Grenze International Journal of Engineering and Technology, January Issue

XAI-Driven Deep Learning Framework for Classifying Indian Music Genres

S. Saranya[1], S.Sangeetha Mariammal[2], Revathi Manoharan[3], Anu Prabhakar[4], Arun Balaji A S[5] and Raahul MJ[6]
[1-4]Assistant Professor, Department of Computer Science and Engineering, Coimbatore Institute of Technology, Coimbatore, India
Email: saranya.s@cit.edu.in, ssangeetha@cit.edu.in, revathimanoharan@cit.edu.in, anuprabhakar@cit.edu.in
[5,6]B. E Student, Department of Computer Science and Engineering, Coimbatore Institute of Technology, Coimbatore, India
Email: arunbalaji2021@gmail.com, mjraahul2004@gmail.com

***Abstract*— In the digital era, vast online music repositories demand efficient classification systems to organize and recommend tracks effectively. Traditional rule-based and statistical methods often fail to capture the intricate tonal and rhythmic patterns, especially in diverse forms of Indian music. This project presents a Deep Learning-based Music Genre Classification (MGC) system enhanced with Explainable Artificial Intelligence (XAI) techniques. The system employs Mel-Frequency Cepstral Coefficients (MFCCs) to extract meaningful frequency-domain features from audio signals, which are then processed by an Artificial Neural Network (ANN) for genre prediction. Five major Indian genres—Bollypop, Carnatic, Ghazal, Semi-Classical, and Sufi—are classified using this model. To enhance interpretability, LIME, SHAP, and Permutation Importance methods are integrated, providing insights into the model's decision-making process and feature contributions. Experimental results demonstrate that the proposed ANN model achieves 99% training accuracy and 81% validation accuracy, outperforming CNN, RNN, LSTM, and hybrid architectures used for comparison. Beyond its technical significance, this system contributes to the preservation and accessibility of Indian musical heritage, supporting applications in music recommendation, digital archiving, and intelligent content organization.**

***Index Terms*— Indian Music, Genre Classification, Deep Learning, ANN, MFCC, Explainable AI, LIME and SHAP.**

I. INTRODUCTION

Music has been an integral part of human civilization, representing cultural identity, emotional expression, and artistic creativity. In the digital era, the exponential growth of online music platforms has resulted in massive repositories of audio data, creating the need for automated and accurate genre classification systems. Efficient categorization not only enhances the accessibility and recommendation of songs but also aids in preserving the cultural diversity embedded in traditional and contemporary music.

Traditional music genre classification methods rely heavily on manual feature engineering and rule-based approaches using features such as pitch, tempo, and rhythm. However, these methods often struggle to capture the complex and nonlinear patterns inherent in audio signals, particularly in Indian music, which is characterized by intricate ragas, rhythmic variations, and diverse instrumental compositions.

Grenze ID: 01.GIJET.12.1.89_1

To address these limitations, recent advancements in deep learning have enabled models to automatically learn discriminative audio features directly from raw or preprocessed signals, significantly improving classification accuracy and scalability.

This paper presents a Deep Learning-based Artificial Neural Network (ANN) model for Indian Music Genre Classification, integrated with Explainable Artificial Intelligence (XAI) techniques to enhance interpretability. The system preprocesses audio signals to extract robust features such as Mel-Frequency Cepstral Coefficients (MFCCs), spectrograms, and amplitude envelopes, which are then used as input to the ANN for genre prediction. The model effectively classifies multiple Indian music genres, including Bollypop, Carnatic, Ghazal, Film-based, and Semi-Classical categories.

To promote model transparency, the proposed framework incorporates Local Interpretable Model-agnostic Explanations (LIME) and SHapley Additive exPlanations (SHAP) for local and global interpretation of feature importance. These techniques provide insights into how specific audio features contribute to genre decisions, improving the explainability and trustworthiness of the model. Experimental results demonstrate that the proposed ANN model achieves 99% training accuracy and 81% validation accuracy, outperforming CNN, RNN, LSTM, and hybrid architectures used for comparison.

The remainder of this paper is organized as follows: Section II reviews related research on music genre classification and explainable AI techniques. Section III describes the proposed system architecture and feature extraction process. Section IV outlines the implementation details and experimental setup of the model. Section V presents the results, and comparative performance analysis. Finally, Section VI concludes the paper and highlights potential directions for future work.

II. Literature Review

Music genre classification has gained significant attention in recent years due to the increasing volume of digital audio data and the growing demand for intelligent music retrieval systems. Numerous studies have explored machine learning and deep learning techniques for automatic genre identification, focusing on improving both classification accuracy and interpretability [13]. Early approaches relied heavily on traditional machine learning techniques such as Support Vector Machines (SVM), K-Nearest Neighbor (KNN), and Decision Trees, primarily using handcrafted audio features like Mel-frequency cepstral coefficients (MFCCs), chroma features, and spectrograms [11, 14]. For instance, Prajwal et al. [7] utilized SVM models on the GTZAN dataset for music genre classification, achieving moderate accuracy; however, their approach was limited in performance compared to deep learning models and suffered from overfitting due to the small dataset. Similarly, Siddiquee et al. [4] developed a KNN-based classifier using Mel spectrogram features, but their study was restricted to a single dataset, limiting the generalizability of their findings. Rajeeva Shreedhara Bhat et al. [8] applied machine learning to log-based Mel spectrograms, achieving high training accuracy but demonstrating a large gap with validation accuracy, highlighting the issue of overfitting in conventional ML approaches.

With the rise of deep learning, Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), including Long Short-Term Memory (LSTM) and Bidirectional LSTM (BiLSTM), have been increasingly employed to automatically extract and learn hierarchical audio features from raw or transformed audio signals [12,15]. Chatterjee et al. [5] compared CNN-based architectures with traditional classifiers, using wavelet and spectrogram features, and reported superior performance of CNNs, illustrating their ability to capture complex time-frequency patterns. Yinhui Yi et al. [3] proposed an LSTM-based method that integrates both time-domain and frequency-domain features such as MFCCs, showing promising results but highlighting the need for validation on larger and more diverse datasets. Similarly, Wijaya et al. [6] used BiLSTM networks fed with MFCC features and achieved high accuracy on the GTZAN dataset; however, LSTM architectures often struggle with highly correlated sequences typical of music data and lack inherent interpretability.

Recent work has also focused on enhancing model interpretability and efficiency. Jenifera et al. [1] introduced a lightweight CNN combined with Explainable AI (XAI) techniques such as LIME, integrating multiple features including amplitude envelope, crest factor, spectrogram, and spectral centroid. Their method demonstrated effective classification with interpretable outputs, although it was only evaluated on GTZAN. Kumari et al. [2] extended the focus to Indian music genres (Carnatic, Ghazal, Semi-classical, Sufi, Bolly-pop) alongside Western genres, using CNN and KNN models with MFCC and spectrogram features; limitations included small dataset size and insufficient genre diversity, suggesting that performance could be improved with larger and more balanced datasets. Zhang [10] implemented a CNN-based feature extraction model achieving 87% accuracy, indicating that architectural complexity and feature selection significantly influence performance. Lau and

Ajoodha [9] conducted a comparative study of CNNs and traditional classifiers using spectrogram and other content-based features, but again, the evaluation was restricted to GTZAN, limiting broader applicability.
The challenges identified in the literature study are:

- over-reliance on small and imbalanced datasets such as GTZAN, which limits generalizability.
- overfitting due to limited data diversity
- Inadequate evaluation on non-Western or diverse music genres
- Difficulties in capturing long-term temporal dependencies in sequential music data
- Lack of interpretability in deep learning models.

Addressing these limitations requires the development of models capable of handling large-scale and diverse datasets, integrating hybrid time-frequency feature representations, and employing explainable AI methods to improve transparency and trustworthiness of classification outcomes.

III. Proposed Methodology

The architecture of the proposed music genre classification system is designed to enable automated, accurate, and interpretable predictions using deep learning and explainable AI techniques. The system pipeline begins with raw audio input, followed by feature extraction, model training, and interpretability analysis.

A. System Architecture

The proposed system combines robust audio feature extraction with deep learning-based classification and explainable AI techniques. MFCC features capture the essential spectral and timbral characteristics of music tracks, while the ANN learns complex patterns for accurate genre classification. The inclusion of LIME, SHAP, and permutation importance ensures transparency and interpretability, making the system a reliable solution for automated music genre recognition and recommendation applications.
As shown in Fig. 1, the system first preprocesses the audio data and extracts relevant features, which are then fed into an Artificial Neural Network (ANN) for multiclass genre classification. To enhance interpretability, LIME, SHAP, and permutation importance methods are integrated to analyze the contribution of individual features to the model's predictions.

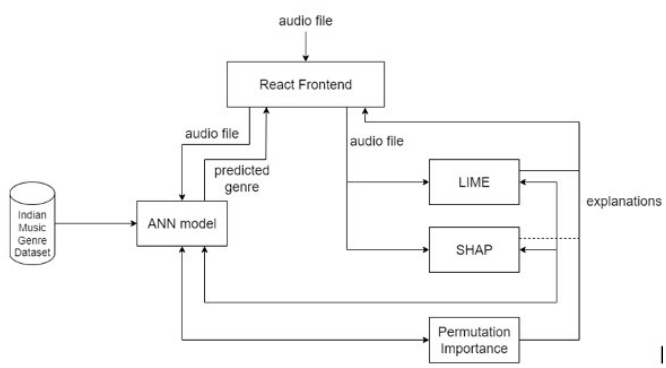

Figure 1. System Architecture

The system uses Mel-Frequency Cepstral Coefficients (MFCCs) to capture both spectral and timbral characteristics of music. For each track, the mean and standard deviation of 40 MFCC coefficients are computed, resulting in an 80-dimensional feature vector representing the audio in a structured numerical format. These feature vectors serve as input to the ANN, enabling the network to learn complex patterns corresponding to different music genres.

B. Dataset

The Indian Music Genre Dataset is used to train and evaluate the system. It contains a representative collection of traditional and contemporary Indian music, covering five major genres: Carnatic, Ghazal, Semi-classical, Sufi, and Bollypop. Key characteristics of the dataset include:

- Balanced samples: Each genre includes labeled audio tracks for consistent learning.
- Feature-rich content: Both time-domain and frequency-domain features are extracted from audio files.
- Structured labels: Each track is assigned a single genre label for supervised multiclass classification.
- Train-test split: The dataset is divided into training and testing subsets to ensure fair performance evaluation.

C. Feature Extraction

Raw audio signals are first converted to mono and sampled at 22,050 Hz using Librosa to ensure uniformity. The feature extraction process aims to capture frequency-domain characteristics relevant for distinguishing music genres.

The Steps involved are:

- Audio Loading: Convert audio files to mono and standardize sampling rate.
- MFCC Extraction: Compute the mean and standard deviation of 40 MFCC coefficients for each track, forming a fixed-size feature vector.
- Visualization and Saving: Plot and save extracted features as images using Matplotlib for analysis and documentation.

D. Deep Learning Model (ANN)

The proposed system employs an ANN to classify music genres based on MFCC-derived features. The architecture consists of:

- Input Layer: Accepts 80-dimensional feature vectors representing mean and standard deviation of MFCCs.
- Hidden Dense Layers: Two fully connected layers with 256 and 128 neurons, respectively, using ReLU activation to learn non-linear patterns.
- Dropout Layers: 40% dropout is applied to prevent overfitting.
- Output Layer: Fully connected layer followed by Softmax activation to predict probabilities for each genre class.

This ANN effectively captures complex relationships in the audio data, offering superior accuracy and generalization compared to traditional machine learning methods.

E. Explainable AI Techniques

To enhance model interpretability, three model-agnostic techniques are integrated:

- LIME (Local Interpretable Model-agnostic Explanations): Provides local explanations for individual predictions and highlights the most influential features for a given track.
- SHAP (SHapley Additive exPlanations): Computes feature contributions to individual predictions, indicating positive or negative influence of each feature on the output.
- Permutation Importance: Measures global and local feature importance by randomly permuting feature values and observing the resulting drop in model performance.

IV. Experimental Setup

The system was implemented on a standard computing platform equipped with Intel Core i5/i7 or AMD Ryzen 5/7 processors, 8–16 GB of RAM, and 256–512 GB SSD storage to ensure sufficient computational capacity for training and evaluation. The software environment consists of Python 3.8+, with libraries and frameworks such as TensorFlow, Keras, NumPy, Pandas, Librosa, and Matplotlib. The system was developed and tested using Google Colab and Visual Studio Code.

The workflow begins with dataset acquisition and preparation. The Indian Music Genre Dataset, comprising 500 audio files evenly distributed across the five genres, was chosen for its diverse representation of Indian music. Each audio file was resampled to 22,050 Hz and converted to mono to standardize the inputs, reduce computational complexity, and remove stereo bias. The preprocessed files were organized into genre-specific directories to facilitate structured access for training and evaluation.

Feature extraction was performed using Mel-Frequency Cepstral Coefficients (MFCCs), which capture both spectral and timbral characteristics of music. The mean and standard deviation of 40 MFCC coefficients were calculated for each track, resulting in 80-dimensional feature vectors. These vectors were standardized using a StandardScaler and divided into training (80%) and validation (20%) sets.

The ANN architecture consists of an input layer that receives the 80-dimensional MFCC vectors, two hidden dense layers with 256 and 128 neurons respectively, each using ReLU activation, and dropout layers with a rate of 40% to prevent overfitting. The output layer comprises five neurons with Softmax activation corresponding to the target music genres.

The model was trained using the Adam optimizer and sparse categorical cross-entropy loss while monitoring validation performance. This design ensures a robust, end-to-end pipeline capable of accurately classifying music genres while supporting interpretability through explainable AI methods.

V. Result Analysis

The experimental evaluation focused on the system's ability to classify audio tracks into five Indian music genres. Performance was primarily measured using accuracy, supplemented by confusion matrices and classification reports for deeper insight into model behavior.

TABLE I. RESULTS OF SPECTROGRAM FEATURES

Model	Training Accuracy	Validation Accuracy
CNN	79%	64%
RNN	21%	20%
LSTM	26%	20%
BiLSTM	22%	20%
CNN-GRU	52%	46%
CNN-BiLSTM	81%	59%
ANN	17%	20%

Initial experiments using spectrogram-based features demonstrated that hybrid CNN models performed better than recurrent models, while ANN and RNN architectures struggled to capture relevant spectral information. As shown in Table I, the CNN-BiLSTM model achieved a training accuracy of 81% and a validation accuracy of 59%, whereas the ANN achieved only 17% training and 20% validation accuracy. These results indicate that spectrogram features alone are insufficient for robust genre classification using traditional ANN architectures.

In contrast, experiments using MFCC-based features significantly improved ANN performance, is shown in Table II. The ANN model achieved a training accuracy of 99% and a validation accuracy of 81%, outperforming other architectures including CNN, RNN, LSTM, BiLSTM, and hybrid models. MFCC features effectively capture the timbral and spectral characteristics of music, enabling the ANN to learn complex patterns distinguishing the five genres. The results confirm that MFCC-based feature representation is more suitable than spectrogram-based features for Indian music genre classification, and that the proposed ANN design provides accurate and reliable predictions.

TABLE II. RESULTS OF MFCC FEATURES

Model	Training Accuracy	Validation Accuracy
CNN	98%	57%
RNN	99%	26%
LSTM	99%	56%
BiLSTM	99%	58%
CNN-GRU	85%	60%
CNN- BiLSTM	99%	58%
ANN	99%	81%

Overall, these findings demonstrate that careful preprocessing, feature extraction, and model design are critical to achieving high performance in music genre classification. The modular design of the system allows for flexibility in integrating different feature representations or experimenting with alternative network architectures, while maintaining robust performance.

The integration of these explainable AI techniques allows for a better understanding of the ANN's decision-making process. LIME and SHAP confirm that specific MFCC coefficients strongly influence genre classification at the track level, while permutation importance identifies which features have the greatest impact on overall model performance. Collectively, these methods ensure that the proposed system is not only accurate but also interpretable, increasing trust in its predictions and supporting its deployment for automated music analysis and recommendation systems.

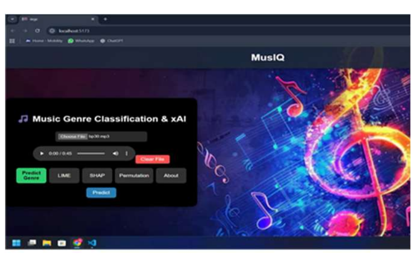

Figure 2. MusIQ Web Application Interface

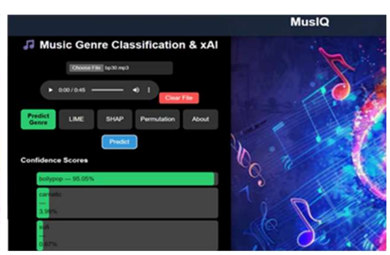

Figure 3. Predicted Genre and Confidence Scores

Fig.2 illustrates the main user interface of the MusIQ web application, which is designed for music genre classification and explainable artificial intelligence. The system allows users to upload audio files, providing a structured interface for submitting music tracks for analysis. Once an audio file is uploaded, the predicted genre and associated confidence scores are displayed, as depicted in Fig.3. The prediction results appear after the user clicks the "Predict" button, offering an immediate and interpretable output. To enhance transparency, the system includes explainable AI functionalities.

Figure 4. Explanation by LIME

Figure 5. SHAP Global Feature Importance Plot

Fig. 4 presents the LIME-based explanation for an uploaded audio track, highlighting the contribution of individual features to the specific prediction. Similarly, Fig.5 shows the SHAP global feature importance plot, which visualizes the overall influence of each feature on the model's classification decisions. Finally, Fig.6 displays the permutation importance analysis, estimating the relevance of each feature by measuring the decrease in model performance when feature values are randomly shuffled. Together, these interface components and interpretability modules provide a user-friendly and informative environment for both music genre classification and model explanation.

Figure 6. Permutation Importance for Model Features

VI. Conclusion and Future Work

A deep learning-based Music Genre Classification System is developed to automatically identify and categorize Indian music tracks into five distinct genres: Carnatic, Ghazal, Semi-Classical, Sufi, and Bollypop. The system leverages a structured pipeline that includes preprocessing of raw audio, extraction of Mel-Frequency Cepstral Coefficients (MFCCs) to capture spectral and timbral characteristics, and classification using an Artificial Neural Network (ANN) with fully connected dense layers and ReLU activations. Dropout layers were employed to prevent overfitting, allowing the model to generalize effectively across diverse tracks. Experimental evaluation demonstrated that the proposed system achieves high accuracy and reliable genre predictions, highlighting its potential for integration into automated music analysis, recommendation engines, and digital music platforms. Furthermore, the integration of explainable AI techniques such as LIME, SHAP, and permutation importance

provides transparency, allowing users to interpret feature contributions and understand the decision-making process of the ANN. This combination of accuracy, robustness, and interpretability confirms the practical viability of deploying deep learning models in music genre classification.
For future work, improvements can focus on enhancing dataset diversity, addressing class imbalance through data augmentation, optimizing model hyper parameters, and enabling deployment in real-time environments such as cloud or edge computing platforms. These enhancements aim to improve generalization, scalability, and applicability of the system across broader music collections.

REFERENCES

[1] E. Jenifera, K. S. Abirami, and M. Rajeshwari, “Deep learning-based method combining time/frequency domain features with Explainable AI (LIME) on a lightweight CNN model for music genre classification,” 2025.
[2] N. Kumari, T. Shukla, S. Kumar, S. Balachandra, “Automatic classification of Indian and Western music genres using MFCC and Spectrogram features with KNN/CNN,” 2021.
[3] Y. Yi, X. Zhu, Y. Yue, and W. Wang, “Music genre classification using LSTM to combine time-domain and frequency-domain (MFCC) features,” 2021.
[4] M. N. A. Siddiquee, M. A. Hossain, and F. Wahida, “Machine learning classifier for automatic music genre prediction using Mel spectrogram features and KNN,” 2023.
[5] R. Chatterjee, K. Agarwal, H. Bajari, R. K. Ghosh, and S. Pramanik, “Deep learning-based music genre classification comparing CNN with traditional classifiers using wavelet and spectrogram features,” 2024.
[6] Wijaya et al., “Genre classification using BiLSTM with MFCC features,” 2022.
[7] P. R., S. Sharma, and P. Naik, “Machine learning approaches for music genre classification on the GTZAN dataset,” 2021.
[8] R. Shreedhara Bhat, R. B. Rohith, and M. K. R., “Machine learning techniques on log-based MEL spectrograms for music genre classification,” 2020.
[9] D. Lau and R. Ajoodha, “Comparison of CNN and traditional classifiers for music genre classification using spectrogram and content-based features,” 2020.
[10] J. Zhang, “Music feature extraction and classification using convolutional neural networks,” 2021.
[11] Z. Liu, "Audio feature extraction and classification technology based on convolutional neural network," J. Electrical Systems, vol. 20, no. 9s, pp. 1425–1431, 2024.
[12] K. K. Jena, S. K. Bhoi, S. Mohapatra, and S. Bakshi, "A hybrid deep learning approach for classification of music genres using wavelet and spectrogram analysis," Neural Computing and Applications, vol. 35, no. 15, pp. 11223–11248, 2023.
[13] R. Sharma and Nisha, "Classification of music genres using neural network," in 11th Int. Conf. System Modeling & Advancement in Research Trends (SMART), 2022, pp. 142–147, doi: 10.1109/SMART55829.2022.10046811.
[14] N. Babu, P. Kumar, J. Mathew, and U. Satija, "Exploration of bonafide and spoofed audio classification using machine learning models," in IEEE 19th India Council Int. Conf., 2022.
[15] G. Tzanetakis and P. Cook, "Musical genre classification of audio signals," IEEE Trans. Speech and Audio Processing, vol. 10, no. 5, pp. 293–302, 2002, doi: 10.1109/TSA.2002.800560.