Music Genre Classification Using Deep Learning

Yugam Parashar
Dept. of CSED
National Institute of Technology
Calicut
Kerala,India
yugam_m220256cs@nitc.ac.in

Subhranil Barua
Dept. of CSED
National Institute of Technology
Calicut
Kerala,India
subhranil_m220294cs@nitc.ac.in

Swapnil Baviskar
Dept. of CSED
National Institute of Technology
Calicut
Kerala,India
swapnil_m220281cs@nitc.ac.in

***Abstract*— The ability to rapidly classify music genres has become increasingly crucial in contemporary society. Therefore, a popular research area is emerged that uses deep learning methods to classify music into different genres. High accuracy can be achieved by deep learning techniques in music genre classification without feature engineering ad fine tuning parameters in traditional machine learning techniques. GT-ZAN music dataset is used for training and testing**

Keywords—deep Learning, GT-ZAN, machine learning, music genre classification

I. INTRODUCTION

In today's world numerous industries have embraced digitalization of their products, including music. Although music is considered abstract, it can also be treated as data. Major music streaming platforms such as Spotify and iTunes utilize algorithms that analyze the similarity between songs to offer recommendations to their users. People are highly enthusiastic about music and tend to listen to songs that suit their mood. The genre of music, such as classical, rock, hip-hop, country, etc., plays a crucial role in this regard. People often organize their playlists by sorting songs into different genres. The Indian music industry is vast and diverse, with numerous songs corresponding to different geographic areas and cultural backgrounds, making it challenging for humans to identify the genre of every song accurately. Music is not just a collection of words; it also conveys valuable information and reflects the realities of life.

II. LITERATURE SURVEY

Albert Jimenez and Ferran Jos ´e published a paper titled "Music genre recognition with deep neural networks" [1]. They utilized deep learning and convolutional neural networks, employing multiframe and transfer learning techniques. The multiframe approach is the primary advantage of their method, where audio files are arranged in a confusion matrix to determine the mean of all frames, leading to more accurate genre classification. Their input features were mel-spectrogram technique and raw audio signals, with a two-layer convolutional network. For training and fine-tuning the dataset, they used optimizers such as adaptive learning rate ADAM and SGD, with ADAM being the preferred choice. However, a drawback of this paper is that it takes longer to process a large amount of data.

Sinno Jialin Pan and Qiang Yang, both Fellows, published a paper titled "Survey on Transfer Learning" [2], in which they utilized techniques from transfer learning, machine learning, and data mining. Transfer learning is a method that enables individuals to apply the knowledge they have acquired in solving various problems in their daily lives. This technique is closely related to multi-task learning, which involves carrying out several tasks simultaneously. Machine learning techniques generally attempt to learn from scratch, while transfer learning typically transfers knowledge from the training set to target data.

The main advantage of transfer learning is that it improves the task's performance across multiple source domains and different features. It is categorized into three types: transductive transfer learning, inductive transfer learning, and unsupervised transfer learning. However, the main drawback is that identifying and correcting negative transfers is challenging. Transfer learning techniques are commonly used in small-scale applications with a limited amount of data, particularly in image and text classification. In the future, these techniques may be used for network analysis and video classification.

George Tzanetakis, a student member of IEEE, and Perry Cook, a member, published a paper titled "Musical Genre Classification of Audio Signals" [3]. The paper utilized audio classification and feature extraction technology, with a focus on signal classification for news, videos, and music in various categories such as sound, speech, and music. The method employed various features, including rhythmic content, pitch content, and texture, and extracted features such as amplitude, frequency, and bandwidth using statistics like mean, autocorrelation, and variance from the dataset. To prevent incorrect classification due to similar features from the same file, separate feature sets were designed for rhythm and harmony. However, a disadvantage of the paper was that it often faced discrimination between speech, sound, and music, and extracting melody and singer voice was a challenging problem.

George Tzanetakis, Student Member, IEEE, and Aaron van den Oord, Sander Dieleman, and Benjamin Schrauwen published a paper titled "Transfer learning by supervised pre-training for audio-based music classification"[4]. Their approach involved using transfer learning and neural networks on the Million Song Dataset (MSD). Specifically, they inserted features and labels from different datasets into a designated space through linear transformations. This method proved to be effective in tag prediction tasks, where the source and target tasks were closely related. However, a disadvantage of this approach is that it only performs well when the source task is also tag prediction and closely connected to the target task.

Yu-Huei Cheng, Pang-Ching Chang, Che-Nan Kuo, published a paper titled "Convolutional Neural Networks Approach for Music Genre Classification" [5], in which they stated that a tool for classifying music genres can save time for individuals involved in this task. They proposed convolutional neural network achieves an accuracy of 83.3% in classifying music genres, which can aid future music genre classification efforts. Moving forward, they aimed to enhance the accuracy of there model and integrate it with streaming media and web crawlers to make it more comprehensive. This will facilitate music beginners and specific musicians in reducing their workload and improving efficiency.

Lakshman Kumar Puppala, Siva Sankar Reddy Muvva, Sudarshan Reddy Chinige, Dr. P.Selvi Rajendran, published a paper titled "A Novel Music Genre Classification Using Convolutional Neural Network" [6], they designed a music genre classification system based on a CNN program. The feature vectors of the system are obtained using Mel-Spectrum, and the built-in librosa package in Python language helps in extracting features, thereby enabling the system to provide good boundaries for organization training. The training accuracy is shown to be 97%, and the testing accuracy is 74% for Mel. Therefore, this approach is entirely user-friendly, as users can input their own custom songs and receive output as classified genres. This method shows promise for organizing a vast database of songs into specific genres.

III. IMPLEMENTATION DETAILS

Data preprocessing

GTZAN data is arranged as folder of genre with name as genre and audio files of duration 30 seconds each.
Files are further split to increase input data.
Audio files are converted to MFCCs.
MFCC refers to Mel Feature Cepstral Coefficients that capture timbral aspect of sound. They are good approximation of our human auditory system. There are 13 to 40 MFCC coefficients and are calculated for each frame.
Librosa[7] open source python library is used to extract MFCCs from audio files and map genre to index.
The processed data is saved in json format.

Modelling

Tensorflow[8] is a open source python library for high performance numerical computation. It simplifies Neural Network implementation. Each deep learning model has learning rate of 0.0001.

Following models are implemented :

- Basic neural network
- Neural network with dropouts
- Convolutional Neural network

Basic neural network: Basic neural network with three hidden layers of 512, 256 and 64 neurons respectively with Rectified Linear Unit(ReLU) activation function.

Neural network with dropout: Neural network with 30% dropout at each layer to address the issue of overfitting with three hidden layers of 512, 256 and 64 neurons respectively with Rectified Linear Unit(ReLU) activation function.

Convolutional Neural network : Convolutional Neural network with three hidden layers each with 32 kernels and grid size of 3 *3 and Rectified Linear Unit(ReLU) activation function. First hidden layer with 3,3 stride. Second hidden layer with 2,2 stride. Third hidden layer with 2,2 stride.

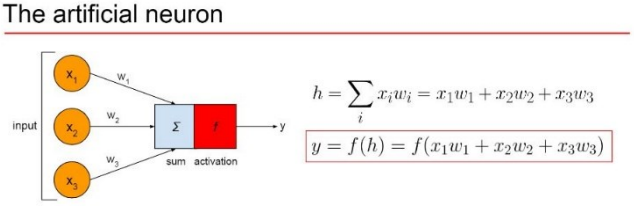

Fig. Basic Neural network

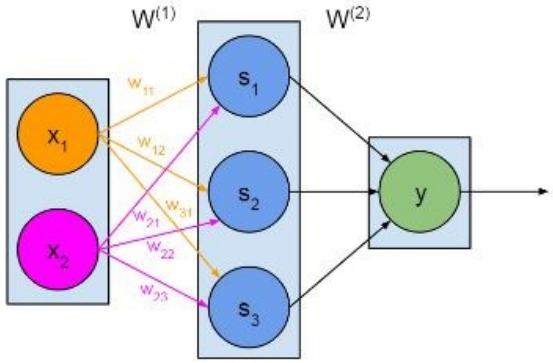

Fig. Neural network with weight

IV. CODE IMPLEMENTATION

Code is available on Github repository:
https://github.com/yug-am/Music-Genre-Classfication

V. RESULTS

Out of above mentioned deep learning models, Convolutional Neural network performed best.

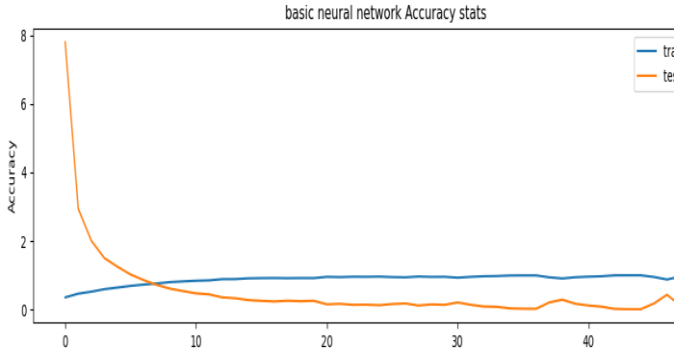

Fig 1. basic Neural Network accuracy plot

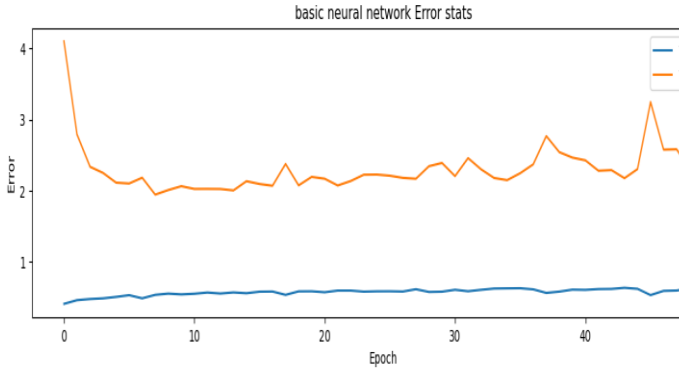

Fig 2. basic Neural Network error plot

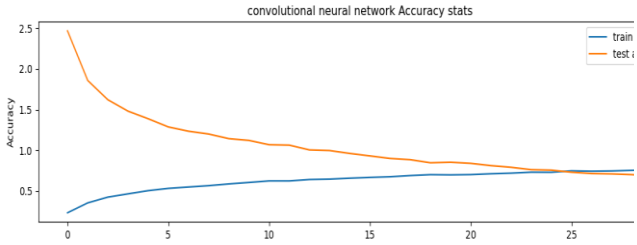

Fig 3. Convolutional Neural Network accuracy plot

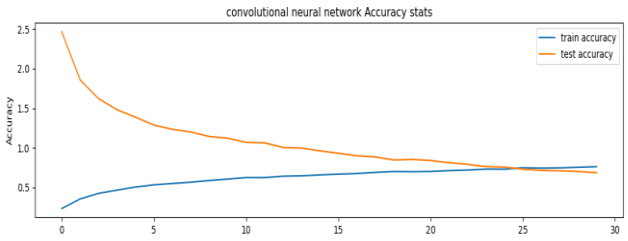

Fig 4. Convolutional Neural Network error plot

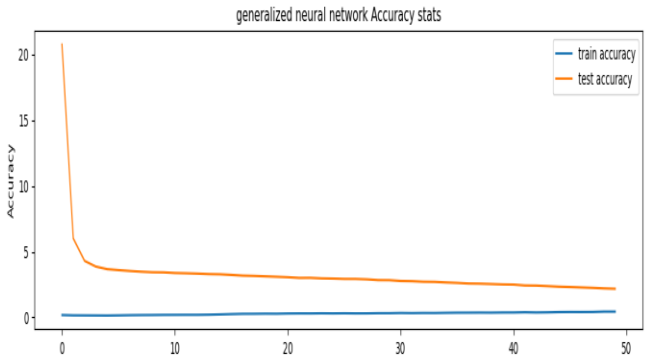

Fig 5. Generalized Neural Network Accuracy Plot

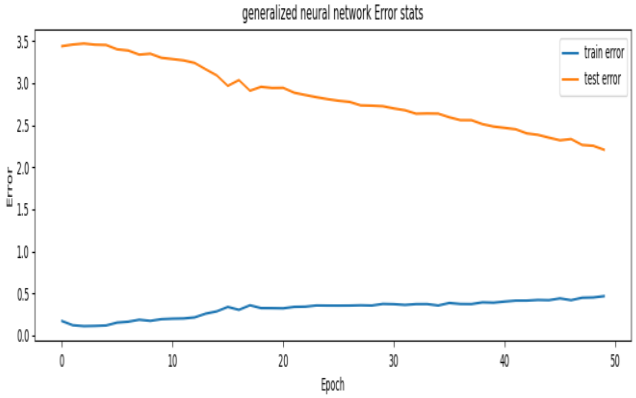

Fig 6. Generalized Neural Network Error Plot

REFERENCES

[1] Albert Jimenez, Ferran Jos´e, “Music genre recognition with deep neural networks”, Universitat Politècnica de Catalunya,2018.

[2] Sinno Jialin Pan and Qiang Yang Fellow, “A Survey on Transfer Learning”, 2010.

[3] Perry Cook, “Musical Genre Classification of Audio Signals”, 2002.

[4] Aaron van den Oord, Sander Dieleman, Benjamin Schrauwen, "Transfer learning by supervised pre-training for audio-based music Classification", 2014.

[5] Yu-Huei Cheng, Pang-Ching Chang, Che-Nan Kuo, "Convolutional Neural Networks Approach for Music Genre Classification," in International Symposium on Computer, Consumer and Control (IS3C),2020.

[6] Lakshman Kumar Puppala, Siva Sankar Reddy Muvva, Sudarshan Reddy Chinige, Dr. P.Selvi Rajendran, "A Novel Music Genre Classification Using Convolutional Neural Network", in 6th International Conference on Communication and Electronics Systems (ICCES-2021), 2021

[7] Librosa librosa python library. Available at: https://librosa.org/doc/latest/index.html (Accessed: March 28, 2023).

[8] Tensorflow PyPI. Available at: https://pypi.org/project/tensorflow/ (Accessed: March 28, 2023).