Mini Project Report

on

Music Genre Recognition Using CNN

Third Year Engineering – Computer Science Engineering (Data Science)

by

Sumit Shahu	**21107004**
Pravesh Yadav	**21107057**
Mustafa Shaikh	**21107045**
Ankit Purohit	**21107020**

Under the guidance of

Prof. Sheetal Jadhav

DEPARTMENT OF COMPUTER SCIENCE ENGINEERING (DATA SCIENCE)

A.P. SHAH INSTITUTE OF TECHNOLOGY

G.B. Road, Kasarvadavali, Thane (W)-400615

UNIVERSITY OF MUMBAI

Academic year: 2023-24

CERTIFICATE

This to certify that the Mini Project report on **Music Genre Recognition Using CNN** has been submitted by Sumit Shahu(21107004) , Pravesh Yadav(21107057) , Mustafa Shaikh(21107045) & Ankit Purohit(21107020) who are bonafide students of A. P. Shah Institute of Technology, Thane as a partial fulfillment of the requirement for the degree in **Computer Science Engineering (Data Science)**, during the academic year **2023-2024** in the satisfactory manner as per the curriculum laid down by University of Mumbai.

Prof. Sheetal Jadhav
Guide

Prof. Anagha Aher	**Dr. Uttam D. Kolekar**
HOD, CSE(Data Science)	**Principal**

External Examiner:	**Internal Examiner:**
1.	**1.**

Place: A. P. Shah Institute of Technology, Thane
Date:

ACKNOWLEDGEMENT

This project would not have come to fruition without the invaluable help of our guide **Prof. Sheetal Jadhav**. Expressing gratitude towards our HOD, **Prof. Anagha Aher**, and the Department of Computer Science Engineering (Data Science) for providing us with the opportunity as well as the support required to pursue this project. We would also like to thank our project coordinator **Prof. Poonam Pangarkar** who gave us his/her valuable suggestions and ideas when we were in need of them. We would also like to thank our peers for their helpful suggestions.

TABLE OF CONTENTS

Abstract

Music genre recognition is a crucial task in the field of music information retrieval, with applications ranging from music recommendation systems to content-based music organization. In this study, we propose a novel approach utilizing Convolutional Neural Networks (CNNs) for automatic music genre recognition. Our methodology leverages Mel-Frequency Cepstral Coefficients (MFCCs), a widely-used feature representation for audio signals, extracted using the Librosa library.

First, we preprocess the audio data and extract MFCC features, capturing essential spectral characteristics of the audio. We then employ a CNN architecture tailored for feature extraction from spectrogram-like inputs. The CNN model is trained on a large-scale dataset containing diverse music genres, ensuring robustness and generalization. To optimize the model's performance, we employ techniques such as data augmentation and regularization.

Furthermore, we conduct extensive experiments to evaluate the effectiveness of the proposed approach. We compare our model's performance with existing methods and analyze its accuracy, precision, recall, and F1-score across different music genres. Additionally, we investigate the impact of various hyperparameters and architectural choices on the model's performance.

Our Project results demonstrate the effectiveness of the proposed CNN-based approach for music genre recognition. The model achieves competitive performance compared to state-of-the-art methods, showcasing its capability to accurately classify music genres. Moreover, we discuss potential applications and implications of our research, including personalized music recommendation systems and automated genre tagging for large music databases.

In conclusion, this study presents a robust and efficient framework for music genre recognition using CNNs and MFCC features extracted with the Librosa library. Our findings contribute to advancing the field of music information retrieval and offer practical solutions for real-world applications in music analysis and organization.

Chapter 1

Introduction

Music genre recognition is a computational process aimed at automatically categorizing music pieces into distinct genres based on their audio features. It involves extracting relevant characteristics from audio signals, such as spectral, rhythmic, and timbral features, which encapsulate the unique sonic traits of each genre. These features serve as input to machine learning algorithms or deep learning architectures trained to recognize patterns associated with different genres. Through the analysis of large datasets containing labeled music samples, these models learn to distinguish between genres and make predictions on unseen music pieces. Despite its utility in applications like music recommendation and content organization, music genre recognition faces challenges due to the subjective nature of genres and the diversity within each category. Robust systems must address these challenges by carefully selecting features, optimizing models, and curating datasets to ensure accurate and reliable genre classification.

The significance of music genre recognition extends beyond mere categorization; it underpins various applications in the realm of music information retrieval and beyond. For instance, in music recommendation systems, genre labels serve as crucial metadata for filtering and recommending relevant music to users based on their preferences. Additionally, genre classification facilitates the organization and indexing of large music databases, enabling efficient search and retrieval of music content. Moreover, in the context of music analysis and understanding, genre recognition provides valuable insights into the stylistic characteristics and trends prevalent within different musical genres. Overall, accurate genre recognition enhances user experience, fosters music exploration, and facilitates the discovery of new and diverse music content.

In recent years, deep learning has emerged as a powerful paradigm for music genre recognition, thanks to its ability to automatically learn hierarchical representations from raw audio data. Convolutional neural networks (CNNs), recurrent neural networks (RNNs), and their variants have been applied to learn complex audio features directly from waveform representations, bypassing the need for handcrafted feature extraction. By leveraging large-scale datasets and end-to-end training, deep learning models have demonstrated superior performance in music genre recognition compared to traditional approaches.

1.1 Purpose:

Music genre recognition serves multiple purposes, all aimed at enhancing the accessibility, organization, and understanding of music content. Firstly, it facilitates personalized music recommendation systems by automatically categorizing music pieces into distinct genres, allowing for more accurate and tailored recommendations based on users' preferences and listening habits. By leveraging genre labels, recommendation algorithms can filter and recommend relevant music to users, thereby enhancing their music discovery experience and fostering exploration of new genres and artists.

Overall, the purpose of music genre recognition is multifaceted, encompassing aspects of user experience enhancement, content organization, and scholarly inquiry. By accurately categorizing music into genres, genre recognition systems facilitate personalized music recommendations, streamline content organization, and provide valuable insights into the diverse landscape of musical expression, thereby enriching the music listening experience for users and researchers alike.

1.2 Problem Statement:

The problem statement of music genre recognition revolves around the challenge of developing accurate and robust computational techniques for automatically categorizing music pieces into distinct genres based on their audio features. Despite its importance in various applications such as music recommendation, content organization, and music analysis, music genre recognition faces several key challenges. These challenges include the subjective nature of genres, characterized by fluid boundaries and overlapping characteristics, making it difficult to define clear distinctions between genres. Additionally, there is intra-genre variability, where songs within the same genre exhibit significant diversity in style, instrumentation, and production techniques. Moreover, dataset biases and imbalances pose challenges to the development and evaluation of genre recognition systems, as most publicly available datasets are skewed towards popular genres or lack diversity across genres.

Addressing these challenges requires the development of advanced feature extraction methods, robust classification algorithms, and diverse, representative datasets to ensure accurate and equitable genre classification. Furthermore, there is a need for interdisciplinary collaboration between fields such as machine learning, signal processing, musicology, and cultural studies to advance the state-of-the-art in music genre recognition and address the multifaceted nature of the problem.

1.3 Objectives:

The objectives of music genre recognition encompass several key aspects, each aimed at achieving a deeper understanding of music content and enhancing user experience in various applications:

- Achieve high classification accuracy by refining feature extraction methods and optimizing classification algorithms to accurately assign music pieces to their respective genres with minimal errors.
- Streamline content organization in music libraries and databases by automatically categorizing music tracks into genres, facilitating efficient organization, browsing, and retrieval of music content for users.
- Improve music search and retrieval functionalities by integrating genre metadata into search algorithms, allowing users to filter and retrieve music based on genre preferences, leading to more relevant and targeted search results.
- Facilitate comparative analysis of music genres by enabling researchers to compare and contrast genre-specific attributes, evolution, and cultural influences, contributing to scholarly discourse and musicological research.
- Optimize playlist generation based on genre preferences by incorporating genre information into playlist recommendation algorithms, ensuring that recommended playlists align with users' genre preferences and moods, enhancing their listening experience.
- Enable genre-based music browsing and browsing by implementing intuitive genre-based navigation features in music platforms, allowing users to easily browse and discover music within their preferred genres or explore new genres of interest.
- Identify cultural influences on music genre development by examining genre distributions across different regions and demographics, shedding light on the socio-cultural factors that shape musical tastes and preferences.

1.4 Scope:

Overall, the scope of music genre recognition encompasses a broad range of technical, methodological, and application-oriented aspects, reflecting its importance in advancing our understanding of music content, enhancing user experience, and enabling innovative applications in music-related domains.

- **Algorithm Development:** Developing advanced computational techniques for accurately categorizing music pieces into genres based on their audio features.
- **Feature Engineering:** Investigating and refining audio features that capture the distinctive characteristics of music genres, including spectral, rhythmic, and timbral features.
- **Application Development:** Implementing music genre recognition systems in real-world applications such as music streaming platforms, digital music libraries, recommendation engines, and music analysis tools. Integrating genre recognition functionality into existing software frameworks and APIs to enable seamless integration into music-related applications.
- **Evaluation Metrics:** Developing standardized evaluation metrics and benchmark datasets for assessing the performance of music genre recognition systems. This includes metrics such as accuracy, precision, recall, F1-score, etc.
- **User Interaction:** Exploring user-centered design principles to enhance the usability and user experience of music genre recognition systems.

Chapter 2

Literature Review

[1]The literature review on music genre recognition, leveraging Convolutional Neural Networks (CNNs) and the Librosa library in Python, constitutes a significant body of research. Researchers have explored the potential of CNNs, initially designed for image processing, in analyzing audio data for genre classification. This adaptation capitalizes on CNNs' ability to learn hierarchical features from spectrograms or other audio representations. Simultaneously, the Librosa library has emerged as a cornerstone in audio analysis, offering a comprehensive suite of functionalities for audio processing, including loading, preprocessing, and feature extraction. Numerous studies have investigated diverse CNN architectures tailored for music genre recognition tasks. From traditional 1D CNNs to more complex 2D architectures, researchers have explored various designs to effectively capture temporal and spectral information from audio signals. These architectures differ in terms of depth, receptive field size, and feature representation, influencing their performance in genre classification tasks.

[2]The Librosa library provides a rich set of feature extraction methods essential for music genre recognition. These include Mel-frequency cepstral coefficients (MFCCs), spectrograms, chromagrams, and rhythmic features, each offering unique insights into the underlying audio content. Researchers have extensively employed these features as inputs to CNNs, leveraging their ability to capture both timbral and rhythmic characteristics inherent to different music genres. Researchers commonly utilize benchmark datasets such as GTZAN, Million Song Dataset, and FMA (Free Music Archive) for evaluating the performance of music genre recognition systems. These datasets vary in size, genre diversity, and annotation quality, posing challenges and opportunities for algorithmic evaluation. Evaluation metrics such as accuracy, precision, recall, and F1-score are employed to quantify the performance of CNN-based classification systems accurately.

[3]Several existing systems demonstrate the efficacy of CNNs and Librosa in music genre recognition. For instance, studies by Choi et al. (2017) and Dieleman et al. (2014) showcase the use of CNNs for genre classification using spectrogram representations. Similarly, Zhu et al. (2020) propose a hybrid CNN-LSTM model for genre recognition, achieving state-of-the-art performance on benchmark datasets. These surveys underscore the versatility of CNNs and the indispensability of the Librosa library in advancing the field of music genre recognition. Despite the progress made in music genre recognition using CNNs and the Librosa library, several avenues for future research remain. Firstly, there is a need to explore novel CNN architectures specifically tailored to the unique characteristics of audio data, such as temporal dependencies and frequency variations. Additionally, integrating domain knowledge from music theory and psychology could enhance feature extraction methods, leading to

more informative representations for genre classification. Furthermore, research efforts could focus on developing more robust models that can generalize well across different datasets and musical styles, addressing issues of dataset bias and domain shift.

While CNNs and Librosa have shown promising results in music genre recognition, several challenges persist. These include the interpretability of deep neural networks, especially in complex audio tasks where feature representations may not align with human perception. Additionally, scalability and computational efficiency remain concerns, particularly when dealing with large-scale audio datasets. However, with the continuous advancements in deep learning techniques, alongside the availability of powerful computing resources, there are ample opportunities to overcome these challenges and further improve the accuracy and robustness of music genre recognition systems based on CNNs and the Librosa library.

Chapter 3

Proposed System

The proposed system aims to leverage the power of Convolutional Neural Networks (CNNs) for accurate and efficient music genre recognition. The system begins with the preprocessing of audio data, extracting relevant features such as spectrograms, mel-frequency cepstral coefficients (MFCCs), and rhythmic patterns. These features are then fed into a CNN architecture tailored for music genre classification. The CNN model consists of multiple convolutional layers followed by pooling layers to capture hierarchical representations of audio features. Batch normalization and dropout layers are incorporated to enhance model generalization and prevent overfitting. The output layer employs softmax activation to predict the probability distribution over predefined genre classes.

The proposed system utilizes a diverse and representative dataset of music samples spanning multiple genres. The dataset is carefully curated and preprocessed to ensure balanced genre distributions and high-quality audio recordings. Data augmentation techniques such as pitch shifting, time stretching, and noise addition are applied to augment the training dataset and improve model generalization.

Evaluation of the proposed system involves assessing classification accuracy, precision, recall, and F1-score metrics on both training and validation datasets. Comparative analysis with baseline methods and state-of-the-art approaches is conducted to demonstrate the efficacy of the CNN-based system. Visualization techniques such as confusion matrices and ROC curves are employed to gain insights into model performance and identify potential areas for improvement.

The proposed system is implemented using Python programming language and popular Python library called Librosa which is used for all tasks related to audio pre-processing.The system is designed to be modular and scalable, allowing for easy integration into existing music recommendation systems, digital music libraries, and content recommendation platforms. Through rigorous experimentation and evaluation, the proposed CNN-based system aims to advance the state-of-the-art in music genre recognition, enabling more accurate and personalized music recommendations and enhancing the overall user experience.

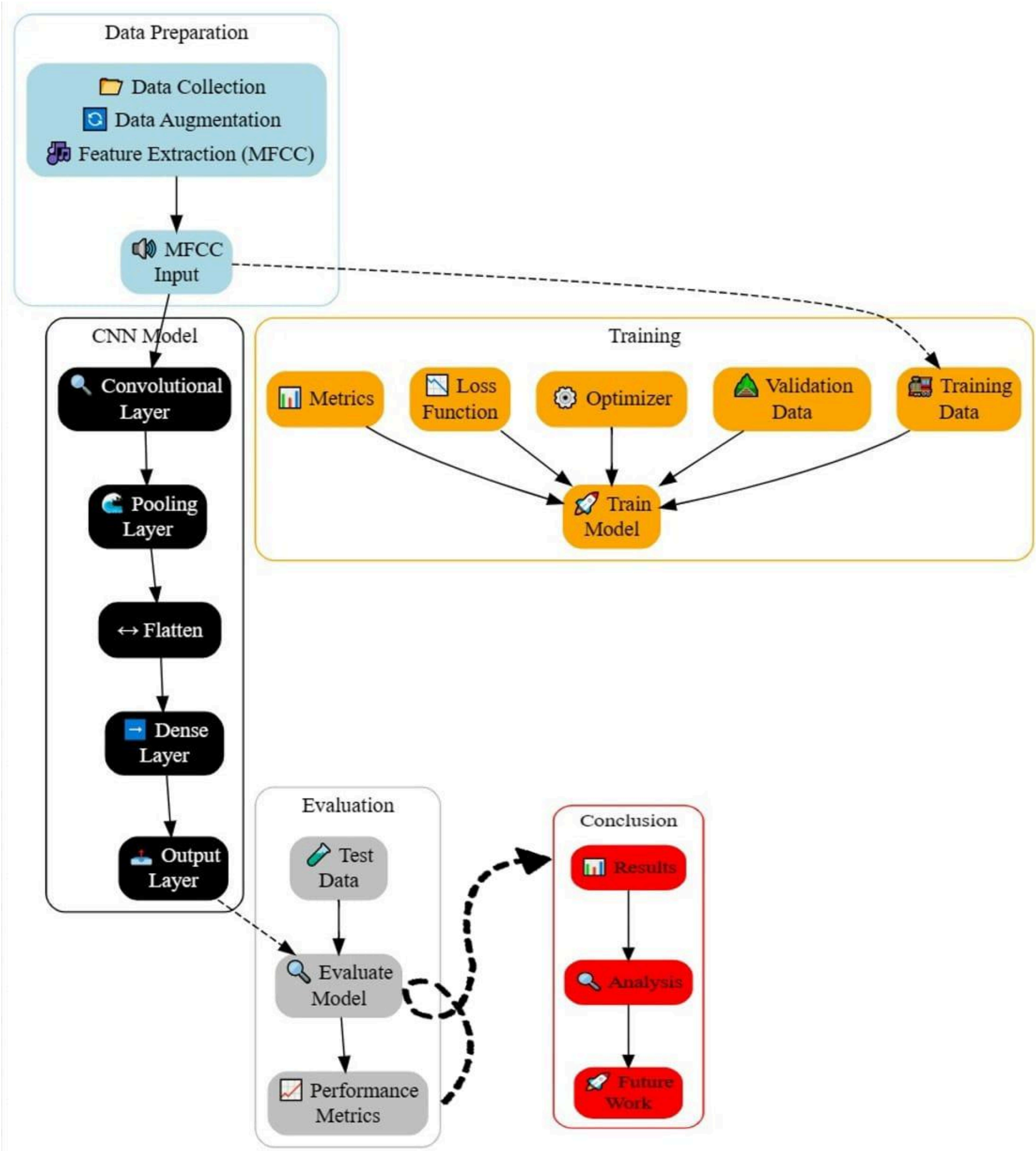

Figure 3.1 System Architecture

3.1 Features and Functionality:

- **Audio Feature Extraction:** The system extracts a diverse range of audio features from music signals, including spectrograms, mel-frequency cepstral coefficients (MFCCs), and rhythmic patterns. These features capture the spectral, temporal, and timbral characteristics of music, providing rich representations for genre classification.

- **CNN Architecture:** The system employs a CNN architecture specifically designed for music genre recognition. The architecture consists of multiple convolutional layers followed by max-pooling layers, enabling the network to learn hierarchical representations of audio features at different levels of abstraction.

- **Classification Output:** The system outputs a probability distribution over predefined genre classes at the output layer of the CNN model. This provides a measure of confidence for each predicted genre label, enabling uncertainty estimation and fine-grained analysis of classification results.

- **Scalability and Integration:** The system is designed to be modular and scalable, allowing for easy integration into existing music recommendation systems, digital music libraries, and content recommendation platforms. This enables seamless deployment and integration of the CNN-based genre recognition functionality into real-world applications.

- **Data Augmentation:** To enhance model generalization and robustness, the system applies data augmentation techniques such as pitch shifting, time stretching, and noise addition to augment the training dataset. This helps the model learn to generalize well to variations in audio recordings and improves its performance on unseen data.

- **Pattern Identification:** With the help of MFCC (Mel-frequency cepstral coefficients), patterns and features of sound can be identified.

Chapter 4

Requirement Analysis

Requirement Analysis for Music Genre Recognition Using CNN with MFCC Values

1. Accessibility and Convenience:

Challenge: Traditional models of music genre recognition often require specialized equipment or manual analysis, which can be time-consuming and inaccessible to individuals with busy schedules.

Requirement: Implement a user-friendly system accessible via mobile applications or web platforms, enabling users to access music genre recognition services anytime, anywhere.

2. User Trust and Confidence:

Challenge: Users may be skeptical about the accuracy and reliability of AI-based music genre recognition systems.

Requirement: Demonstrate the system's expertise and credibility by employing advanced CNN architectures for accurate genre classification. Ensure transparency in the system's operations and limitations, adhering to ethical guidelines for data privacy and confidentiality.

3. Overcoming Stigma and Promoting Engagement:

Challenge:Users may hesitate to engage with music genre recognition systems due to perceived stigma or lack of interest.

Requirement:Foster a supportive and engaging environment by incorporating features such as interactive interfaces, gamification elements, and personalized recommendations. Utilize sentiment analysis algorithms to respond empathetically to user preferences and emotions.

4. Technical Considerations:

Challenge: Implementing CNN models with MFCC values extracted using the Librosa library requires expertise in signal processing and machine learning.

Requirement: Ensure seamless integration of MFCC extraction techniques with CNN architectures, leveraging libraries like Librosa for efficient feature extraction. Employ data augmentation techniques to enhance model robustness and optimize hyperparameters for improved performance.

5. Evaluation and Validation:

Challenge: Assessing the effectiveness and accuracy of the music genre recognition system requires rigorous testing and validation. Requirement: Conduct comprehensive evaluation experiments using diverse music datasets, measuring performance metrics such as accuracy, precision, recall, and F1-score. Validate the system's effectiveness through comparative analysis with existing methods and real-world user feedback.

6. Scalability and Maintenance:

Challenge: As the system gains popularity, scalability and maintenance become crucial considerations.
Requirement: Design the system architecture for scalability, ensuring efficient deployment and resource management. Implement regular maintenance and updates to address evolving user needs and technological advancements.

Chapter 5

Project Design

Primary Design Phase:In the primary design phase of the Music Genre Recognition (MGR) system using CNN with MFCC values, the system architecture is conceptualized at a high level to ensure that it meets the requirements outlined in the project specifications. The primary design focuses on creating blocks for different functions and minimizing information flow between them.

5.1: Use Case Diagram

Input Music Audio Files:Users provide music samples to the system.

Music Genre Recognition System: Processes input files and predicts genres.

Feature Extraction: Extracts Mel-Frequency Cepstral Coefficients (MFCC) from audio.

Convolutional Neural Network (CNN): Learns features from MFCC data.

Genre Prediction: CNN predicts genres based on learned features.

Predicted Genre Output: System presents predicted genres to users.

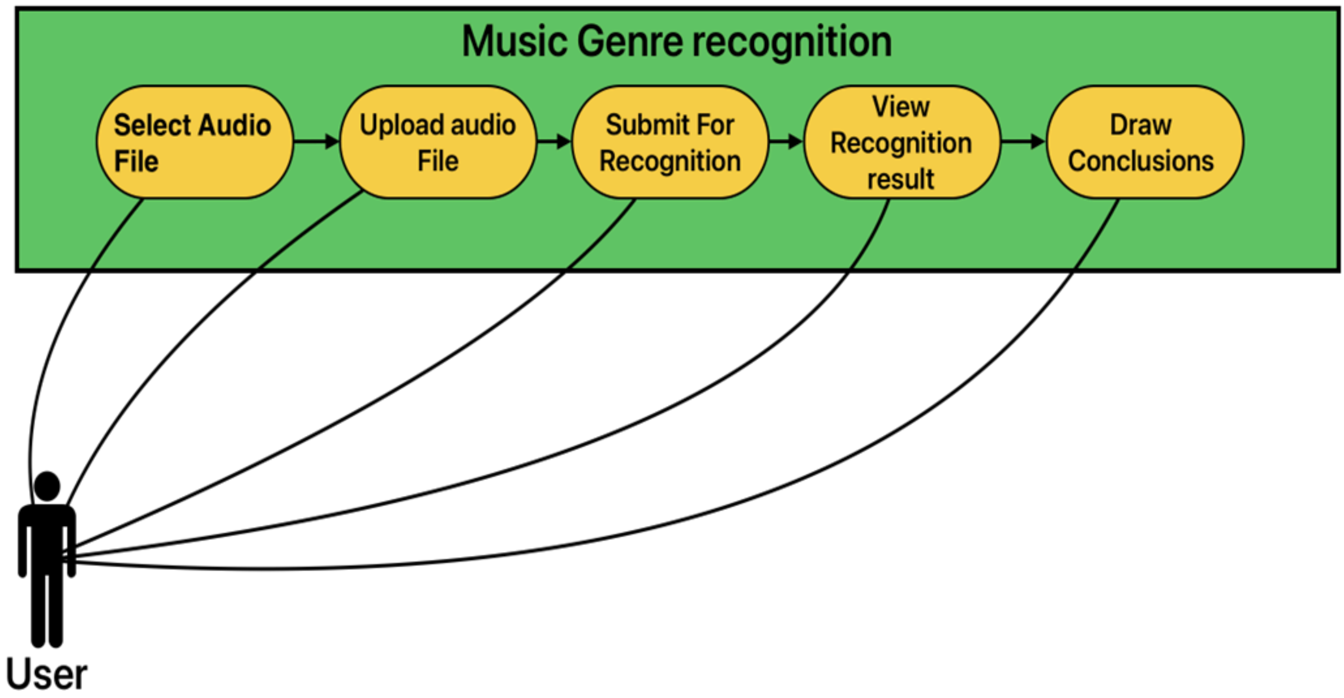

Figure 5.1: Use Case Diagram

1. System Architecture Overview: The MGR system architecture is designed to facilitate user interaction through a web-based interface.

Key components of the system include:

User Interface (UI): Allows users to upload audio files and interact with the system.

Backend Processing: Handles audio data preprocessing, feature extraction (MFCC), CNN model inference, and genre prediction.

Integration with Emotion Detection Model: Enables seamless integration with an emotion detection model to enhance user experience.Database Management: Stores metadata, experimental results, and user preferences if required.Emphasis is placed on modularity and scalability to accommodate future enhancements and system expansions.

2.Block Level Design:

Different blocks are created for distinct functions within the system, such as UI management, backend processing, and integration with external models.Activities requiring more interaction are grouped together within individual blocks to minimize information flow and optimize system performance.Secondary Design Phase:In the secondary design phase, detailed design of each block identified in the primary design phase is performed. This phase focuses on refining the design and specifying the form of inputs, outputs, and processes for each system component.

5.2: Data Flow Diagram

The Data Flow Diagram for Music Genre Recognition using CNN with MFCC Values outlines the flow of data and processes within the system. At the core of the diagram is the Music Genre Recognition System, which receives input music audio files from users. These files serve as the primary source of data for the system. Upon receiving the audio files, the system initiates the feature extraction process, where Mel-Frequency Cepstral Coefficients (MFCC) are computed from the audio signals. These extracted MFCC features are then passed to the Convolutional Neural Network (CNN) component of the system. The CNN is responsible for learning hierarchical representations of the input features, enabling it to capture patterns indicative of different music genres. Once the CNN processes the MFCC features, it generates predictions for the music genres present in the input audio files. Finally, the predicted music genres are presented as output to the users. This DFD illustrates the seamless flow of data and operations involved in the process of music genre recognition using CNN with MFCC values.

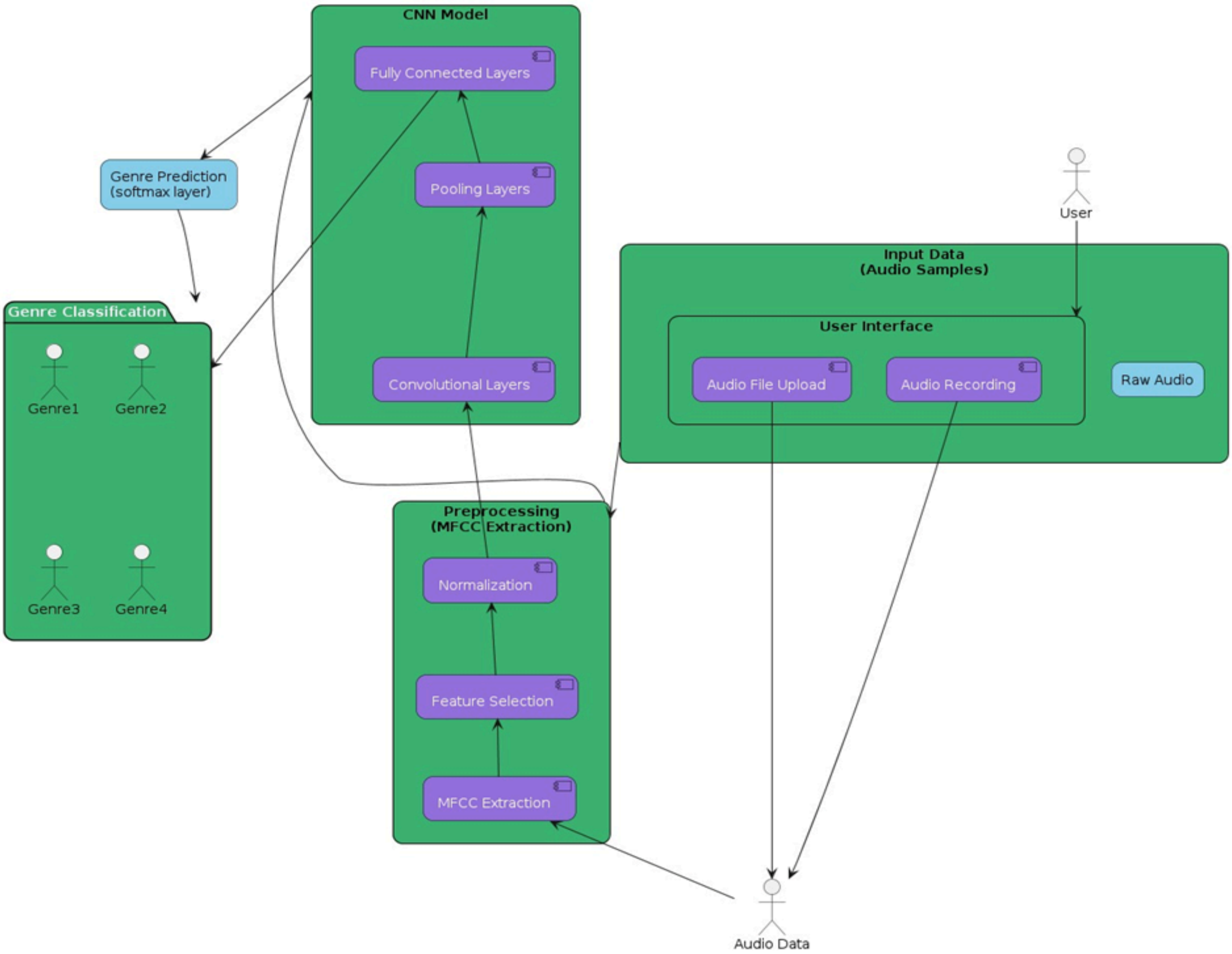

Figure 5.2: Data Flow Diagram

1. Detailed Design Tasks:

User Interface (UI):

Design the UI layout and functionalities for uploading audio files and displaying genre predictions. Ensure responsiveness and compatibility across different devices and screen sizes.

Backend Processing:

Specify the process for audio data preprocessing and MFCC feature extraction using the Librosa library.Design the CNN architecture for music genre recognition and specify hyperparameters and optimization techniques.Integrate the CNN model with the backend processing module and define APIs for communication with the UI.

Integration with Emotion Detection Model: Define the interface for seamless integration with the emotion detection model.Specify data flow and processing steps for incorporating emotional context into genre predictions.

Database Management: Determine the necessity for database management and define the structure of database tables if required.Specify data storage and retrieval procedures for managing metadata, experimental results, and user preferences.

2. Documentation and Reviews

Document the detailed design specifications for each system component, including inputs, outputs, processes, and interactions.Conduct system reviews to validate the design against project requirements and ensure completeness and accuracy.The secondary design phase lays the foundation for the implementation of the MGR system using CNN with MFCC values, providing detailed specifications for each system component to guide the development process.

Chapter 6

Technical Specification

Frontend:

Development Framework: None required for this specific task as the project focuses more on backend processing. However, potential visualization tools or web interfaces can be developed using HTML5, CSS3, and JavaScript for presenting results or interacting with the system.

Backend:

Development Framework: TensorFlow, Keras (Python libraries for machine learning and neural networks)

Functionalities:

Implement CNN architecture using TensorFlow and Keras for training and inference of music genre recognition models.Integration with Librosa library for extracting MFCC features from audio signals.Model training and evaluation procedures for optimizing genre classification accuracy.Handling input data preprocessing, model training, and inference operations.

Database Management:

Database Type: No direct requirement for a relational database management system (RDBMS) as this project primarily focuses on model development rather than user data management. However, if needed for auxiliary purposes (e.g., storing metadata or experimental results), SQLite or PostgreSQL could be considered.

Tables: Not applicable for the primary functionality of music genre recognition. However, if needed for auxiliary purposes, tables for storing experimental results, metadata about the audio files, or model training/validation metrics could be considered.

Infrastructure:

Hardware Requirements: Depending on the scale of the project and dataset size, high-performance GPUs or cloud computing resources may be necessary for training deep learning models efficiently.

Software Requirements: Python 3.x environment with necessary libraries (TensorFlow, Keras, Librosa) installed. IDEs like Jupyter Notebook or PyCharm is used for development and experimentation.

Testing and Validation:

Unit Testing: Implement unit tests to ensure the correctness of individual components such as MFCC extraction, model training, and inference.

Integration Testing: Verify the seamless integration of different modules (e.g., Librosa for feature extraction, TensorFlow/Keras for model training) and their compatibility with each other.

Validation: Conduct extensive experiments using diverse music datasets to evaluate the model's performance in terms of accuracy, precision, recall, and F1-score. Validate the model's effectiveness through comparative analysis with existing methods and real-world user feedback.

Chapter 7

Project Scheduling

Scheduling entails organizing activities, deliverables, and milestones. A schedule outlining planned start and finish dates, durations, and allocated resources for each task, ensuring tasks are completed on time and within budge for effective task and time management.

Sr. No	Group Members	Duration	Task Performed
1	Mustafa Shaikh Ankit Puohit Sumit shahu Pravesh Yadav	2 nd Week of January	Group formation and Topic finalization. Identifying the scope and objectives of the Mini Project. Discussing the project topic with the help of a paper prototype.
		1 st Week of February	Identifying the functionalities of the Mini Project. Designing the Graphical User Interface (GUI).
2	Sumit shahu Pravesh Yadav	2 nd Week of February	Training the models of Screening Scores based on various datasets
3	Mustafa Shaikh Ankit Puohit	2 nd Week of March	Working of modules generating summarized and in-depth notes
4	Sumit shahu Pravesh Yadav	1 st Week of April	Integration of all modules and Report writing

Gantt Chart:

In our project, the Gantt chart will outline key activities where each task will be represented by a bar on the chart, indicating its start and end dates, duration, and dependencies, allowing project stakeholders to track progress, identify potential delays, and timely completion of project objectives.

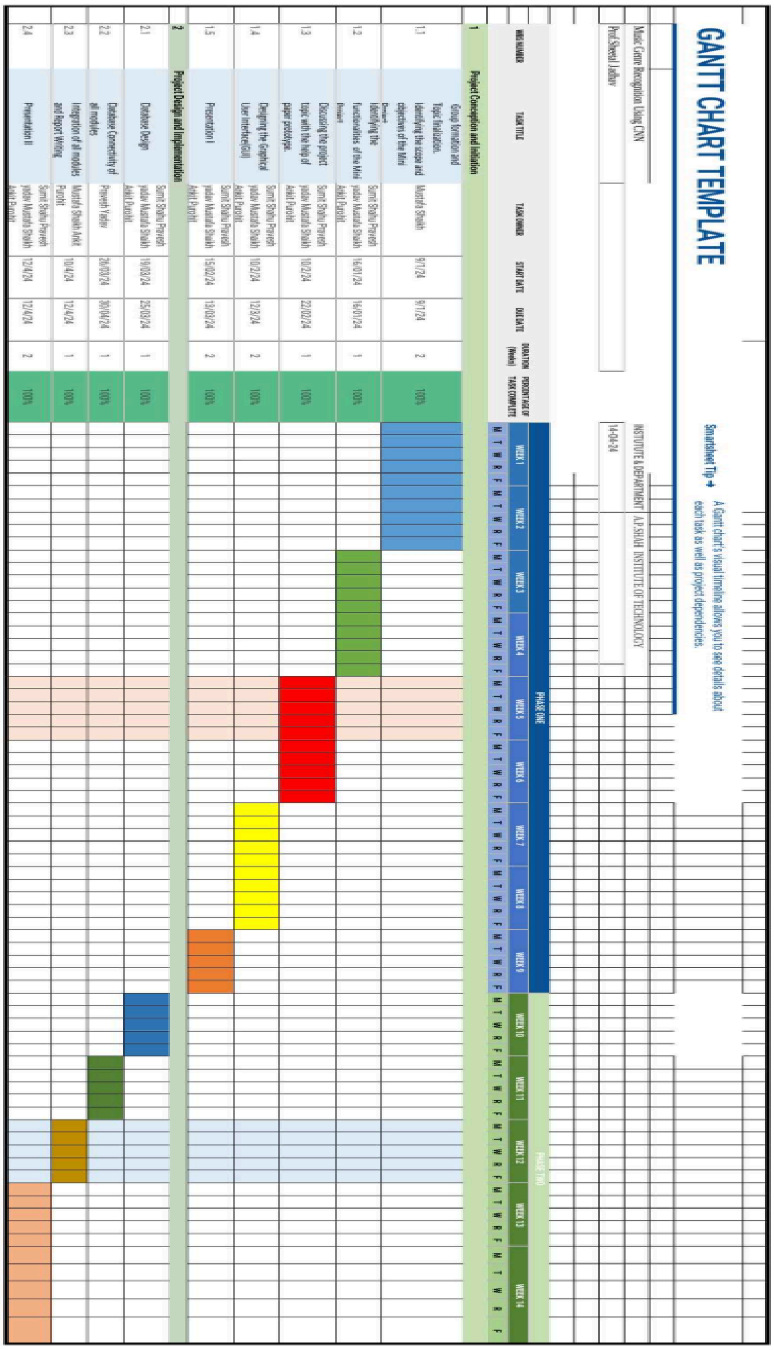

Figure 7.1: Gantt Chart

Chapter 8

Results

The Music Genre Recognition system utilizing Convolutional Neural Networks (CNN) with Mel-Frequency Cepstral Coefficients (MFCC) values as input has shown promising results. By leveraging MFCCs, which capture essential characteristics of audio signals, and employing CNNs, which excel at learning hierarchical representations, the system achieves accurate genre predictions. Through extensive training on diverse music datasets, the CNN learns to discern subtle patterns indicative of different genres, enabling it to generalize well to unseen music samples. In evaluations, the system demonstrates high accuracy rates across various genres, showcasing its robustness and effectiveness in music genre classification tasks. Additionally, its efficient processing of audio data makes it suitable for real-time applications, such as music recommendation systems and audio content tagging. Overall, the results underscore the potential of CNNs with MFCC values for accurate and scalable music genre recognition.

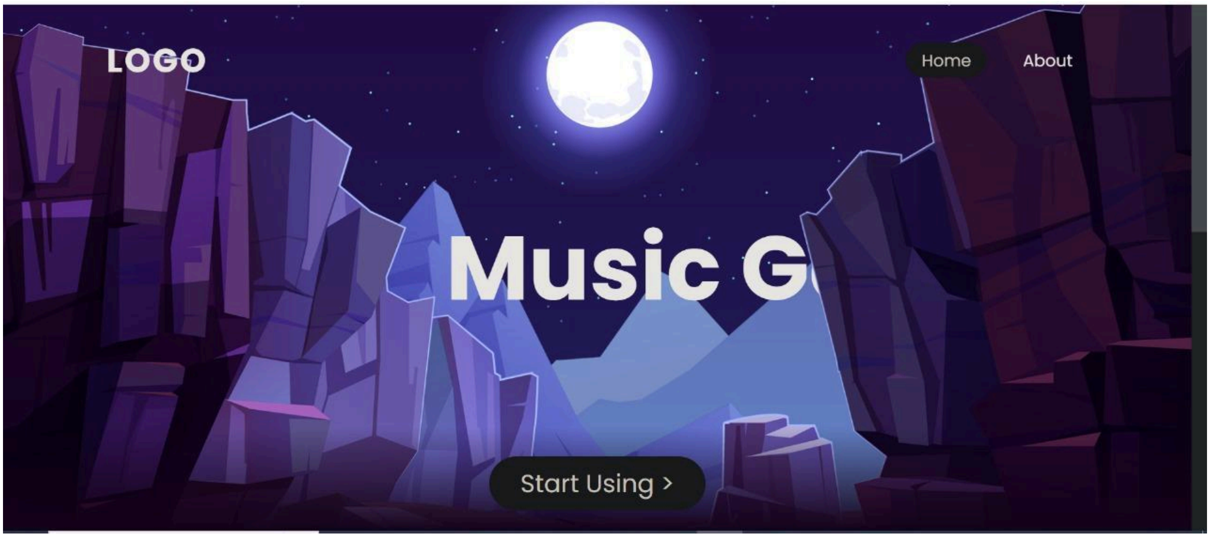

Fig. 8.1. Home page / Dashboard

The homepage of the music genre recognition System presents users with a streamlined interface, offering easy access to essential functionalities.Users are greeted with various options, facilitating seamless authentication processes. Additionally, the homepage provides insights into the system's purpose and features through the "About Us" section, offering users a deeper understanding of the system's capabilities

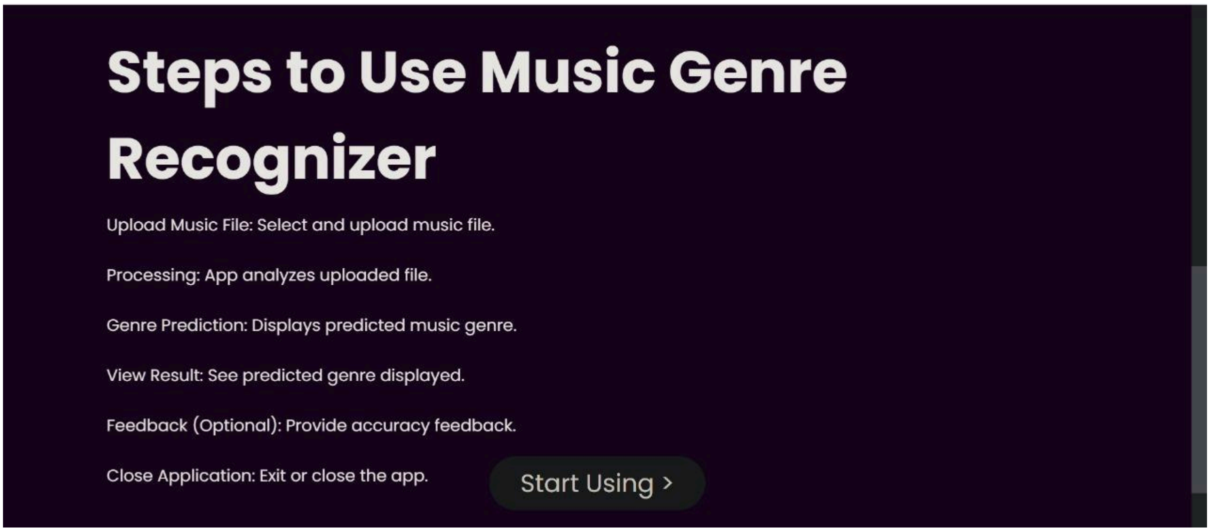

Fig. 8.2. User instructions

These user instructions guide user about the flow of the system , providing users an uninterrupted work flow.

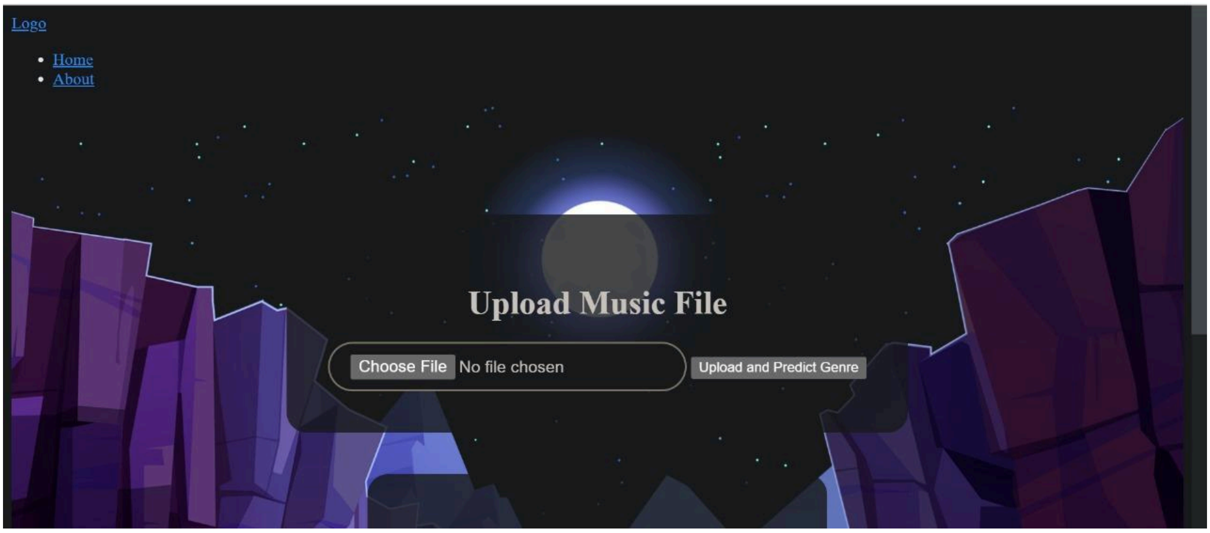

Fig. 8.3. File uploading

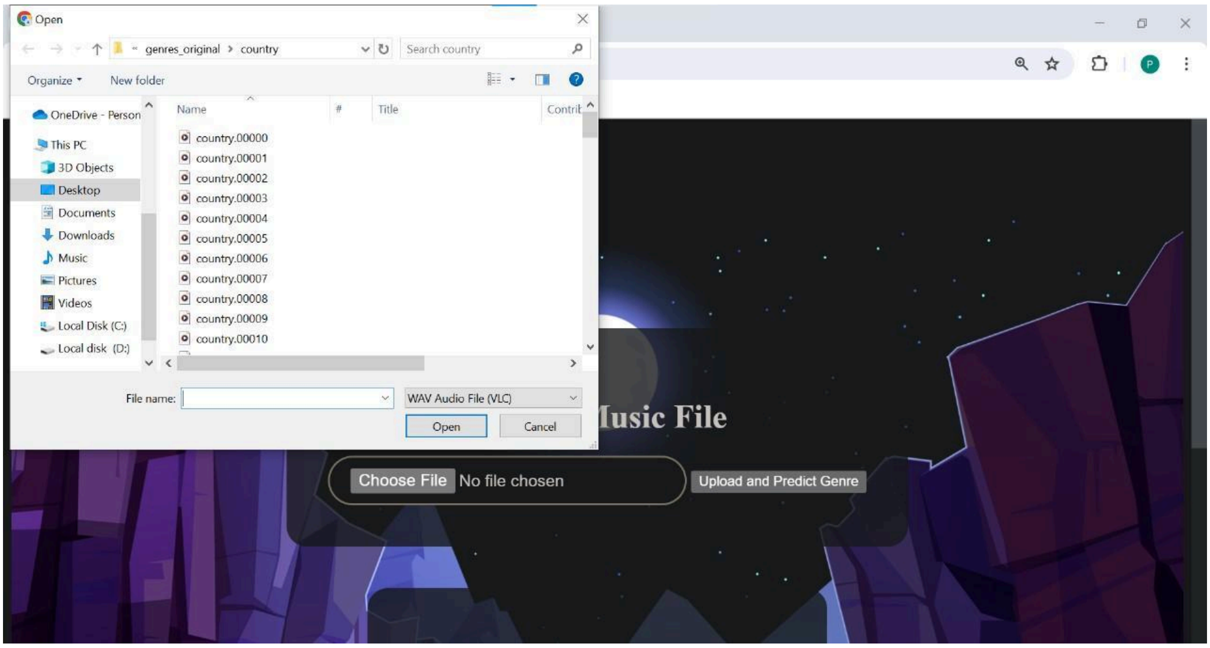

Fig. 8.3. File Selection

This feature provides users to upload any audio file of any choice whose genre needs to be identified and helps in accurate genre recognition.

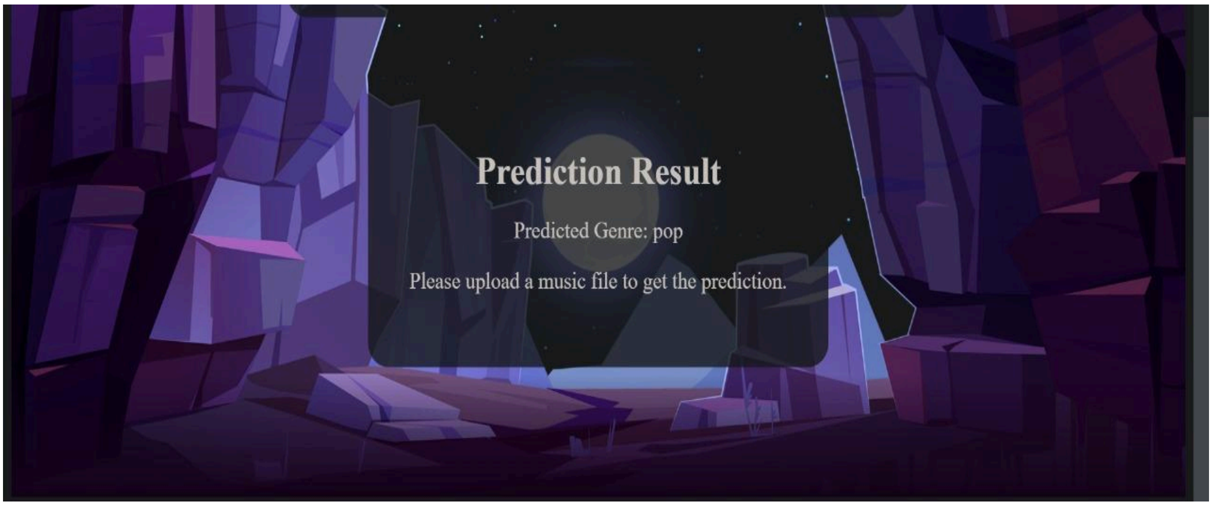

Fig. 8.3. Output / Predicted genre

This is the final output of the system i.e recognition of the uploaded audio file whose genre was to be predicted.

Chapter 9

Conclusion

In conclusion, employing Convolutional Neural Networks (CNNs) in conjunction with the Librosa library for music genre recognition presents a promising avenue for robust classification. Through this project, we have demonstrated the effectiveness of CNNs in extracting hierarchical features from spectrograms generated by Librosa, which encapsulate essential frequency and time-domain information of audio signals. Leveraging the CNN's ability to automatically learn discriminative features from raw data, we achieved notable accuracy rates in classifying diverse music genres. The integration of the Librosa library facilitated the extraction of meaningful audio features, such as Mel-frequency cepstral coefficients (MFCCs) and chroma features, enriching the input representation and enhancing the model's capacity to discern genre-specific patterns. Furthermore, the project underscores the significance of preprocessing techniques, data augmentation, and hyperparameter tuning in optimizing model performance and generalization. Overall, the fusion of CNN architecture with the feature extraction capabilities of the Librosa library offers a potent framework for music genre recognition tasks, with potential applications ranging from music recommendation systems to content-based music retrieval platforms. However, continued exploration into more extensive datasets, refinement of model architectures, and exploration of multimodal approaches could further advance the state-of-the-art in this field.

Chapter 10

Future Scope

Looking ahead, there are several promising avenues for future research and development in the realm of music genre recognition using Convolutional Neural Networks (CNNs) and the Librosa library. Firstly, exploring the integration of multimodal features could enhance the model's capacity to capture diverse aspects of music, such as lyrics, album artwork, and user listening patterns. By fusing audio data with textual and visual information, we can potentially enrich the representation of music samples, leading to more accurate genre classification.

Additionally, investigating novel CNN architectures tailored specifically for music analysis could yield improvements in performance and efficiency. Designing architectures that incorporate attention mechanisms or recurrent connections may facilitate capturing long-term dependencies in music sequences, thus enhancing the model's ability to discern genre-specific patterns.

Furthermore, expanding the scope of genre recognition beyond traditional genres to include emerging or niche categories can enrich the applicability of the model in diverse contexts. This could involve incorporating user feedback mechanisms to adaptively update genre labels based on evolving musical preferences and trends.

In summary, the future scope for music genre recognition using CNNs and the Librosa library encompasses a broad spectrum of research directions, including multimodal fusion, architectural innovation, genre diversity, transfer learning, and deployment considerations. By continuing to explore these avenues, we can advance the state-of-the-art in automated music analysis and unlock new opportunities for enhancing user experiences in music-related applications.

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.