International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 01 | January - 2023 Impact Factor: 7.185 ISSN: 2582-3930

Review on Music Classification using Machine Learning

Asmita Patil
Student
GHRIET,Pune

Pallavi Wankhede
Student
GHRIET,Pune

Aditi Nimbalkar
Student
GHRIET,Pune

Jyoti Y. Deshmukh
Faculty
GHRCEM, Pune

ABSTRACT -

Music is the most popular area, particularly among all people. Most individuals like to listen to music in a specific genre, such as classical, hip-hop, or disco, and they desire an easy way to categorise the music according to their mood. Due to the selection and extraction of relevant data, music genre categorization in music information retrieval (MIR) is a challenging process. The methodology for classifying songs and audio music into the appropriate genres is known as the Music Genre Classification Model(MGCM). A more effective and precise model for this classification has to be created since people's lifestyles are growing more reliant on music, technology and the internet as these things become more affordable to end users. We are using deep learning in the suggested system. Convolution neural networks (CNN) are used to categorise music into different genres. Acoustic feature extraction is the most important process while evaluating music. The model in the proposed system is trained using the GTZAN dataset.

Keywords: Deep Learning, Classification, Convolution Neural Network, Music Information Retrieval, Feature Extraction.

I. INTRODUCTION

There are many different songs accessible now in many different genres. These tunes are calming, upbeat, and happy to listeners. Studies demonstrate that relaxing music promotes mental health, lowers stress levels, and eases anxiety. Additionally, the music industry has seen substantial changes as a result of globalisation since numerous individuals, including musicians and music producers, have drawn inspiration from various musical genres from around the globe to create soulful music. Users can choose from a variety of music as a result.

Numerous music streaming services, like Spotify, Gaana, Prime Music, Youtube Music, etc., have enhanced their song suggestions and categorization methods with the use of new developing technologies. Customers now find it very simple to stream music.

As the name implies, machine learning is simply the training of a machine, or computer programme. Without any explicit programming, we enable this machine to learn a variety of things. It's an intriguing area of artificial intelligence where robots learn from the diverse facts at their disposal.

The categorization of genres is a crucial subject with several practical applications. As the amount of music being released daily continues to soar, especially on internet platforms like Soundcloud and Spotify, the need for accurate meta-data required for database management and search/storage purposes rises in proportion. Any music streaming or purchase service should have the ability to rapidly categorise songs in a playlist or collection by genre and the statistical analysis that accurate labelling of music and audio permits is basically endless. We put into practise many categorization algorithms that accept input from separate sources.

One of the primary categories used to categorise millions of music is genre. The tracks are divided into a few different genres. With the most effective methods and algorithms now available, a futuristic model that improves song classification in the music business must be created for the present and forthcoming generations.

GTZAN DATASET:

Hip-hop, rock, classical, blues, country, disco, jazz, reggae, pop and metal are among the ten genres represented in the GTZAN i.e. Genre Collection dataset, which has 1000 songs each lasting 30 seconds. The most often used dataset for machine learning research on music genre identification (MGR) is one that is publicly available.The dataset has the following folders:

1. Genres original : 10 different genres, each featuring 100 audio files that are all 30 seconds long (the famous GTZAN dataset, the MNIST of sounds)

2. Images original : A visual representation of every audio file. Neural networks (NNs) are one method of data classification since they frequently incorporate some kind of picture representation.

3. 2 CSV files : Contains the audio file's characteristics. A mean and variance computed across numerous characteristics that may be recovered from an audio stream are contained in one file, one for each song (30 seconds long). The songs are separated into three-second audio files in the other file, which otherwise has the same format.

II. LITERATURE SURVEY

Nikki Pelchat and Craig M. Gelowitz[4] suggested a music genre classification system employing a machine learning method called the Convolutional Neural Network (CNN) algorithm in “Neural Network Music Genre Classification," published in 2019. They made use of a dataset made up of 1880 songs from various genres. Each song in the dataset has a 3 minute duration. By segmenting songs into 2.56 second spectrograms, they produced 132,000 tagged spectrogram pictures. Then, they divided the dataset into training, validation,

and test data, each comprising 70% of the total. The CNN algorithm receives these spectrogram pictures as input and uses them to categorise the music by genre. This study's accuracy rate was 67%.

S. Vishnupriya and K. Meenakshi[3] suggested in their 2018 publication "Automatic Music Genre Classification Using Convolutional Neural Network," an automated approach for classifying musical genres (CNN). They fed their system with data from the GTZAN project. The crucial work of feature extraction is carried out during pre-processing, and the Mel Frequency Cepstral Coefficient (MFCC) was chosen as the feature vector. Python's Librosa module is used to extract feature vectors. For MFCC, the feature vector measures 599x13x5. Input is then provided to build spectrograms in the training phase from this database. In this study, learning accuracy with MFCC was 76%.

Haree, a book by Hareesh Bahuleyan (2018). The researchers created a way for automatically recognising music in a user's library by tagging the songs in their collection using machine learning techniques. It investigates both neural networks and conventional machine learning techniques to achieve their objectives. The first method makes use of a Convolutional Neural Network that is completely trained using characteristics from the spectrogram of an audio stream (images). The second method makes use of a number of Machine Learning algorithms to categorise the music into its many genres. ML methods employed include Random Forest, Logistic Regression, Gradient Boosting (XGB), and Support Vector Machines (SVM). Separately comparing the two methods, they found that the VGG-16 CNN model had the highest accuracy. The optimised model, which has an accuracy of 0.894, was built using a VGG-16 CNN and an XGB ensemble classifier.

Yandre M.G. Costaa, Luiz S. Oliveira b, and Carlos N. Silla Jr. c[1] developed a music genre classification system employing Convolution Neural Network and Support Vector Machine method in 2017 publication "An assessment of Convolutional Neural Networks for music classification using spectrograms," . African Music Database, ISMIR, and Latin Music Dataset (LMD) were the three datasets they used. Three 3,277 full-length songs from ten different genres made up LMD. ISMIR 2004 included 1,458 musical compositions. 50% of this dataset was used for testing, and 50% for training. In order to increase accuracy, the output of the two algorithms is blended in this study. The combined effect of CNN and SVM yields an accuracy of 83%.

Sarfaraz Masood, Aadam Saleem, Anshuman Goel (2017) classified Music Genres Using Neural Networks (MGC). The project's objective is to automatically classify each album's music by genre. This study enables the classification of songs in real time, and the parallel architecture that is suggested may be implemented on a multi-processor computer. In order to classify songs by genre, characteristics such as beats, tempo, energy, loudness, speechiness, valence, danceability, acousticness, and DWT are collected using Echonest libraries. These data are then input into a parallel multi-layer perceptron network. The suggested approach had an accuracy of roughly 85% when categorising the songs for the two different and well-known genres of Indian Hindi music, Sufi and Classical.

Elizabeth Nurmiyati Tamatjita and Aditya Wikan Mahastama[2] published a article in the 2016 study "Comparison of Music Genre Classification Using Nearest Centroid Classifier and kNearest Neighbours" In this study, they examined the Nearest Centroid Classifier (NCC) and K-Nearest Neighbors methods for categorising musical genres (KNN). Here, metrics like Zero Crossing Rate (ZCR), Silent Ratio (SR), and Average Energy (E) are taken from audio files to categorise the genre of songs. They used songs from 12 distinct genres in their dataset. Selected characteristics, such as ZCR, SR, and E, provide the categorization with the highest degree of accuracy. With k-Nearest Neighbors, this paper's accuracy rate is 56.3%. (k-NN).

Weibin Zhang (2016) suggested two techniques for improving music genre classification using convolutional neural networks: 1) utilising a technique influenced by residual learning to combine peak- and average pooling to provide higher level neural networks greater statistical information; 2) exploiting shortcut connections to go around one or more. A deep neural network for classification is fed the output of the CNN. Based on a study of two network typologies, our early results.Our preliminary experimental findings, which evaluate two distinct network typologies in the GTZAN data set, indicate that the second of the two techniques mentioned above, in particular, can significantly increase classification accuracy.

Tom L. H. Li and others (2010) has developed a voice segmentation technique and utilised a convolutional neural network to automatically extract musical pattern information. This is a completely unsupervised state. It employs a novel methodology combining quasi-GMM with LSP correlation analysis. The model can handle open-set speakers, online speaker modelling, and real-time segmentation without any prior knowledge. Less trustworthy statistical spectral features, rhythm, and pitch extracted from audio clips lead to less precise models.They thus used a different strategy than CNN and concentrated on musical data, which is comparable to picture data and requires very little prior information. The relevant dataset was GTZAN. There are ten genres in all, each with 100 audio snippets. Each audio clip is 30 seconds long, has a sample frequency of 22050 Hz, and has a 16-bit resolution. We used the WEKA system to analyse musical patterns, and we used a number of categorization models. The classifier's accuracy started off at 84 percent but subsequently increased. In comparison to the MFCC, chroma, and temp characteristics, the CNN features were more dependable and produced superior outcomes. Additionally, it is possible to increase the precision of parallel computation on various genre combinations..

Carlos N. Silla Jr., Celso A. A. Kaestner, and Alessandro L. Koerich (2008). Automatic Classification of Music Genres Using a Machine Learning Method. An innovative technique for automatically categorising musical genres is presented in this research. The suggested technique uses a large number of feature vectors and a pattern recognition ensemble approach and is based on time and space degradation approaches. Despite the fact that classifying music genres is a multi-class assignment, we are able to complete it by merging the output of many binary classifiers (space decomposition). Additionally, the time segments taken from the beginning, middle, and end of the original audio signal are used to decompose the music samples (time-decomposition).

Tzanetakis and associates (also known as Tzanetakis et (2002). They investigated the best way to categorise audio signals into musical genres automatically. They concur that these musical subgenres are merely artificial category designations intended to put musical works together. They are divided into categories based on a few defining traits. The instruments employed, the rhythmic patterns, and, most importantly, the harmonic content of the song all have an impact on these characteristics. Genre hierarchies are frequently utilised to organise enormously huge internet music collections.They have suggested three distinct feature sets: pitch content, rhythmic substance, and timbral texture. By utilising real-world audio collections to train statistical pattern recognition classifiers, researchers were able to analyse suggested features' performance

and relative relevance. Both whole-file and real-time frame-based categorization strategies are examined in this work. Using the suggested feature sets, this model labels roughly 61 percent of the 10 musical genres correctly.

III. SYSTEM ARCHITECTURE

i. Proposed System

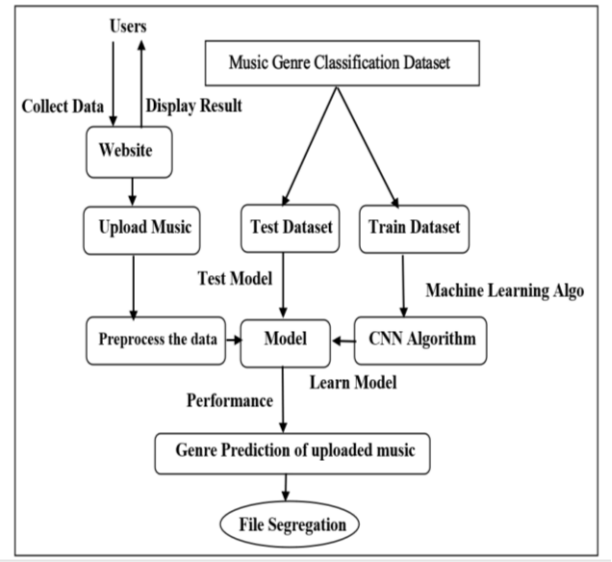

Fig.1: *Proposed System Architecture*

With the help of the GTZAN dataset, we are creating a music genre classification system for the proposed model that will categorise songs according to their genres. After reading the dataset, mel spectrograms—which are an extraction of our song's features—are produced. A file containing the pixel values from spectrograms is given along with the music to be categorised as input to the CNN algorithm.

CNN's algorithm categorises the song into a certain genre using the pixel values. Additionally, our system will automatically classify/place the results of the algorithm, or categorised songs, into the appropriate folder according to their genre; this will assist to obtain a list of songs according to genre. The user will be able to select any genre in accordance with the mood or an event into this genre-based organisation of music into distinct folders. Additionally, appropriately organising the songs in terms of genre will aid the user in selecting the list of songs and managing the music effectively. Consequently, the key modules of this application are as follows:

1. Generation of Spectrogram Files

2. Classification of Musical Genres

3. A website's user interface

4. Put the music in the genre-specific folder.

Mel Scale has a significant role to play in this. It connects a pure tone's perceived frequency, or pitch, to its actual measured frequency. The formula presented below can be used to convert a frequency measured in Hertz to Mel Scale:

$$Mel(f) = 1 + 2595\log(f/700)$$

The following methods for feature extraction in music genre categorization are available:

a. Spectrograms : A spectrogram is a visual representation of a signal's intensity or volume as it changes over time at different frequencies contained in a given waveform. Spectrograms are two-dimensional graphs with a third variable that is represented by colour. Spectrograms are produced using an optical spectrometer. The vertical axis indicates frequency, pitch, or tone, while the horizontal axis represents time. The energy, amplitude, or loudness of audio is represented in the third dimension.

The steps below are what Signal Analyser does to create the audio spectrogram:

i) Equal-length pieces of audio are separated. The segments should be brief enough such that there is minimal variation in the frequency content of the audio within each segment. The portions might overlap or not.
ii) The Short-time Fourier Transform is obtained by windowing each segment and computing its spectrum.
iii) Show the strength of each spectrum in decibels for each section of the spectrogram.

b.MFCC : Mel Frequency Cepstral Coefficients, also known as MFCCs, are frequently employed for feature extraction. Automatic speech recognition uses it. Linear Prediction Coefficients (LPCs) and Linear Prediction Cepstral Coefficient existed before MFCCs (LPCCs). The process of automated speech recognition starts with the extraction of all characteristics. i.e., isolate the elements that are useful for distinguishing linguistic content from other elements such as background noise, emotion, etc. The primary idea is that it pinpoints the human vocal tract's form, which controls what sounds are produced. We can precisely depict the sound being generated if its form can be established. Therefore, it is the responsibility of MFCCs to appropriately portray this data.

The steps to create an MFCC are as follows:
i) Divide the signal into brief (20–40 ms) frames. A standard is regarded as being 25 ms.
ii) Calculating the periodogram estimate of the power spectrum for a single frame - In this case, the discrete Fourier transform of the single frame is used to calculate the periodogram estimate of the power spectrum.
iii) Applying the Mel-spaced filterbank, which consists of 20–40 (26 being the usual) triangle filters, to the result of step 2—a periodogram power spectral estimate—is the final step. In this case, filterbanks have a total of 26 vectors and a 257-length vector.
iv) Take filterbank logarithm - In this step, we are taking the log of each of the 26 energies from step 3.

v) The last step entails finding the discrete cosine transform of the log filterbank energies while preserving the DCT coefficients 2 to 13 and excluding the others.

c.Mel-Spectrogram: A spectrogram in which frequencies are scaled down to mel units is known as a Mel-Spectrogram. The x-axis depicts time, and the y-axis the mel scale. The mel spectrogram is produced after the waveforms of audio files are processed through mel filter banks. Mel-Spectrograms may be produced using the Python language and the Librosa Library.

IV.METHODOLOGY

Convolutional Neural Network

Machine learning includes convolutional neural networks (CNNs), which are a subset of it. It is a subset of the several artificial neural network models that are employed for diverse purposes and data sets. A CNN is a particular type of network design for deep learning algorithms that is used for tasks like image recognition and pixel data processing.

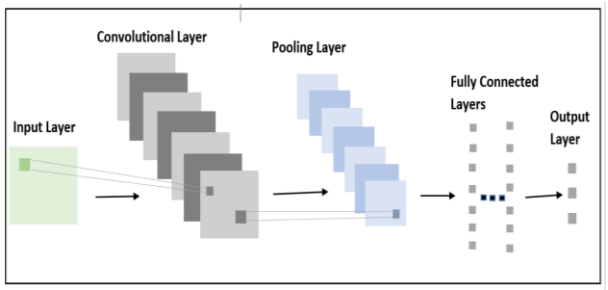

Fig.2: *Convolutional Neural Network*

In deep learning, a convolutional layer, a pooling layer, and a fully connected (FC) layer are the three layers that make up CNN.

Convolutional layer: The central component of a CNN, the convolutional layer is where most calculations take place. The first convolutional layer may be followed by a subsequent convolutional layer. A kernel or filter inside this layer moves over the image's receptive fields during the convolution process to determine if a feature is present.

Pooling layer: The pooling layer, like the convolutional layer, runs a kernel or filter over the input picture. Contrary to the convolutional layer, the pooling layer has fewer input parameters but also causes some information to be lost. Positively, this layer simplifies the CNN and increases its effectiveness.

Fully connected layer: The CNN's FC layer classifies images based on the characteristics that were retrieved from the preceding layers. Fully linked in this context indicates that every activation unit or node of the subsequent layer is connected to every input or node from the preceding layer.

V. CONCLUSION

We are putting out an effective music genre classification method in our system that can categorise the submitted audio file into different genres. To train our system, we use the GTZAN dataset, which contains 1000 songs of different genre. Using the Librosa package that Python provides, features are extracted, and spectrograms are then produced. The CNN Algorithm is used to classify data. The result of CNN's classifier, or the categorised music, is subsequently divided into its appropriate folder. As a conclusion, we can state that the CNN method provided greater classification accuracy when compared to other algorithms like K Nearest Neighbor, SVM etc.

The idea behind this project is to see how to handle sound files, compute sound and audio features from them, run Machine Learning Algorithms on them, and see the results. In a more systematic way, the main aim is to create a machine learning model, which classifies music samples into different genres. It aims to predict the genre using an audio signal as its input.User gets the required audio based on their mood as the outcome.

REFERENCES

[1] Yandre M.G.Costa, Luiz S.Oliveira, Carlos N.Silla Jr., ”An evaluation of Convolutional Neural Networks for music classification using spectrograms”, Applied Soft Computing, Volume 52, March 2017

[2] Elizabeth Nurmiyati Tamatjita, Aditya Wikan Mahastama, “Comparison of Music Genre Classification Using Nearest Centroid Classifier and k-Nearest Neighbours”, 2016 International Conference on Information Management and Technology (ICIMTech), 18 May 2017

[3] S. Vishnupriya, K. Meenakshi, ”Automatic Music Genre Classification using Convolution Neural Network”, 2018 International Conference on Computer Communication and Informatics (ICCCI), January 2018.

[4] Nikki Pelchat, Craig M. Gelowitz, ”Neural Network Music Genre Classification”, Canadian Journal of Electrical and Computer Engineering, Volume: 43, Issue: 3, Summer 2020

[5] Convolutional Neural Network Tutorial by Simplilearn - https://www.simplilearn.com/tutorials/deeplearningtutorial/convolutional-neural-network

[6]Getting started with Django - https://www.djangoproject.com/start/

[7]GTZAN Dataset Site : http://marsyas.info/downloads/datasets.html

[8] Machine Learning GeeksforGeek - https://www.geeksforgeeks.org/machine-learning/

[9] Musical Genre Classification with Convolutional Neural Networks by Leland Roberts : https://towardsdatascience.com/musical-genreclassification-with-convolutionalneural-networks

[10] Understanding the Mel Spectrogram by Leland Roberts - https://medium.com/analyticsvidhya/ understanding-the-mel-spectrogram-fca2afa2ce53