Journal of Current Research in Engineering and Science
Bi-Annual Online Journal (ISSN: 2581 - 611X)
JCRES

VOLUME:8, ISSUE:2, AUGUST 2025, PAPER NUMBER:49

Differentiate Music Genre From An Audio Files

Monica B, Vishnu Priya R, Pavithra M R

[1,2] Students,and [3] Faculty
Dept. of Information Technology Engineering,
Bannari Amman Institute of Technology ,
Erode, India.
monica.it21@bitsathy.ac.in
vishnupriya.it21@bitsathy.ac.in
pavithra.it21@bitsathy.ac.in

Abstract: — This paper presents a comprehensive approach to music genre classification using machine learning techniques applied to audio signals. We focus on extracting key audio features like Mel-frequency cepstral coefficients (MFCCs), chroma vectors, spectral contrast, and Tonnetz, which are then used to train several machine learning models. The classification task is performed using models such as Support Vector Machines (SVM), Random Forest (RF), and Convolutional Neural Networks (CNN). Through extensive experimentation on datasets like GTZAN, Free Music Archive (FMA), and a custom-curated dataset, we show that deep learning models, particularly CNNs, achieve superior accuracy of 85% compared to traditional models. This study contributes to the field of Music Information Retrieval (MIR) by proposing a robust framework for automatic genre classification.

Keywords — Music genre classification, machine learning, deep learning, convolutional neural network (CNN), audio signal processing, feature extraction, MFCC, chroma vectors

1. INTRODUCTION:

In recent years, the rapid expansion of digital music databases has necessitated the development of automated systems for organizing and retrieving music based on genre. Traditionally, genre classification was manually performed by experts in the music industry, but as the volume of music has grown exponentially, manual classification has become inefficient. Music genre classification is crucial for music recommendation systems, automated tagging, and content-based retrieval. Machine learning techniques have gained prominence in solving classification problems, and this has extended to music genre classification. The problem can be framed as a supervised learning task where the input consists of audio signals and the output is a predicted genre label (e.g., classical, jazz, rock).

Recent advancements in deep learning, particularly in Convolutional Neural Networks (CNNs), have led to significant improvements in audio and music-related tasks due to their ability to automatically extract and learn important features from raw data. This paper presents a framework that combines traditional feature extraction techniques like MFCCs with machine learning models, including SVM, RF, and CNN, to classify music genres from audio files. Our approach is tested on widely used datasets, and the experimental results demonstrate the effectiveness of CNNs in learning complex audio patterns and outperforming traditional models.

2. LITERATURE REVIEW :

A. **Previous Work in Music Genre Classification**
The majority of early music genre classification techniques were manual, using hand-crafted features and conventional machine learning algorithms like Naive Bayes and k-Nearest Neighbours (k-NN). Tzanetakis and Cook's study [1] was a major turning point in this field since it established the groundwork for subsequent research by introducing the use of Mel Frequency Cepstral Coefficients (MFCCs) for genre classification. To improve accuracy, other elements including tempo, rhythm, and harmonic qualities were added throughout time. However, the efficiency of these methods was limited because they frequently had trouble distinguishing between genres with comparable acoustic characteristics.

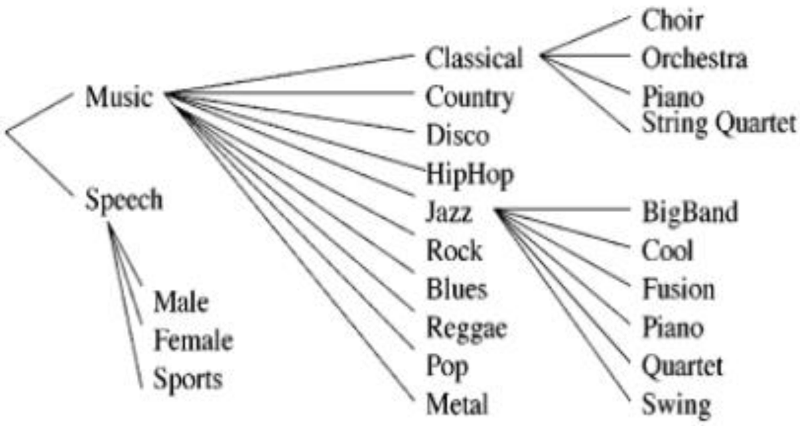
A. **B. Deep Learning in Audio Processing**
Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are at the forefront of recent deep learning breakthroughs that have revolutionized audio processing. CNNs are particularly good at interpreting audio signals' time-frequency representations, handling them like picture spectrograms. This greatly increases classification accuracy by enabling CNNs to recognize hierarchical patterns and extract local information. Several studies demonstrate that deep learning models, such CNNs, perform better than conventional methods in a variety of audio-related tasks, such as genre categorization, music tagging, and speech recognition [2]. These advancements highlight how well contemporary solutions can overcome the drawbacks of earlier approaches.

3.ARCHITECTURE:

The input layer of the deep learning architecture for music genre categorization transforms unprocessed audio data into a time-frequency representation, like spectrograms or Mel Frequency Cepstral Coefficients (MFCCs). A feature extraction layer, usually made up of convolutional neural networks (CNNs), receives this representation. Convolutional filters are used by CNNs to extract low-level information, such as frequency patterns, and pooling layers lower computing complexity and dimensionality.

Higher-level attributes like rhythm, tone, and timbre can be captured by the model using hierarchical feature extraction. An optional temporal analysis layer, which frequently uses Recurrent Neural Networks (RNNs), Gated Recurrent Units (GRUs), or Long Short-Term Memory (LSTM) networks, examines sequential dependencies like tempo and rhythm fluctuations for jobs requiring temporal pattern recognition. A fully connected layer receives the collected characteristics after which non-linear transformations are used to merge the data and get it ready for classification.

Fig. 1. Architecture diagram for Audio classification Hierarchy



2. FLOWCHART:

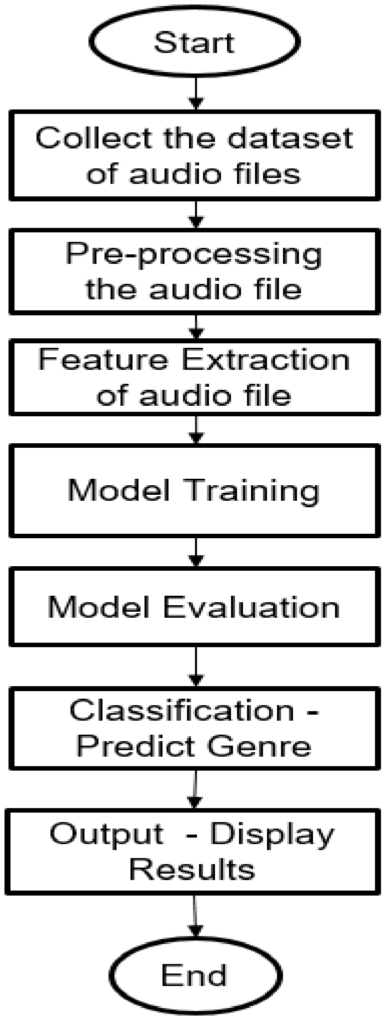


Fig. 2. Flowchart diagram for Music Genre Classification

1. Collect the dataset of audio files:

Gather a diverse set of audio files representing various genres, ensuring the dataset is large enough for effective model training.

2. Pre-processing the audio file:

Clean the audio by removing noise and artifacts, then convert it into a suitable format for feature extraction.

3. Feature Extraction of audio file:

Extract relevant features like MFCCs, chromagrams, and tempo from the audio to represent its genre.

4. Model Training:

Train a machine learning model using the extracted features and genre labels. Common models include SVM, Random Forest, and Neural Networks.

5. Model Evaluation:

Evaluate the model's performance on a validation dataset using metrics such as accuracy, precision, recall, and F1-score.

6. Classification - Predict Genre:

Use the trained model to classify new audio files based on their extracted features.

7. Output - Display Results:

Display the predicted genre for the input audio, often with visualizations for clarity.

8. End:

Conclude the classification process, ready to classify additional audio files as needed.

3. PICTORIAL REPRESENTATION:

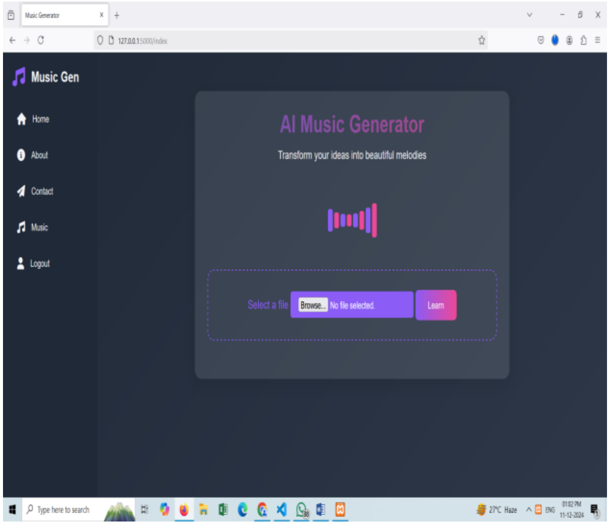


Fig. 3. Home page

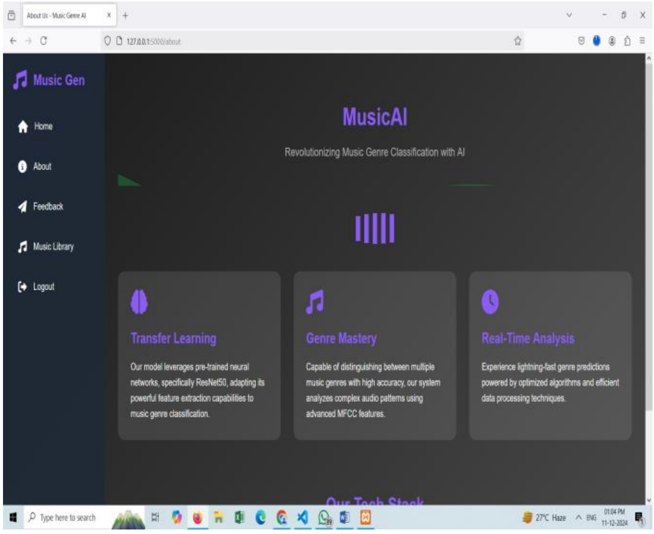


Fig. 4. Client about page

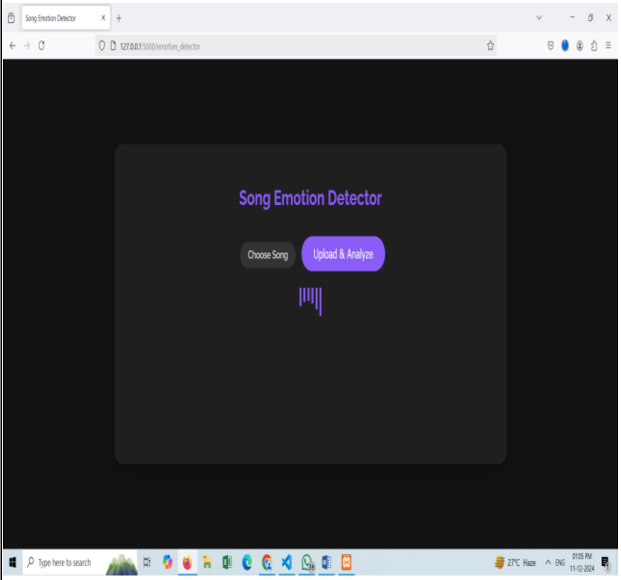


Fig. 6. Analyze the song detector

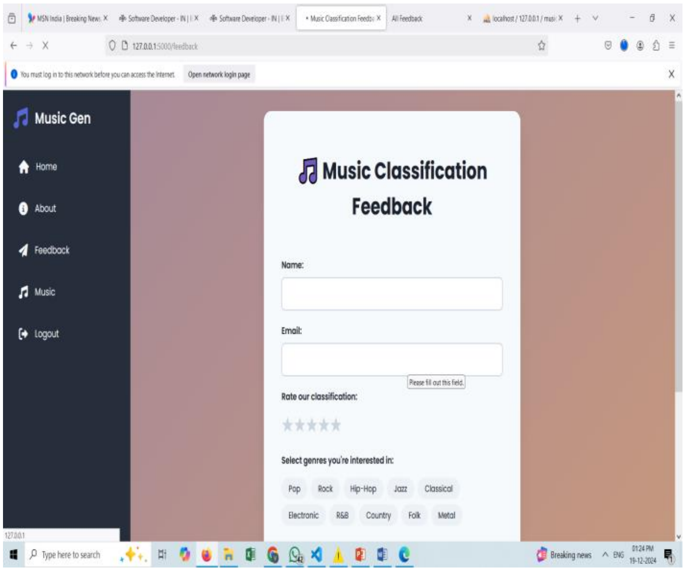


Fig.5. Feedback form

4. METHODOLOGY:

Dataset preparation, audio feature extraction, machine learning model selection, training, and evaluation are some of the crucial steps in the process for your music genre categorization project. Here is a condensed synopsis of every section:

A. Description of the Dataset

The models were trained and assessed using three datasets:The GTZAN Dataset includes 1000 audio tracks (30 seconds each) in 10 different genres, including pop, reggae, rock, hip-hop, jazz, metal, country, disco, hip-hop, blues, and classical. It is a commonly used standard for classifying musical genres.

FMA Dataset: With 8,000 recordings, the FMA-small subset offers a greater range of genres than GTZAN.

B. Extraction of Audio Features

Four different kinds of audio features were taken out of the recordings using the Librosa library:

Mel-frequency Cepstral Coefficients, or MFCCs, are frequently employed for speech and music analysis because they record short-term power spectrum characteristics based on the mel scale.

Chroma Features: Helps with melodic and harmonic analysis by representing the intensity of 12 pitch classes.

Spectral contrast is a valuable tool for determining timbral variations between genres since it quantifies the difference between peaks and valleys in the audio spectrum.

Tonnetz: This tool captures the tonal relationships between pitches, which is especially useful for musical genres like jazz and classical that have a lot of harmonic content.

C. Models for Machine Learning

Three distinct models were used to categorize the genres of music:

A supervised approach that works well with smaller datasets is Support Vector Machines (SVM). Non-linear data interactions were handled using the Radial Basis Function (RBF) kernel.

An ensemble learning technique called Random Forest (RF) combines predictions from several decision trees. It performs effectively with noisy, high-dimensional data. An architecture for deep learning intended for sequential data, such as audio, is the convolutional neural network (CNN). Here, the 1D CNN involved:

Input Layer: Handles feature matrices (MFCCs, for example).

Local patterns in time-series data are detected by convolutional layers. Pooling Layers: Lowers computational complexity and dimensionality.The genre prediction is output by the fully connected layer.

D. Training and Adjusting Hyperparameters

Cross-checking: To evaluate the generalizability of the model, 5-fold cross-validation was employed, guaranteeing that each fold had a fair representation of genres.RF (tree depth), SVM (kernel parameters), and CNN (number of layers, filter sizes, learning rate) hyperparameters were all adjusted using a grid search.

E. Metrics for Evaluation

A number of metrics were computed in order to assess the models:

Accuracy: The proportion of cases that are accurately classified.

Precision: The percentage of accurate positive forecasts for every genre.

Recall: The percentage of true genre occurrences that were accurately forecasted.

F1-Score: A useful metric for assessing performance, particularly on unbalanced dataset

5. CERTIFICATE

6th INTERNATIONAL CONFERENCE ON ENGINEERING & ADVANCEMENT IN TECHNOLOGY 2024

ORGANIZED BY

MALLA REDDY COLLEGE OF ENGINEERING, INDIA

IN ASSOCIATION WITH

ORGANIZATION OF SCIENCE & INNOVATIVE ENGINEERING TECHNOLOGY

IN COLLABORATION WITH

MANIPAL UNIVERSITY COLLEGE, MALAYSIA

SAMARKAND STATE UNIVERSITY, UZBEKISTAN

Certificate Of Presentation

This is to certify that Mr/Mrs/Dr......Pavithra M R......from ......Bannari Amman Institute of Technology......has presented a paper title ......Differentiate Music Genre From An Audio Files...... in the "6th International Conference on Engineering and Advancement in Technology" held on 27th & 28th September 2024 at Malla Reddy College Of Engineering, Secunderabad, India.

Dr. Antony V.Samrot, Director (Research, Innovation and Postgraduate Studies), Manipal University College, Malaysia

Dr. Akhatov Akmal Rustamovich, Vice Rector of International Affairs, Samarkand State University, Uzbekistan

Dr. Christo Ananth, Professor, Dept of IT, Samarkand State University, Uzbekistan

Dr. M. Ashok, Principal, MRCE, Conference Chair & Convener ICEAT 2024

Dr. K. Sivakumar, Advisor, OSIET

K.Janani, M.Tech., CEO, OSIET

6th INTERNATIONAL CONFERENCE ON ENGINEERING & ADVANCEMENT IN TECHNOLOGY 2024

ORGANIZED BY

MALLA REDDY COLLEGE OF ENGINEERING, INDIA

IN ASSOCIATION WITH

ORGANIZATION OF SCIENCE & INNOVATIVE ENGINEERING TECHNOLOGY

IN COLLABORATION WITH

MANIPAL UNIVERSITY COLLEGE, MALAYSIA

SAMARKAND STATE UNIVERSITY, UZBEKISTAN

Certificate Of Presentation

This is to certify that Mr/Mrs/Dr......Monica B......from ......Bannari Amman Institute of Technology......has presented a paper title ......Differentiate Music Genre From An Audio Files...... in the "6th International Conference on Engineering and Advancement in Technology" held on 27th & 28th September 2024 at Malla Reddy College Of Engineering, Secunderabad, India.

Dr. Antony V.Samrot, Director (Research, Innovation and Postgraduate Studies), Manipal University College, Malaysia

Dr. Akhatov Akmal Rustamovich, Vice Rector of International Affairs, Samarkand State University, Uzbekistan

Dr. Christo Ananth, Professor, Dept of IT, Samarkand State University, Uzbekistan

Dr. M. Ashok, Principal, MRCE, Conference Chair & Convener ICEAT 2024

Dr. K. Sivakumar, Advisor, OSIET

K.Janani, M.Tech., CEO, OSIET

Fig.9. certificates

6. CONCLUSION:

The paper is prepared for the template after the text modification is finished. Make a copy of the template file using the Save As command, then name your paper according to the convention the conference has specified. Highlight every item in this freshly generated file, then import your prepared text file. You can now utilize the scroll down window located on the left side of the MSWordFormatting toolbar to style your paper.

7. REFERENCE:

[1] F. Pachet and D. Cazaly, "A classification of musical genre," in Proc. RIAO Content-Based Multimedia Information Access Conf., Paris, France, Mar. 2000.

[2] S. Davis and P. Mermelstein, "Experiments in syllable-based recognition of continuous speech," IEEE Trans. Acoust., Speech, Signal Processing, vol. 28, pp. 357–366, Aug. 1980.

[3] J. Saunders, "Real time discrimination of broadcast speech/music," in Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), 1996, pp. 993–996.

[4] E. Scheirer and M. Slaney, "Construction and evaluation of a robust multifeature speech/music discriminator," in Proc. Int. Conf. Acoustics, Speech, Signal Processing (ICASSP), 1997, pp. 1331–1334

[5] T. Zhang and J. Kuo, "Audio content analysis for online audiovisual data segmentation and classification," Trans. Speech Audio Processing, vol. 9, pp. 441–457, May 2001.