International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 08 Issue: 07 | July 2021 www.irjet.net p-ISSN: 2395-0072

Automatic Music Genre Segregation Using Machine Learning Algorithams

Monika Rokade[1], Prerana D Gayke[2], Sunil S Khatal[3]

[1]Professor, Dept.of Computer Enginering, Sharadchandra Pawar Collage of Enineering, Dumbarwadi, Otur.
[2]M.E Student, Dept.of Computer Enginering, Sharadchandra Pawar Collage of Enineering, Dumbarwadi, Otur.
[3]Professor, Dept.of Computer Enginering, Sharadchandra Pawar Collage of Enineering, Dumbarwadi, Otur.

Abstract - *Music has become the most liked or favourable areanowadays and people want a user-friendly way to classify the musicas per their choice and due to this "Music Genre Classification" cameinto picture. Music genre classification can be imple- mented usingvarious machine learning algorithms. In the proposed system, weare using a deep learning technique .i.e. convolution neural network(CNN) for classi- fying the music in various genres. While analysingmusic, feature extraction is the most pivotal task. Music has become the most liked or favourable areanowadays and people want a user-friendly way to classify the musicas per their choice and due to this "Music Genre Classification" cameinto picture. Music genre classification can be imple- mented usingvarious machine learning algorithms. In the proposed system, weare using a deep learning technique .i.e. convolution neural network(CNN) for classi- fying the music in various genres. While analysingmusic, feature extraction is the most pivotal task.*

***Key Words*: Music gener, Convolution Neural Network, Classifiction.**

1.INTRODUCTION

To develop an automatic classification system for users to ease the process of music genre classification with improved performance and accuracy using machine learn- ing algorithm. After classification the proposed system will segregate the song into various folders. We are providing an user friendly GUI to access the system and enhance the user experience. When a user downloads a song into his system, he/she has to manually save the song into folder of its appropriate genres.

This process is tedious and time consuming. Due to this we are proposing our system, where the user just has to upload the song on the application and the song will get segregated into its respective folder. When a user downloads a song into his system, he/she has to manually save the song into folder of its appropriate genres. This process is tedious and time consuming. Due to this we are proposing our system, where the user just has to upload the song on the application and the song will get segregated into its respective folder.

To develop an automatic classification system for users to ease the process of music genre classification with improved performance and accuracy using machine learn- ing algorithm. After classification the proposed system will segregate the song into various folders. We are providing an user friendly GUI to access the system and enhance the user experience. Generally, people carry their favorite songs on smartphones. Songs can be of various genres. With the help of deep learning techniques, we can provide a classified list of songs to the smartphone user. We will apply deep learning algorithms to create models, which can classify audio files into various genres. After training the model, we will also analyze the performance of our trained model.

1.1 Dataset:

We will use GITZAN dataset, which contains 1000 music files. Dataset has ten types of genres with uniform distribution. Dataset has the following genres: blues, classical, country, disco, hiphop, jazz, reggae, rock, metal, and pop. Each music file is 30 seconds long.

1.2 Process Flow:

Figure 01 represents the overview of our methodology for the genre classification task. We will discuss each phase in detail. We train three types of deep learning models to explore and gain insights from the data. After preprocessing the data, we create our first deep learning model. We construct a Convolution Neural Network model with required input and out units.

The final architecture of our CNN model is shown in Figure 04. We use only spectrogram data for the training and testing. First, we need to convert the audio signals into a deep learning model compatible format. We use two types of formats, which are as follows:

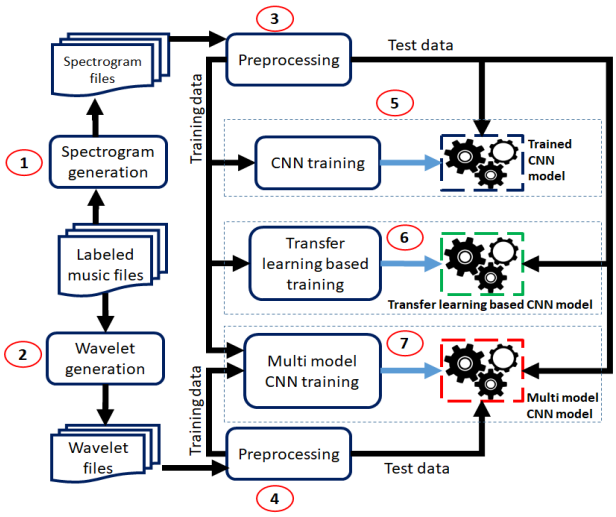

Fig -1: Process Flow

2. TYPES OF DEEP LARNING MODEL

2.1 Spectrogram Generation:

A spectrogram is a visual representation of the spectrum signal frequencies as it varies with time. We use librosa library to transform each audio file into a spectrogram. Figure 02 shows spectrogram images for each type of music genre.

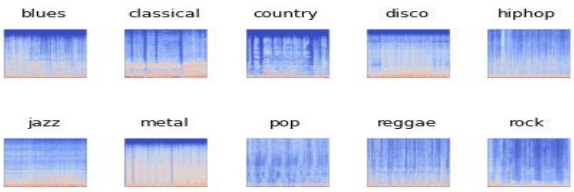

Fig -2: Spectrogram Generation

2.2 Wavelet generation:

The Wavelet Transform is a transformation that can be used to analyze the spectral and temporal properties of non-stationary signals like audio. We use librosa library to generate wavelets of each audio file. Figure 03 shows wavelets of each type of music genre.

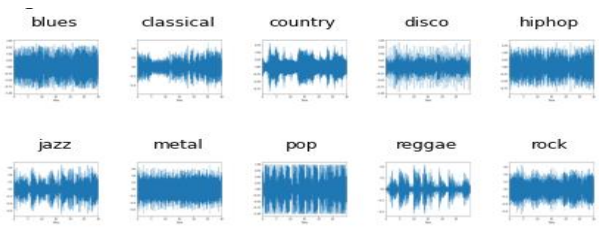

Fig -3: Wavelet Generation

3. LITERATURE SURVEY

Yandre M.G. Costaa, Luiz S. Oliveira b, Carlos N. Silla Jr. c[3] in their paper "An evaluation of Convolutional NeuralNetworks for music classification using spectrograms"published in year 2017 proposed music genre classificationsystem by using Convolution Neural Network and SupportVector Machine algorithm. They used three dataset: LatinMusic Dataset [LMD], ISMIR, African Music Database. LMDconsisted of three 3,277 full-length music pieces of 10different genres. ISMIR 2004 consisted of 1,458 musicpieces. This dataset was divided into 50% training and 50%testing.In this paper output of both algorithms is combinedto get more accuracy. The result obtained is the fusion of CNN and SVM which gives 83% Accuracy.

Nikki Pelchat and Craig M. Gelowitz[2] in their paper"Neural Network Music Genre Classification" published in year 2019 proposed a music genre classification system using a machine learning approach that is Convolutional Neural Network (CNN) algorithm. They used a dataset consisting of 1880 songs of different genres. Length of each song in dataset was of 3 minutes. They generated 132,000 labelled spectrogram images by dividing song into 2.56 second segment spectrograms .Then, they splitted dataset into 70% training data, 20% validation data, and 10% testdata. These spectrogram images are given as input to CNNalgorithm which then classifies the song according to theirgenre. The accuracy obtained in this paper is 67%.

S. Vishnupriya and K. Meenakshi[10] in their paper"Automatic Music Genre Classification using ConvolutionalNeural Network" published in year 2018 proposed an automatic music genre classification system using Convolution Neural Network(CNN). They used GTZANdataset as input to their system. Feature extraction is an important task which is performed in pre-processing andthey used Mel Frequency Cepstral Coefficient (MFCC) as
feature vector. Librosa package is used for feature vectorextraction in python. The size of feature vector is 599x13x5for MFCC. From this database, input is then given to generatespectrograms in the training phase.The learning accuracyobtained using MFCC in this paper is 76%.

Pioneered their work on music genre classification using machine learning algorithm. They created the GTZAN dataset which is till date considered as a standard for genre classification. Changsheng Xu et al .

4. ARCHITECTURE

When a user downloads a song, it gets saved into the downloads folder and if he goes on downloading multiple songs, all the songs will be cluttered there and the user has to segregate them manually into various folders genre-wise which is tedious and time consuming process. Hence we are proposing a system to automate this for users to ease the process of music genre classification with improved performance and accuracy using a machine learning approach.

When a user downloads a song into his system, he/she has to manually save the song into folder of its appropriate genres. This process is tedious and time consuming. Due to this we are proposing our system, where the user just has to upload the song on the website and the song will get segregated into its respective folder.

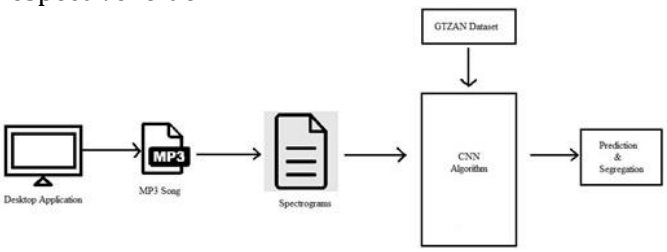

Fig -4: – Architecture Diagram

4. 1 Feature Extraction:

Mel Scale plays an important part. It relates perceived frequency, or pitch, of a pure tone to its actual measured frequency. A frequency which is measured in Hertz (f) can be converted to Mel Scale.

5. ABOUT THE DATASET:

GTZAN Genre Collection dataset was used to perform the classification. The dataset has been taken from the popular software framework MARSYAS. Marsyas (Music Analysis, Retrieval and Synthesis for Audio Signals) is an open source software framework for audio processing with specific emphasis on Music Information Retrieval applications. It has been designed and written by George Tzanetakis (gtzan@cs.uvic.ca). Marsyas has been used for a variety of projects in both academia and industry. Dataset consists of 1000 audio tracks each 30 seconds long. It contains 10 genres(Blues, Classical, Country, Disco, Hip-Hop, Jazz, Metal, Pop, Reggae and Rock), each represented by 100 tracks. The tracks are all 22050Hz Mono 16-bit audio files in .wav format.

Dataset consists of 1000 audio tracks each 30 seconds long. It contains 10 genres(Blues, Classical, Country, Disco, Hip-Hop, Jazz, Metal, Pop, Reggae and Rock), each represented by 100 tracks. The tracks are all 22050Hz Mono 16-bit audio files in .wav format.

5.1 Data Pre-Processing was done in the following manner:

1. Database of the complete collection was created and stored in a .csv file.

2.Feature Vector Extraction is done using the libROSA package in python. libROSA is a python package for music and audio analysis which provides the building blocks necessary to create music information retrieval systems.

3.Each audio file is taken and from that, its feature vector is extracted. The extracted feature vector is called MFCC(Mel-Frequency Cepstral Coefficients). The MFCCs encode the timbral properties of the music signal by encoding the rough shape of the log-power spectrum on the Mel frequency scale. A ZeroCrossings graph is plotted for each audio track. This graph visualizes the number of times the signal crosses zero level.

4. Fourier Transforms are applied on the music signal. A Frequency Spectrum is thus obtained. Mel Scale Filtering is applied on the frequency spectrum to obtain a Mel Frequency Spectrum. A log() function is applied on this Mel Frequency Spectrum which is transformed into Cepstral Coefficients on applying discrete cosine transforms. Finally, the Feature Vector is obtained by finding out the derivatives of the Cepstral Coefficients.

6. CONVOLUTIONAL NEURAL NETWORK:

CNN is a Deep Learning algorithm which can take an input image as input, assign importance (learnable weights and biases) to various aspects/objects in the image and be able to differentiate one from the other. ACNN has various layers such as Convolutional layers, ReLU layers, Pooling layers and a fully connected layer CNN is widely used for image classification because it does automatic feature extraction using convolution.

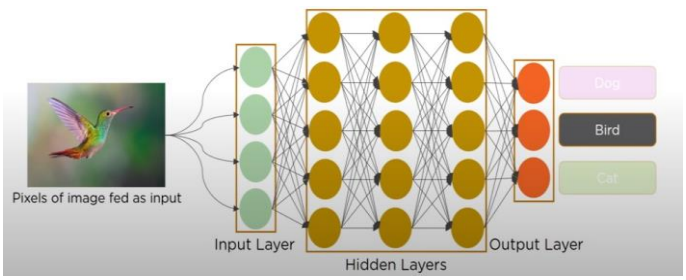

Fig -5: – : Basic CNN Algorithm Structure

1.Input layer accepts the array of pixels as an input.

2.The hidden layer carry out feature extraction by performing certain calculations and manipulation. This layer users matrix filter and performs convolution operation to detect patterns in images and there are

multiple hidden layers like convolution layer, Pooling, ReLU, etc. that performs feature extraction from the image.
3.Output layer i.e. fully connected layer that identifies the object in the image.

6. 1 Layers In CNN:
1.Convolution Layer : It is the main part of CNN. Here whole processing of image is done like all operations of detecting patterns in an image.
2.ReLU (Rectified Linear Unit) : It is an activation function that will output the input directly if it is positive otherwise, it will set output as 0. That is the function returns 0 if it receives any negative input but for any positive value 'x', it returns that value back.
3.Pooling Layer : Polling data or mapping all the data or images and reducing the data together.
4.Fully connected layer : Here the output is generated by taking the input from pooling layer and matching it with the expected output.

7. CONCLUSION:

This research work provides the details of an application which performs Music Genre Classification using Machine Learning techniques. The application uses a Convolutional Neural Network model to perform the classification. A Mel Spectrum of each track from the GTZAN dataset is obtained. This is done by using the libROSA package of python. A piece of software is implemented which performs classification of huge database of songs into their respective genres.

The extension of this work would be to consider bigger data sets and also tracks in different formats(mp3, au etc). Also, with time the style represented by each genre will continue to change. So the objective for the future will be to stay updated with the change in styles of genres and extending our software to work on these updated styles. This work can also be extended to work as a music recommendation system depending on the mood of the person.

ACKNOWLEDGEMENT:

Here, I express my sincere thanks to our guide Prof. MONIKA ROKDE for their constant support and for providing guidance. We are also thankful to all other researchers for their publications.

REFERENCES:

[1] Elizabeth Nurmiyati Tamatjita, Aditya Wikan Mahastama, "Comparison of Music Genre Classification Using Nearest Centroid Classifier and k-Nearest Neighbours", 2016 InternationalConference on Information Management and Technology (ICIMTech),18 May 2017
[2] Machine Learning GeeksforGeek - https://www.geeksforgeeks.org/machine-learning/
[3] GTZAN Dataset Site : http://marsyas.info/downloads/datasets.html
[4] Musical Genre Classification with Convolutional Neural Networks by Leland Roberts : https://towardsdatascience.com/musical-genre-classification-with-convolutionalneural-networks-ff04f9601a74
[5] Understanding the Mel Spectrogram by Leland Roberts -https://medium.com/analyticsvidhya/understanding-the-mel-spectrogram-fca2afa2ce53
[6] Convolutional Neural Network Tutorial by Simplilearn - https://www.simplilearn.com/tutorials/deep-learningtutorial/convolutional-neural-network
[7] Getting started with Django - https://www.djangoproject.com/start/
[8] ISMIR2004 Genre dataset - https://www.upf.edu/web/mtg/ismir2004-genre
[9] The Latin Music Database (LMD) - https://sites.google.com/site/carlossillajr/resources/the-latin-music-database-lmd
[10] Music of Africa dataset - https://research.google.com/audioset/dataset/music_of_africa.html
[11] Understanding Audio data, Fourier Transform, FFT andSpectrogram features for a Speech Recognition System by Kartik Chaudhary https://towardsdatascience.com/understanding-audio-data-fourier-transform-fft-spectrogram-and-speech-recognition-a4072d228520
[12]Monika D.Rokade ,Dr.Yogesh kumar Sharma,"Deep and machine learning approaches for anomaly-based intrusion detection of imbalanced network traffic." IOSR Journal of Engineering (IOSR JEN), ISSN (e): 2250-3021, ISSN (p): 2278-8719
[13]Monika D.Rokade, Dr.Yogesh kumar Sharma"MLIDS: A Machine Learning Approach for Intrusion Detection for Real Time Network Dataset", 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), IEEE
[14] Monika D.Rokade, Dr. Yogesh Kumar Sharma. (2020). Identification of Malicious Activity for Network Packet using Deep Learning. International Journal of

Advanced Science and Technology, 29(9s), 2324 - 2331.
[15]Sunil S.Khatal ,Dr.Yogesh kumar Sharma, "Health Care Patient Monitoring using IoT and Machine Learning.", **IOSR Journal of Engineering (IOSR JEN),** ISSN (e): 2250-3021, ISSN (p): 2278-8719
[16] Sunil S.Khatal ,Dr.Yogesh kumar Sharma, "Data Hiding In Audio-Video Using Anti Forensics Technique ForAuthentication ", IJSRDV4I50349, Volume : 4, Issue : 5
[17] Sunil S.Khatal Dr. Yogesh Kumar Sharma. (2020). Analyzing the role of Heart Disease Prediction System using IoT and Machine Learning. International Journal of Advanced Science and Technology, 29(9s), 2340 - 2346.