Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет інформаційно-аналітичних технологій та менеджменту
(повна назва)

Кафедра прикладної математики
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА
Пояснювальна записка

рівень вищої освіти другий (магістерський)

Дослідження моделей і методів аналізу спектральних характеристик аудіозаписів для виявлення плагіату
(тема)

Виконав:
здобувач 2 року навчання, групи САУм-23-2
Цвіркун О.А.
(прізвище, ініціали)
Спеціальність
124 Системний аналіз
(код і повна назва спеціальності)
Тип програми освітньо-професійна
(освітньо-професійна або освітньо-наукова)
Освітня програма
Системний аналіз і управління
(повна назва освітньої програми)
Керівник доц. Ситникова Ю.В.
(посада, прізвище, ініціали)

Допускається до захисту

Зав. кафедри ПМ _____ Сидоров М.В.
(підпис) (прізвище, ініціали)

2025 р.

Харківський національний університет радіоелектроніки

Факультет _інформаційно-аналітичних технологій та менеджменту
Кафедра _прикладної математики
Рівень вищої освіти _другий (магістерський)
Спеціальність _124 Системний аналіз
(код і повна назва)
Тип програми _освітньо-професійна
(освітньо-професійна або освітньо-наукова)
Освітня програма _Системний аналіз і управління
(повна назва)

ЗАТВЕРДЖУЮ:

Зав. кафедри ПМ _____
(підпис)
“ _25_ ” _листопада_ 2024 р.

ЗАВДАННЯ
НА КВАЛІФІКАЦІЙНУ РОБОТУ

здобувачеві _Цвіркуну Олександру Анатолійовичу
(прізвище, ім’я, по батькові)

1. Тема роботи _Дослідження моделей і методів аналізу спектральних характеристик аудіозаписів для виявлення плагіату

затверджена наказом по університету від 22 листопада 2024 р. № 1228 Ст

2. Термін подання здобувачем роботи до екзаменаційної комісії 6 січня 2025 р.

3. Вихідні дані до роботи _математична модель задачі аналізу спектральних характеристик аудіозаписів для виявлення плагіату

4. Перелік питань, що потрібно опрацювати в роботі
1. Системний аналіз предметної області
2. Вибір і обґрунтування методу розв’язання
3. Програмна реалізація
4. Результати обчислювального експерименту
5. Аналіз можливих застосувань

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій

1. Актуальність теми роботи
2. Постановка задачі
3. Системний аналіз предметної області
4. Метод чисельного аналізу
5. Результати обчислювального експерименту

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Терміни виконання етапів роботи	Примітка
1	Підбір та вивчення технічної літератури за темою роботи	25 листопада – 1 грудня 2024 р.	виконано
2	Вибір та обґрунтування методу	2 – 8 грудня 2024 р.	виконано
3	Розробка алгоритму і програми	9 – 22 грудня 2023 р.	виконано
4	Проведення аналітичних досліджень та розрахунків	23 – 29 грудня 2024 р.	виконано
5	Робота над текстом пояснювальної записки	30 грудня 2024 р. – 9 січня 2025 р.	виконано
6	Представлення роботи на рецензію в ЕК	10 січня 2025 р.	виконано

Дата видачі завдання _25_ листопада 2024 р.

Здобувач _____
(підпис)

Керівник роботи _____ доц. Ситникова Ю.В.
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка: 73 с., 13 рис., 1 дод., 25 джерел.

АНАЛІЗ АУДІОДАНИХ, ВИЯВЛЕННЯ ПЛАГІАТУ, СПЕКТРАЛЬНІ ХАРАКТЕРИСТИКИ, МЕЛ-ЧАСТОТСНІ КАПСТРАЛЬНІ КОЕФІЦІЄНТИ, КОНВУЛЯЦІЙНА НЕЙРОННА МЕРЕЖА, ПЕРЕТВОРЕННЯ ФУР'Є, SIAMESE NETWORK.

Об'єкт дослідження – процес аналізу спектральних характеристик аудіозаписів для автоматизованого пошуку схожих аудіофрагментів з метою виявлення плагіату.

Метою роботи – підвищення ефективності (за критеріями точності та швидкодії) виявлення схожих аудіозаписів для виявлення плагіату шляхом подальшого розвитку методів аналізу спектральних характеристик аудіозаписів та адаптації рішень для їх реалізації.

Методи дослідження – методи спектрального аналізу та обробки сигналів, а також алгоритми глибокого навчання для визначення ступеня схожості аудіозаписів.

У кваліфікаційній роботі досліджується та вдосконалюється метод виявлення плагіату в аудіофайлах шляхом аналізу спектральних характеристик. Проведено системний аналіз предметної області, огляд сучасних підходів до аналізу аудіоданих, а також методів машинного навчання для порівняння аудіофайлів. Використано методи глибокого навчання на основі CNN та модифікованої архітектури Siamese Network.

Розроблено програмне забезпечення для обробки аудіоданих, що включає етапи ресемплінгу, розрахунку MFCC, та порівняння отриманих ембедінгів за допомогою косинусної схожості.

Розроблене програмне забезпечення може бути використано у сфері захисту авторських прав, аудіоаналізу та виявлення схожості між композиціями.

ABSTRACT

Introductory note: 73 pages, 13 figures, 1 appendixes, 25 sources.

AUDIO DATA ANALYSIS, PLAGIARISM DETECTION, SPECTRAL CHARACTERISTICS, MEL-FREQUENCY CEPSTRAL COEFFICIENTS, CONVOLUTIONAL NEURAL NETWORK, SIAMESE NETWORK, FOURIER TRANSFORM.

Object of research – the process of analyzing the spectral characteristics of audio recordings for the automated search of similar audio fragments to detect plagiarism.

Purpose of work – improving the efficiency (in terms of accuracy and performance) of detecting similar audio recordings for plagiarism detection through the further development of methods for analyzing the spectral characteristics of audio recordings and adapting solutions for their implementation.

Methods of research are methods of spectral analysis and signal processing, as well as deep learning algorithms for determining the degree of similarity between audio recordings.

The qualification work investigates and improves the method of plagiarism detection in audio files through the analysis of spectral characteristics. A systematic analysis of the subject area, an overview of modern approaches to audio data analysis, and machine learning methods for audio file comparison were carried out. Deep learning methods based on CNN and a modified Siamese Network architecture.

The software was developed for audio data processing, including stages of resampling, MFCC calculation, and comparison of the resulting embeddings using cosine similarity.

The developed software can be used to protect copyrights, automate audio analysis, and identify similarities between compositions.

ЗМІСТ

ПЕРЕЛІК СКОРОЧЕНЬ, УМОВНИХ ПОЗНАК, ОДИНИЦЬ І ТЕРМІНІВ

API – Application Programming Interface;

CBCD – Content-Based Copy Detection;

CENS – Chroma Energy Normalized Statistics;

CNN – Convolutional Neural Networks;

CPU – Central Processing Unit;

CQT – Constant-Q Transform;

CWT – Continuous Wavelet Transform;

DCT – Discrete Cosine Transform;

DRM – Digital Rights Management;

FFMAP – Fundamental Frequency Map;

GPU – Graphics Processing Unit;

MFCC – Mel-Frequency Cepstral Coefficients;

SCF – Successive Closest Frames;

STFT – Short-Time Fourier Transform;

VGG – Visual Geometry Group.

ВСТУП

Актуальність теми. Актуальність роботи зумовлена зростаючою кількістю аудіоконтенту, доступного в цифрових медіа та на онлайн-платформах. Генеративні моделі можуть використовувати фрагменти існуючих аудіозаписів як патерни, що дає можливість створювати нові звукові доріжки на основі вже існуючих. Це може призвести до використання частин оригінальних творів без дозволу авторів і згоди на це, що викликає зростання потреби в автоматизованих інструментах, здатних ідентифікувати схожі аудіозаписи, виявляти їхні спільні спектральні характеристики та розпізнавати можливі випадки плагіату. Це має важливе значення в епоху швидкої комерціалізації контенту, де питання порушення авторських прав і незаконного використання інтелектуальної власності набувають критичного значення.

Мета і завдання кваліфікаційної роботи. Метою кваліфікаційної роботи є підвищення ефективності (за критеріями точності та швидкодії) виявлення схожих аудіозаписів для виявлення плагіату шляхом подальшого розвитку методів аналізу спектральних характеристик аудіозаписів та адаптації рішень для їх реалізації. Для досягнення поставленої мети необхідно виконати наступні завдання:

– провести огляд і аналіз сучасного стану задачі аналізу спектральних характеристик аудіозаписів з метою пошуку схожих аудіозаписів для виявлення плагіату;

– дослідити наявні моделі та методи, спрямовані на вирішення задачі аналізу спектральних характеристик аудіозаписів з метою пошуку схожих аудіозаписів для виявлення плагіату;

– здійснити формальну постановку задачі аналізу спектральних характеристик для автоматизованого пошуку схожих аудіозаписів;

– вибрати та модифікувати моделі та методи вирішення задачі аналізу спектральних характеристик аудіозаписів з метою виявлення плагіату;

– розробити алгоритм розв’язання задачі аналізу спектральних характеристик аудіозаписів з метою виявлення плагіату;

– розробити прототип програмного забезпечення для розв’язання задачі аналізу спектральних характеристик аудіозаписів з метою виявлення плагіату.

Об’єктом дослідження є процес аналізу спектральних характеристик аудіозаписів для автоматизованого пошуку схожих аудіофрагментів з метою виявлення плагіату.

Предметом дослідження є моделі та методи аналізу спектральних характеристик аудіозаписів, що використовуються для автоматизованого пошуку схожих аудіофрагментів з метою виявлення плагіату.

Методи дослідження. У роботі використовуються методи спектрального аналізу та обробки сигналів, а також алгоритми глибокого навчання для визначення ступеня схожості аудіозаписів.

Публікації. Результати, отримані у кваліфікаційній роботі, було представлено на VI Міжнародній науково-практичній конференції «SCIENTIFIC RESEARCH: MODERN CHALLENGES AND FUTURE PROSPECTS» (м. Мюнхен, Німеччина, 20-22.01.2025) [1].

1 СИСТЕМНИЙ АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ПОСТАНОВКА ЗАДАЧ ДОСЛІДЖЕННЯ

1.1 Системний аналіз задачі аналізу спектральних характеристик аудіозаписів з метою пошуку схожих аудіозаписів для виявлення плагіату

На сьогоднішній день технології обробки аудіо, зокрема методи цифрової обробки сигналів та спектрального аналізу, відіграють ключову роль у вирішенні проблеми виявлення плагіату.

Плагіат у музичній сфері – це несанкціоноване використання або копіювання музичних елементів, таких як мелодія, гармонія, ритм, текст чи аранжування, що є унікальним продуктом творчої діяльності іншого автора або виконавця, з метою видати їх за власні. Це включає запозичення значних частин твору без дозволу та без належного зазначення авторства. Плагіат у музиці порушує авторське право та етичні норми й може призвести до правових наслідків, якщо доведено схожість та відсутність оригінальності в новому творі [2].

Розвиток цифрових технологій і доступність великої кількості аудіоконтенту через онлайн-платформи значно ускладнили захист авторських прав. Реєстрація забезпечує документальне підтвердження, яке свідчить про час і обсяг прав на твір. Однак, незважаючи на офіційну реєстрацію, власник авторських прав зобов’язаний самостійно стежити за випадками можливого плагіату або залучати для цього сторонні організації та служби. Це означає, що правовласник повинен активно моніторити використання свого твору в медіапросторі, оскільки державні органи чи інші інституції автоматично не відслідковують можливі порушення прав.

Також слід звернути увагу на проблеми, пов’язані з обробкою аудіоданих. Вплив навколишнього середовища, якість запису, зміна темпу чи накладання ефектів можуть значно ускладнити виявлення плагіату, що вимагає розробки надійних методів передобробки та фільтрації аудіозаписів. Важливу роль у

цьому відіграє нормалізація звуку та видалення шумів, що забезпечує якісніші результати аналізу.

Окрім цього, для визначення схожості аудіофайлів використовуються методи виявлення особливих точок, які ідентифікують ключові ознаки запису та порівнюють їх з іншими файлами. Особливі точки аудіофайлу можуть виступати як унікальні індикатори, що дозволяють провести аналіз без обов'язкового порівняння усіх спектральних характеристик, а лише ключових ознак, що значно пришвидшує процес.

Першим та найбільш відомим алгоритмом для ідентифікації схожих аудіозаписів став Shazam, який зробив революцію у сфері аудіоаналізу. Ця технологія була розроблена компанією Shazam Entertainment у 2000 році й дозволила впровадити інноваційний підхід до розпізнавання музики. Її основою став алгоритм, що базується на створенні аудіовідбитків для кожного треку в базі даних. Ці аудіовідбитки є компактними репрезентаціями часово-частотних характеристик аудіо та дозволяють швидко й точно ідентифікувати музику навіть за короткими й зашумленими зразками [3].

Спектральний аналіз включає розкладання звукового сигналу на частотні компоненти, що дозволяє отримати набір характеристик, зокрема амплітуду, фазу частотних складових та гармоніки. Методи спектрального аналізу, такі як перетворення Фур'є, короткочасне перетворення Фур'є (Short-Time Fourier Transform, STFT) або вейвлет-перетворення, забезпечують отримання спектрального відбитку, унікального для кожного запису. Таким чином, аналізуючи аудіозаписи як сигнали, можна спостерігати, як зміни в частотному діапазоні відображаються на спектрі, що дає змогу порівнювати записи та виявляти подібності (схожості) навіть, якщо вони зазнали змін у тривалості, гучності, тональності. Це важливо для аналізу записів, які можуть бути оброблені чи модифіковані, щоб уникнути виявлення плагіату.

Для аналізу схожості спектрів аудіозаписів використовуваться різні підходи. Наприклад, методи кореляції дозволяють порівнювати частотні компоненти різних сигналів, щоб виявити спільні або подібні елементи. Більш просуну-

ті підходи включають використання нейронних мереж та машинного навчання для класифікації спектральних ознак, які значно покращують точність та швидкість у порівнянні з класичними методами. Це робить спектральний аналіз потужним інструментом для виявлення плагіату навіть за умови застосування різних маніпуляцій із записом, таких як зміна темпу або накладання додаткових звуків.

Застосування методів глибокого навчання для аналізу спектрограм дозволяє обробляти великі обсяги даних з високою точністю, що важливо для автоматизованих систем перевірки на плагіат у великих базах аудіозаписів. Система може бути побудована на основі підходів класифікації за схожістю спектрограм та кластеризації, що дозволяє групувати записи за їх спектральними характеристиками. Це корисно для формування бібліотек схожих записів, які можуть бути віднесені до однієї музичної тематики чи жанру, або навіть для ідентифікації унікальної звукової сигнатури окремих виконавців.

Гарним прикладом є згорткові нейронні мережі (Convolutional Neural Networks, CNN), які спеціалізуються на обробці спектрограм – графічного представлення спектральних характеристик. Спектрограма перетворює аудіозапис у візуальне представлення звуку, де кожен колір чи інтенсивність вказує на рівень амплітуди певної частоти у конкретний момент часу. CNN здатні автоматично виявляти унікальні патерни та співвідносити їх із наявними шаблонами, що дозволяє ідентифікувати аудіозаписи навіть у випадках їх модифікації.

Основний сценарій для застосування CNN у задачах аналізу спектральних характеристик аудіо передбачає декілька етапів, які дозволяють автоматизувати процес виявлення плагіату, зіставляючи схожі елементи звукових записів.

Спочатку аудіосигнал перетворюється на спектрограму, хромограму або перетворення з постійним Q-фактором (Constant-Q Transform, CQT), що забезпечує двовимірне зображення, де відображається зміна частот у часі. CNN використовує це зображення як вхідний сигнал, оскільки двовимірні структури, такі як спектрограми, підходять для виявлення патернів і структури.

Для навчання CNN створюють великий набір спектрограм або хромограм

схожих і несхожих аудіозаписів. Прикладом може бути використання відкритих аудіобаз, на основі яких можна розділити дані на класи або маркувати зразки для навчання. Крім того, додають різноманітні варіації записів, такі як зміни гучності, темпу або частоти, що дозволяє мережі навчитися розпізнавати плагіат, навіть коли оригінальний запис було модифіковано.

Після підготовки даних CNN навчається знаходити патерни, властиві схожим аудіозаписам, використовуючи шари згортки та об’єднання, щоб виділяти ключові частотні компоненти, гармонії та ритмічні структури. Мережа автоматично визначає суттєві елементи, наприклад, певні гармонічні патерни, характерні для одного чи кількох аудіофрагментів, що може свідчити про їх схожість або навіть пряме копіювання.

CNN зводить спектральне представлення аудіо до вектора характеристик, що містить основні відомості про частотний склад та структуру. Вектори схожих аудіофайлів зазвичай близькі у векторному просторі, і це дає можливість порівнювати аудіо на основі їхніх характеристик. Сучасні системи використовують метрики на кшталт косинусної схожості або евклідової відстані, щоб визначити ступінь подібності між векторами двох аудіозаписів.

Останній крок передбачає порівняння вектора характеристик оброблюваного аудіозапису із записами з бази даних, щоб визначити, чи є вони схожими. При досягненні певного порогу схожості (який визначається залежно від задачі), аудіозапис може бути позначений як потенційно плагіатний.

CNN дозволяють досягти високої точності виявлення схожих записів, оскільки мережа здатна адаптуватися до різних варіацій аудіо, таких як зміна гучності, темпу або додавання шуму. Крім того, CNN можуть навчатись на великих наборах даних і використовувати складні структури для виявлення схожості, що є перевагою над класичними методами, які вимагають ручного виділення характеристик. Це робить CNN ідеальними для задач, де важливо автоматично та точно порівнювати великі обсяги аудіоінформації.

Незважаючи на високу ефективність, CNN мають деякі обмеження. Наприклад, для досягнення високої точності потрібні великі обсяги

тренувальних даних, а також суттєві обчислювальні ресурси. Крім того, навчання моделей на спектрограмах або хромограмах потребує спеціальних підходів для обробки фазової інформації, що є важливою складовою гармонійного аналізу. Сучасні дослідження в цьому напрямі спрямовані на розробку моделей, стійких до різноманітних змін сигналу, та вдосконалення процесу навчання CNN для точнішого виявлення схожих аудіозаписів.

1.2 Аналіз сценаріїв вирішення задачі аналізу спектральних характеристик аудіозаписів з метою пошуку схожих аудіозаписів для виявлення плагіату

На теперішній час тема спектрального аналізу сигналів, в тому числі аудіосигналів є актуально, що підтверджуються дослідженнями у даних сфері.

У [4] досліджено ефективність різних спектральних та ритмічних характеристик для класифікації аудіосигналів за допомогою глибоких CNN. Автори порівняли такі представлення сигналів, як мел-спектрограми, мел-частотні кепстральні коефіцієнти (Mel-Frequency Cepstral Coefficients, MFCC), циклічні темпограми, хромограми на основі STFT, хромограми на основі перетворення з константою (Constant-Q Transform, CQT) та хромограми з нормалізованими енергетичними статистиками (Chroma Energy Normalized Statistics, CENS). Результати експериментів, проведених на наборі даних ESC-50, показали, що мел-спектрограми та MFCC забезпечують значно кращу продуктивність у задачах класифікації аудіо за допомогою CNN порівняно з іншими дослідженими характеристиками.

У [5] досліджується природа фазових розподілів коефіцієнтів STFT аудіосигналів. Автори виявили, що, всупереч поширеному припущенню про рівномірність фазових розподілів, при аналізі фаз за окремими частотами або діапазонами амплітуд ці розподіли можуть бути значно нерівномірними. Це свідчить про те, що загальне припущення про рівномірність фазових розподілів може

приховувати важливі деталі. У статті пояснюється значення цих нерівномірних фазових розподілів, їх походження та вплив форми вікна STFT на нерівномірність фазових розподілів.

У [6] автори досліджують проблему відновлення фази STFT для задач розділення аудіоджерел. Традиційно в таких задачах оцінюють лише амплітуду STFT кожного джерела, а для синтезу сигналу у часовій області використовують фазу вихідної суміші, застосовуючи підхід, подібний до фільтрації Вінера. Однак цей метод може призводити до залишкових перешкод та артефактів у розділених сигналах, особливо коли джерела перекриваються у часово-частотній області. Автори пропонують альтернативний підхід, заснований на моделюванні фази STFT. Вони зазначають, що багато музичних подій складаються з повільно змінюваних синусоїд, для яких приріст фази STFT з часом має специфічну форму. Це дозволяє відновлювати фазу за допомогою техніки розгортання фази після отримання короткочасної оцінки частоти. На основі цих спостережень автори розробили нову ітеративну процедуру розділення джерел, яка мінімізує помилку змішування за допомогою методу допоміжної функції. Ця процедура ініціалізується, використовуючи техніку розгортання фази, що забезпечує оцінки з властивістю часової безперервності. Експерименти, проведені на реалістичних музичних композиціях, показали, що за умови точних оцінок амплітуди запропонований метод перевершує сучасний узгоджений фільтр Вінера, забезпечуючи кращу якість розділених сигналів та зменшуючи залишкові перешкоди. Ця робота підкреслює важливість врахування фазової інформації та використання моделей сигналу для покращення результатів у задачах розділення аудіоджерел.

У [7] автори пропонують новий метод виділення аудіовідбитків, що поєднує сегментацію сигналу з новою схемою побудови аудіовідбитків. Запропонований підхід демонструє стійкість до стиснення та часових зсувів аудіофайлів. Аудіовідбиток – це компактний підпис аудіофайлу, обчислений на основі його основних перцептивних властивостей, який дозволяє ідентифікувати аудіофайл серед набору кандидатів, не розкриваючи інших характеристик файлів. Засто-

сування аудіовідбитків включає моніторинг аудіо на мовних каналах, фільтрацію в peer-to-peer мережах, відновлення метаданих у великих аудіобібліотеках та захист авторських прав у системах управління цифровими правами (Digital Rights Management, DRM). Запропонований алгоритм поєднує метод сегментації з новою схемою побудови фінгерпринтів, забезпечуючи стійкість до стиснення та часових зсувів аудіофайлів.

У [8] представлено дослідження покращення системи виявлення копій аудіо на основі контенту (Content-Based Copy Detection, CBCD). Описують новий метод побудови аудіо відбитків (фінгерпринтів), які формуються з використанням спектрограми. Суть підходу полягає у використанні двох типів параметрів – глобальних та локальних середніх значень спектральних інтенсивностей.

Глобальний підхід базується на обчисленні середнього інтенсивності для всього сигналу, а локальний працює з розділенням спектрограми на дрібні блоки, для яких також обчислюється середнє. Важливим кроком є перетворення спектрограм у двовимірні бінарні зображення та їх подальше представлення у вигляді векторів для спрощення пошуку найближчих сусідів.

Ключовим досягненням є здатність системи ефективно обробляти різноманітні трансформації аудіо, включаючи MP3-компресію, фільтрацію частот, додавання шуму тощо. Автори також аналізують вплив параметра кількості сусідніх кадрів (Successive Closest Frames, SCF), що допомагає зменшити помилки.

Основний акцент дослідження – це підвищення стійкості аудіо-відбитків до спотворень та їх оптимальне комбінування для забезпечення надійного виявлення копій у великих базах даних.

У [9] досліджуються методи виявлення та локалізації часткових збігів у аудіо у різних сценаріях застосування, таких як управління архівами, аналіз трансляцій, пошук медіа і медіа-форенсика. Запропоновано новий алгоритм часткового збігу, який забезпечує високу точність навіть у складних випадках, коли стандартні підходи, засновані на фингерпринтингу, не працюють. Ця методика передбачає два варіанти: перший спрямований на обробку великих на-

борів даних із низькою часовою деталізацією, другий – на детальний аналіз малих наборів даних або окремих об'єктів.

У [9] розглядаються різноманітні способи модифікації аудіо, такі як вирізання, вставлення чи заміна частин треку, а також їхні наслідки для автоматичного виявлення дублікатів. Також розглядаються можливості використання цього алгоритму в аналізі структури програм, моніторингу повторного використання матеріалів і виявленні контенту в різних джерелах.

У [10] представлено новий підхід до створення аудіофінгерпринтів, який демонструє високу стійкість до маніпуляцій аудіоданими, таких як зміни тону, темпу, швидкості, а також до додавання шуму і фільтрації. Запропонований метод використовує фундаментальну частоту для екстракції ознак і створення карти фундаментальної частоти (Fundamental Frequency Map, FFMAP), що є ключовим компонентом у процесі аналізу схожості між аудіо.

Головним недоліком вище згаданих досліджень є зосередження на оптимізації хеш-функції для забезпечення максимально точного пошуку конкретного аудіозапису у великій базі даних. Такий підхід дозволяє ефективно ідентифікувати точні копії або практично незмінені версії записів, але значно обмежує можливості виявлення схожих аудіофрагментів, які можуть відрізнятися за тембром, аранжуванням чи іншими характеристиками. Це створює суттєві обмеження для застосування таких алгоритмів у завданнях, які вимагають аналізу подібностей між аудіозаписами, наприклад, виявлення плагіату або рекомендаційних систем, де важливо враховувати не лише точні відповідники, але й контекстуальну схожість.

1.3 Змістовна та формальна постановка задачі

Основна ідея задачі виявлення плагіату в аудіозаписах полягає в тому, щоб виокремити з аудіосигналів ключові частотно-часові ознаки та визначити, наскільки два (або більше) записи схожі за цими ознаками. Якщо ступінь схо-

жості між будь-якими двома фрагментами різних аудіозаписів перевищує наперед заданий поріг, такі фрагменти позначаються як потенційно плагіатні.

Таким чином, змістовно задача зводиться до:

– збирання набору аудіозаписів у цифровому форматі;

– перетворення кожного фрагмента у вектор ознак, що відображає спектральні характеристики;

– встановлення функції порівняння (міри схожості або відстані) між отриманими векторами;

– визначення порога, перевищення якого свідчить про потенційний плагіат.

З метою здійснення системного виявлення можливого плагіату у множині аудіозаписів доцільно формулювати задачу в термінах формального математичного опису. Нижче наведено основні етапи такої постановки, що охоплюють сегментування вихідних аудіосигналів, побудову векторів спектральних ознак і визначення функції порівняння фрагментів.

Нехай задано множину аудіозаписів (1.1)

$$A = \{a_1, ..., a_2, a_N\}, \tag{1.1}$$

де A – множина аудіозаписів a_i;

a_i – аудіозапис – дискретизований сигнал;

N – кількість аудіозаписів.

Кожен аудіозапис a_i розбивається на фрагменти фіксованої тривалості (1.2)

$$P_i = \{p_{i,1}, p_{i.2}, ..., p_{iK_i}\}, \tag{1.2}$$

де P_i – множина сегментів i-го аудіозапису;

$p_{i,j}$ – j-й сегмент i-го аудіозапису;

K_i – кількість сегментів і-го аудіозаписів.

Для кожного сегмента обчислюються спектральні характеристики (1.3)

$$x_{i,j} = \Phi(p_{i,j}) \in R^d, \tag{1.3}$$

де $x_{i,j}$ – вектор спектральних ознак, що описує характеристики j-й сегменту і-го аудіозапису;

Φ – функція отримання спектральних ознак;

$p_{i,j}$ – j-й сегмент і-го аудіозапису;

R^d – d-вимірний простір дійсних чисел, у якому знаходиться вектор $x_{i,j}$;

d – кількість компонент вектора спектральних ознак.

Тоді функція перевірки на плагіат матиме вигляд (1.4)

$$F(x_{i,j}, x_{l,k}) = \begin{Bmatrix} 0, H(x_{i,j}, x_{l,k}) \geq \delta, \\ 1, H(x_{i,j}, x_{l,k}) < \delta \end{Bmatrix}, \tag{1.4}$$

де $F(x_{i,j}, x_{l,k})$ – функція перевірки на плагіат j-го сегменту i-го аудіозапису та k-го сегменту l-го аудіозапису;

$H(x_{i,j}, x_{l,k})$ – функція порівняння спектральних характеристик плагіат j-го сегменту і-го аудіозапису та k-й сегменту l-го аудіозапису;

δ – порогове значення схожості, яке свідчить про наявність плагіату.

1.4 Постановка задач дослідження

Метою магістерської кваліфікаційної роботи є дослідження моделей і методів вирішення задачі аналізу спектральних характеристик аудіозаписів для виявлення плагіату з метою підвищення ефективності.

Для досягнення поставленої мети у магістерській кваліфікаційній роботі

пропонується вирішити такі задачі дослідження:

– провести системний аналіз предметної області;

– дослідити існуючі методи та методи для вирішення задачі аналізу спектральних характеристик аудіозаписів з метою виявлення плагіату;

– описати математичну модель задачі аналізу спектральних характеристик аудіозаписів з метою виявлення плагіату (формули 1.1–1.4);

– вибрати та модифікувати моделі та методи вирішення задачі аналізу спектральних характеристик аудіозаписів з метою виявлення плагіату;

– розробити алгоритм розв’язання задачі аналізу спектральних характеристик аудіозаписів з метою виявлення плагіату;

– розробити прототип програмного забезпечення для розв’язання задачі аналізу спектральних характеристик аудіозаписів з метою виявлення плагіату.

2 ВИБІР ТА ОБҐРУНТУВАННЯ МЕТОДУ РОЗВ'ЯЗАННЯ

2.1 Класичні методи аналізу спектральних характеристик аудіозаписів

Аналіз спектральних характеристик аудіозаписів є ключовим підходом у задачах пошуку схожих аудіофрагментів для виявлення плагіату, оскільки дозволяє визначити частотні і часові особливості сигналу, що є унікальними для кожного запису. Існують різні методи для спектрального аналізу, і всі їх можна розділити на традиційні алгоритми обробки сигналів та методи, що базуються на машинному і глибокому навчанні. Традиційні методи, як STFT і вейвлет-перетворення, забезпечують детальну інформацію про часово-частотну структуру сигналу. З їхньою допомогою можна отримати спектрограму, яка показує зміни частотного складу сигналу в часі і дає змогу визначати повторювані або схожі патерни у музичних чи мовних фрагментах. Додатково, MFCC дозволяють ще компактніше й ефективніше відображати основні властивості спектра, оскільки базуються на мел-шкалі, що узгоджується з особливостями людського слуху. Такий підхід може спростити подальше порівняння звуків і підвищити стійкість до незначних шумів або відмінностей у гучності.

STFT є основою для аналізу нестаціонарних аудіосигналів, оскільки він надає можливість досліджувати частотні компоненти сигналу у різні моменти часу. У випадку виявлення плагіату часово-частотний підхід на основі STFT дозволяє побудувати спектрограму, яка візуалізує енергетичний розподіл частот, допомагаючи порівнювати аудіофрагменти на основі їхньої подібності в частотному спектрі.

Процес STFT складається з розбиття сигналу на короткі фрагменти (вікна), кожен з яких обробляється окремо. Для кожного фрагмента застосовується перетворення Фур'є, що надає інформацію про частотний склад цього сегмента. Такий підхід генерує спектрограму, де горизонтальна вісь представляє час, вертикальна – частоту, а інтенсивність кольору або яскравість – амплітуду частотного компоненту. Це представлення дозволяє візуально оцінювати і знаходити

схожі частотні патерни між записами.

Формальне визначення STFT приведено в формулі

$$X(t,\omega) = \int_{-\infty}^{\infty} x(\tau) \cdot w(\tau - t) \cdot e^{-i\omega\tau} d\tau,$$

де $X(t,\omega)$ – результат STFT, що є функцією часу t і кутової частоти ω;

$x(\tau)$ – оригінальний сигнал у часовому вимірі;

$w(\tau - t)$ – віконна функція, яка виділяє сегмент сигналу навколо моменту t;

$e^{-i\omega\tau}$ – комплексна експоненціальна функція для перетворення в частотну область, де i – уявна одиниця, а ω – кутова частота.

Дискретне представлення STFT приведено в формулі

$$X[m,k] = \sum_{-\infty}^{\infty} x[n] \cdot w[n - mR] \cdot e^{-i\frac{2\pi}{N}kn},$$

де $X[m,k]$ – значення STFT для m-го часового вікна та k-ї частотної компоненти;

$x[n]$ – дискретний сигнал;

$w[n - mR]$ – віконна функція, зміщена на m-те вікно з кроком R;

N – кількість точок для дискретного перетворення Фур'є, що визначає частотну роздільну здатність.

Для більш складних і детальних досліджень, особливо коли сигнал має різкі зміни у часі, використовується вейвлет-перетворення. Воно забезпечує кращу часову та частотну роздільну здатність, що особливо корисно для складних звукових композицій. На відміну від STFT, вейвлет-перетворення використовує вікна змінної ширини, що дозволяє точніше розрізняти частотні компоненти, які змінюються з часом, зберігаючи при цьому високу часову точність [11].

Вейвлет-перетворення полягає в розкладі сигналу на «вейвлети» –

короткочасні хвилеподібні функції, які локалізовані в часі й частоті. Цей метод використовує функцію-основу, відому як «материнський вейвлет», яка масштабується та зміщується для побудови часово-частотного представлення сигналу. Завдяки цьому вейвлет-перетворення дозволяє отримати більше інформації про локальні особливості сигналу, що є важливим для аналізу аудіозаписів, особливо для нестаціонарних і музичних сигналів [12].

Вейвлет-перетворення надає детальне представлення часово-частотної структури сигналу, що особливо корисно для музичних сигналів, де важливими є локальні особливості – такі як швидкі зміни в частотному вмісті. Для порівняння двох аудіозаписів на предмет плагіату можна обчислити вейвлет-коефіцієнти для кожного запису та порівняти їх за певними метриками (наприклад, кореляцією). Якщо коефіцієнти мають значний збіг у часово-частотних областях, це може вказувати на схожість між записами.

Вейвлет-перетворення забезпечує високу часову роздільну здатність для високочастотних компонентів і високу частотну роздільну здатність для низькочастотних компонентів. Це робить його дуже ефективним для аналізу звуків, які мають швидкі зміни або складну структуру. Однак обробка великих обсягів даних за допомогою вейвлет-перетворення може бути ресурсоємною, і його ефективність залежить від вибору відповідного материнського вейвлету.

Формула безперервного вейвлет-перетворення (Continuous Wavelet Transform, CWT) приведена в формулі

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

де $T(a,b)$ – вейвлет-перетворення сигналу $x(t)$ з масштабом a і зсувом b;

$x(t)$ – оригінальний сигнал у часовому вимірі;

$\psi^*\left(\frac{t-b}{a}\right)$ – комплексне спряження масштабованого та зміщеного материнського вейвлету;

a – параметр масштабування, який впливає на частотну роздільну здатність (малі значення a відповідають високим частотам);

b – параметр зсуву у часі.

MFCC широко використовуються в задачах обробки аудіосигналів. Основна ідея методу полягає у відображенні частотної інформації в простір ознак, що узгоджується з особливостями людського слуху та дає змогу одержати компактне й чутливе представлення аудіосигналу. У низькочастотному діапазоні (до ~700 Гц) зміна частоти сприймається лінійно: люди відчувають кожне збільшення на 100 Гц як однакове. У високочастотному діапазоні (>700 Гц) чутливість знижується, і однакова зміна частоти в герцах сприймається менш інтенсивно. Мел-шкала компресує високі частоти, зменшуючи їхню «вагу» порівняно з низькими.

MFCC отримують послідовно, починаючи з розбиття аудіосигналу на короткі вікна. Після отримання спектру сигналу для кожного сегмента за допомогою STFT, частотна вісь k переводиться в мел-шкалу. Для цього використовуються трикутні мел-фільтри $H_r[k]$. Енергія для кожного фільтра r обчислюється як

$$E[m,r] = \sum_{k=0}^{N-1} \left|X[m,k]\right|^2 \cdot H_r[k],$$

де $E[m,r]$ – енергія r-го мел-фільтра у m-му сегменті;

$H_r[k]$ – трикутний фільтр, що концентрує енергію на певній частотній ділянці;

r – індекс фільтра.

Фільтри $H_r[k]$ визначаються на основі нелінійного перетворення частотної осі f_k у мел-шкалу за формулою

$$M(f) = 2595 \cdot \log_{10}(1 + \frac{f}{700}),$$

де $M(f)$ – частота в мел-шкалі, яка відповідає частоті f в герцах, узгоджена з психоакустичним сприйняттям частот людським слухом;

f – частота в герцах (Гц), яка є фізичною характеристикою звукового сигналу.

Для кожного мел-фільтра логарифмується обчислена енергія

$$M[m,r] = \ln(E[m,r] + \varepsilon), \tag{1.2}$$

де $M[m,r]$ – логарифмована енергія для r-го фільтра в сегменті m;

$E[m,r]$ – енергія r-го мел-фільтра у m-му сегменті;

ε – мала константа, що запобігає обчисленню $\ln(0)$.

Логарифмовані енергії мел-фільтрів піддаються дискретному косинус-перетворенню (Discrete Cosine Transform, DCT), щоб отримати MFCC. Формула для n-го коефіцієнта

$$c_n[m] = \sum_{r=1}^{R} M[m,r] \cdot \cos(\pi n(r - \frac{1}{2}) / R),$$

де $c_n[m]$ – n-й мел-капстральний коефіцієнт для сегмента m;

R – кількість мел-фільтрів;

n – індекс кофецієнта.

Таким чином, використання MFCC дозволяє значно зменшити розмірність ознак у порівнянні зі спектрограмами та іншими спектральними поданнями. Це досягається завдяки кільком ключовим етапам обробки, серед яких виділення енергії через мел-фільтри, логарифмічне перетворення та застосування DCT. Цей підхід концентрує основну інформацію сигналу у кількох коефіцієн-

тах, відкидаючи менш значущі компоненти, що можуть бути надлишковими або шумовими. У результаті зменшується кількість даних, які потрібно аналізувати, зберігаючи при цьому основні характеристики сигналу, пов'язані з тембром, частотним складом і ритмічною структурою. Така компактність робить MFCC особливо ефективними для задач порівняння, класифікації або розпізнавання аудіозаписів.

2.2 Сучасні методи для аналізу спектральних характеристик аудіозаписів

Методи глибокого навчання, зокрема CNN, відкривають нові можливості для задач пошуку схожих аудіозаписів завдяки здатності виявляти складні патерни без необхідності вручного виділення характеристик. Використання CNN дозволяє розглядати аудіофрагменти у вигляді спектрограм як зображення, і таким чином мережа може виявляти схожість між сигналами, навіть якщо вони були змінені через стиснення, фільтрацію чи інші процеси обробки звуку.

CNN є одним із ключових підходів у сучасному глибинному навчанні для розв’язання завдань комп’ютерного зору, розпізнавання мови та багатьох інших. Основу CNN становить послідовне використання шарів згортки, нелінійних активацій, пулінгу (підвибірки) та, зазвичай, одного або декількох повнозв’язних шарів у кінці. Така структура дає змогу модельованому об’єктові «навчитися» багаторівневих ознак (features) без потреби ручного конструювання ознак, що робить згорткові мережі особливо ефективними в різноманітних задачах.

Загальний вхід у CNN можна розглядати як тривимірний тензор X розмірністю $H_{in} \times W_{in} \times D_{in}$, де H_{in} та W_{in} відповідають висоті та ширині даних, а D_{in} – кількості ознак. На вхід мережі цей тензор надходить у згортковий шар, після якого результати обробляються шарами активації, пулінгу, й нарешті – одним або кількома повнозв’язними шарами для отримання кінцевого результату (на-

приклад, класу об’єкта).

Ключова ідея згорткових шарів полягає у тому, що замість повнозв’язного з’єднання кожного вхідного пікселя з усіма нейронами наступного шару, використовуються фільтри (ядра згортки), які мають обмежене рецептивне поле (висоту K_h та ширину K_w) і обробляють лише невеликі фрагменти вхідних даних. Такий підхід дає змогу зберігати просторову структуру вхідного зображення та суттєво зменшує кількість параметрів порівняно з традиційними штучними нейронними мережами [13, 14].

Загальна схема CNN приведена на рисунку 2.1.

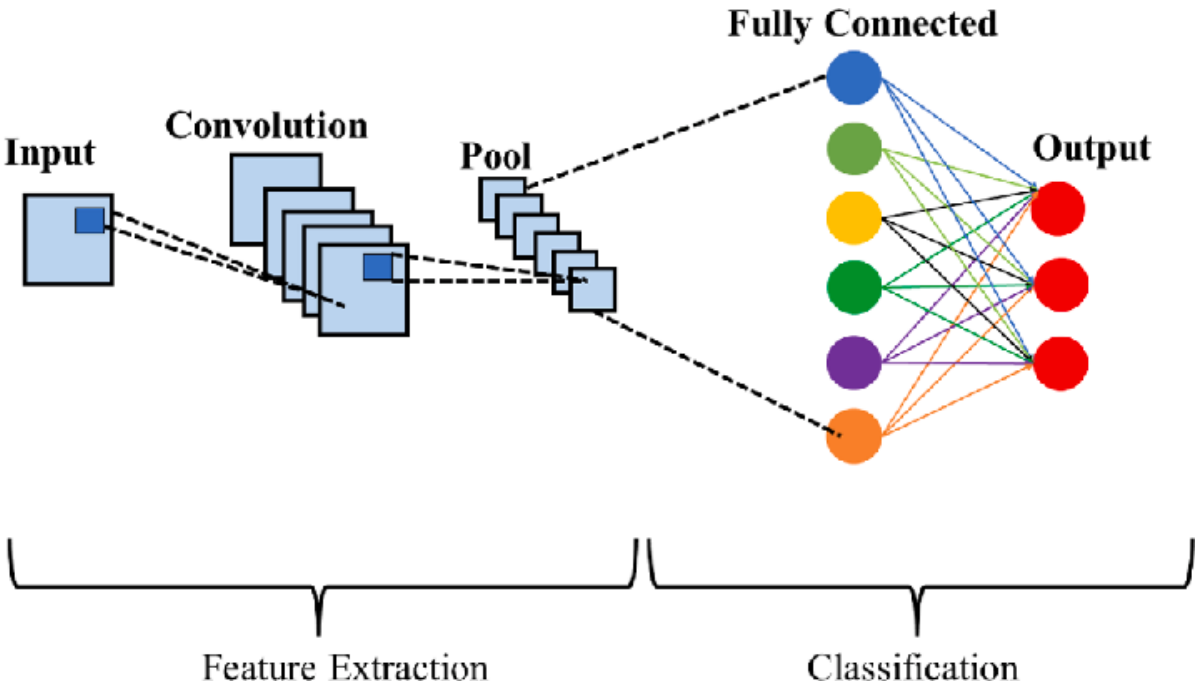

Рисунок 2.1 – Загальна схема CNN

Згортковий шар має набір фільтрів (ядра згортки) $W^{(k)}$, ожне з яких має розмір $K_h \times K_w \times D_{in}$. Якщо кількість фільтрів у шарі дорівнює D_{out}, то кожен фільтр генерує окремий вихідний канал, утворюючи тензор вихідних ознак розміру $H_{out} \times W_{out} \times D_{out}$. Додатково до кожного фільтра додається зміщення (bias) b_k.

У загальному випадку у згортці можуть застосовуватися такі гіперпараметри:

– zero-padding – заповнення вхідного тензора нулями розміром p елементів з кожного боку (у кожному просторому вимірі) для контролю зміни просторових розмірів у вихідному тензорі;

– stride (крок) — крок s, на скільки позицій «зсувається» ядро згортки під час обчислень у кожному просторому вимірі.

Розміри вихідного тензора обчислюється за формулами

$$H_{out} = \left\lfloor \frac{H_{in} - K_h + 2p}{s} \right\rfloor + 1,$$

$$W_{out} = \left\lfloor \frac{W_{in} - K_w + 2p}{s} \right\rfloor + 1,$$

де H_{out} – це висота (кількість рядків) вихідного тензора після згортки;

H_{in} – висота вхідного тензора (наприклад, вихід попереднього шару або початковий тензор);

K_h – висота ядра згортки (kernel);

p – кількість нульових елементів, що додаються до тензора з кожного краю;

s – крок згортки – на скільки елементів зсувається ядро згортки при послідовному обчисленні;

W_{out} – це ширина (кількість стовпців) вихідного тензора після згортки;

W_{in} – ширина вхідного тензора (наприклад, вихід попереднього шару або початковий тензор);

K_w – ширина ядра згортки (kernel).

Сам процес згортки можна описати рівнянням

$$Y(x,y,k) = \sum_{d=1}^{D_{in}} \sum_{u=0}^{K_h-1} \sum_{v=0}^{K_w-1} W(u,v,d,k) \cdot X(s \cdot x + u - p, s \cdot y + v - p, d) + b_k \,,$$

де $Y(x,y,k)$ – елемент вихідного тензора у позиції (x,y) для ознаки k;

D_{in} – кількість ознак;

K_h – висота ядра згортки (kernel);

K_w – ширина ядра згортки (kernel);

$W(u,v,d,k)$ – вага k-го фільтра (ядра згортки) у позиції (u,v) для ознаки d;

$X(h,w,d)$ – елемент вхідного тензора у позиції (h,w) для ознаки d;

b_k – зміщення для k-го фільтра;

p – кількість нульових елементів, що додаються до тензора з кожного краю;

s – крок згортки – на скільки елементів зсувається ядро згортки при послідовному обчисленні;

Таким чином, кожен фільтр «ковзає» по вхідному тензору й обчислює лінійну комбінацію локальних значень, що дає змогу моделі виявляти певні просторові патерни.

Після лінійної операції (зокрема, згортки) зазвичай вводиться нелінійність у вигляді функції активації. Це необхідно для того, щоб нейронна мережа могла апроксимувати складні функції, а не лише лінійні відображення.

Активаційний шар застосовує обрану функцію активації f до вихідного тензора Z, який є результатом згорткової операції або іншого лінійного перетворення. Активація виконується покомпонентно для кожного елемента Z. Обчислення виконується за формулою

$$A(x,y,k) = f(Z(x,y,k)),$$

де $A(x,y,k)$ – значення вихідного тензора після активації в позиції (x,y) для ознаки k;

f – обрана функція активації;

$Z(x,y,k)$ – значення тензора в позиції (x,y) для ознаки k до застосування активації.

ReLU (Rectified Linear Unit) – це одна з найпопулярніших активаційних

функцій у нейронних мережах, яка використовується для внесення нелінійності у модель. Вона проста в реалізації, має низькі обчислювальні витрати та забезпечує високу ефективність у задачах глибокого навчання. Головною особливістю ReLU є те, що вона пропускає позитивні значення без змін, тоді як від'ємні значення переводяться в нуль. Це дозволяє моделі зосереджуватися на корисних ознаках та ігнорувати негативні впливи.

Завдяки своїй поведінці ReLU допомагає вирішувати проблему затухання градієнта, яка характерна для інших активаційних функцій, таких як сигмоїдна чи гіперболічний тангенс [15].

Функція ReLU має вигляд

$$f(x) = \max(0, x),$$

де $f(x)$ –функція ReLU від x;

x – вхідне значення (сигнал, який подається на нейрон).

Для задачі аналізу спектральних характеристик аудіозаписів з метою виявлення плагіату найкраще підходить сигмоїдна функція активації, оскільки вона дозволяє моделі оцінювати наявність подібності між окремими спектральними ознаками у вигляді значень у діапазоні $(0,1)$. Це особливо важливо, оскільки результат можна інтерпретувати як імовірність збігу між спектральними характеристиками двох аудіозаписів, що спрощує подальший аналіз та прийняття рішення про плагіат.

Сигмоїдна функція має вигляд

$$\sigma(x) = \frac{1}{1 + e^{-x}},$$

де $\sigma(x)$ – сигмоїдна функція від x;

x – лінійне поєднання вхідних даних і ваг.

Пулінг виконує операцію зменшення просторових вимірів, агрегації ло-

кальних значень, зменшуючи розмір тензора. Це допомагає скоротити кількість параметрів, знизити ризик перенавчання та зробити модель стійкішою до варіацій у даних. Пулінг виконується через вікно розміру $P_h \times P_w$, яке «ковзає» по тензору з кроком s_p .

У спектральному аналізі важливо не лише визначати пікові значення (як у Max Pooling), але й оцінювати загальний характер амплітудного розподілу у певному діапазоні частот. Average Pooling дозволяє ефективно усереднити локальні варіації, що допомагає моделі зберігати інформацію про загальні тенденції у спектрі.

Також, Average Pooling робить модель менш чутливою до незначних коливань у спектрі, таких як варіації у рівні гучності чи незначні відхилення частот. Це допомагає виявляти подібність навіть між записами, які можуть мати невеликі відмінності через технічні або природні фактори.

Average Pooling згладжує випадкові пікові значення, які можуть бути спричинені шумом у даних. Це особливо важливо для задачі аналізу аудіозаписів, де шум може спотворювати локальні ознаки, але не впливатиме на середнє значення.

Якщо задача вимагає порівнювати два аудіозаписи для виявлення плагіату, Average Pooling зменшує просторову розмірність спектрального представлення, зберігаючи основні особливості кожної локальної області спектру. Завдяки цьому модель буде здатна ідентифікувати схожість між записами навіть у разі незначних відмінностей у частотному розподілі [16].

Таким чином, Average Pooling у задачі аналізу спектральних характеристик аудіозаписів допоможе забезпечити стійкість до шуму, узагальнення спектральних характеристик та виділення глобальних ознак, необхідних для порівняння записів.

Average Pooling обчислюється за формулою

$$Y(x,y,d) = \frac{1}{P_h \times P_w} \sum_{u=0}^{P_h-1} \sum_{v=0}^{P_w-1} X(s_p \cdot x + u, s_p \times y + v, d),$$

де $Y(x, y, d)$ – значення вихідного тензора після застосування пулінгу середнього значення в позиції тензора (x, y) для ознаки d;

P_h – висота вікна пулінгу;

P_w – ширина вікна пулінгу;

$X(h, w, d)$ – елемент вхідного тензора у позиції (h, w) для ознаки d;

s_p – крок (stride), який визначає, на скільки позицій зсувається вікно під час обчислень.

2.3 Адаптація класичних та сучасних методів аналізу спектральних характеристик аудіосигналів для вирішення задачі виявлення плагіату

Кожен із підходів має свої переваги та обмеження, і їх вибір залежить від конкретної задачі, обсягу даних та необхідної точності. Традиційні методи, такі як STFT і вейвлет-перетворення, забезпечують високу роздільну здатність і дозволяють точніше відстежувати зміни частотного складу з часом, проте мають обмеження щодо автоматичного виявлення складних аудіоподібностей. Методи з використанням CNN, навпаки, дозволяють автоматизувати процес порівняння великих обсягів даних, проте потребує значних обчислювальних ресурсів та великих тренувальних наборів для досягнення високої точності.

Ключовим аспектом задачі пошуку схожих аудіозаписів є ефективність спектрального представлення сигналу. Якість цього представлення впливає на точність і швидкість алгоритмів виявлення.

Аудіозаписи можуть бути зроблені в різних умовах, що впливають на їх якість, наприклад, наявність фонових шумів, зміни акустичного середовища, температура і вологість. Всі ці фактори можуть створювати перешкоди або вносити додаткові гармоніки у запис, що ускладнює порівняння записів без належної обробки. Зазвичай аудіо проходить обробку, яка включає стиснення

(зокрема, з втратами), фільтрацію, нормалізацію або додавання спецефектів. Це може впливати на спектральний склад сигналу і створювати додаткові складнощі при порівнянні, оскільки спотворює оригінальні частотні компоненти. Важливу роль відіграє структура самого сигналу, зокрема ритмічні та мелодійні патерни. Зміни в темпі або тональності можуть суттєво змінювати видимий спектральний склад і ускладнювати виявлення подібностей між записами.

Задача аналізу спектральних характеристик аудіозаписів полягає у визначенні ступеня схожості між аудіофрагментами для виявлення можливих випадків плагіату. Такий підхід передбачає кількісну оцінку подібності спектрального складу записів, що дозволяє точно визначати ступінь схожості навіть при варіаціях у темпі, тональності або форматі.

Оскільки MFCC можна інтерпретувати як компактні двовимірні «зображення», CNN є ефективним інструментом для їх класифікації. Зокрема, великі архітектури CNN, такі як Visual Geometry Group (VGG), наприклад, VGG16 або VGG19, що демонструють високу точність розпізнавання завдяки своїй глибокій багаторівневій структурі. Водночас, ці моделі характеризуються значними обчислювальними вимогами. Альтернативою є ResNet, яка, завдяки використанню механізму резидуальних зв’язків («residual connections»), дозволяє тренувати дуже глибокі мережі, долаючи проблему згасання градієнта. Це робить її доцільною для вирішення складних задач, таких як аналіз довших аудіофрагментів або багатокласова класифікація. У випадках, коли обчислювальні ресурси обмежені або потрібна швидка обробка даних, перевагу можуть мати компактні моделі, такі як MobileNet.

Для порівняння двох аудіофрагментів найбільш відповідною є архітектура типу Siamese Network, яка включає пару ідентичних згорткових нейронних мереж із загальними вагами.

Siamese Network – це архітектура нейронної мережі, яка спеціалізується на задачах порівняння двох вхідних об'єктів, зокрема мел-спектрограм. Основна ідея полягає у використанні двох однакових (ідентичних за параметрами)

гілок CNN, що мають спільні ваги. Ці гілки незалежно обробляють два вхідних об'єкти та створюють їх векторні уявлення у просторі ознак, які потім порівнюються для визначення схожості [17]. Загальна схема Siamese Network приведена на рисунку 2.2.

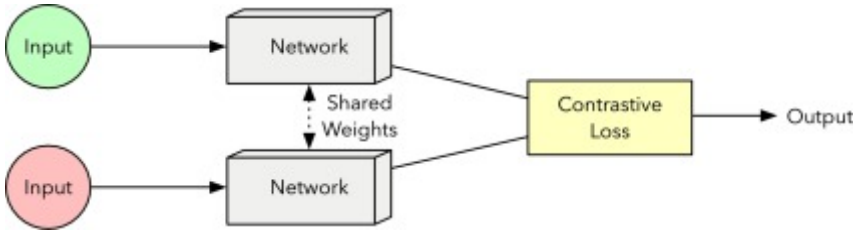

Рисунок 2.2 – Загальна схема Siamese Network

Такий підхід може бути особливо ефективний у задачах виявлення плагіату, оскільки результати мережі відображають не лише категоріальні класи (наприклад, «мова/не мова» або «спів/не спів»), а й безперервну міру схожості. Це значення може бути порівняно із заздалегідь визначеним порогом для прийняття рішення щодо наявності плагіату.

Одним із головних недоліків архітектури Siamese Network є те, що для визначення ступеня схожості необхідно двічі пропустити аудіозаписи через одну і ту ж мережу CNN: один запис із пари обчислюється в реальному часі, інший – кожного разу заново для всіх зразків бази даних. Це стає обчислювально затратним, особливо коли потрібно порівняти новий запис із великою кількістю вже існуючих зразків.

Щоб уникнути таких витрат, архітектуру можна модифікувати. Для цього всі вихідні вектори ознак для записів із бази даних попередньо обчислюються за допомогою тієї ж CNN і зберігаються у базі даних. При цьому збережені дані обраховуються з використанням фіксованих вагових коефіцієнтів, які були отримані під час навчання моделі.

Коли з’являється новий запис, його спектральні характеристики проходять через ту ж CNN для отримання вектору ознак. Потім цей вектор порівнюється з усіма попередньо збереженими векторами ознак у базі даних, використ-

товуючи функцію відстані чи подібності. Таким чином, обчислення в реальному часі обмежуються лише одним проходженням CNN для нового запису та порівнянням із попередньо обчисленими векторами.

Цей підхід значно знижує обчислювальні витрати в реальному часі та дозволяє масштабувати систему навіть для великих баз даних, не втрачаючи точності моделі.

Висновки за розділом 2

Другий розділ присвячений вибору та обґрунтуванню методів вирішення задачі аналізу спектральних характеристик аудіозаписів для виявлення плагіату. Розглянуто метод STFT, який дозволить перейти від часової області до частотно-часового представлення сигналів. Завдяки цьому методу можливо отримати спектрограми аудіофайлів, які є базовими даними для подальшого аналізу. STFT забезпечує високу точність відображення змін у спектральних характеристиках у часі, що особливо важливо для ідентифікації схожості між аудіозаписами.

Розглянуто також метод вейвлет-перетворення для аналізу часово-частотних ознак. Цей підхід дозволяє локалізувати особливості сигналу, але через високу обчислювальну складність та складність порівняння ознак вейвлет-перетворення не було обрано для вирішення поставленої задачі.

Для подальшої оптимізації обробки даних розглянуто метод отримання MFCC, який дозволить зменшити розмірність даних і зберегти ключові спектральні ознаки. MFCC враховують особливості людського сприйняття звуку, що підвищує точність, релевантність аналізу та зменшення обчислювальних витрат. Вибір цього методу забезпечить ефективну підготовку даних для вхідного шару нейронної мережі.

Запропоновано адаптовану архітектуру CNN, яка дозволить аналізувати спектральні характеристики MFCC. Адаптація включає налаштування згортко-

вих і пулінгових шарів для роботи зі спектрально-часовими даними. Це дозволить виявляти локальні та глобальні особливості сигналу, необхідні для оцінки схожості між аудіофрагментами.

Для порівняння спектральних характеристик запропоновано модифікацію Siamese Network із використанням оптимізаційних методів. Модифікація включає спільні ваги для вилучення ознак з MFCC, а також оптимізацію обчислень за рахунок ефективного налаштування функції втрат і регуляризації. Це дозволить скоротити обчислювальні витрати та підвищити точність аналізу.

У кваліфікаційній роботі для розв’язання задачі аналізу спектральних характеристик аудіозаписів обрано підхід на основі модифікованої Siamese Network із використанням CNN. Ця архітектура дозволить ефективно порівнювати мел-спектрограми, вилучаючи ключові ознаки з обох входів за допомогою спільних ваг нейронної мережі.

3 ПРОГРАМНА РЕАЛІЗАЦІЯ

3.1 Обгрунтування використання Python для вирішення задачі аналізу спектральних характеристих аудіо записів для виявлення плагіату

Python є динамічно типізованою мовою програмування високого рівня, що поєднує простий синтаксис із широкими можливостями для наукових обчислень і розробки алгоритмів машинного навчання. Завдяки модульній структурі та численним пакетам Python активно використовується в задачах обробки аудіосигналів, зокрема для оцінювання спектральних характеристик та побудови моделей глибокого навчання на основі CNN [18].

Для реалізації STFT та отримання спектрограм часто застосовують SciPy та NumPy, або інші бібліотеки, які їх використовують.

NumPy забезпечує роботу з багатовимірними масивами й оптимізованими математичними операціями над ними, що є основою швидкої обробки сигналів [19].

SciPy розширює функціонал NumPy та містить низку готових функцій для аналізу й перетворень сигналів (зокрема, модуль scipy.signal зі зручними інструментами для обчислення STFT) [20].

Щоб виокремлювати та аналізувати спектральні ознаки, а також виконувати різні види аудіоопераций (наприклад, ресемплінг або фільтрацію), доцільно використовувати бібліотеку librosa. Ця бібліотека містить готові методи для отримання MFCC, хромаграм тощо. Завдяки цьому можна швидко підготувати навчальні вибірки для задачі виявлення плагіату або класифікації звуків [21].

Побудова й тренування згорткових нейронних мереж у Python зазвичай здійснюється у фреймворках PyTorch або TensorFlow.

PyTorch пропонує динамічне розгортання обчислювального графа та зручний механізм автоматичного диференціювання. Це прискорює експерименти зі складними структурами мереж та спеціальними шарами, орієнтованими на

обробку спектральних даних [22].

TensorFlow (разом з високорівневим інтерфейсом Keras) надає гнучкі засоби для створення, тренування та візуалізації моделей різної складності, зокрема глибоких CNN для аналізу аудіо. Однією з ключових характеристик TensorFlow є можливість легко налаштувати виконання обчислень як на центральному процесорі (Central Processing Unit, CPU), так і на графічному процесорі (Graphics Processing Unit, GPU). Використання GPU дозволяє скоротити час навчання моделей у декілька разів порівняно з обчисленнями лише на CPU, що робить TensorFlow ефективним інструментом для дослідників і практиків. Ця особливість забезпечує широкі можливості для проведення експериментів у локальному середовищі, а також у масштабованих хмарних системах, що дозволяє оптимально використовувати доступні обчислювальні ресурси [23].

Keras є високорівневим програмним інтерфейсом (Application Programming Interface, API) для глибокого навчання, що вирізняється своєю простотою використання та гнучкістю. Його основна перевага полягає у зручності побудови моделей, що досягається завдяки інтуїтивно зрозумілому синтаксису. Це дозволяє як досвідченим фахівцям, так і новачкам швидко створювати складні архітектури нейронних мереж. Keras пропонує широкий набір інструментів для моделювання, таких як вбудовані шари (наприклад, згорткові, рекурентні, нормалізації тощо), оптимізатори, функції втрат і метрики. Крім того, фреймворк забезпечує підтримку налаштування кастомних компонентів, що робить його надзвичайно гнучким для експериментальних і дослідницьких задач. Однією з важливих характеристик Keras є його здатність працювати на основі різних бекендів, таких як TensorFlow, Theano або Microsoft Cognitive Toolkit. Це забезпечує масштабованість і дозволяє ефективно використовувати ресурси, як локальні, так і хмарні [24].

Також варто відзначити вбудовані можливості для навчання моделей, зокрема зручний інтерфейс для використання методів раннього завершення навчання, перевірки точності та візуалізації процесу тренування. Усе це робить Keras потужним інструментом для розробки й експериментування в галузі ма-

шинного навчання.

Для графічного відображення результатів досліджень широко використовується Matplotlib, що дає змогу швидко генерувати зображення спектрограм, криві навчання та інші наочні діаграми. Також у Python-екосистемі доступні інструменти (на кшталт seaborn або plotly) для більш розширеної та інтерактивної візуалізації [25].

Таким чином, Python пропонує зручний та гнучкий підхід до всього циклу обробки аудіоданих: від обчислення STFT та виділення спектральних фіч до розробки й навчання складних моделей на базі згорткових нейронних мереж. Така інтегрованість рішень і різноманітність бібліотек робить Python одним із найефективніших інструментів для виконання завдань з аналізу аудіозаписів та виявлення плагіату.

3.2 Алгоритм розв’язання задачі аналізу спектральних характеристик аудіозаписів з метою виявлення плагіату

Алгоритм розв’язання задачі аналізу спектральних характеристик аудіозаписів з метою виявлення плагіату начає наступні етапи:

– завантаження аудіофайлів;

– поділ аудіофайлів на сегменти;

– обрахування MFCC для кожного сегменту;

– присвоєння міток (labels) завантаженим файлам;

– побудова моделі;

– тренування моделі;

– збереження моделі;

– збереження коефіцієнтів моделі (embedings, векторного представлення);

– завантаження аудіофайлу для перевірки на плагіат;

– поділ аудіофайлу на сегменти;

– обрахування MFCC для кожного сегменту;

– отримання передбаченого векторного представлення для обрахованих MFCC;

– порівняння передбаченого векторного представлення зі збереженими векторними представленнями;

– якщо степінь схожості менше порогового значення, то ваажаємо, що аудіофайл не містить плагіату, до зберігаємо його векторне представлення;

– повернення результату.

Для спрощення обробки та маркування аудіофайлів, розміщаємо їх у теці *Audio,* де кожна тека представляє собою один клас, і всі аудіофайли в ній будуть належати до одного класу. Аудіофайл з назвою *Original* буде вважатись саме оригінальною версією аудіозапису. Для спрощення фільтрації файлів будуть оброблятись тільки файли з розширенням *.mp3*.

Тека *Extra* містить аудіофайли, де кожен з них представляє один клас.

Для того щоб забезпечити узгодженість аудіоданих, коректну обробку та сумісність із алгоритмами машинного навчання, необхідно приводити всі файли до єдиної частоти дискретизації. Для цього застосовується операція ресемплінгу. Це дозволяє уникнути помилок, пов’язаних із різною кількістю зразків на одиницю часу, та гарантує однакове представлення даних під час аналізу чи тренування моделі. У якості частоти дикретизації обрано найбільш популярне значення 44100 Гц. Після операції ресемплінгу розраховується кількість семплів на один сегмент аудіофайлу. Для розрахунку розміру вікна та кроку MFCC використовується розмір вікна перетворення Фур’є 16384, яке є рекомендованим, з метою отримання MFCC високої роздільної здатності.

Далі, для кожного сегменту файлу розраховуються MFCC. Після розрахунку MFCC їх значення значення стандартизуються для всіх сегментів.

Далі, з метою надати більшого значення менш представленим класам під час обчислення функції втрат, щоб модель навчалася ефективно для всіх класів, незалежно від їх частоти в даних, розраховуються ваги класів для врахування дисбалансу між класами у навчальних даних.

Далі, будується модель CNN для обробки спектральних MFCC, завдяки їх

здатності виділяти просторові ознаки в даних.

Архітектура даної CNN включає:

– вхідний шар;

– згорткові шари – для виявлення локальних ознак MFCC з фукцією активації ReLU;

– шари підвибірки – для зменшення розмірності даних;

– повноз’язний шар – для підключення ознак до кінцевого шару класиіфкації;

– вихідний шар – для класифікації з використанням сигмоъдноъ функції активації.

Схема моделі приведена на рисунку 3.1.

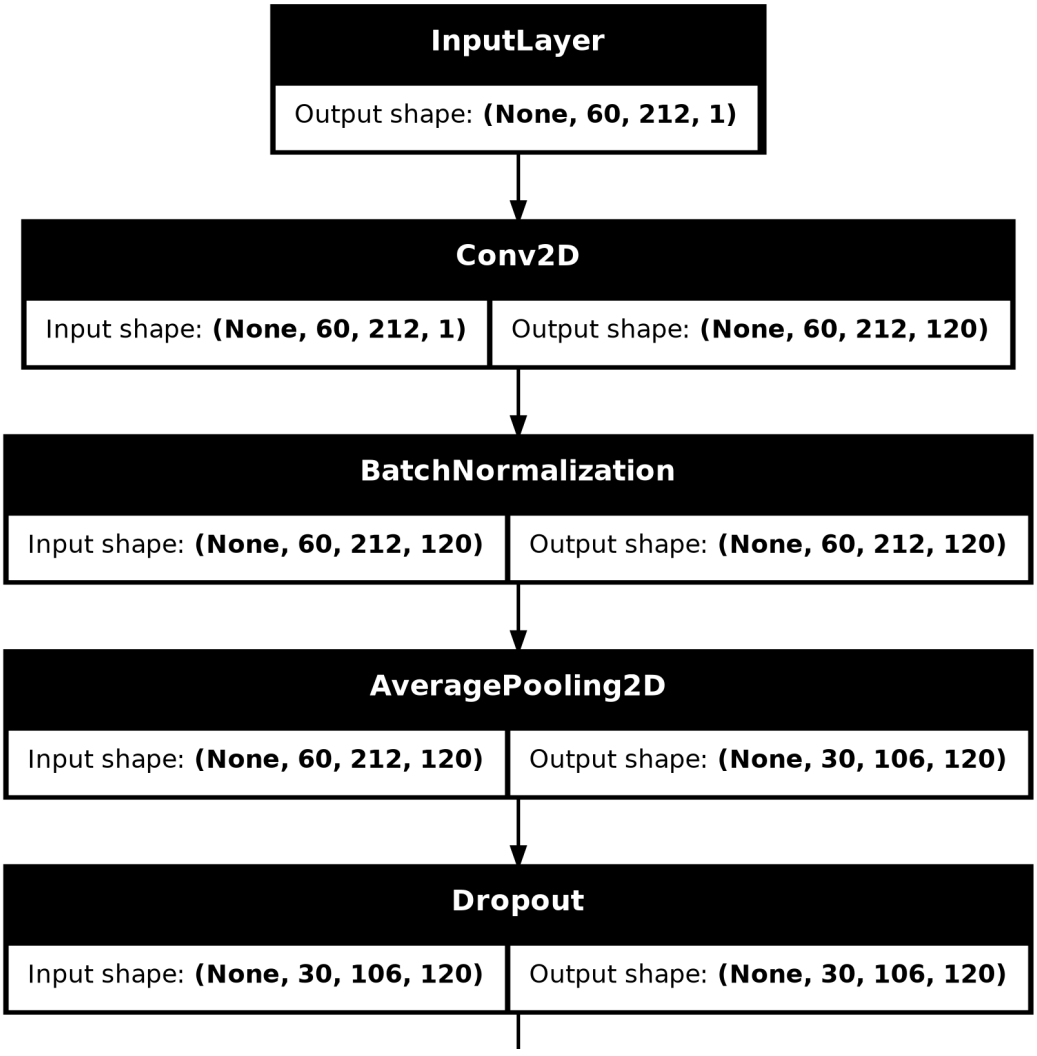

Рисунок 3.1 – Схема моделі

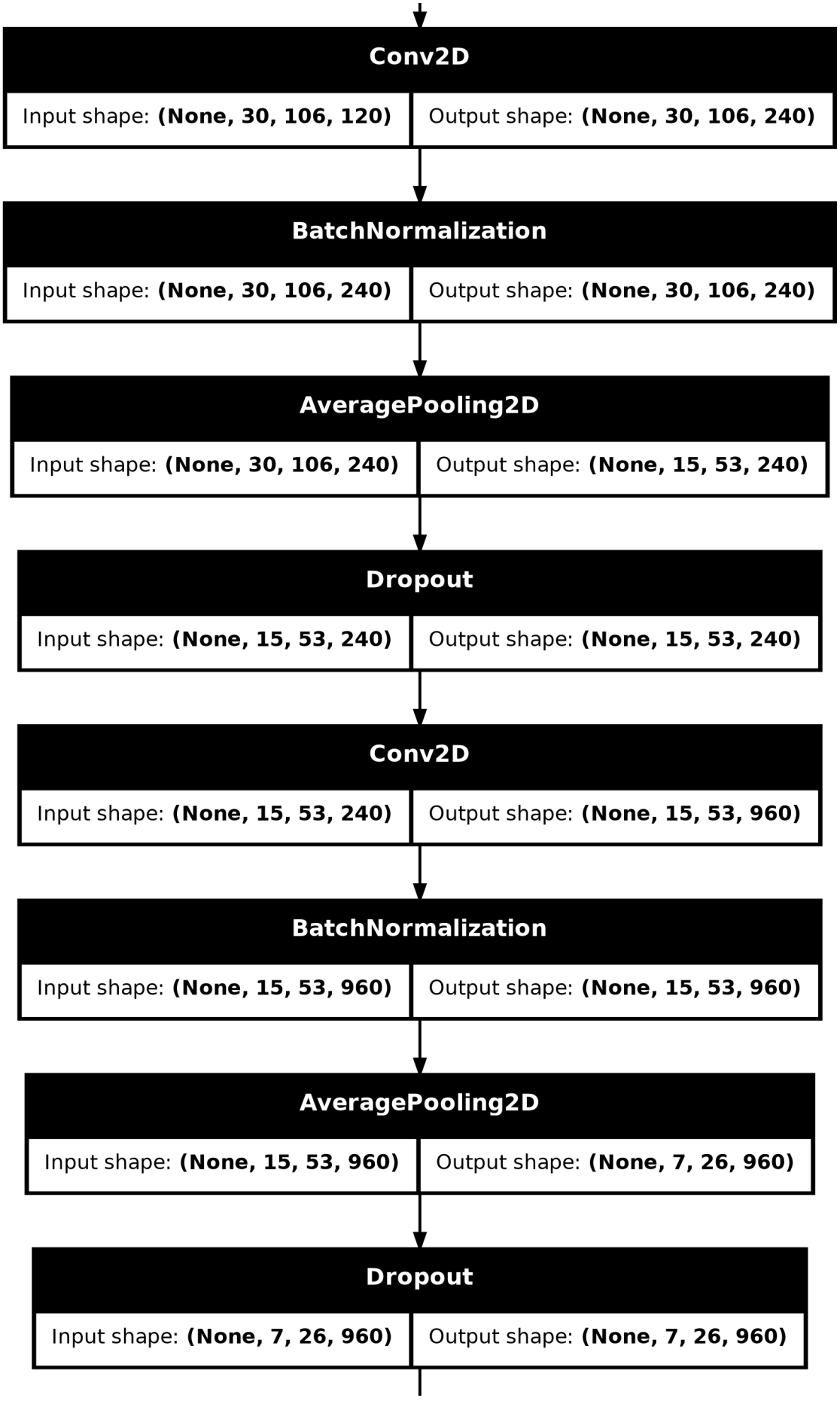

Рисунок 3.1, частина 2

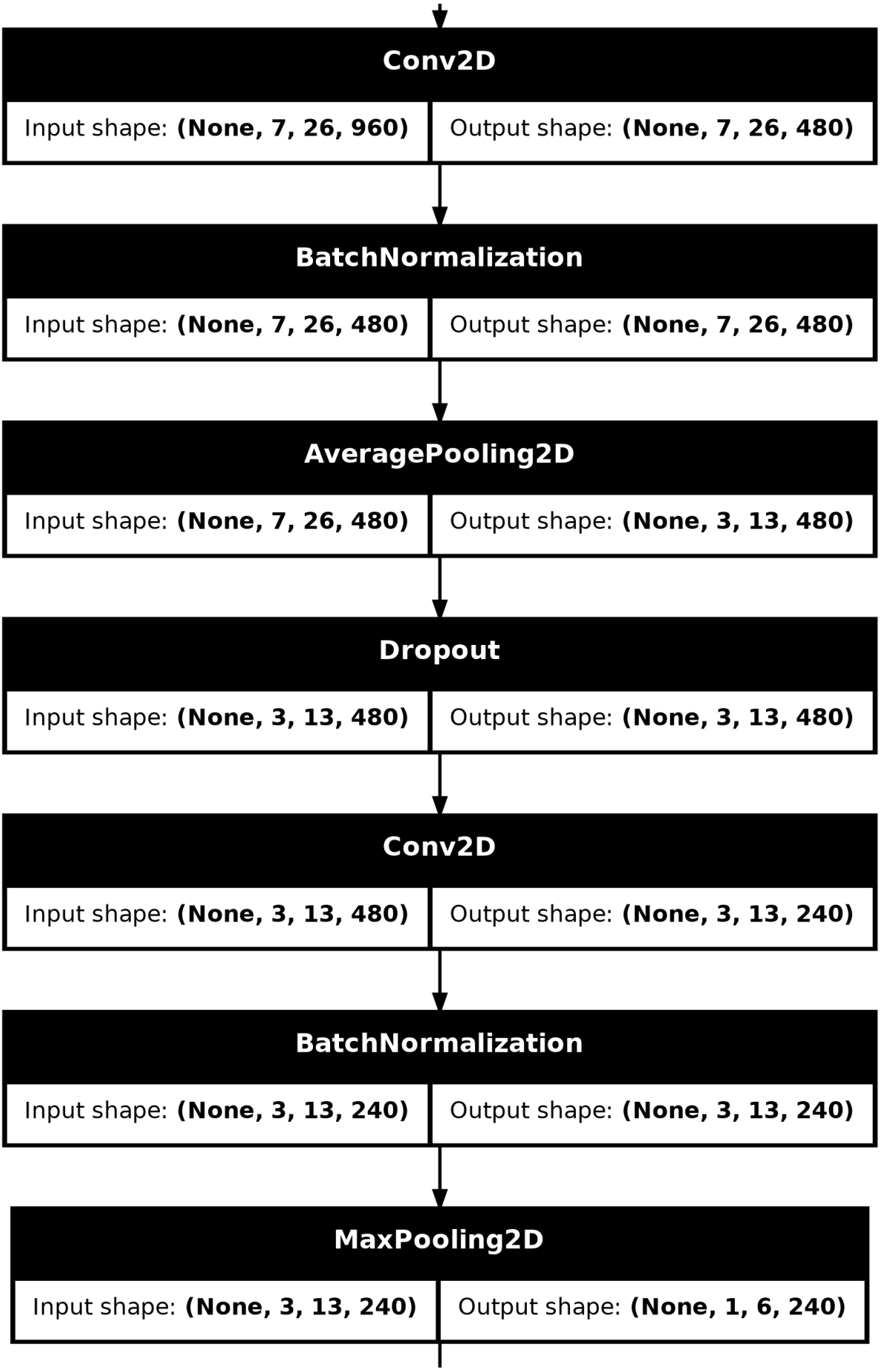

Рисунок 3.1, частина 3

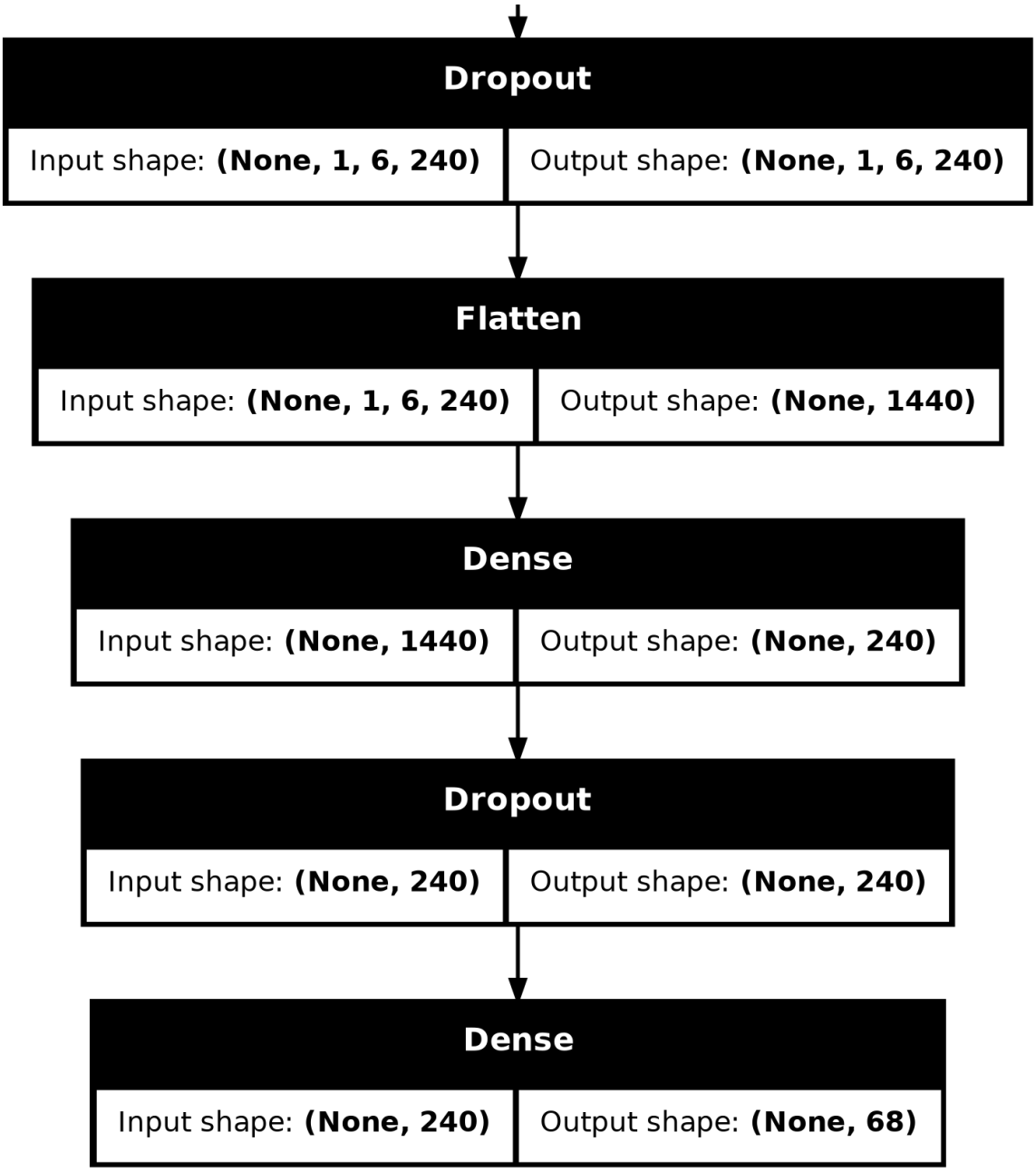

Рисунок 3.1, частина 4

Схема модифікованого варіанту архітектури Siamese Network зображено на рисунку 3.2.

Додавання регуляризації Dropout у модель виконує ключову роль у боротьбі з перенавчанням і покращенні узагальнювальної здатності моделі.

Dropout додається для зменшення залежності моделі від окремих нейронів. Під час навчання випадкове «відключення» певного відсотка нейронів у шарах моделі примушує її знаходити більш стійкі патерни в даних, що зменшує

ризик перенавчання. Це особливо важливо для глибоких моделей, які мають велику кількість параметрів, адже вони більш схильні до надмірного підлаштування під навчальні дані. Разом ці методи дозволяють зробити модель більш стійкою до шуму, покращити її здатність працювати з новими даними та забезпечити стабільність навчання, навіть на складних і неоднорідних наборах даних.

У модифікованій архітектурі Siamese Network було внесено зміни, спрямовані на підвищення точності моделі та її здатності працювати з різними типами вхідних даних. Основна модифікація стосувалася структури мережі, де один із входів, який відповідає за обчислення ембедингів у реальному часі, був збережений у стандартному вигляді. Натомість другий вхід, що отримує ембединги з бази даних, був доповнений окремим набором додаткових шарів для попередньої обробки. Ці зміни дозволяють покращити якість порівняння та підвищити загальну продуктивність моделі в задачах класифікації або виявлення схожості. Схема модифікованої архітектури Siamese Network приведена на рисунку 3.2.

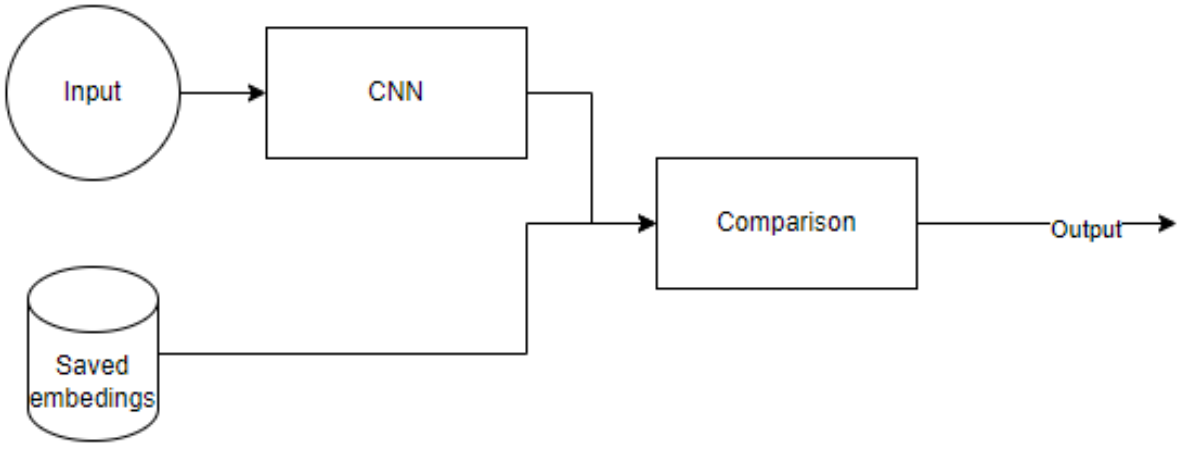

Рисунок 3.2 – Схема модифікованої архітектури Siamese Network

Далі проводиться процес тренування моделі, в якому дані поділяються у співвідношенні 7:3 на тренувальний і валідаційний набори. Це означає, що 70%

даних використовується для навчання моделі, а решта 30% призначається для оцінки її узагальнювальної здатності на даних, які не використовувалися під час навчання. Такий поділ забезпечує баланс між якісним навчанням моделі та надійною оцінкою її продуктивності. Валідаційний набір дозволяє виявляти, чи не перенавчається модель, а також допомагає налаштовувати гіперпараметри, такі як кількість епох, розмір батчу, використання регуляризації тощо. Це важливий етап, що забезпечує коректність і ефективність побудови моделі.

Далі аудіофайл поділяється на сегменти для аналізу, кожен із яких порівнюється зі збереженими ембедингами. Таке розбиття дозволяє обробляти тривалі аудіозаписи покроково, аналізуючи окремі часові фрагменти.

Порівняння виконується за допомогою косинусної функції подібності, яка обчислює схожість між векторами ембедингів. Косинусна подібність визначається як косинус кута між двома векторами у багатовимірному просторі, що дозволяє оцінити, наскільки орієнтація одного вектора близька до іншого, незалежно від їхньої довжини.

Використання косинусної подібності є ефективним у задачах, де важлива семантична схожість, оскільки цей підхід фокусується на напрямку векторів, а не на їхній абсолютній величині. Результат аналізу кожного сегмента дозволяє визначити, чи є збіг між аналізованим аудіо та збереженими ембедингами, а також оцінити рівень подібності.

Косинусоїдальна подібність між двома векторами має вигляд

$$sim(A,B) = \frac{A \cdot B}{\|A\| \cdot \|B\|},$$

де $sim(A,B)$ – функція подібності для векторів A, B;

A, B – вектори.

Після обчислення косинусної подібності для всіх сегментів аудіофайлу та порівняння їх зі збереженими ембедингами, модель повертає результат із найвищим значенням подібності. Це значення відображає рівень схожості між

найбільш відповідним сегментом вхідного файлу та збереженим ембедингом у базі даних. Для кожного сегмента вхідного аудіо обчислюється косинусна подібність із кожним збереженим ембедингом. Після обчислення подібності для всіх пар векторів (сегмент–ембединг) вибирається максимальне значення подібності. Найвищий рівень подібності та відповідний ембединг (або його мітка) повертаються як найкращий збіг. Цей підхід дозволяє ідентифікувати найбільш схожий збережений фрагмент, використовуючи його ембединг.

Алгоритм роботи процедури завантаження файлів та порівняння аудіофайлів на подібність приведені на рисунках 3.3 та 3.4 відповідно.

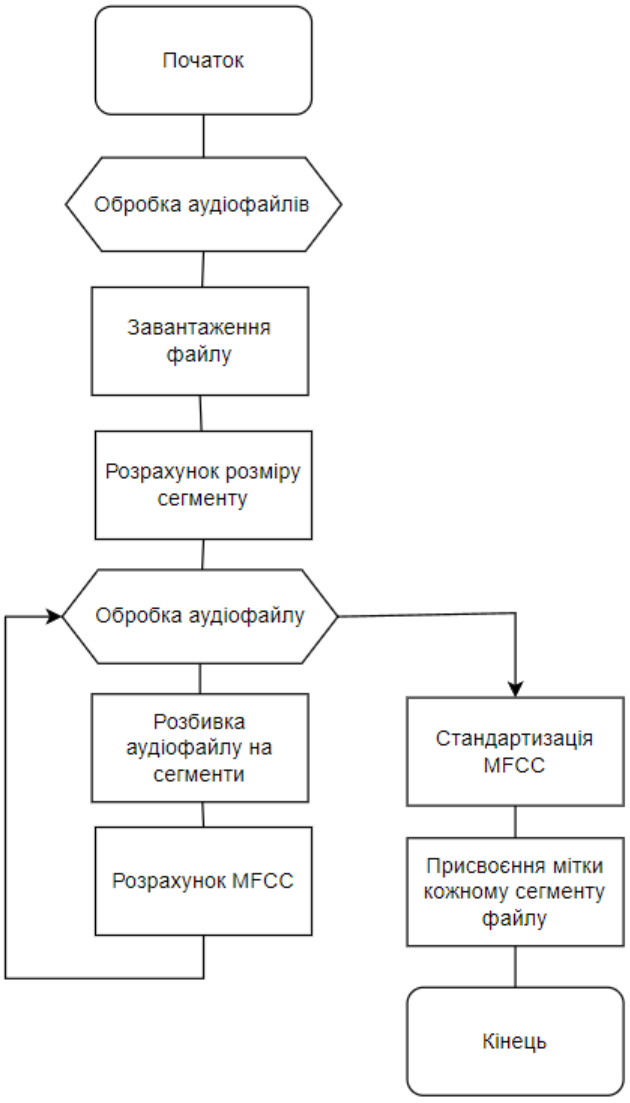

Рисунок 3.3 – Алгоритм роботи процедури завантаження файлів

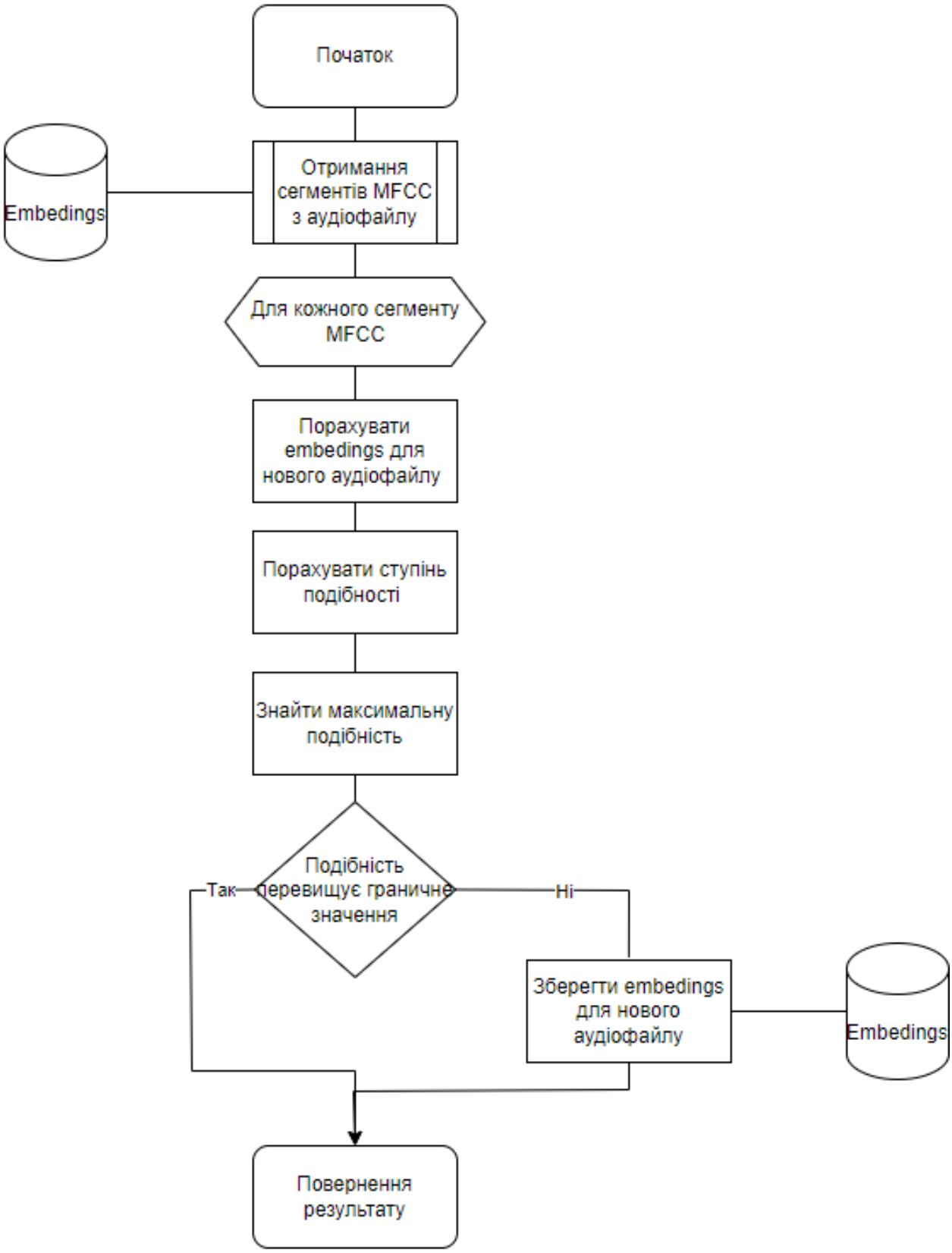

Рисунок 3.4 – Алгоритм роботи процедури порівняння аудіофайлів

Висновки за розділом 3

У 3 розділі було розроблено алгоритм аналізу спектральних характеристик аудіозаписів, який включає ключові етапи обробки даних, побудови моделей та перевірки на плагіат. Для забезпечення узгодженості аудіоданих виконано ресемплінг до частоти дискретизації 44100 Гц. Це дозволило уникнути проблем із різною кількістю зразків на одиницю часу та забезпечило однакове представлення даних.

Для кожного сегмента аудіофайлу було розраховано MFCC із параметрами 16384, що забезпечує високу частотну роздільну здатність. Значення MFCC були стандартизовані, щоб узгодити дані між сегментами та покращити стабільність моделі. Крім того, було розраховано ваги класів для врахування дисбалансу в даних, що дозволило моделі навчатися однаково ефективно для всіх класів.

Побудовано модель CNN для обробки спектральних характеристик, архітектура якої включає вхідний шар, згорткові шари з активацією ReLU, шари підвибірки, повнозв’язний шар і вихідний шар із сигмоїдною активацією. Для зменшення ризику перенавчання модель була доповнена регуляризацією Dropout, що забезпечило стабільність навчання навіть на складних і неоднорідних наборах даних.

У модифікованій архітектурі Siamese Network було додано окремі шари для попередньої обробки ембедингів із бази даних, що дозволило покращити точність порівняння. Для обчислення схожості між сегментами застосовано косинусну подібність, яка забезпечує ефективний аналіз векторних представлень. Ембедінги зберігалися для подальшого використання в задачах перевірки плагіату або класифікації.

4 РЕЗУЛЬТАТИ ОБЧИСЛЮВАЛЬНОГО ЕКСПЕРИМЕНТУ ТА ЇХ АНАЛІЗ

4.1 Тестові приклади

Для проведення обчислювального експерименту були використані тестові дані, що складаються з аудіофайлів різної тривалості, жанрів та джерел. Усі файли були підготовлені шляхом приведення до єдиної частоти дискретизації (44100 Гц) з подальшою сегментацією на рівні інтервали часу. Такий підхід забезпечує узгодженість даних та їхню коректну обробку.

Кожен аудіофайл був асоційований із відповідною міткою (label), яка вказувала на його клас (наприклад, оригінальний трек чи потенційний плагіат). Для спрощення маркування аудіофайли були розподілені по теках, де кожна тека представляє один клас. Аудіофайли з теки *Extra* використовувалися для створення нових класів під час виявлення файлів, які не є плагіатом.

Тестові дані містили:

– оригінальні файли – аудіофайли, які вважались програмою еталонами;

– файли, які містили плагіат – модифіковані записи, які включали кавери від інших виконавців, інші версії треку від оригінального виконавця, наприклад, концетрні записи, та файли записані власноруч;

– невідомі файли – аудіофайли, які раніше не з'являлися у навчальній вибірці, використовувалися для перевірки здатності моделі виявляти нові класи.

Під час експерименту було враховано баланс класів, щоб оцінити точність моделі в умовах рівномірного розподілу даних, а також її стійкість до дисбалансу. Тестові файли були у форматі *.mp3*, що є стандартом для аудіоаналітики, і містили треки різної тривалості від 3 до 8 хвилин.

Такий склад тестових даних дозволив оцінити роботу алгоритму в умовах різноманітності аудіоматеріалів та їх характеристик.

4.2 Аналіз отриманих результатів

З графіка втрат (рис. 4.1) видно, що тренувальні втрати стабільно зменшуються впродовж епох, демонструючи ефективність навчання моделі. Валіднаційні втрати також зменшуються, що свідчить про узгодженість моделі з тестовими даними. Проте на певних етапах спостерігається розходження між тренувальними та валідаційними втратами, що може вказувати на легке перенавчання.

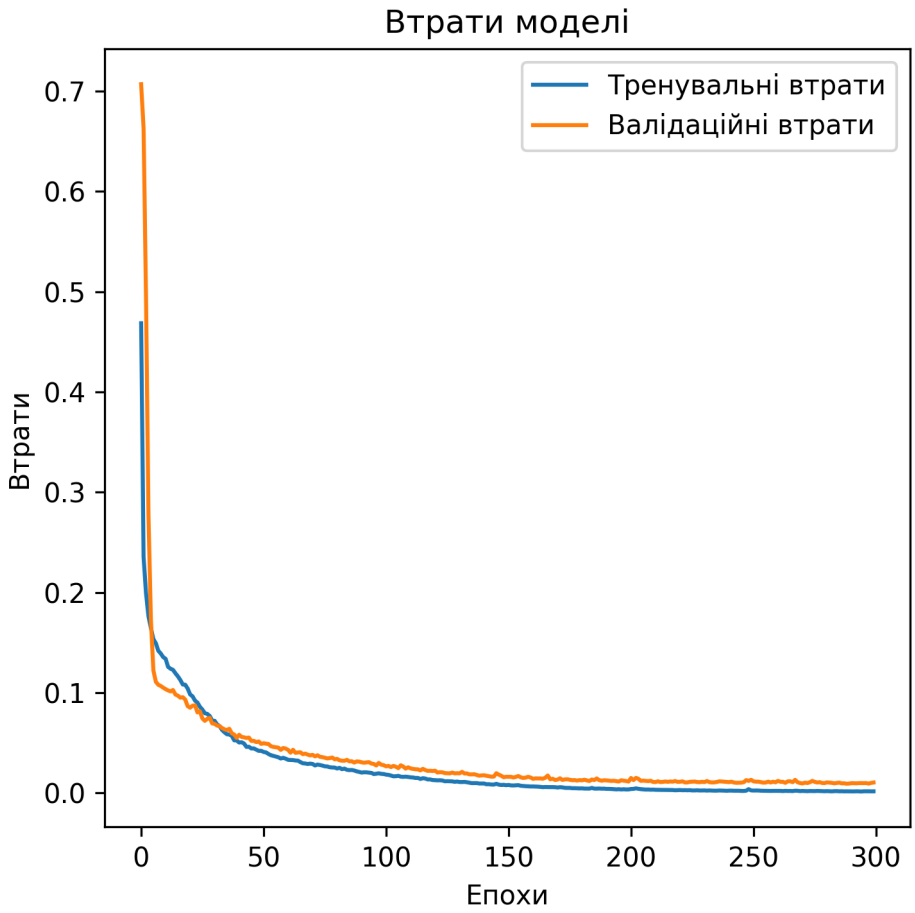

Рисунок 4.1 – Графік втрат

Графік точності (рис. 4.2) демонструє, що модель досягає високої точності на тренувальних даних. Валіднаційна точність є трохи меншою, що може бути наслідком різноманітності або складності тестових даних. Така стабільність показників свідчить про те, що модель ефективно виявляє позитивні класи без

надмірного кількості хибних спрацьовувань.

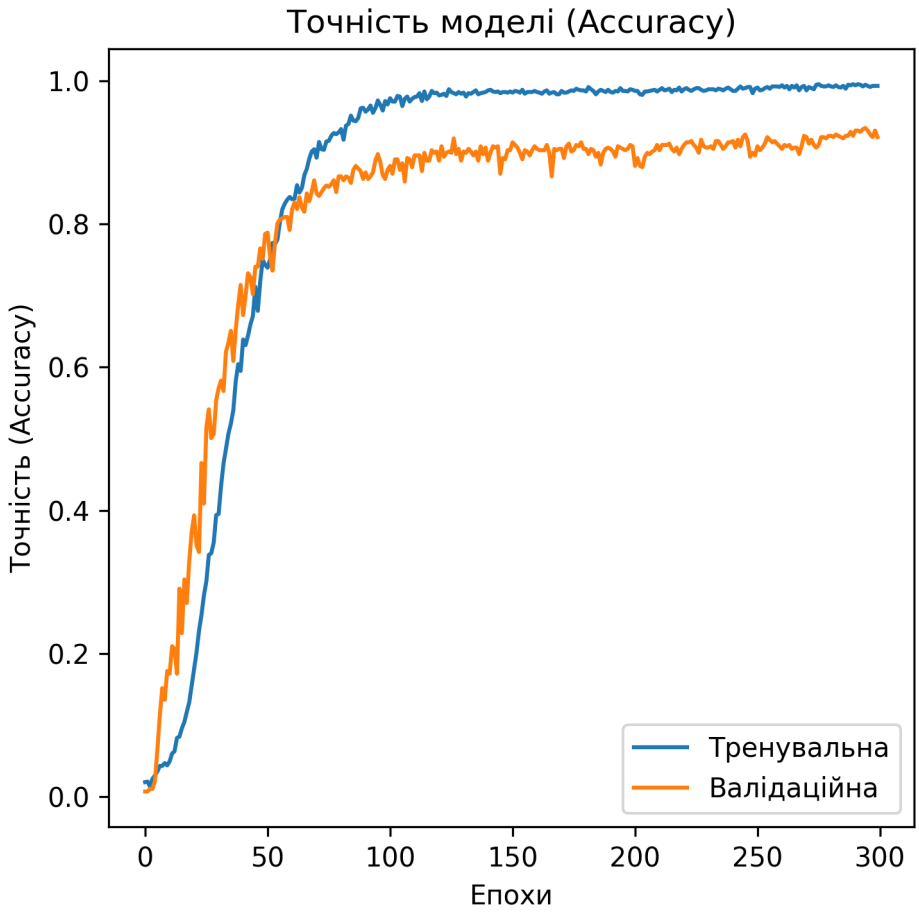

Рисунок 4.2 – Графік точності (Accuracy)

Графік чутливості (рис. 4.3) показує, як модель справляється з виявленням усіх релевантних класів. Значення чутливості на валідаційних даних стабілізуються на рівні, який є трохи нижчим за тренувальні, що може свідчити про потребу у більшій кількості тренувальних прикладів для покращення генералізації.

Графік точності (рис. 4.4) демонструє хорошу узгодженість між тренувальними та валідаційними даними, хоча тренувальні дані демонструють трохи вищі значення. Це є типовою поведінкою для глибоких моделей, які можуть легко адаптуватися до тренувального набору даних.

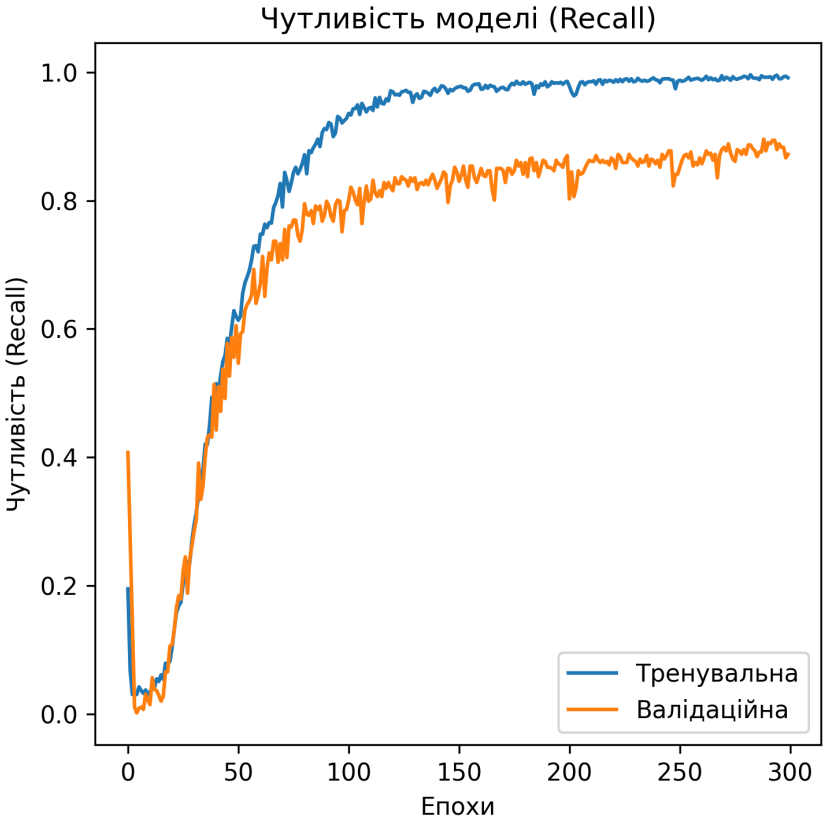

Рисунок 4.3 – Графік чутливості (Recall)

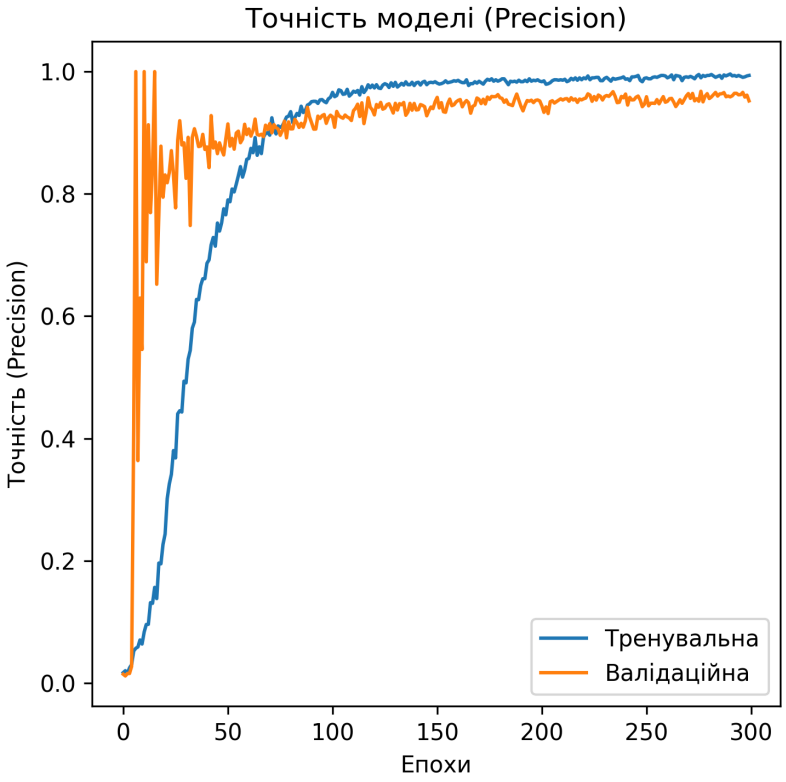

Рисунок 4.4 – Графік точності (Precision)

Візуалізації MFCC для тестових аудіофайлів показують, що модель успішно порівнює спектральні характеристики сегментів. Зокрема, схожість між сегментами оцінюється на основі косинусної подібності, і для кожного тестового аудіофайлу вказана максимальна подібність.

Для файлів «*03. Mother Earth.mp3*» та «*01. Mother Earth.mp3*» MFCC, що зображені на рисунку 4.5 максимальна схожість становить 90.63%, що свідчить про велику ймовірність плагіату. Насправді це один і той же трек від одного й того ж виконавця, тільки 2 різних концертних записи, що й зумовлює високу подібність аудіофайлів.

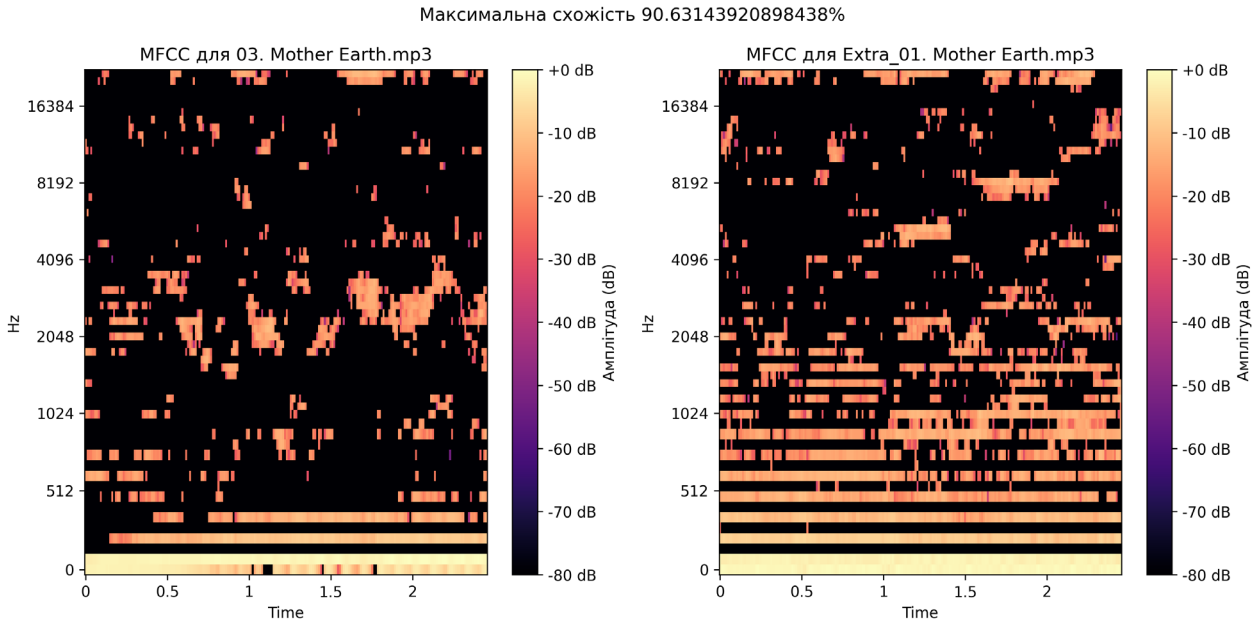

Рисунок 4.5 – MFCC спектрограма для «*03. Mother Earth.mp3*» та «*01. Mother Earth.mp3*»

Для файлів «11 – Tamacun.mp3» та «Tamacun», MFCC яких представлені на рисунку 4.6, спостерігаються помітні відмінності у силі та розподілі амплітудних компонентів. Ці відмінності пояснюються різницею в умовах запису: один файл є студійним оригіналом, тоді як інший – live-запис виконання іншого виконавця. Незважаючи на ці відмінності, модель оцінила схожість на рівні 80.79%, що вказує на значний ступінь подібності між спектральними характеристиками двох аудіозаписів.

Такий високий рівень схожості підтверджує наявність запозичення музичного матеріалу та, відповідно, плагіату. Використання моделі для аналізу цих файлів демонструє її здатність враховувати ключові особливості аудіо, навіть якщо вони виконані в різних умовах, ідентифікуючи основні риси, які залишаються спільними для обох записів.

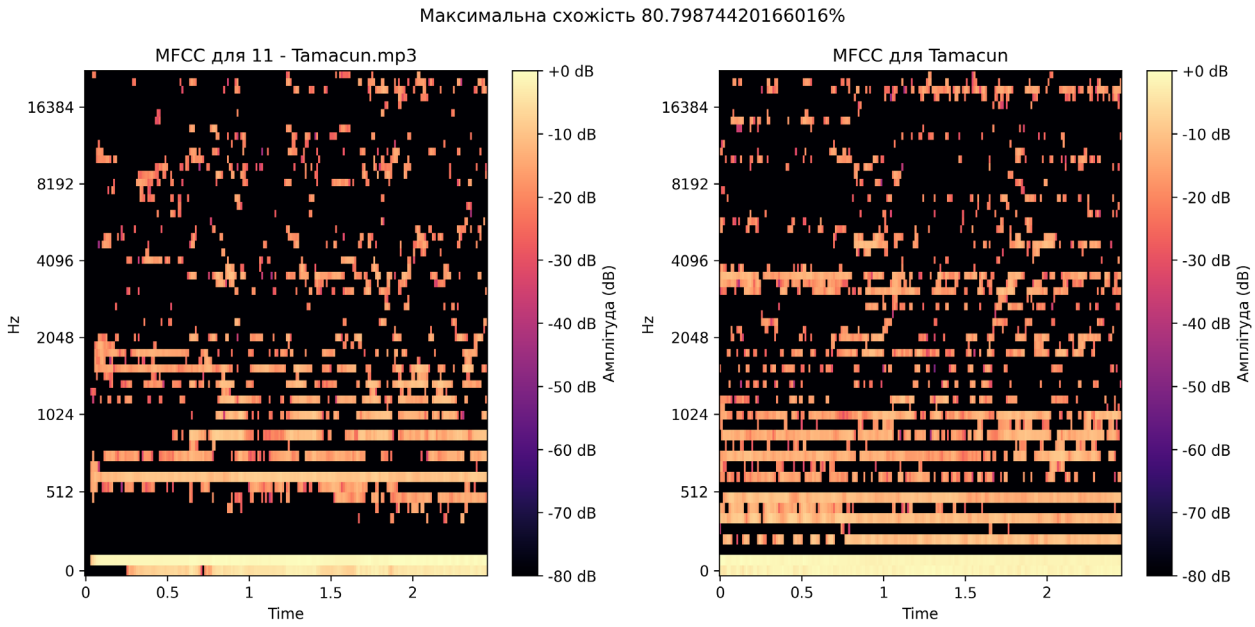

Рисунок 4.6 – MFCC спектрограма для «Tamacun.mp3» та «Tamacun»

Для файлів «05 – Ixtapa.mp3» та «Orion», MFCC яких представлені на рисунку 4.7, модель оцінює рівень схожості на 89.22%, що є дуже високим показником. Такий результат свідчить про наявність сегментів з високим ступенем збігу між спектральними характеристиками цих двох композицій. Враховуючи, що ці треки належать до різних композицій, виконаних різними виконавцями, такий рівень схожості може вказувати на присутність елементів плагіату.

Це також може свідчити про використання спільних музичних тем, мотивів чи гармонійних структур, які є достатньо виразними для моделі, щоб виділити їх як подібні. Такий аналіз дозволяє зробити припущення про можливе запозичення музичного матеріалу чи ідей між композиціями, що потребує додаткового людського аналізу для підтвердження. Висока точність моделі при цьому демонструє її здатність до аналізу та виявлення подібностей навіть у різноп-

ланових творах.

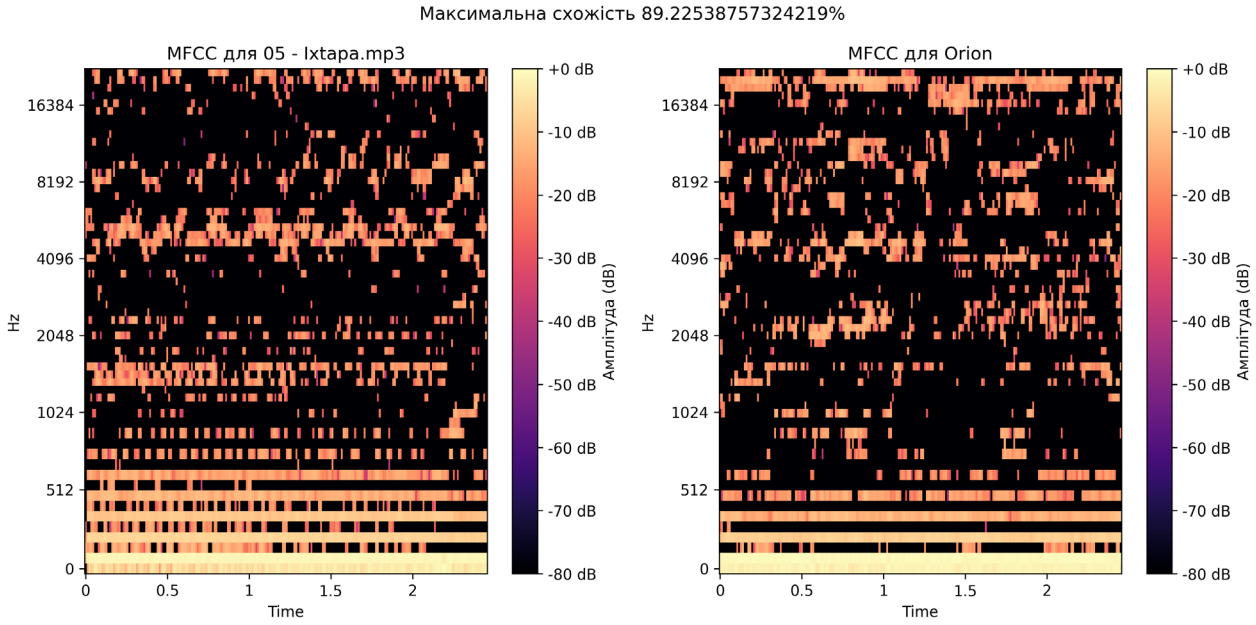

Рисунок 4.7 – MFCC спектрограма для «05 – Ixtapa.mp3» та «Orion»

Висновки за розділом 4

У розділі 4 було проведено обчислювальний експеримент із використанням розробленої моделі для аналізу спектральних характеристик аудіозаписів з метою виявлення плагіату.

Модель демонструє стабільне зменшення втрат як на тренувальних, так і на валідаційних даних. Показники точності, чутливості та загальної точності свідчать про здатність моделі ефективно класифікувати спектральні характеристики аудіофайлів. Досягнута точність на тренувальних даних становила близько 99%, тоді як на валідаційних даних — 88%, що вказує на добру генералізацію моделі. Проведений аналіз підтвердив, що модель здатна виявляти подібності між аудіофайлами з різними умовами запису та різною структурою.

ВИСНОВКИ

У межах дослідження було розроблено та вдосконалено методи аналізу спектральних характеристик аудіозаписів з метою виявлення плагіату. Виконаний аналіз задачі дозволив обґрунтувати вибір сучасних технологій та алгоритмів для обробки аудіоданих. Проведена розробка моделі на основі глибоких нейронних мереж із використанням CNN та модифікованої архітектури Siamese Network для обчислення та порівняння ембедингів спектральних характеристик аудіозаписів.

Отримані результати відповідають сучасному рівню розвитку технічних та наукових знань у галузі машинного навчання та аналізу аудіо. Використання MFCC як вхідних характеристик дозволило підвищити точність і ефективність класифікації та порівняння аудіофайлів.

Розроблена модель була протестована на наборі аудіофайлів різних жанрів і стилів. Експерименти продемонстрували здатність моделі виявляти схожість між аудіозаписами навіть за наявності змін у тональності чи частотному спектрі. Результати показали високий рівень точності та чутливості моделі при аналізі аудіо, що свідчить про її придатність до використання у задачах виявлення плагіату.

Розроблене програмне забезпечення може бути застосоване в галузях, що займаються авторським правом, аудіоаналізом або порівнянням музичних творів. Крім того, методи, реалізовані в межах роботи, можуть бути використані в інших завданнях машинного навчання, що передбачають аналіз нестаціонарних сигналів.

Наукова і практична значущість роботи полягає у розширенні підходів до виявлення плагіату в аудіозаписах, а також у впровадженні ефективних алгоритмів аналізу спектральних характеристик. Подальші дослідження можуть бути спрямовані на оптимізацію моделі для роботи з великими наборами даних, інтеграцію в реальні інформаційні системи.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Цвіркун О.А. Дослідження моделей та методів аналізу спектральних характеристик аудіозаписів з метою виявлення плагіату *VI Міжнародна науково-практична конференця «SCIENTIFIC RESEARCH: MODERN CHALLENGES AND FUTURE PROSPECTS»* (м. Мюнхен, Німеччина, 20-22.01.2025) : зб. матеріалів конференції. MDPC Publishing, 2025. С. 218-221

2. Музичні права: як українським виконавцям захиститися від плагіату URL: https://life.pravda.com.ua/columns/2024/01/18/258909/ (дата звернення 26.09.2024)

3. Wang A. An Industrial Strength Audio Search Algorithm. *4th International Society for Music Information Retrieval Conference*: Proceedings. Baltimore, Maryland, USA. Johns Hopkins University. October 27-30, 2003.

4. Wolf-Monheim, F. Spectral and Rhythm Features for Audio Classification with Deep Convolutional Neural Networks. URL: https://doi.org/10.48550/arXiv.2410.06927 (дата звернення 26.10.2024).

5. Voran, S. Why some audio signal short-time Fourier transform coefficients have nonuniform phase distributions. URL: https://doi.org/10.48550/arXiv.2409.08981 (дата звернення 26.10.2024).

6. Thoshkahna B., Nsabimana F.X., Kalpathi R.R. A Transient Detection Algorithm for Audio Using Iterative Analysis of STFT. *12th International Society for Music Information Retrieval Conference*: Proceedings. ISMIR, , Miami, Florida, USA, October 24-28, 2011. P. 203–208

7. Magron P., Badeau R., David B. Model-Based STFT Phase Recovery for Audio Source Separation. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 2018. №6 (26). P. 1095–1105.

8. Ouali C., Dumouchel P., Gupta V. Robust features for content-based audio copy detection. *Interspeech*: Proceedings. Singapure. 14-18 September 2014. P. 2395–2399.

9. Maksimovic M., Aichroth P., Cuccovillo, L. Detection and localization of

partial audio matches in various application scenarios. *Multimedia Tools and Applications* : Springler, 2021. Vol. 80. P. 22619–22641.

10. Son H., Byun S., Lee S. A Robust Audio Fingerprinting Using a New Hashing Method. *IEEE Access*, Vol. 8, P. 172343–172351.

11. Дубровін В.І. Твердохліб Ю.В., Харченко В.В. Комп'ютерні методи інтелектуальної обробки даних. Запоріжжя: ЗНТУ, 2013. 105 с.

12. Капшій, О.В., Коваль О.І., Русин Б.П. Вейвлет-перетворення у компресії та обробці зображень. Львів: СПОЛОМ, 2008. 208 с.

13. Добровська Л.М., Добровська І.А. Теорія та практика нейронних мереж: навчальний посібник. Київ: НТУУ «КПІ» Видавництво «Політехніка», 2015. 396 с.

14. CNN architectures for large-scale audio classification. / Hershey S., Chaudhuri S., Ellis D.P., Gemmeke J.F., Jansen A., Moore R.C., Plakal M., Platt D., Saurous R.A., Seybold B., Slaney M., Weiss R.J., Wilson K.W. *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP),* 2017. P. 131-135.

15. Improving the Imbalanced Data Accuracy Using CNN and ReLU / Saeed A., Baber J., Abbas M.Z., Sajid A., Razzaq H., Khan A.A. *IETI Transactions on Data Analysis and Forecasting (iTDAF).* Vol. 2 (3). P. 50–58.

16. Audio Steganalysis with Improved Convolutional Neural Network / Lin Y., Wang R., Yan D., Dong L., Zhang X. *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security.* July 2, 2017. P. 210–215.

17. Siamese Neural Network - an overview | ScienceDirect Topics. URL: https://www.sciencedirect.com/topics/computer-science/siamese-neural-network (дата звернення: 24.10.2024).

18. The Python Tutorial – Python 3.13.1 documentation. URL: https://docs.python.org/3/tutorial/index.html (дата звернення: 24.10.2024).

19. What is NumPy? – NumPy v2.2 Manual. URL: https://numpy.org/doc/stable/user/whatisnumpy.html (дата звернення: 24.10.2024).

20. SciPy User Guide – SciPy v1.15.1 Manual. URL: https://docs.scipy.org/doc/scipy/tutorial/index.html (дата звернення: 24.10.2024).

21. Tutorial – librosa 0.10.2 documentation. URL: https://librosa.org/doc/latest/tutorial.html (дата звернення: 24.10.2024).

22. Learn the Basics – PyTorch Tutorials 2.5.0+cu124 documentation. URL: https://pytorch.org/tutorials/beginner/basics/intro.html (дата звернення: 24.10.2024).

23. Guide | TensorFlow Core. URL: https://www.tensorflow.org/guide (дата звернення: 24.10.2024).

24. Getting started with Keras. URL: https://keras.io/getting_started/ (дата звернення: 24.10.2024).

25. Quick start guide – Matplotlib 3.10.0 documentation. URL: https://matplotlib.org/stable/users/explain/quick_start.html#quick-start (дата звернення: 24.10.2024).