

RESEARCH ARTICLE



Classification of Assamese Folk Songs' Melody using Supervised Learning Techniques

OPEN ACCESS**Received:** 17-08-2022**Accepted:** 11-12-2022**Published:** 09-01-2023**Krishnarjun Bora¹, Manash Pratim Barman^{2*}, Arnab N Patowary³, Toralima Bora⁴**¹ Research Scholar, Department of Statistics, Dibrugarh University, Assam, India² Associate Professor, Department of Statistics, Dibrugarh University, Assam, India³ Assistant Professor, Department of Statistics, Assam Agricultural University, Assam, India⁴ Assistant Professor, Centre for Computer Science and Applications, Dibrugarh University, Assam, India

Citation: Bora K, Barman MP, Patowary AN, Bora T (2023) Classification of Assamese Folk Songs' Melody using Supervised Learning Techniques. Indian Journal of Science and Technology 16(2): 89-96. <https://doi.org/10.17485/IJST/v16i2.1686>

* **Corresponding author.**

manash_stats@dibru.ac.in

Funding: None

Competing Interests: None

Copyright: © 2023 Bora et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Published By Indian Society for Education and Environment ([iSee](https://www.isee.org/))

ISSN

Print: 0974-6846

Electronic: 0974-5645

Abstract

Objectives: A melody is made up of several musical notes or pitches that are joined together to form one whole. This experiment aims to develop four models based on the Mel-frequency Cepstral Coefficients (MFCC) to classify the melodies played on harmonium corresponding to five different class of Assamese folk Music. **Methods:** The melodies of five different categories of Assamese folk songs are selected for classification. With the help of expert musicians, these melodies are played in harmonium and audio samples are recorded in the same acoustic environment. 20 MFCC's are extracted from each of the samples and classification of the melodies is done using four supervised learning techniques- Decision Tree Classifier, Linear Discriminant Analysis (LDA), Random Forest Classifier, and Support Vector Machine (SVM). **Findings:** The performance of the fitted models are evaluated using different evaluation techniques and presented. A maximum of 94.17% average accuracy score is achieved under Support Vector Machine. The average accuracy scores of Decision Tree Classifier, Linear Discriminant Analysis (LDA), and Random Forest Classifier are 73.58%, 85.58%, and 86.11% respectively. The models are developed based on 250 samples (50 from each type). However, increasing the training sample size, there is a possibility to improve the performances of the other three models also. **Novelty:** The developed approach for identifying the melodies is based on computational techniques. This work will certainly provide a basis for conducting further computational studies in folk music for any community.

Keywords: Assamese Folk Music; Decision Tree Classifier; Linear Discriminant Analysis; Random Forest Classifier; Support Vector Machine

1 Introduction

Folk literature is an integral part of every society. It reflects the emotions, feelings, and experiences of the individuals of a society. Assam in the North-Eastern part of India is well known for its rich and diversified culture and its richness can be easily noticed in the various forms of folk music as the people of the state belong to different tribes and communities. A variety of music forms are observed among the people of Assam which provides a wonderful way of expressing diverse communities as well as their traditions. The tunes of most of the Assamese folk songs are in pentatonic scale similar to other traditional music of Asia like China, Mongolia, etc.

A melody is made up of several musical notes or pitches that are joined together to form one whole. The majority of compositions are made up of several melodies that interact with one another. However, when we refer to a folk song's melody, we really just mean the precise note transitions of the various folk song forms without any accompaniment. The melodies of a specific class of folk songs have a distinctive note transition pattern which makes it different from those of other types. Researches have been carried out to study the melody similarity of Hindustani Classical music where high-level music features have been used.

In this work, melodies of five major varieties of folk songs are selected for identification of the type using short-term features. Although the problem under consideration and the genre recognition problem are somewhat comparable, they are not exactly the same. A genre of a particular class of music is characterized by different elements such as accompaniment, polyphony, etc. Here our main focus is on the melody only, and it is a major reason for not considering live music or pre-recorded songs. On the other hand, only a few types of folk songs such as lokageet, and Bianam are available in recorded format. Since these few types of folk songs are recorded for commercial purposes, rawness is missing in the melodies of the songs. Melodies of the selected folk songs are played in a harmonium with the help of expert musicians and samples are recorded. The mel frequency cepstrum is found to be highly effective in modeling the subjective pitch and frequency content of audio signals as well as recognizing the structure of music signals. Based on Mel-frequency Cepstral Coefficients (MFCC), a number of musicological researches are carried out to identify genre, singer, music style, musical instruments etc. In this experiment, four popular supervised learning models namely Decision Tree Classifier (DTC), Linear Discriminant Analysis (LDA), Random Forest Classifier, and Support Vector Machine (SVM) are developed using the MFCC (Mel-frequency Cepstral Coefficients) features extracted from the audio signals of the melodies and also the identification performance is assessed using different evaluation techniques.

1.1 Brief introduction of the selected types of folksongs is given below

1.1.1 Bihu Nam

The festival of *Bihu* is the national festival of Assam. The most precious *Bohag bihu* is considered as the soul of Assamese culture. As the essence of spring spreads everywhere in nature, the Assamese people start expressing their happiness and joy by singing *bihu nam* (Bihu Song) and dancing *bihu*. The source of these songs is the nature. *Bihu* songs are an indivisible part of *Hunsori* where people sings and dances in the courtyards of the houses and blesses the house owner and his family. The theme of the *bihu nam* reflects love, hope, and anticipation of people for *bihu*. The melodic structure of the *bihu* songs is different from classical or other modern songs.

1.1.2 Bianam

In human life, marriage is a memorable ceremony. The cheerful circumstances of wedding are the source of *bianam* and the songs portray the mixed feeling of people at a wedding. The nature of these songs is spiritual and are performed in various parts like anointing the bride and groom with "Mah Haldi", "Juron", "pani tola" etc. of a marriage ceremony by elderly women.

1.1.3 Ainam

In every society, religious-traditional beliefs are rooted deeply in the human mind. In the ancient period, it was believed by Assamese people that the causes of Smallpox, Chickenpox, etc. were some angry Devine energy, called *Aai* or *Sitala Devi*. To calm *Aai* prayers in rhythm are made by elderly women that are known as *Ainam*. These songs are rich in poetic beauty as well as in literary value.

1.1.4 Kamrupiya Lokageet

From the ancient period of time people of Kamrup area of Assam used to express their feelings and emotions in the form of songs which are popularly known as *Kamrupia Lokageet*. The subject matter is mainly day to day life of the *Kamrupi* people, the bond of a married lady to her family and some are based on religious aspects like stories of Ramayana, Mahabharata etc.

1.1.5 Goalporiya Lokageet

It is a type of folk music sung by the *Goalporiya* language-speaking people of Assam. The prime theme of these songs is pathos in love (*Biroh*) but not all songs are based on this theme. Some are based on agriculture, worshiping of God, philosophy, the life of *Goalporiya* people etc. This type of music has been kept alive by *Koch Rajbonsi* people in their festivals. Protima Pandey Boruah is one noted personality of Assam, who introduces this music genre to the nation.

1.2 Related Works

Music Information retrieval (MIR) which is an interdisciplinary science of retrieving information from music is a rapidly growing field of research with several real-world applications. The growth of data mining techniques and signal processing have made it easy to study the different features of a piece of music. Classification as well as identification of music, namely genre classification, mood identification, instrument identification etc. have become very popular in the field of the musicological research.

In order to identify melody from standard MIDI files, an algorithm, based on Bayesian maximum-likelihood approach and dynamic programming is used that achieves an overall accuracy of 89%⁽¹⁾. Simonetta et.al. develop a computational approach using the symbolic scores of a piece of music for the identification of melody lines⁽²⁾. This approach is based on a Convolutional Neural Network (CNN) estimating the probability that each note in the score belongs to the melody line. CNN is also found to be an effective algorithm in the identification of music genres also. To classify the songs into their respective genre CNN is also on spectrograms extracted from the songs⁽³⁾. MFCC-based CNN and RNN (Recurrent Neural Network) are used to determine the music style where 93.3% classification accuracy is achieved⁽⁴⁾.

In today's machine learning applications, (Support Vector Machine), SVM is found to be one of the best algorithms for solving different types of classification as well as identification problems. In order to classify musical instruments, Prabavathy et. al. use MFCC features and Sonogram features⁽⁵⁾. Two classifiers used in this work are the SVM and k-NN (k Nearest Neighbourhood) where SVM performs better with a 98% accuracy score for classifying the instruments. An experiment is conducted using SVM, KNN, and Naïve Bayes (NB) classifiers for the classification of music genres on the Spotify music dataset⁽⁶⁾. They find that the SVM classifier has the best classification performance with 80% of accuracy, followed by KNN and NB. SVM and NB-based classifier is also used on audio features and lyrical features to automatically classify the mood of a song⁽⁷⁾. Based on MFCC features and Chroma Reduced Pitch (CRP) features of songs, Sangeetha & Nalini⁽⁸⁾ propose an approach for the identification of singers with the help of SVM. They find that MFCC features are more effective in the identification of singers than CRP features. A classification accuracy of 92.0% is achieved with MFCC-based SVM while only 44.0% is achieved with CRP-based SVM. SVM is also found to be an effective classification technique for analysing RNA-seq datasets from the mosquito *Anopheles gambiae* to predict Malaria Vector Gene Expression where up to 98% of accuracy score is achieved⁽⁹⁾.

2 Methodology

2.1 Collection of Data

The data for this study is the melodies of the selected types of folk songs. For collecting samples for the study four folksingers were contacted and briefed about our objectives. All of them consented to record the songs sung by them. They are requested to sing the songs having variations in their melodic structure. In the next stage, with the help of an expert harmonium player, the melodies of the already recorded songs are played in harmonium and recorded in the .wav format at a sampling rate of 44100 Hz under the same acoustic environment. To maintain homogeneity, all the melodies are played in the same key or scale (Note A) of the same harmonium. A total of 60 songs are recorded in this experiment, out of which, 15 songs from *bihu nam*, 10 songs from *bianam*, 10 songs from *ainam*, 13 songs from *gualporia lokageet*, and 12 songs from *Kamrupia lokageet*. Songs are played in such a manner so that the repetition of the melody is not occur. Only the distinct unique melody of the “*Sthai*” or “*Mukhra*” and of the “*Antara*” are recorded. Therefore, the length of the songs are ranges from 30 sec. to 2 minutes. Each of the songs is then cropped into 10 sec. interval and 20 MFCC's are extracted from each of the sample. 50 samples of length 10 sec. from each class of songs (total 250) are used in the analysis. Since only the unique and distinct melody are considered during recording, repetition of the same melody are not happening in the samples of length 10 sec.

Including the extraction of the features, all the analysis is conducted in Python programming languages. From each of the audio samples, spectrogram is generated using the Matplotlib library which is a numerical extension of NumPy, a fundamental package for scientific computing. Features extraction from the spectrograms is done using librosa, a Python library for music and audio analysis.

2.2 Generation of the Spectrogram

A spectrogram is a visual representation of signal strength over time at various frequencies present in a particular waveform. The horizontal axis represents time while the vertical axis is used to represent the frequency of the signal. A third dimension, color, is used to describe the amplitude (or energy) of a particular frequency at a particular time. In this study, MEL (having MEL frequency bins on the y-axis) spectrogram is extracted from each of the samples. In Figure 1, spectrograms extracted from one audio signal of each type of folk song are shown.

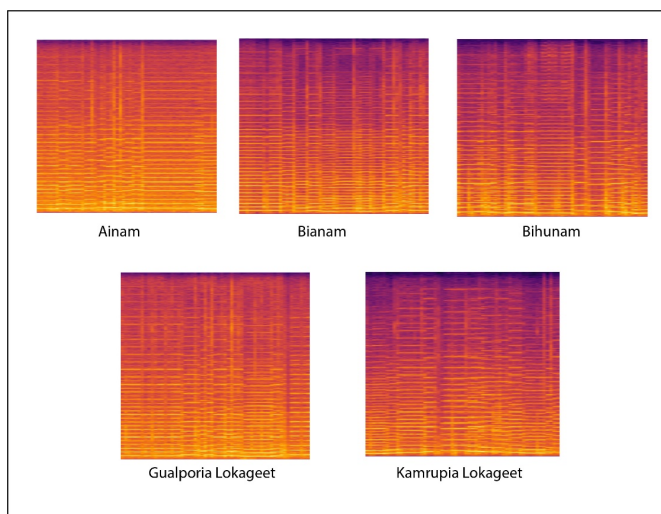


Fig 1. Spectrogram extracted from a single audio sample from each type of folk song

2.3 Mel- frequency Cepstral Coefficients

Mel-frequency Cepstral Coefficients (MFCC) that are introduced by Davis and Mermelstein in 1980, are a set of features (usually 10 to 20) that have wide use in automatic speech and speaker recognition. The representation of a short-term power spectrum of a sound is known as mel-frequency cepstrum (MFC). The coefficients that collectively make up an MFC are called Mel-Frequency Cepstral Coefficients. In other words, these are the cepstral representation of a signal where the frequency bands are distributed according to mel-scale. The MFCCs have been applied in a wide range of audio analyses, and have shown good performance compared to other features^(4,5,8).

2.4 Decision Tree Classifier

A decision tree is a tree-structured multistage classification strategy where each internal node represents a test on an attribute. Each branch represents an outcome of the test. Class label or dependent variable is represented by each leaf node (or terminal node). For a tuple X, the attribute values of the tuple are tested against the decision tree. A path is discovered from the root to a leaf node that holds the tuple's class prediction. A decision tree can be easily converted into a classification rule. As a predictive modeling approach, it has wide applications in statistics, machine learning as well as data mining.

2.5 Linear Discriminant Analysis

Linear Discriminant Analysis or discriminant function analysis is a multivariate technique to find a linear combination of features that characterizes or classifies two or more sets of objects or events. It is also used as a dimensionality reduction technique as a pre-processing step for machine learning. The term discrimination was introduced by R. A. Fisher in the first modern treatment of separative problems⁽¹⁰⁾.

2.6 Support Vector Machines

Support Vector Machine (SVM) is one of the robust and accurate classification algorithms in a wide range of machine learning applications. It was developed by Vladimir Vapnik with his colleagues at AT&T Bell Laboratories. SVM has a sound theoretical

foundation and needs only a dozen training examples. It determines the best hyperplane in the input space that differentiates between the classes. Originally this algorithm was developed for binary classification problems. In the case of multiclass classification, the problem is reduced to multiple binary classification problems.

2.7 Random Forest Classifier

A random forest is a popular supervised machine learning algorithm that can be used for both classification and regression problems. It is an ensemble classification method that produces multiple classifiers using a randomly selected subset of training samples to solve complex problems. Random forest contains a number of decision trees on various subsets of the given dataset and takes the average to improve the accuracy of prediction. Instead of relying on one decision tree, it uses the prediction from each tree, and based on the majority of predictions, the final prediction is made. This algorithm is found to have wide applications in the fields like Banking, Marketing, Medicine, remote sensing, etc.

2.8 Evaluation of the Model Performance

After fitting the models following measures are adopted to evaluate the performances of the fitted model.

Accuracy Score: It is the percentage of correctly classified test samples. The accuracy score is calculated by the formula-

$$Accuracy\ Score = \frac{TP + TN}{TP + TN + FP + FN}$$

Where, TP= True Positive, FP= False Positive

TN= True Negative, FN= False Negative

F-1 Score: It is the harmonic mean of precision and recall value where

$$Precision = \frac{TP}{TP+FP} \quad \text{and} \quad Recall = \frac{TP}{TP+FN}$$

Precision also called positive predictive value gives the value of the fraction of relevant instances among the retrieved instances and recall or sensitivity gives the value of the fraction of relevant instances that were retrieved. Where both the error false positive and false negative are equally serious, a decision is taken on the basis of the F-1 score.

3 Results and Discussion

A multivariate normality test is performed using the Henz-Zirkler test⁽¹¹⁾ on the extracted data and it is found the data are not normal. In order to train the models, 70% of the sampled data are selected randomly and the remaining 30% has been used for testing purposes. Since the splitting of the data is done in a random manner, variation in the accuracy score of the same model is obvious. To fix this problem, we have repeated the process of calculating the accuracy score 100 times for all four models and finally, the average accuracy score along with a 95% confidence interval is calculated which is visualized in Table 1.

Table 1. Average accuracy score of the fitted models along with 95% confidence interval

Fitted Models	Average Accuracy Score	95% Confidence Interval
Decision Tree Classifier	73.58%	71.54% - 75.62%
Linear Discriminant Analysis	85.58%	83.65% - 87.51%
Support Vector Machine	94.17%	92.67% - 95.67%
Random Forest Classifier	86.11%	84.55% - 87.67%

In this experiment, the performance of the Support Vector Machine in the identification of the melodies is quite good with a maximum average accuracy score i.e. 94.17% while the decision tree classifier predicts with the lowest accuracy score of 73.58%. The performance of Linear discriminant analysis and Random forest classifier are very close to each other.

In most of the musicological research work of genre classification or melody classification, the features are extracted from the original version of the recorded song or by using the written scores of the piece of music. Since the folk songs are not available in recorded form and neither their note transitions are available in written format, the melodies are recorded one by one from the primary sources. This characteristic makes this experiment unique from the others. For extraction of the features, instead of singing, all the melodies of the different types of folk songs are played on the same harmonium at the same note 'A' so that homogeneity can be maintained in the other characteristics. This is not possible in the case of the original recorded version of songs. Though some recorded versions of some folk songs are available, rawness is lacking due to the use of different modern musical instruments.

To identify melody, Standard MIDI files, as well as symbolic scores, are used^(1,2) whereas our approach is based on MFCC features. Several pieces of research are performed that use MFCC in order to identify Music style, Musical instruments as well as Singer^(4,5,8). In the identification of singers, MFCC features are found to be better than the CRP features⁽⁸⁾. In this experiment also results indicate that even if the songs are played in the same instrument at the same key, MFCC-based SVM is a very effective model in the identification of the melody correctly.

For a specific randomly chosen test sample, the results are visualized in the confusion matrices in Figure 2.

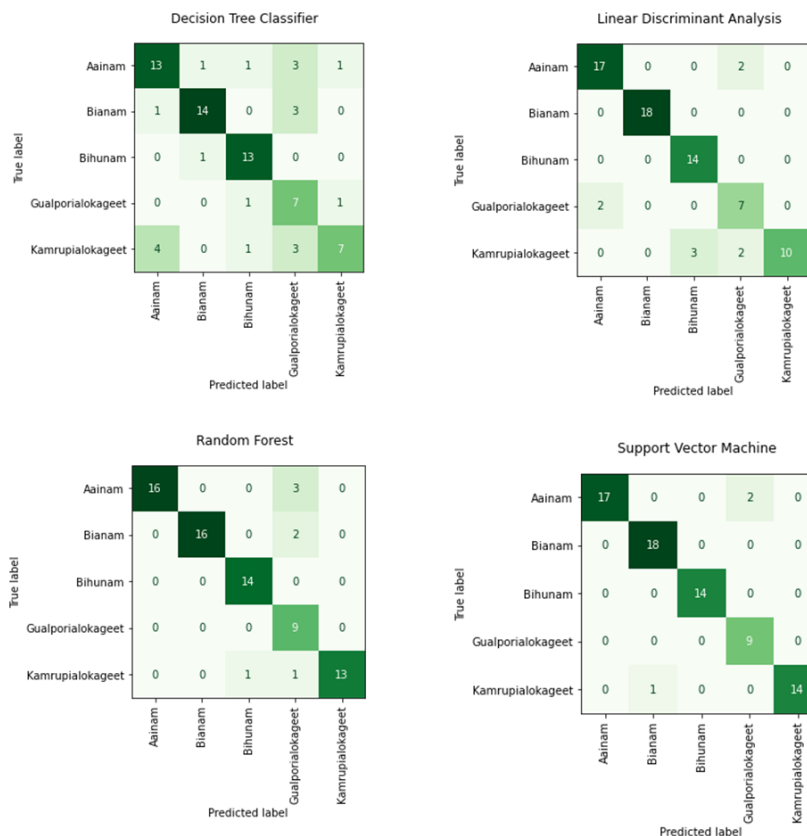


Fig 2. Confusion Matrices for each of the fitted Models

Out of the 19 *Ainams* in the test set only 13 are correctly identified by the decision tree classifier while in the case of the other three classifiers, this value is quite good. In all the four classifiers, misclassification is occurring in the identification of *Gualporialokageet*. In all four cases, *Gualporialokageet* is misclassified as *ainam*. It may be due to the similarity between the note transitions of these two types of songs. The class prediction errors for the four models are shown in Figure 3.

The classification reports for the four fitted models are shown in Figure 4. From this report, if we compare the performances of the four models, it is found that SVM performs the best compared to all three models. In our problem both false positive as well as false negative are equally sensitive, we can choose such a model which have maximum F-1 scores. From Figure 4 it is clear that SVM gives maximum F-1 scores for all types of folks. In all four models, misidentification is occurring between *Gulaporialokageet* and *Ainam*. Similarly, in the case of the Decision tree classifier, LDA and Random Forest Classifier, *Kamrupia Lokageet* are misidentified as *Gualporia lokageet*. False positive cases are happening in identifying each type of melody in the case of the decision tree classifier. The linear discriminant analysis identifies only two classes i.e. *Bianam* and *Kamrupialokageet* with zero false positive cases. If we observe this figure in Radom Forest Classifier, it has three classes namely *Ainam*, *Bianam* and *Kamrupia lokageet* with zero false positive cases. SVM identifies three classes *Ainam*, *Bihunam* and *Kamrupia Lokageet* with zero false positive cases. However, the rate of false negative cases is higher in the Random Forest Classifier compared to SVM.

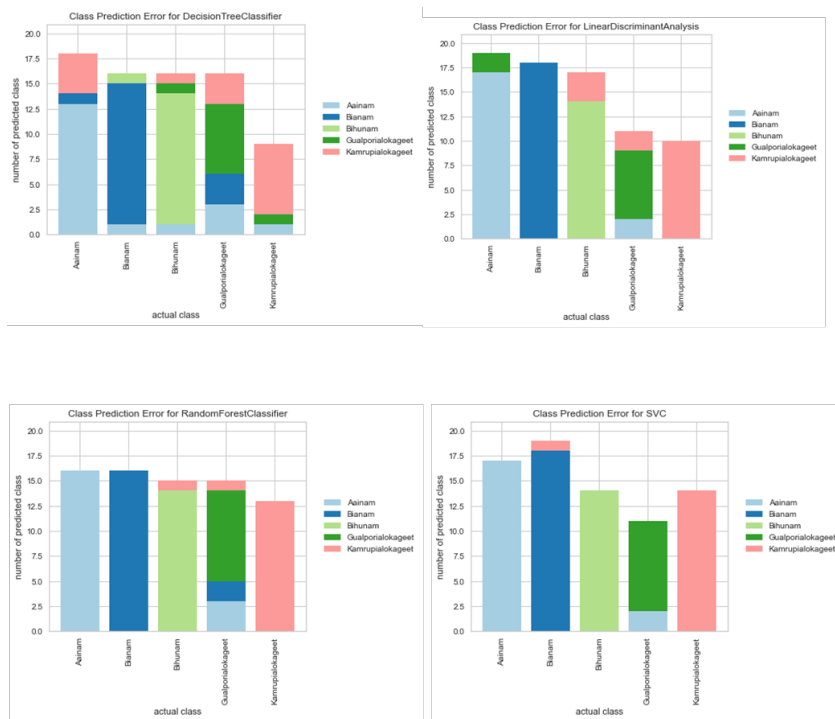


Fig 3. Class prediction error for each of the Models

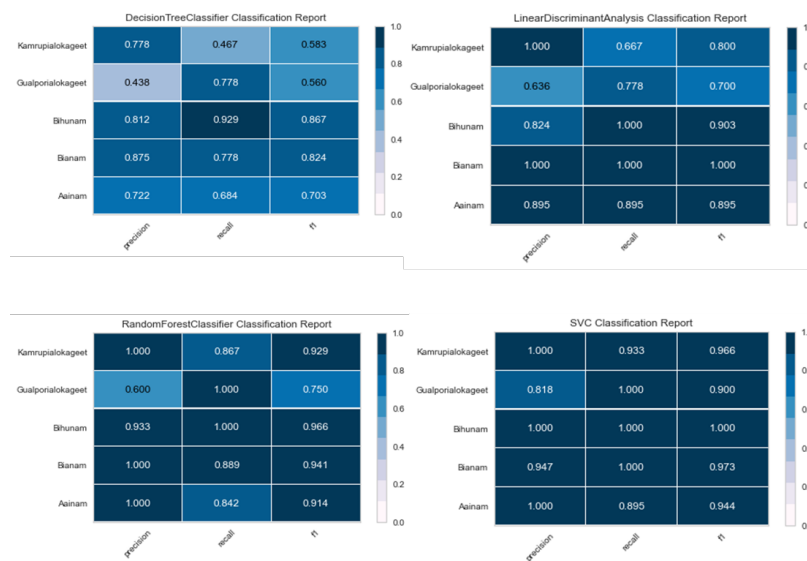


Fig 4. Classification Reports of the four Trained Classifiers

4 Conclusion

Identification of folk song melody by hearing it is not an easy task. Human listeners having knowledge about the particular folk music can readily identify the melody of that type of folk songs. In this experiment, an attempt has been made to develop a computational approach for the identification of the folk song melody using short-term features instead of using the note transition of the melody. This experiment will definitely provide a fair comparison between the popular classification algorithms DTC, LDA, SVM, and Random Forest Classifier in the identification of the solo folk song melody.

One of the biggest challenges in conducting this experiment was the collection of the audio samples for each type of folk under consideration. Since hearing the songs, the melody was played in harmonium to obtain the final audio samples, the process was very tedious. Results from this experiment indicate that the performance of SVM is quite good compared to the other three models. However increasing the training sample size, there is a possibility to improve the performances of the other three models also. The classification accuracy does not depend only on the features selected to fit the models but also on the size of the training set. This work will definitely provide a direction to study the important features in the identification of solo folk song melody and also the optimum size of the training data for each of the models.

5 Acknowledgement

We would like to express our sincere gratitude to the folk music experts Mr. Ajit Bora, Mr. Ghana Kanta Saikia, Mr. Sunil Bhuyan, Mrs. Mamaoni Saika, and Mr. Ranjit Bora for their active participation in the data collection process of the experiment. Without their contribution, it would not be possible to conduct this study.

References

- 1) Jiang Z, Dannenberg RB. Melody Identification in Standard MIDI Files. *Proceedings of the 16th Sound & Music Computing Conference*. 2019. Available from: <https://www.cs.cmu.edu/~rbd/papers/melody-identification-midi-smc2019.pdf>.
- 2) Simonetta F, Chacón CC, Ntalampiras S, Widmer G, G. A Convolutional Approach to Melody Line Identification in Symbolic Scores. In: Proceedings of the 20th International Society for Music Information Retrieval Conference. 2019;p. 924–931. Available from: <https://archives.ismir.net/ismir2019/paper/000114.pdf>.
- 3) Pelchat N, Gelowitz CM. Neural Network Music Genre Classification. *Canadian Journal of Electrical and Computer Engineering*. 2020;43(3):170–173. Available from: <https://doi.org/10.1109/CJECE.2020.2970144>.
- 4) Wang W, Sohail M. Research on Music Style Classification Based on Deep Learning. *Computational and Mathematical Methods in Medicine*. 2022;2022:1–8. Available from: <https://doi.org/10.1155/2022/3699885>.
- 5) Prabavathy* S, Rathikarani V, Dhanalakshmi P. Classification of Musical Instruments using SVM and KNN. *International Journal of Innovative Technology and Exploring Engineering*. 2020;9(7):1186–1190. Available from: <http://doi.org/10.35940/ijitee.G5836.059720>.
- 6) Setiadi DRIM, Rahardwika DS, Rachmawanto EH, Sari C, Irawan A, Kusumaningrum C, et al. Comparison of SVM, KNN, and NB Classifier for Genre Music Classification based on Metadata. *International Seminar on Application for Technology of Information and Communication*. 2020;p. 12–16. Available from: <https://doi.org/10.1109/iSemantic50169.2020.9234199>.
- 7) Tan KR, Villarino ML, Maderazo C. Automatic music mood recognition using Russell's twodimensional valence-arousal space from audio and lyrical data as classified using SVM and Naïve Bayes. *IOP Conference Series: Materials Science and Engineering*. 2019;482:012019. Available from: <https://doi.org/10.1088/1757-899X/482/1/012019>.
- 8) Sangeetha R, Nalini NJ. Singer Identification Using MFCC and CRP Features with Support Vector Machines. In: Das, A, Nayak, J, Naik, B, et al., editors. *Computational Intelligence in Pattern Recognition*;vol. 999. Springer Singapore. 2020;p. 295–306. Available from: https://doi.org/10.1007/978-981-13-9042-5_25.
- 9) Arowolo MO, Adebisi MO, Nnodim CT, Abdulsalam SO, Adebisi AA. An Adaptive Genetic Algorithm with Recursive Feature Elimination Approach for Predicting Malaria Vector Gene Expression Data Classification using Support Vector Machine Kernels. *Walailak Journal of Science and Technology (WJST)*. 2021;18(17). Available from: <https://doi.org/10.48048/wjst.2021.9849>.
- 10) Johnson AR, Wichern DW. Applied Multivariate Statistical Analysis. *Biometrics*. 1988;44(3):920.
- 11) Henze N, Zirkler B. A class of invariant consistent tests for multivariate normality. *Communications in Statistics - Theory and Methods*. 1990;19(10):3595–3617. Available from: <https://doi.org/10.1080/03610929008830400>.