Songklanakarin J. Sci. Technol.
45 (3), 399–406, May – Jun. 2023

http://www.sjst.psu.ac.th

Original Article

A computational approach for identifying assamese folk music instruments

Krishnarjun Bora[1], Manash Pratim Barman[1*], and Arnab N. Patowary[2]

[1] *Department of Statistics, Dibrugarh University, Dibrugarh, Assam, 785004 India*

[2] *Department of Statistics, Assam Agricultural University, Jorhat, Assam, 785013 India*

Received: 11 March 2022; Revised: 6 October 2022; Accepted: 16 March 2023

Abstract

The classification of musical instruments by using a computational technique is a very challenging task. The developments in signal-processing and data-mining techniques have made it feasible to analyse the many musical signal characteristics, which is essential for resolving the classification issues in music. In this work, 12 popular Assamese folk music instruments were selected for identification. Twelve musicians played the instruments and audio samples were recorded, different instantaneous features were extracted, and an effort has been made to identify those instruments using three popular classification techniques - Decision Tree Classifier (DTC), Support Vector Machine (SVM), and Linear Discriminant Analysis (LDA). A performance-based comparison was made among the three classifiers. The proposed sets of features enabled the DTC, SVM and LDA models to achieve average accuracy ratings of 86.9%, 90% and 92.2% respectively. Regarding the performances of the three fitted models in identifying instrumental sounds, this study offers a valid comparison.

Keywords: musical instrument, machine learning, decision tree classifier, support vector machine, linear discriminant analysis

1. Introduction

India is famous for its diversified cultures and traditions. Each part of the country has its own unique culture and traditions, and each culture is conspicuously visible in the various art forms. Assam, situated in the northeastern part of India, has rich cultural resources, including different kinds of traditional music, as the people in this state belong to different tribes and communities. Assamese people practise a range of musical genres, which offers a beautiful means of expressing the varied communities and their traditions. A number of musical instruments are used in performing the different kinds of music prevailing in Assam. Krishnaswami (1971) classified the Indian musical instruments to categories named TATA (stringed instruments), SUSHIRA (wind instruments), AVANADH (percussion instruments like drums covered with skins) and GHANA (ideophones, instruments which are struck against each other like cymbals). All these four kinds of musical instruments are used in the performance of Assamese folk music. The raw materials used to make these instruments are bamboo, leather, soil, buffalo horns, strings, wood, bottle gourds, etc.

This study aimed to develop three models for identifying Assamese folk music instruments, using Decision Tree Classifier (DTC), Support Vector Machine (SVM), and Linear Discriminant Analysis (LDA), and to analyze their performances. Twelve of the most well-known traditional Assamese musical instruments were chosen for classification, and are listed below with their categories.

Dotara, *Ananda lohori*, *Ektara* and *Dogor* are widely played instruments in one of the important type of Assamese folk called *Lokageet*. *Bihu* is the prime festival of Assam. Popular folk music instruments played in *Bihu* songs are *Dhol*, *Pepa*, *Xutuli*, *Gogona* and *Bahi*. *Dhol* is a two-faced drum played with a single bamboo stick. The main part of *Pepa* is specially made from bamboo but with a buffalo horn attached to it, making the sound very unique. *Xutuli* is a wind instrument made from clay or the lower end of a bamboo tree. *Gogona* is

*Corresponding author
Email address: manash_stats@dibru.ac.in

a very unique instrument made of bamboo similar to a jaw harp, which has a vibrating reed. The wide side of *Gogona* is gripped with the lips and the free end is struck repeatedly with the fingers to produce sound. *Bahi* is a kind of flute which is made from bamboo. *Nagara* (or *Negara*) is a combination of two single-faced drums played using two bamboo sticks. It is the major instrument in the performance of spiritual songs called *Negara Naam*. *Khol* (or *mridanga*) is another popular two-faced drum played with free hands. This instrument is played with the religious, spiritual songs known as *Mohapurusia Sangeet*.

2. Related Works

Music data analysis and retrieval has become a very popular research field in the recent years. Previously the clustering and classification of music were performed subjectively using specified features of samples. The rapid progress in signal-processing and data-mining techniques has made it possible to study the computable features of musical signals, which plays an important role in solving the classification and identification problem in music. K-means clustering is a widely used technique to solve clustering and classification problems in music. For the classification of Indonesian traditional music, Jondya and Iswanto (2017) selected the essential musical features using principal component analysis and found four distinct clusters of the selected songs using the K-mean clustering algorithm. Similar work is found in clustering classical, rap, metal and Indian music (Sen, 2014).

Deng, Simmermacher, and Cranefield (2008) studied the features of musical instruments and classified them using K- nearest neighbour algorithm. In this work, PCA and Isomap were used to explore sparse feature collections and examine the residuals of the chosen dimensionality to estimate how many features should be included in a subset.

Marques and Moreno (1999) classified eight musical instruments using two classification algorithms, Gaussian Mixture Models and Support Vector Machines. Here, the SVM gave the best results with an overall error rate of 30% when classifying segments of 0.2 seconds of sound. This work is one of the first applications of SVM to music classification.

Another classification of musical instrument timbres was done by Agostini, Longari and Poolastri (2003) using 117 spectral features. The performances were assessed for SVM, k-NN, Canonical Discriminant Analysis, and Quadratic Discriminant Analysis. SVM and Quadratic Discriminant Analysis performed the best. Tzanetakis and Cook (2002) used a Gaussian mixture model (GMM) and K-means for audio signal based musical genre classification. Hidden Markov Model (HMM) has been found to be one of the successful statistical techniques in solving classification and identification problems. HMM-based classifier was used by Kim, Moreau and Sikora (2004) for speaker recognition and sound classification. Comparing the MFCC and MPEG-7 audio features Xiong, Radhakrishnan, Divakaran and Huang (2003) used HMM, K-NN, GMM, AdaBoost, and SVM techniques for sports audio classification.

In today's machine learning applications, SVM has been among the best algorithms for solving different types of classification problems. For classification of the bass playing style Abeßer, Lukashevich, and Bräuer (2012) used three approaches based on SVM, Classification and Regression Tree (CART), and two pattern similarity measures, with the highest accuracy of 64.8%. Arowolo, Adebiyi, Nnodim, Abdulsalam and Adebiyi (2021) used SVM for analysing RNA-seq dataset from the mosquito Anopheles gambiae to predict Malaria Vector Gene Expression where up to 98 % accuracy was achieved.

Ünal, Bozkurt, and Karaosmanoğlu (2014) used symbolic data for the classification of Turkish makam music. Here, in the first level, the information provided by the n-gram likelihood of the symbolic sequences was used. Then a more detailed identification was achieved using statistical features related to the content of the piece, such as the tonic note, the average pitch level for local excerpts, and the overall pitch progression.

3. Data

All the raw and solo audio samples for each instrument were collected from primary sources. Twelve expert musicians were contacted and briefed about the study goals. All of them consented to play the instruments. Different instruments of the same type may produce different sounds due to differences in size, shape, tuning and build quality. Therefore, we collected samples from more than one instrument of the same type. Thus, for the collection, 35 instruments were used. For each instrument type, musicians were asked to play 50 different melodies or beats, and the sounds were recorded for 20 sec windows in WAV format at 44,100 Hz sampling rate, under the same acoustic environment and otherwise similar conditions. One of the serious obstacles in the data collection process was that due to the limited use of some instruments like *Gogona* or *Xutuli* in the performances, comparatively small numbers of samples were obtained. The number of collected samples for each type of instrument is shown in Table 1.

4. Methodology

All the analyses were performed using Python programming language, including the extraction of features from the audio samples. To generate the spectrograms from each audio sample, we used the Matplotlib library and the numerical extension NumPy, a fundamental package for scientific computing in Python. Extraction of the features from the spectrograms was done using librosa, a Python library for music and audio analysis.

4.1 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 frequencies in the signal. A third dimension, colour, 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) spectrograms were extracted from each of the samples. Spectrograms extracted from one audio signal of each type of instruments are shown in Figure 1.

Table 1. Instruments with their sample sizes

	Instrument	Number of instruments	Sample size
1	*Dotara*	2	50
2	*Ananda Lohori*	2	50
3	*Ektara*	2	50
4	*Bahi*	7	50
5	*Pepa*	3	50
6	*Xutuli*	2	30
7	*Dhol*	3	50
8	*Khol*	3	50
9	*Nagara*	3	50
10	*Dogor*	2	45
11	*Madol*	3	50
12	*Gogona*	3	40
	Total	35	565

4.2 Features extracted from the spectrogram

A brief introduction to the time domain and frequency domain features that have been extracted from each of the spectrograms is given below.

4.2.1 Time domain features

1) Zero Crossing Rate (ZCR)**:** The Zero-Crossing Rate (ZCR) of an audio frame is the rate of sign-changes of the signal during the frame. The ZCR is defined as follows:

$$Z_i = \frac{1}{2W_L}\sum_{n=1}^{W_L} |\ sgn[x_i(n)] - sgn[x_i(n-1)]\ |$$

2) Root Mean Square Energy (RMSE)**:** The energy in a signal is determined as:

$$\sum_{n=1}^{N} |\ x(n)\ |^2$$

Further, the Root Mean Square Value is obtained by:

$$\sqrt{\frac{1}{N}\sum_{n=1}^{N} |\ x(n)\ |^2}$$

It is calculated for all the frames and finally the average and the standard deviation are considered for analysis.

4.2.2 Frequency domain features

1) Mel-Frequency Cepstral Coefficients (MFCC): In sound processing, the representation of a short-term power spectrum of a sound is known as its mel-frequency cepstrum (MFC). The coefficients that collectively make up an MFC are called Mel-Frequency Cepstral Coefficients. These are the cepstral representation of a signal where the frequency bands are distributed according to mel-scale (Weihs, Jannach, Vatilkin, & Rudolph, 2017).

2) Chroma Features**:** The representation of the spectral energy of the 12 pitch classes (C, C#, D, D#, E, F, F#, G, G#, A, A#, B) is termed the Chroma Vector or the Chroma features. Chroma vector coefficients are determined by grouping of short-term windows into 12 bins. Every bin is calculated according to the formula:

$$v_k = \sum_{n \in S_k} \frac{X_i(n)}{N_k}\ ,\ k \in 0,1,2,\ldots,11$$

The respective mean and standard deviation are calculated by aggregating the Chroma vectors across the frames.

3) Spectral Centroid**:** The spectral centroid determines the frequency bin in which the largest spectral energy is concentrated. It is the 'centre of gravity' of the spectrum. The value of spectral centroid, C_i, of the *i*th audio frame is determined by:

$$C_i = \frac{\sum_{k=1}^{wf_L} kX_i(k)}{\sum_{k=1}^{wf_L} x_i(k)}$$

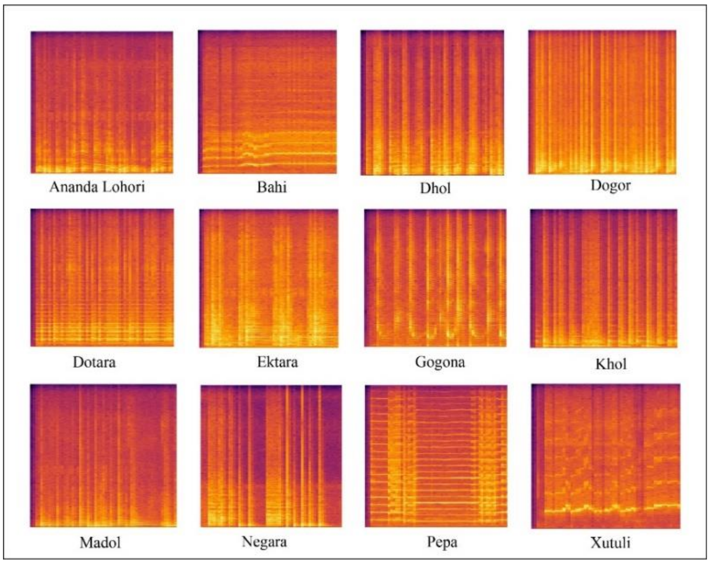

Figure 1. Sample spectrograms for one audio signal of each type of instrument

4) Spectral Band-width**:** Band-width is the difference between the upper and lower frequencies in a continuous band of frequencies. The p^{th} order spectral band-width corresponds to the p^{th} order moment about the spectral centroid (Tjoa, 2017) and is determined by

$$[\sum_{k}(S(k)f(k)-c_i)^p]^{1/p}$$

Here *S(k)* and *f(k)* are respectively the spectral magnitude and the frequency of bin *k*.

5) Spectral Contrast**:** After dividing each frame into a pre-specified number of frequency bands, spectral contrast is defined as the difference between the maximum and minimum magnitudes within each frequency band (Jiang, Lu, Zhang, Tao, & Cai, 2002).

6) Spectral Roll-off**:** Spectral roll-off is the value of the frequency below which a certain amount (85%) of the total energy of the spectrum lies. The user can set this threshold value for the energy.

To have a comparable accuracy score of the fitted models, it is necessary for the dataset to be balanced. In order to make the classes balanced, oversampling was used for *Gogona*, *Dogor* and *Xutuli*.

4.2.3 Models used for identification

The three supervised learning strategies utilised to accomplish the objectives are briefly described below.

1) Decision tree classifier

Decision tree is a popular predictive modelling approach used in statistics, machine learning, and data mining. It 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 assigned to each leaf node (or terminal node). A decision tree can be easily converted into a classification rule. Decision tree learning uses a decision tree as a predictive model, which maps observations about an item to conclusions about the item's target value (Patel & Prajapati, 2018; Wu-Zhou *et al.*, 2008).

2) Support vector machine (SVM)

In different machine learning applications, the support vector machines (SVM) have served as robust and accurate classification algorithms (Vapnik, 1995). This algorithm was developed at the AT&T Bell Laboratories by Vladimir Vapnik with colleagues (Boser, Guyon, & Vapnik, 1992; Drucker, Burges, Kaufman, Smola & Vapnik, 1997). SVM has a sound theoretical foundation and needs only a dozen training examples. SVM determines the best hyperplane in the input space that differentiates between the classes (Arowolo *et al.*, 2021). Originally this algorithm was developed for binary classification problems. For multiclass classification, the problem is reduced to multiple binary classification problems (Duan & Keerthi, 2005).

3) Linear discriminant analysis (LDA)

Linear Discriminant Analysis (LDA) is a very popular multivariate statistical technique for pattern discrimination or classification, as well as for dimensionality reduction problems as a pre-processing step in machine learning. Originally the term discrimination was introduced by R. A. Fisher in the first modern treatment of separative problems (Johnson & Wichern, 2015). In this technique, a linear combination of features is identified that characterizes two or more classes of objects. The resulting linear combination may be used as a linear classifier.

4.2.4 Evaluation of the fitted models

In order to evaluate the performances of the three selected model types, the data were split into two parts, one for training the models and the other for evaluation of the model performance. The following measures were used in the evaluation of the fitted models.

Accuracy Score**:** This is the percentage of correctly classified test samples. It is calculated by the formula (Harikrishnan, 2019).

$$\text{Accuracy score} = \frac{TP+TN}{TP+TN+FP+FN};$$

TP= True Positives, FP= False Positives
TN= True Negatives, FN= False Negatives

ROC Curve**:** An ROC (receiver operating characteristic) is a two-dimensional graph showing the performance of a classification model at all classification thresholds. In this plot the True Positive Rate (TPR) is plotted on the Y axis and the False Positive Rate (FPR) is plotted on the X axis.

$$\text{TPR}=\frac{TP}{TP+FN},\ \text{FPR}=\frac{FP}{FP+TN}$$

This is a useful approach to visualizing, organizing and selecting a classification model based on their performances. The AUC (Area Under the Curve) score indicates the performance of the model.

F-1 score**:** With the help of the predicted outcomes of the fitted models the precision and recall are calculated for each instrument, where

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

In machine learning, precision gives the fraction of relevant instances among the retrieved instances and recall gives the fraction of relevant instances that were retrieved. When both false positive and false negative calls are equally serious, the F-1 score is an effective model evaluation measure, being defined as the harmonic mean of precision and recall.

5. Results and Discussion

In this work 70 % of the total samples were selected randomly and used for training the models; and the remaining

30% of the samples were use for testing. This process was repeated 100 times, so that a confidence interval for the estimates could be constructed. Confusion matrix is considered one of the valid methods for inspecting the performance of the fitted models from a qualitative point of view. For a specific randomly chosen test sample, the model predictions are visualized in three confusion matrices, which are shown in Figure 2.

5.1 Evaluation the models

The accuracy of each fitted model was determined for 100 randomly selected test samples and the 95% confidence intervals for the scores were calculated. The average accuracy along with its 95% confidence interval is presented in Table 2.

The performances of Linear Discriminant Analysis and SVM are quite good, better than that of the Decision Tree Classifier in accuracy. In the work by Marques and Moreno (1999), eight musical instruments were classified using SVM and Gaussian Mixture Model, and the SVM gave the best results with an overall error rate of 30% when classifying 0.2 second segments of sound. In the work by Agostini *et al.* (2003), SVM with RBF kernel gave the best result for recognition of individual instrument, in comparison to the other classifiers that were Quadratic Discriminant Analysis (QDA), Canonical Discriminant Analysis (CDA), and k-nearest neighbours. In the same work, the second-best score was achieved by QDA, with success rates close to those of the SVM. On the other hand, in case of instrument family recognition and sustain/pizzicato classification, QDA surpassed all the other classifiers with its success rate of 81% (Agostini *et al.*, 2003). In the experiment of Setiadi-Trusthi *et al.* (2020), three classifiers namely SVM, KNN, and Naïve Bayes (NB) were used in the classification of music genres of the Spotify music dataset. They found that the SVM classifier had the best classification performance with 80% accuracy, followed by KNN and NB. The model accuracy, however, may vary depending on the issue being investigated. Since every problem has a unique set of features, their amount of information will vary depending on the set of features taken into account.

In our work, the accuracy of SVM was very close to that of the LDA in some samples. However, in most cases the LDA performed better than the SVM. The class prediction errors for three confusion matrices are visualized with histograms in Figure 3.

From Figure 3, it can be observed that most of the misclassification occurs within the same type of instruments that are mentioned in Table 3. *Dogor* is misclassified as Madal and Negara in all three models. Similarly, classification errors are observed among all the Drums. *Pepa*, which is a very unique wind instrument due to its bold vibrating sound, is correctly classified by all three models. Both SVM and LDA classified *Gogona* with zero false positive and false negative rates. Classification errors happen between the two wind instruments *Xutuli* and *Bahi* in all three models. Similarly, misclassifications are observed among *Ananda-lohori*, *Dotara*, and *Ektara* in both Decision Tree Classifier and SVM, while the LDA is performing quite well in identification of these three string instruments.

For a better evaluation of these three fitted models, ROC curves and F-measures are also determined for each instrument.

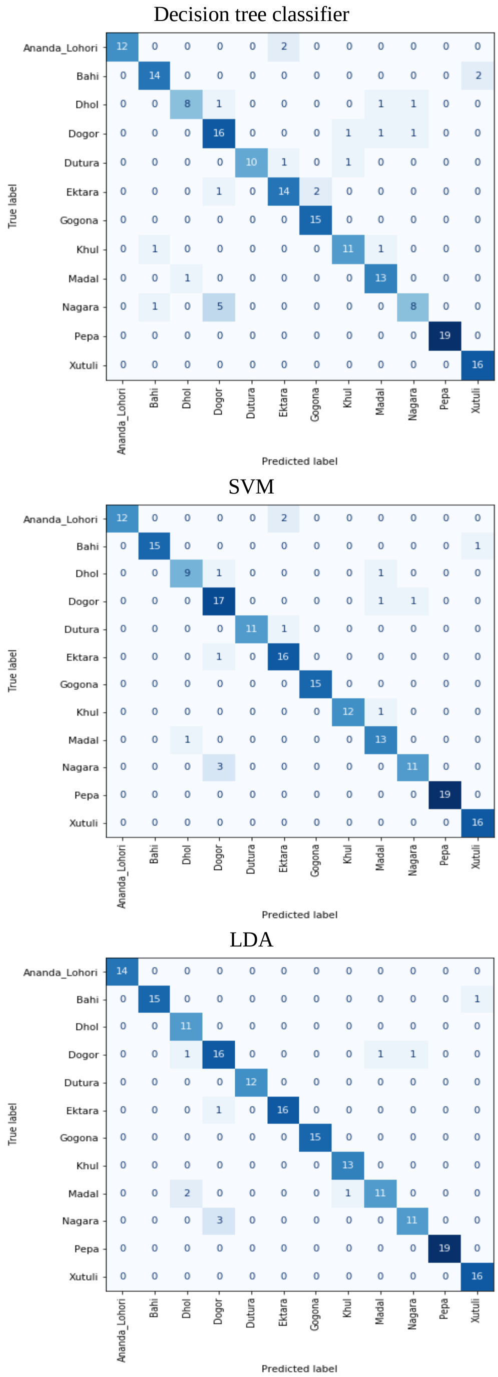

Figure 2. Confusion matrices for three types of trained models

5.2 ROC curves of the fitted models

The ROC curve was constructed for each instrument, and their macro and micro averaging of the three fitted models separately, shown in Figure 4.

Both LDA and SVM provided a better AUC score in comparison to Decision Tree Classifier. The average (both Micro and Macro) AUC scores for Decision Tree Classifier and SVM were 93% and 99% respectively while for LDA, the micro average AUC was 99% and the macro average was 98%.

Table 2. Accuracies of the three types of models

Model	Average accuracy score	95% Confidence interval of the accuracy score
Decision tree classifier	0. 869	0.864 – 0.874
Support vector machine	0.90	0.896 – 0.905
Linear discriminant analysis	0.922	0.919 – 0.926

Table 3. Selected instruments with their categories

String instrument	Wind instrument	Drum	Non-drum percussion instrument
1. *Dotara* or *Dutura* 2. *Ananda Lohori* 3. *Ektara*	4. *Bahi* 5. *Pepa* 6. *Xutuli*	7. *Dhol* 8. *Khol* or *Khul* 9. *Nagara* 10. *Dogor* 11. *Madol*	12. *Gogona*

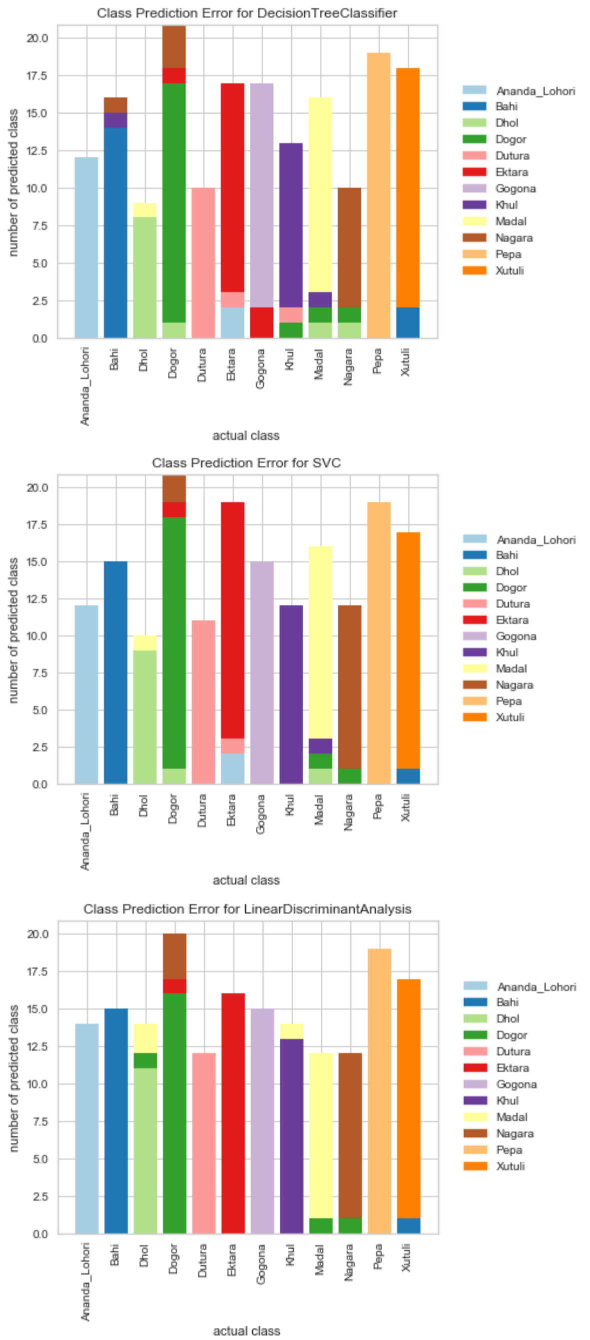

Figure 3. Class prediction error for each model type

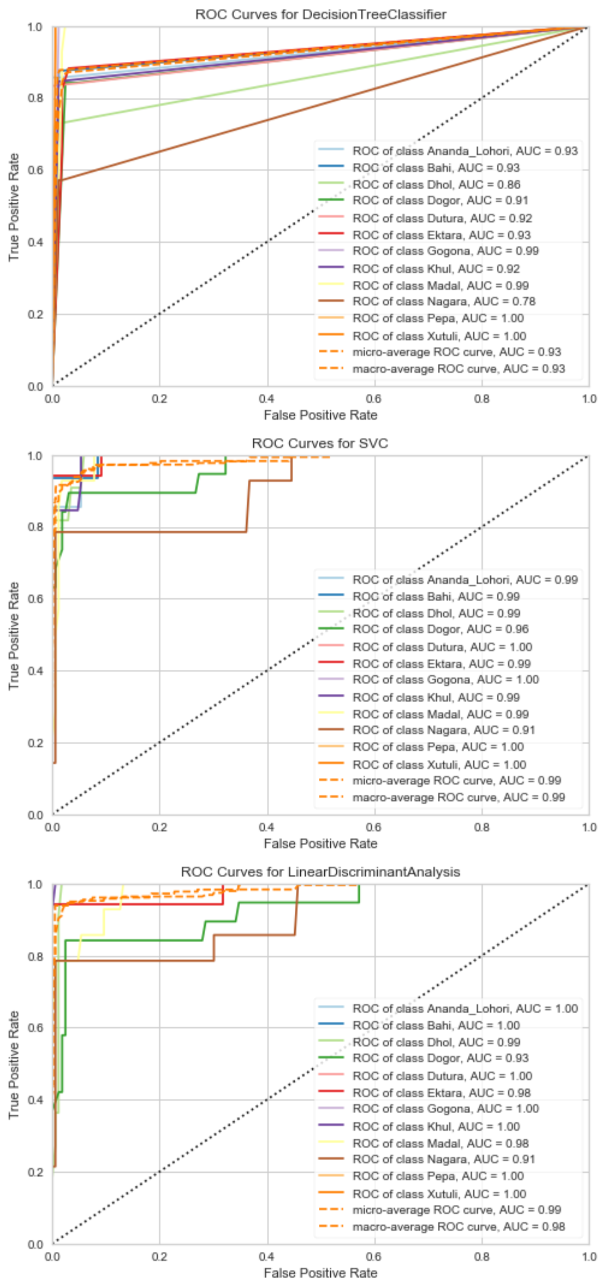

Figure 4. ROC curve for each identified model

5.3 F-1 Score

In our problem both types of error (Type-I and Type-II) were considered equally sensitive, and the F-1 score determined for each instrument is shown in Table 4.

Table 4 shows that both SVM and Linear Discriminant Analysis had better F-1 scores than the Decision Tree Classifier. For *Khol*, *Ektara*, *Dhol* and *Ananda Lohori*, LDA performed better than SVM. On the other hand, for Madal and *Dogor*, F-1 score of SVM was better than that of the LDA. Figure 5 shows the average precision-recall curves for each of the models. In this work, LDA provided the highest average precision in identifying the instruments.

Table 5 summarises the various scores used to assess the performances of the three models.

Table 4. Instrument-wise comparison of F-1 scores for each model

Instrument	F-1 Score: Decision tree classifier	Support vector machine	Linear discriminant analysis
Xutuli	0.941	0.970	0.970
Pepa	1.000	1.000	1.000
Nagara	0.667	0.846	0.846
Madol	0.867	0.867	0.846
Khol	0.846	0.960	0.963
Gogona	0.938	1.000	1.000
Ektara	0.824	0.889	0.970
Dotara	0.909	0.957	1.000
Dogor	0.762	0.829	0.821
Dhol	0.800	0.857	0.880
Bahi	0.875	0.968	0.968
Ananda Lohori	0.923	0.823	1.000

Table 5. Classification report of three models

Model	Average Accuracy	AUC (Micro average)	Precision (micro- average)
Decision tree classifier	0.869	0.93	0.81
Support vector machine	0.90	0.99	0.90
Linear discriminant analysis	0.922	0.99	0.94

6. Conclusions

One of the biggest challenges in conducting this study was the collection of the audio samples for each selected folk music instrument, which was very time-consuming. Also, some instruments are less popular than others. As a result, the musicians of some instruments, like *Ektara* and *Ananda lohori*, are not easily available everywhere. It is not possible to gather a very large number of samples for each instrument, because every sample is recorded with the assistance of several music professionals.

Results from all three selected classification techniques show that LDA and SVM performed significantly better than Decision Tree classifier when using the same set of features. If we compare the performances of LDA and SVM, the LDA performed slightly better than the SVM. The primary drawback of this work is that we are addressing only three classifiers, but there are several statistical methods, including Logistic Regression and Random Forest Classifier etc. that might have been successfully used to tackle the classification issues.

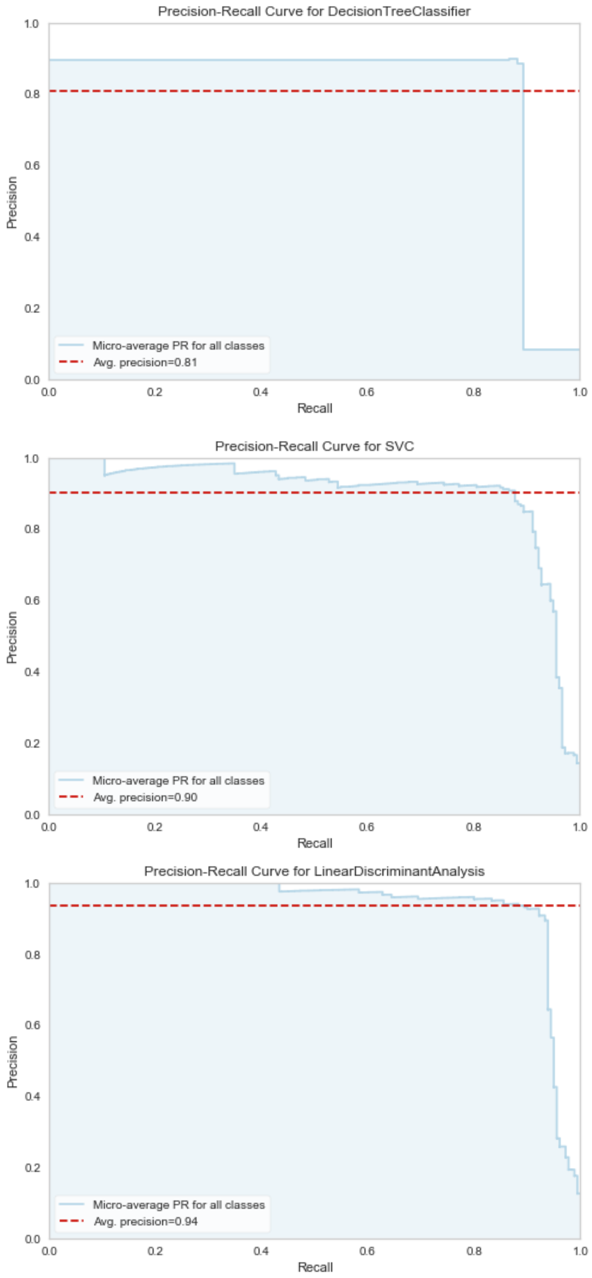

Figure 5. Average precision and recall curves for each model

To the best knowledge of the authors, no prior computational study has been carried out regarding sound-based identification of Assamese musical instruments. This study is a first step in the future research on Assamese music.

Based on this, the next study is expected to find the optimal set of features for identifying solo musical instruments. This work will certainly provide a basis for studying the features of sounds from traditional musical instruments for any community over the country.

Acknowledgements

We would like to convey our heartfelt gratitude to the musicians for their enthusiastic participation in the experiment's data collection process.

References

Abeßer, J., Lukashevich, H., & Bräuer, P. (2012). Classification of music genres based on repetitive basslines. *Journal of New Music Research*, *41*(3), 239-257. doi:10.1080/09298215.2011.641571

Agostini, G., Longari, M., & Poolastri, E. (2003). Musical instrument timbres classification with spectral features. *EURASIP Journal on Advances in Signal Processing, 678(*1*)*, 5-14.

Arowolo, M. O., Adebiyi, M. O., Nnodim, C. T., Abdulsalam, S. O., & Adebiyi, A. A. (2021). An adaptive genetic algorithm with recursive feature elimination approach for predicting malaria vector gene expression data classification using support vector machine kernels. *Walilak Journal of Science and Technology*, *18*(17). doi:10.48048/wjst.2021.9849

Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory* – COLT '92, 144-152.

Deng, D. J., Simmermacher, C., & Cranefield, S. (2008). A study on feature analysis for musical-instrument classification. *IEEE Transactions on System, Man and Cybernetics, Part B (Cybernetics)*, *38*(2), 429-438.

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A., & Vapnik, V. N. (1997). support vector regression machines. *Advances in Neural Information Processing Systems 9*, NIPS 1996, 155–161.

Duan, K., & Keerthi, S. S. (2005). Which is the best multiclass SVM method? An empirical study. In N.C. Oza, R. Polikar, J. Kittler, & F. Roli (Eds.), Multiple Classifier Systems. MCS 2005. *Lecture Notes in Computer Science*: *Volume 3541*. Berlin, Germany: Springer. doi:10.1007/11494683_28

Harikrishnan, N. B. (2019). Confusion matrix, accuracy, precision, recall, F1 score. Retrieved from https://medium.com/analytics-vidhya/confusion-matrix-accuracy-precision-recall-f1-score-ade299cf63cd

Jiang, D. N., Lu, L., Zhang, H. J., Tao, J. H., & Cai, L. H. (2002). Music type classification by spectral contrast feature. *Proceedings of the IEEE International Conference on Multimedia and Expo, 1*, 113-116. doi:10.1109/ICME.2002.1035731.

Johnson, R. A., & Wichern, D. W. (2015). *Applied multivariate statistical analysis* (6^{th} ed.). India: Person India Education Services.

Jondya, A. G., & Iswanto, B. H. (2017). Indonesian's traditional music clustering based on audio features. *Proceeding of 2^{nd} International Conference on Computer Science and Computational Intelligence, ICCSCI, 2017*, 174-181.

Kim, H. G., Moreau, N., & Sikora, T. (2004). Audio classification based on MPEG-7 spectral basis representations. *IEEE Transaction on Circuits and Systems for Video Technology*, *14*(5), 716–725.

Krishnaswami, S. (1971). Musical instruments of India. *Asian Music*, *2*(2), 31-42.

Marques, J., & Moreno, P. (1999). A study of musical instrument classification using Gaussian mixture models and support vector machines. Cambridge Research Laboratory, *Technical Report Series, CRL 99/4*

Patel, H. H., & Prajapati, P. (2018). Study and analysis of decision tree based classification algorithms. *International Journal of Computer Sciences and Engineering*, *6*(10), 74-78.

Sen, A. (2014). Automatic music clustering using audio attributes. *International Journal of Computer Science*, *3*(6), 307-312.

Setiadi, D. R. I. M., Rahardwika, D. S., Rachmawanto, E. H., Sari, C., A., Irawan, C., Kusumaningrum, D. P., Nuri, & Trusthi, S. L. (2020). Comparison of SVM, KNN, and NB classifier for genre music classification based on metadata. *Proceeding of International Seminar on Application for Technology of Information and Communication (iSemantic)*, 12-16. doi:10.1109/iSemantic50169.2020.9234199.

Tjoa, S. (2017). Music information retrieval. Retrieved from https://musicinformationretrieval.com/spectral_features.html.

Tzanetakis, G., & Cook, P. (2002). Musical genre classification of audio signals. *IEEE Transactions on Speech and Audio Processing*, *10*(5), 293–302.

Ünal, E., Bozkurt, B., & Karaosmanoğlu, M. K. (2014). A hierarchical approach to Makam classification of Turkish Makam music using symbolic data. *Journal of New Music Research*, *43*(1), 132-146, doi:10.1080/09298215.2013.870211

Vapnik, V. N. (1995). *The nature of statistical learning theory*. New York, NY: Springer.

Weihs, C., Jannach, D., Vatilkin, I., & Rudolph, G. (2017). *Music data analysis-foundations and applications*. London, England: Taylor and Francis.

Wu, X., Kumar, V., Quinlan, R. J., Ghosh, J., Yang, Q., Motoda, H., . . . Zhou, Z. H. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, *14*(1), 1-37. doi:10.1007/s10115-007-0114-2

Xiong, Z., Radhakrishnan, R., Divakaran, A., & Huang, T. (2003). Comparing MFCC and MPEG-7 audio features for feature extraction, maximum likelihood HMM and entropic prior HMM for Sports Audio Classification. *Proceeding of the International Conference on Multimedia and Expo. ICME'03*. (Cat.No. 03TH8698), July 6-9, Baltimore, MD, USA, 7–400.