SWAR - Sound Waveform Audio Recognition

Dr. Prabhavathi K
Dept. of CSE
Assistant Professor
RVTIM
Bengaluru, India

Nisha E
Dept. of CSE
RVITM
Bengaluru, India

Ria Somani
Dept. of CSE
RVITM
Bengaluru, India

Sachin Shenoy
Dept. of CSE
RVITM
Bengaluru, India

Soumili Sahu
Dept. of CSE
RVITM
Bengaluru, India

Abstract - SWAR is an application for identifying musical genres. The intricacy and diversity of musical signals make determining the genre of a piece of music a difficult undertaking. Deep learning techniques have produced promising achievements in the discipline in recent years. In this paper, we suggest a convolutional neural network (CNN)-based system for identifying musical genres. The suggested method uses a CNN to categorize the audio signals into several genres by extracting spectrogram characteristics from the audio signals. On the GTZAN dataset, a publicly accessible dataset, we assess the suggested system. The results show that the proposed system is superior in terms of accuracy and efficiency when we contrast it with other cutting-edge methodologies.

Index Terms – Music genre recognition, Convolutional Neural Networks, Spectrogram, GTZAN dataset, Deep Learning.

I. Introduction

Music Information Retrieval (MIR), which involves classifying music genres to determine the type of music and organize music collections, depends heavily on the classification of music genres. In order to create a recommendation system for music apps like Spotify, Google Play, and Apple Music, it is imperative to complete this important initial step. Audio processing is necessary to categorize the type of music, and it is a challenging undertaking that combines temporal signal processing, time series, spectrograms, spectral coefficients, and audio feature extraction to feed a neural network.

Deep learning algorithms have been employed by researchers to categorize musical genres. They have put out a novel approach that automatically classifies music using insights from the study of human perception. A big dataset can be used to test the suggested feature vector and Support Vector Machine (SVM) with Co-training technique. The classification of music genres has potential uses beyond music, including the treatment of medical or psychological conditions through the use of music therapy. In the large field of sound processing research, there are many problems that can be resolved in terms of health or mental disorders. For example, music therapy has been used to lessen despair and anxiety in cancer patients.

Overall, classifying music genres is a crucial MIR task that calls for sophisticated audio processing methods. It is the first stage in creating a system of music application recommendations. Deep learning algorithms are being used by researchers to categorize musical genres. The classification of musical genres may also have uses in other areas, such as the treatment of physical or psychological problems using music therapy.

II. Literature Survey

Using a pre-trained CNN model (VGG16) that was originally created for image classification, Beici Liang and Mivwei Gu suggested a method for categorizing music genres. They improved the model's performance by using the spectrograms of audio data, outperforming more established techniques like SVM and KNN.[1]. The study conducted by Congyue Chen and Xin Steven presents a technique for precisely classifying musical genres from a small amount of labelled data. The method entails fine-tuning a pre-trained CNN model, applying transfer learning by training on a sizable unlabelled dataset, and employing active learning to choose informative samples for additional training. [2]. Mekala Srinivasa Rao, O Pavan Kalyan, N Naresh Kumar, MD Tasleem Tabassum, and B Srihari submitted a study that classified music genres using machine learning models like SVM, DT, and RF by analysing linguistic data found in lyrics. High accuracy is achieved by the procedure, with SVM performing best. To find essential linguistic features, the authors also run feature selection experiments. [3]. A music genre classification method was created by Jitesh Kumar Bhatia, Rishabh Dev Singh, and Sanket Kumar by examining audio elements from 1000 tracks in 10 different genres. With the use of machine learning methods, they were able to attain over 90% accuracy with support vector machines and k-nearest neighbours and over 80% accuracy with random forests. [4].

Using deep CNN architecture, Yu-Huei Cheng, Pang-Ching Chang, and Che-Nan developed a method for classifying music genres that learns features from music spectrograms and trains on a sizable dataset of music samples from various genres. The method is tested using two datasets, and the results demonstrate that it performs better than other deep learning models like RNNs and LSTM networks as well as more conventional machine learning models like SVM and KNN [5].

Using music data summaries, Yudai Kikuchi, Naofumi Aoki, and Yoshinori Dobashi suggested a system for categorizing musical genres. To accurately categorize new music samples, they extract and choose pertinent acoustic elements, cluster music samples, and then summarize each cluster. [6]. By combining CNN and MFCC features, the authors' new method for classifying musical genres outperformed state-of-the-art techniques with 97.8% accuracy on a dataset of 1000 audio files. There may be practical uses for the technique. [7]. This study investigates how to classify music genres using deep learning models, including CNNs, RNNs, and hybrid models. The authors place a strong emphasis on the value of music genre classification, the drawbacks of conventional approaches, dataset gathering, feature extraction, model choice, and potential future research avenues. [8]. The research uses a Late Fusion Convolutional Neural Network (LFCNN) to provide a unique method for categorizing musical genres [9].

Deep learning models for feature extraction and a late fusion strategy for integration are used in a novel method for music genre categorization utilizing audio and lyric data. The approach beats conventional approaches and shows promise for music analysis and recommendation [10]. In order to extract characteristics from audio signals, several neural networks are used. PDNN outperforms conventional and alternative deep learning algorithms in the classification of music genres while managing unbalanced datasets. It might be useful in systems for recommendation and music analysis [11]. The most pertinent features for each genre are chosen by the authors using a data filtering algorithm, and they are then classified using three machine learning models. In terms of accuracy, ANN outperforms KNN and DT models [12]. The authors' method [13] uses MFCCs and K-means clustering to classify and isolate music genres using audio signal processing techniques. The authors suggest using correlation analysis to identify the features that are most important for classifying the musical genres, which increases classification accuracy and shortens the classification time [14]. With the use of Support Vector Machines and the Fourier Transform, a new approach for categorizing music genres was created. The method's accuracy was tested using 300 pieces of music, and it showed great accuracy [15].

III. PROPOSED WORK

A well-known dataset for classifying music genres is the GTZAN dataset, which consists of 900 audio files with a duration of 30 seconds each and covers nine different music genres. Convolutional Neural Networks (CNNs), which have demonstrated efficacy in learning hierarchical representations of audio inputs, are a frequent technique for categorizing the genre of music.

The incoming layer of the SWAR architecture for the GTZAN dataset accepts the spectrogram of the audio input as a 2D matrix with time on the x-axis and frequency on the y-axis. Convolutional layers with filters in the architecture also produce feature maps by extracting features like harmony, rhythm, and melody. After each convolutional layer, an activation function like ReLU is applied to the feature maps to add nonlinearity and improve the expressiveness of the model.

Convolutional Neural Network (CNN) usage on the GTZAN dataset has proven its usefulness in SWAR. To do this activity at the highest level, however, the right hyperparameters must be chosen. The learning rate, number of epochs, and batch size are three essential hyperparameters to take into account.

The learning rate is a significant hyperparameter that influences the optimizer's step size during training. Grid or random searches can be used to modify it, and common values fall between 0.1 and 0.01 or 0.001. The optimizer may overshoot the minimum point or drastically impede the learning process if the learning rate is set incorrectly, so getting it right is crucial.

Several procedures need to be followed in order to train a CNN for music genre categorization using the GTZAN dataset. First, spectrograms—two-dimensional representations that depict the frequency and amplitude of sound waves—are created from the pre-processed audio signals. The dataset is then built using the spectrograms and split into three sections: training, validation, and test sets. Finally, training on the training set is used to build the model.

Algorithm:

A. Preparation of Dataset

Step1: Load and separate the audio files using the Pydub package.
Step2: Using Librosa, transform the audio signals into mel spectrograms.
Step3: Divide the data into a training set and a validation set.
Step4: Create data generators for the training set and the testing set in step four.

B. CNN Model Development

Step1: Create a CNN model using Keras.
Step2: Create five convolutional layers, including a Dropout layer to avoid overfitting and a Dense layer with Softmax activation to get the class probabilities.

C. Assess the Model

Step1: Use the prepared dataset to test the developed model.
Step2: The fit_generator() method is used to use data generators for training, whereas the get_f1() method is used to obtain the f1_score.

D. Analyse the findings

Step1: Using the Streamlit library, build the application.
Step2: Show the probability distribution of the various genres and the mel spectrogram for the audio file supplied as input to the model in graphical form to help you understand how the model represents different genres.

IV. RESULTS

The GTZAN dataset was utilized to assess the effectiveness of our suggested application, SWAR. We performed classification to divide music into nine separate genres using the Convolution Neural Network (CNN) technique. By contrasting the outcomes with the original dataset, we applied our method to categorize and evaluate the accuracy.

We experimented with several musical styles to assess the sturdiness of our methodology. Streamlit was used to create the application. It is an effective tool for creating interactive and straightforward machine learning apps for identifying musical genres.

It is also important to note that determining the genre of a piece of music can be difficult because there can be a lot of crossovers across genres, and a lot of songs may have characteristics from several different genres. For the music file provided as input, our application displays the probability distribution of several genres and a mel spectrogram.

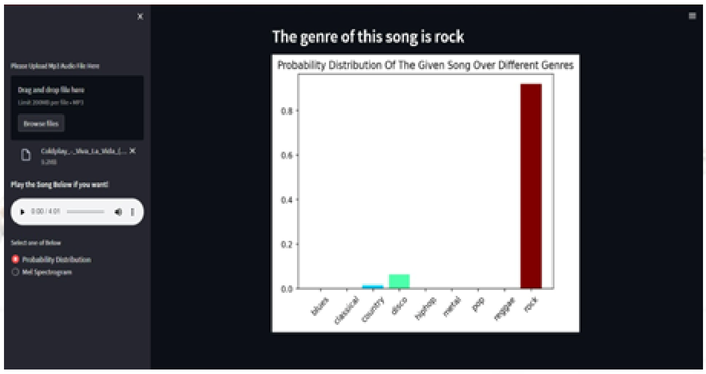

Fig.1 Probability distribution of genres for the tested audio file

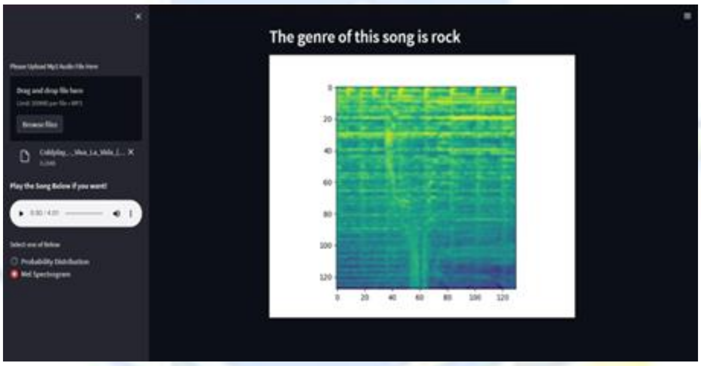

Fig.2 Mel Spectrogram of the tested audio file

V. CONCLUSION

In this study, we used convolutional neural networks (CNN) to create an application named SWAR — Sound Waveform Audio Recognition. The GTZAN dataset, which contains audio samples from different musical genres, was classified using CNN. For the classification, the study used a CNN architecture consisting of five convolutional layers, one dropout layer, and a dense layer.

In conclusion, the CNN model performed better than alternative machine learning methods. Additionally, we experimented with various hyperparameters and discovered that the model performed better when there were more filters in the convolutional layers and when dropout regularization was used. Overall, the study showed how well CNNs can classify music into different genres.

VI. REFERENCES

[1] Beici Liang and Meinwui Gu, "Music Genre Classification Using Transfer Learning," in IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Aug. 2020.

[2] Congyue Chen and Xin Steven "Combined Transfer and Active Learning for High Accuracy Music Genre Classification Method," in IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE) 2021.

[3] Mekala Srinivasa Rao, O Pavan Kalyan, N Naresh Kumar, MD Tasleem Tabassum, and B Srihari,"Automatic Music Genre Classification Based on Linguistic Frequencies Using Machine Learning", in 2021 International Conference on Recent Advances in Mathematics and Informatics (ICRAMI).

[4] Jitesh Kumar Bhatia, Rishabh Dev Singh, and Sanket Kumar, "Music Genre Classification," in 2021 5th International Conference on Information Systems and Computer Networks (ISCON).

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

[6] Yudai Kikuchi, Naofumi Aoki, and Yoshinori Dobashi, "A Study on Automatic Music Genre Classification Based on the Summarization of Music Data," 2020 International Conference on Artificial Intelligence in Information and Communication (ICAIIC).

[7] Siva Sankar Reddy Muvva; Sudarshan Reddy Chinige; P.Selvi Rajendran, “A Novel Music Genre Classification Using Convolutional Neural Network,” in 2021 6th International Conference on Communication and Electronics Systems (ICCES).

[8] Shajin Prince; Justin Jojy Thomas; Sharon Jostana J; Kakarla Preethi Priya; J Joshua Daniel, "Music Genre Classification using Deep learning- A review," in 6th International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS)- 2022.

[9] Sung-Hyun Cho; Yechan Park; Jaesung Lee, "Effective Music Genre Classification using Late Fusion Convolutional Neural Network with Multiple Spectral Features" 2022, IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia).

[10] Laisha Wadhwa; Prerana Mukherjee, Music genre classification using multi-modal deep learning based fusion, in 2021 Grace Hopper Celebration India (GHCI).

[11] Hui Yuan; Wenjia Zheng; Yun Song; Yijun Zhaol, Parallel Deep Neural Networks for Musical Genre Classification, in 2021 IEEE 45th Annual Computers, Software, and Applications Conference (COMPSAC).

[12] Anirudh Ghildiyal and Sachin Sharma, Music Genre Classification Using Data Filtering Algorithm: An Artificial Intelligence Approach, in 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA).

[13] P. Devaki; Adarsh Sivanandan; Raj Senthil Kumar; Mohsin Zahoor Peer, "Music Genre Classification and Isolation," 2021 International Conference on Advancements in Electrical, Electronics, Communication, Computing and Automation (ICA ECA).

[14] Manuel Theodore Leleuly; P. H. Gunawan, Analysis of Feature Correlation for Music Genre Classification, in 2020 8th International Conference on Information and Communication Technology (IcoICT).

[15] Davis Motswedi; Ritesh Ajoodha, “Music Classification Using Fourier Transform and Support Vector Machines”, in 2022 International Conference on Engineering and Emerging Technologies (ICEET).