ESP IJACT
ESP International Journal of Advancements in Computational Technology
Volume 1 / Issue 1 Paper Id: IJACT-V1I1P102 / Page: 5-10

Original Article

Indian Musical Instrument Recognition Using Integrated Mean Method

Seema Chaudhary[1], Dr Sangeeta Kakarwal[2]

[1]*Maharashtra Institute of Technology, Aurangabad(MS)*
[2]*Int. Centre of Excellence in Engineering and Management, Aurangabad(MS)*

Received Date: 03 March 2023 Revised Date: 16 March 2023 Accepted Date: 03 April 2022

Abstract: *Daily, numerous musical works are uploaded on social media platforms. The process of searching for content according to our preferences is time-consuming. One of the emerging research fields that is concerned with the process of extracting content from audio data is known as musical information retrieval. The field of musical information retrieval contains a subfield known as musical instrument recognition. Previous studies had primarily concentrated on a variety of western instruments that belonged to diverse families, such as brass, string, and woodwind instruments. The objective of this research is to categorize different types of musical instruments by making use of the Integrated Mean technique and acoustic features. The experiments make use of homophonic recordings of musicians performing solo on their instruments. Temporal and spectral aspects of sound have been accounted for in acoustic features. The new approach Integrated Mean method provides a vector that combines the mean that was obtained with the features that were extracted. A produced vector is used to categorize the musical instruments of the IRMAS and ISI-500 dataset. The proposed method obtained higher accuracy than using audio features independently. The K-nearest neighbour classifier has been utilized here for the purpose of classification.*

Keywords: *Musical Instruments, Machine Learning, k-NN, IRMAS Dataset, Audio Features, Integrated Mean*

I. INTRODUCTION

One of the most important aspects of human life is music. It has a calming effect on both your mind and body, while also connecting to your spirit. the vast majority of musicians worldwide today promote their work on social media platforms. On the internet, you can listen to music from an extremely wide range of genres. It can be challenging and difficult to find a specific one based on criteria, depending on the genre, performer, instrument, piece of music and so on that, you are looking for. The retrieval of musical information can be used for a different type of purposes[1-3], including the identification of musical instruments, the searching of songs based on their contents, the sorting of audio, the categorization of songs according to their genre, and the recognition of musical performers.

Each instrument produces a sound that is uniquely its own, which is influenced by the type of material it is made of, in addition to the instrument's dimensions and shapes. There are primarily four families of musical instruments, which include percussion instruments, string instruments, brass instruments, and wind instruments. Striking, plucking, and bowing are the three different ways that string instruments can be played, and this further divides string instruments into three groups. The classification of musical instruments into their respective families has been the subject of a significant amount of research. [4–8] It can be difficult to distinguish one instrument from another member of the same family of instruments.

A method that integrates Mean with extracted audio data is proposed in this research, and it is used to conduct an analysis of the timbre of various musical instruments. A k-NN classifier is used to analyze a solo dataset consisting of Indian string instruments after the dataset has been prepared. In addition to it, the benchmark dataset known as IRMAS [9] is taken into consideration during the evaluation process. Within the scope of this part, we have investigated the research that is associated with the field of musical instrument recognition.

The researchers made use of attributes to capture timbre's spectrum envelope, tonal and noise-like range, as well as Spectro temporal evolution. In terms of overall performance, it reveals that 1-NN classifiers are superior to SVM. The accuracy of the results produced by the GMM classifier is lower than that of the results obtained by the k-NN and SVM classifiers respectively [10].

Sarfaraz Masood, Shubham Gupta, and Shadab Khan used a neural network to test wind and stringed instruments[11].

ZCR, Spectral Rolloff, Skewness, Kurtosis, Brightness, Flatness, RMS, and MFCC were measured.

The researchers [12] identified the instruments by listening to the RWC Music Database and University of Iowa Musical Instrument Samples. Linear Discriminant Analysis (LDA) used Random Forest (RF) to achieve instrument recognition rates of 24.8%–82.1% for Iowa monotones' ZCR, Spectral Centroid, Pitch, and Brightness characteristics. Even with RF, RWC monotones have a 54.9% rate, maybe because there are not enough features to cover the enormous database's variability.

This suggested study[13] examined the role of Indian instruments in polyphonic musical signals using samples from the veena, sitar, flute, sarod, harmonium, and shankha. Researchers found 68 factors that identify polyphonic audio signal instruments. Calculating the mean and standard deviation of 34 characteristics per frame yielded these parameters. ZCR, Spectral, MFCC, and Chorma. KNN classifiers classify instruments. Classifiers with 68 parameters will have 68 dimensions. Selecting a suitable feature set reduces dimensions. The features are selected using those 68 factors, each of which can be scored as great, ordinary, or dreadful.

The spectrogram for the same note played on different instruments was shown to have distinct differences [14].The MFCC and spectral features are retrieved, and supervised approaches in addition to unsupervised machine learning algorithms, such as Hierarchical Clustering, are utilized. SVM fared significantly better than the other approaches.

A significant majority of these studies have conducted study on the categorization of western musical instruments. Indian musical instruments have received less attention [15-17]. The vast majority of the work that has been done to classify Indian musical instruments has focused on categorizing instruments that are members of a variety of families [18-20]. In this study, we used supervised learning to categorize Indian musical instruments that belong to the string family.

II. MATERIALS & METHOD

A. Datasets

This research made use of two different types of data. The first dataset is known as IRMAS and the second dataset is comprised of Indian string instruments (ISI) and had been produced from recordings of solo performances of Guitar, Santoor, Sitar, Sarod, Veena, and Violin. The wind, brass, and string families are all represented in the 5927 audio samples that are included in the IRMAS dataset. IRMAS dataset is polyphonic and the ISI-500 dataset of 3000 samples are homophonic music.

B. Audio Features Extraction

The classification of audio depends heavily on the characteristics of the audio. The temporal and spectral characteristics [21] are being considered for the purposes of this experiment. List of features are given in Table1.

When the time-domain feature known as zero crossing rate (ZCR) is evaluated, it performs computations directly on the samples of the signal, without modifying the original audio stream in any way. The ZCR is determined by counting the number of times the waveform of the audio signal goes through zero amplitude [22].

For the computation of spectral characteristics, the discrete Fourier transform (DFT) transformation on the signal is typically used, and the signal spectrum is taken into account.

Table 1: List of Audio Features

Type	Features	No. of Features
Temporal	Zero Crossing Rate (ZCR)	01
Spectral	Spectral Centroid (SC) , Spread Roll-off (SR) , Spectral Spread (SS)	03

C. Integrated Mean Method

The method that is being proposed begins by removing the section of the audio data files that is silent. Audio features are extracted from the signal after it has been processed. From audio samples, audio features such as ZCR, Spectral and Mean are retrieved. Every feature that was extracted is normalized. The Integrated Mean Method is presented in the following figure-1.

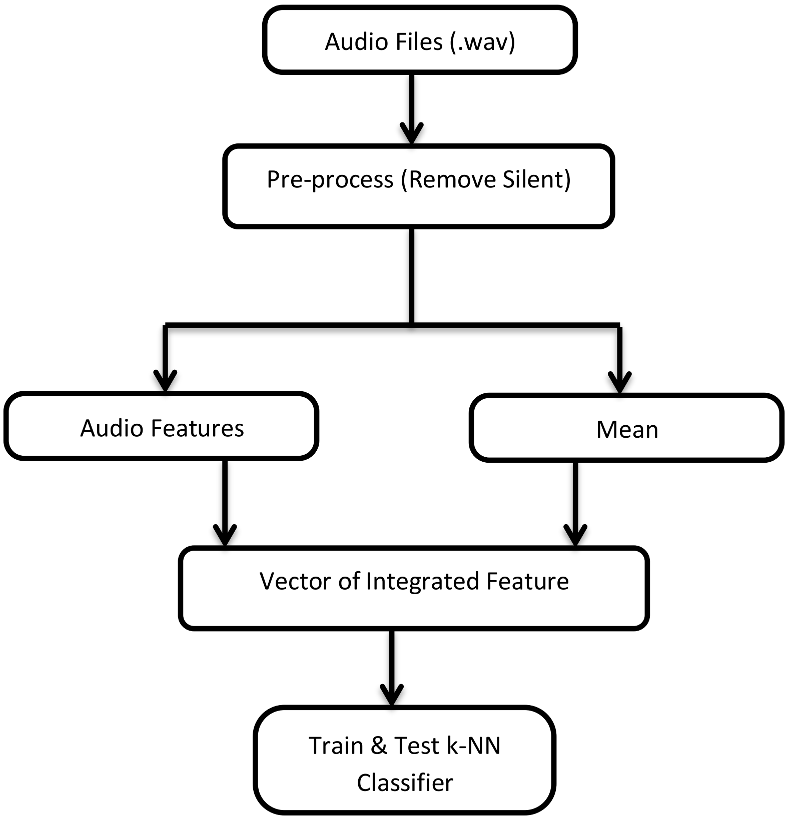

Figure 1: Block diagram of Integrated Mean Method

A vector of audio features that incorporates statistical feature Mean, is computed, and then it is provided to the k-NN classifier[23] as an input. For the purpose of the evaluation of the model, fivefold cross-validation is utilized.

III. RESULTS AND DISCUSSION

The classification of musical instruments has been accomplished by carrying out two separate experiments. In the first experiment, the temporal properties of the instruments have been taken into account, and the results have been analysed. The spectrum properties of the instruments, on the other hand, have been taken into account in the second experiment. IRMAS and ISI-500 datasets are utilized in the experiments that are carried out. The extraction of audio features utilizes the temporal and spectral properties of the features themselves.

The Max-Min approach is utilized in order to accomplish the normalization of the features. When dealing with values that are in the negative range, the absolute value of the lowest negative value is added to each value of that feature, and then Min-Max normalization is utilized. The evaluation is done using a method called five-fold cross-validation[24-25].

Experiment 1: Integrated Mean Method (IMM) using Temporal feature

In this experiment, the first vector of the temporal feature i.e. ZCR is examined. After that, it is integrated with the Mean to classify instruments.

Experiment 2: Integrated Mean Method (IMM) using Spectral features

Spectral features such as Spectral Rolloff, Spectral Centroid and Spectral Spread are evaluated with and without the Integrated Mean method.

The percentage of classifications by k-NN using audio features is presented in Table 2.

Table 2: Classification Accuracy (%) using Audio Features

Vector	IRMAS	ISI-500
ZCR	20.72	35.14
Spectral Centroid (SC)	17.58	46.36
Spectral Rolloff (SR)	18.67	48.86
Spectral Spread (SS)	15.67	48.06
Spectral Features (SC+SR+SS)	43.53	73.75

Table 3 shows the accuracy of k-NN using audio features with Integrated Mean Method.

Table 3: Classification Accuracy (%) using Audio Features with IMM

With Integrated Mean	IRMAS	ISI-500
ZCR	70.16	42.79
Spectral Centroid (SC)	65.70	55.68
Spectral Rolloff (SR)	65.77	57.21
Spectral Spread (SS)	65.45	56.91
Spectral Features (SC+SR+SS)	76.79	75.18

Integrated Mean Method (IMM) and using Independent Features on audio files are compared in the following figure, displays a comparative analysis of their respective levels of accuracy.

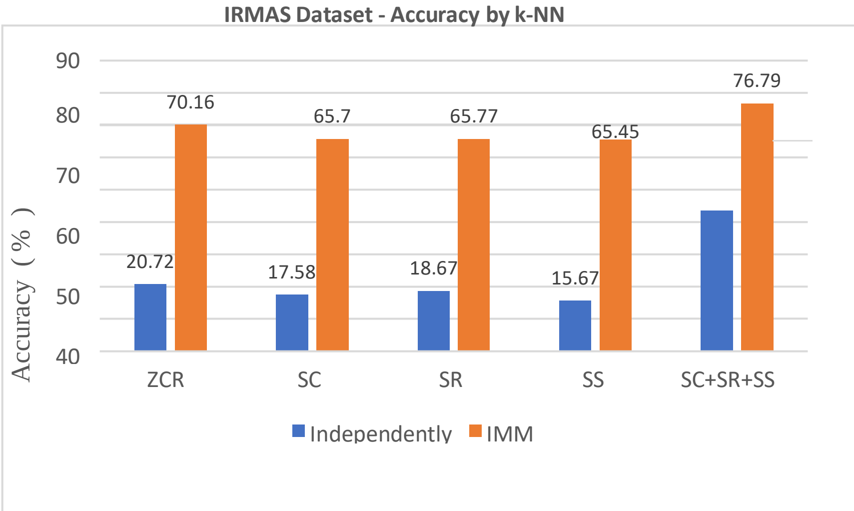

Figure 2: Accuracy by k-NN on IRMAS Dataset

IMM has been able to obtain a higher accuracy using zero crossing rate features up to 70.16 percent. When using IMM, the accuracy of all of the spectral features in IRMAS is 76.79%.

Figure 3 shows a relative analysis of the respective levels of accuracy of both methods on the ISI-500 dataset. In the Indian string instruments dataset, IMM with ZCR has given 42.79% accuracy which is higher than independent ZCR. Similarly, by using IMM improvement in accuracy has been observed for all spectral features.

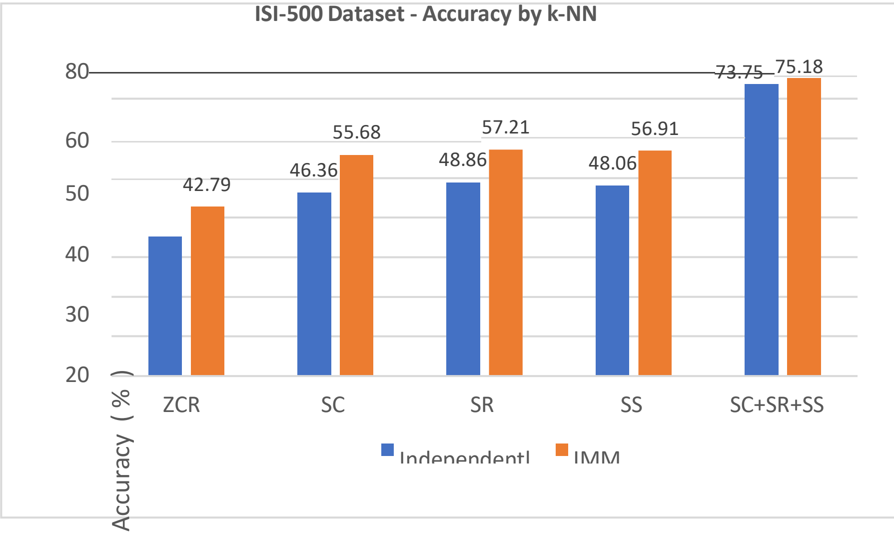

Figure 3: Accuracy by k-NN on ISI-500 Dataset

IV. CONCLUSION

When applied to the IRMAS and ISI-500 datasets, the combination of the spectral features Spectral Centroid, Spectral Rolloff, and Spectral Spread produces a high level of accuracy. In contrast, the ISI-500 dataset only contains string instruments, whereas the IRMAS dataset also includes wind and brass instruments in addition to string instruments. The Integrated Mean approach has given more accuracy than independent features in both the polyphonic IRMAS dataset and the homophonic ISI-500 dataset. The Integrated Mean technique showed a significant improvement in classification accuracy it offered for spectral and temporal audio features.

Interest of Conflicts

There is no conflict of interest concerning the publishing of this paper.

Acknowledgements

We would like to express our gratitude to the Dr. B. A. Marathwada University in India's Department of Computer Science and Information Technology for providing us with a substantial amount of support for our research.

V. REFERENCES

[1] Eghbal-Zadeh, H.; Dorfer, M.; Widmer, G. A COSINE-DISTANCE BASED NEURAL NETWORK FOR MUSIC ARTIST RECOGNITION USING RAW I-VECTOR FEATURES. 2016
[2] Caclin, A.; McAdams, S.; Smith, B. K.; Winsberg, S. Acoustic Correlates of Timbre Space Dimensions: A Confirmatory Study Using Synthetic Tones. J. Acoust. Soc. Am. 2005, 118 (1), 471–482. https://doi.org/10.1121/1.1929229.
[3] Chaudhary, S.; Kakarwal, S. Various Approaches in Musical Instrument Identification: A Review. Int. J. Appl. Evol. Comput. 2019, 10, 1–7. https://doi.org/10.4018/IJAEC.2019040101.
[4] Agostini, G.; Longari, M.; Pollastri, E. Musical Instrument Timbres Classification with Spectral Features. EURASIP J. Adv. Signal Process. 2003, 2003 (1), 1–10. https://doi.org/10.1155/S1110865703210118.
[5] Mazarakis, G.; Tzevelekos, P.; Kouroupetroglou, G. Musical Instrument Recognition and Classification Using Time Encoded Signal Processing and Fast Artificial Neural Networks. In Advances in Artificial Intelligence; Antoniou, G., Potamias, G., Spyropoulos, C., Plexousakis, D., Eds.; Lecture Notes in Computer Science; Springer: Berlin, Heidelberg, 2006; pp 246–255. https://doi.org/10.1007/11752912_26.
[6] Eronen, A.; Klapuri, A. Musical Instrument Recognition Using Cepstral Coefficients and Temporal Features. In 2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100); 2000; Vol. 2, pp II753-II756 vol.2. https://doi.org/10.1109/ICASSP.2000.859069.
[7] Brown, J. C.; Houix, O.; McAdams, S. Feature Dependence in the Automatic Identification of Musical Woodwind Instruments. J. Acoust. Soc. Am. 2001, 109 (3), 1064–1072. https://doi.org/10.1121/1.1342075.
[8] Ghisingh, S.; Mittal, V. K. Classifying Musical Instruments Using Speech Signal Processing Methods. In 2016 IEEE Annual India Conference (INDICON); IEEE: Bangalore, India, 2016; pp 1–6. https://doi.org/10.1109/INDICON.2016.7839034.
[9] IRMAS Dataset: https://zenodo.org/record/1290750
[10] Franz A. de Leon, Kirk Martinez, “Music Genre Classification Using Polyphonic Timbre Models”, IEEE International Conference on Digital Signal Processing,2014
[11] Sarfaraz Masood, Shubham Gupta and Shadab Khan “Novel Approach for Musical Instrument Identification Using Neural Network”, IEEE INDICON, 2015.
[12] Yuta Takahashi, Kazuhiro Kondo, “Comparison of Two Classification Methods for Musical Instrument Identification”, IEEE 3rd Global Conference on Consumer Electronics (GCCE), 2014

[13] S.V. Chandan, Mohan R. Naik, Ashwini, A. Vijay Krishna,"Polyphonic Indian Instrument Identification KNN Classifier",IEEE Conference, 2019

[14] Karthikeya Racharla, Vineet Kumar, Chaudhari Bhushan Jayant, Ankit Khairkar and Paturu Harish,"Predominant Musical Instrument Classification based on Spectral Features",arXiv:1912.02606v2 [eess.AS] 21 Apr 2020 ,https://doi.org/10.1109/SPIN48934.2020.9071125 [15]Sundararajoo, Kohshelan, "Improvement of Audio Feature Extraction Techniques in Traditional Indian String Musical Instrument", Master Thesis, 2016

[16] Akshata Shelke, Abhijit Chitre, "An Effective Feature Calculation For Analysis & Classification Of Indian Musical Instruments Using Timbre Measurement", ICCCT , ACM,2015,pp. 101-105,doi.org/10.1145/2818567.2818586

[17] M. Joshi and S. Nadgir, "Extraction of feature vectors for analysis of musical instruments," 2014 International Conference on Advances in Electronics Computers and Communications, Bangalore, 2014, pp. 1-6, doi: 10.1109/ICAECC.2014.7002391

[18] Burred, J., Röbel, A. and Sikora, T., "Dynamic spectral envelope modelling for timbre analysis of musical instrument sounds", IEEE Transaction on Audio, Speech, Language Processing, Mar. 2010.

[19] Wu J., Vincent E., Raczynski, S.A., Nishimoto, T., Ono N., Sagayama S., "Polyphonic pitch estimation and instrument identification by joint modeling of sustained and attack sounds", IEEE Journal of Selected Topics in Signal Processing, Vol. 5,Oct.,2011.

[20] Kazi F I, Bhalke D G, "Musical Instrument Classification using Higher Order Spectra and MFCC", IEEE International Conference on Pervasive Computing-2015

[21] Url : https://www.jyu.fi/hytk/fi/laitokset/mutku/en/research/materials/mirtoolbox/manual1-7.pdf MIRtoolbox 1.7 - page-175

[22] Peetersa Geoffroy, Bruno L. Giordano, Patrick Susini and Nicolas Misdariis, Stephen McAdams,The Timbre Toolbox: Extracting audio descriptors from musical Signals, .J. Acoustical Society of America,2011

[23] Dr. Sangeeta N. Kakarwal ,Seema Chaudhary, "Analysis of Musical String Instruments using k-NN", IEEE International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing, 2020

[24] A. Ghosal, R. Chakraborty, R. Chakraborty, S. Haty, B. C. Dhara, and S. K. Saha, "Speech/music classification using occurrence pattern of ZCR and STE," in 2009 Third International Symposium on Intelligent Information Technology Application, vol. 3. IEEE, 2009, pp. 435–438.

[25] Seema Chaudhary,Dr.Sangeeta Kakarwal,Dr.R.R.Deshmukh, "Musical Instrument Recognition using Audio Features with Integrated Entropy Method", J. Integr. Sci. Technol. 2021,