International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Application of Elman Back Propagation Neural Network for Automatic Identification of Tabla Strokes in North Indian Classical Music

Shambhavi Shete[1], Saurabh Deshmukh[2]

[1]Career Point University, Kota, Rajasthan, India

[2]Maharashtra Institute of Technology, Aurangabad, Maharashtra, India

Abstract: ***Tabla is the most useful accompanying percussion instrument used in North Indian Classical Music. The homophonic sound produced from Tabla instruments generates multiple harmonics. Therefore, Tabla stroke identification is a challenging task. Tabla stroke identification has various applications such as Tempo Estimation, Beat Tracking, Rhythm Identification, Tala Recognition, Automatic Music Transcription to name a few. This research aims to compare Elman Neural Network (ENN) with various neural network architectures useful for automatic Tabla stroke identification. This comparison would be useful to appropriately select the ENN for the applications of Tabla stroke identification. Studio recordings of 640 Tabla audio excerpts, sampled at 44, 100 Hz sampling frequency are used to train and test the neural networks namely, Feed Forward Back Propagation Neural Network (FFBPNN), Pattern Recognition Neural Network (PRNN), Elman Neural Network (ENN), Cascade Forward Neural Network (CFNN), and Recurrent Neural Network (RNN). The audio features are extracted using traditional Mel Frequency Cepstral Coefficient (MFCC) and Timbral audio descriptors along with MFCC. The Tabla strokes are categorized into two major categories namely, Open and Closed Tabla strokes. The result shows that Tabla stroke identification accuracy is obtained higher for open strokes due to the difference of Attack, Decay, Sustain, and Release values of the strokes. The Tabla stroke identification accuracy of 94.1% is achieved using ENN, for Timbral audio features.***

Keywords: Music Information Retrieval, Timbre, MFCC, Elman Neural Network, North Indian Classical Music

1. Introduction

Sound information retrieval is a branch of sound engineering that deals with information extraction from sound and its utilization for a given application. The Indian classical music is divided into two types, North Indian Classical Music (NICM) and Carnatic Classical Music (South Indian Classical Music). The difference in the two traditions is in terms of style of singing, note presentation, variations in the structure of musical notes, and the accompanying instruments used. In NICM the music is structured around a group of musical notes called Raga [1]. The vocalist or an instrumentalist performs recitation of the raga in different tempos, namely VilambitLaya (Slow Tempo), Madhya Laya (Mid-Tempo), and DrutLaya (Fast Tempo). The vocalist is accompanied by different types of musical rhythm instruments such as Tabla, Pakhawaj, Dholaki, Dhol, Mrudangam, and Dholak. In this research, Tabla musical performances are considered.

Automatic Tabla stroke identification has various applications such as Tempo Detection [2], Tala (Rhythm) Recognition [3], Sound Source Separation [4] [5], Automatic Music Score Generation [6], and Musical Notes/Rhythm Transcription [7] [8]. Automatic identification of Tabla stroke is accomplished using audio features. These features are called audio descriptors. The audio descriptors are the features that uniquely describe audio. There are two major types of audio descriptors, temporal and spectral. Audio descriptors are also classified based on the time at which they are considered, e. g., instantaneous audio descriptors and audio descriptors of the entire audio signal. Mel Frequency Cepstral Coefficient (MFCC) is proven to be one of the best suitable audio features for the applications of speech and music [9]. Timbre is the non-tangible, fourth dimension of a sound. This fourth dimension is multidimensional. The Timbral audio descriptors considered here are Zero Crossing Rate (ZCR), Roll Off, Roughness, Brightness, Irregularity, and Mel Frequency Cepstral Coefficient (MFCC) [10].

The Tabla strokes are identified using various neural networks. The algorithms considered here are Feed Forward Back Propagation Neural Network (FFBPNN) [11], Pattern Recognition Neural Network (PRNN) [12], Elman Neural Network (ENN) [13], Cascade Forward Neural Network (CFNN) [14], and Recurrent Neural Network (RNN) [15]. The aim is to analyze and compare the efficiency of ENN with other neural network architectures to correctly identify/predict Tabla strokes from given excerpts of Tabla audio recording.

2. Literature Survey

North Indian Classical Music (NICM) is based on melody, rhythm, and harmony. Tabla is one of the important accompanying percussion instruments which provide explicit rhythmic structure. Tabla is a set of two drums of different sizes and shapes. Tabla sound is produced by tapping a finger on the drums with hands (left, right, or both). The methods with which the fingers are tapped produces different types of sound called Tabla strokes. The open Tabla strokes are those which are generated through Tabla by hitting the Tabla skin with open fingers leaving a sustaining voice of the stroke being played while the closed Tabla strokes are generated by hitting the Tabla with closed figures stopping all the other overtone that may be generated through the Tabla instrument. These strokes are sequenced together to create different rhythm patterns of Tala. The rhythmic structure of Tabla strokes is called as Tala [16]. Tala prediction is one of the most important research areas in the transcription system [17].

Tala consists of a different number of rhythmic strokes such as Teentala (16 strokes) or Ektala (12 strokes). The Tabla strokes are categorized into three parts. Tabla strokes originated from the left drum, strokes that are originated from the right drum, and combined strokes generated by hitting on both the drums simultaneously. The basic stroke considered for Tabla instrument also includes speedy alteration of both the drum strokes.

When the Tabla player strikes on the skin of Tabla a resonating sound is produced. This generated sound has a homophonic texture. In contrast to the western music where all the musical instruments in an orchestra are played in different melody lines simultaneously producing polyphonic texture of the music. The drum used in western music has different components tuned in different frequencies, however, the resonating sound generated from the left drum, right drum, or both the drum of Tabla produces a non-separable mixture of various harmonic overtones produced from Tabla [18]. Along with this, in NICM, the harmonic sound is produced from various other accompanying musical instruments such as Tanpura and Harmonium. Previously, Automatic Tabla stroke labeling is accomplished by using MFCC as an audio feature and Hidden Markov Model (HMM) as a classifier that achieved Tabla stroke identification accuracy of 93.4%. Complex domain thresholding is applied to the segment time-domain Tabla stroke signal. A Feed Forward Neural Network (FFNN) and Probabilistic Neural Network (PNN) resulted in the highest Tabla stroke identification accuracy.

Another musical rhythm instrument called Mrudangam stroke identification is obtained by considering three parts of each of the stroke of Mrudangam namely onset, attack time, and decay. A three-state 1-mixture HMM was applied for three parts of each Mrudangam strokes. Further, the Mrudangam strokes were transcribed into basic symbols. Different audio features were compared in identifying the Tabla strokes. It was concluded that MFCC was the most suitable audio descriptor when used with a support vector machine (SVM) for the identification of the Tabla strokes [19].

Attempts were made to improve the quality of audio feature extraction and classification. A novel indirect method of audio features selection useful for gender recognition of a singer of NICM uses the concept of a two-pass algorithm. The algorithm back-tracks the utility of an audio descriptor or combinations of various audio descriptors from the accuracy of male or female singer identification. For male singer identification, a combination of ZCR, MFCC, Roughness, and irregularity was found to be useful. For female singer identification, a combination of Roll Off, Brightness, ZCR, and MFCC was found to be most suitable [20]. Typically, the left drum strokes of the Tabla instrument exhibit low values of attack, decay, sustain, and release (ADSR) and is not harmonic if compared with the right Tabla drum.

Musical instrument identification is another application of the music information retrieval technique. Interestingly, musical instruments and a singing voice that sounds like the human ear (e. g., male singing voice vs flute or female singing voice vs violin) are treated as a single instrument. The classification accuracy for a database of mixed audio samples containing singing voice and musical instrument sound was obtained as 91.1% for Timbral audio descriptors [21].

Audio feature selection is a challenging task because it directly affects the accuracy of the Tabla strokes identification system. From the pool of hundreds of audio descriptors, it is difficult and challenging to use appropriate audio descriptors. In a wrap-per approach of audio descriptors selection process hybrid selection algorithm attempts different combinations of audio descriptors in the forward pass. The algorithm considers the classifier accuracy for the elimination of audio descriptors in the backward pass. This method has the advantage of attempting various sets of audio descriptors, however, it increases the system complexity. An accuracy of 96.67% is obtained with a reduced set of audio descriptors ZCR, Roll Off, Brightness, and Irregularity. Timbral Audio descriptors play important role in the singer identification process [22].

3. The Method

The automatic Tabla stroke identification system is as shown in Fig.1. The Tabla stroke identification system uses different neural networks that are trained and tested over the audio excerpts containing Tabla stroke samples. The input audio dataset consists of all the basic Tabla strokes originated from the left drum, right drum, both the drums simultaneously, and alternate patterns of fast Tabla strokes, which are treated as single Tabla strokes in NICM (Ti-Ra-Ki-Ta or Dha-Ge-Na-Ti). The Tabla strokes are manually labeled as open strokes or closed strokes. A total of 640 samples of Tabla strokes are recorded in the studio at a sampling frequency of 44100 Hz and in 16-bit PCM. wav file format. The audio files are preprocessed to remove noise. The audio files are segmented based on the onsets detected from each stroke and segmented into individual strokes. Peak picking algorithm is used to identify and select the stroke locations. The onset detection and peak picking algorithm locate the exact positions of the Tabla strokes which are further segmented. The segmented Tabla strokes are then further fed to Audio Feature extraction methods namely MFCC and Timbre. The Timbre contains Audio Descriptors specified by Olivier [10] The Timbral audio descriptor contains ZCR, Roll-off, Brightness, Roughness, Irregularity, and MFCC.

The Elman Neural Network used here is a two-layer backpropagation network. It is a partial Recurrent Neural Network. This network has additional feedback given from the output of the hidden layer to the input. The ENN, due to this feedback system recognizes and develops temporal patterns. Each time before giving the decision the neuron consider the information of the past internal states.

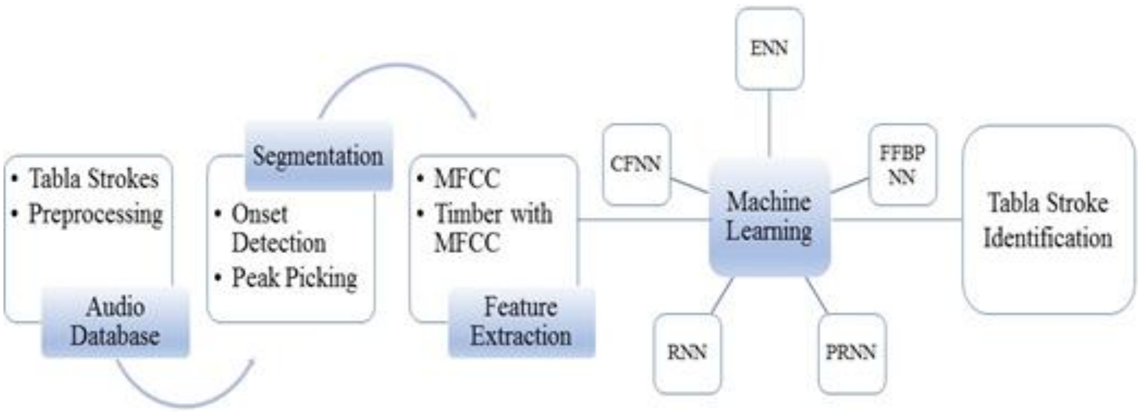

Figure 1: Automatic Tabla Stroke Identification System

The ENN is popularly used in data prediction and classification. The state layer is updated with external input of the network as the previous forward propagation. The feedback from the hidden layer to input is modified through the set of weights. Thus, an automatic adaptation through learning is achieved. In a fully connected ENN each neuron receives input from every other neuron in the network.

To evaluate the performance of ENN, four different Neural Networks along with ENN are trained and tested to identify the Tabla strokes. As a thumb rule, 70% of the input audio excerpts are used to train each neural network and 30% of samples are used to test the network. After testing, each identified Tabla stroke is again manually labeled as open stroke or closed stroke. This labeling is done to analyses the performance of the neural network for open and closed strokes. Out of sixteen basic Tabla strokes considered here, there are eight open Tabla strokes and eight closed strokes.

The performance of each neural network is calibrated by calculating stroke identification accuracy. The percentage accuracy is calculated by multiplying the ratio of the number of Tabla strokes correctly identified to the total number of Tabla strokes given for testing with 100.

4. Experiments and Results

The experiments are carried out on the dataset containing 16 strokes (640 audio excerpts). The audio feature extraction methods namely, MFCC and Timbre with MFCC are used to extract audio features. The system is trained using these audio features and the performances of five neural network architectures are compared. Fig.2 shows a comparison of the performance of neural network architectures for the identification of Tabla strokes.

Feed Forward Back Propagation Neural Network (FFBPNN) is a fully connected neural network with an input layer consisting of neurons collecting all the audio features and an output layer comprises different classes representing different Tabla strokes. The network architecture of (25: 10: 16) implies that 25 audio features in the input layer, 10 neurons in the hidden layer, and 16 neurons in the output layer representing 16 basic Tabla strokes are used. Here, the learning algorithm used is Levenberg-Marquardt (LM). The FFBPNN is one of the efficient neural networks widely used for classification. However, due to back propagation learning, and application of steepest descent to update the weights, the neural network suffers from a slow convergence rate.

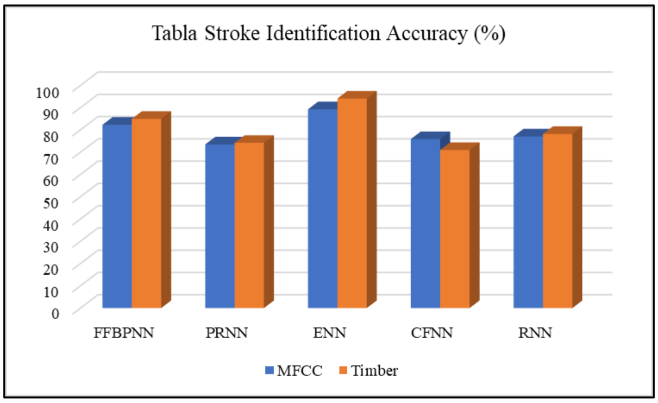

Figure 2: Tabla Stroke Identification Using MFCC and Timbre

Another type of Feed Forward network is Pattern Recognition Neural Network (PRNN) which is trained to classify the inputs according to the target classes. The target data for PRNN is designed such that it contains vectors of all zero values while one at the element that represents the class. In PRNN each input pattern is assigned as a unique label that describes a basic Tabla stroke. The pattern recognition algorithm in supervised learning trains the network for the Target data. A recurrent neural network with back propagation through time (BPTT) learning algorithm uses a feedback connection from a neuron's output to its input. Such kind of RNN is suitable for dealing with a sequence of spectral frames in a spectrogram. Recurrent Neural Network has additional recurrent connections inside each layer. In this network feedback connection from a neuron output to input is given. To deal with audio descriptors that are frame-based, the feedback from the output layer to the hidden layer of RNN simply acts as memory.

The Cascade Forward Neural Network (CFNN) architecture is like FFBPNN except for the network connections such that the outputs of the hidden layer are given as input to every successive hidden layer and output layer. The back propagation learning algorithm of CFNN is like the back propagation learning algorithm of FFBPNN. CFNN implemented here has four layers (25: 10: 10: 16) which

Paper ID: SR22419123931 DOI: 10.21275/SR22419123931

imply that 25 input neurons, 10 hidden layer-1 neurons, 10 hidden layer-2 neurons, and 16 output layer neurons, along with network connections given from hidden layer-1 to hidden layer-2 and output layer. The additional connections given between the hidden layer and the output layer improves the data distribution and increases the neural network generalization. The hidden layer neurons and the output layer neurons use nonlinear and sigmoid activation functions.

The Tabla stroke identification accuracy for all the network architectures is compared using MFCC and Timbral audio descriptors. It is observed that the CFNN architecture that cascades the hidden layer to the output layer does not work well with Timbral audio descriptors. However, for the rest, all the neural networks, the combination of Timbral audio descriptors along with MFCC leads to accurate Tabla stroke identification. The ENN in this system performs well and provides automatic Tabla stroke identification accuracy of 94.1%. The Tabla strokes are of two types, open and closed. The open strokes contain harmonics and have a higher value of the attack, decay, sustain, and release (ADSR) than closed stroke. The combination of Timbral audio descriptors along with MFCC thus proves to be efficient feature extraction to improve the stroke classification accuracy. Due to lower values and abrupt decay of closed stroke, the stroke exhibit similar feature values. Thus, making it difficult to identify the closed strokes accurately.

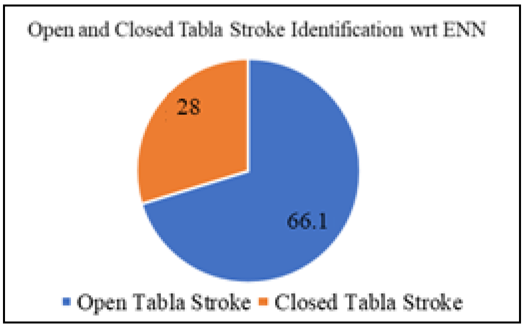

Figure 3: Tabla Stroke Identification Using Timbre

Fig.3 shows the percentage share of open and closed strokes in Tabla stroke identification when Timbral audio descriptors and Elman Neural networks are used. The pie chart shows that out of 16 basic Tabla strokes, the open strokes could be identified more accurately than the closed strokes.

5. Conclusion

Automatic Tabla stroke identification is one of the popular research. For various music genre identification, music transcription and mood identification of song, the basic rhythm of a song needs to be identified first. To identify the rhythm, Tabla strokes should be located and segmented using onset detection and segmentation, respectively.

In this research, we have compared the performances of five different neural network architectures to identify the Tabla strokes. Due to the homophonic nature of the sound, Timbral audio descriptors along with MFCC are used. It is observed that the Elman Neural Network (ENN) gives 94.1% automatic Tabla stroke identification accuracy. The difference in the ADSR values of open and closed Tabla strokes affects the stroke identification accuracy to a great extent. It is also found that open Tabla strokes are easy to identify due to their harmonic structure. However, the closed Tabla strokes due to their abrupt damping are difficult to identify.

References

[1] P. Chordia and A. Rae, "Raag Recognition Using Pitch-Class and Pitch-Class Dyad Distributions, " in International Conference on Music Information Retrieval, ISMIR 2007, Austria, 2007.

[2] S. Bhaduri, S. K. Saha and C. Majumdar, "Matra and Tempo Detection for INDIC Talas, " in Advanced Computing and Informatics Proceedings of the Second International Conference on Advanced Computing, Networking, and Informatics (ICACNI-2014), Switzerland, 2014.

[3] L. C. Martin Richard, "The Rhythmic Organisation of North Indian Classical Music: Tal, Lay and Laykari, " London, November, 1992.

[4] D. Barry, D. Fitzgerald, E. Coyle and B. Lawlor, "Drum source separation using percussive feature detection and spectral modulation, " in IEE Irish Signals and Systems Conference 2005, Dublin, Ireland, 2005.

[5] D. FitzGerald, "Automatic Drum Transcription and Source separation, " ARROW[at]TU Dublin, Dublin, 2004.

[6] S. Gupta, "Discovery of Percussion Patterns from Tabla Solo Recordings, " 20 September 2015.

[7] O. Gillet and R. Gael, "Automatic Labelling of Tabla Signals, " in 4th ISMIR Conference, 2003.

[8] J. Kuriakose, K. Chaitanya, S. Padi, H. Murthy and S. Umayalpuram, "Akshara Transcription of Mrudangam Strokes in Carnatic Music, " in 2015 Twenty First National Conference on Communications (NCC), Mumbai, India, 2015.

[9] "Mel-frequency cepstrum, " Wikipedia, 21 December 2019. [Online]. Available: https: //en. wikipedia. org/wiki/Mel-

[10] O. Lartillot, "MIR ToolBox Manual, " Finnish Centre of Excellence in Interdisciplinary Music Research.

[11] P. R. Kshirsagar and D. G. Akojwar, "Classification and Prediction of Epilepsy using FFBPNN with PSO, " in 2015 IEEE International Conference on Communication Networks (ICCN), Gwalior, 2015.

[12] O. Ahmad AL-Allaf, S. AbdAlKader and A. A. Tamimi, "Pattern Recognition Neural Network for Improving the Performance of Iris Recognition System, " International Journal of Scientific & Engineering Research, vol.4, no.6, April 2013.

[13] S. Nidhyananthan, Selva Kumari, R. Shantha and Shenbagalakshmi, "Assessment of Dysarthric Speech using Elman Back Propagation Network (Recurrent Network) for Speech Recognition, " International Journal of Speech Technology, pp.577-583, July 2016.

[14] B. Warsito, R. Santoso, Suparti and H. Yasin, "Cascade Forward Neural Network for Time Series Prediction, " in IOP Conference Series: Journal of Physics: Conf. Series 1025 (2018) 012097.

[15] R. Vogl, M. Dorfer and P. Knees, "Recurrent Neural Networks for Drum Transcription, " in ISMIR Conference, New York, USA, 2016.
[16] M. Miron, "Automatic Detection of Hindustani Talas, " UniversitatPompeuFabra, Barcelona, 2011.
[17] Shete Shambhavi, Deshmukh Saurabh. (2021) North Indian Classical Music Tabla Tala (Rhythm) Prediction System Using Machine Learning. In: Advances in Speech and Music Technology. Advances in Intelligent Systems and Computing, vol 1320. Springer, Singapore. https: //doi. org/10.1007/978-981-33-6881-1_16
[18] R. Siddharthan, P. Chatterjee and V. Tripathi, "A Study of Harmonic Overtones Produced in Indian Drums, " Physics Education, 1994.
[19] Shete Shambhavi and Deshmukh Saurabh, "Analysis and Comparison of Timbral Audio Descriptors with Traditional Audio Descriptors Used in Automatic Tabla Bol Identification of North Indian Classical Music, " in Proceeding of International Conference on Computational Science and Applications, Algorithms for Intelligent Systems, Pune, 2019.
[20] Saurabh Deshmukh and Sunil Bhirud, "A Novel Method to Identify Audio Descriptors, Useful in Gender Identification from North Indian Classical Music Vocal, " International Journal of Computer Science and Information Technologies, vol.5, no.2, pp.1139-1143, 2014.
[21] Saurabh Deshmukh, D. Bajaj and Sunil Bhirud, "Audio Descriptive Analysis of Singer and Musical InstrumentIdentification in North Indian Classical Music, " IJRET: International Journal of Research in Engineering and Technology, vol.04, no.06, pp.505-508, June 2015.
[22] Saurabh Deshmukh and Sunil Bhirud, "North Indian classical music's singer identification by timbre recognition using MIR toolbox, " International Journal of Computer Applications, vol.91, no.4, pp.1-5, 2014.