

**НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ
«КІЇВСЬКИЙ ПОЛТЕХНІЧНИЙ ІНСТИТУТ»**

Факультет прикладної математики

(повна назва інституту/факультету)

Кафедра прикладної математики

(повна назва кафедри)

«На правах рукопису»
УДК 51-76

«До захисту допущено»

Завідувач кафедри

О. Р. Чертов

(підпис)

(ініціали, прізвище)

«____» 2015 р.

Магістерська дисертація
на здобуття ступеня магістра

зі спеціальності 8.04030101 «Прикладна математика»

(код і назва)

на тему: **АНАЛІЗ ЕЛЕКТРОЕНЦЕФАЛОГРАМ З ВИКОРИСТАННЯМ
КОНВОЛЮТИВНИХ НЕЙРОННИХ МЕРЕЖ**

Виконала: студентка 2-го курсу, групи КМ-31м
(шифр групи)

Коршунова Ірина Анатоліївна

(прізвище, ім'я, по батькові)

(підпис)

Науковий керівник зав. каф., д-р техн. наук, доцент Чертов О. Р.

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

(підпис)

Консультант із нормоконтролю старший викладач Мальчиков В.В.

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

(підпис)

Рецензент професор, д-р техн. наук, проф. Кулаков Ю. О.

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

(підпис)

Засвідчую, що у цій магістерській
дисертації немає запозичень з праць інших
авторів без відповідних посилань.
Студентка _____
(підпис)

Київ – 2015 року

**Національний технічний університет України
«Київський політехнічний інститут»**

Інститут (факультет) Факультет прикладної математики
(повна назва)

Кафедра прикладної математики
(повна назва)

Рівень вищої освіти – другий (магістерський)

Спеціальність 8.04030101 «Прикладна математика»
(код і назва)

ЗАТВЕРДЖУЮ
Завідувач кафедри
О. Р. Чертов
(підпис) (ініціали, прізвище)

« » 2015р.

**ЗАВДАННЯ
на магістерську дисертацію студенту
Коршуновій Ірині Анатоліївні**

(прізвище, ім'я, по батькові)

1. Тема дисертації «АНАЛІЗ ЕЛЕКТРОЕНЦЕФАЛОГРАМ З
ВИКОРИСТАННЯМ КОНВОЛЮТИВНИХ НЕЙРОННИХ МЕРЕЖ»,
науковий керівник дисертації Чертов Олег Романович, д.т.н., доц.,
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

затверджені наказом по університету від «20» березня 2015 р. №785-С

2. Термін подання студентом дисертації «18» червня 2015 р.
3. Об'єкт дослідження: методи класифікації ЕЕГ сигналів
4. Предмет дослідження: конволютивні нейронні мережі, їх архітектури, методи тренування, регуляризації та аналізу.
5. Перелік завдань, які потрібно розробити:
 - проаналізувати підходи, що застосовуються для аналізу багатовимірних часових рядів і, зокрема, ЕЕГ
 - дослідити природу ЕЕГ даних з метою виявлення потенціальних складностей в роботі з ними
 - розробити метод, що здатен вивчати релевантну високорівневу репрезентацію ЕЕГ даних з низькорівневої або сирого ЕЕГ сигналу
 - здійснити апробацію методу та проаналізувати його результати за різними критеріями оцінки

6. Орієнтовний перелік ілюстративного матеріалу:

- ілюстрація операцій конволюції, субдискретизації та з'єднання шарів в конволютивних нейронних мережах
- схематичне зображення архітектур запропонованих моделей
- таблиця порівняння результатів двох архітектур
- навчальні криві під час процесу тренування моделей
- графіки передбачень моделі на тестовій вибірці
- 2D візуалізація ЕЕГ сегментів
- графіки чутливості та специфічності класифікаторів залежно від порогу класифікуючого правила

7. Орієнтовний перелік публікацій:

«Конволютивні нейронні мережі для прогнозування епілептичних методів», VII наукова конференція магістрантів та аспірантів «Прикладна математика та комп’ютинг» ПМК-2015 (Київ, 15-17 квітня 2014 р.);

«Detecting epileptic seizures from raw EEG data using convolutional neural networks», 17та міжнародна конференція «Системний аналіз та інформаційні технології» SAIT-2015 (Київ, 22-25 червня 2015 р.).

8. Консультанти розділів дисертації

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
нормоконтроль	Мальчиков В.В., старший викладач		

9. Дата видачі завдання «15» жовтня 2012 р.

Календарний план

№ з/п	Назва етапів виконання магістерської дисертації	Термін виконання етапів магістерської дисертації	Примітка
1.	Визначити тематику дослідження	15.10.12	
2.	Сформулювати об’єкт та предмет дослідження	15.12.12	
3.	Визначити структуру магістерської дисертації	01.03.12	

4.	Вивчення літератури за тематикою роботи та збір даних	31.03.12	
5.	Дослідження задачі класифікації зображень за допомогою конволютивних нейронних мереж	01.05.12	
6.	Підготовка матеріалів першого розділу магістерської дисертації	01.06.12	
7.	Дослідження ососбливостей ЕЕГ даних	15.09.14	
8.	Розробка методу передбачення епілептичних нападів на основі конволютивних нейронних мереж	01.11.14	
9.	Апробація методу у змаганні «American Epilepsy Society Seizure Prediction Challenge»	17.11.14	
10.	Підготовка матеріалів другого та третього розділів магістерської дисертації	31.12.14	
11.	Імплементація методу класифікації ЕЕГ на основі регресійних моделей, їх аналіз	01.02.15	
12.	Підготовка матеріалів четвертого розділу магістерської дисертації	01.03.15	
13.	Проведення порівняльного аналізу методів	15.03.15	
14.	Підготовка матеріалів четвертого розділу магістерської дисертації	01.04.15	
15.	Підготовка тез доповіді на конференції ПМК-2015 та SAIT-2015	15.04.15	
16.	Оформлення текстової і графічної частини магістерської дисертації	31.05.15	

Студент

(підпис)

Коршунова І.А.

(ініціали, прізвище)

Науковий керівник дисертації

(підпис)

Чертов О.Р.

(ініціали, прізвище)

РЕФЕРАТ

Актуальність теми. Електроенцефалографія має широкий спектр застосувань в медицині, нейродослідженнях, є основою систем нейрокомп'ютерних інтерфейсів та ін. Незважаючи на значний успіх в деяких задачах аналізу ЕЕГ сигналів, наприклад декодування візуальних стимулів, можливість передбачення епілептичних нападів на основі ЕЕГ залишається спірним питанням, але надзвичайно актуальним для 1% населення, що страждає на епілепсію.

Основною складністю в області класифікації ЕЕГ є відсутність універсальних характеристик сигналу, які, при використанні в класифікуючих алгоритмах, здатні забезпечити високу надійність системи. На відміну від звичайних багатовимірних даних, статистичний аналіз яких можливий після зменшення розмірності, ЕЕГ є нестационарним багатоканальним часовим рядом, який в часової області може аналізувати існуючими методами напочуд складно. Саме з цієї причини інженерія характеристик відіграє важливу роль в даній області, і тому сповільнює її прогрес.

В останні роки популярними стали методи на основі глибоких нейронних мереж, що сьогодні тримають першість в задачах, де отримання високорівневої репрезентації даних є ключовим компонентом до успіху. Саме релевантне представлення ЕЕГ сигналів необхідне для їх ефективного аналізу, тому адаптація нейронних мереж в даній області є першочерговою проблемою, рішення якої здатне принести бажаний прогрес.

Об'єктом дослідження є методи класифікації ЕЕГ сигналів

Предметом дослідження є конволютивні нейронні мережі, їх архітектури, методи тренування, регуляризації та аналізу.

Мета роботи: розробити метод, що може класифікувати ЕЕГ сегменти, що відповідають різним станам мозкової активності пов'язаної з епілептичними нападами. Розробити методи оцінки якості класифікації та методи валідації моделей. Провести порівняльний аналіз існуючих методів з запропонованими.

Методи дослідження. В роботі використовуються методи ком-

п'ютерного та імітаційного моделювання, теорії цифрової обробки сигналів та машинного навчання, методи теорії ймовірностей та математичної статистики.

Наукова новизна роботи полягає в наступному:

1. Розроблена новітня архітектура нейронної мережі, яка здатна конкурувати з існуючими методами аналізу ЕЕГ і потенційно має більші можливості для подальшого вдосконалення, оскільки елімінує важливість процесу інженерії ознак

2. Вперше були продемонстровані недоліки валідаційних схем, які використовувались для оцінки моделей для передбачення нападів.

Практична цінність отриманих в роботі результатів полягає в тому, що розроблений метод та його програмна реалізація в перспективі можуть бути застосовані в системах передбачення епілептичних нападів. На жаль, це відбудеться не в скороому майбутньому, оскільки експерименти з розробки подібних систем знаходяться на початкових етапах.

Апробація роботи. Основні положення і результати роботи були представлені на VII науковій конференції “Прикладна математика та комп’ютинг – ПМК’2015” (Київ, 15-17 квітня 2015 р.) та на 17-й міжнародній конференції “System analysis and information technologies” (Київ, 22–25 червня 2015 року). Однак сама робота пов’язана з міжнародним змаганням “American Epilepsy Society Seizure Prediction Challenge” (Kaggle, 25 серпня - 17 листопада 2014), де розроблений метод зайняв 10е місце з 504ох команд.

Структура та обсяг роботи. Магістерська дисертація складається з вступу, п’ятьох розділів, висновків та додатку.

У вступі надано загальну характеристику роботи, виконано оцінку сучасного стану проблеми, обґрунтовано актуальність напрямку досліджень, сформульовано задачу досліджень, показано наукову новизну отриманих результатів та їх цінність.

У першому розділі приведені загальні теоретичні відомості з області машинного навчання, детально описані методи класифікації на основі лінійних моделей та нейронних мереж, коротко представлені необхідні поняття з теорії цифрової обробки сигналів та розглянутий метод візуа-

лізації багатовимірних даних.

У другому розділі була описана суть змагання “American Epilepsy Society Seizure Prediction Challenge”: які дані були використані та за яким критерієм оцінювались моделі передбачення епілептичних нападів.

У третьому розділі детально було пояснено запропонований метод класифікації ЕЕГ, що базується на конволютивних нейронних мережах. Також в даному розділі були проаналізовані результати, отримані різними архітектурами нейронних мереж.

У четвертому розділі описуюється метод на основі лінійних класифікаторів. Наводиться алгоритм валідації моделей, під час якого були проілюстровані властивості ЕЕГ, що значно ускладнюють процес аналізу даного сигналу.

У п'ятому розділі були порівняні методи на основі нейронних мереж та лінійних класифікаторів за різними критеріями, також був зроблений огляд методів, запропонованих іншими учасниками змагання.

У висновках були підсумувані отримані результати роботи.

У додатках наведені ілюстративні матеріали та лістинги програми.

Робота виконана на 94 аркушах, містить 2 додатка та посилання на список використаних літературних джерел з 35 найменувань. У роботі наведено 26 рисунків та 7 таблиць.

Ключові слова: машинне навчання, конволютивні нейронні мережі, ЕЕГ, епілептичні напади.

ABSTRACT

Topic urgency. Electroencephalography has a wide spectrum of applications in medicine, neuroscience; it is a basis for brain-computer interfaces and many others. Despite the significant success in some of the tasks of the EEG analysis, e.g. the decoding of visual stimuli, the possibility of predicting epileptic seizures remains a controversial problem, but it is extremely important for the 1% of people who suffer from epilepsy.

The main issue in classifying EEG is the absence of a universal set of features able to give rise to reliable classification algorithms. For conventional multi-dimensional data statistical analysis is usually possible after the dimensionality reduction. However, EEG is a multi-channel non-stationary time series, for which the analysis in a time-domain analysis is extremely difficult. For this reason, feature engineering still plays an important role, the downside of which is that it slows down the progress in the field.

In recent years deep neural networks became popular; these models are currently breaking state-of-the-art results in almost any problem, where having a good high-level data representation is crucial. Learning an efficient representation of EEG signals would potentially lead to the desired progress.

Object of research is classification of EEG signals and the performance analysis of such systems.

Subject of research is convolutional neural networks, their architectures, training and regularization techniques, analysis of their performance in the context of epileptic seizure prediction.

Research objective: to develop a method that can classify EEG clips corresponding to different states of brain activity associated with epileptic seizures. Develop model validation techniques and analyse performance measures. Perform a comparative study of different methods.

Research methods. Methods of computer modeling and simulation, a theory of digital signal processing and machine learning, methods of probability theory and mathematical statistics.

Scientific novelty of the work is as follows:

1. A novel neural network architecture that is able to compete with existing methods of EEG analysis. It has a great potential for further improvements because it eliminates the importance of feature engineering.

2. For the first time we demonstrated the shortcomings of the validation schemes that were used to evaluate seizure prediction classification models.

Practical value of the results is that proposed method and its implementation in the future can be used in seizure advisory systems. Unfortunately, it will not happen soon, since these experiments are on their early stages.

Approbation. The results were presented at the VII Conference "Applied mathematics and computing - AMC-2015" (Kiev, 15-17 April 2015) and at the 17th International Conference "System analysis and information technologies" (Kiev, 22-25 June 2015). However, the work is related to the international competition "American Epilepsy Society Seizure Prediction Challenge" (Kaggle, 25 August - 17 November 2014), where our method got the 10th place out of 504 teams.

Structure and content of the thesis. Master's thesis consists of an introduction, five chapters, conclusions and appendices.

Introduction provides a general overview of the current state of the problem, the relevance of the research objectives, and the scientific novelty of the results.

The first section contains theoretical basics of the machine learning, detailed description of the some statistical classification methods and neural networks models. Moreover, it provides essential concepts of digital signal processing theory and explains a method for high-dimensional data visualization.

In the second chapter, we describe the essence of the competition "American Epilepsy Society Seizure Prediction Challenge". Here we explore the structure of the dataset and explain how the models were scored.

In the third chapter, we explain our method of classifying EEG with convolutional neural networks. Also in this section, we analyzed the results obtained by different network architectures.

The fourth chapter describes an approach based on linear classifiers. Here we analyse the validation methods while highlighting the properties of

EEG signals, which make the validation procedures unreliable.

In the fifth chapter, using different criteria we compared two methods based on neural networks and linear discriminant analysis. Also we provided an overview of the methods proposed by other competitors.

In the conclusions, we made a summary of our work.

The appendices provide illustrative materials and listings.

The thesis is presented on 94 pages, it contains 2 appendices and references to the literature used (35 items). 26 figures and 7 tables are given in the thesis.

Keywords: machine learning, convolutional neural networks, EEG, epileptic seizures.



Зміст

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ	13
ВСТУП	14
1 ТЕОРЕТИЧНІ ВІДОМОСТІ	18
1.1 Лінійні класифікатори	18
1.1.1 Лінійний дискримінантний аналіз	18
1.1.2 Логістична регресія	20
1.2 Штучні нейронні мережі	21
1.2.1 Багатошаровий перцептрон	21
1.2.2 Тренування ШНМ	22
1.2.3 Алгоритм зворотнього поширення помилки у випадку БШП	24
1.2.4 Функції активації	25
1.2.5 Конволютивні нейронні мережі	27
1.2.6 Перенавчання	30
1.3 Дискретне перетворення Фур'є	32
1.4 t-SNE	33
1.5 Висновки	34
2 AES SEIZURE PREDICTION CHALLENGE	35
2.1 Дані	35
2.2 Оцінка якості класифікації	37
2.3 Висновки	39
3 МЕТОД НА ОСНОВІ КОНВОЛЮТИВНИХ НЕЙРОННИХ МЕРЕЖ	40
3.1 Обробка даних	40
3.2 Архітектури моделей	42
3.3 Тренування моделей	44
3.4 Калібрація передбачень	46
3.5 Результати	46

3.6 Висновки	48
4 МЕТОД НА ОСНОВІ ЛІНІЙНИХ КЛАСИФІКАТОРІВ	49
4.1 Обробка даних	49
4.2 Навчання моделі	50
4.3 Валідація моделі	50
4.4 Аналіз моделі	53
4.5 Висновки	55
5 ПОРІВНЯЛЬНИЙ АНАЛІЗ РОЗРОБЛЕНИХ МЕТОДІВ	57
5.1 Порівняння КНН та ЛДА методів	57
5.2 Аналіз інших підходів	62
5.3 Висновки	64
ВИСНОВКИ	65
ПЕРЕЛІК ПОСИЛАНЬ	68
ДОДАТКИ	72
Додаток А. Лістинг програми	72
Додаток Б. Ілюстративний матеріал	85



ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ

БШП – багатошаровий перцептрон

ЕЕГ – електроенцефалограма

інтеріктальний – період нормальної мозкової активності

КНН – конволютивна нейронна мережа

ЛДА – лінійний дискримінантний аналіз

ЛР – логістична регресія

преіктальний – той, що передує епілептичному нападу

ШНМ – штучна нейронна мережа

ADADELTA – метод градієнтного спуску з адаптивною швидкістю

AES – (American Epilepsy Society) Американське товариство епілепсії

AUC – (area under ROC curve) площа під ROC кривою

ReLU – (rectified linear units) нейрони з ректифікаційною функцією активації

ROC – (receiver operating characteristic) крива помилок класифікатора

TNR – (true negative rate) специфічність класифікатора

TPR – (true positive rate) чутливість класифікатора

ВСТУП

Електроенцефалографія має широкий спектр застосувань в медицині, нейродослідженнях, є основою систем нейро-комп'ютерний інтерфейсу та ін. Незважаючи на значний успіх в деяких задачах аналізу ЕЕГ сигналів, наприклад декодування візуальних стимулів в системах комунікації із зовнішніми електронними присторями [1], існують задачі, можливість розв'язку яких залишається предметом гарячих дебатів, оскільки за останні роки видимого прогресу досягнуто не було [2]. Однією з таких задач, на якій була сконцентрована увага, є передбачення епілептичних нападів.

Епілепсія – один з найпоширеніших неврологічних розладів, який вражає близко 1% населення. Дано хвороба характеризується виникненням спонтанних нападів, форма яких може варіюватись від короткотривалої втрати свідомості до тяжких конвульсій. Приблизно у 2/3 випадках запобігти нападам можна за рахунок медикаментозного лікування, однак для решти пацієнтів медикаменти є недієвими. На додачу, лікарські засоби майже завжди викликають побічні ефекти [3]. Таким чином, люди хворі на епілепсію мають суттєво нижчий рівень життя, що зумовлено постійною тривогою у зв'язку з можливістю раптового епілептичного нападу. Системи передбачення епілептичних нападів мають на меті покращити життя людей, які страждають на епілепсію: пацієнти зможуть уникати небезпечних занять, наприклад, плавання, кермування автомобілем, а також приймати медикаменти лише у разі необхідності. Очевидно, що дані системи потребують надійних алгоритмів, здатних визначати періоди підвищеної ймовірності майбутнього епілептичного нападу.

Епілептичні напади зумовлюються аномальною електричною активністю клітин мозку, тому електроенцефалограма (ЕЕГ) найчастіше використовується для детекції нападів або їх передбачення. ЕЕГ – це багатоканальний запис електричної активності мозку, і особливо цінним матеріалом для дослідження природи епілептичних нападів є інtrakраніальна ЕЕГ (iEEG). У випадку iEEG електроди імплантується безпосередньо в

тканини мозку, що дозволяє отримати якісніший сигнал ніж при екстракраніальному вимірюванні ЕЕГ.

Декілька робіт підтримують гіпотезу про існування особливого періоду, що передує епілептичному нападу. Згідно даних досліджень, цей період характеризується особливими спектральними патернами а також формою ЕЕГ хвиль [4]. Однак, можливість передбачення нападів залишається предметом дискусій, оскільки в попередніх роботах було виявлено багато статистичних недоліків, що зводить на нівець майже 10 років досліджень в даній області [2]. На даний момент не існує алгоритму, який би забезпечив рівень якості розпізнавання різних ЕЕГ станів, доцільного для медичного впровадження.

Кілька років тому провідними науково-дослідними інститутами у галузі епілепсії був заснований International Epilepsy Electrophysiology Portal, що має на меті створення єдиної платформи та баз ЕЕГ даних для прозорої оцінки алгоритмів аналізу ЕЕГ. В рамках даного проекту, було організовано міжнародне змагання “American Epilepsy Society Seizure Prediction Challenge”. Мета змагання полягала у виявленні найкращої моделі для дискримінації преіктальних та інтеріктальних ЕЕГ сегментів – даних, що передують нападу та таких, що не містять слідів епілептичної активності. Змагання мало призовий фонд в 25000\$ і проводилось на базі платформи Kaggle з 25 серпня до 17 листопада, 2014 року [5]. В даній роботі представлений метод, що у даному змаганні посів 10-те місце з 504ох команд.

Основою запропонованого методу [6] є конволютивні нейронні мережі (КНН) – тип глибоких штучних нейронних мереж, що в останні роки революціонізував область комп’ютерного бачення [7, 8]. Для аналізу зображень та відео застосування конволютивних нейромереж є предметом активних досліджень, однак, для аналізу фізіологічних даних, або в ширшому сенсі, багатовимірних часових рядів конволютивні нейромережі і глибокі нейронні мережі майже не були адаптовані [9]. Це пояснюється складністю роботи з сирими, необробленими ЕЕГ даними і, як простіше вирішення проблеми, використання низькорівневої презентації, яка може слугувати входом для нейромережі. Знаходження останньої також

є однією з непростих задач, але саме даний підхід буде детально описаний в даній роботі, оскільки він був застосований у вищезгаданому змаганні.

Як нам відомо існує лише одна робота [10], де КНН були використані для предбачення епілептичних нападів. Однак, всупереч отримання неперевершених результатів, вони викликають сумніви, оскільки автори серед кількох методів обирають найкращий на тестовій множині, таким чином перетворюючи її на валідаційну.

Найбільш близькою роботою до суті змагання є [4], оскільки авторами використовувалась підмножина даних, що входила в тренувальну вибірку для змагання. Однак, організатори не надали інформації які точно дані були використані, що не дозволяє порівняти методи. Більшою мірою процес первинної фільтрації ЕЕГ та ідея використання спектральних характеристик ЕЕГ були адаптовані з [4], хоча вперше детальний аналіз спектральних характеристик для задачі передбачення епілептичних нападів був описаний в [11].

Варто зауважити, що в області аналізу ЕЕГ існують інші проблеми, наприклад детекції епілептичного нападу. Данна проблема є простішою, оскільки іктальний період візуально досить чітко виражений на ЕЕГ, на відміну від преіктального. Проблема детекції ширше адресована в літературі, але використання методів для виявлення нападів напряму у задачі прогнозування нападів не дає бажаних результатів [3].

Однією з важливих задач, яка потенційно може бути використана як складова частина для прогнозування/детекції нападів або в системах нейрокомп'ютерних інтерфейсів, які теж базуються на ЕЕГ, є класифікація форм ЕЕГ хвиль. В [12] описується підхід на основі глибоких нейронних мереж, де входом є сирі ЕЕГ дані. Вивчення репрезентації з сиріх фізіологічних даних (до них відноситься не лише ЕЕГ, але й магнітоелектроенцефалографія, електроміографія, etc.) є багатообіцяючою областью досліджень, оскільки у разі успіху процес інженерії характеристик (англ. feature engineering) стає непотрібним. Очевидно, що головним недоліком підбору характеристик є необхідність досконалого знання предметної області. На разі в задачі аналізу ЕЕГ домінує підхід з викори-

станням стандартного набору ознак і ускладненням класифікуючих алгоритмів, наприклад як в [13]. На жаль, подібні роботи мають невелику наукову цінність, хоча на даний момент дають кращі результати.

В даній роботі однією з головних цілей, не враховуючи розробки нового методу, є демонстрація важливості правильної оцінки алгоритмів прогнозування епілептичних нападів. Участь у “AES Seizure Prediction Challenge” надала нам прекрасну базу для порівняльного аналізу, тому даному змаганню буде приділена велика увага.



1 ТЕОРЕТИЧНІ ВІДОМОСТІ

Машинне навчання – це розділ штучного інтелекту, що займається розробкою моделей здатних самостійно навчатись з даних. Згідно одного із визначень машинного навчання, модель навчається з досвіду E відносно певного класу задач T та міри ефективності P , якщо якість моделі P для задач з класу T , покращується з отриманням досвіду E [14]. У нашому випадку T представляє конкретну задачу бінарної класифікації. Досвід E задається навчальною вибіркою $D = \{(\mathbf{x}_i, y_i)\}_{i=0..N-1}$, де \mathbf{x}_i – реалізація багатовимірної випадкової величини \mathbf{X} , що представляє собою вектори ознак для опису об'єктів, якими ми операємо; y_i – реалізація випадкової величини Y , що відповідає міткам класів, яким належать \mathbf{x}_i . Мірою ефективності P можуть слугувати різноманітні оцінки якості класифікації на тестовій вибірці – даних, які не використовувались під час навчання моделі. Саме здатність моделі узагальнювати вивчені результати на нових, тестових даних є ключовим поняттям, що відрізняє машинне навчання від методів оптимізації.

В наступних підрозділах ми детально описемо моделі класифікації, які лежать в основі запропонованого методу передбачення епілептичних нападів. Для роботи з ЕЕГ нам знадобляться деякі поняття з теорії цифрової обробки сигналів, а також методи візуалізації багатовимірних даних.

1.1 Лінійні класифікатори

1.1.1 Лінійний дискримінантний аналіз

Лінійний дискримінантний аналіз (ЛДА) базується на припущення, що \mathbf{X} – випадковий вектор, розподілений за багатовимірним нормальним законом в межах кожного з класів, з різними середніми значеннями, але

однаковими коваріаційними матрицями:

$$\mathbf{X}|Y=1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}) \quad (1.1)$$

$$\mathbf{X}|Y=0 \sim \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}). \quad (1.2)$$

Позначимо функції щільності розподілів а обох групах як $f_1(\mathbf{x})$ та $f_0(\mathbf{x})$, тоді для даного екземпляру \mathbf{x} , логарифм відношення ймовірностей приналежності до позитивного класу відносно негативного:

$$\begin{aligned} \log \frac{f_1(\mathbf{x})}{f_0(\mathbf{x})} &= \log \frac{\exp(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_1)^t \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_1))}{\exp(-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_0)^t \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}_0))} = \\ &= \mathbf{x}^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0) - \frac{1}{2}(\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)^t \boldsymbol{\Sigma}^{-1}(\boldsymbol{\mu}_1 + \boldsymbol{\mu}_0), \end{aligned}$$

що визначає лінійну границю класифікації. Об'єкт \mathbf{x} в даному випадку буде віднесений до позитивного класу, якщо $\log \frac{f_1(x)}{f_0(x)} > \log \frac{\pi_0}{\pi_1}$. Формула для ймовірності $P(Y=1|\mathbf{x})$ задана як:

$$P(Y=1|\mathbf{x}) = \frac{\pi_1 \cdot f_1(\mathbf{x})}{\pi_1 \cdot f_1(\mathbf{x}) + \pi_0 \cdot f_0(\mathbf{x})}, \quad (1.3)$$

де π_1 і π_0 априорні ймовірності позитивного та негативного класів ($\pi_0 + \pi_1 = 1$, тому можна позначити $\pi_1 = \pi$ і $\pi_0 = 1 - \pi$).

ЛДА має на меті знайти спільний розподіл \mathbf{X} та Y , який параметризується вектором $\boldsymbol{\theta} = \{\pi, \boldsymbol{\mu}_1, \boldsymbol{\mu}_0, \boldsymbol{\Sigma}\}$. Для оцінки даних параметрів використовується метод максимальної правдоподібності, тобто максимізації функції $\ell(\boldsymbol{\theta}; D) = \sum_{i=0}^{N-1} \log(P(y_i, \mathbf{x}_i | \boldsymbol{\theta}))$ відносно $\boldsymbol{\theta}$, де $D = \{(\mathbf{x}_i, y_i)\}_{i=0..N-1}$ – множина пар характеристичних векторів об'єктів \mathbf{x}_i та їх класів y_i з тренувальної вибірки. Нехай в D міститься N_1 позитивних і N_0 негативних екземплярів, тоді процес максимізації правдоподібності дає наступні оцінки:

$$\hat{\pi} = \frac{N_1}{N_1 + N_0}, \quad (1.4)$$

$$\hat{\boldsymbol{\mu}}_1 = \frac{1}{N_1} \sum_{y_i=1} \mathbf{x}_i, \quad \hat{\boldsymbol{\mu}}_0 = \frac{1}{N_0} \sum_{y_i=0} \mathbf{x}_i, \quad (1.5)$$

$$\hat{\Sigma} = \frac{1}{N-2} \left[\sum_{y_i=1} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_1)^t \cdot (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_1) + \sum_{y_i=0} (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_0)^t \cdot (\mathbf{x}_i - \hat{\boldsymbol{\mu}}_0) \right]. \quad (1.6)$$

Вираз $\mathbf{x}^t \boldsymbol{\delta} = \mathbf{x}^t \boldsymbol{\Sigma}^{-1} (\boldsymbol{\mu}_1 - \boldsymbol{\mu}_0)$ називається дискримінантою функцією. Аналізуючи значення $\boldsymbol{\delta}$, можна дізнатись чим відрізняються об'єкти з позитивного та негативного класів.

1.1.2 Логістична регресія

Логістична регресія (ЛР) – це статистичний регресійний метод, що моделює умовну ймовірність належності до позитивного класу як $P(Y = 1 | \mathbf{x}) = \frac{\exp(\alpha + \boldsymbol{\beta}' \mathbf{x})}{1 + \exp(\alpha + \boldsymbol{\beta}' \mathbf{x})}$. Як і в попередньому випадку параметри $\boldsymbol{\theta} = \{\alpha, \boldsymbol{\beta}\}$ оцінюються за допомогою методу максимальної правдоподібності, де функція правдоподібності має вигляд $\ell(\boldsymbol{\theta}; D) = \sum_{i=0}^{N-1} \log(P(y_i | \mathbf{x}_i, \boldsymbol{\theta}))$. Аналітичного розв'язку дана задача в загальному випадку не має, тому застосовуються методи чисельної оптимізації, наприклад метод градієнтного спуску – ітераційний метод пошуку локального мінімуму, що полягає в кроковому переміщенні в напрямі протилежному градієнту функції в поточній точці. Для використання методу градієнтного спуску необхідно задати початкову точку $\boldsymbol{\theta}_0 = \{\alpha_0, \boldsymbol{\beta}_0\}$ а також величину кроку η . На кожній ітерації для максимізації функції правдоподібності використовується наступне правило зміни параметрів $\boldsymbol{\theta}_n = \{\alpha_n, \boldsymbol{\beta}_n\}$:

$$\alpha_{n+1} \leftarrow \alpha_n + \eta \frac{\partial \ell}{\partial \alpha_n}, \quad \boldsymbol{\beta}_{n+1} \leftarrow \boldsymbol{\beta}_n + \eta \frac{\partial \ell}{\partial \boldsymbol{\beta}_n} \quad (1.7)$$

Дана процедура при певних значеннях η сходиться до глобального максимуму функції правдоподібності.

1.2 Штучні нейронні мережі

1.2.1 Багатошаровий перцептрон

Штучна нейронна мережа (ШНМ) була розроблена як математична модель біологічного мозку. Структура ШНМ складається з елементарних вузлів обробки інформації (нейронів), які пов'язані між собою зваженими з'єднаннями (синаптичними з'єднаннями). Виділяють два основні класи ШНМ: рекурентні нейронні мережі, зв'язки яких формують цикли, і мережі прямого поширення – ациклічні мережі. Найпростішою моделлю останнього типу є багатошаровий перцептрон (БШП).

Багатошаровий перцептрон можна розглядати як стек моделей мультикасової логістичної регресії, де нелінійність може бути не лише сігмоїдальною функцією. Метою багатошаровості є вивчення репрезентації даних на різних рівнях абстракції. Приклад БШП з N шарами наведений на Рис. 1.1.

Кожен i -й нейрон в n -ому прихованому шарі обраховує певну нелінійну функцію f (активаційну функцію) від зваженої лінійної комбінації входних імпульсів, як показано на Рис. 1.2. Виходом мережі на n -ому шарі є вектор: $\mathbf{x}^n = f(\mathbf{W}^n \cdot \mathbf{x}^{n-1} + \mathbf{b}^n)$, де \mathbf{x}^{n-1} – вихід шару $n - 1$, \mathbf{W}^n – матриця вагів \mathbf{b}^n – вектор зміщень (англ. biases). Вхід та вихід нейронної мережі позначені як \mathbf{x}^0 та \mathbf{x}^N . В задачах бінарної класифікації, де ми зацікавлені у ймовірності класів, виходом БШП слугує логістична регресія:

$$x^N = P(Y = 1 | \mathbf{x}^0; \boldsymbol{\theta}) = \frac{\exp(\mathbf{b}^N + \mathbf{W}^N \mathbf{x}^{N-1})}{1 + \exp(\mathbf{b}^N + \mathbf{W}^N \mathbf{x}^{N-1})} \quad (1.8)$$

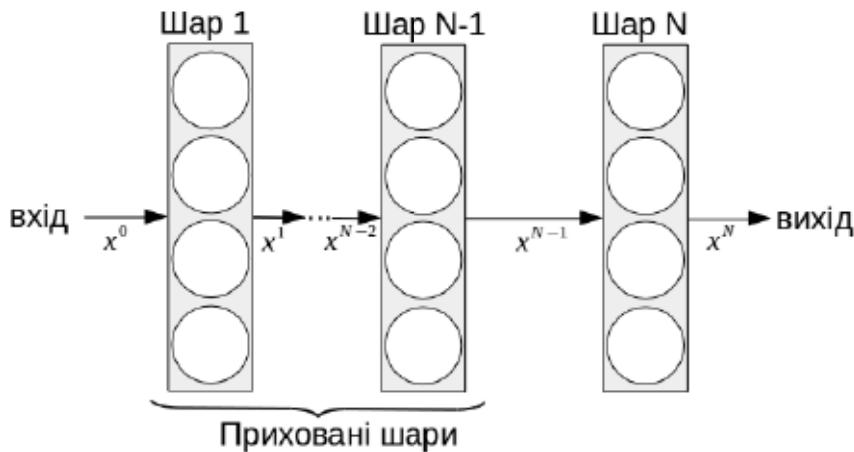


Рисунок 1.1 – Багатошаровий перцептрон

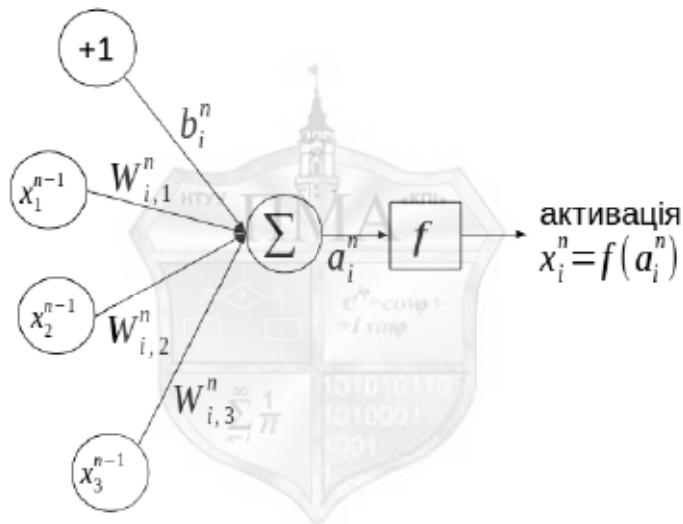


Рисунок 1.2 – Штучний нейрон

1.2.2 Тренування ШНМ

Тренування нейронної мережі полягає у вивченні параметризації $\boldsymbol{\theta}$, що мінімізує функцію помилки – в нашому випадку логарифму від функції правдоподібності, взятої з від’ємним знаком:

$$\ell(\boldsymbol{\theta}; D) = - \sum_{i=0}^{|D|} \log(P(Y = y^{(i)} | x^{(i)}, \boldsymbol{\theta})). \quad (1.9)$$

Як і у випадку логістичної регресії, для мінімізації $\ell(\boldsymbol{\theta}; D)$ можна застосувати метод градієнтного спуску

$$\mathbf{W}^n \leftarrow \mathbf{W}^n - \eta \frac{\partial \ell}{\partial \mathbf{W}^n}, \quad b \leftarrow b^n - \eta \frac{\partial \ell}{\partial b^n}. \quad (1.10)$$

Відмінність від випадку логістичної регресії полягає у тому що $\ell(\boldsymbol{\theta}; D)$ має локальні мініуми та сідлові точки, тому бажаним є знайти точку з достатньо малим значенням цільової функції.

Зазвичай, для оцінки напрямку градієнта використовується підмножина тренувальної вибірки – випадково вибрана міні-група (англ. mini-batch), розмір якої підбирається емпіричним шляхом. Даний розмір регулює наскільки варіативною є оцінка градієнтів, а також впливає на час обчислень [15].

Важливим параметром в процесі навчання є величина кроку η : занадто великі значення η призводять до розбіжності оптимізаційного процесу, занадто малі – до повільного навчання. Для того, щоб полегшити вибір η існує ряд методів, що адаптують η в процесі навчання. Одним із таких методів є ADADELTA [16]. На кожній ітерації t ADADELTA змінює деякий параметер ω за наступним правилом:

$$\Delta \omega_t \leftarrow -\frac{RMS[\Delta \omega]_{t-1}}{RMS[g]_t} \cdot g_t, \quad (1.11)$$

$$\omega_t \leftarrow \omega_t + \Delta \omega_t \quad (1.12)$$

В даній формулі $RMS[g]_t = \sqrt{E[g^2]_t + \varepsilon}$, де $E[g^2]_t = \rho E[g^2]_{t-1} + (1 - \rho)g_t^2$ – середнє значення квадратів градієнтів g^2 , де ρ – коефіцієнт згасання, з яким враховуються попередні градієнти. Аналогічно, $RMS[\Delta \omega]_{t-1} = \sqrt{E[\Delta \omega^2]_{t-1} + \varepsilon}$. У наведених формулах ε – це допоміжна константа порядку 10^{-6} , що забезпечує відмінність знаменника від нуля а також забезпечує нетривіальну зміну параметрів на першій ітерації, де значення $E[\Delta \omega^2]_0$ є нульовим.

ADADELTA присвоює кожному параметру свій динамічний крок. Чисельник даного виразу, $RMS[\Delta\omega]_{t-1}$, прискорює навчання в тих напрямках, де знаки градієнтів зберігаються, і сповільнює навчання в тих напрямках, де градієнти змінюють знак. Така поведінка необхідна для оптимізації функцій з довгими та вузькими долинами в просторі параметрів. Знаменник $RMS[g]_t$, в свою чергу, прагне збільшити крок для малих значень градієнтів і зменшити для великих значень, тим самим збалансуючи прогрес в різних напрямках.

1.2.3 Алгоритм зворотнього поширення помилки у випадку БШП

Градієнти в рівнянні 1.10 обраховуються за допомогою алгоритму зворотнього поширення помилки, основою якого є застосування правила диференціювання складної функції.

Введемо наступну нотацію: $\mathbf{x}^n = f(\mathbf{W}^n \cdot \mathbf{x}^{n-1} + \mathbf{b}^n)$ – вихід на n -ому шарі мережі ($n = 1 \dots N - 1$), де активація одного нейрона i дорівнює $\mathbf{x}_i^n = f(\sum_j \mathbf{W}_{i,j}^n \cdot \mathbf{x}_j^{n-1} + b_i^n)$. Пре-активацію i го шару позначимо як $\mathbf{a}^n = \mathbf{W}^n \cdot \mathbf{x}^{n-1} + \mathbf{b}^n$. Виходом мережі є шар логістичної регресії, у багатовимірному випадку відомої також як софтмакс: $\mathbf{x}^N = softmax(\mathbf{W}^N \cdot \mathbf{x}^{N-1} + \mathbf{b}^N)$. Цільову функцію, обраховану для одного навчального прикладу, який належить до класу y позначимо як $C = -\log p_y$, де p_y відповідає ймовірності класу y , яку потрібно максимізувати.

За стосовуючи рекурсивне правило диференціювання складної функції, можна порахувати градієнти відносно вагів нейронної мережі на всіх шарах від N до 1:

1. зміщення:

$$\frac{\partial C}{\partial \mathbf{b}_i^n} = \frac{\partial C}{\partial \mathbf{a}_i^n} \cdot \frac{\partial \mathbf{a}_i^n}{\partial \mathbf{b}_i^n} = \frac{\partial C}{\partial \mathbf{a}_i^n} \quad (1.13)$$

2. ваги:

$$\frac{\partial C}{\partial \mathbf{W}_{i,j}^n} = \frac{\partial C}{\partial \mathbf{a}_i^n} \cdot \frac{\partial \mathbf{a}_i^n}{\partial \mathbf{W}_{i,j}^n} = \frac{\partial C}{\partial \mathbf{a}_i^n} \cdot \mathbf{x}_j^{n-1} \quad (1.14)$$

3. поширити помилку на попередній $n - 1$ шар:

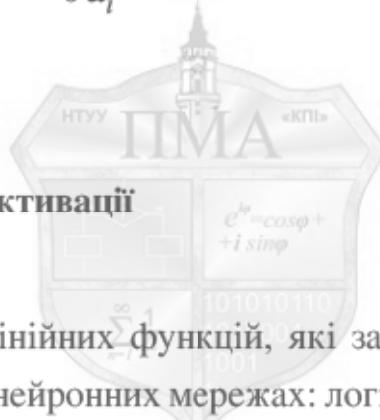
$$\frac{\partial C}{\partial \mathbf{x}_j^{n-1}} = \sum_i \frac{\partial C}{\partial \mathbf{a}_i^n} \cdot \frac{\partial \mathbf{a}_i^n}{\partial \mathbf{x}_j^{n-1}} = \sum_i \frac{\partial C}{\partial \mathbf{a}_i^n} \cdot \mathbf{W}_{i,j}^n \quad (1.15)$$

$$\frac{\partial C}{\partial \mathbf{a}_j^{n-1}} = \frac{\partial C}{\partial \mathbf{x}_j^{n-1}} \cdot \frac{\partial \mathbf{x}_j^{n-1}}{\partial \mathbf{a}_j^{n-1}} = \frac{\partial C}{\partial \mathbf{x}_j^{n-1}} \cdot \frac{d}{d \mathbf{a}_j^{n-1}} f(\mathbf{a}_j^{n-1}) \quad (1.16)$$

і повторити з кроку 1.

Однак, для початку обчислень нам необхідна похідна $\frac{\partial C}{\partial \mathbf{a}_i^N}$ – значення градієнту відносно пре-софтмакс активації. Нескладно довести, що

$$\frac{\partial C}{\partial \mathbf{a}_i^N} = \mathbf{p}_i - I\{y = i\} \quad (1.17)$$



1.2.4 Функції активації

Існує кілька нелінійних функцій, які зазвичай використовуються як функції активації в нейронних мережах: логістична функція $(1 + \exp(-x))^{-1}$, гіперболічний тангенс та лінійна ректифікація – ReLU, що має вираз $\max(0, x)$ (Рис. 1.3) [17, 18].



Рисунок 1.3 – Лінійна ректифікація

Щоб забезпечити вільний потік інформації в нейронній мережі, ва-

ти мають бути правильно ініціалізовані залежно від функції активації. Для $tanh$ рекомендовано вибрати ваги для i го шару з рівномірного розподілу на проміжку $[-\sqrt{\frac{6}{fan_{in}+fan_{out}}}; \sqrt{\frac{6}{fan_{in}+fan_{out}}}]$, де fan_{in} – кількість нейронів на $(i - 1)$ шарі, а fan_{out} – кількість нейронів в шарі i [19]. Для лінійної ректифікації загальновживано практикою є ініціалізація вагів з нормального розподілу $\mathcal{N}(0, 0.01)$. Початкові зміщення для $tanh$ можуть бути нульовими; для ректифікації доцільним є ініціалізація позитивними значеннями, наприклад 0.1, оскільки це змушує нейрони працювати в лінійному режимі з ненульовими градієнтами.

З рівняння 1.16 можна пересвідчитись, що похідна функції помилок відносно пре-активацій містить похідну активаційної функції. На Рис. 1.4 зображеній графік похідної гіперболічного тангенса $\frac{d}{dx}\tanh(x)$, що ілюструє можливу сатурацію нейронів при великих, за абсолютною значенням, пре-активаціях (аналогічно, для логістичної функції). Як результат, помилка не розповсюджується в глибші шари мережі. Дано проблема відома під назвою зникаючого градієнта (англ. vanishing gradient) [15]. Ректифікаційна функція активації не призводить до подібного ефекту, тому вона найчастіше використовується в глибоких нейронних мережах.

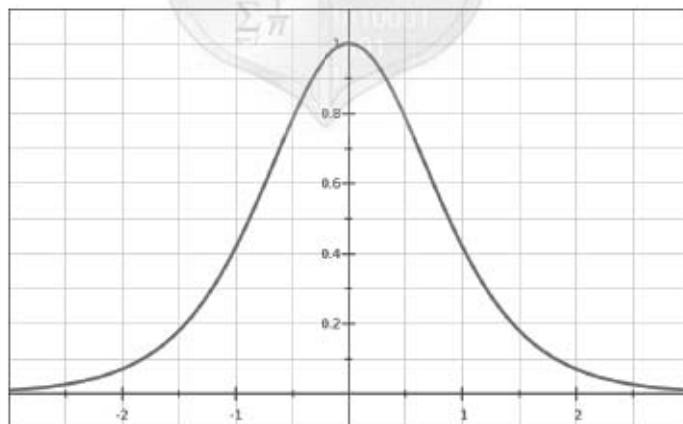


Рисунок 1.4 – Графік похідної гіперболічного тангенса

1.2.5 Конволютивні нейронні мережі

Конволютивні нейронні мережі (КНН)[20] – тип штучних нейронних мереж прямого поширення, які моделюють роботу області кори головного мозку, що відповідає за обробку візуальної інформації. Особливістю даної ділянки мозку є присутність локально чутливих нейронів, які сприймають лише окремі частини зображення, але зв'язані таким чином, що покривають все візуальне поле. У цьому полягає головна відмінність від БШП, де кожен нейрон зв'язаний з усіма нейронами попереднього шару.

Класична архітектура конволютивної нейронної мережі містить три види шарів: конволютивні, субдискретизуючі та шари з повним з'єднанням. В наступних параграфах буде детально описаний кожен з них. Також буде пояснено як змінюється алгоритм зворотнього поширення помилки у випадку включення в мережу конволютивних шарів.

Конволютивні шари здійснюють операцію згортки відносно попереднього шару нейронів, застосовуючи матрицю вагів – ядро конволюції, розмірність якої є меншою за розмірність вхідного шару. В подальшому розглянатиметься випадок, коли конволюція здійснюється лише в одному вимірі, наприклад у часі.

Вихід кожного нейрону після застосування конволюції можна описати наступною формулою:

$$\mathbf{h}_i = f((\mathbf{W} * \mathbf{x})_i + b), \quad (1.18)$$

де $*$ позначає операцію згортки між набором вагів \mathbf{W} та вхідним вектором \mathbf{x} . Рис. 1.5 ілюструє наведену формулу у випадку, коли \mathbf{W} містить 3 параметри (на рисунку одинакові ваги позначені одним типом ліній), в даному випадку ширина візуального поля рівна трьом. Вектор \mathbf{h} , що складається у даному випадку з 4-х нейронів, називається “картою ознак” (англ. feature map). Оскільки використовується одна матриця вагів, \mathbf{h} може кодувати лише один певний признак, наприклад наявність певних ліній в зображенні.

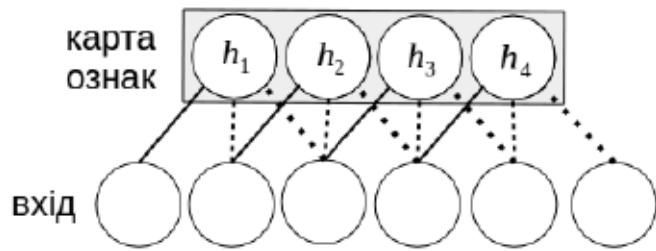


Рисунок 1.5 Одновимірна конволюція з однією картою ознак

Щоб зменшити кількість нейронів у вихідних шарах, на практиці часто використовується конволюція з певним кроком, коли відстань між центрами візуальних полів більша за 1. Приклад конволюції з кроком 2 наведений на рис. 1.6.



Рисунок 1.6 – Конволюція з кроком 2

Для того, щоб сформувати ефективну репрезентацію даних, очевидно, що потрібно закодувати більш ніж одну характеристику. Для цього на кожному шарі нейронної мережі використовують різні карти призnaків.

Рис. 1.7 ілюструє як два коволютивні шари з кількома картами признаків можуть бути поєднані. В шарі $n - 1$ маємо три карти ознак $\mathbf{h}^{n-1,i}$, ($i = 1 \dots 3$), і дві карти $\mathbf{h}^{n,1}, \mathbf{h}^{n,2}$ у шарі n . Відповідні матриці вагів та зміщення позначені як $\mathbf{W}^{n,1}, \mathbf{W}^{n,2}, b^{n,1}, b^{n,2}$ відповідно. Кожен k -й нейрон ($k = 1 \dots 3$) з $\mathbf{h}^{n,1}$ та $\mathbf{h}^{n,2}$ має візуальне поле ширину 2, їх функції активацій можуть бути записані як:

$$\mathbf{h}_k^{n,1} = f\left(\sum_{i=1}^3 \sum_{j=k}^{k+1} \mathbf{W}_{i,j}^{n,1} h_j^{n-1,i} + b^{n,1}\right) \quad (1.19)$$

$$\mathbf{h}_k^{n,2} = f\left(\sum_{i=1}^3 \sum_{j=k}^{k+1} \mathbf{W}_{i,j}^{n,2} h_j^{n-1,i} + b^{n,2}\right) \quad (1.20)$$

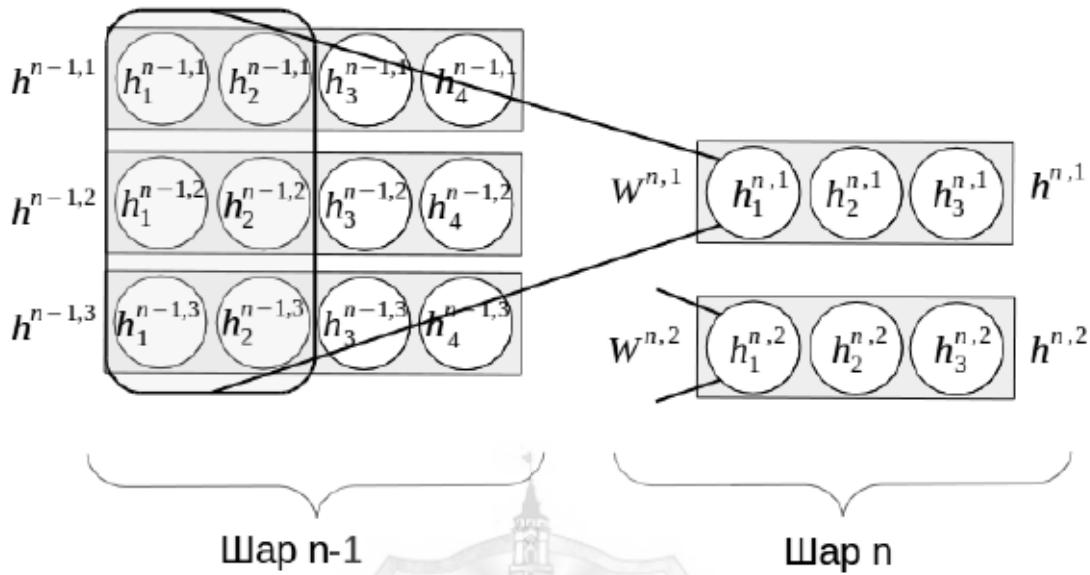


Рисунок 1.7 – З’єднання між конволютивними шарами

Субдискретизуючі шари (англ. subsampling layers) зазвичай включають між конволютивними шарами для зменшення резолюції карт ознак. Як правило, субдискретизуючі шари рахують максимум або середнє значення груп нейронів з попереднього шару. Операція субдискретизації забезпечує додаткову інваріантність відносно трансляцій а також зменшують кількість обчислень для послідовних шарів. Схематичне зображення даної операції наведене на рис. 1.8.

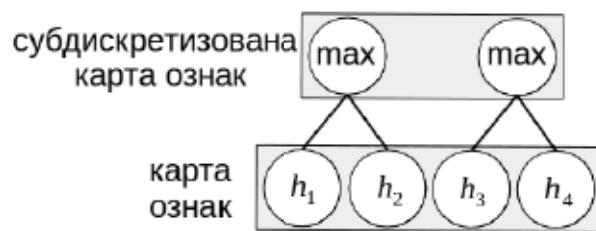


Рисунок 1.8 – Операція субдискретизації

Шари з повним з’єднанням нейронів (щільні шари) використовуються у верхні шарах КНН і є повною аналогією БПШ. Таким чином, дані

шари містять найбільшу кількість параметрів нейромережі. Оскільки в БШП змішується просторова інформація, конволютивні шари не можуть бути включені після щільних шарів.

Тренування КНН потребує внесення певних змін до алгоритму зворотнього поширення помилки в БШП, оскільки в конволютивних шарах нейронної мережі нейрони в межах однієї карти признаків мають однакові ваги, а також візульні поля, що перетинаються. Повернувшись до рис.1.7, можна побачити, що $\mathbf{W}_{i,j}^{n,1}$ і зміщення $b^{n,1}$ присутні в активаціях усіх нейронів з $\mathbf{h}^{n,1}$. Позначивши $\mathbf{a}_k^{n,1}$ як пре-активацію k го нейрона в $\mathbf{h}^{n,1}$ отримуємо наступні формули градієнтів:

$$\frac{\partial C}{\partial b^{n,1}} = \sum_{k=1}^3 \frac{\partial C}{\partial \mathbf{a}_k^{n,1}} \cdot \frac{\partial \mathbf{a}_k^{n,1}}{\partial b^{n,1}} = \sum_{k=1}^3 \frac{\partial C}{\partial \mathbf{a}_k^{n,1}} \quad (1.21)$$

$$\frac{\partial C}{\partial \mathbf{W}_{i,j}^{n,1}} = \sum_{k=1}^3 \frac{\partial C}{\partial \mathbf{a}_k^{n,1}} \cdot \frac{\partial \mathbf{a}_k^{n,1}}{\partial \mathbf{W}_{i,j}^{n,1}} = \sum_{k=1}^3 \frac{\partial C}{\partial \mathbf{a}_k^{n,1}} \cdot \mathbf{h}_j^{n-1,i} \quad (1.22)$$

Для поширення помилки в нижчі шари мережі, нам потрібні похідні $\frac{\partial C}{\partial \mathbf{h}_k^{n-1,i}}$, ($k = 1 \dots 4, i = 1 \dots 3$). Можна помітити, що наприклад, вираз $\mathbf{h}_2^{n-1,i}$ присутній в активаціях чотирьох нейронів з i го шару: $\mathbf{h}_1^{n,1}, \mathbf{h}_2^{n,1}, \mathbf{h}_1^{n,2}, \mathbf{h}_2^{n,3}$. Приймаючи до уваги даний факт, можна записати градієнти як:

$$\frac{\partial C}{\partial \mathbf{h}_k^{n-1,i}} = \sum_{j=1}^2 \sum_{i=k}^{k-1} \frac{\partial C}{\partial \mathbf{a}_i^{n,j}} \cdot \frac{\partial \mathbf{a}_i^{n,j}}{\partial \mathbf{h}_k^{n-1,i}} = \sum_{j=1}^2 \sum_{i=k}^{k-1} \frac{\partial C}{\partial \mathbf{a}_i^{n,j}} \cdot \mathbf{W}_{i,k}^{n,j} \quad (1.23)$$

Вирази інших похідних, які необхідні для поширення помилки в КНН залишаються аналогічними тим, що описані в пункті 1.2.2.

1.2.6 Перенавчання

Незважаючи на чисельні успіхи, нейронні мережі потерпають від кількох недоліків. Наявність багатьох нелінійних шарів робить глибо-

кі мережі дуже потужними обчислювальними моделями, здатними виражати складні залежності між X та Y . Як наслідок, ШНМ склонні до перенавчання (англ. overfitting) – вивчення шуму, присутнього у тренувальній вибірці замість істинної закономірності між X та Y у популяції. Перенавчена модель не здатна узагальнювати свої знання по відношенню до нових даних. Очевидно, що модель повинна бути достатньо потужна, щоб вивчити закономірності в даних і в той же час бути нечутливою до незначних варіацій присутніх у тренувальній множині. Ми прагнемо знайти компроміс, що досягається методами регуляризації. Найпростіший метод регуляризації – введення штрафу за великі значення параметрів, що перешкоджає моделі апроксимувати будь-яку функцію. Формально до виразу помилки додається L_1 або L_2 норма параметрів нейронережів з певним коефіцієнтом λ , що контролює силу регуляризації. Існують також інші методи, що підходять для будь-яких моделей класифікації і мають на меті збільшити здатність моделі до генералізації:

1. збільшення тренувальної вибірки: створити нові екземпляри тренування шляхом трансформації існуючих при збереженні класових міток.
2. комбінування моделей: усереднення передбачень декількох моделей, що найбільш ефективне у випадку, коли моделі відрізняються у своїх передбаченнях. Для того щоб зробити моделі різними, вони повинні мати різні гіперпараметри та/або навчатись на різних даних. Якщо усереднити моделі з приблизно однаковими показниками якості, результатує передбачення гарантовано будуть кращими. Для проблем класифікації найчастіше використовуюється середнє арифметичне або середнє геометричне моделей [21].

Для нейронних мереж існує кілька особливих технік:

1. метод ранньої зупинки [15] відслідковує розмір помилки на валідаційній множині, що не використовується під час тренування ШНМ. Як тільки дана помилка починає зростати, потрібно припинити навчання мережі. В результаті надмірна потужність моделі компенсується малою тривалістю тренування моделі.

2. алгоритм виключення (англ. dropout) [22] найбільш простий та ефективний метод для зменшення проявів перенавчання у ШНМ. Під виключенням розуміється усунення з мережі певних нейронів та інцидентних їм зв'язків. Кожен нейрон незалежно від інших виключається з певною ймовірністю p , де типовим приймається значення $p = 0.5$. Процедура виключення відбувається під час навчання нейронної мережі і щоразу, коли мережа отримує на вхід один екземпляр з тренувальної вибірки, в шарі з N нейронами залишається в середньому $N(1 - p)$. Формально вихід n го шару, під час виключення нейронів з шару $n - 1$ можна записати як:

$$\mathbf{x}_n = f(\mathbf{W}_n \cdot (\mathbf{r}_n \circ \mathbf{x}_{n-1}) + \mathbf{b}_n), \quad (1.24)$$

де \mathbf{r}_n – вектор незалежних випадкових змінних розподілених за законом Бернуллі, де ймовірність нуля дорівнює p ; \circ позначає поелементний добуток. Таким чином, відбувається одночасне тренування максимум 2^N розріджених архітектур зі спільними параметрами. На етапі тестування використовується одна мережа з масштабованими параметрами: $\mathbf{W}_n^{test} = (1 - p) \cdot \mathbf{W}_n$.

1.3 Дискретне перетворення Фур'є

Нехай вектор $a = (a_0, \dots, a_{n-1})$ містить виміри сигналу через рівні проміжки часу протягом періоду T . Дискретне перетворення Фур'є (ДПФ) представляє даний сигнал в частотній області. Коефіцієнти ДПФ, що відповідають комплексним синусоїдам з частотою $f_k = \frac{k}{T}$, задаються наступною формулою:

$$A_k = \sum_{m=0}^{n-1} a_m \exp\left(-2\pi i \frac{mk}{n}\right), k = 0, \dots, n-1 \quad (1.25)$$

В подальшому наступні поняття є важливими:

1. множина $|A_k|$ формує амплітудний спектр, що складається з $n/2 + 1$ компонент для дійсних сигналів.
2. множина $|A_k|^2$ визначає спектр потужності
3. спектральна енергія: $E_k = |A_k|^2$

1.4 t-SNE

t-SNE [23] – нелінійний метод зменшення розмірності даних, який підходить для відображення багатовимірних даних в простір 2ox або 3ox компонент, де дані можуть бути легко візуалізовані. t-SNE моделює багатовимірні об'єкти таким чином, що схожі об'єкти моделюються 2ox вимірними точками, що знаходяться близько і навпаки. Для цього задається метрика схожості, що у випадку t-SNE є ймовірністю, що багатовимірні об'єкти \mathbf{x}_i та \mathbf{x}_j мають ймовірність p_{ij} бути одночасно вибраними з усієї множини $\{\mathbf{x}_1 \dots \mathbf{x}_N\}$:

$$p_{j|i} = \frac{\exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\sigma_i^2)}{\sum_{k \neq i} \exp(-\|\mathbf{x}_i - \mathbf{x}_k\|^2 / 2\sigma_i^2)} \quad (1.26)$$

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2N} \quad (1.27)$$

t-SNE прагне знайти відображення між багатовимірними об'єктами \mathbf{x} та, припустимо, двовимірними об'єктами \mathbf{y} , так щоб найкраще зберегти схожість у 2D просторі. Нехай між точками \mathbf{y}_i та \mathbf{y}_j схожість задається як:

$$q_{j|i} = \frac{(1 + \|\mathbf{y}_i - \mathbf{y}_j\|^2)^{-1}}{\sum_{k \neq l} (1 + \|\mathbf{y}_k - \mathbf{y}_l\|^2)^{-1}} \quad (1.28)$$

Для встановлення бажаного відображення між багатовимірними та 2D об'єктами t-SNE мінімізує дивергенцію Кульбака-Лейблера між розподілами P та Q відносно координат \mathbf{y}_i :

$$KL(P||Q) = \sum_{i \neq j} p_{ij} \log \frac{p_{ij}}{q_{ij}} \quad (1.29)$$

1.5 Висновки

У розділі розглянуто деякі загальні теоретичні відомості з теорії машинного навчання та цифрової обробки сигналів. Детально були пояснені лінійні методи класифікації, такі як логістична регресія та ЛДА. Велика увага була приділена основним поняттям в області штучних нейронних мереж. Були описані дві часто вживані архітектури, а саме БШП та КНН, наведені алгоритми їх навчання, а також підходи для зменшення ефекту перенавчання, що притаманний для глибоких нейронних мереж. Насамкінець, був описаний один з методів візуалізації багатовимірних даних, що допоможе в подальшому зрозуміти труднощі, які виникають у роботі з ЕЕГ даних.



2 AES SEIZURE PREDICTION CHALLENGE

2.1 Дані

Інtrakраніальні ЕЕГ дані для змагання були записані у 7ох пацієн-тів: 5ти собак та 2ох людей, які проходять ЕЕГ моніторинг (надалі ми будемо позначати їх оригінальними мітками, наприклад Dog_1 або Patient_1). Для собак iEEG сигнал був сепльований з частотою дискретизації 400 Гц з 16 електродів. Записи iEEG збирались протягом року і для деяких собак містять до сотні епілептичних нападів. Для людей кількість електродів варіювалась залежно від стану хворого; сигнал був сепльованний з частотою 5000 Гц; період моніторингу складав близько тижня.

На рис. 2.1 зображений принцип роботи системи передбачення нападів. Основою системи є імплантований пристрій (А), що знімає iEEG дані з 16ти електродів (В) і передає дані на персональний пристрій, який завантажує дані на сайт центрального сховища. (С) ілюструє часовий профіль ймовірності наступаючого нападу; щоразу, як дана ймовірність перевищує певний поріг, система вмикає сигнал попередження [4].

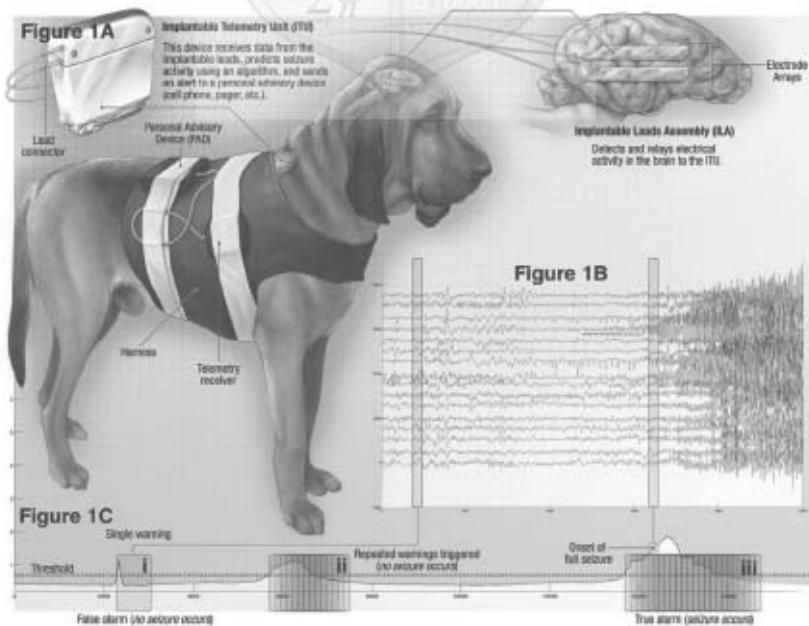


Рисунок 2.1 – Принцип роботи системи передбачення нападів [4]

Відомо, що напади виникають групами, тому доцільно передбачати

лише перший з них. Тому дані для змагання включали лише напади, які відрізняються у часі на більш ніж 4 години. Інтеріктальні дані, що не містять слідів епілептичної активності, були вибрані випадково з усього запису ЕЕГ за умови, що вони відділені від нападу мінімум один тиждень для собак та 4 години для людей [5].

Кожен екземпляр даних представляє сегмент ЕЕГ тривалістю в 10 хвилин. Тренувальні дані згруповани в одногодинні послідовності, в той час як для тестових даних їх час появи невідомий. Преіктальний період починається за одну годину до нападу з 5-ти хвилинним горизонтом: від 1:05 до 0:05 перед настанням нападу як показано на рис. 2.2. Це дозволяє завчасно передбачити напад і прийняти швидкодіючі медикаменти.

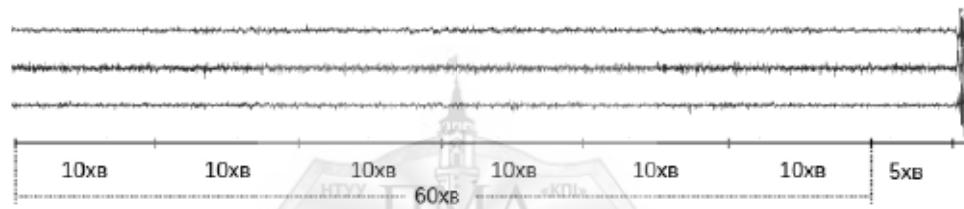


Рисунок 2.2 – Приклад ЕЕГ сигналу (лише 3 канали) з преіктального періоду

Кількість 10-хвилинних сегментів ЕЕГ з преіктального та інтеріктального періодів наведені в таблиці 2.1 (впродовж змагання тестові мітки були невідомими). Тестова множина була розбита невідомим чином на публічну та приватну, що можна в деякому сенсі розглядати як валідаційну та тестову множину в термінах теорії машинного навчання. Джерелом валідації є результат, який платформа Kaggle відображає, коли учасники змагання завантажують файли з передбаченнями (деталі в наступному параграфі).

Таблиця 2.1 – Кількість інтеріктальних/преіктальних сегментів в тренувальній (1) та тестовій (2) множині

	Dog_1	Dog_2	Dog_3	Dog_4	Dog_5	Patient_1	Patient_2
1	480/24	500/42	1440/72	804/97	450/40	50/18	42/18
2	478/24	910/90	865/42	933/57	179/12	183/12	136/14

2.2 Оцінка якості класифікації

В даному змаганні для кожного сегменту ЕЕГ вимагалось передбачити ймовірність того, що даний сегмент належить преіктальному періоду, тобто $P(Y = 1|x)$. Для оцінки якості класифікації використовувалась площа під ROC кривою (AUC), обрахованою водночас для всіх пацієнтів (далі ми будемо називати її глобальною AUC або просто AUC, якщо не зазначено інше). Для подальших пояснень введемо наступні терміни, що стосуються бінарної класифікації, коли замість ймовірності розподілу між класами, ми зацікавлені лише в бінарних мітках 0 (негативний клас) або 1 (позитивний клас):

1. TPR (англ. True Positive Rate, чутливість) – пропорція числа правильно класифікованих позитивних екземплярів до всіх можливих позитивних екземплярів у вибірці (класифікованих як 1 так і 0).
2. TNR (англ. True Negative Rate, специфічність) – пропорція числа правильно класифікованих негативних екземплярів до всіх можливих негативних екземплярів у вибірці.

Якщо виходом класифікатора є ймовірності, їх можна перевести в бінарні мітки за допомогою класифікуючого правила: присвоїти екземпляру позитивну мітку, якщо $P(Y = 1|x) > \delta$ і негативну мітку у протилежному випадку. Наприклад, на виході класифікатора буди отримані ймовірності: $P(Y = 1|x_1) = 0.4$ і $P(Y = 1|x_2) = 0.6$. При $\delta = 0.2$, обидва екземпляри будуть класифіковані як позитивні; якщо ж $\delta = 0.5$, x_1 буде віднесений до негативного класу, а x_2 – до позитивного.

ROC крива [24] відображає TPR відносно $1 - TNR$ при різному порозі класифікації δ . Приклад ROC кривої наведений на Рис. 2.3. Площа під кривою, AUC, дає кількісну інтерпритацію ROC кривої. Для рандомного класифікатора дорівнює $AUC=0.5$, в той час як максимум досягається при 1.0.

Повернемось до критерія, який був використаний у змаганні: площа під ROC кривою, обрахована одночасно для 7-ми пацієнтів. Далі ми

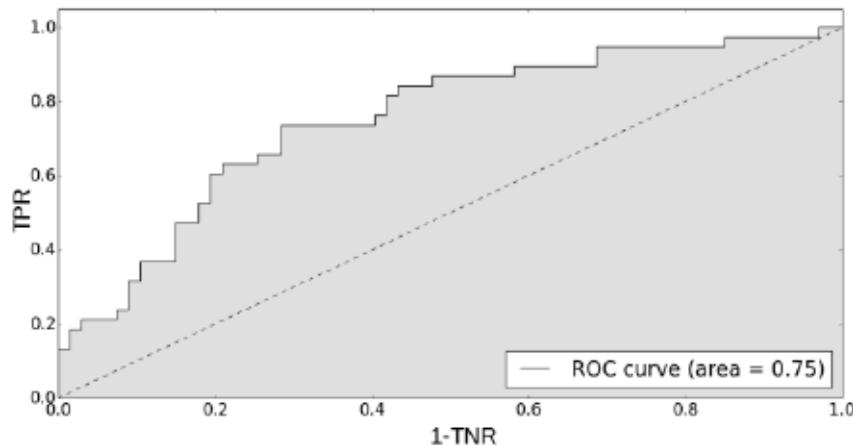


Рисунок 2.3 – Приклад ROC кривої

обговоримо недоліки та переваги такого способу оцінки.

Оскільки мозкова активність є суто індивідуальною, для кожного пацієнта доцільно тренувати окремі моделі. В даному випадку передбачення моделей можуть бути невідкалібровані, що призводить до низьких значень AUC, навіть якщо окремі класифікатори близькі до ідеальних. Пояснимо на прикладі: нехай для першого пацієнта існують наступні пари передбачень $P(Y = 1|x)$ та справжніх міток: (0.1, 0), (0.2, 0), (0.3, 1), (0.4, 1); результуюча AUC в даному випадку рівна 1. Для іншого пацієнта маємо пари (0.6, 0), (0.7, 0), (0.8, 1), (0.9, 1), що теж призводить до максимального значення AUC. Однак, якщо скомбінувати їх воєдино, отримаємо AUC 0.75. Таким чином, використовуючи глобальна AUC вимагає від моделей бути менш залежними від порогу класифікації δ . Дано властивість є дуже цінною на практиці, оскільки у зв'язку з нестационарністю ЕЕГ підібрати постійне оптимальне значення δ не є можливим.

Варто зауважити, що за місяць до кінця змагання організатори дозволили використовувати тестову множину для калібрації передбачень між пацієнтами, чим зробили результати змагання менш прозорими. Частковою мірою це призвело до фокусування на розробці методів калібрації, аніж моделей передбачення. В реальності моделі тренуються та тестиються для кожного пацієнта індивідуально, тому калібрація не несе жодного практичного сенсу.

2.3 Висновки

У даному розділі був детально описаний набір даних, який був використаний у змаганні “AES Seizure Prediction Challenge”. Також був пояснений метод, що використовувався для оцінки якості класифікуючих алгоритмів, були проаналізували його переваги та недоліки. Головним негативним фактором є використання тестової вибірки для калібрації передбачень, що не дозволяє належно оцінити запропоновані алгоритми відносно їх практичної цінності.



3 МЕТОД НА ОСНОВІ КОНВОЛЮТИВНИЙ НЕЙРОННИХ МЕРЕЖ

3.1 Обробка даних

Першим чином частота дискретизації ЕЕГ сигналу для усіх пацієнтів була змінена на 400 Гц, після чого застосувався смуговий фільтр в діапазоні 0.1-180 Гц. Позначимо кількість ЕЕГ каналів як N , тоді кожен 10-хвилинний сегмент матиме розмірність $N \times (600 \text{ sec} \cdot 400 \frac{1}{\text{sec}}) = N \times 240000$.

Попередні роботи довели, що частотні характеристики сигналу є дискримінітними під час виявлення преіктального стану [4], тому був використаний схожий підхід. Кожен 10-хвилинний сегмент був розбитий на 1-хвилинні кадри, що не перетинаються. У кожному кадрі, здійснивши ДПФ, ми порахували десятковий логарифм результуючого амплітудного спектру. На рис. 3.1 показаний приклад (1й канал ЕЕГ інреіктально-го сегменту даних Dog_1) спектrogrammi отриманої за допомогою даної процедури.

Далі спектр був поділений на 6 частотних смуг: дельта (0.1-4 Гц), тета (4-8 Гц), альфа (8-12 Гц), бета (12-30 Гц), низькі гама (30-70 Гц) та високі гама частоти (70-180 Гц). В кожній смузі ми порахували середнє значення відповідних $\log_{10} |A_k|$, як показано на прикладі з рис. 3.2. Таким чином розмірність сегменту даних зменшилась до $N \times 6 \times 10$ (к-ть каналів \times к-ть частотних суг \times к-ть кадрів). Варто зауважити, що оригінальні дані семи пацієнтів займали 106 Gb, після обробки – менш ніж 200 Mb, цей фактор може бути перевагою для пристройів з обмеженою памяттю.

Для багатьох методів класифікації, в тому числі нейронних мереж, ефективна оптимізація цільової функції потребує нормалізації вхідних ознак. Були випробувані дві наступні схеми нормалізації (кожен канал ЕЕГ розглядається окремо):

- представити кожну спектrogrammu у вигляді 60-вимірного вектора; дляожної i -ї компоненти порахувати математичне очікування μ_i та середнє квадратичне відхилення σ_i у тренувальній вибірці; нор-

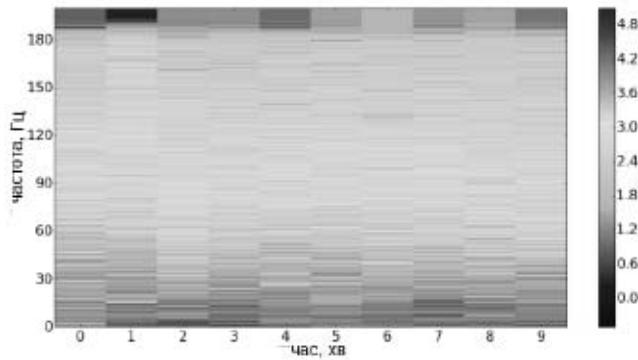


Рисунок 3.1 – Логарифм спектрограми амплітуд ЕЕГ сигналу

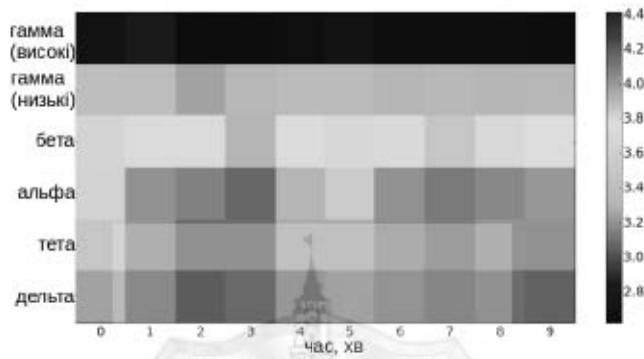


Рисунок 3.2 – Сегментована спектрограма

малізувати кожну компоненту вектора за формулою $\frac{x_i - \mu_i}{\sigma_i}$. Нормалізовуючи дані таким чином, враховується позиція ознаки в часі та частотній смузі.

- декомпозувати 6×10 спектрограму на 10 векторів, що містять по 6 частотних характеристик, нормалізувати 6-вимірні вектори для кожного виміру окремо і скомбінувати вектори назад в матрицю 6×10 . На відміну від попередньої схеми, положення характеристик у часі не враховується.

Як додаткову характеристику, ми використали середнє квадратичне відхилення сигналу, обраховане в 1-хвилинних проміжках $N \times 7 \times 10$. Також варіювалось розбиття частотних смуг, наприклад, дві широкі гама смуги можна поділити на 4 частини: 30-50 Гц, 50-70 Гц, 70-100 Гц and 100-180 Гц.

Кількість тренувальних даних була збільшена за рахунок перетину сегментів у межах однієї години ЕЕГ. Наприклад, з 5-ти хвилинним

перетином новий сегмент формується як показано на рис. 3.3. Данна процедура може виконуватись під час тренування моделі, тому немає необхідності у збереженні додаткових даних.

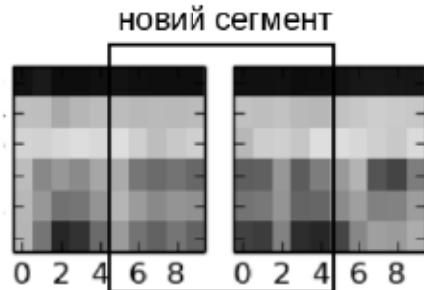


Рисунок 3.3 – Формування нового екземпляру даних

3.2 Архітектури моделей

Оскільки преіктальний період – година до настання епілептичного нападу – був визначений досить довільно, певні ознаки іктальної активності можуть бути частково присутні або повністю відсутні в деяких преіктальних сегментах. Тому бажано, щоб модель знаходила іктальні ознаки в різні моменти часу і комбінувала вивчену інформацію для створення одного передбачення на кожен ЕЕГ сегмент. Данна ідея була реалізована на основі конволютивної нейронної мережі, що виконує одновимірну конволюцію у часі, таким чином, вивчаючи однакові типи ознак з різних часових проміжків в межах сегменту. У глибших шарах мережі, дана локальна інформація поєднується.

Не менш важливою є інформація про зв’язки між ЕЕГ каналами. В [10] пропонується включити попарні ознаки між каналами, наприклад кроскореляції. В даній роботі запропонована архітектура КНН, що автоматично враховує подібні відношення. Для цього конволютивні фільтри в першому шарі КНН повинні бачити всі частотні смуги усіх ЕЕГ каналів.

В даній роботі були розроблені кілька архітектур КНН, в наступних параграфах далі ми надамо їх детальний опис.

Приклад архітектури першого типу зображеній на рис. 3.4. Перший шар (C1) виконує конволюцію у часі, охоплюючи усі 6 частотних смуг усіх N каналів, тому розмір фільтрів дорівнює $(6 \cdot N) \times 1$. C1 містить 16 карт ознак, кожна розмірністю 1×10 ; кількість параметрів в даному шарі: $16 \cdot 6 \cdot N + 16$ (додаткові 16 відповідають зміщенням). Наступний шар (C2) виконує конволюцію фільтрами 16×2 , тому результиуючі карти ознак мають ширину $10 - 2 + 1 = 9$. C2 має 32 карти ознак та $16 \times 2 \times 32 + 32 = 1056$ параметрів. C2 щільно з'єднаний з F3, у якому 128 нейронів, тому на даному рівні $32 \times 9 \times 128 + 128 = 36992$ параметри. Вихідним шаром слугує логістична регресія зі 129 вагами. Всього дана модель містить $\approx 40K$ параметрів, які необхідно вивчити під час тренування.

Оскільки між останнім конволютивним шаром та прихованим шаром F3 присутні всі зв'язки між нейронами, модель враховує часову позицію вивчених ознак у шарі C2. Для даного типу моделей природньою є схема нормалізації, яка розгладає кожну спектrogramу у вигляді вектора, таким чином враховуючи час та частотну смугу кожної входної ознаки.

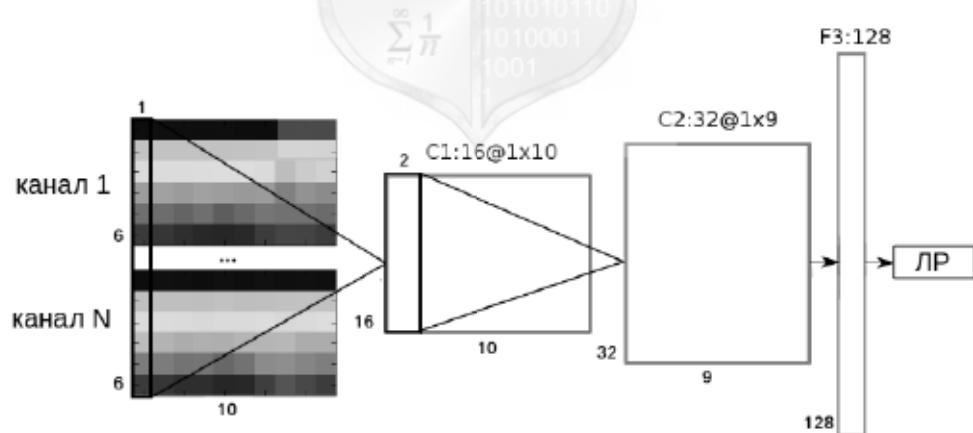


Рисунок 3.4 – Приклад архітектури моделі типу 1 (пояснення в тексті)

Архітектура другого типу (рис. 3.5) додатково містить об'єднуючий шар [25], який рахує статистики (середнє арифметичне та геометричне, максимум, мінімум, дисперсію, L2 норму) вивчених ознак у шарі C2 крізь часову вісь. Таким чином, GP3 містить $6 \times 32 = 192$ нейрони, які з'єднані зі 128 нейронами в шарі F4, тому між GP3 і F4 існує 24704 пара-

метри. Об'єднучий шар не має власних вагів, тому дана архітектура має на $\approx 12K$ параметрів менше ніж попедня архітектура.

Присутність GP3 в даній моделі дозволяє ліквідувати часову інформацію, яка інтуїтивно є нерелевантною. Важливим є факт детекції певних ознак іктальної активності, а не їх місце в 10-хвилинному сегменті. Для даного типу моделей нормалізація відбувалась за схемою, яка розглядає кожну спектrogramу як набір 6-вимірних векторів.

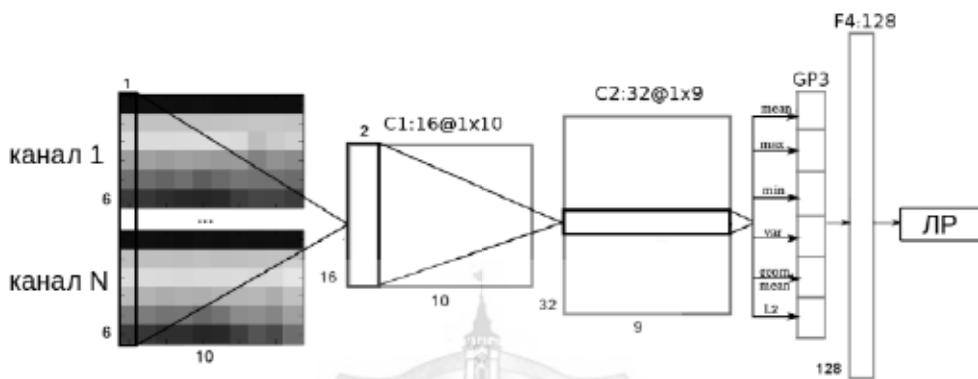


Рисунок 3.5 – Приклад архітектури моделі типу 2 (пояснення в тексті)

В обох типах моделей функцією активації у конволютивних шарах слугувала лінійна ректифікація і, як правило, $tanh$ активація у прихованому шарі. Включення до архітектури субдискретизуючого шару між С1 та С2 погіршувало результативність моделей, на відміну від використання конволюції з кроком 2 у шарі С2.

3.3 Тренування моделей

Тренування моделей здійснювалось методом ADADELTA з параметрами $\rho = 0.95$ та $\varepsilon = 10^{-6}$. Однак, оригінальний алгоритм був модифікований у зв'язку з нестабільним оптимізаційним процесом. Приклад навчальної кривої, що зображає значення цільової функцію відносно кількості ітерацій (одна ітерація відповідає зміні параметрів) наведена на 3.6 сірою лінією. Очевидно, що алгоритм оптимізації не збігається, тому після достатньої кількості ітерацій можливо опинитись в точці з

дуже високим значенням цільової функції. Причин даної поведінки невідомі, також ми не знайшли подібних згадок в літературі.

Якщо переписати оригінальне правило зміни параметрів ADADELTA (рівняння 1.11) як

$$\Delta\omega_t \leftarrow -\eta \cdot \frac{RMS[\Delta\omega]_{t-1}}{RMS[g]_t} \cdot g_t, \quad (3.1)$$

де за замовчуванням $\eta = 1.0$, можна побачити, що контролювати величину кроку можна за рахунок η . Зменшивши η до 0.01, був отриманий задовільний результат, що наведений чорною лінією на рис. 3.6.

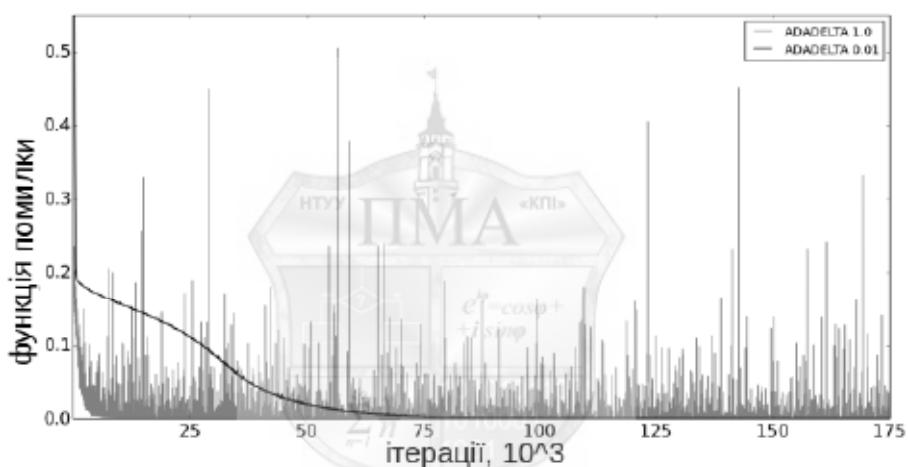


Рисунок 3.6 – Значення цільової функції під час тренування методом ADADELTA при різному навчальному кроці

У щільних шарах кожної моделі ми використовували алгоритм виключення. Спроби використання методу виключення в конволютивних шарах а також виключення випадковим чином ЕЕГ каналів на вході мережі, як правило, погіршувало результати.

Кожна мережа тренувалась протягом 175K ітерацій. Адаптувати метод ранньої зупинки не було можливим, оскільки в даній задачі використання валідаційної множина в даній задачі не дає репрезентативної оцінки. Джерелом валідації наших КНН моделей була публічна тестова множина.

Для імплементації моделі КНН була використана мова програмування Python.

ння Python та компілятор Theano [26, 27]. Theano автоматично здійснює символьне диференціювання а також дозволяє виконувати програму на CPU та GPU. Використовуючи NVIDIA GeForce GTX 980, тренування моделей для усіх семи пацієнтів тривало кілька годин.

3.4 Калібрація передбачень

Щоб компенсувати ефект проблематичної метрики оцінки, організатори змагання дозволили використовувати тестові дані для калібрації тестових передбачень моделей між пацієнтами. Ми використали мін-макс нормалізацію, яка для кожного пацієнта масштабує передбачення p до 0-1 інтервалу за допомогою формули:

$$P_{scaled} = \frac{p - \min(\mathbf{p})}{\max(\mathbf{p}) - \min(\mathbf{p})}, \quad (3.2)$$

де \mathbf{p} є вектором передбачень для одного пацієнта.

Дані трансформації не мають ефекту на AUC, обрахований для кожного з пацієнтів окремо. На практиці це означає, що результативність моделі не змінюється. Однак, при комбінованому AUC калібрація має значний ефект на результат, отриманий у змаганні.

3.5 Результати

Результат, який отримав 10те місце з AUC 0.81292 на публічній та 0.78513 на приватній тестовій вибірці, був отриманий в результаті усереднення передбачень, відкаліброваних за допомогою мін-макс схеми, 11ти КНН моделей. Всі моделі в ансамблі мали об'єднучий шар, але різну кількість карт ознак на кожному шарі, розмірність фільтрів, ймовірність виключення нейронів в щільних шарах, etc.

В таблиці 3.1 наведено кілька цікавих результатів. Данна таблиця містить опис даних, на яких тренувалась модель, опис архітектури та ре-

зультати на публічній та приватній тестовій множині з мін-макс калібрацією та без. Напрклад, для другої моделі ДПФ та середнє квадратичне відхилення ЕЕГ сигналу рахувалось в 1-хвилинних вікнах; для кожного каналу спектrograma була розбита на 6 частотних смуг; між сегментами при тренуванні були використані 9-хвилинні перетини. Дані мережа має 2 конволютивні шари з лінійною ректифікацією як функцією активації, об'єднуючий шар та один щільний шар з $tanh$ активацією; виходом мере жі слугувала логістична регресія. Шар C2 виконує конволюцію з кроком 2. В останніх двох шарах використовується виключення нейронів з ймо вірностями 0.3 та 0.6.

Накраща відкалібрована модель на публічному наборі даних виявилась модель типу 1 – з усіма зв'язками між конволютивним шаром C2 та прихованим шаром F3. Це свідчить, що релевантним для задачі є розташування ознак у часі. Однак, дана модель має труднощі з генералізацією і тому її результат значно нижчий на приватній тестовій вибірці. Варто зауважити, що під час змагання був доступним лише результат на публічній множині (аналог валідаційної), і зазвичай моделі типу 1 перевершували результати моделей типу 2, що містили об'єднучий шар. Але, покладаючись на інтуїцію, ми більшою мірою досліджували моделі другого типу. Як бачимо, кращий результат на публічній вибірці не призводить кращого на приватній, як у випадку моделі 2. Таким чином, використання однієї моделі є ризикованим і створення ансамблів моделей було необхідністю (модель 3 була моделлю з найгіршим результатом серед інших моделей ансаблю). Можна помітити, що калібрування передбачень дає лише незначне покращення на $\approx 1\%$.

Таблиця 3.1 – Результати ROC AUC вибраних архітектур на публічній (валід.) та приватній (тест.) вибірці (пояснення в основному тексті)

№	Дані	Архітектура	Валід./ Тест. (мін-макс калібра- ція)	Валід./ Тест. (без калібрації)
1	6 частот. смуг; 1хв ДПФ вікна	C1: 16@1x10, ReLU C2: 32@1x9, ReLU F3: 128, ReLU, p=0.2 ЛР: p = 0.6	0.81448/ 0.76258	0.79899/ 0.75201
2	6 частот. смуг + стд. відхилення; 1хв ДПФ вікна; 9 хв перетин	C1: 32@1x10, ReLU C2: 64@1x5, ReLU GP3 F4: 512, tanh, p=0.3 ЛР: p=0.6	0.80872/ 0.77182	0.80276/ 0.76432
3	8 частот. смуг + стд. відхилення; 30 сек ДПФ вікна; 5 хв перетин	C1: 16@1x20, ReLU C2: 32@1x10, крок 2, ReLU ^{cosp +} GP3 ^{sinc} F4: 512, tanh, p=0.3 ЛР: $\sum \pi p=0.6$	0.78612/ 0.76133	0.77949/ 0.75465

3.6 Висновки

У розділі описаний запропонований метод на основі КНН, точніше два типи архітектур мережі, які здатні працювати з ЕЕГ даними, представленими набором спектральних ознак сигналу. Був проведений аналіз методу градієнтного спуску з адаптивним кроком ADADELTA, в результаті якого нам довелось модифікувати оригінальний алгоритм у зв'язку з його нестабільною поведінкою при оптимізації цільової функції. Наочностанок, були наведені результати, досягнуті запропонованим методом у змаганні “AES Seizure Prediction Challenge”.

4 МЕТОД НА ОСНОВІ ЛІНІЙНИХ КЛАСИФІКАТОРІВ

Метод на основі лінійних моделей класифікації був запропонований командою, яка зайніяла 2ге місце у змаганні. Деталей для репродукції їх результатів немає, але загальна ідея є наступною:

1. для кожного з пацієнтів побудувати індивідуальний класифікатор на основі лінійної регресії з регуляризацією, де одиницею даних являється 1 хвилина ЕЕГ з оригінального 10-хвилинного сегменту. Залежна змінна була представлена міткою класу 0 або 1.
2. порахувати середнє значення передбачень для кожної 1 хвилини, таким чином формуючи передбачення для 10-хвилинного сегменту
3. нормалізувати передбачення та звести їх за допомогою логістичної функції до (0,1) інтервалу (надалі “софтмакс масштабування”).

Оскільки застосування лінійної регресії для задачі класифікації не є наращеною практикою, ми будемо застосовувати логістичну регресію та ЛДА у якості класифікуючого алгоритму.

4.1 Обробка даних

Перші кроки з перетворення сигналу у спектrogramами є аналогічними до описаних у розділі 3.1 з наступними відмінностями:

1. вікна, в яких рахується ДПФ, мають перетин у 30 секунд. Наприклад, якщо перше вікно містить секунди 0-60 оригінального ЕЕГ сигналу, тоді наступне вікно покриває секунди з 30 до 90. Таким чином, з 10-хвилинного сегменту отримується 19 кадрів.
2. розбиття частотних смуг є наступним: 1Гц смуги в діапазоні частот 1-50Гц, 5Гц смуги в 50-100Гц і 10Гц в діапазоні 100-180Гц. В результаті 67 отримемо частотних смуг на кожен 1-хвилинний кадр.

4.2 Навчання моделі

На відміну від моделей на основі конволютивних неронних мереж, одиницею даних є 1-хвилинні кадри, тому з оригінального ЕЕГ сегменту отримуємо 19 екземплярів відповідного класу, кожен з яких містить ($N \times 67$) спектральних ознак. Після нормалізації векторів, можливо застосувати регуляризовану логістичну регресію або ЛДА, щоб отримати передбачення для кожної хвилини ЕЕГ з тестової множині. Щоб отримати передбачення для 10-хвилинного сегменту, ми усереднюємо передбачення відповідних 1-хвилинних кадрів: $p = \sum_{i=0}^{18} p_{1\text{xb}}^{(i)}$. На додачу, p можуть бути відкаліброкані, спершу нормалізуючи передбачення для кожного пацієнта окремо і трансформуючи логістичною функцією до 0-1 інтервалу (софтмакс нормалізація):

$$P(Y = 1|x)_{scaled} = (1 + \exp(\frac{p - \mu}{\sigma}))^{-1}, \quad (4.1)$$

де μ та σ є значеннями математичного очікування та середнього квадратичного відхилення передбачень у тестовій вибірці. Варто зауважити, що на практиці дані оцінки є невідомими, тому наведений трюк працює лише у описаному змаганні. Надалі буде явно вказуватись, чи даний тип калібрації був застосований.

4.3 Валідація моделі

Валідація моделі використовується для оцінки якості моделі відносно даних, які не були використані під час тренування або/та для підбору гіперпараметрів. Валідація полягає в розбитті оригінальної тренувальної вибірки на тренувальну та валідаційну частини. Кращу оцінку можна отримати процедурою крос-валідації, коли процедура розбиття повторюється кілька разів [28].

Для аналізу ЕЕГ конструкція процедур крос-валідації є важливим аспектом, оскільки кількість даних зазвичай є обмеженою, щоб ма-

ти окрему репрезентативну валідаційну множину. У випадку, коли дані для крос-валідації розбиваються випадковим чином, результуюча валідаційна оцінка не є достовірною оцінкою помилки на тестовій множині з кількох наступних причин, які притаманні ЕЕГ даним.

По-перше, сегменти даних в межах однієї години є більш схожими, ніж сегменти з різних годин. На рис. 4.1 зображений t-SNE графік преіктальних даних для одного з пацієнтів. Для побудови t-SNE ми перетворили $N \times 6 \times 10$ спектрограми у вектори і зменшили їх розмірність до 50ти за допомогою методу головних компонент [28], використовуючи всю тренувальну вибірку. Очевидно, що сегменти з однієї години формують кластери (окреслені на рисунку контурами), що і доводить наше твердження.

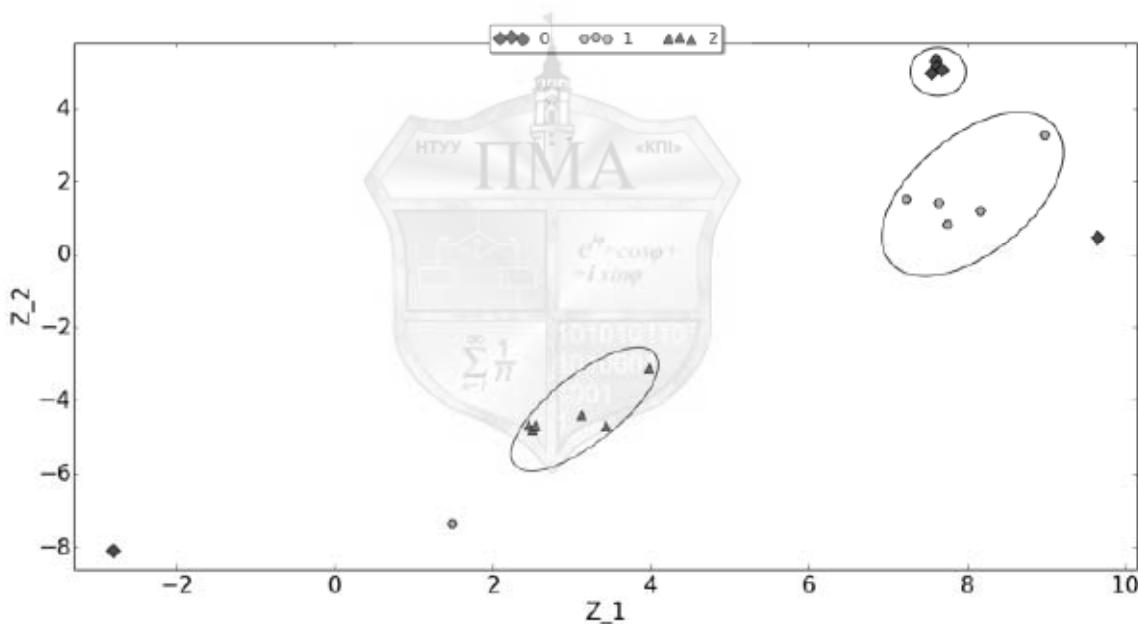


Рисунок 4.1 – Візуалізація преіктальних сегментів методом t-SNE

Випадкове розбиття може віднести сегменти з однієї години в тренувальну та валідаційну вибірки, що призводить до оптимістичних валідаційних оцінок, оскільки модель бачить схожі екземпляри під час тренування. Рішенням даної проблеми є збереження 1-годинних послідовностей лише у тренувальній або валідаційній вибірках.

По-друге, візуалізувавши тренувальні та тестові дані за допомогою t-SNE для одного з пацієнтів, можна переконатися, що розподіл в двох вибірках суттєво відрізняється. Даний феномен в літературі відомий як

коваріативний зсув (англ. covariate shift) [29]. Наразі спроб валідації методів для аналізу ЕЕГ, яка враховує даний фактор, до нашого відому не існує. Можливо, єдиним рішенням боротьби з нестационарністю ЕЕГ є активне навчання класифікаторів [3].

По-третє, кількість преіктальних та інтеріктальних сегментів є не-збалансованою (див. таблицю 2.1), тому розбиваючи дані випадковим чином, в найгіршому випадку тренування або валідація відбуватиметься лише на інтеріктальних даних, що складають переважну більшість. Для уникнення подібної ситуації використовується стратифікація: збереження кількісного відношення між класами у кожній з вибірок.

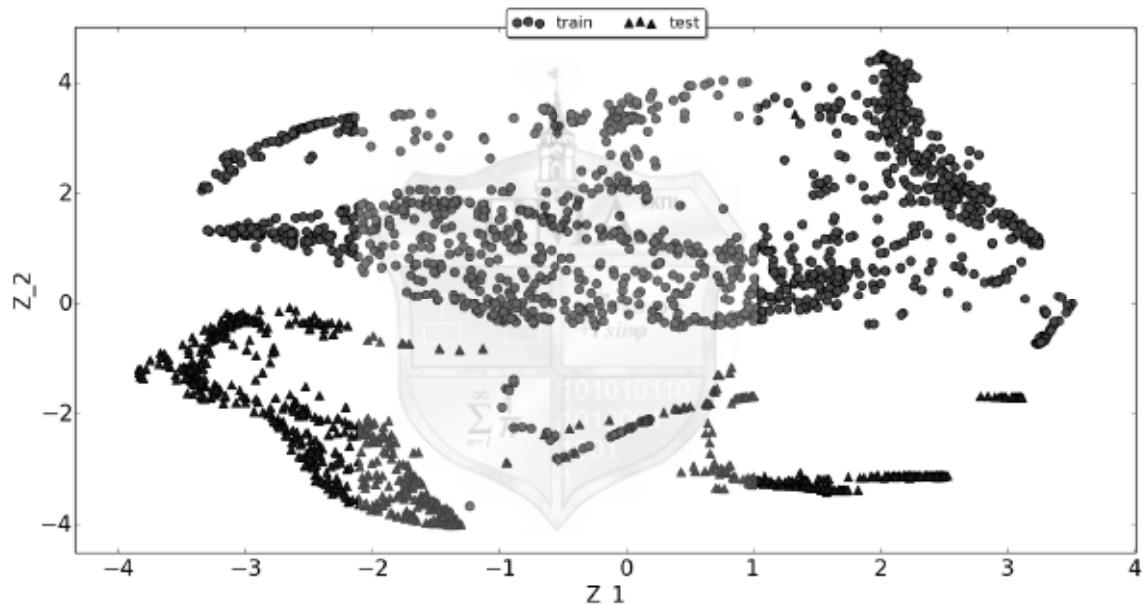


Рисунок 4.2 – Візуалізація тренувальних та тестових сегментів одного з пацієнтів методом t-SNE

Підсумовуючи, запропонований алгоритм крос-валідації виглядає наступним чином: вибрati кiлькiсть валiдацiйних iтерацiй рiвною кiлькостi преiктальнiх сегментiв, тодi кожна валiдацiйна частина мiститиме 1 преiктальну та $\frac{n_{interictal}}{n_{preictal}}$ iнтерiктальнiх послiдовностей. На кожнiй iтерацiї зберiгається предбачення моделi на валiдацiйнiй множинi i рахується глобальний AUC критерiй, сподiваючись що отримана в результатi оцiнка вiдображатиме результатi на тестовiй множинi.

Дана процедуру крос-валідації була використана для порiвняння

ЛДА та логістичної регресії з кількома значеннями параметру λ – сили L_2 регуляризації. Результати наведені в таблиці 4.1. В рамках логістичної регресії результати крос-валідації є узгодженими з результатами на публічній тестовій вибірці, однак різниця порядку 10^{-3} не є надійним індикатором переваги одного методу над іншим. Очевидно, що крос-валідація не може бути використана для порівняння моделей різних типів: логістична регресія з $\lambda = 10$ є найкращою за результатами крос-валідації, але на публічній тестовій множині ЛДА набагато перевершує її. Можливо, даний феномен виникає через неузгодженість AUC та цільових функцій, які оптимізовують ЛДА або логістична регресія. Проблема розробки надійної крос-валідаційної процедури є нетривіальною задачею. Для порівняння в таблиці 4.1 ми також навели результати крос-валідації, при якій множина розбивається випадковим чином, але зі стратифікацією. Легко було передбачити, що дані результати є занадто оптимістичною оцінкою тестової AUC.

Таблиця 4.1 – Результати регуляризованої логістичної регресії та ЛДА на крос-валідації та публічній тестовій множині

	ЛР $\lambda = 0.0$	ЛР $\lambda = 0.1$	ЛР $\lambda = 10.0$	ЛР $\lambda = 100.0$	ЛДА
Крос-валідація	0.85298	0.85433	0.85767	0.85296	0.83749
Публічна вибірка	0.68562	0.69712	0.70820	0.67735	0.74287
Випадкова крос-валідація	0.93356	0.93028	0.91430	0.89769	0.91130

4.4 Аналіз моделі

Згідно правил змагання, заборонялось обирати моделі та окремі ознаки для кожного пацієнта окремо. Наприклад, для одного пацієнта використовувати спектральні ознаки лише з гама смуги, у іншого – ви-

користовувати лише кілька вибраних ЕЕГ каналів. Неприйнятність такого підходу диктується його непрактичністю а також потенційною небезпекою, оскільки обираючи ознаки на порівняно невеликій валідаційній множині кожного пацієнта, в тестовому режимі модель може поводитись непередбачувано погано. Тому бажаним є визначення універсального набору ознак для всіх пацієнтів. Щоб зрозуміти які ознаки є дискримінтивними в даній задачі, ми використаємо ЛДА.

Розпочнемо аналіз з даних, які містять 6 основних спектральних смуг. Для одного з пацієнтів (Dog_1) зображені абсолютні значення дискримінінтних δ коефіцієнтів на рис 4.3. Кожна гістограма представляє 16 ЕЕГ каналів, де 6 прямокутників позначають 6 коефіцієнтів (від 0 до 5), що відповідають дельта, тета, альфа, бета, низьким та високим гама частотним смугам. Неважко помітити що для кількох каналів найбільшими є значення гама коефіцієнтів, тому можна зробити висновок, що саме вони мають найбільшу дискримінтивну силу. Схожа поведінка спостерігається також для інших пацієнтів. Знаючи, що корисна інформація знаходиться у смугах високих частот, ми можемо розбити їх на вужчі смуги і отримати більш релевантні вхідні ознаки. Наприклад, розбити дві гамма смуги на 2 частини (30-50Гц, 50-70Гц, 70-100Гц, 100-180Гц), що в сумі дає 8 частотних смуг для кожного з каналів, а також на 4 частини з межами 30, 40, 50, 60, 70, 85, 100, 140 та 180 Гц, що породжує 12 частотних смуг.

Результати роботи ЛДА на даних з різним розбиттям частот наведені в таблиці 4.2. Бачимо, що використання 8 частотних смуг має на кращі результати на публічній та приватній тестовій вибірці. При дрібнішому розбитті, погіршується генералізація моделі. З даної таблиці також очевидно, що результати на тестовій приватній та публічній вибірках не ідеально корельовані, тому обираючи модель, не можна повністю покладатись на публічну тестову вибірку. Софтмакс калібрація має великий вплив на результативність моделі: відносний приріст AUC після калібрації складає $\approx 7\%$.

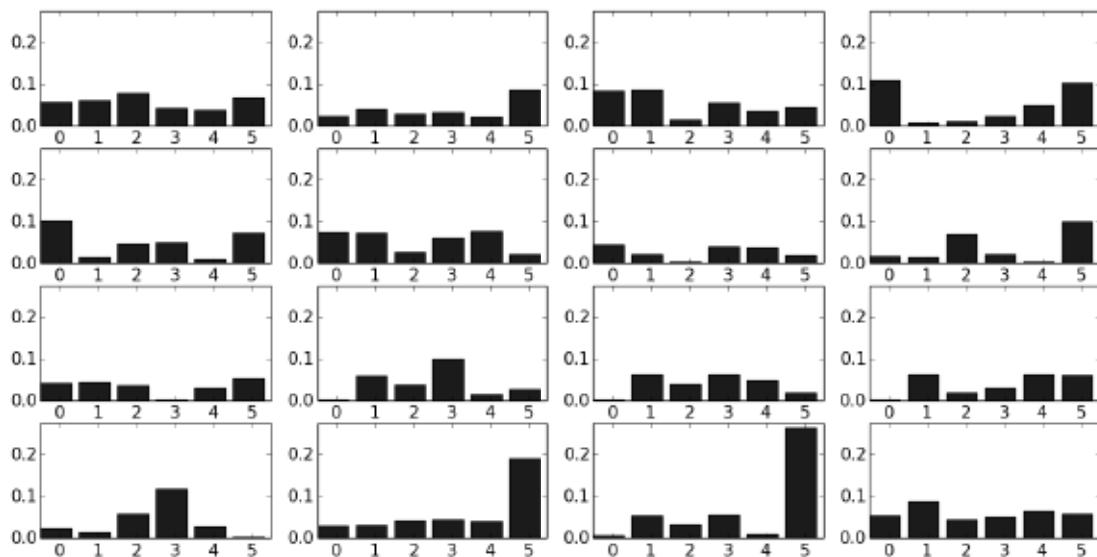


Рисунок 4.3 – Абсолютні величини коефіцієнтів ЛДА, що відповідає бти спектральним ознакам з 16 ЕЕГ каналів

Таблиця 4.2 – AUC, отриманий ЛДА на приватній та публічній тестових вибірках при різній кількості частотних смуг (6, 8, 12, 67)

	Публічний (софтмакс)	Приватний (софтмакс)	Публічний (без калібрації)	Приватний (без калібрації)
6	0.79939	0.78787	0.74287	0.72723
8	0.79894	0.79481	0.75268	0.74224
12	0.80044	0.79539	0.74528	0.72823
67	0.81405	0.79211	0.74693	0.69885

4.5 Висновки

У розділі описаний метод, загальну ідею якого ми запозичили у одного з учасників-переможців змагання, хоча принцип класифікації окремих 1-хвилинних кадрів ЕЕГ та усереднення отриманих передбачень був досить поширенним підходом. Оскільки, деталі оригінального алгоритму автори не надали, була розроблена варіація з використанням ЛДА, яка може відтворити їх результати. Однак, ключем до успіху лінійних класифікаторів є калібрація передбачень моделі, використовуючи тестову

множину. Дано ситуація є прикрою, оскільки вона не відповідає справжнім умовам експлуатації алгоритму для передбачення епілептичних нападів.

Окрім аналізу результатів, були досліджені труднощі, що виникають при роботі з ЕЕГ даними, головним джерелом яких є нестационарність ЕЕГ сигналів. Даний факт унеможливлює використання стандартних валідаційних процедур. Запропонована нами схема крос-валідації, яка враховує природу ЕЕГ, також не дала бажаних результатів.

Насамкінець, за допомогою ЛДА ми дослідили які з ознаки ЕЕГ сигналу є дискримінativними в даній задачі.



5 ПОРІВНЯЛЬНИЙ АНАЛІЗ РОЗРОБЛЕНИХ МЕТОДІВ

5.1 Порівняння КНН та ЛДА методів

Дотепер ми розглядали глобальну AUC, що була єдиною доступною мірою оцінки протягом змагання. Оскільки, після завершення змагання стала відомою класова приналежність для кожного екземпляру тестових даних, в даному розділі ми проаналізуємо результативність представлених моделей для кожного пацієнта окремо.

Далі ми розглянемо фінальний ансамбль КНН моделей, найгіршу модель в ансамблі (номер 4 з таблиці 3.1) і ЛДА модель на основі даних з вісімома частотними смугами. Таблиця 5.1 надає результати моделей для кожного пацієнта обрахованих для всієї тестової множини (публічна та приватна разом). Порівняння ансамблю КНН та ЛДА за допомогою зваженого тесту Ст'юдента[30] (ваги відповідали кількості даних кожного пацієнта) не виявило статистично значимої різниці: p-значення=0.71 і 95% довірчий інтервал ($-0.111, 0.154$). Незважаючи на низький результат відносно глобальної AUC, одна нейронна мережа відносно кожного пацієнта працює не гірше. При порівнянні з ЛДА результат не є статистично значимим: p-значення=0.57 і 95% довірчий інтервал ($-0.082, 0.136$).

Було легко передбачити, що для Patient_1 та Patient_2 результати КНН виявляться невтішними, оскільки кількість тренувальних екземплярів була набагато меншою кількості параметрів мережі. Однак, дані людей складали лише 9% від усього набору, тому ми сфокусувались на суб'єктах Dog_1-Dog_5

Таблиця 5.1 – ROC AUC на тестовій множині порахована для кожного з пацієнтів окремо, де передбачення отримані ансамблем КНН(1), однією КНН (2) та ЛДА (3)

	Dog_1	Dog_2	Dog_3	Dog_4	Dog_5	Patient_1	Patient_2
1	0.79489	0.78937	0.80512	0.90656	0.78864	0.53734	0.58981
2	0.76456	0.81243	0.81547	0.89970	0.73138	0.67304	0.56066
3	0.74294	0.86694	0.74338	0.87601	0.29609	0.73452	0.63971

Рис. 5.1 та 5.2 ілюструють ймовірності присвоєні тестовим сегментам ЕЕГ на виході однієї КНН моделі та ЛДА відповідно (справжні мітки класів позначені різними маркерами). Без огляду на справжні класи, легко помітити, що ймовірності присвоєні КНН щільно розташовані в околі 0 та 1, можна сказати, що дана модель є більш впевненою у своїх передбаченнях незалежно від їх правильності. Це також означає, що майже будь-який поріг класифікації в околі 0.5 буде породжувати однакове значення на глобальній ROC кривій. На відміну від КНН, ймовірності отримані методом ЛДА більш розсіяні та їх межі відрізняються між суб'єктами, наприклад, як у випадку Dog_1 відносно інших пацієнтів. Таким чином ефект зміна порогу класифікації є значимішою. Відмінність між предбаченнями для кожного суб'єкта пояснює чому калібрація приносить ЛДА більшу користь, ніж для КНН, як проілюстровано в таблицях 4.2 та 3.1.

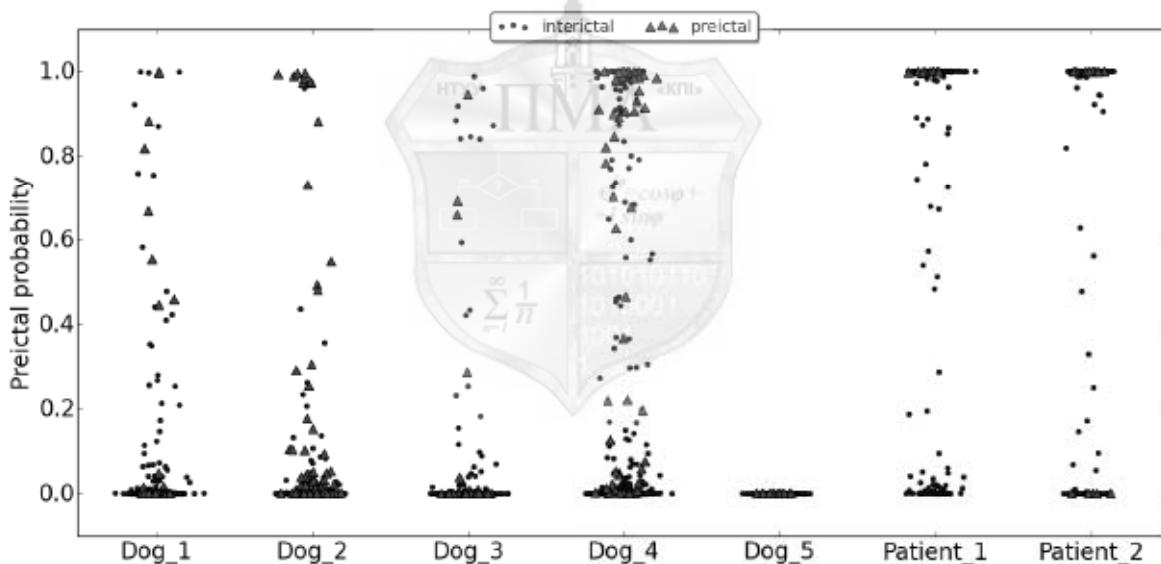


Рисунок 5.1 – Тестові передбачення отримані КНН

На практиці ЛДА є менш робастним методом відносно вибору порогу класифікації: під час тестування можна отримати більше хибнонегативних та хибнопозитивних результатів, ніж на валідаційній множині при підборі порогу. На жаль, даний факт неможливо перевірити, оскільки були надані лише мітки класів для всієї тестової множини, тобто розбиття між публічною та приватною вибірками залишилось невідоме. Також для тестових екземплярів ми не знаємо яким 1-годинним послідовностям вони належать, тому самостійно правильно розбити тестові дані

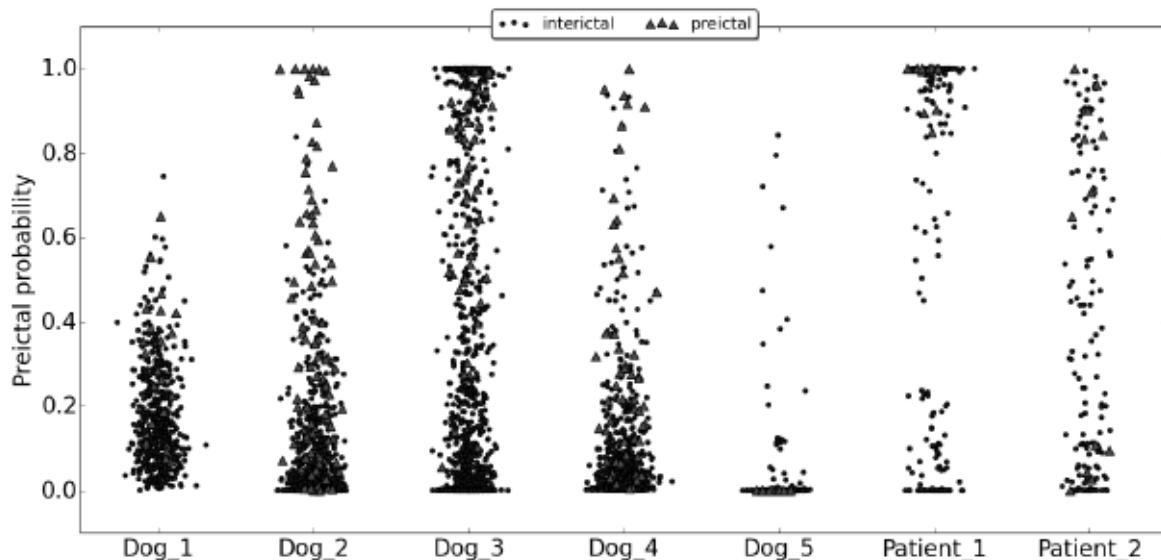


Рисунок 5.2 – Тестові передбачення отримані методом ЛДА

не є можливим. Однак, деяку інтуїцію щодо поведінки методів під час тестування можна отримати аналізуючі графіки TPR та TNR при змінному порозі класифікації. Рис. 5.3 та 5.4 зображують TPR неронної мережі та ЛДА, коли значення порогу змінюється від 0.0 до 1.0 з кроком 0.02. Бачимо, що для більшості суб’єктів та значень порогу ЛДА має кращі показники за КНН, але вони змінюються швидше, ніж для КНН. Вибір найрашції моделі може бути питанням балансування між результативністю та впевненістю, що модель буде поводитись подібним чином в режимі експлуатації. Однак, необхідно знати нижню межу чутливості класифікатора, після якої застосування алгоритму для передбачення нападів стає безглуздим. На практиці, як правило, для передбачення нападу достатньо правильно класифікувати лише один сегмент з преіктально-го періоду, тому виміряний нами TPR теж є не найкращою метрикою. Кількість істиннопозитивних результатів відносно кількості нападів, як запропоновано в [3], є більш доцільною оцінкою якості класифікуючого алгоритму для даної задачі. Як ми зазначали раніше, ми не маємо доста-тньо відомостей про тестову множину, щоб порахувати дані оцінки.

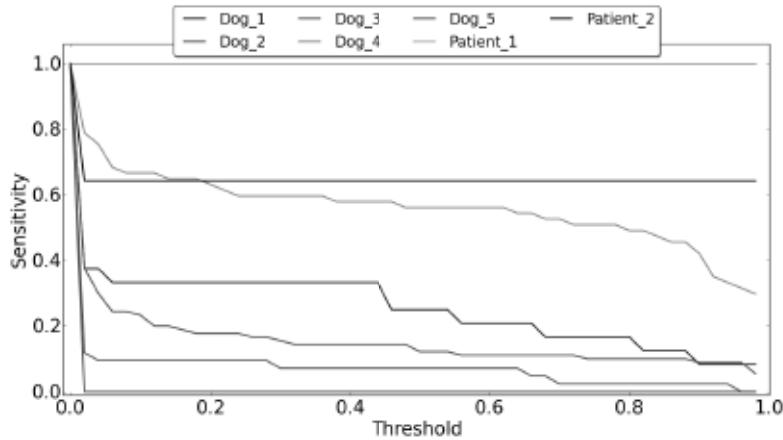


Рисунок 5.3 – Чутливість КНН моделі залежно від порогу класифікації для кожного з пацієнтів

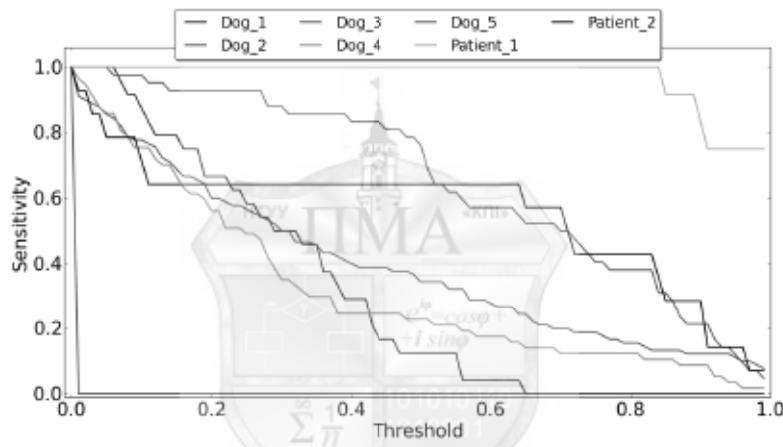


Рисунок 5.4 – Чутливість ЛДА моделі залежно від порогу класифікації для кожного з пацієнтів

Рис. 5.5 та 5.6 зображають аналогічні графіки для TNR. На відміну від ЛДА, TNR криві для КНН є більш пологими. Висока специфічність класифікатора є важливою в задачі передбачення нападів, оскільки велика кількість хибних попереджень призводить до так званої “втоми”, коли пацієнт починає ігнорувати попередження системи [31].

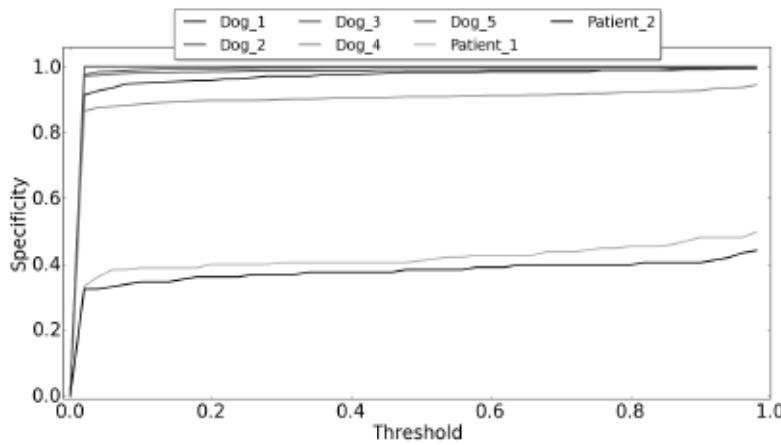


Рисунок 5.5 – Специфічність КНН моделі залежно від порогу класифікації для кожного з пацієнтів

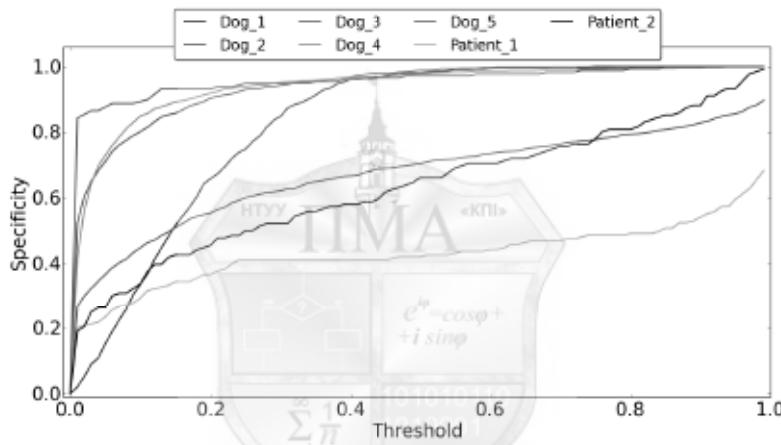


Рисунок 5.6 – Чутливість КНН моделі залежно від порогу класифікації для кожного з пацієнтів

Недоліком нашої КНН моделі є неможливість визначення оптимального значення порогу класифікації на тренувальній множині, як запропоновано в [3]. Це пояснюється тим, що тренувальні ймовірності завжди є близькими до 0 та 1 і будь-який поріг досягає ідеальної класифікації. Для ЛДА ситуація відрізняється, тому оптимізувати поріг можна на тренувальній множині. Критерій, що був використаний під час мінімізації, була відстань між ROC кривою та точкою (0,1) в ROC просторі, що відповідає максимальній чутливості та нульовому рівню хибної класифікації. Результати ми порівняли з КНН при порозі 0.5, результати TPR та TNR обраховані для тестової множини наведені в таблиці 5.2. Не зважаючи на упередженість такого порівнняння, можна стверджувати що

порівняно прості методи, такі як ЛДА, в деяких аспектах перевершують більш складні підходи.

Таблиця 5.2 – Чутливість та специфічність ЛДА та КНН на тестовій мно-жині

	Dog_1	Dog_2	Dog_3	Dog_4	Dog_5	Patient_1	Patient_2
TPR ЛДА	0.792	0.5	1.000	0.719	0.000	1.000	0.643
TPR КНН	0.250	0.122	0.071	0.561	0.000	1.000	0.643
TNR ЛДА	0.423	0.945	0.345	0.869	0.950	0.464	0.721
TNR КНН	0.983	0.999	0.980	0.909	1.000	0.410	0.382

5.2 Аналіз інших підходів

Ми порівняли підходи інших учасників змагання: стандартним методом було тренування класифікаторів для окремих часових вікон тривалістю 15-75 секунд та усереднення отриманих передбачень, подібно до описаного підходу з використанням ЛДА. На додачу до різних частотних характеристик (найчастіше використовувалась спектральна енергія), учасники рахували парні кореляції між ЕЕГ каналами в часовій та частотній області, власні числа даних кореляційних матриць, фрактальні розмірності, статистичні моменти ЕЕГ сигналу та ін. Варто зауважити, що обрахунок деяких з наведених ознак вимагає обчислювальних ресурсів, які можуть бути недоступними у приладу, який аналізує ЕЕГ. Серед популярних класифікаторів застосовувались метод опорних векторів, логістична регресія, рандомний ліс та ін. Можна впевнено стверджувати, що наш метод на основі КНН є оригінальним, оскільки він потребує дуже простих входних ознак і класифікація 10-хвилинних сегментів відбувається агрегацією інформації всередині мережі, а не пост-фактум – після класифікації менших частин з сегменту.

Глибокі нейронні мережі (до 6 прихованих шарів) в комбінації з методом k-найближчих сусідів був запропонований командою з Університету Кардинала Еррери (Universidad CEU Cardenal Herrera), яка виграла Зій приз змагання [32]. Їх найкраща одиночна модель була ШНМ з 5а прихованими шарами і мала результати 0.7937 і 0.7644 на публічній та приватній множині відповідно, що порівнювано з нашими результатами (див. таблицю 3.1). Причиною їх переваги було використання Баєсівської комбінації моделей [33].

З персональної переписки ми дізналися про учасника, який також використовував конволютивні нейронні мережі, і фінішував на 245му місці з результатом 0.60120. Його підхід базувався на біваріативних ознаках як в [10]. Припустимо, що на відміну від статті, де використовувалась невелика кількість ЕЕГ каналів, в даному змаганні метод був недієздатним, оскільки кількість ЕЕГ каналів була більшою. Це ще раз ілюструє, що до результатів досліджень в області аналізу ЕЕГ потрібно ставитись скептично.

В таблиці 5.3 наведені результати топ-10 моделей. Цікавим є факт, що між приватними та публічними заликами був досить великий розрив, наприклад на публічній множині наш результат був шістнадцятим. Даний факт можна пояснити валідацією моделей на публічній вибірці, таким чином перенавчаючи її без змоги генералізації результатів на приватній тестовій множині.

Таблиця 5.3 – Фінальні результати учасників змагання

№	Назва команди	Публічний AUC	Приватний AUC
1	Medrr	0.90316	0.83993
2	QMSDP	0.85951	0.81962
3	Birchwood	0.83869	0.80079
4	ESAI CEU-UCH	0.82488	0.79347
5	Michael Hills	0.86248	0.79251
6	KPZZ	0.82051	0.79136
7	Carlos Fernandez	0.84225	0.79063
8	Isaac	0.84197	0.78863
9	Wei Wu	0.81803	0.78724
10	golondrina	0.82455	0.78513

5.3 Висновки

У розділі наведені результати порівняння методів на основі КНН та ЛДА використовуючи різні метрики оцінки якості бінарної класифікації, які релевантні для задачі перебачення нападів. Однозначного висновку про перевагу одного методу над іншим зробити не можна, оскільки їх результати є комплементарними: інколи можна пожертвувати кращою результативністю ЛДА методу на користь впевненості у тестових результатах, що дає КНН.

Також був зроблений огляд методів, запропонованих іншими учасниками змагання. Наш метод на основі КНН докорінно відрізняється (хоча ми не стверджуємо, що це є перевагою) від усіх інших, тому що розглядає 10-хвилинні сегменти ЕЕГ як одне ціле, а не намагається класифікувати окремі фрагменти сегменту, як наприклад в описаному методі з використанням ЛДА. Головною позитивною рисою нашого методу ми вважаємо використання дуже простих спектральних ознак ЕЕГ сигналу, тобто трудоємний процес пошуку релевантних ознак був майже виключений з процесу вирішення поставленої задачі.

ВИСНОВКИ

В даній магістерській дисертації був запропонований метод на основі конволютивних нейронних мереж для аналізу ЕЕГ даних, зокрема для передбачення епілептичних нападів. Була розроблена новітня архітектура КНН, що здатна ефективно класифікувати ЕЕГ, отримуючи на вході прості спектральні ознаки ЕЕГ сигналу. Ми апробували запропоновану модель в “American Epilepsy Society Seizure Prediction Challenge” – міжнародному змаганні з передбачення епілептичних нападів.

Не вперше КНН були застосовані для аналізу ЕЕГ, однак спираючись на наші результати та результати інших учасників змагання, архітектура представлена в даній роботі найкраще підходить для аналізу багатоканальних ЕЕГ сигналів. Припустимо, що нашу модель можна використовувати для інших фізіологічних даних схожої з ЕЕГ структури, наприклад, магнітоелектроенцефалограм. Однак, наш метод не є універсальним методом аналізу багатовимірних часових рядів будь-якої природи, оскільки важливим етапом залишається вибір початкової репрезентації даних. Наприклад, в задачах економетрики спектральні характеристики часового ряду навряд чи несуть важливу інформацію. Тому наступним важливим етапом досліджень є застосування нейронних мереж безпосередньо до сиріх даних в часовій області, щоб модель автоматично вивчила необхідну репрезентацію даних, релевантну для задачі. В [34] ми описали метод застосування КНН до сиріх ЕЕГ даних, який на даний момент знаходиться в процесі дослідження.

Окрім наведеного методу, був детально проаналізований метод на основі лінійної класифікації. Цікаво, що досить простий ЛДА при порівнянні результатів окремо для кожного пацієнта, може перевершити складніші алгоритми. Однак, однозначного висновку про перевагу ондих методів над іншими зробити не можна. Проаналізувавши чутливість та специфічність методів, можна зробити висновок, що вони є доповнюючими і залежно від вподобань лікаря або пацієнта, можуть бути використані окремо або їх передбачення певним чином можна поєднати. Наш

аналіз, на жаль, був неповним, оскільки для кожного екземпляру тестових даних ми не мали інформації щодо його часу відносно ЕЕГ послідовностей, тому не можна сказати скільки нападів було передбачено. Цікавим було б здійснити оцінку моделі на усіх етапах: натренувати модель, знайти оптимальне значення порогу класифікації на відокремленій валідаційній множині, використовуючи тестову множину оцінити кількість хибних попереджень, хибнонегативних результатів а також час між правильно здійсненим передбаченням нападу та його настанням.

Ми проаналізували труднощі, які виникають з валідацією моделей, коли необхідно підібрати гіперпараметри моделі або оцінити помилку на тестовій множині. Не зважаючи на те що запропонована нами схема крос-валідації не здатна дати надійні оцінки, сподіваємось що вона принесе користь для подальших досліджень. Головним чином, було показано що будь-яка класична схема валідації непридатна у випадку ЕЕГ, головним чином, тому що розподіл даних під час тренування та тестування відрізняється. Можливо, єдиним рішенням боротьби з нестационарністю ЕЕГ є активне навчання класифікаторів.

За допомогою ЛДА ми дослідили які спектральні ознаки мають найбільшу дискримінтивну силу між преіктальними та інтеріктальними сегментами. Було показано, що розбиття смуг високих гама частот на декілька частин дає кращі результати класифікації, однак ми не провели формальних статистичних тестів для визначення важливостей високочастотних ознак. Таким чином, це є одним з напрямків майбутніх досліджень.

Результати American Epilepsy Society Seizure Prediction Challenge показали, що після майже десяти років активних досліджень в області передбачення епілептичних, не існує однозначно найкращого методу, який можна сміливо застосовувати в медичній практиці. Головними проблемами в області залишається відстність уніфікованих наборів даних, неможливість відтворення опублікованих результатів, методологічні помилки в статистичному аналізі результатів.

Було продемонстровано, що метрика, яка використовувалась для оцінки якості класифікації у даному змаганні, є ненадійною з точки зору

визначення найкращого методу. З даної причини, організатори змагання працюють над створенням нового набору даних та в подальшому виконанням глибшого аналізу.



ПЕРЕЛІК ПОСИЛАНЬ

- [1] Kindermans Pieter-Jan. A Bayesian machine learning framework for true zero-training brain-computer interfaces : Ph. D. thesis / Pieter-Jan Kindermans ; Ghent University.— Ghent University. Faculty of Engineering and Architecture, 2014.— P. 207.
- [2] Mormann Florian, Andrzejak Ralph G., Elger Christian E., Lehnertz Klaus. Seizure prediction: the long and winding road // Brain.— 2007.— Vol. 130, no. 2.— P. 314–333.
- [3] Buteneers Pieter. Detection of epileptic seizures: the reservoir computing approach : Ph. D. thesis / Pieter Buteneers ; Ghent University.— Ghent University. Faculty of Engineering and Architecture, 2012.— P. XIX, 154.
- [4] Howbert J. Jeffry, Patterson Edward E., Stead S. Matt et al. Forecasting seizures in dogs with naturally occurring epilepsy // PLoS One.— 2014.— Vol. 1, no. e81920.
- [5] Kaggle. American Epilepsy Society Seizure Prediction Challenge.— <http://www.kaggle.com/c/seizure-prediction>. — 2014.
- [6] Korshunova Iryna, Kopychko Sergii. Convolutional neural networks for epileptic seizure prediction // Applied Mathematics and computing: PMK-2015.— 2015.— P. x-x.
- [7] Krizhevsky Alex, Sutskever Ilya, Hinton Geoffrey E. ImageNet Classification with Deep Convolutional Neural Networks // Advances in Neural Information Processing Systems 25.— Curran Associates, Inc., 2012.— P. 1097–1105.
- [8] Dieleman Sander, van den Oord Aaron, Korshunova Iryna et al. Classifying plankton with deep neural networks. — <http://benanne.github.io/2015/03/17/plankton.html>. — 2015.

- [9] Längkvist Martin, Karlsson Lars, Loutfi Amy. A review of unsupervised feature learning and deep learning for time-series modeling // Pattern Recognition Letters. — 2014. — Vol. 42, no. 0. — P. 11 – 24.
- [10] Mirowski Piotr, Madhavan Deepak, LeCun Yann, Kuzniecky Ruben. Classification of patterns of EEG synchronization for seizure prediction // Clinical neurophysiology. — 2009. — Vol. 120, no. 11. — P. 1927–1940.
- [11] Park Yun, Luo Lan, Parhi Keshab K., Netoff Theoden. Seizure prediction with spectral power of EEG using cost-sensitive support vector machines // Epilepsia. — 2011. — Vol. 52, no. 10. — P. 1761–1770.
- [12] Wulsin D F, Gupta J R, Mani R et al. Modeling electroencephalography waveforms with semi-supervised deep belief nets: fast classification and anomaly measurement // Journal of Neural Engineering. — 2011. — Vol. 8, no. 3. — P. 36–45.
- [13] Pramod Siddharth, Page Adam, Mohsenin Tinoosh, Oates Tim. Detecting Epileptic Seizures from EEG Data using Neural Networks // CoRR. — 2014. — Vol. abs/1412.6502.
- [14] Anderson J.R., Michalski R.S., Carbonell J.G., Mitchell T.M. Machine learning: an artificial intelligence approach. Machine Learning. — Tioga Pub. Co., 1983.
- [15] Bengio Yoshua. Practical recommendations for gradient-based training of deep architectures // CoRR. — 2012. — Vol. abs/1206.5533.
- [16] Zeiler Matthew D. ADADELTA: An Adaptive Learning Rate Method // System Analysis and Information Technology. — 2012. — P. –1–1.
- [17] Nair Vinod, Hinton Geoffrey E. Rectified Linear Units Improve Restricted Boltzmann Machines // Proceedings of the 27th International Conference on Machine Learning (ICML-10), June 21-24, 2010, Haifa, Israel. — 2010. — P. 807–814.

- [18] Glorot Xavier, Bordes Antoine, Bengio Yoshua. Deep Sparse Rectifier Neural Networks // JMLR WCP: Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics (AISTATS 2011). — 2011.
- [19] Glorot Xavier, Bengio Yoshua. Understanding the difficulty of training deep feedforward neural networks // Proceedings of the International Conference on Artificial Intelligence and Statistics. Society for Artificial Intelligence and Statistics. — 2010.
- [20] LeCun Y., Bottou L., Y. Bengio, Haffner P. Gradient-Based Learning Applied to Document Recognition // Proceedings of the IEEE. — 1998. — November. — Vol. 86, no. 11. — P. 2278–2324.
- [21] Hinton G., Tieleman T. Lecture 10.1—Why it helps to combine models. — COURSERA: Neural Networks for Machine Learning. — 2012.
- [22] Srivastava Nitish, Hinton Geoffrey, Krizhevsky Alex et al. Dropout: A Simple Way to Prevent Neural Networks from Overfitting // Journal of Machine Learning Research. — 2014. — Vol. 15. — P. 1929–1958.
- [23] van der Maaten Laurens, Hinton Geoffrey E. Visualizing High-Dimensional Data Using t-SNE // Journal of Machine Learning Research. — 2008. — Vol. 9. — P. 2579–2605.
- [24] ROC Graphs: Notes and Practical Considerations for Researchers : Rep. ; Executor: Tom Fawcett : 2004.
- [25] Dieleman Sander. Recommending music on Spotify with deep learning. — <http://benanne.github.io/2014/08/05/spotify-cnns.html>. — 2014.
- [26] Bergstra James, Breuleux Olivier, Bastien Frédéric et al. Theano: a CPU and GPU Math Expression Compiler // Proceedings of the Python for Scientific Computing Conference (SciPy). — 2010.
- [27] Bastien Frédéric, Lamblin Pascal, Pascanu Razvan et al. Theano: new features and speed improvements. — Deep Learning and Unsupervised Feature Learning NIPS 2012 Workshop. — 2012.

- [28] Bishop Christopher M. Pattern Recognition and Machine Learning (Information Science and Statistics). — Secaucus, NJ, USA : Springer-Verlag New York, Inc., 2006.
- [29] Sugiyama Masashi, Krauledat Matthias, Müller Klaus-Robert. Covariate Shift Adaptation by Importance Weighted Cross Validation // J. Mach. Learn. Res. — 2007. — Vol. 8. — P. 985–1005.
- [30] Bland J Martin, Kerry Sally M. Weighted comparison of means // BMJ. — 1998. — Vol. 316, no. 7125. — P. 129.
- [31] Dolgin Elie. Technology: Dressed to detect // Nature. — 2014. — Vol. 511, no. 7508. — P. 16–17.
- [32] Zamora-Martinez Francisco, Almaraz Francisco Javier Munoz, Pardo Juan. Kaggle-epilepsy. — <https://github.com/ESAI-CEU-UCH/kaggle-epilepsy>. — 2014.
- [33] Monteith Kristine, Carroll James L., Seppi Kevin D., Martinez Tony R. Turning Bayesian model averaging into Bayesian model combination. // IJCNN. — IEEE, 2011. — P. 2657–2663.
- [34] Korshunova Iryna, Kopychko Sergii. Detecting epileptic seizures from raw EEG data using convolutional neural networks // System Analysis and Information Technology: 17-th International Conference SAIT-2015. — 2015. — P. x–x.