International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 18 August 2023 **Revised:** 10 October 2023 **Accepted:** 22 October 2023*

Thaat Classification Using Recurrent Neural Networks with Long Short-Term Memory and Support Vector Machine

Swati Shilaskar[1], Shripad Bhatlawande[2], Shivam Shinde[3], Soham Sattigeri[4]
[1]Dept. of E&TC Engg.
VIT, Pune, India 411038
swati.shilaskar@vit.edu
[2]Dept. of E&TC Engg.
VIT, Pune, India 411038
shripad.bhatlawande@vit.edu
[3]Dept. of E&TC Engg.
VIT, Pune, India 411038
shivam.shinde20@vit.edu
[4]Dept. of E&TC Engg.
VIT, Pune, India 411038
soham.sattigeri20@vit.edu

Abstract—This research paper introduces a groundbreaking method for music classification, emphasizing thaats rather than the conventional raga-centric approach. A comprehensive range of audio features, including amplitude envelope, RMSE, STFT, spectral centroid, MFCC, spectral bandwidth, and zero-crossing rate, is meticulously used to capture thaats' distinct characteristics in Indian classical music. Importantly, the study predicts emotional responses linked with the identified thaats. The dataset encompasses a diverse collection of musical compositions, each representing unique thaats. Three classifier models - RNN-LSTM, SVM, and HMM - undergo thorough training and testing to evaluate their classification performance. Initial findings showcase promising accuracies, with the RNN-LSTM model achieving 85% and SVM performing at 78%. These results highlight the effectiveness of this innovative approach in accurately categorizing music based on thaats and predicting associated emotional responses, providing a fresh perspective on music analysis in Indian classical music.

Keywords- music classification, MFCC, thaats, signal processing, HMM, SVM, RNN-LSTM.

I. INTRODUCTION

Indian classical music has long been celebrated for its intricate melodies and profound emotional expressions, featuring a vast array of ragas that constitute the core of musical compositions [1]. Ragas are intricate melodic structures that form the basis of melody in Indian classical music. Each raga is characterized by a specific set of ascending and descending notes, as well as distinctive tonal patterns, which are traditionally associated with times of day, seasons, and emotions. Ragas serve as a framework for improvisation and musical expression, allowing musicians to create intricate and evocative melodies within the established guidelines of each raga. The task of classifying ragas is essential in music information retrieval, as it involves the automated identification and categorization of these distinctive ragas from audio signals. Diverse methodologies have been explored by researchers in this endeavor, ranging from conventional approaches like support vector machines (SVMs) and hidden Markov models (HMMs) to cutting-edge deep learning techniques that employ convolutional neural networks (CNNs) [2]. Various feature extraction methods, including the utilization of mel-frequency cepstral coefficients (MFCCs), chroma features, and spectral roll-off features, have been harnessed to capture the unique characteristics of ragas within the audio data [3]. By accurately identifying ragas, automatic music recommendation systems can be empowered, music education can be enhanced, and emotional analysis of musical compositions can be facilitated [4]. Such achievements in developing robust music classification models have the potential to revolutionize our interaction with and comprehension of Indian classical music, ushering in a new era of exploration and profound appreciation for this invaluable cultural treasure [5]. However, upon delving deeper into the realm of Indian classical music, it becomes apparent that the framework of ragas represents just one facet of its rich and intricate tapestry. In addition to ragas, another fundamental aspect deserving of attention is the concept of thaats. Thaats constitute a distinct classification system within Indian classical music, offering a unique perspective on the organization and comprehension of musical compositions.

Thaats are served as a complementary approach to ragas, with their advantages over raga classification being worthy of exploration. In contrast to ragas, thaats provide a broader classification of scales and melodic frameworks that encompass multiple ragas within each thaat. This broader perspective can be advantageous for both musicians and musicologists, as it allows for a more comprehensive understanding of the relationships between various ragas and their underlying structural similarities.

Furthermore, thaats offer a simplified framework for learners and enthusiasts to navigate the intricate world of Indian classical music. While ragas can be highly specific and nuanced, thaats provide a broader context that aids in the initial comprehension of the musical landscape. This makes thaats an excellent starting point for those embarking on their journey into the depths of Indian classical music. The significance of thaat classification extends beyond its educational value. Similar to raga classification, it plays a pivotal role in preserving and promoting the rich legacy of Indian classical music. Through the accurate identification of thaats, automatic music recommendation systems can be empowered, music education can be enhanced, and emotional analysis of musical compositions can be facilitated. As the focus shifts from ragas to thaats, a journey of exploration and profound appreciation for this invaluable cultural treasure is undertaken. In this research paper, an exploration of the world of thaats in Indian classical music is initiated, encompassing the examination of their classification, attributes, and the potential inherent in them to enrich the understanding of this timeless musical tradition. Through this endeavor, illumination is aimed to be shed on the significance of thaats and their advantages over traditional raga classification methodologies, ultimately contributing to the continued evolution of research and practice in Indian classical music.

II. RELATED WORK

Researchers, Bidkar et al., put forth an approach for recognizing North Indian classical ragas in instrumental music [6]. Employing a two-step process involving feature selection and SVM classification, their study encompassed a dataset comprising 1200 audio files, each featuring six distinct ragas played on four diverse instruments: sitar, sarod, santoor, and flute. The features extracted encompassed spectral, temporal, and rhythmic attributes. In a parallel endeavor, Nawasalkar et al. introduced an EEG-based stress recognition system, utilizing Indian classical music as a medium [7]. With a 14-channel EEG headset capturing brain signals, the authors extracted key features like power spectral density, entropy, and correlation coefficients. Their evaluation on 20 subjects, each exposed to different Indian classical music ragas, resulted in an accuracy of 88.9%. Moving on, Bargaje's work focused on developing an emotion recognition and classification system for audio, employing genetic algorithms [8]. The authors' study embraced a varied dataset of 100 audio files, including music, speech, and environmental sounds. By extracting spectral, temporal, and rhythmic features and training a genetic algorithm for emotion-based classification, the system achieved an accuracy of 83.2%. The authors of [9] provide a comprehensive overview of the use of deep learning for music signal processing The study explores datasets employed in music deep learning, such as the Million Song Dataset, McGill University Dataset of Music Signals, and Indian Classical Music Dataset. It delves into feature extraction, encompassing mel-frequency cepstral coefficients (MFCCs), chroma, and spectral attributes from music signals. Additionally, the survey discusses methodologies encompassing supervised, unsupervised, and reinforcement learning for training and evaluating music-oriented deep learning models.

Meanwhile, Tejaswini Priyanka et al. presented an emotion-based music recommendation system employing CNN [10]. Their approach involved detecting user emotions through a CNN trained on facial images with seven distinct emotions. Leveraging past listening history and user feedback, the system generated music recommendations tailored to detected emotions, culminating in an accuracy of 80%. The authors, however, omitted analyzing the system's efficiency when employing multiple classifiers, leaving room for further investigation. On another front, Chhetri et al. delved into Carnatic music identification of melakarta ragas [11]. Extracting MFCCs, chroma, and spectral features, the authors achieved 90% accuracy utilizing an SVM classifier on a dataset comprising 1000 audio files. While their system demonstrated competence, optimizing it through an extensive dataset and evaluation across diverse music types could be worthwhile. Suma and Koolagudi embarked on a raga classification journey for Carnatic music [12]. Employing spectral and temporal feature extraction, normalized for unbiased results, and subsequently employing SVM classification, their system achieved an accuracy of 91.5%. The paper [13] authored by H. Purwins et al., conducts an investigation into the integration of deep learning techniques within the realm of audio signal processing. The paper discusses various datasets, including the Million Song Dataset, utilized for training and evaluation, and emphasizes the significance of features such as mel-frequency cepstral coefficients (MFCCs) and chroma features. Additionally, the survey sheds light on methodologies like supervised, unsupervised, and reinforcement learning, showcasing their roles in tasks spanning from speech recognition to music generation.

Additionally, Acharya et al. conducted a comprehensive comparison of various classification models for raga identification in Carnatic classical audio [14]. Evaluating SVM, RF, CNN, and XGBoost, their study crowned XGBoost with the highest accuracy of 97%. Notwithstanding this, the authors

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 18 August 2023 **Revised:** 10 October 2023 **Accepted:** 22 October 2023*

acknowledged limited identification of the number of ragas and the exclusion of instrumental audio from their analysis, leaving scope for further refinement. Deng and Leung presented a music emotion retrieval system based on acoustic features [15]. Their study featured SVM for emotion classification and an innovative approach to expand user queries using acoustic features, enhancing emotion retrieval accuracy. Nonetheless, the authors acknowledged SVM's sensitivity to hyperparameters and data balance, opening avenues for further optimization. Shetty, Achary, and Hegde introduced raga identification in Carnatic music through the Hidden Markov Model (HMM) technique [16]. Featuring feature extraction, HMM state modeling, transition modeling, and classification, their system demonstrated an accuracy of 90%. The paper authored by Shah et al. proposed raga recognition using deep learning [17]. The authors extracted features from the audio signal, preprocessed them, trained a convolutional neural network (CNN) for classification, and achieved an accuracy of 72% on a relatively small dataset. The authors of [18] introduce a deep learning method to recognize melodic patterns in Hindustani classical music. Using a dataset of 1000 labeled recordings, the system employs pitch, loudness, and spectral flatness features, and utilizes a bidirectional LSTM neural network. The system achieves a commendable 90% accuracy on the test set, signifying its potential for automated music learning. The paper proposed by authors of [19] suggests a deep learning model for music melody extraction, achieving state-of-the-art accuracy of 92.3% on the MIREX Melody Extraction Dataset. The model extracts MFCCs, chroma features, tempo features, and beat features from the raw music signal, and uses a deep neural network classifier to predict the melody. In [20], the authors propose a method for automatic music transcription using Fourier transform for both monophonic and polyphonic audio files. The method extracts Fourier transform features from the audio signal and uses a simple rule-based classifier to predict the note values and pitches. The method achieves relatively small percentage errors for note value and pitch detection, but a relatively moderate percentage error for extra note detection in polyphonic files.

The field of raga classification in Indian classical music has made notable strides, yet several challenges persist. One common concern revolves around the size of datasets, which, when small, can impact model generalizability and performance on unseen data. Moreover, class imbalances within datasets may introduce biases and hinder accurate classification. Additionally, certain studies have exclusively focused on specific musical genres within Indian classical music, overlooking the diverse range of ragas present in other musical forms. Evaluating models across a broader spectrum of genres could enhance their practicality and overall robustness. Furthermore, some studies have solely employed a single classifier, limiting the scope of analysis on different model performances. Another essential consideration often disregarded is the computational complexity and efficiency of proposed methods. Optimizations may be required to handle the vast amount of audio data efficiently, especially for real-time or large-scale applications. Lastly, the absence of standardized datasets and evaluation protocols poses challenges in comparing results across studies. Establishing common benchmarks would foster advancements in raga classification research and promote collaborative efforts among researchers. Tackling these challenges promises to unlock more precise and effective raga classification methods, thereby contributing to the preservation and deeper appreciation of the rich heritage embodied in Indian classical music. In light of the enduring challenges in the classification of Indian classical music ragas, this research introduces a groundbreaking perspective that constitutes a fundamental departure from the traditional raga classification approach. Instead of adhering exclusively to raga-based classification, the proposed approach centers on the classification of thaats. This novel perspective offers a transformative means of surmounting several of the constraints encountered in the conventional raga classification method. Thaats, as an organizational structure in Indian classical music, encompassing a collection of related ragas, provide a more comprehensive and adaptable framework for classification. This innovative approach not only holds the potential to resolve issues related to dataset size, class imbalances, and genre diversity but also permits a more all-encompassing examination of the music, emphasizing the inherent interconnections between ragas. Furthermore, the research places a strong emphasis on computational efficiency, ensuring that the shift from raga-centric to thaat-centric classification is not only conceptually well-founded but also practical for real-time and large-scale applications. By pioneering this novel perspective and advocating for the adoption of thaats as the basis for classification, this research paper represents a groundbreaking contribution to the field. This approach will foster progress in raga classification research, enriching the preservation and fostering a deeper appreciation of the opulent heritage embodied in Indian classical music.

III. METHODOLGY

Temporal, spectral, and time-frequency attributes play a central role in the categorization of thaats, a core aspect of music analysis and comprehension. Temporal characteristics capture the evolving nature of musical events over time, including elements like note duration and intervals between onsets. These aspects are vital for distinguishing the unique rhythmic patterns and phrasing within each thaat. Conversely, spectral attributes elucidate the frequency composition of musical signals, facilitating the differentiation of various thaats based on tonal qualities such as pitch and timbre. Time-frequency traits bridge the temporal and spectral dimensions,

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 18 August 2023 **Revised:** 10 October 2023 **Accepted:** 22 October 2023*

enabling the exploration of how frequency components evolve temporally. These attributes are critical for capturing the subtle intricacies present in thaats, offering insights into ornamentation and melodic variations. After extracting these characteristics, they serve as the basis for training classification models. Recurrent Neural Networks (RNNs) featuring Long Short-Term Memory (LSTM) cells are harnessed to capture sequential dependencies and discern intricate temporal patterns within thaats. Meanwhile, Hidden Markov Models (HMMs) are employed to represent the underlying probabilistic structures, aiding in the recognition of thaats. Through the application of these feature extraction techniques and advanced modeling methods, thaats can be accurately categorized, leading to a deeper understanding of their intricate musical structures, as illustrated in Fig. 1.

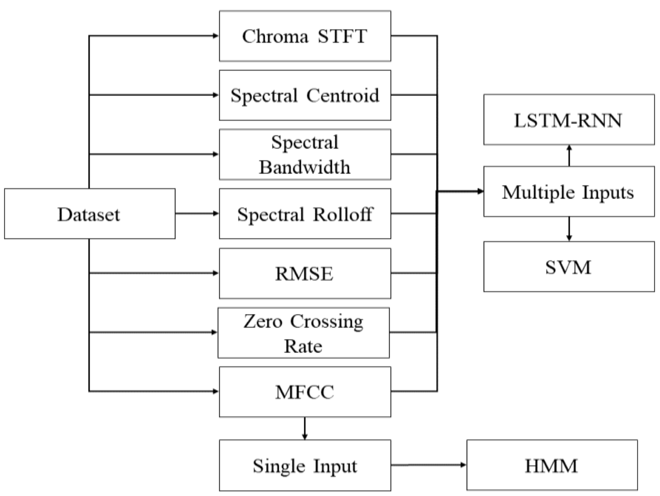

Figure 1. Flowchart representing a step by step procedural flow of the proposed system.

The experimentation transpired within the confines of a Google Collaboratory computational environment. This environment featured 8 gigabytes of RAM and was equipped with an Intel(R) CPU operating at a clock speed of 2.20 gigahertz. The training procedure encompassed both single and multi-channel input data, and it was executed on the Tensor Processing Unit (TPU) resources provided by Google Collaboratory. In this study, a constant Sample Rate of 22050 Hz was considered. Subsequently, the dataset was divided into training and test subsets, maintaining an 80:20 ratio.

A. Data Acquisition

The dataset employed for the purpose of classifying music based on thaats encompasses audio samples from nine unique thaats. Each audio file in the dataset adheres to a uniform 30-second duration, which is achieved by partitioning longer audio recordings into standardized segments. The dataset encompasses a range of thaats as shown in Table I, each with its respective quantity of audio files, including "Bilawal" (364 audio files), "Kalyan" (378 audio files), "Asavari" (125 audio files), "Khamaj" (100 audio files), "Marwa" (496 audio files), "Purvi" (149 audio files), "Bhairav" (273 audio files), "Bhairavi" (335 audio files), and "Todi" (435 audio files).

TABLE I. DATASET CONTAINING INFORMATION ABOUT THE THAATS USED FOR PROPOSED MODEL

Sr No.	Name of the Thaat	Number of Audio Files
1.	Bilawal	364
2.	Kalyan	378
3.	Asavari	125
4.	Khamaj	100
5.	Marwa	496
6.	Purvi	149
7.	Bhairav	273
8.	Bhairavi	335
9.	Todi	435

This dataset has been meticulously organized to facilitate the classification of musical compositions based on their alignment with these nine specific thaats, while maintaining a consistent 30-second duration for each audio segment.

B. Feature Extraction

Temporal and spectral attributes have a profound influence on thaat categorization by capturing the distinct elements of the music. Temporal features, such as the amplitude envelope, reveal how the signal's amplitude changes over time, providing insights into the evolving dynamics and articulation within a thaat. The zero-crossing rate measures how quickly the signal changes polarity, aiding in recognizing the rhythmic intricacies that define thaats, as different thaats often exhibit unique rhythmic patterns. Root mean square energy offers a measure of the overall signal energy, helping to differentiate thaats with varying levels of intensity or loudness. Spectral features, including spectral centroid and spectral bandwidth, contribute to thaat classification by highlighting tonal attributes and timbral distinctions, allowing for the differentiation of thaats with similar note sequences but differing tonal qualities. Band energy ratios offer insights into how energy is distributed across various frequency bands, aiding in identifying tonal centers and harmonically significant areas within the music. In combination, these temporal and spectral attributes act as distinctive markers, facilitating precise thaat classification by capturing the unique temporal and spectral characteristics. Visual representations of the impact of temporal and spectral coefficients on a sample audio file can be seen in Fig. 4 and Fig. 5, respectively.

The variations in melodic patterns of different thaats can be better understood by examining and comparing the graphs

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 18 August 2023 **Revised:** 10 October 2023 **Accepted:** 22 October 2023*

depicting the changes in their amplitude values with respect to time as shown in Fig. 2 and Fig. 3. By taking the intricate technical and graphical attributes into consideration, a comparative analysis of “todi” and “purvi” has been performed. The melodic structure of each thaat is meticulously expounded, encompassing the precise enumeration of swaras, their nomenclature, and their octave placements in accordance with the established conventions of Indian classical music. A meticulous examination of the arohana (ascending) and avarohana (descending) scales for both "todi" and "purvi" is presented, discerning divergences and convergences in their respective swara sequences and identifying any instances of vikrut (altered) swaras. Moreover, the investigation extends to the identification and discussion of noteworthy ragas encapsulating the defining melodic patterns within each thaat. This comparative exploration not only contributes to a comprehensive understanding of "todi" and "purvi" but also facilitates insights into their distinctive musical nuances, enabling a richer appreciation of their role within the Indian classical music tradition.

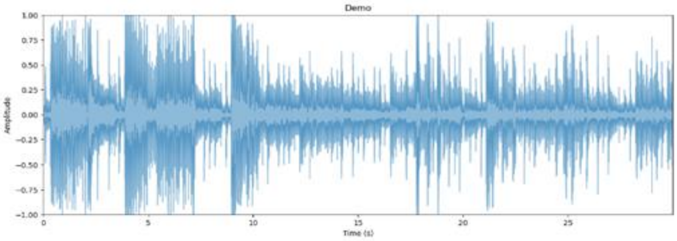

Figure 2. Graphical representation of time varying amplitudes values for todi thaat.

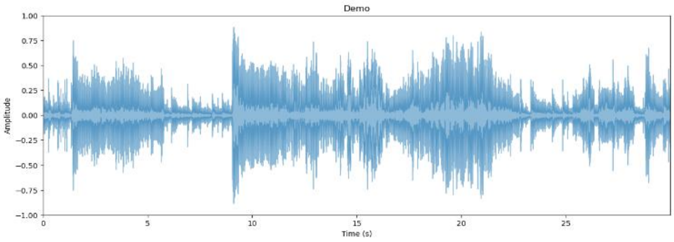

Figure 3. Graphical representation of time varying amplitudes values for purvi thaat.

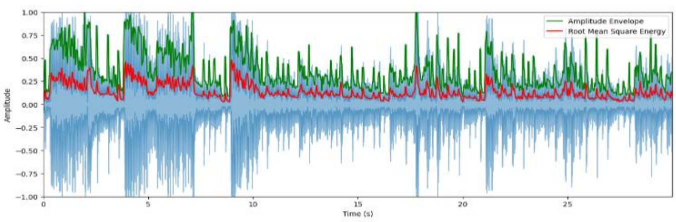

Figure 4. Graphical representation of extraction of temporal features such as amplitude envelope and RMSE on sample audio file.

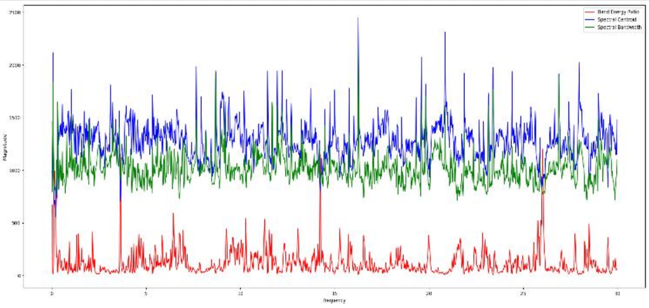

Figure 5. Graphical representation of extraction of spectral features such as spectral centroid and spectral bandwidth on sample audio file.

Time-frequency characteristics play a crucial role in thaat categorization by providing a dynamic representation of how frequencies change over time in music. These characteristics offer valuable insights into the intricate variations in melody, ornamentation, and rhythmic subtleties that are unique to each thaat. The boxplot shown in Fig. 6 represents the range of MFCC feature coefficient for 9 classes. The box plot indicates its contribution in discriminating the 9 classes which plays a major role in the classification. By capturing the evolving spectral content over time, they empower classification models to distinguish the nuanced attributes that define different thaats, significantly improving the accuracy and precision of the classification process. In Fig. 7, a visual representation of how time-frequency characteristics impact a sample audio file is shown.

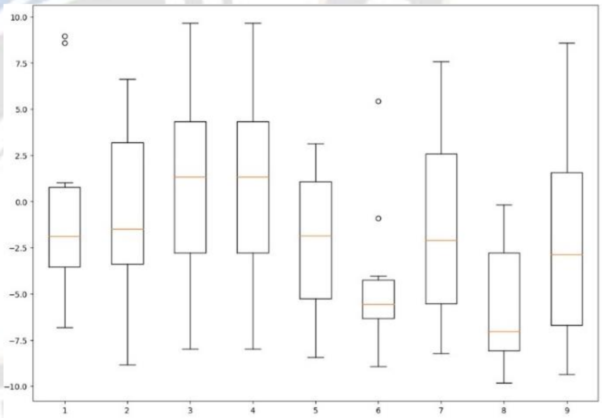

Figure 6. Boxplot representing MFCC feature distribution across each thaat. Numbers on x-axis represent the class of Thaat.

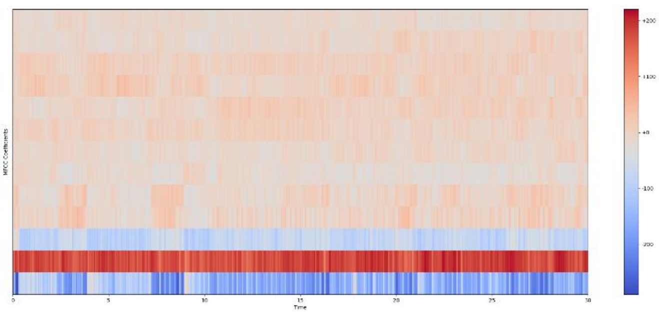

Figure 7. Graphical representation of extraction of time-frequency features on a sample audio file.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 18 August 2023 **Revised:** 10 October 2023 **Accepted:** 22 October 2023*

Some of the significant time-frequency features are as follows:

1) Mel Frequency Cepstral Coefficients: In this study, the primary feature extraction technique is MFCC, a pivotal element in thaat recognition, a fundamental aspect of Indian classical music. MFCC coefficient extraction from audio files entails five critical phases, as illustrated in Fig. 8. These stages encompass Audio Preprocessing, Frame Segmentation, Spectrum Conversion, Frequency Domain Filtering, and Discrete Cosine Transform. The first step involves audio preprocessing, a critical process for enhancing sound quality. Pre-emphasis reduces higher-frequency components, significantly contributing to audio signal refinement. (1) demonstrates the execution of the preprocessing filter, a crucial component of the thaat recognition study.

$$H\,(z) = 1 - bz^{-1} \tag{1}$$

Where b is the slope of the filter. The second phase concentrates on audio segmentation through Frame Blocking and Windowing.

This technique proves highly effective in addressing the edge effect, a common concern in Fourier Transform. In this study, audio is partitioned into ten sections, each with a windowed length of 512. This step ensures manageable portions of audio data for subsequent analysis. Subsequently, the third stage revolves around obtaining the frequency spectrum. The frequency spectrum is obtained by applying the Discrete Fourier Transform (DFT) to each windowed frame. The spectrum resulting from DFT is often extensive. Equation (2) represents the transformation into Fourier space, denoted as X(k), a crucial aspect of the thaat recognition strategy. Following DFT processing, the frequency spectrum often exhibits an extensive range.

$$X\,(k) = \sum_{n=0}^{N-1} x\,(n)\, e^{\frac{-j2\pi nk}{N}}\,; 0 \le k \le N-1 \tag{2}$$

Where N is the number of points, the value of FFT is assumed to be 2048 in this study. To linearize the frequency range and capture pertinent acoustic characteristics, Mel-filter banks are employed. A Mel-filter bank comprises an array of bandpass filters, operating based on the Mel scale as depicted in (3).

$$f_{Mel} = 2595 log_{10}\left(1 + \frac{f}{700}\right) \tag{3}$$

Where f_{Mel} denotes the perceived frequency, and f represents the physical frequency in Hz.

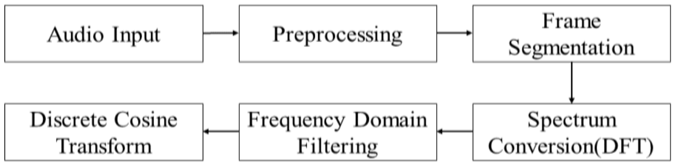

Figure 8. Flowchart representing different processes for extracting MFCC features from audio data.

These filters are then applied to the spectrum. The triangular Mel Weighting filter is multiplied by the spectrum to derive the Mel Spectrum, represented as $s(m)$ as demonstrated in (4).

$$s(m) = \sum_{k=0}^{N-1}[\, |X\,(k)|^2\, H_m(k)\,]\,; 0 \le m \le M-1 \tag{4}$$

Where H_m is weight to k^{th} energy spectrum granting to m^{th} output band, M is total triangular Mel weighting filters. The final phase entails visualizing the Mel Spectrum on a logarithmic scale as illustrated in (5). This process accentuates vital acoustic features essential for distinguishing thaats. Subsequently, the Discrete Cosine Transform (DCT) is executed to extract cepstral coefficients. Cepstral coefficients play a pivotal role in the thaat recognition model, enabling the accurate identification of diverse thaats within Indian classical music.

$$c(n) = \sum_{m=0}^{M-1} log_{10}\big(s(m)\big) \cos\left(\frac{\pi n(m-0.5)}{M}\right) \tag{5}$$

Where C is the number of MFCCs, n = 0, 1, 2, . . . , C − 1, and $c(n)$ are the cepstral coefficients. The value of C is assumed to be 13 in this study. MFCC coefficients play a critical role in identifying thaats in Indian classical music. thaats are characterized by distinct melodic and tonal patterns, attributes effectively captured by MFCCs. These coefficients condense intricate audio information into a succinct representation, suitable for machine learning algorithms. The utilization of MFCCs in the thaat recognition model allows for the identification of subtle variations in tonal patterns and the effective classification and differentiation of various thaats. This approach enriches our understanding and preservation of this culturally rich musical tradition, facilitating the automatic identification and categorization of thaats with exceptional precision.

2) Scalogram: A Scalogram, generated through wavelet transformation, reveals the time-frequency characteristics of a signal. It assists in pinpointing coefficient estimations at specific time-frequency positions, offering a thorough portrayal of the signal. The Scalogram essentially corresponds to the modulus of the multiscale wavelet transformation, enhancing the understanding of the interplay between time and frequency. Due to their spectro-temporal attributes, Scalograms are highly sought-after by neural networks for their ability to map signals effectively while minimizing data

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 18 August 2023 **Revised:** 10 October 2023 **Accepted:** 22 October 2023*

loss. Scalogram gives the insight into the frequency, energy in time which is shown as a function $Ix\ (t, \alpha)$ in (6).

$$Ix(t, \alpha) = |W_x\ (t, \alpha)| = |x + \ \varphi_\alpha\ (t)| \qquad (6)$$

Where $W\_x\ (t, \alpha)$ is the wavelet transform, x is energy, $\varphi_\alpha\ (t)$ is dilated wavelet and $2^\propto$ is frequency. Scalograms provide an intricate view of how the signal's frequency content evolves over time. This detailed representation is crucial for capturing the complex melodic and tonal patterns intrinsic to thaats, thus contributing to precise recognition. Moreover, scalograms facilitate the extraction of precise coefficient estimations at specific time-frequency coordinates. These coefficients contain vital information about the signal's attributes, aiding in the differentiation of various thaats. In summary, Scalograms play a pivotal role in thaat recognition by providing a detailed, high-resolution depiction of a signal's time-frequency attributes.

3) Spectrogram: A Spectrogram serves as a visual representation that illustrates how a signal's strength varies across multiple frequencies throughout its duration within a waveform. In simpler terms, it also reflects the loudness of the signal. It essentially constitutes an intensity plot derived from the magnitude of the Short-Time Fourier Transform (STFT). The STFT is obtained by consecutively performing Fast Fourier Transforms (FFTs) on segments of the data. The process of creating a Spectrogram involves eight primary steps: Pre-emphasis, Frame Blocking, Windowing, Discrete Fourier Transform (DFT), Power Spectrum Density (PSD), Mapping and Normalization, computation of the Short-time Spectrogram, and Linear Superposition. These steps are visually presented in Fig. 9.

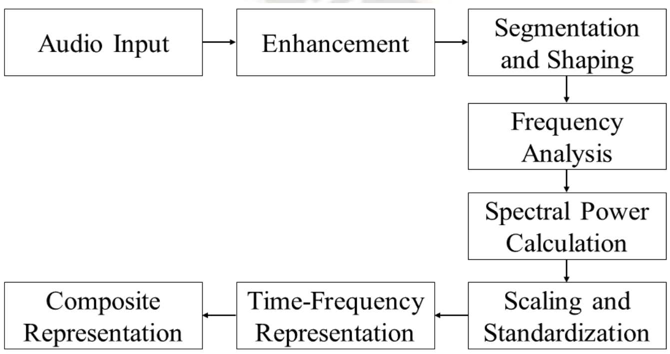

Figure 9. Flowchart representing different processes for extracting spectrogram features from audio

Power Spectral Density$S_x(f)$ of signal $X(t)$ is computed as the Fourier Transform of an autocorrelation function $R_x(\tau)$ as shown in (7).

$$S_x(f) = F\ \{R_x(\tau)\} = \ \int_{-\infty}^{\infty} R_x(\tau) e^{-2j\pi f\tau}\ d\tau \qquad (7)$$

Where $j\ =\ \sqrt(-1)$. Spectrograms play a crucial role in thaat recognition by offering a visual depiction of how audio signals' frequency content changes over time. This visual representation allows for the examination of the unique melodic and tonal patterns specific to various thaats, facilitating their identification and categorization. Spectrograms also extract distinctive features that are valuable for automated recognition systems, making them well-suited for machine learning applications. They provide an objective and quantifiable means of analyzing thaats, reducing subjectivity, and serving as educational aids for musicians and learners, contributing to a deeper comprehension and preservation of the traditions of Indian classical music.

C. Classifier Models

1) Long Term Short Memory-Recurrent Neural Network: To The strategy applied in this classification model employs a Sequential Recurrent Neural Network (RNN) integrated with Long Short-Term Memory (LSTM) cells to categorize thaats in music, utilizing Mel-Frequency Cepstral Coefficients (MFCC) as the input feature. This method encompasses the creation of a step-by-step model that combines LSTM and dense layers. The initial LSTM layer, furnished with 256 units and 'relu' activation, is essential for capturing time-based relationships present in the MFCC data. Subsequently, multiple dense layers with 'relu' activation are introduced to diminish dimensionality and gain insights into abstract patterns. The ultimate dense layer adopts 'softmax' activation, tailored for effective multi-class classification, yielding a probability distribution across the 9 distinct thaats. Model adaptation is facilitated by the fine-tuning of parameters, including the number of LSTM layers, units in each LSTM layer, configurations of dense layers, and the choice of activation functions. The LSTM layer's retention of historical MFCC inputs is pivotal for grasping temporal intricacies in musical content, while the model's architectural adaptability accommodates the complexity inherent in the MFCC data. The application of activation functions and the concluding softmax layer assures the model's output of class probabilities, rendering it highly suitable for multi-class classification. This, in turn, facilitates the precise identification of diverse thaats in music through the lens of their MFCC representations. Fig. 10 illustrates the sequential steps in an RNN-LSTM process.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 18 August 2023 **Revised:** 10 October 2023 **Accepted:** 22 October 2023*

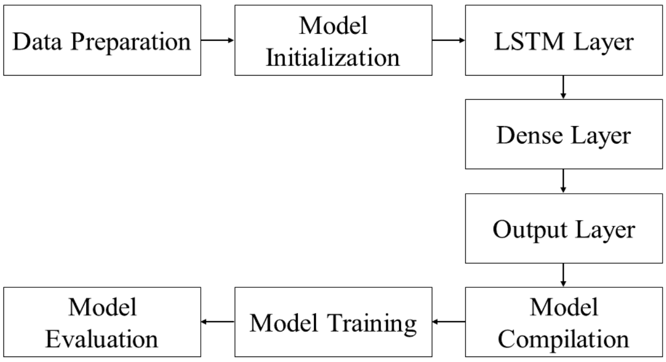

Figure 10. Flowchart representing different steps involved in RNN-LSTM

The model is compiled with an optimizer that utilizes the Adam optimization algorithm and a loss function, which calculates the difference between the predicted class probabilities and the true class labels. The goal is to minimize this difference during the training process as described in (8). The model is trained for 100 epochs with a batch size of 128.

$$J(w) = -\frac{1}{N}\sum_{i=1}^{N}[y_i\, log(x_i) + (1-y_i)\, log(1-\ x_i)] \quad (8)$$

Where J(w) is the loss value, w is the weight of the neural network, y_i is true label, x_i is the predicted label and N is the number of classes.

2) Support Vector Machine: The Support Vector Machine (SVM) is a robust machine learning algorithm employed for classification tasks, known for its efficacy in high-dimensional feature spaces and its ability to handle both linear and non-linear separations. SVM operates by determining a hyperplane that maximizes the margin between different classes, with the margin defined as the perpendicular distance between the hyperplane and the nearest data points, known as support vectors. Equation (9) represents the objective function to minimize the norm (magnitude) of the weight vector (w) while subject to the constraint that data points are correctly classified.

$$min_{w,b}\frac{1}{2}\|w\|^2 \quad (9)$$

The constraint, on the other hand, defines the conditions that the SVM solution must satisfy as shown in (10).

$$y_i(w.x_i + b) \geq 1\ for\ i = 1,2,3 \ldots .N \quad (10)$$

Here, w is the weight vector, b is the bias, x_i are data points, and y_i is the class label (+1 or -1). The objective function seeks to minimize the norm of the weight vector while ensuring that the data points are correctly classified, satisfying the inequality constraint. SVM strives to achieve the maximum margin by selecting the hyperplane that optimally separates the data points of different classes. This margin maximization results in robust decision boundaries, rendering SVM a valuable tool for the accurate classification of music based on thaats, effectively capturing the unique melodic characteristics intrinsic to Indian classical music. Additionally, Principal Component Analysis (PCA) is harnessed for the purpose of dimensionality reduction. Also, the SVM model implemented in this study employs a linear kernel and sets the regularization parameter (C) to 1.0. Following this, the model is trained using the provided training dataset (X_train and y_train), yielding predictions on the test dataset that are subsequently stored in the predictions variable. In the context of classifying different thaats in music, SVM can learn the unique patterns and characteristics of each thaat, effectively separating them based on the extracted features. SVM's robustness and ability to handle high-dimensional data make it a valuable choice for this application. However, the choice of appropriate features and kernel functions is critical to the success of the SVM classifier in this domain.

3) Hidden Markov Model: Hidden Markov Models (HMMs) are deployed in the categorization of thaats within the realm of Indian classical music, signifying distinct musical scales or modes. In this context, HMM states correspond to diverse thaats, while observations emanate from musical pitch values, often denoted as discrete symbols or intervals. The workflow entails the delineation of states and observations, initializing the HMM with probabilities governing state transitions and emissions, executing the forward and backward algorithms to compute probabilities of observation sequences, and employing the Baum-Welch algorithm for parameter estimation during model training. To classify a new musical composition, the HMM assesses the probability of the observation sequence for each thaat, ultimately designating the one with the highest probability as the classification result. Supplementary post-processing techniques may be enlisted to further refine the outcomes. The representation of this process in a flowchart involves the illustration of state transitions, the application of forward and backward algorithms, and the classification stage, which often necessitates the use of specialized software tools tailored for HMM modelling as shown in Fig. 11.

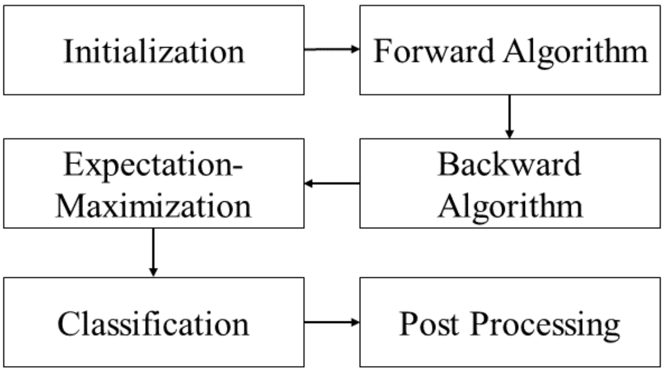

Figure 11. Flowchart representing different steps involved in Hidden Markov Model

Hidden Markov Models (HMMs) play a critical role in the classification of thaats in the domain of Indian classical music, representing distinct musical scales or modes. A foundational equation that underpins this classification process is shown in (11).

$$P(O|\gamma) = \sum_{i=1}^{N}\left(\pi_i . b_i(O_1) . \prod_{t=2}^{T}\left(a_{i,j} . b_j(O_t)\right)\right) \quad (11)$$

In this equation, $P(O|\gamma)$ computes the likelihood of observing a sequence of pitch values (O) given the HMM model (γ), N denotes the number of states in the HMM, each corresponding to a specific thaat, π_i represents the initial state probabilities, indicating the likelihood of commencing with state i, which corresponds to a particular "thaat", O_t signifies the observations, specifically pitch values, at time t, $a_{i,j}$ reflects the state transition probabilities, encapsulating the likelihood of transitioning from state i to state j within the musical composition, $b_j(O_t)$ represents the emission probabilities, elucidating the probability of observing pitch value O_t when in state i. This equation serves as the computational core for assessing the likelihood of different thaats based on the observed pitch values. The thaat with the highest likelihood is designated as the classification outcome. This process is augmented by the application of forward and backward algorithms, which facilitate probability calculations, and the potential incorporation of post-processing methodologies to refine the classification results. Visual representations, such as flowcharts, can elucidate state transitions and algorithmic steps within the HMM framework, often necessitating specialized software tools tailored for HMM modelling.

D. Emotion-Based Classification

The advanced machine learning models, particularly the Recurrent Neural Networks with Long Short-Term Memory (RNN-LSTM), are employed to predict emotional responses derived from the classification of music based on thaats. Utilizing a comprehensive array of audio features, including root mean square energy (RMSE), spectral centroid, and Mel-frequency cepstral coefficients (MFCC), in combination with the recognized thaats, the model is trained to establish a connection between musical attributes and emotional expressions. By identifying subtle tonal fluctuations and rhythmic intricacies, the RNN-LSTM model effectively discerns the emotional nuances linked with each identified thaats category. This intricate process involves a sophisticated mapping mechanism that associates the musical components with specific emotional states such as happiness, sadness, tranquility, and enthusiasm, among others. The integration of this predictive framework not only facilitates accurate music classification based on thaats but also enables the precise anticipation of emotional states. This predictive model demonstrates significant potential for various applications, including the development of music recommendation systems, the creation of mood-based playlists, and the implementation of tailored therapeutic strategies aimed at managing emotions and promoting well-being.

IV. RESULTS AND DISCUSSIONS

Before The feature set comprises a total of 28 features, encompassing both time-domain and spectral characteristics. To achieve audio classification into distinct thaat labels, a deep learning model was employed. This model leveraged a diverse array of time-domain and spectral features, including chromal shift density, spectral centroid, spectral bandwidth, rolloff, zero-crossing rate, and MFCC coefficients spanning from 1 to 20. On a test dataset of audio samples, the model exhibited an impressive overall accuracy of 86%, demonstrating robust performance across all thaat labels, with accuracy rates ranging from 75% to 89%. The architectural configuration of the deep learning model featured six layers, commencing with an LSTM layer comprising 256 units, followed by five dense layers with 256, 128, 64, 64, and 10 units, respectively. This architecture facilitated the model's ability to capture long-term dependencies within the audio data through the LSTM layer while extracting higher-level features and accurately predicting the thaat label via the subsequent dense layers. The thaats todi and purvi have a similar set of features. Due to this and the inavaibility of the data of thaat purvi, the precision of thaat purvi seems to be less apealing. However, this does not pamper the other results in the experiment. These findings are represented in Fig. 12 which underscore the efficacy of employing a sequential deep learning architecture for precise thaat recognition in audio signals.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 10
***Article Received:** 18 August 2023 **Revised:** 10 October 2023 **Accepted:** 22 October 2023*

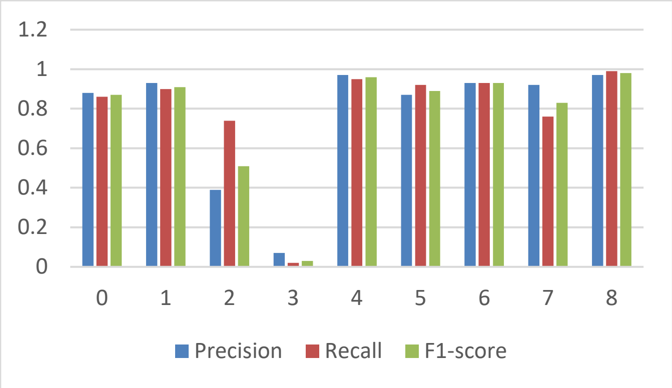

Figure 12. Representation of evaluation parameters for RNN-LSTM model

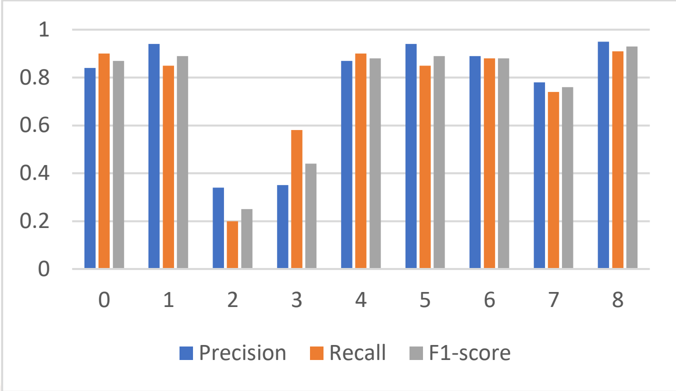

Figure 13. Representation of evaluation parameters for SVM model

Within the context of the provided SVM classifier, a comprehensive analysis of precision, recall, and F1-scores has been undertaken for individual classes. Precision, which gauges the precision of positive predictions, exhibited variance among classes, with the highest values seen in classes 1, 4, 5, and 8, denoting the precision of positive predictions, while classes 2 and 3 displayed relatively diminished precision, indicating a heightened occurrence of false positives. Conversely, recall, representing the model's capacity to correctly identify real positive instances, revealed disparities across classes, with classes 0, 4, 5, 6, and 8 demonstrating superior recall values, implying effective recognition of true positives, whereas classes 2 and 7 displayed diminished recall, signifying the model's incapacity to detect a significant number of positive instances. The F1-score, serving as a harmonizing metric between precision and recall, reflected these patterns, with higher F1-scores in classes 0, 1, 4, 5, and 8, highlighting a harmonious balance between precision and recall, and lower F1-scores in classes 2 and 3, signaling a compromise between precision and recall. Additionally, the weighted average F1-score, offering an encompassing performance measure, stood at 0.75, indicating a moderate overall model performance, while the macro average F1-score, according equal importance to each class, was calculated at 0.76. This analysis underscores the SVM classifier's diverse efficacy across distinct classes as shown in Fig. 13, necessitating potential enhancements, notably within classes 2 and 3, to attain enhanced classification efficacy. Furthermore, it is essential to recognize that SVM, as a traditional machine learning method, excels in structured data with explicit features, whereas recurrent neural networks with Long Short-Term Memory (RNN-LSTM) models, primarily tailored for sequential data, warrant consideration in scenarios involving temporal dependencies and expansive datasets, albeit with heightened computational demands. The selection between SVM and RNN-LSTM hinges on data characteristics and task-specific needs, with SVM being suited for feature-rich structured datasets and RNN-LSTM catering to sequential data replete with temporal associations.

In the current research, the categorization of music using the thaats system, a traditional Indian music classification technique, has been effectively carried out employing the HMM classifier also. The HMM classifier was advantageous due to its ability to model time-based connections, manage variable-length sequences, and handle probabilistic data representations. In comparison to the SVM and RNN-LSTM methods, the HMM approach had limitations in capturing complex sequential patterns and identifying intricate relationships within the music dataset. Consequently, the analysis revealed that the HMM classifier performed less effectively than both the RNN-LSTM and SVM techniques for this specific task of music taxonomy classification.

V. CONCLUSION AND FUTURE SCOPE

Following the in-depth analysis in the study, it becomes apparent that the categorization of music according to thaats, rather than the traditional raga-centric method, presents a fresh and effective way of comprehending the complexities of Indian classical music. A diverse range of audio features, including measures like amplitude envelope and spectral centroid, have facilitated a detailed understanding of thaats, bringing their distinct musical qualities to the forefront. The comparison of various classification models - RNN-LSTM, SVM, and HMM - revealed the superior performance of RNN-LSTM in accurately categorizing musical compositions based on their respective thaats, achieving an impressive accuracy of 85%. This highlights the potential of deep learning methods, particularly RNN-LSTM, in capturing the intricate patterns inherent in the classification of Indian classical music based on thaats. Moreover, the study not only focuses on music classification but also introduces a novel element that predicts emotional responses linked with the identified thaats. By incorporating this feature, the research enables the classification of music into different genres based on mood, adding value to the study and offering potential pathways for the development of personalized music recommendations and mood-based playlists.

In the future, these findings could lay the groundwork for advanced music recommendation algorithms, enriching user experiences by providing personalized music recommendations based on emotional inclinations. Additionally, the study's emphasis on understanding the interplay between thaats and emotions opens possibilities for cross-cultural music analysis, enhancing the understanding of emotional responses to diverse musical compositions across different cultural contexts. Furthermore, integrating this research into the realm of music therapy shows promise in developing tailored interventions that cater to specific emotional needs. This underscores the broader implications of the research for psychology, neuroscience, and overall well-being. Additionally, the insights gained from this research contribute to the preservation and documentation of India's rich cultural musical heritage, ensuring its appreciation and continuation for future generations.

REFERENCES

[1] A. K. Sharma et al., "Classification of Indian classical music with time-series matching deep learning approach," in IEEE Access, vol. 9, pp. 102041-102052, 2021, doi: 10.1109/ACCESS.2021.3093911Bejan, A. (2015). Constructal thermodynamics. Constructal Law & Second Law Conference, Parma, pp. S1-S8.

[2] S. John, M. S. Sinith, S. R. S and L. P. P, "Classification of Indian classical Carnatic music based on raga using deep learning," 2020 IEEE Recent Advances in Intelligent Computational Systems (RAICS), Thiruvananthapuram, India, 2020, pp. 110-113, doi: 10.1109/RAICS51191.2020.9332482.

[3] A. Vasudevan and H. Ramasangu, "A hybrid cluster-classifier model for Carnatic raga classification," 2021 IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT), Bangalore, India, 2021, pp. 1-6, doi: 10.1109/CONECCT52877.2021.9622669.

[4] V. Rajadnya and K. Joshi, "Raga classification based on MFCC and variants," 2021 IEEE 2nd International Conference on Technology, Engineering, Management for Societal impact using Marketing, Entrepreneurship and Talent (TEMSMET), Pune, India, 2021, pp. 1-6, doi: 10.1109/TEMSMET53515.2021.9768314.

[5] P. Dighe, P. Agrawal, H. Karnick, S. Thota and B. Raj, "Scale independent raga identification using chromagram patterns and swara based features," 2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), San Jose, CA, USA, 2013, pp. 1-4, doi: 10.1109/ICMEW.2013.6618238.

[6] A. A. Bidkar, R. S. DeshPande and Y. H. Dandawate, "A novel approach for selection of features for North Indian classical raga recognition of instrumental music," 2018 International Conference On Advances in Communication and Computing Technology (ICACCT), Sangamner, India, 2018, pp. 499-503, doi: 10.1109/ICACCT.2018.8529392.

[7] R. K. Nawasalkar, P. K. Butey, S. G. Deshpande and V. M. Thakare, "EEG based stress recognition system based on Indian classical music," 2015 International Conference on Advances in Computer Engineering and Applications, Ghaziabad, India, 2015, pp. 936-939, doi: 10.1109/ICACEA.2015.7164840.

[8] M. Bargaje, "Emotion recognition and emotion based classification of audio using genetic algorithm - an optimized approach," 2015 International Conference on Industrial Instrumentation and Control (ICIC), Pune, India, 2015, pp. 562-567, doi: 10.1109/IIC.2015.7150805.

[9] L. Moysis et al., "Music deep learning: deep learning methods for music signal processing—a review of the state-of-the-art," in IEEE Access, vol. 11, pp. 17031-17052, 2023, doi: 10.1109/ACCESS.2023.3244620

[10] V. Tejaswini Priyanka, Y. Reshma Reddy, D. Vajja, G. Ramesh and S. Gomathy, "A novel emotion based music recommendation system using CNN," 2023 7th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2023, pp. 592-596, doi: 10.1109/ICICCS56967.2023.10142330.

[11] A. R. Chhetri, K. Kumar, M. P. Muthyala, S. M R and R. A. Bangalore, "Carnatic music identification of Melakarta ragas through machine and deep learning using audio signal processing," 2023 4th International Conference for Emerging Technology (INCET), Belgaum, India, 2023, pp. 1-5, doi: 10.1109/INCET57972.2023.10170568.

[12] Suma, S.M., Koolagudi, S.G. (2015). Raga classification for Carnatic music. In: Mandal, J., Satapathy, S., Kumar Sanyal, M., Sarkar, P., Mukhopadhyay, A. (eds) Information Systems Design and Intelligent Applications. Advances in Intelligent Systems and Computing, vol 339. Springer, New Delhi. doi:10.1007/978-81-322-2250-7_86.

[13] H. Purwins, B. Li, T. Virtanen, J. Schlüter, S. -Y. Chang and T. Sainath, "Deep learning for audio signal processing," in IEEE Journal of Selected Topics in Signal Processing, vol. 13, no. 2, pp. 206-219, May 2019, doi: 10.1109/JSTSP.2019.2908700.

[14] Acharya, S., Devalla, V., Amitesh, O., Ashwini (2021). Analytical comparison of classification models for raga identification in Carnatic classical Audio. In: Biswas, A., Wennekes, E., Hong, TP., Wieczorkowska, A. (eds) Advances in Speech and Music Technology. Advances in Intelligent Systems and Computing, vol 1320. Springer, Singapore. doi:10.1007/978-981-33-6881-1_18.

[15] Deng, J.J., Leung, C.H.C. (2012). Music emotion retrieval based on acoustic features. In: Hu, W. (eds) Advances in Electric and Electronics. Lecture Notes in Electrical Engineering, vol 155. Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-28744-2_22.

[16] Shetty, S., Achary, K.K., Hegde, S. (2012). Raga identification in Carnatic music using Hidden Markov Model technique. In: Krishna, P.V., Babu, M.R., Ariwa, E. (eds) Global Trends in Information Systems and Software Applications. ObCom 2011. Communications in Computer and Information Science, vol 270. Springer, Berlin, Heidelberg. doi:10.1007/978-3-642-29216-3_45.

[17] Shah, D.P., Jagtap, N.M., Talekar, P.T., Gawande, K. (2021). Raga recognition in Indian classical music using deep learning. In: Romero, J., Martins, T., Rodríguez-Fernández, N. (eds) Artificial Intelligence in Music, Sound, Art and Design.

EvoMUSART 2021. Lecture Notes in Computer Science(), vol 12693. Springer, Cham. doi:10.1007/978-3-030-72914-1_17.

[18] Pendyala, Vishnu S., Nupur Yadav, Chetan Kulkarni, and Lokesh Vadlamudi. "Towards building a deep learning based automated indian classical music tutor for the masses." Systems and Soft Computing 4 (2022): 200042. doi:10.1016/j.sasc.2022.200042.

[19] Jiang, Jinwen. "Signal feature extraction of music melody based on deep learning." Traitement du Signal 39, no. 6 (2022): 2203.

[20] Minor, Kelvin A., and Iman H. Kartowisastro. "Automatic music transcription using fourier transform for monophonic and polyphonic audio file." Traitement du Signal 39 no. 4 (2022): 629-635