

Final Project Proposal: Music Genre Classification Using Spectrograms and CNNs

For our final project, we plan to build a music genre classification system using machine learning. The main idea is to take raw audio clips and convert them into spectrograms, which are visual representations of sound intensity across time and frequency. By transforming audio signals into images, we can treat the problem as an image-classification task and apply deep learning models to learn patterns that distinguish different music genres.

We will use the GTZAN Genre Collection, a widely used dataset containing 1,000 audio tracks across ten genres such as Classical, Jazz, Pop, Rock, Hip-Hop, and Metal. Each audio file will be converted into a mel-spectrogram using standard audio preprocessing steps such as STFT and log-frequency scaling. These spectrogram images will serve as the input to our models.

In this project, we plan to test several models. We will start with a simple baseline such as a fully connected neural network trained on flattened spectrograms. After that, we will experiment with multiple convolutional neural network (CNN) architectures to capture time-frequency patterns more effectively. We will compare their performance through train/validation/test accuracy, loss curves, and confusion matrices to evaluate how well each model distinguishes between genres.

If time permits, we may further explore improvements such as data augmentation, batch normalization, or using pre-trained vision models after resizing spectrograms. Our final deliverables will include the trained models, evaluation results, and a short demo showing how the system predicts the genre of an unseen audio clip.