

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет інформаційно-аналітичних технологій та менеджменту

(повна назва)

Кафедра прикладної математики

(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА Пояснювальна записка

рівень вищої освіти другий (магістерський)

Дослідження і класифікація електрофізіологічних
сигналів за допомогою машинного навчання

(тема)

Виконав:

студент 2 курсу, групи САУМ-22-1

Харченко Я. О.

(прізвище, ініціали)

Спеціальність 124 Системний аналіз

(код і повна назва спеціальності)

Тип програми освітньо-професійна

(освітньо-професійна або освітньо-наукова)

Освітня програма Системний аналіз і управління

(повна назва освітньої програми)

Керівник проф. Кіріченко Л.О.

(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри ПМ

(підпис)

Сидоров М.В.

(прізвище, ініціали)

2024 р.

Харківський національний університет радіоелектроніки

Факультет інформаційно-аналітичних технологій та менеджменту

Кафедра прикладної математики

Рівень вищої освіти другий (магістерський)

Спеціальність 124 Системний аналіз

(код і повна назва)

Тип програми освітньо-професійна

(освітньо-професійна або освітньо-наукова)

Освітня програма Системний аналіз і управління

(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри ПМ _____

(підпис)

“ 06 ” листопада 2023 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Харченко Ярославу Олеговичу
(прізвище, ім'я, по батькові)

1. Тема роботи Дослідження і класифікація електрофізіологічних сигналів за допомогою машинного навчання

затверджена наказом по університету від 2 листопада 2023 р. № 1277 Ст

2. Термін подання студентом роботи до екзаменаційної комісії 10 січня 2024 р.

3. Вихідні дані до роботи сигнали електрокардіограми

4. Перелік питань, що потрібно опрацювати в роботі _____

1. Системний аналіз предметної області

2. Вибір і обґрунтування методу розв'язання

3. Програмна реалізація

4. Результати обчислювального експерименту

5. Аналіз можливих застосувань

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

1. Актуальність теми роботи _____

2. Постановка задачі _____

3. Системний аналіз предметної області _____

4. Метод чисельного аналізу _____

5. Результати обчислювального експерименту _____

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Підбір та вивчення технічної літератури за темою роботи	6 – 12 листопада 2023 р.	виконано
2	Вибір та обґрунтування методу	13 – 26 листопада 2023 р.	виконано
3	Розробка алгоритму і програми	27 листопада – 10 грудня 2023 р.	виконано
4	Проведення аналітичних досліджень та розрахунків	11 грудня – 24 грудня 2023 р.	виконано
5	Робота над текстом пояснювальної записки	25 грудня 2023 р. – 9 січня 2024 р.	виконано
6	Представлення роботи на рецензію в ЕК	10 січня 2024 р.	виконано

Дата видачі завдання 6 листопада 2023 р.

Студент _____
(підпис)

Керівник роботи _____ проф. Кіріченко Л.О.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 67 с., 2 табл., 25 рис., 1 дод., 21 джерело.

ЕЛЕКТРОКАРДІОГРАМА (ЕКГ), ЧАСОВІ РЯДИ, КЛАСИФІКАЦІЯ, МАШИННЕ НАВЧАННЯ, ЗГОРТКОВІ НЕЙРОННІ МЕРЕЖІ (CNN), МЕТОД МУЛЬТИФРАКТАЛЬНОГО ФЛУКТУАЦІЙНОГО АНАЛІЗУ (МФДФА), PYTHON, NEUROKIT2, TENSORFLOW, KERAS.

Об'єкт дослідження – завдання класифікації біоелектричних сигналів.

Мета роботи – застосування нейронних мереж та методу мультифрактального флуктуаційного аналізу для дослідження та класифікації біоелектричних сигналів.

Методи дослідження – метод мультифрактального флуктуаційного аналізу (МФДФА) для аналізу даних та згорткові нейронні мережі (CNN) для класифікації.

Результати експерименту демонструють, що обраний метод класифікації виявився досить точним і на тренувальному, і на тестовому наборах даних. Використання мультифрактальних характеристик у контексті дослідження біоелектричних сигналів могло б допомогти відділити корисну інформацію від шуму та покращити точність класифікації. Висновки з цієї роботи можуть бути використані для виявлення відхилень та для класифікації біоелектричних сигналів із застосуванням машинного навчання.

ABSTRACT

Introductory note: 67 pages, 2 tables, 25 figures, 1 appendix, 21 source.

ELECTROCARDIOGRAM (ECG), TIME SERIES, CLASSIFICATION, MACHINE LEARNING, CONVOLUTIONAL NEURAL NETWORKS (CNN), MULTIFRACTAL FLUCTUATION ANALYSIS (MFDFA) METHOD, PYTHON, NEUROKIT2, TENSORFLOW, KERAS.

Object of research – the task of bioelectrical signals classification.

Purpose of work – to apply neural networks and the method of multifractal fluctuation analysis to research and classify bioelectrical signals.

Methods of research –method of multifractal fluctuation analysis (MFDFA) for data analysis and convolutional neural networks (CNN) for classification.

The results of the experiment demonstrate that the chosen classification method proved to be highly accurate on both training and test datasets. The use of multifractal characteristics in the context of bioelectrical signal research could help distinguish useful information from noise and improve classification accuracy. The conclusions from this work can be utilized for anomaly detection and classification of bioelectrical signals using machine learning.

ЗМІСТ

	С.
Перелік скорочень, умовних познач, одиниць і термінів	8
Вступ	9
1 Системний аналіз предметної області та постановка задач дослідження	11
1.1 Системний аналіз задачі класифікації біоелектричних сигналів	11
1.2 Змістовна та формальна постановка задачі	18
1.3 Постановка завдань дослідження	26
2 Вибір та обґрунтування методу розв'язання	27
2.1 Метод детрендованого флуктуаційного аналізу	27
2.2 Мультифрактальний флуктуаційний аналіз	28
2.3 Методи класифікації біоелектричних сигналів	33
2.4 Глибоке навчання як інструмент класифікації	33
2.5 Згорткові нейронні мережі в задачах класифікації часових рядів	34
Висновки за розділом 2	36
3 Програмна реалізація	37
3.1 Python як високорівнева мова програмування	37
3.2 Бібліотеки TensorFlow / Keras та Scikit-learn	38
3.3 Бібліотека NeuroKit2	39
3.4 Алгоритм розв'язання задачі класифікації біоелектричних сигналів	40
3.5 Опис програми	41
Висновки за розділом 3	42
4 Результати обчислювального експерименту та їх аналіз	43
4.1 Вхідні дані.....	43
4.2 Розрахунок мультифрактальних характеристик	46
4.3 Проведення експерименту	48
4.4 Точність роботи класифікатора	51
Висновки за розділом 4	54
Висновки	55

Перелік джерел посилання	56
Додаток А Лістинг програми	58

ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАК, ОДИНИЦЬ І ТЕРМІНІВ

ДФА – метод детрендованого флуктуаційного аналізу;

МФДФА – мультифрактальний детрендований флуктуаційний аналіз;

ЕКГ – електрокардіограма;

ЕЕГ – електроенцефалограма;

ЕМГ – електроміографія;

DNN – глибока нейронна мережа;

CNN – згорткова нейронна мережа;

RNN – рекурентна нейронна мережа;

ReLU – зрізаний лінійний вузол.

ВСТУП

Актуальність теми. Дослідження біоелектричних сигналів з використанням методів машинного навчання відкриває перед медичною галуззю можливість раннього виявлення та діагностики різноманітних захворювань, що дозволяє забезпечити ефективне та вчасне лікування пацієнтів. Використання машинного навчання в аналізі біомедичних сигналів дозволяє автоматизувати та поліпшити процеси обробки та інтерпретації великих обсягів медичної інформації, що сприяє прискоренню наукових досліджень та розробці нових методів лікування. Результати таких досліджень можуть мати важливий вплив на клінічну практику, сприяючи вдосконаленню діагностики та терапії пацієнтів, а також розробці інноваційних медичних технологій.

Дослідження біоелектричних сигналів з використанням машинного навчання є актуальною і важливою темою в сучасній медичній науці. Ці методи відіграють значну роль у діагностиці, моніторингу стану пацієнтів та оптимізації лікування.

Фрактальний аналіз та машинне навчання дозволяють більш точно виявляти закономірності та кореляції у біомедичних даних, що має значення для більш точної діагностики різних захворювань, моніторингу пацієнтів та розробки персоналізованих методів лікування. Крім того, ці методи сприяють пошуку нових наукових відкриттів та інтеграції з сучасними технологіями штучного інтелекту, посилюючи їх ефективність та застосовність у медичній практиці.

Мета та завдання кваліфікаційної роботи. Метою кваліфікаційної роботи є застосування нейронних мереж та методу мультифрактального флуктуаційного аналізу для дослідження та класифікації біоелектричних сигналів. Для досягнення поставленої мети необхідно виконати наступні завдання:

– провести огляд та аналіз сучасного стану задачі класифікації біоелектричних сигналів за допомогою машинного навчання та методів мультифрактального аналізу;

- вивчити основні поняття та властивості мультифрактальних випадкових процесів;
- вивчити метод мультифрактального флуктуаційного аналізу;
- провести огляд методів класифікації біоелектричних сигналів;
- вивчити основні поняття теорії машинного навчання та нейронних мереж;
- провести мультифрактальний аналіз нормальних та аномальних біоелектричних сигналів;
- провести аналіз та класифікацію реальних біоелектричних сигналів за допомогою машинного навчання.

Об'єктом дослідження є завдання класифікації біоелектричних сигналів.

Предметом дослідження є розв'язання задачі класифікації біоелектричних сигналів за допомогою машинного навчання.

Методи дослідження. У кваліфікаційній роботі використовується метод мультифрактального флуктуаційного аналізу (МФДФА) для аналізу даних та згорткові нейронні мережі для класифікації.

Публікації. Результати, отримані у кваліфікаційній роботі, було представлено на 27-му Міжнародному молодіжному форумі «Радіоелектроніка та молодь у ХХІ столітті» (м. Харків, 10-12 травня 2023 р.) [21].

1 СИСТЕМНИЙ АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧ ДОСЛІДЖЕННЯ

1.1 Системний аналіз задачі класифікації біоелектричних сигналів

Об'єкт дослідження – модельні реалізації фрактальних процесів та біоелектричні сигнали.

Предмет дослідження – класифікація біоелектричних сигналів (ЕКГ).

Мета: класифікація реалізацій фрактальних процесів за допомогою машинного навчання з огляду на результати методів мультифрактального аналізу.

Фрактал (лат. fractus – подрібнений, зламаний, розбитий) – складна геометрична фігура, що має властивість самоподібності, тобто складена з кількох частин, кожна з яких подібна до всієї фігури цілком. У ширшому значенні під фракталами розуміють множину точок в евклідовому просторі, що мають дробову метричну розмірність (у сенсі Мінковського або Хаусдорфа), або метричну розмірність, відмінну від топологічної. Класичним зразком фракталу є множина Мандельброта.

Мультифрактал – комплексний фрактал, який може детермінуватися не одним єдиним алгоритмом побудови, а кількома алгоритмами, що послідовно змінюють один одного. Кожен із них генерує патерн зі своєю фрактальною розмірністю. Для опису мультифрактала обчислюють мультифрактальний спектр включає ряд фрактальних розмірностей властивих елементам даного мультифрактала.

Досліджувана система – дослідження мультифрактальних властивостей довільних процесів.

Метафункція системи – вивчення параметрів мультифрактальних сигналів.

Морфологічна модель системи «Дослідження мультифрактальних властивостей випадкових процесів» є сукупністю моделей [1]: модель

зовнішнього середовища, модель типу «чорна скриня» та модель типу «біла скриня». Усі ці системи наочно можна зобразити схематично. При створенні моделі «зовнішнього середовища» враховуються процеси взаємодії системи та навколишнього його зовнішнього середовища. При цьому слід виділити та обґрунтувати параметри, що достовірно відображають взаємозв'язки між елементами зовнішнього середовища та складовими внутрішнього середовища системи. Таких взаємозв'язків існує безліч, але для побудови моделей беруться тільки ті, які мають найбільшу стійкість і мають вирішальне значення для функціонування системи. Модель системи типу «зовнішнього середовища» зображено на рисунку 1.1.



Рисунок 1.1 – Модель системи типу «зовнішнього середовища»

Модель системи типу «чорна скринька» відображає дві важливі властивості системи – цілісність та відокремленість від середовища. Система не є повністю ізольованою від середовища, вона пов'язана із середовищем і за допомогою цих зв'язків взаємодіє з нею (входи та виходи системи). У цій моделі відсутні відомості про внутрішній зміст системи, а задаються, фіксуються та перераховуються лише входні та вихідні зв'язки системи із середовищем. В даному випадку модель є змістовним словесним описом входів та виходів, тоді вона є просто їх списком. Модель системи типу «чорна скриня» представлена на рисунку 1.2.

Склад системи «Дослідження мультифрактальних властивостей випадкових процесів» можна подати у вигляді моделі «білої скриньки» (склад системи) з наступною структурою:

- A_0 аналіз отриманих характеристик наявність мультифрактальної структури вихідного сигналу;
- A_1 одержання характеристик вихідного сигналу;
- A_2 обробка сигналу методом МДФА:
- A_{21} побудова куммулятивного ряду (розбиття на N сегменти завдовжки s);
- A_{22} обчислення флуктуаційної функції для кожного сегмента та усереднюємо результати по ряду;
- A_{23} знаходимо залежність $F(q)$ при різному s ;
- A_3 обчислюємо мультифрактальні характеристики.

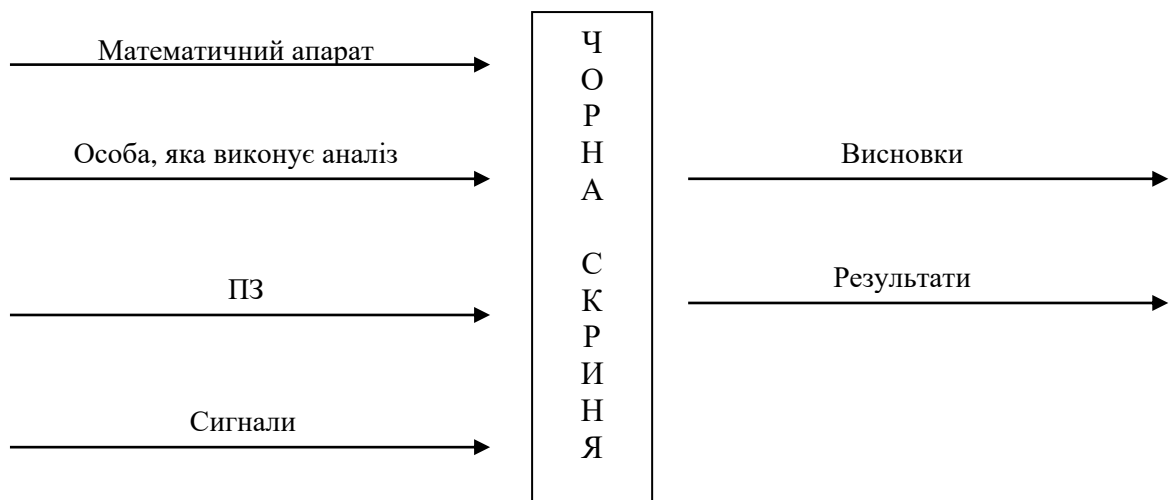


Рисунок 1.2 – Модель системи типу «чорна скриня»

Модель системи типу «біла скриня» (склад системи) показує взаємозв'язок різних структур між собою, графічний опис системи показано рис. 1.3.

Збудуємо функціональну модель системи. Функціональна модель системи є важливою моделлю, за допомогою якої оцінюється логіка функціонування системи та об'єктів, що пов'язують ці функції.

Перша діаграма IDEF0 [1] зображує функціонування системи загалом,

називається контекстної, вона показано рис. 1.4. Для системи "Дослідження мультифрактальних властивостей випадкових процесів" представлена модель декомпозиції основної функції "Провести аналіз біоелектричного сигналу методом МДФА". Кожна функція її відповідає одній з частин системи, що її виконує.

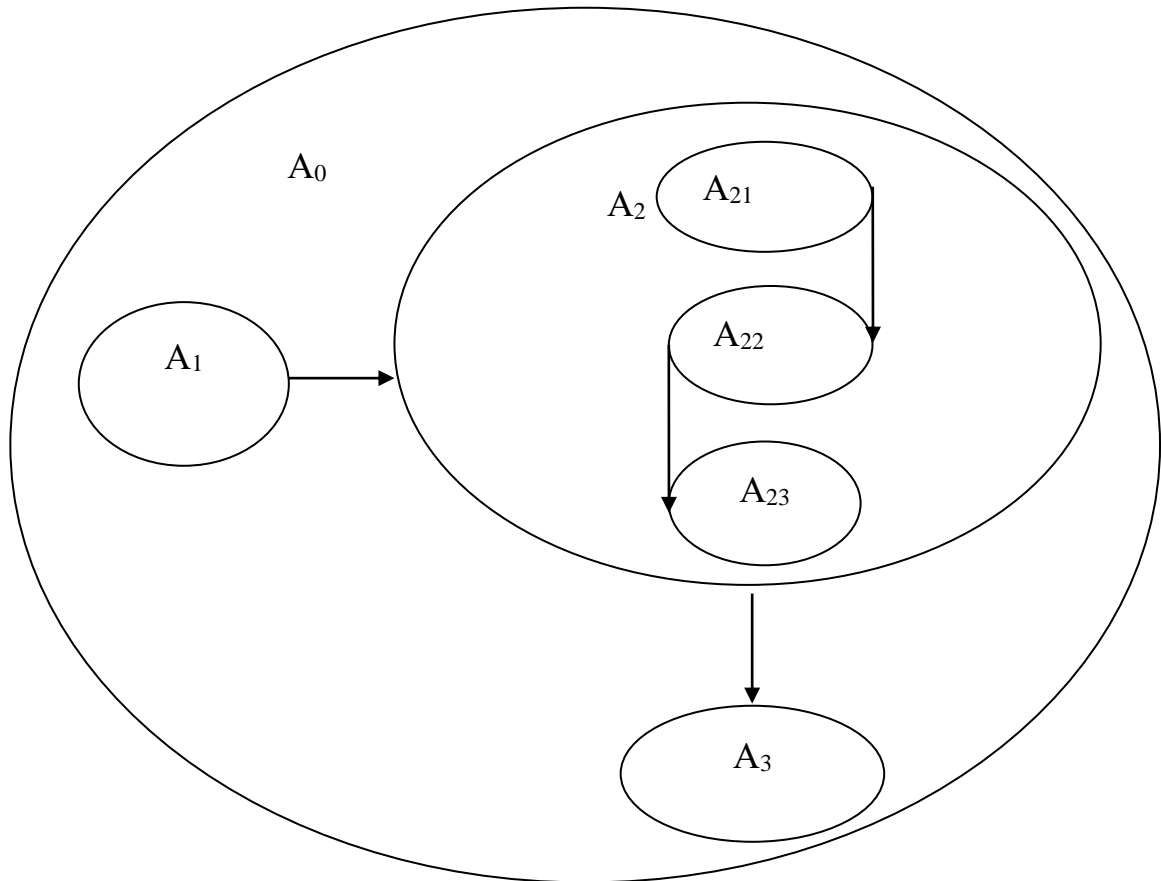


Рисунок 1.3 – Модель системи типу «біла скриня»

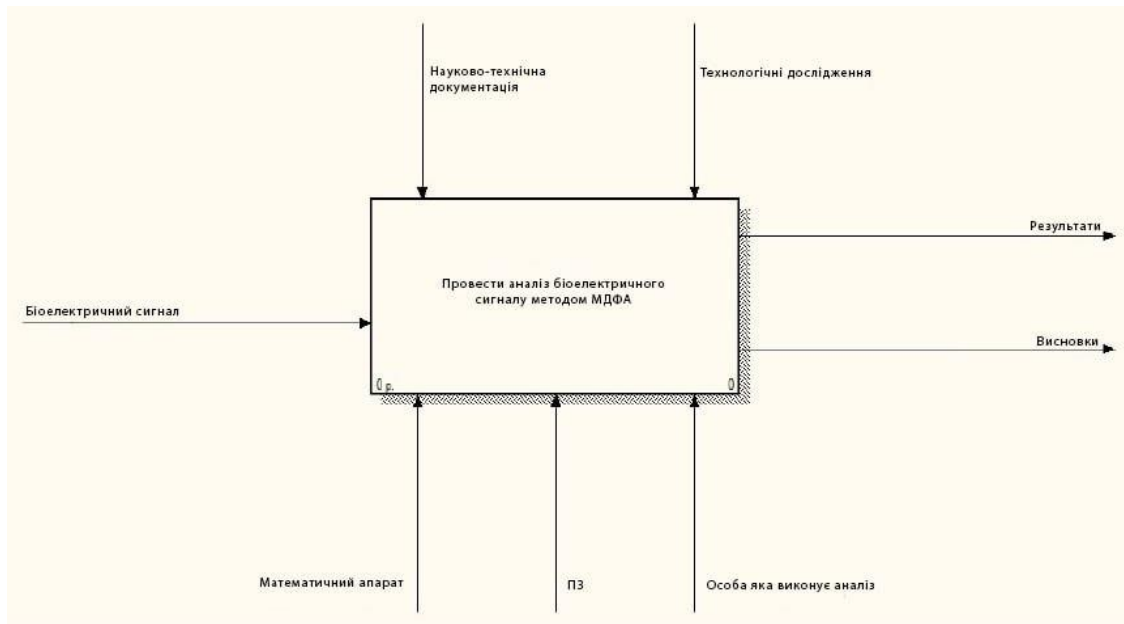


Рисунок 1.4 – Контекстна діаграма, модель IDEF0

Кожна наступна діаграма є більш докладним описом однієї з робіт на попередній діаграмі. Друга діаграма показує основні функції системи. Ці діаграми представлені на рис. 1.5, 1.6.

Зобразимо технологічну модель системи. За допомогою стандарту IDEF3 [1] розглядається процес аналізу та обробки інформації. У цьому стандарті містяться логічні оператори, з допомогою яких показуються альтернативи та місця прийняття рішень у процесі, і навіть з допомогою об'єктів-стрілок, показується тимчасова послідовність робіт у процесі. Ці процеси представлені рисунку 1.7.

Побудуємо інформаційну модель системи. Інформаційна модель є описом потоків даних як документообігу чи інформаційних потоків, тобто. потоків даних, що переміщуються між зовнішніми об'єктами, завданнями та сховищами даних.

Графічне зображення цієї моделі наведено на рисунках 1.8 та 1.9.

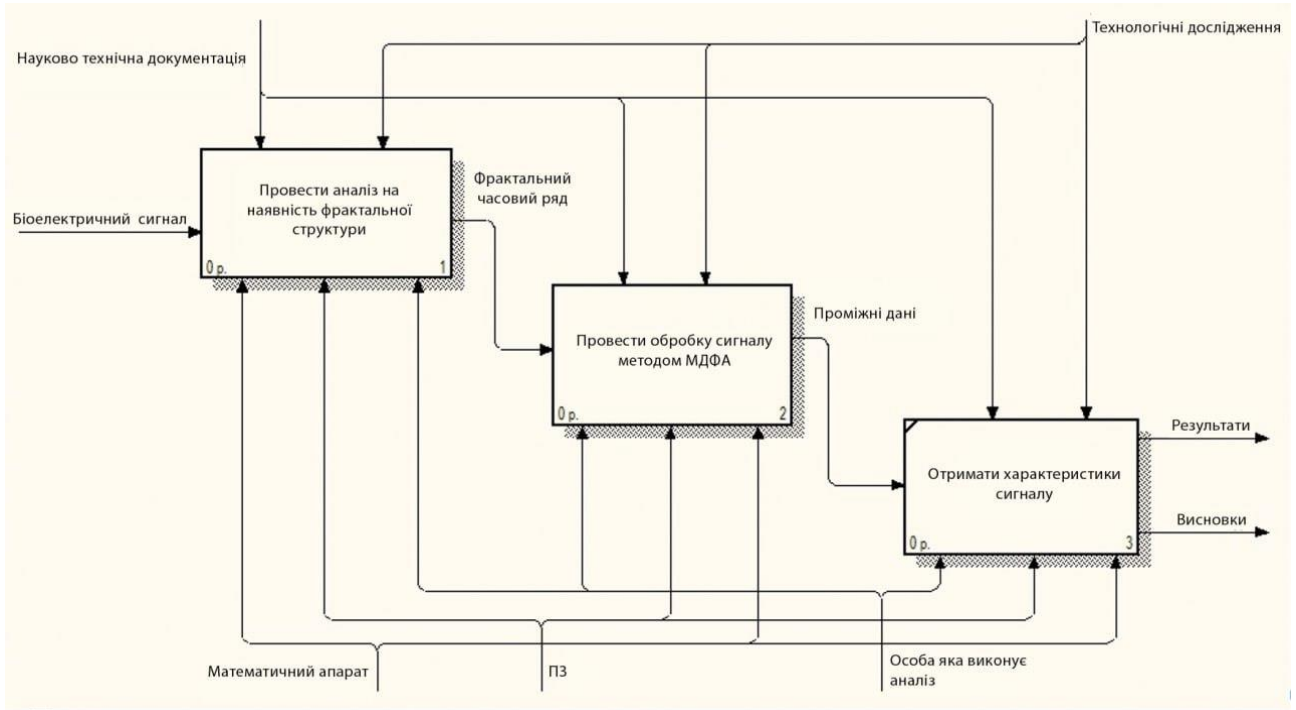


Рисунок 1.5 – Перший рівень декомпозиції моделі IDEF0

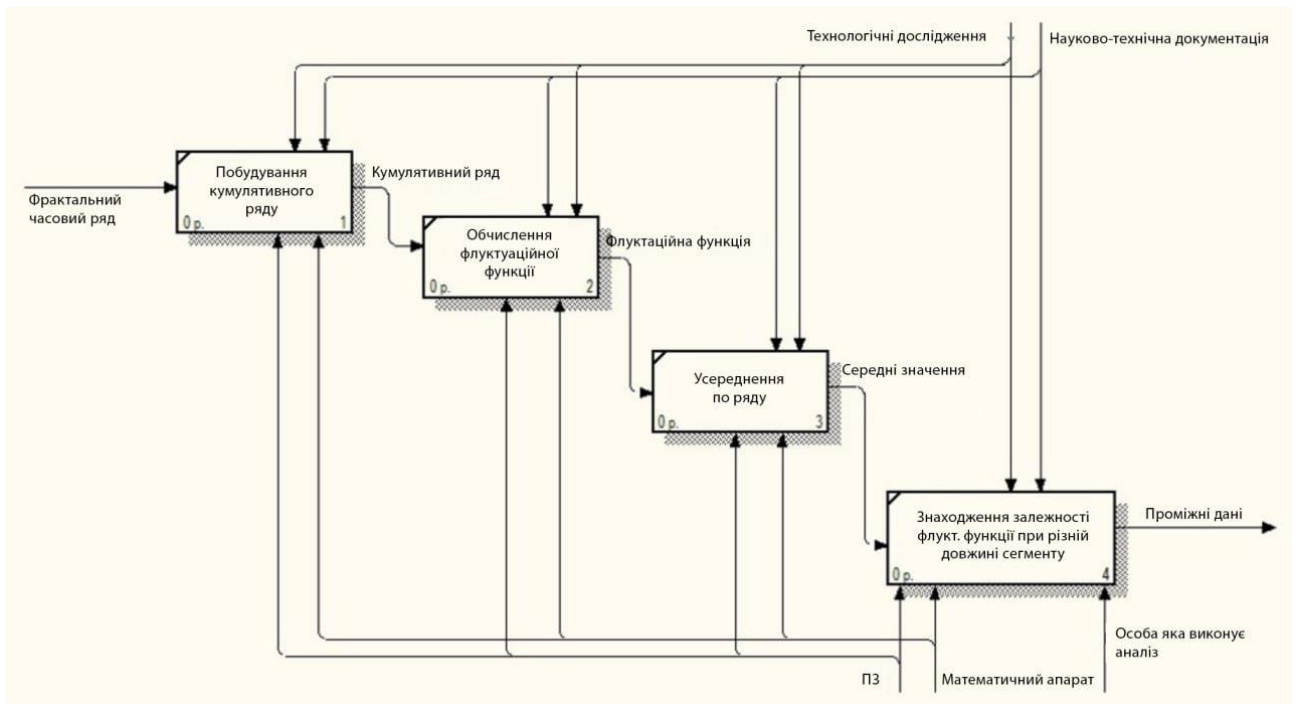


Рисунок 1.6 – Другий рівень декомпозиції моделі IDEF0, декомпозиція блоку "Провести обробку сигналу методом МДФА"

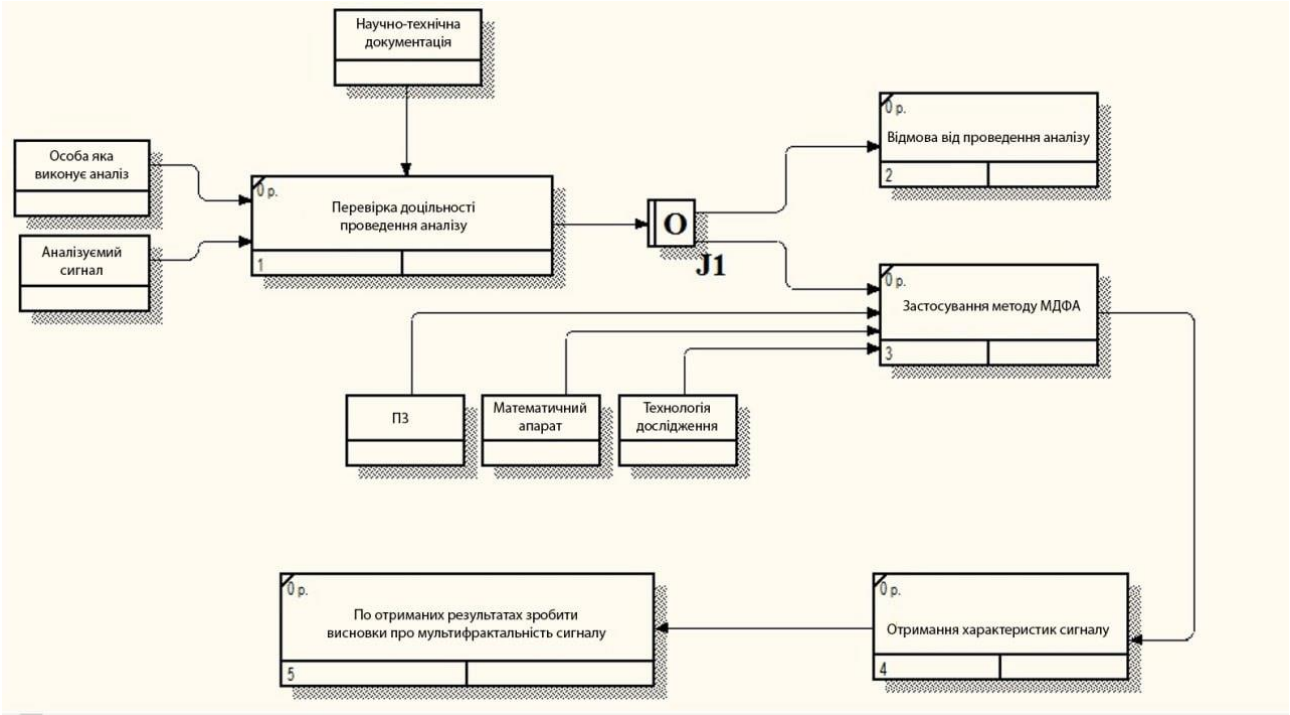


Рисунок 1.7 – Модель IDEF3

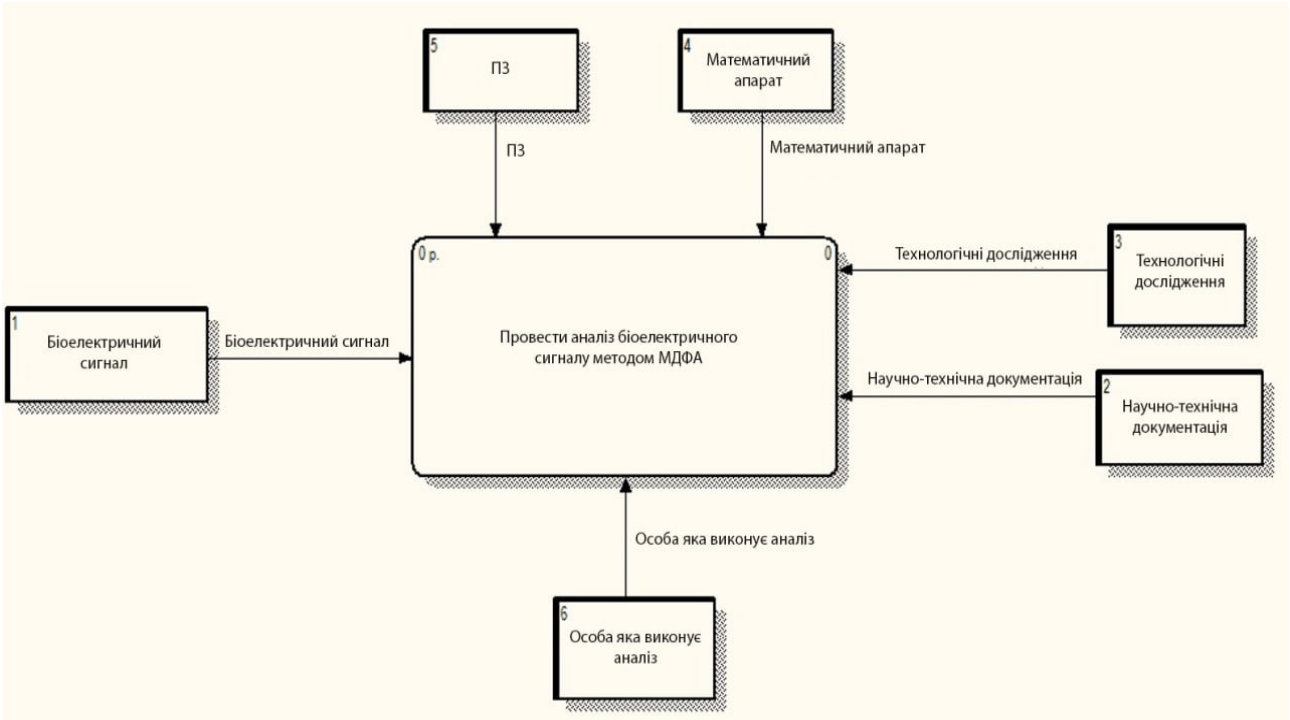


Рисунок 1.8 – Інформаційна модель DFD, перший рівень декомпозиції

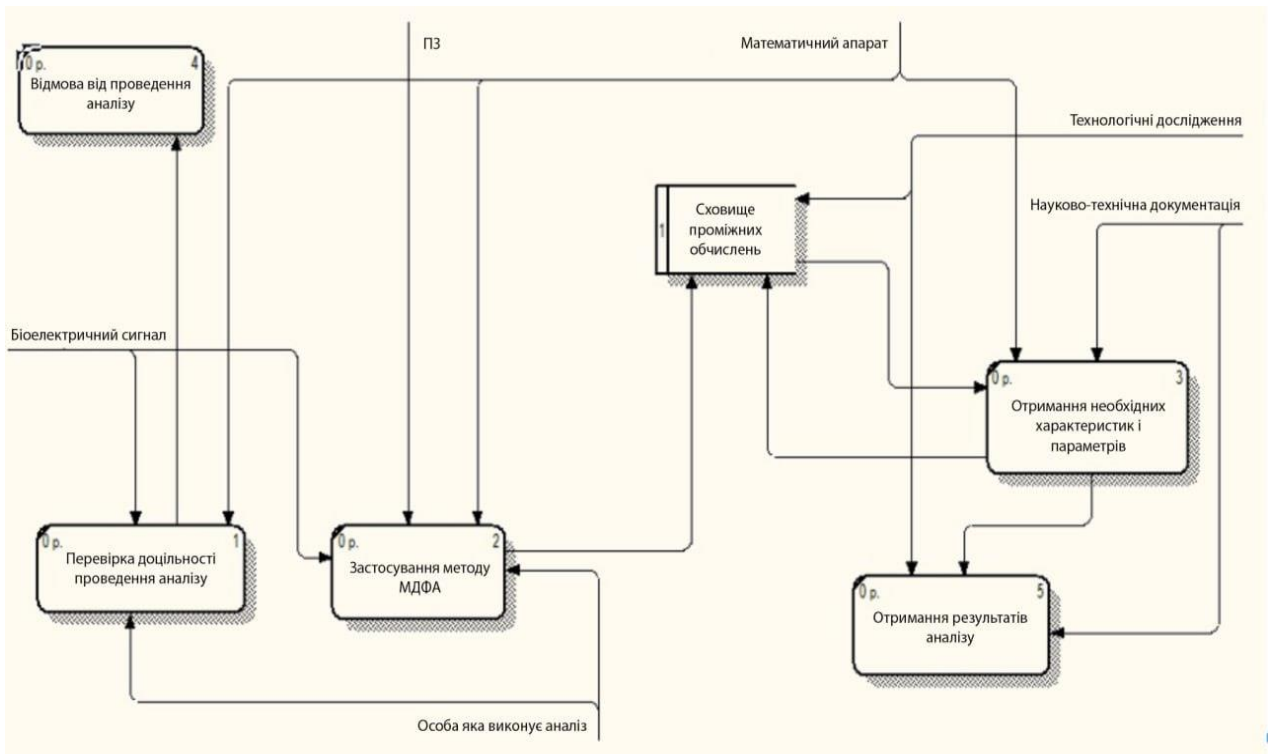


Рисунок 1.9 – Інформаційна модель DFD, другий рівень декомпозиції

1.2 Змістовна та формальна постановка задачі

Поняття фрактал було запроваджено Бенуа Мандельбротом у сімдесяті роки. Термін походить від латинського *fractus*, прикметника від дієслова *frangere* ламати, розбивати на частини. Тобто, як говорить одне з визначень фракталів, фрактали – це структури, які, незважаючи на свою крайню нерегулярність на різних масштабах, виглядають приблизно однаково (зазначимо, що це далеко не єдине визначення).

В даний час визнано, що багато інформаційних, біологічних, фізичних і технологічних процесів мають складну фрактальну структуру. Фрактальний аналіз використовується для моделювання, аналізу та контролю складних систем у різних галузях науки та техніки. Фрактальний аналіз застосовується для прогнозування сейсмічної активності та цунамі, а також допомагає визначити вік геологічних порід у геології, передбачити кризу та ризик за

допомогою фінансових рядів в економіці, вивчити турбулентність та перебіг термодинамічних процесів у фізиці.

Фрактальна геометрія вже чверть століття використовується у біології. Застосування методів фрактального аналізу відкриває нові можливості для вивчення функціональної організації живих систем. Численні експериментальні та клінічні дані дають підстави для висновку про те, що вивчення фрактальної топології різних біологічних систем дозволить закласти основи фрактальної діагностики.

Суб'єкти, які мають фрактальні властивості можна розділити на дві групи: регулярні (монофрактали) і мультифрактальні. Монофрактальні предмети є однорідними навіть тому, що мають єдиний показник масштабування. Їхні фрактальні характеристики залишаються незмінними на будь-якому діапазоні масштабування. Мультифрактальні предмети можна розкласти на сегменти з різними локальними властивостями. Вони характеризуються спектром масштабування показників.

Сигнали, що створюються за допомогою складних саморегульованих біологічних систем мають широкий спектр властивостей, таких як неоднорідність, нелінійність, нестационарність, наявність коливань та ін. Було показано, що для багатьох систем біологічні сигнали мають довгострокові кореляції та фрактальні властивості. Зокрема, мультифрактальні властивості виявлено в багатьох біоелектричних сигналах.

Наведемо приклади можливого застосування фрактального аналізу для медичних та біологічних сигналів:

- а) електрокардіограми виявлення нерегулярних серцевих скорочень, турбулентність кровотоку;
- б) електроенцефалографія. Виявлення по ЕЕГ апоплексичного удару чи епілептичного нападу;
- в) виявлення різноманітних аномалій на ранніх стадіях розвитку плоду, ЕЕГ плоду, пов'язаних зі ступенем розвитку;
- г) підвищення якості зображення на маммограмах;

- д) стиснення медичних сигналів;
- е) томографічні дослідження;
- ж) вивчення мовного сигналу для побудови систем розпізнавання мови.

Інтервальні послідовності серцебиття були одними з перших фізіологічних часових рядів, у яких були виявлені властивості самоподібності [1]. У роботах [2] було досліджено фрактальні та кореляційні властивості часових інтервалів між послідовними ударами серця під час легкого, глибокого та швидкого сну. У роботі [3] було представлено дослідження, яке дозволить використовувати фрактальний аналіз, щоб автоматично класифікувати стадії сну, використовуючи тільки запис електрокардіограми. У роботі [4] були розглянуті статистичні дані, кореляції та фрактальні властивості серцевого ритму у здорових та хворих людей. У роботах [5] порівнювали фрактальні показники варіабельності ритму серця людей, які практикують йогу та початківців.

На сьогоднішній день відомо, що сигнали, пов'язані з серцевою діяльністю, характеризуються не тільки самоподібністю, а й мультифрактальними властивостями, що відображають гетерогенність та нестационарність фізіологічних процесів. Зокрема, це було предметом дослідження [6], де було показано, що частота серцевих скорочень має мультифрактальні властивості, які були різні у здорових людей і тих, хто страждає від різних захворювань. У роботах [7] було досліджено мультифрактальні властивості варіабельності серцевого ритму на основі вейвлет-перетворення.

Численні дослідження присвячені вивченню фрактальних властивостей часових рядів електроенцефалограми (ЕЕГ). У роботах [8] було запропоновано новий метод виявлення інсульту з урахуванням масштабних властивостей ЕЕГ людини. У роботах [9] було вивчено фрактальні характеристики сну ЕЕГ здорових випробуваних. У роботах [10] були досліджені сигнали ЕЕГ птахів, щоб охарактеризувати різні стадії дозрівання мозку птахів. У роботах [11] були проаналізовані фрактальні властивості ЕЕГ сну у чоловіків із гострою

депресією. У роботах [12] досліджували ЕЕГ пацієнтів із хворобою Альцгеймера. У роботах [13] була показана різниця в ЕЕГ між випробуваними із закритими та відкритими очима за допомогою фрактального аналізу. У роботах [3] було запропоновано метод дослідження крос-модальної пластичності сліпих на основі фрактального аналізу ЕЕГ. У роботі [10] показали відмінності в мультифрактальних характеристиках ЕЕГ для нейронної активності здорових і хворих на епілепсію щурів. У роботах [8] було досліджено мультифрактальні властивості багатоканальних записів ЕЕГ. У роботах [3] було проведено мультифрактальний аналіз на стадіях сну та неспання. У роботах [11] було показано, що множину параметрів, що характеризують мультифрактальний спектр як ЕЕГ, так і частоти серцевих скорочень може розрізняти здорові та патологічні стани.

Цей список показує як широко і по-різному представлені фізіологічні фрактальні сигнали. Очевидно, що фрактальні характеристики відбивають суттєві особливості стану організму. Дослідження показують, що мультифрактальні методи можуть бути успішно використані для аналізу фізіологічних сигналів та детерменованих функціональних змін у продуктивності організму. Найчастіше мультифрактальні характеристики розглядалися швидше якісно, ніж кількісно. Однак для застосування результатів фрактального аналізу як знання в медичних експертних системах ми повинні використовувати їх кількісні значення. Метою даної роботи є розробка апарату для практичного застосування мультифрактального аналізу, для можливого використання в якості кількісної оцінки характеристик, зокрема в системах підтримки прийняття рішень.

Самоподібність фрактальних об'єктів полягає у збереженні структури об'єкта при зміні масштабу. Розглянемо основні характеристики мультифрактальної множини. Нехай в загальному випадку мультифрактальний об'єкт займає деяку обмежену область в d -мірному Евклідовому просторі і визначає множина з $N \rightarrow \infty$ точок. Розіб'ємо всю область на кубічні комірки з ребром ε та об'ємом ε^d . Розглянемо узагальнену статистичну суму $Z(q, \varepsilon)$, що

характеризується показником ступеня q , який може набувати будь-яких значень в інтервалі $-\infty < q < +\infty$:

$$Z(q, \varepsilon) = \sum_{i=1}^{N(\varepsilon)} p_i^q(\varepsilon), \quad (1.1)$$

де $p_i(\varepsilon) = \lim_{N \rightarrow \infty} \frac{n_i(\varepsilon)}{N}$;

$n_i(\varepsilon)$ - кількість точок, що потрапила в комірку з номером i ;

$N(\varepsilon)$ – сумарна кількість зайнятих комірок, яка залежить від розміру комірки ε .

Імовірності p_i характеризують відносну заселеність комірок.

У загальному випадку мультифрактальна множина характеризується деякою нелінійною функцією $\tau(q)$, що визначає поведінку статистичної суми $Z(q, \varepsilon)$ при $\varepsilon \rightarrow 0$:

$$Z(q, \varepsilon) \propto \varepsilon^{\tau(q)}. \quad (1.2)$$

Функція $\tau(q)$ зазвичай називається скейлінговою експонентою і визначається як

$$\tau(q) = \lim_{\varepsilon \rightarrow 0} \frac{\ln Z(q, \varepsilon)}{\ln \varepsilon}. \quad (1.3)$$

У разі однорідної фрактальної множини з фрактальною розмірністю D у всіх зайнятих комірках міститься однакова кількість точок, тобто, $p_i(\varepsilon) = p(\varepsilon) = 1/N(\varepsilon)$ і узагальнена статистична сума набуває вигляду:

$$Z(q, \varepsilon) = N^{1-q}(\varepsilon) = \varepsilon^{-D(1-q)},$$

а функція $\tau(q) = (q-1)D$ є лінійною. Якщо розподіл точок по комірках однаковий, то фрактальна множина є неоднорідною, тобто. мультифрактальною, і $\tau(q)$ є нелінійною функцією. При $q \rightarrow +\infty$ основний внесок в узагальнену статистичну суму вносять комірки, що містять найбільшу кількість частинок n_i , отже, що характеризуються найбільшою ймовірністю заповнення p_i . Навпаки, при $q \rightarrow -\infty$ основний вклад у суму дають розріджені комірки з малими значеннями p_i . Таким чином, функція $\tau(q)$ показує, наскільки неоднорідною є множина точок, що досліджується.

Поряд із скейлінговою експонентою $\tau(q)$ для характеристики мультифрактальної множини використовується функція мультифрактального спектру (спектру сингулярностей) $f(\alpha)$. Залежність ймовірності від розміру комірки $p_i(\varepsilon)$ має статечний характер

$$p_i(\varepsilon) \propto \varepsilon^{\alpha_i}, \quad (1.4)$$

де α_i є деякий показник ступеня, взагалі кажучи різний, для різних комірок (показник сингулярності). Для однорідного фрактала всі показники ступеня α_i однакові та рівні фрактальній розмірності D .

Функція мультифрактального спектру $f(\alpha)$ характеризує розподіл ймовірностей різних значень α_i . Якщо величина $n(\alpha)d\alpha$ є ймовірністю того, що α_i перебуває у інтервалі $(\alpha, \alpha + d\alpha)$, тобто. являє собою число комірок i , що мають однакову міру $p_i(\varepsilon)$ з $\alpha_i \in (\alpha, \alpha + d\alpha)$, тоді

$$n(\alpha) \approx \varepsilon^{-f(\alpha)}. \quad (1.5)$$

Таким чином, функція $f(\alpha)$ являє собою фрактальну розмірність одного однорідного фрактального підмножини ξ_α з вихідної множини ξ , що характеризується однаковими ймовірностями заповнення комірок $p_i(\varepsilon) \approx \varepsilon^\alpha$.

Враховуючи вирази (1.1) та (1.5), узагальнену статистичну суму $Z(q, \varepsilon)$ можна виразити через функцію мультифрактального спектру $f(\alpha)$ таким чином:

$$Z(q, \varepsilon) = \sum_{i=1}^{N(\varepsilon)} p_i^q(\varepsilon) \approx \int d\alpha n(\alpha) \varepsilon^{q\alpha} \approx \int d\alpha \varepsilon^{q\alpha - f(\alpha)}.$$

Формально перехід від змінних $\{q, \tau(q)\}$ до змінних $\{\alpha, f(\alpha)\}$ може бути здійснений за допомогою наступних перетворень Лежандра:

$$\left\{ \begin{array}{l} \alpha = \frac{d\tau}{dq}, \\ f(\alpha) = q \frac{d\tau}{dq} - \tau, \end{array} \right. \text{ та } \left\{ \begin{array}{l} q = \frac{df}{d\alpha}, \\ \tau(q) = \alpha \frac{df}{d\alpha} - f. \end{array} \right. \quad (1.6)$$

Самоподібність випадкових процесів полягає у збереженні статистичних характеристик при зміні масштабу часу. Стохастичний процес $X(t)$ є самоподібним з параметром самоподібності H , якщо $a^{-H} X(at)$ описується тими ж кінцевомірними законами розподілів, що і $X(t)$:

$$\text{Law}\{X(t)\} = \text{Law}\{a^{-H} X(at)\}, \quad \forall a > 0, t > 0. \quad (1.7)$$

Параметр H , $0 < H < 1$, який називається показником Херста, є ступенем самоподібності. Поряд із цією властивістю, показник H характеризує міру довгострокової залежності стохастичного процесу.

При значеннях $0,5 < H < 1$ ряд демонструє персистентну (трендостійку) поведінку. Якщо персистентний ряд зростав (зменшувався) у попередній період, то чим ближче показник Херста до 1, тим більшою ймовірністю буде зберігатися тенденція поведінки цього ряду протягом такого ж періоду в

майбутньому. Значення $H = 0,5$ вказує на незалежність (відсутність будь-якої пам'яті про минуле) прирощень часового ряду. Діапазон $0 < H < 0,5$ відповідає антиперсистентним рядам: якщо антиперсистентний ряд характеризувався зростанням у попередньому періоді, то чим ближче показник Херста до 0, тим більшою ймовірністю в наступному періоді почнеться спад.

Легко показати, поклавши $a = 1/t$, що для самоподібного процесу виконується така рівність:

$$\text{Law}\{X(t)\} = \text{Law}\left\{\left(\frac{1}{t}\right)^{-H} X(1)\right\} = \text{Law}\{t^H X(1)\}. \quad (1.8)$$

З огляду на (1.8) початкові моменти самоподібного випадкового процесу можна виразити як

$$\text{M}\left[|X(t)|^q\right] = \text{M}\left[|t^H X(1)|^q\right] = t^{qH} \text{M}\left[|X(1)|^q\right] = C(q) \cdot t^{qH}, \quad (1.9)$$

де величина $C(q) = \text{M}\left[|X(1)|^q\right]$.

Мультифрактальні процеси виявляють гнучкіші скейлінгові закономірності для моментних характеристик. На відміну від самоподібних процесів, де всі моменти $\text{M}\left[|X(t)|^q\right]$ показують однакову масштабну поведінку, для мультифрактальних процесів виконується відношення

$$\text{M}\left[|X(t)|^q\right] = C(q) \cdot t^{qh(q)}, \quad (1.10)$$

де $h(q)$ – узагальнений показник Херста, що є нелінійною функцією, на яку значення $h(q = 2)$ збігається зі значенням ступеня самоподібності H .

Для часових рядів, які відповідають монофрактальному процесу, узагальнений показник Херста $h(q) = H$ не залежить від параметра q .

Узагальнений показник Херста пов'язаний з функцією $\tau(q)$ співвідношенням :

$$\tau(q) = qh(q) - 1. \quad (1.11)$$

Визначення функції мультифрактального спектра $f(\alpha)$ випадкового процесу провадиться за формулами (1.6).

1.3 Постановка завдань дослідження

Метою кваліфікаційної роботи є застосування нейронних мереж та методу мультифрактального флуктуаційного аналізу для дослідження та класифікації біоелектричних сигналів.

Проаналізувавши всі отримані у попередніх трьох підпунктах дані, сформуємо перелік завдань, які ставимо перед собою під час дослідження:

- провести огляд та аналіз сучасного стану задачі класифікації біоелектричних сигналів за допомогою машинного навчання та методів мультифрактального аналізу;
- вивчити основні поняття та властивості мультифрактальних випадкових процесів;
- вивчити метод мультифрактального флуктуаційного аналізу;
- провести огляд методів класифікації біоелектричних сигналів;
- вивчити основні поняття теорії машинного навчання та нейронних мереж;
- провести мультифрактальний аналіз нормальних та аномальних біоелектричних сигналів;
- провести аналіз та класифікацію реальних біоелектричних сигналів за допомогою машинного навчання.

2 ВИБІР ТА ОБҐРУНТУВАННЯ МЕТОДУ РОЗВ'ЯЗАННЯ

2.1 Метод детрендованого флуктуаційного аналізу

Метод детрендованого флуктуаційного аналізу (ДФА) заснований на скейлінговому співвідношенні флуктуаційної функції $F(\tau)$ випадкового процесу, що має довгострокову залежність, і довжини сегмента ряду τ :

$$F(\tau) = (c \cdot \tau)^H. \quad (2.1)$$

Побудуємо покроковий алгоритм знаходження H по часовому ряду.

Крок 1. Для розрахунку необхідно перетворити вихідний ряд $X(t)$, $t = 1, \dots, N$ шляхом знаходження першої різниці: $x(t) = X(t+1) - X(t)$. Якщо вихідний ряд вже є рядом прирощень, цей крок пропускається. Ряд $X(t)$ бажано привести до нульового середнього: $X^*(t) = X(t) - \bar{X}(t)$.

Крок 2. Будується кумулятивний ряд $Y^{cum}(k) = \sum_{t=1}^k X^*(t)$.

Крок 3. Ряд $Y^{cum}(t)$ розбивається на N сегментів, що не перекриваються, довжиною τ і для кожного сегмента $Y^{cum}(t)$ обчислюється флуктуаційна функція

$$F^2(\tau) = \frac{1}{\tau} \sum_{t=1}^{\tau} (Y^{cum}(t) - Y_m(t))^2, \quad (2.2)$$

де $Y_m(t)$ – локальний m -поліноміальний тренд у межах даного сегмента.

Крок 4. Функція $F(\tau)$ усереднюється по всьому ряду $Y^{cum}(t)$:

$$F(\tau) = \sqrt{\frac{1}{N} \sum_{i=1}^N F_i^2(\tau)}. \quad (2.3)$$

Крок 5. Значення τ збільшується та кроки 3–4 повторюються. Слід зазначити, що максимальна довжина сегмента має перевищувати значення $N/4$, інакше побудова міститиме великі похибки.

Крок 6. Будується лінійна регресія методом найменших квадратів на $\log(\tau)$ як незалежної змінної, і $\log F(\tau)$ як залежної змінної (рис. 2.1). Тангенс кута нахилу прямої є оцінкою показника Херста.

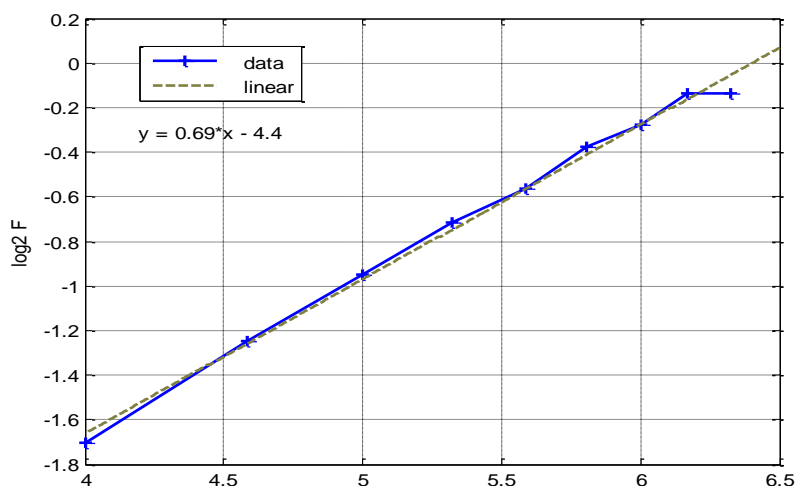


Рисунок 2.1 – Залежність $\log F(\tau)$ від $\log(\tau)$ отримана методом ДФА

2.2 Мультифрактальний флуктуаційний аналіз

Під час проведення мультифрактального флуктуаційного детрендованого аналізу (МФДФА) досліджується залежність флуктуаційної функції $F_q(s)$ від параметра q :

$$F_q(s) = \left\{ \frac{1}{N} \sum_{i=1}^N [F^2(s)]^{\frac{q}{2}} \right\}^{\frac{1}{q}}, \quad (2.4)$$

яка отримана зведенням виразу (2.3) у ступінь q та подальшим усередненням по всіх сегментах. Оскільки при $q = 0$ виразі (2.4) містить невизначеність, то замість нього використовують наступний граничний вираз:

$$F_q(\tau) = \text{Exp} \left\{ \frac{1}{N} \sum_{i=1}^N \ln[F^2(\tau)] \right\}.$$

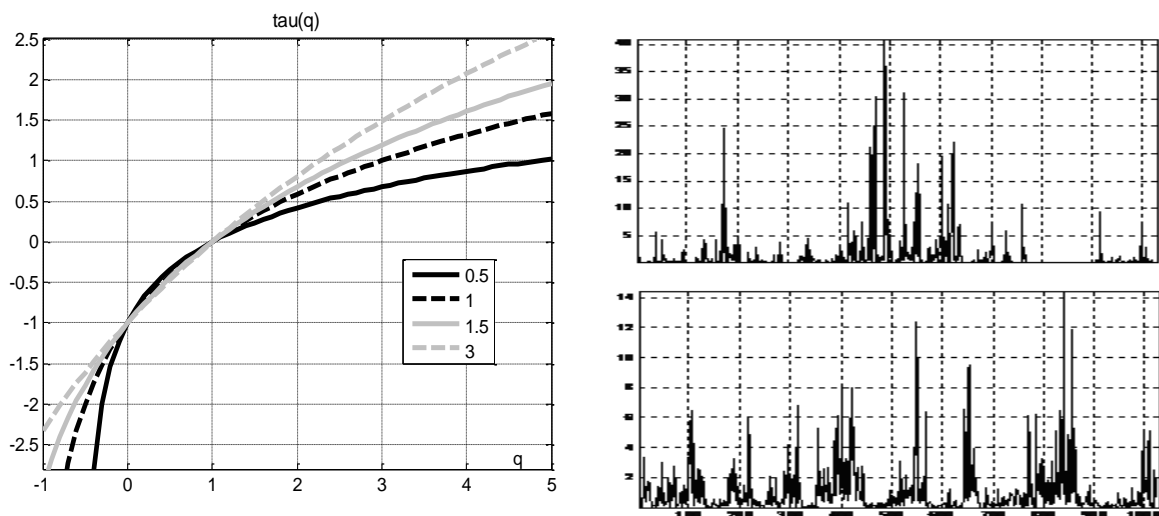


Рисунок 2.2 – Функція $\tau(q)$ для різних значень параметра a бета-розподілу; реалізації біномного каскаду для значень $a = 1$ і $a = 3$

Змінюючи часову шкалу s при фіксованому показнику q знаходимо залежність $F_q(s)$, представляючи її в подвійних логарифмічних координатах. Якщо досліджуваний ряд зводиться до мультифрактальної множини, що виявляє довгострокові залежності, то флуктуаційна функція $F_q(s)$ є ступеневою залежністю

$$F_q(s) \propto s^{h(q)}, \quad (2.5)$$

де $h(q)$ – функція узагальненого показника Херста.

З виразу (2.4) випливає, що $q = 2$ цей показник зводиться до звичайного значення параметра Херста H .

На рисунку 2.3 представлені флуктуаційні функції $F_q(s)$ для монофрактального (ліворуч) та мультифрактального (праворуч) процесів для значень параметра $q = \{-5; -2; 0; 2; 5\}$.

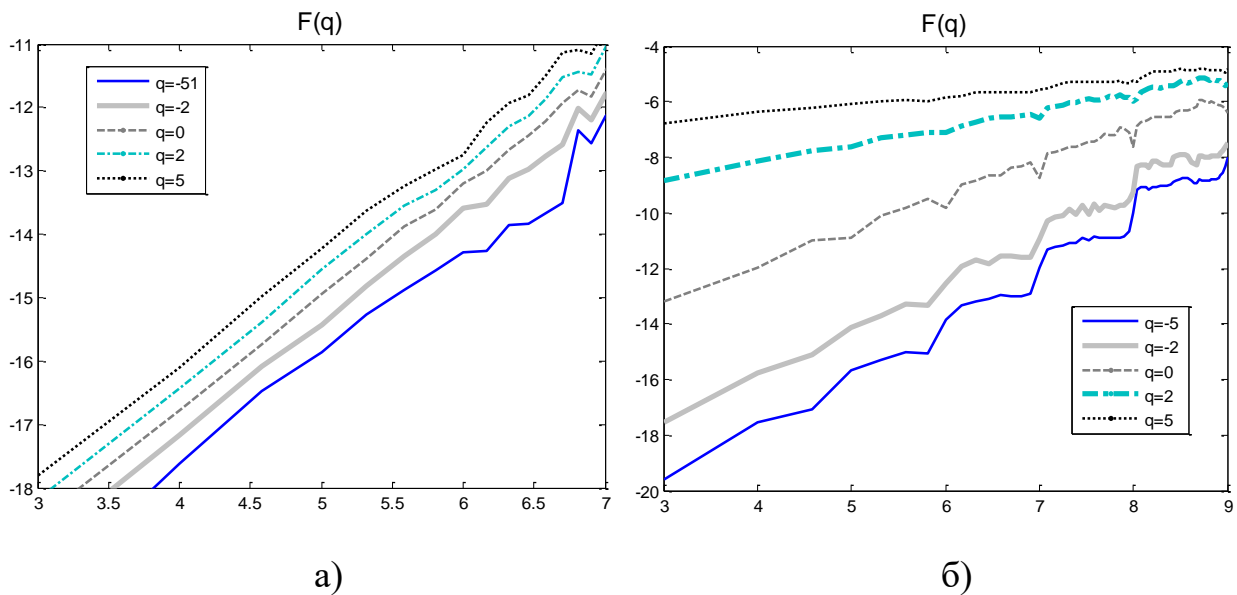


Рисунок 2.3 – $F_q(s)$ для а) монофрактального та б) мультифрактального процесів

На рис. 2.4 представлені функції узагальненого показника Херста $h(q)$, отримані за реалізацією детермінованого біномного каскаду при значеннях вагового коефіцієнта $p_1 = \{0,6; 0,7; 0,8; 0,9\}$. Відповідні значення функцій скейлінгової експоненти $\tau(q)$ показано рисунку 2.1.

Для часових рядів, які відповідають монофрактальній множині, флуктуаційна функція $F_q(s)$ однакова для всіх сегментів, і узагальнений показник Херста $h(q) = H$ не залежить від параметра q . Для мультифрактальних рядів $h(q)$ є нелінійною функцією: при позитивних q основний внесок у функцію $F_q(s)$ дають сегменти, що виявляють великі

відхилення $F^2(s)$, а при негативних q домінують сегменти з малими дисперсіями $F^2(s)$. Таким чином, при негативних значеннях q узагальнений показник Херста $h(q)$ описує сегменти, що виявляють малі флуктуації, а за позитивних – великі.

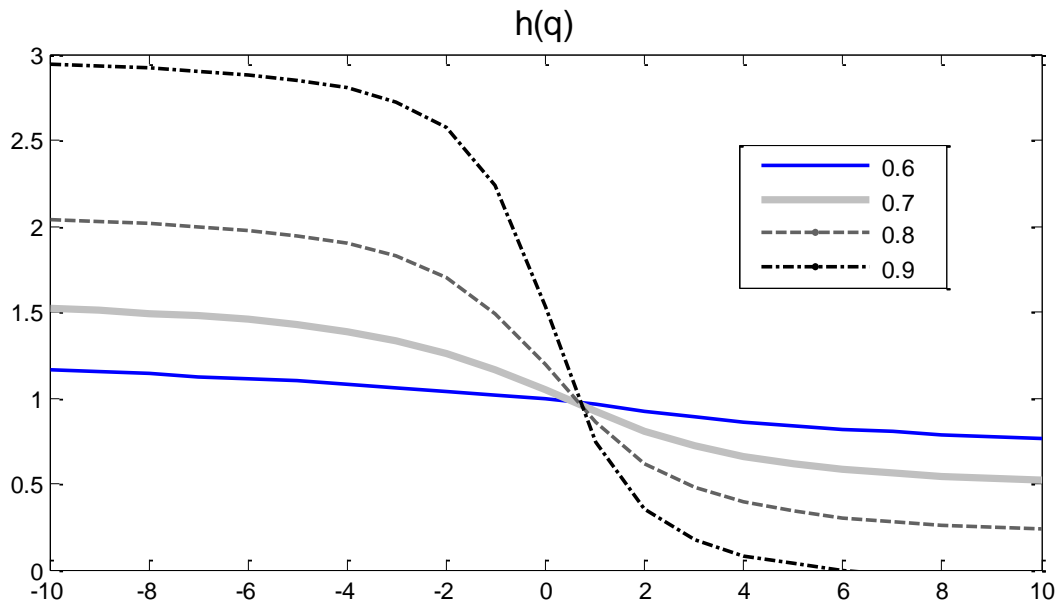


Рисунок 2.4 – Функції $h(q)$ для реалізації детермінованого каскаду при різних значеннях p_1

Виходячи із співвідношень (2.1–2.3) та (2.4,2.5) можна запропонувати наступний покроковий алгоритм оцінювання мультифрактальних характеристик по часовому ряду $X(t)$, $t = 1, 2, \dots, n$.

Крок 1. Для розрахунку необхідно перетворити вихідний ряд на кумулятивний $X^{cum}(k) = \sum_{t=1}^k X(t)$, $k = 1, 2, \dots, n$. Якщо вихідний ряд є кумулятивним, цей крок пропускається.

Крок 2. Встановлюються значення параметра q , у діапазоні яких потрібно знайти значення мультифрактальних характеристик $h(q)$, $\tau(q)$, $f(\alpha)$.

Крок 3. Ряд $X^{cum}(t)$ розбивається на N сегментів, що не перекриваються ,

довжиною τ і для кожного сегмента $X^{cum}(t)$ обчислюється флуктуаційна функція

$$F^2(\tau) = \frac{1}{\tau} \sum_{t=1}^{\tau} (X^{cum}(t) - Y_m(t))^2, \quad (2.6)$$

де $Y_m(t)$ – локальний m -поліноміальний тренд у межах даного сегмента.

Крок 4. Функція $F(\tau)$ усереднюється по всьому ряду $X^{cum}(t)$

$$F(\tau) = \sqrt{\frac{1}{N} \sum_{i=1}^N F_i^2(\tau)}.$$

Крок 5. Обчислюється флуктуаційна функція значення q :

$$F_q(\tau) = \left\{ \frac{1}{N} \sum_{i=1}^N [F_i^2(\tau)]^{\frac{q}{2}} \right\}^{\frac{1}{q}}. \quad (2.7)$$

Крок 6. Значення τ збільшується та кроки 3–5 повторюються.

Крок 7. Будується лінійна регресія методом найменших квадратів на $\log(\tau)$, як незалежної змінної, і $\log F_q(\tau)$, як залежної змінної Тангенс кута нахилу прямої, що апроксимує отримані точки, визначає значення величини $h(q)$ для заданого значення q .

Крок 8. Виконавши кроки 2-6 всім заданих значень параметра q , отримуємо залежність $h(q)$.

Крок 9. Визначаються значення скейлінгової експоненти $\tau(q)$, а функція мультифрактального спектра $f(\alpha)$ перебувають за допомогою перетворення Лежандра.

2.3 Методи класифікації біоелектричних сигналів

Біоелектричні сигнали представляють собою взаємодію біологічних систем, таких як серцебиття, мозкова активність або нервова передача, що вимірюється за допомогою електричних сигналів. Існує кілька методів класифікації біоелектричних сигналів, у тому числі:

- методи на основі статистичних характеристик такі як середнє значення, стандартне відхилення, коефіцієнт варіації тощо;
- методи на основі частотних характеристик таких як спектральна площа, спектральна енергія тощо;
- методи на основі часових характеристик, як максимальне значення, максимальний спад тощо;
- машинне навчання, нейронні мережі, опорні вектори машини (SVM) або алгоритми випадкового лісу;
- глибоке навчання: Глибокі нейронні мережі, такі як згорткові нейронні мережі (CNN) або рекурентні нейронні мережі (RNN);
- технології на основі штучних нейронних мереж.

Вибір методу залежить від конкретного завдання, доступних даних та обмежень, таких як обчислювальна потужність або час обробки. Важливо розуміти, що часто комбінація різних підходів може привести до найкращих результатів.

2.4 Глибоке навчання як інструмент класифікації

Глибоке навчання є галуззю машинного навчання, яка сконцентрована на побудові алгоритмів, які можуть вивчати та репрезентувати дані за допомогою складних архітектур. Ці методи використовують особливості даних на різних рівнях абстракції, що дозволяє їм моделювати високорівневі аспекти даних. Є

декілька типів моделей глибокого навчання, які використовуються для класифікації, включаючи:

- глибокі штучні нейронні мережі (DNN): основний тип мереж глибокого навчання, який складається з трьох основних типів шарів: вхідного, вихідного та прихованого. Кожний шар містить велику кількість паралельних шарів, які виконують різні функції;

- згорткові нейронні мережі (CNN): Ці мережі використовуються переважно для обробки зображень. Вони використовують згорткові шари для вивчення локальних особливостей зображення, а потім редуктивні (pooling) шари для зменшення вимірності даних;

- рекурентні нейронні мережі (RNN): Вони використовуються для обробки послідовних даних, таких як тексти, часові ряди або послідовності амінокислот у білках. RNN дозволяють пов'язувати інформацію з пройдених у часі етапів;

- трансформери: Це новіший тип моделі глибокого навчання, який замінив RNN в багатьох задачах, особливо в області обробки природної мови. Трансформери використовують механізм уваги, що дозволяє моделі зосереджуватись на певних частинах вхідних даних;

Кожна з цих моделей має власні сильні й слабкі сторони та може бути краще підходити для певних типів завдань. При виборі моделі глибокого навчання для класифікації важливо врахувати не лише її потенціал для побудови складних репрезентацій, але й вимоги до обчислювальних ресурсів та наявність великого набору даних для тренування.

2.5 Згорткові нейронні мережі в задачах класифікації часових рядів

Згорткові нейронні мережі (Convolutional Neural Networks, CNN) не тільки ефективно порятує з обробкою зображень, але також можуть бути успішно використані для класифікації часових рядів.

На рис. 2.5. можна побачити типову архітектуру згорткової нейронної мережі. Розглянемо етапи її роботи у контексті часових рядів.

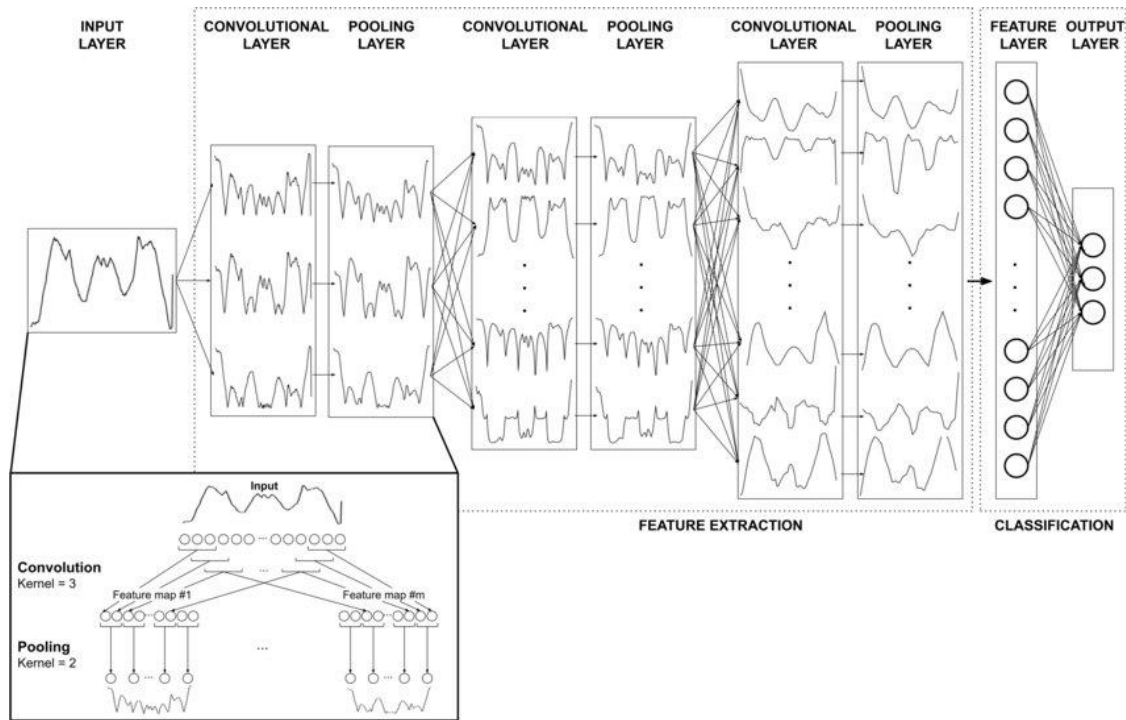


Рисунок 2.5 – Типова архітектура згорткової нейронної мережі класифікатора часових рядів

Згортка: Ядро (або фільтр) "проходить" по вхідному часовому ряду, обчислюючи добуток відповідних точок часу і значень фільтра, після чого сумує результати. Це створює новий часовий ряд, відомий як вихідна карта.

Функція активації: Після вихідної карти застосовується функція активації. Найчастіше в CNN використовується ReLU (Rectified Linear Unit) для додавання нелінійності в мережу.

Редукція (Pooling): Цей етап зменшує часовий ряд з кожної вихідної карти, зберігаючи найбільш інформативну складову. Це зазвичай робиться для запобігання перенавчанню і зменшення обчислювальних витрат.

Повністю зв'язаний шар: Після декількох циклів згортки та пулінгу, дані передаються в повністю зв'язані шари, схожі на ті, що використовуються в звичайних нейронних мережах.

Класифікація: Завершується обробка, коли останній повністю зв'язаний шар використовує функцію активації, наприклад softmax, для виведення прогнозованих класів часового ряду.

У результаті, згорткові нейронні мережі можуть виявляти як низькорівневі, так і високорівневі особливості в часових рядах, створюючи потужну модель для задач класифікації.

Висновки за розділом 2

Було розглянуто метод детрендованого флуктуаційного аналізу та мультифрактальний флуктуаційний аналіз як інструментів які добре підходять для аналізу різних типів часових рядів, таких як фінансові дані, кліматичні зміни, біоелектричні сигнали тощо. Ці методи допомагають виявити складні структури та властивості сигналів, які можуть залишатися непоміченими за допомогою традиційних методів аналізу.

Сучасні технології та досягнення в галузі нейронних мереж та машинного навчання виявляють великий потенціал у сфері аналізу та класифікації біоелектричних сигналів.

Зокрема, згорткові нейронні мережі демонструють високу ефективність у роботі з біоелектричними сигналами, такими як електроенцефалограма (ЕЕГ), електрокардіограма (ЕКГ) та електроміографія (ЕМГ). Їх здатність розпізнавати складні шаблони та залежності дозволяють досягати високої точності при класифікації, що в перспективі може автоматизувати процес діагностики та моніторингу пацієнтів.

3 ПРОГРАМНА РЕАЛІЗАЦІЯ

3.1 Python як високорівнева мова програмування

Python – це високорівнева мова програмування, яка популярна своєю зручною синтаксичною структурою, масштабованістю, відносною легкістю в освоєнні. Він підтримує декілька парадигм програмування, включаючи процедурне, об'єктно-орієнтоване та функціональне програмування.

Сфери застосування Python дуже широкі і містять наукові дослідження, веб-розробку, обробку даних, штучний інтелект, машинне навчання, автоматизацію процесів, вбудовані системи та багато іншого. Завдяки великій кількості бібліотек для всеосяжних задач, Python є універсальною мовою програмування.

Python відомий своєю простим, проте потужним синтаксисом, який особливо корисний для новачків. Python, не в останню чергу, зосереджується на простоті та читабельності коду. Python також допомагає програмістам організувати свій код через функції, класи, модулі та пакети.

Що стосується математичного моделювання та математики в цілому, Python володіє дуже широким спектром можливостей. Python має потужні бібліотеки, такі як NumPy, SciPy, Pandas та Matplotlib, які дозволяють виконувати різноманітні математичні операції. Більше того, Python використовується для розв'язання диференціальних рівнянь, статистичного моделювання, машинного навчання, оптимізації, аналізу складних даних та зручного графічного представлення результатів.

Нарешті, Python стає все більш центральною мовою програмування у сфері науки про дані (Data Science), машинного навчання та штучного інтелекту. Бібліотеки, такі як SciKit-Learn, TensorFlow, Keras, PyTorch надають інструменти для моделювання машинного навчання, включаючи класифікацію, регресію, кластеризацію, Deep Learning та інше.

Всі ці можливості зробили Python незамінним інструментом для науковців, інженерів, аналітиків даних, вчених про штучний інтелект і багатьох

інших. Python об'єднує науковців з усього світу в глобальну спільноту, яка активно підтримує відкриті проекти, організовує конференції та семінари, а також сприяє розвитку навчальних матеріалів і курсів, що показують, як ефективно використовувати Python для наукових цілей.

3.2 Бібліотеки TensorFlow / Keras та Scikit-learn

TensorFlow, розроблена Google, є однією з найвідоміших бібліотек для створення моделей машинного навчання та глибокого навчання. Вона здатна працювати з великими датасетами і розподіляти операції між багатьма GPU або машинами. Це також дозволяє розробникам створювати свої власні моделі та експериментувати з їх архітектурами. TensorFlow надає підтримку різноманітних моделей глибокого навчання, таких як нейронні мережі, рекурентні нейронні мережі (RNN), нейронні мережі згортки (CNN) та автокодувачі.

Keras, з іншого боку, є високорівневим інтерфейсом для TensorFlow та інших бібліотек глибокого навчання, який робить роботу з ними простішою та більш інтуїтивно зрозумілою. Він спрощує процес визначення та тренування моделей, при цьому пропонує простий у використанні API для створення і тренування моделей глибокого навчання. Keras також надає утиліти для візуалізації, опису, трансформації даних, що допомагає в їх попередній обробці, а також має підтримку аугментації.

Scikit-learn – це потужна бібліотека машинного навчання для мови програмування Python, яка дозволяє легко використовувати різноманітні алгоритми для аналізу та обробки даних. Основні особливості включають в себе широкий набір алгоритмів для класифікації, регресії, кластеризації та інших завдань машинного навчання, а також зручний інтерфейс для їх використання. Завдяки простоті використання та широким можливостям, scikit-learn стає ідеальним інструментом як для початківців в машинному навчанні, так і для досвідчених фахівців.

3.3 Бібліотека NeuroKit2

NeuroKit2 - це бібліотека Python розроблена з метою надати комплексні, прозорі та легкі у використанні інструменти аналізу нейрофізіологічних сигналів.

Ось деякі основні функції, які надає NeuroKit2:

- попередня обробка та аналіз сигналів електрокардіограми (ЕКГ);
- обробка та аналіз сигналів електроенцефалограми (ЕЕГ);
- функції психофізіологічного моделювання.

NeuroKit2 досить легко встановлюється і використовується, а її розробники надають багато документації та прикладів для початківців. Один з цілей цього проекту - зробити аналіз біосигналів більш доступним і зрозумілим, а не заплутаним і технічно складним.

NeuroKit2 має в своєму арсеналі потужний інструментарій для обчислення фрактальних та мультифрактальних характеристик біоелектричних сигналів.

Зокрема, метод `fractal_dfa` в комбінації з параметром `multifractal=True` може бути використаний для розрахунку чисельних мультифрактальних характеристик біоелектричних сигналів та їх візуалізації.

Результатом мультифрактального детрендованого флуктуаційного аналізу є узагальнений показник Херста h для різних значень q . Він перетворюється в мультифрактальний скейлінговий показник Тау $\tau_{(q)}$, чия нелінійна залежність від q може бути індикатором мультифрактальності. Звідси ми можемо отримати показник сингулярності H (або α) (також відомий як показник Гольдера) і розмірність сингулярності D (або $f(\alpha)$). Варіація цих показників також відома як спектр мультифрактальної сингулярності (MSP) та зазвичай має вигляд оберненої параболи. Їм вимірюють властивість довгострокової кореляції сигналу. З цих елементів витягуються наступні характеристики:

Ширина (Width). Ширина спектру сингулярності, яка кількісно визначає ступінь мультифрактальності. У випадку монофрактальних сигналів ширина MSP дорівнює нулю, оскільки $h(q)$ не залежить від q .

Пік (Peak). Значення показника сингулярності H , що відповідає вершині сингулярної розмірності D . Вимірює ступінь самоподібності сигналу, а велике значення є показником високого ступеня кореляції між точками даних. Іншими словами, процес є повторюваним і рекурентним.

Середнє (Mean). Середнє максимального і мінімального значень сингулярного показника H , яке кількісно визначає середні коливання сигналу.

Максимум (Max). Значення сингулярного спектру D , що відповідає максимальному значенню сингулярного показника H , що вказує на максимальні коливання сигналу.

Дельта (Delta). Вертикальна відстань між сингулярним спектром D , де сингулярні показники мають мінімальне і максимальне значення. Відповідає діапазону коливань сигналу.

Асиметрія (Assymetry). Асиметричний відносіник (AR) відповідає центральності вершини спектра. $AR = 0.5$ вказує на те, що мультифрактальний спектр симетричний.

Коливання (Fluctuation). Індекс коливань h (hFI) визначається як потужність другої похідної $h(q)$.

Приріст (Increment). Кумулятивна функція квадратних приростів (αCF) узагальнених показників Херста між послідовними моментами є більш стійким показником розподілу узагальнених показників Херста.

3.4 Алгоритм розв'язання задачі класифікації біоелектричних сигналів

Для розв'язання задачі класифікації біоелектричних сигналів, було виконано експеримент на основі застосування згорткової нейронної мережі. Алгоритм виконання експерименту:

- завантаження датасету у середовище для обробки у будь-якому зручному форматі який підлягає серіалізації;
- первинний аналіз даних з метою зрозуміти їх структуру;

- візуалізація окремих зразків ЕКГ для кращого розуміння їх відмінностей;
- проведення мультифрактального аналізу реалізацій ЕКГ, порівняння значень характеристик для різних класів;
- обробка даних: нормалізація, видалення шуму, фільтрація, аргументація та інші трансформації необхідні для подальшого опрацювання;
- дані розділяються на незалежні змінні (X), які складаються з біоелектричних сигналів та цільову змінну (Y), яка складається з цільових класів;
- дані діляться на тренувальний та тестовий набори для подальшого використання;
- побудова моделі згорткової нейронної мережі враховуючи структуру даних, та очікуваний результат;
- тренування моделі “з учителем” на тренувальному наборі даних;
- оцінка ефективності моделі на тренувальних даних на основі таких характеристик як точність, втрати та інших;
- застосування моделі до тестового набору даних;
- оцінка ефективності роботи моделі на тестових даних на основі матриці похибки.

3.5 Опис програми

Застосунок, розроблений за допомогою мови програмування Python в середовищі Colaboratory, що працює на базі Google Cloud. Програма виконує класифікацію сигналів ЕКГ сигналів за заздалегідь визначеними класами, які відповідають здоровому, або одному з кількох аномальних станів.

Мета програми – аналіз та класифікація вхідних біоелектричних сигналів, а саме електрокардіограм (ЕКГ), за допомогою мультифрактального аналізу та нейронної мережі відпоідно до одного з заздалегідь визначених класів на яких натренована нейронна мережа.

Для тренування класифікатора використовується великий набір даних який містить сигнали ЕКГ заздалегідь промарковані фахівцями.

На першому етапі роботи ми проводимо завантаження та аналіз сигналів, а саме візуалізацію нормальних і аномальних сигналів. За допомогою графіків ми можемо побачити візуальні відмінності нормальних сигналів від аномальних.

Також ми можемо порівняти результати мультифрактального аналізу сигналів різних класів, що може дати додаткову інформацію.

Наступним кроком є підготовка даних до використання в глибокому навчанні нейронної мережі.

Формуємо виборки X (features) та Y (targets). Одним з найпоширеніших типів машинного навчання є "навчання з учителем" (supervised learning), де для усіх прикладів в навчальних даних відома відповідь (мітка чи ціль Y). Задача цього типу навчання - навчити модель правильно прогнозувати ці відповіді з нових даних, для яких відповідь невідома, використовуючи ознаки X .

Після етапу навчання моделі на тренувальних даних, ми маємо перевірити її точність на свіжому тренувальному датасеті.

Завершальним етапом іде оцінка точності моделі за допомогою метрик, таких як точність (accuracy), втрати (loss) та матриці помилок (confusion matrix).

Висновки за розділом 3

Для роботи було обрано мову програмування Python як інструмент який має достатній функціонал і не є складним в застосуванні. Також був проведений огляд наявних рішень для проведення мультифрактального аналізу та побудування нейронної мережі, такі як NeuroKit2 та TensorFlow / Keras.

Було розроблено ефективний та вичерпний алгоритм розв'язання задачі класифікації біоелектричних сигналів та сформований опис роботи програми.

4 РЕЗУЛЬТАТИ ОБЧИСЛЮВАЛЬНОГО ЕКСПЕРИМЕНТУ ТА ЇХ АНАЛІЗ

4.1 Вхідні дані

Мета даного експерименту полягає в розробці надійної моделі машинного навчання, яка може класифікувати біоелектричні сигнали серця, отримані за допомогою ЕКГ, у звичайні або аномальні. Це допоможе вчасно виявляти потенційні проблеми з серцем та розпочинати лікування, тим самим покращуючи здоров'я та благополуччя пацієнтів.

Для обчислювального експерименту був обраний датасет «ECG Heartbeat Categorization Dataset». Цей набір даних складається з двох колекцій сигналів серцебиття, отриманих з двох відомих наборів даних класифікації серцебиття - «MIT-BIH Arrhythmia Dataset» та «The PTB Diagnostic ECG Database». Кількість зразків в обох колекціях достатньо велика для тренування глибокої нейронної мережі.

Цей набір даних використовувався для дослідження класифікації серцебиття з використанням архітектур глибоких нейронних мереж та оцінювання деяких можливостей transfer learning на його основі. Сигнали відповідають формам електрокардіограм (ЕКГ) серцебиття у звичайному випадку та випадках, які характеризуються різними аритміями та інфарктом міокарда. Сигнали передпроцесуються та сегментуються, при цьому кожен сегмент відповідає одному удару серця. Структура даних наведена на рис. 4.1.

	0	1	2	3	4	5	6	7	8	9	...	178
0	0.977941	0.926471	0.681373	0.245098	0.154412	0.191176	0.151961	0.085784	0.058824	0.049020	...	0.0
1	0.960114	0.863248	0.461538	0.196581	0.094017	0.125356	0.099715	0.088319	0.074074	0.082621	...	0.0
2	1.000000	0.659459	0.186486	0.070270	0.070270	0.059459	0.056757	0.043243	0.054054	0.045946	...	0.0
3	0.925414	0.665746	0.541436	0.276243	0.196133	0.077348	0.071823	0.060773	0.066298	0.058011	...	0.0

Рисунок 4.1 – Представлення даних тестового датасету

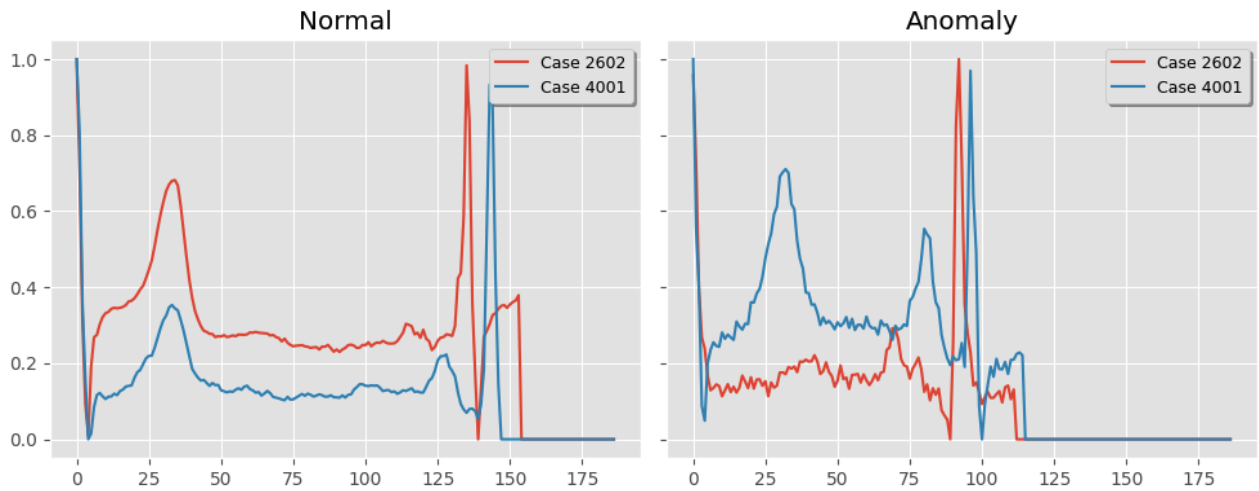


Рисунок 4.2 – Візуалізація сигналів з тестового датасету

Розглянувши графічну репрезентацію сигналів, що наведені на рис. 4.2, ми можемо зробити, що у осіб з нормальними показниками ЕКГ спостерігаються відповідність кривій Гауса (дзвоноподібна). ЕКГ аномальних осіб показує інші типи кривих. Ми використаємо цю інформацію для створення нашої моделі машинного навчання для класифікації.

Згідно з набором даних, нам дано 5 різних класів серцебиття: [N:0, S:1, V:2, F:3, Q:4]:

- N: Нормальні серцеві удари (Non-Ectopic Beats);
- S: Суправентрикулярні ектопічні серцеві удари (Supraventricular Ectopic Beats);
- V: Вентрикулярні ектопічні серцеві удари (Ventricular Ectopic Beats);
- F: Фузійні серцеві удари (Fusion Beats);
- Q: Невідомі серцеві удари (Unknown Beats).

Розглянемо розподіл необроблених даних за допомогою кругової діаграми на рис. 4.3. Можемо зробити висновок, що наші дані упереджені і нам потрібно використати аргументацію даних на них, щоб ми могли видалити упередженість від даних та досягти їх рівномірного розподілу. Для цього ми окремо виділяємо набір даних для кожного класу і застосовуємо метод `resample` бібліотеки `sklearn.utils`.

Метод `resample` часто використовується в задачах машинного навчання для створення вибірок `bootstrap`, збільшення меншинства (`oversampling`) або зменшення більшості (`undersampling`) класів у несбалансованих датасетах.

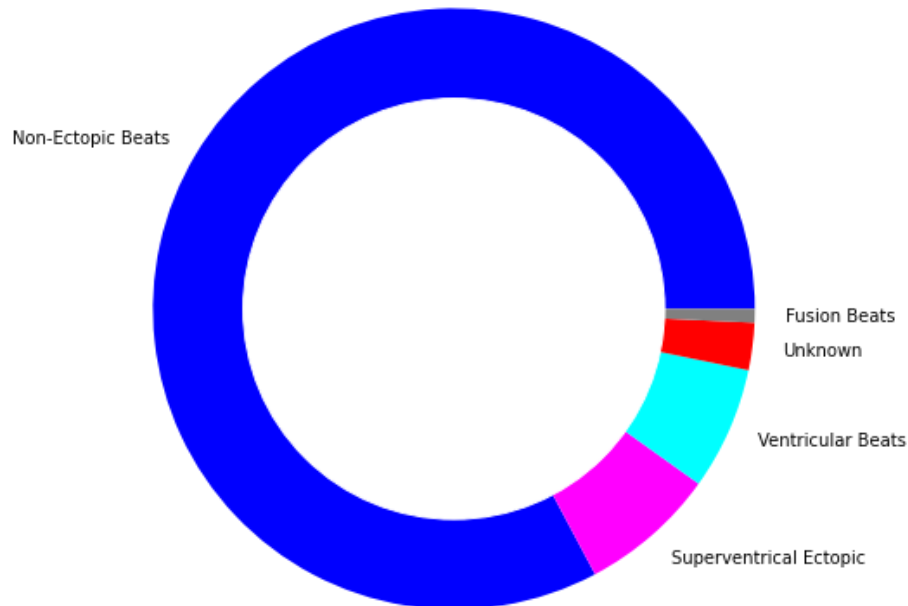


Рисунок 4.3 – Розподілення даних в датасеті в сирому вигляді

```
from sklearn.utils import resample
data_1_resample = resample(data_1, n_samples=20000,
                           random_state=123, replace=True)
data_2_resample = resample(data_2, n_samples=20000,
                           random_state=123, replace=True)
data_3_resample = resample(data_3, n_samples=20000,
                           random_state=123, replace=True)
data_4_resample = resample(data_4, n_samples=20000,
                           random_state=123, replace=True)
data_0 = data_train[data_train[187] == 0].sample(n=20000, random_state=123)

train_dataset = pd.concat([data_0, data_1_resample, data_2_resample, data_3_resample,
                           data_4_resample])
```

Рисунок 4.4 – Використання методу `resample` для нормалізації даних

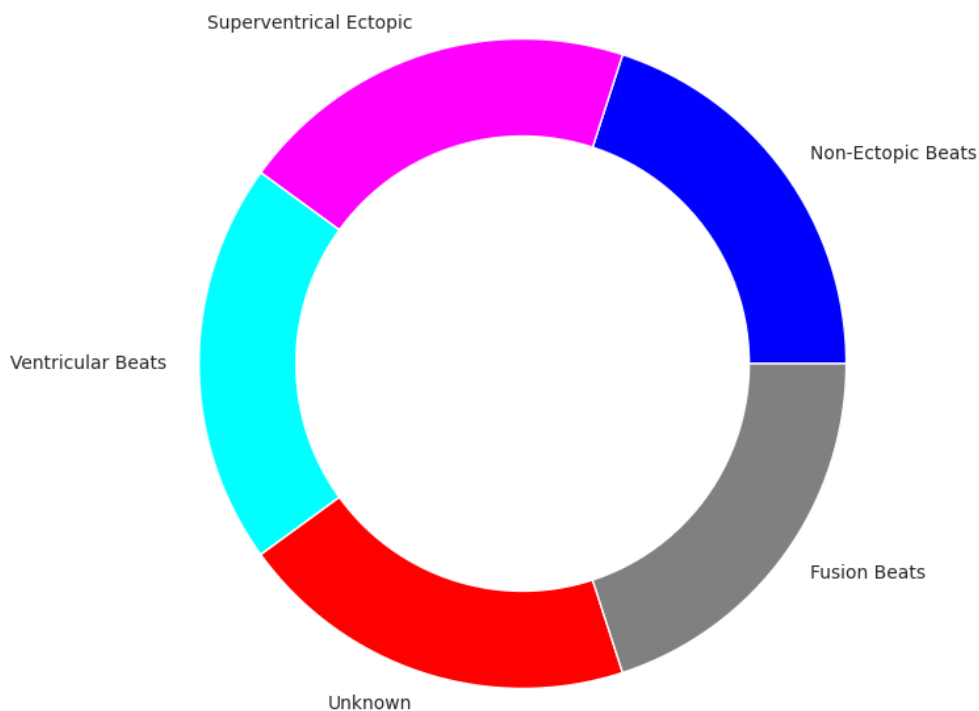


Рисунок 4.5 – Розподілення даних в датасеті після обробки

В результаті застосування методу `resample` на нашому датасеті отримуємо збалансований набір даних наведений на рис. 4.5.

4.2 Розрахунок мультифрактальних характеристик

За допомогою бібліотеки `NeuroKit2` отримаємо мультифрактальні характеристики сигналів різних класів та порівняємо.

Отримані мультифрактальні характеристики наведені в таб. 4.1, а їх графічні репрезентації на рис. 4.6 та 4.7. Можна зробити висновок, що мультифрактальні характеристики виявляють значні розбіжності між різними класами сигналів ЕКГ, це може свідчити про відмінності в структурі та динаміці серцевої активності, що мають відображення в цих сигналах.

Сигнали від пацієнта зі здоровим серцем мають інші характеристики, ніж сигнали від серця з патологіями. Тому велика розбіжність в мультифрактальних характеристиках може бути індикатором наявності патологій.

Таблиця 4.1 – Мультифрактальні характеристики сигналів відповідних різним класам

	N	S	V	F	Q
Width	3.291163	1.255482	0.344544	3.088284	1.147023
Peak	1.757925	1.419068	1.190132	2.334796	1.388251
Mean	1.65308	0.901293	1.362404	2.435106	0.945838
Max	-0.330785	1.0	0.993734	-0.486283	1.0
Delta	0.213775	1.332629	-0.438565	-0.415879	1.007428
Asymmetry	-0.531857	-0.912411	0.0	-0.467519	-0.885706
Fluctuation	0.024696	0.024933	0.000552	0.014264	0.032905
Increment	1.089485	0.377909	0.015145	0.945007	0.407141

Multifractal Detrended Fluctuation Analysis (MFDFA)

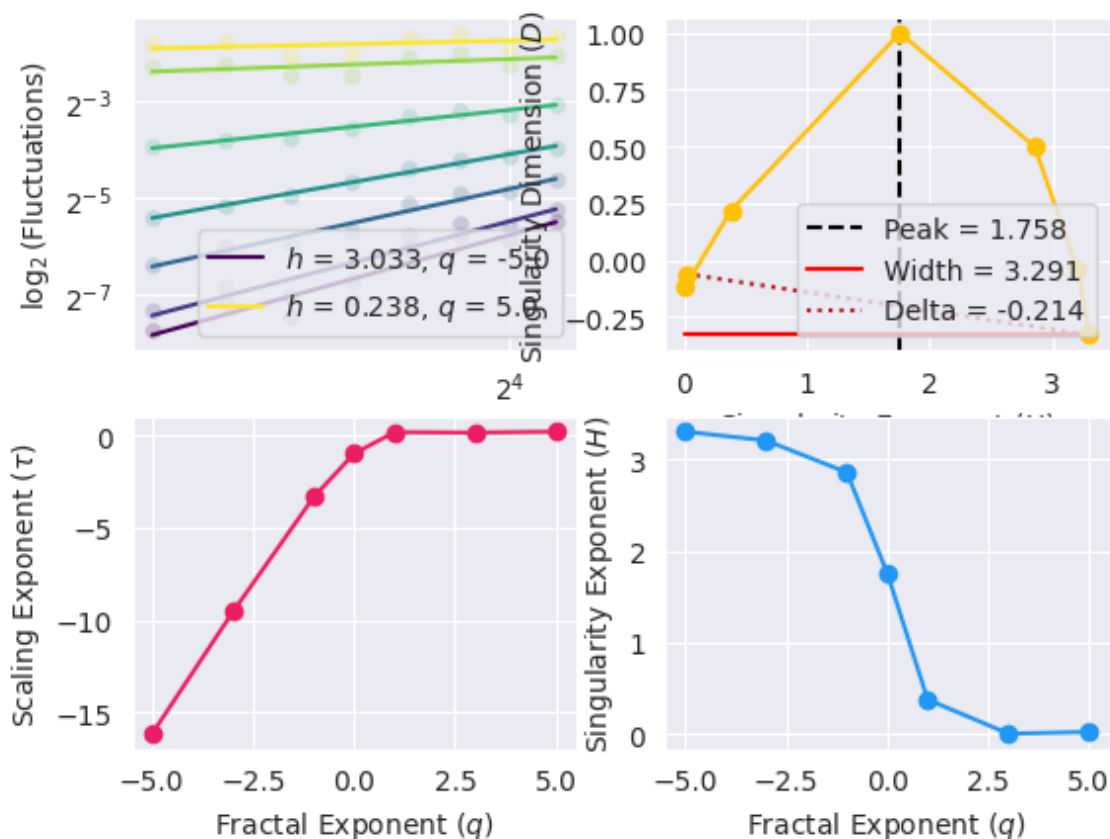


Рисунок 4.6 – Графічна репрезентація мультифрактальних характеристик нормального сигналу

Multifractal Detrended Fluctuation Analysis (MFDFA)

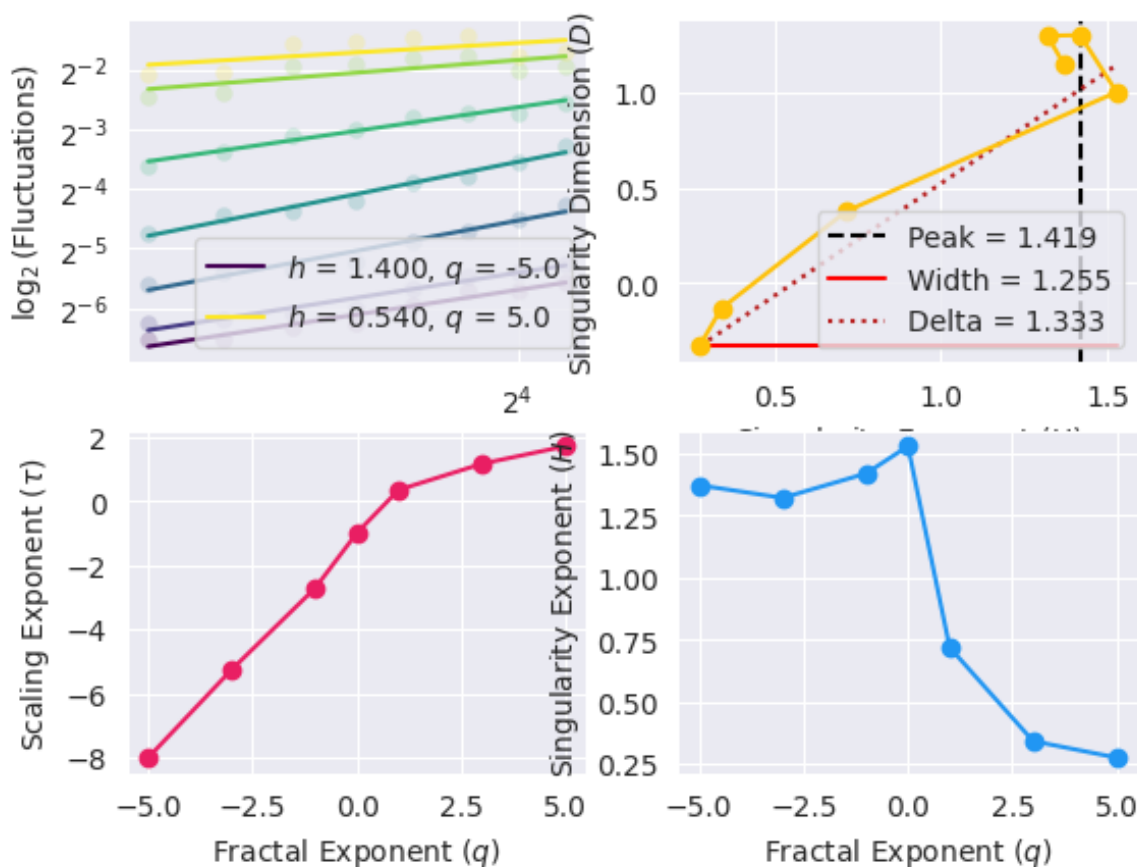


Рисунок 4.7 – Графічна репрезентація мультифрактальних характеристик аномального сигналу

Таким чином, аналіз розбіжностей в таких метриках, як ширина, максимум, середнє, асиметрія або коливання, може бути корисним для виявлення аномальних станів, що вимагають подальшого медичного обстеження або втручання.

4.3 Проведення експерименту

Для початку ми формуємо виборки X (features) та Y (targets). Одним з найпоширеніших типів машинного навчання є "навчання з вчителем" (supervised learning), де для усіх прикладів в навчальних даних відома відповідь

(мітка чи ціль Y). Задача цього типу навчання - навчити модель правильно прогнозувати ці відповіді з нових даних, для яких відповідь невідома, використовуючи ознаки X .

Наступним кроком є розбиття даних на навчальні та тестові набори. Навчальні дані будуть використовуватися для навчання нашої моделі, а тестові дані – для перевірки продуктивності моделі на невідомому наборі даних. Ми використовуємо співвідношення 80-20, тобто 80% даних використовуються для навчання і 20% – для тестування.

Для експерименту використаємо багат шарову згорткову нейронну мережу, через її високу ефективність в задачах класифікації, здатність виявляти локальні патерни та здатність працювати з вхідними даними різного розміру. На вхід нейронна мережа отримує сигнал ЕКГ (одномірний масив), а на вихід віддає клас (ціле число в діапазоні від 1 до 5).

Згорткова нейронна мережа (CNN) складається з декількох основних етапів:

Згортковий шар (Convolution Layer): Цей шар використовує операцію згортки на вхідних даних та фільтру, визначаючи що слід виявити, щоб вивчити важливі особливості сигналу.

Функція активації / нелінійність (ReLU): Цей шар застосовує функцію активації, як-от ReLU, до вихідних даних з згорткового шару.

Pooling або Sub Sampling: Цей шар використовується для зменшення просторового розміру (ширина та висота) вхідних даних.

Повнозв'язний шар (Fully Connected Layer): Це шар, де виконуються подібні до класичної нейронної мережі обчислення. Кожен нейрон в цьому шарі пов'язаний з кожним нейроном в попередньому шарі.

Softmax: Цей шар включає в себе функції активації, які використовуються для класифікаційних задач. Softmax використовується для багатокласової класифікації.

Схема нашої згорткової нейронної мережі отримана за допомогою `plot_model` бібліотеки `tensorflow.keras.utils` наведена на рис. 4.8.

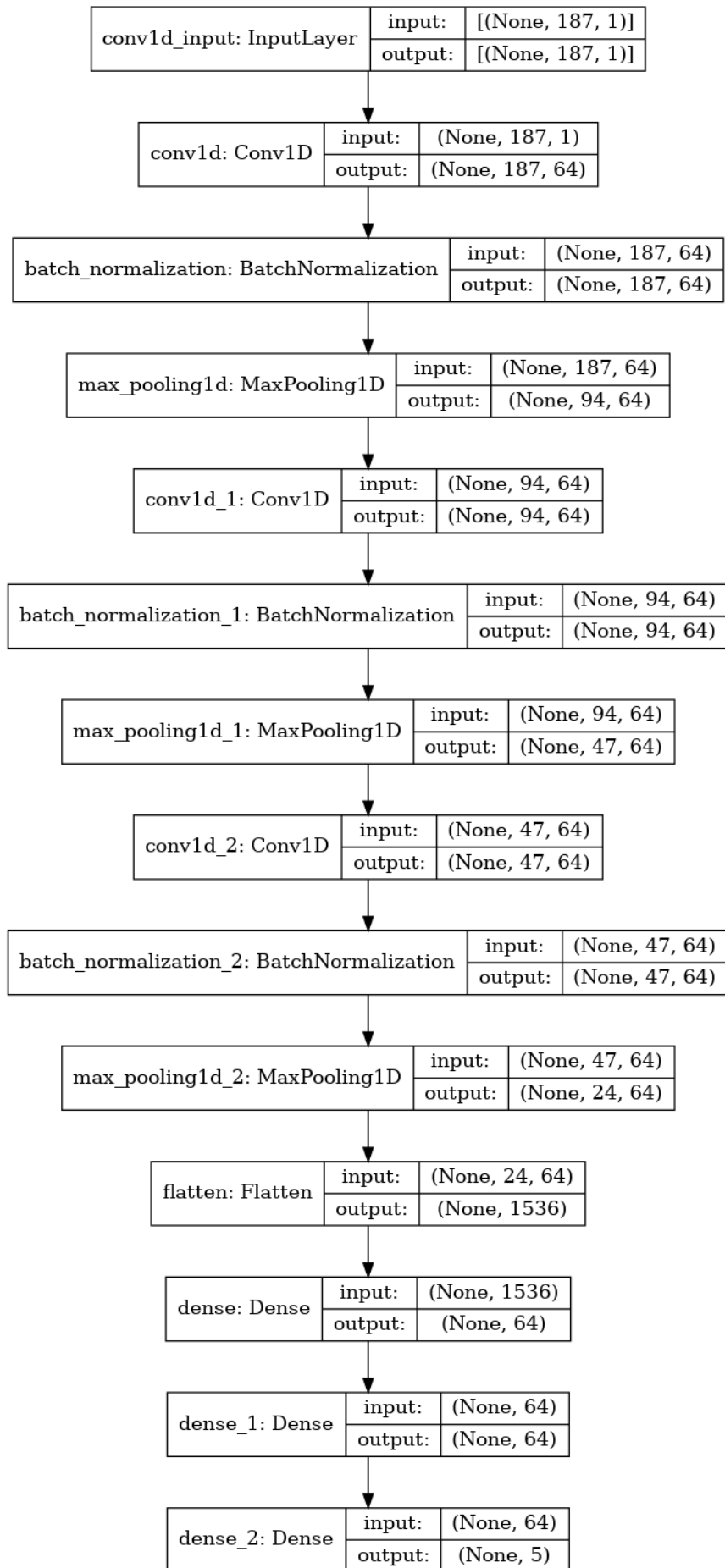


Рисунок 4.8 – Схема згорткової нейронної мережі

4.4 Точність роботи класифікатора

Для оцінки ефективності нейронної мережі (класифікатора), такої як CNN, використовуються різні метрики. Нижче представлений опис найпоширеніших з них.

Accuracy (Точність). Ця метрика вимірює відсоток вірно класифікованих інстанцій. Це один з найпростіших методів оцінки точності моделі. Додатково, "val_accuracy" - це точність на валідаційному наборі даних.

Loss (Втрата): Loss або cost function – це функція, яку оптимізує нейронна мережа під час тренування, намагаючись мінімізувати її за допомогою методів оптимізації, таких як стохастичний градієнтний спуск. Кожна епоха тренування є спробою зменшити значення цієї функції. Додатково, "val_loss" – це втрата на валідаційному наборі даних.

Графічна репрезентація метрик точності роботи класифікатора наведена на рис. 4.9, а втрат на рис. 4.10.

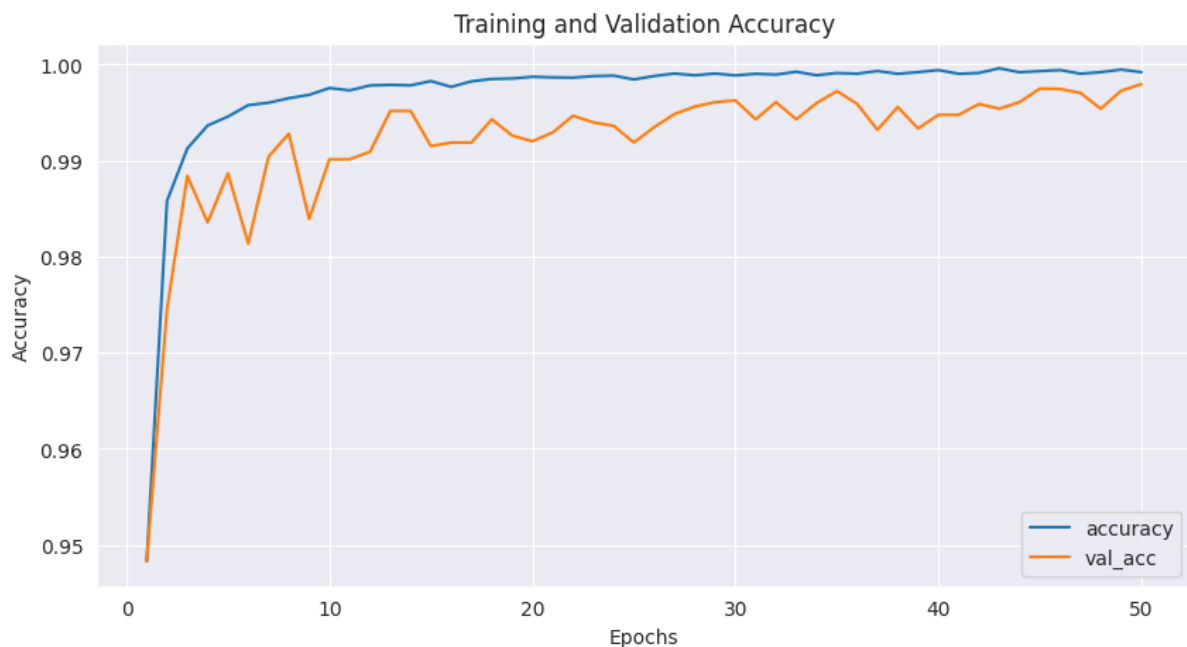


Рисунок 4.9 – Крива зміни точності нейронної мережі під час навчання

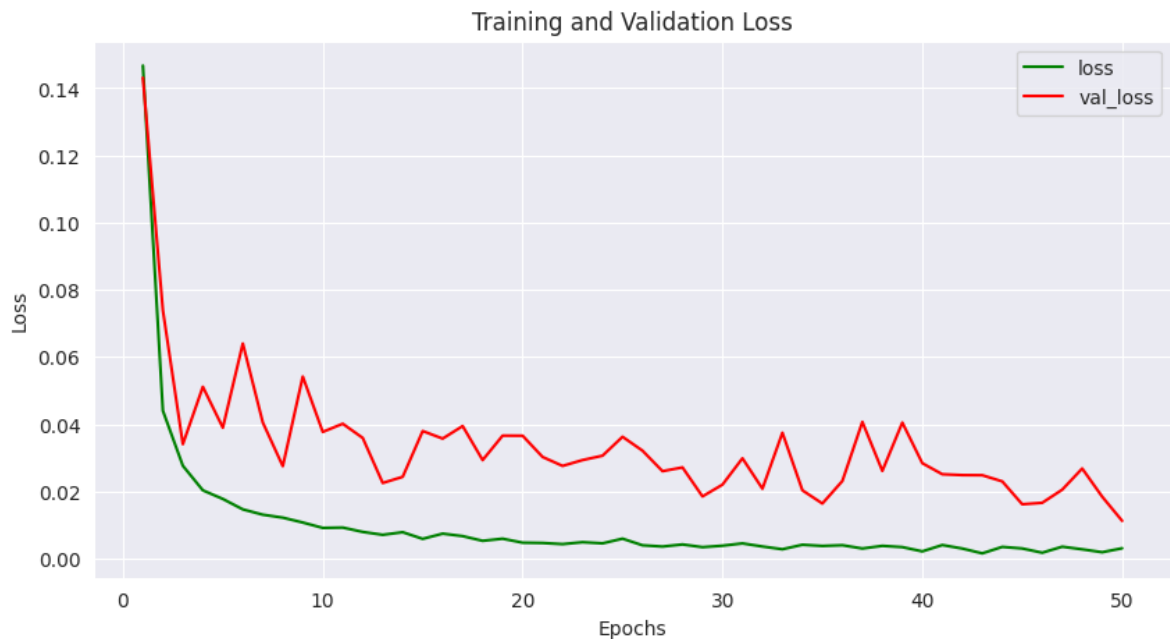


Рисунок 4.10 – Крива зміни втрати нейронної мережі під час навчання

Висока точність і низька втрата свідчать про те, що модель добре працює на навчальних даних.

Однак, потрібно перевірити, чи не перенавчилася та досконально запам'ятала навчальний датасет або чи не підганяє результати під ці конкретні дані наша модель. Якщо це так, вона може мати високу точність і низькі втрати на навчальних даних, але погано справлятися з новими даними або даними валідації.

Також, багато залежить від розподілу та якості даних. Модель може мати високі показники точності та низькі втрати, якщо дані відповідають розподілу навчального датасету. Саме для цього ми виконали попередню аргументацію даних, та привели дані в датасеті до нормального розподілу.

Про високу коректність моделі можна говорити лише тоді, коли вона показує високу точність і низькі втрати і на тестових даних, або даних, які модель раніше не бачила.

Візуалізуємо результати за допомогою матриці помилок, а саме методу `confusion_matrix` пакету `sklearn.metrics`. Результати наведені на рис. 4.11.

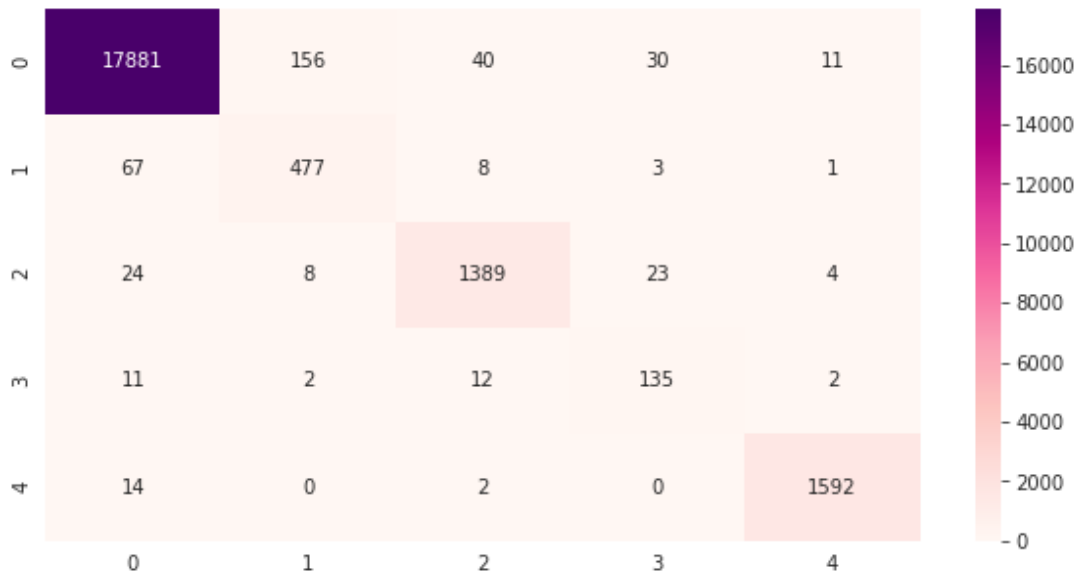


Рисунок 4.11 – Матриця помилок нейронної мережі на тестових даних

Таблиця 4.2 – Результати класифікації та її точність

Клас	True Positives	False Negatives	Точність
N	17881	116	0.994
S	477	166	0.742
V	1389	62	0.957
F	135	56	0.707
Q	1592	18	0.989

Результати класифікації приведені в таблиці 4.2 демонструють доволі високу точність визначення нормальних, вентрикулярних та невідомих серцевих ударів і відносно високу точність визначення супервентрикулярних та фузійних серцевих ударів. Не така висока точність визначення окремих класів може бути пов'язана з наступними причинами:

- висока варіабельність у патернах сигналів цих патологій, що може викликати складнощі для моделі;
- набір даних відповідних цим класам може бути не достатньо вдалим для визначення індивідуальних характеристик класу нейронною мережею;
- перенавчність моделі на інших класах;
- шум в даних, або помилки в оригінальних мітках класів.

Втім, загалом, модель показала вдалу роботу. Можемо припустити, що це результат навчання з «учителем», зокрема тому що цей метод вважається одним з найефективніших при навчанні класифікаторів. Крім того, це може бути результатом ретельної підготовки даних, що знову ж таки сприяє вдалому навчанню моделі.

Висновки за розділом 4

Проведений експеримент показав потенціал використання згорткових нейронних мереж (CNN) у класифікації біоелектричних сигналів. Ми використали дані з датасету ECG Heartbeat Categorization Dataset, вивчили їх структуру та проаналізували мультифрактальні властивості.

Було застосовано різні алгоритми для попередньої обробки даних, включаючи нормалізацію, фільтрацію і збільшення даних, щоб підготувати їх до подальшого аналізу.

Створений нами класифікатор, заснований на CNN був натренований на тренувальному наборі даних де показав високу точність у визначенні належності сигналю до певного класу. Нарешті, модель підтвердила свою ефективність на тестовому наборі даних.

Висока точність результатів підтверджує потенціал глибокого навчання і CNN, зокрема, у задачах класифікації біоелектричних сигналів, автоматично виявляючи не помітні людському оку особливості, які можуть свідчити про наявність аномалій.

ВИСНОВКИ

В ході даної кваліфікаційної роботи було проведено огляд та аналіз сучасного стану задачі класифікації біоелектричних сигналів за допомогою машинного навчання та методів мультифрактального аналізу. Розглянуто основні поняття та властивості мультифрактальних випадкових процесів, методу мультифрактального флуктуаційного аналізу, машинного навчання та нейронних мереж. Проведено огляд методів класифікації біоелектричних сигналів, таких як машинне навчання та нейронні мережі, зокрема, згорткові (CNN).

Було розглянуто задачу класифікації біоелектричних сигналів, використовуючи датасет «ECG Heartbeat Categorization Dataset», який складається з двох відомих наборів даних класифікації серцебиття - «MIT-BIH Arrhythmia Dataset» та «The PTB Diagnostic ECG Database». Ряди розбиті на п'ять класів. В якості класифікатора виступає багатомасштабна згорткова нейронна мережа.

Дані були проаналізовані та попередньо оброблені. Було застосовано ряд алгоритмів, спрямованих на обробку даних, виділення характеристик, притаманних певним компонентам, та нормуванням їх для подальшої класифікації нейронною мережею. Для додаткового аналізу були застосовані методи мультифрактального аналізу.

Результати експерименту демонструють, що обраний метод класифікації виявився досить точним. Висновки з цієї роботи можуть бути використані для виявлення відхилень та для класифікації біоелектричних сигналів із застосуванням машинного навчання.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Kobayashi M., Musha T. 1/f fluctuation of heartbeat period. *IEEE Trans. Biomed. Eng.* 1992. Vol. 29, № 6. P. 456–457.
2. Multifractal detrended fluctuation analysis of non-stationary time series / J.W. Kantelhardt, S.A. Zschiegner, A. Bunde, S. Havlin, E. Koscielny-Bunde, H.E. Stanley. *Physica A*. 2002. № 316. P. 87–114.
3. Al-ani T., Kazbunda D., Novrak D. Automatic sleep scoring based only on electrocardiogram records. Proc. EUROSIM, 2007.
4. Non-Gaussian heart rate as an independent predictor of mortality of chronic heart failure patients with chronic heart failure / Kiyono K., Hayano J., Watanabe E., Struzik Z.R., Yamamoto Y. *Heart Rhythm Society*. 2008. Vol. 5, Issue 2. P. 261-268.
5. Hoshiyama M., Hoshiyama A. Heart Rate Variability Associated with Experienced Zen Meditation. *Computers in Cardiology*. 2008. P. 569-572.
6. From 1/f noise to multifractal cascades in heartbeat dynamics / P. Ivanov, L.N. Amaral, A.L. Goldberger., S. Havlin, M.G. Rosenblum, H. Stanley, Z.R. Struzik *Chaos*. 2001. Vol. 11, № 3. P. 641–652.
7. Methodology for multifractal analysis of heart rate variability: from LF/HF ratio to wavelet leaders / P. Abry, H. Wendt, S. Jaffard, H. Helgason, P. Goncalves, E. Pereira, C. Gharib, P. Gaucherand, M. Doret. *Conf Proc IEEE Eng Med Biol Soc*. 2010.
8. Hwa C.R., Ferree T.C. Stroke detection based on the scaling properties of human EEG. *Physica A: Statistical Mechanics and its Applications*. 2004. Vol. 338, Issue 1. P. 246–254.
9. Shin HB, Jeong DU, Kim EJ. Detrended Fluctuation Analysis on Sleep EEG of Healthy Subjects. *Sleep Med Psychophysiol*. 2007. № 14 (1). P. 42–48.
10. Study of EEG Brain Maturation Signals with Multifractal Detrended Fluctuation Analysis / A. Figliola, E. Serrano, J.A.P. Rostas, M. Hunter, O.A. Rosso. *AIP Conf. Proc.* 2007. Vol. 913, Issue 1. P. 190–195.
11. Detrended fluctuation analysis of resting EEG in depressed outpatients and healthy controls / J.-S. Lee, B.-H. Yang^b, J.-H. Lee, J.-H. Choi, I.-G. Choi, S.-B. Kim. *Clinical Neurophysiology*. 2007. Vol. 118, Issue 11. P. 2489–2496.

12. A Study on the Possible Usefulness of Detrended Fluctuation Analysis of the Electroencephalogram Background Activity in Alzheimer's Disease / Abasolo D., Hornero R., Escudero J., Espino P. *IEEE transactions on biomedical engineering*. 2008. Vol. 55, № 9. P. 2171–2179.

13. Tingting Gao, Dan Wu, Dezhong Yao. EEG Scaling Difference Between Eyes-Closed and Eyes-Open Conditions by Detrended Fluctuation Analysis. *Advances in Cognitive Neurodynamics ICCN*. 2007, 2008. P. 501–504.

14. Kirichenko L., Zinchenko P., Radivilova T. Classification of Time Realizations Using Machine Learning Recognition of Recurrence Plots. Lecture Notes in Computational Intelligence and Decision Making. ISDMCI 2020. *Advances in Intelligent Systems and Computing*. 2021. Vol 1246. Springer, Cham. P. 687-696.

15. Kirichenko L., Radivilova T., Bulakh V. (2020) Binary Classification of Fractal Time Series by Machine Learning Methods. In: Lytvynenko V., Babichev S., Wójcik W., Vynokurova O., Vyshemyrskaya S., Radetskaya S. (eds) Lecture Notes in Computational Intelligence and Decision Making. ISDMCI 2019. *Advances in Intelligent Systems and Computing*. Vol 1020. Springer, Cham. P. 701-711.

16. Time Series Classification. URL: <http://www.timeseriesclassification.com> (дата звернення: 30.10.2023).

17. Python. URL: <https://www.python.org> (дата звернення: 28.10.2023).

18. Neurophysiological Data Analysis with NeuroKit2. URL: <https://neuropsychology.github.io> (дата звернення: 15.11.2023).

19. TensorFlow. URL: <https://www.tensorflow.org> (дата звернення: 10.11.2023).

20. Kaggle: Your Machine Learning and Data Science Community. URL: <https://www.kaggle.com> (дата звернення: 10.11.2023).

21. Харченко Я. О. Аналіз мультифрактальних біоелектричних реалізацій. *27-й Міжнародний молодіжний форум «Радіоелектроніка і молодь у XXI столітті»* : зб. матеріалів форуму. Т. 7. Харків : ХНУРЕ, 2023. С. 221–222.