

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ
імені ІГОРЯ СІКОРСЬКОГО»
ФАКУЛЬТЕТ БІОМЕДИЧНОЇ ІНЖЕНЕРІЇ
(повна назва інституту/факультету)
кафедра БІОМЕДИЧНОЇ КІБЕРНЕТИКИ
(повна назва кафедри)

«До захисту допущено»
В.о. завідувачки кафедри БМК

Світлана АЛХІМОВА
(підпис) (ініціали, ПРІЗВИЩЕ)

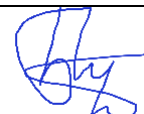
“ ” червня 2025р.

Дипломна робота
на здобуття ступеня бакалавра
за освітньо-професійною програмою
«Комп'ютерні технології в біології та медицині»
зі спеціальності 122 «Комп'ютерні науки»
на тему: **Прогнозування розвитку тремору на основі даних
інерційних сенсорів**

Виконав: студент IV курсу, групи БС-15

ЧЕРВОНЯК БОГДАН ВОЛОДИМИРОВИЧ

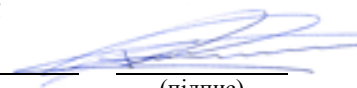
(прізвище, ім'я, по батькові)



(підпис)

Науковий керівник: *асистент каф. біомедичної кібернетики
(БМК) ас. каф. БМК Давидович Ілля Вікторович*

(посада, науковий ступінь, вчене звання, прізвище, ім'я, по ініціали)



(підпис)

Консультант з розділів дипломної роботи:

к.ф.-м.н., доцент КБМК Федорін Ілля Валерійович

(назва розділу) (посада, вчене звання, науковий ступінь, прізвище, ініціали)

(підпис)

Рецензент: *професор каф. біомедичної інженерії, д.м.н.*

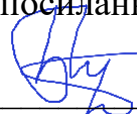
Максименко Віталій Борисович

(посада, науковий ступінь, вчене звання, науковий ступінь, прізвище та ініціали)

(підпис)

Засвідчую, що у цій дипломній роботі
немає запозичень з праць інших авторів
без відповідних посилань.

Студент



(підпис)

Київ – 2025 року

Національний технічний університет України
«Київський політехнічний інститут імені Ігоря Сікорського»
Інститут (факультет) БІОМЕДИЧНОЇ ІНЖЕНЕРІЇ

(повна назва)

Кафедра БІОМЕДИЧНОЇ КІБЕРНЕТИКИ

(повна назва)

Рівень вищої освіти – перший (бакалаврський)

Спеціальність 122 «Комп'ютерні науки»

Освітньо-професійна програма «Комп'ютерні технології в біології та медицині»

(код і назва)

ЗАТВЕРДЖУЮ

В.о. завідувачки кафедри БМК

Світлана АЛХІМОВА

« » квітень 2025 р.

ЗАВДАННЯ
на дипломну роботу студенту

ЧЕРВОНЯКУ БОГДАНУ ВОЛОДИМИРОВИЧУ

(прізвище, ім'я, по батькові)

1. Тема роботи Прогнозування розвитку тремору на основі даних інерційних сенсорів

керівник роботи

Давидович Ілля Вікторович асистент каф. БМК

(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом по університету від «26» травня 2025 р. №1758-с

2. Термін подання студентом роботи 22-23 травня 2025 року

3. Вихідні дані: *Датасет містить сигнали IMU (Inertial Measurement Unit), зібрані за допомогою акселерометрів, вбудованих у смартфони, для виявлення епізодів тремору, пов'язаних із хворобою Паркінсона. У дослідженні взяли участь 31 пацієнт з ХП та 14 здорових людей контрольної групи, які надали дані акселерометра зі своїх смартфонів протягом декількох місяців.*

4. Зміст роботи: *анотації українською та англійською мовами, перелік умовних позначень, символів, одиниць, скорочень і термінів, вступ, чотири розділи основної частини (теоретико-методологічні основи дослідження; методика реалізації моделі прогнозування; збір та оцінка результатів; розробка веб-інтегрованої системи для використання моделей), загальні висновки, список використаних джерел.*

5. Перелік ілюстративного матеріалу (із зазначенням плакатів, презентацій тощо): *презентація на 15 слайдах, 52 рисунки, 3 таблиці.*

6. Консультанти розділів роботи^{1*}

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Дипломної роботи	Федорін Ілля Валерійович, к.ф.-м.н., доцент КБМК ФБМІ	07.04.25	17.05.25

7. Дата видачі завдання **06 квітня 2025р.**

Календарний план

№ з/п	Назва етапів виконання дипломної роботи	Термін виконання етапів роботи	Примітка
1	Отримати завдання за темою ДР на практику	До 24.12.2024р.	виконано
2	Переддипломна практика. Виконання практичної частини ДР та додаткових розділів	1-4 тиждень за графіком практики	виконано
3	Виконання основних розділів ДР (теоретико-методологічні основи дослідження, методика реалізації моделі прогнозування, збір та оцінка результатів, розробка веб-інтегрованої системи для використання моделей.)	Протягом усієї практики	виконано
4	Перевірка ДР науковим керівником	до 20.05.2025	виконано
5	Подання в електронному вигляді ДР на перевірку нормоконтролера та плагіат (StrikePlagiarism.com). Доопрацювання ДР	Не пізніше 23.05.2025	виконано
6	Отримати відгук від керівника ДР	Не пізніше 05.06.2025	виконано
7	Надати пакету документів по ДР та супровідних до неї документів на засідання кафедри		виконано
8	Подання ДР рецензенту. Отримання рецензії.	09.06 – 13.06.2025	виконано
9	Захист ДР в ЕК	16.06 - -21.06.2025	виконано

Студент



(підпис)

Богдан ЧЕРВОНЯК

(ім'я, ПРІЗВИЩЕ)

Керівник ДР

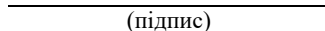


(підпис)

Ілля ДАВИДОВИЧ

(ім'я, ПРІЗВИЩЕ)

Нормоконтролер



(підпис)

Галина КОРНІЄНКО

(ім'я, ПРІЗВИЩЕ)

АНОТАЦІЯ

Дипломна робота за темою «Прогнозування розвитку тремору на основі даних інерційних сенсорів» виконана студентом кафедри біомедичної кібернетики ФБМІ Червоняком Богданом Володимировичем зі спеціальності 122 «Комп'ютерні науки» за освітньо-професійною програмою «Комп'ютерні технології в біології та медицині» та складається зі: вступу; 4 розділів (теоретико-методологічні основи дослідження, методика реалізації моделі прогнозування, збір та оцінка результатів, розробка веб-інтегрованої системи для використання моделей), висновків до кожного з цих розділів; загальних висновків; списку використаних джерел, який налічує 36 джерел та додатків. Загальний обсяг роботи 86 сторінок.

Актуальність теми. Хвороба Паркінсона є однією з найпоширеніших нейродегенеративних патологій, що потребує раннього виявлення тремору для ефективного лікування. Збір даних у реальному часі з мобільних пристроїв із сенсорами дозволяє створювати високоточні моделі для прогнозування розвитку симптомів. Це відкриває нові можливості для автоматизованого моніторингу стану пацієнтів та покращення якості медичної допомоги.

Мета і завдання роботи.

Мета роботи полягає в удосконаленні підходу до прогнозування розвитку тремору у пацієнтів із хворобою Паркінсона шляхом аналізу даних інерційних сенсорів мобільних пристроїв із використанням алгоритмів штучних нейронних мереж.

Її досягнення передбачає вирішення наступних завдань:

1. Визначити теоретичні засади прогнозування тремору на основі інерційних даних.
2. Провести порівняльний аналіз методів обробки та моделювання даних.
3. Підготувати та структурувати вхідні дані для моделювання.
4. Розробити та оптимізувати модель прогнозування тремору.
5. Оцінити ефективність моделі та сформулювати рекомендації для подальших досліджень.

Використані методи. Методи статистичної обробки, , нейронні мережі FNN і DNN, Python, Flask.

Отримані результати. Розроблено модель прогнозування тремору у пацієнтів з хворобою Паркінсона на основі інерційних даних мобільних пристроїв. Створено веб-застосунок, що дозволяє автоматично аналізувати дані, виявляти тремор і демонструє високу точність прогнозування. Ефективність моделей підтверджено за допомогою експеримента.

Публікації. Не заплановано.

Ключові слова. Тремор, хвороба Паркінсона, інерційні сенсори, нейронні мережі, машинне навчання, веб-застосунок.

Бібліографічний опис ДР

Червоняк Б. В. Прогнозування розвитку тремору на основі даних інерційних сенсорів : дипломна роб. бакалавра : 122 Комп'ютері науки / Червоняк Богдан Володимирович. – Київ, 2025. – 86 с

ABSTRACT

The bachelor's thesis titled "Predicting the development of tremor based on inertial sensor data" completed by *Bohdan Volodymyrovych Chervoniak*, a student of the Department of Biomedical Cybernetics at the Faculty of Biomedical Engineering, specializing in 122 "Computer Science" under the educational and professional program "Computer Technologies in Biology and Medicine". The work consists of: an introduction; 4 chapters (theoretical and methodological foundations of the study, implementation methodology of the prediction model, data collection and evaluation of results, development of a web-integrated system for model utilization), conclusions for each chapter; general conclusions, a list of references containing 36 sources, and appendices. The total volume of the work is 86 pages.

Relevance of the topic.

Parkinson's disease is one of the most common neurodegenerative disorders that requires early detection of tremor for effective treatment. Real-time data collection from mobile devices equipped with sensors enables the creation of highly accurate models for predicting symptom development. This opens new opportunities for automated patient monitoring and improving the quality of medical care

Purpose and objectives of the thesis.

The aim of the study is to improve the approach to predicting the progression of tremor in patients with Parkinson's disease by analyzing data from inertial sensors of mobile devices using artificial neural network algorithms. To achieve this goal, the following tasks were set:

1. Define the theoretical foundations of tremor prediction based on inertial data.
2. Conduct a comparative analysis of data processing and modeling methods.
3. Prepare and structure input data for modeling.
4. Develop and optimize the tremor prediction model.
5. Evaluate the model's effectiveness and formulate recommendations for further research.

Methods used. Statistical processing methods, FNN and DNN neural networks, Python, Flask.

Results. A tremor prediction model for patients with Parkinson's disease based on inertial data from mobile devices was developed. A web application was created that allows automatic data analysis, tremor detection, and demonstrates high prediction accuracy. The effectiveness of the models was confirmed experimentally.

Publications. No publications are planned based on this thesis.

Keywords. Tremor, Parkinson's disease, inertial sensors, neural networks, machine learning, web application.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ	9
ВСТУП.....	10
РОЗДІЛ 1 ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ОСНОВИ ДОСЛІДЖЕННЯ	12
1.1 Аналіз проблематики дослідження	12
1.2 Методи аналізу інерційних даних	16
1.3 Моделі машинного навчання	21
Висновки до розділу 1	26
РОЗДІЛ 2 МЕТОДИКА РЕАЛІЗАЦІЇ МОДЕЛІ ПРОГНОЗУВАННЯ	27
2.1 Опис набору даних.....	27
2.2 Обґрунтування вибору моделей прогнозування	36
Висновки до розділу 2	40
РОЗДІЛ 3 ЗБІР ТА ОЦІНКА РЕЗУЛЬТАТІВ	42
3.1 Попередня обробка даних	42
3.2 Проведення експериментів.....	56
3.3 Оцінка ефективності моделей	58
3.4 Розрахунок економічного ефекту	65
Висновки до розділу 3	67
РОЗДІЛ 4 РОЗРОБКА ВЕБ-ІНТЕГРОВАНОЇ СИСТЕМИ ДЛЯ ВИКОРИСТАННЯ МОДЕЛЕЙ.....	69
Висновки до розділу 4	79
ЗАГАЛЬНІ ВИСНОВКИ	80
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ.....	82

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

БД – База даних, це структурована колекція даних, яка зберігається та управляється комп'ютерною системою.

FNN – Персептрони (Feedforward Neural Network).

CNN – Згорткові нейронні мережі (Convolutional Neural Network).

DNN – Глибокі нейронні мережі (Deep Neural Network).

SVM – Метод опорних векторів (Support Vector Machine).

RGB – Червоний, зелений, синій (Red, Green, Blue).

GUI – Графічний інтерфейс користувача (Graphical User Interface).

ReLU – Функція активації (Rectified Linear Unit).

API – Інтерфейс прикладного програмування (Application Programming Interface).

AES – Стандарт шифрування (Advanced Encryption Standard).

TLS – Протокол захисту транспортного рівня (Transport Layer Security).

HTTPS – Захищений протокол передачі гіпертексту (Hypertext Transfer Protocol Secure).

IMU – Інерційний вимірювальний блок (Inertial Measurement Unit).

LSTM – Довга короткочасна пам'ять (Long Short-Term Memory).

RVFL – Мережа випадкових векторних функціональних ланок (Random Vector Functional Link).

PCA – Аналіз головних компонент (Principal Component Analysis).

MAD – Медіанне абсолютне відхилення (Median Absolute Deviation).

ВСТУП

Актуальність дослідження. У сучасних умовах зростання кількості захворювань нервової системи, особливої уваги заслуговує хвороба Паркінсона — прогресуюча нейродегенеративна патологія, яка значно знижує якість життя пацієнтів. Своєчасна діагностика та прогнозування розвитку симптомів, зокрема тремору, є критично важливими для персоналізованого підходу до лікування. Використання інерційних сенсорів, вбудованих у мобільні пристрої, відкриває нові можливості для неперервного збору біомедичних даних у реальному часі. Це забезпечує доступність, достовірність і високу репрезентативність даних. Застосування сучасних методів інтелектуального аналізу, зокрема штучних нейронних мереж, дозволяє моделювати складні нелінійні залежності та виявляти приховані закономірності в сигналах, що надходять від пацієнтів. Такий підхід підвищує точність прогнозування та забезпечує основу для розробки адаптивних стратегій лікування.

Мета і завдання роботи

Метою роботи є удосконалення підходу до прогнозування розвитку тремору у пацієнтів із хворобою Паркінсона шляхом аналізу даних інерційних сенсорів мобільних пристроїв із використанням алгоритмів штучних нейронних мереж.

Її досягнення передбачає вирішення наступних завдань:

- аналіз вітчизняних та зарубіжних джерел щодо сучасних методів прогнозування тремору;
- визначення теоретичних засад прогнозування тремору на основі інерційних даних;
- проведення порівняльного аналізу методів обробки та моделювання даних;
- підготовка та структурування вхідних даних для моделювання;
- розробка та оптимізація моделі прогнозування тремору;
- оцінка ефективності моделі та формулювання рекомендацій для подальших досліджень.

Використані методи

У роботі застосовано методи статистичної обробки даних, частотного аналізу сигналів, цифрової фільтрації, а також алгоритми машинного навчання, зокрема Feedforward Neural Network (FNN) та Deep Neural Network (DNN).

Отримані результати

У результаті дослідження було розроблено і протестовано модель прогнозування тремору, яка забезпечує високу точність класифікації на основі сигналів акселерометрів. Побудовано та протестовано декілька архітектур штучних нейронних мереж, обґрунтовано вибір оптимального підходу. Проведено експериментальну оцінку ефективності моделей та здійснено розрахунок потенційного економічного ефекту. Розроблено веб-інтерфейс для практичного застосування моделі.

Апробація результатів роботи

Апробація результатів роботи не здійснювалась.

Публікації

За результатами виконаної роботи публікації відсутні.

Структура роботи

Дипломна робота за темою «Прогнозування розвитку тремору на основі даних інерційних сенсорів» виконана студентом *Червоняком Богданом Володимировичем* зі спеціальності 122 «Комп'ютерні науки» за освітньо-професійною програмою «Комп'ютерні технології в біології та медицині», побудована за класичним типом та викладена на 86 сторінках машинописного тексту. Вона складається з: вступу; чотирьох розділів (1. Теоретико-методологічні основи дослідження, 2. Методика реалізації моделі прогнозування, 3. Збір та оцінка результатів, 4. Розробка веб-інтегрованої системи для використання моделей), висновків до кожного з цих розділів; загальних висновків; списку використаних джерел, який налічує 36 джерел (36 – на латиниці), та додатків. У роботі представлено 52 рисунки і 3 таблиці.

РОЗДІЛ 1

ТЕОРЕТИКО-МЕТОДОЛОГІЧНІ ОСНОВИ ДОСЛІДЖЕННЯ

1.1 Аналіз проблематики дослідження

Хвороба Паркінсона є прогресуючим нейродегенеративним захворюванням, яке впливає на роботу центральної нервової системи. Головні симптоми цього розладу охоплюють: ригідність (збільшений м'язовий тонус), брадикінезію (зменшення швидкості рухів), постуральну нестабільність (проблеми з утриманням рівноваги) та тремор [25].

Тремор є однією з найраніших ознак розвитку хвороби Паркінсона і відображає не тільки присутність захворювання, але й його прогресування. Регулярне вивчення механізмів появи тремору має важливе значення для ранньої діагностики та створення дієвих лікувальних методик. Основні патофізіологічні індикатори, пов'язані з розвитком тремору, включають:

1. Нестачу дофаміну через дегенеративні зміни в нейронах чорної речовини, що викликає дисбаланс інгібуючих та збудливих сигналів у базальних гангліях, що є первинною причиною тремору.

2. Зменшення нейронної маси та порушення синаптичної комунікації, які разом з дисфункцією зв'язків у моторних центрах мозку, дозволяють кількісно оцінити параметри тремору, включаючи його частотні характеристики.

3. Вікові чинники та генетична схильність, які збільшують ризик розвитку захворювання, при цьому загальний стан здоров'я та супутні захворювання часто пов'язані з появою хвороби Паркінсона [10].

Розглянемо ієрархію факторів-збудників виникнення тремору (рис. 1.1).



Рисунок 1.1 – Збудники виникнення тремору

Тремор, що спостерігається при хворобі Паркінсона, має властивості, які дозволяють його відрізнити: він з'являється, коли людина перебуває в стані спокою, має частоту від 4 до 6 Гц і стає сильнішим під час емоційного стресу, на відміну від тремору, пов'язаного з фізичною активністю. Тремор у випадку паркінсонізму є розповсюдженим і може вражати різні частини тіла, хоча зазвичай спочатку він з'являється в одній кінцівці. Хоча тремор не є обов'язковим симптомом для всіх пацієнтів з хворобою Паркінсона, попри його значення для діагностики [31]. Діагностування хвороби Паркінсона представляє собою значний клінічний виклик через відсутність конкретних біохімічних чи інструментальних маркерів, які б однозначно підтверджували наявність захворювання.

Діагностичний процес ґрунтується на ідентифікації набору клінічних ознак, пов'язаних з цим станом. Для стандартизації оцінки симптомів використовуються перевірені шкали, наприклад, UPDRS. Основною діагностичною проблемою є схожість симптомів хвороби Паркінсона з ознаками вторинного паркінсонізму, прогресуючого над'ядерного паралічу та мультисистемної атрофії, що ускладнює диференційну діагностику через суб'єктивність клінічної оцінки та складність аналізу різних факторів. Іншою діагностичною проблемою є те, що на початкових стадіях симптоми часто неправильно трактуються як знаки фізичної втоми або

вікові зміни, що не привертає належної уваги з боку пацієнтів. Єдиним абсолютно достовірним методом підтвердження діагнозу залишається посмертне гістологічне дослідження мозкових тканин з виявленням характерних тілець Леві, які є агрегатами білка альфа-синуклеїну [21]. Враховуючи зазначені діагностичні обмеження, сучасний підхід до виявлення хвороби Паркінсона включає:

1. Неврологічне оцінювання моторних симптомів з використанням фармакологічних тестів на дофамінергічні препарати, позитивна реакція на які збільшує ймовірність наявності хвороби Паркінсона.

2. Детальний аналіз анамнезу, включаючи сімейну історію для виявлення генетичної схильності, а також вивчення травматичного анамнезу та впливу потенційних нейротоксинів.

3. Застосування передових методів нейровізуалізації, наприклад, магнітно-резонансна та комп'ютерна томографія, які дозволяють виключити вторинні причини паркінсонізму, зокрема пухлини та локальні ураження мозку [3].

Останні роки позначені активним прогресом у розробці новітніх методів діагностики хвороби Паркінсона та аналізу тремору за допомогою передових комп'ютерних технологій та штучного інтелекту, що дозволяє знизити вплив суб'єктивної оцінки на аналіз клінічних даних та підвищити діагностичну точність. Сучасні методи збору та обробки даних включають використання носимих девайсів із вбудованими датчиками для неперервного моніторингу фізіологічних показників в реальному часі, що дозволяє точно фіксувати тремор, брадикінезію та інші рухові порушення. Важливу роль у діагностиці відіграють амплітудно-частотні параметри роботи центральної нервової системи. Також використовується спеціалізоване програмне забезпечення для точного збору інформації про рухову активність хворих, що допомагає виявляти ритмічні моделі та визначати низькочастотний тремор. Крім того, проводяться біомаркерні дослідження, зокрема аналіз рівня альфа-синуклеїну в цереброспінальній рідині, крові та тканинах, що допомагає оцінити ризик розвитку патології. Генетичні маркери, такі як мутації у генах LRRK2 і GBA, можуть свідчити про збільшений ризик захворювання. Функціональні методи нейровізуалізації, включно з сцинтиграфією, дозволяють кількісно оцінити активність дофамінових рецепторів

у базальних гангліях та відрізнити хворобу Паркінсона від інших нейродегенеративних розладів. Для цього застосовуються методи однофотонної емісійної комп'ютерної томографії та позитронно-емісійної томографії [30].

Кількісний аналіз та моніторинг тремору супроводжуються рядом методологічних викликів. Передусім це стосується гетерогенності даних, отриманих за допомогою інерційних датчиків та акселерометрів, що виникає через варіативність тривалості та кількості записів, а також через різну якість записуючих пристроїв. Це призводить до появи шуму на різних рівнях та ускладнює стандартизацію даних для подальшого аналізу. Великою проблемою є також виявлення тремору через його низькочастотний характер при хворобі Паркінсона, який може бути прихованим сигналами екзогенного походження, знижуючи надійність результатів. Виділення важливих компонентів сигналу вимагає високої аналітичної кваліфікації. Також багатофакторний та нелінійний характер розвитку тремору та паркінсонізму значно ускладнює визначення аналітичних закономірностей прогресування захворювання, обмежуючи можливості для розробки профілактичних та лікувальних стратегій. Однією з проблем є обмеженість обсягу даних для аналізу, пов'язану з вимогами до конфіденційності медичної інформації та частими відмовами пацієнтів брати участь у дослідженнях [36].

У розрізі обмежень, які мають традиційні аналітичні методи для аналізу закономірностей розвитку тремору та хвороби Паркінсона, дослідники звертаються до методів машинного навчання для виявлення патернів у даних, зібраних за допомогою інерційних датчиків та акселерометрів. Важливі дослідження в цій області включають застосування тензорної декомпозиції для визначення потенційного фізіологічного індикатора, який показує умереднену кореляцію з прогресуванням хвороби Паркінсона [34], а також аналіз зв'язку між тремором, брадикінезією та хворобою Паркінсона з подальшим створенням кореляційної моделі [29].

У дослідженні [18] представлено результати кількісного аналізу за допомогою інерційних датчиків, що дозволило виявити низькочастотні сигнали, типові для спокійного тремору при паркінсонізмі. Перспективним напрямком є

розробка складних класифікаційних моделей для діагностики тремору та хвороби Паркінсона на основі аналізу медичних зображень, включаючи використання мультимодального класифікатора [26].

Інноваційна стратегія полягає у використанні методів глибокого навчання, які дозволяють виявляти складні закономірності в даних та підвищують точність діагностики тремору та хвороби Паркінсона. У дослідженнях [13, 6] були розроблені моделі глибокого навчання та спеціалізований фреймворк для машинного навчання, що дозволяють діагностувати захворювання на основі комплексного аналізу даних, зібраних під час фізичних вправ пацієнтів.

Важливим напрямком досліджень є створення систем не тільки для діагностики захворювань та тремору, але й для постійного моніторингу стану пацієнтів. Розроблена веб-система для моніторингу стану пацієнтів в домашніх умовах допомагає зменшити фінансові витрати на медичне обслуговування та скоротити час лікування в лікарнях [7].

Найбільш передові дослідження спрямовані на створення моделей, здатних обробляти великі масиви даних з різноманітними параметрами, отриманими з датчиків. Такі моделі мають потенціал класифікувати тремор та паркінсонізм, виходячи з виявлення довготривалих патернів у розподілі даних [20]. Проте, цей підхід вимагає значних обчислювальних ресурсів, що на даний момент обмежує його широке застосування.

1.2 Методи аналізу інерційних даних

Дані, отримані з інерційних сенсорів, представляють собою послідовності, які фіксують рухову активність особи. Ці послідовності дозволяють ідентифікувати різноманітні фізіологічні та патологічні стани через створення моделей для аналізу конкретного часового ряду. Інерційні дані часто застосовують для діагностики тремору у пацієнтів, а аналіз цих даних можна представити у вигляді етапів [19]:

1. Первинна обробка даних, яка включає видалення шумів, артефактів, сегментацію та нормалізацію сигналів;

2. Визначення основних характеристик сигналу шляхом розрахунку та аналізу його статистичних та частотних параметрів.

3. Аналіз розподілу сигналів у часі на основі отриманих статистичних та частотних характеристик. Загалом, існують три основні групи методів аналізу даних з інерційних сенсорів [35]:

1. Методи аналізу в часовому діапазоні;
2. Методи аналізу в частотному діапазоні;
3. Методи аналізу в часово-частотному діапазоні.

Нехай задається дискретний сигнал $x[n], n = 0, 1, \dots, N - 1$, тоді стає можливим обчислення таких статистичних характеристик вхідного сигналу [32]:

Середнє значення:

$$\mu = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \quad (1.1)$$

Дисперсія:

$$\sigma^2 = \frac{1}{N} \sum_{n=0}^{N-1} (x[n] - \mu)^2 \quad (1.2)$$

Пікові значення сигналу:

$$x_{max} = \max_{0 \leq n < N} \{x[n]\}, x_{min} = \min_{0 \leq n < N} \{x[n]\} \quad (1.3)$$

Оцінка періодичності сигналу та ідентифікація його попередніх поведінкових моделей є ключовими, особливо у контексті досліджень тремору та хвороби Паркінсона, де такий аналіз сприяє оцінюванню ритмічності рухів пацієнта, що може бути використано для встановлення зв'язку між тремором та присутністю хвороби Паркінсона. Відомо, що поведінка будь-якої функції біля

певної точки ефективно описується за допомогою похідних, які вказують на зміну поведінки функції через наявність перегину, дозволяючи таким чином ідентифікувати періоди активності або спокою пацієнта та порівнювати їх із моментами виникнення тремору. Для виявлення перегинів введемо таку величину $x'[n]$ – першу похідну дискретного сигналу, згідно з теорією різницьових рівнянь, відомо, що похідна визначається наступним чином [32]:

$$x'[n] = x[n] - x[n - 1] \quad (1.4)$$

Тоді пік сигналу можна визначити аналізуючи такі умови:

$$x'[n - 1] > 0, x'[n] \leq 0 \quad (1.5)$$

Ключовим аспектом аналізу часових характеристик сигналу є використання кореляційного аналізу, зокрема, визначення автокореляції сигналу, що дозволяє виділяти його періодичні компоненти та виявляти поведінкові шаблони в часі, якщо такі існують. Автокореляційну функцію встановлюємо в такий спосіб [32]:

$$R_{xx}[k] = \frac{1}{N - k} \sum_{n=0}^{N-k-1} x[n] \cdot x[n + k], k = \overline{0, N - 1} \quad (1.6)$$

На наступному кроці перейдемо до аналізу сигналу в частотній області. Для цього використовується перетворення Фур'є. Розглянемо дискретне перетворення Фур'є в такому вигляді [32]:

$$X[k] = \sum_{n=0}^{N-1} x[n] \cdot e^{-\frac{j2\pi}{N}kn}, k = \overline{0, N - 1} \quad (1.7)$$

де $X[k]$ – коефіцієнти спектру сигналу, що характеризують його амплітудно-частотну складову.

Цей метод є відповідним для використання, проте при аналізі складних сигналів рекомендується його комбінувати з іншими підходами, адже можливі неточності через втрату деталей про часовий розподіл сигналу. В якості компромісу пропонується використовувати аналіз спектральної щільності сигналу в парі з перетворенням Фур'є, оскільки ця комбінація дозволяє детально оцінити розподіл сигналу та виявити його ключові точки [32]:

$$P_{xx}(f) = \frac{1}{L} \sum_{l=1}^L |X_l(f)|^2 \quad (1.8)$$

де $X_l(f)$ – спектральне представлення l -го сегменту, отримане за допомогою перетворення Фур'є.

При обробці сигналів часто виникає потреба у фільтрації, як високочастотній, так і низькочастотній, для видалення елементів, які не мають значення для аналізу, а лише спричиняють небажану варіабельність. Оглянемо фільтр у загальному вигляді [32]:

$$y[n] \sum_{k=0}^M b[k] \cdot x[n-k] - \sum_{k=1}^N a[k] \cdot y[n-k] \quad (1.9)$$

де $b[k], a[k]$ – коефіцієнти фільтра;

M, N – порядки фільтрації.

На останньому етапі розглянемо методи часово-частотного аналізу сигналів, які є найбільш придатними для роботи з нестационарними сигналами, що часто зустрічаються при зборі даних з сенсорів пацієнтів з різноманітними часовими характеристиками та шумами. Основним і найефективнішим є короткочасне перетворення Фур'є, яке не має недоліків дискретного перетворення завдяки врахуванню часового розподілу сигналу через використання вікна перетворення, що дозволяє аналізувати сигнал на різних інтервалах [32]:

$$STFT\{x\}(t, w) = \int_{-\infty}^{+\infty} x(\tau) \cdot w(\tau - t) \cdot e^{-jw\tau} d\tau \quad (1.10)$$

де $w(\tau - t)$ – віконна функція, що звужує область застосування перетворення.

Застосування такої стратегії дозволяє аналізувати спектральні особливості сигналу у кожному конкретному відрізку, незалежно від його довжини. Проте, у випадку наявності локальних особливостей сигналу, обумовлених його попередньою поведінкою, рекомендується додатково використовувати вейвлет-перетворення разом з короточасним перетворенням Фур'є для ідентифікації характерних особливостей даних, що є важливим для аналізу причин виникнення тремору [32]:

$$CWT_x(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} x(t) \psi^* \left(\frac{t - b}{a} \right) dt \quad (1.11)$$

де a – коефіцієнт масштабування;

$\psi(t)$ – вейвлет-функція;

$\psi^*(\cdot)$ – комплексно спряжена функція до $\psi(t)$;

b – часовий зсув.

Узагальнюючи процес аналізу інерційних даних, доцільно представити його у вигляді блок-схеми (рис. 1.2):



Рисунок 1.2 – Послідовність застосування методів аналізу інерційних даних

Складність використання обговорених методів часто пов'язана з різноманітністю сигналів у досліджуваний вибірці, в таких випадках рекомендується вдаватися до різних стратегій обробки окремих частин датасету, які відрізняються одна від одної, а саме – здійснити класифікацію даних спочатку. Крім того, ідентифікація ключових елементів сигналів може спричинити значне горизонтальне збільшення обсягу вхідних даних, що в подальшому призведе до зростання складності аналізу та обробки даних за допомогою алгоритмів машинного навчання, тому важливо ефективно використовувати обговорені методи тільки в тих випадках, коли це дійсно необхідно, оскільки створення великої кількості характеристик даних не завжди веде до покращення результатів аналізу.

1.3 Моделі машинного навчання

Значна частина досліджень щодо хвороби Паркінсона базується на використанні інтелектуальних підходів, зокрема методів машинного навчання, які застосовуються для обробки даних, зібраних за допомогою інерційних датчиків. Це пов'язано з тим, що складність і структура цих даних зазвичай не дозволяють виконати аналіз за допомогою традиційних аналітичних методів. Використання технік машинного навчання дозволяє автоматично виявляти закономірності у розподілі даних та ідентифікувати фактори, які мають найбільший вплив на розвиток тремору та хвороби Паркінсона.

Архітектура Feedforward Neural Network передає інформацію від вхідного шару до вихідного без використання механізмів зворотного розповсюдження помилок. Таким чином, активація вхідних сигналів відбувається через нелінійну функцію. Незважаючи на свою простоту, ця модель часто виявляється найточнішою для багатьох задач, оскільки вона дозволяє моделювати нелінійні залежності без ризику перенавчання, який може виникати при роботі зі складними даними [8].

Нехай $x \in \mathbb{R}^d$ – вхідний вектор ознак, L – кількість шарів моделі та для кожного шару $l \in L$ параметри задаються у вигляді матриці вагових коефіцієнтів $W^{(l)}$ та вектором зміщення $b^{(l)}$. Тоді для довільного шару пряме поширення можна задати таким чином [8]:

$$\begin{aligned} z^{(l)} &= W^{(l)} \cdot a^{(l-1)} + b^{(l)} \\ a^{(l)} &= \sigma(z^{(l)}) \end{aligned} \quad (1.12)$$

де $a^{(0)}$ – вхідний сигнал;

$a^{(l)}$ – сигнал оброблений l -м шаром мережі;

$\sigma(\cdot)$ – нелінійна функція активації нейронів.

На виході з мережі отримується вектор $\hat{y} = a^{(L)}$, що вказує на ймовірність приналежності кожного спостереження одному із визначених класів.

Архітектуру даної мережі можна представити у вигляді схеми (рис. 1.3).

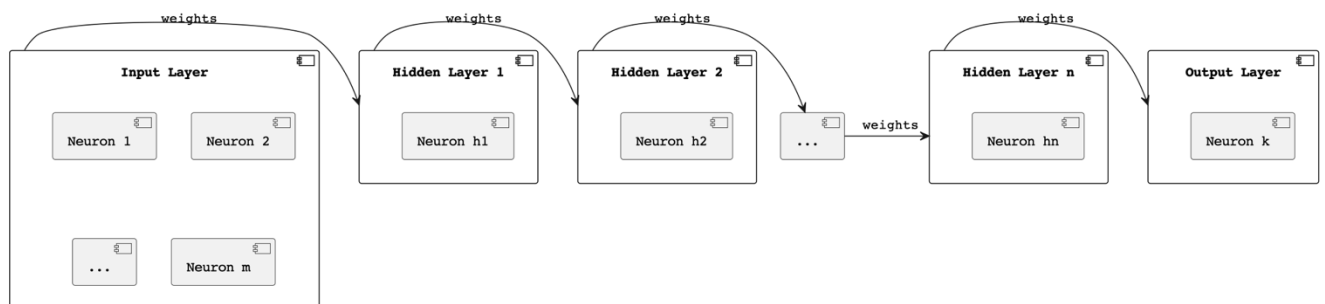


Рисунок 1.3 – Архітектура мережі FNN

Важливими є глибокі нейронні мережі, які покращують можливості мереж FNN завдяки включенню більшої кількості прихованих шарів. Це надає моделям здатність краще апроксимувати дані, узагальнювати закономірності та розпізнавати складні залежності у даних. Глибока архітектура сприяє ефективнішому аналізу взаємозв'язків між вхідними ознаками. Математична модель цих мереж формулюється подібно до моделі FNN, але включає більше шарів, що дозволяє використовувати різноманітні функції активації для виявлення різних залежностей між даними, а також для пошуку закономірностей у розподілі

даних та ознак [16]. Архітектура цих моделей є ідентичною тій, що представлена на рис. 1.3.

Використання ядерних функцій дозволяє більш гнучко та точно формувати гіперплощину, збільшуючи розмірність ознак. Цей метод може забезпечити високу точність та якість класифікації в деяких задачах, але при великій кількості ознак може призвести до вибору неоптимальних рішень або навіть до неможливості роботи алгоритму.

Нехай x_i – вхідні дані, $y_i \in \{-1, 1\}$ – мітки класів, тоді алгоритм вирішує задачу наступним чином [5]:

$$\min_{w,b} \frac{1}{2} \|w\|^2$$

$$y_i(w^T x_i + b) \geq 1, \forall i \quad (1.13)$$

де $\|\cdot\|$ – норма в Евклідовому просторі.

Архітектуру даної моделі доцільно представити таким чином (рис. 1.4):

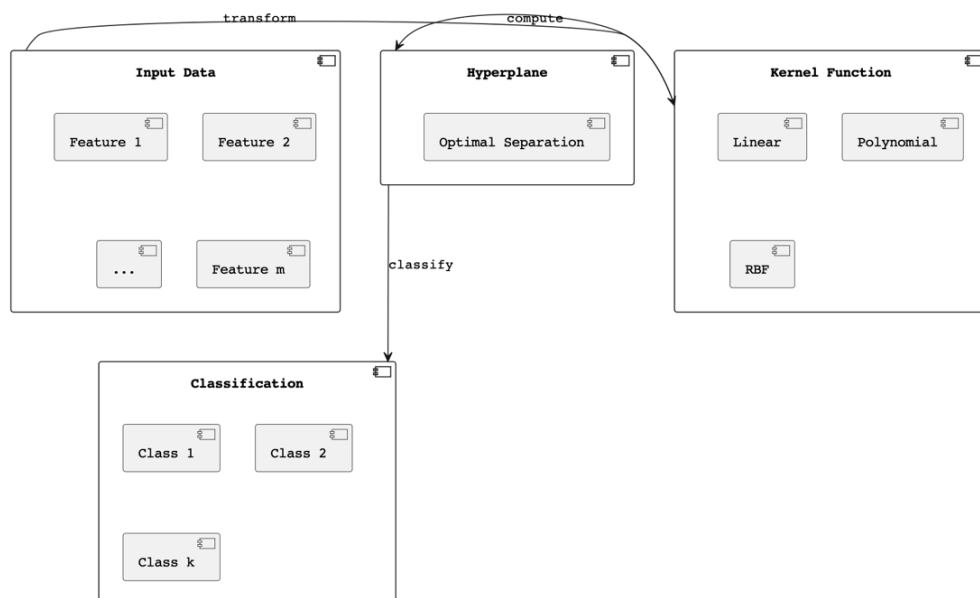


Рисунок 1.4 – Архітектура моделі SVM

Ансамблевий метод класифікації Random Forest працює за рахунок створення множини дерев прийняття рішень, де кожне дерево навчається на заданій підмножині вхідних даних, а процес класифікації фінальних ознак

відбувається за рахунок голосування, тобто підрахунку кількості міток кожного класу для кожного спостереження

Нехай $\{h_1(x), h_2(x), \dots, h_M(x)\}$ – множина дерев, тоді класифікація відбувається за таким правилом [23]:

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \dots, h_M(x)\}$$

Ансамблеву архітектуру можна представити у вигляді діаграми (рис. 1.5):

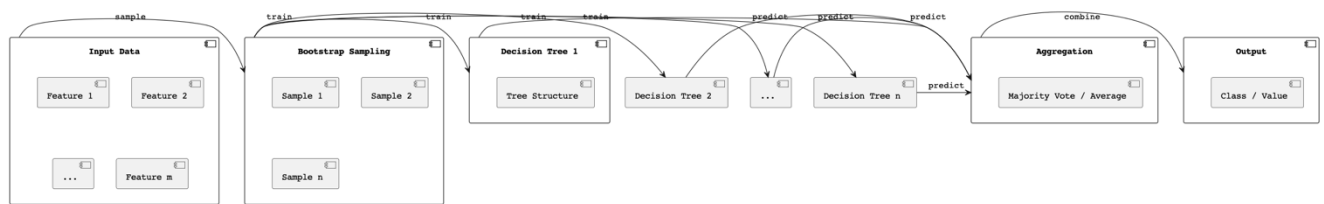


Рисунок 1.5 – Архітектура моделі Random Forest

Long Short-Term Memory призначена для аналізу даних, що представлені у вигляді послідовностей. Моделі оперують сховищами, які використовуються як короткочасна пам'ять моделі, для встановлення складних залежностей між даними й моделювання часових закономірностей розподілу даних. Модель для довільного часу t обчислює таку послідовність величин [33]:

Забувальна складова:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (1.14)$$

Вхідні складові:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (1.15)$$

$$\tilde{c}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (1.16)$$

Оновлення стану:

$$C_t = f_t \cdot C_{t-1} + i_t \cdot \tilde{C}_t$$

(1.17)

Вихідні стани:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$
$$h_t = o_t \cdot \tanh(C_t)$$

(1.18)

де x_t – вхідний сигнал;

h_t – прихований стан;

C_t – стан пам’яті;

$\sigma(z) = \frac{1}{1+e^{-z}}$ – симоїдальна функція активації, що діє як відображення такого вигляду: $\sigma(\cdot): \mathbb{R} \rightarrow [0, 1]$.

Представимо архітектуру моделі у вигляді діаграми (рис. 1.6).

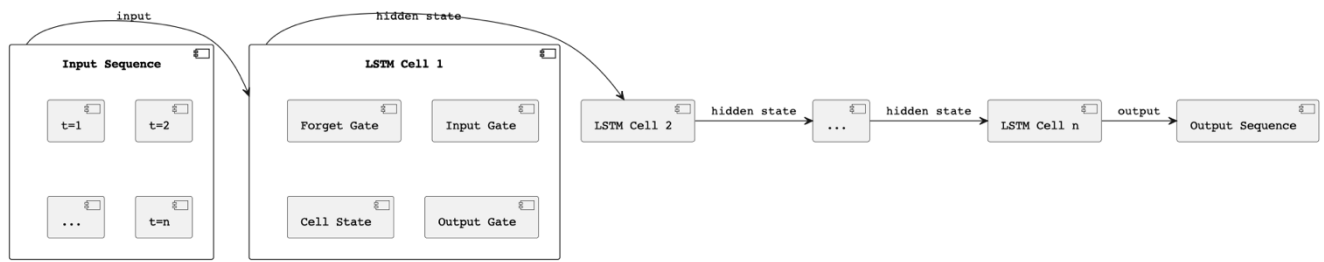


Рисунок 1.6 – Архітектура мережі LSTM

Порівняння розглянутих архітектур в контексті основних показників ефективності наведено в табл. 1.1.

Таблиця 1.1

Порівняння моделей машинного навчання

Модель	Складність архітектури	Час навчання	Інтерпретованість результатів	Необхідний обсяг вхідних даних	Моделювання нелінійних закономірностей
FNN					
DNN					
SVM					
RF					
LSTM					

Як можна побачити з табл. 1.1, моделі FNN та DNN мають найкращі характеристики із розглянутих методів.

Висновки до розділу 1

У першому розділі проведено аналіз теоретико-методологічних основ дослідження хвороби Паркінсона та методів діагностики тремору.

Встановлено, що тремор спокою з частотою 4-6 Гц є одним із найраніших діагностичних маркерів хвороби Паркінсона, причиною якого є дефіцит дофаміну через дегенеративні зміни в нейронах чорної речовини.

Виявлено основні обмеження традиційної клінічної діагностики: суб'єктивність оцінки, складність диференційної діагностики та відсутність специфічних біомаркерів. Єдиним достовірним методом підтвердження діагнозу залишається посмертне гістологічне дослідження.

Проаналізовано сучасні інноваційні підходи, включаючи використання носимих пристроїв з інерційними датчиками, методи нейровізуалізації та застосування штучного інтелекту, які показують значний потенціал для підвищення точності діагностики.

Систематизовано методи аналізу інерційних даних у часовому, частотному та часово-частотному діапазонах. Встановлено важливість статистичних характеристик сигналу для кількісної оцінки тремору.

Досліджено моделі машинного навчання, зокрема глибокі нейронні мережі, які дозволяють виявляти складні закономірності в інерційних даних та забезпечують високу точність класифікації тремору.

Виявлено основні методологічні виклики: гетерогенність даних, наявність шумів, низькочастотний характер тремору та обмеженість обсягів медичних даних.

РОЗДІЛ 2

МЕТОДИКА РЕАЛІЗАЦІЇ МОДЕЛІ ПРОГНОЗУВАННЯ

2.1 Опис набору даних

Вивчена вибірка містить дані про інерційні вимірювання, отримані з акселерометрів мобільних телефонів пацієнтів. Головною перевагою цього набору даних є його збір у реальному часі протягом визначеного періоду, що забезпечує природність поведінки учасників дослідження та збільшує репрезентативність результатів. Збір даних проводився з метою ідентифікації симптомів тремору, які можуть вказувати на розвиток хвороби Паркінсона. Для створення об'єктивного дослідження були обрані як здорові особи, так і пацієнти з діагнозом хвороби Паркінсона, зокрема 31 особа з патологією та 14 здорових. Такий метод дозволяє здійснювати порівняльний аналіз та виявляти статистично значущі відмінності між групами [4]. Основною проблемою аналізу таких даних є обмежений розмір вибірки, особливо з урахуванням кількості учасників, що може призвести до неточностей у результатах та зниження репрезентативності моделі.

Вхідний набір даних складається зі списку словників у форматі `pickle`, де кожен словник детально описує одного пацієнта. `subject_id` є числовим ідентифікатором, який унікально визначає кожного учасника дослідження. Цей ідентифікатор дозволяє точно відстежувати дані, пов'язані з конкретною особою, та проводити аналіз між різними учасниками. `subject_sessions` включає список масивів даних (`numpy.array`), кожен з яких представляє окрему сесію запису даних з акселерометра. Кожен масив має розмір $(N, 4)$, де: Колонка 0 містить часові мітки (`timestamp`) для кожного запису, що дозволяє слідкувати за динамікою сигналу в часі.

Колонки 1-3 відображають величину прискорення у трьох вимірах (x, y, z). Ці дані надають повну картину рухової активності учасника в просторі; Застосування тривимірного запису забезпечує точний аналіз рухових моделей, оскільки різні напрямки можуть містити унікальні характеристики тремору. Різноманітність тривалості сесій відображає варіативність умов збору даних, що

вимагає подальшої стандартизації та нормалізації; `session_datetimes` містить список об'єктів типу `datetime`, які відповідають датам та часам кожної сесії запису даних.

Наявність часових міток дозволяє відновити контекст записів, дослідити сезонні або добові коливання активності, а також аналізувати залежності від часу доби чи тривалості участі у дослідженні. Це критично важливо для визначення динаміки розвитку захворювання та ідентифікації потенційних закономірностей у зміні симптомів; `annotation` містить словник з анотаціями, орієнтованими на тремор, отриманими як від учасника дослідження, так і від лікарів чи експертів.

`updrs16` представляє ціле число, що відображає оцінку тремору згідно з пунктом 16 частини II шкали MDS-UPDRS, визначену самооцінкою пацієнта. Це дозволяє отримати суб'єктивну оцінку інтенсивності тремору, що є важливим для порівняння з об'єктивними даними.

`updrs20_right` та `updrs20_left` це цілі числа в діапазоні $[0,4]$, які вказують на рівень тремору у стані спокою правої та лівої руки відповідно, згідно з оцінкою лікаря за пунктом 20 частини III MDS-UPDRS. Окрема оцінка для кожної руки дозволяє детально аналізувати асиметрію проявів тремору, що є ключовим клінічним показником

`updrs21_right` та `updrs21_left` включають цілі числа в діапазоні $[0,4]$, що описують рівень тремору під час дії або постурального утримання правої та лівої руки відповідно, згідно з оцінкою лікаря, пункт 21 частини III MDS-UPDRS. Ці значення допомагають оцінити вплив дії чи постурального навантаження на інтенсивність тремору, що сприяє розрізненню різних типів рухових розладів; `sp_expert` це бінарне значення (0 або 1), встановлене експертами на основі аналізу сигналів. Значення 1 вказує на наявність тремору, що є ключовим для діагностики та моніторингу хвороби., наявність тремору позначається як 1, а його відсутність – як 0.

`pd_status` також є бінарним індикатором (0 або 1), який показує, чи є індивід хворим на Паркінсонівську хворобу (1) або ні (0). Процес розподілу даних починається з перевірки повноти даних, визначаючи кількість пропущених

значень, адже їх присутність може суттєво вплинути на точність майбутнього аналізу (рис. 2.1).

Data columns (total 9 columns):			
#	Column	Non-Null Count	Dtype
0	subject_id	45 non-null	int64
1	pd_status	45 non-null	int64
2	updrs16	45 non-null	int64
3	updrs20_right	45 non-null	int64
4	updrs20_left	45 non-null	int64
5	updrs21_right	45 non-null	int64
6	updrs21_left	45 non-null	int64
7	sp_expert	45 non-null	int64
8	num_sessions	45 non-null	int64

Рисунок 2.1 – Перевірка невизначених значень в полях датасету

Також дослідимо наявність пропущених значень в усіх полях набору даних (рис. 2.2).

ВІДСУТНІ ЗНАЧЕННЯ:	
subject_id	0
pd_status	0
updrs16	0
updrs20_right	0
updrs20_left	0
updrs21_right	0
updrs21_left	0
sp_expert	0
num_sessions	0
dtype: int64	

Рисунок 2.2 – Кількість пропущених значень в полях датасету

З рис. 2.2 видно, що у датасеті немає відсутніх значень та невизначених типів, що покращує ситуацію з попередньою обробкою даних та дозволить опрацьовувати всі наявні записи. Дослідження статистичних характеристик набору даних, з метою виявлення характеру їх розподілу та визначення можливих залежностей між даними показано на рис. 2.3.

БАЗОВА СТАТИСТИКА:									
	subject_id	pd_status	updrs16	updrs20_right	updrs20_left	updrs21_right	updrs21_left	sp_expert	num_sessions
count	45.000000	45.000000	45.000000	45.000000	45.000000	45.000000	45.000000	45.000000	45.000000
mean	23.000000	0.688889	0.777778	0.244444	0.533333	0.244444	0.422222	0.355556	121.355556
std	13.133926	0.468179	0.901738	0.608857	0.967659	0.608857	0.865734	0.484090	164.265466
min	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	5.000000
25%	12.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	13.000000
50%	23.000000	1.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000	38.000000
75%	34.000000	1.000000	1.000000	0.000000	1.000000	0.000000	1.000000	1.000000	135.000000
max	45.000000	1.000000	4.000000	3.000000	3.000000	3.000000	4.000000	1.000000	535.000000

Рисунок 2.3 – Описова статистика полів датасету

З рис. 2.3. можна помітити, всі анотації мають схожі усереднені значення, що може свідчити про високу точність їх визначення, у процесі дослідження, а також варто зазначити, що кількість сесій доволі сильно варіюється у пацієнтів, що, однозначно, вимагатиме додаткового етапу обробки даних, а саме стандартизації значень даного поля, хоча у 75% пацієнтів кількість сесій, в середньому, однакова, враховуючи той факт, що кількість пацієнтів, вкрай невелика, 25% пацієнтів, що мають набагато більшу кількість сесій можуть або спотворити результати аналізу, або вплинути на точність роботи отриманої моделі.

Необхідно дослідити відсотковий розподіл пацієнтів з контрольної групи та з групи з патологією на факт наявності тремору, це дозволить зрозуміти чи присутня залежність між тремором та хворобою Паркінсона, що, в свою чергу, полегшить подальший аналіз та відбір важливих ознак для побудови моделі (рис. 2.4).

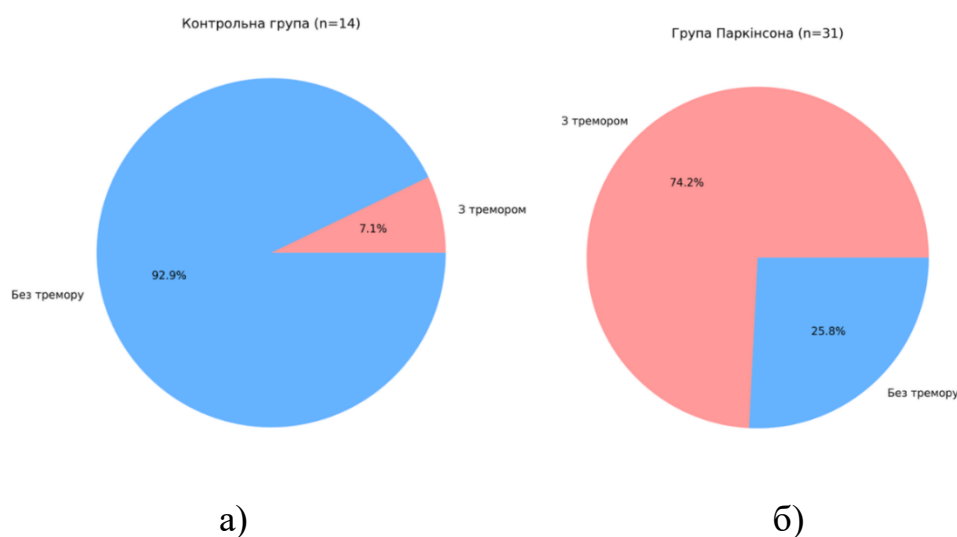


Рисунок 2.4 – Розподіл кількості пацієнтів з тремором між контрольною групою та групою хворих

Дійсно видно, що здорових пацієнтів з тремором практично немає, однак 75% пацієнтів з патологією страждають від тремору, тобто є підстави стверджувати, що наявність тремору, із заданим рівнем ймовірності, вказує на хворобу Паркінсона. Отже, стає можливим застосування частотної обробки сигналів, з метою виявлення закономірності появи тремору у пацієнтів з хворобою Паркінсона.

Проведемо більш деталізоване дослідження статистичного розподілу кількості сесій між досліджуваними. Нерівномірність розподілу значень даного поля може призводити до некоректних результатів дослідження, а отже й унеможливить узагальнення специфіки роботи побудованої моделі. Враховуючи, що кількість досліджуваних незначна, то для отримання репрезентативних, статистично значущих результатів, необхідна якісна попередня обробка даних з сенсорів, інакше вагове значення спостережень тих пацієнтів, що мають більшу частоту, може бути більшим, хоча, насправді, їх вплив не настільки значний як від значень решти пацієнтів. Додатково відобразимо розподіл значень поля кількості сесій (рис. 2.5).

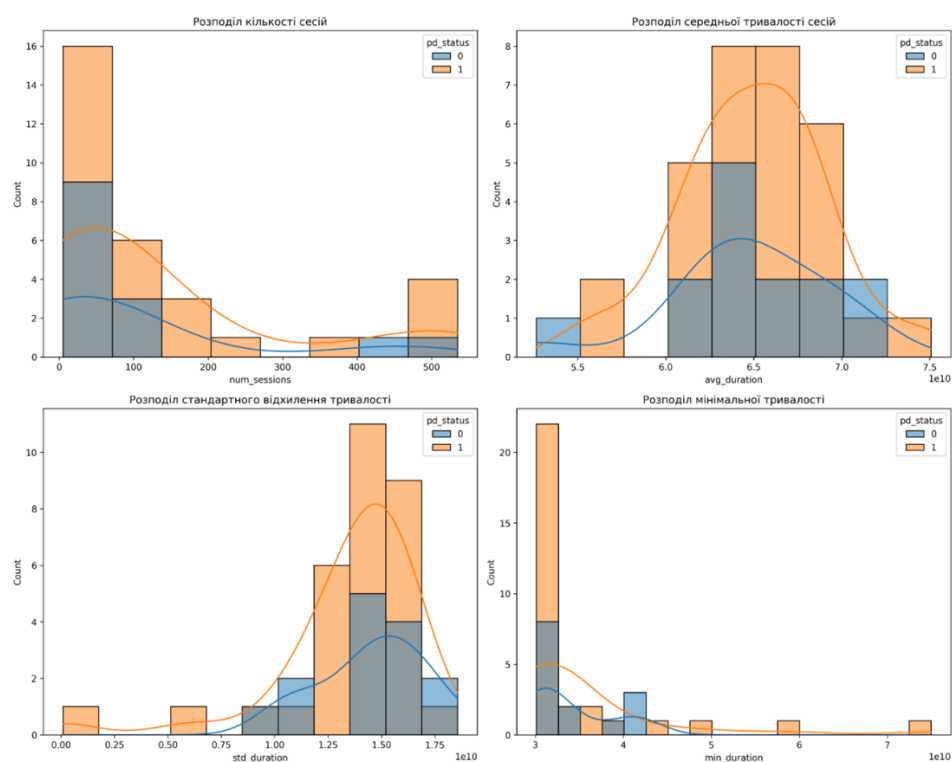


Рисунок 2.5 – Розподіл статистик кількості сесій

Як можна помітити з рис.2.5, характер розподілу даних досліджуваного поля в обох групах однаковий, отже усереднені показники теж мало відрізняються, однак видно, що велика кількість сесій доволі короткі, за тривалістю, а отже це викликає труднощі з подальшим моделюванням, оскільки необхідно визначати важливі характеристичні ознаки за короткий проміжок часу.

Враховуючи нерівномірність природи розподілу кількості сесій між пацієнтами, було прийняте рішення про дослідження поведінки полів анотацій, що вказують на оцінку рівня тремору у пацієнтів. Цей підхід дозволить знайти перетин між кількістю сесій та значеннями анотацій, що дозволить мінімізувати негативний вплив від нерівномірного розподілу значень поля, що описує кількість сесій пацієнтів (рис. 2.6).

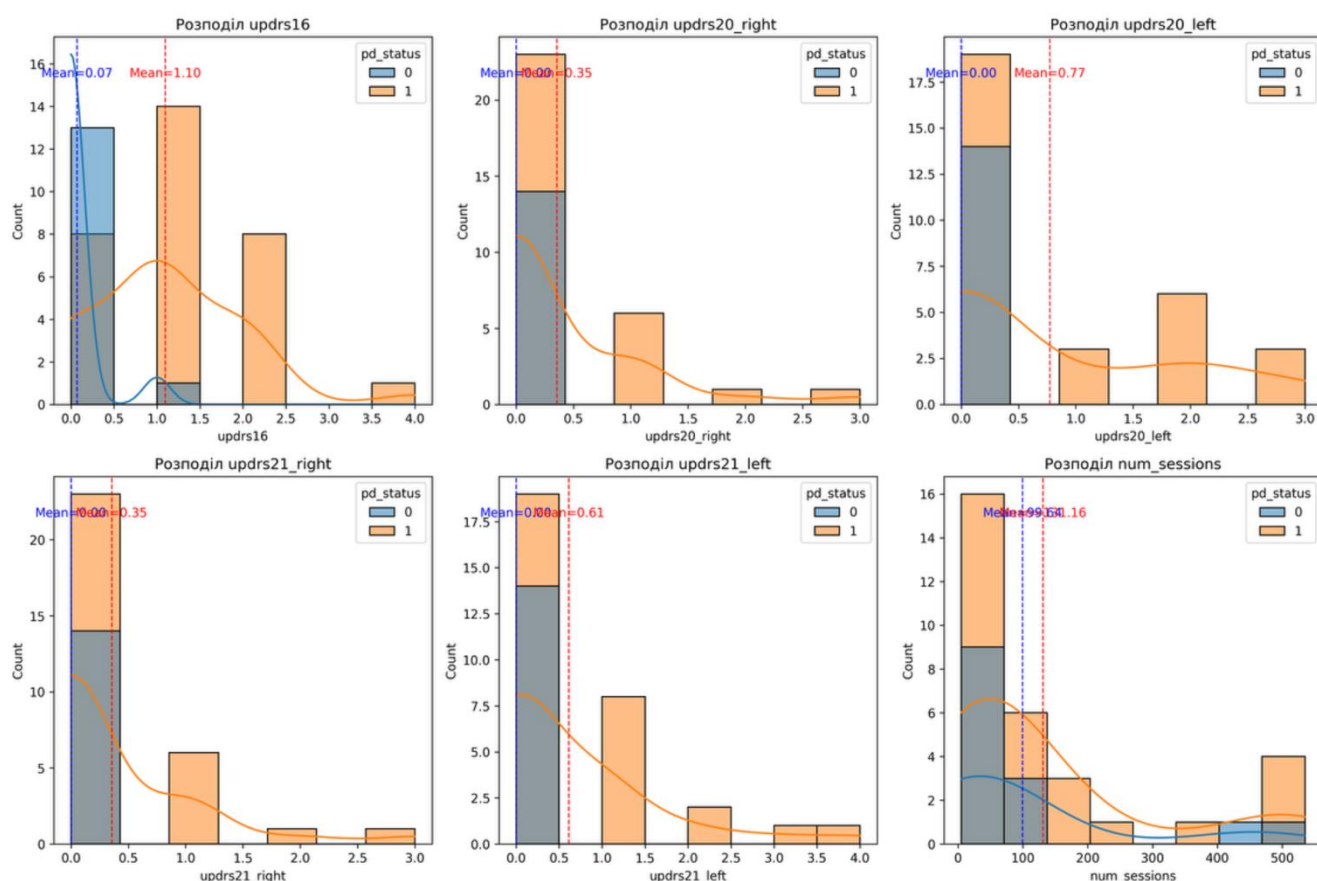


Рисунок 2.6 – Розподіл значень анотацій

З рис. 2.6 можна помітити, що анотації дійсно коректно вказують на наявність тремору у пацієнтів з патологією, що дозволяє аналізувати сигнали

окремо для кожної групи, без обмежень загальності. Звичайно, деякі спостереження з обох груп перетинаються, але причина цьому полягає в тому, що існують здорові пацієнти з тремором, проте їх кількість незначна.

Необхідно провести процес перевірки статистичних гіпотез про статистичну значущість у відмінності між значеннями кожної анотації у контрольній групі та групі хворих. Це дозволить зрозуміти чи дійсно є значущий вплив від перетину (накладання) значень цих анотацій, чи їх можна використовувати в моделі в якості групи індикаторів наявності тремору в пацієнта. Для цього було застосовано U-тест Мана-Уїтні [17] та тест Ст'юдента [13], тести проведені автоматизовано, шляхом написання відповідної програми мовою Python, зведемо дані в зручний формат(табл. 2.1).

Таблиця 2.1

Статистичний аналіз відмінностей між групами

Показник	updrs16		updrs20_right		updrs20_left		updrs21_right		updrs21_left		num_sessions	
	Здорові	Хворі	Здорові	Хворі	Здорові	Хворі	Здорові	Хворі	Здорові	Хворі	Здорові	Хворі
Середнє	0.071 4	1.098 6	0.000 0	0.354 8	0.000 0	0.774 2	0.000 0	0.354 8	0.000 0	0.612 9	99.6429	131.161 3
Стандартне відхилення	0.267 3	0.907 6	0.000 0	0.709 4	0.000 0	1.086 6	0.000 0	0.709 4	0.000 0	0.989 2	157.027 5	169.029 4
t-test	-5.7615		-2.7850		-3.9671		-2.7850		-3.4498		-0.6085	
p-value	0.0000		0.0092		0.0004		0.0092		0.0017		0.5479	
U-test	67.0000		161.0000		133.0000		161.0000		133.0000		169.500	
p-value	0.0001		0.0407		0.0084		0.0407		0.0082		0.2489	
Висновок	Значуща різниця		Значуща різниця		Значуща різниця		Значуща різниця		Значуща різниця		Незначуща різниця	

За результатами проведеного аналізу, стає очевидно, що кількість сесій у кожного пацієнта ніяк не впливатиме на результат, оскільки дана різниця не є статистично значущою. Навпаки, доцільно аналізувати саме значення анотацій, так як і з графіків, і з аналізу помітно, що вони з високою ймовірністю вказують на наявність тремору, а отже можуть бути використані як індикатор при класифікації між групами.

Стає доцільним дослідження статистичного зв'язку між ознаками, окрім можливості аналізу між групами, може з'явитись можливість побудови складних залежностей між групами ознак, а це може підвищити точність моделі. Відобразимо попарні діаграми розсіювання всіх ознак (рис. 2.7).

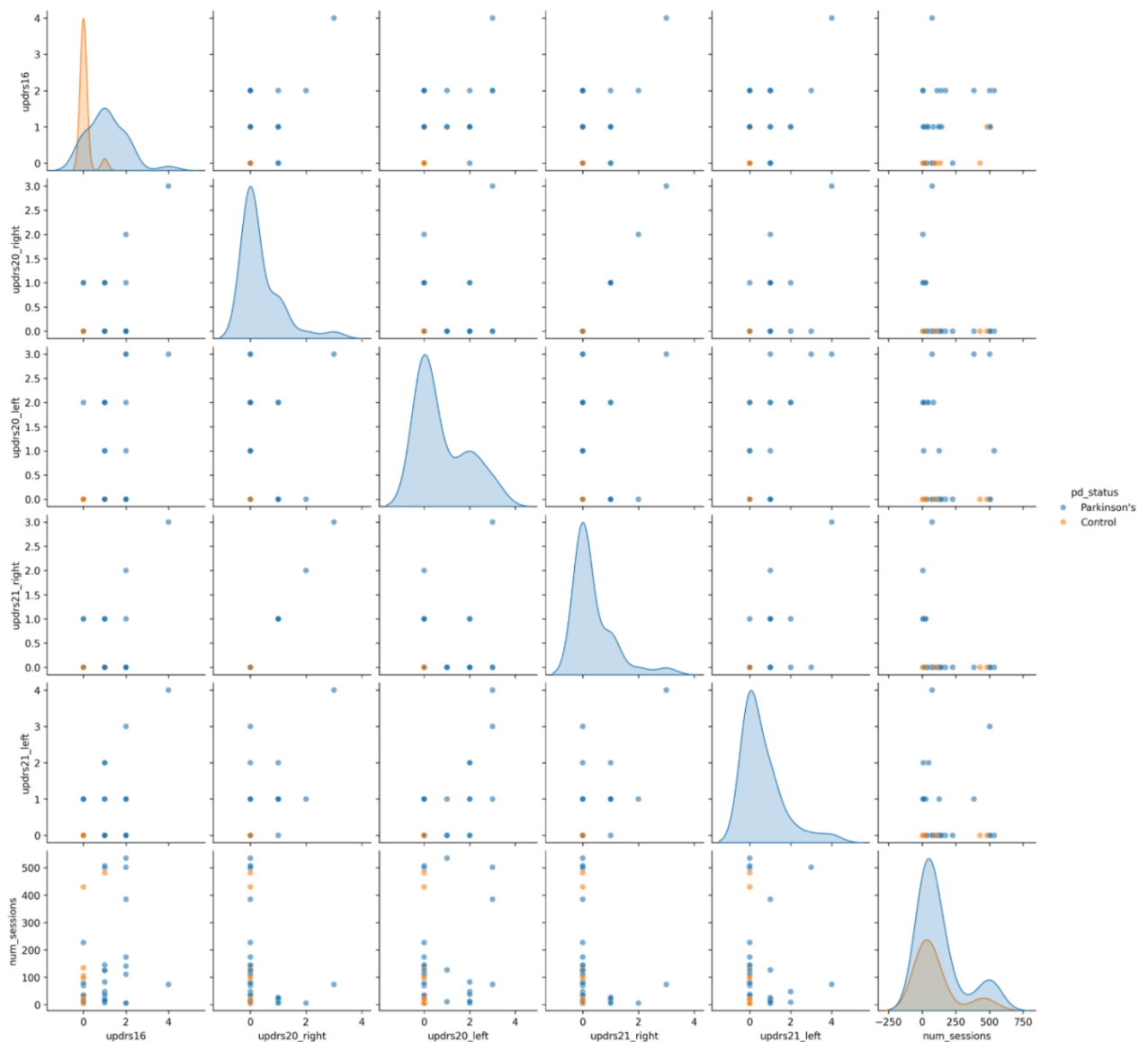


Рисунок 2.7 – Парні графіки розсіювання ознак датасету

Деякі ознаки показують взаємну залежність, тобто підозру на те, що поява тремору у пацієнта може відображатись не лише в одній з кінцівок, а супроводжуватись аналогічним явищем в протилежній кінцівці. Однак для отримання достовірних результатів та підтвердження даного спостереження, доцільно побудувати кореляційну матрицю між ознаками (рис. 2.8).

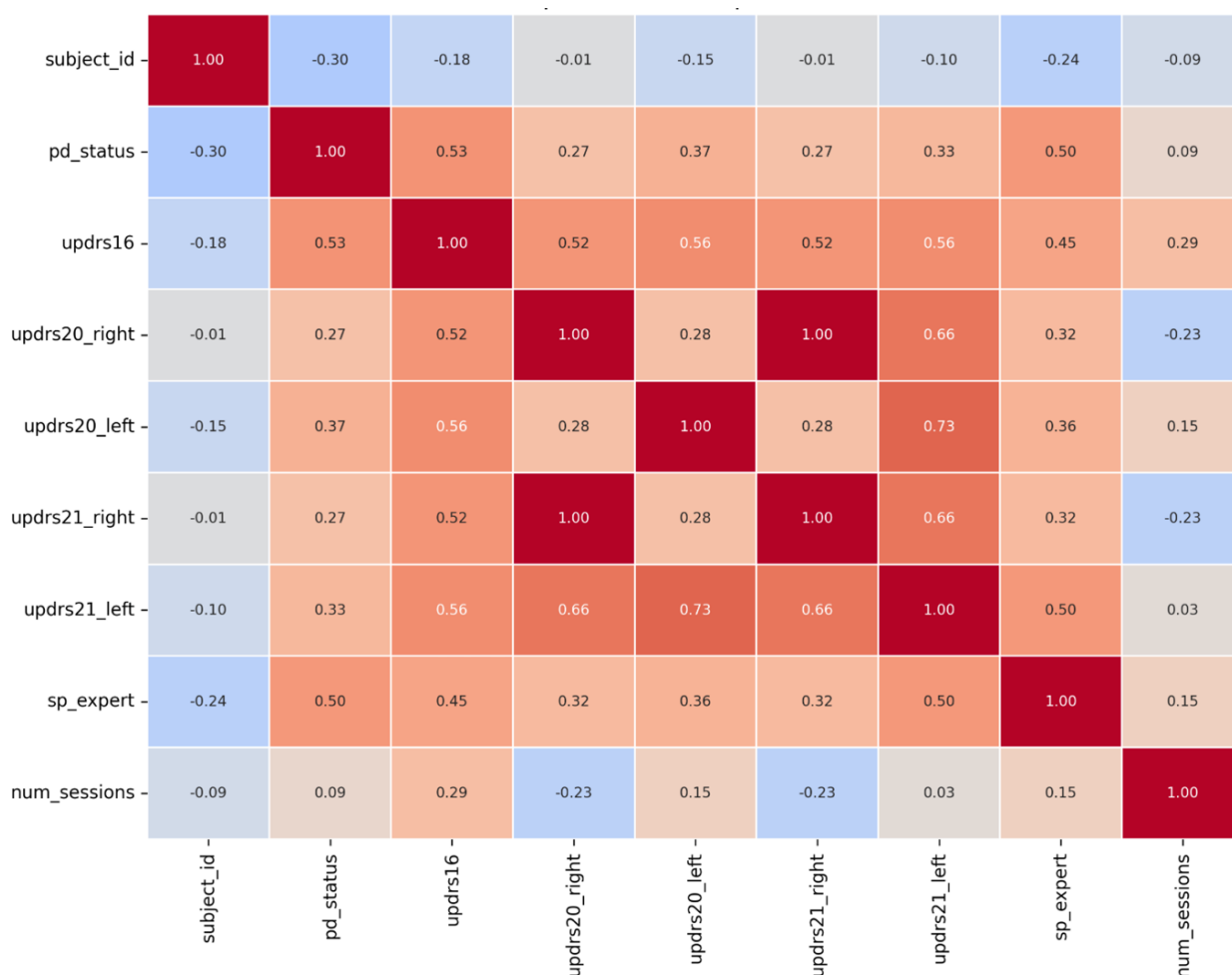


Рисунок 2.8 – Кореляційна матриця ознак датасету

Дійсно є доволі сильна залежність, у пацієнтів, між анотаціями, що вказують на ліві та праві кінцівки, тому, поява тремору в одній з кінцівок повинна супроводжуватись появою тремору в іншій кінцівці для діагностування хвороби Паркінсона.

2.2 Обґрунтування вибору моделей прогнозування

Для вирішення поставленої задачі, обрано дві основні архітектури – FNN та DNN. Базуючись на результатах таблиці 1.1, та враховуючи особливості вхідних даних, а саме їх складність та варіативність. Об'єм ознак змушує використовувати складні архітектурні рішення, з великою кількістю шарів, з метою побудови якісної моделі.

У якості першої архітектури було обрано FNN, яка поєднує в собі класичні повнозв'язні шари та методи регуляризації, що дозволяє даній архітектурі, практично, позбавитись ймовірності перенавчання – явища, що часто виникає на складних даних. Вибір даного архітектурного рішення обумовлений декількома факторами:

1. Залишкові з'єднання які дозволяють ефективно боротись з проблемою зникаючого градієнта у глибоких мережах, що критично для розпізнавання складних закономірностей у часових рядах, яким є дані акселерометра при треморі;
2. Комбінація регуляризаційних технік таких як використання L2-регуляризації, Dropout та BatchNormalization у комплексі дозволяє ефективно боротися з перенавчанням, що особливо важливо для медичних даних, де кількість зразків обмежена, а розмірність ознак висока;
3. LeakyReLU як функція активації на відміну від стандартного ReLU дозволяє пропускати малий градієнт через негативну частину, що запобігає проблемі «вмираючих нейронів» та покращує процес навчання;
4. Поступове зменшення розмірності при якому мережа зменшує розмірність від 512 до 1 нейрона, що дозволяє виділяти ієрархію абстрактних ознак різного рівня.

Для обраної моделі реалізовано архітектурне рішення (рис. 2.9).

```
def create_fnn_model(input_shape, dropout_rate=0.4):
    inputs = layers.Input(shape=input_shape)
    x = layers.Dense(512, kernel_regularizer=regularizers.l2(1e-4))(inputs)
    x = layers.BatchNormalization()(x)
    x = layers.LeakyReLU(alpha=0.1)(x)
    x = layers.Dropout(dropout_rate)(x)
    x = layers.Dense(256, kernel_regularizer=regularizers.l2(1e-4))(x)
    x = layers.BatchNormalization()(x)
    x = layers.LeakyReLU(alpha=0.1)(x)
    x = layers.Dropout(dropout_rate)(x)
    shortcut = x
    res = layers.Dense(256, kernel_regularizer=regularizers.l2(1e-4))(x)
    res = layers.BatchNormalization()(res)
    res = layers.LeakyReLU(alpha=0.1)(res)
    res = layers.Dense(256, kernel_regularizer=regularizers.l2(1e-4))(res)
    res = layers.BatchNormalization()(res)
    x = layers.Add()([shortcut, res])
    x = layers.LeakyReLU(alpha=0.1)(x)
    x = layers.Dense(128, kernel_regularizer=regularizers.l2(1e-4))(x)
    x = layers.BatchNormalization()(x)
    x = layers.LeakyReLU(alpha=0.1)(x)
    x = layers.Dropout(dropout_rate)(x)
    shortcut = x
    res = layers.Dense(128, kernel_regularizer=regularizers.l2(1e-4))(x)
    res = layers.BatchNormalization()(res)
    res = layers.LeakyReLU(alpha=0.1)(res)
    res = layers.Dense(128, kernel_regularizer=regularizers.l2(1e-4))(res)
    res = layers.BatchNormalization()(res)
    x = layers.Add()([shortcut, res])
    x = layers.LeakyReLU(alpha=0.1)(x)
    x = layers.Dense(64, kernel_regularizer=regularizers.l2(1e-4))(x)
    x = layers.BatchNormalization()(x)
    x = layers.LeakyReLU(alpha=0.1)(x)
    x = layers.Dropout(dropout_rate)(x)
    outputs = layers.Dense(1, activation='sigmoid')(x)
    model = models.Model(inputs=inputs, outputs=outputs)
    optimizer = Adam(learning_rate=0.001)
    model.compile(optimizer=optimizer, loss='binary_crossentropy', metrics=['accuracy'])
    return model
```

Рисунок 2.9 – Реалізація архітектури мережі FNN

Архітектура мережі складна та містить велику кількість шарів обробки інформації, такий вибір обумовлений структурою даних та незначною кількістю, тому процес обробки потрібно деталізувати й побудувати в процесі навчання якомога більше залежностей між ознаками. Додатково представлено розроблену архітектуру у вигляді зв'язків між структурними компонентами на рис. 2.10.

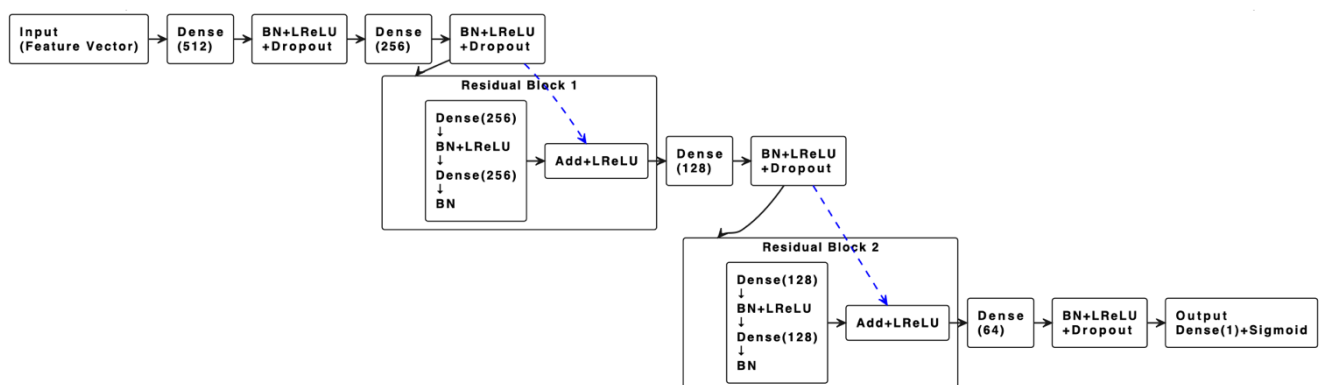


Рисунок 2.10 – Архітектура розробленої мережі FNN

Друга архітектура представляє собою більш класичну глибоку нейронну мережу з послідовним з'єднанням шарів. Такий вибір обґрунтований такими аспектами:

1. Простота та інтерпретованість. Послідовна архітектура без залишкових з'єднань є більш прозорою та легшою для інтерпретації, це важливо для медичних застосувань.
2. Відмінність від першої моделі
3. Пірамідальна структура. Полягає в зменшенні кількості нейронів ($256 \rightarrow 256 \rightarrow 128 \rightarrow 64 \rightarrow 1$) створює ієрархічну структуру, що дозволяє ефективно виділяти ознаки різного рівня абстракції;
4. Регуляризація та використання комбінації L2-регуляризації, Dropout та BatchNormalization. Це дозволяє ефективно боротися з перенавчанням, особливо в умовах обмеженої кількості даних.

Дана архітектура показана на рис. 2.11.

```
def create_dnn_model(input_shape, dropout_rate=0.4):
    model = models.Sequential([
        layers.Input(shape=input_shape),
        layers.Dense(256, activation='relu', kernel_regularizer=regularizers.l2(1e-4)),
        layers.BatchNormalization(),
        layers.Dropout(dropout_rate),
        layers.Dense(256, activation='relu', kernel_regularizer=regularizers.l2(1e-4)),
        layers.BatchNormalization(),
        layers.Dropout(dropout_rate),
        layers.Dense(128, activation='relu', kernel_regularizer=regularizers.l2(1e-4)),
        layers.BatchNormalization(),
        layers.Dropout(dropout_rate),
        layers.Dense(64, activation='relu', kernel_regularizer=regularizers.l2(1e-4)),
        layers.BatchNormalization(),
        layers.Dropout(dropout_rate),
        layers.Dense(1, activation='sigmoid')
    ])
    optimizer = Adam(learning_rate=0.001)
    model.compile(optimizer=optimizer, loss='binary_crossentropy', metrics=['accuracy'])
    return model
```

Рисунок 2.11 – Реалізація архітектури мережі DNN

Дана модель має простішу архітектуру, однак на обмежених даних дозволяє отримувати доволі точні результати, за рахунок логіки своєї роботи. Архітектура зв'язків в моделі представлена на рис. 2.12.

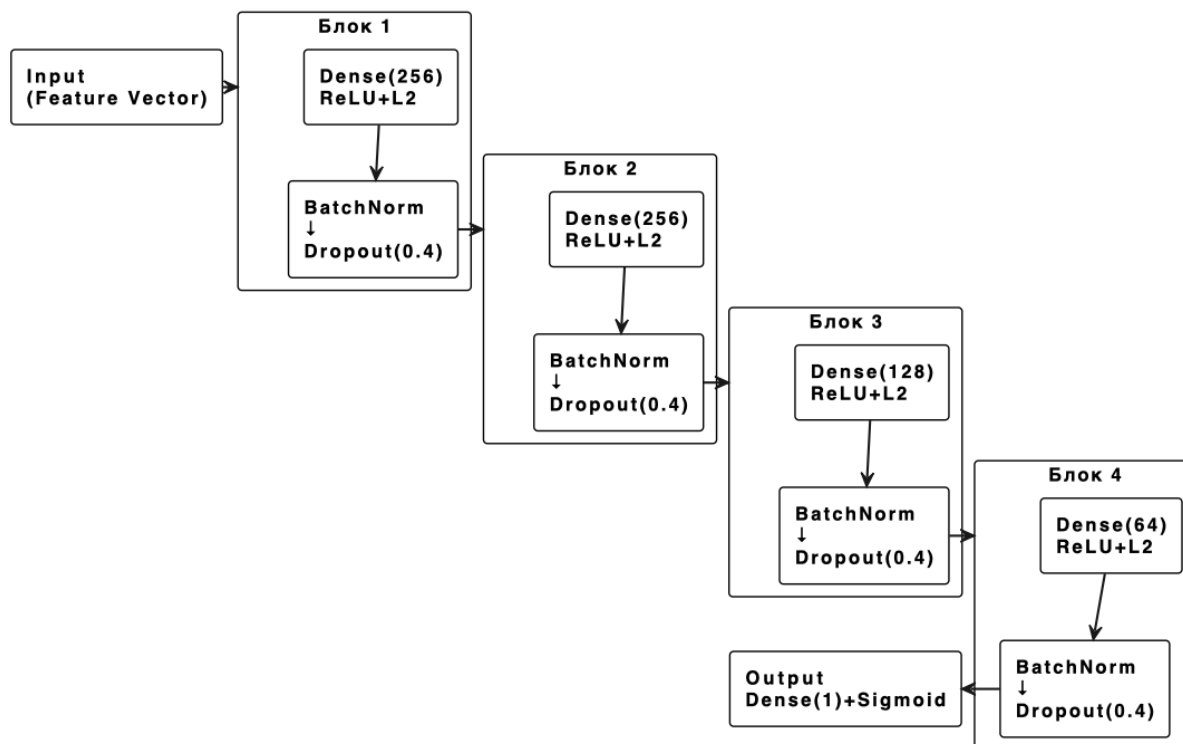


Рисунок 2.12 – Архітектура розробленої мережі DNN

Для побудови ансамблевої архітектури навчання були обрані дві моделі, що обумовлено складністю обробки послідовних даних і необхідністю виявлення глибоких залежностей між ними. Основною метою використання ансамблю є підвищення якості та надійності прогнозування.

Поєднання моделей дозволяє зменшити дисперсію результатів, адже різні архітектури схильні до різних типів помилок на окремих прикладах. Усереднення їхніх прогнозів допомагає компенсувати індивідуальні недоліки кожної з них. Такий підхід також підвищує стабільність результатів, оскільки комплексні методи менш чутливі до випадкової ініціалізації ваг і наявності шумів у даних, що особливо важливо у сфері медичної діагностики.

Ансамбль моделей, які мають різну архітектуру, здатен краще виявляти закономірності у вхідних даних, що позитивно впливає на здатність узагальнення

й забезпечує вищу точність класифікації нових зразків. Крім того, використання ансамблю сприяє зниженню ризику перенавчання, оскільки одночасна регуляризація кількох моделей на обмеженій вибірці даних дозволяє зберігати баланс між точністю та узагальненням.

Для ансамблювання використано метод усереднення прогнозів soft voting [28], який є одним із найпростіших, але водночас ефективним методом об'єднання моделей. Архітектуру ансамблевої моделі наведено на рис. 2.13.

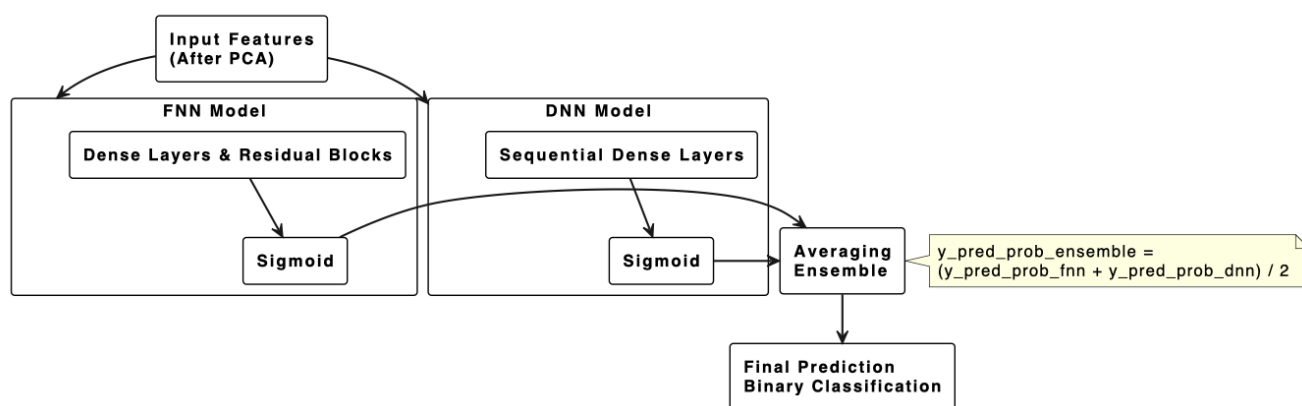


Рисунок 2.13– Архітектура ансамблевої моделі

Побудована архітектура дозволить досягти високої точності вирішення.

Висновок до розділу 2

У даному розділі було проведено детальний аналіз вхідних даних та обґрунтовано вибір архітектурних рішень для створення моделі прогнозування хвороби Паркінсона на основі інерційних вимірювань з акселерометрів мобільних пристроїв.

Аналіз структури даних показав наявність комплексної інформації, що включає тривимірні прискорення, часові мітки, клінічні оцінки за шкалою MDS-UPDRS та експертні анотації. Виявлено значну кореляцію між наявністю тремору та діагнозом хвороби Паркінсона, що підтверджує перспективність використання частотного аналізу сигналів для діагностичних цілей.

Запропоновано використання двох різних архітектур нейронних мереж: ResNet-подібної моделі з залишковими з'єднаннями та класичної послідовної архітектури з пірамідальною структурою. Такий підхід забезпечує баланс між ефективністю навчання глибоких мереж та інтерпретованістю результатів.

Обґрунтовано доцільність застосування ансамблевого методу з використанням soft voting для підвищення надійності та точності прогнозування. Ансамблювання дозволяє компенсувати індивідуальні недоліки окремих моделей та знизити ризик перенавчання на обмеженій вибірці даних.

Визначено основні виклики дослідження, зокрема обмежений розмір вибірки та нерівномірність розподілу кількості сесій між учасниками, що потребує додаткових етапів стандартизації та нормалізації даних для забезпечення статистичної значущості результатів.

РОЗДІЛ 3

ЗБІР ТА ОЦІНКА РЕЗУЛЬТАТІВ

3.1 Попередня обробка даних

Дані потребують якісної та детальної обробки, перед побудовою інтелектуальної моделі. У процесі дослідження було застосовано три етапи обробки та підготовки даних до моделювання: обробка даних акселерометрів; обробка та визначення ознак; підготовка даних до моделювання.

На першому етапі було проведено комплекс направлений на обробку сигналів, побудовано усереднені спектри потужності сигналів, у лінійному та логарифмічному масштабах. Логарифмічна функція досягає екстремумів на тих самих аргументах, що й лінійний аналог, однак дозволяє краще визначати відмінності між різними частотними діапазонами (рис. 3.1).

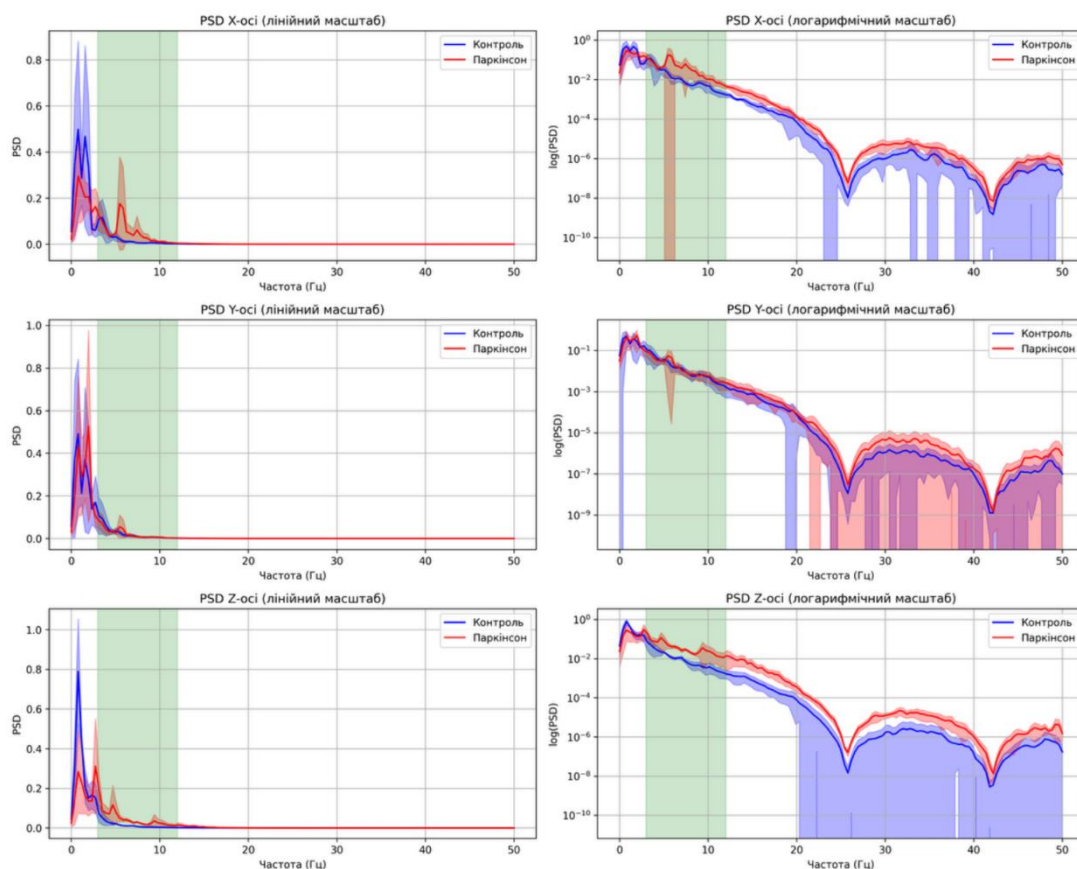


Рисунок 3.1 – Порівняльний аналіз частотних характеристик складових сигналів

Частотні особливості всіх трьох компонент сигналів схожі за своєю поведінкою, що може свідчити важливість значення при аналізі наявності тремору, особливо це видно з логарифмічної шкали, на якій частотний діапазон виникнення тремору проявляється майже однаково для всіх компонент. Важливо порівняти між собою дані акселерометра пацієнтів здорових та з патологією (рис. 3.2).

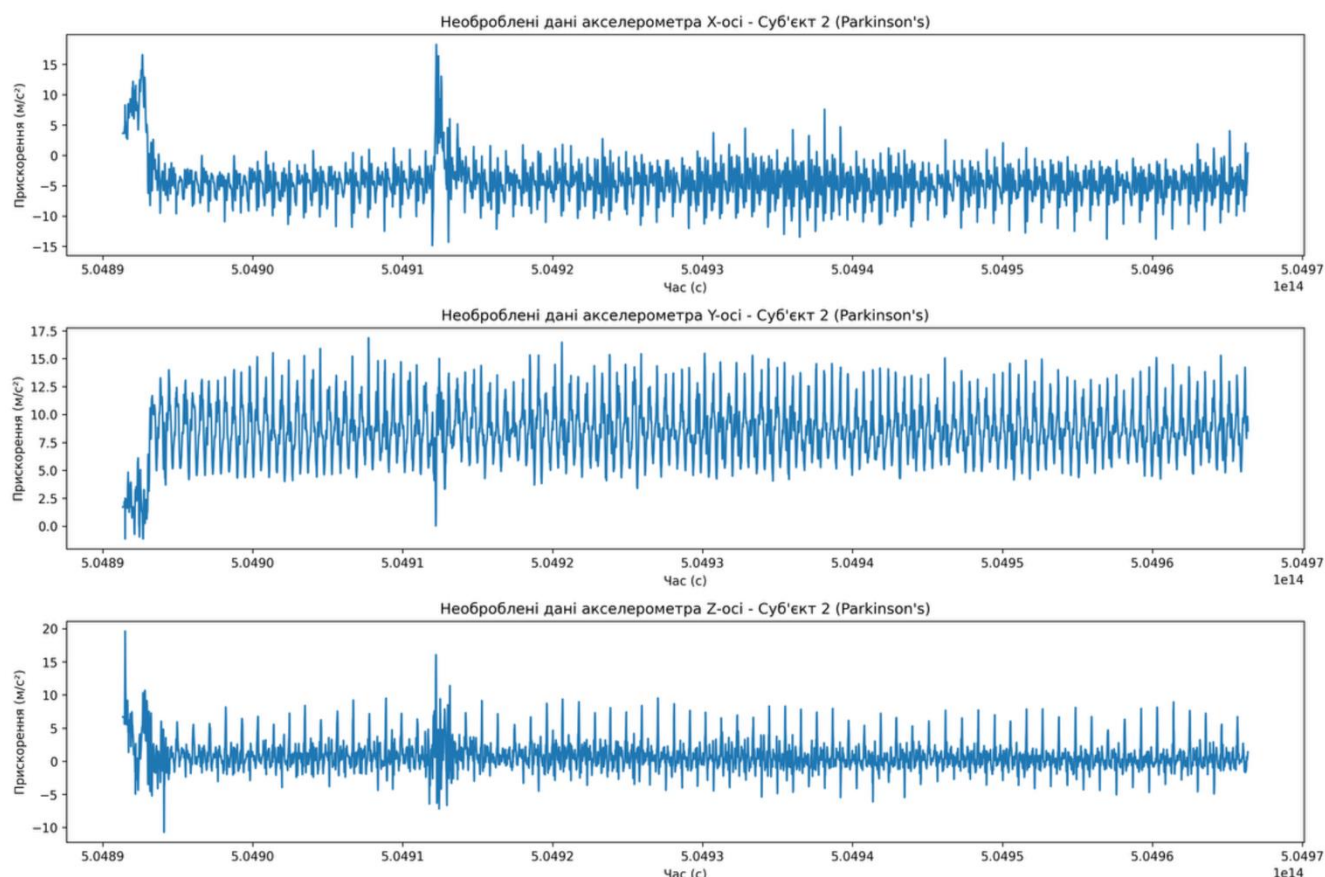


Рисунок 3.2 – Необроблені дані акселерометра пацієнта з хворобою Паркінсона

Даний пацієнт страждає від тремору, це проявляється у зміні прискорення руху кінцівок з високою періодичністю, а отже вказує на тремор в усіх трьох осьових складових. Далі, аналогічно, оберемо випадкового пацієнта з групи здорових та відобразимо дані його акселерометра (рис. 3.3).

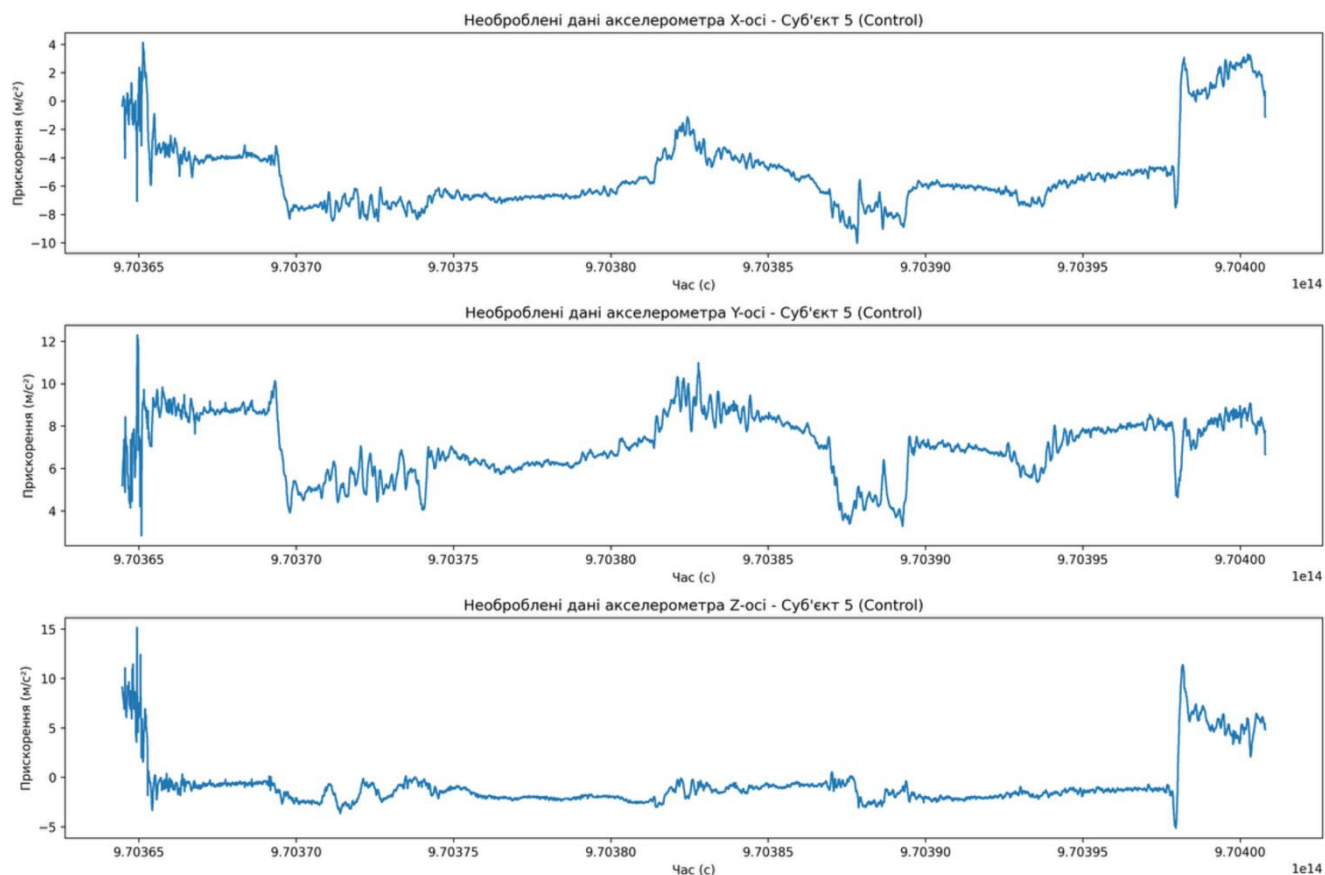


Рисунок 3.3 – Необроблені дані акселерометра здорового пацієнта

Різниця відчутна, у здорового пацієнта прискорення, постійне на більшості часових проміжків, а отже немає скорочень м'язів кінцівок, це свідчить про відсутність тремору.

Для частотної підготовки сигналів, на першому кроці було застосовано два фільтри: низькочастотний та високочастотний фільтр Баттерворта 4-го порядку [27]. Це дозволило зрізати ті частотні діапазони, що не несуть інформації про тремор, найбільш корисними є діапазон 3-12 Гц, а отже застосування двох фільтрів дозволило зменшити обсяг оброблюваних послідовностей даних. Окрім вирізання непотрібних частин сигналу, фільтрація частот дозволяє позбутися дрейфу в сигналі, який часто виникає у стані абсолютного спокою та усунення високочастотних шумів, які виникають під впливом електромагнітних перешкод і можуть негативно впливати на результати моделювання (рис.3.4).

```

for session in acc_data:
    cutoff_hp = 0.5 / nyq
    b, a = signal.butter(4, cutoff_hp, btype='high')
    filtered_data = signal.filtfilt(b, a, session[:, 1:4], axis=0)
    cutoff_lp = 20 / nyq
    b2, a2 = signal.butter(4, cutoff_lp, btype='low')
    filtered_data = signal.filtfilt(b2, a2, filtered_data, axis=0)

```

Рисунок 3.4 – Фільтрація сигналів датасету

Необхідне більше зменшення послідовності даних, оскільки більшість сигналів на початку та вкінці мають нульову частоту, тому, початок запису акселерометрів та його кінець можна видалити, без втрати інформації, необхідної для аналізу. Для цього було застосовано метод усічення сигналу [15] в діапазоні 5 секунд на обох кінцях, це і виконує певну нормалізацію даних (рис. 3.5).

```

trim_samples = int(5 * fs)
if len(filtered_data) > 2 * trim_samples:
    filtered_data = filtered_data[trim_samples:-trim_samples]

```

Рисунок 3.5 – Триммінг сигналів датасету

Вже після скорочення довжини часових рядів, доцільне застосування методів видалення викидів у сигналах. На попередніх кроках застосування цього підходу було б менш ефективним, оскільки тоді було б прибрано ті проміжки, які прибирались фільтрацією, а справжні викиди в потрібному частотному діапазоні залишились би, що спричинило б неточність в подальшій роботі моделі. Для фільтрації було застосовується метод Median Absolute Deviation [11], який є вкрай дієвим при обробці сигналів, тому що, не змінює структуру розподілу даних, а також обмежує вплив викидів й екстремальних значень часової послідовності, які можуть не являться закономірністю, а бути виключенням із правил. Така фільтрація працює не стільки за рахунок видалення спостережень, як за рахунок різкого обмеження їх впливу на поведінку сигналу (рис. 3.6).

```
med = np.median(filtered_data, axis=0)
mad = np.median(np.abs(filtered_data - med), axis=0)
threshold = 3
lower_bound = med - threshold * mad
upper_bound = med + threshold * mad
filtered_data = np.clip(filtered_data, lower_bound, upper_bound)
```

Рисунок 3.6 – Усунення викидів в сигналах датасету

Після видалення викидів зі спостережень можна перейти до згладжування отриманих сигналів, з метою ще більшої стандартизації останніх, оскільки фільтрація та очистка сигналу не здатні видаляти дрібні збурення в даних, було застосовано фільтрацію Савицького-Голея [1], яка здатна згладжувати сигнал не на всьому часовому проміжку, а на заданому вікні, що допомагає зберігати форму розподілу даних і, при цьому, зберігати важливі деталі в сигналах, які відіграють роль при визначенні явища тремору (рис. 3.7).

```
smoothed_data = savgol_filter(filtered_data, window_length=7, polyorder=2, axis=0)
```

Рисунок 3.7 – Згладжування даних у сигналах датасету

Фінальним етапом частотної обробки даних була стандартизація даних, оскільки обробка даних, значення яких знаходяться в різних діапазонах, може призвести до перенавчання та неточних результатів, тому з метою узагальнення характеристик, застосовано стандартизацію значень спостереження, тобто відцентровано навколо нуля, це забезпечує нормалізацію амплітуд сигналів (рис. 3.8).

```
scaler = StandardScaler()
standardized_data = scaler.fit_transform(smoothed_data)
standardized_data -= np.mean(standardized_data, axis=0)
time = np.arange(len(standardized_data)) / fs
processed_session = np.column_stack((time, standardized_data))
processed_data.append(processed_session)
```

Рисунок 3.8 – Стандартизація значень спостережень в датасеті

У результаті проведення частотної обробки даних, було отримано такі сигнали для хворого (рис. 3.9).

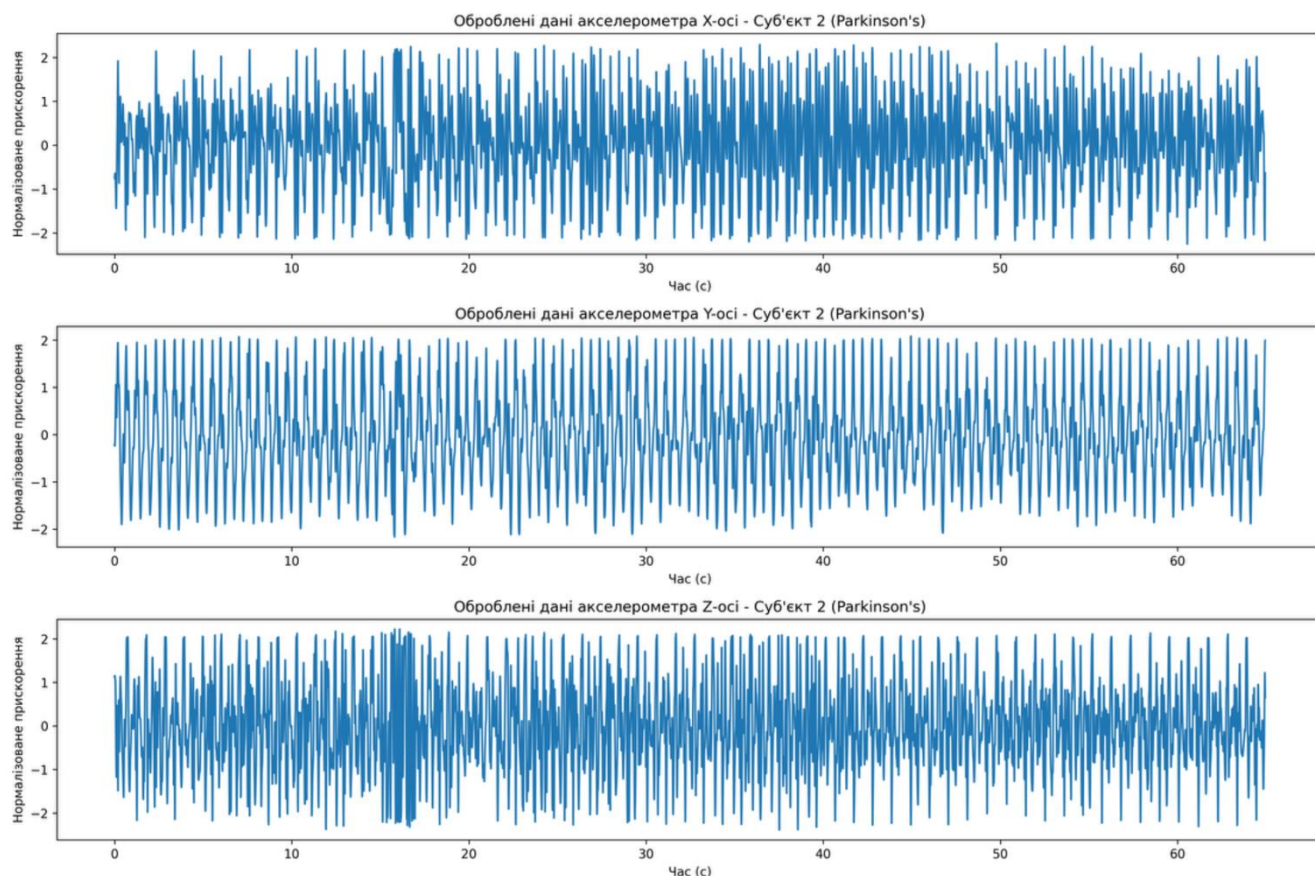


Рисунок 3.9 – Оброблені дані акселерометра пацієнта з хворобою Паркінсона

Порівнюючи цей сигнал з тим, що на рисунку 2.10, бачимо, що аналіз обробленого сигналу дасть кращі результати за рахунок того, що відсутні хаотичні стрибки та незначні коливання. Також відобразимо оброблений сигнал здорового пацієнта (рис. 3.10).

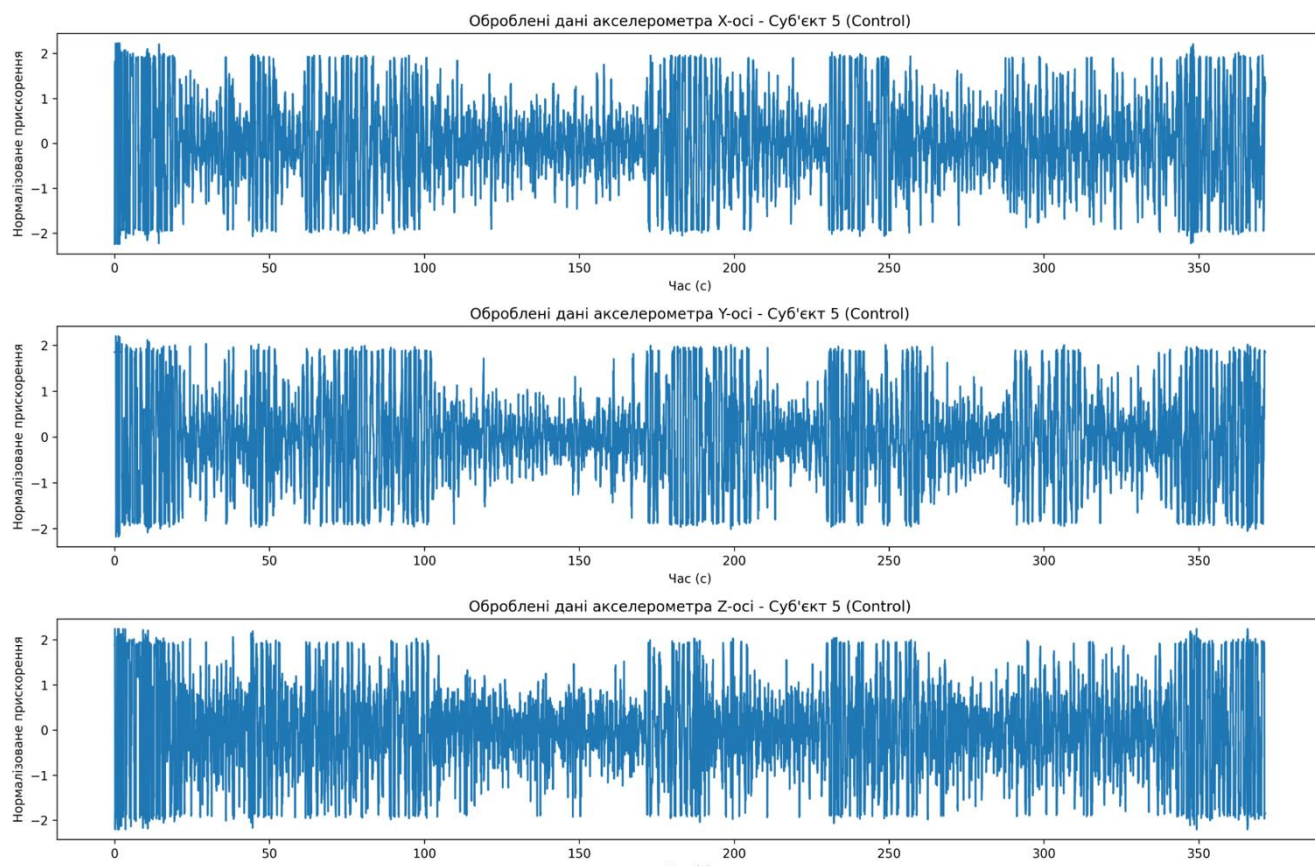


Рисунок 3.11 – Оброблені дані акселерометра здорового пацієнта

Порівнюючи оброблені дані здорового пацієнта з даними на рис. 3.11, поведінка сигналу зрозуміліша, а його розподіл менш хаотичний, за рахунок видалення стрибків та екстремальних значень, що зменшували репрезентативність вибірки й вносили сильну волатильність в частотний діапазон появи тремору.

Також проведено процес формування та відбору ознак, що найкращим чином характеризують сигнали й дають змогу проводити класифікацію наявності тремору. Спершу було проведено спектральний аналіз сигналу з використанням перетворення Фур'є [31], що дозволило отримати спектральну щільність сигналу й виявити пікові значення тремору, це дозволяє фільтрувати звичайні пікові збудження та скорочення м'язів (рис. 3.12).

```

spectral_centroid = np.array([np.sum(freqs * freq_features[:, i]) /
                               np.sum(freq_features[:, i]) if np.sum(freq_features[:, i]) > 0 else 0 for i in range(3)])
rolloff_percent = 0.85
spectral_rolloff = []
for i in range(3):
    cumulative_energy = np.cumsum(freq_features[:, i])
    cutoff = rolloff_percent * cumulative_energy[-1]
    rolloff_freq = freqs[np.where(cumulative_energy >= cutoff)[0][0]]
    spectral_rolloff.append(rolloff_freq)
spectral_rolloff = np.array(spectral_rolloff)

```

Рисунок 3.12 – Застосування спектрального аналізу до даних датасету

Результат застосування спектрального аналізу представлений для хворого та здорового пацієнтів. Для хворого він має такий вигляд на як на рис. 3.13.

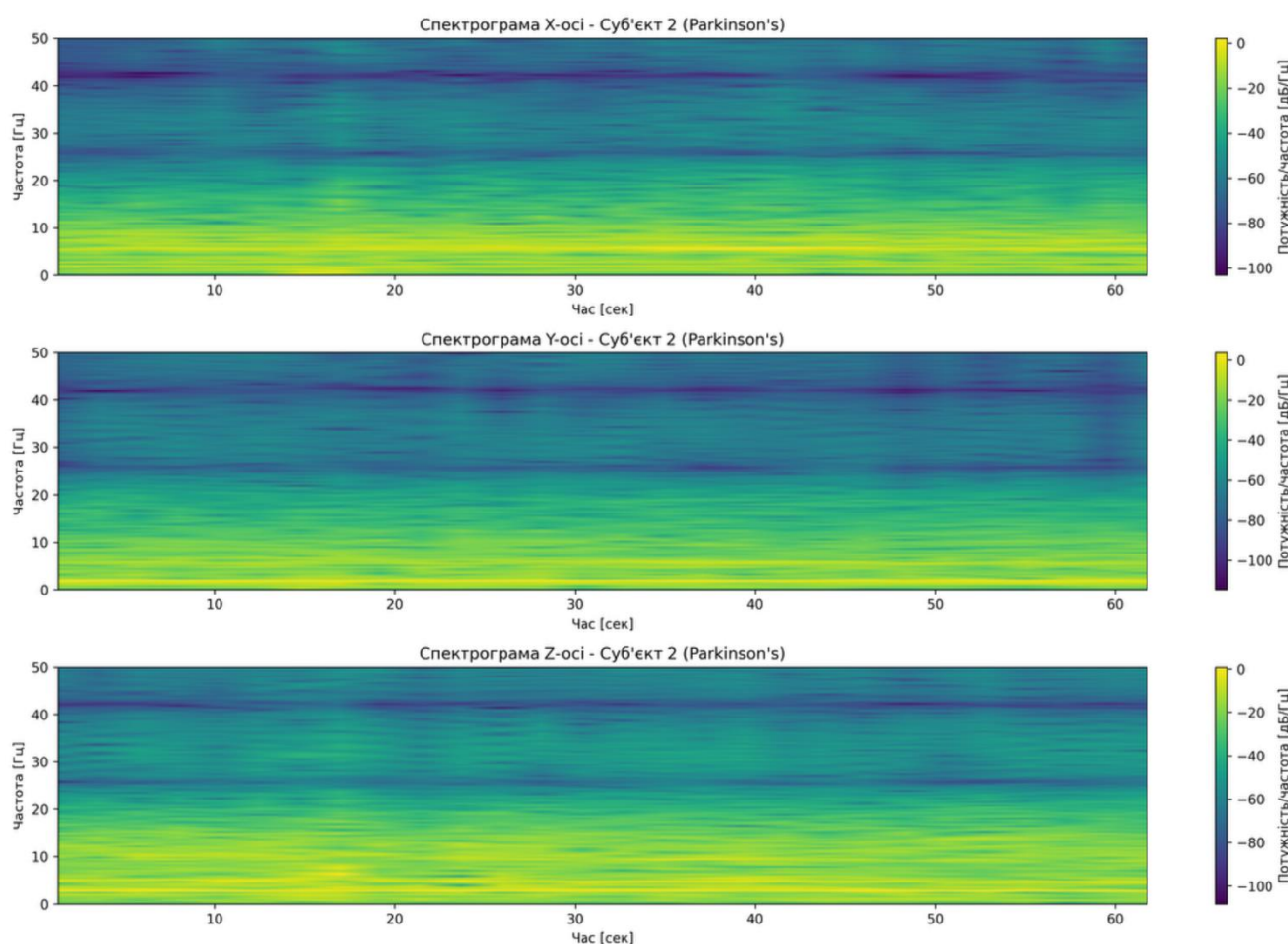


Рисунок 3.13 – Спектрограма обробленого сигналу акселерометра пацієнта з хворобою Паркінсона

На спектрограмі стають помітними найбільш потужні коливання в різному частотному діапазоні, що дозволяє аналізувати коливання акселерометра саме в тих ділянках, де найчастіше проявляється тремор, так видно, що дійсно у хворого

пацієнта в низькому частотному діапазоні потужність коливань найбільша, що вказує на наявність тремору а останнього.

Аналогічно представимо спектрограму для здорового пацієнта (рис. 3.14):

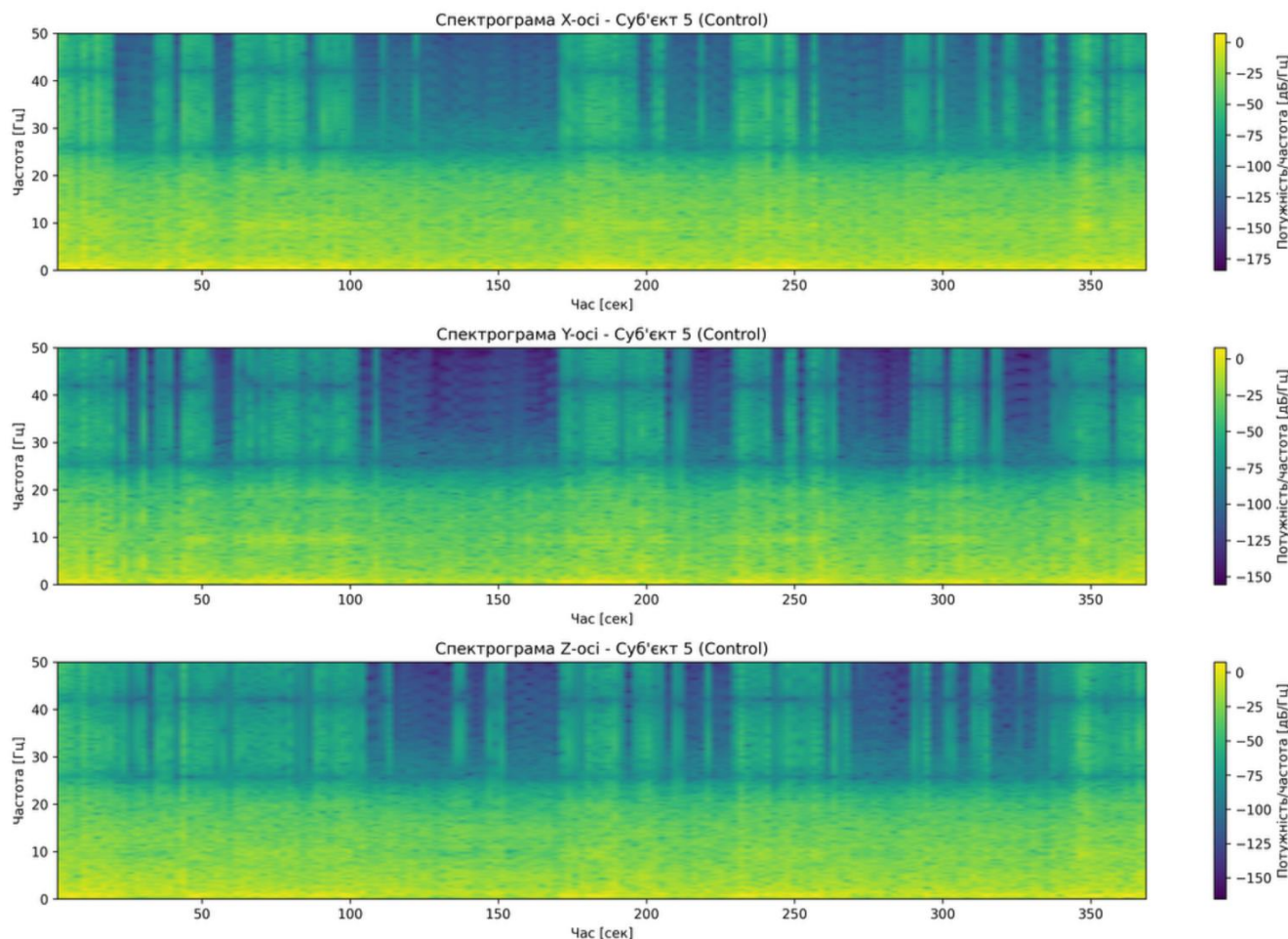


Рисунок 3.14 – Спектрограма обробленого сигналу акселерометра здорового пацієнта

Для здорового пацієнта, розподіл потужності менший, що вказує на відсутність тремору, оскільки пікова потужність практично не досягається в аналізованому частотному діапазоні.

Після цього проведено процес витягнення ознак сигналів, які можна розділити на окремі групи.

Перед створенням та застосуванням алгоритму визначення найбільш важливих ознак для моделювання, проведено додаткову підготовку зібраних даних, для досягнення ще більшого рівня їх узгодженості. В рамках додаткової обробки було застосовано метод Robust Scaler [2], який дозволяє стандартизувати

наявні дані відносно міжквартильного розмаху, а отже й забезпечує ще більшу стійкість датасету до викидів та зберігає структуру розподілу даних, а також підхід Quantile Transformer [22], що найчастіше застосовується саме при обробці біомедичних даних, та дозволяє нормалізувати розподіл даних, при цьому зберігаючи відносні відстані між значеннями спостережень (рис. 3.15).

```
robust_scaler = RobustScaler()
X_acc_scaled = robust_scaler.fit_transform(X_acc)
qt = QuantileTransformer(output_distribution='normal', random_state=42)
X_acc_qt = qt.fit_transform(X_acc_scaled)
```

Рисунок 3.15 – Нормалізація та стандартизація даних

Наступним кроком є застосування підходу PCA [14], що дозволяє зменшувати розмірність ознак, у випадку великої їх кількості. За результатами проведеної обробки даних отримана структура датасету (рис. 3.16).

```
Форма масиву ознак X_acc: (5461, 69)
Форма масиву міток y_acc: (5461,)
Кількість вилучених ознак: 69
```

Рисунок 3.16 – Розмірності масивів даних

Вилучено 69 ознак, що може створювати великі обчислювальні труднощі для моделі, а користі не приносить, за рахунок того, що деякі ознаки мають незначний, або відсутній вплив на результат визначення тремору. Тому для зменшення обчислювального навантаження на модель, необхідно застосування алгоритму PCA, який дозволить зберегти 95% варіативності ознак, тобто залишаться ті ознаки, що описують 95% змін у сигналах пацієнтів. Побудовано графік головних компонент датасету (рис. 3.17).

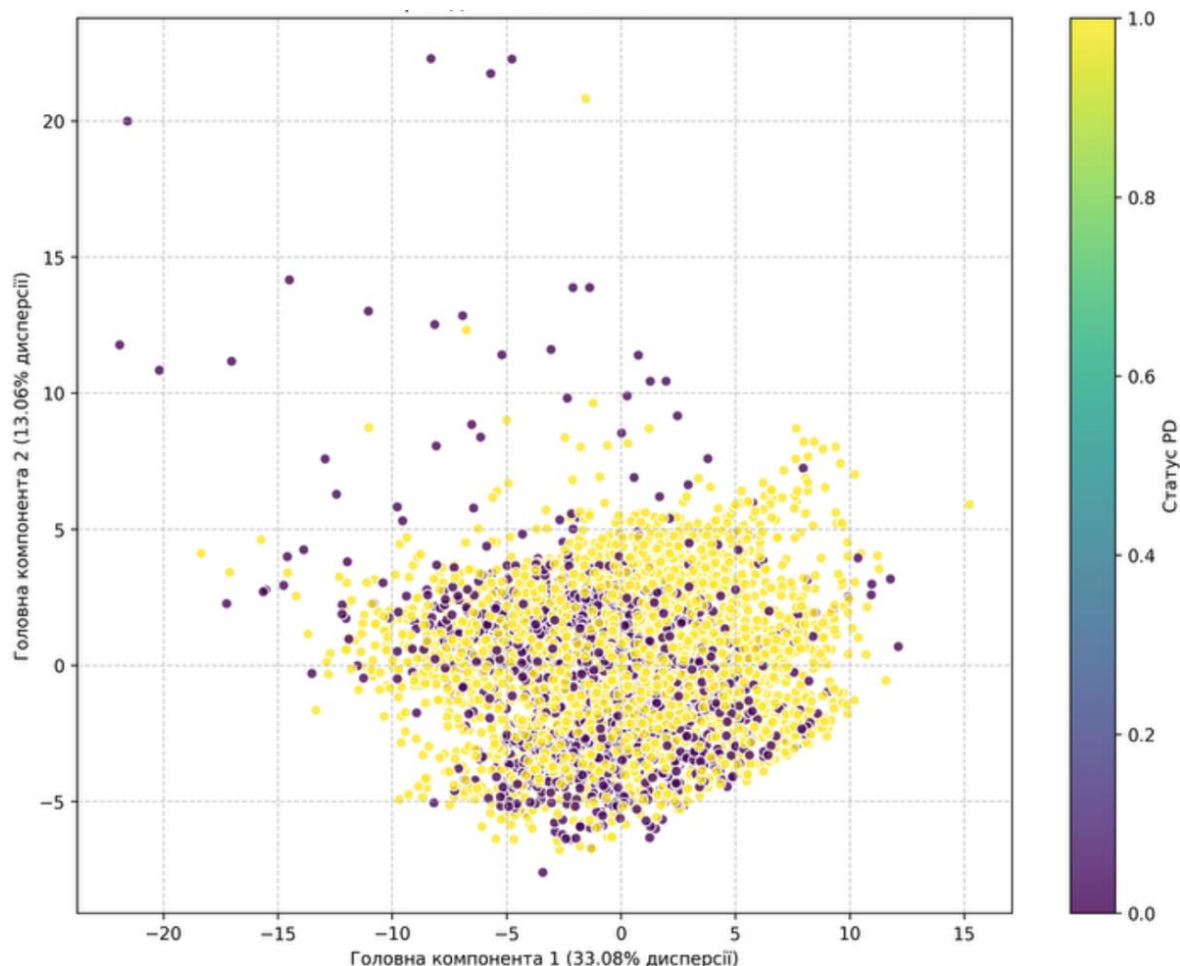


Рисунок 3.17 – Розподіл даних між головними компонентами

Дві компоненти погано підходять для подальшого моделювання через складність досліджуваних даних, оскільки вони описують малу частину варіативності ознак. Тому, для оптимізації процесу, побудовано графік динаміки кумулятивної дисперсії в залежності від різної кількості головних компонент, саме цей підхід дозволить обрати ту конфігурацію, щоб описати варіативність 95% даних і, при цьому, максимально зменшити кількість вхідних ознак (рис. 3.18).

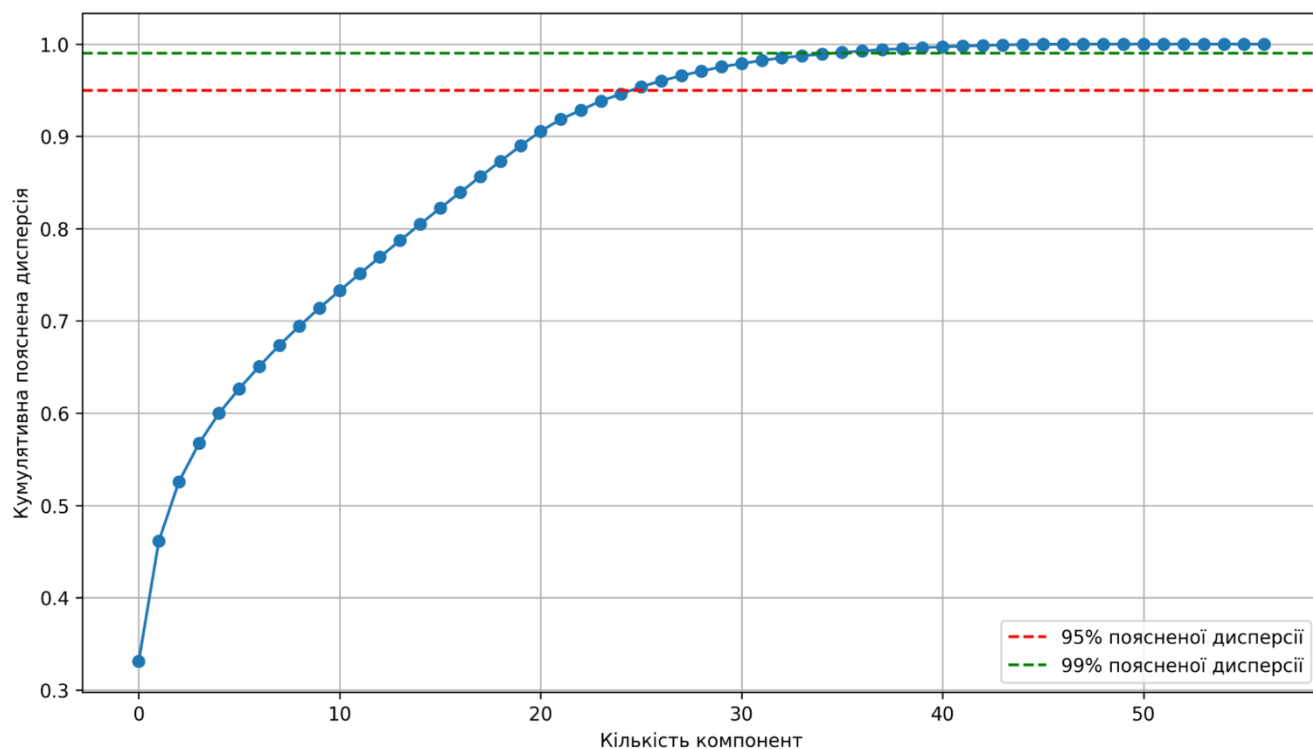


Рисунок 3.18 – Графік кумулятивної поясненої дисперсії за кількістю головних компонент

З даного графіка видно, що близько 23 ознак необхідно, щоб пояснити варіативність даних, на рівні 95%, це було підтверджено реалізованим алгоритмом PCA, після роботи якого отримана розмірність даних (рис. 3.19).

```
Розмір після нормалізації: (5461, 69)
Розмір після PCA: (5461, 23)
```

Рисунок 3.19 – Остаточна розмірність масиву даних

Як бачимо, кількість ознак була зменшена втричі, однак за встановленим припущенням, це не покличе за собою втрату варіативності та репрезентативності в даних.

Використано розглянутий алгоритм машинного навчання для відбору найбільш важливих ознак для моделювання (рис. 3.20).

```

rf_importances = model.feature_importances_
sorted_idx_rf = np.argsort(rf_importances)[::-1]

valid_indices = [i for i in sorted_idx_rf if i < len(feature_names)]
sorted_features_rf = [feature_names[i] for i in valid_indices]
sorted_importances_rf = rf_importances[valid_indices]

```

Рисунок 3.20 – Модель для визначення важливості ознак датасету

У результаті роботи алгоритму, отримано 20 найбільш важливих ознак для моделювання (рис. 3.21).

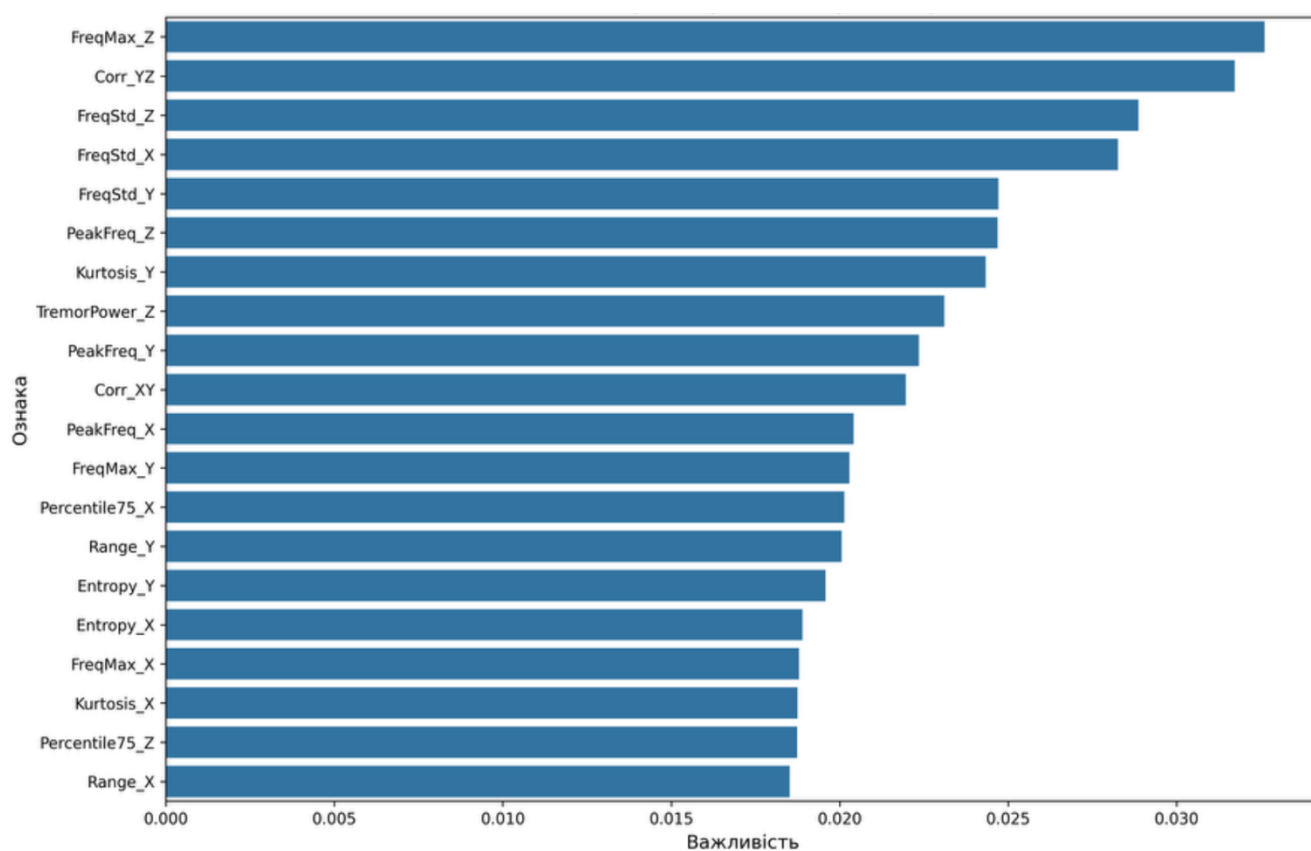


Рисунок 3.21 – Набільш важливі ознаки для визначення тремору

Врахуємо, що серед клінічних ознак, для моделювання, було вибрано всі наявні ознаки, оскільки вони корелюють між собою, а також з високим рівнем ймовірності дозволяють визначати наявність тремору у пацієнта.

На фінальній частині обробки даних, застосовано стратифікації, для розділення вибірки на три категорії: тренувальну, валідаційну та тестову. Структура загального датасету:

1. Тренувальна вибірка: 64%;
2. Валідаційна вибірка: 16%;
3. Тестова вибірка: 20%.

Враховуючи специфічність даних, їх обмежену кількість та складну структуру, створено валідаційну вибірку, яка мінімізує ймовірність виникнення явища перенавчання в моделі, а отже дозволить моделі краще узагальнювати спостереження.

Узагальнення повного процесу обробки даних наведено на діаграмі послідовності (рис. 3.22).

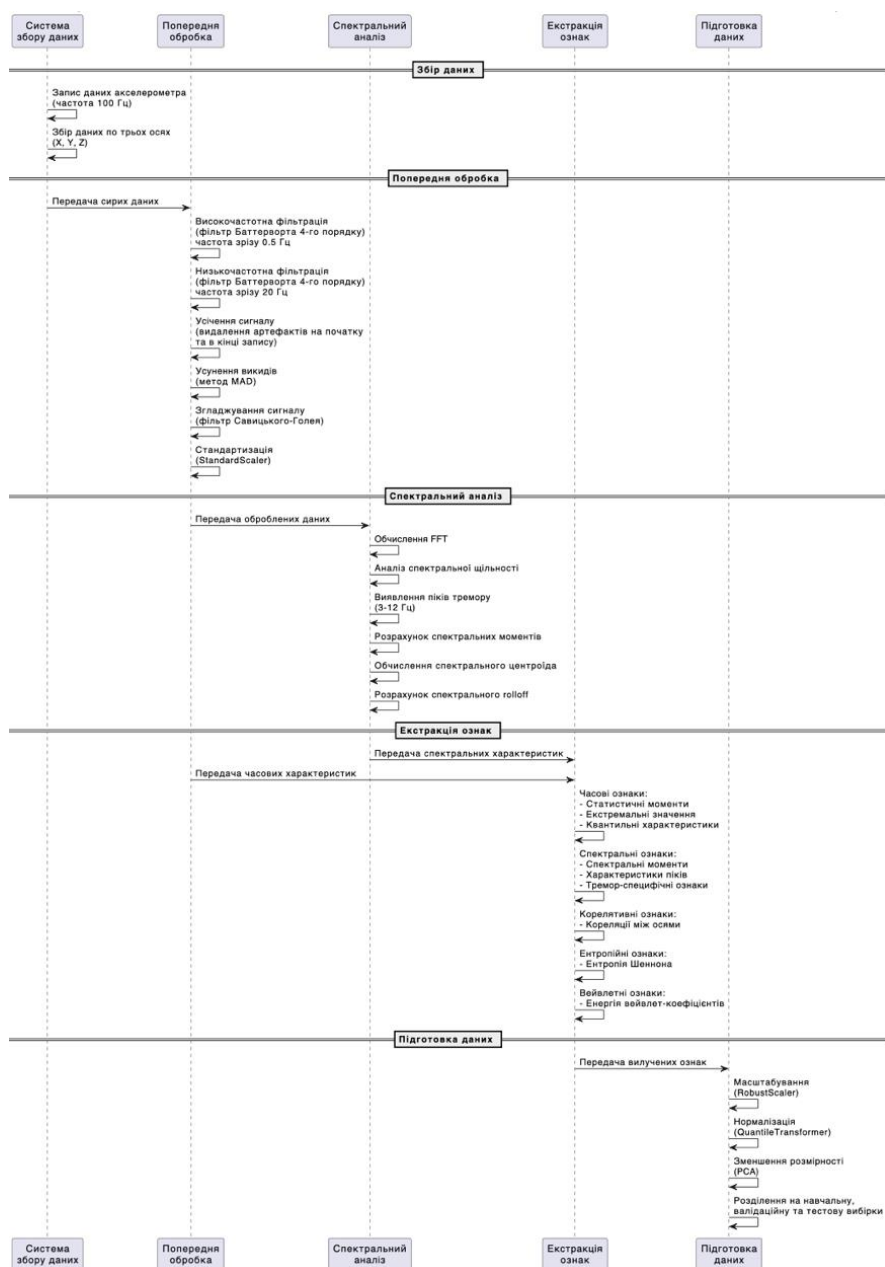


Рисунок 3.22 – Діаграма послідовності обробки та підготовки даних

3.2 Проведення експериментів

Процес навчання моделей побудований так, що кожна реалізована архітектура навчається окремо на повному наборі даних, це дозволяє отримати повноцінну модель опису даних, і, в подальшому, при ансамблюванні, дозволить підвищити точність роботи кожної з моделей окремо.

Процес навчання моделей передбачає налаштування гіперпараметрів та застосування механізмів раннього зупинення для запобігання перенавчанню моделей. Розглянемо специфіку реалізації процесу навчання моделі FNN (рис. 3.23).

```
input_shape = (X_train.shape[1],)
model = create_fnn_model(input_shape)
early_stopping = callbacks.EarlyStopping(monitor='val_loss', patience=20, restore_best_weights=True)
model_checkpoint = callbacks.ModelCheckpoint(os.path.join(MODEL_DIR, 'best_fnn_model.h5'),
                                             monitor='val_loss', save_best_only=True)
history = model.fit(X_train, y_train, epochs=100, batch_size=32, validation_data=(X_val, y_val),
                    callbacks=[early_stopping, model_checkpoint], verbose=1)
```

Рисунок 3.23 – Реалізація процесу навчання моделі FNN

Аналогічно було реалізовано і процес навчання моделі DNN, проте варто звернути увагу на кілька ключових особливостей, притаманних процесу навчання кожної з моделей. Зокрема, було застосовано механізм раннього зупинення з параметром *patience* = 20, що дало змогу своєчасно завершити навчання та уникнути перенавчання. Крім того, під час тренування зберігалися ваги моделі, що відповідали найкращим показникам за метрикою валідаційних втрат.

Для оптимізації процесу навчання відповідно до особливостей архітектури застосовувалися різні розміри пакета: 32 — для FNN та 256 — для DNN. Також встановлено обмеження на кількість епох — не більше 100, що дозволяло ефективно контролювати загальний час тренування.

Доцільним є візуальне представлення динаміки процесу навчання обох моделей. Для цього спершу буде проаналізовано зміну точності на тренувальній та валідаційній вибірках, що ілюструється на рисунку 3.24.

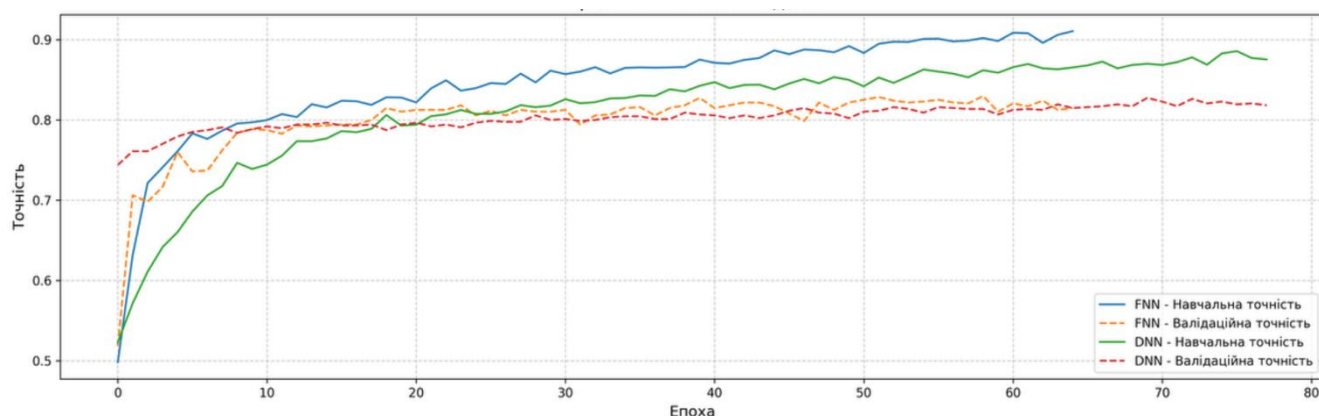


Рисунок 3.24 – Порівняння динаміки точності роботи моделей

Обидві моделі дають майже однаковий результат точності, що вказує на задовільний результат навчання та узагальнення даних. Враховуючи складність структури вхідного вектору, такий процес навчання є адекватним.

Динаміка зміни функції втрат моделей, для оцінки загальної якості роботи архітектур та похибок роботи останніх показано на рис. 3.25.

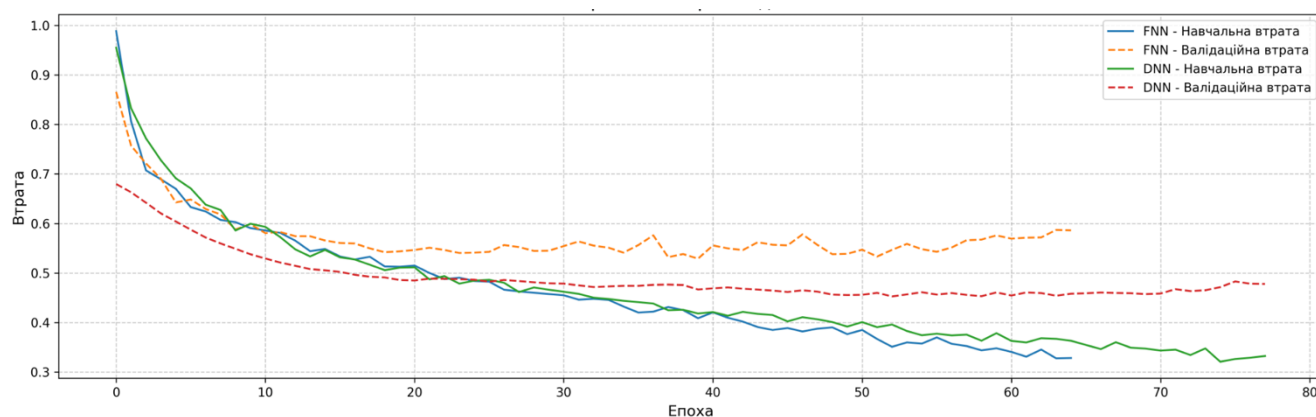


Рисунок 3.25 – Порівняння динаміки втрат моделей

Функціонал якості доволі швидко збігається до мінімуму, це свідчить про якісний процес навчання та обробки даних всередині моделей. Модель DNN навчається краще, однак дає меншу точність. В медичних дослідженнях, з обмеженим обсягом даних та складною структурою останніх, подібні результати є більш ніж задовільними, особливо при обробці даних отриманих у «польових» умовах, а не в процесі лабораторного дослідження.

3.3 Оцінка ефективності моделей

Процес оцінювання якості розробленої ансамблевої архітектури здійснювався поетапно. Спершу аналізувалася ефективність кожної окремої моделі, далі — якість роботи ансамблю, після чого результати порівнювалися з даними зі спорідненого дослідження для отримання об'єктивної оцінки.

Для повного розуміння сильних і слабких сторін моделей використовувалися різні метрики. Ассурасу характеризує загальну точність моделі, тобто співвідношення правильних передбачень до загальної кількості прикладів. Precision відображає точність у виявленні позитивного класу — скільки з передбачених позитивних випадків справді є позитивними. Recall (або чутливість) вказує, яку частину реальних позитивних випадків модель змогла правильно розпізнати. F1-score об'єднує ці дві характеристики у гармонійне середнє, що особливо важливо при дисбалансі класів. AUC-ROC (площа під ROC-кривою) є однією з найцінніших метрик для подібного типу задач, адже дозволяє оцінити здатність моделі розрізняти класи при різних порогах прийняття рішення, що критично при обмеженій кількості даних.

У подальшому аналізі буде розглянута ефективність роботи кожної моделі окремо та ансамблю загалом. Побудуємо ROC-криві (Receiver Operating Characteristic), які демонструють баланс між чутливістю (recall) та специфічністю моделі при зміні порогового значення. Графіки представлені на рис.3.26 – рис.3.28.

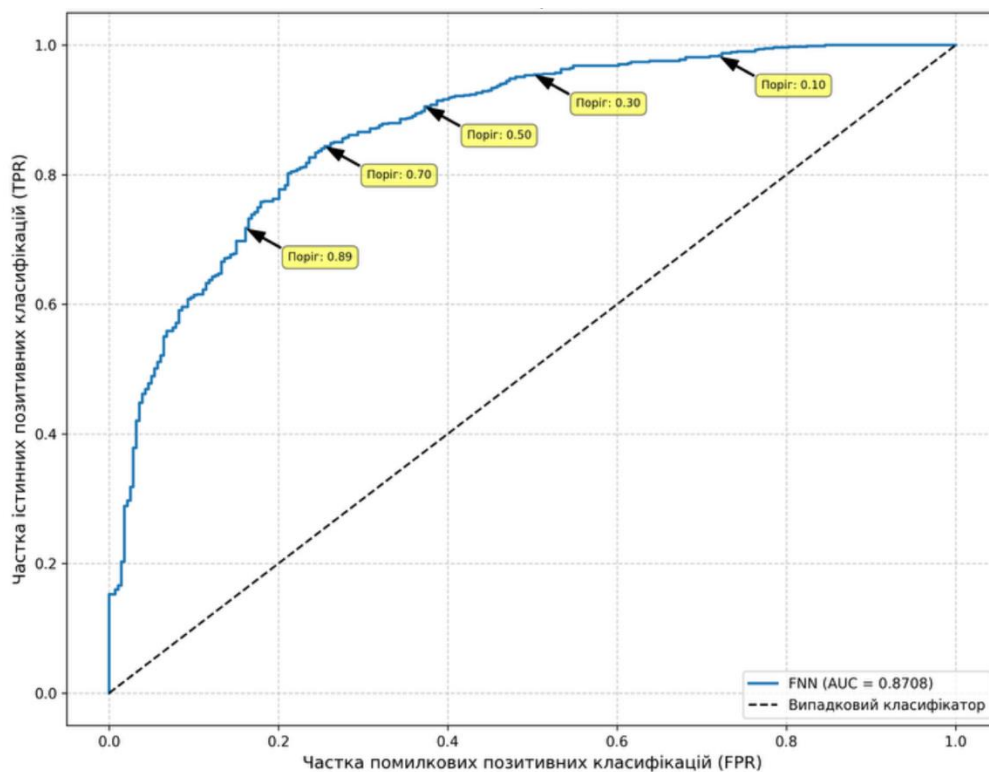


Рисунок 3.26 – ROC-крива для моделі FNN

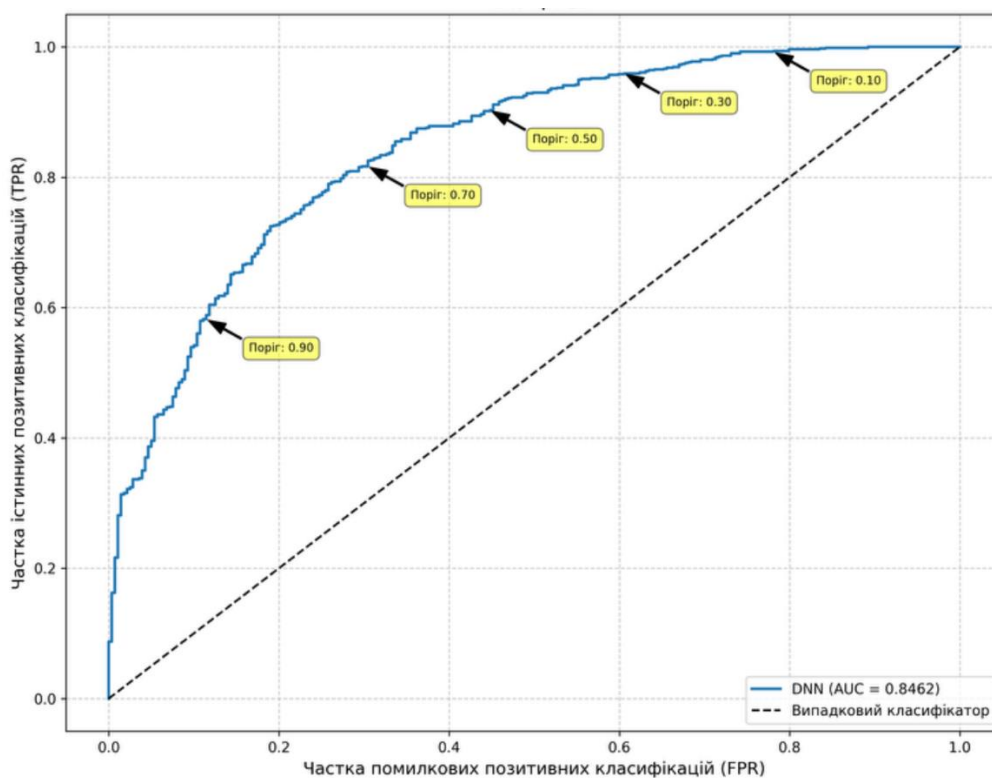


Рисунок 3.27 – ROC-крива для моделі DNN

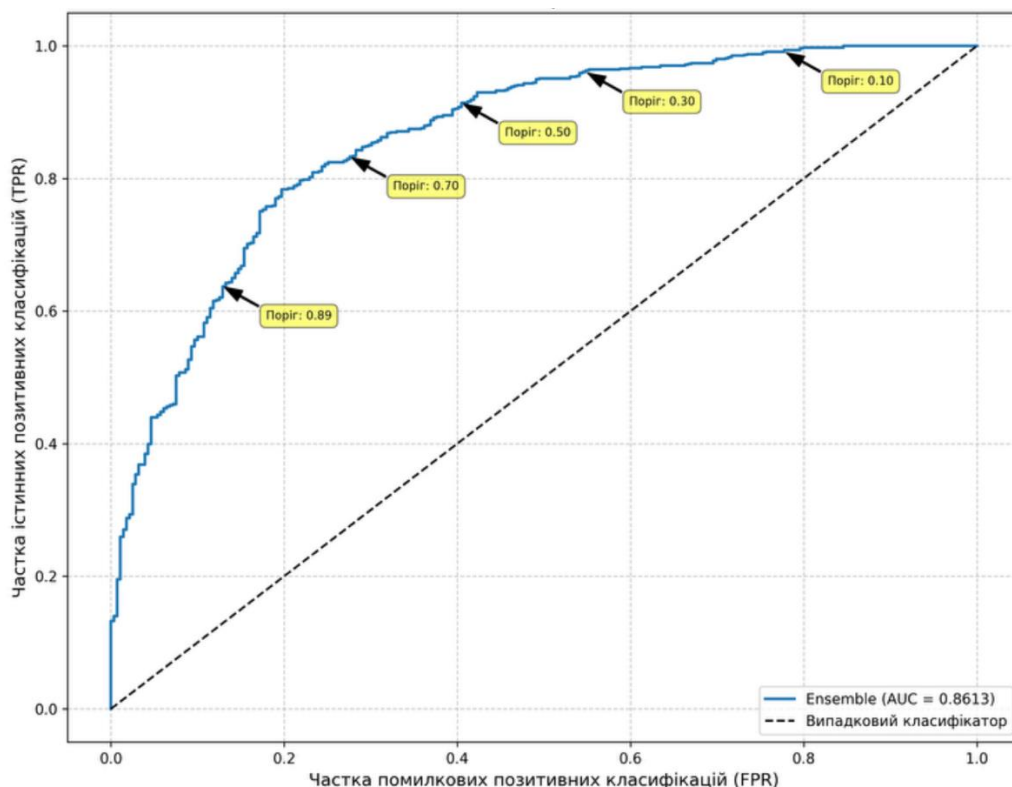


Рисунок 3.28 – ROC-крива для ансамблевої моделі

Для всіх моделей, побудована крива показує високу якість роботи мереж, частка випадкових класифікацій низька, отже узагальнюючі можливості моделей знаходяться на високому рівні, незважаючи на те, що досягти ідеальної траєкторії не вдалось, за рахунок майже нульового коефіцієнта випадкових класифікацій дані моделі працюватимуть краще за аналогічні, з більшим відгуком. А враховуючи обмеженість даних, отримані результати цілком задовільні.

Необхідно побудувати криві точності-повноти, для незбалансованих наборів даних, оскільки вони фокусуються на ефективності моделі стосовно позитивного класу. Оскільки за результатами аналізу даних, було виявлено явний дисбаланс між класами та всередині класів, аналіз поведінки кривої точності-повноти дозволить виявити спроможність побудованих моделей виконувати задачу для довільних вхідних даних пацієнтів а узагальнювати отримані результати, в процесі навчання.

Криві для трьох моделей проілюстровані на рис. 3.29 – рис.3.31.

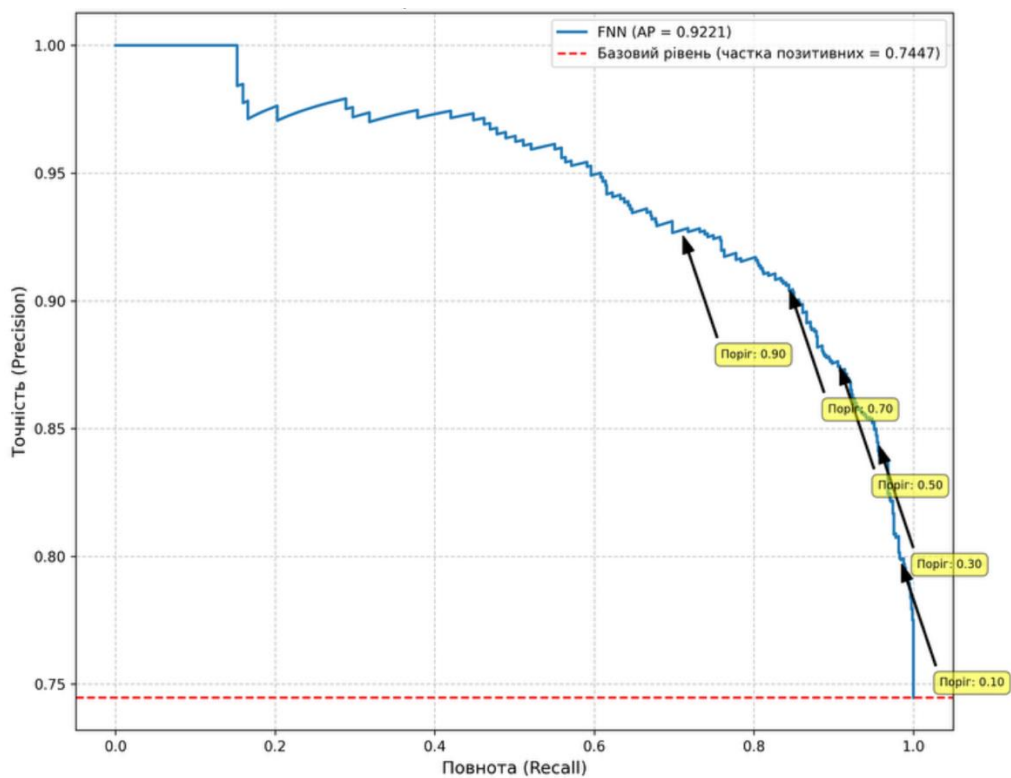


Рисунок 3.29 – Крива Precision-Recall для моделі FNN

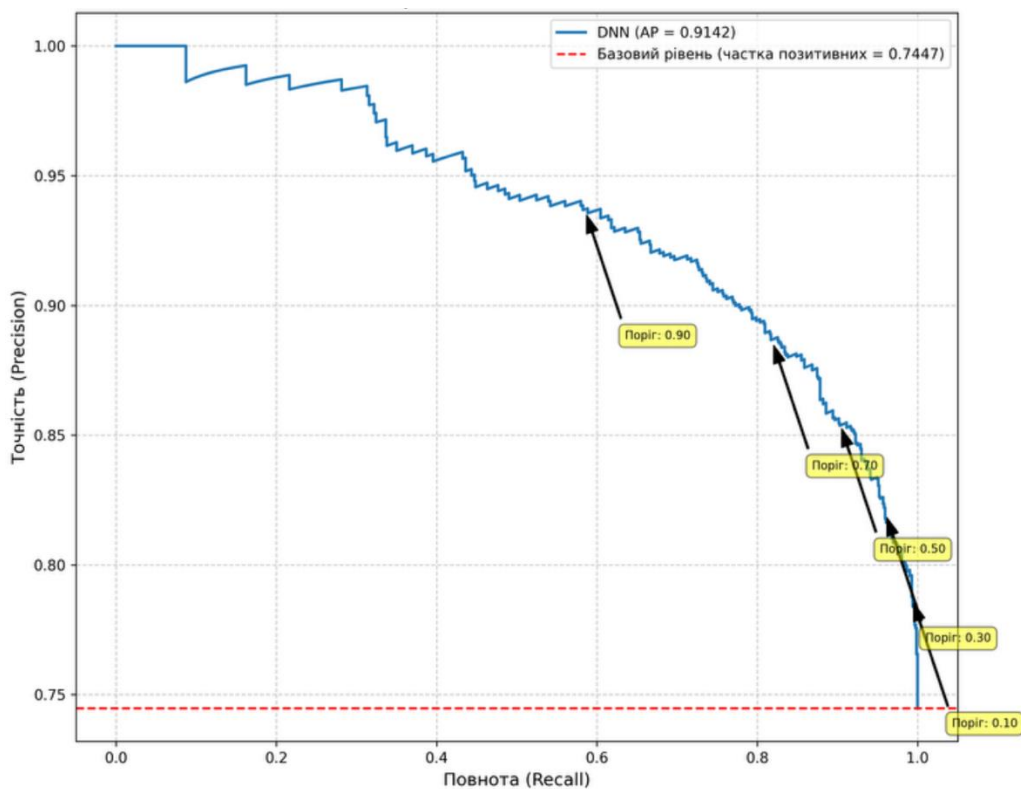


Рисунок 3.30 – Крива Precision-Recall для моделі DNN

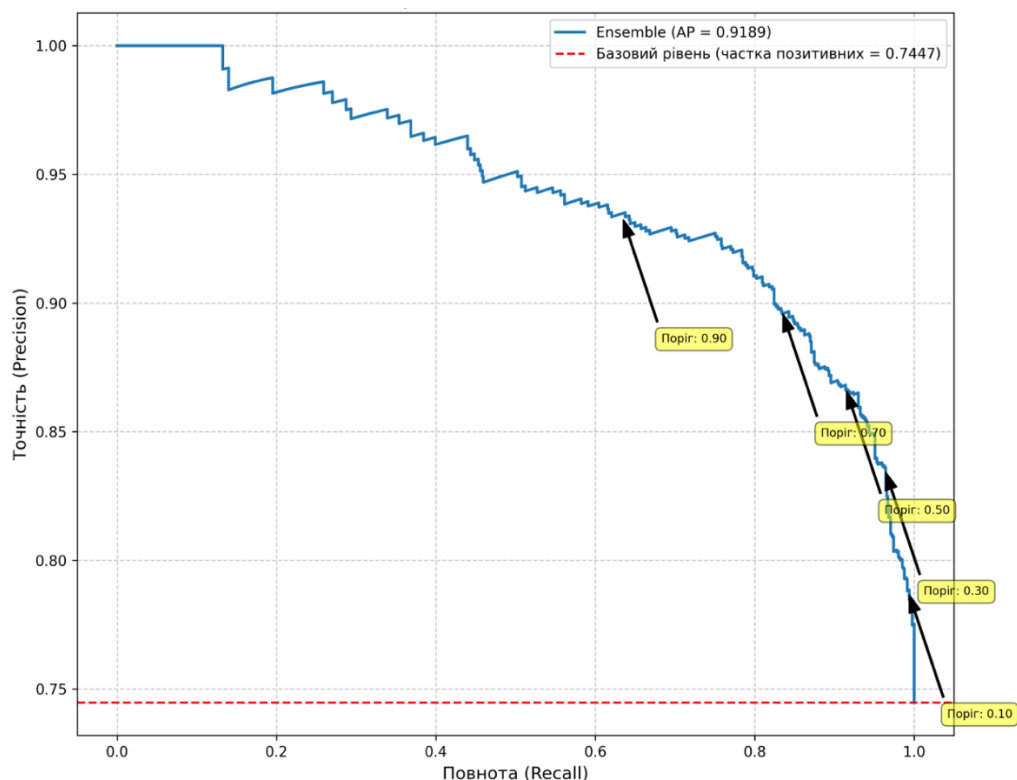


Рисунок 3.31 – Крива Precision-Recall для ансамблевої моделі

Отже, три моделі показують високу середню точність роботи та створюється ідеальний компроміс між точністю роботи та повнотою опису даних. Модель відшуковувала закономірності з огляду на описані ознаки, а не конкретні їх значення, що дозволило використовувати її для довільних вхідних даних із сенсорів.

Дослідимо роботу моделей на тестових даних, за допомогою матриць помилок (рис. 3.32 – рис.3.34).

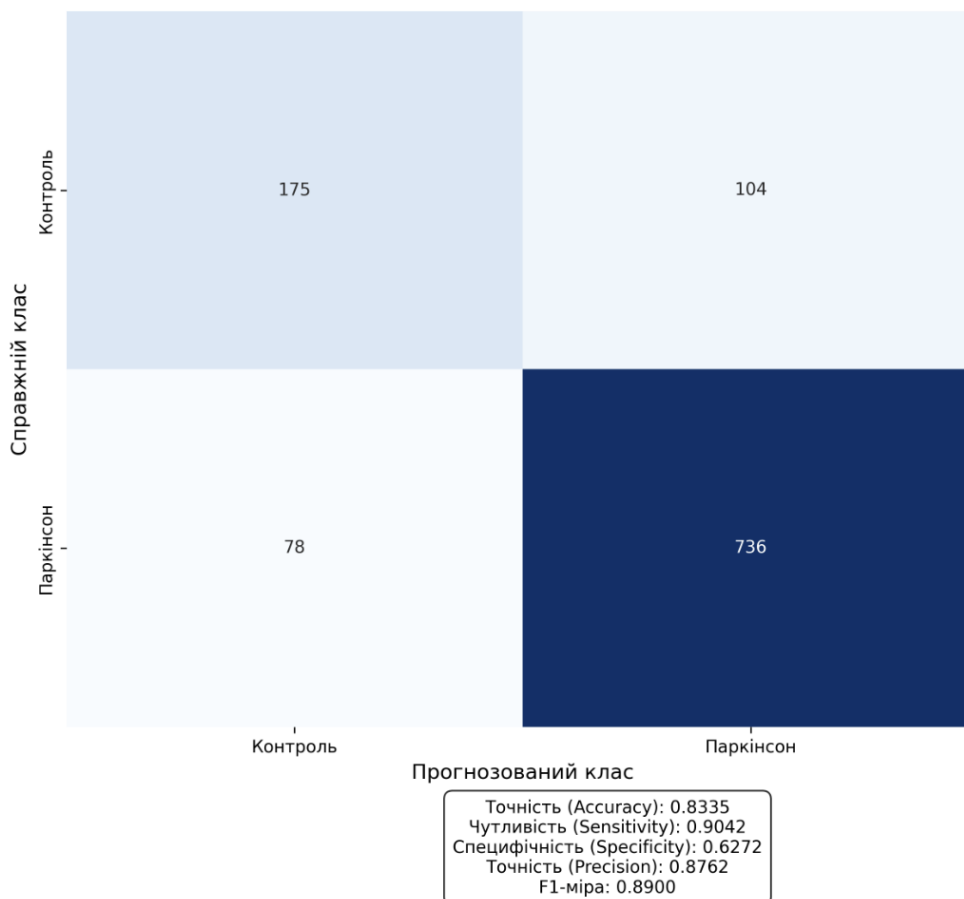


Рисунок 3.32 – Матриця помилок для моделі FNN

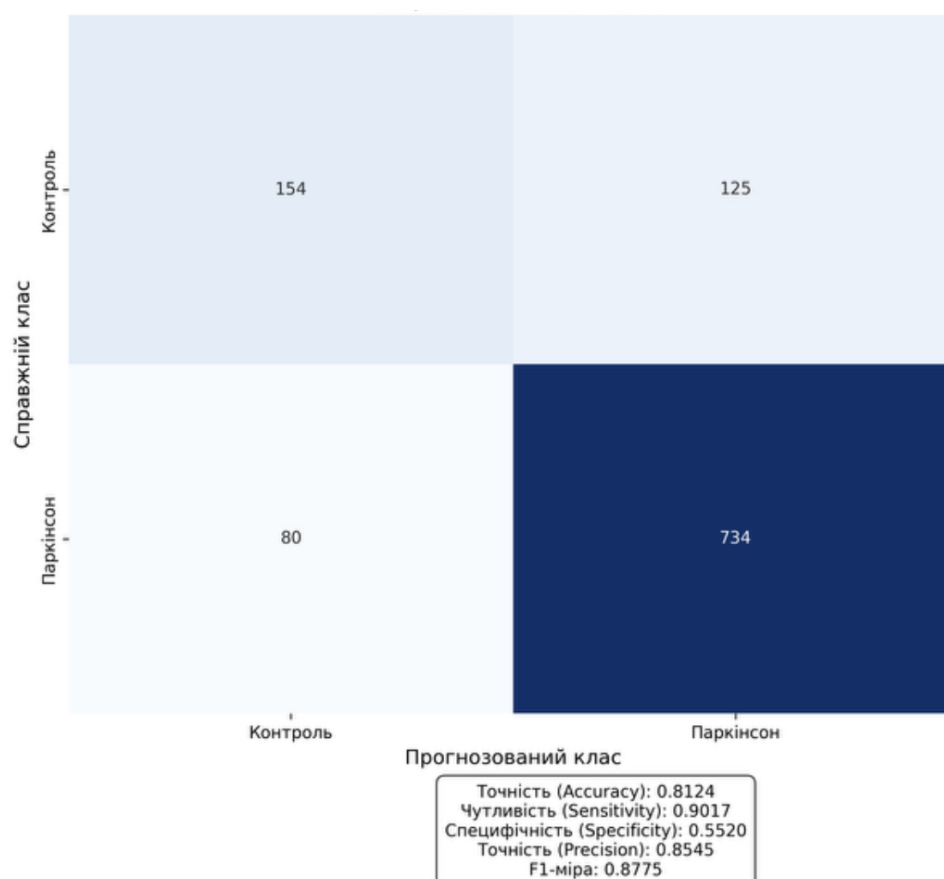


Рисунок 3.33 – Матриця помилок для моделі DNN

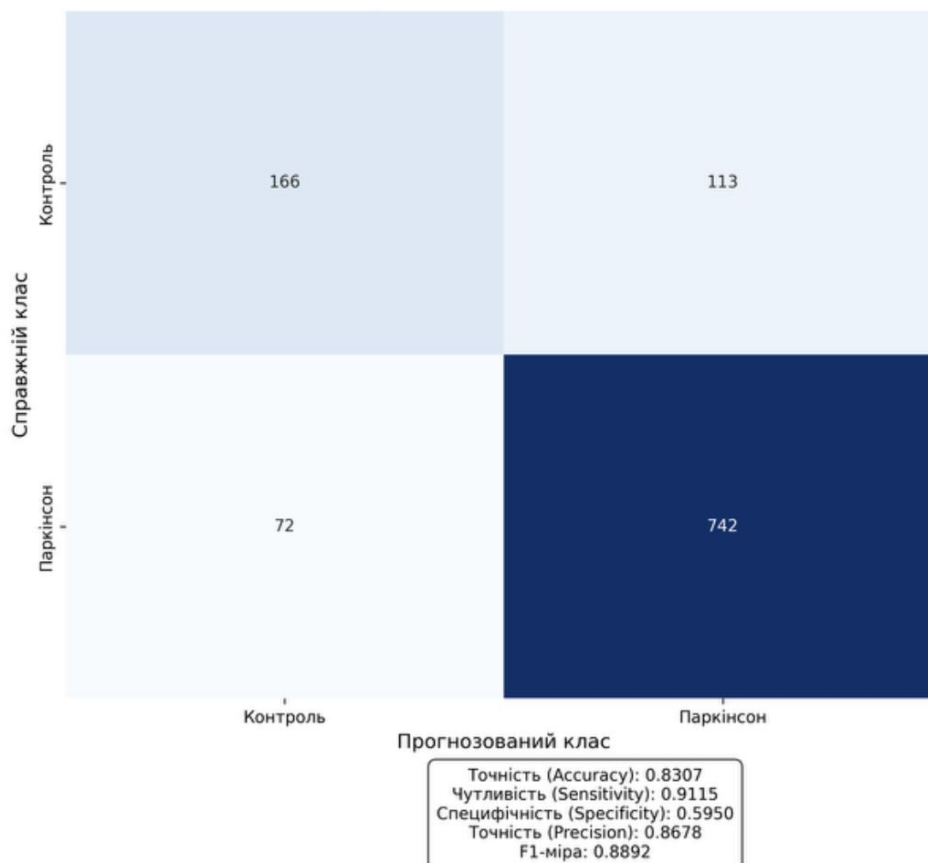


Рисунок 3.34 – Матриця помилок для ансамблевої моделі

Аналізуючи роботу моделей на тестових даних, спостерігаємо різницю в якості роботи окремих моделей та ансамблю. Ансамблева модель дає кращі метрики на тестових даних, це є вкрай важливим аспектом, оскільки від точності роботи моделей, залежить процес розвитку подальших клінічних досліджень.

Порівняємо отримані результати із результатами, що були отримані у дослідженні [9], дане дослідження мало аналогічний вектор розвитку, однак у процесі були використані інші методи аналізу даних, обробки даних та моделі навчання. Всі дані представимо у вигляді таблиці (табл. 3.1)

Аналізуючи дані з табл. 3.1, можна зробити висновок, що розроблені моделі у даному дослідженні дають набагато кращий результат. Однозначно, складна послідовність обробки та підготовки часових та просторових даних, дозволила обраним моделям досягти значних результатів як в точності роботи, так і в репрезентативності на тестовій вибірці, обрані архітектурні рішення є більш комплексними та універсальними для випадків, коли дані мають складну

структуру та обмежену кількість, оскільки здатні будувати закономірності розподілу останніх за рахунок глибших архітектур та зв'язків.

Таблиця 3.1

Порівняння отриманих результатів

Модель	Метрика	Значення
FNN	Ассурасу (точність), %	83.35
DNN		81.24
Ансамбль (FNN + DNN)		83.07
LDA		71.10
NB		71.29
k-NN		73.01
SVM		68.46
SVM rbf		71.73
DT		64.02
Voting model		82.00

3.4 Розрахунок економічного ефекту

Метою даного розділу було дослідження програмного продукту для прогнозування розвитку тремору на основі даних інерційних сенсорів із реалізованим веб-інтерфейсом на Flask. Система призначена для застосування в медичних установах для ранньої діагностики та моніторингу пацієнтів із хворобою Паркінсона та іншими неврологічними порушеннями.

Після постановки задачі було побудовано морфологічну карту (рис. 5.1), яка охоплює варіанти вибору мови програмування, алгоритму прогнозування та способу перевірки точності.

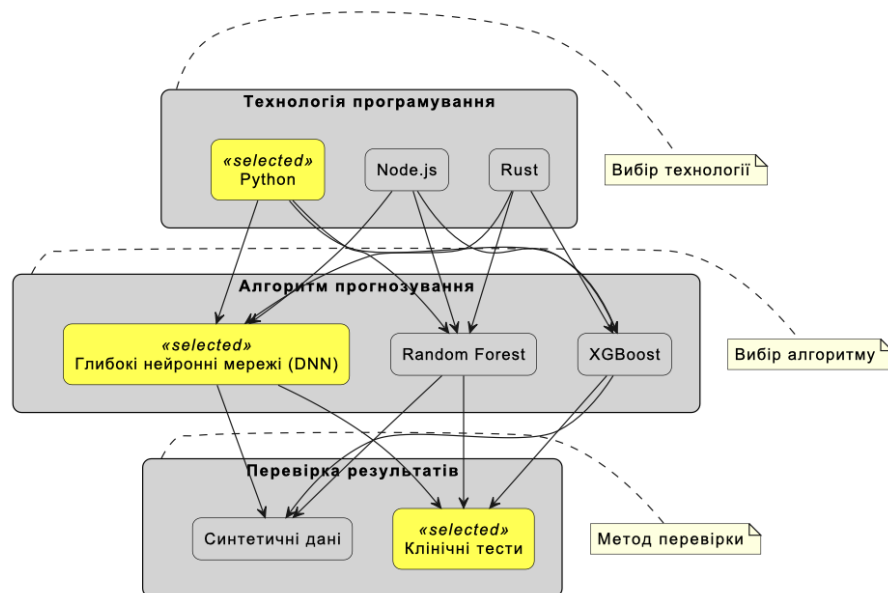


Рисунок 5.1 – Морфологічна карта

У процесі аналізу було розглянуто три варіанти реалізації функціоналу (наприклад: а-б-а, а-в-б), і на основі позитивно-негативної матриці обрано два найперспективніші. Для оцінювання варіантів використано функціонально-вартісний аналіз.

Параметри оцінки:

- X_1 – розмір коду (рядки),
- X_2 – час навчання моделі (хв),
- X_3 – затримка прогнозу (мс),
- X_4 – точність (%)

На основі експертного ранжування (4 експерти), коефіцієнти важливості параметрів склали ($W = 0.5625 > 0.5$ — узгодженість оцінок підтверджена):

$$X_1 = 0.25, \quad X_2 = 0.20, \quad X_3 = 0.30, \quad X_4 = 0.35$$

Під час аналізу якості було розраховано зведений коефіцієнт якості за формулою:

$$KK_i = \sum_{j=1}^4 (\omega_i \cdot V_j)$$

Результати для двох варіантів:

- Варіант 1 (Python + DNN + клінічні тести): **KK = 23.1**

- Варіант 2 (Python + XGBoost + клінічні тести): **KK = 22.05**

Економічний аналіз включав розрахунок трудомісткості та витрат на розробку.

Загальна трудомісткість становить 470.6 людино-годин (35.2 днів на систему прогнозування та 23.6 днів на веб-інтерфейс). У розробці задіяні програміст (оклад 45 000 грн) і Data Scientist (60 000 грн). Загальна заробітна плата за час розробки — близько 116 274 грн.

Таким чином, найбільш доцільним варіантом впровадження є використання Python з глибокими нейронними мережами. Проєкт відповідає вимогам до точності, має оптимальні витрати й забезпечує ефективну роботу в клінічних умовах.

Висновки до розділу 3

У третьому розділі було реалізовано комплексний підхід до обробки даних акселерометрів та оцінено ефективність розроблених моделей прогнозування хвороби Паркінсона.

Етап попередньої обробки даних включав три ключові стадії: обробку сигналів акселерометрів з побудовою спектрів потужності в лінійному та логарифмічному масштабах, визначення та витягнення інформативних ознак, а також підготовку даних до моделювання. Застосування методів усічення сигналу, фільтрації викидів за допомогою Median Absolute Deviation, стандартизації даних з використанням Robust Scaler та нормалізації розподілу через Quantile Transformer забезпечило високу якість підготовлених даних.

Аналіз частотних характеристик підтвердив наявність специфічних особливостей у сигналах пацієнтів з хворобою Паркінсона порівняно зі здоровими особами, що обґрунтовує застосування частотного аналізу для діагностичних цілей. Виявлено схожість поведінки всіх трьох компонент сигналів, що свідчить про важливість комплексного аналізу тривимірних даних.

Проведені експерименти з навчання окремих моделей FNN та DNN з подальшим ансамблюванням показали ефективність обраного підходу.

Застосування механізмів раннього зупинення, оптимальних розмірів пакетів та обмеження кількості епох дозволило уникнути перенавчання та забезпечити стабільність результатів.

Оцінка ефективності за множинними метриками (Accuracy, Precision, Recall, F1-score, AUC-ROC) продемонструвала високу якість роботи розроблених моделей. Ансамблева архітектура показала кращі результати порівняно з окремими моделями, що підтверджує доцільність її використання. Порівняння з результатами аналогічних досліджень засвідчило перевагу запропонованого підходу, зокрема завдяки комплексній послідовності обробки даних та використанню глибших архітектурних рішень.

Отримані результати підтверджують ефективність розробленої методики для діагностики хвороби Паркінсона на основі інерційних вимірювань мобільних пристроїв та демонструють перспективність подальшого розвитку цього напрямку досліджень.

РОЗДІЛ 4

РОЗРОБКА ВЕБ-ІНТЕГРОВАНОЇ СИСТЕМИ ДЛЯ ВИКОРИСТАННЯ МОДЕЛЕЙ

Веб-система реалізована на базі Python із застосуванням фреймворку Flask. Система складається з кількох основних компонентів, що забезпечують її функціональність. Центральним елементом є скрипт `app.py`, який відповідає за запуск веб-сервера та обробку користувацьких запитів. У межах цього скрипту реалізовано маршрути для різних функцій, зокрема: перегляд сторінок розвідувального аналізу, подання сигналів, навчання моделей та здійснення прогнозів.

Для обробки вхідних даних система використовує низку модулів, включаючи масштабування, квантильну трансформацію та метод головних компонент (PCA). На основі цих даних працюють дві попередньо натреновані нейронні мережі — FNN та DNN. Користувач може завантажити сигнали акселерометра через форму або файл, після чого дані проходять обробку: витягуються ознаки, виконується масштабування, а потім здійснюється прогнозування.

Візуалізація результатів забезпечується за допомогою бібліотеки Matplotlib. Вона використовується для наочного представлення сигналів та проведення спектрального аналізу, що дає змогу інтерпретувати результати обробки та оцінити якість моделі.

Створено архітектуру спроектованого додатка (рис. 4.1).

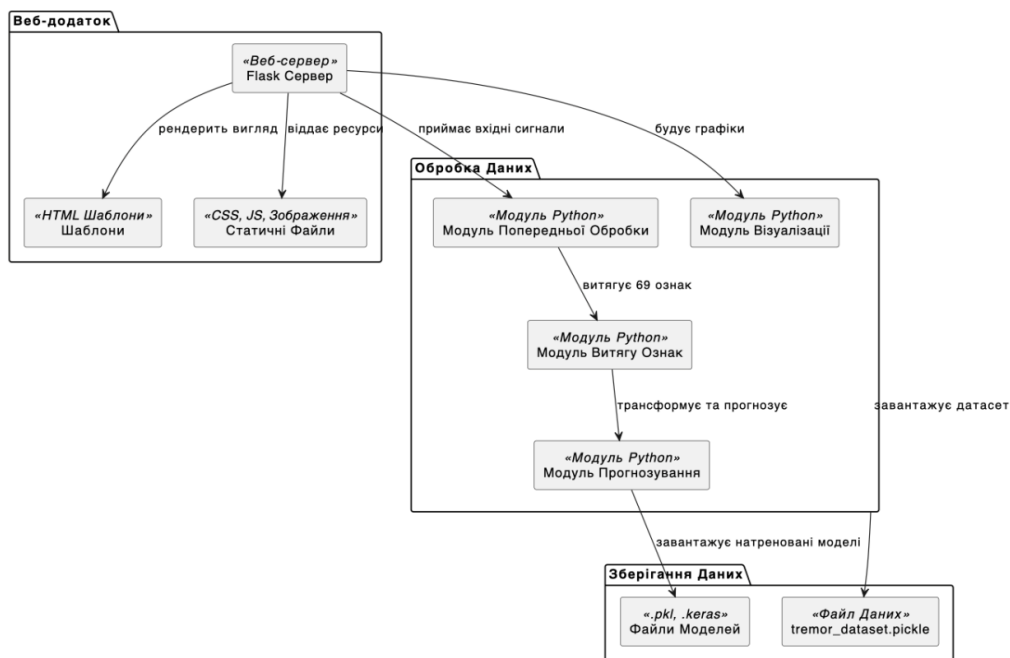


Рисунок 4.1 – Архітектурне рішення веб-додатка

Веб-сервер взаємодіє з HTML-шаблонами та статичними файлами, а також викликає модулі обробки даних для попередньої обробки, витягу ознак та прогнозування. Дані зберігаються у файлах, що дозволяє відокремити логіку обробки від представлення.

Далі доцільно відобразити загальну логіку роботи додатка (рис. 4.2).



Рисунок 4.2 – Логіка роботи веб-додатка

Логіка роботи системи тісно пов'язана з послідовністю викликів між користувачем та сервером. Діаграму послідовності, показано на рис. 4.3.

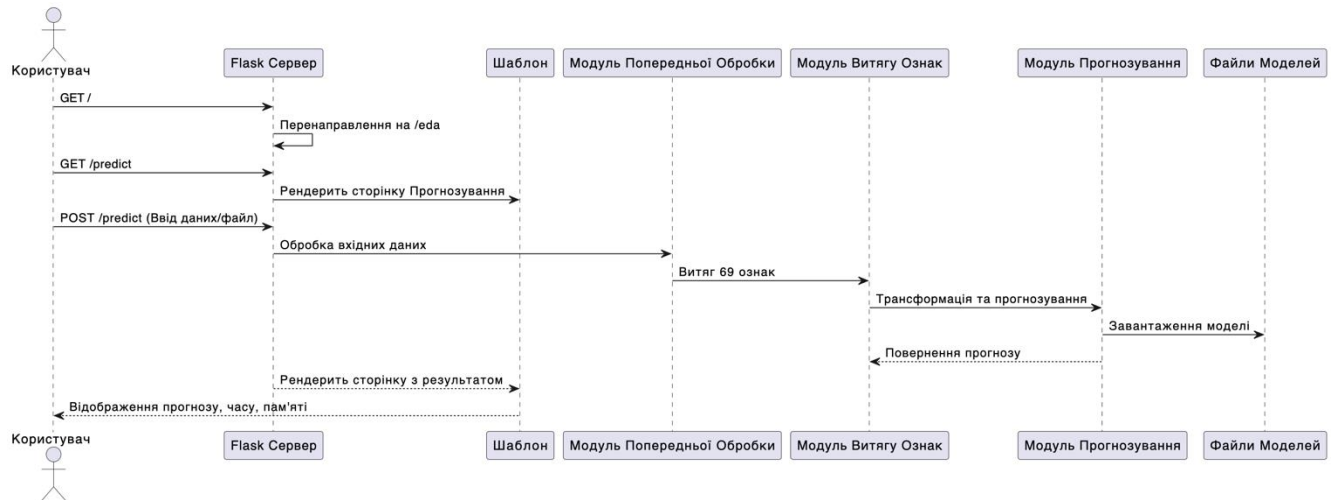


Рисунок 4.4 – Діаграма послідовності роботи веб-додатка

Користувач надсилає запит, сервер обробляє дані, викликаються модулі обробки та прогнозування, і результат повертається користувачу. Діаграму прецедентів вибору користувача в системі наведена на рис. 4.5.



Рисунок 4.5 – Діаграма прецедентів веб-додатка

Діаграма прецедентів показує основні дії, які виконує користувач у системі, включаючи перегляд інформації та роботу з прогнозуванням і файлами. Головна сторінка додатку наведена на рис. 4.6.

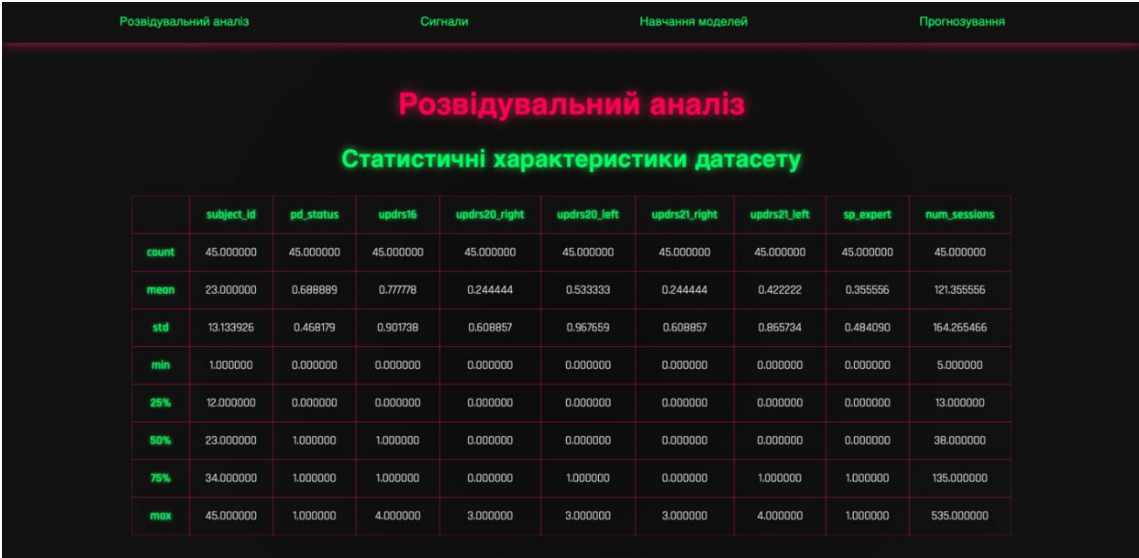


Рисунок 4.6 – Головна сторінка додатку

На головній сторінці представлені результати статистичного аналізу вхідного датасету, відображені у вигляді табличних даних (описової статистики) та графіків (рис. 4.7).

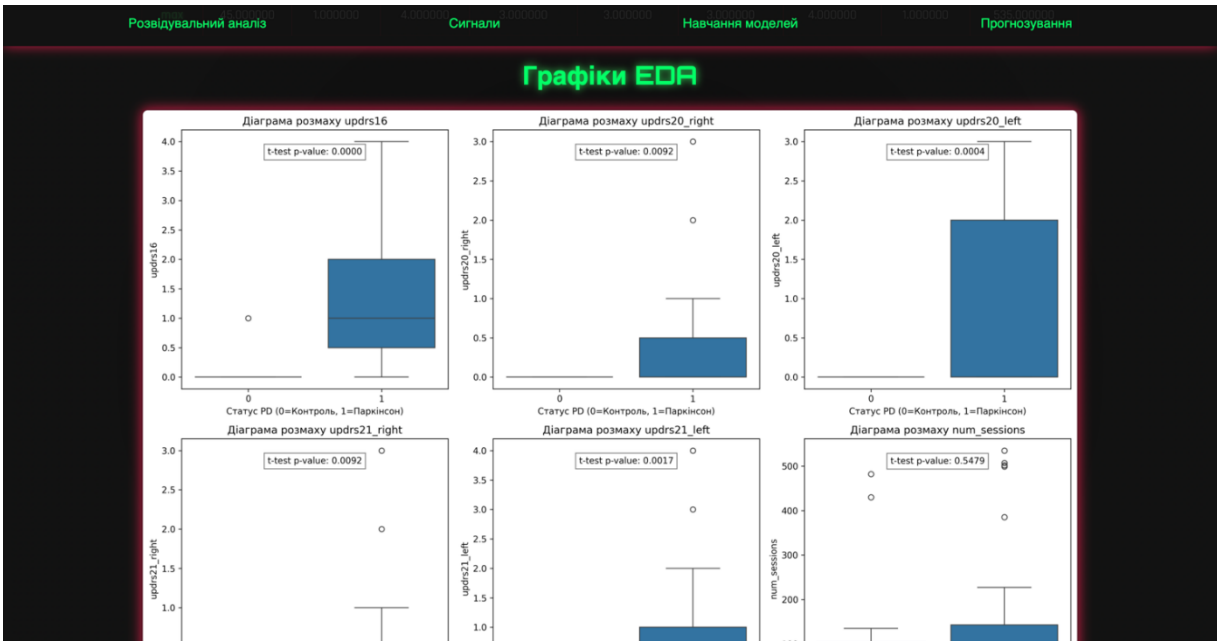


Рисунок 4.7 – Приклад відображення результатів розвідувального аналізу

Друга сторінка містить результати екстракції та обробки сигналів з акселерометрів, представлені результати як оброблених, так і необроблених сигналів для здорових та хворих пацієнтів (рис. 4.8).

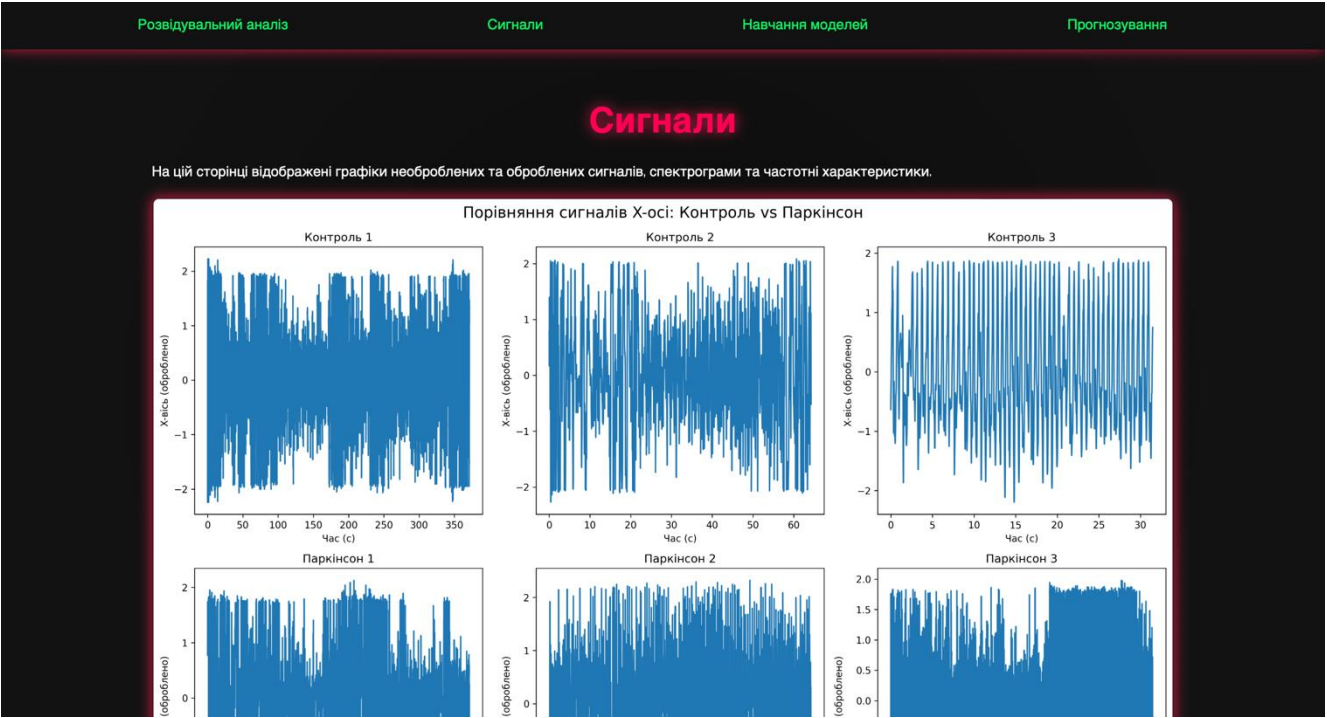


Рисунок 4.8 – Сторінка з сигналами з акселерометрів

Представлені відображені графіки (рис. 4.9).

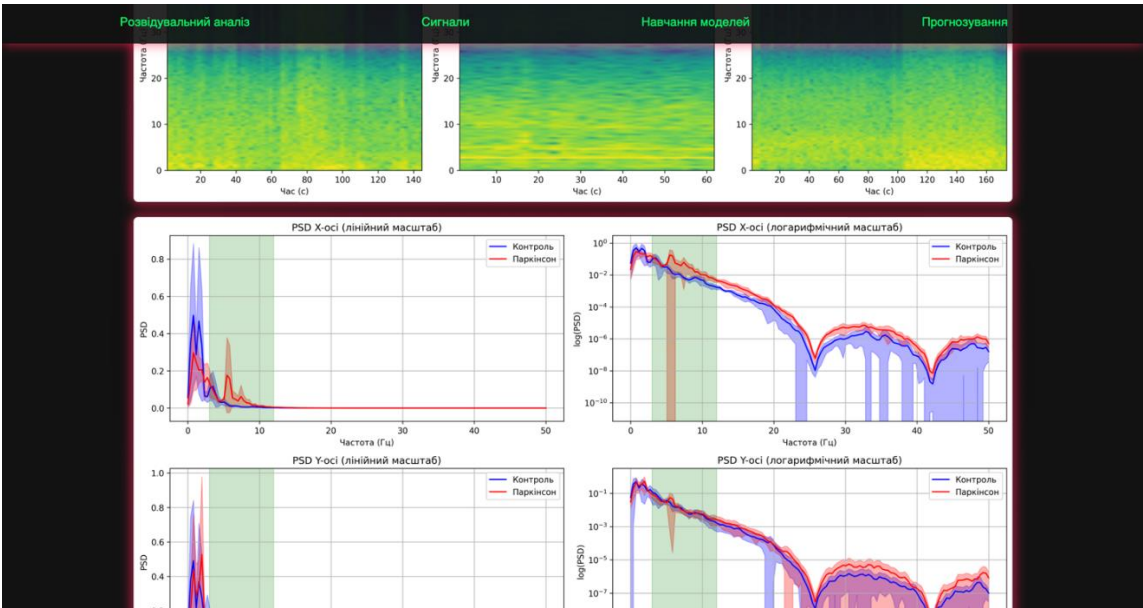


Рисунок 4.9 – Відображення графіків на сторінці сигналів

Наступна сторінка присвячена результатам навчання та тестування розроблених моделей (рис. 4.10).

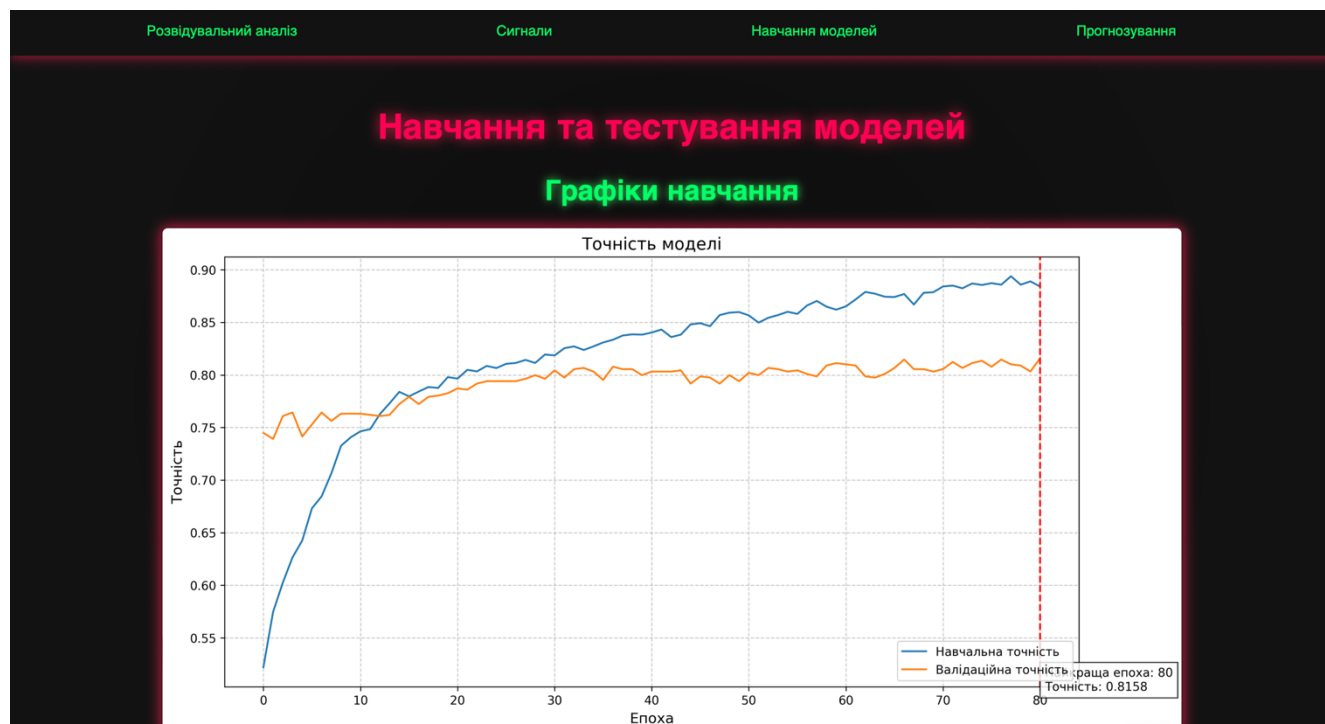


Рисунок 4.10 – Сторінка з результатами навчання та тестування моделей

Найбільш функціональна сторінка дозволяє виконувати прогнозування явища тремору для вхідних даних сигналу, що внесені користувачем (рис. 4.11).

Рисунок 4.11 – Сторінка прогнозування тремору

Користувач може вносити дані сигналу у вигляді окремих компонент або текстового документу та обирати модель (рис. 4.12).

Рисунок 4.12 – Форма для внесення вхідних даних

Також розроблена інструкція для користувачів, яка описує формат вхідних даних та їх специфіку (рис. 4.13).

Інструкція: Вхідні дані сигналів ✕

Введіть три вектори сигналів, які відповідають вимірюванням акселерометра по осях X, Y та Z. Серверна частина автоматично проведе обробку і витягне 69 ознак, а далі за допомогою pipeline (RobustScaler → QuantileTransformer → PCA) зведе їх до 23 компонент, які використовуються для прогнозування тремору.

Вимоги до даних:

- **Три вектори (X, Y, Z):** Кожен вектор повинен містити однакову кількість зразків. Наприклад, якщо вектор X має 100 значень, то вектори Y та Z теж повинні містити 100 чисел.
- **Формат введення:** Введіть числові значення, розділені комами, використовуючи крапку для десяткових дробів (наприклад: 0.12, -0.34, 0.56, ...).
- **Альтернативно:** Ви можете завантажити текстовий файл, де кожен з трьох рядків відповідає одній осі: перший - X, другий - Y, третій - Z.

Що описують вектори:

- **Вектор X:** Сигнал з акселерометра по осі X (одиниці виміру: g або m/s^2) - описує горизонтальні рухи або коливання.
- **Вектор Y:** Сигнал по осі Y - характеризує вертикальні рухи.
- **Вектор Z:** Сигнал по осі Z - описує рухи вперед/назад або інші напрямки залежно від орієнтації пристрою.

Обробка даних:

- Сформується 20-матриця, де перший стовпець - часовий ряд (розраховується автоматично при частоті дискретизації 100 Гц), а наступні три - значення сигналів X, Y та Z.

Рисунок 4.13 – Інструкція користувача

На двох тестових випадках протестовано роботу сервісу, перший сигнал отриманий у хворої людини, дані передаються у текстовому файлі (рис. 4.14).

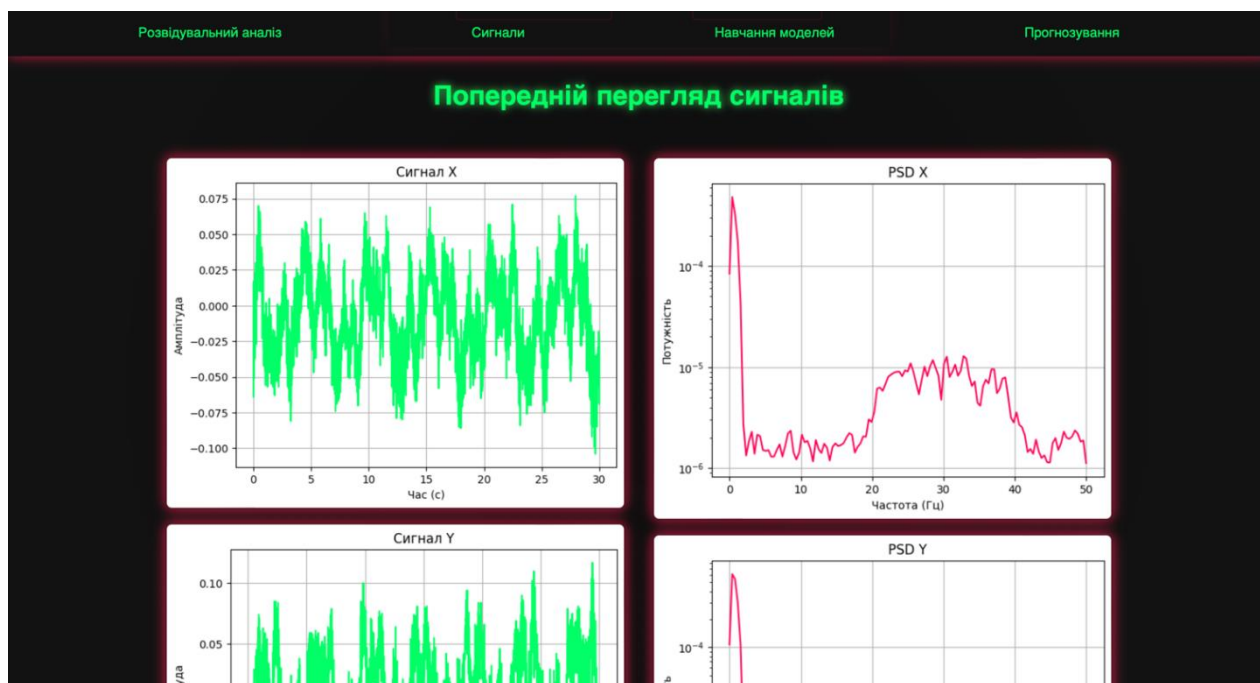


Рисунок 4.15 – Графік розподілу введено сигналу хворого пацієнта

Після завантаження даних, користувач може переглянути обраний сигнал та візуально оцінити його розподіл. Після цього обирається одна з трьох моделей та виконується прогнозування (рис. 4.16 – рис.4.18).

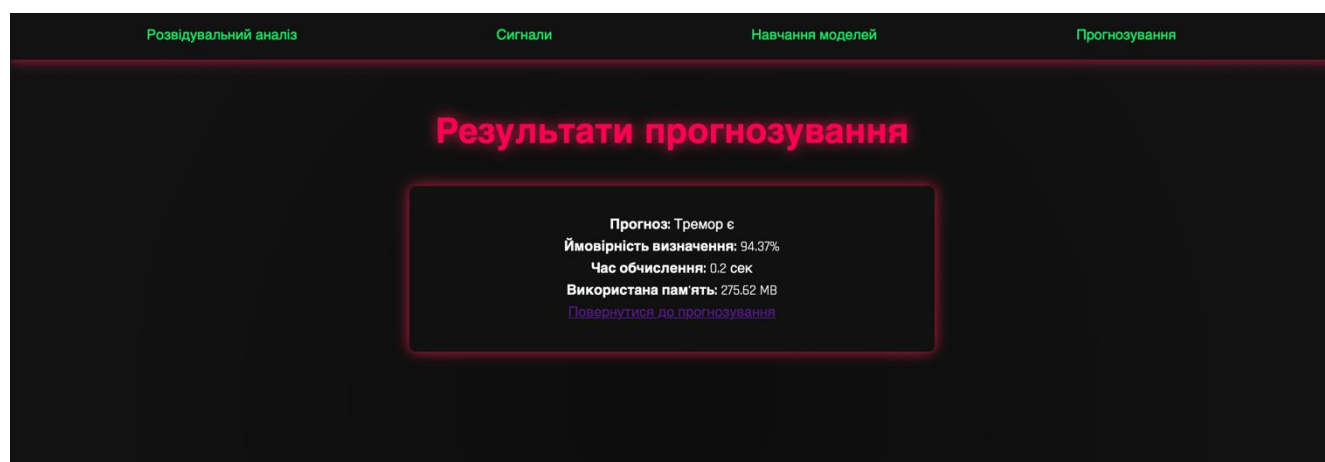


Рисунок 4.16 – Результат прогнозування моделлю FNN

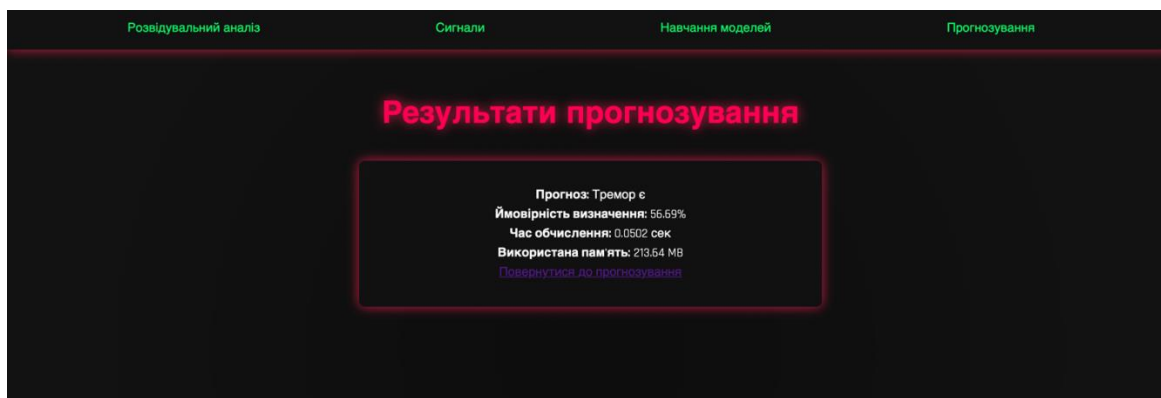


Рисунок 4.17 – Результат прогнозування моделлю DNN

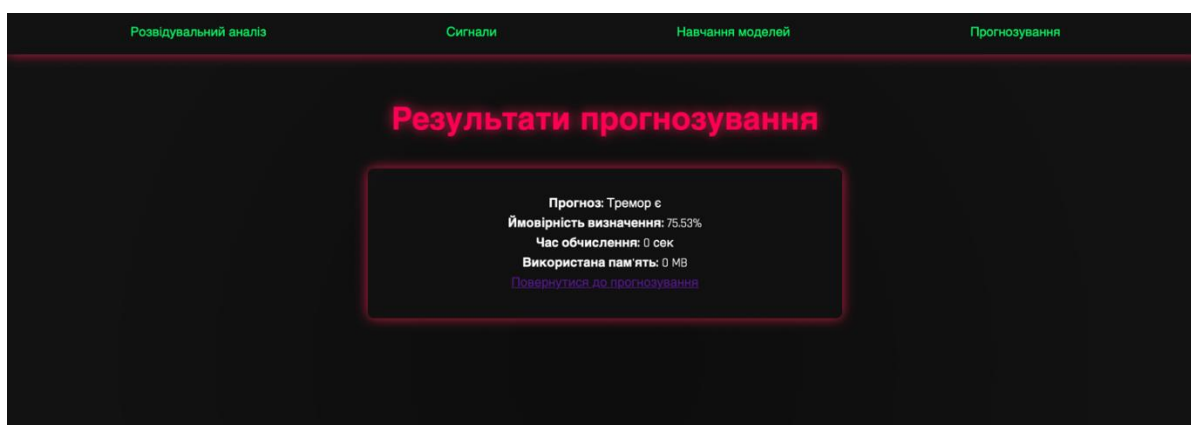


Рисунок 4.18 – Результат прогнозування моделлю Ensemble

Для вхідного сигналу хворої людини всі моделі коректно визначили наявність тремору. Аналіз здорової людини наведений на рис. 4.19.

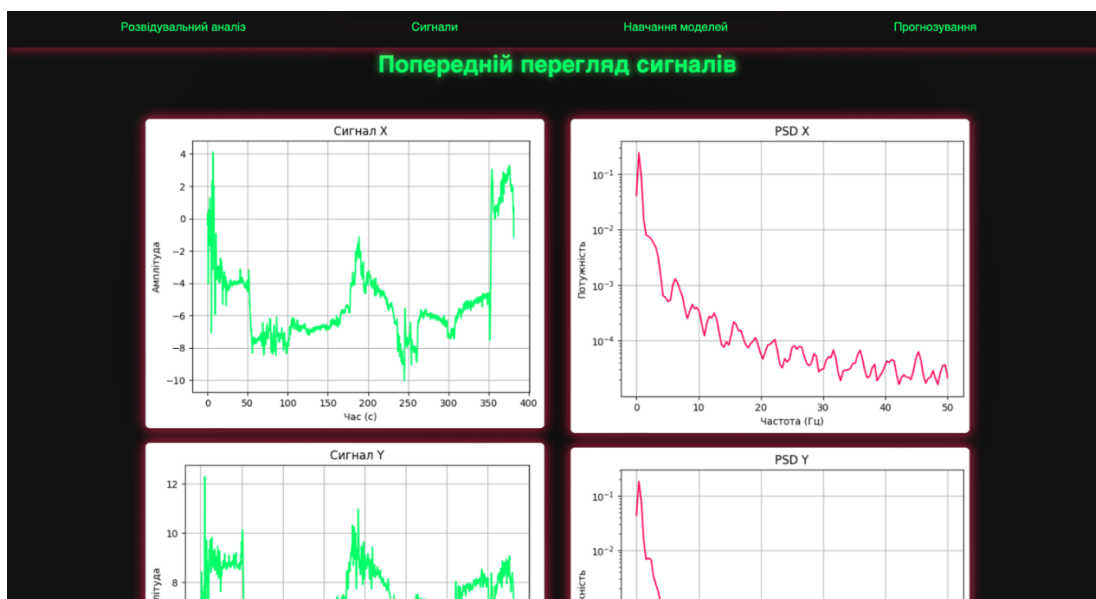


Рисунок 4.19 – Графік розподілу введено сигналу здорового пацієнта

Передамо вхідного сигналу на кожну з моделей показана на рис. 4.20 – рис.4.22.

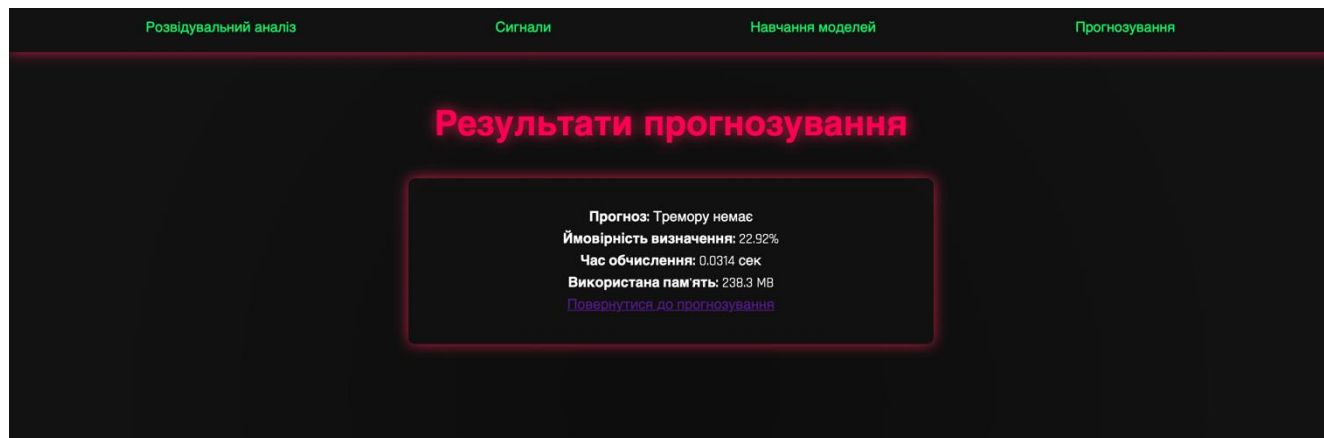


Рисунок 4.20 – Результат прогнозування моделлю FNN

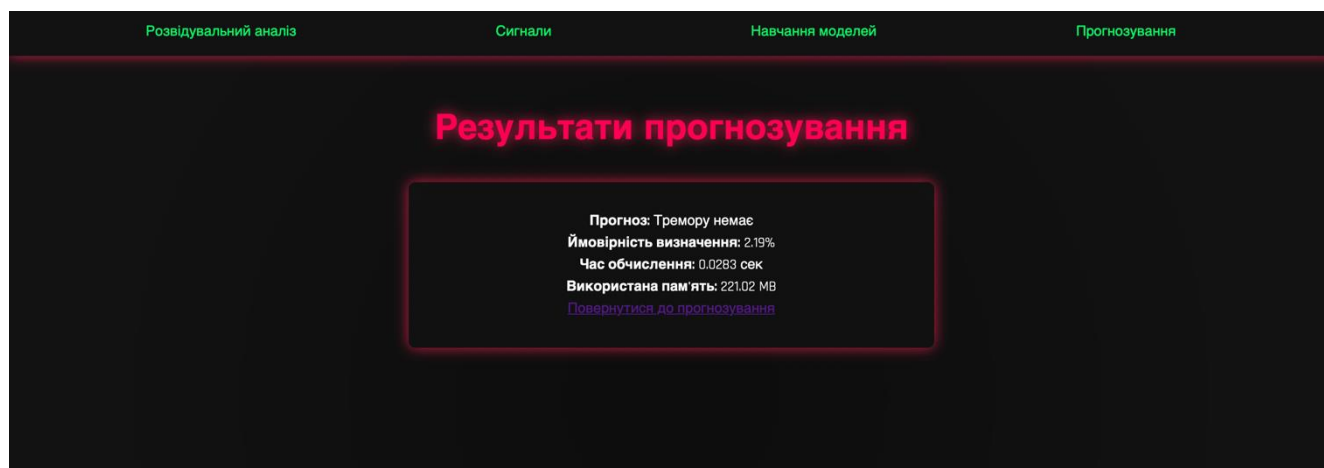


Рисунок 4.21 – Результат прогнозування моделлю DNN

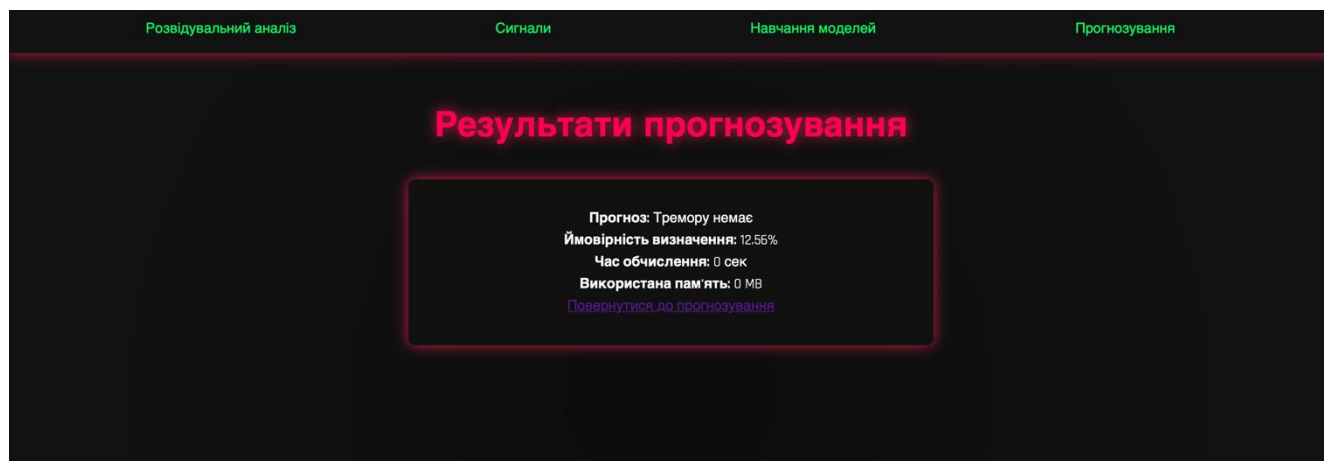


Рисунок 4.22 – Результат прогнозування моделлю Ensemble

Отже, для вхідного сигналу здорової людини, три моделі показали коректні результати. Важливо зазначити, що модель FNN може давати результати з більшою точністю, ансамблева модель працює набагато швидше та вимагає менше витрат пам'яті, особливо враховуючи незначну різницю в точності роботи моделей.

Висновки до розділу 4

У цьому розділі було розроблено функціональну веб-систему для обробки та аналізу сигналів із застосуванням моделей штучного інтелекту. Система поєднує сучасні інструменти машинного навчання, методи обробки даних і візуалізації, що забезпечує її ефективність і зручність для користувача. Реалізовано гнучку архітектуру, яка дозволяє масштабувати систему та адаптувати її під різні задачі прогнозування.

ЗАГАЛЬНІ ВИСНОВКИ

У дипломній роботі було проведено комплексне дослідження по застосуванню апарату штучних нейронних мереж та алгоритмів машинного навчання для прогнозування тремору у пацієнтів здорових та з хворобою Паркінсона, на основі даних, отриманих з інерційних сенсорів. У результаті виконання роботи було досягнуто таких результатів:

У першому розділі було досліджено основні фізіологічні та біомедичні аспекти розвитку та проходження тремору й хвороби Паркінсона, було визначено основні причини появи даного явища, та симптоми, що дозволяють передбачити потенційний розвиток патології. Також було досліджено специфіку роботи з інерційними даними, особливості їх накопичення та найбільш популярні методи обробки останніх. На заключному етапі першого розділу, було описано найбільш популярні архітектурні рішення для аналізу часових рядів та побудови моделей класифікації й прогнозування, було порівняно їх сильні та слабкі сторони, що дозволило обрати дві найбільш адекватні, для поставленої задачі, архітектури: FNN та DNN.

У другому розділі було розроблено та обґрунтовано комплексне рішення, направлене на обробку вхідних даних, отриманих від інерційних сенсорів: було представлено підхід, який дозволяє проводити обробку сигналів на різних рівнях: часовому, частотному та частотно-часовому. У результаті опису набору даних, було прийняте рішення про розробку конвеєрної моделі підготовки даних до подальшого моделювання, спершу було визначено розподіли основних полів датасету й перевірено статистичну значущість різниці ознак між групами (хворі та здорові пацієнти). Це дало змогу перейти до аналізу часових даних, у процесі проведеного аналізу було проаналізовано як частотно-амплітудні характеристики сигналів, так й проведено спектральний аналіз, що дозволив виявити закономірності в даних, а це в свою чергу, дозволило коректно побудувати моделі прогнозування та обґрунтувати архітектурне рішення.

У третьому розділі було проведено навчання розроблених моделей та ансамблевого методу, з метою оцінки якості проведеного дослідження. Процес

навчання був оптимізований, з метою уникнення перенавчання моделей. За результатами навчання були побудовані графіки точності та функції втрат моделей, що показали високу якість роботи використаних архітектурних рішень, а також їх здатність до узагальнення. Після цього було протестовано моделі з метою виявлення їх репрезентативності за допомогою кривих ROC-AUC, Precision-Recall, ці результати показали, що розроблені моделі можуть бути застосовані для розв'язання реальних задач, оскільки не враховують специфіки конкретного набору даних, а узагальнюють закономірності розподілу останніх. На заключному етапі було порівняно отримані результати, з аналогічними, що були отримані в іншому дослідженні, результати показали, що розроблені моделі і процеси обробки даних дозволили отримати набагато більш точні і якісніші результати.

В четвертому розділі було розроблено веб-інтегровану систему, що дозволяє користувачам взаємодіяти з розробленими моделями, шляхом внесення вхідного сигналу на веб-сервері. Система також дозволяє проводити аналіз набору даних, для навчання моделей та оцінювати їх якість навчання та тестування. Функціонал системи дозволяє також порівнювати кількісні показники якості роботи моделей в динамічних, прикладних умовах. Отже, результати досягнуті в проведеному дослідженні, можуть бути використані при розв'язанні реальних клінічних задач.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Abusrir A. H., Elsekaily W., Bohlega S. Tremor in Parkinson's disease: from pathophysiology to advanced therapies // Tremor and Other Hyperkinetic Movements. – 2022. – Т. 12. – С. 29.
2. Acharya D. et al. Application of adaptive Savitzky–Golay filter for EEG signal processing // Perspectives in science. – 2016. – Т. 8. – С. 677-679.
3. Ahsan M. M. et al. Effect of data scaling methods on machine learning algorithms and model performance // Technologies. – 2021. – Т. 9. – №. 3. – С. 52.
4. Alexandros Papadopoulos, Konstantinos Kyritsis, Sevasti Bostantjopoulou, Lisa Klingelhofer, Kallol Ray Chaudhuri and Anastasios Delopoulos. IMU data captured unobtrusively and in-the-wild by Parkinson's disease patients and healthy controls. Zenodo, Oct. 25, 2019. doi: 10.5281/zenodo.3519213.
5. Caramia C. et al. IMU-based classification of Parkinson's disease from gait: A sensitivity analysis on sensor location and feature selection // IEEE journal of biomedical and health informatics. – 2018. – Т. 22. – №. 6. – С. 1765-1774.
6. Chauhan V. K., Dahiya K., Sharma A. Problem formulations and solvers in linear SVM: a review // Artificial Intelligence Review. – 2019. – Т. 52. – №. 2. – С. 803-855.
7. Chen B. R. et al. A web-based system for home monitoring of patients with Parkinson's disease using wearable sensors // IEEE transactions on biomedical engineering. – 2010. – Т. 58. – №. 3. – С. 831-836.
8. Chén O. Y. et al. Building a machine-learning framework to remotely assess Parkinson's disease using smartphones // IEEE Transactions on Biomedical Engineering. – 2020. – Т. 67. – №. 12. – С. 3491-3500.
9. Chumakova E. V. et al. Comparison of the Application of FNN and LSTM Based on the Use of Modules of Artificial Neural Networks in Generating an Individual Knowledge Testing Trajectory // Journal Européen des Systèmes Automatisés. – 2023. – Т. 56. – №. 2.

10. Dirkx M. F. et al. The nature of postural tremor in Parkinson disease // *Neurology*. – 2018. – T. 90. – №. 13. – C. e1095-e1103.
11. Fitriyah H., Budi A. S. Outlier detection in object counting based on hue and distance transform using median absolute deviation (MAD) // 2019 international conference on sustainable information engineering and technology (SIET). – IEEE, 2019. – C. 217-222.
12. Ibrahim A. et al. Real-time voluntary motion prediction and Parkinson's tremor reduction using deep neural networks // *IEEE Transactions on Neural Systems and Rehabilitation Engineering*. – 2021. – T. 29. – C. 1413-1423.
13. Kim T. K. T test as a parametric statistic // *Korean journal of anesthesiology*. – 2015. – T. 68. – №. 6. – C. 540-546.
14. Kurita T. Principal component analysis (PCA) // *Computer vision: a reference guide*. – Cham: Springer International Publishing, 2021. – C. 1013-1016.
15. Lakhe A. S. et al. Spectral trimming technique: a new approach for suppressing motion artefacts in stress electrocardiography // *Journal of Medical Engineering & Technology*. – 2020. – T. 44. – №. 6. – C. 338-345.
16. Li G. et al. Understanding error propagation in deep learning neural network (DNN) accelerators and applications // *Proceedings of the international conference for high performance computing, networking, storage and analysis*. – 2017. – C. 1-12.
17. MacFarland T. W. et al. Mann–whitney u test // *Introduction to nonparametric statistics for the biological sciences using R*. – 2016. – C. 103-132.
18. Panebianco G. P. et al. Analysis of the performance of 17 algorithms from a systematic review: Influence of sensor position, analysed variable and computational approach in gait timing estimation from IMU measurements // *Gait & posture*. – 2018. – T. 66. – C. 76-82.
19. Patel S. et al. Monitoring motor fluctuations in patients with Parkinson's disease using wearable sensors // *IEEE transactions on information technology in biomedicine*. – 2009. – T. 13. – №. 6. – C. 864-873.

20. Peng X. et al. Multi-Scale and Multi-Level Feature Assessment Framework for Classification of Parkinson's Disease State From Short-Term Motor Tasks // IEEE Transactions on Biomedical Engineering. – 2025.
21. Peterson R. A., Cavanaugh J. E. Ordered quantile normalization: a semiparametric transformation built for the cross-validation era // Journal of applied statistics. – 2020.
22. Pham T. D., Yan H. Tensor decomposition of gait dynamics in Parkinson's disease // IEEE Transactions on Biomedical Engineering. – 2017. – T. 65. – №. 8. – C. 1820-1827.
23. Pirker W. et al. Pharmacological treatment of tremor in Parkinson's disease revisited // Journal of Parkinson's Disease. – 2023. – T. 13. – №. 2. – C. 127-144.
24. Rigatti S. J. Random forest // Journal of Insurance Medicine. – 2017. – T. 47. – №. 1. – C. 31-39.
25. Roux B. Transition rate theory, spectral analysis, and reactive paths // The Journal of Chemical Physics. – 2022. – T. 156. – №. 13.
26. Salarian A. et al. Quantification of tremor and bradykinesia in Parkinson's disease using a novel ambulatory monitoring system // IEEE Transactions on biomedical engineering. – 2007. – T. 54. – №. 2. – C. 313-322.
27. Salur M. U., Aydın İ. A soft voting ensemble learning-based approach for multimodal sentiment analysis // Neural Computing and Applications. – 2022. – T. 34. – №. 21. – C. 18391-18406.
28. Shahed J., Jankovic J. Exploring the relationship between essential tremor and Parkinson's disease // Parkinsonism & related disorders. – 2007. – T. 13. – №. 2. – C. 67-76.
29. Shi J. et al. Cascaded multi-column RVFL+ classifier for single-modal neuroimaging-based diagnosis of Parkinson's disease // IEEE Transactions on Biomedical Engineering. – 2018. – T. 66. – №. 8. – C. 2362-2371.
30. Shouran M., Elgamli E. Design and implementation of Butterworth filter // International Journal of Innovative Research in Science, Engineering and Technology. – 2020. – T. 9. – №. 9. – C. 7975-7983.

31. Staudemeyer R. C., Morris E. R. Understanding LSTM--a tutorial into long short-term memory recurrent neural networks // arXiv preprint arXiv:1909.09586. – 2019.
32. Thenganatt M. A., Louis E. D. Distinguishing essential tremor from Parkinson's disease: bedside tests and laboratory evaluations // Expert review of neurotherapeutics. – 2012. – T. 12. – №. 6. – C. 687-696.
33. Vaillancourt D. E., Slifkin A. B., Newell K. M. Regularity of force tremor in Parkinson's disease // Clinical Neurophysiology. – 2001. – T. 112. – №. 9. – C. 1594-1603.
34. Wacker M., Witte H. Time-frequency techniques in biomedical signal analysis // Methods of information in medicine. – 2013. – T. 52. – №. 04. – C. 279-296.
35. Zhao Y., Horemuz M., Sjöberg L. E. Stochastic modelling and analysis of IMU sensor errors // Archiwum Fotogrametrii, Kartografii i Teledetekcji. – 2011. – T. 22. – C. 437-449.
36. Zhong Y. et al. A review on pathology, mechanism, and therapy for cerebellum and tremor in Parkinson's disease // npj Parkinson's Disease. – 2022. – T. 8. – №. 1. – C. 82.