

# CS543: Music Genre Recognition through Audio Samples

**Vedant Choudhary**

Rutgers University

New Brunswick, New Jersey 08901

vedant.choudhary@rutgers.edu

**Aditya Vyas**

Rutgers University

New Brunswick, New Jersey 08901

aditya.vyas@rutgers.edu

## Abstract

This project is part of our coursework for CS543. We aim to implement music genre recognition through audio samples using deep learning techniques. The idea of the project is that audio samples provide advanced features which are more accurate at recognizing a genre than using lyrics. Audio samples can be converted to image spectrogram, reducing the problem to that of an image classification. This allows Convolutional Neural Networks to find features of the audio sample. The data set considered is Free Music Archive (FMA) data set which consists of 106,574 tracks, 16,916 artists, 15,234 albums, 161 genres.

## 1 Introduction

Recognizing genre of a music through lyrics has been a well researched area, but is still a naive approach. It suffers from the limitation that it does not use any musical feature of the song, like the frequency of tones or pattern of beats. In our project, we are trying to overcome this limitation by incorporating audio samples and deriving features from it to present a better, more accurate genre representation of a song. We go along the project by using the *small* dataset of FMA, which has 8,000 tracks of 30s, and 8 balanced genres.

### 1.1 Problem Statement

Music Information Retrieval (MIR) is a wide research area. It encompasses fields like machine learning, signal processing and musical theory. Music Genre Recognition (MGR), the topic for our project is one of the most widely tackled problem in MIR. According to (Costa et al., 2011), a genre of a song is a categorical variable created by humans to facilitate easy grouping of songs

into their respective style of music. With the advent and boom of internet services, music streaming applications are coming up rapidly and it is no longer feasible to manually tag a song with its genre. Further, there are so many sub-genres that it is difficult to manually assign a list of genres to songs. This is how music genre recognition softwares are aiding such streaming applications to maintain a proper meta data of songs in their databases. MGR further helps categorize songs based on genres, which can be used as a music collection for users to listen to. Additionally, in applications such as recommending music, or generating music, genre is used as an important feature, further emphasizing the importance of correctly recognizing genres.

## 2 Literature Review

It is not a part of the intermediate report and will be completed in the next stage.

## 3 Data

After exploring the internet for different possible open source datasets, we finalize **FMA** dataset as the final dataset for the project. There are other possible datasets related to music available on net, but lack some of the qualities which we need for our problem. Shown below are the possible datasets and their limitations:

### 3.1 Million Songs Dataset

Million songs dataset is one of the largest music datasets available on net, however, it does not provide audio clips which are of utmost importance to us. It asks the user to download music clips from online streaming sites like 7digital, which just adds on to the complexity of the problem. Further, it's dataset does not allow to recognize genre, and mostly focuses on topics like segmentation, automatic tagging, year recognition, lyrics etc.

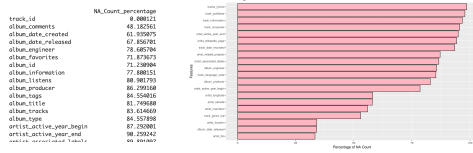


Figure 1: Missing values in Tracks data

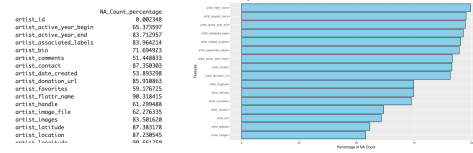


Figure 2: Missing values in Artists data

## 3.2 GTZAN Dataset

It is one of the widely used datasets available for music genre recognition available on net. However, it only has 1000 audio samples for 10 genres i.e. 100 samples per genre. Further, it has lot of faults like mislabeling, repetitions, and distortions (Defferrard et al., 2016). Then, the data was collected in 2001-2002, a time when there was not much electronic music available. Right now, electronic music is one of the most listened genre, so it's inclusion in analysis is of high importance. Lastly, it misses metadata. More information about it's cons can be found in (Sturm, 2013).

## 3.3 Free Music Archive - FMA

The data set used for our problem is FMA dataset. It is a huge dataset of audio samples and meta-data, consisting of 917 GB of licensed audio from 106,574 tracks and 161 genres (Defferrard et al., 2016). However, due to processing power limitations, we will analyze and solve the problem on *small* dataset, which is like GTZAN data, containing 8,000 tracks of 30s, and 8 balanced genres.

dataset	tracks	genres	length	size
small	8,000	8	30	7.4
medium	25,000	16	30	23
large	106,574	161	30	98
full	106,574	161	278	917

Table 1: Subsets of FMA

### 3.3.1 Exploratory data analysis on FMA

In this section, we present why the dataset chosen for the project is a feasible one and explain more about its meta data and correctness.

There are 3 major datasets – *genre.csv*, *artists.csv*, *tracks.csv* – which need to be explored to get a better understanding of the FMA dataset and how it can be used for our project.

We first look at snippets of data files and find out how many missing values are present in different columns. It is an important step to

genre_id	num_tracks	parent	title	top_level
1	11	86931	381	Avant-Garde
2	21	52711	01	International
3	31	17521	01	Blues
4	41	41261	01	Jazz
5	51	41061	01	Classical
6	61	9141	381	Novelty
7	71	2171	281	Comedy
8	81	8681	01	Old-Time / Historic
9	91	19871	01	Country
10	101	138451	01	Pop
11	111	3671	141	Disco
12	121	329231	01	Rock
13	131	7381	01	Easy Listening
14	141	14991	01	Soul-R&B
15	151	344191	01	ET-Soul-R&B
NA_Count_percentage				
genre_id				0.0
num_tracks				0.0
parent				0.0
title				0.0
top_level				0.0

genre_id	num_tracks	parent	title	top_level
genre_id	0.0			
num_tracks	0.0			
parent	0.0			
title	0.0			
top_level	0.0			

Figure 3: Genre data visualization

establish the authenticity of data since a large number of missing values in certain columns can hinder the project modelling phase. Ex- for our project, it is of utmost importance that there should be minimal missing data in columns such as "track\_id", "genre\_id", "artist\_id", "album\_id" etc. for us to be able to go ahead with the project. Although, Figures 1 and 2 have a lot of columns with missing value. It is of no concern to us, because these columns do not convey any meaning towards genre recognition. Having missing values in "artist.bio" does not impact our analysis and project.

On exploring *genre.csv*, we find out that there are no null values in the data 3. This finding validates our decision to go forward with FMA because genre data is of utmost importance. Fig 4 shows how tracks are diversified under genres. Due to image readability, we show track counts

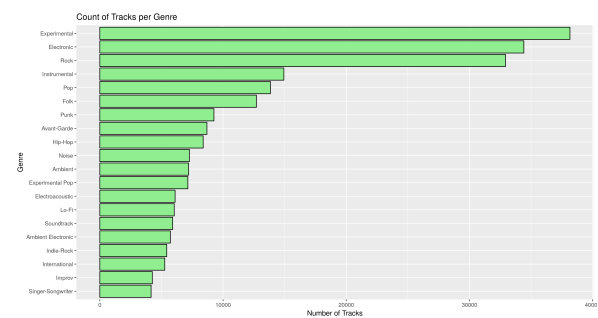


Figure 4: Tracks per genre

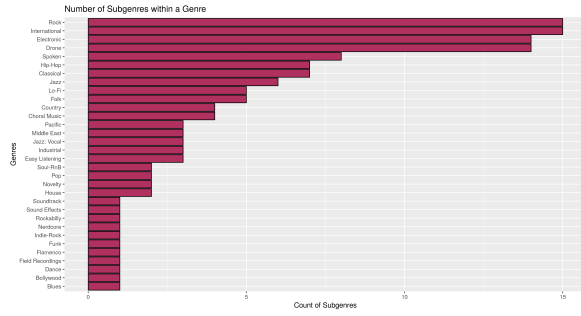


Figure 5: No. of sub-genres of parent genres

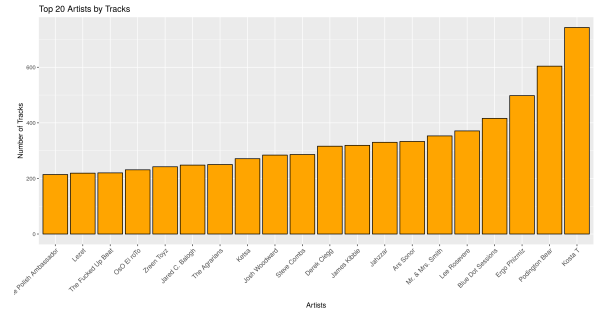


Figure 7: Top 20 Artists

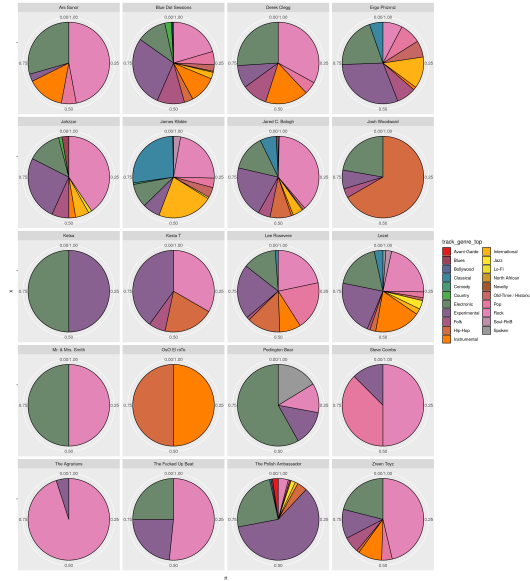


Figure 6: Top Genres of Top Artists

only for top 20 genres. It further shows that the dataset is in sync with current musical trends. As stated above, electronic music is one of the most listened genre in today's world and FMA dataset correctly takes that into account. This inclusion of electronic music is something which lacks in GTZAN dataset.

Fig. 5 shows the count of sub-genres in parent genres and Fig. 6 shows the variety of genre among the tracks of top artists. Both these figures highlight the detailed level-data available for genre recognition tasks. Additionally, Fig. 7 displays the top 20 artists present in the dataset. It can be noticed that the figure validates the fact that there is no biased artist variable in the dataset.

With this thorough analysis of data, it is evident that FMA is a good option for conducting MGR.

## 4 Proposed Approach

Till now, we have discussed our idea of using audio samples to achieve music genre recognition

capability without touching upon how to achieve that. In this section, we will try to go through the solution we are proposing.

Lyrics-based genre recognition achieves sub-par performance and hence, we aim to achieve higher accuracy by using audio features, primarily spectrograms of a track.

Spectrograms are image representation of the spectrum of frequencies in a track as it varies with time. In the final paper, we will explain this concept in much more detail. For now, we want to establish the fact that we aim to harness this image representation of a track and use them as an input for our convolutional neural network.

Convolutional neural networks(CNNs) are considered the paragon of any image processing technique used nowadays because of it's ability to learn image features accurately and therefore, we aim to harness the power of CNNs on our input of spectrograms. The architecture of CNN is yet to discuss, but we will try out different architectures ranging from basic ones to more complex ones.

Our main aim is to not just generate a single genre per song, but generate a continuous genre prediction through the song. This stems from the nature of songs: although they are categorized to some parent genre, it contains elements of a lot of different genres. We believe a continuous prediction of genre for the song will do justice to the input spectrograms (as they provide frequency plots per second). As of now, we believe an LSTM (Long short-term memory) model can be applied for this sequential data processing.

We hope to present the final outcome in the form of a visualized web page or application, where the audio sample is played and the model predicts genre through time. When the full audio is played, the model would be able to display a pie chart or bar plot showing the confidence of each genre recognized in the song.

## Acknowledgments

*PySpark*, *R* and *Python* have been used to run all the exploratory data analysis on this project. The libraries used are *seaborn*, *ggplot2*, *matplotlib* and *pandas*. The document has been made on *ACL2018* template on *Overleaf*.

## References

- Y. M. G. Costa, L. S. Oliveira, A. L. Koerich, and F. Gouyon. 2011. Music genre recognition using spectrograms. In *2011 18th International Conference on Systems, Signals and Image Processing*, pages 1–4.
- Michal Defferrard, Kirell Benzi, Pierre Vandergheynst, and Xavier Bresson. 2016. Fma: A dataset for music analysis.
- Bob Sturm. 2013. The gtzan dataset: Its contents, its faults, their effects on evaluation, and its future use.