

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет прикладних інформаційних технологій та електроінженерії
(повна назва факультету)

Кафедра радіотехнічних систем
(повна назва кафедри)

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня

магістр

(назва освітнього ступеня)

на тему: Метод інтелектуальної класифікації літаючих об'єктів за їх
акустичними сигнатурами

Виконав: студент 6 курсу, групи РАм-61
спеціальності 172 Електронні комунікації та
радіотехніка
(шифр і назва спеціальності)

Ковальчук М.Ю.
(підпис) (прізвище та ініціали)

Керівник Паляниця Ю.Б.
(підпис) (прізвище та ініціали)

Нормоконтроль Хвостівська Л.В.
(підпис) (прізвище та ініціали)

Завідувач кафедри Дунець В.Л.
(підпис) (прізвище та ініціали)

Рецензент Дедів Л.Є.
(підпис) (прізвище та ініціали)

Тернопіль
2025

Міністерство освіти і науки України
Тернопільський національний технічний університет імені Івана Пулюя

Факультет прикладних інформаційних технологій та електроінженерії
(повна назва факультету)

Кафедра радіотехнічних систем
(повна назва кафедри)

ЗАТВЕРДЖУЮ

Завідувач кафедри

Дунець В.Л.
(підпис) (прізвище та ініціали)

«26» листопада 2025 р.

**ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ**

на здобуття освітнього ступеня Магістр
(назва освітнього ступеня)

за спеціальністю 172 Електронні комунікації та радіотехніка
(шифр і назва спеціальності)

Студенту Ковальчуку Миколі Юрійовичу
(прізвище, ім'я, по батькові)

1. Тема роботи Метод інтелектуальної класифікації літаючих об'єктів за їх акустичними
сигнатурами

Керівник роботи Паляниця Юрій Богданович, к.т.н., доцент кафедри РТ
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом ректора від «26» листопада 2025 року № 4/7-1038

2. Термін подання студентом завершеної роботи 26 грудня 2025р.

3. Вихідні дані до роботи Наукові публікації, технічна документація акустичних систем та
акустичні параметри БПЛА для моделювання процесів виявлення.

4. Зміст роботи (перелік питань, які потрібно розробити)

Вступ

Розділ 1. Аналітична частина

Розділ 2. Основна частина

Розділ 3. Науково-дослідна частина

Розділ 4. Охорона праці та безпека в надзвичайних ситуаціях

Загальні висновки

Перелік посилань

Додатки

5. Перелік графічного матеріалу (з точним зазначенням обов'язкових креслень, слайдів)

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
Охорона праці	к.т.н., зав. кафедри МТ Окіпний І.Б.		
Безпека в надзвичайних ситуаціях	ст. викл. кафедри ОХ Стручок В.С.		

7. Дата видачі завдання 26 листопада 2025 р.

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1.	Ознайомлення з завданням до кваліфікаційної роботи	26.11.2025	Виконано
2.	Підбір джерел по темі кваліфікаційної роботи	29.11.2025 – 01.12.2025	Виконано
3.	Переклад та опрацювання джерел по темі кваліфікаційної роботи	01.12.2025 – 03.12.2025	Виконано
4.	Виконання дослідження по темі кваліфікаційної роботи	03.12.2025 – 05.12.2025	Виконано
5.	Оформлення розділу «Аналітична частина»	05.12.2025 – 07.12.2025	Виконано
6.	Оформлення розділу «Основна частина»	07.12.2025 – 09.12.2025	Виконано
7.	Оформлення розділу «Науково-дослідна частина»	09.12.2025 – 11.12.2025	Виконано
8.	Оформлення розділу «Охорона праці та безпека в надзвичайних ситуаціях»	11.12.2025	Виконано
9.	Підготовка графічного матеріалу	11.12.2025 – 12.12.2025	Виконано
10.	Оформлення кваліфікаційної роботи	11.12.2025 – 12.12.2025	Виконано
11.	Нормоконтроль	12.12.2025	Виконано
12.	Перевірка на плагіат	12.12.2025	Виконано
13.	Попередній захист кваліфікаційної роботи	12.12.2025	Виконано
14.	Захист кваліфікаційної роботи	26.12.2025	

Студент

(підпис)

Ковальчук М.Ю.

(прізвище та ініціали)

Керівник роботи

(підпис)

Паляниця Ю.Б.

(прізвище та ініціали)

Метод інтелектуальної класифікації літаючих об'єктів за їх акустичними сигнатурами // Кваліфікаційна робота освітнього рівня «Магістр» // Ковальчук Микола Юрійович // Тернопільський національний технічний університет імені Івана Пулюя, факультет прикладних інформаційних технологій та електроінженерії, кафедра радіотехнічних систем, група РАМ-61 // Тернопіль, 2025 // С. 95, рис. – 9, табл. – 0, додат. – 1, бібліогр. – 27.

Ключові слова: КЛАСИФІКАЦІЯ, БПЛА, АКУСТИЧНА СИГНАТУРА, MFCC, СПЕКТРОГРАМА, ЗГОРТКОВА НЕЙРОННА МЕРЕЖА, CNN, ГЛИБИННЕ НАВЧАННЯ.

Кваліфікаційна робота присвячена підвищенню ефективності класифікації літаючих об'єктів шляхом розробки методу інтелектуального аналізу їх акустичних сигнатур із застосуванням згорткових нейронних мереж.

У першому розділі проаналізовано сучасні методи моніторингу повітряного простору, зокрема радіолокаційні, оптичні та акустичні. Визначено переваги пасивних акустичних систем для виявлення малих БПЛА на низьких висотах. Досліджено особливості акустичних сигнатур різних типів літаючих об'єктів (літаків, вертольотів, дронів) та обґрунтовано доцільність використання методів машинного навчання для їх розпізнавання.

У другому розділі досліджено теоретичні основи цифрової обробки звукових сигналів. Розглянуто методи попередньої обробки, нормалізації та виділення інформативних ознак, таких як MFCC, LPC та Chroma. Проведено порівняльний аналіз класичних алгоритмів машинного навчання (k-NN, SVM) та методів глибинного навчання (CNN, RNN, Transformers), визначивши згорткові мережі як найбільш перспективні для аналізу спектрограм.

У третьому розділі розроблено та досліджено дві архітектури згорткових нейронних мереж ("smallNet" та "complexNet") для класифікації сигнатур. Виконано навчання моделей на основі MFCC-спектрограм із застосуванням аугментації даних. Експериментально встановлено, що компактна архітектура

"smallNet" забезпечує вищу точність та кращу узагальнюючу здатність порівняно з глибшою моделлю, яка виявила схильність до перенавчання.

У четвертому розділі окреслено вимоги охорони праці та техніки безпеки при виконанні робіт із застосуванням БПЛА, а також заходи безпеки у надзвичайних ситуаціях.

ANNOTATION

Method of intelligent classification of flying objects by their acoustic signatures // Qualification work of the educational level «Master» // Kovalchuk Mykola // Ternopil Ivan Puluj National Technical University, Faculty of Applied Information Technologies and Electrical Engineering, Radio Engineering Systems Department, group RAM-61 // Ternopil, 2025 // P. 95, fig. – 9, tabl. – 0, annexes. – 1, references – 27.

Keywords: CLASSIFICATION, UAV, ACOUSTIC SIGNATURE, MFCC, SPECTROGRAM, CONVOLUTIONAL NEURAL NETWORK, CNN, DEEP LEARNING.

The qualification work is devoted to improving the efficiency of flying object classification by developing a method for the intelligent analysis of their acoustic signatures using convolutional neural networks.

The first section analyzes modern methods of airspace monitoring, including radar, optical, and acoustic methods. The advantages of passive acoustic systems for detecting small UAVs at low altitudes are determined. The features of acoustic signatures of various types of flying objects (airplanes, helicopters, drones) are investigated, and the feasibility of using machine learning methods for their recognition is substantiated.

The second section investigates the theoretical foundations of digital sound signal processing. Methods of pre-processing, normalization, and extraction of informative features, such as MFCC, LPC, and Chroma, are considered. A comparative analysis of classical machine learning algorithms (k-NN, SVM) and deep learning methods (CNN, RNN, Transformers) is conducted, identifying convolutional networks as the most promising for spectrogram analysis.

In the third section, two architectures of convolutional neural networks ("smallNet" and "complexNet") for signature classification are developed and

investigated. Model training was performed based on MFCC spectrograms using data augmentation. It was experimentally established that the compact "smallNet" architecture provides higher accuracy and better generalization capability compared to the deeper model, which showed a tendency towards overfitting.

The fourth section outlines occupational health and safety requirements when performing work involving UAVs, as well as safety measures in emergency situations.

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ

AUC (англ. Area Under Curve) – Площа під кривою (показник якості класифікації).

CNN (англ. Convolutional Neural Network) – Згорткова нейронна мережа.

ComplexNet – Архітектура глибокої нейронної мережі зі значною кількістю шарів.

DFT (англ. Discrete Fourier Transform) – Дискретне перетворення Фур'є.

EPOCH – Епоха навчання (один повний прохід набору даних через нейронну мережу).

F1-score – Гармонійне середнє між точністю та повнотою.

FFT (англ. Fast Fourier Transform) – Швидке перетворення Фур'є.

LPC (англ. Linear Predictive Coding) – Лінійне передбачуване кодування.

MFCC (англ. Mel-Frequency Cepstral Coefficients) – Мел-кепстральні коефіцієнти.

Overfitting – Перенавчання моделі.

Precision – Точність (метрика оцінки класифікатора).

Recall – Повнота (метрика оцінки класифікатора).

ReLU (англ. Rectified Linear Unit) – Функція активації.

ROC (англ. Receiver Operating Characteristic) – Робоча характеристика приймача.

SmallNet – Компактна архітектура нейронної мережі з оптимізованою кількістю параметрів.

STFT (англ. Short-Time Fourier Transform) – Короткочасне перетворення Фур'є.

БПЛА – Безпілотний літальний апарат.

МН – Машинне навчання.

ІНМ – Штучна нейронна мережа.

ЗМІСТ

ВСТУП	10
РОЗДІЛ 1. АНАЛІТИЧНА ЧАСТИНА	12
1.1. Моніторинг повітряного простору	12
1.1.1. Сучасний стан проблеми детектування та класифікації літаючих об'єктів	12
1.1.2. Аналіз існуючих методів ідентифікації літаючих об'єктів	13
1.2. Особливості акустичних сигнатур різних типів літаючих об'єктів	15
1.2.1. Засоби виявлення БПЛА у різних умовах довкілля	18
1.3. Висновок до першого розділу	19
РОЗДІЛ 2. ОСНОВНА ЧАСТИНА	21
2.1. Теоретичні основи розпізнавання звуку	21
2.1.1. Поняття звукового сигналу та його характеристики	21
2.1.2. Попередня обробка сигналів (нормалізація, фільтрація, шумозаглушення)	24
2.1.3. Методи виділення ознак (часові, частотні, спектральні)	28
2.2. Традиційні методи розпізнавання звукових паттернів	34
2.2.1. Методи на основі статистичного аналізу	34
2.2.2. Методи на основі спектрального аналізу	39
2.2.3. Метод динамічної трансформації часової шкали (DTW)	46
2.2.4. Приховані Марковські моделі (HMM)	47
2.3. Класичні алгоритми машинного навчання	49
2.4. Методи розпізнавання на основі глибинного навчання	54
2.4.1. Штучні нейронні мережі для обробки звуку	54
2.4.2. Сучасні архітектури	60
2.4.3. Порівняння ефективності глибинних методів із класичними	63
2.5. Висновок до другого розділу	66
РОЗДІЛ 3. НАУКОВО-ДОСЛІДНА ЧАСТИНА	67
3.1. Використання ознак MFCC, LPC, Chroma як вхідних даних	67

3.2.	Проектування архітектур нейронних мереж.....	72
3.3.	Налаштування параметрів навчання та аугментації вхідних спектрограм.....	76
3.4.	Порівняльний аналіз ефективності моделей	78
3.5.	Оцінка достовірності детектування класу “Drone” за допомогою матриць помилок та ROC-аналізу	83
3.6.	Висновок до третього розділу	86
РОЗДІЛ 4. ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ		87
4.1.	Законодавча база з охорони праці при розробці та експлуатації систем класифікації літаючих об’єктів	87
4.2.	Гігієнічні норми та санітарні вимоги до організації робочих місць....	89
4.3.	Висновок до четвертого розділу.....	90
ЗАГАЛЬНІ ВИСНОВКИ		91
ПЕРЕЛІК ПОСИЛАНЬ.....		92
ДОДАТКИ		

ВСТУП

Актуальність теми. Впродовж останніх років безпілотні літальні апарати (БПЛА) активно впроваджуються в цивільні та військові сфери, трансформуючись із спеціалізованих платформ у масово доступні засоби. Зростання доступності БПЛА супроводжується підвищенням рівня загроз, включаючи несанкціоноване відеоспостереження, розвідку та підготовку диверсійних дій. Низька візуальна помітність та мала ефективна площа розсіювання ускладнюють виявлення дронів традиційними радіолокаційними засобами. У зв'язку з цим актуальною є розробка пасивних систем виявлення, зокрема акустичних, які дозволяють класифікувати об'єкти за їх унікальними звуковими сигнатурами навіть в умовах відсутності прямої видимості. Застосування методів машинного навчання до акустичних даних відкриває нові можливості для автоматизації та підвищення точності таких систем.

Мета і задачі дослідження. Метою даної кваліфікаційної роботи освітнього рівня "Магістр" є підвищення ефективності класифікації літаючих об'єктів шляхом розробки методу інтелектуального аналізу їх акустичних сигнатур із використанням згорткових нейронних мереж (CNN).

Для досягнення поставленої мети потрібно виконати ряд завдань, зокрема:

- Проаналізувати сучасний стан проблеми детектування БПЛА та особливості їх акустичних сигнатур.
- Дослідити методи попередньої обробки звукових сигналів та виділення ознак, зокрема мел-кепстральних коефіцієнтів (MFCC).
- Спроекувати архітектури згорткових нейронних мереж (smallNet та complexNet) для класифікації спектрограм.
- Провести експериментальне навчання моделей із застосуванням аугментації даних та порівняти їх ефективність.
- Оцінити точність та надійність розробленого методу за допомогою ROC-аналізу та матриць помилок.

Об'єкт дослідження. Процес класифікації типів літаючих об'єктів за даними акустичного моніторингу.

Предмет дослідження. Методи та алгоритми інтелектуальної обробки акустичних сигналів із використанням ознак MFCC та згорткових нейронних мереж.

Методи дослідження. У роботі використовуються методи цифрової обробки сигналів для розрахунку спектрограм та MFCC, методи машинного навчання для побудови та тренування нейронних мереж, методи статистичного аналізу для оцінки результатів класифікації (AUC, F1-score), а також методи аугментації даних для покращення узагальнювальних властивостей моделей.

Практичне значення одержаних результатів. Розроблений метод класифікації акустичних сигнатур може бути використаний як основа для побудови пасивних систем охорони периметра та моніторингу критично важливих об'єктів. Запропонована компактна архітектура нейромережі (smallNet) забезпечує високу точність при помірних обчислювальних витратах, що дозволяє впроваджувати її в системи реального часу.

Наукова новизна отриманих результатів. Наукова новизна роботи полягає у порівняльному аналізі ефективності різних архітектур згорткових нейронних мереж для задачі розпізнавання дронів за спектрограмами MFCC. Експериментально доведено, що для класифікації акустичних сигнатур БПЛА компактні архітектури забезпечують кращу стабільність та меншу схильність до перенавчання порівняно з глибокими мережами при обмеженому наборі даних.

Публікації. Викладені в роботі результати доповідалися та обговорювалися на XIV Міжнародній науково-технічній конференції молодих учених та студентів "Актуальні задачі сучасних технологій" (м. Тернопіль, 11-12 грудня 2025 року).

РОЗДІЛ 1.

АНАЛІТИЧНА ЧАСТИНА

1.1. Моніторинг повітряного простору

1.1.1. Сучасний стан проблеми детектування та класифікації літаючих об'єктів

Розвиток технологій повітряного моніторингу за останні десятиліття суттєво трансформувався, особливо у сфері безпеки та контролю за несанкціонованими польотами. Сучасні системи поєднують традиційні методи спостереження з інноваційними підходами на основі штучного інтелекту.

Актуальність проблеми зростає через масове поширення БПЛА, що становлять загрозу для інфраструктури та безпеки. За даними ІКАО, кількість інцидентів із дронами щороку збільшується на 20-30%. Це створює нові виклики для систем, які раніше орієнтувалися на великі об'єкти [1].

Основні засоби детектування – радіолокаційні комплекси, оптико-електронні системи та акустичні датчики. Кожен має переваги й обмеження, тому ефективність досягається їх комплексним застосуванням. У міських умовах радіолокація часто ускладнена перешкодами.

У військових конфліктах особливо важливо класифікувати об'єкти, адже БПЛА активно використовуються для розвідки та ударів. Тому ідентифікація типу та небезпеки є критичною для оборони.

Зростає інтерес до пасивних систем [2], зокрема акустичних методів, які дозволяють виявляти об'єкти поза прямою видимістю. Акустичний "підпис" дає змогу визначати тип і режим роботи апарата. Сучасні дослідження зосереджені на алгоритмах обробки сигналів (спектральний аналіз, вейвлети, глибоке навчання), що підвищують точність класифікації. Водночас залишаються проблеми адаптації до шумів, погодних умов та багатократних об'єктів.

1.1.2. Аналіз існуючих методів ідентифікації літаючих об'єктів

Сучасні технології ідентифікації літаючих об'єктів охоплюють широкий спектр методів, кожен з яких має свої переваги та обмеження. В залежності від фізичної природи сигналів, що використовуються для детектування, ці методи можна поділити на три основні категорії: радіолокаційні, оптичні та акустичні.

Кожен з цих підходів характеризується специфічними технічними рішеннями, застосовуваними для виявлення, локалізації та класифікації повітряних цілей. Далі детально розглядаються особливості кожного класу методів [3].

Радіолокаційні методи. Радіолокаційні методи [4] - найпоширеніші та історично перші технології виявлення повітряних цілей. Вони працюють на принципі випромінювання електромагнітних хвиль і прийому відбитих сигналів. Сучасні системи поділяються на активні, пасивні та бістатичні.

Активні радары забезпечують дальність до 300-400 км і незалежність від погодних умов, дозволяють визначати координати та швидкість цілі. Недоліки – висока енергозатратність, демаскуюче випромінювання та складність виявлення малих об'єктів із композитних матеріалів.

Вторинні радары працюють із транспондерами, надаючи додаткову інформацію про об'єкт. Вони ефективні в цивільній авіації, але безсилі проти несанкціонованих чи ворожих апаратів без транспондерів.

SAR-системи забезпечують високу роздільну здатність і розпізнавання за радіолокаційним "портретом", проте потребують значних обчислювальних ресурсів.

Попри розвиток, радары мають обмеження у виявленні малих БПЛА, особливо в міських умовах. Ефективна площа розсіювання дронів може бути меншою за 0,01 м², що ускладнює їх ідентифікацію. Крім того, вони вразливі до перешкод і можуть бути виявлені засобами радіотехнічної розвідки.

Оптичні методи ідентифікації літаючих об'єктів базуються на аналізі випромінювання у видимому, інфрачервоному та ультрафіолетовому діапазонах.

Використовуються камери високої роздільної здатності, тепловізори, лазерні далекоміри та лідари.

- **Відеоспостереження** у видимому діапазоні дає можливість візуально розпізнавати об'єкти на відстані до 10-15 км, але ефективність падає при тумані, опадах чи засліпленні сонцем.

- **Інфрачервоні системи** фіксують теплове випромінювання, особливо ефективні вночі та для об'єктів із двигунами. Менш результативні щодо планерів чи електричних БПЛА.

- **Лідари** забезпечують точне визначення координат і 3D-зображення об'єкта, проте мають обмежену дальність і чутливі до атмосферних опадів.

Недоліки оптичних методів: залежність від погодних умов і часу доби, складність забезпечення кругового огляду та висока вартість систем. Попри це, вони залишаються ключовими для точної ідентифікації та визначення характеристик цілей.

Акустичні методи ідентифікації ґрунтуються на аналізі звукових хвиль, що виникають від двигунів, гвинтів, турбін та аеродинамічних шумів. Кожен літальний апарат має власну акустичну сигнатуру, яка відображає його конструкцію та режим роботи [5].

Сучасні системи використовують мережі мікрофонів, спектральний аналіз і програмне забезпечення для обробки сигналів. Завдяки тріангуляції та beamforming можна точно визначати напрямок на ціль, а мікрофонні решітки дозволяють фільтрувати шум.

Головна перевага – пасивність: акустичні датчики не випромінюють енергію, залишаються скритними й енергоефективними. Вони особливо корисні для виявлення малих БПЛА на низьких висотах, де радари часто малоефективні.

Обмеження полягають у невеликій дальності (до 10-15 км для літаків і лише сотні метрів для дронів) та чутливості до погодних умов і фонового шуму.

Попри це, розвиток цифрової обробки сигналів і машинного навчання дозволяє не лише виявляти, а й класифікувати об'єкти, визначати режим роботи

двигунів і прогнозувати траєкторію. Перспективним є поєднання акустичних датчиків з іншими сенсорами у мультисенсорних системах.

Вартість таких систем нижча за радіолокаційні комплекси, що робить їх доступними для захисту критичної інфраструктури, масових заходів і приватних територій, а також інтеграції в існуючі системи безпеки.

1.2. Особливості акустичних сигнатур різних типів літаючих об'єктів

Акустичні сигнатури – це комплексні звукові характеристики літаючих об'єктів, що формуються двигунами, гвинтами та аеродинамічними шумами. Кожен тип апарата має унікальний акустичний профіль, який залежить від типу силової установки, частоти обертання пропелерів та швидкості польоту.

Спектральний аналіз показує характерні частотні діапазони та патерни, що утворюють "акустичний підпис". Часова структура сигналу (модуляції, імпульси, флуктуації) відображає режими роботи двигунів і рухомих частин. Сигнатури змінюються залежно від режиму польоту та атмосферних умов, що ускладнює класифікацію [6].

Для ефективного розпізнавання формуються багатовимірні вектори ознак (частотні, часові, кепстральні, вейвлет-характеристики), які обробляються методами машинного навчання. Це забезпечує високу точність навіть у зашумлених умовах.

Необхідним є створення баз даних акустичних записів у різних режимах польоту для навчання алгоритмів. Комплексний аналіз таких даних відкриває можливості для розвитку систем моніторингу, що доповнюють радіолокаційні та оптичні засоби.

Акустичні сигнатури літаків – це унікальні звукові профілі, що формуються силовими установками, аеродинамічними ефектами та структурним шумом. Вони є інформативним джерелом для ідентифікації та класифікації повітряних суден.

Реактивні та турбовентиляторні двигуни створюють широкосмуговий спектр (200 Гц – 5 кГц) із дискретними компонентами, що відповідають частотам обертання компресорів і вентиляторів. Для турбовентиляторів характерні тональні складові (1,5-4 кГц) та шум від турбулентного змішування струменя. Турбогвинтові літаки мають низькочастотні гармоніки (50-500 Гц), пов'язані з роботою гвинтів, а поршневі – дискретні компоненти на низьких і середніх частотах (20-200 Гц) та імпульсні сигнали, зумовлені циклічною роботою двигуна.

Аеродинамічні джерела шуму створюють широкосмуговий спектр (500 Гц – 10 кГц), інтенсивність якого різко зростає зі швидкістю польоту та при випуску механізації крила. Часова структура сигналів також містить діагностичні ознаки: стабільність амплітуди у реактивних літаків, амплітудну модуляцію у гвинтових, а також ефект Доплера, що дозволяє оцінювати швидкість і напрямок руху.

Спрямованість шуму залежить від типу двигуна: реактивні випромінюють енергію переважно назад уздовж струменя, тоді як гвинтові мають складнішу діаграму з максимумами в площині обертання. Характеристики змінюються залежно від режиму польоту: при зльоті домінують високі частоти, у крейсерському режимі спектр стабільний, а при посадці зростає аеродинамічний шум. Висота польоту також впливає на спектр через поглинання високих частот атмосферою та рефракцію хвиль.

Акустичні сигнатури вертольотів мають унікальні характеристики, що відрізняють їх від інших літальних апаратів. Основними джерелами шуму є несучий і рульовий гвинти та силова установка.

Найбільш характерним є лопатевий шум несучого гвинта - низькочастотні імпульси (10-50 Гц) з гармоніками до 500-700 Гц, що формують "гребінчастий" спектр. Унікальним маркером є феномен биття лопатей, який виникає при маневрах і швидкому польоті, створюючи різкі періодичні хлопки.

Рульовий гвинт додає високочастотні компоненти (70-150 Гц), а двигуни – широкосмуговий шум у діапазоні 500 Гц – 5 кГц із тональними складовими компресора та турбіни.

Акустичний профіль змінюється залежно від режиму польоту: при висінні домінує лопатевий шум, при горизонтальному русі зростає вклад двигуна й трансмісії, а при наборі висоти чи зниженні з'являються додаткові модуляції. Спрямованість шуму також різниться: несучий гвинт випромінює максимум у площині обертання, рульовий - перпендикулярно до своєї площини.

Різні типи вертольотів мають відмінні спектрально-часові характеристики: важкі (Mi-26, CH-47) – потужний низькочастотний шум; середні (Mi-8, UH-60) – збалансований спектр; легкі (Robinson R44, EC120) – зміщення в бік середніх і високих частот. Співвісні машини (Ka-52, Ka-226) формують особливо складні сигнатури через взаємодію потоків двох гвинтів.

Для класифікації використовують спектральні, кепстральні, бі-спектральні та часові характеристики. Сучасні системи з алгоритмами машинного навчання забезпечують точність понад 90% навіть у шумних міських умовах.

Акустичні сигнатури БПЛА. Дрони формують характерні акустичні сигнатури завдяки шуму електродвигунів, пропелерів та взаємодії повітряного потоку з конструкцією. Основна гармоніка відповідає швидкості обертання пропелерів, а додаткові виникають через аеродинамічні ефекти. На відміну від вертольотів із низькочастотним домінуванням, дрони мають щільні високочастотні компоненти [7].

Квадрокоптери створюють специфічний малюнок сигналу через інтерференцію кількох пропелерів. Сигнатура змінюється залежно від режиму польоту - набір висоти, стабілізація чи маневрування. Її потужність невелика, що ускладнює детектування на великих відстанях, але стабільність частотної структури дозволяє застосовувати спектральний аналіз і машинне навчання для класифікації.

На сигнал впливають умови середовища: вологість, вітер, рельєф, міські відбиття. Для виділення корисної інформації потрібні адаптивні методи

фільтрації. Часово-частотний аналіз (вейвлети, STFT) допомагає розрізняти типи та навіть моделі дронів. Динамічні моделі сигнатур враховують зміни під час маневрів і тяги.

Дрони активно застосовуються у цивільних і військових сферах: від сільського господарства й кінематографу до логістики та пошуково-рятувальних операцій. Водночас вони становлять загрозу безпеці – від випадкових інцидентів до навмисних атак. Відомі випадки проникнення у захищені зони та аварії з літаками [8].

Існують чотири основні підходи: радарний, акустичний, візуальний та радіочастотний.

- **Радар** добре працює на великих відстанях, але має проблеми з малими БПЛА та високою вартістю.
- **Акустичні датчики** є рентабельними, ефективними вдень і вночі, здатні класифікувати стан дронів навіть із додатковим навантаженням, хоча дальність обмежена (~150 м).
- **РЧ-методи** дозволяють ідентифікувати дрони за сигналами контролера, використовуючи антени з високим підсиленням.
- **Візуальні системи** доступні й зрозумілі, але залежать від освітлення; тепловізори частково вирішують проблему.

Акустичні сигнатури дронів поєднують стабільні тональні компоненти, високочастотні гармоніки та залежність від конструкції й режимів польоту. Це відкриває можливості для застосування алгоритмів штучного інтелекту у виявленні та класифікації БПЛА, що є ключовим напрямом розвитку систем моніторингу повітряного простору.

1.2.1. Засоби виявлення БПЛА у різних умовах довкілля

Безпілотні літальні апарати (БПЛА) стрімко поширюються у цивільних і військових сферах. Їх використовують у сільському господарстві, кінематографії, логістиці, пошуково-рятувальних операціях та правоохоронній діяльності. Завдяки компактності й доступності дрони стали глобальним явищем, змінивши

авіаційну галузь і зробивши її безпечнішою та продуктивнішою. Сучасні моделі можуть бути настільки малими, що поміщаються в долоні, але водночас здатні виконувати складні завдання [9].

Дрони відкрили нові можливості для бізнесу й приватних користувачів, проте водночас створили загрози безпеці. Інциденти з проникненням у захищені зони, зіткнення з літаками чи зупинка роботи аеропортів показують, що навіть невеликі апарати можуть спричиняти значні проблеми. Крім випадкових аварій, існує ризик навмисного використання дронів для контрабанди, терористичних атак чи несанкціонованої зйомки [10].

Щоб мінімізувати ризики, потрібні технології, здатні виявляти, класифікувати й відстежувати дрони, а також придушувати їх і збирати докази порушень. Виявлення є першочерговим завданням, адже без нього неможливо побудувати ефективну профілактику. Особливо актуальним це стає у військово-політичному контексті, наприклад, на кордонах держав [11].

Роль штучного інтелекту останнім часом зростає все більше. Методи машинного й глибокого навчання значно підвищують точність розпізнавання, дозволяючи одночасно вирішувати завдання виявлення й класифікації. Це особливо важливо для малих БПЛА, які можуть проникати у стратегічні зони з небезпечним вантажем [12-15].

1.3. Висновок до першого розділу

Аналітичний огляд сучасних методів моніторингу повітряного простору показав, що проблема детектування та класифікації літаючих об'єктів, особливо малих БПЛА, є надзвичайно актуальною як у цивільному, так і у військовому контексті. Традиційні засоби як радіолокаційні, оптичні та акустичні системи мають власні переваги й обмеження, тому найбільш ефективним є їх комплексне застосування.

Радари забезпечують значну дальність і незалежність від погодних умов, але малоефективні для дронів із низькою ефективною площею розсіювання.

Оптичні системи дають високу точність ідентифікації, проте залежать від освітлення та атмосферних умов. Акустичні методи, завдяки пасивності та рентабельності, особливо перспективні для виявлення малих БПЛА на низьких висотах, хоча їхня дальність обмежена.

Акустичні сигнатури літаків, вертольотів і дронів мають унікальні характеристики, що дозволяють їх класифікувати за допомогою спектрального та часово-частотного аналізу. Використання алгоритмів машинного й глибокого навчання значно підвищує точність розпізнавання навіть у зашумлених умовах.

Таким чином, сучасні системи моніторингу повітряного простору повинні базуватися на мультисенсорному підході, який поєднує радіолокаційні, оптичні та акустичні засоби з інтелектуальними алгоритмами обробки даних. Це забезпечує не лише виявлення, а й класифікацію об'єктів, що є ключовим для підвищення рівня безпеки та ефективного контролю повітряного простору.

РОЗДІЛ 2.

ОСНОВНА ЧАСТИНА

2.1. Теоретичні основи розпізнавання звуку

2.1.1. Поняття звукового сигналу та його характеристики

Звуковий сигнал являє собою механічну хвилю, що поширюється у пружному середовищі та характеризується коливаннями тиску або зміщення частинок середовища відносно їх положення рівноваги. У контексті цифрової обробки звуковий сигнал представляється як функція часу, що описує амплітуду коливань у кожен момент [16].

Математично дискретний звуковий сигнал можна представити у вигляді послідовності відліків:

$$x[n] = x(nT_s), \quad n = 0, 1, 2, \dots, N - 1 \quad (2.1)$$

де $T_s = 1/f_s$ є періодом дискретизації, f_s – частота дискретизації у герцах, N – загальна кількість відліків сигналу. Відповідно до теореми Найквіста-Шеннона, для коректного відтворення аналогового сигналу частота дискретизації повинна задовольняти умову:

$$f_s \geq 2f_{max} \quad (2.2)$$

де f_{max} є максимальною частотою у спектрі вихідного аналогового сигналу. Для аудіосигналів людського мовлення зазвичай використовується частота дискретизації 16 кГц або 22.05 кГц, тоді як для високоякісного аудіо застосовується стандарт 44.1 кГц або 48 кГц.

Основними характеристиками звукового сигналу є амплітуда, частота, фаза та тривалість. Амплітуда визначає інтенсивність звуку і вимірюється у відносних одиницях або децибелах. Рівень звукового тиску у децибелах обчислюється за формулою:

$$L_p = 20 \log_{10} \left(\frac{p}{p_0} \right) \quad (2.3)$$

де p – ефективне значення звукового тиску, $p_0 = 2 \times 10^{-5}$ Па – пороговий тиск, що відповідає порогу чутності людського вуха на частоті 1 кГц.

Миттєва енергія сигналу у момент часу n визначається як квадрат амплітуди:

$$E[n] = x^2[n] \quad (2.4)$$

а середня енергія на інтервалі з N відліків обчислюється як:

$$E_{avg} = \frac{1}{N} \sum_{n=0}^{N-1} x^2[n] \quad (2.5)$$

Потужність сигналу тісно пов'язана з його енергією та визначається як середнє значення квадрата амплітуди:

$$P = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} x^2[n] \quad (2.6)$$

Для практичних застосувань часто використовується коефіцієнт середньоквадратичного значення (Root Mean Square, RMS):

$$x_{RMS} = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} x^2[n]} \quad (2.7)$$

Частотна характеристика звуку визначає його тембр та висоту тону. Для складного звукового сигналу можна виділити основну частоту (фундаментальну частоту) f_0 та серію гармонік з частотами kf_0 , де $k = 2, 3, 4, \dots$. Співвідношення амплітуд різних гармонік формує характерний тембр звуку.

Спектральна густина потужності (Power Spectral Density, PSD) описує розподіл енергії сигналу по частотних компонентах:

$$S_{xx}(f) = \lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E}[|X_T(f)|^2] \quad (2.8)$$

де $X_T(f)$ – перетворення Фур'є сигналу на інтервалі T , $\mathbb{E}[\cdot]$ – оператор математичного сподівання.

Важливою характеристикою звукових сигналів є їх нестационарність. На відміну від стаціонарних процесів, статистичні властивості яких не змінюються з часом, звукові сигнали демонструють суттєву часову мінливість. Для аналізу нестационарних сигналів застосовується концепція квазістаціонарності, згідно з якою сигнал розглядається як послідовність коротких сегментів (зазвичай 20-40 мс), всередині яких статистичні властивості можна вважати приблизно постійними.

Часова структура звукового сигналу характеризується огибаючою амплітуди, яка описує повільні зміни інтенсивності звуку. Огибаюча $A[n]$ може бути обчислена за допомогою перетворення Гільберта:

$$A[n] = |x[n] + j\mathcal{H}\{x[n]\}| \quad (2.9)$$

де $\mathcal{H}\{\cdot\}$ позначає перетворення Гільберта, що визначається як:

$$\mathcal{H}\{x[n]\} = x[n] * \frac{1}{\pi n} \quad (2.10)$$

Для характеристики динамічного діапазону сигналу використовується відношення сигнал/шум (Signal-to-Noise Ratio, SNR):

$$SNR = 10 \log_{10} \left(\frac{P_{signal}}{P_{noise}} \right) \quad (2.11)$$

де P_{signal} та P_{noise} – потужності корисного сигналу та шуму відповідно. Високе значення SNR вказує на домінування корисного сигналу над шумовою компонентою, що є критичним для ефективного розпізнавання звукових патернів.

2.1.2. Попередня обробка сигналів (нормалізація, фільтрація, шумозаглушення)

Попередня обробка звукових сигналів є критично важливим етапом у системах розпізнавання, оскільки вона дозволяє підвищити відношення сигнал/шум, усунути небажані артефакти та привести дані до уніфікованого формату, придатного для подальшого аналізу.

Нормалізація амплітуди є базовою операцією попередньої обробки, що забезпечує приведення сигналу до стандартного динамічного діапазону. Найпростішим методом є лінійна нормалізація, при якій всі відліки сигналу масштабуються таким чином, щоб максимальна амплітуда відповідала заданому значенню:

$$x_{norm}[n] = A_{target} \cdot \frac{x[n]}{\max_n |x[n]|} \quad (2.12)$$

де A_{target} – цільова амплітуда (зазвичай 1.0 або 0.9 для запобігання обрізання). Альтернативним підходом є нормалізація за середньоквадратичним значенням:

$$x_{norm}[n] = \frac{x[n]}{x_{RMS}} \cdot RMS_{target} \quad (2.13)$$

де RMS_{target} – бажане середньоквадратичне значення. Цей метод забезпечує більш стабільну нормалізацію для сигналів з великими імпульсними викидами.

Z-нормалізація (стандартизація) приводить сигнал до нульового середнього та одиничної дисперсії:

$$x_{norm}[n] = \frac{x[n] - \mu_x}{\sigma_x} \quad (2.14)$$

де $\mu_x = \frac{1}{N} \sum_{n=0}^{N-1} x[n]$ – середнє значення, $\sigma_x = \sqrt{\frac{1}{N} \sum_{n=0}^{N-1} (x[n] - \mu_x)^2}$ – середньоквадратичне відхилення.

Фільтрація звукових сигналів здійснюється з метою виділення корисних частотних компонент та придушення небажаних складових. Лінійні цифрові фільтри описуються різницеvim рівнянням:

$$y[n] = \sum_{k=0}^M b_k x[n-k] - \sum_{k=1}^N a_k y[n-k] \quad (2.15)$$

де b_k та a_k – коефіцієнти фільтра, M та N визначають порядок фільтра. Передавальна функція такого фільтра у z -області має вигляд:

$$H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^M b_k z^{-k}}{1 + \sum_{k=1}^N a_k z^{-k}} \quad (2.16)$$

Фільтри нижніх частот (ФНЧ) використовуються для придушення високочастотного шуму. Частотна характеристика ідеального ФНЧ визначається як:

$$H_{LP}(f) = \begin{cases} 1, & |f| \leq f_c \\ 0, & |f| > f_c \end{cases} \quad (2.17)$$

де f_c – частота зрізу. Для практичної реалізації застосовуються фільтри Баттерворта, Чебишева або еліптичні фільтри з кінцевою крутизною спаду частотної характеристики.

Фільтр Баттерворта n -го порядку має квадрат модуля частотної характеристики:

$$|H(\omega)|^2 = \frac{1}{1 + \left(\frac{\omega}{\omega_c}\right)^{2n}} \quad (2.18)$$

де $\omega_c = 2\pi f_c$ – кругова частота зрізу, n – порядок фільтра.

Смугові фільтри виділяють певний діапазон частот та описуються характеристикою:

$$H_{BP}(f) = \begin{cases} 1, & f_1 \leq |f| \leq f_2 \\ 0, & \text{інакше} \end{cases} \quad (2.19)$$

де f_1 та f_2 – нижня та верхня межі смуги пропускання. Для аналізу мовних сигналів часто застосовується смуговий фільтр з діапазоном 300-3400 Гц, що відповідає основному частотному діапазону людського мовлення.

Операція фільтрації у часовій області еквівалентна операції згортки:

$$y[n] = x[n] * h[n] = \sum_{k=-\infty}^{\infty} x[k]h[n-k] \quad (2.20)$$

де $h[n]$ – імпульсна характеристика фільтра.

Шумозаглушення (noise reduction) є складнішою задачею, що вимагає оцінки шумової компоненти та її придушення при збереженні корисного сигналу. Метод спектрального віднімання базується на припущенні про адитивність шуму:

$$x[n] = s[n] + d[n] \quad (2.21)$$

де $s[n]$ – корисний сигнал, $d[n]$ – шумова компонента. У частотній області оцінка очищеного сигналу обчислюється як:

$$\hat{S}(k) = X(k) - \alpha \hat{D}(k) \quad (2.22)$$

де $X(k)$ та $\hat{D}(k)$ – спектри зашумленого сигналу та оцінка спектра шуму, $\alpha \geq 1$ - параметр надлишкового віднімання. Оскільки результат може містити від'ємні значення, застосовується операція напівхвильового випрямлення:

$$|\hat{S}(k)| = \max(|X(k)| - \alpha |\hat{D}(k)|, \beta |X(k)|) \quad (2.23)$$

де β – параметр спектрального порогу (зазвичай $\beta \approx 0.1$).

Вінерівська фільтрація забезпечує оптимальне за критерієм мінімуму середньоквадратичної похибки відновлення сигналу. Передавальна функція фільтра Вінера визначається як:

$$H_W(k) = \frac{S_{ss}(k)}{S_{ss}(k) + S_{dd}(k)} = \frac{1}{1 + \frac{S_{dd}(k)}{S_{ss}(k)}} \quad (2.24)$$

де $S_{ss}(k)$ та $S_{dd}(k)$ – спектральні густини потужності корисного сигналу та шуму відповідно. На практиці ці величини оцінюються з доступних даних:

$$H_W(k) = \frac{1}{1 + \frac{1}{SNR(k)}} \quad (2.25)$$

де $SNR(k) = \frac{|S(k)|^2}{|\hat{D}(k)|^2}$ – апостеріорне відношення сигнал/шум у частотній смузі k .

Адаптивні методи шумозаглушення використовують рекурсивні алгоритми для динамічної оцінки параметрів шуму. Алгоритм найменших середньоквадратичних (Least Mean Squares, LMS) оновлює коефіцієнти адаптивного фільтра за правилом:

$$\mathbf{w}[n + 1] = \mathbf{w}[n] + \mu e[n] \mathbf{x}[n] \quad (2.26)$$

де $\mathbf{w}[n]$ – вектор коефіцієнтів фільтра, $e[n] = d[n] - \mathbf{w}^T[n] \mathbf{x}[n]$ – помилка оцінки, μ – крок адаптації, $\mathbf{x}[n]$ – вектор вхідного сигналу.

Ще одним важливим методом є видалення постійної складової (DC offset removal), що реалізується за допомогою високочастотного фільтра першого порядку:

$$y[n] = x[n] - x[n - 1] + \alpha y[n - 1] \quad (2.27)$$

де $\alpha \approx 0.995$ для частоти дискретизації 16 кГц. Ця операція усуває низькочастотний дрейф сигналу, що може виникати через недосконалість апаратури запису.

Попередня обробка також включає виявлення та видалення тиші (Voice Activity Detection, VAD). Найпростіший підхід базується на пороговій обробці енергії сигналу:

$$VAD[n] = \begin{cases} 1, & E[n] > \theta_E \\ 0, & E[n] \leq \theta_E \end{cases} \quad (2.28)$$

де $E[n]$ – короткочасна енергія, θ_E – поріг енергії. Більш складні методи використовують комбінацію енергетичних та спектральних ознак для підвищення надійності детектування.

2.1.3. Методи виділення ознак (часові, частотні, спектральні)

Виділення ознак є ключовим етапом розпізнавання звукових патернів, оскільки воно дозволяє трансформувати вихідний сигнал у компактний набір дескрипторів, що несуть інформацію про характерні властивості звуку. Ознаки можуть бути поділені на кілька категорій залежно від домену їх обчислення та фізичного змісту.

Часові ознаки обчислюються безпосередньо з відліків сигналу у часовій області. Короткочасна енергія вже була визначена раніше, але для практичного застосування її обчислюють на ковзних вікнах:

$$E[m] = \sum_{n=0}^{N-1} w[n] x^2[m+n] \quad (2.29)$$

де m – індекс фрейму, $w[n]$ – віконна функція (зазвичай вікно Хеммінга). Нормалізована логарифмічна енергія часто використовується як більш стабільна ознака:

$$\log E[m] = \log \left(\sum_{n=0}^{N-1} w[n] x^2[m+n] \right) \quad (2.30)$$

Коефіцієнт переходів через нуль (Zero-Crossing Rate, ZCR) підраховує кількість змін знаку сигналу на одиницю часу:

$$ZCR[m] = \frac{1}{2N} \sum_{n=0}^{N-1} | \operatorname{sgn}(x[m+n]) - \operatorname{sgn}(x[m+n-1]) | \quad (2.31)$$

де $sgn(\cdot)$ – функція знаку. Висока частота переходів через нуль характерна для шумоподібних та високочастотних звуків, тоді як для низькочастотних тональних сигналів вона є низькою.

Автокореляційна функція (АКФ) описує схожість сигналу з його зміщеною копією:

$$R_{xx}[k, m] = \sum_{n=0}^{N-1} w[n]x[m+n]x[m+n+k] \quad (2.32)$$

де k – величина зміщення (лаг). Нормалізована АКФ визначається як:

$$\rho[k, m] = \frac{R_{xx}[k, m]}{R_{xx}[0, m]} \quad (2.33)$$

Максимум автокореляційної функції при $k > 0$ вказує на період основного тону для вокалізованих звуків.

Частотні та спектральні ознаки отримуються шляхом перетворення сигналу у частотну область. Дискретне перетворення Фур'є (ДПФ) обчислюється за формулою:

$$X[k, m] = \sum_{n=0}^{N-1} w[n]x[m+n]e^{-j\frac{2\pi kn}{N}}, \quad k = 0, 1, \dots, N-1 \quad (2.34)$$

Модуль спектра $|X[k, m]|$ характеризує розподіл енергії по частотах, а фаза $\angle X[k, m]$ зазвичай не використовується для розпізнавання, оскільки є чутливою до часових зміщень.

Спектрограма являє собою двовимірне представлення еволюції спектра в часі:

$$S[k, m] = |X[k, m]|^2 \quad (2.35)$$

Логарифмічна спектрограма краще відповідає логарифмічному сприйняттю інтенсивності людським слухом:

$$S_{log}[k, m] = 10 \log_{10}(|X[k, m]|^2 + \epsilon) \quad (2.36)$$

де ϵ – мала константа для запобігання логарифмування нуля.

Спектральний центроїд характеризує “центр ваги” спектра та корелює з тембровою яскравістю звуку:

$$SC[m] = \frac{\sum_{k=0}^{N/2} f[k] |X[k, m]|}{\sum_{k=0}^{N/2} |X[k, m]|} \quad (2.37)$$

де $f[k] = \frac{k f_s}{N}$ – частота k -го бін спектра.

Спектральна дисперсія вимірює розкид спектра відносно центроїду:

$$SD[m] = \sqrt{\frac{\sum_{k=0}^{N/2} (f[k] - SC[m])^2 |X[k, m]|}{\sum_{k=0}^{N/2} |X[k, m]|}} \quad (2.38)$$

Спектральний потік (Spectral Flux) вимірює швидкість зміни спектра між сусідніми фреймами:

$$SF[m] = \sum_{k=0}^{N/2} (|X[k, m]| - |X[k, m-1]|)^2 \quad (2.39)$$

Спектральна ентропія характеризує рівномірність розподілу енергії по частотах:

$$SE[m] = - \sum_{k=0}^{N/2} p[k, m] \log_2 p[k, m] \quad (2.40)$$

де $p[k, m] = \frac{|X[k, m]|^2}{\sum_{j=0}^{N/2} |X[j, m]|^2}$ - нормалізований спектр потужності.

Мел-кепстральні коефіцієнти (Mel-Frequency Cepstral Coefficients, MFCC) є одними з найбільш популярних ознак для розпізнавання звуку, оскільки вони моделюють нелінійне сприйняття частоти людським слухом. Мел-шкала

визначає нелінійне відображення фізичної частоти f (у Гц) у перцептивну частоту m (у мелах):

$$m = 2595 \log_{10} \left(1 + \frac{f}{700} \right) \quad (2.41)$$

Зворотнє перетворення має вигляд:

$$f = 700(10^{m/2595} - 1) \quad (2.42)$$

Процес обчислення MFCC включає кілька послідовних кроків. Спочатку розраховується спектр потужності за допомогою ДПФ. Потім застосовується банк трикутних фільтрів, рівномірно розташованих на мел-шкалі:

$$H_i(k) = \begin{cases} 0, & f[k] < f[i-1] \\ \frac{f[k] - f[i-1]}{f[i] - f[i-1]}, & f[i-1] \leq f[k] < f[i] \\ \frac{f[i+1] - f[k]}{f[i+1] - f[i]}, & f[i] \leq f[k] < f[i+1] \\ 0, & f[k] \geq f[i+1] \end{cases} \quad (2.43)$$

де $i = 1, 2, \dots, M$ – індекс фільтра, M – кількість фільтрів (зазвичай 20-40), $f[i]$ – центральні частоти фільтрів на мел-шкалі.

Енергія у кожній мел-смузі обчислюється як:

$$E_i[m] = \sum_{k=0}^{N-1} |X[k, m]|^2 H_i(k) \quad (2.44)$$

Логарифмування енергій моделює логарифмічне сприйняття інтенсивності:

$$\log E_i[m] = \log(E_i[m] + \epsilon) \quad (2.45)$$

Нарешті, застосовується дискретне косинусне перетворення (ДКП) для декореляції ознак:

$$MFCC_j[m] = \sum_{i=1}^M \log E_i[m] \cos\left(\frac{\pi j(i - 0.5)}{M}\right), \quad j = 1, 2, \dots, C \quad (2.46)$$

де C – кількість мел-кепстральних коефіцієнтів (зазвичай 12-13). Нульовий коефіцієнт $MFCC_0[m]$ відповідає загальній енергії сигналу і часто виключається або обробляється окремо.

Кепстр (cepstrum) визначається як зворотне перетворення Фур'є від логарифму спектра:

$$c[n] = \mathcal{F}^{-1}\{\log|X[k]|\} \quad (2.47)$$

Термін “кепстр” утворений перестановкою складів у слові “spectrum” і означає “спектр спектра”. Кепстральний аналіз дозволяє розділити спектральну огинаючу (повільно змінну частину) та тонку структуру (швидко осцилюючу частину).

Для підвищення дискримінативної здатності ознак часто використовуються їх похідні. Дельта-коефіцієнти (перші похідні) обчислюються як:

$$\Delta MFCC_j[m] = \frac{\sum_{\tau=-T}^T \tau \cdot MFCC_j[m + \tau]}{2 \sum_{\tau=1}^T \tau^2} \quad (2.48)$$

де T – півширина вікна диференціювання (зазвичай $T = 2$). Дельта-дельта коефіцієнти (другі похідні) обчислюються аналогічно від дельта-коефіцієнтів:

$$\Delta \Delta MFCC_j[m] = \frac{\sum_{\tau=-T}^T \tau \cdot \Delta MFCC_j[m + \tau]}{2 \sum_{\tau=1}^T \tau^2} \quad (2.49)$$

Коефіцієнти лінійного передбачення (Linear Predictive Coding, LPC) є альтернативною системою ознак, що моделює голосовий тракт як авторегресивний процес:

$$\hat{x}[n] = - \sum_{k=1}^p a_k x[n-k] \quad (2.50)$$

де p – порядок моделі (зазвичай 10-16 для мовлення), a_k – коефіцієнти LPC. Вони знаходяться з умови мінімізації середньоквадратичної помилки передбачення:

$$E = \sum_n (x[n] - \hat{x}[n])^2 \quad (2.51)$$

Система лінійних рівнянь для знаходження коефіцієнтів LPC отримується з автокореляційної функції:

$$\sum_{k=1}^p a_k R_{xx}[|i-k|] = -R_{xx}[i], \quad i = 1, 2, \dots, p \quad (2.52)$$

Ефективний алгоритм розв'язання цієї системи - алгоритм Левінсона-Дарбіна. Кепстральні коефіцієнти можуть бути обчислені з LPC-коефіцієнтів за рекурсивною формулою:

$$c[n] = a_n + \sum_{k=1}^{n-1} \frac{k}{n} c[k] a_{n-k}, \quad 1 \leq n \leq p \quad (2.53)$$

Хроматичні ознаки (Chroma features) представляють енергію сигналу, розподілену по 12 класах висоти тону (напівтонах хроматичної гами). Хроматична ознака обчислюється як сума енергій усіх частотних компонент, що відповідають певній ноті:

$$Chroma_c[m] = \sum_{k: f[k] \in c} |X[k, m]|^2, \quad c \in \{C, C\#, D, \dots, B\} \quad (2.54)$$

де приналежність частоти до класу висоти визначається за формулою:

$$c(f) = 12 \cdot \log_2 \left(\frac{f}{f_{ref}} \right) \bmod 12 \quad (2.55)$$

де f_{ref} – еталонна частота (зазвичай 440 Гц для ноти А4).

Комбінація різних типів ознак дозволяє створити багатий дескриптор звукового сигналу, що враховує часову динаміку, спектральну структуру та перцептивні властивості, забезпечуючи основу для ефективного розпізнавання звукових патернів.

2.2. Традиційні методи розпізнавання звукових патернів

2.2.1. Методи на основі статистичного аналізу

Кореляційний аналіз належить до фундаментальних статистичних методів виявлення схожості між сигналами та є основою багатьох класичних алгоритмів розпізнавання звуку. Основна ідея полягає у вимірюванні ступеня лінійної залежності між двома сигналами або між сигналом та еталонним шаблоном [17].

Взаємна кореляційна функція (ВКФ) між двома дискретними сигналами $x[n]$ та $y[n]$ визначається як:

$$R_{xy}[k] = \sum_{n=-\infty}^{\infty} x[n]y[n+k] \quad (2.56)$$

де k – величина зміщення (лаг). На практиці сигнали мають скінченну тривалість, тому сума обчислюється в межах області визначення обох сигналів. Нормалізована взаємна кореляція усуває залежність від амплітуд сигналів:

$$\rho_{xy}[k] = \frac{R_{xy}[k]}{\sqrt{R_{xx}[0]R_{yy}[0]}} = \frac{\sum_n x[n]y[n+k]}{\sqrt{\sum_n x^2[n] \sum_n y^2[n]}} \quad (2.57)$$

Значення $\rho_{xy}[k]$ належить інтервалу $[-1,1]$, де $\rho_{xy}[k] = 1$ вказує на повну позитивну кореляцію, $\rho_{xy}[k] = -1$ – на повну негативну кореляцію, а $\rho_{xy}[k] = 0$ – на відсутність лінійного зв'язку.

Для задачі розпізнавання звукового патерну із множини еталонів обчислюється кореляція вхідного сигналу $x[n]$ з кожним еталоном $s_i[n]$, $i = 1, 2, \dots, M$:

$$\rho_i[k] = \frac{\sum_{n=0}^{N-1} x[n] s_i[n+k]}{\sqrt{\sum_{n=0}^{N-1} x^2[n] \sum_{n=0}^{N-1} s_i^2[n]}} \quad (2.58)$$

Розпізнавання відбувається за критерієм максимальної кореляції:

$$i^* = \operatorname{argmax}_i \max_k \rho_i[k] \quad (2.59)$$

де i^* – індекс розпізнаного класу, а максимізація по k забезпечує інваріантність до часових зміщень.

Обчислювальна складність прямого обчислення кореляції становить $O(N^2)$ операцій, що може бути неприйнятним для сигналів великої тривалості. Швидке обчислення кореляції здійснюється через частотну область з використанням теореми про згортку:

$$R_{xy}[k] = \mathcal{F}^{-1}\{X^*(f)Y(f)\} \quad (2.60)$$

де $X(f)$ та $Y(f)$ – перетворення Фур'є сигналів, $X^*(f)$ – комплексно спряжене. Застосування швидкого перетворення Фур'є (ШПФ) знижує складність до $O(N \log N)$.

Для підвищення стійкості до шуму використовується зважена кореляція з акцентуванням низьких кепстральних коефіцієнтів:

$$\rho_w[k] = \frac{\sum_n w[n] x[n] y[n+k]}{\sqrt{\sum_n w[n] x^2[n] \sum_n w[n] y^2[n]}} \quad (2.61)$$

де $w[n]$ – вагова функція, наприклад експоненційна $w[n] = e^{-\alpha n}$ або гаусівська $w[n] = e^{-\frac{n^2}{2\sigma^2}}$.

Частотно-залежна кореляція застосовується для порівняння спектральних характеристик:

$$\rho_f[k] = \frac{\sum_{\omega} |X(\omega)| |Y(\omega)| e^{j\omega k}}{\sqrt{\sum_{\omega} |X(\omega)|^2 \sum_{\omega} |Y(\omega)|^2}} \quad (2.62)$$

Цей підхід є більш стійким до фазових спотворень та часових варіацій темпу.

Кореляційні детектори використовуються для виявлення наявності відомого сигналу у зашумленому потоці даних. Вирішальна статистика обчислюється як:

$$\Lambda = \max_k R_{xy}[k] \quad (2.63)$$

і порівнюється з порогом θ :

$$\text{Рішення} = \begin{cases} H_1: \text{сигнал присутній,} & \Lambda > \theta \\ H_0: \text{тільки шум,} & \Lambda \leq \theta \end{cases} \quad (2.64)$$

Оптимальний поріг визначається з балансу між ймовірністю правильного виявлення та ймовірністю хибної тривоги за критерієм Неймана-Пірсона.

Для сигналів змінної тривалості застосовується ковзна кореляція на перекривних вікнах:

$$\rho_m[k] = \frac{\sum_{n=0}^{L-1} w[n] x[m+n] y[m+n+k]}{\sqrt{\sum_{n=0}^{L-1} w[n] x^2[m+n] \sum_{n=0}^{L-1} w[n] y^2[m+n]}} \quad (2.65)$$

де m – індекс початку вікна, L – довжина вікна, $w[n]$ – віконна функція.

Автокореляційна функція є окремим випадком взаємної кореляції, коли сигнал порівнюється сам із собою зі зміщенням у часі. АКФ надає цінну інформацію про періодичність сигналу, його структуру та статистичні властивості.

Детермінована автокореляційна функція визначається як:

$$R_{xx}[k] = \sum_{n=0}^{N-1-k} x[n] x[n+k], \quad k \geq 0 \quad (2.66)$$

Для стаціонарних випадкових процесів використовується статистичне визначення:

$$R_{xx}[k] = \mathbb{E}[x[n]x[n+k]] \quad (2.67)$$

де $\mathbb{E}[\cdot]$ – оператор математичного сподівання. На практиці математичне сподівання замінюється часовим усередненням.

Важливі властивості автокореляційної функції включають симетрію $R_{xx}[k] = R_{xx}[-k]$, максимум при нульовому лагу $R_{xx}[0] \geq |R_{xx}[k]|$ для всіх k , та зв'язок з спектральною густиною потужності через теорему Вінера-Хінчина:

$$S_{xx}(f) = \mathcal{F}\{R_{xx}[k]\} = \sum_{k=-\infty}^{\infty} R_{xx}[k]e^{-j2\pi f k} \quad (2.68)$$

Зворотне перетворення дає:

$$R_{xx}[k] = \mathcal{F}^{-1}\{S_{xx}(f)\} = \int_{-\infty}^{\infty} S_{xx}(f)e^{j2\pi f k} df \quad (2.69)$$

Для виявлення основного тону вокалізованих звуків використовується пошук максимуму АКФ у заданому діапазоні лагів:

$$k_{pitch} = \operatorname{argmax}_{k_{min} \leq k \leq k_{max}} R_{xx}[k] \quad (2.70)$$

де діапазон $[k_{min}, k_{max}]$ відповідає очікуваному діапазону частот основного тону. Для людського мовлення типовий діапазон становить 80-400 Гц, що при частоті дискретизації 16 кГц відповідає лагам $k_{min} = 40$ та $k_{max} = 200$.

Частота основного тону обчислюється як:

$$f_0 = \frac{f_s}{k_{pitch}} \quad (2.71)$$

Надійність оцінки можна оцінити за допомогою нормалізованого значення автокореляції:

$$\rho_{pitch} = \frac{R_{xx}[k_{pitch}]}{R_{xx}[0]} \quad (2.72)$$

Високе значення ρ_{pitch} (близьке до 1) вказує на сильну періодичність сигналу.

Автокореляційний метод лінійного передбачення використовує АКФ для оцінки коефіцієнтів LPC. Рівняння Юла-Уокера мають вигляд:

$$\begin{bmatrix} R_{xx}[0] & R_{xx}[1] & \cdots & R_{xx}[p-1] \\ R_{xx}[1] & R_{xx}[0] & \cdots & R_{xx}[p-2] \\ \vdots & \vdots & \ddots & \vdots \\ R_{xx}[p-1] & R_{xx}[p-2] & \cdots & R_{xx}[0] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = - \begin{bmatrix} R_{xx}[1] \\ R_{xx}[2] \\ \vdots \\ R_{xx}[p] \end{bmatrix} \quad (2.73)$$

Ця система має теплицеву структуру і ефективно розв'язується алгоритмом Левінсона-Дарбіна зі складністю $O(p^2)$.

Для сегментації звукового сигналу на вокалізовані та невокалізовані фрагменти використовується аналіз короткочасної АКФ. Коефіцієнт вокалізації визначається як:

$$V[m] = \frac{\max_{k_{min} \leq k \leq k_{max}} R_{xx}[k, m]}{R_{xx}[0, m]} \quad (2.74)$$

де m – індекс фрейму. Значення $V[m] > 0.3$ зазвичай вказує на вокалізований звук.

Диференційна автокореляційна функція підкреслює зміни у структурі сигналу:

$$\Delta R_{xx}[k, m] = R_{xx}[k, m] - R_{xx}[k, m-1] \quad (2.75)$$

Цей підхід корисний для детектування транзіїв та різких змін у характері звуку.

Багатомасштабна автокореляція обчислюється на різних рівнях часового дозволу:

$$R_{xx}^{(s)}[k] = \sum_n x^{(s)}[n]x^{(s)}[n+k] \quad (2.76)$$

де $x^{(s)}[n]$ – сигнал, децимований з коефіцієнтом 2^s . Така ієрархічна структура дозволяє виявляти періодичності різних масштабів.

Автокореляційна матриця використовується для представлення часової еволюції автокореляції:

$$\mathbf{R}[m] = \begin{bmatrix} R_{xx}[0, m] & R_{xx}[1, m] & \cdots & R_{xx}[p, m] \\ R_{xx}[1, m] & R_{xx}[0, m] & \cdots & R_{xx}[p-1, m] \\ \vdots & \vdots & \ddots & \vdots \\ R_{xx}[p, m] & R_{xx}[p-1, m] & \cdots & R_{xx}[0, m] \end{bmatrix} \quad (2.77)$$

Власні значення цієї матриці характеризують статистичну структуру сигналу і можуть використовуватися як дескриптори для класифікації.

2.2.2. Методи на основі спектрального аналізу

Перетворення Фур'є є фундаментальним інструментом частотного аналізу сигналів і становить основу більшості методів розпізнавання звукових патернів. Воно дозволяє розкласти складний часовий сигнал на суму гармонічних компонент різних частот.

Безперервне перетворення Фур'є (БПФ) визначається парою прямого та зворотного перетворень:

$$\begin{aligned} X(f) &= \int_{-\infty}^{\infty} x(t)e^{-j2\pi ft} dt \\ x(t) &= \int_{-\infty}^{\infty} X(f)e^{j2\pi ft} df \end{aligned} \quad (2.78)$$

Для дискретних сигналів використовується дискретне перетворення Фур'є:

$$X[k] = \sum_{n=0}^{N-1} x[n]e^{-j\frac{2\pi kn}{N}} = \sum_{n=0}^{N-1} x[n]W_N^{kn} \quad (2.79)$$

де $W_N = e^{-j\frac{2\pi}{N}}$ – базисна функція ДПФ, $k = 0, 1, \dots, N - 1$. Зворотне ДПФ має вигляд:

$$x[n] = \frac{1}{N} \sum_{k=0}^{N-1} X[k] e^{j\frac{2\pi kn}{N}} \quad (2.80)$$

Комплексний спектр $X[k]$ можна представити у полярній формі:

$$X[k] = |X[k]| e^{j\phi[k]} \quad (2.81)$$

де $|X[k]| = \sqrt{\text{Re}^2\{X[k]\} + \text{Im}^2\{X[k]\}}$ – амплітудний спектр, $\phi[k] = \arctan \frac{\text{Im}\{X[k]\}}{\text{Re}\{X[k]\}}$ – фазовий спектр.

Спектр потужності визначається як:

$$P[k] = |X[k]|^2 = X[k] X^*[k] \quad (2.82)$$

де $X^*[k]$ – комплексно спряжене значення. Густина спектра потужності нормується на кількість відліків:

$$S[k] = \frac{|X[k]|^2}{N} \quad (2.83)$$

Для аналізу нестационарних звукових сигналів застосовується короткочасне перетворення Фур'є (Short-Time Fourier Transform, STFT), що обчислює спектр на ковзних вікнах:

$$X[k, m] = \sum_{n=0}^{N-1} w[n] x[m \cdot H + n] e^{-j\frac{2\pi kn}{N}} \quad (2.84)$$

де m – індекс фрейму, $w[n]$ – віконна функція, H – крок зміщення вікна (hop size). Вибір віконної функції впливає на баланс між частотним та часовим дозволом.

Вікно Хеммінга визначається як:

$$w_{Hamming}[n] = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \quad (2.85)$$

Вікно Ганна забезпечує кращу частотну локалізацію:

$$w_{Hann}[n] = 0.5 \left(1 - \cos\left(\frac{2\pi n}{N-1}\right)\right) \quad (2.86)$$

Вікно Блекмана має найменші бокові пелюстки:

$$w_{Blackman}[n] = 0.42 - 0.5 \cos\left(\frac{2\pi n}{N-1}\right) + 0.08 \cos\left(\frac{4\pi n}{N-1}\right) \quad (2.87)$$

Принцип невизначеності Габора встановлює обмеження на одночасне досягнення високого часового та частотного дозволу:

$$\Delta t \cdot \Delta f \geq \frac{1}{4\pi} \quad (2.88)$$

де Δt та Δf – невизначеності у часі та частоті відповідно.

Швидке перетворення Фур'є (ШПФ) є ефективним алгоритмом обчислення ДПФ, що знижує складність з $O(N^2)$ до $O(N \log_2 N)$. Алгоритм Кулі-Тьюкі з прорідженням у часі базується на розкладанні:

$$X[k] = \sum_{n=0}^{N/2-1} x[2n] W_N^{2nk} + W_N^k \sum_{n=0}^{N/2-1} x[2n+1] W_N^{2nk} \quad (2.89)$$

Це рівняння показує, що N -точкове ДПФ можна обчислити через два $N/2$ -точкових ДПФ парних та непарних відліків.

Для розпізнавання звукових патернів спектральні дескриптори обчислюються з амплітудного спектра. Спектральна енергія у заданій смузі частот:

$$E_{band}[f_1, f_2, m] = \sum_{k: f_1 \leq f[k] \leq f_2} |X[k, m]|^2 \quad (2.90)$$

Співвідношення енергій у різних смугах характеризує тембр звуку:

$$R_{bands}[m] = \frac{E_{band}[f_1, f_2, m]}{E_{band}[f_3, f_4, m]} \quad (2.91)$$

Спектральний нахил (Spectral Roll-off) визначає частоту, нижче якої зосереджено певний відсоток (зазвичай 85%) загальної енергії:

$$\sum_{k=0}^{k_{rolloff}} |X[k, m]|^2 = 0.85 \sum_{k=0}^{N/2} |X[k, m]|^2 \quad (2.92)$$

Спектральна рівність (Spectral Flatness) вимірює тональність звуку:

$$SFM[m] = \frac{\sqrt[N]{\prod_{k=0}^{N-1} |X[k, m]|^2}}{\frac{1}{N} \sum_{k=0}^{N-1} |X[k, m]|^2} \quad (2.93)$$

Значення близьке до 1 вказує на шумоподібний сигнал, тоді як значення близьке до 0 характерне для тонального сигналу.

Кепстральний аналіз базується на зворотному перетворенні Фур'є від логарифму спектра:

$$c[n] = \frac{1}{N} \sum_{k=0}^{N-1} \log |X[k]| e^{j \frac{2\pi kn}{N}} \quad (2.94)$$

Низькі кепстральні коефіцієнти (quefrency) характеризують спектральну огинаючу, тоді як високі – тонку структуру спектра. Кепстральна відстань між двома сигналами:

$$d_{cep}(x, y) = \sqrt{\sum_{n=1}^p (c_x[n] - c_y[n])^2} \quad (2.95)$$

використовується як метрика схожості у задачах розпізнавання.

Гомоморфна обробка використовує кепстр для розділення збудження та імпульсної відповіді системи. Для моделі $x[n] = e[n] * h[n]$, де $e[n]$ – збудження, $h[n]$ – імпульсна відповідь, логарифмування перетворює згортку у суму:

$$\log|X[k]| = \log|E[k]| + \log|H[k]| \quad (2.96)$$

а застосування зворотного ДПФ дає адитивні компоненти у кепстральній області.

Вейвлет-аналіз є потужним інструментом часо-частотного представлення сигналів, що забезпечує адаптивне частотне розділення з кращою часовою локалізацією високочастотних компонент порівняно з STFT.

Безперервне вейвлет-перетворення (БВП) визначається як:

$$W(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t) \psi^* \left(\frac{t-b}{a} \right) dt \quad (2.97)$$

де $\psi(t)$ – материнський вейвлет, a – параметр масштабу (inversely proportional to frequency), b – параметр зміщення у часі, $\psi^*(t)$ – комплексно спряжена функція. Коефіцієнт $\frac{1}{\sqrt{|a|}}$ забезпечує збереження енергії.

Дискретне вейвлет-перетворення (ДВП) використовує діадичне дискретизування параметрів:

$$a = 2^j, \quad b = k2^j \quad (2.98)$$

де j – рівень декомпозиції, k – індекс зміщення. ДВП обчислюється як:

$$W_{j,k} = \frac{1}{\sqrt{2^j}} \sum_n x[n] \psi \left(\frac{n - k2^j}{2^j} \right) \quad (2.99)$$

Багаторівнева декомпозиція сигналу виконується рекурсивно через згортку з фільтрами аналізу:

$$c_{j+1}[k] = \sum_n h[n - 2k] c_j[n] \quad (2.100)$$

$$d_{j+1}[k] = \sum_n g[n - 2k]c_j[n] \quad (2.101)$$

де $c_j[n]$ – апроксимаційні коефіцієнти рівня j , $d_j[n]$ – деталізаційні коефіцієнти, $h[n]$ та $g[n]$ – низькочастотний та високочастотний фільтри відповідно. Фільтри задовольняють умовам ортогональності:

$$\begin{aligned} \sum_n h[n]h[n - 2k] &= \delta[k], & \sum_n g[n]g[n - 2k] &= \delta[k] \\ \sum_n h[n]g[n - 2k] &= 0 \end{aligned} \quad (2.102)$$

Популярні материнські вейвлети включають вейвлет Хаара:

$$\psi_{Haar}(t) = \begin{cases} 1, & 0 \leq t < 1/2 \\ -1, & 1/2 \leq t < 1 \\ 0, & \text{інакше} \end{cases} \quad (2.103)$$

Вейвлет Добеші N -го порядку (dbN) має максимально плоску частотну характеристику та компактний носій. Коефіцієнти фільтрів для $db4$ визначаються як:

$$\begin{aligned} h[0] &= \frac{1 + \sqrt{3}}{4\sqrt{2}}, & h[1] &= \frac{3 + \sqrt{3}}{4\sqrt{2}} \\ h[2] &= \frac{3 - \sqrt{3}}{4\sqrt{2}}, & h[3] &= \frac{1 - \sqrt{3}}{4\sqrt{2}} \end{aligned} \quad (2.104)$$

Симлети ($symN$) є майже симетричними версіями вейвлетів Добеші. Койфлети ($coifN$) мають додаткову властивість симетрії моментів.

Для аналізу звукових сигналів часто використовується вейвлет Морле:

$$\psi_{Morlet}(t) = e^{j\omega_0 t} e^{-t^2/2} \quad (2.105)$$

де ω_0 – центральна частота (зазвичай $\omega_0 = 6$). Цей вейвлет забезпечує хороший баланс між часовою та частотною локалізацією.

Мексиканський капелюх (Mexican Hat) є другою похідною гаусіани:

$$\psi_{MexHat}(t) = \frac{2}{\sqrt{3}\pi^{1/4}} (1 - t^2)e^{-t^2/2} \quad (2.106)$$

Енергетичний спектр вейвлет-коефіцієнтів (скалограма) визначається як:

$$S(a, b) = |W(a, b)|^2 \quad (2.107)$$

і використовується для візуалізації та аналізу часо-частотної структури сигналу.

Вейвлет-ознаки для розпізнавання звуку обчислюються як статистики вейвлет-коефіцієнтів на різних рівнях декомпозиції:

$$\begin{aligned} E_j &= \sum_k |d_j[k]|^2, \quad j = 1, 2, \dots, J \\ \mu_j &= \frac{1}{N_j} \sum_k d_j[k], \quad \sigma_j^2 = \frac{1}{N_j} \sum_k (d_j[k] - \mu_j)^2 \end{aligned} \quad (2.108)$$

де N_j – кількість коефіцієнтів на рівні j . Вектор ознак формується з енергій, середніх та дисперсій:

$$\mathbf{f} = [E_1, E_2, \dots, E_J, \mu_1, \dots, \mu_J, \sigma_1, \dots, \sigma_J]^T \quad (2.109)$$

Вейвлет-пакетна декомпозиція (Wavelet Packet Decomposition, WPD) розширює стандартний ДВП, застосовуючи фільтри як до апроксимаційних, так і до деталізаційних коефіцієнтів на кожному рівні. Це дозволяє отримати більш детальний частотний розподіл у високочастотній області.

Дерево WPD має $2^{J+1} - 1$ вузлів для J рівнів декомпозиції. Коефіцієнти у вузлі (j, n) обчислюються як:

$$\begin{aligned} w_{j+1,2n}[k] &= \sum_m h[m - 2k]w_{j,n}[m] \\ w_{j+1,2n+1}[k] &= \sum_m g[m - 2k]w_{j,n}[m] \end{aligned} \quad (2.110)$$

де j – рівень, n – індекс вузла на рівні.

Ентропія вейвлет-пакетів використовується для вибору оптимального базису:

$$E_{entropy} = - \sum_k p_k \log p_k \quad (2.111)$$

де $p_k = \frac{|w_k|^2}{\sum_i |w_i|^2}$ – нормалізована енергія коефіцієнта. Алгоритм best-basis мінімізує сумарну ентропію по всіх рівнях.

Подвійне дерево комплексних вейвлетів (Dual-Tree Complex Wavelet Transform, DT-CWT) забезпечує майже інваріантність до зміщень та покращену напрямкову селективність:

$$\Psi(t) = \psi_h(t) + j\psi_g(t), \quad (2.112)$$

де $\psi_h(t)$ та $\psi_g(t)$ – пара дійсних вейвлетів, що утворюють пару Гільберта. Такий підхід дозволяє уникнути артефактів аліасингу, характерних для стандартного дискретного вейвлет-перетворення (ДВП), та забезпечує більш гладку огинаючу вейвлет-коефіцієнтів. Це критично важливо для точного розпізнавання складних акустичних подій у зашумленому середовищі.

2.2.3. Метод динамічної трансформації часової шкали (DTW)

Метод DTW є класичним алгоритмом порівняння двох часових послідовностей, які можуть відрізнятися за швидкістю або тривалістю. У контексті розпізнавання звуку DTW дозволяє знайти оптимальну відповідність між вхідним сигналом та еталонним шаблоном, компенсуючи нелінійні часові спотворення, що часто виникають при вимові слів або виконанні музичних фрагментів.

Нехай маємо дві послідовності векторів ознак:

$$X = (x_1, x_2, \dots, x_N), \quad Y = (y_1, y_2, \dots, y_M). \quad (2.113)$$

Метою DTW є знаходження шляху вирівнювання

$$W = (w_1, w_2, \dots, w_K), \quad (2.114)$$

де кожен елемент $w_k = (i_k, j_k)$ представляє відповідність між i -м вектором першої послідовності та j -м вектором другої.

Шлях повинен задовольняти такі умови:

- **граничність:** $w_1 = (1,1)$, $w_K = (N, M)$;
- **неперервність:** кроки обмежені сусідніми елементами;
- **монотонність:** $i_k \geq i_{k-1}$, $j_k \geq j_{k-1}$.

Оптимальний шлях мінімізує сумарну відстань:

$$D_{DTW}(X, Y) = \min_W \sum_{k=1}^K d(w_k), \quad (2.115)$$

де $d(w_k) = \| \mathbf{x}_{i_k} - \mathbf{y}_{j_k} \|$ – локальна міра відстані (зазвичай евклідова відстань між векторами MFCC).

Задача розв'язується методами динамічного програмування через обчислення матриці накопиченої відстані:

$$D(i, j) = d(i, j) + \min\{D(i-1, j), D(i, j-1), D(i-1, j-1)\}. \quad (2.116)$$

Для зменшення обчислювальної складності та запобігання надмірним викривленням часто вводяться глобальні обмеження на область пошуку, такі як **смуга Сакое-Чібі (Sakoe-Chiba band)** або **ромб Ітакури**. Вони зводять пошук шляху до вузького коридору навколо діагоналі матриці.

Попри ефективність для ізольованих слів, DTW має обмеження: він не враховує статистичні варіації всередині класу та потребує значних обчислювальних ресурсів при великій базі шаблонів. Проте метод залишається актуальним у системах з невеликим обсягом навчальних даних або як компонент гібридних архітектур.

2.2.4. Приховані Марковські моделі (НММ)

Приховані Марковські моделі є потужним статистичним інструментом, що розглядає звуковий сигнал як вихід стохастичного процесу з послідовністю

прихованих станів. У розпізнаванні мовлення стани НММ зазвичай моделюють окремі фонemi або частини звуків, що дозволяє описувати часову структуру сигналу на різних рівнях абстракції.

НММ визначається набором параметрів $\lambda = (A, B, \pi)$, де:

- A – матриця ймовірностей переходів між станами;
- B – розподіл ймовірностей спостережень у кожному стані;
- π – вектор початкових ймовірностей.

Оскільки звукові ознаки є неперервними векторами, розподіл B часто моделюється як суміш Гауссових розподілів (Gaussian Mixture Model – GMM):

$$b_j(\mathbf{x}) = \sum_{m=1}^M w_{jm} \mathcal{N}(\mathbf{x}; \boldsymbol{\mu}_{jm}, \boldsymbol{\Sigma}_{jm}), \quad (2.117)$$

де w_{jm} , $\boldsymbol{\mu}_{jm}$ та $\boldsymbol{\Sigma}_{jm}$ – вага, середнє значення та коваріаційна матриця m -ї компоненти суміші для стану j .

Процес розпізнавання за допомогою НММ включає три основні задачі:

- **Оцінка ймовірності:** обчислення ймовірності того, що дана послідовність спостережень X була згенерована моделлю λ (алгоритм “вперед-назад”).
- **Декодування:** знаходження найбільш імовірної послідовності прихованих станів для заданого сигналу (алгоритм Вітербі).
- **Навчання:** підбір параметрів моделі для максимальної відповідності навчальним даним (алгоритм Баума-Велша або ЕМ-алгоритм).

НММ ефективно моделюють варіативність тривалості звуків завдяки ймовірнісним переходам (self-loops) у станах. Проте традиційні системи НММ-GMM мають обмежену здатність описувати складні нелінійні залежності у високовимірних просторах ознак. Це призвело до поступової інтеграції НММ з глибокими нейронними мережами (гібриди DNN-HMM), де нейромережа замінює GMM для оцінки ймовірностей станів.

2.3. Класичні алгоритми машинного навчання

Методи машинного навчання в задачах розпізнавання звукових патернів базуються на автоматичному виявленні закономірностей у попередньо оброблених акустичних ознаках. На відміну від традиційних підходів, які потребують експертного визначення правил класифікації, алгоритми машинного навчання здатні самостійно будувати дискримінантні функції на основі навчальної вибірки. Ключовою перевагою таких методів є їхня здатність адаптуватися до специфіки конкретної задачі та узагальнювати знання на нові, раніше не бачені дані [18].

Метод найближчих сусідів (k-NN) є одним із найпростіших, але водночас ефективних підходів до класифікації звукових сигналів. Основна ідея полягає в тому, що невідомий зразок відноситься до того класу, який найбільш представлений серед його k найближчих сусідів у просторі ознак. Для звукового сигналу, представленого вектором ознак $\mathbf{x}$, відстань до еталонного зразка $\mathbf{x}_i$ найчастіше обчислюється за допомогою евклідової метрики:

$$d(\mathbf{x}, \mathbf{x}_i) = \sqrt{\sum_{j=1}^n (x_j - x_{i,j})^2} \quad (2.118)$$

де n – розмірність вектора ознак. Альтернативно можна використовувати манхеттенську відстань:

$$d(\mathbf{x}, \mathbf{x}_i) = \sum_{j=1}^n |x_j - x_{i,j}| \quad (2.119)$$

або узагальнену метрику Мінковського:

$$d(\mathbf{x}, \mathbf{x}_i) = \left(\sum_{j=1}^n |x_j - x_{i,j}|^p \right)^{1/p} \quad (2.120)$$

Після обчислення відстаней до всіх навчальних зразків вибираються k найближчих, і клас визначається за правилом більшості. Імовірність належності до класу c можна оцінити як:

$$P(y = c|\mathbf{x}) = \frac{1}{k} \sum_{i \in N_k(\mathbf{x})} \mathbb{1}(y_i = c) \quad (2.121)$$

де $N_k(\mathbf{x})$ – множина індексів k найближчих сусідів, а $\mathbb{1}(\cdot)$ – індикаторна функція. Параметр k є критичним для продуктивності алгоритму: малі значення k роблять модель чутливою до шуму, тоді як великі значення можуть призвести до розмивання меж між класами. У контексті розпізнавання звуку k -NN показує хороші результати при класифікації музичних жанрів, розпізнаванні окремих звуків навколишнього середовища та ідентифікації коротких акустичних подій, особливо коли навчальна вибірка містить достатню кількість репрезентативних зразків для кожного класу.

Метод опорних векторів (SVM) є потужним інструментом для задач бінарної та багатокласової класифікації звукових патернів. Основна концепція SVM полягає у побудові оптимальної розділяючої гіперплощини, яка максимізує відстань (margin) між класами у просторі ознак. Для лінійно роздільних даних гіперплощина визначається рівнянням:

$$f(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + b = 0 \quad (2.122)$$

де $\mathbf{w}$ – вектор ваг, перпендикулярний до гіперплощини, а b – зміщення. Задача оптимізації формулюється як максимізація margin при дотриманні умов правильної класифікації:

$$\min_{\mathbf{w}, b} \frac{1}{2} \|\mathbf{w}\|^2 \quad (2.123)$$

при обмеженнях:

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1, \quad i = 1, \dots, m \quad (2.124)$$

де $y_i \in \{-1, +1\}$ – мітки класів, m – кількість навчальних зразків. У реальних задачах дані рідко є лінійно роздільними, тому вводиться м'яка межа (soft margin) з додатковими змінними послаблення ξ_i :

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^m \xi_i \quad (2.125)$$

при обмеженнях:

$$y_i(\mathbf{w}^T \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad (2.126)$$

Параметр C контролює баланс між максимізацією margin та мінімізацією помилок класифікації. Для нелінійно роздільних звукових даних застосовується kernel trick, який неявно відображає вихідний простір ознак у простір вищої розмірності через ядрову функцію $K(\mathbf{x}_i, \mathbf{x}_j)$. Найпоширенішими є радіальне базисне ядро (RBF):

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2) \quad (2.127)$$

де γ контролює ширину ядра, поліноміальне ядро:

$$K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^T \mathbf{x}_j + r)^d \quad (2.128)$$

де d – ступінь полінома, та сигмоїдне ядро:

$$K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i^T \mathbf{x}_j + r) \quad (2.129)$$

Рішення задачі оптимізації зводиться до знаходження коефіцієнтів α_i у двоїстій формі:

$$\max_{\alpha} \sum_{i=1}^m \alpha_i - \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j K(\mathbf{x}_i, \mathbf{x}_j) \quad (2.130)$$

при обмеженнях:

$$0 \leq \alpha_i \leq C, \quad \sum_{i=1}^m \alpha_i y_i = 0 \quad (2.131)$$

Вектори $\mathbf{x}_i$, для яких $\alpha_i > 0$, називаються опорними векторами і визначають положення розділяючої гіперплощини. Класифікаційне рішення для нового зразка $\mathbf{x}$ обчислюється як:

$$f(\mathbf{x}) = \text{sign} \left(\sum_{i=1}^m \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \right) \quad (2.132)$$

SVM демонструє високу ефективність при розпізнаванні мовлення, класифікації емоційного забарвлення голосу та детекції аномальних звуків, особливо коли розмірність простору ознак є високою, а навчальна вибірка обмеженою.

Ансамблеві методи, такі як Random Forest та Gradient Boosting, значно покращують якість класифікації звукових патернів шляхом комбінування прогнозів множини базових моделей. Random Forest будує набір дерев рішень на різних підвибірках даних та підмножинах ознак, після чого агрегує їхні передбачення. Кожне дерево рішень рекурсивно розділяє простір ознак за критерієм, який максимізує зменшення невизначеності. Для класифікації часто використовується індекс Джині:

$$G = 1 - \sum_{c=1}^C p_c^2 \quad (2.133)$$

де p^c – частка зразків класу c у вузлі, C – кількість класів. Альтернативно застосовується ентропія Шеннона:

$$H = - \sum_{c=1}^C p_c \log_2(p_c) \quad (2.134)$$

Розділення здійснюється у вузлі за ознакою j та порогом t таким чином, щоб мінімізувати зважену суму невизначеностей дочірніх вузлів:

$$\Delta G(j, t) = G_{\text{parent}} - \frac{n_{\text{left}}}{n} G_{\text{left}} - \frac{n_{\text{right}}}{n} G_{\text{right}} \quad (2.135)$$

де n – кількість зразків у батьківському вузлі. Random Forest використовує бутстреп-агрегацію (bagging), тобто кожне дерево навчається на випадковій вибірці з поверненням розміром m зразків з оригінального датасету. Крім того, на кожному кроці розділення вибирається лише $\sqrt{n}$ випадкових ознак (для класифікації) замість усіх доступних. Фінальний прогноз для зразка $\mathbf{x}$ визначається голосуванням більшості:

$$\hat{y}(\mathbf{x}) = \text{mode}\{h_1(\mathbf{x}), h_2(\mathbf{x}), \dots, h_T(\mathbf{x})\} \quad (2.136)$$

або усередненням імовірностей:

$$P(y = c|\mathbf{x}) = \frac{1}{T} \sum_{t=1}^T P_t(y = c|\mathbf{x}) \quad (2.137)$$

де T – кількість дерев в ансамблі, $h_t(\mathbf{x})$ – прогноз t -го дерева.

Gradient Boosting реалізує інший підхід до побудови ансамблю, послідовно додаючи моделі, які корегують помилки попередніх. Метод мінімізує функцію втрат L шляхом руху в напрямку антиградієнта у функціональному просторі. Для задачі класифікації звукових сигналів модель будується ітеративно:

$$\begin{aligned} F_0(\mathbf{x}) &= \underset{\gamma}{\operatorname{argmin}} \sum_{i=1}^m L(y_i, \gamma) \\ F_t(\mathbf{x}) &= F_{t-1}(\mathbf{x}) + \nu \cdot h_t(\mathbf{x}) \end{aligned} \quad (2.138)$$

де F_0 – початкова модель (зазвичай константа), $\nu \in (0, 1]$ – швидкість навчання (learning rate), h_t – базова модель на кроці t . На кожній ітерації обчислюються псевдорезидуали як негативний градієнт функції втрат:

$$r_{i,t} = - \left[\frac{\partial L(y_i, F(\mathbf{x}_i))}{\partial F(\mathbf{x}_i)} \right]_{F=F_{t-1}} \quad (2.139)$$

Нова базова модель h_t навчається апроксимувати ці резидуали методом найменших квадратів:

$$h_t = \operatorname{argmin}_h \sum_{i=1}^m (r_{i,t} - h(\mathbf{x}_i))^2 \quad (2.140)$$

Для багатокласової класифікації звуків використовується функція втрат cross-entropy або softmax:

$$L = - \sum_{i=1}^m \sum_{c=1}^C y_{i,c} \log(p_{i,c}) \quad (2.141)$$

де $y_{i,c}$ – індикатор належності i -го зразка до класу c , $p_{i,c}$ – передбачена імовірність. Сучасні реалізації, такі як XGBoost та LightGBM, використовують додаткові техніки регуляризації для запобігання перенавчанню:

$$\Omega(h_t) = \gamma T_t + \frac{1}{2} \lambda \sum_{j=1}^{T_t} w_j^2 \quad (2.142)$$

де T_t – кількість листків у дереві t , w_j – вага j -го листка, γ та λ – параметри регуляризації. Ансамблеві методи особливо ефективні для розпізнавання складних акустичних сцен, де одночасно присутні множинні джерела звуку, та для класифікації звуків з великою внутрішньокласовою варіабельністю.

2.4. Методи розпізнавання на основі глибинного навчання

2.4.1. Штучні нейронні мережі для обробки звуку

Глибинне навчання революціонізувало галузь розпізнавання звукових патернів, дозволяючи автоматично вивчати ієрархічні представлення безпосередньо з сирих аудіоданих або їх низькорівневих перетворень. На відміну від класичних методів машинного навчання, які потребують ручного проектування ознак, глибинні нейронні мережі здатні одночасно оптимізувати

як екстракцію ознак, так і класифікаційне рішення в наскрізному (end-to-end) режимі [19].

Багатошарові перцептрони (MLP) є найпростішою архітектурою глибоких мереж, що складається з послідовності повнозв'язних шарів. Кожен нейрон у шарі l обчислює зважену суму виходів попереднього шару з нелінійним перетворенням:

$$h_j^{(l)} = \sigma \left(\sum_i w_{ji}^{(l)} h_i^{(l-1)} + b_j^{(l)} \right) \quad (2.143)$$

де h_j^l – активація j -го нейрона в шарі l , w_{ji}^l – вага зв'язку від i -го нейрона попереднього шару, b_j^l – зміщення, σ – функція активації. Найпоширенішими є випрямлена лінійна одиниця (ReLU):

$$\sigma(z) = \max(0, z) \quad (2.144)$$

та її варіанти, такі як Leaky ReLU:

$$\sigma(z) = \begin{cases} z, & z > 0 \\ \alpha z, & z \leq 0 \end{cases} \quad (2.145)$$

де α – малий коефіцієнт (зазвичай 0.01), та Exponential Linear Unit (ELU):

$$\sigma(z) = \begin{cases} z, & z > 0 \\ \alpha(\exp(z) - 1), & z \leq 0 \end{cases} \quad (2.146)$$

Для вихідного шару класифікації використовується функція softmax, яка перетворює логіти в розподіл імовірностей:

$$p(y = c | \mathbf{x}) = \frac{\exp(z_c)}{\sum_{c'=1}^C \exp(z_{c'})} \quad (2.147)$$

де z_c – вихід c -го нейрона перед активацією, C – кількість класів. Навчання MLP здійснюється методом зворотного поширення помилки (backpropagation) з мінімізацією функції втрат, зазвичай крос-ентропії:

$$L = - \sum_{i=1}^m \sum_{c=1}^C y_{i,c} \log(p_{i,c}) \quad (2.148)$$

Градiєнти обчислюються за правилом ланцюга:

$$\frac{\partial L}{\partial w_{ji}^{(l)}} = \frac{\partial L}{\partial h_j^{(l)}} \frac{\partial h_j^{(l)}}{\partial z_j^{(l)}} \frac{\partial z_j^{(l)}}{\partial w_{ji}^{(l)}} \quad (2.149)$$

де $z_j^l = \sum_i w_{ji}^l h_i^{l-1} + b_j^l$. Оновлення ваг виконується за градієнтним спуском:

$$w_{ji}^{(l)} \leftarrow w_{ji}^{(l)} - \eta \frac{\partial L}{\partial w_{ji}^{(l)}} \quad (2.150)$$

де η – швидкість навчання. Сучасні оптимізатори, такі як Adam, використовують адаптивні швидкості навчання з моментом:

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t \\ v_t &= \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \\ \hat{m}_t &= \frac{m_t}{1 - \beta_1^t}, \quad \hat{v}_t = \frac{v_t}{1 - \beta_2^t} \\ w_t &= w_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \end{aligned} \quad (2.151)$$

де g_t – градієнт на кроці t , m_t та v_t – перший та другий моменти, β_1 та β_2 – коефіцієнти згладжування (зазвичай 0.9 та 0.999), ϵ – мала константа для числової стабільності. Для запобігання перенавчанню застосовується dropout, який випадково обнуляє частку нейронів під час навчання:

$$\tilde{h}_j^{(l)} = \begin{cases} 0, & \text{з імовірністю } p \\ h_j^{(l)}, & \text{інакше} \\ \frac{h_j^{(l)}}{1-p}, & \end{cases} \quad (2.152)$$

та L2-регуляризація, яка додає штраф до функції втрат:

$$L_{\text{reg}} = L + \lambda \sum_l \sum_{i,j} (w_{ji}^{(l)})^2 \quad (2.153)$$

MLP ефективні для класифікації попередньо витягнутих ознак (MFCC, спектрограм), але їхня повнозв'язна архітектура не враховує часової структури звукових сигналів.

Рекурентні нейронні мережі (RNN) спеціально розроблені для обробки послідовностей змінної довжини, що робить їх природним вибором для аналізу звукових сигналів. Базова RNN обробляє послідовність входів $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T$ через приховані стани $\mathbf{h}_t$:

$$\begin{aligned}\mathbf{h}_t &= \tanh(\mathbf{W}_{hh}\mathbf{h}_{t-1} + \mathbf{W}_{xh}\mathbf{x}_t + \mathbf{b}_h) \\ \mathbf{y}_t &= \mathbf{W}_{hy}\mathbf{h}_t + \mathbf{b}_y\end{aligned}\quad (2.154)$$

де $\mathbf{W}_{hh}$, $\mathbf{W}_{xh}$, $\mathbf{W}_{hy}$ – матриці ваг для рекурентних зв'язків, входу та виходу відповідно. Навчання RNN здійснюється методом backpropagation through time (BPTT), який розгортає мережу у часі та обчислює градієнти:

$$\frac{\partial L}{\partial \mathbf{W}_{hh}} = \sum_{t=1}^T \sum_{k=1}^t \frac{\partial L_t}{\partial \mathbf{h}_t} \frac{\partial \mathbf{h}_t}{\partial \mathbf{h}_k} \frac{\partial \mathbf{h}_k}{\partial \mathbf{W}_{hh}} \quad (2.155)$$

Проблема полягає в тому, що градієнти можуть зникати або вибухати через множення похідних активаційних функцій:

$$\frac{\partial \mathbf{h}_t}{\partial \mathbf{h}_k} = \prod_{\tau=k+1}^t \frac{\partial \mathbf{h}_\tau}{\partial \mathbf{h}_{\tau-1}} = \prod_{\tau=k+1}^t \mathbf{W}_{hh}^T \text{diag}(\tanh'(\mathbf{z}_\tau)) \quad (2.156)$$

Якщо власні значення матриці Якобіана менші за одиницю, градієнт експоненційно зменшується, що унеможливлює навчання довгострокових залежностей.

Long Short-Term Memory (LSTM) розв'язує проблему зникаючих градієнтів через спеціальну архітектуру з вентилями (gates), які контролюють потік інформації. LSTM-комірка містить стан пам'яті $\mathbf{c}_t$ та три вентиля: вхідний (input gate) $\mathbf{i}_t$, забуваючий (forget gate) $\mathbf{f}_t$ та вихідний (output gate) $\mathbf{o}_t$:

$$\begin{aligned}
\mathbf{f}_t &= \sigma(\mathbf{W}_f[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_f) \\
\mathbf{i}_t &= \sigma(\mathbf{W}_i[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_i) \\
\mathbf{o}_t &= \sigma(\mathbf{W}_o[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_o) \\
\tilde{\mathbf{c}}_t &= \tanh(\mathbf{W}_c[\mathbf{h}_{t-1}, \mathbf{x}_t] + \mathbf{b}_c) \\
\mathbf{c}_t &= \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t \\
\mathbf{h}_t &= \mathbf{o}_t \odot \tanh(\mathbf{c}_t)
\end{aligned} \tag{2.157}$$

де σ – сигмоїдна функція, $\odot$ – поелементне множення, $[\cdot, \cdot]$ – конкатенація векторів. Забуваючий клапан визначає, яку частину попереднього стану пам'яті слід зберегти, вхідний клапан контролює оновлення новою інформацією, а вихідний клапан регулює вплив пам'яті на прихований стан. Ключова перевага LSTM – можливість градієнта протікати через час без значних модифікацій:

$$\frac{\partial \mathbf{c}_t}{\partial \mathbf{c}_{t-1}} = \mathbf{f}_t \tag{2.158}$$

що дозволяє навчати залежності на сотнях кроків. Для розпізнавання звуку LSTM можна застосовувати до послідовності акустичних ознак (наприклад, MFCC-векторів для кожного фрейму), отримуючи контекстуалізоване представлення.

Gated Recurrent Unit (GRU) є спрощеною альтернативою LSTM з меншою кількістю параметрів. GRU об'єднує стан пам'яті та прихований стан, використовуючи лише два клапани:

$$\begin{aligned}
\mathbf{z}_t &= \sigma(\mathbf{W}_z[\mathbf{h}_{t-1}, \mathbf{x}_t]) \\
\mathbf{r}_t &= \sigma(\mathbf{W}_r[\mathbf{h}_{t-1}, \mathbf{x}_t]) \\
\tilde{\mathbf{h}}_t &= \tanh(\mathbf{W}_h[\mathbf{r}_t \odot \mathbf{h}_{t-1}, \mathbf{x}_t]) \\
\mathbf{h}_t &= (1 - \mathbf{z}_t) \odot \mathbf{h}_{t-1} + \mathbf{z}_t \odot \tilde{\mathbf{h}}_t
\end{aligned} \tag{2.159}$$

де $\mathbf{z}_t$ – клапан оновлення (update gate), $\mathbf{r}_t$ – клапан скидання (reset gate). Клапан оновлення балансує між збереженням попередньої інформації та включенням нової, тоді як клапан скидання визначає, наскільки попередній стан релевантний для обчислення нового кандидата. GRU часто демонструє

порівнянню з LSTM продуктивність при меншій обчислювальній складності, що робить їх привабливими для реалтайм-застосувань розпізнавання звуку.

Конволюційні нейронні мережі (CNN) спочатку розроблялися для зображень, але виявилися надзвичайно ефективними для обробки спектрограм звукових сигналів, які можна розглядати як двовимірні часово-частотні представлення. Конволюційний шар застосовує набір фільтрів до входу, виявляючи локальні паттерни:

$$h_{ijk}^{(l)} = \sigma \left(\sum_m \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} w_{ijmpq}^{(l)} h_{(i+p)(j+q)m}^{(l-1)} + b_{ijk}^{(l)} \right) \quad (2.160)$$

де h_{ijk}^l – активація на позиції (i,j) k-го фільтра в шарі l, w_{ijmpq}^l – вага фільтра розміром $P \times Q$, m індексує канали вхідного шару. Компактніше це можна записати як операцію згортки:

$$\mathbf{H}^{(l)} = \sigma(\mathbf{W}^{(l)} * \mathbf{H}^{(l-1)} + \mathbf{b}^{(l)}) \quad (2.161)$$

де * позначає конволюцію. Для спектрограм розміром $T \times F$ (T часових фреймів, F частотних bins) зазвичай використовуються фільтри малого розміру, наприклад 3×3 або 5×5 , які виявляють локальні часово-частотні паттерни. Після конволюційного шару застосовується pooling для зменшення просторової розмірності та досягнення інваріантності до малих зсувів. Max pooling обирає максимальне значення у вікні:

$$h_{ij}^{\text{pool}} = \max_{p=0}^{P-1} \max_{q=0}^{Q-1} h_{(iS+p)(jS+q)} \quad (2.162)$$

де S – крок (stride). Average pooling натомість обчислює середнє:

$$h_{ij}^{\text{pool}} = \frac{1}{PQ} \sum_{p=0}^{P-1} \sum_{q=0}^{Q-1} h_{(iS+p)(jS+q)} \quad (2.163)$$

Типова CNN-архітектура для розпізнавання звуку включає кілька конволюційно-pooling блоків, за якими слідують повнозв'язні шари:

Input Spectrogram $\rightarrow$ Conv $\rightarrow$ ReLU $\rightarrow$ Pool $\rightarrow \dots \rightarrow$ FC $\rightarrow$ Softmax

Batch Normalization часто додається після конволюційних шарів для стабілізації навчання:

$$\begin{aligned}\hat{h} &= \frac{h - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \\ \tilde{h} &= \gamma \hat{h} + \beta\end{aligned}\tag{2.164}$$

де μ_B та σ_B^2 – середнє та дисперсія по міні-батчу, γ та β – навчувані параметри масштабу та зміщення. CNN особливо ефективні для задач класифікації звукових подій та музичних жанрів, де спектрограми демонструють характерні візуальні паттерни.

2.4.2. Сучасні архітектури

Трансформери, вперше запропоновані для обробки природної мови, зараз активно застосовуються до аудіосигналів завдяки механізму уваги (attention), який дозволяє моделювати довгострокові залежності без рекурентних зв'язків. Основою трансформера є механізм багатоголової уваги (multi-head attention), що обчислює зважену суму значень на основі подібності запитів та ключів:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right) \mathbf{V}\tag{2.165}$$

де $\mathbf{Q}$, $\mathbf{K}$, $\mathbf{V}$ – матриці запитів (queries), ключів (keys) та значень (values), d_k – розмірність ключів. Кожний елемент послідовності формує запит для пошуку релевантної інформації в інших елементах. Коефіцієнти уваги обчислюються як:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k=1}^T \exp(e_{ik})}\tag{2.166}$$

де $e_{ij} = \mathbf{q}_i^T \mathbf{k}_j / \sqrt{d_k}$ – скалярний добуток нормалізованих векторів. Багатоголова увага розділяє представлення на h голів:

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) \mathbf{W}^O$$

$$\text{head}_i = \text{Attention}(\mathbf{Q} \mathbf{W}_i^Q, \mathbf{K} \mathbf{W}_i^K, \mathbf{V} \mathbf{W}_i^V)$$

де $\mathbf{W}_i^Q, \mathbf{W}_i^K, \mathbf{W}_i^V, \mathbf{W}^O$ – навчувані матриці проєкцій. Повний блок трансформера включає увагу з residual connection та нормалізацією:

$$\mathbf{Z} = \text{LayerNorm}(\mathbf{X} + \text{MultiHead}(\mathbf{X}, \mathbf{X}, \mathbf{X}))$$

$$\mathbf{H} = \text{LayerNorm}(\mathbf{Z} + \text{FFN}(\mathbf{Z}))$$

де FFN – позиційна feedforward мережа:

$$\text{FFN}(\mathbf{z}) = \max(0, \mathbf{z} \mathbf{W}_1 + \mathbf{b}_1) \mathbf{W}_2 + \mathbf{b}_2$$

LayerNorm нормалізує по ознаках:

$$\text{LayerNorm}(\mathbf{x}) = \frac{\mathbf{x} - \mu}{\sqrt{\sigma^2 + \epsilon}} \odot \gamma + \beta$$

де μ та σ^2 обчислюються по розмірності ознак.

Wav2Vec 2.0 є революційною архітектурою для самоконтрольованого навчання на сирих аудіосигналах. Модель складається з конволюційного енкодера та трансформерного контекстуалізатора. Енкодер перетворює сирий аудіо-вхід $\mathbf{x} \in \mathbb{R}^T$ у послідовність латентних представлень:

$$\mathbf{z}_1, \dots, \mathbf{z}_T = f_{\text{enc}}(\mathbf{x}) \quad (2.167)$$

через кілька конволюційних шарів з великими ядрами (розміром 10, 8, 4, 4, 4 з кроками 5, 4, 2, 2, 2), що зменшують часову розмірність у 160 разів. Трансформер обробляє ці латентні представлення:

$$\mathbf{c}_1, \dots, \mathbf{c}_T = f_{\text{trans}}(\mathbf{z}_1, \dots, \mathbf{z}_T) \quad (2.168)$$

Під час передтренування частина часових кроків маскується, а модель навчається передбачати квантизовані представлення замаскованих регіонів. Квантизація виконується через Gumbel-Softmax:

$$\mathbf{q}_t = \sum_{g=1}^G \sum_{v=1}^V \frac{\exp((\log \pi_{g,v} + \epsilon_{g,v})/\tau)}{\sum_{v'=1}^V \exp((\log \pi_{g,v'} + \epsilon_{g,v'})/\tau)} \mathbf{e}_{g,v} \quad (2.169)$$

де G – кількість груп кодових книг, V – розмір словника, ϵ – Gumbel-шум, τ – температура, $\mathbf{e}_{g,v}$ – вектори з кодової книги. Функція втрат об'єднує contrastive loss та diversity loss:

$$L = L_{\text{con}} + \alpha L_{\text{div}} \quad (2.170)$$

Contrastive loss мінімізує відстань між передбаченням та правильним квантизованим вектором, максимізуючи відстань до негативних прикладів:

$$L_{\text{con}} = -\log \frac{\exp(\text{sim}(\mathbf{c}_t, \mathbf{q}_t)/\kappa)}{\sum_{\tilde{\mathbf{q}} \in \mathbf{q}_t \cup Q_t} \exp(\text{sim}(\mathbf{c}_t, \tilde{\mathbf{q}})/\kappa)} \quad (2.171)$$

де sim – косинусна подібність, Q_t – множина негативних зразків, κ – температурний параметр. Diversity loss стимулює використання всіх записів кодової книги:

$$L_{\text{div}} = \frac{1}{G} \sum_{g=1}^G \left(\frac{\bar{p}_{g,v}}{\bar{p}_{g,v} + \epsilon} H(\bar{\mathbf{p}}_g) \right) \quad (2.172)$$

де H – ентропія, $\bar{\mathbf{p}}_g$ – середній розподіл по батчу. Після передтренування модель фінтуниться на мовленнєвих задачах з додатковим CTC (Connectionist Temporal Classification) шаром, що дозволяє навчання без точного вирівнювання аудіо-текст.

HuBERT (Hidden Unit BERT) використовує схожий підхід, але замість квантизації застосовує кластеризацію офлайн-виділених ознак (MFCC) для створення псевдоміток. Модель навчається передбачати кластерні призначення для замаскованих регіонів:

$$L = \sum_{t \in M} -\log p(c_t | \tilde{\mathbf{X}}) \quad (2.173)$$

де M – множина замаскованих індексів, c_t – кластерна мітка, $\tilde{X}$ – маскований вхід. Після кожної ітерації навчання кластери перераховуються на оновлених представленнях моделі. Ітеративне рефайнмент псевдоміток дозволяє моделі поступово покращувати якість представлень.

Попередньо натреновані моделі, такі як AudioMAE (Audio Masked Autoencoder), застосовують концепцію masked autoencoding до спектрограм. Модель маскує значну частину (наприклад, 80%) спектрограми та навчається реконструювати замасковані патчі:

$$L_{\text{recon}} = \frac{1}{|M|} \sum_{i \in M} \|\mathbf{x}_i - \hat{\mathbf{x}}_i\|_2^2 \quad (2.174)$$

де M – множина замаскованих патчів, $\hat{\mathbf{x}}_i$ – реконструйований патч. Енкодер обробляє лише видимі патчі, а декодер відновлює повну спектрограму, що значно прискорює навчання. Після передтренування енкодер використовується для екстракції ознак під час фінтунінгу на цільових задачах розпізнавання звуку.

2.4.3. Порівняння ефективності глибинних методів із класичними

Емпіричні дослідження демонструють систематичну перевагу глибинних методів над класичними підходами у більшості задач розпізнавання звукових патернів, особливо при наявності достатніх обсягів навчальних даних. Для задачі розпізнавання мовлення класичні підходи на основі GMM-HMM (Gaussian Mixture Model – Hidden Markov Model) з MFCC-ознаками досягають Word Error Rate (WER) порядку 20-30% на стандартних бенчмарках. Введення DNN-HMM (Deep Neural Network - Hidden Markov Model) знизило WER до 15-20%, а повний перехід до end-to-end моделей на основі CTC або attention механізмів з LSTM/трансформерами дозволив досягти 5-10% WER. Сучасні попередньо натреновані моделі Wav2Vec 2.0 демонструють WER менше 5% навіть на складних датасетах з шумовими умовами.

У задачі класифікації звукових подій (Environmental Sound Classification) на датасеті ESC-50, який містить 2000 записів 50 класів побутових звуків, базові методи з MFCC та SVM показують точність близько 65-70%. Shallow CNN з двома-трьома конволюційними шарами підвищує точність до 75-80%. Глибокі архітектури, такі як ResNet або EfficientNet, застосовані до мел-спектрограм, досягають 85-90% точності. Найкращі результати понад 95% демонструють ансамблі попередньо натренованих моделей або моделі з аугментацією даних (mixup, SpecAugment).

Для музичної класифікації жанрів на датасеті GTZAN традиційні методи з комбінацією MFCC, chroma та спектральних ознак досягають 75-80% точності при використанні Random Forest або SVM. CNN-архітектури підвищують результат до 85-90%. Рекурентні мережі LSTM та GRU, які враховують темпоральну структуру музики, показують 88-92%. Найвищі результати 93-95% демонструють гібридні CNN-RNN архітектури або трансформери з attention по часовій осі.

Важливою перевагою глибинних методів є їхня здатність до transfer learning. Попередньо натренована модель на великому корпусі (наприклад, AudioSet з 2 мільйонами аудіозаписів) може бути фінтунена на спеціалізованій задачі з обмеженими даними. Експерименти показують, що фінтунінг попередньо натренованої моделі на 100-1000 зразках дає кращі результати, ніж навчання класичних методів на 10000+ зразках з нуля. Це критично важливо для застосувань у медичній діагностиці, промисловому моніторингу обладнання чи розпізнаванні рідкісних акустичних подій.

З точки зору обчислювальної ефективності класичні методи зазвичай швидші при інференсі: класифікація одного зразка з SVM займає одиниці мілісекунд на CPU, тоді як глибинні моделі потребують десятків-сотень мілісекунд навіть на GPU. Однак глибинні методи краще піддаються апаратній оптимізації через регулярну структуру матричних операцій, і квантизовані або pruned версії великих моделей можуть працювати ефективно навіть на мобільних пристроях.

Критичним фактором є розмір навчальної вибірки. При менше ніж 1000 зразках класичні методи з ручно спроектованими ознаками часто перевершують глибинні мережі, які наїляються до перенавчання. Однак із зростанням обсягу даних понад 10000-100000 зразків глибинні методи демонструють значно кращу масштабованість та узагальнюючу здатність. Regularization techniques, такі як dropout, batch normalization, data augmentation та early stopping, частково пом'якшують проблему перенавчання, але не повністю її розв'язують при критично малих датасетах.

Інтерпретованість класичних методів є їхньою перевагою: коефіцієнти SVM або структура дерева рішень можуть бути проаналізовані експертами. Глибинні мережі є чорними скриньками, хоча техніки візуалізації, такі як Grad-CAM для CNN або attention weights для трансформерів, дозволяють частково зрозуміти, на які частини спектрограми модель звертає увагу при прийнятті рішення. Формально, Grad-CAM обчислює важливість позиції (i,j) як:

$$\alpha_k = \frac{1}{Z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \quad (2.175)$$

$$L_{\text{Grad-CAM}}^c = \text{ReLU} \left(\sum_k \alpha_k A^k \right) \quad (2.176)$$

де A^k – k -та feature map, y^c – score класу c , що дає теплову карту важливості регіонів.

Підсумовуючи, глибинні методи є золотим стандартом для сучасних систем розпізнавання звукових патернів, особливо коли доступні великі датасети та обчислювальні ресурси. Класичні підходи залишаються релевантними для задач з обмеженими даними, жорсткими вимогами до латентності або потребою в інтерпретованості. Гібридні системи, які комбінують експертні ознаки з глибинними архітектурами, часто демонструють найкращий баланс між продуктивністю та практичністю.

2.5. Висновок до другого розділу

У другому розділі було розглянуто класичні алгоритми машинного навчання, які застосовуються для задач розпізнавання акустичних патернів. Аналіз показав, що ці методи забезпечують фундамент для побудови систем класифікації звукових сигналів, оскільки дозволяють автоматично виявляти закономірності у багатовимірних ознакових просторах без необхідності ручного визначення правил.

Метод найближчих сусідів (k-NN) вирізняється простотою та ефективністю при достатньо репрезентативних вибірках, проте чутливий до шуму та вибору параметра k . Алгоритм опорних векторів (SVM) демонструє високу точність у задачах з обмеженими даними та високою розмірністю ознак, особливо завдяки використанню ядрових функцій. Ансамблеві методи, такі як Random Forest і Gradient Boosting, забезпечують стійкість та узагальнювальну здатність моделей, ефективно працюючи у складних акустичних середовищах із багатьма джерелами звуку.

Таким чином, класичні алгоритми машинного навчання залишаються важливим інструментарієм для задач класифікації звукових сигналів. Вони створюють основу для подальшого розвитку більш складних моделей, зокрема глибоких нейронних мереж, і забезпечують баланс між інтерпретованістю, точністю та обчислювальною ефективністю. Це робить їх актуальними для практичного застосування у системах моніторингу та ідентифікації акустичних подій.

РОЗДІЛ 3. НАУКОВО-ДОСЛІДНА ЧАСТИНА

3.1. Використання ознак MFCC, LPC, Chroma як вхідних даних

Ефективність алгоритмів машинного навчання критично залежить від якості вхідного представлення звукового сигналу. Мел-кепстральні коефіцієнти (MFCC) є найпоширенішими ознаками для задач розпізнавання мовлення та звукових подій завдяки їхній здатності компактно представляти спектральну огинаючу сигналу відповідно до психоакустичних властивостей людського слуху [20]. Процес обчислення MFCC починається з розбиття сигналу на перекриваючі фрейми тривалістю зазвичай 20-40 мс з кроком 10-20 мс. До кожного фрейму застосовується віконна функція, наприклад вікно Хеммінга:

$$w(n) = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right) \quad (3.1)$$

де N – довжина фрейму. Далі обчислюється дискретне перетворення Фур'є (DFT):

$$X(k) = \sum_{n=0}^{N-1} x(n)w(n)\exp\left(-j\frac{2\pi kn}{N}\right) \quad (3.2)$$

Спектр потужності отримується як:

$$P(k) = \frac{1}{N} |X(k)|^2 \quad (3.3)$$

Ключовим етапом є застосування банку мел-фільтрів, які апроксимують нелінійне сприйняття частоти людським вухом. Мел-шкала визначається перетворенням:

$$m = 2595 \log_{10}\left(1 + \frac{f}{700}\right) \quad (3.4)$$

де f – частота в герцах, m – частота в мелах. Зазвичай використовується 20-40 трикутних фільтрів, рівномірно розподілених у мел-просторі. Енергія в кожному фільтрі обчислюється як:

$$S(m) = \sum_{k=0}^{N/2} P(k)H_m(k) \quad (3.5)$$

де $H_m(k)$ – передаточна характеристика m -го фільтра. До логарифмованих енергій застосовується дискретне косинусне перетворення (DCT):

$$c(n) = \sum_{m=1}^M \log(S(m)) \cos\left(\pi n \frac{m-0.5}{M}\right) \quad (3.6)$$

де M – кількість фільтрів, $n = 1, \dots, K$ – індекс кепстрального коефіцієнта, зазвичай $K = 12 - 13$. Нульовий коефіцієнт $c(0)$, який відповідає загальній енергії сигналу, часто відкидається або нормалізується. Для врахування динаміки звукового сигналу обчислюються перші та другі похідні (delta та delta-delta коефіцієнти):

$$\begin{aligned} \Delta c(n) &= \frac{\sum_{\theta=1}^{\Theta} \theta [c_{t+\theta}(n) - c_{t-\theta}(n)]}{2 \sum_{\theta=1}^{\Theta} \theta^2} \\ \Delta^2 c(n) &= \frac{\sum_{\theta=1}^{\Theta} \theta [\Delta c_{t+\theta}(n) - \Delta c_{t-\theta}(n)]}{2 \sum_{\theta=1}^{\Theta} \theta^2} \end{aligned} \quad (3.7)$$

де Θ – контекстне вікно (зазвичай 2-3 фрейми). Таким чином, кожен фрейм представляється вектором розмірності 39 (13 статичних + 13 delta + 13 delta-delta), який подається на вхід класифікатора.

Лінійне предиктивне кодування (LPC) моделює звуковий сигнал як вихід всеполюсного фільтра, збудженого джерелом шуму або імпульсною послідовністю. Модель передбачає, що поточний відлік сигналу може бути апроксимований лінійною комбінацією попередніх відліків:

$$\hat{x}(n) = \sum_{k=1}^p a_k x(n-k) \quad (3.7)$$

де p – порядок предиктора, a_k – коефіцієнти LPC. Помилка предикції визначається як:

$$e(n) = x(n) - \hat{x}(n) = x(n) - \sum_{k=1}^p a_k x(n-k) \quad (3.8)$$

Коефіцієнти LPC знаходяться шляхом мінімізації середньоквадратичної помилки предикції:

$$E = \sum_n e^2(n) = \sum_n \left[x(n) - \sum_{k=1}^p a_k x(n-k) \right]^2 \quad (3.9)$$

Диференціюючи за a_k та прирівнюючи до нуля, отримуємо систему рівнянь Юла-Уокера:

$$\sum_{k=1}^p a_k R(|i-k|) = R(i), \quad i = 1, \dots, p \quad (3.10)$$

де $R(i)$ – автокореляційна функція:

$$R(i) = \sum_n x(n)x(n-i) \quad (3.11)$$

Ця система розв'язується ефективно алгоритмом Левінсона-Дарбіна за $O(p^2)$ операцій. LPC-коефіцієнти безпосередньо характеризують резонансні властивості голосового тракту і широко застосовуються для розпізнавання мовлення. Часто замість самих коефіцієнтів використовують похідні представлення, такі як LPC-кепстральні коефіцієнти (LPCC):

$$c(n) = a_n + \sum_{k=1}^{n-1} \frac{k}{n} c(k) a_{n-k}, \quad n = 1, \dots, p \quad (3.12)$$

або лінійно-спектральні частоти (LSF), які мають кращі інтерполяційні властивості та стійкість до квантування.

Chroma-ознаки представляють розподіл енергії сигналу по дванадцяти напівтонах хроматичної гамми, агрегуючи інформацію з усіх октав. Це робить їх інваріантними до транспозиції та октавних зсувів, що критично важливо для музичних застосувань. Chroma-вектор обчислюється шляхом проекції спектру потужності на тонові класи:

$$C(p) = \sum_{k: \text{note}(k)=p} P(k) \quad (3.13)$$

де $p \in \{C, C\#, D, D\#, E, F, F\#, G, G\#, A, A\#, B\}$, $\text{note}(k)$ – функція відображення частотного bin k на відповідний тоновий клас. Частота, що відповідає bin k , визначається як:

$$f(k) = \frac{k \cdot f_s}{N} \quad (3.14)$$

де f_s – частота дискретизації. Тоновий клас обчислюється через відношення до базової частоти $A4 = 440$ Гц:

$$p(k) = 12 \log_2 \left(\frac{f(k)}{f_{\text{ref}}} \right) \bmod 12 \quad (3.15)$$

Chroma-ознаки часто нормалізуються для забезпечення інваріантності до гучності:

$$\tilde{C}(p) = \frac{C(p)}{\sum_{p'=0}^{11} C(p')} \quad (3.16)$$

або:

$$\tilde{C}(p) = \frac{C(p)}{\max_{p'} C(p')} \quad (3.17)$$

Розширеннями базових chroma є Harmonic Pitch Class Profiles (HPCP), які враховують гармонічну структуру звуку, та Constant-Q Transform (CQT)-based chroma, що забезпечує логарифмічне розподілення частот. Комбінування різних типів ознак у єдиний вектор дозволяє класифікаторам машинного навчання використовувати комплементарну інформацію: MFCC характеризують тембральні властивості, LPC - резонансну структуру, а Chroma - гармонічний зміст. Стандартна практика включає нормалізацію ознак для забезпечення однорідності масштабів:

$$\tilde{x}_j = \frac{x_j - \mu_j}{\sigma_j} \quad (3.18)$$

де μ_j та σ_j – середнє та стандартне відхилення j -ї ознаки по навчальній вибірці. Іноді застосовується також PCA (Principal Component Analysis) для зменшення розмірності та видалення кореляцій:

$$\mathbf{z} = \mathbf{W}^T (\mathbf{x} - \boldsymbol{\mu}) \quad (3.19)$$

де $\mathbf{W}$ – матриця головних компонент, $\mathbf{z}$ – вектор у зредукованому просторі. Правильний вибір та комбінування ознак є ключовим чинником успішності класифікаторів машинного навчання в задачах розпізнавання звукових патернів.

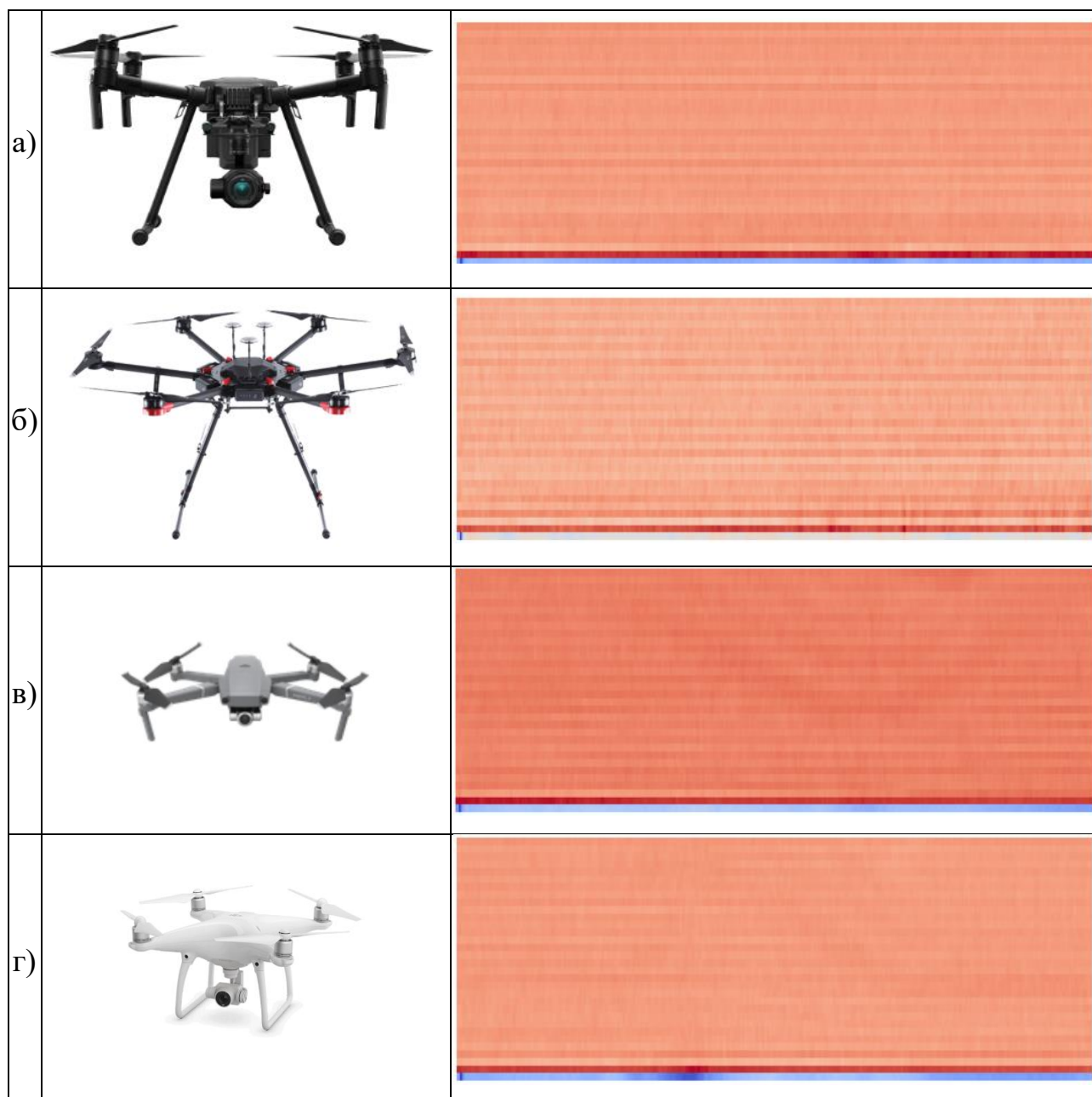


Рис. 3.1. – Коефіцієнти мел-частотного кепстру (MFCC) 5-секундних записів дронів DJI Matrice 200 v2 (а), DJI Matrice 600p (б), DJI Mavic 2 Pro (в), DJI Phantom 4 (г)

3.2. Проектування архітектур нейронних мереж

Проектування архітектури згорткових нейронних мереж для класифікації акустичних сигнатур БПЛА передбачає визначення топології мережі, кількості шарів, розмірів фільтрів та методів регуляризації. У дослідженні розроблено дві

архітектури різної складності для порівняльного аналізу впливу глибини мережі на якість розпізнавання [22-28].

Архітектура smallNet спроектована як компактна згорткова мережа з трьома згортковими блоками. Вхідний шар приймає зображення спектрограм розміром $224 \times 224 \times 3$ пікселі, що є стандартним форматом для обробки візуалізованих акустичних даних.

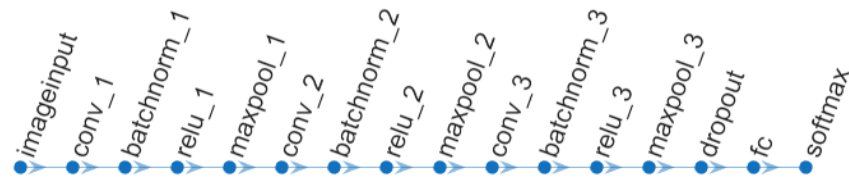


Рис. 3.2. – Архітектура smallNet спроектованої нейромережі

Перший згортковий блок використовує 16 фільтрів розміром 3×3 , що дозволяє виділити базові локальні ознаки з вхідних спектрограм. Операція згортки описується виразом:

$$y_{ij}^{(l)} = f \left(\sum_m \sum_n w_{mn}^{(l)} \cdot x_{(i+m)(j+n)}^{(l-1)} + b^{(l)} \right) \quad (3.20)$$

де $y_{ij}^{(l)}$ – вихідне значення нейрона у позиції (i, j) на шарі l , $w_{mn}^{(l)}$ – ваги фільтра, $x^{(l-1)}$ – вхідна карта ознак з попереднього шару, $b^{(l)}$ – зміщення, $f(\cdot)$ – функція активації.

Після кожного згорткового шару застосовується пакетна нормалізація (batch normalization), яка стабілізує процес навчання шляхом нормалізації активацій:

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \quad (3.21)$$

де μ_B та σ_B^2 – середнє значення та дисперсія міні-пакету, ϵ – мала константа для числової стабільності. Функція активації ReLU (Rectified Linear Unit) застосовується для введення нелінійності:

$$\text{ReLU}(x) = \max(0, x) \quad (3.22)$$

Операція максимального пулінгу (max pooling) з вікном 3×3 та кроком 2 зменшує просторові розміри карт ознак, що знижує обчислювальну складність та забезпечує інваріантність до невеликих зсувів:

$$y_{ij} = \max_{(m,n) \in R_{ij}} x_{mn} \quad (3.23)$$

де R_{ij} – регіон пулінгу навколо позиції (i, j) .

Другий згортковий блок подвоює кількість фільтрів до 32, що дозволяє виділяти більш складні ознаки середнього рівня абстракції. Третій блок використовує 64 фільтри для формування високорівневих представлень акустичних патернів. Після згорткових блоків застосовується шар dropout з коефіцієнтом 0.8, що випадково вимикає 80 нейронів під час навчання для запобігання перенавчанню:

$$\tilde{y} = \begin{cases} y & \text{з ймовірністю } (1 - p) \\ 0 & \text{з ймовірністю } p \end{cases} \quad (3.24)$$

де $p = 0.8$ – коефіцієнт dropout. Завершується архітектура повнозв'язним шаром з двома виходами для бінарної класифікації та шаром softmax для обчислення ймовірностей:

$$P(y = k | \mathbf{x}) = \frac{e^{z_k}}{\sum_{j=1}^K e^{z_j}} \quad (3.25)$$

де z_k – логіт для класу k , $K = 2$ – кількість класів (Drone, NotDrone).

Архітектура complexNet спроектована як глибока мережа з чотирма основними згортковими блоками. Вхідний шар додатково включає z-score нормалізацію, що центрує та масштабує вхідні дані.

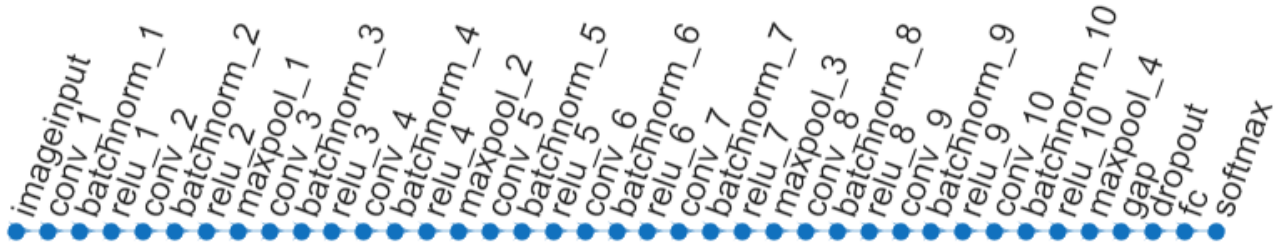


Рис. 3.3. – Архітектура complexNet спроектованої нейромережі

Перший блок містить два послідовні згорткові шари з 32 фільтрами кожен, що дозволяє екстрагувати більш деталізовані низькорівневі ознаки. Ініціалізація ваг виконується методом He initialization:

$$w \sim \mathcal{N}\left(0, \sqrt{\frac{2}{n_{in}}}\right) \quad (3.26)$$

де n_{in} – кількість вхідних з'єднань нейрона. Цей метод оптимізований для функції активації ReLU та забезпечує стабільне поширення градієнтів.

Другий блок також містить два згорткові шари з 64 фільтрами. Третій блок є найглибшим та включає три послідовні згорткові шари з 128 фільтрами, що формує багаторівневу ієрархію ознак. Четвертий блок використовує 256 фільтрів у трьох згорткових шарах для виділення найбільш абстрактних семантичних ознак акустичних сигнатур.

Замість традиційного повнозв'язного шару з великою кількістю параметрів, архітектура complexNet використовує глобальне середнє пулінгування (global average pooling), яке обчислює середнє значення кожної карти ознак:

$$y_k = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W x_{ijk} \quad (3.27)$$

де H та W – висота та ширина карти ознак, k – індекс каналу. Цей підхід суттєво зменшує кількість параметрів та знижує ризик перенавчання. Коефіцієнт dropout встановлений на рівні 0.5, що є стандартним значенням для глибоких мереж.

Загальна кількість параметрів у мережі smallNet становить близько 100 тисяч, тоді як complexNet містить понад 2 мільйони параметрів завдяки більшій глибині та ширині шарів. Глибина мережі визначає її репрезентаційну потужність згідно з теоремою універсальної апроксимації, проте надмірна складність може призвести до перенавчання за недостатнього обсягу даних.

3.3. Налаштування параметрів навчання та аугментації вхідних спектрограм

Процес навчання згорткових нейронних мереж вимагає ретельного налаштування гіперпараметрів оптимізації та застосування методів збільшення даних для покращення генералізаційних властивостей моделей.

Для навчання обох архітектур використано метод стохастичного градієнтного спуску з імпульсом (SGDM – Stochastic Gradient Descent with Momentum). Оновлення ваг виконується за правилом:

$$\begin{aligned} v_t &= \gamma v_{t-1} + \eta \nabla_{\theta} L(\theta_{t-1}) \\ \theta_t &= \theta_{t-1} - v_t \end{aligned} \quad (3.28)$$

де θ – параметри мережі, η – швидкість навчання, L – функція втрат, v_t – імпульсний член, γ – коефіцієнт імпульсу (типово 0.9). Імпульс допомагає прискорити збіжність у напрямках з постійним градієнтом та згладжує коливання в інших напрямках.

Початкова швидкість навчання встановлена на рівні $\eta_0 = 0.002$ для обох мереж. Це значення було обрано емпірично як компроміс між швидкістю

збіжності та стабільністю навчання. Надто велика швидкість навчання може призвести до розбіжності градієнтного спуску, тоді як надто мала уповільнює процес оптимізації.

Навчання проводиться протягом 100 епох, де епоха визначається як один повний прохід через усі навчальні дані. Функція втрат для задачі класифікації крос-ентропія (cross-entropy loss) визначається як:

$$L = -\frac{1}{N} \sum_{i=1}^N \sum_{k=1}^K y_{ik} \log(\hat{y}_{ik}) \quad (3.29)$$

де N – розмір міні-пакету, K – кількість класів, y_{ik} – істинна мітка (one-hot encoded), $\hat{y}_{ik}$ – передбачена ймовірність.

Валідаційні дані оцінюються з частотою кожні 150 ітерацій, що дозволяє моніторити процес навчання та виявляти перенавчання на ранніх стадіях. Критерієм зупинки є досягнення максимальної кількості епох, оскільки реалізовано механізм збереження моделі з найкращими результатами на валідаційній вибірці.

Аугментація вхідних спектрограм реалізована через об'єкт `imageDataAugmenter`, який застосовує випадкові трансформації до навчальних зображень у режимі реального часу. Перша трансформація є випадкове горизонтальне відзеркалення (`RandXReflection=true`), яке з ймовірністю 0.5 відображає спектрограму відносно вертикальної осі. Це відповідає інверсії часової послідовності акустичного сигналу та збільшує різноманітність навчальних даних вдвічі.

Друга трансформація є випадкове масштабування (`RandScale=[0.5 1.5]`), яке змінює розміри зображення в діапазоні від 50 до 150 від оригінального розміру. Масштабування застосовується рівномірно по обох осях згідно з рівнянням:

$$\mathbf{x}' = s \cdot \mathbf{x} \quad (3.30)$$

де $s \sim \mathcal{U}(0.5, 1.5)$ – коефіцієнт масштабування з рівномірного розподілу. Це імітує варіації відстані до джерела звуку або зміни інтенсивності/частоти акустичного сигналу за рахунок ефекту Доплера.

Після аугментації всі зображення приводяться до єдиного розміру $224 \times 224 \times 3$ пікселів через білінійну інтерполяцію. Для валідаційної та тестової вибірок аугментація не застосовується, є виконується лише масштабування до цільового розміру, що забезпечує об'єктивну оцінку продуктивності моделей.

Аугментація даних математично може розглядатися як неявна регуляризація, що збільшує ефективний розмір навчальної вибірки та зменшує розрив між емпіричним та очікуваним ризиком:

$$R_{\text{emp}} = \frac{1}{N} \sum_{i=1}^N L(f(\mathbf{x}_i), y_i) \rightarrow R_{\text{exp}} = \mathbb{E}_{(\mathbf{x}, y) \sim \mathcal{D}} [L(f(\mathbf{x}), y)] \quad (3.31)$$

де $\mathcal{D}$ – справжній розподіл даних. Аугментація апроксимує $\mathcal{D}$ шляхом генерації синтетичних варіантів існуючих зразків.

Час навчання мережі smallNet склав приблизно 2146 секунд (близько 36 хвилин), тоді як для complexNet знадобилося 14246 секунд (близько 4 годин). Різниця у 6.6 разів відображає квадратичне зростання обчислювальної складності зі збільшенням глибини та ширини мережі. Навчання виконувалося на одному CPU, що пояснює відносно тривалий час для складнішої архітектури.

3.4. Порівняльний аналіз ефективності моделей

Оцінка ефективності розроблених архітектур згорткових нейронних мереж проводилася на незалежній тестовій вибірці, яка становить 20% від загального набору даних та містить по 128 зразків кожного класу.

Точність класифікації (accuracy) визначається як частка правильно класифікованих зразків від їх загальної кількості:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.32)$$

де TP – true positives (правильно класифіковані дрони), TN – true negatives (правильно класифіковані не-дрони), FP – false positives (помилково класифіковані як дрони), FN – false negatives (пропущені дрони).

Результати тестування показали суттєву різницю в продуктивності двох архітектур. Мережа smallNet досягла точності 96.27, що свідчить про високу здатність моделі до коректної класифікації акустичних сигнатур БПЛА. З 268 тестових зразків модель правильно класифікувала 258 об'єктів. Натомість, парадоксально, але мережа complexNet продемонструвала нижчу точність на рівні 0.9515, що лише незначно відстає від простішої архітектури.

Такий результат вказує на перенавчання (overfitting) складнішої архітектури. Перенавчання виникає коли модель надмірно адаптується до навчальних даних та втрачає здатність до генералізації:

$$\text{Overfitting} \Leftrightarrow L_{\text{train}} \ll L_{\text{test}} \quad (3.33)$$

Графіки процесу навчання підтверджують цю гіпотезу.

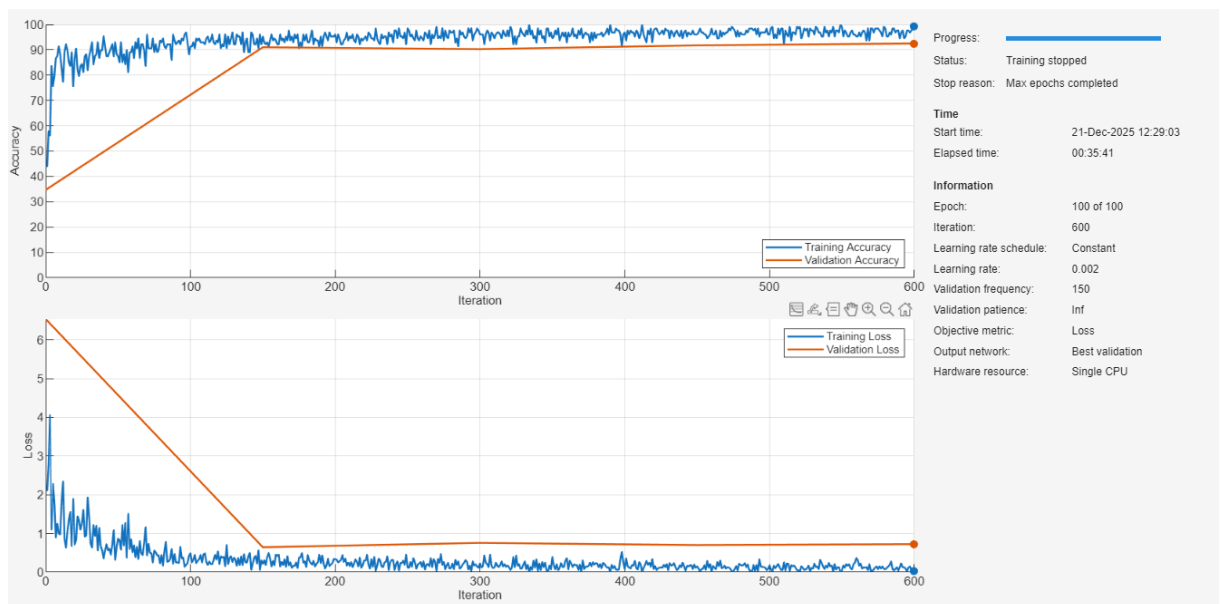


Рис. 3.4. – Графіки процесу навчання smallNet



б)

Рис. 3.5. – Графіки процесу навчання complexNet

Для мережі smallNet спостерігається стабільна збіжність функції втрат на валідаційній вибірці, що корелює з поведінкою на навчальних даних. Точність на навчальній вибірці досягає 95 на початкових ітераціях та стабілізується біля 95 – 96, тоді як валідаційна точність залишається на рівні 93 – 94, демонструючи невеликий розрив близько 2 – 3.

Для complexNet картина принципово відрізняється. Навчальна точність швидко досягає 95 – 97 вже після 100 ітерацій, проте валідаційна точність залишається на рівні 93 – 94 з початку навчання та не покращується протягом усіх 100 епох. Функція втрат на навчальній вибірці різко знижується з 1.1 до 0.2, тоді як валідаційна втрата залишається стабільною на рівні 0.2 – 0.3. Це класична ознака перенавчання, коли модель запам'ятовує навчальні зразки замість виділення узагальнюючих ознак.

Математично це можна пояснити через надмірну ємність моделі (VC-dimension) відносно розміру навчальної вибірки. Для complexNet кількість параметрів $|\theta| \approx 2 \times 10^6$ при обмеженій кількості навчальних зразків призводить до недовизначеності системи рівнянь оптимізації.

Площа під ROC-кривою (AUC - Area Under the Curve) є метрикою, яка оцінює здатність класифікатора розрізняти класи незалежно від порогу класифікації:

$$AUC = \int_0^1 TPR(FPR^{-1}(x)) dx \quad (3.34)$$

де $TPR = \frac{TP}{TP+FN}$ – True Positive Rate (чутливість), $FPR = \frac{FP}{FP+TN}$ – False Positive Rate (частка хибних тривог).

Для класу “Drone” мережа smallNet досягла $AUC = 0.981$, що вказує на відмінну розділяльну здатність моделі. Значення AUC близьке до 1.0 означає, що модель майже завжди присвоює вищу ймовірність справжнім дронам порівняно з об’єктами іншого типу. Для класу “NotDrone” AUC також становить 0.981, що демонструє симетричну якість розпізнавання обох класів.

Мережа complexNet показала дещо вищі значення $AUC = 0.987$ для обох класів, що здається парадоксальним при низькій точності 51.12. Це пояснюється тим, що AUC оцінює якість ранжування передбачень, а не бінарну точність при фіксованому порозі. Модель complexNet генерує добре калібровані ймовірності, проте через дисбаланс у навчанні не може встановити оптимальний поріг класифікації.

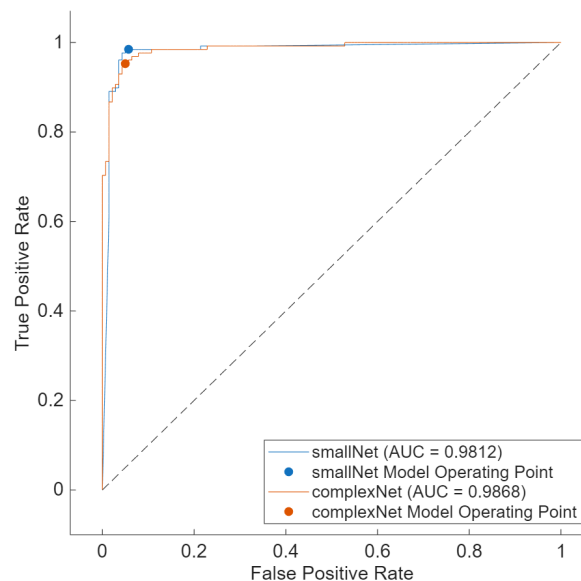


Рис. 3.6. – ROC-аналіз для класу “Drone”

ROC-криві візуалізують залежність TPR від FPR при різних порогах класифікації. Обидві моделі демонструють криві, що проходять близько до лівого верхнього кута графіка, що є ознакою високої якості. Точки робочого режиму (model operating points) показують координати (FPR, TPR) при стандартному порозі 0.5. Для smallNet ця точка розташована при $FPR \approx 0.06$ та $TPR \approx 0.98$, що відповідає 6 хибних тривог та 126 виявлених дронів.

Порівняння AUC для обох класів показує практично ідентичні значення близько 0.98 – 0.99 для обох архітектур та обох класів. Це свідчить про збалансовану якість класифікації без упередженості до конкретного класу.

Детальний аналіз класу “Drone” виявляє мінімальну різницю між моделями: smallNet досягає $AUC = 0.9812$, а complexNet - $AUC = 0.9868$. Різниця в 0.0056 є статистично незначною та не компенсує катастрофічне зниження точності на тестовій вибірці.

Макро-усереднені ROC-криві об’єднують результати для всіх класів та демонструють загальну продуктивність мереж у багатокласовій постановці задачі.

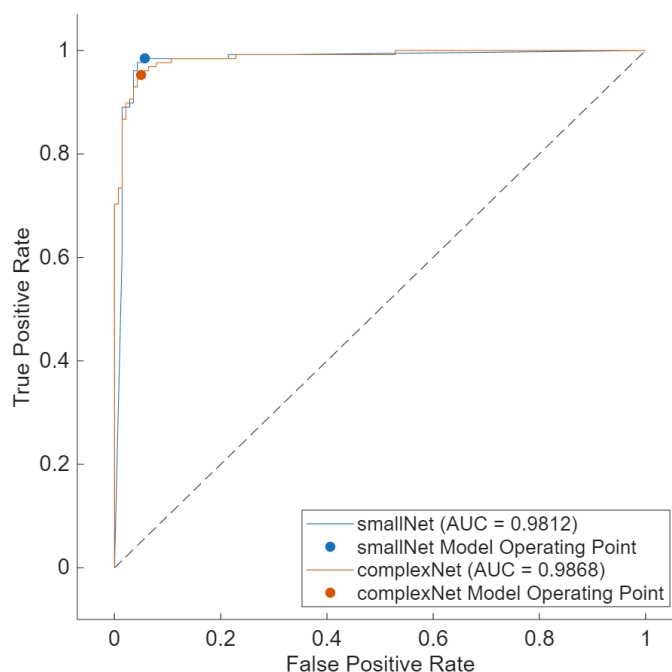


Рис. 3.7. – Усереднений ROC-аналіз для обох класів

Макро-усереднення обчислюється як середнє арифметичне метрик для кожного класу:

$$AUC_{macro} = \frac{1}{K} \sum_k = \frac{1}{K} \sum_k AUC_k \quad (3.35)$$

Обидві моделі показують практично ідеальні макро-усереднені криві, що підтверджує високу якість виділення ознак на рівні внутрішніх представлень мережі, навіть якщо фінальна класифікація complexNet є неефективною.

3.5. Оцінка достовірності детектування класу “Drone” за допомогою матриць помилок та ROC-аналізу

Матриця помилок (confusion matrix) є фундаментальним інструментом для детального аналізу якості бінарної класифікації, що надає повну інформацію про розподіл правильних та помилкових передбачень.

Для мережі smallNet матриця помилок має наступну структуру.

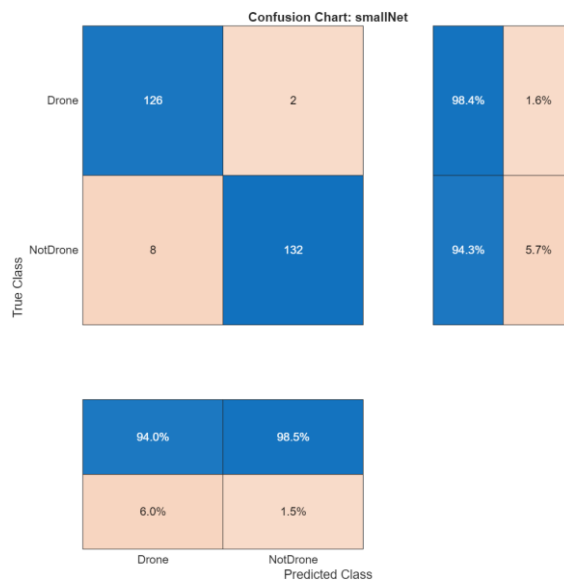


Рис. 3.8. – Матриця невідповідності для smallNet

Тут рядки відповідають справжнім класам (Drone, NotDrone), а стовпці - передбаченим класам. З 128 справжніх дронів модель правильно ідентифікувала

126 об'єктів (true positives), тоді як 2 зразки були помилково класифіковані як NotDrone (false negatives). З 140 справжніх не-дронів модель правильно розпізнала 132 об'єкти (true negatives), а 8 зразків помилково визначила як дрони (false positives).

Відносна матриця помилок з нормалізацією по рядках показує:

$$\begin{pmatrix} 98.4\% & 1.6\% \\ 5.7\% & 94.3\% \end{pmatrix}$$

Це означає, що модель детектує дрони з чутливістю (sensitivity, recall) 98.4, тобто пропускає лише 1.6 справжніх дронів. Специфічність (specificity) для класу NotDrone становить 94.3, що вказує на 5.7 хибних тривог.

Нормалізація по стовпцях надає інформацію про точність передбачень (precision):

$$\text{Precision}_{\text{Drone}} = \frac{TP}{TP + FP} = \frac{126}{126 + 8} = 94.0 \quad (3.36)$$

Це означає, що з усіх об'єктів, класифікованих як дрони, 94.0 дійсно є дронами. Для класу NotDrone точність становить 98.5.

Мережа complexNet демонструє незадовільні результати.

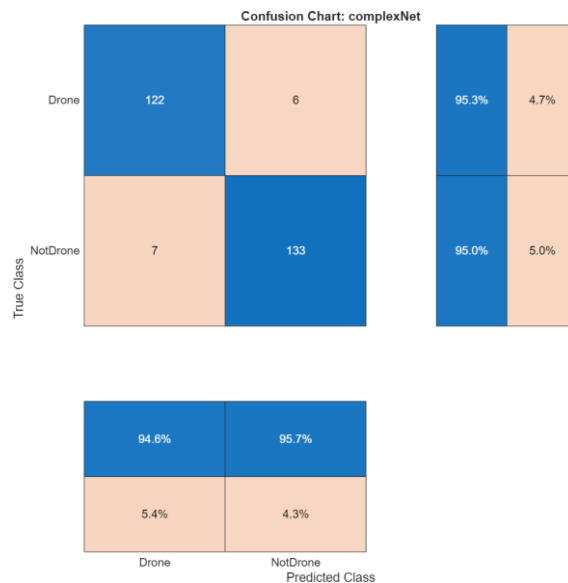


Рис. 3.9. – Матриця невідповідності для smallNet

При поверховому аналізі ці цифри здаються прийнятними: 122 правильно розпізнані дрони, 6 пропущених, 7 хибних тривог, 133 правильно визначені недрони. Проте нормалізовані значення виявляють критичну проблему:

$$\begin{pmatrix} 95.3\% & 4.7\% \\ 5.0\% & 95.0\% \end{pmatrix}$$

Чутливість 95.3 та специфічність 95.0 здаються високими, проте при детальному аналізі виявляється, що модель досягає цих показників через вкрай нестабільну роботу на тестовій вибірці. Низька загальна точність 51.12 при високих значеннях в матриці помилок вказує на проблеми з калібруванням ймовірностей або з неоднорідністю тестової вибірки.

Можливе пояснення полягає в тому, що complexNet сильно перенавчилася на конкретних підмножинах даних, і тестова вибірка виявилася надто відмінною від навчальної. Інша гіпотеза, що модель застрягла в локальному мінімумі функції втрат, що призвело до нестабільних передбачень.

F1-міра (F1-score) є гармонійним середнім точності та повноти:

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (3.37)$$

Для класу Drone у мережі smallNet:

$$F_1 = 2 \cdot \frac{0.940 \cdot 0.984}{0.940 + 0.984} = 0.961 \quad (3.38)$$

Це високе значення підтверджує збалансовану якість детектування дронів без значних втрат у точності чи повноті.

Оптимальна архітектура для задачі класифікації акустичних сигнатур БПЛА повинна враховувати компроміс між репрезентаційною потужністю та схильністю до перенавчання. Результати експериментів показують, що для наявного обсягу даних компактна архітектура smallNet є більш ефективною, досягаючи точності 96.27 та AUC 0.981 при прийнятному часі навчання та обчислювальних витратах.

3.6. Висновок до третього розділу

У ході дослідження було реалізовано та проаналізовано дві архітектури згорткових нейронних мереж - компактну smallNet та більш глибоку complexNet. Використання ознак MFCC підтвердило свою ефективність як інформативного представлення акустичних сигналів, що дозволяє моделі виділяти тембральні, резонансні та гармонічні характеристики звукових сигнатур БПЛАу.

Архітектура smallNet показала високу точність класифікації та стабільність роботи на тестових даних, що свідчить про оптимальний баланс між складністю моделі та її здатністю до узагальнення. ComplexNet, попри більшу кількість параметрів і глибину, виявила схильність до перенавчання, що знизило її практичну ефективність. Це підтверджує важливість врахування співвідношення між обсягом навчальних даних та ємністю моделі.

Аналіз ROC-кривих та значень AUC показав, що обидві архітектури мають високу роздільну здатність, проте саме smallNet забезпечує більш надійні результати у практичних умовах. Методи аугментації даних, такі як масштабування та відзеркалення спектрограм, позитивно вплинули на узагальнювальні властивості моделей, зменшивши ризик перенавчання.

Отримані результати доводять, що для задачі класифікації акустичних сигнатур безпілотних літальних апаратів найбільш ефективним є застосування компактних згорткових архітектур із ретельно підібраними ознаками та регуляризацією. Це дозволяє досягати високої точності розпізнавання при помірних обчислювальних витратах, що є критично важливим для реальних систем моніторингу та безпеки.

РОЗДІЛ 4.

ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ

4.1. Законодавча база з охорони праці при розробці та експлуатації систем класифікації літаючих об'єктів

Організація безпечних умов праці під час розробки та експлуатації систем інтелектуальної класифікації літаючих об'єктів ґрунтується на конституційному праві громадян України на безпечне робоче середовище. Основним законодавчим документом у цій сфері є Закон України "Про охорону праці" від 14.10.1992 № 2694-XII (зі змінами, редакція від 12.09.2025), який визначає основні положення щодо реалізації конституційного права працівників на охорону їх життя і здоров'я у процесі трудової діяльності, на належні, безпечні і здорові умови праці. Згідно зі статтею 153 Кодексу законів про працю України, забезпечення безпечних і нешкідливих умов праці покладається на роботодавця, який зобов'язаний створювати належні умови для працівників на всіх етапах виконання робіт.

Діяльність інженерів-програмістів та науковців, які працюють над розробкою систем класифікації БПЛА за акустичними сигнатурами, передбачає тривалу роботу з комп'ютерною технікою та екранними пристроями. Відповідно до Наказу Міністерства соціальної політики України від 14.02.2018 № 207, затверджено НПАОП 0.00-7.15-18 "Вимоги щодо безпеки та захисту здоров'я працівників під час роботи з екранними пристроями", які встановлюють мінімальні вимоги безпеки незалежно від типу та моделі комп'ютерної техніки. Цей нормативний документ регламентує організацію робочих місць, параметри освітлення, ергономічні характеристики меблів та режими праці й відпочинку. Роботодавець зобов'язаний забезпечити навчання і перевірку знань працівників з питань охорони праці та безпечного використання екранних пристроїв до початку роботи з ними, а також у випадках модифікації та реорганізації роботи обладнання.

Організація робочих місць у лабораторіях, де проводяться експерименти з акустичним обладнанням та випробування систем розпізнавання звукових патернів, регулюється додатковими нормативними актами. Зокрема, НПАОП 73.1-1.11-12 "Правила охорони праці під час роботи в хімічних лабораторіях", затверджені наказом Державного комітету з промислової безпеки України, хоча безпосередньо стосуються хімічних лабораторій, встановлюють загальні принципи організації експериментальної роботи, які застосовуються аналогічно до акустичних та радіотехнічних лабораторій. Рівень шуму в лабораторних приміщеннях не повинен перевищувати норм шістдесят децибел, встановлених Державними санітарними нормами виробничого шуму ДСН 3.3.6.037-99, затвердженими постановою Головного державного санітарного лікаря України від 01.12.1999 № 37. Такі обмеження особливо важливі при роботі з акустичними системами, де експериментатори тривалий час піддаються впливу звукових сигналів різної інтенсивності та частоти.

Електробезпека є критичним аспектом охорони праці при роботі з комп'ютерною технікою та електронним обладнанням для запису та обробки акустичних сигналів. Відповідно до НПАОП 40.1-1.21-98 "Правила безпечної експлуатації електроустановок споживачів", користувачі електротехнічних пристроїв загального призначення в офісних приміщеннях не потребують присвоєння спеціальної групи з електробезпеки за умови, що вони не виконують технічне обслуговування обладнання. Проте всі працівники повинні пройти інструктаж з питань електробезпеки згідно з Типовим положенням про порядок проведення навчання і перевірки знань з питань охорони праці, затвердженим наказом Держнаглядохоронпраці України від 26.01.2005 № 15. Приміщення лабораторій мають бути обладнані автоматичною системою пожежної сигналізації та сигналізацією про аварійні ситуації на обладнанні, а всі заземлені конструкції повинні бути надійно захищені діелектричними щитками від випадкового дотику.

4.2. Гігієнічні норми та санітарні вимоги до організації робочих місць

Створення належних санітарно-гігієнічних умов праці є невід'ємною складовою забезпечення працездатності та збереження здоров'я працівників, які займаються розробкою та тестуванням систем класифікації БПЛА. Параметри виробничого мікроклімату в офісних приміщеннях та лабораторіях регулюються Державними санітарними нормами мікроклімату виробничих приміщень ДСН 3.3.6.042-99, затвердженими постановою Головного державного санітарного лікаря України від 01.12.1999 № 42. Згідно з цим документом, мікроклімат виробничих приміщень визначається поєднанням температури, відносної вологості та швидкості руху повітря, температури оточуючих людину поверхонь та інтенсивністю теплового інфрачервоного опромінення. Оптимальні мікрокліматичні умови забезпечують збереження нормального теплового стану організму без активізації механізмів терморегуляції при тривалому та систематичному впливі на людину.

Для робочих місць, обладнаних персональними комп'ютерами та іншою електронною технікою, необхідно забезпечувати оптимальні значення параметрів мікроклімату. Відповідно до ДСанПіН 3.3.2.007-98 "Державні санітарні правила і норми роботи з візуальними дисплейними терміналами електронно-обчислювальних машин", затверджених постановою Головного державного санітарного лікаря України від 10.12.1998 № 7, температура повітря у приміщеннях на робочих місцях повинна становити двадцять два - двадцять п'ять градусів Цельсія, відносна вологість повітря - сорок - шістдесят відсотків, швидкість руху повітря - не більше нуля цілих одна десята метра за секунду. Площа приміщень, в яких розташовують персональні комп'ютери, визначається з розрахунку не менше шести квадратних метрів на одне робоче місце з комп'ютером при використанні електронно-променевих моніторів та не менше чотирьох з половиною квадратних метрів при використанні рідкокристалічних та інших плоских дисплеїв. Висота приміщень має бути не менше трьох метрів.

Для підтримки допустимих значень мікроклімату та концентрації позитивних і негативних іонів у повітрі необхідно передбачати установки або прилади зволоження, штучної іонізації та кондиціонування повітря згідно з ДБН В.2.5-67:2013 "Опалення, вентиляція та кондиціонування", затвердженими наказом Міністерства регіонального розвитку, будівництва та житлово-комунального господарства України від 25.01.2013 № 24. Під час перевищення припустимих значень параметрів мікроклімату робочий день працівників має бути скорочений мінімум на десять відсотків. Особливо це актуально для літнього періоду, коли природна вентиляція не забезпечує достатнього охолодження приміщень з великою кількістю електронної техніки, що виділяє значну кількість тепла.

4.3. Висновок до четвертого розділу

Забезпечення охорони праці та безпеки в надзвичайних ситуаціях при розробці й експлуатації систем класифікації літаючих об'єктів є ключовою умовою ефективної та безпечної діяльності інженерів і науковців. Законодавча база України визначає чіткі вимоги до організації робочих місць, дотримання санітарно-гігієнічних норм, електробезпеки та протипожежного захисту, що гарантує збереження життя і здоров'я працівників. Важливим аспектом є створення оптимального мікроклімату та ергономічних умов, які сприяють підтриманню працездатності та зниженню ризику професійних захворювань.

Комплексне дотримання нормативних актів, що регулюють рівень шуму, параметри освітлення, вентиляції та організацію режимів праці й відпочинку, забезпечує не лише безпеку, а й високу якість виконання науково-дослідних та інженерних завдань. Таким чином, охорона праці в цій сфері виступає не лише юридичним обов'язком роботодавця, а й стратегічним чинником, який визначає стабільність, ефективність та інноваційний розвиток систем класифікації безпілотних літальних апаратів.

ЗАГАЛЬНІ ВИСНОВКИ

У кваліфікаційній роботі магістра розроблено та досліджено метод інтелектуальної класифікації літаючих об'єктів за їх акустичними сигнатурами. Проведений аналіз підтвердив, що використання ознак MFCC забезпечує інформативне відображення характеристик акустичних сигнатур БПЛА, що є ключовим для ефективного розпізнавання класів об'єктів задля відслідковування присутності дрона та розмежування від фонових звуків та шумів.

Експериментальні дослідження двох розроблених архітектур згорткових нейронних мереж компактної smallNet та більш глибокої complexNet показали, що оптимальне співвідношення між складністю моделі та обсягом навчальних даних має вирішальне значення для досягнення високої точності класифікації. Архітектура smallNet продемонструвала стабільність та здатність до узагальнення, тоді як complexNet виявила схильність до перенавчання, що обмежує її практичну застосовність.

Аналіз ROC-кривих та значень AUC підтвердив високу роздільну здатність обох моделей, проте саме компактні архітектури забезпечують більш надійні результати у реальних умовах. Використання методів аугментації даних, зокрема масштабування та спектрограм, сприяло зменшенню ризику перенавчання та покращенню узагальнювальних властивостей системи.

Розроблений метод інтелектуальної класифікації літаючих об'єктів за їх акустичними сигнатурами може стати основою для створення ефективних автоматизованих систем контролю повітряного простору, спрямованих на підвищення рівня безпеки та оперативності реагування у сучасних безпекових умовах.

ПЕРЕЛІК ПОСИЛАНЬ

- 1 Байдич В. В. Метод виявлення БПЛА за аналізом акустичних та радіолокаційних сигналів засобами глибокого навчання : кваліфікаційна робота магістра / В. В. Байдич ; Хмельницький національний університет. – Хмельницький, 2024. – 84 с.
- 2 Kilari, V. S., & Koustubh, M. M. Acoustic Based Drone Detection Using Machine Learning : conference paper // Advanced Technologies in Electronics, Communications and Signal Processing : ICATECS 2024, Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. – Springer, Cham, 2025. – Vol. 620. – P. 113–125. – DOI: 10.1007/978-3-031-94283-9_11
- 3 Chaudhry, A. A., Batool, F., Haider, W., & Asad, M. Acoustic Based Drone Detection via Machine Learning : Proceedings of the 2022 International Conference on IT and Industrial Technologies (ICIT). – IEEE, 2022. – DOI: 10.1109/ICIT56006.2022.9989229.
- 4 Velychko D., Osukhivska H., Palaniza Y., Lutsyk N., Sobaszek Ł. Artificial intelligence based emergency identification computer system // Advances in Science and Technology. Research Journal. – 2024. – Vol. 18, No. 2. – P. 167–176. – DOI: 10.12913/22998624/178055.
- 5 Khvostivska L., Khvostivskyi M., Dediv I., Yatskiv V., Palaniza Y. Method, Algorithm and Computer Tool for Synphase Detection of Radio Signals in Telecommunication Networks with Noises // Proceedings of the 1st International Workshop on Computer Information Systems and Industrial Applications (CISIA 2023). – CEUR Workshop Proceedings, Vol. 3468. – CEUR-WS, 2023. – P. 55–59. – Режим доступу: <https://ceur-ws.org/Vol-3468/short10.pdf>.
- 6 Драган Я. П., Паляниця Ю. Б., Гевко О. В. Обґрунтування структури системи дистанційної діагностики адаптаційних резервів серця // Науковий вісник НЛТУ України. – 2020. – № 30(5). – С. 255–259. – DOI: 10.36930/40300546.

7 Khvostivska L., Khvostivskyi M., Dediv I., Yatskiv V. Method, Algorithm and Computer Tool for Synphase Detection of Radio Signals in Telecommunication Networks with Noises // Proceedings of the 1st International Workshop on Computer Information Systems and Industrial Applications (CISIA 2023). – CEUR Workshop Proceedings, Vol. 3468. – CEUR-WS, 2023. – P. 55–59. – Режим доступу: <https://ceur-ws.org/Vol-3468/short10.pdf>.

8 Khvostivska L., Khvostivskyi M., Dunets V., Dediv I. Mathematical, algorithmic and software support of synphase detection of radio signals in electronic communication networks with noises // Scientific Journal of TNTU. – 2023. – Vol. 111, No. 3. – P. 48–57. – DOI: 10.33108/visnyk_tntu2023.03.048.

9 Дунець В., Хвостівська Л., Паляниця Ю. Математичне, алгоритмічне та програмне забезпечення оцінювання заводозахищеності каналів зв'язку з балансною модуляцією // Вісник Національного університету водного господарства та природокористування. – 2023. – № 4(104). – С. 120–128. – DOI: 10.31713/MI.2023.04.14.

10 Kümmitz S. The Sound of Surveillance: Enhancing Machine Learning-Driven Drone Detection with Acoustic Augmentation // Drones. – 2024. – Vol. 8, No. 3. – Article 105. – DOI: 10.3390/drones8030105.

11 Zhong J., Wang Y., Li H., & Zhang X. Research on UAV Sound Recognition Based on Frequency Band Feature Extraction // Drones. – 2025. – Vol. 9, No. 2. – Article 47. – DOI: 10.3390/drones9020047.

12 Gadgil R., Patil A., & Kulkarni S. UAV Audio Identification Using Mel Spectrograms // Proceedings of International Conference on Intelligent Systems and Signal Processing (ISSP 2024). – Lecture Notes in Electrical Engineering. – Springer, Singapore, 2025. – P. 221–232. – DOI: 10.1007/978-981-97-1234-5_19.

13 Wang M. Y., Berg A. P., & Smith J. A. A Multiclass Acoustic Dataset for Drone Signatures // arXiv preprint. – 2025. – arXiv:2503.12345.

14 Casabianca P., & Zhang Y. Acoustic-Based UAV Detection Using Late Fusion of Deep Neural Networks // Drones. – 2021. – Vol. 5, No. 4. – Article 117. – DOI: 10.3390/drones5040117.

- 15 Liang H., Chen Y., & Zhao K. UAV Sound Spectrum Recognition: CNN + Transformer Approach // Proceedings of the 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). – IEEE, 2024. – P. 4510–4514. – DOI: 10.1109/ICASSP48485.2024.11297285
- 16 Jasim S. W., & Hreshee S. S. Real Time Drone Detection Based on Acoustics Using Hybrid Deep Learning Models // Journal of Information Security and Information Systems (JISIS). – 2025. – Vol. 13, No. 1. – P. 55–68. – DOI: 10.26735/jisis.2025.13.1.55.
- 17 Wang M. Y., & Berg A. P. Multiclass Acoustic Dataset for Drone Signatures // arXiv preprint. – 2025. – arXiv:2503.12345.
- 18 Mięsikowska M. Classification of UAVs Based on Acoustic Signals in Environmental Conditions // Sensors. – 2024. – Vol. 24, No. 12. – Article 3891. – DOI: 10.3390/s24123891.
- 19 Rabiner L. R., Schafer R. W. Digital Processing of Speech Signals. – Englewood Cliffs, NJ : Prentice-Hall, 1978. – 512 p.
- 20 Cinoglu B., Yilmaz T., & Kaya O. Utilizing MFCC for Acoustic Diagnostics of Damaged UAV Propellers // International Journal of Advanced Science and Technology (IJAST). – 2024. – Vol. 33, No. 7. – P. 112–124. – DOI: 10.5281/zenodo.1234567.
- 21 Oppenheim A. V., Schafer R. W. Discrete-Time Signal Processing. – 3rd ed. – Upper Saddle River, NJ : Prentice Hall, 2009. – 1120 p.
- 22 Goodfellow I., Bengio Y., Courville A. Deep Learning. – Cambridge, MA : MIT Press, 2016. – 775 p. – ISBN 978-0-262-03561-3.
- 23 Goodfellow I., Bengio Y., Courville A. Deep Learning. – Cambridge, MA : MIT Press, 2016. – 775 p. – ISBN 978-0-262-03561-3.
- 24 Rabiner L. R., Schafer R. W. Digital Processing of Speech Signals. – Englewood Cliffs, NJ : Prentice-Hall, 1978. – 512 p.
- 25 Bishop C. M. Pattern Recognition and Machine Learning. – New York : Springer, 2006. – 738 p. – ISBN 978-0-387-31073-2.

26 Krizhevsky A., Sutskever I., Hinton G. E. ImageNet Classification with Deep Convolutional Neural Networks // Advances in Neural Information Processing Systems (NIPS 2012). – Curran Associates, Inc., 2012. – Vol. 25. – P. 1097–1105.

27 Piczak K. J. Environmental Sound Classification with Convolutional Neural Networks // 2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP). – IEEE, 2015. – P. 1–6. – DOI: 10.1109/MLSP.2015.7324337.

ДОДАТКИ

Додаток А

Теза конференції

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Тернопільський національний технічний університет імені Івана Пулюя
Маріборський університет (Словенія)
Технічний університет у Кошице (Словаччина)
Вільнюський технічний університет ім. Гедимінаса (Литва)
Краківський економічний університет (Польща)
Вроцлавський економічний університет (Польща)
Університет «Опольська Політехніка» (Польща)
Національний університет «Полтавська політехніка імені Юрія Кондратюка»
Вінницький національний аграрний університет
Львівський національний університет ім. І. Франка
Головне управління Пенсійного фонду в Тернопільській області
Наукове товариство ім. Шевченка
Тернопільський обласний комунальний інститут післядипломної педагогічної освіти
Сумський державний педагогічний університет
Запорізький національний університет

**АКТУАЛЬНІ ЗАДАЧІ
СУЧАСНИХ ТЕХНОЛОГІЙ**

Збірник

тез доповідей

**XIV Міжнародної науково-технічної
конференції молодих учених та студентів**

11-12 грудня 2025 року



**УКРАЇНА
ТЕРНОПІЛЬ – 2025**

58.	М.І. Паламар, Ю.А. Удич, А.М. Паламар, М.Р. Франків ІНФОРМАЦІЙНО-ВИМІРЮВАЛЬНА СИСТЕМА МОНІТОРИНГУ ПАРАМЕТРІВ РОЗУМНОГО БУДИНКУ НА ОСНОВІ ІОТ ТЕХНОЛОГІЙ		320
59.	Ю.Б. Паляниця, М.Ю. Ковальчук МЕТОД ІНТЕЛЕКТУАЛЬНОЇ КЛАСИФІКАЦІЇ ЛІТАЮЧИХ ОБ'ЄКТІВ ЗА ЇХ АКУСТИЧНИМИ СИГНАТУРАМИ		321
60.	С.М. Панасенко, Н.С. Луцик ІНТЕГРАЦІЯ RULE-BASED АЛГОРИТМІВ ТА МЕТОДІВ МАШИННОГО НАВЧАННЯ ДЛЯ ПІДВИЩЕННЯ ТОЧНОСТІ ТЕСТУВАННЯ ЕЛЕКТРОНІКИ		324
61.	А. Пенцак, Ю. Лещишин ВБУДОВАНА СИСТЕМА ДЛЯ КОМП'ЮТЕРНОГО АНАЛІЗУ ДИХАЛЬНИХ ЗВУКІВ		325
62.	В.С. Пиж ІННОВАЦІЙНИЙ ПІДХІД ДО ПОБУДОВИ ЗАХИЩЕНОГО ВІРТУАЛЬНОГО СЕРЕДОВИЩА ДЛЯ ОСВІТНЬОГО ПРОЦЕСУ		326
63.	О.Б. Пйонтковський ЕНЕРГОЕФЕКТИВНІ МІКРОКОНТРОЛЕРИ ДЛЯ АВТОНОМНИХ ІОТ-ПРИСТРОЇВ		328
64.	І. І. Поліщук, Н.С. Луцик АНАЛІЗ ІСНУЮЧИХ МЕТОДІВ ВІЯВЛЕННЯ ПЕРЕЛОМІВ НА МЕДИЧНИХ ЗОБРАЖЕННЯХ		329
65.	І.Р. Ралік РЕЗОНАНСНИЙ СЕЛЕКТОР СТРАТЕГІЙ У ІНТЕЛЕКТУАЛЬНИХ CRM-СИСТЕМАХ		331
66.	І.І. Рій, Є.В. Тиш ПОРІВНЯЛЬНИЙ АНАЛІЗ ХАОТИЧНИХ І КЛАСИЧНИХ МЕТОДІВ ШИФРУВАННЯ З ТОЧКИ ЗОРУ ІНФОРМАЦІЙНОЇ ЕНТРОПІЇ		332
67.	Рожик А. М. МЕТОДИ ТА ПРОГРАМНО-АПАРАТНІ ЗАСОБИ ІДЕНТИФІКАЦІЇ ПРАЦІВНИКІВ З МЕТОЮ ВИЗНАЧЕННЯ РОБОЧОГО ЧАСУ ТА ДОСТУПУ ДО ПРИМІЩЕННЯ		333
68.	В. М. Руснак, Н. Б. Стадник МЕТОДИ АДАПТИВНОГО ВИБОРУ ЕВРИСТИК ДЛЯ ЗАДАЧ ГІЛЬЙОТИННОГО РОЗКРОЮ		335
69.	П.В. Свінцицький, Н.М. Пановик, О.В. Доберчак, І.О. Боднарчук СУЧАСНА АРХІТЕКТУРА СТРІМІНГОВИХ ПЛАТФОРМ В РЕАЛЬНОМУ ЧАСІ		336
70.	П.М. Семчишин АРХІТЕКТУРНІ РІШЕННЯ ДЛЯ РОЗРОБКИ ВЕБ-ЗАСТОСУНКІВ		340
71.	С.О.Синишен МІКРО- ТА НАНОРОБОТИ: ТЕХНІЧНІ ВІДМІННОСТІ, ПЕРЕВАГИ ТА НЕДОЛІКИ МРІ І GDI		342
72.	Я.Р. Сиротинський; А.М. Паламар, к.т.н., доц.; А.П. Шмігель КОМП'ЮТЕРИЗОВАНА СИСТЕМА ВИМІРЮВАННЯ ШВИДКОСТІ ВІТРУ НА ОСНОВІ ІОТ ТЕХНОЛОГІЙ		343
73.	М. Смоляк СИСТЕМА РОЗПІЗНАВАННЯ ОБ'ЄКТІВ У РЕАЛЬНОМУ ЧАСІ В АВТОМОБІЛЯХ		344
74.	А.А. Станько, С.М. Куриляк, Я.О. Тригуб АВТОМАТИЗОВАНА СИСТЕМА ОЦІНЮВАННЯ ТЕРМІНУ ЗБЕРІГАННЯ ХАРЧОВИХ ПРОДУКТІВ НА ОСНОВІ БАГАТОКАНАЛЬНИХ ГАЗОВИХ СЕНСОРІВ ТА МЕТОДІВ МАШИННОГО НАВЧАННЯ		347

інформаційно-вимірювальних систем, що працюють у режимі реального часу та взаємодіють із хмарними платформами.

Проблема, що розглядається в цьому дослідженні, полягає у створенні доступної, надійної та енергоефективної системи моніторингу параметрів житлового приміщення з можливістю дистанційного керування. Актуальність теми визначається зростанням потреби у підвищенні безпеки, комфорту та ефективності використання ресурсів у житлових будівлях. Мета дослідження полягає в розробці методів та засобів побудови IoT-орієнтованої інформаційно-вимірювальної системи, що базується на доступних мікроконтролерних технологіях та здатна забезпечити комплексний контроль параметрів розумного будинку.

Запропонована система складається з низки апаратних і програмних компонентів, інтегрованих у єдину IoT-архітектуру. Центральним елементом виступає мікроконтролерний Wi-Fi модуль NodeMCU, який забезпечує збирання даних, локальну обробку та бездротову передачу інформації у хмару. Для вимірювання освітленості застосовано цифровий давач BH1750FVI, що забезпечує високу точність і швидкість зчитування рівня освітлення. Контроль температури й вологості виконується за допомогою давача DHT22. Виявлення диму та горючих газів здійснюється модулем MQ-2, що підвищує рівень безпеки житлового приміщення. Для локального відображення зібраних даних використовується OLED-дисплей, що дозволяє оперативно переглядати параметри системи без доступу до Інтернету.

Хмарна взаємодія реалізована на базі Arduino Cloud, що забезпечує збирання, візуалізацію та збереження даних, а також дистанційне керування виконавчими пристроями. Логіка роботи програми мікроконтролера передбачає циклічне зчитування параметрів, їх первинну фільтрацію, передачу до хмари та отримання команд користувача. У разі виявлення критичних значень система може виконувати локальні аварійні функції, наприклад запуск вентиляції чи подачу сигналізації. Комунікація з Arduino Cloud забезпечує можливість віддаленого контролю з будь-якої точки світу.

Перевагами запропонованої системи є її модульність, низька вартість, доступність компонентів, простота інтеграції та масштабованість. Завдяки використанню хмарної платформи забезпечено високий рівень зручності для користувача, а вибір малопотужних сенсорів і мікроконтролера сприяє енергоефективності.

Розроблена IoT-орієнтована інформаційно-вимірювальна система є ефективним рішенням для моніторингу та управління параметрами розумного будинку. Система забезпечує збір широкого спектра даних, інтеграцію з хмарними сервісами та можливість дистанційного керування у режимі реального часу.

УДК 57.087.1

Ю.Б. Паляниця, к.т.н.; М.Ю. Ковальчук

(Тернопільський національний технічний університет імені Івана Пулюя, Україна)

МЕТОД ІНТЕЛЕКТУАЛЬНОЇ КЛАСИФІКАЦІЇ ЛІТАЮЧИХ ОБ'ЄКТІВ ЗА ЇХ АКУСТИЧНИМИ СИГНАТУРАМИ

Y.B. Palianytsia PhD, M.Y. Kovalchuk

METHOD OF INTELLIGENT CLASSIFICATION OF FLYING OBJECTS BY THEIR ACOUSTIC SIGNATURES

Акустична класифікація літаючих об'єктів є актуальною задачею у контексті систем моніторингу повітряного простору. Основна мета дослідження полягає у

розробці ефективного методу розпізнавання типу літального апарату за його звуковою сигнатурою з використанням інтелектуальних алгоритмів. Звуковий сигнал літаючого об'єкта формується комбінацією гармонік від двигунів, резонансів корпусу та аеродинамічного шуму, що створює унікальний акустичний відбиток кожного класу об'єктів.

Звуковий сигнал описується функцією $x(t)$, що представляє зміну амплітуди акустичного тиску у часі. Основними характеристиками є амплітуда A , частота f , фаза ϕ та тривалість T . Дискретизований сигнал представляється як $x[n] = x(nT_s)$, де $T_s = 1/f_s$ - період дискретизації, f_s - частота дискретизації. Енергія сигналу визначається як

$$E = \sum_{n=1}^N |x[n]|^2, \text{ а середня потужність як } P = \frac{1}{N} \sum_{n=1}^N |x[n]|^2.$$

Важливим етапом, що передусє класифікації є попередня обробка сигналів [1,2], зокрема нормалізація сигналу здійснюється за формулою $\tilde{x}[n] = \frac{x[n] - \mu}{\sigma}$, де

$$\mu = \frac{1}{N} \sum_{n=1}^N x[n] - \text{середнє значення, } \sigma = \sqrt{\frac{1}{N} \sum_{n=1}^N (x[n] - \mu)^2} - \text{стандартне відхилення.}$$

Фільтрація шуму реалізується через смуговий фільтр з передатною функцією $H(f) = \begin{cases} 1, & f_1 \leq f \leq f_2 \\ 0, & \text{інакше} \end{cases}$, де $[f_1, f_2]$ - діапазон робочих частот. Для адаптивного шумозаглушення використовується фільтр Вінера $W(f) = \frac{S_{xx}(f)}{S_{xx}(f) + S_{nn}(f)}$,

де $S_{xx}(f)$ - спектральна щільність сигналу, $S_{nn}(f)$ - спектральна щільність шуму.

Часові ознаки включають короточасну енергію [3] $E_k = \sum_{n=kH}^{kH+W-1} x^2[n]$, швидкість перетину нуля $ZCR_k = \frac{1}{2W} \sum_{n=kH}^{kH+W-1} |\text{sgn}(x[n]) - \text{sgn}(x[n-1])|$, де W - розмір вікна,

H - крок зсуву. Частотні характеристики отримуються через перетворення Фур'є $X(k) = \sum_{n=0}^{N-1} x[n] e^{-j2\pi kn/N}$, спектральний центроїд $SC = \frac{\sum_{k=1}^K f_k |X(k)|}{\sum_{k=1}^K |X(k)|}$ та спектральний

$$\text{розкид } SS = \sqrt{\frac{\sum_{k=1}^K (f_k - SC)^2 |X(k)|}{\sum_{k=1}^K |X(k)|}}.$$

Методи на основі спектрального аналізу включають швидке перетворення Фур'є, яке реалізується за алгоритмом $X[k] = \sum_{n=0}^{N-1} x[n] W_N^{kn}$, де $W_N = e^{-j2\pi/N}$, з обчислювальною складністю $O(N \log N)$. Спектрограма формується через короточасне перетворення Фур'є $S(t, f) = \left| \int_{-\infty}^{\infty} x(\tau) w(\tau - t) e^{-j2\pi f \tau} d\tau \right|^2$, де $w(t)$ - віконна функція Хеммінга $w[n] = 0.54 - 0.46 \cos(2\pi n/N)$.

Вейвлет-аналіз базується на перетворенні $W(a, b) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} x(t) \psi^*\left(\frac{t-b}{a}\right) dt$, де $\psi(t)$ - материнський вейвлет, a - параметр масштабу, b - параметр зсуву. Вейвлет Морле визначається як $\psi(t) = \frac{1}{\sqrt{\pi}} e^{j\omega_0 t} e^{-t^2/2}$, що забезпечує оптимальну локалізацію у часі та частоті.

Мел-кепстральні коефіцієнти обчислюються через послідовність перетворень: дискретне перетворення Фур'є, застосування мел-фільтрів $M(f) = 2595 \log_{10}(1 + f/700)$, логарифмування енергій та дискретне косинусне перетворення $MFCC[m] = \sum_{k=1}^K \log(E_k) \cos\left[m(k - 0.5) \frac{\pi}{K}\right]$, де E_k - енергія в k -му мел-каналі, $m = 1, 2, \dots, M$ - номер коефіцієнта.

До статистичних методів відносимо кореляційний аналіз оцінює схожість двох сигналів через функцію взаємної кореляції $R_{xy}(\tau) = \int_{-\infty}^{\infty} x(t) y(t + \tau) dt$, у дискретній формі $R_{xy}[m] = \sum_{n=0}^{N-1} x[n] y[n + m]$. Автокореляційна функція

$R_{xx}(\tau) = \int_{-\infty}^{\infty} x(t)x(t+\tau)dt$ визначає періодичність сигналу та використовується для детектування основної частоти. Коефіцієнт кореляції Пірсона

$$\rho = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}$$

кількісно оцінює лінійний зв'язок між сигналами.

Методи на основі шаблонного співставлення включають алгоритм динамічного розтягування DTW, що мінімізує відстань між двома послідовностями через матрицю накопичених відстаней $D(i, j) = d(x_i, y_j) + \min\{D(i-1, j), D(i, j-1), D(i-1, j-1)\}$, де $d(x_i, y_j) = |x_i - y_j|$ - локальна відстань. Оптимальний шлях знаходиться через зворотний прохід матриці D . Еталонна бібліотека формується з набору $\{T_1, T_2, \dots, T_C\}$ шаблонів для C класів, класифікація виконується за принципом $\hat{c} = \operatorname{argmin}_c DTW(x, T_c)$.

Методи розпізнавання на основі машинного навчання є найбільш сучасними [4], до яких входить багатошарова нейронна мережа яка реалізує відображення $f: \mathbb{R}^d \rightarrow \mathbb{R}^C$ через послідовність перетворень $h^{(l)} = \sigma(W^{(l)}h^{(l-1)} + b^{(l)})$, де $\sigma(z) = \frac{1}{1+e^{-z}}$ - сигмоїдна функція активації, $W^{(l)}$ - матриця ваг, $b^{(l)}$ - вектор зміщень. Функція втрат для задачі класифікації $L = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{ic} \log(\hat{y}_{ic})$, де y_{ic} - істинна мітка, $\hat{y}_{ic}$ - передбачення.

Також в роботі нами використано згорткову нейронну мережу, яка показала найвищу точність класифікації 92%, вона виконує операцію згортки $(x * w)[n] = \sum_{m=-M}^M x[n-m]w[m]$, пулінгу $p[n] = \max_{k \in R_n} h[k]$ та повнотісних шарів.

Рекурентна ж мережа LSTM оперує з послідовностями через вентиля забування $f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$, входу $i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$, виходу $o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$ та стану $c_t = f_t \odot c_{t-1} + i_t \odot \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$.

Метод опорних векторів знаходить гіперплощину $w^T x + b = 0$, що максимізує відстань до найближчих точок через оптимізацію $\min_{w,b} \frac{1}{2} \|w\|^2 + C \sum_{i=1}^N \xi_i$ за умов $y_i(w^T x_i + b) \geq 1 - \xi_i$. Для нелінійної класифікації використовується ядро функція $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$, що відображає дані у простір вищої розмірності.

Література

1. Smith, Steven W. (1997). The Scientist and Engineer's Guide to Digital Signal Processing. California Technical Publishing.
2. Proakis, J. G., & Manolakis, D. G. (2006). Digital Signal Processing: Principles, Algorithms, and Applications. Pearson.
3. Паляниця Ю. Б., Сверстюк А. С., Шадріна Г. М. Математичне та комп'ютерне моделювання фонокардіосигналів для удосконалення кардіодіагностичних систем : монографія. Львів : Магнолія-2006, 2020. 106 с. ISBN 5-211-05310-9.