



# Machine Learning-Based Classification of Khasi Musical Instruments for Cultural Preservation

<sup>1</sup>Mebanpynshai Lyngdoh Marshilong, <sup>2</sup>Sushanta Kabir Dutta,

<sup>1</sup>IoT Engineer, <sup>2</sup>Associate Professor

<sup>1</sup>Electronic And Communication Engineering,

<sup>1</sup>North Eastern Hill University, Shillong, India

## Abstract

Cultural folk music often faces negligence in front of Technological advancements resulting in the erosion of traditional musical heritage. This study aims to preserve the characteristics of Khasi folk musical instruments through music information retrieval (MIR) techniques to integrate Khasi folk music into the modern digital Technology era. We focus on extracting audio features like “Attack-Delay-Sustain-Release (ADSR)”, “Zero-Crossing-Rate (ZCR)”, “Spectral Centroid”, “Spectral Rolloff” and “Spectral flux” from the sound produced by the Khasi folk musical instruments. The extracted features will be used to train the machine learning model and preserved for future use and reference. Our model used in this study consists of a bunch of Classifiers and we focus our analysis on 7 different types of Khasi folk musical instruments. We obtain a precision of 91%, a recall of 90%, and an F1-score of 90%.

**Index term:** Khasi, Khasi folk musical instruments, machine learning, Music Information Retrieval(MIR)

## 1. Introduction

The Khasi people are an ethnic group of Meghalaya in North-Eastern India and their folk music is known as Khasi folk music. The evolution of Khasi folk music is as ancient and profound as the Khasi tribe itself. The music which evolved from Hima Shyllong which is one of the oldest native states is old and tenable. authentic identity over the centuries. This rich musical tradition is integral to the

Khasi cultural heritage evolved through various stages from folk to a more structured traditional form.

Khasi folk music lays the foundational structure upon which traditional Khasi music has grown. It provides a glimpse into the basic rhythmic patterns that have progressively evolved, becoming more structured and classical. These patterns are characterized by a high degree of organization and technicality, including unwritten drum syllables that have standardized the playing techniques of Khasi drums. The spontaneous and creative nature of Khasi folk music reflects the way of life, culture, and traditions of the Khasi community.

Traditionally, Khasi music is transmitted orally, learned by ear, and passed down from generation to generation. There is no formal technique or conscious awareness of form and construction among folk singers.

The unique characteristics of Khasi folk music, with its rich rhythms and spontaneous nature, are deeply intertwined with the instruments used by the Khasi people. These instruments are not merely tools for creating music but are central to the Khasi community's cultural expression and traditional practices. They embody the musical heritage crucial in ceremonies, storytelling, and daily life. (1)

This paper proposes using Music Information Retrieval (MIR) techniques to extract information from the sound produced by Khasi folk musical instruments. The extracted data, referred to as features, will be stored for future reference and study, thereby aiding in preserving these musical traditions. These features will also be used to train a machine-learning model for classifying different Khasi folk musical instruments. (7)

The classification will be based on the technique used to produce sound and the materials used to construct the instruments. (2) Commonly there are many Khasi folk musical instruments; some of them are used for classification in this paper. 7 different musical instruments from four different families were used for the experiments, such as Duitara from String instruments, Besli from wind instruments, Singphong from Reed instruments, Ksing Shynrang, Ksing Kynthei, Bom, and Pdiah from Membrane instruments(1). A total of 252 music samples were collected.

## 2. Literature Review

An audio signal consists of a large number of features. However, it is required to extract the features that are relevant to the problem we are trying to solve- [3]. The features used in this paper represent both the temporal features (Time domain features) and the Timbral features(Spectral features), such as Attack-Delay-Sustain-Release (ADSR), Zero-Crossing-Rate (ZCR), spectral flux, spectral roll-off, and spectral centroid. According to [4] a Supervised machine learning model is a model where the

algorithm generates a function that maps inputs to desired outputs. One standard formulation of supervised learning tasks is the classification problems. In this paper we tested the following Supervised ML models: Decision Trees Classifier, Random Forest Classifier, Multi-Layer Perceptron Classifier, and K-Nearest Neighbors Classifier.

### 2.1 Extracted Features

#### A. Attack Delay Sustain Release (ADSR)

ADSR is the acronym for Attack(A), Decay(D), Sustain(S) and Release(R). These parameters shape the sound by controlling the amplitude(loudness) over time. The time instant when the note starts is termed the onset time. The amplitude reaches its peak after a certain amount of time. The time in between is termed as Attack Time. The amplitude starts to drop after the peak and the time instant when the envelope amplitude drops to 25% of its maximum value is termed as forward position. The rate of decrease is known as the Roll-off rate which defines the slope of the decay portion of the envelope. The amplitude remains constant along with time in Sustain time after which the amplitude reduces to almost zero.

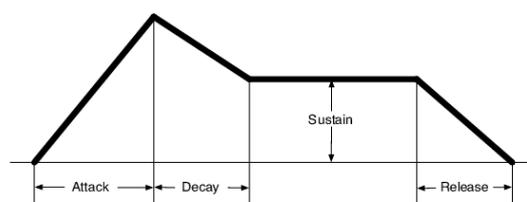


Fig 1.1: ADSR Envelope

#### B. Zero-crossing rate

Zero-crossing rate is the rate of change from positive to negative and vice-versa along the signal. It gives the smoothness of the signal wave. ZCR is calculated using the equation.

$$ZCR = \frac{1}{2W} \sum_{n=1}^W | \text{sgn}[x(n)] - \text{sgn}[x(n-1)] |$$

where  $\text{sgn}(\cdot)$  is the sign function i.e

$$\text{sgn}[x(n)] = \begin{cases} 1, & x(n) \geq 0 \\ -1, & x(n) < 0 \end{cases}$$

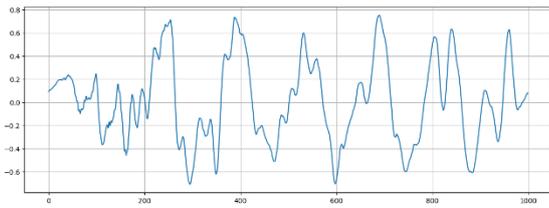


Fig 1.2: ZCR of Ksing Shynrang

### C. Spectral Centroid

Spectral Centroid is defined as the ‘center of gravity’ of the magnitude spectrum. It gives the sound signal a bright sensation. It can be calculated as the weighted mean of the spectral frequencies. We calculate the FFT of the signal along with the average energy weighted by the sum of spectrum amplitudes within one frame. The spectral centroid measures the spectral energy distribution in steady-state portions of tone.

$$\text{Spectral Centroid} = \frac{\sum_{n=0}^{N-1} f(n)x(n)}{\sum_{n=0}^{N-1} x(n)}$$

where,

$f(n)$  = magnitude of the FFT for a frame  $n$ .

$x(n)$  = index of the frequency bin

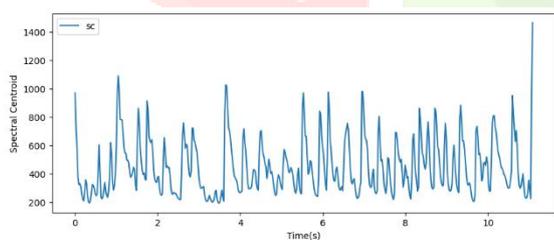


Fig 1.3: Spectral Centroid of Ksing Shynrang.

### D. Spectral Roll-off

Spectral Roll-off is defined as the frequency bin  $M$  below which 85% of the magnitude distribution is concentrated. Frequency bins are the intervals between samples in the frequency domain. The equation is given as,

$$\sum_{n=0}^M = 0.85 \times \sum_{n=0}^N f(n)$$

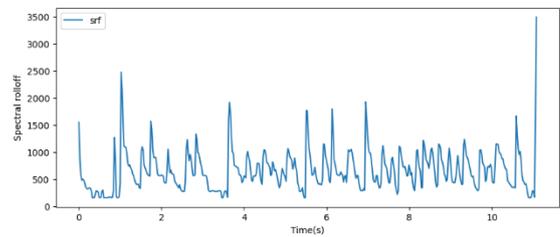


Fig 1.4: Spectral Roll-off of Ksing Shynrang

### E. Spectral Flux

Spectral Flux is defined as the rate of change in the power spectrum. It is the measure of how quickly the power spectrum changes from frame to frame. It is calculated by comparing the power spectrum of successive frames.

$$\text{Spectral flux} = \sum_{k=2}^K |M(f_x) - M(f_{x-1})|^2$$

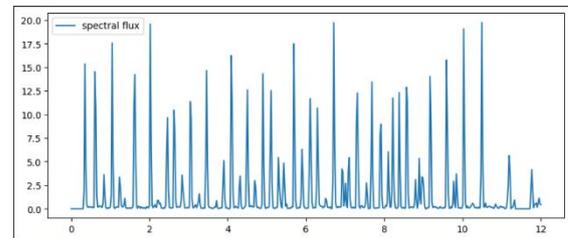


Fig 1.5: Spectral flux of Ksing Shynrang.

## 2.2 Classifiers

### A. Random Forest Classifier:

It is a classifier that contains several decision trees on various subsets of the given datasets and takes an average to improve the predictive accuracy of those datasets. The greater the number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

### B. K-Nearest Neighbors(KNN):

It is one of the simplest Machine Learning algorithms. Its algorithm assumes the similarity between the new data and available data and puts the new case into the category that is most similar to the available categories. This algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well-suited category by using the K-NN algorithm.

### C. Decision Tree Classifier:

It is a Supervised learning technique that uses a tree structure classifier where internal nodes represent the feature of the dataset, branches represent the decision rule and each leaf node represents the outcome. It is a graphical representation of getting all the possible solutions to a problem based on given conditions.

### D. Multi-layer Perceptron

MLP is a supervised learning algorithm that learns a function  $f: R^m \rightarrow R^o$  by training on a dataset, where  $m$  is the number of dimensions for input and  $o$  is the number of dimensions for output. Given a set of features  $X = x_1, x_2, \dots, x_m$  and a target  $y$ , it can learn a nonlinear function approximator for either classification or regression.

## 3. Data Collection and Processing

The first crucial step of this proposed paper involves collecting datasets for seven different Khasi folk musical instruments. Data collection was conducted in various locations, and all the recordings for this proposed paper were self-collected to ensure authenticity, with no external data sources used.

Post-collection, the recorded audio data required cleaning to ensure accuracy. Using audacity, ambient noise was filtered out and the recordings were normalized to achieve uniform amplitude levels. Each recording was segmented into four-second clips, preparing them for feature extraction. The final dataset comprises 252 recordings (36 for each of the seven instruments). These were divided into training and testing datasets, with 80% allocated for training and 20% of the total datasets for testing.

The cleaned datasets were preserved for further study and analysis of Khasi folk musical instruments. These datasets can later be used to compare Khasi instruments with other globally used musical instruments and to study the unique characteristics and sounds of the Khasi instruments themselves.

## 4. Experimental Setup and Result

For the collection of datasets, we used a recording setup that includes a UR 22C audio interface with a sampling frequency of 192KHz, a Shure microphone, and a laptop. Music Information Retrieval (MIR), the science responsible for extracting musical information, was used for feature extraction from the audio datasets. For this task, we utilized "Librosa," a Python module specialized in MIR.

A total of 252 datasets were collected, with 49 used for testing and the remaining for training. We selected several pre-trained machine learning models from Sklearn, including the Random Forest Classifier, Decision Tree Classifier, MLP Classifier, and KNN Classifier, for classification purposes.

```
print(f"Accuracy of Random Forest Classifier:\t\t {accuracy1}")
print(f"Accuracy of Decision Tree Classifier:\t\t {accuracy2}")
print(f"Accuracy of MLP Classifier:\t\t\t {accuracy3}")
print(f"Accuracy of KNeighbors Classifier:\t\t {accuracy4}")
```

Accuracy of Random Forest Classifier:	0.8775510204081632
Accuracy of Decision Tree Classifier:	0.7755102040816326
Accuracy of MLP Classifier:	0.7959183673469388
Accuracy of KNeighbors Classifier:	0.8571428571428571

**Fig4.1:** Accuracy Score of the different ML models

After evaluating all four models with the same datasets and under identical conditions, the RandomForestClassifier emerged as the best-performing model without any parameter tuning.

Having identified the RandomForestClassifier as the best-performing model in our initial evaluations, we proceed to enhance its performance through parameter tuning. To achieve optimal results, we employ GridSearchCV, a tool for cross-validating our model and identifying the best parameters.

After applying the tuned parameters and making predictions against the same test data, the RandomForestClassifier achieves an accuracy score of 89.79%, which is a significant improvement over the untuned model.

```
[CV] END bootstrap=True, max_depth=2, max_features=sqrt, min_samples_leaf=1, min_samples_split=5, n_estimators=100
[CV] END bootstrap=True, max_depth=2, max_features=sqrt, min_samples_leaf=1, min_samples_split=5, n_estimators=100
[CV] END bootstrap=True, max_depth=2, max_features=sqrt, min_samples_leaf=1, min_samples_split=5, n_estimators=100
pre_rf.predict(x_test)
acc = accuracy_score(y_test, pre)
print(f"Accuracy: {acc}")
Accuracy: 0.8979591836734694
```

**Fig4.2:** The accuracy score of a Tuned Random Forest Classifier model

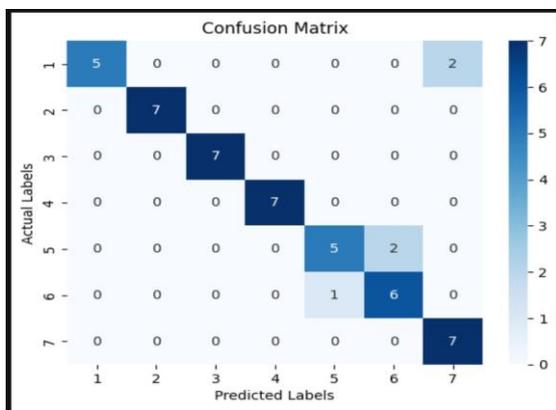
For further evaluation, we run a code to obtain a classification report of a tuned RandomForestClassifier model and from the classification report, we further understand the performance of our model, getting insight into how well the model performs across different instrument classes.

```
[56]: print("Classification Report:")
print(classification_report(y_test, pre))
```

Classification Report:							
	precision	recall	f1-score	support			
1	1.00	0.71	0.83	7			
2	1.00	1.00	1.00	7			
3	1.00	1.00	1.00	7			
4	1.00	1.00	1.00	7			
5	0.83	0.71	0.77	7			
6	0.75	0.86	0.80	7			
7	0.78	1.00	0.88	7			
accuracy				0.90	49		
macro avg				0.91	0.90	0.90	49
weighted avg				0.91	0.90	0.90	49

**Fig4.3:** Classification report of tuned Random Forest Classifier model

We also obtain the confusion matrix which visualizes the model's performance, showing the counts of true positive, true negative, false positive, and false negative predictions. This helps in understanding the model's strengths and weaknesses in classifying different musical instruments.



**Fig 4.4:** Confusion matrix heat map of tuned Random Forest Classifier model.

To facilitate interaction with our trained model, we develop a web-based user interface using the "Streamlit" library. The UI allows users to upload audio files in formats such as '.wav', '.mp3', and '.ogg'. Upon uploading, the extract\_features function is called to extract features from the audio file, and the model predicts the instrument based on these features.

## 4. Summary and Conclusion

This paper demonstrates how technology can be utilized to preserve Khasi folk musical instruments, which risk being overlooked due to technological advancements. Using Artificial Intelligence, we can document and analyze these instruments, ensuring their rich cultural heritage is not lost. The extracted features are crucial as they provide detailed information for various tasks, including the classification of musical instruments through Machine Learning models. By preserving these features, we can facilitate further study and research while introducing the sounds of these instruments to the technological world.

Our work shows that by extracting features and implementing ML, we can develop a classification model specifically for Khasi folk musical instruments—a novel approach for these instruments. This achievement underscores the potential for using technology to preserve cultural heritage. We hope this paper encourages others to integrate their cultural traditions with technological advancements, ensuring that unique folk music and instruments are celebrated and preserved in the modern world. By doing so, we aim to honor and maintain our cultural heritage while embracing the future.

## 6. Bibliography

1. **Syiem, Lapynshai.** *The Evolution of Khasi Music: A study of classical content.* Shillong: s.n., 2012.
2. *Classification of Musical Instruments using SVM.* **S. Prabavathy, V. Rathikarani, P. Dhanalakshmi.** 2020, International Journal of Innovative Technology and Exploring Engineering (IJITEE).
3. *Music Genre classification.* **Pandey, Parul.** 2018.
4. **Taiwo O A.** *Types of Machine Learning.* 2010.
5. **Midhula Vijayan, Mohan R, Preeth Raguraman.** *LibROSA-based assessment tool for music information retrieval system.* 2019.

6. **Diego A. Cruz, Cristian C. Cristancho, and Jorge E. Camargo.** *Automatic Identification of Traditional Colombian Music Genres Based on Audio Content Analysis and Machine Learning Techniques.* 2019.

7. **Marshilong, M. L., Moirangthem, B.S., and Yongyo, H.M.** *Features Extraction and Classification of Khasi Musical Instruments.* Unpublished undergraduate thesis. Department of Electronics and Communication Engineering, North-Eastern Hill University, 2024.

