Music Genre Classification

Submitted By:

Abhishek Kumar Singh	1805210001	Supervisor : Dr. Upendra Kumar
Akash Saha	1805210005	Co-Supervisor : Mr. Ajeet Shukla
Shivam Yadav	1805210049	

Introduction

- Music Genre Classification means to automatically classify different musical genres from audio files.
- We will classify these audio files using their low-level features of frequency and time domain.
- We have used the GTZAN genre classification dataset which is the most recommended dataset for the music genre classification project .

Motivation

- Recommend new music based on user's listening history
- Automatically organise music libraries
- Create reproducible music

Objective

- To classify audio files (in ".wav" or ".mp3" format) into 10 musical genres: blues, classical, country, disco, hip-hop, jazz, metal, pop, reggae, and rock

- Try to introduce a web interface to upload music files to find their genres.

Methodology

Dataset

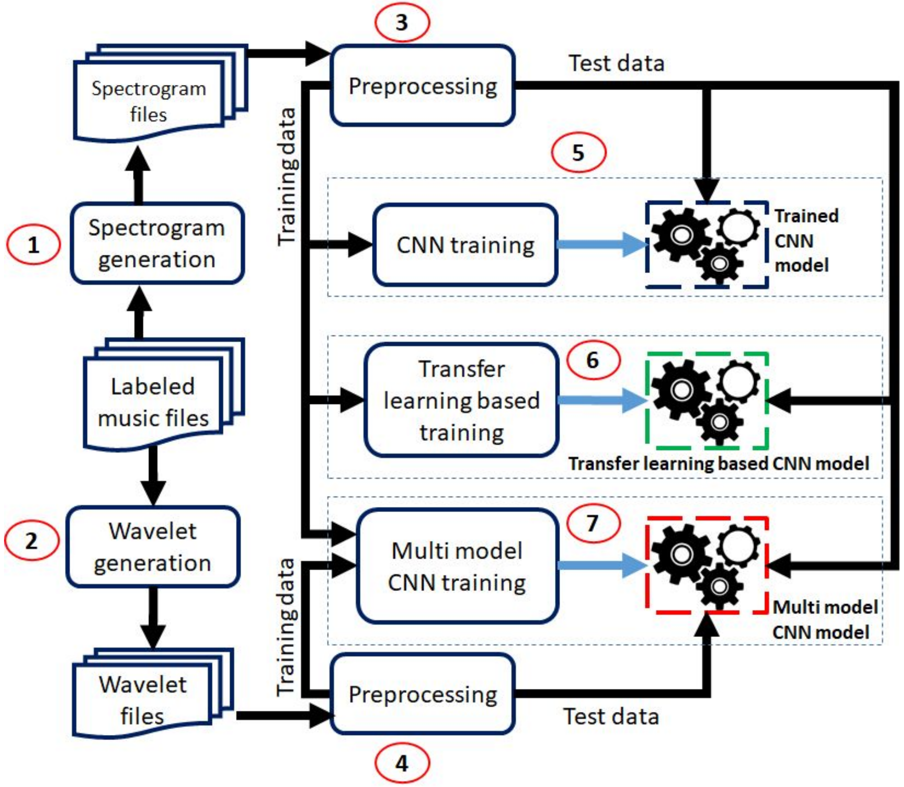
- For this project, the dataset that we will be working with is the GTZAN Genre Classification dataset
- It consists of 1,000 audio tracks, each 30 seconds long. It contains 10 genres, each represented by 100 tracks.
- The 10 genres are as follows: Blues, Classical, Country, Disco, Hip-hop, Jazz, Metal, Pop, Reggae, Rock

Folders in the dataset

- Genres original — A collection of 10 genres with 100 audio files each, all having a length of 30 seconds
- Images original — A visual representation of each audio file. One way to classify data is through neural networks because NN's usually take in some sort of image representation.
- 2 CSV files — Containing features of the audio files. One file has for each song (30 seconds long) a mean and variance computed over multiple features that can be extracted from an audio file. The other file has the same structure, but the songs are split before into 3-second audio files.

Flow

The below figure represents the overview of our methodology for the genre classification task



❑ Spectrogram generation:

A spectrogram is a visual representation of the spectrum signal frequencies as it varies with time. We will use the librosa library to transform each audio file into a spectrogram.

❑ Wavelet generation:

The Wavelet Transform is a transformation that can be used to analyze the spectral and temporal properties of non-stationary signals like audio. We will use the librosa library to generate wavelets of each audio file.

Preprocessing and Feature Extraction

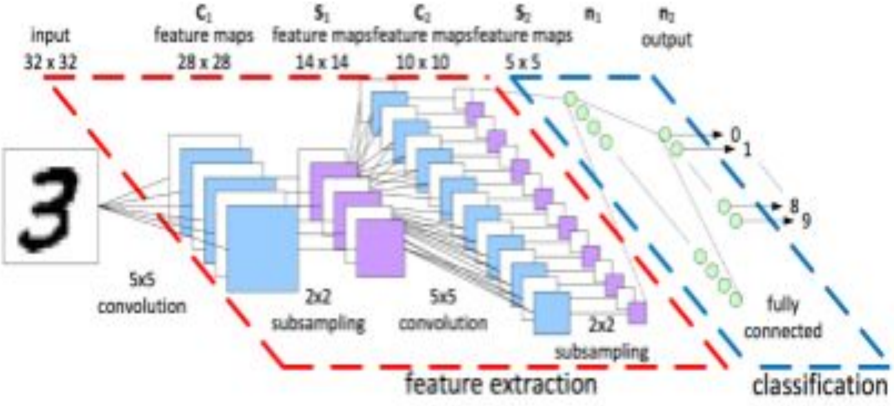
- ❑ After generating spectrograms and wavelets, we will apply general image preprocessing steps to generate training and testing data. Preprocessing of data is required before we finally train the data.

- ❑ We can't have text in our data if we're going to run any kind of model on it. So before we can run a model, we need to make this data ready for the model. To convert this kind of categorical text data into model-understandable numerical data, we use the LabelEncoder class.

Model Training

- We will pass both spectrogram and wavelet data into the CNN model for the training in this experiment. We are using the late-fusion technique in this multimodal training.

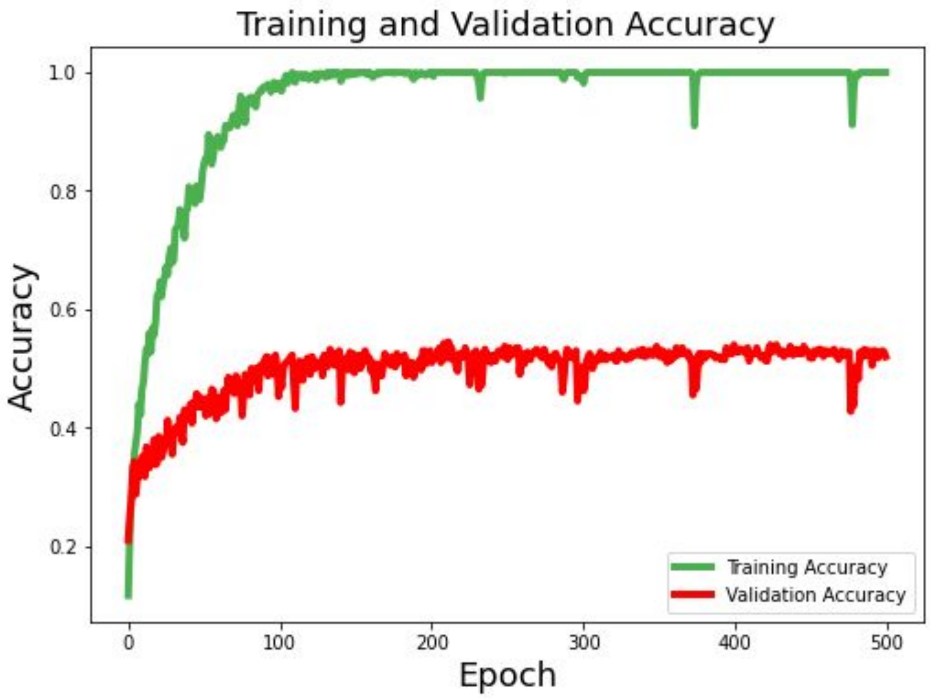
- We will use a Pytorch framework to design our Convolutional Neural Network (CNN) with 3 layers of convolutions and a final fully connected layer with softmax activation with 10 outputs (for 10 genres).

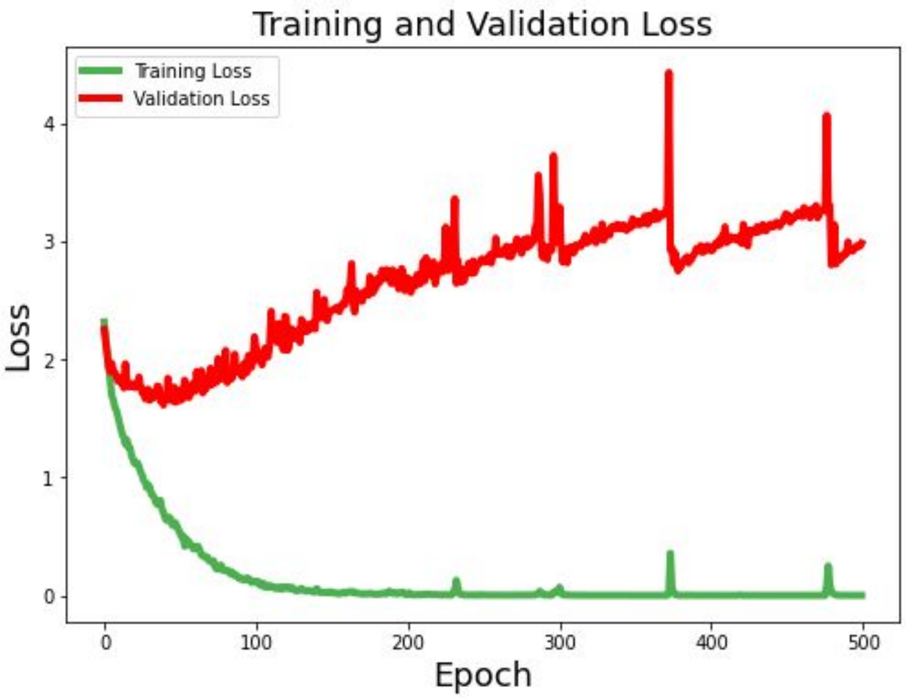


Experimental Results

- Accuracy was used as the main performance metric for our model
- The accuracy of our model is 76%. Accuracy kept on increasing with each epoch and became maximum at epoch 210 and became constant after that.
- Our model is able to accurately recognize rap and rock music because they tend to have very distinctive sounds and vocal stylings.
- Classical and jazz music are both typically instrumental and use similar instruments, causing the model to easily confuse them and thus cause worse performance with both genres

Accuracy and Loss graph





Training epochs data

```
Epoch: 85     Train loss : 1.37     Valid loss : 1.47     Train acc : 0.52     Valid acc : 0.48
Epoch: 86     Train loss : 1.37     Valid loss : 1.51     Train acc : 0.49     Valid acc : 0.48
Epoch: 87     Train loss : 1.44     Valid loss : 1.51     Train acc : 0.50     Valid acc : 0.47
Epoch: 88     Train loss : 1.43     Valid loss : 1.42     Train acc : 0.54     Valid acc : 0.52
Epoch: 89     Train loss : 1.33     Valid loss : 1.40     Train acc : 0.56     Valid acc : 0.51
Epoch: 90     Train loss : 1.29     Valid loss : 1.39     Train acc : 0.57     Valid acc : 0.54
Epoch: 91     Train loss : 1.28     Valid loss : 1.39     Train acc : 0.57     Valid acc : 0.53
Epoch: 92     Train loss : 1.27     Valid loss : 1.41     Train acc : 0.55     Valid acc : 0.53
Epoch: 93     Train loss : 1.31     Valid loss : 1.44     Train acc : 0.53     Valid acc : 0.52
Epoch: 94     Train loss : 1.32     Valid loss : 1.49     Train acc : 0.49     Valid acc : 0.47
Epoch: 95     Train loss : 1.41     Valid loss : 1.46     Train acc : 0.53     Valid acc : 0.50
Epoch: 96     Train loss : 1.36     Valid loss : 1.45     Train acc : 0.52     Valid acc : 0.50
Epoch: 97     Train loss : 1.36     Valid loss : 1.49     Train acc : 0.52     Valid acc : 0.50
Epoch: 98     Train loss : 1.40     Valid loss : 1.36     Train acc : 0.56     Valid acc : 0.51
Epoch: 99     Train loss : 1.27     Valid loss : 1.37     Train acc : 0.57     Valid acc : 0.53
Epoch: 100    Train loss : 1.26     Valid loss : 1.40     Train acc : 0.55     Valid acc : 0.52
Epoch: 101    Train loss : 1.27     Valid loss : 1.42     Train acc : 0.54     Valid acc : 0.51
Epoch: 102    Train loss : 1.32     Valid loss : 1.35     Train acc : 0.57     Valid acc : 0.55
Epoch: 103    Train loss : 1.24     Valid loss : 1.33     Train acc : 0.58     Valid acc : 0.54
Epoch: 104    Train loss : 1.22     Valid loss : 1.41     Train acc : 0.55     Valid acc : 0.53
Epoch: 105    Train loss : 1.31     Valid loss : 1.40     Train acc : 0.56     Valid acc : 0.52
Epoch: 106    Train loss : 1.27     Valid loss : 1.39     Train acc : 0.56     Valid acc : 0.53
Epoch: 107    Train loss : 1.30     Valid loss : 1.31     Train acc : 0.60     Valid acc : 0.56
Epoch: 108    Train loss : 1.20     Valid loss : 1.32     Train acc : 0.59     Valid acc : 0.54
Epoch: 109    Train loss : 1.20     Valid loss : 1.35     Train acc : 0.58     Valid acc : 0.57
Epoch: 110    Train loss : 1.22     Valid loss : 1.37     Train acc : 0.57     Valid acc : 0.55
Epoch: 111    Train loss : 1.25     Valid loss : 1.33     Train acc : 0.59     Valid acc : 0.55
Epoch: 112    Train loss : 1.19     Valid loss : 1.32     Train acc : 0.60     Valid acc : 0.55
Epoch: 113    Train loss : 1.20     Valid loss : 1.34     Train acc : 0.59     Valid acc : 0.54
Epoch: 114    Train loss : 1.21     Valid loss : 1.45     Train acc : 0.54     Valid acc : 0.51
Epoch: 115    Train loss : 1.33     Valid loss : 1.38     Train acc : 0.57     Valid acc : 0.53
Epoch: 116    Train loss : 1.27     Valid loss : 1.34     Train acc : 0.60     Valid acc : 0.56
Epoch: 117    Train loss : 1.20     Valid loss : 1.40     Train acc : 0.56     Valid acc : 0.53
Epoch: 118    Train loss : 1.28     Valid loss : 1.34     Train acc : 0.60     Valid acc : 0.55
Epoch: 119    Train loss : 1.19     Valid loss : 1.29     Train acc : 0.61     Valid acc : 0.57
```

Conclusion

In this project, we built a machine learning model to classify music into 10 different genres based on their acoustic signature. We have used Convoluted Neural Network(CNN) and have achieved an accuracy of 76%

Our next step is to build and deploy a website where anyone can upload a song/sound and get the genre of the uploaded music/sound.

References

- Hareesh Bahuleyan. Music genre classification using machine learning techniques. CoRR, abs/1804.01149, 2018.

- François Chollet et al. Keras. https://keras.io, 2015.

- Mingwen Dong. The convolutional neural network achieves human-level accuracy in music genre classification. CoRR, abs/1802.09697, 2018.

- Y. Panagakis, C. Kotropoulos, and G. R. Arce. Music genre classification via sparse representations of auditory temporal modulations. In 2009 17th European Signal Processing Conference, pages 1–5, Aug 2009.

- F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M.
Music Genre Classification and Recommendation by Using Machine Learning
Techniques Innovations in Intelligent Systems and Applications Conference (ASYU), IEEE.
the Posted: 2018

THANK YOU