

Міністерство освіти і науки України  
Харківський національний університет радіоелектроніки

Факультет Інформаційно-аналітичних технологій та менеджменту  
(повна назва)

Кафедра Інформатики  
(повна назва)

## КВАЛІФІКАЦІЙНА РОБОТА Пояснювальна записка

рівень вищої освіти другий (магістерський)

**ДОСЛІДЖЕННЯ РОЗПІЗНАВАННЯ ШУМІВ ТРАНСПОРТНИХ  
ЗАСОБІВ З ВИКОРИСТАННЯМ ВЕЙВЛЕТ-ПЕРЕТВОРЕННЯ  
ТА ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ**  
(тема)

Виконав:  
здобувач 2 року навчання,  
групи ІНФМ-24-2  
Зубков А. О.  
(прізвище, ініціали)

Спеціальність 122 Комп'ютерні науки  
(код і повна назва спеціальності)

Тип програми освітньо-професійна

Освітня програма Інформатика  
(повна назва освітньої програми)

Науковий керівник проф. Шафроненко А. Ю.  
(посада, прізвище, ініціали)

Допускається до захисту

Завідувач кафедри інформатики \_\_\_\_\_  
(підпис)

Кобилін О. А.  
(прізвище, ініціали)

2025 р.

## Харківський національний університет радіоелектроніки

Факультет Інформаційно-аналітичних технологій та менеджментуКафедра ІнформатикиРівень вищої освіти другий (магістерський)Спеціальність 122 Комп'ютерні науки  
(код і повна назва)Тип програми освітньо-професійнаОсвітня програма Інформатика  
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри \_\_\_\_\_  
(підпис)

« \_\_\_\_\_ » \_\_\_\_\_ 2025 р.

**ЗАВДАННЯ**  
НА КВАЛІФІКАЦІЙНУ РОБОТУздобувачеві Зубкову Артему Олеговичу  
(прізвище, ім'я, по батькові)1. Тема роботи Дослідження розпізнавання шумів транспортних засобів з використанням вейвлет-перетворення та згорткових нейронних мереж

затверджена наказом університету від 14 листопада 2025 року № 1045Ст

2. Термін подання здобувачем роботи до екзаменаційної комісії 10 грудня 2025 р.

3. Вихідні дані до роботи Методи часово-частотного аналізу і вейвлет-перетворень, архітектури згорткових нейронних мереж, методи стиснення, літературні джерела щодо розпізнавання акустичних сигналів та класифікації транспортних шумів, програмні засоби для вейвлет-аналізу та глибокого навчання, набір аудіозаписів транспортних засобів різних типів для формування датасету.

4. Перелік питань, що потрібно опрацювати в роботі \_\_\_\_\_

1. Аналіз природи та характеристик транспортних акустичних шумів, огляд традиційних методів аналізу акустичних сигналів.

2. Дослідження теоретичних основ вейвлет-перетворення та аналіз властивостей різних типів вейвлетів для застосування до транспортних шумів.

3. Розробка та дослідження методів стиснення спектрограм з метою зменшення обчислювальної складності.

4. Формування набору даних акустичних записів дев'яти класів транспортних засобів.

5. Розробка архітектур згорткових нейронних мереж, адаптованих для класифікації часово-частотних представлень транспортних шумів, включаючи базову архітектуру та спеціалізовані архітектури для різних методів стиснення.

6. Проведення комплексних експериментальних досліджень для різних комбінацій типів вейвлетів та методів стиснення з оцінкою точності класифікації та швидкості операцій вейвлет-перетворень, нормалізації та стиснень.

7. Порівняльний аналіз отриманих результатів, виявлення оптимальних комбінацій типу вейвлет-перетворень, методів стиснення та загорткових нейронних мереж.

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (п.5 включається до завдання за рішенням випускової кафедри) актуальність проблеми розпізнавання транспортних акустичних шумів, об'єкт та мета дослідження, математичні вирази вейвлет-перетворення для різних типів вейвлетів, приклади візуалізованих спектрограм для різних класів транспортних засобів до та після стиснення, архітектура розробленої згорткової нейронної мережі з відображенням шарів та їх параметрів, матриці помилок (confusion matrices) різних комбінацій вейвлетів та методів стиснення, таблиці з результатами обчислювальної ефективності вейвлет-перетворень та методів стиснення, висновки та практичні рекомендації, перспективи подальших досліджень.

6. Консультанти розділів роботи (п.6 включається до завдання за наявності консультантів згідно з наказом, зазначеним у п.1)

Найменування розділу	Консультант (посада, прізвище, ім'я, по батькові)	Позначка консультанта про виконання розділу	
		підпис	дата

### КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Строк / терміни виконання етапів роботи	Примітка
1	Отримання завдання на кваліфікаційну роботу	29.09.2025	
2	Аналіз завдання, підбір літератури	30.09.25-07.10.25	
3	Аналіз літератури з досліджуваної проблеми	08.10.25-14.10.25	
4	Особливості вейвлетів, стиснень, CNN архітектур	15.10.25-20.10.25	
5	Дослідження вейвлетів, стиснень, CNN архітектур	21.10.25-27.10.25	
6	Програмна реалізація	28.10.25-05.11.25	
7	Обґрунтування отриманих результатів	06.11.25-11.11.25	
8	Оформлення пояснювальної записки	12.11.25-14.11.25	
9	Перевірка на нормоконтроль	19.11.25-10.12.25	
10	Перевірка на плагіат	20.11.25-10.12.25	
11	Рецензування	21.11.25-10.12.25	
12	Підготовка презентації та доповіді	21.11.25-22.12.25	
13	Занесення роботи в електронний архів	21.11.25-22.12.25	
14	Попередній захист кваліфікаційної роботи	01.12.25-22.12.25	

Дата видачі завдання 29 вересня 2025 р.

Здобувач \_\_\_\_\_  
(підпис)

Керівник роботи \_\_\_\_\_  
(підпис)

проф. Шафроненко А. Ю.  
(посада, прізвище, ініціали)

## РЕФЕРАТ

Пояснювальна записка до кваліфікаційної роботи: 120 с., 2 табл., 8 рис., 2 дод., 24 джерела.

АКУСТИЧНІ ШУМИ, ВЕЙВЛЕТ-ПЕРЕТВОРЕННЯ, ГЛИБОКЕ НАВЧАННЯ, РОЗПІЗНАВАННЯ ЗВУКІВ, ТРАНСПОРТНІ ЗАСОБИ, ЧАСОВО-ЧАСТОТНИЙ АНАЛІЗ, AUDIO CLASSIFICATION, CNN MODEL, DATA COMPRESSION, FEATURE EXTRACTION, SIGNAL PROCESSING.

Об'єктом дослідження є процес розпізнавання акустичних шумів транспортних засобів у міському середовищі з використанням часово-частотного аналізу та глибокого навчання.

Метою дослідження є розробка ефективних методів розпізнавання шумів транспортних засобів на основі вейвлет-перетворення та згорткових нейронних мереж.

Використано методи вейвлет-аналізу сигналів, методи стиснення даних, згорткові нейронні мережі та статистичний аналіз.

Наукова новизна полягає у виявленні впливу типів вейвлетів, методів стиснення та структур загорткових нейронних мереж на точність розпізнавання шумів та визначенні оптимальних параметрів обробки.

Взаємозв'язок з іншими роботами. Дослідження базується на сучасних досягненнях цифрової обробки сигналів і глибокого навчання, розвиваючи часово-частотні методи акустичної класифікації.

Рекомендації щодо використання. Результати можуть застосовуватись у системах моніторингу міського шуму, інтелектуальних, транспортних і системах безпеки.

У результаті встановлено, що вейвлет FBSP з методом стиснення AVG1 забезпечує найвищий баланс між точністю та швидкістю розпізнавання, далі Resize та AVG, тоді як Statistics найменш ефективний через втрату інформації.

## ABSTRACT

Explanatory note to the qualification work: 120 pages, 2 tables, 8 figures, 2 appendixes, 24 sources.

ACOUSTIC NOISE, WAVELET TRANSFORM, DEEP LEARNING, SOUND RECOGNITION, VEHICLES, TIME-FREQUENCY ANALYSIS, AUDIO CLASSIFICATION, CNN MODEL, DATA COMPRESSION, FEATURE EXTRACTION, SIGNAL PROCESSING.

The object of the research is the process of recognizing acoustic noise from vehicles in urban environments using time-frequency analysis and deep learning.

The aim of the research is to develop effective methods for recognizing vehicle noise based on wavelet transforms and convolutional neural networks.

Methods of wavelet signal analysis, data compression techniques, convolutional neural networks, and statistical analysis were used.

Scientific novelty lies in identifying the influence of wavelet types, compression methods, and CNN architectures on noise recognition accuracy and determining the optimal processing parameters.

Interconnection with other works. The research builds on modern advances in digital signal processing and deep learning, extending time-frequency methods for acoustic classification.

Recommendations for using the results of the work. The results can be applied in urban noise monitoring systems, intelligent, transport and safety systems.

As a result of the research, it was found that the FBSP wavelet with the AVG1 compression method provides the highest balance between accuracy and recognition speed. Resize and AVG are little worse, while Statistics is the least effective due to information loss.

## ЗМІСТ

Перелік умовних позначень, символів, одиниць, скорочень і термінів .....	9
Вступ.....	11
1 Огляд основних методів розпізнавання транспортних акустичних шумів. 14	14
1.1 Природа та характеристики транспортних акустичних шумів .....	14
1.2 Традиційні методи аналізу акустичних сигналів.....	16
1.3 Вейвлет-перетворення в аналізі акустичних сигналів.....	17
1.4 Згорткові нейронні мережі для розпізнавання образів.....	19
1.5 Інтеграція вейвлет-аналізу та глибокого навчання .....	22
1.6 Огляд існуючих досліджень у галузі розпізнавання транспортних шумів .....	24
1.7 Постановка задачі дослідження.....	26
2 Методологія дослідження розпізнавання транспортних шумів .....	28
2.1 Типи вейвлет-перетворень та їх математичний опис .....	28
2.1.1 Вейвлет Морле (Morlet).....	29
2.1.2 Вейвлет Mexican Hat (Mexh) .....	30
2.1.3 Комплексний вейвлет Морле (CMOR) .....	31
2.1.4 Вейвлет Frequency B-Spline (FBSP) .....	32
2.1.5 Комплексний гаусівський вейвлет (CGAU) .....	33
2.1.6 Вейвлет Шеннона (SHAN) .....	34
2.1.7 Порівняльний аналіз властивостей вейвлетів .....	35
2.2 Методи стиснення часово-частотних представлень.....	38
2.2.1 Метод зміни розміру з інтерполяцією (Resize) .....	39
2.2.2 Метод статистичної агрегації (Statistics) .....	40
2.2.3 Метод обрізання (Crop) .....	43
2.2.4 Комбінований метод Crop + Max Pooling.....	44
2.2.5 Метод усереднення блоків (Average Blocks, AVG) .....	45

2.2.6 Модифікований метод усереднення блоків (AVG1).....	46
2.2.7 Метод сингулярного розкладу (SVD).....	47
3 Створення датасету, згорткових нейронних мереж та тестування.....	49
3.1 Формування та характеристика експериментального набору даних.	49
3.1.1 Вибір класів транспортних засобів та обґрунтування структури датасету .....	49
3.1.2 Технічні параметри запису та обробки аудіоданих .....	53
3.1.3 Формування фінального датасету .....	55
3.1.4 Попередня обробка аудіосигналів перед вейвлет-перетворенням .....	56
3.2 Розроблені архітектури згорткових нейронних мереж .....	58
3.2.1 Базова архітектура нейронних мереж для більшості методів .	58
3.2.2 Архітектура для методу Statistics.....	63
3.2.3 Архітектура для методу Crop+MaxPooling .....	65
3.2.4 Використання архітектури EfficientNetB0 .....	66
3.3 Результати експериментальних досліджень .....	68
3.3.1 Методологія проведення експериментів та метрики оцінювання.....	69
3.3.2 Результати для вейвлета Morlet з різними методами стиснення .....	71
3.3.3 Результати для вейвлета Mexican Hat з методом AVG.....	75
3.3.4 Результати для вейвлетів CMOR та FBSP .....	77
3.3.5 Результати для вейвлета CGAU з різними методами .....	79
3.3.6 Результати для вейвлета Shannon .....	81
3.3.7 Результати з використанням архітектури EfficientNetB0 .....	83
3.3.8 Аналіз обчислювальної ефективності вейвлет-перетворень та методів стиснення .....	84
3.3.9 Порівняльний аналіз та узагальнення результатів.....	87

Висновки.....	93
Перелік джерел посилання .....	96
Додаток А Програмний код для досліджень .....	99
Додаток Б Вигляд спектрограм отриманих в результаті вейвлет-перетворень та стиснень .....	117

## ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

- Adam – Adaptive Moment Estimation (адаптивна оцінка моментів)
- AdaGrad – Adaptive Gradient Algorithm (алгоритм адаптивного градієнта)
- AVG – Average Blocks (метод усереднення блоків)
- Batch Normalization – батч-нормалізація
- B-spline – B-сплайн
- CGAU – Complex Gaussian wavelet (комплексний гаусівський вейвлет)
- CMOR – Complex Morlet wavelet (комплексний вейвлет Морле)
- CNN – Convolutional Neural Network (згорткова нейронна мережа)
- CPU – Central Processing Unit (центральний процесор)
- Crop – метод обрізання
- Cross-Entropy – крос-ентропія
- CWT – Continuous Wavelet Transform (неперервне вейвлет-перетворення)
- Data Augmentation – аугментація даних
- Dropout – метод регуляризації випадковим відключенням нейронів
- EarlyStopping – рання зупинка навчання
- FBSP – Frequency B-Spline wavelet (вейвлет частотного B-сплайна)
- Flatten – операція розгортання тензора
- GPU – Graphics Processing Unit (графічний процесор)
- Inference – передбачення, виведення
- Leaky ReLU – модифікована функція активації
- LSTM – Long Short-Term Memory (довга короткострокова пам'ять)
- Max Pooling – операція максимального об'єднання
- Mexh – Mexican Hat wavelet (вейвлет мексиканський капелюх)
- MFCC – Mel-Frequency Cepstral Coefficients (мел-частотні кепстральні коефіцієнти)
- NumPy – бібліотека для роботи з масивами в Python
- Padding – доповнення нулями

- ReduceLROnPlateau – зменшення швидкості навчання при плато
- ReLU – Rectified Linear Unit (випрямлена лінійна одиниця)
- ResNet – Residual Network (залишкова мережа)
- Resize – метод зміни розміру з інтерполяцією
- RGB – Red Green Blue (червоний зелений синій)
- RMS – Root Mean Square (середньоквадратичне значення)
- RMSprop – Root Mean Square Propagation (алгоритм оптимізації)
- SELU – Scaled Exponential Linear Unit (масштабована експоненціальна лінійна одиниця)
- SHAN – Shannon wavelet (вейвлет Шеннона)
- Softmax – функція активації для багатокласової класифікації
- Squeeze-and-Excitation – механізм уваги в нейронних мережах
- Statistics – метод статистичної агрегації
- SVD – Singular Value Decomposition (сингулярний розклад)
- TensorFlow – фреймворк для машинного навчання

## ВСТУП

Актуальність роботи полягає у зростаючій потребі автоматизації процесів моніторингу та управління міським транспортним середовищем. Сучасні мегаполіси характеризуються постійним збільшенням транспортних потоків, що призводить до підвищення рівня шумового навантаження та створює нові виклики для систем безпеки дорожнього руху. Традиційні методи контролю та аналізу транспортних потоків, що базуються переважно на візуальних спостереженнях, не завжди є достатньо ефективними в умовах обмеженої видимості, складних погодних умов або при великій щільності руху.

Розпізнавання транспортних засобів за акустичними характеристиками відкриває нові можливості для створення інтелектуальних транспортних систем. Акустичні сигнали, на відміну від візуальних даних, не залежать від освітлення, можуть реєструватися в умовах обмеженої видимості та дозволяють здійснювати моніторинг у режимі реального часу без порушення приватності громадян. Це робить акустичні методи особливо привабливими для широкого спектру застосувань, включаючи контроль дорожнього руху, екологічний моніторинг, системи раннього попередження про надзвичайні ситуації та оптимізацію міської інфраструктури.

Однак розпізнавання транспортних шумів є складною задачею через високу варіативність акустичних характеристик різних транспортних засобів, вплив фонових шумів міського середовища, ефекти відбиття та реверберації, а також перекриття сигналів від декількох джерел. Традиційні методи спектрального аналізу, такі як перетворення Фур'є, мають обмеження при обробці нестационарних сигналів, оскільки не дозволяють одночасно аналізувати часові та частотні характеристики. Це призводить до втрати важливої інформації про динаміку зміни акустичних параметрів у часі.

Вейвлет-перетворення є потужним інструментом часово-частотного аналізу, який дозволяє подолати обмеження класичних методів [1]. На відміну від перетворення Фур'є, вейвлет-аналіз забезпечує можливість локалізації

особливостей сигналу як у часовій, так і в частотній області. Це особливо важливо для аналізу транспортних шумів, які характеризуються швидкими змінами спектральних характеристик, пов'язаними з режимами роботи двигунів, швидкістю руху, прискоренням та гальмуванням.

Поєднання вейвлет-перетворення з сучасними методами глибокого навчання, зокрема згортковими нейронними мережами, створює нові можливості для автоматичного розпізнавання та класифікації транспортних шумів. Згорткові нейронні мережі (CNN) довели свою ефективність у задачах розпізнавання образів та обробки складноструктурованих даних. Їх застосування до часово-частотних представлень акустичних сигналів дозволяє автоматично виявляти характерні ознаки різних типів транспортних засобів без необхідності ручного конструювання ознак.

Проте існує низка технічних викликів, пов'язаних з інтеграцією вейвлет-аналізу та глибокого навчання. Результатом вейвлет-перетворення акустичного сигналу є матриця значних розмірів, де одна вісь відповідає частотному спектру, а інша – часовій розгортці. Для типового аудіосигналу тривалістю одну секунду з частотою дискретизації 44100 Гц при використанні 128 частотних компонент отримуємо матрицю розміром  $128 \times 16384$ . Така висока розмірність даних створює серйозні обчислювальні труднощі при навчанні нейронних мереж: збільшується кількість параметрів моделі, зростають вимоги до обчислювальних ресурсів та пам'яті, виникає ризик перенавчання та суттєво подовжується час навчання.

Тому критично важливим є вибір адекватних методів стиснення та попередньої обробки даних, які дозволяють зменшити розмірність вхідних даних зі збереженням найбільш інформативних характеристик сигналу. До таких методів належать операції обрізання (Crop), зміни розміру з інтерполяцією (Resize), статистичної агрегації (Statistics) та підвибірки (Pooling). Кожен з цих методів має свої переваги та недоліки, і їх ефективність може значно варіюватися залежно від типу акустичних сигналів та вимог до точності розпізнавання.

Крім того, вибір типу вейвлет-функції також суттєво впливає на результати розпізнавання. Різні вейвлети (Морле, FBSP, CMOR та інші) мають різні властивості щодо часової та частотної локалізації, що визначає їх здатність виявляти певні характеристики сигналів. Комплексні вейвлети дозволяють отримувати як амплітудну, так і фазову інформацію, що може бути критично важливим для розрізнення близьких за спектральним складом джерел шуму.

Огляд сучасного стану показує, що питання оптимального поєднання типів вейвлет-перетворень, методів стиснення даних та архітектури нейронних мереж для задачі розпізнавання транспортних шумів залишається недостатньо дослідженим. Більшість існуючих робіт зосереджуються на окремих аспектах цієї проблеми, не надаючи комплексного порівняльного аналізу різних підходів. Відсутність систематичних досліджень впливу параметрів попередньої обробки на якість розпізнавання ускладнює вибір оптимальних рішень при проектуванні практичних систем.

Наукова задача, що виявляється з аналізу сучасного стану, полягає у необхідності комплексного дослідження ефективності різних комбінацій вейвлет-перетворень та методів стиснення даних для задачі розпізнавання транспортних акустичних шумів за допомогою згорткових нейронних мереж. Необхідно встановити, які типи вейвлетів найкращим чином виявляють характерні особливості різних транспортних засобів, які методи стиснення забезпечують оптимальний баланс між зменшенням обчислювальної складності та збереженням інформативності даних, а також як ці фактори взаємодіють між собою та впливають на кінцеву точність розпізнавання.

Вирішення цієї наукової задачі дозволить сформувати науково обґрунтовані рекомендації щодо проектування ефективних систем розпізнавання транспортних шумів, що матиме практичне значення для розробки інтелектуальних транспортних систем, систем екологічного моніторингу та забезпечення безпеки дорожнього руху в сучасних містах.

# 1 ОГЛЯД ОСНОВНИХ МЕТОДІВ РОЗПІЗНАВАННЯ ТРАНСПОРТНИХ АКУСТИЧНИХ ШУМІВ

## 1.1 Природа та характеристики транспортних акустичних шумів

Акустичні шуми транспортних засобів є складними нестационарними сигналами, що формуються внаслідок роботи численних механічних компонентів, взаємодії з навколишнім середовищем та аеродинамічних ефектів [2]. Розуміння фізичної природи цих шумів є необхідною передумовою для розробки ефективних методів їх розпізнавання [3].

Основними джерелами акустичних шумів транспортних засобів є двигуни внутрішнього згоряння або електродвигуни, трансмісія, система випуску відпрацьованих газів, взаємодія шин з дорожнім покриттям, аеродинамічний шум від обтікання повітряними потоками корпусу транспортного засобу. Кожне з цих джерел має специфічні спектральні характеристики та часову динаміку.

Шум двигуна внутрішнього згоряння характеризується виразною гармонічною структурою, пов'язаною з частотою обертання колінчастого валу та кількістю циліндрів. Основна частота визначається кількістю робочих тактів за одиницю часу, а в спектрі присутні численні гармоніки. При зміні режиму роботи двигуна частотний спектр динамічно змінюється, що відображає прискорення або уповільнення транспортного засобу.

Електричні транспортні засоби генерують принципово інший акустичний профіль, що характеризується відсутністю низькочастотних компонент від згоряння палива та присутністю високочастотного шуму від електродвигунів та інверторів. Це створює додаткові виклики для універсальних систем розпізнавання.

Шум від контакту шин з дорожнім покриттям залежить від швидкості руху, типу шин, стану дорожнього покриття та має більш широкосмуговий характер. Його інтенсивність зростає зі збільшенням швидкості руху і може

домінувати над іншими джерелами на високих швидкостях. Спектральні характеристики цього типу шуму містять інформацію не тільки про транспортний засіб, але й про умови руху.

Аеродинамічний шум стає суттєвим при швидкостях понад 60-80 км/год. і характеризується турбулентним характером з широким частотним спектром. Він залежить від геометрії транспортного засобу, наявності відкритих вікон та інших факторів, що впливають на обтікання повітряними потоками.

Різні типи транспортних засобів мають характерні акустичні профілі. Мотоцикли характеризуються високим рівнем шуму двигуна з яскраво вираженими гармоніками в середньо та високочастотному діапазоні. Легкові автомобілі мають більш збалансований спектр з акцентом на низько та середньочастотні компоненти. Вантажні автомобілі та автобуси генерують потужні низькочастотні компоненти через великі об'єми двигунів та масу транспортного засобу. Залізничний транспорт характеризується ритмічними ударами від контакту коліс з рейками та специфічним шумом від електричних тягових двигунів.

Важливою характеристикою транспортних шумів є їх нестационарність. На відміну від стаціонарних джерел шуму, параметри транспортних шумів постійно змінюються через зміну режимів роботи двигуна, швидкості руху, відстані до джерела та умов поширення звуку. Це вимагає використання методів аналізу, здатних відслідковувати часову еволюцію спектральних характеристик.

Додаткову складність створює міське акустичне середовище, яке характеризується високим рівнем фонового шуму, ефектами відбиття та реверберації від будівель та інших поверхонь, ефектом Доплера при русі джерел звуку відносно приймача. У реальних умовах акустичні сигнали від різних транспортних засобів можуть перекриватися, створюючи складні суперпозиції, що ускладнює задачу розпізнавання окремих джерел.

## 1.2 Традиційні методи аналізу акустичних сигналів

Класичні підходи до аналізу акустичних сигналів базуються переважно на спектральних методах, які дозволяють виявляти частотні характеристики сигналів та використовувати їх для класифікації.

Перетворення Фур'є є фундаментальним інструментом спектрального аналізу, що дозволяє представити сигнал у частотній області. Дискретне перетворення Фур'є розкладає сигнал на суму синусоїдальних компонент різних частот, що дозволяє виявити домінуючі частотні складові. Для практичного застосування використовується швидке перетворення Фур'є, яке забезпечує ефективне обчислення з алгоритмічною складністю  $O(n \log n)$ .

Однак перетворення Фур'є має суттєве обмеження: воно надає інформацію про частотний склад сигналу в цілому, але не дозволяє визначити, як частотні характеристики змінюються в часі. Для нестационарних сигналів, якими є транспортні шуми, ця інформація є критично важливою, оскільки динаміка зміни спектру містить важливі дискримінаційні ознаки.

Для подолання цього обмеження було розроблено короткочасне перетворення Фур'є, яке розбиває сигнал на короткі часові фрагменти та застосовує перетворення Фур'є до кожного фрагменту окремо. Це дозволяє отримати часово-частотне представлення сигналу у вигляді спектрограми. Однак короткочасне перетворення Фур'є має фундаментальне обмеження, пов'язане з принципом невизначеності: підвищення часової роздільної здатності призводить до погіршення частотної роздільності і навпаки. Розмір вікна аналізу є компромісом між цими двома характеристиками.

Mel-частотні кепстральні коефіцієнти є класичними ознаками, що широко використовуються для розпізнавання мовлення та інших акустичних сигналів. MFCC базуються на нелінійній шкалі частот, що відповідає особливостям людського сприйняття звуку. Обчислення MFCC включає застосування перетворення Фур'є, перехід до мел-шкали, логарифмування енергії в мел-фільтрах та обернене дискретне косинусне перетворення. MFCC компактно

представляють спектральну огинаючу сигналу і є ефективними для багатьох задач класифікації звуків.

Методи на основі лінійного передбачувального кодування моделюють мовний тракт як лінійний фільтр та визначають його параметри, що також можуть використовуватися як ознаки для класифікації. Кепстральний аналіз дозволяє розділити вклад джерела збудження та резонансних характеристик акустичного тракту.

Для класифікації на основі отриманих ознак традиційно використовувалися методи машинного навчання, такі як метод  $k$ -найближчих сусідів, метод опорних векторів, дерева рішень та їх ансамблі, наївний баєсівський класифікатор, гаусівські моделі сумішей. Ці методи вимагають ретельного проектування ознак та їх попередньої обробки, а також застосування інструментів швидкого пошуку метричних даних, характерних для структурних методів класифікації [4,5].

Однак традиційні методи мають ряд недоліків при застосуванні до транспортних шумів. По-перше, вони вимагають ручного конструювання ознак, що потребує глибокої експертної підготовки та може не враховувати всіх релевантних характеристик сигналів. По-друге, фіксована часово-частотна роздільність не дозволяє оптимально аналізувати сигнали з різними часовими масштабами. По-третє, чутливість до шумів та спотворень обмежує надійність роботи в реальних умовах міського середовища.

### 1.3 Вейвлет-перетворення в аналізі акустичних сигналів

Вейвлет-перетворення є потужним інструментом часово-частотного аналізу, який дозволяє подолати багато обмежень традиційних методів. Основна ідея вейвлет-аналізу полягає у розкладанні сигналу за системою базисних функцій, що отримуються шляхом зсуву та масштабування материнської вейвлет-функції.

Неперервне вейвлет-перетворення визначається як згортка сигналу з масштабованою та зсувнутою версією материнського вейвлета. Параметр масштабу відповідає частотній характеристиці, а параметр зсуву визначає часове положення. Це забезпечує адаптивну часово-частотну роздільність: високу частотну роздільність на низьких частотах та високу часову роздільність на високих частотах.

Така адаптивна роздільність є природною для багатьох сигналів, включаючи транспортні шуми, оскільки низькочастотні компоненти зазвичай є більш тривалими і потребують кращої частотної локалізації, тоді як високочастотні компоненти часто є короткочасними і потребують кращої часової локалізації.

Існує широкий спектр вейвлет-функцій, кожна з яких має специфічні властивості та призначення. Вейвлет Морле є комплексною функцією, що являє собою комплексну експоненту, модульовану гаусівською огинаючою. Цей вейвлет забезпечує хороший баланс між часовою та частотною локалізацією і є особливо ефективним для виявлення гармонічних компонент сигналів. Його застосування до транспортних шумів дозволяє чітко виділити періодичні структури, пов'язані з роботою двигунів.

Вейвлет FBSP комбінує гнучкість B-сплайнів з частотною локалізацією. Він характеризується гладкістю в часовій області та компактністю в частотній області, що робить його ефективним для аналізу сигналів з плавними змінами спектральних характеристик.

Комплексний вейвлет Морле надає як амплітудну, так і фазову інформацію, що є критично важливим для повного опису коливальних процесів. Фазова інформація може містити додаткові дискримінаційні ознаки, що не відображаються в амплітудному спектрі.

Вейвлет Хаара є найпростішим вейвлетом і ефективний для виявлення різких змін у сигналі. Вейвлети Добеші характеризуються компактним носієм та певною кількістю нульових моментів, що робить їх ефективними для

стиснення даних. Мексиканський капелюх використовується для виявлення локальних екстремумів та особливостей сигналів.

Вибір материнського вейвлета суттєво впливає на результати аналізу. Критеріями вибору є подібність форми вейвлета до характерних елементів сигналу, необхідний баланс між часовою та частотною роздільністю, потреба в фазовій інформації, обчислювальна ефективність. Для транспортних шумів, що характеризуються як гармонічними, так і імпульсними компонентами, комплексні вейвлети типу Морле зазвичай демонструють кращі результати.

Результатом вейвлет-перетворення є часово-частотна матриця, де рядки відповідають різним масштабам (частотам), а стовпці – часовим позиціям. Ця матриця, яку часто називають скалограмою, наочно показує, як енергія сигналу розподіляється в часово-частотній площині.

Переваги вейвлет-перетворення для аналізу транспортних шумів включають адаптивну роздільність, що відповідає природі сигналів, можливість виявлення короткочасних подій та перехідних процесів, стійкість до деяких типів шумів та спотворень, відсутність обмежень на стаціонарність сигналу. Вейвлет-аналіз особливо ефективний для сигналів з різними часовими масштабами, що характерно для транспортних шумів.

#### 1.4 Згорткові нейронні мережі для розпізнавання образів

Згорткові нейронні мережі представляють клас глибоких нейронних мереж, спеціально розроблених для обробки даних з регулярною структурою, таких як зображення, відео або часово-частотні представлення сигналів. CNN здатні автоматично виявляти ієрархічні ознаки з мінімальною попередньою обробкою даних, що було успішно продемонстровано не лише в аудіо, але й у задачах розпізнавання емоцій та класифікації кулінарних страв на зображеннях [6].

Основна ідея згорткових мереж полягає у використанні операції згортки замість повнозв'язного матричного множення в деяких шарах мережі. Згорткова операція застосовує набір навчених фільтрів до вхідних даних, виявляючи локальні паттерни. Кожен фільтр сканує вхідні дані та обчислює скалярний добуток зі своїми ваговими коефіцієнтами в кожній позиції.

Ключовими компонентами згорткової нейронної мережі є згорткові шари, які виконують виявлення локальних ознак через застосування фільтрів, функції активації, що вносять нелінійність і дозволяють мережі моделювати складні залежності, шари підвибірки, які зменшують просторову розмірність даних та забезпечують інваріантність до невеликих зсувів, повнозв'язні шари, які виконують класифікацію на основі виявлених ознак.

Згортковий шар застосовує множину фільтрів до вхідних даних. Кожен фільтр характеризується певним розміром рецептивного поля та множиною вагових коефіцієнтів, що визначають, які паттерни він виявляє. У результаті застосування фільтра отримується карта ознак, що показує наявність відповідного паттерну в різних частинах вхідних даних.

Важливою властивістю згорткових шарів є спільне використання ваг: один і той же фільтр застосовується до всіх позицій вхідних даних. Це значно зменшує кількість параметрів моделі порівняно з повнозв'язними мережами та забезпечує інваріантність до зсувів виявлених паттернів.

Функції активації вносять нелінійність у модель. Найпоширенішою є функція ReLU, яка обнуляє від'ємні значення та залишає без змін додатні. ReLU прискорює навчання та допомагає уникнути проблеми зникаючих градієнтів. Альтернативними є функції Leaky ReLU, ELU, SELU та інші.

Шари підвибірки зменшують просторову розмірність карт ознак, зберігаючи найважливішу інформацію. Max pooling вибирає максимальне значення в кожному вікні, що забезпечує інваріантність до невеликих локальних деформацій. Average pooling обчислює середнє значення, що дає більш згладжене представлення.

Повнозв'язні шари в кінці мережі агрегують інформацію з усіх карт ознак і виконують остаточну класифікацію. Кожен нейрон цих шарів з'єднаний зі всіма нейронами попереднього шару, що дозволяє комбінувати виявлені локальні ознаки для прийняття глобального рішення про клас вхідних даних.

Архітектура типової згорткової нейронної мережі зазвичай складається з послідовності блоків, кожен з яких містить згортковий шар, функцію активації та шар підвибірки. Глибина мережі, тобто кількість таких блоків, визначає складність ознак, які мережа здатна виявити. Перші шари виявляють прості ознаки, такі як краї та текстури, а глибші шари комбінують їх у більш складні паттерни.

Для навчання згорткових нейронних мереж використовується метод зворотного поширення помилки в поєднанні з градієнтним спуском або його модифікаціями. Функція втрат вимірює розбіжність між передбаченнями мережі та справжніми мітками класів. Для задач класифікації зазвичай використовується крос-ентропійна функція втрат.

Однією з ключових проблем навчання глибоких мереж є перенавчання, коли модель добре працює на навчальних даних, але погано узагальнює на нові дані. Для боротьби з цим використовуються методи регуляризації, зокрема Dropout, що випадково відключає частину нейронів під час навчання, L2-регуляризація, яка штрафує великі значення ваг, Data Augmentation, що штучно збільшує розмір навчальної вибірки, Batch Normalization, що нормалізує активації між шарами.

Batch Normalization є особливо важливою технікою, яка нормалізує вхідні дані кожного шару, забезпечуючи середнє значення близьке до нуля та стандартне відхилення близьке до одиниці. Це прискорює навчання, дозволяє використовувати більші швидкості навчання та зменшує чутливість до ініціалізації ваг.

Для оптимізації процесу навчання розроблено численні алгоритми адаптивного градієнтного спуску. Adam (Adaptive Moment Estimation) комбінує ідеї моментуму та адаптивної швидкості навчання для різних параметрів.

RMSprop адаптує швидкість навчання на основі ковзного середнього квадратів градієнтів. AdaGrad збільшує швидкість навчання для параметрів з рідкісними оновленнями.

Застосування згорткових нейронних мереж до часово-частотних представлень акустичних сигналів є природним, оскільки такі представлення мають структуру, подібну до зображень. Скалограми та спектрограми можна розглядати як двовимірні зображення, де один вимір відповідає частоті, інший – часу, а інтенсивність пікселя кодує амплітуду або енергію сигналу.

Згорткові фільтри здатні автоматично виявляти характерні часово-частотні паттерни, які відрізняють різні класи транспортних засобів. Наприклад, гармонічна структура шуму двигуна може бути виявлена фільтрами, налаштованими на вертикальні смуги в часово-частотному представленні, тоді як перехідні процеси виявляються фільтрами, чутливими до локалізованих у часі енергетичних піків.

### 1.5 Інтеграція вейвлет-аналізу та глибокого навчання

Поєднання вейвлет-перетворення та згорткових нейронних мереж створює синергетичний ефект, де переваги обох підходів доповнюють один одного. Вейвлет-аналіз забезпечує ефективне представлення нестационарних сигналів у часово-частотній області, а CNN автоматично виявляють складні дискримінаційні ознаки в цих представленнях.

Такий гібридний підхід дозволяє уникнути необхідності ручного конструювання ознак, що є трудомістким процесом та вимагає глибоких експертних знань. Натомість, мережа сама навчається виявляти найбільш релевантні характеристики сигналів безпосередньо з часово-частотних представлень.

Існує декілька підходів до інтеграції вейвлет-перетворення в архітектуру нейронних мереж. Найпростішим є використання вейвлет-перетворення як

етапу попередньої обробки, де акустичний сигнал спочатку перетворюється на скалограму, яка потім подається на вхід звичайної згорткової мережі. Цей підхід є найбільш поширеним завдяки своїй простоті та ефективності.

Альтернативним підходом є інтеграція вейвлет-перетворення безпосередньо в архітектуру мережі у вигляді вейвлет-шарів. У такому підході вагові коефіцієнти згорткових фільтрів ініціалізуються відповідно до форми вейвлет-функцій і можуть або залишатися фіксованими, або налаштовуватися в процесі навчання. Це дозволяє мережі адаптувати параметри вейвлет-перетворення до специфіки задачі.

Мультирозділова архітектура використовує паралельні гілки обробки з різними типами вейвлетів або різними параметрами перетворення. Результати з різних гілок потім об'єднуються для прийняття рішення. Такий підхід дозволяє використовувати комплементарну інформацію з різних представлень сигналу.

Однак інтеграція вейвлет-аналізу з глибоким навчанням стикається з рядом викликів. Основною проблемою є висока розмірність результатів вейвлет-перетворення. Для типового аудіофрагменту тривалістю одну секунду з частотою дискретизації 44100 Гц скалограма може мати розмір  $128 \times 16384$  або більше. Така велика кількість вхідних даних призводить до експоненційного зростання кількості параметрів мережі, що створює проблеми з обчислювальною ефективністю та ризиком перенавчання.

Для вирішення цієї проблеми застосовуються різноманітні методи стиснення та зменшення розмірності часово-частотних представлень чи стиснення структурного опису [7]. Вибір адекватного методу стиснення є критично важливим, оскільки він повинен зберігати найбільш інформативні характеристики сигналу, відкидаючи надлишкові або незначущі дані.

Метод обрізання полягає у виділенні найбільш інформативної частини скалограми (спектрограми) – центральної області або певного частотного діапазону, характерного для цільових класів. Обрізання по часовій осі зменшує кількість відліків, зберігаючи повний спектр, а по частотній – дозволяє зосередитись на потрібних смугах.

Метод зміни розміру з інтерполяцією (Resize) використовує масштабування зображень для скалограм. Білінійна чи бікубічна інтерполяція зменшує розмір матриці, зберігаючи структуру розподілу енергії. Метод простий у реалізації та ефективний на практиці.

Статистична агрегація перетворює часово-частотну матрицю у набір статистичних показників – середнє, дисперсію, асиметрію, ексцес, мінімум і максимум. Такий підхід створює компактне представлення, але може втрачати просторову інформацію про розподіл енергії.

Операції підвибірки (max або average pooling) зменшують розмір скалограми. Max pooling зберігає найвиразніші особливості сигналу, тоді як average pooling дає більш згладжене представлення. Поєднання обрізання та підвибірки забезпечує гнучке керування розміром даних.

Вибір методу стиснення залежить від задачі, ресурсів і вимог до точності. Надмірне стиснення призводить до втрати важливої інформації, а недостатнє – ускладнює обчислення.

## 1.6 Огляд існуючих досліджень у галузі розпізнавання транспортних шумів

За останні роки опубліковано значну кількість робіт, присвячених застосуванню методів машинного навчання та глибокого навчання для розпізнавання акустичних шумів, зокрема транспортних засобів.

Ранні дослідження зосереджувалися на використанні традиційних методів обробки сигналів та класичних алгоритмів машинного навчання. Застосування MFCC в поєднанні з методом опорних векторів або гаусівськими моделями сумішей демонструвало прийнятні результати для задач з обмеженою кількістю класів. Однак точність цих методів суттєво знижувалася при збільшенні різноманітності класів та в умовах високого рівня фонового шуму.

З розвитком глибокого навчання почалося активне застосування нейронних мереж для акустичної класифікації. Рекурентні нейронні мережі, зокрема LSTM, що здатні запам'ятовувати довгострокові залежності, використовувалися для аналізу часових послідовностей акустичних ознак.

Згорткові нейронні мережі довели свою ефективність при застосуванні до спектрограм та мел-спектрограм акустичних сигналів [8]. Численні дослідження показали, що CNN здатні автоматично виявляти складні часово-частотні паттерни і досягати високої точності розпізнавання без необхідності ручного конструювання ознак.

Деякі дослідження зосередилися на застосуванні вейвлет-перетворення для акустичної класифікації [9]. Було показано, що вейвлет-скалограми в поєднанні з класичними методами класифікації можуть перевершувати підходи, базовані на перетворенні Фур'є, особливо для нестационарних сигналів.

Гібридні архітектури, що поєднують різні типи нейронних мереж, також досліджувалися. Комбінація згорткових та рекурентних шарів дозволяє використовувати як просторові, так і часові залежності в даних [10,11]. Механізми уваги дозволяють мережі автоматично фокусуватися на найбільш інформативних частинах сигналу.

У контексті транспортних шумів було проведено дослідження з класифікації різних типів транспортних засобів за акустичними характеристиками. Деякі роботи зосереджувалися на розпізнаванні конкретних типів транспорту, таких як автомобілі різних класів, мотоцикли або залізничний транспорт. Інші роботи розглядали більш широкий спектр класів, включаючи авіаційний транспорт та фонові шуми міського середовища.

Важливим напрямком є дослідження стійкості методів розпізнавання до різних типів спотворень та шумів. Реальні умови міського середовища характеризуються високою варіативністю акустичних умов, наявністю ефектів відбиття та реверберації, змінним рівнем фонового шуму [12]. Застосування методів аугментації даних, таких як додавання шуму, зміна швидкості відтворення, реверберації, дозволяє підвищити робастність моделей [13,14].

Трансферне навчання також показує свою ефективність у задачах акустичної класифікації [15]. Використання попередньо навчених на великих наборах даних мереж з подальшим донавчанням на специфічних даних дозволяє досягти хороших результатів навіть при обмеженому обсязі навчальних даних.

Однак аналіз існуючих досліджень виявляє кілька прогалин. По-перше, недостатньо систематичних досліджень щодо порівняння різних типів вейвлет-перетворень для задачі розпізнавання транспортних шумів. Більшість робіт використовує один конкретний тип вейвлета без обґрунтування цього вибору.

По-друге, питання оптимального стиснення часово-частотних представлень залишається недостатньо вивченим. Різні методи стиснення впливають не тільки на обчислювальну ефективність, але й на якість розпізнавання, однак комплексних досліджень цього впливу для транспортних шумів практично немає.

По-третє, більшість досліджень використовує відносно невеликі набори даних або обмежену кількість класів, що ускладнює оцінку реальної ефективності методів. Створення репрезентативних наборів даних, що охоплюють широкий спектр транспортних засобів та умов запису, залишається актуальною задачею.

По-четверте, недостатньо уваги приділяється інтерпретованості моделей глибокого навчання. Розуміння того, які саме характеристики сигналів використовуються мережею для класифікації, важливе як для оцінки надійності моделей, так і для їх подальшого вдосконалення.

## 1.7 Постановка задачі дослідження

Таким чином, аналіз сучасного стану досліджень у галузі розпізнавання транспортних акустичних шумів виявляє актуальність комплексного дослідження ефективності різних підходів до часово-частотного аналізу в

поєднанні зі згортковими нейронними мережами. Незважаючи на численні роботи в цій галузі, залишаються невирішеними питання оптимального вибору типу вейвлет-перетворення, методів стиснення даних та їх взаємного впливу на якість розпізнавання.

Об'єктом дослідження є процес розпізнавання акустичних шумів різних типів транспортних засобів з використанням методів часово-частотного аналізу та глибокого навчання.

Метою дослідження є визначення оптимальних комбінацій типів вейвлет-перетворень та методів стиснення даних для ефективного розпізнавання транспортних акустичних шумів за допомогою згорткових нейронних мереж.

Для досягнення поставленої мети необхідно вирішити такі завдання:

- сформувати репрезентативний набір даних акустичних записів різних типів транспортних засобів та фонових шумів міського середовища;
- дослідити характеристики різних типів вейвлет-перетворень при застосуванні до транспортних акустичних шумів та оцінити їх здатність виявляти дискримінаційні ознаки різних класів транспортних засобів;
- розробити та дослідити ефективність різних методів стиснення часово-частотних представлень з точки зору балансу між зменшенням обчислювальної складності та збереженням інформативності даних;
- розробити архітектуру згорткової нейронної мережі, адаптовану для класифікації часово-частотних представлень транспортних шумів, з використанням сучасних методів регуляризації та оптимізації навчання;
- провести комплексні експериментальні дослідження та виконати порівняльний аналіз отриманих результатів для всіх можливих комбінацій типів вейвлетів та методів стиснення, оцінити точність і швидкість розпізнавання для кожного класу транспортних засобів та визначити оптимальні підходи та закономірності.

## 2 МЕТОДОЛОГІЯ ДОСЛІДЖЕННЯ РОЗПІЗНАВАННЯ ТРАНСПОРТНИХ ШУМІВ

### 2.1 Типи вейвлет-перетворень та їх математичний опис

Вейвлет-перетворення є потужним інструментом для аналізу нестационарних сигналів, який дозволяє одночасно досліджувати як частотні, так і часові характеристики. Основою вейвлет-аналізу є материнська вейвлет-функція  $\psi(t)$ , яка задовольняє умову допустимості та має нульове середнє значення.

Неперервне вейвлет-перетворення (CWT) сигналу  $x(t)$  визначається як:

$$W(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{+\infty} x(t) \psi^* \left( \frac{t-b}{a} \right) dt, \quad (2.1)$$

де  $a$  – параметр масштабу, що визначає частотні характеристики аналізу;

$b$  – параметр зсуву, що визначає часове положення вейвлета;

$\psi^*(t)$  – комплексно-спряжена материнська вейвлет-функція;

$\frac{1}{\sqrt{|a|}}$  – множник, що забезпечує нормалізацію енергії.

Параметр масштабу  $a$  обернено пропорційний частоті: малі значення  $a$  відповідають високим частотам (стиснутий вейвлет), великі значення  $a$  відповідають низьким частотам (розтягнутий вейвлет). Це забезпечує адаптивну часово-частотну роздільність, де високочастотні компоненти аналізуються з високою часовою роздільністю, а низькочастотні – з високою частотною роздільністю. На теперішній час існує багато різновидів вейвлетів, призначених для аналізу гармонічних сигналів, речових, різких акустичних сплесків. Так як обрані джерела акустичних сигналів генерують шуми, які містять складові різної природи, то однозначно обрати найкращий вейвлет за його математичним описом неможливо. Тому для досліджень оберемо

сукупність вейвлетів, що дозволяють ефективно аналізувати сигнали в частотній області.

### 2.1.1 Вейвлет Морле (Morlet)

Вейвлет Морле є одним з найпоширеніших вейвлетів для аналізу акустичних сигналів завдяки своїм чудовим частотним характеристикам [16,17]. Він являє собою комплексну експоненту, модульовану гаусівською огинаючою.

Материнська функція вейвлета Морле визначається як:

$$\psi(t) = \frac{1}{\sqrt{\pi}} e^{i\omega_0 t} e^{-\frac{t^2}{2}}, \quad (2.2)$$

де  $\omega_0$  – центральна частота вейвлета.

Дійсна частина вейвлета Морле має вигляд:

$$\psi_{Re}(t) = \frac{1}{\sqrt{\pi}} \cos(\omega_0 t) e^{-\frac{t^2}{2}}. \quad (2.3)$$

Основні властивості вейвлета Морле:

- комплексна функція, що дозволяє отримувати інформацію як про амплітуду, так і про фазу сигналу;
- чудова локалізація в частотній області завдяки синусоїдальній складовій;
- гаусівська огинаюча забезпечує добру локалізацію в часовій області;
- оптимальний баланс між часовою та частотною невизначеністю;
- особливо ефективний для виявлення гармонічних та квазігармонічних компонент сигналу.

Вейвлет Морле широко використовується для аналізу сигналів з виразною частотною структурою, що робить його природним вибором для транспортних шумів, де важливу роль відіграють гармоніки двигунів та інших механічних компонентів.

Зв'язок між масштабом  $a$  та псевдочастотою  $f$  для цього вейвлета:

$$f = \frac{\omega_0}{2\pi a f_s}, \quad (2.4)$$

де  $f_s$  – частота дискретизації сигналу.

### 2.1.2 Вейвлет Mexican Hat (Mexh)

Вейвлет Mexican Hat, також відомий як вейвлет Рікера, є дійсною функцією, що визначається як друга похідна гаусівської функції. Своєю назвою він зобов'язаний формі, що нагадує поперечний переріз мексиканського капелюха-сомбреро.

Математичний вираз для вейвлета Mexican Hat:

$$\psi(t) = \frac{2}{\sqrt{3}\sqrt{\pi}} (1 - t^2) e^{-\frac{t^2}{2}}. \quad (2.5)$$

Основні характеристики вейвлета Mexican Hat:

- дійсна симетрична функція;
- має нульове середнє та два бокових лепестки;
- чудова локалізація як у часовій, так і в частотній області;
- ефективний для виявлення локальних екстремумів та особливостей сигналу;
- добре підходить для детектування різких змін та перехідних процесів;
- має компактний носій, що сприяє швидким обчисленням.

Вейвлет Mexican Hat особливо корисний для виявлення імпульсних компонент у транспортних шумах, таких як удари, клацання та інші короткочасні події. Його симетрична форма забезпечує однакову чутливість до зростаючих та спадних фронтів сигналу.

Частотна характеристика Mexican Hat має один виразний пік, що робить його придатним для аналізу сигналів у певному частотному діапазоні. Масштабування цього вейвлета дозволяє адаптувати його до різних частотних компонент аналізованого сигналу.

### 2.1.3 Комплексний вейвлет Морле (CMOR)

Комплексний вейвлет Морле є узагальненням класичного вейвлета Морле з додатковими параметрами, що дозволяють гнучко налаштовувати його властивості. В бібліотеці PyWavelets він позначається як «*cmorBw-Fc*», де  $B_w$  – параметр ширини смуги,  $F_c$  – центральна частота.

Загальна форма комплексного вейвлета Морле:

$$\psi(t) = \frac{1}{\sqrt{\pi B_w}} e^{i2\pi F_c t} e^{-\frac{t^2}{B_w}}, \quad (2.6)$$

де  $B_w$  – контролює ширину вейвлета у часовій області;

$F_c$  – визначає центральну частоту комплексної експоненти.

Ключовими властивостями CMOR є:

- комплекснозначна функція, що надає амплітудну та фазову інформацію;
- параметр  $B_w$ , що дозволяє контролювати компроміс між часовою та частотною роздільністю;
- параметр  $F_c$ , що визначає центральну частоту аналізу;
- гнучкість у налаштуванні для специфічних характеристик сигналу;
- чудова частотна локалізація для аналізу гармонічних структур.

Комплексний вейвлет Морле особливо ефективний для транспортних шумів, оскільки дозволяє не тільки виявити присутність певних частотних компонент, але й відстежити їх фазові характеристики, що може бути критично важливим для розрізнення подібних джерел шуму.

#### 2.1.4 Вейвлет Frequency B-Spline (FBSP)

Вейвлет FBSP поєднує гладкість B-сплайнів з частотною локалізацією синусоїдальних функцій. Він позначається як «*fbspM-Bw-Fc*», де  $M$  – порядок B-сплайна,  $Bw$  – ширина смуги,  $Fc$  – центральна частота.

Материнська функція FBSP визначається як:

$$\psi(t) = B_M(t) \cdot \sin(2\pi F_c t), \quad (2.7)$$

де  $B_M(t)$  – B-сплайн порядку  $M$ .

B-сплайн порядку  $M$  обчислюється рекурсивно:

$$B_0(t) = \begin{cases} 1, & |t| \leq \frac{1}{2} \\ 0, & |t| > \frac{1}{2} \end{cases}. \quad (2.8)$$

$$B_M(t) = B_{M-1}(t) * B_0(t), \quad (2.9)$$

де  $*$  – позначає операцію згортки.

Основні характеристики вейвлета FBSP:

- дійсна функція з гладкою формою завдяки b-сплайнам;
- чудовий баланс між часовою та частотною локалізацією;
- гладкість у часовій області зменшує артефакти аналізу;
- компактність у частотній області забезпечує добре частотне розділення;
- ефективний для аналізу сигналів з плавними змінами спектра;

– менша чутливість до високочастотного шуму порівняно з вейвлетами-похідними.

Вейвлет FBSP особливо придатний для аналізу транспортних шумів з поступовими змінами частотних характеристик, таких як прискорення або уповільнення транспортного засобу. Його гладкість дозволяє уникнути хибних деталей, які можуть виникати при використанні більш нерегулярних вейвлетів.

Частотна характеристика FBSP має добре виражений головний пік з обмеженими боковими лепестками, що забезпечує чітке розділення частотних компонент сигналу.

### 2.1.5 Комплексний гаусівський вейвлет (CGAU)

Комплексний гаусівський вейвлет є узагальненням гаусівських вейвлетів з комплексною модуляцією. Він визначається як похідна певного порядку від комплексної гаусівської функції.

Загальна форма комплексного гаусівського вейвлета  $n$ -го порядку:

$$\psi_n(t) = C_n \frac{d^n}{dt^n} (e^{-t^2} e^{i\omega t}), \quad (2.10)$$

де  $C_n$  – нормувальний коефіцієнт;

$n$  – порядок похідної;

$\omega$  – частота модуляції.

Основні властивості комплексного гаусівського вейвлета:

- комплекснозначна функція;
- дуже гладка завдяки гаусівській огинаючій;
- чудова локалізація у часовій області;
- порядок похідної визначає кількість осциляцій;
- вищі порядки мають кращу частотну локалізацію;
- ефективний для аналізу гладких, повільно змінюваних сигналів.

Вибір порядку похідної визначає характеристики вейвлета:

- низькі порядки ( $n=1,2$ ) – краща часова локалізація, менше осциляцій;
- середні порядки ( $n=3,4$ ) – баланс між часовою та частотною роздільністю;
- високі порядки ( $n=5,6$ ) – краща частотна локалізація, більше осциляцій.

Для транспортних шумів комплексний гаусівський вейвлет корисний при аналізі плавних змін акустичних характеристик, таких як зміна тональності при зміні швидкості руху або режиму роботи двигуна.

#### 2.1.6 Вейвлет Шеннона (SHAN)

Вейвлет Шеннона базується на функції *sinc* ( $\sin(x)/x$ ) та забезпечує чудову локалізацію в частотній області. Він визначається через перетворення Фур'є з прямокутною частотною характеристикою.

Материнська функція вейвлета Шеннона:

$$\psi(t) = \frac{\sin(2\pi B t) - \sin(\pi B t)}{\pi t} e^{i\pi B t}, \quad (2.11)$$

де  $B$  – параметр ширини смуги.

Ключові характеристики вейвлета Шеннона:

- ідеальна локалізація в частотній області (прямокутна частотна характеристика);
- необмежений у часовій області (функція *sinc* спадає повільно);
- забезпечує точне частотне розділення без витоку енергії між смугами;
- ефективний для спектрального аналізу з чітким розділенням частот;
- комплекснозначна функція з фазовою інформацією.

Переваги для аналізу транспортних шумів:

- чітке розділення частотних смуг дозволяє точно ідентифікувати гармоніки двигунів;

- відсутність витоку енергії між частотами забезпечує точніші вимірювання;

- комплексна природа дозволяє аналізувати фазові співвідношення.

Недоліки:

- повільне спадання у часовій області може призводити до артефактів на краях сигналу;

- вимагає довших сигналів для коректного аналізу;

- чутливий до розривів та різких змін у сигналі.

Вейвлет Шеннона особливо корисний, коли потрібна висока точність частотного розділення, наприклад, для розрізнення близьких гармонік у спектрі шуму двигуна.

### 2.1.7 Порівняльний аналіз властивостей вейвлетів

У процесі вибору вейвлет-функцій для аналізу транспортних шумів важливим етапом є порівняльний аналіз їх основних характеристик, що визначають ефективність подальшого застосування методу. До таких характеристик належать часова та частотна локалізація, комплексність, гладкість функції, кількість осциляцій, обчислювальна складність, а також загальна придатність певного типу вейвлета для аналізу сигналів різної структури.

Одним із ключових параметрів є часова локалізація, що відображає здатність вейвлета точно визначати момент виникнення події в сигналі. Ця властивість є критичною при аналізі нестационарних або імпульсних процесів, характерних для транспортних систем. Вейвлет Mexican Hat вирізняється компактною симетричною формою, що забезпечує чітку часову локалізацію подій. Гаусівські вейвлети низьких порядків (CGAU) характеризуються гладкою гаусівською огибаючою з невеликою кількістю осциляцій, що також сприяє точному визначенню моментів появи сигналів. Вейвлет FBSP,

побудований на основі В-сплайнової огибаючої, забезпечує добрий компроміс між часовою роздільною здатністю та стійкістю до шуму, що робить його універсальним для аналізу короткотривалих подій.

Не менш важливою характеристикою є частотна локалізація, яка визначає здатність вейвлета розділяти близькі частотні компоненти сигналу. Вейвлет Shannon має ідеальну прямокутну частотну характеристику, завдяки чому забезпечує високу точність частотного розділення. Вейвлети Morlet і CMOR, до складу яких входить синусоїдальна компонента, характеризуються чудовими частотними властивостями та придатні для дослідження сигналів із виразними тональними складовими. FBSP, завдяки В-сплайновій структурі, демонструє збалансовані часово-частотні характеристики, що робить його ефективним для аналізу процесів із плавною зміною спектра.

Важливою особливістю, яка визначає інформативність аналізу, є комплексність вейвлет-функції. Вейвлети можуть бути дійсними або комплексними. До комплексних належать Morlet, CMOR, CGAU та Shannon, які дозволяють досліджувати як амплітудну, так і фазову складову сигналу. Такий підхід забезпечує більш повний опис коливальних процесів, оскільки фазова інформація може містити додаткові діагностичні ознаки. До дійсних вейвлетів належать Mexican Hat та FBSP, які оперують лише амплітудними характеристиками сигналу, що спрощує обчислення та підвищує стійкість до випадкових шумів.

Ще однією суттєвою характеристикою є гладкість функції вейвлета, що впливає на стабільність результатів аналізу та чутливість до шумових компонентів. Найбільш гладкими вважаються гаусівські вейвлети (CGAU) та вейвлет FBSP, побудований на основі В-сплайнів. Вейвлети Morlet і CMOR мають середній рівень гладкості, що дозволяє їм ефективно поєднувати точність і стійкість. Меншу гладкість демонструють Mexican Hat і Shannon, структура яких містить виражені коливання або функції типу *sinc*. Як правило, гладкіші вейвлети забезпечують більш стабільне відтворення основних компонентів сигналу та меншу чутливість до високочастотного шуму.

Кількість осциляцій у структурі вейвлет-функції визначає співвідношення між часовою та частотною роздільною здатністю. Вейвлети з малою кількістю осциляцій, такі як Mexican Hat (що має два бокові піки) або CGAU низьких порядків, є ефективними для аналізу коротких імпульсів і швидкоплинних процесів. Вейвлети Morlet, CMOR та FBSP характеризуються середньою кількістю осциляцій, забезпечуючи баланс між часовою та частотною локалізацією. У свою чергу, Shannon та CGAU високих порядків мають значну кількість осциляцій, що підвищує їх частотну вибірковість, але водночас потребує аналізу довших сигналів для коректного відтворення спектральної структури.

З практичної точки зору важливим аспектом є обчислювальна складність вейвлет-перетворення. Простішими у реалізації вважаються Mexican Hat та FBSP, які не потребують складних комплексних операцій. Середній рівень складності притаманний вейвлетам Morlet і CGAU, тоді як CMOR і Shannon відзначаються підвищеними обчислювальними витратами через наявність комплексних складових та додаткових параметрів. Вибір вейвлета з певною складністю визначається вимогами до швидкодії та точності аналізу в конкретних умовах застосування.

У контексті придатності для різних типів сигналів кожен вейвлет має свою специфічну область ефективного використання. Вейвлети Morlet і CMOR доцільно застосовувати для аналізу гармонічних структур, характерних для шумів двигунів та сигналів із вираженими тональними компонентами, а також при дослідженні процесів частотної модуляції. Вейвлет Mexican Hat ефективний для виявлення імпульсів, різких змін і перехідних процесів, що часто спостерігаються в акустичних сигналах транспортних систем. Вейвлет FBSP доцільно використовувати для аналізу сигналів із плавними змінами спектра, зокрема при дослідженні процесів прискорення чи уповільнення, оскільки він відзначається стійкістю до шумів. Гаусівський вейвлет CGAU добре підходить для аналізу гладких, повільно змінюваних сигналів, а також для робастного аналізу зашумлених даних і виявлення тенденцій у

спектральних характеристиках. Вейвлет Shannon доцільно використовувати для задач, що потребують високої точності частотного розділення та ідентифікації близьких гармонічних компонент, зокрема під час спектрального аналізу сигналів.

Отже, порівняльний аналіз властивостей вейвлетів визначає їх основні переваги та обмеження з урахуванням особливостей транспортних шумів. Кожен із розглянутих вейвлетів має специфічні характеристики, що забезпечують різний баланс між часовою та частотною локалізацією, гладкістю, складністю обчислень та стійкістю до шуму.

## 2.2 Методи стиснення часово-частотних представлень

Результатом вейвлет-перетворення акустичного сигналу є двовимірна матриця коефіцієнтів  $W(a,b)$  розміром  $N\_scales \times N\_timepoints$ , де  $N\_scales$  – кількість масштабів (частотних компонент),  $N\_timepoints$  – кількість часових відліків. Для типового аудіофрагменту тривалістю 0,372(16384/44100) секунди з частотою дискретизації 44100 Гц при використанні 128 масштабів розмір матриці становить  $128 \times 16384$ . Така висока розмірність створює серйозні обчислювальні проблеми при використанні згорткових нейронних мереж.

Основні проблеми, пов'язані з високою розмірністю:

- велика кількість вхідних параметрів призводить до експоненційного зростання кількості ваг у мережі;
- збільшуються вимоги до обсягу пам'яті GPU/CPU;
- подовжується час навчання та передбачення;
- зростає ризик перенавчання через надлишкову параметризацію;
- ускладнюється регуляризація та оптимізація моделі.

Тому критично важливим є застосування ефективних методів стиснення, які зменшують розмірність даних зі збереженням найбільш інформативних характеристик сигналу.

## 2.2.1 Метод зміни розміру з інтерполяцією (Resize)

Метод `Resize` базується на техніках масштабування зображень, застосованих до часово-частотної матриці вейвлет-коефіцієнтів. Основна ідея полягає в перетворенні вхідної матриці розміром  $128 \times 16384$  у матрицю стандартного розміру  $128 \times 128$  за допомогою інтерполяційних методів.

Зробимо математичний опис методу. Нехай  $W$  – вихідна матриця розміром  $(h_1, w_1) = (128, 16384)$ , потрібно отримати матрицю  $W'$  розміром  $(h_2, w_2) = (128, 128)$ . Коефіцієнти масштабування:

$$s_h = \frac{h_1}{h_2}, s_w = \frac{w_1}{w_2}. \quad (2.12)$$

Для даного випадку,  $s_h = 1.0$  (висота не змінюється),  $s_w = 128$ .

Для кожної точки  $(i, j)$  результуючої матриці обчислюється відповідна координата у вихідній матриці:

$$x = j \cdot s_w, y = i \cdot s_h. \quad (2.13)$$

Переваги методу `Resize`:

- зберігає загальну структуру часово-частотного розподілу енергії;
- гладка інтерполяція зменшує артефакти стиснення;
- відносно проста реалізація з використанням стандартних бібліотек;
- всі частотні компоненти представлені в результуючій матриці;
- зберігає топологічні властивості вихідного представлення.

Недоліки:

- втрата детальної часової інформації через усереднення;
- можливе розмиття різких переходів у сигналі;
- не враховує різну інформативність різних ділянок сигналу;
- фіксований розмір виходу може не бути оптимальним для всіх типів сигналів.

### 2.2.2 Метод статистичної агрегації (Statistics)

Метод Statistics радикально змінює підхід до представлення даних, перетворюючи повну часово-частотну матрицю в набір статистичних характеристик для кожної частотної смуги.

Концепція методу така, що для кожного масштабу (частотної компоненти) обчислюється набір статистичних показників на основі всіх часових відліків цього масштабу:

$$\{W(a_i, b_1), W(a_i, b_2), \dots, W(a_i, b_N)\}. \quad (2.14)$$

Набір статистичних характеристик включає:

– середнє значення (Mean) – відображає середню енергію на даній частоті протягом усього сигналу:

$$\mu_i = \frac{1}{N} \sum_{j=1}^N W_{ij}; \quad (2.15)$$

– стандартне відхилення (Standard Deviation) – характеризує варіативність амплітуди на даній частоті:

$$\sigma_i = \sqrt{\frac{1}{N} \sum_{j=1}^N (W_{ij} - \mu_i)^2}; \quad (2.16)$$

– коефіцієнт асиметрії (Skewness) – вимірює асиметрію розподілу значень відносно середнього:

$$\gamma_i = \frac{1}{N\sigma_i^3} \sum_{j=1}^N (W_{ij} - \mu_i)^3; \quad (2.17)$$

– коефіцієнт ексцесу (Kurtosis) – характеризує «гостроту» розподілу та наявність викидів:

$$\kappa_i = \frac{1}{N\sigma_i^4} \sum_{j=1}^N (W_{ij} - \mu_i)^4 - 3; \quad (2.18)$$

– середньоквадратичне значення (RMS) – відображає загальну енергію сигналу на даній частоті:

$$\text{RMS}_i = \sqrt{\frac{1}{N} \sum_{j=1}^N W_{ij}^2}; \quad (2.19)$$

– енергія (Energy) – сумарна енергія частотної компоненти:

$$E_i = \sum_{j=1}^N W_{ij}^2; \quad (2.20)$$

– мінімальне значення (Min):

$$W_{min,i} = \min_{j=1,\dots,N} W_{ij}; \quad (2.21)$$

– максимальне значення (Max):

$$W_{max,i} = \max_{j=1,\dots,N} W_{ij}; \quad (2.22)$$

– медіана (Median) – робастна оцінка центральної тенденції, менш чутлива до викидів:

$$\text{Med}_i = \text{median}\{W_{i1}, W_{i2}, \dots, W_{iN}\}; \quad (2.23)$$

– перший квантиль (Q25) – значення, нижче якого знаходиться 25% спостережень:

$$Q_{25,i} = \text{percentile}_{25}\{W_{i1}, W_{i2}, \dots, W_{iN}\}; \quad (2.24)$$

– третій квантиль (Q75) – значення, нижче якого знаходиться 75% спостережень:

$$Q_{75,i} = \text{percentile}_{75}\{W_{i1}, W_{i2}, \dots, W_{iN}\}; \quad (2.25)$$

– середнє абсолютне відхилення (MAD) – робастна міра варіативності:

$$\text{MAD}_i = \frac{1}{N} \sum_{j=1}^N |W_{ij} - \mu_i|. \quad (2.26)$$

В результаті, з матриці розміром  $128 \times 16384$  цей метод формує матрицю  $128 \times 12$ , де кожен рядок містить 12 статистичних характеристик для відповідної частотної компоненти.

Переваги методу:

– максимальне стиснення даних (редукція розмірності більш ніж у 1000 разів);

– компактне векторне представлення;

– мінімальні обчислювальні витрати при навчанні мережі;

– робастність до часових зсувів сигналу;

– статистичні характеристики мають чітку інтерпретацію.

Недоліки:

– повна втрата інформації про часову структуру сигналу;

– неможливість виявлення локальних часових подій;

– втрата інформації про послідовність та динаміку змін;

- агрегація може маскувати важливі короточасні характеристики;
- критична втрата просторової інформації для згорткових мереж.

### 2.2.3 Метод обрізання (Crop)

Метод Crop базується на гіпотезі, що найбільш інформативна частина сигналу зосереджена в певній часовій ділянці, зазвичай у центрі. Цей метод вирізає фрагмент заданої ширини з часово-частотної матриці.

Зробимо математичний опис методу. Нехай  $W$  – вихідна матриця розміром  $(h, w) = (128, 16384)$ , потрібно отримати матрицю  $W'$  розміром  $(h, w') = (128, 1024)$ .

Центральне обрізання формується наступним чином:

$$\text{start\_idx} = \left\lfloor \frac{w-w'}{2} \right\rfloor. \quad (2.27)$$

$$\text{end\_idx} = \text{start\_idx} + w'. \quad (2.28)$$

Результуюча матриця:

$$W' = W[:, \text{start\_idx}:\text{end\_idx}]. \quad (2.29)$$

Переваги методу Crop:

- збереження повної частотної інформації (всі масштаби присутні);
- збереження часової роздільності у вибраній ділянці;
- відсутність інтерполяції та пов'язаних з нею артефактів;
- простота та швидкість обчислень;
- можливість адаптивного вибору найінформативнішої ділянки.

Недоліки:

- втрата інформації з відкинутих ділянок;

- чутливість до вибору положення вікна обрізання;
- може пропустити важливі події поза вікном;
- фіксований розмір може не відповідати структурі різних сигналів;
- неефективність при рівномірному розподілі інформації по часу.

#### 2.2.4 Комбінований метод Crop + Max Pooling

Цей метод поєднує обрізання з операцією max pooling, що широко використовується в згорткових нейронних мережах. Ідея полягає в послідовному застосуванні обрізання та підвибірки.

Метод складається з двох етапів:

- етап 1 – обрізання до розміру  $128 \times 1024$ :

$$W_{crop} = \text{Crop}(W, \text{target\_width} = 1024); \quad (2.30)$$

- етап 2 – застосування Max Pooling з ядром  $4 \times 2$ . Max pooling розбиває матрицю на неперекривні блоки розміром  $(pool\_h, pool\_w)$  та вибирає максимальне значення в кожному блоці.

Для блоку з координатами  $(i, j)$ :

$$W'_{pool}[i, j] = \max_{m=0, \dots, p_h-1} \max_{n=0, \dots, p_w-1} W_{crop}[i \cdot p_h + m, j \cdot p_w + n], \quad (2.31)$$

де  $p_h$  – висота ядра пулінгу;

$p_w$  – ширина ядра пулінгу.

Переваги комбінованого методу:

- збереження найбільш виражених характеристик сигналу;
- інваріантність до невеликих часових зсувів;
- зменшення чутливості до локального шуму;
- суттєве зменшення розмірності для подальшої обробки;

– виділення домінуючих піків енергії.

Недоліки:

- втрата детальної інформації через агресивну підвибірку;
- можливість втрати важливих, але не максимальних значень;
- складність налаштування розмірів ядра пулінгу;
- накопичення втрат від двох послідовних операцій стиснення;
- необхідність адаптації архітектури мережі.

### 2.2.5 Метод усереднення блоків (Average Blocks, AVG)

Метод усереднення блоків є новим підходом, розробленим спеціально для ефективного стиснення часово-частотних представлень вейвлет-перетворення. Він забезпечує оптимальний баланс між зменшенням розмірності та збереженням інформативності даних.

Концепція даного методу наступна. Вихідна матриця  $W$  розміром  $(N_{scales}, N_{timepoints}) = (128, 16384)$  розбивається на  $N_{blocks}$  послідовних блоків однакового розміру  $(N_{scales}, block\_size)$ , де  $block\_size$  – цільовий розмір виходу. Для кожного рядка (частотної компоненти) обчислюється середнє значення по всіх блоках. Розрахунок кількості блоків:

$$N_{blocks} = \left\lfloor \frac{N_{timepoints}}{block\_size} \right\rfloor = \left\lfloor \frac{16384}{128} \right\rfloor = 128. \quad (2.32)$$

Обрізання до кратного розміру:

$$W_{trunc} = W[:, 0: N_{blocks} \times block\_size]. \quad (2.33)$$

Розбиття на блоки:

$$B[i, k, j] = W_{trunc}[i, k \times block\_size + j], \quad (2.34)$$

де  $i$  – індекс частоти;

$k$  – індекс блоку;

$j$  – індекс всередині блоку.

Усереднення блоків по вісі  $k$ :

$$W'[i, j] = \frac{1}{N_{blocks}} \sum_{k=0}^{N_{blocks}-1} B[i, k, j]. \quad (2.35)$$

Результуюча матриця представляє блок розміром 128 часових відліків, що характеризує середню структуру сигналу на цьому масштабі. Якщо сигнал має періодичну або квазіперіодичну структуру, метод AVG ефективно виявляє цю періодичність.

Переваги методу AVG:

- оптимальний баланс між стисненням та збереженням інформації;
- збереження часової структури на масштабі блоку;
- усереднення зменшує вплив локального шуму та випадкових викидів;
- виявлення періодичних та квазіперіодичних структур;
- робастність до часових варіацій сигналу;
- гарна сумісність з архітектурами snp (розмір  $128 \times 128$ ).

Недоліки:

- втрата інформації про довготривалі тренди;
- можлива втрата унікальних короткочасних подій;
- фіксований розмір блоку може не завжди відповідати характерним масштабам сигналу.

## 2.2.6 Модифікований метод усереднення блоків (AVG1)

Метод AVG1 є варіацією базового методу AVG з іншим розміром блоку, орієнтованим на виявлення більш детальних короткочасних структур.

На відміну від AVG, який використовує блоки по 128 відліків, AVG1 – по 10 відліків:

$$N_{blocks} = \left\lfloor \frac{16384}{10} \right\rfloor = 1638. \quad (2.36)$$

У AVG1 більша ніж у AVG деталізація часової структури:

- виявлення швидких змін амплітуди;
- краще розділення короткочасних подій;
- чутливість до високочастотної модуляції.

AVG1 також має більше усереднення по блоках:

- сильніше придушення нерегулярного шуму;
- виділення стабільних повторюваних структур;
- стійкість до випадкових викидів.

Це робить AVG1 особливо ефективним для сигналів з: швидкими переходами, високочастотною модуляцією, короткочасними повторюваними структурами та імпульсним характером і повинна давати вищу точість після навчання.

## 2.2.7 Метод сингулярного розкладу (SVD)

Метод SVD (Singular Value Decomposition) застосовує сингулярний розклад матриці для виділення найбільш значущих компонент та редукції розмірності.

Він дозволяє представити будь-яку вагову матрицю  $W$  ( $m, n$ ) нейронної мережі (наприклад, 128 на 16384) як добуток:

$$W = U\Sigma V^T, \quad (2.37)$$

де  $U$  – матриця лівих сингулярних векторів розміром  $(m, m) = (128, 128)$ ;

$\Sigma$  – діагональна матриця сингулярних значень розміром  $(m, n)$ , які впорядковані за спаданням і вказують на важливість відповідних компонент;

$V^T$  – транспонована матриця правих сингулярних векторів розміром  $(n, n)$ .

Впорядкованість робить SVD ідеальним для стиснення: замість збереження повної матриці, використовується лише апроксимація низького рангу  $k$ , зберігаються тільки перших  $k$  сингулярних значень та відповідні сингулярні вектори:

$$W_k = U_k \Sigma_k V_k^T. \quad (2.38)$$

Вибір оптимального рангу  $k$  базується на критерії збереження енергії, де  $k$  обирається таким чином, щоб частка збереженої інформації  $\eta_k$ , перевищувала заданий поріг (наприклад 95%):

$$\eta_k = \frac{\sum_{i=1}^k \sigma_i^2}{\sum_{i=1}^r \sigma_i^2}. \quad (2.39)$$

Можна використовувати компактне представлення, що дасть матрицю розміром  $(128, k)$ :

$$W' = U_k \Sigma_k. \quad (2.40)$$

Метод SVD має переваги у вигляді оптимальної апроксимації, збереження значущих компонент сигналу, зменшення шуму та виявлення прихованих структур. Він також дозволяє адаптивно вибирати рівень стиснення. Однак його недоліки включають високу обчислювальну складність, можливу втрату локальних особливостей, знищення просторової структури, складність інтерпретації сингулярних векторів та необхідність налаштування параметра  $k$ .

### 3 СТВОРЕННЯ ДАТАСЕТУ, ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ ТА ТЕСТУВАННЯ

#### 3.1 Формування та характеристика експериментального набору даних

Створення якісного та репрезентативного набору даних є фундаментальною основою для розробки ефективної системи розпізнавання транспортних акустичних шумів [18]. Якість кінцевих результатів безпосередньо залежить від того, наскільки добре навчальні дані відображають реальне різноманіття транспортних звуків, з якими система може зіткнутися в практичних умовах експлуатації. У рамках цього дослідження було сформовано спеціалізований набір даних, що охоплює широкий спектр типів транспортних засобів та акустичних умов, характерних для міського середовища.

##### 3.1.1 Вибір класів транспортних засобів та обґрунтування структури датасету

При формуванні набору даних особлива увага приділялася вибору класів транспортних засобів, які найбільш часто зустрічаються в міському середовищі та створюють значний внесок у загальний акустичний фон. Вибір 9 класів був обумовлений необхідністю охопити максимально широкий спектр типових джерел транспортного шуму, що мають різні акустичні характеристики та представляють різні категорії транспорту.

Перший клас представлений звуками літаків, які характеризуються специфічним низькочастотним гулом від реактивних двигунів або пропелерів. Акустичний профіль авіаційного транспорту відрізняється високою інтенсивністю та характерною динамікою зміни частоти під час зльоту, польоту та посадки. Звуки літаків зазвичай мають виражену низькочастотну складову в діапазоні від 50 до 500 Гц, що пов'язано з роботою великих турбін та

аеродинамічними ефектами. Водночас присутні також високочастотні компоненти, що виникають від турбулентності повітряних потоків та механічних систем літака.

Другий клас охоплює фонові шуми міського середовища, які включають різноманітні звуки, що не пов'язані безпосередньо з конкретними транспортними засобами. Це можуть бути звуки вітру, шум дерев, розмови людей, звуки будівельної техніки на відстані, загальний міський гул від віддалених джерел. Включення цього класу є критично важливим для того, щоб система навчилася розрізняти специфічні транспортні шуми від загального акустичного фону. Без цього класу існував би високий ризик хибних спрацьовувань, коли система помилково ідентифікувала б фонові звуки як транспортні засоби. Фонові шуми характеризуються широкосмуговим спектром з відсутністю виражених гармонічних структур, що типово для механічних джерел звуку.

Третій клас представлений велосипедами, які створюють специфічні звуки, пов'язані переважно з механічним тертям ланцюгової передачі, обертанням коліс та взаємодією шин з дорожнім покриттям. На відміну від моторизованих транспортних засобів, велосипеди не мають двигунів внутрішнього згоряння, тому їх акустичний профіль значно тихіший та має інші спектральні характеристики. Звуки велосипедів зосереджені переважно в середньому та високочастотному діапазоні, мають імпульсний характер, пов'язаний з обертанням педалей та ланцюга. Розпізнавання велосипедів є складним завданням через низьку інтенсивність їх акустичного сигналу порівняно з іншими класами транспорту.

Четвертий клас включає автобуси, які є представниками громадського транспорту з дизельними двигунами великого об'єму. Автобуси створюють характерний потужний низькочастотний шум від роботи двигуна, що доповнюється звуками від пневматичної системи гальм, відчинення та зачинення дверей, роботи компресорів кондиціонування. Акустичний спектр автобусів має виражені гармоніки в діапазоні від 100 до 700 Гц, що відповідає

частоті обертання колінчастого валу дизельного двигуна. Великі розміри та маса автобусів також впливають на характер звуку від взаємодії шин з дорогою, створюючи додаткові низькочастотні компоненти.

П'ятий клас складається з легкових автомобілів, які є найбільш численною категорією транспорту в міських умовах. Сучасні легкові автомобілі можуть мати як бензинові, так і дизельні двигуни, а також гібридні силові установки. Їх акустичний профіль характеризується середньою інтенсивністю з балансом між низько, середньо та високочастотними компонентами. Звук роботи двигуна легкового автомобіля зазвичай має основну частоту в діапазоні від 50 до 300 Гц залежно від об'єму двигуна та режиму роботи. При прискоренні спостерігається характерне зміщення частотного спектру в бік вищих частот, що відображає збільшення обертів двигуна.

Шостий клас представлений вертольотами, які мають унікальний акустичний підпис, що суттєво відрізняється від інших типів транспорту. Основним джерелом звуку вертольота є несучий гвинт, що створює характерний ритмічний шум з частотою, пов'язаною з кількістю лопатей та швидкістю обертання. Звук вертольота має виражену періодичну структуру з основною частотою в діапазоні від 5 до 800 Гц та численними гармоніками. Також присутній широкосмуговий шум від турбулентності повітря навколо обертаючихся лопатей. Акустичний профіль вертольота містить як низькочастотні басові компоненти від лопатей, так і високочастотний шум від двигунів та редукторів.

Сьомий клас складається з мотоциклів, які характеризуються специфічним високоінтенсивним звуком двигуна з яскраво вираженими гармоніками. Мотоциклетні двигуни зазвичай мають менший об'єм порівняно з автомобільними, але працюють на значно вищих обертах, що призводить до формування характерного високочастотного звуку. Спектр мотоциклетного шуму зосереджений переважно в діапазоні від 300 до 1000 Гц з численними яскраво вираженими гармоніками. Багато мотоциклів мають модифіковані системи випуску, що додатково посилює специфічний звук двигуна.

Акустичний сигнал мотоцикла також містить характерні імпульсні компоненти, пов'язані з циклами роботи циліндрів.

Восьмий клас включає поїзди, які представляють залізничний транспорт з унікальним набором акустичних характеристик. Звуки поїздів включають ритмічні удари від контакту коліс з рейками, гул від електричних тягових двигунів або дизель-електричних установок, свист від аеродинамічного обтікання, а також характерний металевий скрип при проходженні поворотів. Частота ударів коліс про стики рейок залежить від швидкості руху та довжини рейкових плетей, створюючи характерну ритмічну структуру звуку. Електричні поїзди також генерують специфічний високочастотний гул від інверторів та трансформаторів. Загальний акустичний спектр поїздів є широкосмуговим з виразними піками на певних частотах, пов'язаних з механічними резонансами конструкції.

Дев'ятий клас представлений вантажними автомобілями, зокрема великими фурами, які мають потужні дизельні двигуни великого об'єму та створюють інтенсивний низькочастотний шум. Вантажні автомобілі характеризуються масивною конструкцією, що впливає на всі аспекти їх акустичного профілю. Звук двигуна фури має виражені низькочастотні компоненти в діапазоні від 50 до 200 Гц, що створює відчуття потужного гулу. Додатковими джерелами шуму є турбокомпресор, що створює характерний високочастотний свист, система випуску відпрацьованих газів, пневматична система гальм, яка періодично генерує різкі шиплячі звуки при спрацьовуванні. Шум від взаємодії багатьох коліс з дорожнім покриттям також вносить значний внесок у загальний акустичний профіль.

Вибір саме цих 9 класів забезпечує репрезентативне охоплення основних категорій транспорту, що зустрічаються в міських та приміських умовах. Кожен клас має достатньо відмінні акустичні характеристики, щоб забезпечити можливість їх автоматичного розрізнення, водночас деякі класи мають певну подібність, що створює складність для системи класифікації та дозволяє адекватно оцінити її дискримінаційну здатність. Приклади візуалізованих

частотних спектрів та спектрограм деяких транспортних засобів, отриманих за допомогою вейвлет-перетворення без стиснення, зображено на рисунку 3.1.

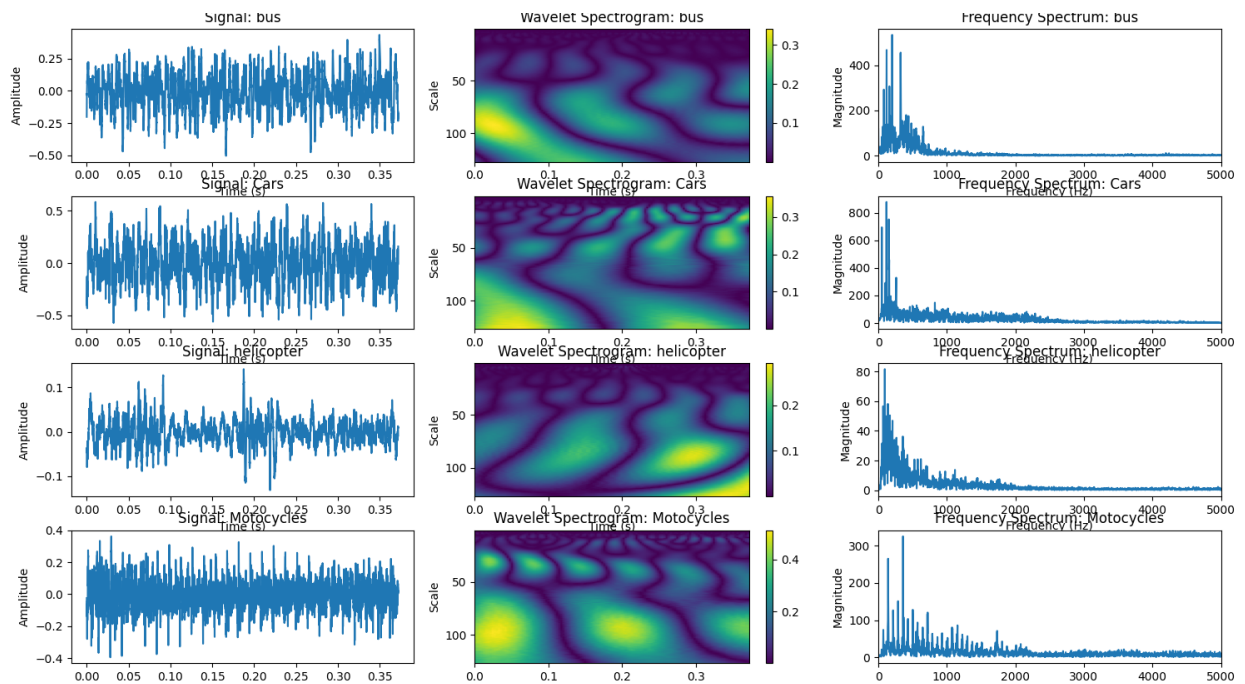


Рисунок 3.1 – Приклади сигналів, їх спектрів та спектрограми вейвлета Morlet

### 3.1.2 Технічні параметри запису та обробки аудіоданих

Якість вхідних аудіоданих є критично важливим фактором для подальшого аналізу та класифікації. У цьому дослідженні використовувалася частота дискретизації 44100 Гц, що є стандартом для аудіо компакт-дисків та забезпечує достатню частотну смугу для аналізу акустичних сигналів. Згідно з теоремою Найквіста-Шеннона, частота дискретизації 44100 Гц дозволяє коректно відтворювати частотні компоненти до 22050 Гц, що повністю покриває діапазон чутності людського вуха та охоплює всі акустично значущі компоненти транспортних шумів.

Тривалість кожного аудіофрагмента становила 0,372 секунди, що відповідає приблизно одній третині секунди. Вибір такої тривалості був зумовлений необхідністю балансу між достатньою часовою роздільністю для

аналізу динаміки сигналу та обмеженнями обчислювальних ресурсів. Фрагмент тривалістю близько третини секунди є достатнім для того, щоб зафіксувати характерні особливості звуку транспортного засобу, включаючи кілька циклів роботи двигуна, характерні модуляції та перехідні процеси. Водночас така відносно коротка тривалість дозволяє ефективно обробляти великі обсяги даних без надмірних вимог до пам'яті та обчислювальної потужності.

При частоті дискретизації 44100 Гц та тривалості 0,372 секунди кожен аудіофрагмент містить 16384 дискретних відліки. Це забезпечує високу деталізацію сигналу в часовій області та дозволяє застосовувати вейвлет-перетворення з достатньою часовою роздільністю для виявлення короткочасних особливостей сигналу.

Всі аудіозаписи були представлені в моно форматі, тобто з одним аудіоканалом. Вибір моно запису замість стерео був обумовлений кількома факторами. По-перше, для задачі класифікації типу транспортного засобу просторова інформація про положення джерела звуку не є критично важливою, оскільки класифікація базується на спектральних та часових характеристиках самого звуку, а не на його локалізації в просторі. По-друге, використання моно запису зменшує обсяг даних вдвічі порівняно зі стерео, що знижує вимоги до пам'яті та прискорює обробку. По-третє, в реальних умовах практичного застосування системи розпізнавання часто використовуються моно мікрофони, тому навчання на моно даних забезпечує кращу відповідність між навчальними умовами та умовами експлуатації.

Для кожного з 9 класів транспортних засобів було зібрано по 2500 оригінальних аудіозаписів. Таким чином, загальна кількість оригінальних записів у наборі даних становила 22500. Ця кількість була обрана як компроміс між необхідністю мати достатньо великий та різноманітний набір даних для навчання глибоких нейронних мереж та практичними обмеженнями на збір та обробку аудіоматеріалу. Велика кількість записів на клас забезпечує достатнє різноманіття акустичних умов, відстаней до джерела, режимів роботи транспортних засобів та фонових умов.

Важливим аспектом формування датасету була необхідність забезпечити різноманітність умов запису в межах кожного класу. Записи звуків літаків включали як звуки під час зльоту та набору висоти, так і звуки крейсерського польоту та посадки. Записи автомобілів охоплювали різні режими руху, від холостого ходу до активного прискорення, від низьких до високих швидкостей. Фонові шуми були записані в різних локаціях міського середовища, включаючи парки, житлові райони, ділові центри. Така різноманітність умов запису є критично важливою для того, щоб навчена модель мала хорошу узагальнювальну здатність та могла коректно класифікувати транспортні шуми в умовах, які не були представлені в навчальних даних.

### 3.1.3 Формування фінального датасету

У результаті підготовки даних загальна кількість аудіоспектрограм у датасеті становила 22500, що відповідає кількості оригінальних аудіозаписів, оскільки для кожного запису створювалася одна спектрограма без додаткової фрагментації. Для кожного з дев'яти класів містилося по 2500 прикладів, що забезпечувало повну збалансованість датасету. Збалансованість класів є критично важливою для коректного навчання моделі класифікації, оскільки при незбалансованому датасеті модель може навчитися надавати перевагу більш представленим класам.

Розподіл даних на навчальну, валідаційну та тестову вибірки було виконано з використанням стратифікованого підходу, що гарантує однакові пропорції класів у кожній підвибірці. Навчальна вибірка містила 60% даних, тобто 13500 прикладів, по 1500 на кожен клас. Валідаційна вибірка становила 20%, що дорівнює 4500 прикладів, або 500 на клас. Тестова вибірка також містила 20% даних з аналогічним розподілом по класах.

Вибір такого співвідношення між навчальною, валідаційною та тестовою вибірками обумовлений необхідністю забезпечити достатній обсяг даних для

навчання моделі та водночас мати незалежні підвибірки для валідації та фінальної оцінки. Навчальна вибірка є найбільшою, що дозволяє ефективно оптимізувати параметри моделі. Валідаційна вибірка використовується для контролю якості під час навчання та налаштування гіперпараметрів. Тестова вибірка повністю ізольована від процесу навчання і використовується виключно для фінальної оцінки моделі.

Критично важливою була вимога забезпечити, щоб дані з одного аудіофайлу не потрапляли одночасно в різні вибірки. Оскільки для кожного аудіофайлу створюється одна спектрограма, розподіл здійснювався на рівні оригінальних записів, що гарантує незалежність між навчальним, валідаційним та тестовим наборами.

#### 3.1.4 Попередня обробка аудіосигналів перед вейвлет-перетворенням

Перед застосуванням вейвлет-перетворення всі аудіофрагменти проходили процедуру попередньої обробки, метою якої було приведення сигналів до стандартизованого вигляду та зменшення впливу факторів, не пов'язаних безпосередньо з характеристиками джерела звуку. Першим кроком попередньої обробки була нормалізація амплітуди сигналу. Нормалізація виконувалася шляхом ділення кожного відліку сигналу на максимальне абсолютне значення в цьому фрагменті. Така процедура приводить амплітуду сигналу до діапазону від -1 до 1.

Нормалізація амплітуди є необхідною процедурою з кількох причин. По-перше, оригінальні записи могли бути зроблені з різною загальною гучністю залежно від відстані до джерела звуку, налаштувань записуючого обладнання, умов поширення звуку. Без нормалізації ці відмінності в загальному рівні гучності могли б стати домінуючим фактором при класифікації, маскуючи справжні спектральні характеристики, що визначають тип транспортного засобу. По-друге, нормалізація забезпечує стабільність вейвлет-перетворення та

наступних операцій обробки, оскільки всі сигнали мають однаковий динамічний діапазон. По-третє, нормалізація відповідає природі задачі класифікації, де цікавить саме спектральна структура та часова динаміка звуку, а не його абсолютна гучність.

Після нормалізації амплітуди до кожного фрагмента застосовувалося неперервне вейвлет-перетворення. Вейвлет-перетворення виконувалося з використанням 128 масштабів, що охоплювали широкий діапазон частот від низьких до високих. Вибір саме 128 масштабів був обумовлений необхідністю забезпечити достатню частотну роздільність для виявлення детальних особливостей спектру транспортних шумів, водночас утримуючи обчислювальну складність на прийнятному рівні. 128 масштабів дозволяють отримати детальне представлення частотної структури сигналу з роздільністю, достатньою для розрізнення близьких гармонік та виявлення тонких відмінностей між класами.

Результатом застосування вейвлет-перетворення до кожного аудіофрагмента була двовимірна матриця вейвлет-коефіцієнтів [19]. Перший вимір цієї матриці, розміром 128, відповідав частотним компонентам або масштабам вейвлет-перетворення. Менші значення масштабу відповідають вищим частотам, а більші значення масштабу відповідають нижчим частотам. Другий вимір матриці, розміром 16384 відліки, відповідав часовій розгортці сигналу. Таким чином, кожен елемент матриці вейвлет-коефіцієнтів містив інформацію про енергію сигналу на певній частоті в певний момент часу, що забезпечувало одночасне часово-частотне представлення акустичного сигналу.

Матриця вейвлет-коефіцієнтів містила комплексні значення, які включали як амплітудну, так і фазову інформацію про сигнал. Однак для подальшої обробки згортковими нейронними мережами використовувалися модулі комплексних коефіцієнтів, що представляють амплітудну інформацію. Вибір амплітудної інформації був обумовлений тим, що саме розподіл енергії сигналу в часово-частотній області є найбільш інформативною характеристикою для класифікації типу транспортного засобу.

Після обчислення модулів вейвлет-коефіцієнтів виконувалася додаткова нормалізація отриманої матриці. Нормалізація виконувалася методом мінімум-максимум. Така процедура приводила всі значення в матриці до діапазону від 0 до 1. Додатково, для запобігання ділення на нуль в випадку вироджених сигналів, до знаменника додавалася мала константа порядку  $10^{-8}$ . Ця нормалізація вейвлет-коефіцієнтів забезпечувала однорідність вхідних даних для нейронної мережі та сприяла стабільності процесу навчання.

### 3.2 Розроблені архітектури згорткових нейронних мереж

Розробка архітектури згорткової нейронної мережі для задачі класифікації транспортних шумів вимагала врахування специфіки часово-частотних представлень, отриманих в результаті вейвлет-перетворення. На відміну від класичних задач комп'ютерного зору, де вхідними даними є природні зображення з трьома кольоровими каналами, в нашому випадку вхідними даними є скалограми, які мають специфічну структуру з частотною віссю та часовою віссю. Ця специфіка вимагала розробки спеціалізованих архітектур, адаптованих до різних методів стиснення часово-частотних представлень.

#### 3.2.1 Базова архітектура нейронних мереж для більшості методів

Для методів стиснення, які в результаті формують вихідні матриці стандартного розміру  $128 \times 128$  елементів, була розроблена універсальна архітектура згорткової нейронної мережі, що складається з послідовності згорткових блоків з поступовим збільшенням глибини представлення. Ця архітектура базується на класичних принципах побудови глибоких згорткових

мереж, де кожен наступний шар виявляє більш абстрактні та складні ознаки на основі простіших ознак, виявлених попередніми шарами.

Вхідний шар мережі приймає тензор розміром (128, 128, 1), де перші два виміри відповідають просторовим вимірам скалограми (частота та час), а третій вимір представляє єдиний канал даних, що містить амплітудну інформацію вейвлет-коефіцієнтів. Така структура вхідних даних аналогічна до монохромних зображень у комп'ютерному зорі, що дозволяє застосувати перевірені архітектурні рішення з області обробки зображень, адаптуючи їх до специфіки акустичних даних.

Перший згортковий блок мережі починається з шару двовимірної згортки, що містить 32 фільтри з розміром ядра  $3 \times 3$ . Вибір невеликого розміру ядра  $3 \times 3$  є стандартною практикою в сучасних глибоких нейронних мережах, оскільки такі фільтри здатні ефективно виявляти локальні паттерни, водночас мінімізуючи кількість параметрів мережі. Кількість фільтрів 32 у першому шарі є достатньою для виявлення базових ознак, таких як локальні зміни енергії, переходи між частотними смугами, часові градієнти. Згортковий шар використовує режим *padding='same'*, що означає доповнення вхідних даних нулями по краях таким чином, щоб розмір виходу співпадав з розміром входу. Це дозволяє зберегти просторову роздільність даних та уникнути втрати інформації з крайових областей скалограми.

Після згорткового шару застосовується функція активації ReLU (Rectified Linear Unit), що вносить нелінійність у модель. Функція ReLU перетворює від'ємні значення в нулі, залишаючи додатні значення без змін. Ця проста, але ефективна функція активації вирішує проблему зникаючих градієнтів, що характерна для сигмоїдальних функцій активації, та забезпечує швидку збіжність навчання.

Наступним компонентом блоку є шар батч-нормалізації (Batch Normalization), який нормалізує активації кожного міні-батча під час навчання. Батч-нормалізація обчислює середнє значення та стандартне відхилення для активацій в межах батча і використовує ці статистики для нормалізації даних до

розподілу з нульовим середнім та одиничною дисперсією. Додатково, батч-нормалізація вводить два навчані параметри для кожного каналу, що дозволяють мережі відновити необхідний масштаб та зсув розподілу, якщо це потрібно для оптимальної роботи. Використання батч-нормалізації приносить численні переваги: прискорення навчання через можливість використання вищих швидкостей навчання, зменшення чутливості до ініціалізації ваг, додаткову регуляризацию, що зменшує потребу в інших методах регуляризації.

Після батч-нормалізації застосовується операція max pooling з вікном розміру  $2 \times 2$  та кроком 2. В результаті просторові розміри даних зменшуються вдвічі по кожному виміру, що призводить до зменшення розміру з  $128 \times 128$  до  $64 \times 64$  після першого блоку.

Операція max pooling виконує кілька важливих функцій: зменшує обчислювальну складність наступних шарів через зменшення просторової роздільності, забезпечує інваріантність до невеликих локальних зсувів та деформацій, збільшує рецептивне поле наступних згорткових шарів, дозволяючи їм аналізувати більші за розміром області вхідних даних.

Завершує перший блок шар dropout з ймовірністю відключення 0,25. Dropout є методом регуляризації, що випадково відключає (обнуляє) 25% нейронів під час кожної ітерації навчання [20]. Це змушує мережу не покладатися на окремі нейрони і розвивати більш розподілені представлення, що покращує узагальнювальну здатність моделі. Під час тестування dropout не застосовується, але виходи масштабуються відповідно до ймовірності відключення, щоб компенсувати різницю між навчанням та inference (передбаченням).

Другий згортковий блок має аналогічну структуру до першого, але містить 64 фільтри замість 32. Подвоєння кількості фільтрів на кожному наступному рівні є стандартною практикою в глибоких згорткових мережах і відображає ієрархічну природу виявлення ознак. Перші шари виявляють прості, низькорівневі ознаки, такі як краї та локальні текстури, тоді як глибші шари комбінують ці прості ознаки у більш складні та абстрактні представлення.

Збільшення кількості фільтрів компенсує зменшення просторової роздільності через pooling, дозволяючи мережі зберігати або навіть збільшувати загальну інформаційну ємність представлення.

Після другого згорткового блоку з його компонентами (згортка, ReLU, батч-нормалізація, max pooling  $2 \times 2$ , dropout 0,25) просторовий розмір даних зменшується до  $32 \times 32$ , а кількість каналів збільшується до 64. На цьому рівні абстракції мережа вже виявляє більш складні паттерни, що можуть відповідати характерним часово-частотним структурам, специфічним для певних типів транспортних засобів, таким як періодичні модуляції від роботи двигуна або характерні енергетичні піки в певних частотних діапазонах.

Третій згортковий блок продовжує цю ієрархію з 128 фільтрами. Після обробки третім блоком розмір даних стає  $16 \times 16 \times 128$ . На цьому рівні мережа виявляє ще більш абстрактні та високорівневі ознаки, що можуть представляти комбінації різних акустичних характеристик, специфічних для класів транспортних засобів. Наприклад, комбінація певних гармонічних структур з певною часовою динамікою може бути характерною саме для мотоциклів, відрізняючи їх від автомобілів.

Четвертий згортковий блок містить 256 фільтрів і зменшує розмір даних до  $8 \times 8 \times 256$ . Це найглибший рівень згорткової обробки в базовій архітектурі, де представлення даних є найбільш абстрактним та високорівневим. На цьому етапі мережа сформувала компактне, але інформативне представлення вхідної скалограми, де кожен з 256 каналів включає інформацію про присутність або відсутність певних складних акустичних паттернів, релевантних до класифікації.

Після завершення послідовності згорткових блоків тривимірний тензор розміром  $8 \times 8 \times 256$  необхідно перетворити в одновимірний вектор для подальшої обробки повнозв'язними шарами. Це досягається за допомогою операції Flatten, яка розгортає всі елементи тензору в єдиний вектор довжиною 16384 елементів ( $8 \times 8 \times 256 = 16384$ ). Цей вектор є фіксованим представленням

вхідної скалограми у високовимірному просторі ознак, де кожна координата кодує присутність певної високорівневої характеристики.

Перший повнозв'язний шар містить 512 нейронів і перетворює 16384-вимірний вектор ознак у 512-вимірне представлення. Повнозв'язний шар виконує афінне перетворення з наступною нелінійною активацією ReLU. Кожен з 512 нейронів має з'єднання з усіма 16384 входами, що дозволяє йому аналізувати глобальні комбінації виявлених згортковими шарами ознак. Після активації ReLU застосовуються батч-нормалізація та dropout з ймовірністю 0,5. Вищий рівень dropout у повнозв'язних шарах (0,5 замість 0,25 у згорткових блоках) є типовою практикою, оскільки повнозв'язні шари мають значно більше параметрів і більш схильні до перенавчання.

Другий повнозв'язний шар зменшує розмірність до 256 нейронів, продовжуючи процес вилучення найбільш релевантних для класифікації ознак. Структура з поступовим зменшенням розмірності (512 до 256) створює звужувальну архітектуру, що сприяє формуванню компактного, але інформативного представлення. Цей шар також включає активацію ReLU, батч-нормалізацію та dropout 0,5.

Фінальний вихідний шар містить 9 нейронів, що відповідає кількості класів у задачі класифікації. Цей шар використовує функцію активації softmax, яка перетворює логіти (сирі виходи нейронів) у розподіл ймовірностей. Функція softmax нормалізує виходи таким чином, що вони стають невід'ємними та сумуються до одиниці, що дозволяє інтерпретувати їх як ймовірності належності вхідного зразка до кожного з 9 класів. Під час inference клас з найвищою ймовірністю вибирається як передбачення моделі.

Загальна кількість параметрів у базовій архітектурі становить приблизно 10 мільйонів. Більшість цих параметрів зосереджена у повнозв'язних шарах, особливо в першому повнозв'язному шарі, що з'єднує 16384 входів з 512 виходами, що дає  $16384 \times 512 = 8.4$  мільйони ваг тільки в цьому одному шарі. Така велика кількість параметрів надає моделі високу виразну потужність, але

також вимагає значного обсягу навчальних даних та ретельної регуляризації для запобігання перенавчанню.

### 3.2.2 Архітектура для методу Statistics

Метод стиснення Statistics формує значно менші вхідні матриці розміром  $128 \times 12$ , де 128 рядків відповідають частотним компонентам, а 12 стовпців містять статистичні характеристики розподілу енергії в часовій області для кожної частоти. Така радикальна редукція просторової розмірності вимагає суттєвої модифікації архітектури мережі, щоб ефективно працювати з компактними, але високоінформативними статистичними дескрипторами.

Вхідний шар приймає тензор розміром (128, 12, 1). Мала ширина входу (12 стовпців) створює специфічні виклики для архітектури, оскільки застосування стандартних операцій max pooling швидко призведе до вироджування просторового виміру. Тому архітектура для методу Statistics використовує більш обережний підхід до зменшення просторової роздільності.

Перший згортковий блок містить 32 фільтри з ядром  $3 \times 3$ , ReLU активацію, батч-нормалізацію, max pooling  $2 \times 1$  та dropout 0,25, аналогічно до базової архітектури. Після цього блоку розмір даних стає  $64 \times 12 \times 32$ . На цьому етапі та подальших, часовий вимір, що дорівнює 12, достатньо малий і не потребує зменшення при операціях пулінгу.

Другий згортковий блок з 64 фільтрами та аналогічною структурою зменшує розмір до  $32 \times 12 \times 64$ .

Третій згортковий блок містить 128 фільтрів з ядром  $3 \times 3$  і не включає операцію max pooling. Замість цього після згортки, активації, батч-нормалізації та dropout розмір даних залишається  $32 \times 12 \times 128$ . Відсутність pooling дозволяє зберегти просторову структуру, яка на цьому етапі вже є мінімальною.

Замість стандартної операції Flatten, яка призвела б до вектора розміром  $32 \times 12 \times 128 = 49152$ , архітектура для методу Statistics використовує Global

Average Pooling 2D. Ця операція обчислює середнє значення по всіх просторових позиціях для кожного каналу окремо, що призводить до вектора розміром 128 (по одному значенню для кожного з 128 каналів). Global Average Pooling має кілька важливих переваг: драматичне зменшення кількості параметрів у наступних повнозв'язних шарах, природна регуляризація через усереднення [21], інваріантність до точного просторового положення ознак, відсутність додаткових параметрів для навчання.

Після Global Average Pooling 128-вимірний вектор обробляється послідовністю повнозв'язних шарів. Перший повнозв'язний шар містить 256 нейронів з ReLU активацією, батч-нормалізацією та dropout 0,5. Другий повнозв'язний шар містить 128 нейронів з аналогічною структурою. Фінальний вихідний шар містить 9 нейронів з softmax активацією.

Загальна кількість параметрів у архітектурі для методу Statistics становить приблизно 450 тисяч, що майже в 20 разів менше порівняно з базовою архітектурою. Така значна редукція кількості параметрів пояснюється насамперед використанням Global Average Pooling замість Flatten. У базовій архітектурі перехід від згорткових шарів до повнозв'язних створював вузьке місце з мільйонами параметрів, тоді як в архітектурі для Statistics цей перехід є набагато більш компактним завдяки попередній агрегації просторової інформації.

Менша кількість параметрів має як переваги, так і недоліки. З одного боку, модель навчається швидше, вимагає менше пам'яті, менш схильна до перенавчання на обмежених датасетах. З іншого боку, обмежена ємність моделі може бути недостатньою для повного відображення складності взаємозв'язків між статистичними дескрипторами та класами транспортних засобів, що може призвести до нижчої точності класифікації порівняно з архітектурами, що працюють з більш детальними часово-частотними представленнями.

### 3.2.3 Архітектура для методу Crop+MaxPooling

Даний метод Crop формує вхідні матриці розміром  $128 \times 1024$ , що значно ширші порівняно з базовою архітектурою. Ця асиметрія між висотою та шириною вхідних даних вимагає спеціального підходу до проектування архітектури, зокрема використання асиметричних ядер max pooling для збалансованого зменшення просторових розмірів.

Вхідний шар приймає тензор розміром  $(128, 1024, 1)$ . Якби застосовувався стандартний max pooling з квадратним ядром  $2 \times 2$ , після чотирьох згорткових блоків розмір даних став би приблизно  $8 \times 64$ , що означає, що часовий вимір все ще був би майже в 10 разів більшим за частотний. Така нерівномірність просторових розмірів не є оптимальною для ефективної роботи згорткових фільтрів та призводить до неефективного використання параметрів моделі.

Щоб вирішити цю проблему, архітектура для методу Crop використовує асиметричні ядра max pooling, які зменшують часовий вимір більш агресивно, ніж частотний. Перший згортковий блок містить 32 фільтри  $3 \times 3$ , ReLU, батч-нормалізацію, але max pooling виконується з ядром  $2 \times 8$  (2 по частоті, 8 по часу). Це зменшує розмір з  $128 \times 1024$  до  $64 \times 128$ , вже значно більш збалансований.

Другий згортковий блок з 64 фільтрами також використовує асиметричний max pooling  $2 \times 8$ , зменшуючи розмір до  $32 \times 16$ . Тепер співвідношення сторін становить приблизно 2:1, що є прийнятним для подальшої обробки. Третій блок з 128 фільтрами використовує менш агресивний max pooling  $2 \times 4$ , отримуючи розмір  $16 \times 4$ . Четвертий блок з 256 фільтрами не використовує max pooling, залишаючи розмір  $16 \times 4 \times 256$ .

Далі архітектура може використовувати або Global Average Pooling для отримання 256 вимірного вектора, або Flatten для отримання вектора розміром  $16 \times 4 \times 256 = 16384$ . У випадку використання Flatten, подальша структура повнозв'язних шарів аналогічна до базової архітектури: Dense(512), Dense(256),

Dense(9). У випадку Global Average Pooling структура повнозв'язних шарів може бути спрощена, оскільки вхідний вектор вже є компактним.

Використання асиметричних ядер pooling є ключовою інновацією архітектури для методу Stop, що дозволяє ефективно обробляти вхідні дані з великим співвідношенням сторін, застосовуючи принципи цифрової обробки до часово-частотних представлень [22]. Альтернативою могло б бути використання асиметричних згорткових фільтрів (наприклад,  $3 \times 5$  замість  $3 \times 3$ ), але практика показує, що комбінація стандартних квадратних фільтрів з асиметричним pooling забезпечує кращу гнучкість та продуктивність.

### 3.2.4 Використання архітектури EfficientNetV0

У межах дослідження, окрім створення архітектур нейронних мереж з нуля, була розглянута можливість використання готових ефективних згорткових моделей, серед яких окреме місце посіла EfficientNetV0. Мережі сімейства EfficientNet є одними з найбільш збалансованих та результативних згорткових архітектур сучасності. Їх популярність пояснюється здатністю забезпечувати значно вищу точність при суттєво меншій кількості параметрів та обчислень у порівнянні з класичними моделями, такими як ResNet, Inception чи MobileNet. Наприклад, ResNet для досягнення високої точності потребує десятків мільйонів параметрів і має значну глибину, тоді як Inception використовує складні багатогілкові блоки з великою кількістю внутрішніх операцій. MobileNet, хоча й орієнтована на обчислювальну ефективність, у багатьох випадках поступається EfficientNet за точністю. EfficientNet досягає кращого компромісу завдяки систематичному масштабуванню архітектури та використанню оптимізованих MBConv-блоків з механізмом Squeeze-and-Excitation. У результаті модель EfficientNetV0 демонструє високу продуктивність при мінімальному розмірі, що робить її однією з найпривабливіших архітектур для задач класифікації з обмеженими ресурсами.

Позначення B0 у назві моделі вказує на те, що це базова та найменша версія сімейства EfficientNet. Вона була отримана шляхом автоматичного пошуку архітектур і використовується як початкова модель, на основі якої будуються масштабовані варіації B1-B7. Ці старші моделі мають більшу глибину, ширину та роздільність вхідних зображень, що дозволяє підвищувати точність, але водночас збільшує обчислювальну складність та вимоги до пам'яті. Вибір EfficientNetB0 у рамках дослідження зумовлений її оптимальним співвідношенням між точністю, швидкістю та компактністю, а також малим частотним спектром у 128 одиниць для більших моделей. Вона майже не потребує збільшених розмірів вхідних даних та є найбільш раціональним варіантом у випадках, коли модель має бути достатньо легкою, але водночас здатною демонструвати якісну класифікацію складних структур сигналів.

На відміну від поширених підходів трансферного навчання, у роботі не використовувалися попередньо навчені ваги EfficientNetB0, отримані на наборі даних ImageNet. Це пояснюється тим, що часово-частотні спектральні представлення транспортних шумів суттєво відрізняються від природних зображень, на яких навчалася базова модель. Статистична природа спектрограм не відповідає структурі RGB-фотографій, що могло б призвести до некоректної генералізації та неефективного перенесення ознак на початкових шарах. Тому застосування EfficientNetB0 у дослідженні ґрунтувалося виключно на її архітектурних перевагах, тоді як навчання здійснювалося з нуля без використання зовнішніх попередньо навчальних ваг.

Важливою технічною проблемою була невідповідність формату даних: EfficientNetB0 очікує на вході кольорові зображення, тоді як у дослідженні використовувалися одноканальні скалограми. Серед можливих рішень цієї проблеми існують підходи, пов'язані з модифікацією першого шару мережі або частковим збереженням ваг через усереднення каналів. Проте в рамках роботи був застосований найпростіший та найстабільніший підхід – реплікація одноканальних спектрограм у три однакових канали, що забезпечує відповідність формату без втручання в структуру моделі. Такий спосіб не

спотворює вихідні дані та дозволяє працювати з архітектурою EfficientNetB0 без необхідності зміни її ваг або оптимізатора.

Кількість параметрів у моделі на основі EfficientNetB0 становить приблизно 4,5 мільйони, з яких близько 4 мільйонів припадає на базову модель EfficientNetB0, і приблизно 0,5 мільйона – на додані класифікаційні шари. Це значно менше, ніж у базовій архітектурі, розробленій з нуля (приблизно 8,75 мільйона параметрів), що робить EfficientNetB0-based модель більш компактною і потенційно швидшою під час inference. Крім компактності, EfficientNetB0 має доведену ефективність: у численних бенчмарках на стандартному наборі зображень модель показує високу точність порівняно з важчими архітектурами. Наприклад, EfficientNetB0 демонструє приблизно 77,1% точності на ImageNet, при приблизно 5,3 мільйонів параметрів, що порівнянно з результатами таких важчих мереж, як ResNet-50, але при значно меншій обчислювальній складності. І якщо взяти легкі (з невеликою кількістю параметрів, низькою обчислювальною складністю) моделі, як-от MobileNetV2, вони мають меншу точність за тих же або подібних умов, а великі моделі – такі як Inception v3 – потребують значно більше параметрів і ресурсів, аби забезпечити невелике підвищення точності. Таким чином, для задач із обмеженими обчислювальними ресурсами або потребою в швидкому inference, EfficientNetB0 забезпечує найкращий компроміс: достатню точність при невеликому розмірі та відносно невеликих витратах ресурсів.

### 3.3 Результати експериментальних досліджень

Після завершення розробки архітектур та підготовки даних було проведено всебічні експериментальні дослідження для оцінки ефективності різних комбінацій типів вейвлет-перетворень та методів стиснення. Дослідження охопили шість різних типів вейвлетів (Morlet, Mexican Hat, CMOR, FBSP, CGAU, Shannon) та сім методів стиснення (Resize, Statistics,

Crop, Crop+MaxPooling, AVG, AVG1, SVD), що дало багато різних комбінацій для тестування. Для яких був виконан повний цикл навчання моделі з оцінкою на тестовій вибірці.

### 3.3.1 Методологія проведення експериментів та метрики оцінювання

У ході дослідження для реалізації та навчання моделей було використано мову програмування Python та фреймворк TensorFlow, який забезпечив ефективне виконання обчислень на основі тензорних операцій і підтримку апаратного прискорення. Процес навчання моделі здійснювався у середовищі Google Colab, що надало доступ до обчислювальних ресурсів хмарної інфраструктури, зокрема графічних процесорів (GPU), а також дало можливість зручно і швидко відтворити та масштабувати експерименти.

Всі експерименти проводилися в однакових умовах для забезпечення справедливості порівняння. Навчання кожної моделі виконувалося з використанням оптимізатора Adam з початковою швидкістю навчання 0,001. Розмір батча становив 16 зразків, що є компромісом між стабільністю градієнтів (більші батчі дають більш стабільні оцінки градієнта) та обчислювальною ефективністю (менші батчі дозволяють виконувати більше оновлень ваг за епоху).

Максимальна кількість епох була встановлена на рівні 100, але фактично жодна з моделей не навчалася повні 100 епох завдяки використанню механізму EarlyStopping. Цей механізм моніторить значення функції втрат на валідаційній вибірці та зупиняє навчання, якщо протягом 8 послідовних епох не спостерігається покращення. Водночас EarlyStopping зберігає ваги моделі з найкращим значенням валідаційної функції втрат, що дозволяє уникнути деградації якості через перенавчання на пізніх стадіях тренування.

Додатково використовувався механізм ReduceLROnPlateau, який динамічно зменшує швидкість навчання при виявленні плато в покращенні

валідаційної функції втрат. Якщо протягом 4 послідовних епох валідаційна функція втрат не покращується, швидкість навчання зменшується вдвічі. Це дозволяє моделі робити більш тонкі налаштування ваг на пізніх стадіях навчання, коли великі кроки градієнтного спуску можуть перестрибувати через оптимум.

Як функцію втрат використовувалася категоріальна cross-entropy, стандартна для задач багатокласової класифікації. Ця функція вимірює розбіжність між розподілом ймовірностей, передбачених моделлю, та справжнім розподілом (one-hot encoded мітки класів). Мінімізація цієї функції еквівалентна максимізації логарифмічної правдоподібності коректної класифікації.

Для комплексної оцінки якості класифікації для кожної навченої моделі будувалася confusion matrix (матриця помилок) на тестовій вибірці. Confusion matrix є квадратною матрицею розміром  $9 \times 9$ , де рядки відповідають справжнім класам, а стовпці – передбаченим класам. Елемент матриці в позиції  $(i, j)$  показує кількість зразків класу  $i$ , які були класифіковані як клас  $j$ . Діагональні елементи матриці відповідають коректним класифікаціям, тоді як позадіагональні елементи представляють помилки різних типів.

Для зручності порівняння та аналізу матриць помилок були нормалізовані таким чином, що сума елементів в кожному рядку дорівнює одиниці. Це дозволяє інтерпретувати елементи матриці як ймовірності класифікації зразка справжнього класу  $i$  як клас  $j$ .

Базова версія програмного коду для проведення досліджень була додана у дотаток А.

Щоб побачити візуальне перетворення оригінальних спектрограм (рис. 3.2 (а)) в скалограми (після стиснення) (рис. 3.2 (б)), були відображені по одному зразку для частини класів.

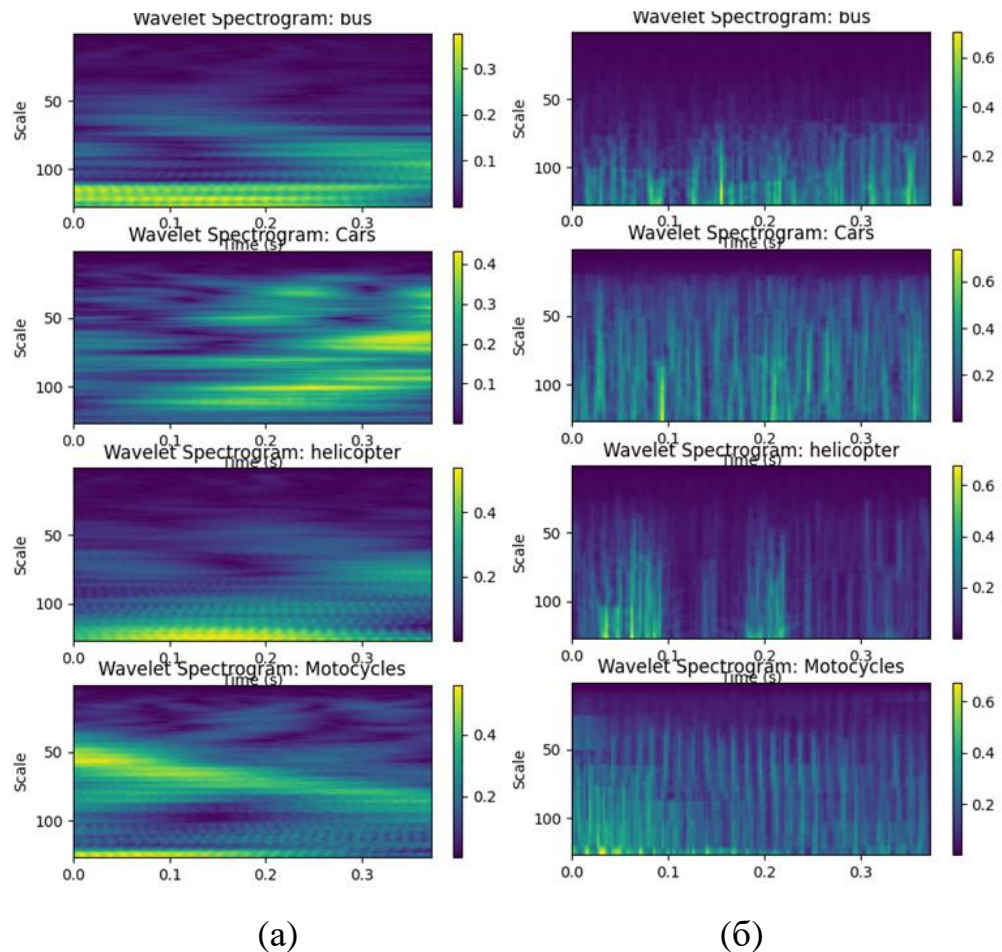


Рисунок 3.2 – Приклад спектрограм з вейвлетом FBSP:

(а) – до стиснення; (б) – після стиснення

### 3.3.2 Результати для вейвлета Morlet з різними методами стиснення

Вейвлет Morlet був обраний як базовий для початкових експериментів завдяки своїм добре відомим властивостям та широкому використанню в аналізі акустичних сигналів. Експерименти з вейвлетом Morlet дозволили оцінити відносну ефективність різних методів стиснення та встановити базовий рівень продуктивності для подальших порівнянь. Візуально можна побачити різницю між методами стиснення з цим вейвлетом на спектрограмах, що на рисунку Б.1. Всі спектрограми були візуалізовані зі всіма методами стиснення, незалежно від використаних з тим чи іншим вейвлетом, окрім комбінованого

методу стиснення Crop+MaxPooling, так як він загалом не дає можливості побачити результати пулінга окремої карти ознак в процесі навчання.

Метод Resize в комбінації з вейвлетом Morlet показав добрі загальні результати з середньою точністю класифікації 0,86 по всіх класах. Аналіз результатів по окремих класах виявив значну варіативність якості розпізнавання. Найвищу точність було досягнуто для класу Train (поїзд) з значенням 0,95, що свідчить про високу відмінність акустичного профілю поїздів від інших класів транспорту. Характерні ритмічні удари від контакту коліс з рейками створюють унікальний акустичний підпис, який ефективно виявляється вейвлет-аналізом та успішно розпізнається нейронною мережею.

Клас Background (фон) також демонстрував високу точність розпізнавання 0,94, що пояснюється фундаментальними відмінностями між фоновими шумами та специфічними транспортними звуками. Фонові шуми характеризуються більш широкосмуговим та менш структурованим спектром без виражених гармонічних компонент, що робить їх легко відрізняваними від механічних джерел звуку.

Висока точність 0,92 була досягнута для класу Cars (легкові автомобілі), що свідчить про ефективність вейвлет-перетворення Morlet у виявленні характерних гармонічних структур, пов'язаних з роботою автомобільних двигунів.

Для класу Bicycle (велосипед) точність становила 0,83, це можна вважати не поганим результатом з урахуванням специфічних труднощів розпізнавання цього класу. Велосипеди генерують значно тихіші звуки порівняно з моторизованими транспортними засобами, і їх акустичний сигнал легко маскується фоновим шумом. Досягнута точність 0,83 свідчить про те, що навіть такі слабкі сигнали містять достатньо інформації для успішної класифікації при використанні вейвлет-аналізу та глибокого навчання.

Класи Helicopter (вертольот) та Truck (фура) показали точність 0,80 та 0,86 відповідно. Вертольоти мають специфічний періодичний акустичний профіль від обертання лопатей, який добре виявляється вейвлет-перетворенням.

Фури характеризуються потужними низькочастотними компонентами від дизельних двигунів великого об'єму, що також ефективно представляються вейвлет-коефіцієнтами.

Найнижча точність у комбінації Morlet+Resize спостерігалася для класу Motorcycles (мотоцикли) з значенням 0,73. Це може бути пов'язано з великою варіативністю акустичних характеристик мотоциклів різних типів, від скутерів до потужних спортивних мотоциклів з модифікованими системами випуску.

Метод Statistics з вейвлетом Morlet показав значно гірші результати з середньою точністю лише 0,43. Це драматичне зниження якості пояснюється втратою практично всієї інформації про часову структуру сигналу при перетворенні часово-частотної матриці в набір статистичних дескрипторів. Для транспортних шумів динаміка зміни спектральних характеристик в часі несе критично важливу інформацію, яка повністю втрачається при статистичній агрегації.

Аналіз результатів по окремих класах для методу Statistics виявляє загальну деградацію продуктивності з точністю в діапазоні від 0,32 до 0,54. Навіть відносно легкі для розпізнавання класи, такі як поїзди та фоновий шум, демонструють точність лише 0,38 та 0,54 відповідно, що лише трохи краще за випадкове відгадування для 9 класів (яке дало б точність 0,11). Це підтверджує, що статистичні характеристики розподілу енергії в часі, хоча і несуть певну інформацію про джерело звуку, є недостатніми для надійної класифікації транспортних шумів.

Метод Crop показав проміжні результати з середньою точністю 0,62. Це краще за метод Statistics (0,43), але гірше за метод Resize (0,86). Така проміжна продуктивність пояснюється специфікою методу Crop, який зберігає детальну часово-частотну інформацію, але тільки для обмеженої центральної частини сигналу. Якщо найбільш інформативні характеристики транспортного звуку зосереджені в цій центральній частині, метод працює добре. Однак якщо важливі акустичні події відбуваються на початку або в кінці фрагменту, вони втрачаються при обрізанні.

Аналіз результатів по класах для методу Crop виявляє також значну неоднорідність. Клас Train зберігає відносно високу точність 0,85, тобто характерні ритмічні структури звуків поїздів присутні протягом усього сигналу і успішно захоплюються центральною частиною. Клас Cars з точністю 0,77, що є достатнім, але нижчим порівняно з методом Resize. Найбільше страждають класи Motorcycles (0,39) та Truck (0,39), де точність впадає майже до рівня випадкового відгадування. Це може вказувати на те, що акустичні характеристики цих класів мають більш виражену часову динаміку з важливими подіями, розподіленими по всій довжині сигналу, а не зосередженими в центрі.

Комбінований метод Crop+MaxPooling з вейвлетом Morlet показав середню точність 0,60, що лише трохи гірше за простий метод Crop. Ідея цього методу полягала в тому, щоб залишити більшу часову частину спектрограм після Crop, і потім стискати її операцією max pooling, з відповідними коефіцієнтами, нейронних мереж. Тож така зміна не допомогла, результати не показали бажаного покращення.

Детальний аналіз confusion matrix для методу Crop+MaxPooling виявляє цікаві патерни помилок. Найвища точність 0,90 досягається для класу Train, що підтверджує стійкість характеристик залізничного транспорту до різних видів обробки даних. Клас Background демонструє точність 0,72, що є нижчим порівняно з іншими методами. Класи Helicopter (0,71) та Cars (0,70) показують помірну точність. Найбільші труднощі спостерігаються для класу Truck з точністю лише 0,27, що вказує на те, що важливі низькочастотні характеристики вантажних автомобілів можуть бути втрачені при комбінації обрізання та пулінгу.

Метод AVG (усереднення блоків) з вейвлетом Morlet показав найкращі результати серед усіх досліджених комбінацій для цього вейвлета з середньою точністю 0,85. Це свідчить про ефективність підходу в розбитті довгого часового сигналу на блоки та усередненні їх.

Для транспортних шумів, особливо тих, що пов'язані з роботою двигунів, існує виражена квазіперіодична структура, пов'язана з циклами роботи циліндрів. Усереднення блоків ефективно виділяє цю періодичну компоненту, водночас зменшуючи випадковий шум та нерегулярні події. Це пояснює, чому метод AVG досягає такої високої точності – він оптимально налаштований на виявлення періодичних характеристик, які є домінуючими в акустичних профілях більшості моторизованих транспортних засобів.

Аналіз результатів по класах наступний. Клас Train досягає точності 0,97, що є одним з найвищих результатів у всьому дослідженні. Клас Background демонструє точність 0,95, що також є відмінним результатом. Класи Cars (0,88) та Bicycle (0,91) показують дуже добру точність. І навіть найскладніший клас Motorcycles досягає точності 0,79, що значно краще порівняно з іншими методами стиснення. Єдиним винятком є клас Truck з нижчою точністю 0,67.

Цікаво порівняти метод AVG з методом Resize, які обидва формують вихідні матриці однакового розміру  $128 \times 128$  і використовують ту ж саму архітектуру нейронної мережі. Метод AVG показує лише трохи гіршу середню точність (0,85 проти 0,86 для Resize). Усереднення блоків може для деяких класів більш підходити та краще зберігати корисну інформацію порівняно з інтерполяцією для певних типів сигналів з виразною періодичною структурою.

### 3.3.3 Результати для вейвлета Mexican Hat з методом AVG

Вейвлет Mexican Hat був протестований в комбінації з методом AVG для оцінки його ефективності в порівнянні з вейвлетом Morlet. Mexican Hat має інші властивості порівняно з Morlet – він є дійсним симетричним вейвлетом, ефективним для виявлення локальних екстремумів та різких змін у сигналі, на відміну від комплексного Morlet, орієнтованого на виявлення гармонічних структур. Візуально можна побачити різницю між методами стиснення з цим вейвлетом на рисунку Б.2.

Комбінація Mexican Hat+AVG показала середню точність 0,83 – це дуже близько до результату Morlet+AVG (0,85). Це свідчить про те, що обидва типи вейвлетів можуть ефективно представляти транспортні шуми, хоча виявляють різні аспекти сигналів.

Клас Train, як і з Morlet, з відмінною точністю 0,97, дав з вейвлетом Mexican Hat точність 0,98. Це можна пояснити тим, що Mexican Hat ефективно виявляє імпульсні компоненти в звуках поїздів, пов'язані з ударами коліс об рейки. Ці імпульси створюють різкі локальні екстремуми в сигналі, які є природною областю застосування Mexican Hat вейвлета.

Клас Background показує точність 0,97, що також є відмінним результатом і навіть краще за Morlet+AVG (0,95). Фонові шуми часто містять різноманітні імпульсні та перехідні події, які добре характеризуються Mexican Hat перетворенням. Клас Cars демонструє точність 0,92, що є дещо вищим за результат з Morlet (0,88). Звуки автомобілів містять як гармонічні компоненти від двигуна, так і імпульсні події від взаємодії з дорогою, і Mexican Hat може краще виявляти ці імпульсні компоненти.

Однак для деяких класів Mexican Hat демонструє гіршу продуктивність порівняно з Morlet. Клас Helicopter показує точність 0,70, що нижче за результат Morlet+AVG (0,90). Звуки вертольотів характеризуються виразною періодичною структурою від обертання лопатей, яка краще представляється гармонічно-орієнтованим вейвлетом Morlet, ніж імпульсно-орієнтованим Mexican Hat. Аналогічно, клас Motorcycles демонструє точність 0,69, що нижче за Morlet (0,79).

Клас Bus показує несподівано низьку точність 0,67 з Mexican Hat, що значно гірше за більшість інших класів. Це може бути пов'язано з тим, що характерний низькочастотний гул дизельних двигунів автобусів має виражену гармонічну структуру, яка краще представляється вейвлетом Morlet. Mexican Hat, будучи орієнтованим на виявлення імпульсів та різких змін, може бути менш чутливим до цих плавних гармонічних компонент.

Цікавим є результат для класу Truck з точністю 0,82, що значно краще за результат Morlet+AVG (0,67). Це можна пояснити тим, що звуки вантажних автомобілів містять численні імпульсні компоненти від взаємодії багатьох коліс з дорогою, спрацьовування пневматичної системи, механічні удари в трансмісії. Mexican Hat ефективно виявляє ці імпульсні події, що призводить до кращого розпізнавання класу Truck.

Загалом, порівняння вейвлетів Morlet та Mexican Hat виявляє, що вибір оптимального вейвлета залежить від характеру акустичних сигналів конкретного класу. Для класів з виразною гармонічною структурою (вертольоти, мотоцикли, автобуси) кращим є Morlet, тоді як для класів з більшою кількістю імпульсних компонент (поїзди, фури) Mexican Hat може давати кращі результати.

### 3.3.4 Результати для вейвлетів CMOR та FBSP

Вейвлети CMOR (Complex Morlet) та FBSP (Frequency B-Spline) були протестовані з методом Resize для оцінки їх ефективності в порівнянні з базовим вейвлетом Morlet. CMOR є узагальненням Morlet з додатковими параметрами для контролю частотної та часової локалізації, тоді як FBSP поєднує гладкість B-сплайнів з частотною локалізацією синусоїдальних функцій. Візуально можна побачити різницю між методами стиснення з вейвлетом CMOR на рисунку Б.3 та з вейвлетом FBSP на рисунку Б.4.

Комбінація CMOR+Resize показала середню точність 0,86, що ідентично до результату Morlet+Resize. Тобто додаткові параметри CMOR не приносять значних переваг для даної задачі при використанні типових налаштувань. Аналіз по окремих класах виявляє схожий розподіл точності з невеликими варіаціями. Клас Train демонструє точність 0,97, клас Background – 0,93, клас Cars – 0,88.

Найбільша різниця спостерігається для класу Airplane, де CMOR досягає точності 0,87 порівняно з 0,75 для Morlet. Це суттєве покращення може бути пов'язано з тим, що звуки літаків мають складну частотну структуру з широким спектром від низькочастотного гулу турбін до високочастотного шуму від аеродинаміки. Додаткова гнучкість CMOR в налаштуванні частотно-часової локалізації дозволяє краще представити цю складну структуру.

Комбінація FBSP+Resize показала середню точність 0,88, що є найкращим результатом серед усіх протестованих комбінацій вейвлетів з методом Resize. Це вказує на те, що властивості вейвлета FBSP, зокрема його гладкість та збалансована часово-частотна локалізація, є особливо придатними для аналізу транспортних шумів. Для більшості класів результати чудові.

Клас Train досягає точності 0,97, клас Background – 0,96, що є одними з найвищих результатів у дослідженні. Клас Cars демонструє точність 0,93, що значно краще за результати з Morlet (0,92) та CMOR (0,88). Особливо помітне покращення спостерігається для класу Bicycle з точністю 0,92 порівняно з 0,83 для Morlet. Це свідчить про те, що гладкість FBSP та його здатність ефективно представляти сигнали з плавними змінами спектру є перевагою для аналізу слабких та зашумлених сигналів велосипедів.

Клас Helicopter демонструє точність 0,88 з FBSP, що є хорошим результатом, хоча трохи нижче за деякі інші комбінації. Клас Bus досягає точності 0,85, що є солідним результатом. Навіть найскладніший клас Motorcycles показує точність 0,77 з FBSP, що краще за більшість інших комбінацій. Клас Truck демонструє точність 0,86, що також є добрим результатом.

Цікаво відзначити, що FBSP демонструє більш збалансовану продуктивність по класах порівняно з іншими вейвлетами. Різниця між найкращим (Train – 0,97) та найгіршим (Motorcycles – 0,77) класом становить 0,20, тоді як для Morlet+Resize ця різниця становила 0,22 (від 0,95 для Train до 0,73 для Motorcycles). Така більш рівномірна продуктивність вказує на універсальність FBSP для різних типів акустичних сигналів.

FBSP також був протестований з іншими методами стиснення. Комбінація FBSP+Crop показала середню точність 0,78, що є добрим результатом, але нижче за FBSP+Resize (0,88). Це підтверджує, що метод Resize краще зберігає інформацію порівняно з методом Crop, незалежно від типу використаного вейвлета. Аналіз по класах для FBSP+Crop виявляє схожі патерни: високу точність для Train (0,94), Background (0,88) та Cars (0,86), та нижчу точність для Motorcycles (0,60) та Truck (0,73).

Комбінація FBSP+SVD, де використовувався метод сингулярного розкладу для стиснення, показала середню точність 0,66. Це значно нижче за FBSP+Resize, що вказує на те, що метод SVD, незважаючи на свою математичну елегантність та здатність виділяти основні компоненти, призводить до втрати важливої інформації для класифікації транспортних шумів. Можливо, SVD занадто агресивно стискає дані або виділяє компоненти, що не є оптимальними для розрізнення класів транспортних засобів.

Комбінація FBSP+AVG показала середню точність 0,88, що ідентично до FBSP+Resize. Це цікавий результат, що підтверджує еквівалентність цих двох методів стиснення з точки зору збереження інформації, релевантної для класифікації. Аналіз по класах виявляє дуже близькі результати з невеликими варіаціями. Це свідчить про те, що як інтерполяція (Resize), так і усереднення блоків (AVG) ефективно зберігають характерні особливості часово-частотних представлень, отриманих з використанням вейвлета FBSP.

### 3.3.5 Результати для вейвлета CGAU з різними методами

Комплексний гаусівський вейвлет (CGAU) був протестований з трьома різними методами стиснення: Resize, AVG та AVG1. Цей вейвлет характеризується високою гладкістю завдяки гаусівській огиноючій та є ефективним для аналізу гладких, повільно змінюваних сигналів. Візуально

можна побачити різницю між методами стиснення з цим вейвлетом на спектрограмах, що на рисунку Б.5.

Комбінація CGAU+Resize показала середню точність 0,68, що є значно нижче за результати більшості інших вейвлетів. Це вказує на те, що властивості CGAU не є оптимальними для аналізу транспортних шумів, які часто містять швидкі зміни та нестационарні компоненти. Аналіз по класах виявляє дуже неоднорідну продуктивність. Клас Background демонструє точність 0,93, що є хорошим результатом і свідчить про те, що CGAU ефективно відрізняє гладкі фонові шуми від структурованих транспортних звуків.

Клас Train показує точність 0,96, що також є відмінним результатом. Однак для більшості інших класів точність значно нижча. Клас Cars демонструє точність 0,87, клас Truck – 0,66, клас Bus – 0,70. Найнижча точність спостерігається для класу Motorcycles з значенням 0,32, що практично не краще за випадкове відгадування. Це значне зниження точності для мотоциклів може бути пов'язано з тим, що звуки мотоциклів містять різкі імпульсні компоненти та швидкі зміни частоти, які погано представляються гладким гаусівським вейвлетом. Клас Airplane також демонструє низьку точність 0,52, що вказує на труднощі CGAU в аналізі складних широкосмугових сигналів авіаційного транспорту. Комбінація CGAU+AVG показала середню точність 0,72, що є покращенням порівняно з CGAU+Resize (0,68). Це свідчить про те, що метод усереднення блоків може трохи компенсувати недоліки CGAU через виділення періодичних структур у сигналі. Аналіз по класах виявляє подібні патерни з покращенням для більшості класів. Клас Background зберігає високу точність 0,93, клас Train демонструє точність 0,94. Клас Cars показує точність 0,88, що є прийнятним результатом. Однак проблемні класи залишаються проблемними і з методом AVG. Клас Motorcycles демонструє точність 0,47, що є покращенням порівняно з Resize (0,32), але все ще дуже низьким. Клас Helicopter показує точність 0,68, клас Bus – лише 0,48. Ці результати підтверджують, що фундаментальні властивості CGAU не є оптимальними для транспортних шумів з виразною динамікою та швидкими змінами. Комбінація CGAU+AVG1, де

використовувався модифікований метод усереднення з меншими блоками, показала середню точність 0,78, що є значним покращенням порівняно з CGAU+AVG (0,72) та CGAU+Resize (0,68). Це цікавий результат, що свідчить про важливість вибору розміру блоку для усереднення. Менші блоки в методі AVG1 дозволяють краще зберегти короточасні динамічні характеристики сигналу, що компенсує недостатню часову роздільність CGAU. Аналіз по класах для CGAU+AVG1 виявляє покращення для всіх класів. Клас Background демонструє точність 0,94, клас Train – 0,96, що залишаються на високому рівні. Клас Cars показує точність 0,87, клас Bicycle – 0,83. Найбільше покращення спостерігається для раніше проблемних класів: Motorcycles досягає точності 0,59 (порівняно з 0,32 для Resize), Bus – 0,66 (порівняно з 0,70 для Resize), Helicopter – 0,72 (порівняно з 0,55 для Resize). Ці результати показують, що комбінація вейвлета з методом стиснення має критичне значення для загальної продуктивності системи. Навіть вейвлет з не оптимальними властивостями може дати прийнятні результати при правильному виборі методу стиснення, і навпаки, хороший вейвлет може показати погані результати з неадекватним методом стиснення.

### 3.3.6 Результати для вейвлета Shannon

Вейвлет Shannon, що характеризується ідеальною прямокутною частотною характеристикою, був протестований з методами Resize, AVG та AVG1. Цей вейвлет теоретично забезпечує найкращу можливу частотну локалізацію, але має повільне спадання у часовій області, що може створювати артефакти на краях сигналу. Візуально можна побачити різницю між методами стиснення з цим вейвлетом на спектрограмах, що на рисунку Б.6.

Комбінація SHAN+Resize показала середню точність 0,85, що є дуже хорошим результатом і близьким до найкращих комбінацій у дослідженні. Це підтверджує теоретичні переваги Shannon вейвлета для частотного аналізу.

Точне розділення частотних компонент без витоку енергії між смугами дозволяє нейронній мережі ефективно виявляти характерні частотні сигнатури різних транспортних засобів.

Аналіз по класах виявляє стабільно високу продуктивність. Клас Train демонструє видатну точність 0,98, що є одним з найвищих результатів у всьому дослідженні. Клас Background – 0,94, Cars – 0,91, Bicycle – 0,88. Клас Airplane демонструє точність 0,81, що є хорошим результатом для цього складного класу. Класи Helicopter (0,83), Bus (0,73) та Truck (0,82) показують прийнятну точність. Навіть клас Motorcycles досягає точності 0,76, що краще за результати багатьох інших комбінацій.

Комбінація SHAN+AVG показала середню точність 0,74, що є нижче за SHAN+Resize (0,85). Це несподіваний результат, оскільки для інших вейвлетів метод AVG зазвичай давав результати, порівнянні або кращі за Resize. Зниження продуктивності може бути пов'язано з тим, що повільне спадання Shannon вейвлета у часовій області створює проблеми при розбитті на блоки та усередненні. Крайові ефекти від кожного блоку можуть накопичуватися та спотворювати результуюче усереднене представлення.

Аналіз по класах для SHAN+AVG виявляє значне зниження точності для більшості класів порівняно з SHAN+Resize. Клас Train зберігає високу точність 0,97, клас Background – 0,93. Однак клас Cars показує точність 0,89, клас Airplane – 0,71, клас Bus – лише 0,55. Клас Motorcycles демонструє точність 0,45, що є значно гірше за результат з Resize (0,76). Клас Truck показує точність 0,62, що також є зниженням порівняно з 0,82 для Resize.

Комбінація SHAN+AVG1 показала середню точність 0,86, що є покращенням порівняно з SHAN+AVG (0,74) і навіть трохи краще за SHAN+Resize (0,85). Це підтверджує, що менші блоки в методі AVG1 краще справляються з артефактами від повільного спадання Shannon вейвлета. Менший розмір блоку означає менші крайові ефекти, що призводить до більш чистого усередненого представлення.

Аналіз по класах для SHAN+AVG1 виявляє покращення для всіх класів порівняно з SHAN+AVG. Клас Train демонструє точність 0,97, клас Background – 0,94, клас Cars – 0,92. Клас Airplane показує точність 0,84, що є значним покращенням порівняно з SHAN+AVG (0,71). Клас Motorcycles досягає точності 0,75, що наближається до результату SHAN+Resize (0,76). Класи Bus (0,75), Helicopter (0,86) та Truck (0,84) також демонструють хорошу точність.

### 3.3.7 Результати з використанням архітектури EfficientNetB0

Для оцінки потенційних переваг використання сучасних архітектур загорткових нейромереж було проведено експерименти з EfficientNetB0 в комбінації з вейвлетом FBSP та методом AVG1. Вибір саме цієї комбінації був обумовлений тим, що FBSP+AVG показав одні з найкращих результатів серед власнорозроблених архітектур, а AVG1 показав кращі результати ніж AVG, і було цікаво перевірити, чи може ця архітектура поркащити результати.

Комбінація FBSP+EfficientNetB0+AVG1 показала середню точність 0,87, що є дуже хорошим результатом, але не краще за FBSP+AVG з власнорозробленою архітектурою (0,88).

Аналіз по класах виявляє цікаві відмінності в розподілі точності порівняно з власнорозробленою архітектурою. Клас Train демонструє видатну точність 0,99, що є найвищим результатом для цього класу в усьому дослідженні. Це свідчить про те, що складна архітектура EfficientNetB0 з механізмами Squeeze-and-Excitation особливо ефективна для виявлення характерних паттернів залізничного транспорту. Клас Background показує точність 0,96, клас Cars – 0,93, що також є відмінними результатами.

Помітне покращення спостерігається для класу Airplane з точністю 0,84 порівняно з 0,82 для базової архітектури FBSP+Resize. Клас Bicycle демонструє точність 0,90, клас Helicopter – 0,88. Клас Bus показує точність 0,80, що є

прийнятним результатом. Клас Motorcycles досягає точності 0,66, що є нижче за деякі інші комбінації, але все ще прийнятним. Клас Truck демонструє точність 0,89, що є одним з найкращих результатів для цього класу.

Цікаво відзначити, що EfficientNetB0 демонструє більш збалансовану продуктивність по класах порівняно з базовою архітектурою. Різниця між найкращим (Train – 0,99) та найгіршим (Motorcycles – 0,66) класом становить 0,33, що є більшим за деякі інші комбінації. Однак для більшості класів (7 з 9) точність перевищує 0,85, що свідчить про високу надійність класифікації.

Час навчання моделі з EfficientNetB0 був значно довшим порівняно з базовою архітектурою – приблизно 40-50 епох для повної збіжності порівняно з 25-35 епохами для базової архітектури. Це пов'язано з більшою складністю EfficientNetB0.

Враховуючи лише незначне погіршення точності (0,87 проти 0,88 для найкращих власнорозроблених архітектур) при значно більшій обчислювальній складності та часі навчання, використання EfficientNetB0 не виявилось істотно вигідним для даної задачі. Тобто, для специфічних типів даних, таких як спектрограми, використання продвинутої архітектури, що приймає на вхід кольорові зображення, не дає суттєвого покращення.

### 3.3.8 Аналіз обчислювальної ефективності вейвлет-перетворень та методів стиснення

Окрім точності класифікації, важливим аспектом практичного застосування розробленої системи є обчислювальна ефективність різних етапів обробки даних. Час виконання вейвлет-перетворення суттєво залежить від типу використаного вейвлета та методів стиснення. Для комплексної оцінки було проведено вимірювання часу виконання вейвлет-перетворень, нормалізації та методів стиснення. Результати наведено в таблиці 3.1.

Таблиця 3.1 – Час вейвлет-перетворень, нормалізації та стиснень

№	Wavelet	Wavelet time, ms	Norm time, ms	Resize time, ms	Crop time, ms	AVG time, ms	AVG1 time, ms	SVD time, ms	Statistic s time, ms
1	2	3	4	5	6	7	8	9	10
1	Morlet	82	10	689	0,001	1,4	12	35	240
2	CMOR	85	10	689	0,001	1,4	12	35	240
3	Mexh	30	10	689	0,001	1,4	12	35	240
4	FBSP	34	10	689	0,001	1,4	12	35	240
5	CGAU	28	10	689	0,001	1,4	12	35	240
6	SHAN	102	10	689	0,001	1,4	12	35	240

Найшвидшими серед використаних вейвлетів, виявилися вейвлети CGAU (28 мс) та Mexican Hat (30 мс), що пояснюється їх відносно простою математичною структурою. Вейвлет CGAU базується на похідних гаусівської функції, які обчислюються ефективно, а Mexican Hat є другою похідною гаусіана з компактним носієм, що дозволяє швидко виконувати згортку з сигналом.

Вейвлет FBSP демонструє час обчислення 34 мс, що також є дуже хорошим результатом. Це свідчить про те, що комбінація В-сплайнів з синусоїдальною модуляцією не створює значного обчислювального навантаження, незважаючи на більшу складність порівняно з простими вейвлетами. Ефективна реалізація В-сплайнів через рекурсивні формули дозволяє досягти високої швидкості обчислень.

Вейвлети Morlet та CMOR показують час виконання 82 мс та 85 мс відповідно, що приблизно втричі повільніше за найшвидші вейвлети. Найповільнішим виявився вейвлет Shannon з часом виконання 102 мс.

Важливо відзначити, що різниця в часі виконання вейвлет-перетворень (від 28 до 102 мс) є суттєвою з точки зору обробки великих обсягів даних. Однак для застосувань реального часу всі досліджені вейвлети є достатньо швидкими, оскільки навіть найповільніший Shannon дозволяє обробляти майже 10 аудіофрагментів на секунду, що значно перевищує швидкість надходження даних в типових сценаріях моніторингу транспорту.

Час нормалізації вейвлет-коефіцієнтів становить 10 мс і є однаковим, як і стиснення, для всіх типів вейвлетів, оскільки ця операція залежить тільки від розміру матриці коефіцієнтів, а не від природи вейвлета. Нормалізація включає обчислення мінімуму та максимуму матриці, віднімання та ділення, що виконуються ефективно завдяки векторизованим операціям.

Серед методів стиснення спостерігається ще більша варіативність часу виконання. Метод Crop є найшвидшим з часом виконання близьким до 0,001 мс, що практично не додає обчислювального навантаження, так як це операція індексування масиву без будь-яких обчислень над даними – вибирається підмножина стовпців з вихідної матриці.

Метод AVG демонструє час виконання 1,4 мс, що також є дуже швидким. Розбиття матриці на блоки та усереднення по осі блоків реалізується через операцію reshape та mean в NumPy, що оптимізовані на низькому рівні.

Метод AVG1 з меншими блоками вимагає 12 мс для виконання, що майже в 10 разів повільніше за базовий метод AVG через більшу кількість блоків для усереднення, але це все ще швидко.

Метод SVD виявився одним з найповільніших з часом виконання 35 мс. Незважаючи на оптимізовані реалізації SVD в бібліотеках лінійної алгебри, обчислення сингулярних значень та векторів залишається відносно дорогою операцією.

Метод Statistics показує час виконання 240 мс, що робить його найповільнішим серед усіх досліджених методів стиснення. Необхідність обчислення 12 різних статистик для кожного з 128 масштабів призводить до значного обчислювального навантаження.

Метод Resize з найгіршим часом виконання 689 мс використовує бікубічну інтерполяцію з бібліотеки scikit-image, яка включає згладжування анти-аліасингу для запобігання артефактам при зменшенні розміру. Ці додаткові операції роблять Resize найповільнішим методом з часом виконання 689 мс, що майже в 500 разів повільніше за метод AVG.

Для застосувань реального часу різниця в швидкості стиснення є критичною. Метод AVG дозволяє обробляти понад 700 аудіофрагментів на секунду, тоді як Resize обробляє лише 1,45 фрагменти на секунду. Це означає, що система на основі AVG може аналізувати безперервний аудіопотік практично без затримок, тоді як система на основі Resize буде мати значне відставання від реального часу.

### 3.3.9 Порівняльний аналіз та узагальнення результатів

Комплексний аналіз всіх проведених експериментів дозволяє виявити кілька ключових закономірностей та зробити практичні рекомендації щодо вибору оптимальних комбінацій вейвлетів та методів стиснення для розпізнавання транспортних шумів.

Для узагальнення отриманих результатів було створено відповідну таблицю 3.2.

Таблиця 3.2 – Результати експериментів

Вейвлет	Метод стиснення	Літак	Фон	Велосипед	Автобус	Автомобіль	Гелікоптер	Мотоцикл	Поїзд	Фура	Середнє
1	2	3	4	5	6	7	8	9	10	11	12
Morlet	Resize	0,75	0,94	0,83	0,77	0,92	0,8	0,73	0,95	0,86	0,86
Morlet	Statistics	0,42	0,54	0,46	0,41	0,51	0,33	0,47	0,38	0,32	0,43
Morlet	Crop	0,64	0,67	0,66	0,51	0,77	0,63	0,39	0,85	0,39	0,62
Morlet	Crop + Max pooling	0,43	0,72	0,58	0,53	0,7	0,71	0,62	0,9	0,27	0,6
Morlet	AVG	0,84	0,95	0,91	0,78	0,88	0,9	0,79	0,97	0,67	0,85
Mexh	AVG	0,87	0,97	0,91	0,67	0,92	0,7	0,69	0,98	0,82	0,83
CMOR	Resize	0,87	0,93	0,88	0,82	0,88	0,81	0,75	0,97	0,86	0,86
FBSP	Resize	0,82	0,96	0,92	0,85	0,93	0,88	0,77	0,97	0,86	0,88

Продовження таблиці 3.2

1	2	3	4	5	6	7	8	9	10	11	12
FBSP	Crop	0,75	0,88	0,87	0,62	0,86	0,81	0,6	0,94	0,73	0,78
FBSP	SVD	0,65	0,74	0,62	0,57	0,8	0,61	0,46	0,92	0,57	0,66
FBSP	AVG	0,87	0,93	0,92	0,77	0,93	0,88	0,75	0,97	0,87	0,88
FBSP + Efficient NetB0	AVG1	0,84	0,96	0,9	0,8	0,93	0,88	0,66	0,99	0,89	0,87
CGAU	Resize	0,52	0,93	0,58	0,7	0,87	0,55	0,32	0,96	0,66	0,68
CGAU	AVG	0,68	0,93	0,66	0,48	0,88	0,68	0,47	0,94	0,79	0,72
CGAU	AVG1	0,71	0,94	0,83	0,66	0,87	0,72	0,59	0,96	0,78	0,78
SHAN	Resize	0,81	0,94	0,88	0,73	0,91	0,83	0,76	0,98	0,82	0,85
SHAN	AVG	0,71	0,93	0,86	0,55	0,89	0,71	0,45	0,97	0,62	0,74
SHAN	AVG1	0,84	0,94	0,88	0,75	0,92	0,86	0,75	0,97	0,84	0,86

Найкращі загальні результати були досягнуті з вейвлетом FBSP в комбінації з методами Resize або AVG1, що дали середню точність 0,88. Це підтверджує теоретичні переваги FBSP – його гладкість завдяки B-сплайновій огинаючій забезпечує стійкість до шуму, а хороша частотна локалізація від синусоїдальної складової дозволяє ефективно виявляти характерні гармонічні структури транспортних шумів. Враховуючи також відносно швидкий час виконання вейвлет-перетворення FBSP (34 мс), ця комбінація є оптимальною для більшості практичних застосувань.

Вейвлет Morlet також показав дуже хороші результати, особливо в комбінації з методами Resize (0,86) та AVG (0,85). Це підтверджує його статус як універсального вейвлета для аналізу акустичних сигналів. Комплексна природа Morlet дозволяє отримувати інформацію як про амплітуду, так і про фазу, хоча в даному дослідженні використовувалася лише амплітудна інформація. Час виконання Morlet перетворення (82 мс) не самий довгий, але є помітно повільнішим за FBSP.

Вейвлет Shannon також продемонстрував конкурентоспроможні результати з середньою точністю 0,85-0,86 залежно від методу стиснення. Його ідеальна частотна локалізація виявилася корисною для точного розділення

частотних компонент транспортних шумів. Однак повільне спадання у часовій області створило проблеми з методом AVG, що призвело до зниження продуктивності для цієї комбінації. А також він є найповільнішим вейвлетом з часом виконання 102 мс, що робить його менш привабливим для застосувань з обмеженими обчислювальними ресурсами.

Вейвлет Mexican Hat показав добрі результати для класів з виразними імпульсними компонентами (Train – 0,98, Truck – 0,82), але гірші результати для класів з домінуючими гармонічними структурами (Helicopter – 0,70, Motorcycles – 0,69). Це підтверджує його спеціалізацію на виявленні локальних екстремумів та різких змін. Перевагою Mexican Hat є дуже швидке виконання (30 мс).

Вейвлет CMOR показав результати, подібні до базового Morlet, з середньою точністю 0,86 та відносно високим часом виконання (85 мс). Додаткові параметри CMOR не принесли значних переваг для даної задачі, що може вказувати на те, що типові налаштування вже є близькими до оптимальних, або що перевага від додаткової гнучкості компенсується складністю налаштування параметрів.

Вейвлет CGAU хоч і має найменший час виконання (28 мс), але показав найгірші загальні результати серед усіх досліджених вейвлетів з середньою точністю в діапазоні 0,68-0,78 залежно від методу стиснення. Його надмірна гладкість виявилася недоліком для аналізу динамічних транспортних шумів з швидкими змінами.

Серед методів стиснення найкращий баланс між точністю та швидкістю загалом продемонстрував метод AVG1, який забезпечив стабільно високу точність при дуже малому часі виконання (12 мс). Він перевершив базовий AVG для вейвлетів CGAU (0,78 проти 0,72) і SHAN (0,86 проти 0,74), а також досяг майже ідентичних результатів для FBSP (0,87 проти 0,88). Завдяки поєднанню високої якості відновлення та прийнятної швидкості обчислення AVG1 є оптимальним вибором у більшості випадках.

Метод AVG забезпечував високу швидкодію (1,4 мс) та хороші показники точності (0,83-0,88 залежно від вейвлета), що робить його ефективним для систем реального часу та обробки великих обсягів даних. Особливо хороші результати AVG показав у поєднанні з вейвлетами Morlet, Mexh та FBSP (0,85; 0,83; 0,88 відповідно). Однак для деяких вейвлетів, зокрема CGAU та SHAN, його точність виявилася нижчою, ніж у методу AVG1, що свідчить про чутливість AVG до втрати деталей через надмірне згладжування.

Метод Resize продемонстрував одну з найвищих точностей – на рівні 0,85-0,88 для Morlet, CMOR, FBSP та SHAN, проте його ключовим недоліком є вкрай великий час виконання (689 мс). Така обчислювальна витратність фактично унеможлиблює використання Resize у системах реального часу або при роботі з великими вибірками. Тому він може бути доцільним лише в тих випадках, коли максимальна точність має пріоритет над швидкістю та ресурсною ефективністю.

Метод Stop показав значно гірші результати з середньою точністю 0,62-0,78 залежно від вейвлета. Втрата інформації з початку та кінця сигналу виявилася критичною для багатьох класів, особливо тих, де важливі акустичні події розподілені по всій довжині фрагменту. Цей метод має найменший час виконання (0,001 мс), але низька точність робить його непридатним для практичного використання, може бути корисним тільки в специфічних ситуаціях, коли відомо де зосереджена найбільш інформативна частина сигналу.

Комбінований метод Stop+MaxPooling показав схожі або трохи гірші результати порівняно з простим Stop, що вказує на те, що агресивна підвибірка призводить до додаткової втрати інформації. Операція max pooling, хоча і зберігає найбільш виражені піки енергії, відкидає інформацію про їх точне розташування та взаємні співвідношення, що виявляється важливим для класифікації.

Метод Statistics показав найгірші результати серед усіх досліджених методів з середньою точністю лише 0,43 і значним часом виконання (240 мс).

Повна втрата інформації про часову структуру сигналу виявилася фатальною для задачі класифікації транспортних шумів і робить цей метод абсолютно непривабливим варіантом.

Метод SVD показав проміжні результати з середньою точністю близько 0,66 та часом виконання 35 мс. Хоча сингулярний розклад теоретично є оптимальною апроксимацією низького рангу в сенсі мінімізації похибки реконструкції, виявилось, що виділені ним компоненти не є оптимальними для задачі класифікації. Це підтверджує, що для машинного навчання важливо не мінімізувати похибку реконструкції, а максимізувати дискримінаційну здатність представлення.

Аналіз по окремих класах виявляє, що деякі класи є значно легшими для розпізнавання, ніж інші, незалежно від використаної комбінації вейвлета та методу стиснення. Клас Train стабільно демонструє найвищу точність (0,85-0,99 залежно від комбінації) завдяки унікальним характеристикам залізничного транспорту. Клас Background також легко розпізнається (0,72-0,97) через фундаментальні відмінності від структурованих транспортних звуків.

Клас Motorcycles виявився найскладнішим для розпізнавання з точністю в широкому діапазоні від 0,32 до 0,79 залежно від комбінації. Висока внутрішньокласова варіативність мотоциклів, від малокубатурних скутерів до потужних спортбайків, створює значні труднощі для навчання узагальненої моделі. Клас Truck також демонструє високу варіативність результатів (0,27-0,89), що може бути пов'язано з різноманітністю типів вантажних автомобілів.

Класи Cars, Bicycle, Bus, Helicopter та Airplane показують проміжні результати з помірною варіативністю залежно від комбінації вейвлета та методу стиснення. Для цих класів вибір оптимальної комбінації має значний вплив на кінцеву точність класифікації.

Аналіз confusion matrices виявляє типові патерни помилок. Найчастіші помилки виникають між подібними класами: Cars та Bus часто плутаються

через схожі характеристики двигунів; Motorcycle та Cars також можуть плутатися, особливо для малокубатурних мотоциклів; Airplane та Helicopter іноді класифікуються неправильно через спільні риси авіаційного транспорту.

Цікаво, що помилки класифікації не є симетричними. Наприклад, Motorcycle частіше помилково класифікується як Cars, ніж Cars як Motorcycle. Це може бути пов'язано з тим, що деякі мотоцикли (з великими двигунами) мають акустичні характеристики, ближчі до автомобілів, тоді як автомобілі рідко звучать як типові мотоцикли.

Загалом, результати дослідження демонструють, що поєднання вейвлет-перетворення з глибоким навчанням є дуже ефективним підходом для розпізнавання транспортних шумів. Досягнуті рівні точності 0,85-0,88 для найкращих комбінацій є достатніми для багатьох практичних застосувань, таких як моніторинг транспортних потоків, системи підрахунку транспортних засобів, аналіз акустичного забруднення міського середовища.

## ВИСНОВКИ

Таким чином, у кваліфікаційній роботі досліджено методи розпізнавання транспортних акустичних шумів з використанням вейвлет-перетворення та згорткових нейронних мереж, та вирішено такі завдання:

- проведено аналіз літературних джерел щодо методів розпізнавання акустичних сигналів та класифікації транспортних шумів, що дало можливість виявити сучасний стан дослідженої проблематики;

- проведено детальний аналіз шести типів вейвлет-перетворень (Morlet, Mexican Hat, CMOR, FBSP, CGAU, Shannon) та семи методів стиснення часово-частотних представлень (Resize, Statistics, Crop, Crop+MaxPooling, AVG, AVG1, SVD), що дало можливість встановити їх математичні властивості, переваги та недоліки для подальшого експериментального дослідження ефективності різних комбінацій;

- розроблено спеціалізовані архітектури згорткових нейронних мереж, адаптовані до специфіки різних методів стиснення часово-частотних представлень, що дало можливість ефективно обробляти вхідні дані різної розмірності та структури, від компактних статистичних дескрипторів ( $128 \times 12$ ) до детальних часово-частотних матриць ( $128 \times 1024$ );

- сформовано репрезентативний набір даних, що включає 9 класів транспортних засобів (літак, велосипед, автомобіль, автобус, фура, гелікоптер, мотоцикл, поїзд, фоновий шум) з 22500 аудіофрагментів загальною кількістю, що дозволило забезпечити достатню різноманітність акустичних умов для навчання та тестування глибоких нейронних мереж;

- впроваджено комплексну методологію попередньої обробки аудіосигналів, що включає нормалізацію амплітуди, застосування вейвлет-перетворення з 128 масштабами та нормалізацію вейвлет-коефіцієнтів, що дало можливість привести сигнали до стандартизованого вигляду та забезпечити стабільність навчання нейронних мереж;

– проведено всебічні експериментальні дослідження приблизно 20 комбінацій різних типів вейвлетів та методів стиснення з використанням єдиної методології навчання та тестування, що дозволило об'єктивно порівняти їх ефективність та виявити оптимальні підходи для розпізнавання транспортних шумів;

– здійснено порівняльний аналіз власнорозроблених архітектур згорткових нейронних мереж з готовою архітектурою EfficientNetB0, що дало можливість оцінити їх якість та відмінності один до одного;

– побудовано та проаналізовано матриці помилок для всіх досліджених комбінацій, що дозволило оцінити точності розпізнавання і виявити типові патерни помилок класифікації, встановити найбільш складні для розпізнавання класи транспортних засобів та зрозуміти причини плутанини між схожими класами.

– проаналізовано та порівняно часові характеристики обчислень для кожного вейвлет-перетворення, нормалізації, методу стиснення та архітектур нейромереж, що дало змогу оцінити їхню придатність для практичних застосувань і визначити методи з найменшою затримкою обробки.

У рамках кваліфікаційної роботи встановлено, що найвищу абсолютну точність (0,88) на тестових вибірках забезпечили комбінації з вейвлетом FBSP у поєднанні з методами Resize та AVG. Вейвлети Morlet, CMOR та Shannon також продемонстрували високу ефективність із середньою точністю 0,85-0,86.

Серед методів стиснення найкращий баланс між точністю та швидкістю продемонстрував метод AVG1. Він забезпечив високу точність, перевищивши базовий AVG, а також показав майже однакову ефективність для FBSP (0,87 проти 0,88 для AVG). При цьому його час виконання становив лише 12 мс, що є на три порядки швидше за Resize (689 мс). Загалом метод AVG також показав хороші результати точності (0,83-0,88) при надзвичайно високій швидкодії (1,4 мс), однак за рядом вейвлетів поступався у точності AVG1 через втрату деталей при надмірному усередненні. Метод Statistics показав найгірші результати (середня точність 0,43) через повну втрату інформації про часову

структуру сигналу, що підкреслює критичну важливість збереження динаміки зміни спектральних характеристик для успішної класифікації.

Аналіз результатів по окремих класах виявив, що клас Train (поїзд) є найлегшим для розпізнавання з точністю до 0,99, завдяки унікальним характеристикам залізничного транспорту. Клас Motorcycles (мотоцикли) виявився найскладнішим з точністю в діапазоні 0,32-0,79 залежно від комбінації, що пов'язано з високою внутрішньокласовою варіативністю акустичних характеристик різних типів мотоциклів.

Використання готової однієї з найкращих передових архітектур загорткових мереж EfficientNetB0 – не показало покращення точності (0,87), а навчання було дуже довгим, тож в даному дослідженні, порівняно з найкращими власнорозробленими архітектурами (0,88) вона не є ефективним рішенням для навчання на скалограмах, і можливо потребує додаткових досліджень з різними варіаціями вхідних даних для покращення точності розпізнавання.

Наукова новизна роботи полягає у комплексному порівняльному дослідженні ефективності різних типів вейвлет-перетворень в поєднанні з різними методами стиснення часово-частотних представлень для задачі класифікації транспортних акустичних шумів, що дозволило виявити оптимальні комбінації параметрів обробки та встановити критичну важливість збереження часово-частотної структури сигналу для успішного розпізнавання, а також в розробці власної високоефективної архітектури нейроної мережі. Розроблено та експериментально перевірено спеціалізовані архітектури згорткових нейронних мереж, адаптовані до специфіки різних методів стиснення, що забезпечують ефективну обробку вхідних даних різної розмірності та структури.

Результати роботи апробовано у вигляді 2 тез доповідей під час ІХ Міжнародної студентської наукової конференції «Діджиталізація науки як виклик сьогодення» [23], VIII Всеукраїнської студентської наукової конференції «Формування сучасної науки: методика та практика» [24].

**ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ**

1. Rhif, M., Ben Abbes, A., Farah, I. R., Martínez, B., & Sang, Y. (2019). Wavelet transform application for/in non-stationary time-series analysis: A review. *Applied sciences*, 9(7), 1345.
2. Shabbir, A., Cheema, A. N., Ullah, I., Almanjahie, I. M., & Alshahrani, F. (2024). Smart city traffic management: Acoustic-based vehicle detection using stacking-based ensemble deep learning approach. *IEEE access*, 12, 35947-35956.
3. Tvoroshenko, I., & Dziubenko, M. (2020). Modern methods of analysis of the movement scheme using video detection of vehicles.
4. Daradkeh, Y. I., Gorokhovatskyi, V., Tvoroshenko, I., & Zeghid, M. (2022). Tools for fast metric data search in structural methods for image classification. *IEEE Access*, 10, 124738-124746.
5. Gorokhovatskyi, V., Tvoroshenko, I., Yakovleva, O., Hudáková, M., & Gorokhovatskyi, O. (2024). Application a committee of Kohonen neural networks to training of image classifier based on description of descriptors set. *IEEE Access*, 12, 73376-73385.
6. Pomazan, V., Tvoroshenko, I., & Gorokhovatskyi, V. (2023). Development of an application for recognizing emotions using convolutional neural networks.
7. Gorokhovatskyi, V., Chmutov, Y., Tvoroshenko, I., & Kobylin, O. (2025). Reducing computational costs by compressing the structural description in image classification methods. *Advanced Information Systems*, 9(1), 5-12.
8. Wang, H., Zou, Y., & Chong, D. (2020). Acoustic scene classification with spectrogram processing strategies. *arXiv preprint arXiv:2007.03781*.
9. Zamorano, M., Gómez, M. J., Castejon, C., & Carboni, M. (2025). Analysis of Acoustic Emission Waveforms by Wavelet Packet Transform for the Detection of Crack Initiation Due to Fretting Fatigue in Solid Railway Axles. *Applied Sciences*, 15(15), 8435.

10. Zhang, Z., Xu, S., Zhang, S., Qiao, T., & Cao, S. (2021). Attention based convolutional recurrent neural network for environmental sound classification. *Neurocomputing*, *453*, 896-903.
11. Gorokhovatskyi, V., & Tvoroshenko, I. (2024). Transforming image descriptions as a set of descriptors to construct classification features.
12. Remolina Soto, M. S., Amaya Guzmán, B., Aya-Parra, P. A., Perdomo, O. J., Becerra-Fernandez, M., & Sarmiento-Rojas, J. (2025). Intelligent Classification of Urban Noise Sources Using TinyML: Towards Efficient Noise Management in Smart Cities. *Sensors*, *25*(20), 6361.
13. Fernández-Gallego, M. P., & Toledano, D. T. (2021). A study of data augmentation for increased ASR robustness against packet losses. In *IberSPEECH*.
14. YAKOVLEVA, Olena, et al. Face Detection for Video Surveillance-based Security System. In: *COLINS* (3). 2023. p. 69-86.
15. Shin, S., Kim, J., Yu, Y., Lee, S., & Lee, K. (2021). Self-supervised transfer learning from natural images for sound classification. *Applied Sciences*, *11*(7), 3043.
16. Phan, D. T., Huynh, T. A., Pham, V. T., Tran, C. M., Mai, V. T., & Tran, N. Q. (2025). Optimal Scalogram for Computational Complexity Reduction in Acoustic Recognition Using Deep Learning. *arXiv preprint arXiv:2505.13017*.
17. Tvoroshenko, I. S., & Gorokhovatskyi, V. O. (2019). Intelligent classification of biophysical system states using fuzzy interval logic. *Telecommunications and Radio Engineering*, *78*(14).
18. Gorokhovatskyi, V., Tvoroshenko, I., Yakovleva, O., & Hudáková, M. (2025). Image description compression in classification structural methods. *IEEE Access*, *13*, 43631-43641.
19. Lyashenko, V., Kobylín, O., & Shafronenko, A. (2019, September). Wavelet analysis and decomposition into color spaces in researching of human fluorescently labeled images tissues. In *2019 IEEE 8th International Conference on Advanced Optoelectronics and Lasers (CAOL)* (pp. 618-621). IEEE.

20. Gorokhovatskyi, V., Tvoroshenko, I., Kobylin, O., & Vlasenko, N. (2023). Search for visual objects by request in the form of a cluster representation for the structural image description. *Advances in Electrical and Electronic Engineering*, 21(1), 19.

21. Dogan, Y. (2023). A new global pooling method for deep neural networks: Global average of top-k max-pooling. *Traitement du signal*, 40(2), 577-587.

22. Кобилін, О. А., & Творошенко, І. С. (2021). Методи цифрової обробки зображень.

23. Зубков, А. (2025) Дослідження розпізнавання акустичних шумів транспортних засобів з використанням вейвлет-перетворення та нейронних мереж. У IX Міжнародній студентській науковій конференції «Діджиталізація науки як виклик сьогодення», Жовтень 3, 2025, Харків. С. 131-133.

24. Зубков, А. (2025) Наука і технології розпізнавання транспортних шумів: вейвлет-перетворення, стиснення та нейронні мережі. У VIII Всеукраїнській студентській науковій конференції «Формування сучасної науки: методика та практика», Жовтень 17, 2025, Львів. С. 255-257.