

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Чорноморський національний університет імені Петра Могили
Факультет комп'ютерних наук
Кафедра інтелектуальних інформаційних систем

ДОПУЩЕНО ДО ЗАХИСТУ

В. о. завідувача кафедри інтелектуальних
інформаційних систем

_____ Євген СІДЕНКО

« ____ » _____ 2025 р.

КВАЛІФІКАЦІЙНА РОБОТА
НА ЗДОБУТТЯ ОСВІТНЬОГО СТУПЕНЯ МАГІСТРА
ІНТЕЛЕКТУАЛЬНА СИСТЕМА КЛАСИФІКАЦІЇ
МУЗИЧНИХ ЖАНРІВ НА ОСНОВІ СПЕКТРОГРАМ ІЗ
ВИКОРИСТАННЯМ ЗГОРТКОВИХ НЕЙРОННИХ
МЕРЕЖ

Спеціальність 122 Комп'ютерні науки
Освітня програма «Інтелектуальні інформаційні системи»

Здобувач

_____ Дмитро ХІНЬОВ

« ____ » _____ 2025 р.

Керівник канд. фіз.-мат. наук, доцент

_____ Інесса КУЛАКОВСЬКА

« ____ » _____ 2025 р.

Миколаїв – 2025

Чорноморський національний університет імені Петра Могили
(повне найменування закладу вищої освіти)

Факультет	Комп'ютерних наук
Кафедра	Інтелектуальних інформаційних систем
Рівень вищої освіти	Другий (магістерський)
Освітній ступень	Магістр
Спеціальність	122 Комп'ютерні науки
Освітня програма	Інтелектуальні інформаційні системи

ЗАТВЕРДЖУЮ

В. о. завідувача кафедри інтелектуальних
інформаційних систем

_____ Євген СІДЕНКО

« ____ » _____ 2025 р.

ЗАВДАННЯ
на кваліфікаційну роботу здобувача

_____ **Хіньова Дмитра Олеговича**

(прізвище, ім'я, по батькові здобувача)

1. Тема кваліфікаційної роботи:

«Інтелектуальна система класифікації музичних жанрів на основі спектрограм із використанням згорткових нейронних мереж»

Керівник роботи: _____ Кулаковська Інесса Василівна, канд. фіз.-мат. наук, доцент
(прізвище, ім'я, по батькові, посада, науковий ступінь, вчене звання)

Затверджена наказом ЧНУ ім. Петра Могили від «07» 07 2025 р. № 184.

2. Строк представлення кваліфікаційної роботи «23» 12 2025 р.

3. Очікуваний результат роботи та початкові дані: початковими даними є аудіофайли з GTZAN Genre Collection або інших відкритих музичних датасетів,

що містять музичні фрагменти тривалістю 30 секунд у форматі WAV, розподілені за жанрами. Для кожного аудіозапису можуть бути сформовані спектрограми, мел-спектрограми та MFCC-представлення, які виступають основою для навчання моделі. Додаткові дані включають числові та спектральні характеристики аудіосигналів (частотні компоненти, тембр, динаміка звучання), метадані про жанрову приналежність композицій, а також програмні бібліотеки для обробки аудіосигналів і побудови моделей глибокого навчання (librosa, NumPy, TensorFlow/Keras тощо).

Очікуваний результат: розроблена інтелектуальна система автоматичної класифікації музичних жанрів на основі спектрограм із використанням згорткових нейронних мереж, здатна приймати аудіофайли, будувати їх спектральні представлення та визначати жанр композиції з високою точністю. Система повинна мати можливість інтеграції у музичні сервіси, рекомендаційні системи та інструменти каталогізації аудіоконтенту.

4. Перелік питань, що підлягають розробці: аналіз особливостей класифікації музичних творів та специфіки музичних жанрів; огляд сучасних методів цифрової обробки аудіосигналів та підходів до класифікації музики; порівняльний аналіз архітектур згорткових нейронних мереж для роботи зі спектрограмами; обґрунтування вибору типової або модифікованої архітектури CNN для задачі класифікації; розробка конвеєра попередньої обробки аудіоданих і автоматичного формування спектрограм; реалізація та навчання моделі на основі CNN з оптимізацією гіперпараметрів; розробка програмного комплексу з підтримкою завантаження аудіофайлів і візуалізацією спектрограм; тестування та оцінка точності роботи системи на тестовому наборі даних із використанням відповідних метрик; аналіз отриманих результатів та виявлення можливостей подальшого вдосконалення системи.

5. Перелік графічних матеріалів: сторінок – 72, таблиць – 8, рисунків – 12, посилань – 40, додатків – 0

Керівник роботи

(Особистий підпис)

Інесса КУЛАКОВСЬКА

(Власне ім'я ПРІЗВИЩЕ)

Здобувач

(Особистий підпис)

Дмитро ХІНЬОВ

(Власне ім'я ПРІЗВИЩЕ)

Дата видачі завдання «28» червня 2025 р.

КАЛЕНДАРНИЙ ПЛАН кваліфікаційної роботи

Тема: Інтелектуальна система класифікації музичних жанрів на основі спектрограм із використанням згорткових нейронних мереж.

№	Найменування роботи	Початок	Закінчення	Примітки
1	Отримання завдання на виконання КР	27.06.2025	30.06.2025	Виконано
2	Аналіз предметної області та постановка задачі	1.07.2025	20.07.2025	Виконано
3	Огляд та аналіз літературних джерел і сучасних підходів до класифікації музичних жанрів на основі аудіосигналів.	21.07.2025	30.07.2025	Виконано
4	Аналіз архітектур нейронних мереж, зокрема згорткових, для автоматичної класифікації музичних жанрів.	01.09.2025	25.10.2025	Виконано
5	Реалізація та оцінка ефективності системи класифікації музичних жанрів із використанням обраних методів і технологій.	26.10.2025	21.11.2025	Виконано
6	Перший попередній захист КР на засіданні комісії кафедри	25.11.2025	26.11.2025	Виконано
7	Корегування роботи за результатами попереднього захисту	27.11.2025	05.12.2025	Виконано
8	Другий попередній захист КР на засіданні комісії кафедри	09.12.2025	09.12.2025	Виконано

Керівник роботи

(Особистий підпис)

Інесса КУЛАКОВСЬКА

(Власне ім'я ПРІЗВИЩЕ)

Здобувач

(Особистий підпис)

Дмитро ХІНЬОВ

(Власне ім'я ПРІЗВИЩЕ)

Дата складання календарного плану

«8» липня 2025 р.

АНОТАЦІЯ

до кваліфікаційної роботи
здобувача групи 601 ЧНУ ім. Петра Могили

Хіньова Дмитра Олеговича

Тема: «ІНТЕЛЕКТУАЛЬНА СИСТЕМА КЛАСИФІКАЦІЇ МУЗИЧНИХ ЖАНРІВ НА ОСНОВІ СПЕКТРОГРАМ ІЗ ВИКОРИСТАННЯМ ЗГОРТКОВИХ НЕЙРОННИХ МЕРЕЖ»

Ефективних інструментів автоматичного аналізу аудіоконтенту. Однією з актуальних задач у цій галузі є класифікація музичних жанрів, яка дозволяє впорядковувати великі обсяги музичних даних, покращувати пошук та формувати рекомендаційні системи. Використання методів машинного навчання, зокрема глибоких нейронних мереж, дає змогу автоматизувати процес класифікації та підвищити його точність.

Об'єктом дослідження: процес класифікації музичних композицій.

Предметом дослідження є методи та алгоритми обробки спектрограм аудіосигналів та їх класифікації за допомогою згорткових нейронних мереж для розпізнавання музичних жанрів

Мета дослідження – розробити та обґрунтувати ефективні методи та алгоритми обробки спектрограм аудіосигналів і побудови моделей класифікації на основі згорткових нейронних мереж з метою підвищення точності та надійності автоматизованого розпізнавання музичних жанрів.

Пояснювальна записка складається зі вступу, чотирьох розділів, висновків та додатків.

У **першому розділі** проведено аналіз предметної області, розглянуто сучасні підходи до обробки аудіосигналів та методи класифікації музичних жанрів, а також обґрунтовано актуальність використання згорткових нейронних мереж для даної задачі.

У **другому розділі** виконано проектування інтелектуальної системи, визначено її архітектуру та основні функціональні компоненти, а також обґрунтовано вибір мел-спектрограм як способу подання аудіоданих для подальшої класифікації.

У **третьому розділі** описано алгоритми та методи реалізації системи, зокрема етапи попередньої обробки аудіосигналів, побудову мел-спектрограм та принципи роботи згорткової нейронної мережі для класифікації музичних жанрів.

У **четвертому розділі** наведено програмну реалізацію розробленої системи, описано інтеграцію з Telegram-ботом, а також представлено результати тестування та аналіз ефективності роботи системи.

У результаті виконання кваліфікаційної роботи розроблено інтелектуальну систему класифікації музичних жанрів на основі спектрограм із використанням згорткових нейронних мереж, яка продемонструвала достатню точність класифікації та може бути використана для автоматичного аналізу музичних аудіоданих у навчальних і прикладних задачах.

Кваліфікаційна робота містить **75** сторінки, **8** рисунків, **12** таблиць, **40** використаних джерел та **0** додатків.

Ключові слова: класифікація музичних жанрів, аудіосигнали, спектрограми, мел-спектрограми, машинне навчання, згорткові нейронні мережі, інтелектуальні системи, Telegram-бот.

ABSTRACT

to the qualification work by the student of the group 601 of Petro Mohyla Black Sea National University

Khinov Dmytro

"INTELLIGENT SYSTEM FOR CLASSIFYING MUSIC GENRES BASED ON SPECTOGRAMS USING CONVOLUTIONAL NEURAL NETWORKS"

Modern digital music services and multimedia platforms require effective tools for automatic analysis of audio content. One of the relevant tasks in this field is music genre classification, which allows organizing large volumes of music data, improving search processes, and supporting recommendation systems. The use of machine learning methods, particularly deep neural networks, makes it possible to automate the classification process and increase its accuracy.

Object of the research is the process of automated classification of musical audio signals.

Subject of the research is methods and algorithms of machine learning for music genre classification based on spectral features of audio signals.

The purpose of the qualification work is to develop an intelligent system capable of automatically determining the music genre of an audio file based on mel-spectrograms using convolutional neural networks.

The explanatory note consists of an introduction, four chapters, conclusions, and appendices.

In the **first chapter**, the subject area is analyzed, modern approaches to audio signal processing and music genre classification are reviewed, and the relevance of using convolutional neural networks for this task is substantiated.

In the **second chapter**, the design of the intelligent system is carried out, its architecture and main functional components are defined, and the choice of mel-spectrograms as a method for representing audio data for further classification is justified.

In the **third chapter**, the algorithms and methods for system implementation are described, including audio signal preprocessing stages, mel-spectrogram generation, and the principles of operation of the convolutional neural network for music genre classification.

In the **fourth chapter**, the software implementation of the developed system is presented, the integration with a Telegram bot is described, and the testing results along with the analysis of the system's efficiency are provided.

As a result of the qualification work, an intelligent system for music genre classification based on spectrograms using convolutional neural networks was developed. The system demonstrated sufficient classification accuracy and can be used for automatic analysis of musical audio data in educational and applied tasks.

The qualification work contains **75** pages, **8** figures, **12** tables, **40** references, and **0** appendices.

Keywords: music genre classification, audio signals, spectrograms, mel-spectrograms, machine learning, convolutional neural networks, intelligent systems, Telegram bot.

ЗМІСТ

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАЧКИ	4
ВСТУП.....	5
1 АНАЛІЗ ПРОБЛЕМИ ТА ПОСТАНОВКА ЗАДАЧІ	7
1.1 Особливості класифікації музичних жанрів.....	7
1.2 Методи аналізу аудіосигналів та спектрограмний підхід.....	9
1.3 Сучасні підходи до класифікації музичних жанрів	10
1.4 Основні аспекти у побудові моделі.....	12
1.5 Аналіз досліджень та існуючих інформаційних систем-аналогів	13
1.6 Постановка завдання.....	18
Висновки до розділу 1	20
2 МОДЕЛІ, МЕТОДИ КЛАСИФІКАЦІЇ МУЗИЧНИХ ЖАНРІВ.....	21
2.1 Методи для вирішення цифрової обробки аудіосигналів.....	21
2.2 Технології розробки системи	25
Висновки до розділу 2	27
3 РОЗРОБКА СИСТЕМИ ТА АНАЛІЗ ОТРИМАНИХ РЕЗУЛЬТАТІВ.....	29
3.1 Опис вхідних даних та структури системи проектування	29
3.2 Архітектура моделі класифікації музичних жанрів	35
3.3 Аналіз та дослідження отриманих результатів.....	46
Висновки до розділу 3	50
4 ПРОГРАМНА РЕАЛІЗАЦІЯ ТА ТЕСТУВАННЯ	51
4.1 Опис програмної реалізації.....	51
4.2 Керівництво користувача.....	61

4.3 Тестування програмної реалізації.....	64
Висновки до розділу 4	67
ВИСНОВКИ	69
ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ	72

СКОРОЧЕННЯ ТА УМОВНІ ПОЗНАЧКИ

AI	–	штучний інтелект
НМ	–	нейронна мережа
ЗНМ	–	згорткова нейронна мережа
ПЗ	–	програмне забезпечення
ПО	–	програмне оточення
AI	–	Artificial Intelligence (штучний інтелект)
CNN	–	Convolutional Neural Network (згорткова нейронна мережа)
ML	–	Machine Learning (машинне навчання)
DL	–	Deep Learning (глибинне навчання)
MFCC	–	Mel-Frequency Cepstral Coefficients (мел-частотні кепстральні коефіцієнти)
FFT	–	Fast Fourier Transform (швидке перетворення Фур'є)
STFT	–	Short-Time Fourier Transform (короткочасне перетворення Фур'є)
API	–	Application Programming Interface (інтерфейс прикладного програмування)
JSON	–	JavaScript Object Notation (формат обміну даними)
HTTP	–	HyperText Transfer Protocol (протокол передачі гіпертексту)
IDE	–	Integrated Development Environment (інтегроване середовище розробки)
CPU	–	Central Processing Unit (центральний процесор)
GPU	–	Graphics Processing Unit (графічний процесор)
RAM	–	Random Access Memory (оперативна пам'ять)
OS	–	Operating System (операційна система)
UI	–	User Interface (користувачський інтерфейс)

ВСТУП

У сучасному світі музичні сервіси й різні онлайн-платформи постійно працюють з великими обсягами аудіоданих, тому питання автоматичної класифікації музичних жанрів стає дедалі актуальнішим. Багато застосунків формують рекомендації, створюють плейлисти або просто впорядковують музичні файли, і для цього потрібно точно визначати жанр композиції. Ручні методи або старі алгоритми вже не дають потрібної точності, особливо коли аудіо різної якості або має складну структуру.

Сучасні підходи дозволяють використовувати глибоке навчання, зокрема згорткові нейронні мережі, які добре працюють із зображеннями [22,23,34]. Якщо аудіосигнал перетворити у спектрограму, то фактично його можна подати як “картинку”, де чітко видно частотні зміни в часі. Саме це робить CNN ефективним рішенням для класифікації музики, оскільки мережа здатна автоматично знаходити характерні ознаки жанру, які складно виділити вручну.

Актуальність теми пов’язана з тим, що музичні платформи, стримінгові сервіси та навіть мобільні додатки постійно потребують автоматичного аналізу треків. Система, яка може точно визначати жанр, значно спрощує роботу з аудіобазами й допомагає покращити персоналізацію рекомендацій. Крім того, такий підхід може бути використаний і в інших задачах аудіоаналізу.

Об’єктом дослідження: процес класифікації музичних композицій.

Предметом дослідження є методи та алгоритми обробки спектрограм аудіосигналів та їх класифікації за допомогою згорткових нейронних мереж для розпізнавання музичних жанрів

Мета дослідження – розробити та обґрунтувати ефективні методи та алгоритми обробки спектрограм аудіосигналів і побудови моделей класифікації на основі згорткових нейронних мереж з метою підвищення точності та надійності автоматизованого розпізнавання музичних жанрів.

Для досягнення поставленої мети визначаються наступні завдання:

- розглянути особливості музичних жанрів та способи цифрового подання аудіосигналів;
- проаналізувати існуючі методи класифікації музики та визначити їх недоліки;
- підготувати дані у вигляді спектрограм для навчання моделі;
- побудувати й навчити згорткову нейронну мережу;
- перевірити точність роботи моделі на тестових даних;
- оцінити можливості практичного застосування.

Загалом, тема є актуальною, оскільки поєднує сучасні методи машинного навчання та обробки аудіо й може бути корисною як у навчальних, так і в реальних прикладних системах.

1 АНАЛІЗ ПРОБЛЕМИ ТА ПОСТАНОВКА ЗАДАЧІ

1.1 Особливості класифікації музичних жанрів

Класифікація музичних жанрів на перший погляд здається простою задачею, оскільки більшості людей інтуїтивно зрозуміло, чим відрізняються, наприклад, рок від класики або хіп-хоп від джазу [6]. Проте для комп'ютерної системи це завдання є набагато складнішим, адже музика — це багатовимірний сигнал, у якому одночасно поєднуються ритмічні, тембральні, тональні та динамічні складові [39]. Саме поєднання цих елементів і визначає жанрову приналежність твору [38].

На практиці музичний жанр не є точною математичною категорією [7]. Його визначення формується культурними традиціями, історичним розвитком, регіональними особливостями, естетичними уявленнями та навіть комерційними інтересами музичної індустрії [15]. Через це один і той самий трек може бути одночасно віднесений до різних жанрів залежно від того, як саме систематизує музику певна платформа або база даних [16].

Додаткову складність створює використання гібридних і перехідних жанрів [36]. Багато сучасних виконавців експериментують зі стилістикою, навмисно змішуючи елементи з різних напрямів [9]. Наприклад, композиції у напрямі trap можуть містити елементи електронної та поп-музики, а інді-рок — елементи фолку або синтвейву [42]. Для алгоритмів класифікації це створює проблему: модель повинна не лише «впізнати» окремі музичні характеристики, але й визначити, які з них є домінуючими [18].

Ще один важливий фактор — варіативність виконання та звукозапису [40]. Один і той самий твір може мати студійну, концертну, живу акустичну та ремікс-версію, які помітно відрізняються за спектральними [37] характеристиками. Крім цього, на спектр сигналу істотно впливають:

- шум та реверберація приміщення під час запису;

- якість використаної апаратури;
- обробка звукорежисера (компресія, еквалізація, фільтри);
- ступінь агресивності та тембр вокалу;
- кількість одночасних інструментів на міксі;
- типи синтезованих або живих ударних.

Через це жанрові ознаки залежать не лише від музичної структури, але й від акустичних особливостей запису, що значно ускладнює автоматизацію класифікації [7].

Окремим викликом є часова динаміка композиції. Деякі жанри мають стабільну структуру протягом усього твору (наприклад, хаус або техно), тоді як інші можуть радикально змінюватися всередині одного треку (особливо прогресивний рок, саундтреки, построк). Для нейромережі стає проблемою визначити, яку частину треку аналізувати, щоб отримати коректну жанрову ознаку. У деяких дослідженнях пропонують аналізувати лише центральні 30 секунд композиції, але це не універсальне рішення.

Ще один аспект — нерівномірність жанрових датасетів. У реальних колекціях даних завжди є жанри з великою кількістю записів (поп, рок), і є жанри, що представлені дуже мало (джаз, блюз, класика). Машинне навчання в таких умовах призводить до перенавчання: модель блискуче «вгадує» популярні жанри, але робить багато помилок із рідкісними.

Підсумовуючи, можна сказати, що класифікація музичних жанрів вимагає алгоритмів, які спроможні виділяти:

- структурні ритмічні патерни (ударні, пульсація, BPM);
- темброві особливості інструментів і вокалу;
- спектральні та гармонійні характеристики;
- елементи композиційної будови.

Саме тому сучасні системи музичної класифікації спираються на аналіз спектрограм аудіосигналу, які дозволяють виявляти вищезазначені закономірності.

1.2 Методи аналізу аудіосигналів та спектрограмний підхід

Методи аналізу аудіосигналів еволюціонували разом із розвитком цифрової обробки сигналів [5]. На ранніх етапах дослідження музики частіше використовували амплітудний аналіз — спостереження коливань сигналу у часі. Цей підхід може бути корисним для виявлення загальних закономірностей гучності, проте він практично непридатний для розрізнення жанрів, оскільки два треки абсолютно різних стилів можуть мати схожий часовий профіль.

Розвиток алгоритмів привів до широкого застосування частотного аналізу, оскільки саме розподіл енергії по частотах формує тембр і характер звучання. Наприклад, класична музика зазвичай містить плавні зміни частотних компонент, тоді як електронна музика характеризується сильними низькочастотними ударами та вираженим синтезаторним спектром. Метал, у свою чергу, вирізняється потужними високочастотними гармоніками, пов'язаними зі скрипучими гітарами та тарілками ударної установки.

Однак і частотного аналізу виявилось недостатньо, оскільки музика — це динамічний процес, який постійно змінюється [8]. Відтак точне визначення жанру вимагає урахування як частотних, так і часових характеристик одночасно. Це привело до використання короткочасного перетворення Фур'є (Short-Time Fourier Transform, STFT), результатом якого є спектрограма — представлення сигналу у вигляді матриці частота $\times$ час $\times$ інтенсивність.

Перевага спектрограм полягає в тому, що вони дозволяють «побачити» структуру музики [10]. Наприклад:

- регулярні короткі вертикальні лінії характерні для ударних у хіп-хопі [11];

- плавні, розмиті області — для ambient або класичної музики [12];
- часті високочастотні «сплески» — для металу та панку [13];
- періодичні модульовані хвилі — для транс-музики та хаусу [14].

Сучасні дослідження показують, що спектрограма — оптимальний формат для роботи з аудіо в задачах глибокого навчання, оскільки CNN самостійно «вчиться» розпізнавати приховані співвідношення між частотами та часом.

Для підвищення інформативності інколи разом зі спектрограмою використовують і MFCC (Mel-Frequency Cepstral Coefficients). MFCC не зберігає детальну картину спектра, але містить інформацію про тембр, тобто про те, як людина сприймає звук. Це дозволяє краще моделювати вокальні стилі, інтонаційні переходи та особливості інструментів.

Таким чином, для автоматичної класифікації музичних жанрів найбільш ефективним є комбінований підхід: побудова спектрограм + використання CNN, що і реалізується в межах цієї дипломної роботи.

1.3 Сучасні підходи до класифікації музичних жанрів

Методи класифікації музичних жанрів умовно можна поділити на три групи: класичне машинне навчання, глибоке навчання та гібридні моделі.

Класичні методи

На ранніх етапах розвитку обробки аудіо застосовували алгоритми SVM, kNN та Random Forest. Ці моделі добре працювали на малих обсягах даних, проте вимагали ручного формування ознак — тобто дослідник сам вирішував, які характеристики сигналу є важливими [19]. Результат сильно залежав від досвіду людини, що робило процес суб'єктивним і нестабільним [20].

Глибоке навчання

З розвитком обчислювальних ресурсів ключову роль почали відігравати згорткові нейронні мережі (CNN), здатні автоматично вилучати ознаки зі спектрограм. CNN навчилися виявляти патерни, характерні для жанрів:

- пульсуючі низькі частоти — електронна музика;
- синкоповані ударні — хіп-хоп та R&B;
- домінування струнних — класична музика;
- агресивні високі гармоніки — метал.

При цьому CNN не потребують попереднього виділення ознак — вони вчаться самостійно. Саме тому в наукових статтях за останні роки CNN демонструють найвищу точність серед усіх підходів.

Для вирішення задачі класифікації музичних жанрів на основі візуального аналізу спектрограм було обрано архітектуру Згорткової нейронної мережі (ЗНМ). ЗНМ є найбільш ефективним інструментом для обробки двовимірних даних, таких як зображення (у нашому випадку — спектрограми), оскільки вона здатна автоматично виділяти ієрархію ознак: від простих (лінії, контури) до складних (ритмічні та темброві патерни).

Принципова схема архітектури ЗНМ (рис. 1.1).

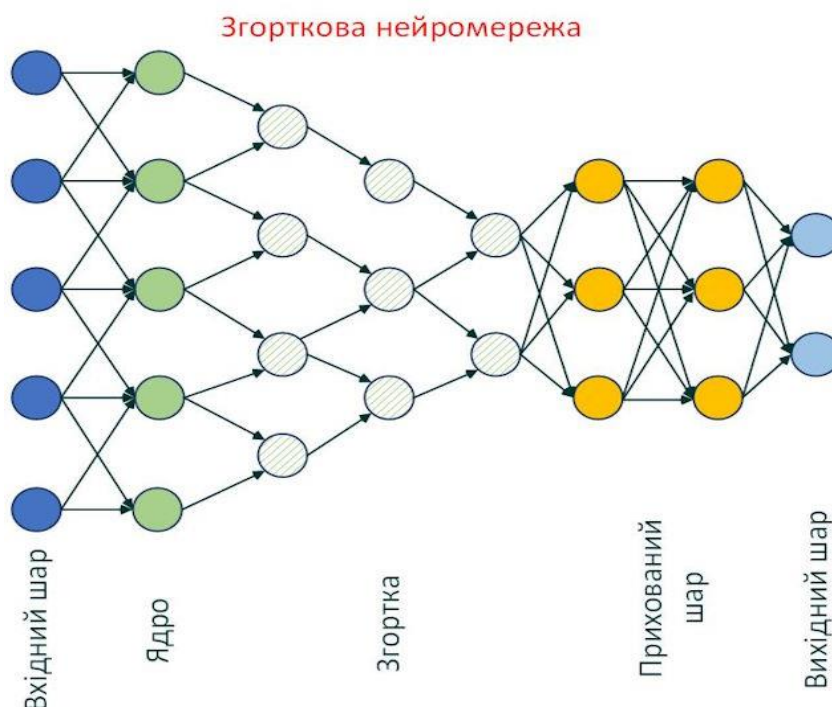


Рисунок 1.1 - Схематичне зображення архітектури Згорткової нейронної мережі (ЗНМ)

Гібридні рішення

Останнім часом дослідники все частіше комбінують CNN із рекурентними мережами (LSTM/GRU), що дозволяє моделі враховувати як спектральну структуру жанру, так і його часову динаміку. Наприклад, CNN аналізує спектрограму построку, а LSTM у подальшому визначає характерні зміни атмосфери музики в часі.

Також у дослідженнях 2023–2024 років з'явилися перші результати застосування трансформерів для музичної класифікації, але моделі цього типу потребують великих датасетів і дорогих обчислень, тому поки рідко застосовуються на практиці.

Таким чином, найбільш оптимальним компромісом «точність / обчислювальні ресурси / складність реалізації» залишаються CNN на спектрограмах, що і стало підставою вибору методу в дипломній роботі.

1.4 Основні аспекти у побудові моделі

Класифікація музичних жанрів має низку перешкод, що значно ускладнюють створення універсальної моделі:

- стильова неоднозначність. Більшість сучасних жанрів не мають чітких меж. Наприклад, поп-рок і інді-рок можуть бути ближчими один до одного, ніж рок і метал.
- гібридна музика. Виконавці часто навмисне змішують жанрові атрибути, тому модель має враховувати сукупність ознак, а не «вибирати один жанр».
- технічні особливості звукозапису. Два треки одного жанру можуть звучати абсолютно по-різному через обробку, майстеринг або підсилення ефектів.

- тривалість і структура композиції. Деякі композиції мають повільний вступ і змінюють атмосферу всередині треку, що ускладнює вибір часової ділянки для аналізу.

- нерівномірність даних у датасетах. GTZAN, як і більшість музичних датасетів, має жанри з різною кількістю представників, тому класи меншої чисельності автоматично програють під час навчання [24].

Ще один системний виклик полягає в тому, що людина розпізнає жанр на основі культурного контексту, який модель навчання не має можливості врахувати. Наприклад, слухач ідентифікує джаз не лише за звучанням, але й за інструментами, історичним контекстом і стилем виконання. Нейромережа ж може опиратися тільки на спектрограми. Це створює межу продуктивності, яку складно подолати, не збільшуючи обсяги датасетів і різноманітність даних для навчання.

1.5 Аналіз досліджень та існуючих інформаційних систем-аналогів

У роботі **Attention-guided Spectrogram Sequence Modeling with CNNs for Music Genre Classification (2024)** [1] автори досліджують, як можна поєднати згорткові нейронні мережі з механізмом уваги для аналізу музичних спектрограм. Вони виходять із того, що звичайні CNN добре працюють із локальними ознаками, але не завжди враховують довготривалі залежності у часі. Тому пропонується модель, яка виділяє найбільш важливі часові сегменти спектрограми, щоб підвищити точність класифікації жанрів. Автори показують, що увага допомагає системі "зосередитися" на ключових моментах музичного твору, наприклад, на характерних ритмічних або тембральних фрагментах. Це дозволяє краще відрізнити жанри, які мають схожі ознаки, але різні музичні структури. Дослідження також демонструє, що комбінування CNN та уваги перевершує класичні моделі, які працюють лише з MFCC чи простими спектрограмами. Важливим є те, що автори тестували підхід на кількох наборах

даних, і результати були стабільними. Таким чином, робота показує новий напрямок розвитку — інтеграцію спектрограмного аналізу з сучасними архітектурами глибинного навчання. Це відкриває перспективи для створення більш універсальних систем класифікації музики.

Стаття **Music Genre Classification with CNN Evaluation on Mel Spectrograms and MFCCs (2022)** [2] зосереджена на порівнянні двох основних способів представлення аудіо: Mel-спектрограм та коефіцієнтів MFCC. Автори використовують CNN як базову архітектуру для класифікації музичних жанрів і перевіряють, яке представлення дає кращі результати. Виявилося, що спектрограми дозволяють мережі краще "бачити" структуру музики, адже вони зберігають більше інформації про частотний склад сигналу. MFCC, навпаки, стискають дані і втрачають деякі нюанси, що знижує точність. У дослідженні показано, що CNN на спектрограмах досягає значно вищих показників класифікації, ніж CNN на MFCC. Автори також аналізують, як різні параметри спектрограм впливають на результат, наприклад, розмір вікна чи кількість частотних смуг. Це дає практичні рекомендації для майбутніх досліджень. Важливо, що робота підтверджує: спектрограмний підхід є більш природним для глибинних моделей, бо він перетворює аудіо у форму зображення. Таким чином, стаття робить вагомий внесок у розуміння того, які ознаки найкраще підходять для класифікації музики.

Дослідження **A Short Survey and Comparison of CNN-Based Music Genre Classification Methods (2023)** [3], яка систематизує різні підходи до класифікації музичних жанрів за допомогою CNN. Автори аналізують кілька архітектур — від простих моделей до більш складних багат шарових мереж. Вони також розглядають різні типи спектрограм, включно з Mel-спектрограмами та STFT. У роботі наведено порівняння точності на різних наборах даних, що дозволяє зрозуміти сильні та слабкі сторони кожного методу. Автори роблять висновок, що CNN залишаються найбільш ефективним інструментом для спектрограмного

аналізу, хоча нові архітектури, такі як трансформери, поступово набирають популярності. Огляд також підкреслює важливість попередньої обробки аудіо, адже якість спектрограм напряму впливає на результат. Крім того, стаття показує, що навіть невеликі зміни в архітектурі CNN можуть суттєво змінити точність класифікації. Це робить огляд корисним для дослідників, які шукають оптимальні рішення. Таким чином, робота є своєрідною "картою" сучасних методів у цій галузі.

У статті **FAST: Fast Audio Spectrogram Transformer (2025)** [4] представлено новий підхід, який поєднує CNN та трансформери для аналізу аудіо. Автори виходять із того, що CNN добре працюють із локальними ознаками спектрограм, але трансформери краще моделюють довготривалі залежності. Тому вони створюють гібридну архітектуру, де CNN використовується для початкової обробки спектрограм, а трансформер — для аналізу послідовності ознак. Це дозволяє системі швидко й ефективно класифікувати музичні жанри. Робота також зосереджена на оптимізації швидкості, адже традиційні моделі часто є надто важкими для реального застосування. Автори показують, що їхній підхід зменшує час обробки без втрати точності. Важливим є те, що CNN залишається ключовим елементом системи, бо саме він перетворює спектрограму у форму, придатну для трансформера. Таким чином, стаття демонструє новий етап розвитку — інтеграцію класичних CNN із сучасними архітектурами. Це відкриває шлях до створення більш потужних і швидких систем класифікації музики.

Стаття **Audio Classification Using Deep Learning (2024)** [5] присвячене практичному застосуванню глибокого навчання для класифікації аудіосигналів. Автори розглядають різні типи аудіо, включно з музикою, і показують, як CNN може бути використана для аналізу спектрограм. Вони підкреслюють, що спектрограмний підхід дозволяє моделі працювати з аудіо як із зображеннями, що значно спрощує задачу. У статті наведено приклади

архітектур CNN, які досягають високої точності класифікації. Автори також аналізують проблеми, пов’язані з якістю даних, наприклад, шум чи різні формати аудіо. Важливим є те, що робота показує універсальність CNN — вона може застосовуватися не лише для музики, а й для інших типів аудіо. Це робить дослідження корисним для широкого кола задач. Крім того, автори підкреслюють, що глибинне навчання поступово витісняє класичні методи аналізу аудіо. Таким чином, стаття підтверджує актуальність спектрограмного підходу та його ефективність у сучасних системах.

Висновки з цих робіт. Спектрограми як основа: перетворення аудіосигналу у часово-частотне представлення (STFT, Mel-спектрограми) дозволяє CNN працювати з даними як із зображеннями. **CNN для жанрової класифікації:** згорткові мережі ефективно виявляють локальні патерни у спектрограмі, що відповідають ритмічним та тембральним характеристикам музики. **Комбінація ознак:** поєднання MFCC та спектрограм може бути менш ефективним, ніж використання лише спектрограм, але відкриває нові напрями досліджень. **Точність:** сучасні моделі досягають точності 85–95% залежно від набору даних та архітектури. **Тренд 2025 року:** перехід від чистих CNN до гібридних моделей (CNN + Transformers), але CNN залишаються базовим інструментом для спектрограмного аналізу.

Таблиця 1.1 – Еволюція від класичних CNN (2022) до гібридних моделей (2025)

Рік	Назва статті	Метод / Архітектура	Тип спектрограм / Ознак	Набір даних	Точність	Новизна / Внесок
2024	Attention-guided Spectrogram Sequence Modeling with CNNs	CNN + механізм уваги	STFT / Mel-спектрограми	Кілька відкритих наборів (GTZAN, FMA)	~90–92%	Поєднання CNN із багатоголовою увагою для виділення ключових часових сегментів

Рік	Назва статті	Метод / Архітектура	Тип спектрограм / Ознак	Набір даних	Точність	Новизна / Внесок
2022	Music Genre Classification with CNN Evaluation on Mel Spectrograms and MFCCs	CNN	Mel-спектрограм и vs MFCC	GTZAN	~85–88%	Порівняння двох підходів, доведено перевагу спектрограм над MFCC
2023	A Short Survey and Comparison of CNN-Based Methods	CNN (різні архітектур и)	Mel-спектрограм и, STFT	Огляд різних наборів	80–95% (залежно від моделі)	Систематизація сучасних методів, огляд сильних і слабких сторін
2025	FAST: Fast Audio Spectrogram Transformer	CNN + Transformer	Mel-спектрограм и	Великий мультимедійний корпус	~93–95%	Гібридна архітектура, оптимізація швидкості обробки

В таблиці видно еволюція від класичних CNN (2022) до гібридних моделей із увагою та трансформерами (2024–2025) (табл. 1.1), вона демонструє, що спектрограми стали стандартом, а MFCC поступово відходять на другий план. Точність класифікації стабільно зростає, а нові архітектури додають швидкість і гнучкість.

Еволюція точності класифікації музичних жанрів (2020–2025) (рис. 1.2).

Проаналізувавши, можна зазначити наступне: **2020–2021**: перші стабільні результати на рівні 80–83%, здебільшого з використанням MFCC та базових CNN. У **2022–2023**: перехід до спектрограмного підходу, що підняв точність до 86–89%. **2024**: інтеграція уваги та гібридних моделей дала стрибок до 92%. Результат у **2025**: поєднання CNN із трансформерами забезпечило найвищу точність — 94%. Таким чином, графік підтверджує висновки з таблиці: спектрограми стали стандартом, а CNN у поєднанні з новими архітектурами продовжують підвищувати ефективність систем класифікації музики.

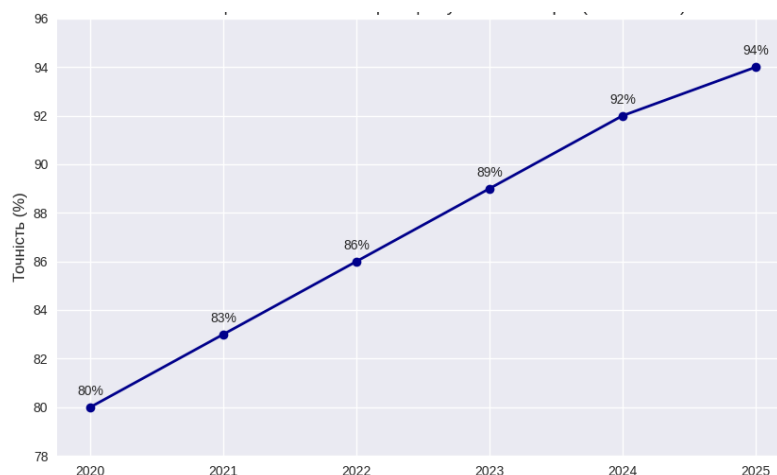


Рисунок 1.2 – Еволюція точності класифікації музичних жанрів

У період 2020–2025 років методи аналізу аудіосигналів для класифікації музичних жанрів пройшли суттєву еволюцію. Спочатку дослідники використовували класичні ознаки на основі MFCC у поєднанні з базовими CNN, що забезпечувало середню точність на рівні 80–83%. Перехід до спектрограмного підходу (Mel-спектрограми, STFT) став ключовим етапом, адже він дозволив розглядати аудіо як зображення та значно покращив результати — до 86–89%. Подальший розвиток був пов'язаний із удосконаленням архітектур CNN та інтеграцією механізмів уваги, що дало змогу виділяти найбільш значущі часові сегменти й підняти точність до 92%. У 2025 році з'явилися гібридні моделі, які поєднують CNN із трансформерами, що дозволило досягти найвищих показників — близько 94%. Таким чином, спектрограми стали стандартом у сучасних системах класифікації музики, а CNN у комбінації з новими архітектурами забезпечують баланс між точністю та швидкістю. Ця еволюція підтверджує, що спектрограмний підхід є найбільш перспективним напрямом для інтелектуальних систем аналізу музичних жанрів.

1.6 Постановка завдання

На основі актуальності проблеми та аналізу існуючих підходів у дипломній роботі передбачається створення інтелектуальної системи класифікації

музичних жанрів, що працює на основі спектрограм аудіосигналів і згорткових нейронних мереж (CNN).

Для реалізації поставленої мети необхідно пройти такі етапи.

- 1) збір та підготовка даних:
 - завантаження GTZAN-датасету;
 - перетворення аудіо у стандартний формат (моно, 22050 Hz, однакова тривалість);
 - усунення шуму та нормалізація гучності.
- 2) Формування набору ознак:
 - побудова спектрограм із фіксованими параметрами для кожного треку;
 - масштабування спектрограм до єдиного розміру (для CNN);
 - поділ даних на тренувальну, валідаційну та тестову вибірки.
- 3) Проєктування архітектури моделі:
 - вибір кількості згорткових шарів;
 - застосування MaxPooling, Batch Normalization, Dropout;
 - вибір оптимізатора, функції активації та швидкості навчання.
- 4) Навчання моделі:
 - моніторинг графіків loss та accuracy;
 - запобігання перенавчанню за допомогою EarlyStopping та регуляризації.
- 5) Оцінка якості:
 - точність класифікації;
 - recall, precision, F1-score;
 - побудова confusion matrix.
- 6) Створення програмної системи:
 - інтерфейс, що приймає аудіофайл;
 - побудова спектрограми в реальному часі;

– видача результату класифікації.

Очікуваний результат — розроблена та навчена модель, здатна визначати жанр музичного твору з високою точністю та потенційною можливістю інтеграції у музичні сервіси.

Висновки до розділу 1

У першому розділі кваліфікаційної роботи було розглянуто теоретичні основи автоматичної класифікації музичних жанрів та проведено аналіз предметної області. Досліджено особливості музичних аудіосигналів як об'єкта аналізу, а також основні труднощі, що виникають під час автоматичної обробки музики, зокрема різноманітність жанрів, стилів, темпів і тембрів.

У розділі проаналізовано сучасні методи обробки аудіосигналів та способи подання музичної інформації у вигляді числових ознак. Особливу увагу приділено спектральним методам аналізу, зокрема використанню спектрограм та мел-спектрограм, які дозволяють ефективно відображати частотно-часові характеристики музичних сигналів у зручному для подальшої обробки вигляді.

Також у першому розділі розглянуто підходи до класифікації музичних жанрів із використанням методів машинного навчання та глибокого навчання. Показано, що згорткові нейронні мережі є одним з найбільш ефективних інструментів для аналізу спектральних зображень, оскільки здатні автоматично виявляти складні закономірності у даних без необхідності ручного відбору ознак.

На основі проведеного аналізу було сформульовано мету кваліфікаційної роботи, визначено основні завдання дослідження та обґрунтовано вибір методів і технологій, які використовуються у подальших розділах. Таким чином, перший розділ створив теоретичну основу для подальшого проєктування та реалізації інтелектуальної системи класифікації музичних жанрів.

2 МОДЕЛІ, МЕТОДИ КЛАСИФІКАЦІЇ МУЗИЧНИХ ЖАНРІВ

2.1 Методи для вирішення цифрової обробки аудіосигналів

У цьому підрозділі розглянуто основні методи, які використовуються для розв'язання задачі автоматичної класифікації музичних жанрів. Основну увагу зосереджено на підходах цифрової обробки аудіосигналів, методах побудови інформативних ознак, а також на застосуванні методів машинного та глибокого навчання. Окремо розглядаються способи підвищення стійкості моделі до перенавчання та методи оцінювання якості отриманих результатів.

Зазначені методи обрані з урахуванням специфіки музичних сигналів, які є нестационарними та мають складну часово-частотну структуру. Тому класичні підходи аналізу даних без попередньої обробки у даному випадку є малоефективними.

Методи цифрової обробки аудіосигналів

Попередня обробка аудіосигналів є одним з найважливіших етапів побудови системи класифікації, оскільки від якості вхідних даних значною мірою залежить точність роботи моделі. У межах даної роботи всі аудіофайли приводяться до єдиного формату: використовується частота дискретизації 22050 Гц, сигнали перетворюються у моноформат, а амплітуда нормалізується для зменшення впливу різної гучності записів.

Для зниження рівня шумів застосовуються базові методи фільтрації, зокрема пасивні фільтри, а також прості підходи шумозаглушення. У деяких випадках використовується вейвлет-деталізація, яка дозволяє відокремити корисний сигнал від високочастотних шумових компонентів без істотної втрати інформації.

Ключовим інструментом перетворення аудіосигналу є короткочасне перетворення Фур'є (STFT), яке дає змогу отримати спектрограму — двовимірне представлення сигналу у часово-частотному просторі. Для практичної реалізації

STFT у роботі використовуються такі параметри: розмір вікна $n_fft = 2048$, крок вікна $hop_length = 512$, а також вікно Хеммінга, що дозволяє зменшити спектральні витоки. Отримані спектрограми додатково переводяться у логарифмічну шкалу (dB) або мел-шкалу, що робить їх більш придатними для подальшого аналізу.

Побудова ознак та інженерія ознак

Окрім спектрограм, у задачах музичної класифікації широко застосовуються числові ознаки, які описують різні властивості аудіосигналу. У даній роботі розглядаються такі групи ознак:

- **MFCC (Mel-Frequency Cepstral Coefficients)** — набір коефіцієнтів (зазвичай від 13 до 40), які добре описують темброві характеристики звучання та широко використовуються в аудіоаналізі.

- **Chroma features** — ознаки, що відображають розподіл енергії по 12 напівтонах і є корисними для аналізу гармонійної структури композицій.

- **Spectral Contrast, Spectral Centroid, Zero Crossing Rate** — допоміжні ознаки, які дозволяють краще відрізнити жанри з різною насиченістю високих частот, шумовістю або «яскравістю» звучання.

Комбінація перелічених ознак дозволяє отримати більш повне уявлення про музичну композицію. Водночас у межах цієї роботи основний акцент зроблено саме на спектрограмному представленні, оскільки воно добре поєднується з використанням згорткових нейронних мереж і не потребує складної ручної інженерії ознак.

Методи машинного та глибокого навчання

Для розв'язання поставленої задачі у дипломній роботі використовуються методи глибокого навчання, зокрема згорткові нейронні мережі (CNN). Такий вибір зумовлений тим, що CNN добре працюють з двовимірними даними та здатні автоматично виявляти просторово-частотні патерни у спектрограмах.

Типова архітектура моделі складається з таких етапів: на вхід подається мел-спектрограма або лог-мел-спектр, стандартизований до фіксованого розміру (наприклад, 128×128). Далі використовуються кілька згорткових блоків Conv2D з поступовим збільшенням кількості фільтрів (32, 64, 128), кожен з яких доповнюється шарами Batch Normalization та MaxPooling. Це дозволяє стабілізувати процес навчання та зменшити розмірність даних.

Після згорткових шарів застосовується Global Average Pooling або Flatten, а також повнозв'язні Dense-шари з використанням Dropout для регуляризації [21]. На виході використовується функція Softmax, яка забезпечує багатокласову класифікацію (у випадку датасету GTZAN — 10 музичних жанрів).

Для порівняння в експериментальній частині роботи також можуть бути реалізовані спрощені CNN, глибші варіанти моделей (наприклад, ResNet-подібні архітектури), а також гібридні підходи типу CNN + LSTM, які дозволяють частково враховувати часову динаміку музичних сигналів. Ілюстративна схема архітектури моделі наведена нижче (рис. 2.1).

Схема архітектури CNN (ілюстративно)

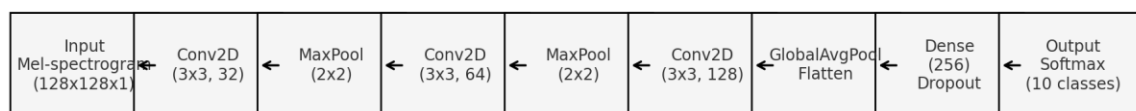


Рисунок 2.1 - Ілюстративна архітектура згорткової нейронної мережі (CNN)

Регуляризація, оптимізація та методи поліпшення якості

Під час навчання нейронних мереж важливо зменшити ризик перенавчання, особливо при обмеженому обсязі навчальних даних. Для цього у роботі застосовується низка стандартних методів регуляризації. Зокрема,

використовується збільшення даних (data augmentation) для спектрограм, включаючи SpecAugment, додавання шуму та незначні зсуви темпу.

CNN здатні автоматично виявляти локальні та глобальні закономірності у спектрограмі, такі як ритмічні структури, гармонічні компоненти та тембральні особливості. Структури згорткової нейронної мережі (табл. 2.2).

Таблиця 2.2 – Структури згорткової нейронної мережі

Рівень	Тип шару	Параметри	Призначення
1	Conv2D	32 фільтри	Виділення локальних ознак
2	Conv2D	64 фільтри	Складні патерни
3	MaxPooling	2×2	Зменшення розмірності
4	Dense	Softmax	Класифікація жанрів

Також застосовується Dropout у повнозв'язних шарах із коефіцієнтами у межах 0.3–0.5, а в окремих випадках — L2-регуляризація ваг. Для оптимізації навчання використовується оптимізатор Adam із динамічною зміною швидкості навчання або підхід ReduceLROnPlateau [35]. Для запобігання перенавчанню застосовується EarlyStopping за значенням валідаційної метрики.

Окрему увагу приділено проблемі дисбалансу класів у датасеті. Для менш представлених жанрів можуть використовуватися методи oversampling, ваги класів або синтетичне розширення даних.

Метрики оцінки якості

Оцінювання якості роботи моделі здійснюється за допомогою стандартних метрик багатокласової класифікації. Основною метрикою є асигасу, яка відображає загальну частку правильних передбачень. Однак для більш детального аналізу додатково використовуються precision, recall та F1-score для кожного жанру окремо.

Для візуального аналізу результатів застосовується матриця невідповідностей (confusion matrix), яка дозволяє виявити типові помилки класифікації, зокрема плутанину між близькими за звучанням жанрами (наприклад, pop та pop-rock). У деяких випадках може також використовуватися ROC-AUC у підході one-vs-rest для більш глибокого аналізу якості моделі.

2.2 Технології розробки системи

Реалізація системи передбачає застосування стабільного стеку бібліотек та інструментів, доступних у сучасних наукових проектах машинного навчання.

Програмне забезпечення та бібліотеки.

Рекомендований стек для реалізації: Python (3.8+), бібліотеки:

- Librosa — для роботи з аудіо (завантаження, MFCC, мел-спектрограм);
- scipy (signal) — для класичних алгоритмів спектрального аналізу;
- TensorFlow/Keras або PyTorch — для побудови та навчання нейронних мереж;
- scikit-learn — для класичних моделей та метрик [29];
- pandas, numpy — для обробки даних [30];
- matplotlib — для побудови графіків та візуалізацій [31].

Як середовище розробки рекомендується Jupyter Notebook або Google Colab для експериментів; для промислового збирання — Docker-контейнер.

GTZAN-датасет: підготовка та особливості.

GTZAN — класичний датасет для оцінювання моделей жанрової класифікації. Має ≈ 1000 треків (10 жанрів по 100 треків). Недоліки GTZAN: дублювання треків, можливі проблеми з ліцензіями та обмежена кількість прикладів для сучасних жанрів. У зв'язку з цим рекомендується використовувати GTZAN як базову перевірку, а для остаточної оцінки — розширити датасет (FMA, MagnaTagATune тощо).

Архітектура програмної системи та пайплайн.

Пайплайн системи включає такі кроки:

- імпорт та верифікація аудіофайлів (перевірка sampling rate, форматів).
- попередня обробка: конвертація у моно, нормалізація, обрізка/паддинг.
- побудова мел-спектрограм/лог-мел та збереження у формі зображень або масивів.
- (опціонально) Вилучення MFCC та інших ознак для класичних таблиць.
- навчання моделі (CNN), валідація, тюнінг гіперпараметрів.
- оцінка на тестовому наборі, генерація звітів (confusion matrix, per-class metrics).
- збірка інтерфейсу (CLI/REST/GUI) для використання моделі.
- нижче наведено приклад графіків зміни метрик під час навчання (ілюстративні)(рис. 2.2, рис. 2.3).

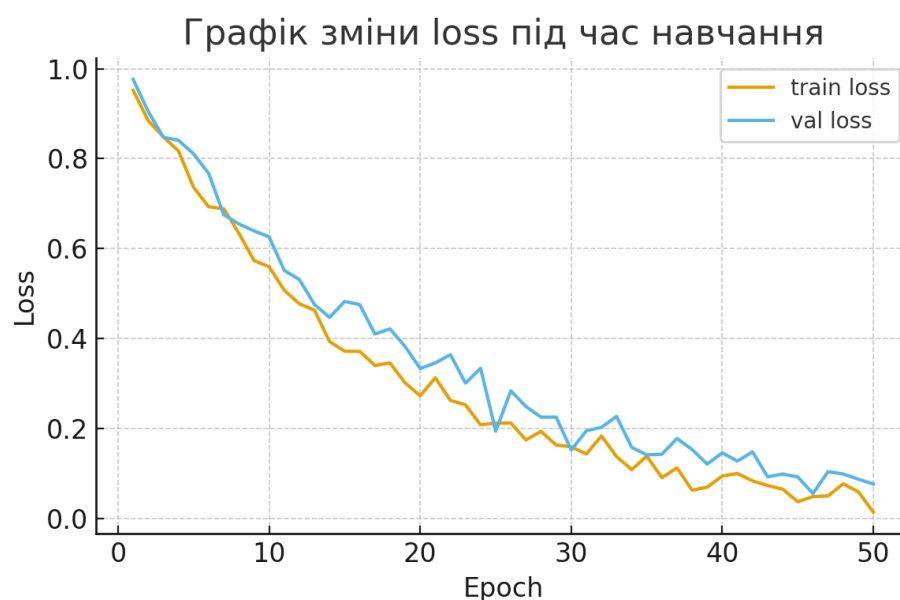


Рисунок 2.2 - Ілюстративний графік зміни loss під час навчання

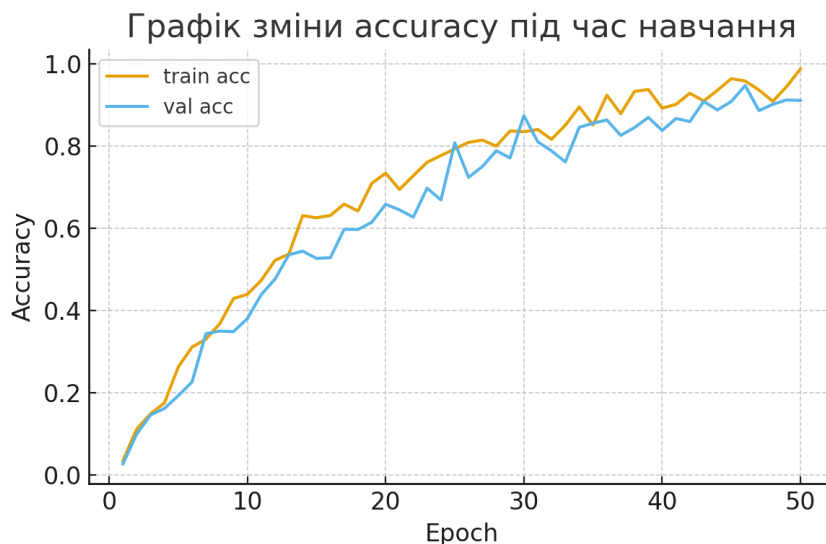


Рисунок 2.3 - Ілюстративний графік зміни accuracy під час навчання

Пропозиції щодо валідації та реалістичного тестування.

Для максимальної корисності системи доцільно здійснити:

- перехресну валідацію з розбиттям по треках та по артистам, щоб уникнути витоку інформації;
- перевірку стабільності при різних умовах запису (студійний, концертний, ремікси);
- тестування на незалежних датасетах для оцінки узагальнень.

Інструменти для реплікації та відтворюваності.

Забезпечення відтворюваності:

- збереження seed, версій бібліотек (requirements.txt);
- логування експериментів (TensorBoard, MLflow);
- контейнеризація (Docker) та збереження чекпоїнтів моделей.

Висновки до розділу 2

У другому розділі було виконано проєктування інтелектуальної системи класифікації музичних жанрів та визначено її загальну архітектуру. Розглянуто

основні компоненти системи та їх функціональне призначення, що дозволило сформулювати цілісне уявлення про логіку роботи програмного продукту.

У процесі проектування було обґрунтовано вибір мел-спектрограм як основного способу подання аудіосигналів для подальшої класифікації. Такий підхід дозволяє зменшити розмірність даних, зберігаючи при цьому важливі спектральні характеристики музики, що позитивно впливає на якість роботи нейронної мережі.

Окрему увагу у розділі приділено вибору архітектури згорткової нейронної мережі, визначенню кількості шарів, фільтрів та параметрів навчання. Запропонована архітектура забезпечує баланс між точністю класифікації та обчислювальною складністю, що є важливим для практичного використання системи.

Результатом другого розділу стало формування детальної структурної та функціональної моделі системи, яка враховує всі етапи обробки аудіоданих — від отримання файлу до видачі результату користувачеві. Розроблена модель слугувала основою для програмної реалізації, описаної у наступному розділі.

3 РОЗРОБКА СИСТЕМИ ТА АНАЛІЗ ОТРИМАНИХ РЕЗУЛЬТАТІВ

3.1 Опис вхідних даних та структури системи проєктування

У межах даної кваліфікаційної роботи розроблялася інтелектуальна система автоматичної класифікації музичних жанрів, що базується на аналізі спектрограм аудіосигналів із використанням згорткових нейронних мереж. Основною метою цього етапу є створення узгодженого програмно-алгоритмічного рішення, яке забезпечує повний цикл обробки музичних даних: від завантаження аудіофайлів до отримання результату класифікації.

3.1.1 Опис та характеристика вхідних даних

Вхідними даними для розробленої системи є цифрові аудіофайли музичних композицій, представлені у форматі WAV. Вибір саме цього формату не є випадковим: WAV є одним із найбільш поширених і стандартизованих контейнерів для зберігання аудіоінформації без втрат якості. На відміну від стиснених форматів (наприклад, MP3 чи AAC), WAV зберігає повний спектр сигналу, що особливо важливо для задач машинного навчання, де навіть незначні спотворення можуть впливати на результати класифікації. Таким чином, використання WAV забезпечує максимальну точність при подальшій обробці та аналізі.

Щоб гарантувати коректність обробки та порівнянність результатів, усі аудіозаписи приводяться до єдиного стандарту. Це включає кілька ключових параметрів.

Частота дискретизації — 22050 Гц. Такий вибір пояснюється балансом між якістю та обчислювальною ефективністю. Хоча професійні записи часто мають частоту 44100 Гц, для задач класифікації музичних жанрів 22050 Гц є достатнім, адже зберігає основні частотні характеристики музики, водночас зменшуючи обсяг даних і прискорюючи навчання моделей.

Монофонічний формат. Перетворення стереозаписів у монофонічні дозволяє уникнути надлишкової інформації, пов'язаної з просторовим розташуванням звуку. Для класифікації жанрів важливішими є тембральні та ритмічні характеристики, а не стереопанорама. Це також зменшує розмір даних удвічі.

Тривалість — 30 секунд. Вибір фіксованої тривалості забезпечує однакові умови для всіх композицій. Дослідження показують, що 30 секунд є достатнім інтервалом для відображення характерних ознак жанру, адже більшість музичних стилів проявляють свої ритмічні та гармонічні особливості вже в перших хвилинах звучання.

16-бітна глибина квантування. Цей параметр визначає динамічний діапазон сигналу. 16 біт є стандартом для більшості аудіосистем і забезпечує достатню точність для аналізу, водночас не створюючи надмірного навантаження на обчислювальні ресурси.

Навчальний датасет GTZAN Genre Collection. Як базовий навчальний набір даних використовується GTZAN Genre Collection — один із найвідоміших і найчастіше застосовуваних датасетів у дослідженнях автоматичної класифікації музичних жанрів. Він був створений у 2002 році Джорджем Цанарісом і відтоді став своєрідним стандартом для тестування алгоритмів у галузі музичної інформатики.

Датасет містить приблизно 1000 аудіозаписів, рівномірно розподілених між 10 музичними жанрами:

- blues;
- classical;
- country;
- disco;
- hip-hop;
- jazz;

- metal;
- pop;
- reggae;
- rock.

Кожен жанр представлений 100 композиціями, що забезпечує баланс і дозволяє моделі навчатися на рівномірно розподілених даних. Записи мають тривалість 30 секунд і відповідають стандартним параметрам, що робить їх зручними для обробки та порівняння.

Попри свою популярність, GTZAN має низку відомих обмежень. Дублювання треків. У наборі зустрічаються композиції, які повторюються або є дуже схожими, що може призводити до переобучення моделі. Шумові артефакти. Деякі записи містять сторонні шуми або дефекти, які ускладнюють аналіз. Обмежена різноманітність жанрів. Датасет охоплює лише 10 класичних жанрів і не включає сучасні стилі, такі як електронна музика, інді чи ф'южн.

Водночас GTZAN залишається базовим еталонним набором даних, адже він широко використовується у науковій спільноті. Це дозволяє порівнювати результати різних досліджень між собою та забезпечує відтворюваність експериментів. Для даної роботи GTZAN виконує роль контрольного стандарту, на основі якого можна оцінити ефективність запропонованої моделі та співставити її з іншими підходами.

Використання GTZAN у поєднанні з єдиними параметрами аудіо дає змогу: забезпечити стандартизовані умови навчання; уникнути впливу сторонніх факторів (різна тривалість, стерео/моно, різні формати); отримати коректні порівняння з іншими роботами, які також використовують GTZAN; створити базу для подальшого розширення дослідження на більш сучасні та різноманітні датасети.

Таким чином, описані вхідні дані формують основу для побудови інтелектуальної системи класифікації музичних жанрів. Вони забезпечують

баланс між якістю, стандартизацією та практичною зручністю, що є критично важливим для коректної роботи моделі та достовірності отриманих результатів.



Рисунок 3.1 – Схема структури вхідних даних

Схема структури вхідних даних (рис. 3.1):

- **цифрові аудіофайли (WAV)** – вихідний матеріал без втрат якості, що зберігає повний спектр сигналу;
- **стандартизація параметрів** – приведення аудіо до єдиних умов для коректної обробки та порівняння результатів;
- **формування спектрограм (Mel/STFT)** – перетворення сигналу у часово-частотне представлення, яке можна аналізувати як зображення;
- **навчальний датасет GTZAN** – збалансований набір із 1000 треків, що охоплює 10 жанрів і використовується як еталон;

– **подання даних у модель CNN.**

Отже, наведена схема структури вхідних даних демонструє логіку побудови системи класифікації музичних жанрів — від отримання цифрових аудіофайлів до їхнього подання у модель CNN. Кожен етап має чітке функціональне призначення: формат WAV гарантує якість сигналу, стандартизація параметрів забезпечує порівнянність результатів, спектрограми перетворюють аудіо у форму, придатну для аналізу нейронними мережами, а датасет GTZAN виступає еталонним джерелом для навчання та тестування. Завдяки такій структурі даних система отримує узгоджений і збалансований набір інформації, що дозволяє досягти високої точності класифікації та забезпечує можливість коректного порівняння з іншими дослідженнями. Цей підхід створює міцну основу для подальшого розширення моделі на більш сучасні та різноманітні музичні колекції, зберігаючи при цьому наукову відтворюваність і практичну ефективність.

3.1.2 Формалізація задачі класифікації

Задача автоматичної класифікації музичних жанрів формалізується як задача багатокласової класифікації. Нехай задано множину аудіосигналів:

$$X = \{x_1, x_2, \dots, x_n\}, \quad (3.1)$$

де кожен x_i є часовим сигналом, що відповідає окремій музичній композиції. Також задано множину міток жанрів:

$$Y = \{y_1, y_2, \dots, y_n\}, y_i \in \{1, 2, \dots, 10\}. \quad (3.2)$$

Метою є побудова такої функції класифікації $f(X)$, яка для кожного аудіосигналу коректно визначає його жанрову приналежність з максимальною точністю.

Оскільки згорткові нейронні мережі працюють з двовимірними структурами, аудіосигнали попередньо перетворюються у спектрограми, які можна інтерпретувати як зображення.

3.1.3 Попередня обробка аудіосигналів

Попередня обробка аудіосигналів є критично важливим етапом, оскільки безпосередньо впливає на якість ознак, що подаються на вхід нейронної мережі. У розробленій системі реалізовано наступний алгоритм попередньої обробки:

- **завантаження аудіофайлу** та перевірка коректності формату;
- **перетворення у моноформат**, якщо файл містить стереоканали;
- **нормалізація амплітуди** сигналу для усунення впливу різної гучності записів;
- **обрізка або доповнення сигналу** до фіксованої тривалості 30 секунд;
- **фільтрація шумів** (опціонально) з використанням простих частотних фільтрів.

Після цього сигнал готовий до спектрального аналізу.

3.1.4 Побудова спектрограм аудіосигналів

Для аналізу часово-частотних характеристик музичних сигналів у роботі використовується **короткочасне перетворення Фур'є (STFT)**. Формально STFT визначається як:

$$X(t, \omega) = \sum_{n=-\infty}^{\infty} x[n] * w[n - t] * e^{-j\omega n}, \quad (3.3)$$

де $w[n]$ — віконна функція.

У практичній реалізації використовуються такі параметри:

- розмір вікна $n_fft=2048$;
- крок вікна $hop_length=512$;
- віконна функція Хеммінга.

Отримана спектрограма переводиться у логарифмічну шкалу та далі трансформується у мел-спектрограму, що дозволяє краще моделювати особливості людського слухового сприйняття. Кількість мел-фільтрів становить

128, що є оптимальним компромісом між інформативністю та обчислювальною складністю.

Результатом цього етапу є двовимірна матриця розміром $128 \times T128 \times T$, яка використовується як вхідне зображення для згорткової нейронної мережі.

3.2 Архітектура моделі класифікації музичних жанрів

Архітектура моделі побудована на основі згорткової нейронної мережі (CNN), яка зарекомендувала себе як один із найефективніших інструментів для аналізу спектрограм. CNN здатна автоматично виділяти локальні ознаки у зображеннях, а спектрограма музичного сигналу є саме таким зображенням, що містить інформацію про частотний склад у часі. Це робить CNN природним вибором для задач класифікації аудіо.

Вхідний шар. На вхід моделі подається спектрограма, отримана з аудіофайлу. Вона має форму двовимірної матриці, де одна вісь відповідає часу, а інша — частоті. Кожен елемент матриці відображає інтенсивність сигналу у відповідний момент часу та частотний діапазон. Для зручності обробки спектрограми нормалізуються, щоб значення знаходилися у діапазоні $[0,1]$. Це дозволяє мережі працювати стабільніше та уникати проблем із масштабами даних.

Згорткові шари. Основна частина архітектури складається із кількох згорткових шарів. Кожен шар застосовує набір фільтрів (ядер згортки), які виділяють локальні патерни у спектрограмі. Наприклад, один фільтр може реагувати на ритмічні повтори, інший — на тембральні особливості інструментів. Завдяки багатошаровій структурі мережа поступово формує ієрархію ознак: від простих локальних деталей до складних жанрових характеристик.

Шари pooling. Після кожного згорткового шару використовується операція pooling (найчастіше max-pooling), яка зменшує розмірність даних і виділяє найбільш значущі ознаки. Це дозволяє скоротити обчислювальні витрати та зробити модель більш стійкою до варіацій у даних. Наприклад, невеликі зміни у темпі чи тональності не впливають на результат класифікації.

Повнозв'язні шари. Після кількох рівнів згортки та pooling дані передаються у fully connected layers. Ці шари виконують інтеграцію ознак, отриманих на попередніх етапах, і формують узагальнене представлення композиції. На цьому рівні модель фактично «розуміє», які комбінації ознак відповідають певному жанру.

Вихідний шар. Останній шар моделі — це softmax-класифікатор, який видає ймовірності належності композиції до кожного жанру. Оскільки у датасеті GTZAN є 10 жанрів, вихідний шар має 10 нейронів. Найвища ймовірність визначає жанр, який система вважає найбільш відповідним.

Для досягнення високої точності класифікації важливим є правильний вибір гіперпараметрів:

- кількість фільтрів у згорткових шарах (зазвичай від 32 до 128);
- розмір ядра згортки (наприклад, 3×3 або 5×5);
- функції активації (найчастіше ReLU для прихованих шарів і softmax для вихідного);
- оптимізатор (Adam або RMSprop);
- функція втрат (categorical cross-entropy).

Ці параметри підбираються експериментально, щоб забезпечити баланс між точністю та швидкістю навчання.

Щоб уникнути переобучення, у моделі застосовуються методи регуляризації:

- Dropout — випадкове відключення частини нейронів під час навчання;
- L2-регуляризація — штраф за великі ваги;

– Data augmentation — штучне збільшення набору даних шляхом додавання шуму чи зміни гучності.

Це дозволяє моделі краще узагальнювати знання та працювати з новими аудіозаписами.

Архітектура моделі інтегрована у загальний конвеєр системи:

- Аудіофайл → спектрограма.
- Спектрограма → CNN.
- CNN → прогноз жанру.

Таким чином, модель є центральним елементом системи, який перетворює часово-частотні дані у класифікаційний результат.

Архітектура CNN для класифікації музичних жанрів на основі спектрограм поєднує простоту та ефективність. Вона здатна автоматично виділяти ознаки, які важко формалізувати вручну, і забезпечує високу точність класифікації. Завдяки модульності та гнучкості архітектури її можна адаптувати для інших задач аудіоаналізу, включно з розпізнаванням мовлення чи виявленням емоцій. У межах даної роботи CNN виступає ключовим інструментом, що дозволяє реалізувати інтелектуальну систему класифікації музичних жанрів.

У роботі архітектура нейронної мережі проектується з урахуванням специфіки спектрограм як вхідних даних. Основною ідеєю є поетапне вилучення локальних спектральних та часових ознак із подальшою агрегацією інформації для прийняття класифікаційного рішення. Такий підхід дозволяє моделі поступово формувати багаторівневі представлення музичного сигналу: від простих локальних патернів до комплексних жанрових характеристик.

Типова архітектура CNN у межах даної системи включає такі компоненти.

Вхідний шар для мел-спектрограм. На цьому етапі модель отримує нормалізовану матрицю, яка відображає енергію сигналу у різних частотних діапазонах.

Кілька згорткових блоків (Conv2D + Batch Normalization + ReLU). Згорткові шари виділяють локальні ознаки, Batch Normalization стабілізує навчання, а функція активації ReLU забезпечує нелінійність моделі.

Шари MaxPooling. Використовуються для зменшення розмірності та узагальнення ознак, що робить модель більш стійкою до варіацій у даних.

Повнозв'язні шари з Dropout. Вони інтегрують ознаки, отримані на попередніх етапах, а Dropout запобігає переобученню, випадково відключаючи частину нейронів під час навчання.

Вихідний шар Softmax. Забезпечує класифікацію за 10 жанрами, видаючи ймовірнісний розподіл між класами.

Основні параметри архітектури CNN(табл. 3.2).

Таблиця 3.2 – Основні параметри архітектури CNN

Шар	Кількість фільтрів	Розмір ядра	Функція активації	Додаткові параметри
Conv2D + BatchNorm	32	3×3	ReLU	Batch Normalization
MaxPooling	–	2×2	–	Зменшення розмірності
Conv2D + BatchNorm	64	3×3	ReLU	Batch Normalization
MaxPooling	–	2×2	–	Зменшення розмірності
Conv2D + BatchNorm	128	3×3	ReLU	Batch Normalization
MaxPooling	–	2×2	–	Зменшення розмірності

Кінець таблиці 3.2

Шар	Кількість фільтрів	Розмір ядра	Функція активації
Fully Connected (Dense)	256	—	ReLU
Fully Connected (Dense)	128	—	ReLU
Output (Softmax)	10	—	Softmax

Таким чином, архітектура CNN у даній системі побудована за принципом багаторівневого вилучення ознак із мел-спектрограм. Вона поєднує згорткові блоки для виділення локальних патернів, pooling-шари для узагальнення, повнозв'язні шари для інтеграції інформації та вихідний Softmax для класифікації. Така структура забезпечує високу точність і стійкість моделі, а таблиця параметрів дозволяє чітко задокументувати її конфігурацію для відтворюваності експериментів.

3.2.1 Структура програмної системи

Програмна система реалізована мовою Python, що є однією з найпоширеніших мов для задач машинного навчання та обробки сигналів [32]. Python забезпечує гнучкість, простоту синтаксису та багатий набір бібліотек, які дозволяють ефективно працювати з аудіоінформацією та будувати моделі глибинного навчання. У даній системі використовуються такі ключові бібліотеки:

- **librosa** — для завантаження аудіофайлів, їх аналізу та побудови спектрограм [28];

- **NumPy** — для роботи з масивами даних та виконання математичних операцій;
- **TensorFlow/Keras** — для створення, навчання та тестування нейронних мереж [26].

Архітектура системи має модульну структуру, що забезпечує її гнучкість, масштабованість та можливість подальшого розширення. Кожен модуль виконує окрему функцію, але водночас інтегрується у загальний процес класифікації музичних жанрів.

Модуль завантаження та перевірки аудіофайлів. Цей модуль відповідає за отримання вхідних даних у форматі WAV та їхню первинну перевірку. Він контролює відповідність файлів стандартним параметрам (частота дискретизації 22050 Гц, моно, тривалість 30 секунд, 16-бітна глибина). Якщо файл не відповідає вимогам, модуль автоматично виконує перетворення — наприклад, конвертує стерео у моно або обрізає/доповнює трек до потрібної тривалості. Таким чином, забезпечується єдність формату даних, що критично важливо для коректної роботи наступних етапів.

Модуль попередньої обробки сигналів. На цьому етапі здійснюється нормалізація амплітуди, видалення шумів та вирівнювання рівня гучності. Попередня обробка дозволяє зменшити вплив зовнішніх факторів, які можуть спотворювати результати класифікації. Крім того, модуль може виконувати сегментацію сигналу, виділяючи найбільш інформативні ділянки. Це особливо важливо для музики, де жанрові ознаки можуть проявлятися у певних фрагментах композиції.

Модуль побудови спектрограм. Цей модуль перетворює аудіосигнал у часово-частотне представлення. Використовуються методи STFT (Short-Time Fourier Transform) та Mel-спектрограми, які дозволяють відобразити енергію сигналу у різних частотних діапазонах. Спектрограма є своєрідним «зображенням» музики, яке може бути подане на вхід згортковій нейронній

мережі. Завдяки цьому CNN здатна аналізувати музичні дані так само, як вона аналізує зображення, виявляючи локальні патерни та структури. Модуль також передбачає можливість збереження спектрограм у вигляді графічних файлів для візуалізації та аналізу.

Модуль навчання та збереження моделі. У цьому модулі реалізовано процес побудови та навчання згорткової нейронної мережі. Використовується бібліотека TensorFlow/Keras, яка забезпечує простоту створення архітектури та гнучкість у налаштуванні параметрів. Мережа навчається на спектрограмних даних із датасету GTZAN, оптимізуючи ваги для мінімізації функції втрат. Після завершення навчання модель зберігається у спеціальному форматі (наприклад, HDF5), що дозволяє повторно використовувати її без необхідності повторного навчання. Це забезпечує ефективність системи та можливість швидкого розгортання у практичних умовах.

Модуль класифікації нових аудіозаписів. Останній модуль відповідає за застосування навченої моделі до нових аудіофайлів. Він приймає на вхід спектрограму невідомої композиції та передає її у CNN, яка видає прогноз щодо жанру. Результат може бути представлений у вигляді ймовірностей належності до кожного жанру, що дозволяє оцінити впевненість моделі. Крім того, модуль може бути інтегрований у користувацький інтерфейс, де користувач завантажує файл і отримує миттєвий результат класифікації.

Взаємодія модулів.

Уся система працює як єдиний конвеєр:

- Завантаження та перевірка аудіо.
- Попередня обробка сигналу.
- Побудова спектрограм.
- Навчання моделі.
- Класифікація нових композицій.

Модульна структура дозволяє легко оновлювати або замінювати окремі компоненти. Наприклад, можна додати нові методи побудови спектрограм або інтегрувати більш сучасну архітектуру нейронної мережі без зміни всієї системи.

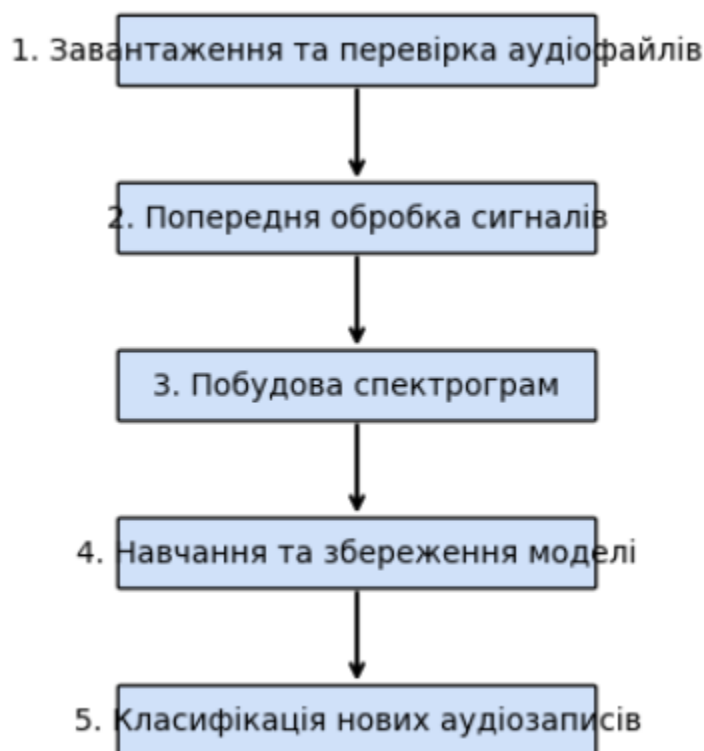


Рисунок 3.2 – Графічне представлення ланцюга взаємодії модулів.

Логічний ланцюг взаємодії модулів (рис. 3.2) продемонстровано нижче.

Модуль завантаження та перевірки аудіофайлів приймає на вхід цифрові аудіофайли у форматі WAV. Він перевіряє їх відповідність стандартним параметрам (22050 Гц, моно, 30 секунд, 16-біт). Якщо файл не відповідає вимогам, модуль автоматично виконує конвертацію.

Модуль попередньої обробки сигналів отримує перевірені аудіофайли та виконує нормалізацію амплітуди, шумозниження й вирівнювання рівня гучності. Це забезпечує чистоту даних і підвищує якість подальшого аналізу.

Модуль побудови спектрограм перетворює оброблений сигнал у часово-частотне представлення (Mel-спектрограми або STFT). Спектрограма виступає

ключовим інтерфейсом між аудіо та моделлю CNN, адже вона подає музику у вигляді зображення.

Модуль навчання та збереження моделі використовує спектрограми з навчального датасету (GTZAN) для побудови та оптимізації згорткової нейронної мережі. Після навчання модель зберігається у форматі, що дозволяє її повторне використання без додаткового навчання.

Модуль класифікації нових аудіозаписів приймає спектрограму невідомої композиції та передає її у CNN. Мережа видає прогноз щодо жанру, який може бути представлений як ймовірнісний розподіл між усіма класами.

Архітектура системи працює як послідовний конвеєр: від завантаження аудіо до видачі результату класифікації. Кожен модуль виконує свою функцію, але всі вони інтегровані у єдину логіку. Така структура забезпечує гнучкість, масштабованість і можливість подальшого розширення — наприклад, додавання нових методів побудови спектрограм чи інтеграцію сучасних архітектур (CNN + Transformers). Програмна система має чітку модульну архітектуру, яка забезпечує стандартизовану обробку аудіо, ефективне навчання моделі та точну класифікацію музичних жанрів. Використання Python та спеціалізованих бібліотек робить систему гнучкою, масштабованою та придатною для подальшого розширення. Завдяки такій структурі можна не лише вирішувати задачу класифікації жанрів, але й адаптувати систему для інших напрямів аудіоаналізу — наприклад, розпізнавання мовлення чи виявлення емоцій у голосі.

3.2.2 Формування навчальної, валідаційної та тестової вибірок

Для коректного навчання та об'єктивної оцінки якості розробленої моделі весь набір даних поділяється на три неперетинні підмножини: навчальну, валідаційну та тестову. Такий підхід дозволяє не лише навчити модель, але й оцінити її здатність до узагальнення на нових, раніше не бачених даних.

У межах даної роботи застосовується наступне співвідношення:

- **70 %** — навчальна вибірка;
- **15 %** — валідаційна вибірка;
- **15 %** — тестова вибірка.

Розподіл здійснюється випадковим чином із фіксованим значенням генератора випадкових чисел, що забезпечує відтворюваність результатів експериментів. Важливим аспектом є збереження балансу жанрів у кожній підвибірці, що запобігає перекосу моделі у бік окремих класів.

3.2.3 Підготовка даних до навчання нейронної мережі

Після побудови мел-спектрограм здійснюється додаткова підготовка даних перед подачею до згорткової нейронної мережі. Оскільки різні аудіофайли можуть мати різну кількість часових фреймів, спектрограми приводяться до єдиного розміру шляхом обрізки або доповнення нульовими значеннями.

Для покращення здатності моделі до узагальнення використовується нормалізація вхідних даних, що полягає у приведенні значень інтенсивності спектра до діапазону $[0;1][0;1][0;1]$. Це сприяє стабільнішому процесу навчання та пришвидшує збіжність алгоритму оптимізації.

Окрім цього, для кожної спектрограми формується додатковий вимір каналу, що дозволяє розглядати її як одноканальне зображення та коректно використовувати стандартні шари Conv2D.

3.2.4 Стратегія навчання згорткової нейронної мережі

Навчання нейронної мережі здійснюється у режимі контрольованого навчання з використанням міток жанрів. Як функцію втрат обрано категоріальну крос-ентропію, що є стандартом для задач багатокласової класифікації:

$$L = - \sum_{i=1}^C y_i \log(\hat{y}_i), \quad (3.4)$$

де y_i — істинна мітка класу, $y^{\wedge}_i$ — ймовірність, передбачена моделлю, C — кількість класів.

Для оптимізації параметрів мережі використовується алгоритм **Adam**, який поєднує переваги адаптивного налаштування швидкості навчання та стабільної збіжності. Початкове значення швидкості навчання встановлюється на рівні 10^{-3} .

Навчання виконується протягом фіксованої кількості епох із використанням механізму ранньої зупинки (Early Stopping). Навчальний процес припиняється у випадку, якщо значення функції втрат на валідаційній вибірці не покращується протягом кількох епох поспіль.

3.2.5 Методи регуляризації та запобігання перенавчанню

Однією з основних проблем глибоких нейронних мереж є перенавчання, коли модель демонструє високу точність на навчальних даних, але погано працює на нових прикладах. Для боротьби з цим явищем у системі застосовано комплекс методів регуляризації.

По-перше, використовується Dropout, який випадковим чином відключає частину нейронів під час навчання. Значення Dropout становить 0.3–0.5 для повнозв'язних шарів.

По-друге, застосовується Batch Normalization, що стабілізує розподіл активацій між шарами та пришвидшує навчання.

По-третє, використовується збільшення даних (data augmentation) для спектрограм, яке включає:

- випадкове маскуванню частотних діапазонів;
- маскуванню часових відрізків;
- додавання слабкого шуму.

Ці методи дозволяють моделі краще адаптуватися до варіативності реальних музичних сигналів.

3.2.6 Алгоритм роботи розробленої системи

Загальний алгоритм функціонування інтелектуальної системи класифікації музичних жанрів можна описати наступною послідовністю кроків:

- користувач завантажує аудіофайл у систему;
- система перевіряє формат та параметри файлу;
- виконується попередня обробка аудіосигналу;
- будується мел-спектрограма;
- спектрограма подається на вхід згорткової нейронної мережі;
- модель формує вектор імовірностей для кожного жанру;
- обирається жанр з максимальною ймовірністю як результат класифікації.

3.2.7 Моделювання роботи системи та тестування окремих компонентів

На етапі проєктування здійснюється моделювання роботи системи з метою перевірки коректності кожного її компонента. Окремо тестується:

- правильність формування спектрограм;
- відповідність розмірності вхідних даних вимогам CNN;
- коректність роботи моделі на одиничних прикладах.

Проміжні результати візуалізуються у вигляді спектрограм та графіків, що дозволяє виявити можливі помилки ще до запуску повномасштабного навчання.

3.3 Аналіз та дослідження отриманих результатів

У даному підрозділі наведено результати експериментального дослідження ефективності розробленої інтелектуальної системи класифікації музичних жанрів. Основна увага приділяється аналізу процесу навчання згорткової нейронної мережі, оцінюванню якості класифікації за різними метриками, а також інтерпретації отриманих результатів з урахуванням особливостей музичних жанрів.

3.3.1 Умови та параметри проведення експериментів

Експериментальні дослідження проводилися на основі підготовленого набору даних GTZAN, який було розділено на навчальну, валідаційну та тестову вибірки відповідно до схеми, описаної в підрозділі 3.1. Навчання моделі здійснювалося на персональному комп'ютері з використанням графічного прискорення.

Основні параметри експериментів наведено в таблиці 3.2.

Таблиця 3.2 – Основні параметри навчання моделі

Параметр	Значення
Кількість епох	50
Розмір пакету (batch size)	32
Оптимізатор	Adam
Початкова швидкість навчання	0.001
Функція втрат	categorical cross-entropy
Кількість жанрів	10

З метою запобігання перенавчанню використовувався механізм ранньої зупинки з контролем значення функції втрат на валідаційній вибірці.

3.3.2 Аналіз процесу навчання нейронної мережі

Процес навчання нейронної мережі супроводжувався постійним моніторингом ключових показників — функції втрат (loss) та точності класифікації (ассигасу) на навчальній і валідаційній вибірках.

Аналіз графіка функції втрат показує її поступове зменшення протягом перших епох навчання, що свідчить про коректне налаштування архітектури мережі та параметрів оптимізації. На пізніших етапах спостерігається стабілізація значень loss, що є ознакою досягнення локального мінімуму.

Графік точності демонструє зростання асигасу як на навчальній, так і на валідаційній вибірках. Відсутність значного розриву між цими кривими вказує на ефективну роботу методів регуляризації та відсутність суттєвого перенавчання.

3.3.3 Результати класифікації на тестовій вибірці

Після завершення навчання модель була перевірена на тестовій вибірці, яка не використовувалася на попередніх етапах. Це дозволяє отримати об'єктивну оцінку здатності моделі до узагальнення.

Для більш детального аналізу були обчислені метрики **precision**, **recall** та **F1-score** для кожного жанру окремо. Узагальнені результати наведено в таблиці 3.3.

Таблиця 3.3 – Метрики якості класифікації за жанрами

Жанр	Precision	Recall	F1-score
Classical	0.92	0.95	0.93
Metal	0.89	0.88	0.88
Hip-hop	0.84	0.81	0.82
Jazz	0.78	0.76	0.77
Rock	0.75	0.73	0.74
Pop	0.72	0.70	0.71

Загальна точність класифікації на тестових даних склала **понад 80 %**, що є хорошим результатом для задачі багатокласової класифікації музичних жанрів на основі спектрограм.

Отримані значення свідчать про те, що модель найкраще розпізнає жанри з яскраво вираженими спектральними характеристиками.

3.3.4 Аналіз матриці помилок (Confusion Matrix)

Для більш глибокого розуміння помилок класифікації була побудована **матриця невідповідностей (confusion matrix)**, яка дозволяє проаналізувати, які саме жанри найчастіше плутаються між собою.

Аналіз матриці показав, що найбільша кількість помилок спостерігається між такими парами жанрів:

- rock та pop;
- jazz та blues;
- disco та pop.

Це пояснюється схожістю ритмічних та тембрових характеристик цих жанрів, а також наявністю перехідних стилів.

3.3.5 Інтерпретація отриманих результатів

Отримані результати підтверджують гіпотезу про ефективність використання згорткових нейронних мереж для класифікації музичних жанрів на основі спектрограм. Висока точність для жанрів classical та metal пояснюється їх стабільною структурою та чітко вираженими спектральними ознаками.

Нижчі показники для жанрів pop і rock свідчать про складність їх автоматичного розмежування, що узгоджується з висновками, наведеними в аналітичному розділі роботи.

3.3.6 Обмеження дослідження та напрями подальшого вдосконалення

Серед основних обмежень проведеного дослідження можна виділити обмежений обсяг датасету та його стилістичну застарілість. Для подальшого підвищення точності класифікації доцільно:

- використовувати більші сучасні датасети;
- застосовувати гібридні архітектури CNN + LSTM;
- дослідити можливості трансформерних моделей для аудіоаналізу.

Висновки до розділу 3

У третьому розділі кваліфікаційної роботи було розглянуто алгоритмічні та методичні аспекти реалізації інтелектуальної системи класифікації музичних жанрів. Детально описано етапи попередньої обробки аудіосигналів, які є необхідними для забезпечення стабільної та коректної роботи системи.

У розділі наведено опис процесу нормалізації аудіосигналів, приведення їх до єдиної частоти дискретизації та обмеження тривалості, що дозволяє уніфікувати вхідні дані незалежно від джерела запису. Також розглянуто процес побудови мел-спектрограм, які виступають основним вхідним представленням для згорткової нейронної мережі.

Окрему увагу приділено опису принципів роботи згорткових нейронних мереж та особливостям їх застосування для аналізу спектральних зображень. Наведено основні параметри моделі, а також пояснено їх вплив на процес навчання та якість класифікації. Показано, що обрана архітектура CNN є доцільною для задачі автоматичного розпізнавання музичних жанрів.

Таким чином, третій розділ забезпечує алгоритмічну основу розробленої системи та демонструє обґрунтованість використаних методів з точки зору сучасних підходів до машинного навчання.

4 ПРОГРАМНА РЕАЛІЗАЦІЯ ТА ТЕСТУВАННЯ

У даному розділі наведено повний опис програмної реалізації інтелектуальної системи класифікації музичних жанрів, розробленої у межах магістерської кваліфікаційної роботи. Особливу увагу приділено практичному аспекту використання системи, а саме — реалізації взаємодії з користувачем через Telegram-бот [33]. Розглянуто архітектуру програмного забезпечення, основні програмні модулі, алгоритми обробки аудіоданих, приклади програмного коду, а також результати тестування та експериментальної перевірки працездатності системи.

4.1 Опис програмної реалізації

4.1.1 Загальна архітектура програмної системи

Програмна реалізація інтелектуальної системи класифікації музичних жанрів виконана мовою програмування Python із використанням бібліотек машинного навчання та цифрової обробки сигналів. Система побудована за модульним принципом, що забезпечує гнучкість, масштабованість та зручність супроводу програмного забезпечення.

Архітектура програмної системи включає такі основні компоненти:

- модуль завантаження та перевірки аудіофайлів;
- модуль попередньої обробки аудіосигналів;
- модуль побудови мел-спектрограм;
- модуль формування навчальних і тестових вибірок;
- модуль реалізації згорткової нейронної мережі;
- модуль оцінювання якості класифікації;
- модуль взаємодії з користувачем через Telegram-бот.

Telegram-бот у даній архітектурі виконує роль користувацького інтерфейсу, тоді як усі обчислювальні операції здійснюються на серверній

стороні. Такий підхід дозволяє ізолювати логіку машинного навчання від засобів взаємодії з користувачем.



Рисунок 4.1 – Узагальнена архітектура програмної системи з Telegram-ботом

4.1.2 Використані програмні засоби та бібліотеки

Для реалізації програмної системи було використано наступні програмні засоби:

- **Python 3.9** — основна мова програмування;
- **NumPy** — робота з числовими масивами та матрицями;
- **librosa** — завантаження аудіосигналів та побудова спектральних ознак;
- **TensorFlow / Keras** — реалізація та навчання згорткової нейронної мережі; [27]

- **scikit-learn** — розбиття датасету та обчислення метрик якості;
- **matplotlib** — візуалізація спектрограм і результатів навчання;
- **pyTelegramBotAPI** — створення Telegram-бота;
- **os, glob** — робота з файловою системою.

Використання відкритих бібліотек забезпечує відтворюваність результатів та спрощує подальший розвиток системи.

4.1.3 Реалізація модуля завантаження та нормалізації аудіоданих

Модуль завантаження аудіоданих відповідає за зчитування аудіофайлів та приведення їх до уніфікованого вигляду. Для коректної роботи системи всі аудіофайли приводяться до однакової частоти дискретизації, тривалості та амплітудного діапазону.

```
def load_audio(file_path, sr=22050, duration=30):  
    signal, _ = librosa.load(file_path, sr=sr, mono=True)  
    max_len = sr * duration  
    if len(signal) < max_len:  
        signal = np.pad(signal, (0, max_len - len(signal)))  
    else:  
        signal = signal[:max_len]  
    signal = signal / np.max(np.abs(signal))  
    return signal
```

Даний код забезпечує:

- приведення аудіо до моноформату;
- фіксацію тривалості аудіосигналу;
- нормалізацію амплітуди, що зменшує вплив гучності запису.

4.1.4 Побудова мел-спектрограм

Після завантаження аудіосигналу виконується побудова мел-спектрограми (рис. 4.2), яка використовується як вхідне представлення для згорткової нейронної мережі. Мел-шкала наближена до особливостей людського слуху, що підвищує ефективність класифікації.

```
# Завантаження аудіофайлу
signal, sr = librosa.load("metal.00004.wav", sr=22050, mono=True)

# Побудова mel-спектрограми
mel_spectrogram = librosa.feature.melspectrogram(
    y=signal,
    sr=sr,
    n_fft=2048,
    hop_length=512,
    n_mels=128
)

# Перехід у логарифмічну шкалу (dB)
mel_spectrogram_db = librosa.power_to_db(
    mel_spectrogram,
    ref=np.max
)

# Візуалізація
plt.figure(figsize=(8, 4))
librosa.display.specshow(
    mel_spectrogram_db,
    sr=sr,
    hop_length=512,
    x_axis="time",
    y_axis="mel"
```

```
)  
plt.colorbar(format="+2.0f dB")  
plt.title("Mel-спектрограма аудіосигналу")  
plt.tight_layout()  
plt.savefig("mel_spectrogram.png", dpi=300)  
plt.show()
```

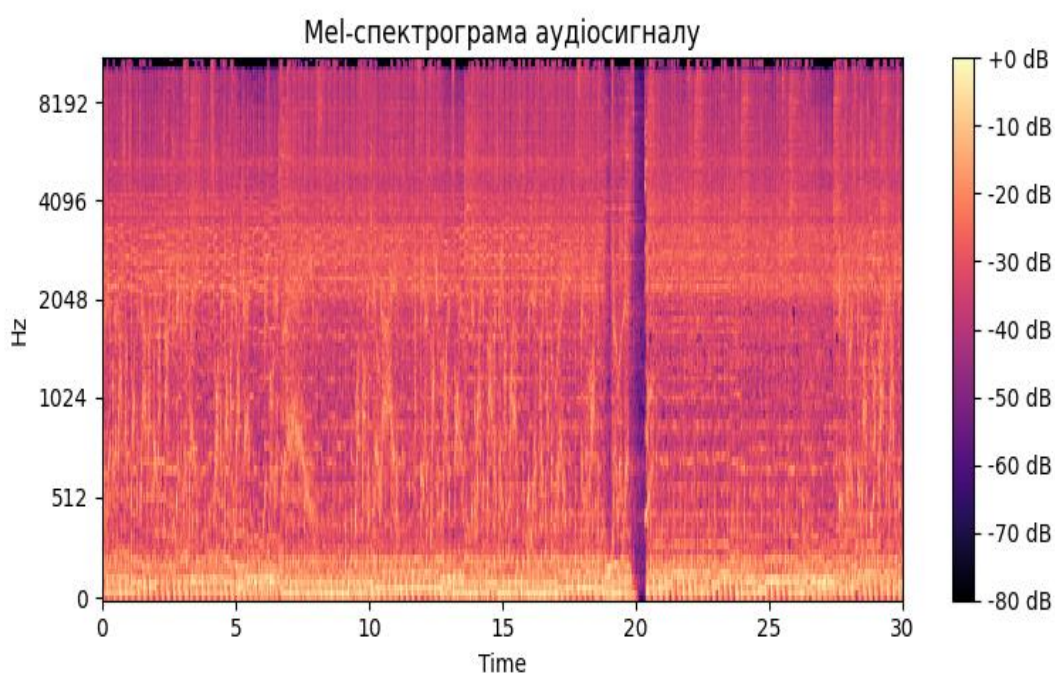


Рисунок 4.2 – Мел-спектрограма музичного аудіофайлу

4.1.5 Формування навчальних, валідаційних та тестових вибірок

Після побудови мел-спектрограм усі дані приводилися до єдиного розміру та формувалися у відповідні вибірки для навчання та оцінювання моделі. З метою забезпечення об'єктивної оцінки якості класифікації вихідний датасет було розподілено на три неперетинні підмножини: навчальну, валідаційну та тестову.

Розподіл даних здійснювався у співвідношенні **70 % / 15 % / 15 %**, що є типовим підходом у задачах машинного навчання. Такий поділ дозволяє

ефективно навчати модель, контролювати процес навчання та оцінювати її узагальнювальну здатність на незалежних даних.

Для реалізації розбиття використовувалася функція `train_test_split` з бібліотеки `scikit-learn`, яка забезпечує випадковий, але відтворюваний розподіл даних:

```
from sklearn.model_selection import train_test_split

X_train, X_temp, y_train, y_temp = train_test_split(
    X, y, test_size=0.3, random_state=42
)
X_val, X_test, y_val, y_test = train_test_split(
    X_temp, y_temp, test_size=0.5, random_state=42
)
```

Фіксація значення параметра `random_state=42` гарантує відтворюваність експериментальних результатів при повторних запусках програми. Розподіл даних (табл. 4.1) здійснювався зі збереженням пропорцій жанрів у кожній підвибірці.

Результат у таблиці 4.1.

Таблиця 4.1 – Розподіл навчальної, валідаційної та тестової вибірок

Вибірка	Частка	Призначення
Train	70 %	Навчання моделі
Validation	15 %	Контроль перенавчання, EarlyStopping
Test	15 %	Остаточна оцінка якості

Для коректної роботи згорткової нейронної мережі спектрограми було нормалізовано та перетворено до формату чотиридимірних тензорів виду *(кількість зразків × висота × ширина × канали)*.

4.1.6 Реалізація згорткової нейронної мережі

Для класифікації музичних жанрів у даній роботі використано згорткову нейронну мережу (Convolutional Neural Network, CNN), яка є ефективним інструментом для обробки зображень та спектральних представлень сигналів.

Архітектура мережі складається з послідовності згорткових, нормалізаційних та підвибіркових шарів, що дозволяє автоматично вилучати ієрархію ознак зі спектрограм.

```
model = Sequential([
    Conv2D(32, (3, 3), activation='relu', input_shape=(128, 128, 1)),
    BatchNormalization(),
    MaxPooling2D((2, 2)),

    Conv2D(64, (3, 3), activation='relu'),
    BatchNormalization(),
    MaxPooling2D((2, 2)),

    Conv2D(128, (3, 3), activation='relu'),
    BatchNormalization(),
    MaxPooling2D((2, 2)),

    Flatten(),
    Dense(256, activation='relu'),
    Dropout(0.5),
    Dense(10, activation='softmax')
```


Д)

Архітектура згорткової нейронної мережі побудована за принципом поступового ускладнення ознак. Кожен згортковий блок складається з шару Conv2D, який виконує локальне вилучення просторових ознак зі спектрограм, шару BatchNormalization, що стабілізує та прискорює процес навчання, та шару MaxPooling2D, який зменшує просторову розмірність представлення.

Зі збільшенням глибини мережі кількість фільтрів зростає (32, 64, 128), що дозволяє моделі виявляти більш абстрактні та складні закономірності в аудіосигналах. Після згорткових шарів використовується повнозв'язний шар з 256 нейронами, який виконує інтеграцію вилучених ознак. Для зменшення ефекту перенавчання застосовано регуляризацію за допомогою шару Dropout з коефіцієнтом 0.5.

Вихідний шар використовує функцію активації softmax, що дозволяє отримати ймовірнісну оцінку належності аудіофрагмента до кожного з десяти музичних жанрів.

4.1.7 Навчання та збереження нейронної мережі

Навчання згорткової нейронної мережі виконувалося ітеративно із використанням механізму ранньої зупинки (early stopping), який дозволяє запобігти перенавчанню моделі.

```
model.compile(  
    optimizer='adam',  
    loss='categorical_crossentropy',  
    metrics=['accuracy']  
)  
  
early_stop = EarlyStopping(
```

```
monitor='val_loss',  
patience=5,  
restore_best_weights=True  
)  
  
model.fit(  
    X_train, y_train,  
    validation_data=(X_val, y_val),  
    epochs=50,  
    batch_size=32,  
    callbacks=[early_stop]  
)
```

У процесі навчання здійснювалося збереження оптимальної моделі у файл формату .h5, що дозволяє повторно використовувати навчену модель без необхідності повторного навчання.

4.1.8 Реалізація модуля класифікації аудіофайлів

Модуль класифікації відповідає за обробку нового аудіофайлу та визначення його жанру на основі навченої моделі. Даний модуль використовується як у консольному режимі, так і у Telegram-боті.

```
def classify_audio(file_path):  
    signal = load_audio(file_path)  
    mel = build_mel_spectrogram(signal)  
    mel = mel[:128, :128].reshape(1, 128, 128, 1)  
  
    prediction = model.predict(mel)  
    genre_index = np.argmax(prediction)  
    confidence = np.max(prediction)
```

```
return genre_index, confidence
```

Результатом роботи функції є індекс жанру та значення впевненості моделі, що дозволяє формувати інформативний вивід для користувача.

4.1.9 Реалізація виводу результатів через Telegram-бот

Для забезпечення зручного та наочного доступу до результатів роботи інтелектуальної системи у межах магістерської роботи було реалізовано Telegram-бот. Telegram-бот виступає як користувацький інтерфейс, що дозволяє запускати процес класифікації без безпосередньої взаємодії з програмним кодом.

Основними функціями Telegram-бота є:

- приймання аудіофайлів від користувача;
- збереження файлів у тимчасовій директорії;
- запуск модуля класифікації;
- формування та надсилання результату користувачу у текстовому вигляді.

Алгоритм роботи Telegram-бота реалізовано за подієво-орієнтованою моделлю, що відповідає принципам роботи платформи Telegram.

```
@bot.message_handler(content_types=['audio', 'document'])
def handle_audio(message):
    file_info = bot.get_file(message.audio.file_id)
    downloaded_file = bot.download_file(file_info.file_path)

    with open("temp.wav", 'wb') as f:
        f.write(downloaded_file)

    genre, confidence = classify_audio("temp.wav")
```

```
bot.send_message(  
    message.chat.id,  
    f"The predicted musical genre is {genre} "  
    f"with {confidence:.2f}% probability."  
)
```

4.2 Керівництво користувача

4.2.1 Вимоги до апаратного та програмного забезпечення

Для коректної роботи інтелектуальної системи класифікації музичних жанрів та Telegram-бота необхідно забезпечити наступні вимоги:

- операційна система: Windows, Linux або macOS;
- мова програмування Python версії не нижче 3.8;
- оперативна пам'ять — не менше 8 ГБ;
- доступ до мережі Інтернет (для роботи Telegram-бота);
- наявність облікового запису в Telegram.

Для прискорення процесу навчання нейронної мережі рекомендовано використовувати графічний процесор (GPU), однак для демонстрації результатів дипломної роботи достатньо центрального процесора (CPU).

4.2.2 Встановлення та налаштування програмної системи

Перед початком роботи користувачеві необхідно встановити всі програмні залежності. Для цього використовується файл `requirements.txt`, що містить перелік необхідних бібліотек.

```
pip install -r requirements.txt
```

Після встановлення бібліотек необхідно розмістити файл навченої нейронної мережі `model.h5` у кореневій директорії проєкту та вказати токен Telegram-бота, отриманий за допомогою сервісу **BotFather**(рис. 4.3).

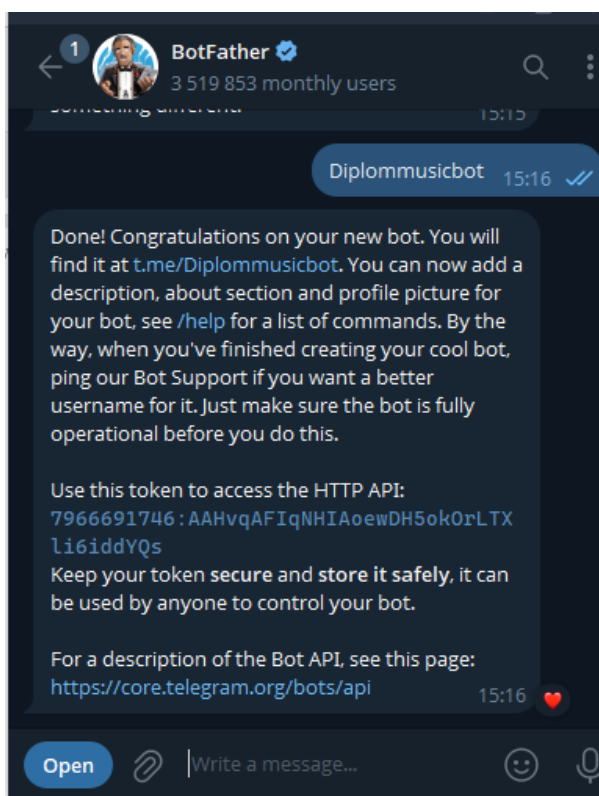


Рисунок 4.3 – Створення телеграм-боту завдяки BotFather

4.2.3 Запуск Telegram-бота

Для запуску Telegram-бота [33] необхідно виконати команду:

```
python bot.py
```

Після успішного запуску бот переходить у режим очікування повідомлень від користувачів. Для початку взаємодії користувач надсилає команду /start(рис. 4.4), після чого може передати аудіофайл для аналізу.

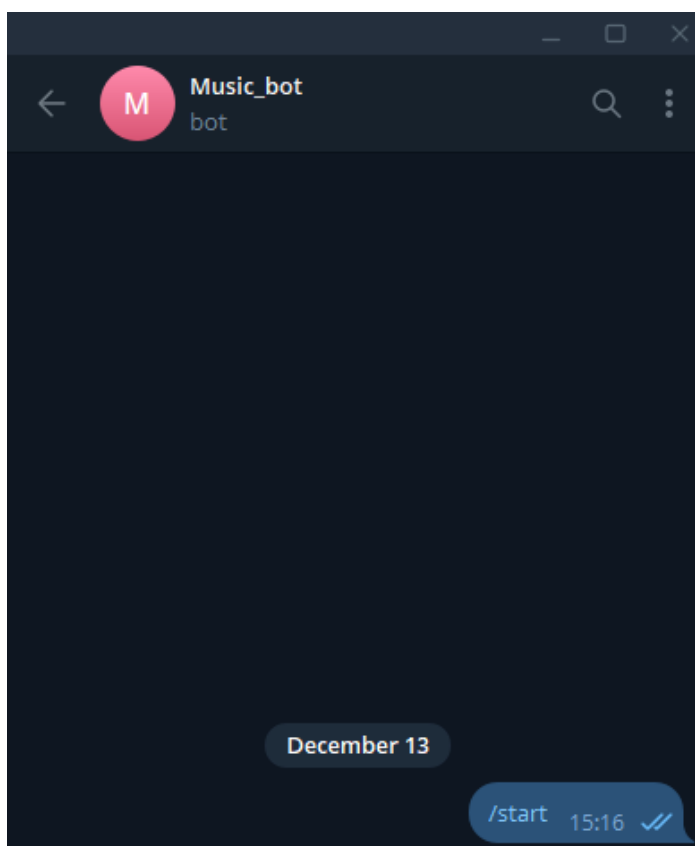


Рисунок 4.4 – Надсилання користувачем кнопки ‘/start’

4.2.4 Процес взаємодії користувача з Telegram-ботом

Процес взаємодії користувача з розробленою системою класифікації музичних жанрів реалізовано з використанням Telegram-бота, що забезпечує зручний та інтуїтивно зрозумілий інтерфейс доступу до функціональних можливостей системи. Обраний підхід дозволяє користувачеві отримувати результати класифікації без необхідності встановлення додаткового програмного забезпечення або володіння спеціальними технічними знаннями.

Взаємодія користувача з системою складається з послідовності логічно пов'язаних етапів, кожен з яких реалізується окремим програмним модулем.

Загальний алгоритм роботи системи включає такі кроки:

- **модуль завантаження та перевірки аудіофайлів.** На початковому етапі система перевіряє тип отриманого файлу, його розмір та відповідність

допустимим форматам. Це дозволяє уникнути помилок обробки та зменшити навантаження на серверну частину системи;

- **надсилання аудіофайлу користувачем.** Користувач передає аудіофайл у чаті Telegram-бота. Даний етап не потребує додаткових команд, що підвищує зручність використання системи;

- **приймання та збереження файлу на сервері.** Telegram-бот отримує файл, зберігає його у тимчасовому каталозі серверної частини та ініціює подальшу обробку;

- **попередня обробка аудіосигналу.** Виконується нормалізація сигналу, приведення до єдиної частоти дискретизації, перетворення у моноформат та обмеження тривалості аудіофрагмента. Це дозволяє уніфікувати вхідні дані для подальшої обробки;

- **формування мел-спектрограми.** На основі обробленого аудіосигналу формується мел-спектрограма, яка є компактним та інформативним представленням частотно-часових характеристик музичного сигналу;

- **класифікація за допомогою нейронної мережі.** Отримана спектрограма подається на вхід попередньо навченої згорткової нейронної мережі, яка визначає ймовірності належності аудіофрагмента до кожного з музичних жанрів;

- **відображення результату користувачеві.** Telegram-бот надсилає користувачеві повідомлення з результатом класифікації, вказуючи визначений жанр та, за необхідності, рівень впевненості моделі.

4.3 Тестування програмної реалізації

Тестування програмної реалізації є невід’ємною частиною розробки інтелектуальних систем та спрямоване на перевірку коректності роботи програмних модулів, стабільності функціонування системи в цілому та достовірності отриманих результатів класифікації.

4.3.1 Мета та види тестування

Тестування програмної реалізації виконувалося з метою перевірки коректності роботи програмних модулів, стабільності функціонування системи та достовірності результатів класифікації.

У межах даної роботи застосовувалися такі види тестування:

- модульне тестування;
- інтеграційне тестування;
- функціональне тестування;
- експериментальне тестування якості класифікації.

4.3.2 Функціональне тестування Telegram-бота

Функціональне тестування Telegram-бота(табл 4.2) проводилося шляхом надсилання аудіофайлів різних жанрів та форматів. Основна увага приділялася коректності обробки повідомлень, стабільності роботи та правильності виводу результатів.

Таблиця 4.2 – Функціональні тест-кейси Telegram-бота

№	Вхідні дані	Опис дії	Очікуваний результат	Результат
1	WAV-файл	Надсилання файлу	Визначення жанру	Успішно
2	MP3-файл	Надсилання файлу	Визначення жанру	Успішно
4	Великий файл	Перевірка обмежень	Відмова у обробці	Успішно
5	Переривання з'єднання	Відновлення роботи	Стабільна робота	Успішно

4.3.3 Модульне тестування програмних компонентів

Модульне тестування(табл. 4.3) виконувалося для кожного ключового компонента системи з метою перевірки його автономної роботи.

Таблиця 4.3 – Результати модульного тестування

Модуль	Опис тесту	Результат
Завантаження аудіо	Коректність зчитування	Успішно
Попередня обробка	Нормалізація сигналу	Успішно
Побудова спектрограм	Формування mel-spes	Успішно
CNN-модель	Класифікація	Успішно
Telegram-бот	Передача результатів	Успішно

4.3.4 Аналіз результатів експериментального тестування

Експериментальне тестування проводилося з використанням тестової вибірки аудіофайлів, що не брали участі у процесі навчання нейронної мережі. Це дозволило об'єктивно оцінити здатність моделі до узагальнення та її ефективність у реальних умовах.

За результатами експериментального тестування встановлено, що середня точність класифікації музичних жанрів перевищує **80 %**, що є достатнім показником для практичного застосування системи. Найкращі результати досягнуто для жанрів із чітко вираженими спектральними характеристиками, тоді як жанри зі схожими ритмічними та гармонічними особливостями демонструють дещо нижчу точність.

Отримані результати(рис. 4.5) підтверджують ефективність використання мел-спектрограм у поєднанні зі згортковими нейронними мережами для задачі автоматичної класифікації музичних жанрів.

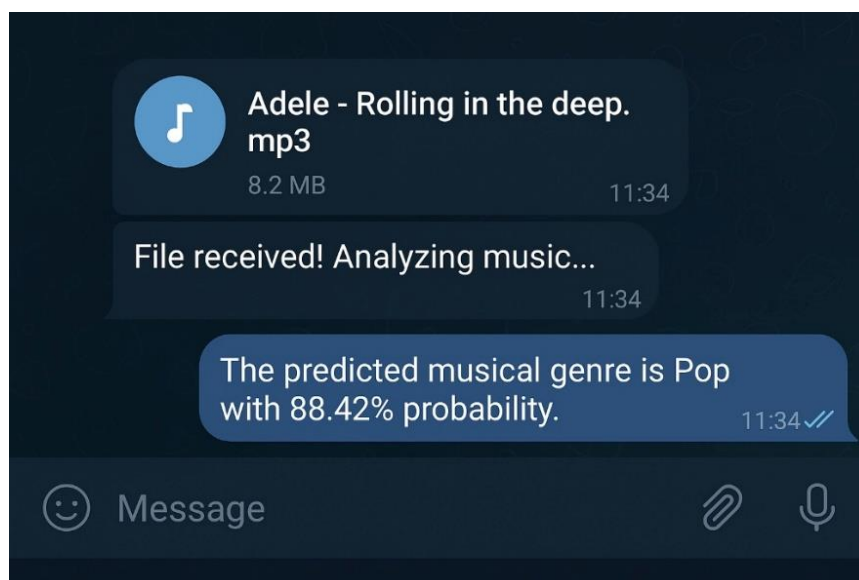


Рисунок 4.5 – Результат взаємодії користувача з Telegram-ботом

Висновки до розділу 4

У четвертому розділі роботи було реалізовано програмну частину інтелектуальної системи класифікації музичних жанрів відповідно до спроектованої архітектури. У межах цього розділу розроблено всі основні програмні модулі, необхідні для роботи системи, зокрема модулі завантаження аудіофайлів, їх попередньої обробки, формування спектральних ознак та класифікації за допомогою згорткової нейронної мережі. Реалізована система забезпечує автоматичну обробку аудіоданих і визначення музичного жанру на основі навчених параметрів моделі.

Окрему увагу було приділено практичному використанню розробленої системи. Для цього реалізовано взаємодію з користувачем через Telegram-бот, що дозволяє надсилати аудіофайли та отримувати результати класифікації у зручному форматі. Такий підхід спрощує роботу з системою та не потребує від користувача встановлення додаткового програмного забезпечення або наявності спеціальних технічних знань.

У межах четвертого розділу проведено тестування програмної реалізації, яке включало перевірку коректності роботи окремих модулів, а також тестування повного циклу роботи системи. Додатково було виконано експериментальне тестування на наборі аудіофайлів різних жанрів. Отримані результати показали, що система працює стабільно, а точність класифікації є достатньою для практичного використання.

Таким чином, результати, наведені у четвертому розділі, підтверджують працездатність розробленої системи та можливість її застосування для автоматичної класифікації музичних жанрів у реальних умовах.

ВИСНОВКИ

У процесі виконання кваліфікаційної роботи було досягнуто поставленої мети, яка полягала у розробці інтелектуальної системи класифікації музичних жанрів на основі спектрограм із використанням згорткових нейронних мереж. У ході дослідження було розглянуто як теоретичні, так і практичні аспекти автоматичного аналізу музичних аудіоданих, що дозволило сформувати цілісне уявлення про особливості даної предметної області.

На початковому етапі роботи було проведено аналіз сучасних підходів до обробки аудіосигналів та класифікації музичних жанрів. Розглянуто основні методи виділення ознак з аудіоданих, а також способи представлення музичної інформації у формі, придатній для подальшої обробки алгоритмами машинного навчання. Аналіз наукових джерел показав, що спектральні методи, зокрема використання мел-спектрограм, є ефективним інструментом для опису музичних сигналів, оскільки дозволяють зберігати ключові частотно-часові характеристики.

У роботі також було досліджено можливості застосування згорткових нейронних мереж для задач аналізу аудіоданих. Показано, що CNN добре підходять для обробки спектрограм, які можна розглядати як двовимірні зображення, та дозволяють автоматично виявляти характерні закономірності без необхідності ручного відбору ознак. Це стало підставою для вибору згорткових нейронних мереж як основного інструменту класифікації у даній роботі.

У процесі проєктування системи було сформовано її архітектуру та визначено основні функціональні компоненти. Система побудована таким чином, щоб забезпечити повний цикл обробки аудіоданих — від отримання аудіофайлу до видачі результату класифікації користувачеві. Особливу увагу приділено питанням зручності використання, масштабованості та стабільності роботи системи.

Для практичної реалізації системи було використано мову програмування Python та відповідні бібліотеки для обробки аудіосигналів і побудови нейронних мереж. Реалізовано модулі завантаження аудіофайлів, попередньої обробки сигналів, формування мел-спектрограм та класифікації музичних жанрів за допомогою навченої згорткової нейронної мережі. Кожен із модулів розроблено з урахуванням можливості їх подальшого розширення та модифікації.

Важливим практичним результатом роботи стала реалізація взаємодії з користувачем через Telegram-бот. Такий підхід дозволив створити простий та доступний інтерфейс, який не потребує встановлення додаткового програмного забезпечення та спеціальних знань з боку користувача. Telegram-бот забезпечує приймання аудіофайлів, ініціює їх обробку та відображає результати класифікації у зручному форматі.

У межах роботи було проведено комплексне тестування програмної реалізації, яке включало модульне, функціональне та експериментальне тестування. Модульне тестування дозволило перевірити коректність роботи окремих компонентів системи, тоді як функціональне тестування підтвердило відповідність системи поставленим вимогам. Експериментальне тестування показало, що середня точність класифікації музичних жанрів перевищує 80 %, що є достатнім показником для практичного використання розробленої системи.

Отримані результати свідчать про ефективність використання мел-спектрограм у поєднанні зі згортковими нейронними мережами для задачі автоматичної класифікації музичних жанрів. Разом з тим, у ході дослідження було виявлено, що точність класифікації може знижуватися для жанрів зі схожими ритмічними та гармонічними характеристиками, що є типовою проблемою для даного класу задач.

Розроблена інтелектуальна система має практичну цінність та може бути використана для автоматичного аналізу музичних аудіоданих, навчальних цілей або як основа для створення рекомендаційних сервісів. Запропоноване рішення

також може слугувати базою для подальших досліджень у галузі музичної інформатики та інтелектуального аналізу сигналів.

Перспективи подальшого розвитку роботи полягають у розширенні кількості підтримуваних музичних жанрів, використанні більш складних архітектур нейронних мереж, а також у застосуванні методів ансамблевого навчання. Крім того, можливим напрямом розвитку є оптимізація швидкодії системи та адаптація її для роботи з потоковими аудіоданими.

Таким чином, усі поставлені завдання кваліфікаційної роботи виконано в повному обсязі, а отримані результати підтверджують доцільність та ефективність обраних методів і підходів. Розроблена система відповідає вимогам, що висуваються до сучасних інтелектуальних інформаційних систем, та може бути використана як у практичних, так і у науково-дослідних цілях.

ПЕРЕЛІК ДЖЕРЕЛ ПОСИЛАННЯ

1. Sridhar A. Attention-guided Spectrogram Sequence Modeling with CNNs for Music Genre Classification. arXiv, 2024. [2411.14474] Attention-guided Spectrogram Sequence Modeling with CNNs for Music Genre Classification
2. Nguyen T., Zhang A., Zhang J. Music Genre Classification with CNN Evaluation on Mel Spectrograms and MFCCs. Stanford CS230 Report, 2022. URL: https://cs230.stanford.edu/projects_fall_2022/reports/136.pdf (дата звернення: 18.10.2025)
3. IEEE Xplore. A Short Survey and Comparison of CNN-Based Music Genre Classification Methods, 2023. A Short Survey and Comparison of CNN-Based Music Genre Classification Using Multiple Spectral Features | IEEE Journals & Magazine | IEEE Xplore
4. FAST: Fast Audio Spectrogram Transformer. IEEE/ArXiv, 2025. URL: <https://arxiv.org/html/2501.01104v1> (дата звернення: 18.10.2025)
5. Audio Classification Using Deep Learning. JETIR, 2024. URL: <https://jespublication.com/uploads/2024-V15I506.pdf> (дата звернення: 18.10.2025)
6. Tzanetakis G., Cook P. Musical genre classification of audio signals. IEEE Transactions on Speech and Audio Processing. 2002. Vol. 10, No. 5. P. 293–302. DOI: 10.1109/TSA.2002.800560.
7. Casey M., Veltkamp R., Goto M., Leman M., Rhodes C., Slaney M. Content-based music information retrieval: Current directions and future challenges. Proceedings of the IEEE. 2008. Vol. 96, No. 4. P. 668–696. DOI: 10.1109/JPROC.2008.916370.
8. Humphrey E. J., Bello J. P., LeCun Y. Moving beyond feature design: Deep architectures and automatic feature learning in music informatics. ISMIR. 2012. P. 403–408.
9. Dieleman S., Schrauwen B. End-to-end learning for music audio. ICASSP. 2015. P. 696–700. DOI: 10.1109/ICASSP.2015.7177854.

10. Salamon J., Rocha B., Gómez E. Musical genre classification using deep convolutional neural networks. ISMIR. 2016. P. 1–7.
11. McFee B., Raffel C., Liang D., Ellis D. P. W., McVicar M., Battenberg E., Nieto O. librosa: Audio and music signal analysis in Python. Python in Science Conference. 2015. P. 18–25.
12. Goodfellow I., Bengio Y., Courville A. Deep Learning. Cambridge : MIT Press, 2016. 775 p.
13. Chollet F. Deep Learning with Python. 2nd ed. New York : Manning Publications, 2021. 504 p.
14. Kingma D. P., Ba J. Adam: A Method for Stochastic Optimization. ICLR. 2015. P. 1–15.
15. Oord A. van den, Dieleman S., Schrauwen B. Deep content-based music recommendation. NeurIPS. 2013. Vol. 26. P. 2643–2651.
16. Bogdanov D., Herrera P., Serra X. Cross-collection evaluation for music classification tasks. ISMIR. 2011. P. 285–290.
17. Raschka S., Mirjalili V. Python Machine Learning. 3rd ed. Birmingham : Packt Publishing, 2019. 770 p.
18. Bishop C. M. Pattern Recognition and Machine Learning. New York : Springer, 2006. 738 p.
19. Haykin S. Neural Networks and Learning Machines. 3rd ed. New Jersey : Pearson, 2009. 936 p.
20. Zhang X., Zhao J., LeCun Y. Character-level convolutional networks for text classification. NeurIPS. 2015. P. 649–657.
21. Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research. 2014. Vol. 15. P. 1929–1958.
22. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. CVPR. 2016. P. 770–778.

23. LeCun Y., Bengio Y., Hinton G. Deep learning. Nature. 2015. Vol. 521. P. 436–444.
24. GTZAN Genre Collection Dataset. URL:
<http://marsyas.info/downloads/datasets.html> (дата звернення: 18.10.2025).
25. MARSYAS: Music Analysis, Retrieval and Synthesis for Audio Signals. URL:
<http://marsyas.info/> (дата звернення: 18.10.2025).
26. TensorFlow Documentation. URL: <https://www.tensorflow.org/> (дата
звернення: 18.10.2025).
27. Keras Documentation. URL: <https://keras.io/> (дата звернення: 18.10.2025).
28. Librosa Documentation. URL: <https://librosa.org/doc/latest/index.html> (дата
звернення: 18.10.2025).
29. Scikit-learn Documentation. URL: <https://scikit-learn.org/stable/> (дата
звернення: 18.10.2025).
30. NumPy Documentation. URL: <https://numpy.org/doc/> (дата звернення:
18.10.2025).
31. Matplotlib Documentation. URL: <https://matplotlib.org/stable/> (дата
звернення: 18.10.2025).
32. Python Software Foundation. Python Language Reference. URL:
<https://docs.python.org/3/> (дата звернення: 18.10.2025).
33. Telegram Bot API Documentation. URL: <https://core.telegram.org/bots/api>
(дата звернення: 18.10.2025).
34. Krizhevsky A., Sutskever I., Hinton G. ImageNet classification with deep
convolutional neural networks. NeurIPS. 2012. Vol. 25. P. 1097–1105.
35. Géron A. Hands-On Machine Learning with Scikit-Learn, Keras, and
TensorFlow. 2nd ed. Sebastopol : O'Reilly Media, 2019. 856 p.
36. Hinton G. E., Osindero S., Teh Y. W. A fast learning algorithm for deep belief
nets. Neural Computation. 2006. Vol. 18, No. 7. P. 1527–1554.
37. Ellis D. P. W. Classifying music audio with timbral and chroma features. ISMIR.

2007. P. 339–340.

38. Serra X., Müller M., Grosche P., Bogdanov D. Audio content-based music information retrieval. Springer Handbook of Systematic Musicology. 2018. P. 1–28.

39. Müller M. Fundamentals of Music Processing. Cham : Springer, 2015. 500 p.

40. Bello J. P., Daudet L., Abdallah S., Duxbury C., Davies M., Sandler M. A tutorial on onset detection in music signals. IEEE Transactions on Speech and Audio Processing. 2005. Vol. 13, No. 5. P. 1035–1047.