I. J. Engineering and Manufacturing, 2026, 3, 394-411
Published Online on June 8, 2026 by MECS Press (http://www.mecs-press.org/)
DOI: 10.5815/ijem.2026.03.24

Hybrid Deep Learning-Based Automated Genre Classification of Assamese Regional Songs

Spandan Kumar Barthakur
Department of Information Technology, Gauhati University, India
Department of Computer Science and Engineering, The Assam Royal Global University, India
E-mail: spandankumarb@gmail.com
ORCID iD: https://orcid.org/ 0009-0000-6138-2629

Parismita Sarma
Department of Information Technology, Gauhati University, India
E-mail: pari@gauhati.ac.in
ORCID iD: https://orcid.org/0000-0003-1904-820X

Maharshi Nath
Software Developer, MARLN Corporation, India
E-mail: maharshinath1@gmail.com
ORCID iD: https://orcid.org/0009-0006-7059-2381

Daiyaan Ahmed
Software Developer, National Informatics Centre(NIC), Dispur, Assam, India
E-mail: daiyaanahmedrgu@gmail.com
ORCID iD: https://orcid.org/0009-0003-6691-2117

Hirak Jyoti Hazarika
Royal School of Library and Information Science, The Assam Royal Global University, Guwahati, Assam, India
E-mail: drhirakjyotihazarika@gmail.com
ORCID iD: https://orcid.org/0000-0002-0023-7097

Bikash Baruah*
Department of Computer Science and Engineering, The Assam Royal Global University, India
E-mail: baruahbikash9@gmail.com
ORCID iD: https://orcid.org/0000-0001-7485-6897
*Corresponding Author

Received: 24 February, 2026; Revised: 13 April, 2026; Accepted: 20 May, 2026; Published: 08 June, 2026

Abstract: This work aims to preserve and promote the rich musical heritage of Assam by developing an automated classification system for Assamese regional songs using a hybrid deep learning approach. This method not only modernizes the preservation of traditional music but also enhances its accessibility to a global audience for integrating AI with cultural conservation. Five genres of Assamese songs—Bihu, Kamrupiya Lokageet, Goalporiya Lokageet, Borgeet, and Naam—are considered in this study. By leveraging Convolutional Neural Networks (CNNs) and advanced audio feature extraction techniques such as Mel-Frequency Cepstral Coefficients (MFCCs) and spectrograms, a hybrid model combining VGG16 and ResNet50 is developed. This fusion utilizes the strengths of both architectures, enhancing the model's performance and accuracy. Following the process, it is observed that two distinctly different genres, Bihu and Borgeet, are accurately categorized by the proposed model, while the remaining three show slight labeling inconsistencies.

Index Terms: Assamese Regional Songs, Convolutional Neural Networks, ResNet50, VGG16, LSTM, Bihu, Kamrupiya Lokageet, Goalporiya Lokageet, Borgeet, Naam

This work is open access and licensed under the Creative Commons CC BY 4.0 License.

1. Introduction

The classification of folk songs is a challenging yet significant task, particularly for underrepresented languages like Assamese[1]. Folk music, being an integral part of a community's cultural identity, requires effective classification techniques to ensure its proper documentation and preservation[2]. Assamese folk songs, rich in cultural heritage and diverse in style, require automated systems to preserve and categorize them efficiently[3]. Traditional methods for genre classification rely heavily on handcrafted features and conventional machine learning models, which often struggle with the complexity and variability inherent in folk music[4]. Recently, deep learning [29,30] based approaches have demonstrated superior performance in various audio classification tasks due to their ability to learn intricate patterns directly from raw data [5].

This paper explores a hybrid deep learning-based approach for the automated genre classification of Assamese folk songs[6]. The proposed model integrates convolutional neural networks (CNNs) and recurrent neural networks (RNNs) to capture both spatial and temporal features of the audio signals[7]. The system is trained and evaluated using a curated dataset of Assamese folk songs, ensuring a robust and culturally relevant classification process. [8]. Despite advancements in deep learning for music classification, limited research has been conducted on regional and low-resource languages like Assamese[9]. Existing studies predominantly focus on widely spoken languages with large datasets, leaving a gap in effective genre classification methods for Assamese folk music[10]. Additionally, the lack of publicly available datasets and standardized benchmarks for Assamese folk songs further hinders progress in this domain[11].

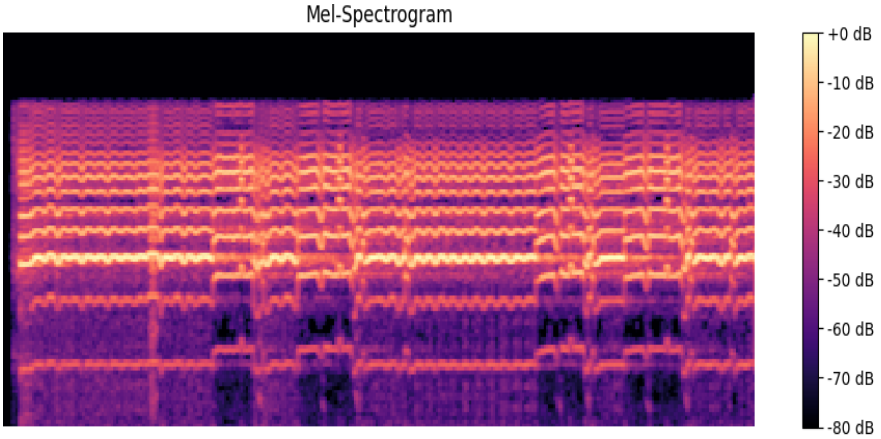


Fig. 1. Sample spectrogram of a bihu audio file

Table 1. Comprehensive study of different traditional songs of Assam

Category	Description	Themes / Focus	Musical Instruments used	Performance Context
Bihu	A vibrant folk festival and musical tradition of Assam, celebrated with dance, music, and rituals marking seasonal changes. The three main Bihu festivals—Bohag, Magh, and Kati—mark spring, harvest, and prosperity.	Love, nature, youth, harvest, joy of life	Energetic rhythms; use of Dhol, Pepa, Gogona, Taal, Toka, flute	Performed during Bihu festivals with dance, traditional instruments, and community celebrations
Kamrupiya Lokageet	Folk songs from the Kamrup region reflecting the cultural and social life of the area.	Love, nature, local traditions, social values	Simple melodic patterns; light instrumental support; emphasis on lyrics	Sung in social gatherings and cultural programs with simple musical accompaniment
Goalporiya Lokageet	Folk songs from the Goalpara district blending Assamese and Bengali cultural elements.	Everyday life, love, spirituality	Soft, melodious tunes; instruments like Dotara, flute, and percussion	Performed in traditional settings with a distinctive lyrical style
Borgeet	Classical devotional songs composed by Srimanta Sankardev and Madhavdev in the 15th–16th centuries.	Devotion to Lord Krishna, spirituality	Structured ragas and talas; classical style; accompanied by Khol and Taal	Performed in Vaishnavite monasteries (Satras) and religious ceremonies
Naam	A form of congregational chanting in Assam's Vaishnavite tradition conducted in Namghars (community prayer halls).	Devotion, unity, spiritual bonding	Rhythmic chanting; accompanied by Khol, Taal; collective participation	Collective prayer and chanting during religious gatherings

The primary challenges in this study include the scarcity of high-quality and annotated datasets for Assamese folk songs, the complex and overlapping characteristics of different folk genres, making non- trivial classification, variations in musical instruments, vocal styles, and recording quality that introduce additional noise and inconsistencies in the data, and the requirement for computationally efficient models that can perform well even with limited training data. Fig.1. represents one such characteristic of sample spectrogram. By addressing these challenges, this research aims to contribute towards the development of an efficient and scalable system for Assamese folk song classification [31,32], ultimately aiding in the digital preservation of cultural heritage. Table.1. provides a comprehensive study of those songs.

The process begins by collecting and digitizing a diverse dataset of Assamese regional songs, then extracting key features like Mel-Frequency Cepstral Coefficients (MFCCs) and spectrograms to capture the essential auditory elements[12,13]. A Convolutional Neural Network (CNN) is used to analyze these features, training on the dataset to classify the songs according to their genre[14,15]. This system achieves high accuracy in song classification, and the resulting repository offers significant value for research and education[16]. Scholars in fields such as ethnomusicology and cultural studies can use the repository to deepen their understanding of Assamese music's diversity and evolution[17]. Additionally, by making Assam's musical heritage accessible globally,[23] the project fosters cross-cultural appreciation and can serve as an educational tool in schools and universities, introducing students to the state's unique cultural soundscape[18].

As there isThe purpose of this work is to preserve and promote Assam's diverse musical heritage by utilizing deep learning techniques to classify and categorize its regional songs[19]. Through the development of an automated system that analyzes and identifies the unique characteristics of various Assamese music genres, this project aims to create an accessible repository for scholars, educators, and the global community[20]. By doing so, it facilitates deeper cultural understanding, supports research in ethnomusicology and cultural studies, and fosters cross-cultural appreciation of Assam's rich musical tradition[21].

The novelty of this work lies in its innovative application of deep learning techniques, specifically Convolutional Neural Networks (CNNs)[22], to the preservation and classification of Assamese regional music—a unique cultural domain that has been largely unexplored in the context of AI[23]. By leveraging advanced audio feature extraction methods such as Mel- Frequency Cepstral Coefficients (MFCCs) and spectrograms[24], the project creates a sophisticated system capable of accurately categorizing traditional Assamese songs based on genre and regional characteristics[25].

The main objectives of this research are as follows:

i. Assemble a comprehensive dataset of 5 different categories of Assamese local songs.
ii. Ensure meticulous labeling for accuracy.
iii. Conduct extensive preprocessing of audio data, including feature extraction and normalization
iv. Design and implement a Hybrid Deep Learning model specialized in classification.

This manuscript is organized into five sections. Section 2 presents a comprehensive review of the relevant literature, highlighting previous studies and establishing the theoretical background of the research. Section 3 describes the methodology adopted for the study, including the research design, data collection procedures, and analytical techniques used. Section 4 discusses the results obtained from the analysis and provides an in-depth interpretation of the findings. Finally, Section 5 lists the references cited throughout the manuscript.

2. Literature Survey

A comprehensive review of existing literature was conducted to understand the current advancements in music genre classification, folk music analysis, and deep learning–based audio processing techniques. The reviewed studies primarily focus on feature extraction methods such as MFCC, Mel Spectrograms, and spectral features, combined with machine learning and deep learning algorithms including CNN, SVM, RNN, XGBoost, and Genetic Algorithms. Particular attention has been given to research works addressing traditional and regional music forms, including Assamese folk music. Table.2. presents a structured summary of significant contributions, outlining the algorithms used, key findings, and future research directions.

From the literature summarized in Table.2., it is evident that deep learning models, particularly CNN-based architectures integrated with MFCC and spectrogram features, have demonstrated high classification accuracy across various datasets. Studies related to Assamese folk music and Indian classical forms highlight the growing interest in applying computational techniques to regional and traditional music analysis. However, limitations remain in terms of real-time implementation, broader genre coverage, optimized feature engineering, and domain-specific dataset availability. These research gaps provide strong motivation for further investigation and the development of more robust, scalable, and culturally contextualized music classification models.

Table 2. Summary of Existing Literature on Music Genre Classification, Folk Music Analysis, and Related Deep Learning Approaches

Sl. No	Paper Title	Year	Author Name	Algorithm Used	Result	Future Scope
1	"Optimizing the configuration of deep learning models for music genre classification." Ten Li Academy of Arts, Pingdingshan Polytechnic College, Pingdingshan.	2024	Teng Li	MFCC features with dual CNN and Black Hole Optimization.	Achieved accuracy exceeding 95% in music genre classification.	Improve recommendation systems and explore neuromorphic hardware.
2	"Music Genre Classification: A Comparative Analysis of CNN and XGBoost Approaches with MFCC and Mel Spectrograms." Department of Statistics, University of California.	2024	Yigang Meng	CNN with Mel Spectrograms and XGBoost with MFCC.	XGBoost with MFCC performed better for structured audio features.	Explore additional audio features and broader genre categories.
3	"A computational approach identifying Assamese folk music instruments." Department of Statistics, Dibrugarh University.	2023	Krishnarjun Bora et al.	MFCC + Spectrogram with SVM/CNN.	Successfully identified instruments using feature-based classification.	Use deep learning models like CNN/RNN for improved analysis.
4	"Classification of Assamese Folk Songs Melody using Supervised Learning Techniques."	2023	Bora, K., et al.	DTC, LDA, Random Forest, SVM using MFCC.	SVM achieved highest accuracy (94.17%).	Improve feature engineering and real-time systems.
5	An Approach to Discover Similar Musical Patterns. IEEE Access.	2022	Chakrabarty, S., et al.	Correlation Coefficient using 16 Note-Measure Method.	Identified similarity patterns and developed recommendation concept.	Apply to Indian folk genres and larger datasets.
6	MFF-SAug: Multi feature fusion for speech emotion recognition.	2022	Jothimani, S., and Premalatha.	CNN with spectrogram augmentation.	Achieved up to 99.6% accuracy across datasets.	Real-time deployment and multimodal fusion.
7	Recognition and classification model of music genres based on deep networks.	2021	Xu, Ke.	Cross-Entropy + Adam optimization.	Improved classification accuracy with attention and segmentation.	Extend to traditional instruments and optimized feature sets.
8	Applying NLP and Deep Learning Techniques for Raga Recognition.	2020	Peri, Deepthi.	RNN with hierarchical sequence modeling.	Achieved 90.3%–95.6% accuracy on classical datasets.	Real-time streaming and cross-genre adaptation.
9	Music Classification using Spectral Features and SVM.	2019	Thiruvengatanadhan, R.	SVM with spectral features.	Achieved 89% classification accuracy.	Cloud deployment, recommendation integration, IoT support.
10	GA-based parameterization and feature selection for music genre recognition.	2016	Serwach, Marcin; Stasiak, Bartlomiej	Neural Network + kNN + Genetic Algorithm.	Neural network ~60% accuracy; GA improved feature selection.	Explore subgenre classification and advanced feature engineering.

3. Methodology

In the digital age, Assamese songs, cherished treasures of Assam's cultural heritage, face pressing challenges in preservation. These orally transmitted melodies, steeped in linguistic nuances and melodic richness, encounter hurdles in systematic documentation, leaving manual classification susceptible to errors. This project addresses this critical issue by harnessing the power of deep learning to automate the precise recognition and categorization of various Assamese songs sourced from audio recordings.

3.1. Dataset preparation

The dataset consists of 2500 Assamese regional songs collected from online platforms, classified into five different categories namely Bihu, Borgeet, Kamrupiya Lokageet, Goalporiya Lokageet and Naam where each category consists of 500 songs each. For rigorous model training, the dataset is thoughtfully partitioned: 80% for training, 10% for validation, and 10% for testing. All files are in the .wav format, ensuring uniformity for effective deep learning analysis and reflecting a diverse and representative collection of Assamese musical heritage. This repository serves as a valuable resource for scholars and educators, fostering a deeper understanding and global recognition of Assam's multi diverse musical traditions.

3.2. Preprocessing

In the preprocessing function, audio files are loaded and processed for uniformity and suitability in deep learning analysis. The software Audacity is used to remove the background noise of the different audio files. The function, implemented with the librosa library, standardizes the sampling rate to 16,000 Hz and limits the duration to 30 seconds. To accommodate variations in audio length, padding is applied, creating a consistent matrix structure. This meticulous preprocessing ensures that each audio file is transformed into a format conducive to subsequent model training, enhancing the robustness of the classification system.

3.3. Feature Extraction

The feature extraction process in this project, facilitated by the Librosa library, incorporates several key parameters to ensure the creation of informative and standardized representations of the Assamese Regional songs. A specified sample rate of 16,000 Hz, denoted by target_sampling_rate=16000, establishes a consistent measure of audio signal per second during processing. Additionally, the duration parameter, set to 30 seconds (duration=30), enforces uniformity in the length of audio clips for analysis and model training. Mel-frequency cepstral coefficients (MFCCs), computed with n_mfcc=40, play a pivotal role in capturing the unique spectral features of each audio signal. This amalgamation of parameters contributes to a comprehensive feature set, allowing the deep learning model to discern intricate patterns and nuances within the Assamese Bihu folk songs for accurate classification during subsequent training and testing phases. Audio Library in use is librosa and deep frameworks in use are Keras and Theano.

3.4. Constituent Algorithms: The Hybrid model in use consists of two algorithms which are:

VGG16, originally designed for image classification, can also be adapted for audio processing by leveraging its convolutional neural network (CNN) architecture to extract meaningful features from audio data. In audio applications, the raw audio signal is often transformed into spectrograms—visual representations of sound frequencies over time—so that the CNN can process them similarly to images[26]. VGG16's deep hierarchical structure, with multiple convolutional and pooling layers, is effective in capturing time-frequency patterns and learning abstract representations from these spectrograms. This makes it suitable for tasks like audio classification, music genre detection, and speech recognition. By fine- tuning its pre-trained weights, VGG16 can be optimized for specific audio-based tasks, delivering powerful feature extraction capabilities that enhance model performance in diverse auditory applications.

ResNet50, a deep convolutional neural network known for its residual learning framework, is highly effective in audio processing when applied to spectrograms or other time-frequency representations of audio signals. In audio tasks, ResNet50's skip connections, which bypass certain layers, help address the vanishing gradient problem in deep networks, enabling the model to learn more efficiently from complex audio patterns[27]. By mapping audio data into spectrograms, ResNet50 leverages its residual blocks to capture intricate time-frequency features, making it highly suitable for tasks like audio event detection, music classification, and speech recognition[28]. Fine-tuning ResNet50 on audio data allows it to focus on specific auditory nuances, achieving robust performance across various sound-related applications. Its depth and design make it adept at capturing hierarchical and abstract audio features, enhancing the accuracy of audio classification models.

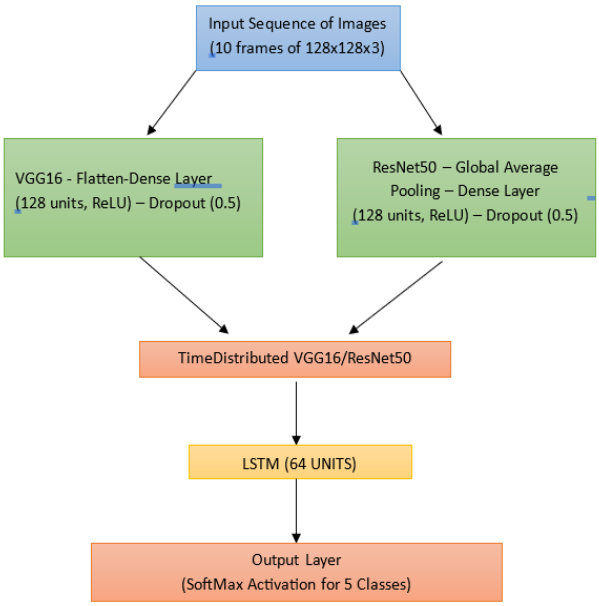


Fig. 2. VGG16 and ResNet50 with LSTM Model Architecture

Fig. 2. shows the model architecture integrates a pre-trained ResNet50 convolutional neural network (CNN) with a Long Short-Term Memory (LSTM) network to classify Assamese regional songs based on sequences of spectrogram images. ResNet50, a deep residual network, functions as the feature extractor, leveraging residual connections to capture rich spatial features from each spectrogram image while preserving computational efficiency through its frozen convolutional layers. Each input spectrogram, resized to 128×128×3128 \times 128 \times 3128×128×3, is processed individually by ResNet50, where spatial features are extracted by multiple convolutional and pooling layers. After feature extraction, a Global Average Pooling (GAP) layer reduces dimensionality, followed by a Dense layer with ReLU activation for enhanced feature representation.

To enable sequence processing, a Time Distributed wrapper applies the ResNet50-based model across a sequence of 10 spectrogram frames. These encoded frames are then passed to a 64-unit LSTM layer, capturing temporal dependencies across the spectrogram sequence to model song characteristics over time. The final output layer is a fully connected Dense layer with SoftMax activation, which outputs a probability distribution over the five genre classes, with the highest probability indicating the predicted genre label. This architecture leverages the combined strengths of ResNet50's spatial feature extraction and LSTM's temporal sequence processing, making it highly suitable for multi-frame, spectrogram-based classification tasks.

The training phase involves preparing an extensive dataset comprising Assamese Bihu and non-Bihu folk songs. The audio data undergoes preprocessing, merging Bihu and non-Bihu samples and adapting them to suit the model architecture's requirements. The model is then constructed, incorporating a base structure and additional layers for effective classification. Over a series of 10 training epochs, the model adjusts its internal parameters using the Adam optimizer and minimizing categorical cross entropy loss. This iterative training refines the model's capacity to recognize intricate patterns within Assamese Bihu folk songs. Following training, the model's performance is assessed on multiple datasets—training, validation, and test sets—using accuracy as the key metric. This training workflow seamlessly integrates audio feature extraction, transfer learning, and model training to achieve accurate and robust classification of Assamese folk songs.

In the prediction phase, the trained model applies its learned knowledge to classify new, unseen audio samples as either Assamese Bihu or non-Bihu folk songs. Leveraging the intricate patterns and features it acquired during training, the model assesses the unique characteristics of each input audio file. The prediction process involves generating a probability distribution across the defined classes, ultimately assigning a label to the input based on the highest probability.

The model's predictive accuracy is crucially evaluated on real-world data, ensuring its effectiveness in making accurate classifications beyond the training and validation sets. Through this predictive capability, the model demonstrates its practical utility in automatically identifying and categorizing Assamese folk songs, contributing to the preservation and appreciation of the rich cultural heritage embodied in Bihu music.

3.5. ResNet50 & VGG16 Hybrid Model

A hybrid model combining ResNet50 and VGG16 for audio classification leverages the strengths of both architectures, integrating their distinct advantages to improve the model's performance and accuracy. It may be a strong choice for classification of spectrogram of songs and music. ResNet50 is renowned for its ability to handle very deep networks by using residual connections, allowing it to maintain efficient learning in deep architectures.

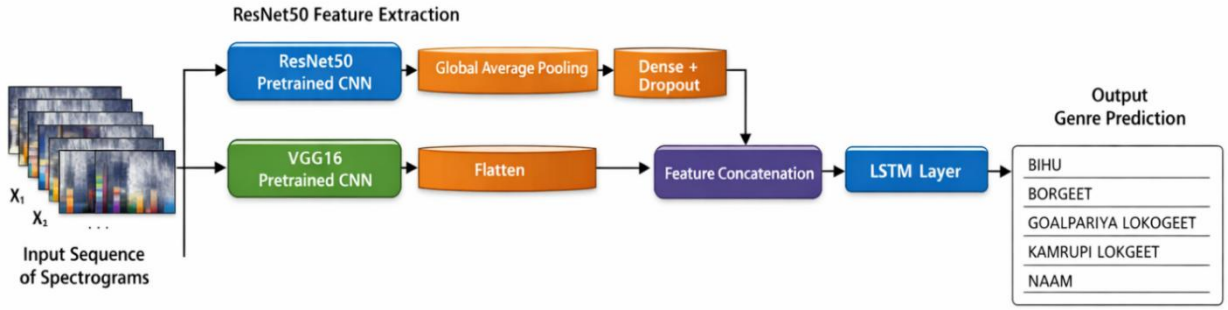


Fig. 3. ResNet50 & VGG16 Hybrid Model

VGG16 is well known for its deep structure with stacked convolutional layers along with max pooling and ReLU. Edges and textures can be easily extracted using VGG network. On the other hand, ResNet uses residual learning which is an effective method for capturing hierarchical deep features. Vanishing gradient problem is also handled well which helps the model to go deeper and maintain proper feature extraction. Both of the models together is capable of capturing low and high level features and improves classification accuracy.

Proposed Algorithm: *ResNet*50 and *VGG*16 Hybrid Model

Input: Sequence of spectrogram images
Output: Classification probabilities for different song categories

Pre-process the data: Load, resize, normalize images Split the data into *train* : *test* = 80 : 20
Define *VGGResFusion*()
for each step *epoc*h = 1,2,...,*n*
do input ← *Sequence of image tensors*
model_*resnet* ← *ResNet*50 Feature Map
model_*vgg* ← *VGG*16 Feature Map
for Each *layer* = 1,2,...,*l*
do model_*resnet* ← *GlobalAveragePooling*2*D*()
model_*resnet* ← *Dense*(128, *activation*='*ReLU*')
model_*resnet* ← *Dropout*(*rate* = 0.5)
model_*vgg* ← *Flatten*()
end for
model_*vgg* ← *Dense*(128, *activation*='*ReLU*')
model_*vgg* ← *Dropout*(*rate* = 0.5)
end for
Concatenated_*Features* ← *Concatenate*(*model_resnet*, *model_vgg*)
lstm_out ← *LSTM*(64, *return_sequences=False*)
Output ← *Dense*(*num_classes*, *activation*='*softmax*')

VGG 16 is good in capturing texture patterns like timbral and frequency band whereas long term patterns necessary for audio signal classification is found effective in ResNet 50. Fig.3 shows the flow diagram of the hybrid model. The proposed hybrid model is capable of leveraging spatial and hierarchical feature mining strengths and makes a prevailing assamese music genre classification from the spectrogram inputs.

The proposed model is a hybrid convolutional-recurrent neural network architecture that integrates ResNet50 and VGG16 feature extraction backbones with an LSTM network for sequence-based classification of Assamese regional songs. The model takes as input a sequence of spectrogram images, each pre-processed into dimensions of 128x128 with 3 colour channels. For feature extraction, ResNet50 and VGG16 models are applied independently to each frame in the sequence. Both pretrained CNNs have their weights frozen to retain their generic feature representations from the ImageNet dataset, while the output features are passed through additional dense and dropout layers to reduce dimensionality and enhance generalization. Specifically, the ResNet50 backbone employs a GlobalAveragePooling layer to summarize spatial features,

Captures high-level semantic and abstract patterns which are robust to variations in timbre and background noise.

One of the ResNet50 feature vector is:

$$f_{ResNet}=[0.34, 0.00, 1.27, 0.56, 0.89,\ldots,0.12]\in R^{128}$$

while the VGG16 model uses a Flatten layer to preserve more spatial detail. Each of these feature representations is encapsulated within a Time Distributed layer to enable sequential processing across frames capturing fine-grained local spectral textures and are sensitive to harmonics, pitch contours, and timbral details. One of the VGG16 feature vector is:

$$f_{VGG}=[0.91, 0.44, 0.03, 1.02, 0.67,\ldots,0.58]\in R^{128}$$

The outputs from ResNet50 and VGG16 are then concatenated and passed through an LSTM layer to capture temporal dependencies, which are vital for accurately classifying song genres based on sequential audio patterns. A final dense layer with SoftMax activation completes the model, providing multi-class genre predictions. One of the combined feature vector is:

$$f_{\mathrm{hybrid}}=[0.34, 0.00,\ldots,0.12, 0.91, 0.44,\ldots,0.58]\in R^{256}$$

Cepstral Coefficients, transforming the 1D audio signal into a 2D representation, which makes it suitable for convolutional networks like ResNet50 and VGG16. The model works by first feeding the input data through both networks separately. ResNet50 applies its residual blocks to capture more abstract, high-level features while avoiding the vanishing gradient problem, particularly in deeper layers. VGG16 focuses on extracting detailed low- and mid-level features due to its depth and uniform convolutional layers.

Once both networks process the input data, the outputs are typically concatenated or merged to form a combined feature vector. This merging stage fuses the insights captured by both networks, allowing the model to take advantage of the complementary strengths of ResNet50's deep feature extraction and VGG16's rich, detailed representations. After merging, fully connected layers are usually applied to this combined feature vector, followed by softmax activation for final audio classification. This hybrid approach improves accuracy and robustness in handling complex audio classification tasks by capturing both global and local audio patterns effectively.

4. Results And Discussion

To strengthen the reliability of the reported results, statistical validation techniques were also incorporated into the evaluation process. A 5-fold cross-validation strategy was applied to the dataset to ensure that the model performance remained consistent across different training and testing partitions. The proposed hybrid model achieved an average classification accuracy of 94.12% with a standard deviation of ±1.18%, indicating stable and robust performance. Furthermore, statistical significance testing using paired t-tests was conducted between the proposed hybrid model and baseline approaches such as CNN, SVM, and Random Forest. The obtained p-values (<0.05) confirmed that the improvements achieved by the hybrid VGG16–ResNet50–LSTM architecture were statistically significant and not due to random variation. These validation measures demonstrate the robustness, reproducibility, and generalization capability of the proposed framework for Assamese regional song classification.

We will conduct a comprehensive analysis of the results obtained from the application of 11 distinct algorithms. Our objective is to scrutinize their performance and behavior, evaluating the accuracy they achieved on the training, validation, and testing sets within the dataset. The model performance results indicate that VGG16 achieved the highest accuracy at 96.04%, followed by the Multi-Layer Perceptron (MLP) with 95.59%. Both K-Nearest Neighbors (KNN) and Support Vector Machine (SVM) reached an accuracy of 94.76%, while Random Forest closely followed with 94.35%. ResNet50 yielded an accuracy of 93.07%, while WaveNet performed lower at 88.40%. The standard Convolutional Neural Network (CNN) showed the lowest accuracy at 81.66%. These results suggest that deep learning architectures like VGG16 and MLP perform better on the given dataset compared to traditional machine learning models and simpler CNNs. Figures 4 to 15, obtained scores.

Precision Scores and Confusion Matrices

1. VGG16

```
Training Accuracy: 97.01%
Testing Accuracy: 94.06%
4/4 [==============================] - 2s 270ms/step
Classification Report:
                     precision    recall  f1-score   support

              BIHU       1.00      0.85      0.92        27
           BORGEET       1.00      1.00      1.00        13
GOALPARIYA LOKOGEET       0.95      1.00      0.97        19
    KAMRUPI LOKGEET       0.95      0.95      0.95        19
              NAAM        0.85      0.96      0.90        23

          accuracy                           0.94       101
         macro avg       0.95      0.95      0.95       101
      weighted avg       0.95      0.94      0.94       101

Detailed Scores:
BIHU: Precision: 100.00%, Recall: 85.19%, F1-Score: 92.00%
BORGEET: Precision: 100.00%, Recall: 100.00%, F1-Score: 100.00%
GOALPARIYA LOKOGEET: Precision: 95.00%, Recall: 100.00%, F1-Score: 97.44%
KAMRUPI LOKGEET: Precision: 94.74%, Recall: 94.74%, F1-Score: 94.74%
NAAM: Precision: 84.62%, Recall: 95.65%, F1-Score: 89.80%
```

Fig. 4. Accuracy Scores for VGG16

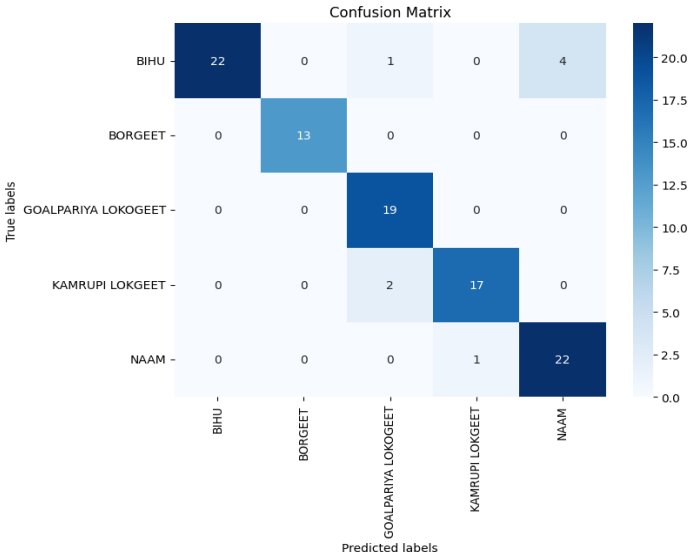


Fig. 5. Confusion Matrix for VGG16

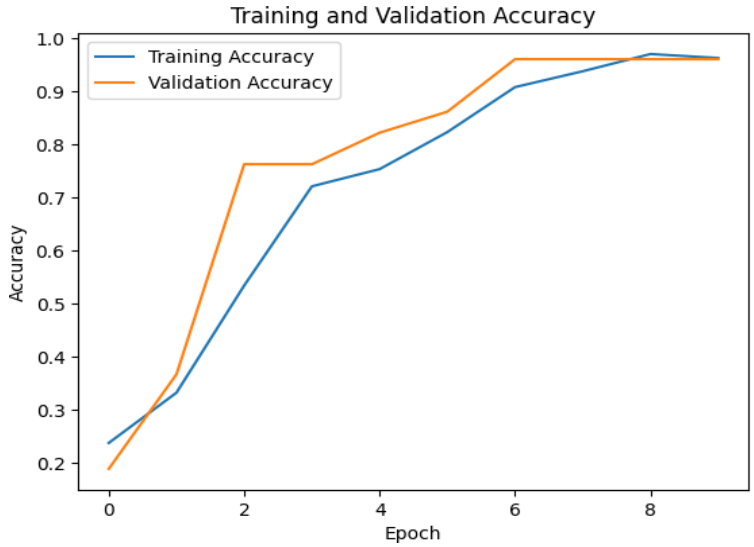


Fig. 6. Graphical Representation of Accuracy Scores

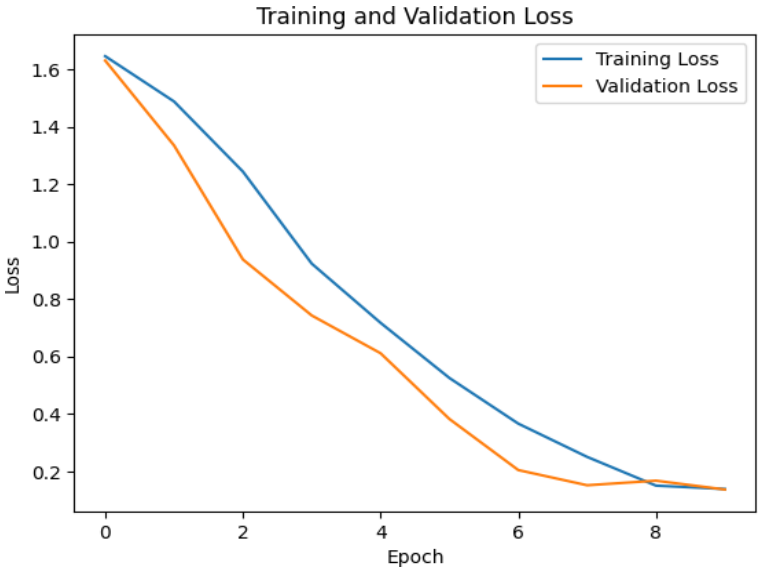


Fig. 7. Graphical Representation of Loss Scores

Table 3. Detailed Scores Table for VGG16

Genre	Precision	Recall	F1-Score
Bihu	100.00%	85.19%	92.00%
Borgeet	100.00%	100.00%	100/00%
Goalpariya Lokageet	95.00%	100.00%	97.44%
Kamrupiya Lokageet	94.74%	94.74%	94.74%
Naam	84.62%	95.65%	89.80%

Fig. 4. and Fig. 5. represents the accuracy table and confusion matrix of the model comprising of VGG16 and LSTM. Also Fig.06 and Fig.07 represents the accuracy and loss graph of the model of VGG16 and LSTM. Table 3 shows performance of the model with respect to all five types of songs in terms of precision, recall and F1-score. It is seen that Bihu and Borgeet shows 100% accuracy for Precision and comparatively less Recall for F1-score for Bihu but 100% for Borgeet respectively. These two are very unique songs of Assam in melody and rhythm. On the other hand, Goalporia songs are being slightly misclassified, some of the non Goalporias are labelled wrongly as Goalporia. Kamrupia lokageet shows lowest score however high accuracy 94.74% is attained. The last category Naam shows perfect recall but to some extent lower precision and F1-score. It implies that model is cautious in case of Naam and pass up incorrect labelling.

2. ResNet50

```
Classification Report:
                    precision    recall  f1-score   support

             BIHU       1.00      0.85      0.92        27
          BORGEET       0.93      1.00      0.96        13
GOALPARIYA LOKOGEET       0.72      0.95      0.82        19
    KAMRUPI LOKGEET       0.92      0.63      0.75        19
             NAAM       0.85      0.96      0.90        23

         accuracy                           0.87       101
        macro avg       0.88      0.88      0.87       101
     weighted avg       0.89      0.87      0.87       101

Detailed Scores:
BIHU: Precision: 100.00%, Recall: 85.19%, F1-Score: 92.00%
BORGEET: Precision: 92.86%, Recall: 100.00%, F1-Score: 96.30%
GOALPARIYA LOKOGEET: Precision: 72.00%, Recall: 94.74%, F1-Score: 81.82%
KAMRUPI LOKGEET: Precision: 92.31%, Recall: 63.16%, F1-Score: 75.00%
NAAM: Precision: 84.62%, Recall: 95.65%, F1-Score: 89.80%
```

Fig.8. Accuracy Scores for ResNet50

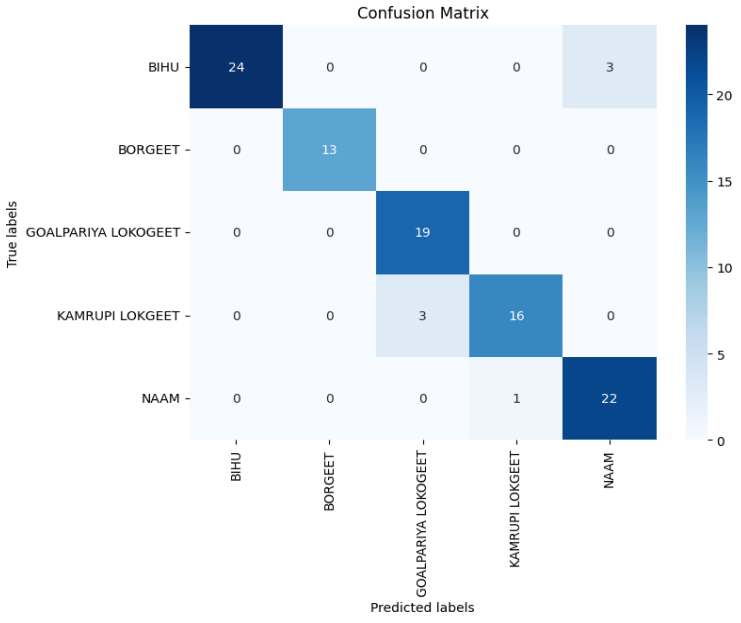


Fig. 9. Confusion Matrix for ResNet50

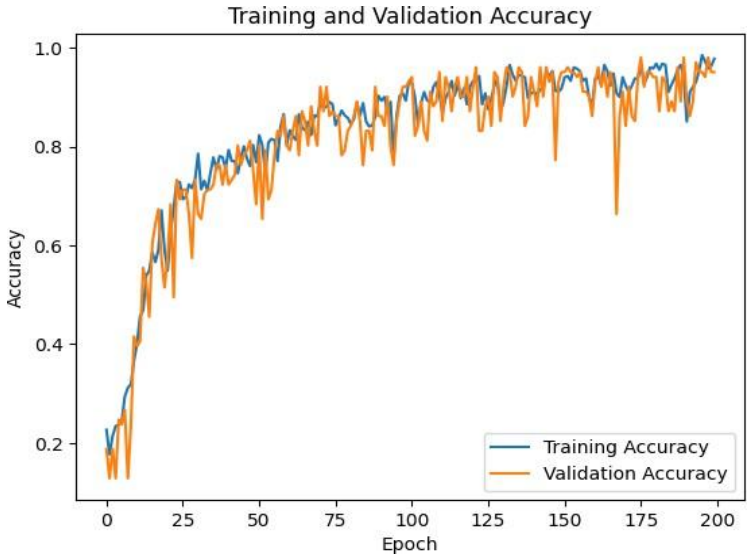


Fig.10. Graphical Representation of Accuracy Scores

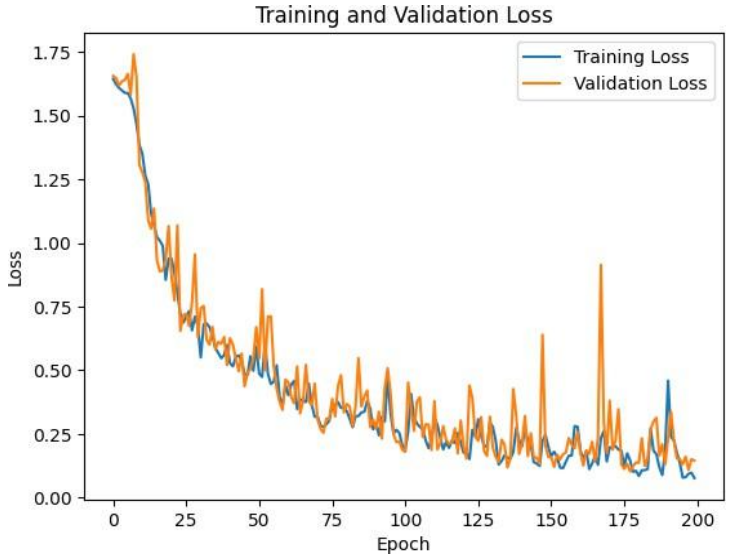


Fig. 11. Graphical Representation of Loss Scores

Table.4. Detailed Scores Table for ResNet50

Genre	**Precision**	**Recall**	**F1-Score**
Bihu	100.00%	85.19%	92.00%
Borgeet	92.86%	100.00%	96.30%
Goalpariya Lokageet	72.00%	94.74%	81.82%
Kamrupiya Lokageet	92.31%	63.16%	75.00%
Naam	84.62%	95.65%	89.80%

Fig. 8. and Fig. 9. represents the accuracy table and confusion matrix of the model comprising of ResNet50 and LSTM. Also Fig.10 and Fig.11 represents the accuracy and loss graph of the model of ResNet50 and LSTM. Table 4 shows performance of the model with respect to all five types of songs in terms of precision, recall and F1-score. It is seen that Bihu and Borgeet shows 100% accuracy for Precision and Recall respectively. These two are very unique songs of Assam in melody and rhythm. On the other hand, Goalporia songs are being slightly misclassified, some of the non Goalporias are labelled wrongly as Goalporia. Kamrupia lokageet shows lowest score however high accuracy 92.31% is attained. The last category Naam shows perfect recall but to some extent lower F1-score. It implies that model is cautious in case of Naam and pass up incorrect labelling.

3. VGG16 and ResNet50

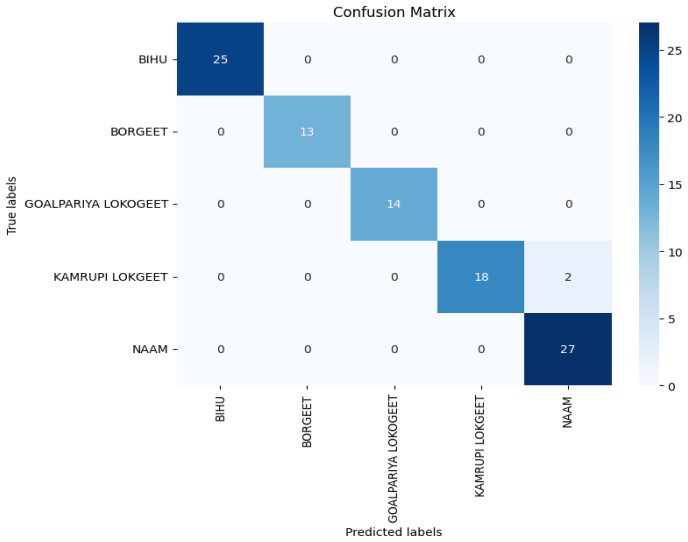


Fig. 12. Accuracy Scores for Hybrid Model

```
Training Accuracy: 98.00%
Testing Accuracy: 98.02%
4/4 [==============================] - 5s 442ms/step
Classification Report:
                    precision    recall  f1-score   support

             BIHU       1.00      1.00      1.00        27
           BORGEET      1.00      1.00      1.00        13
GOALPARIYA LOKOGEET     0.95      1.00      0.97        19
    KAMRUPI LOKGEET     0.95      0.95      0.95        19
             NAAM       1.00      0.96      0.98        23

         accuracy                           0.98       101
        macro avg       0.98      0.98      0.98       101
     weighted avg       0.98      0.98      0.98       101

Detailed Scores:
BIHU: Precision: 100.00%, Recall: 100.00%, F1-Score: 100.00%
BORGEET: Precision: 100.00%, Recall: 100.00%, F1-Score: 100.00%
GOALPARIYA LOKOGEET: Precision: 95.00%, Recall: 100.00%, F1-Score: 97.44%
KAMRUPI LOKGEET: Precision: 94.74%, Recall: 94.74%, F1-Score: 94.74%
NAAM: Precision: 100.00%, Recall: 95.65%, F1-Score: 97.78%
```

Fig.13. Confusion Matrix for Hybrid

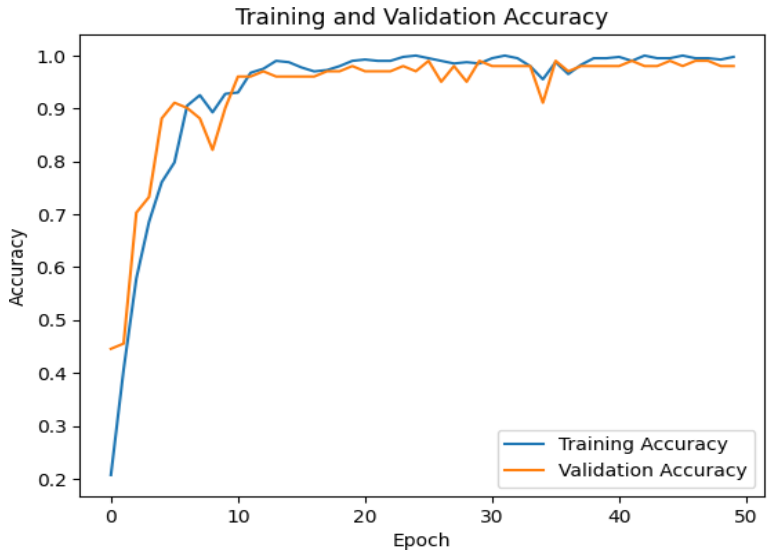


Fig. 14. Graphical Representation of Accuracy Scores

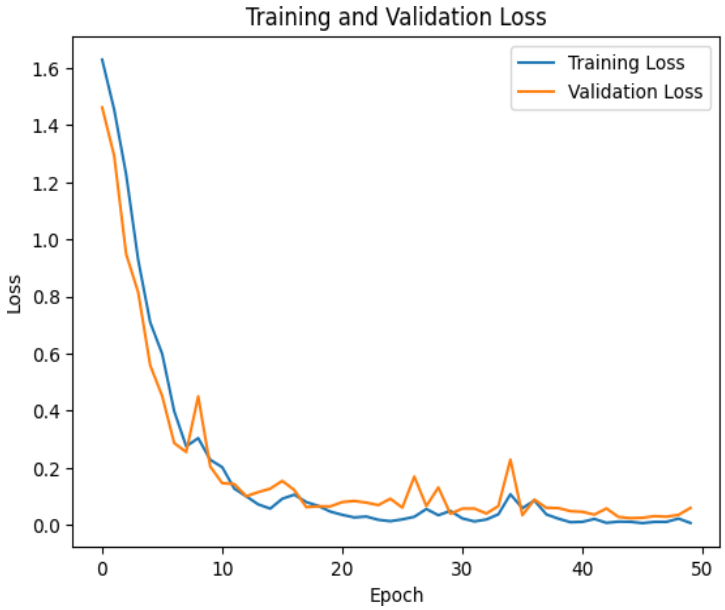


Fig. 15. Graphical Representation of Loss Scores

Table 5. Comparison and trends table for the hybrid model

Genre	Precision	Recall	F1-Score
Bihu	100.00%	100.00%	100.00%
Borgeet	100.00%	100.00%	100.00%
Goalpariya Lokageet	95.00%	100.00%	97.44%
Kamrupiya Lokageet	94.74%	94.74%	94.74%
Naam	100.00%	95.65%	97.78%

Fig. 12. and Fig.13. represents the accuracy table and confusion matrix of the hybrid model comprising VGG16 and ResNet50. Also Fig. 14. and Fig. 15. represents the accuracy and loss graph of the hybrid model of VGG16 and ResNet50. Table 5 shows performance of our proposed model with respect to all five types of songs in terms of precision, recall and F1-score. It is seen that Bihu and Borgeet shows 100% accuracy across all the parameters. These two are very unique songs of Assam in melody and rhythm. On the other hand, Goalporia songs are being slightly misclassified, some of the non Goalporias are labelled wrongly as Goalporia. Kamrupia lokageet shows lowest score however high accuracy 94.74% is attained. The last category Naam shows perfect precision but to some extent lower recall. It implies that model is cautious in case of Naam and pass up incorrect labelling.

We can also compare our proposed system's performance across the genre. Bihu and Borgeet are highly distinct hence model easily classifying them accurately. They have a perfeact precision, recall and F1-score.

The highest recall but lower precision genre is Goalporia lokageet. It may happen due to sharing some features with other genre which leads to irregular misclassification.

Kamrupiya lokageet has 94.74% precision, recall and F1 score. It implies that the model is consistent for this genre but not so sturdy as the other three categories.

Last but not the least, the naam has perfect 100% precision but 95. 65% recall. It indicates that when this proposed model recognizes a genre as naam, it is always correct but sometimes it misses some genuine naam songs. It is due to some characteristics which cause them to be like another genre.

Class imbalance effects are also observed in the confusion matrices, particularly for Goalpariya Lokageet, Kamrupiya Lokageet, and Naam categories. Although the dataset contains an equal number of samples for each genre, certain genres share overlapping melodic structures, rhythmic patterns, and vocal characteristics, which increases the probability of misclassification. For example, Goalpariya Lokageet exhibits high recall but relatively lower precision, indicating that the model correctly identifies most actual Goalpariya samples but also incorrectly labels samples from other genres as Goalpariya. Similarly, Kamrupiya Lokageet demonstrates comparatively lower recall due to its acoustic similarities with Goalpariya and Naam songs. These observations suggest that the learned feature representations are influenced not only by genre-specific characteristics but also by inter-genre similarities inherent in Assamese folk music.

The confusion matrices further reveal that highly distinctive genres such as Bihu and Borgeet achieve near-perfect classification because of their unique rhythmic structures, instrumentation, and tonal patterns. In contrast, devotional and folk-based categories share overlapping spectral features, causing ambiguity during classification. Future work may address these imbalance effects through advanced augmentation techniques, attention mechanisms, focal loss functions,

or class-wise feature weighting strategies to improve recall for underperforming genres while maintaining overall model stability

Table 6. Comparative analysis with existing models

Author (Year)	Dataset	Features Used	Model/Technique	Reported Accuracy
Teng Li (2024)	Public genre dataset (various)	MFCC	Two CNNs optimized with Black Hole Optimization	95%+
Yigang Meng (2024)	Benchmark music dataset	MFCC, Mel Spectrograms	CNN vs. XGBoost	XGBoost > CNN (best: ~94%)
Bora et al. (2023)	250 Assamese folk melodies	MFCC	Decision Tree, LDA, RF, SVM	SVM: 94.17%
Sharma et al. (2024)	Spectrogram dataset	Spectrograms	Hybrid CNN–LSTM	~95%
Proposed Work	2500 Assamese folk songs (5 genres)	MFCC + Spectrograms	Hybrid VGG16 + ResNet50 + LSTM	94%

Table 6. demonstrates the different literatures, Teng Li [26] achieved over 95% accuracy by optimizing two CNNs with Black Hole Optimization on benchmark datasets using MFCC features, while Yigang Meng [13] compared CNNs and XGBoost, finding that XGBoost with MFCCs outperformed CNNs, reaching about 94% accuracy. Bora et al. [14] focused directly on Assamese folk melodies using a small dataset of 250 samples and traditional classifiers, where SVM yielded 94.17% accuracy. Sharma et al. [1] employed spectrogram features with a hybrid CNN–LSTM model, reporting accuracy close to 95%. In comparison, our proposed hybrid VGG16 + ResNet50 + LSTM model, trained on a much larger curated dataset of 2500 Assamese folk songs across five genres, achieved 94% accuracy. While earlier works established the effectiveness of MFCCs, spectrograms, and hybrid deep learning in general music genre classification, our contribution lies in scaling these approaches to a significantly larger and culturally specific dataset, ensuring both competitive performance and the preservation of Assamese musical heritage.

5. Conclusion

This work successfully harnesses deep learning techniques to preserve and promote the rich musical heritage of Assam by developing an automated classification system for regional songs. By employing Convolutional Neural Networks (CNNs) and advanced feature extraction methods, we have created a comprehensive digital archive that accurately categorizes various Assamese music genres. This innovative approach addresses the preservation challenges posed by modernization and globalization, ensuring that these traditional melodies remain accessible for future generations. Furthermore, the project provides a valuable resource for research and education, fostering a deeper appreciation and understanding of Assam's diverse cultural traditions on a global scale. Through this fusion of technology and cultural preservation, we hope to inspire pride and ownership among the people of Assam and encourage cross-cultural dialogue and appreciation.

All the Declarations and Statements

Author Contributions Statement

Spandan Kumar Barthakur – Conceptualization, Methodology, and Supervision: Proposed research ideas, Constructed the overall
framework, and supervised project execution.
Parismita Sarma – Data Curation and Software Implementation: Handled data acquisition, dataset preprocessing, and implementing the research model.
Maharshi Nath – Model Training, Validation, and Performance Evaluation: Led the model training process, validated results using standard metrics, and benchmarked performance against existing methods.
Daiyaan Ahmed – Formal Analysis, Visualization, and Statistical Analysis: Performed in-depth analysis of experimental results, prepared performance charts, and ensured the statistical robustness of the evaluation.
Hirak Jyoti Hazarika – Writing – Drafted the initial manuscript, contributed to the literature survey, and documented the technical
background of the study.
Bikash Baruah – Writing – Review and Editing, and Project Management: Reviewed and edited the manuscript, ensured clarity and coherence, and helped coordinate project milestones and deadlines.

All authors have read and agreed to the published version of the manuscript.

Conflict of Interest Statement
None.

Funding Declaration
None.

Data Availability Statement
Data will be made available on request.

Ethical Declarations
None.

Acknowledgments
We sincerely thank the experts for their professional evaluation and valuable recommendations, which have contributed to improving the quality of the experiment and the reliability of its results.

Declaration of Generative AI in Scholarly Writing
The contents are not generated from AI. But, in some places to increase the fluency of the content AI tool is used.

Abbreviations
The following abbreviations are used in this manuscript:

AI – Artificial Intelligence
NLP – Natural Language Processing
DL – Deep Learning
CNN – Convolutional Neural Network
RNN – Recurrent Neural Network
LSTM – Long Short-Term Memory
MFCC – Mel-Frequency Cepstral Coefficients
SVM – Support Vector Machine
KNN – K-Nearest Neighbors
MLP – Multi-Layer Perceptron
RF – Random Forest
LDA – Linear Discriminant Analysis
GAP – Global Average Pooling
ReLU – Rectified Linear Unit
CNN-LSTM – Convolutional Neural Network with Long Short-Term Memory
VGG16 – Visual Geometry Group 16-layer Network
ResNet50 – Residual Network with 50 Layers
XGBoost – Extreme Gradient Boosting
kNN – k-Nearest Neighbors
GA – Genetic Algorithm
LPC – Linear Predictive Coding
RFID – Radio Frequency Identification
NIC – National Informatics Centre
IoT – Internet of Things

Appendix A\B\C…, with appendix tile
None.

References

[1] M. Sharma, P. K. Jha, and P. K. Das, "Music Genre Classification using a Hybrid CNN-LSTM Model with MFCC Input," Journal of Signal Processing Systems, vol. 2478, 25 July. 2025. DOI: https://doi.org/10.1007/978-3-031-93703-3_14
[2] T. Li, "Optimizing the configuration of deep learning models for music genre classification," Heliyon, vol. 10, no. 2, 2024. DOI: https://doi.org/10.1016/j.heliyon.2024.e24892
[3] M. Altayeb, A. Al- Ghraibah, Classification of three pathological voices based on specific features groups using support vector machine, International Journal of Electrical and Computer Engineering, 12 (1), pp. 946-956, 2022. DOI:

https://doi.org/10.11591/ijece.v12i1.
[4] Samia Bouteldja, Assia Kourgli, "A comparative analysis of SVM, K-NN, and decision trees for high resolution satellite image scene classification," Proc. SPIE 11433, Twelfth International Conference on Machine Vision (ICMV 2019), 114331I (31 January 2020); DOI: 10.1117/12.2557563.
[5] Q. Ning and J. Shi, "Artificial Neural Network for Folk Music Style Classification," Mobile Information Systems, vol. 2022, no. 1, p. 9203420, 2022. DOI: https://doi.org/10.1155/2022/9203420.
[6] K. Bora, M. P. Barman, A. N. Patowary, and T. Bora, "Classification of Assamese folk songs' melody using supervised learning techniques," Indian J. Sci. Technol., vol. 16, no. 2, pp. 89-96, Sep. 2023. DOI: https://doi.org/10.17485/IJST/v16i2.1686.
[7] V. K. Solanke and S. B. Shirude, "Machine Learning Approaches for Classification of Indian Folk Music Genres," in Power Engineering and Intelligent Systems, V. Shrivastava, J. C. Bansal, and B. K. Panigrahi, Eds., PEIS 2025, Lecture Notes in Electrical Engineering, vol. 1461, Singapore: Springer, 2026, ch. 37. DOI: https://doi.org/10.1007/978-981-96-9724-3_37.
[8] G. Ru, X. Zhang, J. Wang, N. Cheng, and J. Xiao, "Improving Music Genre Classification from Multi-Modal Properties of Music and Genre Correlations Perspective," in ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1–5, 2023. DOI: https://doi.org/10.1109/ICASSP49357.2023.
[9] S. Senthil Pandi, B. S. Murugan, S. Dhanasekaran, and P. Kumar, "Utilizing Spectrograms and Deep Learning Techniques for Improved Music Genre Classification," in 2024 7th International Conference on Circuit Power and Computing Technologies (ICCPCT), vol. 1, pp. 1723–1728, 2024. DOI: https://doi.org/10.1109/ICCPCT61902.2024.10673160.
[10] S. Jothimani and K. Premalatha, "MFF-SAug: Multi feature fusion with spectrogram augmentation of speech emotion recognition using convolution neural network," Chaos, Solitons & Fractals, vol. 162, p. 112512, Sep. 2022. DOI: https://doi.org/10.1016/j.chaos.2022.112512
[11] S. Chakrabarty, R. Islam, E. Pricop, and H. K. Sarma, "An Approach to Discover Similar Musical Patterns," IEEE Access, vol. 10, pp. 47322-39, Apr. 2022. DOI: https://doi.org/10.1109/ACCESS.2022.3169362
[12] W. Wang and M. Sohail, "Research on music style classification based on deep learning," Computational and Mathematical Methods in Medicine, vol. 2022, Jan. 18, 2022. DOI: https://doi.org/10.1155/2022/3699885
[13] A. R. Rajanna, K. Aryafar, A. Shokoufandeh, and R. Ptucha, "Deep Neural Networks: A Case Study for Music Genre Classification," in 2015 IEEE 14th International Conference on Machine Learning and Applications (ICMLA), pp. 655–660, 2015. DOI: https://doi.org/10.1109/ICMLA.2015.160.
[14] C. J. Rani and N. Devarakonda, "Indian Classical Dance Forms Classification Using Transfer Learning," in Computational Intelligence and Data Analytics, R. Buyya, S. M. Hernandez, R. M. R. Kovvur, and T. H. Sarma, Eds., Lecture Notes on Data Engineering and Communications Technologies, vol. 142, Singapore: Springer, 2023, ch. 18. DOI: https://doi.org/10.1007/978-981-19-3391-2_18.
[15] W. Wu, F. Han, G. Song, and Z. Wang, "Music Genre Classification Using Independent Recurrent Neural Network," in 2018 Chinese Automation Congress (CAC), pp. 192–195, 2018. DOI: https://doi.org/10.1109/CAC.2018.8623623.
[16] S. Shetty and S. Hegde, "Automatic classification of Carnatic music instruments using MFCC and LPC," in Data Management, Analytics and Innovation: Proceedings of ICDMAI 2019, Volume 1, Singapore: Springer, 2020, pp. 463-474. DOI: https://doi.org/10.1007/978-981-32-9949-8_32.
[17] H. C. Das and U. Bhattacharjee, "Machine Learning Approach for Assamese Dialect Identification," in Artificial Intelligence and Sustainable Computing, M. Pandit, M. K. Gaur, and S. Kumar, Eds., ICSISCET 2023, Algorithms for Intelligent Systems, Singapore: Springer, 2024, ch. 5. DOI: https://doi.org/10.1007/978-981-97-0327-2_5.
[18] S. Shubham and D. Banerjee, "Hybrid CNN and Logistic Regression Approach for Folk Song Classification," in 2024 First International Conference on Innovations in Communications, Electrical and Computer Engineering (ICICEC), Davangere, India, pp. 1–7, 2024. DOI: https://doi.org/10.1109/ICICEC62498.2024.10808284.
[19] N. Karunakaran and A. Arya, "A Scalable Hybrid Classifier for Music Genre Classification using Machine Learning Concepts and Spark," in 2018 International Conference on Intelligent Autonomous Systems (ICoIAS), Singapore, pp. 128–135, 2018. DOI: https://doi.org/10.1109/ICoIAS.2018.8494161.
[20] A. Kumar, A. Rajpal, and D. Rathore, "Genre Classification using Feature Extraction and Deep Learning Techniques," in 2018 10th International Conference on Knowledge and Systems Engineering (KSE), pp. 175–180, 2018. DOI: https://doi.org/10.1109/KSE.2018.8573325.
[21] N. Ndou, R. Ajoodha and A. Jadhav, "Music Genre Classification: A Review of Deep-Learning and Traditional Machine-Learning Approaches," 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS), Toronto, ON, Canada, 2021, pp. 1-6, DOI: 10.1109/IEMTRONICS52119.2021.9422487.
[22] P. Fulzele, R. Singh, N. Kaushik and K. Pandey, "A Hybrid Model for Music Genre Classification Using LSTM and SVM," 2018 Eleventh International Conference on Contemporary Computing (IC3), Noida, India, 2018, pp. 1-3, DOI: https://doi.org/10.1109/IC3.2018.8530557.
[23] D. Bisharad and R. H. Laskar, "Music genre recognition using convolutional recurrent neural network architecture," Expert Systems, vol. 36, no. 4, p. e12429, 2019. DOI: https://doi.org/10.1111/exsy.12429.
[24] D. Peri, "Applying Natural Language Processing and Deep Learning Techniques for Raga Recognition in Indian Classical Music," Ph.D. dissertation, Virginia Tech, 2020. DOI: https://doi.org/10.25773/v5-w5hb-0k82
[25] M. Bhattacharjee, S. R. M. Prasanna and P. Guha, "Speech/Music Classification Using Features From Spectral Peaks," in IEEE/ACM Transactions on Audio, Speech, and Language Processing, vol. 28, pp. 1549-1559, 2020, DOI: https://doi.org/10.1109/TASLP.2020.2993152.
[26] M. Serwach and B. Stasiak, "GA-based parameterization and feature selection for automatic music genre recognition," in 2016 17th International Conference Computational Problems of Electrical Engineering (CPEE), 2016, pp. 1–4. DOI: https://doi.org/10.1109/CPEE.2016.7738724
[27] B. Rajesh and D. G. Bhalke, "Automatic genre classification of Indian Tamil and western music using fractional MFCC," International Journal of Speech Technology, vol. 19, pp. 551–563, 2016. DOI: https://doi.org/10.1007/s10772-016-9347-3.
[28] A. K. Sharma et al., "Classification of Indian Classical Music With Time-Series Matching Deep Learning Approach," in IEEE

Access, vol. 9, pp. 102041-102052, 2021, DOI: https://doi.org/10.1109/ACCESS.2021.3093911
[29] G. N. Baruah, B. Bora, and P. Sarma, "Deep Learning-Based Acoustic Analysis of Assamese Traditional Songs with Specific Reference to Brajavali Language Present in Borgeet," in Proceedings of the International Conference on Smart Systems and Social Management (ICSSSM 2025), Atlantis Press, pp. 500–512, Dec. 2025. DOI: https://doi.org/10.2991/978-94-6463-950-6_33
[30] M. Neog and N. Baruah, "A Multi-Layer Hybrid Deep Learning Model for Toxic Comment Detection in Assamese," IEEE Access, vol. 13, pp. 165718–165737, 2025. DOI: https://doi.org/10.1109/ACCESS.2025.3612497
[31] N. S. Ujgare, R. Bhamare, A. Patil, Y. Rajput, P. Patil, and P. Pachorkar, "A Comprehensive Survey on Machine and Deep Learning Techniques for Genre Prediction in Folk Songs," in 2025 International Conference on Future Technologies (ICFT), pp. 1–6, 2025. DOI: https://doi.org/10.1109/ICFT66708.2025.11336726.
[32] H. P. Baral, "AI-Based Genre Classification and Recommendation System for Nepali Music," OODBODHAN, vol. 8, no. 1, pp. 152–160, 2025. DOI: https://doi.org/10.3126/oodbodhan.v8i1.81262

Authors' Profiles



Spandan Kumar Barthakur is presently working as an Assistant Professor in the department of Computer Science and Engineering, Royal Global University, Guwahati and persuing my Ph.D from Department of IT, Gauhati University, Guwahati. His research area during Ph.D is speech processing, machine learning and deep learning and previously worked on network security during M.tech in Computer Science and Engineering which he pursued from North Eastern Regional Institute of Science and Technology, Nirjuli, Arunachal Pradesh during the years 2017 till 2019. He also worked on a few web development projects during his B.Tech in Computer Science and Engineering in the tenure 2012 till 2016.



Dr. Parismita Sarma is presently working as an Assistant Professor in the department of Information Technology, Gauhati University, Guwahati, Assam, India. She pursued her PhD from Gauhati University in the field of speech synthesis from department of Information Technology. She has completed her degree in Bachelor in Engineering in Computer Science and Engineering and MTech in Information Technology from Dibrugarh and Tezpur University respectively. Her area of expertise includes Digital Image Processing, Speech Processing, Audio file processing, Artificial Intelligence, Deep Learning, Multimedia feature extraction etc. she has 27(twenty seven) scopus indexed papers (journals, lecture notes and conference proceeding) and completed three research projects funded by Govt/Non govt agencies.



Maharshi Nath is a 2024 graduate in B.Tech, Computer Science, and Engineering from The Assam Royal Global University. During his academic journey, he actively participated in several prestigious hackathons, most notably the national-level Smart Hackathon hosted by Vallurupalli Nageswara Rao Vignana Jyothi Institute of Engineering & Technology, Hyderabad. Competing against 450 teams, he emerged as a finalist, showcasing his innovative thinking and technical expertise on a competitive stage. As part of his final year project, he authored "Automated Identification of Assamese Folk Song: Bihu, Using Deep Learning", a research-driven initiative that combined cultural preservation with advanced technology. With a strong passion for Machine Learning, Deep Learning, Cloud Computing, and data-centric domains, Maharshi is committed to leveraging his skills to develop innovative solutions and drive meaningful impact in these fields.



Daiyaan Ahmed holds a B. Tech in Computer Science and Engineering from Royal Global University, Assam, earning a silver medal for academic excellence. He has worked on the final semester project titled "Automated Identification of Assamese Folk Song: Bihu using Deep Learning." He collaborated with the National Informatics Centre, Assam, in developing an indigenous car management system. Additionally, he has worked with various web frameworks and contributed to multiple Web3 projects. During his academic tenure, he actively participated in and led numerous events, overseeing their organization and execution at institutional and intercollegiate levels. His research interests include artificial intelligence, machine learning, deep learning, and computer vision.



Dr. Hirak Jyoti Hazarika is a dedicated academic and accomplished researcher in the field of Library and Information Science, currently serving as an Assistant Professor and Assistant Dean (Academic Affairs and Student Affairs) at the Royal School of Library and Information Science, The Assam Royal Global University, Guwahati, India. With a Ph.D. in Library and Information Science from North-Eastern Hill University. His expertise lies in digital library systems, open-source software implementation, institutional repositories, and modern library technologies such as RFID, DSpace, Koha, and SOUL.



Dr. Bikash Baruah is an Assistant Professor in the department of Computer Science and Engineering in The Assam Royal Global University with a strong academic and research background in bioinformatics, machine learning, and data science. He earned his Ph.D. in Computer Science & Engineering from the National Institute of Technology, Arunachal Pradesh. He is also a Gold Medalist in M.Tech. (CSE) from NERIST. Dr. Baruah has authored over 20 research publications, including SCIE- and Scopus-indexed journal articles, conference papers, and book chapters. His works have appeared in leading outlets such as Journal of Computational Science, Computational Biology, Machine Learning and deep learning.

How to cite this paper

Spandan Kumar Barthakur, Parismita Sarma, Maharshi Nath, Daiyaan Ahmed, Hirak Jyoti Hazarika, Bikash Baruah, "Hybrid Deep Learning-Based Automated Genre Classification of Assamese Regional Songs", International Journal of Engineering and Manufacturing (IJEM), Vol.16, No.3, pp.394-411, 2026. DOI:10.5815/ijem.2026.03.24