УДК 004.89:004.932.2

РОЗРОБКА АЛГОРИТМУ ТА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ КЛАСИФІКАЦІЇ МУЗИЧНИХ ЖАНРІВ У ПРОСТОРІ МЕЛ-ЧАСТОТНИХ КЕПСТРАЛЬНИХ КОЕФІЦІЄНТІВ

Калашнікова П. В.[1]; Полякова М. В., д.т.н., доц.[2]
[1,2] Національный університет "Одеська політехніка"
[1,2] Україна, Одеса
[1] poloniakalashnikova@gmail.com; [2] marinapolyakova943@gmail.com

Анотація. *У роботі досліджено архітектури згорткових нейронних мереж та рекурентних нейронних мереж, зокрема довга короткочасна пам'ять, для класифікації музичних жанрів. Проведено попередню обробку аудіофайлів та отримання ознак у вигляді мел-частотних кепстральних коефіцієнтів. Обрано архітектури згорткової нейронної мережі та довготривалої короткочасної пам'ятті для класифікації музичних жанрів, визначено оптимальні гіперпараметри та методи оцінки точності мережі. Проаналізовано результати навчання та тестування мережі на наборі даних. В роботі також розглянуто можливості застосування обраних мереж для реальних задач класифікації музичних жанрів.*

Ключові слова: *згорткові та рекурентні нейронні мережі; класифікація музичних жанрів; мел-частотні кепстральні коефіцієнти, машинне навчання.*

Вступна частина. Класифікація музичних жанрів визначається як задача класифікації у просторі векторів аудіосигналів на основі їх характеристик. Музичні стрімінгові сервіси, такі як Spotify і Apple Music, застосовують алгоритми класифікації для автоматичного створення персоналізованих плейлистів та рекомендацій нової музики. Це забезпечується за допомогою методів машинного навчання, таких як глибоке навчання, які використовуються для аналізу та класифікації великих обсягів аудіоданих.

Актуальність задачі класифікації музичних жанрів обумовлена зростанням обсягів аудіоконтенту та необхідністю його ефективної організації, особливо в стрімінгових сервісах, таких як Spotify та Apple Music. У цій роботі розглядається підхід до класифікації музичних жанрів на основі МЧКК та глибоких нейронних мереж.

Задача полягає в створенні мережі для класифікації музичних жанрів з високою точністю. У загальному випадку, високою точністю вважається показник, що перевищує 90%. Проте, в задачах класифікації музичних жанрів, де дані можуть бути дуже різноманітними і складними, точність на рівні 80-85% може також вважатися високою.

На вхід подаються аудіофайли у форматі WAV, які обробляються для отримання МЧКК. Процес включає перетворення аудіофайлів на МЧКК, розбиття на сегменти, та проходження сегментів через ЗНМ або ДКЧП. Вихід – точна класифікація жанрів музичних треків.

У дослідженні застосовано набір даних GTZAN Dataset, що містить 1000 аудіофайлів у форматі WAV (10 жанрів по 100 треків тривалістю 30 с). Дані поділені на вибірки: 80% – навчання, 10% – валідація, 10% – тест [1]. Алгоритм обробки включає завантаження аудіофайлів, обчислення МЧКК для кожного сегмента, створення навчальної та тестової вибірок, тренування мережі та оцінку її точності. У коді реалізовано вибір між ЗНМ і ДКЧП для класифікації музичних жанрів.

Метою дослідження є порівняння якості класифікації музичних жанрів за аудіо даними у просторі мел-частотного кепстрального перетворення. Результати порівняння дозволяють

визначити, який класифікатор доцільно застосовувати для організації музичного контенту відповідно до вподобань слухачів.

Основна частина. Використання згорткових (ЗНМ) та рекурентних (ДКЧП) нейронних мереж для класифікації музичних жанрів дозволяє ефективно обробляти як просторові, так і часові аспекти аудіосигналів, що забезпечує високу точність класифікації. ЗНМ виділяють важливі ознаки з аудіосигналів, представлених у вигляді спектрограм або МЧКК, вони показали високу ефективність у задачах розпізнавання образів та класифікації аудіосигналів, а їх використання для класифікації музичних жанрів у просторі МЧКК [1]. Тоді як ДКЧП враховують часові залежності між кадрами, що є критично важливим для музичних даних. Вони здатні запам’ятовувати довготривалі залежності, працювати з послідовностями різної довжини та зменшувати затримки, що критично для застосувань у реальному часі [2].

У процесі розробки алгоритму класифікації музичних жанрів за допомогою нейронних мереж важливим етапом є підготовка та вибір ознак, які використовуються для навчання мережі. Починають з підготовки даних, які представляють собою аудіофайли різних музичних жанрів. Ефективний спосіб представлення аудіосигналів - використання мел-частотних кепстральних коефіцієнтів (МЧКК), що дозволяють виділити основні характеристики звукового сигналу і зменшити його розмірність [3].

Аудіофайли тривалістю 30 с поділяються на короткі сегменти для збільшення навчальних прикладів. Для шкірного сегмента обчислюються крейдяно-частотні кепстральні коефіцієнти (МЧКК): сигнал розбивається на вікна, виконується дискретне перетворення Фур'є, обчислення спектральної енергії, застосування крейд-фільтрів, логарифмування та дискретне косинусне перетворення. Отримані МЧКК зберігаються з мітками жанрів і використовуються як вхідні дані для навчання нейромереж.

Архітектура ЗНМ складається зі згорткових шарів (32–128 фільтрів), шарів об'єднання (pooling) та двох Dropout (0.3) для зменшення перенавчання. Після глобального об'єднання застосовується повнозв'язкова куля на 512 нейронів (ReLU) та вихідна Softmax-куля. Навчання також проводиться з оптимізатором Adam та функцією втрат binary_crossentropy. Графіки втрат і точності використовуються для контролю стабільності та виявлення ознак перенавчання. [4]

Архітектура ДКЧП містить дві послідовні кулі по 64 одиниці, які моделюють довготривалі залежності у часових даних, після чого додається повнозв'язна куля з активацією ReLU та вихідна Softmax-куля, що повертає ймовірності жанрів. Модель компілюється з оптимізатором Adam, функцією втрат sparse_categorical_crossentropy та метрикою accuracy, тренується протягом 50–150 епох з розміром пакета 32. [5]

Обробку виконували на ноутбуці ASUS VivoBook (Intel Core i7 10th Gen, 16 ГБ RAM, 1 ТБ пам'яті) з використанням Python 3.8 та бібліотек NumPy, Librosa, Keras, TensorFlow.

Графіки втрат та точності для ЗНМ на тренувальній та валідаційній вибірках протягом епох навчання показують зміну показників (рис. 1). Втрати та точність змінюються протягом навчання, що дозволяє спостерігати за процесом та виявляти ознаки перенавчання.

Для моделі ЗНМ точність на тестовій вибірці досягла 88% після 50 епох навчання та за 150 епох точність зросла до 90%. Графік втрат показує, що втрати на тренуванні та валідації поступово зменшуються протягом епох навчання, але з часом на валідації відбувається деяке зростання втрат після приблизно 100 епох, що може вказувати на початок перенавчання моделі.

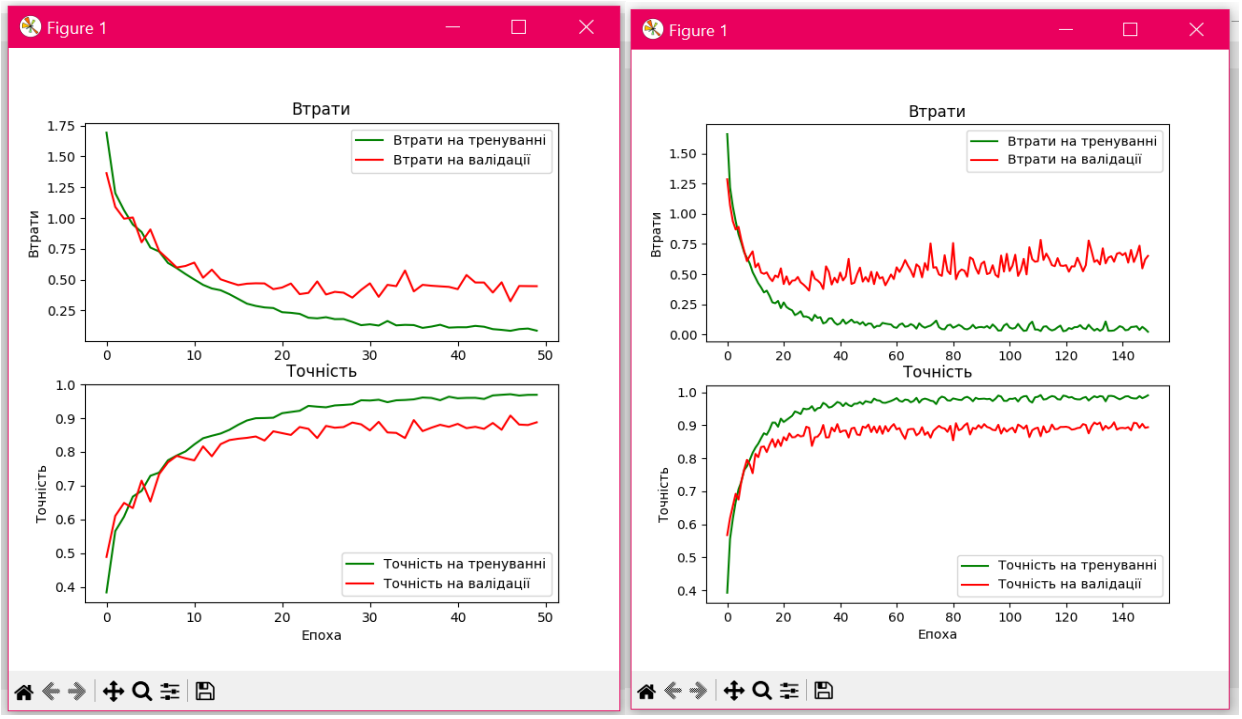

Рисунок 1 – Графіки втрат та точності на тренувальній та валідаційній вибірках для ЗНМ на 50 та 150 епохах

У порівнянні з моделлю ЗНМ, модель ДКЧП показує трохи іншу точність: графік вказує на швидше зменшення втрат на тренуванні та валідації протягом епох (рис. 2).

Після 50 епох навчання точність ДКЧП становить 81%, а після 150 епох вона зменшується до 80%. Це вказує на те, що мережа досягла певного рівня насичення під час навчання, і наступні епохі не вплинули позитивно на результат.

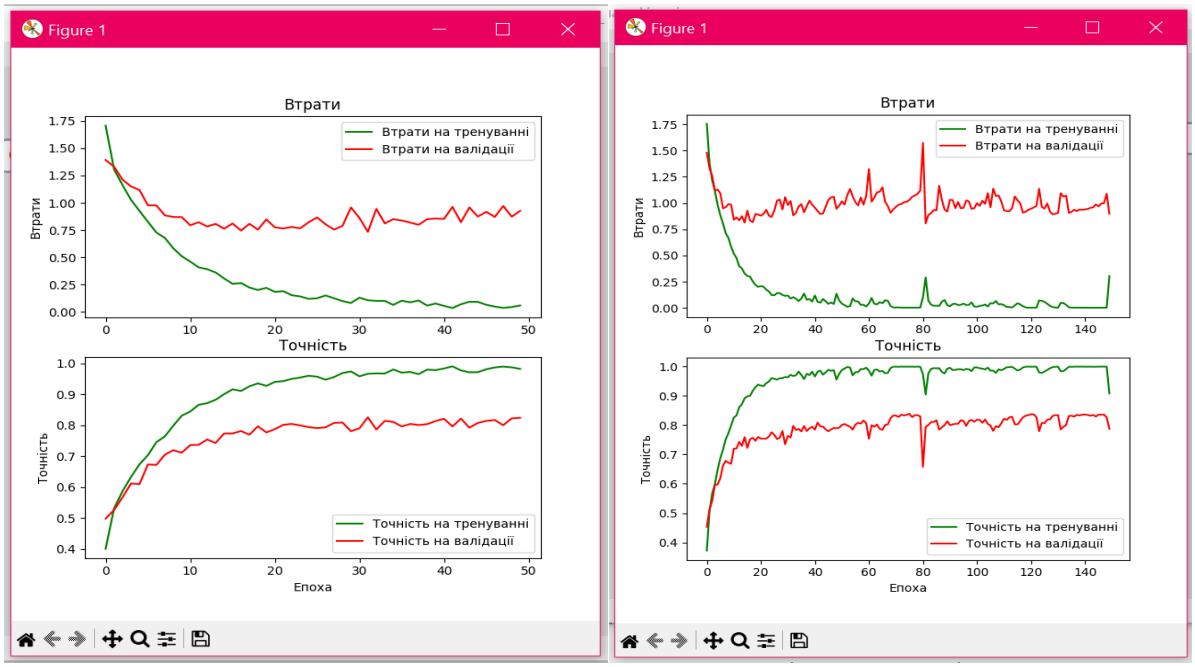

Рисунок 2 – Графіки втрат та точності на тренувальній та валідаційній вибірках для ДКЧП на 50 та 150 епохах

Висновки. Дослідження результатів застосування розроблених алгоритму та програмного забезпечення для визначення музичних жанрів для реальних аудіо записів показало, що зі збільшенням кількості епох точність зростає, проте з часом з'являється небезпека перенавщання. ДКЧП мережа досягла певного рівня насичення під час навчання раніше за ЗНМ, і додаткові епохи не призвели до подальшого поліпшення результатів.

Розробка класифікації музичних жанрів за допомогою ЗНМ та ДКЧП є значним кроком вперед, але подальші можливості дослідження в цій області можуть включати комбінацію цих двох методів, що відкриває нові горизонти для автоматизації та вдосконалення процесів аналізу музики. Такий підхід може бути використаний для створення більш точних і ефективних систем рекомендацій, що забезпечить покращений користувацький досвід на музичних платформах.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

[1] Genussov, M., & Cohen, I. (2010). *Musical genre classification of audio signals using geometric methods.* In *2010 18th European Signal Processing Conference (EUSIPCO)* (pp. 497–501). IEEE.

[2] Graves, A., Mohamed, A.-r., & Hinton, G. (2013). *Speech recognition with deep recurrent neural networks.* In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6645–6649). IEEE.

[3] Humphrey, E. J., Bello, J. P., & LeCun, Y. (2012). *Moving beyond feature design: Deep architectures and automatic feature learning in music informatics.* In *Proceedings of the 13th International Society for Music Information Retrieval Conference (ISMIR)* (pp. 403–408).

[4] Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. arXiv preprint arXiv:1412.3555.

[5] Tang, C. P., Chan, W. K., & Cheng, L. M. (2018). Music genre classification using a hierarchical long short term memory (LSTM) model. In Third International Workshop on Pattern Recognition (pp. 334–340). SPIE.

Kalashnikova P. V., Polyakova M. V.

Development of an Algorithm and Software for Music Genre Classification in the Space of Mel-Frequency Cepstral Coefficients

Abstract. *The paper investigates the architectures of convolutional neural networks and recurrent neural networks, in particular long short-term memory, for the classification of musical genres. Preprocessing of audio files and obtaining features in the form of mel-frequency cepstral coefficients is carried out. The architectures of convolutional neural networks and long-term short-term memory are selected for the classification of musical genres, the optimal hyperparameters and methods for assessing the accuracy of the network are determined. The results of training and testing the network on the dataset are analyzed. The paper also considers the possibilities of applying the selected networks to real problems of classifying musical genres.*

Keywords: *convolutional and recurrent neural networks; classification of musical genres; mel-frequency cepstral coefficients, machine learning.*