Music Genre Classifier using Machine Learning

Asmita Patil
Student
GHRIET, Pune
asmitap941@gmail.com

Pallavi Wankhede
Student
GHRIET, Pune
pallaviwankhede7117 @gmail.com

Aditi Nimbalkar
Student
GHRIET, Pune
aditinimbalkar1740 @gmail.com

Jyoti Y. Deshmukh
Faculty
GHRCEM, Pune
jyoti1584@gmail.com

ABSTRACT -

Music is a worldwide language with the ability to evoke emotions and unite people. With the proliferation of digital music platforms and the availability of vast music collections, the need for effective organization and exploration of music has become increasingly important. One crucial aspect of music organization is genre classification, which involves categorizing music tracks into specific genre classes based on their sonic characteristics, musical elements, and stylistic attributes. Most people like listening to music in a certain genre, such as classical, hip-hop, or disco, and they want an easy way to categorize the music based on their mood. People's lifestyles are becoming increasingly dependent on music, technology, and the internet as these items become more affordable to end users, that requires the development of a more effective and precise model for this categorization. Deep learning is used in the proposed system. Convolutional neural networks (CNN) are used to classify music into various genres. The suggested system's model is trained using the GTZAN dataset.

Keywords: Deep Learning, Classification, Convolution Neural Network, MFCC, Mel Spectrogram, ResNet18.

I. INTRODUCTION

There are several songs available presently in a variety of genres. Listeners will find these songs relaxing, lively, and pleasant. Relaxing music has been shown in studies to improve mental health, reduce stress, and reduce anxiety. Furthermore, the music business has seen significant changes as a result of globalization, as many people, including artists and music producers, have taken inspiration from diverse musical styles from across the world to create soulful music. As a result, users may select from a wide range of music.

Many music streaming services, such as Spotify, Gaana, Prime Music, and YouTube Music, have improved their song recommendations and classification processes by using new scientific technologies. Customers may now stream music with ease.

Machine learning, as the name indicates, is simply the teaching of a machine, or computer software. We enable this machine to learn a number of things without any explicit programming. It's a fascinating field of artificial intelligence in which robots learn from the many information available to them.

The classification of genres is an important topic with several practical applications. As the volume of music produced everyday increases, particularly on internet platforms like Soundcloud and Spotify, so does the requirement for precise meta-data for database administration and search/storage purposes. Any music streaming or purchasing service should be able to quickly categorize songs in a playlist or collection by genre.

Genre is one of the key categories used to categorize millions of pieces of music. The tunes are separated into several genres. With the most effective methodologies and algorithms currently accessible, a futuristic model that enhances song classification in the music industry for current and future generations must be developed.

GTZAN DATASET:

GTZAN dataset includes 10 genres such as hip-hop, rock, classical, blues, country, disco, jazz, reggae, pop, and metal. A publicly accessible dataset is the most often used dataset for machine learning research on music genre recognition (MGR). The dataset is divided as follows:

1. Original genres: Ten separate genres, each with 100 files and 30-second audio samples (the famous GTZAN dataset, the MNIST of sounds).

2. Original images: A visual depiction of each audio file. Because they usually integrate some form of image representation, neural networks (NNs) are one way of data categorization.

3. CSV files: Contains information about the audio file. One file, one for each song (30 seconds long), has a mean and variance computed across several variables that may be retrieved from an audio stream. In the other file, which otherwise has the same structure, the songs are divided into three-second audio files.

II. LITERATURE SURVEY

Flexer and Schnitzer [1] in their (2006)'s publication "Music genre classification using support vector machines and hidden Markov models" focuses on the use of support vector machines (SVM) and hidden Markov models (HMM) for music

genre classification. The objective of the study is to investigate the combination of SVMs and HMMs for music genre classification. The authors aim to leverage the complementary strengths of these two algorithms to improve the accuracy of genre classification. The paper provides results that demonstrate the effectiveness of the SVM-HMM approach for music genre classification. It shows improved classification accuracy compared to using SVMs or HMMs individually.

Zhang and Hu's [2] in their (2008)'s publication "Music genre classification using support vector machines" focuses on the use of support vector machines (SVM) for music genre categorization. The objective of the study is to investigate the effectiveness of SVMs for music genre classification tasks. The authors aim to classify music audio into different genres using SVMs and evaluate the performance of the approach. They extract various audio features from music signals, such as pitch, rhythm, and spectral features. These features are used to represent the audio content of music tracks. The trained SVM model is used to classify new, unseen music tracks into their respective genres. The paper provides results indicating the effectiveness of SVMs for music genre classification.

L. Yang, R. Jin [3] and R. Pan (2011)'s paper "Music Genre Classification Using Multi-Label Learning" investigates the use of multi-label learning techniques for music genre classification. The objective of the study is to investigate the effectiveness of multi-label learning techniques for music genre classification. Unlike traditional single-label classification, multi-label classification allows assigning multiple genre labels to a music track, considering the possibility of a track belonging to multiple genres simultaneously. They used standard evaluation metrics for multi-label classification, such as precision, recall, F1 score, and Hamming loss. The performance of the algorithms is compared, and the effectiveness of the multi-label approach is analyzed.

S. Lee and K. Lee's [4] (2013) paper "Ensemble Methods for Music Genre Classification" focuses on the use of ensemble methods for music genre classification. The objective of the study is to explore ensemble methods for music genre classification and investigate their effectiveness in improving classification accuracy compared to individual classifiers. The paper provides results that demonstrate the effectiveness of ensemble methods for music genre classification. It shows that the ensemble classifiers outperform the individual classifiers in terms of classification accuracy, highlighting the benefits of combining multiple classifiers to leverage their complementary strengths.

S. Park and J. Lee's [5] (2018) paper "Music Genre Classification Using Convolutional Recurrent Neural Networks" focuses on the use of Convolutional Recurrent Neural Networks (CRNN) for music genre classification. The objective of the study is to investigate the effectiveness of CRNNs for music genre classification. The authors aim to leverage the combined power of convolutional and recurrent neural networks to capture both local and temporal features in music audio. It shows that the CRNN model outperforms or achieves competitive performance compared to other baseline models or state-of-the-art methods, indicating the capability of CRNNs in capturing both local and temporal information for genre classification.

Panda, Meher, and Mallick's [6] (2019) paper titled "A comparative analysis of machine learning techniques for music genre classification" focuses on conducting a comparative analysis of various machine learning techniques for music genre classification. The objective of the study is to compare and analyze the performance of various machine learning techniques for music genre classification. The authors explore multiple machine learning algorithms for music genre classification, such as k-nearest neighbors (KNN), support vector machines (SVM), decision trees, random forests, and naive Bayes. They apply these techniques to a labeled dataset of music samples, where each sample is associated with a specific genre label. The paper provides results that showcase the comparative analysis of the machine learning techniques for music genre classification. It highlights the performance of each technique in terms of classification accuracy and provides insights into the strengths and limitations of the algorithms.

S. Hou et al. [7] (2020) published research named "Exploring CNN Architectures for Music Genre Classification" that looks into the impact of several convolutional neural network (CNN) architectures on music genre classification performance. The objective of the study is to explore and compare various CNN architectures for music genre classification. The authors utilize multiple CNN architectures, including VGGNet, ResNet, and DenseNet, as the backbone models for music genre classification. The paper provides results that demonstrate the impact of CNN architectures on music genre classification performance.

S. Kim and T. Park's [8] (2020) publication "Music Genre Classification Using Deep Learning: A Survey" presents an overview of deep learning approaches used in music genre categorization. The objective of the survey is to provide a comprehensive review of deep learning approaches for music genre classification. It aims to analyze and summarize the deep learning models, architectures, and methodologies used in the field. The paper discusses different deep learning models and architectures, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and their variants (e.g., Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs)). The paper addresses the challenges and limitations of deep learning-based music genre classification, such as data scarcity, high dimensionality, and interpretability issues. It also highlights potential research directions, including cross-genre analysis, hybrid models, and real-time applications.

III. SYSTEM ARCHITECTURE

i. Proposed System

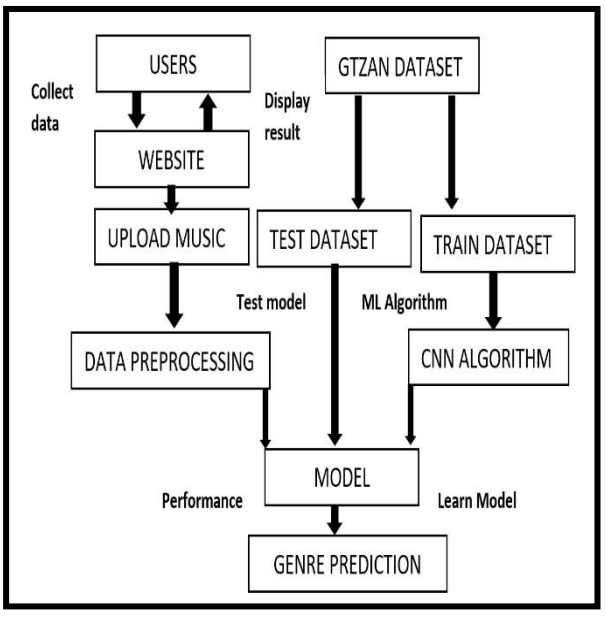

Fig.1: Proposed System Architecture

We are developing a music genre categorization system for the proposed model using the GTZAN dataset, which would categorize songs based on their genres. Mel spectrograms, which are an extraction of our song's characteristics, are generated after reading the dataset. Along with the music to be classified, a file containing the pixel values from spectrograms is provided as input to the CNN algorithm.

Using the ResNet 18 Model, CNN's algorithm classifies the song into a certain genre. When a user enters a (.WAV) or (.MP3) file into the programme, our system will automatically determine the genre of the music. The user will be able to choose any genre file and upload it as input to the online application. Furthermore, the user will be able to determine the proportion of various genres in a certain audio file. And the genre with the highest proportion in a certain song will be the final product. As a result, the following are the application's major modules:

1. Creating Spectrogram Files.

- Train the GTZAN Dataset.

- We get Spectrogram image using Librosa library in .png format based on frequency and amplitude.

2. Genre classification in music.

- Feature Extraction using MFCC.

- Processes the Spectrograms in batches.

- Load a pre-trained ResNet18 model of CNN from the torchvision library.

- Test the Data.

3. The user interface of a website.

4. Prediction of the genre of any (.WAV) or (.MP3) audio file.

Mel Scale plays an important function in this. It compares a pure tone's perceived frequency, or pitch, to its actual measured frequency. The following formula may be used to convert a frequency measured in Hertz to Mel Scale:

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

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

a. Spectrograms: A spectrogram is a visual depiction of the intensity or volume of a signal as it varies over time at various frequencies contained in a given waveform. Spectrograms are two-dimensional graphs that include a third variable indicated by color. An optical spectrometer is used to create spectrograms. The horizontal axis represents time, whereas the vertical axis represents frequency, pitch, or tone. The third dimension represents acoustic energy, amplitude, or loudness.

The following are the procedures that Signal Analyzer takes to generate the audio spectrogram:

i) Audio segments of equal duration are separated. The segments should be brief enough that the frequency content of the audio within each segment varies minimally.
ii) To acquire the Short-time Fourier Transform, window each segment and compute its spectrum.
iii) For each region of the spectrogram, show the intensity of each spectrum in db.

b. MFCC: Mel-frequency cepstral coefficients (MFCCs) are utilized as audio features to capture the spectral characteristics of music tracks. The process involves dividing the audio signal into frames, converting them to the frequency domain using the Fourier transform, applying the mel-scale filterbank to approximate human auditory perception, taking the logarithm of the filterbank energies, and finally applying the discrete cosine transform (DCT) to obtain the MFCCs. By extracting MFCCs from music tracks, the classifier can capture important information related to tonal patterns, rhythmic structures, and timbral properties. These coefficients serve as a condensed representation of the spectral content of the audio and enable the classifier to learn discriminative features specific to different music genres. MFCCs are widely used in music genre classification due to their effectiveness in capturing genre-related characteristics and improving the accuracy of genre classification models.

c. Mel-Spectrogram: A Mel-Spectrogram is a spectrogram in which frequencies are scaled down to mel units. The x-axis represents time, while the y-axis represents the mel scale. The mel spectrogram is generated after audio waveforms are processed by mel filter banks. The Python programming language and the Librosa Library may be used to create Mel-Spectrograms.

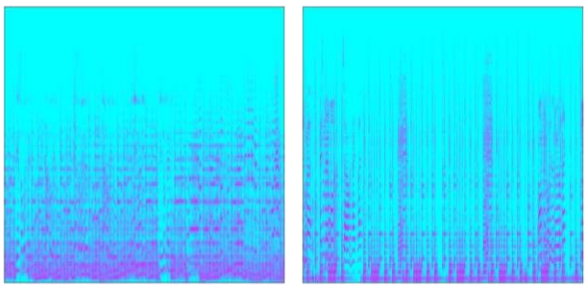

Fig.2: *Mel – Spectrogram of Blues and Rock genre.*

IV.METHODOLOGY

i. Convolutional Neural Network:

Convolutional neural networks (CNNs) are a subset of multiple artificial neural network models used for various applications and data sources. A CNN is a form of deep learning network design that is utilized for tasks such as image recognition and pixel data processing. CNN is made up of three layers in deep learning: a convolutional layer, a pooling layer, and a fully connected (FC) layer.

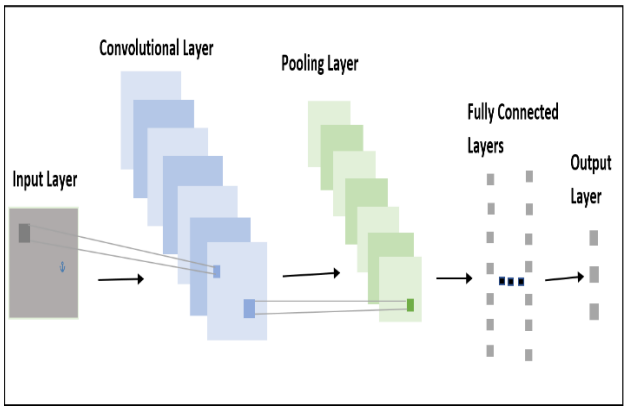

Fig.3: *Convolutional Neural Network*

Convolutional layer: The convolutional layer is the primary component of a CNN and is where the majority of the computations take place. An additional convolutional layer may be added after the initial convolutional layer. During the convolution process, a kernel or filter inside this layer travels through the image's receptive fields to detect whether or not a feature is present.

Pooling layer: Like the convolutional layer, the pooling layer applies a kernel or filter to the input image. The pooling layer, in contrast to the convolutional layer, has fewer input parameters but also causes some information to be lost. This layer simplifies the CNN and improves its efficacy.

Fully connected layer: The CNN's FC layer classifies pictures based on the attributes gathered from previous levels. In this sense, fully linked means that every activation unit or node in the succeeding layer is coupled to every input or node in the preceding layer.

ii. Librosa:

Librosa is a popular Python package for music analysis and feature extraction. It has various capabilities and tools for working with audio data, such as loading and storing audio files, creating spectrograms and mel spectrograms, computing beat and tempo information, extracting harmonic and percussive components, and many more. Librosa has utilities for reading and writing audio files in various formats such as WAV, MP3, and FLAC. Librosa has methods for extracting audio characteristics like MFCCs (Mel-Frequency Cepstral Coefficients), chroma features, and spectral contrast features that may be used as input for machine learning models.

iii. ResNet18:

The ResNet18 "Residual Network" is a form of Convolutional Neural Network (CNN). It has had widespread application in picture recognition and classification applications. It particularly refers to an 18-layer ResNet design with a convolutional layer, many residual blocks, and a fully connected layer at the end. It may also be used to classify music by considering audio signals as spectrograms, which are pictures that depict the frequency spectrum of an audio source across time. To use ResNet18 for music categorization, first transform the audio signals into spectrograms, which divide the signal into tiny pieces and compute the frequency spectrum of each segment. These spectrograms may then be given as input into ResNet18, where the convolutional layers extract features from the spectrograms and the fully connected layers categorize the features into other categories, such as music genres or mood.

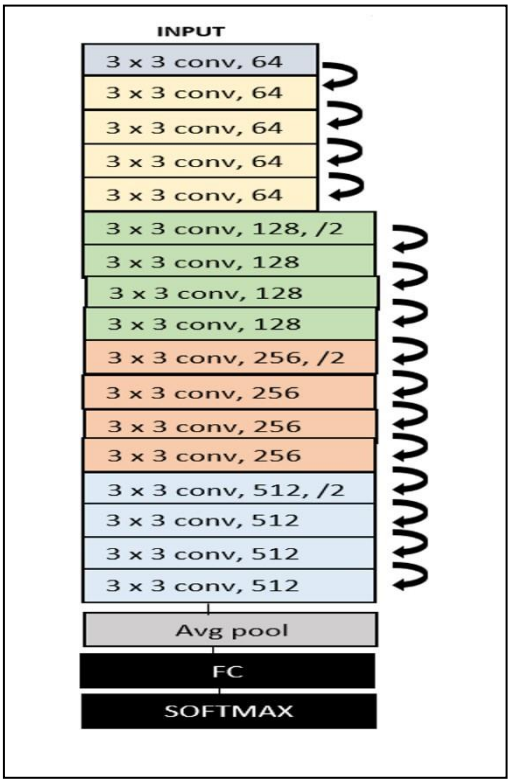

Fig.4: *ResNet 18 Model*

iv. Streamlit:

Streamlit is a Python open-source framework for developing web apps and data visualizations. Users may easily create online apps that connect with machine learning models, data visualizations, and other data analysis tools using Streamlit. It has an easy-to-use API that allows users to construct interactive widgets, charts, tables, and other visualizations. Streamlit is compatible with several data science libraries, including NumPy, Pandas, Matplotlib, and Scikit-Learn, as well as machine learning frameworks like TensorFlow and PyTorch.

V. CONCLUSION

In our system, we are providing an excellent music genre categorization method that can classify the provided audio files into distinct genres. We utilize the GTZAN dataset, which comprises 1000 songs of various genres to train our system. Using Python's Librosa module, features are extracted and spectrograms are generated. To categorize data, the CNN Algorithm's ResNet 18 model is employed. A web application is created that allows users to upload any (.mp3) or (.wav) audio file. The anticipated genre is returned to the user as the final result.

The goal of this project is to learn how to work with sound files, calculate sound and audio attributes from them, run Machine Learning Algorithms on them, and analyze the results. The main goal is to construct a machine learning model that categorises music samples into distinct genres in a more systematic manner. It attempts to guess the genre by using an audio sample as input. As a result, the user receives the appropriate genre for the respective audio file uploaded on the interface.

VI. RESULT

User can upload any audio file of (.MP3) or (.WAV) format on the web application. The audio file size is limited upto 200MB per file. The app is executed using Streamlit library. Once the audio file is uploaded as input by the user, the user is able to play the audio. And as final result the user gets the predicted genre of the respective audio file uploaded.

Comparison of CNN algorithm with other various algorithms used for Music Genre Classifier is as follows:

Classifier	Accuracy	Flexibility	Ease of Use	Training Time	Features
Convolutional Neural Network (CNN)	High	High	High	High	Image-based approach, good for spectrogram-based features.
Support Vector Machine (SVM)	High	Low	Medium	Medium	Feature-based classification, good for structured genres.
Recurrent Neural Network	High	High	Low	High	Sequential data processing, captures temporal dependencies.
K-Nearest Neighbor (KNN)	Medium	High	High	Medium	Instance-based classification, simple and intuitive.
Decision Tree	Medium	High	High	Medium	Hierarchical structure, interpretable model.

REFERENCES

[1] Flexer, A., & Schnitzer, D. (2006). "Music genre classification using support vector machines and hidden Markov models." In Proceedings of the 7th International Conference on Music Information Retrieval (ISMIR) (pp. 438-443).

[2] Zhang, C., & Hu, G. (2008). "Music genre classification using support vector machines." In Proceedings of the 9th International Society for Music Information Retrieval Conference (ISMIR) (pp. 605-610).

[3] Yang, L., Jin, R., & Pan, R. (2011). "Music genre classification using multi-label learning." In Proceedings of the 20th ACM International Conference on Information and Knowledge Management (CIKM) (pp. 1907-1910).

[4] "Ensemble Methods for Music Genre Classification" by S. Lee and K. Lee (2013)

[5] "Music Genre Classification Using Convolutional Recurrent Neural Networks" by S. Park and J. Lee (2018).

[6] Panda, R., Meher, S., & Mallick, P. K. (2019). "A comparative analysis of machine learning techniques for music genre classification." In Proceedings of the 2nd International Conference on Advanced Computational and Communication Paradigms (ICACCP) (pp. 1-6).

[7] "Exploring CNN Architectures for Music Genre Classification" by S. Hou et al. (2020).

[8] Kim, S., & Park, T. (2020). "Music genre classification using deep learning: A survey." IEEE Access, 8, 205520-205531.

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

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

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

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