MUSIC GENRE CLASSIFICATION

Team Information:

Name	Student ID	Email
Bala Naga Narasimha Reddy Machha	11671010	BNNarasimhaReddyMachha@my.unt.edu
Navya Bade	11648430	NavyaBade@my.unt.edu
Yuvaraj Sriramoju	11710243	YuvarajSriramoju@my.unt.edu
Sreeja Bommineni	11648419	Sreejabommineni@my.unt.edu

Problem Statement

Music genre classification aims to categorize songs into different genres such as rock, jazz, classical, and hip-hop using machine learning models. This task helps in enhancing user experience in music streaming platforms by improving recommendation systems and can be helpful in organizing large music libraries. By classifying the music perfectly, music platforms such as Spotify, Amazon Music etc., can offer personalized suggestions, better search functionality, and automated playlist generation, ensuring users discover music aligned with their preferences.

The main challenges in this problem include handling noisy data, class imbalance and genre overlap, where songs exhibit characteristics of multiple genres. Additionally, processing large-scale audio datasets efficiently and extracting relevant features like MFCCs or chroma from raw audio can be some technical challenges.

Objective and Research Questions

Objective: The main objective of this project is to classify music into various genres by using neural networks. The goal is to explore various feature extraction techniques and train the model to get better accuracy. This can be useful in music recommendation system.

Research Questions:

1. Does the model's accuracy is affected if the features are reduced or by using normalization?
2. Will there be a chance to increase the model's accuracy to classify if we use data augmentation method for small datasets?
3. What audio features will be useful in classifying the music genres more accurately?
4. Can the feature selection method be used to improve the readability of the model and help the model reduce overfitting?

Dataset Description

Source: We are using GTZAN dataset from Kaggle. The audio files were collected in 2000-2001 from a variety of sources including personal CDs, radio, microphone recordings.

Dataset Link: https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification/data/

Size: The dataset has 10 music genres, each containing 100 audio files of 30 seconds each. The total dataset is of size 1.41 GB.

Characteristics: The dataset has audios of 10 genres and spectrogram images of respective audio files. The dataset also contains a csv file of extracted features mean and average value of different extracted features from the audio files for 30 seconds and 3 seconds. But due to computational constraints, we plan to use only a part of the dataset.

Data Pre-Processing

After collecting the data, we perform pre-processing steps. In this step, we make sure that all the audio files in the dataset are usable by checking if there are any missing or corrupted files. In case of corruption, we are going to repair or remove those files. We will also make sure that pixel values are normalized for better training by rescaling. We also plan to handle any class imbalance by applying oversampling and under sampling for the classes.

Proposed Feature Engineering Techniques

Feature Creation:

1. **Frequency-Domain Features:** Feature likes Spectral Centroid, Spectral Bandwidth, Spectral Roll-Off capture information from the frequency spectrum of the signal.
2. **Time-Domain Features:** Features such as Zero Crossing Rate (ZCR), Root Mean Square Energy (RMSE), Temporal Entropy can be extracted from the raw audio signals and give information about the variations in amplitudes.
3. **Mel-Spectrogram:** We plan to convert the raw audio data to Spectrogram image using Mel Scale. This generates the Spectrogram images which are visual representation of the frequency content of audio signal.

Feature Transformation:

1. **Log Transformation:** Applying log transformations on the features which have high variance. Log Transformation helps in reducing skewness.
2. **Normalization Techniques:** We plan to normalize the pixel values of the spectrogram image to ensure that the model treats all features equally by using Normalization Techniques such as Min-Max Normalization, Z-Score Normalization etc.

3. **Image Processing Techniques:** We also intend to apply image processing techniques such as color grading, smoothing, edge filtering to the processed spectrogram image. This helps the model to learn the differentiate the features and learn effectively.

Feature Selection:

Generally, Data contains many features. Training the model by using all the features can misguide the model as non-relevant features might result in false learning. So, we use different methods such as correlation analysis, Variance Thresholding, Recursive Feature Elimination (RFE) etc. to select the best feature for training the model.

Baseline Model and Evaluation Plan

Baseline Model: We initially plan to use basic Convolutional Neural Networks (CNN) as baseline model.

Evaluation Plan:

We will use common evaluation metrics such as Accuracy, Precision, recall and F-1 score. If the final dataset has imbalance, then, we will focus on improving precision and recall avoiding biases toward more frequent genres. \

We will compare the results of our model with both normal spectrogram images generated from the raw audio files and with the engineered features images.

Project Timeline

Milestone 1: Project Proposal and Data Collection (10/19/2024)

The initial step of the project is developing the project proposal that outlines the goals of the project, the objectives and methodology. And GTZAN dataset is sourced from Kaggle.

Milestone 2: Data Preprocessing (10/26/2024)

The next task is data preprocessing, where we clean the dataset, handle missing values, converting the audio files into suitable formats.

Milestone 3: Feature Engineering (11/05/2024)

In this task, we will apply various proposed feature engineering techniques on the processed data, which can help in better accuracy.

Milestone 4: Model Development (11/20/2024)

During this milestone, we develop the model using appropriate algorithm. And then the model will be trained on the dataset that we have considered.

Milestone 5: Model Evaluation and Final Report Preparation (12/02/2024)

In this milestone where we will evaluate the built model using the test datasets, we are going to analyze the performance metrics. And we are going to document our findings and insights into the final project report.

Milestone 6: Video Recording and Final Submission (12/07/2024)

This is the final milestone where we create a presentation in which we summarize the project and record a video in which we will explain the methodology and results.

We are going to submit the final project report, code and the video.

Anticipated Challenges and Risks

1. **Data Quality Issues**
   Challenge: The GTZAN dataset may have missing data, low-quality audio files, and noise data which can affect the performance of the model.
   Strategy for Overcoming:
   - Data Cleaning needs to be done before being used.
   - Consider augmenting the dataset with additional audio files if needed.
2. **Computational Constraint**
   Challenge: Feature extraction from the audio files and training the model may require more computational resources which leads to possible slowdowns or failures sometimes.
   Strategy for Overcoming:
   - Using GPU acceleration where possible.
   - Testing on a smaller dataset subset to ensure that model runs smoothly.
3. **Model Overfitting**
   Challenge: I some cases, if the parameters are not set correctly, model may be overfit to the training data which leads to poor performance.
   Strategy for Overcoming:
   - Use K-fold cross-validation to assess model performance
   - Some genres can be similar to each other. Understanding genre similarity can help adjust model's loss function

References

1. https://www.kaggle.com/datasets/andradaolteanu/gtzan-dataset-music-genre-classification/
2. Music Genre Classification: Training an AI model, Keoikantse Mogonediwa, Academy of computer science, University of Johannesburg, Auckland Park JHB 2092, South Africa.
3. Music Genre Classification via Machine Learning, Category: Audio and Music, Li Guo(liguo94), Zhiwei Gu(zhiweig), Tianchi Liu(kitliu5).
4. Music Genre Classification using Machine Learning Techniques, Hareesh Bahuleyan.