

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

Факультет комп'ютерно-інформаційних систем і програмної інженерії
(повна назва факультету)

Кафедра комп'ютерних систем та мереж
(повна назва кафедри)

КВАЛІФІКАЦІЙНА РОБОТА

на здобуття освітнього ступеня

Магістр

(назва освітнього ступеня)

на тему: **Методи та засоби комп'ютеризованого відстеження міграції птахів на основі аналізу звукових сигналів**

Виконав: студентка 6 курсу, групи СІм-62
спеціальності 123 «Комп'ютерна інженерія»

(шифр і назва спеціальності)

Сороківська Н.В.

(підпис)

(прізвище та ініціали)

Керівник

ЯЦИШИН В.В.

(підпис)

(прізвище та ініціали)

Нормоконтроль

Луцик Н.С.

(підпис)

(прізвище та ініціали)

Завідувач кафедри

Осухівська Г.М.

(підпис)

(прізвище та ініціали)

Рецензент

Деркач М.В.

(підпис)

(прізвище та ініціали)

Тернопіль
2024

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

Факультет комп'ютерно-інформаційних систем і програмної інженерії
(повна назва факультету)

Кафедра комп'ютерних систем та мереж
(повна назва кафедри)

ЗАТВЕРДЖУЮ
Завідувач кафедри
Осухівська Г.М.
(підпис) (прізвище та ініціали)
«11» листопада 2024 р.

ЗАВДАННЯ НА КВАЛІФІКАЦІЙНУ РОБОТУ

на здобуття освітнього ступеня магістр
(назва освітнього ступеня)

за спеціальністю 123 «Комп'ютерна інженерія»
(шифр і назва спеціальності)

студенту Сороківській Надії Віталіївній
(прізвище, ім'я, по батькові)

1. Тема роботи Методи та засоби комп'ютеризованого відстеження міграції птахів на основі аналізу звукових сигналів

Керівник роботи Яцишин Василь Володимирович, к.т.н., доц.
(прізвище, ім'я, по батькові, науковий ступінь, вчене звання)

Затверджені наказом ректора від « 29 » листопада 2024 року № 4/7-1144

2. Термін подання студентом завершеної роботи 16.12.2024

3. Вихідні дані до роботи Масив даних у вигляді аудіозаписів формату .wav обсерваторії Tautenburg, De та інформація щодо його вмісту, методи штучного інтелекту для класифікації масивів даних

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

Вступ

1 Аналітичний огляд предметної області, засоби опрацювання аудіоданих та їх оцифрування

2 Розробка алгоритмів та проектування системи

3 Аналіз і узагальнення результатів практичних експериментів

4 Охорона праці та безпека в надзвичайних ситуаціях

Висновки

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

1. Актуальність і мета дослідження.

2. Завдання, об'єкт і предмет, наукова новизна і практична цінність дослідження.

3. Процеси опрацювання аудіоданих та їх оцифрування

4. Порівняння методів моніторингу міграцій птахів

5. Алгоритм робити над вирішенням завдання

6. Аналіз метаданих отриманих аудіосигналів

7. Порівняння методів машинного навчання для вирішення класу завдань

8. Висновки

6. Консультанти розділів роботи

Розділ	Прізвище, ініціали та посада консультанта	Підпис, дата	
		завдання видав	завдання прийняв
<i>Охорона праці</i>	<i>Осухівська Г.М., зав.каф. КС</i>	<i>05.12.2024</i>	
<i>Безпека в надзвичайних ситуаціях</i>	<i>Стручок В.С., ст. викладач каф. ОХ</i>		

7. Дата видачі завдання 11.11.2024

КАЛЕНДАРНИЙ ПЛАН

№ з/п	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1.	<i>Аналіз існуючих методів та засобів розв'язання науково-технічних проблем, що відповідають тематиці кваліфікаційної роботи.</i>	<i>20.11.2024р. – 30.11.2024р.</i>	<i>Викон.</i>
2.	<i>Аналітичний огляд області досліджень</i>	<i>01.12.2024р – 03.12.2024р.</i>	<i>Викон.</i>
3.	<i>Розробка алгоритмів та проектування системи</i>	<i>04.12.2024р – 07.12.2024р.</i>	<i>Викон.</i>
4.	<i>Аналіз і узагальнення результатів практичних експериментів</i>	<i>07.12.2024р – 12.12.2024р.</i>	<i>Викон.</i>
5.	<i>Охорона праці та безпека в надзвичайних ситуаціях</i>		<i>Викон.</i>
6.	<i>Оформлення пояснювальної записки і графічного матеріалу</i>	<i>15.12.2024р.</i>	<i>Викон.</i>
7.	<i>Попередній захист кваліфікаційної роботи магістра</i>	<i>16.12.2024</i>	<i>Викон.</i>
8.	<i>Захист кваліфікаційної роботи магістра</i>	<i>23.12.2024</i>	<i>Викон.</i>

Студент

(підпис)*Сороківська Н.В.*

(прізвище та ініціали)

Керівник роботи

(підпис)*Яцишин В.В.*

(прізвище та ініціали)

АНОТАЦІЯ

Сороківська Н.В. «Методи та засоби комп'ютеризованого відстеження міграції птахів на основі аналізу звукових сигналів». Робота на здобуття кваліфікаційного ступеня магістра: спец. 123 — комп'ютерна інженерія / наук. кер. В.В. Яцишин. Тернопіль: Тернопільський національний технічний університет імені Івана Пулюя, 2024. — 81 с.

Ключові слова: пташиний спів, спектрограма, штучний інтелект, CQT, IoT, моніторинг, опрацювання аудіосигналів, опрацювання даних.

Кваліфікаційна робота присвячена дослідженню методів розпізнавання пташиного співу в умовах підвищеного шумового фону за допомогою сучасних методів аналізу звуку. У роботі здійснено порівняння різних підходів до обробки аудіоданих, зокрема використання мел-спектрограм і CQT-спектрограм, а також аналіз ефективності навчання моделей із використанням попередньо натренованих і власноруч налаштованих нейронних мереж. Система базується на концепції Інтернету речей, що дозволяє збирати та обробляти дані в режимі реального часу.

Отримані результати демонструють потенціал застосування таких систем у моніторингу біорізноманіття та екологічному аналізі, а також підвищують ефективність реагування на надзвичайні ситуації, спричинені природними катастрофами. Робота підкреслює перспективи інтеграції таких систем у малогабаритні комп'ютерні пристрої для забезпечення автономного функціонування в польових умовах.

ANNOTATION

Sorokivska N.V. "Methods and tools for computerized tracking of bird migration based on sound signal analysis" Master's Thesis: specialty 123 — Computer Engineering / Academic Advisor: V.V. Yatsyshyn. Ternopil: Ternopil Ivan Puluj National Technical University, 2024. — 81 pages.

Keywords: bird song, spectrogram, artificial intelligence, CQT, IoT, monitoring, audio signal processing, data processing.

This master's thesis focuses on studying methods for recognizing bird songs in high-noise environments using modern sound analysis techniques. The work compares different approaches to audio data processing, including the use of mel spectrograms and CQT spectrograms, and evaluates the training efficiency of models using both pretrained and custom-configured neural networks. The system is based on the Internet of Things concept, enabling real-time data collection and processing.

The results highlight the potential application of such systems in biodiversity monitoring and ecological analysis, as well as in improving response efficiency to emergencies caused by natural disasters. The thesis emphasizes the prospects for integrating such systems into compact computing devices to ensure autonomous operation in field conditions.

ЗМІСТ

ВСТУП.....	7
РОЗДІЛ 1	10
АНАЛІТИЧНИЙ ОГЛЯД ОБЛАСТІ ДОСЛІДЖЕНЬ.....	10
1.1. Методи роботи із аудіосигналами у комп'ютеризованих системах	10
1.2. Методи отримання та передачі даних для систем моніторингу.....	15
1.3. Огляд наявних рішень для поставленого класу задач	23
1.4. Висновки до розділу	28
РОЗДІЛ 2	29
РОЗРОБКА АЛГОРИТМІВ РОБОТИ І ПРОЕКТУВАННЯ СИСТЕМИ.....	29
2.1. Опис алгоритму опрацювання даних	29
2.2. Ефективні застосування моделей ШІ для роботи із аудіоданими.....	38
2.3. Виклики у аналізі та опрацюванні наборів даних.....	43
2.4 Висновки до розділу	46
РОЗДІЛ 3	47
АНАЛІЗ І УЗАГАЛЬНЕННЯ РЕЗУЛЬТАТІВ ПРАКТИЧНИХ ЕКСПЕРЕМЕНТІВ	47
3.1. Застосовані методи попереднього опрацювання даних та аналіз їх ефективності	47
3.2. Застосовані моделі та аналіз їх результатів.....	54
3.3. Оцінка результатів та перспективи використання.....	57
3.4. Висновки по розділу	59
РОЗДІЛ 4	60
ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ	60
4.1 Охорона праці	60
4.2. Вплив надзвичайних ситуацій природного характеру на надійність роботи засобів відстеження.....	61
ВИСНОВКИ.....	66
СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ	68
ДОДАТКИ.....	71
Додаток А	72
Додаток Б.....	78

ВСТУП

Актуальність роботи. Інтеграція алгоритмів машинного навчання у вивченні біорізноманіття та моніторингу екосистем відкриває нові можливості для аналізу великих обсягів даних без необхідності значних людських ресурсів чи спеціальної підготовки. Це дозволяє проводити точніші й масштабніші дослідження екологічного стану середовища та поведінки організмів, сприяючи збереженню природних ресурсів. Поєднання біології й сучасних технологій дозволяє вирішувати наукові задачі з якісно новим підходом, що робить дослідження вкрай актуальними.

Мета дослідження. Дослідити шляхи підвищення точності і автоматизації ідентифікації видів птахів із використанням аналізу аудіо файлів засобами нейронних мереж і сучасних методів машинного навчання. Проаналізувати як дана робота сприятиме покращенню наукових досліджень у галузі орнітології та збереження біорізноманіття.

Задачі дослідження:

1. Проаналізувати сучасні методи збору й обробки даних для екологічного моніторингу. Провести огляд і дослідження сучасних моделей штучного інтелекту, таких як BirdNet та інші відомі системи, що застосовуються для ідентифікації видів птахів.

2. Дослідити можливості застосування алгоритмів машинного навчання для обробки аудіоданих та ефективність існуючих моделей у контексті розпізнавання нічних міграційних покликів птахів. Визначити, чи краще використовувати наявні моделі, або створити нову, що враховує специфіку нічного моніторингу.

3. Розробити модель для автоматичного розпізнавання пташиних голосів, яка враховує реальні акустичні умови.

4. Провести оцінку ефективності розробленої системи на основі отриманих результатів. Дослідити потенційні напрямки впровадження розробленої моделі для масштабованого моніторингу міграцій птахів, зокрема

для інтеграції у великі мережі датчиків або використання науково-дослідницькими установами та організаціями, що займаються збереженням біорізноманіття.

Об'єкт дослідження: Процес автоматичного моніторингу біорізноманіття за допомогою обробки звукових сигналів.

Предмет дослідження: Методи та алгоритми машинного навчання для аналізу спектрограм пташиних голосів.

Методи дослідження:

1. Аналіз літератури – для вивчення сучасних тенденцій у застосуванні штучного інтелекту для моніторингу біорізноманіття.
2. Обробка аудіоданих – використання спектральних методів (наприклад, постійна-Q трансформація) для отримання спектрограм звукових сигналів.
3. Моделювання – застосування нейромереж, зокрема CNN і RNN, для класифікації звукових сигналів.
4. Експериментальний аналіз – оцінка точності й продуктивності моделі на тестових даних.

Наукова новизна дослідження:

- уперше запропоновано застосування адаптованої архітектури глибоких нейронних мереж для розпізнавання пташиних голосів із урахуванням незбалансованості даних та впливу фонових шумів, що дало змогу вдосконалити існуючі методи аудіоаналізу в умовах реальних екосистем;
- за рахунок запропонованої моделі архітектури нейронних мереж, набули подальшого розвитку комп'ютеризовані системи збору та інтелектуального аналізу аудіосигналів, що дало змогу забезпечити більш високу ефективність розпізнавання птахів під час їх міграції

Практичне значення результатів. Результати роботи можуть бути використані для автоматизованого моніторингу популяцій птахів, дослідження екологічних змін і розробки стратегій для збереження біорізноманіття. Розроблені методики також можуть бути інтегровані у системи моніторингу природного середовища.

Публікації. Результати дипломної роботи апробовані на XIII міжнародній науково-технічній конференції молодих учених і студентів «Актуальні задачі сучасних технологій» (11-12 грудня 2024 р.) Тернопільського національного технічного університету імені Івана Пулюя та на XII науково-технічній конференції Тернопільського національного технічного університету імені Івана Пулюя «Інформаційні моделі, системи та технології» (18-19 грудня 2024 року) у вигляді тез конференцій.

1. Сороківська Н.В., Яцишин В.В. Використання адаптованих моделей штучного інтелекту для ідентифікації птахів у природних умовах. Матеріали XIII міжнародної науково - технічної конференції молодих учених і студентів «Актуальні задачі сучасних технологій» (11-12 грудня 2024 р.) Тернопільського національного технічного університету імені Івана Пулюя. Тернопіль: ТНТУ. 2024. С. 478.

2. Сороківська Н.В., Яцишин В.В. Дослідження актуальності моніторингу біорізноманіття за допомогою комп'ютеризованих систем. Матеріали XII науково технічної конференції Тернопільського національного технічного університету імені Івана Пулюя «Інформаційні моделі, системи та технології» (18-19 грудня 2024 року). Тернопіль: ТНТУ. 2024. С. 142.

Структура роботи. Робота складається з пояснювальної записки та графічної частини. Пояснювальна записка складається із вступу, 4 розділів, висновків, списку використаних джерел та додатків. Обсяг роботи: пояснювальна записка – 81 аркушів формату А4, графічна частина – 8 аркушів формату А1.

РОЗДІЛ 1

АНАЛІТИЧНИЙ ОГЛЯД ОБЛАСТІ ДОСЛІДЖЕНЬ

1.1. Методи роботи із аудіосигналами у комп'ютеризованих системах

Аудіосигнали, що генеруються джерелами звуку і передаються через середовище, зокрема повітря або воду, можуть бути записані за допомогою мікрофонів і перетворені в цифрову форму для подальшого аналізу. Обробка аудіосигналів у комп'ютерних системах – це складний багатоступеневий процес, що охоплює перетворення звукових даних у цифровий формат, їх обробку, аналіз і подальше використання, наприклад, для класифікації чи розпізнавання.

Сигнал який є поширюваним у середовищі коливанням, реєструється мікрофоном. Його початкова форма – аналогова, тобто безперервна у часі та значеннях амплітуди. Для обробки на комп'ютері він перетворюється у цифрову форму через дискретизацію у часі та квантування амплітуди. Під час дискретизації сигнал вимірюється через рівні проміжки часу, де частота дискретизації повинна бути вдвічі вищою за максимальну частоту сигналу відповідно до теореми Найквіста, щоб уникнути втрати інформації. Наприклад, для людського слуху, який охоплює діапазон 20 Гц-20 кГц, використовується частота дискретизації 44,1 кГц. Квантування амплітуди полягає у переведенні значень сигналу у дискретні рівні з певною розрядністю, що дозволяє отримати точне числове представлення сигналу.

Звукові хвилі під час поширення залежать від властивостей середовища, таких як густина, температура чи тиск, і характеризуються явищами поглинання, розсіювання, фоновим шумом і дифракцією. Поглинання посилюється для високочастотних хвиль, що призводить до їх швидкого згасання, а наявність фонових шумів може ускладнити подальшу обробку сигналу. Частотні характеристики сигналів відіграють важливу роль у їхній класифікації. Наприклад, пташині вокалізації зазвичай лежать у межах чутного діапазону (20 Гц – 20 кГц), але аналіз може охоплювати також ультразвукові чи інфразвукові сигнали, залежно від об'єкта дослідження[7].

Процес цифрової обробки аудіосигналів включає кілька ключових етапів. Перший етап, як зазначалось раніше, представляє перетворення аналогового сигналу на цифровий. Процес дискретизації представляє рис. 1.1, на якому показано як безперервний сигнал розбивається на послідовність дискретних значень у часі, і квантування, при якому амплітуда сигналу переводиться у скінченну кількість рівнів. Ці процеси дозволяють зберігати сигнал у вигляді цифрових даних, придатних для комп'ютерного аналізу.

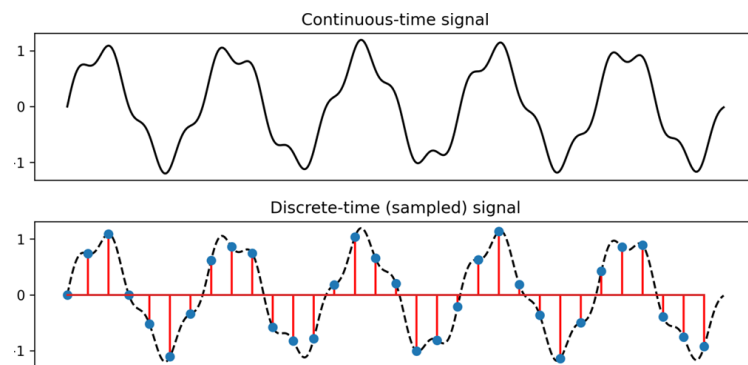


Рис. 1.1. Дискретизація аудіосигналу для опрацювання у КС[7]

За частотою дискретизації визначається, як часто трапляються амплітуди сигналу, що відповідно впливає на точність відтворення звуку. Для збереження інформації про вихідний сигнал використовується теорема Найквіста-Шеннона, яка вимагає, щоб частота дискретизації була вдвічі більшою за максимальну частоту сигналу.

Після перетворення сигналу у цифровий формат він репрезентується у часовій, частотній або час-частотній областях. У часовій області сигнал відображається як графік амплітуди у часі, однак ця форма не дозволяє оцінити частотний склад сигналу. Для цього використовують спектральний аналіз за допомогою швидкого перетворення Фур'є, що розкладає сигнал на частотні компоненти. Ефективна реалізація цього методу – швидке перетворення Фур'є (FFT), яке широко використовується завдяки своїй швидкості й точності. Однак варто зазначити, що це перетворення дає лише глобальне уявлення про сигнал і не враховує часової мінливості частотних компонентів. Для розв'язання цієї проблеми використовуються віконні методи аналізу, які розбивають сигнал на

короткі інтервали часу для детальнішого вивчення локальних особливостей, як представлено на рис. 1.2. У такій трансформації висока частотна точність на довгих вікнах і висока часова точність на коротких вікнах. Тобто фіксоване вікно — компроміс між точністю часу та частоти.

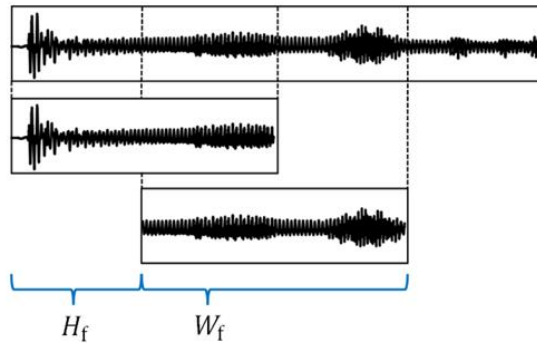


Рис. 1.2. Опрацювання аудіо із розбиттям на інтервали

Після попередньої обробки здійснюється аналіз спектральних характеристик сигналу. Для цього широко застосовуються методи, такі як Швидке Перетворення Фур'є (FFT) або перетворення Мел-частотної кепстральної характеристики (MFCC). FFT дозволяє отримати амплітудно-частотний спектр сигналу, тоді як MFCC — це популярний метод, який використовується для вилучення ознак, що імітують людське сприйняття звуку.

Спектрограма ж більш інформативна, оскільки поєднує часову та частотну інформацію і створюється за допомогою короткочасного перетворення Фур'є або вейвлет-перетворення, де вісь часу, частоти та амплітуди відображається у вигляді кольорової інтенсивності. Вейвлети дають спектрограму, де точність змінюється залежно від масштабу і на низьких частотах (довгі хвилі) вікно довше, а отже точніше видно частотну інформацію, натомість у високих частотах (короткі хвилі) вікно коротше і краще видно часові зміни.

У порівнянні методів цього етапу опрацювання звуків, як "стандартний" варіант з однаковим вікном для всіх частот, можна виокремити STFT. Процес опрацювання сигналів із цим підходом зображено на рис. 1.3. Ця трансформація добре підходить для сигналів з постійними частотами і коли немає великої варіативності.

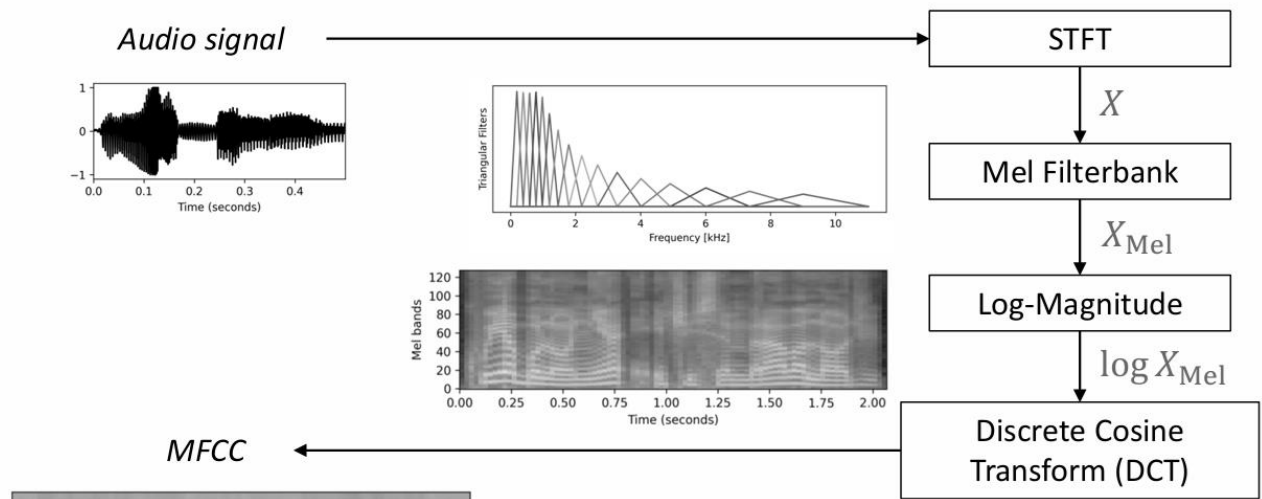


Рис. 1.3. Процес опрацювання звуку із STFT[7]

Рисунок показує загальний процес обробки аудіосигналів для отримання MFCC (Mel-Frequency Cepstral Coefficients). Далі застосовується короткочасне перетворення Фур'є для переходу від часового представлення до частотного представлення сигналу і спектр сигналу пропускається через фільтри, розташовані за Mel-шкалою, що відповідає людському сприйняттю звуку, амплітуди перетворюються у логарифмічний масштаб для відображення енергії сигналу. Далі здійснюється дискретне косинусне перетворення для зменшення кількості коефіцієнтів та усунення кореляції. Для опрацювання сигналів також можливі застосування дещо інших методів таких як вейвлет-перетворення та constant-q.

Вейвлет-перетворення краще адаптоване для нестационарних сигналів, де частоти швидко змінюються у часі (наприклад, звуки птахів, вибухи), оскільки метод дозволяє досліджувати їх одночасно у часовій та частотній областях. На відміну від класичного перетворення Фур'є, яке дає інформацію лише про частотний склад сигналу без прив'язки до часу, вейвлет-перетворення забезпечує локалізований аналіз, тобто дозволяє визначати, у які моменти часу виникають певні частоти.

Constant-Q перетворення також може бути використане, оскільки вбачає у собі, що співвідношення між частотою та шириною вікна має бути постійне. А отже, на низьких частотах CQT дає високу частотну роздільну здатність й низьку часову, а на високих частотах — навпаки, що добре для музики та природних

звуків (наприклад, пташиного співу). Тож така спектрограма буде зручна для звуків, де важлива логарифмічна шкала частот, у біоакустичних дослідженнях чи аналізі музики оскільки забезпечить потрібну у деяких випадках деталізацію на низьких частотах, але "розмиє" високі частоти.

На наступному етапі виконується подальша обробка сигналу, що критично важливо для підвищення якості даних перед аналізом. Цей етап спрямований на очищення сигналу від завад, нормалізацію параметрів і виділення корисних особливостей, які слугують основою для точного розпізнавання або класифікації. Аудіосигнали, особливо ті, що записані в природних умовах, часто містять значний рівень фонового шуму — це може бути вітер, інші тварини, транспорт або технічні перешкоди. Для видалення шуму застосовуються цифрові фільтри. Низькочастотні фільтри дозволяють відсікати високочастотні завади, а високочастотні фільтри допомагають видаляти низькочастотний гул або вібрації. Смугові фільтри ізолюють корисний діапазон частот, характерний для цільового сигналу. Наприклад, для звуків пташиного співу можна обмежити частоти в межах діапазону в якому вони трапляються у наборі даних, оскільки ці дані визначені. Також часто застосовуються методи спектрального віднімання, де шумова компонента оцінюється і віднімається від сигналу. Для більш складних випадків використовують адаптивні фільтри та методи машинного навчання, що можуть навчатися на фоні та цільовому сигналі.

Оскільки амплітуда сигналів може значно варіюватися через зміни відстані до джерела звуку, гучність або характеристики запису, виконується нормалізація. Цей процес приводить амплітуду сигналу до єдиного масштабу, що дозволяє порівнювати різні записи та уникати спотворень у подальшому аналізі. Наприклад, амплітуда може бути масштабована до діапазону від -1 до 1.

При порівнянні кількох аудіосигналів важливо забезпечити їх часову синхронізацію. Для цього використовують методи перетворення часової шкали (Time Stretching) або віконного вирівнювання, щоб сигнали починалися з однакової точки відліку. Це особливо важливо для автоматичного розпізнавання голосів або побудови моделей на основі нейронних мереж.

Після очищення та нормалізації сигналу виконується виділення ключових характеристик, які описують аудіосигнал. Мел-кепстральні коефіцієнти, про які йшлося раніше – найпопулярніший метод, що забезпечує компактне представлення звуку у вигляді сету коефіцієнтів, які описують спектральні особливості сигналу.

Таким чином, після отримання спектрограм аудіосигнали проходять обробку, яка включає видалення шуму, нормалізацію, фільтрацію, виділення ознак та побудову моделей для їх класифікації чи розпізнавання. Цей процес забезпечує високу точність аналізу й адаптацію системи до реальних умов, де звук може бути нечітким чи зашумленим.

Основою для автоматизованого аналізу звуків у біоакустичних дослідженнях є обробка аудіосигналів включно з ідентифікацією видів тварин на основі їхніх вокалізацій. Сучасні комп'ютеризовані системи дозволяють вирішувати складні завдання аналізу аудіоданих, навіть у випадках наявності сильного фонового шуму чи варіативності сигналів між особинами одного виду. Однак, ці завдання супроводжуються низкою викликів, таких як необхідність обробки великих обсягів даних, обмеженість доступу до якісно анотованих записів і врахування фізичних властивостей поширення звуку в природному середовищі.

1.2. Методи отримання та передачі даних для систем моніторингу

Для досягнення ефективності у відстеженні птахів під час нічних перельотів використовують кілька різних технологій, кожна з яких має свої переваги та обмеження. Основні методи включають радар, спостереження «із місяцем», освітлювальні установки, інфрачервоні камери та запис польотних звуків. Ці методи порівнюються за такими параметрами, як визначення напрямку польоту, швидкості, висоти, кількості птахів та їх видів.

На базах офіційних обсерваторій використовуються високоточні мікрофонні системи та комплекси з кількох пристроїв запису, які можуть працювати в різних діапазонах частот. Такі системи зазвичай включають

стаціонарні масиви мікрофонів, які забезпечують високоточний збір звукових даних із великих площ. Вони дозволяють локалізувати джерела звуку у просторі завдяки триангуляції або ж методами обробки часових затримок сигналів. У місцях де найбільш очікувані спостереження перелітних птахів, як ось поблизу озер, на яких птахи зупиняються чи вздовж поширених маршрутів міграцій, встановлюються високочутливі мікрофони для реєстрації сигналів у різних діапазонах: від інфразвуку до ультразвуку (вище 20 кГц, як для моніторингу кажанів). Такі мікрофони мають низький рівень шуму та забезпечують максимальну деталізацію сигналу.

Раніше для проведення моніторингу використовувались прості методи спостереження за птахами на тлі місяця. Цей метод дозволяв орнітологам отримувати обмежену інформацію про напрямок та швидкість птахів, але не дозволяє визначати висоту їх польоту або кількість птахів. Він також не дає можливості ідентифікувати види птахів.

Використання потужних прожекторів може допомогти відстежувати птахів у нічному небі. Цей метод добре працює для визначення напрямку та висоти, проте менш ефективний для оцінки швидкості та кількості птахів. Також поширене застосування радарів, такий підхід один із найефективніших інструментів для моніторингу нічних польотів птахів. Він дозволяє точно визначати напрямок польоту, швидкість, висоту, а також може оцінити кількість птахів у повітрі. Однак радар, так само як і прожектор, не здатний ідентифікувати види птахів, оскільки працює переважно на основі фізичних характеристик руху об'єктів, а не їх біологічних ознак.

Інфрачервоні камери використовуються для виявлення теплових слідів птахів у нічному небі. Вони ефективні для визначення напрямку польоту, проте їх обмеження стосуються визначення швидкості, висоти та кількості птахів. Цей метод також не дозволяє ідентифікувати види птахів. Таким чином, для вирішення саме завдання із ідентифікації та класифікації перелітних птахів, метод запису звуків птахів – єдиний, що дозволяє з великою точністю визначати види птахів, оскільки багато видів мають унікальні вокалізації. Однак варто зазначити, що цей метод не дає інформації про напрямок, швидкість, висоту або

кількість птахів у польоті, тому для повноти аналізу отримуваних даних потрібно продумувати алгоритми для майбутньої інтеграції цього методу із, наприклад, даними радарів тощо. Порівняльна характеристика розглянутих методів приведена на рис. 1.4.

Bird monitoring methods @ night					
	direction	velocity	altitude	numbers	species
Radar	+	+	+	o	-
Moonwatching	o	o	o	o	-
Sky beamer	+	o	+	o	-
Infrared cameras (thermal)	+	o	o	o	-
Flight call recording	-	-	-	o	+

Рис. 1.4. Порівняння методів відстеження птахів

На рисунку представлено, що найбільший обсяг даних забезпечують радарні технології, що робить цей метод надзвичайно популярним у дослідженнях міграції птахів. Один із лідерів у цій сфері – Швейцарська орнітологічна станція в Земпаху (Swiss Ornithological Station Sempach), яка є ключовим європейським центром із застосування радарного моніторингу. У своїй роботі станція активно використовує метеорологічні радари для спостереження за нічними перельотами птахів.

Основна перевага радарного моніторингу полягає у можливості безперервного відстеження міграції на значних висотах, особливо в нічний час, коли традиційні візуальні методи стають неефективними. Завдяки радарам можна отримати важливі параметри, такі як швидкість, висота та напрямок руху птахів, що дозволяє детально вивчати міграційні шляхи та вплив навколишніх умов на їх поведінку. Приклад застосування радарів представлено на рис. 1.5.

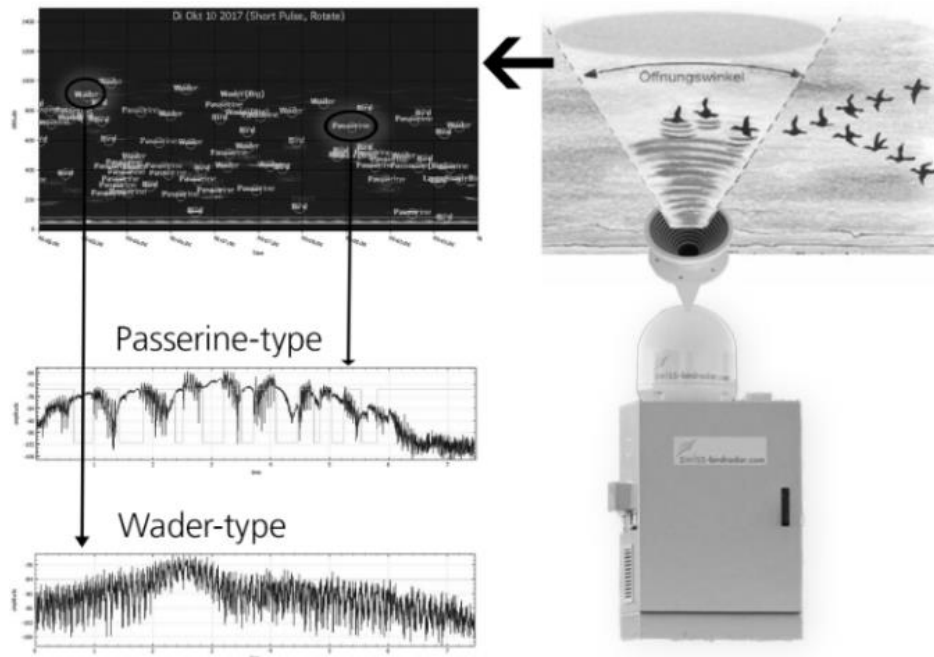


Рис. 1.5. Аналіз методів збору даних із радарів

Попри те, що радары не можуть ідентифікувати конкретні види птахів, їх ефективність зростає при поєднанні з іншими методами, наприклад, акустичними записами, що допомагає створити більш повну картину міграційних процесів. Крім того, Швейцарська орнітологічна станція активно досліджує наслідки змін клімату та антропогенних факторів на міграційні маршрути, використовуючи дані, отримані завдяки радарам. Це дозволяє оперативно збирати, аналізувати та інтерпретувати великі обсяги інформації для наукових цілей.

Однак масштабування таких станцій у більшій кількості обмежене через кілька чинників, як от вартість обладнання, оскільки високоточні мікрофони, масиви рекордерів та системи обробки даних що входять у склад однієї такої станції може сягати десятків або навіть сотень тисяч доларів. Формат зібраних результатів часто призводить до того що отримувані обсерваторіями записи залишаються неопрацьованими зокрема і через свій об'єм. Тривалі записи з високою точністю потребують високопродуктивних серверів для зберігання та обробки, що неможливе без значних ресурсів та кваліфікованих спеціалістів, які можуть інтерпретувати результати. Наявність кваліфікованих спеціалістів також сповільнює процеси досліджень, оскільки часто для налізу даних отриманих із

обсерваторій задіюються експерти вузьких галузей, оскільки на даному етапі, ще відсутні методи автоматичної інтерпретації матеріалів. Черговою проблемою виступає енергозалежність більшості із пристроїв, оскільки радари та сервери для опрацювання отримуваних даних потребують постійного підключення та стабільності у джерелах енергозабезпечення.

Таким чином, обсерваторії для збору звукових сигналів у великих масштабах – потужні інструменти для біоакустичних досліджень, але їх висока вартість, складність налаштування та потреба у фахівцях обмежують можливість широкого використання. Такі системи оптимальні для точкових досліджень у ключових природоохоронних регіонах, де необхідно забезпечити максимальну точність даних.

Для зменшення витрат процесу дослідження, поєднуються великі наукові обсерваторії та спеціалізовані менші станції із пристроями для аналізу звукових сигналів, що дозволяє створити ефективну систему моніторингу перелітних птахів, приклад такого обладнання зображено на рис. 1.6.



Рис. 1.6. Обладнання для віддаленого запису

Використання автономних аудіорекордери, як от AudioMoth, забезпечує можливість тривалого збору звукових даних у природних умовах. Такі пристрої можуть бути розміщені на значній території для охоплення різних міграційних шляхів, працюючи в режимі автоматичного запису. Програмне налаштування дозволяє оптимізувати роботу рекордерів: записувати сигнали у визначені періоди, як-от уночі, коли активність птахів найвища, зменшуючи таким чином обсяги непотрібних даних та енергоспоживання.

Розподілені обчислювальні системи та мобільні мережі, зокрема технології Multi-Access Edge Computing (MEC), дозволяють обробляти дані безпосередньо на місці їх збору. Контролери для мікрофонів можуть автоматично фільтрувати аудіосигнали, відсіюючи сторонній шум та непотрібні звуки. Це дає змогу передавати лише релевантну інформацію, наприклад потенційні фрагменти із вмістом потрібних звуків, із певного діапазону частот тощо, для подальшої перевірки спеціалістами, зменшуючи навантаження на мережу та обсяг збережених даних. Алгоритми обробки в реальному часі також здатні розпізнавати ключові події чи звуки, що робить процес більш ефективним і швидким.

У масштабних наукових проєктах аудіопристрої можуть використовуватися спільно з радарними системами обсерваторій для комплексного аналізу. У цьому випадку аудіозаписи доповнюють зібрані дані, дозволяючи ідентифікувати види птахів за їхніми голосами. Зібрані аудіофайли можуть аналізуватися за допомогою алгоритмів машинного навчання та акустичних моделей, що автоматично розпізнають звуки окремих видів.

Для локального застосування моніторингові пристрої можуть мати дещо інший масштаб проте зберігати свою ефективність та слугувати першочерговій цілі. Із розвитком сфери інтернету речей, появились можливості створювати бюджетні компактні та доступні локальні моніторингові станції, які можуть використовуватись у польових умовах, на приміських територіях або в приватних секторах. Такі системи підходять для ентузіастів, любителів птахів і невеликих дослідницьких проєктів завдяки своїй доступності, енергоефективності та можливості автономної роботи.

Одними з найпопулярніших платформ для створення таких систем є Raspberry Pi та Arduino. Ці пристрої дозволяють інтегрувати мікрофони, сенсори та модулі для передачі даних. Цифрові мікрофони з інтерфейсами I²C та I²S можуть забезпечити високу якість запису та стійкість до шуму в складних умовах навколишнього середовища, мікроконтролери та одноплатні комп'ютери Raspberry Pi Zero або ESP32 використовуються для керування датчиками та обробки отриманих даних у реальному часі, а мережеві технології передачі даних, зокрема: протоколи LoRaWAN і MQTT – дозволяють передавати дані на великі відстані при мінімальному енергоспоживанні. Це робить можливим збір інформації у віддалених регіонах з обмеженим доступом до традиційних мереж.

Такі аудіосистеми можуть автоматично записувати та передавати звукові сигнали у центральну базу або хмарне сховище, де відбувається їх аналіз. Алгоритми штучного інтелекту можуть ідентифікувати звуки конкретних видів птахів та фільтрувати фонові шуми, що дозволяє отримати чисті дані про активність птахів на певній території.

Окрім звукових пристроїв, у локальних станціях активно застосовуються сенсори навколишнього середовища для моніторингу природних умов. Такі пристрої широко використовуються в агропромисловості, але можуть бути адаптовані для досліджень природи:

- датчики температури та вологості (DHT22, BME280) – дозволяють отримувати дані про мікроклімат на території досліджень, що важливо для вивчення впливу середовища на активність птахів;
- датчики освітленості (BH1750) – допомагають фіксувати зміну освітлення, що корисно для розуміння поведінки нічних видів;
- датчики тиску повітря – використовуються для аналізу погодних умов, що можуть впливати на міграційні маршрути;
- датчики рівня шуму – дозволяють оцінити фоновий шум, що особливо актуально у приміських районах з високим рівнем антропогенного впливу.

Мережі передачі даних для локальних систем LoRaWAN (Long Range Wide Area Network) та MQTT (Message Queuing Telemetry Transport) виконують функції обміну даними, різниця між цими протоколами у радіусі їх дії, тож в

залежності від умов в яких розміщатимуться пристрої можна використовувати один із них. LoRaWAN підтримує дальню передачу даних, до кількох кілометрів у відкритій місцевості, з мінімальним енергоспоживанням та ідеально підходить для пристроїв, які працюють у польових умовах автономно на акумуляторах чи сонячних панелях. MQTT може бути використаний якщо потрібна безперервна передача. Цей протокол полегшує обмін даними між пристроями та центральними серверами і дозволяє швидко передавати невеликі пакети даних, що важливо для обмежених каналів зв'язку. Ще одним із варіантів – використання вже наявних Wi-Fi та мобільні мереж (3G/4G/5G). Вони можуть бути інтегровані у систему в приміських районах або місцях із доступом до інтернет-мережі для миттєвої передачі даних на хмарні платформи.

Такий варіант використання має можливість набути масштабного застосування завдяки ентузіасти можуть налаштовувати локальні системи моніторингу у власних садах, на сільськогосподарських полях чи в природних зонах. Мікрофони, підключені до Raspberry Pi, можуть записувати звуки вночі для аналізу активності видів у певний період року, а використання датчиків температури та вологості допомагатимуть відстежувати зміни кліматичних умов і їхній вплив на біорізноманіття. Датчики шуму та освітленості надаватимуть корисну інформацію, для визначення рівня впливу людської діяльності на середовище існування птахів.

Використання сенсорів та IoT-девайсів не є чимось новим у дослідженнях, деякі із них також набули цінність у промислових застосуваннях. Наприклад, сенсори вологості ґрунту, такі як YL-69 або Decagon, використовуються для оптимізації поливу в агросекторі, пристрої для моніторингу мікроклімату в теплицях або відкритих полях – допомагають регулювати зрошення та інші процеси, компактні системи з сенсорами температури, вимірюють вологість та тиск для моніторингу погодних умов у реальному часі, а пристрої, що відслідковують зростання, стан ґрунту та рівень освітлення, допомагають з оптимізацією вирощування культур.

Використання малогабаритних пристроїв таких, як локальні IoT-системи для моніторингу природи має ряд переваг:

- можливість тривалої роботи на сонячних батареях або акумуляторах;
- використання недорогих компонентів, що дозволяє ентузіастам створювати власні системи;
- легкість інтеграції нових сенсорів та можливість модернізації систем;
- обробка та передача лише релевантної інформації, що знижує навантаження на мережі.

Таким чином, локальні IoT-системи на основі малогабаритних плат і сенсорів – доступне та ефективне рішення для моніторингу птахів і навколишнього середовища. Вони відкривають широкі можливості для ентузіастів, дослідників і малих громадських проєктів, дозволяючи з мінімальними ресурсами отримувати цінні дані про стан природних екосистем.

1.3. Огляд наявних рішень для поставленого класу задач

Існує низка спеціалізованих платформ та інструментів для збору й аналізу звукових сигналів птахів, які використовуються для ідентифікації видів, основні з яких показано на рис. 1.7. Один із найвідоміших – це BirdNet, система розпізнавання пташиних голосів, що працює на основі глибокого навчання. BirdNet використовує аудіодані, зібрані з відкритих джерел, включно з записами ентузіастів та користувачів, які завантажують звуки птахів на різні платформи. Аналогічно працює інструмент Merlin від Корнельської лабораторії орнітології, що також базується на відкритих даних і призначений для автоматичної ідентифікації голосів птахів у природному середовищі.

Raven Pro і Kaleidoscope Pro – програмні інструменти для аналізу біоакустичних даних, які дозволяють обробляти записи та класифікувати види за допомогою попередньо налаштованих алгоритмів. Крім цього, платформа Xeno-Santo слугує онлайн-базою даних, де ентузіасти з усього світу завантажують записи пташиних голосів з інформацією про їхнє місцезнаходження, час та видову належність.



Рис. 1.7. Огляд готових рішень

Хоча ці платформи й інструменти і корисні для загального моніторингу пташиних голосів, їхні дані мають суттєві обмеження для конкретного завдання дослідження міграційних кличів. По-перше, отримані дані – неточні, оскільки вони зібрані здебільшого не професійними орнітологами, а аматорами та ентузіастами. Це призводить до ймовірних помилок у визначенні видів та маркуванні записів. По-друге, моделі, натреновані на цих даних, орієнтовані на звичайний спів птахів, що кардинально відрізняється від їх ж міграційних кличів, які коротші, монотонніші та мають іншу акустичну структуру. Таким чином, використання таких інструментів для дослідження міграційних кличів не надто неефективне. По-третє, при практичному тестуванні цих систем було виявлено, що вони погано справляються зі сторонніми шумами, такими як природні звуки вітру, дощу, комах, або техногенний шум, що часто присутні в записах, отриманих у польових умовах. Шумові перешкоди значно знижують точність роботи моделей, які не були адаптовані для подібних умов. Результати роботи попередньо налаштованої моделі BirdNet на отриманому із орнітологічної станції наборі даних представлено на рис. 1.8.

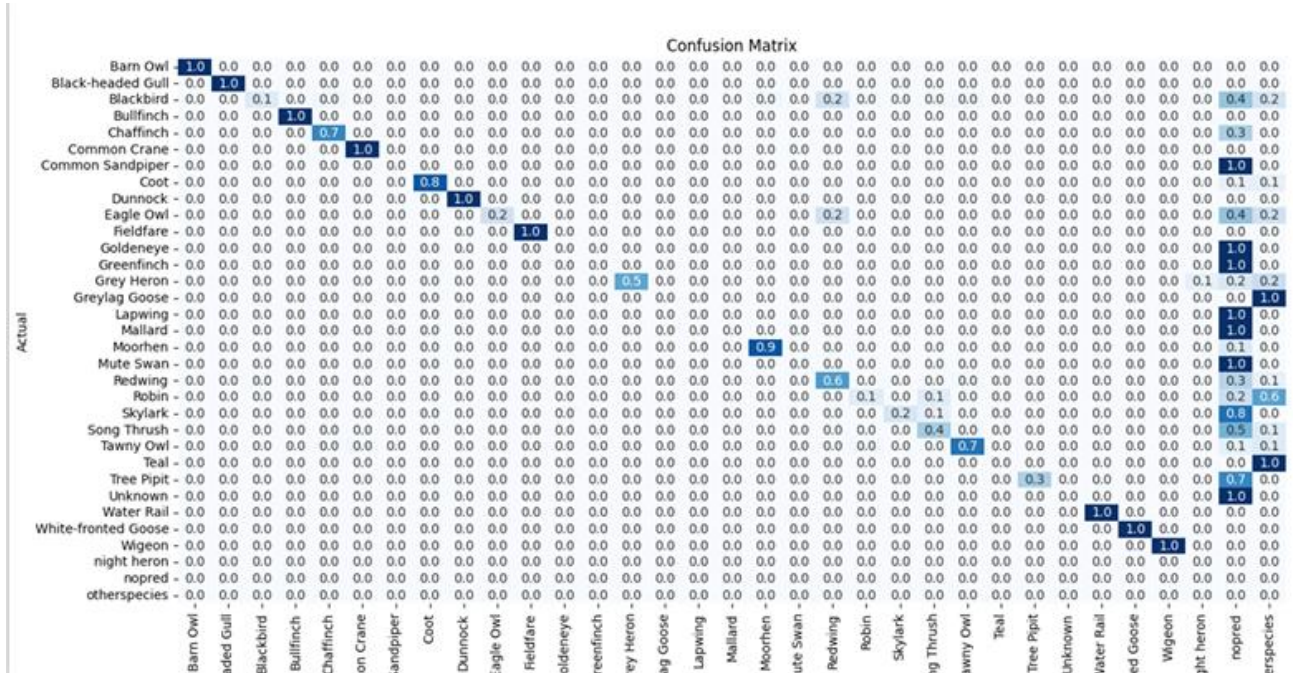


Рис. 1.8. Результати передбачення готової моделі

Отже, платформи, що використовують відкриті дані, такі як BirdNet, Merlin, Kaleidoscope Pro чи записи з Xeno-Canto, не можуть забезпечити необхідної якості для ідентифікації міграційних кличів. Їхні моделі потребують додаткового навчання на спеціалізованих датасетах, що містять записи саме міграційних звуків птахів, а також повинні бути оптимізовані для роботи в умовах значного фонових шумів та обмеженої якості аудіоданих. Також, з рисунку видно, що модель справляється із ідентифікацією лише деяких, доволі унікальних класів птахів, проте не справляється із більшістю даних що отримані із датасету. Хоче це і вказує у першу чергу на те, що ця модель непристосована для виконуваного класу завдань, результати теж свідчать про нерепрезентативність датасету, і вказує на необхідність модифікації даних різними методами препроцесінгу.

На разі доступні дані із однієї обсерваторії, вони обмежені як за обсягом, так і за кількістю видів птахів, що фіксуються на аудіозаписах. Такі дані можуть містити позначки про конкретні види, що дозволяє здійснити початкове маркування для навчання проміжних моделей. Проте через обмежену кількість зразків і нерівномірне представлення різних типів птахів ці моделі будуть потребувати подальшого донавчання або перевчання на значно більших і

різноманітніших датасетах. Розподіл поточного набору представлений на рис. 1.9. Зокрема, важливо зібрати записи з інших обсерваторій чи регіонів для розширення обсягу даних і забезпечення репрезентативності модельованих звуків.

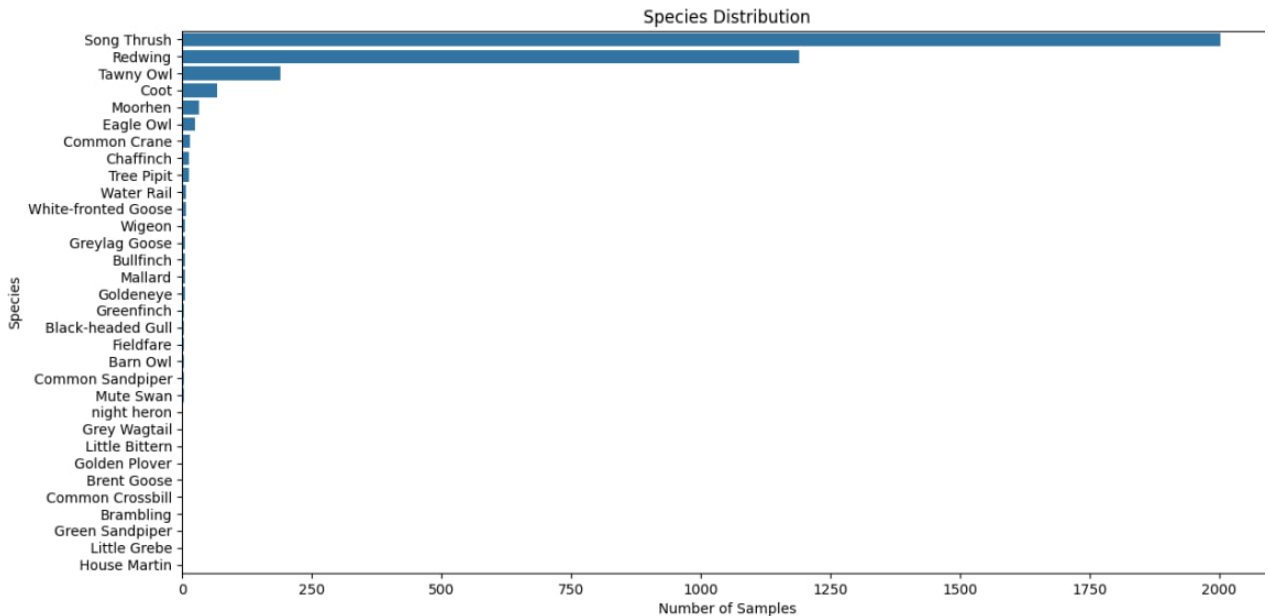


Рис. 1.9. Отримані дані для вирішення завдання

У процесі моніторингу перелітних птахів за допомогою аудіопристроїв отримуються різноманітні дані, що складаються зі звукових записів певної тривалості. Зазвичай тривалість таких звукових фрагментів залежить від налаштувань системи збору: вона може варіюватися від кількох секунд до хвилин, залежно від активності середовища, пропускної здатності системи та обсягу доступної пам'яті. У випадку довготривалих спостережень, таких як нічний моніторинг птахів, записи часто виконуються у вигляді безперервних фрагментів протягом визначених часових проміжків, наприклад, у нічні години. Отримані дані містять як звуки цільових об'єктів, тобто голоси птахів різних видів, так і значну кількість фонового шуму, який може включати звуки вітру, дощу, комах, технічний шум від людської діяльності або інші природні та антропогенні сигнали. На цьому етапі записи є "сирими" і потребують попередньої обробки для відокремлення релевантних сигналів. У результаті

аналізу отриманого набору даних бачимо розподіл тривалостей записів, як показано на рис. 1.10.

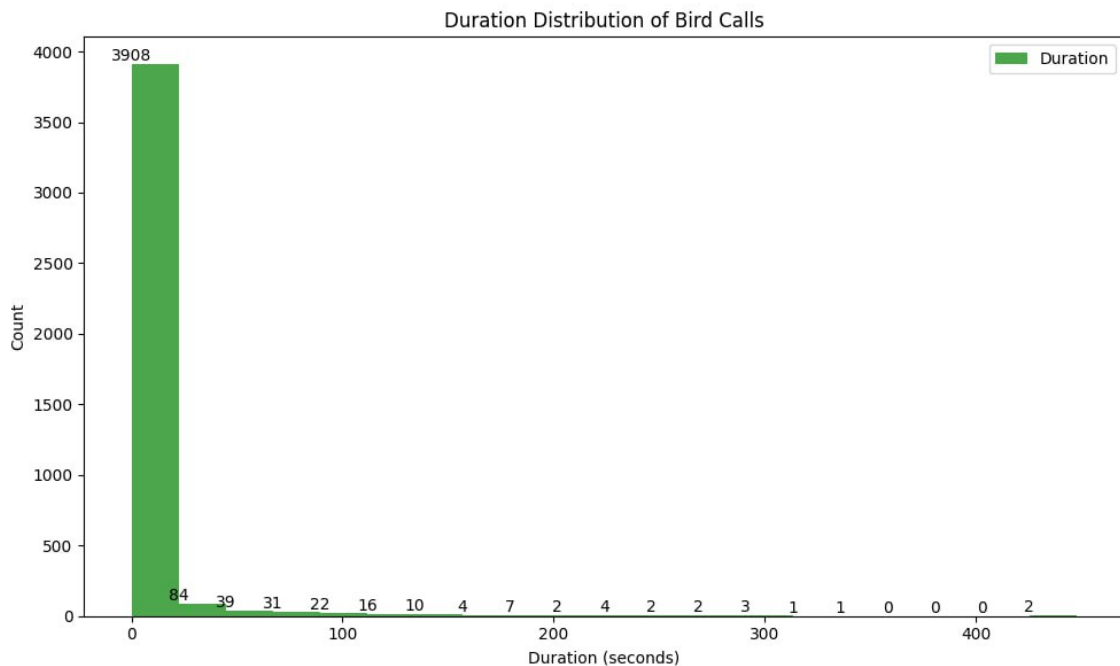


Рис. 1.10. Розподіл тривалості записуваних даних

З рисунку видно, що записувані дані здебільшого мають тривалість менше 1 хвилини, проте трапляються також записи тривалістю більше 3 хвилин, що доволі розбалансовує дані та ускладнює їх подальше опрацювання. Для ефективного використання даних у процесі тренування моделей машинного навчання необхідно забезпечити їх однакову тривалість, оскільки наявні записи мають значний розрив у часових рамках від коротких відрізків тривалістю 10 секунд до великих записів понад 3 хвилини у процесі роботи з ними виникатимуть серйозні проблеми як на етапі попередньої обробки даних, так і під час їх використання для навчання моделей. А можливість обрізання файлів доволі обмежена, хоча видається найпростішим способом нормалізації даних, оскільки призводить до ризику втрати ключової інформації. Якщо ми випадково обріжемо саме той момент, де зафіксований міграційний клич птаха, це знецінить увесь запис, що критично позначиться на якості тренування моделі.

1.4. Висновки до розділу

Для задачі розпізнавання міграційних ключів, які часто короткі, слабкі та розташовані нерівномірно в часі, обрізання особливо небезпечне, оскільки такі сигнали можуть бути єдиною корисною частиною запису.

Проте, на основі таких проміжних даних можливе тестування різних підходів для аналізу й обробки звукових сигналів. Ефективними можуть виявитися методи автоматичного розпізнавання звуків на основі алгоритмів машинного навчання, зокрема нейронних мереж, таких як згорткові нейронні мережі (CNN), які добре працюють із акустичними спектрограмами[8]. Попереднє перетворення звуків у спектрограми дозволяє перетворити аудіодані у візуальні представлення, що полегшує подальше розпізнавання та класифікацію. Також важливо застосувати методи фільтрації шуму для зменшення впливу небажаних сигналів і покращення якості вхідних даних для аналізу. У перспективі тестування різних архітектур моделей дозволить визначити, які саме підходи будуть найбільш ефективними для подальшого опрацювання більш масштабних і складних датасетів. Таким чином, отримані обмежені дані хоча і виступають важливим етапом для розробки методології обробки й аналізу, проте лише проміжним і їх необхідно розширювати та адаптувати для побудови повноцінних моделей на основі великої кількості репрезентативних записів у подальшому.

РОЗДІЛ 2

РОЗРОБКА АЛГОРИТМІВ РОБОТИ І ПРОЕКТУВАННЯ СИСТЕМИ

2.1. Опис алгоритму опрацювання даних

Структурований підхід важливий при роботі із науковими завданнями, оскільки він забезпечує послідовність, чіткість і ефективність на кожному етапі роботи та допомагає із деталізацією та описом кроків для відтворення результатів та їх масштабування у подальшому. Діаграма із рис. 2.1 демонструє логічний ланцюг процесів — від збору даних до їх обробки та валідації. Поділ завдання на окремі етапи (збір сигналів, оцифрування, попередня обробка, аналіз) дозволяє оптимізувати роботу, визначити слабкі місця й застосувати відповідні методи для їх вирішення.

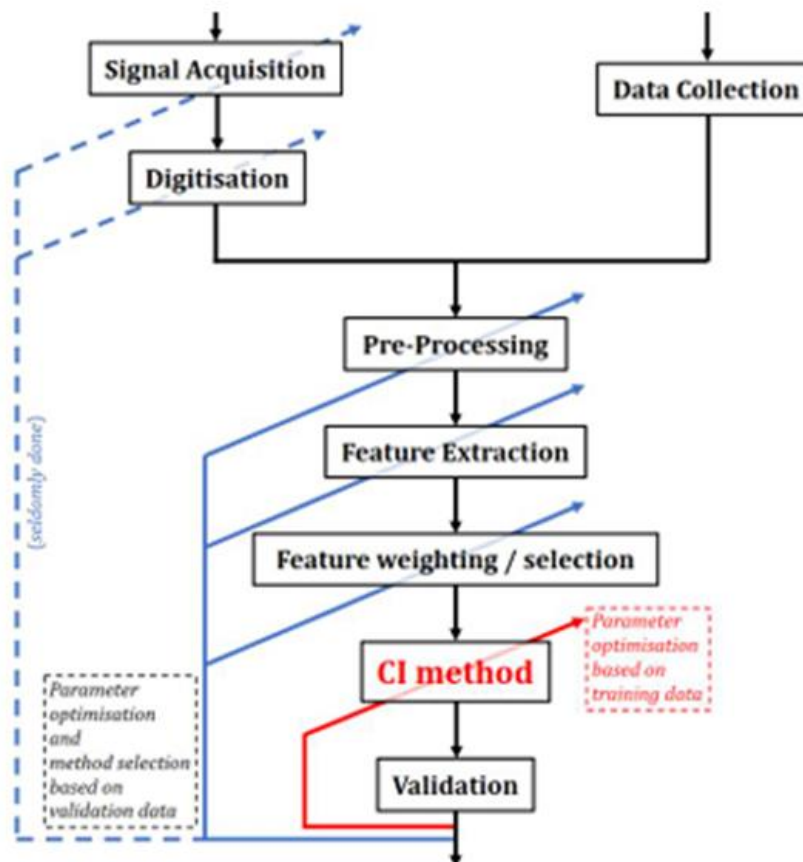


Рис. 2.1. Алгоритм роботи із завданням[8]

Такий підхід дозволяє впорядкувати вхідні дані для точного аналізу, покроково усувати проблеми через те що кожен етап чітко визначає аспекти які потрібно вдосконалити або виправити. Розділення завдання на підпроцеси допомагає краще розподілити обчислювальні ресурси та час.

Етапи роботи над завданням відповідно до представленої схеми включають такі ключові позиції, як збір даних (Data Collection), зчитування та оцифрування сигналів (Signal Acquisition та Digitisation). На першому етапі відбувається аналітичний збір інформації щодо мігруючих птахів. Цей процес охоплює кілька напрямів:

- географічний аналіз та поширення птахів – збираються дані про види птахів, які здійснюють міграцію через певні регіони, зокрема територію Тюрингії (Німеччина), де розташована обсерваторія із якої отримано спостереження. Дані включають інформацію про місця розташування, чисельність популяцій та типові маршрути їх переміщення;
- сезонні дані – аналізуються часові рамки міграцій, що дозволяє визначити пікові періоди активності конкретних видів. Ця інформація допоможе оптимізувати роботу системи збору аудіоданих;
- популяційні характеристики – включає інформацію про чисельність популяцій, видове різноманіття та зміни у їх динаміці. Дані можуть бути зібрані з відкритих орнітологічних джерел, а також наукових публікацій.

Отримані аналітичні дані являють собою основу для налаштування обладнання та визначення пріоритетних параметрів для збору аудіосигналів. Вони дозволяють зрозуміти, на які види птахів слід орієнтуватися, у який час найкраще проводити спостереження та як адаптувати алгоритми для специфічних міграційних кличів.

Паралельний етап із зображеної схеми – зчитування та оцифрування сигналів (Signal Acquisition та подальший – Digitisation). Тут відбувається безпосередній збір аудіоданих за допомогою мікрофонів, встановлених у польових умовах або на території обсерваторії. Сам процес поділяється на кілька кроків:

– фіксація сигналів – мікрофони фіксують звукові сигнали у навколишньому середовищі, що включають як цільові звуки (міграційні кличі птахів), так і природні шуми (вітер, дощ, інші тварини тощо). Важливо використовувати цифрові мікрофони з високою чутливістю, здатні працювати в умовах зовнішнього середовища та фіксувати слабкі сигнали, ці умови якраз забезпечені на орнітологічних та інших дослідницьких станціях, тому, в порівнянні із методами які застосовані у раніше описаних моделях, можна вважати дані бездоганними, тобто як такі що не містять хибних класифікацій, оскільки зібрані науковими дослідниками та, багаторазово перевірені;

– оцифрування сигналів (Digitisation) – аналогові звукові сигнали перетворюються у цифрові дані за допомогою аналогово-цифрових перетворювачів (ADC). На цьому етапі важливими параметрами є частота дискретизації (sampling rate) і бітова глибина (bit depth), які впливають на точність та обсяг даних. Частота дискретизації повинна бути достатньо високою для збереження якості аудіосигналів, щоб не втратити характерні ознаки кличів птахів (наприклад, 44.1 кГц або вище). Бітова глибина ж визначає динамічний діапазон сигналу — важливо фіксувати як гучні, так і слабкі звуки;

– збереження у цифрових форматах – оцифровані сигнали записуються у форматах, зручних для обробки комп'ютерними системами (наприклад, WAV або FLAC), оскільки ці формати забезпечують мінімальну втрату даних. Збереження у форматах із компресією, як MP3, не є прийнятним через втрату важливих акустичних характеристик.

Окрім цього, можна використовувати автоматичне маркування часу для кожного запису, що дозволяє пов'язувати отримані сигнали з конкретними часовими інтервалами, що важливо для аналізу міграційних патернів. Оцифрування та вигляд оцифрованих даних у графічному представленні подано на рис. 2.2.

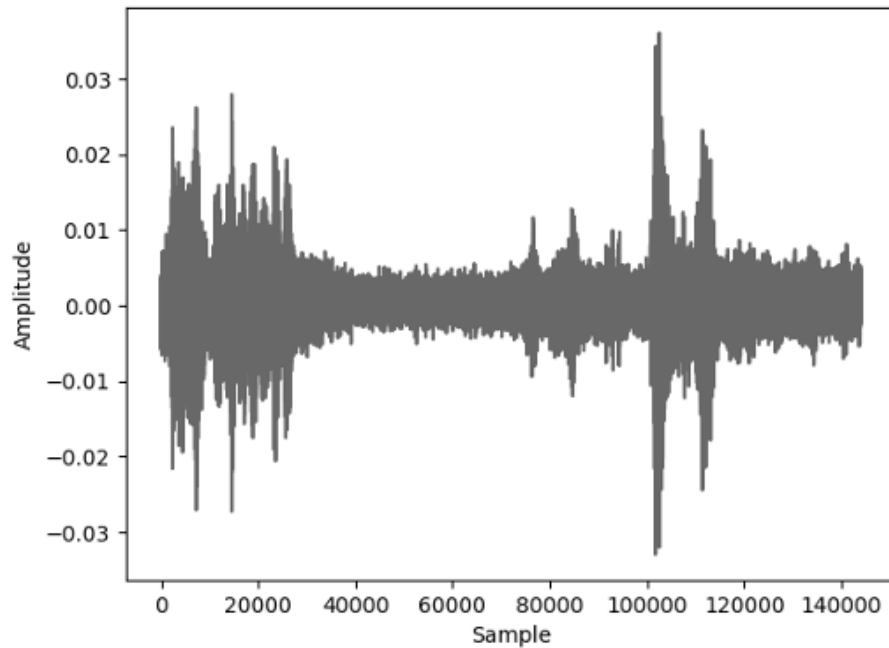


Рис. 2.2. Представлення результатів звуку із вхідних аудіо

Після збору та оцифрування сигналів відбувається попередня обробка, необхідна для підготовки даних до подальшого аналізу та застосування методів машинного навчання. Етап препроцесінгу – один із найбільш критичних і трудомістких у роботі над підготовкою даних для моделей машинного навчання, оскільки від нього залежить якість подальшого навчання та ефективність самої моделі. Весь процес попереднього працювання даних представлено на рис. 2.3.

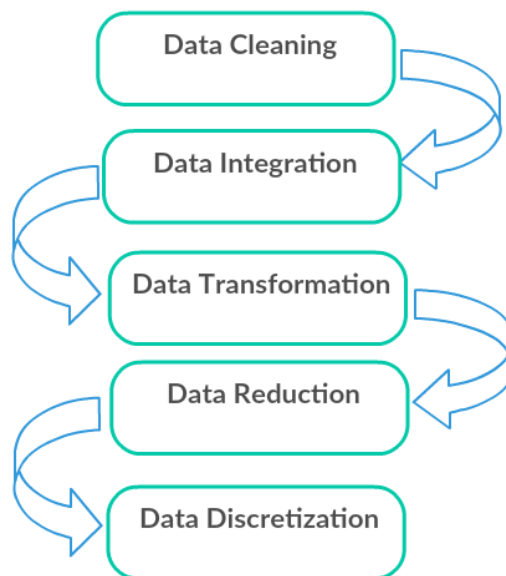


Рис. 2.3. Етапи препроцесінгу даних

На початковому етапі препроцесінгу найочевиднішим підходом є видалення шумів зі звукових сигналів за допомогою технік *noise cancellation*. Основна ідея полягає у фільтрації сторонніх шумів, таких як вітрові пориви, дощ, фонові звуки природи або техногенний шум, які можуть ускладнювати розпізнавання пташиних звуків. Методологічно це часто виконується за допомогою алгоритмів спектральної віднімальної фільтрації або фільтрів нижніх і верхніх частот. Однак на практиці виникає питання ефективності такого підходу, адже повне усунення шумів може також призвести до втрати важливих характеристик сигналу або його часткової деградації. Крім того, досвід застосування існуючих моделей, таких як BirdNet або Merlin, вказує на те, що попередні методи очищення даних від шумів не завжди забезпечують стійку роботу в реальних умовах, що призводить до погіршення результатів на польових записах.

Наступним логічним кроком препроцесінгу було б виділення фрагментів, які містять виключно звуки птахів. Для цього використовуються методи детекції акустичних подій (*Audio Event Detection, AED*), що базуються на енергетичних характеристиках сигналу або на машинному навчанні, зокрема за допомогою моделей класифікації та детекції. Виділення фрагментів дозволяє сконцентрувати дані на релевантних ділянках, зменшуючи кількість непотрібної інформації. Однак на цьому етапі виникає проблема різної тривалості аудіозаписів. Дані можуть мати фрагменти довжиною від кількох секунд до кількох хвилин, що робить їх непридатними для прямого використання у більшості моделей машинного навчання, які потребують фіксованих розмірів вхідних даних. Найочевидніше рішення – застосування техніки *padding* (наприклад, *zero padding*), щоб вирівняти довжину сигналів, але це може призвести до зростання розмірів спектрограм та ускладнення обчислювальної обробки. Спектрограми з різними розмірами ускладнюють роботу моделей через обмеженість обчислювальних ресурсів, особливо якщо навчання відбувається на локальних потужностях замість високопродуктивних серверів.

Крім того, як описувалось в попередньому розділі, залишається актуальною проблема незбалансованості датасету. У реальних умовах записи пташиних звуків мають значну диспропорцію між частотою зустрічання різних видів. Деякі види птахів трапляються у записах значно частіше, тоді як інші — дуже рідко. Ця ситуація типова для природних досліджень, особливо якщо записи були зібрані вручну орнітологами, оскільки процеси прослуховування та відбору даних доволі ресурсномісткі. Наявність такого дисбалансу значно впливає на навчання моделей, оскільки вони схильні до переобчислення на користь більш поширених класів. Для вирішення цієї проблеми на етапі препроцесінгу часто застосовуються техніки oversampling для рідкісних класів, такі як синтетичне збільшення даних через методи SMOTE (Synthetic Minority Over-sampling Technique) або техніки генерації додаткових варіантів сигналу шляхом аугментації (наприклад, зміни амплітуди, додавання штучного шуму, зміна швидкості відтворення).

Таким чином, препроцесінг включає кілька критичних кроків, які передбачають послідовне вирішення проблем шумів, виділення релевантних фрагментів, нормалізацію тривалості записів та боротьбу з дисбалансом даних. Кожен з цих підходів має свої обмеження та вимагає ретельної оцінки ефективності, оскільки будь-яка зміна на цьому етапі може вплинути на точність і продуктивність кінцевої моделі машинного навчання.

На етапах Feature Extraction та Feature Weighting/Selection відбувається перехід від оброблених сирих даних до обчислювально-зрозумілих та релевантних для моделі ознак, що критично важливе для якісного навчання.

Видобування вектору ознак набору даних зосереджене на трансформації аудіосигналів у набір характеристик, що описують важливі аспекти звуків птахів. Для цього використовуються згадувані раніше методи обробки звукових даних, наприклад ми можемо представити дані у спектрограмах, із якими добре працюватиме такий метод машинного навчання як згорткові нейромережі. Приклад спектрограми, яку отримуємо у результаті опрацювання фрагменту із аудіозапису що містить звуки перелітних птахів, представлено на рис. 2.4.

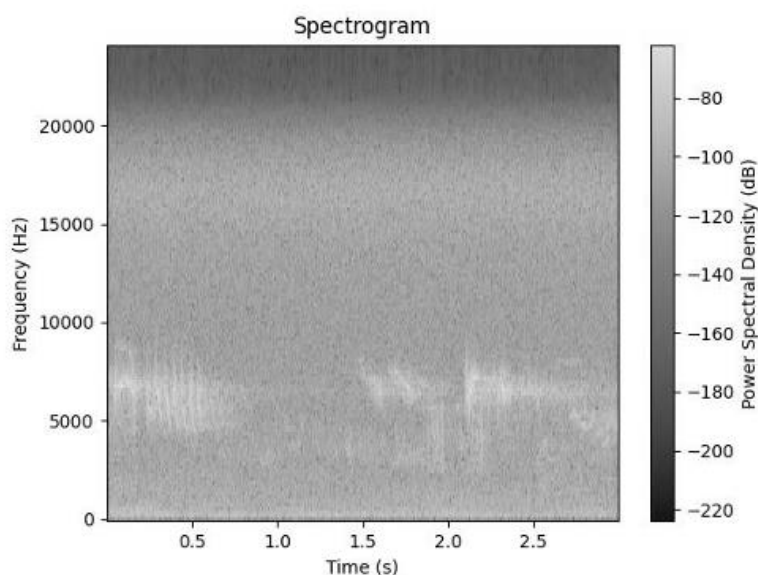


Рис. 2.4. Інтерпретація звукових хвиль у вигляді спектрограми

Окрім спектрограм, ми також можемо використовувати інші дескриптори, такі як енергетичні показники, часові характеристики, а також обчислення параметрів гармонічності сигналу[9]. Також важливий аспект аналізу ефективності методів, що застосовувалися для схожих класів задач. Наприклад, у задачах біоакустики чи розпізнавання мовлення. Аналізуючи попередні дослідження, можна визначити, що методи, орієнтовані на обробку "чистих" голосових записів, не завжди підходять для даних із даного дослідження. Це пов'язано з тим, що міграційні кличі птахів мають короткочасний та хаотичний характер у порівнянні зі звичайними співами, які, як правило, більш структурованіші. Відповідно, під час вибору ознак потрібно враховувати специфіку даних і експериментувати з нестандартними методами, такими як побудова часових характеристик, які можуть краще описувати короткі звуки.

На наступному етапі (Feature Weighting/Selection) акцент робиться на відборі найбільш значущих ознак, які мають найбільший внесок у навчання моделі. Відбір ознак потрібний, оскільки дані, отримані на етапі Feature Extraction, можуть містити надмірну кількість інформації або нерелевантні ознаки, що лише збільшують обчислювальну складність і ризик перенавчання моделі. Тут застосовуються такі підходи, як аналіз важливості ознак на основі статистичних методів або алгоритмів машинного навчання. Також

використовуються методи регуляризації (L1-регуляризація, як у LASSO), які дозволяють зменшити вагу малозначущих ознак.

Частиною цього процесу є і порівняння ефективності методів, що використовувались для схожих задач. Якщо певні підходи довели свою результативність у задачах біоакустичного моніторингу, їх можна застосовувати як базові стратегії.

Таким чином, на цих двох етапах здійснюватиметься перетворення «сирих» даних у формати, що можуть використовуватись моделями і оптимізація наборів ознак із відкиданням зайвих чи малоефективних. Аналіз попередніх рішень для схожих задач дозволяє використовувати відомі методи як відправну точку, але специфіка завдання потребує подальшого тестування та налаштування під особливості звукових сигналів міграційних кличів.

На заключних кроках роботи, якими згідно із рисунку виступають використання методів машинного навчання (CI method) та їх валідація, відбувається безпосереднє застосування моделей для розв'язання поставленого завдання, а також оцінка їхньої ефективності на основі обраних метрик, застосовуються різні підходи, що відповідають специфіці задачі. Однозначно, найбільш поширені моделі – це нейронні мережі (зокрема, згорткові нейронні мережі, CNN), які дозволяють ефективно обробляти спектрограми завдяки їхній здатності виявляти просторово-часові патерни. Альтернативно можуть застосовуватись традиційні алгоритми машинного навчання, як-от Random Forest, Support Vector Machines (SVM) або Gradient Boosting, які працюють із вручну відібраними ознаками чи pretrained models, як от BirdNet, ResNet, YAMNet, які потрібно «доналаштувати» на датасеті, проте не потрібно обирати параметрів самої моделі. Вибір методу значною мірою залежить від обсягу даних, їхньої якості та обчислювальних ресурсів.

Етап валідації критичний для оцінки узагальнюючої здатності моделі, для нього дані розділяються на тренувальний, валідаційний та тестовий набори, що дозволяє уникнути перенавчання та об'єктивно оцінити роботу моделі на невідомих даних. Основні метрики, що застосовуються – точність (accuracy), F1-score та recall, оскільки у завданнях біоакустики важливо звести до мінімуму

пропуск сигналів птахів. Крім того, в реальних умовах проводиться аналіз продуктивності моделі на даних із шумами, оскільки чистота аудіозаписів недостатня.

Зворотні стрілки, показані на діаграмі, ілюструють ітераційний характер процесу роботи. На практиці рідко вдається побудувати ефективну модель з першого разу, навіть якщо обрана стратегія видається оптимальною в теорії. Часто виникають ситуації, коли результати валідації демонструють недостатню продуктивність моделі, що вказує на необхідність повернення на попередні етапи для доопрацювання. Наприклад, низька точність чи висока кількість помилок можуть бути наслідком некоректно обраних ознак, недостатньої якості препроцесингу чи неправильно відібраних методів роботи із «зашумленими» даними. У такому випадку доводиться переглядати підходи до Feature Extraction або Feature Selection, вносячи зміни у методи обробки сигналів та балансування даних.

Ще одним поширеним сценарієм є необхідність повернення на етап збору даних, коли результати моделювання вказують на те, що існуючі дані неповні або неадекватні для навчання. Наприклад, це може траплятись через значну незбалансованість класів, недостатню кількість записів міграційних кличів чи їхню низьку якість. В таких випадках виникає необхідність у дозбируванні даних — процесі, який доволі тривалий і ресурсозатратний, але без нього неможливо досягти очікуваних результатів. Цей етап включає повторний збір сигналів, можливо, у нових локаціях чи часових рамках, що дозволить збільшити обсяг репрезентативних даних для навчання.

Таким чином, хоча основний цикл процесу сконцентрований на тренуванні та валідації моделі, як це показано на діаграмі, реальна робота над завданням вимагає постійного повернення на попередні етапи для вдосконалення методів роботи з даними. Ітеративний підхід забезпечує поступове доопрацювання кожного кроку, враховуючи специфічні виклики, які виникають на практиці, а саме: необхідність більш точного препроцесингу, оптимізації ознак та адаптації методів до реальних умов застосування.

2.2. Ефективні застосування моделей ШІ для роботи із аудіоданими

Загальна ефективність методів машинного навчання значною мірою залежить від специфіки даних і характеристик звукових сигналів. Для перетворення аудіоданих у формат, придатний для навчання моделей, найчастіше звукові сигнали конвертуються у спектрограми, мел-спектрограми чи інші часово-частотні представлення, які дозволяють виявити приховані патерни та особливості сигналу. Проте, навіть за наявності ефективного препроцесингу, не всі методи машинного навчання демонструють однакову продуктивність для задач аналізу звукових даних. Традиційні алгоритми, як-от Random Forest, SVM чи k-Nearest Neighbors (kNN), виявилися менш ефективними у випадках, коли дані мають складні часові та просторові залежності. Вони часто вимагають ручного вибору ознак, що може призвести до втрати важливої інформації, особливо у випадках коротких, фрагментованих або зашумлених сигналів. На противагу їм, глибокі нейронні мережі, зокрема згорткові нейронні мережі, і рекурентні нейронні мережі, включно з LSTM чи GRU варіаціями продемонстрували високу ефективність завдяки здатності автоматично виявляти складні патерни у спектрограмі та ефективно працювати із послідовностями даних. Залежність між змінами у кількості даних та ефективності застосування методів машинного навчання представлено на рис. 2.5.

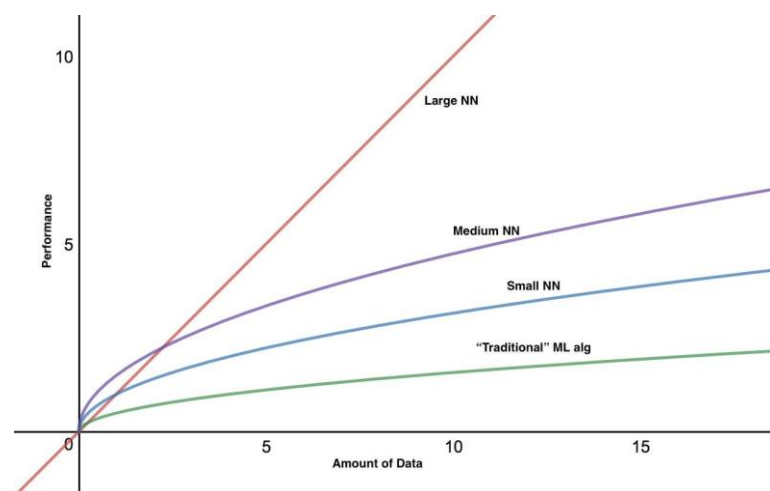


Рис. 2.5. Кореляція між кількістю даних та результатами перформансу

З рис. 2.5 видно, що із об'ємними завданнями краще справляються моделі на основі нейромереж. Наприклад, CNN добре підходять для аналізу спектральних зображень, оскільки здатні обробляти часово-частотні особливості сигналу, тоді як RNN дозволяють враховувати послідовні часові залежності у даних, що є важливим для динамічних звукових сигналів. Комбінація цих двох підходів, як у CRNN (Convolutional Recurrent Neural Networks) – показала себе особливо успішною для задач аудіокласифікації, такий підхід здебільшого і використовується у раніше згаданих pretrained моделях. Загальні кроки у підборі моделей для роботи із такими задачами включають:

- попередню обробку даних для усунення шумів, нормалізації сигналів та їх трансформації у відповідний формат;
- вибір методів представлення даних, таких як спектрограми чи MFCCs, що забезпечують оптимальну вхідну інформацію для моделей.
- підбір моделей машинного навчання на основі характеру даних. Для невеликих обсягів ефективні традиційні алгоритми, а для великих та складних наборів краще підходять глибокі нейронні мережі.
- балансування датасету, оскільки значна частина аудіозавдань «страждає» від дисбалансу класів, що впливає на якість навчання.
- тестування продуктивності моделі на реальних умовах із зашумленими сигналами, оскільки ідеально чисті дані рідко трапляються на практиці.

Серед ключових проблем, які впливають на ефективність методів машинного навчання – чутливість до шумів у природньому середовищі, а також висока варіативність сигналів, що ускладнює їх класифікацію. Методи, які показують високу точність на синтетичних або лабораторних даних, часто виявляються неефективними на реальних записах, що вимагає додаткових етапів адаптації моделей чи доопрацювання даних.

Таким чином, застосування машинного навчання для аналізу звукових сигналів потребує системного підходу, що поєднує ефективний препроцесинг, обґрунтований вибір методів обробки та тестування моделей в умовах, наближених до реальних.

Для вибору правильної стратегії роботи та моделі, варто спробувати застосувати моделі які вже довели свою ефективність. Наприклад, у науковому дослідженні 2021 року, яке стосується розробки методу детекції апное розглядалось використання глибоких нейронних мереж, що може бути корисним у сфері діагностики обструктивного апное сну (OSA)[10]. Це дослідження має значний потенціал як для медичних установ, так і для розробки домашніх систем моніторингу сну.

Важливо відзначити, що група працювала з даними, зібраними в клінічних умовах, де застосовувалося до п'яти мікрофонів для реєстрації аудіосигналів. Це дозволило оцінити вплив позиції мікрофона на якість детекції. У випадку із визначенням у контексті «природнього» середовища збору даних наприклад, використання аудіосигналів у домашніх умовах або з меншою кількістю мікрофонів може демонструвати значно інші результати. Проте, очікується, що навіть при змінних умовах якісний збір даних та правильний вибір методів обробки забезпечуватимуть високоточний результат.

Завдяки застосуванню глибоких нейронних мереж, дослідники змогли відмовитися від ручного вилучення ознак з аудіосигналів. Вони використали комбіновану архітектуру нейромережі, що складається зі згорткових нейронних мереж та рекурентних шарів, повна архітектура представлена на рисунку 2.6.

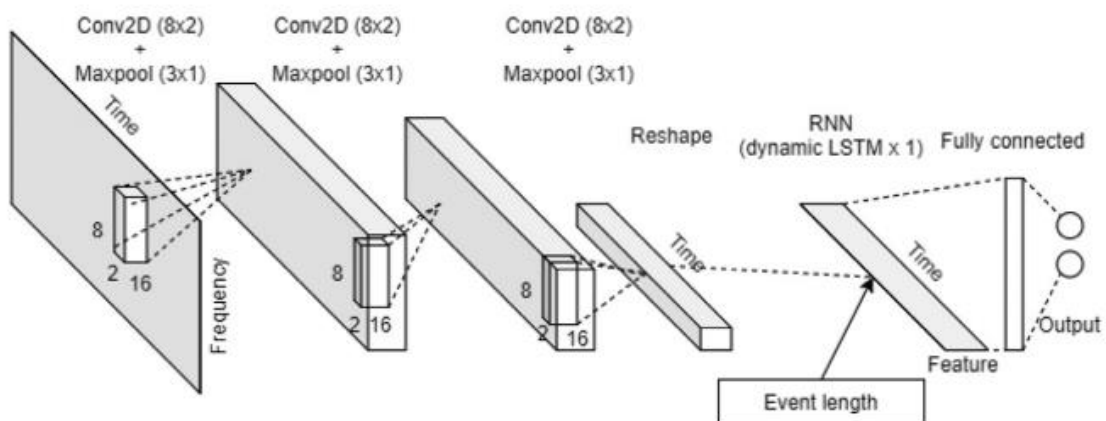


Рис. 2.6. Пропонована у дослідженні архітектура моделі[10]

Використання CNN дозволило автоматично виділяти просторові ознаки зі спектрограм, у той час як RNN допомагала обробляти часову послідовність

аудіосигналів. Для інтерпретації звукових сигналів у цифровому вигляді було використано CQT-спектрограми як вхідні дані для навчання моделі, оскільки вони забезпечують кращу роздільну здатність у низькочастотних діапазонах, де переважно знаходились очікувані звуки. Цей підхід може бути застосований і у інших сферах, як от аналізу короткотривалих звуків, що створюють птахи у час міграції, адже аудіосигнали часто мають низькочастотні характеристики, які важливо враховувати.

Також, дослідження показало, що CQT-спектрограми у поєднанні з CNN-RNN дозволяють досягти середньої точності 95.3%, що є значно вищим за результати традиційних методів розпізнавання. Для порівняння, дослідники також протестували Mel-спектрограми, які зазвичай використовуються в задачах розпізнавання мовлення. Хоча Mel-спектрограми показали трохи нижчу ефективність, їх включення в аналіз допомагає зрозуміти, як вибір типу спектрограм впливає на результат навчання, що важливо для інших досліджень, оскільки підкреслює необхідність тестування різних видів спектрограм для досягнення оптимального результату в конкретних умовах.

Ще одна особливість розглянутого підходу в обробці полягала у рішенні застосовувати фільтрацію низьких частот для усунення шуму та нормалізували тривалість подій до стандартного значення (3.5 секунди) за допомогою zero-padding. Це гарантувало стабільність моделі при обробці аудіофрагментів різної довжини. Крім того, автоматична детекція звукових подій супроводжувалася «ручною» валідацією, що забезпечило якість даних для тренування моделі.

Загалом підхід дослідників демонструє ефективне поєднання сучасних методів глибокого навчання та продуманої обробки даних для детекції апное. Це дослідження є корисним для описуваного у цій роботі дослідження, оскільки надає цінну інформацію про застосування нейронних мереж у обробці аудіоданих, а також показує, як вибір правильних архітектур і форматів представлення даних може значно покращити результативність моделі. Аналіз оглянутих методів дозволяє визначити ключові кроки для розробки ефективної системи: від вибору типу спектрограм до обробки даних та оптимізації

архітектури нейронної мережі, таблиця порівнянь різних підходів і їх ефективності представлені на рис. 2.7.

Method	Feature/Representation	Microphone placement	Dataset	Results
Hidden Markov models (HMMs)	MFCC-based feature vectors	Bedside table	6 subjects, 1.5 h from each subject, 1 h for training, 0.5 h for testing	Sensitivity, 82%-89%.
Robust linear regression	Spectral features	15cm over the patient's head	Full night recordings from 18 simple snorers and 12 OSA patients	Sensitivity and PPV, Simple snorer: 97.3%, 99.6%; OSA patients: Mixed, 90.2%, 98.7%; OSA patients:86.8%, 93.8%.
Unsupervised fuzzy C-means clustering	Feature extraction using principal component analysis (PCA)	Hang in the air	A short period of the entire night recording from 23 OSA patients and 7 simple snorers	Sensitivity and PPV, Simple snorers and OSA: 93.1%, 95.9% OSA: 91.6%, 97.2% Simple snorers: 96.4%, 93.5%
AdaBoost classifier	Time-related features and spectral-related features (127 features)	1 meter above the head	Full night recordings from 67 subjects (42 for validation)	Accuracy: 98.4%; Sensitivity: 98.1%; Specificity: 98.2%.
Artificial Neural Network	A set of features for speech analysis	Between 40cm to 70cm from the face of the patients	Full night recordings from 34 subjects, 21 subjects for training, 13 subjects for testing	Accuracy: 86-89%; Sensitivity: 82-87%; Specificity: 87-89% PPV: 85-88%; NPV: 85-91%.
SoundNet: a deep 1D convolutional architecture	Raw audio signal	A table 50cm away from the volunteer's head	Part of full night recordings from 10 subjects	Average PPV: 81.82%
Recurrent neural network	MFCCs	Approximately 70 cm from the top end of the bed (M1)	Part of full night recordings from 20 subjects (11 for training, 3 for validation and 6 for testing)	Accuracy: 95%; Sensitivity: 92%; Specificity: 98%.
Quadratic classifier	Features related to formant positioning and magnitude ratios	Upper side of the shoulder, within arm's reach (<50cm)	Full night recordings after pre-selection from 18 subjects without preexisting respiratory diseases or sleep disorders.	Accuracy: 95.07%; Sensitivity: 98.58%; Specificity: 94.62%.
CNN + RNN	CQT spectrogram	Approximately 70 cm from the top end of the bed (M1)	Part of full night recordings from 38 subjects	Accuracy: $95.3 \pm 0.5\%$; Sensitivity: $92.2 \pm 0.9\%$; Specificity of $97.7 \pm 0.4\%$

Рис. 2.7. Порівняння різних моделей для класу завдань[10]

На основі таблиці, що містить порівняння попередніх досліджень, можна провести детальний аналіз відмінностей у методах, представленні даних, позиціонуванні мікрофонів та використанні різних моделей машинного навчання.

Кожен підхід використовує різні особливості, репрезентації даних та місця розташування мікрофонів, що впливає на результати продуктивності. Hidden Markov Models (HMMs) ґрунтуються на MFCC-векторах ознак та використовують мікрофон, розташований біля ліжка, досягаючи чутливості в діапазоні 82%-89%. Натомість метод robust linear regression, що спирається на спектральні ознаки, працює з мікрофоном, розташованим 15 см над головою пацієнта, забезпечуючи високу чутливість і позитивне передбачувальне значення (PPV) для простих пацієнтів(97.3%, 99.6%), але трохи нижчі показники для тих

у кого діагностоване обструктивне апное. Для unsupervised fuzzy C-means clustering використовуються ознаки, екстраговані методом головних компонент (PCA), також мікрофон підвішаний над піддослідними. Цей підхід показує чутливість та PPV на рівні 93.1%-97.2% залежно від групи.

Класифікатор AdaBoost фокусується на часових та спектральних ознаках, зібраних на відстані 1 метра над головою пацієнта, та демонструє високі показники точності (98.4%), чутливості (98.1%) і специфічності (98.2%). Artificial Neural Network обробляє набір ознак для мовного аналізу з мікрофоном, розташованим на відстані 40-70 см від обличчя пацієнтів. Продуктивність цієї моделі коливається в межах 86%-89% за точністю, при чутливості до 87% та специфічності 87%-89%. Модель SoundNet на основі глибокої 1D згорткової архітектури працює з неопрацьованими аудіосигналами, отриманими з мікрофона, встановленого на столі за 50 см від голови об'єктів дослідження. Цей метод демонструє середній PPV на рівні 81.82%.

Recurrent Neural Network (RNN) використовує MFCC-ознаки з мікрофоном, розташованим приблизно за 70 см від об'єкту спостереження. Вона показує значні результати: точність 95%, чутливість 92% та специфічність 98%. У випадку цього дослідження, мікрофон було розташовано з боку від пацієнтів на відстані до 50 см. Ця модель забезпечує точність у 95.07%, чутливість 98.58% та специфічність 94.62%. Нарешті, комбінація CNN і RNN із використанням CQT-спектрограми працює з тими ж умовами розміщення мікрофона, як і RNN. Цей метод показує найвищу точність у $95.3 \pm 0.5\%$, а також чутливість $92.2 \pm 0.9\%$ та специфічність $97.7 \pm 0.6\%$. Загалом різні моделі демонструють високу продуктивність, особливо при використанні комбінованих глибоких архітектур та специфічних особливостей.

2.3. Виклики у аналізі та опрацюванні наборів даних

Одна з головних проблем, яка спостерігається в більшості з досліджень – недостатність або обмеженість датасетів. У багатьох випадках аналіз виконується на досить невеликих вибірках, що може призводити до обмеженої узагальненості

результатів. Наприклад, Hidden Markov Models (HMMs) застосовувалися лише на 6 суб'єктах, що є явно недостатнім для створення моделі, яка може ефективно працювати на різноманітних даних з реального світу. Подібна ситуація простежується і з методами, такими як SoundNet і RNN, де кількість записів також обмежена (10-20 суб'єктів). Незважаючи на високу точність, отримані результати можуть бути завищеними через недостатню варіативність у даних. Малі вибірки часто не містять достатньої кількості аномалій чи особливих випадків, що зазвичай присутні у реальних умовах [6].

Інша важлива проблема – "стерильність" умов, у яких проводився запис аудіоданих як для навчання моделей, так і для тестових вибірок. Майже всі дослідження виконувались у контрольованих лабораторних умовах, де сторонні шуми або артефакти були мінімізовані або повністю відсутні. Наприклад, записи велися при фіксованому розташуванні мікрофонів (зазвичай на відстані від 15 см до 1 метра від пацієнта), що не відображає реальних умов у житлових приміщеннях чи лікарнях, де можуть бути численні перешкоди у вигляді фонового шуму, руху людей, розмов або інших джерел звуку. У таких "ідеальних" умовах моделі мають змогу навчатися на чистих, високоякісних даних, але їх продуктивність може суттєво знизитись при застосуванні в реальному житті, коли дані будуть значно менш структурованими.

Крім того, відсутність різноманітності в аудіозаписах, як-от врахування лише певної групи пацієнтів могла також вплинути на результати. Це призводить до того, що модель може бути надмірно "пристосованою" до навчального набору, але не буде ефективно працювати на інших категоріях пацієнтів або в умовах з іншими характеристиками звуку. Наприклад, для штучних нейронних мереж, які демонстрували хорошу продуктивність на невеликих наборах, відсутність розмаїття в даних, включно з пацієнтами різного віку, статі або з різними особливостями сну, могла стати ще однією причиною потенційної перетренованості.

Усі ці аспекти підкреслюють важливість розширення датасетів при виконанні будь яких дослідів пов'язаних із машинним навчанням, додавання різноманітних умов, а також тестування моделей у менш контрольованих

середовищах. Тільки з таким підходом можна отримати релевантну оцінку продуктивності методів і забезпечити їхню надійність для практичного застосування.

Ще однією проблемою є уніфікація даних, коли записи мають різну тривалість і якість. Навіть якщо вдасться нормалізувати дані, обробка коротких і довгих записів одночасно може спричинити дисбаланс у навчанні моделі. Довгі записи, що містять значну кількість фонових шумів, створюють додаткове "засмічення" даних, оскільки не вся інформація в них є релевантною. З іншого боку, короткі записи містять менше інформації для аналізу, що може призвести до зниження точності моделі при розпізнаванні складних сигналів. Внаслідок цього виникає необхідність у додатковому фільтруванні та сегментації даних, що надто трудомістко і потребує додаткових ресурсів для автоматизації процесу.

Окрім того, великий розкид у тривалості записів ускладнює побудову навчальних батчів під час тренування моделей. У машинному навчанні важливо, щоб батчі мали однакові розміри й були структурованими для рівномірного завантаження даних у GPU або CPU. Якщо дані мають різну тривалість, кожен батч доводиться "підганяти" під найдовший запис у ньому, що призводить до надмірного споживання пам'яті й значно уповільнює процес навчання.

Альтернативний підхід — використання zero padding, як показано на рис. 2.8.

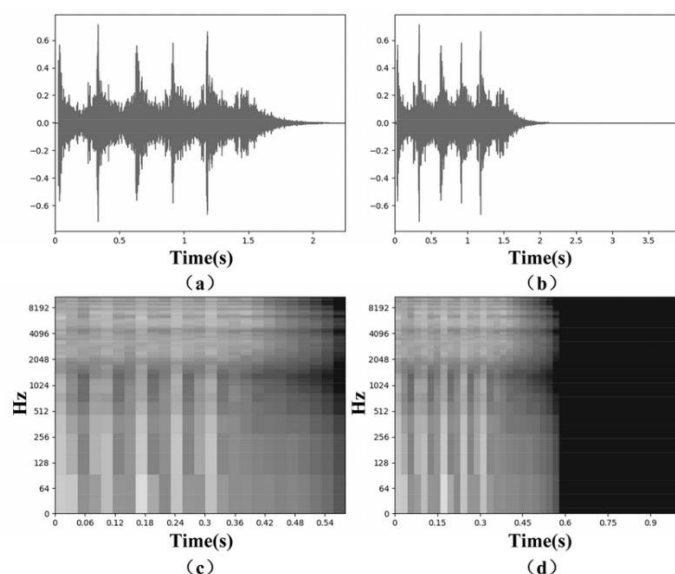


Рис. 2.8. Застосування техніки «доповнення» на спектрограмах

Тобто підхід вбачає додавання нульових значень до коротших записів, щоб довести їх до єдиної тривалості. Проте цей метод також має значні недоліки. По-перше, додавання нулів змінює структуру спектрограм аудіофайлів, які зазвичай використовуються як вхідні дані для моделей машинного навчання. Внаслідок цього виникають дуже великі розміри тензорів, що ускладнює їх обробку.

2.4 Висновки до розділу

Для моделей глибокого навчання, таких як CNN або RNN, великі спектрограми значно збільшують вимоги до обчислювальних ресурсів і пам'яті. Для роботи з такими підходами на наявних записах потрібна розробка спеціальних підходів до обробки аудіоданих, які б враховували їх нерівномірну тривалість. Одним із можливих рішень може бути автоматична сегментація довгих записів на менші фрагменти з дотриманням певних критеріїв (наприклад, виявленням звукових активностей через порогові значення). Однак цей процес також потребує додаткових алгоритмів для виявлення релевантних сигналів, щоб уникнути випадкового втручання фонових шумів у критичні моменти. І лише після цього можна переходити до тренування моделей машинного навчання з використанням більш структурованих і чистих даних.

РОЗДІЛ 3

АНАЛІЗ І УЗАГАЛЬНЕННЯ РЕЗУЛЬТАТІВ ПРАКТИЧНИХ ЕКСПЕРЕМЕНТІВ

3.1. Застосовані методи попереднього опрацювання даних та аналіз їх ефективності

Для підготовки аудіоданих до аналізу та побудови моделі проведено ряд кроків, починаючи від обробки вихідних файлів формату WAV та екстракції метаданих до застосування методів балансування та нормалізації даних. Вихідні файли містять назви, з яких за допомогою відповідної таблиці кодів видів птахів було створено датасет у форматі CSV. Структура отриманого файлу включала наступні поля: назва файлу (filename), юліанська дата (juliandate), місце запису (loc), нижня та верхня частота сигналу (low_freq і high_freq), час початку та закінчення запису (start, end), а також вид птаха (species). Ця інформація основою для подальшої обробки.

На основі отриманих метаданих проводиться аналіз тривалості покликів та розподілу кількості записів за видами птахів. Зокрема, було визначено, що значна кількість записів мала тривалість менше однієї хвилини (1825 записів), тоді як виклики тривалістю понад дві хвилини були рідкісними (47 записів). Також було помічено значну диспропорцію в кількості записів різних видів птахів. Наприклад, спів дрозда співочого був зареєстрований 2003 рази, тоді як деякі види, як-от сірий плиска чи в'юрок, мали лише по одному запису. Ця нерівномірність у даних потребує застосування технік балансування.

Для збільшення ефективності обробки даних і зменшення дисбалансу у вибірках було прийнято рішення використовувати комбіновані підходи: over- та undersampling. Спочатку, для довгих записів тривалістю понад одну хвилину було застосовано поділ на менші сегменти. Записи, які тривали від однієї до двох хвилин, розділялися на два фрагменти, а ті, що перевищували дві хвилини, — на три частини. Це дозволило збільшити кількість записів, зберігаючи основну інформацію, і уникнути необхідності переглядати всі довгі записи вручну. Також

цей підхід базувався на перевіреному факті, що у довгих записах спів птахів був зафіксований кілька разів.

Для нормалізації вибірок застосовувалася технологія undersampling із використанням алгоритму кластеризації k-means. Цей метод дозволив збалансувати кількість зразків у класах шляхом відбору представників з найбільш надмірно представлених класів, зберігаючи різноманітність даних та уникаючи втрати ключової інформації. Функція із застосуванням методу приведена на рис. 3.1.

```
def undersample_class(df, class_name, audio_dir, target_size=500):
    """
    Undersample a specific class using clustering to retain diverse records.
    """
    class_records = df[df['species_code'] == class_name]
    features = []
    indices = []

    for _, row in class_records.iterrows():
        audio_path = audio_dir / row['filename']
        try:
            feature = simple_audio_features(audio_path)
            features.append(feature)
            indices.append(row.name)
        except Exception as e:
            print(f"Error processing {audio_path}: {e}")

    if len(features) > target_size:
        kmeans = KMeans(n_clusters=target_size, random_state=42)
        kmeans.fit(features)
        labels = kmeans.labels_

        # Select one record per cluster
        selected_indices = []
        for cluster in range(target_size):
            cluster_indices = [idx for idx, lbl in zip(indices, labels) if lbl == cluster]
            if cluster_indices:
                selected_indices.append(cluster_indices[0])
    else:
        # Keep all records if less than or equal to target_size
        selected_indices = indices

    return selected_indices
```

Рис. 3.1. Застосування технік опрацювання даних із k-means методом

З метою розширення малих класів у датасеті було використано кілька підходів до oversampling. Для аудіозаписів було застосовано ковзне вікно, яке створювало нові зразки шляхом накладення сегментів із частковим перекриттям. Це дозволило збільшити кількість записів у рідкісних класах, зберігаючи часові та частотні характеристики сигналів. Окрім цього, для спектрограм, що були

отримані із сирих аудіозаписів, застосовувалась технологія SMOTE (Synthetic Minority Oversampling Technique). Цей метод синтетично генерував нові зразки для малих класів, створюючи комбінації існуючих даних і забезпечуючи кращий розподіл у багатовимірному просторі ознак. Результати балансування даних приведено на рис. 3.2.

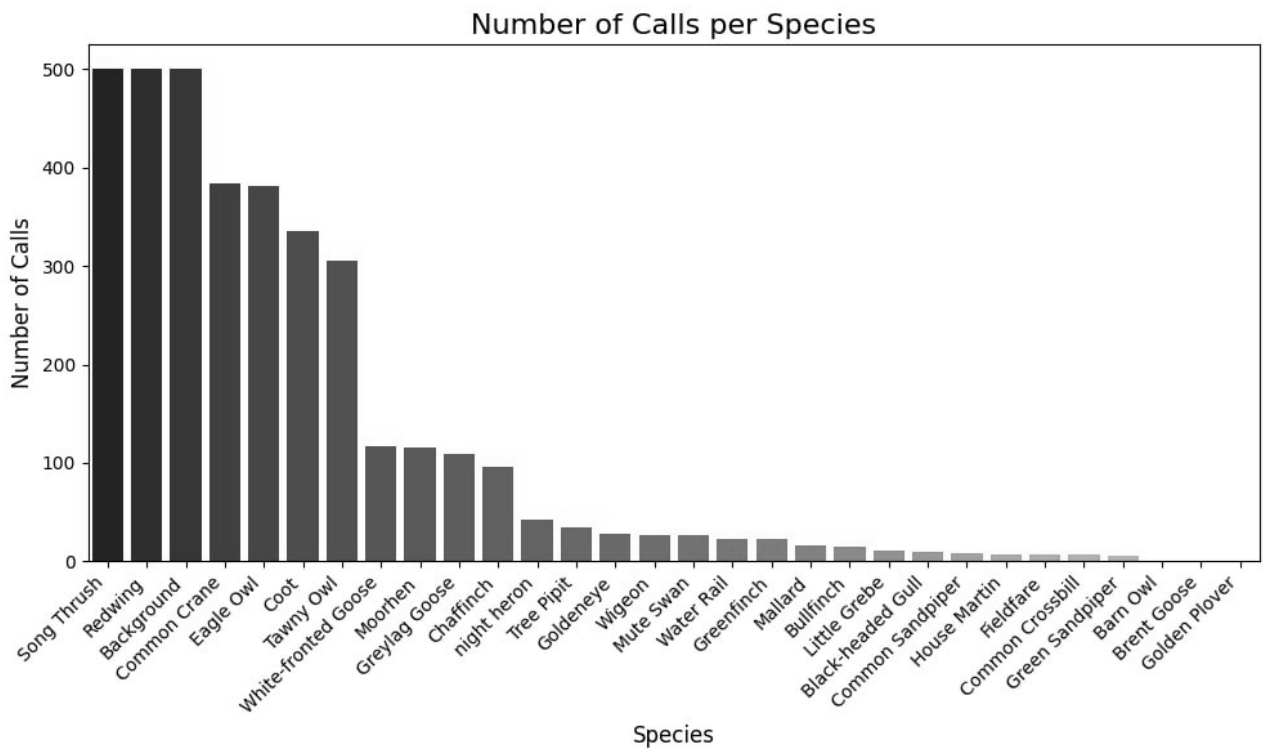
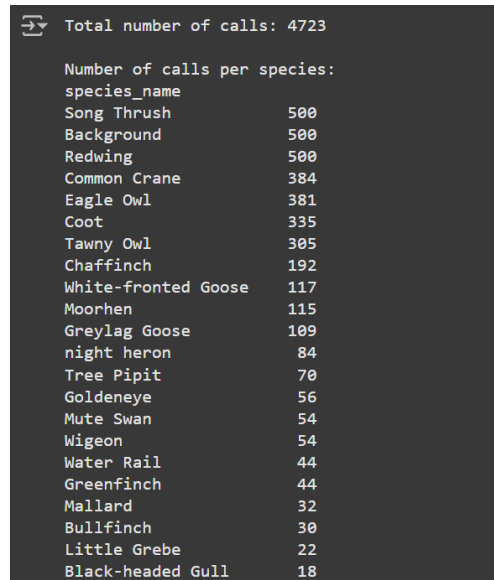


Рис. 3.2. Розподіл даних після застосування технік оверсамплінгу

Проведені заходи підготовки та балансування датасету дали змогу зменшити дисбаланс у даних та підготувати більш репрезентативний набір для тренування моделей. Це, своєю чергою, створило умови для коректної оцінки продуктивності моделей на різних класах і забезпечило більшу стійкість під час застосування в реальних умовах. Проте із графіку чітко видно що для найменш поширених класів кількість записів суттєво не змінилась, це відбувається через особливості методу синтетичної генерації семплів, оскільки для алгоритму SMOTE чи ADASYN потрібно аби в класу було хоч 5 прикладів, на основі яких створюються нові, синтетичні приклади, а оскільки в наявному наборі лише 1 запис – такий підхід не може надати практичного вирішення проблеми навіть у

разі створення дублікатів записів, оскільки в плані частот та звукових хвиль вони не різнитимуться. Отриманий в результаті даного етапу масив класів подано на рис. 3.3.



```

Total number of calls: 4723

Number of calls per species:
species_name
Song Thrush           500
Background            500
Redwing               500
Common Crane          384
Eagle Owl             381
Coot                  335
Tawny Owl             305
Chaffinch             192
White-fronted Goose   117
Moorhen               115
Greylag Goose         109
night heron           84
Tree Pipit            70
Goldeneye             56
Mute Swan             54
Wigeon                54
Water Rail            44
Greenfinch            44
Mallard               32
Bullfinch             30
Little Grebe          22
Black-headed Gull      18
  
```

Рис. 3.3. Цифрове представлення розподілу набору даних

У процесі підготовки даних для тренування моделі також додано спеціальний клас, який складається з записів шумів із оточення, в якому здійснювалися оригінальні записи пташиного співу. Це дозволило впевнитись, що модель навчається розпізнавати саме пташині звуки, а не фонові шуми, такі як автомобільний рух, дощ, звуки комах або інші природні та антропогенні фактори.

Для створення цього класу були використані дані з оригінальних аудіозаписів. Фрагменти, що не містили зазначених у метаданих періодів із пташиним співом, обрізались із початку та кінця кожного файлу. Ці обрізані частини містили типові фонові шуми для середовища, де проводився запис, і, ймовірно, не включали значущих біоакустичних сигналів.

Після обрізання ці фрагменти конкатинувались, з метою формування вибірки. З отриманого набору було відібрано 500 найбільш відповідних за довжиною файлів, які відповідали середній тривалості записів основного датасету. Цей обмежений розмір вибірки дозволив досягти балансу між

ефективністю використання шумових даних і уникненням надлишкової їхньої кількості, що могло б створити диспропорцію в датасеті.

Додавання цього класу шумів важливе для підготовки даних, оскільки забезпечує краще розуміння моделлю специфіки пташиного співу та допомагатиме у зменшенні ризику "навчання" моделі на некоректних ознаках, пов'язаних виключно з особливостями оточення. Цей підхід також сприяє підвищенню точності моделі та її здатності до узагальнення під час роботи із записами, отриманими в різних умовах.

У рамках підготовки та тестування даних для класифікації пташиного співу були створені мел-спектрограми та constant-Q спектрограми для заданих фрагментів аудіозаписів. Обидва типи спектрограм були протестовані на різних моделях, як pretrained так і власних. Мета цього етапу полягала у визначенні, який підхід до представлення даних забезпечує кращу продуктивність моделей на основі отриманих записів.

Результати показали, що вже на перших епохах навчання між моделями, які працювали з мел-спектрограмами та constant-Q спектрограмами, спостерігалася різниця. Зокрема, моделі, що використовували мел-спектрограми, демонстрували початкову точність, яка фактично відповідала випадковому вибору класів. У той же час constant-Q спектрограми забезпечували стартову точність до 10% вищу. Це свідчить про те, що constant-Q спектрограми краще зберігають релевантну інформацію про частотні характеристики пташиного співу, що полегшує навчання моделей.

Хоча ця різниця на ранніх етапах навчання не є критичною, вона підтверджує висновки, отримані в попередніх дослідженнях, які згадувалися в цій роботі. Такі результати підкреслюють важливість вибору способу перетворення сирих аудіоданих у спектральне представлення для оптимізації навчання моделей та підвищення їхньої ефективності у задачах біоакустичного аналізу.

Для кращого аналізу ефективності певного виду спектрограм для класів із набору даних, варто розглянути розподіл частот які зустрічаються у записах, демонстрацію показано на рис. 3.4.

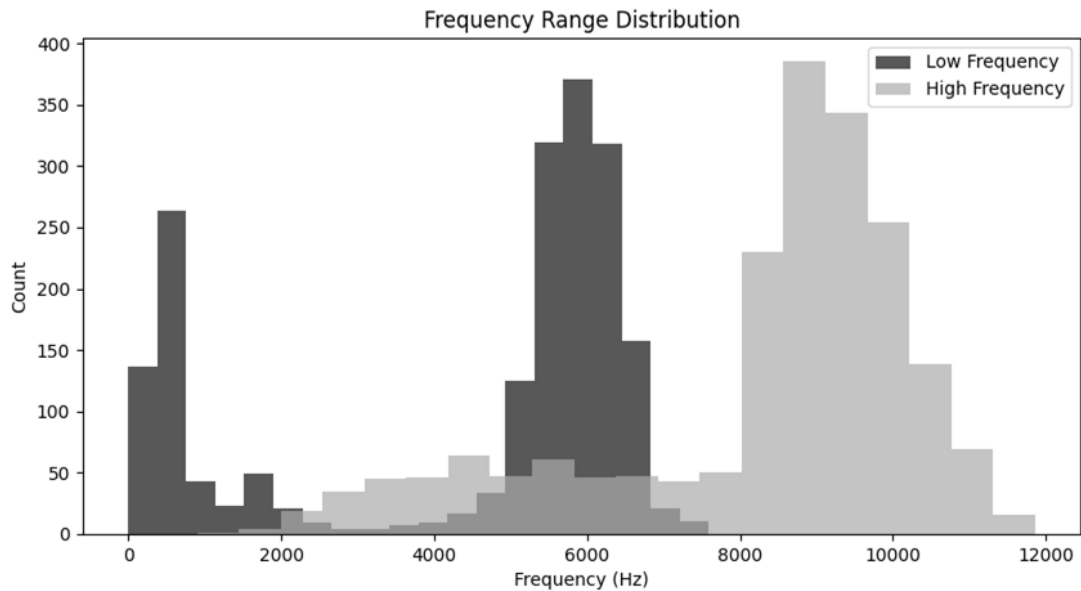


Рис. 3.4. Розподіл частот із даних аудіофайлів

Графік розподілу частот у датасеті демонструє дві чіткі групи: низькочастотний та високочастотний діапазони. Низькочастотна частина, що концентрується в області до 6000 Гц, переважно відповідає звукам із більшою енергією, які характерні для типових вокалізацій птахів. У той же час високочастотний діапазон, що розташований у межах від 6000 до 12000 Гц, містить не лише шуми або зовнішні звуки, як-от звуки комах, але й специфічні кличі птахів, які вони видають під час міграції. Такі кличі відрізняються від звичайних вокалізацій птахів, що спостерігаються в їхньому природному середовищі проживання. Під час міграції птахи видають короткі, часто високочастотні сигнали, які дозволяють їм орієнтуватися в просторі та підтримувати зв'язок із зграєю. Це створює додаткову складність у роботі з аудіоданими, адже високочастотні звуки можуть бути переплутані з фоновими шумами, такими як звук комах чи інші природні сигнали, що перебувають у тому ж діапазоні частот. Така подібність між корисними сигналами і шумами ускладнює застосування алгоритмів приглушення шумів, оскільки подібні підходи часто знижують енергію не лише небажаних сигналів, але й самих пташиних кличів. Цей факт пояснює, чому вибір спектрального представлення даних є надзвичайно важливим у завданнях класифікації пташиного співу. У цьому контексті, мел-спектрограми, адаптовані для людського сприйняття звуку,

віддають перевагу низькочастотному діапазону, що може спричиняти втрату критичної інформації у високих частотах. Натомість constant-Q спектрограми, завдяки рівномірному розподілу частотного розрізу, краще зберігають високочастотні компоненти, що є важливими для точного аналізу пташиних сигналів. Таким чином, розподіл частот на графіку відображає складність і водночас важливість точного аналізу високочастотного спектру, який містить як ключову інформацію про пташині кличі, так і значну частку фонових шумів.

Продовжуючи попередню думку, варто відзначити, що застосування моделей на зразок BirdNet і Raven, хоча і є ефективним для загальних завдань класифікації пташиного співу, стикається із суттєвими обмеженнями у випадку міграційних досліджень. Ці обмеження пов'язані насамперед із тим, як моделі взаємодіють з частотними характеристиками звуків, представленими в спектрограмах, і як вони обробляють рідкісні класи. BirdNet і Raven, будучи високорозвиненими інструментами для ідентифікації видів птахів, ефективно працюють із сигналами, які відповідають типовим вокалізаціям птахів, часто зосередженим у низько- та середньочастотних діапазонах. Ці моделі побудовані на основі великих навчальних датасетів, де пріоритет надається частотним і тембральним особливостям звуків, типовим для видів у їхньому природному середовищі. Однак, для міграційних умов, де вокалізації часто зміщені у високочастотний діапазон або мають інший характер (короткі, сигналоподібні звуки), подібні моделі втрачають частину своєї точності.

Що стосується феномену високої точності для певних видів, таких як barn owl, навіть за умов обмеженої репрезентації у датасеті, це пов'язано з двома факторами. По-перше, вокалізації цього виду мають унікальні акустичні особливості, які легко виділяються серед інших сигналів навіть за невеликої кількості зразків. Наприклад, низькочастотні гармонійні або шиплячі звуки, характерні для класу barn owl, мають високу розрізняваність, що дозволяє практично всім розглянутим моделям ефективно класифікувати їх незалежно від репрезентації.

3.2. Застосовані моделі та аналіз їх результатів

Для навчання моделей класифікації пташиного співу було використано кілька архітектур, що дозволило порівняти їхню ефективність у роботі з даними різного формату та ступеня підготовки. Зокрема, використовувались як наперед натреновані моделі, так і власноруч налаштовані архітектури.

Наперед натреновані моделі, такі як ResNet і YAMNet, продемонстрували значну перевагу у швидкості конвергенції завдяки наявності попередніх знань із великих зовнішніх датасетів. ResNet використовує передачу знань із задач класифікації зображень, зокрема спектрограм, і забезпечує стабільну продуктивність навіть на невеликих наборах даних. YAMNet, навпаки, розроблена спеціально для задач обробки аудіо і працювала на базі аудіоспектрограм, що спрощувало її інтеграцію в задачі біоакустичного аналізу. Приклад використання моделей подано на рис. 3.5.

```

yamnet_model = hub.load('https://tfhub.dev/google/yamnet/1')

def extract_features_and_labels(df):
    """Extract YAMNet embeddings and labels from audio files."""
    features = []
    labels = []
    for _, row in df.iterrows():
        try:
            audio_binary = tf.io.read_file(row['filepath'])
            waveform, _ = tf.audio.decode_wav(audio_binary, desired_channels=1, desired_samples=SAMPLE_RATE)
            waveform = tf.squeeze(waveform, axis=-1)
            scores, embeddings, _ = yamnet_model(waveform)
            features.append(tf.reduce_mean(embeddings, axis=0).numpy())
            labels.append(row['encoded_label'])
        except Exception as e:
            print(f"Error processing {row['filepath']}: {e}")
    return np.array(features), np.array(labels)

print("Extracting features...")
X_train, y_train = extract_features_and_labels(train_df)
X_valid, y_valid = extract_features_and_labels(valid_df)
X_test, y_test = extract_features_and_labels(test_df)

print(f"Feature shapes: Train={X_train.shape}, Valid={X_valid.shape}, Test={X_test.shape}")

def build_model(input_dim, num_classes):
    model = tf.keras.Sequential([
        tf.keras.layers.Input(shape=(input_dim,)),
        tf.keras.layers.Dense(128, activation='relu'),
        tf.keras.layers.Dropout(0.3),
        tf.keras.layers.Dense(64, activation='relu'),
        tf.keras.layers.Dropout(0.3),
        tf.keras.layers.Dense(num_classes, activation='softmax')
    ])

```

Рис. 3.5. Застосування моделі YAMNet

Однак обмеження таких моделей полягають в недостатній гнучкості налаштування параметрів, таких як вибір типу спектрограми або особливості попередньої обробки аудіосигналу. Це робить їх менш адаптованими до специфіки наявного датасету.

Власно-налаштовані моделі, зокрема архітектури, побудовані на основі згорткових нейронних мереж і рекурентних нейронних мереж, дозволяють експериментувати з різними вхідними форматами даних. Наприклад, для мел-спектрограм і constant-Q спектрограм можна оптимізувати архітектуру мережі під конкретні характеристики цих представлень. Це включає вибір параметрів шарів, типу активації, нормалізації та інших компонентів, що значно підвищувало ефективність навчання. Крім того, власні моделі дозволяють краще враховувати специфіку задачі, наприклад, створення спеціалізованих шарів для роботи з часово-частотними характеристиками спектрограм. Використання однієї з таких моделей представлено на рис. 3.6.

```
# Step 8: Model definition
def build_model(input_shape, num_classes):
    inputs = tf.keras.layers.Input(shape=input_shape)

    x = tf.keras.layers.Conv2D(32, (3, 3), activation='relu')(inputs)
    x = tf.keras.layers.MaxPooling2D((2, 2))(x)
    x = tf.keras.layers.Conv2D(64, (3, 3), activation='relu')(x)
    x = tf.keras.layers.MaxPooling2D((2, 2))(x)
    x = tf.keras.layers.Conv2D(128, (3, 3), activation='relu')(x)
    x = tf.keras.layers.MaxPooling2D((2, 2))(x)

    attention = tf.keras.layers.GlobalAveragePooling2D()(x)
    attention = tf.keras.layers.Dense(128, activation='relu')(attention)
    attention = tf.keras.layers.Dense(num_classes, activation='softmax')(attention)

    model = tf.keras.models.Model(inputs, attention)
    model.compile(optimizer='adam', loss='sparse_categorical_crossentropy', metrics=['accuracy'])
    return model

# Step 9: Build and train the model
input_shape = (*TARGET_SHAPE, 1) # CQT spectrogram dimensions with channel
model = build_model(input_shape, N_CLASSES)
model.summary()

history = model.fit(
    train_generator,
    validation_data=valid_generator,
    epochs=EPOCHS
```

Рис. 3.6. Створення власної моделі на базі CNN

Процес налаштування власних моделей у цьому дослідженні дозволяє забезпечити високу гнучкість у роботі з даними та адаптацію під специфічні завдання класифікації пташиного співу. Спершу проводиться підготовка даних: із CSV-файлу зчитуються метадані, що містять інформацію про шлях до аудіофайлів, мітки видів птахів та інші параметри.

Наступним кроком виступає розділення даних на навчальну, валідаційну та тестову вибірки у співвідношенні 70:15:15. Це забезпечує рівномірний розподіл класів у кожній із вибірок, що зменшує також ризик перенавчання моделі на окремих класах. Для обробки аудіофайлів використовувалась функція `sqt` з бібліотеки `librosa`, яка створювала спектрограми у форматі Constant-Q трансформації. Важлива частина обробки – приведення спектрограм до фіксованого розміру, щоб вони відповідали вхідному формату моделі. Спектрограми обрізаються або доповнювались до потрібних розмірів, що робить їх сумісними із архітектурою мережі.

Дані подаються у модель за допомогою спеціального генератора, який створює батчі спектрограм та міток. Це дозволяє ефективно обробляти великі обсяги даних без перевантаження пам'яті. Генератор динамічно обробляє спектрограми, додаючи канал вимірювання, щоб вони були сумісними з обгортковими шарами мережі.

Архітектура моделі включає кілька згорткових шарів, які витягують просторові ознаки зі спектрограм. Для покращення ефективності класифікації доцільно використовувати механізм уваги, що дозволить моделі зосереджуватись на найважливіших частинах спектрограми. На виході моделі застосовано шар `softmax` для отримання ймовірностей кожного класу. Модель треновано із використанням оптимізатора `'Adam'` і функції втрат `'sparse_categorical_crossentropy'`, що забезпечує стабільне навчання навіть при великій кількості класів.

Під час навчання постійно відслідковувались ключові метрики, такі як точність і втрати на валідаційній вибірці, що дозволяє у подальшому коригувати параметри в процесі. Результати навчання оцінюються на тестовій вибірці за допомогою класифікаційного звіту та матриці похибок. Отримані результати

демонстрували стабільне зростання точності на кожній епосі, хоча процес навчання вимагає значно більше часу, ніж у попередньо-натренованих моделей, такими як ResNet та YAMNet. Незважаючи на те, що pretrained моделі демонструють високу точність уже на початкових етапах, власні моделі показують значний потенціал до вдосконалення при подальшому тренуванні, як показано на рис. 3.7.

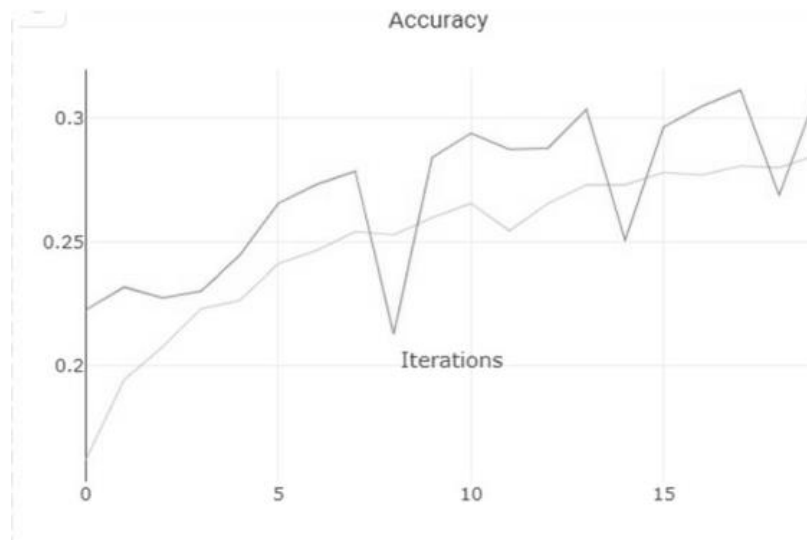


Рис. 3.7. Зростання точності у оцінках моделі із тривалістю навчання

У результаті, налаштовані власні моделі є найбільш перспективними для виконання завдання класифікації пташиного співу, оскільки дозволяють врахувати специфіку даних і забезпечують високий рівень гнучкості в роботі з параметрами. Це стало важливим внеском у дослідження, демонструючи переваги комплексного підходу до побудови моделей машинного навчання.

3.3. Оцінка результатів та перспективи використання

Усі протестовані моделі, як власно натреновані, так і попередньо натреновані демонструють незадовільні результати для певних класів, що представлено на рисунку 3.8, що вказує на важливість перегляду підходів до попереднього опрацювання даних.

	precision	recall	f1-score	support
b	0.49	0.50	0.49	48
mh	0.11	0.06	0.08	17
ms	1.00	0.25	0.40	4
na	1.00	0.33	0.50	6

Рис. 3.8. Представлення метрик щодо певних класів у наборі даних

Ця таблиця містить результати оцінки класифікаційної моделі за метриками точності (precision), повноти (recall), значення F1 (f1-score) та кількості прикладів у кожному класі (support). Аналіз цих результатів дозволяє зробити кілька важливих висновків про продуктивність моделі та проблеми з класифікацією.

Для класу «b» точність – 0.49, повнота: 0.50, значення F1: 0.49. Тобто, результати є середніми. Модель часто помиляється як при віднесенні прикладів до цього класу, так і у пропусках реальних прикладів, що належать до нього. Кількість прикладів (support = 48) свідчить, що модель мала дані для навчання, але все одно продуктивність залишається незадовільною. Це може вказувати на те, що дані цього класу перетинаються за характеристиками з іншими класами.

Для класу «mh» точність – 0.11, повнота: 0.06, а значення F1: 0.08. Це дуже низькі значення для всіх метрик. Вони свідчать про значну проблему з розпізнаванням класу.

Клас «ms» та «na» представлений точністю 1.00, повнота: 0.25, F1: 0.40. І хоча модель демонструє високу точність, дуже низька повнота свідчить про те, що вона майже не визначає приклади цього класу. Це вказує на проблему з дисбалансом даних (support = 4). Настільки менша кількість прикладів для класів "ms" та "na" значно впливає на продуктивність моделі, що проявляється у високій точності, але дуже низькій повноті. Низький рівень значення F1-score для класу "mh" свідчить про те, що дані цього класу, ймовірно, мають значну схожість з іншими класами або недостатньо добре представлені у спектрограмах.

3.4. Висновки по розділу

Однією з можливих причин наявності класів що погано класифікуються може бути невідповідність використовуваних методів перетворення аудіосигналів у спектральне представлення вимогам моделей. Наприклад, застосування мел-спектрограм або constant-Q спектрограм без попередньої нормалізації аудіоданих може обмежувати ефективність збереження ключових акустичних характеристик сигналу. Це особливо помітно для класів із низькою представленістю в наборі даних, де недостатня кількість зразків ускладнює навчання моделей.

Застосування нормалізації перед виконанням перетворень у спектрограми може бути потенційним способом поліпшення якості спектрального представлення. Нормалізація дозволяє вирівняти амплітуди сигналів, зменшуючи вплив гучності запису чи незначних змін у динамічному діапазоні, що могло б забезпечити більш стабільні характеристики для моделі. Однак навіть після реалізації таких підходів результати залишаються незадовільними для класів із невеликою кількістю екземплярів у наборі даних. Це підкреслює, що самі по собі попередні обробки не можуть повністю розв'язати проблему, яка глибше пов'язана з балансом і репрезентативністю навчального датасету.

Моделі часто стикаються з труднощами у випадках, коли класи представлені кількома записами, адже навіть при нормалізації або інших формах підготовки даних їхня здатність до генералізації для таких класів залишається обмеженою. У таких ситуаціях можуть бути корисними методи аугментації даних, які штучно збільшують кількість прикладів для слабопредставлених класів, або використання підходів перенесення навчання (transfer learning), де модель адаптується до специфіки завдання з мінімальною кількістю нових даних.

Однак, навіть ці методи не завжди можуть гарантувати задовільний результат, особливо у задачах, де класи мають суттєво схожі акустичні характеристики або знаходяться у частотно складному діапазоні, де значна частина інформації перекривається із шумами чи іншими сигналами.

РОЗДІЛ 4

ОХОРОНА ПРАЦІ ТА БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ

4.1 Охорона праці

Кваліфікаційна робота магістра присвячена дослідженню методів і засобів моніторингу звуків для аналізу пташиного співу в умовах підвищеного шумового фону, що має велике значення для біологічних досліджень і моніторингу біорізноманіття. При виконанні експериментальних досліджень і роботи з апаратно-програмними комплексами потрібно дотримуватись нормативних вимог з охорони праці, техніки безпеки та інших заходів безпеки, наприклад у поводженні із електроприладами. Основними регламентуючими документами охорони праці при роботі з електронним обладнанням є НПАОП 0.00-7.15-18 «Вимоги щодо безпеки та захисту здоров'я працівників під час роботи з екранними пристроями» і ДБН В.2.5-28-2018, що визначають норми природного і штучного освітлення.

У польових умовах робота передбачає застосування мобільних комп'ютерних пристроїв і мікроконтролерів. Для безпечної експлуатації цієї техніки враховано:

- захист від впливу зовнішніх факторів (вологість, пил, температура);
- використання портативних джерел живлення, які відповідають стандартам безпеки, зокрема СТУ EN 60950-1:2015;
- мінімізацію фізичних ризиків при монтажі, демонтажі та експлуатації обладнання.

При розміщенні обладнання у приміщеннях обсерваторій чи біологічних лабораторій забезпечено належні параметри мікроклімату: температура повітря 18–24°C, відносна вологість 40–60%, рівень шуму не більше 50 дБ. Також передбачено захист електрообладнання від перенапруги і короткого замикання через використання стабілізаторів і джерел безперебійного живлення.

У приміщеннях із обладнанням дотримуються правил освітлення: для природного освітлення використовується північна або північно-східна

орієнтація вікон, а штучне освітлення забезпечують енергоефективні люмінесцентні лампи. Поверхні стін та стелі покриті матеріалами з матовим відбиванням, що запобігає створенню відблисків на моніторах.

Екранні пристрої систем розміщено на відстані 600–700 мм від очей користувача, що відповідає ергономічним вимогам. Робочі місця обладнані стільцями, які регулюються за висотою і нахилом спинки, а також столами, що забезпечують достатній простір для комфортної роботи.

Для польових досліджень з використанням мікрофонів і сенсорів передбачено вологозахищені корпуси і захист обладнання від механічних пошкоджень. У разі роботи на відкритих майданчиках забезпечується захист від опадів і прямого сонячного випромінювання.

Дотримання цих вимог дозволить забезпечити безпечні умови праці для дослідників, знизити ризики негативного впливу обладнання на їхнє здоров'я та створити комфортні умови для проведення біологічних досліджень і моніторингу міграції птахів.

4.2. Вплив надзвичайних ситуацій природного характеру на надійність роботи засобів відстеження

Надзвичайні ситуації природного характеру, такі як повені, урагани, землетруси, лісові пожежі та інші катастрофи, значно впливають на надійність роботи засобів відстеження. Вони можуть викликати пошкодження або руйнування обладнання, переривання каналів зв'язку, втрату доступу до джерел живлення та створювати перешкоди для отримання даних. Наприклад, сильні вітри та опади здатні пошкоджувати антени, датчики та інші елементи систем спостереження, що ускладнює отримання і передачу інформації. Також природні катаклізми можуть змінювати характеристики місцевості, такі як рельєф або рослинний покрив, що впливає на якість та точність роботи сенсорів і камер. Часткове або повне руйнування інфраструктури створює труднощі в доступі до систем для технічного обслуговування та ремонту, а значні зони ураження можуть перешкоджати мобільності обслуговуючого персоналу. Електромагнітні

збурення, спричинені блискавками або магнітними бурями, можуть викликати збої в роботі систем зв'язку і датчиків, знижуючи точність переданих даних або роблячи їх недоступними. У районах, що постраждали від надзвичайних ситуацій, нерідко виникають проблеми з доступом до інтернету або мобільного зв'язку, що ще більше ускладнює моніторинг. Інтенсивні опади або пилові бурі можуть створювати візуальні перешкоди, що обмежує ефективність систем спостереження, які базуються на візуальному аналізі. В умовах підвищеної вологості або екстремальних температур технічні пристрої можуть зазнавати прискореного зношування, що скорочує їх термін служби та підвищує ризик виходу з ладу. Водночас відсутність оперативного обслуговування в умовах надзвичайної ситуації унеможливорює швидке усунення пошкоджень і відновлення роботи систем.

Система DymoNet має значний потенціал для вирішення задач забезпечення надійності роботи засобів відстеження в умовах надзвичайних ситуацій природного характеру[20]. Однією з ключових переваг цієї технології є її здатність автоматично створювати гнучкі та динамічні мережі передачі даних, які можуть адаптуватися до змінюваних умов середовища. Це особливо важливо в ситуаціях, коли звичайна інфраструктура зв'язку зазнала пошкоджень або стала недоступною.

DymoNet дозволяє пристроям у мережі автономно встановлювати зв'язки між собою, використовуючи маршрути, які мінімізують затримки передачі даних і максимально враховують доступні ресурси. Завдяки цьому система здатна забезпечити передачу критично важливої інформації навіть у випадках серйозних руйнувань інфраструктури або значних географічних змін. У надзвичайних ситуаціях така функціональність дозволяє оперативно збирати дані з постраждалих територій і передавати їх для подальшого аналізу, що сприяє прийняттю швидких і точних рішень.

Важливою перевагою DymoNet є можливість інтеграції з широким спектром пристроїв, зокрема сенсорами, дронами, мобільними пристроями та станціями моніторингу. Ця сумісність дозволяє створювати розподілені мережі, які охоплюють великі території та забезпечують безперервний потік даних,

необхідних для оцінки масштабу надзвичайної ситуації та координації рятувальних операцій.

Ще одним перспективним аспектом є здатність DymoNet працювати в умовах обмежених ресурсів, наприклад, при низькому рівні заряду батареї або обмеженій пропускну здатності. Використання ефективних алгоритмів маршрутизації дозволяє зберігати працездатність мережі навіть за мінімальної доступності енергоресурсів. Крім того, технологія здатна працювати в умовах високої мобільності вузлів, наприклад, коли дрони або рятувальні транспортні засоби пересуваються територією.

Використання DymoNet відкриває перспективи для автоматизації моніторингу в режимі реального часу в районах, що постраждали від стихійного лиха. Це може включати аналіз змін у довкіллі, виявлення пошкоджень інфраструктури та моніторинг стану постраждалих. У майбутньому така технологія може бути вдосконалена за рахунок інтеграції з системами штучного інтелекту для автоматичного визначення пріоритетів передачі даних або адаптації до специфічних умов катастрофи. Схема архітектури даної системи подана на рис. 4.1.

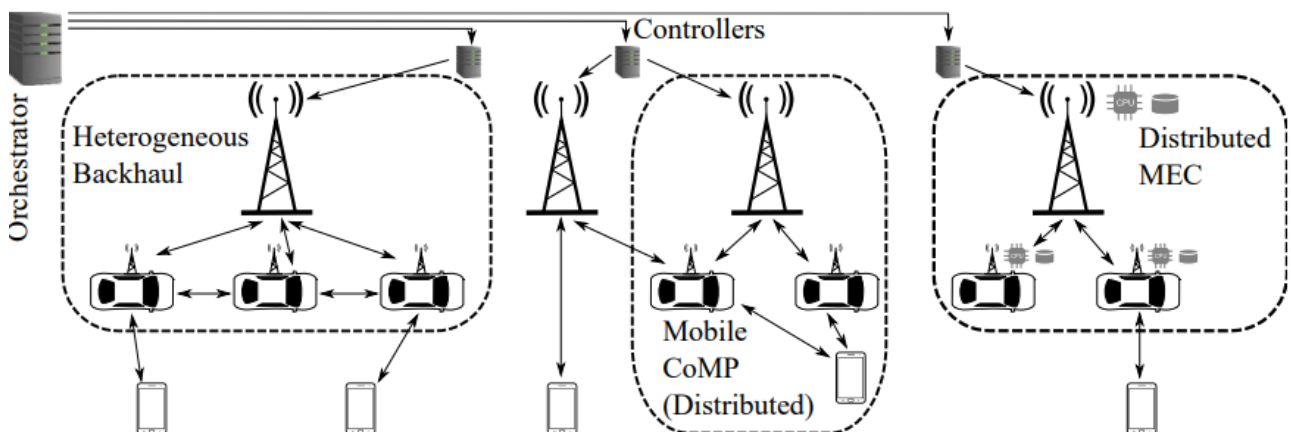


Рис. 4.1. Архітектура DymoNet[20]

Таким чином, DymoNet є потужним інструментом для підвищення стійкості та надійності систем відстеження в екстремальних умовах. Її

застосування може значно покращити здатність реагувати на надзвичайні ситуації, а також забезпечити ефективну комунікацію і передачу даних, що критично важливо для успішного подолання наслідків природних катаклізмів.

Також у надзвичайних ситуаціях можуть бути використані Unicast та VANETs, ці технології можуть бути ефективними для забезпечення зв'язку та передачі даних у надзвичайних ситуаціях, зокрема в умовах, де інфраструктура пошкоджена або недоступна [21]. Unicast передбачає передачу даних від одного відправника до одного конкретного отримувача, забезпечуючи точність і мінімальне навантаження на мережу. Ця технологія особливо ефективна для адресованих повідомлень, таких як команди для рятувальників, передача критичних даних між сенсорами та центральними станціями або надсилання медичних запитів. Завдяки прямому зв'язку Unicast забезпечує цілеспрямовану доставку інформації, що є важливим у ситуаціях, де кожна секунда має значення.

VANETs, або мережі транспортних засобів, є динамічними ad-hoc мережами, що дозволяють обмінюватися даними між транспортними засобами або іншими мобільними вузлами без централізованої інфраструктури. Ця технологія створена для роботи в умовах високої мобільності вузлів і забезпечує швидке розгортання, стійкість до пошкоджень мережі та адаптацію до змін. У критичних ситуаціях VANETs можуть використовуватися для координації рятувальних операцій, передаючи геолокаційну інформацію, статус виконання завдань чи зведення про небезпечні зони. Вони також підходять для збору даних із сенсорів, які фіксують стан довкілля або наявність постраждалих, а також для забезпечення навігації в реальному часі, зокрема обміну даними про стан доріг, перешкоди чи обхідні шляхи. Приклад використання таких мереж подано на рис. 4.2.



Рис. 4.2. Застосування мереж передачі даних у надзвичайних ситуаціях

Об'єднання можливостей Unicast і VANETs забезпечує гнучкість і ефективність у роботі мереж надзвичайного реагування. Наприклад, Unicast у межах VANETs дозволяє передавати персоналізовані повідомлення між окремими вузлами, такими як рятувальні транспортні засоби чи сенсори. Це дає змогу мінімізувати обсяг даних, що передаються, і забезпечити швидкий доступ до потрібної інформації. Разом ці технології утворюють стійкі та адаптивні мережі, здатні забезпечувати обмін критично важливою інформацією навіть у найбільш складних умовах. Їх використання відкриває нові можливості для підвищення ефективності реагування на надзвичайні ситуації, особливо в зонах із сильно пошкодженою інфраструктурою.

Застосування сучасних підходів у передачах даних значно полегшує роботу у надзвичайних ситуаціях, та дає змогу екстреним службам ефективно та швидко реагувати на виклики, що постають перед ними через критичні кліматичні умови чи інші стихійні явища[22,23]. Впровадження та масштабування систем IoT сприяє розвитку безпекових аспектів, оскільки системи які використовуються із комерційними цілями також містять у собі знаряддя які можуть бути використані для покращення чи впровадження методів охорони та безпеки життєдіяльності.

ВИСНОВКИ

У процесі виконання дослідження було реалізовано кілька важливих етапів, спрямованих на побудову системи автоматичного розпізнавання пташиного співу за допомогою глибоких нейронних мереж та спектрального аналізу аудіосигналів. Виконані завдання включали розробку власних моделей, тестування їх на мел-спектрограмах і спектрограмах constant-Q, а також порівняння результатів із попередньо натренованими моделями, такими як ResNet та YAMNet. У результаті було виявлено кілька важливих аспектів і нюансів. Згідно із поставленими до роботи завданнями було виконано наступні вимоги:

1. Проведено аналіз сучасних методів збору й обробки даних для екологічного моніторингу. Виявлено, що автоматизація процесів розпізнавання пташиного співу може значно знизити витрати та підвищити ефективність моніторингу біорізноманіття завдяки мінімізації участі людини. Особливу увагу приділено методам спектрального аналізу аудіосигналів, зокрема мел-спектрограмам і спектрограмам constant-Q, а також нейромережевим підходам.

2. Досліджено можливості застосування алгоритмів машинного навчання для обробки аудіоданих. Проведено порівняння власних моделей із попередньо натренованими архітектурами, такими як ResNet і YAMNet. Виявлено, що моделі, які використовують спектрограми constant-Q, демонструють кращу початкову точність у порівнянні з моделями на основі мел-спектрограм. Це підкреслює значення обраного способу представлення аудіо для точності навчання.

3. Розроблено модель для автоматичного розпізнавання пташиних голосів, яка враховує реальні акустичні умови. Власні моделі продемонстрували потенціал для розпізнавання пташиного співу навіть в умовах підвищеного шуму. Однак час їхнього навчання виявився значно довшим у порівнянні з натренованими моделями, що потребує оптимізації архітектури й гіперпараметрів.

4. Проведено оцінку ефективності розробленої системи на основі отриманих результатів. Балансування класів і розширення датасету через збільшення навчальних прикладів дозволили уникнути роботи з початковим дисбалансом та суттєво покращили навчання моделей. Крім того, використання методів перевірки фонових шумів зменшило ризик навчання на сторонніх звуках.

Таким чином, розроблена система розпізнавання пташиного співу може стати основою для створення малогабаритних пристроїв з енергоефективними алгоритмами, що дозволить її широке впровадження у сферу екологічного моніторингу. Подальша автоматизація включає розробку систем, які об'єднують всі етапи від збору даних, їх фільтрації, аналізу, аж до навчання моделей та їх валідації. Це дозволить розробникам і дослідникам зосередитись на вдосконаленні алгоритмів та впровадженні нових методів обробки даних для ще точнішого й ефективнішого розпізнавання пташиного співу.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Луцик Н.С., Луцків А.М., Осухівська Г.М., Тиш Є.В. Програма та методичні рекомендації з проходження практики за тематикою кваліфікаційної роботи для студентів спеціальності 123 «Комп'ютерна інженерія» другого (магістерського) рівня вищої освіти усіх форм навчання. Тернопіль. ТНТУ. 2024. 45 с.
2. Луцик Н. С., Луцків А. М., Осухівська Г. М., Тиш Є. В. Методичні рекомендації до виконання кваліфікаційної роботи магістра для студентів спеціальності 123 «Комп'ютерна інженерія» другого (магістерського) рівня вищої освіти усіх форм навчання. Тернопіль. ТНТУ. 2024. 44 с.
3. Варавін А.В., Лещишин Ю.З., Чайковський А.В. Методичні вказівки до виконання курсового проєкту з дисципліни «Дослідження і проєктування комп'ютерних систем та мереж» для здобувачів другого (магістерського) рівня вищої освіти спеціальності 123 «Комп'ютерна інженерія» усіх форм навчання. Тернопіль. ТНТУ, 2024. 32 с.
4. Сороківська Н.В., Яцишин В.В. дослідження актуальності моніторингу біорізноманіття за допомогою комп'ютеризованих систем. Матеріали XII науково-технічної конференції Тернопільського національного технічного університету імені Івана Пулюя «Актуальні задачі сучасних технологій» (11-12 грудня 2024 року). Тернопіль: ТНТУ. 2024. С. 478
5. Сороківська Н.В., Яцишин В.В. Використання адаптованих моделей штучного інтелекту для ідентифікації птахів у природних умовах. Матеріали XII науково-технічної конференції Тернопільського національного технічного університету імені Івана Пулюя «Інформаційні моделі системи та технології» (18-19 грудня 2024 року). Тернопіль: ТНТУ. 2024. С. 142.
6. Ribeiro M. T., et al. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. ACM Conference on Knowledge Discovery and Data Mining, 2016.
7. Abeßer J., Schmitt T., A review of deep learning based methods for acoustic scene classification. 2020.

8. Prof. Dr. Martin Golz. Lectures on Computational Intelligence in Applied Computer Science. Schmalkalden, SUAS, 2023.
9. Krajewski K., Batliner A., Golz M. Acoustic sleepiness detection: Framework and Validation of a speech-adapted pattern recognition approach. European Conference of Behaviour Research Methods. 2009, 41(3), pp. 795-804.
10. Xie J., et al. Audio-based Snore Detection Using Deep Neural Networks. IEEE Journal of Biomedical and Health Informatics, 2021.
11. Xia, X.; Togneri, R.; Sohel, F.; Zhao, Y.; Huang, D. A Survey: Neural Network-Based Deep Learning for Acoustic Event Detection. In Circuits, Systems, and Signal Processing; Springer: Berlin, Germany, 2019; pp. 3433–3453.
12. Dang, A.; Vu, T.H.; Wang, J.C. Acoustic Scene Classification using Convolutional Neural Networks and Multi-Scale Multi-Feature Extraction. In Proceedings of the IEEE International Conference on Consumer Electronics (ICCE), Hue City, Vietnam, 18–20 July 2018.
13. Applying Machine Learning in Audio Signal Analysis. URL: <https://medium.com/octave-john-keells-group/applying-machine-learning-in-audio-signal-analysis-e6cd013783d4> (дата звернення 04.12.2024).
14. Roche F., Hueber T., Limier S., Girin L. Autoencoders for music sound modelling: a comparison of linear, shallow, deep, recurrent and variational model. 2018.
15. Stowell D., et al. Bird Species Recognition in Audio. IEEE Journal of Selected Topics in Signal Processing, 2019.
16. Singh J. An introduction to audio processing and machine learning using Python. The pyAudioProcessing library classifies audio into different categories and genres. URL: <https://opensource.com/article/19/9/audio-processing-machine-learning-python> (дата звернення 01.12.2024).
17. Schmidhuber J. Deep learning in neural networks: An overview. Neural Networks. Vol. 61. N. 1. January 2015. pp. 85–117
18. Satya Mallick. Machine Learning for IoT Sensors: A Practical Guide. URL: <https://learnopencv.com/machine-learning-for-iot-sensors/> (дата звернення 03.12.2024).

19. Mohammed M., El-Shafie, H., Munir M. Development and Validation of Innovative Machine Learning Models for Predicting Date Palm Mite Infestation on Fruits. *Agronomy* 2023. 13(2):494.

20. Sarang Narkhede. Understanding AUC - ROC Curve. URL: <https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5> (дата звернення 01.12.2024)

21. Laura Finarelli, Falko Dressler, Marco Ajmone Marsan and Gianluca A. Rizzo, "Mobile Networks on the Move: Optimizing Moving Base Stations Dynamics in Urban Scenarios," *Proceedings of 15th IEEE Vehicular Networking Conference (VNC 2024)*, Kobe, Japan, May 2024, pp. 101–104.

22. Falko Dressler, Florian Klingler and Gianluca A. Rizzo, "Dynamic Mobile Base Stations in 5G Networks – The Moving Network Paradigm," in *The 5G Italy Book 2019: a Multiperspective View of 5G*, Marco Ajmone Marsan, Nicola Blefari Melazzi, Stefano Buzzi and Sergio Palazzo (Eds.), *Consorzio Nazionale Interuniversitario per le Telecomunicazioni (CNIT)*, 2019, pp. 477–492.

23. Yatsyshyn V., Pastukh O., Palamar A., Zharovskyi R. Technology of relational database management systems performance evaluation during computer systems design. *Scientific Journal of TNTU (Tern.)*. Vol. 109. No 1. 2023. pp. 54–65.

24. Pastukh O., Yatsyshyn V. Development of software for neuromarketing based on artificial intelligence and data science using high-performance computing and parallel programming technologies. *Scientific Journal of TNTU. Tern.: TNTU*, 2024. Vol 113. No 1. P. 143–149.

25. Yatsyshyn V., Pastukh O., Lutskevych A., Tsymbalistyy V., Martsenko N. A Risks management method based on the quality requirements communication method in Agile approaches. *CEUR Workshop Proceedings*. 2022. Vol. 3309. *Information Technologies: Theoretical and Applied Problems (ITTAP 2022)* : proc. of the 2nd Intern. Workshop, Ternopil, Ukraine, November 22-24, 2022. pp. 1-10.

26. Навчальний посібник «ТЕХНОЕКОЛОГІЯ ТА ЦИВІЛЬНА БЕЗПЕКА. ЧАСТИНА «ЦИВІЛЬНА БЕЗПЕКА»» / автор-укладач В.С. Стручок. Тернопіль: ФОП Паляниця В. А. 156 с.

27. Методичний посібник для здобувачів освітнього ступеня «магістр» всіх спеціальностей денної та заочної (дистанційної) форм навчання «БЕЗПЕКА В НАДЗВИЧАЙНИХ СИТУАЦІЯХ» / В.С. Стручок –Тернопіль: ФОП Паляниця В. А.156 с.

28. НПАОП 0.00-7.15-18 «Вимоги щодо безпеки та захисту здоров'я працівників під час роботи з екранними пристроями». Київ. 2018.

29. Катренко Л.А., Катренко А.В. Охорона праці в галузі комп'ютингу. Львів: Магнолія-2006. 2012. 544 с.

ДОДАТКИ

Додаток А
Тези конференцій

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Тернопільський національний технічний університет імені Івана Пулюя (Україна)
Університет імені П'єра і Марії Кюрі (Франція)
Маріборський університет (Словенія)
Технічний університет у Кошице (Словаччина)
Вільнюський технічний університет ім. Гедимінаса (Литва)
Наукове товариство ім. Т.Шевченка

АКТУАЛЬНІ ЗАДАЧІ СУЧАСНИХ ТЕХНОЛОГІЙ

**Збірник
тез доповідей**

**ХІІІ Міжнародної науково-практичної
конференції молодих учених та студентів
11-12 грудня 2024 року**



**УКРАЇНА
ТЕРНОПІЛЬ – 2024**

45. **М.В. Дрогобицький, А.І. Фіялка, Н.С. Луцик** 466
 АНАЛІЗ МЕТОДІВ ІНТЕГРАЦІЇ MICROGRID ДЛЯ ДИНАМІЧНОГО
 БАЛАНСУВАННЯ НАВАНТАЖЕННЯ ЗАГАЛЬНИХ ЕЛЕКТРИЧНИХ МЕРЕЖ
46. **М.В. Королевич; А.М. Галабіцький; С.Т. Гавриць** 467
 ДОСЛІДЖЕННЯ ЧУТЛИВОСТІ КОНДУКТОМЕТРИЧНОГО ДАВАЧА
47. **М.Г. Білик, І.Г.Білик, І.В. Чихіра** 469
 РОЗРОБКА АВТОМАТИЗОВАНОЇ СИСТЕМИ КЕРУВАННЯ ЛІНІЇ ПОДАЧІ
 МЕТАЛЕВОГО ПРОФІЛЮ НА БАЗІ КОНТРОЛЕРА ФІРМИ SIEMENS
48. **М.І. Богунський** 470
 ОПТИМІЗАЦІЯ АЛГОРИТМІВ ПОШУКУ ІНФОРМАЦІЇ НА ВЕБ-САЙТІ З
 ВИКОРИСТАННЯМ МЕТОДІВ МАШИННОГО НАВЧАННЯ
49. **М.О. Скоробогата, Л.П. Дмитроца** 472
 ВПЛИВ ГЕНЕРАТИВНОГО ШІ НА МАРКЕТИНГОВІ КОМУНІКАЦІЇ КОМПАНІЇ
 "TECHNOVAAPP"
50. **Микитишин А.Г., к.т.н., доц., Чуревич Б.В., Дильовий О.О.** 474
 ПРОМИСЛОВІ СИСТЕМИ ЕНЕРГОМОНІТОРИНГУ НА ОСНОВІ ІНТЕРНЕТУ
 РЕЧЕЙ
51. **Н. М. Головецький, І.О. Баран,** 477
 ОГЛЯД ІСНУЮЧИХ ПРОЕКТІВ LPWAN
52. **Н. Сороківська, В.Яцишин** 478
 КОМП'ЮТЕРИЗОВАНІ СИСТЕМИ МОНІТОРИНГУ БІОРИЗНОМАНІТТА З
 ВБУДОВАНИМИ ІНТЕЛЕКТУАЛЬНИМИ СЕРВІСАМИ
53. **Н.М. Топольницький, Р.М. Небесний** 479
 ЗАСТОСУВАННЯ GEOMETRY NODES В ПРОГРАМНОМУ ПАКЕТІ BLENDER ДЛЯ
 ЗАДАЧ ПРОЦЕДУРНОГО МОДЕЛЮВАННЯ
54. **Н.М. Топольницький, Р.М. Небесний** 480
 ЗАСТОСУВАННЯ GEOMETRY NODES В ПРОГРАМНОМУ ПАКЕТІ BLENDER ДЛЯ
 ЗАДАЧ ПРОЦЕДУРНОГО МОДЕЛЮВАННЯ
55. **О.А. Заяць** 471
 МЕТОДИ ТА ІНСТРУМЕНТИ АНАЛІЗУ І КЛАСИФІКАЦІЇ ВИМОГ ДО
 ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ
56. **О.А. Саган, К.Б. Швирло** 482
 АНАЛІЗ SQL-ІН'ЄКЦІЙ:ТИПИ АТАК ТА МЕТОДИ ЗАХИСТУ
57. **О.В. Смолій; А.Г. Микитишин к.т.н.; І.В. Булич** 456
 ОПТИМІЗАЦІЯ ВКЛЮЧЕННЯ НАВАНТАЖЕННЯ ПРИ РЕЗЕРВНОМУ ЖИВЛЕННІ
 НА МАЛОМУ ВИРОБНИЧОМУ ПІДПРИЄМСТВІ ТА ЙОГО ІНТЕГРАЦІЯ З
 ТЕХНОЛОГІЧНИМ ПРОЦЕСОМ
58. **О.В. Шкварок; А.О. Курило; І.Р. Козбур; Ю.Б. Капаціла** 483
 РОЗРОБКА АВТОМАТИЗОВАНОЇ СИСТЕМИ РОЗУМНОГО БУДИНКУ НА БАЗІ
 ІНТЕРНЕТУ РЕЧЕЙ ТА БЕЗДРОТОВИХ МЕРЕЖ
59. **О.Д. Пакон, Р.М. Небесний** 485
 АНАЛІЗ ЕФЕКТИВНОСТІ ВПРОВАДЖЕННЯ MDM СИСТЕМ ДЛЯ
 КОРПОРАТИВНОГО СЕРЕДОВИЩА
60. **О.Д. Пакон, Р.М. Небесний** 487
 ОСНОВНІ ВИКЛИКИ ДЛЯ СИСТЕМ КЕРУВАННЯ МОБІЛЬНИМИ ПРИСТРОЯМИ

УДК 004.8-9: 656.07

Н. Сороківська, В.Яцишин, к.т.н., доц.

Тернопільський національний технічний університет імені І. Пулюя, Україна

КОМП'ЮТЕРИЗОВАНІ СИСТЕМИ МОНІТОРИНГУ БІОРІЗНОМАНІТТЯ З ВБУДОВАНИМИ ІНТЕЛЕКТУАЛЬНИМИ СЕРВІСАМИ

N

S

o

r

o

k

r

r

r

r

u

u

u

u

u

u

u

u

u

u

h

u

u

u

u

u

u

u

u

u

u

u

u

u

u

u

u

u

u

u

u

u

u

u

u

u

u

u

u

u

u

u

u

u

u

u

COMPUTERIZED SYSTEMS OF BIODIVERSITY MONITORING USING INTELLIGENT SERVICES

Моніторинг біорізноманіття є важливим аспектом збереження екосистем у контексті глобальних екологічних змін. У сучасних дослідженнях усе частіше застосовуються комп'ютеризовані системи, які базуються на компонентах інтелектуального аналізу природних даних. В якості таких компонентів доцільно використовувати програмно реалізовані алгоритми машинного навчання, які здатні забезпечувати високу точність та ефективність при розв'язанні задач класифікації об'єктів живої природи. Наприклад, на основі аналізу набору аудіозаписів звуків птахів проводиться їхня ідентифікація.

Сучасні технології, як-от радари, інфрачервоні камери, мікрофони з високою чутливістю забезпечують процес збору природних даних. Водночас, алгоритми машинного навчання відкривають нові можливості для аналізу великих обсягів інформації. Застосування глибоких нейронних мереж за наявності великих наборів аудіо-даних дають змогу забезпечити високу точність розпізнавання звуків у природі. Доказом ефективності застосування підходу deep learning для задачі виявлення подій на основі аналізу звукових сигналів є дослідження [1], де точність розпізнавання досягає 95,3%. Запропонована автором структура глибокої нейронної мережі передбачає використання трьох згорткових шарів (CNN) та одного шару рекурентного (RNN). Це дало змогу забезпечити автоматичне добування ознак із двовимірних спектральних представлень звуку, отриманих через постійне Q- перетворення (CQT).

Окрім цього, критично важливо при проведенні досліджень екосистем, розуміти та виявляти ті характеристики спектрограм звукових сигналів, які є визначальними при ідентифікації конкретного виду птахів. Так, у [2], автор наводить два методи, які дозволяють підвищити прозорість моделей і сприяє довірі до їх використання в наукових і практичних застосуваннях. Подібні підходи можуть бути використані для аналізу аудіозаписів пташиних співів, що забезпечить автоматичну ідентифікацію видів птахів навіть за наявності фонових шумів. Симбіоз методів визначення важливості ознак у спектрограмі звукових сигналах та підходу глибоких нейронних мереж відкриває можливість моніторингу популяцій птахів у важкодоступних регіонах та віддалених екосистемах, що на сьогодні є актуальним і не до кінця розв'язаним завданням.

Таким чином, комп'ютеризовані системи з вбудованими інтелектуальними сервісами на основі алгоритмів штучного інтелекту формують перспективний напрям при проведенні моніторингу біорізноманіття. Розробка та впровадження таких систем забезпечує автоматизацію процесів збору та аналізу даних, сприяє ефективності прийняття рішень щодо збереження екосистем.

Література

1. Xie J., et al. Audio-based Snore Detection Using Deep Neural Networks. IEEE Journal of Biomedical and Health Informatics, 2021
2. Ribeiro M. T., et al. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. ACM Conference on Knowledge Discovery and Data Mining, 2016.

**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ТЕРНОПІЛЬСЬКИЙ НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ
імені ІВАНА ПУЛЮЯ
НАУКОВЕ ТОВАРИСТВО ім. ШЕВЧЕНКА**

МАТЕРІАЛИ

ХІІ НАУКОВО-ТЕХНІЧНОЇ КОНФЕРЕНЦІЇ

**«ІНФОРМАЦІЙНІ МОДЕЛІ,
СИСТЕМИ ТА ТЕХНОЛОГІЇ»**



18-19 грудня 2024 року

**ТЕРНОПІЛЬ
2024**

I. О. Кухар, Г. М. Осухівська ІНТЕГРАЦІЯ ВІДНОВЛЮВАЛЬНИХ ДЖЕРЕЛ ЕНЕРГІЇ В РОЗУМНІ ЕЛЕКТРИЧНІ МЕРЕЖІ З ВИКОРИСТАННЯМ ПЕРЕДОВИХ ТЕХНОЛОГІЙ	
I. O. Kukhar, H. Osukhivska INTEGRATION OF RENEWABLE ENERGY SOURCES INTO SMART GRIDS USING ADVANCED TECHNOLOGIES	135
Ю. Лешишин, А. Герасименко, О. Герасименко КОМП'ЮТЕРНА СИСТЕМА МОНІТОРИНГУ РАДІОЗВ'ЯЗКУ ВІД БПЛА	
Yu. Leshchysyn, A. Herasymenko, O. Herasymenko COMPUTER SYSTEM FOR MONITORING RADIO COMMUNICATIONS FROM UAVS	136
А. Луцків, А. Люлька ЗАСТОСУВАННЯ АЛГОРИТМІВ БАЛАНСУВАННЯ НАВАНТАЖЕННЯ В ПРОЦЕСАХ СУЧАСНИХ ОПЕРАЦІЙНИХ СИСТЕМ	
A. Lutskevych, A. Liulka APPLICATION OF LOAD BALANCING ALGORITHMS IN THE PROCESSES OF MODERN OPERATING SYSTEMS	137
Я. О. Мишаківський МЕТОДИ РОЗПІЗНАВАННЯ СТАТІ ТА ВІКУ ЗА ЗОБРАЖЕННЯМ ОСОБИ	
Ia. O. Myshakivskiy METHODS OF GENDER AND AGE RECOGNITION FROM THE IMAGE OF A PERSON	138
Я. О. Мишаківський МЕТОДИ ДЛЯ РЕАЛІЗАЦІЇ АЛГОРИТМУ РОЗПІЗНАВАННЯ СТАТІ ТА ВІКУ ЛЮДИНИ У ВІДЕОПОТОЦІ ДАНИХ	
Ia. O. Myshakivskiy METHODS FOR IMPLEMENTING AN ALGORITHM FOR RECOGNITION OF A PERSON'S GENDER AND AGE IN A VIDEO DATA STREAM	140
М. Є. Олійник, І. В. Мудрий, Н. С. Луцк МЕТОДИ ТА ЗАСОБИ ОПТИМАЛЬНОГО РОЗПОДІЛУ ЗАВДАНЬ В КОМП'ЮТЕРИЗОВАНИЙ СИСТЕМІ БЕЗПЛОТНОЇ ДОСТАВКИ	
M. Y. Oliinyk, I. V. Mudryi, N. S. Lutsyk METHODS AND TOOLS FOR OPTIMAL TASK ALLOCATION IN A COMPUTERIZED UNMANNED DELIVERY SYSTEM	141
Н. Сороківська, В. Яцишин ВИКОРИСТАННЯ АДАПТОВАНИХ МОДЕЛЕЙ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ІДЕНТИФІКАЦІЇ ПТАХІВ У ПРИРОДНИХ УМОВАХ	
N. Sorokivska, V. Yatsyshyn APPLICATION OF ADAPTED ARTIFICIAL INTELLIGENCE MODELS FOR BIRD IDENTIFICATION IN NATURAL ENVIRONMENTS	142
А. А. Станько, А. В. Гончаренко, І. В. Журик РОЗУМНІ МІСТА: ІНТЕГРАЦІЯ ТЕХНОЛОГІЙ, ДАНИХ І СТРАТЕГІЇ СТАЛОГО РОЗВИТКУ МІСЬКОЇ ЕКОСИСТЕМИ	
A. A. Stanko, A. V. Honcharenko, I. V. Zhuryk SMART CITIES: INTEGRATING TECHNOLOGIES, DATA AND STRATEGIES FOR SUSTAINABLE URBAN ECOSYSTEM DEVELOPMENT	143
А. А. Станько, С. З. Кульчицький, Ю. В. Срібний ФОРМУВАННЯ КЕРОВАНИХ ОЗЕР ДАНИХ: МЕТОДОЛОГІЇ, ІНСТРУМЕНТИ ТА ПРАКТИЧНІ АСПЕКТИ	
A. A. Stanko, S. Z. Kulchytskyi, Y. V. Sribnyi FORMATION OF MANAGED DATA LAKES: METHODOLOGIES, TOOLS AND PRACTICAL ASPECTS	144
А. Чуба, В. Тимошук, А. Микитишин, В. Шиманська СУЧАСНІ ПІДХОДИ ДО МОНІТОРИНГУ ТА УПРАВЛІННЯ В DOCSIS-МЕРЕЖАХ	
A. Chuba, V. Tymoshchuk, A. Mykytyshyn, V. Shymanska MODERN APPROACHES TO MONITORING AND MANAGEMENT IN DOCSIS NETWORKS	145

УДК 004.8-9: 656.07

Н. Сороківська, В. Яцишин, к.т.н., доц.

Тернопільський національний технічний університет імені І. Пулюя, Україна

ВИКОРИСТАННЯ АДАПТОВАНИХ МОДЕЛЕЙ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ІДЕНТИФІКАЦІЇ ПТАХІВ У ПРИРОДНИХ УМОВАХ

N. Sorokivska, V. Yatsyshyn, Ph.D. Assoc. Prof.

APPLICATION OF ADAPTED ARTIFICIAL INTELLIGENCE MODELS FOR BIRD IDENTIFICATION IN NATURAL ENVIRONMENTS

Розробка та впровадження комп'ютеризованих систем з вбудованими інтелектуальними сервісами на основі алгоритмів штучного інтелекту забезпечує автоматизацію процесів збору та аналізу даних, а також сприяє ефективності прийняття рішень щодо збереження екосистем. Це обумовлено актуальністю методів спостереження на які накладають обмеження стаціонарність пунктів, що охоплюють лише невеликі території, або вимагають значних людських ресурсів. Інші обмеження включають складність роботи в умовах фонових шумів, недостатню точність моделей і труднощі у зборі якісних даних. Апаратні пристрої з вбудованими, попередньо навченими моделями машинного навчання, здатні автоматично виконувати ідентифікацію видів птахів, оцінювати кількість особин і збирати інші дані. Це дає змогу оптимізувати часові затрати і людський ресурс при проведенні процедур аналізу. Як наслідок, впровадження технологій машинного навчання дозволяє масштабувати процес спостереження за птахами. Основною проблемою при створенні інтелектуальної системи є побудова моделі, яка б дозволила точно визначати усі наявні у наборі даних види птахів. Це, в першу чергу, пов'язано із незбалансованістю даних, оскільки певні види досліджуваних об'єктів представлені недостатньо і також наявні фонові шуми, які складно усувати. Для боротьби з дисбалансом даних застосовуються техніки resampling, зокрема oversampling, для рідкісних класів птахів. Також може використовуватися метод ковзного вікна для створення додаткових зразків із частковим перекриттям. Паралельно проводиться аугментація даних, додаються штучні варіації аудіофайлів, фонових шумів у вигляді окремих файлів, що підвищує стійкість моделі до реальних умов. Щоб надати даним зручний для аналізу формат, аудіозаписи перетворюються на спектрограми за допомогою різних методів. Важливим етапом є коректний розподіл набору даних на навчальну, валідаційну й тестову вибірки, що забезпечить об'єктивність оцінки роботи моделі і буде враховувати критичність кількості екземплярів рідкісних класів при навчанні моделі. Описані вище процедури спрямовані на забезпечення якісного представлення аудіосигналів у моделі і створення умов для її ефективного навчання.

Попередньо навчені моделі, такі як ResNet, що були адаптовані для аудіо задач, виявляються конкурентними завдяки їхній здатності працювати з великими наборами ознак. У деяких експериментах ResNet перевершує моделі CNN за точністю, особливо коли дані спектрограм є складними або містять шум. Водночас BirdNet, спеціалізована система для ідентифікації птахів, демонструє хороші результати для конкретних завдань класифікації. Однак існують обмеження її застосування для загальних задач, оскільки вона сильно орієнтована на типові для своєї архітектури сценарії.

Література

1. He, K., Zhang, X., Ren, S., & Sun, J. Deep Residual Learning for Image Recognition. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
2. Knight, E. C., et al. Acoustic monitoring reveals long-term population trends of threatened bats. Biological Conservation, 2019.

Додаток Б

Скрипт для запуску моделі на навчання

```

import os
import pandas as pd
import numpy as np
import librosa
import librosa.display
import tensorflow as tf
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.utils.class_weight import compute_class_weight
from sklearn.metrics import classification_report, confusion_matrix
from clearml import Task, Logger
import pickle

# Initialize ClearML Task
task = Task.init(project_name="Audio Classification",
task_name="Attention-based CQT Model")

# Parameters
SAMPLE_RATE = 1600
BATCH_SIZE = 64
EPOCHS = 50
DATA_DIR = "/content/drive/MyDrive/mastert/data/preprocessed"
ANNOTATIONS_FILE =
"/content/drive/MyDrive/mastert/data/metadata/sliding_window_annotations.
csv"
TARGET_SHAPE = (128, 1024) # Full-length spectrogram dimensions

# Step 1: Load annotations and filter valid files
annotations = pd.read_csv(ANNOTATIONS_FILE)
annotations['filepath'] = annotations['filename'].apply(lambda x:
os.path.join(DATA_DIR, x))
annotations = annotations[annotations['filepath'].apply(os.path.isfile)]

# Step 2: Handle rare classes and ensure stratification feasibility
min_samples_per_class = 2
label_counts = annotations['species_code'].value_counts()

# Classes with fewer samples than `min_samples_per_class` will be grouped
into "Other"
rare_classes = label_counts[label_counts < min_samples_per_class].index
annotations['species_code'] = annotations['species_code'].apply(
    lambda x: 'Other' if x in rare_classes else x
)

# Encode labels after handling rare classes

```

```

label_encoder = LabelEncoder()
annotations['encoded_label'] =
label_encoder.fit_transform(annotations['species_code'])
class_names = label_encoder.classes_
N_CLASSES = len(class_names)

# Log class names to ClearML
Logger.current_logger().report_text(f"Class names: {class_names}")

# Train-validation-test split
train_df, temp_df = train_test_split(
    annotations,
    test_size=0.3,
    stratify=annotations['encoded_label'],
    random_state=42
)
valid_df, test_df = train_test_split(
    temp_df,
    test_size=0.5,
    stratify=temp_df['encoded_label'],
    random_state=42
)

# Step 3: Compute spectrograms
def compute_full_cqt(filepath, target_shape=(128, 1024), n_bins=84,
bins_per_octave=12, sample_rate=1600):
    try:
        audio, _ = librosa.load(filepath, sr=sample_rate, mono=True)
        cqt = librosa.cqt(audio, sr=sample_rate, n_bins=n_bins,
bins_per_octave=bins_per_octave)
        cqt_db = librosa.amplitude_to_db(np.abs(cqt), ref=np.max)

        if cqt_db.shape[1] < target_shape[1]:
            pad_width = target_shape[1] - cqt_db.shape[1]
            cqt_db = np.pad(cqt_db, ((0, 0), (0, pad_width)),
mode='constant')
        elif cqt_db.shape[1] > target_shape[1]:
            cqt_db = cqt_db[:, :target_shape[1]]

        return cqt_db
    except Exception as e:
        print(f"Error processing {filepath}: {e}")
        return None

# Step 4: Data generator
class AudioDataGenerator(tf.keras.utils.Sequence):
    def __init__(self, dataframe, batch_size, shuffle=True,
target_shape=TARGET_SHAPE):
        self.dataframe = dataframe
        self.batch_size = batch_size
        self.shuffle = shuffle
        self.target_shape = target_shape

```

```

        self.indices = np.arange(len(self.dataframe))
        if self.shuffle:
            np.random.shuffle(self.indices)

    def __len__(self):
        return len(self.dataframe) // self.batch_size

    def __getitem__(self, idx):
        batch_indices = self.indices[idx * self.batch_size:(idx + 1) *
self.batch_size]
        batch_data = self.dataframe.iloc[batch_indices]

        X = []
        y = []

        for _, row in batch_data.iterrows():
            cqt_db = compute_full_cqt(row['filepath'],
target_shape=self.target_shape)
            if cqt_db is not None:
                X.append(cqt_db)
                y.append(row['encoded_label'])

        if len(X) == 0:
            return self.__getitem__((idx + 1) % len(self))

        X = np.array(X)
        X = np.expand_dims(X, axis=-1) # Add channel dimension for CNN
        y = np.array(y)

        return X, y

    def on_epoch_end(self):
        if self.shuffle:
            np.random.shuffle(self.indices)

# Instantiate data generators
train_generator = AudioDataGenerator(train_df, batch_size=BATCH_SIZE)
valid_generator = AudioDataGenerator(valid_df, batch_size=BATCH_SIZE)
test_generator = AudioDataGenerator(test_df, batch_size=BATCH_SIZE,
shuffle=False)

# Step 5: Build the attention-based model
def build_attention_model(input_shape, num_classes):
    inputs = tf.keras.layers.Input(shape=input_shape)
    x = tf.keras.layers.Conv2D(32, (3, 3), activation='relu')(inputs)
    x = tf.keras.layers.MaxPooling2D((2, 2))(x)
    x = tf.keras.layers.Conv2D(64, (3, 3), activation='relu')(x)
    x = tf.keras.layers.MaxPooling2D((2, 2))(x)
    x = tf.keras.layers.Conv2D(128, (3, 3), activation='relu')(x)
    x = tf.keras.layers.MaxPooling2D((2, 2))(x)
    x = tf.keras.layers.Reshape((x.shape[1] * x.shape[2], x.shape[3]))(x)
    attention_scores = tf.keras.layers.Dense(1, activation='softmax')(x)

```



```

        attended_features = tf.keras.layers.Multiply()([x, attention_scores])
        attended_summary =
tf.keras.layers.GlobalAveragePooling1D()(attended_features)
        outputs = tf.keras.layers.Dense(num_classes,
activation='softmax')(attended_summary)
        model = tf.keras.models.Model(inputs, outputs)
        model.compile(optimizer='adam',
loss='sparse_categorical_crossentropy', metrics=['accuracy'])
        return model

# Step 6: Build and train the model
input_shape = (*TARGET_SHAPE, 1)
model = build_attention_model(input_shape, N_CLASSES)
model.summary()

# Compute class weights
class_weights = compute_class_weight(
    class_weight='balanced',
    classes=np.unique(train_df['encoded_label']),
    y=train_df['encoded_label']
)
class_weights_dict = dict(enumerate(class_weights))

# ClearML logging during training
class ClearMLCallback(tf.keras.callbacks.Callback):
    def on_epoch_end(self, epoch, logs=None):
        logs = logs or {}
        for metric, value in logs.items():
            Logger.current_logger().report_scalar(metric, "train",
iteration=epoch, value=value)

clearml_callback = ClearMLCallback()

history = model.fit(
    train_generator,
    validation_data=valid_generator,
    epochs=EPOCHS,
    class_weight=class_weights_dict,
    callbacks=[clearml_callback]
)

# Evaluate on test data
test_loss, test_accuracy = model.evaluate(test_generator)
predictions = model.predict(test_generator)
y_true = np.concatenate([y for _, y in test_generator], axis=0)
y_pred = np.argmax(predictions, axis=1)
# Classification report
report = classification_report(y_true, y_pred, target_names=class_names,
output_dict=True)
Logger.current_logger().report_table("Classification Report", "Test",
iteration=EPOCHS, table=report)
# ...

```