Music Genre Classification

Rajeeva Shreedhara Bhat[#1],Rohit B. R.[#2] ,Mamatha K. R.[#3]

[3]*Assistant Professor*
Information Science and Engineering,
B M S College Of Engineering,
Bengaluru, India

Abstract: *A music genre is a conventional category that identifies some pieces of music as belonging to a shared tradition or set of conventions. It is to be distinguished from musical form and musical style. Music can be divided into different genres in many different ways. The popular music genres are Pop, Hip-Hop, Rock, Jazz, Blues, Country and Metal.*

Categorizing music files according to their genre is a challenging task in the area of music information retrieval (MIR). Automatic music genre classification is important to obtain music from a large collection. It finds applications in the real world in various fields like automatic tagging of unknown piece of music (useful for apps like Saavn, Wynk etc.).

Companies nowadays use music classification, either to be able to place recommendations to their customers or simply as a product. Determining music genres is the first step in the process of music recommendation. Most of the current music genre classification techniques use machine learning techniques.

The same principles are applied in Music Analysis also. Machine Learning techniques have proved to be quite successful in extracting trends and patterns from the large pool of data.

I. Introduction

In today's world, an individual's music collection generally contains hundreds of songs, while the professional collection normally contains tens of thousands of music files. Music databases are incessantly gaining reputation in relations to specialized archives and private sound collections. With improvements in internet services and increase in network bandwidth there is also an increase in number of people accessing the music database. Dealing with extremely large music databases is exhausting and time consuming.

Most of the music files are stored according to the song title or the artist name. This may cause trouble in searching for a song related to a specific genre. The popular music genres are Blues, Classical, Country, Disco, Hip-Hop, Jazz, Metal, Pop, Reggae and Rock.

Music has also been divided into Genres and sub genres not only on the basis on music but also on the lyrics as well. This makes music genre classification difficult. Also the definition of music genre has changed over time. For instance, pop songs that were made fifty years ago are different from the pop songs we have today. Fortunately, the progress in music data and its storage has improved considerably over the past few years.

Since manually classifying each track of a large music database according to their genre is a tedious task, Machine Learning Techniques to perform Automatic Music Genre Classification are used.

Rest of this paper is organized as follows. Section II deals with the Literature Review done to write this paper, this section consists of the important takeaways gathered after studying different works in the same field. Section III provides an overview about the Dataset that was used to carry out this research work. Section IV deals with the high level design of the work. Section V deals with the structure of the Convolutional neural network which performs the classification. Section VI contains the experiment results. Section VII consists of the conclusion and future work.

II. Literature Review

Music Genre Classification is an area which has attracted the interest of many researchers. This section will provide details about some of the research work already done in this field. Vishnupriya S and K Meenakshi [1] have proposed a Neural Network Model to perform the classification. Tzanetakis and Cook [2] pioneered their work on music genre classification using machine learning algorithm. They created the GTZAN dataset which is till date considered as a standard for genre classification. Changsheng Xu et al [3] have shown how to use support vector machines (SVM) for this task. Matthew Creme,

Charles Burlin, Raphael Lenain from Stanford University [4] have used 4 different methods to perform the classification. They have used Support Vector Machines, Neural Networks, Decision Trees and K-Nearest Neighbours methods to perform classification. Tao [5] shows the use of restricted Boltzmann machines and arrives to better results than a generic multilayer neural network by generating more data out of the initial dataset, GTZAN.

After carrying out the above mentioned literature survey, Convolutional Neural Network is used to perform classification and the details of the same are explained in the following sections.

III. About the Dataset

GTZAN Genre Collection dataset was used to perform the classification. The dataset has been taken from the popular software framework MARSYAS. Marsyas (Music Analysis, Retrieval and Synthesis for Audio Signals) is an open source software framework for audio processing with specific emphasis on Music Information Retrieval applications. It has been designed and written by George Tzanetakis (gtzan@cs.uvic.ca). Marsyas has been used for a variety of projects in both academia and industry.

Dataset consists of 1000 audio tracks each 30 seconds long. It contains 10 genres(Blues, Classical, Country, Disco, Hip-Hop, Jazz, Metal, Pop, Reggae and Rock), each represented by 100 tracks. The tracks are all 22050Hz Mono 16-bit audio files in .wav format.

Table 1. Distribution of the Dataset

Genre	Number of tracks
Blues	100
Classical	100
Country	100
Disco	100
Hip-Hop	100
Jazz	100
Metal	100
Pop	100
Reggae	100
Rock	100
Total	**1000**

Data Pre-Processing was done in the following manner:

1. Database of the complete collection was created and stored in a .csv file.

2. Feature Vector Extraction is done using the libROSA package in python as shown in figure 1. libROSA is a python package for music and audio analysis which provides the building blocks necessary to create music information retrieval systems.

3. Each audio file is taken and from that, its feature vector is extracted. The extracted feature vector is called MFCC(Mel-Frequency Cepstral Coefficients). The MFCCs as shown in figure 3 encode the timbral properties of the music signal by encoding the rough shape of the log-power spectrum on the Mel frequency scale. A Zero-Crossings graph is plotted as shown in figure 4 for each audio track. This graph visualizes the number of times the signal crosses zero level.

4. As shown in the following figure, Fourier Transforms are applied on the music signal. A Frequency Spectrum is thus obtained. Mel Scale Filtering is applied on the frequency spectrum to obtain a Mel Frequency Spectrum. A log() function is applied on this Mel Frequency Spectrum which is transformed into Cepstral Coefficients on applying discrete cosine transforms. Finally, the Feature Vector is obtained by finding out the derivatives of the Cepstral Coefficients.

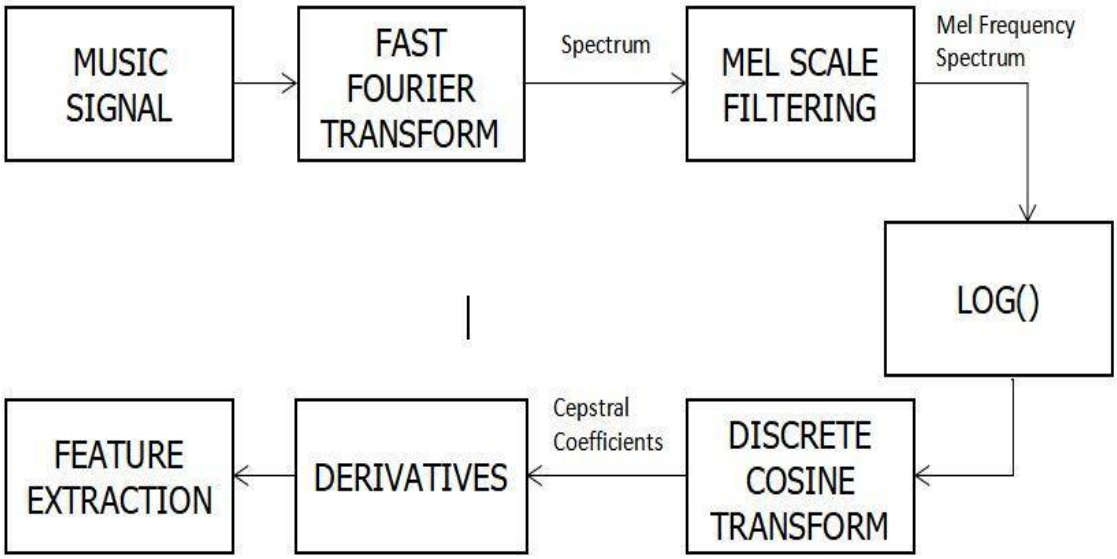

Fig 1: Extraction of Feature Vector

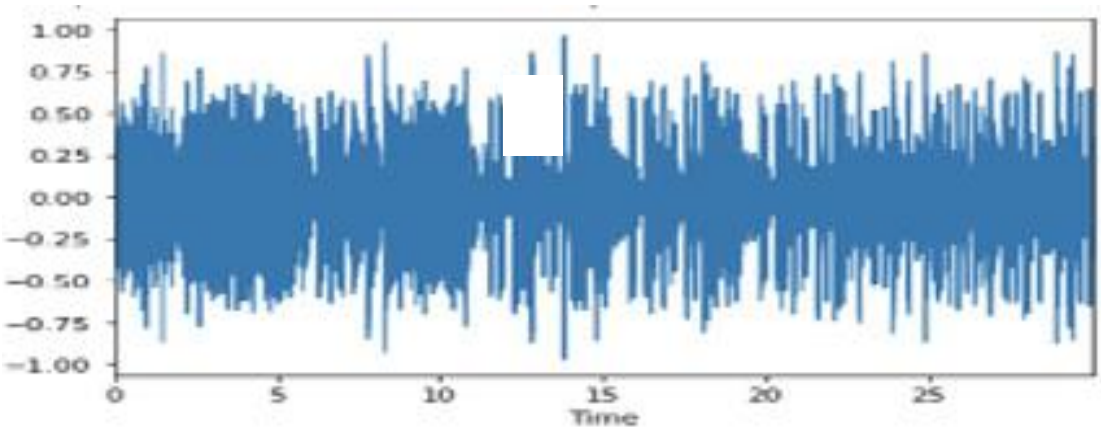

Fig 2: Sample Frequency Spectrum

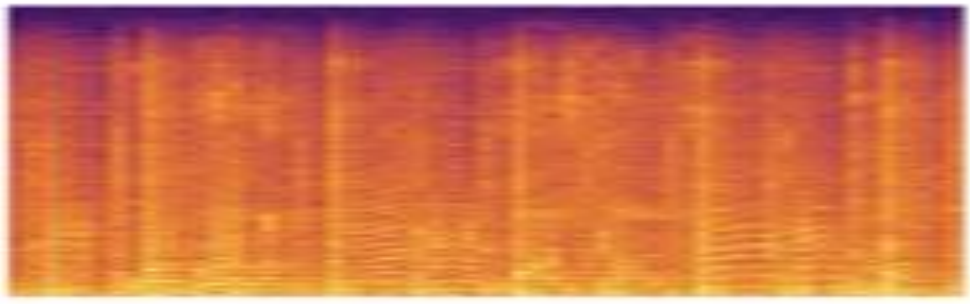

Fig 3:Sample Mel Frequency Spectrum

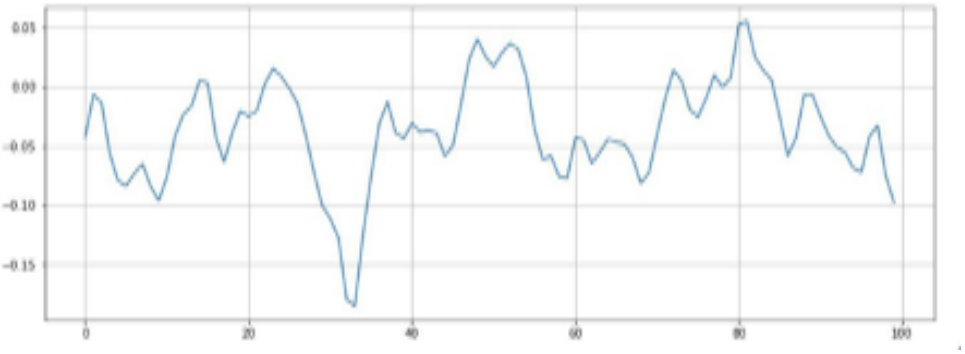

Fig 4: Zero Crossings

IV. High Level Design

1. The dataset is split into two parts, Training data and Test data.
2. Each track from the train dataset is pre-processed and a feature vector is extracted for the same. A Feature Vector Database is generated from the extracted feature vectors.
3. The Neural Network model is trained using the obtained feature vector database.
4. Each track from the test dataset is also pre-processed and a feature vector is extracted for the same.
5. The trained Neural Network model operates on the feature vector obtained at the end of step 4 to perform classification on test data.
6. Finally, output is genre of the music track.

High Level Design of the system is shown in figure 5:

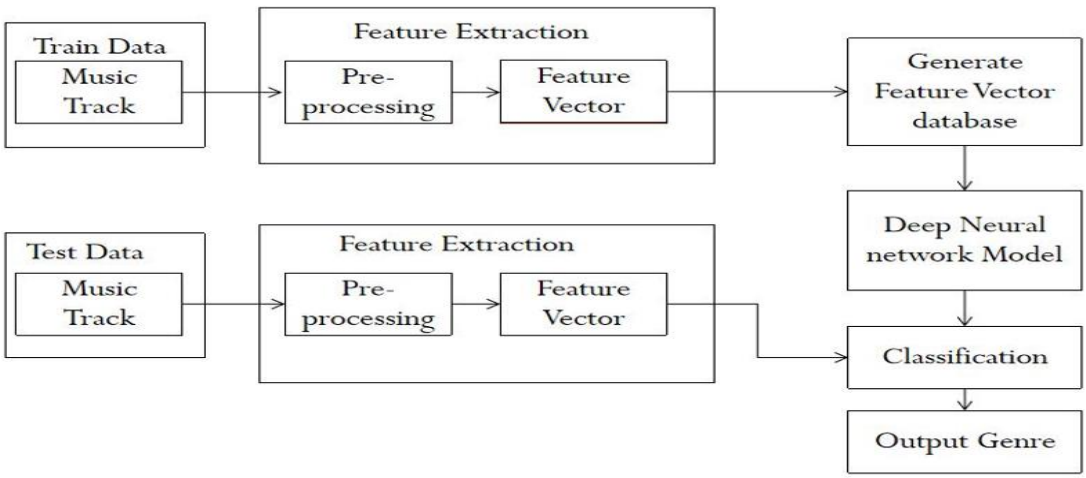

Fig 5: High Level Design

V. Convolutional Neural Network

CNN is a Deep Learning algorithm which can take an input image as input, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. A CNN has various layers such as Convolutional layers, ReLU layers, Pooling layers and a fully connected layer as shown in figure 6. CNN is widely used for image classification because it does automatic feature extraction using convolution.

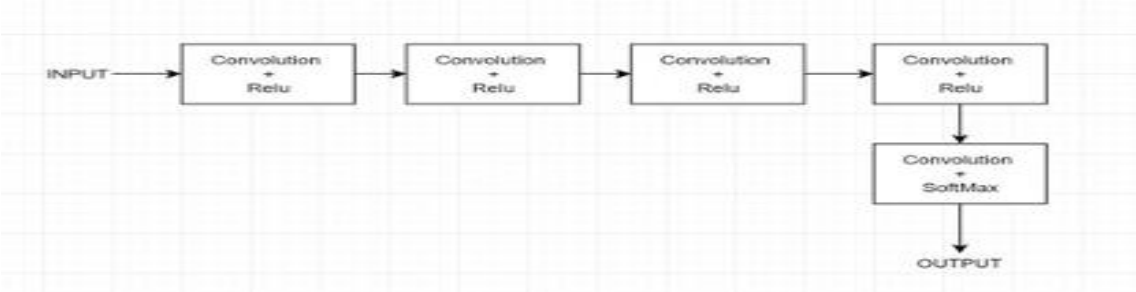

Fig 6: Architecture of CNN used in model

VI. Results

To train the Convolutional Neural Network, an 80% - 20% splitting strategy was used for training and testing respectively.

The accuracy of the model is calculated using:

$$\text{ACCURACY} = \frac{\text{No of songs correctly classified}}{\text{Total no of songs}} * 100$$

For the GTZAN dataset, the model we used achieved a training accuracy of about 98% and validation accuracy of 73% as shown in figure 7 and figure 8.

Python was the language used to develop the model. A number of packages such as keras, numpy, pandas were used to build the model. Experiment is done on the google colab platform. Tensorflow package is used for deep-learning.

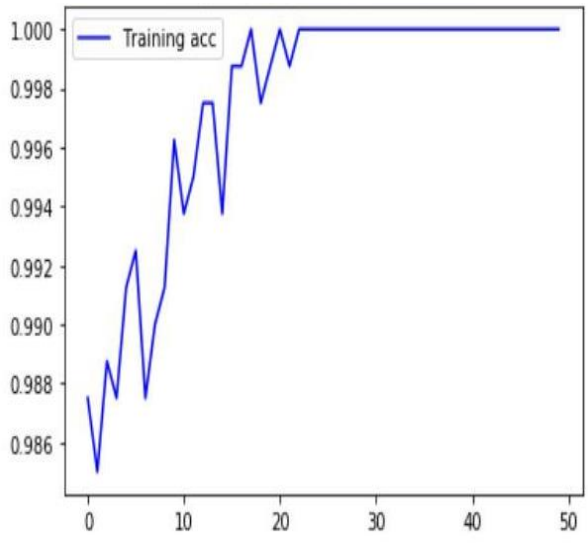

Fig 7: Accuracy

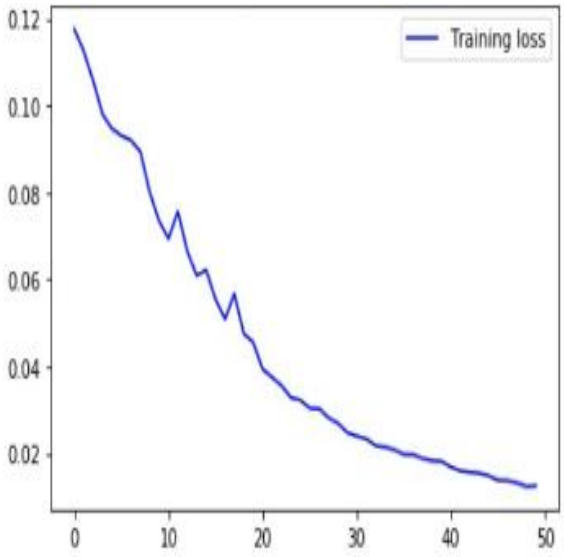

Fig 8: Loss

VII. Conclusion and Future Work

This research work provides the details of an application which performs Music Genre Classification using Machine Learning techniques. The application uses a Convolutional Neural Network model to perform the classification. A Mel Spectrum of each track from the GTZAN dataset is obtained. This is done by using the libROSA package of python. A piece of software is implemented which performs classification of huge database of songs into their respective genres.

The extension of this work would be to consider bigger data sets and also tracks in different formats(mp3, au etc). Also, with time the style represented by each genre will continue to change. So the objective for the future will be to stay updated with the change in styles of genres and extending our software to work on these updated styles. This work can also be extended to work as a music recommendation system depending on the mood of the person.

REFERNECES

[1] Vishnupriya S, and K.Meenakshi, "*Automatic Music Genre Classification using Convolution Neural Network*", IEEE Conference 2018.

[2] G.Tzanetakis and P. Cook. "*Musical genre classification of audio signals, Speech and Audio Processing*", IEEE Transactions, July 2002.

[3] Chandsheng Xu, Mc Maddage, Xi Shao, Fang Cao, and Qi Tan, "*Musical genre classification using support vector machines*", IEEE Proceedings of International Conference of Acoustics, Speech, and Signal Processing, Vol. 5, pp. V-429-32, 2003.

[4] Matthew Creme, Charles Burlin, Raphael Lenain, "*Music Genre Classification* ", Stanford University, December 15, 2016.

[5] T. Feng, "*Deep learning for music genre classification*", 2014.

[6] Daniel Grzywczak, Grzegorz Gwardys, "*Deep image features in music information retrieval*", 10th international conference, AMT 2014, Warsaw, Poland, August 11-14, 2014 proceedings, pp 187-199.

[7] Muhammad Asim Ali, Zain Ahmed Siddqui, "*Automatic Music Genres Classification using Machine Learning*", International Journal of Advanced Computer Science and Applications, Vol 8, No 8, 2017.

[8] Hareesh Bahuleyan, "*Music Genre Classification using Machine Learning Techniques*", University of Waterloo, ON, Canada, 2018

[9] Sam Clark, Danny Park, Adrien Guerard, "*Music Genre Classification using Machine Learning Techniques*", 2012.

[10] Eve Zheng, Melody Moh, Teng-Sheng Moh, "*Music Genre Classification: A N-gram based Musicological Approach*", 7th International Advance Computing Conference, 672-677, 2017.

[11] http://marsyas.info/downloads/datasets.html