Original Article

A Computational Approach for Identification of Assamese Folk Musical Instruments

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

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

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

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

Abstract

Classification of musical instruments using the computational technique is a very challenging task. The development of signal-processing and data-mining techniques has 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 musical instruments are selected for identification. Twelve musicians play the instruments and audio samples are recorded, different instantaneous features are extracted and 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 is made among the three classifiers. The proposed sets of features enable the models DTC, SVM and LDA to achieve average accuracy ratings of 86.9%, 90% and 92.2% respectively. Regarding the performance of the three fitted models in identifying instrumental sound, this study will offer a valid comparison.

Keywords: Musical Instrument, Machine learning, Decision Tree Classifier, Support Vector Machine, Linear Discriminant Analysis

1. Introduction:

Indian culture is famous due to its diversified cultures and traditions. Each part of the country has its own unique culture and tradition, and each culture is conspicuously visible in its different art forms. Assam situated in the northeastern part of India has rich cultural resources, including different kinds of traditional music as the people of the 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 the performance of different kinds of folks prevailing in Assam. Krishnaswami (1971) classified the Indian musical instruments under the heads, namely 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 etc.). All four kinds of musical instruments are used in the performance of Assamese folk. The raw materials used to make these instruments are bamboo, leather, soil, buffalo horn, string, wood, bottle gourd, etc.

This research aims to develop three models for identifying Assamese folk musical instruments using Decision Tree Classifier, Support Vector Machine (SVM), and Linear Discriminant Analysis (LDA), and to analyze the performance. 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 instruments played along with *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 the buffalo horn attached to it, makes the sound very unique. *Xutuli* is a wind instrument made from clay or the lower end of a bamboo tree. *Gogona* is 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 by two bamboo sticks. It is the major instrument in the performance of spiritual songs called *Negara Naam*. *Khol* (or *Khol*) is another popular two-faced drum played with free hands. This instrument is played with the religious, spiritual songs *Mohapurusia Sangeet***.**

2. Related Work:

Music data analysis and retrieval have become a very popular research field in recent years. Previously the clustering and classification of music were performed using manually specified features of samples. The rapid progress of signal-processing and data-mining techniques has made it possible to study the different features of musical signals, which plays an important role in solving the classification and identification problem in music. K-Mean clustering is one of the widely used techniques to solve clustering and classification problems in music. For the classification of Indonesian traditional music, Jondya and Iswanto (2017) select the essential musical features using principal component analysis and find 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) study the features of musical instruments and classify them using K- nearest neighbour classification algorithm. In this work, PCA and Isomap were used to explore the sparseness of the feature space 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 is done by Agostini, Longari and Poolastri (2003) using 117 spectral features. The performance was assessed using SVM, k-NN, canonical Discriminant analysis, and quadratic discriminant analysis. SVM and quadratic discriminant analysis performed the best. Tzanetakis and Cook (2002) used the Gaussian mixture model (GMM) and K-mean algorithm for audio signal musical genre classification. Hidden Markov Model (HMM) is found to be one of the successful statistical techniques for solving classification and identification problems. HMM-based classifier is 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) use HMM, K-NN, GMM, AdaBoost and SVM techniques for sports audio classification.

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

Ünal, Bozkurt and Karaosmanoğlu (2014) use 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 is used. Then a more detailed identification is 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 are collected from primary sources. Twelve expert musicians are contacted and briefed about the study's goals. All of them consented to play the instruments. Different instruments of the same type may produce different sounds due to their size, shape, tuning and build quality. Therefore we collect samples from more than one instrument of the same type. Thus for the collection 35 instruments are used. For each instrument type, musicians are asked to play 50 different melodies or beats and sounds were recorded for 20 sec. window in .WAV format in a sampling rate of 44100 Hz under the same acoustic environment and same condition. One of the serious obstacles in the data collection process was that due to the limited use of some instruments like *Gogona*, *Xutuli* in the performance, comparatively small number of samples were obtained. The number of collected samples for each type of instrument is shown in the table-2.

4. Methodology:

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

Generation of the Spectrogram:

A spectrogram is a visual representation of a signal's 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, 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) spectrogram is extracted from each of the samples. Spectrograms extracted from one audio signal of each type of instruments are shown in Figure 1.

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 are given below.

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 according to the following equation:

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

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 frame and finally the average and the standard deviation is considered for analysis.

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 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 as Chroma Vector or Chroma features. Chroma vector coefficients are determined by grouping of short-term window into 12 bins. Every bin is calculated according to the formula:

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

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

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

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

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

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

Where *S(k)* and *f(k)* are respectively spectral magnitude and 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 of the energy.

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

Models used for Identification:

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

1. Decision Tree Classifier

A decision tree is one of the popular predictive modelling approaches 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 represented by 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, support vector machines (SVM) are one of the robust and accurate classification algorithms (Vapnik, 1995). This algorithm was developed at 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 classes (Arowolo et al., 2021). Originally this algorithm was developed for binary classification problems. For multiclass classification, the multiclass 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 application as well as for dimensionality reduction problems as a pre-processing step for 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 that characterizes two or more classes of objects. The resulting linear combination may be used as a linear classifier.

Evaluation of the Fitted Models:

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

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

Accuracy score = $\frac{TP+TN}{TP+TN+FP+FN}$; TP= True Positive, FP= False Positive

TN= True Negative, FN= False Negative

ROC Curve:

An ROC curve (receiver operating characteristic curve) is a two dimensional graph showing the performance of a classification model at all classification thresholds.

In this plot True Positive Rate (TPR) is plotted on the Y axis and False Positive Rate (FPR) is plotted on the X axis where-

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

It is a useful technique for visualizing, organizing and selecting classification model based on their performance. 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 value is calculated for each instrument where,

$$\textbf{Precision} = \frac{TP}{TP+FP} \qquad \textbf{Recall} = \frac{TP}{TP+FN}$$

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

5. Results and Discussion:

In this work 70 % of the total sample have been selected randomly and used for training the models and the remaining 30% of the sample for testing purposes. This process is repeated 100 times so that confidence interval for the estimates can be constructed. Confusion matrix is considered as one of the valid method for inspecting the performance of the fitted models from a qualitative point of view. For a specific randomly chosen test sample, the models' prediction is visualized in three confusion matrices, which are shown in figure-2.

Evaluation the Model:

The accuracy scores for each fitted models are determined for 100 randomly selected test samples and 95% confidence intervals for the scores are calculated. The average accuracy score along with 95% confidence interval is presented in table-3.

The performance of Linear Discriminant Analysis and SVM is quite good than the Decision Tree Classifier in terms of their accuracy scores. In the work by Marques and Moreno (1999), eight musical instruments are classified using SVM and Gaussian Mixture Model, where SVM gives the best results with an overall error rate of 30% when classifying segments of 0.2 seconds of sound. In the work by Agostini *et al.* (2003), SVM with RBF kernel gives the best result for recognition of individual instruments in comparison to the other classifiers- 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 SVM's performances. On the other hand, in case of instrument family recognition and sustain/pizzicato classification, QDA overcame all other classifiers with a success rate of 81% (Agostini *et al.*, 2003). In the experiment of Setiadi-Trusthi et al. (2020), three classifiers SVM, KNN, and Naïve Bayes (NB) are used for the classification of music genres on the Spotify music dataset. They find that the SVM classifier has the best classification performance with 80% of accuracy, followed by KNN and NB. The model's accuracy, however, may vary depending on the issue being investigated. Since every problem has a unique set of features, the amount of information will vary depending on the set of features being taken into account.

In our work, the accuracy score of SVM was very close to LDA in some samples. However in most of the cases, LDA performs better than SVM. The class prediction error for three confusion matrices are visualized with help of 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-1. *Dogor* is misclassified as Madal and Negara in all three models. Similarly classification error is 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 rate. A classification error is happening between the two wind instruments *Xutuli* and *Bahi* in all three models. Similarly, misclassification is observed among *Ananda-lohori*, *Dotara*, and *Ektara* in both Decision Tree Classifier and SVM while LDA is performing quite good in identification of these three String Instruments.

For better evaluation of these three fitted models, ROC curve and F-measure are also determined for each instruments.

ROC curve of the fitted models:

The ROC curve is constructed for each instrument and their macro and micro averaging of three fitted models separately, shown in figure-4.

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

F-1 Score:

In our problem both the errors (Type-I and Type-II) are equally sensitive, the F-1 score is determined for each of the instruments which are shown in table 4.

Table: 4 shows that both SVM and Linear Discriminant Analysis have better F-1 score than Decision Tree Classifier. For *Khol*, *Ektara*, *Dhol* and *Ananda Lohori*, LDA is performing better than SVM. On the other hand, for Madal and *Dogor*, F-1 score of SVM is better than LDA. Figure 5 shows the average precision-recall curves for each of the models. In this work, LDA provides the maximum average precision for the identification of the instruments.

Table 5 summarises the various scores that are used to assess the performance of the three models.

6. Conclusion:

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

Results from all three selected classification techniques show that LDA and SVM performs significantly better than Decision Tree classifier under the same set of features. If we compare the performance of LDA and SVM, LDA performs slightly better than 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 may have been successfully used to tackle classification issues in a wide range of domains.

As far the knowledge of the reasercher is concerned, no computational study has been carried out in Assamese musical instruments. This study is expected to be the first step of the future research of Assamese music. Based on this, the next research is expected to find the optimal set of features for identifying solo musical instruments. This work will certainly provide a basis for studying features of 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: https://doi.org/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, MIT Press.

Duan, K., & Keerthi, S. S. (2005). Which Is the Best Multiclass SVM Method? An Empirical Study. In: Oza N.C., Polikar R., Kittler J., Roli F. (eds) Multiple Classifier Systems. MCS 2005. *Lecture Notes in Computer Science*: *Vol. 3541*. Springer, Berlin, Heidelberg. doi: https://doi.org/10.1007/11494683_28

Harikrishnan, N. B. (2019, December 10). 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. IEEE International Conference on Multimedia and Expo*: 113-116 vol.1, doi: 10.1109/ICME.2002.1035731.

Johnson, R. A., & Wichern, D. W.(2015). *Applied Multivariate Statistical Analysis*. (Sixth Edition). India: Person India Education Services Pvt. Ltd.

Jondya, A. G., & Iswanto, B. H. (2017). Indonesian's Traditional Music Clustering Based on Audio Features. *2nd International Conference on Computer Science and*

Computational Intelligence, ICCSCI; October 13-14, 2017, Bali, Indonesia, 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. *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*. Springer, New York

Weihs, C., Jannach, D., Vatilkin, I., & Rudolph, G.(2017). *Music Data Analysis-Foundations and Applications*. London, New York : Taylor & Francis Group.

Wu, X., Kumar, V., Quinlan, R.J., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. j., Ng, A., Liu, B., Yu, P.S., & Zhou, Z. H. (2008). Top 10 algorithms in data mining. *Knowledge and Information Systems*, 14(1), 1-37. doi: https://doi.org/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.

Table

Table 1: Selected Instruments with their original classes

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*

Table 2: Instrument with their sample sizes.

	Instrument	**Number of Instrument**	**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

Table 3: Accuracy Score of the three models

Model	**Average Accuracy Score**	**95% Confidence Interval of the Accuracy Scores**
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 4: Instrument wise comparison of F-1 Score for each model

Instruments	**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

Models	**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

figure

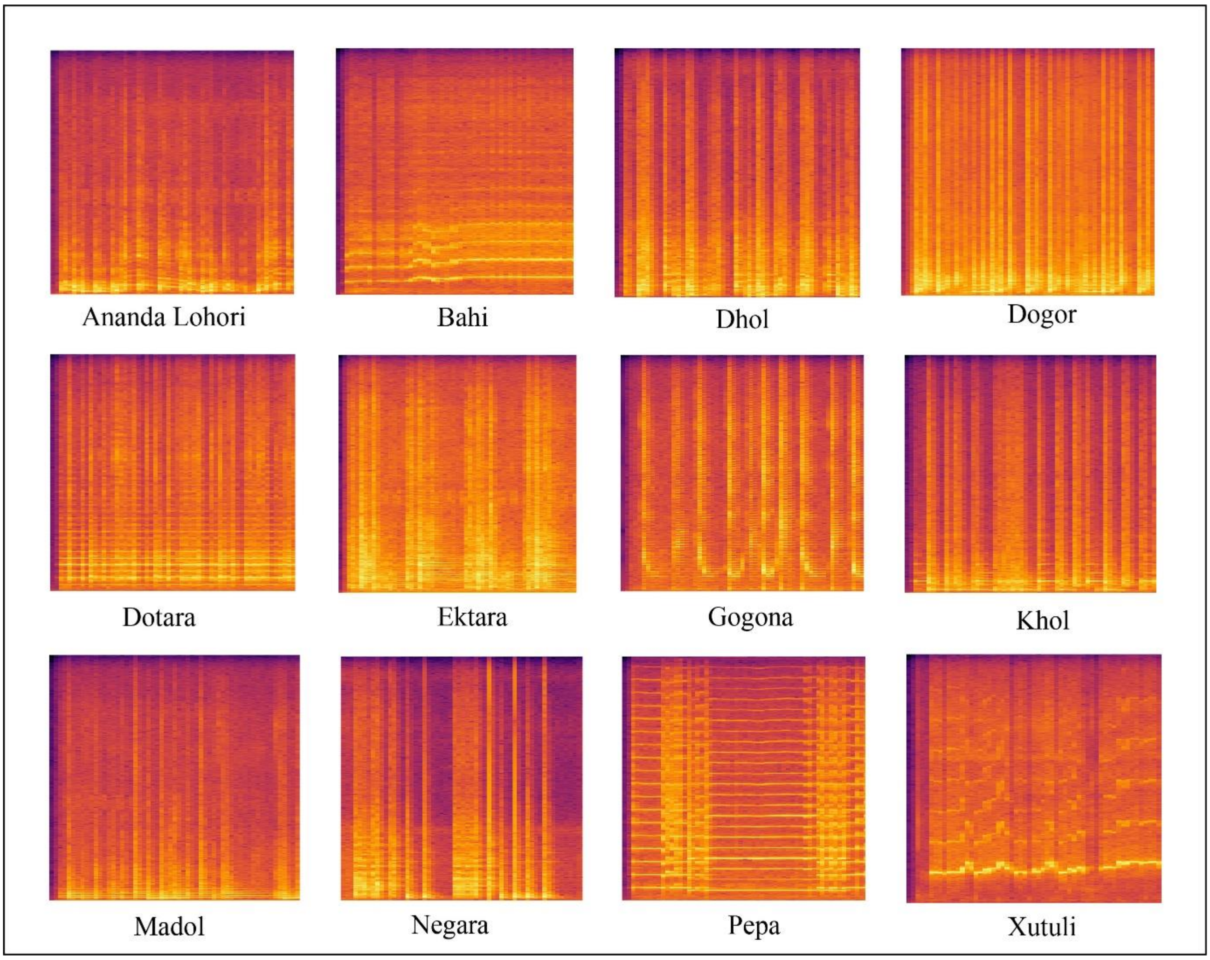
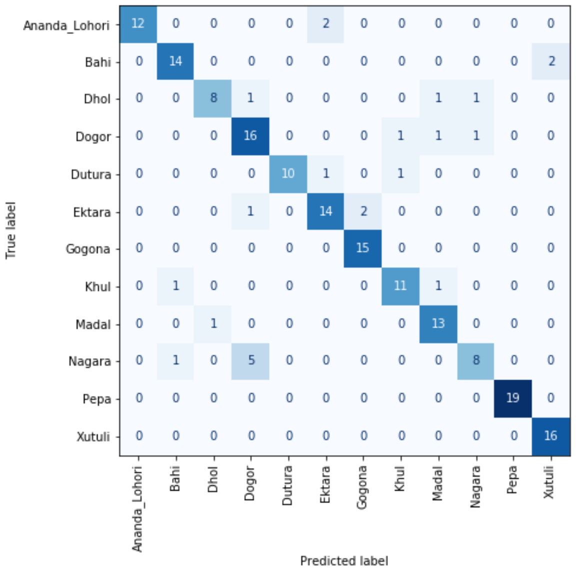
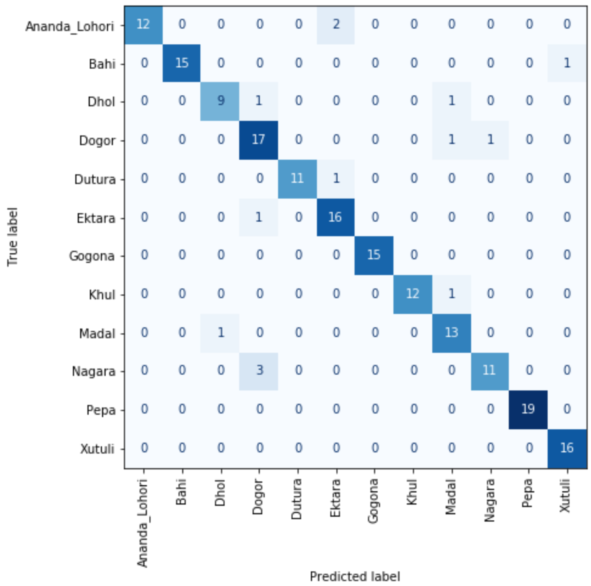


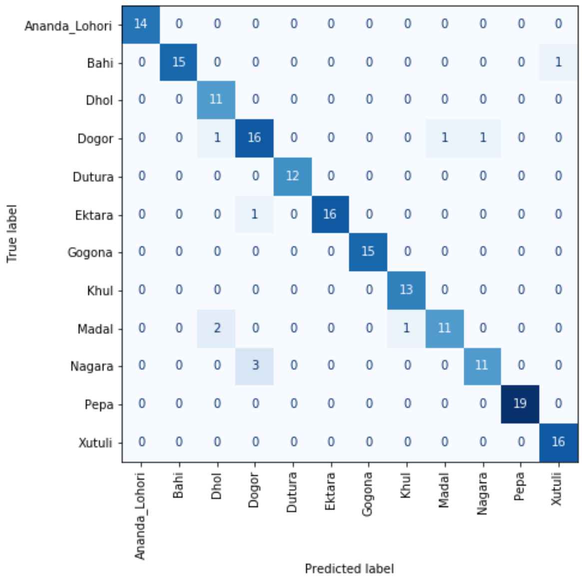
Figure 1: Sample Spectrograms for one audio signal of each type of instrument



Decision Tree Classifier



SVM



LDA

Fig-2. Confusion Matrices for three models.

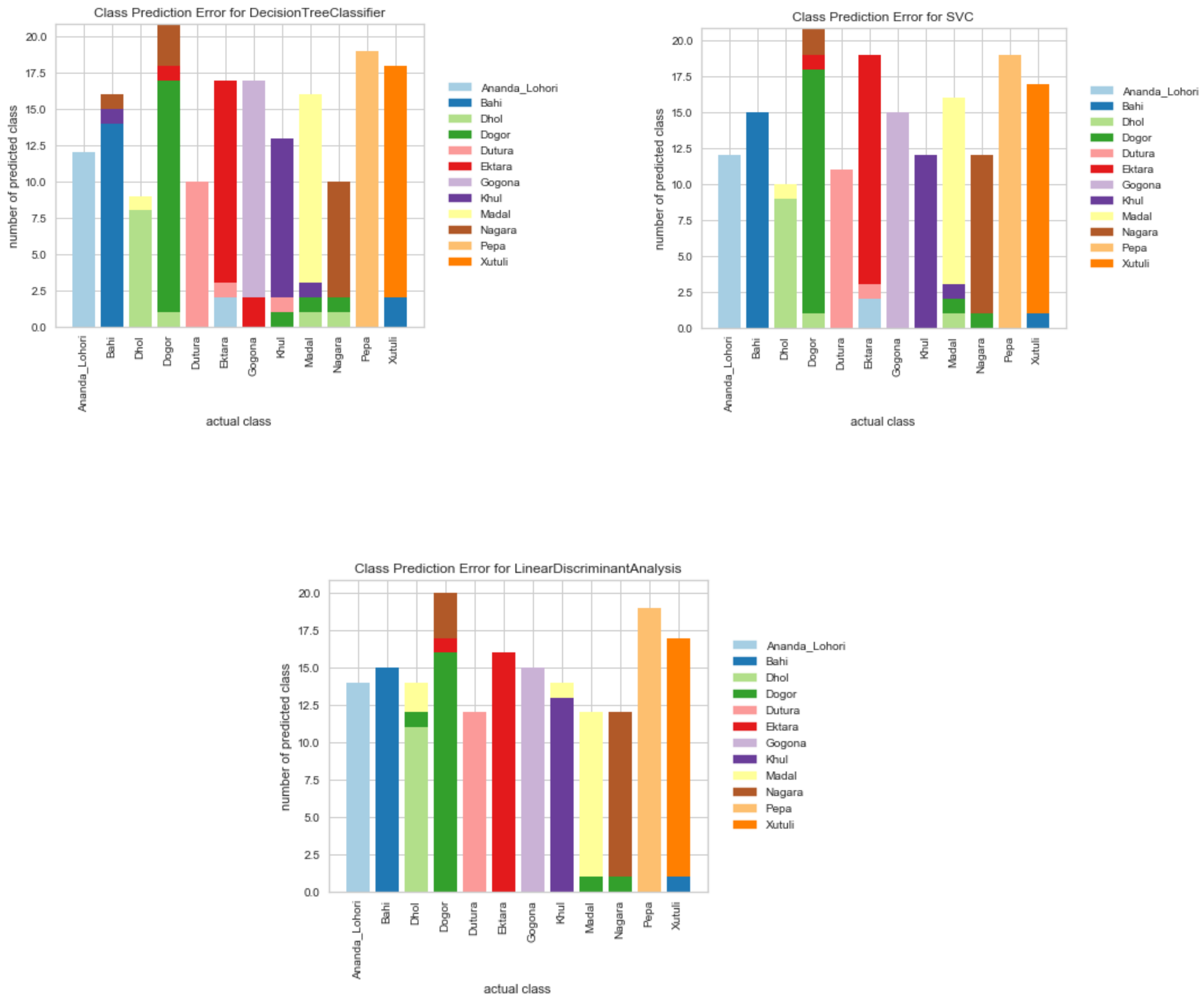


Figure 3: Class prediction Error for each Model

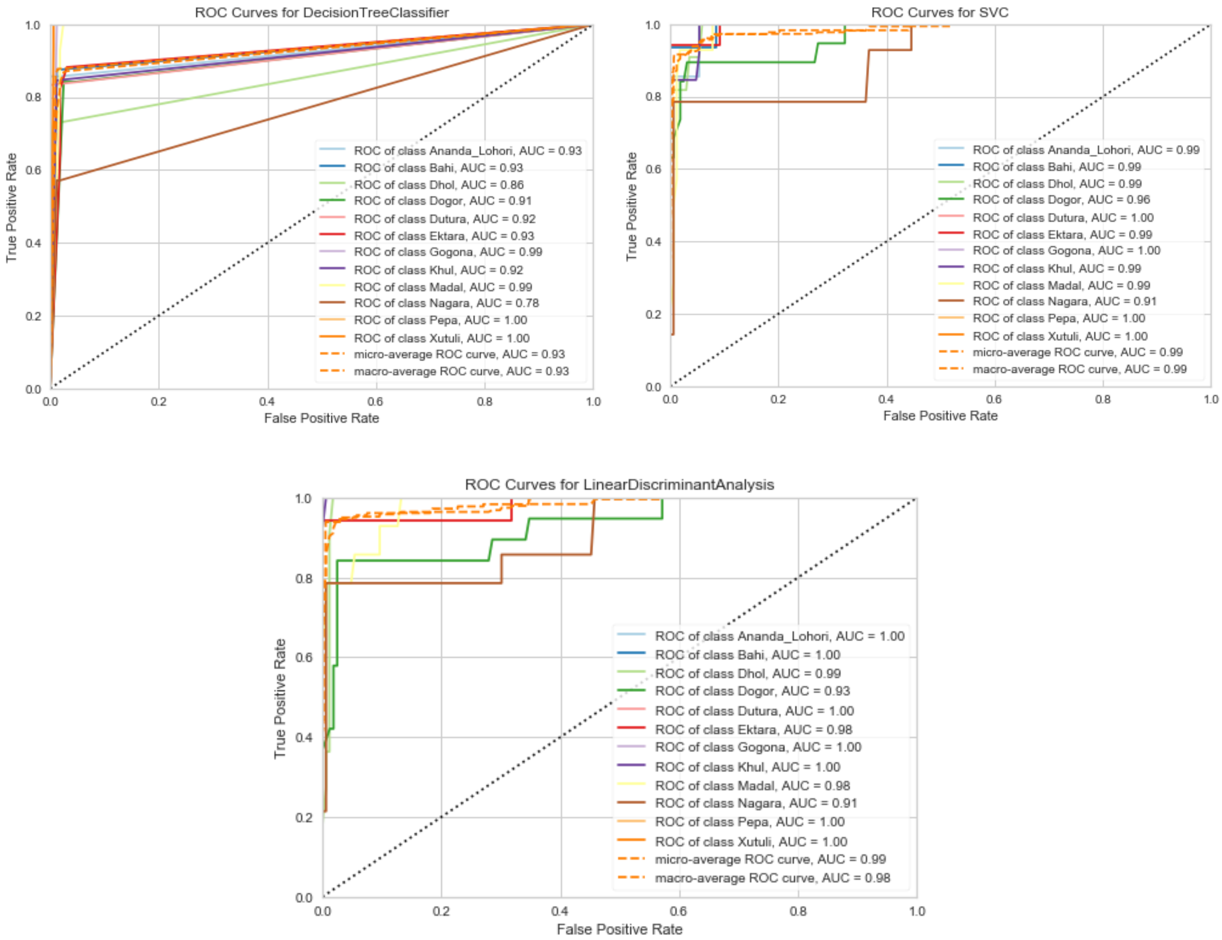


Figure 4: ROC curves for each Models

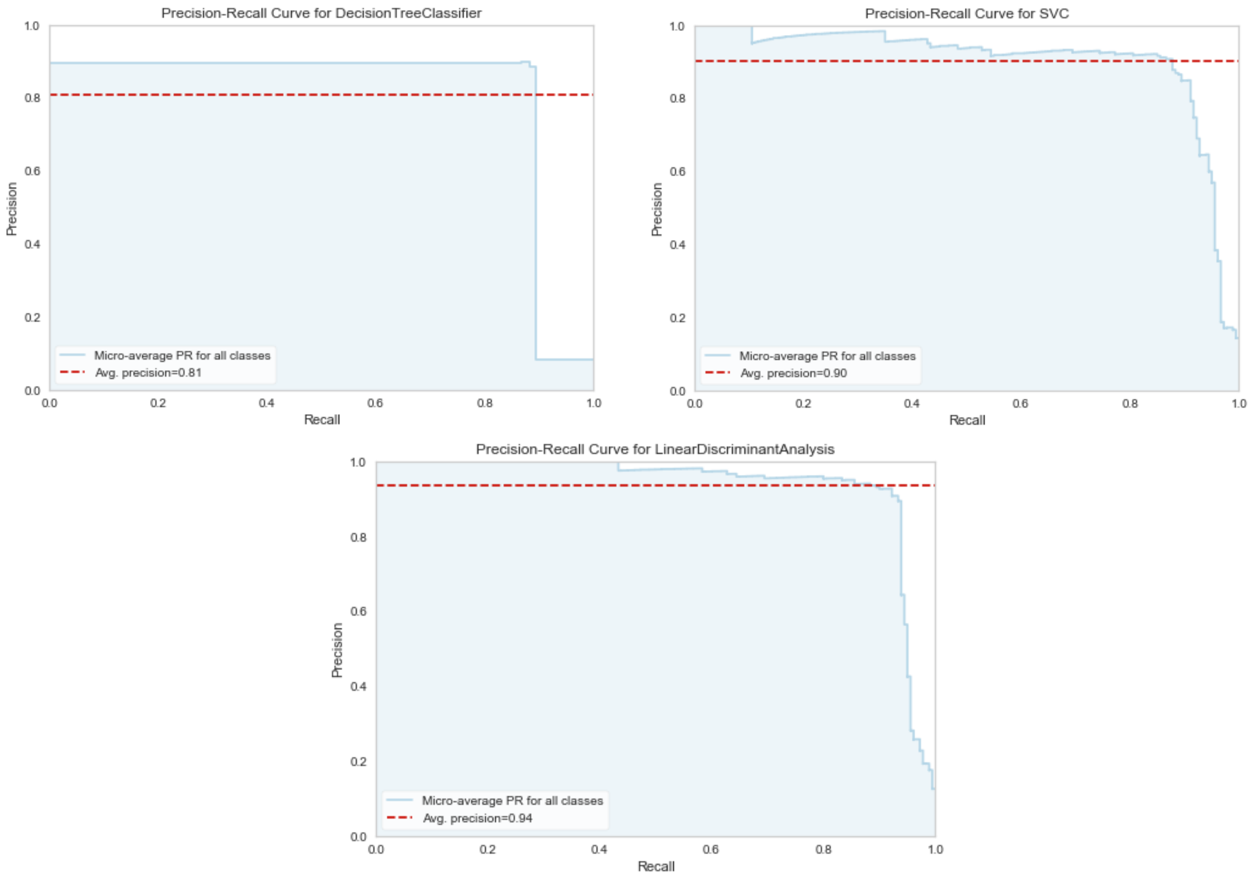


Figure 5: Average Precision Recall Curves for each model

List of Changes

Serial No.	Reviewer's Suggestions	Author's Changes
1.	**Reviewer #1**: Please find more references instead of repeated use of "Jondya & Siswanto(2007)".	Instead of repeating the reference of "Jondya & Siswanto(2007)" a new book (Weihs et al., 2017) is added in the references.
2.	**Reviewer #1**: Please add the clustering result and interpretation to the abstract and conclusion.	As the number of words in the abstract is limited, the author included only the best result among the fitted models. Now the prediction scores of all the models are also included as suggested by the reviewer.
3.	**Reviewer #2:** The model efficiency is very high, please give the reasons relating to the other papers.	A quick comparison is made between the results of the fitted models and the other relevant work as suggested by the reviewers. The model accuracy may vary depending upon the problem under study. Because each problem has some specific set of features and accordingly the amount of information will vary depending upon the set of features under consideration.
4.	**Reviewer #2:** In abstract section, Please add.. 'What benefit is provided by research?"	The abstract is modified as suggested by both reviewers. The benefit of this piece of work is included in the abstract.
5.	**Reviewer #2:** Explain or more discussion of the limitations of using three popular classification techniques.	The primary drawback of using only the three models is included in the conclusion section as suggested by the reviewer.
6.		The "Acknowledgements" section is added to the manuscript. Additionally, the authors have rectified a few minor grammatical errors. To maintain the word limit of the manuscript some lines are reconstructed without changing the meaning. The description of the three models is shortened and three references- Abbas, (2021), Fawcett, T. (2006), Sharma et al. (2013) are removed.