

Multimodal Deep Learning Architecture for Hindustani Raga Classification

¹ Stella PASCHALIDOU and ² Ioanna MILIARESIS

¹ Hellenic Mediterranean University, E. Daskalaki, Rethymno, Greece
Tel.: +30 2831021918

² Ionian University, Plateia Tsirigoti 7, Corfu 49100, Greece
Tel.: +30 2661087860

E-mail: pashalidou@hmu.gr, imiliaresi@ionio.gr

Received: 27 May 2023 Accepted: 23 June 2023 Published: 26 June 2023

Abstract: In this paper, our key aspect is the design of a deep learning architecture for the classification of Hindustani (classical North Indian music) ragas (music modes). In an attempt to address this task, we propose a modular deep learning architecture designed to process data from two modalities, comprising audio recordings and metadata. Our bipolar classifier utilizes convolutional and feed forward neural networks and incorporates spectral information of audio data and metadata descriptors tailored to the peculiar melodic characteristics of Hindustani music. In specific, audio recordings as well as manually annotated and automatically extracted metadata were utilized for audio samples of both Hindustani improvisations and compositions available in the Saraga open dataset of Indian art music. Experiments are conducted on two Hindustani ragas, namely Yaman and Bhairavi. Results indicate that the integration of multimodal data increases the classification accuracy of the classifier in comparison to simply using audio features. Additionally, for the specific task of raga classification the use of the swaragram feature, which is customized for Hindustani music, outperforms the effectiveness of audio features that are commonly used in Eurocentric music genres.

Keywords: Hindustani raga identification, Deep learning, Convolutional neural networks, Multimodal.

1. Introduction

This work focuses on machine listening identification of ragas in Hindustani music, which is the north Indian style of classical Indian music. Tzanetakis was the one to highlight the importance in expanding data-driven research to cover non-Eurocentric music traditions too [1], aiming—among other things—at improving the current state of the art in MIR (Music Information Retrieval).

Hindustani music is a so-called oral music tradition, which means that it is based on the transmission of knowledge from teacher to student through direct imitation rather than written instructions of music notation. It relies heavily on improvisation, which is rule-based and conforms to the so-called raga system. Raga refers to a concept of

melodic mode lying between scale and tune [2], and it provides a tonal framework for composition and improvisation. A mode is a collection of musical notes (the degrees of a scale) coupled with rules about how to treat them. In other words, in contrast to the way a scale is performed and perceived in Eurocentric music genres, a raga cannot be merely understood based on the set of used notes; given the same set, a different raga could be performed, the differentiation of which relies on other musical features.

For a concrete explanation of the term, following readings are suggested: [3-5]. However, defining what exactly constitutes a raga is not a straightforward task [6]. Loosely, one can say that a raga refers to a group of important features, including the scale (discrete pitches), the notes' sequence, their hierarchy, relative duration, importance, micro-tonal intonation and

ornamentation, types of melodic movement, specific phrases, mood, etc. To further perplex matters, there are other attributes that extend beyond purely musical features and contribute to the rendering of different ragas by performers—and their identification by listeners—that would ideally also need to be taken into account. These may include elements of mood and aesthetics, such as distinct mental images that adhere to specific ragas as depicted in the so-called Ragamala miniature paintings, but have not been considered in our work.

For these reasons, mode classification tasks, such as those of Hindustani ragas, fundamentally differ to those applied to scale or genre identification tasks in Eurocentric music genres. The following offers a short introduction to Hindustani music, which highlights the peculiar characteristics of the music genre and justifies the selection of extracted features in our implementation.

Hindustani music follows a movable tonic system, which means that the tonic (Do) is defined by the most comfortable pitch area of the singer; therefore, the pitches of scale-degrees (swaras) need to be computationally deduced based on the (either manually or automatically) extracted tonic frequency. The soloist is invariably accompanied by a constantly sounding drone instrument (called tanpura), which is used to create a sonic canvas as a tonal reference, from which the pitch of the tonic can be both perceived and computationally extracted. Manual annotations for the tonic frequency provided by the used corpus have been considered in this paper.

Three registers (octaves) are commonly used in performance, the low, the middle and the high. There are seven basic scale degrees, from which the first (tonic) and 5th are never altered. The 2nd, 3rd, 6th and 7th degrees can be lowered by a semitone and the 4th degree can be raised by a semitone. All degrees of a raga scale need to be computed relative to the tonic frequency. There is a great sense of pitch accuracy by Hindustani musicians. However, notes are not considered exactly fixed in pitch and—in contrast to Western styles of music—the exact position of the semitones, excluding the natural 4th, natural 5th and octave, can vary slightly according to the specific raga [3]. Another peculiar characteristic of ragas is the intricate micro-tonal embellishment (shrutis) of scale-degrees as well as the abundance of smooth melodic glides by which musicians often connect discrete notes; the space between notes is often considered more important than the discrete notes themselves [7]. It is for this reason, that pitch identification in Hindustani music is an especially challenging task [8], and a fundamental aspect that differentiates computational approaches of note extraction in Hindustani music from those used for Eurocentric music genres.

A raga can be further characterized by its overall melodic outline in terms of ascent and descent, whereby different degrees of the scale may be used or omitted in ascending versus descending order according to the raga. In most cases, the tonic and the

5th function as strong anchor points, while two notes, unique to individual ragas, gain attention and special treatment and contribute to the special mood of each raga. Pitch distribution plots of a Hindustani improvisation reflect the tonal hierarchy of a raga.

Characteristic motifs, key phrases and ornaments constitute prominent aspects of a raga, and therefore also strong cues (the ‘signature’) for raga identification. It is for this reason, that we take advantage of the manual annotations of raga key phrases in the used corpus, which are described as a sequence of notes. However, one needs to emphasize here that, while the problem of automatic melodic phrase detection has been attempted for Western music, Hindustani music presents additional challenges due to the conception and treatment of notes in a pitch continuum [9]. In specific, phrases and motifs cannot be simply recognized in terms of discrete note sequences, as is the case in Western styles of music, but would rather need to be considered as melodic shapes as represented by computed pitch contours [10]. For a more extensive contribution to the discussion of pertinent features and proposed computational methods of feature extraction the reader can refer to [11] and [12].

To address the complexity of raga identification, in this paper we formulate a methodology for machine listening raga classification, which is based on customized features for the specific task utilizing deep learning architectures. The current paper extends our previous work [13] by relying on an audio feature that is tailor-made for Hindustani music—called swaragram—which have been recently introduced in research literature.

Raga identification methods are expected to find application in music education, aiding the oral tradition of music training, as well as in Music Information Retrieval tasks, such as the automatic labelling of unknown audio material and hence the development of automatically annotated repositories of Hindustani music.

In this study we focus on two ragas of Hindustani music, namely Yaman and Bhairavi, with each having the tone material that is illustrated in Table 1.

Table 1. Basic information for Yaman and Bhairavi. Upper case letters denote the higher (sharper) pitch alternative, while lower case letters denote the lower pitch (flatter) alternative of the note.

Raga	Tone material	Characteristic phrases
Yaman	S R G M P D N	MDNR\DS NRGMP\RG- RGM/RMG GMDNDP
Bhairavi	S r g m P d n	Sg-mPmgmrS gmPdP mg-SrS

What follows in this paper is organized in five sections. The next section offers an overview and discussion of related work. The third section describes the data corpus where experiments were conducted, and details the proposed methodology, mostly focusing on the feature extraction stage and the network architecture. All experiments and results are reported in the fourth section, leading to the concluding part of the work.

2. Related Work

Automatic raga identification is a challenging task in music information retrieval research due to complexity and diversity of Indian music. The lack of a standardized notation system and the variability in performance by different musicians between different ragas increases the complexity of their accurate identification. However, recent studies report progress in developing architectures for automatic raga identification with the usage of machine learning and deep learning techniques. Most studies in this field focus on Carnatic (south Indian classical) music ([14-22]) and less are reported on Hindustani ([23-29]), while some papers refer to both genres ([30-34, 8, 35-37]). There is still significant room for improvement in this field, mainly in the Hindustani genre and further research is needed to develop robust algorithms; this is where the current paper targets. Conventional classifiers have been used in Carnatic and Hindustani music for this task and have reported promising results. In [8] k-Nearest Neighborhood and Support vector machine classifiers have been used. Gaussian Mixture Model based Hidden Markov Models (HMMs), discrete HMMs and classification trees have been also utilized in [38].

Recent studies incorporate machine learning techniques and involve various neural network architectures. Convolutional neural networks are reported as well-suited architectures for raga identification tasks in ([23, 15, 24, 22]). Recurrent neural networks and Long-short-term neural networks were utilized in [14] and [26] respectively. Transfer learning aggregated methods [39] and supervised learning techniques ([40, 22]) have also reported good results.

The aforementioned architectures have been tested in conjunction with the development of effective feature extraction techniques to improve accuracy and robustness. Several features have been defined, conveying information about the identity, mood, and characteristics of a raga. Patterns of melodic phrases were utilized in [10] using vector space modeling.

Audio features conveying temporal and spectral information are considered to be the most important and provide good results. Various features extracted from audio recordings have been tested, mainly Mel-spectrogram, Mel-frequency Cepstrum Coefficients (MFCC), chromagram, modgdgram, pitch and timbre features ([32, 26, 41, 16, 34]). The specific features are aimed at Eurocentric music genres, and in a way don't

successfully capture the melodic structure of Indian classical music. In a recently study [17], an alternative audio feature (swaragram) that is tailored to the distinct characteristics of Hindustani music was utilized by a residual neural network (ResNets) for raga identification tasks, and it has reported an accuracy of 79.9 % for the Carnatic and 90.4 % for the Hindustani music genre.

In this work, we define a two-class raga classification task on recordings of Hindustani music, namely Yaman and Bhairavi. The specific problem has been addressed in [25], where experiments were carried out on 341 web-extracted audio clips, 60sec-long each, for both instrumental and vocal music, 194 of which in Yaman and 147 in Bhairavi. In the proposed method conventional classifiers, K-Nearest Neighbor (KNN) and Support vector machine (SVM) classifiers were used, achieving a 92 % classification accuracy.

In our method, a deep learning architecture consisting of convolutional and fully connected networks is utilized to process data of two modalities; audio-based features and metadata. The proposed performance optimization scheme processes audio signals as an image representation of the swaragram Indian music-specific audio feature. The proposed classifier achieves a classification accuracy of 98.22%, thus outperforming the state-of-the-art score in the specific classification task.

3. Material and Method

3.1. Data Corpus

The most widely used databases of classical Indian music are the CompMusic Art Indian music dataset—and its subset, Dunya Corpora, for Music Information Research in Indian Art Music [33]—and the Saraga music research corpus of Indian Art Music [42]. Notwithstanding the mp3-quality of the Saraga music research corpus, it was preferred, as it served the scope of this paper in raising the raga classification accuracy by incorporating metadata, which was possible because it includes additional information other than audio.

A Hindustani performance involves a number of sections, which can be different according to the specific sub-genre, and may include a non-metrical melodic improvisational introduction, rhythmic improvising sections and one or more compositions. All sections of a single performance adhere to the same raga and were included in our implementation.

Hindustani music is monophonic, which means that a performance will typically entail the melodic development by a single performer or the accompaniment by another melodic instrument in unison with the soloist (e.g. by a harmonium, which will repeat or imitate phrases of the soloist's melodic line with a slight delay). It can be performed either vocally or by an instrument, and in the former case a vocalist can employ either lyrics or non-lexical

syllables. The chosen corpus includes all above-mentioned cases.

In specific, the corpus comprises 107 multitrack (rarely available in Music Information Retrieval) audio recordings of 68 different ragas by 11 singers of Hindustani music. The recordings are accompanied by editorial metadata and time-aligned structural, melodic, and rhythmic (the latter not used for our purposes) annotations, which are either extracted automatically or created manually by expert listeners.

The editorial metadata includes foremost the raga name extracted from the cover-art, manual annotations by Classical Indian Music experts (such as start/end time-stamps of individual sections, section name and melodic transitions, with time-stamped melodic phrase annotations being flagged according to raga-importance as characteristic of the raga or not), and automatically extracted annotations (such as the tonic and the sequence of time-stamped pitches for the lead musician). From the multi-track recordings, the track corresponding to the lead vocalist was only used for the purposes of this work 44100 Hz sampling rate, 16 bit PCM.

For our experiments, we use a subset of the entire corpus, specifically targeting two ragas of interest, namely Yaman and Bhairavi. The subset includes all three predominant classical forms of Hindustani music, namely Dhrupad, Khayal, as well as the rather semi-classical form of Thumri. That results in an unbalanced dataset with Bhairavi being the minority class. Table 2 provides an overview of information concerning the data structure of the Saraga subset which was used for our experiments.

Table 2. Dataset description.

Saraga music research corpus of Indian Art Music 107 multitrack audio recordings of 68 different ragas by 11 singers of Hindustani vocal music	
Mp3 multi-track recordings only the track corresponding to the lead vocalist	Editorial metadata and time-aligned structural, melodic and rhythmic annotations
Training-Testing dataset Two classes	
Yaman 361 samples	Bhairavi 202 samples

3.2. Feature Extraction

At the feature extraction stage, we construct a complex input vector with audio-based features and metadata descriptors (automatically extracted as well as manually annotated in the Saraga music research corpus), which is fed into the classifier.

Given that raga identification is predominantly substantiated upon the melodic structure, as opposed

to the strongly harmony-based structure of Eurocentric music genres, the information that was considered to be most relevant to the task of raga classification among all available meta-data were the c-tonic (automatic extraction), pitch values (automatic extraction) and characteristic melodic phrases (manual annotations). Metadata-drawn descriptors are part of the input vector. All numerical parameters are represented as a number and all categorical features are transformed to numerical by mapping the state “presence” to the 0 numerical value and “non presence” to 1.

To optimize our system’s performance, we improve upon our previous study by exploiting the swaragram audio feature, which is tailored to the needs of classical Indian music. Mel-scale based features and swaragram have showed high robustness across datasets of various deep learning models for music genre classification tasks [35, 43] and have been favored in this paper for the raga classification task too.

In music information retrieval, the analysis of the chromagram is often selected. A chromagram is a visual representation of energies in the 12 semitones (or chromas) of the musical octave. However, chromagram lacks information about the pitch of the tonic, which is fundamental for music genres that are based on a movable tonic system. Hence, it also lacks information about the pitch class distribution (the relative pitch information of all scale-degrees in relation to the tonic), as well as the micro-tonal embellishment of scale-degrees in Hindustani music. For this reason, the swaragram (SGram), which is a shruti-based chromagram, is considered as better suited for Hindustani music and raga classification tasks [43].

To serve one of the key aspects of our research—that is the comparison on the processing of MFCC and swaragram features—we also calculate MFCC coefficients from the raw audio signal. The extraction process for Mel frequency cepstrum coefficients and the swaragram is illustrated in Fig. 1.

At the first step, all audio recordings are pre-processed. The audio file format is converted from mp3 to wav and all signals are normalized and re-sampled to 44100 Hz. The initial recordings are segmented into clips of 3 seconds. Recordings for both classes (Yaman and Bhairavi) vary heavily in duration, ranging between 1081.678 sec and 718.915 sec and as a result, 361 and 202 audio clips are produced for each class respectively.

To calculate MFCC features, all audio clips are parsed with a 40 ms long moving window with a hop size of 20 ms. A standard zero padding procedure is applied. At each frame, the Discrete Fourier Transform (DFT) is computed and it is given as input to a mel-filter-bank, with each mel-filter performing a weighted sum of the magnitude of the DFT coefficients that lie inside its frequency range. The logarithm of each filterbank output and the discrete cosine transform of the logarithms is calculated. From the set of Mel frequency cepstrum coefficients

(MFCCs), the first coefficient is discarded (as recommended in bibliography), while the following 12 MFCCs are included and the rest are discarded too.

In order to extract the swaragram, the pitch of the tonic—for each specific audio clip from the corpus—is first computed, from which all remaining scale-degree pitches are also extracted. Then, a logarithmic pooling is performed followed by a 12 note (swara) binning. In parallel, a Short-Term Fourier Transform (STFT) and the log-scale frequency Spectrogram are also calculated.

To calculate the swaragram, we follow the algorithmic implementation that is described in [44], with reference pitch set to 69, reference frequency value set to 440 Hz, sample rate set to 22050 Hz and with the number of fft points set to 2048. This results in a 2-D one channel image representation of the audio sample with dimensions [198, 130, 1]. Two representative images of swaragrams for Yaman (a,b) and Bhairavi (c,d) respectively are displayed in contrast in the following images of Fig. 2.

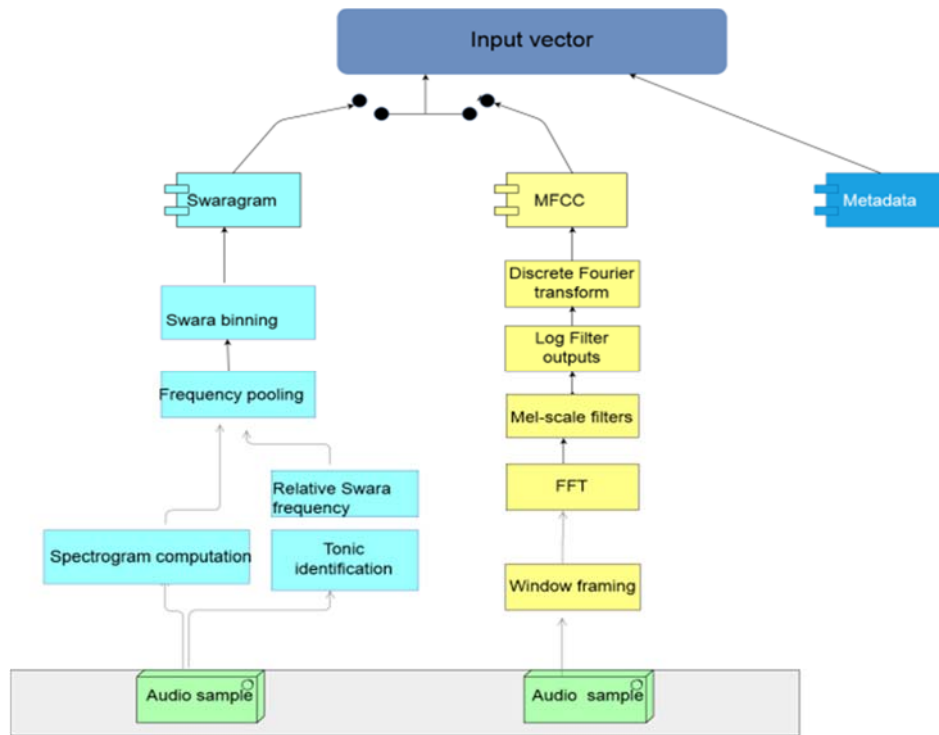


Fig.1. Illustration of the input feature vector composition with swaragram or MFCC feature extraction process.

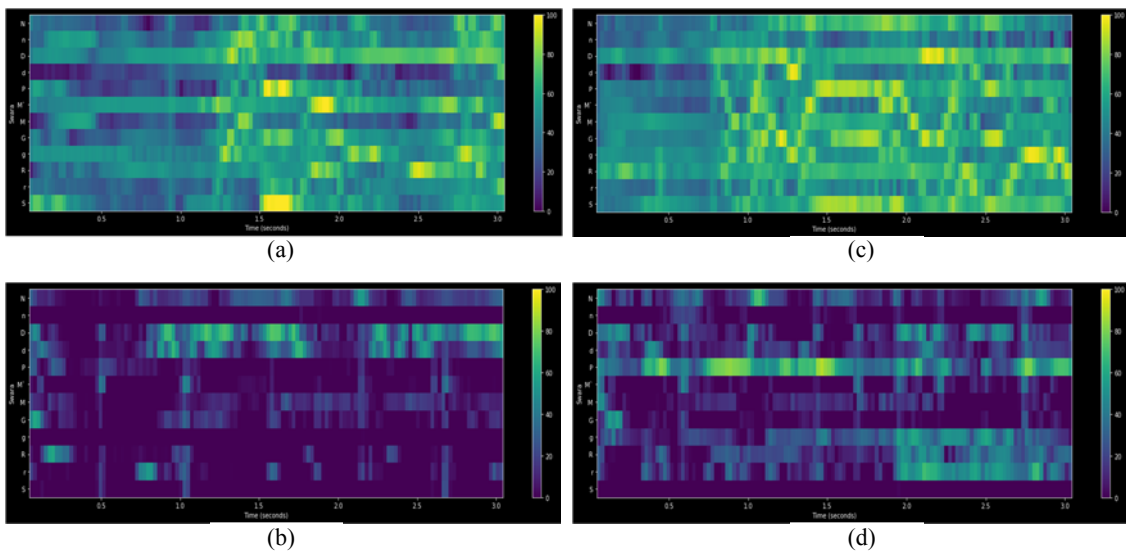


Fig. 2. Illustration of swaragram images for melodic movements in raga Yaman and Bhairavi.

3.3. Network Architecture

The proposed classifier is a deep neural network architecture designed to address the problem of raga identification. The architecture is based on the idea of multimodal learning, which allows the network to learn with reference to two modalities, rather than processing audio signal only. This is achieved by adding two separate branches that process audio and metadata.

As displayed in Fig. 3, the overall system architecture consists of three modules. The basic building blocks are convolutional and fully connected sub-networks. A convolutional (CNN) neural sub-network processes swaragrams, a fully connected (FCN) neural sub-network processes metadata, and a final fully connected sub-network merges information and yields the final classification decision.

The structure of the CNN sub-network involves a series of convolutional layers with different filter sizes and numbers of filters, followed by batch normalization and activation functions. At the end, a max pooling layer is added to reduce the spatial dimension of the feature maps of the last four subsequently convolutional layers, where rectified

linear unit is used as an activation function. The initial convolutional layer consists of 64 filters of size 3x3 and stride 2, followed by a batch normalization layer. Subsequently, three convolutional layers follow with 64, 32 and 32 filters of size 3x3, 5x5, 5x5 respectively and stride size 2. Each convolutional layer is followed by a batch normalization layer, which is used to normalize the contributions to each convolutional layer for every mini-batch. At the end, a max-pooling layer with a pool shape of 2x2 is used to average the size of the feature maps. The output of the final layer is flattened into a 1D vector and passed through the final fully connected layer to produce the final prediction.

The second sub-network processes information from metadata utilizing two sequentially fully connected layers with 16 and 12 nodes each.

The embeddings produced by the modules are fed to a final fully connected layer of 1024 nodes and are followed by a two output softmax layer in order for the classification decision to be drawn. Table 3 summarizes the dimensions of the architecture and displays the output of each layer when swaragram is processed.

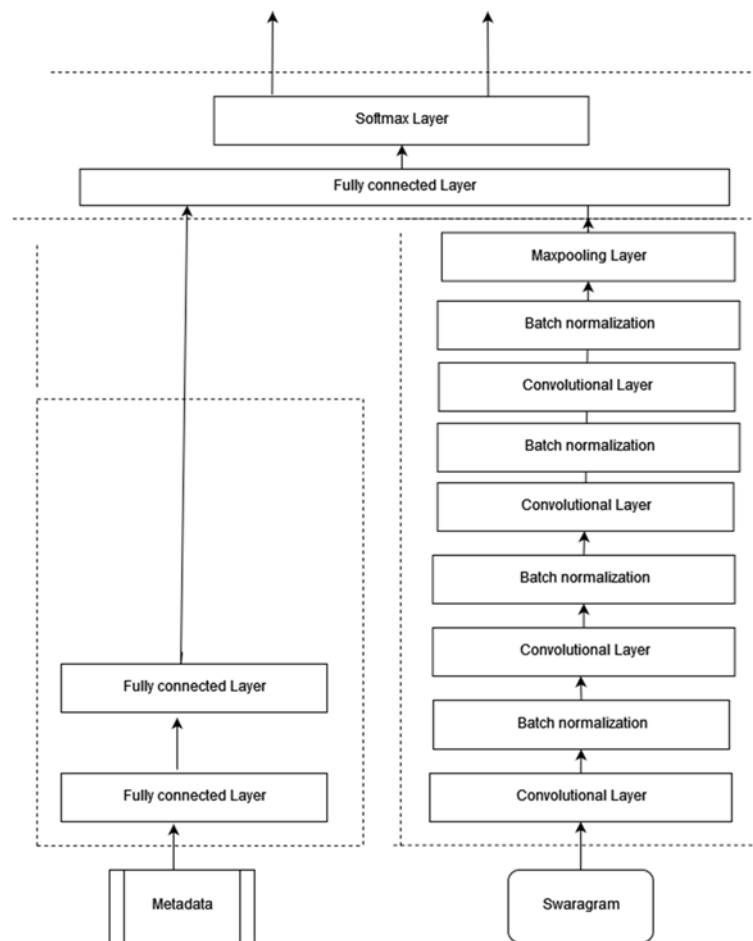


Fig. 3. Block diagram of the overall network architecture consisting of three sub-networks. A convolutional branch that processes audio features, a feed forward branch for metadata descriptors and a final feed forward stage that yields the classification decision.

Table 3. Network dimensionality.

Layer	Number of Filters	Kernel size	Stride	Output
Input	-	-	-	[198, 130, 1]
Conv2D	64	3×3	1×1	[198, 130, 64]
Batch_normalization	-	-	-	[198, 130, 64]
Conv2D	64	3×3	1×1	[198, 130, 64]
Batch_normalization	-	-	-	[198, 130, 64]
Conv2D	32	5×5	1×1	[198, 130, 32]
Batch_normalization	-	-	-	[198, 130, 32]
Conv2D	32	5×5	1×1	[198, 130, 32]
Batch_normalization	-	-	-	[198, 130, 32]
MaxPooling2D	-	-	-	[99, 65, 32]
Dense1	-	-	-	16
Dense2	-	-	-	12
Dense3	-	-	-	1024

4. Experiments and Results

4.1. Ablation Study

The ablation study of the proposed architecture involved several key components on model architecture and training procedure, with systematic experimentations on the role of sub-networks, the number of layers, the trade-off between depth and performance and any other design choices. In addition, ablation experiments were conducted on the effect of the dimensionality of the feature maps, the stride and zero padding values.

To support the training and testing of the network, backpropagation method was applied. To fine tune the model and define the model hyperparameters, training was performed for 300 epochs with the Adam gradient descent algorithm and a learning rate of 0.0001. Categorical cross entropy loss was optimized towards validation accuracy. To prevent network overfitting, we applied an early-stopping criterion of 10 epochs and dropout regularization, with a dropout value set to 0.2 for the fully connected layers and 0.4 for the convolutional layers. The network was evaluated using a 5-fold cross-validation scheme. The training-testing subsets of the entire dataset are defined by 90 % -10 % splitting percentages.

Training is performed on the subset of the Saraga dataset described in subsection 3.1, which is an unbalanced dataset. Training with an unbalanced dataset is challenging, as it can lead to poor performance of the model on the minority class. To address the class imbalance problem, the minority class was oversampled by replicating its audio samples with the usage of the random oversampling algorithm.

Experiments were designed to thoroughly evaluate the performance of the architecture and compare it with the state-of-the-art model. We conducted extensive experiments with different variations of the architecture, for three alternative system setups. The experimental setups were designed to serve the main directions of our study; in specific, the comparison of MFCC coefficients versus the swaragram audio feature, as well as the comparison between the

combined used of audio features and metadata versus using only audio-extracted features. To evaluate the contribution of each modality on the classification task, we experimented with bimodal and unimodal models. A unimodal classifier, where the CNN sub-network processes image representations of the audio signal, compares the performance of MFCC and swaragrams. The overall model configuration, where all sub-networks are present, evaluates whether the integration of metadata descriptors can improve the classification accuracy and generalization capability.

Our previous work [13] reported that information obtained from Mel frequency Cepstrum coefficients analysis effectively serves Hindustani raga discrimination tasks. Their performance is now compared with swaragram audio representations, and we rely on the single CNN network configuration to figure out which feature set is more effective for raga identification. The bimodal configurations experiment on the relative emphases to be placed on the two modalities.

4.2. Experimental Results

Overall, the experimental results of all model variations demonstrate that the bipolar architecture is highly effective for raga identification tasks and achieves state-of-the-art performance on the specific task.

To gain insight on the modalities and features impact we study the results presented in Table 4. All results, reported in the table, are expressed in terms of testing accuracy, in order to compare with the best reported result in the specified task. When a unimodal classifier, where MFCCs feature vectors are processed is used, classification accuracy for all five folds in terms of % scores [94.44, 96.29, 88.88, 96.29, 95.37], with an average value of 94.25. The unimodal classifier, where swaragrams are processed, reports a higher classification accuracy of 96.24, [96.45, 96.27, 95.72, 96.41, 96.36], which is a significant improvement over the previous performance.

Table 4. Experimental results in terms of accuracy for all folds.

Classifier	Input vector	Accuracy (%)	Average (%)
Unimodal	MFCC	[94.44, 96.29, 88.88, 96.29, 95.37]	994.25
Unimodal	Swaragram	[96.45, 96.27, 95.72, 96.41, 96.36]	996.24
Bimodal	MFCC + metadata	[97.14, 98.07, 99.19, 97.31, 95.37]	997.41
Bimodal	Swaragram + metadata	[97.74, 98.27, 99.38, 97.36, 98.48]	998.22

The classification accuracy values (%) for the bimodal classifier when MFCC coefficients are processed raises at [97.14, 98.07, 99.19, 97.31, 95.37] with an average score of 97.41. A combination of the swaragram feature and metadata, further increases the performance up to 98.22 (%), [97.74, 98.27, 99.38, 97.36, 98.48].

Experimental evaluation shows that fusion of audio features and metadata achieve the best classification performance. The classification accuracy results clarify the statement that representing Indian music with swaragrams, offers more information to a deep learning architecture for raga identification. The proposed method consistently outperforms the existing method of the reference paper [25], by 6.22% units, even though results are reported on a smaller dataset.

Hence, we further investigation should be made to the limitations that poor network interpretability imply, since it hinders the robustness evaluation of each network layer, as well as the models adaptability and transferability to alternative tasks on different data collections.

5. Conclusions

This work challenges machine driven raga identification with the application of a deep learning architectures. Our architecture utilizes convolutional and fully connected neural networks, for the classification task of two Hindustani ragas, namely Yaman and Bhairavi. The research focuses on more sophisticated audio features that are better suited to Hindustani music rather than generic audio features typically used in Eurocentric types of music. Our experimental results reported that the swaragram audio feature can serve as a 2-D image representation of the audio signal, which—processed by a convolutional network—achieves high classification accuracy for Hindustani ragas. The classification capability of such an architecture increases with the integration of metadata, providing evidence that fusion of data from multiple modalities into deep learning architectures increases the accuracy of machine listening techniques in raga identification tasks. It has been shown, that integration of swaragram representations of the audio signals and metadata has increased the ability of discriminating between Yaman and Bairavi ragas of Hindustani music, achieving a classification accuracy of 98.22%.

The classification capability of the proposed methodology should be further tested on several Indian datasets, with collections of various artists and types of instruments. We plan to expand the classifiers' performance on extended multiclass classification tasks, with the inclusion of different raga classes, on both Carnatic and Hindustani genres. Integration of more modalities, where face expressions, motion or emotional situation of the musician can be assessed, expresses the main direction of our future research, where the final scope is the formulation of a robust, multiclass classifier for raga identification.

References

- [1]. G. Tzanetakis, Computational ethnomusicology: a music information retrieval perspective, in *Proceedings of the Joint International Computer Music Conference / Sound and Music Computing Conference (ICMC/SMC)*, Athens, Greece, 14-20 September 2014, pp. 112–117.
- [2]. R. Powers and S. India, *The New Grove Dictionary of Music and Musicians*, Macmillan, London, 2001.
- [3]. J. Bor, S. Rao, W. Van Der Meer, and J. Harvey, *The raga guide: a survey of 74 hindustani ragas*, Nimbus Records with Rotterdam Conservatory of Music, 1999.
- [4]. R. Widdess, *The rāgas of early Indian music: Modes, melodies, and musical notations from the Gupta period to c. 1250*, Clarendon Press, Oxford, 1995.
- [5]. Melody, *Music in Motion*, 17-May-2013. [Online]. Available [Accessed: 23-Feb-2023]: <https://autrimncpa.wordpress.com/melodic-concepts/>.
- [6]. W. Kaufmann, *The Ragas of North India*. Bloomington: Asian Studies Research Institute, International Affairs Center, 1968.
- [7]. W. Van Der Meer and S. Rao, What you hear isn't what you see: The representation and cognition of fast movements in Hindustani music, in *Proceedings of the International Symposium Frontiers of Research on Speech and Music*, Baripada, India, 11-12 November 2006, pp. 12–20.
- [8]. A. Bhat, K. G. Krishnan, V. Mahesh, and V. K. Ananthapadmanabha, Deep learning approach to joint identification of instrument pitch and raga for Indian classical music, in *Advances in Speech and Music Technology*, Springer International Publishing, Cham, 2023, pp. 159–174.
- [9]. J. C. Ross and P. Rao, Detection of raga-characteristic phrases from Hindustani classical music audio, in *Proceedings of the 2nd CompMusic Workshop*, Istanbul, Turkey, 12-13 July 2012, pp. 133–138.
- [10]. S. Gulati, J. Serrà, V. Ishwar, S. Sentürk, and X. Serra, Phrase-based rAga recognition using vector space modeling, in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal*

- Processing (ICASSP)*, Shanghai, China, 20-25 March 2016, pp. 66–70.
- [11]. Y. Singh and A. Biswas, Computational approaches for Indian classical music: A comprehensive review, in *Advances in Speech and Music Technology*, Springer International Publishing, Cham, 2023, pp. 91–118.
- [12]. A. K. Datta, S. S. Solanki, R. Sengupta, S. Chakraborty, K. Mahto, and A. Patranabis, *Signal analysis of Hindustani classical music*. Singapore: Springer, 2017.
- [13]. I. Miliarezi and S. Paschalidou, A deep learning architecture for automatic raga identification, in *Proceedings of the 4th International Conference on Advances in Signal Processing and Artificial Intelligence (ASPAAI' 2022)*, Corfu, Greece, 19-21 October 2022, pp. 39-42.
- [14]. S. Madhusudhan and G. Chowdhary, DeepSRGM - sequence classification and ranking in Indian classical music via deep learning, in *Proceedings of Music Information Retrieval Conference (ISMIR)*, Delft, Netherlands, 4-8 November 2019, pp. 533-540.
- [15]. A. Anand, Raga identification using convolutional neural network, in *Proceedings of the Second International Conference on Advanced Computational and Communication Paradigms (ICACCP' 2019)*, Majitar, India, 25-28 February 2019, pp. 1-6.
- [16]. P. Kavitha, J. Charles, and L. S. Lekamge, A deep learning based approach for Janya raga classification in Carnatic music, in *Proceedings of the 6th SLAAI International Conference on Artificial Intelligence (SLAAI-ICAI' 2022)*, Kelaniya, Colombo, Sri Lanka, 1-2 December 2022, pp. 1-6.
- [17]. R. Sridhar and T. Geetha, Raga Identification of Carnatic music for Music Information Retrieval, *International Journal of Recent Trends in Engineering*, Vol. 1, Issue 1, 2009, pp. 571–574.
- [18]. S. Samsekai Manjabhat, S. G. Koolagudi, K. S. Rao, and P. B. Ramteke, Raga and tonic identification in Carnatic music, *J. New Music Res.*, Vol. 46, Issue 3, 2017, pp. 229–245.
- [19]. A. Bhat, A. V. Krishna, and S. Acharya, Analytical comparison of classification models for raga identification in Carnatic classical instrumental polyphonic audio, *SN Computer Science*, Vol. 1, 6, 2020, pp. 1-9.
- [20]. B. T. Rao, S. Chinnam, P. L. Kanth, and M. Gargi, Automatic Melakarta Raaga Identification System: Carnatic Music, *International Journal of Advanced Research in Artificial Intelligence*, Vol. 1, Issue 4, 2012, pp. 43–48.
- [21]. S. John, M. S. Smith, R. S. Sudheesh, and P. P. Lalu, Classification of Indian classical Carnatic music based on raga using deep learning, in *Proceedings of the IEEE Recent Advances in Intelligent Computational Systems (RAICS' 2020)*, Thiruvananthapuram, India, 3-5 December 2020, pp. 110-113.
- [22]. R. Rajan and S. Sivan, Multi-channel CNN-Based Raga Recognition in Carnatic Music Using Sequential Aggregation Strategy, *Circuits, Systems, and Signal Processing*, 2023, pp. 1–24.
- [23]. M. S. Siji John, R. S. Sudheesh, and P. P. Lalu, Classification of Indian classical Hindustani music based on raga using deep learning, in *Proceedings of the IEEE Recent Advances in Intelligent Computational Systems (RAICS' 2020)*, Thiruvananthapuram, India, 3-5 December 2020, pp. 110–113.
- [24]. A. Vasudevan and H. Ramasangu, A hybrid cluster-classifier model for Hindustani raga classification, in *Proceedings of the IEEE International Conference on Electronics, Computing and Communication Technologies (CONECCT' 2021)*, Bangalore, India, 9-11 July 2021, pp. 1–6.
- [25]. D. Joshi, J. Pareek, and P. Ambatkar, Indian classical raga identification using machine learning, in *Proceedings of the International Semantic Intelligence Conference (ISIC)*, New Delhi, India, February 2021, pp. 259–263.
- [26]. S. Chowdhuri, Phononet: multi-stage deep neural networks for raga identification in Hindustani classical music, in *Proceedings of the International Conference on Multimedia Retrieval*, 2019, pp. 197–201.
- [27]. D. P. Shah, N. M. Jagtap, P. T. Talekar, and K. Gawande, Raga Recognition in Indian Classical Music Using Deep Learning, in *Artificial Intelligence in Music, Sound, Art and Design, Proceedings of the 11th International Conference (EvoMUSART' 2022)*, Madrid, Spain, April 20–22, 2021, pp. 248–263.
- [28]. V. S. Pendyala, N. Yadav, C. Kulkarni, and L. Vadlamudi, Towards building a Deep Learning based Automated Indian Classical Music Tutor for the Masses, *Systems and Soft Computing*, Vol. 4, Issue 200042, 2022, pp. 1-8.
- [29]. S. Banerjee, G. Hota, R. Sanyal, and M. K. Rath, Two step recognition of Raags in Hindustani classical music using supervised deep learning, in *Proceedings of the 1st IEEE International Conference on Data, Decision and Systems (ICDDS' 2022)*, Bangalore, India, 2-3 December 2022, pp. 1-5.
- [30]. C. Alyani and A. Balwant, Raga identification techniques for classifying Indian classical music: A survey, *International Journal of Signal Processing Systems*, Vol. 5, Issue 4, 2017, pp. 1-6.
- [31]. R. Shastry and A. Bai, Signal processing and machine learning approaches and evaluation for Indian Classical music, in *Proceedings of the IOP Conference Series: Materials Science and Engineering*, Vol. 1116, No. 1, Jeju Island, Korea, 12-14 March 2021, pp. 012188.
- [32]. C. Kalyani and A. Balwant, Analyzing acoustics of Indian music audio signal using timbre and pitch features for raga identification, in *Proceedings of the 3rd International Conference on Imaging, Signal Processing and Communication (ICISPC' 2019)*, 2019, pp. 42–46.
- [33]. A. Srinivasamurthy, K. Gopala, S. Koduri, V. Gulati, and X. Ishwar, Corpora for music information research in Indian art music, in *Proceedings of the Joint International Computer Music Conference / Sound and Music Computing Conference (ICMC/SMC)*, Athens, Greece, 14-20 September 2014, pp. 1029-1036.
- [34]. P. Dighe, P. Agarwal, and H. Karnick, Siddhartha Thota, and Bhiksha Raj, Scale independent raga identification using chromagram patterns and swara based features, in *Proceedings of the IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, San Jose, US, 15-19 July 2013, pp. 1–4.
- [35]. Y. Singh and A. Biswas, Swaragram based residual neural architecture for raga identification in Indian classical music, in *Proceedings of the 12th International Conference on Computing Communication and Networking Technologies (ICCCNT' 2021)*, Khargpur, India, July 6-8, 2021, pp. 1-6.

- [36]. S. Patil, R. P. Bhavsar, and B. V. Pawar, A survey on computational approaches for raga identification, in *Proceedings of the IEEE World Conference on Applied Intelligence and Computing (AIC' 2022)*, Sonbhadra, India, 17-19 June 2022, pp. 85-92.
- [37]. P. Sutar, M. Saitwal, and S. Wakchaure, Rohit Shinde, and Poonam Sawdekar, A survey on raga identification using machine learning techniques, *International Journal of Innovative Research in Technology (IJIRT)*, Vol. 6, Issue 11, 2020, pp. 691-694.
- [38]. S. Alekh, Automatic Raga Recognition in Hindustani Classical Music, *arXiv:1708.02322*, 2017.
- [39]. S. Sreejith and R. Rajan, Rāga recognition in indian carnatic music using transfer learning, in *Proceedings of the Fourth International Conference on Microelectronics, Signals & Systems (ICMSS)*, Kollam, India, 18-19 November 2021, pp. 1–5.
- [40]. V. Kaimal and S. Barde, Implementation of a supervised learning model for raga identification in carnatic music, in *Proceedings of the Asian Conference on Innovation in Technology (ASIANCON' 2021)*, Pune, India, 27-29 August 2021, pp. 1-6.
- [41]. M. Hemali Dodia and P. N. P. Pandey, Identification of raga by machine learning with chromagram, *International Research Journal of Engineering and Technology (IRJET)*, Vol. 7, Issue 5, 2020, pp. 3220-3225.
- [42]. A. Srinivasamurthy, S. Gulati, R. Caro Repetto, and X. Serra, Saraga: Open datasets for research on Indian Art Music, *Empir. Musicol. Rev.*, Vol. 16, No. 1, 2021, pp. 85–98.
- [43]. Y. Singh and A. Biswas, Robustness of musical features on deep learning models for music genre classification, *Expert Systems with Applications*, Vol. 199, Issue 116879, 2022, pp. 1-14.
- [44]. Singh, Yeshwant and Kumar, Ranjeet and Biswas, Anupam, Swaragram: Shruti-Based Chromagram for Indian Classical Music, *Advances in Speech and Music Technology*, Springer, Singapore, 2021, pp. 109–118.



Published by International Frequency Sensor Association (IFSA) Publishing, S. L., 2023 (<http://www.sensorsportal.com>).

Open Access Book



Advances in Artificial Intelligence: Reviews

Sergey Y. Yurish, Editor



Artificial intelligence has been one of the fastest-growing technologies in recent years. The market growth is mainly driven by factors such as the increasing adoption of cloud-based applications and services, growing big data, and increasing demand for intelligent virtual assistants. Various end-use industries have also employed artificial intelligence such as retail and business analysis that has also boosted the demand in this market. The major restraint for the market is the limited number of artificial intelligence technology experts. The Book Series on 'Advances in Artificial Intelligence: Reviews' has been launched with the aim to fill-in this gap.

The first book volume from the 'Advances in Artificial Intelligence: Reviews' Book Series contains 11 chapters written by 21 contributors from academia and industry from 10 countries: Algeria, Germany, India, Iran, Israel, Russia, Slovenia, South Africa, Tunisia and USA.

1

http://www.sensorsportal.com/HTML/IFSA_Publishing.htm