FOREX Publication
Open Access | Rapid and quality publishing

International Journal of
Electrical and Electronics Research (IJEER)
Research Article | Volume 12, Issue 1 | Pages 73-82 | e-ISSN: 2347-470X

Indian Classical Music Recognition using Deep Convolution Neural Network

Swati Aswale[1*], Dr. Prabhat Chandra Shrivastava[2], Dr. Ratnesh Ranjan[3] and Seema Shende[4]

[1]*Research Scholar, G.H. Raisoni University, Amravati, India; swatiaswale31@gmail.com*
[2]*Assistant Professor, JK Institute of Applied Physics Allahabad University, India; prabhatphd@gmail.com*
[3]*Assistant Professor, G.H. Raisoni College of Engineering and Management, Pune, India; ratnesh.ranjan.ece.gmail.com*
[4]*Research Scholar, G.H. Raisoni University, Amravati, India; seemashende18.gmail.com*

***Correspondence:** Swati Aswale; swatiaswale31@gmail.com

ABSTRACT- A divine approach to communicate feelings about the world occurs through music. There is a huge variety in the language of music. One of the principal variables of Indian social legacy is classical music. Hindustani and Carnatic are the two primary subgenres of Indian classical music. Models have been trained and taught to distinguish between Carnatic and Hindustani songs. This paper presents Indian classical music recognition based on multiple acoustic features (MAF) consisting of various statistical, spectral, and time domain features. The MAF provides the changes in intonation, timbre, prosody and pitch of the musical speech due to different ragas. The lightweight DCNN is used to improve the representation of the raga sound and to provide higher order abstract level features. The overall performance of the raga type is estimated using various performance metrics, including accuracy, precision, recall and F1-score. The proposed DCNN achieves an accuracy, precision, recall, and F1-score of 89.38%, 0.89, 0.89, and 0.89, respectively, for eight raga classifications. The extensive experimentation on eight classical ragas has shown a noteworthy improvement over the traditional state of art.

Keywords: Music recognition, Indian Raga Classification, Deep Convolution Neural Network, Spectral Features, Speech Recognition.

ARTICLE INFORMATION

Author(s): Swati Aswale, Dr Prabhat Chandra Shrivastava, Dr. Ratnesh Ranjan and Seema Shende;

Received: 28/10/2023; **Accepted**: 13/01/2024; **Published**: 05/02/2024;
e-ISSN: 2347-470X;
Paper Id: IJEER 2810-29;
Citation: 10.37391/IJEER.120112
Webpage-link:
https://ijeer.forexjournal.co.in/archive/volume-12/ijeer-120112.html

crossref member
CROSSREF.ORG
THE CITATION LINKING BACKBONE

Publisher's Note: FOREX Publication stays neutral with regard to Jurisdictional claims in Published maps and institutional affiliations.

1. INTRODUCTION

Music is one of the most vital components of India's Immaterial Culture legacy, which holds significant importance worldwide. Indian classical music consists of two major subgenres: Hindustani and Carnatic. Hindustani is rehearsed in the northern region of India, while Carnatic is prevalent in the southern region [1]. This musical distinction between the two subgenres dates back to the 16th century. Hindustani music encompasses vocal styles like Thumri, Tarana, Dhrupad, Khayal, Dadra and Gazals, whereas Carnatic music includes Kalpnaswaram, Niraval, Ragam Thana Pallavi and Alpana. Carnatic music employs instruments such as veena, mandolin and mridangam to perform 72 ragas, while Hindustani music uses instruments like sarangi, tabla, santoor and sitar and focuses on six major ragas [2]. Carnatic has a single, directed chanting style, whereas Hindustani boasts various sub-styles. In Hindustani music, the vocal aspect takes precedence, whereas both vocal and instrumental elements are equally emphasized in Carnatic [3].

The recent development in music history and retrieval involves the categorization of Raga Music based on Indian Classical Music. The abundance of musical data available on the internet enables this research. However, audio processing, particularly in speech and music, is still in the early stages of development. This study discusses speech processing as a potential foundation for classifying classical music using tools such as MFCCs, Spectrograms, and Scalograms. The research analyzes and extracts sound and music characteristics to categorize various musical genres. The study addresses the early stages of musical signal analysis, including pitch class profiles and acoustic trait-based statistical measurements. Encouraging findings and performance comparisons are provided. Future research will utilize various computer techniques to study diverse musical genres. The analysis of music not only helps us understand society's history and the cultures from which it has evolved but also contributes to the creation of scientific models. While most research focuses on Western music, there are also studies examining the sound of Indian classical music. Carnatic music and Hindustani music are the two primary subgenres of Indian traditional music, each having a significant fan base. Carnatic Music is notably more sophisticated in the manner of notes presented and structured [4]. Raga and Talam are the foundations of Indian classical music, where ragas are more intricate than Western music regarding song and scale. Ragas are composed of notes organized to evoke specific moods. Swara is the musical term for a note in Carnatic music, with each note associated with a specific frequency [5]. In Carnatic music, a song's rhythm is based on Talam, which denotes the order of syllables and the musical composition's pace. Talam is indicated through hand gestures in Carnatic music. This study uses Raga patterns to distinguish between Hindustani and Carnatic classical music, considering the musical note as the fundamental building block of Indian traditional music. The intervals between notes, known as swara, are characterized by the ratios of their fundamental frequencies. There are seven melodic notes: Sa, Ri, Ga, Ma, Pa, Dha and Ni, with frequencies that can be additionally isolated into semitones or microtones. Hindustani and Carnatic music use different scale types, namely, the 12-note scale and the 16-note scale [1][4][8].

Hindustani music employs a 12-note scale, while Carnatic music uses a 16-note scale. Carnatic music has identified 12 different frequency components [9][10]. Various deep master frameworks have been presented for music recognition in the past. For instance, Dipti Joshi's work with two classifiers, KNN and SVM, on Ragas Yaman and Bhairavi achieved a 90% result on the database [11]. Choi et al. proposed a music labeling plan given CNN [12]. Abdul et al. combined 2-D DCNN with a Mel-spectrogram of the music sign to give inert elements and a higher component portrayal capacity [13]. Other researchers have utilized techniques like CNN, CRNN, LSTM, and hybrid classification approaches to study and classify music genres [14] to [21]. Their studies offer valuable insights into understanding and categorizing Indian classical music. Indian classical music is widely recognized and streamed on social media and community platforms. Indian raga classification is also challenging due to the variability in the languages and corpora of the songs. Still, very few researchers have focused on the classification of Indian Ragas. Thus, there is a need to analyze the different Indian ragas that play an imperative role in the music industry. Musical speech has a wide variety of intonation, pitch, timbre, and prosody, which is essential to capturing and describing the distinctiveness of the signal. The musical speech representation is challenging because of the variety of ragas in Indian classical music. Thus, this work presents the multiple acoustic features that combine the spectral, time domain, and voice quality features for describing the musical signal. The previous systems have used heavy and complicated DL frameworks, increasing the systems' computational intricacy. Thus, there is a need to provide a lightweight DL architecture that needs lower trainable parameters and lower computational intricacy.

This paper offers Indian Raga grouping utilizing a DCNN. The commitments of the proposed work are summed up as follows:

- Representation of the Indian raga music signal using a set of spectral, temporal, and voice quality features to characterize the impact of raga on the voice signal.
- Representation of musical speech using multiple acoustic features that encompasses spectral, time-domain, and voice quality features to characterize the distinctiveness of the Indian ragas.
- Implementation of a five-layered DCNN to improve the distinctiveness of the traditional multiple acoustic features for enhancement in raga classification.

The proposed raga classification is assessed for eight Indian ragas such as Asawari, Bageshwari, Bhairavi, Bhoopali, Darbari, Malkans, Sarang and Yaman based on accuracy, precision, recall and F1-score.

The rest of the article is arranged as follow: *Section 2* provides the methodology in details that focuses on various spectral features, temporal features, voice quality features and the DCNN model used for proposed Raga classification. *Section 3* depicts the details about experimental results and discussions on the results, Later, *section 4* gives concise findings and the future scope for potential improvement in the proposed system.

░ 2. PROPOSED METHODOLOGY

2.1 MTMFCC Features

In the generalized MFCC, a Hamming window with higher variance is utilized, but it may fail to capture subtle variations in the speech signal frames. The signal is filtered during the stage of pre-emphasis to reduce the noise. In the multi-taper windowing technique, the entire signal is divided into frames of 40ms. Signal is converted in frequency domain using DFT. The linearly scaled signal is subsequently transformed to Mel frequency, which corresponds to understanding human hearing. The transformed signal is changed to the time domain using DCT to reduce signal redundancies. For feature extraction, 13 cepstral coefficients are chosen, as they are computed after log filter-bank power has been calculated over the frames. *Figure 2* illustrates the MT-technique MFCC. In comparison, the multi-taper MFCC utilizes different tappers with varying characteristics for windowing the signal, which catch minute variations in the signal [22-23].

The sign weighted ceptrum estimator (SWCE) outcomes in decrease errors fee as contrasted to MFCC, is given by equation 1[29][30].

$$wP(j) = \sqrt{\frac{2}{N+1}} sin(\frac{\pi p(j+1)}{N+1}), j = 0,1,\ldots.,N-1. \qquad (1)$$

Where, N denotes number of frames, *wp* is tapper window, and p=1,2,3,….,M .

The weights of SWCE tapers are computed using *equation 2* [31].

$$\lambda(p) = \frac{cos(\frac{2\pi(p-1)}{M/2})+1}{\sum_{P=1}^{M}(cos(\frac{2\pi(p-1)}{M/2})+1)}, p = 1,2,\ldots,M. \qquad (2)$$

Here, λ(p) is weight of pth taper, M represent variety of tapers and p =1,2,3,……., M. The power spectral density (PSD) of the dysarthric signal over the various tapering home windows is calculated the use of equation three [29][30].

$$SMT(m,k) = \sum_{p=1}^{M} \lambda(p) \left|\sum_{j=0}^{N-1} wp(j)s(m,j)e^{\frac{2\pi ik}{N}}\right|^2 \qquad (3)$$

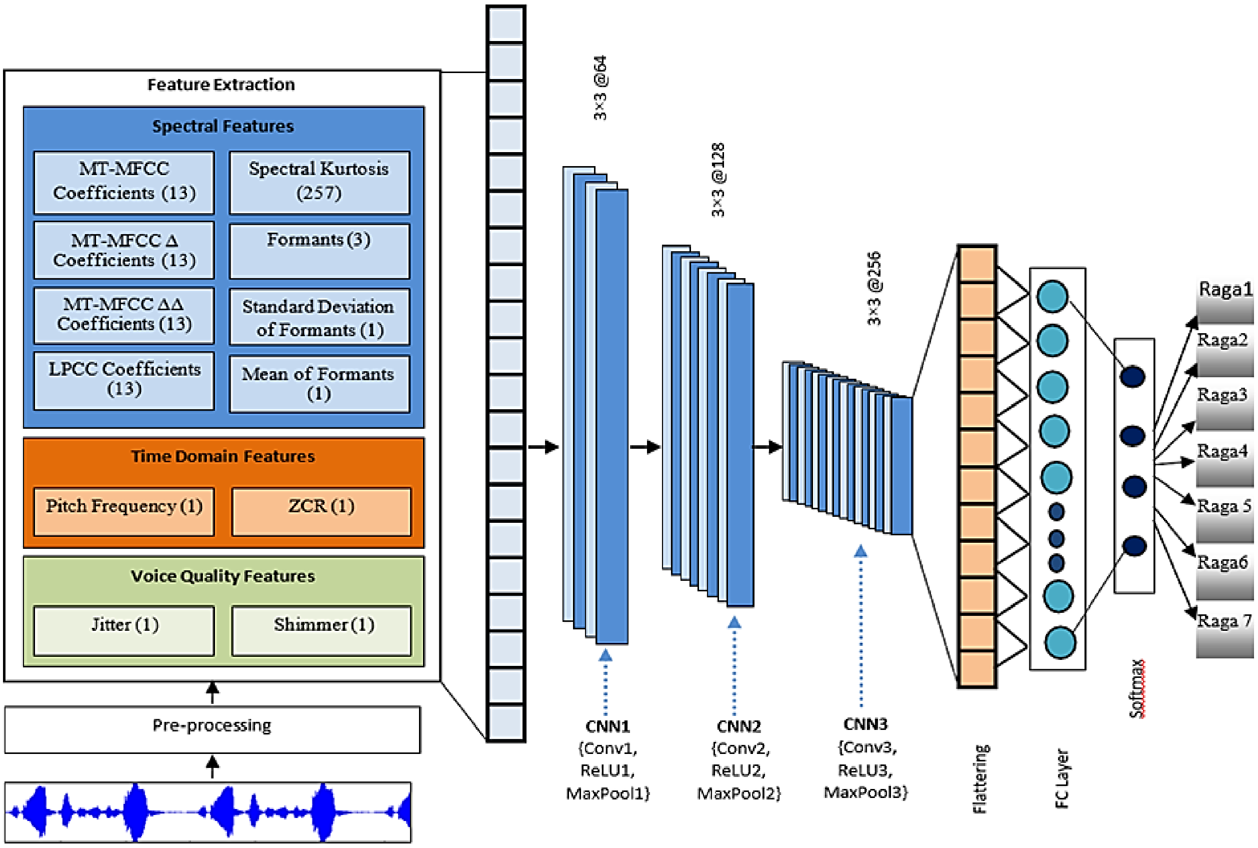


Figure 1: shows diagram of the proposed work that encompasses DCNN framework for Indian Raga classification.

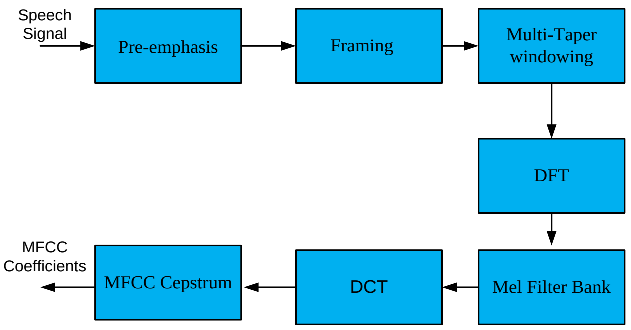


Figure 2. MFCC flow diagram [33]

Discrete Cosine Transform (DCT) transformed Speech signal into the frequency domain signal. The spectral features capture modifications in the spectral domain of the voice signal due to intonation, prosodic and node changes in the Indian classical ragas. The feature set comprises 13 MTMFCC coefficients, 13 Δ-MTMFCC coefficients and 13 ΔΔ-MTMFCC coefficients. The MTMFCC coefficients reflect the cepstral changes in energy over the frames of the voice signal. The Δ and ΔΔ coefficients capturing changes in the pitch of the voice signal.

2.2 LPCC

Different Indian ragas depict different emotions. To record information related to emotions expressed through vocal tract characteristics, this study employs LPCCs (Linear Predictive Cepstral Coefficients). A frame shift of 10ms and a 10th order LP analysis on the speech signal provide thirteen LPCCs for each speech frame of 20ms [24][25].

The LPC considers the previous samples knowledge for the estimation of future nth coefficients using *equation 4*.

$$x(n) = a_1x(n-1) + a_2x(n-2) + a_3x(n-3) + \ldots\ldots\ldots + a_px(n-p), \quad (4)$$

Where a_1, a_2, ….., a_p are the constants over the music signal. The error between actual sample x (n) and predicted $\hat{x}(n)$ is computed using *equation 5*.

$$x(n) = x(n) - \hat{x}(n) = x(n) - \sum_{k=1}^{p} a_k s(n-k) \quad (5)$$

To find the distinctive predictive constants, the sum of the squared difference of error (e_n) between $\hat{x}(n)$ and x(n) is calculated utilizing *Equation (6)*. Here, m characterizes the total samples in music frame.

$$e_n = \sum_m \left| x(m) - \sum_{k=1}^{p} a_k x(m-k) \right|^2 \quad (6)$$

The LPCC are estimated by resolving *equations (7)– (10).*

$$\frac{dE_n}{da_k} = 0 \quad \text{for } k = 1,2,3,\ldots\ldots..p \quad (7)$$

$$C_0 = \log_e(p) \quad (8)$$

$$\mathrm{LPCC_m} = \mathrm{a_m} + \sum_{k=1}^{m-1} \frac{k}{m} \mathrm{C_k a_{m-k}}; \qquad \text{for } 1 < m < p \qquad (9)$$

$$\mathrm{LPCC_m} = \sum_{k=m-p}^{m-1} \frac{k}{m} \mathrm{C_k a_{m-k}}; \qquad \text{for } m > p \qquad (10)$$

2.3 Formants

Formants are contributing to defining the resonance produced by the aroha (gradual increase of voice) and avroha (gradual diminution of voice) during raga singing. The study considers three formant frequencies along with the standard deviation and mean of the formants [26][27]. The formants, its mean and standard deviation are provided by *equation 11-13.*

$$\mathrm{frm} = \{f_1,\ f_2,\ f_3\} \qquad (11)$$

$$\mathrm{frm_u} = \frac{f_1 + f_2 + f_3}{3} \qquad (12)$$

$$\mathrm{frm_\sigma} = \sqrt{\frac{\sum_{i=1}^{3}(f_i - \mathrm{frm_u})^2}{3}} \qquad (13)$$

2.4 Pitch Frequency

The voice signal's fundamental period is known as pitch. It represents the perceived equivalent frequency and reflects that frequency when vocal cords vibrate for producing sound. The pitch represents the voice texture of the singer while singing ragas [28][29].

3. ZCR

The ZCR allows assessing both voiced and unvoiced data, providing insights into the number of times the waveform switches polarity in a given time period [30][31].

The sign gives positive one and zero for positive amplitude and negative amplitude respectively as given in *equation 14.*

$$\mathrm{ZCR_t} = \frac{1}{2}\left(\sum_{n=1}^{N}(\mathrm{sign}(x[n]) - \mathrm{sign}(x[n-1]\right) \qquad (14)$$

3.1 Jitter and Shimmer

Jitter and shimmer are terms used to describe variations in amplitude and frequency of the emotional signal caused by periodic vibrations of vocal cord. They represent the breathiness, hoarseness and roughness of the emotional voice [32]. The jitter (Jt) and shimmer (Sh) are computed using *equation 15 and 16* where A, T and N represents the peak-to-peak amplitude, time period and number of periods.

$$\mathrm{Jt} = \frac{1}{N-1}\sum_{i=1}^{N-1}|T_i - T_{i+1}| \qquad (15)$$

$$\mathrm{Sh} = \frac{\frac{1}{N-1}\sum_{i=1}^{N-1}|A_i - A_{i+1}|}{\frac{1}{N}\sum_{i=1}^{N} A_i}, \qquad (16)$$

3.2 Spectral Kurtosis

The spectral kurtosis (SK) shows transients and their spectral domain positions. It illustrates how arousal and emotion valence affect the speech spectrum's non-Gaussianity or flatness around its centroid. *Equation 17* estimates voice signal spectral kurtosis.

$$\mathrm{SK} = \frac{\sum_{k=b1}^{b2}(f_k - \mu_1)^4 s_k}{(\mu_2)^4 \sum_{k=b1}^{b2} s_k} \qquad (17)$$

Here, μ_1 and μ_2 symbolizes the spectral centroid and spectral spread, respectively, s_k is spectral value over k bins, and b_1 and b_2 are the inferior and higher limits of the bins where SK of speech is estimated.

3.3 DCNN Model

The proposed DCNN accepts an acoustic feature vector with dimensions of 318×1, which encompasses various features of Indian raga. The Convolution operation of DCNN provides the discriminant features of the input acoustic features. It involves sliding the convolution filter (fk) over the acoustic feature vector, one sample at a time. The convolution operation for feature vector $feat$, with N number of features, and the convolution kernel fk with dimensions of 1×3, is given by *equation 18*. To maintain the dimensions of the convolution feature map, the input feature vector is zero-padded with 2 zero values. The ReLU layer improves the non-linearity of the signal by eliminating negative values, as described in *equation 19*. The MaxPool layer reduces the feature dimensions and addresses the problem of over fitting, as mentioned in equation 20[11].

$$\boldsymbol{Rconv(k) = \sum_{n=1}^{N} feat(n).fk(n-k)} \qquad (18)$$

$$R\,Re\,L\,U(x,y) = max(\,0, Rconv(k)) \qquad (19)$$

$$RMP(i) = \max_{i=1:N-1,} R\,Re\,L\,U(i\!:i+1) \qquad (20)$$

The Softmax layer functions as the final classification layer in the DCNN. It takes the entire input vector and converts it into an output vector, where each value represents the probability of the input sample having a place with a particular class. The Softmax function ensures that the probabilities are normalized, meaning that the addition of all probabilities in the output vector is equal to one. By using the Softmax function (as described in *equation 21*), the DCNN can provide a probability distribution over the classes, allowing it to make confident predictions about the input samples class memberships.

$$\sigma(z)i = \frac{e^{zi}}{\sum_{j=1}^{k} e^{zj}} \qquad (21)$$

Softmax function is similar to sigmoid function only difference is that softmax consider vector as the input whereas sigmoid value considers scale value. The sigmoid function can be given by *equation 22.*

$$\sigma(z)i = \frac{1}{1+e^{-z}} \qquad (22)$$

After passing through the Softmax layer, the precise raga is recognized based on the record of the greatest Value in the neuron of the Softmax output layer. The index with the highest probability corresponds to the predicted raga class. Furthermore, the DCNN is trained using three different learning algorithms: ADAM, SGDM and RMSProp. These learning algorithms play a crucial role in adjusting the model's parameters in training process, enabling it to learn from this data and make accurate predictions. The DCNN configurations are provided in *table 1*. The training configuration and hyper-parameters of DCNN are provided in *table 2*.

Table 1: Configurations of DCNN

Layer	Number of Filters	Stride	Activation Maps	Padding
Input	-	-	1×318	-
Conv1	3×3×64	1	1×318×64	[1,1]
ReLU1	-	1	1×318×64	-
MaxPool1	-	2	1×159×64	-
Conv1	3×3×128	1	1×159×128	[1,1]
ReLU1	-	1	1×159×128	-
MaxPool1	-	2	1×79×128	-
Conv1	3×3×256	1	1×79×256	[1,1]
ReLU1	-	1	1×79×256	-
MaxPool1	-	2	1×39×256	-
FC Layer	-	-	1×8×9984	-
Softmax layer	-	-	1×8	-

Table 2: Training parameters of DCNN

Parameter	Specification
Optimization algorithm	ADAM, SGDM and RMSPROP
Learning rate	0.01
Epoch	200
System	CPU-20GB RAM
Training Samples	70%
Testing Samples	30%
Mini-batch size	32

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

The proposed scheme's results are evaluated using various quantitative and qualitative metrics, which include recall, precision, F1-score and accuracy. The equations of quantitative and qualitative metrics [11] are as follows:

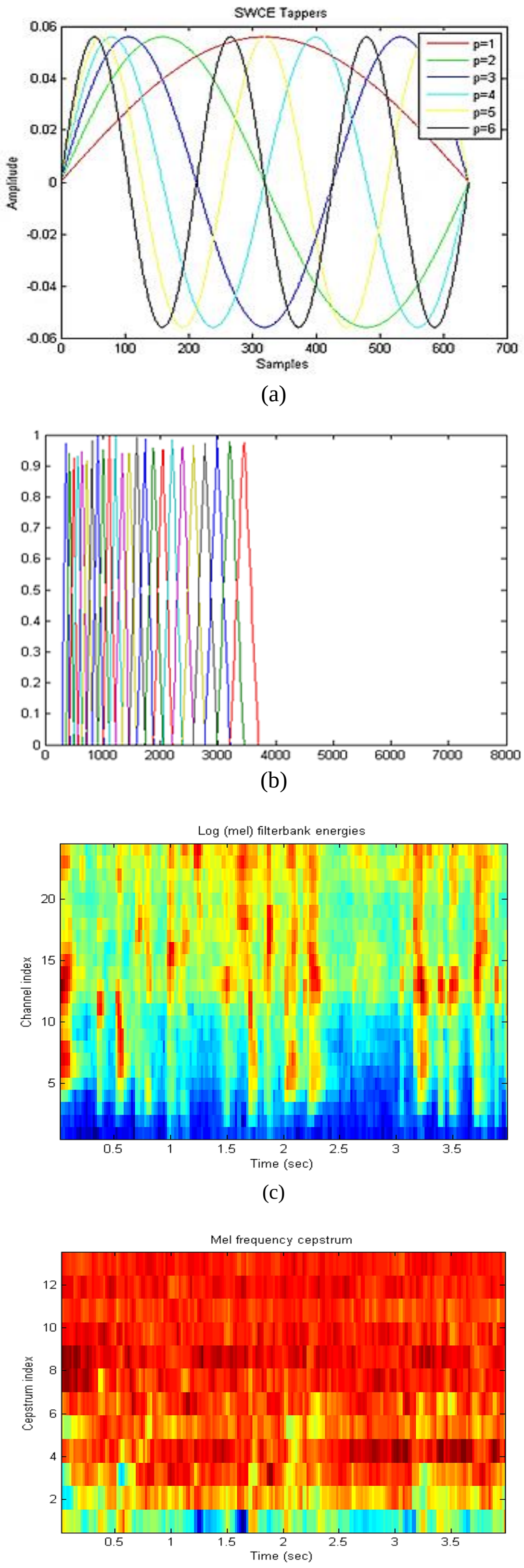
$$Pr\ e\ cision = \frac{TP}{TP+FP} \tag{23}$$

$$Re\ c\ all = \frac{TN}{TN+FN} \tag{24}$$

$$Accuracy(\%) = \frac{TP+TN}{TP+TN+FP+FN} X100 \tag{25}$$

$$F1 - Score = \frac{2*Pr\ ecision*Re\ call}{Pr\ ecision+Re\ call} \tag{26}$$

The MT-MFCC visualization stages shown in *figure 3. Figure 3(a)* provides the representation of the SWCE multi-taper windows required for the MT-MFCC. *Fig. 3(b)* illustrates the response of triangular filter bank and *fig 3(c and d)* depicts the Mel log spectrogram and MFCC cepstrum.



(a)

(b)

(c)

(d)

Figure 3. Experimental results (a) Multi-taper Windows (b) Mel clear out bank (c) Mel log filter bank electricity (d) Mel frequency Cepstrum

The training performance of DCNN is shown in *figure 4* (training accuracy) and *figure 5* (training loss).

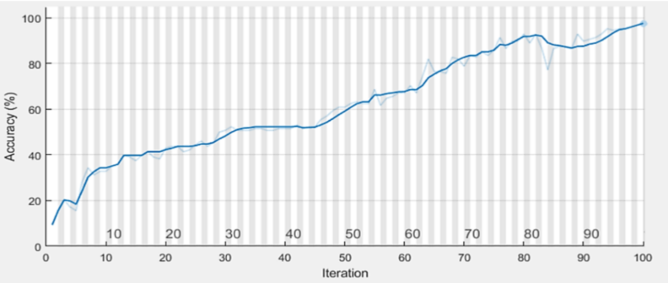


Figure 4. Training accuracy for proposed DCNN

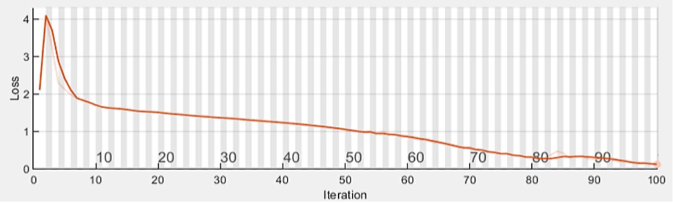


Figure 5. Training loss for the proposed DCNN

The DCNN demonstrates an overall accuracy of 89.38%, outperforming the SGDM (78.68%) and RMSProp (61.88%) learning algorithms. The ADAM optimization algorithm combines the strengths of SGDM (good performance for sparse gradient problems in natural language processing) and RMSProp (ability to work better for noisy and non-stationary signals). The DCNN with ADAM optimization shows significant improvements in accuracy, achieving 2.17% and 29.91% higher performance compared to DCNN with SGDM and RMSProp, respectively, for the eight-class raga classification. Furthermore, the DCNN with ADAM achieves the highest accuracy of 100% for Asawari, Bhairavi, Malkans and Yaman ragas. However, it achieves the lowest accuracy of 41.41% for the Bageshwari raga.

Table 3: Performance of proposed DCNN for raga classification

Raga	ADAM				SGDM				RMSProp			
	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score	Accuracy	Precision	Recall	F1-Score	Accuracy
Asawari	0.73	1.00	0.84	100.00	1.00	0.75	0.86	75.00	1.00	0.69	0.82	69.23
Bageshwari	1.00	0.71	0.83	71.43	0.56	1.00	0.71	100.00	0.40	0.33	0.36	33.33
Bhairavi	1.00	1.00	1.00	100.00	1.00	0.50	0.67	50.00	0.67	1.00	0.80	100.00
Bhoopali	0.75	0.86	0.80	85.71	0.65	0.92	0.76	91.67	0.63	0.63	0.63	62.50
Darbari	1.00	0.85	0.92	84.62	0.64	0.78	0.70	77.78	0.25	0.17	0.20	16.67
Malkans	0.83	1.00	0.91	100.00	1.00	0.60	0.75	60.00	0.57	0.80	0.67	80.00
Sarang	0.85	0.73	0.79	73.33	0.75	0.75	0.75	75.00	0.75	0.33	0.46	33.33
Yaman	1.00	1.00	1.00	100.00	1.00	1.00	1.00	100.00	0.77	1.00	0.87	100.00
Average	0.89	0.89	0.89	89.39	0.82	0.79	0.77	78.68	0.63	0.62	0.60	61.88

The Precision, Recall, F1-score and Accuracy of DCNN for different learning algorithms are visualized in *fig. 6-9* respectively. The DCNN-ADAM provides good balance between qualitative (precision) and quantitative (recall) of the raga classification. It provides overall F1-score of (0.89) which is superior over DCNN-SGDM (0.77) and DCNN- RMSProp (0.60).

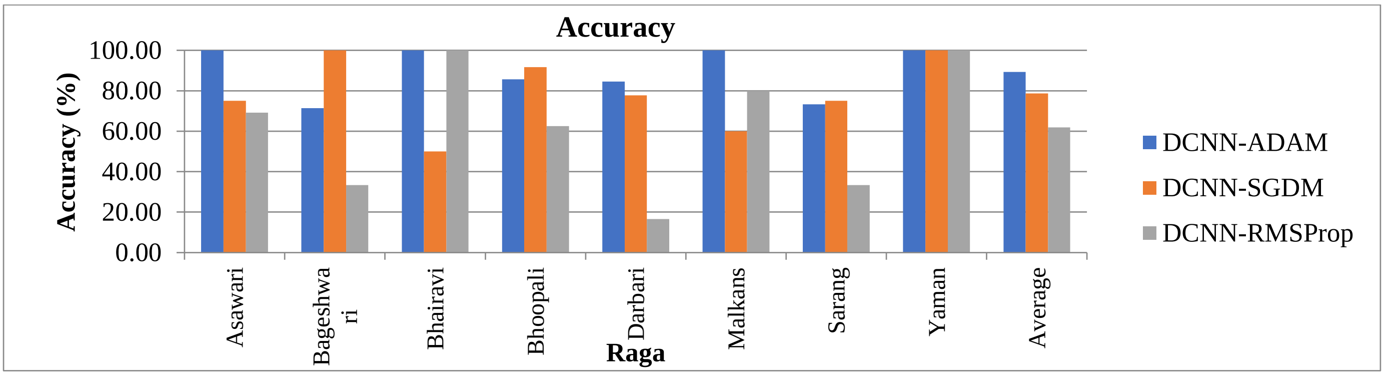


Figure 6. Accuracy of DCNN for raga classification

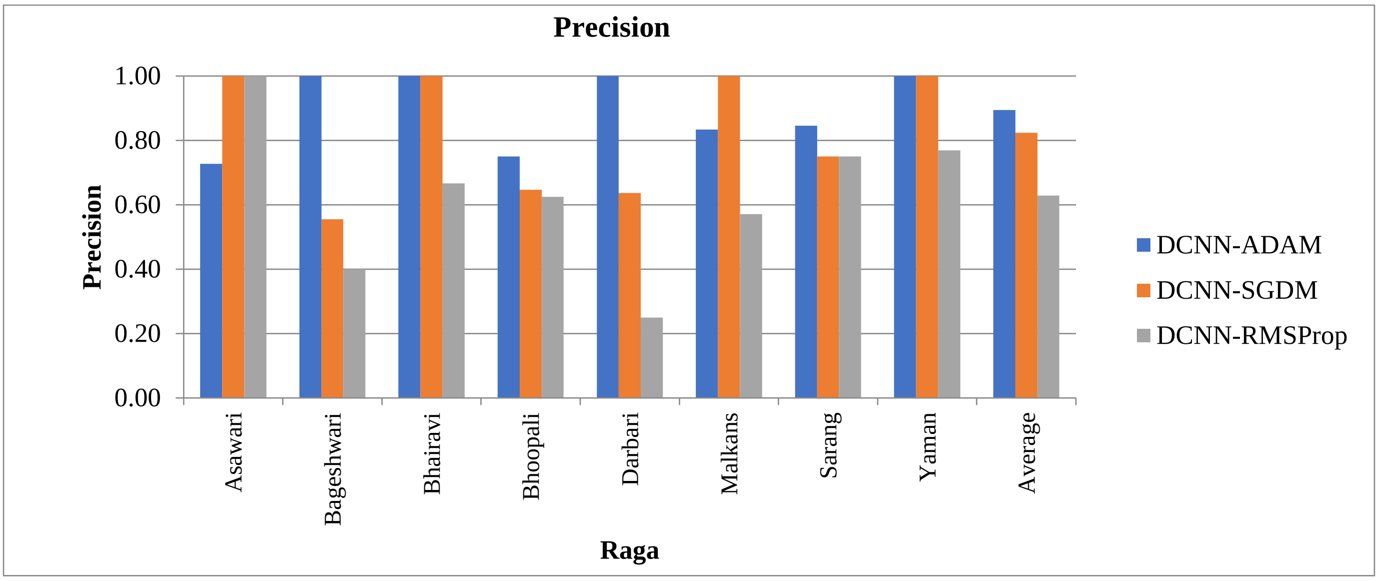


Figure 7. Precision of DCNN for raga classification

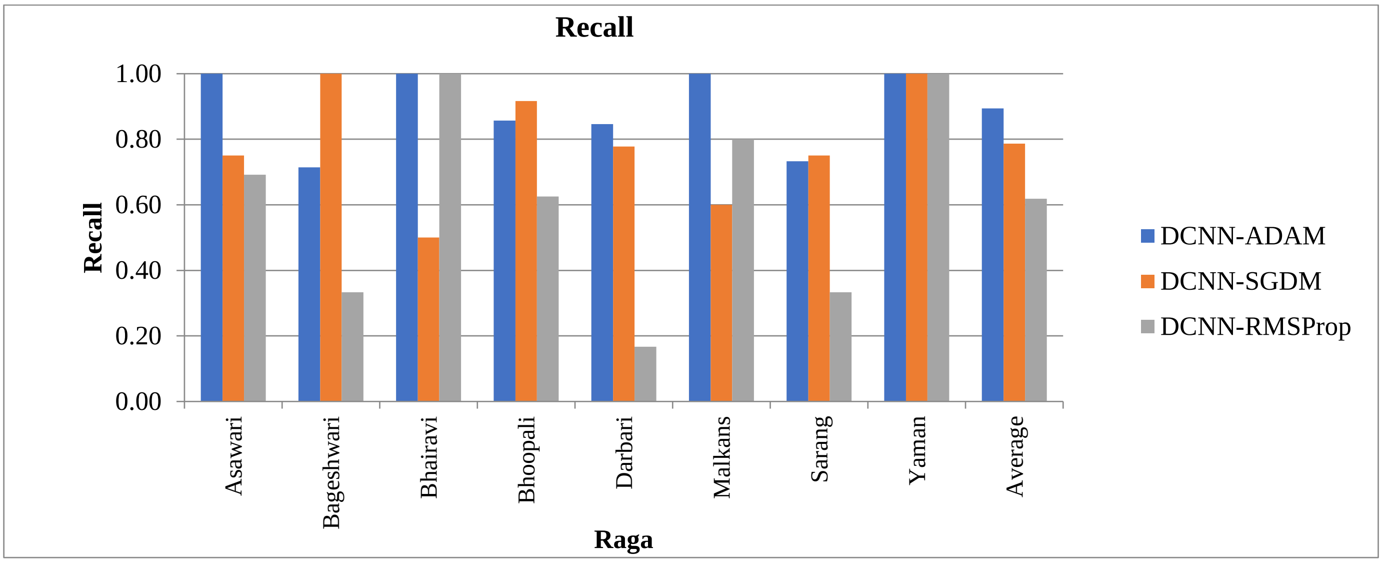


Figure. 8. Recall of DCNN for raga classification

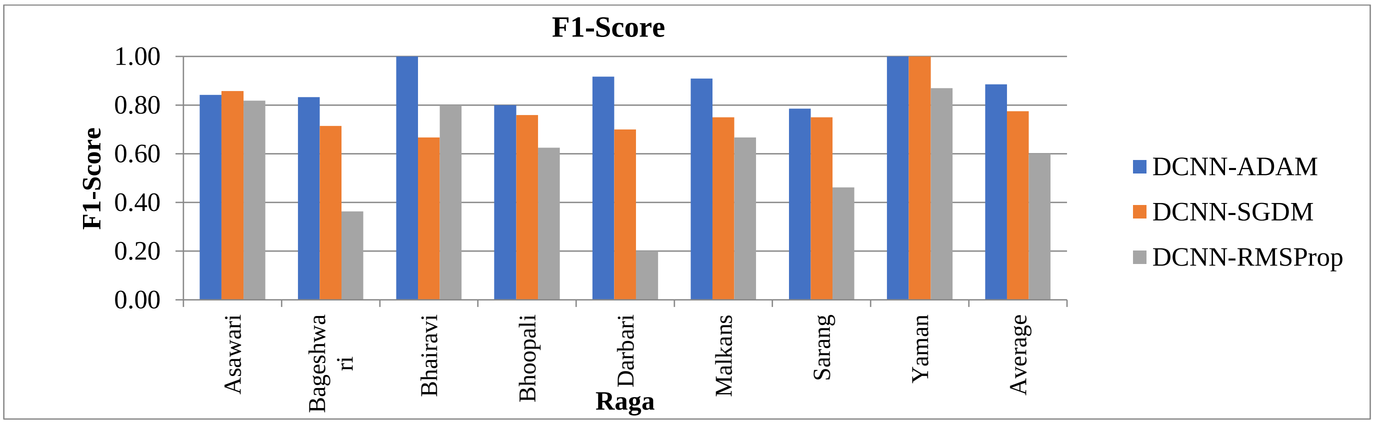
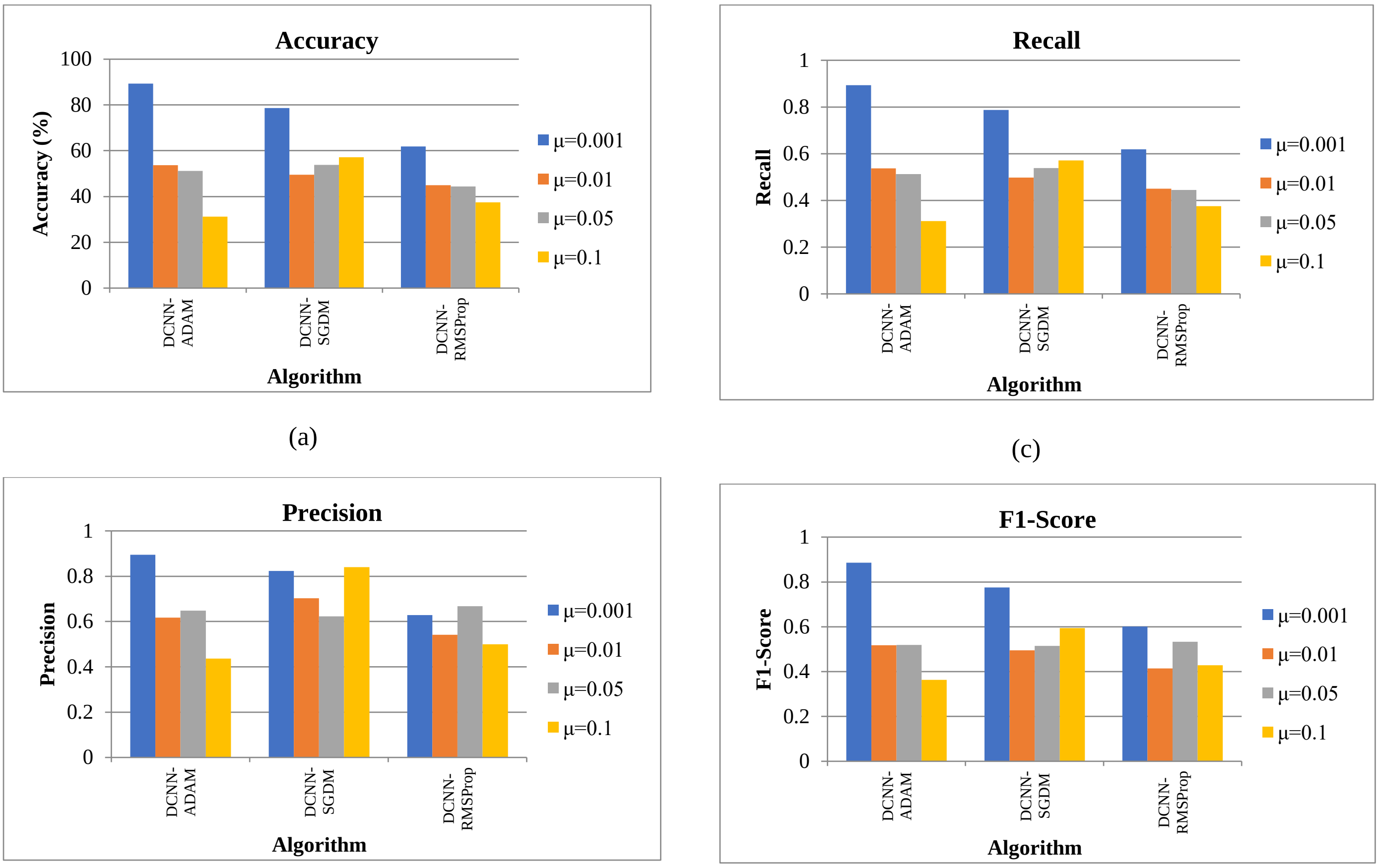


Figure. 9. F1-score of DCNN for raga classification

Further, the results of proposed DCNN for raga classification are evaluated for the different initial learning art of ADAM, SGDM and RMSProp algorithm. The results of the proposed algorithms are observed to be superior to the learning rate of 0.001 compared with 0.01, 0.05, and 0.1. The smaller value of the initial learning rate provides better testing results and minimizes the error in training effectively as given in *figure 10*.



(a)

(c)

(b)

(d)

Figure 10. Performance of proposed DCNN for different learning rate

Table 4: Performance of proposed DCNN for raga classification

Author and Year	Feature Extraction	Classifier	Accuracy (%)	Ragas
Joshi et al. (2021) [39]	Power spectrogram, Rolloff, MFCC, Spectral bandwidth, ZCR	KNN (K=1)	98.00	2 ragas (Yaman and Bhairavi)
Gaurav Pandey et al. (2003) [40]	-	HMM-String Matching	77.00	Kalyani,Yaman and Bhupali
Choi, K., Fazekas, G. (2016).[12]	Mel Spectrogram feature set, Accuracy	CNN	84.00	4 ragas
Elbir, A. and Aydin, N. (2020).[15]	Prcision, Recall, F1-Score	KNN	84.00	Different Music type
Proposed Approach	Proposed Feature set	KNN	78.00	8 ragas
		SVM	84.20	8 ragas
		RF	84.50	8 ragas
		DCNN	89.38	8 ragas

The output of the proposed system is compared with the conventional state of arts used for Indian classical music recognition as given in *table 4*. It provided 89.34% accuracy for the 8 ragas for suggested lightweight DCNN architecture, which is superior over the RF (84.50%), SVM (84.20%), and KNN (78%), respectively. The previous methods have considered only 2 or 4 ragas for the experimental evaluations. Proposed system's performance compared with the conventional machine learning classifiers to analyze the effect of set of features on classifiers such as Support Vector machine (SVM), K-Nearest

Neighbor (SVM) and Random Forest (RF). It provides superior accuracy for 8 class classification of ragas.

4. CONCLUSION AND FUTURE SCOPES

This paper aims to show an Indian raga classification system that utilizes various acoustic features, including spectral, time domain and voice quality features, along with a (DCNN). The use of the proposed DCNN-ADAM results in an overall accuracy of 89.38% for classifying eight ragas, namely Asawari, Bageshwari, Bhairavi, Bhoopali, Darbari, Malkans, Sarang and Yaman. The lightweight DCNN architecture helps capture various intonation changes, pitch variations, and discriminative attributes in classical music, contributing to the improved distinctiveness of low-level acoustic features. In the future, the proposed model's output can be generalized more for multiple Indian languages. In the future, the issue of data scarcity can be tackled using the data augmentation technique.

REFERENCES

[1] R. Sridhar and T. V. Geetha, ‘‘Swara indentification for south indian classical music,’’ in Proc. 9th Int. Conf. Inf. Technol. (ICIT), Dec. 2006, pp. 143–144.

[2] R. Sridhar and T. V. Geetha, ‘‘Music information retrieval of carnatic songs based on carnatic music singer identification,’’ in Proc. Int. Conf. Comput. Electr. Eng., Dec. 2008, pp. 407–411.

[3] G. Pandey, C. Mishra, and P. Ipe, ‘‘TANSEN: A system for automatic raga identification,’’ IICAI, Dec. 2003, pp. 1350–1363.

[4] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler,‘‘A tutorial on onset detection in music signals,’’ IEEE Trans. Speech Audio Process,vol.13, no. 5, pp. 1035–1047, Sep. 2005.

[5] A. Klapuri and M. Davy, Signal Processing Methods for Music Transcription. New York, NJ, USA: Springer-Verlag, 2006.

[6] P. Chordia, ‘‘Automatic raag classification of pitch-tracked performances using pitch-class and pitch- class dyad distributions,’’ in Proc. ICMC, 2006, pp. 1–7.

[7] G. E. Poliner, D. P. W. Ellis, A. F. Ehmann, E. Gomez, S. Streich, and B. Ong, ‘‘Melody transcription from music audio: Approaches and evaluation,’’ IEEE Trans. Audio, Speech Lang. Process., vol. 15, no. 4, pp. 1247–1256, May 2007.

[8] S. Samsekai Manjabhat, S. G. Koolagudi, K. S. Rao, and P. B. Ramteke, ‘‘Raga and tonic identification in carnatic music,’’ J. New Music Res., vol. 46, no. 3, pp. 229–245, Jul. 2017.

[9] Theory of Indian Music, Pankaj, New Delhi, India, 1999.

[10] S. Shetty and S. Hegde, ‘‘Automatic classification of carnatic music instruments using MFCC and LPC,’’ in Data Management, Analytics and Innovation. Singapore: Springer, 2020, pp. 463-474.

[11] Joshi Dipti, Jyoti Pareek, and Pushkar Ambatkar. "Indian Classical Raga Identification using Machine Learning." In ISIC'21: International Semantic Intelligence Conference, February 25-27, 2021, New Delhi, India, pp. 259-263. 2021

[12] Choi, K., Fazekas, G. and Sandler, M. (2016). Automatic tagging using deep convolutional neural networks.

[13] Abdul, A., Chen, J., Liao, H.-Y. and Chang, S.-H. (2018). An emotion-aware personalized music recommendation system using a convolutional neural networks approach, Applied Sciences 8: 1103.

[14] Chang, S., Abdul, A., Chen, J. and Liao, H. (2018). A personalized music recommendation system using convolutional neural networks approach, 2018 IEEE International Conference on Applied System Invention (ICASI), pp. 47–49.

[15] Elbir, A. and Aydin, N. (2020). Music genre classification and music recommendation by using deep learning, Electronics Letters 56(12): 627–629.

[16] Jiang, M., Yang, Z. and Zhao, C. (2017). What to play next? arnn-based music recommendation system, 2017 51st Asilomar Conference on Signals, Systems, and Computers, pp. 356–358.

[17] Tao, Y., Zhang, Y. and Bian, K. (2019). Attentive context-aware music recommendation, 2019 IEEE Fourth International Conference on Data Science in Cyberspace (DSC), pp. 54–61.

[18] Fulzele, P., Singh, R., Kaushik, N. and Pandey, K. (2018). A hybrid model for music genre classification using lstm and svm, 2018 Eleventh International Conference on Contemporary Computing (IC3), pp1-3.

[19] Adiyansjah, Alexander, G. and Derwin, S. (2019). Music recommender system based on genre using convolutional recurrent neural networks, Procedia Computer Science 157: 99–109.

[20] Irene, R. T., Borrelli, C., Zanoni, M., Buccoli, M. and Sarti, A. (2019). Automatic playlist generation using convolutional neural networks and recurrent neural networks, 2019 27th European Signal Processing Conference (EUSIPCO), pp. 1–5.

[21] Kim, H., Kim, G. Y. and Kim, J. Y. (2019). Music recommendation system using human activity recognition from accelerometer data, IEEE Transactions on Consumer Electronics 65(3): 349–358.

[22] Prabhat Chandra Shrivastava, Prashant Kumar, Manish Tiwari, Amit Dhawan, “Efficient Architecture for the Realization of 2-D Adaptive FIR Filter Using Distributed Arithmetic. Circuits Syst Signal Process, Issue Date March 2021, Volume 40, pp 1458–1478 https://doi.org/10.1007/s00034-020- 01539-y, (SCI, Impact Factor-2.25).

[23] Prashant Kumar, Prabhat Chandra Shrivastava, Manish Tiwari and Ganga Ram Mishra, “High- Throughput, Area-Efficient Architecture of 2-D Block FIR Filter Using Distributed Arithmetic Algorithm” Circuits System & Signal Processing, Springer., Issue Date-March 2019, Volume 38, Issue 3, pp 1099–1113, https://doi.org/10.1007/s00034-018-0897-2, (SCI, Impact Factor-2.25).

[24] Prashant Kumar, Prabhat Chandra Shrivastava, Manish Tiwari, and Amit Dhawan. "ASIC Implementation of Area-Efficient, High-Throughput 2-D IIR Filter Using Distributed Arithmetic", Circuits System & Signal Processing, Springer., Issue Date-July 2018, Volume 37, Issue 7, pp 2934– 2957, https://doi.org/10.1007/s00034-017-0698-z (SCI, Impact Factor-2.25).

[25] Prabhat Chandra Shrivastava, Prashant Kumar, Manish Tiwari, “Hardware Realization of 2-D General Model State Space Systems”, International Journal of Engineering and Technology (IJET), ISSN (Online): 0975-4024, Vol 9 No, Pages: 3996-4005, 5 Oct-Nov 2017, DOI: 10.21817/ijet/2017/v9i5/170905301 (Scopus Index Impact Factor-1.998).

[26] Alam, Md Jahangir, Tomi Kinnunen, Patrick Kenny, Pierre Ouellet, and Douglas O’Shaughnessy. "Multitaper MFCC and PLP features for speaker verification using i-vectors." Speech communication 55, no. 2 (2013): 237-251.

[27] Mansouri, Arash, and Eduardo Castillo-Guerra. "Multitaper MFCC and normalized multitaper phase- based features for speaker verification." SN Applied Sciences 1, no. 4 (2019): 1-18.

[28] Chowdhury, Anurag, and Arun Ross. "Fusing MFCC and LPC features using 1D triplet CNN for speaker recognition in severely degraded audio signals." IEEE transactions on information forensics and security 15 (2019): 1616-1629.

[29] Chauhan, Neha, Tsuyoshi Isshiki, and Dongju Li. "Speaker recognition using LPC, MFCC, ZCR features with ANN and SVM classifier for large input database." In 2019 IEEE 4th International Conference on Computer and Communication Systems (ICCCS), pp. 130-133. IEEE, 2019.

[30] Welling, Lutz, and Hermann Ney. "Formant estimation for speech recognition." IEEE Transactions on Speech and Audio Processing 6, no. 1 (1998): 36-48.

[31] Zhang, Yang, Tess Koerner, Sharon Miller, Zach Grice-Patil, Adam Svec, David Akbari, Liz Tusler, and Edward Carney. "Neural coding of formant-exaggerated speech in the infant brain." Developmental science 14, no. 3 (2011): 566-581.

[32] Levin, Herman, and William Lord. "Speech pitch frequency as an emotional state indicator." IEEE Transactions on Systems, Man, and Cybernetics 2 (1975): 259-273.

[33] Savchenko, A. V., and V. V. Savchenko. "A method for measuring the pitch frequency of speech signals for the systems of acoustic speech analysis." Measurement Techniques 62, no. 3 (2019): 282-288.

[34] Ghosal, Arijit, Rudrasis Chakraborty, Ractim Chakraborty, Swagata Haty, Bibhas Chandra Dhara, and Sanjoy Kumar Saha. "Speech/music classification using occurrence pattern of zcr and ste." In 2009 Third International Symposium on Intelligent Information Technology Application, vol. 3, pp. 435-438. IEEE, 2009.

[35] Banchhor, Sumit Kumar, and Arif Khan. "Musical instrument recognition using zero crossing rate and short-time energy." Musical Instrument 1, no. 3 (2012): 1-4.

[36] Farrús, Mireia, and Javier Hernando. "Using jitter and shimmer in speaker verification." IET Signal Processing 3, no. 4 (2009): 247-257.

[37] Becker, Alyssa S., and Peter J. Watson. "The Use of Vibrato in Belt and Legit Styles of Singing in Professional Female Musical-Theater Performers." Journal of Voice (2022).

[38] “Multilingual Indian Musical Type Classification” Mrs. Swati P. Aswale, Prabhat Chandra Shrivastava, Dr. Roshani Bhagat, Vikrant B. Joshi, Mrs. Seema M. Shende, conference paper 5th International Conference on VLSI, Communication and Signal Processing (Via Online mode), Volume, Year 2022.

[39] Joshi, Dipti, Jyoti Pareek, and Pushkar Ambatkar. "Indian Classical Raga Identification using Machine Learning." In ISIC, pp. 259-263. 2021.