

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ  
«КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ  
ІМЕНІ ІГОРЯ СІКОРСЬКОГО»

Кваліфікаційна наукова  
праця на правах рукопису

**ОЛЕКСІЙ АРТУР ОЛЕГОВИЧ**

УДК 004.96

**ДИСЕРТАЦІЯ**

**МЕТОДИ ТА ПРОГРАМНІ ЗАСОБИ АНАЛІЗУ АКУСТИЧНИХ  
СИГНАЛІВ НА ОСНОВІ НЕЙРОМЕРЕЖЕВИХ МОДЕЛЕЙ**

121 – Інженерія програмного забезпечення

12 – Інформаційні технології

Подається на здобуття наукового ступеня  
доктора філософії

Дисертація містить результати власних досліджень. Використання ідей,  
результатів і текстів інших авторів мають посилання на відповідне джерело

\_\_\_\_\_ А.О. Олексій

Наукові керівники:

ВЕРЛАНЬ Андрій Анатолійович,  
доктор технічних наук, професор  
ВАРАВА Іван Андрійович,  
кандидат технічних наук

## АНОТАЦІЯ

**Олексій А. О.** Методи та програмні засоби аналізу акустичних сигналів на основі нейромережових моделей – Кваліфікаційна наукова праця на правах рукопису.

Дисертація на здобуття наукового ступеня доктора філософії з галузі знань 12 Інформаційні технології за спеціальністю 121 Інженерія програмного забезпечення. – Національний технічний університет України «Київський політехнічний інститут імені Ігоря Сікорського», Київ, 2025.

Дисертаційна робота присвячена розробці науково-методичного апарату аналізу акустичних сигналів водного середовища на основі нейронних мереж.

Метою дисертаційної роботи є підвищення точності класифікації та підвищення продуктивності обробки при аналізі акустичних сигналів.

Об'єктом досліджень є процеси функціонування програмних засобів обробки і аналізу акустичних сигналів.

Предметом досліджень є методи, алгоритми та програмні засоби обробки та аналізу акустичних сигналів.

Обробка та аналіз акустичних даних підводного середовища є складною та затратною задачею через складні умови навколишнього середовища. Наприклад, стандартні процедури обробки сигналів не впораються з поставленою задачею через фактори, такі як наявність численних джерел шумів, нерівні границі водного каналу, низьке співвідношення сигнал/шум, великий обсяг отриманої інформації тощо. Більшість класичних методів використовують спрощені припущення для зручної математичної обробки, що неухильно призводить до неадекватних результатів. У зв'язку з вищезазначеним виникає потреба в альтернативному підході до обробки підводних акустичних сигналів, яким є нейронні мережі.

Вони здатні до асоціативної пам'яті, коли патерни запам'ятовуються та порівнюються з вхідними даними. Вони можуть адаптуватися до навколишнього середовища, регулюючи свої вільні параметри та навіть змінюючи свою структуру чи стратегію розв'язання проблем. Нейронні мережі добре справляються з нечіткою інформацією. Отже, вони мають очевидні переваги перед традиційними методами у галузі розпізнавання образів. Нейронні мережі - це масивно паралельні та розподілені обробники. Вони можуть забезпечити набагато вищу швидкість обчислень, маючи потенціал для реалізації складних обчислень у режимі реального часу. Вони стійкі до збоїв, оскільки складаються з великої кількості нейронів, кожен з яких має переважно локальний зв'язок. Пошкодження декількох нейронів чи їх синаптичних зв'язків не повинно суттєво погіршувати загальну продуктивність. Нейронні мережі, як правило, є нелінійними системами через те, що кожен нейрон у нейронній мережі є нелінійною одиницею обробки. Деякі нейронні мережі є універсальними апроксиматорами в тому сенсі, що вони можуть апроксимувати будь-який неперервний та нелінійний вхід-вихід з будь-якою бажаною точністю. З цих причин нейронні мережі особливо добре справляються з розв'язанням нелінійних задач.

У останні роки нейронні мережі успішно використовувалися для обробки підводних акустичних сигналів у багатьох прикладних застосуваннях. Таким чином, нейронні мережі є задовільною технологією для побудови уніфікованого засобу для аналізу інформації, використання якого допоможе досягти високої гнучкості та ефективності, обробки великих обсягів даних та адаптації до зовнішніх умов.

Проблеми аналізу акустичних сигналів водного середовища були розглянуті в роботі вчених O'Neill J., Roberts S., Smith, J., Miller, R., Wang, X., Ashraf, H., Song, Y., Chen, J., Liu, J., Liu, C. та українських вчених В. А. П'ятаковича, А. М. Василенка, О. В. Хотинського, А. І. Верланя, Є.М. Мачуський.

Робота виконувалась в рамках договору: № Д/020.01/0204.02/58/2020 від 23 грудня 2020 р. між КПІ ім. Ігоря Сікорського та Інститутом Океанографічного Приладобудування Академії наук провінції Шандунь (КНР) про виконання наукової теми “Дослідження інтелектуальних комп’ютерних моделей та алгоритмів аналізу сигналів морського середовища”.

Одержані результати запропонованого методу полягають у наступному:

**Удосконалено метод** аналізу акустичних сигналів водного середовища, а саме, класифікації на основі нейронних мереж шляхом застосування багатомасштабної згортки та середнього пулінгу до нейромережі CNN-SOP, що надає можливість підвищити точність класифікації на 3% для датасетів з низьким вмістом фонових шумів та на 9% для датасетів з високим вмістом фонових шумів.

**Отримав подальший розвиток** метод придушення шумів акустичних сигналів водного середовища зашумлених фоновими шумами на основі нейромережі UWAR GAN шляхом зменшення кількості шарів та розміру вхідного тензора, що уможливорює збереження якості придушення шумів, а саме характеристик SNR та RMSE очищеного сигналу, зменшивши тривалість тренування в 10 разів.

**Вперше запропоновано архітектуру** програмного забезпечення аналізу акустичних сигналів водного середовища на основі нейронних мереж, характерною особливістю якої є застосування адаптивного створення нейромережевих моделей, залежно від характеристик вхідних даних, що дає змогу підвищити точність класифікації сигналів з фоновими шумами на 2%.

За результатами тестування на різних датасетах акустичних сигналів водного середовища, покращений метод CNN SOP показав кращі результати, а саме вищу точність класифікації.

Результати дослідження є вагомим внеском у розвитку теоретичних та практичних дослідженнях аналізу акустичних сигналів водного середовища.

Перспективними напрямками подальших досліджень є подальше покращення методу шляхом імплементації та застосування нейромережі для розширення вибірки акустичних сигналів водного середовища та нейромережі для визначення напрямку приходу акустичного сигналу у водному середовищі.

**Ключові слова:** інженерія програмного забезпечення, програмний нейромережевий засіб, інформаційні технології, штучний інтелект, машинне, навчання, нейромережі, акустичні сигнали водного середовища, згорткові нейромережі, пулінг другого рівня, аналіз акустичних сигналів, багатомасштабна згортка, генеративно-змагальні мережі, алгоритми обробки даних, обробка сигналу, алгоритм.

## ABSTRACT

***Oleksii A. O.*** Methods and software tools for analyzing acoustic signals based on neural network models– Qualification research paper in manuscript form.

Dissertation for obtaining the degree of Doctor of Philosophy in the field of knowledge 12 – Information Technologies, specialty 121 – Software Engineering.  
– National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, 2025.

This dissertation is dedicated to the development of a scientific and methodological framework for the analysis of underwater acoustic signals using neural networks.

The purpose of this research is to improve the classification accuracy and increase the processing performance in the analysis of acoustic signals.

The object of research is the functioning processes of software tools for acoustic signal processing and analysis.

The subject of research is the methods, algorithms, and software tools for processing and analyzing acoustic signals.

The processing and analysis of acoustic data in the underwater environment is a complex and resource-intensive task due to the challenging environmental conditions. Most classical methods rely on simple assumptions for convenient mathematical processing, which leads to inadequate results. Considering this, an alternative approach to underwater acoustic signal processing are neural networks.

Neural networks possess associative memory, allowing them to store and compare patterns with input data. They can adapt to environmental conditions by adjusting their free parameters and even modifying their structure or problem-solving strategy. Neural networks effectively handle fuzzy information and thus offer clear advantages over traditional methods in pattern recognition. Neural networks are massively parallel and distributed processors capable of achieving significantly higher computational speeds and performing complex calculations in real-time. They are fault-tolerant, as they consist of numerous neurons, each with

predominantly local connections. Damage to a few neurons or their synaptic connections should not significantly degrade overall performance. Neural networks are typically nonlinear systems because each neuron within them is a nonlinear processing unit. Some neural networks are universal approximators, meaning they can approximate any continuous and nonlinear input-output function with arbitrary precision. Due to these properties, neural networks excel in solving nonlinear problems.

In recent years, neural networks have been successfully applied to underwater acoustic signal processing in numerous practical applications. Thus, neural networks serve as a promising technology for building a unified tool for information analysis, providing high flexibility and efficiency, the ability to process large data volumes, and adaptation to external conditions.

The problems of underwater acoustic signal analysis have been explored in the works of scientists such as O'Neill J., Roberts S., Smith J., Miller R., Wang X., Ashraf H., Song Y., Chen J., Liu J., Liu C., and Ukrainian researchers V. A. Piatakovich, A. M. Vasylenko, O. V. Khotynskyi, A. I. Verlan, and Ye. M. Machuskyi.

This work was carried out under contract № Д/020.01/0204.02/58/2020, since 23th of December 2020 between the Igor Sikorsky Kyiv Polytechnic Institute and the Institute of Oceanographic Instrumentation of the Shandong Academy of Sciences (China) on the scientific research project “Investigation of Intelligent Computer Models and Algorithms for Marine Environment Signal Analysis.”

The obtained results of the proposed method include the following:

A method for analyzing acoustic signals of the water environment, namely, classification based on neural networks, was improved by applying multiscale convolution and mean pooling to the CNN-SOP neural network, which makes it possible to increase the classification accuracy by 3% for datasets with low background noise and by 9% for datasets with high background noise.

The method of noise suppression of acoustic signals of the water environment noisy with background noise based on the UWAR GAN neural network was further

developed by reducing the number of layers and the size of the input tensor, which makes it possible to preserve the quality of noise suppression, namely the characteristics of SNR and RMSE of the cleaned signal, reducing the training time by 10 times.

For the first time, an architecture of software for analyzing acoustic signals of the water environment based on neural networks is proposed, a characteristic feature of which is the use of adaptive creation of neural network models, depending on the characteristics of the input data, which makes it possible to increase the accuracy of classification of signals with background noise by 2%.

Based on testing results on various datasets of underwater acoustic signals, the improved CNN SOP method demonstrated superior performance, specifically achieving higher classification accuracy.

The research findings represent a significant contribution to both theoretical and practical studies in the field of underwater acoustic signal analysis.

Future research directions include further improving the method by implementing and applying neural networks for expanding underwater acoustic signal datasets and developing a neural network for estimating the direction of arrival of underwater acoustic signals.

**Keywords:** software engineering, neural network-based software tool, information technology, artificial intelligence, machine, learning, neural networks, acoustic signals of water environment, convolutional neural networks, second level pooling, acoustic signal analysis, multiscale convolution, generative adversarial networks, data processing algorithms, signal processing, algorithm.



## СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА

*Наукові праці, в яких опубліковані основні наукові результати дисертації:*

1. Олексій А. О., Верлань А. А. Огляд та порівняння методів машинного навчання для розпізнавання гідроакустичних сигналів. Науковий журнал «Інфокомунікаційні та комп'ютерні технології», № 1 (03). 2022. С. 113 – 120. <https://doi.org/10.36994/2788-5518-2022-01-03-18>.
  2. Олексій А. О., Верлань А. А. Застосування багатошарового персептрона для аналізу акустичних сигналів у водному середовищі/ Науковий журнал «Зв'язок», Випуск №3 (163), 2023. С. 40 – 45. <https://doi.org/10.31673/2412-9070.2023.039000>.
  3. Олексій А., Пуха Г. Створення датасету акустичних сигналів водного середовища для тренування нейромережі для придушення шумів. Науковий журнал «Інформаційні технології та суспільство», Випуск 2 (13). 2024. С.56-60. <https://doi.org/10.32689/maup.it.2024.2.8>
  4. Олексій, А. О. і Верлань А. А., 2024. Поліпшений метод аналізу акустичних сигналів водного середовища на основі згорткової нейромережі SOP. Вісник Вінницького політехнічного інституту, №6, с. 129-134. <https://doi.org/10.31649/1997-9266-2024-177-6-129-134>
- Наукові праці, які засвідчують апробацію матеріалів дисертації:*
5. Олексій, А. О. і Верлань А. А., 2022. Огляд нейромережевих підходів для аналізу акустичних сигналів водного середовища. Теоретичні і прикладні проблеми фізики, математики та інформатики: зб. тез XX Всеукраїнської науково-практичної конференції студентів, аспірантів та молодих вчених, 15 червня 2022 р., м. Київ, 347-350.
  6. Олексій, А. О., Верлань А. А. Нейромережеві підходи до генерації акустичних сигналів водного середовища, Сучасні проблеми наукового забезпечення енергетики: зб. тез XX Міжнародної науково-практичної

конференції молодих вчених і студентів, Т. 2, м. Київ, 25–28 квітня 2023 року с. 128-129.

7. Олексій, А. О., Верлань А. І. 2023. Аналіз акустичних сигналів водного середовища з застосуванням нейромережевих методів. Future of Work: Technological, Generational and Social Shifts: 2nd International Scientific and Practical Internet Conference., May 11-12, 2023, Dnipro, Ukraine, p. 67-68.
8. Олексій, А. О., Варава І. А. Створення датасету акустичних сигналів водного середовища. Системи та технології зв'язку, інформатизації та кібербезпеки: актуальні питання і тенденції розвитку: матеріали III Міжнародної науково-технічної конференції, 30 листопада 2023 року, м. Київ, с. 229.
9. Олексій, А. О., Верлань А. А. Огляд нейромережевих підходів для придушення шумів водного середовища. Системи та засоби штучного інтелекту: тези доповідей Міжнародної наукової молодіжної школи., 10-11 жовтня 2023 р. м. Київ, с. 42-45.
10. Олексій, А. О., Варава І. А. 2023. Аналіз акустичних сигналів водного середовища з застосуванням нейромережевих моделей та створення датасетів акустичних сигналів водного середовища. Сучасні аспекти інженерії програмного забезпечення: збірник тез Першої міжнародної науково-практичної конференції, 14 грудня 2023 року, м. Київ, с. 100-103.
11. Олексій, А. О., Верлань А. І. Аналіз акустичних сигналів водного середовища з застосуванням нейромережевих моделей та створення датасетів акустичних сигналів водного середовища. Сучасні аспекти інженерії програмного забезпечення: збірник тез Другої міжнародної науково-практичної конференції, 13 листопада 2024 року, м. Київ, с. 60-64.

## ЗМІСТ

ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ .....	14
ВСТУП.....	15
РОЗДІЛ 1. ДОСЛІДЖЕННЯ УМОВ, МІСЦЯ ВИКОРИСТАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ В ПРОЦЕСІ АНАЛІЗУ АКУСТИЧНИХ СИГНАЛІВ МОРСЬКОГО СЕРЕДОВИЩА .....	21
1.1. Аналіз акустичних шумів у морському середовищі .....	21
1.2. Аналіз моделей випромінювальних шумів надводних і підводних об'єктів .....	23
1.3. Стан сучасного програмного забезпечення для аналізу гідроакустичних систем .....	24
1.4. Аналіз сучасних наукових методів аналізу гідроакустичних сигналів 36	
1.5. Огляд методів машинного навчання, що можуть бути застосовані для аналізу акустичних сигналів водного середовища .....	39
1.6. Дослідження різних нейромережових алгоритмів з метою використання для аналізу акустичних сигналів морського середовища з урахуванням його особливостей .....	45
1.6.1. Екстремальна машина самонавчання .....	46
1.6.2. Згорткові нейронні мережі (CNN) .....	47
1.6.3. Рекурентні нейромережі.....	50
1.6.4. Нейромережа довгої короткострокової пам'яті (LSTM).....	50
1.6.5. Машина Больцмана .....	52
1.6.6. Мережа глибокої довіри .....	53
1.6.7. Автоенкодер.....	55
1.7. Огляд нейромережових підходів для обробки гідроакустичних сигналів	58
1.7.1. Згорткові нейромережі .....	58
1.7.2. Автоенкодери.....	62
1.7.3. Багатошарові перцептрони .....	65
1.8. Нейромережі для придушення шумів .....	66
1.9. Обґрунтування вибору нейромереж для виконання задачі аналізу акустичних сигналів водного середовища за задачі аналізу .....	81
Висновки до розділу 1 .....	85
РОЗДІЛ 2. ФОРМАЛІЗАЦІЯ ТА МАТЕМАТИЧНЕ ПРЕДСТАВЛЕННЯ МЕТОДІВ АНАЛІЗУ АКУСТИЧНИХ СИГНАЛІВ ВОДНОГО СЕРЕДОВИЩА .....	87

2.1. Формалізація задачі класифікації акустичних сигналів водного середовища .....	87
2.1.1. Постановка задачі класифікації акустичних сигналів .....	87
2.1.2. Представлення акустичного сигналу в різних областях .....	90
2.1.3. Метрики оцінки результатів класифікації акустичних сигналів водного середовища .....	91
2.1.4. Можливі виклики при виконанні задачі класифікації акустичних сигналів водного середовища .....	94
2.2. Формалізація задачі придушення фонових шумів в акустичних сигналах водного середовища .....	94
2.2.2. Загальне представлення архітектури нейронної мережі.....	95
2.2.3. Функція втрат, подання вхідних даних та вибір алгоритму оптимізації .....	96
2.2.4. Опис алгоритму навчання та функцій втрат нейромережі для придушення шумів .....	96
2.2.5. Метрики оцінки придушення фонових шумів водного середовища .....	98
2.3. Математичне представлення згорткової нейромережі SOP.....	100
2.3.1. Опис попередньої обробки вхідних характеристик до нейромережі .....	100
2.3.2. Покращений метод класифікації акустичних сигналів водного середовища на основі нейромережі CNN SOP .....	102
2.4. Опис нейромережі для придушення шумів UWAR-GAN.....	105
2.4.1. Математичне представлення нейромереж для придушення фонових шумів водного середовища .....	105
2.4.2. Попередня обробка даних методу для придушення шумів.....	109
Висновки до розділу 2.....	110
РОЗДІЛ 3. ЕКСПЕРИМЕНТАЛЬНА ПЕРЕВІРКА МЕТОДІВ АНАЛІЗУ АКУСТИЧНИХ СИГНАЛІВ ВОДНОГО СЕРЕДОВИЩА.....	112
3.1. Розробка датасетів акустичних сигналів водного середовища .....	112
3.1.1. Огляд датасетів акустичних сигналів водного середовища .....	112
3.1.2. Створення датасетів акустичних сигналів водного середовища .	114
3.1.3. Процес навчання нейромережі придушення шумів водного середовища	117
3.2. Підбір гіперпараметрів покращеної згорткової нейромережі CNN-SOP	120
3.3. Проведення перевірки моделі CNN SOP.....	122
3.4. Проведення перевірки моделі UWAR-GAN .....	127

3.5. Результати придушення шумів та класифікації.....	130
3.5.1. Результати придушення шумів акустичних сигналів водного середовища нейромережею UWAR GAN .....	130
3.5.2. Оцінка результатів класифікації .....	133
Висновки до розділу 3.....	134
РОЗДІЛ 4. ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ АНАЛІЗУ АКУСТИЧНИХ СИГНАЛІВ ВОДНОГО СЕРЕДОВИЩА .....	135
4.1. Розроблене програмне забезпечення.....	135
4.1.2. Модуль тренування нейромережі для придушення шумів .....	137
4.1.3. Модуль виконання придушення шумів нейромережею .....	140
4.1.4 Модуль тренування нейромережі для класифікації акустичних сигналів .....	142
4.1.5. Модуль виконання класифікації акустичних сигналів .....	145
4.1.6. Модуль аналізу акустичних сигналів .....	147
4.2. Адаптивний пайплайн для створення нейромережевих моделей.....	149
4.3. Алгоритмічне представлення програмного засобу аналізу акустичних сигналів водного середовища .....	152
4.3.1. Опис основного алгоритму модуля виконання класифікації акустичних сигналів .....	152
4.3.2. Опис основного алгоритму модуля виконання придушення фонових шумів у акустичних сигналах .....	154
4.3.3. Опис основного алгоритму тренування нейромережі для придушення фонових шумів водного середовища .....	156
4.3.4. Опис основного алгоритму модуля виконання тренування нейромережі класифікації акустичних сигналів водного середовища ..	158
Висновки до розділу 4.....	161
ВИСНОВКИ .....	164
ДОДАТОК А .....	178
ДОДАТОК Б .....	180
ДОДАТОК В.....	182

## ПЕРЕЛІК УМОВНИХ СКОРОЧЕНЬ

CNN – згорткова нейронна мережа

SOP – пулінг другого порядку

MSC – багатомасштабна згортка

SNR – співвідношення сигналу та шуму

PSNR – пікове співвідношення сигналу та шуму

RMSE – середньоквадратична похибка

CQT – перетворення з постійним Q

STFT – Віконне перетворення Фур'є

ISTFT – Обернене віконне перетворення Фур'є

GAN – генеративно-змагальні мережі

VRAM – відеопам'ять

## ВСТУП

**Актуальність роботи.** Обробка акустичних даних у підводному середовищі є складним завданням через шум, низьке співвідношення сигнал/шум та великий обсяг даних. Традиційні методи часто неефективні через спрощені припущення. Нейронні мережі пропонують альтернативний підхід, оскільки вони можуть адаптуватися до умов, працювати з нечіткими даними, здійснювати складні обчислення в реальному часі та стійко функціонувати навіть при часткових пошкодженнях. Вони є універсальними апроксиматорами для розв'язання нелінійних задач, що робить їх ідеальними для аналізу підводних акустичних сигналів. Нейронні мережі забезпечують гнучкість, ефективність, обробку великих обсягів даних та автономність, що робить їх оптимальним рішенням для аналізу інформації в цьому середовищі.

На підставі проведеного аналізу виявлено існування протиріч між:

- вимогою ефективної обробки та аналізу великого обсягу одержуваної інформації в реальному часі;
- необхідністю аналізу даних за умови наявності численних джерел сигналів із сильними перешкодами або низьким відношенням сигнал/шум;
- вимогою швидкої обробки отриманих даних;
- високою точністю аналізу (класифікації) сигналів.

Ці протиріччя лежать в основі наукового завдання щодо *розробки методів і програмних засобів аналізу акустичних сигналів водного середовища, які дозволять виявляти сигнал цільового об'єкта за низького відношення сигнал/шум.*

**Зв'язок роботи з науковими програмами, планами, темами.** Робота виконувалася на кафедрі інженерії програмного забезпечення в енергетиці рамках договору № Д/020.01/0204.02/58/2020 від 23 грудня 2020 р. між НТУУ КПІ ім. Ігоря Сікорського та Інститутом океанографічного приладобудування Академії наук провінції Шаньдунь (КНР) щодо виконання наукової теми

"Дослідження інтелектуальних комп'ютерних моделей та алгоритмів аналізу сигналів морського середовища".

**Мета і завдання дослідження.**

**Метою дисертаційної роботи** є підвищення точності класифікації та підвищення продуктивності обробки при аналізі акустичних сигналів.

Для досягнення мети в було сформульовано такі **наукові завдання**:

- розробка архітектури програмного засобу аналізу акустичних сигналів водного середовища на основі нейронних мереж сигналів водного середовища;
- удосконалення методів аналізу, а саме класифікації акустичних сигналів водного середовища на основі нейронних мереж;
- надання подальшого розвитку методу придушення шумів акустичних сигналів водного середовища, зашумлених фоновими шумами на основі нейронних мереж;

**Об'єкт досліджень** - процеси функціонування програмних засобів обробки та аналізу акустичних сигналів.

**Предмет досліджень** - методи та програмні засоби обробки та аналізу акустичних сигналів.

**Методи дослідження:** методи системного та порівняльного аналізу (для визначення актуальності та постановки мети дисертаційного дослідження). Методи аналізу на основі Data Science (для придушення фонових шумів та класифікації). Методи цифрової обробки сигналів (для створення наборів даних акустичних сигналів).

**Наукова новизна одержаних результатів** полягає в наступному:

**Удосконалено метод** аналізу акустичних сигналів водного середовища, а саме класифікації на основі нейронних мереж шляхом застосування багатомасштабної згортки та середнього пулінгу до нейромережі CNN-SOP, що надає можливість підвищити точність класифікації на 3% для датасетів з



низьким вмістом фонових шумів та на 9% для датасетів з високим вмістом фонових шумів.

**Отримав подальший розвиток** метод придушення шумів акустичних сигналів водного середовища зашумлених фоновими шумами на основі нейромережі UWAR GAN шляхом зменшення кількості шарів та розміру вхідного тензора, що уможливило збереження якості придушення шумів, а саме характеристик SNR та RMSE очищеного сигналу, зменшивши тривалість тренування в 10 разів.

**Вперше запропоновано архітектуру** програмного забезпечення аналізу акустичних сигналів водного середовища на основі нейронних мереж, характерною особливістю якої є застосування адаптивного створення нейромережевих моделей, залежно від характеристик вхідних даних, що дає змогу підвищити точність класифікації сигналів з фоновими шумами на 2%.

**Практичне значення одержаних результатів** полягає в покращенні результатів класифікації акустичних сигналів завдяки застосуванню багатомасштабної згортки, середнього пулінгу, адаптивного створення нейромережевих моделей класифікатора, а також у підвищенні продуктивності придушення шумів акустичних сигналів за рахунок спрощення архітектури та зменшення розміру вхідного тензора.

#### **Впровадження одержаних результатів:**

Результати досліджень прийняті до впровадження в конструкторському бюро «Шторм» (акт реалізації від 25.03.2025 р.); в навчальному процесі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» (акт реалізації від 07.04.2025 р.) при викладанні дисципліни «Методології розробки інтелектуальних комп'ютерних програм» для студентів освітньо-кваліфікаційного рівня «Бакалавр» спеціальності 121 «Інженерія програмного забезпечення».

**Особистий внесок здобувача.** Дисертаційна робота є результатом самостійного наукового дослідження. Усі представлені наукові результати,

викладені в дисертації, є результатом праці здобувача. Було опубліковано наукові роботи у співавторстві, що є безпосереднім внеском автора. У дисертаційній роботі використані лише ті результати, які є індивідуальним внеском здобувача.

У статтях зі співавторами здобувачеві належать такі результати:

У роботі [1] проведено огляд основних методів машинного навчання та обґрунтовано переваги й необхідність вибору нейронних мереж у порівнянні з іншими методами машинного навчання. Також зроблено огляд різних способів представлення даних для аналізу акустичних сигналів.

У роботі [2] розглянуто метод класифікації акустичних сигналів водного середовища, що передбачав використання запропонованого багатоповільного перцептрона. Він продемонстрував високу точність класифікації при роботі з акустичними семплами водних шумів, що мають частоту дискретизації 16 000 Гц.

У роботі [3] зроблено огляд різних датасетів водного середовища, описано процес створення датасету акустичних сигналів та представлено попередні результати придушення шумів.

У роботі [4] розглянуто метод аналізу акустичних сигналів водного середовища на основі згорткової нейромережі CNN-SOP, а також обґрунтовано необхідність покращення цього методу. Описано вдосконалення підходу, а отриману модель протестовано на різних датасетах. Результати тестування показали вищу точність класифікації у всіх перевірених випадках.

У роботі [5] здійснено огляд різних нейромереж для класифікації акустичних сигналів водного середовища та зроблено припущення щодо вибору моделі, яка відповідає заданим критеріям.

У роботі [6] зроблено припущення щодо необхідності застосування нейромереж для розширення вибірки акустичних сигналів та виконано огляд робіт, присвячених цьому питанню.

У роботі [7] розглянуто можливість створення нейромережевого комплексу для розширення вибірки акустичних сигналів водного середовища та їх подальшої класифікації.

У роботі [8] обґрунтовано необхідність створення датасету для тестування акустичних сигналів водного середовища з достатнім рівнем фонових шумів.

У роботі [9] зроблено огляд нейромережевих методів придушення шумів у водному середовищі та визначено оптимальну модель, яка відповідає заданим критеріям.

У роботі [10] описано створення датасетів акустичних сигналів водного середовища та запропоновано покращити результати класифікації за допомогою нейромережевого методу придушення шумів.

У роботі [11] розглянуто нейромережевий класифікатор акустичних сигналів водного середовища на основі згорткової нейромережі CNN-SOP, запропоновано та обґрунтовано його вдосконалення.

#### **Апробація результатів дисертації:**

Результати та основні положення роботи доповідалися та обговорювалися на:

XX Всеукраїнська науково-практична конференція «Теоретичні і прикладні проблеми фізики, математики та інформатики», Міжнародний електронний науково-практичний журнал Wayscience: «FUTUTRE OF WORK: TECHNOLOGICAL GENERATIONAL AND SOCIAL SHIFTS», XX міжнародна науково-практична конференція молодих вчених та студентів «Сучасні проблеми наукового забезпечення енергетики», XXI Міжнародна науково-практична конференція молодих вчених та студентів «СУЧАСНІ ПРОБЛЕМИ НАУКОВОГО ЗАБЕЗПЕЧЕННЯ ЕНЕРГЕТИКИ», XXIII Міжнародна науково-технічна конференція ШТУЧНИЙ ІНТЕЛЕКТ ТА ІНТЕЛЕКТУАЛЬНІ СИСТЕМИ AIIS'2023 Міжнародна наукова молодіжна школа «СИСТЕМИ ТА ЗАСОБИ ШТУЧНОГО ІНТЕЛЕКТУ», III

МІЖНАРОДНА НАУКОВО-ТЕХНІЧНА КОНФЕРЕНЦІЯ «Системи і технології зв'язку, інформатизації та кібербезпеки: актуальні питання і тенденції розвитку», I Міжнародна науково-практична конференція «Сучасні аспекти інженерії програмного забезпечення».

**Публікації.** За матеріалами дисертації опубліковано 11 робіт, з яких 4 – це статті у періодичних журналах, що входять до переліку фахових видань, затверджених МОН України за спеціальністю дисертації та 7 – публікації у матеріалах конференцій.

## **РОЗДІЛ 1. ДОСЛІДЖЕННЯ УМОВ, МІСЦЯ ВИКОРИСТАННЯ МЕТОДІВ МАШИННОГО НАВЧАННЯ В ПРОЦЕСІ АНАЛІЗУ АКУСТИЧНИХ СИГНАЛІВ МОРСЬКОГО СЕРЕДОВИЩА**

Аналіз гідроакустичних сигналів є важливою складовою морської акустики, оскільки він забезпечує ідентифікацію, класифікацію та локалізацію об'єктів у водному середовищі. Завдання цього напрямку включають дослідження фізичних характеристик звукових хвиль, методів їх аналізу та практичного застосування в умовах різних шумових середовищ.

### **1.1. Аналіз акустичних шумів у морському середовищі**

Навколишній шум в океанському середовищі змінюється з часом через різноманітні джерела, зокрема вітер, дощ, рух суден і біологічну активність. Ці фактори можуть змінюватися в часовому діапазоні від кількох хвилин до кількох днів. Погодні зміни відбуваються поступово, тоді як у прибережних районах шум може різко змінюватися через місцеву активність невеликих суден, особливо у вихідні дні.

Заміри шуму в глибоководних районах показують його мінливість на рівні 6 дБ із періодом від кількох годин до кількох днів. Шум, спричинений морським станом, демонструє схожу варіативність, але з довгими часовими сталими [12].

Рівень акустичного шуму на мілководді на 5 дБ вищий, ніж у глибоководді за однакових умов швидкості вітру та частоти. Проте можливі випадки, коли рівень шуму на великих глибинах виявляється таким самим або навіть нижчим, ніж на мілководді. Ця відмінність залежить від поєднання вітрозалежних і незалежних від вітру компонентів шуму [14]. Наприклад, на низьких частотах у глибокій воді рівень шуму зменшується з глибиною, особливо на частотах, спричинених суднами. Найменший рівень шуму спостерігається на дні океану або в його безпосередній близькості.

Водночас, на високих частотах, зокрема понад 10 кГц, рівень шуму швидко зменшується через поглинання звуку водою. Це явище особливо помітне для теплового шуму, який визначається внутрішніми процесами середовища [13].

Рельєф океанського дна суттєво впливає на рівень акустичного шуму та його варіації. Дослідження, проведені поблизу Гренландії, показали, що навіть за різних профілів глибин на відстані кількох миль рівень шуму на великих глибинах і низьких частотах може відрізнятися до 10 дБ. Батиметрія може як підвищувати, так і знижувати рівень шуму.

Морріс [16] описав випадок, коли відбиття звуку від морського дна на великих глибинах призводило до підвищення рівня шуму на 5 дБ. Інші дослідження також показали, що зміна швидкості звуку в різних шарах води впливає на просторовий розподіл шуму, зокрема на суднових маршрутах [17].

Напрямок поширення шуму в океані залежить від численних факторів, зокрема джерел шуму та умов середовища, таких як швидкість звуку у воді. Шум від суден, розташованих на поверхні або поблизу неї, має виражену анізотропію, особливо на низьких частотах. Це означає, що напрямок поширення звукових хвиль може суттєво змінюватися залежно від географічного розташування суден та гідрофонів.

Дослідження показали, що в реальних умовах анізотропія шуму впливає на кореляцію між гідрофонами, що має важливе значення для моніторингу підводного акустичного середовища [16].

Шум в Арктиці поділяється на два основні типи: шум під паковим льодом та шум у зоні краєвого льодовика. Перший тип зазвичай має низький рівень, оскільки відсутній хвильовий вплив, а шум формується переважно внаслідок тріщин льоду.

У зоні краєвого льодовика, де присутні активні хвилі, сильний вітер і судноплавство, рівень шуму суттєво зростає. Переміщення приймача з

льодовикового покриву у відкритий океан дозволяє спостерігати значні зміни в частотному спектрі шуму, зокрема на частотах від 100 Гц до 1 кГц [12].

Шум поділяється на два основні типи: широкосмуговий і вузькосмуговий. Широкосмугові шуми охоплюють великий діапазон частот, тоді як вузькосмугові шуми концентруються на певних частотах.

Оскільки обробка цих видів шуму потребує різних підходів, важливо правильно їх ідентифікувати для подальшого аналізу. Широкосмугові компоненти часто мають менші пікові значення, тоді як вузькосмугові компоненти можуть бути стабільними і фіксованими на певних частотах, наприклад, частоті обертання пропелера або резонансних частотах [14, 15].

## **1.2. Аналіз моделей випромінювальних шумів надводних і підводних об'єктів**

Гвинтовий шум виникає через обертання лопатей, кавітацію та резонансні коливання. Шум обертання лопатей спричинений зміною тяги під час їхнього проходження через нерівності потоку, що проявляється у вигляді гармонічного ряду частот. Кавітація, яка виникає за високих швидкостей потоку, призводить до утворення газових бульбашок, що викликають імпульсний шум широкосмугового характеру. Цей тип шуму домінує на високих частотах і часто визначає загальний рівень шуму для поверхневих суден [15].

Машинний шум, який виникає від усіх рухомих частин судна, спричинений зворотно-поступальними рухами деталей, вибухами в циліндрах дизельних двигунів та взаємодією лопатей турбін. Цей шум може бути як вузькосмуговим, так і широкосмуговим, залежно від характеру джерела. Наприклад, турбулентність у гідравлічних системах і насосах зазвичай створює широкосмуговий шум, тоді як періодичні вибухи в двигунах— вузькосмуговий [12].

Для точних оцінок рівня шуму від суден Дональд Росс у своїй роботі запропонував рівняння для розрахунку акустичних характеристик суден, зокрема для великого судна масою 100 000 тонн. Сумарний шум судна концентрується в нижній частині спектра і зростає зі збільшенням швидкості. Гвинт є основним джерелом шуму на частотах вище 100 Гц, тоді як на нижчих частотах значну роль відіграють дизельні двигуни та генератори [15].

### **1.3. Стан сучасного програмного забезпечення для аналізу гідроакустичних систем**

SpectraPLUS [18] — це потужне програмне забезпечення для аналізу сигналів, призначене головним чином для загального акустичного та вібраційного аналізу, яке пропонує широкий набір функцій, таких як реєстрація сигналів у реальному часі, спектральний аналіз за допомогою БПФ (FFT) та візуалізація даних. Програма підтримує налаштування частоти дискретизації та запис даних у реальному часі через сумісні звукові карти та USB-аудіопристрої. Така адаптивність робить її придатною для різноманітних застосувань, включаючи моніторинг довкілля, промислову акустику та біомедичні дослідження. До її розширених можливостей візуалізації належать детальні спектрограми, 3D-графіки типу "водоспад" та налаштовувані дисплеї, що дозволяють користувачам детально спостерігати та аналізувати характеристики сигналів. Ці функції, у поєднанні з інструментами для створення користувацьких фільтрів, застосування смугових і шумознижувальних фільтрів, а також вимірювання таких показників, як загальна гармонічна спотвореність (THD), роблять SpectraPLUS ефективним інструментом для дослідження складних акустичних даних.

SpectraPLUS розроблено з урахуванням зручності використання. Її інтуїтивно зрозумілий інтерфейс гарантує, що навіть користувачі з обмеженим досвідом у аналізі сигналів зможуть легко працювати з інструментами та функціями. Програма дозволяє здійснювати обширне



логування даних та їх відтворення, що дає змогу захоплювати й аналізувати акустичні дані з часом. Крім того, сумісність із різними форматами файлів, такими як .WAV, .TXT та .CSV, підвищує її гнучкість, полегшуючи інтеграцію даних та співпрацю з іншими платформами. Іншою примітною функцією є підтримка скриптів, яка дозволяє автоматизувати повторювані завдання та оптимізувати робочі процеси, що є значною перевагою під час роботи з великими наборами даних.

Попри свою універсальність, SpectraPLUS має обмежене застосування для аналізу підводних акустичних сигналів. Хоча програмне забезпечення чудово справляється із загальною обробкою акустичних сигналів, воно не має спеціалізованих функцій, адаптованих для аналізу сигналів водного середовища, таких як класифікація звуків морських ссавців, аналіз навколишнього шуму чи алгоритми локалізації джерел звуку. Крім того, програмі бракує інтеграції сучасних інструментів, таких як машинне навчання, які стають дедалі важливішими для виконання завдань, пов'язаних із класифікацією шуму та ідентифікацією джерел. Залежність від зовнішнього обладнання, наприклад звукових карт або USB-пристроїв для збору даних, також може створювати труднощі в підводному середовищі, де потрібні спеціалізовані гідрофони. Така залежність від сумісного обладнання може ускладнити використання програми в морських умовах, де важливі конкретні характеристики сигналів і налаштування збору даних.

Хоча SpectraPLUS пропонує високоточні візуалізації та потужні можливості фільтрації, відсутність спеціалізованих функцій для морських досліджень обмежує її цінність для сфокусованого підводного аналізу. Наприклад, інструменти для генерації спектрограм і фільтрації сигналів можуть бути застосовані для аналізу підводного шуму, проте програмі не вистачає глибини налаштувань або спеціалізації, необхідних для задач аналізу акустичних сигналів водного середовища, таких як аналіз частотних діапазонів, специфічних для окремих видів водної фауни, або інтеграція моделей класифікації.

Крім того, комерційна ліцензія SpectraPLUS може стати бар'єром для академічних або неприбуткових дослідників з обмеженими бюджетами, знижуючи її доступність у цих секторах.

У цілому, SpectraPLUS є потужним інструментом для загального акустичного аналізу, пропонуючи широкі можливості візуалізації та обробки даних. Однак її обмеження у вирішенні унікальних завдань підводної акустики підкреслюють необхідність більш спеціалізованого програмного забезпечення. Платформа, яка поєднує сильні сторони SpectraPLUS із цільовими функціями для морських середовищ, такими як класифікація на основі машинного навчання та сумісність із пристроями для підводного запису, могла б заповнити критичну прогалину в галузі аналізу підводних акустичних сигналів. Хоча SpectraPLUS може бути корисним для загальних підводних досліджень, її ширший вплив у галузі морської обмежений відсутністю інструментів, орієнтованих на специфічні завдання, і адаптивності до потреб підводних досліджень.

Underwater Acoustic Signal Analysis Toolkit [19], розроблений Кірком Бьєневу молодшим, — це програмне забезпечення на базі MATLAB, створене для вирішення специфічних потреб підводних акустичних досліджень. Цей інструментарій був створений у рамках зусиль консорціуму Littoral Acoustic Demonstration Center Gulf Ecological Monitoring and Modeling (LADC-GEMM) для підвищення ефективності аналізу даних у морській біоакустиці. Програмне забезпечення спеціалізується на візуалізації, фільтрації та аналізі підводних акустичних даних, зокрема на обробці сигналів, таких як кліки морських ссавців. Воно призначене для роботи з великими наборами даних, зібраних у ході довгострокового підводного моніторингу, і пропонує інструменти, які значно спрощують раніше ручні процеси.

Underwater Acoustic Signal Analysis Toolkit включає графічний інтерфейс користувача (GUI), який забезпечує доступне та інтерактивне середовище для дослідників. Цей інтерфейс дозволяє завантажувати дані у

різних форматах, включаючи бінарні та текстові файли, і виконувати такі операції, як генерація часових рядів, спектрограм, спектральної щільності потужності та застосування фільтрів для зниження шуму. Візуалізація часових рядів дає змогу користувачам ефективно навігувати даними та виділяти конкретні області інтересу. Ця функція є особливо корисною для аналізу вокалізацій морських ссавців, оскільки вона дозволяє детально досліджувати акустичні події. Генерація спектрограм — ще одна ключова особливість, яка забезпечує високоточний аналіз у часо-частотному просторі за допомогою STFT. Програмне забезпечення також інтегрує вейвлет-аналіз для більш точного представлення частот, що забезпечує кращу локалізацію високочастотних сигналів у часі порівняно з традиційними методами Фур'є.

Важливою перевагою цього інструментарію є здатність автоматизувати складні розрахунки та процеси каталогізації. Наприклад, він може розраховувати метрики сигналів, такі як пікова частота, ширина смуги пропускання та характеристики щільності спектральної потужності, які є критично важливими для класифікації видів. Ці можливості доповнюються алгоритмами для фільтрації шуму через смугові фільтри, використовуючи техніки, такі як фільтри Баттерворта, які пригнічують небажані частоти, зберігаючи при цьому сигнали, що цікавлять. Модульна архітектура програмного забезпечення дозволяє інтегрувати нові інструменти, що дозволяє дослідникам адаптувати його до змінних аналітичних потреб.

Одним із найвизначніших аспектів Underwater Acoustic Signal Analysis Toolkit є його спрямованість на підвищення продуктивності досліджень. Завдяки автоматизації багатьох ручних завдань, раніше необхідних для обробки даних, програмне забезпечення значно скорочує час, потрібний для аналізу сигналів. Наприклад, каталогізація шумів морських ссавців — процес, який раніше вимагав значних зусиль вручну, тепер може бути виконана за кілька хвилин. Ця ефективність зробила інструментарій цінним ресурсом для дослідників та освітніх програм, де його використовують для навчання студентів аналізу сигналів.

Попри свої переваги, інструментарій має певні обмеження. Його залежність від MATLAB як середовища розробки створює серйозний бар'єр, оскільки ліцензія MATLAB може бути дорогою для деяких дослідників, особливо в академічних або неприбуткових організаціях. Крім того, хоча Underwater Acoustic Signal Analysis Toolkit є високоефективним для конкретних підводних досліджень, його основна увага зосереджена на аналізі сигналів морських ссавців, і він може не задовольнити ширші потреби підводної акустики, наприклад, аналіз даних сонарів або загальний моніторинг шуму. Також, попри інтеграцію передових методів візуалізації та фільтрації, програмі бракує інструментів на основі машинного навчання для автоматичної класифікації чи прогнозування, які стають стандартом у галузі акустичних досліджень.

У цілому, Underwater Acoustic Signal Analysis Toolkit забезпечує високоспеціалізоване рішення для аналізу акустичних сигналів водного середовища, відповідаючи на критичні потреби в ефективному та точному аналізі даних. Програма відзначається успіхами в обробці сигналів морських ссавців, візуалізації у часо-частотному просторі та автоматизації каталогізації, що робить її незамінним інструментом для дослідників у цій вузькій галузі. Однак її орієнтація на конкретні застосунки та залежність від MATLAB можуть обмежити ширше впровадження. Майбутня версія програмного забезпечення, яка інтегрує можливості машинного навчання та підтримує ширший спектр підводних акустичних даних, могла б зробити його більш універсальним і придатним для використання у різних галузях. Наразі Underwater Acoustic Signal Analysis Toolkit закриває важливу нішу в дослідженнях акустичних аналізів водного середовища, проте зростаючі вимоги галузі вказують на можливості для подальшого розвитку та інновацій.

Sigview [20] — це універсальне програмне забезпечення, призначене для аналізу сигналів, яке пропонує широкий набір функцій, що підходять для обробки акустичних даних загального призначення. Його високо цінують за зручний інтерфейс і можливість працювати з різними типами сигналів, що

робить його доступним як для початківців, так і для досвідчених дослідників. Sigview підтримує візуалізацію та аналіз сигналів за допомогою таких інструментів, як генерація спектрограм, обробка даних у реальному часі та виконання математичних операцій над сигналами. Завдяки можливості обробляти дані в реальному часі програмне забезпечення ідеально підходить для застосувань, які потребують негайного зворотного зв'язку, наприклад, аналізу вібрацій, аналізу мовлення та моніторингу навколишнього шуму.

Однією з ключових особливостей Sigview є потужна функція генерації спектрограм, яка дозволяє візуалізувати частотний вміст сигналу у часі з регульованими параметрами роздільної здатності та масштабування. Ця функціональність є важливою для ідентифікації частотних шаблонів у складних сигналах, зокрема в підводних акустичних середовищах. Програмне забезпечення також підтримує створення 3D спектрограм, що забезпечує додатковий рівень розуміння тимчасових і спектральних характеристик акустичних сигналів. Ці візуальні інструменти доповнюються розширеними опціями фільтрації, такими як смуговий, низькочастотний та високочастотний фільтри, що дозволяє користувачам виділяти певні частотні діапазони для детального аналізу.

Sigview є гнучким інструментом завдяки вбудованій підтримці скриптів, яка дозволяє автоматизувати повторювані завдання та розширювати функціонал програмного забезпечення. Ця можливість є особливо корисною для дослідників, які працюють із великими наборами даних, оскільки вона зменшує час і зусилля, необхідні для виконання рутинних аналізів. Крім того, програмне забезпечення підтримує різноманітні формати файлів, що забезпечує сумісність з іншими інструментами та полегшує інтеграцію в існуючі робочі процеси. Його математичний інструментарій дозволяє виконувати складні операції над сигналами, такі як перетворення Фур'є та вейвлет-перетворення, статистичний аналіз і згладжування даних, що робить його потужним інструментом для детальної характеристики сигналів.

Однак, попри свої сильні сторони, Sigview має певні обмеження при застосуванні в підводних акустичних дослідженнях. Хоча програмне забезпечення чудово справляється із загальним аналізом сигналів, йому бракує спеціалізованих інструментів і функцій, адаптованих для підводного середовища. Наприклад, у ньому немає вбудованої підтримки біоакустичного аналізу, такого як класифікація специфічних для видів звуків або аналіз навколишнього шуму, що є важливими для морських досліджень. Крім того, воно не містить можливостей машинного навчання, які дедалі частіше використовуються в акустичних дослідженнях для автоматичного розпізнавання шаблонів і завдань класифікації. Ці недоліки обмежують його ефективність у вирішенні унікальних завдань обробки підводних сигналів.

Ще одним обмеженням є залежність програмного забезпечення від зовнішнього обладнання для збору даних. Хоча Sigview сумісний із широким спектром пристроїв, процес інтеграції може бути складним для дослідників, які використовують спеціалізоване обладнання для запису підводних сигналів, таке як гідрофони. Ця залежність може створювати додаткові труднощі в налаштуванні програмного забезпечення для морських застосувань, де конфігурації збору даних можуть значно відрізнятись.

Отже, Sigview є потужним і гнучким інструментом для загального аналізу сигналів, пропонуючи вражаючий набір функцій для візуалізації, фільтрації та математичної обробки. Його зручний дизайн і можливості налаштування роблять його цінним ресурсом для дослідників у різних галузях. Однак відсутність спеціалізованих інструментів для підводної акустики обмежує його корисність у морській біоакустиці та суміжних сферах. Хоча програму можна адаптувати для деяких підводних застосувань, дослідникам у цій галузі, ймовірно, доведеться доповнювати Sigview іншими програмами, розробленими для морських середовищ. Для повного задоволення потреб підводних акустичних досліджень майбутні версії Sigview могли б отримати вигоду від включення спеціалізованих біоакустичних функцій і підтримки сучасних технік машинного навчання.

Попри ці обмеження, Sigview залишається надійним і ефективним вибором для загального аналізу акустичних даних, вирізняючись своєю гнучкістю та простотою у використанні.

DaViS Underwater Acoustic Data Visualization Software [21], розроблене Abyssens, є спеціалізованим інструментом для візуалізації та аналізу підводних акустичних даних. Це програмне забезпечення орієнтоване на надання дослідникам інтуїтивної та інтерактивної платформи для вивчення складних підводних акустичних середовищ. Його головна перевага полягає у здатності перетворювати акустичні сигнали на детальні й інформативні візуальні представлення, що дозволяє користувачам отримувати глибше розуміння підводних звукових ландшафтів, розподілу шуму та сигналів.

Однією з ключових особливостей DaViS є його розширені можливості створення спектрограм. Програмне забезпечення забезпечує високоточну візуалізацію у часо-частотному просторі, яка є важливою для аналізу підводних звуків, таких як вокалізації морських ссавців, шум кораблів і навколишній шум довкілля. Інтерактивні спектрограми дозволяють користувачам масштабувати окремі частотні діапазони чи часові сегменти, щоб детально досліджувати інформацію. Ця функція особливо корисна для ідентифікації тонких шаблонів у акустичних даних, наприклад, наявності біоакустичних сигналів або впливу антропогенного шуму на морські екосистеми. DaViS також підтримує багатовимірну візуалізацію даних, зокрема побудову 3D-графіків, що додає аналітичній глибини. Такі візуалізації дозволяють дослідникам оцінювати тимчасові та просторові характеристики акустичних сигналів більш комплексно. Можливість накладати кілька шарів даних, наприклад, інтенсивність шуму та частотний вміст, підвищує корисність програмного забезпечення для порівняння та кореляції різних акустичних характеристик. Ця функція особливо цінна для досліджень, у яких важливо розуміти взаємодію між різними джерелами підводного шуму. Ще однією перевагою DaViS є зручний інтерфейс, який спрощує процес імпорту, візуалізації та інтерпретації акустичних даних.

Програмне забезпечення розроблене таким чином, щоб задовольнити потреби користувачів із різними рівнями досвіду, роблячи його доступним як для досвідчених дослідників, так і для новачків у сфері підводної акустики. Його оптимізовані робочі процеси знижують складність обробки даних, дозволяючи дослідникам зосередитися на аналізі, а не на подоланні технічних труднощів.

Однак DaViS здебільшого зосереджене на візуалізації й не включає деяких розширених інструментів для обробки сигналів і аналітики, необхідних для всебічних підводних акустичних досліджень. Хоча воно пропонує виняткові можливості для візуального вивчення даних, програмне забезпечення не має функцій для глибокої обробки сигналів, фільтрації або класифікації на основі методів машинного навчання. Ці обмеження означають, що DaViS більше підходить як допоміжний інструмент для візуалізації даних, ніж як самостійне рішення для таких завдань, як ідентифікація видів, придушення шуму чи локалізація джерел. Дослідникам часто доводиться поєднувати його з більш потужними платформами для обробки сигналів, щоб отримати повний аналітичний конвеєр. Крім того, програмне забезпечення, здається, не підтримує автоматизацію роботи з великими наборами даних або пакетну обробку, що може бути недоліком для проектів із великими обсягами інформації. Відсутність вбудованої підтримки інтеграції власних алгоритмів або скриптів також може обмежити його гнучкість для дослідників, яким потрібне більш адаптоване рішення для специфічних дослідницьких завдань. Хоча DaViS пропонує видатні інструменти візуалізації, його орієнтація на графічну експлорацію позначається на аналітичній глибині, що може обмежувати його застосування у більш інтенсивних із точки зору даних чи дослідженнях, заснованих на машинному навчанні.

Отже, DaViS Underwater Acoustic Data Visualization Software є високоефективним інструментом для дослідження й інтерпретації підводних акустичних даних за допомогою передових методів візуалізації. Його



інтуїтивний інтерфейс, інтерактивна генерація спектрограм і можливості багатовимірної візуалізації роблять його безцінним ресурсом для дослідників, які прагнуть глибше зрозуміти підводні звукові ландшафти. Проте відсутність розширених функцій обробки, класифікації та автоматизації обмежує його корисність як комплексного аналітичного інструментарію. DaViS найкраще підходить як допоміжне програмне забезпечення для візуалізації акустичних даних, яке працює в тандемі з більш аналітично потужними платформами, щоб забезпечити повноцінне рішення для підводних акустичних досліджень. Його роль у покращенні інтерпретації складних підводних наборів даних є беззаперечною, але майбутні версії могли б отримати вигоду від інтеграції розширених функцій обробки сигналів і аналітики, щоб задовольнити зростаючі потреби у сфері підводної акустики.

SignalSOFT [22] — це спеціалізоване програмне забезпечення, призначене для аналізу сигналів сонарів та обробки підводних акустичних даних, яке пропонує передові інструменти, спеціально адаптовані для вирішення завдань у підводних середовищах. Воно відзначається здатністю обробляти дані сонарів і акустичні вимірювання довкілля, що робить його особливо придатним для застосувань у морських дослідженнях, військово-морських операціях та екологічному моніторингу. Основна сила SignalSOFT полягає у фокусі на підводній акустиці та розширеному наборі функцій для аналізу й візуалізації сигналів сонарів. Програмне забезпечення забезпечує інструменти для створення детальних спектрограм, застосування передових фільтрів і виділення окремих акустичних подій, що є критично важливим для аналізу підводного шуму та інтерпретації сигналів сонарів.

Однією з ключових особливостей SignalSOFT є його розширені можливості шумозаглушення та фільтрації, які є необхідними в підводних умовах, де сигнали часто маскуються фоновим шумом. Програмне забезпечення підтримує налаштовувані смугові, високочастотні та низькочастотні фільтри, що дозволяє користувачам ізолювати частотні діапазони, які їх цікавлять, і ефективно зменшувати шумові перешкоди. Ці

функції особливо цінні для дослідників і операторів, які працюють із системами сонарів, оскільки вони дозволяють ідентифікувати та виділяти слабкі або приховані сигнали в умовах шуму. Крім того, SignalSOFT пропонує інструменти для вимірювання характеристик сигналу, таких як інтенсивність, спектральний розподіл і часові шаблони, забезпечуючи користувачам детальну інформацію про підводні акустичні явища.

Візуалізаційні можливості SignalSOFT також є значною перевагою. Програмне забезпечення створює високоточні спектрограми та графіки часових рядів, дозволяючи користувачам із високою точністю аналізувати частотні й часові характеристики акустичних сигналів. Інтерактивний графічний інтерфейс забезпечує зручність використання, пропонуючи інструменти для масштабування, прокрутки та фокусування на окремих сегментах акустичних даних. Ці функції особливо корисні для завдань, пов'язаних із виявленням короточасних подій, оцінкою розповсюдження сигналу та візуалізацією акустичних характеристик підводних об'єктів чи явищ. Здатність програмного забезпечення створювати детальні та налаштовувані візуальні уявлення акустичних даних підвищує його корисність як для дослідницьких, так і для практичних завдань.

Попри свої сильні сторони, вузька спеціалізація SignalSOFT на сонарих і екологічних акустичних дослідженнях обмежує його універсальність для ширшого використання в підводній акустиці. Хоча програмне забезпечення чудово підходить для аналізу сигналів сонарів і шуму, йому бракує спеціалізованих функцій для біоакустичних досліджень, таких як інструменти для ідентифікації вокалізацій морських ссавців або аналізу біологічних звукових ландшафтів. Крім того, SignalSOFT не інтегрує сучасних методів машинного навчання, які стають все більш важливими для автоматичної класифікації та ідентифікації джерел сигналів у підводній акустиці. Відсутність функціоналу машинного навчання зменшує адаптивність програмного забезпечення для передових досліджень, у яких необхідні методи, засновані на даних.

Ще одним обмеженням SignalSOFT є його вузька спеціалізація, яка, хоча й корисна для цільових застосувань, знижує гнучкість для загальних досліджень у сфері підводної акустики. Програмне забезпечення в першу чергу розроблене для користувачів, які зосереджені на системах сонарів, що може зробити його менш привабливим для дослідників, які працюють у сфері біоакустики чи інших морських наук. Крім того, відсутність відкритих опцій для налаштування або підтримки інтеграції зовнішніх алгоритмів може бути недоліком для користувачів, яким потрібен більш адаптивний інструмент для специфічних дослідницьких завдань. Хоча SignalSOFT забезпечує потужну функціональність для застосувань, орієнтованих на сонари, його вузька спрямованість і обмежені можливості налаштування можуть обмежувати його корисність для дослідників, які шукають більш комплексне програмне рішення.

Таким чином, SignalSOFT є потужним і високоспеціалізованим інструментом для аналізу сигналів сонарів і підводної акустичної обробки. Його розширені можливості фільтрації шуму, детальні візуалізації та орієнтація на сонарну акустику роблять його відмінним вибором для застосувань у військово-морських операціях, екологічному моніторингу та дослідженнях сонарів. Однак його вузька спеціалізація та відсутність функцій для біоакустики чи машинного навчання обмежують його застосовність для ширших досліджень підводної акустики. SignalSOFT найкраще підходить для користувачів із цільовими потребами в аналізі сонарів, проте дослідникам в інших галузях, ймовірно, доведеться доповнювати його іншими програмами для задоволення своїх специфічних вимог. У міру розвитку досліджень у сфері підводної акустики інтеграція машинного навчання та розширення універсальності програмного забезпечення могли б значно покращити корисність SignalSOFT у ширшому спектрі застосувань.

#### **1.4. Аналіз сучасних наукових методів аналізу гідроакустичних сигналів**

Статистичні та ймовірнісні моделі відіграють важливу роль в аналізі підводних акустичних сигналів через змінність та часто шумність морського середовища. Ці моделі створені для отримання корисної інформації з комплексних наборів даних, що дозволяє застосовувати їх у таких задачах, як класифікація сигналів, локалізація джерел звуку та моніторинг навколишнього середовища.

Авторегресійні (AR) моделі є основоположними у аналізі часових рядів і часто використовуються для моделювання підводних акустичних сигналів. Ці моделі виражають поточне значення сигналу як лінійну комбінацію його попередніх значень із додаванням похибки. Такий підхід ефективно описує часову динаміку сигналів, що є цінним для задач, як-от прогнозування майбутньої поведінки сигналу та виявлення шаблонів у акустичних даних. Наприклад, AR-моделі широко застосовуються в обробці сигналів сонарів для прогнозування спектральних характеристик звукових хвиль, що сприяє виявленню та ідентифікації підводних об'єктів. Їх простота та ефективність роблять їх популярним вибором для реального часу в підводній акустиці [23].

Приховані марковські моделі (НММ) є ще одним важливим інструментом в аналізі підводних акустичних сигналів. НММ – це ймовірнісні моделі, які передбачають, що система переходить між скінченою кількістю прихованих станів, кожен із яких пов'язаний із ймовірнісним розподілом спостережуваних акустичних сигналів. Ця характеристика робить НММ особливо придатними для задач класифікації, таких як ідентифікація морських видів за їх вокалізаціями або виявлення певних типів суден за їх акустичними підписами. Темпоральна природа НММ дозволяє їм враховувати послідовні залежності в акустичних даних, забезпечуючи надійну продуктивність навіть у шумних умовах. Дослідження демонструють

ефективність НММ у класифікації викликів морських ссавців і розрізненні біологічних та антропогенних звуків у підводному середовищі [24].

Гаусівські змішані моделі (GMM) широко використовуються для кластеризації та класифікації в підводній акустиці. Ці моделі представляють розподіл ймовірностей даних як суміш кількох гаусівських компонентів, кожна з яких відповідає потенційному класу або кластеру. У підводному середовищі GMM застосовуються для моделювання розподілу джерел шуму або класифікації акустичних сигналів на окремі категорії. Наприклад, GMM успішно використовуються для розрізнення фонових шумів і сигналів, згенерованих морськими суднами або біологічними джерелами. Адаптивність GMM до змінного акустичного середовища та їх здатність обробляти перекриваючі розподіли роблять їх потужним інструментом для аналізу підводних сигналів [25].

Крім цих моделей, байєсівські методи стали значущим підходом у підводній акустиці. Байєсівські методи оцінювання враховують попередні знання про середовище та статистичні властивості сигналів для здійснення ймовірнісних висновків. Ці методи особливо корисні для локалізації акустичних джерел, де вони враховують невизначеності в параметрах середовища та похибки вимірювань. Байєсівські моделі застосовувалися для локалізації джерел звуку в умовах мілководдя, досягаючи високої точності навіть за складних умов [26].

Ще одним інноваційним підходом є використання ймовірнісних графічних моделей (PGM), які представляють складні взаємозв'язки між змінними за допомогою графової структури. PGM застосовуються в підводній акустиці для таких задач, як виявлення несправностей у сонарних системах і моніторинг навколишнього середовища. Моделюючи залежності між змінними, ці моделі дозволяють здійснювати всебічний аналіз і прийняття рішень у невизначених середовищах. Дослідження підкреслюють потенціал

PGM для підвищення надійності та ефективності підводних акустичних систем [27].

Фільтри Калмана та частинкові фільтри є додатковими ймовірнісними інструментами, що застосовуються в підводній акустиці. Фільтри Калмана є оптимальними для лінійних систем із гаусівським шумом і широко використовуються для відстеження підводних апаратів або оцінювання параметрів середовища. Частинкові фільтри, зі свого боку, розширюють рамки фільтрів Калмана до нелінійних та негаусівських систем, що робить їх придатними для складних підводних сценаріїв. Ці фільтри відіграють ключову роль у таких додатках, як відстеження цілей та покращення сигналів у реальному часі в шумних підводних середовищах [28].

Машинне навчання - це технологія, в якій машина (комп'ютер) використовує дані для автоматичного виявлення і навіть прогнозування прихованих характеристики або закономірності [29]. Машинне навчання у задачах аналізу акустичних сигнатур водного середовища є потужним інструментом, що дає змогу обробляти великі масиви даних, виявляти приховані закономірності та робити прогнози зі значно вищою точністю, ніж за допомогою традиційних методів. Насамперед оцифровані сигнали підлягають фільтрації від шуму та перетворенню у зручний формат, наприклад у вигляді спектрограм. Далі з отриманих даних виділяють найінформативніші ознаки, які передаються на вхід алгоритмам машинного навчання (нейронним мережам, методам класифікації чи регресії). Після навчання та тестування моделі стають здатними виявляти або розрізняти специфічні події, такі як звуки різних морських мешканців, шум від суден чи підводні вибухи, а також відстежувати довготривалі зміни у підводному акустичному фоні. Це суттєво підвищує ефективність досліджень, адже зменшує затрати часу й ресурсів, одночасно покращуючи якість та точність аналізу.

### 1.5. Огляд методів машинного навчання, що можуть бути застосовані для аналізу акустичних сигналів водного середовища

Метод опорних векторів (SVM) — це метод класифікації, який належить до групи граничних методів і визначає класи за допомогою меж просторів. Опорними векторами є об'єкти множини, що лежать на цих межах. Класифікація вважається вдалою, якщо простір між межами порожній.

Метод опорних векторів — це група алгоритмів, що застосовуються для розв'язання завдань регресійного аналізу та класифікації. Він належить до алгоритмів навчання з учителем.

Характерною особливістю методу опорних векторів є мінімізація емпіричної помилки класифікатора та максимізація проміжку між класами. Також цей метод називають класифікатором із максимальним проміжком.

Суть методу полягає в перетворенні вхідних векторів у простір вищої розмірності та пошуку роздільної гіперплощини з максимальним проміжком. Відбувається побудова двох паралельних гіперплощин по обидва боки гіперплощини, що розділяє досліджувані класи. Візуалізацію роботи методу опорних векторів можна побачити на рис. 1.1.

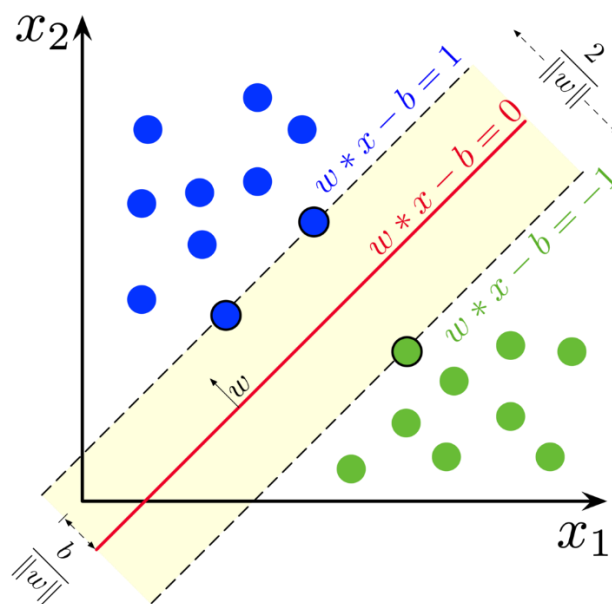


Рис. 1.1. Візуалізація роботи методу опорних векторів [31].

В якості роздільної гіперплощини буде обрана така площина, що максимізує відстань до паралельних гіперплощин.

Для цього алгоритму правильним буде припущення, що чим більша відстань між паралельними площинами, тим меншою буде середня помилка класифікатора [32].

Метод опорних векторів нестійкий до шуму у вхідних даних та чутливий до стандартизації даних. Якщо навчальна вибірка містить аномальні значення (викиди), його використання може бути нераціональним.

SVM добре працює з нелінійними залежностями без перенавчання, досягаючи високої точності та узагальнення у завданнях розпізнавання [33]. Це один із найшвидших алгоритмів знаходження оптимальних вирішальних функцій.

Метод розв'язує задачу квадратичного програмування у просторах, де рішення завжди єдине. Він виявляє роздільну площину максимальної ширини, що надалі дозволяє точніше класифікувати об'єкти. Метод опорних векторів був застосований у роботі [34] і показав точність 89,5%.

Дерево прийняття рішень, яке також може називатися деревом класифікацій або регресійним деревом — це засіб підтримки прийняття рішень, що використовується в статистиці та аналізі даних для побудови прогнозних моделей. Він допомагає у розв'язанні завдань класифікації та регресії.

Дерева прийняття рішень є простими аналітичними моделями, оскільки правила генеруються природною мовою. Наприклад: «Якщо реклама залучила 1000 клієнтів, то вона добре налаштована». Правила формуються шляхом узагальнення множини окремих спостережень (навчальних прикладів), що описують предметну область. Тому їх називають індуктивними правилами, а сам процес навчання — індукцією дерев рішень [35].



У навчальній множині для прикладів має бути задано цільове значення, так, як дерева рішень — моделі, створені на основі навчання з учителем. За типом цільової змінної виділяють два типи дерев: дерево класифікації — коли цільова змінна дискретна та дерево регресії — коли цільова змінна безперервна.

Структура дерева починається з кореневого вузла, який не має жодних вхідних гілок. Гілки, що виходять з кореневого вузла, потрапляють до внутрішніх вузлів, відомих як вузли прийняття рішень. На основі існуючих ознак ці вузли виконують оцінки і формують однорідні підмножини, які позначаються листовими вузлами, що представляють усі можливі результати набору даних.

Простий випадок: приклади потрапляють у вузол, проходять перевірку і розбиваються на дві підмножини:

- перша — ті, які задовольняють встановленому правилу;
- друга — ті, які не задовольняють встановленому правилу.

Далі до кожної підмножини знову застосовується правило, і процедура повторюється. Це триває до тих пір, поки не буде досягнута умова зупинки алгоритму. Останній вузол, до якого не застосовується перевірка та розбиття, стає листом. Відображення дерева прийняття рішень представлено на рис. 1.2.

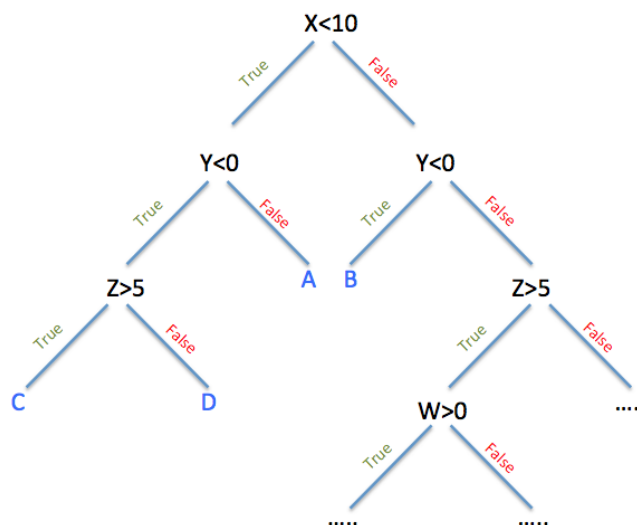


Рис. 1.2. Дерево прийняття рішень [36]

Лист визначає рішення для кожного вхідного прикладу. Для дерева класифікації лист означає клас, що асоціюється з вузлом, а для дерева регресії — містить прогнозне значення (чи інтервал) цільової змінної. У листку міститься не правило, а підмножина об'єктів, які відповідають усім правилам гілки, що до нього веде. Приклад потрапляє до листка, якщо відповідає всім правилам на шляху до нього. До кожного листка веде лише один шлях, що забезпечує єдність розв'язку [37].

Даний метод є простим і зрозумілим, не потребує підготовки вхідних даних і надає можливість оцінити надійність моделі з використанням тестів. При застосуванні для ідентифікації гідроакустичних сигналів метод показує досить високі показники повноти та точності результату. Однак серед недоліків алгоритму можна віднести складність інтерпретації деяких концептів. Ще одним мінусом є необхідність перенавчати модель при значному розростанні розмірів дерева. Метод дерева прийняття рішень був застосований у роботі [34] і показав точність 86.8%.

Нейромережа — це обчислювальна система, влаштована за принципом, близьким до роботи мозку тварини. Нейромережі використовуються для вирішення ряду задач, таких як розпізнавання мовлення та зображень. Класична мережа використовує тришарову структуру прямого зв'язку для вирішення завдань класифікації.

Інформація потрапляє в нейронну мережу через вхідний рівень, який є первинним зовнішнім шаром. Кінцевим рівнем, через який проходить інформація, є вихідний рівень. Вхідний та вихідний рівні можуть мати або не мати додаткових шарів між ними. Шари, якщо вони є, що розташовані між вхідним і вихідним шарами, називаються прихованими шарами. Нейронна мережа вважається «глибокою» нейронною мережею, якщо вона має кілька прихованих шарів. Як правило, кожен нейрон будь-якого шару взаємопов'язаний з усіма нейронами сусідніх шарів. Кожен рівень нейронної мережі перетворює вхідні дані в більш просту форму для подальшої

інтерпретації та класифікації вмісту. Велика кількість шарів означає, що може існувати більше шляхів для переміщення інформації через мережу, що потенційно дозволяє виконувати досить складні завдання. Приклад тришарової архітектури нейромережі представлено на рис. 1.3.

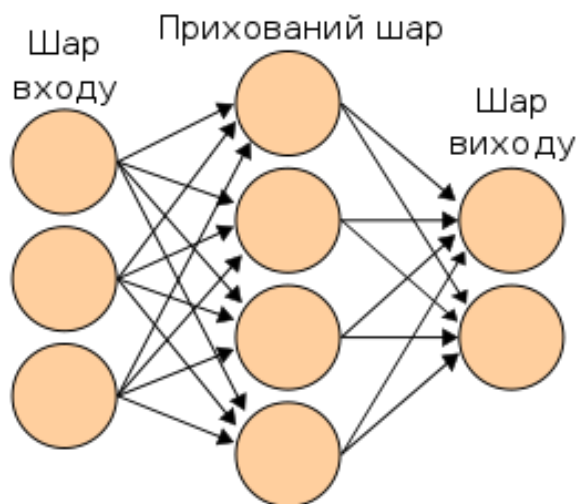


Рис. 1.3. Тришарова архітектура нейромережі [38]

У мережах глибокого навчання кожен рівень вузлів тренується за окремим набором функцій на основі результатів попереднього рівня. Чим глибше інформація проходить через мережу, тим складніші ознаки можуть розпізнаватися вузлами, оскільки вони узагальнюють функції попередніх шарів. Один із основних алгоритмів, що використовується для навчання нейромережі, — це алгоритм зворотного поширення помилки. Ідея цього методу полягає в тому, що сигнал помилки поширюється від виходів нейромережі до її входів, тобто у зворотному напрямку. Його алгоритм являє собою ітеративний алгоритм градієнтного спуску, який використовується для мінімізації помилки роботи нейромережі шляхом послідовної корекції ваг нейронів. На кожній ітерації вагові коефіцієнти змінюються таким чином, аби покращити виконання поточного завдання

Однак алгоритм зворотного поширення помилки має два недоліки. Перший — процес навчання проходить дуже повільно. Другий — наявність

локальних мінімумів. Причини цього полягають у складній нелінійній функції втрат у ваговому просторі та відсутності оптимальності при виборі швидкості навчання, що пов'язано з різними ваговими коефіцієнтами. Для алгоритму зворотного поширення існує багато модифікацій. Нейронні мережі були застосовані в роботі [40] та показали точність в межах від 95% до 98%.

Можна зробити висновок, що нейронні мережі мають низку переваг, що роблять їх доцільним вибором для аналізу акустичних сигналів водного середовища. Вони здатні апроксимувати складні нелінійні залежності в даних, що є особливо важливим для обробки сигналів із шумами, реверберацією та змінами у частотних компонентах. Порівняно з деревами прийняття рішень та методом опорних векторів, нейронні мережі демонструють кращу ефективність при роботі з великими багатовимірними наборами даних, типовими для аналізу спектральних і часово-частотних характеристик.

Крім того, вони відзначаються здатністю до роботи з великими обсягами інформації. На відміну від методу опорних векторів, де можлива проблема перенавчання на великих наборах даних, нейронні мережі при збільшенні обсягу даних лише покращують точність аналізу, що є великою перевагою для обробки гідроакустичних сигналів. Використання спеціалізованих архітектур, таких як згорткові нейронні мережі для аналізу спектрограм або рекурентні мережі для часових рядів, дозволяє враховувати часово-частотні закономірності, що є ключовими для таких задач.

Завдяки багат шаровій структурі, нейронні мережі забезпечують високу стійкість до зашумлених даних, тоді як традиційні алгоритми часто демонструють зниження ефективності в таких умовах. Вони також легко інтегруються з сучасними методами, наприклад, із застосуванням CQT або використанням кепстральних коефіцієнтів, що дозволяє отримувати якісніші ознаки для класифікації. Їх потенціал досягнення високої точності є значним: як зазначено в статті, вони демонструють точність 95–98%, що перевищує показники SVM та дерев прийняття рішень. Нейронні мережі також

дозволяють створювати налаштовувані архітектури, оптимізовані для специфічних задач. Наприклад, використання мульти-масштабних згорток або механізмів уваги дозволяє адаптувати модель під конкретні особливості сигналів. Завдяки універсальності, гнучкості та високій продуктивності, нейронні мережі є найкращим вибором для задач аналізу акустичних сигналів водного середовища, особливо за умови високої складності даних, великих обсягів інформації та високих вимог до точності.

#### **1.6. Дослідження різних нейромережевих алгоритмів з метою використання для аналізу акустичних сигналів морського середовища з урахуванням його особливостей**

Зазвичай нейрони організовані в кілька шарів, особливо в глибокому навчанні. Нейрони одного шару з'єднані лише з нейронами безпосередньо попереднього та безпосередньо наступного шарів. Шар, що отримує зовнішні дані, називається вхідним шаром. Шар, що формує остаточний результат, — вихідним шаром. Між ними може бути нуль або більше прихованих шарів. Між двома шарами можливі різні схеми з'єднань. Вони можуть бути повністю з'єднані, коли кожен нейрон в одному шарі підключений до кожного нейрона в наступному шарі. Інша схема передбачає, що група нейронів в одному шарі з'єднується з одним нейроном у наступному шарі, що зменшує загальну кількість нейронів у цьому шарі. Нейрони, які мають лише такі зв'язки, утворюють орієнтований ациклічний граф і формують так звані мережі прямого поширення. Багат шаровий персептрон є популярним рішенням у машинному навчанні та застосовується для розв'язання завдань у різних галузях, таких як розпізнавання мовлення, розпізнавання зображень та машинний переклад.

### 1.6.1. Екстремальна машина самонавчання

Екстремальна машина навчання (ELM) — це нейронна мережа з прямим поширенням, що має один або кілька шарів прихованих вузлів. Приховані вузли можуть бути призначені випадковим чином і ніколи не оновлюватися або успадковуватися від попередніх моделей без змін. У більшості випадків ваги прихованих вузлів обчислюються за один крок, що, по суті, призводить до надзвичайно швидкого процесу навчання. За словами розробників, ці моделі здатні забезпечувати високу узагальнюючу здатність і навчатися в тисячі разів швидше, ніж мережі із зворотним поширенням помилки. Результати досліджень також показують, що такі моделі можуть перевершувати машини опорних векторів (SVM) у задачах класифікації та регресії.

ELM — це швидкий метод навчання мереж із прямим поширенням та одним прихованим шаром. Вхідний шар приймає характеристики даних, але не виконує обчислень, тоді як вихідний шар є лінійним і не містить функції активації або зсуву. Метод ELM встановлює ваги вхідного шару  $w$  та зсуви  $b$  випадковим чином і ніколи їх не коригує. Оскільки вхідні ваги фіксовані, вихідні ваги залежать лише від вихідних даних (на відміну від методу навчання зворотного поширення помилки). Це дозволяє отримати аналітичне рішення, яке не потребує ітерацій.

Таке рішення також є лінійним і обчислюється дуже швидко для лінійного вихідного шару. Випадкові ваги вхідного шару покращують узагальнюючі властивості моделі, оскільки вони забезпечують практично ортогональні (слабо корельовані) ознаки прихованого шару. Рішення лінійної системи завжди знаходиться в діапазоні вхідних ваг. Якщо діапазон ваг рішення обмежений, ортогональні вхідні ваги забезпечують більший об'єм простору рішень у цих межах.

Менші норми ваг зазвичай роблять систему стійкішою до шуму, оскільки помилки на вході не підсилюються на виході лінійної системи з малими коефіцієнтами. Як наслідок, випадковий прихований шар створює слабо корельовані ознаки, що дозволяє отримати рішення з низькою нормою та високою узагальнюючою здатністю.

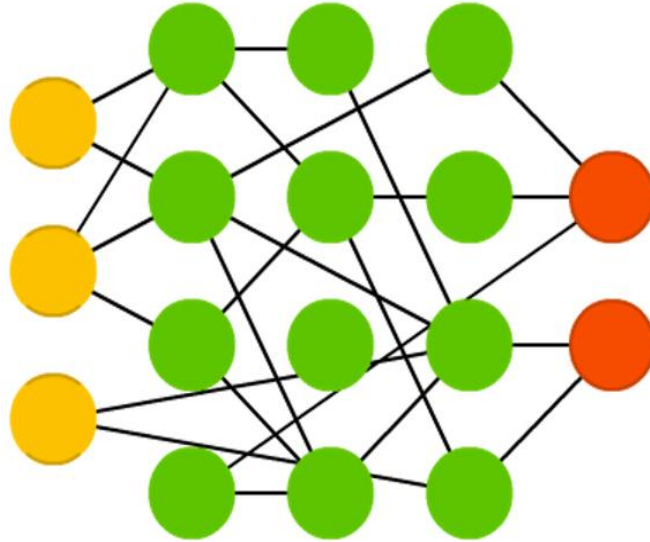


Рис. 1.4. Модель екстремальної машини навчання [41]

Області застосування ELM включають класифікацію та регресійні задачі. Крім того, ELM також успішно використовується для розпізнавання образів, прогнозування, діагностики, обробки зображень та в інших галузях[42].

### 1.6.2. Згорткові нейронні мережі (CNN)

Згорткові нейронні мережі відрізняються від інших нейронних мереж високою продуктивністю в обробці зображень, мовлення або аудіосигналів. Вони складаються з трьох основних типів шарів: згорткового шару, шару підгрупування та повнозв'язного шару. Згортковий шар — це перший шар у згортковій мережі. Після згорткових шарів можуть іти додаткові згорткові шари або шари підгрупування, а останнім є повнозв'язний шар. З кожним наступним шаром збільшується складність згорткової нейронної мережі, і

вона ідентифікує все більш складні частини зображення. Попередні шари фокусуються на простих характеристиках, таких як кольори та контури.

У міру проходження зображення через згорткові шари мережа починає розпізнавати складніші елементи або форми об'єкта, поки зрештою не ідентифікує цільовий об'єкт. На рис. 1.5 представлено модель згорткової нейронної мережі.

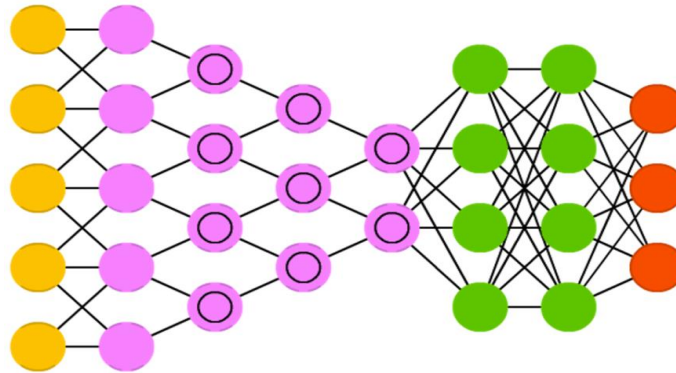


Рис 1.5. Модель згорткової нейронної мережі [41]

Блок 1D-згортки є шаром, який використовується для виявлення особливостей у векторі. Такий блок складається з певної кількості фільтрів, де кожен фільтр має заданий розмір. Операція згортки виконується між вхідним вектором і фільтром, у результаті чого на виході формується новий вектор із кількістю каналів, що дорівнює кількості фільтрів. Кожне значення у тензорі потім проходить через функцію активації, яка додає нелінійність. Зазвичай фільтри переміщуються по ширині вектора на один елемент за раз — це називається горизонтальним кроком, і користувач може змінювати цей параметр. Чим більший крок, тим менший вихідний вектор. Це можна використати для зменшення кількості параметрів і пам'яті, але це також призводить до втрати інформації. За допомогою підшивки (padding) можна додати один або кілька нульових пікселів навколо вхідного вектора, щоб збільшити його ефективний розмір. Згорткові шари за замовчуванням зменшують розмірність тензора. Якщо багато згорткових шарів додаються один до одного, вихідний тензор поступово зменшується, доки не стане надто



малим для подальшого використання. Щоб запобігти цьому, можна додати підшивку перед згортковим шаром, щоб зберегти розмір тензора. Зазначену архітектуру можна застосовувати для автоматичного розпізнавання мовлення, моніторингу електрокардіограм у реальному часі та виявлення структурних пошкоджень у цивільній інфраструктурі на основі вібрацій [43].

Wavenet здатний синтезувати природно звучне мовлення. Ця авторегресійна модель формує ймовірнісний розподіл наступного зразка, враховуючи певний фрагмент попередніх зразків. Вся послідовність зразків генерується шляхом послідовної подачі раніше згенерованих зразків назад у модель — це забезпечує часову неперервність у результируючій хвильовій формі звукових сигналів. Wavenet є адаптацією генеративної моделі PixelCNN для зображень, перенесеною в аудіосферу. Він зберігає багато особливостей PixelCNN, зокрема каузальність, згорткові блоки та дискретний розподіл виходів softmax. Оскільки висока часова складність Wavenet була основним обмеженням, у запропонованій моделі її авторегресійний характер було усунуто.

Логічним продовженням асиметричних розширених згорток Wavenet є збільшення довжини фільтра до 3 і виконання симетричного заповнення в кожному розширеному шарі. Якщо зразок, який потрібно покращити, знаходиться в центрі рецептивного поля, це подвоює контекст навколо нього та усуває каузальний зв'язок. Модель отримує однаковий доступ до зразків у минулому та майбутньому, що покращує точність прогнозу. Замість цього багатомодальний розподіл вихідного сигналу може вносити артефакти у відфільтрований сигнал. Це дозволяє припустити, що унімодальні прогнози (з Гаусовим розподілом вихідного сигналу) є більш придатними для цієї задачі. З цих причин запропонована модель передбачає необроблений звук — без попередньої обробки. Вихід моделі не є самим сигналом, а його ймовірнісним розподілом. Крім того, модель навчається шляхом мінімізації функції втрат регресії. Як наслідок, ця модель є не генеративною, а дискримінаційною. Її

можна використовувати для розпізнавання мовлення або зменшення шуму в аудіосигналі [44].

### **1.6.3. Рекурентні нейромережі**

Рекурентна нейронна мережа (RNN) — це тип штучних нейронних мереж, що працює з послідовними або часовими даними. RNN застосовуються для вирішення завдань, пов'язаних із послідовними або часовими залежностями, таких як машинний переклад, обробка природної мови, розпізнавання мовлення та генерація підписів до зображень. Їх використовують у популярних додатках, таких як Siri, голосовий пошук і Google Translate. Рекурентні нейронні мережі, так само як мережі прямого поширення та згорткові нейронні мережі, навчаються на основі навчальних даних. Основною відмінністю RNN є наявність пам'яті, оскільки вони враховують інформацію з попередніх входів, що впливає на поточний вхід і вихід. Якщо традиційні глибокі нейронні мережі припускають, що входи та виходи незалежні один від одного, то вихід RNN залежить від попередніх елементів у послідовності. Хоча майбутні події також можуть бути корисними для прогнозування, однонаправлені рекурентні нейронні мережі не враховують майбутню інформацію у своїх прогнозах.

### **1.6.4. Нейромережа довгої короткострокової пам'яті (LSTM)**

Стандартні рекурентні нейронні мережі (RNN) відстежують довгострокові залежності у вихідних послідовностях. Однак проблема стандартних RNN полягає в тому, що під час зворотного поширення градієнти можуть або наближатися до нуля, або зростати до нескінченності через використання чисел з обмеженою точністю в обчисленнях. Рекурентні нейронні мережі, що використовують блоки LSTM, частково вирішують

проблему зникнення градієнта, оскільки блоки LSTM дозволяють градієнтам ефективно передаватися без значних змін.

Загальна архітектура LSTM-мережі містить основний осередок (cell), який є частиною модуля пам'яті, а також три регулятори потоку інформації, що називаються "воротами". В модулі LSTM існують три типи воріт:

- Вхідні ворота (input gate) — контролюють, яка частина інформації з вхідних даних буде збережена в пам'яті.
- Вихідні ворота (output gate) — визначають, яка частина інформації впливатиме на прогнозоване значення.
- Ворота забування (forget gate) — регулюють ступінь, з яким попередня інформація залишається в осередку пам'яті.

Основна функція осередку (cell) — зберігати залежності між елементами вхідної послідовності. LSTM-мережі застосовуються в багатьох галузях, серед яких розпізнавання мовлення [45]. На рис. 1.6 представлено архітектуру нейронної мережі LSTM.

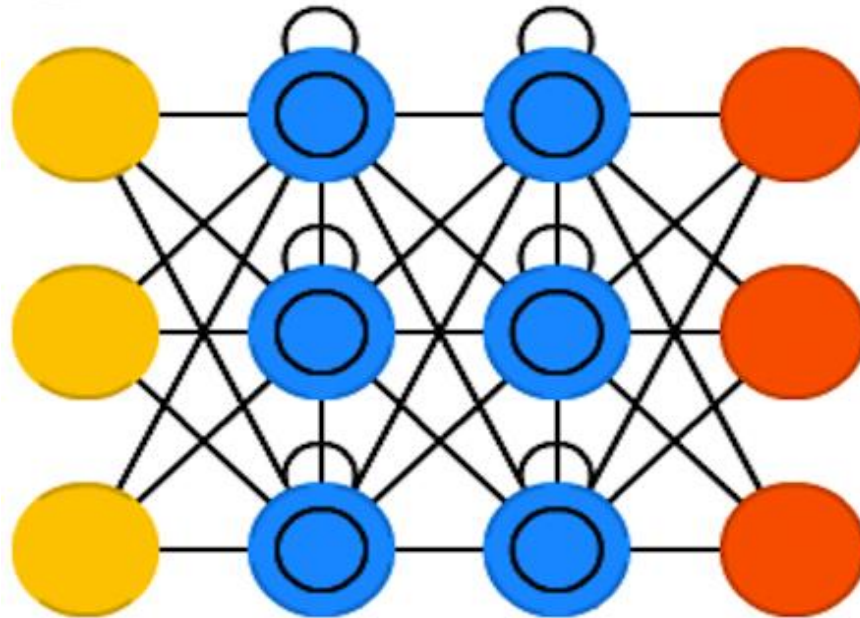


Рис. 1.6. Архітектури нейронної мережі LSTM [41]

### 1.6.5. Машина Больцмана

Машина Больцмана є однією з різновидів стохастичних рекурентних нейронних мереж. Її можна розглядати як стохастичний генеративний варіант мережі Хопфілда. Ця мережа використовує алгоритм імітації відпалу для навчання і стала першою нейронною мережею, здатною навчати внутрішні представлення та розв'язувати складні комбінаторні задачі. В архітектурі машини Больцмана немає чіткого розмежування між вхідним і вихідним шарами. Фактично, вихідний шар відсутній.

Вузли в машинах Больцмана поділяються на видимі та приховані. Видимі вузли приймають вхідні дані, тоді як ті ж вузли, що отримують вхід, повертають реконструйований вхід як вихід. Це здійснюється за рахунок двонаправлених ваг, які поширюються у зворотному напрямку і формують вихід на видимих вузлах. Важливою перевагою є те, що кожен вузол з'єднаний з усіма іншими вузлами, навіть усередині одного шару (наприклад, кожен видимий вузол з'єднаний з іншими видимими вузлами, а також із прихованими вузлами). Усі зв'язки є двонаправленими, а ваги симетричні.

Навіть якщо навчання в машинах Больцмана загалом є непрактичним, його можна зробити ефективнішим у обмеженій машині Больцмана (RBM), яка не дозволяє внутрішніх зв'язків між прихованими та видимими блоками. Після навчання одного RBM, активації його прихованих вузлів можна використовувати як вхідні дані для навчання RBM вищого рівня. Цей метод об'єднання RBM дозволяє ефективно навчати кілька рівнів прихованих модулів і є одним із найпоширеніших у глибокому навчанні. Один із прикладів практичного використання RBM — це розпізнавання мовлення, зменшення розмірності даних [46], класифікація [47], вивчення ознак [48]. На рис. 1.7 представлено архітектуру нейронної мережі машини Больцмана.

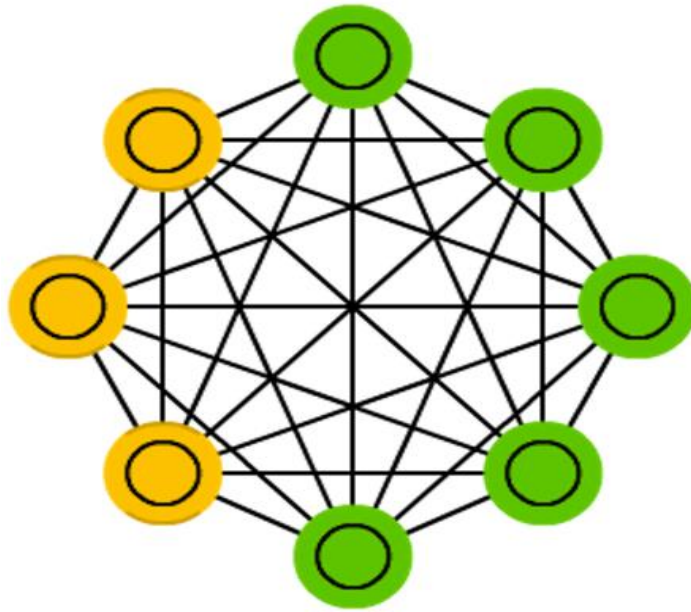


Рис. 1.7. Модель машини Больцмана [41]

#### 1.6.6. Мережа глибокої довіри

Мережі глибокої довіри (DBN) — це ймовірнісні генеративні моделі, які складаються з кількох шарів стохастичних прихованих змінних. Приховані змінні можуть мати двійкові значення і часто називаються прихованими вузлами або детекторами ознак. Два верхні шари мають ненаправлені, симетричні зв'язки між собою і формують асоціативну пам'ять. Між рештою пар сусідніх шарів є лише направлені зв'язки від верхнього до нижнього. Стани вузлів у нижньому шарі представляють вектори даних.

Два найважливіших властивості мереж глибокої довіри:

- Існує ефективна послойова процедура для налаштування ваг низхідних шарів, яка визначає, як змінні в одному шарі залежать від змінних в верхньому шарі.
- Після навчання прихованих змінних, в кожному шарі можуть бути виведені значення за один прохід знизу вгору, який розпочинається зі спостережуваного вектора даних в нижньому шарі і використовує ваги в зворотньому напрямку.

Мережі глибокої довіри навчаються по одному шару за раз, обробляючи значення прихованих змінних в одному шарі в момент, коли вони виводяться з даних для навчання наступного шару. Це ефективне, жадібне навчання може супроводжуватися або поєднуватися з іншими процедурами навчання, які точніше налаштовують всі ваги для покращення генеративних характеристик всієї мережі.

Дискримінативне тонке налаштування можна виконати шляхом додавання останнього шару змінних, які представляють бажані вихідні дані, а також використання градієнтів помилок зворотного поширення. Коли мережі з багатьма прихованими шарами застосовуються до високоструктурованих вхідних даних, таких як зображення, зворотне поширення градієнтів працює набагато краще, якщо детектори ознак у прихованих шарах ініціалізуються за допомогою навчання мережі глибокої довіри, що моделює структуру вхідних даних. Мережу глибокої довіри можна використовувати для розв'язання завдань від класифікації зображень і розпізнавання мовлення до класифікації аудіо [49]. На рис. 1.8 представлено архітектуру мережі глибокої довіри.

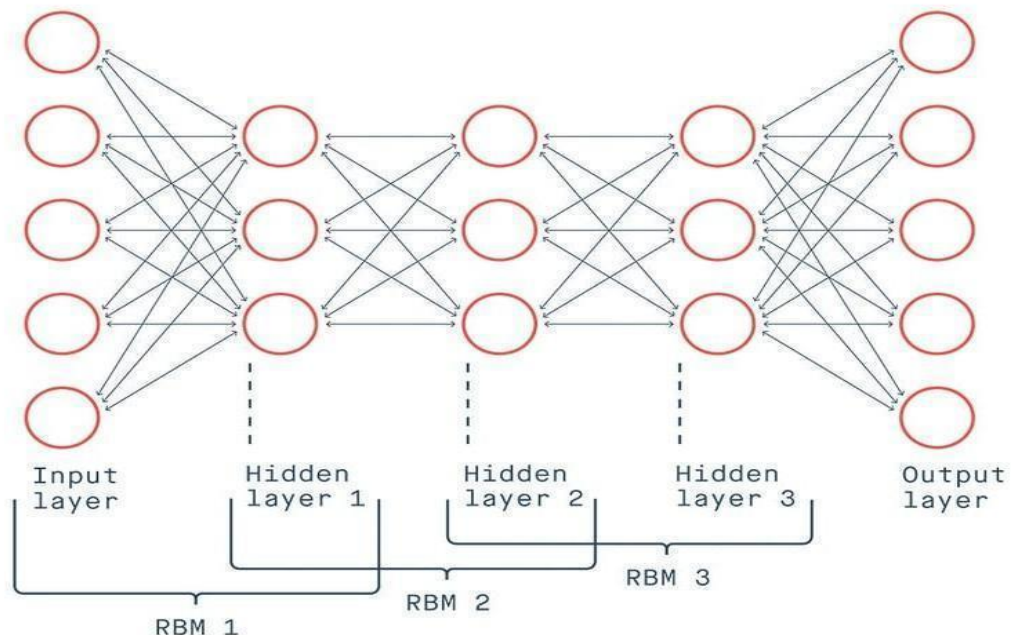


Рис. 1.8. Архітектура мережі глибокої довіри [49]

### 1.6.7. Автоенкодер

Автоенкодери — це нейронні мережі прямого поширення, які відтворюють вхідний сигнал на виході. Вони мають прихований шар, що представляє собою код, який описує модель. Автоенкодери створюються так, щоб не мати можливості точно відтворювати вхід на виході. Зазвичай їх обмежують у розмірності коду (вона менша, ніж розмірність сигналу) або штрафують за надмірну активацію нейронів у коді. Через втрати під час кодування вхідний сигнал відтворюється з помилками, але для їх мінімізації мережа навчається виділяти найважливіші ознаки. Автоенкодер проводить апроксимацію даних, зосереджуючи увагу на основних характеристиках вхідного сигналу. Він визначається через вхідний шар, внутрішнє представлення та вихідний шар.

Розріджені автоенкодери пропонують альтернативний спосіб введення інформаційного "вузького місця", не вимагаючи зменшення кількості нейронів у прихованих шарах. Замість цього функція втрат модифікується так, щоб штрафувати надмірні активації нейронів у прихованому шарі. Для кожного спостереження мережа навчається кодуванню та декодуванню таким чином, щоб активувалася лише невелика кількість нейронів. Це відрізняється від традиційної регуляризації, оскільки зазвичай регуляризації піддають ваги мережі, а не активації нейронів. На рисунку 1.9 представлено архітектуру розрідженого автоенкодера.

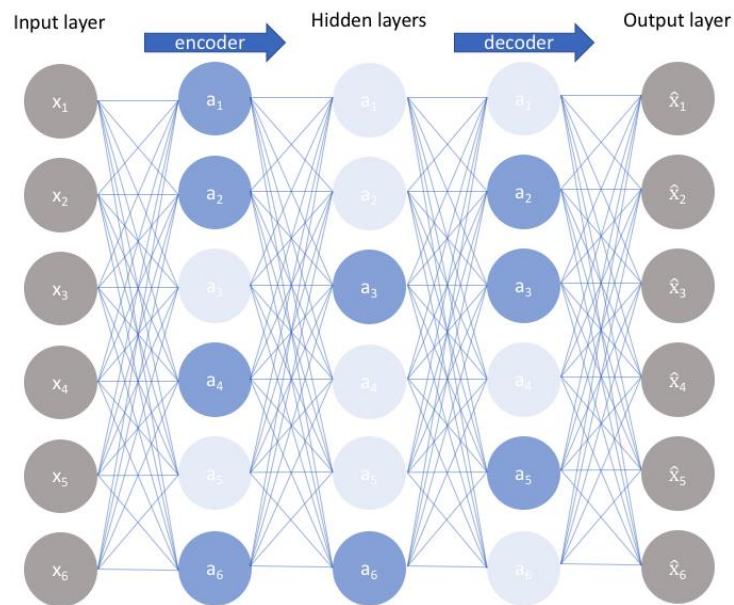


Рис. 1.9. Архітектура розрідженого автоенкодера [50]

Активація окремих вузлів навченої моделі залежить від вхідних даних: різні вхідні дані призводять до активації різних вузлів мережі. Одним із наслідків цього є те, що мережа чутливо реагує на окремі вузли прихованого шару, відповідно до атрибутів вхідних даних. Неповний автоенкодер використовує всю мережу для кожного спостереження. Розріджений автоенкодер вибірково активує різні області мережі залежно від вхідних даних. В результаті обмежується здатність мережі запам'ятовувати вхідні дані, але зберігається можливість ефективного вилучення ознак. Це дозволяє розглядати представлення латентного простору та регуляризацію мережі окремо. Таким чином, вибір представлення латентного простору (тобто розміру кодування) узгоджується зі змістом даних, одночасно застосовуючи регуляризацію шляхом введення обмеження розрідженості.

Існують два основних способи, за допомогою яких можна накласти обмеження розрідженості. Обидва підходи включають вимірювання активацій прихованого шару для кожної навчальної партії та модифікацію функції втрат для штрафування надмірних активацій. Представлення даних формується так, щоб сприяти розрідженості, що покращує продуктивність при виконанні завдань класифікації [51].



Шумопридушуючий автоенкодер додає шум до вхідних даних і навчається його усувати. Таким чином, вдається уникнути простого копіювання вхідних даних у їхній незмінній формі та замість цього вивчати їхні особливості. Ці автоенкодері приймають частково пошкоджений вхідний сигнал і навчаються відновлювати початковий непошкоджений сигнал. Модель вивчає векторне поле, яке відображає вхідні дані у простір меншої розмірності, що описує природні дані та допомагає усувати доданий шум. Таким чином, кодер виокремлює найважливіші характеристики та навчається створювати більш стійке представлення даних. На практиці основною метою шумоподавальних автоенкодерів є очищення пошкоджених вхідних даних або зменшення шуму.

У роботі [52] шумопридушуючий автоенкодер використовувався для класифікації зображень та аудіоданих. У роботі [53] автоенкодер застосовувався для відновлення пошкоджених сигналів стану транспортного засобу. На рисунку 1.10 зображено модель шумопридушуючого автоенкодера.

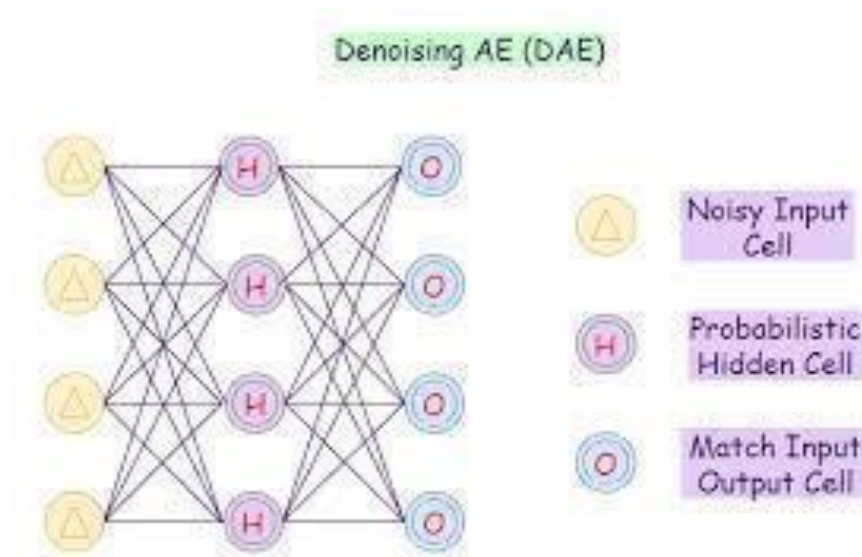


Рис. 1.10. Модель шумоподавального автоенкодера [54]

## 1.7. Огляд нейромережових підходів для обробки гідроакустичних сигналів

### 1.7.1. Згорткові нейромережі

У роботі [55] була запропонована нейромережева архітектура UATC-DensNet (Underwater Acoustic Signal Recognition Based on Dense Convolutional Neural Network), яка базується на використанні щільної згорткової нейромережі (Dense CNN). Архітектура нейромережі розроблена для повторного використання попередніх карт ознак, що дозволяє уникнути затухання градієнта. Наявність шарів зі з'єднанням "skip-connection" дає змогу використовувати всі попередні карти ознак і уникнути проблеми затухання градієнта, яка виникає при накладенні багатьох послідовних шарів згорткової обробки та активації. Крім того, класифікатор на основі CNN потребує менше обчислювальних ресурсів завдяки адаптивній структурі, яка зменшує кількість навчальних параметрів. На вхід подається представлення аудіосигналу у часовій області. Архітектура запропонованого методу зображена на рисунку 1.11.

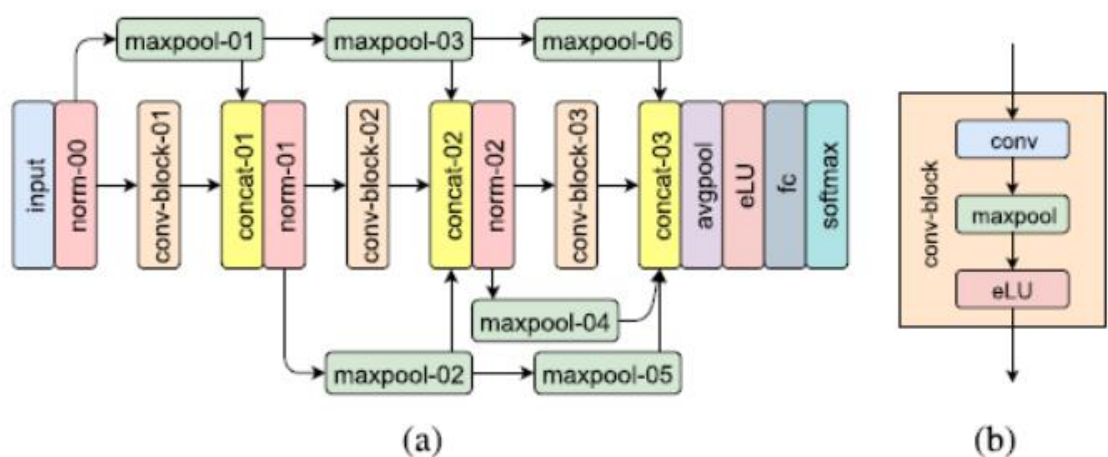


Рис. 1.11. Архітектура нейромережі UATC-DensNet [55]

У роботі [56] запропоновано поєднання згорткової нейронної мережі (CNN) з методом покращення спектра LOFAR. Для розв'язання проблеми обривів у спектрі LOFAR, які можуть виникати через вплив шумів або низьке співвідношення сигнал/шум, розроблено метод покращення спектра LOFAR на основі багаторівневого алгоритму прийняття рішень. Покращений спектр LOFAR використовується як вхідний сигнал для CNN. З підводного акустичного сигналу, отриманого шляхом багаторівневого аналізу спектра LOFAR, розробляється структура згорткової нейронної мережі, яка враховує його особливості. У цій структурі CNN застосовуються деякі ідеї початкового модуля, який використовує різні розміри ядер згорткової обробки та оцінює важливість характеристик глобального та локального розподілу інформації. Ця архітектура дозволяє вибирати оптимальні ядра згорткової обробки та ядра групування для ефективного виділення ознак. Мережа ефективно виявляє позиційні залежності спектральних ліній на різних частотах у спектрі LOFAR.

Дані, що використовувалися у цій роботі, були поділені на дві частини:

- База даних ShipsEar, що містить записи підводних акустичних сигналів.
- Набір із чотирьох типів сигналів, що генеруються шумами.

База даних була розділена на чотири категорії відповідно до типу та розміру суден. Дата, погодні умови під час запису, координати та стан руху суден, кількість, глибина і потужність гідрофонів варіювалися між записами. Результати показали, що метод LOFAR-CNN досяг максимальної точності розпізнавання на рівні 95,22%, що перевищує точність традиційних методів. Цей підхід дозволяє ефективно класифікувати дані за різних рівнів співвідношення сигнал/шум і в різних умовах навколишнього середовища.

У роботі [57] представлено новий метод, що інтегрує CNN та LSTM. Згорткова нейронна мережа використовує одновимірну згортку для обробки вхідних даних. Оскільки характеристики підводного акустичного сигналу судна змінюються з часом, можна застосувати нейромережу LSTM, щоб обробляти як характеристики поточного моменту, так і збережену інформацію

з попередніх моментів. Поєднання 1D-CNN та мережі LSTM дає змогу системі швидко адаптуватися до змін сигналу та підвищити точність розпізнавання.

Частина CNN складається з двох послідовних згорткових шарів та двох шарів групування. Шар групування використовує максимальне об'єднання. Після нього застосовується шар виключень. Мережа LSTM складається з одного шару LSTM та одного шару виключень. Нарешті, вихідні дані передаються у повнозв'язаний шар для класифікації. Архітектура нейромережі зображена на рисунку 1.12.

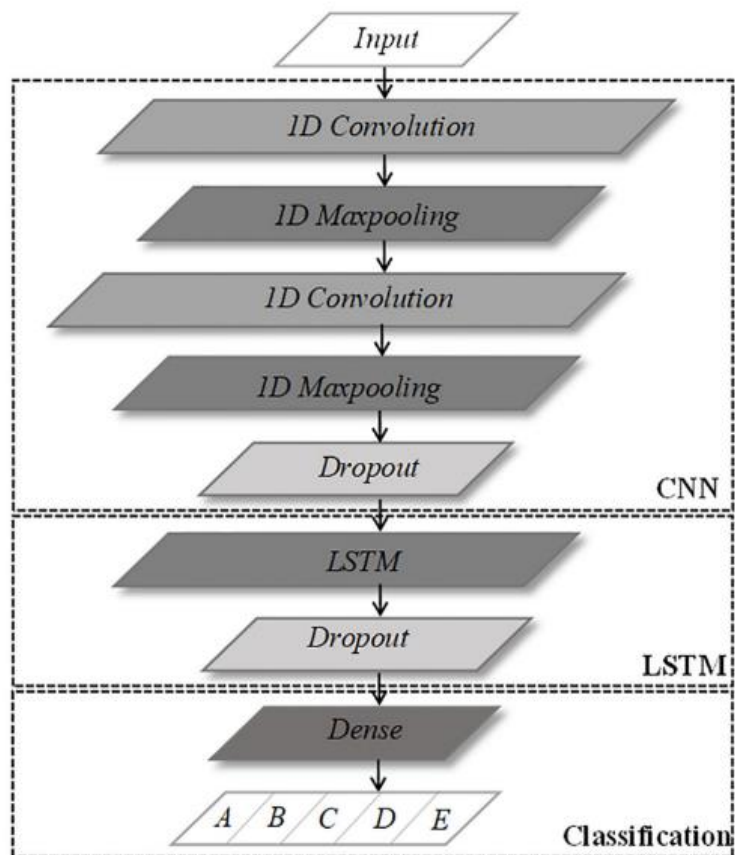


Рис. 1.12. Модель CNN LSTM [57]

Датасет містить 90 акустичних сигналів, записаних в реальних умовах середовища. Деякі записи містили значний рівень шуму або порожні сегменти, тому вони були оброблені шляхом видалення шуму або часткового видалення низькоякісних фрагментів аудіозаписів. Результати класифікації запропонованого методу були порівняні із звичайною згортковою нейронною мережею (CNN) та мережею довготривалої короткочасної пам'яті (LSTM). За

результатами експериментів на наборі даних підводних суден, середня точність роботи нейромережі склала 92%, що на 8% вище, ніж у CNN, і на 15% вище, ніж у LSTM.

У роботі [58] було запропоновано нову систему класифікації, що поєднує архітектуру CNN із SOP для виділення часових кореляцій у часово-частотному представленні акустичних сигналів. Згорткові шари використовуються для навчання локальних характеристик за допомогою набору ядерних фільтрів. На вхід подаються характеристики, отримані шляхом CQT. Замість максимального групування запропонований оператор SOP дозволяє вивчати взаємозв'язки між різними фільтрами CNN, використовуючи часову траєкторію ознак CNN для кожного частотного піддіапазону. Для збереження частотних відмінностей корелюючі ознаки кожної частоти зберігаються окремо. Результати об'єднуються, нормалізуються за допомогою квадратного кореня та L2-нормалізації, після чого подаються на Softmax-класифікатор. Архітектура нейромережі CNN з групуванням другого порядку представлена в таблиці 1.1.

Таблиця 1.1. Архітектура нейромережі CNN з групуванням другого порядку [58]

Softmax 5
Dense 1024
Norm. 4096
SOP 16× 16× 16
Conv. 16× 16× 16
Conv. 32× 32× 8
CQT input (64× 64)

У ході експериментів використовували датасет, що містить акустичні сигнали, записані за допомогою одного гідрофона. Гідрофон встановлювали на глибинах 50 м, 70 м, 100 м, 150 м та 200 м під рівнем моря. Записи

здійснювали для п'яти різних суден, що відрізнялися за вагою, розміром, конструкцією гвинта та системою приводу. Під час кожного запису судно розташовувалося на відстані від -500 м до +500 м. Роботу нейромережі порівнювали з результатами глибокого автоенкодера та нейромережі глибокої довіри. Отримані результати продемонстрували, що запропонована нейромережа забезпечує високу точність класифікації, перевищуючи показники інших моделей щонайменше на 8%.

### 1.7.2. Автоенкодери

У роботі [59] представлено кооперативний метод глибокого навчання, що поєднує глибоку мережу довготривалої короточасної пам'яті (DLSTM) та глибокий автоенкодер (DAE). Спочатку використовується автоенкодер LSTM, який навчає шари DLSTM. На другому етапі навчену модель DLSTM та Softmax-класифікатор застосовують для класифікації підводних шумів. Блок-схема методу наведена на рисунку 1.14.

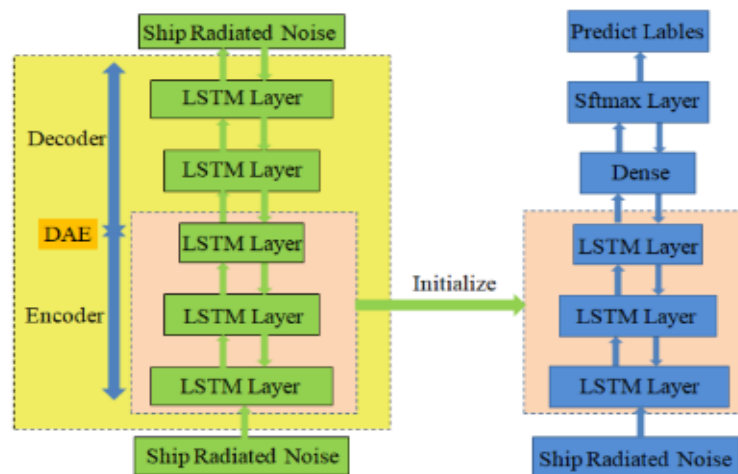


Рис. 1.14. Відображення роботи кооперативного методу глибокого навчання [59]

Перевірку запропонованого методу здійснювали на датасеті, що містить реальні записи підводних звуків пасажирських та вантажних суден, а також фоновий шум. Запропонований метод порівнювали з глибоким автоенкодером

(DAE) та глибокою мережею довготривалої короткочасної пам'яті (DLSTM). Отримані результати засвідчили найвищу точність класифікації, що становила 90%.

У роботі [60] автор запропонував модель складеного розрідженого автоенкодера (SSAE), яка включає кілька прихованих шарів розрідженого автоенкодера та Softmax-класифікатор. Розріджений автоенкодер дозволяє формувати розріджене представлення початкових даних. Для покращення дискримінаційних характеристик запропоновано використання багатoshарової структури SSAE, що будується шляхом об'єднання вхідного та прихованого шарів розрідженого автоенкодера поетапно. При цьому шари декодера для кожного окремого розрідженого автоенкодера видаляються. Модель нейромережі SSA зображена на рисунку 1.15.

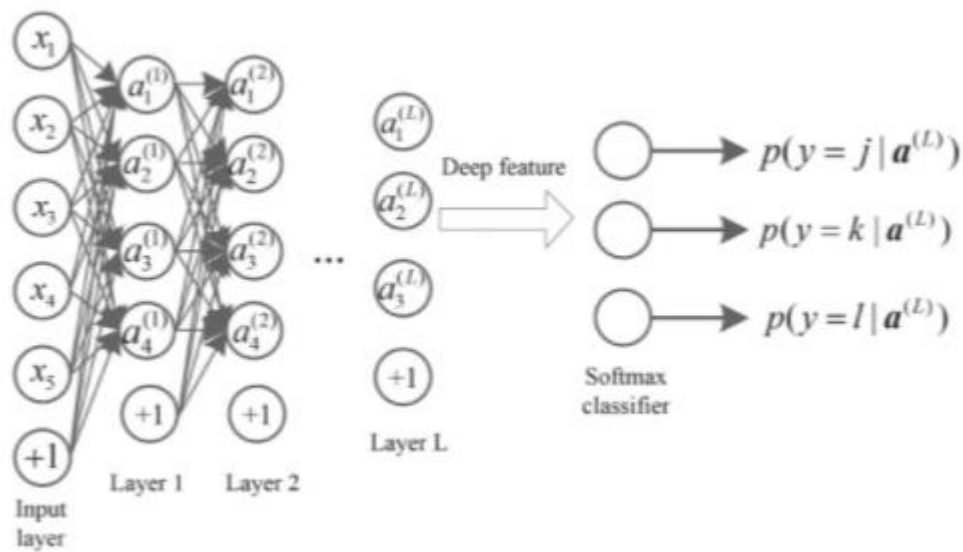


Рис. 1.15. Відображення нейромережевої моделі SSAE [60]

Датасет містив записи, отримані на різних глибинах. На вхід подавали спектральні та вейвлет-представлення сигналів. Результати порівнювали з мережею глибокої довіри та методом опорних векторів. Запропонований метод продемонстрував найвищу точність класифікації, що становила 92%. Підхід виявився стійким до змін шуму та дозволив покращити загальні

характеристики нейромережі, забезпечуючи високі результати за різних співвідношень сигнал/шум.

У роботі [61] запропоновано метод нейромережевого сепарабельного згорткового автоенкодера. Модель складається з мережі енкодера та декодера, де модифікована версія Xception використовується як основа для енкодера. Основна концепція запропонованої архітектури полягає у розділенні міжканальних та просторових кореляцій у картах ознак. Такий підхід підвищує ефективність навчання нейромережі та покращує точність класифікації. У кожному блоці та шарі визначено кількість фільтрів і їхній розмір. У фінальному шарі використовується Softmax-класифікатор.

Метод ґрунтується на роздільній згортці по глибині, яка передбачає виконання просторової згортки окремо для кожного каналу вхідного сигналу. Архітектура нейромережі представлена в таблиці 1.2.

Таблиця 1.2. Шари сепарабельного згорткового автоенкодера [61]

Layer Type	Configuration
Encoder	
Convolution + BatchNorm	Filters: $64 \times (3 \times 3)$ , ReLU
Convolution + BatchNorm	Filters: $64 \times (3 \times 3)$ , ReLU
Xception Block	Filter: 128
Xception Block	Filters: 256
4 × Middle Block	
ReLU	
Xception Block	Filters: 256
SeperableConv Block	Filters: 256
SeperableConv Block	Filters: 256
globalAvgPooling	
Fully connected	10 units; Softmax classifier
Decoder	
Upsampling	Factor: $4 \times 4$
Convolution + BatchNorm	Filters: $128 \times (3 \times 3)$ , ReLU
Upsampling	Factor: $3 \times 4$
Convolution + BatchNorm	Filters: $64 \times (3 \times 3)$ , ReLU
Upsampling	Factor: $2 \times 2$
Convolution + BatchNorm	Filters: $32 \times (3 \times 3)$ , ReLU
Upsampling	Factor: $2 \times 2$
Convolution	Filters: $1 \times (3 \times 3)$ , ReLU

Датасет містить понад 47 годин реальних підводних записів, отриманих від 265 суден, що належать до чотирьох різних класів. Записи охоплюють різні пори року, умови моря та рівні шуму. Вхідні дані представлені у часово-



частотному форматі у вигляді зображень, що включають кепстр, мел-спектрограму, мел-частотні кепстральні коефіцієнти, гамматон-частотні спектральні коефіцієнти та вейвлет-пакети. Метод порівнювали з іншими моделями, такими як метод опорних векторів (SVM), алгоритм k-найближчих сусідів (KNN), наївний байєсівський класифікатор (Naive Bayes), випадковий ліс (RF), глибока нейронна мережа (DNN) та згорткова нейронна мережа (CNN). Запропонована модель досягла точності 77% при використанні перетворення з постійним Q. Висока гнучкість нейромережі дозволяє застосовувати її для класифікації великих і різноманітних наборів акустичних сигналів у реальних умовах, забезпечуючи відносно високу точність результатів.

### **1.7.3. Багатошарові персептрони**

У роботі [62] запропоновано підхід, що поєднує багатошаровий персептрон (MLP) та автоенкодер на основі обмеженої машини Больцмана. Метод використовує нормалізований частотний спектр сигналу як вхідні дані, застосовує кодувальник на базі обмеженої машини Больцмана для автоматичного кодування та послідовного витягнення глибокої структури даних. Отримані ознаки класифікуються за допомогою багатошарової нейронної мережі. У дослідженні використовувався п'ятишаровий персептрон, навчений методом зворотного поширення помилок. Експериментальне тестування проводилося на двох наборах даних, записаних у реальних умовах на різних глибинах, де метод продемонстрував середню точність класифікації на рівні 93%.

У роботі [63] запропоновано використання багатошарового персептрона, навчання якого здійснюється за допомогою алгоритму рою сальпів. Загалом, цей тип нейромережі демонструє високу ефективність у змінному середовищі, проте його точність класифікації може бути нестабільною, а процес навчання іноді призводить до застрягання у локальних

мінімумах. Для вирішення цих проблем було використано алгоритм рою сальпів, який покращує процес оптимізації та підвищує надійність навчання. Архітектура нейромережі представлена на рисунку 1.16.

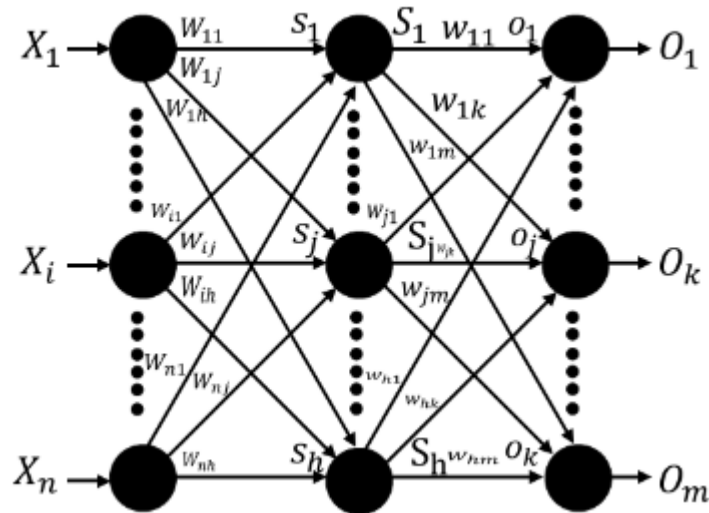


Рис. 1.16. Архітектура багатошарового персептрона [63]

У дослідженні застосовувався трьохшаровий персептрон, що працював із вхідними даними у вигляді вектора мел-частотних кепстральних коефіцієнтів, що сприяло зменшенню розмірності ознак. Під час тестування цього підходу використовувалися дані з різними рівнями відношення сигнал/шум. Проведене порівняння з іншими методами навчання нейромережі показало, що запропонований підхід досяг найвищої точності класифікації, яка склала 97%.

## 1.8. Нейромережі для придушення шумів

Стаття [64] представляє складну систему шумозаглушення, спеціально розроблену для вирішення проблем підводного фонового шуму (UWAN) в акустичних сигналах. У центрі цього дослідження лежить архітектура GAN, яка унікально інтегрує як амплітудний, так і фазовий спектри з представлень STFT підводних акустичних сигналів [65]. Традиційні методи шумозаглушення, такі як статистичні методи, наприклад, фільтрація Вінера,

або алгоритми на основі вейвлетів, часто стикаються з залишковими артефактами, музичним шумом або обмеженою ефективністю при низьких співвідношеннях SNR. Більше того, ці традиційні методи часто ігнорують фазовий компонент акустичних сигналів, який відіграє критичну роль у збереженні якості звуку. Щоб вирішити ці обмеження, дослідження представляє UWAR-GAN, умовну GAN (сGAN) систему, призначену для шумозаглушення підводних сигналів, зберігаючи при цьому основні акустичні характеристики.

Архітектура UWAR-GAN складається з генератора на основі U-net і дискримінатора PatchGAN. Генератор використовує структуру енкодера-декодера з пропускними з'єднаннями, що дозволяє зберігати дрібні деталі з вхідного шумового спектрограми, одночасно навчаючись реконструювати чисті версії. На відміну від багатьох традиційних підходів на основі GAN, які зосереджуються лише на амплітудному спектрі, UWAR-GAN включає як амплітудний, так і фазовий спектри, що значно покращує якість реконструйованого звуку. Дискримінатор, реалізований у стилі PatchGAN, оцінює менші ділянки згенерованого виходу, що дозволяє моделі зосередитися на локальній спектральній узгодженості, а не лише на глобальних шаблонах.

Критичним внеском цієї моделі є включення функції втрат кореляції, яка забезпечує високу кореляцію вихідного спектрограми з чистим еталоном як за амплітудними, так і за фазовими вимірами. Ця втрата особливо ефективна у збереженні тональних компонентів звуків корабельних двигунів, навіть у середовищах з дуже низьким SNR. Функція втрат поєднує адверсаріальну втрату, L1 втрату реконструкції та втрату кореляції, де кожен компонент точно налаштований з певними вагами для балансування адверсаріальної точності, піксельного рівня точності та спектральної узгодженості.

Набір даних, використаний для навчання та оцінки, включає як реальні, так і змодельовані підводні акустичні записи, що охоплюють різні умови навколишнього шуму, включаючи дощ, хвилі та турбулентність. Перетворення STFT застосовується до сирих аудіосигналів для перетворення їх у часово-частотні представлення, які потім подаються у GAN систему. Через численні експерименти та порівняльний аналіз UWAR-GAN демонструє значні покращення порівняно з базовими методами. Метрики, такі як SNR, Structural Similarity Index (SSIM), Mean Opinion Score (MOS) та RMSE, підкреслюють надійність та ефективність моделі. Порівняно з традиційним методом фільтрації Вінера, UWAR-GAN досягає покращення SNR на 45.1%, збільшення SSIM на 40.7% та аналогічно вражаючих результатів у MOS та RMSE. Навіть у порівнянні з іншими системами шумозаглушення на основі GAN, UWAR-GAN показує стабільну перевагу, особливо у складних умовах з низьким SNR, таких як -10 дБ.

Дослідження підкреслює, що включення фазової інформації не є просто додатковим, а є невід'ємною частиною досягнення високоякісної реконструкції звуку. Без збереження фази денойзовані сигнали часто страждають від тональних невідповідностей та чутних спотворень. Крім того, дослідження підкреслює значення адверсаріальної кореляційної втрати, яка є ключовим компонентом у вдосконаленні здатності мережі відокремлювати тональні звуки корабельних двигунів від фонового шуму.

На закінчення, UWAR-GAN представляє значний прорив у шумозаглушенні підводних акустичних сигналів, поєднуючи надійну архітектурну конструкцію з інноваційними функціями втрат та всебічним обробленням як амплітудного, так і фазового спектрів. Ця робота не тільки створює потужний інструмент для аналізу шуму підводних суден, але й закладає основу для майбутніх досліджень у реальному часі застосувань шумозаглушення в морському середовищі.

Стаття [66] представляє передовий підхід до зниження шуму підводних акустичних сигналів, вирішуючи обмеження традиційних методів шумозаглушення в складних і динамічних морських середовищах. Традиційні підходи, такі як вейвлет-декомпозиція, емпірична модова декомпозиція та статистична фільтрація, часто не справляються з нестабільною природою підводного шуму, який може значно змінюватися через біологічну активність, змінні океанічні умови та механічні перешкоди. Ці традиційні методи зазвичай покладаються на ручне виділення ознак та попередньо задані припущення щодо характеристик шуму, що призводить до зниження продуктивності в непередбачуваних умовах. Щоб подолати ці недоліки, дослідження представляє архітектуру Fully Convolutional Encoder-Decoder Neural Network (FCEDN) на основі глибокого навчання, здатну навчатися високорівневим ознакам безпосередньо з часово-частотних представлень підводних акустичних сигналів. У центрі запропонованого методу лежить перетворення акустичних хвиль у часовій області в вейвлет-спектрограми низькочастотного аналізу. На відміну від стандартних часово-частотних перетворень, таких як короткочасне перетворення Фур'є (STFT), вейвлет-підхід забезпечує більш адаптивне часово-частотне представлення, захоплюючи як дрібні деталі, так і ширші тенденції сигналу в різних частотних діапазонах. Це перетворення забезпечує збереження критичних підводних ознак сигналу, таких як гармоніки двигуна корабля або інші тональні компоненти, під час попередньої обробки. Спектрограми служать вхідними даними для FCEDN, де етап енкодера стискає високовимірні дані спектрограми в низьковимірне представлення ознак. Енкодер використовує конволюційні шари з різними розмірами ядер для виділення як локальних, так і глобальних ознак зі спектрограми, дозволяючи мережі навчатися складним взаємозв'язкам між шумом і чистими сигналами. Етап декодера потім відновлює чисту спектрограму з закодованої карти ознак за допомогою транспонованих конволюційних шарів, які збільшують дані до їх початкових розмірів. Ця структура дозволяє мережі ефективно навчатися та зворотньо

відтворювати процес спотворення шуму, зберігаючи структурну цілісність сигналу. Важливо, що дизайн енкодера-декодера уникає інформаційного вузького місця, яке часто зустрічається в простіших архітектурах автоенкодера.

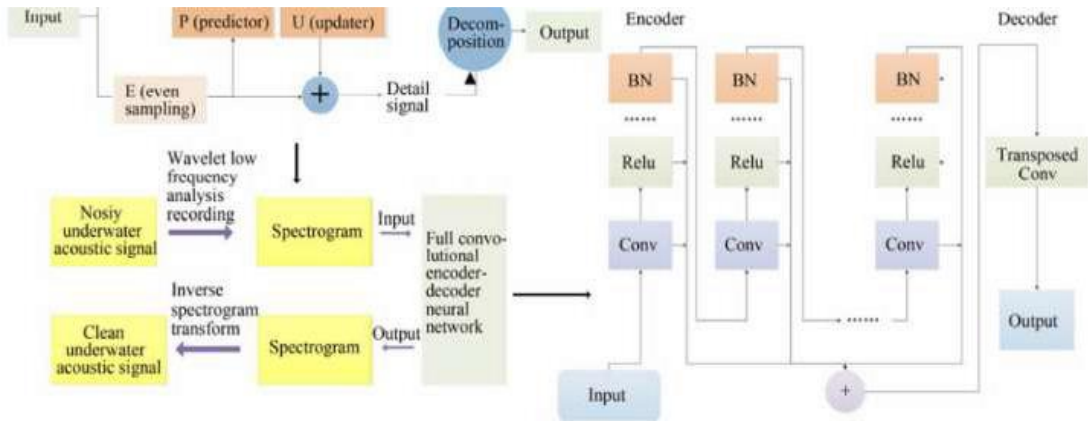


Рис. 1.17. Архітектура нейромережі FCEDN [66]

Модель також вводить підхід на основі відображення, замінюючи традиційні методи маскування. Ця стратегія відображення підвищує здатність FCEDN до узагальнення різних типів шуму без покладання на фіксовані припущення щодо розподілу шуму. Навчання FCEDN включає мінімізацію функції втрат, яка балансує точність реконструкції та спектральну узгодженість. Набір даних, використаний для навчання, включає зразки з набору даних ShipsEar, відомого підводного акустичного набору даних, що містить записи шуму, випромінюваного кораблями, в різних умовах навколишнього середовища та рівнях шуму. Під час навчання мережа навчається мінімізувати різницю між передбаченими чистими спектрограмами та їх відповідними еталонними зразками. Оптимізатор Adam використовується для налаштування параметрів мережі, а модель навчається за допомогою міні-пакетів для прискорення збіжності. Оцінка FCEDN показує значні покращення в продуктивності шумозаглушення порівняно з традиційними та сучасними підходами глибокого навчання. Основні метрики, такі як SNR та масштабно-інваріантне відношення сигнал/шум (SI-SNR), використовуються для кількісної оцінки ефективності моделі. Результати

показують, що FCEDN досягає збільшення на 10,02 дБ у SNR та 9,5 дБ у SI-SNR, підкреслюючи її здатність відновлювати чисті сигнали навіть у дуже шумних умовах. Ці покращення пояснюються здатністю мережі зберігати як локальну, так і глобальну контекстну інформацію протягом усього процесу кодування та декодування. Однією з ключових сильних сторін FCEDN є її здатність справлятися з динамічними шумовими середовищами. На відміну від традиційних методів, які часто вимагають повторного навчання або ручного налаштування параметрів для нових умов шуму, FCEDN демонструє сильні здатності до узагальнення в різних підводних акустичних середовищах. Крім того, повністю конволюційна природа мережі дозволяє їй обробляти сигнали довільної довжини без необхідності фіксованих розмірів вхідних даних, що робить її дуже адаптивною для реального використання в системах підводного моніторингу. На закінчення, FCEDN представляє значний прорив у зниженні шуму підводних акустичних сигналів. Використовуючи вейвлетні часово-частотні представлення, добре структуровану архітектуру енкодера-декодера та інноваційний підхід на основі відображення, FCEDN успішно долає багато обмежень традиційних методів шумозаглушення. Її здатність ефективно відокремлювати шум від цільових сигналів у різних підводних умовах підкреслює її потенціал для застосування в морському спостереженні, системах підводного зв'язку та моніторингу навколишнього середовища.

Стаття [67] досліджує інноваційний підхід до вирішення нелінійного сліпого розділення джерел (BSS) в підводних акустичних середовищах, де шум, випромінюваний кораблями, часто спотворюється складними нелінійними ефектами. У підводних умовах сигнали, захоплені гідрофонами, зазвичай забруднені шумом, що походить від природних джерел, механічних вібрацій та біологічної активності. Ці спотворення ще більше посилюються ефектом повітряних бульбашок, створених гвинтами кораблів, які вводять значні нелінійності в процес акустичної пропagaції. Традиційні методи BSS, які в основному покладаються на лінійні припущення, не можуть ефективно

розділяти ці сигнали, особливо в присутності сильних нелінійних спотворень. Щоб подолати ці виклики, дослідження представляє архітектуру RANN, яка інтегрує Bi-directional Long Short-Term Memory (Bi-LSTM) мережі з механізмами уваги для моделювання як локальних тимчасових залежностей, так і глобальних контекстуальних взаємозв'язків в акустичних сигналах. У центрі RANN лежить гібридна нейронна архітектура, призначена для вирішення як тимчасових, так і спектральних складностей підводних сигналів. Шари Bi-LSTM відповідають за захоплення короткострокових і довгострокових тимчасових залежностей, дозволяючи моделі розуміти послідовні взаємозв'язки в часовій області сигналу. Архітектура нейромережі RANN представлена на рисунку 1.18.

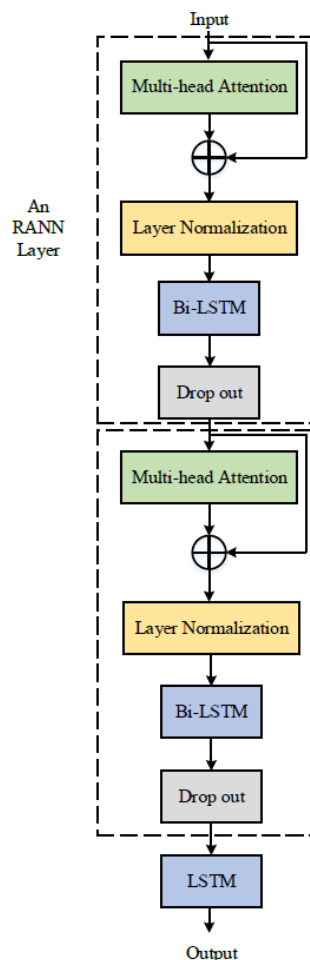


Рис. 1.18. Архітектура нейромережі RANN [67]



На відміну від стандартних LSTM, Bi-LSTM обробляють вхідну послідовність як у прямому, так і в зворотному напрямках, дозволяючи мережі одночасно використовувати контекстну інформацію з минулого та майбутнього. Ця двонаправлена обробка особливо ефективна у вирішенні тимчасових неоднозначностей та ізоляції компонентів сигналу, спотворених нелінійними ефектами пропагації. Доповнюючи шари Bi-LSTM, механізм багатоголової уваги відмінно захоплює глобальні залежності в просторі ознак сигналу. Механізми уваги працюють, призначаючи динамічні ваги різним частинам вхідної послідовності, дозволяючи моделі зосереджуватися на найбільш релевантних ознаках, мінімізуючи вплив нерелевантного шуму. Дизайн багатоголової уваги розширює цю можливість, дозволяючи мережі одночасно звертати увагу на кілька підмножин ознак, кожна з яких обробляється через окрему голову уваги. Це забезпечує, що модель може захоплювати різноманітні шаблони та взаємозв'язки в даних, навіть у присутності нелінійних взаємодій. Архітектура RANN складається з двох шарів рекурентних модулів уваги, де кожен модуль поєднує шар Bi-LSTM, шар багатоголової уваги та залишкові з'єднання для полегшення потоку градієнтів та запобігання зникненню градієнтів під час навчання. Залишкові з'єднання також допомагають зберігати критичні низькорівневі ознаки через глибші шари. Після двох рекурентних модулів уваги додається фінальний шар LSTM для уточнення представлення сигналу та отримання вихідних розділених сигналів. Для навчання та оцінки RANN автори використовували набір даних ShipsEar, який містить записи шуму, випромінюваного кораблями. Набір даних був доповнений для імітації нелінійних спотворень, викликаних бульбашками, використовуючи математичні моделі, отримані з рівняння Релея-Плессета та моделей пост-нелінійного (PNL) змішування. Ці моделі імітують взаємодію акустичних хвиль з повітряними бульбашками, захоплюючи нелінійні варіації тиску та їх вплив на отриманий сигнал. Набір даних був розділений на навчальні, валідаційні та тестові підмножини у співвідношенні 7:1:2 для забезпечення надійної оцінки на невидимих даних.

Під час процесу навчання модель оптимізувала спеціальну функцію втрат, впливала на середньоквадратичну помилку (MSE), коефіцієнти кореляції та відношення сигнал/спотворення (SDR). Оптимізатор, використаний для навчання, був Adam, і модель була точно налаштована для запобігання перенавчання через шари відсіву та нормалізації пакетів. Вхідні послідовності були сегментовані на менші пакети, кожен з яких складався з фіксованої кількості часових кроків, забезпечуючи обчислювальну ефективність та стабільні оновлення градієнтів під час навчання. Результати оцінки показали, що RANN значно перевершує існуючі методи для підводного акустичного BSS, включаючи традиційні лінійні алгоритми BSS та інші підходи на основі глибокого навчання, такі як стандартні мережі LSTM та архітектури Transformer. Метрики, такі як середньоквадратична помилка (MSE), коефіцієнт кореляції та відношення сигнал/спотворення (SDR), використовувалися для кількісної оцінки продуктивності моделі. RANN досягла нижчих значень MSE, вищих коефіцієнтів кореляції та кращих показників SDR порівняно з конкуруючими моделями, підкреслюючи її ефективність у точному розділенні спотворених сигналів з нелінійних сумішей. Однією з ключових сильних сторін RANN є її здатність балансувати локальне виділення ознак через шари Bi-LSTM та глобальне контекстуальне розуміння через багатоголову увагу. Хоча Bi-LSTM відмінно моделюють тимчасові шаблони та залежності в коротких послідовностях, механізм уваги компенсує проблему ефективного захоплення довгострокові залежності. Разом ці компоненти дозволяють RANN ефективніше справлятися зі складними, нестабільними та нелінійними характеристиками підводних акустичних сигналів, ніж традиційні методи. На закінчення, RANN представляє значний прорив у галузі нелінійного сліпого розділення джерел для підводних акустичних сигналів. Поєднуючи шари Bi-LSTM для тимчасового моделювання та механізми багатоголової уваги для вилучення глобального контексту, RANN пропонує надійне рішення для розділення шуму, випромінюваного кораблями, у складних підводних середовищах. Її

висока продуктивність за кількома метриками оцінки демонструє її потенціал для реальних застосувань у підводному спостереженні, виявленні кораблів та системах обробки акустичних сигналів.

Підхід до шумозаглушення на основі статті [68] зосереджується на використанні генеративних змагальних мереж (GAN) для очищення підводних акустичних сигналів, зокрема звуків двигунів кораблів, шляхом видалення фонового шуму. Система називається Ambient Noise-Free GAN (AN-GAN), спеціалізована структура, розроблена для вирішення проблем шумних підводних середовищ. GAN добре підходять для цього завдання завдяки своїй здатності генерувати реалістичні розподіли даних з шумних вхідних сигналів. Архітектура AN-GAN складається з двох основних компонентів: генератора та дискримінатора. Генератор є повністю конволюційною мережею, яка навчається створювати чисті аудіосигнали шляхом перетворення шумних вхідних сигналів. Він працює в часовій області, генеруючи вихідні сигнали, що імітують основну структуру чистих аудіосигналів. Дискримінатор, мережа на основі патчів, оцінює ці вихідні сигнали, порівнюючи їх з шумними сигналами. Замість того, щоб безпосередньо розрізняти чисті та шумні сигнали, дискримінатор порівнює вимірювання (наприклад, шумові патерни) для оцінки автентичності згенерованого аудіо. Архітектура нейромережі AN-GAN представлена в таблиці 1.3.

Таблиця 1.3. Архітектура нейромережі AN-GAN [68]

Layer	Output Shape	Layer	Output Shape
Input $x$ or $G(z)$	$(b, 16384, 1)$	Input $z \sim \text{Gaussian}(-1, 1)$	$(b, 100)$
Conv1D (Stride=4)	$(b, 4096, 1)$	Dense	$(b, 256)$
LReLU ( $\alpha = 0.2$ )	$(b, 4096, 1)$	Reshape	$(b, 16, 16)$
Conv1D (Stride=4)	$(b, 1024, 2)$	ReLU	$(b, 16, 16)$
LReLU ( $\alpha = 0.2$ )	$(b, 1024, 2)$	Trans Conv1D (Stride=4)	$(b, 64, 8)$
Conv1D (Stride=4)	$(b, 256, 4)$	ReLU	$(b, 64, 8)$
LReLU ( $\alpha = 0.2$ )	$(b, 256, 4)$	Trans Conv1D (Stride=4)	$(b, 256, 4)$
Conv1D (Stride=4)	$(b, 64, 8)$	ReLU	$(b, 256, 4)$
LReLU ( $\alpha = 0.2$ )	$(b, 64, 8)$	Trans Conv1D (Stride=4)	$(b, 1024, 2)$
Conv1D (Stride=4)	$(b, 16, 16)$	ReLU	$(b, 1024, 2)$
LReLU ( $\alpha = 0.2$ )	$(b, 16, 16)$	Trans Conv1D (Stride=4)	$(b, 4096, 1)$
Reshape	$(b, 256)$	ReLU	$(b, 4096, 1)$
Dense	$(b, 16)$	Trans Conv1D (Stride=4)	$(b, 16384, 1)$
		Tanh	$(b, 16384, 1)$

AN-GAN використовує процес, натхненний AmbientGAN, де генератор імітує чисте аудіо, а потім спотворює його за допомогою попередньо визначеної моделі шуму. Цей змодельований шумний сигнал подається в дискримінатор разом із реальними зашумленими сигналами. Мета генератора - створити чисті аудіосигнали, спотворені версії яких не відрізняються від реальних шумних вхідних сигналів. Навчання мережі включає змагальне навчання, де генератор і дискримінатор поступово покращуються, змагаючись один з одним. Набір даних, використаний у цьому дослідженні, - це ShipsEar, який містить записи різних звуків двигунів кораблів та фонового підводного шуму. Чисті сигнали отримані з записів кораблів, що проходять поблизу, з мінімальним фоновим шумом, тоді як зразки фонового шуму записані незалежно. Ці записи попередньо обробляються на аудіофрагменти тривалістю одну секунду, щоб забезпечити узгоджені розміри вхідних даних для навчання. Генератор і дискримінатор оптимізуються за допомогою методів на основі градієнтів, причому генератор прагне мінімізувати розбіжність між своїм виходом і справжнім розподілом даних. Продуктивність AN-GAN перевіряється за допомогою реальних і змодельованих зашумлених сигналів. Результати демонструють його здатність покращувати відношення сигнал/шум (SNR) порівняно з традиційними методами шумозаглушення, такими як вейвлетне шумозаглушення, фільтрація Вінера та емпірична модова декомпозиція. Структура досягає середнього покращення SNR на понад 6% порівняно з методами на основі GAN і значно перевершує традиційні статистичні методи, особливо в сценаріях з високим рівнем шуму. Цей підхід на основі GAN ефективно демонструє потенціал змагальних мереж у шумозаглушенні аудіо, особливо в складних підводних середовищах, де отримання чистих навчальних даних не є практичним. Використовуючи парадигму змагального навчання, AN-GAN вдається навчитися надійним представленням чистих аудіосигналів безпосередньо з шумних наборів даних, що є значним досягненням у обробці підводних акустичних сигналів.

Стаття [69] - це модель на основі глибокого навчання, призначена для зниження шуму в підводних акустичних сигналах. Вона вирішує проблему розділення цільових сигналів від фонового шуму, використовуючи механізм обробки з двома шляхами, який працює в часових і просторових вимірах. Ця архітектура базується на сильних сторонах рекурентних нейронних мереж (RNN) для моделювання послідовних даних і вводить інновації для обробки як локальних, так і глобальних залежностей у сигналі. Фреймворк DPRNN починається з фази виділення ознак, де одновимірний згортковий нейронний мережа (1D CNN) виділяє ознаки вищого порядку з сирого вхідного сигналу. Ці ознаки сегментуються на фіксовані за довжиною частини, які потім об'єднуються у тривимірний тензор. Цей тензор служить вхідними даними для DPRNN, дозволяючи мережі обробляти сигнал у керованих сегментах, зберігаючи часову структуру. Основою DPRNN є стратегія обробки з двома шляхами, яка включає внутрішню та міжчасткову операції. Внутрішня обробка зосереджується на виділенні локальних ознак у кожному сегменті вхідного тензора. Це досягається за допомогою двонаправлених RNN, які моделюють дрібнозернисті часові взаємозв'язки в кожному часткові. Після незалежної обробки кожного сегмента, міжчасткова операція бере на себе захоплення глобальних залежностей між частками. Цей крок включає однобічні RNN, які аналізують, як різні сегменти сигналу пов'язані один з одним, дозволяючи мережі реконструювати узгоджене представлення всього сигналу. Після обробки з двома шляхами DPRNN застосовує нормалізацію та залишкові з'єднання для стабілізації процесу навчання та покращення навчання ознак. Ці залишкові з'єднання забезпечують ефективну передачу градієнтів під час навчання, пом'якшуючи проблеми, такі як зникнення або вибух градієнтів. Оброблені ознаки потім передаються через повністю зв'язані шари для подальшого уточнення вивчених представлень. Для реконструкції очищеного сигналу мережа використовує механізм перекриття-додавання. Цей метод агрегує вихідні дані окремих часток, забезпечуючи безперервний перехід між сегментами та зберігаючи цілісність аудіосигналу.

Реконструйований сигнал оцінюється за допомогою метрик, таких як SNR та масштабно-інваріантне SNR (SI-SNR), які оцінюють здатність моделі покращувати якість цільового сигналу, пригнічуючи фоновий шум. DPRNN оцінюється на наборі даних ShipsEar, колекції підводних аудіозаписів, що містять звуки двигунів кораблів, змішані з різними фоновими шумами. Набір даних попередньо обробляється на мічені зразки, і DPRNN навчається за допомогою завдань з різними рівнями складності шуму. Результати показують, що DPRNN значно покращує SNR, досягаючи покращень до 12,02 дБ та 9,48 дБ у конкретних завданнях. Ці покращення підкреслюють здатність моделі узагальнювати різні рівні шуму та підтримувати надійну продуктивність у реалістичних підводних сценаріях. Поєднуючи сильні сторони обробки з двома шляхами та рекурентних архітектур, DPRNN ефективно вирішує проблеми зниження шуму підводних акустичних сигналів. Її здатність захоплювати як локальні, так і глобальні залежності в сигналі робить її потужним інструментом для застосувань шумозаглушення в складних середовищах.

У статті [70] розроблена для шумозаглушення підводних акустичних сигналів. Вона вирішує притаманні проблеми шумних середовищ, поєднуючи переваги механізмів уваги на основі трансформерів та двошляхової обробки. Модель структурована для захоплення як короткострокових, так і довгострокових залежностей у сигналі, пропонуючи комплексний підхід до зниження шуму.

Процес починається з модуля вилучення ознак, який використовує одно-вимірну згорткову нейронну мережу (Conv1D). Цей модуль перетворює сирий сигнал у часовій області, що містить як цільові, так і шумові компоненти, у високовимірне представлення. Вилучені ознаки потім діляться на підпоследовності фіксованої довжини, які реорганізуються у тривимірний тензор. Цей тензор стає вхідними даними для подальшої двошляхової обробки.

Основна інновація DPTN полягає в її двошляховій архітектурі, яка складається з внутрішньо-трансформера та міжтрансформера. Внутрішньо-трансформер зосереджується на моделюванні короткострокових залежностей у кожній підпоследовності, ефективно захоплюючи локальні шаблони, на які впливає шум. На відміну від цього, міжтрансформер працює через підпоследовності, вивчаючи довгострокові залежності та глобальні відносини у сигналі. Такий подвійний підхід дозволяє мережі комплексно обробляти сигнал, вирішуючи як локалізовані спотворення, так і ширші контекстуальні невідповідності.

У блоках трансформера механізм уваги відіграє вирішальну роль. Кожен шар трансформера включає багатоголову самоувагу, що дозволяє моделі одночасно звертати увагу на різні частини последовності. Цей механізм забезпечує, що мережа ідентифікує та підкреслює значущі ознаки, такі як ті, що представляють цільовий сигнал, зменшуючи при цьому компоненти шуму. Позиційне кодування додається до вхідних ознак, щоб надати інформацію про порядок последовності, що є критичним для точного моделювання часових відносин.

Залишкові з'єднання та нормалізація шару застосовуються протягом блоків трансформера для стабілізації навчання та покращення потоку градієнтів. Мережа також включає шар прямого проходження в кожному трансформері, додатково уточнюючи ознаки, вилучені механізмом уваги. Це поєднання багатоголової уваги, залишкового навчання та шарів прямого проходження дозволяє DPTN ефективно захоплювати складну динаміку підводних акустичних сигналів.

Після обробки ознак через двошляховий трансформер, денойзоване представлення передається до модуля реконструкції. Цей модуль використовує техніку перекриття-додавання, агрегуючи оброблені підпоследовності для реконструкції повної довжини денойзованого сигналу. Транспонована згорткова мережа служить декодером, перетворюючи

оброблені ознаки назад у початкову область сигналу, зберігаючи очищений цільовий сигнал. Відображення роботи методу DPTN представлена на рисунку 1.19.

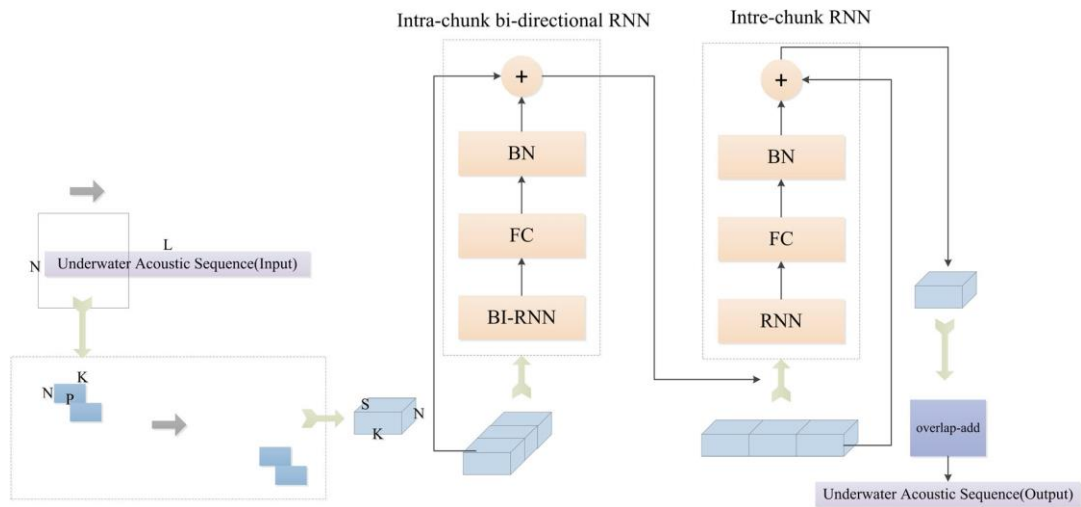


Рис. 1.19. Відображення роботи методу DPTN [70]

Продуктивність DPTN оцінюється за допомогою наборів даних ShipsEar та DeepShip, які включають різноманітні підводні записи з різними рівнями шуму. Завдання розроблені з підвищеною складністю для перевірки стійкості моделі та здатності до узагальнення. Результати показують значні покращення SNR та масштабно-інваріантному SNR (SI-SNR), демонструючи ефективність моделі у зниженні шуму та покращенні чіткості цільового сигналу.

Інтегруючи механізми уваги на основі трансформерів з двошляховою обробкою, DPTN представляє значний прорив у шумозаглушенні підводних акустичних сигналів. Її здатність обробляти як короткострокові, так і довгострокові залежності, у поєднанні з надійними методами вилучення та реконструкції ознак, робить її потужним інструментом для складних завдань зниження шуму в водних середовищах.



### **1.9. Обґрунтування вибору нейромереж для виконання задачі аналізу акустичних сигналів водного середовища залачі аналізу**

У роботі [55] запропонований метод UATC-DenseNet був використаний для класифікації даних із різними рівнями SNR. Архітектура цієї нейромережі здатна обробляти великі обсяги даних та демонструє високу точність класифікації, однак її ефективність у складних умовах досліджена не була. Для отримання повнішої оцінки необхідно перевірити її роботу на датасеті, записаному в більш варіативних і складних умовах.

Метод LOFAR-CNN запропонований в статті [52] для вирішення проблеми класифікації даних з високим значенням сигнал/шум та можливих розривів у спектрі LOFAR. Запропонований метод показав високу точність класифікації, стійкість до даних з низьким співвідношенням сигнал/шум та можливість роботи з великим обсягом даних, але в статті не було проведено дослідження роботи в складних умовах. У роботі [53] датасет був вручну змінений шляхом видалення деяких низькоякісних сигналів або частин порожніх сегментів сигналу вручну. Подібні дії не дозволяють говорити про повністю автоматизований процес класифікації та не дозволяють вибрати цей метод як можливий варіант рішення. Запропонована в роботі [53] архітектура CNN+SOP має високу здатність до узагальнення, ефективно працюючи з даними, записаними на різних глибинах та має стійкість до різних співвідношень сигнал/шум. Запропонована архітектура має гарні можливості для роботи з великими обсягами даних та показує високу точність класифікації, що дозволяє вибрати це рішення як одне з можливих.

У роботі [59] кооперативний метод глибокого навчання був протестований на датасеті, записаному в складних та змінних умовах, проте його ефективність щодо даних із різним співвідношенням сигнал/шум не була вивчена. Сепарабельний згортковий автоенкодер із дослідження [61] відповідає більшості визначених вимог, за винятком точності класифікації.

Для його подальшого використання необхідно дослідити можливості підвищення точності класифікації.

Архітектура, запропонована в роботі [58], найбільше відповідає визначеним критеріям. Нейромережа складеного розрідженого автоенкодера забезпечує високу точність класифікації, демонструє стабільні результати на різних глибинах і при зміні значень SNR, а також здатна ефективно працювати з великими обсягами даних.

Багатошаровий персептрон, навчений за допомогою алгоритму рою сальпів [63], продемонстрував хороші результати для різних значень SNR та в умовах змінного середовища. Однак використаний набір даних є обмеженим і створеним у контрольованих умовах, що ставить під сумнів його ефективність при роботі з реальними даними. Крім того, запропонована архітектура не є глибокою, що може ускладнити її застосування для великих обсягів інформації. Використання багатошарового персептрона та обмеженої машини Больцмана [62] не було протестовано на даних із різними значеннями SNR, хоча автори зазначають, що це може стати предметом подальших досліджень.

У ході аналізу було розглянуто різні нейромережеві підходи до обробки акустичних сигналів. У більшості досліджень використовувалися три основні типи архітектур: згорткові нейронні мережі, автоенкодери та багатошарові персептрони. Всі запропоновані моделі ґрунтуються на цих підходах, тому їх можна класифікувати за відповідними критеріями. З урахуванням визначених вимог до нейромережевих рішень сформульовано ключові критерії оцінки, серед яких: здатність працювати зі складними даними, адаптація до змін SNR, ефективна обробка великих обсягів інформації та висока точність класифікації. Відповідність розглянутих нейромереж визначеним критеріям подано в таблиці 1.4.

Таблиця 1.4. Порівняння розглянутих підходів за обраними критеріями

	Точність класифі- кації	Робота зі складними даними	Використання при різних співвідно- шеннях	Використання великих обсягів даних
UATC-DensNet	+	-	+	+
LOFAR-CNN	+	-	+	+
CNN LSTM	+	-	+	+
CNN SOP	+	+	+	+
Кооперативний метод глибокого навчання	+	+	-	+
SSAE	+	+	+	+
SCAE	-	+	+	+
RBM+MLP	+	+	-	+
MLP+Salm swap	+	+	+	-

Архітектура CNN SOP є найкращим варіантом для цієї задачі. На відміну від інших моделей, вона ефективно працює з даними, записаними на різних глибинах і за різних умов шуму, забезпечуючи високу точність класифікації навіть у складних середовищах. Використання SOP дозволяє краще виявляти часові кореляції у часово-частотному представленні сигналів, що є критично важливим для аналізу підводних акустичних даних. CNN SOP також має хорошу здатність узагальнення, що дає змогу адаптуватися до нових умов, зберігаючи точність. Це стало можливим завдяки інтеграції згорткових шарів, які виявляють локальні характеристики сигналу, та SOP, що захоплює взаємозв'язки між частотними компонентами.

Такий підхід дозволяє неймережі зберігати частотні особливості сигналів, підвищуючи їхню класифікацію навіть за низького SNR. Ефективність архітектури підтверджена експериментальними результатами, які показали, що CNN SOP перевершує інші підходи, такі як LOFAR-CNN або CNN+LSTM, як за точністю, так і за стійкістю до шумів. Це робить її оптимальним вибором для задач класифікації акустичних сигналів водного середовища.

Порівнюючи дві статті, UWAR-GAN та AN-GAN, стаття UWAR-GAN виявляється кращим вибором на основі кількох критичних критеріїв: методологічної інновації, глибини оцінки, практичної ефективності та стійкості до різних умов шуму.

Підхід UWAR-GAN демонструє значну новизну, інтегруючи як амплітудні, так і фазові спектри у процес видалення шуму, що часто ігнорується в інших підходах на основі GAN. Цей двоспектральний підхід підвищує здатність моделі зберігати важливі деталі сигналу, особливо в умовах низького SNR, де фазова інформація відіграє важливу роль у збереженні аудіо якості. Крім того, архітектура мережі включає генератор U-net зі з'єднаннями пропуску, дискримінатор патчів та втрати кореляції змагального характеру, що узгоджує шумні та чисті спектри, досягаючи кращих кількісних та якісних результатів.

Оціночні метрики, такі як SNR, SSIM, RMSE та MOS, вказують на явну перевагу UWAR-GAN. Він досягає покращення SNR на 45,1%, SSIM на 40,7%, MOS на 64,3% та RMSE на 61,1% порівняно з традиційними статистичними методами, такими як фільтрація Вінера. Порівняно з іншими моделями на основі GAN, UWAR-GAN все ще перевершує їх з покращеннями на 26,5%, 39,1%, 7,6% та 35,3% за цими метриками відповідно. Ці результати підкреслюють його стійкість та адаптивність до різних умов шуму, включаючи екстремальні умови низького SNR.

На відміну від цього, метод зниження шуму на основі DPRNN

зосереджується на внутрішньо- та міжчасткових часових залежностях, використовуючи рекурентні нейронні мережі. Хоча він ефективний, його залежність від обробки послідовностей обмежує його адаптивність у захопленні тонких спектральних деталей, особливо у присутності фазових спотворень. Відсутність явного моделювання фази є помітним недоліком порівняно з комплексним підходом UWAR-GAN.

Крім того, UWAR-GAN ретельно оцінюється за допомогою реальних та синтетичних наборів даних, включаючи сценарії з невідомими типами суден та кількома моделями UWAN. Його здатність узагальнювати ці умови демонструє вищий рівень стійкості порівняно з моделлю DPRNN, яка показує зниження продуктивності на невідомих даних.

Загалом, UWAR-GAN виділяється завдяки своїй методологічній інновації, надійній архітектурі, комплексній оцінці та стабільній продуктивності в різних умовах підводного шуму, що робить його кращим вибором для завдань видалення підводного акустичного шуму.

## **Висновки до розділу 1**

Аналіз акустичних сигналів водного середовища є складним завданням через значну варіативність шумів, різноманітність джерел сигналів та їхню мінливість у просторі та часі. Було здійснено огляд умов та методів використання машинного навчання для аналізу акустичних сигналів морського середовища, зокрема методів на основі нейронних мереж. Визначено основні етапи аналізу інформації з використанням нейромережевих підходів, а також досліджено та обґрунтовано ефективні методи обробки акустичних сигналів морського середовища за допомогою нейронних мереж.

Існуючі методи, такі як статистичні моделі та класичні алгоритми машинного навчання, не завжди забезпечують достатню точність у

зашумлених умовах морського середовища. Використання нейронних мереж дозволяє суттєво підвищити точність класифікації та ідентифікації об'єктів завдяки здатності адаптуватися до складних нелінійних закономірностей у даних. Розробка та вдосконалення методів аналізу акустичних сигналів на основі нейронних мереж є критичною для вирішення задач класифікації та виявлення сигналів у реальних умовах морського середовища.

Наукове завдання полягає в покращенні методів аналізу акустичних сигналів водного середовища, зокрема у розробці ефективнішого методу класифікації на основі нейромережі CNN-SOP. Також перспективним напрямом є створення методу аналізу акустичних сигналів, зашумлених фоновими шумами, із використанням поєднання нейромережі UWAR-GAN для придушення шумів та CNN-SOP для класифікації очищених сигналів.

Розробка прототипу програмного забезпечення дозволить перевірити та валідувати запропоновані методи, оцінити їхню ефективність у реальних умовах. Важливим науковим завданням є розробка архітектури програмного засобу для аналізу акустичних сигналів водного середовища на основі нейронних мереж.

Створення тестових наборів даних із різними типами шумів є необхідною умовою для навчання та оцінки моделей, а також підвищує їхню стійкість до реальних умов. У межах дослідження слід виконати такі наукові завдання: створити тестові набори даних для оцінки функціонування програмного засобу аналізу акустичних сигналів водного середовища та провести тестування запропонованих методів для перевірки їхньої ефективності в умовах реального середовища.

## РОЗДІЛ 2. ФОРМАЛІЗАЦІЯ ТА МАТЕМАТИЧНЕ ПРЕДСТАВЛЕННЯ МЕТОДІВ АНАЛІЗУ АКУСТИЧНИХ СИГНАЛІВ ВОДНОГО СЕРЕДОВИЩА

### 2.1. Формалізація задачі класифікації акустичних сигналів водного середовища

#### 2.1.1. Постановка задачі класифікації акустичних сигналів

- $X$  — простір усіх можливих вхідних акустичних сигналів, де кожен сигнал  $x \in X$  може бути представлений як хвильова форма у часовій області або як часово-частотне представлення;
- $Y$  — простір можливих вихідних класів, де кожен  $y \in Y$  і відповідає певній категорії.

Задача полягає у побудові функції класифікації:

$$f : X \rightarrow Y, \quad (2.1)$$

яка відображає вхідний акустичний сигнал  $X$  у відповідну мітку класу  $Y$ .

У реальних підводних середовищах акустичні сигнали рідко записуються умовах мізерно малої присутності фонових шумів. Вони майже завжди містять шуми від різних природних і техногенних джерел. Цей зв'язок можна математично виразити як:

$$x_{\text{noisy}} = x_{\text{clean}} + n, \quad (2.2)$$

де  $x_{\text{clean}}$  — чистий акустичний сигнал, що є ідеальним представленням без шумів,  $n$  — фоновий шум, що включає шуми середовища, технічні перешкоди або інші сигнали,  $x_{\text{noisy}}$  — спостережуваний зашумлений сигнал, який є комбінацією чистого сигналу та шуму.

Задача полягає в оптимізації функції втрат, що мінімізує помилку класифікації. При тренуванні даної нейронерережі була використана крос-ентропія [71]:

$$L_{\text{class}}(\Theta) = -\frac{1}{N} \sum_{i=1}^N \sum_{y \in Y} \mathbf{1}_{y_i=y} \log P(y | x_i; \Theta), \quad (2.3)$$

де  $\Theta$  — параметри моделі,  $\mathbf{1}_{y_i=y}$  — індикаторна функція, що дорівнює 1, якщо передбачена мітка відповідає справжній,  $P(y | x_i; \Theta)$  — ймовірність того, що сигнал належить до класу  $y$ , за заданими параметрами моделі.

Втрати перехресної ентропії вимірюють різницю між істинними мітками класів та передбаченими ймовірностями класів з шару softmax. Вона карає неправильні класифікації, порівнюючи логарифм передбачених ймовірностей з фактичними мітками. Втрати обчислюються шляхом підсумовування всіх класів і усереднення по всьому набору даних. Нижчі значення цієї функції втрат свідчать про кращу роботу моделі.

Щоб знайти оптимальні параметри моделі  $\Theta$ , необхідно мінімізувати функцію втрат  $L(\Theta)$ . Це зазвичай робиться за допомогою алгоритмів оптимізації, таких як стохастичний градієнтний спуск [72] або Adam [73]. Ці алгоритми обчислюють градієнт функції втрат відносно параметрів  $\Theta_i$  оновлюють параметри ітеративно:

$$\Theta_{i+1} = \Theta_i - \eta \nabla_{\Theta} L(\Theta_i), \quad (2.4)$$

де  $\eta$  - це швидкість навчання, а  $\nabla_{\Theta} L(\Theta_i)$  - це градієнт функції втрат відносно  $\Theta$ .

Для тренування цієї нейронмережі був застосований оптимізатор Adam. Оптимізатор Adam поєднує в собі імпульс (який допомагає згладжувати оновлення) та адаптивну швидкість навчання (яка масштабує оновлення на основі останніх градієнтів). Вона підтримує експоненціально спадаючі середні значення як минулих градієнтів, так і їх квадратів, коригує зсув і оновлює



параметри, використовуючи ці уточнені оцінки. Цей підхід покращує швидкість збіжності та стабільність моделей глибокого навчання. Тренування Adam відбувається наступним чином

1. Обчислити градієнт функції втрат відносно параметрів моделі  $\theta_t$ :

$$g_t = \nabla_{\theta} L(\theta_t), \quad (2.5)$$

де  $g_t$  градієнт функції втрат  $L$  на часовому кроці  $t$ .

2. Оновити зміщену оцінку першого моменту (експоненціально спадаюче середнє минулих градієнтів):

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t, \quad (2.6)$$

де  $m_t$  оцінка першого моменту (ковзне середнє градієнтів),  $\beta_1$  контролює швидкість спадання минулих градієнтів (за замовчуванням: 0.9).

3. Оновити зміщену оцінку другого моменту (експоненціально спадаюче середнє квадратів градієнтів):

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2, \quad (2.7)$$

де  $v_t$  оцінка другого моменту (ковзне середнє квадратів градієнтів),  $\beta_2$  контролює швидкість спадання квадратичних градієнтів (за замовчуванням: 0.999).

4. Корекція зсуву для першого та другого моментів:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t}, \quad (2.8)$$

де  $\hat{m}_t$  та  $\hat{v}_t$  оцінки з поправкою на зсув для  $m_t$  та  $v_t$ , відповідно.

## 5. Оновлення параметрів моделі:

$$\theta_{t+1} = \theta_t - \alpha \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \delta}, \quad (2.9)$$

де  $\theta_t$  параметри моделі на часовому кроці  $t$ .,  $\alpha$  швидкість навчання,  $\delta$  невелика константа (за замовчуванням:  $g_t^2$ ) для запобігання діленню на нуль.

### 2.1.2. Представлення акустичного сигналу в різних областях

Акустичні сигнали у часовому домені відображають їх зміни з плином часу через амплітуду, фазу та часовий профіль. Амплітуда визначає інтенсивність сигналу, фаза — стадію коливань, а часовий профіль демонструє зміну амплітуди з часом. Часові характеристики дозволяють аналізувати еволюцію сигналу, але мають обмеження: вони можуть бути нечутливими до дрібних змін, не враховувати частотний склад і чутливі до шумів [74, 75].

Частотні характеристики описують спектральний склад сигналу через амплітуду залежно від частоти, частотний діапазон, піки та смуги. Вони показують інтенсивність сигналу на різних частотах, але обмежені вузьким діапазоном, нечутливістю до дрібних змін і впливом шуму [76].

Часово-частотні характеристики поєднують динаміку змін у часі та частоті, дозволяючи аналізувати спектральний склад сигналу протягом часу. Спектрограми надають візуальне уявлення змін частот і амплітуд, що робить їх цінним інструментом для аналізу акустичних явищ у водному середовищі, реалістичної генерації сигналів та досліджень їх властивостей [77].

### 2.1.3. Метрики оцінки результатів класифікації акустичних сигналів водного середовища

Продуктивність моделі класифікації часто оцінюється за допомогою різних метрик, залежно від застосування та специфічних характеристик даних підводних сигналів. Загальні метрики включають:

- Ассурасу [78]: Частка правильних передбачень від загальної кількості передбачень.
- Precision [79], повнота [80] та F1-оцінка [81]: Особливо корисні при роботі з незбалансованими наборами даних або специфічними класами сигналів, які важливіше виявити.
- Матриця плутанини [82]: Матриця, яка візуалізує продуктивність моделі класифікації, показуючи справжні позитиви, справжні негативи, хибні позитиви та хибні негативи.

Ассурасу (Точність класифікації) — це метрика, що показує відсоток правильно класифікованих зразків серед усіх передбачень моделі. Вона є базовим і часто використовуваним показником для оцінки ефективності класифікаторів, особливо у випадках, коли дані збалансовані, і кількість зразків у кожному класі приблизно однакова.

$$F_1 = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)}, \quad (2.10)$$

де  $TP$  (True Positives) — кількість правильно передбачених позитивних прикладів,  $TN$  (True Negatives) — кількість правильно передбачених негативних прикладів,  $FP$  (False Positives) — кількість негативних прикладів, помилково класифікованих як позитивні,  $FN$  (False Negatives) — кількість позитивних прикладів, помилково класифікованих як негативні.

Перевагою Ассурасу є простота розрахунку та інтерпретації, що легко обчислюється та зрозуміла навіть для користувачів без глибоких знань у машинному навчанні.

- Універсальність застосування: Підходить для широкого спектра завдань класифікації.

- Загальна оцінка продуктивності: Дає швидкий огляд того, наскільки добре модель працює на всьому наборі даних.

Precision (Точність) — це міра, яка показує, яка частка об'єктів, передбачених як позитивні, насправді є позитивними. Вона особливо важлива в задачах, де помилкове виявлення позитивного класу (False Positive) може мати серйозні наслідки, наприклад, у медицині або системах безпеки.

$$Precision = \frac{TP}{(TP + FP)}, \quad (2.11)$$

де  $TP$  (True Positives) — кількість правильно передбачених позитивних випадків,  $FP$  (False Positives) — кількість негативних випадків, помилково класифікованих як позитивні.

До переваг Precision можна віднести зменшення помилкових спрацювань. Вона допомагає мінімізувати кількість хибно позитивних результатів. Також, Precision є ключовою для завдань, де важливішою є якість позитивних прогнозів, ніж їх кількість. У поєднанні з Recall, Precision дозволяє краще зрозуміти продуктивність моделі, особливо в умовах незбалансованих даних.

Recall (Повнота) — це метрика, що показує, яка частка реальних позитивних випадків була правильно ідентифікована моделлю. Вона є важливою в задачах, де критично мінімізувати кількість пропущених позитивних випадків (False Negatives), наприклад, у медичній діагностиці або системах виявлення загроз.

$$Recall = \frac{TP}{(TP + FN)}, \quad (2.12)$$

де  $TP$  (True Positives) — кількість правильно передбачених позитивних випадків,  $FN$  (False Negatives) — кількість позитивних випадків, помилково класифікованих як негативні.

$Recall$  передбачає меншу кількість пропущених позитивних випадків. Це допомагає мінімізувати кількість помилково класифікованих позитивних зразків як негативні. Вона є важливою для сценаріїв, де пропуск позитивного випадку може мати серйозні наслідки. Також,  $Recall$  добре підходить для оцінки моделей, коли позитивний клас значно менший за негативний.

$F1$ -міра є показником для оцінки продуктивності моделі класифікації, особливо у випадках, коли дані є незбалансованими, і важливо знайти баланс між точністю та повнотою.  $F1$ -міра є саме гармонійним середнім між точністю та повнотою і обчислюється за наступною формулою:

$$F_1 = 2 \times \frac{(Precision \times Recall)}{(Precision + Recall)}. \quad (2.13)$$

Переваги  $F1$ -міри полягають у збалансованій оцінці, оскільки вона враховує як  $Precision$ , так і  $Recall$ , забезпечуючи більш об'єктивну оцінку продуктивності моделі. Вона є особливо корисною для незбалансованих даних, оскільки залишається ефективною навіть тоді, коли один клас значно переважає інший. Крім того,  $F1$ -міра стійка до компромісів між  $Precision$  і  $Recall$ , надаючи рівну вагу обом метрикам і допомагаючи уникнути ситуацій, коли одна з них суттєво домінує.

Для класифікації підводних сигналів також важливо оцінювати стійкість моделі до шуму та інших умов навколишнього середовища. Метрики, такі як  $SNR$  та крива приймача-оператора ( $ROC$ ) [83], можуть використовуватися для оцінки продуктивності за різних рівнів шуму та умов.

#### **2.1.4. Можливі виклики при виконанні задачі класифікації акустичних сигналів водного середовища**

У класифікації підводних акустичних сигналів може виникнути ряд викликів. Підводні середовища за своєю природою шумні через наявність як природних звуків, таких як природні, так і штучних джерела. Моделі класифікації сигналів повинні бути стійкими до таких типів шуму та інтерференції. Отримання позначених підводних акустичних даних може бути дорогим і складним. Техніки збільшення даних та генерації синтетичних даних часто використовуються для покращення продуктивності моделі. Також, підводні сигнали можуть значно змінюватися залежно від умов навколишнього середовища: глибини, температури та солоності, що ускладнює завдання класифікації цих сигналів.

#### **2.2. Формалізація задачі придушення фонових шумів в акустичних сигналах водного середовища**

Підводні акустичні сигнали є важливими для роботи морських апаратів, включаючи комунікацію, навігацію та моніторинг навколишнього середовища. Проте ці сигнали часто спотворюються фоновим шумом, який виникає через морське життя, хвилі та антропогенні фактори, що створює труднощі для точного аналізу сигналів [84, 85]. Традиційні методи придушення шуму, такі як фільтрація часто є недостатніми у високо динамічних підводних умовах. Нейронні мережі виявилися надійною альтернативою, оскільки вони здатні адаптивно навчатися та вилучати ознаки із зашумлених даних [86, 87]. Цей розділ досліджує математичну основу придушення шуму за допомогою нейронних мереж та їх застосування у морських апаратах.

### 2.2.1. Представлення проблеми задачі придушення фонових шумів в акустичних сигналах водного середовища

Прийнятий підводний акустичний сигнал  $r(t)$  може бути змодельований як:

$$r(t) = s(t) + n(t), \quad (2.14)$$

де  $r(t)$ : Прийнятий сигнал,  $s(t)$ : Бажаний акустичний сигнал,  $n(t)$ : Фоновий шум.

Мета — відновити чистий сигнал  $s(t)$  із  $r(t)$ . Нейронні мережі забезпечують функцію відображення  $f_\theta$ , параметризовану вагами мережі, для оцінки  $s^{(t)}$ :

$$s^{(t)} = f_\theta(r(t)). \quad (2.15)$$

Це відображення досягається шляхом навчання мережі мінімізувати різницю між  $s(t)$  та  $s^{(t)}$  на наборі прикладів.

### 2.2.2. Загальне представлення архітектури нейронної мережі

Системи на основі нейронних мереж для придушення підводного шуму зазвичай використовують архітектури, такі CNN або моделі на основі Transformer. Однією з поширених архітектур є згортковий автоенкодер:

1. Кодувальник витягує ієрархічні особливості із зашумленого вхідного сигналу.
2. Боттлнек кодує особливості сигналу у стислому форматі.
3. Декодувальник відновлює чистий сигнал із прихованого подання.

Математично процес можна представити як:

$$s^{(t)} = D(E(r(t); \theta_E); \theta_D), \quad (2.16)$$

де  $E$  та  $D$  — це відповідно функції кодувальника та декодувальника, а  $\theta_E$  та  $\theta_D$  — їх параметри.

### **2.2.3. Функція втрат, подання вхідних даних та вибір алгоритму оптимізації**

Зашумлений сигнал  $r(t)$  часто перетворюють у часово-частотну область, наприклад у спектрограму  $R(f, t)$ , для кращого витягання особливостей. Спектрограма обчислюється як:

$$R(f, t) = |F(r(t))|^2, \quad (2.17)$$

де  $F$  позначає STFT [88].

Параметри нейронної мережі  $\theta$  можуть бути оптимізовані оптимізуються за допомогою методів, що базуються на градієнті. В даному випадку був застосований алгоритм Adam.

### **2.2.4. Опис алгоритму навчання та функцій втрат нейромережі для придушення шумів**

В рамках тренування алгоритмом GAN присутні два основні компоненти - генератор та дискримінатор. Генератор - це нейронна мережа, яка генерує екземпляри даних, а дискримінатор намагається визначити їхню достовірність. Модель дискримінатора виявляє, чи є екземпляр даних справжнім або фальшивим. Модель генератора намагається обдурити дискримінатор і тренується на більшій кількості даних, щоб отримати правдоподібні результати. Ця архітектура є змагальною, оскільки генератор і дискримінатор працюють один проти одного з протилежними цілями.

Загальна функція втрат складається з кількох функцій:  $L_{cGAN}$ ,  $L_{gen}$ ,  $L_{corr}$ . Основною метою cGAN є гра в гру між двома гравцями - генератором



G та дискримінатором D. Функція втрат cGAN спрямована на мінімізацію змагальної функції втрат LcGAN, яка формується на основі концепції гри двох гравців. Генератор G намагається створити вихідні дані, які важко розрізнити для дискримінатора D від реальних чистих даних X. Функція втрат визначена в рівнянні 2.42 і включає дві частини: одну, пов'язану з реальними даними, і іншу, пов'язану зі створеними даними.

$$\min_G \max_D L_{cGAN}(G, D) = E_{X,Y} [\log D(X, Y)] + E_{X,Y} [\log(1 - D(X, G(Y)))] \quad (2.18)$$

де  $G$  – результат роботи генератора,  $D$  – результат роботи дискримінатора,  $Y$  – зашумлений вхід,  $G(Y)$  – фальшивий вихід,  $X$  – істинне значення.

Втрата збереження генератора  $L_{gen}$  була використана для поліпшення якості виходів генератора вводиться додатковий термін втрати  $L_{gen}$ . Ця втрата спрямована на забезпечення збереженості між виходами генератора та чистими цільовими даними.

$$L_{gen} = E_{X,Y} [\log(1 - D(X, G(Y)))] \quad (2.19)$$

Функція втрат  $L_{gen}$ , використовує L1-норму для вимірювання різниці між виходом генератора та чистими даними. Це допомагає покращити якість створених виходів та уникнути візуальних артефактів.

Була використана функція втрат на основі кореляції,  $L_{corr}$ , щоб використовувати кореляційні зразки, присутні в часово-частотних представленнях підводної акустики. Ця втрата використовує кореляційну матрицю між шумовими та чистими зображеннями у частотному домені. Кореляційна матриця вказує на відповідність між різними частотними компонентами.

$$R_{X,Y}(i, j) = \frac{Cov(X_i, Y_j)}{\sigma_{X_i} \sigma_{Y_j}} \quad (2.20)$$

де  $\sigma_{X_i}\sigma_{Y_j}$  - стандартні відхилення для  $X$  та  $Y$ ,  $Cov(X_i, Y_j)$  - коваріаційна матриця.

Визначена кореляційну матрицю, і вона використовується для обчислення втрати на основі кореляції  $L_{\text{corr}}$ . Ця втрата спонукає генератор точно відтворювати бажані частотні компоненти.

$$L_{\text{corr}}(G) = E_{X,Y} \| R_{X,Y} - R_{G(Y),Y} \|_1, \quad (2.21)$$

Завершальна цільова функція, як вказано в 2.46, комбінує ці три втрати -  $L_{\text{cGAN}}$ ,  $L_{\text{gen}}$  і  $L_{\text{corr}}$  - з регуляризаційними коефіцієнтами  $\gamma$  і  $\lambda$ . Значення цих коефіцієнтів налаштовуються під час навчання з урахуванням важливості кожної складової втрати. В результаті, для перших 100 епох значення  $\gamma$  і  $\lambda$  були рівні 1. Після 100 епох, значення для коефіцієнтів були підвищені - значення  $\gamma$  дорівнювало 15 і  $\lambda$  дорівнювало 10. Остаточна функція втрат є сумою попередньо описаних функцій втрат та визначається за формулою:

$$L = L_{\text{cGAN}}(G, D) + \gamma L_{\text{gen}}(G) + \lambda L_{\text{corr}}(G), \quad (2.22)$$

### 2.2.5. Метрики оцінки придушення фонових шумів водного середовища

Ефективність придушення шуму на основі нейронних мереж вимірюється за допомогою таких метрик, як SNR [89], RMSE [90], стандартне відхилення [91], PSNR [92].

SNR характеризує співвідношення потужності корисного сигналу до потужності шуму. Цей показник дозволяє оцінити, наскільки домінує сигнал над шумом у системі, і зазвичай виражається в децибелах (дБ).

$$SNR = 10 \cdot \log_{10} \left( \frac{P_{\text{signal}}}{P_{\text{noise}}} \right), \quad (2.23)$$

де  $P_{\text{signal}}$  – потужність корисного сигналу,  $P_{\text{noise}}$  – потужність шуму

RMSE — це міра, що використовується для оцінки середньої величини помилки між прогнозованими або відновленими значеннями та істинними значеннями сигналу. RMSE визначається як квадратний корінь із середньоквадратичної похибки (MSE). Формула для RMSE:

$$RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^N (x(n) - \hat{x}(n))^2}, \quad (2.24)$$

де  $x(n)$  — істинне значення сигналу,  $\hat{x}(n)$  — відновлене або прогнозоване значення сигналу, а  $N$  — загальна кількість дискретних відліків.

PSNR — це метрика, яка використовується для оцінки якості відновленого або обробленого сигналу шляхом порівняння його з оригінальним сигналом. Вона визначається через відношення максимальної можливої потужності сигналу до потужності шуму, який спотворює цей сигнал. Формула для розрахунку  $PSNR$  у випадку дискретного сигналу визначається як:

$$PSNR = 10 \cdot \log_{10} \left( \frac{MAX^2}{MSE} \right), \quad (2.25)$$

де  $MAX$  — максимальне можливе значення сигналу, а  $MSE$  — середньоквадратична похибка між оригінальним сигналом  $x(n)$  і спотвореним сигналом  $\hat{x}(n)$ , що обчислюється за формулою:

$$MSE = \frac{1}{N} \sum_{n=1}^N (x(n) - \hat{x}(n))^2, \quad (2.26)$$

де  $N$  — кількість дискретних відліків сигналу. PSNR виражається в децибелах (dB) і використовується для оцінки втрат інформації після обробки сигналу, зокрема в задачах шумозаглушення або компресії.

Стандартне відхилення  $\sigma$  характеризує розсіювання значень навколо середнього. Воно обчислюється як квадратний корінь із середнього значення квадратів відхилень від середнього значення:

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}, \quad (2.27)$$

де

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i, \quad (2.28)$$

де  $x_i$  – значення сигналу (або виміряні дані),  $N$  – кількість точок даних, а  $\mu$  – середнє значення сигналу.

Метрики перцептивної якості, наприклад, PESQ для суб'єктивної оцінки [93], не були застосовані при оцінці результатів.

## **2.3. Математичне представлення згорткової нейромережі SOP**

### **2.3.1. Опис попередньої обробки вхідних характеристик до нейромережі**

Для виділення ознак у підводних нестационарних сигналах ефективнішим є підхід у часово-частотній області. CQT дозволяє перетворювати сигнал із часової області в часово-частотну область з високою частотною роздільною здатністю на низьких частотах, що надає більше деталей в порівнянні з STFT. У даному дослідженні CQT використовується для отримання часово-частотних ознак, які подаються на вхід запропонованої згорткової моделі. У цій роботі пропонується поєднання багатомасштабної згортки та SOP для захоплення часових кореляцій в ознаках, що покращує можливості розрізнення різних цілей за допомогою вхідних даних CQT.

Для дискретного сигналу в часовому домені, CQT визначається наступним чином:

$$X^{CQ}(k, n) = \sum_{j=n-|N_k/2|}^{n+|N_k/2|} x(j) a_k^*(j - n + N_k / 2), \quad (2.29)$$

де  $k=1,2,\dots, K$  представляє  $K$  частотних бінів CQT, а  $a_k^*(n)$  є комплексним спряженням базисної функції.  $N_k$  позначає довжину вікна, яка є змінною.

Центральна частота  $k$ -го біна визначається як:

$$f_k = f_1 2^{\frac{k-1}{B}}, \quad (2.30)$$

де  $f_1$  є центральною частотою найнижчого частотного біна, а  $B$  - кількість бінів на октаву, що визначає компроміс між часовою та частотною роздільною здатністю CQT. У цьому дослідженні пропонується застосувати CQT для отримання часово-частотних ознак, які використовуються в CNN. Ці ознаки обчислюються з множини кадрів за наступним виразом:

$$K = B \left( \log_2 \frac{f_{max}}{f_1} + 1 \right), \quad (2.31)$$

де  $f$  є центральною частотою найвищого частотного біна.

Пропонується використовувати CQT для отримання часово-частотних ознак для моделі CNN. Часово-частотна ознака CQT отримується наступним чином:

$$X = \{X^1, X^2, \dots, X^N\}, \quad (2.32)$$

де

$$X^i = 20 |X^{CQ}(i)|, \quad (2.33)$$

де  $X^i$  є ознакою CQT для  $i$ , а  $N$  позначає загальну кількість.  $X^{CQ}(i) \in R^K$  є комплексним вектором CQT з  $K$  частотних бінів, що представляє  $i$ -ту характеристику.

### 2.3.2. Покращений метод класифікації акустичних сигналів водного середовища на основі нейромережі CNN SOP

Часово-частотне представлення акустичних сигналів демонструє сильні часові кореляції, що допомагають розрізняти різні цілі. У запропонованому підході пропонується метод SOP для ознак CNN, щоб захопити ці часові кореляції з вхідних даних CQT.

Карти ознак CNN  $H^L$  отримані з ознак CQT  $X$ , де  $h$  та  $w$  представляють частотні смуги та часові кадри вхідних даних CQT. Для кожної частотної смуги траєкторія часової ознаки  $m$ -ої карти ознак позначена як

$$s^m = [s_1^m, s_2^m, \dots, s_w^m] \in R^w, \quad (2.34)$$

Покращений метод передбачає застосування багатомасштабної згортки замість звичайних згорткових шарів, що дозволяє ефективніше вилучає потрібні характеристики у вхідних даних, та вилучаючи характеристики різних масштабів. Застосування середнього пулінгу забезпечило зменшення розмірності вхідних даних. Додатково, повнозв'язний шар з функцією активації ReLU та пакетною нормалізацією забезпечив поліпшення узагальнювальної можливості та гнучкості нейромережі.

Таблиця 2.1. Архітектура покращеної нейромережі CNN SOP

Номер шару	Назва шару	Розмір вхідного тензора
1	CQT input	(64× 64)
2	Multiscale conv (3x3, 5x5, 7x7)	(3x3, 5x5, 7x7)
3	Avg. pooling	
4	SOP	16× 16× 16
5	Norm.	1024
6	Dense	1024
7	Softmax	5

Шар MultiScaleConv застосовує три згорткові фільтри з різними розмірами ядер (3x3, 5x5 і 7x7) паралельно, після чого над результатом виконується конкатенація та пакетна нормалізація.

Кожна згортка застосовує ядро до вхідного тензора для вилучення ознак. Розмір ядра визначає область вхідних даних, яку дивиться фільтр, тоді як крок (stride) і заповнення (padding) визначають, як ядро переміщується по вхідним даним:

$$x_{ij} = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} w_{mn} \cdot x_{i+m, j+n} + b, \quad (2.35)$$

де  $x_{ij}^{(k)}$  — вхідне значення на позиції  $(i, j)$ ,  $x_{i+m, j+n}$  — значення ядра (фільтра) згортки для фільтра розміру  $(i+m, j+n)$ ,  $w_{mn}^{(k)}$  — значення ядра (фільтра) згортки для фільтра розміру  $k \times k$ , де  $k \in 3, 5, 7$ ,  $b^{(k)}$  — зміщення (bias) для відповідного фільтра розміру  $k \times k$ ,  $k$  — розмір ядра згортки.

Далі відбувається конкатенація, що дозволяє об'єднати вилучені характеристики в один тензор:

$$x = [x1, x2, x3]. \quad (2.36)$$

Після конкатенації йде нормалізація за мініпакетами:

$$x_{ij} = \frac{x_{ij} - \mu}{\sqrt{\sigma^2 + \delta}}, \quad (2.37)$$

де  $\mu$  — середнє значення,  $\sigma^2$  — дисперсія, а  $\delta$  — мала константа для стабільності.

Отримані характеристики подаються на середній пулінг для зменшення розмірності вхідних даних, водночас зберігаючи важливу інформацію:

$$y_{ij} = \frac{1}{k^2} \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} x_{i+m, j+n}, \quad (2.38)$$

де  $x_{ij}$  — вхідне значення на позиції  $(i, j)$ ,  $y_{ij}$  — вихід після пулінгу,  $k$  — розмір вікна (наприклад,  $k=2$ ).

Оператор пулінгу другого порядку визначається як:

$$SOP(s^j, s^k) = \sum_{i=1}^w s_i^j s_i^k, \quad (2.39)$$

де  $SOP(s_i^j, s_i^k)$  представляє часові кореляції між траєкторіями ознак  $s^j$  та  $s^k$  з  $j$ -ої та  $k$ -ої карт ознак.

Оператор SOP захоплює взаємодії фільтрів згортки протягом часу. Для карт ознак використовується матриця часових ознак  $S \in R^{c \times w}$ , і оператор SOP в матричній формі:

$$a^{(i)} = f_{L+1}(z^{(i)}; W_{L+1}), \quad (2.40)$$

де  $SOP(S) \in R^{c^2}$  є симетричною позитивно визначеною матрицею, яка захоплює часові кореляції для однієї частотної смуги.

На відміну від інших стратегій пулінгу, які сумують кореляції по всьому зображенню, ми зберігаємо результати SOP для всіх частотних смуг, щоб підтримувати частотні відмінності для класифікації. Кінцева ознака SOP складається з  $h$  операторів SOP, що відповідають висоті карт ознак.

Нормалізація значно покращує продуктивність глибоких мереж. Ми застосовуємо квадратний корінь поелементно та нормалізацію до операторів SOP. Оператори SOP перетворюються у вектор  $p \in R^l$ , де  $l = c \times c \times h$ . Цей вектор обробляється через квадратний корінь поелементно:

$$x_{sr} = (\sqrt{x_1}, \sqrt{x_2}, \dots, \sqrt{x_n}), \quad (2.41)$$

та  $l_2$  нормалізацію:



$$x_{norm} = \frac{x}{\|x\|_2} . \quad (2.42)$$

Обчислювальна складність запропонованої стратегії SOP для карт ознак CNN розміру  $h \times w \times c$  складає  $O(hwc^2)$ , що аналогічно білінійному пулінгу, тоді як максимальний пулінг має складність  $O(hwc)$ .

Отриманий вектор об'єднання другого порядку  $z$  потім вводиться в шар softmax для класифікації після щільного шару. Оцінки класів для  $i^{th}$  вибірки  $z^{(i)}$  до категорії  $j$  можна обчислити наступним чином:

$$p(y^{(i)} = j | a^{(i)}; \theta) = \frac{e^{\theta_j^{T_a(i)}}}{\sum_{t=1}^u e^{\theta_t^{T_a(i)}}} . \quad (2.43)$$

## 2.4. Опис нейромережі для придушення шумів UWAR-GAN

### 2.4.1. Математичне представлення нейромереж для придушення фонових шумів водного середовища

Генератор створений на базі архітектури U-net та відповідає за генерацію чистих зразків шляхом фільтрації, тоді як дискримінатор на базі PatchNet відрізняє реальні реальні та очищені зразки. Часово-частотне представлення сигналу проходить через серію блоків енкодера. Кожен блок енкодера складається з згорткового шару, за яким слідує функція нормалізації та активаційна функція leaky ReLU. Після енкодера, отримані характеристики проходять через шар Bottleneck, який представляє стиснуте подання цих даних. Блоки декодера відновлюють чисту спектрограму. Кожен блок декодера складається з деконволюційного згорткового шару, функції нормалізації та активаційних функції ReLU. Після виконання роботи, застосовується skip-connection для поєднання особливостей енкодера з

декодером. Це допомагає фіксувати як низькорівневі, так і високорівневі особливості, сприяючи збереженню важливих характеристик сигналу.

Ця нейронна мережа є архітектурою генератора, схожою на U-net, і призначена для генерації або відновлення зображень (або спектрограм) на основі вхідних даних. Вона складається з енкодера, вузького місця, боттлнека і декодера.

Енкодер складається з чотирьох послідовних згорткових шарів, які поступово зменшують просторові розміри вхідних даних, одночасно збільшуючи кількість каналів. Архітектура нейромережі представлена на рисунку 2.1.

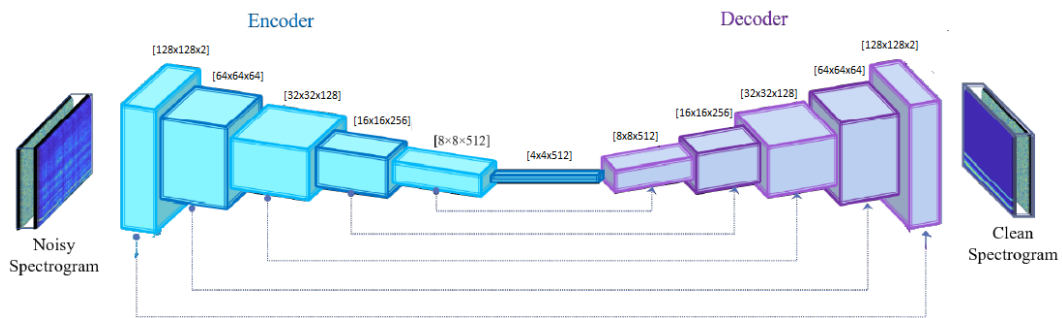


Рис 2.1. - Архітектура генератора U-net [64]

Формула для згорткового шару має наступний вигляд:

$$d_i = f(W_i^T * [d_{i-1}, e_{n-i}] + b_i), \quad (2.44)$$

де  $e_i$  — вихід  $i$ -го шару енкодера,  $W_i$  — ваги згортки,  $m*$  — операція згортки,  $b_i$  — зміщення (bias),  $f$  — функція активації LeakyReLU.

Шар є найглибшим рівнем мережі і застосовує додатковий згортковий шар для обмеження просторового представлення даних:

$$e_b = W_b * e_b + b, \quad (2.45)$$

де  $e_b$  — вихід bottleneck,  $W_b$  — ваги згортки bottleneck,  $b$  — зміщення.

Декодер відновлює дані, використовуючи транспоновані згорткові шари. Кожен шар декодера об'єднується (concatenation) з відповідним шаром енкодера для збереження просторового контексту:

$$d_i = f(W_i^T * [d_{i-1}, e_{n-i}] + b_i), \quad (2.46)$$

де  $d_i$  — вихід  $i$ -го шару декодера,  $W_i^T$  — ваги транспонованої згортки,  $[d_{i-1}, e_{n-i}]$  — конкатенація виходу попереднього декодера і відповідного енкодера,  $b_i$  — зміщення.

Останній транспонований згортковий шар відновлює вихідний розмір і кількість каналів і представлений формулою 2.34:

$$out = W_{out}^T * [d_5, e_1] + b_{out}, \quad (2.47)$$

де  $out$  — вихід мережі,  $W_{out}^T$  — ваги вихідного шару транспонованої згортки,  $b_{out}$  — зміщення.

Ця архітектура дозволяє поєднати локальні та глобальні ознаки даних завдяки механізму *скір-зв'язків* між енкодером і декодером. Модель ефективно реконструює або генерує вхідні дані, зберігаючи важливі деталі та контекст.

Ця нейронна мережа є дискримінатором, що використовується в архітектурі GAN. Дискримінатор класифікує вхідні дані як реальні або синтезовані.

Мережа складається з шести згорткових шарів зі зменшенням розмірності на кожному шарі. Після кожного шару, крім останнього, застосовуються:

- LeakyReLU — функція активації, що допомагає уникнути проблеми "вмирання нейронів".

- Batch Normalization — стабілізує навчання та прискорює його.

Останній шар використовує функцію активації Sigmoid, яка переводить вихід у діапазон  $[0, 1]$ , інтерпретований як ймовірність того, що вхід є реальним. Архітектура нейромережі PatchGAN представлена на рисунку 2.2.

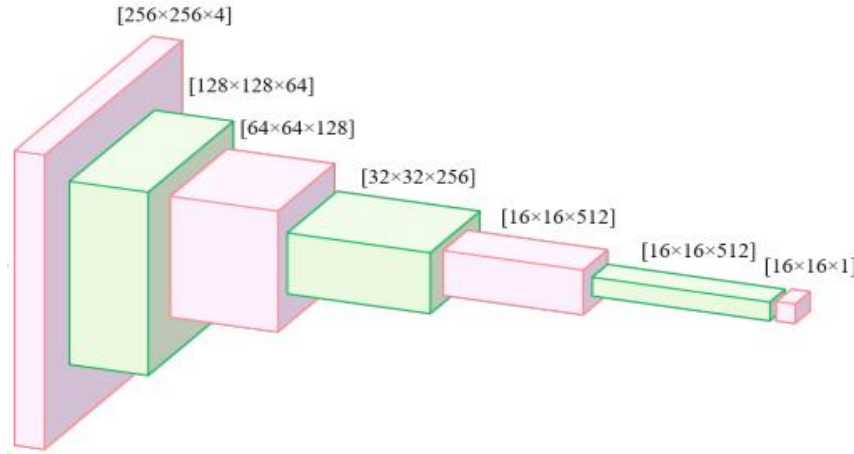


Рис. 2.2. Архітектура лискримінатора PatchGAN [64]

На кожному згортковому шарі операція обчислюється як:

$$f_i = \sigma(W_i * f_{i-1} + b_i), \quad (2.48)$$

де  $f_i$  — вихід  $i$ -го шару,  $W_i$  — ядро згортки,  $*$  — операція згортки,  $b_i$  — зміщення,  $\sigma$  — функція активації LeakyReLU крім останнього шару, де застосовується Sigmoid. Для кожного каналу нормалізація виглядає так:

$$y = \gamma \hat{x} + \beta \quad (2.49)$$

$\gamma$ ,  $\beta$  — параметри навчання.

$$\hat{x} = \frac{x - \mu}{\sqrt{\sigma^2 + \delta}} \quad (2.50)$$

де  $x$  — вхід до нормалізації,  $\mu$  — середнє значення в мініпартії,  $\sigma^2$  — дисперсія в мініпартії,  $\delta$  — мале значення для уникнення ділення на нуль

Останній шар виконує згортку і застосовує функцію Sigmoid:

$$D(x) = \sigma(W_{out} * f_{last} + b_{out}), \quad (2.51)$$

де  $D(x)$  — ймовірність того, що вхід є справжнім,  $\sigma$  — функція активації Sigmoid.

На завершення вихід усереднюється по всіх просторових вимірах:

$$\hat{D}(x) = \frac{1}{N} \sum_{i=1}^N \hat{D}(x)_i, \quad (2.52)$$

де  $N$  — кількість просторових точок (пікселів),  $\hat{D}(x)_i$  — вихід Sigmoid для  $i$ -тої точки.

Вхідний тензор  $x$  передається через послідовність згорткових шарів з активацією LeakyReLU і нормалізацією. Останній шар застосовує активацію Sigmoid. Вихідний тензор усереднюється для отримання скаляра, що представляє ймовірність реальності вхідного зображення.

$$\hat{D}(x) = \frac{1}{N} \sum_{i=1}^N \sigma(W_{\{out\}} * f_{\{last\}} + b_{\{out\}}), \quad (2.53)$$

Дискримінатор забезпечує ефективне навчання шляхом вивчення локальних та глобальних ознак вхідного зображення завдяки каскаду згорток та нормалізації. Він забезпечує зворотний зв'язок для генератора, допомагаючи покращити якість очищених зображень.

#### 2.4.2. Попередня обробка даних методу для придушення шумів

Був створений датасет на базі датасету ShipsEar. Ці сигнали складаються з записів, які включають як бажаний сигнал водних суден, так і небажаний фоновий шум. файли були сегментовані на фрагменти розміром 2 секунди кожен. Були обрані фрагменти з низькою присутністю небажаних шумів. Далі, для відібраних сегментів були обрані шуми, що дозволило отримати набір зашумлених даних з різним співвідношенням SNR. В результаті, отриманий сегмент має високий рівень сторонніх шумів, тоді як початковий семпл представляє підводний сигнал з мінімальною кількістю шумів.

Спочатку, записані аудіосигнали перетворюються в спектрограми за допомогою методів, таких як STFT. Спектрограми надають часово-частотне представлення аудіосигналів, де вертикальна вісь відображає частоту, горизонтальна вісь відображає час, а інтенсивність кольору відображає величину сигналу на кожній точці. Спектрограми розділяються на магнітудні та фазові компоненти. Зашумлені спектрограми складаються в пари з відповідними чистими спектрограмами. Зашумлена спектрограма подається на вхід, тоді як чиста спектрограма представляє реальні дані та подається на дискримінатор для порівняння з результатом роботи генератора. Датасет розділяється на набори для навчання та валідації та у співвідношенні 80% до 20%.

## **Висновки до розділу 2**

Задача класифікації акустичних сигналів формалізується як оптимізація функції втрат для мінімізації помилок моделі, використовуючи методи машинного навчання, включаючи згорткові нейронні мережі. Особливий акцент зроблено на використанні часово-частотного представлення, зокрема CQT, яке забезпечує високу частотну роздільність на низьких частотах. Поєднання багатомасштабних згорткових шарів та пулінгу другого порядку

підвищує точність класифікації завдяки кращому вилученню ознак та врахуванню часових кореляцій.

У задачах придушення шуму використано підхід на основі GAN, де генератор на базі U-net забезпечує відновлення чистого сигналу, а дискримінатор PatchGAN розрізняє реальні та синтетичні дані. Функції втрат, включаючи кореляційну втрату та регуляризацію, сприяють підвищенню якості очищення сигналів навіть за умов сильного шумового забруднення. Результати показують, що інтеграція цих методів у задачі аналізу підводних акустичних сигналів сприяє досягненню високої ефективності у класифікації та придушенні шуму, що відкриває перспективи для використання у морських дослідженнях та прикладних завданнях.

## **РОЗДІЛ 3. ЕКСПЕРИМЕНТАЛЬНА ПЕРЕВІРКА МЕТОДІВ АНАЛІЗУ АКУСТИЧНИХ СИГНАЛІВ ВОДНОГО СЕРЕДОВИЩА**

### **3.1. Розробка датасетів акустичних сигналів водного середовища**

Через складність підводного середовища, особливості морських об'єктів та обмеження, які накладає обладнання, ефективність виявлення може суттєво знижуватися за традиційних підходів. Дослідження показують, що глибоке навчання має значний вплив на різні сфери застосування [94], зокрема на обробку даних морського середовища. Одним із можливих ефективних способів покращення таких даних є використання нейромереж. Задача придушення шумів є досить складною та вимагає великого обсягу якісних даних. Крім того, для навчання нейромереж GAN необхідні датасети, що містять як достовірні, так і спотворені зразки. Однак такі датасети не є широко доступними, що ускладнює їх використання та потребує додаткових зусиль для створення. Отже, ця робота спрямована на аналіз підходів до створення датасетів для водного середовища, визначення найоптимальнішого методу, формування власного датасету та оцінку його ефективності під час тренування нейромережі.

#### **3.1.1. Огляд датасетів акустичних сигналів водного середовища**

Існують різні підходи до створення, обробки та застосування датасетів для водного середовища. У цьому дослідженні здійснено огляд датасетів, які використовувалися для класифікації та придушення шумів, а також оцінено доцільність їх застосування для розв'язання задачі усунення фонових перешкод.

Одним із підходів є створення датасету шляхом збору зразків у реальних умовах за допомогою гідрофона, як це було реалізовано в дослідженні [61]. Такий метод є доволі ресурсозатратним і потребує значного часу, при цьому



отримані результати не завжди відповідають очікуванням. У рамках цього підходу вдалося зібрати велику кількість зразків, однак запропонована модель не продемонструвала високої ефективності при класифікації. Основною проблемою виявилася незбалансованість датасету, що ускладнило поділ суден на відповідні класи. Крім того, для задачі придушення шумів такий набір даних малопридатний, оскільки він не містить пар чистих і зашумлених зразків, необхідних для навчання відповідних моделей.

Інший підхід передбачає створення записів у штучних умовах, як це було зроблено в дослідженні [94]. Було розроблено шість типів гвинтів та створено спеціальне обладнання для симуляції кавітаційного тунелю, що дозволило вимірювати згенеровані шуми. Для запису фонових шумів використовувалася втулка замість справжнього корабельного гвинта. Основною перевагою такого методу є можливість контрольованого запису сигналів для конкретних типів корабельних гвинтів. Однак відтворити всі нюанси реального водного середовища в лабораторних умовах практично неможливо. Тому алгоритми, які тренуються на таких даних, можуть демонструвати низьку точність класифікації при роботі з реальними шумами.

Ще один метод, розглянутий у роботі [55], передбачає використання реальних записів кораблів, які експерти поділили на 12 класів, після чого вони були штучно зашумлені білим шумом із різними значеннями SNR. Цей датасет використовувався для класифікації сигналів водного середовища. Хоча він може бути корисним для тестування нейромереж, штучне зашумлення білим шумом не відповідає реальним акустичним умовам підводного середовища. Внаслідок цього алгоритми, треновані на таких даних, втрачають ефективність при роботі з реальними шумами.

Перспективним підходом є використання датасету Shipsear, як це показано в дослідженні [61]. У цьому випадку шумові записи кораблів із низьким рівнем фонових перешкод були штучно зашумлені відібраними фоновими шумами таким чином, щоб отримати необхідне співвідношення

SNR. У порівнянні з іншими методами цей підхід базується на вже створеному та перевіреному датасеті, на якому було навчено низку моделей, що продемонстрували високі результати. Саме тому він був обраний як основа для створення нового датасету. Для вирішення задачі придушення шумів і класифікації у цьому дослідженні використано датасети Shipsear та Storm.

### **3.1.2. Створення датасетів акустичних сигналів водного середовища**

Для задачі придушення шумів використовується неймережа UWAR-GAN [64]. Її основою є архітектура GAN [95], яка передбачає змагальне тренування генератора та дискримінатора. У ролі дискримінатора застосовується неймережа PatchGAN, вперше представлена в дослідженні [96]. Вона є згортковим бінарним класифікатором, що оцінює якість відфільтрованих зразків. Генератором виступає неймережа U-net [97], яка виконує фільтрацію зашумлених сигналів. Фільтрація реалізується за допомогою згорткового автоенкодера з пропускними з'єднаннями. Вхідними даними є магнітудні та фазові спектрограми зашумлених семплів. Після придушення шумів отримані спектрограми піддаються оберненому швидкому перетворенню Фур'є, у результаті чого формується сигнал із приглушеними фоновими шумами [3].

У дослідженні використовувалися два датасети водного середовища: Shipsear і Storm, причому останній було розроблено в межах науково-дослідницької роботи [98]. У датасеті Storm міститься п'ять класів кораблів, кожен з яких представлений годинним записом акустичних сигналів. Шуми кораблів у цьому датасеті характеризуються низьким рівнем фонових шумів, а велика тривалість записів забезпечує значний обсяг даних для тренування неймережі. Датасет Shipsear містить приблизно 80 суден, поділених на чотири класи. Він також включає супровідний файл із детальним описом кожного запису, що містить інформацію про судно, умови запису та якість

сигналу. Важливою перевагою Shipsear є наявність окремих записів фонових шумів, які відокремлені від шумів кораблів. Представлено чотири типи фонових шумів: потік води, вітер, дощ і удари хвиль. Завдяки цьому можна створювати датасети з різними комбінаціями шумів та моделювати різні акустичні умови.

Під час відбору сигналів із низьким рівнем шуму важливо забезпечити достатню потужність дискретних компонентів корабельного шуму, щоб їхня енергія суттєво перевищувала рівень фонового шуму. Водночас, комбінування таких сигналів із фоновими шумами мало забезпечити достатньо низьке SNR. Усього було відібрано близько 2000 комбінованих зразків із середнім значенням SNR, яке сягало -10 дБ. Однак не всі зразки мали достатню амплітуду, що свідчить про слабкість окремих сигналів. Через це для зразків із низьким SNR відбиралися сигнали з більшою амплітудою, максимальне значення якої перевищувало 0,03[3].

Було згенеровано представлення акустичних сигналів водного середовища, яке включає спектрограму, нормалізовану потужність спектру, логарифмічне представлення спектру та амплітуду сигналу.

Для створення семплів шумів водного середовища використовувався датасет Shipsear, у якому містяться як шуми кораблів, так і фонові шуми водного середовища. Для узгодження з семплами шумів кораблів кожен аудіофайл фонових шумів було сегментовано на відрізки тривалістю 2 секунди[3]. Оскільки початкової кількості шумових зразків виявилось недостатньо, було вирішено комбінувати різні типи шумів між собою. Це дозволило значно збільшити різноманіття шумових варіацій та створити ширший набір комбінацій шумів кораблів із фоновими шумами, що мали низький рівень SNR. Візуалізацію цього представлення наведено на рис. 3.1.

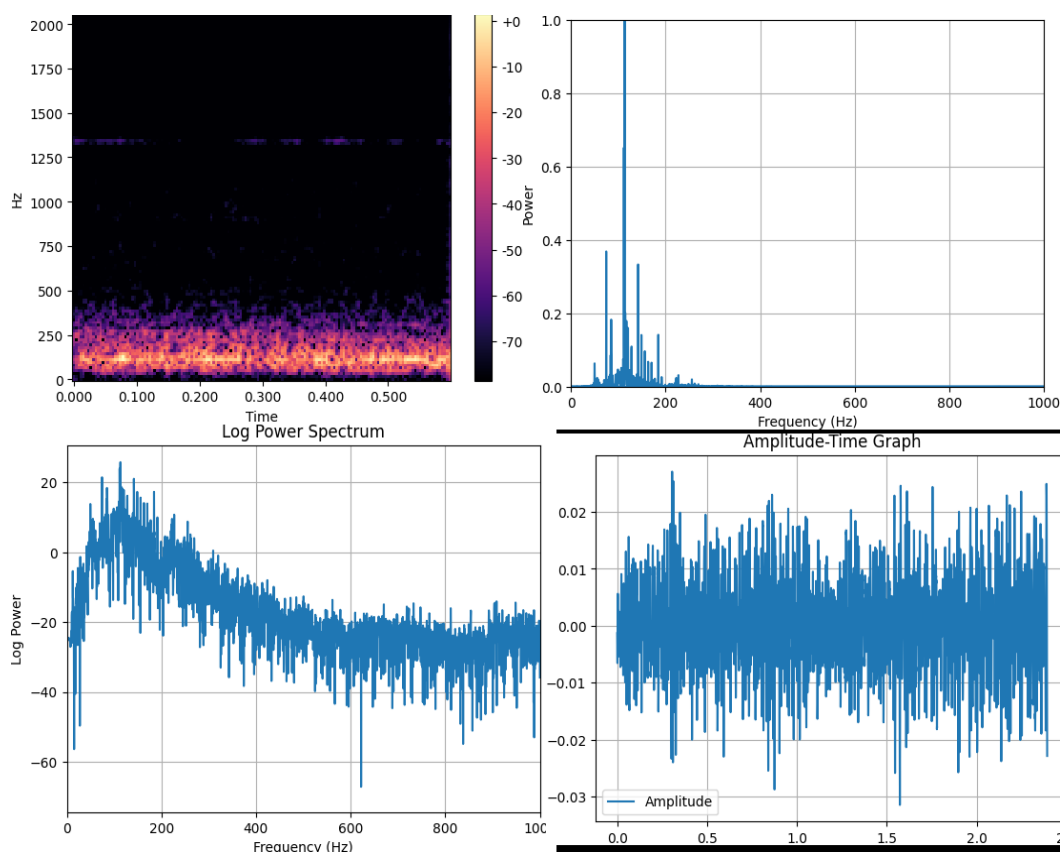


Рис. 3.1. Приклад представлень семпла акустичного сигналу

Для об'єднання чистих і зашумлених зразків, а також комбінування різних шумових компонентів застосовувалася бібліотека AudioSegment, що є частиною модуля PyDub [99]. Її функціональні можливості забезпечують гнучке управління обробкою аудіосигналів. Зокрема, для поєднання шумових компонентів використовувався метод *overlay*, який дозволяє синхронізувати аудіосигнали шляхом одночасного накладання декількох шарів. Такий підхід створює ефект акустичної багатошаровості, що є недосяжним при простому з'єднанні звуків шляхом конкатенації. Створений датасет було оброблено, після чого отримані дані були збережені у вигляді файлів *.pt*, що містять магнітудні та фазові спектрограми, які надалі використовувалися для тренування нейромережі. Алгоритмічне представлення процесу комбінування двох сигналів зображено на рис. 3.2.

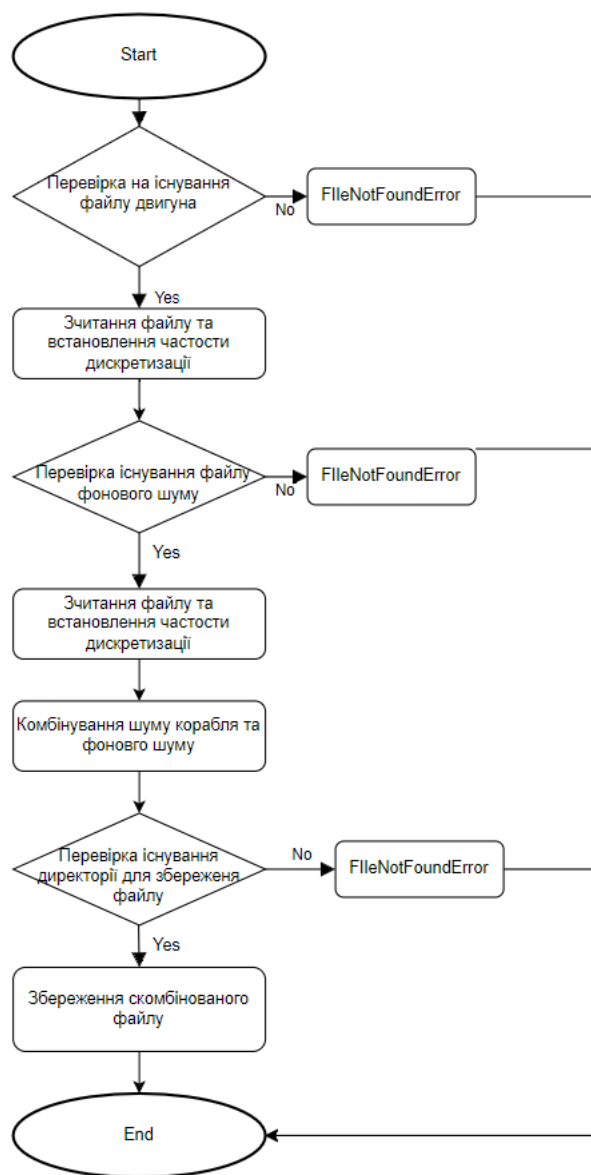


Рис. 3.2. Блок-схема алгоритмічного представлення процесу комбінування двох сигналів

### 3.1.3. Процес навчання нейромережі придушення шумів водного середовища

Навчання нейромережі виконувалося з використанням відеокарти Nvidia GeForce RTX 3070 Mobile та оптимізатора Adam із темпом навчання 0.0002. Тренування проводилося з розміром пакету 2 та тривало близько 12 000 epoch. Загальний час навчання становив близько двох днів. Генератор оновлювався після кожного п'ятого оновлення дискримінатора.

На початкових етапах навчання, до 100 епох, генератор був недостатньо тренуваним для коректного генерування цільових чистих зображень. Відповідно до останніх досліджень, у цей період внесок параметрів регуляризації встановлювався рівним 1 для  $\gamma$  та  $\lambda$ . Після 100 епох, коли генератор досяг достатнього рівня навчання, значення параметрів регуляризації змінювалися:  $\gamma$  збільшувалося до 15, а  $\lambda$  – до 10, що було обрано ґрунтуючись на дослідженнях в цій області [92, 100]. Натренована нейромережа була протестована на семплах, які не використовувалися під час тренування. Отримані результати порівнювалися із зашумленими та чистими зразками корабельних сигналів.

У контексті обробки підводних акустичних сигналів вдосконалення архітектури нейронних мереж відіграє вирішальну роль у покращенні можливостей придушення шуму. Це дослідження розширює попередні методології шляхом оптимізації представлення вхідних даних та архітектури мережі для кращого пристосування до характеристик підводних акустичних сигналів. Оригінальний метод, який використовував частоту дискретизації 16 000 Гц і тензорне представлення 256x256, забезпечував високу роздільну здатність часово-частотного представлення, але накладав значні обчислювальні витрати. Для подальшого розвитку цього підходу було впроваджено оптимізований фреймворк, який використовує зменшену частоту дискретизації до 4,096 Гц, зберігаючи при цьому надійність вилучення спектральних характеристик.

Зменшення частоти дискретизації безпосередньо впливає на частотно-часову репрезентацію, що вимагає модифікацій у перетворенні спектрограми та вхідних розмірах мережі. В рамках цього вдосконалення розмір тензора було адаптовано з 256x256 до 128x128, щоб забезпечити ефективне представлення акустичних особливостей, зменшивши при цьому обчислювальні витрати. Таке коригування ґрунтується на спостереженні, що переважаючі спектральні компоненти підводних сигналів існують у нижчих

частотних діапазонах, що робить початкову високу частоту дискретизації непотрібною для ефективного вивчення особливостей.

На додаток до модифікації представлення вхідних даних, це дослідження вдосконалює архітектуру мережі, оптимізуючи її глибину для узгодження з новими розмірами вхідних даних. Оригінальна модель включала структуру U-Net з п'ятьма шарами дискретизації, де кожен шар поступово зменшував просторові розміри карт ознак. Враховуючи переглянутий розмір тензора, було запроваджено нову адаптацію шляхом зменшення кількості шарів дискретизації з п'яти до чотирьох. Це вдосконалення зберігає важливий ієрархічний процес виділення об'єктів, запобігаючи при цьому надмірному стисненню, яке в іншому випадку може призвести до втрати інформації. Запропонована модифікація підвищує ефективність мережі, підтримуючи оптимальний баланс між збереженням ознак та обчислювальною складністю.

Ця розробка ґрунтується на попередніх дослідженнях у галузі обробки підводних акустичних сигналів, демонструючи, що оптимізована стратегія дискретизації в поєднанні з адаптованою архітектурою мережі може досягти порівнянної ефективності придушення шуму зі зниженими обчислювальними вимогами. Отримані результати підтверджують ефективність адаптації моделей глибокого навчання до специфічних характеристик сигналів, тим самим сприяючи розвитку нейромережових методів придушення підводних шумів. Емпірична перевірка запропонованих удосконалень підтвердила, що переглянута архітектура не тільки зберігає здатність придушення шуму, але й підвищує ефективність обробки, що є значним кроком вперед в еволюції методологій обробки підводних акустичних сигналів.

### **3.2. Підбір гіперпараметрів покращеної згорткової нейромережі CNN-SOP**

Оригінальна нейромережа містила два згорткові шари з функцією активації ReLU, шар SOP, шар нормалізації та вихідний шар. Для покращення її роботи було вирішено застосувати багатомасштабну згортку, що дозволяє ефективніше аналізувати різнорівневі характеристики акустичних сигналів. Такий підхід дає змогу нейромережі адаптивно працювати з різними масштабами деталей у сигналі, що є критично важливим при класифікації акустичних сигналів водного середовища, де частоти можуть значно варіюватися, а шуми ускладнюють процес розпізнавання.

Було протестовано кілька варіантів багатомасштабної згортки з різними розмірами ядер, що включали поєднання згорток розмірами  $3 \times 3$ ,  $5 \times 5$ , розширений варіант із додатковим ядром  $7 \times 7$ , а також ще більш комплексний підхід із включенням ядра  $9 \times 9$ . Параметри інших шарів у нейромережі адаптувалися відповідно до зміненої розмірності даних.

Під час першого експерименту було застосовано дві згортки з ядрами  $3 \times 3$  і  $5 \times 5$ . Перша згортка містила один вхідний канал, вісім вихідних каналів, розмір ядра 3, крок 1 та просторове доповнення 1. Друга згортка мала аналогічну конфігурацію, але з розміром ядра 5 та просторовим доповненням 2. Навчання цієї архітектури показало точність класифікації 96%.

У наступному експерименті була додана третя згортка розміром  $7 \times 7$ , що дозволило отримати точність до 96%. Додавання четвертої згортки розміром  $9 \times 9$  не дало подальшого покращення результатів, оскільки точність не збільшилась. Враховуючи однакові результати першого та другого експериментів, було обрано варіант із двома згортками, оскільки він забезпечував швидше навчання нейромережі.



Таблиця 3.1. Параметри та результати експериментів

Номер експ.	Попередні згортки	Розмір згортки	Розмір ядра	Крок	Просторове доповнення	Точність клас.	Час виконання
1	-	3x3	3	1	1	96%	360
		5x5	5	1	2		
2	3x3, 5x5	7x7	7	1	3	96%	836
3	3x3, 5x5, 7x7	9x9	9	1	4	94%	919

Для подальшого покращення роботи нейромережі було вирішено провести експерименти щодо вибору типу пулінгу після шару багатомасштабної згортки. Застосування пулінгу на цьому етапі може значно підвищити ефективність моделі, оскільки зменшення розмірності вхідних даних спрощує подальшу обробку та знижує обчислювальні витрати. Додатково пулінг допомагає виділити найбільш значущі ознаки, що формуються після згорткових операцій, підвищуючи стійкість моделі до варіацій у вхідних даних та зменшуючи ризик перенавчання.

Було розглянуто два підходи до пулінгу: максимальний пулінг та середній пулінг. Використання MaxPool дозволяє виділяти найбільш виразні особливості сигналу, що робить модель більш стійкою до шуму та варіацій у вхідних даних. Крім того, цей метод знижує обчислювальні витрати за рахунок скорочення кількості оброблюваних параметрів, хоча водночас може втрачати слабкі сигнали, які містять корисну інформацію. У свою чергу, застосування AvgPool сприяє згладжуванню локальних варіацій, що підвищує стабільність виділених ознак і зменшує вплив випадкових шумів. Цей підхід є ефективним у випадках, коли важливо зберегти всю інформацію про область сигналу, оскільки він використовує середнє значення замість вибору найбільшої активації. Водночас він менш ефективний для виявлення найбільш релевантних особливостей у сигналі. Результати обчислень з пакетною нормалізацією представлені в таблиці 3.2.

Таблиця 3.2. Обчислювальні експерименти щодо пакетної нормалізації

Номер експ.	Тип пулінгу	Крок	Просторове доповнення	Точність клас.
1	Максимальний пулінг	2	2	96
2	Середній пулінг	2	2	97

Далі для обраної нейромережі було проведено обчислювальні експерименти із застосуванням пакетної нормалізації. У дослідженні використовувалися підібрані на попередніх етапах багатомасштабні згортки в поєднанні із середнім пулінгом. Результати оцінки роботи нейромережі з пакетною нормалізацією після повнозв'язного шару наведено в таблиці 3.3.

Таблиця 3.3. Експерименти пов'язані з вибором пакетної організації

Номер експ.	Пакетна нормалізація	Точність клас.
1	+	98
2	-	97

### 3.3. Проведення перевірки моделі CNN SOP

Точність моделі оцінюється шляхом аналізу її здатності правильно класифікувати акустичні сигнали на основі числових метрик, що відображають різні аспекти її роботи. Метрики Precision і Recall відображають, наскільки точно модель ідентифікує позитивні класи та яку частку правильних прогнозів вона видає серед усіх передбачених результатів [101]. Для забезпечення балансу між цими показниками використовується F1-score, який є гармонійним середнім Precision і Recall. Оцінка точності здійснюється на тестовому наборі даних, що не використовувався під час навчання. Результати перевірки наведено в таблиці 3.4.

Таблиця 3.4. Перевірка метрик якості тренування нейромережі

	precision	recall	f1-score	Support
0	0.99	0.99	0.99	80
1	0.97	0.97	0.97	80
2	0.97	0.97	0.97	80
3	1.00	1.00	1.00	80
accuracy			0.98	
macro avg	0.98	0.98	0.98	320
weighted avg	0.98	0.98	0.98	320

Також була побудована матриця плутанини для аналізу результатів тренування класифікатора. Абсолютна більшість тестових семплів була правильно класифікована, що свідчить про високу ефективність моделі. Графічне представлення матриці плутанини наведено на рисунку 3.3.

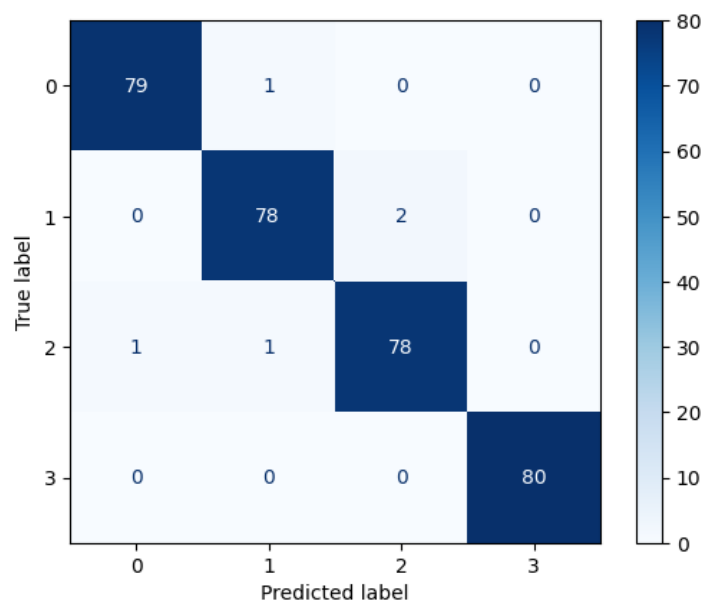


Рис. 3.3. Матриця плутанини, при перевірці на тестовій вибірці

Отримані результати демонструють високу якість класифікації на контрольній вибірці. Загальна точність становить 98%, що вказує на здатність моделі коректно класифікувати більшість зразків. Значення Precision, Recall та

F1-score для всіх класів знаходяться в діапазоні від 97% до 100%, що підтверджує збалансовану продуктивність нейромережі. Найкращий результат спостерігається для третього класу, де модель досягла ідеальної точності та повноти. Для першого та другого класів значення дещо нижчі за 97%, що може свідчити про мінімальні перехресні помилки. Показники *macro* та *weighted average* також підтверджують стабільну роботу класифікатора, оскільки всі середні значення знаходяться на рівні 98%. Оскільки підтримка кожного класу є рівномірною, модель не демонструє значного перекосу у бік будь-якої категорії. Це свідчить про її здатність узагальнювати знання на нових даних та забезпечувати високу якість класифікації.

Обчислювальна складність нейромережі є важливим параметром, що визначає ефективність її роботи в умовах обмежених ресурсів. Вона оцінюється за кількістю параметрів у моделі, що безпосередньо впливає на її розмір і швидкість обчислень. Час, необхідний для ідентифікації певної кількості семплів, дозволяє визначити реальну швидкодію моделі, особливо у випадку її використання в системах реального часу. Додатково враховуються обчислювальні витрати, зокрема кількість операцій із плаваючою комою (FLOPs) та використання пам'яті. Це є критично важливим фактором для ефективного розгортання моделі на апаратних пристроях.

Таблиця 3.5. Параметри нейромережі CNN SOP

Параметри	Значення
Загальна кількість параметрів	4,206,196
Треновані параметри	4,206,196
Нетреновані параметри	0
Загальна кількість множень-додавань (МБ)	98.54
Розмір вхідних даних (МБ)	0.16
Розмір проходу вперед/назад (МБ)	13.27
Розмір параметрів (МБ)	16.82
Оцінений загальний розмір (МБ)	30.26
Кількість параметрів	4,206,196
FLOPs	102,492,160

Модель містить 4,206,196 параметрів, що вказує на її відносно велику складність, особливо для нейромережі, яка працює з двовимірними акустичними ознаками.

Обчислювальна складність у термінах FLOPs становить 102.49 MFLOPs, що є відносно високим показником для моделі, яка може використовуватися у реальному часі. Основна частина обчислювальних витрат, ймовірно, припадає на SOP та Linear шари, оскільки SOP працює з великими тензорами, а Linear шари виконують матричні операції великого розміру.

Використання пам'яті складає 16MB, що є доволі оптимальним для сучасних апаратних платформ, таких як GPU або навіть високопродуктивні CPU. Проте, загальна оцінка пам'яті для проходу вперед-назад вказує на можливість оптимізації, наприклад, через використання менших повнозв'язних шарів або альтернативних підходів для SOP-перетворень.

Дивлячись на розмір проходу вперед/назад 13.27 MB, значна частина пам'яті витрачається на обробку проміжних тензорів під час навчання. Це може призводити до високого споживання VRAM під час тренування, що варто враховувати при розгортанні моделі на обмежених ресурсах. Модель демонструє достатньо високу обчислювальну складність через велику кількість параметрів, особливо у повнозв'язних шарах. Вона може бути ефективною у випадках, коли є достатні обчислювальні ресурси. Загалом, модель є потужною та здатною до ефективної класифікації.

Перевірка універсальності моделі дозволяє визначити, наскільки вона здатна працювати у різних умовах експлуатації. Для цього проводиться тестування на семплах, зашумлених білим шумом із різними рівнями відношення сигнал/шум. Оцінка SNR дозволяє визначити, наскільки модель чутлива до фонових перешкод і чи зберігає вона точність при зміні характеристик вхідного сигналу. Визначення узагальнюючої здатності моделі

допомагає оцінити її ефективність при зміні параметрів вхідних даних, що критично важливо для її застосування у різних реальних сценаріях.

Датасет було зашумлено білими шумами. Всього було проведено 3 експерименти. В результаті, були отримані семпли з рівнем шуму 0, -5 -10дБ. Кожен з семплів було подано в якості вхідних даних до запропонованого методу класифікації. Результати класифікації при різних значеннях SNR представлена у таблиці 3.6.

Таблиця. 3.6. Результати роботи нейромережі для класифікації з різним рівнем шуму

№	Співвідношення SNR	Точність класифікації
1	-10	74%
2	-5	81%
3	0	89%

Експерименти показали, що точність класифікації суттєво залежить від рівня шуму у вхідних даних. Чим вище SNR, тим кращі результати демонструє запропонована нейромережева модель. За SNR = 0 дБ модель досягає точності 89%, тоді як при SNR = -10 дБ точність знижується до 74%. Це свідчить про здатність нейромережі успішно розрізняти класи за високих значень SNR, проте за значного рівня шуму її класифікаційна ефективність суттєво знижується.

Результати експериментів підтверджують, що запропонована модель забезпечує високу точність класифікації для малошумних або помірно зашумлених сигналів. Водночас її стійкість до сильно зашумлених сигналів, де SNR сягає -10 дБ є обмеженою, що вимагає додаткових підходів до підвищення її надійності. Таким чином, модель демонструє високу ефективність за  $SNR \geq -5$  дБ, проте потребує оптимізації для класифікації сигналів із низьким співвідношенням сигнал/шум.

### 3.4. Проведення перевірки моделі UWAR-GAN

Оцінка нейронної мережі для придушення шуму демонструє значну здатність покращувати якість сигналу. PSNR, широко використовувана міра якості реконструкції сигналу, демонструє значне збільшення з -2,05 дБ до придушення до 7,73 дБ після придушення. Це свідчить про те, що мережа ефективно покращує чіткість сигналу за рахунок зменшення шумових перешкод. Аналогічно, SNR, яке відображає баланс між бажаним сигналом і фоновим шумом, наближається до нейтрального рівня, що підкреслює здатність моделі придушувати небажані шуми, зберігаючи цілісність сигналу. RMSE, яка кількісно вимірює відхилення від чистого сигналу, зменшується з 0,0301 до 0,0098, підтверджуючи, що нейронна мережа успішно мінімізує спотворення і підвищує загальну точність сигналу.

Окрім стандартних метрик оцінки сигналу, лог-спектральна відстань (LSD), яка вимірює спектральну схожість між чистим та обробленим сигналами, демонструє суттєве зменшення з 0,2211 до 0,0487. Це свідчить про те, що модель ефективно зберігає спектральні характеристики, зменшуючи при цьому шумові компоненти. Середня залишкова потужність шуму значно зменшується, що свідчить про те, що мережа ефективно зменшує фонові завади, зберігаючи при цьому природне представлення сигналу. Чутливість, параметр, що відображає реакцію мережі на варіації вхідних сигналів, залишається стабільною, забезпечуючи стабільність придушення шуму в різних умовах.

Аналіз вагових розподілів показує, що мережа підтримує добре збалансований розподіл параметрів протягом навчання. Середні значення, стандартне відхилення та діапазон значень ваг залишаються стабільними, що свідчить про добре навчену модель, яка не демонструє нестабільності або екстремальних відхилень ваг. Така узгодженість у розподілі ваг сприяє стійкості та надійності моделі в обробці сигналів за різних шумових умов.

Час виведення дещо збільшується з 0,0184 секунди до 0,0242 секунди, що свідчить про те, що модель виконує додаткову обробку для придушення шуму, залишаючись при цьому обчислювально ефективною. Використання пам'яті залишається постійним на рівні приблизно 953,74 МБ, що свідчить про те, що модель підтримує оптимальний баланс між складністю та ефективністю, не накладаючи додаткових обчислювальних вимог.

Загалом, нейронна мережа ефективно покращує чіткість сигналу, зменшує шум і зберігає ключові спектральні компоненти. Покращення PSNR, зменшення RMSE та значне зменшення лог-спектральної відстані демонструють здатність моделі генерувати чистіші та розбірливіші сигнали. Стабільність вагового розподілу та стабільна ефективність пам'яті підтверджують адекватність процесу навчання. Результати показують, що модель успішно пригнічує шум, зберігаючи деталі сигналу, що робить її добре придатною для практичних застосувань, які потребують якісної обробки акустичних сигналів.

Таблиця 3.7. Параметри нейромережі U-net

Параметри	Значення
Total Parameters	16658562
Trainable Parameters	16658562
Non-Trainable Parameters	0
Multiply-Accumulate Operations	4.04 GMac
FLOPs (2 * MACs)	8.08 FLOPs
Input Data Size (MB)	0.12
Forward/Backward Pass Size (MB)	0.25
Parameters Size (MB)	63.55
Inference Time	0.010275 seconds
GPU Memory Allocated	331.92 MB
GPU Memory Cached	1084.00 MB
Estimated Total Size (MB)	63.92 MB
FLOPs (Floating Point Operations)	4035903488 MACs

Навчена нейронна мережа демонструє добре структурований баланс між обчислювальною ефективністю та складністю моделі. З загальною кількістю



параметрів 16 658 562, всі з яких піддаються навчанню, модель має високу здатність до вивчення складних репрезентацій, забезпечуючи адаптивність до завдань придушення шуму. Відсутність параметрів, що не піддаються навчанню, вказує на те, що вся мережа робить свій внесок у навчання, максимізуючи оптимізаційний потенціал.

Модель виконує приблизно 4,04 мільярда операцій множення-акумуляції (MAC), що відповідає 8,08 мільярдів операцій з плаваючою комою (FLOP), що відображає обчислювальні потреби для обробки кожної вхідної вибірки. Незважаючи на це, час виведення залишається ефективним на рівні 0,0103 секунди, демонструючи, що модель може забезпечити швидку продуктивність у реальному часі без надмірних обчислювальних витрат.

Розподіл пам'яті добре оптимізовано, модель займає приблизно 63,92 МБ, включаючи вхідні дані, зберігання параметрів та проміжні активації. Ефективно використовується пам'ять графічного процесора: 331,92 МБ виділяється під час виведення, а загальний обсяг кешу становить 1,08 ГБ, що свідчить про те, що модель працює в межах розумних апаратних обмежень. Розмір прямого і зворотного проходу 0,25 МБ свідчить про ефективний процес зворотного поширення, що забезпечує плавне оновлення градієнта під час навчання.

Загалом, нейронна мережа демонструє адекватність як з точки зору пропускної здатності, так і з точки зору обчислювальної ефективності. Велика кількість параметрів, що навчаються, гарантує, що модель може вивчати складні патерни придушення шуму, зберігаючи при цьому керований обсяг пам'яті та час виведення. Ефективне використання ресурсів графічного процесора в поєднанні з високою швидкістю обробки підтверджує, що навчена мережа добре оптимізована для практичного розгортання в додатках реального або близького до реального часу.

### 3.5. Результати придушення шумів та класифікації

#### 3.5.1. Результати придушення шумів акустичних сигналів водного середовища нейромережею UWAR GAN

Дані різних класів датасету були подані на вхід до запропонованої нейромережі для придушення шумів. Результати були оцінені по наступним критеріям: співвідношення сигнал-шум, середньоквадратична похибка, пікове співвідношення сигнал-шум, стандартне відхилення. На вхід нейромережі подавались семпли з низьким співвідношенням SNR, що сягало -10 дБ. Результати придушення шумів для різних класів датасету Storm представлені в таблиці 3.8.

Таблиця 3.8. Результати придушення фонових шумів в семплах датасету Storm

	До придушення шумів	Після придушення
SNR (дБ)	-10.79	-0.58
RMSE	2.7	0.8
Пікове співвідношення	0.81	11.02
Стандартне відхилення	0.0285	0.0039

Після придушення фонових шумів якість акустичних сигналів значно покращилася. Підвищення значення SNR свідчить про покращення якості сигналу, оскільки потужність корисного сигналу починає переважати над рівнем шуму. Це означає, що сигнал стає чіткішим і менш спотвореним, що сприяє його кращому сприйняттю та обробці. Загалом, після обробки якість сигналу значно зростає, що підтверджується зменшенням RMSE та стандартного відхилення. Візуальне представлення оброблених семплів наведено на рисунку 3.4.

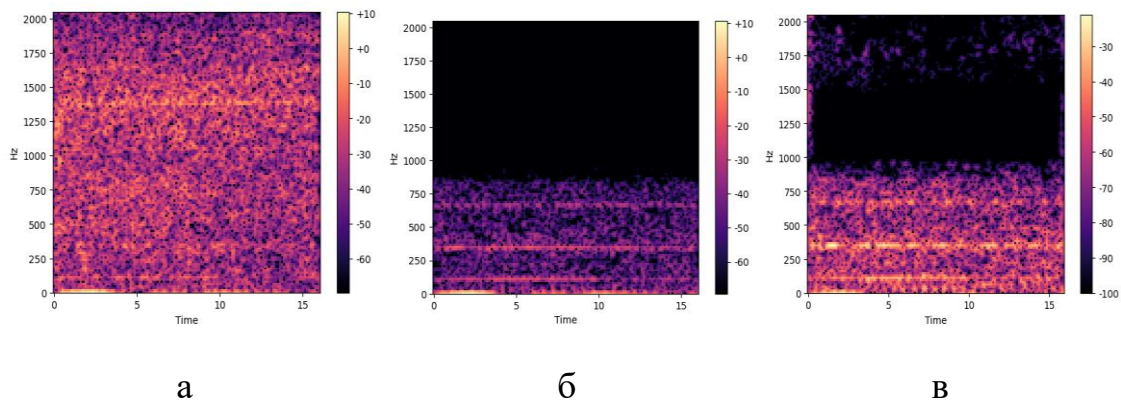


Рис 3.4. Відображення семплів датасету Storm: до придушення шумів (а), після придушення шумів (б), до зашумлення фоновими шумами (в)

Додатково було здійснено перевірку результатів придушення шумів на семплах датасету Shipsear, що мали значення SNR -10 дБ. Результати обробки сигналів цього датасету наведено в таблиці 3.9.

Таблиця 3.9. Результати придушення фонових шумів в семплах датасету Shipsear

	До придушення шумів	Після придушення шумів
Співвідношення сигнал-шум (дБ)	-10.67	-0.81
Середньоквадратична похибка	2.4	0.8
Пікове співвідношення сигнал-шум (дБ)	1.9	11.1
Стандартне відхилення	0.025	0.003

Після придушення фонових шумів якість акустичних сигналів у датасеті Shipsear значно покращилася. Значення SNR підвищилося з -10,67 дБ до -0,81 дБ, а PSNR зріс із 1,9 дБ до 11,1 дБ. Середньоквадратична похибка зменшилася з 2,4 до 0,8, а стандартне відхилення – з 0,025 до 0,003. Візуалізація семплу до та після обробки наведена на рисунку 3.6/

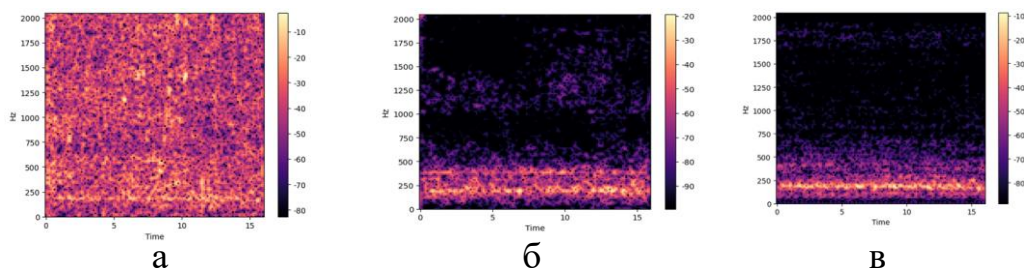


Рис 3.6. Відображення семплів датасету Shipteam: до придушення шумів (а), після їх придушення (б), до зашумлення фоновими шумами (в)

Також, була здійснена перевірка результатів на штучно згенерованому датасеті. Результати придушення шумів на штучному датасеті представлені у таблиці 3.10.

Таблиця 3.10. Результати придушення шумів на штучному датасеті представлені

	До придушення шумів	Після придушення шумів
Співвідношення сигнал-шум	-10.349	-0.241
Середньоквадратична похибка	1.05	0.33
Пікове співвідношення сигнал-шум	3.183	9.623
Стандартне відхилення	1.09	0.1

Також було виконане придушення штучних шумів.

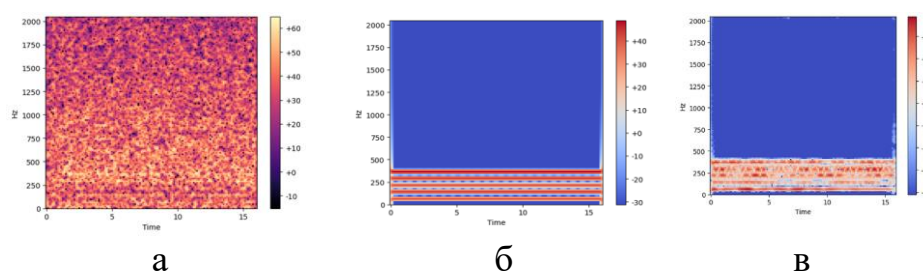


Рис 3.7. Відображення семплів датасету Shipteam: до придушення шумів (а), після їх придушення (б), до зашумлення фоновими шумами (в)

Після придушення фонових шумів у штучно згенерованих сигналах спостерігається покращення якості, аналогічне до попередніх експериментів. SNR зросло з -10,349 до -0,241, що свідчить про суттєве зменшення рівня шуму відносно корисного сигналу. Пікове значення SNR підвищилося з 3,183 до 9,623, вказуючи на покращення максимальної амплітуди сигналу відносно шуму. RMSE зменшилася з 1,05 до 0,33, що означає зниження різниці між очищеним та оригінальним сигналами. Стандартне відхилення знизилося з 1,09 до 0,1, що вказує на стабільність та однорідність очищеного сигналу. Показники підтверджують ефективність процедури придушення шумів та покращення якості обробленого сигналу.

### 3.5.2. Оцінка результатів класифікації

Отримані результати підтверджують ефективність процедури придушення шумів та покращення якості обробленого сигналу. Оригінальний і покращений методи були протестовані на трьох основних датасетах та їхніх модифікованих версіях із доданими шумами. Візуалізацію результатів наведено на рисунку 3.8.

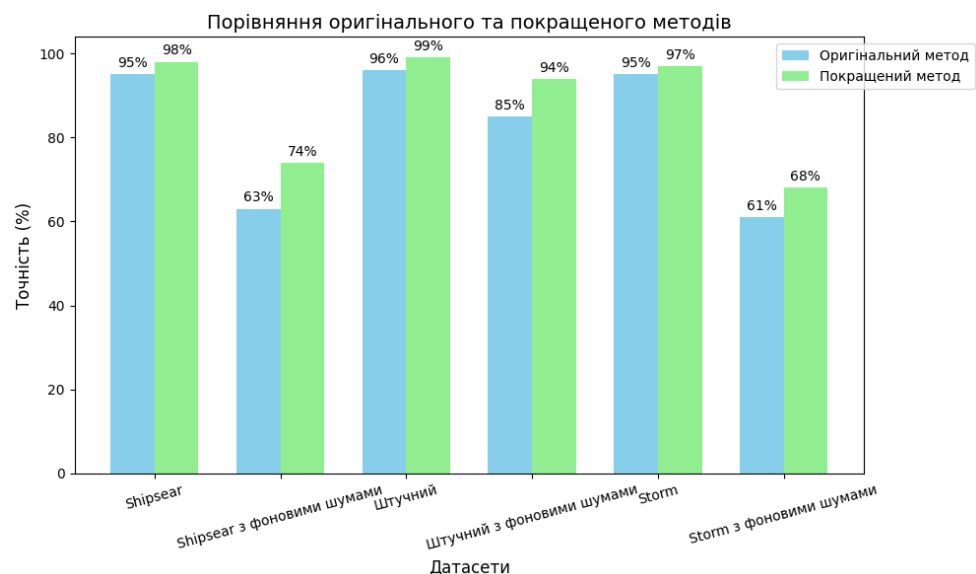


Рис 3.8. Порівняння точності роботи покращеного методу SOP в порівнянні з оригінальним методом

### Висновки до розділу 3

Було обґрунтовано складність задачі придушення шумів водного середовища та необхідність наявності достатньої кількості даних для роботи з ними. Також обґрунтовано актуальність формування датасетів для тренування нейромережі з придушення шумів. Було виконано огляд літератури, де розглянуто різні датасети водного середовища. Обрано та описано датасети Shipsear та Storm. Також описано процес відбору чистих семплів водного середовища, процес комбінації семплів, вказано отримані семпли, критерії оцінювання та результати застосування отриманого датасету для тренування нейромережі з придушення шумів[3].

В результаті проведеного дослідження удосконалено метод класифікації акустичних сигналів водного середовища та отримано вищі результати класифікації у порівнянні з оригінальним методом. Запропонований метод ефективніше вилучає потрібні характеристики у вхідних даних, застосовуючи багатомасштабну згортку та вилучаючи характеристики різних масштабів. Застосування середнього пулінгу забезпечило зменшення розмірності вхідних даних. Додатково доданий шар з функцією активації ReLU та пакетною нормалізацією забезпечив поліпшення здатності до узагальнення нейромережі. Удосконалена нейромережа показала вищі результати у порівнянні з оригінальним методом, зокрема — підвищення точності класифікації[4].

## **РОЗДІЛ 4. ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ АНАЛІЗУ АКУСТИЧНИХ СИГНАЛІВ ВОДНОГО СЕРЕДОВИЩА**

### **4.1. Розроблене програмне забезпечення**

Для можливості користувача зручно користуватись розробленими методами, було розроблене програмне забезпечення. Програмне забезпечення для аналізу, придушення шумів і класифікації акустичних сигналів водного середовища структуроване у вигляді окремих модулів, кожен із яких виконує певну функцію в обробці аудіоданих.

Модуль `prediction_module` забезпечує нейромережеве розпізнавання класів акустичних сигналів. Він обробляє запити користувача, готує дані та виконує прогнозування, використовуючи попередньо навчені моделі. Система виконує перетворення вхідного сигналу у відповідний формат, проводить його попередню обробку та класифікує, використовуючи глибоку нейронну мережу, яка враховує спектральні особливості звукових хвиль.

Модуль `denoising_module` відповідає за очищення акустичних сигналів від фонового шуму, покращуючи їхню якість для подальшого аналізу. Він отримує запити на фільтрацію сигналів, виконує їхню попередню обробку та передає у нейромережу для видалення шумів. Система використовує генеративно-змагальну нейронну мережу, яка навчається реконструювати чистий сигнал, відділяючи його від завад. Після обробки забезпечується зворотне перетворення сигналу до вихідного формату.

Модуль `denoise_training_module` дозволяє навчати модель для придушення шумів на основі підготовлених акустичних даних. Він формує навчальні вибірки, проводить їхнє попереднє опрацювання та запускає процес тренування генеративно-змагальної мережі. Під час навчання система адаптивно налаштовує параметри моделей і використовує алгоритми оцінки якості очищених сигналів. Архітектура програмного забезпечення аналізу акустичних сигналів водного середовища представлена на рисунку 4.1.

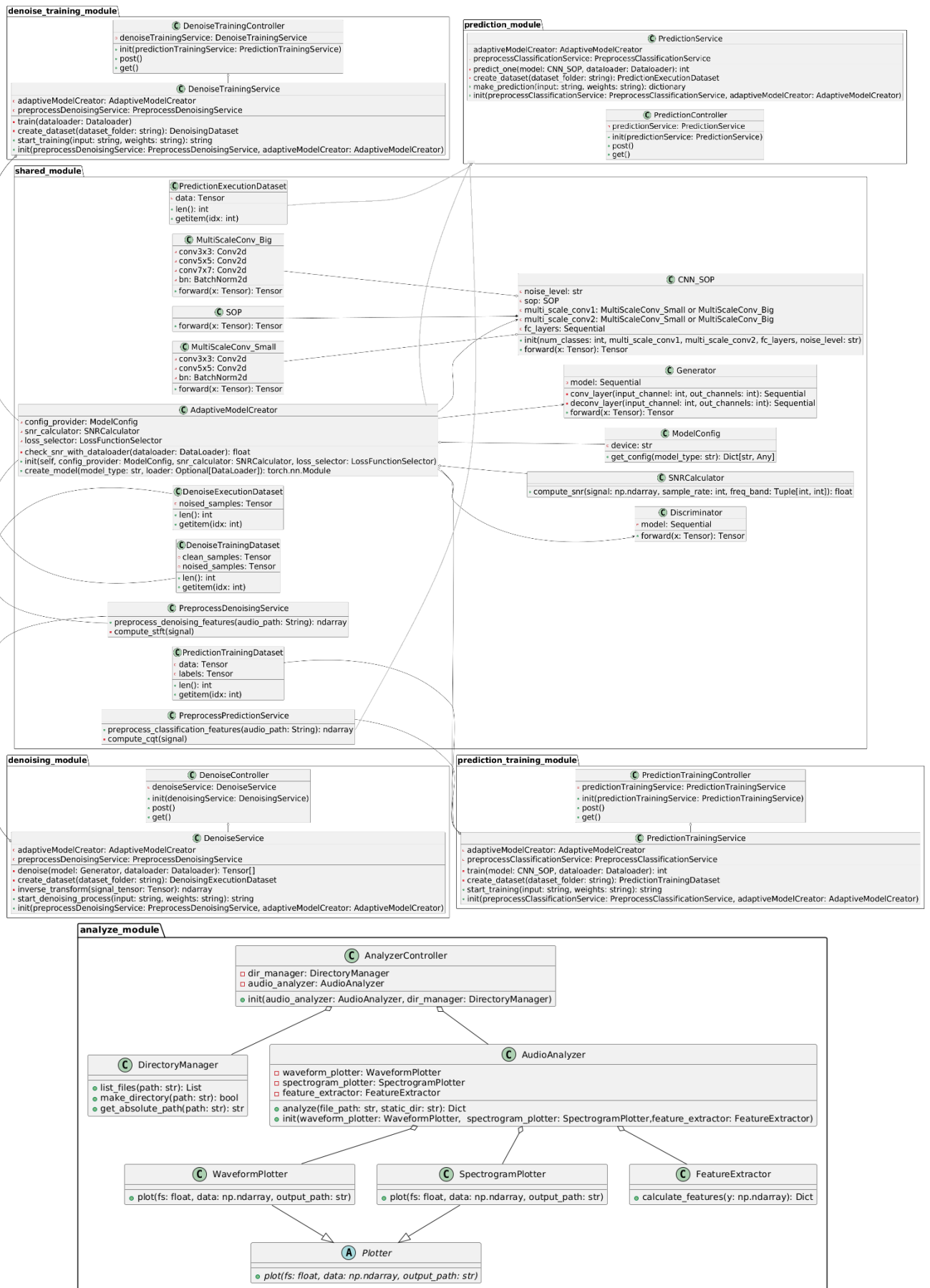


Рис. 4.1. – Архітектура програмного забезпечення аналізу акустичних сигналів водного середовища



Модуль `prediction_training_module` забезпечує процес навчання нейронної мережі для класифікації акустичних сигналів. Він створює набір даних для тренування, виконує підготовку характеристик і запускає процес навчання. Система адаптує конфігурацію моделі відповідно до характеристик акустичних середовищ, дозволяючи підвищити точність розпізнавання класів.

Класи модуля `shared_module` включають допоміжні сервіси, які використовуються всіма модулями. Вони забезпечують попередню обробку сигналів, реалізують адаптивне створення моделей та виконують спектральні перетворення, такі як STFT і CQT. Система оцінює рівень шуму в сигналах, адаптивно підбираючи архітектуру мережі залежно від середовища запису. Мережі включають багатомасштабні згортки, глибокі згорткові структури та генеративні моделі для придушення шумів.

Модуль `analyze_module` дозволяє проводити комплексний аналіз аудіофайлів. Він здійснює візуалізацію звукових хвиль у вигляді графіків та спектрограм, а також відображає характеристики сигналу. Аналізатор керує файлами, виконує необхідні перетворення та надає користувачу детальну інформацію про структуру акустичних даних.

Загалом, така архітектура забезпечує ефективну взаємодію між придушенням шумів, класифікацією та аналізом акустичних сигналів. Вона дозволяє працювати з великими наборами даних, адаптувати моделі до різних умов та отримувати точні результати як у розпізнаванні класів, так і у відновленні чистих сигналів.

#### **4.1.2. Модуль тренування нейромережі для придушення шумів**

Модуль `denoise_training_module` відповідає за навчання нейромережі для шумозаглушення акустичних сигналів. Він керує створенням датасету, вибором моделі, підготовкою даних та безпосереднім процесом тренування.

Клас DenoiseTrainingController приймає зовнішні HTTP-запити, передаючи їх у DenoiseTrainingService для виконання навчання. Структура модуля тренування нейромережі для придушення шумів представлена на рисунку 4.2.

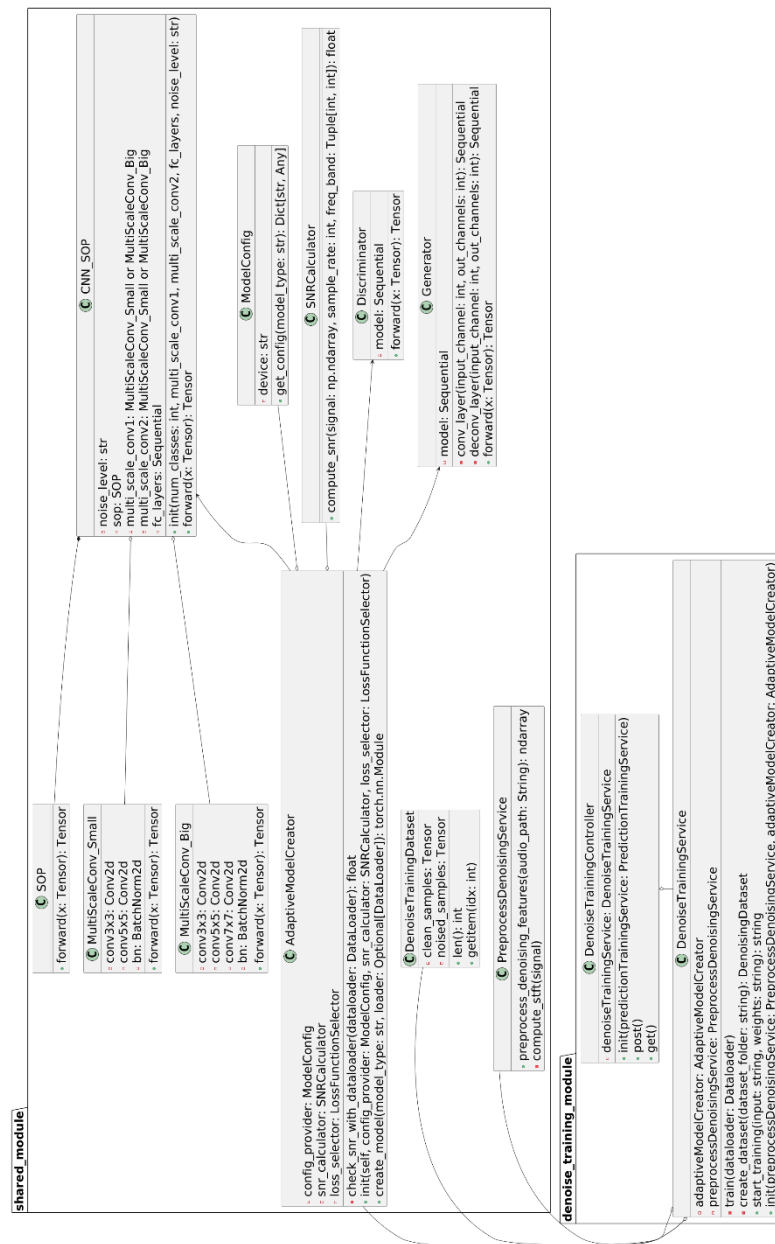


Рис. 4.2. Структура модуля тренування нейромережі для придушення шумів та зв'язок цього модуля з класами інших модулів.

Метод `post()` ініціює тренувальний процес, тоді як `get()` може використовуватися для перевірки його статусу.

Клас `DenoiseTrainingService` реалізує основну логіку навчання. Він використовує `PreprocessDenoisingService` для обробки аудіофайлів перед подачею в нейромережу та `AdaptiveModelCreator` для вибору відповідної архітектури. Метод `start_training()` приймає вхідні дані у вигляді файлу з вагами, ініціює навчання та повертає статус виконання. Метод `train()` здійснює навчання генератора на основі даних у `Dataloader`, навчаючи модель шумозаглушення.

Модуль `shared_module` надає важливі компоненти для реалізації тренувального процесу.

Клас `PreprocessDenoisingService` виконує передобробку аудіосигналів, генеруючи STFT-представлення, яке використовується як вхід для нейромережі. Це дозволяє мережі працювати безпосередньо із частотними компонентами сигналу.

Клас `AdaptiveModelCreator` забезпечує вибір та ініціалізацію моделі шумозаглушення. Він використовує `SNRCalculator` для оцінки рівня шуму у вхідних даних та приймає рішення щодо конфігурації моделі. Якщо рівень шуму високий, обирається складніший генератор, тоді як для менш зашумлених сигналів модель може бути спрощеною.

Клас `Generator` реалізує архітектуру генеративної нейромережі для шумозаглушення. Він містить згорткові та деконволюційні шари для обробки спектрального представлення та реконструкції очищеного сигналу.

Клас `Discriminator` використовується в разі застосування генеративно-змагального навчання. Він допомагає покращити якість шумозаглушення, розрізняючи очищені та реальні чисті сигнали, змушуючи генератор створювати більш природні результати.

Клас `DenoiseTrainingDataset` представляє навчальний набір, що містить як зашумлені, так і чисті зразки, необхідні для порівняння та навчання нейромережі.

### 4.1.3. Модуль виконання придушення шумів нейромережею

Модуль `denoise_module` реалізує процес шумозаглушення акустичних сигналів, забезпечуючи управління запитами, підготовку даних, вибір моделі та виконання шумозаглушення

Клас `DenoiseController` є точкою входу для зовнішніх запитів, приймаючи HTTP-запити та передаючи їх для обробки в `DenoiseService`. Метод `post()` запускає процес шумозаглушення, тоді як `get()` може використовуватися для отримання статусу обробкою GET запитів.

Клас `DenoiseService` відповідає за обробку вхідних даних, створення нейромережі, виконання придушення шумів та реконструкцію очищеного сигналу. Він використовує `PreprocessDenoisingService` для перетворення аудіофайлів у необхідне представлення та `AdaptiveModelCreator` для вибору відповідної генеративної моделі. Метод `start_denoising_process()` приймає шлях до вхідних даних і ваги моделі, виконує шумозаглушення та повертає відфільтрований сигнал. Метод `denoise()` застосовує модель `Generator` до кожного зразка в `Dataloader`, а `inverse_transform()` перетворює очищений тензор у вихідний аудіосигнал. Модуль `shared_module` забезпечує основні компоненти, необхідні для процесу шумозаглушення. Клас `PreprocessDenoisingService` виконує підготовку вхідного сигналу, обчислюючи STFT, що є вхідним представленням для нейромережі. Це дозволяє моделі працювати з часово-частотними характеристиками сигналів. Клас `AdaptiveModelCreator` вибирає та ініціалізує відповідну нейромережу для шумозаглушення. В рамках створення нейромережі для придушення шумів `Generator`, адаптивне створення нейромережі не застосовується. Структура модуля придушення шумів та його зв'язки з класами інших модулів представлена на рисунку 4.3.

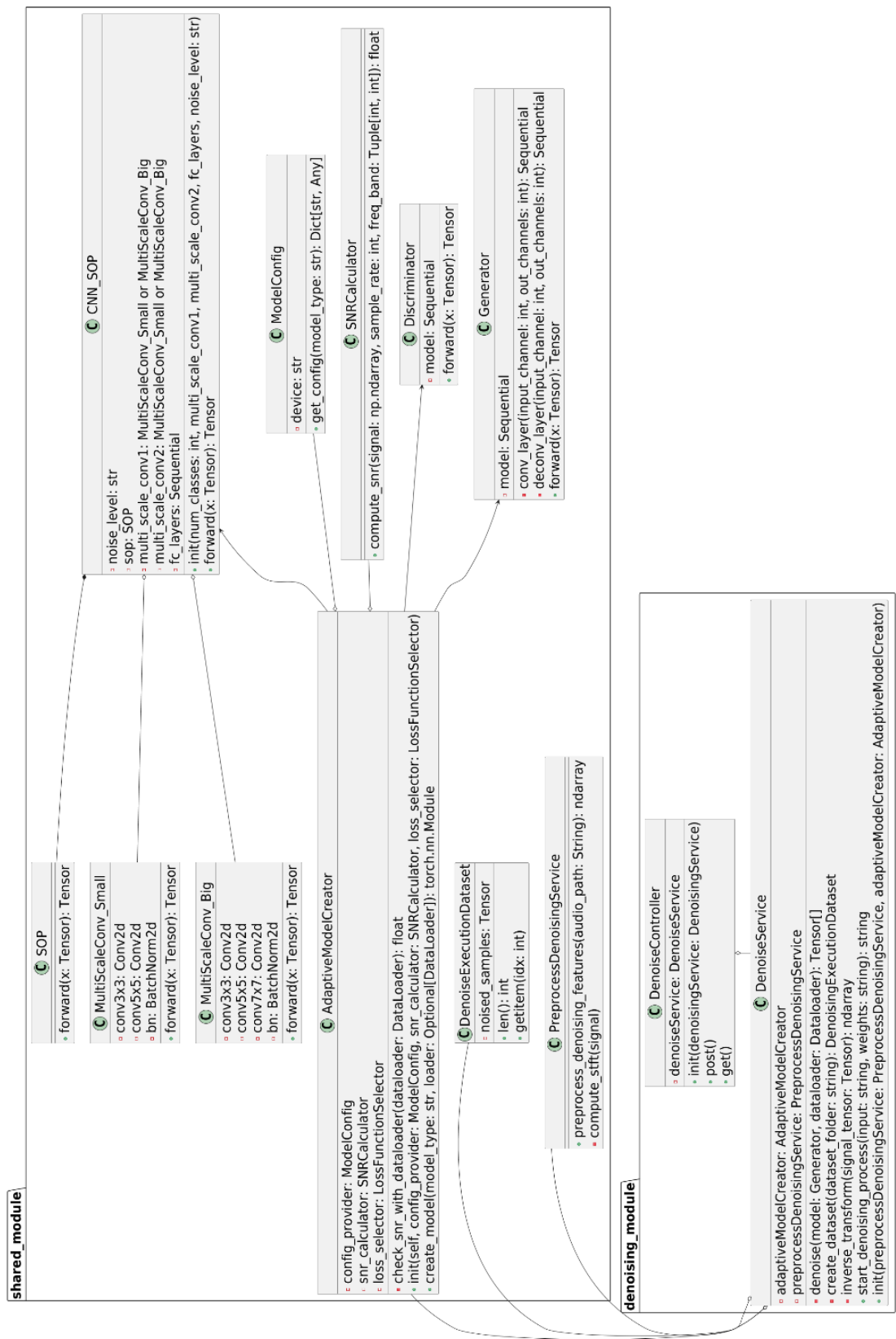


Рис. 4.3. Структура модуля придушення шумів та його зв'язки з класами інших модулів

Клас `Generator` реалізує архітектуру нейромережі для шумозаглушення, використовуючи згорткові та деконволюційні шари для обробки спектрального представлення та відновлення аудіосигналу. Його вихід використовується для формування очищеного спектру, який потім реконструюється у тимчасову область.

Клас `Discriminator` застосовується в разі використання генеративно-змагального навчання, допомагаючи покращити якість очищених сигналів шляхом оцінки їх відповідності до чистих зразків.

Клас `DenoiseExecutionDataset` представляє набір даних, що містить зашумлені зразки для подачі в модель, забезпечуючи доступ до кожного з них у відповідному форматі.

Таким чином, `denoise_module` має виконує обробку вхідних запитів та тренування, використовуючи сервіси `shared_module` для обробки даних, створення моделі.

#### **4.1.4 Модуль тренування нейромережі для класифікації акустичних сигналів**

Модуль `prediction_training_module` призначений для навчання моделей класифікації акустичних сигналів. Він керує створенням навчального датасету, вибором моделі залежно від рівня шуму, навчальним процесом та управлінням гіперпараметрами. Клас `PredictionTrainingController` є інтерфейсом для отримання запитів на навчання. Він взаємодіє з `PredictionTrainingService`, передаючи запити на старт тренування. Метод `post()` ініціює навчальний процес, тоді як `get()` може використовуватися для отримання статусу тренування. Структура модуля тренування класифікатора та його зв'язки з класами інших модулів представлена на рисунку 4.4.

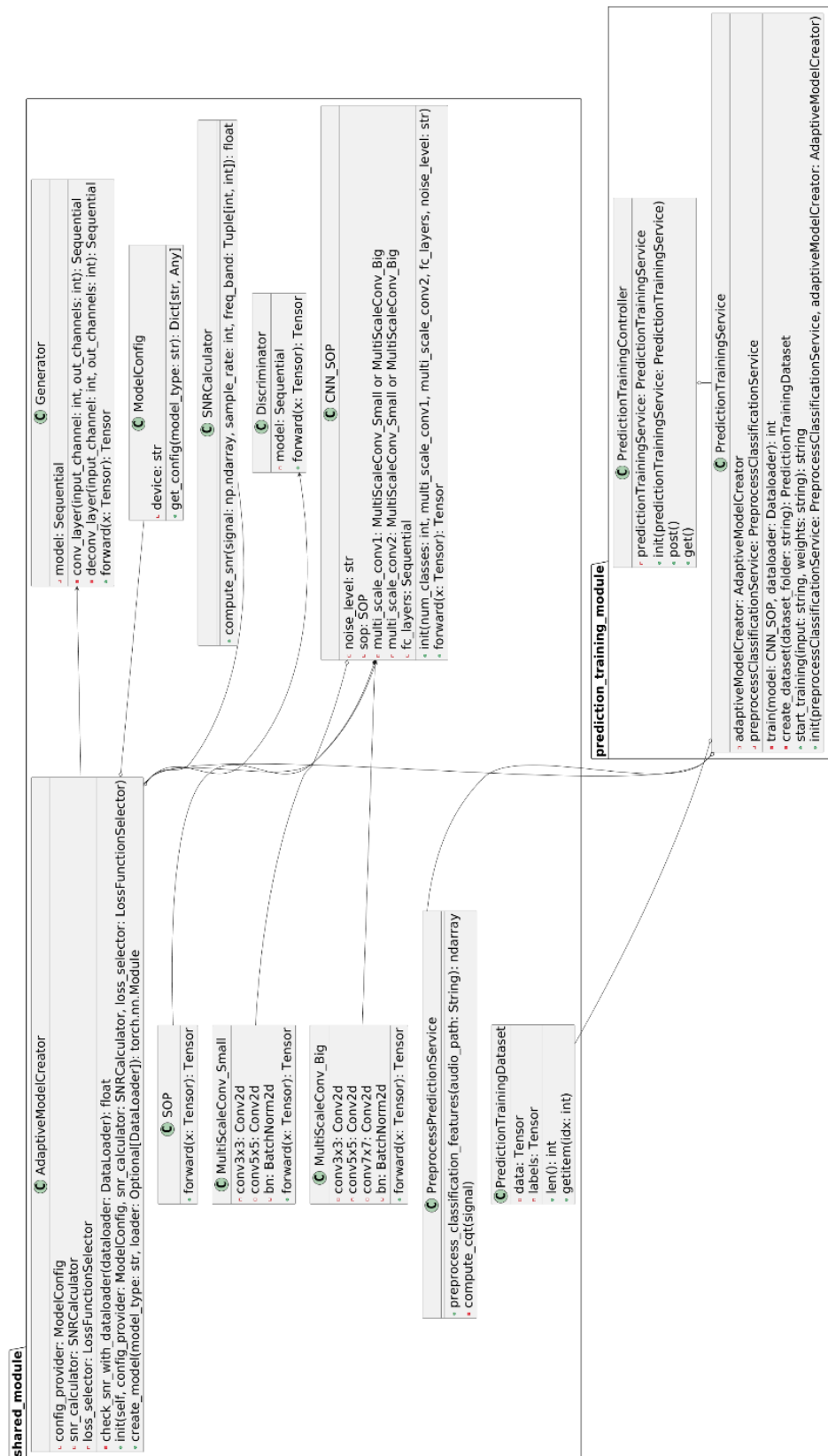


Рис. 4.4. Структура модуля тренування класифікатора та його зв'язки з класами інших модулів

Клас `PredictionTrainingService` виконує основну логіку навчання моделі. Він використовує `PreprocessClassificationService` для обробки аудіосигналів перед подачею в мережу та `AdaptiveModelCreator` для створення відповідної моделі, залежно від рівня шуму. Метод `start_training()` приймає шлях до вхідних даних та ваги моделі, запускає навчання та повертає результат. Метод `train()` здійснює навчання моделі `CNN_SOP`, використовуючи підготовлений датасет та задані параметри.

Модуль `shared_module` забезпечує ключові компоненти для роботи модуля навчання. Клас `PreprocessPredictionService` виконує підготовку аудіофайлів для навчання, обчислюючи CQT, що є необхідним вхідним представленням для класифікаційної мережі.

Клас `AdaptiveModelCreator` відповідає за створення відповідної моделі, аналізуючи рівень шуму у вхідних даних. Він використовує `SNRCalculator` для оцінки відношення сигнал/шум і на основі цього вибирає параметри архітектури. Якщо рівень шуму високий, модель отримує більш складні багатомасштабні згортки та ще один збільшений розширений повнозв'язний шар. В іншому випадку застосовується спрощена версія архітектури. Класи `Generator` та `Discriminator` не створюються в рамках тренування класифікатора.

Клас `ClassificationDataset` представляє набір даних, що використовується під час навчання моделі. Він містить тензори `data` для збереження вхідних даних та `labels` для відповідних міток класів. Метод `len()` повертає кількість зразків у наборі даних, а метод `getitem(idx)` дозволяє отримати окремий елемент за індексом, що використовується під час завантаження батчів у `Dataloader`.

`PredictionTrainingComponent` напряму керує `PredictionTrainingService`, який відповідає за підготовку даних, викликає `PreprocessPredictionService` для обчислення ознак та використовує `CNN_SOP` для виконання класифікації. Модель `CNN_SOP` містить рівні `SOP` та `MultiScaleConv`, що дозволяють



їїобробляти складні акустичні сигнали. PredictionTrainingService також працює з PredictionTrainingDataset, який містить вхідні дані для навчання.

#### **4.1.5. Модуль виконання класифікації акустичних сигналів**

Модуль prediction\_module реалізує процес прогнозування для акустичних сигналів, забезпечуючи управління запитами, обробку вхідних даних, створення моделей і виконання передбачення. Він містить два основні класи: PredictionService та PredictionComponent, які тісно взаємодіють між собою та з іншими модулями.

Клас PredictionController є основною точкою входу для зовнішніх запитів. Він отримує HTTP-запити та передає їх для обробки в PredictionService. Метод post() приймає запити на виконання передбачення, тоді як get() використовується для обробки GET запитів. Структура модуля

Клас PredictionService є основною бізнес-логікою модуля, керуючи створенням датасету, виконанням передбачення та обробкою аудіосигналів. Він використовує PreprocessPredictionService для вилучення значень CQT, а також AdaptiveModelCreator для створення відповідної нейромережі. Метод make\_prediction() приймає аудіофайл і ваги моделі, виконує передобробку та передбачення, повертаючи результати у вигляді словника. Метод predict\_one() застосовує модель CNN\_SOP до окремого зразка даних, отриманого з DataLoader.

Модуль shared\_module забезпечує базову функціональність, необхідну для передбачення. Клас PreprocessPredictionService виконує передобробку аудіосигналу, зокрема вилучає тензори CQT з часового представлення вхідних даних моделі CNN\_SOP. Клас AdaptiveModelCreator створює відповідну нейромережу, ґрунтуючись на конфігурації пристрою та назві вхідних ваг. Він використовує ModelConfig для отримання параметрів моделі та SNRCalculator для оцінки рівня шуму у вхідних даних. В рамках проведення класифікації вхідних даних, модель створюється без урахуванням результатів SNRCalculator. Клас CNN\_SOP реалізує архітектуру згорткової нейромережі,

для якої здійснюється підбір кількості повнозв'язних шарів та однієї з багатомасштабних згорток - MultiScaleConv\_Small або MultiScaleConv\_Big, що дозволяє створювати модель залежно від типу обраних ваг. Моделі Generator і Discriminator в даному випадку не були створені. Структура придушення шумів та його зв'язки з класами інших модулів представлена на рисунку 4.5.

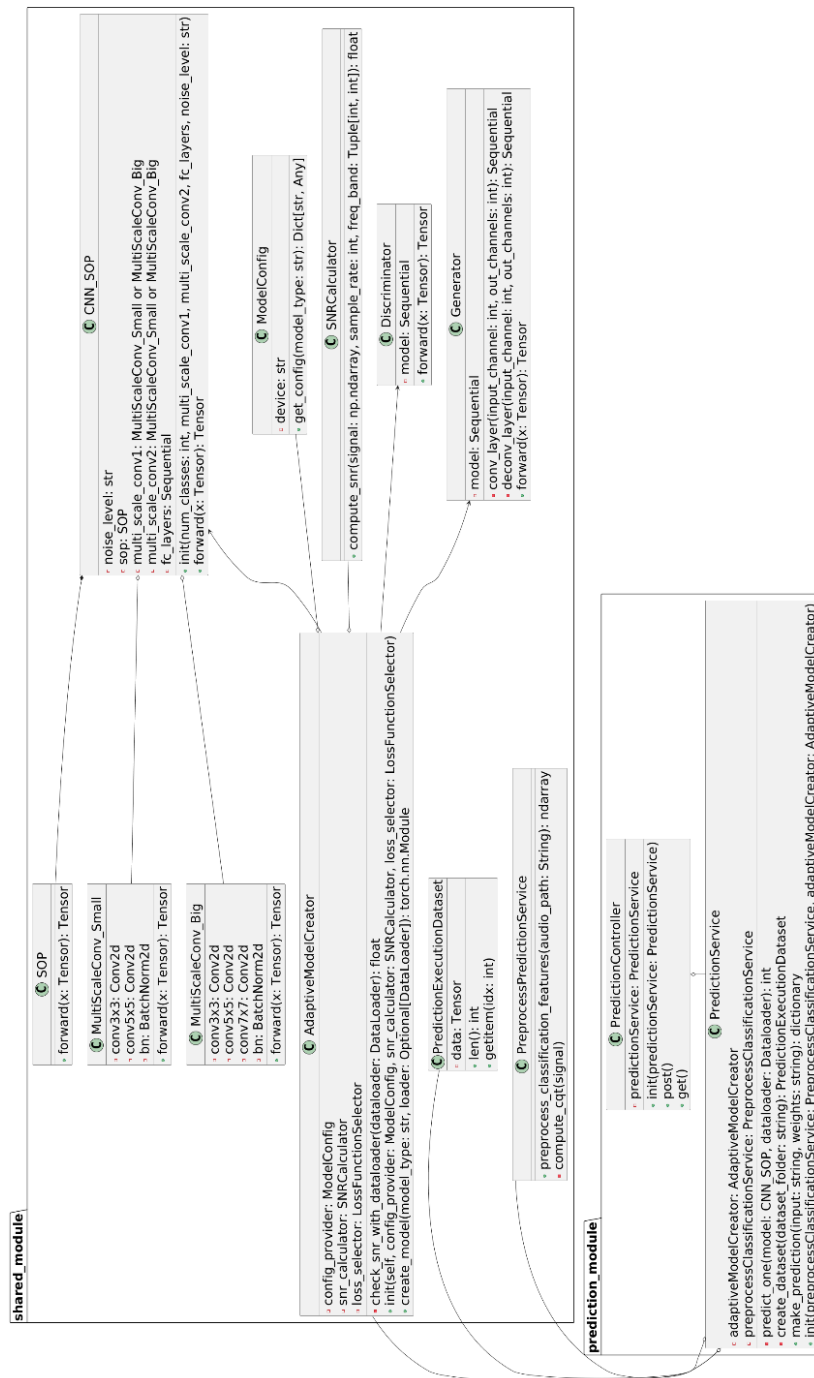


Рис. 4.5. Структура модуля класифікації та його зв'язки з класами інших модулів

Клас `PredictionExecutionDataset` представляє набір даних, який використовується для передбачення, забезпечуючи доступ до індивідуальних зразків.

`PredictionComponent` безпосередньо залежить від `PredictionService` для обробки запитів. У свою чергу, `PredictionService` покладається на `PreprocessPredictionService` для підготовки функцій і `PredictionExecutionDataset` для зберігання структурованих даних. На схемі також показано, як `PredictionService` взаємодіє з `AdaptiveModelCreator`, що підкреслює його роль у адаптивному створенні нейронних мереж CNN-SOP.

#### **4.1.6. Модуль аналізу акустичних сигналів**

Модуль `analyze_module` відповідає за аналіз аудіофайлів, їхню візуалізацію та обчислення характеристик. Основним керуючим класом є `AnalyzerController`, який координує роботу підлеглих компонентів. Він взаємодіє з `DirectoryManager`, що займається управлінням файлами та директоріями, а також з `AudioAnalyzer`, який проводить безпосередній аналіз аудіосигналів. `DirectoryManager` надає можливість отримувати список файлів у вказаній директорії, створювати нові папки та отримувати абсолютний шлях до файлів. `AudioAnalyzer` відповідає за обробку аудіосигналів та використовує три підлеглі компоненти: `WaveformPlotter`, `SpectrogramPlotter` і `FeatureExtractor`. `WaveformPlotter` створює графік форми сигналу, а `SpectrogramPlotter` будує спектрограму, що дозволяє візуалізувати частотний розподіл сигналу. Обидва класи успадковують абстрактний клас `Plotter`, який визначає єдиний інтерфейс для візуалізації аудіоданих.

`FeatureExtractor` виконує обчислення різних характеристик сигналу, таких як статистичні параметри або інші показники, що можуть використовуватися для подальшого аналізу або класифікації. Основний метод

analyze у AudioAnalyzer обробляє аудіофайл, керує створенням відповідних візуалізацій та витягує характеристики, повертаючи результати у вигляді словника analyze у AudioAnalyzer обробляє аудіофайл, керує створенням відповідних візуалізацій та витягує характеристики, повертаючи результати у вигляді словника, що включає різні характеристики сигналу. Структура класів модуля аналізу представлена на рисунку 4.6.

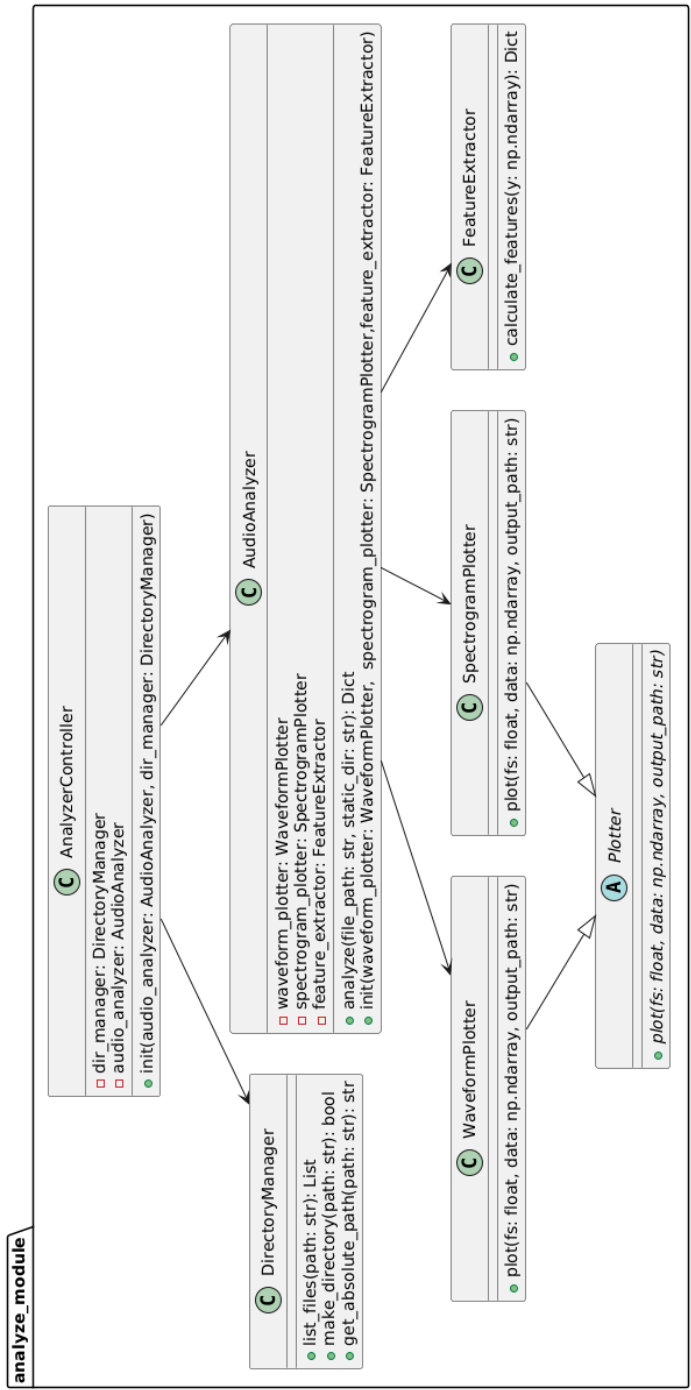


Рис. 4.6. - Структура модуля аналізу акустичних сигналів

## 4.2. Адаптивний пайплайн для створення нейромережових моделей

Підводні акустичні сигнали характеризуються значною мінливістю якості через коливання рівнів навколишнього шуму. Для вирішення цієї проблеми описано адаптивну архітектуру нейронної мережі, яка динамічно змінює свою структуру залежно від SNR вхідного сигналу. Система організована в три модульні компоненти – SNRCalculator, ModelConfig та AdaptiveModelCreation, – які разом дозволяють вибирати оптимальну складність моделі для надійної та ефективної класифікації акустичних сигналів. Нижче детально описано кожен компонент та обґрунтовано переваги SNR-орієнтованої адаптації архітектури з погляду стійкості, узагальнення й обчислювальної ефективності в умовах підводної акустики.

Надійне оцінювання SNR є першим кроком у роботі адаптивної архітектури. SNRCalculator обчислює значення SNR для вхідного сигналу за допомогою методу Велча оцінювання спектральної густини потужності. Метод Велча сегментує сигнал, застосовує STFT і усереднює отримані спектральні оцінки, знижуючи варіації в оцінках спектру. Конкретно, SNRCalculator інтегрує PSD у визначеній частотній смузі сигналу, наприклад, 100–300 Гц і порівнює отриману енергію з загальною спектральною енергією сигналу, формуючи таким чином смуговий показник SNR. Це дозволяє надійно оцінити рівень шуму навіть у динамічних умовах.

ModelConfig – це абстрактний структурний рівень, який визначає і зберігає конфігурації різних нейромережових моделей, що використовуються в системі. Він забезпечує уніфікований інтерфейс для створення моделей із заданими параметрами архітектури. Цей модуль відокремлює визначення архітектур від адаптаційної логіки, що спрощує підтримку коду й забезпечує послідовність конфігурацій різних моделей.

AdaptiveModelCreation – основний компонент, що динамічно будує нейронну мережу, адаптовану до рівня шуму. Цей модуль має два режими роботи – вибір моделі на основі SNR та вибір моделі на основі назви ваг. Вибір

моделі на основі SNR полгає в отриманні сигналів , для яких обчислюється значення SNR в класі SNRCalculator. Далі, залежно від значення SNR, створюється відповідна архітектура CNN\_SOP. На рисунку 4.7. представлені класи, що використовуються в рамках для адаптивного створення моделей:

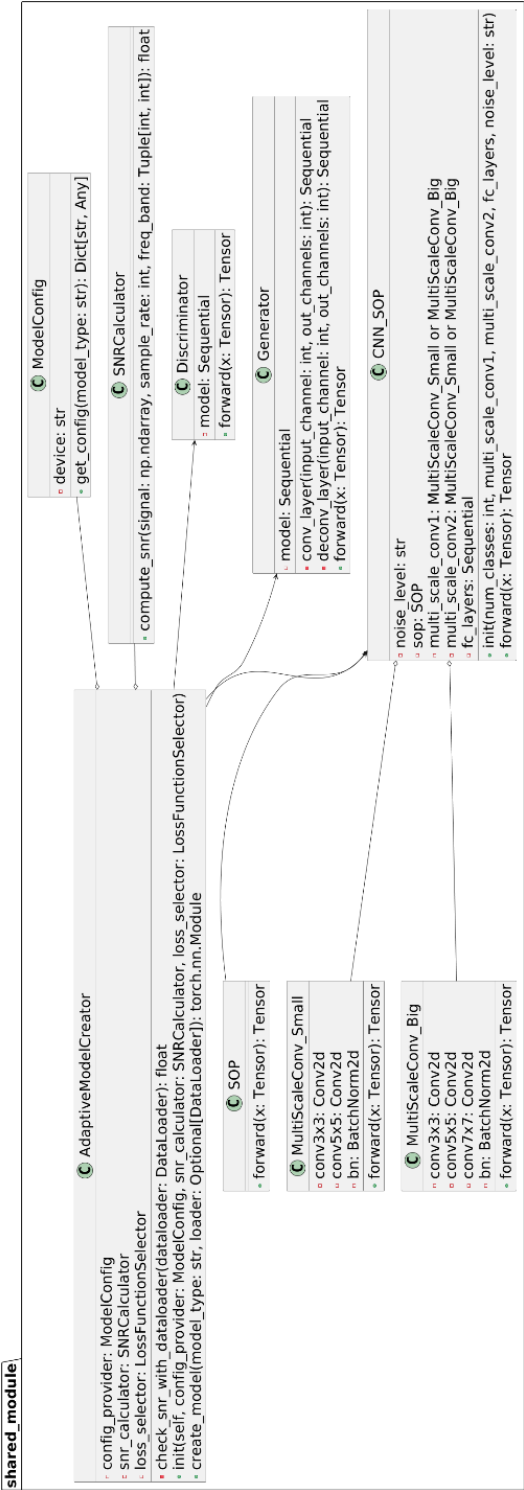


Рис 4.7. Структура класів адаптивного створення моделей

Для низького рівня шуму створюється простіша модель із меншою кількістю шарів та меншою кількістю нейронів. Для умов високого шуму будується складніша модель, що передбачає більшу кількість згорток в MSC та більшу кількість повнозв'язних шарів. Завантаження попередньо навчених моделей для класифікації передбачає визначення типу архітектури за назвою вхідних ваг і створює відповідну конфігурацію моделі перед завантаженням ваг. Цей адаптивний підхід дозволяє збалансувати складність моделі й умови шуму, забезпечуючи ефективність роботи системи з даними що записані в різних умовах підводного середовища. Для задач тренування нейромереж для придушення шумів та виконання придушення шумів, адаптивне створення моделей не застосовується.

Науковий результат дисертації полягає у створенні архітектури програмного забезпечення для аналізу акустичних сигналів водного середовища, що імплементує адаптивне створення нейромережових моделей. Запропонований підхід дозволяє значно зменшити зв'язність системи, забезпечуючи модульність та розділення функціональності, що полегшує розширення та підтримку програмного комплексу. Оптимізація коду досягається завдяки централізованому зберіганню параметрів та моделей, що усуває необхідність їх розкиданого розміщення у різних сервісах та підсистемах.

Динамічна зміна архітектури моделі на основі SNR продиктована значною мінливістю підводних акустичних середовищ. Цей адаптивний підхід дозволяє підвищити стійкість до шуму завдяки використанню складнішої MSC в умовах сильного шуму, що покращує точність класифікації через ефективне виділення ознак, стійких до шумових перешкод. Крім того, завдяки використанню двох спеціалізованих моделей, які оптимізовані відповідно до різних рівнів шуму, система демонструє кращу узагальнювальну здатність, що підвищує її загальну точність у різних акустичних умовах. В умовах низького шуму використовується простіша

архітектура, яка значно знижує витрати обчислювальних ресурсів, що особливо важливо в ресурсно обмежених підводних системах. Завдяки цій спеціалізації моделей, система також забезпечує підвищену стійкість до перенаванчання, оскільки дозволяє оптимально налаштовувати регуляризацію залежно від умов середовища. Нарешті, загальна структура системи забезпечує високу модульність і розширюваність, що спрощує впровадження нових методів і дозволяє легко адаптувати її до майбутніх вимог. Отже, адаптивна архітектура моделі забезпечує ефективне використання нейронних мереж для аналізу підводних акустичних сигналів, поєднуючи високу продуктивність із економією ресурсів залежно від умов експлуатації.

#### **4.3. Алгоритмічне представлення програмного засобу аналізу акустичних сигналів водного середовища**

##### **4.3.1. Опис основного алгоритму модуля виконання класифікації акустичних сигналів**

Діаграма описує процес роботи модуля прогнозування, починаючи від запиту клієнта та закінчуючи отриманням передбачених результатів. Процес ініціюється, коли Client надсилає запит на прогнозування. Prediction Service приймає запит і виконує початкову ініціалізацію сервісу. Після цього Adaptive Model Creation створює CNN\_SOP, яка буде використовуватися для виконання прогнозів. Далі Prediction Service завантажує ваги моделі. Якщо під час цього кроку виникає помилка, вона обробляється відповідно, і процес завершується. На наступному етапі File Processing здійснює операцію формування списку аудіофайлів. Якщо на цьому етапі виникає помилка при зчитуванні списку файлів, система обробляє її та завершує роботу. Якщо все працює коректно, починається процес обробки кожного аудіофайлу по черзі. Файл завантажується, після чого Preprocessing проводить витягнення характеристик звуку та їхню подальшу обробку. Підготовлені характеристики передаються в Model, де модель переводиться в режим оцінювання, а потім



виконує прогнозування. Отримані результати прогнозу передаються назад у Prediction Service, який зберігає їх для подальшої обробки. Алгоритм виконання класифікації акустичних сигналів представлено на рисунку 4.8.

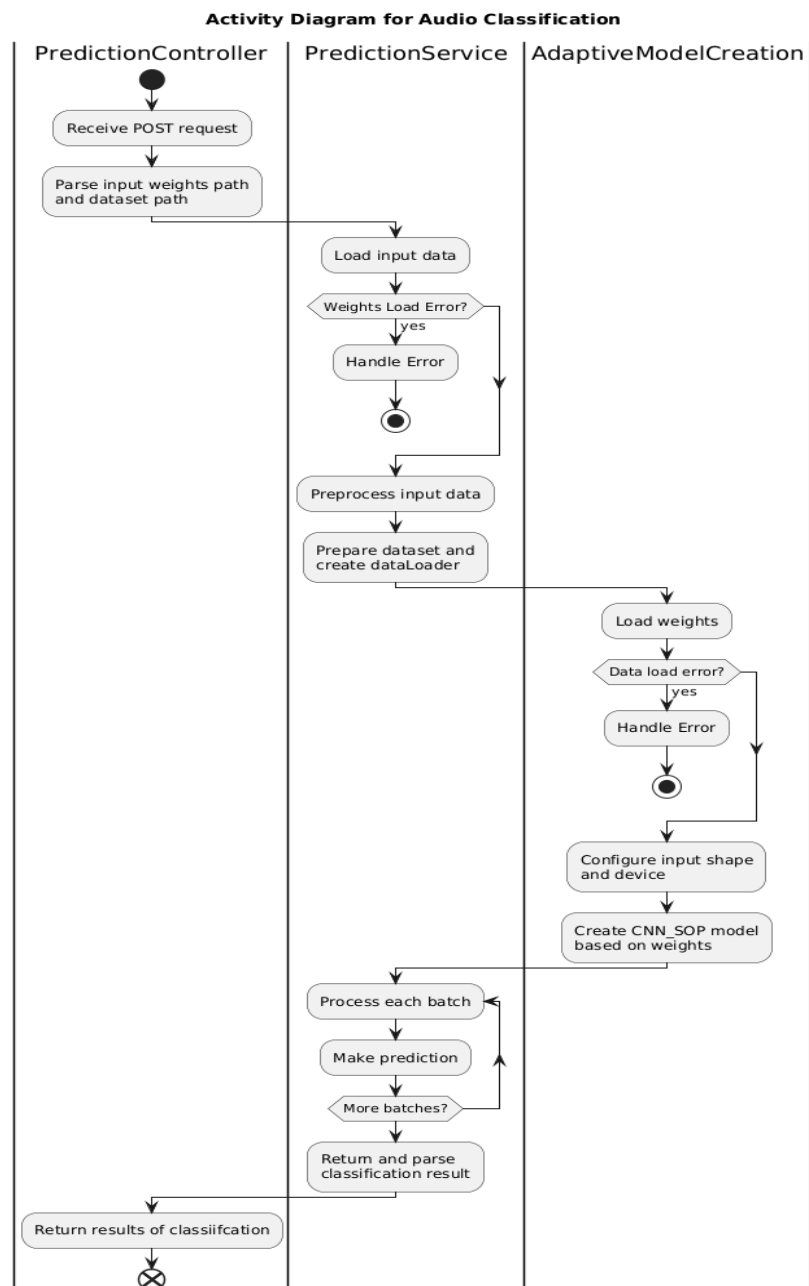


Рис. 4.8. Алгоритм класифікації акустичних сигналів

Процес триває, поки не буде опрацьовано всі доступні аудіофайли. Коли обробка завершується, Prediction Service компілює всі результати в єдиний набір вихідних даних і передає їх Client, який отримує остаточне передбачення. Після цього процес завершується.

#### **4.3.2. Опис основного алгоритму модуля виконання придушення фонових шумів у акустичних сигналах**

Процес видалення шуму з аудіо починається, коли користувач отримує доступ до сторінки денойзингу через веб-інтерфейс. Користувач вказує каталог, що містить вхідний набір даних із зашумленими аудіозразками, і надає шлях до файлу з попередньо навченими вагами нейронної мережі. Після заповнення необхідних полів користувач надсилає запит на запуск операції придушення шумів.

Після отримання запиту `DenoisingController` витягує та аналізує шляхи до набору даних і файлу ваг. Далі система завантажує вхідні дані та намагається отримати вказані ваги моделі. Якщо під час цього процесу виникає помилка, система обробляє проблему та зупиняє виконання. В іншому випадку вона продовжує попередню обробку вхідних даних, структурує набір даних і ініціалізує `DataLoader` для ефективної пакетної обробки.

На наступному етапі `AdaptiveModelCreation` завантажує попередньо навчені ваги, перевіряючи їхню коректну відповідність моделі. Якщо процес завантаження ваг не вдається, система фіксує помилку і припиняє виконання. У разі успіху модель налаштовується шляхом визначення форми вхідних даних і вибору обчислювального пристрою, залежно від доступності. Далі створюється модель генератора, яка пройшла ініціалізацію та готова до придушення шумів.

Далі, в `DenoiseService` розпочинає придушення шумів, де зашумлені аудіозразки обробляються пакет за пакетом. Кожен пакет проходить через створену модель, який виконує придушення фонових шумів. Цей ітераційний процес триває, доки всі доступні пакети не будуть оброблені.

Після завершення процесу придушення, отримані тензори очищених сигналів проходять `ISTFT` для перетворення тензорів в сигнал в часовій

області. Кожен перетворений сигнал зберігається у новоствореній папці. Процес триває, доки всі очищені сигнали не будуть збережені.

Після успішної обробки всіх аудіозразків формується повідомлення про успіх. На рисунку 4.9. представлений алгоритм придушення фонових шумів.

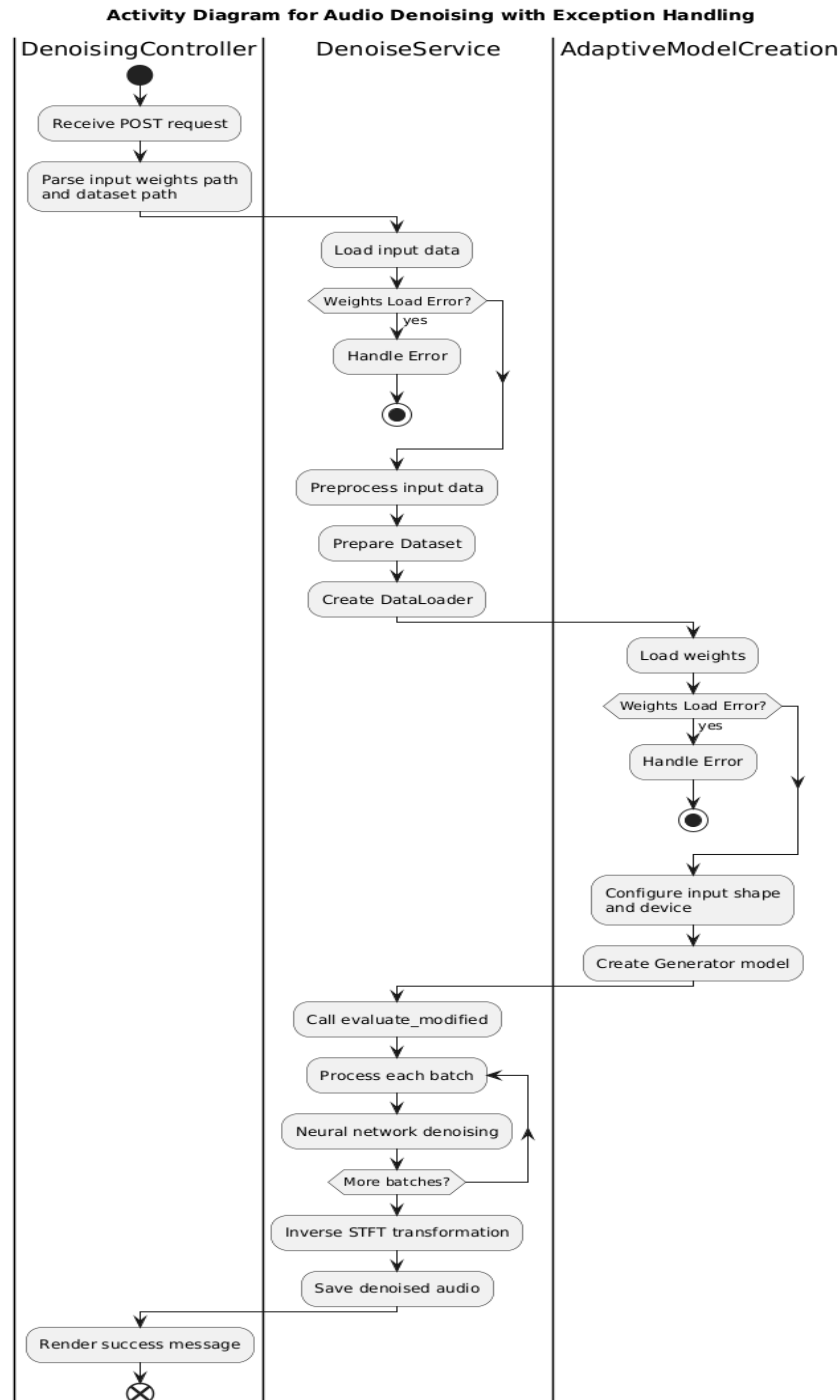


Рис 4.9. Алгоритм придушення шумів акустичних сигналів водного середовища

#### **4.3.3. Опис основного алгоритму тренування нейромережі для придушення фонових шумів водного середовища**

Процес навчання модуля денойзингу починається з підготовки даних, де завантажуються пари чистих і зашумлених аудіозразків. Якщо під час завантаження файлів виникає помилка, система належним чином обробляє її та припиняє виконання, щоб запобігти подальшій обробці некоректних даних. Після успішного завантаження набір даних розділяється на навчальний і валідаційний набори, що дозволяє моделі навчатися на частині даних, залишаючи іншу частину для оцінки.

Після підготовки даних система переходить до етапу створення моделі. Generator і Discriminator налаштовуються з відповідними вхідними розмірами та обчислювальними пристроями, забезпечуючи архітектуру, необхідну для змагального навчання. Після ініціалізації моделей починається етап попередньої обробки, під час якого аудіозразки нормалізуються для приведення їх у стандартний формат. Далі сигнали перетворюються у спектрограми, що спрощує їхню обробку нейронною мережею. Після цього набір даних структурується у мініпакети, що дозволяє ефективно навчання на невеликих частинах даних. Фаза навчання проходить у вигляді ітераційного циклу, де кожен мініпакет із зашумленими та чистими даними обробляється. Generator створює очищені вихідні сигнали, які потім оцінюються Discriminator. Градієнти Discriminator скидаються, і його значення втрати розраховуються окремо для реальних і згенерованих даних перед підрахунком загальних втрат. Виконується зворотне поширення помилки, оновлюючи ваги Discriminator для покращення його здатності розрізняти реальні та згенеровані зразки. Періодично, згідно з визначеною умовою, також скидаються градієнти Generator, і обчислюються кілька компонентів втрат, включаючи adversarial loss, pixel loss і correlation loss. Ці втрати вимірюють ефективність Generator у створенні якісних очищених сигналів, зберігаючи при цьому їхню початкову структуру. Ваги Generator потім оновлюються через обчислення оберненого

проходу неймережі . Цикл тренування протягом епохи триває доти, доки всі мініпакети в епосі не будуть оброблені, після чого Generator переходить у режим оцінювання. Алгоритм тренування неймережі для придушення фонових шумів водного середовища представлений на рисунку 4.10.

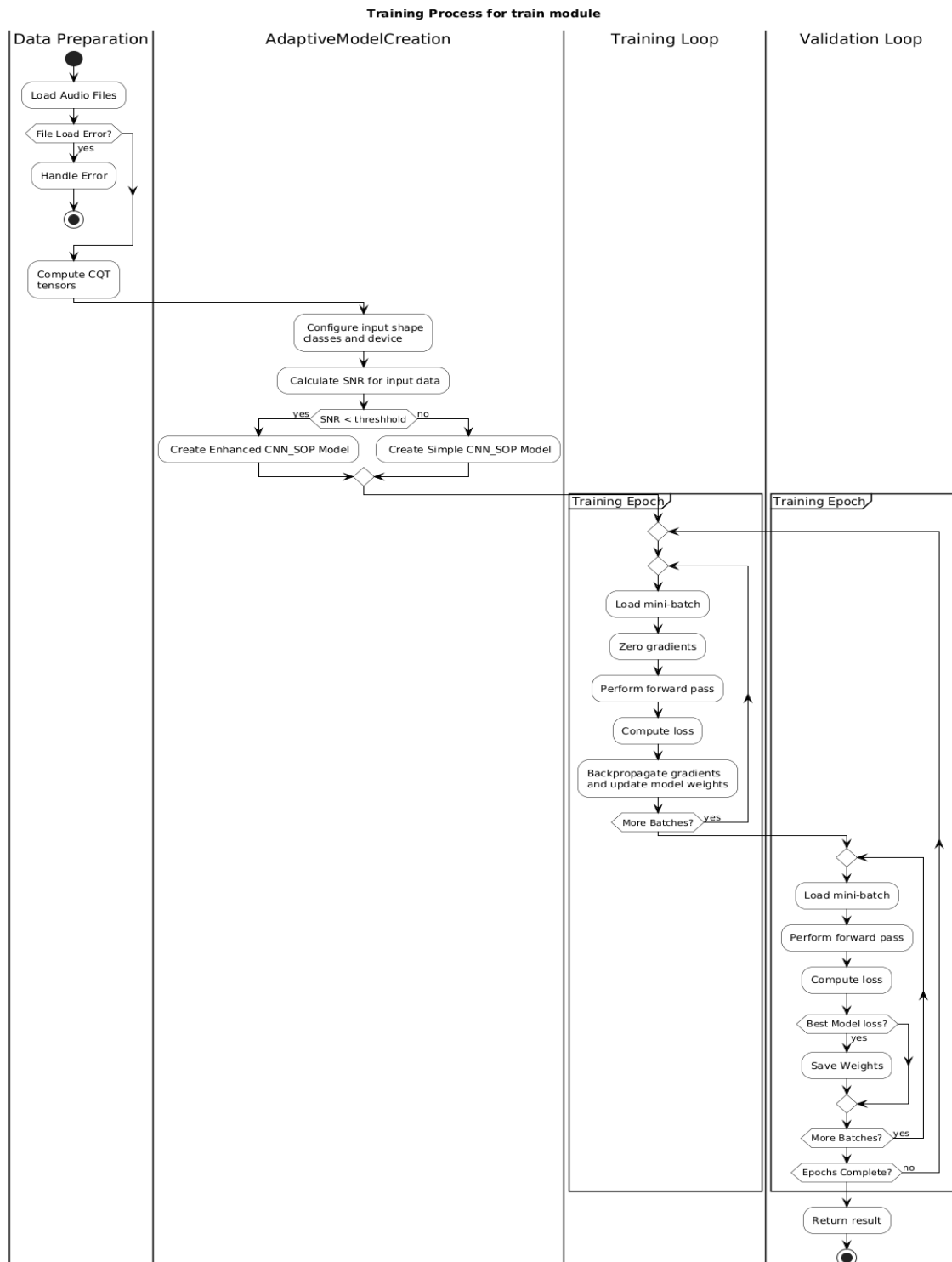


Рис. 4.10. Алгоритм тренування неймережі для придушення шумів

Під час валідаційної фази завантажуються мініпакети валідаційних даних, і Generator обробляє їх, створюючи очищені виходи. Discriminator оцінює як реальні, так і згенеровані зразки, обчислюючи свої відповідні втрати. Втрати Generator також розраховуються аналогічно процесу навчання. Якщо Generator досягає кращого значення втрат ніж Discriminator, його ваги зберігаються, щоб зафіксувати оптимальний стан моделі. Аналогічно, якщо Discriminator досягає значення функції втрат меншої, ніж найкраще значення для Generator, його ваги також зберігаються. Протягом усього процесу накопичуються валідаційні втрати, що дозволяє оцінити загальну продуктивність моделі на валідаційному наборі даних. Навчання триває протягом заданої кількості епох, повторюється процес навчання і валідації, поки не буде досягнуто задану кількість епох. Після завершення навчання система генерує результат тренування.

Навчання триває, доки не буде досягнута задана кількість епох. Після завершення процесу система повертає результат навчання, що підсумовує результат моделі протягом усіх епох. Процес завершується, гарантуючи, що навчені моделі та їхні результати готові до розгортання або подальшого аналізу.

#### **4.3.4. Опис основного алгоритму модуля виконання тренування нейромережі класифікації акустичних сигналів водного середовища**

Процес структурований у кілька фаз, забезпечуючи покроковий робочий процес для підготовки даних, налаштування моделі, навчання, валідації та фінального звітування.

Процес починається з підготовки даних, де система спочатку завантажує аудіофайли в пам'ять. Якщо під час завантаження виникає проблема, активується механізм обробки помилок, що запобігає продовженню навчального процесу. Якщо файли успішно завантажено, аудіодані

розбиваються на тренувальний та валідаційний набори, що дозволяє зберегти частину даних для оцінки моделі.

Після підготовки набору даних система переходить до адаптивного створення моделі, де спочатку налаштовує основні вхідні параметри, такі як форма вхідних даних, тип пристрою та кількість вихідних класів. Для вибору відповідної архітектури моделі значення SNR. Якщо SNR нижче за заданий поріг, створюється покращена модель CNN\_SOP, яка містить додаткові шари та механізми для роботи в зашумлених умовах. Якщо ж SNR перевищує поріг, обирається спрощена модель CNN\_SOP для підвищення обчислювальної ефективності.

Далі система переходить до фази передобробки, де аудіодані нормалізуються, щоб забезпечити узгоджені рівні амплітуди у вибірці. Потім застосовується SQT для отримання частотного представлення аудіосигналів. Після перетворення отримані дані розбиваються на мініпакети, що дозволяє навчальному процесу обробляти невеликі частини набору даних за раз. Коли мініпакети підготовлено, починається процес навчання.

Навчальний цикл у межах епохи навчання виконується багаторазово для ітеративної обробки набору даних. Для кожного мініпакета система завантажує дані з тренувального набору, переміщує їх на обчислювальний пристрій та скидає градієнти, щоб забезпечити коректне зворотне поширення помилки. Потім модель виконує прямий прохід, генеруючи передбачення на основі вхідних даних. Система обчислює функцію втрат, що визначає, наскільки добре передбачення моделі відповідають фактичним міткам. На етапі зворотного поширення помилки параметри моделі оновлюються шляхом обчислення градієнтів і коригування ваг. У міру просування навчання система обчислює значення функції втрат, обчислює передбачення та оновлює кількість правильно класифікованих зразків. Ці передбачення зберігаються для подальшого аналізу. Тренувальний цикл триває, поки всі мініпакети в поточній епосі не будуть оброблені, після чого система переходить до фази валідації. Тренування класифікатор представлене на рисунку 4.11.

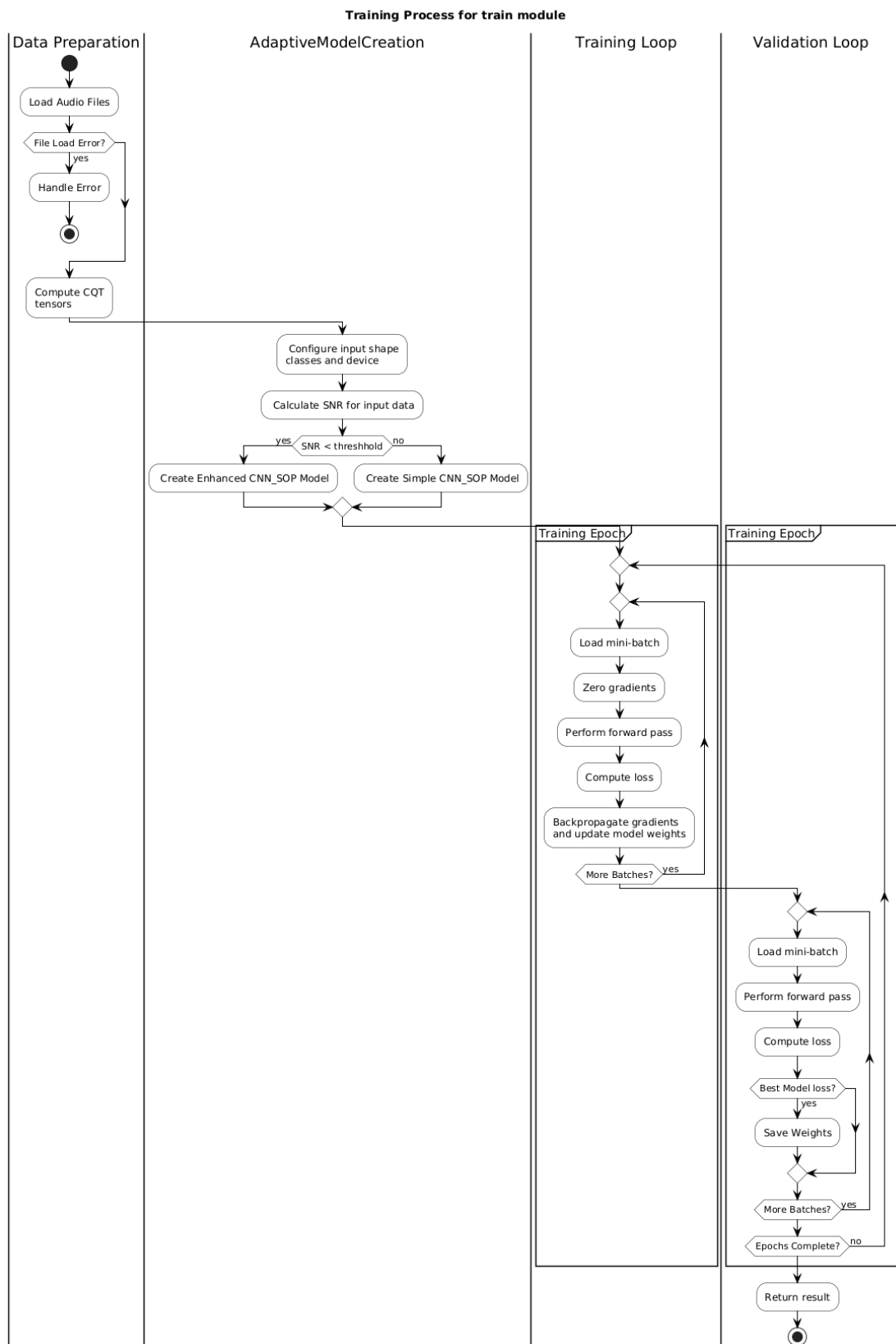


Рис. 4.11. Алгоритм тренування нейромережі для класифікації акустичних сигналів водного середовища



Під час валідації система спочатку вимикає обчислення градієнтів за допомогою, що зменшує використання пам'яті та підвищує ефективність. Цикл валідації обробляє мініпакети аналогічно навчанню: спочатку завантажує мініпакет із валідаційного набору, переміщує його на обчислювальний пристрій і виконує прямий прохід для генерації передбачень. Далі обчислюється валідаційна функція втрат, щоб оцінити, наскільки добре модель узагальнює знання на нових зразках. Якщо поточне значення втрат є найнижчим серед усіх попередніх, система зберігає ваги моделі, забезпечуючи збереження найкращої версії моделі. Цикл валідації триває, поки всі мініпакети не будуть оброблені.

Повний цикл навчання та валідації повторюється протягом 300 епох, що дозволяє моделі поступово вдосконалювати ваги та покращувати точність класифікації. Після завершення всіх епох система переходить у фазу завершення, де генерує фінальний звіт про навчання. У ньому підсумовуються ключові метрики продуктивності, зокрема графіки функції втрат, тенденції точності та результати валідації. Після цього процес завершується.

## **Висновки до розділу 4**

У результаті дослідження було розроблено програмне забезпечення для аналізу акустичних сигналів водного середовища, що складається з декількох модулів, кожен з яких виконує певну функцію у процесі обробки сигналів. Програмне забезпечення побудоване на основі Flask із використанням Python та бібліотек для машинного навчання, що забезпечує його ефективність та гнучкість.

Архітектура програмного забезпечення включає:

- Модуль тренування нейромережі для придушення шумів, де у користувача є можливість обрати датасет для тренування та параметри

тренування та провести його з можливістю зупинки або припинення тренування.

- Модуль тренування нейромережі для класифікації, де у користувача є подібні можливості до модуля тренування нейромережі для придушення шумів.

- Модуль придушення шуму, що приймає оброблені сигнали та видаляє з них небажані шуми.

- Модуль класифікації, який виконує класифікацію обраних семплів та виконує їх розподіл за класами.

- Модуль аналізу даних, що дозволяє користувачеві обирати та переглядати отримані результати.

- Спільний модуль, що містить пайплайн адаптивного створення нейронних мереж та містить класи, що можуть спільно використовуватись багатьма модулями.

Завдяки модульній структурі система може бути легко адаптована до нових задач акустичного аналізу. Реалізована архітектура дозволяє ефективно обробляти акустичні сигнали та отримувати високоточні результати в реальному часі.

У рамках роботи було розроблено пайплайн для адаптивного створення нейромережних моделей, який дозволяє автоматично конфігурувати моделі перед навчанням залежно від характеристик вхідних даних. Запропонована структура покращує масштабованість та ефективність роботи, зменшуючи необхідність ручного налаштування параметрів та забезпечуючи централізоване керування конфігураціями та покращує точність класифікації при ньькому співвідношенні SNR.

Алгоритмічне представлення програмного засобу включає чітко визначені модулі для класифікації та придушення шумів, що дозволяють гнучко обробляти акустичні сигнали у реальному часі. Запропоновані алгоритми забезпечують ефективне використання ресурсів, оптимізують

роботу з великими обсягами даних та підвищують точність класифікації та денойзингу.

Розроблене програмне забезпечення має потенціал для застосування в галузях моніторингу підводного середовища, акустичної діагностики та аналізу шумових характеристик у різних середовищах.

## ВИСНОВКИ

У результаті дисертаційного дослідження вирішено наукове завдання *розробка методів та програмних засобів аналізу акустичних сигналів водного середовища, для виявлення та аналізу сигналу цільового об'єкта з низьким відношенням рівня шуму.*

В дисертації одержано наступні результати:

Одержані результати запропонованого методу полягають у наступному. Удосконалено метод аналізу, а саме класифікації акустичних сигналів водного середовища на основі згорткової нейромережі SOP, яка, на відміну від оригінальної, містить багатомасштабну згортку, що дозволяє вилучати характеристики з різною роздільною здатністю; середній пулінг, що дозволяє зменшувати розмірності вхідних даних; та модифікований повнозв'язний шар з функцією активації ReLU, додавши до нього пакетну нормалізацію, що забезпечує кращу узагальнюючу здатність.

Подальшого розвитку набув метод придушення фонових шумів акустичних сигналів водного середовища на основі нейромережі UWAR-GAN для вилучення сигналу водного судна шляхом придушення фонових шумів водного середовища у зашумленому сигналі.

Вперше розроблено архітектуру програмного забезпечення для аналізу акустичних сигналів водного середовища, що реалізує адаптивний пайплайн, що передбачає адаптивне створення нейромережевих моделей, залежно від характеристик вхідних даних. Програмне забезпечення має можливість виконання придушення фонових шумів у зашумлених вхідних сигналах з застосуванням нейронних мереж та класифікацію вхідних сигналів водного середовища. Також, програмне забезпечення надає можливість тренувати запропоновані нейромережі для придушення та аналізу водного середовища на різних вибірках даних, аналізувати обрані семпли акустичних сигналів водного середовища, виводячи різні характеристики обраних сигналів.

Мета досліджень, яка полягала в покращенні методів аналізу акустичних сигналів суден водного середовища на основі нейромережових моделей, а саме підвищенні точності класифікації, після застосування запропонованих методів, досягнута.

Перспективними напрямками подальших досліджень є подальше покращення методу шляхом імплементація та застосування нейромережі для розширення вибірки акустичних сигналів водного середовища та нейромережі для визначення нейромережі для визначення напрямку приходу акустичного сигналу водного середовища.

Результати дослідження свідчать про науковий і практичний внесок у розв'язання задач аналізу акустичних сигналів водного середовища шляхом впровадження сучасних нейромережових технологій. У межах роботи запропоновано нові методи, що базуються на використанні генеративно-змагальної нейромережі UWAR-GAN для придушення шумів і згорткових нейронних мереж із пулінгом другого порядку CNN SOP для класифікації очищених сигналів. Вперше реалізоване інтегроване рішення, яке забезпечує високоякісний аналіз акустичних даних за умов складних шумових середовищ, демонструючи покращення метрик точності.

Проведена математична формалізація задач класифікації та придушення шумів дала змогу адаптувати моделі до специфіки водного середовища, враховуючи особливості акустичних характеристик. Значна увага приділена створенню та аналізу синтетичних і природних датасетів, що дозволяє тестувати та вдосконалювати розроблені методи в умовах реальних задач.

Результати дослідження продемонстрували ефективність поєднання багатомасштабної згортки, шару SOP та методів часово-частотного представлення даних, що забезпечило підвищення точності класифікації навіть за умов низького співвідношення сигнал/шум. Модифікація нейромережі UWAR GAN дає можливість якісно очищувати фонові шуми за менший проміжок часу. Запропоновані архітектури суттєво розширюють

можливості нейромережових технологій у задачах аналізу акустичних сигналів.

Практичне значення роботи полягає у створенні програмного забезпечення, яке може бути використане для задач морської навігації, гідроакустики, екологічного моніторингу та інших галузей. Висока автономність розроблених систем, їх стійкість до шумів і здатність до адаптації відкривають нові можливості для розв'язання складних прикладних задач у реальних умовах.

Таким чином, виконана робота є вагомим внеском у розвиток теорії та практики обробки акустичних сигналів і створює основу для подальших досліджень у сфері штучного інтелекту та аналізу даних у морському середовищі.

## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Олексій А.О., Верлань А.А. Огляд та порівняння методів машинного навчання для розпізнавання гідроакустичних сигналів. Науковий журнал «Інфокомунікаційні та комп'ютерні технології», № 1 (03). 2022. С. 113 – 120. <https://doi.org/10.36994/2788-5518-2022-01-03-18>.
2. Олексій А.О., Верлань А. А. Застосування багат шарового перцептрона для аналізу акустичних сигналів у водному середовищі/ Науковий журнал «Зв'язок», Випуск №3 (163), 2023. С. 40 – 45. <https://doi.org/10.31673/2412-9070.2023.039000>.
3. Олексій А., Пуха Г. Створення датасету акустичних сигналів водного середовища для тренування нейромережі для придушення шумів. Науковий журнал «Інформаційні технології та суспільство», Випуск 2 (13). 2024. С.56-60. <https://doi.org/10.32689/maup.it.2024.2.8>.
4. Олексій, А. О. і Верлань А. А., 2024. Поліпшений метод аналізу акустичних сигналів водного середовища на основі згорткової нейромережі SOP. Вісник Вінницького політехнічного інституту, №6, с. 129-134. <https://doi.org/10.31649/1997-9266-2024-177-6-129-134>.
5. Олексій, А. О. і Верлань А. А., 2022. Огляд нейромережевих підходів для аналізу акустичних сигналів водного середовища. Теоретичні і прикладні проблеми фізики, математики та інформатики., зб. тез XX Всеукраїнської науково-практичної конференції студентів, аспірантів та молодих вчених, 15 червня 2022 р., м. Київ, с. 347-350.
6. Олексій, А. О., Верлань А.І. Нейромережеві підходи до генерації акустичних сигналів водного середовища., Сучасні проблеми наукового забезпечення енергетики: зб. тез XX Міжнародної науково-практичної конференції молодих вчених і студентів, Т. 2, м. Київ, 25–28 квітня 2023 року с. 128-129.

7. Олексій, А. О., Верлань А.І. 2023. Аналіз акустичних сигналів водного середовища з застосуванням нейромережевих методів. Future of Work: Technological, Generational and Social Shifts: 2nd International Scientific and Practical Internet Conference., May 11-12, 2023, Dnipro, Ukraine, p. 67-68.
8. Олексій, А. О., Верлань А.І. Створення датасету акустичних сигналів водного середовища. Системи та технології зв'язку, інформатизації та кібербезпеки: актуальні питання і тенденції розвитку: матеріали III Міжнародної науково-технічної конференції, 30 листопада 2023 року, м. Київ, с. 229.
9. Олексій, А. О., Верлань А.І. Огляд нейромережевих підходів для придушення шумів водного середовища. Системи та засоби штучного інтелекту: тези доповідей Міжнародної наукової молодіжної школи., 10-11 жовтня 2023 р. м. Київ, с. 42-51.
10. Олексій, А. О., Варава І.А. 2023. Аналіз акустичних сигналів водного середовища з застосуванням нейромережевих моделей та створення датасетів акустичних сигналів водного середовища. Сучасні аспекти інженерії програмного забезпечення: збірник тез Першої міжнародної науково-практичної конференції, 14 грудня 2023 року, м. Київ, с. 100-103.
11. Олексій, А. О., Верлань А.І. Аналіз акустичних сигналів водного середовища з застосуванням нейромережевих моделей та створення датасетів акустичних сигналів водного середовища. Сучасні аспекти інженерії програмного забезпечення: збірник тез Другої міжнародної науково-практичної конференції, 13 листопада 2024 року, м. Київ, с..
12. Urick, R. J. (1983). Principles of Underwater Sound. McGraw-Hill.
13. Kinsler, L. E., Frey, A. R., Coppens, A. B., & Sanders, J. V. (1999). Fundamentals of Acoustics. Wiley.
14. Wenz, G. M. (1962). Acoustic ambient noise in the ocean: Spectra and sources. The Journal of the Acoustical Society of America, 34(12), 1936–1956. <https://doi.org/10.1121/1.1909155>.
15. Ross, D. (1976). Mechanics of Underwater Noise. Pergamon Press.



16. Morris, R. (1981). Influence of seabed topography on underwater noise propagation. *Ocean Acoustics Journal*, 8(3), 45–58.
17. Tolstoy, I., & Clay, C. S. (1987). *Ocean Acoustics: Theory and Experiment in Underwater Sound*. Springer-Verlag.
18. SpectraPLUS: <https://www.cetaceanresearch.com/signal-analysis/index.html>.
19. Bienvenu, K. (2017). Underwater Acoustic Signal Analysis Toolkit. University of New Orleans. Available at: ScholarWorks@UNO
20. Sigview: <https://www.sigview.com/>.
21. DaViS Underwater Acoustic Data Visualization Software: [https://www.abysens.fr/davis\\_underwater-acoustic\\_data\\_visualisation\\_software/](https://www.abysens.fr/davis_underwater-acoustic_data_visualisation_software/).
22. SignalSOFT: [https://sonartech.net/product-detail.php?prd=yM3XwjXmk\\_UHMjyCjggfeilsSNK12X0ejsgbnNSxYKo](https://sonartech.net/product-detail.php?prd=yM3XwjXmk_UHMjyCjggfeilsSNK12X0ejsgbnNSxYKo).
23. Hamilton, J.D. (1994). *Time Series Analysis*. Princeton: Princeton University Press.
24. Rabiner, L.R. (1989). A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition. *Proceedings of the IEEE*, 77(2), pp.257-286. <https://doi.org/10.1109/5.18626>.
25. Reynolds, D.A., and Rose, R. C. (1995). Robust Text-Independent Speaker Identification Using Gaussian Mixture Speaker Models. *IEEE Transactions on Speech and Audio Processing*, 3(1), pp.72-83. <https://doi.org/10.1109/89.365379>.
26. Gemba, K.L., Meyer, F., and Stinco, P. (2021). Bayesian Acoustic Source Localization in Shallow Water Environments. *Journal of the Acoustical Society of America*, 149(4), pp.2521-2535.
27. Koller, D., and Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. Cambridge, MA: MIT Press.
28. Doucet, A., de Freitas, N., and Gordon, N. (2001). *Sequential Monte Carlo Methods in Practice*. New York: Springer.

29. Murphy, K. (2012). *Machine Learning: a Probabilistic Perspective* (1sted). Cambridge, MA, US.
30. Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer, A, MIT Press.
31. Liu, Hongji & Tian, Yadong & Yu, Donghong. (2024). Prediction of cardiovascular and cerebrovascular diseases based on machine learning models. *Applied and Computational Engineering*. 46. 35-44. <https://doi.org/10.54254/2755-2721/46/20241068>.
32. T. Hastie, R. Tibshirani, and J. H. Friedman, *The elements of statistical learning: data mining, inference, and prediction*, Second. New York, NY, USA: Springer, 2009. ISBN: 978-0387848570.
33. Jagtap, S., Wagh, J., Kasabekar, A., Mogal, K., Nene, T. and Shiravale, S. (2024) 'Identification of Musical Notes using Support Vector Machine', *Nanotechnology Perceptions*, 20(S6). <https://doi.org/10.62441/nano-ntp.vi.1331>.
34. H. Robotham, J. Castillo, P. Bosch and J. Perez-Kallens, "A comparison of multi-class support vector machine and classification tree methods for hydroacoustic classification of fish-schools in Chile", *Fisheries Res.*, vol. 111, no. 3, pp. 170-176, Nov. 2011. <https://doi.org/10.1016/j.fishres.2011.07.010>.
35. Telesphera (2021) 'Методи для виявлення і діагностика несправностей стосовно IoT'. Available at: <https://www.telesphera.net/blog/viyavlennya-i-diegnostika-iot.html>.
36. Cardoso, J. M. (2019). *Unsupervised representation learning for anomaly detection on neuroimaging. Application to epilepsy lesion detection on brain MRI* (Doctoral dissertation, Université de Rennes 1). Samnioti, A., Kanakaki, E.M., Fotias, S.P. and Gaganis, V., 2023. p. 14.
37. Rapid Hydrate Formation Conditions Prediction in Acid Gas Streams. *Fluids* 2023, 8, p. 226. <https://doi.org/10.3390/fluids8080226>.

38. Wikipedia contributors (n.d.) Neural network (machine learning). Wikipedia, The Free Encyclopedia. [https://en.wikipedia.org/wiki/Neural\\_network\\_\(machine\\_learning\)](https://en.wikipedia.org/wiki/Neural_network_(machine_learning)).
39. Кушнір, Н.О., Локтікова, Т.М., Морозов, А.В. and Юрченко, В.О., 2022. Використання згорткових нейронних мереж у задачах розпізнавання та класифікації об'єктів зображень. С. 96. [https://doi.org/10.26642/ten-2022-1\(89\)-93-100](https://doi.org/10.26642/ten-2022-1(89)-93-100).
40. Chin-Hsing Chen, Jiann-Der Lee, and Ming-Chi Lin, Classification of underwater signals using wavelet transforms and neural networks. Mathematical and Computer Modelling, 27(2): pp. 47-60. (1998).
41. Asimov Institute (n.d.) The neural network zoo. <https://www.asimovinstitute.org/neural-network-zoo/>.
42. Ding, S., Xu, X., & Nie, R. (2014). Extreme learning machine and its applications. Neural Computing and Applications, 25(3), 549-556. <https://doi.org/10.1007/s00521-013-1522-8>.
43. Kiranyaz, S., Avci, O., Abdeljaber, O., Ince, T., Gabbouj, M., & Inman, D. J. (2021). 1D convolutional neural networks and applications: A survey. Mechanical systems and signal processing, 151, 107398. <https://doi.org/10.1016/j.ymssp.2020.107398>.
44. D. Rethage, J. Pons and X. Serra, "A Wavenet for Speech Denoising," 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018, pp. 5069-5073, <https://doi.org/10.1109/ICASSP.2018.8462417>.
45. Graves, Alex; Mohamed, Abdel-rahman; Hinton, Geoffrey. Speech Recognition with Deep Recurrent Neural Networks (англ.) // Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on: journal. - 2013. - P. 6645-6649. <https://doi.org/10.48550/arXiv.1303.5778>.
46. Hinton, G. E.; Salakhutdinov, R. R. (2006). "Reducing the Dimensionality of Data with Neural Networks" (PDF). Science. 313 (5786): 504–507. <https://doi.org/10.1126/science.1127647>.

47. Larochelle, H.; Bengio, Y. (2008). Classification using discriminative restricted Boltzmann machines (PDF). Proceedings of the 25th international conference on Machine learning - ICML '08. p.536. <https://doi.org/10.1145/1390156.1390224>.
48. Coates, Adam; Lee, Honglak; Ng, Andrew Y. (2011). An analysis of single-layer networks in unsupervised feature learning (PDF). International Conference on Artificial Intelligence and Statistics (AISTATS).
49. Sarikaya, R., Hinton, G. E., & Deoras, A. (2014). Application of deep belief networks for natural language understanding. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 22(4), 778-784. <https://doi.org/10.1109/TASLP.2014.2303296>
50. Vinay, B. (2023) 'Deep Belief Networks', Medium, 5 July. Available at: <https://medium.com/@vinaybv1ai/deep-belief-networks-490e2efb5e2a>
51. Venkatesan, R. Shelke, Maya Sharmila, L. &Aishwarya, R. Umadevi, G. &Ishwarya, M. (2024). Authentication Based Remote System for the Interaction of Reservation Scheme of Cloud with IoT Networks. Wireless Personal Communications. <https://doi.org/10.1007/s11277-024-11231-y>.
52. Makhzani, A., & Frey, B. (2013). K-sparse autoencoders. arXiv preprint arXiv:1312.5663. <https://doi.org/10.48550/arXiv.1312.5663>.
53. Vincent, Pascal; Larochelle, Hugo (2010). "Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion". Journal of Machine Learning Research. 11: 3371–3408.
54. He, J., Liu, L., Zhang, C., Zhao, K., Sun, J., & Li, P. (2018). Deep denoising autoencoding method for feature extraction and recognition of vehicle adhesion status. Journal of Sensors, 2018. <https://doi.org/10.1155/2018/5419645>.
55. V.-S. Doan, T. Huynh-The and D.-S. Kim, "Underwater acoustic target classification based on dense convolutional neural network", IEEE Geosci. Remote Sens. Lett., Oct. 2020, pp. 1-5. <https://doi.org/10.1109/LGRS.2020.3029584>.

56. Chen, J., Liu, J., Liu, C., Zhang, J., Han, B.: Underwater target recognition based on multi-decision lofar spectrum enhancement: a deep learning approach. arXiv preprint arXiv:2104.12362 (2021) <https://doi.org/10.48550/arXiv.2104.12362>.
57. Han XC, Ren C, Wang L, Bai Y (2022) Underwater acoustic target recognition method based on a joint neural network. PLoS ONE 17(4): e0266425. <https://doi.org/10.1371/journal.pone.0266425>.
58. X. Cao, R. Togneri, X. Zhang and Y. Yu, Convolutional Neural Network With Second-Order Pooling for Underwater Target Classification, in IEEE Sensors Journal, vol. 19, no. 8, pp. 3058-3066, 15 April 2019, <https://doi.org/10.1109/JSEN.2018.2886368>.
59. A new cooperative deep learning method for underwater acoustic target recognition / Yang Honghui, Xu Guanghui, Yi Shuzhen, and Li Yiqing // OCEANS 2019-Marseille / IEEE. - 2019. - P. 1-4. <https://doi.org/10.1109/OCEANSE.2019.8867490>.
60. Underwater target classification at greater depths using deep neural network with joint multiple-domain feature / Cao Xu, Zhang Xiaomin, Togneri Roberto, and Yu Yang // IET Radar, Sonar & Navigation. - 2019. - Vol. 13, no. 3. - P. 484–491. <https://doi.org/10.1049/iet-rsn.2018.5279>.
61. Muhammad Irfan, Zheng Jiangbin, Shahid Ali, Muhammad Iqbal, Zafar Masood, Umar Hamid, DeepShip: An underwater acoustic benchmark dataset and a separable convolution based autoencoder for classification, Expert Systems with Applications, Volume 183, 2021, 115270, ISSN 0957-4174. <https://doi.org/10.1016/j.eswa.2021.115270>.
62. Luo, X.; Feng, Y. An Underwater Acoustic Target Recognition Method Based on Restricted Boltzmann Machine. Sensors 2020, 20, 5399. <https://doi.org/10.3390/s20185399>.
63. Khishe Mohammad, Mohammadi Hassan. Passive sonar target classification using multi-layer perceptron trained by salp swarm algorithm // Ocean

Engineering.— 2019.—Vol. 181.—P. 98–108.  
<https://doi.org/10.1016/j.oceaneng.2019.04.0>.

64. Ashraf, H., Jeong, Y. and Lee, C. H., 2021. Underwater ambient-noise removing GAN based on magnitude and phase spectra. *IEEE Access*, 9, pp.24513-24530. <https://doi.org/10.1109/ACCESS.2021.3051263>.
65. Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in neural information processing* 912 systems, pages 2672–2680, 2014.
66. Song, Y., Chu, Q., Liu, F., Wang, T. and Shen, T., 2023. Underwater Acoustic Signal Noise Reduction Based on a Fully Convolutional Encoder-Decoder Neural Network. *Journal of Ocean University of China*, 22(6), pp.1487-1496. <https://doi.org/10.1007/s11802-023-5458-z>.
67. Song, R., Feng, X., Wang, J., Sun, H., Zhou, M. and Esmail, H., 2024. Underwater Acoustic Nonlinear Blind Ship Noise Separation Using Recurrent Attention Neural Networks. *Remote Sensing*, 16(4), p.653. <https://doi.org/10.3390/rs16040653>.
68. Ashraf, H., Shah, B., Soomro, A.M., Halim, Z. and Shah, S.K., 2022. Ambient-noise free generation of clean underwater ship engine audios from hydrophones using generative adversarial networks. *Computers and Electrical Engineering*, 100, p.107970. <https://doi.org/10.1016/j.compeleceng.2022.107970>.
69. Luo, Y., Chen, Z. and Yoshioka, T., 2020, May. Dual-path rnn: efficient long sequence modeling for time-domain single-channel speech separation. In *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 46-50). IEEE.
70. Song, Y., Liur, F. and Shen, T., 2022. Method of underwater acoustic signal denoising based on dual-path transformer network. *IEEE Access*.
71. S. W. Abeyruwan, D. Sarkar, F. Sikder, and U. Visser, “Semi-automatic extraction of training examples from sensor readings for fall detection and

- posture monitoring,” IEEE Sensors Journal, vol. 16, no. 13, pp. 5406–5415, 2016.
72. Мартсенюк. В. П. (2020) 'Огляд методів оптимізації: градієнтний спуск та його варіації', Актуальні задачі сучасних технологій, 2, сс. 44–45.
  73. D. P. Kingma, J. Ba, "Adam: A Method for Stochastic Optimization", International Conference on Learning Representations (ICLR), 2015. N <https://doi.org/10.48550/arXiv.1412.6980>.
  74. Jones, M., and Perez, L. (2020). Time-Frequency Analysis of Acoustic Signals. Cambridge: Cambridge University Press.
  75. Kuchеров, O., Novikov, I., and Smirnov, A. (2023). "Time-domain Characteristics of Marine Acoustic Signals." Journal of Marine Acoustics, 12(3), 145-162.
  76. Novikov, I., et al. (2019). Basics of Underwater Acoustics. Kyiv: Polytechnic Press.
  77. Smith, J., Miller, R., and Wang, X. (2021). "Frequency Domain Signal Analysis in Marine Environments." Acoustics Research Letters, 15(2), 98-110.
  78. Yacouby, R. and Axman, D., 2020. Probabilistic Extension of Precision, Recall, and F1 Score for More Thorough Evaluation of Classification Models. Proceedings of the First Workshop on Evaluation and Comparison of NLP Systems, pp.79-91.
  79. Sitarz, M., 2022. Extending F1 Metric, Probabilistic Approach. Advances in Artificial Intelligence and Machine Learning.
  80. Narasimhan, H., Ramaswamy, H.G., Tavker, S.K., Khurana, D., Netrapalli, P. and Agarwal, S., 2022. Consistent Multiclass Algorithms for Complex Metrics and Constraints. Advances in Neural Information Processing Systems. <https://doi.org/10.48550/arXiv.2210.09695>.
  81. Dyrland, K., Lundervold, A.S. and Porta Mana, P.G.L., 2023. Does the Evaluation Stand Up to Evaluation? A First-Principle Approach to the Evaluation of Classifiers. Machine Learning Journal. <https://doi.org/10.48550/arXiv.2302.12006>.

82. Crall, J., 2023. The MCC Approaches the Geometric Mean of Precision and Recall as True Negatives Approach Infinity. *Journal of Machine Learning Research*. <https://doi.org/10.48550/arXiv.2305.00594>.
83. Fawcett, T. (2006) 'An introduction to ROC analysis', *Pattern Recognition Letters*, 27(8), pp. 861-874. <https://doi.org/10.1016/j.patrec.2005.10.010>.
84. J. A. O'Neill, J. L. M. K. Eldridge, Signal detection and classification of underwater acoustics *IEEE Transactions on Signal Processing*, vol. 61, no. 12, pp. 3124-3132, 2013.
85. S. J. Roberts, Improving classification of acoustic signals in marine environments with generative adversarial networks, *IEEE Transactions on Signal Processing*, vol. 66, no. 5, pp. 1273-1285, 2018.
86. Hayes, M. (1996). *Statistical Digital Signal Processing and Modeling*. Wiley.
87. Khobotov, A.G., Kalinina, V.I., Khil'ko, A.I. and Malekhanov, A.I., 2024. Underwater Acoustic Signal Detection against the Background of Non-Stationary Sea Noise. *Journal of Marine Science and Engineering*, 12(9), p.1540. <https://doi.org/10.3390/jmse12091540>.
88. Oppenheim, A.V. and Schaffer, R.W., 2009. *Discrete-Time Signal Processing*. 3rd ed. Upper Saddle River, NJ: Pearson, pp. 693-698..
89. Proakis, J.G. and Manolakis, D.G., 2006. *Digital Signal Processing: Principles, Algorithms, and Applications*. 4th ed. Upper Saddle River, NJ: Pearson Prentice Hall, pp. 62-65.
90. Willmott, C. J. and Matsuura, K., 2005. Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, 30(1), pp.79-82.
91. Kenney, J.F. and Keeping, E.S. (1962) 'Standard deviation and its applications in statistics', *Mathematics of Statistics*, 3rd edn, New York: Van Nostrand
92. Salomon, D. (2007) *Data Compression: The Complete Reference*. 4th edn. London: Springer, pp. 411-412.
93. Manolakis, D., Ingle, V., & Kogon, S. (2005). *Statistical and Adaptive Signal Processing*. Artech House.

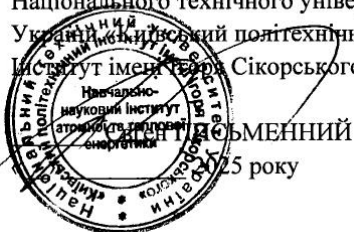


94. Er, M.J., Chen, J., Zhang, Y. and Gao, W., 2023. Research challenges, recent advances, and popular datasets in deep learning-based underwater marine object detection: A review. *Sensors*, 23(4), p.1990. <https://doi.org/10.3390/s23041990>
95. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y., 2014. Generative adversarial nets. *Advances in neural information processing systems*, 27. <https://doi.org/10.48550/arXiv.1406.2661>.
96. Isola, P., Zhu, J.Y., Zhou, T. and Efros, A.A., 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125-1134. <https://doi.org/10.1109/CVPR.2017.632>
97. Ronneberger, O., Fischer, P. and Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18* (pp. 234-241). Springer International Publishing. <https://doi.org/10.48550/arXiv.1505.04597>.
98. Вимірювальні системи та програмне забезпечення для морських охоронних систем і дослідницьких полігонів : звіт про НДР (заключ.) НТУУ "КПІ" ; кер. роб. Мачуський Є.М. - К., 2012. - 104 л. + відеосюжет + CD-ROM. - Д/б №2429-п.
99. Robert, J., Webbie, M., & others. (2018). Pydub. GitHub. Retrieved from <http://pydub.com/>.
100. Amjad Almahairi, Sai Rajeswar, Alessandro Sordoni, Philip Bachman, and Aaron Courville. Augmented cyclegan: Learning many-to-many mappings from unpaired data. *arXiv preprint arXiv:1802.10151*, 2018. <https://doi.org/10.48550/arXiv.1802.10151>.
101. Кононова, К.Ю., 2020. Машинне навчання: методи та моделі: підручник для бакалаврів, магістрів та докторів філософії спеціальності 051 «Економіка». Харків: ХНУ імені ВН Каразіна, 301, р.34.

## ДОДАТОК А

“ПОГОДЖЕНО”

Директор Навчально-наукового  
Інститут атомної та теплової енергетики  
Національного технічного університету  
України «Київський політехнічний  
інститут імені Ігоря Сікорського»



ОЛЕКСІЙ КРАТКО

2025 року

“ЗАТВЕРДЖУЮ”

Директор  
Особливого конструкторського бюро



Олександр КРАТКО

2025 року

### АКТ

Про реалізацію результатів дисертаційної роботи ОЛЕКСІЯ Артура Олеговича на тему  
*«Методи та програмні засоби аналізу акустичних сигналів на основі нейромережевих  
моделей»*

Результати наукових досліджень ОЛЕКСІЯ Артура Олеговича, а саме розроблене спеціальне програмне забезпечення для знешумлення гідроакустичних сигналів використано в науково-дослідницьких та господарських договорах, що ведуться в Особливому конструкторському бюро «ШТОРМ» Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського».

Дане програмне забезпечення дозволяє обробляти сигнали гідроакустичних станцій, підвищуючи їх якість та проводити класифікацію об'єктів шумовипромінювання.

На експериментальних даних було продемонстровано, що розроблений метод придушення шуму за допомогою нейронної мережі показав збільшення відношення «сигнал/шум» на 10 децибел.

В.о. завідувача кафедри Інженерії  
програмного забезпечення в енергетиці  
д.т.н., професор

Олександр КОВАЛЬ

ЗАТВЕРДЖУЮ

проректор з навчальної роботи

Національного технічного університету України

«Київський політехнічний інститут імені Ігоря

Сікорського»

Тетяна ЖЕЛЯСКОВА

«07 09 2025 року

### АКТ

про впровадження в навчальний процес Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського» наукових положень і результатів дисертаційної роботи ОЛЕКСІЯ Артура на тему «Методи та програмні засоби аналізу акустичних сигналів на основі нейромережових моделей» на здобуття наукового ступеня доктора філософії за спеціальності 121 «Інженерія програмного забезпечення».

Комісія в складі голови - в.о. завідувача кафедри Інженерії програмного забезпечення в енергетиці д.т.н., професора Олександра КОВАЛЯ та членів комісії: професора кафедри Інженерії програмного забезпечення в енергетиці д.т.н., професора Олега БАРАБАША та професора кафедри Інженерії програмного забезпечення в енергетиці д.т.н., доцента Андрія МУСІЄНКА, склала цей акт про те, що наукові положення та результати дисертаційної роботи на здобуття наукового ступеня доктора філософії аспіранта Артура ОЛЕКСІЯ на тему «Методи та програмні засоби аналізу акустичних сигналів на основі нейромережових моделей», а саме архітектуру програмного забезпечення аналізу на основі нейронних мереж, метод класифікації акустичних сигналів на основі нейромережі CNN SOP та метод придушення шумів акустичних сигналів на основі нейромережі UWAR GAN введено в навчальний процес кафедри Інженерії програмного забезпечення в енергетиці Навчально-наукового інституту атомної та теплової енергетики при викладанні дисципліни «Методології розробки інтелектуальних комп'ютерних програм» спеціальності 121 «Інженерія програмного забезпечення» за освітнім рівнем підготовки бакалавра (2024-2025):

Лекція: Мережа Хопфілда. Мережа «Машини Больцмана». Мережа Хемінга. Мережа мережної моделі з двонаправленою асоціативною пам'яттю. Мережа адаптивної резонансної теорії (ART). Згорткові нейронні мережі.

Практичне заняття: Практичне ознайомлення з дослідницькими програмами на базі ШІ. Дослідження демонстраційних програмних продуктів

Голова комісії

В.о. завідувача кафедри інженерії програмного  
Забезпечення в енергетиці

доктор технічних наук, професор

Члени комісії:

Професор кафедри інженерії програмного  
забезпечення в енергетиці

доктор технічних наук, професор

Професор кафедри інженерії програмного

Забезпечення в енергетиці

доктор технічних наук, доцент

Олександр КОВАЛЬ

Олег БАРАБАШ

Андрій МУСІЄНКО

## ДОДАТОК Б

### СПИСОК ПУБЛІКАЦІЙ ЗДОБУВАЧА

*Наукові праці, в яких опубліковані основні наукові результати дисертації:*

1. Олексій А.О., Верлань А.І. Огляд та порівняння методів машинного навчання для розпізнавання гідроакустичних сигналів. Науковий журнал «Інфокомунікаційні та комп'ютерні технології», № 1 (03). 2022. С. 113 – 120. <https://doi.org/10.36994 / 2788-5518-2022-01-03-18>.
2. Олексій А.О., Верлань А. А. Застосування багатошарового персептрона для аналізу акустичних сигналів у водному середовищі/ Науковий журнал «Зв'язок», Випуск №3 (163), 2023. с. 40-45.  
<https://doi.org/10.31673/2412-9070.2023.039000>.
3. Олексій А., Пуха Г. Створення датасету акустичних сигналів водного середовища для тренування нейромережі для придушення шумів. Науковий журнал «Інформаційні технології та суспільство», Випуск 2 (13). 2024. С.56-60. <https://doi.org/10.32689/maup.it.2024.2.8>.
4. Олексій, А. О. і Верлань А. А., 2024. Поліпшений метод аналізу акустичних сигналів водного середовища на основі згорткової нейромережі SOP. Вісник Вінницького політехнічного інституту, №6, с. 129-134. <https://doi.org/10.31649/1997-9266-2024-177-6-129-134>.

*Наукові праці, які засвідчують апробацію матеріалів дисертації:*

5. Олексій, А. О. і Верлань А. А., 2022. Огляд нейромережевих підходів для аналізу акустичних сигналів водного середовища. Теоретичні і прикладні проблеми фізики, математики та інформатики: зб. тез XX Всеукраїнської науково-практичної конференції студентів, аспірантів та молодих вчених, 15 червня 2022 р., м. Київ, с. 347-350.
6. Олексій, А. О., Верлань А.І. Нейромережеві підходи до генерації акустичних сигналів водного середовища., Сучасні проблеми наукового

забезпечення енергетики: зб. тез XX Міжнародної науково-практичної конференції молодих вчених і студентів, Т. 2, м. Київ, 25–28 квітня 2023 року с. 128-129.

7. Олексій, А. О., Верлань А.І. 2023. Аналіз акустичних сигналів водного середовища з застосуванням нейромережевих методів. Future of Work: Technological, Generational and Social Shifts: 2nd International Scientific and Practical Internet Conference., May 11-12, 2023, Dnipro, Ukraine, p. 67-68.
8. Олексій, А. О., Верлань А.І. Створення датасету акустичних сигналів водного середовища. Системи та технології зв'язку, інформатизації та кібербезпеки: актуальні питання і тенденції розвитку: матеріали III Міжнародної науково-технічної конференції, 30 листопада 2023 року, м. Київ, с. 229.
9. Олексій, А. О., Верлань А.І. Огляд нейромережевих підходів для придушення шумів водного середовища. Системи та засоби штучного інтелекту: тези доповідей Міжнародної наукової молодіжної школи., 10-11 жовтня 2023 р., м. Київ, с. 42-5.
10. Олексій, А. О., Варава І.А. 2023. Аналіз акустичних сигналів водного середовища з застосуванням нейромережевих моделей та створення датасетів акустичних сигналів водного середовища. Сучасні аспекти інженерії програмного забезпечення: збірник тез Першої міжнародної науково-практичної конференції, 14 грудня 2023 року, м. Київ, с. 100-103.
11. Олексій, А. О., Верлань А.І. Аналіз акустичних сигналів водного середовища з застосуванням нейромережевих моделей та створення датасетів акустичних сигналів водного середовища. Сучасні аспекти інженерії програмного забезпечення: збірник тез Другої міжнародної науково-практичної конференції, 13 листопада 2024 року, м. Київ, с. 60-64.

## ДОДАТОК В

### ЛІСТИНГИ КОДУ ПРОГРАМНИХ РЕАЛІЗАЦІЙ

Лістинг програмного коду модуля класифікації акстичних сигналів водного середовища

Predict.html

```
<!DOCTYPE html>
```

```
<html>
```

```
<head>
```

```
<meta charset="utf-8">
```

```
<meta name="viewport" content="width=device-width, initial-scale=1,  
shrink-to-fit=no">
```

```
<title>Select Dataset Directory</title>
```

```
<script>
```

```
function openFileManager() {
```

```
    document.getElementById('directory-picker').click();
```

```
}
```

```
function handleFiles() {
```

```
    var input = document.getElementById('directory-picker');
```

```
    if (input.files.length > 0) {
```

```
        var path = input.files[0].webkitRelativePath;
```

```
        var directory = path.substring(0, path.indexOf("/"));
```

```
        document.getElementById('data_dir').value = directory;
```

```

    }
}
</script>

<style>
.btn {
    display: inline-block;
    padding: 10px 20px;
    background-color: #3498db;
    color: white;
    text-decoration: none;
    border: none;
    border-radius: 5px;
    font-size: 16px;
    cursor: pointer;
    transition: background-color 0.3s;
}

.btn:hover {
    background-color: #2980b9;
}

</style>

</head>

<body>

<a class="btn" href="/">Home</a>

<div class="container">

<h1 class="mt-5">Select Dataset Directory</h1>

```

```

{% with messages = get_flashed_messages() %}

{% if messages %}

    <ul class="flashes">

        {% for message in messages %}

            <li>{{ message }}</li>

        {% endfor %}

    </ul>

{% endif %}

{% endwith %}

<form method="post" action="/predict" enctype="multipart/form-data">

    <div class="form-group">

        <label for="weights">Choose weights</label>

        <input type="file" class="form-control" id="weights" name="weights"
readonly required>

        <label for="data_dir">Choose dataset directory</label>

        <input type="text" class="form-control" id="data_dir" name="data_dir"
readonly required>

        <input type="file" id="directory-picker" webkitdirectory directory
style="display: none;" onchange="handleFiles()">

        <button class="btn" type="button" class="btn btn-secondary mt-2"
onclick="openFileManager()">Select Directory</button>

    </div>

    <button class="btn" type="submit" class="btn btn-
primary">Predict</button>

</form>

<div></div>

{% if text_answer %}

```



```

<ul>

    {% for key, value in text_answer.items() %}

        <li><strong>{{ key }}:</strong> {{ value[0] }}</li>

    {% endfor %}

</ul>

{% else %}

    <p>No data available.</p>

{% endif %}

</body>

</html>

```

PredictionComponent.py

```

class PredictionComponent(MethodView):

    def __init__(self):

        self.predictionService = PredictionService()

    def remove_last_folder(self, path):

        return "\\".join(path.split("\\")[:-1])

    def get(self):

        return render_template('predict.html', text_answer=None)

    def post(self):

```

```

current_file_path = os.path.abspath(__file__)

current_dir = os.path.dirname(current_file_path)

app_directory = os.path.dirname(current_dir)

input_dataset = request.form['data_dir']
weights = request.form['weights']

weights_path = app_directory + '\\weights\\' + weights
models_path = app_directory + '\\datasets\\'+ input_dataset

text_answer =
self.predictionService.make_prediction(weights_path,models_path)

return render_template('predict.html', text_answer=text_answer)

```

PredictionService.py

```

import numpy as np
from tqdm import tqdm
import os
import torch
import random
import librosa
from app.shared_module.models.classifier.CNN_SOP import CNN_SOP

```

```

from app.shared_module.services.preprocess_classifier import
PreprocessClassifier

import os

from flask import render_template, request


class PredictionService:

    def __init__(self, preprocessClassificationService: P
reprocessClassificationService):

        self.preprocessClassificationService = preprocessClassificationService

        self.device = torch.device('cuda') if torch.cuda.is_available() else
torch.device('cpu')

        self.model = CNN_SOP().to(self.device)


    def predicted_one(self, model, test_feature):

        model.eval()

        vector = test_feature.to(self.device)

        with torch.no_grad():

            preds = self.model(vector)

            _, predicted = torch.max(preds.data, 1)

            preds = predicted.detach().cpu().numpy()


        return preds


    def make_prediction(self, weights_path, models_path):

        samples = {}

        predictions = {}

```

```

model = CNN_SOP().to(self.device)
model.load_state_dict(torch.load(weights_path))

files = os.listdir(models_path)
for file in files:
    sample, _ = librosa.load(os.path.join(models_path, file), sr=4096)
    processed_sample =
self.preprocessClassifier.preprocess_classification_features(sample)
    predictions[file] = self.predicted_one(model, processed_sample)

return predictions

```

PreprocessClassifier.py

```
import torch
```

```
import librosa
```

```
import numpy as np
```

```
class PreprocessClassifier:
```

```
    def preprocess_classification_features(self, signal):
```

```
        cqt_abs = self.compute_cqt(signal)
```

```
        return torch.Tensor(np.array(cqt_abs[:64,
:64])).float().unsqueeze(0).unsqueeze(0)
```

```
    def compute_cqt(self, signal):
```

```
        sr=4096
```

```

fmin=4

fmax=1000

bins_per_octave=8

frame_length=23

num_frames=64

cqt = librosa.cqt(signal, hop_length=128, sr=sr, fmin=fmin,
                   n_bins=bins_per_octave * (int(np.log2(fmax / fmin)) + 1),
                   bins_per_octave=bins_per_octave
                   )

return np.abs(cqt)

```

CNN\_SOP.py

```

import torch

import torch.nn as nn

import torch.nn.functional as F

from ..classifier.MultiScaleConv import MultiScaleConv

from ..classifier.SOP import SOP

```

```

class CNN_SOP(nn.Module):

    def __init__(self, num_classes=5):

        super(CNN_SOP, self).__init__()

        self.multi_scale_conv1 = MultiScaleConv(1)

        self.multi_scale_conv2 = MultiScaleConv(24)

        self.sop = SOP()

```

```

self.bn = nn.BatchNorm1d(1024)

self.fc1 = nn.Linear(4096, 1024)
self.fc2 = nn.Linear(1024, num_classes)

def forward(self, x):
    x = self.multi_scale_conv1(x)
    x = F.avg_pool2d(x, kernel_size=2, stride=2)
    x = F.relu(x)
    x = self.multi_scale_conv2(x)
    x = F.avg_pool2d(x, kernel_size=2, stride=2)
    x = F.relu(x)
    x = self.sop(x)
    x = self.fc1(x)
    x = self.bn(x)
    x = F.relu(x)
    x = self.fc2(x)
    x = F.softmax(x, dim=1)
    return x

```

MultiScaleConv.py

```

import torch

import torch.nn as nn

import torch.nn.functional as F

```

```

class MultiScaleConv(nn.Module):

    def __init__(self, in_channels):

        super(MultiScaleConv, self).__init__()

        self.conv3x3 = nn.Conv2d(in_channels, 8, kernel_size=3, stride=1,
padding=1)

        self.conv5x5 = nn.Conv2d(in_channels, 8, kernel_size=5, stride=1,
padding=2)

        self.bn = nn.BatchNorm2d(16)


    def forward(self, x):

        x1 = F.relu(self.conv3x3(x))

        x2 = F.relu(self.conv5x5(x))

        x = torch.cat([x1, x2], dim=1)

        x = self.bn(x)

        return x

```

SOP.py

```

class SOP(nn.Module):

    def __init__(self):

        super(SOP, self).__init__()


    def forward(self, x):

        batch_size, channels, height, width = x.size()

        x = x.permute(0, 2, 1, 3).contiguous().view(-1, channels, width)

        sop = torch.bmm(x, x.permute(0, 2, 1))

        sop = sop.view(batch_size, height, -1)

        sop = sop.permute(0, 2, 1).contiguous().view(batch_size, -1)

```

```
sop = torch.sqrt(F.relu(sop)) - torch.sqrt(F.relu(-sop))  
sop = F.normalize(sop, p=2, dim=1)  
return sop
```