International Journal of Future Generation Communication and Networking
Vol.14, No. 1, (2021), pp. 4425 - 4430

Identification of Music Genre using Convolutional Neural Network

Agrawal Shreyash#1, S. P. Dhanure#2, P. P. Rathod#3, Challawar Ashay#4, Gadekar Pritesh#5

[1,2,3,4,5,]*Dept.of E&TC Engg., Smt.Kashibai Navale College of Engineering,Pune,Savitribai Phule Pune University, Pune*

[1]shreyash.agrawal.77198@gmail.com
[2]sudhir.dhanure_skncoe@sinhgad.edu
[3]pravinkumar.rathod_skncoe@sinhgad.edu
[4]ashaychallawar@gmail.com
[5]priteshgadekar26@gmail.com

***Abstract*—**
This paper proposes the implementation of a Convolutional Neural Network (CNN) to classify music into its respective genre. The digital entertainment industry is booming right now and people are mostly listening to music online these days. Music plays a key role in our lives. The quantity of music being released on internet platforms is huge. But to manually classify music files is a hectic task for human beings. There is also a good chance of error in case of classification done by humans. The increase amount of work in this field recently has brought a great demand for automatic music genre classification. This paper presents a deep learning approach using Convolutional Neural Network (CNN). CNN model is trained using GTZAN dataset with ten musical genres and spectrogram of each audio file. The model essentially uses a neural specification to classify the music file into 10 different genres. CNN has good ability to work with various musical patterns. CNN has been mostly used in recent approaches and it is more efficient than standard machine learning approaches.

Keywords— *Music genre classification, convolutional neural network, spectrogram, GTZAN, deep learning.*

I. INTRODUCTION

Music performs an essential role in our lives. Music unites like-minded people and is a link between communities. Communities are usually identified by the types of songs that they have composed or may have heard. Different communities and groups listen different sorts of music. One main feature that separates one quite music from another is that the genre of the music.

As the number of songs keeps on increasing, people find it relatively hard to manage the songs of their taste. Since taking note of music online has become very convenient for people, because of the increase of online music streaming services like Spotify, iTunes, etc. users expect the music recommendation by the service. To make this possible, we need to check people's listening options and determine the genre they are listening to. This is the best way. Due to the rapid growth of the digital show business, automatic classification of music genres has acquired significant prominence in recent years. One way to effectively classify the song is based on genre. Classification of genre are often very valuable to elucidate some interesting problems like creating song references, tracking down related songs and music labelling.

II. LITERATURE SURVEY

Snigdha Chillara, Kavitha A. S., Shweta A. Neginhal, Shreya Haldia, Vidyullatha K. S. in [1] did a survey in music genre classification using machine learning algorithms, where comparison of different music genre classification techniques with machine and deep learning algorithms is given. Spectrogram-based models and Feature based models are used to find out the best model for

ISSN: 2233-7853 IJFGCN

classification. CNN model was determined to be the best model over others like RNN, ANN, and LR. Subset of Free Music Archive [FMA] database is used. Librosa, a Python library is used for feature extraction purpose. Hareesh Bahuleyan in [2] proposed the work which gives an approach to classify music automatically, Comparison of the performance of two classes of models is done. CNN model is trained end-to-end using spectrogram and the second approach utilizes various ML algorithms like Logistic Regression, Random Forest, etc. VGG-16 CNN model gave highest accuracy of 89%. In this research 'Audio set' dataset is used which consists of mp4 files. The mp4 files are converted into the desired wav format. Tom LH Li, Antoni B. Chan, Andy HW. Chun in [3] made an effort to understand the main features which actually contribute to build the optimal model for Music Genre Classification. A methodology to automatically extract musical patterns features from audio music is been done. In this CNN has the strong capacity to capture informative features from the varying musical patterns. Due to this the approach is made to CNN, where the musical data have similar characteristics to image data. The dataset considered here is GTZAN which consists of 10 genres of 100 audio clips each. The musical patterns are evaluated using WEKA tool. Bryan Lansdown, Dr. Shan He in [4] gave comparison of machine learning algorithms in their suitedness to the task of music genre classification. Classification was based upon 54 features extracted from each song of the dataset, most of which pertained to spectral information. The average genre scores across the standard dataset were as follows: Electronic:60%, Folk: 73%, Hip-Hop: 74%, Pop: 41%, Rock: 73%. R. Thiruvengatanadhan in [5] proposed a technique of Music Genre Classification using GMM (Gaussian mixture model), a new technique that uses support vector machines to classify songs is described. Parametric or non-parametric methods are used to model the distribution of feature vectors. The music data is collected from music channels using a TV tuner card. A total dataset of 100 different songs is recorded, which is sampled at 22 kHz and encoded by 16-bit. In this work fixed length frames with duration of 20 ms and 50 percentages overlap (10 ms) are used. GMM method has good performance in musical genre classification scheme and is very effective and the accuracy rate is 94%.

III. PROPOSED SYSTEM

The proposed system helps classify music under the genres of the GTZAN dataset. Convolutional Neural Network (CNN) is a deep learning model which delivers great performance with images and basically saves sizeable number of efforts and time.

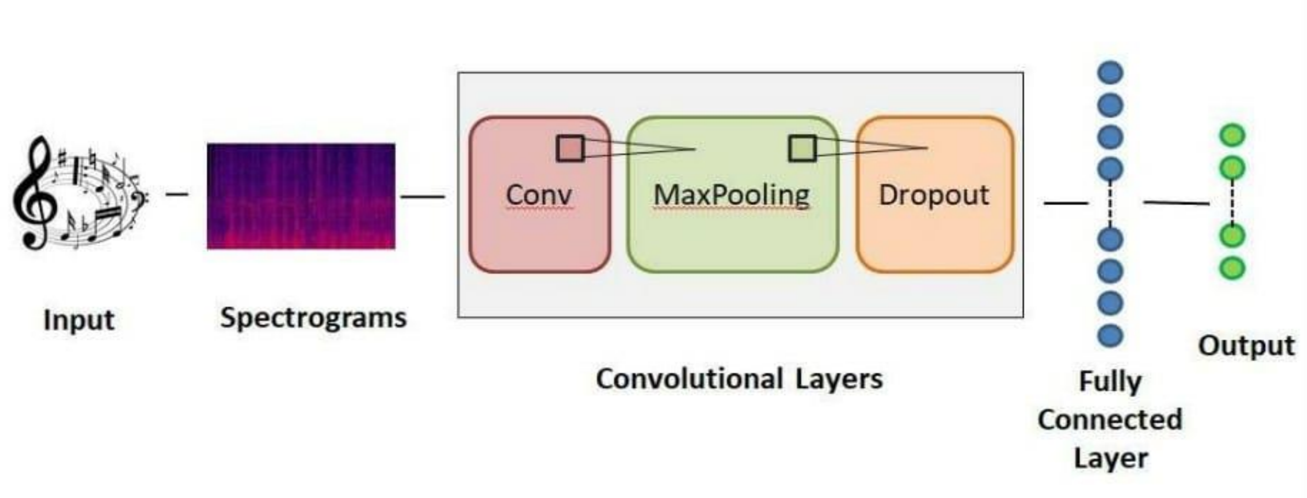

Fig. 1 Working of convolutional neural network

ISSN: 2233-7853 IJFGCN

Input will be presented in the form of audio file in .wav format. Dataset used here to train and test the model is GTZAN which consist of 1000 different songs with 10 different genres (100 songs/genre). Each song will get converted to its respective spectrogram which is nothing but a 2-D representation of an audio signal, having time on x-axis and frequency on y-axis. The table provided below shows different genres in the GTZAN dataset.

TABLE I
GENRES IN GTZAN DATASET

Sr. No.	GENRES	No. of Songs
1	BLUES	100
2	CLASSICAL	100
3	POP	100
4	ROCK	100
5	COUNTRY	100
6	DISCO	100
7	HIP-HOP	100
8	JAZZ	100
9	REGGAE	100
10	METAL	100

The next step is to gather some parameters or features from the dataset. This step includes computing Mel spectrogram of every audio file in the dataset and also computing its MFCC(Mel Frequency Cepstral Coefficients). Mel spectrogram is generated using Mel scale. MFCC takes audio samples as input and after processing, it calculates coefficients unique to that particular sample. Also, several numeric features of the audio file are computed using Librosa library. The numeric features included are first 13 mfccs, zero-crossing rate, spectral centroid and spectral rolloff. This Mel spectrogram data and numeric features will be used to train and test the CNN model. Out of 1000, 800 songs are trained and 200 are tested.

Convolution Neural Network (CNN) is a deep learning module used to classify a short segment of the spectrogram, which is passed to hidden layers which consists of convolution, max pooling and dropout. The data is split into training and testing set of 800 and 200 songs respectively. CNN then predicts the computed genre of the music. CNN gave the training accuracy of 80% and testing accuracy of 62%. As number of epochs are increased, model learns more but too much epochs would lead to overfitting, so increasing the number of epochs would likely not improve the model. Convolutional neural network has its origin from the study of biological neural system. Humans can find and recognize patterns without having to re-learn the concept. Convolutional Neural Net can also do this because it recognizes an object as an object even when it appears in some different way.

IV. CNN ARCHITECTURE

At the heart of it all, convolutional neural networks can be seen as a neural network that uses the same copy of the same neuron. This allows the network to work with a lot of different hidden layers and perform on highly computational models. In simple terms, CNN can take an image as an input and it consists of multiple hidden layers of artificial neurons. Pre-Processing required in convolution networks is much lower as compared to other classification algorithms. As in primitive methods filters used are hand-engineered having enough training, convolution networks have the ability to automatically learn these filters. The result of this is specific features that can be detected anywhere

ISSN: 2233-7853 IJFGCN

on input images. The architecture of this networks is similar to that of connectivity pattern of neurons in the Human Brain. The following figure shows how hidden layers work in CNN.

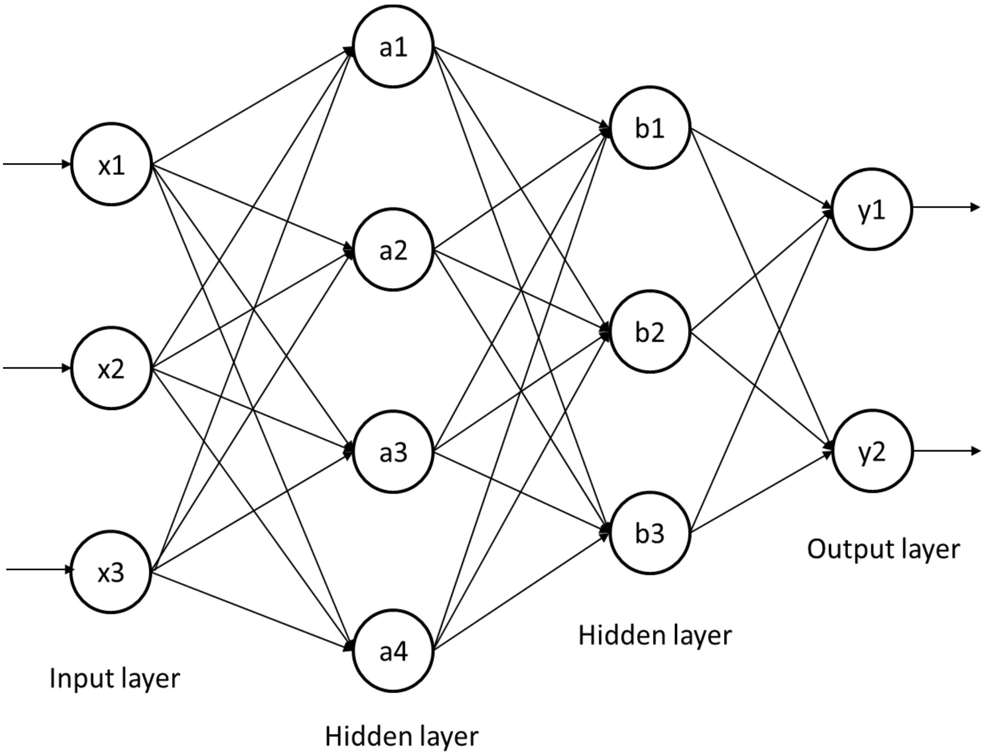

Fig 2 Layers of CNN

Convolutional layers are not only applied to input data but they can also be applied to the output of other layers. More data needs more layers to be able to learn more complicated patterns. In multiple layers the filters that work on raw data will extract low-level features and the extraction of features will be at high-level as depth of the network is increased. There are some basic and important layers that are used in CNN.

A. Convolution layer

This is the first layer of the network which is used to extract different features from the input spectrogramic image. In this layer all the mathematical operations of convolution is perfomed using the filters. In convolution layer the implementation starts with a larger image and ends with a smaller array that helds records of which sections of image were most interesting. In this the information about the image like the corners and edges is extracted.

B. Maxpooling and Droput layer

Convolution layer is followed by pooling layer in most of the cases as the fundamental aim of this layer is to decrease the size of convolved feature and subsequently reduces the computational cost. This is done by reducing the connections between layers and independent operation on each feature map. This layer reduces the size of the array while keeping the most interesting features. Dropout layer is used just in case of over-fitting. To overcome this, few neurons are randomly dropped from the network which simultenously reduces the size of model.

C. Fully-Connected layer

ISSN: 2233-7853 IJFGCN

By convolution and maxpooling a large image is reduced in a small array. That array can be used as input to another layer named as fully connected layer. This layer is the final layer in which the main purpose of classification is carried out thus it is just before the final output.

The basic idea of these layers is to summarize a large image until the final result is achieved.

V. Future Scope

This work was done for quick implementation so it has a lot of scope for improvement. The dataset used in this study has
1000 songs of 10 different genres. Training and testing using larger dataset in future might improve CNN's ability to classify music into its respective genre. Few songs result in weaker analysis and over fitting of some algorithms which can be improved with large dataset. Also using more features of the audio file could help to enhance the CNN accuracy.

VI. Conclusion

The proposed system is a technique for classifying music genres. The exponential growth within digital industry causes an urgent need for effective genre classification for users as well as content provider. The system used is robust and cost effective because open-source tools like Python, Jupyter can be used. And it is also easy to implement as it won't need much higher specifications. In case of classification done by human being there is a good chance of error because of the different nature of every individual. If we have some system which will categorize music into its respective genres automatically, we would save a large number of human efforts and time. CNN model used gives better accuracy i.e., 80% training accuracy than conventional machine learning algorithms. Misclassification may occur in some cases due to complex audio tracks but overall CNN gives better accuracy. So, to justify, it would be better to have an automatic genre classifier for music.

References

[1] Snigdha Chillara, Kavitha A S, Shweta A Neginhal, Shreya Haldia, Vidyullatha K S, “Music Genre Classification using Machine Learning Algorithms: A Comparison” International Research Journal of Engineering and Technology, Volume 06, Issue 05, May 2019.

[2] Hareesh Bahuleyan, “Music Genre Classification using Machine Learning Techniques" University of Waterloo, Canada, Apr 2018.

[3] Tom LH. Li, Antoni B. Chan, Andy HW. Chun, “Automatic Musical Pattern Feature Extraction using Convolutional Neural Network” International Multiconference of Engineers and Computer Scientists, Volume 01, March 2010.

[4] Bryn Lansdown, Dr. Shan He. “Machine Learning for Music Genre Classification” University of Birmingham, September 2019.

ISSN: 2233-7853 IJFGCN

[5] George Tzanetakis, Perry Cook, "Music Genre Classification of Audio Signals" IEEE Transactions on speech and audio processing, Volume 10, Issue 05, July 2002.
[6] S. Lippens, J. P. Martens, T. De Mulder, G. Tzanetakis, "A Comparison of Human and Automatic Musical Genre Classification." IEEE International Conference on Acoustics, Speech and Signal Processing, 2004.
[7] Gabriel Gessle and Simon Akesson" A comparative analysis of CNN and LSTM for music genre classification" Degree Project in technology, Stockholm, Sweden 2019.
[8] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting", Journal of Machine Learning Research, 2014.
[9] R. Thiruvengatanadhan "Music Genre Classification using GMM. (Gaussian mixture model)" (IRJET) volume 05 Issue:10 October 2018.
[10] Gandharva Deshpande, Sushain Bhat, "Audio Genre Classification using Neural Networks" IRJET, Volume 06, Issue 08, August 2019.
[11] Nirmal Vekariya, Hemang Vyas. Nirav Dedhiya, "Music Information Retrieval and Genre Classification using Machine and Deep Learning Techniques" IRJET, Volume 07, Issue 07, July 2020.
[12] Weibin Zhang, Wenkang Lei, Xiangmin Xu, Xiaofeng Xing, "Improved Music Genre Classification using Convolutional Neural Networks" Interspeech 2016, September 2016.
[13] Corey Kereliuk, Bob L. Sturm, Jan Larsen "Deep Learning and Music Adversaries" IEEE, Issue:2015.
[14] Rajeeva Shreedhara Bhat, Rohit B. R., Mamatha K. R. "Music genre classification" IJCMS journal Volume 7, Issue 1, 2020.
[15] Sam Clarke, Danny park, Audrien Guerard" Music Genre Classification Using Machine Learning Techniques"2012.
[16] Tao Feng, "Deep learning for music genre classification", 2014.
[17] Muhammad Asim Ali, Zain Ahmed Siddqui, "Automatic Music Genres Classification using Machine Learning", International Journal of Advanced Computer Science and Applications, Vol 8, No 8, 2017.
[18] Meng A. Ahrendt, P. & Larsen J. "Improving music genre classification by short-time feature integration", ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, V, 497–500, 2005.
[19] Michael I. Mandel and Daniel P.W. Ellis, "Song-level Features and Support Vector Machines for Music Classification", Queen Mary, University of London, 2005.
[20] S. Sigtia and S. Dixon, "Improved music feature learning with deep neural networks," in Acoustics, Speech and Signal Processing (ICASSP), IEEE International Conference on IEEE, pp. 6959–6963, 2014.

ISSN: 2233-7853 IJFGCN