International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Benchmarking Machine Learning and Transformer Models on a Custom Marathi Folk Song Lyrics Corpus

Nitin S. Ujgare, Apurwa Patil, Roshni Bhamare, Yogeshwari Rajput, Pranjal Patil
Department of Information Technology MVP's KBTCOE, Nashik, India

Abstract **- Marathi folk songs represent a rich cultural heritage encompassing diverse genres such as Powada, Lavani, Bharud, Bhajan, etc. However, the classification of these folk songs remains a challenging task due to linguistic variations and the limited availability of structured resources. This paper presents a comparative study of classical machine learning models and a transformer-based approach for Marathi folk song genre classification using textual features. One of the key contributions of this work is a custom Marathi folk song dataset, annotated with genre, built specifically to support supervised learning in a domain that has received very little attention so far. For the classical models, TF-IDF is used for feature extraction, while the transformer model (MuRIL) learns contextual representations directly from the text. The models evaluated include Logistic Regression, Naive Bayes, Linear SVC, and MuRIL. The results show that among all the evaluated models, Linear SVC achieved the highest accuracy of 72.50%, followed by Logistic Regression (68.75%), MuRIL (48.11%), and Naive Bayes (40.00%). This trend indicates that in low-resource scenarios with limited training data, classical machine learning models can outperform the transformer-based approaches. Overall, the findings reflect both the strengths and current limitations of automated folk song classification, while also highlighting its potential for supporting the preservation and analysis of regional cultural heritage.**

Keywords Marathi Folk Songs, Text Classification, Cultural Heritage Preservation, Machine Learning, TF-IDF

I. INTRODUCTION

Music genre prediction is a well-known problem in Music Information Retrieval (MIR), with practical applications ranging from recommendation systems and playlist organization to digital archiving. Most of the work done so far has centered on audio-based features. Folk music presents additional challenges such as significant overlap in themes, linguistic styles, and cultural context. Expert knowledge has long been the backbone of folk song classification.The automatic classification of folk songs remains difficult due to overlapping stylistic patterns, linguistic diversity, and the scarcity of well-structured, annotated datasets [1]. Moreover, genre boundaries in folk traditions are not rigid; they are often loosely defined and overlapping, which further complicates the task of automated classification.

The problem becomes more pronounced in the context of Marathi folk music. The language has limited NLP resources, and publicly available, well-annotated datasets are hard to come by. Existing research in MIR has mostly focused on Western music and high-resource languages, leaving regional Indian folk traditions with little representation in automated analysis systems. Recent studies, including [2], have traced the shift in genre prediction methods from traditional feature-based techniques to transformer-based models. Even so, applying these approaches to regional folk music Marathi in particular remains limited, largely due to data scarcity and linguistic variability. To address this gap, this paper presents a curated dataset of Marathi folk song lyrics annotated with genre, regional origin, and historical context. Multiple machine learning models are evaluated comparatively using TF-IDF-based textual features.

The study examines the performance of classical machine learning approaches under low-resource conditions, with particular attention to linguistic diversity and class imbalance. Results show that while the models achieve acceptable classification performance, data scarcity, genre overlap, and regional variation remain persistent challenges for generalization. Progress in these areas is critical for developing systems capable of supporting the preservation and analysis of regional cultural heritage

A. Key Contribution

The contributions of this work are summarized as follows:

1. A curated and preprocessed dataset of Marathi folk song lyrics prepared to facilitate supervised learning for genre classification.
2. A supervised text classification pipeline is developed using TF-IDF feature representation, and Linear Support Vector Classifier (Linear SVC) is identified as the most effective model for Marathi folk song genre prediction

II. RELATED WORK

Music genre classification, particularly for folk songs, has been widely studied using a variety of computational approaches. Early research primarily focused on traditional machine learning techniques such as Support Vector Machines (SVM), Decision Trees, Na¨ıve Bayes, and Random Forests. These methods relied on hand crafted features including Mel-Frequency Cepstral Coefficients (MFCCs) and textual representations. For example, Li and Zhang [3] applied SVM and Decision Trees for Anhui folk song classification, achieving an accuracy of 70–75%. Similarly, Lalmuanzuala and Chhakchhuak [4] used Na¨ıve Bayes and Random Forests for Mizo folk songs, reporting approximately 72% accuracy. Patil and Joshi [5] further demonstrated that machine learning models

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

can capture linguistic variations in Marathi dialects. In addition, studies by McKay and Fujinaga [6] showed that lyrics alone can be effective for genre classification, while Karydis and Theodoridis [7] improved performance by combining audio and textual features. However, these approaches depend heavily on manual feature engineering and of ten struggle to model complex patterns in music data.

With the advancement of deep learning, more sophisticated models have been introduced to automatically learn feature representations. Techniques such as Long Short-Term Memory (LSTM), Bidirectional LSTM (Bi-LSTM), and hybrid CNN-RNN architectures have shown improved capability in capturing temporal and contextual information. Karami and Ahmadi [8] utilized LSTM-based models and achieved accuracies in the range of 81–83%, while Zhao and Chen [9] proposed a CNN-BiGRU model with attention mechanisms, reaching up to 84% accuracy. Other hybrid approaches, including CNN combined with Logistic Regression [10] and CNN-RNN frameworks [11], have also demonstrated improved performance over traditional machine learning methods. Attention-based models [12] and multimodal systems integrating audio and lyrics [13] further enhanced classification accuracy. Despite these improvements, deep learning models typically require large annotated datasets and significant computational resources, which limits their applicability in low-resource settings.

More recently, transformer-based architectures have gained prominence due to their ability to model long-range dependencies in sequential data. Models such as BERT and its variants have shown strong performance in text-based classification tasks. However, their effectiveness in folk music classification is constrained by data scarcity and computational requirements, particularly for regional languages.

Despite the progress, there remains a significant gap in the development of efficient and scalable models for Marathi folk song classification using limited textual data. This motivates the approach proposed in this work.

III. METHODOLOGY

This section offers an overview of the methods employed in this study

A. Dataset

The dataset consists of 530 Marathi folk song lyrics collected from multiple authentic sources, including folk music archives, academic collections, cultural organizations, and manually transcribed field recordings. All lyrics are written in the Marathi language using the Devanagari script to ensure linguistic consistency. The dataset covers 12 distinct folk genres such as Powada, Lavani, Bhajan, Bharud, and Koli Geet, etc. representing a wide cultural diversity. Each song is annotated with its corresponding genre label, which serves as the primary target variable for the classification task

In addition to genre labels, the dataset also includes supplementary metadata such as region and historical context, which are retained for future research and extended analysis. However, the current study focuses exclusively on genre-based classification. The dataset is organized in a structured tabular format containing attributes such as Title, Lyrics, Genre, Region, and History.

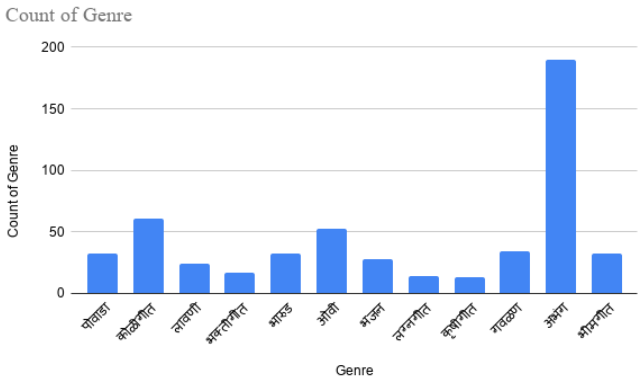

Fig. 1. Distribution of Samples across different Marathi folk song genres

The Fig 1. shows the distribution of samples across different Marathi folk song genres, highlighting a clear class imbalance in the dataset. A few dominant classes contain a significantly higher number of samples compared to several underrepresented classes.

B. Data Preparation and Preprocessing

The dataset was first organized into input lyrics and corresponding genre labels. Basic preprocessing was applied to normalize the textual data and remove formatting inconsistencies. Genre labels were cleaned to eliminate extra whitespace and line-break variations, ensuring consistent class representation across all 12 categories. The lyrics were then used directly for feature extraction without aggressive linguistic normalization, so that the original lexical and contextual characteristics of Marathi folk songs were preserved.

C. Feature Extraction

A. TF-IDF Based Feature Representation

For classical machine learning models such as Logistic Regression, Naive Bayes, and Linear SVC, the lyrical text was transformed into numerical feature vectors using Term Frequency--Inverse Document Frequency (TF-IDF). A word-level n-gram range of (1,2) was used to capture both individual terms and short contextual phrases. Terms with document frequency lower than 2 were excluded to reduce noise, and the maximum feature limit was set to 25,000. In the present dataset, this resulted in an effective vocabulary size of 2753 features. This sparse representation captures word importance across the corpus and is suitable for linear classifiers.

B. Transformer Based Feature Extraction

For the transformer-based model, feature extraction was performed using MuRIL (\texttt{google/muril-base-cased}), a pretrained multilingual transformer designed for Indian languages. The input lyrics were tokenized into subword units with a maximum sequence length of 128 tokens. Unlike TF-IDF, MuRIL generates contextual embeddings in which each token representation depends on its surrounding context. These embeddings are processed through transformer encoder layers, and the resulting sequence representation is used for the classif-

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

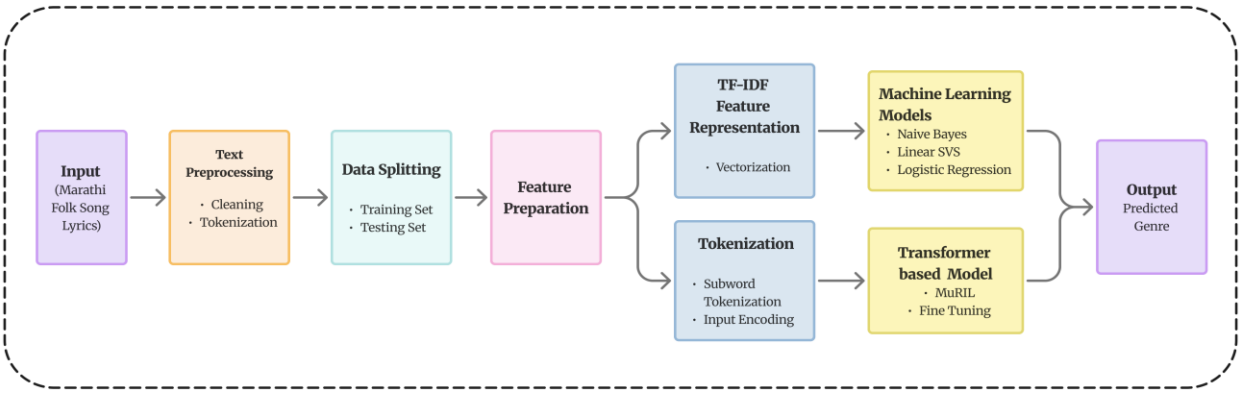

Fig. 2. Proposed System Pipeline

ication via the linear output layer corresponding to the 12 genre classes. The model is fine-tuned on the labeled dataset to adapt pretrained knowledge to the task.

D. Proposed Text Classification Pipeline

Fig. 2, demonstrates the workflow of the system processes short Marathi lyrical inputs through a structured classification pipeline. Initially, the input text undergoes preprocessing, including cleaning, tokenization. The processed data is then split into training and testing sets in the ratio of 85:15. Feature extraction is performed using TF-IDF to convert text into numerical vectors representing word importance. These features are fed into machine learning models along with transformer-based models for contextual understanding. During fine-tuning of the MuRIL model, the AdamW optimizer is employed to update model weights efficiently and improve generalization. The trained models generate predictions, and the best-performing model is used to classify the input into the appropriate folk song genre.

E. Models

The classification task is formulated as a multi-class text classification problem, where each input lyric is assigned to one of 12 distinct Marathi folk song genres.

To address this task, both traditional machine learning models and a transformer-based model are implemented and compared. The models considered in this study include:

- Naïve Bayes (NB) as a probabilistic baseline
- Logistic Regression (LR) as a linear classifier
- Linear Support Vector Classifier (Linear SVC), well-suited for high-dimensional sparse feature spaces
- MuRIL, a pretrained transformer model for contextual text representations.

F. Algorithm for Proposed Model

Input: *Dataset* $D = \{(x_i, y_i)\}_{\{i=1\}}^{\{N\}}$, *where xi represents Marathi lyrics and yi represents genre labels*

Output: Predicted labels yˆi and evaluation metric

Step 1: Dataset Preparation

Extract input texts and labels as

$$X = \{x_i\}, \qquad Y = \{y_i\}$$

Step 2: Text Preprocessing

Apply cleaning and normalization to each input lyric.

$$x_i' = \phi(x_i)$$

Step 3: Train-Test Split

Divide the dataset into training and testing sets:

$$\left(X_{\{train\}}, X_{\{test\}}, Y_{\{train\}}, Y_{\{test\}}\right) = Split(X', Y)$$

Step 4: Feature Extraction and Model Training

If model = NB, LR, Linear SVC:

- Apply TF-IDF

$$X_{train}^{v} = V(X_{train})$$
$$X_{test}^{v} = V(X_{test})$$

- Train the model:

$$M = Train\left(X_{\{train\}}^{\{(v)\}}, Y_{\{train\}}\right)$$

Else if model = MuRIL

- Tokenize the text:

$$X_{\{test\}}^{\{(t)\}} = Tokenize\left(X_{\{test\}}\right)$$

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

- Fine-tune the model:

$$M = FineTune\left(X_{\{train\}}^{\{(t)\}}, Y_{\{train\}}\right)$$

Step 5: Prediction

For ML models: $\{\hat{Y}\} = M\left(X_{\{test\}}^{\{(v)\}}\right)$

For MuRIL: $\{\hat{Y}\} = M\left(X_{\{test\}}^{\{(t)\}}\right)$

The algorithm presents a multi-class classification framework for Marathi folk song lyrics. The process begins with preprocessing the input text, including cleaning and chunking to improve data representation. For classical machine learning models, TF-IDF is used to convert text into numerical feature vectors, which are then used to train Logistic Regression, Naive Bayes, and Linear SVC classifiers.

In parallel, a transformer-based approach using MuRIL processes tokenized text to generate contextual embeddings, which are fine-tuned for genre classification. Predictions are generated on the test dataset, and model performance is evaluated using standard metrics such as accuracy, precision, recall, and F1-score.

IV. EXPERMENTS AND RESULTS

This section presents the evaluation of the proposed Marathi folk song classification system. The performance of four models-Naive Bayes, Logistic Regression, Linear SVC, and the transformer-based MuRIL model-is examined and compared. The objective is to analyze how these different approaches perform on a relatively small, domain-specific dataset and determine the most suitable method for this task.

For accessibility, the Marathi Folk Song dataset is available on both IEEE Dataport and Kaggle platforms [13], [14].

A. Training

The experiments were conducted on a CPU-based system with limited computational resources. Classical machine learning models, including Logistic Regression, Naive Bayes, and Linear SVM, were trained using TF-IDF feature representations. The TF-IDF vectorization resulted in feature vectors of dimensionality 2753.

Logistic Regression and Naive Bayes were implemented using default hyperparameters, while Linear SVM was configured with a regularization parameter $C = 1.0$ and class balancing enabled.

For deep learning, the MuRIL (Multilingual Representations for Indian Languages) transformer model (\texttt{google/muril-base-cased}) was fine-tuned for the multi-class classification task. The model was trained using the AdamW optimizer with a learning rate of 2×10^{-5}, a batch size of 8, and for 3 epochs. Input sequences were tokenized using a maximum sequence length of 128 tokens.

B. Performance Evaluation

Table~1 shows that Linear SVC achieved the best overall performance among the evaluated models, with a test accuracy of 72.50\% and the highest F1-score of 59.85\%. Logistic Regression also performed competitively, but with slightly lower generalization. Naive Bayes produced substantially weaker results, indicating that the strong feature independence assumption is not well suited to Marathi lyrical text. Although the training accuracy for MuRIL is not reported due to computational constraints, and evaluation is based on test performance. It provides contextual representations, its performance remained lower than the best classical models under the present experimental conditions. This suggests that, for low-resource genre classification tasks with limited training data, TF-IDF combined with linear classifiers remains a strong and efficient baseline.

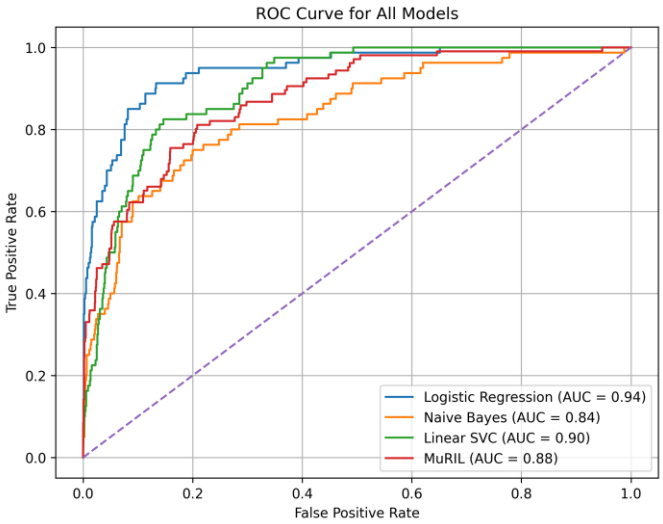

Fig. 3 ROC Curve for Selected Models

Fig. 3. The Receiver Operating Characteristic (ROC) curves are used to evaluate the discriminative ability of the classification models across all classes. Since the problem involves multi-class classification, a One-vs-Rest (OvR) strategy is adopted,. where each class is evaluated against all other classes independently

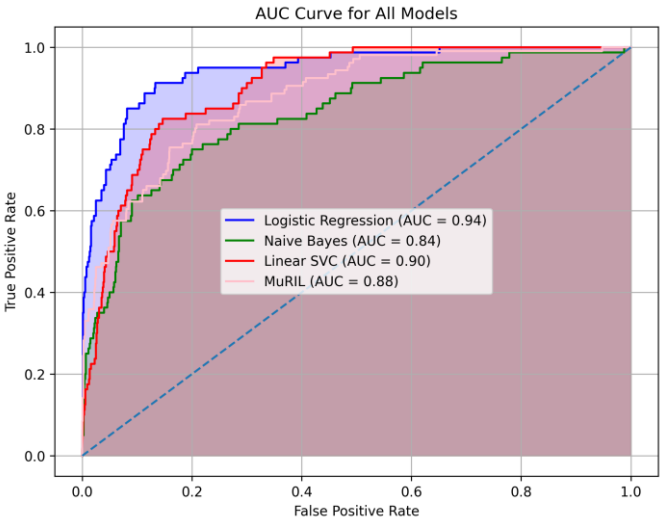

Fig. 4. AUC Curve for Selected Models

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

TABLE 1. Performance Comparison of Selected Models

Models	Train Acc.	Test Acc.	Precision	Recall	Specificity	F1-Score
Logistic Regression	96.21	68.75	59.51	59.97	96.93	57.48
Naive Bayes	45.43	40.00	62.76	11.11	92.20	8.45
Linear SVC	98.66	72.50	72.50	59.67	97.05	59.85
MuRIL	-	48.11	48.11	17.42	94.21	11.38

Fig.4. Area Under the Curve (AUC) provides a single scalar value summarizing the overall performance of the model. Higher AUC values indicate better ranking ability of the classifier in distinguishing between classes

The ROC-AUC analysis highlights that model performance should not be judged solely based on accuracy, especially in the presence of class imbalance, as ROC-AUC captures the underlying separability of classes more effectively.

Fig.5. The confusion matrix for the Linear SVC model shows that most predictions are correctly classified, especially for the more frequent classes, as seen along the diagonal. Some errors occur in the less represented classes, which suggests that class imbalance affects the model's ability to generalize across all classes.

A noticeable gap is observed between training and test accuracy, particularly for Linear SVC and Logistic Regression. For example, Linear SVC achieves a training accuracy of 98.66% but a test accuracy of 72.50%. This indicates that although the model fits the training data effectively, its generalization is constrained by the relatively small dataset size, sparse TF-IDF feature space, and class imbalance. The results suggest that the learned decision boundaries capture dominant lexical patterns strongly, but minority classes remain difficult to separate reliably.

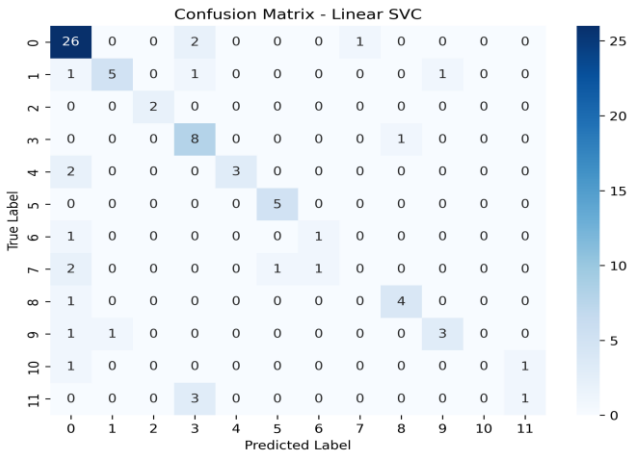

Fig. 5. Confusion Matric - Linear SVC

C. *Computational Cost*

Table 2 shows that traditional models like Naïve, Logistic Regression, and Linear SVC have low computational cost with fast training times due to simple TF-IDF features. Naive Bayes is the most efficient, while Linear SVC balances speed and performance. In contrast, MuRIL has a higher computational cost due to its deep architecture and hardware-dependent training requirements.

Although MuRIL introduces contextual text representation, its performance remains limited in the present study. The relatively low macro-level metrics suggest that the model is biased toward dominant classes, which is likely caused by class imbalance, small dataset size, and limited fine-tuning duration. This indicates that transformer-based models may require larger and more balanced datasets to fully exploit their representational advantages in low-resource Marathi text classification.

TABLE 2. Model Configuration and Training Time

Model	Parametrs and Vectorization	Time(s)
Logistic Regression	C = 1.0, TF-IDF (2753)	0.175
Naive Bayes	α = 1.0, TF-IDF	0.009
Linear SVC	C = 1.0, TF-IDF	0.035
MuRIL	lr=2e −5 , Tokenizer (128)	-

V. CONCLUSION

This work presents a machine learning-based approach for classification of Marathi folk songs, supported by the development of a custom annotated dataset. A comparative evaluation of Naive Bayes, Logistic Regression, Linear SVC, and a transformer-based model was conducted under a consistent experimental setup.

Among the evaluated models, Linear SVC achieved the best overall performance, demonstrating its effectiveness for short-text classification using TF-IDF features. Logistic Regression

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

also showed competitive results, while Naive Bayes exhibited limited predictive capability. Although the transformer-based MuRIL model provides contextual representation, its performance was constrained by the relatively small dataset size and higher computational requirements.

Overall, the results indicate that traditional machine learning approaches remain strong and efficient solutions for low-resource language tasks. This work contributes toward the organization and preservation of regional cultural content and provides a foundation for future improvements, including larger datasets, improved handling of class imbalance, and more advanced multimodal capabilities.

ACKNOWLEDGMET

The authors would like to sincerely thank everyone who provided the necessary support and infrastructure to complete the research.

REFERENCES

[1] Y. Kuwahara and H. Takahashi, “Quantitative analysis of traditional folk songs from shikoku district,” in Proc. Int. Soc. for Music Information Retrieval Conf. (ISMIR), 2018, pp. 774–781

[2] 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), 2025, pp. 1–6.

[3] H. Li and Y. Zhang, “Research on music feature extraction and machine learning classification algorithm for anhui folk songs,” in Proc. Int. Conf. on Artificial Intelligence and Big Data. IEEE, 2019, pp. 123–128.

[4] C. McKay and I. Fujinaga, “Text-based classification of western folk music lyrics,” Journal of New Music Research, vol. 46, no. 3, pp. 232–245, 2017.

[5] I. Karydis and S. Theodoridis, “Multimodal folk music classification using canonical correlation analysis,” Signal Processing, vol. 174, pp. 107–118, 2020.

[6] S. Karami and M. Ahmadi, “Pmg-net: Persian music genre dataset and lstm classifier,” in Proc. Int. Conf. on Signal Processing and Multimedia. ACM, 2021, pp. 89–96.

[7] L. Zhao and X. Chen, “Folk melody generation based on cnn-bigru and self-attention,” Journal of Information Hiding and Multimedia Signal Processing, vol. 12, no. 2, pp. 77–86, 2021.

[8] F. Wang and J. Liu, “Hybrid cnn and logistic regression approach for folk song classification,” Multimedia Tools and Applications, vol. 81, no. 3, pp. 2159–2173, 2022.

[9] K. Choi, G. Fazekas, and M. Sandler, “Automatic music genre classification using convolutional recurrent neural networks,” IEEE Transactions on Audio, Speech, and Language Processing, vol. 27, no. 9, pp. 1749– 1761, 2019.

[10] C. Xu and Y. Yang, “Attention mechanisms for music genre classification,” Neurocomputing, vol. 356, pp. 89–100, 2019.

[11] S. Oramas, O. Nieto, and X. Serra, “Multimodal deep learning for music genre classification,” ACM Transactions on Multimedia Computing, vol. 16, no. 3, pp. 1–23, 2020.

[12] S. Oramas, O. Nieto, and X. Serra, “Multimodal deep learning for music genre classification,” ACM Transactions on Multimedia Computing, vol. 16, no. 3, pp. 1–23, 2020.

[13] "Marathi Folk Songs Dataset," IEEE Dataport. Available: https://ieee-dataport.org/documents/marathi-folk-songs-dataset

[14] [Y] Y. Rajput, "Marathi Folk Songs Dataset," Kaggle. Available: https://www.kaggle.com/datasets/yogeshwarirajput/marathi-folk-songs-dataset