International Journal of Engineering Trends and Technology
ISSN: 2231–5381 / https://doi.org/10.14445/22315381/IJETT-V73I2P113
Volume 73 Issue 2, 155-165, February 2025

Original Article

Raga Net: A Novel Deep Learning Framework for Indian Raga Recognition based on Deep Convolution Neural Network and Long Short-term Memory

Swati Pravin Aswale[1], Wani Patil[2], Shrikant Chavate[3]

[1,2,3]*Electronics and Telecommunication Department, G.H. Raisoni University, Maharashtra, India.*
[1]*D. Y. Patil College of Engineering, Maharashtra, India.*

[1]*Corresponding Author : swatiaswale31@gmail.com*

Received: 06 November 2024 Revised: 23 January 2025 Accepted: 03 February 2025 Published: 21 February 2025

Abstract - *Indian Raga classification is challenging because of the vast disparity in the music's swaras, pitch, melody, style and intonation. This paper presents the Indian Raga classification using a Deep Convolution Neural Network and Multiple Acoustic Features. The multiple acoustic features include particular spectral, temporal, and voice quality attributes of the musical voice that describe the uniqueness of different ragas. A Novel hybrid Archimedes Optimization Algorithm in light of Multi-Attribute Utility Theory (AoA-MAUT) is utilized to choose salient and distinctive features from Multi Acoustic Features. Further, it uses the novel RagaNet, which combines Parallel Deep Convolution Neural Network (PDCNN) to depict the spectral properties of musical features and Long Short-Term Memory (LSTM) for providing temporal and long-term dependencies of the musical features. The proposed AoA-MAUT-RagaNet-based Indian classical raga recognition results in a general accuracy of 91.71%, a precision of 0.93, a recall of 0.90, and an F1-score of 0.91, which is superior compared to the traditional state of arts.*

Keywords - *Raga identification, Archimedes Optimization Algorithm, Multi-Attribute utility theory, Deep Convolution Neural Network, Multiple acoustic features.*

1. Introduction

Raga is an imperative aspect of Indian classical music-composition and rendition of Indian classical music hugely dependent on these ragas. Ragas is often considered the grammar of Indian classical music. Indian old-style music is broadly categorized into Carnatic and Hindustani classical music [1-2]. Swaras are the critical constituents of ragas. All ragas have some ‘Swara’, but different oscillations and intonation variations create different moods. In Indian classical music, the swarms are relative to each other and do not hold the same pitch, unlike in Western music. A music render can set his pitch for swara 'sa', and all remaining swaras are relative to it. There are seven swaras: ‘sa’, ‘re’, ‘ga’, ‘ma’, ‘pa’, ‘dha’ and ‘ni’. The most noticeable swaras are called vadi, and the second most prominent swara is called samvadi [3-4]. The ascend in swaras pitch is called ‘ohana’, whereas descending in swarms pitch is known as ‘avrohana’. Various raga identification techniques have been presented, including pitch tracking, rule-based, and model-based methods. However, raga recognition is very challenging because of the vast disparity in tempos, pitch, ohana, tonic pitch of render and style of render performance. Raga recognition is crucial in music retrieval, recommendation systems and composing, practicing and learning ragas. It further boosts the music experience of music appreciators [5-6]. In traditional raga recognition, the ragas are identified during musical performance and the teaching and learning process. In the teaching-learning process of Indian Established music, the teacher is called a 'Guru', and the learner is known as ‘Shishya’. The listener of Indian classical music often recognizes ragas based on ‘pakad’, which describes the musical phrases or set of phrases. However, it isn't straightforward for beginners to identify the ragas based on swaras from music. Swaras are essential features of the raga and provide the music's melodic details. Swara features fail to give the music signal's temporal characteristics [7-8]. Many musical recognition systems have been presented in the decade, focusing on pop, jazz, and modern music. However, the Indian classical musical ragas greatly impact the musical industry, which has been explored less [9]. The computerized automation of the raga recognition system will help in musical schools, song recording, artificial music generation, and finding the uniqueness of the raga styles. The huge variety in the Indian classical musical gharanas, rendition style, and vast ragas and disparities in the vocals lead to the need for automatic raga recognition. The proposed system aims to recognize the raga to improve the generalization capability of the system, which makes the system robust in terms of the variance in singer style, gender, gharana, and regional and religious influence on voice.

The main commitments of this article are summed up as follows:

- Raga voice representation uses Multiple Acoustic Features that depict spectral, temporal and voice quality features of voice
- Feature selection using AoA-MAUT to choose the noticeable attributes of MAFs to limit the computational complexity of the DCNN structure.
- Novel RAGANet based on parallel DCNN-LSTM-based hierarchical feature representation of the music voice highlights to improve the uniqueness of Indian ragas.

The outstanding article is organized as follows: Section 2 describes the connected work on music and raga recognition using ML and DL. Section 3 provides a brief methodology encompassing feature representation, feature selection and range classification. Further, section 4 explains the exploratory outcomes and conversations. Ultimately, section 5 provides the conclusion and future extent of the work.

2. Related Work

Different techniques have been presented for music recognition, focusing on the Indian and Western musical genres of different languages. Choi et al. [11] suggested full CNN (FCNN) for music recognition to boost the feature representation. It resulted in an accuracy of 89% for real-time datasets and showed compatibility with larger datasets. Abdul et al. [12] presented DCNN to further develop a hierarchical feature representation of the musical voice. It resulted in an accuracy of 95.35% for the Million Song Dataset (MSD). Further, Chang et al. [13] used MFCC spectrograms to upgrade the feature description capability of the DCNN, which gave 90 % accuracy for the MSD dataset.

It has shown sensitivity to the disparity in geographic location, emotions, and ambience factors. Jiang et al. [14] suggested that LSTM enhances the long-term dependency of musical voice, which has 86.4% accuracy for raw speech. Sainath et al. [15] presented a combination of CNN, LSTM and DNN as (CLDNN) to improve the local representation at distinct levels and enhance long-term feature representation. The CLDCNN enhanced the feature distinctiveness and yielded a word error rate of 82.6 % for the larger in-house dataset. Al Mamum et al. [16] presented a basic scheme for Bangla music recognition. It has shown superior results compared with the conventional ML classifies but results in lower inter-class and intra-class disparity. Irene et al. suggested that RNN-LSTM be used to get to the next level of temporal feature depiction and long-term dependency on voice. It assists with limiting the vanishing gradient issue. It has given a root mean square error of 0.0638 for the Art of Mix-2011 dataset. Further, Adiyansjah et al. [17] provided convolution RNN (CRNN) to enhance traditional CNN's spectral and temporal representation. It has given an F1-score of 0.74 for the free Music Archive dataset. However, the method's effectiveness is limited due to disparity in the musical voice's intonation, style, timbre and prosody. Fulzele et al. [18] explored LSTM+SVM for music recognition that combines the long-term dependency of the LSTM and the discrimination power of SVM. It has given 89% accuracy for the million song dataset (MSD), superior to individual SVM and LSTM. Tao et al. [19] presented adjusted LSTM to grab the voice's distinguishing features and temporal context. It resulted in a recall rate of 0.778 for the real-time dataset. It fails to depict the spectral and voice quality properties of the music. Karunakaran and Arya [20] investigated the scalable hybrid classifier to minimize blurry classification of the SVM, KNN, naïve Bayes and neural network. It gave 90% accuracy for the GTZAN Genre collection (1000) dataset. However, the hybrid combination results in higher computational intricacy of the system.

Frank Zalkow et al. [21] proposed a deep Croma model for musical voice recognition that uses chroma features of the voice. It resulted in an accuracy of 87.10% for the musical theme recognition. It has given better results for monophonic music but provides poor results for polyphonic music. Further, Ng Wing et al. [22] suggested CNN and NetVLAD for the musical genre classification. It uses multi-level features and provides long-term dependencies of the features. It provided an accuracy of 92.46% for ISMIR-2004, 96.50% for GTZAN and 95.50% for the extended ballroom dataset, respectively.

The proposed method needs extensive hyperparameter tuning and fails to grab local connectivity at different levels. Elbir and Aydin [23] presented a deep neural network that used an artificial dropout layer to minimize the validation error. It comes about with an accuracy of 97.6% for the GTZAN dataset. From the extensive survey, it is identified that the previous music recognition systems have majorly focused on Western music in English languages. Very little concentration is given to Indian classical music recognition. The earlier music recognition systems are subjected to the timbre, intonation, singer style, language and variations in the music pitch [24-29]. The DL-based schemes have shown imperative enhancement over the ML-based music recognition systems. However, the DL-based music recognition schemes are challenging because of network complexity, extensive hyper-parameter tuning, larger trainable parameters and the computational intricacy of the systems.

3. Methodology

The proposed raga identification flow diagram is shown in Figure 1. It encompasses raga preprocessing, feature extraction, feature selection, feature representation and classification using the proposed RAGANet.

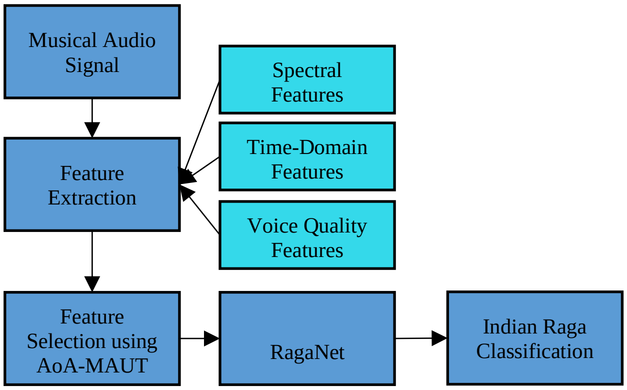


Fig. 1 Flow diagram of proposed raga classification scheme

In the feature extraction stage, various spectral, temporal and voice quality features are extracted to describe the uniqueness of the ragas. The AoA algorithm selects the prominent feature to minimize the redundant and essential features. The MAUT algorithm is utilized to decide the weights of the AoA algorithm. Further, DCNN is utilized for the feature representation and classification of Indian ragas.

3.1. Multiple Acoustic Features

The MAF includes Spectral Domain Features (SDF), Time Domain Features (TDF) and Voice Quality Features (VQF). The spectral features consist of MFCC, MFCC Δ, MFCC ΔΔ, LPCC, Spectral Kurtosis (SK), Formats (FM), standard deviation of FMs and variance of FMs. The MFCC provides the energy features of the musical voice in the spectral domain [30-35]. The first- and second-order derivatives (MFCC Δ and MFCC ΔΔ) show changes in the pitch of the ragas. LPCC provides the emotional attributes of the ragas, which represent the emotional specific features of the raga rendition. Spectral kurtosis provides the transient sequence and its location in the frequency domain.

Table 1. MAFs Details

Features Types	Feature Name	Exact Features	Total Features
SDF	MFCC	13	314
	MFCC Δ	13	
	MFCC ΔΔ	13	
	LPCC	13	
	SK	257	
	FM	3	
	Standard deviation of FMs	1	
	Mean of FMs	1	
TDF	ZCR	1	2
	PF	1	
VQF	Jitter	1	2
	Shimmer	1	
Total Features			318

Formants are frequencies over the maximum power of the raga signal is accumulated. The ZCR and Pitch Frequency (PF) depict the raga music's noisiness measure and fundamental frequency, respectively. The jitter and shimmer offer variations in time and frequency due to raga. The combination of MAFs helps to combine the various characteristics of the ragas together to construct distinguishing features [36-38].

4. Aoa-Basex Feature Selection

In this part, the consequences of exploration are made sense of and simultaneously are given a thorough conversation. Results can be introduced in figures, diagrams, tables, and other things that allow the reader to see them without any problem. The conversation can be made in a few sub-sections. It is unequivocally recommended that correlation with results from other distributed articles be given to give additional background information and to reinforce the case of curiosity. The AOA is utilized for effective feature selection to enhance the distinguishing capability of raga features. The random weights of the fitness function lead to the disparity in fitness criteria and often result in poorly optimized solutions. The MAUT algorithm is employed to choose fitness function weights based on correlation, entropy and ratio of inter to intra-class variance of the raga features. The AOA is based on the physics Archimedes principle, which states, "When the object is lowered into liquid, then the buoyant force applied on the item by the fluid is straightforwardly relative to the object's mass. If the object's mass is larger than the applied force, then the object sinks into the fluid. If the applied force and mass of the object are the same, then the object floats on the surface of the fluid. The buoyant force is affected by the object's mass, density, volume and acceleration [39-42]. The critical goal of AOA is to achieve the equilibrium condition. The possible combination of the features is considered one object. The AOA principle is described by Equations 1-5.

$$F_b = M_o \tag{1}$$

$$p_b v_b a_b = p_o v_o a_o \tag{2}$$

$$a_o = \frac{p_b v_b a_b}{p_o v_o} \tag{3}$$

$$M_b - M_r = M_o \tag{4}$$

$$p_b v_b a_b - p_r v_r a_r = p_o v_o a_o \tag{5}$$

Here, pb and po are densities of fluid and objects, Fb stands for buoyant force, Mo, Mb and Mr are the weights (mass) of object, fluid and neighboring object, vb and vo are volumes of fluid and objects and ab and ao are accelerations of fluid and objects, respectively Assuming other power is applied on the item because of the impact of neighborhood object(r) then the balance state will be characterized by equation 4 and 5.

4.1. AOA Algorithm

Step 1: Initialization phase

The object represents the possible set of features, which is initialized randomly by considering the lower and upper bounds using Equation 6.

$$O_i = lb_i + rand \times (ub_i - lb_i); i = 1,2, \dots, N \quad (6)$$

$$den_i = rand \quad (7)$$

$$vol_i = rand \quad (8)$$

$$acc_i = lb_i + rand \times (ub_i - lb_i) \quad (9)$$

Here, Oi is with individual items in the population of N items. UB and LB are the upper and lower limits of the exploration space. The densities, volume, and acceleration are arbitrarily initialized using Equations 7, 8, and 9.

Step 2: Evaluate the fitness of each item and find the best item with the optimal fitness value using equation 10.

$$Fitness = w_1 * CR + w_2 * EN + w_3 * RIN \quad (10)$$

Here, CR denotes the correlation of features, EN depicts the entropy of features, and RIN provides the ratio of inter-class to intra-class changeability of the features. The weights $w1$, $w2$ *and* $w3$ suggest the decision weightage of each variable for the feature selection objective function such that $w1 + w2 + w3 = 1$.

The weights $w1$, $w2$ *and* $w3$ are chosen in light of the AHP algorithm and are used as decision criteria. Assign $denbest$, $volbest$ and accbest based on best fitness value.

Step 3: Change the den and vol of the ith objects for tth iteration with the help of Equations 11 and 12, respectively.

$$den_i^{t+1} = den_i^t + rand \times (den_{best} - den_i^t) \quad (11)$$

$$vol_i^{t+1} = vol_i^t + rand \times (vol_{best} - vol_i^t) \quad (12)$$

Here, $denbest$, $volbest$ and rand signify the density, volume of the best item and evenly spread random number, respectively.

Step 4: Calculate the Density decreasing factor (d) and Transfer operator (TF). The TF operator helps convert search space from exploration to exploitation, which surges gradually with time. When two objects collide, they try to attain the equilibrium state. In AOA, the TF represents this equilibrium condition as given in Equation 13.

$$T\,F = \exp\left(\frac{t - t_{max}}{t_{max}}\right) \quad (13)$$

Here, t and $tmax$ indicate present and maximum iterations. The value of d gradually declines with time to accomplish the balance in global and local search, as described in equation 14. The appropriate tuning of TF and d significantly impacts harmonizing exploration and exploitation.

$$d^{t+1} = exp\left(\frac{t_{max} - t}{t_{max}}\right) - \left(\frac{t}{t_{max}}\right) \quad (14)$$

The exploration stage (TF<=0.5) considers the collision of two objects and updates the acc is improved by Equation 15.

$$acc_i^{t+1} = \frac{den_{mr} + vol_{mr} \times acc_{mr}}{den_i^{t+1} \times vol_i^{t+1}} \quad (15)$$

The exploitation stage (TF>0.5) assumes no collision of objects, and acc is revised by using the previous best acc, vel and den, as given in Equation 16.

$$acc_i^{t+1} = \frac{den_{best} + vol_{best} \times acc_{best}}{den_i^{t+1} \times vol_i^{t+1}} \quad (16)$$

Equation 17 is used to normalize the usefulness of acc, where the lower bound (lb) and upper bound (u) for normalization are chosen to be 0.1 and 0.9, respectively.

$$acc_{i-norm}^{t+1} = u \times \frac{acc_i^{t+1} - \min(acc)}{\max(acc) - \min(acc)} + l \quad (17)$$

The object's positions are revised using equations 18 and 19, considering exploration and exploitation. Here, C1 and C2 are chosen as 2 and 6. The value of T continuously surges and is estimated by

$$T = C3 * TF$$

$$O_i^{t+1} = O_i^t + C_1 \times rand \times acc_{i-norm}^{t+1} \times d \times (O_{rand} - O_i^t) \quad (18)$$

$$O_i^{t+1} = O_{best}^t + F \times C_2 \times rand \times acc_{i-norm}^{t+1} \times d \times (T \times O_{best} - O_i^t) \quad (19)$$

Equations 20 and 21 characterize the objects' direction deviation, where C4 is an arbitrary variable between 0 and 1.

$$F = \begin{cases} +1 \; if \; P \le 0.5 \\ -1 \; if \; P > 0.5 \end{cases} \quad (20)$$

$$P = 2 * rand - C4 \quad (21)$$

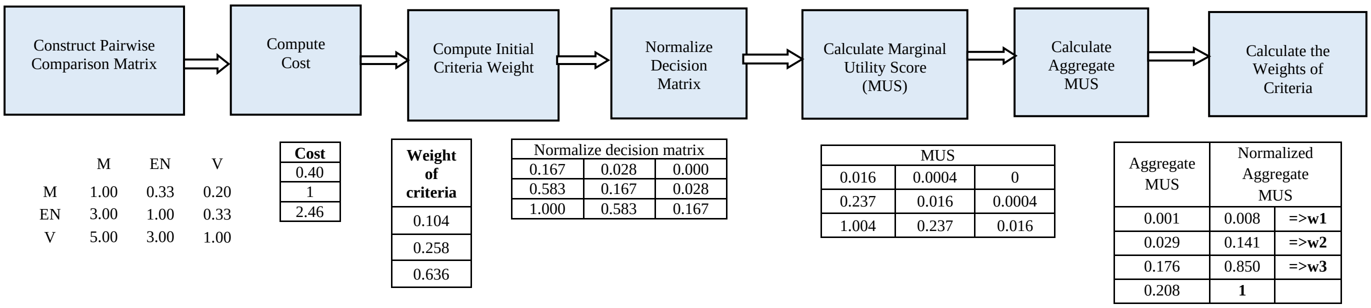
MAUT algorithm is utilized to choose the weights of the AoA algorithm. The pairwise comparison matrix (A_{ij}) decides the initial importance of the three criteria metrics of the features, such as mean (M), entropy (EN) and variance of the features. The variance of features is given very high importance over entropy and the mean of the features. The variance is preferred over entropy. The pairwise comparison matrix is normalized using equation 22. The Marginal Utility Score (MUS of every criterion is computed using exponential Equation 23 [43-44].

$$A_{ij}^{N} = \frac{A_{ij} - \min A_{ij}}{\max A_{ij} - \min A_{ij}} \tag{22}$$

$$MUS_{ij} = \frac{e^{(A_{ij}^{N})^2} - 1}{1.71} \tag{23}$$

The aggregate MUS is computed using equation 24, where w_j denotes the weight of the criteria.

$$MUS_i = \sum_j MUS_{ij} * w_j \tag{24}$$



	M	EN	V
M	1.00	0.33	0.20
EN	3.00	1.00	0.33
V	5.00	3.00	1.00

Cost
0.40
1
2.46

Weight of criteria
0.104
0.258
0.636

Normalize decision matrix		
0.167	0.028	0.000
0.583	0.167	0.028
1.000	0.583	0.167

MUS		
0.016	0.0004	0
0.237	0.016	0.0004
1.004	0.237	0.016

Aggregate MUS	Normalized Aggregate MUS	
0.001	0.008	**=>w1**
0.029	0.141	**=>w2**
0.176	0.850	**=>w3**
0.208	**1**	

Fig. 2 weight decision of AoA using MAUT algorithm

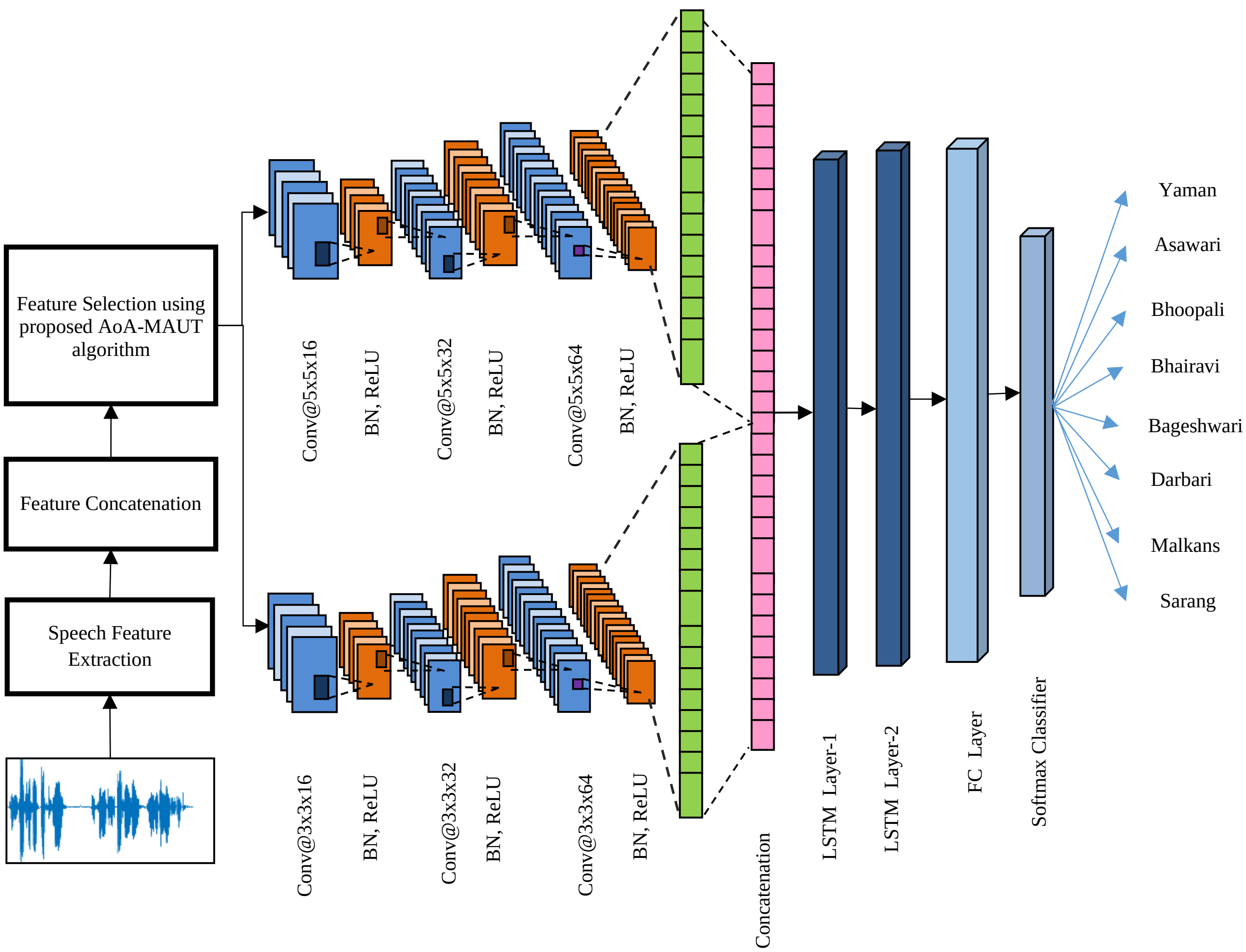


Fig. 3 Framework for proposed RagaNet for indian classical music recognition [49]

The proposed RagaNet consists of PDCNN and LSTM. The PDCNN provides the spectral representation of musical attributes and enhances the generalization capability of the raga recognition systems. The LSTM describes the long-term dependency of the musical features and provides a temporal representation of the musical features. The Parallel DCNN consists of two parallel arms of lightweight DCNN with 3×3, 5×5 and 7×7 filters at every parallel arm. Each parallel arm includes three convolution layers with 16, 32 and 64 filters. The convolution layers are trailed by the rectified linear unit layer (ReLU) to boost the non-linearity of the convolution features. The features of the last ReLU layers of all parallel arms are flattened and concatenated together [45-48]. The concatenated features are provided to two LSTM layers, with 100 hidden layers in each LSTM layer to enhance the temporal depiction of the musical features. The system of the proposed RagaNet is illustrated in Figure 3.

5. Result and Discussion

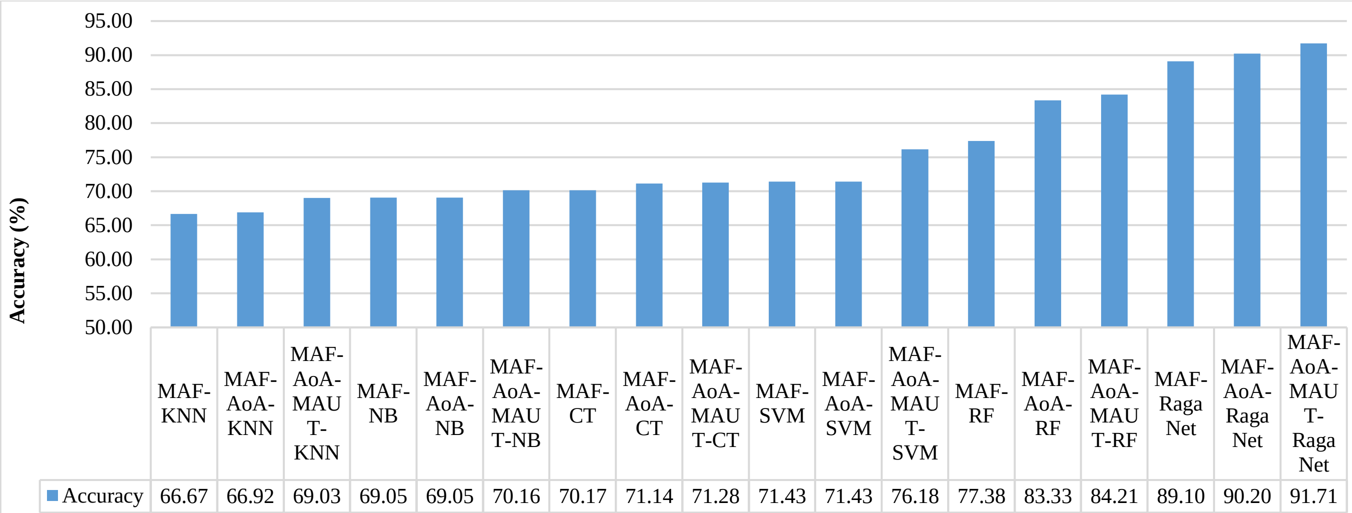
The effectiveness of the suggested raga identification technique is estimated for eight Indian ragas: Asawari, Bhoopali, Bhairavi, Bageshwari, Darbari, Malkans, Sarang and Yaman. The suggested system is simulated on MATLAB 2022 software. The effectiveness of the proposed DL-based framework is contrasted with traditional ML classifiers such as K-Nearest Neighbour (KNN), Naïve Bayes (NB), Classification Tree (CT), linear Support Vector Machine (SVM) and Random Forest (RF) as given in Table 2. Table 2 offers a comprehensive overview of the performance measurements—precision, Recall, F1-Score, and accuracy-across various classification methods. These methods are evaluated in a classification task, each representing a different approach to the problem. Beginning with simpler models like MAF-KNN and MAF-NB, we observe moderate performance levels in Precision, Recall and F1-Score, with Accuracy ranging from 66.67% to 69.05%. Performance metrics show a noticeable improvement as we move towards more feature selection techniques, such as those incorporating AoA and MAUT.

Models like MAF-AoA-MAUT-SVM and MAF-AoAMAUT-DCNN demonstrate high Precision, Recall and F1-Score, coupled with significantly improved Accuracy, surpassing 76%. Random Forest (RF) stands out among the ensemble methods, particularly when augmented with AoA and MAUT features. MAF-AoA-RF and MAFAoA-MAUT-RF exhibit exceptional performance, achieving an F1-Score and Accuracy of 0.83 and above, with the latter surpassing 84%.

This underscores the effectiveness of ensemble techniques in capturing complex relationships within the data. Moreover, the RagaNet models, especially those enriched with AoA and MAUT, showcase remarkable performance. The MAFAoA-MAUT-DCNN method leads the pack with an impressive accuracy of 91.71%. The accuracy of the recommended Raga recognition system is shown in Figure 4, whereas precision, recall, and F1 scores are provided in Figure 5. Figure 6 provides the effect of feature selection for different features selected using the AoAMAUT algorithm. It gives a superior accuracy of 91.71% for 200 features. As the number of features increases beyond 200, the redundancy of features increases, leading to poorer accuracy than the prominent 200 features.

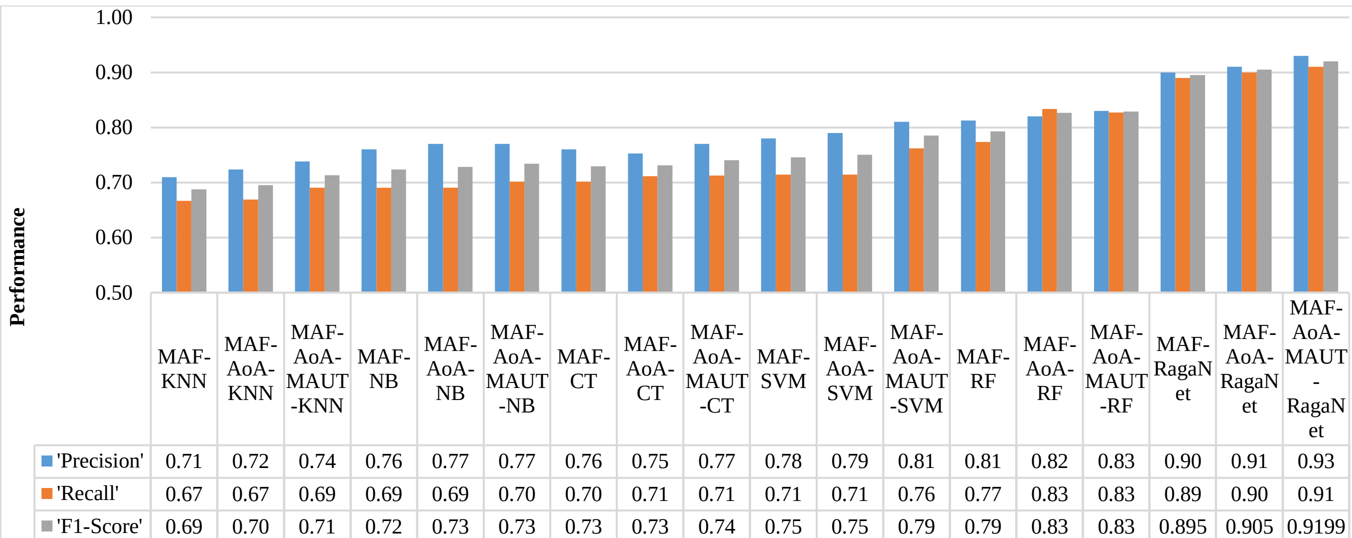
Table 2. Results comparison of the proposed method

Method	Precision	Recall	F1-Score	Accuracy
MAF-KNN	0.71	0.67	0.69	66.67
MAF-AoA-KNN	0.72	0.67	0.70	66.92
MAF-AoA-MAUT-KNN	0.74	0.69	0.71	69.03
MAF-NB	0.76	0.69	0.72	69.05
MAF-AoA-NB	0.77	0.69	0.73	69.05
MAF-AoA-MAUT-NB	0.77	0.70	0.73	70.16
MAF-CT	0.76	0.70	0.73	70.17
MAF-AoA-CT	0.75	0.71	0.73	71.14
MAF-AoA-MAUT-CT	0.77	0.71	0.74	71.28
MAF-SVM	0.78	0.71	0.75	71.43
MAF-AoA-SVM	0.79	0.71	0.75	71.43
MAF-AoA-MAUT-SVM	0.81	0.76	0.79	76.18
MAF-RF	0.81	0.77	0.79	77.38
MAF-AoA-RF	0.82	0.83	0.83	83.33
MAF-AoA-MAUT-RF	0.83	0.83	0.83	84.21
MAF-RagaNet	0.90	0.89	0.89	89.10
MAF-AoA-RagaNet	0.91	0.90	0.90	90.20
MAF-AoA-MAUT-RagaNet	0.93	0.91	0.91	91.71



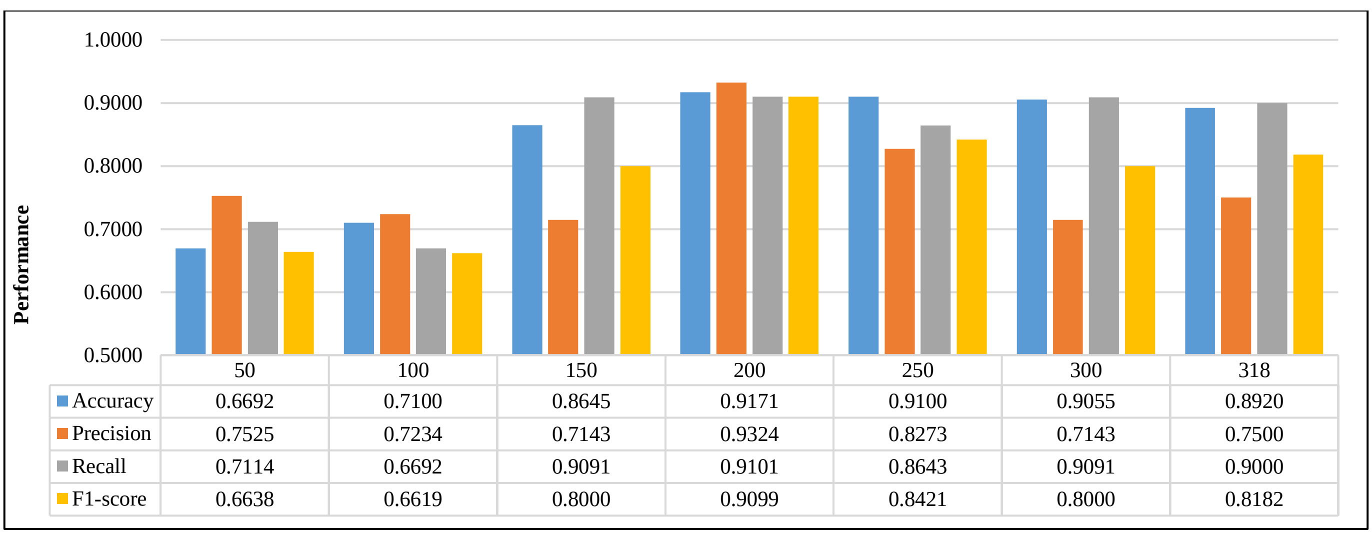
	MAF-KNN	MAF-AoA-KNN	MAF-AoA-MAUT-KNN	MAF-NB	MAF-AoA-NB	MAF-AoA-MAUT-NB	MAF-CT	MAF-AoA-CT	MAF-AoA-MAUT-CT	MAF-SVM	MAF-AoA-SVM	MAF-AoA-MAUT-SVM	MAF-RF	MAF-AoA-RF	MAF-AoA-MAUT-RF	MAF-RagaNet	MAF-AoA-RagaNet	MAF-AoA-MAUT-RagaNet
Accuracy	66.67	66.92	69.03	69.05	69.05	70.16	70.17	71.14	71.28	71.43	71.43	76.18	77.38	83.33	84.21	89.10	90.20	91.71

Fig. 4 Accuracy comparison of the proposed raga classification scheme



	MAF-KNN	MAF-AoA-KNN	MAF-AoA-MAUT-KNN	MAF-NB	MAF-AoA-NB	MAF-AoA-MAUT-NB	MAF-CT	MAF-AoA-CT	MAF-AoA-MAUT-CT	MAF-SVM	MAF-AoA-SVM	MAF-AoA-MAUT-SVM	MAF-RF	MAF-AoA-RF	MAF-AoA-MAUT-RF	MAF-RagaNet	MAF-AoA-RagaNet	MAF-AoA-MAUT-RagaNet
'Precision'	0.71	0.72	0.74	0.76	0.77	0.77	0.76	0.75	0.77	0.78	0.79	0.81	0.81	0.82	0.83	0.90	0.91	0.93
'Recall'	0.67	0.67	0.69	0.69	0.69	0.70	0.70	0.71	0.71	0.71	0.71	0.76	0.77	0.83	0.83	0.89	0.90	0.91
'F1-Score'	0.69	0.70	0.71	0.72	0.73	0.73	0.73	0.73	0.74	0.75	0.75	0.79	0.79	0.83	0.83	0.895	0.905	0.9199

Fig. 5 Precision, Recall, and F1-score comparison of the proposed raga classification scheme



	50	100	150	200	250	300	318
Accuracy	0.6692	0.7100	0.8645	0.9171	0.9100	0.9055	0.8920
Precision	0.7525	0.7234	0.7143	0.9324	0.8273	0.7143	0.7500
Recall	0.7114	0.6692	0.9091	0.9101	0.8643	0.9091	0.9000
F1-score	0.6638	0.6619	0.8000	0.9099	0.8421	0.8000	0.8182

Fig. 6 Effect of feature selection using AoA-MAUT

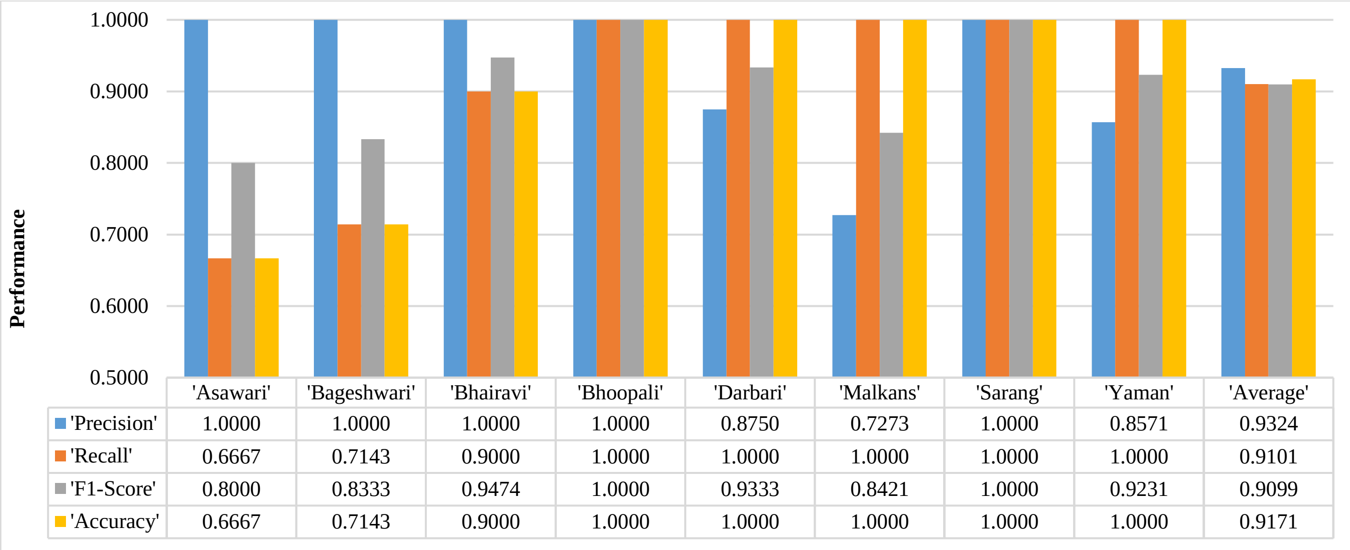


Fig. 7 Results for various ragas using proposed RagaNet

The results of the proposed RagaNet for 200 features selected using AoA-MAUT are provided in Figure 7. It offers the highest accuracy of 100% for Bhoopali, Darbari, Malkans, Yaman, and Sarang, whereas the lowest accuracy of 66.67% for Asawari. The average accuracy for eight classes is 91.71%. The proposed RagaNet provides better results for the Ragas with high pitch variation and style variations. The outcomes of the suggested plan were contrasted with established techniques that used ML and DL frameworks for the raga classification. The suggested scheme gives a superior overall accuracy of 91.71% for eight-class raga classification, as given in Table 3. The VGG16 and ResNet50 provided 55.07% and 68.08% accuracy for 2-class raga classification, respectively. The Deep BiLSTM RNN resulted in an overall accuracy of 78% for 5 Hindustani ragas. The CNN-based system has given 94% accuracy for 5 Carnatic ragas classifications. The RagaNet helps to achieve the generalization capability by utilizing two parallel arms of the DCNN, which helps to learn the multilevel hierarchical features with the help of different filter kernels. The feature selection increases the feature depiction capability of the raw raga signal and helps to improve the spectral, temporal, and voice quality depiction of the classical music sounds. The feature selection also helps minimize the system's redundant features and computational complexity. The RagaNet offers an overall accuracy of 88.20% for MAFs, 90% for MAF-AoA, and 91.71% for MAF-AoA-MAUT, respectively. The MAUT-based weight optimization of the AoA algorithm assists in selecting the salient features from MAFs having higher mean, entropy, and variance. The system can be used in many applications, such as music education, cultural preservation, music therapy, and music recording. It can be extended to recognize and correct Indian classical playback singing and generating classical raga synthetically. It also helps in studying the note and pitch changes in the voice in different Indian musical Gharanas. The suggested system's effectiveness is limited due to the standard dataset's availability.

Table 3. Comparison of the proposed method with the previous state of arts

Authors and Years	Method	Number of ragas	Accuracy
Joshi et al. (2021) [29]	SVM, KNN	2 ragas (Yaman and Bhairavi)	KNN=98%, SVM=94%
John S. (2020) [26]	CNN	5 ragas (Carnatic Ragas are Bhairavi, Hindolam, Ragamalika, Sankarabharanam and Thodi)	94%
Pendyala et al. (2022) [27]	Deep BiLSTM RNN with MFCC input	5 Hindustani classical ragas	78%
Sharma et al. (2021) [28]	ResNet50	2 class (Hindustani and Carnatic)	68.08%
	VGG16		55.07%
Proposed Method	MAF-RagaNet	8 Indian ragas (Asawari, Bageshwari, Bhairavi, Bhoopali, Darbari, Malkans, Sarang and Yaman)	88.20%
	MAF-AoA-RagaNet		90%
	MAF-AoA-MAUT-RagaNet		**91.71%**

The system's performance is challenging due to the musical notes in the raga and vast changes in the prosodic features. The lower dataset leads to the class imbalance problem and shows a significant disparity in the recall and precision of the system.

6. Conclusion and Future Scope

Thus, this work introduces Indian Classical music recognition using multiple acoustic features and a Deep Convolution neural network. The MAFs depict the raga variations over the voice through temporal, spectral and voice quality attributes and assist in grabbing the long-term dependencies of the signal. The AoA-MAUT-based feature selection scheme provides the prominent feature selection. The MAUT algorithm helps to decide the proper weights of the objective function. It is seen that the AoA-MAUT-based feature selection assists with choosing the salient features by limiting the redundant and non-important features to enhance the raga feature distinctiveness. The suggested MAF-AoA-MAUT-RagaNet-based Indian raga classification provides an accuracy of 91.71%, precision of 0.93, recall of 0.90 and F1-score of 0.91.

The recommended system gives better classification accuracy compared with MAF-RagaNet and MAF-AoA-RagaNet. In the future, more focus can be given to developing a generalized raga identification system that can classify more ragas of both Carnatic and Hindustani styles.

The future focus can be on real-time raga recognition and cross-cultural raga classification. The system's effectiveness can be boosted by inculcating interpretability and explainability in the DL framework.

References

[1] Rajeswari Sridhar, and T.V. Geetha, "Swara Indentification for South Indian Classical Music," *9th International Conference on Information Technology (ICIT'06)*, Bhubaneswar, India, pp. 143-144, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[2] Rajeswari Sridhar, and T.V. Geetha, "Music Information Retrieval of Carnatic Songs Based on Carnatic Music Singer Identification," *2008 International Conference on Computer and Electrical Engineering*, Phuket, Thailand, pp. 407-411, 2008. [CrossRef] [Google Scholar] [Publisher Link]

[3] Gaurav Pandey, Chaitanya Mishra, and Paul Ipe, "TANSEN: A System for Automatic Raga Identification," *Indian International Conference on Artificial Intelligence*, pp. 1350-1363, 2003. [Google Scholar]

[4] J.P. Bello et al., "A Tutorial on Onset Detection in Music Signals," *IEEE Transactions on Speech and Audio Processing*, vol. 13, no. 5, pp. 1035-1047, 2005. [CrossRef] [Google Scholar] [Publisher Link]

[5] Anssi Klapuri, and Manuel Davy, *Signal Processing Methods for Music Transcription*, Springer New York, NY, 2006. [CrossRef] [Google Scholar] [Publisher Link]

[6] Parag Chordia, "Automatic Raag Classification of Pitch-Tracked Performances Using Pitch-Class and Pitch-Class Dyad Distributions," *International Conference on Mathematics and Computing (ICMC)*, pp. 314-321, 2006. [Google Scholar] [Publisher Link]

[7] Graham E. Poliner et al., "Melody Transcription from Music Audio: Approaches and Evaluation," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 4, pp. 1247-1256, 2007. [CrossRef] [Google Scholar] [Publisher Link]

[8] S. Samsekai Manjabhat et al., "Raga and Tonic Identification in Carnatic Music," *Journal of New Music Research*, vol. 46, no. 3, pp. 229-245, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[9] Ram Avtar 'Vir, *Theory of Indian Music*, Pankaj Publications, New Delhi, India, 1999. [Publisher Link]

[10] Surendra Shetty, and Sarika Hegde, "Automatic Classification of Carnatic Music Instruments Using MFCC and LPC," *Proceedings of Data Management, Analytics and Innovation*, pp. 463-474, 2020. [CrossRef] [Google Scholar] [Publisher Link]

[11] Keunwoo Choi, George Fazekas, and Mark Sandler, "Automatic Tagging Using Deep Convolutional Neural Networks," *Arxiv*, pp. 1-7, 2016. [CrossRef] [Google Scholar] [Publisher Link]

[12] Ashu Abdul et al., "An Emotion-Aware Personalized Music Recommendation System Using a Convolutional Neural Networks Approach," *Applied Sciences*, vol. 8, no. 7, pp. 1-16, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[13] Shun-Hao Chang et al., "A Personalized Music Recommendation System Using Convolutional Neural Networks Approach," *IEEE International Conference on Applied System Invention*, Chiba, Japan, pp. 47-49, 2018. [CrossRef] [Google Scholar] [Publisher Link]

[14] Miao Jiang, Ziyi Yang, and Chen Zhao, "What to Play Next? Arnn-Based Music Recommendation System," *2017 51st Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, pp. 356-358, 2017. [CrossRef] [Google Scholar] [Publisher Link]

[15] Tara N. Sainath et al., "Convolutional, Long Short-Term Memory, Fully Connected Deep Neural Networks," *IEEE International Conference on Acoustics, Speech and Signal Processing*, South Brisbane, QLD, Australia, pp. 4580-4584, 2015. [CrossRef] [Google Scholar] [Publisher Link]

[16] Md Afif Al Mamun et al., "Bangla Music Genre Classification Using Neural Network," *2019 8th International Conference System Modeling and Advancement in Research Trends*, Moradabad, India, pp. 397-403, 2019. [CrossRef] [Google Scholar] [Publisher Link]

[17] Adiyansjah, Alexander A.S. Gunawan, and Derwin Suhartono, "Music Recommender System Based on Genre Using Convolutional Recurrent Neural Networks," *Procedia Computer Science*, vol. 157, pp. 99-109, 2019. [CrossRef] [Google Scholar] [Publisher Link]
[18] Prasenjeet Fulzele et al., "A Hybrid Model for Music Genre Classification Using LSTM and SVM," *2018 Eleventh International Conference on Contemporary Computing*, Noida, India, pp. 1-3, 2018. [CrossRef] [Google Scholar] [Publisher Link]
[19] Yu Tao, Yuanxing Zhang, and Kaigui Bian, "Attentive Context-Aware Music Recommendation," *2019 IEEE Fourth International Conference on Data Science in Cyberspace*, Hangzhou, China, pp. 54-61, 2019. [CrossRef] [Google Scholar] [Publisher Link]
[20] Nagamanoj Karunakaran, and Arti Arya, "A Scalable Hybrid Classifier for Music Genre Classification Using Machine Learning Concepts and Spark," *2018 International Conference on Intelligent Autonomous Systems*, Singapore, pp. 128-135, 2018. [CrossRef] [Google Scholar] [Publisher Link]
[21] Frank Zalkow, and Meinard Müller, "Using Weakly Aligned Score-Audio Pairs to Train Deep Chroma Models for Cross-Modal Music Retrieval," *Proceedings of the 21st ISMIR Conference*, Montreal, Canada, pp. 184-191, 2020. [Google Scholar] [Publisher Link]
[22] Wing W.Y. Ng, Weijie Zeng, and Ting Wang, "Multi-Level Local Feature Coding Fusion for Music Genre Recognition," *IEEE Access*, vol. 8, pp. 152713-152727, 2020. [CrossRef] [Google Scholar] [Publisher Link]
[23] A. Elbir, and N. Aydin, "Music Genre Classification and Music Recommendation by Using Deep Learning," *Electronics Letters*, vol. 56, no. 12, pp. 627-629, 2020. [CrossRef] [Google Scholar] [Publisher Link]
[24] Rosilde Tatiana Irene et al., "Automatic Playlist Generation Using Convolutional Neural Networks and Recurrent Neural Networks," *2019 27th European Signal Processing Conference (EUSIPCO)*, A Coruna, Spain, pp. 1-5, 2019. [CrossRef] [Google Scholar] [Publisher Link]
[25] Hyoung-Gook Kim, Gee Yeun Kim, and Jin Young Kim, "Music Recommendation System Using Human Activity Recognition From Accelerometer Data," *IEEE Transactions on Consumer Electronics*, vol. 65, no. 3, pp. 349-358, 2019. [CrossRef] [Google Scholar] [Publisher Link]
[26] Siji John et al., "Classification of Indian Classical Carnatic Music Based on Raga Using Deep Learning," *2020 IEEE Recent Advances in Intelligent Computational Systems*, Thiruvananthapuram, India, pp. 110-113, 2020. [CrossRef] [Google Scholar] [Publisher Link]
[27] Vishnu S. Pendyala et al.,"Towards Building a Deep Learning Based Automated Indian Classical Music Tutor for the Masses," *Systems and Soft Computing*, vol. 4, 2022. [CrossRef] [Google Scholar] [Publisher Link]
[28] Akhilesh Kumar Sharma et al., "Classification of Indian Classical Music with Time-Series Matching Deep Learning Approach," *IEEE Access*, vol. 9, pp. 102041-102052, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[29] Dipti Joshi, Jyoti Pareek, and Pushkar Ambatkar, "Indian Classical Raga Identification Using Machine Learning," *International Semantic Intelligence Conference*, New Delhi, India, pp. 259-263, 2021. [Google Scholar] [Publisher Link]
[30] Kishor Barasu Bhangale, and Mohanaprasad Kothandaraman, "Survey of Deep Learning Paradigms for Speech Processing," *Wireless Personal Communications*, vol. 125, no. 2, pp. 1913-1949, 2022. [CrossRef] [Google Scholar] [Publisher Link]
[31] Kishor Barasu Bhangale, and Mohanaprasad Kothandaraman, "A Review on Speech Processing Using Machine Learning Paradigm," *International Journal of Speech Technology*, vol. 24, no. 2, pp. 367-388, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[32] Kishor Bhangale, and Mohanaprasad Kothandaraman, "Speech Emotion Recognition Based on Multiple Acoustic Features and Deep Convolutional Neural Network," *Electronics*, vol. 12, no. 4, pp. 1-17, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[33] Neha Chauhan, Tsuyoshi Isshiki, and Dongju Li, "Speaker Recognition Using LPC, MFCC, ZCR Features with ANN and SVM Classifier for Large Input Database," *2019 IEEE 4th International Conference on Computer and Communication Systems*, Singapore, pp. 130-133, 2019. [CrossRef] [Google Scholar] [Publisher Link]
[34] Yulia Oganian et al., "Vowel and Formant Representation in the Human Auditory Speech Cortex," *Neuron*, vol. 111, no. 13, pp. 2105-2118, 2023. [CrossRef] [Google Scholar] [Publisher Link]
[35] Anagha Sonawane, M.U. Inamdar, and Kishor B. Bhangale, "Sound Based Human Emotion Recognition Using MFCC & Multiple SVM," *2017 International Conference on Information, Communication, Instrumentation and Control*, Indore, India, pp. 1-4, 2017. [CrossRef] [Google Scholar] [Publisher Link]
[36] Bryce J. Dietrich, Matthew Hayes, and Diana Z. O'brien, "Pitch Perfect: Vocal Pitch and the Emotional Intensity of Congressional Speech," *American Political Science Review*, vol. 113, no. 4, pp. 941-962, 2019. [CrossRef] [Google Scholar] [Publisher Link]
[37] Seema Chaudhary, Sangeeta Kakarwal, and Ratnadeep Deshmukh, "Musical Instrument Recognition Using Audio Features with Integrated Entropy Method," *Journal of Integrated Science and Technology*, vol. 9, no. 2, pp. 92-97, 2021. [Google Scholar] [Publisher Link]
[38] Abraham Woubie, Lauri Koivisto, and Tom Bäckström, "Voice-Quality Features for Deep Neural Network Based Speaker Verification Systems," *2021 29th European Signal Processing Conference (EUSIPCO)*, Dublin, Ireland, pp. 176-180, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[39] Fatma A. Hashim et al., "Archimedes Optimization Algorithm: A New Metaheuristic Algorithm for Solving Optimization Problems," *Applied Intelligence*, vol. 51, pp. 1531-1551, 2021. [CrossRef] [Google Scholar] [Publisher Link]
[40] Abeer S. Desuky et al., "EAOA: An Enanced Archimedes Optimization Algorithm for Feature Selection in Classification," *IEEE Access*, vol. 9, pp. 120795-120814, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[41] Ahmed Fathy et al., "Archimedes Optimization Algorithm Based Maximum Power Point Tracker for Wind Energy Generation System," *Ain Shams Engineering Journal*, vol. 13, no. 2, pp. 1-18, 2022. [CrossRef] [Google Scholar] [Publisher Link]

[42] Lifang Zhang et al., "Ensemble Wind Speed Forecasting with Multi-Objective Archimedes Optimization Algorithm and Sub-Model Selection," *Applied Energy*, vol. 301, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[43] Susmita Bandyopadhyay, "A Novel Multi-Criteria Decision Analysis Technique Considering Various Essential Characteristics," *Research Square*, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[44] José Ramón San Cristóbal Mateo, *Multi Criteria Analysis in the Renewable Energy Industry*, 1st ed., Springer London, 2012. [CrossRef] [Google Scholar] [Publisher Link]

[45] Kishor B. Bhangale, and Mohanaprasad Kothandaraman, "Speech Emotion Recognition Using the Novel PEmoNet (Parallel Emotion Network)," *Applied Acoustics*, vol. 212, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[46] Kishor Bhangale, and K. Mohanaprasad, "Speech Emotion Recognition Using Mel Frequency Log Spectrogram and Deep Convolutional Neural Network," *International Conference on Futuristic Communication and Network Technologies*, pp. 241-250, 2021. [CrossRef] [Google Scholar] [Publisher Link]

[47] Ashish Pradhan, and Archit Yajnik, "Parts-of-Speech Tagging of Nepali Texts with Bidirectional LSTM, Conditional Random Fields and HMM," *Multimedia Tools and Applications*, vol. 83, no. 4, pp. 9893-9909, 2023. [CrossRef] [Google Scholar] [Publisher Link]

[48] Yunus Korkmaz, and Aytug Boyaci, "Hybrid Voice Activity Detection System Based on LSTM and Auditory Speech Features," *Biomedical Signal Processing and Control*, vol. 80, 2023. [CrossRef] [Google Scholar] [Publisher Link]