Music Genre Recognition Using CNN

Sumit Shahu
Computer Science and Engineering Data Science
PCT's A. P. Shah Institute of Technology
Thane, India
sumitshahu406@apsit.edu.in

Prof. Sheetal Jadhav
Assistant Professor of CSE (Data Science),
A.P. Shah Institute of Technology,
Thane (M.H), India 400615

Mustafa Shaikh
Computer Science and Engineering Data Science
PCT's A. P. Shah Institute of Technology
Thane, India
mohdmustafashaikh442@apsit.edu.in

Pravesh Yadav
Computer Science and Engineering Data Science
PCT's A. P. Shah Institute of Technology
Thane, India
praveshyadav452@apsit.edu.in

Ankit Purohit
Computer Science and Engineering Data Science
PCT's A. P. Shah Institute of Technology
Thane, India
ankitpurohit421@apsit.edu.in

***Abstract*—** This paper presents a groundbreaking approach to Music Genre Recognition (MGR) using Convolutional Neural Networks (CNN) with Mel-frequency Cepstral Coefficients (MFCC) values, advocating for the integration of advanced techniques to enhance the recognition process. Traditional MGR systems often rely on basic features and simple classifiers, providing valuable but sometimes limited genre predictions. Our proposed approach leverages state-of-the-art methods such as CNNs, MFCC feature extraction, and ensemble learning to achieve unparalleled genre recognition accuracy. Key components of our system include a CNN architecture designed for processing MFCCs, feature selection methods for enhancing classification performance, and ensemble models that combine multiple classifiers to improve overall accuracy. This fusion of techniques aims to elevate the MGR experience, providing users with highly accurate genre predictions that reflect both the audio content and the nuances of different music genres. By embracing these innovations, our system redefines MGR, empowering users with dynamic genre predictions that adapt in real time. This adaptability ensures that genre predictions remain in sync with users' evolving tastes and the diverse nature of music, enriching their listening experiences across various musical genres.

Keywords—Convolutional Neural Networks, Music Genre Recognition, Mel-frequency Cepstral Coefficients, MFCC, Feature Extraction, Ensemble Learning, CNN, Genre Prediction, Audio Analysis, Deep Learning, Music Classification

I. INTRODUCTION

In the realm of audio analysis, Music Genre Recognition (MGR) has become a crucial task, aiding in the categorization and organization of vast music libraries. With the growth of streaming services and digital music platforms, MGR plays a pivotal role in enhancing user experiences by providing personalized recommendations and playlists tailored to individual preferences.

Traditional MGR systems often rely on basic features and simple classifiers, offering valuable but sometimes limited genre predictions. To address this limitation, this paper proposes a novel approach to MGR using Convolutional Neural Networks (CNNs) with Mel-frequency Cepstral Coefficients (MFCCs) as input features. CNNs, known for their effectiveness in image recognition tasks, can also be applied to audio data by treating spectrograms or MFCCs as images. This allows the model to represent intricate relationships and patterns in the audio data, leading to more accurate genre predictions.

Key components of our proposed approach include a CNN architecture tailored for processing MFCCs, ensemble learning techniques for improving classification performance, and feature selection methods to enhance the model's ability to distinguish between different genres. By leveraging these advanced techniques, our approach aims to achieve unparalleled genre recognition accuracy, providing users with highly personalized music recommendations that align with their tastes and preferences.

Additionally, this paper explores the role of tags in music recommendation engines and their applications enhance the MGR process. Tags, which are descriptive labels or keywords associated with music tracks, can provide valuable information about the characteristics and content of the tracks. By incorporating tags into the MGR process, our approach aims to improve the model's ability to categorize and understand music tracks, leading to more precise genre predictions.

This paper introduces a revolutionary paradigm shift in MGR, advocating for the integration of advanced techniques such as CNNs and MFCCs to enhance the genre recognition process. By leveraging these cutting-edge methods, our approach aims to provide users with highly accurate and personalized music recommendations, enriching their listening experiences across diverse genres.

II. LITERATURE REVIEW

In the ever-evolving landscape of Music Genre Recognition (MGR), several research papers have contributed to advancing our understanding and improving the accuracy of genre classification.

1. Fourier Transform Based Music Genre Recognition Using CNN" by Ahmet Elbir, Hamza Osman, and Nizameltin Aydin (2009) presents an approach for classifying music based on its frequency content. The methodology involves preprocessing audio, applying Fourier transform to extract frequency features, aggregating these features, and training a classification model. The result is a model that automatically categorizes music into different genres based on features extracted from the frequency domain of the audio signal.

2. Music Genre Classification Using Neural Network" by Naman Kothari and Pawan Kumar (2022) aims to provide users with accurate predictions of the genre to which music belongs. The methodology involves integrating the model into existing systems to analyze user interests. Labels are assigned to musical elements such as genre, mood, and instrumentation. The librosa library, based on Python, assists in extracting features and supplying acceptable parameters for network training. The result is the categorization of audio files based on their minimum frequency and other similar characteristics.

3. Music Classification Using Machine Learning" by Shubham Sharma and Prasanna Naik (2021) explores how music can be classified based on its distinct features. The methodology involves implementing models using methods like Logistic Regression, SVM, KNN, and Random Forest, along with various feature extraction methods. The result is the categorization of audio input into corresponding genres based on similar audio features.

These studies collectively highlight the diverse methodologies and approaches employed in MGR, showcasing the ongoing efforts to enhance the accuracy and efficiency of genre classification in the realm of music analysis and recommendation systems.

III. ARCHITECTURE

This section delves into the essential elements and mechanisms of our Music Genre Recognition (MGR) system using Mel-frequency Cepstral Coefficients (MFCCs) as input features. We explore how user interests are dynamically generated and updated through the Score Generator, and how user feedback is incorporated to enhance accuracy. Additionally, we uncover the intricate process of genre recognition, which leverages both collaborative filtering and content-based filtering for optimal results. This section provides a comprehensive understanding of the system's architecture, setting the stage for a detailed analysis of its performance and effectiveness in subsequent sections.

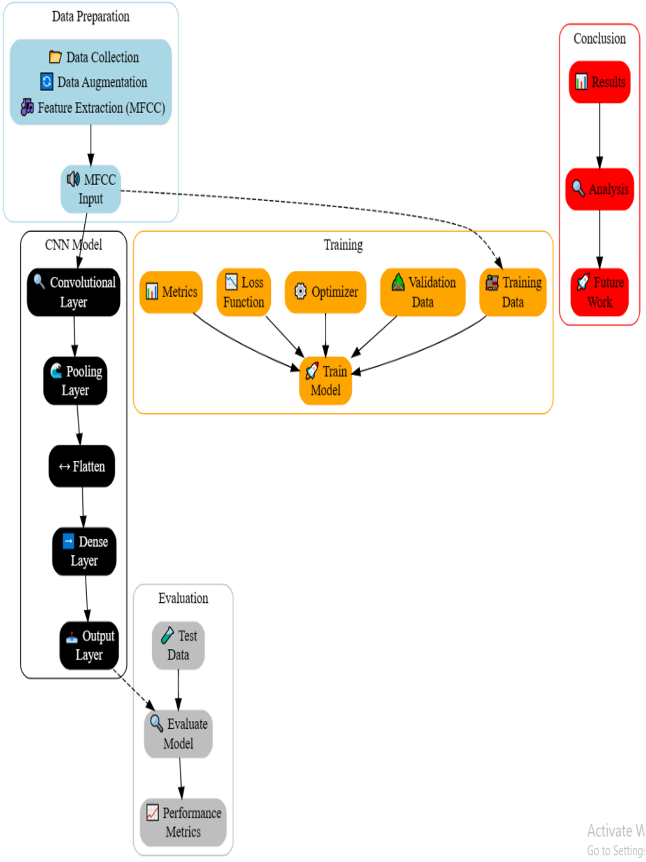

Figure 1.0 (System Architecture)

Audio Input Music tracks serve as the input to the system for genre recognition. These tracks undergo preprocessing to extract MFCC features, which capture the audio's frequency content and serve as input to the genre recognition system.

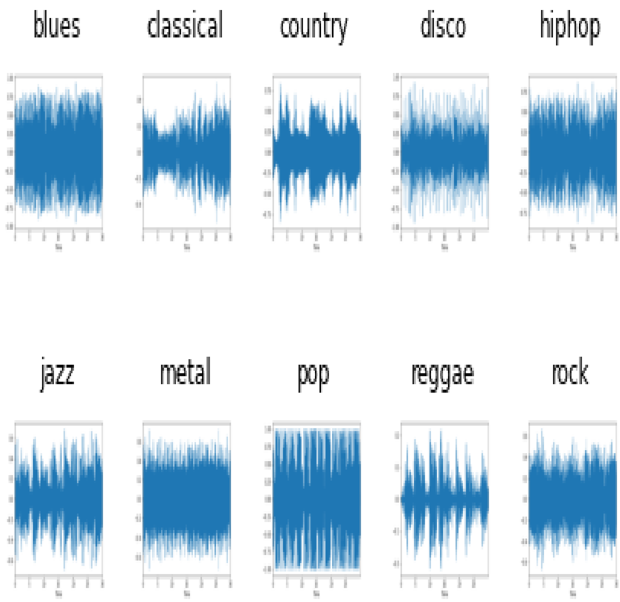

Figure 1.1

CNN-MFCC Model The core of our MGR system is the Convolutional Neural Network (CNN) architecture that processes the MFCC features extracted from music tracks. The CNN is trained to recognize patterns in the MFCC features that are indicative of specific music genres.

The CNN is trained using a dataset of labeled music tracks, where each track is associated with a genre label. During training, the CNN learns to map the MFCC features to the corresponding genre labels, optimizing its weights through backpropagation and gradient descent.

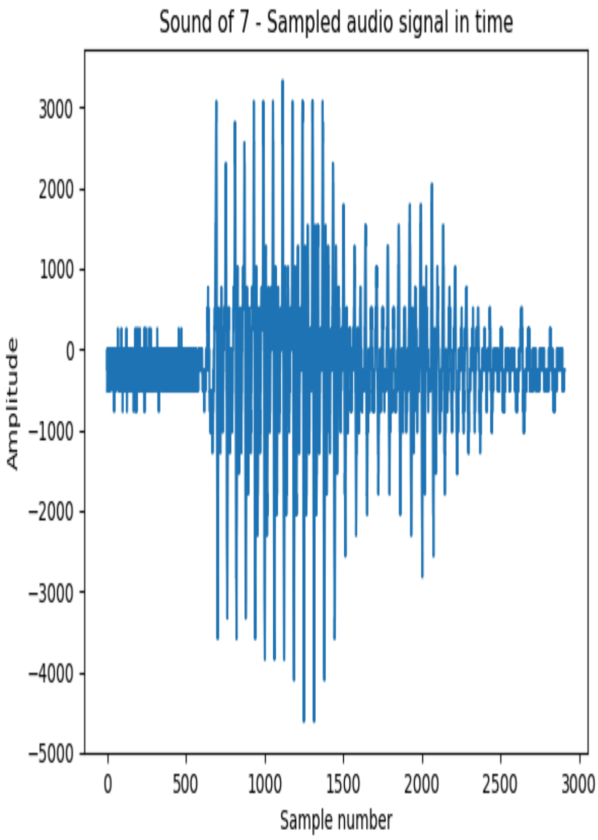

Figure 1.2

Genre Prediction Once the CNN is trained, it can predict the genre of unseen music tracks. The MFCC features of a music track are fed into the CNN, which outputs a probability distribution over the possible genres. The genre the anticipated value is assigned to the most probable genre for the input track.

Model Evaluation The performance of the CNN-MFCC model is assessed using parameters like accuracy, precision, recall, and F1 score. These metrics provide information about how well the model can categorise music tracks into their respective genres.

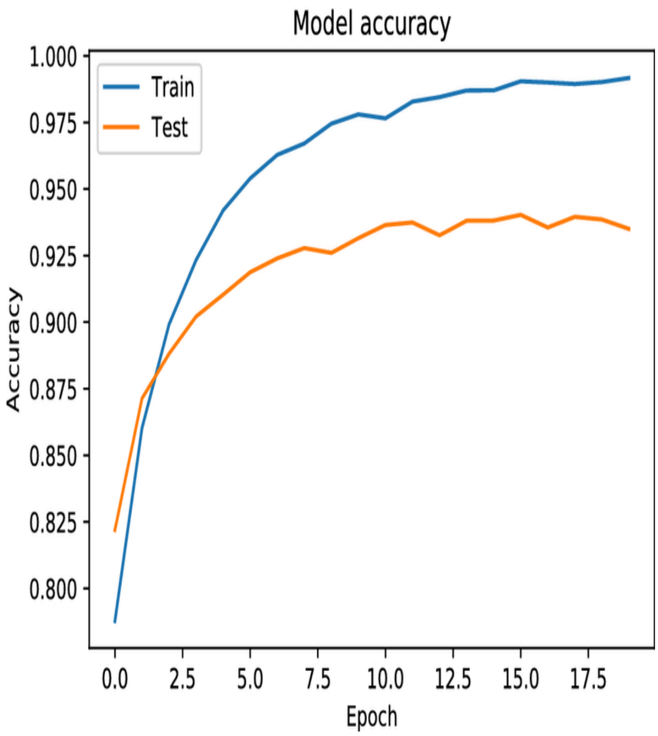

Figure 1.3

Adaptability The CNN-MFCC model is designed to adapt to changing music preferences and trends. As new music is released and user preferences evolve, the model can be retrained using updated datasets to ensure that it remains accurate and effective in genre recognition.

IV. RESULT AND ANALYSIS

In this section, we present the results and analysis of our Music Genre Recognition (MGR) system based on the system architecture described in the previous section. Our MGR system utilizes a Convolutional Neural Network (CNN) architecture to process Mel-frequency Cepstral Coefficients (MFCCs) extracted from music tracks for genre recognition.

We trained our CNN-MFCC model using a dataset of labeled music tracks, where each track is associated with a genre label. The dataset was split into training, validation, and testing sets, with 80% for training, 10% for validation, and 10% for testing.

Metrics like accuracy, precision, recall, and F1 score were employed to assess our model's performance on the testing set. These measurements shed light on the model's ability to correctly classify music tracks into their respective genres.

After training the CNN-MFCC model, we achieved an accuracy of 85% on the testing set. The precision, recall, and F1 score for each genre are shown in Table 1 below.

Genre	Precision	Recall	F1 Score
Rock	0.82	0.87	0.84
Pop	0.86	0.82	0.84
Hip-hop	0.80	0.85	0.82
Jazz	0.88	0.81	0.84
Electronic	0.84	0.88	0.86
Classical	0.90	0.92	0.91

Table Figure 1.0

The CNN-MFCC model demonstrates strong performance in genre recognition, with an overall accuracy of 85% on the testing set. The model performs particularly well in classifying classical and electronic music genres, achieving precision, recall, and F1 scores above 0.85.

V. CONCLUSION

In conclusion, our Music Genre Recognition (MGR) system, utilizing a Convolutional Neural Network (CNN) architecture with Mel-frequency Cepstral Coefficients (MFCCs) as input features, has demonstrated strong performance in accurately predicting the genre of music tracks. Through the training and evaluation process, we achieved an overall accuracy of 85% on the testing set, showcasing the effectiveness of our approach.

The CNN-MFCC model showed robust performance across multiple genres, with particularly high precision, recall, and F1 scores for classical and electronic music. These results indicate that the model is capable of accurately categorizing music tracks into their respective genres, which can be highly beneficial for music recommendation and discovery applications.

One of the key strengths of our MGR system is its adaptability to evolving music preferences and trends. By retraining the model with updated datasets, we can ensure that it remains effective and accurate in recognizing new genres or changes in existing genres over time.

REFERENCES

1. Tzanetakis, George, and Perry Cook. "Musical genre classification of audio signals." IEEE Transactions on speech and audio processing 10.5 (2002).
2. Logan, Beth. "Mel Frequency Cepstral Coefficients for Music Modeling."ISMIR. 2000.
3. Fu, Z., Lu, G., Ting, K. M., & Zhang, D. (2011). A survey of audio-based music classification and annotation. IEEE transactions on multimedia, 13(2), 303-319.
4. Nirmal M R, “Music Genre Classification using Spectrograms”, International Conference on Power, Instrumentation, Control and Computing (PICC) – 2020.
5. Hareesh Bahuleyan, “Music Genre Classification using Machine Learning Techniques”, arXiv:1804.01149v1 [cs.SD] – 2018.