# Music Genre Classification using Convolution Neural Network

Deep Saran Masanam
Department of Electrical Engineering,
Rochester Institute of Technology
Rochester, NY, 14623, USA
dm7079@g.rit.edu

***Abstract*— Music genre classification has become important in this rapid growing world. In order to have a good playlist based on the users interest many companies are doing extensive research in this field so to provide a better overall experience to the user. The main objective of this paper is to design a classifier which can detect the genre of the music clip provided as the test subject. Many techniques have been developed to tackle this problem namely conventional as well as advanced deep learning techniques. Feature extraction is an important part of the design and Mel Frequency Cepstral Coefficients is used as a feature vector. deep learning approach i.e. convolution neural network has been used in the design. A comparison between support vector machines and Convolution neural networks is done and tabulated at the end of the papers which shows that neural nets have dominance in accuracy and better prediction of the test data.**

***Index Terms*— Convolution neural networks, Deep learning, Mel Frequency Cepstral Coefficients (MFCC), Music Genre classification**

## I. INTRODUCTION

In this rapidly advancing world, there is an immense growth in technology. it can be either automation or entertainment, both are equally important. Nowadays everything is being implemented with machine learning/ deep learning algorithms and making things easy to use. Music genre classification is a similar process which improvises the playlist of the user based on his requirements i.e. it develops the playlist based on the most repeated genre and classifies them. However, there are many challenges linked to this task, most common issue is copyrights and another challenging task is handling the large datasets. But these issues cannot be tackled on a small scale.

Music genre classification can be divided into two major parts: Feature extraction and classification. In the first part various features are to be extracted from the dataset such that we can get good accuracy as the classifier depends on these features. There are many papers published on classifiers, i.e. which classifiers gives out best results for different datasets or the dataset which we have used in this paper which is further discussed in the literature review section of this paper.

Neural networks are best classifier which can deal with huge amount of training datasets when compared to the conventional datasets as stated in different papers. Mel-frequency cepstral coefficients (MFCC) these are used as considered as the best feature vectors for audio type datasets.

The paper is further structured as follows. Section II as literature review which shows an insight about the existing and related works done in this field. Section III focuses on the method which has been implemented. Section IV gives an insight on the expected results and Section V is the conclusion.

## II. LITERATURE REVIEW

Many researchers have done phenomenal work related to genre classification which can give amazing insights on the work and provide a foundation for further research. GTZAN and Million data song dataset have been used in [1], librosa package in python has been used as it is specific to audio analysis these extracted featured are named as MFCC's. Many papers like [1],[3],[5],[6],[7] have considered Mel Frequency Cepstral Coefficients (MFCC) and have shown in their results that by using MFCC the achieved high accuracies.

Genre classification by lyrics is implemented in [2] i.e. Natural language processing, where he combined the results obtained from classification of lyrics and artwork of the image to get at most results. In [2] the author implemented both Bag-of-word and TFIDF as natural language processing techniques for lyrical data and compared then with various machine learning algorithms. For image classification i.e. artwork of the album they used Convolution neural networks. Which is a unique way to combine the results and propose a combined classification result. Where as in [3] they have done a comparative study showcasing different machine learning algorithms on music file dataset (GTZAN). Which shows that SVM has the highest accuracy with MFCC as feature vector. By observing the results from different papers, it appears that MFCC are powerful feature vectors which increase the accuracy of the classifiers.

Long-Short Term Memory (LSTM) has been used as a core component in [4]. It is comparative study between neural network/ Multilayer perceptron (MLP), Support vector machine and K-nearest neighbor. Based on the results they proposed a method with the combination of MLP and LSTM which have given highest accuracy among all the algorithms implemented. [4] They also stated that for MLP to perform at

its peak it requires 10 or more layers. Similarly, in [5] they proposed a new method which has a combination of both LSTM and support vector machines to get high accuracy. They have individually trained both the classifiers and combined them using posterior probabilities obtained from the classifiers using sum rule which gave a combined accuracy of 89%.

In [6] various features like zero crossing rate, spectral centroid, spectral contrast, spectral bandwidth, spectral roll-off and Mel-frequency Cepstral Coefficients-MFCC are used with various machine learning and deep learning techniques like K-nearest neighbor, Support vector machine, Naive Bayes, Decision Tree and Random Forest. [6] is an extensive comparison between the features on different classifiers. Out all the features mentioned above MFCC have again proven to give highest accuracy. Whereas in [7] they proposed a model which is based on the main four features roll-off, excitation source, normalized autocorrelation peal strength and MFCC called Gaussian Mixture Model (GMM) which is said to be very useful in forming online music libraries with good accuracy.

## III. PROPOSED METHOD

### A. Music Dataset

The proposed system is used to classify the GTZAN dataset into different genres, it consists of 1000 songs as shown in table 1.

Table 1: Distribution of Genre in 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 |

GTZAN and Million song dataset (MSD) are very popular datasets available but MSD is large dataset around 280Gb due to which the processing time is very high so GTZAN is used for the proposed method.

### B. System Design

The data is split into 80-20 ratio i.e. 80% for training the model and 20% for testing the data. Now we have a dataset of 10 genre each of 80 tracks for training our model. Now the data is sent to the next phase i.e. feature extraction where the feature vector is Mel Frequency Cepstral Coefficient.

MFCC are derived as follows:

1. Take Fourier transforms of signal
2. Map the power spectral densities onto Mel scale
3. Take log of these powers at Mel frequencies
4. Discrete cosine transforms of theses Mel log powers
5. The obtained amplitudes of resulting spectrum are MFCC

After which the features are fed to the Convolution neural nets where the model is trained accordingly.

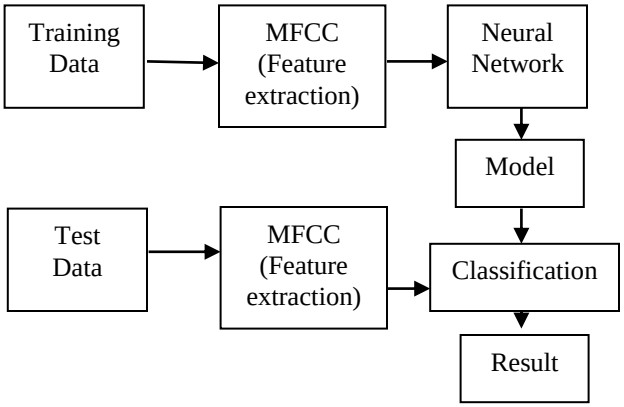


Fig 1: Flow chart of the process

### C. Support Vector Machines

SVM's in short are supervised learning algorithms which are used for regression and classification of data in addition to this it uses a kernel trick for mapping into higher dimensional features. It constructs a hyper plane i.e. a separation between data to linearly classify them into different classes.

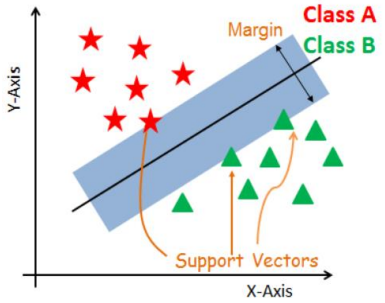


Fig 2: Support vector machines structure

As shown in the above picture the hyperplanes radius is determined by the support vectors.

### D. Convolution Neural Network

Convolutional Neural Network (CNN), it includes one or extra convolutional layers and then proceeded through one or more absolutely connected layers as in a preferred multilayer neural community. Every neuron gets inputs from the characteristic vectors after which they are dot product with the weights which are then surpassed directly to the subsequent layers.

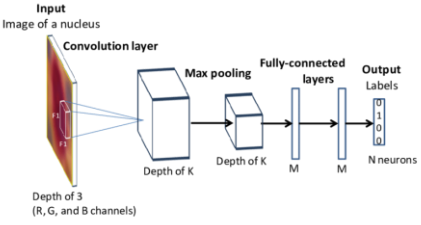


Fig 3: CNN block diagram

The three main layer sin CNN is Convolutional layer, pooling layer and fully connected layer. Maxpooling partitions the input signals into a set of nonoverlapping matrix and for each sub-region output is maximum value

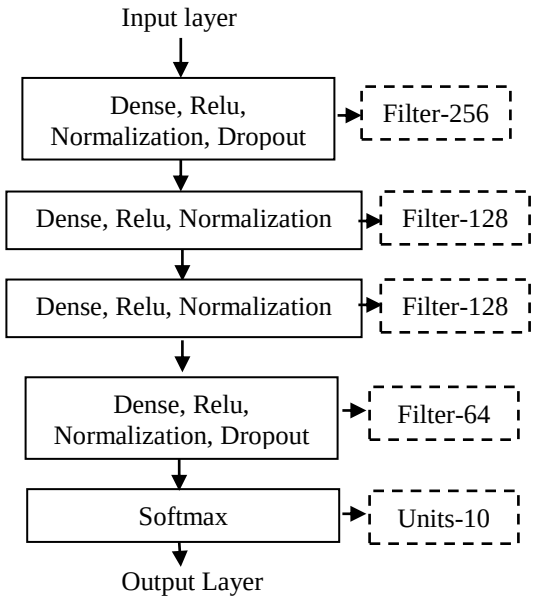


Fig 3: Flow chart of the used Neural network

Dense layer: A linear operation in which every input is attached to each output with the aid of a weight. usually observed via a non-linear activation characteristic.
Pooling layer: Replace each patch in the input with a single output, which is the maximum of the input patch.
Normalization layer: Scale the input so that the output has near to a zero mean and unit standard deviation, to allow for faster and more resilient training.
Relu Layer: Rectified linear unit is an activation function which is defined as the positive part of the function i.e. also known as ramp function.
Softmax layer: It is also known as softargmax or normalized function. It is a smooth approximation of the max function. It is often used a final layer in neural networks.

## IV. EXPECTED RESULTS

The Model is expected to classify with good accuracy and provide good test classification results.

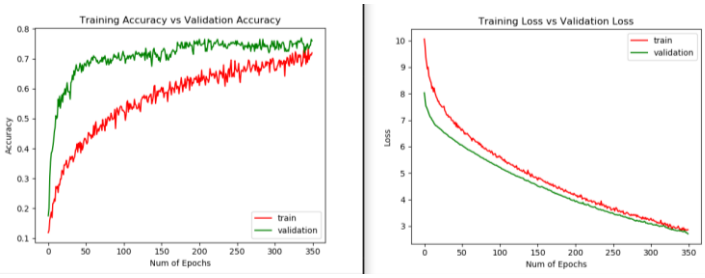


Fig 4: Graphs of validation and training

The above graph shows both the two graphs which shows the training and validation of accuracies, loss.

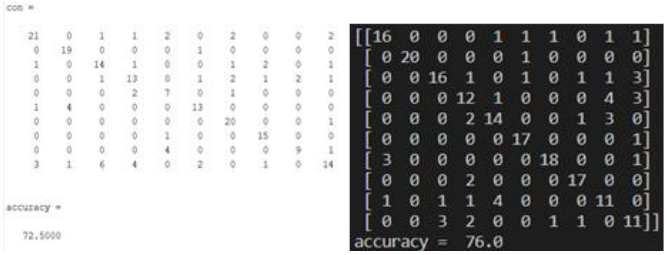


Fig 5: Confusion matrices of both SVM and CNN's

Left shows the confusion matrix of SVM and the right part shows the confusion matrix with accuracy of CNN.

## V. CONCLUSION

As per the classification done on the GTZAN data on both the classifiers, it is observed that accuracies are comparable, but CNN has high accuracy i.e. SVM has an accuracy of 72.5% and CNN has an accuracy of 76% as shown in the confusion matrix in the above section.

## VI. REFERENCES


[1] K. VISHNUPRIYA S, "Automatic Music Genre Classification using Convolution Neural Network," in *International Conference on Computer Communication and Informatics*, Coimbatore, INDIA, 2018.

[2] A. R. D. R. Akshi Kumar, "Genre Classification using Feature Extraction and Deep Learning," in *10TH INTERNATIONAL CONFERENCE ON KNOWLEDGE AND SYSTEMS ENGINEERING (KSE)*, 2018.

[3] S. B. J. C. K. G. S. Pradeep Kumar D, "A Comparative Study of Classifiers for Music Genre Classification based on Feature Extractors," in *IEEE*, 2016.

[4] R. O. A. a. N. M. C. T. Rene Josiah M. Quinto, "Jazz Music Sub-Genre," in *IEEE Region 10 Conference*, Malaysia, 2017.

[5] R. S. N. K. a. K. P. Prasenjeet Fulzele, "A Hybrid Model For Music Genre Classification Using LSTM And SVM," in *Eleventh International Conference on Contemporary Computing*, Noida, India, 2018.

[6] H. B. Ç. M. E. İ. B. Ö. a. N. A. Ahmet Elbir, "Music Genre Classification and Recommendation by Using Machine Learning Techniques," in *IEEE*, 2018.

[7] C. K. a. R. Kumar, "Study and Analysis of Feature Based Automatic Music Genre Classification Using Gaussian Mixture Model," in *Proceedings ofthe International Conference on Inventive Computing and Informatics*, 2017.