# Automatic Music Mood Classification using Multi-class Support Vector Machine based on Hybrid Spectral Features

**Anitha Krishnaiah[1]\***        **Parameshachari Bidare Divakarachari[1]**

[1]*Department of Telecommunication Engineering,*
*GSSS Institute of Engineering and Technology for Women, Mysuru, India*
* Corresponding author's Email: anithak@gsss.edu.in

**Abstract:** Automatic Raga identification is an important step in computational musicology as it is helpful for indexing Indian music, classifying, recommending tunes, etc. However, automatic raga recognition is a difficult task as it requires guidance from the expertise training and each of the ragas is identified based on the characteristic phrase and pitch collection. However, the identification or extraction of appropriate musical sample features was difficult to enhance the success rate of the classifier. The feature identification and extraction were considered as the major issue in the existing models. To overcome such an issue, the proposed hybrid spectral feature extraction technique extracted and combined the spectral features of audios such as Spread, centroid, skewness, Mel Frequency Cepstrum Coefficients (MFCCs) and Linear Predictor Coefficients (LPCs) reduced the dimensionality complexity enhanced the success rate of the Multi-Support Vector Machine (MSVM) classifier for mood classification. The CompMusic dataset is utilized for this research work, where 4 types of ragas namely Sindhu Bhairavi, Darbari, Saveri, and Sri Raga, are considered for the mood classification of ragas. From these ragas, tone based features are extracted based on the amplitude of each raga samples and are fed to the MSVM classifier for mood classification of 4 different types are Sympathy, Serious, Peace, and Sad. The classification results for the proposed hybrid spectral feature extraction obtained better accuracy of 97.53% when compared to the existing Hidden Markov Model (HMM) obtained accuracy as 95.3%, Convolution Neural Network (CNN) model as 94%, and Gaussian Mixture Model (GMM) as 95 %.

**Keywords:** Indian classical music, Linear predictor coefficients, Mean centered clustering, Mel frequency cepstrum coefficients, Multi-support vector machine.

## 1. Introduction

Indian Classical Music (ICM) has two main branches namely Carnatic and Hindustani music which originated in the Indian subcontinent. Raga is the collection of melodic expressions that consist of identifiable melodic movements and the arrangement of tones gives rise to musical compositions and also music improvisations. The ragas are composed based on the moods, area associated with the pitches and the relations among other pitches. The ragas are identified based on the levels of pitches and based on the pitches of tones, their relations among them convey which particular kind of emotion the raga belongs to [1, 2]. The set of pitch arrangements give rise to a set of ragas which establish a mood of the raga or the flavor of raga [3, 4]. The emotion is classified based on the 2 step process that includes feature extraction and classification. The feature extraction process characterizes the set of parameters that belongs to different algorithms. The classifier is used to relate the features that were obtained using feature extraction and classifies them into various emotions to minimize the prediction error [5]. The audio signal is usually changing continuously wherein each signal varies with amplitude, frequency, or both. The signals present in audio are having different amplitudes because of the tone variations while playing music or audio. The audio features are used to categorize the low-level features and mid-level features. The low-level features are extracted using various signal processing techniques while the mid-level features are extracted based on the low-

level features [6]. However, the Music mood classification has several issues such as the labeling of moods which tend to vary according to an individual's perspective. This problem will give rise to ambiguity in mood classification while identifying the raga to which the type of mood belongs to. Secondly the ambiguous moods such as sad or depressed are difficult to identify, as most of the tone variations will be similar. Next, the noise is one of the distortions that degrade the audio content, so removal of these disturbances in the audio is an important task. Also, the lack of large and unbiased datasets in the public domain is a major challenge in music mood classification as all of these challenges required supervision from the expert [7]. The present research has described 4 types of mood taxonomies for 4 different types of ragas namely Sindhu Bhairavi, Darbari, Saveri, and Sri Raga which are selected from the CompMusic dataset that consists of 70 types of ragas [8]. The present research has considered 4 types of mood categories Sympathy, Serious, Peace, and Sad for classification as all these emotions negatively impact a person's daily life, be it personal or professional life. By knowing these negative moods, the present research helps to determine in which mood the person is in thereby enabling an individual to come back to the normal mental state. Music has a unique strength to establish one's attitude and also improves the concentration of individuals' which impacts positivity in everyone's life [9].

This research is organized as follows: Section 2 presents a literature survey of the existing methods involved for mood identification in ragas. Section 3 describes the proposed hybrid feature extraction techniques with the help of block diagrams and equations and Section 4 illustrates the results and discussion section that consists of two subsections: quantitative analysis and comparative analysis. The conclusion and future work of this research study is given in Section 5.

## 2. Literature review

Sinith [2] performed Raga recognition using Fibonacci series based pitch distribution in ICM. The developed model sets up an interesting relation between Fibonacci series and Just Intonation in ICM has been found. The Discrete Pitch Contour (DPC) based Hidden Markov Model (HMM) on a table derived using Fibonacci series was introduced for ICM and the pitch contours effectively represented the melody and acts as a tool for music information retrieval. However, the DPC have not considered the diverse pitch contours of same notes in different ragas that lowered the performances.

Rajib Sarkar [10] developed raga identification from Hindustani classical music signals using compositional properties. The developed model introduced a co-occurrence matrix for summarizing the mood aspects. The audio clip is considered a set of features that identified automatically the properties of raga they belong to. The audio clip type is classified based on the features were fed into Support Vector Machine (SVM). However, the classification of audios also required manual identification of raga with high expertise suggestions, which was not easy.

Makarand Velankar [11] developed melodic pattern recognition in Indian classical music for raga identification. The developed model extracted the pitch-based features as it was used for identifying the melodic level using the autocorrelation method. The results showed that the existing models required varied duration of time for accurate identification. However, the results were not enough for identifying the raga accurately as the limited number of raga samples were utilized for the research.

Sanchali Das [12] developed a computational model for mood classification of music by integrating an asymptotic approach with machine learning techniques. The developed model used audio features forms each of the songs that were analyzed mathematically which gave better results with a similar range of subclasses. However, the ground truth values for amplitude-based features limited the performance of the model using the Kokborok dataset.

Medina [13] developed an emotional classification of music using neural networks with the MediaEval dataset resulted the values of valence and arousal showed imbalanced classification results. However, the required characteristics of the dataset that are: size, class balance and quality of the annotations, still needed more improvement for achieving good performance in terms of accuracy.

Chandanpreet Kaur and Ravi Kumar [14] developed a Mean Centered Clustering (MCC) that was used for melody classification using time and frequency-domain supervised clustering. The developed research included preprocessing techniques and conventional classifiers such as Artificial Neural Network (ANN) and SVM was better melody classification. However, the identification or extraction of appropriate musical sample features was further needed to enhance the success rate of the classifier.

Siji John [15] performed classification of ICM based data on automatic Raga recognition using Deep Learning model through audio signal processing for music signal synthesis, audio dataset management and music therapy. The developed model utilized pitch contour for selecting Ragas in Carnatic music

that used key features and model implementation was done by using Convolutional Neural Network (CNN). However, the most challenging task was faced during pattern recognition using pitch contour based algorithm to obtain pitch sequences using Parsel-mouth library.

S. Samsekai Manjabhat [16] performed raga and tonic identification in Carnatic music using different classifiers such as feedforward neural network model, Gaussian Mixture Models and decision trees. The developed mathematical based approach used the parameters of pitch performed tones identification. However, the accuracy of the system clearly demonstrates that the proposed features capture more raga specific knowledge than just the set of notes that created overfitting problems.

## 3. Proposed method

The block diagram of the proposed method is shown in Fig. 1. Firstly, the database having distinct ragas undergo the preprocessing stage which is performed using the normalization technique. The preprocessed data is undergone for hybrid feature extraction using the feature parameters obtained from the Spectral feature extraction technique. The features are classified for distinct ragas using Multi SVM and the classes are framed into 4 types of emotions that are: Peace, Serious, Sadness and Sympathy.

### 3.1 Data collection

The Indian Music dataset consists of two datasets that are known as Carnatic Music Dataset (CMD) and Hindustani music dataset (HMD) [15]. The dataset comprises of full-length audio recordings having their respective raga labels and are useful for evaluating the approaches for automatic raga recognition in Indian art music. In this research, CMD is available in the Comp Music group which consists of high-quality audio recordings taken from ICM concerts. The dataset consists of 124 hours of time duration that includes 480 audio recordings, which are available in mp3 format. The 12 full lengths of each raga recording are present in the dataset, which is having 40 types of distinct ragas. These types of diverse ragas are having distinct and numerous features composition. Here, diverse refers to the melodic attributes, therefore, training and evaluation of the model using the dataset give better results. HMD comprises 116 hours of audio recordings and editorial metadata, which is presented among 300 recordings having 30 types of ragas. For the present research, 4 types of ragas namely Sindhu Bhairavi, Darbari, Saveri, and Sri Raga are used. The pre-processing stage for the proposed method is used to improve the data quality of a signal that reduces the unwanted distortions and also enhances the signal for further processes. The pre-processing techniques used in this proposed hybrid feature extraction method is explained in the following section.
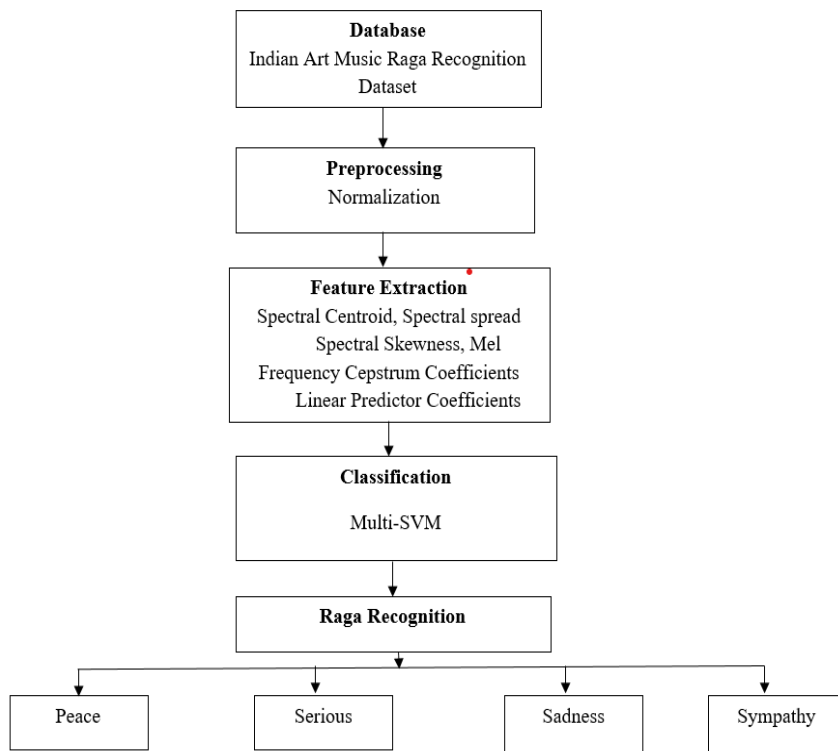


Figure. 1 Block diagram of the proposed hybrid feature extraction

## 3.2 Pre-processing using normalization

The ICM dataset consisting of ragas are usually contaminated with noises. These noise sources are generated by electronic instruments that create various colors of noises. These noises are generated by unpitched frequencies which give rise to the white noise component. The ICM has a distinct type of vocal, non-vocal tonic based performers and in few moments the tonic notes vary distinctly. The higher tonic ranges create loudness or noises and normalization of such loudness is very important to avoid misclassification. A constant amount of gain for an audio recording gives rise to a target level for an Amplitude so that the same amount of gain is spread all over the recorded audio that keeps the Signal to Noise ratio (SNR) and dynamics unchanged. The tracked audio signal is normalized for all the audios present in the ICM due to which the louder signals are constant throughout. The normalization methodology is very effective to remove white noise that also helps to enhance the audio quality significantly. The general formula of image normalization methodology is given in Eq. (1).

$$AN = (A - Min) + \frac{newMax - newMin}{Max - Min} + newMin \quad (1)$$

where, $A$ is represented as the original audio signal, $AN$ is stated as a new audio signal, ($Min = 0, Max = 1$) is specified as pitch amplitude range of original audio and ($newMax, newMin$) is stated as pitch amplitude range of the new audio signal. The prediction of amplitude dynamic range is very important before signal acquisition as it makes it easier by knowing the amplitude limit. The predefined range normally lies in the range of 0 to 1 for transforming the signal to statistical perception range thereby dividing the range as maximum and minimum amplitude dynamic range values. The preprocessed audio signals now undergo the feature extraction process, to get the tone based features extracted from them so that the type of raga in audio can be determined. Fig. 2 and Fig. 3 shows the normalized signals for Sri raga and Darbari raga using the normalization process.
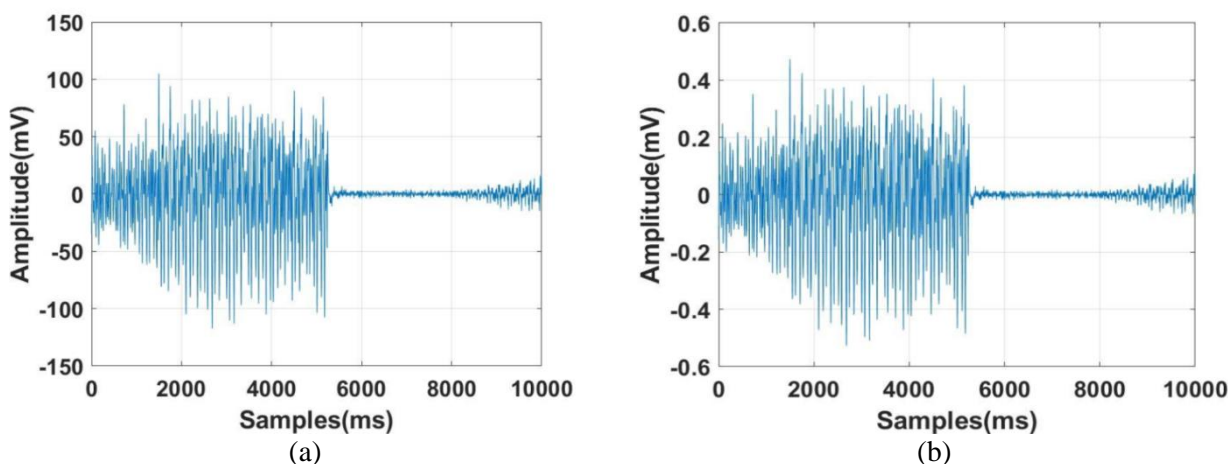


(a)                                          (b)

Figure. 2: (a) Input signal for Sri raga and (b) Normalized signal for Sri raga



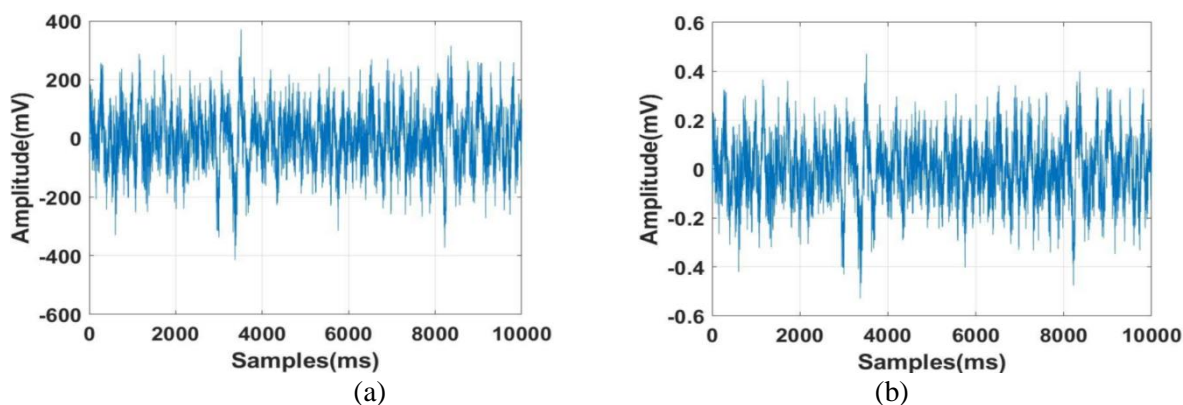(a)                                          (b)

Figure. 3: (a) Input signal Darbari raga and (b) Normalized signal for Darbari raga

## 3.3 Hybrid feature extraction using spectral features

The next step after preprocessing is the feature extraction using spectral features. The spectral features are also the frequency-based features that are determined by converting time signals into frequency domain signals. The signals are converted into fundamental frequency components by using Fourier Transform. These frequency-based components are used for pitch identification, note identification, rhythm detection and frequency component of the type of melodies. The spectral features are used to characterize the features which serve as a physical correlation for attributes like coloration and timbre. The spectral features used in the research are the spectral centroid, spectral spread, spectral skewness, MFCC and LPC.

### 3.3.1. Spectral centroid

The spectral centroid is defined as the center of gravity that is used to rate the most frequently occurring map which determines the timbre parameters. The spectral centroid represents the range of the frequency and the energy on the spectrum is dependent. Spectral Centroid is similar to the weighted mean and is expressed as shown in Eq. (2).

$$f_c = \frac{\sum_k S(k)f(k)}{\sum_k S(k)} \qquad (2)$$

Where,
$f_c$ Is the Centre frequency
$S(k)$ is a weighted mean of bin 'k'
$f(k)$ Is the Weighted frequency of bin 'k'

From Eq. (2), it is shown that the normalized centroid function is rated highest at the center which reduces gamma tone parameters for audio features. This feature is used for indicating the brightness of the audio sample that results from the spectral center of gravity. The centroid is calculated to normalize or balance the signals throughout so that the frequency throughout will be obtained. An advantage of using this feature is that it can balance the low spectral centroid from the high spectral centroid.

### 3.3.2. Spectral spread

The spectral spread is the average deviation around the rate map of the centroid that is associated with the signal bandwidth. The noises are having a large spectral spread which is having distinct tonal sounds that are isolated with a low spectral spread.

The spectral spread feature is similar to the spectral centroid which is used to normalize the highest-rated map center frequency ranging between 0 and 1. The spectral spread feature is used to determine the amount of the frequency spread across the spectrum which is similar to the frequency bandwidth. The features differentiate among the low spectral spread and high spectral spread which is expressed as shown in Eq. (3).

$$Spread, \sigma = \sqrt{\sum_{i=1}^{N}[(f_i - f_c)^2 \cdot w_i]} \qquad (3)$$

Where,
$w_i$ is the weight of $i^{th}$ frequency component $(f_i)$
$f_c$ is the centroid of the spectrum obtained by using Eq. (2).

### 3.3.3. Spectral skewness

The spectral skewness is the measure of spectrum symmetry with respect to the relative arithmetic mean. If the audio is having silent tones the spectral skewness will be 0 and if the high voiced signals are present in the audio, then it is 1 or if the frequency of the audio is very low it is -1. The features used to get an idea about the timbre of sound and the range of these type of signal range from 0, positive value and negative value. Here the positive value indicates the content of frequency skewed towards the right of the mean and the negative value is the frequency skewed towards the left of the mean.

### 3.3.4. Mel frequency cepstrum coefficients

The Mel scale is the perceived pitch or frequency of the pure tone that measures the actual frequency. As human beings are very intelligent in detecting small changes in the low level of pitches it is very important to consider for the present research. MFCC is used for speech processing and music classification can model the subjective frequency contents of audio signals. An expression for the approximation of Mel from physical frequency is expressed as shown in the following Eq. (4).

$$f_{mel} = 2595 \, log_{10}(1 + \frac{f}{700}) \qquad (4)$$

The equation for the DFT in MFCCs is represented as shown in Eq. (5).

$$S_i(k) = \sum_{n=1}^{N} s_i(n) h(n) e^{\frac{-j2\pi kn}{N}} \qquad (5)$$

Where $1 \leq k \leq K$

107

From the equation, $h(n)$ is the hamming window for $N$ sample

$K$ is the length of the DFT
$s_i(n)$ that ranges from 1 to 400
$i$ is the range of number of frames
$S_i(k)$ is the complex DFT

The MFCC spectrum is the product of magnitude spectrum $X(k)$ which explained using the following Eq. (6).

$$s(m) = \sum_{k=0}^{N-1}[|X(k)|^2 H_m(k)] \quad (6)$$

$$0 \leq m \leq M - 1$$

From the equation
$M$ is the total number of Mel weighting filters
$H_m(k)$ is the weight for $k^{th}$ energy spectrum for the output band $'m'$

### 3.3.5. Linear predictor coefficients

The LPCs are used to minimize the energy consumption of error signals. For the process of audio or music recognition, the autocorrelation method is used to improve the computational complexity or efficiency for inherent stability. The proposed method uses the covariance method that guarantees the inherent stability by using the Linear Predictor synthesis filter. Initially, the speech signal $s(n)$ is multiplied with the window $w(n)$, which gives the result as a windowed speech segment $s_w(n)$. The windowed speech signal is expressed as shown in Eq. (7).

$$s_w(n) = s(n)w(n) \quad (7)$$

Once the speech signal is obtained the residual signal of energy is minimized. The residual energy is obtained as $E_P$ in Eqs. (8) and (9).

$$E_P = \sum_{n=-\infty}^{\infty} e^2(n) \quad (8)$$

$$= \sum_{n=-\infty}^{\infty}\left(s_w(n) + \sum_{k=1}^{p} a_k s_w(n-k)\right)^2 \quad (9)$$

The values of $a_k$ are minimized by the $E_P$ sets
$E_P$ is the partial derivative of the energy $E$
where the LPC parameters are set to 0. The equation is expressed as shown in Eq. (10).

$$\frac{\partial E_P}{\partial a_k} = 0 \quad (10)$$

$$1 \ll k \ll p$$

Where,
$p$ is the linear equation of unknown parameters represented as $a_1 \dots, a_p$ expressed in Eqs. (11) and (12).

$$E_P = \sum_{k=1}^{p} a_k \sum_{n=-\infty}^{\infty} s_w(n-i)s_w(n-k) \quad (11)$$

$$E_P = -\sum_{n=-\infty}^{\infty} s_w(n-i)s_w(n) \quad (12)$$

$$1 \leq i \leq p$$

Thus, the aforementioned spectral feature parameters are combined and the proposed hybrid feature extraction technique is processed. Based on the hybrid feature parameters, the spectral frequency-based features are extracted from the audio signals. The hybrid feature is expressed in Eq. (13).

$$F = f_c + f_{sp} + f_{sk} + f_{mel} + f_{lpc} \quad (13)$$

Where,
$F$ is the hybrid spectral feature
$f_c$ is the spectral centroid feature
$f_{sp}$ is the spectral spread feature
$f_{sk}$ is the spectral skewness feature
$f_{mel}$ are the MFCCs
$f_{lpc}$ are the LPCs

The extracted hybrid spectral features are now fed as an input for SVM in order to classify the features to obtain multi mood classes for ragas.

### 3.4 Classification using multi-class SVM

The extracted spectral features are fed to the multiclass SVM for obtaining the classification rate in terms of accuracy that classifies the total number of ragas acquired from the CompMusic dataset. To perform specific classification, arousal and valence problems were identified during two classifications based on high and low tones of ragas. The Multiclass SVM classified ragas into 4 types of emotions are Sadness, Sympathy, Peace, and Serious, directly. The results obtained showed both valence and arousal cases, which yields higher accuracy than the direct four-class classification. But, there seems to be no apparent difference between arousal and valence classification. Next, the model-based scheme is composed of two-level binary classifiers. An advantage of using MSVM is that its average accuracy was better achieved using multi-class in a single step process that lowered the computational complexity of the system.

The SVM is applied to perform binary classification that assigns the data points either 1 or 0.

For multiclass classification, the same principle is utilized. The multiclass problem is broken down into multiple binary classification cases that followed one-vs-one multiclass classification for all other classes. The number of classifiers necessary for one-vs-one multiclass classification can be retrieved with the following Eq. (14).

$$Number\ of\ classifiers = \frac{n\,(n-1)}{2} \qquad (14)$$

Where $n$ is the number of classes considered From Eq. (14), the number of SVM classifiers used to determine the classes for SVM is given by solving the following Eq. (15)

$$\min_{w^{ij},b^{ij},\xi^{ij}} \frac{1}{2}\left(w^{ij}\right)^T\left(w^{ij}\right) + C\sum_t \zeta_t^{ij} = 0 \qquad (15)$$

The solutions obtained for the above mentioned Eq. (15) are as follows:

$$\left(w^{ij}\right)^T\phi(x_t) + b^{ij} \geq 1 - \xi_t^{ij}\ if\ y_t = i$$

$$\left(w^{ij}\right)^T\phi(x_t) + b^{ij} \leq -1 + \xi_t^{ij}\ if\ y_t = j$$

$$\xi_t^{ij} \geq 0$$

Where $\xi_t^{ij}$ is the distance to the correct margin where $t = 0,1\dots n$

$C$ is the regularization parameter

$\phi(x_t)$ Transformed input space vector

$\left(w^{ij}\right)^T\left(w^{ij}\right)$ Product is the normal vector

$b^{ij}$ Denotes the bias parameter which denotes correctly predicted

$w^{ij}$ is finding the optimal value.

If $\left(w^{ij}\right)^T\phi(x_t) + b^{ij}$ says $x$ is in the $i^{th}$ class, then vote $i^{th}$ class as one. Otherwise, increase $j^{th}$ class by one. Then we predict that $x$ is in the class with the largest vote. The voting approach described above is also called the "Max Wins" strategy. In this way, all 4 moods are classified using MSVM.

# 4. Results and discussion

The proposed hybrid feature extraction for mood classification of ragas is performed using MATLAB 2018a operating in i7 core processor. The installed memory RAM is 16GB operating at 4.20 GHz system frequency with 64-bit Operating system. Initially, each of the ragas is divided into 10 segments for experimentation and 70% of segments are taken for

the training and 30% of segments are taken for testing. The present result section is divided into sections namely quantitative analysis and comparative analysis. The performance measures that are used for evaluating the classification results are as follows

i. **Accuracy**:
Accuracy is defined as the ratio of correctly predicted observations to the total number of observations. The accuracy is calculated by using the Eq. (16).

$$Accuracy = \frac{(TP+TN)}{(TP+TN+FP+FN)} \qquad (16)$$

ii. **Specificity**:
Specificity measures the proportion of negatives that are identified correctly using the following Eq. (17).

$$Specificity = \frac{TN}{TN+FP} \times 100 \qquad (17)$$

iii. **Sensitivity**
Specificity measures the proportion of positives that are identified correctly using the following Eq. (18).

$$Sensitivity = \frac{TP}{TP+FN} \qquad (18)$$

iv. **Positive predictive value (PPV):**
PPV is defined as the ratio of true positive to the sum of the true positive and False Positive values which is expressed in Eq. (19).

$$PPV = \frac{TP}{TP+FP} \qquad (19)$$

Where,
TP is True Positive, FP is False Positive, TN is True Negative and FN is False Negative values.

## 4.1 Quantitative analysis

The proposed hybrid spectral feature extraction technique is used to extract spectral features that are fed for the classifier known as Multi SVM for mood classification of the raga. The results are evaluated in terms of Accuracy, Specificity, Sensitivity, and PPV for the proposed hybrid spectral feature extraction as shown in Table 1. The features are Spectral Centroid, spread, skewness, MFCC and LPC that are extracted and fed for the respective classifiers such as ANN, Deep Neural Network (DNN), K-Nearest Neighbor (KNN) and MSVM. The results obtained for all the 5 five spectral features are tabulated in Table 1. These

features are combined to form hybrid spectral features that give better results of performance when compared with each spectral feature. The existing methods failed to consider more number features for the mood classification in ragas. The results obtained from the classifiers are shown in Table 1 and the graphical representation for the results are shown in Fig. 4 and 5.

## 4.2 Comparative analysis

The comparative analysis is performed for existing MLP and MCC models against the proposed Hybrid Spectral Feature Extraction methods that evaluated the results with CompMusic dataset for mood classification of ragas. Sinith [2] developed the Discrete Pitch Contour (DPC) based Hidden Markov Model (HMM) on a table derived using Fibonacci series failed to consider when the same notes in different ragas performed DPC that showed deviations and lowered the performance. Similarly, Siji John [15] utilized CNN for automatic raga classification that faced problem during raga recognition. However, the most challenging task was faced during pattern recognition using pitch contour based algorithm to obtain pitch sequences present in Parsel-mouth library. Therefore, Samsekai

Manjabhat [16] performed automatic raga recognition using Gaussian Mixture mathematical models demonstrated the features captured more raga specific than the set of notes. However, the accuracy of the system clearly demonstrated that the proposed features captured more raga specific knowledge than just the set of notes that created overfitting problems. Whereas, the proposed hybrid features extraction with Multi-SVM approach processed performs dimensionality reduction overcome several problems such as performance enhancement. The accuracy of the system clearly demonstrated that the proposed features capture more raga specific knowledge than just the set of notes.

The classification results for the proposed hybrid spectral feature extraction obtained better accuracy of 97.53% when compared to the existing Fibonacci series based pitch distribution HMM obtained accuracy as 95.3%, CNN model as 94%, and Gaussian Mixture Model as 95 %. The proposed model used hybrid spectral features that were formed by combining the spectral features. The proposed model was more robust and showed better mood classification results. The tabulation of the results for the existing and the proposed hybrid spectral feature extraction method is shown in Table 2.

Table 1. Results for the classifiers using spectral frequencies

| Features | Classifiers | Accuracy (%) | Sensitivity (%) | Specificity (%) | PPV (%) |
|---|---|---|---|---|---|
| Spectral Centroid | ANN | 72.98 | 58.32 | 67.61 | 66.29 |
| | DNN | 86.91 | 83.68 | 81.18 | 80.79 |
| | KNN | 90.9 | 78.13 | 86.34 | 86.2 |
| | MSVM | 87.52 | 97.29 | 89.55 | 91.07 |
| Spectral spread | ANN | 61.56 | 52.17 | 52.81 | 53.13 |
| | DNN | 72.28 | 83.6 | 78.19 | 80.58 |
| | KNN | 86.35 | 85.25 | 82.29 | 80.54 |
| | MSVM | 82.16 | 87.67 | 93.02 | 83.55 |
| Spectral Skewness | ANN | 73.64 | 61.99 | 68.61 | 64.88 |
| | DNN | 82.8 | 86.67 | 81.21 | 84.83 |
| | KNN | 87.91 | 83.39 | 90.28 | 87.13 |
| | MSVM | 90.95 | 92.97 | 95.54 | 97.38 |
| Mel frequency Cepstrum Coefficients | ANN | 66.65 | 62.68 | 69.35 | 57.13 |
| | DNN | 79.86 | 74.14 | 75.37 | 73.72 |
| | KNN | 91.68 | 82.65 | 89.79 | 83.45 |
| | MSVM | 90.36 | 89.18 | 87.03 | 96.19 |
| Linear Predictor Coefficients | ANN | 72.32 | 57.2 | 70.4 | 55.9 |
| | DNN | 73.67 | 71.91 | 70.57 | 72.66 |
| | KNN | 89.59 | 86.46 | 93.93 | 89.39 |
| | MSVM | 83.36 | 91 | 85.88 | 91.54 |
| Hybrid | ANN | 75.75 | 65.27 | 72 | 68.16 |
| | DNN | 87.38 | 86.85 | 85 | 85.69 |
| | KNN | 91.88 | 87.63 | 94 | 90.66 |
| | MSVM | 95.75 | 97.64 | 97.5 | 97.53 |
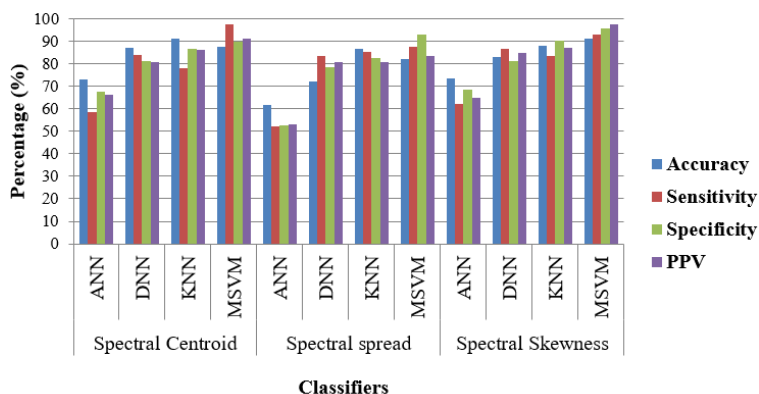
Figure. 4 Graphical representation of results for the spectral features centroid, spread and skewness for distinct classifiers
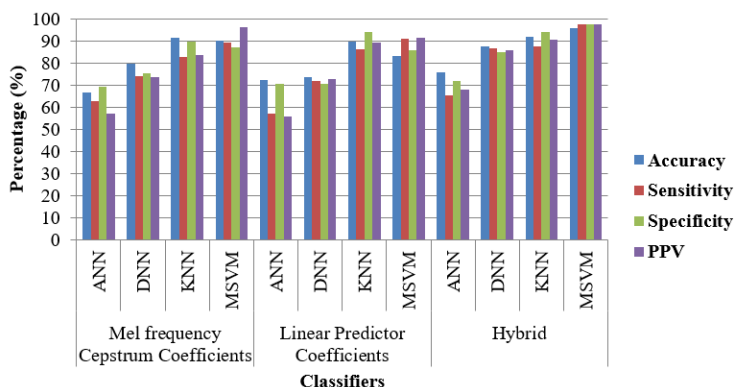


Figure. 5 Graphical representation of results for the MFCCs, LPCs and hybrid spectral features for distinct classifiers

Table 2. Accuracy values for the existing and the proposed methods

| Dataset | Authors | Methodology | Accuracy (%) | Error values |
|---|---|---|---|---|
| CompMusic | Sinith [2] | Fibonacci series based pitch distribution Hidden Markov Model | 95.3 | 4.7 |
| | John [15] | Convolutional Neural Network | 94 | 6 |
| | Samsekai Manjabhat [16] | Gaussian Mixture Model | 95 | 5 |
| | Proposed | Hybrid spectral Feature Extraction with MSVM classifier | 97.53 | 2.47 |

## 5. Conclusion

The ragas are identified based on the levels or tones of pitches and types of emotion conveyed based on their relations. The identification or extraction of appropriate musical sample features was further needed to enhance the success rate of the classifier problem. Initially, the ragas were extracted from the CompMusic database for the research. The white noises in the audio signals were contaminated, which further corrupted the classification of moods. The Normalization process removed the unwanted noises for further processing. The proposed hybrid spectral feature extraction technique extracted features of signals which were represented as a combination of Spread, centroid, skewness, MFCCs and LPCs which are denoted as ($F = f_c + f_{sp} + f_{sk} + f_{mel} + f_{lpc}$),

for an audio signal f. The extracted tone based features were fed to the Multi-Support Vector Machine (MSVM) classifier for classifying music moods into 4 different categories are Sympathy, Serious, Peace, and Sad. The classification results for the proposed hybrid spectral feature extraction obtained better accuracy of 97.53% when compared to the existing Fibonacci series based pitch distribution Hidden Markov Model obtained accuracy as 95.3%, CNN model as 94%, and Gaussian Mixture Model as 95 %. In the future, different types of mood can be used for classification.

## Conflicts of Interest

The authors declare no conflict of interest.

## Author Contributions

The paper conceptualization, methodology, software, validation, formal analysis, investigation, resources, data curation, writing—original draft preparation, writing—review and editing, visualization, have been done by 1st author. The supervision and project administration, have been done by 2nd author.

## References

[1] S. Sharma, A. Sasidharan, V. Marigowda, M. Vijay, S. Sharma, C. S. Mukundan, L. Pandit, and N. R. Masthi, "Indian Classical music with incremental variation in tempo and octave promotes better anxiety reduction and controlled mind wandering–a randomised controlled EEG study", *EXPLORE*, 2020.

[2] M. S. Sinith, S. Tripathi, and K. V. V. Murthy, "Raga recognition using fibonacci series based pitch distribution in Indian Classical Music", *Applied Acoustics*, Vol. 167, pp. 107381, 2020.

[3] S. Chapaneri and D. Jayaswal, "Deep Gaussian processes for music mood estimation and retrieval with locally aggregated acoustic Fisher vector", *Sādhanā*, Vol. 45, No. 1, pp. 1-13, 2020.

[4] S. Das, S. Satpathy, S. Debbarma, and B. K. Bhattacharyya, "Data analysis on music classification system and creating a sentiment word dictionary for Kokborok language", *Journal of Ambient Intelligence and Humanized Computing*, pp. 1-12, 2019.

[5] D. Chaudhary, N. P. Singh, and S. Singh, "Automatic music emotion classification using hashtag graph", *International Journal of Speech Technology*, Vol. 22, No. 3, pp. 551-561, 2019.

[6] S. Mo and J. Niu, "A novel method based on OMPGW method for feature extraction in automatic music mood classification", *IEEE Transactions on Affective Computing*, 2017.

[7] A. K. Patel, S. S. Chouhan, and R. Niyogi, "Using Crowd Sourced Data for Music Mood Classification", *International Conference on Big Data Analytics*, pp. 363-375, 2018.

[8] B. G. Patra, D. Das, and S. Bandyopadhyay, "Labeling data and developing supervised framework for hindi music mood analysis", *Journal of Intelligent Information Systems*, Vol. 48, No. 3, pp. 633-651, 2017.

[9] R. Sarno, J. A. Ridoean, D. Sunaryono, and D. R. Wijaya, "Classification of music mood using MPEG-7 audio features and SVM with confidence interval", *International Journal on Artificial Intelligence Tools*, Vol. 27, No. 05, pp. 1850016, 2018.

[10] R. Sarkar, S. K. Naskar, and S. K. Saha, "Raga identification from Hindustani classical music signal using compositional properties", *Computing and Visualization in Science*, Vol. 22, No. 1-4, pp. 15-26, 2019.

[11] M. Velankar, A. Deshpande, and P. Kulkarni, "Melodic pattern recognition in Indian classical music for raga identification", *International Journal of Information Technology*, pp. 1-8, 2018.

[12] S. Das, B. K. Bhattacharyya, and S. Debbarma, "Building a computational model for mood classification of music by integrating an asymptotic approach with the machine learning techniques", *J. Ambient Intell. Human Comput*, 2020.

[13] Y. O. Medina, J. R. Beltrán, and S. Baldassarri, "Emotional classification of music using neural networks with the MediaEval dataset", *Personal and Ubiquitous Computing*, pp. 1-13, 2020.

[14] C. Kaur and R. Kumar, "Mean centred clustering: improving melody classification using time-and frequency-domain supervised clustering", *Sādhanā,* Vol. 44, No. 2, pp. 40, 2019.

[15] S. John, M. S. Sinith, R. S. Sudheesh, and P. P. Lalu, "Classification of Indian Classical Carnatic Music Based on Raga Using Deep Learning", In: *Poc. of the Conf. 2020 IEEE Recent Advances in Intelligent Computational Systems (RAICS),* pp. 110-113, 2020.

[16] S. S. Manjabhat, S. G. Koolagudi, K. S. Rao, and P. B. Ramteke, "Raga and tonic identification in carnatic music", *Journal of New Music Research*, Vol. 46, No. 3, pp. 229-245, 2017.