Genre Classification of Music Tracks Using the GTZAN Dataset

Project Category: Image Classification/Other (Audio Classification)

Team Members: Inês Laborinho Schwartz 128119, Tristan Bertrand

The objective of this project is to develop a machine learning model capable of accurately classifying music tracks into different genres using the GTZAN dataset. Genre classification is a fundamental problem in the field of music information retrieval and has significant applications in music recommendation systems, music library organization, and streaming services. Accurate genre classification can enhance user experience by enabling personalized music recommendations and efficient music discovery. Lastly, being able to categorize individual characteristics of a dataset (ei. music tracks) into a greater theme (genre, in this case) is a very relevant and useful skill that machine learning models can assist with.

The GTZAN dataset is a publicly available dataset consisting of 1000 audio tracks categorized into 10 genres: blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, and rock. Each genre contains 100 tracks, each with a duration of 30 seconds. This dataset was chosen because it provides a standard benchmark for music genre classification and because it presents a variety of challenges around audio and machine learning which we did not get a chance to work with too much during the class.

Several challenges will be inherent to this project. Feature extraction is crucial, as identifying the most relevant features from audio signals is essential for accurate genre classification. Generalization is another challenge, where the model must perform well on unseen data and avoid overfitting on the training set. Additionally, establishing robust evaluation metrics is necessary to measure the model's performance comprehensively. For instance, training the dataset to differentiate similar genres such as rock versus metal can be complex, and songs that belong to multiple genres might be better classified by outputting probabilities for each genre.

The methodology of this project begins with exploratory data analysis (EDA) to understand the dataset better. This involves statistical analysis and visualization of the data. Data preprocessing follows, which includes converting audio files into a consistent format and sample rate, and extracting audio features. The most commonly used method we identified for **feature extraction is the Mel-Frequency Cepstral Coefficients (MFCCs)**, so we will start by using that.

For model selection, we will implement a **Convolutional Neural Network (CNN)** to classify genres directly from spectrogram images generated from audio tracks. The dataset will be **split into training, validation, and test sets**. We will perform hyperparameter tuning using **cross-validation on the training set** to optimize model performance.

To evaluate the results, we will employ several performance metrics: **accuracy, precision, recall, F1-score, and a confusion matrix**. Accuracy will measure the proportion of correctly classified instances, while precision and recall will help understand the model's performance in predicting each genre accurately. The F1-score will ideally balance both metrics. The confusion matrix will visualize the classification results and identify misclassifications. Cross-validation will ensure the robustness of the model and its ability to generalize to new data.

Ultimately, this project aims to provide insights into the effectiveness of machine learning approaches for genre classification and demonstrate the practicality of such applications.