Musical Genre Classification using Convolutional Neural Networks

R. Thiruvengatanadhan

Abstract*: *Music has likewise been separated into Genres and sub sorts on the premise on music. To show that, we contrast the outcomes acquired and a Convolutional Neural Network (CNN). Experiments were conducted on Marsyas databases with distinct characteristics for genre classification. The proposed CNN results in better accuracy in music genre classification.

Keywords*: *Music genre Classification, Convolutional Neural Network (CNN).

I. INTRODUCTION

Musical classes perception has driven a few specialists to recommend the meaning of another sort characterization plot only for the reasons for music data recovery [1]. Music sort names are valuable classifications to compose and characterize melodies, collections, and craftsmen into more extensive gatherings that share comparable melodic qualities [2]. Music sorts have been broadly utilized for music order, from physical music stores to real time features. Programmed music sort order along these lines is a generally investigated point [3]. This makes arrangement harder. To cause things more to confuse the meaning of music sort may have very much changed after some time [4]. For instance, rock songs that we have today.

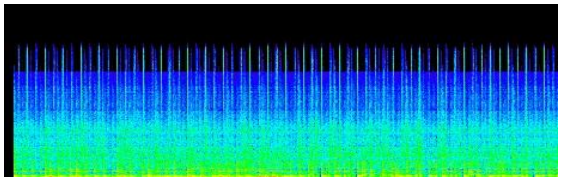

Fig. 1.Spectrogram of Jazz music signal

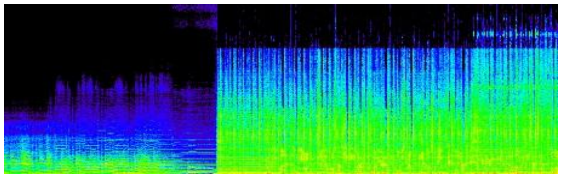

Fig. 2.Spectrogram of Pop music signal

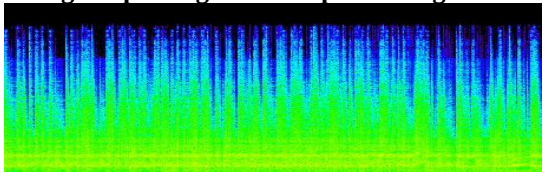

Fig. 3.Spectrogram of Metal music signal

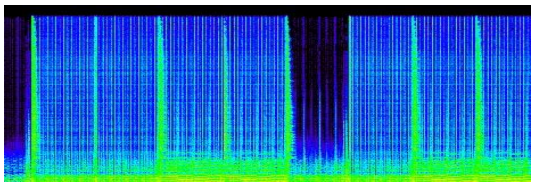

Fig. 4.Spectrogram of Disco music signal

Above Figures show the various music genre and visual spectrogram of music signals. The spectrograms obtained from audio signal [5]. Fig. 5 shows the proposed work using CNN.

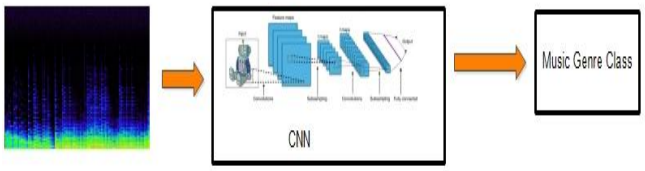

Fig. 5.Block Diagram of Proposed Work

II. CONVOLUTIONAL NEURAL NETWORK (CNN)

CNNs to extract musical pattern features in audio. However, their experimental results showed that the proposed models did not generalise very well to unseen testing data [6].

CNNs have been effectively utilized for different music characterization errands, for example, music labeling [7], kind grouping [8], and client thing inert component forecast for proposal. CNNs expect highlights that are in various degrees of order and can be removed by convolutional parts. The various leveled highlights are found out to accomplish a given assignment during administered preparing. Fig. 6 shows the architecture of CNN.

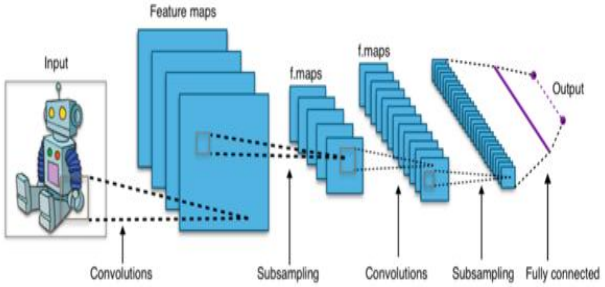

Fig. 6.CNN architecture

Revised Manuscript Received on November 01, 2020.
***** Correspondence Author
R.Thiruvengatanadhan*, Department of Computer Science and Engineering, Faculty of Engineering and Technology, Annamalai University, Annamalainagar, TamilNadu, India. Email: thiruvengatanadhan01@gmail.com

III. EXPERIMENTAL RESULTS

A. Dataset Collection

The data is collected from Marsyas databases with distinct characteristics for genre classification. In our work dataset consists of 480 audio tracks each 30 seconds long. It contains 4 genres. The tracks are all 22050Hz Mono 16-bit audio files in .wav format.

Table- I: Classes and Number of Samples in the Marsyas Database

Classes	Samples
Jazz	160
Pop	80
Metal	100
Disco	140

B. Convolutional Neural Network (CNN)

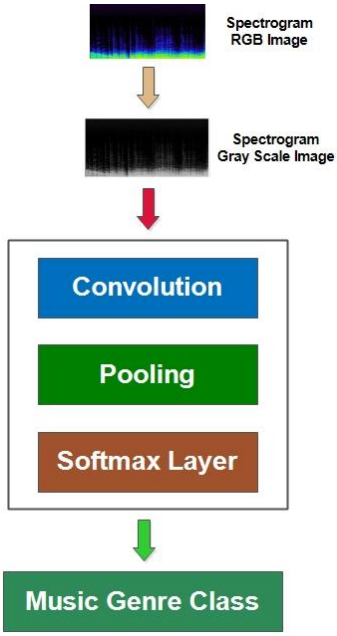

Fig. 7.Block Diagram of Proposed System in CNN

Fig. 7 shows the proposed music genre classification system using CNN. The input image 128 x 128 x 3 pixels. The dataset of 400 images CNN algorithm to train. CNN calculation need involvement with engineering plan, and need to investigate continuously in the commonsense application, so as to acquire the most reasonable for specific application design of CNN. Based on gray image as the input of 256 × 800, To classify whether a pixel in a time-frequency plane belongs to a tempo contour, patches from the key regions. Fig. 8 shows the patches extracted.

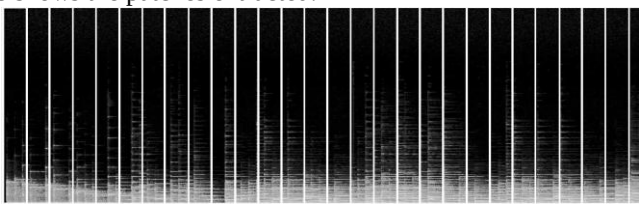

Fig. 8.CNN PATCHES EXTRACTED

Downsampling function to reduce dimensionality in the pooling layer. The CNN structure is show in the Table II. In this work tensorflow is used to implementation of CNN and Librosa is used for audio library in python.

Table- II: Structure of CNN

Output size	Layer type
1×256×800	Input
1×256×16	Patch
64 feature maps @ 256×16	Convolutional
64 feature maps @ 128×8	Max Pooling
64 feature maps @ 128×8	Convolutional
64 feature maps @ 64×4	Max Pooling
500	Fullyconnected
4	Output

To show the confusion matrix of CNN model is plotted in Fig. 9.

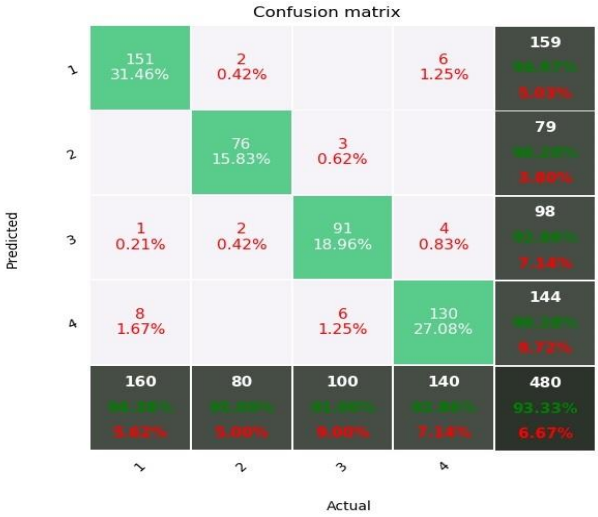

Fig. 9.Confusion Matrix for Proposed CNN Model

IV. CONCLUSION

In this paper, the CNN the music signal is converted into spectrogram image as input. The CNN shows the good result in musical genre classification scheme is very effective and the accuracy rate is 93.33%.

REFERENCES

1. M E ElAlami, S M K Tobar, S M Kh and Eman A Esmaeil., "Design of Simulation System based on Digital Processing of Sound to Identify the Genres of Arabic Music", International Journal of Computer Applications 175(26):36-39, October 2020.
2. Sturm, B. L. "A survey of evaluation in music genre recognition" In International Workshop on Adaptive Multimedia Retrieval, pages 29–66, 2012.
3. Serwach, M., & Stasiak, B., "GA-based parameterization and feature selection for automatic music genre recognition" In Proceedings of 2016 17th International Conference Computational Problems of Electrical Engineering, CPEE 2016.
4. Dijk, L. Van., Radboud Universiteit Nijmegen Bachelor thesis "Information Science Finding musical genre similarity using machine learning techniques", 1–25, 2014.
5. Y. M. G. Costa, L. S. Oliveira, A. L. Koerich, and F. Gouyon." Music genre 607 recognition using spectrograms." In International Conference on Systems, 608 Signals and Image Processing, pages 161-154, 2011.
6. Samarth Sarin and Siba Panda. "A Convolution Neural Network for Classification of Indian Faces", International Journal of Computer Applications 177(24):7-11, December 2019.

7. Keunwoo Choi, George Fazekas, and Mark Sandler, "Automatic tagging using deep convolutional neural net-works," arXiv preprint arXiv:1606.00298, 2016.
8. Keunwoo Choi, George Fazekas, Mark Sandler, and Jeonghee Kim, "Auralisation of deep convolutional neu-ral networks: Listening to learned features," in Proceed-ings of the 16th International Society forMusic Informa-tion Retrieval Conference, ISMIR, 2015, pp. 26–30.