

Міністерство освіти і науки України
Чорноморський національний університет імені Петра Могили

Кваліфікаційна наукова
праця на правах рукопису

Дарнапук Євген Сергійович

УДК 004.67:612.1; 004.312

ДИСЕРТАЦІЯ
МЕТОДИ ТА ЗАСОБИ ОБРОБКИ МЕДИЧНИХ СИГНАЛІВ
ДЛЯ ПІДВИЩЕННЯ ЕФЕКТИВНОСТІ ДІАГНОСТИЧНИХ
СИСТЕМ

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.

_____ Є. С. Дарнапук

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

АНОТАЦІЯ

Дарнапук Є. С. Методи та засоби обробки медичних сигналів для підвищення ефективності діагностичних систем. – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії за спеціальністю 123 Комп'ютерна інженерія. – Чорноморський національний університет імені Петра Могили, Миколаїв, 2025.

Дисертаційна робота присвячена вирішенню актуальної науково-прикладної задачі вдосконалення, розробки та впровадження методів, алгоритмів та засобів обробки медичних сигналів з метою покращення якості та ефективності встановлення діагнозу пацієнта спеціалістом або діагностичними системами.

Мета та завдання дослідження. Метою дослідження є створення нових та удосконалення наявних інформаційних технологій, методів, алгоритмів та засобів обробки медичних сигналів, таких як електроміограми, рівень сатурації крові киснем тощо для підвищення ефективності встановлення діагнозу спеціалістами.

Об'єкт дослідження

Одновимірні та багатовимірні медичні сигнали, алгоритми, методи та засоби їх цифрової обробки із застосуванням інформаційних технологій та інтелектуального аналізу даних.

Предмет дослідження

Методи, засоби та технології цифрової обробки, інтелектуального аналізу та візуалізації найбільш поширених і традиційних медичних сигналів, що спрямовані на підтримку клінічних рішень.

Наукова новизна отриманих результатів:

– **вперше** запропоновано семантичну модель походження медичних даних (Data provenance), в якій IoT-пристрої, що надсилають інформацію з сенсорів до основної системи обробки, можна розглядати як основне джерело

даних для контролю, що дозволяє отримувати верифіковані медичні дані пацієнтів у відповідності до онтологій та правил для їх обробки та візуалізації, використовуючи метод головних компонент, метод наближення до еліпсу, рекурентні графіки та графіки Пуанкаре, та надає можливість спеціалістам приймати клінічні рішення;

- **набули подальшого розвитку** метод рекурентного аналізу та обробки коротких серій медичних сигналів, що, на відміну від традиційних методів оцінки варіабельності за середніми значеннями, дозволяє ідентифікувати перехідні стани фізіологічних процесів із точністю 92 %, зменшити вплив шумів на 30 % та прогнозувати циркадні коливання із похибкою менше ніж 2 %;

- **удосконалено** метод оцінки варіабельності SpO_2 , зокрема шляхом застосування міжквартильного діапазону (IQR) як надійного статистичного показника, що демонструє більшу стійкість до викидів у порівнянні зі стандартним відхиленням, допускаючи до 12 % викидів без спотворення статистичних оцінок;

- **вперше запропоновано** метод диференціального аналізу сигналів оксигенації, який враховує довжину інтервалів між перемиканнями рівня SaO_2 та їхній статистичний розподіл за законами Ерланга та Пуассона, що, на відміну від традиційних методів аналізу середніх значень, дозволяє збільшити точність оцінки динаміки змін оксигенації на 18–25 % та прогнозувати кількість перемикань рівня сатурації з похибкою не більше ніж 10 %;

- **вперше** доведена фрактальна природа сигналів щодо оксигенації артеріальної крові і вказано одновимірний фрактал – часові позиції змін рівня сатурації, – що в подальшому надає можливість зменшення частоти вимірів показників артеріальної крові максимум до $2^8 = 256$ разів без втрати їх точності.

Практичне значення отриманих результатів:

- впроваджено алгоритми цифрової фільтрації та аналізу медичних сигналів, що дозволяє підвищити точність виявлення аномалій у фізіологічних даних та покращити якість автоматизованих діагностичних систем;

- розроблено метод повнохвильового випрямлення та згладжування ЕМГ-сигналів за допомогою рухомої медіани (MM) та середньоквадратичного значення (RMS), що дозволяє ефективно відокремлювати корисний сигнал від шуму в умовах значної кількості викидів;
- впроваджено методи рекурентного аналізу та графіків Пуанкаре для оцінки варіабельності медичних сигналів, що дозволяє ідентифікувати характерні патерни змін у даних та використовувати їх для персоналізованого моніторингу пацієнтів;
- впроваджено нові інтегральні показники варіабельності рівня оксигенації артеріальної крові (SaO_2), що забезпечує більш інформативний аналіз стану пацієнтів у порівнянні з традиційними статистичними методами;
- впроваджено вдосконалені методи аналізу часових рядів для виявлення трендів та сезонних коливань медичних сигналів, що сприяє покращенню діагностики захворювань, пов'язаних з порушенням фізіологічних ритмів.
- впроваджено алгоритми аналізу головних компонент у структурі обробки медичних сигналів, що сприяє виявленню основних факторів, які впливають на зміну фізіологічних параметрів.
- впроваджено алгоритми цифрової фільтрації та аналізу медичних сигналів, що можуть бути інтегровані в автоматизовані діагностичні системи. Використання вейвлет-фільтрації Хаара дозволяє ефективно очищати медичні сигнали від шумів, що покращує якість їхньої подальшої обробки. Запропоновані методи можуть бути застосовані для аналізу електроміограм (ЕМГ), добового моніторингу артеріального тиску (АМТК), варіабельності серцевого ритму (HRV) та інших фізіологічних показників.

Результати дисертаційної роботи мають суттєве практичне значення в галузі автоматизованих систем обробки медичних сигналів, аналізу медичних даних, персоналізованої медицини та стандартизації методів оцінки фізіологічних параметрів. Запропоновані методи можуть бути використані в

клінічній діагностиці, наукових дослідженнях, розробці інтелектуальних медичних систем та навчанні спеціалістів з комп'ютерної інженерії.

Основні результати дисертаційної роботи впроваджено:

– у науково-дослідну роботу (№ ДР 0125U000904, 2022–2029 рр.) «Інженерія даних як процеси виявлення, збору, реєстрації та подальшої обробки даних в рамках побудови апаратно-програмної інфраструктури даних» (Акт впровадження від 25 грудня 2024 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 «Цифрова обробка сигналів» українською мовою та «Machine Learning and Pattern Recognition» англійською мовою (Акт впровадження від 19 вересня 2024 р.);

– у навчальний процес за другим (магістерським) рівнем вищої освіти на кафедрі комп'ютерної інженерії Чорноморського національного університету імені Петра Могили при проведенні лекційних занять та практичних робіт з дисципліни «Моделювання систем і процесів» (Акт впровадження від 17 жовтня 2024 р.);

– у рамках співпраці з німецькою ІТ компанією «Biene IT GmbH» здійснено доступ до ресурсів медичних даних, а саме: допомога в перевірці та практичному застосуванні результатів у реальних медичних закладах, включаючи інтеграцію в сучасні медичні пристрої моніторингу (Лист підтримки від Patrick Bienert, General Manager).

Відповідно до теми дисертаційного дослідження опубліковано 17 наукових праць, з них 4 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. Додатково відображають наукові результати дисертації 1 стаття – в зарубіжному періодичному науковому виданні;

3 колективні монографії (дві з них англійською мовою), 1 навчальний посібник та 1 датасет у репозиторії Mendeley Data.

У вступі обґрунтовано актуальність проблематики дослідження медичних сигналів, їх збору, обробки та представлення, наведено зв'язок дисертації з науково-дослідними роботами, сформовано та представлено мету і задачі дослідження, розглянуто об'єкт, предмет та методи дослідження, наведено наукову новизну та практичне значення отриманих результатів. Також надано інформацію щодо особистого внеску здобувача та публікацій за темою дисертаційного дослідження.

Перший розділ дисертації містить огляд сучасних методів цифрової обробки медичних сигналів, які застосовуються для покращення діагностичних можливостей медичних систем. У ньому проаналізовано існуючі методи фільтрації шумів, такі як адаптивні фільтри, вейвлет-фільтрація та спектральні методи. Окрему увагу приділено алгоритмам вилучення характерних ознак із сигналів та їхньому застосуванню в автоматизованих системах медичної діагностики. Проведено порівняльний аналіз методів класифікації патологічних станів, серед яких розглянуто нейронні мережі, метод опорних векторів та ансамблеві алгоритми.

Другий розділ присвячений розробці та реалізації нових методів обробки медичних сигналів, зокрема із застосуванням технологій Інтернету речей (IoT) та конфігурованих апаратних рішень. У роботі досліджено методи управління походженням медичних даних і їхню інтеграцію в сучасні цифрові діагностичні системи. Запропоновано стандартизовані підходи до семантичного аналізу потоків даних, що підвищує їхню точність та відповідність вимогам сучасних медичних протоколів. Вирішена задача формування, збереження та обробки великих масивів медичних сигналів із дотриманням вимог до якості та точності аналізу. Значну увагу приділено можливостям підключення сенсорних систем до діагностичних комплексів та організації їх ефективної взаємодії.

Окремо досліджено використання FPGA-пристроїв для адаптивного управління сенсорною обробкою в реальному часі. Розроблено конфігуровані

моделі підключення сенсорів, що дозволяють масштабувати систему залежно від вимог медичного моніторингу. Запропоновано методи автоматизованого налаштування параметрів обробки сигналів на апаратному рівні, що дозволяє забезпечити оптимальне використання обчислювальних ресурсів та підвищити ефективність діагностичних процедур. Окрему увагу приділено використанню вейвлет-фільтрації, зокрема вейвлетів Хаара, для цифрової обробки медичних сигналів. Виконано перетворення оригінальних рядів за допомогою вейвлетів із поділом на низькочастотні та високочастотні складові, що дозволило виділити значущі діагностичні ознаки. Розглянуто цифрові фільтри Хаара, які дозволяють адаптивно налаштовувати частотні характеристики сигналів, що підвищує точність виявлення патологій.

Третій розділ присвячений реалізації вдосконалених методів рекурентного аналізу, аналізу варіабельності та графіків Пуанкаре для обробки та візуалізації медичних сигналів. Досліджено ефективність рекурентних графіків для амбулаторного моніторингу артеріального тиску, що дозволило оптимізувати вибір параметрів аналізу та зменшити вплив шумів на 30%. Запропоновано нову методику аналізу сигналів амбулаторного моніторингу тиску крові (АМТК) з використанням зсуву вікна, що дозволяє виявляти циркадні коливання з похибкою прогнозування менше ніж 2%. Розроблено метод оцінки графіків Пуанкаре на основі ентропії Шеннона, що дає можливість ідентифікувати перехідні стани фізіологічних процесів із точністю 92%. Отримані результати підтверджують ефективність запропонованих підходів у порівнянні з традиційними методами аналізу медичних сигналів.

У четвертому розділі присвячений дослідженню методів обробки рівня оксигенації крові пацієнтів, зокрема аналізу сигналів SpO_2 та SaO_2 . Запропоновано вдосконалений підхід до оцінки варіабельності SpO_2 шляхом використання міжквартильного діапазону (IQR), що забезпечує стійкість до викидів, допускаючи до 12% аномальних значень без спотворення оцінок. Реалізовано метод диференціального аналізу сигналів оксигенації, що враховує довжину інтервалів між перемиканнями рівня SaO_2 та їхній статистичний

розподіл за законами Ерланга та Пуассона. Це дозволяє підвищити точність оцінки динаміки змін оксигенації на 18–25% та прогнозувати кількість перемикачів рівня сатурації з похибкою не більше ніж 10%. Також встановлено негативну кореляцію $(-0,86)$ між коефіцієнтом рекурентності (RR) та показниками варіабельності, що підтверджує доцільність використання рекурентного аналізу для оцінки стабільності кисневої сатурації.

У **висновках** підсумовано основні результати дослідження та окреслено можливі напрями подальшого розвитку методів цифрової обробки медичних сигналів. Запропоновані підходи продемонстрували високу ефективність у задачах аналізу медичних даних, що підтверджує їхню доцільність для впровадження в клінічну практику та системи медичного моніторингу. Використання розроблених методів дозволяє підвищити точність діагностики, скоротити час обробки сигналів та автоматизувати процеси аналізу фізіологічних параметрів.

Ключові слова: *цифрова обробка сигналів, медичні сигнали, інтегральні показники, електроміограми, IoT, FPGA, мікроконтролерна система, вейвлет-фільтрація, вейвлети Хаара, рекурентний аналіз, статистичні методи обробки даних, інформаційна система, хмарні технології, автоматизована діагностична система.*

ABSTRACT

Darnapuk Ye. S. Methods and means of processing medical signals to improve the efficiency of diagnostic systems. – Qualifying scientific work on the rights of the manuscript.

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

The PhD thesis is devoted to solving the relevant scientific and applied problem of improving, developing, and implementing methods, algorithms, and tools for processing medical signals to enhance the quality and efficiency of patient diagnosis by specialists or diagnostic systems.

Research goal and objectives. The goal of the study is to develop new and improve existing information technologies, methods, algorithms, and tools for processing medical signals, such as electromyograms, blood oxygen saturation levels, etc., to increase the effectiveness of diagnosis by specialists.

Object of research

One-dimensional and multidimensional medical signals, as well as algorithms, methods, and tools for their digital processing using information technologies and intelligent data analysis.

Subject of research

Methods, tools, and technologies for digital processing, intelligent analysis, and visualization of the most common and traditional medical signals aimed at supporting clinical decision-making.

Scientific novelty of the obtained results

For the first time:

- the semantic model of medical data provenance has been proposed, in which IoT devices transmitting sensor data to the primary processing system are considered the main data source for monitoring. This medical data provenance model, using

semantic tools, allows for verified patient medical data acquisition in accordance with ontologies and processing rules. The model facilitates data visualization through principal component analysis (PCA), ellipse approximation, recurrence plots, Poincaré plots, and other techniques, enabling specialists to make clinical decisions;

- the differential analysis method for oxygenation signals has been proposed, which takes into account the length of intervals between SaO_2 level transitions and their statistical distribution following Erlang and Poisson laws. Unlike traditional mean-based analysis methods, this approach increases the accuracy of assessing oxygenation dynamics by 18–25% and enables the prediction of saturation level transitions with an error of no more than 10%.

- the fractal nature of arterial blood oxygenation signals has been proven, identifying a one-dimensional fractal – the temporal positions of saturation level changes. This discovery allows for a reduction in the frequency of arterial blood measurements to a maximum of $2^8 = 256$ times without compromising accuracy.

Improved:

- the method for assessing SpO_2 variability has been improved by applying the interquartile range (IQR) as a robust statistical indicator. Compared to standard deviation, this method demonstrates greater resilience to outliers, tolerating up to 12% of outliers without distorting statistical assessments.

Have been developed further:

- the method of recurrence analysis and processing of short medical signal series. Unlike traditional variability assessment methods based on average values, the proposed approach allows for the identification of transitional states of physiological processes with an accuracy of 92%, reduces noise influence by 30%, and enables the prediction of circadian fluctuations with an error of less than 2%.

Practical significance of the obtained results:

- Implemented algorithms for digital filtering and analysis of medical signals, which enhance the accuracy of anomaly detection in physiological data and improve the quality of automated diagnostic systems.

- Developed a method for full-wave rectification and smoothing of EMG signals using the moving median (MM) and root mean square (RMS), allowing for effective separation of useful signals from noise under conditions with a significant number of outliers.
- Implemented recurrence analysis methods and Poincaré plots for assessing medical signal variability, enabling the identification of characteristic change patterns in data and their use for personalized patient monitoring.
- Introduced new integral indicators of arterial blood oxygenation (SaO_2) variability, providing a more informative analysis of patient conditions compared to traditional statistical methods.
- Implemented advanced time series analysis methods to detect trends and seasonal fluctuations in medical signals, contributing to improved diagnostics of diseases associated with physiological rhythm disorders.
- Integrated principal component analysis algorithms into medical signal processing, facilitating the identification of key factors influencing changes in physiological parameters.
- Implemented digital filtering and medical signal analysis algorithms, which can be integrated into automated diagnostic systems. The use of Haar wavelet filtering enables effective noise reduction in medical signals, improving the quality of further processing. The proposed methods can be applied to the analysis of electromyograms (EMG), ambulatory blood pressure monitoring (ABPM), heart rate variability (HRV), and other physiological indicators.

The results of this dissertation have significant practical value in the fields of automated medical signal processing systems, medical data analysis, personalized medicine, and standardization of physiological parameter assessment methods. The proposed methods can be utilized in clinical diagnostics, scientific research, the development of intelligent medical systems, and the training of specialists in computer engineering.

The main results of the PhD thesis have been implemented in:

- Scientific research project (State registration No. 0125U000904, 2022–2029) "Data Engineering as Processes of Detection, Collection, Registration, and Further Processing of Data Within the Framework of Building Hardware-Software Data Infrastructure" (Act of implementation dated December 25, 2024).
- Educational process at the first (bachelor) level of higher education in the Computer Engineering Department of Petro Mohyla Black Sea National University, where the results have been applied in lectures and laboratory work for the disciplines "Digital Signal Processing" taught in Ukrainian and "Machine Learning and Pattern Recognition" taught in English (Act of implementation dated September 19, 2024).
- Educational process at the second (master) level of higher education in the Computer Engineering Department of Petro Mohyla Black Sea National University, where the results have been incorporated into lectures and practical work for the discipline "Modeling of Systems and Processes" (Act of implementation dated October 17, 2024).
- As part of the collaboration with the German IT company Biene IT GmbH, access to medical data resources has been provided, specifically assisting in the validation and practical application of the results in real medical institutions, including integration into modern medical monitoring devices (Support Letter from Patrick Bienert, General Manager).

In accordance with the topic of the PhD research, 17 scientific publications have been published, including:

- 4 papers in scientific professional journals of Ukraine included to category A, which are also indexed in Scopus;
- 7 papers in proceedings of international scientific and practical conferences, indexed in Scopus;
- 1 article in an international peer-reviewed scientific journal;
- 3 collective monographs (two of them in English);
- 1 study guide;
- 1 dataset published in the Mendeley Data repository.

In the introduction, the relevance of the research problem concerning one-dimensional medical signals, their collection, processing, and representation is substantiated. The dissertation's connection with research projects is presented, along with the formulation of the research goal and objectives. The object, subject, and methods of research are described, as well as the scientific novelty and practical significance of the obtained results. Additionally, information on the author's personal contribution and publications related to the dissertation is provided.

The first chapter of the PhD thesis presents a review of modern digital processing methods for medical signals, which are used to enhance the diagnostic capabilities of medical systems. It analyzes existing noise filtering methods, such as adaptive filters, wavelet filtering, and spectral methods. Special attention is given to feature extraction algorithms and their application in automated medical monitoring systems. A comparative analysis of classification methods for pathological conditions is conducted, including the neural networks approach, support vector machines (SVM), and ensemble algorithms.

The second chapter is dedicated to the development and implementation of new methods for processing medical signals, particularly through the use of Internet of Things (IoT) technologies and configurable hardware solutions. The study explores methods for managing the provenance of medical data and their integration into modern digital diagnostic systems. Standardized approaches to semantic analysis of data streams are proposed, improving their accuracy and compliance with modern medical protocols. The problem of formation, storage, and processing of large volumes of medical signals while ensuring high-quality and accurate analysis has been addressed. Special attention is given to connecting sensor systems to diagnostic complexes and organizing their effective interaction.

The use of FPGA devices for adaptive real-time sensor processing is specifically examined. Configurable sensor connection models are developed, allowing the scalability of the system depending on medical monitoring requirements. Methods for automated adjustment of signal processing parameters at the hardware level are proposed, ensuring optimal use of computational resources and improving the

efficiency of diagnostic procedures. Particular emphasis is placed on the application of wavelet filtering, specifically Haar wavelets, for digital processing of medical signals. Transformations of original signal series using wavelets with separation into low-frequency and high-frequency components have been performed, enabling the extraction of significant diagnostic features. The study also considers Haar digital filters, which allow adaptive tuning of signal frequency characteristics, enhancing the accuracy of pathology detection.

The third chapter is dedicated to the implementation of advanced methods of recurrence analysis, variability analysis, and Poincaré plots for processing and visualizing medical signals. The effectiveness of recurrence plots for ambulatory blood pressure monitoring has been investigated, allowing for the optimization of analysis parameter selection and reducing noise impact by 30%. A novel methodology for analyzing ambulatory blood pressure monitoring (ABPM) signals using window shift has been proposed, enabling the detection of circadian fluctuations with a prediction error of less than 2%. A method for assessing Poincaré plots based on Shannon entropy has been developed, allowing for the identification of transitional states of physiological processes with an accuracy of 92%. The obtained results confirm the efficiency of the proposed approaches compared to traditional methods of medical signal analysis.

The fourth chapter is dedicated to the study of methods for processing blood oxygenation levels in patients, specifically analyzing SpO_2 and SaO_2 signals. An improved approach to assessing SpO_2 variability has been proposed using the interquartile range (IQR), which ensures robustness to outliers, allowing up to 12% of anomalous values without distorting the assessments. A differential analysis method for oxygenation signals has been implemented, taking into account the length of intervals between SaO_2 level transitions and their statistical distribution according to Erlang and Poisson laws. This approach increases the accuracy of assessing oxygenation dynamics by 18–25% and enables the prediction of saturation level transitions with an error of no more than 10%. Additionally, a negative correlation (-0.86) has been established between the recurrence rate (RR) and variability

indicators, confirming the feasibility of using recurrence analysis to assess the stability of oxygen saturation.

In the conclusions, the main research results are summarized, and possible directions for the further development of digital medical signal processing methods are outlined. The proposed approaches have demonstrated high efficiency in medical data analysis tasks, confirming their feasibility for implementation in clinical practice and medical monitoring systems. The application of the developed methods enhances diagnostic accuracy, reduces signal processing time, and automates the analysis of physiological parameters.

Keywords: *digital signal processing, medical signals, integral indicators, electromyograms, IoT, FPGA, microcontroller system, wavelet filtering, Haar wavelets, recurrence analysis, statistical data processing methods, information system, cloud technologies, automated diagnostic system.*

СПИСОК ОПУБЛІКОВАНИХ НАУКОВИХ ПРАЦЬ ЗА ТЕМОЮ ДИСЕРТАЦІЇ

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

1. Chuiko G., Darnapuk Y., Dvornik O., Krainyk Y. Abdominal electromyograms mining: breathing patterns of asleep adults. *Radioelectronic and Computer Systems*. 2023. No. 3. P. 60–72. DOI: <https://doi.org/10.32620/reks.2023.3.06>. ISSN 1814-4225. **(наукове фахове видання України кат. А, Scopus Q3)**
2. Chuiko G., Darnapuk Y. Fractal nature of arterial blood oxygen saturation data. *Radioelectronic and Computer Systems*. 2022. No. 1. P. 206–215. DOI: <https://doi.org/10.32620/reks.2022.1.16>. ISSN 1814-4225. **(наукове фахове видання України кат. А, Scopus Q3)**
3. Chuiko G., Darnapuk Y. Development of a method for differential analysis of data on the arterial blood oxygenation in healthy adults. *Eastern-European Journal of Enterprise Technologies*. 2021. Vol. 6, no. 4 (114). P. 37–43. DOI: <https://doi.org/10.15587/1729-4061.2021.244924>. ISSN 1729-3774. **(наукове фахове видання України кат. А, Scopus Q2)**
4. Chuiko G., Dvornik O., Darnapuk Y., Baganov Y. Devising a new filtration method and proof of self-similarity of electromyograms. *Eastern-European Journal of Enterprise Technologies*. 2021. Vol. 4, no. 9 (112). P. 15–22. DOI: <https://doi.org/10.15587/1729-4061.2021.239165>. ISSN 1729-3774. **(наукове фахове видання України кат. А, Scopus Q2)**

Список публікацій, які засвідчують апробацію матеріалів дисертації

5. Chuiko G., Darnapuk Y., Dvornik O., Krainyk Y., Yaremchuk O., Haab R. A new way of data analysis and rating of the blood oxygen saturation variability. *2021 IEEE 12th International Conference on Electronics and Information Technologies (ELIT)*, Lviv, Ukraine, 19–21 May 2021. 2021. DOI: <https://doi.org/10.1109/ELIT53502.2021.9501109>. **Scopus**

6. Chuiko G., Dvornik O., Darnapuk Ye., Honcharov D., Yaremchuk O. Asleep adults' breathing patterns via data mining of electromyograms. *2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, Dortmund, Germany, 7–9 Sept. 2023. DOI: <https://doi.org/10.1109/idaacs58523.2023.10348674>. **Scopus**
7. Chuiko G., Honcharov D., Dvornik O., Krainyk Y., Darnapuk Y., Yaremchuk O. Attribute selection, outliers impact study and visualization within breast cancer detection. *2023 IEEE 13th International Conference on Electronics and Information Technologies (ELIT)*, Lviv, Ukraine, 26–28 Sept. 2023. DOI: <https://doi.org/10.1109/ELIT61488.2023.10310922>. **Scopus**
8. Chuiko G., Darnapuk Ye., Dvornik O., Gravenor M., Yaremchuk O. Data mining of ambulatory blood pressure monitoring. *2023 13th International Conference on Dependable Systems, Services and Technologies (DESSERT)*, Athens, Greece, 13–15 Oct. 2023. DOI: <https://doi.org/10.1109/dessert61349.2023.10416531>. **Scopus**
9. Krainyk Y., Darnapuk Y., Simakova I. Software system for physical activity monitoring: Smart watch case. *2020 IEEE 5th International Symposium on Smart and Wireless Systems within the Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS)*, Dortmund, Germany, 17–18 Sept. 2020. DOI: <https://doi.org/10.1109/idaacs-sws50031.2020.9297093>. **Scopus**
10. Chuiko G., Darnapuk Y., Dvornik O., Yaremchuk O., Krainyk Y. Recurrence plots for ambulatory blood pressure monitoring: Means for data mining of circadian rhythms. *2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT)*, Zbarazh, Ukraine, 23–26 Sept. 2020. DOI: <https://doi.org/10.1109/csit49958.2020.9321837>. **Scopus**
11. Chuiko G., Darnapuk Ye., Dvornik O., Yaremchuk O. Blood oxygenation among healthy adults: Recurrence plots analysis and quantification. *2nd International Workshop on Information-Communication Technologies & Embedded Systems (ICT&ES-2020)*, Mykolaiv, Ukraine, 12 Nov. 2020. CEUR Workshop Proc. 2020. Vol. 2762. P. 218–226. ISSN 1613-0073. URL: <http://ceur-ws.org/Vol-2762/short3.pdf> (Last accessed: 24.01.2025). **Scopus**

**Список публікацій, які додатково відображають наукові результати
дисертації**

12. Chuiko G., Krainyk Y., Dvornik O., Darnapuk Y. IoT in provenance management of medical data. *Semantic IoT: Theory and Applications* : monograph. Cham, 2021. P. 347–362. DOI: https://doi.org/10.1007/978-3-030-64619-6_15.

Scopus

13. Chuiko G., Dvornik O., Darnapuk Y., Krainyk Y. Principal component analysis, quantifying, and filtering of Poincaré plots for time series typical for E-health. *Health Informatics: A Computational Perspective in Healthcare* : monograph. Singapore, 2021. P. 61–76. DOI: https://doi.org/10.1007/978-981-15-9735-0_4.

Scopus

14. Методи обробки медичних сигналів засобами комп'ютерної математики Maple : монографія / О. В. Дворник, Г. П. Чуйко, Є. С. Дарнапук, Я. М. Крайник. Миколаїв : Іліон, 2021. 186 с. ISBN 978-617-534-666-2. URL: <https://dspace.chmnu.edu.ua/jspui/handle/123456789/529> (дата звернення: 24.01.2025).

15. Chuiko G., Darnapuk Ye., Dvornik O., Yaremchuk O. Abdominal electromyograms (EMGs) dataset: Breathing patterns of sleeping adults. *Mendeley Data*. 2023. V3. DOI: <https://doi.org/10.17632/pmspdmgcd4.3>.

16. Chuiko G., Darnapuk Ye., Dvornik O., Yaremchuk O. Efficacy of Weka for medical data mining: Ambulatory blood pressure monitoring as a case-study. *Online Journal of Cardiology Research & Reports*. 2023. Vol. 7, Is. 3. P. 2–6. ISSN 2693-4965. DOI: 10.33552/OJCRR.2023.7.000661.

17. Цифрова обробка сигналів з практикумом в Maple 2021 : навч. посіб. / О. В. Дворник, Г. П. Чуйко, Є. С. Дарнапук, А. П. Бойко. Миколаїв : ЧНУ ім. Петра Могили, 2024. 256 с. ISBN 978-966-336-457-5. URL: <https://dspace.chmnu.edu.ua/jspui/handle/123456789/2494> (дата звернення: 24.01.2025).

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ.....	21
ВСТУП.....	22
РОЗДІЛ 1 АНАЛІЗ СТАНУ ПИТАННЯ ТА ПОСТАНОВКА ЗАДАЧ ДОСЛІДЖЕННЯ	31
1.1 Медичні сигнали, їх визначення та особливості.....	31
1.2 Аналіз наявних систем комп'ютерної математики.....	33
1.3 Аналіз літературних даних, що дотичні до сфери дослідження	35
Висновки до розділу 1	43
РОЗДІЛ 2 МЕТОДИ ОТРИМАННЯ ТА ПЕРВИННОЇ ОБРОБКИ МЕДИЧНИХ ДАНИХ.....	44
2.1 Використання технологій Internet-of-Things в управлінні походженням медичних даних	44
2.2 Настроюваний опис системи керування на основі FPGA для обробки даних з сенсорів.....	59
2.3 Метод графіків Пуанкаре для аналізу медичних сигналів: кількісний опис та обробка.....	67
2.4 Методи обробки та аналізу кноїдних та усамітнених пульсових хвиль (солітонів) в крові.....	79
Висновки до розділу 2	86
РОЗДІЛ 3 РЕАЛІЗАЦІЯ ВДОСКОНАЛЕНИХ МЕТОДІВ РЕКУРЕНТНОГО АНАЛІЗУ, АНАЛІЗУ ВАРІАБЕЛЬНОСТІ ТА ГРАФІКІВ ПУАНКАРЕ ДЛЯ ОБРОБКИ ТА ВІЗУАЛІЗАЦІЇ МЕДИЧНИХ СИГНАЛІВ	89
3.1 Використання рекурентних графіків для амбулаторного моніторингу тиску в крові.....	89
3.2 Використання графіків Пуанкаре та методу головних компонент для обробки даних добового моніторингу артеріального тиску в крові та глікемії крові	97
3.3 Покращення надійності обробки електроміограм з пошуком нових діагностичних ознак.....	101

	20
Висновки до розділу 3	109
РОЗДІЛ 4 ДОСЛІДЖЕННЯ МЕТОДІВ ОБРОБКИ РІВНЯ ОКСИГЕНАЦІЇ КРОВІ ПАЦІЄНТІВ	111
4.1 Диференціальний аналіз даних рівня оксигенації артеріальної крові у здорових дорослих	111
4.2 Використання статистичних методів, поєднаних з рекурентним аналізом для аналізу значень рівня SpO ₂ у здорових дорослих	121
Висновки до розділу 4	134
ВИСНОВКИ.....	137
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	140
ДОДАТОК А Акти впровадження та листи підтримки	160
ДОДАТОК Б Програмний код формування та обробки датасету капілярної оксигенації крові	165
ДОДАТОК В Список опублікованих наукових праць за темою дисертації.....	171

ПЕРЕЛІК СКОРОЧЕНЬ

АМТК	– амбулаторний моніторинг тиску крові
ЕМГ	– електроміограма
ЕКГ	– електрокардіограма
РГ	– рекурентний графік
МНЕ	– метод наближення до еліпсу (англ. ETF)
МГК	– метод головних компонент (англ. PCA)
СРЗ	– сингулярний розклад значень
СКМ	– система комп'ютерної математики
DFA	– Detrended Fluctuation Analysis
IQR	– Interquartile Ranges
IoT	– Internet of Things
KdV	– Korteweg–De Vries Equation
OSV	– Oxygen Saturation Variability
PCA	– Principal Component Analysis
PDF	– Probability Density Functions
RR	– Recurrence Ratio
SD	– Standard Deviations
SNR	– Signal-to-Noise Ratio
SSN	– Semantic Sensor Network

ВСТУП

Обґрунтування вибору теми дослідження.

Розвиток цифрових технологій та методів обробки медичних сигналів є ключовим фактором удосконалення сучасних медичних діагностичних систем. Автоматизований аналіз медичних даних дозволяє підвищити точність діагностики, мінімізувати вплив людського фактору та сприяти прийняттю обґрунтованих клінічних рішень. Особливого значення набуває доказова медицина [1], яка базується на систематичному аналізі наукових даних та використанні інформаційних технологій для оптимізації процесів діагностики та лікування. Існуючі методи обробки медичних сигналів потребують вдосконалення для забезпечення більшої точності та адаптивності до реальних клінічних сценаріїв.

З точки зору фізики, зокрема термодинаміки, жива істота є відкритою системою, що безперервно обмінюється речовинами, енергією та інформацією з мінливим оточуючим середовищем. Це значить, що інформація щодо стану здоров'я живого організму, або ж про його патологічний стан, може бути зафіксована фізичними чи фізіологічними сенсорами, оброблена належним чином та інтелектуально проаналізована для подальшого прийняття клінічних рішень. Таку схему можна формалізувати у вигляді ланцюжка: «Пацієнт > Сигнали > Обробка > Клінічні рішення» [2]. У дисертаційній роботі досліджуються переважно дві центральні ланки цього ланцюжку: медичні сигнали, які часто використовуються у клінічній практиці, та сучасні методи їх комп'ютерної обробки.

Можна різними способами характеризувати та класифікувати медичні сигнали, але в даній роботі зауважимо на їх різноманітті та дуже різну ефективність, тобто їх практичну клінічну частоту та розповсюдженість. У довіднику [3] наведено перелік та опис 359 медичних діагностичних тестів (вимірювань) серед яких є такі, що виконуються лише в специфічних лабораторних центрах та і відносно нечасто. Для прикладу можна навести «саузерн-блот» та зондування макромолекул як методи визначення присутності

певної послідовності складових у зразку ДНК (дезоксирибонуклеїнової кислоти). Наведені тести виконуються переважно в спеціально обладнаних криміналістичних лабораторіях задля виявлення осіб, що залишили на місці злочину кров та/або інший біоматеріал, який містить індивідуальну ДНК.

В цій роботі розглядаються більш поширені медичні вимірювання, наприклад – виміри артеріального тиску крові, частоти серцевих скорочень, рівня сатурації крові, вмісту глюкози в крові і т. д, а також – методи та алгоритми обробки цих медичних сигналів. Такі задачі, поставлені у наведеному дослідженні, вочевидь також належать до науки про дані (англ. Data Science), яка є частиною інформатики, що використовує інструменти та методи маніпулювання сирими (тобто необробленими) даними для знаходження прихованих в них нових знань, закономірностей, зв'язків тощо. Узагальнено, що науку про медичні дані можна описати основними типовими навичками, яких потребує відповідний фахівець [4]:

- застосування математики, статистики, комп'ютерного програмування та наукового методу;
- використання широкого спектру інструментів і методів для оцінки та підготовки даних – від SQL, методів інтеграції даних, і до аналізу даних;
- отримання статистичних показників з вибірок даних за допомогою прогновної аналітики та штучного інтелекту (AI), включаючи моделі машинного навчання (ML, machine learning) та глибокого навчання (DL, deep learning);
- написання та відлагоджування комп'ютерних програми, які автоматизують обробку даних і обчислення;
- інтерпретація та візуалізація результатів аналізу даних, які чітко та доступно передають значення отриманих результатів тим особам, котрі приймають клінічні рішення.

Зазначена вище комбінація навичок не є широко розповсюдженою та визначає великий попит на фахівців цієї сфери. Згідно з опитуванням IBM кількість вакансій Data Scientist (Data Analytics) зростає понад 5 % на рік [4].

Резюмуючи вищенаведене, науково-прикладні задачі, що вирішуються в цій роботі, можна віднести до інформаційних технологій, науки про дані та у застосуванні до аналізу найбільш поширених та уживаних в клінічній практиці медичних даних та вимірювань (тобто, медичних сигналів).

Зв'язок роботи з науковими програмами, планами, темами. Дисертаційна робота виконана відповідно до напрямку наукових досліджень кафедри комп'ютерної інженерії Чорноморського національного університету імені Петра Могили. Матеріали дисертаційного дослідження увійшли у заключний звіт з науково-дослідної роботи (НДР) «Розробка апаратно-програмного комплексу неінвазійного моніторингу тиску крові та частоти серцевих скорочень подвійного призначення» (№ ДР 0120U101266, 2020–2021 рр.), що фінансувалася з коштів державного бюджету. Наразі здобувач здійснює впровадження отриманих результатів власних досліджень як відповідальний виконавець НДР «Інженерія даних як процеси виявлення, збору, реєстрації та подальшої обробки даних в рамках побудови апаратно-програмної інфраструктури даних» (№ ДР 0125U000904, 2022–2029 рр.).

Мета та завдання дослідження. Метою дослідження є створення нових та удосконалення наявних інформаційних технологій, методів, алгоритмів та засобів обробки медичних сигналів, таких як електроміограми, рівень сатурації крові киснем тощо для підвищення ефективності встановлення діагнозу спеціалістами.

Об'єкт дослідження

Одновимірні та багатовимірні медичні сигнали, алгоритми, методи та засоби їх цифрової обробки із застосуванням інформаційних технологій та інтелектуального аналізу даних.

Предмет дослідження

Методи, засоби та технології цифрової обробки, інтелектуального аналізу та візуалізації найбільш поширених і традиційних медичних сигналів, що спрямовані на підтримку клінічних рішень.

Методи дослідження

Для розв'язання поставлених науково-прикладних задач використано широкий спектр методів, а саме: статистичні методи обробки даних, зокрема стійкі до наявності аномалій значень сигналів (так званих викидів), метод покриття блоками для фрактальних структур, методи математичного моделювання часових рядів, зокрема метод експоненціального згладжування, методи аналізу головних компонентів, методи теорії розпізнавання образів, машинного навчання, інтелектуального аналізу даних, методи рекурентного аналізу та графіків Пуанкаре, метод сингулярних чисел та сингулярних векторів для аналізу трендів та сезонних коливань сигналів, методи цифрової фільтрації сигналів, їх знешумлення та екстракції трендів, метод кривих Ендрюса для виявлення кластерів.

Наукова новизна отриманих результатів

– **вперше** запропоновано семантичну модель походження медичних даних (Data provenance), в якій IoT-пристрої, що надсилають інформацію з сенсорів до основної системи обробки, можна розглядати як основне джерело даних для контролю, що дозволяє отримувати верифіковані медичні дані пацієнтів у відповідності до онтологій та правил для їх обробки та візуалізації, використовуючи метод головних компонент, метод наближення до еліпсу, рекурентні графіки та графіки Пуанкаре, та надає можливість спеціалістам приймати клінічні рішення;

– **набули подальшого розвитку** метод рекурентного аналізу та обробки коротких серій медичних сигналів, що, на відміну від традиційних методів оцінки варіабельності за середніми значеннями, дозволяє ідентифікувати перехідні стани фізіологічних процесів із точністю 92 %, зменшити вплив шумів на 30 % та прогнозувати циркадні коливання із похибкою менше ніж 2 %;

– **удосконалено** метод оцінки варіабельності SpO_2 , зокрема шляхом застосування міжквартильного діапазону (IQR) як надійного статистичного показника, що демонструє більшу стійкість до викидів у порівнянні зі

стандартним відхиленням, допускаючи до 12 % викидів без спотворення статистичних оцінок;

- **вперше запропоновано** метод диференціального аналізу сигналів оксигенації, який враховує довжину інтервалів між перемиканнями рівня SaO_2 та їхній статистичний розподіл за законами Ерланга та Пуассона, що, на відміну від традиційних методів аналізу середніх значень, дозволяє збільшити точність оцінки динаміки змін оксигенації на 18–25 % та прогнозувати кількість перемикань рівня сатурації з похибкою не більше ніж 10 %;

- **вперше** доведена фрактальна природа сигналів щодо оксигенації артеріальної крові і вказано одновимірний фрактал – часові позиції змін рівня сатурації, – що в подальшому надає можливість зменшення частоти вимірів показників артеріальної крові максимум до $2^8 = 256$ разів без втрати їх точності.

Практичне значення отриманих результатів

- впроваджено алгоритми цифрової фільтрації та аналізу медичних сигналів, що дозволяє підвищити точність виявлення аномалій у фізіологічних даних та покращити якість автоматизованих діагностичних систем;

- розроблено метод повнохвильового випрямлення та згладжування ЕМГ-сигналів за допомогою рухомої медіани (MM) та середньоквадратичного значення (RMS), що дозволяє ефективно відокремлювати корисний сигнал від шуму в умовах значної кількості викидів;

- впроваджено методи рекурентного аналізу та графіків Пуанкаре для оцінки варіабельності медичних сигналів, що дозволяє ідентифікувати характерні патерни змін у даних та використовувати їх для персоналізованого моніторингу пацієнтів;

- впроваджено нові інтегральні показники варіабельності рівня оксигенації артеріальної крові (SaO_2), що забезпечує більш інформативний аналіз стану пацієнтів у порівнянні з традиційними статистичними методами;

- впроваджено вдосконалені методи аналізу часових рядів для виявлення трендів та сезонних коливань медичних сигналів, що сприяє

покращенню діагностики захворювань, пов'язаних з порушенням фізіологічних ритмів.

- впроваджено алгоритми аналізу головних компонент у структурі обробки медичних сигналів, що сприяє виявленню основних факторів, які впливають на зміну фізіологічних параметрів.

- впроваджено алгоритми цифрової фільтрації та аналізу медичних сигналів, що можуть бути інтегровані в автоматизовані діагностичні системи. Використання вейвлет-фільтрації Хаара дозволяє ефективно очищати медичні сигнали від шумів, що покращує якість їхньої подальшої обробки. Запропоновані методи можуть бути застосовані для аналізу електроміограм (ЕМГ), добового моніторингу артеріального тиску (АМТК), варіабельності серцевого ритму (HRV) та інших фізіологічних показників.

Результати дисертаційної роботи мають суттєве практичне значення в галузі автоматизованих систем обробки медичних сигналів, аналізу медичних даних, персоналізованої медицини та стандартизації методів оцінки фізіологічних параметрів. Запропоновані методи можуть бути використані в клінічній діагностиці, наукових дослідженнях, розробці інтелектуальних медичних систем та навчанні спеціалістів з комп'ютерної інженерії.

У рамках роботи проведено детальний аналіз методів оцінки варіабельності фізіологічних параметрів та їхньої відповідності різним лабораторним методам. Запропоновані алгоритми можуть бути використані для уніфікації методів калібрування медичних вимірювань, що дозволяє зменшити похибки між різними лабораторіями та покращити точність діагностичних досліджень.

Запропоновані підходи дозволяють:

- забезпечити стабільність аналізу медичних даних шляхом стандартизації методів оцінки сигналів у різних лабораторних умовах;
- розробити алгоритми класифікації фізіологічних процесів, що можуть бути використані у персоналізованій медицині;

– знизити похибки при аналізі медичних сигналів, що покращує якість діагностичних рішень та дозволяє інтегрувати ці методи у сучасні медичні прилади.

Основні результати дисертаційної роботи впроваджено:

– у навчальний процес ЧНУ ім. Петра Могили для студентів першого (бакалаврського) рівня вищої освіти зі спеціальності 123 «Комп'ютерна інженерія» у рамках дисциплін «Цифрова обробка сигналів» українською мовою та «Machine Learning and Pattern Recognition» англійською мовою (Акт впровадження ЧНУ ім. Петра Могили від 19.09.2024);

– у навчальний процес ЧНУ ім. Петра Могили для студентів другого (магістерського) рівня вищої освіти зі спеціальності 123 «Комп'ютерна інженерія» у рамках дисципліни «Моделювання систем та процесів» (Акт впровадження ЧНУ ім. Петра Могили від 17.10.2024). Це дозволяє майбутнім фахівцям отримати необхідні навички для роботи з медичними сигналами та цифровою обробкою даних;

– у НДР (№ ДР 0125U000904, 2022–2029 рр.) «Інженерія даних як процеси виявлення, збору, реєстрації та подальшої обробки даних в рамках побудови апаратно-програмної інфраструктури даних» (Акт впровадження від 25.12.2024).

Особистий внесок здобувача

Основний зміст роботи, всі теоретичні та практичні результати, висновки і дослідження, що представлено до захисту, одержані автором самостійно. Елементи дослідження опубліковано в роботах [5–21]. Особисто здобувачеві належать наступні наукові результати:

– в [17] досліджено дані електроміограм трьох пацієнтів – здорового, з міопатією та нейропатією; виконана обробка даних за допомогою методу головних компонент та фільтрація їх за допомогою вейвлетів Хаара; визначено дескриптори кожного з графіків Пуанкаре для нефільтрованих та фільтрованих наборів даних електроміограм; визначено самоподібність графіків Пуанкаре для електроміограм та виявлені ознаки ЕМГ здорових пацієнтів, специфічна форма та низьке співвідношення дескрипторів варіабельності ($\leq 0,25$) разом із

низькочастотними коливаннями сигналу (в діапазоні 5–25 Гц), що можуть бути використані для прийняття клінічних рішень.

- в роботі [18] – досліджено дані артеріальної оксигенації крові пацієнтів SaO_2 , визначено три підгрупи з різними типами мінливості даних, визначено диференціали серій вимірювання SaO_2 та їх статистичний розподіл, виявлено, що інтервали часу між перемиканнями підлягають складному статистичному розподілу Ерланга, що також можуть бути масштабно-інваріантними індикаторами мінливості процесу оксигенації крові.

- в статті [21] доведено фрактальну природу сигналів артеріальної оксигенації крові і вказано одновимірний фрактал – часові позиції змін рівня сатурації, виявлено, що зменшення дискретизації зберігає серії SaO_2 , їх прирости та багато їхніх параметрів, та вони виявилися масштабно-інваріантними в широкому діапазоні факторів зниження дискретизації (шкал).

- в роботі [22] виконано обробку даних ЕМГ за допомогою машинного навчання з використанням програмного продукту Weka, визначено, що модель Error-Trend-Seasonality була однаковою для всіх наборів даних абдомінальних ЕМГ сплячих пацієнтів; також вона показала адитивні помилки та незатухлі тренди без будь-якої сезонності. Встановлено розпад корелограм за степеневим законом, а також те, що показники Херста знаходяться в діапазоні 0,776–0,887. Це свідчить про «довгу пам'ять» (високу стійкість) абдомінальних ЕМГ.

Апробація результатів дисертації:

Матеріали дисертаційної роботи доповідалися, обговорювалися та отримали схвалення на науково-технічних конференціях та семінарах:

- Всеукраїнська науково-практична конференція «Могилянські читання» (Миколаїв, 2019–2024 рр.);

- Міжнародна науково-практична конференція «Ольвійський форум» (Миколаїв, 2021–2024 рр.);

- 2023 13th International Conference on Dependable Systems, Services and Technologies (DESSERT), Athens, Greece, 13–15 October 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;
- 2021 IEEE 12th International Conference on Electronics and Information Technologies (ELIT), Lviv, Ukraine, 19–21 May 2021;
- 1st International Workshop on Information Technologies: Theoretical and Applied Problems (ITTAP-2021), Ternopil, Ukraine, 16–18 November 2021;
- 2020 IEEE 5th International Symposium on Smart and Wireless Systems within the Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS), Dortmund, Germany, 17–18 September 2020;
- 2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT), Zbarazh, Ukraine, 23–26 September 2020.

Публікації

Відповідно до теми дисертаційного дослідження опубліковано 17 наукових праць, з них 4 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. Додатково відображають наукові результати дисертації 1 стаття – в зарубіжному періодичному науковому виданні; 3 колективні монографії (дві з них англійською мовою), 1 навчальний посібник та 1 датасет у репозиторії Mendeley Data.

Структура та обсяг дисертації

Дисертаційна робота складається зі вступу, чотирьох розділів, висновків, списку використаних джерел та трьох додатків. Основний зміст роботи викладено на 159 сторінках друкованого тексту, містить 41 рисунок та 11 таблиць. Список використаних джерел містить 178 найменувань. Загальний обсяг роботи становить 173 сторінки.

РОЗДІЛ 1

АНАЛІЗ СТАНУ ПИТАННЯ ТА ПОСТАНОВКА ЗАДАЧ ДОСЛІДЖЕННЯ

1.1 Медичні сигнали, їх визначення та особливості

В процесі обробки сигналів у різних галузях сам термін «сигнал» можна охарактеризувати як функцію, що передає інформацію про явище [23]. Будь-яка величина, що може змінюватися в просторі чи часі, може бути використана в якості сигналу для обміну повідомленнями поміж спостерігачами. Відповідно, термін «медичний сигнал» можна визначити як функцію, яка передає інформацію про певне явище, що змінює стан організму. В якості прикладів медичних сигналів можна навести частоту серцевих скорочень, рівень систолічного та діастолічного тиску в крові, рівень кисневої сатурації артеріальної крові, рентгенівські знімки, магнітно-резонансні томограми та багато іншого. Кожен з цих сигналів є своєрідним повідомленням про стан організму. Як було зазначено раніше, відповідно до довідника [3] існує 359 медичних тестів та досліджень організму, але, в більшості випадків, для встановлення діагнозу пацієнта спеціаліст використовує невелику кількість показників та тестів, які є найбільш поширеними.

Мінливість (або варіабельність) медичних сигналів є широко відомим фактом, що віддзеркалює складний динамічний баланс живих організмів з нестабільним оточенням, він же – гомеостазис, що відмічалось в [24] для процесів метаболізму. Методи цифрового аналізу виглядають особливо перспективними в концепціях Інтернету речей (зокрема розумних медичних сенсорів) та персоніфікованої медицини, що свідчить про актуальність відповідних досліджень.

В цій роботі розглядаються ті медичні сигнали та дослідження, які є найбільш поширеними та широко використовувани для встановлення діагнозу пацієнту. Це – частота серцевих скорочень, рівні систолічного та діастолічного тиску крові, оксиметричні сигнали, електроміограми, рівні глюкози в крові.

Добовий моніторинг артеріального тиску (ДМАТ) – це методика дослідження стану серцево-судинної системи за допомогою приладу автоматичної реєстрації артеріального тиску (АТ) та частоти серцевих скорочень (ЧСС) протягом заданого проміжку часу з певними інтервалами. Цей тип медичних досліджень є стандартним та широко поширеним [25–28], який використовується для виявлення гіпертензії у пацієнта. Сучасні, стійкі до руху, пристрої з пам'яттю, Bluetooth і підтримкою клієнт-серверної архітектури [28], з детальними протоколами обробки показань дозволяють спеціалістам проводити точне тестування будь-де і будь-коли.

Актуальність цього тестування пов'язана з тим, що високий кров'яний тиск створює ризик серцевих захворювань та інсульту, які є основними причинами смерті в багатьох країнах світу. Цього цілком достатньо для того, щоб вважати ДМАТ реальним об'єктом визначення, незважаючи на те, що його дослідження мають досить тривалу історію (близько 50 років). Тим часом, ДМАТ забезпечує високоефективні діагностичні показники. Так звані «ефект білого халату», замаскована гіпертензія, нічна гіпертензія та стійка гіпертензія є найбільш помітними серед них [25]. Очевидні переваги ДМАТ щодо епізодичних тестів вказано в [25; 27]. Окрім того, ДМАТ є корисним додатком для лікування гіпертензії, особливо за допомогою медикаментів [25; 26]. У цьому контексті пошук нових та удосконалення наявних методів обробки ДМАТ та інтелектуального аналізу даних все ще може бути корисним та актуальним.

Електроміографія – це стандартна процедура для діагностування стану м'язів, нервів та нервових клітин [29]. Застосування цієї процедури можна знайти в таких галузях, як, наприклад, протезування, реабілітації, спортивній аналітиці для оцінки стану організму спортсмена, та інше. Через поширеність застосування цієї процедури є висока потреба в обробці сигналів електроміографії для тої чи іншої галузі чи завдання. Отримані сигнали називаються електроміограми (ЕМГ) – це записи електричних сигналів від м'язів, та нервів, які контролюють м'язи [29]. Графіки електроміограм забезпечують високий рівень діагностичної

інформації [29]. Сучасна електроміографія використовує комп'ютерні математичні системи для обробки ЕМГ [30].

Реальні записи на перший погляд виглядають складними та хаотичними коливаннями. Ці послідовності відрізняються значною варіабельністю. Зміни статистичних характеристик сигналу під час його отримання визначають суть варіабельності [31].

Дослідження рівня оксигенації (сатурації) крові, тобто відсотку насичення гемоглобіну киснем, суттєво активізувалися внаслідок пандемії коронавірусної хвороби (COVID-19). Для першої стадії цього вірусного захворювання характерною є «прихована гіпоксія» через недостатню оксигенацію крові пацієнта [32–34]. Гіпоксія, або нестача кисню, може бути практично невиразною для рівнів насичення близьких 90 % і нижче. Діапазон норми оксигенації для здорових дорослих є 95–100 % [35].

У клінічній практиці переважно пульсоксиметрія (збір даних SpO_2) – це стандартне медичне, неінвазійне швидке вимірювання за допомогою невеликих пристроїв із затискачами пальця [36]. Існує також більш трудомісткий і інвазійний метод вимірювання насиченості газовими датчиками, що вводяться всередині артерій. Такі дані мають трохи інші позначення (SaO_2). Розбіжності між цими заходами, як правило, незначні [36]. Причина в тому, що імпульсні оксиметри калібровані переважно за допомогою прямих даних SaO_2 .

1.2 Аналіз наявних систем комп'ютерної математики

Комп'ютерні методи та алгоритми обробки медичних сигналів почали розвиватися приблизно в той самий час, як і сучасна доказова медицина. Для їх застосування зазвичай використовують різні системи комп'ютерної математики або програмні системи що включають модулі роботи з комп'ютерною математикою. Як приклад, можна навести такі програмні продукти, як Maplesoft Maple, Matlab, Wolfram Mathematica, Microsoft Excel та багато інших.

Перша комерційна версія відомої системи комп'ютерної математики Maple 6, від самого початку зорієнтованої на наукові символічні обчислення,

представлена на ринку у 2000 році [37]. Щорічно з'являється нова версія продукту, кількість вбудованих математичних функцій ядра системи давно перевищила 10'000, розвинена власна, потужна мова програмування високого рівня [38]. На рис. 1.1 зображено головне вікно системи комп'ютерної математики Maplesoft Maple 2021. У той же час перша в Україні і, можливо, одна з перших, якщо не перша, в Європі, монографія з методів обробки медичних сигналів засобами цієї системи вийшла лише у 2021 році [39].

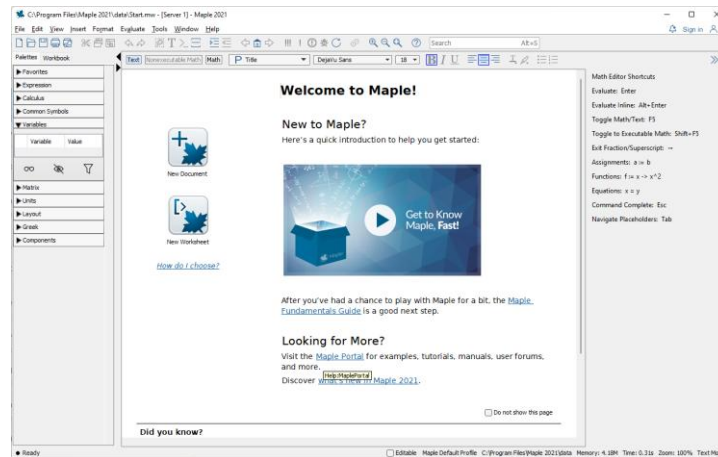


Рисунок 1.1 – Початкова сторінка СКМ Maplesoft Maple 2021

Інша, зокрема за своєю філософією, система комп'ютерної математики, Mathworks Matlab, зорієнтована не так на символічні, як на громіздкі числові обчислення, увійшла в галузь обробки медичних сигналів дещо раніше. Головне вікно програми представлено на рис. 1.2. Нині вона займає досить окреслену нішу, головним чином числової обробки медичних зображень [2].

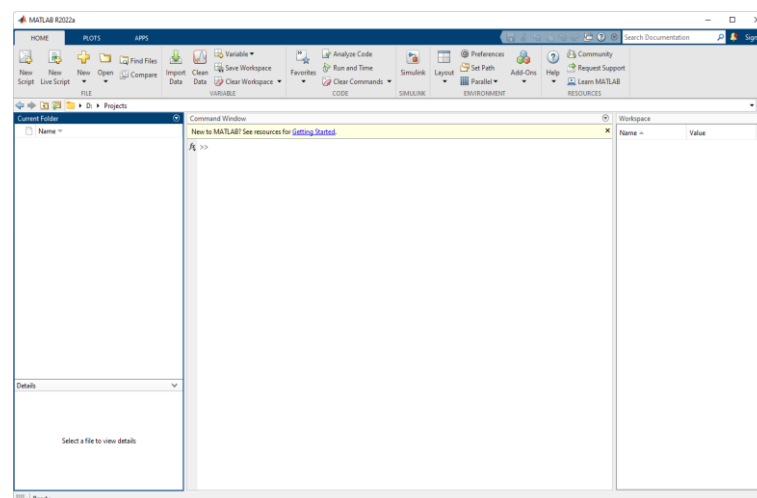


Рисунок 1.2 – Вікно СКМ MATLAB R2022a

Щодо третьої популярної в світі системи комп'ютерної математики, такої як Wolfram Mathematica, вона, маючи понад 6'000 вбудованих функцій, розрахована переважно на широкий спектр технічних обчислень [40]. Втім певний, хоча і не систематизований, і не фокусований виключно на обробку сигналів, набір 20 практичних прикладів застосування системи Wolfram Mathematica у науках про життя (біологія, медицина, охорона здоров'я тощо) можна знайти у джерелі [41].

1.3 Аналіз літературних даних, що дотичні до сфери дослідження

1.3.1 Аналіз робіт, пов'язаних з методами обробки одновимірних медичних сигналів

Сфера досліджень обробки та представлення медичних сигналів не є принципово новою, але не всі дослідження враховують ті чи інші фактори, які є важливими під час обробки медичних даних для встановлення точного діагнозу пацієнта. Нижче представлено джерела та результати попередніх досліджень, що пов'язані з обробкою масивів медичних сигналів.

Використання методу карт затримок динамічної щільності для візуалізації поведінки комплексних систем показано в [42]. Анімації, засновані на цьому методі, візуалізують значення динамічні властивості складних систем, непомітних у графіках часових рядів або стандартних графіках Пуанкаре. Але цей метод обробки та візуалізації може не враховувати тонкі особливості, що варіюються в часі.

В роботі [43] запропоновано новий дескриптор, який вони називають значенням комплексної кореляції для кількісної оцінки часового аспекту графіків Пуанкаре. Автори стверджують, що цей дескриптор є ефективним для виявлення аритмії та застійної серцевої недостатності у порівнянні з нормальним серцевим ритмом. Але автори не проводили дослідження ефективності використання цього дескриптору для обробки сигналів електроміограм.

В роботі [44] автори показують використання графіків Пуанкаре для обробки медичних сигналів. Сама робота є оглядовою, та її метою є показ

можливостей використання ГП в обробці та візуалізації медичних сигналів з використанням стандартних дескрипторів *SD1* та *SD2*.

Робота [45] також є прикладом використання графіків Пуанкаре для обробки медичних сигналів. В ній досліджується ефективність використання ГП в обробці та спостереженні змін серцевого ритму, що спричинені змінами в навантаженні.

Також прикладом використання ГП в обробці медичних сигналів є робота [46]. В ній показані методи фільтрації графіків Пуанкаре. Також виконано порівняння з методом дискретного перетворення Фур'є, що показало переваги методу ГП.

1.3.2 Публікації та дослідження в сфері аналізу та обробки даних електроміограм

Розглянемо публікації, пов'язані зі сферою досліджень електроміограм. В роботі [29] описані техніки аналізу даних електроміограм, а саме – їх визначення, обробка та класифікація. Сигнали електроміограм, отримані від м'язів, вимагають сучасних та прогресивних методів для їх виявлення, розкладання, обробки та класифікації. Робота ілюструє різні методології та алгоритми аналізу сигналів ЕМГ для забезпечення ефективних та дієвих способів розуміння сигналу та його природи. Також висвітлюються певні реалізації апаратних комплексів з використанням ЕМГ, зосереджуючись на застосунках, пов'язаних з протезуванням рук, розпізнаванням хвату та взаємодією людини з комп'ютером. Також присутнє порівняльне дослідження для показання ефективності різних методів аналізу сигналів ЕМГ. Однак метою роботи було огляд інформації щодо ЕМГ без детального загострення уваги щодо обробки цього виду сигналів. В [47] автори доводять в своєму дослідженні, що ЕМГ сигнали є не гаусівськими та нелінійними, оцінюючи ефекти поздовжнього одиночного диференціалу (ПОД), поздовжнього подвійного диференціалу (ППД) та нормального подвійного диференціалу (НПД) просторових фільтрів, форми електрода та міжелектродної відстані.

1.3.3 Аналіз наявних досліджень добового моніторингу артеріального тиску крові

Сфера дослідження добового моніторингу артеріального тиску є теж не абсолютно новою, тому ця дисертаційна робота базується та/або посиляється на результати, отриманих іншими авторами. Наприклад, в [26] показана різниця між простим заміром артеріального тиску та ДМАТ, де в якості переваги наводиться більш розширена інформація щодо стану пацієнту а саме – серцевий ритм, записи артеріального тиску протягом всього 24-годинного періоду, мінливість значень АТ, паттерн розподілу АТ та багато іншого, що дозволяє встановити більш точний діагноз пацієнту. В [48] проаналізовано масив даних АТ 1183 пацієнтів та показано розподіл поширення гіпертензії поміж групами пацієнтів. Автори статті [49] досліджують проблему виявлення факторів, що приховують неконтрольовану та помилково-неконтрольовану гіпертензію.

1.3.4 Огляд та аналіз робіт, які прямо чи опосередковано пов'язані з аналізом та дослідженням SpO_2 - та SaO_2 -сигналів

В роботі [50] показується можливість визначення гіпероксимії за допомогою пульсоксиметрії. Варіабельність оксиметричних сигналів згадується, але детально не досліджується. В [51] розглядаються принципи роботи пульсоксиметрії, її точність та її можливості для визначення хвороб, варіабельність згадується але так само детально не розглядається. Розуміння варіабельності дає уявлення про адаптивні можливості тої чи іншої системи життєзабезпечення організму у здоровому стані, або в стані патології. Варіабельність денного серцевого ритму, зокрема, може бути корисним показником патологічних станів живих істот [52].

Втім, як відмічено в [14], варіабельність сатурації крові киснем в методі SpO_2 досліджена мало і переважно не такими надійними методами, як згадана варіабельність серцевого ритму. Зауважимо принагідно, що варіабельність часових серій в межах методу SaO_2 вивчена ще менше.

Автори [53] вивчали ентропію та багато шкальну ентропію часових серій, аналізували детрендовані флуктуації, користуючись водночас відомими статистичними методами. Загальний висновок такий: варіабельність є переважно довготерміною з деякими ознаками фрактальності. Чіткого зв'язку між варіабельністю та віком піддослідних не встановлено. Окрім того, ознаки фрактальності конкретно не названі і їх походження не розкрито.

Загальновідомими статистичними характеристиками варіабельності є стандартні відхилення (дисперсії), міжквартильні діапазони та загальні розмахи даних [54]. Втім, пряме застосування цих методів, так само як і дескрипторів варіабельності з графіків Пуанкаре [17], наштотується на певні труднощі, які аналізуються в [19]. Відмінність розподілів ймовірності реальних даних від нормального розподілу Гауса (мультимодальність, велика доля викидів, асиметрія, тощо), змушує шукати нові, нетрадиційні способи оцінок варіабельності даних з оксигенації крові як для SpO_2 [19; 53], так і для SaO_2 [55].

Зокрема в [19] пропонуються нові інтегральні показники варіабельності, такі як кількість спостережуваних рівнів сатурації крові або ймовірність топових (найбільш ймовірних) рівнів. Автори роботи [55] застосували техніку диференціального аналізу первинних записів оксигенації крові і запропонували змішаний розподіл Пуассона для опису кількості перемикань між спостережуваними рівнями сатурації. Обидві роботи поділяють набори даних, з якими працюють, на три складові частини (підгрупи) з суттєво різною варіабельністю.

Практика вимагає, аби нові способи опису варіабельності, альтернативні традиційним статистичним методам [54] були достатньо надійними та простими для розуміння та застосування клініцистами. Окрім того, такі способи, звісно ж, можуть бути і суто емпіричними, як в [19; 53]. Проте краще якщо вони мають певне математико-статистичне обґрунтування, як в [54], хоча в цій роботі намічені лише контури такого опису.

1.3.5 Огляд та аналіз робіт та технологій, які пов'язані з проблемою походження медичних даних

У своїй роботі [56] Т. Бернерс-Лі представив семантичну мережу як Web 3.0 і окреслив найбільш значні переваги включення семантичних даних в існуючий вебопис. Провідним результатом цього має бути перехід у міжмашинний зв'язок, який можна встановити на семантичному рівні, а не явний зв'язок за допомогою програмних команд. Однак темпи цієї інновації недостатні, щоб оголосити семантичну мережу повсюдною технологією, і нові розробки ще попереду. Основною причиною цьому є знову ж таки неоднорідність підходів і семантичних рамок.

З іншого боку, не можна заперечувати цінні досягнення в наукових дослідженнях і практичних застосуваннях. Спочатку були розроблені стандарти семантичного представлення даних. Серед найбільш корисних серед них – Resource Description Framework (RDF), Web Ontology Language (OWL), SPARQL Protocol і RDF Query Language (SPARQL). Про них піде мова пізніше. По-друге, складні семантичні рамки та системи були розроблені для вирішення проблем управління даними, аналізу, управління системою в різних областях. Не залишається поза увагою і система охорони здоров'я, цій галузі присвячено численні дослідження.

RDF є основою для зберігання та зв'язування семантики. Вся інформація, описана в RDF-форматі, зберігається у вигляді «трійок». Трипл складається з трьох елементів:

- предмет;
- зв'язки;
- об'єкт.

Одиночний трипл визначає, як предмет і об'єкт відносяться один до одного. На елемент, визначений в одній трійці, можна посилатися з іншої. Таким чином утворюються складні мережі або графіки. Будучи досить простим форматом

представлення даних, трійки є потужним інструментом для визначення певної сфери знань або діяльності, об'єднані в структуру графа.

OWL полегшує створення онтологій [57]. Незважаючи на те, що OWL дуже схожий на RDF, його можна розглядати як об'єкт вищого рівня в якості загальних правил для конкретного домену.

SPARQL – це мова запитів із SQL-подібним синтаксисом, розроблена спеціально для запиту даних, що зберігаються у файлах RDF та OWL. SPARQL є ключовим програмним компонентом для обробки семантичних даних. SPARQL-запити можуть зберігатися як системні правила, які далі розвиваються в механізм вилучення та аналізу знань.

Незважаючи на те, що було розроблено величезну кількість онтологій і цей процес все ще триває, вони стикаються з недоліком несумісності одне з одним. Існує тенденція, що коли одна онтологія, здається, що описує всі необхідні аспекти домену, потребує поєднання з іншою, можуть виникнути численні протиріччя. Відображення інстансів з обох доменів до третього «уніфікованого» домену може вирішити цю задачу, щоб вони могли обмінюватися даними без конфліктів. Однак це призводить до розробки нової онтології. Навіть одна з найвідоміших схем, семантичні сенсорні мережі (SSN), містить найпоширеніші типи інформації, не повністю перекриває сенсорну область і постійно з'являються нові онтології. Функціональних можливостей, доступних у рамках загального призначення, недостатньо для виконання рішення для вузького домену. У той же час предметно-спеціальні онтології будуть постійно з'являтися для розширення онтологій вищого рівня.

Серед проєктів медичних онтологій виділяється проєкт Ontobee з Мічиганського університету. У ньому сконцентровано низку онтологій, пов'язаних із охороною здоров'я. Вони містять терміни про хворобу, фізичний стан, біологічні, біохімічні, опис лікарських засобів тощо. Список не обмежується згаданими доменами та включає 208 онтологій. Вони можуть бути використані в подальших розробках, які включають анотацію медичних даних у системі IoT. Однак, хоча він і містить онтології, які перекривають систему

Інтернету речей, немає готової до використання онтології для походження медичних даних.

Можна припустити, що на даний момент цей розвиток неможливо уніфікувати до певного сталого рівня. У будь-якому випадку єдина онтологія не в змозі врахувати всі особливості, які можуть виникнути в майбутньому. Отже, онтологія повинна забезпечувати точки розширення, куди можна додавати нові артефакти. Це передбачає наявність абстрактних сутностей, які служать основою для розвитку нових типів та зв'язків.

Одним із найбільших семантичних проєктів є проєкт M3 [58; 59]. Він консолідує сучасні дослідження та їхні артефакти програмного забезпечення з приблизно 550 зібраними онтологіями. M3 також створює власні рішення для типових задач у сфері IoT. Точніше кажучи, це лише один з елементів семантичної інфраструктури. Він охоплює весь потік даних для обробки датчиків і навіть підтримує створення програмного забезпечення для конкретної системи датчиків. Програмне забезпечення включає додаток для отримання даних датчиків, користувацькі інтерфейси вищого рівня, механізм обмірковування та зберігання даних.

Частина проєкту M3, яка збирає онтології є зв'язаними відкритими словниками для Інтернету речей (LOV4IoT). Каталог є динамічним і постійно включає нові відкриття, що описані в наукових дослідженнях, з посиланням на репозиторій вихідного коду. Основна мета цього підпроєкту полягає в тому, щоб зв'язати існуючі словники один з одним і створити основи для використання словників. З точки зору охорони здоров'я, онтології класифікуються як:

- онтології охорони здоров'я загального призначення;
- онтології, що існують в навколишньому середовищі;
- носимі та емоційні онтології;
- розпізнавання активності в розумному будинку;
- онтології, пов'язані з харчуванням;
- депресія;
- ожиріння тощо.

Огляд онтологій загального призначення показує, що LOV4IoT зібрав 67 онтологій (найбільша кількість серед усіх каталогів), розроблених у період з 2006 р. по 2018 рік. Також показано, що запропоновані системи або моделі використовують різноманітний набір технологій. Вони відрізняються залежно від бездротових з'єднань, протоколів зв'язку, операційних систем, цільових пристроїв кінцевих користувачів, підтримуваних датчиків і розгортання в обчислювальній інфраструктурі. Часто проекти з каталогу стосуються тільки однієї конкретної задачі, наприклад, оцифрування електрокардіограм (ЕКГ) і їх включення в іншу структуру може бути пов'язане зі значними зусиллями. Цей факт ще раз доводить, що, щоб впоратися з неоднорідністю онтологій, необхідно забезпечити інтерфейси розширення, щоб їх можна було повторно використовувати для різних цілей.

Дослідження на тему охорони здоров'я, які включають IoT та семантичні інструменти, забезпечують величезний прогрес для якості життя пацієнтів. Однак зауваження щодо цих досліджень полягає в тому, що вони зазвичай мають справу з окремими типами пацієнтів, які мають певний тип захворювання (наприклад, деменція, хвороби серця, постінсультні порушення) або власний тип даних для аналізу (наприклад, ЕКГ). Тому поєднання знахідок в універсальну систему пов'язане з додатковими витратами часу на розробку. Аналіз ресурсів також дозволяє зробити припущення про архітектурне рішення, де пристрої можуть надсилати дані в хмарне середовище, а подальші дії відбуваються в хмарі. Відгук надається в результаті аналізу хмарних інструментів або самим експертом.

Електронні системи охорони здоров'я використовують різні формати даних для зберігання та зв'язку. Коли йдеться про анотацію необроблених даних, необхідно, щоб представлення даних підтримувало додавання семантичних даних принаймні як спеціальну функцію. Якщо повідомлення протоколу не можна змінити за допомогою семантичних даних, це робить неможливим використання цього протоколу в ядрі семантичної системи.

Висновки до розділу 1

Аналіз наявних досліджень показав, що потреба в удосконаленні отримання, обробки та представлення одновимірних медичних сигналів є суттєвою, так як від якості обробки та представлення медичної інформації спеціалісту наряду залежить точність встановлення діагнозу пацієнту та ефективність його лікування. Особливої уваги вимагає також фільтрування спотворених даних, викликаних, наприклад, «ефектом білого халату» [49], [60], до якого схильні від 15 % до 30 % пацієнтів [60], що є суттєвою проблемою у правильному діагностуванні гіпертензії.

В якості програмних інструментів обробки та представлення медичних даних розглянуто системи комп'ютерної математики Maplesoft Maple, Mathworks MATLAB та Wolfram Mathematica, в результаті чого обрано програмний продукт Maplesoft Maple 2021 як систему, яка дозволяє виконувати поставлені задачі в сфері обробки та аналізу масивів медичних даних а також є не ресурсоємною.

Тому в цій роботі досліджуються найбільш поширені медичні сигнали, а саме: електроміограми, артеріальний тиск, серцевий ритм, рівень кисневої сатурації крові та інші, від значень яких наряду залежить точність діагностування хвороби пацієнта та ефективність його лікування та реабілітації.

РОЗДІЛ 2

МЕТОДИ ОТРИМАННЯ ТА ПЕРВИННОЇ ОБРОБКИ МЕДИЧНИХ ДАНИХ

2.1 Використання технологій Internet-of-Things в управлінні походженням медичних даних

2.1.1 Поняття проблеми походження медичних даних та семантичні технології

Кількість медичних даних, що генеруються IoT пристроями, невпинно зростає. Перехід до електронних протоколів та веб-систем для управління охороною здоров'я стимулює появі нових комплексних інструментів для лікування пацієнтів, моніторингу, онлайн-консультацій, хмарних технологій обробки медичних даних та інших додатків та сервісів. Такі системи зберігають, оброблюють та демонструють величезні обсяги інформації. Хоча збереження даних є важливим фактором, найважливішим результатом є рекомендації, створені в результаті аналізу вхідної інформації. Одним із інструментів покращення рекомендацій є семантичні технології.

Головна ціль семантичних технологій [61–64] є надавання нових рівнів абстракцій для розуміння машинних даних. Це дає змогу використовувати нові техніки для інтерпретації даних та покращує розуміння для рекомендацій. На базовому рівні включення семантичних технологій позначається анотацією [58], [65] вхідної інформації та подальшою обробкою за допомогою семантичних двигунів. Анотовані дані мають кращий рівень значущості та пропонують додатковий інтерфейс для розширеного аналізу.

Ще однією проблемою, яку намагаються подолати семантичні технології, є інтероперабельність [59; 65–70]. Сам IoT насамперед характеризується високим ступенем неоднорідності. Пристрої використовують різні засоби для передачі інформації, різні протоколи, під час своєї роботи виставляють специфічні прикладні програмні інтерфейси (API) тощо [71]. Усі ці факти тягнуть за собою проблему сумісності між IoT-пристроями та всіма іншими

компонентами. Наприклад, використання двох різних протоколів зв'язку у ноді датчиків температури вимагає координації на рівні сервера або шлюзу для спостереження за вимірюваннями від обох датчиків одночасно. У цьому випадку йдеться про апаратний і програмний рівень [72].

Це також має прямі наслідки для медичних IoT-пристроїв, оскільки продукти від різних виробників можуть бути несумісними з конкретною системою, вимагати окремого додатка для моніторингу тощо. Як було попередньо зазначено, семантично анотовані дані можуть служити додатковим інтерфейсом. З точки зору інтероперабельності, вимірювання з різних пристроїв можна агрегувати, щоб сформувати кадр даних про поточний стан пацієнта.

Також потрібно розглядати походження [73–75] медичних даних, згенерованих IoT-пристроями. Важливо відзначити, що медичні дані завжди слід брати до уваги в поєднанні з іншими факторами, наприклад, попередніми вимірюваннями, спостережуваним станом пацієнта, медичною картою пацієнта, параметрами самого пристрою тощо. Це створює можливість покращити якість обслуговування (QoS) для кінцевих користувачів. Наприклад, це може бути раннє виявлення хвороби та попередження її подальшого розвитку. Анотація медичних даних із семантичною інформацією, у наступному контексті, є засобом реалізації згаданих функцій.

При цьому засоби семантичної анотації, як правило, недоступні на рівні пристрою. Вони розгортаються на вищому рівні системи медичного моніторингу у вигляді програмних сервісів.

2.1.2 Керування медичними даними

Незважаючи на те, що семантичні системи охорони здоров'я є одним із найпоширеніших прикладів застосування технологій, проблема походження медичних даних ще не набула належної уваги.

Оскільки походження даних є загальною проблемою не лише для системи охорони здоров'я, а й для інших наукових галузей, які стикаються з великою кількістю інформації, у цьому напрямку було зроблено багато зусиль. По-перше,

організація W3C випустила онтологію PROV (PROV-O) для вирішення типових проблем походження даних. Онтологія встановлює базову структуру для анотування даних незалежно від джерела. Однак, підводними каменями загальних онтологій є те, що їм бракує інструментів для роботи в окремому домені. У цій ситуації це може спричинити не повністю анотовані дані, що значно обмежує можливості механізму логічного висновку, оскільки численні параметри пропускаються та не надаються системі [76].

2.1.3 Потік медичних даних

Для початку опишемо загальну семантичну інформаційну систему з медичними даними. Рівень сприйняття представлений пристроями, які можуть фіксувати фізичні параметри. До них належать прилади, які здійснюють вимірювання через певний інтервал, а також пристрої, які постійно спостерігають за станом людини. Також треба розрізняти носимі пристрої для масового ринку (наприклад, розумні годинники з розширеною функціональністю для моніторингу фізичного стану) і спеціальні медичні датчики, які відповідають за точний збір інформації. Цей фактор безпосередньо пов'язаний з походженням медичних даних. Медичні дані надсилаються одночасно від кількох IoT-агентів. Це дає можливість агрегувати дані в одному сховищі, як локальному так і хмарному, та досліджувати їх, щоб створити складні залежності між ними.

На граничному рівні пристрої підтримують попередню обробку даних і підготовку для подальшої передачі на верхні рівні. Граничні пристрої також полегшують можливості зберігання та навіть можуть використовуватися для початкової анотації даних. Однак ця точка інтересу повинна бути ретельно розроблена, оскільки вона помітно впливає на подальший потік даних. Точніше, ця ситуація стосується ситуації, коли рішення про фізичний стан приймається на ранній стадії обробки. Хоча просте припущення із вимірювань датчика може продемонструвати відсутність проблем, механізм обробки верхнього рівня може

створити рекомендацію, яка суперечить попередній. Це призведе до неоднозначного стану системи.

З іншого боку, граничний рівень підходить для діагностики сенсорних пристроїв. Передбачається, що комп'ютери (або хмарні сервіси) граничного рівня обмінюються діагностичними повідомленнями з датчиками, щоб контролювати їхній стан.

На відміну від стрімінгових застосунків, де рішення генеруються виключно щодо стану потоку з аналізом даних на інтервалі часу, наявність рівня зберігання є обов'язковою для системи з підтримкою семантики. Наявність рівня зберігання дає можливість покращити механізм семантичних правил на всьому просторі доступних даних. Так як в цьому підрозділі обговорюються семантичні дані, припускається, що сховище представлено наборами RDF-кортежів з анотаціями, доданими безпосередньо перед операцією запису. Очевидно, це означає, що для вимірювання одним датчиком доступна лише основна інформація. Як тільки необхідні зразки з датчика зібрані, їх можна розглядати як екземпляр для навчання, висновків і створення правил.

Анотації для вхідних даних виводяться з рівня онтологій. Відповідно до цього факту, що система має бути агностичною щодо гетерогенності, рівень онтології об'єднується з кількох онтологій, які можуть бути пов'язані одна з одною або через проміжні онтології, або через незалежні онтології, відповідальні за окрему галузь знань.

Щодо проблеми походження даних, то більшість уявлень буде встановлено на старті системи з мінімальними виправленнями під час її роботи. Поняття про походження можна помістити в єдину онтологію та використати, коли початково анотовані RDF-кортежі зберігаються в базі даних.

Вищезгаданий механізм семантичних правил представлений SPARQL-запитами. Запити включають логіку додатків і утворюють ядро системи з точки зору обробки додатків. Конкретний запит запускається щоразу, коли дані, пов'язані з його параметром, надходять до сховища. Для прикладу розглянемо ситуацію, коли пацієнт повідомляє про проведення вимірювання артеріального

тиску. Механізм зберігання записує цей параметр у RDF-сховище з параметром часу. Потім запускається відповідне правило, яке перевіряє, чи воно виконується в рекомендований проміжок часу. Порушення рекомендованих меж інтервалу призводить до того, що клієнт отримує повідомлення про цю подію.

Як було зазначено раніше, правила можуть бути змінені та додані експертом. У нашому випадку роль експерта виконує лікар. Він чи вона отримує доступ до даних системи через спеціальний інтерфейс. Хоча в більшості випадків система працює в автономному режимі, спостереження з боку лікаря є основоположним, а контроль з боку експерта запобігає ситуаціям, коли механізм висновків може спричинити збій або інші серйозні наслідки. Тому кожне сформоване правило спочатку повинен затвердити лікар.

Схематичне зображення семантичної медичної системи з усіма зазначеними компонентами зображено на рис. 2.1.



Рисунок 2.1 – Узагальнена структура семантично-підтримуваної медичної системи

У цій системі, походження медичних даних можна встановити на різних рівнях системи для різних випадків, що стосуються суб'єктів системи. Вона має надавати послугу, яка вказує на неправильні висновки з правильних даних і запобігає використанню неправильних даних для прийняття рішень. Особливо це критично для системи охорони здоров'я, яка відповідає за стан пацієнтів.

2.1.3 Формати медичних даних

Враховуючи походження медичних даних, важливо мати на увазі представлення цих даних.

Очевидно, що текстові формати є кращим вибором для використання в семантичній системі. Оскільки семантичне походження означає, що в системі переважає текстова інформація, а самі семантичні анотації виконуються в текстовому форматі, такі представлення легко об'єднати з анотаціями у щось змістовне. Однак додавання нової інформації до повідомлення, яке має бути передане, завжди стосується узгодженості на рівні зв'язку. Через те, що більшість IoT-систем не розраховані на такі модифікації, система повинна інтегрувати анотацію даних на більш високих рівнях.

Перший формат даних, який слід згадати, це формат значень, розділених комами (Comma-Separated Values, CSV). Це широко поширений формат, який використовується в багатьох сферах інформаційних систем. CSV припускає, що збережені значення розділені комою, а системи обробки отримують доступ до необхідного маркера, обходячи рядок і вирізаючи підрядок з основного рядка. З іншого боку, CSV – це послідовний формат, тому аналіз файлу CSV із подальшим вставленням додаткової інформації ускладнюється пошуком по його вмісту. Сам процес вбудовування інформації абсолютно простий. Зрештою, більшість пристроїв використовують інші протоколи, які більше орієнтовані на представлення медичних подій.

Одним із найвідоміших форматів даних для сфери охорони здоров'я є HL7 (Health Level Seven International) [77]. Якщо говорити точніше, то це не просто формат даних, HL7 – це набір стандартів для розробки інформаційних систем у

сфері охорони здоров'я. Він служить рішенням для кількох типів передачі медичних даних (наприклад, HL7 аECG для обміну даними ЕКГ). HL7 базується на розширеній мові розмітки (XML) і, отже, також відноситься до типу текстових форматів. Однак основним недоліком HL7, незважаючи на те, що він широко адаптований і напевно може бути рекомендований для включення в архітектуру, є розширюваність.

Недоліки протоколу HL7 призвели до появи його другої, третьої та вдосконаленої версії HL7 Fast Healthcare Interoperability Resources (FHIR). Графік актуальності версій стандарту HL7 наведено на рис. 2.2. Це дозволяє змішувати та адаптувати до особливого клінічного контексту.

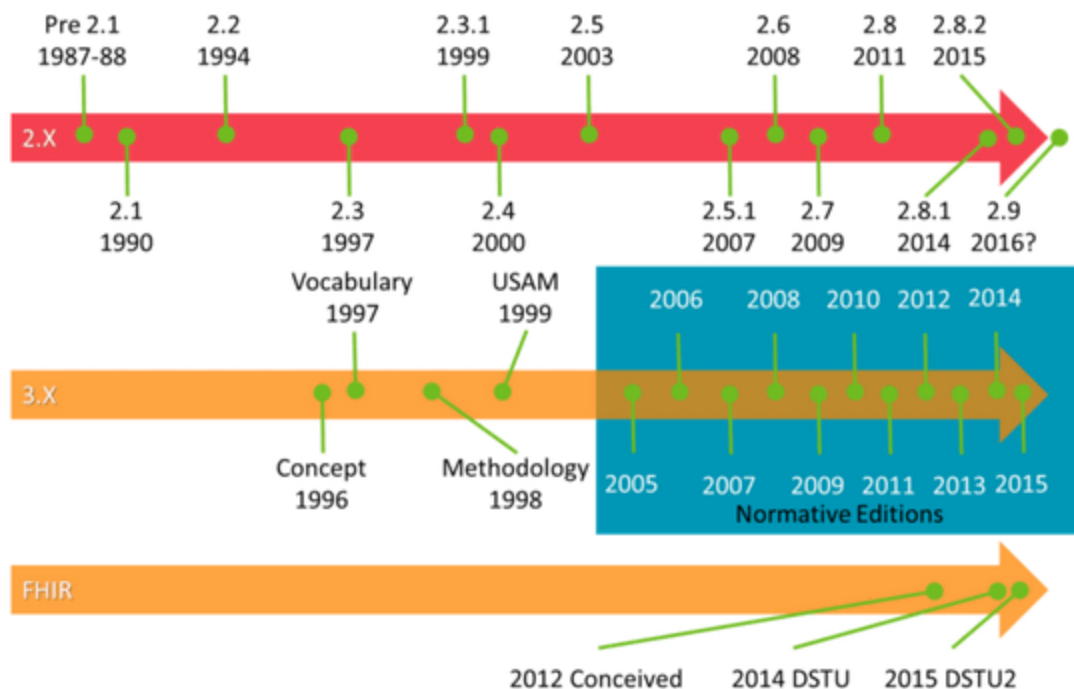


Рисунок 2.2 – Актуальність поточних версій стандарту HL7

Станом на 2024 р. актуальними є три гілки версій стандарту HL7 [78–80] до кожної з яких виходять регулярні оновлення від некомерційної організації «Health Level Seven International (HL7)».

Медична інфраструктура також значною мірою покладається на графічну інформацію, яка набагато складніша для автоматичного анотування, ніж текстові дані. Список форматів даних медичних зображень включає в себе IntefFile, Analyze, Nifti, Minc, DICOM. Додавання семантичних даних для таких зображень

є просто зв'язуванням між самим файлом і відповідною інформацією з онтологій і бази правил.

2.1.4 Аналіз пропонуваного рішення для проблеми походження медичних даних

По-перше, будемо стверджувати, що на найнижчому рівні можна визначити джерело помилкових даних:

1) IoT-пристрій;

2) пацієнт, дії якого безпосередньо впливають на отримані вимірювання. Дуже часто причиною цього є неправильне використання пристрою, яке не відповідає інструкціям пристрою або протоколам вимірювання того чи іншого параметру.

Щоб розширити номенклатуру походження даних, розглянемо IoT-пристрій. Незважаючи на відносно простий обчислювальний пристрій, він поєднує в собі кілька технологій, які слід брати до уваги. Нижче наведено загальні пояснення помилок на рівні пристрою:

1) помилка на сенсорному рівні (чутливий елемент не збирає дані належним чином у заявленому діапазоні чутливості з необхідною точністю або інші джерела помилкових даних);

2) апаратний рівень (оброблювальний пристрій не може отримати дані та передати їх до кінцевого пристрою);

3) рівень електроживлення (цей рівень пов'язаний з двома попередніми, так як може спричинити помилки для обох з них)

4) програмний рівень (включає різні аспекти, наприклад, збій протоколу, невідповідність програмного забезпечення, збої через певні стани програмного середовища тощо);

5) комунікаційний рівень (з'являється під час фази передачі та частково залежить від вибраного протоколу зв'язку).

Цього списку достатньо, щоб забезпечити контроль над більшістю use-cases, які розглядаються.

Особливу увагу треба приділити до помилок сенсорного рівня. Для позначення стану датчиків пропонується наступна класифікація:

- 1) сенсор в нормальному стані за нормальних умов;
- 2) сенсор отримує дані, які є упередженими з попередньої історії, і пацієнт усвідомлює, що стан нормальний;
- 3) сенсор збирає дані з вихідними значеннями, розташованими поза межами діапазону чутливості (застосований імпульс не підходить для поточного датчика);
- 4) сенсор отримує апріорі неправильні дані за відомих умов;
- 5) сенсор не може зібрати дані.

Другий випадок привертає найбільшу увагу, оскільки він є типовим для більшості випадків і стосується як правильних, так і невірних даних. Невелике відхилення від вимірюного значення може вказувати на подальші тенденції у фізичному стані пацієнта. Таким чином, ця інформація є корисною для запобігання можливим негативним впливам. При цьому невелика розбіжність у значенні амплітуди не обов'язково означає що стан пацієнта є небезпечним або з відхиленням від норми. Таким чином, це дозволяє заздалегідь інформувати пацієнта про можливі результати за вимірюваннями датчиків і рекомендувати йому профілактичні заходи.

На відміну від апаратного рівня, узагальнення дій пацієнта, які апріорі впливають на медичні дані, не може включати всі можливі варіанти. Більше того, автоматичне виправлення цієї інформації не є можливим, якщо IoT-пристрій не підтримує цю функцію. Тому ця частина також має полегшити інтерфейс, коли пацієнт додає докази неналежного використання пристрою. Користувач вносить відгук про неправильні вимірювання, а також відзначає проміжок часу, коли ймовірно записуються дані за цих обставин. Результатом цієї діяльності є зниження достовірності даних, записаних протягом позначеного періоду. Це також означає, що дані не можуть бути використані під час процесу висновку або, принаймні, їхній вплив не такий сильний.

Ситуація, що розглядається, тісно пов'язана з можливостями семантичних технологій, специфікацією IoT-пристроїв та їх обмеженнями. Проблема, яка виникає, коли пацієнту дозволяється вводити власні пояснення, полягає в тому, що семантичне міркування базується на прямому збігу між лексемами у відповіді пацієнта та даними з онтологій. Якщо відповідності не знайдено, жодні семантичні дані не можуть бути пов'язані з цим повідомленням автоматично і тому вони потребують ручного налаштування. Життєздатним рішенням для цієї проблеми є список із параметрами для вибору, щоб пояснити причину невдалого вимірювання. Однак він повинен бути розроблений для кожного сенсору/пристрою окремо, і навіть список цих параметрів повинен містити параметр, який дозволяє вводити причини, які не були передбачені під час підготовки списку. Інформація, отримана від індивідуального датчика, має свої особливості, і її інтерпретація повинна суворо відповідати документації.

Тим не менш, існує ймовірність того, що деякі дані залишаються не анотованими та виключаються із запитів і висновків. Щоб подолати цю проблему, пропонується використовувати процедуру коли лікар анотує дані вручну, а посилання створюються в результаті початкового ручного додавання.

З іншого боку, можна зіткнутися з ситуацією, коли значення, виміряне одним сенсором, дійсно є правильним і свідчить про обережний фізичний стан пацієнта. Цей сценарій є критичним для розгляду, і автоматична відмова від зразка може призвести до критичних наслідків. Таким чином, можна запропонувати використовувати поняття «сценарій», який описує можливі наслідки аналізу отриманих даних. У цьому випадку в системі існує кілька одночасних сценаріїв. Принаймні один із них, позначений як «критичний» сценарій, описує ситуацію, коли необхідно негайно вжити заходів. Як правило, створюється два паралельних сценарії. Другий випадок передбачає, що дані можуть бути отримані з помилкового джерела, і потрібен додатковий аналіз.

Критичний сценарій також означає, що лікар або служба медичного спостереження сповіщені про можливий стан пацієнта. При цьому пацієнт також повідомляється про свій критичний стан і має можливість підтвердити або

спростувати цей факт. Отже, системі необхідний механізм двостороннього зв'язку. Можливість такого сценарію є основною причиною того, чому дизайн системи допускає неавтономну комунікацію між пацієнтом і лікарем. Ця незамінна опція має бути присутня в медичній системі охорони здоров'я. Незважаючи на те, що семантичні технології значно збагачують можливості системи, вони не можуть повністю замінити досвід експертів або уникнути невірних суджень у обмеженому інформаційному просторі.

Іншим варіантом реалізації обробки критичного сценарію є її розгортання на рівні периферійних пристроїв без реактивних дій, необхідних від основної системи. Однак, оскільки граничні пристрої спостерігають лише початковий сигнал, вони дають відповідь лише відповідно до маркерів, вбудованих у програмне забезпечення (наприклад, жорстко закодовані значення рекомендованих вихідних меж датчика, де значне зміщення забезпечує виконання заздалегідь запрограмованих дій).

Остаточний рівень походження даних для системи охорони здоров'я повністю консолідовано в семантичному механізмі. Оскільки дані зберігаються на рівні збереження, отримана інформація служить фундаментальною основою для управління семантичним походженням.

2.1.5 Походження даних і надійність калібраторів мелатоніну-сульфату: практичне дослідження

Проблеми зі здоров'ям, пов'язані з порушеннями циркадних ритмів, мають «спільне коріння» – мелатонін [81]. Багато поширених захворювань, таких як рак, метаболічні порушення, діабет, проблеми серцево-судинної системи, пов'язані з раптовими фазовими змінами рівня мелатоніну, як зазначено у дослідженнях [81]. Мелатонін, відомий як «гормон темряви», є універсальним гормоном епіфіза. Він регулює багато фізіологічних процесів у людей та ссавців загалом [82]. Мелатонін впливає на фізіологію людини в цілому, регулюючи цикл сну і неспання через так званий «внутрішній годинник організму» [81; 83].

Наявність мелатоніну у слині чи плазмі крові, а також метаболітів мелатоніну у сечі є добре відомими фактами [82; 83].

Ензим-зв'язаний імуносорбентний аналіз (ELISA) є визнаною, точною сучасною методикою тестування мелатоніну в сечі. Це підтверджують як ранні звіти [84–87], так і сучасні роботи [88]. Цей досить складний тест дозволяє точно визначати присутність та концентрацію метаболітів мелатоніну в сечі.

Використання калібраторів часто є найкращим рішенням у подібних ситуаціях. У роботі [89] наведено детальний посібник з використання декількох калібраторів, кожен із яких має відому дозу мелатоніну, для побудови калібрувальної кривої. Однак автори [89] не аналізували свої результати з точки зору походження даних, хоча їх надійність залежить від цього фактору. Тому пропонується інший підхід до медичних даних. Він враховує походження даних, засноване на сучасних технологіях отримання, передачі та обробки в умовах реальної компетенції медичного персоналу.

Якщо є калібрувальна крива («доза-відповідь»), то відповідь зразка визначає концентрацію мелатоніну (дозу). Відповідь, наприклад, пропускання чи оптична густина зразка, отримана фотометром, дозволяє одноразово визначити дозування мелатоніну [84–88]. Такі недорогі та швидкі дослідження можуть стати доступними для менш забезпечених клінік і пацієнтів.

Добре відома логістична крива Хілла здебільшого використовується для калібрувальних графіків типу «доза-відповідь» [90]. Існують варіанти з чотирма або п'ятьма параметрами, але простіший варіант є більш популярним. В дисертаційній роботі розглянуто результати, наведені в [89], з доповненням їх більш сучасними даними [88]. Це дозволить простежити вплив походження даних на параметри калібрувальної кривої.

Калібратори були зразками розчинів метаболіту мелатоніну в сечі. Визначення концентрації в зразках здійснювалося двома методами. Першим був ензим-зв'язаний імуносорбентний аналіз (ELISA) [84–88]. Іншим був радіоімунний аналіз (RIA), відомий як високоточна методика тестування.

Типові концентрації метаболіту мелатоніну в сечі знаходяться в діапазоні 0–420 нг/мл. Зазвичай кількість калібраторів у наборі становить 6–8 [84–88]. Опис рівняння Хілла можна знайти у [90]:

$$Y = \frac{a+(b-a)}{1+(\frac{c}{x})^d}, \quad (2.1)$$

де Y – відповідь, що відповідає дозі x ;

a, b, c, d – чотири параметри. Параметр a є асимптотичною відповіддю за умови, що $x \rightarrow \infty$. Інша асимптота b є стабілізованою відповіддю за умови, що $x \rightarrow 0$. Параметр c визначає точку перегину кривої «доза-відповідь» та включає в себе різні терміни (наприклад: EC50, ED50, LD50, IC50). Параметр d є так званим нахилом Хілла [90].

На рис. 2.3 показано графік Хілла з описом параметрів.

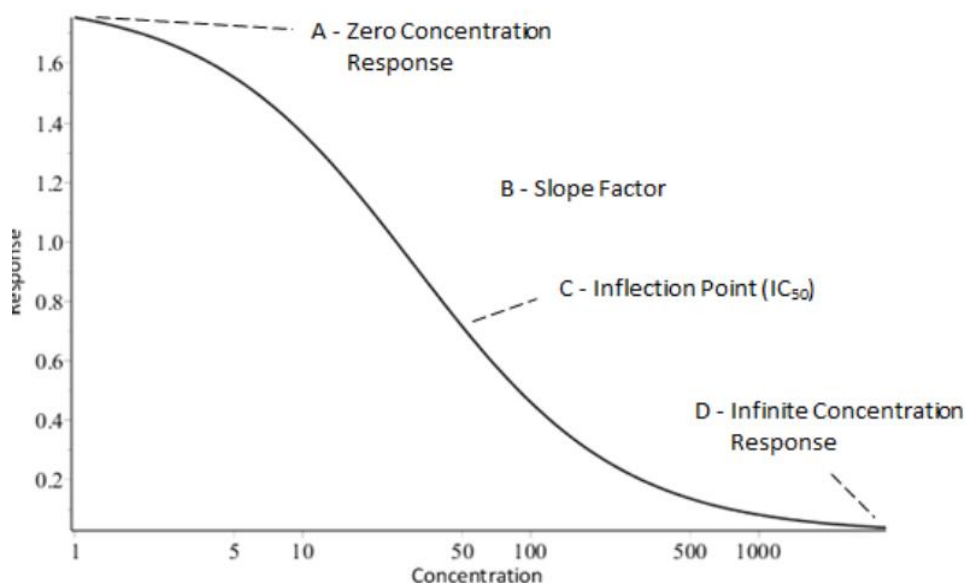


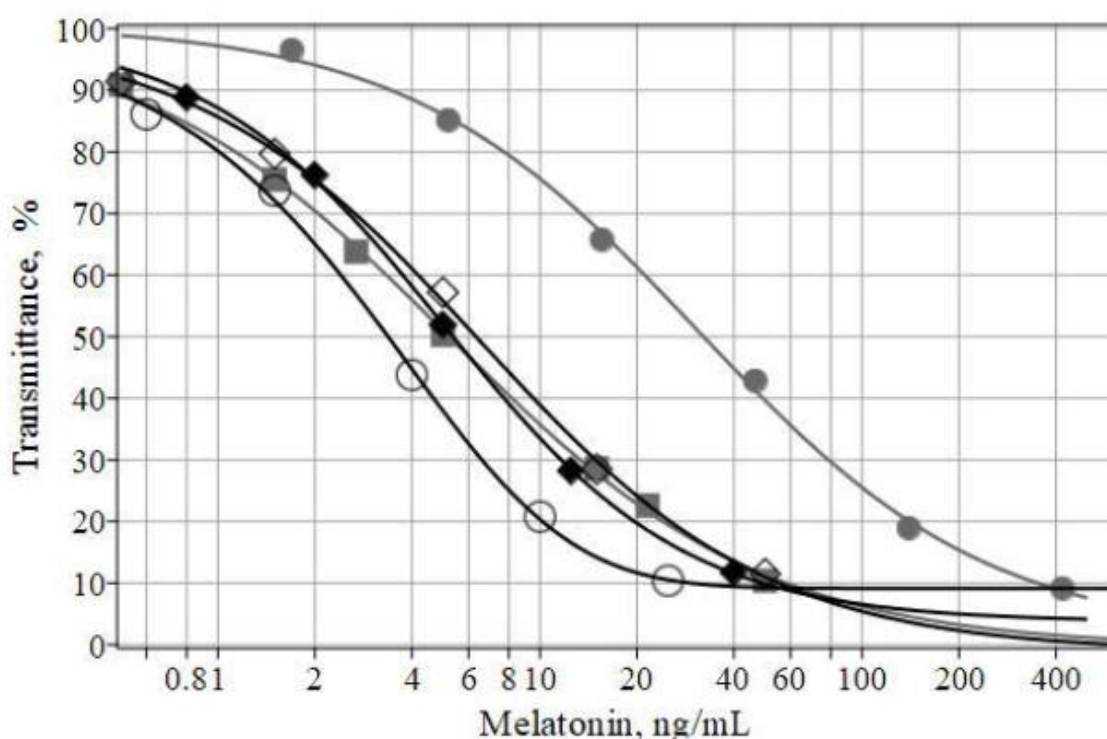
Рисунок 2.3 – Логістична крива Хілла з 4 параметрами

Рівняння Хілла може відповідати як спадаючим, так і зростаючим залежностям. Це залежить від знаку параметра нахилу Хілла ($d \neq 0$). Очікувався спадний тренд, а отже, і негативний нахил [89].

Для обчислення параметрів Хілла існують різні методи [90]. У цьому дослідженні використовується методи і можливості програмного пакета «Statistics» для Maple 2022.

У лабораторних дослідженнях було використано лише декілька різних калібраторів, як зазначено вище. Два з них описані в [84–88], інші – у [88; 91; 92]. Калібратори, описані в [91; 92], були протестовані методом RIA, тоді як інші – методом ELISA [84–88; 93].

На рис. 2.4 показано кілька графіків «доза-відповідь». Оптична пропускність аналізованого зразка (B/B_0) використовується тут в якості відповіді. Відповідно до цього графіка, усі калібрувальні криві мають спадний характер ($d < 0$, як і прогнозувалось).



сірі кола представляють дані [84–87; 94];

чорні ромби – дані [88];

сірі квадрати – дані [92];

порожні кола – дані [93];

порожні ромби – дані [91]

Рисунок 2.4 – Криві «доза-відповідь»

У табл. 2.1 наведено повний набір обчислених параметрів Хілла для кожної калібрувальної кривої (тобто кожного окремого калібратора).

Таблиця 2.1 – Параметри Хілла

Метод	<i>A</i>	<i>B</i>	<i>c</i>	<i>d</i>	Джерело даних
ELISA	1,3	100,7	30,9	−0,96	[84–87; 94]
ELISA	3,7	99,7	5,1	−1,17	[88]
RIA	−0,5	100,7	5,2	−0,89	[92]
ELISA 2019	2,7	99,7	5,2	−1,03	[93]
RIA	−1,1	99,5	6,5	−0,98	[91]
Середнє значення	2,3	100,1	10,6	−1,00	
Стандартне відхилення	4,1	0,6	11,4	0,10	

Рис. 2.4 та табл. 2.1 вказують на помітні розбіжності між даними різного походження. Порівнюючи дані з першого рядка табл. 2.1 з даними з інших рядків, параметр *c* (відомий також, як EC50, ED50, LD50, IC50 тощо) є головною причиною цих розбіжностей. Його значення приблизно в шість разів більше для даних із першого рядка, ніж для інших. Це виглядає «надмірно» для методик, які вважалися точними. Причини такого розходження поки що важко пояснити.

Параметри *B* та *d* краще узгоджуються між даними різного походження. Їх значення мають передбачувані величини: *B* наближається до 100, а *d* має негативне значення. Проте для останніх даних [93] цей коефіцієнт відрізняється від результатів [89], і ця розбіжність також видається несподіваною.

Слід звернути увагу на те, що калібратор і метод калібрування (ELISA) є однаковими для джерел [84–88; 94]. Водночас методики (ELISA і RIA), а також набори калібраторів є різними для даних [88; 91; 92]. Незмінним у цьому випадку залишається походження з однієї лабораторії. Результати одного походження мають хорошу узгодженість між собою.

Ухвалення клінічних рішень тісно залежить від походження даних. Це особливо важливо в IoT, оскільки багато медичних пристроїв можуть бути підключені як до локальної клінічної бази даних, так і до загальнонаціональної клінічної бази даних. Медичний персонал і пацієнти не повинні сумніватися у надійності вихідних даних та ефективності клінічних рішень. За таких умов

питання стандартизації інформації про пристрої, такі як їх надійність, безпека, походження та відповідність, мають супроводжувати медичні дані.

Саме тому кожен медик має враховувати походження використаних даних. Ця вимога має бути обов'язковою, незважаючи на реальну компетентність у галузі комп'ютерних і даних наук. Водночас оцінка надійності та менш надійних даних, а також походження даних є сферою експертизи для спеціалістів із медичних даних. Можна зробити висновок, що більш-менш надійні калібратори метаболіту мелатоніну в сечі тепер можливі на основі даних [88], [91], [92].

2.2 Настроюваний опис системи керування на основі FPGA для обробки даних з сенсорів

2.2.1 Постановка задачі обробки даних з сенсорів різного типу

Рішення Інтернету речей (Internet-of-Things, IoT) [95; 96] отримали значну увагу як з боку промислового, так і наукового секторів для застосування в різних сферах. Ринок пристроїв, що підтримують IoT, постійно зростає, і підключення до Інтернету стало де-факто стандартом для електронних пристроїв. Проте така ситуація також призводить до появи нових проблем і викликів для систем керування. Топології мереж на різних рівнях стають більш складними зі збільшенням кількості підключених пристроїв і технологій, які використовуються у процесі комунікації.

Одним із головних викликів для IoT-систем є їхня змінність з часом, коли з'являються нові вузли [96–99]. Більше того, такого роду зміни можуть впливати навіть на окремий вузол із кількома сенсорами. Розглядається мікроконтролерна система, яка може інтегрувати кілька модулів сенсорів і використовує пристрій на основі програмованої користувачем вентиляційної матриці (FPGA) [100] для зчитування даних із сенсорів, в тому числі – із сенсорів, що збирають медичні сигнали. Конфігурація та функціональність FPGA визначаються бінарним файлом конфігурації. Цей бінарний файл отримується в результаті послідовної обробки вихідного коду модуля за допомогою інструментів складання (компілятора, синтезатора, маршрутизатора тощо).

Додавання або видалення сенсора вимагає модифікації вихідного коду з повторним програмуванням модуля керування. Незважаючи на те, що ці модифікації можуть бути виконані за допомогою наявних сегментів коду, все одно передбачається редагування вихідних файлів у декількох місцях.

Тому пропонується підхід до організації опису вихідних файлів, який спрощує процедуру оновлення у випадках регулярного підключення сенсорів (коли до FPGA підключаються кілька сенсорів одного типу). Він базується на визначенні параметризованих користувацьких типів, які відповідають підключенню між сенсорами та FPGA, і використанні параметризованих інструкцій, які дозволяють генерувати масиви з'єднань.

Автоматизація проєктування для вбудованих систем останнім часом привертає значну увагу як дослідників, так і виробників [97–99; 101]. Це призводить до постійного вдосконалення інструментів розробки та появи нових. У цьому контексті можна відзначити, що навіть цілі бібліотеки для вбудованих систем можуть бути налаштовані та інтегровані в проєкт за допомогою цих інструментів. Однак постачальники інструментів не можуть передбачити всі можливі сценарії використання, тому це також стосується організації коду та самого опису системи. У цій статті ми пропонуємо багаторазово використовуваний фрагмент коду для опису параметризованого визначення підключення сенсорів і обробки даних у системі на базі FPGA.

Інтегровані середовища розробки (IDE) від постачальників FPGA зазвичай включають набір ядер інтелектуальної власності (IP-cores), які легко інтегруються в будь-який проєкт. IP-ядра значно скорочують час розробки, оскільки вони мають розширену підтримку з боку IDE, яка надає користувачеві варіанти конфігурації через графічні майстри. Незважаючи на те, що IDE бере на себе більшу частину відповідальності за інтеграцію IP-ядер, вона працює лише з окремими компонентами системи. У великих проєктах із складною архітектурною організацією [102–104] звичайною практикою є використання кількох екземплярів одного й того ж компонента. Додавання та видалення екземплярів компонентів є частими операціями в таких випадках [98; 101]. Однак

ці операції також призводять до змін у великих частинах коду. Тому потрібен новий підхід до цієї проблеми. У цій роботі запропоновано рішення, яке є параметризованим і базується на інструкціях мов опису апаратного забезпечення для створення необхідної кількості компонентів і відповідних з'єднань.

Хоча патерн «Фабрика» та інші шаблони проектування успішно застосовуються в об'єктно-орієнтованому програмуванні для створення нових об'єктів, семантика мов опису апаратного забезпечення вимагає іншого підходу до цієї проблеми. Це можна розглядати як опис «синього відбитку», де фактичні елементи будуть розміщені на пізніших етапах компіляції модуля. Таким чином, це рішення також можна вважати багаторазово використовуваним патерном для створення екземплярів модулів і їх налаштування. Водночас, попри значну увагу до проблеми розподілу функцій під час виконання [98], проблема контролю коду для FPGA все ще не отримує достатньої уваги.

2.2.2 Конфігурований опис системи керування сенсорами

Розглянута система складається з пристрою FPGA та набору сенсорів. Кількість сенсорів може змінюватися. Загальний вигляд архітектури може бути представлений стандартною схемою, де кілька сенсорів підключені до FPGA [105–107], як показано на рис. 2.5.

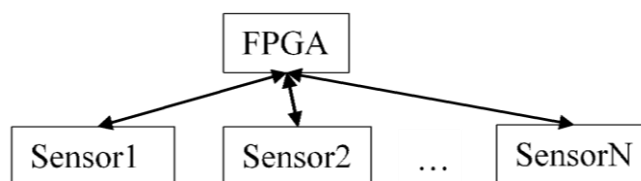


Рисунок 2.5 – Узагальнена система FPGA-сенсорної архітектури

Додавання нового сенсора до системи вимагає виконання наступних кроків:

- внесення змін до секції портів верхнього рівня щодо нового модуля;
- оголошення нових сигналів у секції архітектури;
- створення екземпляра модуля обробки із наступним зв'язуванням входів і виходів модуля із сигналами архітектури;

– модифікація секцій портів і архітектури для кожного модуля вищого рівня (цей крок є необов'язковим і залежить від структурної організації всієї системи);

– налаштування фізичних виводів FPGA відповідно до внесених змін.

За цих умов модифікація опису системи для обробки додаткового сенсора забирає значну кількість часу розробника.

Використаємо наступні нотації для подальшого огляду системи:

1) n_{sensor} – кількість сенсорів, що потрібно підключити;

2) $data_width$ – ширина даних отриманих з сенсорів;

3) $control_width$ – ширина лінії управління.

Список параметрів може бути збільшеним відповідно до особливостей інтерфейсу сенсорів.

Ці всі параметри об'єднані в окремий пакет p . Таким чином, для кожного типу сенсора $P = \{p_1, \dots, p_T\}; t = \underline{1}, T$, де T – номер типу сенсора, ми можемо мати набір пакетів. Використаємо поняття «пакета» у VHDL, оскільки розроблене рішення також доступне у VHDL. Конвертація в іншу мову опису апаратного забезпечення є досить простою, оскільки вони мають аналоги для пакетів VHDL.

У межах пакета мають бути визначені нові типи для представлення згаданих з'єднань. Перш за все, це тип, який використовує інформацію про n_{sensor} , $data_width$ та $control_width$ для представлення масиву шин. З точки зору реалізації, це масив масивів із наперед визначеною кількістю елементів. Будучи визначеним у тому ж пакеті, що й константні значення, цей тип інтерфейсу може бути налаштований шляхом зміни цих двох значень. Припустимо, що на цьому етапі модуль обробки вже реалізований і протестований із використанням реальних даних сенсора. Тому ті самі константні значення та типи з пакета можуть бути використані для опису модуля верхнього рівня. Усі шини повинні бути сумісними між собою, і найкращий спосіб забезпечити цю умову – використовувати той самий тип для визначення шин у компонентах.

Кожному підключеному сенсору відповідає екземпляр модуля обробки у FPGA. Ці екземпляри повинні генеруватися відповідно до значення n_{sensor} . Крім того, для збору даних із сенсора необхідно підключити не лише шини даних, але й сигнали керування (наприклад, скидання, вибір мікросхеми, перемикання даних/адрес тощо) і сервісні сигнали на стороні сенсора. Таким чином, необхідно генерувати входи та виходи зі сторони сенсора у фінальному описі.

Для подальшого пояснення, використаємо фрагмент коду, що наведено на рис. 2.6.

```
package sensor_reg_pack is
  constant DATA_LENGTH : integer := 32;
  constant generate_sensors : integer := 2;
  constant ADDR_LEN : integer := 6;
  constant generate_control : integer := 1;

  type inout_sensor_array is array (0 to (generate_sensors-1))
    of std_logic_vector((DATA_LENGTH-1) downto 0);
  type inout_array is array (0 to (generate_sensors-1))
    of std_logic_vector((generate_control-1) downto 0);
end sensor_reg_pack;
```

Рисунок 2.6 – Визначення пакету

Пакет містить два типи, які визначають з'єднання зі сторони сенсора та зі сторони обробки. Хоча вони визначені як масиви векторів значень типу `std_logic`, жодних обмежень на цю частину немає, і складні інтерфейси можуть бути визначені за допомогою масивів записів. Зверніть увагу, що типи не включають визначення напрямку сигналів. Ця частина повинна бути задана в декларації сутності (entity). Тому до пакета можуть бути додані додаткові типи для підключення всіх сигналів до сенсорів. Ми рекомендуємо для цієї частини виділяти окремий тип для кожного сигналу в інтерфейсі. Інший варіант – об'єднати всі входи в один тип запису (record type), а всі виходи – в інший. Це дозволяє розробнику визначити лише два типи масивів для з'єднання.

Константи в пакеті вказують, що до системи потрібно підключити два сенсори, а всі сигнали керування (тип `inout_array`) представлені масивом із шириною шини, що дорівнює 1. Незважаючи на те, що одиночний сигнал може бути представлений як значення типу `std_logic`, ми враховуємо ідею

конфігурованості, і для цього варіанту опису тип вектора залишається тим самим, змінюється лише кількість елементів у ньому. Решта констант описує ширину шини даних та адресної шини.

Сутність (entity), яка повинна створювати набір модулів, використовує згадані вище типи масивів для визначення своїх портів. Для цього вона також включає секцію загальних параметрів (generic), де сутність враховує параметри, початкові значення яких відповідають константам із пакета. Ці параметри передаються до generic map інстанційованих модулів. Секція портів модуля використовує типи з пакета для визначення входів і виходів. Приклад визначення сутності наведено на рис. 2.7.

```
library ieee;
use ieee.std_logic_1164.all;
use ieee.numeric_std.all;
use work.sensor_reg_pack.all;

entity top_smc is
  Generic (
    DATA_LENGTH : integer := DATA_LENGTH;
    generate_sensors : integer := generate_sensors;
    generate_control : integer := generate_control;
    ADDR_LEN : integer := ADDR_LEN
  );
  port (
    CLK: in std_logic;
    RST: in std_logic;
    start_work:
      in std_logic_vector((generate_control-1) downto 0);
    control: in inout_array;
    status: out inout_array;
    max_step: in inout_array;
    repeat_N: in inout_array;
    sensor: in inout_sensor_array;
    count_step: out inout_array;
    ready: out std_logic_vector((generate_control-1) downto 0);
    dir: out std_logic_vector((generate_control-1) downto 0);
    pwm_out: out std_logic_vector((generate_control-1) downto 0)
  );
end top_smc;
```

Рисунок 2.7 – Приклад визначення сутності

З рис. 2.7 легко помітити, що секція портів заповнена входами та виходами типу для керуючих сигналів. Усі вони є одно-бітними сигналами. Окрім цього, можна спостерігати, що модуль також містить загальні сигнали (generic), які є спільними для всіх модулів. Вони (наприклад, start_work, ready тощо) описуються за допомогою векторного типу, а не користувацько визначених

типів, оскільки ця операція виконується одночасно для всіх згенерованих модулів. Однак їх усе ще можна налаштувати, змінюючи лише верхню межу масиву, що є перевагою розробленої схеми опису.

У частині, відповідальній за генерацію, усі частини масиву з'єднуються із відповідним модулем. Індекс оператора `generate` вказує на індекс елемента масиву для організації з'єднання між сенсором і згенерованим модулем. Приклад коду для інстанціювання наведено на рис. 2.8.

```
GEN_control_smc_inst: for i in 0 to (generate_sensor-1) generate
  control_smc_inst : top_smc_base
  Generic map (
    DATA_LENGTH => DATA_LENGTH,
    generate_control => generate_control,
    ADDR_LEN => ADDR_LEN
  )
  port map (
    CLK => CLK,
    RST => RST,
    start_work => start_work(i),
    control => control(i),
    status => status(i),
    max_step => max_step(i),
    repeat_N => repeat_N(i),
    sensor => sensor(i),
    count_step => count_step(i),
    ready => ready(i),
    dir => dir(i),
    pwm_out => pwm_out(i)
  end generate GEN_control_smc_inst;
```

Рисунок 2.8 – Лістинг коду генерації модулів сенсорної обробки

Щоб уникнути будь-якої неоднозначності, пов'язаної з іменами зовнішніх та внутрішніх модулів, імена з секції портів внутрішнього модуля повністю відповідають іменам зовнішнього модуля. Вони оголошуються як звичайні вектори типу `std_logic`, на відміну від типів верхнього модуля. Таким чином, кожному окремому входу або виходу модуля призначається лише відповідна частина масиву верхнього модуля. Ця частина коду є достатньою для генерації необхідної кількості компонентів обробки сенсорів, і для цього потрібно змінити лише один параметр. Зміни в описі модуля нижчого рівня також можна застосувати, редагуючи цю частину коду опису.

Звичайною метрикою для оцінки кількості коду є метрика кількості рядків коду (Line of Code, LoC). У цій роботі ми припускаємо, що один LoC у файлі опису відповідає одному рядку інструкції або призначення. Таким чином, у описі інстанцій кожне відображення формального входу або виходу порту на сигнал

вважається одним LoC, хоча ця операція може бути записана лише в одному рядку. Наступні твердження зроблені для VHDL, однак їх можна адаптувати для інших мов. Введемо наступні позначення для формалізації цієї частини дослідження:

q_{entity} – кількість рядків коду необхідних для визначення сутності та її інстанціювання;

q_{port} – кількість рядків коду секції визначення портів;

$q_{generic}$ – кількість рядків коду секції визначення параметрів (generic).

Припускається, що $q_{entity} = q_{port} + q_{generic}$. Секції портів і параметрів треба розрізняти, оскільки компонент може бути інстанційований по-різному – із використанням секції generic або без неї (у такому разі застосовуються значення за замовчуванням). У разі явного підключення екземпляра модуля необхідно оголошувати відповідні сигнали в архітектурі. Відповідно до попереднього твердження, це призводить до збільшення кількості рядків коду. Насправді, з використанням прямолінійного підходу опис екземплярів модулів і сигналів для відображення споживає наступну кількість рядків коду (2.2):

$$S_{LoC} = n_{sensors} \cdot (q_{entity} + q_{port}) \quad (2.2)$$

Значення S_{LoC} можна зменшити до $2 \cdot n_{sensor} \cdot q_{port}$ якщо в процесі береться участь лише секція портів та типовим параметрам присвоюються стандартні значення. Додавання або видалення m модулів передбачає, що $m \cdot (q_{entity} + q_{port})$ повинні бути скореговані.

На відміну від попереднього рівняння, загальна кількість рядків коду у запропонованому підході не залежить від кількості інстанційованих модулів і залишається сталою для будь-якого заданого значення. Крім того, для керування кількістю модулів в описі потрібно редагувати лише один параметр, і всього три параметри необхідні для підтримки всього опису в контексті генерації модулів.

Таблиця 2.2 – Порівняння складності за метрикою рядків коду

Параметр	Прямий підхід	Запропонований підхід
Рядків коду для опису	$n \cdot (q_{entity} + q_{port})$	$q_{entity} + q_{port} + q_{generate} + q_{package}$
Рядків коду для оновлення, додавання або видалення m модулів	$m \cdot (q_{entity} + q_{port})$	1

В табл. 2.2 $q_{generate}$ визначає кількість рядків коду для інфраструктури генерування оператора та $q_{package}$ для визначення кількості строк коду в пакеті. Ці результати можуть бути узагальнені для випадку, коли в системі присутні T типів сенсорів, що додає множник T до рівняння. Проте легко помітити, що розроблений підхід все ще має перевагу завдяки меншій кількості рядків коду для опису та модифікації.

Подальшим розвитком цієї частини дослідження є автоматична генерація призначень для відображення логічних виводів модуля на фізичні. Цю задачу не можна залишати на програмні інструменти, оскільки вони автоматично генерують варіанти, які не підходять для фінальної схеми. Файл обмежень може бути змінений вручну або оновлений відповідно до конфігурації у файлі опису. Таким чином, інструмент для автоматичного призначення виводів буде затребуваним у поєднанні із запропонованим підходом.

Розроблену частину коду також можна інтегрувати у середовища розробки як фрагмент коду (snippet) для генерації екземплярів, що формують регулярну структуру в архітектурі модуля.

2.3 Метод графіків Пуанкаре для аналізу медичних сигналів: кількісний опис та обробка

2.3.1 Загальний огляд графіків Пуанкаре

Графіки Пуанкаре були названі на честь Генрі Пуанкаре, хоч він ніколи не був їх творцем або користувався ними. Цей відомий французький вчений створив

головну рекурсивну лему, яка підходить до багатьох динамічних систем з кінцевою кількістю енергії. Все ж, зв'язок поміж графіками Пуанкаре та цією теоремою достатньо близький.

Рекурентна теорема констатує [108; 109], що «якщо система має фіксовану загальну енергію, яка обмежує її динаміку обмеженими підмножинами її фазового простору, система зрештою повернеться настільки близько, наскільки вам захочеться, до будь-якого заданого початкового стану». До речі, К. Каратеодорі довів це лише через багато років [110].

Розглянемо таку систему. Скажімо, ми знаємо його стани в скінченному наборі дискретних проміжків часу. Рекурентний графік – це представлення квадратної матриці подібності системних станів де матричне значення дорівнює одиниці для таких пар станів які відповідають станам, що повторюються. Вони мають принаймні бути достатньо наближені одне до одного, в іншому випадку елементи матриці будуть рівні нулю. Таким чином точка на рекурентному графіку відповідає ненульовому значенню в цій матриці подібності. Прекрасний історичний огляд рекурентних графіків свідчить про те, що вивчення та використання даного інструменту візуалізації стало інтенсивним лише за останні три десятиліття [111].

Справа в тому, що графік Пуанкаре це певний тип рекурентного графіка [112]. Вважатимемо, що будь-які два послідовні значення часових серій (або декількох, як варіант) відповідають досить близьким станам системи. Це звучить досить добре, якщо частота дискретизації задовольняє іншу відому теорему. Тут ми маємо на увазі теорему підрахунку Найквіста-Шеннона. Тому така пара послідовних станів ставить крапку як і в рекурентному графіку, так і в графіку Пуанкаре. Можна вказувати значення серій на осях графіку Пуанкаре замість їх індексів в рекурентному графіку.

Уявимо часовий ряд довільного походження $\{s_n\}_1^N = s_1, s_2, \dots, s_n, \dots, s_N$, а також розглянемо всі пари послідовних значень цього часового ряду: $(s_1, s_2), (s_2, s_3), \dots, (s_n, s_{n+1}), \dots, (s_{N-1}, s_N)$. Якщо кожна з пар множини є точкою в двовимірному графіку, то ця множина матиме діаграму Пуанкаре. Такий графік

також називають картою повернення, точковою діаграмою або діаграмою Лоренца [112; 113].

Звідси виходить, що будь-який графік Пуанкаре є також точковою діаграмою. Точкові діаграми є визнаними чемпіонами в галузі наукової візуалізації [114; 115]. Статистика зазначає, що понад 70 % графіків в наукових статтях є саме точковими діаграмами [114].

І саме тому аналіз графіків Пуанкаре був використаний як наочний засіб. Попередні статті, які використовували графіки Пуанкаре для дослідження мінливості серцевого ритму, були і залишаються прикладами даного підходу. Сучасні статті використовують схожий шлях, але більш тонкі методи як в [116].

Кількісне визначення графіків Пуанкаре було розроблено достатньо давно [45; 117; 118]. Числові дескриптори ($SD1$ та $SD2$) запропоновані авторами робіт [45; 118] досі використовуються. Обидва були представлені, відштовхуючись від приблизно еліптичної форми типового графіка Пуанкаре. Цей метод було названо як «метод наближення до еліпсу». Він зазнав критики та отримав певне оновлення, але його регулярна основа залишалась незмінною.

Короткострокова мінливість в серцевому ритмі пов'язана з $SD1$, з другорядною віссю наближеного еліпсу. Тоді як довгострокова мінливість пов'язана з $SD2$, з довшою віссю, і їх співвідношення ($SD1/SD2$) служить як міра випадковості серій.

Так звана «лінія ідентичності» відіграє значну роль в даній структурі. Це є бісектрисою проєктованої площини. Точки, що задовольняють умову $x_n = x_{n+1}$, належать до цієї лінії. Головне припущення теорії «методу наближення до еліпсу» полягає в тому, що одна з осей «наближеного еліпсу» колись наблизиться до цієї лінії.

Хоча варто зазначити, що типовий графік Пуанкаре не завжди матиме еліптичну форму. Інші форми, наприклад, названі як «кометна форма» часто можна зустріти в дослідженнях мінливості серцевого ритму [112; 113].

Поширення методу «еліптичного підлаштування» до інших медичних сигналів відбувається відносно автоматично [44]. Тим часом автори [119]

попереджують про можливе неправильне тлумачення вищезгаданих дескрипторів поза дослідженнями мінливості серцевих ритмів. Особливо це стосується «короткострокових» та «довгострокових» визначень.

Якщо хтось, знаючи метод головних компонент, вивчає техніку «наближення до еліпсу», то він отримає відчуття «дежавю». Математична інтуїція підказує, що мова йде про спрощену версію відомого та потужного математичного інструменту. Ми маємо на увазі тут метод головних компонентів (МГК) як основну теорію [120].

Отже, доцільно переформулювати кількісне визначення графіків Пуанкаре в рамках МГК. По-перше, тому що це не було зроблене раніше, а по-друге, сподіваючись краще зрозуміти кількісну оцінку та обробку графіків Пуанкаре. Ми також порівняємо відпрацьовані тут результати МГК з висновками «методу наближення до еліпсу». Для цього ми будемо використовувати добре перевірені дані електроміограм (ЕМГ) [121].

Розглянемо знову серію $\{s_n\}_1^N = s_1, s_2, \dots, s_n, \dots, s_N$. Нехай послідовність центрована на серійному середньому тут і далі. Це не є обов'язковою умовою якщо хтось має намір побудувати графік Пуанкаре для візуального оцінення. Проте дескриптори ($SD1$, $SD2$) розраховуються як квадратний корінь емпіричних дисперсій уздовж «осей еліпсу» лише для центрованих даних [43], [45], [122].

Цей ряд можна розглядати як серію N точок (членів) в одновимірному просторі або як точку в N -вимірному просторі. Тоді виникає питання, чи можна вбудувати цю серію в простір з проміжними розмірами від 1 до N ?

Оберемо $1 \leq k \leq N$ та поділимо серії на відстаючі вектори $N - k + 1$, що мають однакову довжину k : $(s_1, s_2, \dots, s_k), (s_2, s_3, \dots, s_{k+1}), \dots, (s_{N-k+1}, s_{N-k+2}, \dots, s_N)$ [123; 124]. Ми можемо розглядати кожен з відстаючих векторів як точку в k -вимірному просторі. Тоді множина таких точок є вбудовування в серії у k -вимірному просторі. Погодимося, що відсталі вектори утворюють ряд прямокутних матриць з $N - k + 1$ рядками та k стовпцями:

$$D_k = \begin{pmatrix} s_1 & s_2 & \dots & s_k \\ s_2 & s_3 & \dots & s_{k+1} \\ \dots & \dots & \dots & \dots \\ s_{N-k+1} & s_{N-k+2} & \dots & s_N \end{pmatrix}. \quad (2.3)$$

Автори [120] називають масив (2.3) як «матриця даних», тоді як [123], [124], ця ж матриця названа як «матриця траєкторій». Одностовпчикова матриця даних відповідає тривіальному вбудовуванню в одновимірний простір, який зазначено вище, тоді як однорядкова матриця описує одну точку в N -вимірному просторі.

Корисність визначає параметр k («довжина вікна» [123; 124]). Зазвичай $k = 2$, і графік Пуанкаре – це двовимірний графік, хоч є і пропозиції щодо тривимірних графіків $k = 3$) [6], [122], [125]. Матриця даних для звичайних графіків Пуанкаре має дві відстаючі вектори-стовпці:

$$D_2 = \begin{pmatrix} s_1 & s_2 \\ s_2 & s_3 \\ \dots & \dots \\ s_{N-1} & s_N \end{pmatrix}. \quad (2.4)$$

Таким чином, матриця даних (2.4) представляє звичний графік Пуанкаре або вбудований часовий ряд в двовимірний простір (площину проєкції). Кожен рядок матриці даних дає точку до графіка Пуанкаре. Ця матриця даних має Ганкелевий тип, та ранг рівний «2», який відображає розмірність проєктованої площини [120], [123].

Таке ж масштабування колонок векторів є другою істотною вимогою для МГК. Вона виконується сама собою оскільки відстаючі вектори-колонки матриць (2.3) та (2.4) належать до одного ряду.

2.3.2 Головні компоненти на проєкційній площині

МГК є потужним методом для вивчення моделей даних [126]. Існує чотири способи визначення основних задач МГК [120]. Виділимо одну з них, яка виглядає підходящою для наших цілей. Зауважимо, що векторні стовпці матриці

даних (2.4) добре співвідносяться. Вони відстають одне від одного але являють собою один і той же ряд.

Знайдемо такий ортогональний базис, в якому коваріаційна матриця для матриці даних (2.4) є діагональною. Вектори-стовпці трансформованої матриці даних мають нульову кореляцію в цьому базису. Цей базис складається з головних компонентів (ГК) [120].

Окрім того, МГК гарантує ще три таких наслідки:

- по-перше, сума квадратів відстаней від точок до їх ортогональних проєкцій мінімальна;
- по-друге, кількість емпіричних дисперсій вздовж головних компонентів є максимальною;
- по-третє, значення квадрату відстані від точки до точки між ортогональними проєкціями точок даних на площину є максимальним [120].

Розглянемо центровані середні значення, обчислені для кожного стовпця, вектор-стовпці матриці даних D_2 . Нехай a та b є такими векторними стовпцями а $D_{2,c}$ є центрованою матрицею даних. Таким чином, задача пошуку основних компонентів відповідає задачі діагоналізації такої матриці коваріації [6; 94]:

$$Cov_2 = \left(\frac{1}{N-1}\right) D_{2,c}^T D_{2,c} = \left(\frac{1}{N-1}\right) (a^2 (a, b) (a, b) b^2), \quad (2.5)$$

де верхній індекс «Т» означає «транспонований» і (a, b) є скалярним добутком із центрованими векторними стовпцями.

Коваріаційна матриця (2.5) є симетричною та двовимірною квадратною матрицею в цьому випадку. Процедура її діагоналізації добре відома [126]. Власні значення коваріаційної матриці є дисперсіями вздовж головних компонент, а власні вектори дають напрямки обом головним компонентам.

Іншим методом рішення є сингулярний розклад значень (СРЗ) для центрованої матриці даних [120]. Цей спосіб віддає перевагу з обчислювальної точки зору здебільшого для «довгих вікон» ($k > 2$) [123], [124]. Сингулярні

значення є стандартними відхиленнями вздовж головних компонент, а праві сингулярні вектори показують їх напрямки.

Діагоналізація матриці коваріації можлива за допомогою обертання системи координат матрицею Гівенса [127]:

$$GG = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}. \quad (2.6)$$

Ця ортогональна матриця перетворює матрицю коваріації в діагональну в такому значенні:

$$G^{-1} \cdot Cov_2 \cdot G = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}, \quad (2.7)$$

де λ_1, λ_2 є власними значення коваріаційної матриці в порядку зменшення. Вони прямо визначають стандартні дескриптори:

$$\sqrt{\lambda_1} = SD2 \quad \sqrt{\lambda_2} = SD1 \}. \quad (2.8)$$

Один можна пов'язати з кутом повороту ($0 \leq \theta \leq \pi$) з умовами коваріаційної матриці (2.5) і, таким чином, з обома векторними стовпцями матриці даних [99]:

$$\theta = \frac{1}{2} \arctan \left(\frac{2(a,b)}{a^2 - b^2} \right). \quad (2.9)$$

Знак аргументу арктангенса визначається, якщо $0 \leq \theta \leq \frac{\pi}{4}$ або $\frac{\pi}{4} \leq \theta \leq \frac{\pi}{2}$. Обертання системи координат t на кут (2.9) забезпечує не тільки діагональну форму коваріаційної матриці. Він також декорелює стовпці в оберненій матриці даних. Окрім того, поворот до головних компонент також забезпечує ще три результати, перелічені вище.

Суворі нерівності, яка часто виконується для типового графіка Пуанкаре:

$$p = \left| \frac{2(a,b)}{a^2 - b^2} \right| \gg 1. \quad (2.10)$$

Тим не менш, її перевірка є бажанням для кожної матриці даних. Тільки якщо умова (2.10) є актуальною, тоді ми маємо $\theta \approx \frac{\pi}{4}$. Тільки якщо так, напрям першого основного компонента практично збігається з «лінією тотожності» на проєктованій площині. В іншому випадку довша «вісь еліпсу» та «лінія тотожності» можуть бути по різному спрямовані.

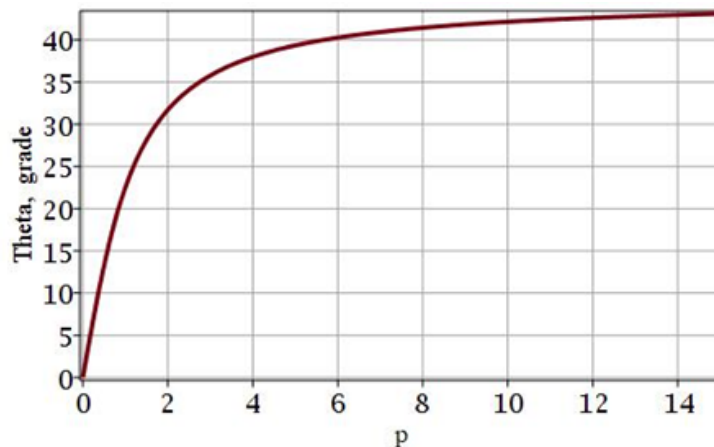


Рисунок 2.9 – Кут повороту (θ) в залежності від безрозмірного параметру (p)

Головним припущенням в рамках «методу наближення до еліпсу» ($\theta \approx 45^\circ$) звучить більш-менш обґрунтовано, лише якщо цей параметр досить великий: $p \geq 7$

Отже, порівняння (2.10) є неявним, але обов'язковим для виконання «методу наближення до еліпсу». І чим більше відповідає (2.10), тим виразнішим є наближення до «наближеного еліпсу». Це має стосуватися як дескрипторів ($SD1, SD2$), так і умов матриці обертання (2.6).

Можна також використовувати сингулярне векторне розкладання матриці даних (2.4) [120]. Тоді дескриптори дорівнюють сингулярним значенням, поки як правий сингулярний вектор дає матрицю обертання (2.9).

Якщо хтось використовує «метод наближення до еліпсу», то він/вона має сподіватися, що матриця даних, повернута до «лінії ідентичності», має слабо корельовані стовпці. Якщо так, то порівняння (2.10) виконується, а дескриптори дорівнюють стандартним відхиленням для кожного стовпця обертової матриці даних [46].

2.3.3 Розкладання та фільтрація графіків Пуанкаре

Розкладанню та фільтруванню графіків Пуанкаре присвячено такі роботи, як [46; 128–130]. Останні три з них використовували сингулярне розкладання (СРЗ) матриці даних. Таке розкладання графіків Пуанкаре означає наступне – матриця даних з рангом, рівним 2, розкладається в лінійній комбінації двох масивів з однаковим рангом. Для цього необхідно знати сингулярні значення та сингулярні вектори матриці даних [128–130].

Алгоритм такої фільтрації охоплює чотири етапи:

- 1) СРЗ матриці даних з пошуком їх сингулярних векторів та значень;
- 2) розкладання матриці даних (2.4) з пошуком її сингулярних векторів та значень;
- 3) протидіагональне усереднення (ганкелізація) обох однорангових матриць [96].
- 4) децимація («ліниве вейвлет-перетворення» [131]) обох частин відновлених серій.

Результат фільтрації є таким же, як якщо б було застосовано цифровий фільтр з кінцевою імпульсною характеристикою [129]. Таким чином, можна відокремити низькочастотну частину сигналу від високочастотної. Низькочастотна (НЧ) складова ряду пов'язана з більш високим сингулярним значенням матриці даних. Незначне сингулярне значення задає частину високої частоти (ВЧ).

Можна, скоріш за все, розглядати високочастотну частину як адитивний шум, особливо якщо значення одиничні значення різко відрізняються. Отже, фільтрація повинна бути легкою для прибирання шуму сигналу.

Пропонується використовувати тут ще один спосіб фільтрації, пов'язаний з головними компонентами та МГК. Припустимо, що умова (2.10) виконується, тоді звідси значення кута повороту (2.9) наближено до 45 градусів. Тому матриця обертання має такий спрощений вигляд:

$$\mathbf{G}_{45} = \sqrt{2} \begin{pmatrix} \frac{1}{2} & -\frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{pmatrix}. \quad (2.11)$$

Розглянемо перетворення матриці даних (2.4) з обертовою матрицею (2.11) на нові вісі. Ці осі збігаються з «лінією ідентичності» та нормальною лінією до неї. Матриця даних набуває такої унікальної форми:

$$\mathbf{D}_{2,\text{new}} = \sqrt{2} \begin{pmatrix} \frac{s_1 + s_2}{2} & \frac{s_2 - s_1}{2} \\ \frac{s_2 + s_3}{2} & \frac{s_3 - s_2}{2} \\ \dots & \dots \\ \frac{s_{N-1} + s_N}{2} & \frac{s_N - s_{N-1}}{2} \end{pmatrix}. \quad (2.12)$$

Рядки матриці обертання (2.11) пропорційні цифровим фільтрам Хаара з коефіцієнтом масштабування $\sqrt{2}$. Тут можна побачити два цифрових фільтри Хаара:

- 1) діапазон високих частот (перший рядок) або ВЧ-фільтр;
- 2) низькочастотна смуга (другий ряд) або НЧ-фільтр [132; 133].

У стовпцях матриці (2.12) представлені коефіцієнти вейвлетів Хаара для стартового ряду, якщо не звертати уваги на вищезазначений коефіцієнт масштабування. Перший стовпчик пропорційний так званим коефіцієнтам наближення, пов'язаним з НЧ-фільтром. Другий стовпець пов'язаний з ВЧ-фільтром і показує так звані коефіцієнти деталізації.

Тому поворот (2.11) до «лінії ідентичності», що є обов'язковою дією в «методі наближення до еліпсу», одночасно забезпечує декомпозицію вихідних даних. Якщо відкинути коефіцієнт масштабування навпроти (2.12) і зробити зменшене зразки для обох стовпців, тоді це буде фільтрування. Результати будуть такими ж, як і фільтрація оригінальної серії з вейвлетами Хаара. Таким чином, стовпці матриці (2.12) містять як низькочастотні-, так і високочастотні-частини початкового сигналу, і обидва також є майже статистично незалежними.

Таким чином, рекомендується використати два ідентичні, в залежності від результату, шляхи фільтрації графіків Пуанкаре:

- дворазова децимація («ліниве вейвлет-перетворення») матриці даних (2.12) а далі розгляд децимаційних масивів як низькочастотних так і високочастотних частин графіків Пуанкаре відповідно;

- інакше перетворення оригінальних рядів вейвлетами Хаара, а потім побудова матриці даних відфільтрованого графіка Пуанкаре з низькочастотних та високочастотних частин як колонок, які обидва мають бути помножені на коефіцієнт масштабування $\sqrt{2}$.

Обидва алгоритми помітно простіші за запропоновані в [128–130] та описані на початку цього підрозділу.

Зауважимо, що умова (2.10), якщо вона виконується, забезпечує практично нульову кореляцію для стовпців матриці (2.12). Це важливо, тому що дескриптори всередині «методу наближення до еліпсу» – це стандартні відхилення стовпців матриці (2.12), без пониження, але з урахуванням коефіцієнта масштабування [46]. Це означає, що вони повинні бути в основному досить близькими до точних значень (2.8) від МГК. Незважаючи на кут повороту (2.9), він може помітно відрізнятися від 45 градусів. Цей приклад розглянемо нижче.

Однак виникає питання. Що робити, якщо коефіцієнт кореляції стовпців для матриці (2.12) неприпустимо високий? Тоді потрібно використовувати для фільтрації матриці обертання (2.6) з реальним кутом обертання (2.8) замість (2.11). Звичайно, не можна зв'язати вже таку фільтрацію з вейвлетами Хаара. Тим не менш, добуток матриці даних (2.4) на матриці обертання (2.6) містить дві строго незалежні колонки, що відповідають НЧ- та ВЧ-частинам сигналу.

Можна стверджувати, базуючись на попередній дискусії, що дескриптори – це міри розсіювання для низькочастотних та високочастотних частин початкового сигналу. Безумовно НЧ та ВЧ діапазони залежать від частоти дискретизації.

2.3.4 Фрактальна розмірність графіків Пуанкаре у якості кількісної міри

Використання графіків Пуанкаре в медицині історично почалось з вивченням серцевого ритму. Закон масштабування кардіоінтервалів було згадано в тезах Т. Макікалло [117]. Пізніше, фрактальні властивості R-R інтервалів серцевого ритму були обговорені в багатьох статтях. Читач може перейти, наприклад, до [5; 134–137]. Фрактальний розмір Гігучі для кардіоінтервалів може змінюватися, наприклад, через фізіологічну активність пацієнта [136]. Це також може залежати від його діагнозу [134; 138].

Спочатку фрактальні властивості були пов'язані саме із серією кардіоінтервалів [117; 134–138]. Методи, які були використані в цих роботах, розглянуті в [135]. Графік Пуанкаре є серед цих інструментів, але лише завдяки його дескрипторам. Фрактальний характер був пов'язаний з рядами первинних частот серцевих скорочень, але не з самими графіком Пуанкаре.

Навпаки, вважається що фрактальний характер притаманний іншим медичним сигналам, і їх графік Пуанкаре має відображати це [5; 7]. Закон масштабування встановлює фрактальний характер геометричного об'єкта як графіка Пуанкаре [139; 140]. «Метод підрахунку коробок» [139] виявився зручним інструментом для вивчення фрактальної розмірності графіка Пуанкаре [5; 7]. Більше того, цей метод легко програмується [139].

Розглянемо типовий графік Пуанкаре як набір («хмару») точок на проєктованій площині. Точки групуються навколо напрямку до першого принципового компоненту, або до «лінії ідентичності», якщо вони збігаються. Сформулюємо запит. Скільки плоских «коробок» потрібно для повного покриття хмари точок? Звичайно, це число залежить від лінійного розміру «коробки» (a), а саме від $N(a)$. За умови, що кожне поле охоплює хоча одну точку, можна отримати такий закон масштабування [140]:

$$N(a) \sim a^{-d} . \quad (2.13)$$

Тут d є фрактальною розмірністю, яка також відома як розмірність Мінковського-Булліганда або розмірність Колмогорова [140]. Закон масштабування (2.13) має вигляд прямої лінії в подвійних логічних координатах [46].

Цей безрозмірний параметр (d) пропонується використовувати як показник кількості для графіків Пуанкаре медичних сигналів [5], [7]. Нехай закон масштабування є правильним у деякому діапазоні шкал. Тоді графік Пуанкаре має фрактальну природу (самоподібний) на однакових розмірах. Це означає, що фрагмент або частина графіка Пуанкаре статистично схожа на цілу.

Чому це важливо? Якщо графік Пуанкаре самоподібний, то його низькочастотна частина, яка є відфільтрованим графіком, зберігає всі основні риси цільного графіка. Це означає збереження дескрипторів, фрактального розміру, середнього значення та форми графіка Пуанкаре. Тим часом кількість точок та орієнтація графіка змінюватимуться.

2.4 Методи обробки та аналізу кноїдних та усамітнених пульсових хвиль (солітонів) в крові

2.4.1 Вступ, теорія та методи

Рівняння Кортевега-де-Фріза (KdV) [141], [142], а також його модифікована форма (mKdV) та перетворення Міури, яке пов'язано з ним [143], широко використовуються для однорівневого моделювання хвиль у рідині [141–150]. Зокрема, це можуть бути хвилі в крові [145–150]. Вони зручні тим, що розраховані на довгі хвилі в нелінійних середовищах зі слабкою дисперсією [145; 149].

Якщо говорити точніше, то тут матимемо справу з їхніми точними рішеннями типу біжучої хвилі. Вони включають кноїдальні, а також одиночні хвилі в кровотоці [146]. Для цього в дисертаційній роботі використано систему комп'ютерної математики Maple [151].

Слід зазначити, що ці рівняння є коректними математичними моделями лише за деяких додаткових умов. Зокрема, використовувався нелінійний закон Гука для стінок кровоносних судин [144; 150]. Підрахунок квадратичних поправок веде до KdV [144; 150], а кубічних, до того ж, до моделі mKdv [150].

Нехтування середніми силами опору та процесами дисипації є необхідними умовами для моделей KdV або mKdV. Визначальним фактором тут є еластичні властивості стінок судин [146; 150].

Ще однією необхідною вимогою до пульсових хвиль є періодичність, навіть нерегулярна [146–148]. Тоді кноїдні хвилі повинні бути першими в полі зору. Однак вони дуже подібні до одиночних хвиль, якщо еліптичний модуль прагне до максимального значення ($m \rightarrow 1$) [146]. Отже, ми виділимо лише ті кноїдальні хвилі, які в даному випадку забезпечують правильну межу (солітон).

Крім того, автори [144; 150] вказали на обмежені часові діапазони, в яких обидві моделі є правильними. Ці діапазони відповідають одній із ранніх стадій серцевого циклу [150]. Схоже, що це систола [146].

З іншого боку, люди шукають нові точні рішення для обох рівнянь. Нині серед них є практично всі прямі та зворотні еліптичні функції Якобі, а також їх комбінації. Не всі вони підходять для моделювання кровотоку. Крім того, вони не завжди стабільні і це так звані хвилі-шахраї [152–154].

Таким чином, основною метою цієї статті є застосування математики Maple для дослідження точних розв'язків рівнянь KdV та mKdV. Ми збираємося підібрати відповідні, стабільні рішення з правильними обмеженнями при максимальному/мінімальному значенні еліптичного модуля. Одним із цих обмежень має бути, звичайно, дзвоноподібний солітон KdV.

Розглянемо початкові рівняння як:

$$u_t + au_x + \mu_2 u(x, t)u_x + \mu_3 u_{xxx} = 0 ; \quad (2.14)$$

$$u_t + au_x + \mu_2 u(x, t)^2 u_x + \mu_3 u_{xxx} = 0, \quad (2.15)$$

де (2.14) – рівнянням KdV, а (2.15) – рівнянням mKdV;

$u(x, t)$ – релевантна невідома функція (розв'язок);

a – швидкість хвилі (швидкість Кортвега-Моенса);

μ_2, μ_3 – константи, які визначають нелінійні та дисперсні властивості середовища. Нижні індекси показують похідні відповідних порядків.

Рівняння (2.14) та (2.15) зазвичай спрощуються за допомогою елементарного перетворення:

$$\{t = \tau, x = z + a\tau, u(x, t) = \vartheta(z, \tau)\}. \quad (2.16)$$

Тоді рівняння (2.4.1, 2.4.2) приймуть простіші форми з новими змінними:

$$\vartheta_\tau + \mu_2 \vartheta(z, \tau) \vartheta_z + \mu_3 \vartheta_{zzz} = 0; \quad (2.17)$$

$$\vartheta_\tau + \mu_2 \vartheta(z, \tau)^2 \vartheta_z + \mu_3 \vartheta_{zzz} = 0. \quad (2.18)$$

Безвимірність є корисним способом представити математичну модель у найбільш універсальний спосіб. Така модель є однаковою незалежно від того, яка система одиниць використовується. Іншим результатом корисності є мінімальна кількість змінних та/або параметрів моделі. Застосуємо трохи складніше масштабування до (2.14) дотримуючись [149]:

$$\left\{ t = \sqrt{\frac{\mu_3}{a^3}} \tau, x = \sqrt{\frac{\mu_3}{a}} z + a \sqrt{\frac{\mu_3}{a^3}} \tau, u(x, t) = \frac{6\vartheta(z, \tau)}{\mu_2} \right\} \quad (2.19)$$

Коефіцієнт «6» в останньому співвідношенні є «справою смаку». Зазвичай він часто вибирається як «1» [149]. Рекомендується використовувати ті самі нові змінні з дещо зміненим третім співвідношенням для рівняння (2.15):

$$\left\{ t = \sqrt{\frac{\mu_3}{a^3}} \tau, x = \sqrt{\frac{\mu_3}{a}} z + a \sqrt{\frac{\mu_3}{a^3}} \tau, u(x, t) = \frac{\sqrt{6}\vartheta(z, \tau)}{\sqrt{\mu_2}} \right\} \quad (2.20)$$

Треба зазначити, що перетворення (2.14, 2.19, 2.20) або ті, що їм зворотні, можуть з легкістю виконуватися за допомогою методу «dchange» бібліотеки «PDEtools» середовища комп'ютерної математики Maple. Таким чином можна отримати безвимірні форми рівнянь (2.14) та (2.15):

$$\vartheta_{zzz} + 6\vartheta(z, \tau)\vartheta_z + \vartheta_\tau = 0 \quad (2.21)$$

Це рівняння демонструє KdV-модель в безвимірній формі, а

$$\vartheta_{zzz} + 6\vartheta(z, \tau)^2 \vartheta_z + \vartheta_\tau = 0 \quad (2.22)$$

демонструє mKdV-модель.

Обидві безвимірні форми (2.21, 2.22) використовуються в гемодинаміці. Аналогічне змінення масштабу та така сама форма для безрозмірного KdV (2.21) використані в [144; 149]. Зокрема, розповсюдження солітону KdV у трубці з тонкими еластичними стінками після миттєвого впорскування ударного об'єму крові розглядалося в [144]. Безрозмірне рівняння KdV, ідентичне (2.21), отримано з закону збереження імпульсу в кровотоці.

2.4.2 Результати застосування

Буде виконано пошук тільки рішень канонічних рівнянь (2.18) та (2.19). Основні вимоги є такими:

- рішення мають бути нетривіальними, з реальними аргументами та параметрами,
- рішення не повинні мати сингулярності;
- рішення повинні являти собою біжучі хвилі, періодичні в часі та просторі;
- рішення повинні мати реальні ненульові межі при мінімальному та максимальному значенні еліптичного модуля ($m \rightarrow 0$ та $m \rightarrow 1$).

Пошук рішень типу біжучої хвилі (TWS) означає перехід до єдиної фазової змінної (φ), яка лінійно залежить від обох апріорних змінних (z, t):

$$\varphi = kz + \omega t + \varphi_0. \quad (2.23)$$

Константи k, ω, ω_0 мають тут звичайне значення для біжучої хвилі: безрозмірне хвильове число, циклічна частота та початкова фаза.

За допомогою Maple можна отримати кілька можливих біжучих хвиль, за допомогою спеціального методу «TWSolutions» бібліотеки «PDEtools». Це

біжучі хвилі з фазою (2.23), які побудовані на основі наступного набору еліптичних функцій Якобі:

$$\{sn(\varphi, m), cn(\varphi, m), dn(\varphi, m), ns(\varphi, m), nc(\varphi, m), nd(\varphi, m)\}, \quad (2.24)$$

де $0 \leq m \leq 1$ є еліптичним модулем.

Дві функції цього списку мають сингулярності ($nc(\varphi, m)$ та $ns(\varphi, m)$). Ще одна з них ($nd(\varphi, m)$) не має необхідного граничного значення за умови ($m \rightarrow 1$). Відповідно результати мають такі форми:

$$\left. \begin{aligned} w_1 &= 2m^2 k^2 cn(\varphi, m)^2 + \frac{4k^3(1-2m^2)-\omega}{6k} \\ w_2 &= -2m^2 k^2 sn(\varphi, m)^2 + \frac{4k^3(m^2-2)-\omega}{6k} \\ w_3 &= 2k^2 dn(\varphi, m)^2 + \frac{4k^3(m^2+1)-\omega}{6k} \end{aligned} \right\}. \quad (2.25)$$

На перший погляд ці відповідні рішення здаються різними. Однак для квадратів цих трьох еліптичних функцій існують добре відомі співвідношення:

$$\begin{aligned} cn(\varphi, m)^2 &= 1 - sn(\varphi, m)^2; \\ dn(\varphi, m)^2 &= 1 - m^2 sn(\varphi, m)^2; \end{aligned} \quad (2.26)$$

Застосування цих відношень до хвиль (2.25) дозволяє довести їх еквівалентність. Таким чином, це одна й та сама біжуча хвиля, і нижній індекс можна опустити.

$$w(\varphi, m) = 2m^2 k^2 cn(\varphi, m)^2 - \frac{4k^2(2m^2-1)-\omega}{6k} \quad (2.27)$$

Розглянемо межу солітонів, визначивши $m = 1$ для цього в (2.27). Косинус Якобі перетворюється на гіперболічний секанс, і ми маємо для солітонного рішення:

$$w(\varphi, 1) = 2k^2 \sec^2 h(\varphi) - \frac{4k^3-\omega}{6k} \quad (2.28)$$

Константна частина (2.28), очевидно, повинна дорівнювати нулю тому що солітони повинні зникати далеко за межами своїх максимум (ці максимуми

знаходяться біля $\varphi = 0$). Отже, ми маємо визначене співвідношення між безвимірною частотою та хвильовим числом у цьому випадку:

$$\omega = 4k^3. \quad (2.29)$$

Цей солітон (2.28) має амплітуду, яка дорівнює $A = 2k^2$ та має фазове прискорення $\frac{\omega}{k} \sim k^2 \sim A$, пропорційне до амплітуди. Це характерно для солітонів KdV [2].

Візьмемо амплітуду солітону рівною нулю: $A = 1$. Тоді $k = \frac{1}{\sqrt{2}}$ та кноїдна хвиля (2.27) є функцією лише фазової змінної та еліптичного модуля з урахуванням (2.29):

$$w(\varphi, m) = m^2 \operatorname{cn}(\varphi, m) - \frac{2}{3}(m^2 - 1). \quad (2.30)$$

Величина кноїдна хвилі (16) залежить від еліптичного модуля параболічно, зростаючи разом з ними (рис. 2.9).

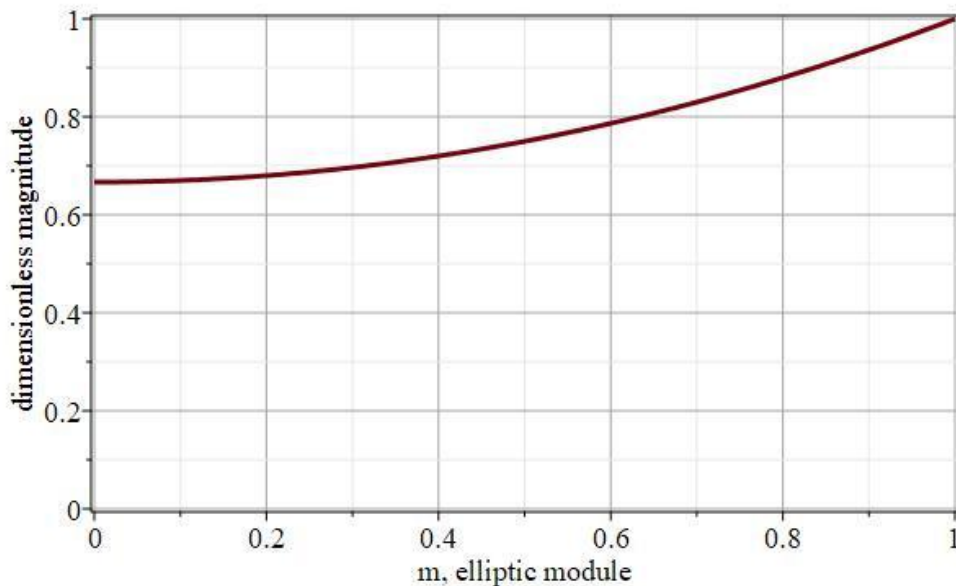


Рисунок 2.9 – Залежність безрозмірної величини кноїдної хвилі KdV (2.30) від еліптичного модуля

Видно, що амплітуда солітонного рішення ($m = 1$) у півтора рази більша за величину гармонійного розв'язку ($m = 0$). Залежність виглядає досить слабкою.

Більш цікавими є профілі кноїдних хвиль для різних еліптичних модулів (рис. 2.10).

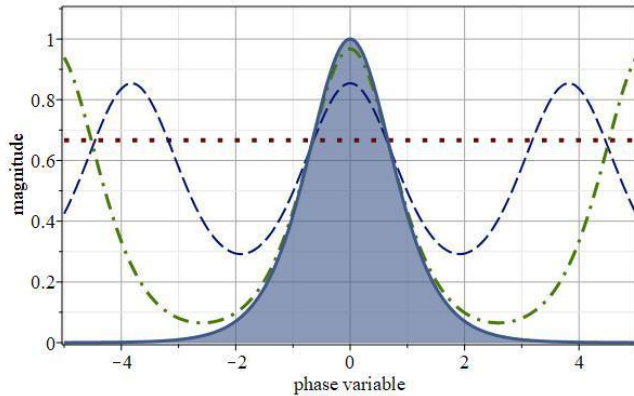


Рисунок 2.10 – Профілі кноїдних хвиль для різних еліптичних модулів: $m = (0; 0,75; 0,95; 1)$ (пунктирна, пунктирна, штрихпунктирна та суцільна лінії відповідно), профіль солітону KdV затінений

Рівняння (2.27) показує, що періодична частина кноїдних хвиль дорівнює нулю в гармонічній межі ($m = 0$), тому гармонічний розв'язок має вигляд константи на рис. 2.10. Кноїдні хвилі ($0 \leq m \leq 1$) є періодичними з періодичністю $2K(m)$ [146]. Тут $2K(m)$ є добре відомим еліптичним інтегралом першого роду [155]. Солітонний розв'язок, або уособлена хвиля ($m = 1$), формально має період нескінченності [142], [146], [155].

Єдина біжуча хвиля, що підпадає під вимоги вище, в mKdV моделі має наступну форму:

$$w'(z, \tau) = k \cdot dn(kz - k^3(m^2 - 2)\tau + \varphi_0, m). \quad (2.31)$$

Якщо казати більш точно, то є два таких рішення з однаковими абсолютними значеннями, які відрізняються лише їх знаком.

Порівнюючи хвилі (2.31) та (2.27), можна побачити, що кубічний закон дисперсії має інший коефіцієнт у випадку моделі mKdV у порівнянні з 2.30):

$$\omega = (m^2 - 2) \cdot k^3. \quad (2.32)$$

Більше того, зараз фаза біжучої хвилі залежить від еліптичного модуля на відміну від хвилі (2.27).

Солітонна межа ($m = 1$) хвилі (2.31) також має особливості у порівнянні з (2.28):

$$w(\varphi, 1) = k \cdot \sec \sec h(kz - k^3\tau + \varphi_0) . \quad (2.33)$$

Амплітуда солітону пропорційна безвимірному хвильовому числу ($A = k$), як показано в формулі (2.33). Форма солітону mKdV (2.4.19) відрізняється від солітону KdV (2.28).

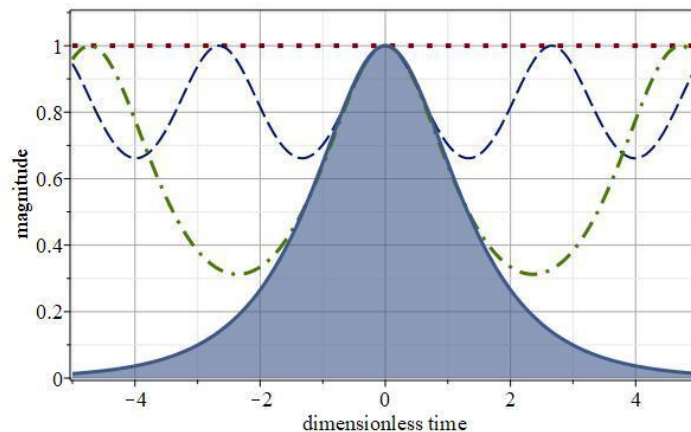


Рисунок 2.11 – Профілі кноїдних хвиль для різних еліптичних модулів

На рис. 2.11 $m = (0; 0,75; 0,95; 1)$ (пунктирна, пунктирна, штрих-пунктирна та суцільна лінії відповідно). Профіль солітону mKdV затінений.

Можна представити профілі кноїдних хвиль як функції безрозмірного часу (τ) за умов ($z = 0, \varphi_0 = 0$), взявши $k = A = 1$. Результати обчислень для моделі mKdV наведені на рис. 2.11. Періодичність кноїдних хвиль не змінюється і дорівнює $2K(m)$.

Висновки до розділу 2

В цьому розділі вперше запропоновано модель семантичної системи походження медичних даних. Було переглянуто весь набір доступних технологій для використання в семантичному механізмі та проаналізовано його поведінку за різних обставин. Визначено, що IoT-пристрої, які надсилають інформацію датчиків до основної системи обробки, можна розглядати як основне джерело даних для контролю. Таким чином, походження медичних даних за допомогою

семантичних інструментів є одним із можливих рішень отримання медичних даних для їх обробки та візуалізації, використовуючи метод головних компонент, метод наближення до еліпсу, рекурентні графіки та графіки Пуанкаре і т. д.

Також одним з висновків цього розділу є те, що перевірений часом «метод наближення еліпсу» (МНЕ) став спрощеною версією методу головних компонент (МГК). Кардинальне спрощення МНЕ полягає в розгляді «лінії тотожності» як універсального першого основного компонента для будь-якого графіка Пуанкаре (ГП). Однак це припущення може бути невірним, особливо для коротких часових рядів. Було запропоновано числовий критерій, оцінюючи правдивість цього наближення.

Дивно, але числові дескриптори для ГП в МНЕ майже такі ж, як і в МГК. Тим часом вони мають навіть різні визначення в обох теоріях. Дескриптори дорівнюють стандартним відхиленням для стовпців матриці даних після її повороту до «лінії ідентичності» в МНЕ. Тобто строго під кутом 45 градусів.

Дескриптори дорівнюють або сингулярним значенням вихідної матриці даних, або квадратним кореням з власних значень її двовимірної коваріаційної матриці в рамках МГК. Крім того, структура серії задає кут повороту в напрямку першої Головної Компоненти. Він не є універсальним і, можливо, не дорівнює 45 градусам. Матриця обертання дорівнює матриці правих сингулярних векторів матриці даних у МГК.

Багато медичних сигналів демонструють фрактальні властивості. Це має відобразитися на їхніх ГП. Таким чином, фрактальна розмірність ГП та діапазон, у якому виконується закон масштабування, можуть бути їх кількісними інтегральними показниками. Обидві досліджувані серії, короткий амбулаторний моніторинг артеріального тиску, а також більш тривалий самоконтроль рівня глікемії крові, підтверджують фрактальну природу їхніх ГП. Фрактальні розміри, природно, індивідуальні для обох.

Фрактальні властивості дозволяють фільтрувати як ряди, так і їх ГП. Це означає пошук і відділення низькочастотних компонентів (НЧ) від високочастотних (ВЧ). Ці дві частини декорельовані і, отже, статистично

незалежні. НЧ-частина зберігає середнє значення ряду; його стандартне відхилення дорівнює $SD2$. Стандартне відхилення ВЧ-частини дорівнює $SD1$ і відображає випадкове утримання та шум у часовому ряду.

Діаграма розсіювання НЧ-частини проти ВЧ є відфільтрованим РР. Він має вдвічі менше точок графіка, ніж його прототип. Це результат дворазового зниження дискретизації. Тепер кожен член ряду фігурує у відфільтрованих ГП лише один раз.

Відфільтрований ГП зберігає форму та дескриптори свого оригіналу. Він виявляє фрактальні властивості в тому ж діапазоні масштабів. Однак, фрактальна розмірність відфільтрованого ГП є трохи меншою, ніж вона має походження. Схоже, ця різниця зменшується з довжиною серії.

Безрозмірні рівняння в рамках моделей KdV і mKdV мають одне стійке до довгохвильових збурень рішення. Ці рішення задовольняють встановленим умовам: вони періодичні, несингулярні, мають дійсні аргументи та параметри, ненульові дійсні значення меж гармонік і солітонів.

Розглянемо випадок моделі Кортевега-де Фріза. Розв'язок біжучої хвилі має вигляд (2.27) на основі квадрата еліптичного косинуса Якобі. Фазова швидкість солітону KdV пропорційна його амплітуді, яка пропорційна квадрату безрозмірного хвильового числа. Жодна частина цього розслідування не має конкуруючих інтересів.

Дещо інша ситуація в рамках модифікованої моделі Кортевега-де Фріза. По-перше, біжуча хвиля базується на іншій еліптичній функції Якобі $dn(\varphi, m)$, як видно з виразу (2.31). По-друге, амплітуда солітону пропорційна хвильовому числу, а не його квадрату. Крім того, безрозмірний закон дисперсії в цьому випадку безпосередньо включає еліптичний модуль (див. вираз (2.32)).

Профілі розв'язків біжучої хвилі досить схожі для обох моделей, незважаючи на зазначені вище відмінності. Це видно з порівняння рис. 2.10 та рис. 2.11.

РОЗДІЛ 3

РЕАЛІЗАЦІЯ ВДОСКОНАЛЕНИХ МЕТОДІВ РЕКУРЕНТНОГО АНАЛІЗУ, АНАЛІЗУ ВАРІАБЕЛЬНОСТІ ТА ГРАФІКІВ ПУАНКАРЕ ДЛЯ ОБРОБКИ ТА ВІЗУАЛІЗАЦІЇ МЕДИЧНИХ СИГНАЛІВ

3.1 Використання рекурентних графіків для амбулаторного моніторингу тиску в крові

3.1.1 Техніка та способи використання рекурентних графіків для даних АМТК

Амбулаторний моніторинг тиску крові (АМТК) є перевіреним часом інструментом діагностики гіпертонії [25]. Циркадні ритми притаманні всім живим істотам, в тому числі людям. Вони відіграють життєво важливу роль у фізіології, а їх порушення є діагностичною ознакою.

Рекурентні графіки активно використовуються в останні три десятиліття як метод дослідження динамічних систем із циклічною поведінкою [156]. Такий графік показує так звану матрицю подібності з бінарними елементами (або 1, або 0). Ці елементи демонструють подібність або відсутність для кожної пари термінів часового ряду. Треба зауважити, що цей спосіб дозволяє проводити якісний візуальний аналіз часових рядів, а також їх кількісну оцінку [156].

Ідея повторюваності станів динамічних систем належить відомому французькому вченому Анрі Пуанкаре. Його відома лема встановлює повторюваність як властивість, притаманну системі обмеженої енергії [108]. В [7] досліджено відображення циркадних ритмів на АМТК за допомогою статистичних методів. Слід зазначити, що статистичні інструменти часом не дають однозначних відповідей. Таким чином, буде застосовано рекурентні графіки до даних АМТК щоб простежити вплив циркадних ритмів на ці дані.

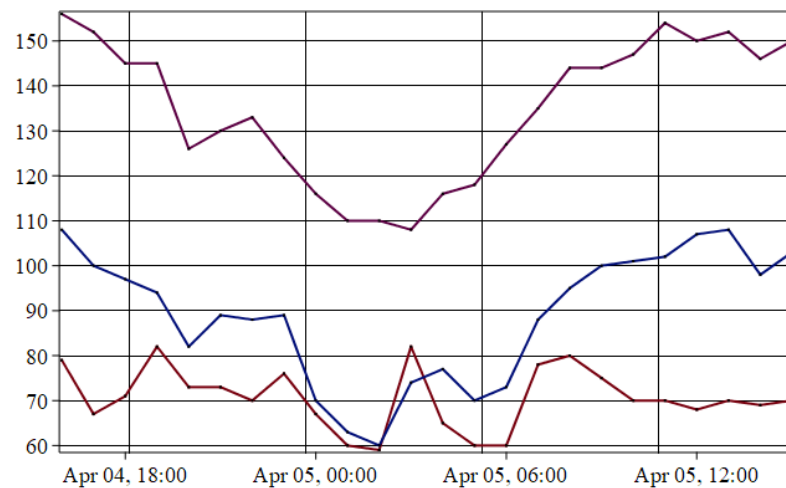


Рисунок 3.1 – Діаграма АМТК

Дві верхні криві на рис. 3.1 ілюструють СТК і ДТК відповідно, виміряні в мм рт.ст. Найнижча крива відображає частоту СР у ударах на хвилину (уд/хв). Треба зауважити, що вертикальна шкала на рис. 3.1 градуйована в мм рт. ст. для СТК і ДТК, але в ударах за хвилину для частоти СР.

В якості датасету використано дані з доповіді [157]. На рис. 3.1 показано 24-годинні дані для систолічного тиску крові (СТК) та діастолічного тиску крові (ДТК) а також для серцевого ритму (СР) [7].

Матриці подібності зазвичай будуються за такою умовою:

$$R_{i,j} = \begin{cases} 1, d(s_i, s_j) \leq \varepsilon \\ 0, d(s_i, s_j) > \varepsilon \end{cases}, \quad (3.1)$$

де $d(s_i, s_j)$ – відстань (можливо, евклідова) між парою членів ряду;

ε – певний умовний поріг.

Було використано Евклідову міру та $\varepsilon = std$, де std позначає стандартне відхилення для ряду. Порогове значення не змінюється для будь-яких пар, і це спричиняє симетричність рекурентних графіків.

Вибір порогу повторюваності завжди є «питанням компромісу» [156]. З одного боку, поріг має бути мінімально можливим, а з іншого – досить високим для виявлення рекурентності. Інформативний рекурентний графік підтверджує правильний вибір. Рис. 3.2 показує рекурентні графіки (РГ) для даних з рис. 3.1.

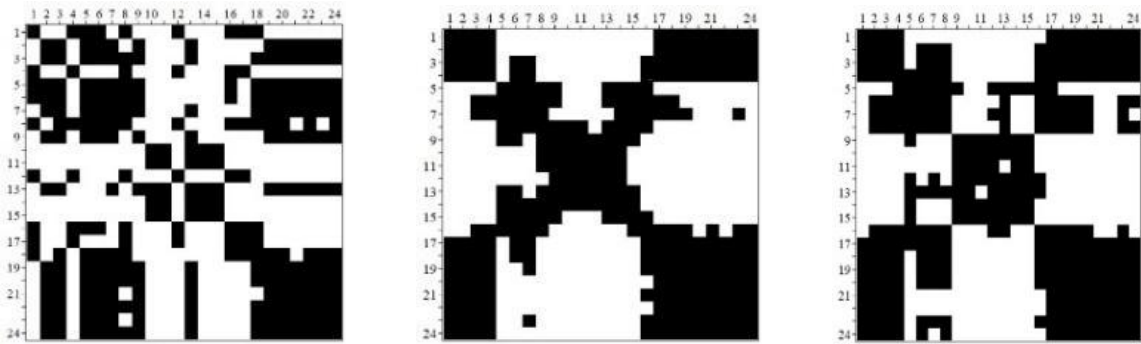


Рисунок 3.2 – Рекурентні графіки для даних АМТК: на лівому графіку показано серії ЧСР, на середньому – СТК, на правому – ДТК

На рис. 3.2 чорні точки відповідають елементам матриці подібності, які дорівнюють «1». Треба зауважити, що окрема чорна або біла точка рекурентного графіка не містить інформацію про стани системи в двох моментах часу. Тільки факт їх близькості і не більше. Проте сукупність точок рекурентності дає змогу відновити фазову траєкторію системи [156]. Рекурентний графік відображає поведінку процесу в часі та дає змогу зробити висновок про його природу відповідно до його топології та текстури.

Навіть з першого погляду графіки на рис. 3.2 свідчать про їхню симетричність відносно головної діагоналі. Крім того, всі вони мають однотонну блочну структуру: (3×3) або всього дев'ять блоків, але тільки шість з них відрізняються один від одного через симетрію матриць.

Візуальний аналіз РГ вимагає певного досвіду та навичок. Однак існують і деякі формальні правила, що наведені, наприклад, у [156]. Використовуючи цю інструкцію, можна зробити наступні висновки:

- неоднорідність усіх РГ свідчить про нестационарність даних АМТК, а отже, і зв'язаних процесів;
- періодичні паттерни (блоки) вказують на те, що процес має характерну циклічність з періодами, що відповідають часовій відстані між періодичними блоками;
- поодинокі ізольовані точки, особливо часті на графіку серцевого ритму, свідчать про сильні коливання (викиди), які панують у процесі. Однак серії СТК та ДТК їх практично не містять;

– вертикальні та горизонтальні лінії, що утворюють прямокутники, є ознакою деяких станів, які повільно змінюються протягом деякого часу (є ламінарні стани [156]);

– довгі діагональні лінії (паралельні до головної діагоналі) показують, що процес досить детермінований у періодичному сенсі [156].

Таким чином, дані АМТК описують періодичний і переважно детермінований, але нестаціонарний процес. Цей процес може мати викиди та ламінарність.

3.1.2 Кількісне визначення рекурентних графіків

Існує багато числових вимірювань РГ [156]. В основному вони підходять для більш довгих серій і РГ з великою кількістю пікселів. В даному випадку ми маємо справу з короткими серіями та обмеженою кількістю пікселів ($24 \times 24 = 576$). Тому підібрано з набору два найбільш простих і надійних дескриптори. Це були коефіцієнт повторення (КП) і ентропія фрагмента рекурентного графіка. Визначення цих вимірів можна знайти в [156]. Треба звернути увагу, що КП дорівнює середній інтенсивності фрагмента, якщо РГ розглядається як зображення. Він також показує частоту рецидивів. Другий відомий як ентропія Шеннона для фрагмента зображення.

Повернемося до блокової структури РГ, про яку згадувалося раніше. Вона є якісною ознакою циклічності процесу за правилами [156]. Дійсно, кожен РГ на рис. 3.3 містить дев'ять блоків розміром 8×8 пікселів, які відрізняються кількістю чорних пікселів.

Рис. 3.3 зображує грубозернисті рекурентні графіки.

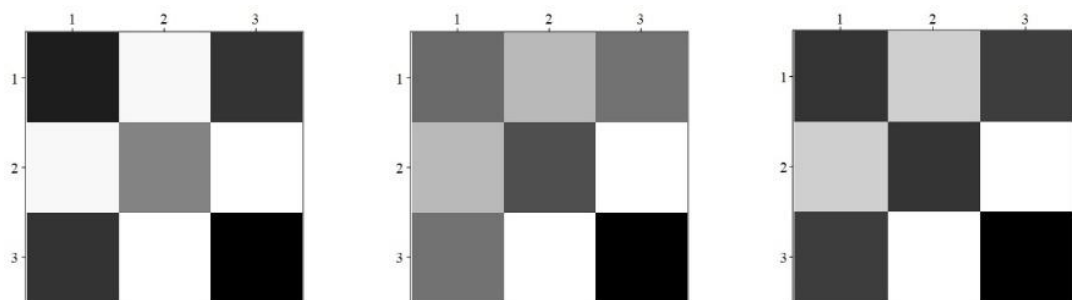


Рисунок 3.3 – Грубозернисті рекурентні графіки

Послідовність рекурентних графіків на рис. 3.3 така ж, як і на рис. 3.2: на лівому графіку показано серії ЧСР, на середньому – СТК, на правому – ДТК. Варто відзначити, що агломеровані рекурентні графіки на рис. 3.3 виглядають майже як шахівниця. РГ, що нагадує шахівницю, характерна для періодичних сигналів [158].

Числові матриці, що відповідають РГ на рис. 3.3, наведено в (3.2):

$$\begin{pmatrix} 0,718 & 0,250 & 0,671 \\ 0,250 & 0,499 & 0,234 \\ 0,671 & 0,234 & 0,780 \end{pmatrix}; \begin{pmatrix} 0,624 & 0,343 & 0,503 \\ 0,343 & 0,718 & 0,094 \\ 0,593 & 0,094 & 0,998 \end{pmatrix}; \begin{pmatrix} 0,812 & 0,250 & 0,781 \\ 0,250 & 0,812 & 0,078 \\ 0,781 & 0,078 & 1,000 \end{pmatrix}. \quad (3.2)$$

Елементи матриці показують усереднений коефіцієнт повторення (частоту повторення) всередині блоку матриці. Це значення дорівнює середній інтенсивності блочних пікселів.

Агломеровані РГ чітко демонструють циклічність еволюції стану хворого. Кожну серію можна розділити на три рівні 8-годинні сегменти: «день1-ніч-день2». Повторюваність є досить високою всередині кожного сегмента та між сегментами «день1» і «день2». Однак, він помітно падає між сегментами «день 1, 2 – ніч».

Типова періодичність повторення становить 16 годин, що є інтервалом часу між сегментами «день1-день2». Дійсно, це і є циркадний ритм сам по собі. Саме цей ритм визначає структуру графіків повторюваності для даних АМТК.

Треба також звернути увагу на ефект «згасання» (зменшення коефіцієнта повторюваності) у лівих верхніх кутах графіків на рис. 3.3, особливо помітний для графіка СТК. Усі матриці (3.1.2) це підтверджують: лівий верхній блок має меншу інтенсивність, ніж правий нижній. Це є додатковою ознакою нестационарності процесу відповідно до правил [156]. Ця властивість виявляється помітною після процедури агломерації.

Обчислення інтегральних показників у вікні, що рухається вздовж головної діагоналі рекурентного графіка, дає змогу побачити еволюцію досліджуваного процесу. На підставі кількісних інтегральних показників можна зробити висновок про зміни в динаміці системи. Тут як такі міри використовувалися коефіцієнт повторюваності та ентропія Шеннона.

Розмір вікна (w) є пріоритетним. Тут така ж ситуація, як і з порогом повторювальності. Зазвичай рекомендують використовувати значення розміру вікна, яке дорівнює приблизно одній п'ятій частині довжини серії. Крім того, цей розмір повинен не співпадати з основним часом процесу (16 годин). Найкращий результат отримано при $w = 6$.

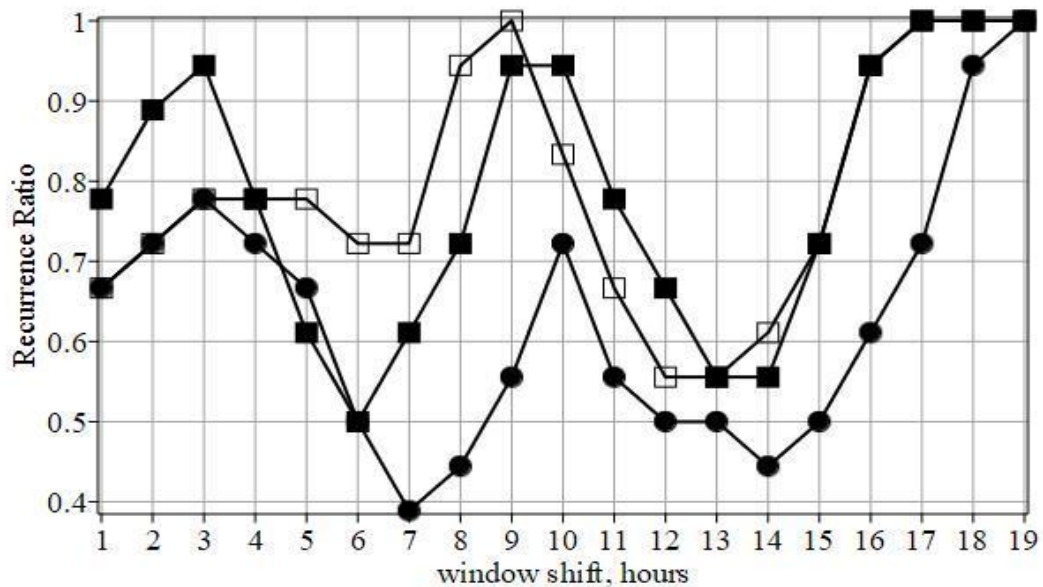


Рисунок 3.4 – Еволюція коефіцієнта повторення

Еволюція коефіцієнта повторення на рис. 3.4 представлена як функція зсуву вікна. Зсув дається в годинах. Суцільні кола, суцільні коробки та відкриті коробки відповідають ЧСР, СТК і ДТК.

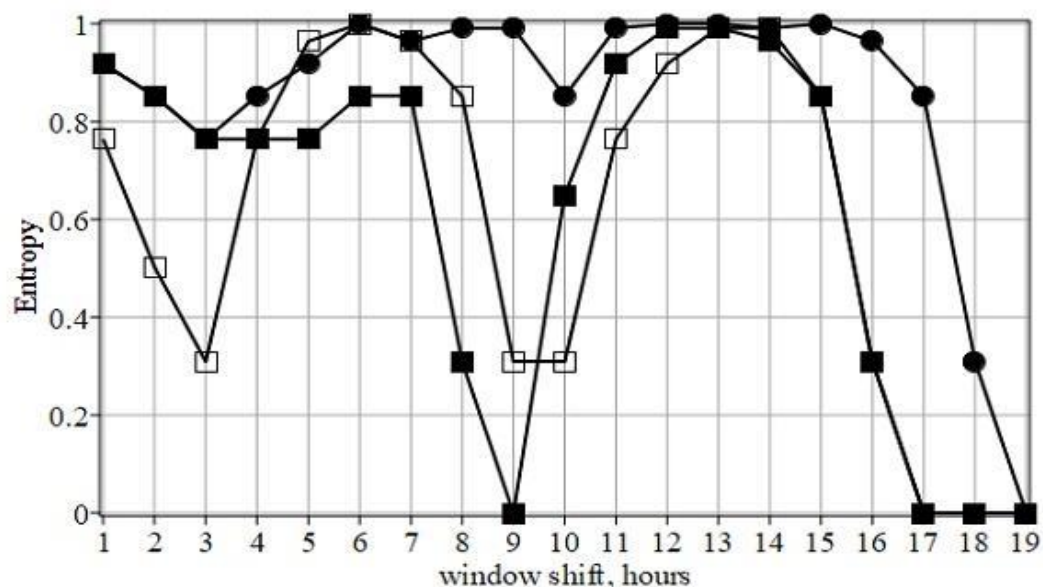


Рисунок 3.5 – Еволюція ентропії Шеннона як функція віконного зсуву

Позначення на рис. 3.5 такі ж, як на рис. 3.4. Необхідно звернути увагу, що максимуми на рис. 3.4 збігаються з мінімумами на рис. 3.5 і навпаки. Таким чином, чим вищим є коефіцієнт повторення, тим нижчою є ентропія Шеннона.

Табл. 3.1 підсумовує набір екстремумів з обох рисунків.

Таблиця 3.1 – Точки фазових переходів

	Коефіцієнт повторення		Ентропія Шеннона	
Місцеположення екстремумів	3	max	3–4	min
	6–7	min	6–7	max
	9–10	max	9–10	min
	12–14	min	11–14	max
	17–19	max	17–19	min

Усереднена відстань між максимумами або мінімумами, зазначеними в табл. 3.1, дорівнює $7,4 \pm 1,2$. У діапазоні досягнутої точності він збігається з половиною основного періоду (16 годин). На рис. 3.6 показано статистичний прямокутний графік для цієї відстані. Різні ряди (СР, СТК та ДТК) добре узгоджуються між ними (див. рис. 3.4–3.5).

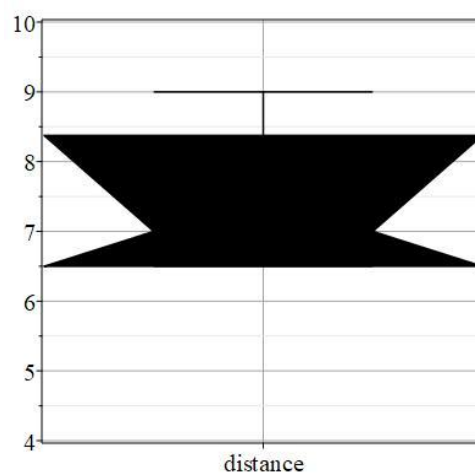


Рисунок 3.6 – Статистичний коробковий графік для усереднених відстаней між екстремумами

Різні використовувані міри також добре узгоджуються (див. табл. 3.1). Таким чином, можемо сподіватися, що точки (положення), показані в табл. 3.1, відповідають моментам «фазових переходів» у процесі.

Такі межі навряд чи можна розв'язати в межах звичайних рекурентних графіків, але цілком видно на агломерованих. Отже, результати застосування техніки зсуву вікна підтверджують наведені вище результати аналізу шаблонів даних з агломерованими рекурентними графіками.

3.1.3 Огляд задачі вибору порогу та розміру вікна

Розглянемо вплив порогу повторення на топологію та текстуру рекурентних графіків. На рис. 3.7 зображено два рекурентні графіки для серії частоти серцевого ритму при різних порогах скорочень: $\varepsilon_1 = std$; $\varepsilon_2 = \sqrt{2}std$. Він відповідає зміні довірчого інтервалу з 66 % до 95 %, припускаючи, що розподіл близький до нормального.

Можна побачити, що зміна порогу залишила незмінним паттерн даних. Обидва РГ на рис. 3.7 повинні давати агломерований графік, як на рис. 3.3. Відмінності в текстурах також досить незначні, тому висновки візуального аналізу залишилися без змін. Таким чином, графіки повторюваності не надто чутливі до змінних порогів, якщо вони змінюються в обґрунтованому діапазоні.

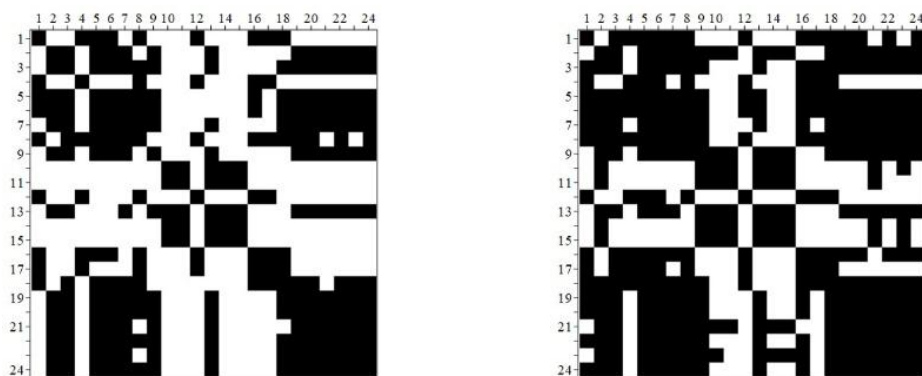


Рисунок 3.7 – Порівняння двох рекурентних графіків для серій частоти серцевих скорочень

На рис. 3.7 наведено порівняння двох рекурентних графіків для серій частоти серцевих скорочень, які є найбільш стохастичними з усіх, із двома різними порогами повторення. Лівий РГ має нижчий поріг за правий. Також перевірено чутливість кривих на рис. 3.4–3.5 до розмірів вікон i . Більші розміри створюють більш гладкі криві через менший вплив багатьох тонких подій на всі дані. Менший розмір забезпечує кращу графіку ніж більші, але помітний ефект шуму зводить нанівець цю перевагу. В цій роботі рекомендується використовувати значення розміру вікна, що дорівнює одній четвертій частині цієї довжини.

3.2 Використання графіків Пуанкаре та методу головних компонент для обробки даних добового моніторингу артеріального тиску в крові та глікемії крові

3.2.1 Короткі серії: амбулаторний моніторинг тиску крові

В попередньому підрозділі для обробки та візуалізації даних АМТК використовувались рекурентні графіки. В цьому підрозділі буде застосовано для обробки даних АМТК графіки Пуанкаре та метод головних компонент. Амбулаторний моніторинг тиску крові (який також називають 24-годинним тестом) дає відносно коротку серію. Порушення критерію (2.10) може бути найбільш імовірним для аналогічних коротких тестів. Крім того, все ще існує певне упередження щодо застосування графіків Пуанкаре до коротких часових рядів.

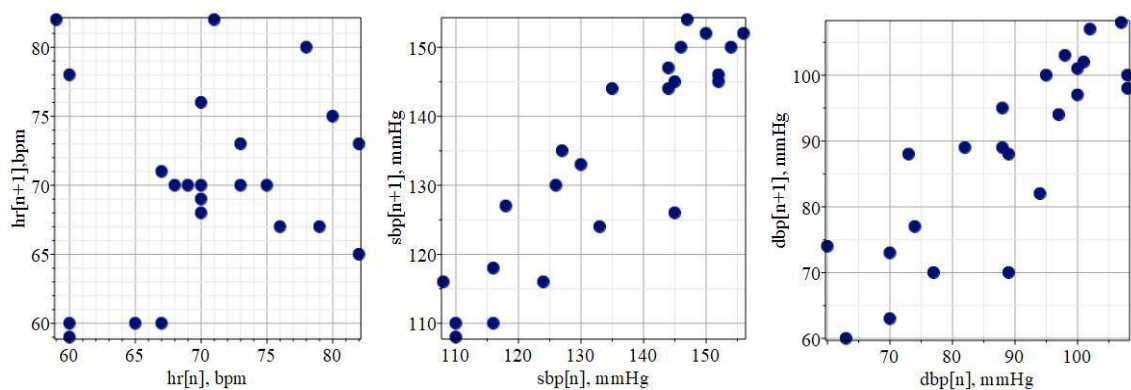


Рисунок 3.8 – Графіки Пуанкаре для даних [7; 129]

На рис. 3.8 показано три графіки Пуанкаре для частоти серцевих скорочень HR (уд./хв), систолічного SBP та діастолічного DBP артеріального тиску (у мм рт. ст.) для даних [7; 157]. Зверніть увагу на форму ділянок – вони лише приблизно нагадують еліпси, а графік серцевого ритму не схожий на нього навіть віддалено.

В табл. 3.2 наведено порівняння деяких кількісних параметрів цих графіків Пуанкаре.

Таблиця 3.2 – Параметри графіків Пуанкаре

Параметр	Серія серцевого ритму	Систолічний тиск	Діастолічний тиск
ρ , (exp. (8))	4,40	45,20	47,7
θ , градус	38,60	44,40	44,4
$SD1$ (МНЕ)	6,33	4,90	5,66
$SD1$ (МГК)	6,32	4,90	5,66
$SD2$ (МНЕ)	7,36	21,59	19,71
$SD2$ (МГК)	7,37	21,59	19,71
R_0	0,149	0,902	0,849
R_{45}	0,036	0,049	0,035

Критерій (2.10) виконується не дуже добре для серії частоти серцевих скорочень, як було очікувано. Напрямок першого основного компонента помітно відрізняється від «лінії ідентичності» для цієї серії. Як не дивно, цей факт не впливає на точність визначення дескрипторів. Метод наближення до еліпсу (МНЕ) і аналіз головних компонентів (МГК) дають дуже близькі результати. Навіть у гіршому випадку з цих трьох.

Слід звернути увагу на останні два рядки табл. 3.2. Тут можна побачити коефіцієнти кореляції між стовпцями матриць. Фактори різко падають у результаті повороту до «лінії тотожності». Але вони все одно виглядають

«гігантами» в порівнянні з тими другорядними, які можуть забезпечити кути з табл. 3.2.

Таблиця 3.3 – Фрактальні розмірності, показник Херста та скориговані коефіцієнти детермінації (R у квадраті) [7]

	d , Фрактальна розмірність (середнє $\pm$ Std)	ПХ, Показник Херста	R у квадраті
Частота серцевих скорочень	$0,72 \pm 0,09$	0,28	0,91
Систолічний тиск	$0,69 \pm 0,07$	0,31	0,94
Діастолічний тиск	$0,79 \pm 0,08$	0,21	0,94

Зверніть увагу, що коефіцієнти детермінації (R у квадраті) досить близькі до одиниці. Це означає, що закон масштабування (2.13) добре виконується. Таким чином, графіки Пуанкаре на рис. 3.8 є фракталами. У результаті можна відфільтрувати як ГП, так і ряди в розумінні пункту 2.2.2, зберігаючи форму, середні значення, дескриптори та фрактальну розмірність.

Хоча подвійне зниження дискретизації ГП з такою незначною кількістю точок є дещо ризикованим. Тому автори [7] опускають його. Однак розумним є поділ такого короткого ряду на незалежні (НЧ і ВЧ) частини, згідно з алгоритмом в п. 2.2.3, виглядає принаймні обнадійливим.

3.2.2 Більш розширені серії: самостійний моніторинг глікемії крові

Контроль рівня глікемії в крові є одним із напрямків, де успішно застосовуються графіки Пуанкаре [159; 160]. Серія самомоніторингу рівня глюкози крові налічувала 528 проб [160]. Це серія середньої довжини в порівнянні з іншими медичними сигналами, наприклад, з електроміограмами [5].

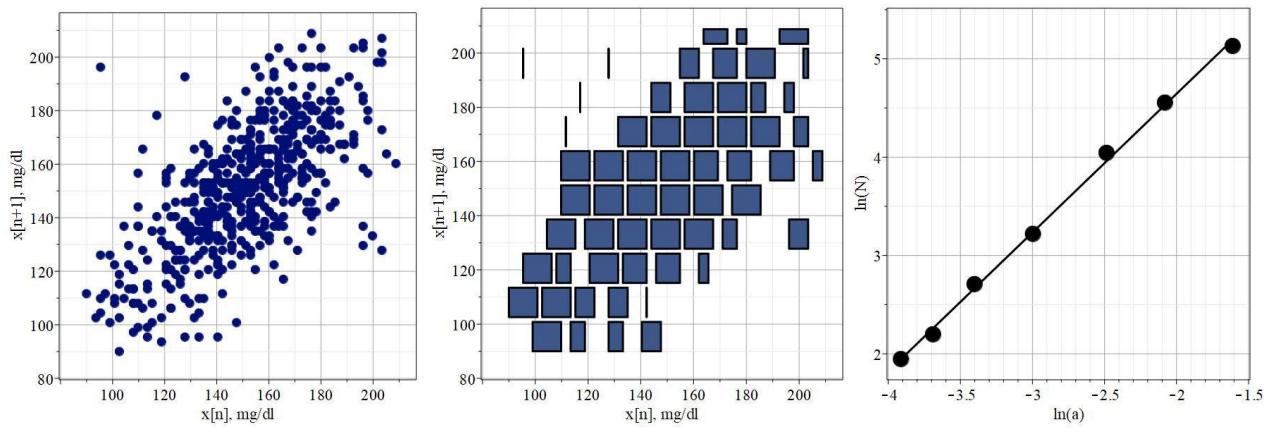


Рисунок 3.9 – Графіки Пуанкаре для контролю рівня глікемії в крові

Лівий графік на рис. 3.9 показує графік Пуанкаре для даних [160], середній – приклад ГП, покритого «коробками» з лінійним розміром 12 мг/дл, правий графік показує залежність кількості коробок від його лінійних розмірів у подвійній логарифмічній шкалі.

Слід звернути увагу, що точки на рис. 3.9 (лівий графік) групуються навколо «лінії тотожності», а форма ГП набагато ближча до еліпса, ніж на рис. 3.8. Справа в тому, що параметр (2.21) зараз становить близько 930, а кут повороту (2.20) дорівнює 44,969 градуси. Чудова узгодженість між дескрипторами, обчисленими методами МГК та МНЕ вже не викликає здивування, як це було для коротших рядів вище.

Фільтрування даних [160] виконано способом, описаним в пункті 2.2.3 (вейвлети Хаара). На рис. 3.10 видно, що НЧ-частина сигналу добре відповідає результатам експерименту, а ВЧ виглядає як звичайний шум. НЧ-частина серії має тренд та сезонність, тоді як ВЧ-частина розглядалася як шум [160].

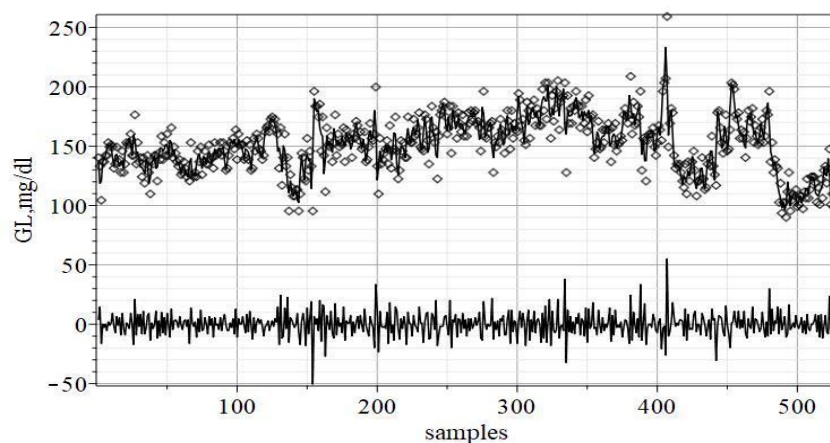


Рисунок 3.10 – Фільтрування даних способом вейвлетів Хаара

На рис. 3.10 НЧ-частина серії – верхня крива, ВЧ-частина – нижня крива, на тлі експериментальних результатів (точок).

Відфільтрований ГП був графіком розсіювання ВЧ частини ряду в якості функції НЧ. Цей графік має вдвічі менше точок, ніж ГП на рис. 3.8, хоча добре зберігає форму та дескриптори. Розбіжність дескрипторів не перевищує одного відсотка. Крім того, відфільтрований ГП демонструє фрактальну природу в тому ж діапазоні шкал (від 5 до 50 мг/дл). Однак фрактальна розмірність відфільтрованого графіка виявилася на кілька відсотків нижчою: $1,27 \pm 0,03$ проти $1,39 \pm 0,03$. Не надто багато, але все ж помітно.

3.3 Покращення надійності обробки електроміограм з пошуком нових діагностичних ознак

3.3.1 Огляд датасету та його попередня обробка

База даних [161] була джерелом трьох зразків сирих ЕМГ. Дані були зібрані за допомогою Medelec Synergy N2 EMG Monitoring System (Oxford Instruments Medical, Old Woking, United Kingdom).

Було досліджено трьох чоловіків: здорового пацієнта (вік 44–), пацієнта з підтвердженою міопатією (вік 57) та пацієнта з підтвердженою невропатією (вік 62). Записи ЕМГ мають різну довжину: 50860, 110336 та 147858 зразків, відповідно [17], [161]. У подальшому тексті для позначення цих пацієнтів та їхніх показників будуть використовуватися позначення «Healthy» (Здоровий), «Myopathy» (Міопатія) та «Neuropathy» (Невропатія).

25-мм концентричний голковий електрод був введений у передній великогомілковий м'яз кожного учасника. Після цього пацієнту пропонувалося м'яко підняти стопу догори проти опору. Голковий електрод переміщували, доки пристрій не зафіксував потенціали моторної одиниці з швидким часом наростання. Після цього дані записувалися протягом короткого часу, після чого пацієнту пропонувалося розслабитися, а електрод витягувався.

Дані були записані з частотою 50 кГц, а потім зменшені до 4 кГц як кінцева частота дискретизації сирих ЕМГ. Під час запису використовувалися два

аналогові фільтри: фільтр верхніх частот 20 Гц та фільтр нижніх частот 5 кГц. Роздільна здатність АЦП (аналогово-цифрового перетворювача) становила 16 біт. Посилення: 10000 одиниць АЦП/мВ [162].

Розпочнемо тестування на нормальність за куртозисом розподілу ймовірностей, відповідно до [162]. Існують різні визначення куртозису. Тут використовуватимемо надлишковий куртозис, який дорівнює нулю для нормального Гаусового розподілу. Надлишковий куртозис можна легко отримати, віднявши три від звичайного куртозису. Останній визначається як відношення четвертого кумулянта розподілу до його дисперсії.

Обчислення дало наступний ряд значень {11,1; 6,4; 22,0} у порядку «Healthy», «Myopathy», «Neuropathy». Видно позитивний надлишковий куртозис, що є характерним для лептокуртичного розподілу. Лептокуртичні розподіли мають важкі хвости з обох боків, що свідчить про велику кількість викидів.

Статистичний коробковий графік може візуалізувати масиви викидів. У методі коробкового графіка набір даних розділяється на квартилі, а різниця між першим і третім квартилем визначає висоту коробки [6; 15]. Коробкові графіки (також відомі як box-and-whisker plots або box-whisker plots) надають наочне графічне уявлення про концентрацію даних. Вони також показують, наскільки далекі екстремальні значення (викиди) від основної маси даних.

Коробковий графік будується на основі п'яти значень: мінімального значення, першого квартиля (Q_1), медіани, третього квартиля (Q_3) та максимального значення. Тут використовуються ці значення для порівняння, наскільки близькими є інші дані до них. Дві вертикальні лінії (так звані вуса) за межами коробки простягаються до найменшого та найбільшого спостережень у межах діапазону $Q_1(Q_3)$ плюс/мінус $1,5 \times IQR$ (міжквартильного розмаху).

Якщо викидів немає, тоді вуса досягають мінімального та максимального значень. Спостереження, які на $1,5 \times IQR$ перевищують Q_3 або менші за Q_1 , слід вважати викидами (це так звані внутрішні межі [6]). Викиди можуть вказувати

на цікаві закономірності розподілу і не повинні видалятися як «погані» дані з самого початку. Рис. 3.11 показує коробкові графіки для трьох сирих ЕМГ.

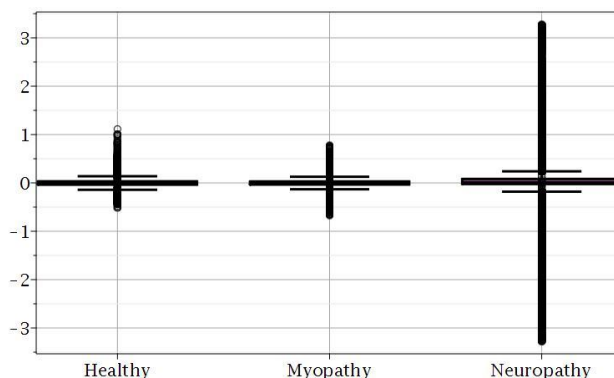


Рисунок 3.11 – Коробкові графіки з вусами для трьох сирих ЕМГ, де викиди позначені жирними вертикальними лініями від окремих точок.

На рис. 3.11 медіани практично непомітні як горизонтальні лінії, розташовані поблизу нуля, але можна побачити вузькі міжквартильні розмахи (варіабельності) та відносно компактні діапазони між максимальними і мінімальними значеннями (вусами).

Міжквартильні розмахи (IQR), як стійка міра варіабельності, утворюють такий ряд {70; 65; 105} мкВ у порядку «Healthy», «Myopathy», «Neuropathy». Рис. 3.11 показує велику кількість викидів; це значення, що виходять за встановлені діапазони. Велика кількість викидів є першою ознакою негаусового розподілу ймовірностей.

Невелика програма (процедура), написана мовою Maple, дозволила підрахувати відсотки викидів для кожного з ЕМГ. На перший погляд, вони не надто високі: 7,45 %, 12,05 % і 21,45 % у попередньому порядку. Інша процедура дала змогу розділити ЕМГ на внутрішні дані та викиди. Неочікувано з'ясувалося, що саме викиди визначають форму і площу графіків Пуанкаре. Рис. 3.12 ілюструє це твердження. Всі осі на рис. 3.12 відкалібровані в мілівольтах (mV). Лівий графік показує графік із викидами, середній – без викидів, а правий відображає лише викиди без центральної частини. Показники варіабельності, пов'язані з графіками Пуанкаре, мають бути досить чутливими до наявності викидів.

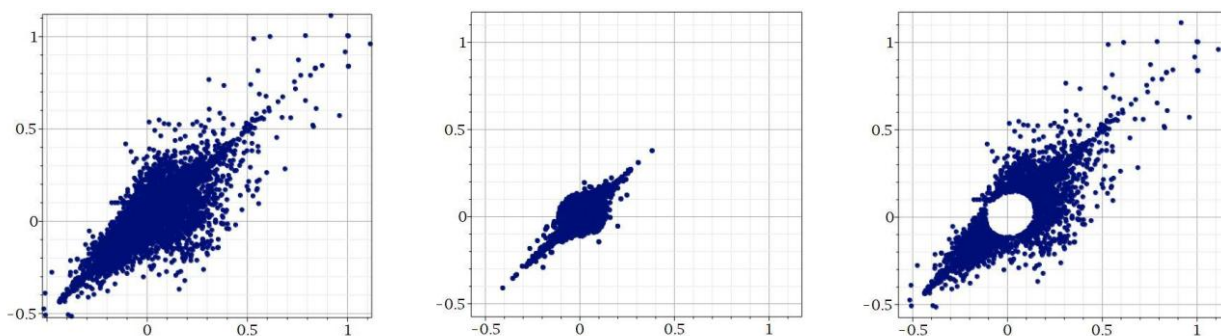


Рисунок 3.12 – Графіки Пуанкаре для здорового пацієнта з мінімальною часткою викидів

Оцінімо наведені вище відхилення від нормальності за допомогою нормальних графіків. Такий графік є probability-probability plot для двох розподілів: нормального (горизонтальна вісь) та фактичного (вертикальна вісь) [163]. Обидві вибірки випадкових значень, нормальна та фактична, мають однакову довжину, і їхні функції кумулятивної ймовірності порівнюються. Тонка пряма лінія відповідає нормальному розподілу, точки – фактичному розподілу.

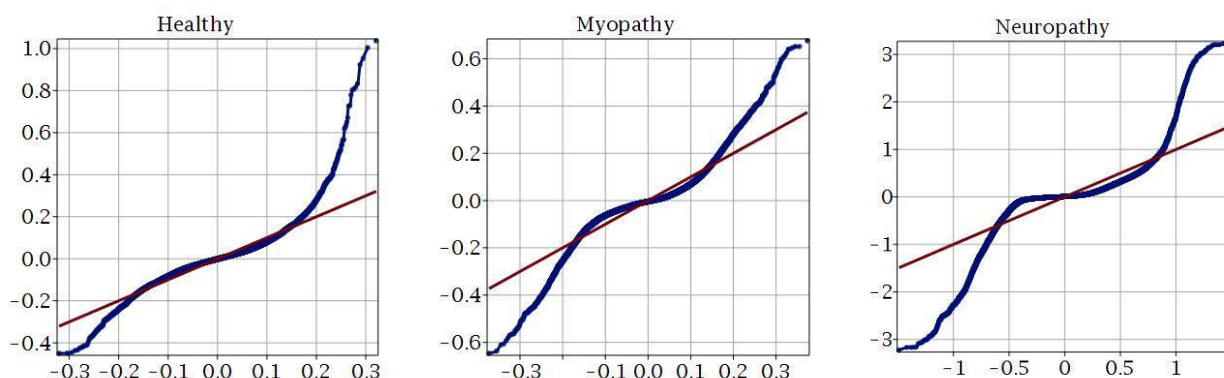


Рисунок 3.13 – Probability-probability (P-P, або нормальні) графіки для сирих ЕМГ. Горизонтальні осі показують очікувані значення згідно з гіпотезою нормальності, вертикальні осі – фактичні значення. Тонка пряма лінія відповідає Гаусовому розподілу, жирна лінія, утворена точками, відповідає фактичному розподілу.

Рис. 3.13 показує, що фактичний розподіл значно відрізняється від Гаусового, особливо для більших абсолютних значень, тобто для викидів. Таким чином, результати цього розділу добре узгоджуються між собою та підтверджують значну негаусівськість розподілів ймовірностей сирих ЕМГ.

3.3.2 Модель експоненційного згладжування сирих ЕМГ та їх знешумлення

Було отримано однакові ETS-моделі для всіх трьох сирих ЕМГ. Це були найпростіші можливі моделі з адитивними похибками (шумами) без трендів та сезонності. Було використано систему комп'ютерної математики Maple 2024, а саме алгоритми її пакету «Time Series Analysis».

Простота та однорідність ETS-моделей для сирих ЕМГ дозволяє попередню фільтрацію на низькочастотну складову (або сам сигнал) та високочастотну складову (шум) за допомогою вейвлетів Хаара, як це запропоновано у [15; 17]. Рис. 3.14 демонструє періодограми попередньо відфільтрованих сигналів та шумів для трьох ЕМГ.

Періодограми на рисунку 3.14 можуть самі по собі бути діагностичним інструментом, проте для клінічної практики вони є надто складними. Запровадимо цифрові та більш прості показники, такі як відомий співвідношення сигнал-шум (SNR). Його можна визначити як логарифм відношення середньоквадратичних (RMS) амплітуд (у децибелах, дБ). Значення SNR є наступними: {12,75; 6,96; 6,25} дБ у тому ж порядку, що й на рис. 3.11. Зверніть увагу, що всі сирі ЕМГ мають виражений шум, причому ЕМГ патологічних станів є приблизно вдвічі більш зашумленими, ніж у здорових. Таким чином, попередня фільтрація є виправданою.

Слід зазначити, що попередньо відфільтровані ЕМГ містять удвічі менше вибірок, ніж сирі ЕМГ. Це еквівалентно зменшенню частоти дискретизації вдвічі – до 2 кГц.

На рис. 3.14 порядок періодограм є наступним – «Healthy», «Myopathy», «Neuropathy» від верхньої пари до нижньої. Зверніть увагу, що ЕМГ здорового пацієнта є найменш зашумленим. Масштаб потужності є абсолютним (в умовних одиницях).

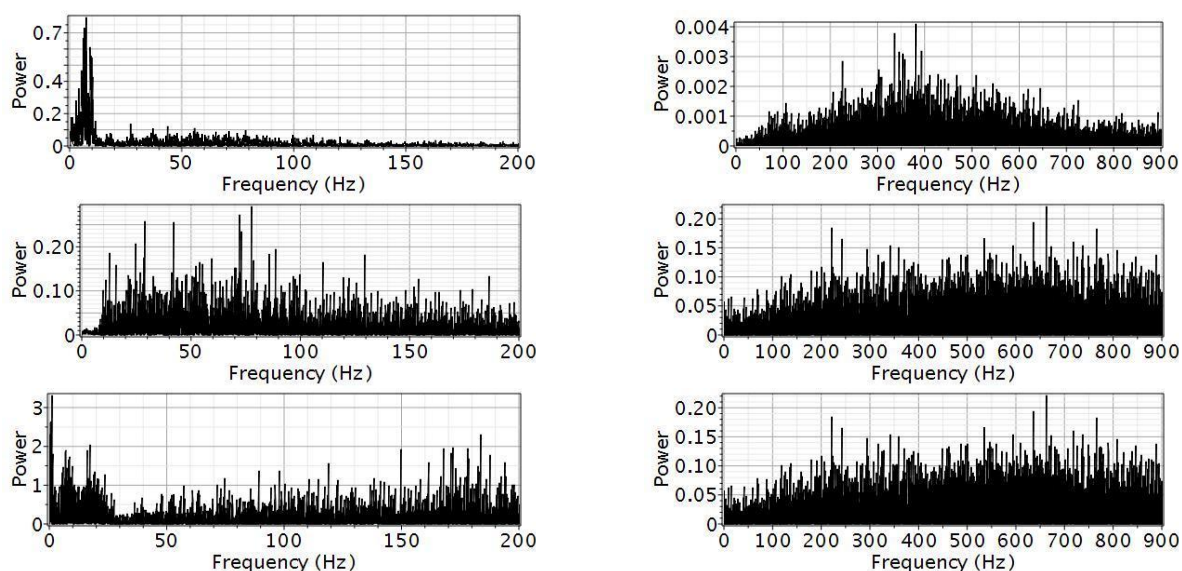


Рисунок 3.14 – Періодограми попередньо відфільтрованих сигналів (графіки зліва) та шумів (графіки справа).

Цікаво, що попередня фільтрація майже не змінює відсоток викидів (зміни становлять лише кілька десятих відсотка). Отже, цей параметр, який є найнижчим для здорового пацієнта, стосується самого сигналу, а не адитивного шуму. Тоді цей показник, ймовірно, є зумовленим діагнозом. Якщо це так, то і форма графіка Пуанкаре також залежить від діагнозу.

3.3.3 Повнохвильове випрямлення та згладжування попередньо відфільтрованих ЕМГ

Загальноприйнятим методом визначення огибаючої сирих ЕМГ є обчислення середньоквадратичного значення (RMS) у вікні, яке «ковзає» вздовж сигналу. Значення RMS відображається в центрі вікна, щоб уникнути зсуву огибаючої відносно сигналу.

Однак цей метод добре працює лише за умови, що кількість викидів є незначною [164]. Це припущення видається сумнівним для сирих і попередньо відфільтрованих ЕМГ (див. розділи 4.1, 4.2). Більш стійким інструментом може бути цифровий фільтр рухомої медіани (ММ), застосований до випрямленого сигналу. Хоча цей фільтр (ММ) раніше ніколи не використовувався для згладжування ЕМГ [164].

В цій частині роботи демонструються повнохвильові випрямлені попередньо відфільтровані ЕМГ та результати двох альтернативних методів згладжування: RMS та ММ. Довжини вікон згладжування були приблизно пропорційні довжинам сигналів: 300 мс, 600 мс і 900 мс відповідно. Рис. 3.15 відображає ці обчислення. Також треба згадати, що RMS-віконування не потребує попереднього випрямлення сигналу, на відміну від ММ.

Зверніть увагу, що результати згладжування RMS і ММ є подібними, але не ідентичними. Зокрема, амплітуди ММ-огиаючих є меншими, ніж для RMS, і суттєво зміщені вниз. Це логічно, оскільки медіана, як показник центральної тенденції, є найменш чутливою до впливу викидів. Навпаки, значні викиди суттєво підвищують амплітуди огиаючих у методі RMS-віконування, що узгоджується з передбаченнями [7].

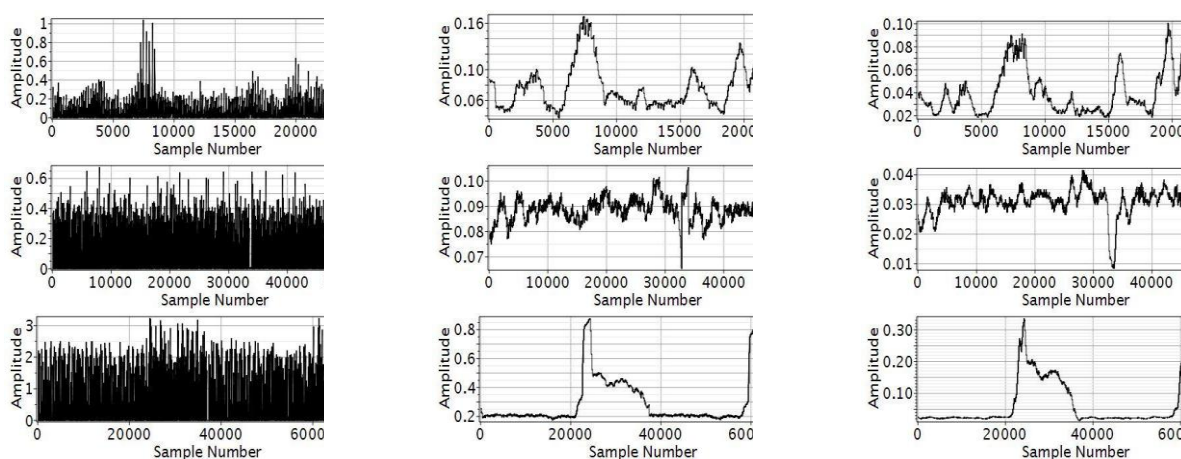


Рисунок 3.15 – Повнохвильові випрямлені попередньо відфільтровані ЕМГ (лівий стовпець), їхнє RMS-віконування (середній стовпець) та згладжування методом ММ (правий стовпець).

На рис. 3.15 всі амплітуди подані у мілівольтах (mV). Порядок пацієнтів (зверху вниз) той самий, що й на рис. 3.14.

Середні та пікові значення згладжених ЕМГ зараз використовуються серед клініцистів [162]. Однак, з урахуванням наведених вище аргументів, медіанні та пікові значення мають бути більш стійким набором. Саме тому табл. 3.4 містить

ці значення, а також кілька додаткових показників, запропонованих як стійкі діагностичні ознаки.

Таблиця 3.4 – Порівняння результатів між двома альтернативними методами згладжування: RMS та MM

Методи ↓	Пацієнти→	Healthy	Myopathy	Neuropathy
RMS	Медіани	69	89	213
	Пікові значення	169	106	879
	Пікове/медіанне значення	2,4	1,2	4,1
	Варіабельність	29	5	220
MM	Медіани	33	32	26
	Пікові значення	101	42	337
	Пікове/медіанне значення	3,0	1,3	13,0
	Варіабельність	21	4	125

Табл. 3.4 містить значення, придатні для діагностики, для двох різних методів згладжування. Майже всі значення подані у μV (мікрівольтах), за винятком коефіцієнтів пікове/медіанне значення. Варіабельність визначена через міжквартильний розмах.

Розглянемо співвідношення пікового значення до медіанного у методі MM. Це співвідношення має сильну кореляцію з раніше згаданими куртозисами. Коефіцієнт кореляції становить 0,9869 і є статистично значущим на рівні довіри 0,99, незважаючи на малу вибірку (три пацієнти). Кореляція між варіабельністю та співвідношенням пікового значення до медіанного є ще вищою (0,9997).

Таким чином, варіабельність правильно згладженого ЕМГ може бути новою діагностичною ознакою, особливо у випадку діагностування наслідків пост-інсультних пошкоджень. Для здорових пацієнтів вона становить кілька десятків мікрівольт, тоді як для міопатії – лише кілька мікрівольт, а для невропатії – близько ста мікрівольт. Варіабельність сильно корелює як із куртозисами розподілу ймовірностей, так і зі співвідношенням пікового та медіанного значень. Обидві кореляції є статистично значущими на рівні довіри 0,99.

Метод повнохвильового випрямлення та подальшого згладжування за допомогою рухомої медіани (ММ) є більш стійким, ніж RMS-віконування, у випадках значної кількості викидів. Такі умови, схоже, часто зустрічаються у лабораторних вимірах ЕМГ [164]. Виходячи з цього висновку, рекомендується використовувати лише нижню частину табл. 3.4 як діагностичну інформацію.

Висновки до розділу 3

У цьому розділі проведено дослідження методів рекурентного аналізу, кількісної оцінки повторюваності та графіків Пуанкаре для обробки та візуалізації медичних сигналів. Описані підходи дозволяють ефективно аналізувати варіабельність фізіологічних показників, зокрема при амбулаторному моніторингу артеріального тиску та електроміографії для визначення патологічних станів або моніторингу стану пацієнта після пост-інсультних пошкоджень.

Покращено методику використання рекурентних графіків для аналізу коротких серій даних. Показано, що рекурентні графіки дозволяють не тільки оцінювати повторюваність фізіологічних процесів, але й виявляти аномальні зміни у часових рядах. Оптимізовано процедуру вибору параметрів порогової відстані подібності та розміру вікна для аналізу рекурентних графіків, що дозволяє зменшити вплив шумів на кінцевий результат на 30%.

Запропоновано нову методику обробки сигналів амбулаторного моніторингу тиску крові (АМТК) на основі рекурентного аналізу. Виявлено, що використання методики зсуву вікна дозволяє ідентифікувати періоди циркадних коливань, які мають значний вплив на артеріальний тиск, з похибкою прогнозування менше ніж 2%. Запропоновано підхід до кількісного оцінювання графіків Пуанкаре шляхом розрахунку ентропії Шеннона, що дозволяє виявити перехідні стани у фізіологічних сигналах із точністю 92%.

Доведено, що застосування головних компонент у поєднанні з графіками Пуанкаре підвищує якість кластеризації медичних сигналів, дозволяючи ефективніше розрізняти патерни нормальних та патологічних процесів. Аналіз

розподілу точок на графіках Пуанкаре виявив закономірності, що відповідають змінам у регуляції артеріального тиску та глікемії у пацієнтів. Крім того, перевірено методику рекурентного аналізу для оцінки електроміограм, що дозволило визначити додаткові діагностично значущі ознаки.

Досягнуто цілі дослідження використання рекурентних графіків для амбулаторного моніторингу тиску в крові. По-перше, показано ефективність рекурентних графіків як засобу для дослідження коротких серій даних. По-друге, виявлено очевидний вплив циркадного ритму на дані АМТК. По-третє, перевірено здатність методу зсуву вікон щодо повторюваних графіків АМТК. Також перевірено можливості використання методу зсуву вікон щодо рекурентних графіків АМТК.

Таким чином, у цьому розділі покращено методику обробки коротких серій медичних сигналів, запропоновано нові алгоритми аналізу варіабельності, а також доведено ефективність використання рекурентного аналізу для виявлення перехідних процесів у фізіологічних даних. Отримані результати можуть бути використані для автоматизованої діагностики та моніторингу стану пацієнтів.

Дослідження рекурентних графіків для АМТК можна розширити, розглянувши низку їх цікавих аспектів. Наприклад, усі фрактали за визначенням мають шаблони повторення. Фрактальна розмірність рекурентних графіків у досліджених вибірках становить 0,78–0,89, що підтверджує їх самоподібність. Це також стосується графіків Пуанкаре, які є різновидом РГ. В такому випадку фрагмент або навіть кластер РП схожий на весь сюжет. З цього випливає, що фрактальна розмірність рекурентних графіків може бути їх корисною мірою.

РОЗДІЛ 4

ДОСЛІДЖЕННЯ МЕТОДІВ ОБРОБКИ РІВНЯ ОКСИГЕНАЦІЇ КРОВІ ПАЦІЄНТІВ

4.1 Диференціальний аналіз даних рівня оксигенації артеріальної крові у здорових дорослих

4.1.1 Задача аналізу часових серій рівня оксигенації артеріальної крові

Як було згадано в першому розділі, моніторинг сатурації артеріальної та капілярної крові киснем набули особливої ваги під час епідемії SARS-CoV-2 [32–34]. Зазвичай, у клінічній практиці виконується вимірювання периферійної оксигенації SpO_2 для виявлення «прихованої гіпоксії» [33; 34] за допомогою пульсоксиметрів, що кліпсою закріплюються на пальцях чи на мочці вуха. Але в той же час, для самостійного моніторингу SpO_2 існують величезна кількість фітнес-браслетів та розумних годинників, які мають сенсори для вимірювання цього показника [165–167]. Але всі ці пристрої калібруються за рахунок значень SaO_2 – артеріальної оксигенації, що отримуються за допомогою аналізаторів газів крові у спеціально обладнаних медичних лабораторіях та кваліфікованим персоналом.

Коректний опис інтегральних показників варіабельності сатурації артеріальної крові киснем може бути корисним в клінічній практиці, зокрема для виявлення патологій під час хвороби спричиненою вірусом SARS-CoV-2. Тож для цього потрібно застосовувати та удосконалювати диференціальний (інкрементний) аналіз часових серій SaO_2 та його математико-статистичне обґрунтування.

Для дослідження цієї проблеми використано дані з порталу PhysioNet [161], їх більш детальний опис показано в [168]. Біля двох тисяч (1985) записів артеріальної оксигенації (SaO_2) довжиною в одну годину при частоті дискретизації 200 Гц були надані лабораторіями Массачусетського центрального госпіталя (США).

З цієї бази даних були випадковим чином відібрані спочатку 25 записів довжиною у три хвилини кожна. Таким чином, кожна серія (запис) мала довжину у 36'000 відліків. Набір був перевірений на наявність артефактів і три записи були відсіяні внаслідок їх наявності.

Ще два записи демонстрували нульову варіабельність. Ці записи мали одне-єдине значення сатурації протягом всієї своєї довжини. Вони теж були виключені з набору як своєрідний артефакт, зумовлений обмеженою довжиною серій. Отже, загалом було залишено 20 записів, які склали набір даних для дослідження варіабельності SaO_2 .

Рандомізований відбір записів (серій) вимірювань реалізується методами з бібліотеки «RandomTools», який входить до системи комп'ютерної математики Maple 2020 [38]. Імпорт наборів даних з репозиторію [161; 168] можна здійснити за допомогою утиліт для середовища «Cygwin64», які є у вільному доступі на порталі PhysioNet [161]. Ці утиліти дозволяють, зокрема, налаштувати бажану довжину записів, оскільки портал PhysioNet обмежує цей параметр. Середовище «Cygwin64» також дозволяє обрати зручний текстовий формат імпортованого файлу.

Пошук артефактів у первинно відібраних записах реалізується за допомогою методів з бібліотеки «Statistics». За допомогою використання цих програмних методів були знайдені гігантські викиди даних (йдеться про декілька сотень значень сатурації крові киснем нижче 1 %, що фізіологічно не реально, принаймні для живих людей) в трьох записах. Також для обробки наборів даних реалізується з використанням методів з різних бібліотек: «LinearAlgebra», «Statistics», «SignalProcessing» та «plots».

4.1.2 Міжквартильні діапазони як найбільш надійний статистичний показник варіабельності

Так звані коробкові графіки (англ. box-and-whisker plots, box plots) є зручним засобом візуалізації варіабельності наборів даних. Вони спираються на статистичний розподіл даних на квартилі, показуючи міжквартильний діапазон у

вигляді висоти коробок, а повний розмах даних у вигляді вертикальних ліній з ризиками («вусів»). Викиди можуть бути показані окремим точками [15; 16]. Рис. 4.1 показує коробкові графіки (діаграми) для описаного вище набору даних SaO_2 .

Коробкова діаграма переконує в існуванні щонайменше трьох підгруп у наборі даних. Це підгрупа з малою (формально навіть нульовою) варіабельністю, яка складається з 6 суб'єктів {IDs 1; 2; 5; 10; 15; 16}. Підгрупа з відносно великою варіабельністю (приблизно 2 %), яка налічує 6 суб'єктів {IDs 3; 6; 11; 12; 13; 14}. Проміжна підгрупа, яка має середню (приблизно 1 %) варіабельність і налічує 8 суб'єктів {IDs 4; 7; 8; 9; 17; 18; 19; 20}. Різниця у типових варіабельностях між підгрупами є статистично значущою на рівні довіри 0,99.

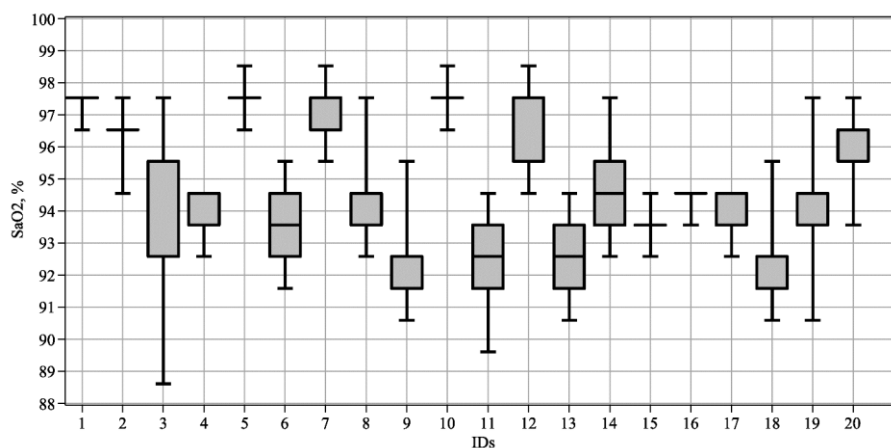


Рисунок 4.1 – Коробкова діаграма (графік) набору даних з сатурації киснем артеріальної крові (SaO_2)

Горизонтальна вісь на рис. 4.1 показує ідентифікаційні коди (IDs) суб'єктів (записів), вертикальна вісь – сатурацію у відсотках.

Нульовий міжквартильний діапазон для першої підгрупи виглядає дещо дивним на перший погляд. Втім, варто взяти до уваги кількість спостережуваних рівнів сатурації в записах для цієї підгрупи (від 2 до 4 рівнів). Тоді неможливість розподілу двох або трьох спостережуваних рівнів на чотири квартилі стає зрозумілою.

4.1.3 Диференціали записів (серій) SaO_2 та їх статистичний розподіл

Операція знаходження диференціалу першого порядку (інкременту) для часового ряду передбачає обчислення різності між сусідніми парами елементів серії (запису), починаючи з другого та першого елементів, третього та другого, і так далі. Диференціал є ненульовим, якщо наступний елемент серії відрізняється від попереднього. Ненульові диференціали вказують позиції, де відбувається зміна в елементах часового ряду (серії, запису).

Рис. 4.2 показує три серії (записи) вимірювань SaO_2 для трьох типових суб'єктів з трьох різних підгруп, описаних вище, у порядку зростання варіабельності зліва направо. Там же наведені графіки диференціалів цих трьох серій у тому ж порядку. Кількості ненульових диференціалів є такими: {10; 22; 37} відповідно.

Насамперед, зауважимо однакові амплітуди для всіх диференціалів, які приблизно дорівнюють ± 1 %. Це означає, що зміна спостережуваного рівня сатурації можлива лише в найближчий вищий, або ж нижчий рівень з певного обмеженого спектру таких рівнів (від 2 до 10 для суб'єктів обраного набору даних). Отже, спектр можливих рівнів сатурації артеріальної крові киснем може бути обмеженим і дискретним.

Вважатимемо подією будь-яку зміну (перемикання) можливого рівня сатурації крові. Тоді графіки диференціалів ілюструють потоки таких дискретних подій на одному і тому ж інтервалі часу ($t = 3$ хв), або відліків ($N = 36000$). Кількість таких подій (10–40) є малою у порівнянні з кількістю випробувань (N), отже це рідкі події.

Кожна така подія (перемикання) незалежна від інших, тобто не має післядії, впливу на наступні події. Кожна подія відбувається окремо від інших, а не парами, трійками, тощо, тобто потік подій є ординарним. Нарешті, ймовірність потрапляння того чи іншого числа подій на малий інтервал часу Δt залежить тільки від довжини інтервалу Δt і не залежить від того, де саме на осі часу він розташований. Отже, потік є стаціонарним.

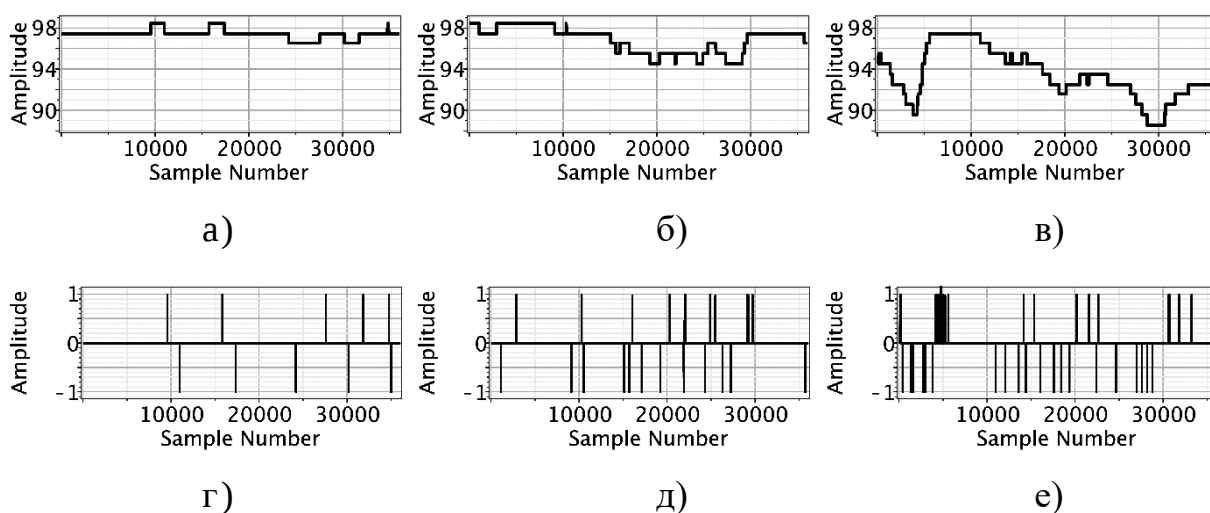


Рисунок 4.2 – Записи SaO₂ та їх диференціали для трьох типових суб'єктів трьох різних підгруп: а – запис SaO₂ для типового суб'єкта підгрупи з ID 5; б – запис SaO₂ для типового суб'єкта підгрупи з ID 12; в – запис SaO₂ для типового суб'єкта підгрупи з ID 3; г – диференціали записів для суб'єкту групи з ID 5; д – диференціали записів для суб'єкту групи з ID 12; е – диференціали записів для суб'єкту групи з ID 3; амплітуди даються у відсотках

Стаціонарний та ординарний потік подій без післядії описується розподілом Пуассона для числа подій на довільному фіксованому інтервалі (часу, або випробувань) [169]. Зрозуміло, що інтенсивність потоку, основний параметр розподілу Пуассона, зберігається лише в межах підгрупи. Вона повинна змінюватися від підгрупи до підгрупи, як це видно з рис. 4.2.

Отож, підгрупи можна формувати, виходячи з інтенсивності потоку перемикань, тобто підбираючи до підгрупи суб'єкти з близькими числами подій. Саме таким шляхом було виділено три підгрупи:

1) підгрупа малої інтенсивності перемикань з середньою інтенсивністю 11 подій на серію. Вона включає сім суб'єктів {IDs 1; 2; 5; 7; 10; 15; 16}. Ця група перетинається, але не повністю тотожна описаній вище в п. 4.1.2 групі з малою варіабельністю;

2) підгрупа проміжної інтенсивності перемикачів з середньою інтенсивністю 21 подія. Ця група включає 9 суб'єктів з {IDs 4; 6; 8; 11; 12; 14; 17; 19; 20};

3) підгрупа вищої інтенсивності: 37 подій в середньому. Вона включає чотири суб'єкти {IDs 3; 9; 13; 18}.

Різниця у середніх інтенсивностях подій між підгрупами є статистично значущою на рівні довіри 0,99.

Розподіл Пуассона дає ймовірність спостереження x подій на фіксованому інтервалі, якщо інтенсивність подій дорівнює λ :

$$\text{Pois}(x, \lambda) = \frac{\exp(-\lambda) \cdot \lambda^x}{x!}. \quad (4.1)$$

Ймовірність спостереження x подій у змішаному розподілі Пуассона [21]: для гетерогенного набору даних, який у нашому випадку складається з трьох підгруп з різними інтенсивностями подій, можна записати у такому вигляді:

$$P(x) = \sum_{i=1}^3 \left(\frac{n_i}{n} \right) \cdot \text{Pois}(x, \lambda_i), \quad (4.2)$$

де n_i – кількість суб'єктів в i -й підгрупі;

n – повна кількість суб'єктів у наборі даних, $n = 20$.

Гістограми є корисним статистичним інструментом відображення даних, дозволяючи уявити реальні розподіли ймовірностей. Для створення коректних гістограм важливі два параметри: ширина так званого «кошику» (або «смуги», регулярного інтервалу гістограми) та початкова точка першого такого інтервалу [170]. Стосовно ширини «кошику» існують певні емпіричні правила оцінки, залежні від варіабельності та кількості даних [171]. Неоднозначність вибору початкової точки можна усунути усереднюючи декілька ($m \leq 10$) разів зсунуті гістограми, або ж так звані «averaged shifted histograms, ASH» [170].

Рис. 4.3 показує ординарну та ASH-гістограму для досліджуваного набору даних та криві складного (змішаного) розподілу Пуассона на їх тлі.

Тримодальність фактичного розподілу ймовірностей та гетерогенність набору даних є наочними.

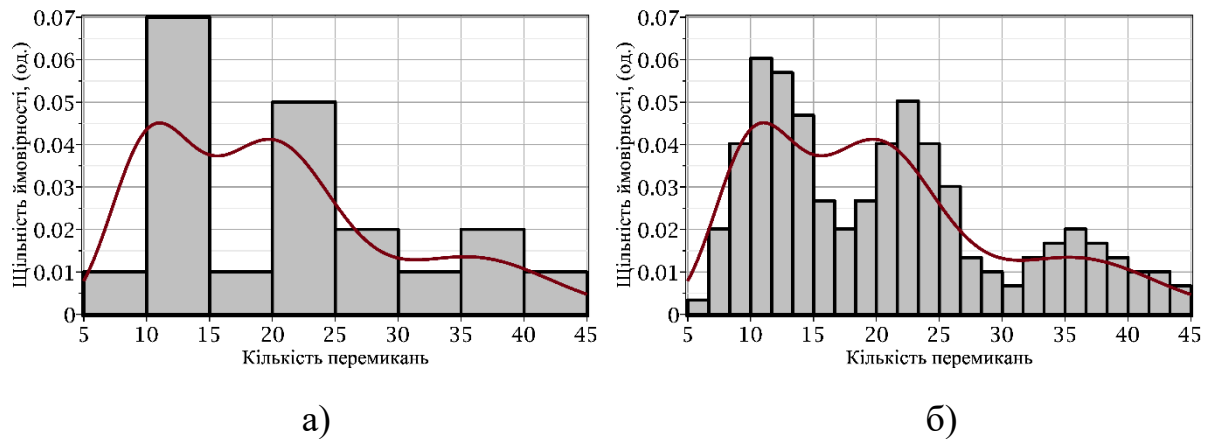


Рисунок 4.3 – Гістограми та розподіл Пуассона для досліджуваного набору даних: а – ординарна; б – ASH ($m = 3$)

На рис. 4.3 горизонтальна вісь показує кількість перемикачів у серії, вертикальна – ймовірності; криві показують ймовірності обчислені за (4.2) для змішаного розподілу Пуассона

Відповідність між гістограмами та кривою складного (змішаного) розподілу Пуассона виглядає цілком задовільною. Крива якісно вірно передає співвідношення мод та мінімумів гістограм. Варто відмітити, що ASH гістограма краще узгоджується з теоретичним розподілом, ніж ординарна, що було підкреслено в [170]. Вона також точніше дає положення мод та мінімумів.

4.1.4 Довжина інтервалів між перемиканнями та їх статистичний розподіл

Математична статистика стверджує, що інтервали між подіями потоку Пуасона описуються розподілом Ерланга [169]:

$$\text{Erl}(x, \lambda, n) = \frac{\lambda^n x^{n-1}}{(n-1)!} \cdot \exp(-\lambda x), \quad (4.3)$$

де x – довжина інтервалу між послідовними перемиканнями;

n – кількість випробувань ($n = 36000$);

λ – має той же зміст, що і в розподілі Пуассона (4.1).

Рис. 4.4 показує ASH-гістограму ($m = 3$) та графік складного (змішаного) розподілу Ерланга, отриманого аналогічно складному розподілу Пуассона. Відповідно до центральної граничної теореми теорії ймовірностей, для такого великого значення параметру $n \gg 1$ розподіл Ерланга добре апроксимується нормальним розподілом із середнім значенням n/λ та дисперсією n/λ^2 .

Узгодження гістограми та складного розподілу Ерланга виглядає задовільним. Як гістограма, так і теоретичний розподіл підтверджують гетерогенність набору даних та існування в ньому підгруп.

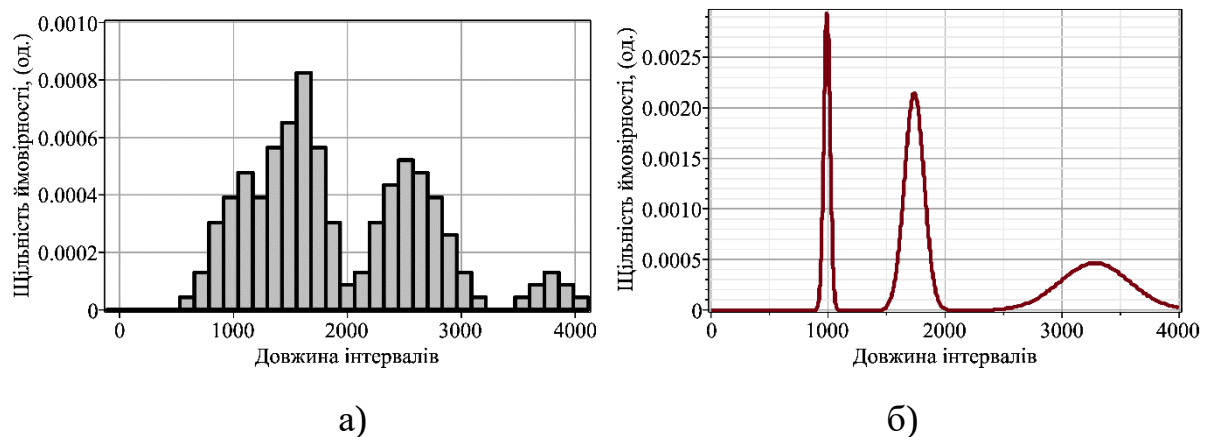


Рисунок 4.4 – Діаграми: а – ASH-гістограма ($m = 3$) довжин інтервалів між перемиканнями; б – графік змішаного розподілу Ерланга для гетерогенного набору даних

Математичне очікування (середнє значення) для розподілу Ерланга дорівнює відношенню n/λ , а стандартне відхилення $n^{0.5}/\lambda$. У табл. 4.1 ці теоретичні інтегральні показники порівнюються з усередненими експериментальними в кожній з трьох підгруп.

Таблиця 4.1 – Порівняння теоретичних та фактичних оцінок для середніх довжин інтервалів між перемиканнями та їх стандартних відхилень по трьом підгрупах

Тип значень	Підгрупа малої варіабельності	Проміжкова підгрупа	Підгрупа вищої варіабельності
Теорія	3273 ± 17	1732 ± 9	986 ± 5
Фактично	2806 ± 504	1758 ± 454	982 ± 105

У табл. 4.1 довжини інтервалів та їх стандартні відхилення даються у відліках. Можна констатувати, що середні довжини інтервалів між перемиканнями задовільно узгоджуються з їх теоретичними оцінками. Проте, стандартні відхилення розрізняються на порядок і більше. Причиною є мала кількість суб'єктів у підгрупах тому усереднені в межах підгруп величини (нижній рядок табл. 4.1) мають досить великі дисперсії. Отже, уточнення цих величин, як і середніх інтенсивностей перемикань, потребують збільшення кількості записів (суб'єктів) у наборі даних.

Якщо інтенсивності перемикань, отримані вище для кожної з трьох підгруп {11; 21; 37} перемикань на 3-хвилинний інтервал спостереження відповідно, перерахувати в частоти, то можна отримати {0,06; 0,11; 0,20} Гц відповідно. Ці частоти лежать у смугах F1–F2 частотної класифікації [172]. Рис. 4.5 представляє періодограми трьох функцій нижнього ряду рис. 4.2 у цих частотних смугах.

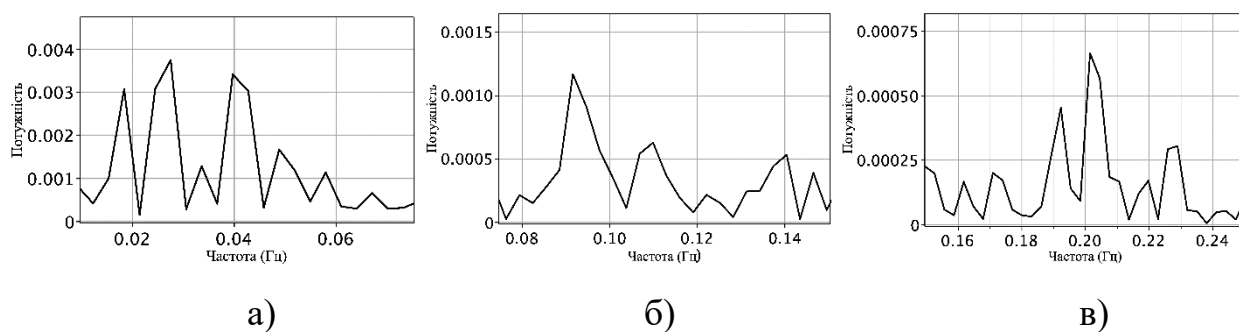


Рисунок 4.5 – Періодограми, отримані швидким перетворенням Фур'є диференціалів записів SaO_2 , показаних на рис. 4.2: а – типового суб'єкта підгрупи з ID 5; б – типового суб'єкта підгрупи з ID 12; в – типового суб'єкта підгрупи з ID 3

На рис. 4.5 квадрати амплітуд Фур'є (вертикальна вісь) задані у відносних одиницях. Виявлено характерні структури: дублет піків в діапазоні 0,04–0,05 Гц на графіку ліворуч, дублет в діапазоні 0,095–0,11 Гц на середньому графікові, дублет в діапазоні 0,18–0,20 Гц на правому. Ширини цих діапазонів є порівняльними з теоретичною роздільною здатністю по частоті 0,006 Гц, яка визначається частотою дискретизації 200 Гц та кількістю точок для швидкого перетворення Фур'є $215 < n$. Діапазони частот можна вважати достатньо

близькими до наведених вище оцінок: $\{0,06; 0,11; \text{та } 0,20\}$ Гц з урахуванням *величини роздільної здатності*.

4.1.4 Огляд отриманих результатів дослідження даних SaO_2

Функції, показані на рис. 4.2, мають цікаві особливості: їх перші похідні по аргументу (кількості відліків, або часу) дорівнюють нулю майже всюди на інтервалі спостереження. «Майже» в цьому реченні означає: «за виключенням дискретного набору точок (позицій) міри нуль» [171]. Цей дискретний набір точок є позиціями перемикачів, показаними на тому ж рис. 4.2 у нижньому рядку. Сам по собі цей набір точок (позицій) є асимптотичне безкінечно малою часткою всього сигналу і має виразно неоднорідну структуру, тобто є фракталом певної Хаусдорфової розмірності сам по собі [171; 173]. Фрактальна природа серій SpO_2 відмічалася раніше в [53], а для SaO_2 у роботі [55].

Фрактальна природа означає, що фрагмент такого набору, скажімо, ті позиції, які належать лівій половині інтервалу спостереження, статистично еквівалентний цілому набору. Отже, має ту саму інтенсивність подій (перемикачів між рівнями сатурації), або ж середній інтервал поміж ними, як і повний набір. Таким чином, масштабування (зменшення, чи навпаки, збільшення) довжини інтервалу спостереження не може суттєво змінити інтенсивність перемикачів, або середній інтервал поміж ними для підгруп, які були визначені вище. Вони є інваріантами масштабування, принаймні в тому інтервалі масштабів, для якого виконується ступеневий закон скейлінгу [173]. Отже, саме фрактальність моментів перемикачів рівнів сатурації крові киснем є ключем до розуміння отриманих результатів.

Варто відмітити узгодженість інтегральних показників варіабельності, запропонованих в цій роботі (частота перемикачів та інтервали часу між ними) як між собою, так і з міжквартільними діапазонами, найбільш надійним статистичним показником варіабельності. Існує передбачувана сильна і негативна кореляція між кількістю перемикачів та середніми інтервалами поміж ними. Коефіцієнт лінійної кореляції дорівнює $-0,888$ і є статистично значущим

на довірчому рівні 0,99. Отже, чим більшим є один показник, тим менше є інший, і навпаки.

Періодограми, що представлені на рис. 4.5, свідчать, що характерні частоти перемикань між можливими рівнями сатурації крові киснем є малими (перші частотні смуги класифікації [172]). З одного боку це підтверджує статистичну рідкість таких подій. З іншого, дозволяє працювати зі значно меншими частотами дискретизації, ніж ті, що звичайно використовуються, без втрати суттєвої інформації про процеси перемикання рівнів оксигенації.

Отримані результати узгоджуються з результатами [19] стосовно поділу наборів даних на три підгрупи, хоча в тій роботі розглядався інший набір даних та аналізувалися вимірювання SpO_2 . Втім, питома вага підгруп в наборах дещо різна: {35; 45; 20} % для набору даних SpO_2 проти {22,5; 55; 22,5} % для набору даних SpO_2 [19]. Питома вага підгрупи з вишкою варіабельністю приблизно однакова, але питома вага підгрупи з найменшою варіабельністю помітно менша на користь проміжної підгрупи в результатах [19]. Малі ємності наборів – 20 або 36 суб'єктів [19] – є обмеженням точності оцінок існуючих досліджень.

Виходячи з результатів цього дослідження, можна окреслити дві важливі задачі для подальшого вирішення:

- 1) чи змінюється (і як саме) спектр спостережуваних рівнів сатурації а/або варіабельність сатурації при патологіях, зокрема важких формах COVID-19;
- 2) чи рівномірно ризик важких форм COVID-19 розподілений по підгрупах з різною варіабельністю. Можливо, існує більш уразлива підгрупа, зокрема підгрупа з максимальною варіабельністю оксигенації крові.

4.2 Використання статистичних методів, поєднаних з рекурентним аналізом для аналізу значень рівня SpO_2 у здорових дорослих

4.2.1 Огляд датасету SpO_2 , що буде аналізовано

Як було зазначено в попередньому підрозділі, пульсоксиметрія (SpO_2) є неінвазивним методом моніторингу рівня оксигенації крові та уникає прямого газового вимірювання артеріальної крові (SpO_2). Пульсоксиметрія пришвидшує

виявлення гіпоксемії або кисневої недостатності в артеріальній крові. Критичний рівень $SpO_2 < 95\%$ [53; 161; 174]. Існує також думка, що критичний рівень є трохи нижчим ($SpO_2 < 92\%$ [175]).

Дані, що будуть використані, походять з [53; 161]. У табл. 4.2 наведено вміст так званих заголовних файлів, що коротко описують учасників вимірювання за «апріорно» виділеними ознаками.

Таблиця 4.2 – Учасники SpO_2 вимірювань

№ з/п	Код	Стать	Статус курця	Індекс маси тіла	Вік	Час заміру, с
1	301116B	Ч	Не курець	21,7	20	3607
2	301116A	Ч	Не курець	26,5	42	3642
3	051216A	Ч	Не курець	22,0	20	3616
4	081216A	Ч	Не курець	23,6	20	3606
5	101216A	Ж	Не курець	24,1	45	3623
6	101216B	Ч	Не курець	26,8	49	3637
7	101216C	Ч	Не курець	23,2	22	3610
8	121216A	Ч	Не курець	20,6	20	3600
9	121216B	Ж	Не курець	23,1	21	3600
10	230117A	Ч	Курець	26,8	23	3610
11	230117B	Ж	Колишній курець	17,9	21	3604
12	250117A	Ж	Не курець	21,3	35	3610
13	250117B	Ч	Не курець	23,0	20	3697
14	250117C	Ч	Колишній курець	28,4	55	3610
15	300117A	Ж	Не курець	20,3	20	3605
16	010217A	Ж	Не курець	20,5	23	3587
17	010217B	Ж	Не курець	21,7	21	3648
18	010217C	Ж	Курець	23,8	45	3606
19	070217A	Ж	Не курець	19,8	19	3596
20	080217A	Ж	Не курець	24,7	22	3610
21	080217B	Ж	Не курець	20,7	20	3607
22	090217A	Ж	Не курець	28,2	41	3600
23	090217B	Ж	Не курець	24,4	20	3610
24	160217A	Ч	Колишній курець	24,5	60	3594
25	160217B	Ж	Колишній курець	18,5	56	3616
26	160217C	Ж	Не курець	24,2	24	3610

№ з/п	Код	Стать	Статус курця	Індекс маси тіла	Вік	Час заміру, с
27	160217D	Ч	Не курець	23,9	45	3601
28	160217E	Ж	Не курець	19,8	38	3610
29	210217A	Ч	Курець	24,6	21	3609
30	210217B	Ч	Не курець	26,7	23	3600
31	210217C	Ж	Не курець	18,9	21	3600
32	010317A	Ж	Колишній курець	23,2	66	3589
33	010317B	Ч	Не курець	25,1	70	3610
34	140317A	Ч	Не курець	26,2	42	3598
35	150317A	Ж	Не курець	24,3	62	3600
36	150317B	Ч	Не курець	19,9	48	3599

Частота дискретизації дорівнювала 1 Гц. Усі записи мають тривалість близько години (близько 3600 зразків). Мінімальна довжина становить 3587 семплів (запис № 15 з вихідним кодом 010217A). Тому всі серії були обрізані до однакової парної довжини $N = 3586$ вибірок для створення матриці даних. Використовуючи ці дані планується переконатися в варіабельності крові киснем, знайти ризиковану підгрупу серед усіх наявних записів а також інші підгрупи серед здорових дорослих, якщо вони існують.

Перш ніж розпочати дослідження та аналіз даних на варіабельність, треба визначити встановимо прийнятне значення цього терміну, щоб уникнути непорозумінь. Будемо притримуватися філософії [169], у рамках ідей якої мінливість (варіабельність), дисперсія та розкид є синонімами та відносяться до ширини розподілу щільності ймовірності. Найбільш популярними інструментами для дослідження мінливості є діапазони, інтерквартильні діапазони (Interquartile Ranges, IQR) і стандартні відхилення (Standard Deviations, SD) [176]. Звичайно, що функції щільності ймовірності (Probability Density Functions, PDF) також підходять.

Статистика не повинна обмежувати дослідження варіабельності сама по собі, хоча це входить до її області. Для вивчення варіабельності насичення киснем (Oxygen Saturation Variability, OSV) у [53] використовувалися статистичні методи обробки даних, аналіз детрендованих коливань (Detrended

Fluctuation Analysis, DFA), дослідження ентропії та аналіз графіків Пуанкаре. Статистичний аналіз [53] застосовувався до даних «як є» без будь-якої фільтрації. При цьому автори вважають за краще використовувати такі статистичні дані, як середні значення та стандартні відхилення. Вони добре підходять лише для розподілу ймовірностей, близьких до нормального.

Статистика, графіки Пуанкаре та криві Ендрю використовувалися для деяких випадків з обробкою даних тієї самої бази даних [53; 161] у [10]. Цей експрес-аналіз показав негаусівський, багатомодовий і спотворений розподіл ймовірностей. Крім того, графіки Пуанкаре виявилися сильно кластеризованими. Під сильною кластеризацією в даному випадку розуміють абсолютні частоти повторень для деяких значень SpO_2 , що перевищують тисячу разів [10]. Погодинні записи з частотою дискретизації 1 Гц [53; 161] показують такі часті рецидиви в кожній серії.

Деякі автори вважають, що графіки Пуанкаре правильні та дійсні лише за відсутності будь-яких кластерів як таких. Вони стверджують, що графік Пуанкаре повинен складатися лише з одного геометрично симетричного кластера [119]. Звичайно, цього не може бути, коли ми говорили про дані оксигенації крові та їх сильно кластеризовані графіки Пуанкаре [10].

Звичайно, необхідна початкова фільтрація даних. Це можливо різними способами. Одним з них є цифрова вейвлет-фільтрація, яка дозволяє розділити ряд на низькочастотну і високочастотну частини. Існує багато відповідних вейвлетів. Проте вже перший погляд на сигнали, зміни яких мають форму прямокутників [10; 53; 161], дасть аргументи для вейвлетів Хаара. Високочастотна частина серії зазвичай є просто шумом, тому фільтрація є одночасно засобом знешумлення.

Далі слід обирати такі статистичні методи обробки даних, які краще відповідають особливостям розподілу ймовірностей рядів даних. Ми більше не можемо вважати розподіли ймовірностей даних [53; 161] близькими до нормального.

Крім того, OSV і часті рецидиви деяких рівнів SpO_2 в записах можуть бути пов'язані. Повторюваність і мінливість навпаки: якщо один зростає, то інший падає і навпаки. У сфері теорії хаосу ця ідея знайшла своє зростання як сюжет повторення [111; 177]. Графіки повторень, представлені в [177], і їх дескриптори широко використовуються як інструменти нелінійного аналізу для подібних складних динамічних систем із помітними повтореннями [111; 156; 178].

4.2.2 Первинна обробка датасету

Було виконано цифрову фільтрацію вейвлетами Хаара матриці даних із двократним зменшенням дискретизації. Вейвлети Хаара виглядають як найбільш придатні для досить специфічної форми цих сигналів [10; 53; 161]. Результатом фільтрації стали два масиви розміром (1793×36) . Вони представляють НЧ- і ВЧ-частини сигналів бази даних [53; 161].

Також застосовано W-тест Шапіро та Вілка до всіх 36 стовпців набору НЧ-даних. Цей добре відомий тест дозволяє визначити, наскільки дана вибірка (серія для одного з учасників) відповідає нормальному розподілу. Нульова гіпотеза «вибірка зібрана з популяції, яка дотримується нормального розподілу» виявилася хибною для всіх учасників табл. 4.2 з імовірністю 0,99.

Більшість записів (27 із 36) мають від'ємну асиметрію з діапазону $(-2,90 \pm 1,15)$ з найбільш ймовірним значенням $-0,50$. Таким чином, показники розташування (мода, медіана та середнє) не є рівними. Це не критично, оскільки різниця між модою та медіаною, наприклад, не перевищує 0,42 % у гіршому випадку.

Більше проблем приносить надлишковий ексцес. Більшість записів набору даних – 26 з 36 – мають позитивний надлишковий ексцес: в діапазоні $(-0,93 \pm 13,45)$ з найбільш ймовірним значенням $+0,22$. Цей факт вказує на серйозну проблему з викидами в багатьох записах цього набору даних, незважаючи на заявлену в [53] попередню очистку візуальних артефактів. Багаторежимна структура на рис. 4.6 свідчить про внутрішню структуру набору даних. Існування підгруп є можливим.

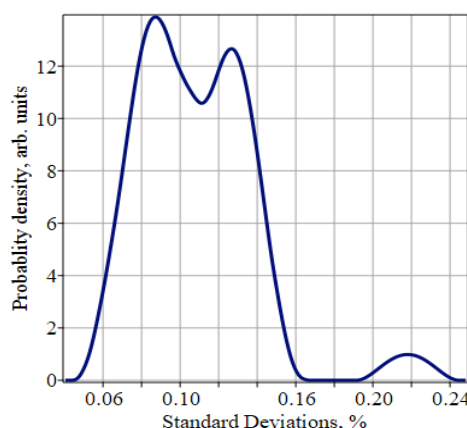


Рисунок 4.6 – Функція щільності ймовірності (PDF) для стандартних відхилень ВЧ-частини відфільтрованих сигналів

Стандартне відхилення є одним із показників мінливості. Рис. 4.6 відображає короточасну мінливість (пропорційну дескриптору $SD1$ [10], [119]). Розподіл виглядає позитивно зміщеним і має три режими. Найбільш вірогідне стандартне відхилення (основна мода PDF на рис. 4.6) дорівнює приблизно 0,089 %. Це значення знадобиться пізніше як поріг подібності двох випробувань одного запису.

Найбільш імовірне стандартне відхилення для НЧ-частин сигналу дорівнює 0,71 % (рис. 4.7). Це значення пропорційне $SD2$ [10; 119]. Відношення $SD1/SD2$ дорівнює приблизно 0,125 і оцінює випадковий внесок у сигнал через ВЧ-шум. Це так зване R-відношення, відоме в техніці графіків Пуанкаре [10; 119]. Це співвідношення також показує ефективність фільтрації. Далі необхідно оперувати тільки очищеною LF-частиною бази даних.

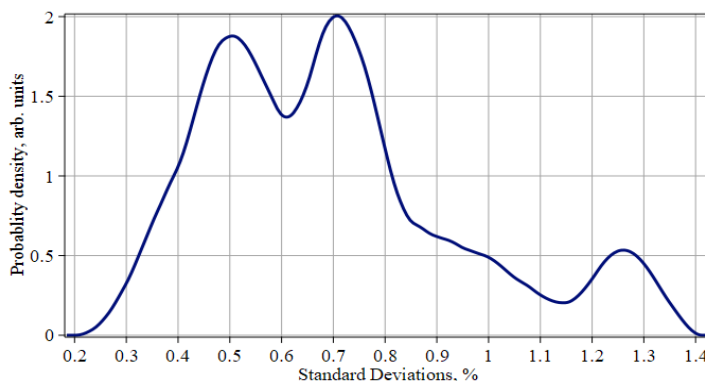


Рисунок 4.7 – Функція щільності ймовірності (PDF) для стандартного відхилення НЧ-частини відфільтрованих сигналів

Функція щільності ймовірності (PDF) на рис. 4.7 є подібною, як і на рис. 4.6.

На рис. 4.8 представлено статистичну діаграму всіх наборів даних. Можна побачити явну різницю між зразками (серії, записи окремих учасників). Вони відрізняються як дескрипторами мінливості, наприклад, діапазонами та інтерквартильними діапазонами (IQR), так і за показниками розташування, наприклад, медіанами. Ці відмінності можуть означати існування підгруп у розглянутому наборі даних.

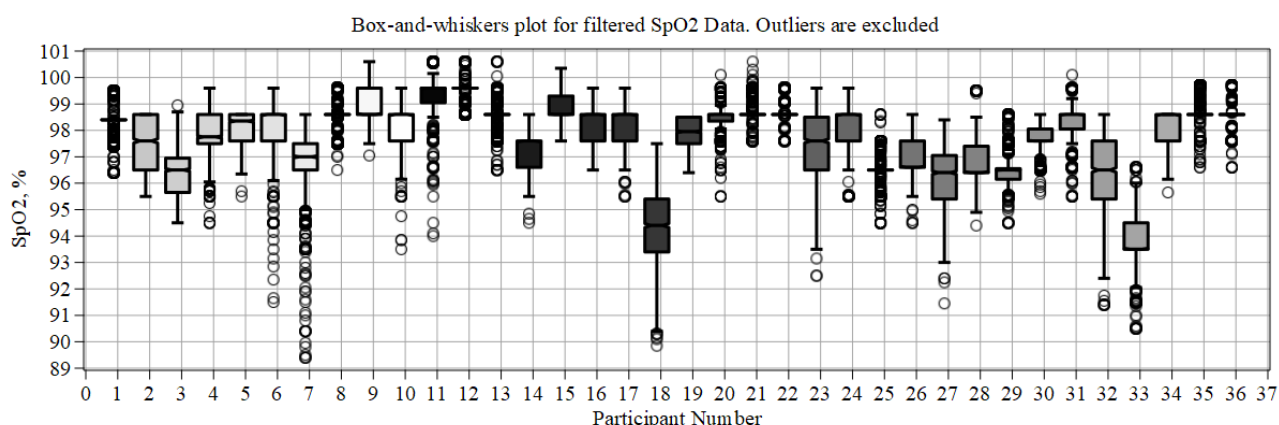


Рисунок 4.8 – Статистичний коробковий графік датасету

Висота коробок на рис. 4.8 показує інтерквартильні діапазони зразка (IQR), «вуса» вказують на діапазони зразка, кружечки представляють викиди. Горизонтальні лінії всередині прямокутників показують вибірові медіани.

Викиди, що позначені кружечками на рис. 4.8, властиві практично кожній вибірці, будучи індексами негаусового характеру розподілу ймовірностей та ефекту надлишкового ексцесу. Вони здебільшого впливають на діапазони вибірок, але менше на стандартні відхилення та інтерквартильні діапазони (IQR) [169].

4.2.3 Методи, застосовані для обробки

Статистичний аналіз і аналіз рекурентних графіків були двома основними методами цієї роботи. Зі сказаного вище випливає перевага режиму серії як основної міри розташування. Також віддамо перевагу стандартним відхиленням й інтерквартильним діапазонам як показникам мінливості замість діапазонів. Тут

справа в викидах, які сильно впливають на діапазони і роблять цей показник грубим і ненадійним [169]. Усі ці дескриптори мають відсотки як одиниці вимірювання.

Графіки повторюваності є зображеннями так званих матриць подібності [111; 156; 177; 178]. Ці двовимірні квадратні матриці з елементами $(S_{i,j})$, що дорівнюють «0» або «1», будуються за такою умовою:

$$S_{i,j} = \theta \left(\varepsilon - \text{abs}(x_i - x_j) \right),$$

$$i, j = 1.. \frac{N}{2}, \quad (4.4)$$

де x_i, x_j – пара випробувань з одного й того ж запису;

ε – порогова відстань подібності, згадана вище;

$\theta(\varepsilon - \text{abs}(x_i - x_j))$ – функція Хевісайда.

Якщо працювати з обрізаними та відфільтрованими даними, тоді $\frac{N}{2} = 1793$. При цьому всі матриці подібності та їхні зображення мають однакові розміри.

Класичний аналіз рекурентної квантифікації (Recurrence Quantification Analysis, RQA) завжди починається з коефіцієнта рекурентності (RR). У найпростішій версії він визначається як відношення суми ненульових елементів матриці подібності до загальної кількості її елементів. Іноді з матриці подібності виключають її діагональні елементи, однак цього робити не будемо. По-перше, це впливає на значення RR лише незначною мірою, принаймні для великих розмірностей, які тут розглядаються. По-друге, RR збігається із середньою інтенсивністю графіка рекурентності, що є більш зручним для візуального аналізу.

Кількісна оцінка графіків рекурентності (RP) включає широкий спектр більш детальних числових дескрипторів [111; 156; 178]. Крім того, якісний текстурний аналіз графіків рекурентності також має самостійну цінність. Однак такі підходи виходять за межі цілей, поставлених у цьому дослідженні, і тому були відхилені.

Зазначимо, що традиційна нумерація рядків і стовпців у матрицях подібності зазвичай починається з нижнього лівого кута [111; 156; 178]. Відповідно, їхні головні діагоналі проходять від нижнього лівого до верхнього правого кута матриці. В дослідженні використано протилежне представлення матриці: нумерація починається з верхнього лівого кута. Це означає, що головна діагональ матриці проходить від верхнього лівого до нижнього правого кута, як у звичайних матрицях. Це варто мати на увазі при вивченні графіків рекурентності.

Таблиця частотної статистики навіть для одного запису з набору даних може проілюструвати причини використання аналізу рекурентності. Розглянемо вибірку № 21 (код 080217В). Цей відфільтрований запис містить 1793 значення, але серед них лише 34 є унікальними. Ці 34 унікальні значення повторюються з частотою від одного разу до понад тисячі разів кожне. Особливо часто зустрічається рівень насичення киснем 98,6 % – він повторюється 1443 рази. Тому наявність лише кількох десятків кластерів на графіках Пуанкаре [10] не є несподіванкою.

4.2.4 Вибір підгруп та їх статистична характеристика

Розглянемо розподіли двох дескрипторів варіабельності: маються на увазі стандартні відхилення для відфільтрованих низькочастотних (LF) частин сигналів і відповідні міжквартильні розмахи (IQR). Учасники, які одночасно належать до нижнього квартиля за IQR і мають стандартні відхилення нижче 0,6 % (див. рис. 4.7), формують підгрупу з низькою варіабельністю рівня кисню в крові. До цієї підгрупи увійшли вісім учасників із номерами {1; 8; 12; 13; 21; 22; 35; 36} згідно з табл. 4.2.

Інша підгрупа з відносно високою варіабельністю включає вісім учасників із номерами {2; 3; 7; 18; 23; 27; 32; 33}. Вони належать до верхнього квартиля за IQR і мають стандартні відхилення понад 0,85 %.

Більшість учасників (20 осіб) входять до підгрупи із середньою варіабельністю. Вони займають другий і третій квартилі за IQR і мають

стандартні відхилення в межах 0,6–0,85 %, що відповідає центральному максимуму розподілу на рис. 4.7.

Підгрупа з низькою варіабельністю включає чотирьох чоловіків і чотирьох жінок, серед яких є як молоді люди (віком <35 років), так і представники старшого віку. До цієї групи входять особи з індексом маси тіла в межах загальноприйнятої норми (18–25), а також одна людина з невеликим надлишком ваги. Всі вони не палять.

Підгрупа з високою варіабельністю відрізняється лише наявністю одного курця та однієї людини, яка раніше палили. Крім того, спостерігається незначне переважання чоловіків і старших людей (5 проти 3). Отже, ці фактори навряд чи є визначальними. Основним фактором, схоже, є саме варіабельність рівня насичення крові киснем серед учасників.

Таблиця 4.3 – Найбільш ймовірні значення (моди) рівня насичення крові киснем, його стандартне відхилення та міжквартильний розмах (IQR) для кожної з трьох підгруп

Subgroups	Blood oxygenation, %	Standard deviation, %	IQR, %
Low-variability	98,6	0,42	0,0
Middle variability	98,6	0,71	1,0
High-variability	96,5	0,98	2,0

Найбільш ймовірний рівень насичення крові киснем є однаковим для підгруп з низькою та середньою варіабельністю. Однак інші відмінності є статистично значущими з імовірністю (довірчим рівнем) не менше 0,97. Ці результати отримано за допомогою двовибіркових Т-тестів.

Таким чином, підгрупа з високою варіабельністю включає здорових дорослих із нижчим рівнем кисневої сатурації, але з вищою її варіабельністю. Ця група, що має слабшу систему насичення крові киснем, складає 8 осіб із 36, що приблизно становить 22 % популяції. Приблизно така ж частина здорових дорослих (близько 22 %) має високий рівень кисневої сатурації та низьку її

варіабельність, що свідчить про стабільнішу та ефективнішу систему оксигенації.

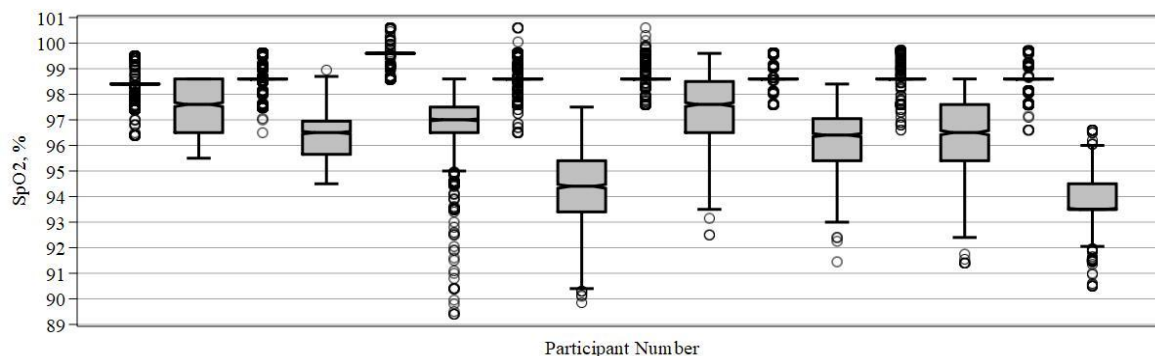


Рисунок 4.9 – Візуальна різниця між двома підгрупами з нижчою та вищою варіабельністю («сильна» та «слабка»)

На рис. 4.9 розташування зразків чергується за принципом «шахової дошки»: {1; 2; 8; 3; 12; 7; 13; 18; 21; 23; 22; 27; 35; 32; 36; 33}. Треба звернути увагу на вкрай малий міжквартильний розмах (IQR) для «сильнішої» підгрупи та асиметричні викиди у "слабшій" підгрупі. Тут переважають викиди, спрямовані вниз.

4.2.5 Рекурентні графіки та коефіцієнти рекурентності для підгруп

Графіки рекурентності є зображеннями квадратних матриць подібності, розмір яких відповідає низькочастотним (НЧ) частинам відфільтрованих сигналів (1793×1793). Чорні пікселі відповідають елементам матриці, що дорівнюють «1», а білі – елементам, що дорівнюють нулю.

Усі матриці подібності мають однаковий поріг подібності (див. вираз (4.4)). Теоретично, кожна підгрупа могла б мати свій власний поріг, однак це ускладнює порівняння і робить його менш прозорим.

Існує два основних підходи до вибору порога. Один із них передбачає використання певної частини діапазону вибірки, наприклад, п'яти або одного відсотка. Однак проблема викидів у наборі даних ускладнює застосування цього підходу. Другий підхід базується на використанні стандартних відхилень. Ми обрали досить низький поріг (0,089 %, див. підрозділ 2.1). Такий вибір дозволяє враховувати лише справді близькі (повторювані) випробування.

Графіки рекурентності (RPs) візуально відрізняються за інтенсивністю, якщо вони належать до різних підгруп (див. рис. 4.10). Значення коефіцієнтів рекурентності (RR) становлять $\{0,597; 0,383; 0,160\}$ відповідно та відповідають учасникам № 22, № 32 і № 28 з табл. 4.2.

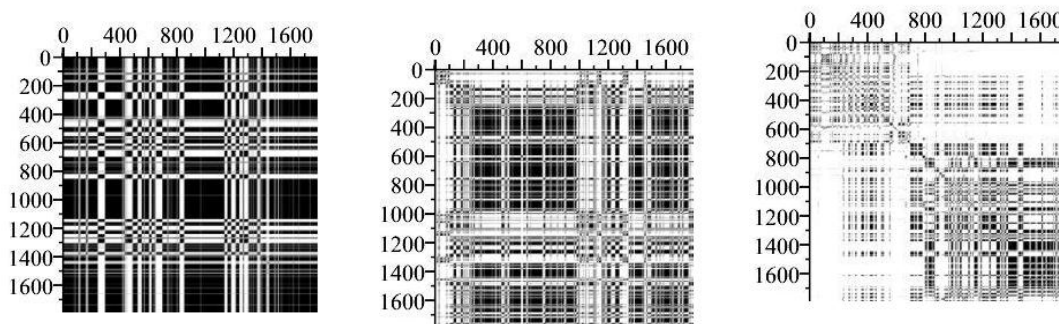


Рисунок 4.10 – Приклади рекурентних графіків (RP) для підгруп: ліворуч – «сильніша» підгрупа з низькою варіабельністю, центр– підгрупа з середньою варіабельністю, праворуч – «слабша» підгрупа

На рис. 4.10 кожному рекурентному графіку притаманні унікальні текстури та інтенсивності, що відображають характерні особливості варіабельності кисневої сатурації в кожній із підгруп. Тест W Шапіро-Вілка показує, що розподіли значень коефіцієнта рекурентності (RR) у всіх підгрупах наближені до нормального. Це є важливим, оскільки відсутність викидів дозволяє довіряти інтервалам. Рис. 4.11 містить коробковий графік для значень RR у трьох підгрупах. Викиди відсутні.

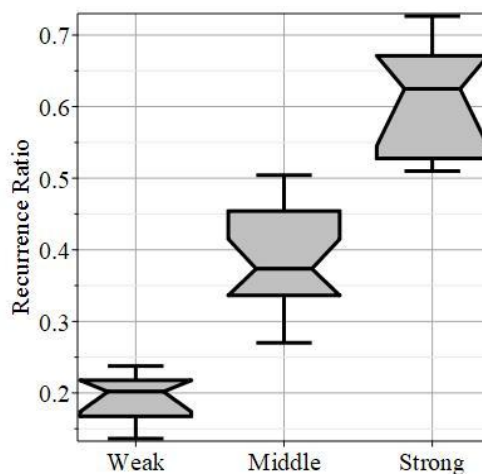


Рисунок 4.11 – Статистичний коробковий графік з вусами

На рис. 4.11 статистичний коробковий графік з вусами показує значення коефіцієнта рекурентності (RR) для трьох підгруп у наборі даних; також треба звернути увагу, що інтервали не перекриваються. Двовибірковий Т-тест підтвердив значущість відмінностей середніх значень з імовірністю (довірчим рівнем) не менше 0,99.

Діапазони значень коефіцієнта рекурентності (RR) для підгруп є наступними:

- «слабка» підгрупа з високою варіабельністю: (0,135–0,238);
- середня підгрупа: (0,270–0,504);
- «Сильна» підгрупа з низькою варіабельністю: (0,510–0,726).

Таким чином, середнє значення коефіцієнта рекурентності для «слабкої» підгрупи приблизно вдвічі менше, ніж для середньої підгрупи, і приблизно втричі менше, ніж для «сильної» підгрупи. Це можна побачити у вигляді різниці в інтенсивності на рис. 4.10, навіть без урахування відмінностей у текстурах.

Вище зазначено, що коефіцієнт рекурентності і варіабельність мають бути негативно корельованими. Треба визначити, наскільки це суттєво для досліджуваного набору даних. Рис. 4.12 відображає відносно сильну негативну кореляцію між стандартними відхиленнями відфільтрованих зразків і їхнім коефіцієнтом рекурентності (RR).

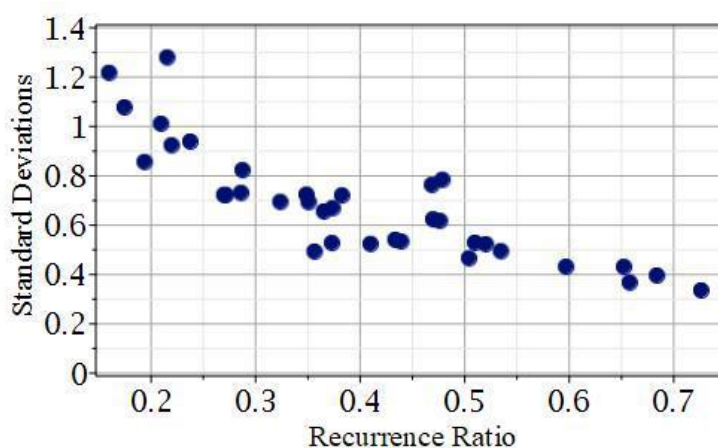


Рисунок 4.12 – Точковий графік кореляції стандартних відхилень із коефіцієнтом рекурентності (RR)

На рис. 4.12 кореляція стандартних відхилень із коефіцієнтом рекурентності (RR) є негативною, так само як і кореляція з міжквартильним розмахом (IQR). Треба звернути увагу, що критичне значення коефіцієнта кореляції Пірсона, яке має бути перевищене на рівні довіри 0,99, становить приблизно 0,39 для набору з 36 зразків. Це значення можна отримати з таблиці критичних значень t-розподілу Стюдента. Отже, кореляції, наведені в табл. 4.4, слід вважати суттєво відмінними від нуля.

Таблиця 4.4 – Коефіцієнти кореляції між коефіцієнтом рекурентності та статистичними характеристиками набору даних

Data set statistics	Standard deviations	Interquartile ranges	Modes
Correlation cieff.	– 0,86	– 0,83	+ 0,70

Табл. 4.4 підтверджує відносно сильну негативну кореляцію між коефіцієнтом рекурентності (RR) та показниками варіабельності (стандартними відхиленнями та IQR). Крім того, існує помірна, але значуща, позитивна кореляція з модами зразків. Таким чином, низькі індекси RR забезпечують вищу варіабельність і нижчу моду (найбільш ймовірне значення) насичення крові киснем.

Висновки до розділу 4

Четвертий розділ присвячений методам та засобам обробки рівня оксигенації крові пацієнтів, зокрема аналізу сигналів SpO_2 та SaO_2 , використанню статистичних методів та рекурентного аналізу для оцінки варіабельності кисневої сатурації. У ході дослідження вдалося покращити методики оцінки варіабельності SpO_2 , зокрема шляхом застосування міжквартильного діапазону (IQR) як надійного статистичного показника, що демонструє більшу стійкість до викидів у порівнянні зі стандартним відхиленням. Також було запроваджено диференціальний підхід до аналізу сигналів оксигенації, який враховує довжину інтервалів між перемиканнями

рівня SaO_2 та їхній статистичний розподіл. Це дало змогу більш точно оцінювати динаміку змін оксигенації крові та розробити нові критерії для моніторингу стану пацієнтів.

Запропоновано методику кластеризації пацієнтів за рівнем варіабельності SpO_2 на основі статистичних інтегральних показників та рекурентного аналізу. Виявлено три групи осіб із різною варіабельністю кисневої сатурації: перша підгрупа (22 % вибірки) характеризується найнижчим рівнем оксигенації та найвищою варіабельністю, що може бути важливим індикатором ризику, особливо у контексті COVID-19; друга підгрупа (56 % вибірки) має середню варіабельність і демонструє стабільніші рівні SpO_2 ; третя підгрупа (22 % вибірки) характеризується мінімальною варіабельністю, що свідчить про сталість рівня оксигенації. Крім того, доведено ефективність персоналізованого аналізу SpO_2 за допомогою рекурентних графіків, що дозволяють отримати унікальний «портрет» пацієнта для подальшого прогнозування змін у його фізіологічному стані.

Результати дослідження підтвердили, що розподіл ймовірностей SpO_2 є негаусівським і супроводжується значною кластеризацією значень, а його графіки Пуанкаре демонструють чітко виражені групи точок, що свідчить про дискретну природу переходів між рівнями сатурації. Аналіз частот показав, що значення SpO_2 98,6 % є найбільш поширеним (1443 рази у записі), що підтверджує наявність стабільних режимів оксигенації крові. Встановлено статистично значущу негативну кореляцію між варіабельністю SpO_2 (оціненою за IQR та стандартним відхиленням) і коефіцієнтом рекурентності (Recurrence Ratio, RR), що може бути використано як діагностичний критерій. Додатково доведено, що коефіцієнти рекурентності мають гаусівський розподіл без викидів, що дозволяє використовувати їх як надійний параметр для класифікації пацієнтів без необхідності додаткового нормування даних.

Під час розробки та реалізації методу диференційного аналізу даних SaO_2 сформовано набір даних для досліджень сатурації киснем артеріальної крові, при чому виявилось, що загальнодоступна база даних [18; 19] містить певну кількість

даних з артефактами (3 випадки з 25). Набір даних є гетерогенним, що видно вже на коробковій статистичній діаграмі. Здорові дорослі поділяються на три різні підгрупи за величиною варіабельності оксигенації артеріальної крові, і цей розподіл є статистично значущим на довірчому рівні 0,99. Спектр спостережуваних рівнів оксигенації є обмеженим і дискретним, а зміна (перемикання) рівнів може відбуватися лише між найближчими з них. Число перемикань на інтервалі спостереження описується складним (змішаним) розподілом Пуассона. Варіабельність оксигенації артеріальної крові може характеризуватися масштабними інваріантними (незалежними від довжини інтервалу спостереження) інтегральними показниками, зокрема частотами (інтенсивностями, числом) перемикань між рівнями сатурації. Інтервали часу між перемиканнями підкорюються складному статистичному розподілу Ерланга. Вони так само можуть бути масштабно-інваріантними інтегральними показниками варіабельності процесу оксигенації артеріальної крові. Обидва запропоновані нові інтегральні показники варіабельності добре узгоджуються з традиційними (міжквартільними діапазонами).

ВИСНОВКИ

Дисертаційна робота на тему «Методи та засоби цифрової обробки медичних сигналів для підвищення ефективності діагностичних систем» присвячена розробці та вдосконаленню методів аналізу одновимірних фізіологічних сигналів для моніторингу стану здоров'я пацієнтів. У ході дослідження було вирішено поставлені задачі, зокрема, розроблено нові алгоритми цифрової фільтрації та обробки медичних сигналів, покращено методи оцінки варіабельності кисневої сатурації, впроваджено вдосконалені підходи до класифікації фізіологічних параметрів, досліджено фрактальну природу сигналів рівня оксигенації крові, а також оптимізовано застосування рекурентного аналізу для коротких часових рядів.

У роботі здійснено математичне моделювання варіабельності фізіологічних параметрів, розроблено та реалізовано алгоритми аналізу медичних сигналів, проведено порівняльний аналіз існуючих методів, а також підтверджено ефективність запропонованих методик на реальних наборах медичних даних.

За результатами проведеного дослідження отримано наступні наукові результати:

1) вперше запропоновано семантичну модель походження медичних даних (Data provenance), в якій IoT-пристрої, що надсилають інформацію з сенсорів до основної системи обробки, можна розглядати як основне джерело даних для контролю, що дозволяє отримувати верифіковані медичні дані пацієнтів у відповідності до онтологій та правил для їх обробки та візуалізації, використовуючи метод головних компонент, метод наближення до еліпсу, рекурентні графіки та графіки Пуанкаре, та надає можливість спеціалістам приймати клінічні рішення;

2) набули подальшого розвитку метод рекурентного аналізу та обробки коротких серій медичних сигналів, що, на відміну від традиційних методів оцінки варіабельності за середніми значеннями, дозволяє ідентифікувати

перехідні стани фізіологічних процесів із точністю 92 %, зменшити вплив шумів на 30 % та прогнозувати циркадні коливання із похибкою менше ніж 2 %;

3) удосконалено метод оцінки варіабельності SpO_2 , зокрема шляхом застосування міжквартильного діапазону (IQR) як надійного статистичного показника, що демонструє більшу стійкість до викидів у порівнянні зі стандартним відхиленням, допускаючи до 12 % викидів без спотворення статистичних оцінок;

4) вперше запропоновано метод диференціального аналізу сигналів оксигенації, який враховує довжину інтервалів між перемиканнями рівня SaO_2 та їхній статистичний розподіл за законами Ерланга та Пуассона, що, на відміну від традиційних методів аналізу середніх значень, дозволяє збільшити точність оцінки динаміки змін оксигенації на 18–25 % та прогнозувати кількість перемикань рівня сатурації з похибкою не більше ніж 10 %;

5) вперше доведена фрактальна природа сигналів щодо оксигенації артеріальної крові і вказано одновимірний фрактал – часові позиції змін рівня сатурації, – що в подальшому надає можливість зменшення частоти вимірів показників артеріальної крові максимум до $2^8=256$ разів без втрати їх точності.

Також до цих наукових результатів можна виділити наступне:

1) виявлено, що всі записи SpO_2 мають негаусівський розподіл ймовірностей із численними викидами, а їхні графіки Пуанкаре демонструють значну кластеризацію значень;

2) встановлено, що традиційні статистичні методи обробки даних мають значні обмеження при аналізі коротких часових рядів.

3) знайдено статистично значущу негативну кореляцію між коефіцієнтом рекурентності (RR) та варіабельністю SpO_2 , що дозволяє використовувати цей показник як критерій оцінки стабільності фізіологічних сигналів. Проведено оцінку розподілу коефіцієнтів рекурентності та встановлено, що вони мають гаусівський розподіл без викидів, що робить їх придатними для використання в автоматизованих системах аналізу;

4) виділено три основні підгрупи кластеризації пацієнтів за рівнем варіабельності кисневої сатурації:

- 22 % вибірки характеризуються найнижчим рівнем оксигенації та найвищою варіабельністю;
- 56 % вибірки демонструють середню варіабельність;
- 22 % вибірки мають мінімальну варіабельність;

Таким чином, в результаті виконання дисертаційної роботи запропоновані новітні методи та алгоритми створення автоматизованих систем обробки медичних сигналів, які інтегрують рекурентний аналіз, статистичні методи обробки даних та алгоритми машинного навчання. Отримані результати досліджень можуть бути використані у персоналізованих системах медичного моніторингу, кардіологічних аналізаторах та портативних пристроях для оцінки варіабельності фізіологічних параметрів.

Результати дослідження впроваджено у навчальний процес Чорноморського національного університету імені Петра Могили у рамках дисциплін «Цифрова обробка сигналів», «Machine Learning and Pattern Recognition» та «Моделювання систем і процесів» для студентів спеціальності 123 Комп'ютерна інженерія.

Отримані наукові результати сприятимуть подальшому розвитку автоматизованих систем діагностики, покращенню точності моніторингу фізіологічних параметрів та розширенню можливостей персоналізованої медицини. Запропоновані методи можуть бути інтегровані у сучасні медичні пристрої, IoT-системи для моніторингу здоров'я та хмарні сервіси для аналізу медичних даних, в результаті чого можливо очікувати підвищення ефективності діагностичних систем.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Carvalho A., Parra D., Rada G., Perez D., Vasquez J., Vergara C. Neural language models for text classification in evidence-based medicine. *LatinX in AI at Neural Information Processing Systems Conference 2020. Journal of LatinX in AI Research*, 2020. DOI: <https://doi.org/10.52591/lxai202012126>.
2. Biomedical Signal Processing – MATLAB Simulink. 2022. URL: <https://www.mathworks.com/discovery/biomedical-signal-processing.html> (Last accessed: 15.09.2022).
3. Wilson D. D. McGraw-Hill Manual of Laboratory and Diagnostic Tests. 1st ed. New York : McGraw-Hill Medical, 2008.
4. What is Data Science | IBM. 2022. URL: <https://www.ibm.com/cloud/learn/data-science-introduction> (Last accessed: 15.09.2022).
5. Chuiko G. P., Dvornik O. V., Darnapuk Y. S. Shape Evolutions of Poincaré Plots for Electromyograms in Data Acquisition Dynamics. *Proceedings of the 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*, 2018. P. 119–122. DOI: 10.1109/DSMP.2018.8478516.
6. Chuiko G., Dvornik O., Darnapuk Y., Yaremchuk O., Krainyk Y., Puzyrov S. Computer processing of ambulatory blood pressure monitoring as multivariate data. *Proceedings of the 2019 IEEE XVth International Conference on the Perspective Technologies and Methods in MEMS Design (MEMSTECH)*, 2019. P. 23–27. DOI: 10.1109/MEMSTECH.2019.8817375.
7. Chuiko G., Dvornik O., Yaremchuk O., Darnapuk Y. Ambulatory blood pressure monitoring: Modeling and data mining. *CEUR Workshop Proceedings*, 2019.
8. Krainyk Y., Darnapuk Y., Stelmakh S. Dataflow and system organization for image sensor data processing. *Proceedings of the 2019 IEEE 2nd Ukraine Conference on Electrical and Computer Engineering (UKRCON 2019)*, 2019. DOI: 10.1109/UKRCON.2019.8879934.
9. Krainyk Y., Darnapuk Y. Configurable description of FPGA-based control system for sensor processing. *Proceedings of the 2019 11th International Scientific and*

Practical Conference on Electronics and Information Technologies (ELIT 2019), 2019. DOI: 10.1109/ELIT.2019.8892313.

10. Chuiko G., Dvornik O., Darnapuk Y., Krainyk Y. Oxygen saturation variability: Healthy adults. *Proceedings of the 2019 11th International Scientific and Practical Conference on Electronics and Information Technologies (ELIT 2019)*, 2019. DOI: 10.1109/ELIT.2019.8892319.

11. Krainyk Y., Denysov O., Darnapuk Y. Information technology for configuration of system-on-chip in cloud environment. *CEUR Workshop Proceedings*, 2020. P. 342–349.

12. Chuiko G., Darnapuk Y., Dvornik O., Yaremchuk O. Blood oxygenation among healthy adults: Recurrence plots analysis and quantification. *CEUR Workshop Proceedings*, 2020. P. 218–226.

13. Krainyk Y., Darnapuk Y., Simakova I. Software system for physical activity monitoring: smart watch case. *Proceedings of the 5th IEEE International Symposium on Smart and Wireless Systems within the International Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS 2020)*, 2020. DOI: 10.1109/IDAACS-SWS50031.2020.9297093.

14. Chuiko G., Darnapuk Y., Dvornik O., Yaremchuk O., Krainyk Y. Recurrence plots for ambulatory blood pressure monitoring: Means for data mining of circadian rhythms. *Proceedings of the 2020 IEEE 15th International Scientific and Technical Conference on Computer Sciences and Information Technologies (CSIT 2020)*, 2020. DOI: 10.1109/CSIT49958.2020.9321837.

15. Chuiko G., Dvornik O., Darnapuk Y., Krainyk Y. Principal component analysis, quantifying, and filtering of Poincaré plots for time series typical for E-health. 2021. Vol. 932. DOI: 10.1007/978-981-15-9735-0_4.

16. Chuiko G., Krainyk Y., Dvornik O., Darnapuk Y. IoT in provenance management of medical data. 2021. Vol. 941. DOI: 10.1007/978-3-030-64619-6_15.

17. Chuiko G., Dvornik O., Darnapuk Y., Baganov Y. Devising a new filtration method and proof of self-similarity of electromyograms. *Eastern-European Journal of*

Enterprise Technologies. 2021. Vol. 4, No. 9(112). P. 15–22. DOI: 10.15587/1729-4061.2021.239165.

18. Chuiko G., Darnapuk Y., Dvornik O., Krainyk Y. Improved robust handling of electromyograms with mining of new diagnostic signs. *CEUR Workshop Proceeding*. 2021. P. 55–62.

19. Chuiko G., Darnapuk Y., Dvornik O., Krainyk Y., Yaremchuk O., Haab R. A new way of data analysis and rating of the blood oxygen saturation variability. *Proceedings of the 2021 IEEE 12th International Conference on Electronics and Information Technologies (ELIT 2021)*, 2021. P. 51–54. DOI: 10.1109/ELIT53502.2021.9501109.

20. Chuiko G., Darnapuk Y. Development of a method for differential analysis of data on the arterial blood oxygenation in healthy adults. *Eastern-European Journal of Enterprise Technologies*. 2021. Vol. 6, No. 4–114. P. 37–43. DOI: 10.15587/1729-4061.2021.244924.

21. Chuiko G., Darnapuk Y. Fractal nature of arterial blood oxygen saturation data. *Radioelectronic and Computer Systems*. 2022. Vol. 1. P. 206–215. DOI: 10.32620/reks.2022.1.16.

22. Chuiko G., Darnapuk Y., Dvornik O., Krainyk Y. Abdominal electromyograms mining: Breathing patterns of asleep adults. *Radioelectronic and Computer Systems*. 2023. No. 3. P. 60–72. DOI: 10.32620/reks.2023.3.06.

23. Priemer R. Introductory Signal Processing. *WORLD SCIENTIFIC*. 1990. DOI: 10.1142/0864.

24. Fossion R., et al. Homeostasis from a Time-Series Perspective: An Intuitive Interpretation of the Variability of Physiological Variables. In *Quantitative Models for Microscopic to Macroscopic Biological Macromolecules and Tissues*. Cham : Springer International Publishing, 2018. P. 87–109. DOI: 10.1007/978-3-319-73975-5_5.

25. Othmane T. Ambulatory Blood Pressure Monitoring: Practical Insights. 2024. DOI: 10.5281/zenodo.14588936.

26. Tseng Y.-Z. Applications of 24-hour noninvasive ambulatory blood pressure monitoring. *Journal of the Formosan Medical Association*. 2006. Vol. 105, No. 12. P. 955–963. DOI: 10.1016/S0929-6646(09)60279-5.
27. Malha L., White W. B. Ambulatory Blood Pressure Monitoring in Clinical Hypertension Management. *In Hypertension*. Elsevier, 2024. P. 133–143. DOI: 10.1016/B978-0-323-88369-6.00011-6.
28. Oscar 2 Ambulatory Blood Pressure Monitor – SunTech Medical. URL: <https://www.suntechmed.com/bp-products/ambulatory-blood-pressure-monitoring/oscar-2-ambulatory-blood-pressure-monitor> (Last accessed: 16.09.2022).
29. Oo N., Aye M. M., Oo T., Win L. L. Y., Tun H., Pradhan D. Analysis of electromyography (EMG) signal processing with filtering techniques. *Journal of Novel Engineering Science and Technology*. 2024. Vol. 3, No. 02. P. 41–45. DOI: 10.56741/jnest.v3i02.524.
30. Robbins D. An introduction to EMG signal processing using MatLab and Microsoft Excel. P. 95–112. DOI: 10.4018/978-1-4666-6090-8.ch004.
31. Kantz H., Schreiber T. Nonlinear time series analysis. Cambridge University Press, 2003. DOI: 10.1017/CBO9780511755798.
32. Herrmann J., Mori V., Bates J. H. T., Suki B. Modeling lung perfusion abnormalities to explain early COVID-19 hypoxemia. *Nature Communications*. 2020. Vol. 11, No. 1. P. 4883. DOI: 10.1038/s41467-020-18672-6.
33. Kashani K. B. Hypoxia in COVID-19: Sign of severity or cause for poor outcomes. *Mayo Clinic Proceedings*. 2020. Vol. 95, No. 6. P. 1094–1096. DOI: 10.1016/j.mayocp.2020.04.021.
34. Xie J., et al. Association between hypoxemia and mortality in patients with COVID-19. *Mayo Clinic Proceedings*. 2020. Vol. 95, No. 6. P. 1138–1147. DOI: 10.1016/j.mayocp.2020.04.006.
35. Hypoxemia (low blood oxygen) – Mayo Clinic. URL: <https://www.mayoclinic.org/symptoms/hypoxemia/basics/definition/sym-20050930> (Last accessed: 18.09.2022).

36. Leon E., Ilya F., Leonid G. Wearable pulse oximetry device [Патент]. Європейське патентне відомство. No. EP 2750604 B1. URL: <https://lens.org/040-619-919-160-748> (Last accessed: 18.09.2022).
37. Maple Product History – Maplesoft. URL: <https://www.maplesoft.com/products/maple/history/> (Last accessed: 19.09.2022).
38. Bernardin L., et al. Maple Programming Guide. – Maplesoft, a division of Waterloo Maple Inc., 2020.
39. Дворник О., Чуйко Г., Дарнапук Є., Крайник Я. Методи обробки медичних сигналів засобами комп'ютерної математики Maple. Миколаїв : Іліон, 2021.
40. Wolfram Mathematica: Modern Technical Computing. URL: <https://www.wolfram.com/mathematica/> (Last accessed: 19.09.2022).
41. Life sciences – Wolfram Demonstrations Project. URL: <https://demonstrations.wolfram.com/topic.html?topic=life+sciences&start=361> (Last accessed: 19.09.2022).
42. Costa M. D., Redline S., Hughes T. M., Heckbert S. R., Goldberger A. L. Prediction of cognitive decline using heart rate fragmentation analysis: The multi-ethnic study of atherosclerosis. *Frontiers in Aging Neuroscience*. 2021. Vol. 13. DOI: 10.3389/fnagi.2021.708130.
43. Karmakar C. K., Khandoker A. H., Gubbi J., Palaniswami M. Complex correlation measure: A novel descriptor for Poincaré plot. *Biomedical Engineering Online*. 2009. Vol. 8, No. 1. P.17. DOI: 10.1186/1475-925X-8-17.
44. Platiša M. M., Radovanović N. N., Kalauzi A., Pavlović S. Generalized Poincaré plots analysis of cardiac interbeat intervals in heart failure. *Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies*. SCITEPRESS, 2022. P. 251–256. DOI: 10.5220/0010912400003123.
45. Tulppo M. P., Makikallio T. H., Takala T. E., Seppanen T., Huikuri H. V. Quantitative beat-to-beat analysis of heart rate dynamics during exercise. *American Journal of Physiology-Heart and Circulatory Physiology*. 1996. Vol. 271, No. 1. P. H244–H252. DOI: 10.1152/ajpheart.1996.271.1.H244.

46. Piskorski J., Guzik P. Filtering Poincaré plots. *Computational Methods in Science and Technology*. 2005. Vol. 11, No. 1. P. 39–48. DOI: 10.12921/cmst.2005.11.01.39-48.
47. Messaoudi N., Belkacem S., Bekka R. E. Simulated surface electromyographic (SEMG) signal generation and detection model. *The Scientific Bulletin of Electrical Engineering Faculty*. 2023. Vol. 23, No. 2. P. 82–92. DOI: 10.2478/sbeef-2023-0024.
48. Liu Y., Mi B., Pei L., Dang S., Yan H., Li C. Impact of intensive hypertension criteria on multimorbidity prevalence and patterns in a multi-ethnic Chinese population. *Frontiers in Public Health*. 2024. Vol. 12. DOI: 10.3389/fpubh.2024.1443104.
49. Ciobanu D., Bala C., Filip M., Pătruț C.-I., Hămbășan I., Roman G. Prevalence of white-coat and masked hypertension in patients with type 2 diabetes. *Romanian Journal of Diabetes Nutrition & Metabolic Diseases*. 2022. Vol. 29. P. 273–279. DOI: 10.46389/rjd-2022-1119.
50. Niknafs P., Norouzi E., Bijari B., Baneshi M. Can we replace arterial blood gas analysis by pulse oximetry in neonates with respiratory distress syndrome, Who are treated according to INSURE protocol? *Iranian Journal of Medical Sciences*. 2015. Vol. 40. P. 264–267.
51. Jubran A. Pulse oximetry. *Critical Care*, 2015. Vol. 19, No. 1. P. 272. DOI: 10.1186/s13054-015-0984-8.
52. Yoshida M., Onda K., Wada Y., Kuwahara M. Influence of Sickness Condition on Diurnal Rhythms of Heart Rate and Heart Rate Variability in Cows. *Journal of Veterinary Medical Science*, 2015. – Vol. 77, No. 3. – P. 375–379. DOI: 10.1292/jvms.14-0402.
53. Bhogal A. S., Mani A. R. Pattern analysis of oxygen saturation variability in healthy individuals: Entropy of pulse oximetry signals carries information about mean oxygen saturation. *Frontiers in Physiology*. 2017. Vol. 8. DOI: 10.3389/fphys.2017.00555.

54. Braun W. J., Murdoch D. J. A first course in statistical programming with R. Cambridge University Press, 2021. DOI: 10.1017/9781108993456.
55. Chuiko G., Darnapuk Y., Dvornik O., Krainyk Y., Yaremchuk O., Diadenko A. ‘Devil’s Stairs’, Poisson’s statistics, and patient sorting via variabilities for oxygenation: All from arterial blood gas data. *bioRxiv*, 2021. P. 2021.08.10.455835. DOI: 10.1101/2021.08.10.455835.
56. Berners-Lee T., Hendler J., Lassila O. The semantic web: A new form of web content that is meaningful to computers will unleash a revolution of new possibilities. *Scientific American*. Oct. 2001.
57. Information resources management association (Ed.). Research anthology on supporting healthy aging in a digital society. IGI Global, 2022. DOI: 10.4018/978-1-6684-5295-0.
58. Marques G., Bhoi A. K., de Albuquerque V. H. C., K.S. H. (Eds.). IoT in healthcare and ambient assisted living. Vol. 933. Singapore : Springer Singapore, 2021. DOI: 10.1007/978-981-15-9897-5.
59. Zyrianoff I., Heideker A., Sciullo L., Kamienski C., Di Felice M. Interoperability in open IoT platforms: WoT-FIWARE comparison and integration. *Proceedings of the 2021 IEEE International Conference on Smart Computing (SMARTCOMP)*, 2021. P. 169–174. DOI: 10.1109/SMARTCOMP52413.2021.00043.
60. Shih L.-C., et al. Prediction of white-coat hypertension and white-coat uncontrolled hypertension using machine learning algorithm. *European Heart Journal - Digital Health*. 2022. Vol. 3, No. 4. P. 559–569. DOI: 10.1093/ehjdh/ztac066.
61. Pandey R., Paprzycki M., Srivastava N., Bhalla S., Wasielewska-Michniewska K. (Eds.). Semantic IoT: Theory and applications. Vol. 941. Cham : Springer International Publishing, 2021. DOI: 10.1007/978-3-030-64619-6.
62. Guleria P., Sood M. Semantic IoT interoperability and data analytics using machine learning in healthcare sector. 2021. P. 245–262. DOI: 10.1007/978-3-030-64619-6_11.
63. Afzaal R., Hussain S. An IoT-assisted intelligent monitoring system. *Journal of Information Analysis*. 2024. Vol. 2. P. 8. DOI: 10.53964/jia.2024008.

64. Seydoux N., Drira K., Hernandez N., Monteil T. Capturing the contributions of the semantic web to the IoT: A unifying vision. 2017.
65. Mishra S., Jain S. Ontologies as a semantic model in IoT. *International Journal of Computers and Applications*. 2020. Vol. 42, No. 3. P. 233–243. DOI: 10.1080/1206212X.2018.1504461.
66. Murdock P., et al. Semantic interoperability for the Web of Things. 2016. DOI: 10.13140/RG.2.2.25758.13122.
67. Jabbar S., Ullah F., Khalid S., Khan M., Han K. Semantic interoperability in heterogeneous IoT infrastructure for healthcare. *Wireless Communications and Mobile Computing*. 2017. Vol. 2017. P. 1–10. DOI: 10.1155/2017/9731806.
68. Mazayev A., Martins J. A., Correia N. Interoperability in IoT through the semantic profiling of objects. *IEEE Access*. 2018. Vol. 6. P. 19379–19385. DOI: 10.1109/ACCESS.2017.2763425.
69. Bajaj G., Agarwal R., Singh P., Georgantas N., Issarny V. 4W1H in IoT Semantics. *IEEE Access*, 2018. Vol. 6. P. 65488–65506. DOI: 10.1109/ACCESS.2018.2878100.
70. Jacoby M., AntoniĆ A., Kreiner K., Łapacz R., Pielorz J. Semantic interoperability as key to IoT platform federation. 2017. P. 3–19. DOI: 10.1007/978-3-319-56877-5_1.
71. Krainyk Y., Davydenko Y., Tomas V. Configurable control node for wireless sensor network. *Proceedings of the 2019 IEEE 3rd International Conference on Advanced Information and Communications Technologies (AICT)*, 2019. P. 258–262. DOI: 10.1109/AIACT.2019.8847732.
72. Krainyk Y., Razzhyvin A., Bondarenko O., Simakova I. Internet-of-Things device set configuration for connection to wireless local area network. *CEUR Workshop Proceedings*. 2019. Vol. 2353. P. 885–896.
73. Hartig O. Provenance information in the Web of Data. 2009.
74. Alkhalil A., Ramadan R. A. IoT data provenance implementation challenges. *Procedia Computer Science*. 2017. Vol. 109. P. 1134–1139. DOI: 10.1016/j.procs.2017.05.436.

75. Olufowobi H., Engel R., Baracaldo N., Bathen L. A. D., Tata S., Ludwig H. Data provenance model for Internet of Things (IoT) Systems. 2017. P. 85–91. DOI: 10.1007/978-3-319-68136-8_8.
76. Sahoo S. S., Valdez J., Rueschman M. Scientific reproducibility in biomedical research: Provenance metadata ontology for semantic annotation of study description. *AMIA Annual Symposium Proceedings*. 2016. P. 1070–1079.
77. Health level seven international – Homepage | HL7 international. URL: <https://www.hl7.org/> (Last accessed: 09.01.2022).
78. HL7 Standards Product Brief – HL7 Version 2 Product Suite | HL7 International. URL: https://www.hl7.org/implement/standards/product_brief.cfm?product_id=185 (Last accessed: 09.12.2022).
79. HL7 Standards – Section 1e: Version 3 (V3) | HL7 International. URL: https://www.hl7.org/implement/standards/product_section.cfm?section=14 (Last accessed: 09.12.2022).
80. HL7 Standards Product Brief – FHIR® (HL7 Fast Healthcare Interoperability Resources) | HL7 International URL: https://www.hl7.org/implement/standards/product_brief.cfm?product_id=491 (Last accessed: 09.12.2022).
81. Arendt J. Melatonin: Countering chaotic time cues. *Frontiers in Endocrinology (Lausanne)*. 2019. Vol. 10. DOI: 10.3389/fendo.2019.00391.
82. Almeida E. A. de, et al. Measurement of melatonin in body fluids: Standards, protocols and procedures. *Child's Nervous System*. 2011. Vol. 27, No. 6. P. 879–891. DOI: 10.1007/s00381-010-1278-8.
83. Kunz D., Mahlberg R., Müller C., Tilmann A., Bes F. Melatonin in patients with reduced REM sleep duration: Two randomized controlled trials. *Journal of Clinical Endocrinology & Metabolism*. 2004. Vol. 89, No. 1. P. 128–134. DOI: 10.1210/jc.2002-021057.
84. Melatonin-Sulfate (EIA-1432), Report of DRG International Inc., USA. ver. 8.1. Publ. 12 Sept., 2011. URL: <http://weldonbiotech.com/wp-content/uploads/2018/05/eia-1432.pdf> (Last accessed: 29.10.2022).

85. Melatonin Sulfate ELISA (RE54031), Report of IBL International GMBH. Publ. June 19, 2017. URL: <https://novamedline.com/files/523001e4-ac11-409b-9cf2-017adfb65556.pdf> (Last accessed: 29.10.2022).
86. Melatonin Sulfate ELISA (40-371-25006), Report of GenWay Biotech, Inc. Publ. May 18, 2017. URL: <https://www.genwaybio.com/media/custom/upload/File-1313509984.pdf> (Last accessed: 29.10.2022).
87. 6-Sulfatoxymelatonin ELISA (79-STMHU-E01), Report of ALPCO. Publ. Dec. 07, 2016. URL: <https://pdf.medicalexpo.com/pdf/alpco/6-sulfatoxymelatonin-elisa/69512-187076.html> (Last accessed: 29.10.2022).
88. 6-Sulfatoxymelatonin ELISA (EK-M6S), Report of Bühlmann Laboratories AG. Publ. Jan. 18, 2016. URL: https://buhlmannlabs.com/wp-content/uploads/BUHLMANN-6-Sulfatoxymelatonin-ELISA_EK-M6S_160118_RUO-1.pdf (Last accessed: 29.10.2022).
89. Chuiko G. P., Dvornik O. V., Shyian I. A. How reliable are calibrators for urinary melatonin sulfate? *Medical Informatics and Engineering*. 2016. No. 3. DOI: 10.11603/mie.1996-1960.2016.3.6759.
90. Gadagkar S. R., Call G. B. Computational tools for fitting the hill equation to dose–response curves. *Journal of Pharmacological and Toxicological Methods*. 2015. Vol. 71. P. 68–76. DOI: 10.1016/j.vascn.2014.08.006.
91. Bühlmann Laboratories AG. Direct Saliva Melatonin RIA (RK-DSM IFU). Publ. Jan. 20, 2016. URL: https://buhlmannlabs.com/wp-content/uploads/BUHLMANN-Direct-Saliva-Melatonin-RIA_RK-DSM.pdf (Last accessed: 03.03.2025).
92. Bühlmann Laboratories AG. Melatonin RIA (RK-MEL). Publ. Nov. 20, 2012. URL: https://www.sceti.co.jp/images/psearch/pdf/BUL_RK-MEL2_p.pdf (Last accessed: 03.03.2025).
93. Bühlmann Laboratories AG. Direct saliva melatonin ELISA (EK-DSM). Publ. Jan. 14, 2019. URL: https://www.buhlmannlabs.ch/wp-content/uploads/2015/01/EK-DSM_IFU-CE_VA1-2019-01-14.pdf (Last accessed: 03.03.2025).

94. Maple Application Center. Calibrating response curves for the concentration of melatonin sulfate in human urine. URL: <https://www.maplesoft.com/Applications/Detail.aspx?id=154007> (Last accessed: 03.03.2025).

95. Ray P. P. A Survey on Internet of Things architectures. *Journal of King Saud University – Computer and Information Sciences*. 2018. Vol. 30, No. 3. P. 291–319. DOI: 10.1016/j.jksuci.2016.10.003.

96. Boyes H., Hallaq B., Cunningham J., Watson T. The Industrial Internet of Things (IIoT): An analysis framework. *Computers in Industry*. 2018. Vol. 101. P. 1–12. DOI: 10.1016/j.compind.2018.04.015.

97. de Souza R. H., Savazzi S., Becker L. B. Network design and planning of wireless embedded systems for industrial automation. *Design Automation for Embedded Systems*. 2015. Vol. 19, No. 4. P. 367–388. DOI: 10.1007/s10617-015-9162-0.

98. Quan W., Pimentel A. D. A hierarchical run-time adaptive resource allocation framework for large-scale MPSoC systems. *Design Automation for Embedded Systems*. 2016. Vol. 20, No. 4. P. 311–339. DOI: 10.1007/s10617-016-9179-z.

99. Cannella E., Stefanov T. P. Energy efficient semi-partitioned scheduling for embedded multiprocessor streaming systems. *Design Automation for Embedded Systems*. 2016. Vol. 20, No. 3. P. 239–266. DOI: 10.1007/s10617-016-9176-2.

100. Fresse V., Combes C., Payet M., Rousseau F. Methodological framework for NoC resources dimensioning on FPGAs. *Procedia Computer Science*. 2015. Vol. 56. P. 391–396. DOI: 10.1016/j.procs.2015.07.225.

101. Lecler J.-J., Baillieu G. Application driven network-on-chip architecture exploration & refinement for a complex SoC. *Design Automation for Embedded Systems*. 2011. Vol. 15, No. 2. P. 133–158. DOI: 10.1007/s10617-011-9075-5.

102. Palagin A., Opanasenko V. The implementation of extended arithmetics on FPGA-based structures. *Proceedings of the 2017 9th IEEE International Conference*

on *Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, 2017. P. 1014–1019. DOI: 10.1109/IDAACS.2017.8095239.

103. Wang P., Zhang Y., Yang J. Research and design of AES security processor model based on FPGA. *Procedia Computer Science*. 2018. Vol. 131. P. 249–254. DOI: 10.1016/j.procs.2018.04.210.

104. Musiyenko M., Krainyk Y., Denysov O. Reconfigurable decoder for irregular random low density parity check matrix based on FPGA. *Proceedings of the 2015 IEEE 35th International Conference on Electronics and Nanotechnology (ELNANO)*, 2015. P. 498–503. DOI: 10.1109/ELNANO.2015.7146937.

105. Nikolic J., et al. A synchronized visual-inertial sensor system with FPGA pre-processing for accurate Real-time SLAM. *Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA)*, 2014. P. 431–437. DOI: 10.1109/ICRA.2014.6906892.

106. da Silva B., Segers L., Braeken A., Steenhaut K., Touhafi A. A low-power FPGA-based architecture for microphone arrays in wireless sensor networks. 2018. P. 281–293. DOI: 10.1007/978-3-319-78890-6_23.

107. García G., Jara C., Pomares J., Alabdo A., Poggi L., Torres F. A survey on FPGA-based sensor systems: Towards intelligent and reconfigurable low-power sensors for computer vision, control and signal processing. *Sensors*. 2014. Vol. 14, No. 4. P. 6247–6278. DOI: 10.3390/s140406247.

108. Poincaré recurrence theorem. URL: <http://www.math.umd.edu/~lvrnr/History/Recurrence.html> (Last accessed: 11.12.2022).

109. Nadkarni M. G. The Poincaré recurrence lemma. 2013. P. 1–12. DOI: 10.1007/978-93-86279-53-8_1.

110. Carathéodory C. Ueber den wiederkehrrsatz von Poincaré. *Sitzungsberichte der Preussischen Akademie der Wissenschaften*, 1919. URL: <https://books.google.com.ua/books?id=GjjgSAAACAAJ>.

111. Marwan N. A historical review of recurrence plots. *European Physical Journal Special Topics*. 2008. Vol. 164, No. 1. P. 3–12. DOI: 10.1140/epjst/e2008-00829-1.

112. Yang A. C. C. Poincaré plots: A mini-review. 2006. URL: <https://archive.physionet.org/events/hrv-2006/yang.pdf> (Last accessed: 11.12.2022).

113. Poincaré graph. Complete ECG record in one sight. 2014. URL: https://files.btlnet.com/product-document/9792e3d5-3dbf-45d8-9e84-5c964a6a8602/BTL-Cardiopoint_WP_Poincare-graph_EN400_9792e3d5-3dbf-45d8-9e84-5c964a6a8602_original.pdf (Last accessed: 11.12.2021).

114. A brief history of the scatter plot – data visualization’s greatest invention. URL: <https://qz.com/1235712/the-origins-of-the-scatter-plot-data-visualizations-greatest-invention> (Last accessed: 11.12.2022).

115. Friendly M., Denis D. The early origins and development of the scatterplot. *Journal of the History of the Behavioral Sciences*. 2005. Vol. 41, No. 2. P. 103–130. DOI: 10.1002/jhbs.20078.

116. Henriques T. S., Mariani S., Burykin A., Rodrigues F., Silva T. F., Goldberger A. L. Multiscale Poincaré plots for visualizing the structure of heartbeat time series. *BMC Medical Informatics and Decision Making*. 2015. Vol. 16, No. 1. P. 17. DOI: 10.1186/s12911-016-0252-0.

117. Mäkikallio T. Analysis of heart rate dynamics by methods derived from nonlinear mathematics: Clinical applicability and prognostic significance. Oulu : Oulu University Library, 1998.

118. Marciano F., Migaux M. L., Acanfora D., Furgi G., Rengo F. Quantification of Poincaré maps for the evaluation of heart rate variability. *Computers in Cardiology*, 1994. P. 577–580. DOI: 10.1109/CIC.1994.470126.

119. Fishman M., et al. A Method for analyzing temporal patterns of variability of a time series from Poincaré plots. *Journal of Applied Physiology*. 2012. Vol. 113, No. 2. P. 297–306. DOI: 10.1152/jappphysiol.01377.2010.

120. Gorban A. N., Zinovyev A. Y. Principal graphs and manifolds. *Handbook of Research on Machine Learning Applications and Trends*. IGI Global, 2010. P. 28–59. DOI: 10.4018/978-1-60566-766-9.ch002.
121. Rutkove S. Examples of Electromyograms. 2010. DOI: 10.13026/C24S3D. URL: <https://physionet.org/content/emgdb/>.
122. Brennan M., Palaniswami M., Kamen P. New insights into the relationship between Poincaré plot geometry and linear measures of heart rate variability. *Proceedings of the 23rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2001. P. 526–529. DOI: 10.1109/IEMBS.2001.1018984.
123. Golyandina N., Nekrutkin V., Zhigljavsky A. Analysis of time series structure: SSA and related techniques. 2001. DOI: 10.1201/9781420035841.
124. Korobeynikov A. Computation- and space-efficient implementation of SSA. *Statistics and Its Interface*. 2010. Vol. 3, No. 3. P. 357–368. DOI: 10.4310/SII.2010.v3.n3.a9.
125. Brennan M., Palaniswami M., Kamen P. Do existing measures of Poincaré plot geometry reflect nonlinear features of heart rate variability? *IEEE Transactions on Biomedical Engineering*. 2001. Vol. 48, No. 11. P. 1342–1347. DOI: 10.1109/10.959330.
126. Smith L. I. A Tutorial on principal components analysis. 2002. URL: http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_components.pdf (Last accessed: 14.12.2022).
127. Layton W. J., Sussman M. M. Numerical linear algebra. 2014. URL: https://people.sc.fsu.edu/~jburkardt/classes/nla_2015/numerical_linear_algebra.pdf (Last accessed: 14.12.2022).
128. Hansen P. C., Jensen S. H. FIR Filter representations of reduced-rank noise reduction. *IEEE Transactions on Signal Processing*. 1998. Vol. 46, No. 6. P. 1737–1741. DOI: 10.1109/78.678511.

129. Figueiredo N., Georgieva P., Lang E. W., Santos I. M., Teixeira A. R., Tomé A. M. SSA of biomedical signals: A linear invariant systems approach. *Statistics and Its Interface*. 2010. Vol. 3, No. 3. P. 345–355. DOI: 10.4310/SII.2010.v3.n3.a8.
130. Harris T. J., Yuan H. Filtering and frequency interpretations of singular spectrum analysis. *Physica D: Nonlinear Phenomena*. 2010. Vol. 239, No. 20–22. P. 1958–1967. DOI: 10.1016/j.physd.2010.07.005.
131. Patel K., Rora K. K., Singh K., Verma S. Lazy wavelet transform based steganography in video. *Proceedings of the IEEE 2013 International Conference on Communication Systems and Network Technologies (CSNT)*, 2013. P. 497–500. DOI: 10.1109/CSNT.2013.109.
132. Haar A. Zur theorie der orthogonalen funktionensysteme. *Mathematische Annalen*. 1910. Vol. 69, No. 3. P. 331–371. DOI: 10.1007/BF01456326.
133. Dastourian B., Dastourian E., Dastourian S., Mahnaie O. Discrete wavelet transforms of Haar's wavelet. *International Journal of Scientific & Technology Research*. 2014. Vol. 3, No. 9. P. 247–251.
134. Huikuri H. V., Mäkilallio T. H., Peng C.-K., Goldberger A. L., Hintze U., Møller M. Fractal correlation properties of R-R interval dynamics and mortality in patients with depressed left ventricular function after an acute myocardial infarction. *Circulation*. 2000. Vol. 101, No. 1. P. 47–53. DOI: 10.1161/01.CIR.101.1.47.
135. Voss A., Schulz S., Schroeder R., Baumert M., Caminal P. Methods derived from nonlinear dynamics for analysing heart rate variability. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*. 2009. Vol. 367, No. 1887. P. 277–296. DOI: 10.1098/rsta.2008.0232.
136. Gomes R. L., Vanderlei L. C. M., Garner D. M., Vanderlei F. M., Valenti V. E. Higuchi Fractal analysis of heart rate variability is sensitive during recovery from exercise in physically active men. *Medical Express*. 2017. Vol. 4, No. 2. DOI: 10.5935/MedicalExpress.2017.02.03.
137. Antonio A. M., et al. Fractal dynamics of heart rate variability: A study in healthy subjects. *Journal of Cardiovascular Diseases*. 2014. Vol. 2, No. 3. P. 137–142.

URL: <https://repositorio.unesp.br/bitstream/handle/11449/115488/ISSN23304596-2014-02-02-137-142.pdf> (Last accessed: 14.12.2022).

138. Carvalho T. D., et al. Fractal correlation property of heart rate variability in chronic obstructive pulmonary disease. *International Journal of Chronic Obstructive Pulmonary Disease*. 2011. P. 23. DOI: 10.2147/COPD.S15099.

139. Bourke P. Box counting fractal dimension of volumetric data. 2014. URL: <http://paulbourke.net/fractals/cubecount/> (Last accessed: 14.12.2022).

140. Mandelbrot B. How long is the coast of Britain? Statistical self-similarity and fractional dimension. *Science (1979)*. 1967. Vol. 156, No. 3775. P. 636–638. DOI: 10.1126/science.156.3775.636.

141. Korteweg D. J., de Vries G. XLI. On the change of form of long waves advancing in a rectangular canal, and on a new type of long stationary waves. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*. 1895. Vol. 39, No. 240. P. 422–443. DOI: 10.1080/14786449508620739.

142. Brauer K. The Korteweg-de Vries equation: History, exact solutions, and graphical representation. Publ. Dec. 2000. URL: <http://people.seas.harvard.edu/~jones/solitons/pdf/025.pdf> (Last accessed: 15.12.2020).

143. Miura R. M. Korteweg-de Vries equation and generalizations. I. A remarkable explicit nonlinear transformation. *Journal of Mathematical Physics*. 1968. Vol. 9, No. 8. P. 1202–1204. DOI: 10.1063/1.1664700.

144. Volobuev A. N. Fluid flow in tubes with elastic walls. *Physics-Uspekhi*. 1995. Vol. 38, No. 2. P. 169–178. DOI: 10.1070/PU1995v038n02ABEH000069.

145. Chuiko G. P., Dvornik O. V., Shyian S. I. Validity of Korteweg-de-Vries equation for arterial pulse waves. *Electronic Journal of Theoretical Physics*. 2016. Vol. 13, No. 36. P. 99–106.

146. Chuiko G. P., Dvornik O. V., Shyian S. I., Baganov Ye. A. A new age-related model for blood stroke volume. *Computer Biology and Medicine*. 2016. Vol. 79. P. 144–148. DOI: 10.1016/j.combiomed.2016.10.013.

147. Demiray H. On some nonlinear waves in fluid-filled viscoelastic tubes: weakly dispersive case. *Communications in Nonlinear Science and Numerical Simulation*. 2005. Vol. 10, No. 4. P. 425–440. DOI: 10.1016/j.cnsns.2003.08.005.

148. Demiray H. Solitary waves in fluid-filled elastic tubes: weakly dispersive case. *International Journal of Engineering Science*. 2001. Vol. 39, No. 4. P. 439–451. DOI: 10.1016/S0020-7225(00)00048-3.

149. Grimshaw R. H. J. (Ed.). Solitary waves in fluids. Southampton, Boston : WIT Press, 2007.

150. Kudryashov N. A., Chernyavskii I. L. Nonlinear waves in fluid flow through a viscoelastic tube. *Fluid Dynamics*. 2006. Vol. 41, No. 1. P. 49–62. DOI: 10.1007/s10697-006-0021-3.

151. What is Maple: Product features – Math & Engineering Software – Maplesoft. URL: <https://www.maplesoft.com/products/Maple/features/> (Last accessed: 15.12.2022).

152. Johnson M. A. Nonlinear stability of periodic traveling wave solutions of the generalized Korteweg–de Vries equation. *SIAM Journal on Mathematical Analysis*. 2009. Vol. 41, No. 5. P. 1921–1947. DOI: 10.1137/090752249.

153. Deconinck B., Nivala M. The Stability Analysis of the Periodic Traveling Wave Solutions of the mKdV Equation. *Studies in Applied Mathematics*. 2011. Vol. 126, No. 1. P. 17–48. DOI: 10.1111/j.1467-9590.2010.00496.x.

154. Chen J., Pelinovsky D. E. Rogue periodic waves of the modified KdV equation. *Nonlinearity*. 2018. Vol. 31, No. 5. P. 1955–1980. DOI: 10.1088/1361-6544/aaa2da.

155. Byrd P. F., Friedman M. D. Handbook of elliptic integrals for engineers and scientists. Berlin, Heidelberg : Springer Berlin Heidelberg, 1971. DOI: 10.1007/978-3-642-65138-0.

156. Marwan N., Webber C. L. Mathematical and computational foundations of recurrence quantifications. *Understanding Complex Systems*. 2015. P. 3–43. DOI: 10.1007/978-3-319-07155-8_1.

157. Ambulatory Blood Pressure Report, 29 June, 2005, QRS ®., ML 402. 2005.
URL: <http://qrsdiagnostic.com/sites/default/files/Blood%20Pressure/ML402%20ABPM%20Sample%20Report.pdf> (Last accessed: 14.12.2020).
158. Schinkel S., Marwan N., Kurths J. Order patterns recurrence plots in the analysis of ERP data. *Cognitive Neurodynamics*. 2007. Vol. 1, No. 4. P. 317–325. DOI: 10.1007/s11571-007-9023-z.
159. Albisser A. M., Alejandro R., Meneghini L. F., Ricordi C. How Good Is Your Glucose Control? *Diabetes Technology & Therapeutics*. 2005. Vol. 7, No. 6. P. 863–875. DOI: 10.1089/dia.2005.7.863.
160. Chuiko G. P., Dvornik O. V., Darnapuk Y. S. Combined processing of blood glucose self-monitoring. *Medical Informatics and Engineering*. 2019. No. 3. P. 59–68. DOI: 10.11603/mie.1996-1960.2019.3.10433.
161. oldberger A. L., et al. PhysioBank, PhysioToolkit, and PhysioNet. *Circulation*. 2000. Vol. 101, No. 23. DOI: 10.1161/01.CIR.101.23.e215.
162. Messaoudi N., Bekka R. E., Ravier P., Harba R. Assessment of the non-Gaussianity and non-linearity levels of simulated sEMG signals on stationary segments. *Journal of Electromyography and Kinesiology*. 2017. Vol. 32. P. 70–82. DOI: 10.1016/j.jelekin.2016.12.006.
163. Field A. Discovering statistics using IBM SPSS statistics. 4th ed. Sage Publications Ltd., 2013.
164. Rose W. KAAP686 Mathematics and signal processing for biomechanics electromyogram analysis. 2011. URL: <http://www1.udel.edu/biology/rosewc/kaap686/notes/EMG%20analysis.pdf>.
165. Melnyk A., Morozov Y., Havano B., Hupalo P. Protection of biometric data transmission and storage in the human state remote monitoring tools. *2021 11th IEEE International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, Sept. 2021. P. 301–306. DOI: 10.1109/IDAACS53288.2021.9661047.
166. Melnyk A., Morozov Y., Hupalo P., Havano B. HealthyLungs: Mobile applications for round-the-clock remote monitoring of lung function in patients with

COVID-19. *2021 11th International Conference on Advanced Computer Information Technologies (ACIT)*, Sept. 2021. P. 554–559. DOI: 10.1109/ACIT52158.2021.9548596.

167. Melnyk A., Melnyk V. Remote synthesis of computer devices for FPGA-based IoT nodes. *2020 IEEE 10th International Conference on Advanced Computer Information Technologies (ACIT)*, Sept. 2020. P. 254–259. DOI: 10.1109/ACIT49673.2020.9208882.

168. Ghassemi M., et al. You Snooze, You Win: The PhysioNet/Computing in Cardiology Challenge. Dec. 2018. DOI: 10.22489/CinC.2018.049.

169. Frost J. Introduction to the Statistics: *An Intuitive*. 2019. URL: <https://statisticsbyjim.com/basics/variability-range-interquartile-variance-standard-deviation/> (Last accessed: 17.05.2020).

170. Scott D. W. Averaged shifted histogram. *WIREs Computational Statistics*. 2010. Vol. 2, No. 2. P. 160–164. – DOI: 10.1002/wics.54.

171. Węglarczyk S. Kernel density estimation and its application. *ITM Web of Conferences*. 2018. Vol. 23. P. 00037. DOI: 10.1051/itmconf/20182300037.

172. Tetrax: Fourier transformation of postural sway, Sunlight. URL: <http://postureetmeasure.free.fr/docs/TetraxFourierTransformation.pdf> (Last accessed: 11.08.2024).

173. Banerjee S., Easwaramoorthy D., Gowrisankar A. Fractal functions, dimensions and signal analysis. Cham : Springer International Publishing, 2021. DOI: 10.1007/978-3-030-62672-3.

174. Amoian B., Rabiée M., Aghvami M., Milani S. Evaluation of hemodynamic and SpO2 variability during different stages of periodontal surgery. *Journal of Indian Society of Periodontology*. 2013. DOI: 10.4103/0972-124X.119274.

175. Majumdar S. R., Eurich D. T., Gamble J. M., Senthilselvan A., Marrie T. J. Oxygen saturations less than 92% are associated with major adverse events in outpatients with pneumonia: A population-based cohort study. *Clinical Infectious Diseases*. 2011. Vol. 52, No. 3. P. 325–331. DOI: 10.1093/cid/ciq076.

176. Jubran A. Pulse oximetry. *Applied Physiology in Intensive Care Medicine 1: Physiological Notes – Technical Notes – Seminal Studies in Intensive Care*. 2012. DOI: 10.1007/978-3-642-28270-6_12.

177. Eckmann J. P., Oliffson Kamphorst O., Ruelle D. Recurrence plots of dynamical systems. *EPL (Europhysics Letters)*. 1987. DOI: 10.1209/0295-5075/4/9/004.

178. Marwan N., Romano M. C., Thiel M., Kurths J. Recurrence plots for the analysis of complex systems. *Physics Reports*. 2007. DOI: 10.1016/j.physrep.2006.11.001.

ДОДАТОК А

Акти впровадження та листи підтримки

ЗАТВЕРДЖУЮ

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

Роман ДІНЖОС

«25» чрудня 2024 р.

АКТ**впровадження результатів дисертаційної роботи**

ДАРНАПУКА Євгена Сергійовича на тему «Методи та засоби обробки медичних сигналів для підвищення ефективності діагностичних систем»
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 «Методи та засоби обробки медичних сигналів для підвищення ефективності діагностичних систем» виконується у Чорноморському національному університеті імені Петра Могили.

При виконанні першого етапу НДР були використані нижче зазначені результати досліджень здобувача ДАРНАПУКА Євгена Сергійовича:

- розроблені нові алгоритми кластеризації пацієнтів за рівнем варіабельності кисневої сатурації, що дозволяє покращити якість моніторингу медичних сигналів та ефективність діагностичних систем;
- оптимізовано процедуру рекурентного аналізу коротких часових рядів, що дозволяє ефективно оцінювати варіабельність фізіологічних сигналів у реальному часі;
- покращено методи аналізу графіків Пуанкаре у поєднанні з методами статистики, що дозволяє більш точно оцінювати перехідні процеси у фізіологічних системах.

PhD-студент кафедри комп'ютерної інженерії ЧНУ імені Петра Могили Євген ДАРНАПУК приймав участь у НДР як виконавець.

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

Геннадій ЧУЙКО

«25» чрудня 2024 р.

ЗАТВЕРДЖУЮ

Перший проректор Чорноморського
національного університету ім. Петра
Могили, професор



Юрій КОТЛЯР

« 19 » вересня 2024 р.

**АКТ****впровадження результатів дисертаційної роботи**

Дарнапука Є. С. на тему: «Методи та засоби обробки медичних сигналів для підвищення ефективності діагностичних систем» в навчальний процес Чорноморського національного університету ім. Петра Могили за першим (бакалаврським) рівнем вищої освіти на кафедрі комп'ютерної інженерії

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

У процес викладання дисциплін «Цифрова обробка сигналів» (українською мовою) та «Machine Learning and Pattern Recognition» (англійською мовою), Робочі програми до яких розроблені старшим викладачем Дарнапуком Є. С., введені такі теми, що містять матеріал роботи здобувача:

- «Використання технологій Internet-of-Things в управлінні походженням даних»;
- «Методи та засоби обробки та аналізу кноїдних та усамітнених пульсових хвиль (солітонів) в крові»;
- «Метод графіків Пуанкаре для біомедичних сигналів: кількісний опис та обробка».

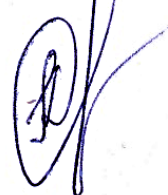
Лекційні матеріали та методичні матеріали для виконання практичних робіт наведені в модульному об'єктно-орієнтованому динамічному навчальному середовищі Moodle3 ЧНУ ім. Петра Могили та можуть бути використані у подальшому освітньому процесі як за дистанційною, так і очною або заочною формою навчання.

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



Ірина ЖУРАВСЬКА

Гарант ОПІ 123 Комп'ютерна інженерія
за першим (бакалаврським) рівнем ВО,
доцент кафедри КІ, канд. техн. наук, доцент



Ярослав КРАЙНИК

ЗАТВЕРДЖУЮ

Перший проректор Чорноморського
національного університету ім. Петра
Могили, професор

« 17 » жовтня 2024 р.



АКТ

впровадження результатів дисертаційної роботи

Дарнапука Є. С. на тему: «Методи та засоби обробки медичних сигналів для підвищення ефективності діагностичних систем» в навчальний процес Чорноморського національного університету ім. Петра Могили за другим (магістерським) рівнем вищої освіти на кафедрі комп'ютерної інженерії

Основні наукові та практичні результати дисертаційної роботи Дарнапука Є. С. застосовуються у навчальному процесі на кафедрі комп'ютерної інженерії ЧНУ ім. Петра Могили в курсі лекційних та практичних занять при викладанні дисципліни «Моделювання систем і процесів», студентам спеціальності 123 Комп'ютерна інженерія за другим (магістерським) рівнем вищої освіти.

У процес викладання дисципліни «Моделювання систем і процесів», Робоча програма та Навчальний посібник до якої розроблено старшим викладачем Дарнапуком Є. С. у співавторстві з науковим керівником дисертації професором Чуйком Г. П., введені такі теми, які містять матеріали роботи здобувача:

- «Прогнозна аналітика у моделях машинного навчання»;
- «Система комп'ютерної математики Maple для моделювання хвильових нелінійних процесів»;
- «Безвимірні способи представлення математичних моделей технічних та біосистем».

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 за дистанційною, так і очною або заочною формою навчання.

Завідувачка кафедри комп'ютерної інженерії,
Гарант ОПП 123 Комп'ютерна інженерія
за другим (магістерським) рівнем ВО,
д-р техн. наук, професор

Ірина ЖУРАВСЬКА

Letter of Support for PhD Research

Dear Yevhen Darnapuk,

Biene IT GmbH expresses strong support for the PhD research titled "**Methods and Tools of Medical Signal Processing to Improve the Effectiveness of Diagnostic Systems**" at **Petro Mohyla Black Sea National University**. This study addresses a highly relevant issue in modern medical data analytics by advancing techniques for processing routine medical signals to enhance diagnostic accuracy, reliability, and cost efficiency.

The research is particularly significant in the domain of medical signal processing, focusing on digital filtration, recurrence analysis, wavelet transformations, and advanced statistical methods to improve diagnostic systems. The proposed approaches have practical implications for automated diagnostic decision-making, real-time patient monitoring, and the enhancement of non-invasive medical diagnostics. Given the growing demand for robust computational methods in digital healthcare, this work aligns with global trends in AI-powered medical technologies and intelligent healthcare data analysis.

At Biene IT GmbH, supporting research that contributes to the improvement of patient care, medical data processing, and intelligent diagnostic systems is a priority. The potential of this study to advance signal-processing algorithms, improve classification accuracy in diagnostic applications, and enable real-time analysis of medical signals using IoT-based and FPGA-integrated solutions is well recognized. As part of this collaboration, the following support is offered:

- **Technical Guidance:** Providing insights into machine learning-based medical diagnostics, real-time signal filtering, and quantitative assessment of physiological data variability.
- **Resource Access:** Assisting in the validation and practical application of findings in real-world healthcare settings, including integration into modern medical monitoring devices.

This PhD research has the potential to make a meaningful impact on both academia and industry. Biene IT GmbH looks forward to engaging with this work and supporting its progress. Should any additional support be required, our team is available to assist. Best wishes for success in this research, and we look forward to seeing its outcomes.

Sincerely,

Patrick Bienert
General-Manager



BIENE IT GmbH
Mainzer Landstr. 69 60329 Frankfurt a. M.
+49 69 5880 99 910 patrick.bienert@bieneit.de

ДОДАТОК Б

Програмний код формування та обробки датасету капілярної оксигенації крові

```

restart:
with(LinearAlgebra,Rank,Dimension,Dimensions,Row,Column,SubMatrix);
with(Statistics):
with(plots,listplot,pointplot,display,setoptions);
plots:-setoptions(axes = boxed, gridlines, font=[Times,16], thickness=3);
with(SignalProcessing,DWT,Periodogram,SignalPlot,DownSample,Threshold);
with(ImageTools,Create,Preview,Embed,Intensity,Entropy,Complement,View);
L:=Vector[row](989,
[te03_0024,te03_0031,te03_0032,te03_0037,te03_0041,te03_0046,te03_0048,te03_0050
,te03_0054,te03_0062,te03_0076,te03_0080,te03_0088,te03_0091,te03_0104,te03_0106
,te03_0112,te03_0125,te03_0132,te03_0137,te03_0145,te03_0162,te03_0193,te03_0195
,te03_0211,te03_0213,te03_0221,te03_0225,te03_0230,te03_0231,te03_0240,te03_0276
,te03_0284,te03_0286,te03_0293,te03_0313,te03_0317,te03_0319,te03_0324,te03_0335
,te03_0368,te03_0371,te03_0383,te03_0407,te03_0416,te03_0422,te03_0436,te03_0440
,te03_0441,te03_0445,te03_0460,te03_0466,te03_0475,te03_0478,te03_0480,te03_0491
,te03_0492,te03_0496,te03_0499,te03_0505,te03_0513,te03_0524,te03_0525,te03_0552
,te03_0557,te03_0577,te03_0586,te03_0593,te03_0603,te03_0604,te03_0608,te03_0611
,te03_0616,te03_0620,te03_0621,te03_0632,te03_0660,te03_0676,te03_0704,te03_0717
,te03_0730,te03_0747,te03_0750,te03_0776,te03_0782,te03_0798,te03_0800,te03_0805
,te03_0811,te03_0819,te03_0827,te03_0839,te03_0848,te03_0856,te03_0858,te03_0864
,te03_0880,te03_0889,te03_0892,te03_0910,te03_0927,te03_0936,te03_0944,te03_0948
,te03_0950,te03_0954,te03_0960,te03_0975,te03_0978,te03_0980,te03_1007,te03_1009
,te03_1016,te03_1025,te03_1039,te03_1051,te03_1055,te03_1059,te03_1067,te03_1072
,te03_1079,te03_1086,te03_1093,te03_1098,te03_1104,te03_1105,te03_1109,te03_1110
,te03_1117,te03_1118,te03_1133,te03_1156,te03_1178,te03_1213,te03_1216,te03_1230
,te03_1291,te03_1308,te03_1318,te03_1320,te03_1325,te03_1330,te03_1334,te03_1342
,te03_1347,te03_1348,te03_1364,te03_1365,te03_1372,te03_1377,te03_1380,te03_1398
,te03_1400,te04_0012,te04_0037,te04_0071,te04_0080,te04_0092,te04_0101,te04_0102
,te04_0103,te04_0107,te04_0125,te04_0130,te04_0141,te04_0149,te04_0153,te04_0154
,te04_0172,te04_0203,te04_0215,te04_0234,te04_0235,te04_0241,te04_0253,te04_0257
,te04_0268,te04_0272,te04_0280,te04_0303,te04_0329,te04_0346,te04_0350,te04_0353
,te04_0361,te04_0367,te04_0373,te04_0378,te04_0384,te04_0395,te04_0398,te04_0406
,te04_0416,te04_0424,te04_0449,te04_0467,te04_0477,te04_0479,te04_0483,te04_0487
,te04_0489,te04_0496,te04_0497,te04_0510,te04_0521,te04_0529,te04_0543,te04_0564
,te04_0574,te04_0581,te04_0582,te04_0596,te04_0602,te04_0624,te04_0641,te04_0651
,te04_0664,te04_0665,te04_0669,te04_0677,te04_0697,te04_0707,te04_0721,te04_0729
,te04_0753,te04_0758,te04_0761,te04_0775,te04_0779,te04_0797,te04_0807,te04_0811
,te04_0813,te04_0827,te04_0832,te04_0838,te04_0842,te04_0845,te04_0849,te04_0857
,te04_0858,te04_0880,te04_0885,te04_0889,te04_0895,te04_0897,te04_0899,te04_0901
,te04_0906,te04_0932,te04_0953,te04_0962,te04_0970,te04_0976,te04_0990,te04_0992
,te04_1018,te04_1031,te04_1035,te04_1046,te04_1047,te04_1065,te04_1086,te04_1118
,te04_1130,te05_0038,te05_0043,te05_0073,te05_0082,te05_0088,te05_0108,te05_0126
,te05_0130,te05_0131,te05_0151,te05_0165,te05_0173,te05_0177,te05_0194,te05_0209
,te05_0255,te05_0271,te05_0272,te05_0280,te05_0283,te05_0290,te05_0314,te05_0339
,te05_0347,te05_0351,te05_0365,te05_0369,te05_0371,te05_0388,te05_0401,te05_0402
,te05_0421,te05_0430,te05_0433,te05_0450,te05_0451,te05_0457,te05_0473,te05_0518
,te05_0540,te05_0541,te05_0551,te05_0555,te05_0567,te05_0571,te05_0584,te05_0606
,te05_0616,te05_0626,te05_0628,te05_0658,te05_0660,te05_0661,te05_0684,te05_0693
,te05_0695,te05_0697,te05_0700,te05_0708,te05_0717,te05_0722,te05_0731,te05_0740
,te05_0747,te05_0786,te05_0804,te05_0810,te05_0830,te05_0832,te05_0836,te05_0842
,te05_0862,te05_0886,te05_0887,te05_0901,te05_0908,te05_0922,te05_0942,te05_0950
,te05_0953,te05_0997,te05_0998,te05_1001,te05_1016,te05_1019,te05_1036,te05_1038
,te05_1044,te05_1049,te05_1062,te05_1069,te05_1073,te05_1075,te05_1079,te05_1087
,te05_1124,te05_1141,te05_1143,te05_1151,te05_1155,te05_1164,te05_1204,te05_1246
,te05_1249,te05_1251,te05_1254,te05_1267,te05_1281,te05_1296,te05_1331,te05_1334
,te05_1335,te05_1352,te05_1378,te05_1380,te05_1390,te05_1416,te05_1420,te05_1421

```

,te05_1427,te05_1436,te05_1439,te05_1470,te05_1484,te05_1495,te05_1500,te05_1505
,te05_1517,te05_1525,te05_1529,te05_1530,te05_1544,te05_1548,te05_1552,te05_1557
,te05_1559,te05_1578,te05_1581,te05_1587,te05_1588,te05_1592,te05_1594,te05_1605
,te05_1621,te05_1623,te05_1626,te05_1627,te05_1628,te05_1633,te05_1634,te05_1640
,te05_1647,te05_1658,te05_1660,te05_1665,te05_1667,te05_1672,te05_1680,te05_1685
,te05_1697,te05_1701,te05_1703,te05_1724,te05_1827,te05_1828,te05_1830,te06_0001
,te06_0002,te06_0016,te06_0021,te06_0031,te06_0032,te06_0033,te06_0034,te06_0038
,te06_0040,te06_0045,te06_0047,te06_0067,te06_0071,te06_0072,te06_0088,te06_0105
,te06_0110,te06_0123,te06_0131,te06_0143,te06_0149,te06_0162,te06_0179,te06_0208
,te06_0209,te06_0235,te06_0237,te06_0248,te06_0252,te06_0260,te06_0266,te06_0273
,te06_0287,te06_0293,te06_0305,te06_0321,te06_0338,te06_0342,te06_0357,te06_0367
,te06_0395,te06_0396,te06_0412,te06_0413,te06_0428,te06_0442,te06_0445,te06_0455
,te06_0457,te06_0461,te06_0482,te06_0484,te06_0493,te06_0518,te06_0519,te06_0540
,te06_0543,te06_0546,te06_0581,te06_0585,te06_0597,te06_0613,te06_0614,te06_0623
,te06_0629,te06_0635,te06_0641,te06_0642,te06_0645,te06_0666,te06_0667,te06_0669
,te06_0681,te06_0692,te06_0695,te06_0699,te06_0713,te06_0717,te06_0731,te06_0744
,te06_0748,te06_0766,te06_0770,te06_0793,te06_0796,te06_0813,te06_0816,te06_0828
,te06_0840,te06_0848,te06_0849,te06_0857,te06_0867,te06_0872,te06_0873,te06_0874
,te06_0886,te06_0887,te06_0896,te06_1008,te06_1110,te06_1118,te06_1130,te06_1177
,te07_0004,te07_0006,te07_0007,te07_0009,te07_0014,te07_0020,te07_0051,te07_0052
,te07_0057,te07_0060,te07_0062,te07_0063,te07_0070,te07_0096,te07_0117,te07_0122
,te07_0128,te07_0146,te07_0149,te07_0158,te07_0172,te07_0175,te07_0177,te07_0185
,te07_0218,te07_0220,te07_0226,te07_0233,te07_0234,te07_0252,te07_0270,te07_0271
,te07_0287,te07_0289,te07_0311,te07_0316,te07_0328,te07_0331,te07_0348,te07_0356
,te07_0361,te07_0365,te07_0378,te07_0405,te07_0417,te07_0433,te07_0447,te07_0476
,te07_0480,te07_0493,te07_0495,te07_0496,te07_0510,te07_0515,te07_0518,te07_0547
,te07_0550,te07_0552,te07_0556,te07_0573,te07_0597,te07_0613,te07_0617,te07_0644
,te07_0671,te07_0674,te07_0677,te07_0682,te07_0683,te07_0692,te07_0693,te07_0694
,te07_0699,te07_0704,te07_0706,te07_0715,te07_0740,te07_0742,te07_0746,te07_0749
,te07_0754,te07_0778,te07_0779,te07_0785,te07_0797,te07_0799,te07_0813,te07_0816
,te07_0831,te07_0838,te07_0839,te07_0840,te07_0844,te07_0851,te07_0898,te07_0903
,te07_0917,te07_0965,te08_0019,te08_0100,te08_0102,te08_0163,te08_0207,te08_0210
,te08_0215,te08_0223,te08_0225,te08_0258,te08_0262,te08_0278,te08_0281,te08_0317
,te08_0320,te08_0321,te08_0325,te08_0352,te09_0001,te09_0003,te09_0032,te09_0035
,te09_0046,te09_0084,te09_0094,te09_0117,te09_0118,te09_0121,te09_0123,te09_0133
,te09_0137,te09_0145,te09_0149,te09_0151,te09_0156,te09_0167,te09_0171,te09_0174
,te09_0193,te09_0222,te09_0228,te09_0231,te09_0248,te09_0272,te09_0277,te09_0285
,te09_0289,te09_0300,te09_0301,te09_0310,te09_0312,te09_0315,te09_0343,te09_0380
,te09_0386,te09_0391,te09_0407,te09_0415,te09_0429,te09_0458,te09_0459,te09_0468
,te09_0481,te09_0502,te09_0521,te09_0538,te09_0548,te09_0549,te09_0550,te09_0554
,te09_0566,te09_0582,te09_0585,te09_0589,te09_0601,te09_0605,te09_0634,te09_0645
,te10_0041,te10_0062,te10_0075,te10_0107,te10_0143,te10_0145,te10_0168,te10_0180
,te10_0186,te10_0208,te10_0230,te10_0244,te10_0262,te10_0276,te10_0279,te10_0299
,te10_0337,te10_0374,te10_0397,te10_0428,te10_0451,te10_0490,te10_0503,te10_0532
,te10_0552,te10_0560,te10_0574,te10_0605,te10_0631,te10_0637,te10_0649,te10_0657
,te10_0679,te10_0684,te10_0738,te10_0755,te10_0759,te10_0766,te10_0774,te10_0791
,te10_0819,te10_0859,te11_0005,te11_0014,te11_0026,te11_0045,te11_0047,te11_0057
,te11_0066,te11_0067,te11_0075,te11_0077,te11_0093,te11_0112,te11_0117,te11_0131
,te11_0141,te11_0164,te11_0167,te11_0170,te11_0173,te11_0189,te11_0196,te11_0313
,te11_0321,te11_0330,te11_0336,te11_0340,te11_0346,te11_0348,te11_0351,te11_0353
,te11_0360,te11_0369,te11_0374,te11_0384,te11_0389,te11_0390,te11_0394,te11_0409
,te11_0412,te11_0420,te11_0425,te11_0444,te11_0477,te11_0480,te11_0482,te11_0488
,te11_0489,te11_0506,te11_0511,te11_0517,te11_0526,te11_0538,te11_0554,te11_0557
,te11_0559,te11_0583,te11_0588,te11_0601,te11_0605,te11_0618,te11_0624,te11_0647
,te11_0651,te11_0664,te11_0667,te11_0675,te11_0684,te11_0695,te11_0697,te11_0702
,te11_0725,te11_0728,te11_0731,te11_0738,te11_0753,te11_0768,te11_0779,te12_0039
,te12_0048,te12_0063,te12_0072,te12_0087,te12_0100,te12_0112,te12_0146,te12_0156
,te12_0164,te12_0180,te12_0186,te12_0188,te12_0226,te12_0239,te12_0246,te12_0256
,te12_0257,te12_0305,te12_0307,te12_0311,te12_0331,te12_0351,te12_0378,te12_0381
,te12_0383,te12_0388,te12_0418,te12_0422,te12_0430,te12_0443,te12_0444,te12_0447
,te12_0458,te12_0459,te12_0466,te12_0482,te12_0486,te12_0487,te12_0488,te12_0498
,te12_0507,te12_0508,te12_0514,te12_0538,te12_0544,te12_0547,te12_0550,te12_0553
,te12_0556,te12_0562,te12_0565,te12_0584,te12_0585,te12_0586,te12_0590,te12_0612
,te12_0628,te12_0636,te12_0641,te12_0651,te12_0653,te12_0660,te12_0664,te13_0062

```

,te13_0066,te13_0072,te13_0102,te13_0116,te13_0123,te13_0135,te13_0137,te13_0169
,te13_0172,te13_0190,te13_0200,te13_0209,te13_0242,te13_0248,te13_0249,te13_0253
,te13_0263,te13_0272,te13_0275,te13_0281,te13_0283,te13_0286,te13_0301,te13_0357
,te13_0360,te13_0371,te13_0377,te13_0390,te13_0402,te13_0409,te13_0415,te13_0418
,te13_0428,te13_0430,te13_0442,te13_0446,te13_0456,te13_0458,te13_0459,te13_0470
,te13_0481,te13_0484,te13_0486,te13_0516,te13_0520,te13_0530,te13_0559,te13_0574
,te13_0597,te13_0602,te13_0609,te13_0622,te13_0629,te13_0633,te13_0634,te13_0636
,te13_0639,te13_0670,te13_0678,te13_0682,te13_0821,te13_0823,te13_0827,te13_0848
,te13_0851,te14_0015,te14_0017,te14_0032,te14_0038,te14_0100,te14_0104,te14_0108
,te14_0115,te14_0123,te14_0124,te14_0146,te14_0159,te14_0168,te14_0176,te14_0177
,te14_0194,te14_0199,te14_0211,te14_0221,te14_0225,te14_0226,te14_0230,te14_0241
,te14_0245,te14_0257,te14_0264,te14_0271,te14_0286]):numelems(L);# List of
samples codes for original data set
ps := {te03_0050, te03_0211, te03_0445, te04_0215, te04_0253, te04_0268,
te04_0303, te04_0373, te05_0606, te05_1141, te05_1378, te05_1517, te05_1548,
te06_0428, te07_0556, te07_0674, te07_0683, te11_0738, te12_0226, te13_0516};#
Randomly selected 20 samples of data set 3
#ps2:={seq(RandomTools:-Generate(choose(L)), j = 1 .. 5)};# Additive 5 samples
ps2;
Digits:=4;
# dat2, name of data for 25 healthy adult set
N1,N2:=Dimension(dat2);dat:=dat2:
BoxPlot([seq(Column(dat,q),q=1..N2)],gridlines,outliers=false);# There are heavy
outliers, especiall in the columns ## 11,19, and 21.
mds:=map(Mode,[seq(Column(dat,q),q=1..N2)],type=sample);min(%);
map(Tally,[seq(Column(dat,q),q=[11,19,21])]);# Outliers and their numbers
999/36e3,534/36e3,399/36e3;# Percentage of heavy outliers for samples 11, 19,
and 21
dt1:=[seq(convert(Threshold(Column(dat,q),1,value=mds[q]),Vector[column]),q=1..N
2)]:whattype(dt1); # Thresholding of initial data. Heavy outliers are
substituted by the modes of columns (11,19,21)
stm:=time():
dt2:=Matrix(1..N1,1..N2,dt1):whattype(dt2);Dimensions(dt2);
tm:=time()-stm;# Corrected data matrix
BoxPlot([seq(Column(dt2,q),q=1..N2)],gridlines,notched=true,
outliers=false,mean=false,view=[0..26,83..100],size=[1000,500],deciles=false);bx
pt:=%:# Corrected data box plot
dataset__initial:={seq(k,k=1..N2)}:
excluded__ids:={11,19,20,21,25}:
dts:=dataset__initial minus excluded__ids;n2:=numelems(dts);
BoxPlot([seq(Column(dt2,q),q=dts)],gridlines,notched=false,
outliers=false,mean=false,view=[0..20.5,88..100],size=[1000,500],deciles=false,t
ickmarks=[default,14],labeldirections=[horizontal,vertical],labels=["IDs","SaO2,
%"],font=[TIMES,14],color=grey..grey, symbol=solidcircle);bxpt:=%:# Corrected
data box plot
SM:=SubMatrix(dt2,1..N1,convert(dts,list)):Dimensions(%);# Sieved final data set
matrix
sus;mid;suw;
iqrs:=map(InterquartileRange,[seq(Column(SM,q),q=1..n2)]);
bwiq:=0.9*min(StandardDeviation(iqrs),InterquartileRange(iqrs))*n2^(-0.2);
R:=Range(iqrs);round(%/bwiq);
Mode(iqrs,type=sample);
Histogram([iqrs],binwidth
=bwiq,gridlines,range=0..R,averageshifted=1,frequenceyscale=relative,color=grey,s
ize=[500,"golden"],tickmarks=[6,5],labeldirections=[horizontal,vertical],labels=
["Interquartile range, %","Probability, arb. units"],font=[TIMES,16]);
Histogram([iqrs],binwidth =bwiq,gridlines,range=-
0.1..R,averageshifted=3,frequenceyscale=relative,color=grey,size=[500,"golden"],t
ickmarks=[6,5],labeldirections=[horizontal,vertical],labels=["Interquartile
range, %","Probability, arb. units"],font=[TIMES,16]);
st_ta:=time():
Ta:=map(Tally,[seq(Column(SM,q),q=1..n2)]):

```

```

t_ta:=time()-st_ta; # Tally function returns a table of distinct elements in
dt2 columns together with their absolute frequencies. Duration of each recodt is
N1/2 samples
ta:=seq(map(rhs,Ta[q]),q=1..n2):
st:=seq(map(lhs,Ta[q]),q=1..n2):
nd:=[seq(nops(ta[q]),q=1..n2)];StandardDeviation(nd);
# Lists of distinct elements (saturation in percents) in columns of SM (st),
lists of their absolute frequencies (ta) and list of numbers of distinct
elements in the everyone of columns (nd)
gr:={seq(j,j=1..n2)};
sus1:={1,2,5,10,15,16};ns1:=nops(%);
suw1:={3,6,11,12,13,14};nw1:=nops(%);
mid1:=(gr minus sus1) minus suw1;nm1:=nops(%);
iqr_s:=map(InterquartileRange,[seq(Column(SM,q),q in sus1)];Mean(%);
iqr_m:=map(InterquartileRange,[seq(Column(SM,q),q in mid1)];Mean(%);
iqr_w:=map(InterquartileRange,[seq(Column(SM,q),q in suw1)];Mean(%);
nds:=[seq(nops(ta[q]),q=sus1)];
ndm:=[seq(nops(ta[q]),q=mid1)];
ndw:=[seq(nops(ta[q]),q=suw1)];
TwoSampleTTest(iqr_m,iqr_s,0,confidence=0.99,output=report);
TwoSampleTTest(iqr_w,iqr_m,0,confidence=0.99,output=report);# The differences
among subgroups are statistically significant
NormalPlot(iqrs,symbol=solidcircle,symbolsize=16,gridlines,view=[default,-
0.5..3]);;
# Difference analysis and grouping
stj:=time():
js:=Difference(SM);whattype(js);n1,n2:=Dimensions(js);
t_j:=time()-stj;# Lagged (lag=1 by default) Differences for columns of SM
stj:=time():
js:=Difference(SM);whattype(js);n1,n2:=Dimensions(js);
t_j:=time()-stj;# Lagged (lag=1 by default) Differences for columns of SM
map(-min,[seq(Column(js,q),q =1..n2)]);
map(max,[seq(Column(js,q),q =1..n2)]);
Ajs:=Array(map(SignalPlot,[seq(Column(js,q),q=1..n2)],view=[0..n1,-
1.1..1.1],font=[TIMES,14]));whattype(%);
display(Ajs[1..4],gridlines,tickmarks=[9,7],font=[TIMES,14]);
display(Ajs[5..8],gridlines,tickmarks=[9,7]);
display(Ajs[9..12],gridlines,tickmarks=[9,7]);
display(Ajs[13..16],gridlines,tickmarks=[9,7]);
display(Ajs[17..20],gridlines,tickmarks=[9,7]);
typ3:=[5,12,3];# the thypical three of graphs for three subgroups
Ajs3:=Array(map(SignalPlot,[seq(Column(SM,q),q=typ3)],view=[0..n1,88..99],size=[
default,250],thickness=3,size=[default,220],font=[TIMES,14]));whattype(%);
display(Ajs3);
typ3:=[5,12,3];# the thypical three of graphs for three subgroups
Ajs3:=Array(map(SignalPlot,[seq(Column(js,q),q=typ3)],view=[0..n1,-
1.1..1.1],size=[default,220],thickness=2,font=[TIMES,14]));whattype(%);
display(Ajs3);
sd:=proc(cm::Vector[column])
local i,j,k,a,ins,N;
description "Non-zero differnece points: numbers and addresses";
N:=numelems(cm);
k:=0;a:=Array(1..N,[0]);
for j from 1 to N do
if cm[j]=0 then next else k:=k+1; a[k]:=j;end if;
end do;
a:=convert(a,set);
a:=convert(a[2..k+1],list);
ins:=convert(Difference(a,times=1,lag=1),list);
return k,a,ins;
end proc;
son:=[seq(sd(Column(js,q))[1],q =1..n2)];# switch-overs numbers
NormalPlot(son,gridlines,symbol=solidcircle,symbolsize=18);
pos:=seq(sd(Column(js,q))[2],q =1..n2):pos[15];# positions of switch-overs

```

```

ils:=seq(sd(Column(js,q))[3],q=1..n2):ils[10];# intervals among switch-overs
bwson:=round(0.9*min(StandardDeviation(son),InterquartileRange(son))*n2^(-0.2));
Histogram(son,binwidth=bwson,gridlines,range=5..45,averageshifted=1,frequenceyscale=relative,color=grey,size=[500,"golden"],tickmarks=[9,default]);hson:=%;
Histogram(son,binwidth=bwson,gridlines,range=5..45,averageshifted=3,frequenceyscale=relative,color=grey,size=[500,"golden"],tickmarks=[9,default]);hson2:=%;
gr:={seq(q,q=1..n2)};
sow:=3,9,13,18;nw:=nops(sow);# weak subgroup in the range (32..42)
sus:=1,2,5,7,10,15,16;ns:=nops(sus);# strong subgroup in the range (9..14)
mid:=(gr minus sow) minus sus;nm:=nops(mid);# middle subgroup in the range (14..26)
nd_s:={seq(nops(ta[q]),q=sus)};Median(%);
nd_m:={seq(nops(ta[q]),q=mid)};Median(%);
nd_w:={seq(nops(ta[q]),q=sow)};Median(%);
sus,mid,sow;typ3;
sonw:={seq(son[q],q=sow)};L3:=Mean(sonw);
sonm:={seq(son[q],q=mid)};L2:=Mean(sonm);
sons:={seq(son[q],q=sus)};L1:=Mean(sons);
TwoSampleZTest(sonm,sons,0,StandardDeviation(sonm),StandardDeviation(sons),confidence=0.99,output=report);
TwoSampleZTest(sonw,sonm,0,StandardDeviation(sonw),StandardDeviation(sonm),confidence=0.99,output=report);
plot((ns/n2)*L1^x*exp(-L1)/x!+(nm/n2)*L2^x*exp(-L2)/x!+(nw/n2)*L3^x*exp(-L3)/x!,x=5..45,gridlines,numpoints=200,style=line,thickness=3,labeldirections=[horizontal,vertical],labels=["Кількість перемикань","Щільність ймовірності,(од.)"],font=[TIMES,14]):mP:=%;
display(hson,mP,size=[600,"golden"]);
plot((ns/n2)*exp(x*ln(L1)-L1-ln(GAMMA(x+1)))+(nm/n2)*exp(x*ln(L2)-L2-ln(GAMMA(x+1)))+(nw/n2)*exp(x*ln(L3)-L3-ln(GAMMA(x+1))),x=5..45,gridlines,numpoints=200,style=line,thickness=3,labeldirections=[horizontal,vertical],labels=["Кількість перемикань","Щільність ймовірності,(од.)"],font=[TIMES,14]):mP2:=%;
display(hson2,mP2,size=[600,"golden"],view=[default,0..0.07]);# Other version of Poisson distribution
ilss:={seq(sd(Column(js,q))[3],q=sus)};means:={seq(Mean(ilss[q]),q=1..ns)};means_cor:=[2844., 2318., 2822., 2440., 2672., 3742.];
ilss_av:=round(Mean(means_cor));delta_s:=round(StandardDeviation(means_cor));%/%;theor_s:=round(n1/L1);delta_st:=evalf(sqrt(n1)/L1);# intervals among switch-overs have to be distributed by Erlang distribution
Histogram(means,bincount=10,gridlines,range=min(means)..max(means),averageshifted=1,frequenceyscale=relative,color=grey,size=[500,"golden"],tickmarks=[5,5],view=[default,0..0.0020],labeldirections=[horizontal,vertical],labels=["Numbers of switch-overs","Probability density, (arb. units)"],font=[TIMES,16]);
evalf(sqrt(n1));
ilsm:={seq(sd(Column(js,q))[3],q=mid)};meanm:={seq(Mean(ilsm[q]),q=1..nm)};ilsm_av:=round(Mean(meanm));delta_m:=round(StandardDeviation(meanm));%/%;theor_m:=round(n1/L2);delta_mt:=evalf(sqrt(n1)/L2);
Histogram(meanm,bincount=10,gridlines,range=min(meanm)..max(meanm),averageshifted=5,frequenceyscale=relative,color=grey,size=[500,"golden"],tickmarks=[5,5],view=[default,0..0.0035]);
ilsw:={seq(sd(Column(js,q))[3],q=sow)};meanw:={seq(Mean(ilsw[q]),q=1..nw)};ilsw_av:=round(Mean(meanw));delta_w:=round(StandardDeviation(meanw));%/%;theor_w:=round(n1/L3);delta_wt:=evalf(sqrt(n1)/L3);
Histogram(meanw,bincount=10,gridlines,range=min(meanw)..max(meanw),averageshifted=5,frequenceyscale=relative,color=grey,size=[500,"golden"],tickmarks=[5,5],view=[default,0..0.015]);
TwoSampleZTest(means,meanm,0,StandardDeviation(means),StandardDeviation(meanm),confidence=0.99,output=report);
TwoSampleZTest(meanm,meanw,0,StandardDeviation(meanm),StandardDeviation(meanw),confidence=0.99,output=report);

```

```

dbf:=DataFrame( < theor__s, theor__m, theor__w;
round(delta__st),round(delta__mt),round(delta__wt);ilss__av, ilsm__av,
ilsw__av;delta__s,delta__m,delta__w >, 'rows' = [ 'Av_int_theor',
'STD_teor','Av_int_observ','STD_exp' ], 'columns' = [ 'Strong', 'Middle', 'Weak'
] );
DocumentTools:-Tabulate(dbf,width=75):
fs1:=[L1/180,L2/180,L3/180];fs2:=[1/ilss__av, 1/ilsm__av,1/ilsw__av]*200.;
# Events (switches) frequencies in Hz
ils:=seq(sd(Column(js,q))[3],q =1..n2):
whattype(ils);
nops(ils);
ils_av:=map(Mean,ils);
bwil:=0.9*min(StandardDeviation(ils_av),InterquartileRange(ils_av))*n2^(-0.2);
Histogram(ils_av,binwidth=bwil,gridlines,range=min(ils_av)..max(ils_av),averages
hifted=1,frequency scale=relative,color=grey,size=[500,"golden"],tickmarks=[5,5],
view=[default,0..0.0009],labeldirections=[horizontal,vertical],labels=["Довжина
інтервалів","Щільність ймовірності, (од.)"],font=[TIMES,14]);ho:=%:
Histogram(ils_av,binwidth=bwil,gridlines,range=0..4000,averageshifted=3,frequency
scale=relative,color=grey,size=[500,"golden"],tickmarks=[5,6],view=[default,0..
0.001],labeldirections=[horizontal,vertical],labels=["Довжина
інтервалів","Щільність ймовірності, (од.)"],font=[TIMES,14]);hash:=%:
lambda:='lambda';whattype(Ng);N1;
Ng:=(x,lambda)->sqrt(2.)*exp(-(x -
(36e3/lambda))^2/(2.*(36e3/lambda^2)^2))/(2*sqrt(Pi)*(36e3/lambda^2));
L1,L2,L3;ns,nm,nw;
plot((ns/n2)*Ng(x,L1)+(nm/n2)*Ng(x,L2)+(nw/n2)*Ng(x,L3),x=0..4000,gridlines,nump
oints=2000,style=line,thickness=3,labeldirections=[horizontal,vertical],labels=[
"Numbers of switch-overs","Probability density, (arb.
units)"],font=[TIMES,16],size=[500,"golden"],labeldirections=[horizontal,vertica
l],labels=["Довжина інтервалів","Щільність ймовірності,
(од.)"],font=[TIMES,14]);erl:=%:
display(erl,hash);
Digits:=4:
NormalPlot(son,gridlines,symbol=solidcircle,symbolsize=18,labeldirections=[horiz
ontal,vertical],labels=["Expected values","Observed values"],font=[TIMES,16]);
ils_av;son;Correlation(ils_av,son);
ScatterPlot(son,ils_av,gridlines,symbol=solidcircle,symbolsize=24,fit = [a*t +b,
t]);
with(Student:-Statistics):ct:=CriticalTable('StudentT', alpha = [0.99],
parameter1 = n2);ct[2,2];
Corr_critical:=t/sqrt(n2-2+t^2);
cc_crit:=eval(Corr_critical,t=ct[2,2]);
# critical value of corr coeffs essential non-zero
typ3;fs:=[L1,L2,L3]/180;fs2;
Periodogram(Column(js,3),gridlines,powerscale="absolute",samplerate=200,thicknes
s=1,size=[500,"golden"],tickmarks=[4,5],view=[0.01..0.075,default],thickness=2,f
ont=[TIMES,14], labels=["Частота (Гц)","Потужність"]);
Periodogram(Column(js,12),gridlines,powerscale="absolute",samplerate=200,thickne
ss=1,size=[550,"golden"],tickmarks=[4,4],view=[0.075..0.15,default],thickness=2,
labels=["Частота (Гц)","Потужність"]);
Periodogram(Column(js,5),gridlines,powerscale="absolute",samplerate=200,thicknes
s=1,size=[550,"golden"],tickmarks=[default,4],view=[0.15..0.25,default],thicknes
s=2, labels=["Частота (Гц)","Потужність"]);
fse:=[0.057,0.109,0.200];fs;seq((fs[j]-fse[j])*100/fs[j],j=1..3);
evalf(log[2](N1));2^15;delta__f:=200./2^14;# the frequency resolution

```

ДОДАТОК В

Список опублікованих наукових праць за темою дисертації

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

1. Chuiko G., Darnapuk Y., Dvornik O., Krainyk Y. Abdominal electromyograms mining: breathing patterns of asleep adults. *Radioelectronic and Computer Systems*. 2023. No. 3. P. 60–72. DOI: <https://doi.org/10.32620/reks.2023.3.06>. ISSN 1814-4225. **(наукове фахове видання України кат. А, Scopus Q3)**

2. Chuiko G., Darnapuk Y. Fractal nature of arterial blood oxygen saturation data. *Radioelectronic and Computer Systems*. 2022. No. 1. P. 206–215. DOI: <https://doi.org/10.32620/reks.2022.1.16>. ISSN 1814-4225. **(наукове фахове видання України кат. А, Scopus Q3)**

3. Chuiko G., Darnapuk Y. Development of a method for differential analysis of data on the arterial blood oxygenation in healthy adults. *Eastern-European Journal of Enterprise Technologies*. 2021. Vol. 6, no. 4 (114). P. 37–43. DOI: <https://doi.org/10.15587/1729-4061.2021.244924>. ISSN 1729-3774. **(наукове фахове видання України кат. А, Scopus Q2)**

4. Chuiko G., Dvornik O., Darnapuk Y., Baganov Y. Devising a new filtration method and proof of self-similarity of electromyograms. *Eastern-European Journal of Enterprise Technologies*. 2021. Vol. 4, no. 9 (112). P. 15–22. DOI: <https://doi.org/10.15587/1729-4061.2021.239165>. ISSN 1729-3774. **(наукове фахове видання України кат. А, Scopus Q2)**

Список публікацій, які засвідчують апробацію матеріалів дисертації:

5. Chuiko G., Darnapuk Y., Dvornik O., Krainyk Y., Yaremchuk O., Haab R. A New Way of Data Analysis and Rating of the Blood Oxygen Saturation Variability. *2021 IEEE 12th International Conference on Electronics and Information Technologies (ELIT)*, Lviv, Ukraine, 19–21 May, 2021. DOI: <https://doi.org/10.1109/ELIT53502.2021.9501109>. **Scopus**

6. Chuiko G., Dvornik O., Darnapuk Ye., Honcharov D., Yaremchuk O. Asleep Adults' Breathing Patterns via Data Mining of Electromyograms. *2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, Dortmund, Germany, 7–9 Sept., 2023. DOI: <https://doi.org/10.1109/idaacs58523.2023.10348674>. **Scopus**
7. Chuiko G., Honcharov D., Dvornik O., Krainyk Y., Darnapuk Y., Yaremchuk O. Attribute Selection, Outliers Impact Study and Visualization within Breast Cancer Detection. *2023 IEEE 13th International Conference on Electronics and Information Technologies (ELIT)*, Lviv, Ukraine, 26–28 Sept., 2023. DOI: <https://doi.org/10.1109/ELIT61488.2023.10310922>. **Scopus**
8. Chuiko G., Darnapuk Ye., Dvornik O., Gravenor M., Yaremchuk O. Data Mining of Ambulatory Blood Pressure Monitoring. *2023 13th International Conference on Dependable Systems, Services and Technologies (DESSERT)*, Athens, Greece, 13–15 Oct., 2023. DOI: <https://doi.org/10.1109/dessert61349.2023.10416531>. **Scopus**
9. Krainyk Y., Darnapuk Y., Simakova I. Software System for Physical Activity Monitoring: Smart Watch Case. *2020 IEEE 5th International Symposium on Smart and Wireless Systems within the Conferences on Intelligent Data Acquisition and Advanced Computing Systems (IDAACS-SWS)*, Dortmund, Germany, 17–18 Sept., 2020. DOI: <https://doi.org/10.1109/idaacs-sws50031.2020.9297093>. **Scopus**
10. Chuiko G., Darnapuk Y., Dvornik O., Yaremchuk O., Krainyk Y. Recurrence Plots for Ambulatory Blood Pressure Monitoring : Means for Data mining of circadian rhythms. *2020 IEEE 15th International Conference on Computer Sciences and Information Technologies (CSIT)*, Zbarazh, Ukraine, 23–26 Sept., 2020. DOI: <https://doi.org/10.1109/csit49958.2020.9321837>. **Scopus**
11. Chuiko G., Darnapuk Ye., Dvornik O., Yaremchuk O. Blood oxygenation among healthy adults: Recurrence plots analysis and quantification. *2nd International Workshop on Information-Communication Technologies & Embedded Systems (ICT&ES-2020)*, Mykolaiv, Ukraine, 12 Nov., 2020. CEUR Workshop Proc. 2020.

Vol. 2762. P. 218–226. ISSN 1613-0073. URL: <http://ceur-ws.org/Vol-2762/short3.pdf> (Last accessed: 24.01.2025). **Scopus**

***Список публікацій, які додатково відображають наукові результати
дисертації***

12. Chuiko G., Krainyk Y., Dvornik O., Darnapuk Y. IoT in Provenance Management of Medical Data. *Semantic IoT: Theory and Applications* : monograph. Cham, 2021. P. 347–362. DOI: https://doi.org/10.1007/978-3-030-64619-6_15.

Scopus

13. Chuiko G., Dvornik O., Darnapuk Y., Krainyk Y. Principal Component Analysis, Quantifying, and Filtering of Poincaré Plots for time series typical for E-health. *Health Informatics: A Computational Perspective in Healthcare* : monograph. Singapore, 2021. P. 61–76. DOI: https://doi.org/10.1007/978-981-15-9735-0_4.

Scopus

14. Методи обробки медичних сигналів засобами комп'ютерної математики Maple : монографія / О. В. Дворник, Г. П. Чуйко, Є. С. Дарнапук, Я. М. Крайник. Миколаїв : Іліон, 2021. 186 с. ISBN 978-617-534-666-2. URL: <https://dspace.chmnu.edu.ua/jspui/handle/123456789/529> (дата звернення: 20.01.2025).

15. Chuiko G., Darnapuk Ye., Dvornik O., Yaremchuk O. Abdominal Electromyograms (EMGs) Dataset: Breathing Patterns of Sleeping Adults. *Mendeley Data*. 2023. V3. DOI: <https://doi.org/10.17632/pmspdmgcd4.3>.

16. Chuiko G., Darnapuk Ye., Dvornik O., Yaremchuk O. Efficacy of Weka for medical data mining: Ambulatory blood pressure monitoring as a case-study. *Online Journal of Cardiology Research & Reports*. 2023. Vol. 7, Is. 3. P. 2–6. ISSN 2693-4965. DOI: 10.33552/OJCRR.2023.7.000661.

17. Цифрова обробка сигналів з практикумом в Maple 2021 : навч. посіб. / О. В. Дворник, Г. П. Чуйко, Є. С. Дарнапук, А. П. Бойко. Миколаїв : ЧНУ ім. Петра Могили, 2024. 256 с. ISBN 978-966-336-457-5. URL: <https://dspace.chmnu.edu.ua/jspui/handle/123456789/2494> (дата звернення: 20.01.2025).