Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп'ютерної інженерії та управління
(повна назва)

Кафедра Комп'ютерних інтелектуальних технологій та систем
(повна назва)

КВАЛІФІКАЦІЙНА РОБОТА

Пояснювальна записка

рівень вищої освіти другий (магістерський)

Інтелектуальна система класифікації стилів музики.

Виконав:

студент 2 курсу, групи КІТм-21-1
Ушаков, М. Р.
Спеціальність 123 Комп'ютерна інженерія
Тип програми освітньо-професійна
Освітня програма Комп'ютерні інтелектуальні технології
Керівник проф. Н. Г. Аксак

Допускається до захисту _____
(підпис)

Зав. кафедри _____ проф. Руденко О.Г.
(підпис)

2022 р.

Харківський національний університет радіоелектроніки

Факультет Комп’ютерної інженерії та управління

Кафедра Комп’ютерних інтелектуальних технологій та систем

Рівень вищої освіти другий (магістерський)

Спеціальність 123 – Комп’ютерна інженерія

Тип програми освітньо-професійна

Освітня програма Комп'ютерні інтелектуальні технології

ЗАТВЕРДЖУЮ:
Зав. кафедри _____
(підпис)
“____” _____ 20__ р.

ЗАВДАННЯ

НА КВАЛІФІКАЦІЙНУ РОБОТУ

студентові Ушакову Матвію Романовичу
(прізвище, ім’я, по батькові)

1. Тема роботи Інтелектуальна система класифікації стилів музики

затверджена наказом по університету від “ 07 ” листопада 2022 р. № 1455 Ст

2. Термін подання студентом роботи до екзаменаційної комісії 10 грудня 2022 р.

3. Вхідні дані до роботи

1) Мова програмування Python
2) Середовище розробки Visual Studio Code

4. Перелік питань, що потрібно опрацювати в роботі

1) Аналіз музики для розуміння комп’ютером. 2) Підбір класифікації жанрів.
3) Розробка моделі нейронних мереж для класифікації жанрів музики.
5) Створення інтелектуальних систем різних типів. 6) Тестування інтелектуальних систем
7) Порівняння інтелектуальних систем. 8) Висновки

5. Перелік графічного матеріалу із зазначенням креслеників, схем, плакатів, комп'ютерних ілюстрацій (слайдів) _____

1) Слайди-презентація

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

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

КАЛЕНДАРНИЙ ПЛАН

№	Назва етапів роботи	Термін виконання етапів роботи	Примітка
1	Аналіз предметної області та актуальності	08.11 – 13.11	
	класифікації музичних стилів		
2	Огляд сучасної літератури за напрямком	13.11 – 20.11	
	магістерської роботи		
3	Аналіз способів виділення музичних функцій	21.11 – 23.11	
4	Проектування моделей нейронних систем	24.11 – 26.11	
5	Розробка моделі нейронних систем та	26.11 – 30.12	
	навчання		
6	Оформлення пояснювальної записки	01.12 – 07.12	
7	Оформлення графічного матеріалу	08.12 – 10.12	

Дата видачі завдання 07 листопада 2022 р.

Студент _____
(підпис)

Керівник роботи _____ _____
(підпис) (посада, прізвище, ініціали)

РЕФЕРАТ

Пояснювальна записка кваліфікаційної роботи: 82 с., 27 рис., 2 табл., 2 дод., 63 джерела.

НЕЙРОННІ МЕРЕЖІ, RNN, Bi-RNN, CNN, РОЗПІЗНАВАННЯ АУДІО, ГЛИБОКЕ НАВЧАННЯ, МОДЕЛЬ, ЗГОРТКОВА МЕРЕЖА, РЕКУРСИВНА МЕРЕЖА.

Метою даної роботи є розробка системи розпізнавання музикальних жанрів за допомогою нейронних мереж.

У роботі пропонується система розпізнавання музики з використанням нейронних мереж. Під час роботи було представлено та порівняно два типи нейронної мережі, а саме: згорткова нейронна мережа, рекурсивна нейронна мережа.

Об’єктом дослідження є система класифікації стилів музики.

Предметом дослідження є процес вилучення музичних функцій та характеристик необхідних для класифікації її жанрів через непостійність звукового сигналу яким і є музика, а також розпливчастим поняттям кожного жанру.

ABSTRACT

Master’s thesis: 82 pages, 27 figures, 2 tables, 2 appendices, 63 sources.

NEURAL NETWORKS, RNN, Bi-RNN, CNN, AUDIO RECOGNITION, DEEP LEARNING, MODEL, CONVULSIONAL NETWORK, RECURSIVE NETWORK.

The purpose of this work is to develop a system for recognizing musical genres using neural networks.

The paper proposes a music recognition system using neural networks. During the work, two types of neural network were presented and compared, namely: convolutional neural network, recursive neural network.

The object of research is the system of classification of music styles**.**

The subject of the study is the process of extracting musical functions and characteristics necessary for the classification of its genres due to the impermanence of the sound signal, which is music, as well as the vague concept of each genre.

Міністерство освіти і науки України
Харківський національний університет радіоелектроніки

Факультет Комп’ютерної інженерії та управління

Кафедра Комп’ютерних інтелектуальних технологій та систем

АНОТАЦІЯ

КВАЛІФІКАЦІЙНОЇ РОБОТИ

рівень вищої освіти другий (магістерський)

Інтелектуальна система класифікації стилів музики

.

Виконав:
студент 2 курсу, групи КІТм-21-1
Ушаков, М. Р.
Спеціальність 123 Комп’ютерна інженерія
Тип програми освітньо-професійна
Освітня програма Комп’ютерні інтелектуальні технології
Керівник проф., Н.Г. Аксак

2022 р.

АНОТАЦІЯ

Ушаков Матвій Романович. Інтелектуальна система класифікації стилів музики. – Магістерська кваліфікаційна робота.

Актуальність теми дослідження.

За останні роки, зі швидким розвитком та інноваціями Інтернету та мультимедійних технологій, цифрова музика давно стала основною формою слухання музики людьми, що також сприяє зростанню попиту на оцінку музики.

Музичний стиль зараз є одним з найбільш часто використовуваних атрибутів класифікації для керування та зберігання цифрових музичних баз даних, а також є одним з основних елементів пошуку та класифікації, які використовуються більшістю музичних онлайн-сайтів. Ефективність ручного методу маркування, що використовується при пошуку ранньої музичної інформації, більше не може задовольнити потреби менеджменту в умовах величезної кількості музичних даних, і це дуже легко може споживати багато робочої сили та часу. Як наслідок, створення алгоритмів класифікації музичних стилів є критичним для досягнення мети автоматичної класифікації музичних стилів.

Класифікація музичних жанрів є важливою галуззю пошуку музичної інформації. Правильна класифікація музики має велике значення для підвищення ефективності пошуку музичної інформації.

Технологія класифікації музичних стилів може додавати теги стилів до музики на основі вмісту. Коли мова заходить про дослідження та впровадження таких аспектів, як ефективна організація, підбір та рекомендації щодо музичних ресурсів, це має вирішальне значення. Методи класифікації музичних стилів використовує широкий діапазон акустичних характеристик. Розробка характеристик вимагає музичних знань, і характеристики різних класифікаційних завдань не завжди узгоджені. Швидкий розвиток нейронних мереж і технології великих даних забезпечив новий спосіб кращого вирішення проблеми класифікації музичних стилів.

Метою даної роботи є розробка системи розпізнавання музикальних жанрів за допомогою нейронних мереж.

Об’єктом дослідження є нейронна мережа класифікації музичних стилів.

Предметом дослідження є способи та результати вилучення музичних функцій та характеристик необхідних для класифікації її жанрів через непостійність звукового сигналу яким і є музика, а також розпливчастим поняттям кожного жанру. Невід’ємною частиною цього дослідження є використання тих функцій які можна отримати при обробці музики в нейронних мережах, задля отримання більш чітких результатів.

Методи дослідження: вивчення та отримання музичних функцій задля отримання розуміння як їх можна використовувати в аналогії з іншими типами даних, створення декількох видів нейронних систем з різними типами навчання та обробки інформації задля отримання найкращого результату при порівнянні їх.

Практична цінність отриманих результатів. Результати виконаної роботи можуть бути використані в подальшому для кращої систематизації музикальних творів і музики взагалі. Також ці результати можуть бути корисними при пошуку окремого типу композицій, наприклад для деякого дослідження. І найголовніше ці результати можуть бути використані в майбутньому при покращенні інших алгоритмів, таких як наприклад використовує застосунок Shazam.

У першому розділі зроблено аналіз предметної області та літератури пов’язаної з цією роботою. Таким чином ми дізналися що класифікація музики по музичним стилям може бути досить розпливчастою.

Через те що досі використовується метод ручного маркування стало зрозуміло актуальність цієї роботи. В даний час музична класифікація в основному включає класифікацію тексту та класифікацію на основі музичного змісту. Класифікація тексту в основному базується на інформації про музичні метадані, наприклад про співака, тексти пісень, автора пісень, вік, назву музики та іншу позначену текстову інформацію. Перевагами цього методу класифікації є легкість у застосуванні, простота експлуатації та швидкість пошуку, але недоліки також очевидні.

Перш за все, цей метод покладається на музичні дані, позначені вручну, що

потребує багато робочої сили, а ручне маркування важко уникнути проблем із неправильним маркуванням музичної інформації. По-друге, цей текстовий метод не включає звукові дані самої музики.

Аудіодані включають багато ключових характеристик музики, таких як висота, тембр і мелодія. Ці характеристики майже неможливо позначити текстом; і на основі класифікації вмісту витягуються ознаки вихідних музичних даних, і витягнуті дані ознак використовуються для навчання класифікатора, щоб досягти мети класифікації музики.

Поява глибокого навчання перенесла технологію класифікації музики на новий етап розвитку. Глибоке навчання широко використовується в обробці зображень, розпізнаванні мовлення та інших сферах, і його продуктивність у багатьох завданнях перевершує традиційні методи машинного навчання. Взагалі тема автоматичного розпізнавання жанрів не нова. Люди в різних країнах використовують штучні методи для оцінки музичних жанрів і стилів з 1990-х років.

В даний час широко використовувані характеристики музичних сигналів, що включають в основному короткочасну швидкість переходу через нуль, короткочасну енергію, коефіцієнт лінійного передбачення, частотний спектр, потік, зворотний коефіцієнт частоти Мел, центроїд спектру та спектральний контраст. Оскільки ці характеристики знаходяться як у часовій, так і в частотній області, вони можуть певною мірою відображати характеристики музичного сприйняття висоти, ритму, тембру та гучності.

У другому розділі були детально досліджені можливі способи отримання музичних функцій, а саме:

1. Кепстральний коефіцієнт Мела – походить від автоматичного розпізнавання мови, де він був використаний із великим успіхом. Кепстральний коефіцієнт Мела певною мірою створено відповідно до принципів слухової системи людини, але також мають бути компактним представленням амплітудного спектру та з урахуванням обчислювальної складності. Мел-спктральний аналіз хоч і програє з деякими моделями, але його все одно вважають одним із кращих способів отримання музичних функцій.

2. Частота тону є одним з основних способів класифікації музики. Аудіосигнал, що складається з різних тонів, можна розглядати як звукову послідовність. Коливання тембру містить емоції композитора під час створення музичного твору, а тон визначається частотою тону. Частота тону – це по суті голос музики; тому це дуже важливий параметр обробки сигналу.

3. Пік резонансу - це інша назва форманта. Він відноситься до явища, при якому енергія, що міститься в певному звуковому каналі, збільшується в результаті явища резонансу звукового сигналу.

4. Розподіл енергії в діапазоні частот відноситься до розподілу енергії, якою володіє сегмент аудіо сигналу, який містить таку інформацію, як сила та частота звукового сигналу.

5. Застосування багатоагентних систем до експресивного виконання музики.

У третьому розділі пропонується моделі різних систем розпізнавання музикальних стилів. Перша це рекурсивна нейронна мережа нейронна двостороння мережа, зокрема, підсумовує дані часової області, щоб модель могла дізнатися про часову послідовність музики. Оскільки музичні характеристики музичного твору можуть мати різний вплив на музичну категорію в різний час, механізм уваги використовується для призначення різної ваги-уваги вихідним сигналам циклічної нейронної мережі в різний час і для поєднання послідовних характеристик.

Рекурсивна нейронна мережа може фіксувати внутрішню структуру, приховану в послідовності з часом. Сам звуковий сигнал можна розглядати як часову послідовність. Використання рекурсивної нейронної мережі для обробки музики може фіксувати просторову залежність аудіосигналу у часовому вимірі. Звуковий спектр також розширений у часовому вимірі. Карту ознак після одновимірної згортки можна розглядати як послідовність ознак часу, тому використання рекурсивної нейронної мережі для обробки характеристик звукового спектру також може відігравати ту ж роль. Щоб краще охопити різноспрямовану залежність у вимірі часу в послідовності музичних функцій і бути близькою до сприйняття музики мозком, буде використовуватися двостороння рекурсивна нейронна мережа для моделювання [45] музичної послідовності.

Інша нейронна мережа це згоркова нейронна мережа. Триканальне матричне подання зображення подається в згорткову нейронну мережу, яка навчена прогнозувати клас зображення. У цій роботі звукова хвиля може бути представлена у вигляді спектрограми, яку, у свою чергу, можна розглядати як зображення. Завдання рекурсивної нейронної мережі – за допомогою спектрограми передбачити жанрову мітку (один із дев'яти класів).

Після створення двох нейронних мереж їх було навчено на одному датасеті задля того щоб зрозуміти яка нейронна мережа краще підходить задля того щоб визначати музикальний стиль. Задля цього було нейромережі було навчено п'ять разів задля отримання найвищого результату. В результаті цих експериментів було запропоновано систему розпізнавання музичних стилів з використанням нейромережі яка показала найкращій результат.

Четвертий розділ присвячений розробці двох систем для розпізнавання стилю музики, а також результатам роботи. В ході четвертого розділу було розроблено дві системи, згадані у третьому розділі, а саме: згорткова нейронна мережа, та рекурсивна нейронна мережа. Для навчання нейронної мережі класифікації був використаний датасет GTZAN для обох видів нейронних мереж.

Після визначення найкращої системи для використання було створено інтелектуальну систему розпізнавання музичних стилів.

НЕЙРОННІ МЕРЕЖІ, КЛАСИФІКАЦІЯ, МУЗИЧНІ СТИЛІ, ЗГОРТКОВІ НЕЙРОННІ МЕРЕЖІ, РЕКУРСИВНІ НЕЙРОННІ МЕРЕЖІ, ІНТЕЛЕКТУАЛЬНА СИСТЕМА, ДАТАСЕТ НАВЧАННЯ

Публікації здобувача за темою роботи:

1. Ушаков М. Модель програмування Map-Reduce для користувацьких обчислювальних машин. *Інформаційні технології в сучасному світі: дослідження молодих вчених:* Міжнародна науково-практична конференція молодих учених, аспірантів та студентів тези доповідей, (м. Харків, лютий 2020 р.). – Харків: ХНЕУ імені Семена Кузнеця, 2020. С.7.

2. Аксак Н., Ушаков М. Проектування мультиагентів з використанням платформи JADE *“Інформаційні технології та системи”.*Матеріали Міжнародної науково-практичної конференції: тези доповідей, 8-9 квітня 2021 р. – Х.: ХНЕУ імені Семена Кузнеця, 2021. – С.4.

3. Ушаков М. Мультиагентні системи для моделювання компонентів інтернету речей. *Інформаційні технології в сучасному світі: дослідження молодих вчених:* Міжнародна науково-практична конференція молодих учених, аспірантів та студентів тези доповідей, (м. Харікв, 17 – 18 лютого 2022 року). – Харків: ХНЕУ імені Семена Кузнеця, 2022. С.4.

4. Ушаков М. Огляд методів синхронізації та управління багатоагентними системами. *Методи та засоби обчислювального інтелекту:* матеріали XXIV міжнар. молод. форуму, м. Харків, 7-9 квіт. 2020 р. Харків, 2020. С. 175-176.

5. Axak N., Korablyov M., Ushakov M. Cloud Architecture for Remote Medical Monitoring // IEEE Proceedings of the 15th International conference «Computer Sciences and Information Technologies», (CSIT-2020). – Zbarazh-Lviv, Ukraine, September 23-26, 2020. – vol. 1, pp. 344-347.

6. Axak N., Serdiuk N., Ushakov M., Korablyov M. Development of System for Monitoring and Forecasting of Employee Health on the Enterprise. //COLINS. – 2020. – С. 979-992.

7. Axak N., Korablyov M., Ushakov M. The Development of a Multi-Agent System for Controlling an Autonomous Robot // Proceedings of the 17th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume I, (ICTERI-2021). – Kherson, Ukraine, 2021. – vol. 1, pp. 96-105.

ЗМІСТ

ПЕРЕЛІК УМОВНИХ ПОЗНАЧЕНЬ, СИМВОЛІВ, ОДИНИЦЬ, СКОРОЧЕНЬ І ТЕРМІНІВ

RNN – рекурсивна нейронна мережа

CNN – згорткова нейронна мережа

МЧКК – Мел-частотні Кепстральні Коефіцієнти

ДКП – дискретне косинусне перетворення

ВАФ – виявлення автокореляції

РСАФ – різниця середньої амплітуди

RGB – червоний зелений синій

МАС – мультиагентна система

ВСТУП

Технологія класифікації музичних стилів може додавати теги стилів до музики на основі вмісту. Коли справа доходить до дослідження та впровадження таких аспектів, як ефективна організація, пошук та рекомендації щодо музичних ресурсів, це дуже важливо.

Традиційні методи класифікації музичних стилів використовують широкий спектр акустичних характеристик. Розробка характеристик вимагає музичних знань, а характеристики різних завдань класифікації не завжди узгоджені. Швидкий розвиток нейронних мереж і технологій великих даних надав новий спосіб кращого розв'язання проблеми класифікації музичних стилів.

У цій роботі пропонується методи визначення музичних стилів за допомогою різних типів нейронних мереж, та їх порівняльна характеристика задля визначення найкращого типу нейронної мережі для визначення.

Алгоритм класифікації музичних стилів виділяє два типи ознак як класифікаційні характеристики музичних стилів: темброві та мелодичні ознаки, тому що метод класифікації на основі згорткової нейронної мережі ігнорує час аудіо. В результаті запропоновано модель музичної класифікації на основі одновимірної згортки повторюваної нейронної мережі, яку поєднано з одновимірною згорткою та двосторонньою рекурентною нейронною мережею.

Для кращого представлення властивостей музичного стилю, до виводу застосовуються різні ваги.

1. АНАЛІЗ ПРЕДМЕТНОЇ ОБЛАСТІ ТА ЛІТЕРАТУРНИЙ ОГЛЯД

1.1 Аналіз предметної області

Музика – це звуковий сигнал, що складається з певного ритму та мелодії, гармонії або злиття музичних інструментів за певним правилом, і це дивовижне та різнобічне мистецтво, яке містить і зображає майже всі людські емоції[1-3], і в залежності від стилю відрізняється і сприйняття кожної емоції або настрою на який буде впливати музика на людину. Це може бути танці при енергійності музичного стилю або смуток не достатку тієї самої енергійності чи інших показників.

Музика різноманітна; вона складається з різних елементів, таких як поєднання мелодії, ритму та гармонії відповідно до певних правил художніх форм. Розуміння музики різних форм часто вимагає певних базових знань, а не як стандарт класифікації музики, тому майже всі музичні медіа-платформи використовують текстові мітки як основу класифікації музики або пошуку. Музичні ярлики — це текстові дескриптори, які виражають музичні властивості у високих вимірах, наприклад «щасливий» і «сумний» для вираження емоцій, а також «електронний» і «блюзовий» для вираження музичних стилів [4, 5].

Різні характеристики, сформовані унікальними ритмами, тембром, мелодіями та іншими елементами музичних творів, називаються музичними стилями[6-8](хоча і при недостатньому розумінні усіх елементів це поняття може бути досить розпливчатим), такими як поширена рок-музика(мелодія, що має монотонний одноманітний ритм, імпровізаційність, підвищена емоційність та специфічний склад виконавців) [9], класична музика(прозора та ясна мелодія, певне чітке членування музичної тканини) [10], джаз(мелодія характерна швидким темпом і складними імпровізаціями, заснований на зміні гармонії, а не мелодії) та інші.

За останні роки, зі швидким розвитком та інноваціями Інтернету та мультимедійних технологій(це поняття варіюється починаючи з нових музикальних систем створення музики, що для цієї роботи є більш суміжною темою, до

створення інтелектуальних систем моніторингу здоров’я працівників на підприємстві) [11-14], цифрова музика(метод подання звуку як числових значень, які використовують різні формати аудіо-кодування для зберігання аудіо інформації, вони створюються шляхом перетворення аналогових даних у цифрові) [14, 15] давно стала основною формою слухання музики людьми, що також сприяє зростанню попиту на оцінку музики. Музичний стиль зараз є одним з найбільш часто використовуваних атрибутів класифікації для керування та зберігання цифрових музичних баз даних, а також є одним з основних елементів пошуку та класифікації, які використовуються більшістю музичних онлайн-сайтів.

Ефективність ручного методу маркування, що використовується при пошуку ранньої музичної інформації, більше не може задовольнити потреби менеджменту в умовах величезної кількості музичних даних, і це дуже легко може споживати багато робочої сили та часу. Як наслідок, створення алгоритмів класифікації музичних стилів є критичним для досягнення мети автоматичної класифікації музичних стилів [16-18].

Класифікація музичних жанрів [19–21] є важливою галуззю пошуку музичної інформації. Правильна класифікація музики має велике значення для підвищення ефективності пошуку музичної інформації. В даний час музична класифікація в основному включає класифікацію тексту та класифікацію на основі музичного змісту. Класифікація тексту в основному базується на інформації про музичні метадані, наприклад про співака, тексти пісень, автора пісень, вік, назву музики та іншу позначену текстову інформацію. Перевагами цього методу класифікації є легкість у застосуванні, простота експлуатації та швидкість пошуку, але недоліки також очевидні. Перш за все, цей метод покладається на музичні дані, позначені вручну, що потребує багато робочої сили, а ручне маркування важко уникнути проблем із неправильним маркуванням музичної інформації. По-друге, цей текстовий метод не включає звукові дані самої музики. Аудіодані включають багато ключових характеристик музики, таких як висота, тембр і мелодія. Ці характеристики майже неможливо позначити текстом; і на основі класифікації вмісту витягуються ознаки вихідних музичних даних, і витягнуті дані ознак

використовуються для навчання класифікатора, щоб досягти мети класифікації музики. Тому класифікація музики на основі змісту також стала гарячою точкою дослідження в останні роки. Виходячи з цього, напрям дослідження даної статті також базується на змістовній класифікації музики [22].

Класифікація музичних стилів є важливою галуззю у сфері пошуку музичної інформації, яка була глибоко вивчена, оскільки автоматичний алгоритм класифікації музичних стилів має вищезгадану практичну цінність. Цифрова обробка сигналів випадковий процес, теорія музики та інші теорії в основному використовуються в алгоритмічних дослідженнях класифікації музичних стилів для математичного опису та вираження пов'язаних з жанром характеристик у музичних сигналах для формування різних типів музичних характеристик. Алгоритм машинного навчання[23-25] використовується для вивчення характеристик розподілу ознак різних жанрів для отримання класифікатора.

Нарешті, властивість фрагмента аудіо сигналу задається як вхід класифікатора, а стиль класифікатора визначається відповідно до апостеріорної ймовірності. Серед них структура ознаки визначає верхню межу продуктивності алгоритму класифікації, а ефективний метод представлення може максимізувати точність результату класифікації. Тому велика кількість вчених зосереджується на інженерному зв’язку музичних сигналів.

Однак у вивченні музичних сигналів є дві основні труднощі: з одного боку, музика містить складну й абстрактну інформацію, таку як емоції, ритми, інструменти та акорди, які часто важко виразити у штучно створених рисах; з іншого боку, музика містить складну й абстрактну інформацію.

У порівнянні зі звичайними голосовими сигналами музика має складнішу частотну композицію та багатшу темброву інформацію. Тому деякі звичайні методи обробки голосових сигналів неможливо просто застосувати, і потрібно розробляти спеціальні алгоритми відповідно до характеристик музичних сигналів.

За останні роки глибоке навчання [26-28] досягло видатних результатів у сферах зображення [29], мовлення та обробки природної мови. Все більше і більше дослідників намагаються навчитися хорошому вираженню музичних сигналів через

глибокі нейронні мережі, замінюючи попереднє ручне вилучення.

Характеристики підвищення продуктивності алгоритму мають важливе теоретичне значення. На даний момент Spotify, найбільша у світі платформа сервісу справжніх потокових музичних сервісів, успішно застосувала глибоке навчання до своєї системи рекомендацій щодо музики. Тому обробка музичних сигналів на основі глибокого навчання може сприяти розвитку музичних платформ і надавати користувачам кращий досвід обслуговування, а також має величезну економічну цінність і цінність для дослідження.

Поява глибокого навчання перенесла технологію класифікації музики на новий етап розвитку. Глибоке навчання широко використовується в обробці зображень, розпізнаванні мовлення та інших сферах, і його продуктивність у багатьох завданнях перевершує традиційні методи машинного навчання. Вчені також почали використовувати технологію глибокого навчання для вивчення суміжних питань у сфері пошуку музичної інформації, тому необхідно досліджувати методи класифікації музики на основі глибокого навчання, щоб покращити ефект класифікації музики [30].

1.2 Літературний огляд

Різні жанри мають різні музичні стилі, і ідентифікація музичних стилів або музичних жанрів широко досліджувалася з моменту їх виникнення. Люди в різних країнах використовують штучні методи для оцінки музичних жанрів і стилів з 1990-х років. «Проєкт музичної хромосоми» є найбільш відомим.

Основна мета цього проєкту полягає в тому, щоб музичні експерти розділили музику на різні типи на основі їх знань та розуміння музичних технологій. Однак, зіткнувшись із величезною кількістю даних, штучні методи є незрілими через обмежені технічні умови, а різні експерти мають дещо різне розуміння різних музичних жанрів, тому на проєкт витратили багато фінансових та матеріальних ресурсів. Скеровані цією ситуацією, люди почали робити дослідження алгоритмів автоматичної класифікації для розпізнавання музичних жанрів.

У 1995 році Беньямін Матітяхо та Ферст [31] запропонували метод аналізу музичних сигналів у частотній області. Спочатку над аудіоданими виконується швидке перетворення Фур'є, а потім виконується логарифмічне перетворення масштабу, щоб використовувати отримані дані як характерні дані. Навчання проводилося в нейронній мережі [32–33], що містить два прихованих шари, і остаточно ідентифіковано два музичних жанри: класичну та поп-музику.

Пізніше американські дослідники запропонували алгоритм класифікації.

Цей метод переважно обчислює середнє значення, дисперсію та коефіцієнт автокореляції на основі масивних музичних даних, щоб додатково проаналізувати характеристики музики, такі як гучність та висота, які люди можуть легко відчути. Отримані ознаки потім ідентифікуються та класифікуються за допомогою деяких класифікаторів.

Згодом цей алгоритм був значно розкручений, і люди почали намагатися використовувати деякі вдосконалені алгоритми для класифікації музичних жанрів на основі цього алгоритму.

У 2002 році Цанетакіс і Кук [34] надали новий алгоритм класифікації. Цей метод спочатку виділяє акустичні особливості, які в основному містять три типи акустичних особливостей, тембр музики, ритм і зміст висоти; автори прийняли модель суміші Гауса та К.

Як класифікатор використовується метод близькості. Оскільки вилучені акустичні ознаки, як правило, мають більшу розмірність, алгоритм вибору ознак використовується для зменшення розмірності ознак, щоб полегшити обчислення, і водночас видаляється деяка незначна зайва інформація. Варто зазначити, що через численні музичні жанри раніше в академічних колах не існувало відносно фіксованого стандарту класифікації. Після новаторських результатів досліджень Цанетакіса десять музичних жанрів, що містяться в наборі даних GTZAN, використаному Джорджем Цанетакісом, стали музичною інформацією. Стандарт класифікації був загальновизнаним у полі пошуку.

Оскільки результати дослідження Джорджа Цанетакіса заклали для нас велику основу, пізніші вчені в галузі автоматичного розпізнавання музичних жанрів

зосереджувалися переважно на двох аспектах. З одного боку, вони внесли відповідні покращення у вибір вилучення музичних ознак і розмірність векторів ознак. З іншого боку, вони покращили вибір алгоритму класифікації. Виділення музичних ознак є дуже важливою частиною розпізнавання музичного жанру. Якщо виділені ознаки не можуть відображати основні характеристики музики, то ефект класифікації музики, безсумнівно, буде дуже поганим. Скарінгелла та ін. [35] розділив характеристики музичного сигналу на три категорії: висота, тембр і ритм.

Нарешті, деякі моделі та відповідні алгоритми використовуються для визначення та класифікації музичних жанрів.

З розвитком комп'ютерних технологій машинне навчання також почало застосовуватися до класифікації музичних жанрів. У 2003 році Чаншен Ху [36] досліджували класифікацію музичних жанрів, використовуючи різні музичні характеристики.

Порівнявши метод K-найближчого сусіда, умовне випадкове поле та алгоритми моделі Маркова, він виявив, що алгоритм класифікації ефекту розпізнавання SVM є найбільш ефективним.

Застосовуючи теорію вейвлет-перетворення, Тао Лі[37] використовували статистичні методи для розрахунку статистичних значень вейвлет-коефіцієнтів у поєднанні з моделями класифікації, які зазвичай використовуються в машинному навчанні, такими як LDA, GMM та KNN, щоб отримати хороші результати класифікації.

У 2011 році, щоб отримати більш важливі музичні характеристики, Панагакіс [38] вперше запропонував неконтрольований метод зменшення розмірності. Завдяки експериментальним результатам було виявлено, що цей метод має значний ефект у виділенні музичних особливостей порівняно з попередніми методами.

В даний час широко використовувані характеристики музичних сигналів включають в основному короткочасну швидкість переходу через нуль, короткочасну енергію, коефіцієнт лінійного передбачення, частотний спектр, потік, зворотний коефіцієнт частоти Mel, центроїд спектру та спектральний контраст. Оскільки ці характеристики знаходяться як у часовій, так і в частотній області, вони можуть

певною мірою відображати характеристики музичного сприйняття висоти, ритму, тембру та гучності. Процес виділення музичних ознак зазвичай полягає в тому, щоб спочатку виконати кадрову обробку вихідного аудіосигналу, потім виконати відповідні обчислення на основі математичної статистичної значущості ознак і, нарешті, використовувати обчислені результати як навчальні дані класифікатора у формі вектори.

Оскільки виділення музичних ознак базується на аналізі музичного сигналу, поточні методи аналізу музичного сигналу на основі аудіо в основному включають методи аналізу в часовій області та методи аналізу в частотній області. Так званий метод аналізу в часовій області полягає в аналізі та підрахунку стану форми хвилі музичного сигналу за часовим виміром.

Аналіз у частотній області перетворює музичний сигнал у часовій області в частотну за допомогою перетворення Фур'є, тому можна отримати багато корисних функцій у частотній області, наприклад, Мел до загального коефіцієнта, спектральний центроїд, частоту висоти тону, енергію піддіапазону, спектрограму , і т.д. Стаття [39] показує разом коефіцієнт Mel-to-Pop і частоту висоти, центроїд спектру, енергію піддіапазону та інші перцептивні характеристики для формування високовимірного вектора ознак. В алгоритмі класифікації музики в основному використовуються традиційні методи машинного навчання, такі як опорні векторні машини, моделі суміші Гауса, дерева рішень, найближчі сусіди, приховані Маркова та штучні нейронні мережі [40, 41]. Крім того, є деякі вдосконалення вищевказаних алгоритмів. Наприклад, стаття [42] додає генетичний алгоритм до моделі суміші Гауса, що покращує точність класифікації за експериментальними результатами.

Таким чином, метою даної роботи є розробка системи розпізнавання музикальних жанрів за допомогою нейронних мереж.

Для досягнення поставленої мети необхідно виконати наступні завдання:

1. Визначити способи вилучення музичних функцій для подальшого використання в навчанні нейронних мереж.

2. Створити моделі для нейронних мереж різних типів задля визначення кращого типу нейромережі для визначення музичного стилю.

3. Навчити моделі використовуючи визначені музичні функції при обробці музичних даних датасету.

4. Зробити порівняльну характеристику різних нейронних мереж після закінчення навчання, задля розуміння яка з представлених систем підходить для цього найраще.

2. ОПИС МОДЕЛІ ТА РОЗРОБКА СТРУКТУРИ СИСТЕМИ КЛАСИФІКАЦІЇ СТИЛІВ МУЗИКИ

2.1 Опис музикальних елементів

Музика містить три елементи: висоту, ритм і тембр. Мелодія і гармонія музики можуть утворюватися шляхом поєднання і перетворення висоти; темп пов'язаний з артикуляцією, яка контролює швидкість і перехід музики; тембр — це якість звуку для сприйняття звуку, що використовується для розрізнення різних типів звуків для створення нот, і кожен інструмент має свій унікальний тембр.

Поєднання цих трьох елементів може утворити інші елементи в музиці. Наприклад, декілька різних звуків, які грають одночасно, стають гармонією, а скоординований ефект, отриманий різними висотами звуку в різний час, стає мелодією.

Ці елементи за допомогою різних поєднань формують унікальний музичний стиль, передаючи радість, хвилювання, смуток та інші емоції, утворюючи, таким чином, різні жанри.

Вокальна музика, яка ґрунтується на вокальному співі, та інструментальна музика, яка ґрунтується на виконанні інструментів, є двома основними видами музики. Прикладами цих двох форм є хор і соло у вокальній музиці, а також соло, концерт та симфонія в інструментальній музиці. Інструментальну та вокальну музику можна поєднувати для створення широкого діапазону музичних жанрів.

2.1.1 Вилучення музичних функцій

Існує велика кількість функцій, які можна отримати з аудіосигналу.

Наприклад, у [43] виділено чотири групи ознак:

1) Короткочасні характеристики витягуються з коротких вікон часу, порядку 10-100 мс, протягом яких аудіосигнал вважається стаціонарним. Ці

особливості описують елементарні компоненти музики, такі як тон або тембр, отримані з музичного інструменту.

2) Довгострокові ознаки описують довші сегменти музичної композиції, які отримані шляхом поєднання вікон короткочасних ознак. Такі знаки дозволяють отримати інформацію про ритмічні та часові особливості композиції.

3) До семантичних ознак належать ознаки, які мають пряме тлумачення для кінцевого користувача: темп, тональність, настрій та інші семантичні властивості музики.

4) Композиційні особливості характеризують структуру композиції в цілому. Це може бути як розподіл інтервалів висоти і гучності тону, так і інтервали зміни тональності і ритму.

В аудіоаналізі виділення ознак – це процес вилучення життєво важливої інформації з (фіксованого розміру) часового інтервалу оцифрованого аудіосигналу. Математично, вектор ознак x_n у дискретний час n можна обчислити за допомогою функції F на сигналі s як:

$$x_n = F\left(w_0 s_{n-(N-1)}, \dots, w_{N-1} s_n\right), \quad \beta \in \mathbb{R}^{10} \tag{2.1}$$

де $w_0, w_1, \dots, w_{N-1}$ — коефіцієнти віконної функції; N — розмір кадру.

Розмір кадру є мірою масштабу часу функції. Зазвичай не обов'язково мати x_n для кожного значення n, тому між кадрами використовується розмір переходу M.Весь процес показано на рисунку 2.1. З точки зору обробки сигналу, використання розміру стрибка означає зменшення дискретизації сигналу xn, який тоді містить лише члени:

$$\dots, x_{n-2M}, x_{n-M}, x_n, x_{n+M}, x_{n+2M}, \dots \tag{2.2}$$

Функція вікна множиться на сигнал, щоб уникнути проблем через кінцевий розмір кадру.

Прямокутне вікно з амплітудою 1 відповідає обчисленню характеристик без

вікна, але має серйозні проблеми з явищем спектрального витоку і рідко використовується. Для цього може бути використано так зване вікно Хеммінга, яке має бічні пелюстки зі значно нижчою величиною, але могли бути використані інші функції вікна.

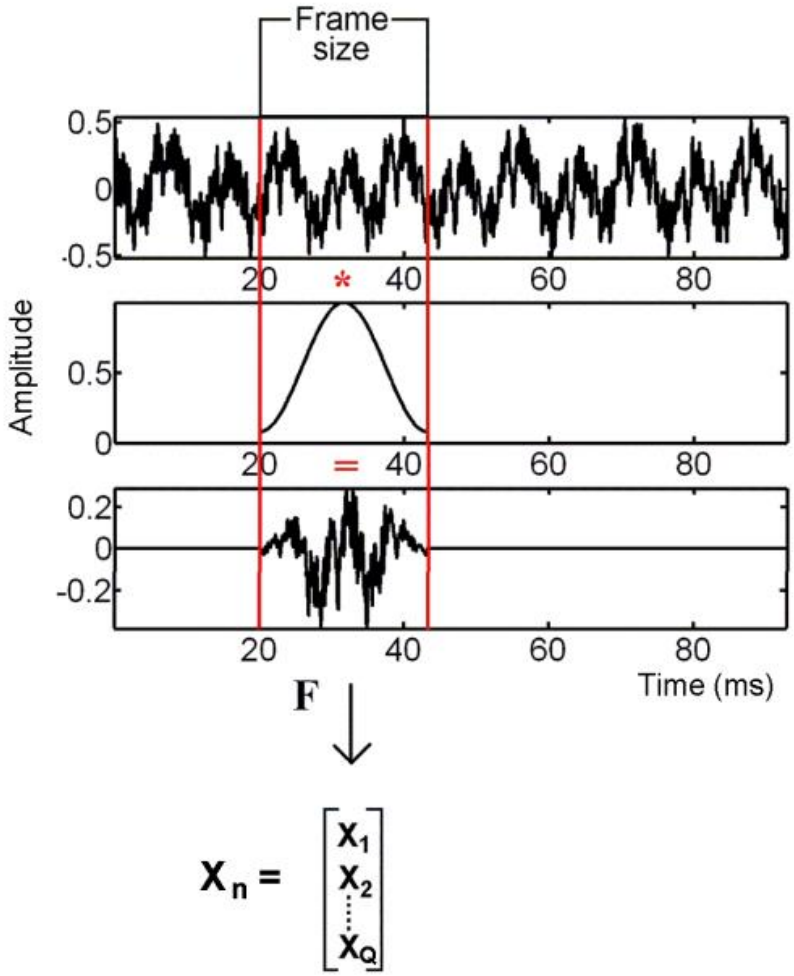

Рисунок 2.1 - Ілюстрація традиційного процесу виділення ознак

На рисунку 2.2 показано результат дискретного перетворення Фур'є для сигналу з вікном Хеммінга та без нього, і переваги вікна Хеммінга легко побачити. Вікно Хеммінга можна знайти

$$w_n = 0.54 - 0.46\,cos\left(\frac{2\pi n}{N-1}\right) \qquad n = 0, \dots, N-1 \tag{2.3}$$

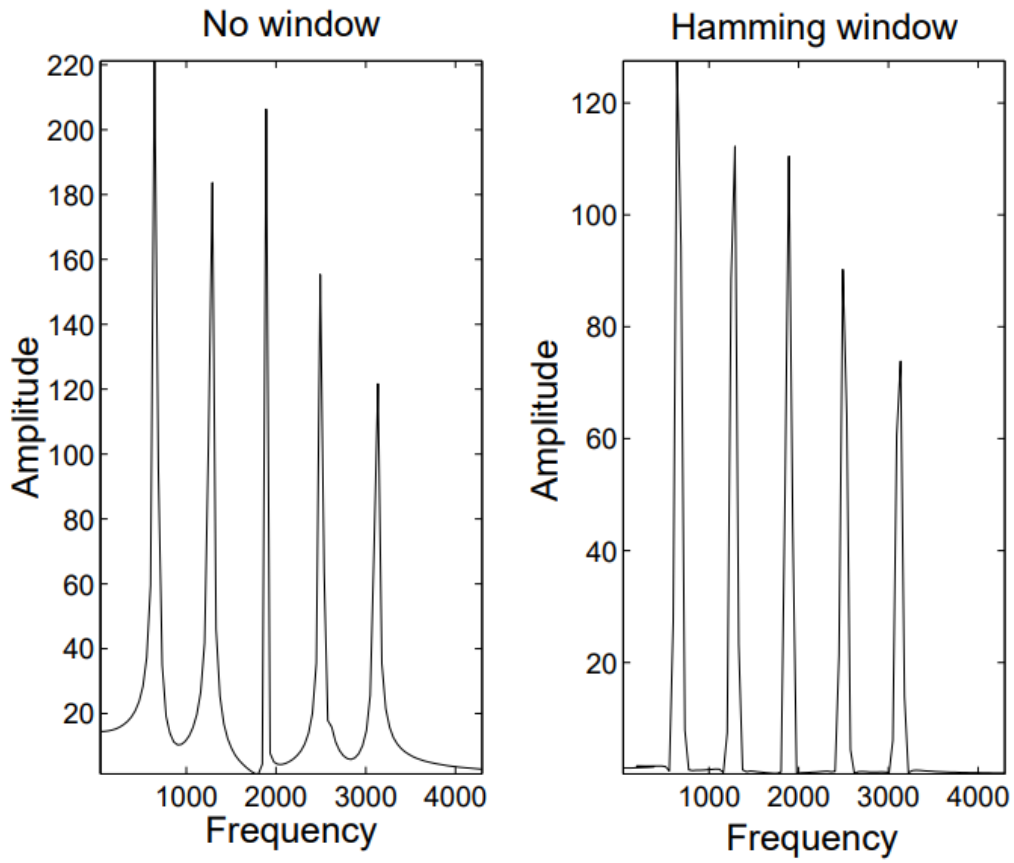

Рисунок 2.2 – Частотний спектр гармонічного сигналу з основною частотою та чотирма обертонами

У цій роботі використовується кілька методів виділення характеристик для вилучення тембрових особливостей нижньої музики (коефіцієнти кепстру Мела) та особливостей мелодії середніх музичних характеристик (частота висоти, форманта та енергія смуги) з вихідного звукового сигналу, а потім буде складається з цих ознак.

Навчальний набір пояснює, як використовувати навчальну систему класифікації для підвищення точності системи класифікації.

2.1.2 Кепстральний коефіцієнт Мела

Мел-Частотні Кепстральні Коефіцієнти (МЧКК) походить від автоматичного розпізнавання мови [44], де він був використан із великим успіхом. Спочатку він був запропоновані в [45]. Вони стали дуже популярними в спільноті Пошуку Музичної

Інформації, де його успішно використовували [46, 47] для класифікації музичних жанрів, і для категоризації в релевантні для сприйняття групи, такі як настрій і сприймана складність[48].

МЧКК певною мірою створені відповідно до принципів слухової системи людини [49], але також мають бути компактним представленням амплітудного спектру та з урахуванням обчислювальної складності. У [50] стверджується, що вони моделюють тембр у музиці. [51] порівнюють їх із слуховими характеристиками з більш точними (і вимогливими до обчислень) моделями, але все одно вважають МЧКК кращими.

Рисунок 2.3 – Ілюстрація розрахунку мел-частотних кепстральних коефіцієнтів (МЧКК)

Рисунок 2.3 ілюструє побудову функцій МЧКК. Відповідно до рівняння 2.1, виділення ознаки можна описати як функцію F на кадрі сигналу. Після застосування вікна Хеммінга до фрейму ця функція містить наступні 4 кроки:

1. Дискретне перетворення Фур'є

Першим кроком є виконання дискретного перетворення Фур’є на кадрі. Для розміру кадру N це призводить до N (комплексних) коефіцієнтів Фур’є. Ця фаза зараз відкидається, оскільки вважається, що вона мало цінна для людського розпізнавання мови та музики. Це призводить до N-вимірного спектрального представлення кадру.

2. Масштабування Мела

Люди впорядковують звуки в музичній шкалі від низького до високого за допомогою перцептивного атрибута, який називається висотою. Висота тону синуса

тісно пов’язана з фізичною величиною частоти та основною частотою для складного тону. Проте шкала висоти не має такого ж інтервалу, як частотна шкала. Мел-шкала — це оцінка співвідношення між сприйнятою висотою тону та частотою, яка визначається шляхом прирівнювання 1000 мел до синусового тону 1000 Гц при 40 дБ. Він використовується в обчисленні МЧКК для перетворення частот у спектральному представленні в перцептивну шкалу тону. Зазвичай крок мел-масштабування має форму групи фільтрів (перекриваються) трикутних фільтрів у частотній області та з центральними частотами, які рознесені мел. Стандартний набір фільтрів показано на малюнку 2.4. Отже, цей етап мел-масштабування також є згладжуванням спектра та зменшенням розмірності вектора ознак.

3. Визначення гучності за допомогою перцептривного атрибуту

Подібно до висоти, люди впорядковують звук від тихого до гучного за допомогою перцептивного атрибута гучності. Перцептивна гучність досить близько відповідає фізичній мірі інтенсивності. Хоча інші величини, такі як частота, смуга пропускання та тривалість, впливають на сприйману гучність, прийнято пов’язувати гучність безпосередньо з інтенсивністю. Таким чином, співвідношення часто апроксимується як $L \propto I^{0.3}$, де L – це гучність, а I – інтенсивність (степеневий закон Стівенса). Це стверджується, напр. [49], що перцептивна гучність також може бути апроксимована логарифмом інтенсивності, хоча це не дуже схоже на згаданий раніше степеневий закон. Це перцептивна мотивація для етапу логарифмічного масштабування при вилученні МЧКК. Ще одна мотивація для логарифмічного масштабування в аналізі мови полягає в тому, що його можна використовувати для деконволюції повільно змінної модуляції та швидкого збудження з періодом висоти тону [52].

4. Дискретне косинусне перетворення

На останньому етапі дискретне косинусне перетворення (ДКП) використовується як обчислювальне недорогий метод декореляції мел-спектральних логарифмічних коефіцієнтів. Це свідчить про те, що ДКП фактично можна використовувати для декореляції. Зазвичай використовується лише підмножина базисних функцій ДКП, і результатом є ще нижчий вектор ознаки МЧКК.

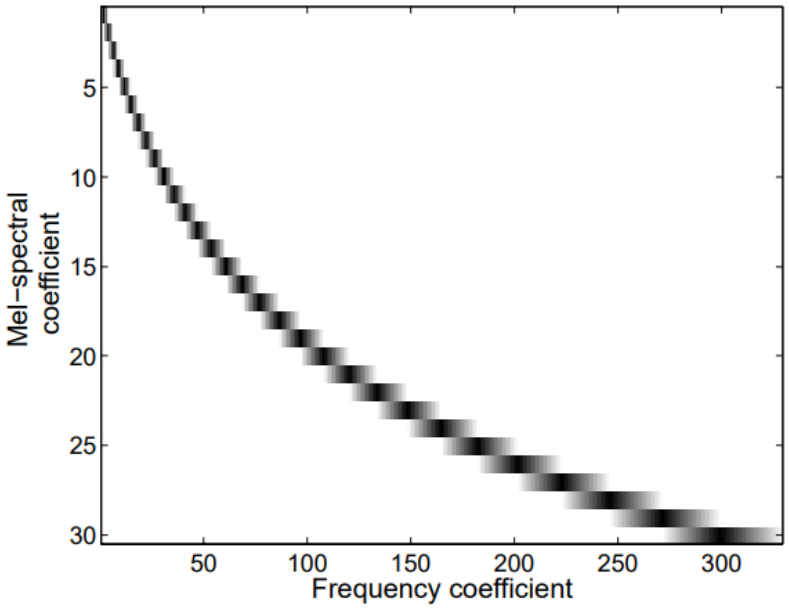

Рисунок 2.4 – Ілюстрація матриці, яка використовується для перетворення лінійної частотної шкали в логарифмічну Мел-шкалу під час обчислення частотних кепстральних коефіцієнтів Мела

Кепстральний коефіцієнт Мела має хорошу протишумну здатність і високу швидкість розпізнавання. У сучасних дослідженнях мовного сигналу він став широко використовуваним характеристичним параметром[28]. Спочатку імпортується аудіосигнал, виконується обробка кадру та вікно для сигналу та використовується перетворення Фур'є для перетворення сигналу часової області в сигнал частотної області:

$$x(k) = \sum_{k=0}^{N-1} x(n) e^{-\frac{j2\pi nk}{N}}, 0 \leq k \leq N-1. \quad (2.4)$$

Вхідний сигнал представлений *x*, а потужність вхідного сигналу при n представлена $x(x), n = 0,1, \dots, J$ де J — довжина сигналу. У дискретному перетворенні Фур'є кількість точок, які воно виконує, позначається *N*.

Енергетичний спектр розрахується, а потім передається. Для реалізації методу передачі використовується набір трикутних фільтрів масштабу Мела. Ключовим параметром цієї форми фільтра є центральна частота, яка позначається як *f(m)*. Для

кожного трикутного фільтра вихідна енергія групи обчислюється і виражається в логарифму:

$$S(m) = \ln(\sum_{k=1}^{N-1}|x(k)|^2 H_m(k)), 0 \le m \le M-1. \quad (2.5)$$

По-друге, щоб обчислити параметри МЧКК, спосіб досягнення цього полягає в виконанні дискретного косинусного перетворення (ДКП):

$$C(n) = \sum_{m=0}^{N-1} S(m) \cos\left(n\pi\left(m - \frac{0.5}{m}\right)\right), n = 1,2,\dots,L. \quad (2.6)$$

2.1.3 Частота тону

Аудіосигнал, що складається з різних тонів, можна розглядати як звукову послідовність. Коливання тембру містить емоції композитора під час створення музичного твору, а тон визначається частотою тону. Частота тону - це голос; тому це дуже важливий параметр обробки сигналу. Вилучення звукової частоти має враховувати короткочасну стабільність сигналу від людини, що говорить.

В даний час найпоширенішими методами є: виявлення автокореляції (ВАФ), різниця середньої амплітуди (РСАФ), видалення піків тощо.

У цій роботі метод визначення автокореляційної функції вибирається для вилучення частоти тону з огляду на стабільність і плавність сигналу висоти. Короткочасна автокореляційна функція *Rn(k)* мовного сигналу *s(m)* визначається як:

$$R_n(k) = \sum_{m=0}^{N-k-1} S_n(m) S_n(m+k), \quad (2.7)$$

де N - довжина вікна, доданого мовним сигналом; $S_n(m)$ – це сегментований мовний сигнал вікна, перехоплений мовним сигналом *s(m)* через вікно з довжиною N і визначається як:

$$S_n(m) = s(m) w(n-m). \quad (2.8)$$

Функція автокореляції основної частини аудіокліпу матиме очевидні піки, а високочастотні тони не очевидні порівняно з основною. Таким чином, визначити, чи

це основний тон чи високочастотний тон, можна, виявивши очевидний пік, а частоту висоти можна витягти, виявивши відстань між сусідніми піками.

2.1.4 Пік резонансу

Резонансна частота — це інша назва форманта. Він відноситься до явища, при якому енергія, що міститься в певному звуковому каналі, збільшується в результаті явища резонансу звукового сигналу.

Голосовий тракт зазвичай можна розглядати як рівномірно розподілену звукову трубку, а резонанс вібрації звукової трубки в різних положеннях є звуковим процесом. Форма форманта зазвичай пов’язана з будовою голосового тракту. Зі зміною структури голосового тракту змінюватиметься і форма форманта. Для сегмента мовного сигналу різні емоції відповідають різним формам каналу.

Тому частота формант може бути використана як важливий параметр розпізнавання емоцій мовного сигналу.

2.1.5 Розподіл енергії в діапазоні частот

Розподіл енергії в діапазоні частот відноситься до розподілу енергії, якою володіє сегмент аудіосигналу, який містить таку інформацію, як сила та частота звукового сигналу. Він має сильну кореляцію з красою та емоціями музики. У галузі музики, аналізуючи характеристики розподілу енергії в діапазоні частот, можна отримати красу та емоційність звукового сигналу. Припустимо, є музичний відрізок довжиною *M*, який містить характеристики голосу різних інструментів і людських голосів. Тепер ми хочемо знайти один із піддіапазонів у частотній області музичного сегмента, від a до $a+N$, який енергийний. По-перше, згідно з перетворенням Фур’є, вихідний музичний сигнал часової області *f(t)* перетворюється на сигнал частотної області *F(t)*:

$$F(t) = \int f(t)e^{-jwt}dt. \tag{2.9}$$

Розподіл енергії E в діапазоні частот дорівнює:

$$E = \frac{1}{N}\sum_{a}^{a+N} |F(t)|^2. \quad (2.10)$$

2.1.6 Застосування багатоагентних систем до алгоритмічної композиції та експресивного виконання музики

Окрім дослідження корисності застосування імітаційної багатоагентної парадигми [56–62], також доцільним є дослідження різноманітністей в цій методології: особливо для творчого застосування, такого як музичне виконання.

Властивості багатоагентної системи (MAS) для комп'ютеризованої композиційної системи (ККС) виконання дозволяють генерувати виразне виконання як частину композиційного процесу та створюють нові мелодичні структури [63]. ККС складається з набору агентів малого та середнього розміру (від 2 до 16), у якому кожен агент може виконувати монофонічні мелодії та вивчати монофонічні мелодії від інших агентів. Кожен агент має афективний стан («штучний емоційний стан»), який впливає на те, як він виконує музику для інших агентів; наприклад, «щасливий» агент буде виконувати «веселішу» музику.

Виконання агента передбачає не лише композиційні зміни в музиці, але й додає менші зміни на основі алгоритмів експресивного виконання музики для гуманізації. Кожен агент ініціалізується мелодією, що містить ту саму одну ноту, і протягом періоду взаємодії створюються довші мелодії за допомогою взаємодії агента. Агенти вивчатимуть мелодії, які їм виконують інші агенти, лише якщо афективний зміст мелодії схожий на їхній поточний афективний стан; вивчені мелодії об'єднуються до кінця поточної мелодії. Кожен агент у суспільстві вивчає свою власну зростаючу мелодію під час процесу взаємодії.

Агенти формують «думку» інших агентів, які виконують для них, залежно від того, наскільки виконавець може сприяти розвитку їхніх мелодій. Ці думки впливають на те, з ким вони будуть спілкуватися в майбутньому. ККС не є відображенням взаємодії кількох агентів із музичними функціями, а фактично

використовує музику для передачі агентами емоцій. Незважаючи на відсутність явного мелодичного інтелекту в ККС, показано, що система генерує нетривіальну висоту мелодії в результаті емоційного спілкування між агентами. Мелодії також мають ієрархічну структуру, засновану на виниклій соціальній структурі багатоагентної системи, а ієрархічна структура є результатом нової структури соціальної взаємодії агента.

Мелодії також мають ієрархічну структуру, засновану на виниклій соціальній структурі багатоагентної системи, а ієрархічна структура є результатом нової структури соціальної взаємодії агента. Інтерактивні гуманізації створюють мікровідхилення синхронізації та гучності в мелодії, які, як показано, виражають її ієрархічну генеративну структуру без потреби в програмному забезпеченні структурного аналізу, яке часто використовується в комп’ютерній гуманізації музики.

3. РОЗРОБКА ТА АНАЛІЗ НЕЙРОННИХ МЕРЕЖ ДЛЯ СТВОРЕННЯ ІНТЕЛЕКТУАЛЬНОЇ СИСТЕМИ

3.1 Опис рекурсивної нейронної мережі глибокого навчання

Діапазон звуків є послідовним у часовому вимірі, а інформація про час у музиці ігнорується завдяки простому використанню структури згортки.

Одновимірна конфігурація має місце у вимірі часу, а також ігнорує взаємозв'язок послідовності між властивостями звукового спектру різних часових кадрів, фіксуючи локальні характеристики звукового спектру. Відношення музичної послідовності не можна ефективно моделювати лише за допомогою одновимірної згортки.

Таким чином, можна об'єднати запропоновану структуру згортки з однією ДНК та двосторонньою рекурентною нейронною мережею та зробити модуль класифікації на основі повторюваної мережі однієї ДНК та використати механізм уваги для переміщення нейронної мережі в різний час. Вихід має різну вагу уваги, тому характеристики музичного стилю краще представлені

Рекурентна нейронна двостороння мережа, зокрема, підсумовує дані часової області, щоб модель могла дізнатися про часову послідовність музики. Оскільки музичні характеристики музичного твору можуть мати різний вплив на музичну категорію в різний час, механізм уваги використовується для призначення різної ваги-уваги вихідним сигналам циклічної нейронної мережі в різний час і для поєднання послідовних характеристик.

RNN може фіксувати внутрішню структуру, приховану в послідовності з часом. Сам звуковий сигнал можна розглядати як часову послідовність. Використання RNN для обробки музики може фіксувати просторову залежність аудіо сигналу у часовому вимірі. Звуковий спектр також розширений у часовому вимірі. Карту ознак після одновимірної згортки можна розглядати як послідовність ознак часу, тому використання RNN для обробки характеристик звукового спектру

також може відігравати ту ж роль. Щоб краще охопити різноспрямовану залежність у вимірі часу в послідовності музичних функцій і бути близькою до сприйняття музики мозком, буде використовуватися Bi-RNN для моделювання[45] музичної послідовності.

Bi-RNN не тільки враховує попередні вхідні дані, але й останні вхідні дані також можуть бути корисними для моделювання даних. На рисунку 2.5 показана структура Bi-RNN. Підрахунок ваг цієї мережі виглядає так:

$$\overleftarrow{H}^i = f(W'X^i + V'\overleftarrow{H}^{i+1}) \tag{3.1}$$

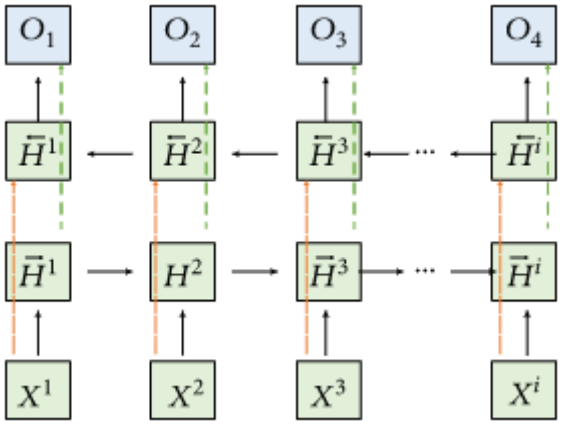

Рисунок 2.5 – Структура мережі Bi-RNN

Далі потрібно додати пряму і задню частину кожного кроку мережі, щоб отримати кінцевий мережевий результат:

$$O_i = U\vec{H}^i + U'\overleftarrow{H}^i. \tag{3.2}$$

Для категорії музики, що відповідає цій функції, конкретна функція звукового спектру, яка з’явилася під час різних музичних часів, може відрізнятися. Механізм фокусування може кожен раз обчислювати вагу послідовності характеристик і кожного разу зважено підсумовувати характеристики за вагою.

Пропонується модель уваги з послідовною структурою, як показано на рисунку 2.6. Оскільки вихідний сигнал O_i Bi-RNN представляє подання ознак, засвоєне в моделі класифікації. Модель уваги використовує лінійне перетворення для обчислення показника уваги. Формула розрахунку виглядає так:

$$e_i = w_i^T O_i \,, \tag{3.3}$$

де e представляє оцінку уваги, присвоєну i-му вектору ознак,
O_i — i-й вектор ознак,

$$E = [e_1, e_2, \dots, e_T] \tag{3.4}$$

Потім отримана оцінка уваги нормалізується, щоб створити розподіл ймовірності уваги на представленні ознаки:

$$a_i = soft\max(E) = \frac{\exp(e_i)}{\sum_{j=1}^{T} \exp(e_j)}\,, \tag{3.5}$$

де a_i представляє ймовірність уваги, присвоєну моделлю i-му вектору ознак.

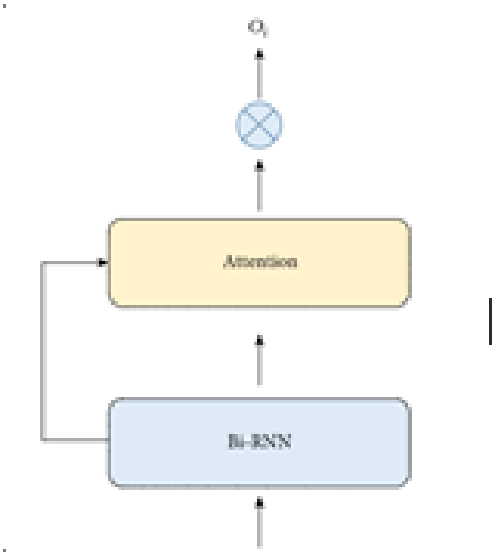

Рисунок 3.2– Схема механізму уваги послідовної структури

Після того, як згортковий шар навчиться з різних звукових спектрів, тепер

можна отримати карти ознак з абстрактними ознаками високих рівнів.

Функціональні карти можуть бути розширені в часі для досягнення послідовностей перетворювальних ознак, а послідовності згортки для моделювання музичної послідовності вводяться в Bi-RNN. Потім, використовуючи вагу фокуса, мережа навчиться зважувати послідовність функцій, яку виконує Bi-RNN у підсумку, кілька разів інтегруючи вихідні дані Bi-RNN і переводити його на повністю підключений шар музики.

3.2 Згорткова нейронна мережа глибокого навчання

Використовуючи глибоке навчання ми можемо виконати завдання класифікації музичних жанрів без необхідності ручного вводу. Згорткові нейронні мережі (CNN) широко використовуються для завдання класифікації зображень. Триканальне (RGB) матричне подання зображення подається в CNN, яка навчена прогнозувати клас зображення. У цій роботі звукова хвиля може бути представлена у вигляді спектрограми, яку, у свою чергу, можна розглядати як зображення. Завдання CNN – за допомогою спектрограми передбачити жанрову мітку (один із дев’яти класів).

Спектрограма — це двовимірне представлення сигналу, що має час на осі абсцис і частоту на осі y. Колірна карта використовується для кількісного визначення величини даної частоти в межах заданого часового вікна. У цьому дослідженні кожен аудіосигнал перетворювався на Мел-спектрограму (з частотними розділами на осі Y).

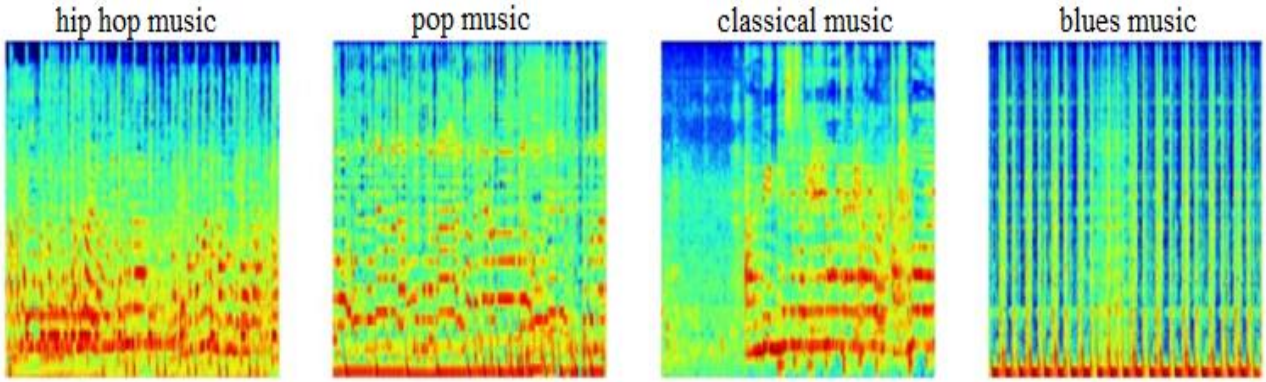

Рисунок 3.3 – Приклад Мел-спектрограм сгенеровано під час створення нейронної мережі

З рисунка 3.3 можна зрозуміти, що в спектрограмах звукових сигналів, що належать до різних класів, існують деякі характерні закономірності. Отже, спектрограми можна розглядати як «зображення» та надавати як вхідні дані для CNN, яка показала хорошу продуктивність у задачах класифікації зображень. Кожен блок у CNN складається з наступних операцій:

• Згортка: цей крок включає ковзання матричного фільтра (скажімо, розміром 3x3) над вхідним зображенням, яке має розміри: *ширина зображення* * *висота зображення*. Фільтр спочатку розміщується на матриці зображення, а потім ми обчислюємо поелементне множення між фільтром і частиною зображення, що перекривається, з подальшим підсумовуванням, щоб отримати значення ознаки. Ми використовуємо багато таких фільтрів, значення яких «вивчаються» під час навчання нейронної мережі через зворотне поширення.

• Об’єднання: це спосіб зменшити розмірність карти ознак, отриманої в результаті кроку згортки, формально відомого як процес зменшення вибірки. Наприклад, шляхом максимального об’єднання з розміром вікна 2x2 ми зберігаємо лише елемент із максимальним значенням серед 4 елементів карти функцій, які охоплені цим вікном. Ми продовжуємо переміщати це вікно по карті функцій із заздалегідь визначеним кроком.

Нелінійна активація: операція згортки є лінійною, і щоб зробити нейронну мережу потужнішою, нам потрібно ввести деяку нелінійність. Для цього ми можемо застосувати функцію активації.

Модель складається з 5 згорткових блоків, за якими йде зведений шар, який перетворює двовимірну матрицю в одновимірний масив, за яким слідує повністю зв’язаний шар, який виводить ймовірність того, що задане зображення належить кожному з можливі заняття.

Останній рівень нейронної мережі виводить ймовірності класу для кожної з восьми можливих міток класу. Втрата крос-ентропії обчислюється, як показано

нижче:

$$H = \sum_{c=1}^{M} y_{o,c} * \log p_{o,c} \quad (3.6)$$

де M – кількість класів;

$y_{o,c}$ — двійковий показник, значення якого дорівнює 1, якщо спостереження o належить до класу c, і 0 в іншому випадку;

$p_{o,c}$ – передбачена моделлю ймовірність того, що спостереження o належить до класу c.

Ці втрати використовуються для зворотного поширення помилки, обчислення градієнтів і, таким чином, оновлення ваг мережі. Цей ітераційний процес триває до тих пір, поки втрати не досягнуть мінімального значення.

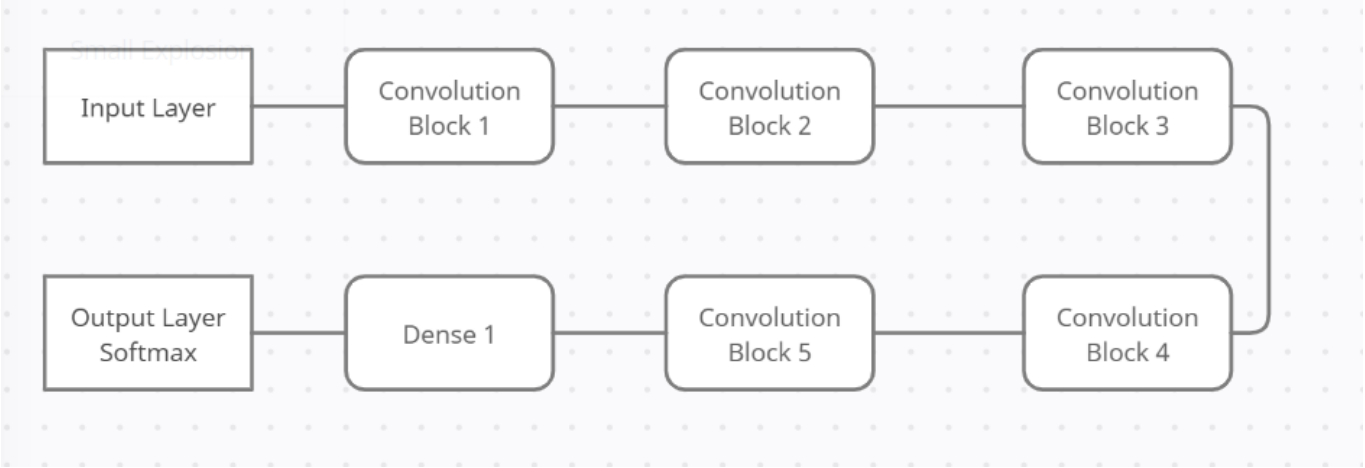

Рисунок 3.4 – Структура згорткової нейронної мережі

3.3 Порівняльна характеристика згорткової ті рекурсивної нейронних мереж

Для навчання моделі був використаний датасет GTZAN, який складається з 1000 композицій. Набір даних випадковим чином розбивається на набір для навчання (80%), набір для перевірки (10%), набір для тестування (10%).

Як показують результати, ми досягли 90% точності для моделі RNN і 92% точності для моделі CNN.

Таблиця 3.1 – Експерименти та результати моделі рекурсивної нейронної мережі.

№	Епохи	Точність	Втрата
1	30	74%	72%
2	30	67%	90%
3	30	74%	71%
4	100	75%	69%
5	30	90%	32%

Таблиця 3.2 – Експерименти та результати моделі згорткової нейронної мережі

№	Епохи	Точність	Втрата
1	30	67%	88%
2	30	64%	98%
3	30	67%	88%
4	100	69%	85%
5	30	92%	23%

На рис. 3.4 показано продуктивність навчання моделі RNN у шостому експерименті, а на рис. рис. 3.5 показано її для моделі CNN у п’ятому експерименті.

Незважаючи на те, що була досягнена 90% точность для RNN і 92% для CNN, коли ми подивимося на перші чотири експерименти моделей, то побачимо, що вони погано працювали на тестовому наборі даних. Для обох моделей перший експеримент не показав належних результатів на тестовому наборі даних, і втрати становили понад 70% в обох випадках.

Щоб дослідити проблему, у другому експерименті було навчено та перевірено моделі на меншому наборі даних, але результати не покращилися; точність була знижена, а втрати збільшені. Під час навчання моделей у першому та другому експериментах точність перевірки була вищою, ніж точність навчання, що вказує на те, що набір даних не є добре рандомізованим.

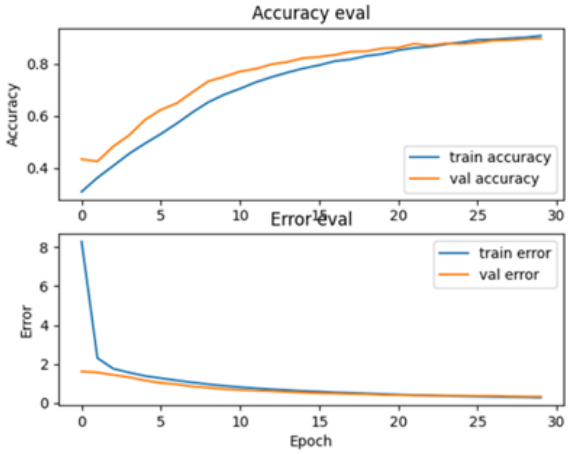

Рисунок 3.4 – Навчальний процес п’ятого експерименту для моделі RNN

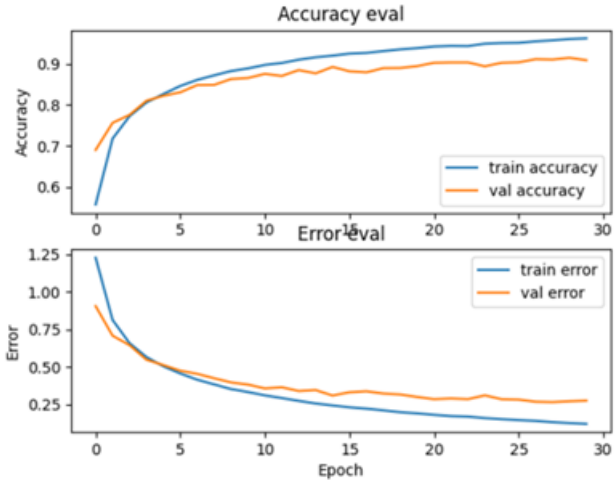

Рисунок 3.5 – П'ятий експериментальний навчальний процес для моделі CNN

У третьому експерименті алгоритм рандомізації набору даних було змінено на алгоритм перехресної перевірки. У цьому експерименті набір даних було рівномірно розподілено для кожного жанру. Цього разу результати показали вищу точність перевірки, ніж точність навчання, подібну до другого експерименту. Для четвертого

експерименту було збільшено кількість епох до 100 для обох моделей. Точність перевірки моделі RNN була нижчою, ніж підготовка, але вона погано виявила жанри. На відміну від RNN, результат моделі CNN був подібним до попереднього експерименту, точність перевірки була вищою, ніж точність навчання.

Щоб отримати довші зразки потрібно було зменшити кількість сегментації. Тому ми повернулися до вилучення функцій у п’ятому експерименті та змінили кількість зразків до 30 із кожної 30-секундної музики. Розділивши кожну 30-секундну музику на 30 семплів, кожен семпл отримав 1 секунду. Зразки довше 1 секунди не потрібно було витягувати, оскільки набір даних був би набагато меншим для експериментів. Крім того, через те, що довжина набору даних стала б меншою, ніж у попередніх експериментах, модель могла переповнюватись. Таким чином, знову кількість епох було змінено до 30.

Озираючись на експерименти, найкращий результат для моделі RNN був у п’ятому експерименті. Ми досягли 90% точності з 32% втратою. Крім того, найкращий результат для моделі CNN був у п’ятому експерименті, оскільки він мав менший рівень помилок, ніж інші експерименти, що означає, що він зробив менше помилок у тестовому наборі даних. Модель мала 92% точності з 23% втратою.

Отже, модель CNN показала 92% точності, а модель RNN досягла 90% точності, що свідчить про те, що модель CNN перевершила модель DNN у розпізнаванні музичних жанрів.

3.4 Створення системи розпізнавання стилю музики

Для створення системи буде використана модель представлена на рисунку 3.6.

Працюватиме вона наступним чином:

1. Користувач вибирає пісню, якої він хоче дізнатися жанр.

2. Відправляє ці дані на сервер за допомогою додатку(додаток може бути буль якого типу, починаючи від звичайного веб сайту, до Telegram бота, якого і буде використано)

3. Сервер повинен зберегти цей файл у директорію з якої нейронна мережа

буде отримувати цей файл на обробку

4. Нейронна мережа зчитує музичний файл і зберігає його в обробку.
5. Нейронна мережа передбачає можливий жанр відправленої пісні.
6. Отриманий результат повертається до користувача.

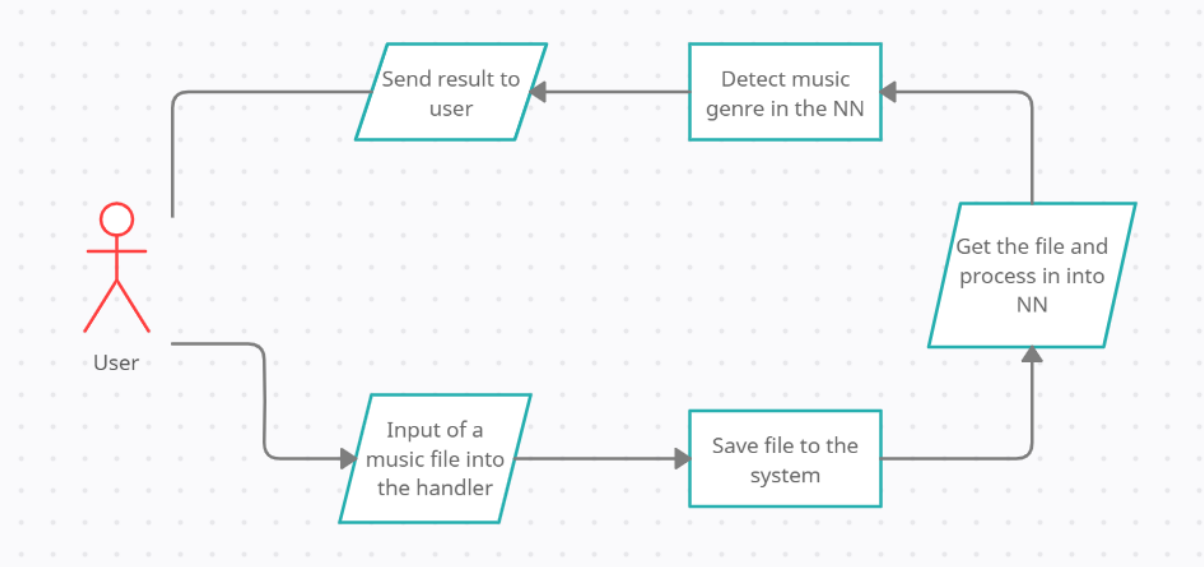

Рисунок 3.6 – Модель системи розпізнавання жанру музики

Для зв’язку з користувачем використовується Telegram бот, а задля кращого та більш точного визначення стилю музики було використано навчену згорткову нейронну мережу.

4. ПРОГРАМНА РЕАЛІЗАЦІЯ ТА РЕЗУЛЬТАТИ ВИКОНАННЯ

4.1 Створення та навчання нейронної мережі згоркового типу

Я збираюся використовувати набір даних GTZAN, як і було сказано раніше. Набір даних містить 10 жанрів, а саме: блюз, класика, кантрі, диско, хіп-хоп, джаз, метал, поп, реггі, рок. Я буду використовувати усі жанри окрім джазу, бо генерація спектрограм для нього некоректна.

Перш ніж розділити аудіофайли, треба створити порожні каталоги для кожного жанру.

Код нижче створює масив назв жанрів та за допомогою команди `os.makedirs` створює пусті папки де будуть зберігатися розділені аудіофайли та пов’язані до них спектрограми

```
    genres = 'blues classical country disco pop hiphop metal reggae rock'
    genres = genres.split()
    for g in genres:
      path_audio = os.path.join('/content/audio3sec',f'{g}')
      os.makedirs(path_audio)
      path_train          =          os.path.join('/content/gdrive/My Drive/spectrograms3sec/train',f'{g}')
      path_test           =          os.path.join('/content/gdrive/My Drive/spectrograms3sec/test',f'{g}')
      os. makedirs(path_train)
      os. makedirs(path_test)
```

Рисунок 4.1 – Створення папок для музичних файлів та спектрограм

Після чого аудіофайли розбиваються на аудіосегменти. Отже, тут є три цикли. Спочатку ми переглядаємо наявні у нас жанри, потім для цього жанру другий цикл переглядає кожен його аудіофайл, а третій цикл розділяє аудіофайл на 10 частин і зберігає його в порожніх каталогах, які ми створили для аудіофайлів раніше. Цього можливо добитися завдяки бібліотеці AudioSegment, яка тут і використовується. Це все зроблено для того щоб у майбутньому при вилучені Мел-спектрограм ми могли

отримати більшу вибірку для навчання. Значення розділення аудіофайлу, яке зараз встановлено на 3 секунди також було змінено протягом тестування функціоналу задля порівняльної характеристики вище.

```
i = 0
for g in genres:
  j=0
  print(f"{g}")
  for filename in os.listdir(os.path.join('/content/gdrive/MyDrive/kunal/genres',f"{g}")):

song=os.path.join(f'/content/gdrive/MyDrive/kunal/genres/{g}',f'{filename}')
    j = j+1
    for w in range(0,10):
      i = i+1
      t1 = 3*(w)*1000
      t2 = 3*(w+1)*1000
      newAudio = AudioSegment.from_wav(song)
      new = newAudio[t1:t2]
      new.export(f'/content/audio3sec/{g}/{g+str(j)+str(w)}.wav', format="wav")
```

Рисунок 4.2 – Розбиття датасету аудіофайлів на менші частини

Тепер для генерації Мел-спектрограм ми будемо використовувати бібліотеку `librosa`, яка сгенерує спектрограму для кожного файла свореного після розбиттся датасету на менші файли. Як можна побачити в реалізації нижче ми беремо кожну назву стилю та шукаємо файл у відповідній директорії, після того як було отримано файл ми створюємо Мел-спектрограму для кожного файлу за допомогою функції `librosa.feature.melspectrogram(y=y,sr=sr)`, потім генеруємо з кожної спектрограми зображення і зберігаємо його у відповідну папку. Таким чином ми заповнюємо пусті директорії які було створено на початку програми.

```
for g in genres:
  j = 0
  print(g)
  for filename in os.listdir(os.path.join('/content/audio3sec',f"{g}")):
    song = os.path.join(f'/content/audio3sec/{g}',f'{filename}')
    j = j+1
```

```
        y,sr = librosa.load(song,duration=3)
        #print(sr)
        mels = librosa.feature.melspectrogram(y=y,sr=sr)
        fig = plt.Figure()
        canvas = FigureCanvas(fig)
        p = plt.imshow(librosa.power_to_db(mels,ref=np.max))
        plt.savefig(f'/content/gdrive/My
Drive/spectrograms3sec/{g}/{g+str(j)}.png')
```

Рисунок 4.3 – Створення Мел-спектрограм для кожного створеного аудіофайлу

Наступним кроком буде розділення наших файлів на тренувальний набір даних, та на перевірочний набір. Наші повні дані знаходяться в каталозі для тренування, тому нам потрібно взяти частину повних даних і перемістити їх у наш тестовий каталог. Для кожного жанру ми випадково перемішуємо назви файлів, вибираємо 100 найпопулярніших імен файлів і переміщуємо їх у каталог тестування/перевірки. Ми підготували наші дані таким чином, щоб використовувати дивовижний `ImageDataGenerator` у `keras`, він дуже корисний для навчання моделі на великих наборах даних.

```
directory ="/content/gdrive/MyDrive/spectrograms3sec/train/"
for g in genres:
  filenames = os.listdir(os.path.join(directory,f"{g}"))
  random.shuffle(filenames)
  test_files = filenames[0:100]

  for f in test_files:
    shutil.move(directory + f"{g}"+ "/" + f,"/content/gdrive/My
Drive/spectrograms3sec/test/" + f"{g}")
```

Рисунок 4.4 – Створення тестового датасету для нейронної мережі

Далі ми створимо генератори даних як для навчання, так і для тестування.

Метод `flow_from_directory()` автоматично виводить мітки за допомогою нашої структури каталогу та кодує їх відповідним чином.

`ImageDataGenerator` спрощує навчання на великих наборах даних, використовуючи той факт, що під час навчання модель навчається лише на одному пакеті за крок, тому під час навчання генератор даних просто завантажує один пакет

у пам’ять за раз, тому ресурси пам’яті не вичерпуються.

```
train_dir = "/content/gdrive/My Drive/spectrograms3sec/train/"
train_datagen = ImageDataGenerator(rescale=1./255)
train_generator = train_datagen.flow_from_directory(train_dir,target_size=(288,432),color_mode="rgba",class_mode='categorical',batch_size=128)
validation_dir = "/content/gdrive/My Drive/spectrograms3sec/test/"
vali_datagen = ImageDataGenerator(rescale=1./255)
vali_generator = vali_datagen.flow_from_directory(validation_dir,target_size=(288,432),color_mode='rgba',class_mode='categorical',batch_size=128)
```

Приклад 4.5 – Створення генераторів даних як для навчання, так і для тестування

Далі потрібно ініціалізувати модель згорткової нейронної мережі. Модель буде складатися з п’яти згорткових шарів, де перший буде вхідним шаром, і кожен шар буде мати розмір ядра 3*3, з кроком 1. Нижче можна побачити приклад одного з згорткових шарів.

```
X = Conv2D(8,kernel_size=(3,3),strides=(1,1))(X_input)
X = BatchNormalization(axis=3)(X)
X = Activation('relu')(X)
X = MaxPooling2D((2,2))(X)
```

Рисунок 4.6 – Створення першого згорткового шару

Після цього треба додати шар виключення щоб запобігти перенавчання, а також щільний шар з активацією Softmax для виведення ймовірностей стилю. Після створення всіх шарів можна створити модель де першим параметром будуть вхідні дані, другим вихідні дані і останнім назва моделі.

```
X = Dropout(rate=0.3)
  X = Dense(classes, activation='softmax', name='fc' + str(classes))(X)
  model = Model(inputs=X_input,outputs=X,name='GenreModel')
```

Рисунок 4.7 – Створення шару виключення, щільного шару, а також генерація моделі

Останнім кроком буде навчання моделі на датасеті зробленого з Мел-

спектрограм. Наступний код використовується для прорахування точності з якою буде збережена інформація про музикальний стиль. Воно приймає у себе передбачене значення, та реальне після якого повертає похибку на яке повинно змінитися вихідне значення.

```
    def get_f1(y_true, y_pred):
        true_positives = K.sum(K.round(K.clip(y_true * y_pred, 0,
1)))
        possible_positives = K.sum(K.round(K.clip(y_true, 0, 1)))
        predicted_positives = K.sum(K.round(K.clip(y_pred, 0, 1)))
        precision = true_positives / (predicted_positives +
K.epsilon())
        recall = true_positives / (possible_positives + K.epsilon())
        f1_val = 2*(precision*recall)/(precision+recall+K.epsilon())
        return f1_val
```

Приклад 4.8 – Створення функції врахування точності моделі

І після чого треба згенерувати модель і встановити кількісь епох для генерації, у цьому випадку це 70.

get_f1() використовується для обчислення f1_score, а для використання генераторів даних під час навчання ми використовуємо метод fit_generator().

4.1.1 Результати роботи програми

Після першого, тестового, навчання протягом 70 епох я отримав точність навчання 89%, а під час перевірки – 84%. Отже, ми отримали справді хорошу точність під час перевірки, це завдяки розділенню аудіо файлів на 10 рівних частин.

Я також навчив модель на оригінальному наборі даних GTZAN, який містив 1000 аудіо файлів по 30 секунд кожен і отримав точність близько 50%, тому розділення аудіо файлів на частини значно підвищило нашу точність.

Тепер я виберу кілька пісень і спробую визначити стиль пісні за допомогою створеної моделі

Прикладом буде пісня: Enter The Sandman - Metallica.

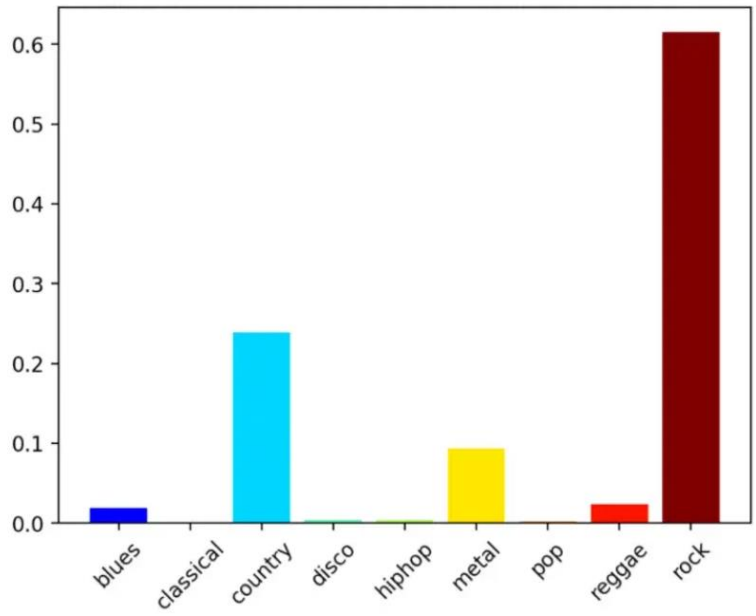

Рисунок 4.9 – Визначення жанру при використанні пісні Enter The Sandman - Metallica

Як можна побачити на ілюстрації (Рисунок 4.1) нейронна мережа з дуже великою точністю визначила жанр пісні, а саме, рок.

4.2 Створення та навчання нейронної мережі рекурсивного типу

Як і для попередньої моделі ми будемо використовувати датасет GTZAN. Для початку треба перетворити датасет на дані які можуть бути використані нейронною мережею. На відміну від згроткової нейромережі для навчання рекурсивної ми будемо використовувати дані перетворені в JSON об'єкти. Для цього ми так само будемо розділяти кожний сампл на п'ять частин та зберігати отримані дані у JSON.

```
    for n in range(num_segment):
        start = samples_per_segment * n
        finish = start + samples_per_segment
        mfcc   =   librosa.feature.mfcc(y[start:finish],   sample_rate,
n_mfcc = num_mfcc, n_fft = n_fft, hop_length = hop_length)
        mfcc = mfcc.T #259 x 13
            if len(mfcc) == num_mfcc_vectors_per_segment:
            data["mfcc"].append(mfcc.tolist())
```

```
            data["labels"].append(i-1)
            print("Track Name ", file_path, n+1)

        with open(json_path, "w") as fp:
            json.dump(data, fp, indent = 4)
```

Рисунок 4.10 – Вилучення музичної інформації та збереження її у JSON файл

Після того як тренувальні дані було згенеровано, потрібно створити та навчити рекурсивну нейронну мережу. Для початку потрібно поділити датасет на тренувальні, валідаційні та тестові значення. Як показано у прикладі 4.7 спочатку датасет завантажується як масив Мел-спектрограм та музикальних стилів пов’язаних з ними, потім ці дані розділяються за допомогою бібліотеки `sklearn.model_selection`, а саме метода `train_test_split`, який розділяє дані у розмірі за заданими відсотками. Спочатку ми зберігаємо тестові дані, та зберігаємо тренувальні дані, потім з тренувальних даних ми виділяємо валідаційні дані.

```
    def prepare_datasets(test_size,val_size):

        #load the data
        x, y = load_data(data_path)
        x_train,        x_test,        y_train,        y_test        =
train_test_split(x,y,test_size = test_size)
        x_train,        x_val,        y_train,        y_val        =
train_test_split(x_train,y_train,test_size = val_size)

        return x_train, x_val, x_test, y_train, y_val, y_test
```

Рисунок 4.11 – Завантаження датасету та розділення його на тренувальні, тестові та валідаційні дані

Після цього ми повинні створити модель. Для цього ми використаємо параметри які присутні в датасеті для створення першого шару моделі з 64 нейронами, другий шар також буде містити у собі 64 нейрони, третій та четвертий шари будуть щільними, при цьому третій шар буде мати 64 нейрони та використовувати функцію ReLU, а четвертий має 10 нейронів та функцію Softmax. Після того як модель було створено до нейронної мережі додається оптимізатор та

вона компілюється

```
    def build_model(input_shape):

        model = tf.keras.Sequential()

        model.add(tf.keras.layers.Bidirectional(64,    input_shape    =
input_shape, return_sequences = True))
        model.add(tf.keras.layers.Bidirectional(64))

        model.add(tf.keras.layers.Dense(64, activation="relu"))

        model.add(tf.keras.layers.Dense(10,activation = "softmax"))
        return model

        optimiser = tf.keras.optimizers.Adam(lr=0.001)
        model.compile(optimizer=optimiser,
                      loss='sparse_categorical_crossentropy',
                      metrics=['accuracy'])
```

Рисунок 4.12 – Створення моделі рекурсивної згорткової мережі та її компіляція

Після того як модель буда успішно створена та скомпільована нам потрібно навчити модель на даних які ми приготували. В цьому випадку ми будемо тренувати модель 50 епох з одночасною вибіркою у розмірі 32. Після цього модель перевіряється на тестових даних та зберігається у вигляді файлу .h5

```
    history  =  model.fit(x_train,  y_train,  validation_data=(x_val,
y_val), batch_size=32, epochs=50)
    model.save("model_RNN.h5")
```

Рисунок 4.13 – Навчання рекурсивної нейронної мережі та збереження навчених даних

Після усіх цих кроків залишилося лише протестувати нашу модель на існуючих мизичних творах. Для цього нам потрібно з отримати пісню та розділити її на частини по 30 секунд, після чого для кожної частини сгденерувати Мел-спектрограми, після чого для кожної спектрограми буде зроблено прогноз можливого жанру, за допомогою навеної моделі, і ці дані будуть збережені в масив

передбачень. З цього масиву для кожної ітерації фрагмента пісні нам потрібно отримати максимальне значення, це значення і буде нашим передбаченням відносного цього відрізка пісні. Його буде збережено в масив можливих передбачень.

```
    for n in range(num_segment):
        start = samples_per_segment * n
        finish = start + samples_per_segment
        mfcc = librosa.feature.mfcc(y[start:finish], sample_rate,
n_mfcc = num_mfcc, n_fft = n_fft, hop_length = hop_length)
        mfcc = mfcc.T
        mfcc = mfcc.reshape(1, mfcc.shape[0], mfcc.shape[1])
        array = model.predict(mfcc)*100
        array = array.tolist
    class_predictions.append(array[0].index(max(array[0])))
```

Рисунок 4.14 – Вилучення можливих значень стилю пісні

Після чого буде створенно ключ-значення, де буде рахуватися найбільше входження спецефічного жанру. Найбільше входження і буде тим жанром який є головним у пісні.

```
    occurence_dict = {}
        for i in class_predictions:
            if i not in occurence_dict:
                occurence_dict[i] = 1
            else:
                occurence_dict[i] +=1

            max_key = max(occurence_dict, key=occurence_dict.get)
            prediction_per_part.append(classes[max_key])

        #print(prediction_per_part)
        prediction = max(set(prediction_per_part), key =
prediction_per_part.count)
        print(prediction)
```

Приклад 4.11 – Вилучення найбільш імовірного жанру пісні

4.2.1 Результати роботи програми

Після першого, тестового, навчання протягом 50 епох я отримав точність

навчання 94%, а під час перевірки – 75%.

Тепер я виберу кілька пісень і спробую визначити стиль пісні за допомогою створеної моделі

Прикладом буде пісня: Viva La Vida – Coldplay

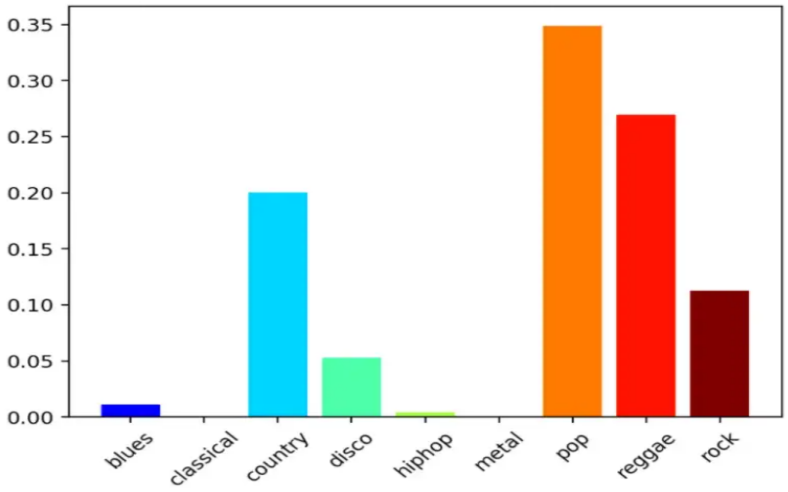

Рисунок 4.15 – Визначення жанру при використанні пісні Viva La Vida – Coldplay

4.3 Розробка Telegram боту для розпізнавання музичного стилю

Для того щоб ми могли використовувати Telegram бота задля розпізнавання музичного стилю ми повинні створити в ньому систему яка буде завантажувати файл у систему, а потім викликати нейронну мережу і визначати його стиль.

По перше ми повинні отримати файл з повідомленням яке було відправлено, та завантиажити його у систему за допомогою методу `get_file().download()`, де у параметри методів ми передаємо саме повідомлення, а саме документ прив'язаний до нього, та файл у який він буде записуватись. В той час програма перевіряє чи присутній файл у дерикторії, і якщо та то викликає метод передбачення.

```
with open("music/file.mp3", 'wb') as f:
    context.bot.get_file(update.message.document).download(out=f)
```

```
for filename in os.listdir("music"):
    genre = predict(os.path.join("music", filename), 'r')
bot.send_message(chat_id=user_id, text=genre)
```

Рисунок 4.16 – Опрацьовування відправленого користувачом музичного файлу

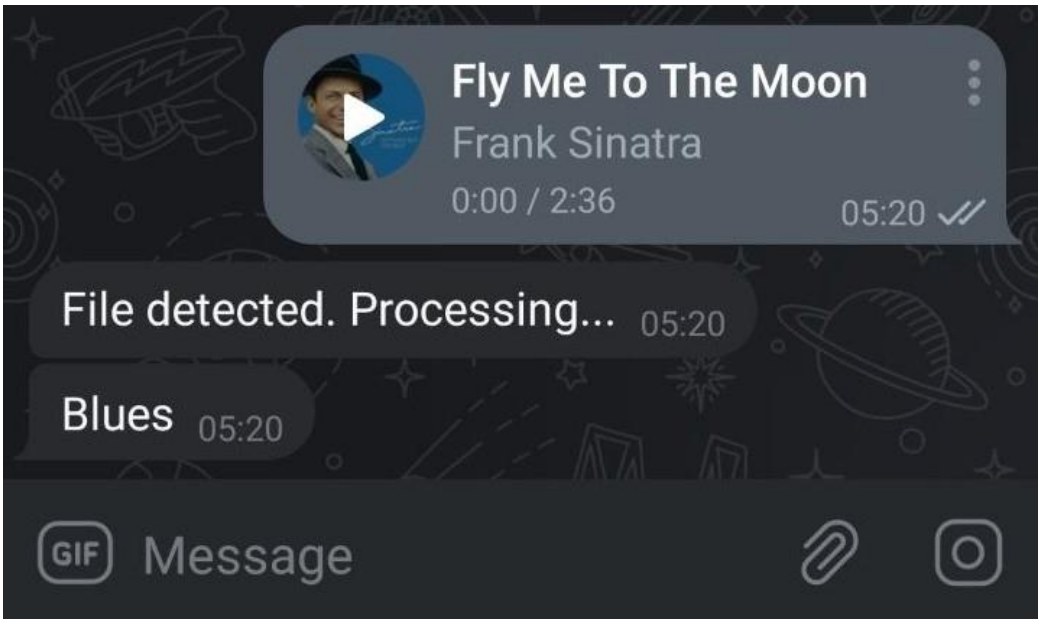

Рисунок 4.17 – Визначення жанру пісні за допомогою Telegram бота

ВИСНОВКИ

В ході кваліфікаційної роботи було розроблено інтелектуальну систему для розпізнавання музичних стилів на основі вибору кращого типу нейронної мережі для розв’язання поставленої проблеми, а саме для розпізнавання музикальних стилів. Для цього було розроблено систему рекурсивної нейронної мережі та нейронної мережі згорткового типу глибокого навчання. В результаті роботи було встановлено що використання згорткової мережі більш ефективне для розпізнавання стилів музики.

Під час виконнаня роботи було розроблено дві нейронних мережі розпізнавання: для рекурсивної нейронної системи було запропоновано алгоритм класифікації музики на основі вилучення музики та глибокої нейронної мережі. Як класифікаційні параметри для музики, алгоритм спочатку виділяє два типи характеристик, темброву характеристику та ознаку мелодії. Метод класифікації на основі згорткової нейронної мережі ігнорує час аудіо. Для згорткової нейронної мережі було створено алгоритм розпізнавання використовуючи Мел-спектрограмму, яку згорткова нейромережа використовувала як зображення, для чого згорткові нейронні мережі підходять якнайкраще.

Після опрацьовування результатів передбачення нейронних мереж було створено систему розпізнавання музичних стилів використовуючи Telegram бота як користувацький інтерфейс для виклику передбачення згорткової нейронної мережі

ПЕРЕЛІК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Reybrouck M., Eerola T. Music and its inductive power: a psychobiological and evolutionary approach to musical emotions //Frontiers in Psychology. – 2017. – Т. 8. – С. 494.

2. Burger B. et al. Relationships between perceived emotions in music and music-induced movement //Music Perception: An Interdisciplinary Journal. – 2012. – Т. 30. – №. 5. – С. 517-533.

3. Nagel F. et al. EMuJoy: Software for continuous measurement of perceived emotions in music //Behavior Research Methods. – 2007. – Т. 39. – №. 2. – С. 283-290.

4. An Y., Sun S., Wang S. Naive Bayes classifiers for music emotion classification based on lyrics //2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS). – IEEE, 2017. – С. 635-638.

5. Pandeya Y. R., Lee J. Deep learning-based late fusion of multimodal information for emotion classification of music video //Multimedia Tools and Applications. – 2021. – Т. 80. – №. 2. – С. 2887-2905.

6. Klimek P., Kreuzbauer R., Thurner S. Fashion and art cycles are driven by counter-dominance signals of elite competition: quantitative evidence from music styles //Journal of the Royal Society Interface. – 2019. – Т. 16. – №. 151. – С. 20180731.

7. Guimaraes P. et al. A comparison of identification methods of Brazilian music styles by lyrics //Proceedings of the Fourth Widening Natural Language Processing Workshop. – 2020. – С. 61-63.

8. Schneider B. Community and language in transnational music styles: symbolic meanings of Spanish in salsa and reggaetón //Contested Communities. – Brill, 2017. – С. 237.

9. Nobile D. Double-tonic complexes in rock music //Music Theory Spectrum. – 2020. – Т. 42. – №. 2. – С. 207-226.

10. Weiß C. et al. Investigating style evolution of Western classical music: A computational approach //Musicae Scientiae. – 2019. – Т. 23. – №. 4. – С. 486-507.

11. El Saddik A. Digital twins: The convergence of multimedia technologies //IEEE multimedia. – 2018. – Т. 25. – №. 2. – С. 87-92.

12. Kotevski Z., Tasevska I. Evaluating the potentials of educational systems to advance implementing multimedia technologies //International Journal of Modern Education and Computer Science (IJMECS). – 2017. – Т. 9. – №. 1. – С. 26-35.

13. Kozina Z. L. et al. Multimedia technologies as a means of training athletes in student basketball //Health, sport, rehabilitation. – 2018. – Т. 4. – №. 4. – С. 50-61.

14. Axak, N., Serdiuk, N., Ushakov, M., & Korablyov, M. (2020). Development of System for Monitoring and Forecasting of Employee Health on the Enterprise. In COLINS (pp. 979-992).

15. Fleischer R. If the song has no price, is it still a commodity?: Rethinking the commodification of digital music //Culture Unbound. – 2017. – Т. 9. – №. 2. – С. 146-162.

16. Riemer K., Johnston R. B. Disruption as worldview change: A Kuhnian analysis of the digital music revolution //Journal of Information Technology. – 2019. – Т. 34. – №. 4. – С. 350-370.

17. Zhang J. Music feature extraction and classification algorithm based on deep learning //Scientific Programming. – 2021. – Т. 2021.

18. Gan J. Music feature classification based on recurrent neural networks with channel attention mechanism //Mobile Information Systems. – 2021. – Т. 2021.

19. Lee J. et al. SampleCNN: End-to-end deep convolutional neural networks using very small filters for music classification //Applied Sciences. – 2018. – Т. 8. – №. 1. – С. 150.

20. Oramas S. et al. Multimodal deep learning for music genre classification //Transactions of the International Society for Music Information Retrieval. 2018; 1 (1): 4-21. – 2018.

21. Oramas S. et al. Multi-label music genre classification from audio, text, and images using deep features //arXiv preprint arXiv:1707.04916. – 2017.

22. Rosner A., Kostek B. Automatic music genre classification based on musical instrument track separation //Journal of Intelligent Information Systems. – 2018. – Т. 50. – №. 2. – С. 363-384.

23. Anisetty M. D. S. et al. Content-based music classification using ensemble of classifiers //International Conference on Intelligent Human Computer Interaction. –

Springer, Cham, 2018. – С. 285-292.

24. Zhang J. et al. Lightweight deep network for traffic sign classification //Annals of Telecommunications. – 2020. – Т. 75. – №. 7. – С. 369-379.

25. Bahuleyan H. Music genre classification using machine learning techniques //arXiv preprint arXiv:1804.01149. – 2018.

26. Lau D. S., Ajoodha R. Music Genre Classification: A Comparative Study Between Deep Learning and Traditional Machine Learning Approaches //Proceedings of Sixth International Congress on Information and Communication Technology. – Springer, Singapore, 2022. – С. 239-247.

27. Gu Y. et al. Deep learning based cell classification in imaging flow cytometer //ASP Transactions on Pattern Recognition and Intelligent Systems. – 2021. – Т. 1. – №. 2. – С. 18-27.

28. Chu W., Ho P. S., Li W. An adaptive machine learning method based on finite element analysis for ultra low-k chip package design //IEEE Transactions on Components, Packaging and Manufacturing Technology. – 2021. – Т. 11. – №. 9. – С. 1435-1441.

29. Sun W. et al. Prediction of cardiovascular diseases based on machine learning //ASP Transactions on Internet of Things. – 2021. – Т. 1. – №. 1. – С. 30-35.

30. Jiang Y. et al. A novel negative-transfer-resistant fuzzy clustering model with a shared cross-domain transfer latent space and its application to brain CT image segmentation //IEEE/ACM transactions on computational biology and bioinformatics. – 2020. – Т. 18. – №. 1. – С. 40-52.

31. Nam J. et al. Deep learning for audio-based music classification and tagging: Teaching computers to distinguish rock from bach //IEEE signal processing magazine. – 2018. – Т. 36. – №. 1. – С. 41-51.

32. Matityaho B., Furst M. Neural network based model for classification of music type //Eighteenth Convention of Electrical and Electronics Engineers in Israel. – IEEE, 1995. – С. 4.3. 4/1-4.3. 4/5.

33. Ning X. et al. Feature refinement and filter network for person re-identification //IEEE Transactions on Circuits and Systems for Video Technology. – 2020. – Т. 31. – №. 9. – С. 3391-3402.

34. Cai W. et al. TARDB-Net: triple-attention guided residual dense and BiLSTM networks for hyperspectral image classification //Multimedia Tools and Applications. – 2021. – Т. 80. – №. 7. – С. 11291-11312.

35. Tzanetakis G., Cook P. Musical genre classification of audio signals //IEEE Transactions on speech and audio processing. – 2002. – Т. 10. – №. 5. – С. 293-302.

36. Scaringella N., Zoia G., Mlynek D. Automatic genre classification of music content: a survey //IEEE Signal Processing Magazine. – 2006. – Т. 23. – №. 2. – С. 133-141.

37. Xu C. et al. Musical genre classification using support vector machines //2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings.(ICASSP'03). – IEEE, 2003. – Т. 5. – С. V-429.

38. Li T., Ogihara M., Li Q. A comparative study on content-based music genre classification //Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval. – 2003. – С. 282-289.

39. Panagakis I., Benetos E., Kotropoulos C. Music genre classification: A multilinear approach //ISMIR. – 2008. – С. 583-588.

40. McKinney M., Breebaart J. Features for audio and music classification. – 2003.

41. Zhang X. et al. An improved encoder-decoder network based on strip pool method applied to segmentation of farmland vacancy field //Entropy. – 2021. – Т. 23. – №. 4. – С. 435.

42. Yang Z. L. et al. VAE-Stega: linguistic steganography based on variational auto-encoder //IEEE Transactions on Information Forensics and Security. – 2020. – Т. 16. – С. 880-895.

43. Molla K. I., Hirose K. On the effectiveness of MFCCs and their statistical distribution properties in speaker identification //2004 IEEE Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems, 2004. (VCIMS). – IEEE, 2004. – С. 136-141.

44. Claus Weihs, Uwe Ligges, Fabian M¨orchen, and Daniel M¨ullensiefen. Classification in music research. Advances in Data Analysis and Classification, 1(3):255–291, Dec 2007.

45. Albadr M. A. A. et al. Mel-frequency cepstral coefficient features based on standard deviation and principal component analysis for language identification systems //Cognitive Computation. – 2021. – Т. 13. – №. 5. – С. 1136-1153.

46. Kang S. I., Lee S. M. Improvement of speech/music classification based on RNN in EVS codec for hearing aids //Journal of rehabilitation welfare engineering & assistive technology. – 2017. – Т. 11. – №. 2. – С. 143-146.

47. Rabiner L., Juang B. H. Fundamentals of speech recognition. – Prentice-Hall, Inc., 1993.

48. Davis S., Mermelstein P. Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences //IEEE transactions on acoustics, speech, and signal processing. – 1980. – Т. 28. – №. 4. – С. 357-366.

49. Mandel M. I., Ellis D. P. W. Song-level features and support vector machines for music classification. – 2005.

50. Kim H. G., Sikora T. Audio spectrum projection based on several basis decomposition algorithms applied to general sound recognition and audio segmentation //2004 12th European Signal Processing Conference. – IEEE, 2004. – С. 1047-1050.

51. Pohle T., Pampalk E., Widmer G. Evaluation of frequently used audio features for classification of music into perceptual categories //Proceedings of the Fourth International Workshop on Content-Based Multimedia Indexing. – 2005. – Т. 162.

52. Logan B. Mel frequency cepstral coefficients for music modeling //In International Symposium on Music Information Retrieval. – 2000.

53. Aucouturier J. J., Pachet F. Representing musical genre: A state of the art //Journal of new music research. – 2003. – Т. 32. – №. 1. – С. 83-93.

54. Lippens S., Martens J. P., De Mulder T. A comparison of human and automatic musical genre classification //2004 IEEE international conference on acoustics, speech, and signal processing. – IEEE, 2004. – Т. 4. – С. iv-iv.

55. Rabiner L. R. Digital processing of speech signals. – Pearson Education India, 1978.

56. Axak N., Korablyov M., Ushakov M. Cloud Architecture for Remote Medical Monitoring // IEEE Proceedings of the 15th International conference «Computer Sciences

and Information Technologies», (CSIT-2020). – Zbarazh-Lviv, Ukraine, September 23-26, 2020. – vol. 1, pp. 344-347.

57. Axak N., Serdiuk N., Ushakov M., Korablyov M. Development of System for Monitoring and Forecasting of Employee Health on the Enterprise. //COLINS. – 2020. – С. 979-992.

58. Axak N., Korablyov M., Ushakov M. The Development of a Multi-Agent System for Controlling an Autonomous Robot // Proceedings of the 17th International Conference on ICT in Education, Research and Industrial Applications. Integration, Harmonization and Knowledge Transfer. Volume I, (ICTERI-2021). – Kherson, Ukraine, 2021. – vol. 1, pp. 96-105.

59. Ушаков М. Модель програмування Map-Reduce для користувацьких обчислювальних машин. *Інформаційні технології в сучасному світі: дослідження молодих вчених:* Міжнародна науково-практична конференція молодих учених, аспірантів та студентів тези доповідей, (м. Харікв, лютий 2020 р.). – Харків: ХНЕУ імені Семена Кузнеця, 2020. С.7.

60. Аксак Н., Ушаков М. Проектування мультиагентів з використанням платформи JADE *"Інформаційні технології та системи"*.Матеріали Міжнародної науково-практичної конференції: тези доповідей, 8-9 квітня 2021 р. – Х.: ХНЕУ імені Семена Кузнеця, 2021. – С.4.

61. Ушаков М. Мультиагентні системи для моделювання компонентів інтернету речей. *Інформаційні технології в сучасному світі: дослідження молодих вчених:* Міжнародна науково-практична конференція молодих учених, аспірантів та студентів тези доповідей, (м. Харікв, 17 – 18 лютого 2022 року). – Харків: ХНЕУ імені Семена Кузнеця, 2022. С.4.

62. Ушаков М. Огляд методів синхронізації та управління багатоагентними системами. *Методи та засоби обчислювального інтелекту:* матеріали XXIV міжнар. молод. форуму, м. Харків, 7-9 квіт. 2020 р. Харків, 2020. С. 175-176.

63. Kirke A. Application of intermediate multi-agent systems to integrated algorithmic composition and expressive performance of music. 2011.