Machine Learning and Noise Reduction Techniques for Music Genre Classification

Omkar chavan[1], Nikhil Kharade[2], Amol chaudhari[3], Nikhil Bhalke[4], Pravin Nimbalkar[5]

[1-4]Student, Department of Computer Engineering, JSPM'S Imperial college of Engineering and research Pune, Maharashtra, India
[5]Professor, Department of Computer Engineering, JSPM'S Imperial college of Engineering and research Pune, Maharashtra, India

---***---

Abstract - *Classification of music files in the field of music information retrieval based on the genre is one of the challenging task to perform manually. In this study we have applied machine learning and deep learning techniques to classify music genre automatically, in our approach we have applied deep learning based convolutional neural networks to predict genre label of music with the use of spectrogram. We have proposed the classification by reducing the noise using audacity noise reduction in that we have used Fourier analysis to improve audio frequency spectrum. Similarly, we have also applied supervised machine learning models using features and compared their performances.*

Key Words: Convolutional neural network, spectrogram, Fast fourier transform (noise reduction)

1. INTRODUCTION

With the growth of online content for different commercial enterprises music is something which is more popular and should be available on tip of categories which have different set of audience , it is not easy to sort these music which in bunch to make it easy for individuals to listen.

So, one way to categorize and organize the music is in the music genre. Genre is a way of identifying the music on some characteristic like rhythmic structure, harmonic content, Emotional content instrumentation and other factors. Based on genre online music streaming services like Saavan, Spotify and Wynk music are in demanding position for such automated classification.

In this approach we have explored the application of machine learning algorithms to identify and classify the genre of a given audio file primarily we perform the noise reduction on given audio file by using noise reduction techniques. Then the audio file with less noise is provided as input to first major model Convolutional neural networks, which is trained end to end on MTL spectrogram of the audio signal.

2. RELATED WORK

It is observed that the last couple of year investigation about music information is not discovered, we observe from the following.

(I).In Hareesh Bahuleyan(2018),**Music Genre Classification using Machine Learning Techniques [1]** it gives the idea about Relegation utilizing Machine Learning techniques, the work conducted gives an approach to relegate music automatically by providing tags to the musical compositions present in the user's library. It explores both Neural Network and traditional method of utilizing Machine Learning algorithms and to achieve their goal. The first approach uses Convolutional Neural Network which is trained end to culminate utilizing the features of Spectrograms (images) of the audio signal. The second approach uses sundry Machine Learning algorithms like Logistic Regression, Random forest etc, where it utilizes hand-crafted features from time domain and frequency domain of the audio signal. The manually extracted features like Mel-Frequency Cepstral Coefficients (MFCC), Chroma Features, Spectral Centroid etc are habituated to relegate the music into its genres utilizing ML algorithms like Logistic Regression, Desultory Forest, Gradient Boosting (XGB), Support Vector Machines (SVM). By comparing the two approaches discretely they came to a conclusion that VGG-16 CNN model gave highest precision. By constructing ensemble classifier of VGG-16 CNN and XGB the optimised model with 0.894 precision was achieved.

(2) In Srishti Sharma, Prasenjeet Fulzele, Indu Sreedevi, **Novel Hybrid Model for Music Genre Classification based on Support Vector Machine.** This paper suggested that we use the Support Vector Machine Algorithm for Music Genre Classification. They have considered the acoustic properties for the genre classification. They achieved an accuracy of 87%.

(3). In Haojun Li, Siqi Xue, Jialun Zhang **Combining CNN and Classical Algorithms for Music Genre Classification.** In this paper they proposed of using various classification algorithms like SVM, Logistic Regression and Random Forest algorithms. What they proposed is that we can use CNN's output as features for the other Algorithms. They found that at times the accuracy of Classical Algorithms exceeded the accuracy of Neural Networks.

(4). Iyad Lahsen Cherif, Abdesselem Kortebi, **On using eXtreme Gradient Boosting (XGBoost) Machine Learning algorithm for Home Network Traffic Classification.** In this paper mostly classification is performed on traffic and its flow. XG boost classification algorithm is best applied and has quite better accuracy in classifying as compared to other supervised machine learning techniques, hence we got the brief idea about eXtreme Gradient Boosting for classification in terms of music and audio with good accuracy.

(5).Jitendra Kumar Jaiswal, Rita Samikannu, **Application of Random Forest Algorithm on Feature Subset Selection and Classification and Regression**. In this study we observed that random forest classification handles the classification smoothly with more higher number of features, as music genre is type of classification where every detail spectral features are to be considered to classify, hence random forest best fits for classifying and representing the differing features of processed audio and related dataset features.

3. PROPOSED SYSTEM

We have split the working of the system into 3 different phases, namely: Noise Reduction Phase, Spectogram generation phase and Deep learning Convolution neural networks this phases will be operated sequentially.

3.1 Dataset

It is the basic requirement for cnn to train which gives complete training to model to perform classification. We have used free available gtzn for MIR. Following table shows the details.

Table -1: Music files used for training

Sr.no	Genre name	Count
1.	Blues	100
2.	Rock	100
3.	Pop	100
4.	Reggae	100
5.	Metal	100
6.	Jazz	100
7.	Disco	100
8.	Country	100
9.	Classical	100
10.	Hiphop	100

3.2 Noise Reduction Phase

The Noise Reduction algorithm uses Fourier analysis. It finds the spectrum of pristine tones that make up the background noise in the susurration segment that you culled - that's called the "frequency spectrum" of the sound. That forms a dactylogram of the static background noise in your sound file. When you reduce noise from the sound holistically, the algorithm finds the frequency spectrum of each short segment of sound. Any pristine tones that aren't adequately louder than their average levels in the dactylogram are reduced in volume. That way, (verbally express) a guitar note or an overtone of the singer's voice are preserved, but his and other steady noises can be minimized. The general technique is called spectral noise gating.

The first pass of noise reduction is done over just noise. For each windowed sample of the sound, we take a Expeditious Fourier Transform (FFT) [6] utilizing a Hann window and then statistics, including the mean potency, are tabulated for each frequency band.

During the noise reduction phase, those statistics and the Sensitivity setting determine a threshold for each frequency band. We commence by setting a gain control for each frequency band such that if the sound has exceeded the threshold, the gain is set to 0 dB, otherwise the gain is set lower to the Noise Reduction slider setting (e.g. -18 dB), to suppress the noise.

The gain controls are then applied to the involute FFT of the signal and the inverse FFT applied, followed by another Hann window. The output signal is then pieced together utilizing overlap/integrate of one-fourth the window size. Following is real time plot of fourier transform:

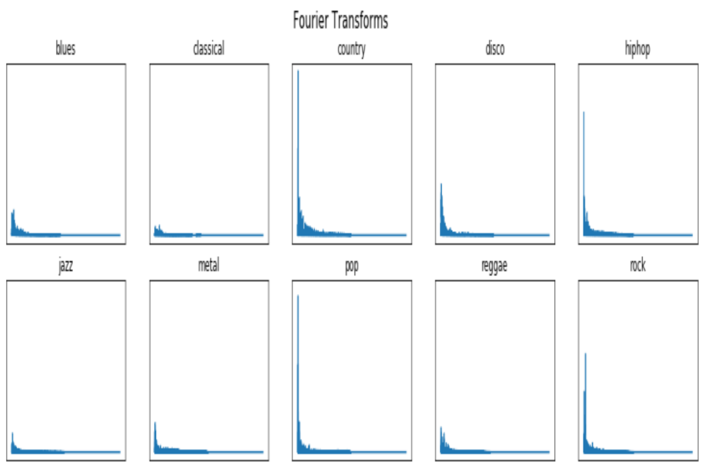


Fig-1: Real time Fourier transforms of genres

3.3 Spectrogram Generation

A spectrogram is a 2D representation of a signal, having time on the x-axis and frequency on the y-axis. In this study, each audio signal was converted into a MEL spectrogram (having MEL frequency bins on the y-axis). The parameters used to engender the puissance spectrogram utilizing STFT. Following are real time spectrogram and plots for genres.

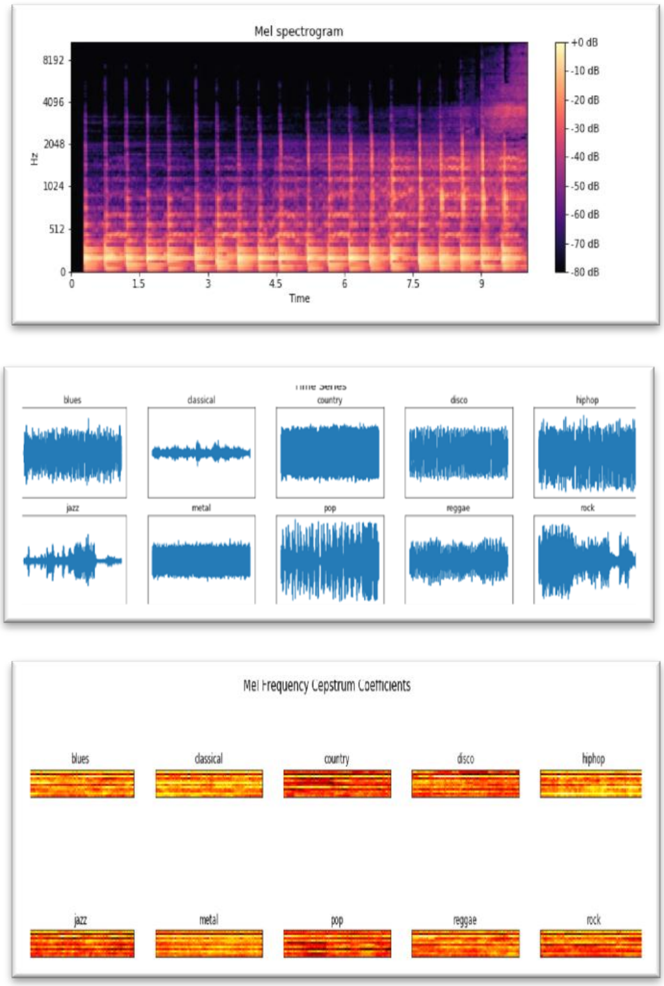


Fig -2: Real Time Spectogram and plots of genres

3.4 Deep Learning Model Phase

With deep learning algorithms, we can achieve the task of music genre relegation without hand-crafted features. Convolutional neural networks (CNNs) prove to be a great cull for relegating images. The 3-channel (R-G-B) matrix of an image is given to a CNN which then trains itself on those images. In this study, the sound wave can be represented as a spectrogram, which can be treated as an image.

3.4.1 Convolutional Neural Networks (CNN)

[3]We can optically discern that there subsist some characteristic patterns in the spectrograms of the audio signals belonging to different classes. Hence spectrograms can be considered as 'images' and can be given as input to a CNN.

Operations of CNN:

Each block in a CNN consists of the following operations:

Convolution: This step involves a matrix filter (verbally express 3x3 size) that is moved over the input image which is of dimension image width x image height. The filter is first placed on the image matrix and then we compute an element-sagacious multiplication between the filter and the overlapping portion of the image, followed by a summation to give a feature value.

Pooling: This method is utilized to reduce the dimension of the feature map obtained from the convolution step. By max pooling with 2x2 window size, we only retain the element with the maximum value among the 4 elements of the feature map that are covered in this window. We move this window across the feature map with a predefined stride.

Non-linear Activation: The convolution operation is linear and in order to make the neural network more potent, we require to introduce some non-linearity. For this purport, we can apply an activation function such as Rectifier Linear Unit (ReLU) on each element of the feature map. The model consists of 3 convolutional blocks (conv base), followed by a flatten layer which converts a 2D matrix to a 1D array, which is then followed by a plenarily connected layer, which outputs the probability that a given image belongs to each of the possible classes.

Max Pooling layer: The feature maps engendered by the convolution layer are forwarded to the pooling layer. We employ 1-max pooling function on a feature map to reduce it to a single most ascendant feature. Pooling on B × A feature maps results in B × A features that will be joined to compose a feature vector inputted to the final SoftMax layer. This pooling strategy offers a unique advantage. That is, albeit the dimensionality of the feature maps varies depending on the length of audio events and the width of the filters, the pooled feature vectors have the same size of A × B.

4. Architecture

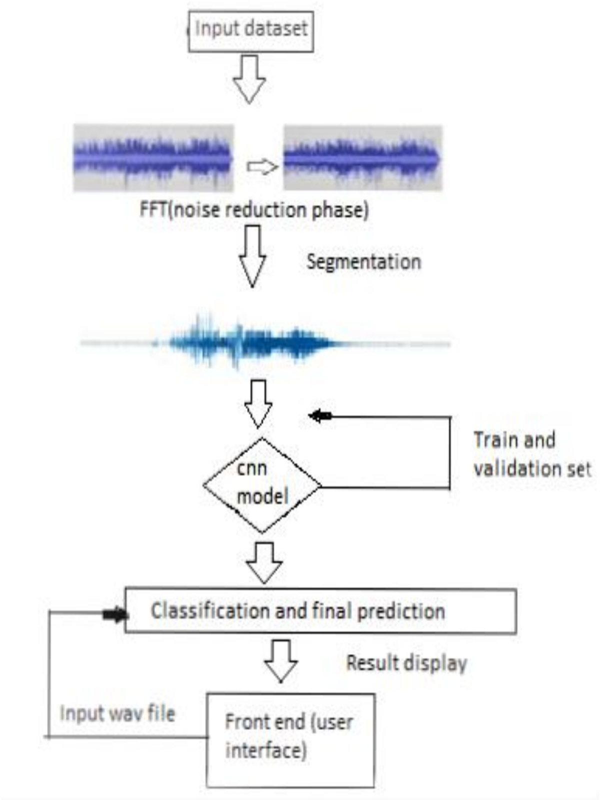
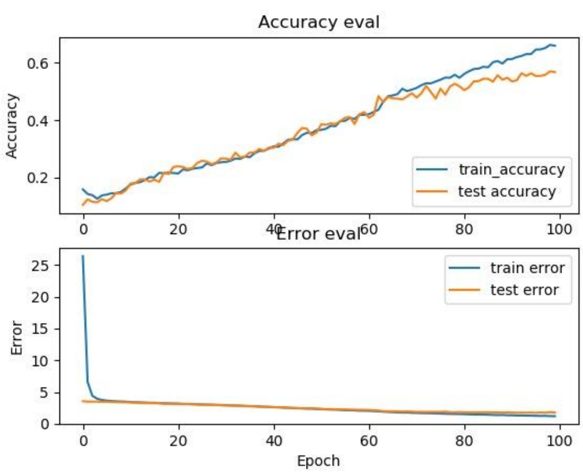


Fig-3: Architecture of proposed system

5. Result and Discussion

To evaluate the model we have used **accuracy** as metrics, it refers to a percentage of correctly classified test samples.

Our real time Evaluation graphs show how validation accuracy is improved over training.



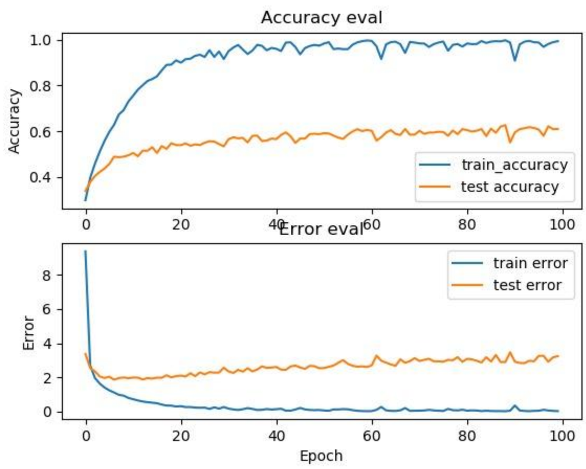


Chart-1: Evaluation for accuracy and error with respect to epochs

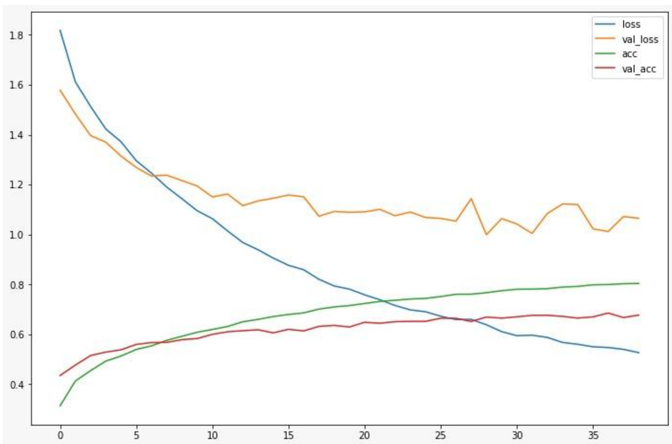
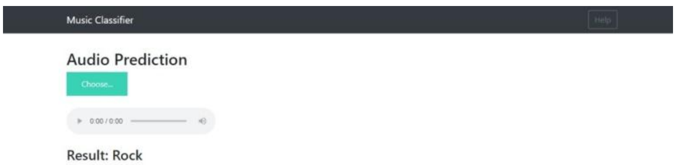


Chart-2: Validation loss and validation accuracy

Hence we have performed classification with the **convolutional neural network**, which has given us accuracy of **about 88%** following are real time images for two genres, we have used wav file as input as of now on which our training is based.



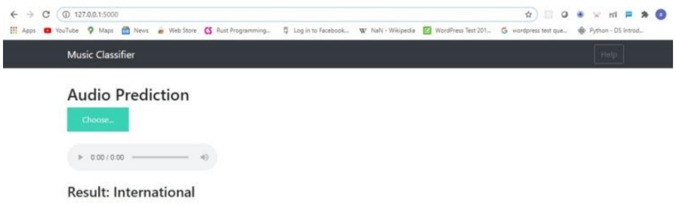


Fig -4: Real time results and outcome

6. CONCLUSIONS

In this paper we have proposed is that the integration of noise reduction module in the system will increase the probability of getting higher accuracy compared to just using a model.

The inclusion of noise reduction on the dataset will definitely improve the accuracy of the model. Also, this inclusion can be used on raw music files.

REFERENCES

1. Hareesh Bahuleyan."Music Genre Classification using Machine learning Techniques". University Of Waterloo.ON, Canada.

2. SrishtiSharma, Prasenjeet Fulzele, Indu Sreedevi."Novel Hybrid Model for Music Genre Classification based on Support Vector Machine".Delhi Technology University, IEEE.

3. Haojun Li,Siqi Xue, Jialun Zhang ."Combining CNN and Classical Algorithms for Music Genre Classification ``.Department of CS, Stanford University, ICME, Stanford University.

4. Iyad Lahsen Cherif, Abdesselem Kortebi."On using eXtreme Gradient Boosting (XGBoost) Machine Learning algorithm for Home Network Traffic Classification ``.Orange Labs Network, 2019 IEEE

5. Jitendra Kumar Jaiswal, Rita Samikannu. "Application of Random Forest Algorithm on Feature Subset Selection and Classification and Regression. VIT University, IEEE.

6. D. P. Gadekar, N. P. Sable, A. H. Raut, "Exploring Data Security Scheme into Cloud Using Encryption Algorithms" International Journal of Recent Technology and Engineering (IJRTE), Published By:Blue Eyes Intelligence Engineering & Sciences Publication, ISSN: 2277-3878, Volume-8 Issue-2, July2019, DOI: 10.35940/ijrte.B2504.078219

7. Ahmet Elbir, Hamza Osman ølhan, Gorkem Serbes,Nizamettin AydÕn."Short Time Fourier Transform Based Music Genre Classification ``.Yildiz Technical University, Istanbul, Turkey, 2018 IEEE.

8. Sable Nilesh Popat*, Y. P. Singh," Efficient Research on the Relationship Standard Mining Calculations in Data Mining" in Journal of Advances in Science and Technology | Science & Technology, Vol. 14, Issue No. 2, September-2017, ISSN 2230-9659.