

Шифр «Предиктор ПТСР»

**Визначення предикторів діагнозу посттравматичного стресового розладу
на основі параметрів моделей зв'язності BOLD-сигналів структур мозку**

АНОТАЦІЯ

Наукова робота за темою «Визначення предикторів діагнозу посттравматичного стресового розладу на основі параметрів моделей зв'язності BOLD-сигналів структур мозку» складається зі вступу; 3 розділів, висновків до кожного з розділів, загальних висновків та списку використаних джерел, який налічує 26 позицій. Загальний обсяг роботи становить 30 сторінок (без урахування списку використаних джерел) і включає 12 рисунків та 4 таблиці.

Актуальність теми

Актуальність дослідження полягає у складності виявлення маркерів посттравматичного стресового розладу (ПТСР) на фізіологічному рівні серед багатьох взаємопов'язаних процесів, які можуть виступати індикаторами порушень. З огляду на часте поєднання у пацієнтів кількох психічних розладів визначення релевантних предикторів діагнозу є необхідною умовою для підвищення ефективності діагностики.

Мета і завдання

Метою роботи є вдосконалення методів діагностики ПТСР на основі аналізу функціональної зв'язності мозкових структур, а також моделювання їх активності з метою виявлення ключових маркерів. Для досягнення поставленої мети необхідно порівняти сучасні методи діагностики, провести аналіз даних функціональної МРТ та побудувати моделі для виявлення предикторів.

Використані методи

Огляд наукових джерел, статистичний аналіз, оцінка функціональної зв'язності, регуляризаційне моделювання, автоматизація обчислень мовою Python, оцінка моделей за агрегованими метриками.

Отримані результати

Відібрані параметри моделей першого порядку, отримані за допомогою програмного продукту, були використані як вектори вхідних ознак для GMDH-LR класифікатора з метою побудови простіших, але не менш продуктивних моделей бінарної класифікації. У результаті запропоновано нову технологію попереднього

діагностування ПТСР, що має практичний потенціал для використання у клінічній діагностиці.

Ключові слова: посттравматичний стресовий розлад, функціональна магнітно-резонансна томографія, машинне навчання, функціональний аналіз, предиктори діагнозу.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ	2
ВСТУП.....	3
РОЗДІЛ 1 ОГЛЯД ІСНУЮЧИХ МЕТОДІВ ВИЯВЛЕННЯ ПРЕДИКТОРІВ ДІАГНОСТУВАННЯ ПТСР.....	4
Висновок до розділу 1	5
РОЗДІЛ 2 ЗАСОБИ РЕАЛІЗАЦІЇ МЕТОДІВ АНАЛІЗУ ЗВ'ЯЗНОСТІ ДЛЯ ВИЗНАЧЕННЯ ПРЕДИКТОРІВ ПТСР.....	6
2.1 Теоретичні основи лінійно-регресійних моделей.....	6
2.2 Методи оптимізації вагових коефіцієнтів	7
2.3 Критерії оцінювання якості регресійних моделей.....	8
2.4 Регуляризаційні методи в обробці значних обсягів даних	10
2.5 Інструменти програмної реалізації.....	12
Висновок до розділу 2	12
РОЗДІЛ 3 ПРОГРАМНА РЕАЛІЗАЦІЯ ТА МЕТОДИКА РОБОТИ ПЗ З АНАЛІЗУ VOLD-СИГНАЛІВ.....	13
3.1 Огляд даних дослідження.....	13
3.2 Попередня обробка даних дослідження.....	14
3.3 Аналіз функціональної зв'язності	15
3.4 Програмна реалізація пошуку предикторів діагнозу	20
3.5 Оцінка результатів роботи програми	25
Висновок до розділу 3	27
ЗАГАЛЬНІ ВИСНОВКИ	28
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	29

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

AUC — площа під кривою ROC, метрика оцінки якості класифікації

BOLD (Blood Oxygen Level Dependent) — сигнал рівня кисню в крові; використовується у фМРТ

F1-міра — метрика, що враховує точність і повноту

FDR (False Discovery Rate) — метод корекції результатів множинних статистичних тестів

GMDH-LR (Group Method of Data Handling – Logistic Regression) — метод моделювання з груповим урахуванням аргументів

GPU — графічний процесор, забезпечує паралельну обробку даних

HC (Healthy Control) — контрольна група

LR — лінійна регресія

MRMR (Minimum Redundancy Maximum Relevance) — метод селекції ознак мінімальної надмірності та максимальної інформативності

NIfTI — формат зберігання нейровізуалізаційних даних

PCL-5 — шкала оцінки симптомів ПТСР

МРТ — магнітно-резонансна томографія

ПТСР (PTSD) — посттравматичний стресовий розлад

ФМРТ — функціональна магнітно-резонансна томографія

ВСТУП

Однією з латентних проблем сучасного суспільства є недооцінювання впливу психічних розладів на життя людини. На відміну від фізичних ушкоджень, функціональні порушення мозкової активності часто залишаються поза увагою. Водночас психічний стан безпосередньо впливає на фізичне здоров'я, що зумовлює необхідність комплексного підходу до його діагностики. Згідно з природою нейронної активності, виявлення характерних змін у функціональній зв'язності мозкових структур є надскладним завданням через високу варіативність і частку хаотичних процесів. Однак завдяки стрімкому розвитку сучасних технологій, стає можливим виявлення інформативних зв'язків, що можна вважати потенційними предикторами психічних порушень. У зв'язку з цим, у роботі пропонується покращена методологія аналізу функціональної зв'язності мозку для виявлення предикторів діагнозу ПТСР на основі параметрів математичних моделей.

Мета і завдання

Метою роботи є вдосконалення методів діагностики ПТСР на основі аналізу функціональної зв'язності мозкових структур, а також моделювання їх активності з метою виявлення ключових маркерів. Для досягнення поставленої мети передбачено:

- 1) порівняти сучасні методи діагностики;
- 2) провести аналіз даних функціональної МРТ;
- 3) побудувати моделі для виявлення ключових предикторів.

РОЗДІЛ 1

ОГЛЯД ІСНУЮЧИХ МЕТОДІВ ВИЯВЛЕННЯ ПРЕДИКТОРІВ ДІАГНОСТУВАННЯ ПТСР

Впродовж життя людина спостерігає за навколишнім світом та активно взаємодіє з ним, на основі чого формується унікальний досвід. Навчання на помилках і численні спроби ведуть до майстерності, формуючи особистість. Проте досвід не завжди сприяє зростанню: травматичний досвід може завдати шкоди, залишаючи глибокий слід у житті людини. Його прояви часто нагадують про себе через знайомі звуки, запахи, місця або слова [21].

Дослідження підтверджують, що травматичний досвід змінює функціональну взаємодію між центрами регуляції страху та емоційного контролю, зокрема базолатеральною мигдалеподібною корою (BLA), орбітофронтальною корою (OFC), гіпокампом і вентромедіальною префронтальною корою (vmPFC). Такі зміни фіксуються у пацієнтів різних вікових груп після пролонгованої експозиційної терапії [10]. Зниження здатності до емоційної обробки провокує соціальні ускладнення у людей із ПТСР.

Магнітно-резонансна томографія (МРТ) дозволяє досліджувати функціональні особливості мозку та широко застосовується для діагностики психічних розладів, зокрема ПТСР. Глибокі нейронні мережі (DNN) надають можливість одночасного аналізу функціональних і структурних змін. У 2018 році в Університеті Дрекслея застосували згорткові нейронні мережі (CNN) для прогнозування біологічного віку мозку пацієнтів із ПТСР, що показало потенційне пришвидшене старіння тканин префронтальної кори та гіпокампу [16]. Оцінювання здійснювалося через коефіцієнти кореляції Пірсона та середньоквадратичну помилку. Результати продемонстрували зменшення обсягу сірої речовини та зміни функціональної активності лімбічної системи, що відповідає за емоційну регуляцію, вказуючи на потенційні предиктори діагнозу.

Водночас успішність використання CNN безпосередньо залежить від інформативної вибірки та значних обчислювальних ресурсів, тому актуальність застосування простіших моделей зростає. Так, у 2019 році дослідження *“Predicting Posttraumatic Stress Disorder Risk: A Machine Learning Approach”* охопило 90 осіб, які пережили травму за місяць до початку, та щоденно нотували симптоми через мобільний додаток Metricwire [25, 15]. Основним критерієм ПТСР був бал ≥ 33 за шкалою PCL- 5 [17].

Для прогнозування ризику порівнювали SVM з різними ядрами та логістичну регресію (LR). Найвища ефективність була у RBF-SVM (AUC = 0.846) із деякою тенденцією до перенавчання (AUC на тесті = 0.818). Лінійна SVM показала кращу здатність до узагальнення (AUC = 0.833). Видалення неінформативних ознак підвищило AUC до 0.882. Метод ефективний без GPU, якщо обсяг даних достатній. LR також продемонструвала близький результат (AUC = 0.799), підтверджуючи продуктивність параметричних моделей при обмеженій вибірці.

Таким чином, поєднання традиційних методів аналізу з вдосконаленою логістичною регресією та багаторівневим підходом дозволяє глибше досліджувати зв'язки між структурами мозку, що відкриває перспективи для персоналізованої діагностики та профілактики ПТСР.

Висновок до розділу 1

Розділ узагальнює сучасні підходи до діагностики ПТСР та порівнює результати багаторічних досліджень щодо виявлення предикторів. Було помічено, що складні моделі, зокрема глибокі нейронні мережі, потребують значних обсягів даних та обчислювальних ресурсів, не забезпечуючи діагностичних критеріїв. Натомість параметричні моделі, завдяки своїй простоті, демонструють стабільний зв'язок між нейрональними показниками та наявністю ПТСР. Що підкреслює необхідність у методології, яка поєднує інтерпретованість та динамічність аналізу, зміщуючи акцент від традиційних підходів до багаторівневого оцінювання функціональної активності мозку пацієнта.

РОЗДІЛ 2

ЗАСОБИ РЕАЛІЗАЦІЇ МЕТОДІВ АНАЛІЗУ ЗВ'ЯЗНОСТІ ДЛЯ ВИЗНАЧЕННЯ ПРЕДИКТОРІВ ПТСР

Часові послідовності відомі своєю унікальністю, адже, на відміну від статичних даних, кожна точка, впорядкована в часі, містить інформацію, яка в загальному представленні формує цілісну картину поведінки спостережуваного явища.

Моделювання часових рядів вимагає детального аналізу закономірностей, що дозволяє дослідити природу взаємозв'язків між послідовними значеннями та обрати відповідний підхід.

У загальному випадку можна виділити два основні типи взаємозв'язку: *лінійний* та *нелінійний*. Лінійні залежності легше піддаються моделюванню, оскільки їхня повторюваність у часі є вищою порівняно з нелінійними [8, 9].

Серед досліджень, розглянутих у попередньому розділі, переважно спостерігалися лінійні зв'язки, що нашо вухе на думку про доцільність застосування лінійної регресії — одного з класичних методів аналізу, що дозволить не лише оцінити взаємодію між структурами головного мозку, а й вивчити динаміку процесів у часі [25].

2.1 Теоретичні основи лінійно-регресійних моделей

Лінійна регресія — це традиційний підхід у статистиці спрямований на моделювання зв'язку між залежною змінною (y) та незалежною змінною (x) [13]. Вона поділяється на два типи:

Однофакторна лінійна регресія описується рівнянням (2.1) і використовується для прогнозування на основі однієї змінної.

$$y = a_0 + a_1x \quad (2.1)$$

Багатофакторна регресія залучає кілька незалежних змінних, дозволяючи вивчати їх як окремий, так і сукупний вплив (2.2).

$$y = a_0 + a_1x_1 + a_2x_2 + \dots + a_nx_n \quad (2.2)$$

Аналіз зв'язків між структурами мозку проводиться попарно, тож однофакторна модель є оптимальною для простої інтерпретації та оцінки впливу однієї змінної без впливу інших.

У нашій роботі змінні представлені як часові послідовності, тому застосовується *ковзна лінійна регресія* — модифікований підхід для аналізу підмножин часового ряду (t) (2.3). Вона фіксує динаміку тенденцій, кореляцій і аномалій, покращує прогнозування. Метод передбачає вибір розміру вікна: менші — чутливі до змін, більші — згладжують шум. У кожному вікні обчислюються параметри регресії a_0 й a_1 , що дозволяє відстежити зміну взаємозв'язків у часі [24]. У результаті, формується набір параметричних значень моделей.

$$y_t = a_0 + a_1x_t. \quad (2.3)$$

2.2 Методи оптимізації вагових коефіцієнтів

Навчання моделі лінійної регресії спрямоване на визначення оптимальних значень вагових коефіцієнтів a_0 та a_1 . Коефіцієнти, або іншими словами параметри, дозволяють отримати рівняння регресійної моделі у такий чин, щоб значення прогнозу були максимально наближені до реальних даних, оскільки від цього залежить її здатність до узагальнення [9].

Оцінка вагових коефіцієнтів (2.4) відбувається на основі мінімізації значення залишкової суми квадратів (RSS). Даний показник можна вважати критерієм, що дозволяє встановити міру відхилення прогнозованих значень від фактичних. Процес мінімізації RSS базується на поступовому коригуванні вагових

коефіцієнтів так, щоб залишки, тобто різниця між реальними та прогнозованими значеннями, були зведені до мінімуму (2.5). Таким чином, параметри, при яких значення RSS досягає мінімального рівня, формують високу якість моделі [7].

$$RSS = (y_1 - a_0 - a_1x_1)^2 + (y_2 - a_0 - a_1x_2)^2 + \dots + (y_n - a_0 - a_1x_n) \quad (2.4)$$

Розглянутий підхід має назву — *метод найменших квадратів* (МНК), який є одним із найбільш відомих та ефективних способів оцінки параметричних значень лінійної регресії, формальне представлення якого можна подати наступним чином [9].

$$J(a_0, a_1) = \min_{a_0, a_1} \left(\frac{1}{n} \sum_{i=1}^n (y_i - (a_0 + a_1x_i))^2 \right) \quad (2.5)$$

2.3 Критерії оцінювання якості регресійних моделей

У результаті навчання отримане представлення функції моделі містить ретельно відібрані значення вагових коефіцієнтів, що дозволяють отримати високі показники точності. Однак реальні результати можна побачити лише безпосередньо на тестовому наборі даних. В обох випадках необхідно належним чином здійснювати оцінювання показників, що демонструє модель. Існує безліч підходів до оцінювання якості апроксимації функції, що описує модель. У випадку однофакторного моделювання для пошуку предикторів діагнозу доцільно застосовувати метрики, як *середня абсолютна помилка* (MAE) та *середньоквадратична помилка* (MSE), які дозволяють оцінити варіативність діагнозу [6].

Легке та інтуїтивне оцінювання забезпечує MAE, що обчислює середню абсолютну відстань між фактичним і прогнозованим значеннями (2.6). Метод менш чутливий до викидів, що призводить до згладжувального ефекту

штрафування, але обмежує здатність виявляти значні відхилення, які можуть впливати на відбір чинників моделі [7].

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \widehat{y}_i| \quad (2.6)$$

За великого обсягу даних ефективний відбір ознак є складним завданням. MAE дає лише загальне уявлення про точність. Натомість MSE, зводячи відхилення до квадрата, максимізує значення помилки, що ефективно підганяє функцію втрат у необхідному напрямку (2.7). Крім того, її похідна є гладкішою, ніж у MAE, яка містить розриви та створює складності в оптимізації [6].

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \widehat{y}_i)^2 \quad (2.7)$$

Утім, в умовах однофакторного регресійного моделювання кожен предиктор розглядається окремо, тож якщо один має велику дисперсію цільової змінної, а інший — малу, MSE відобразить різницю лише через масштаб, а не якість прогнозу. Що ускладнює аналіз, адже MSE є масштабозалежною. Рішенням виступає NMSE — *нормалізоване представлення MSE* (2.8), яке забезпечує коректне попарне порівняння незалежно від масштабу y -змінної. Врахування дисперсії цільової змінної спрощує інтерпретацію результатів, оскільки NMSE лежить в інтервалі $\sigma_n^2 \in [0; +\infty)$, що дозволяє об'єктивно оцінити відносну якість кожного предиктора [26].

$$\sigma_n^2 = \frac{\sum_{i=1}^n (y_i - \widehat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (2.8)$$

2.4 Регуляризаційні методи в обробці значних обсягів даних

У даній роботі моделюються воксельні дані високої розмірності, де виражена *мультиколінеарність* ускладнює виявлення інформативних ознак. Оцінки параметрів стають нестабільними, тому класичне МНК-моделювання втрачає ефективність. Рішенням є *регуляризація* — методика, що додає штрафний член до функції втрат, стабілізує вагові коефіцієнти та запобігає перенавчанню. Тож, розглянемо два основні підходи — Lasso та Ridge-регресії, які дозволяють оптимізувати пошук мінімуму функції, утворюючи відповідну траєкторію градієнтного спуску.

Lasso-регресія (від англ. *Least Absolute Shrinkage and Selection Operator*) пропонує рішення даного питання завдяки штрафного члену $L1$ -норми (2.9) [11]. Штрафне значення базується на сумі модулів коефіцієнтів, помножених на параметр регуляризації λ , який контролює силу обмеження.

$$J(a_0, a_1) = \min_{a_0 a_1} \left(\frac{1}{n} \sum_{i=1}^n (y_i - (a_0 + a_1 x_i))^2 + \lambda |a_1| \right) \quad (2.9)$$

Одна з ключових переваг Lasso-регресії — це здатність до автоматичного витіснення неінформативних ознак. Із зростанням λ частина коефіцієнтів може набувати нульових значень, що призводить до стрімкого спрощення моделі, та особливо актуально для даних з великою кількістю змінних.

Водночас Lasso має певні обмеження. Зокрема, у випадках сильно корельованих предикторів метод зазвичай залишає в моделі лише одну змінну з групи, ігноруючи інші, що потенційно може погіршити точність передбачення [11].

Ridge-регресія або гребінцева регресія передбачає додавання до функції втрат $L2$ -регуляризаційного штрафу (2.10) [11]. Значення штрафного члена являє

собою суму квадратів коефіцієнтів, помножених на параметр регуляризації λ , що дозволяє регулювати ступінь скорочення оцінених параметрів.

$$J(a_0, a_1) = \min_{a_0 a_1} \left(\frac{1}{n} \sum_{i=1}^n (y_i - (a_0 + a_1 x_i))^2 + \lambda a_1^2 \right) \quad (2.10)$$

На відміну від Lasso-регресії, Ridge не здійснює безпосередньо відбір групи типових ознак, оскільки не відбувається повного занулення жодного з коефіцієнтів. Замість цього метод лише зменшує їх значення, знижуючи вплив менш важливих змінних. Таким чином, Ridge-регресія особливо ефективна у ситуаціях, коли спостерігається мультиколінеарність між вхідними змінними, оскільки дозволяє рівномірно розподілити вагу між корельованими ознаками [11].

Нижче наведене зображення, що ілюструє як регуляризація впливає на траєкторію пошуку оптимального розв'язку задачі лінійної регресії (рис. 2.2).

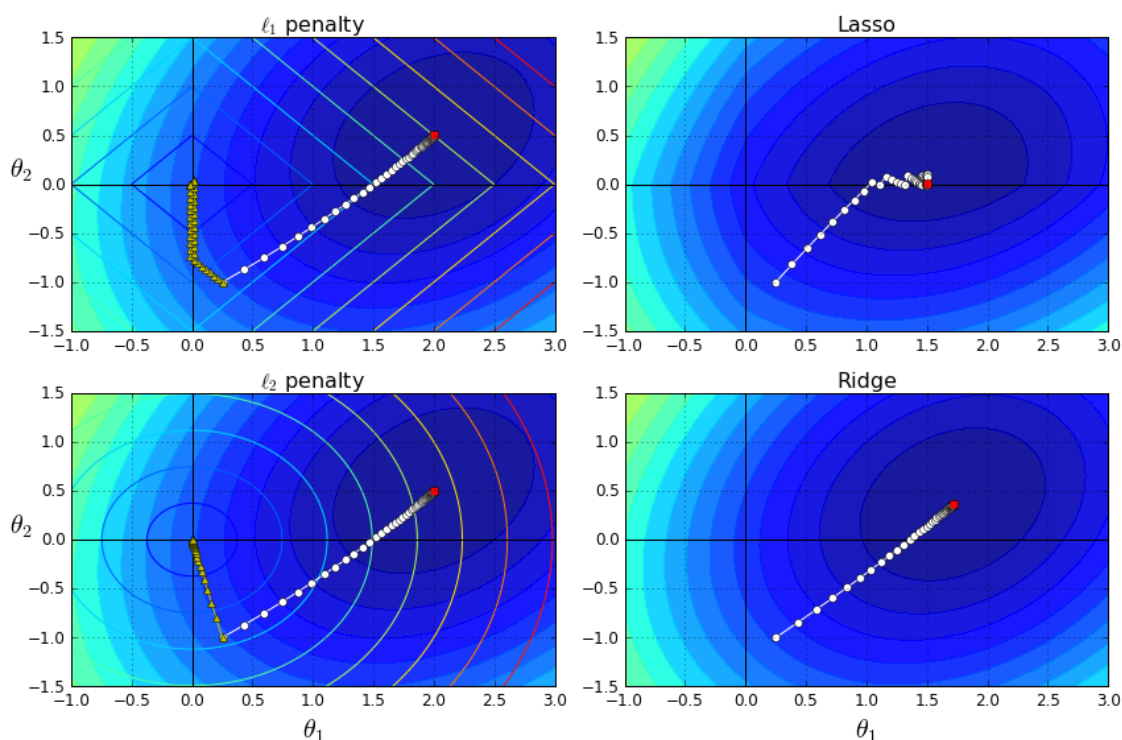


Рисунок 2.2 — Зміна траєкторії градієнтного спуску в залежності від методу регуляризації [12]

У кожному варіанті траєкторія пошуку оптимального рішення, що позначена на рисунку білими точками, змінюється залежно від форми регуляризаційного обмеження, накладеного на вектор початкових коефіцієнтів моделі [9].

Вибір між Lasso та Ridge залежить від структури даних і поставлених завдань, де перший метод — кращий для відбору групи змінних, другий — для підвищення стабільності при високій автокорельованості.

2.5 Інструменти програмної реалізації

Для виконання поставлених завдань використовується високорівнева мова програмування Python, яка завдяки простоті синтаксису, інтерпретації та мінімалістичній структурі коду забезпечує ефективну реалізацію рішень різного рівня складності [23].

Для реалізації задач дослідження використовується: локальне (PyCharm Professional), хмарне (Google Colab) та ізольоване (Anaconda Navigator) середовище розробки [5, 23].

Ключові пакети та бібліотеки, що використовуються для обробки, аналізу та візуалізації даних — NumPy, Pandas, Seaborn, SciPy, Statsmodels, Scikit-learn та Setuptools [4, 19].

Висновок до розділу 2

У розділі проаналізовано методи, що складають основу дослідження функціонального зв'язку та регіональних порушень у структурі мозкової активності. Розглянуто підходи до моделювання на основі нестационарних біологічних часових послідовностей, а також визначені технології програмної реалізації методології: мова програмування, середовища розробки, ключові бібліотеки, пакети для обробки й аналізу даних, тощо.

РОЗДІЛ 3

ПРОГРАМНА РЕАЛІЗАЦІЯ ТА МЕТОДИКА РОБОТИ ПЗ З АНАЛІЗУ BOLD-СИГНАЛІВ

Провідна технологія сьогодення – магнітно-резонансна томографія (МРТ), завдяки якій можливість багаторівневого дослідження людського організму стала реальною [20]. Ще в минулому столітті структурний аналіз був еталоном у галузі прийняття діагностичних рішень, однак точність ставилася під сумнів, потребуючи тим, детальнішого вивчення процесів, що відбуваються в організмі. З часом, завдяки подальшим дослідженням та вдосконаленню можливостей обчислюваних ресурсів з'явився функціональний аналіз. Нарешті, стало можливим оцінювати не лише наслідки прихованих порушень, а й причини їх виникнення, аналізуючи зміни у динаміці. Так на світ з'явилася функціональна магнітно-резонансна томографія (фМРТ).

Оскільки дана робота спрямована на виявлення зон інтересу головного мозку та дослідження зв'язків між ними, доцільно провести аналіз даних, що характеризують зміни рівня кисневого насичення крові протягом фіксованого часу. Дані, отримані за допомогою фМРТ та надані ДУ «Всеукраїнський центр материнства та дитинства НАМН України» у NIfTI форматі - це перший крок на шляху до поставленої мети.

3.1 Огляд даних дослідження

Анонімізовані дані отримано від 24 пацієнтів, з яких 12 мали підозру на посттравматичний стресовий розлад, у інших 12 типових порушень не спостерігалось. Обидві групи проходили сканування у стані спокою протягом 8 хвилин з часовою дискретизацією 3 секунди, що склало 140 часових відліків BOLD-сигналу для кожного вокселю головного мозку. Що дозволило досягнуто загального розуміння просторової організації мозку кожного з пацієнтів.

3.2 Попередня обробка даних дослідження

Серед зібраних даних яскраво прослідковується унікальність анатомічної будови для кожного пацієнта, що обумовлює необхідність стандартизації з метою якісного міжгрупового порівняння. У такому випадку, реалізувати задум можливо за допомогою анатомічного шаблону MNI152 [1].

Як згадувалося раніше, головна мета полягає у виділенні зон інтересу (ROI), які демонструють патологічну поведінку та формують основу для успішної діагностики. З приводу цього, була проведена сегментація воксельної складової згідно британської нейроанатомічної карти на потенційні зони, загальне значення яких склало – 164 [1].

На даному етапі, маніпулювати даними простіше, проте через нестационарну природу BOLD-сигналів легко зіткнутися з мультиколінеарністю у вихідних даних, що ймовірно призведе до помилкових результатів діагностування. Найчастіше це зумовлено паралельною активністю різних фізіологічних систем. Тож, щоб коректно виокремити зони, що містять корисну складову, сегментовані дані пропускаються через низькочастотний фільтр, регресійний аналіз та фінальне згладжування (рис. 3.1) [1].

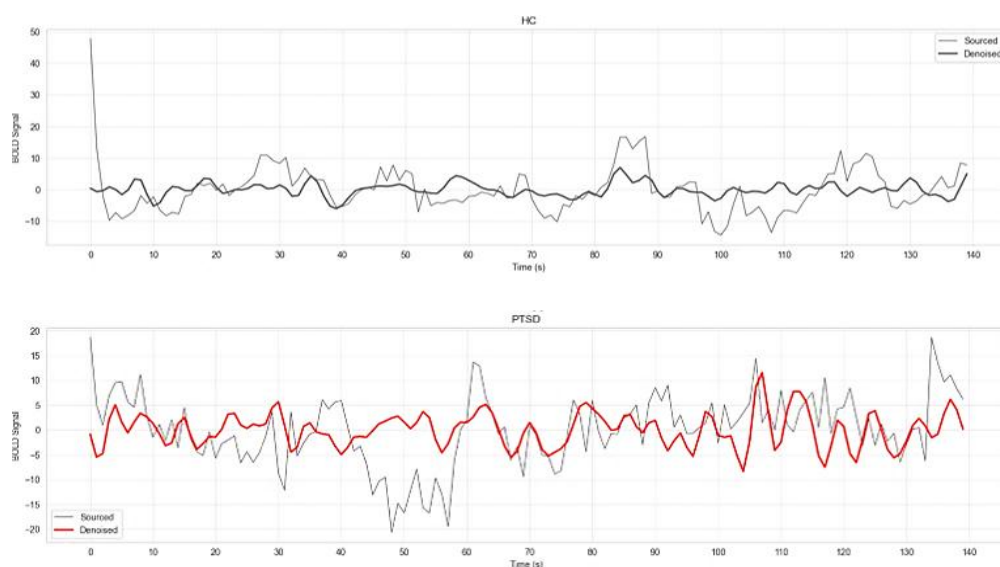


Рисунок 3.1 – BOLD-активність пари структур головного мозку до та після обробки на прикладі представників обох груп

3.3 Аналіз функціональної зв'язності

У результаті попередньої обробки були отримані матриці значень часових рядів, що відображають тенденції зміни рівня кисневого насичення крові в узагальнених 164 структурах головного мозку впродовж 140 часових відліків. Хоча розмірність даних була максимально зменшена, подавати отриманий обсяг даних (140 x 164 x 24) на опрацювання класифікатору з метою виділення маркерів діагнозу марно, оскільки якість такої системи не буде задовільною. У зв'язку з цим, варто продовжити аналіз даних, зосередившись на функціональній зв'язності між ними. Представлені результати аналізу базуватимуться на порівнянні двох груп, де HC – контрольна група, PTSD – група з посттравматичним розладом.

Як відомо, BOLD-сигнали є чудовими індикаторами нейроактивності, тому, розрахувавши кореляцію за Пірсоном можна побачити силует мозкової активності у фіксований момент часу (рис. 3.2). Таким чином, загальна картина спостережуваного явища підтверджує, що сумніви з приводу гіпер- або гіпоактивності внаслідок порушень функціонування окремих ділянок мозку не безпідставні. Взаємодія між ними є патологічною, хоча і різного ступеню прояву.

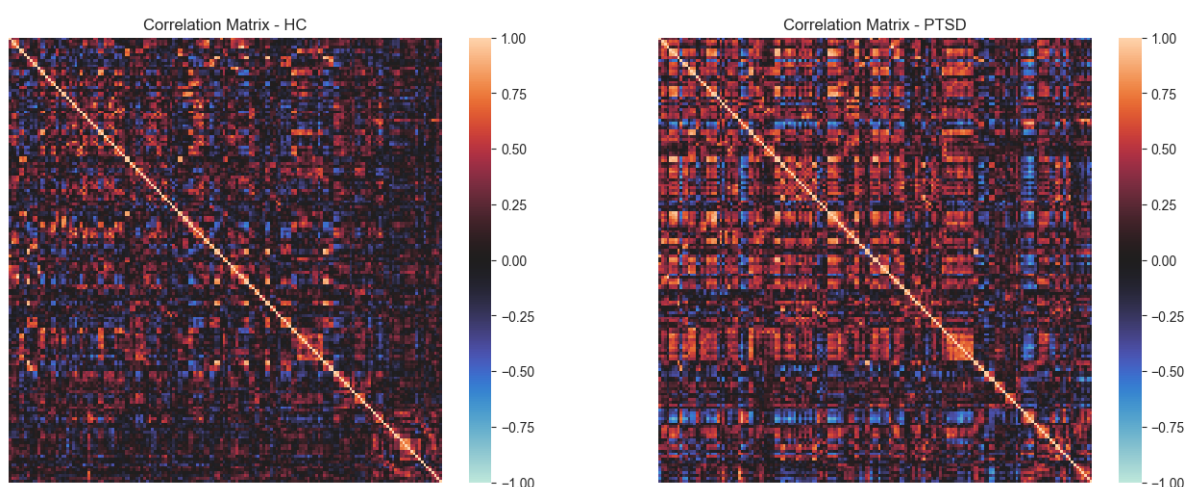


Рисунок 3.2 – Коефіцієнти кореляції між структурами головного мозку

Хоча дане припущення здається очевидним з боку візуального сприйняття, воно все ж таки потребує статистичного обґрунтування. У такому разі можна спробувати провести множинне тестування, яке дозволить прийняти або відкинути гіпотезу щодо наявності статистично значущої різниці у функціонуванні пар структур головного мозку між здоровою та хворою групами.

Перед початком тестування, необхідно врахувати особливості розподілу коефіцієнтів поміж груп та обрати відповідний статистичний тест, який найкраще підійде до даних.

Моделюючи розподіл даних, можна стверджувати, що функції, які їх описують, не близькі до нормального розподілу. Однак їх можливо скоригувати, максимально наблизивши, завдяки перетворенню Фішера (рис. 3.3), унаслідок чого точність тестування має покращитися, що ми і перевіримо згодом [1].

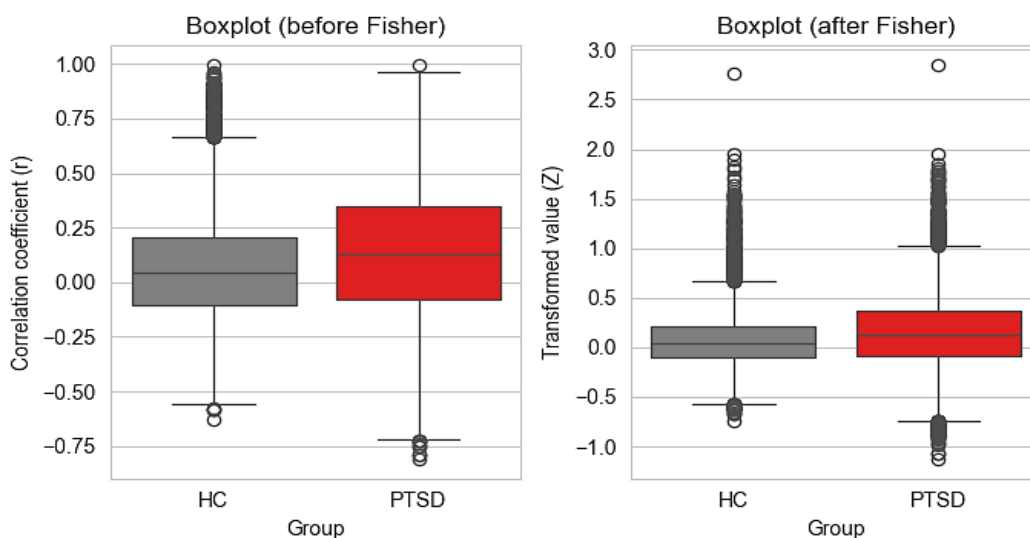


Рисунок 3.3 – Розподіл до та після набуття нормальності

T-тест незалежного тестування було обрано після повторної перевірки даних на нормальність на основі результатів іншого відомого тесту Шапіро-Вілко, де р-значення оцінки значущості в середньому було набагато меншим за встановлене. З метою уникнення помилки першого роду, додатково проводиться корекція, відома як FDR-корекція, яка контролює ймовірність її появи під час множинного тестування [2]. Найбільш стійкі результати представлені нижче (табл. 3.1).

Таблиця 3.1

Тестування гіпотези

№	Region1	Region2	p-value	tstat
0	DorsalAttention.FEF (L)	LG l	0.000009	-6.115097
1	DorsalAttention.FEF (L)	TOFusC l	0.000133	-4.833656
2	SMA L (...)	Cerebl l	0.000165	4.734979
3	DorsalAttention.FEF (L)	TOFusC r	0.000238	-4.568444
4	SensoriMotor.Lateral (L)	Cerebl l	0.000249	4.548821

З першого погляду результати тестування виглядають вагомими, про що свідчать малі р-значення, отримані вже після корекції множинних порівнянь. Іншими словами, різниця з боку функціонування ділянок мозку, що відповідають за перцептивну обробку (*LR l*), когнітивний контроль (*TOFusC l* й *TOFusC r*) та сенсорно-моторну координацію (*Cerebl l*) між групами є статистично значущою, що узгоджується з нейрофізіологічними змінами у мозку людини, яка страждає на ПТСР.

Водночас загальний відсоток зв'язків, серед яких у ході тестування випадково значущими може виявитися половина, та складність встановлення порогового значення прийняття гіпотези про відмінність неминуче ведуть до невизначеності даного підходу.

Окрім цього, властивість психофізіологічних процесів полягає в тому, що вони є динамічними, а отже взаємодія між областями мозку не може бути лише синхронною, як демонструє кореляційний аналіз у конкретний момент часу. Таким чином, аналіз, що враховує лише статичні взаємодії, не дозволить якісно визначити ключові предиктори, необхідні для досягнення високої точності діагностування.

У зв'язку з цим, доцільно дослідити функціональну зв'язність у часі, що дозволить встановити характер регіональної взаємодії у динаміці – зокрема, визначити, які зв'язки призводять до збудження або гальмування.

Таким чином, розглянемо механізм нейрональної взаємодії на прикладі пари структур, яка внаслідок тестування проявила себе найстійкіше, заклавши ідею

про те, що гіперактивність *LG I* може бути пов'язана з підвищеною активністю *DorsalAttention.FEF (L)*.

Нижче представлена синхронізація сигналів між зазначеної парою структур в момент максимальної взаємодії (рис. 3.4).

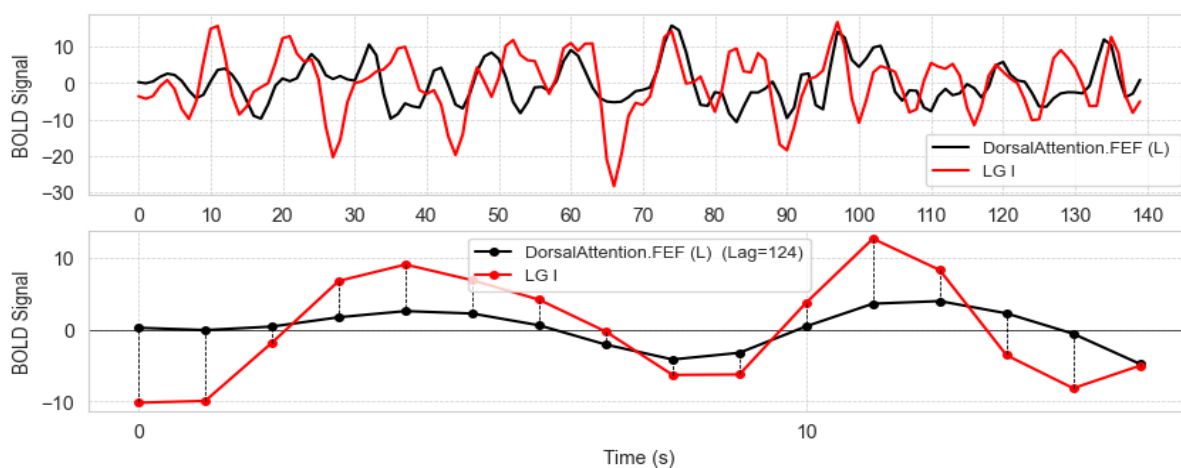


Рисунок 3.4 – Синхронізація активності BOLD-сигналів

Таким чином, завдяки наявності певної патерності у поведінці сигналів з'являється можливість прогнозування функціонального зв'язку між ними, узагальнюючи їхні властивості у просторі часових рядів. Тож звернувшись до вже відомого, неодноразово перевіреного на практиці методу класичного статистичного аналізу – однофакторної лінійної регресії, спробуємо змодельовати декілька моделей для п'яти різних значень часової затримки.

Також, з метою вивчення рівня відхилення моделі від реальних значень, побудуємо стовпчасту діаграму значень нормалізованої середньоквадратичної помилки, що надасть розуміння з приводу точності отриманих моделей та сформує попередні висновки з боку якості процесу апроксимації функції.

У результаті вищезгаданих маніпуляцій отримуємо нижче представлені моделі та значення помилки (рис. 3.5).

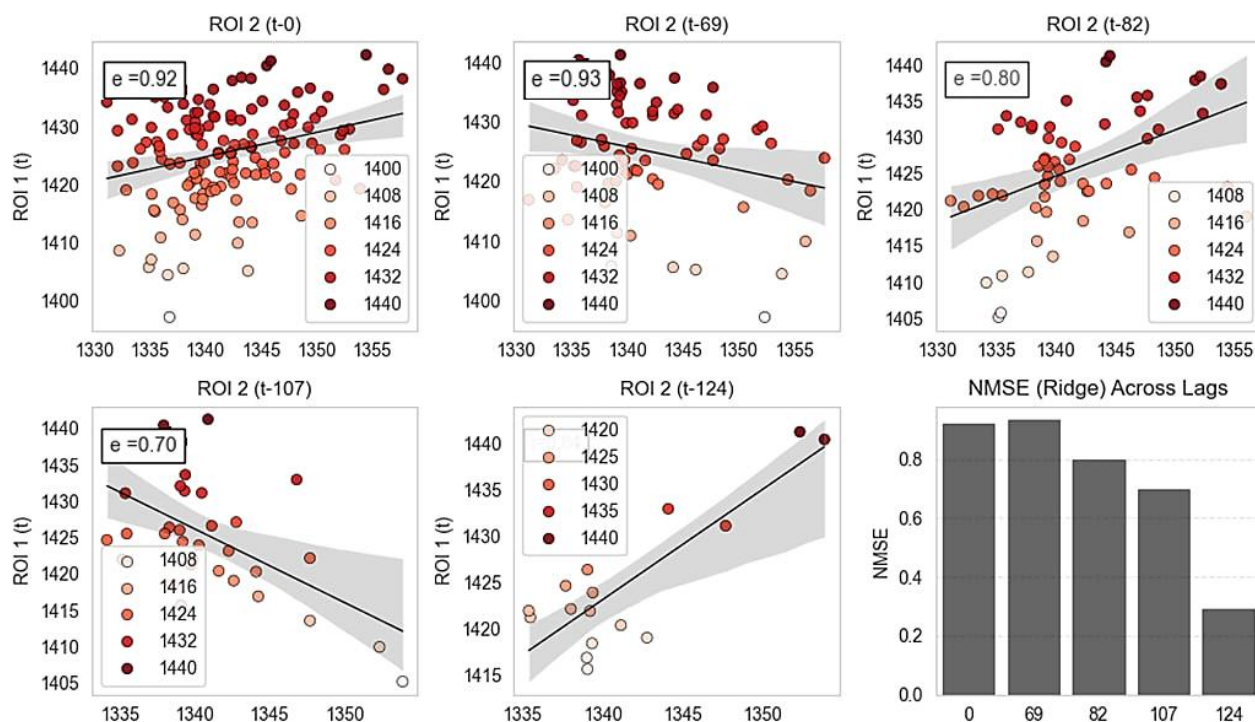


Рисунок 3.5 – Моделі зв'язності першого порядку

Детальна характеристика найкращої моделі з часовою затримкою рівною 124 наведена нижче (рис. 3.6).

OLS Regression Results						
Dep. Variable:	y	R-squared:	0.708			
Model:	OLS	Adj. R-squared:	0.687			
Method:	Least Squares	F-statistic:	33.89			
Date:	Mon, 17 Feb 2025	Prob (F-statistic):	4.43e-05			
Time:	18:19:26	Log-Likelihood:	-45.114			
No. Observations:	16	AIC:	94.23			
Df Residuals:	14	BIC:	95.77			
Df Model:	1					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
const	-169.9457	273.941	-0.620	0.545	-757.491	417.600
x1	1.1889	0.204	5.822	0.000	0.751	1.627
Omnibus:	2.918	Durbin-Watson:	0.550			
Prob(Omnibus):	0.232	Jarque-Bera (JB):	1.518			
Skew:	-0.451	Prob(JB):	0.468			
Kurtosis:	1.789	Cond. No.	3.39e+05			

Рисунок 3.6 – Статистичні показники OLS-регресії

Результуюче значення моделювання, помилки NMSE, що склало 0.29, демонструє наявність помірної, але також і суттєвої залежності між *DorsalAttention.FEF (L)* та *LG l*. Що каже про відносно невелику середню похибку передбачення моделі та прийнятно для нейрофізіологічних даних із підвищеним рівнем вмісту артефактів.

Значення $\beta = 1.1889$ ($p < 0.001$) стверджує, що гіперактивність *DorsalAttention.FEF (L)* безпосередньо впливає на *LG l*, підсилюючи її реакцію.

T-статистика для коефіцієнта $x1$, рівна 5.822, вказує на внесок значного вкладу до моделі, а довірчий інтервал для $\beta \in [0.751, 1.627]$ підтверджує стабільність результату. У свою чергу, низьке значення Durbin-Watson ($DW = 0.550$) натякає на можливі залишкові ефекти, що автокорелюють, до яких може належати, наприклад, вплив лімбічної системи [22].

Зважаючи на все вище сказане, можна вважати, що розглянуті зони інтересу є фізіологічними маркерами порушення функціонування окремих ділянок мозку при ПТСР.

У зв'язку з цим, найкращим рішенням є автоматизація аналізу даних активності BOLD-сигналів, що пришвидшить пошук предикторів діагнозу та сформулюють базу для бінарного класифікатора.

3.4 Програмна реалізація пошуку предикторів діагнозу

Під час проведення аналізу функціональної складової даних було отримано змістовні результати. Використана методологія дійсно виявляє інформативні закономірності в даних, запобігаючи генеруванню випадкових передбачень. У цьому контексті вислів “*The power of the machine learning method is not based on chance*” вдало описує дане спостереження [3]. З огляду на все вище сказане, можна спробувати узагальнити ефект, прискоривши процес дослідження шляхом програмної реалізації розглянутих етапів аналізу.

Обмежений обсяг вибірки ускладнює процес аналізу, оскільки будь-яка модель є вимогливою до якості вхідних даних. У таких випадках, як правило, або генерують зразки синтетично, або тимчасово призупиняють дослідження. Дійсно, на відміну від навчання з вчителем, методи навчання без вчителя часто розглядають малу вибірку як неінформативну. Тому, щоб отримати максимум з наявних даних, перед навчанням моделей доцільно розділити їх таким чином, щоб кожна група містила приблизно однаковий відсоток інформативної складової. Існує багато методів раціонального розбиття даних, проте безпосередньо метод розбиття за критерієм *Махаланобіса* вважається одним з найбільш оптимальних рішень [9].

У результаті такого розподілу утворюються 3 підвибірки: *навчальна*, *валідаційна* та *тестова*.

Наступний етап передбачає визначення структури моделі першого порядку, яка буде в змозі отримати змістовне представлення щодо патернів відмінного функціонування мозкових структур групи з ПСТР.

З першого погляду, метод лінійного моделювання з бібліотеки Scikit-learn — *LinearRegression()* — найліпше підходить до здійснення задуманого. Утім, він не враховує таку наявну особливість даних, як мультиколінеарність.

Натомість методи *Lasso()* та *Ridge()* вирішують дану проблему, випадково зануляючи параметри під час перебору. Тому, порівнюючи результати тестування обох методів, було зроблено вибір у бік гребінцевої, тобто *Ridge*, регресії, яка менш радикально усуває потенційні параметри моделей порівняно з *Lasso* [11].

Не менш важливою є оцінка вихідних моделей. Найбільш поширеними метриками оцінки якості вважається середньоквадратична помилка (MSE), середня абсолютна помилка (MAE), а також похідні. У нашому випадку, було обрано NMSE, *нормалізовану середньоквадратичну помилку*, оскільки вона з легкістю масштабує значення помилки відносно дисперсії даних, що робить її незамінною для міжгрупового порівняння [26].

Далі розглянемо особливості реалізації методології на програмному рівні.

Активність структур мозку моделюємо за допомогою лінійного моделювання з L2-регуляризацією, що дозволяє визначити оптимальні параметри для кожної пари структур у момент пікової взаємодії. Даний підхід забезпечує індивідуалізацію результатів для кожного пацієнта та їхнє впорядковане представлення.

Моделювання узагальнює результати, виділяючи релевантні параметри та формуючи тенденції, характерні для всієї вибірки. Після завершення аналізу дані зберігаються у стандартизованому вигляді, що дозволяє їх повторне використання без втрати інформації. Отримані масиви параметрів містять значення відібраних моделей, тоді як альтернативні моделі, які не несли корисного навантаження, позначені як *NaN*.

Часткове представлення вихідних даних, що описують взаємозв'язок регіонів активності головного мозку пацієнта з ПТСР, включає: значення кореляційних коефіцієнтів, які пояснюють силу взаємодії між парами регіонів (рис. 3.7); значення порогів активації, що визначають мінімальний рівень, необхідний для активації регіонів (рис. 3.8); а також значення нормалізованої середньоквадратичної помилки, які характеризують здатність моделей до навчання (рис. 3.9).

```
pd.DataFrame(A1[0]).head() #a1
```

	0	1	2	3	4	5	6	7	8	9	...	154	155	156	157
0	NaN	NaN	-0.263683	NaN	NaN	-0.144676	-0.236114	-0.315989	NaN	0.224963	...	NaN	0.533985	NaN	0.777564
1	-0.286081	NaN	NaN	NaN	NaN	-0.147893	0.168456	0.383839	NaN	0.315969	...	NaN	NaN	NaN	0.939726
2	NaN	NaN	NaN	NaN	NaN	-0.135650	-0.245140	NaN	NaN	NaN	...	NaN	NaN	NaN	-0.835566
3	-0.313791	0.209900	NaN	NaN	NaN	-0.096851	-0.154309	NaN	NaN	0.377415	...	NaN	0.281306	NaN	0.713106
4	-0.239017	0.184857	-0.142561	NaN	NaN	NaN	NaN	NaN	NaN	-0.245299	...	NaN	NaN	NaN	1.035474

Рисунок 3.7 – Матриця значень кореляційних коефіцієнтів

```
pd.DataFrame(A0[0]).head() #a0
```

	0	1	2	3	4	5	6	7	8	9 ...	154	155	
0	NaN	NaN	807.391890	NaN	NaN	776.921759	972.372700	998.488723	NaN	591.617664	...	NaN	358.423364
1	777.899231	NaN	NaN	NaN	NaN	778.187798	727.914994	576.000198	NaN	537.815589	...	NaN	NaN
2	NaN	NaN	NaN	NaN	NaN	777.061480	988.041973	NaN	NaN	NaN	...	NaN	NaN
3	871.126312	424.069051	NaN	NaN	NaN	771.099563	959.810843	NaN	NaN	408.802208	...	NaN	444.072580
4	793.746732	456.172692	759.820579	NaN	NaN	NaN	NaN	NaN	NaN	920.914994	...	NaN	NaN

Рисунок 3.8 – Матриця значень порогів активації

```
pd.DataFrame(SE[0]).head() #se
```

	0	1	2	3	4	5	6	7	8	9 ...	154	155	156	157	158	
0	NaN	NaN	0.885257	NaN	NaN	0.959007	0.860313	0.930667	NaN	0.930029	...	NaN	0.932548	NaN	0.946163	NaN
1	0.934529	NaN	NaN	NaN	NaN	0.962657	0.942169	0.915746	NaN	0.877396	...	NaN	NaN	NaN	0.932976	NaN
2	NaN	NaN	NaN	NaN	NaN	0.975730	0.901477	NaN	NaN	NaN	...	NaN	NaN	NaN	0.957126	NaN
3	0.909903	0.955420	NaN	NaN	NaN	0.982774	0.938804	NaN	NaN	0.803718	...	NaN	0.981339	NaN	0.954320	NaN
4	0.973727	0.982715	0.985606	NaN	NaN	NaN	NaN	NaN	NaN	0.954411	...	NaN	NaN	NaN	0.955469	0.929112

Рисунок 3.9 – Матриця значень нормалізованої середньоквадратичної помилки

З одного боку, можна було б зупинитися на досягнутому та передати на вхід до класифікатора отримані вектори значень A_0 та A_1 . З іншого боку, можливо покращити ефект аналізу завдяки комбінаційному підходу. Так завдяки поєднанню основної методології з *селекційним відбором ознак* (mRMR) можливо значно зменшити загальне число вихідних предикторів та залишити тільки ті, які включатимуть характерні ознаки відмінної роботи мозку групи з ПСТР порівняно зі здоровою групою. Детальніше про методологію розписано у науковій статті “*Class-Oriented Features Selection Technology in Medical Images Classification Problem on the Example of Distinguishing Between Tuberculosis Sensitive and Resistant Forms*” [14].

Застосувавши далі селекційний аналіз до контейнерів значень A_0 та A_1 , а також до попередньо розрахованих та нормалізованих за Фішером кореляційних матриць S , отримуємо 30 ознак, які вже справді можна вважати предикторами діагнозу [1]. Результат селекційного аналізу виглядає наступним чином (рис. 3.10).

#	Column	Non-Null Count	Dtype	#	Column	Non-Null Count	Dtype
0	A0-6608	24 non-null	float64	16	A1-22424	24 non-null	float64
1	A0-14983	24 non-null	float64	17	A1-17895	24 non-null	float64
2	A0-16863	24 non-null	float64	18	X-17306	24 non-null	float64
3	A0-23857	24 non-null	float64	19	A0-16373	24 non-null	float64
4	X-7644	24 non-null	float64	20	A1-7679	24 non-null	float64
5	A0-24408	24 non-null	float64	21	A1-15074	24 non-null	float64
6	A0-2652	24 non-null	float64	22	A1-5624	24 non-null	float64
7	A0-10855	24 non-null	float64	23	X-14996	24 non-null	float64
8	A0-17382	24 non-null	float64	24	X-11846	24 non-null	float64
9	A0-23625	24 non-null	float64	25	X-5218	24 non-null	float64
10	X-14935	24 non-null	float64	26	X-22007	24 non-null	float64
11	A0-24122	24 non-null	float64	27	X-5878	24 non-null	float64
12	A0-10825	24 non-null	float64	28	X-9388	24 non-null	float64
13	A0-11852	24 non-null	float64	29	X-13422	24 non-null	float64
14	A0-24347	24 non-null	float64	30	Y	24 non-null	int8
15	A0-23342	24 non-null	float64	dtypes: float64(30), int8(1)			

Рисунок 3.10 – Перелік найінформативніших ознак

На базі отриманих векторів значень з’являється можливість формування фінального *класифікатора*, що утворюється завдяки комбінації бінарних логістичних регресій за принципом *методу групового урахування аргументів* (МГУА). Іншими словами, підхід передбачає розбиття задачі на набір бінарних класифікаторів, де кожна модель навчається за принципом «один клас проти всіх», а потім їх результати комбінуються для визначення остаточного класу. Що у свою чергу, дозволяє зменшити обсяг параметрів, усуваючи неінформативні ознаки, та підвищити точність класифікації без втрати узагальнюючої здатності моделі. Опис підходу у деталях можна знайти у публікації “*Multiclass Classifier Based on Binary Logistic Regressions Obtained According to the Principles of GMDH*” [18].

Результуючі значення коефіцієнтів, отримані для двох вихідних моделей побудованих на основі найбільш релевантних ознак представлені нижче (табл. 3.2).

Таблиця 3.2

Параметри моделей бінарного класифікатора

parameters	values
x5, x1	0.0059, -316.05
x12, x5, x1	0.089, 0.00013, -86.28

3.5 Оцінка результатів роботи програми

У попередньому розділі розглянута методологія була успішно програмно реалізована та поєднана з іншими підходами для виявлення предикторів діагнозу, на основі яких отримано моделі бінарної класифікації (3.1) та (3.2). Нижче графічно представлені матриці демонструють відповідність діагностичних рішень еталонним значенням (рис. 3.11 – рис. 3.12).

$$Y = \frac{1}{1 + e^{-z}}, \quad (3.1)$$

$$z = 0,0059 x_1 \cdot x_5 - 316,05,$$

де

x_1 — параметр моделі a_0 , вагове значення якого описує силу функціональної взаємодії між ділянками “*lateral part of visual network right*” та “*caudate right*”;

x_5 — параметр моделі a_0 , вагове значення якого описує силу функціональної взаємодії між ділянками “*frontal pole (left hemisphere)*” та “*central opercular cortex (right hemisphere)*”.

$$Y = \frac{1}{1 + e^{-z}}, \quad (3.2)$$

$$z = 0,089 \cdot x_5 \cdot x_{12} \cdot 0,000013 \cdot x_1 \cdot x_5 - 86,28,$$

де

x_{12} — коефіцієнт кореляції a_1 , що описує силу функціональної взаємодії між ділянками “*inferior frontal gyrus, pars triangularis (left hemisphere)*” та “*vermis*” (частина мозочка).

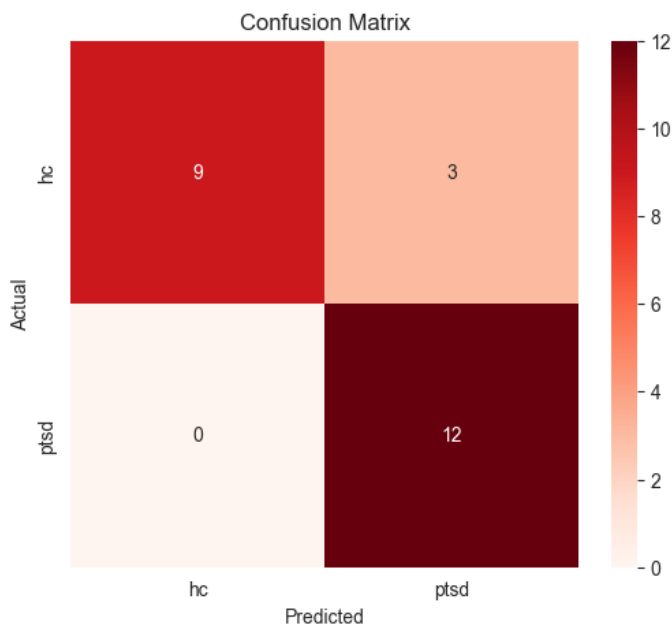


Рисунок 3.11 – Матриця невідповідностей для моделі 3.1 на повній вибірці

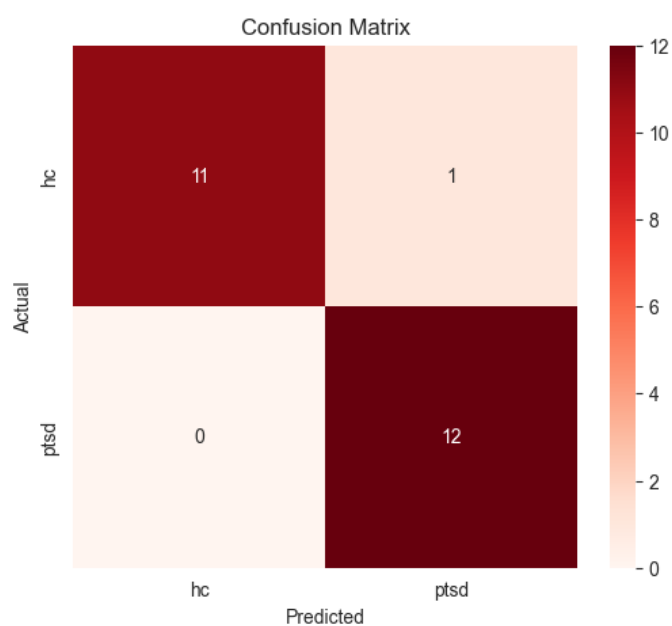


Рисунок 3.12 – Матриця невідповідностей для моделі 3.2 на повній вибірці

Для поглибленого аналізу здатності моделей до навчання доцільно розрахувати метрики *f1-міру*, *точність*, *чутливість* та *специфічність*. Для цього використовуються результати розбиття вибірки на навчальний, валідаційний і тестовий набори. Додатково проводиться оцінка на повному наборі даних, що дозволяє визначити здатність моделей до узагальнення (табл. 3.3 – табл. 3.4).

Таблиця 3.3

Показники моделі 3.1

	F1-міра	Точність	Чутливість	Специфічність
Навчальна	1,0	1,0	1,0	1,0
Валідаційна	0,87	0,83	1,0	0,67
Тестова	0,8	0,75	1,0	0,54
Повна	0,92	0,92	1,0	0,83

Таблиця 3.4

Показники моделі 3.2

	F1-міра	Точність	Чутливість	Специфічність
Навчальна	1,0	1,0	1,0	1,0
Валідаційна	0,86	0,83	1,0	0,67
Тестова	1,0	1,0	0,98	1,0
Повна	0,96	0,96	1,0	0,93

Кожна з моделей продемонструвала гарні результати за всіма показниками. Водночас модель (3.2) на валідаційному та тестовому наборах має дещо вищі значення специфічності та f1-міри порівняно з моделлю (3.1), що свідчить про зниження ймовірності хибнопозитивної класифікації та краще розпізнання патерності серед часових рядів. Показник чутливості для обох моделей залишається високим, що є необхідною умовою успішного виявлення позитивних випадків з найінформативнішими ознаками.

Висновок до розділу 3

У даному розділі була розглянута програмна реалізація рішення, побудованого на засадах методології, розглянутої у попередньому розділі. Використання анатомічного шаблону MNI152, алгоритмів машинного навчання, таких як Ridge, вдосконаленої логістичної регресії за МГУА та інших подібних технологій дозволило успішно автоматизувати обробку даних BOLD-сигналів і створити моделі для бінарної класифікації діагнозу ПСТР.

ЗАГАЛЬНІ ВИСНОВКИ

У результаті дослідження була отримана унікальна методологія, покладена в основу програмного рішення відбору функціональних маркерів-предикторів діагнозу ПТСР. Особливість удосконаленого алгоритму полягає в багаторівневому аналізі даних фМРТ, що зосереджується на дослідженні та моделюванні функціональної зв'язності, завдяки чому стає можливим отримання векторів найбільш інформативних параметрів-ознак, які не лише дозволяють виявити ключові зв'язки між структурами мозку, а й зрозуміти, як саме відбувається перебіг патологічних процесів у часовій динаміці.

Розроблені моделі продемонстрували високу ефективність у виявленні предикторів ПТСР, про що кажуть кількісно високі показники класифікації фінальної моделі (f1-міра: 0,8–1,00; точність: 0,75–1,00; чутливість: 0,98–1,00; специфічність: 0,54–1,00 на тестовій вибірці). Регуляризаційні методи забезпечили достатню стабілізацію моделей з метою уникнення перенавчання — що є частою проблемою при роботі з багаторозмірними даними через підвищений ризик наявної мультиколінеарності.

Таким чином, програмне рішення, створене в рамках роботи, автоматизує процес аналізу даних, містить інструменти для попередньої обробки даних, моделювання та візуалізації результатів.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. ...

...

...

...

...

2. Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 57(1), 289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>

3. Chen, Y., Zhu, G., Liu, D., Liu, Y., Zhang, X., Du, T., & Zhang, J. (2022). Seed-Based Connectivity Prediction of Initial Outcome of Subthalamic Nuclei Deep Brain Stimulation. *Neurotherapeutics*, 19(2), 608–615. <https://doi.org/10.1007/s13311-022-01208-9>

4. Comparative Analysis of Data Visualization Libraries Matplotlib and Seaborn in Python. (2021). *International Journal of Advanced Trends in Computer Science and Engineering*, 10(1), 277–281. <https://doi.org/10.30534/ijatcse/2021/391012021>

5. Drymonitis, A. (2023). Writing Your First Python Programs. *The Python Audio Cookbook*, 11–42. <https://doi.org/10.4324/9781003386964-2>

6. Draper, N. R., & Smith, H. (1998). *Applied Regression Analysis*. Wiley Series in Probability and Statistics. <https://doi.org/10.1002/9781118625590>

7. Everything You Need to Know About Linear Regression [Електронний ресурс]. – Режим доступу до ресурсу: <https://medium.com/analytics-vidhya/everything-you-need-to-know-about-linear-regression-750a69a0ea50>

8. Forecasting Based on Time Series by States. (2018). *Advanced Time Series Data Analysis*, 447–491. Portico. <https://doi.org/10.1002/9781119504818.ch7>

9. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning*. In Springer Series in Statistics. Springer New York.

<https://doi.org/10.1007/978-0-387-84858-7>

10. Hinojosa, C. A., George, G. C., & Ben-Zion, Z. (2024). Neuroimaging of posttraumatic stress disorder in adults and youth: progress over the last decade on three leading questions of the field. *Molecular Psychiatry*, 29(10), 3223–3244.

<https://doi.org/10.1038/s41380-024-02558-w>

11. Lasso and Ridge Regression — Regularization Techniques Explained for Beginners [Электронный ресурс]. – Режим доступа до ресурсу:

<https://medium.com/@seshu8hachi/lasso-and-ridge-regression-regularization-techniques-explained-for-beginners-be568ac6c0a3>

12. Lasso or Ridge for correlated variables [Электронный ресурс]. – Режим доступа до ресурсу:

<https://stackoverflow.com/questions/42904211/lasso-or-ridge-for-correlated-variables>

13. Linear Regression in Machine learning [Электронный ресурс] – Режим доступа до ресурсу: <https://www.geeksforgeeks.org/ml-linear-regression/>

14. ...

...

...

... 2021 IEEE ... International Conference on ...

...

...

15. Metricwire Inc. [Электронный ресурс]. – Режим доступа до ресурсу:

<https://metricwire.com/>

16. Niu, X., Liang, H., & Zhang, F. (2018). Brain age prediction for post-traumatic stress disorder patients with convolutional neural networks: a multi-modal neuroimaging study. 2018 Conference on Cognitive Computational Neuroscience.

[https://www.semanticscholar.org/paper/Brain-age-prediction-for-post-traumatic-stress-with-Niu-](https://www.semanticscholar.org/paper/Brain-age-prediction-for-post-traumatic-stress-with-Niu-Liang/efb078e37107d0364e53913614ee654b7afac1e9?utm_source=direct_link)

[Liang/efb078e37107d0364e53913614ee654b7afac1e9?utm_source=direct_link](https://www.semanticscholar.org/paper/Brain-age-prediction-for-post-traumatic-stress-with-Niu-Liang/efb078e37107d0364e53913614ee654b7afac1e9?utm_source=direct_link)

17. PTSD Checklist for DSM-5 (PCL-5) - National Center for PTSD [Электронный

ресурс]. – Режим доступа до ресурсу:

<https://www.ptsd.va.gov/professional/assessment/adult-sr/ptsd-checklist.asp>

18. ...(2023).

...

...

19. Rayhan, Rajan & Rayhan, Abu & Kinzler, Robert. (2023). Exploring the Power of Data Manipulation and Analysis: A Comprehensive Study of NumPy, SciPy, and Pandas. <http://dx.doi.org/10.13140/RG.2.2.22390.16968>

20. Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E. J., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., & Matthews, P. M. (2004). Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage*, 23, S208–S219. <https://doi.org/10.1016/j.neuroimage.2004.07.051>

21. The Body Keeps the Score: Brain, Mind and Body in the Healing of Trauma, by Bessel A. van der Kolk, MD. (2015). *Journal of Sandplay Therapy*, 24(1). <https://doi.org/10.61711/jst.2015.24.1s.060>

22. The Durbin-Watson Test: Definition & Example [Электронный ресурс]. – Режим доступа до ресурсу: <https://www.statology.org/durbin-watson-test/>

23. The Top Programming Languages 2024 [Электронный ресурс]. – Режим доступа до ресурсу: <https://spectrum.ieee.org/top-programming-languages-2024>

24. Understanding Time Series Rolling Linear Regression [Электронный ресурс]. – Режим доступа до ресурсу: <https://medium.com/@aleksej.gudkov/understanding-time-series-rolling-linear-regression-404e171d31e5>

25. Wshah, S., Skalka, C., & Price, M. (2019). Predicting Posttraumatic Stress Disorder Risk: A Machine Learning Approach. *JMIR Mental Health*, 6(7), e13946. <https://doi.org/10.2196/13946>

26. ...

...
...
...
...
...