

# **CHAPTER 1**

## **INTRODUCTION**

People go crazy on hearing songs now-a-days. The songs they hear can differ according to their mood. So the genre of music such as jazz, classical, rock, hiphop, country,...etc plays an important role. People usually have the habit of sorting songs by different genres in their playlist. Indian people speak different languages, follow various cultures according to their geographical area. Hence, Indian music industry is a huge and vast area having numerous songs corresponding to their geographical area, finding their genre by human capabilities is quite a difficult task. Music is not just a collection of words, it conveys many useful information, reality of life. Moreover, it is one of the online platforms which connects thousands of minds together. Modern world has brought many improvements in the field of music. For instance, the classical songs that were composed sixty years ago are different from the classical songs that are composed today.

The extraction of acoustic features that are the best estimators in the classification of genres. This method follows a single or multi-label classification or in some stages, regression stage required for building that system. A music genre is a category of conventional which derives from the frames of music, feature extraction depends on signal processing and relevant features using time or frequency domain audio representation. This modified features are given as input. To know the most relevant feature to perform the task, deep neural network is approached. The word genre generally represents or relates the style or category of music, art and literature. The result of the music genre is obtained by predicting the music genre of audio signals in the form of a wav file.

The aim of this project is to identify the type of music genre. Millions of people like to hear songs in their day to day life. Most people of all ages listen to music in their leisure time. While analyzing, the features are identified and it is visualized for the clear review of the process. Here, a convolutional neural network algorithm is used to analyze the data set where the accuracy level is high. By implementing these methods, it gives a high accuracy rate while testing and training a lot of data.

## **CHAPTER 2**

### **LITERATURE SURVEY**

#### **2.1 Music genre recognition with deep neural networks**

Albert Jimenez, Ferran Jos 'e in 2018 published "Music genre recognition with deep neural networks"[1]. The methods they used are deep learning and convolutional neural networks, multiframe and transfer learning. The main advantage of this is the multiframe approach. Audio files are placed in a confusion matrix to find the mean of all frames. So the classification of genre is more accurate. They have used a two layer convolutional network with mel-spectrogram technique and also raw audio signals as their input features.

For training and fine-tuning the dataset optimizers like adaptive learning rate ADAM and SGD are used. Comparing the results ADAM is chosen to use. On the other hand, when it uses a large amount of data, it takes more time for processing, which is a drawback of this paper.

#### **2.2 Survey on Transfer Learning**

Sinno Jialin Pan and Qiang Yang Fellow in 2010 published "Survey on Transfer Learning"[2]. They have used transfer learning, machine learning and data mining techniques. Transfer learning is a technique in which people learn to apply the knowledge they have gained so far in every problem they face in their day to day life. The transfer learning technique is closely related to multi- task learning methods, in which multiple tasks are carried out simultaneously. Machine learning techniques always try to learn from scratch. On the other hand, transfer learning usually transfer knowledge from the training set to target data.

The advantage of this method is across multiple dispersals of the source and the task is improved with different features. It is of three categories; transductive transfer learning, inductive transfer learning and unsupervised transfer learning. There should be a clear difference between source and target domain. The drawback of this is identification and rectification of negative transfer is difficult. transfer learning technique is used in small applications with less number of datas in the field of image and text classification. As a future work it will be used for network analysis, video classification.

### **2.3 Transfer learning by supervised pre-training for audio-based music classification**

Aaron van den Oord, Sander Dieleman, Benjamin Schrauwen in 2014 had published the paper "Transfer learning by supervised pre-training for audio-based music Classification"[3]. They have used transfer learning and neural networks. They have used the Million song dataset(MSD). The transfer learning method has been explored by inserting features of specified audio and labels from various datasets into an allocated space with linear transformations.

That transfer learning works good especially when the main task is tag prediction, i.e. when the target task and source are closely related. The main disadvantage is that the transfer learning method does the given job specifically good only when the source task is tag prediction, i.e. both source and target task are nearly connected.

### **2.4 Musical Genre Classification of Audio Signals**

George Tzanetakis, Student Member, IEEE, and Perry Cook, Member in 2002 published "Musical Genre Classification of Audio Signals"[4]. They used audio classification and feature extraction technology. The different features used in this method are rhythmic content, pitch content and texture. This system mostly targets the signal classification having news, videos in various categories like sound, speech and music. A new method for non-speech song classification which is used to recognise sound effects and instrumental sounds separately. Using the dataset, features such as amplitude, frequency and bandwidth are found using the statistics of mean, autocorrelation and variance.

The main advantage is that to control incorrect higher rate of accuracy which occurs when similar features from the same file in which frames from those files could not be divided between training and testing data. Separate feature sets are designed in order to establish better results for rhythm and harmony. The disadvantage of this paper is that there always arises a discrimination between speech, sound and music. Melody extraction and singer voice extraction is a hard problem in this paper.