International Journal of Future Generation Communication and Networking
Vol. 13, No. 2s, (2020), pp. 18–24

# Musical Instrument Sound Classification using Deep Convolutional Neural Network

Ankur Aaditya[1], R. S. Kothe[2], Aman Srivastava[3], C Rangrajan Kumar[4]

*[1,2,3,4] Dept. of E & TC Engg., Smt. Kashibai Navale College of Engineering, Savitribai Phule Pune University, Pune*

*[1]an.kur10skn@gmail.com*
*[2]rskothe@gmail.com*
*[3]amansrivastava383@gmail.com*
*[4]rangrajan8899@gmail.com*

***Abstract***

*MFCC is one of the mostly used feature in automatic musical instrument sound classification. We also used some higher order spectral features such as spectral centroid, spectral flux and spectral kurtosis along with MFCC. The MFCC features resembles human auditory system thus gives better accuracy for identification. These features together comprise to make the model robust and accurate. CNN is used here for classification purpose. It detects the features and builds a network to identify different instruments.*

***Keywords***— *MFCC, CNN, Features, MIR*

## I. Introduction

Music is one of greatest creation of human mankind in course of history. Music speaks what cannot be expressed, makes mind peaceful. Music is used not only for entertainment and pleasure but also for psychological effects and treatments. Music makes mind peaceful, happy and relaxed. Musical instrument classification deals with recognition and classification of musical instruments. The demand for online access of music data on internet is increasing day by day. Therefore, the need for development of computational tools for the analysis, summarization, classification, and indexing of music data has increased.

Over a decade, Music Information Retrieval (MIR) has been making efforts to provide a reliable solution for several music related tasks. Musical instruments sound classification is a sub-task of MIR and deals with identification of individual musical instruments and its family. Musical instrument classification problem consists of three steps: pre-processing, feature extraction and classification. Majority of the research on musical instrument classification is focused on feature extraction and feature analysis, of which feature extraction is vital. Finding compact, efficient and robust feature set is a major challenge in instrument classification. In addition, selection of a suitable classifier plays an important role in improving classification accuracy.

Here CNN is applied to extracted feature vectors of raw audio data. Both spectral and temporal features are used for finding better accuracy.

## II. Literature survey

In the proposed system, the model tries to solve the trouble of MIR. The use of features gaining knowledge of system based on the network of deep convolution. Music signal is represented as spectrograms which is given as input to deep CNNs. The CNN model consisting of 3 convolutionary layers, 3 fully connected layers and a softmax layer extract features of spectrograms and predictions of ten instrumental classes is done. Prediction described by the CNN model for individual spectrogram performance as signal to update the acceptance values for all instrumentals. [1].

This paper presents an enhanced deep convolutional neural network for recognizing musical instrument from large musical database. MFCC feature extraction algorithm is used to extract the

ISSN: 2233-7857 IJFGCN

features from the sample musical instruments. The extracted features are trained for predicting the instrument exactly. The proposed algorithm classifies the instrument according to the extracted features. The result shows that the proposed algorithm classifies the musical instrument with the accuracy of 97.5% [2]

In the proposed work, music instrument recognition using deep convolution neural network is accomplished. The system receives input in the form of sampled audio signal which further converted to Mel spectrogram form. The network receives input in the matrix representation form of Mel spectrogram. This work is accomplished using eight layers deep convolution neural network [3].

Two state-of-the-art source separation methods was firstly evaluated and showed that on multitimbral audio data, analysing the harmonic and solo streams can be beneficial compared to the mixed audio data. For the specific use-case of jazz solo instrument classification, which involves classifying six instruments with high timbral similarity, combining solo/accompaniment source separation and transfer learning methods seems to lead to AIR models with better generalization to unseen data [4].

Musical instrument classification method using convolutional neural networks (CNNs) was presented. Unlike the traditional methods, they investigated a scheme for classifying musical instruments using the learned features from CNNs. To create the learned features from CNNs,not only conventional spectrogram image was used, but also multi resolution recurrence plots (MRPs)were proposed that contain the phase information of a raw input signal.By combining the proposed MRPs and spectrogram images with a multi-column network, the performance of proposed classifier system improved over a system that uses only a spectrogram [5].

Identification of polyphonic sound in music is a very complex and difficult task. It auto-tags piece of musical sound by the instrument, and then search that piece of music in music database by the instrument. It can be used in creating other application such assource separation, genre recognition, music transcription, and instrument specific equalizations. The methods were reviewed for the work which also includes recent developments in the convolutional neural network. The major problem with this learning models was that they required large amount of annotated data[6].

In the proposed system instrument identification with convolutional neural networks three different experiments were designed. For all approaches, a multi-track classical music dataset incorporating 13 different instruments was employed. After applying the herein developed crosstalk reduction method on the spot microphone tracks, the individual instrument activities were first determined to then binarize, features wereextracted in form of Mel-spectrograms from the according stereo mixtures. To examine the effects of frame size to classification performance, labels as well as features were generated with different frame sizes. After the training phase, the learned model was able to predict the instrument composition for the given frame size.While the former can improve classification for small frame sizes by utilizing classifiers which were trained on larger frames but only classify centreframes, the latter did not show any enhancements [7].

In this paper, the performance of western musical instruments was studied which uses the higher order moments, such as skewness and kurtosis. Supplementary statistical information of the signal over the conventional first or second order momentswas contained in high order moments. Higher-order moments provide segment level features. Experimental results proved that low level features integrated using higher order moments improves the recognition accuracy [8].

The classification of musical instruments extraction of timber from sound signal has been subject of several experimentation, where various spectro-temporal parameters have been proposed and compared. Direct classification and hierarchical classification are two discrimination approaches of classification strategies.Both strategies are based on factthat the feature vector is static and is used during the whole treatment process. From this paper, author experimented on echelons classification where the feature vector is dynamic and changes depending on each level and each node of the echelons tree. The feature vector was optimized and was determined with the sequential backward

ISSN: 2233-7857 IJFGCN

selection (SBS) algorithm. Using a large database (RWC), the results show, a score gain in musical instrument recognition performances with the proposed approach [9].

From this paper, the role of various features with different classifiers on automatic classification of musical instruments was examined. Piano, flute, trumpet, guitar, xylophone and violin were classified using various attributes and classifiers.For this purpose, spectral features like spectral centroid, spectral slope, spectral spread, spectral kurtosis, spectral skewness and spectral roll-off were used along with zero crossing rate, autocorrelation and Mel Frequency Cepstral Coefficients (MFCC) were used. The classification of musical instruments accuracy depended on these features was studied for different classifiers. The analysis also confirmed the selection of features and classifiers with the results [10].

In this paper, the author presented a new innovative method for quantitative estimation of the musical instrument categories which composes a music piece. The two methods used were for a wavelet-based music source (i.e., musical instrument) separation algorithm. In the first step,the musical instruments which composes music are separated using a source separation technique based on wavelet packets. For the second step, a categorization algorithm based on support vector machines was applied to approximate the musical type of each of the musical instruments recognize in the first step. This method performssuccessfully when evaluated on the publicly available Iowa Musical Instrument Database. [11].

A new acoustic model using decision trees (DTs) is used in this paper as replacements for Gaussian mixture models (GMM). It computes the observation likelihoods for a given hidden Markov model state in a speech recognition system. DTs are selected by author because of their advantageous properties, such as that they do not impose restrictions on the number or types of features, and that they automatically perform feature selection. High-level information such as gender or contexts can be directly incorporated into DTAMs (Decision Tree based Acoustic Models) using equal or decoding questions. Feature usage analysis is also better in DTAM than GMM. Context-dependent DT-based models are highly compact compared to conventional GMM-based acoustic models [12].

This paper presents an empirical study on classical musical instrument classification. A number of experiments was used to assess the proposed methodology of feature extraction and evaluation. Experiments involving features like MFCC, MPEG-7 Audio Descriptors, kNN, SVM, GMM etc. are studied. The paper was aimed to detect instruments in solo passages. Total 20 Musical instruments from 4 families have been used for this research work [13].

This paper presents the use of kNN classifier for automatic classification of musical instruments. The signal is segmented and framed with 441 samples in each frame. Then the signal was passed through hamming window which reduces signal to zero from both sides. This ensures that processing is concentrated on middle part of signal in time domain. Feature used for classification is MFCC as it represents something near to human auditory system [14].

This paper presents an analysis on cepstral features and spectral features. Cepstral features such as MFCC is used. 13 coefficients were found necessary and spectral features were extracted using MIR toolbox. 8 musical instruments were used for the experimentation. Only coefficient 1 was found useful for MFCC. The experimentation concluded that only Roll-off and centroid features are useful for classification [15].

## III.METHODOLOGY

Various attempts have been made to construct automatic musical instrument classification system. Researchers have used different approaches and scopes, achieving different performances. But, the literature survey study states that, finding reduced and efficient feature scheme is a major challenge for musical instrument classification. In order to achieve this objective, we have proposed automatic musical instrument classification using MFCC features using CNN for isolated musical instrument

sounds. This work can be extended for polyphonic musical instrument sound classification. In polyphonic musical instrument classification, several instruments are played simultaneously.
The timbral and similar features like spectral decrease, spectral centroid, spectral flux, spectral kurtosis etc., can be extracted and used in same way for classification of polyphonic musical sound signals. The proposed automatic musical instrument classification system is depicted. The input to the system is musical instrument sound in the form of wave file and output is name of instrument and its family. We applied pre-processing to the raw file to reduce redundancies using methods like Zero Crossing Rate, trimming the signals to small length. The complete block diagram is described here.

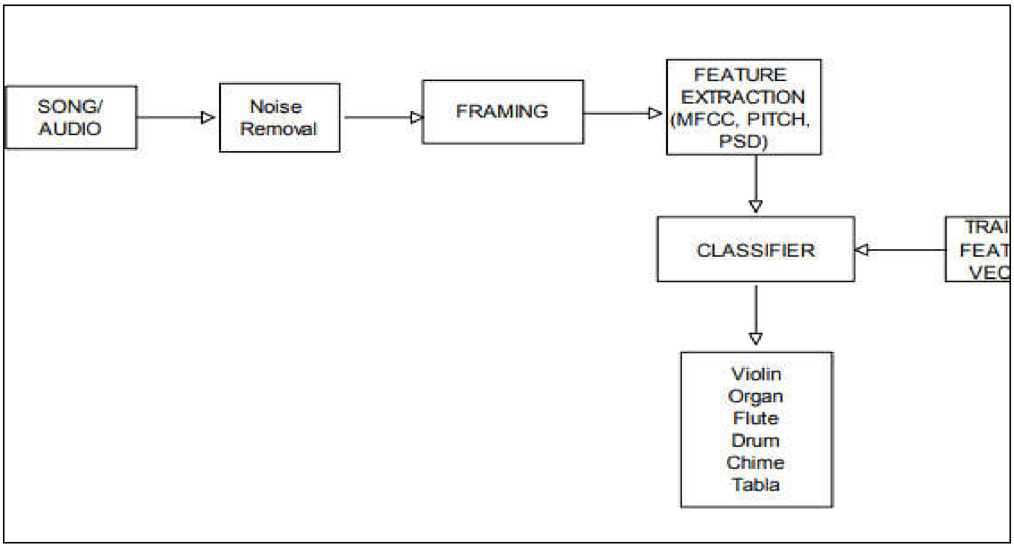


Fig. 1 Block Diagram of Proposed Model

*1)Pre-Processing:* The dataset is firstly trimmed to lengths of few microseconds. It helps in training easily and then framing is done. We also applied Zero Crossing Rate to make sure the void part of signal is removed before processing.

*2)Feature Extraction:* After the pre-processing is applied, the main aspect is to extract features which are useful and robust.Features such as Spectral Centroid, Spectral Kurtosis, Spectral Spread, Spectral Roll-off, MFCC, Time Peak Envelope etc are used to make sure that classification is done with better accuracy between in-class and inter class instrument family.

*3)Spectral Spread:*It is the standard deviation around the spectral centroid. It represents the “instantaneous bandwidth” of the spectrum.It indicates dominance of a tone.

*4)Spectral Centroid:* It is the frequency-weighted sum normalized by the weighted sum. The spectral centroid represents the “center of gravity” of the spectrum. It is used to represent the indication of brightness. Genre classification and music analysis is generally done using this feature.

*5)Spectral Kurtosis:*It is a measure of flatness, or non-Gaussianity, of the spectrum around its centroid. It is computed from fourth order moment.

*6)MFCC:*MFCC are the short time power spectral representation of a signal and represents psychoacoustic property of the human auditory system. MFCC has been used extensively in speech analysis over past few decades and recently received more attention in music analysis. Here, attempts have been made to identify and classify musical instruments using MFCC features. Computation of MFCC features for a segment of music signal consists of pre-emphasis, framing, windowing, FFT, triangular scale band pass filtering and DCT which are: - Pre-emphasis, Frame Blocking.

*7)CNN:*A convolutional neural network (CNN) is a specific type of artificial neural network that uses perceptron, a machine learning unit algorithm, for supervised learning, to analyse data. CNNs apply to image processing, natural language processing and other kinds of cognitive tasks. A convolutional neural network is also known as a ConvNet,a convolutional neural network (CNN, or ConvNet) is a class of deep neural networks, most commonly applied to analysing visual imagery. There are various steps such as: - Convolutional Operation, Pooling, Flattening and Full Connection.

## IV.EXPERIMENTS AND RESULTS

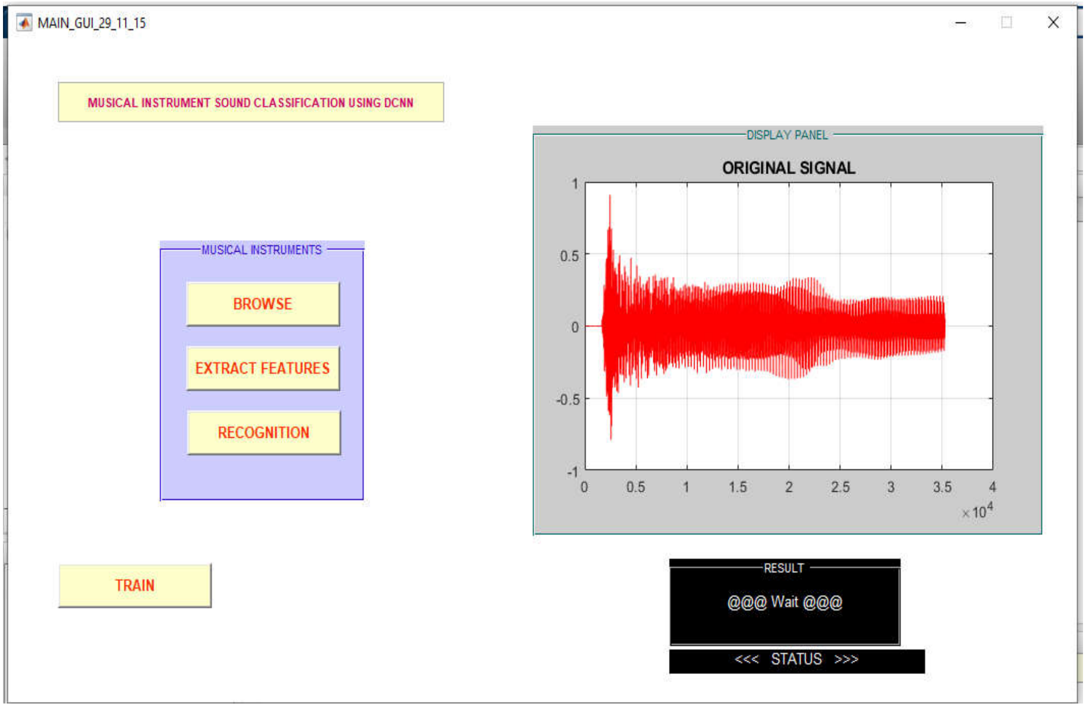


Fig. 2Original Signal

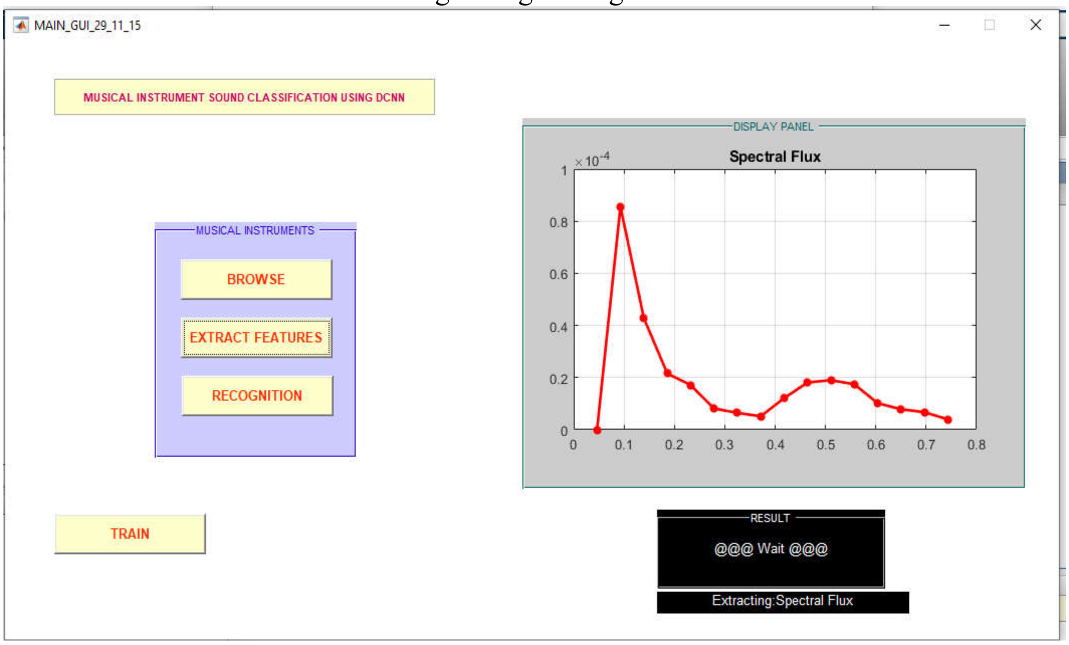


Fig. 3Feature Extraction (Spectral Flux)

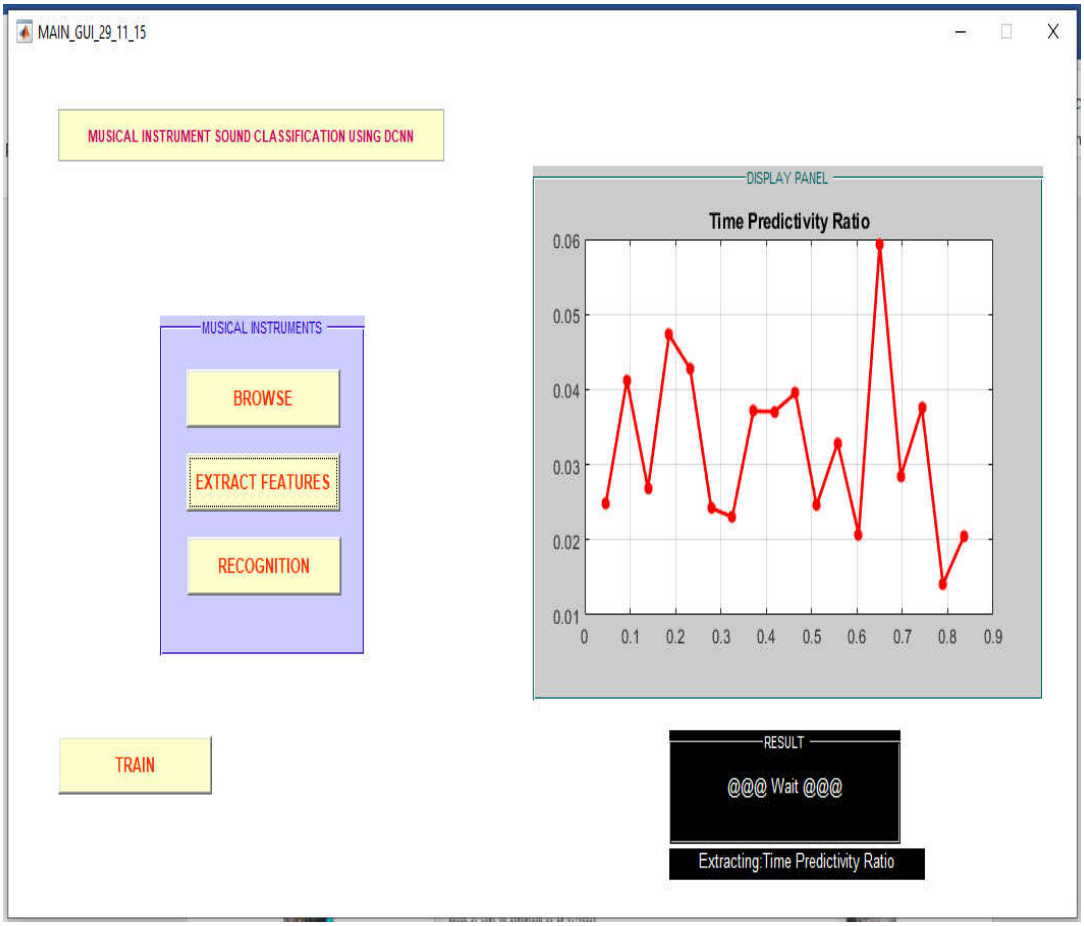


Fig. 4Feature Extraction (Time Predictivity Ratio)

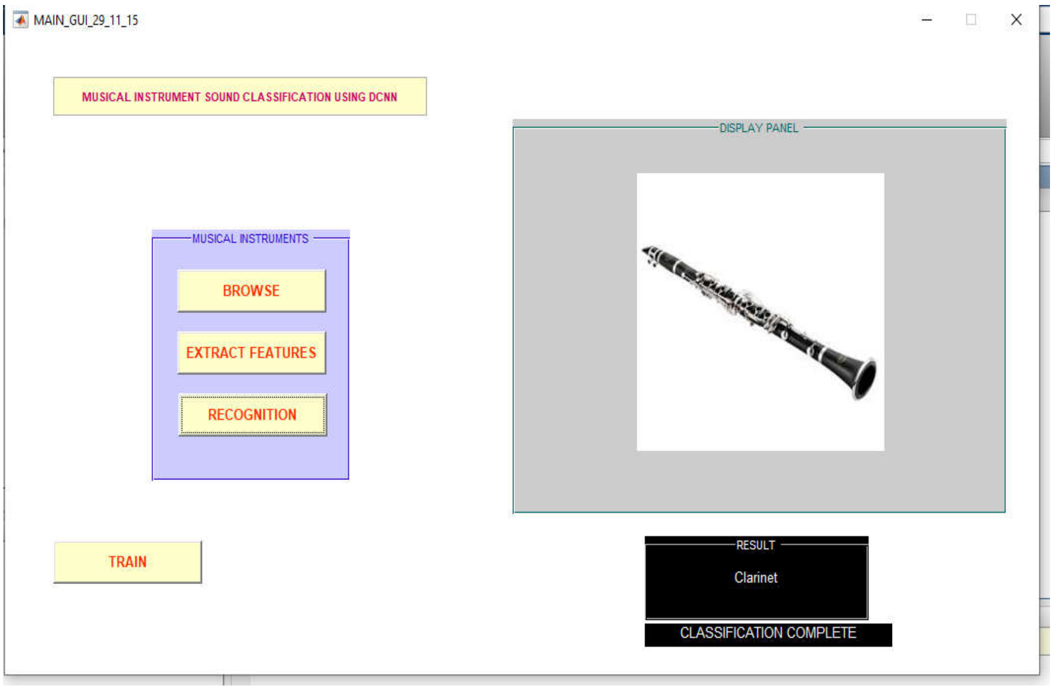


Fig. 5Recognition (Clarinet)

## V. CONCLUSIONS

An epic component extraction plot for programmed order of melodic instruments sound utilizing MFCC highlights have been proposed. The proposed highlights expand the between-class variety and limit the inside class variety and increment the separating capacity of the highlights with same computational multifaceted nature as MFCC. The proposed highlights catch the dynamic time fluctuating nature of sound sign due to trill like portion premise capacity and help to catch acoustic qualities of music sound flag superior to MFCC. CNN has been used for prediction and classification.

## REFERENCES

[1] G. Jawaherlalnehru and S. Jothilakshmi,” Music Instrument Recognition from Spectrogram Images Using Convolution Neural Network”, International Journal of Innovative Technology and Exploring Engineering (IJITEE), Volume-8 Issue-9, July 2019
[2] S. Prabavathy, V. Rathikarani and P. Dhanalakshmi, “An Enhanced Musical Instrument Classification using Deep Convolutional Neural Network”, International Journal of Recent Technology and Engineering (IJRTE), Volume-8 Issue-4, November 2019.
[3] Arun Solanki and Sachin Pandey"Music instrument recognition using deep convolutional neural networks",School of Information and Communication Technology,Gautam Buddha University, Greater Noida, Uttar Pradesh 201308, India , 2019
[4] Juan S. Gomez, Jakob Abeber and Estefania Cano, “Jazz Solo Instrument Classification with Convolutional Neural Networks, Source Separation, and Transfer Learning”, Proceedings of the 19th ISMIR Conference, Paris, France, September 23-27, 2018.
[5] Taejin Park and Taejin Lee, “Musical Instrument Sound Classification with Deep Convolutional Neural Network using Feature Fusion Approach”, Electronics and Telecommunications Research Institute (ETRI), Republic of Korea,2017
[6] Siddharth Bhardwaj, “Audio Data Augmentation with respect toMusical In- strument Recognition”,Universitat Pompeu Fabra,2017
[7] Nevo Segal, “Automatic Musical Instrument Recognition Using Convolutional Neural Networks”, MSc in Digital Signal Processing.2016
[8] Bhalke, C. Rao, and D. Bormane, (2013) “Musical Instrument Recognition Using Higher Order Moments,” Digit. Signal Process.vol. 5, no. 4, pp. 133–138.
[9] E. Hall, H. Ezzaidi, and M. Bahoura, (2012) “Study of Feature Categories for Musical Instrument Recognition State of Art,” Springer-VerlagBerlin Heidelb., pp. 152–161
[10] Chandwadkar, D M, Sutaone, M, (2012) “Role of Features and Classifiers on Accuracy of Identification of Musical Instruments,” CISP2012— Proceedings, pp. 66–70.
[11] M. Akamineand J. Ajmera, (2012) “Decision tree-based acoustic models for speech recognition,” EURASIP J. Audio, Speech, Music Process., vol. 1, pp. 2-10.
[12] E. Hall, H. Ezzaidi, and M. Bahoura, (2012) “Hierarchical Parametrisation and Classification for Musical Instrument Recognition,” 11th Int. Conf. Inf. Sci. Signal Process. Their Appl. Main Tracks.
[13] D. Deng, C. Simmermacher, and S. Cranefield, (2008) “A Study on Feature Analysis for Musical Instrument Classification,” IEEE Trans. Syst. Man. Cy- bern., vol. 38, no. 2, pp. 429–438.
[14] M. S. Nagawade and V. R. Ratnaparkhe, "Musical instrument identification using MFCC," 2017 2nd IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT), Bangalore, 2017, pp. 2198-2202.
[15] G. S. R., B. S. S. and S. S. D., "Cepstral (MFCC) Feature and Spectral (Timbral) Features Analysis for Musical Instrument Sounds," 2018 IEEE Global Conference on Wireless Computing and Networking (GCWCN), Lonavala, India, 2018, pp. 109-113.

ISSN: 2233-7857 IJFGCN