко Г. П. ЕКГ-модуль ad8232: будова, характеристики і застосування. Актуальні завдання медичної, біологічної фізики та інформатики: тези доп. II Всеукр. наук.-практ. конф., Вінниця, 7 квітня 2023 р. С. 113–114. URL: <https://dspace.vnmu.edu.ua/handle/123456789/6161> (дата звернення: 12.01.2026).

76. Li C., Zhang H., Chen J., Shao S., Li X., Yao M., Zheng Y., Wu R., Shi J. Deep learning radiomics of ultrasonography for differentiating sclerosing adenosis from breast cancer. *Clinical Hemorheology and Microcirculation*. 2022. Vol. 84. P. 153–163. DOI: 10.3233/CH-221608

77. Ramani R. G., Vanitha N. S., Valarmathy S. The Pre-Processing Techniques for Breast Cancer Detection in Mammography Images. *International Journal of Image, Graphics and Signal Processing*. 2013. Vol. 5. P. 47–54. DOI: 10.5815/ijigsp.2013.05.06

78. Sudre C. H., Li W., Vercauteren T. K. M., Ourselin S., Cardoso M. J. Generalised Dice overlap as a deep learning loss function for highly unbalanced segmentations. International Workshop on Deep Learning in Medical Image Analysis International Workshop on Multimodal Learning for Clinical Decision Support. Int. conf., Quebec City, 2017. P. 240–248. DOI: 10.1007/978-3-319-67558-9_28

79. Suzuki Y., Hanaoka S., Tanabe M., Yoshikawa T., Seto Y. Predicting Breast Cancer Risk Using Radiomics Features of Mammography Images. *Journal of Personalized Medicine*. 2023. Vol. 13, iss. 11. P. 1528. DOI: 10.3390/jpm13111528

80. Ling X., Alexander G. S., Molitoris J., Choi J., Schumaker L., Tran P., Mehra R., Gaykalova D. A., Ren L. Radiomic biomarkers of locoregional recurrence: prognostic insights from oral cavity squamous cell carcinoma preoperative CT scans. *Frontiers in Oncology*. 2024. Vol. 14. 17 p. DOI: 10.3389/fonc.2024.1380599

81. Shickel B., Tighe P. J., Bihorac A., Rashidi P. Deep EHR: A Survey of Recent Advances in Deep Learning Techniques for Electronic Health Record (EHR)

Analysis. *IEEE Journal of Biomedical and Health Informatics*. 2017. Vol. 22, iss. 5. P. 1589–1604. DOI: 10.1109/JBHI.2017.2767063

82. Nguyen N. M., Ngo D. Van. Comparative analysis of boosting algorithms for predicting personal default. *Cogent Economics & Finance*. 2025. Vol. 13, №. 1. P. 2465971. DOI: 10.1080/23322039.2025.2465971

83. Chen D., Ma G., Yang E., Yang Y., Wang C., Sun Z., Liang H., Xu X., Wei J.-H. Cadinane-type sesquiterpenoid dimeric diastereomers hibisceusones A-C from infected stems of *Hibiscus tiliaceus* with cytotoxic activity against triple-negative breast cancer cells. *Bioorganic chemistry*. 2022. Vol. 127. P. 105982. DOI: 10.1016/j.bioorg.2022.105982

84. Liu Z., Lin Y., Cao Y., Hu H., Wei Y., Zhang Z., Lin S., Guo B. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Int. conf., October, 11–17, 2021. P. 9992–10002. DOI: 10.48550/arXiv.2103.14030

85. Aburass S., Dorgham O. M., Shaqsi J. Al, Rumman M. A., Al-Kadi O. Vision Transformers in Medical Imaging: a Comprehensive Review of Advancements and Applications Across Multiple Diseases. *Journal of Imaging Informatics in Medicine*. 2025. Vol. 38. P. 3928–3971. DOI: 10.1007/s10278-025-01481-y

86. Xie J., Girshick R. B., Farhadi A. Unsupervised Deep Embedding for Clustering Analysis. Proceedings of the 33rd International Conference on Machine Learning ICML 2016. Int. conf., New York, USA, June 19–24, 2016. Vol. 1. P. 740–749. DOI: 10.48550/arXiv.1511.06335

87. Bercea C. I., Wiestler B., Rueckert D., Schnabel J. A. Towards universal unsupervised anomaly detection in medical imaging. *arXiv preprint arXiv:2401.10637*. 2024. 13 p. DOI: 10.48550/arXiv.2401.10637

88. Cai H. Shapley value-based class activation mapping for improved explainability in neural networks. *The Visual Computer*. 2025. Vol. 41, iss. 10. P. 7249–7267. DOI: 10.1007/s00371-025-03803-1

89. Haripriya R., Khare N., Pandey M. Privacy-preserving federated learning for collaborative medical data mining in multi-institutional settings. *Scientific Reports*. 2025. Vol. 15. P. 12482. DOI: 10.1038/s41598-025-97565-4

90. Кандиба І. О., Гончаров Д. С., Гончарова Н. В. Інтеграція інструментарію Weka до багаторівневої системи моніторингу стану здоров'я людини. *Наука і техніка сьогодні*. 2025. № 5 (46). DOI: 10.52058/2786-6025-2025-5(46)-1571-1584. URL: <https://perspectives.pp.ua/index.php/nts/article/view/24521>.

91. Кандиба І. О., Гончаров Д. С., Гончарова Н. В. Архітектура багаторівневої системи моніторингу стану здоров'я людини. *Вчені записки Таврійського національного університету імені В. І. Вернадського. Серія Технічні науки*. 2024. № 2. С. 78–84. DOI: 10.32782/2663-5941/2024.2/11. URL: https://tech.vernadskyjournals.in.ua/journals/2024/2_2024/13.pdf.

92. Кандиба І.О., Фісун М. Т., Горбань Г. В., Антипова К. О., Гончаров Д. С. Аналіз даних в системі моніторингу стану здоров'я людини засобами Python. *Наука і техніка сьогодні*. 2024. № 5(33). С. 1162–1176. DOI: 10.52058/2786-6025-2024-5(33)-1162-1175. ISSN: 2786-6025

93. Гончаров Д. С. Діагностика раку грудей: ІОТ-система з FPGA та WEKA-аналітикою. *Технічні науки та технології*. 2025. № 3 (41). С. 185–192. DOI: 10.25140/2411-5363-2025-3(41)-185-192. URL: <http://tst.stu.cn.ua/article/view/345197> (дата звернення 11.01.2026).

94. Гончаров Д. С. Обробка даних з приладів за допомогою weka на прикладі раку молочної залози. *Ольвійський форум – 2023: стратегії країн Причорноморського регіону в геополітичному просторі: тези доповідей: Технічні науки/ Миколаїв: ЧНУ, 15–18 червня 2023*. С. 13–17. URL: <https://dspace.chmnu.edu.ua/jspui/handle/123456789/1194>.

95. Гончаров Д. С. Інтелектуальні моделі та методи моніторингу стану здоров'я людини на основі даних сенсорних мереж. *Вчені записки Таврійського національного університету імені В. І. Вернадського. Серія: Технічні науки*. 2025. № 4. С. 80–86. DOI: 10.32782/2663-5941/2025.4.2/11. URL: https://www.tech.vernadskyjournals.in.ua/journals/2025/4_2025/part_2/13.pdf.

96. Chuiko G., Honcharov D. Breast cancer dataset from Coimbra: Pre-ratings of its value to machine learning and diagnosis. *Digital Technologies Research and Applications*. 2025. Vol. 4, no. 2. P. 182–193. DOI: 10.54963/dtra.v4i2.1348. URL: <https://ojs.ukscip.com/index.php/dtra/article/view/1348> ISSN:2754-5687.

97. Chuiko G., Dvornik O., Darnapuk Y., Honcharov D., Krainyk Y., Yaremchuk O. Attribute selection, outliers impact study and visualization within breast cancer detection. Proceedings of the 2023 IEEE 13th International Conference on Electronics and Information Technologies. Int. conf., Lviv, Ukraine, Sept. 26–28 September 2023. P. 1–5. DOI: 10.1109/elit61488.2023.10310922.

98. Breast Cancer Coimbra. *UC Irvine. Machine Learning Repository*: вебсайт. URL: <https://archive.ics.uci.edu/dataset/451/breast+cancer+coimbraWisconsin%20Breast%20Cancer%20Dataset> (дата звернення 11.01.2026).

99. Chuiko G., Honcharov D. Dimensionality cutback and deep learning algorithms efficacy as to the breast cancer diagnostic dataset. *Radioelectronic and Computer Systems*. 2024. Vol. 2024, no. 4. P. 91–98. DOI: 10.32620/reks.2024.4.08. URL: <https://nti.khai.edu/ojs/index.php/reks/article/view/2652> ISSN: 2663-2012.

100. Чуйко Г. П., Гончаров Д. С., Бушук М. С. Відбір атрибутів та скорочення розмірності даних щодо раку молочної залози з Knowledge flowe. Могілянські читання – 2024: тези доповідей Всеукр. наук.-практ. конф. Комп'ютерні науки. Технічні науки, Миколаїв, 6–10 листоп. 2024. С. 152–155. URL: <https://dspace.chmnu.edu.ua/jspui/handle/123456789/2507>.

101. Breast Cancer Wisconsin (Diagnostic) Data Set. *Kaggle*: вебсайт. URL: <https://www.kaggle.com/datasets/uciml/breast-cancer-wisconsin-data> (дата звернення 11.01.2026).

102. Chuiko G., Dvornik O., Darnapuk Y., Honcharov D., Yaremchuk O. Asleep adults' breathing patterns via data mining of electromyograms. Proceedings of the IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS). Int. conf., Dortmund, Germany, 2023. P. 550–554. DOI: 10.1109/IDAACS58523.2023.10348674. URL: <https://ieeexplore.ieee.org/document/10348674> (дата звернення 11.01.2026).

103. Гончаров Д. С., Гончарова Н. В. Огляд та застосування оптичного датчика MAX30105 у медицині. Free and Open Source Software (FOSS-2023) : тези доп. XI Міжнар. наук.-практ. конф. / Харків. нац. економ. ун-т ім. Семена Кузнеця, Харків, 14–16 лютого 2023 р. Харків : ХНЕУ ім. Семена Кузнеця, 2023. С. 51–52. URL: <http://repository.hneu.edu.ua/bitstream/123456789/29041/1/foss-2023-theses.pdf> (дата звернення: 12.01.2026).

104. Гончаров Д. С. Емпірична оцінка якості вимірювань датчика пульсу шляхом статистичної обробки даних. Ольвійський форум – 2024: стратегії країн Причорноморського регіону в геополітичному просторі: тези доповідей: Технічні науки/ Миколаїв: ЧНУ, 20-23 червня 2024. С. 146–151. URL: <https://dspace.chmnu.edu.ua/jspui/handle/123456789/1194> (дата звернення 11.01.2026).

105. Гончаров Д. С. Застосування камер в медичних роботах-маніпуляторах. Інформаційні технології та інженерія : тези доповідей: Всеукраїнська наук.-практ. конф. молодих вчених, аспірантів і студентів. Миколаїв: ЧНУ, 29–02 лютого, 2024. С. 74–76 URL: <https://files.znu.edu.ua/files/Bibliobooks/Inshi79/0058873.pdf#page=74> (дата звернення 11.01.2026).

106. Гончаров Д. С., Чуйко Г. П. ЕКГ-модуль ad8232: будова, характеристики і застосування. Актуальні завдання медичної, біологічної фізики та інформатики: тези доп. II Всеукр. наук.-практ. конф., Вінниця, 7 квітня 2023 р. С. 113–114. URL: <https://dspace.vnmu.edu.ua/handle/123456789/6161> (дата звернення 11.01.2026).

107. Гончаров Д. С. Стиснення інформації у системах моніторингу стану здоров'я: методи сімейства LZx. Інформаційні технології та інженерія : тези доповідей Всеукр. наук.-практ. конф. молодих вчених, аспірантів і студентів, Миколаїв, 7–10 лютого 2023 р. С. 65–67. URL: <https://dspace.chmnu.edu.ua/jspui/handle/123456789/606> (дата звернення 11.01.2026).

ДОДАТОК А
Акти впровадження

ЗАТВЕРДЖУЮ

Перший проректор
Чорноморського національного
університету ім. Петра Могили,
д-р іст. наук, професор



Юрій КОТЛЯР

2025 р.

АКТ

впровадження результатів дисертаційної роботи

Гончарова Д. С. на тему: «Моделі та методи побудови спеціалізованих систем моніторингу стану здоров'я людини» в навчальний процес Чорноморського національного університету ім. Петра Могили на кафедрі комп'ютерної інженерії

Основні наукові та практичні результати дисертаційної роботи Гончарова Д. С. застосовуються у навчальному процесі на кафедрі комп'ютерної інженерії ЧНУ ім. Петра Могили в курсі лекційних та практичних занять при викладанні дисциплін «Сенсори, перетворювачі» та «Machine Learning and Pattern Recognition» студентам спеціальності 123 Комп'ютерна інженерія за першим (бакалаврським) рівнем вищої освіти.

У процес викладання дисциплін «Сенсори, перетворювачі» (українською мовою) та «Machine Learning and Pattern Recognition» (англійською мовою), розроблених викладачем Гончаровим Д. С., введені такі теми, які містять матеріал роботи здобувача:

- «Використання технологій Internet-of-Medical-Things для збору та попередньої обробки даних серцевого ритму»;
- «Використання алгоритмів зниження розмірності та глибокого навчання для обробки датасетів біомедичних даних».

Лекційні матеріали та методичні матеріали для виконання практичних робіт наведені в модульному об'єктно-орієнтованому динамічному навчальному середовищі Moodle3 ЧНУ ім. Петра Могили та можуть бути використовувані у подальшому освітньому процесі як очною або заочною формою навчання.

Завідувачка кафедри комп'ютерної інженерії,
д-р техн. наук, професор

Ірина ЖУРАВСЬКА

Гарант освітньої програми 123,
доцент кафедри КІ, канд. техн. наук, доцент

Ярослав КРАЙНИК

ЗАТВЕРДЖУЮ

Ректор Чорноморського
національного університету
ім. Петра Могили

Леонід КЛИМЕНКО

" 26 " листопада 2022 р.

АКТ

впровадження результатів дисертаційної роботи

Гончарова Д. С. на тему: «Моделі та методи побудови спеціалізованих систем моніторингу стану здоров'я людини» при виконанні держбюджетної НДР «Розробка модулів автоматизації бездротових приладів відновлення пост-інфарктних, пост-інсультних пацієнтів в індивідуальних умовах віддаленої реабілітації» (№ держ. реєстрації 0121U109898; керівник НДР д-р техн. наук, проф. Трунов О. М., термін виконання роботи 01.01.2021–31.12.2022)

Держбюджетна НДР «Розробка модулів автоматизації бездротових приладів відновлення пост-інфарктних, пост-інсультних пацієнтів в індивідуальних умовах віддаленої реабілітації» виконувалась у Чорноморському національному університеті ім. Петра Могили.

При виконанні НДР здобувачем Гончаровим Д. С. було проаналізовано різні методи стиснення даних за допомогою алгоритмів стиснення LZ77,78, LZW та RLE з метою вдосконалення та використання у медичній галузі. Встановлено, що кодування текстових даних за допомогою засобів української мови в середньому має надлишковість на 20–25 % більшу ніж кодування аналогічних даних засобами англійської мови. Протягом дослідження було проаналізовано, що стиснення алгоритмами LZ78 над LZ77 ефективніше у 1,5 разу. Був виконаний порівняльний аналіз алгоритмів із даними різних типів та проаналізований коефіцієнт стиснення, для визначення ефективності даних методів.

PhD-студент кафедри комп'ютерної інженерії за спеціальністю 123 Комп'ютерна інженерія, фахівець I кат. НДЧ ЧНУ ім. Петра Могили Гончаров Д. С., брав участь у НДР як виконавець.

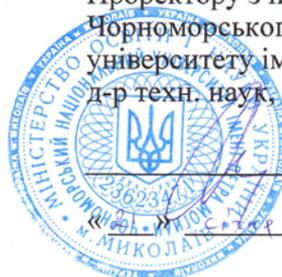
Керівник НДР,
проф. каф. автоматизації
та комп'ютерно-інтегрованих технологій,
д-р техн. наук, проф.

Олександр ТРУНОВ

« 26 » листопада 2022 р.

ЗАТВЕРДЖУЮ

Проректору з наукової роботи
Чорноморського національного
університету імені Петра Могили
д-р техн. наук, проф.



Роман ДІНЖОС

_____ 2026 р.

АКТ**впровадження результатів дисертаційної роботи**

ГОНЧАРОВА Дениса Сергійовича на тему «Моделі та методи побудови спеціалізованих систем моніторингу стану здоров'я людини» при виконанні науково-дослідної роботи № держ. реєстрації 0125U000904; науковий керівник НДР д-р фіз.-мат. наук, проф. Геннадій ЧУЙКО, термін виконання роботи 01.01.2022–31.12.2029

НДР «Інженерія даних як процеси виявлення, збору, реєстрації та подальшої обробки даних в рамках побудови апаратно-програмної інфраструктури даних» виконується у Чорноморському національному університеті імені Петра Могили.

При виконанні першого етапу НДР були використані нижче зазначені результати досліджень здобувача ГОНЧАРОВА Дениса Сергійовича:

– впроваджено метод безвратного стиснення медичних зображень із вибором JPEG-LS; що сприяє зменшенню обсягу передавання та зберігання візуалізаційних даних із збереженням діагностично значущих деталей;

– впроваджено модель контуру інтеграції «edge – сервер – аналітика WEKA»: після декомпресії дані зберігаються в БД (MySQL), конвертуються у формат ARFF та використовуються для діагностики алгоритмами RandomForest або J48, що забезпечує практичну придатність підходу для телемедичних сценаріїв;

PhD-студент кафедри комп'ютерної інженерії ЧНУ імені Петра Могили Денис ГОНЧАРОВ брав участь у НДР як виконавець.

Керівник НДР,
проф. кафедри комп'ютерної інженерії,
д-р фіз.-мат. наук, проф.

Геннадій ЧУЙКО

«21» січня 2026 р.

ДОДАТОК Б

Список опублікованих наукових праць за темою дисертації

Наукові праці, в яких опубліковані основні результати дисертації та
відповідають п. 8 Постанови КМУ від 12.01.2022 № 44

1. Chuiko G., Honcharov D. Dimensionality cutback and deep learning algorithms efficacy as to the breast cancer diagnostic dataset. *Radioelectronic and Computer Systems*. 2024. Vol. 2024, no. 4. P. 91–98. DOI: 10.32620/reks.2024.4.08. URL: <https://nti.khai.edu/ojs/index.php/reks/article/view/2652> (Last accessed: 12.01.2026). **ISSN: 2663-2012 Scopus(Q3)**

2. Chuiko G., Honcharov D. Breast cancer dataset from Coimbra: Pre-ratings of its value to machine learning and diagnosis. *Digital Technologies Research and Applications*. 2025. Vol. 4, no. 2. P. 182–193. DOI: 10.54963/dtra.v4i2.1348. URL: <https://ojs.ukscip.com/index.php/dtra/article/view/1348> (Last accessed: 12.01.2026). **ISSN: 2754-5687 Scopus**

3. Chuiko G., Dvornik O., Darnapuk Y., Honcharov D., Yaremchuk O. Asleep adults' breathing patterns via data mining of electromyograms. *Proceedings of the IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, Dortmund, Germany, 2023. P. 550–554. DOI: 10.1109/IDAACS58523.2023.10348674. URL: <https://ieeexplore.ieee.org/document/10348674> **ISSN: 2770-4254 Scopus**

4. Кандиба І. О., Гончаров Д. С., Гончарова Н. В. Інтеграція інструментарію Weka до багаторівневої системи моніторингу стану здоров'я людини. *Наука і техніка сьогодні*. 2025. № 5 (46). DOI: 10.52058/2786-6025-2025-5(46)-1571-1584. URL: <https://perspectives.pp.ua/index.php/nts/article/view/24521> (дата звернення: 12.01.2026). **кат. Б**

5. Кандиба І. О., Гончаров Д. С., Гончарова Н. В. Архітектура багаторівневої системи моніторингу стану здоров'я людини. *Вчені записки Таврійського національного університету імені В. І. Вернадського. Серія «Технічні науки»*. 2024. № 2. С. 78–84. DOI: 10.32782/2663-5941/2024.2/11. URL: https://tech.vernadskyjournals.in.ua/journals/2024/2_2024/13.pdf **кат. Б**

6. Гончаров Д. С. Інтелектуальні моделі та методи моніторингу стану здоров'я людини на основі даних сенсорних мереж. *Вчені записки Таврійського національного університету імені В. І. Вернадського. Серія: Технічні науки*. 2025. № 4. С. 80–86. DOI: 10.32782/2663-5941/2025.4.2/11. URL: https://www.tech.vernadskyjournals.in.ua/journals/2025/4_2025/part_2/13.pdf кат Б

7. Гончаров Д. С. Діагностика раку грудей: ІОТ-система з FPGA та WEKA-аналітикою. *Технічні науки та технології*. 2025. № 3 (41). С. 185–192. DOI: 10.25140/2411-5363-2025-3(41)-185-192. URL: <http://tst.stu.cn.ua/article/view/345197> кат. Б

8. Кандиба І.О., Фісун М. Т., Горбань Г. В., Антипова К. О., Гончаров Д. С. Аналіз даних в системі моніторингу стану здоров'я людини засобами Python. *Наука і техніка сьогодні*. 2024. № 5(33). С. 1162–1176. DOI: 10.52058/2786-6025-2024-5(33)-1162-1175. ISSN: 2786-6025 кат. Б

Список публікацій, які засвідчують апробацію матеріалів дисертації

– матеріали конференцій, індексовані в наукометричній базі Scopus

9. Chuiko G., Dvornik O., Darnapuk Y., Honcharov D., Krainyk Y., Yaremchuk O. Attribute selection, outliers impact study and visualization within breast cancer detection. *2023 IEEE 13th International Conference on Electronics and Information Technologies*, Lviv, Ukraine, 26–28 September 2023. P. 1–5. DOI: 10.1109/elit61488.2023.10310922.

– інші матеріали конференцій

10. Гончаров Д. С. Стиснення інформації у системах моніторингу стану здоров'я: методи сімейства LZx. *Інформаційні технології та інженерія* : тези доповідей Всеукр. наук.-практ. конф. молодих вчених, аспірантів і студентів, Миколаїв, 7–10 лютого 2023 р. С. 65–67. URL: <https://dspace.chmnu.edu.ua/jspui/handle/123456789/606> (дата звернення: 12.01.2026).

11. Гончаров Д. С., Гончарова Н. В. Огляд та застосування оптичного датчика MAX30105 у медицині. *Free and Open Source Software (FOSS-2023)* : тези

доп. XI Міжнар. наук.-практ. конф. / Харків. нац. економ. ун-т ім. Семена Кузнеця, Харків, 14–16 лютого 2023 р. Харків : ХНЕУ ім. Семена Кузнеця, 2023. С. 51–52. URL: <http://repository.hneu.edu.ua/bitstream/123456789/29041/1/foss-2023-theses.pdf> (дата звернення: 12.01.2026).

12. Гончаров Д. С., Чуйко Г. П. ЕКГ-модуль ad8232: будова, характеристики і застосування. *Актуальні завдання медичної, біологічної фізики та інформатики* : тези доп. II Всеукр. наук.-практ. конф., Вінниця, 7 квітня 2023 р. С. 113–114. URL: <https://dspace.vnmu.edu.ua/handle/123456789/6161> (дата звернення: 12.01.2026).

13. Гончаров Д. С. Обробка даних з приладів за допомогою weka на прикладі раку молочної залози. *Ольвійський форум – 2023: стратегії країн Причорноморського регіону в геополітичному просторі: тези доповідей: Технічні науки*, Миколаїв, 15–18 червня, 2023. С. 13–17. URL: <https://dspace.chmnu.edu.ua/jspui/handle/123456789/1194> (дата звернення: 12.01.2026).

14. Гончаров Д. С. Емпірична оцінка якості вимірювань датчика пульсу шляхом статистичної обробки даних. *Ольвійський форум – 2024: стратегії країн Причорноморського регіону в геополітичному просторі: тези доповідей: Технічні науки* / Миколаїв, 20–23 червня 2024 / ЧНУ ім. Петра Могили. С. 146–151. URL: <https://dspace.chmnu.edu.ua/jspui/handle/123456789/1194> (дата звернення: 12.01.2026).

15. Гончаров Д. С. Застосування камер в медичних роботах-маніпуляторах. *Інформаційні технології та інженерія* : тези доповідей: Всеукраїнська наук.-практ. конф. молодих вчених, аспірантів і студентів. Миколаїв, 29–02 лютого, 2024 / ЧНУ ім. Петра Могили. С. 74–76 URL: <https://files.znu.edu.ua/files/Bibliobooks/Inshi79/0058873.pdf#page=74> (дата звернення: 12.01.2026).

16. Чуйко Г. П., Гончаров Д. С., Бушук М. С. Відбір атрибутів та скорочення розмірності даних щодо раку молочної залози з Knowledge flowe. *Могілянські читання – 2024: тези доповідей Всеукр. наук.-практ. конф.*

Комп'ютерні науки. Технічні науки, Миколаїв, 6–10 листоп. 2024 / ЧНУ ім.
Петра Могили. С. 152–155. URL:
<https://dspace.chmnu.edu.ua/jspui/handle/123456789/2507> (дата звернення:
12.01.2026).