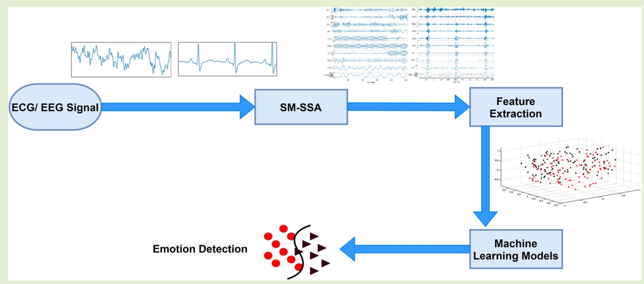
A Computerized Approach for Automatic Human Emotion Recognition Using Sliding Mode Singular Spectrum Analysis

Lakhan Dev Sharma, *Member, IEEE*, and Abhijit Bhattacharyya, *Member, IEEE*

***Abstract*—Emotion is a biological process owing to the alteration manifested in the human neurophysiological system, triggered consciously or unconsciously by an external stimulus. Physiological signals such as electroencephalogram (EEG) and electrocardiogram (ECG) measure the neuronal and cardiac activities associated with different emotional states. The complex biological activities result in a highly nonstationary nature of the acquired physiological signals, requiring advanced signal processing and machine learning (ML) techniques to identify and classify hidden patterns. This paper proposes a three-step process for comprehensive analysis and classification of human emotional states. The first step comprises decomposing physiological signals into reconstructed components (RCs) using sliding mode singular spectrum analysis (SM-SSA). In the second step, the discriminatory features such as information potential (IP) and centered correntropy (CEC) were computed from the extracted RCs. Afterward, the extracted features were considered input to various ML classifiers to discriminate human emotional states in the third step. The proposed method was studied over two publicly available databases, DREAMER and AMIGOS, for emotion analysis and achieved the highest classification accuracy of 92.38% in classifying human emotions. The obtained experimental results indicate that the proposed method can identify different human emotional states and yield better performance than existing emotion recognition methods.**



***Index Terms*—Human emotions, sliding mode singular spectrum analysis (SM-SSA), information potential (IP), centered correntropy (CEC), classifiers.**

I. Introduction

EMOTION is an intense mental activity having impact on memory [1], perception [2], attention [3], and decision-making [4]. In recent years, automation in human emotion analysis has gathered researchers' attention in human-computer interactions. For natural and unambiguous interaction, robots and computers must have the capability of emotion processing. Emotions are associated with many mental disorders like autism, hyperactivity, depression, and game addiction [5], [6]. Affective computation focuses mainly on pattern recognition techniques for the analysis of human emotions. Two important non-physiological methods of emotion recognition are speech [7] and facial recognition [8]. However, both these modalities may not represent the actual mental state because the non-physiological response can be manipulated or hidden. Whereas the physiological signals, namely, respiration (RSP), galvanic skin response (GSR), temperature, electrocardiogram (ECG), and electroencephalogram (EEG) can represent the real emotional states of an individual [9], [10]. In exposure to an emotional situation, the physiological reaction would be reflected on the autonomous nervous system (ANS) and may alter the balance between parasympathetic and sympathetic branches [11]. In particular, ECG and EEG have shown a strong correlation between their waveforms and emotional attributes. Also, the use of multimodal signals has shown improved recognition accuracy [12]. In addition, the acquisition of EEG and ECG signals are inexpensive, fast, and noninvasive, making them suitable tools for studying human emotions in response to given stimuli [13]. Emotion recognition requires implementing several steps such as the computation of discriminatory features, feature selection, and classification [6], [14], [15]. The extraction of

Manuscript received August 10, 2021; revised October 11, 2021; accepted October 11, 2021. Date of publication October 26, 2021; date of current version November 30, 2021. The associate editor coordinating the review of this article and approving it for publication was Prof. Rosario Morello. *(Corresponding author: Abhijit Bhattacharyya.)*

Lakhan Dev Sharma is with the School of Electronics Engineering, VIT-AP University, Amaravati, Andhra Pradesh 522237, India (e-mail: devsharmalakhan@gmail.com).

Abhijit Bhattacharyya is with the Department of Electronics and Communication Engineering, National Institute of Technology Hamirpur, Hamirpur 177005, India (e-mail: abhijit@nith.ac.in).

Digital Object Identifier 10.1109/JSEN.2021.3120787

discriminative features from EEG and ECG signals involving signal processing has gained much interest in recent literature.

Several methods have been proposed in the literature for human emotion recognition using EEG and ECG signals. In [16] modality fusion strategy was adopted by employing EEG signals, pupillary response, and gaze distance along with a support vector machine (SVM) classifier for human emotion recognition. In [17], the authors presented a study on human emotion recognition using EEG signals, comprising common spatial pattern (CSP) based feature extraction and linear SVM classifier. The SVM classifier was also explored in another work [18] for human emotion recognition. In [19] emotion classification was carried out with time-frequency (TF) domain features extracted from different TF representations, namely Zhao-Atlas-Marks transform, Hilbert Huang spectrum, the spectrogram of EEG signals. In [20], features were extracted from fast Fourier transform (FFT) and fed to Bayes' probabilistic classifier for human emotion recognition using EEG signals. The differential entropy feature was employed in [21] for recognizing human emotions. In [22], deep belief networks (DBNs) were applied for finding the critical frequency bands and channels for classifying emotions into positive, neutral, and negative emotions. In [23], the spatial-temporal recurrent neural network was used for classifying human emotions. In [24], authors extracted 18 distinct linear and non-linear features for classifying human emotions using SVM classifier. In [25], multivariate empirical mode decomposition (MEMD) was employed to extract TF domain features for the classification of human emotions with an artificial neural network classifier. In [6], Fourier-Bessel series expansion-based empirical wavelet transform (FBSE-EWT) was used for generating multivariate Hilbert marginal spectrum. Further, time and frequency domain entropy features were computed and fed to an autoencoder-based random forest (ARF) classifier for classifying different human emotions.

A detailed review of the existing literature substantiates that there has been an emerging research trend in investigating discriminatory features and suitable machine learning (ML) algorithms for human emotion recognition. Despite many studies, most emotion recognition works are limited with their efficacy, making them not fully reliable techniques for emotion recognition. Moreover, only a few studies have considered multimodal physiological signals. Hence, the literature supports further investigation of the patterns inherent in the multimodal physiological signals and their classification for human emotion recognition.

Motivated by this, we propose a new method for human emotion recognition using multimodal physiological signals. The proposed method is composed of three steps as depicted in Fig. 1. In the first step, the physiological signals are decomposed into reconstructed components (RCs) using sliding mode singular spectrum analysis (SM-SSA) for analyzing the signals in different multi-resolution levels. The SM-SSA was proposed in [26], is a data-driven nonstationary signal decomposition method, and was used in [27] for the diagnosis of sleep apnea, in [28] for the analysis of focal EEG signals. The SM-SSA extends the concept of the singular spectrum analysis (SSA) method to nonstationary signals with amplitude and frequency modulated (AM-FM) components. Due to the presence of a sliding window, SM-SSA can accurately estimate/track the signal information at different oscillatory scales [26]. Physiological signals (e.g., EEG and ECG) are nonstationary and require multiscale analysis that can be carried out using the SM-SSA method. In the second step, two distinct features, namely information potential (IP) and centered correntropy (CEC), are extracted from the obtained narrow-band RCs. These features characterize and discriminate different types of human emotions that have been classified using different ML techniques, namely, the K-nearest neighbor (KNN) and an SVM. The proposed method has been studied on two publicly available multimodal human emotion databases.

The rest of the paper is organized as follows: Section II describes the used databases. The methodology is described in section III. Section IV and V present the results and discussion, respectively followed by conclusion in section VI.

II. Databases Used for Emotion Recognition

A. DREAMER Database

DREAMER is a publicly available database comprising of ECG and EEG signal recordings during the influence of audio-visual stimuli from movie clips [12]. These movie clips feature scenes from various movies that elicit a wide range of emotions. From these 18 film clips, two of each targeted one of the following nine emotions: amusement, anger, excitement, fear, happiness, sadness, calmness, surprise, and disgust. Film clips of mean length $\approx$ 199 seconds were considered, which is adequate to elicit single emotions. On a five-point scale, twenty-three volunteers participated and self-evaluated the emotions in terms of dominance, arousal, and valence. The recorded EEG and ECG signals consist of fourteen (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4) and two channels, respectively. The EEG signal was sampled at 128 Hz, and the ECG signal was sampled at 256 Hz.

B. AMIGOS Database

AMIGOS is a publicly available database comprising ECG, EEG, and GSR signals recordings during the influence of audio-visual stimuli from movie clips [29]. Movies with a sufficiently long emotional segment ($\approx$ 20 min) were used. Forty volunteers (13 female) participated and self-evaluated the emotions in terms of dominance, arousal, and valence on a scale from 1 to 9 and selected basic emotions (neutral, sadness, happiness, surprise, anger, fear, and disgust) felt. The EEG and ECG signals consist of fourteen (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4) and two channels, respectively. The EEG and ECG signals were sampled at 128 Hz and 256 Hz, respectively. One's emotional state can change over time, primarily when more prolonged video stimuli are used. Only the recordings captured during the last 30 seconds of each film clip were considered and segmented into 2 seconds duration epochs for analysis from both the databases. This work uses a 3-D emotional model to classify emotions into high and low category. In the DREAMER and

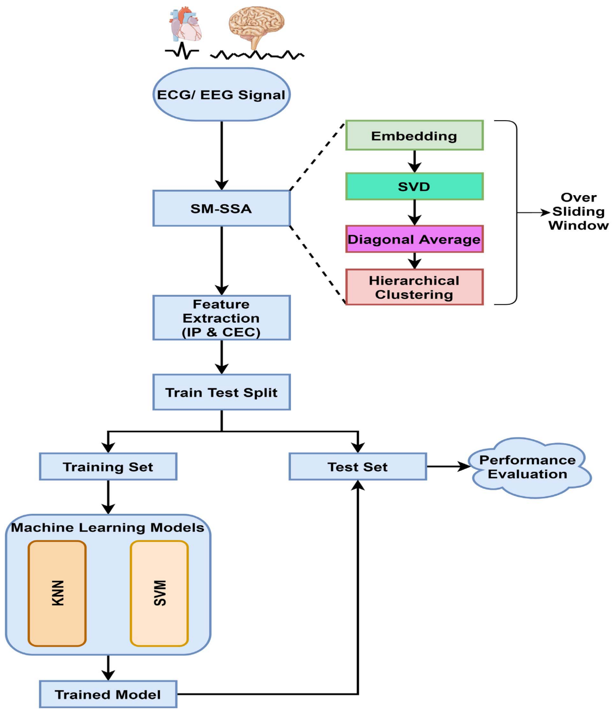


Fig. 1. Overview of the proposed model for emotion recognition.

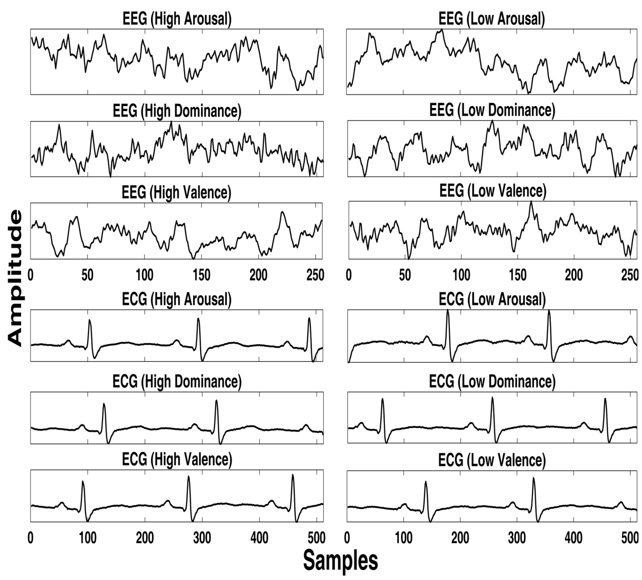


Fig. 2. Sample EEG and ECG signals of the DREAMER database that were recorded during the high and low emotional states.

AMIGOS database, annotation $\geq$ 4 and 7 is considered high, respectively. The class-specific sample EEG and ECG signals are shown in Fig. 2.

III. Methodology

A. Sliding Mode Singular Spectrum Analysis

The SM-SSA is an adaptive multi-component signal processing technique suitable for the analysis of non-stationary signals [26]. It decomposes signals with AM-FM components into a set of narrowband RCs. As a parametric method, it requires the number of components prior set at each sample of the analyzed signal. We have applied the automatic SSA (autoSSA) technique over an overlapping sliding window and decomposed emotional EEG and ECG signals into a set of informative narrowband RCs in this work. The steps followed for the evaluation of autoSSA of the signal $\{w(n),\ n = 1 \ldots N\}$ can be summarised as follows [26]: The one dimensional emotional EEG/ECG signal (N-samples) is mapped to a trajectory matrix S consisting of $K = N - L + 1$ lagged column vectors.

$$S = \begin{bmatrix} w(1) & w(2) & \cdots & w(K) \\ w(2) & w(3) & \cdots & w(K+1) \\ \vdots & \vdots & \ddots & \vdots \\ w(L) & w(L+1) & \cdots & w(N) \end{bmatrix} \tag{1}$$

The embedding dimension representing the length of each column vector is considered as L. The parameter L is an integer (in the range of $[2, N-1]$) and can be chosen depending on the time series's particularity. Afterward, the singular value decomposition (SVD) expands the $L \times K$ dimensional trajectory matrix S (rank $R \leq min(L,\ K)$) of emotion signal into a sum of weighted orthogonal matrices, expressed as:

$$S = UXV^T = \sum_{i=1}^{R} S_i \ with\ S_i = \sigma_i u_i v_i^T \tag{2}$$

where $X = diag(\sigma_1, \cdots, \sigma_R)$. The parameters u_i, v_i, and σ_i represent the left singular vector, right singular vector, and i^{th} singular value of the trajectory matrix S, respectively. The singular values are arranged in the descending order. In the next step, diagonal averaging also known as Hankelization process is performed to each ith matrix $S_i = [s(t, q)]$, t and q representing the indices of the elements of the matrix S_i. The i^{th} component $x_i(n)$ can be obtained by the diagonal averaging of the matrix S_i and represented as,

$$x_i(n) = \begin{cases} \dfrac{1}{n} \displaystyle\sum_{l=1}^{n} s(l, n-l+1), & \text{for} 1 \leq n < L. \\ \dfrac{1}{L} \displaystyle\sum_{l=1}^{L} s(l, n-l+1), & \text{for} L \leq n \leq K. \\ \dfrac{1}{N-n+1} \displaystyle\sum_{l=n-K+1}^{L} s(l, n-l+1), & \text{for} K+1 \leq n \leq N. \end{cases} \tag{3}$$

In the subsequent step of autoSSA, hierarchical clustering is performed on the components $(\{x_1(n), x_2(n), \cdots x_R(n)\})$ with a fixed r number of clusters. At the end of the clustering, the components belonging to m^{th} cluster are assigned as the index T_m where $m = 1, 2, \cdots r$. Finally, each cluster components are summed up for obtaining m^{th} RC of emotion EEG/ECG signal which is mathematically represented as,

$$w_m(n) = \sum_{i \in T_m} x_i(n) \tag{4}$$

In conventional SSA, the trajectory matrices in equation (2) are used for grouping, in which singular vectors contain the information [30]. It is presumed that singular vectors are belonging to a particular group share some common characteristic/ information of the signal component. Therefore,

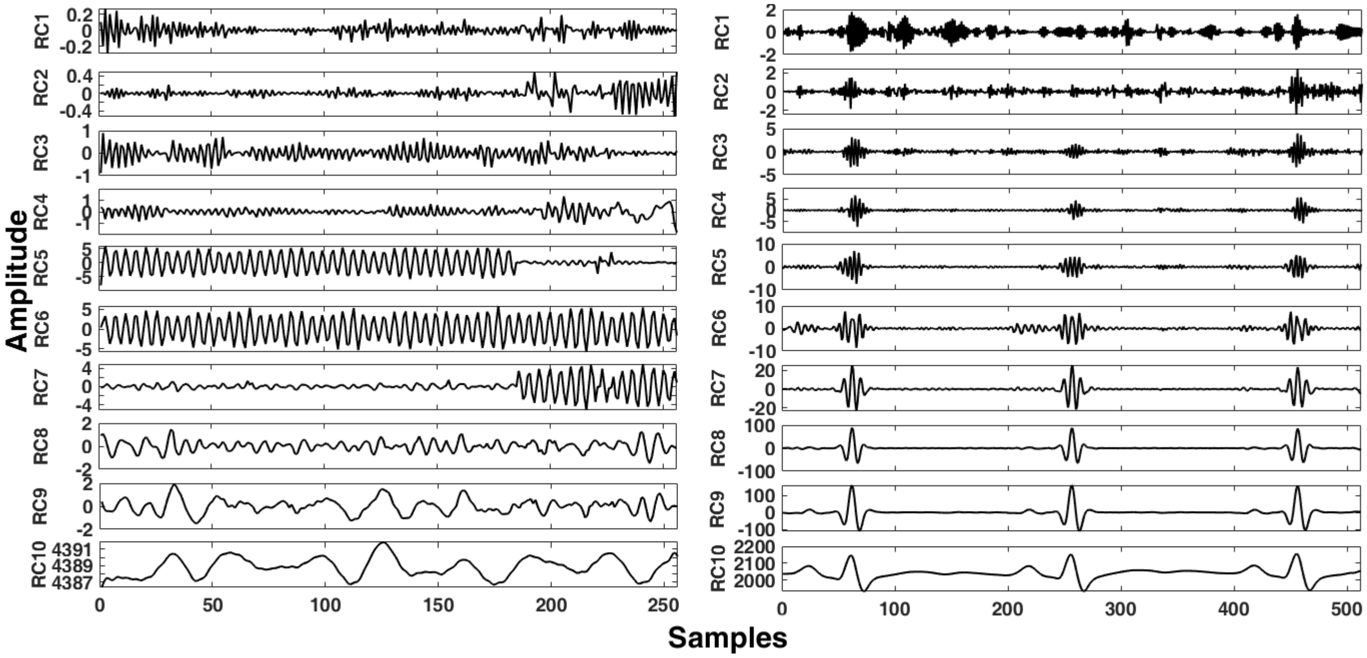


Fig. 3. Extracted RCs for EEG (left column) and ECG (right column) signal using SM-SSA.

the correlation of singular vectors and singular spectrum coherence are taken into consideration while grouping [31]. In contrary to this, autoSSA uses hierarchical clustering in the final step for obtaining the RCs. Finally, the computation steps for SM-SSA for decomposing the emotion signals can be summarised through the following steps: (I) A M length sliding window is considered, which takes a step size of δ samples for reaching the next frame. The number of RCs at each sample is considered $p(n)$, having the same length of the input signal $w(n)$. (II) With the total number of frames $N-M+1$, the frame index f changes its value from 1 to $N-M+1$ by taking a step size of δ, specifying the center (n_c) and end (n_e) of each frame as $f+\frac{(M-1)}{2}$ and $f-M+1$, respectively. (III) Considering the embedding dimension L, the RCs are computed from each frame using the autoSSA algorithm. For every frame, we sort the computed RCs in accordance with their mean frequencies [32]. (IV) For the sample range of $[1, n_c]$ (belonging to frame index $f=1$), the components are directly assigned with the sorted RCs extracted using autoSSA. (V) For subsequent frames ($f>1$), the values of the components at sample index $f-1+n_c$ are assigned as the values of the respective RCs at sample index n_c. The step (V) continues till we reach the last frame ($f=N-M+1$), for which the components in the sample range $[f-1+n_c, N]$ are assigned the value of RCs in the sample range $[n_c, M]$ obtained using autoSSA. The SM-SSA method discussed has been employed for decomposing emotion physiological signals into RCs, facilitating the computation of features in multiple oscillatory levels of the analyzed signals. In this work, $L=10$ and $M=57$ have been chosen, which resulted in improved classification performance. Fig. 3 presents the extracted 10 RCs for both EEG (left column) and ECG (right column) signals using SM-SSA method. It can be observed that RCs contain information in multiple resolution levels of the signal, which can be crucial for analyzing human emotions. The comparative singular spectrum plots for high/low dimensions ($L=10$, $M=57$) of emotions are shown in Fig. 4. These plots represent energy contribution by i-th eigentriple (σ_i, u_i, v_i) in the trajectory matrix [26]. The dissimilarities between the high and low emotion dimensions can be observed in the figures, which have been unitized in this work for SM-SSA-based emotion recognition.

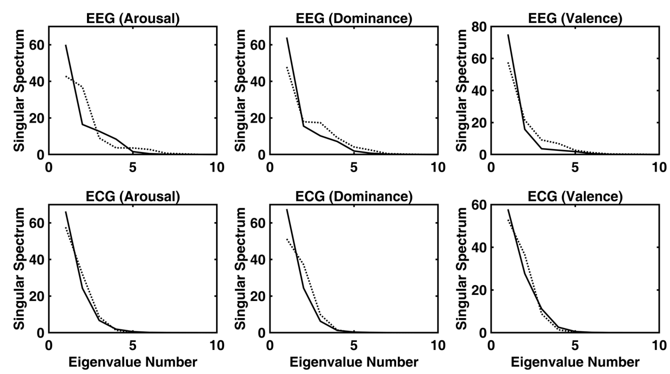


Fig. 4. Singular spectrum plots for the signals recorded during different emotional states (solid line: high emotion, dotted line: low emotion).

B. Feature Extraction From the RCs

The computed features reduce the input data's dimension and become more useful in discriminating the human emotional states. One of the most efficient methods for EEG-based emotion recognition has been demonstrated to be entropy-based pattern learning. The human brain is an inherently complex and dynamic system. Entropy measures can be used to assess the irregularity, randomness, and complexity of physiological data [33]. They have a potent ability to extract information from EEG/ECG waves indicating clinical/physiological implications [34]. Inspired from the facts as mentioned earlier, we have computed entropy-based discriminatory features, namely IP and CEC, from each RC of the considered physiological signals; they are described as follows:

1) Information Potential (IP): IP is the non-parametric estimator of Renyi's quadratic entropy [14]. For a random variable x_n, IP can be estimated using equation (5) [35].

$$IP(X) = \frac{1}{N^2}\sum_{p=1}^{N}\sum_{q=1}^{N} K_\sigma(x_q - x_p) \tag{5}$$

Here, x_p and x_q are the p^{th} and q^{th} samples of the data set. N is the total number of samples and $K_\sigma(x_q - x_p)$ represents the Gaussian kernel function. IP is invariant to the mean of the underlying density of the samples. Due to this, when IP is utilized for supervised learning, the mean of the error signal is not necessarily zero, which is a requirement for most applications.

2) Centered Correntropy (CEC): Correntropy (CoC) is a generalization of correlation that extracts the second-order information and higher-order moments of the joint distribution. It is a generalized similarity measure and computes nonlinear correlation among multiple delayed samples of the signal [36].

$$CoC(X) = \frac{1}{N-m+1}\sum_{n=m}^{N} K_\sigma(x_n - x_{n-m}) \tag{6}$$

$$C\hat{o}C(X) = \frac{1}{N^2}\sum_{n=1}^{N}\sum_{n=m}^{N} K_\sigma(x_n - x_{n-m}) \tag{7}$$

Here, m represents delay. To lessen the influence of DC bias, the mean value of the correntropy $C\hat{o}C(X)$ is subtracted from the $CoC(X)$ to attain the centered correntropy $(CEC(l))$. In this work, we have used $m = 2$ and the Gaussian kernel function.

$$CEC(l) = CoC(X) - C\hat{o}C(X) \tag{8}$$

C. Detection of Human Emotions Using Classifiers

Binary classification is the way toward arranging the given set into two groups utilizing a classification strategy. In this work, classification of different emotional states has been performed using the features extracted from the RCs. The generated feature matrix can be denoted as $F \in R^{m\times n}$; m is the number of instances representing high and low emotional states, n is the number of computed features for classification. The j^{th} instance is considered as 0 for the low emotional state and 1 for the high emotional state, respectively. The training and testing of the classifiers have been carried out using the ten-fold cross-validation method. Furthermore, classification performance parameters, namely accuracy (Acc = (TP+TN)/(TP+FN+TN+FP)), sensitivity (Se = TP/(TP+FN)), specificity (Sp = TN/(TN+FP)), false-positive rate (FPR = FP/(FP+TN)), false-negative rate (FNR = FN/(TP+FN)), precision (Pr = TP/(TP + FP)), and J statistic (J stat% = Se% + Sp% −100) have been computed for evaluating the performance of the proposed method. These performance parameters are computed from true-positive (TP), false-positive (FP), true-negative (TN), and false-negative (FN) values.

We have used two different classifiers, namely KNN and SVM, to classify high and low arousal, valence, and dominance states. The KNN is a supervised, non-parametric, and instance-based lazy ML algorithm based on the concept of 'K' nearest neighbours. The use of the local approximation method simplifies its learning process [37]. In contrast, SVM constructs a hyperplane for the separation of data into different classes based on support vectors, the distance between support vectors, and decision boundary [38]. Further, to handle real-life data samples with non-linear decision boundaries, SVM uses kernel functions that transform the data into multi-dimensional space.

IV. Experimental Results

This work was implemented on MATLAB R2020b in a computer with an Intel i5 processor @ 1.8 GHz and 16 GB RAM. Two supervised ML models were tested for the classification of emotional states. Classifier parameters were optimized following the grid search method based on the best classification accuracy. The KNN classifier was employed with the Euclidean, Minkowski distance functions, and K ∈ {1, 3, 5, 7, 9, 11} were used. For the SVM classifier, the features were mapped using the radial basis function (RBF) kernel. The kernel scale (σ) parameter was varied from 0.5 to 10 in an interval of 0.5. Table I and II present the experimental results for the proposed emotion recognition method on the DREAMERS and AMIGOS databases, respectively. The tables exhibit the classification performance in terms of the Acc, FNR, FPR, and Pr for all three emotional dimensions: arousal, dominance, and valance.

Table I shows that in the DREAMER database, KNN outperformed SVM in classifying low and high states of each of the three emotional dimensions; the obtained classification Accs are 92.06%, 92.38%, and 92.3%, respectively. The performance of the SVM classifier is marginally less than KNN in classifying the emotional states. The obtained Accs using SVM are 92.01%, 92.29%, and 91.53% in classifying low and high states of arousal, dominance, and valance, respectively, which are less than 1% compared to the corresponding accuracies obtained using the KNN classifier. Comparing computational complexities, for 'm' number of features and 'n' number of training samples, the KNN with Euclidean distance has an overall complexity of $O(mn + n\log(n))$ [39], whereas, Abdiansah and Wardoyo [40] deduced that SVM has a complexity of $O(n^3)$. The classification performance was also analyzed in terms of the FPR and FNR, which should be low for better classification. One can observe that the KNN classifier achieved FNR and FPR of 8.61% and 6.7% in classifying low and high dominance states. The reported FNR and FPR values using KNN for the other two binary classification tasks were also significantly low. The obtained FNR and FPR values using the SVM classifier are slightly higher than the corresponding values using the KNN classifier. However, the highest FPR and the lowest FNR of 16.2% and 3.45% were obtained using the SVM classifier in classifying low and high valance states. The highest Pr of 94.03% was

TABLE I
EXPERIMENTAL RESULTS OBTAINED USING PROPOSED EMOTIONS RECOGNITION METHOD ON DREAMER DATABASE

E	Model	Acc%	FNR%	FPR%	Pr%
A	KNN (K=3, Euclidean)	**92.06**	10.53	5.92	92.15
	SVM ($\sigma = 1$, RBF)	92.01	11.53	5.24	92.92
D	KNN (K=3, Euclidean)	**92.38**	8.61	6.7	92.66
	SVM ($\sigma = 1$, RBF)	92.29	8.61	6.88	92.47
V	KNN (K=3, Euclidean)	**92.3**	11.08	5.5	91.31
	SVM ($\sigma = 1$, RBF)	91.53	16.2	3.45	94.03

E=Emotions, A=Arousal, D=Dominance, V=Valance.

TABLE II
EXPERIMENTAL RESULTS OBTAINED USING PROPOSED EMOTIONS RECOGNITION METHOD ON AMIGOS DATABASE

E	Model	Acc%	FNR%	FPR%	Pr%
A	KNN (K=3, minkowski)	**92.01**	39.11	3.17	74.8
	SVM ($\sigma = 1$, RBF)	89.87	14.25	6.92	90.58
D	KNN (K=3, Euclidean)	**89.65**	25.15	6.06	78.17
	SVM ($\sigma = 1.5$, RBF)	87.31	47.49	2.6	85.41
V	KNN (K=3, Euclidean)	**85.26**	35.64	7.71	73.72
	SVM ($\sigma = 1.5$, RBF)	81.97	67.01	1.57	87.57

E=Emotions, A=Arousal, D=Dominance, V=Valance.

achieved using the SVM classifier in classifying low and high valance states.

The proposed method is also studied on the AMIGOS database, and the results are illustrated in Table II. One can see that KNN achieved the classification Accs of 92.01%, 89.65%, and 85.26% in classifying low and high states of arousal, dominance, and valence, respectively. The corresponding classification Accs obtained using an SVM classifier are 89.87%, 87.31%, and 81.97%, respectively, which are around 2% lesser than the obtained Acc values using the KNN classifier. In the table, one can also notice that though the obtained FPR values are low, both the classifiers report high FNR values. The SVM achieved the highest Pr of 90.58% in classifying low and high arousal states. The corresponding FPR and FNR values are 6.92% and 14.25%, respectively.

Fig. 5 and 6 further analyze the classification performance in terms of the Se, Sp, and J stat% values obtained using the proposed method on DREAMER and AMIGOS databases, respectively. J stat% is an index that gives equal weightage to false negative and false positive values [41]. It can be seen that obtained Se, Sp, and J stat% values are higher using both KNN and SVM classifier in the DREAMER database compared to the values obtained in the AMIGOS database. Looking at the J stat% values, it is evident that KNN is the best-suited classification model for all the three emotional dimensions on DREAMER and AMIGOS databases.

The performance of the proposed emotion detection method is further verified using the receiver operating characteristics (ROC) curves. The ROC curve is a standard performance measure for classification problems, and the area under the

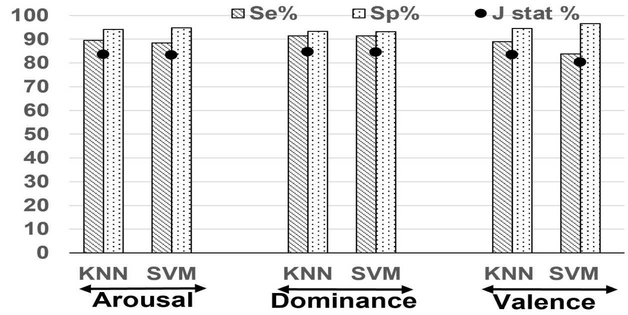


Fig. 5. Se%, Sp%, and J stat% of best fitted models for emotion recognition on DREAMER database.

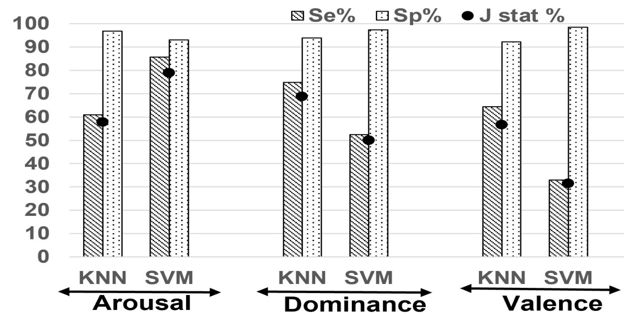
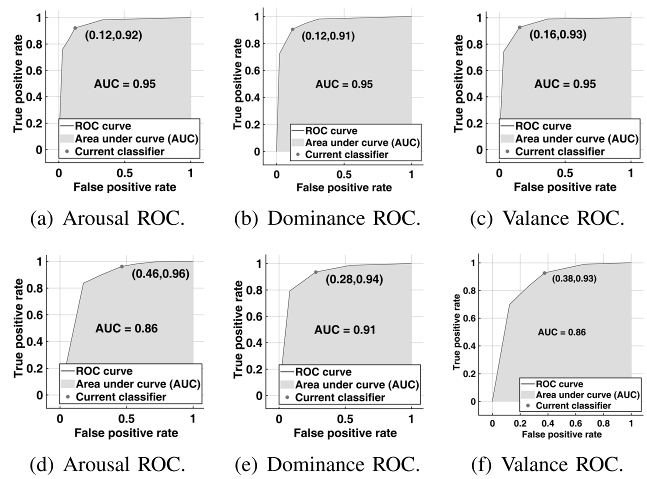


Fig. 6. Se%, Sp%, and J stat% of best fitted models for emotion recognition on AMIGOS database.



(a) Arousal ROC. (b) Dominance ROC. (c) Valance ROC.
(d) Arousal ROC. (e) Dominance ROC. (f) Valance ROC.

Fig. 7. ROC for best fitted models for emotion recognition. (a)–(c): DREAMER, (d)–(f): AMIGOS.

curve (AUC) represents the degree of separability. The higher value of the AUC confirms the better performance of the classifier. Fig. 7 shows the ROC plots of best-fitted models (tuning parameters mentioned in Table I and II) for both the databases and for all three emotional dimensions. In the DREAMER database, AUC was observed as 0.95 for all three emotional dimensions. In the AMIGOS database, AUC was computed as 0.86 in classifying low and high arousal and valence states, whereas, for dominance, AUC was relatively high (0.91).

A. Experiments on Tuning Parameters of SM-SSA

Embedding dimension (L) and sliding window (M) are two crucial tuning parameters of SM-SSA. The proposed method was evaluated for different values of these parameters, and the

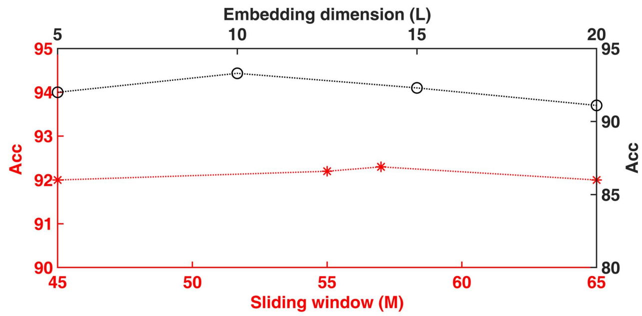


Fig. 8. Plot of variation of emotion classification Acc for different tuning parameters of SM-SSA method.

TABLE III
THE ROLE OF RCS IN CLASSIFYING EMOTIONAL STATES

	Acc%	
Emotion	Without SM-SSA	With SM-SSA
Arousal	76.40	92.06
Dominance	77.60	92.38
Valence	75.50	93.30

results are presented graphically in Fig. 8. An improved result was obtained for $M = 57$ and $L = 10$; thus, these values were used throughout this work.

B. Experiment to Analyze the Role of RCs

This method uses SM-SSA for the decomposition of physiological signals before feature extraction and classification. To emphasize the role of RCs in emotion recognition, feature extraction and classification have been performed without employing the decomposition method (using the same parameters as described in Table I). The results are presented in Table III. The improved results with SM-SSA signify the effectiveness of RCs in emotion recognition.

C. Spatial Analysis of the Emotional Features

We have generated the topographic maps using the features computed from EEG signals to investigate the spatial brain activity during the emotional states, namely, high arousal (HA), low arousal (LA), high dominance (HD), low dominance (LD), high valance (HV), and low valence (LV). Topographic maps of a randomly selected subject from DREAMER database are shown in Fig. 9 and 10. Whereas, for the AMIGOS database, they are presented in Fig. 11 and 12. These figures exhibit activity level variations in various areas of the brain. In Fig. 9, one can observe a noticeable difference throughout the cerebral cortex in IP value for arousal and occipital lobe for valence. Fig. 10 depicts low CEC for HA and high CEC in case of LV. Throughout the cerebral cortex, Fig. 11 and 12 show changes in IP and CEC features for arousal and in the frontal lobe for dominance and valence. These plots can be considered as visual features for analyzing human emotions.

D. Comparison With Existing State-of-the-Art Works

In Table IV, the proposed method's emotion classification performance is compared with other existing ML/ deep learning (DL) based emotion classification methods. The table

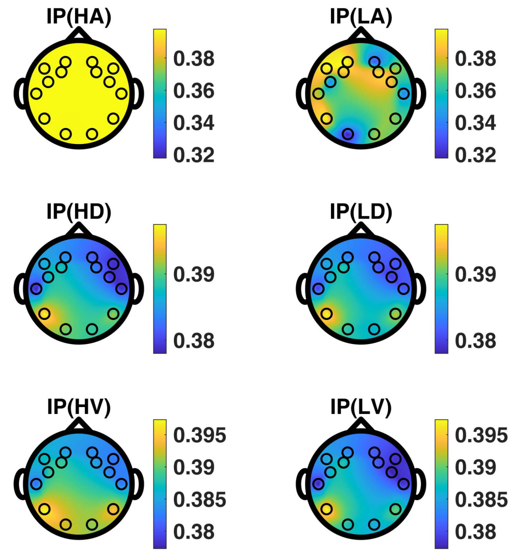


Fig. 9. DREAMER: Topographic map for IP feature.

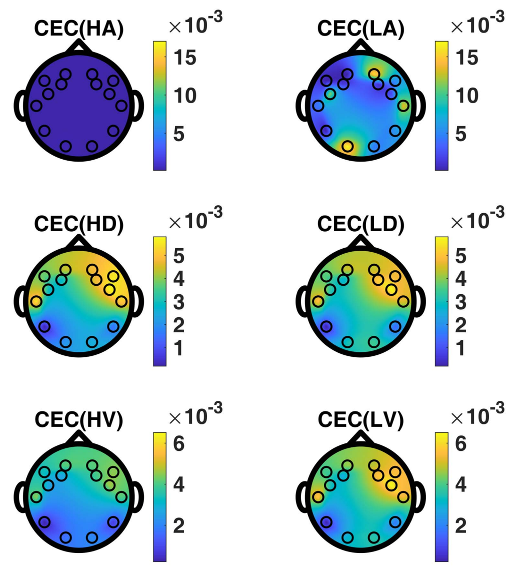


Fig. 10. DREAMER: Topographic map for CEC feature.

presents details of the databases, method used, and performance in terms of Acc. Harper and Southern [4] proposed Bayesian DL based model for the prediction of emotions. They achieved the highest classification Acc of 90% in classifying high and low valence states on the AMIGOS database. Santamaria-Granados *et al.* [42] proposed a model for emotion detection using TF based features and convolutional neural network (CNN). They reported the highest Acc of 76% for arousal detection on the AMIGOS database. Song *et al.* [43] presented a method for emotion recognition using dynamical graph CNN (DGCNN) and achieved an average Acc of $\approx 85\%$ on DREAMER database. Cheng *et al.* [44] proposed an emotion recognition model using deep forest that has the advantage of lesser training time and hyperparameter tuning requirement compared to deep neural networks (DNNs). The authors obtained the highest Acc of 90.41% in detecting arousal states. Katsigiannis and Ramzan [12] developed the DREAMER database and used an ML-based

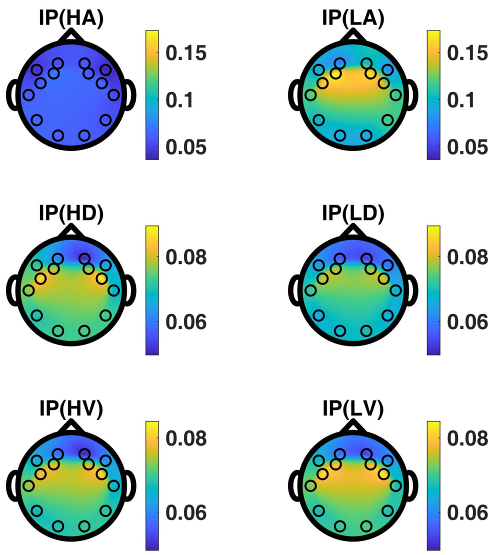


Fig. 11. AMIGOS: Topographic map for IP feature.

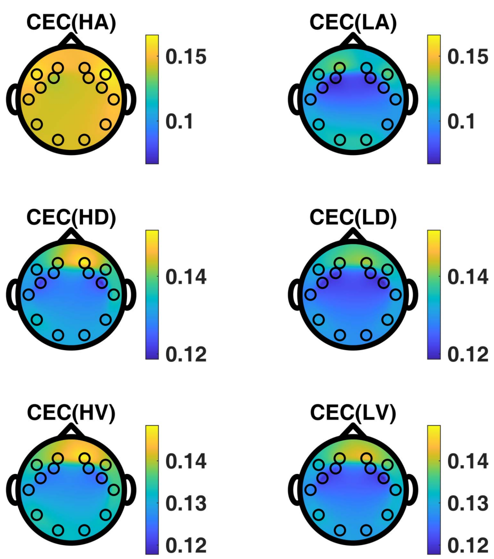


Fig. 12. AMIGOS: Topographic map for CEC feature.

model for emotion recognition using ECG and EEG signals features. They reported a classification Acc of $\approx 62\%$ using the SVM classifier. Bhattacharyya *et al.* [6] reported an average Acc of $\approx 85\%$ in emotion recognition using multivariate MFBSE-EWT and ARF classifier. Li *et al.* [45] used a spectrogram image to capture the signal's TF information. Finally, attention-based bidirectional long short-term memory recurrent neural networks followed by DNN have been used to predict the emotions. Our model has achieved the highest classification Acc of 92.38% in classifying high and low dominance states. It can be noticed that the proposed method has outperformed the existing state-of-art works in recognizing all the emotional states.

V. Discussion

This work proposes an automated human emotion recognition system that can categorize arousal, dominance, and valence into high or low levels. During different emotional

TABLE IV

Comparative Analysis of Proposed Method's Performance to That of Other Existing Methods in Terms of Acc Values

Method	Database	HA-LA	HD-LD	HV-LV
IP, CEC Features, ML Models (This Work)	AMIGOS	92.01	89.65	85.26
	DREAMER	92.06	92.38	92.30
Bayesian DL [4]	AMIGOS	–	–	90.00
	DREAMER	–	–	86.00
TF Features, CNN Models [42]	AMIGOS	76.00	–	75.00
DGCNN [43]	DREAMER	84.54	85.02	86.23
Deep Forest [44]	DREAMER	90.41	89.89	89.03
ECG-EEG Features, ML Models [12]	DREAMER	62.32	61.84	62.49
MFBSE-EWT, ARF [6]	DREAMER	85.40	84.50	86.20
Spectrogram, DL Models [45]	AMIGOS	83.30	–	79.40

HA = High Arousal, LA= Low Arousal, HD = High Dominance, Low Dominance, HV = High Valance, LV= Low Valence.

states, the powers of EEG rhythms over the scalp either decrease or increase [46]–[48]. In the high arousal state, the power levels for δ, θ, and γ-rhythms increase at the posterior location [46], [48]. Whereas the reduction in α-rhythm power level was quantified at high arousal state [47]. Similarly, in response to angry emotion stimuli, the mid-line EEG channels experience high δ-rhythm power level [49]. The reduction in γ-rhythm activity is observed with increase in valance level for spectral component above 51 Hz [50]. The increased power level was reported for θ-rhythms of frontal-medial EEG channels with positive music stimuli [51]. Further, it is worth mentioning that physiological signals, namely, EEG and ECG, exhibit nonlinear and nonstationary properties that can better be analyzed using a signal adaptive decomposition method. Furthermore, the appearance of emotional reactions of the subjects is time-varying and shows frequency dependency in the recorded signals [52]. Thus, in order to meet the challenges as mentioned earlier and to retrieve emotional information, we employ the subspace-based nonstationary signal decomposition method, SM-SSA. The SM-SSA captures the dynamics and nonstationarity of emotion signals over a sliding window using the formed trajectory matrix, inherent SVD, and hierarchical clustering. To the best of the author's knowledge, the SM-SSA has been explored for the first time in this work for extracting frequency-dependent emotional information (in the form of RCs) from the physiological signals. In contrast to the traditional frequency bands (e.g. EEG rhythms), the RCs are adaptively extracted based on the information present in the signal. Afterward, the IP and CEC features are computed from the RCs that measure the key emotional information over different signal adaptive oscillatory scales. The boxplots of randomly selected IP and CEC features for all three emotional states are shown in Fig. 13. These plots exhibit the inter-class distribution of the considered features. One can notice that interquartile ranges of the features are different for different high and low emotion dimensions, which lead to improved classification performance. Finally, the extracted emotion features are classified using supervised ML models with tuned parameters. A comprehensive study has been carried out to select the model parameters to achieve the best classification accuracy. The experimental outcomes substantiate the usefulness of the proposed SM-SSA-based framework in classifying the emotions of the 3D emotional space model.

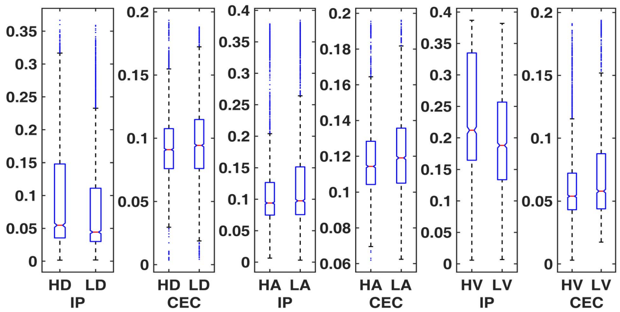


Fig. 13. Boxplots depicting the separability of selected features for the considered emotional states. (HD, LD, HA, LA, HV, and LV stand for high dominance, low dominance, high arousal, low arousal, high valence, and low valence, respectively).

It is worth mentioning here that various DL based models have been unitized in the literature for emotion recognition [4], [23], [43], [53]–[55]. However, armed with the advantages of automated feature engineering, DL-based models have limitations like: Difficulty in comprehending the output of a DL model; hence interpretation and study of the significance of the extracted features are arduous. In addition, due to the complexity of the data models, training of DL models is quite costly and requires system with high computing capabilities.

On the other hand, our work has the following limitations that can be the future scope for further research: Optimization techniques can be used for the setting of model parameters. Noise and artifact removal techniques can be employed as a preprocessing step. Additionally, despite the promising potential of the proposed SM-SSA-based emotion recognition framework, further validation of its reliability is necessary that can be carried out over a large-scale experimental database.

VI. Conclusion

This work presents a computerized multi-modal approach for human emotion recognition using the SM-SSA method. The physiological signals, namely, EEG and ECG, were decomposed using SM-SSA into RCs, which were further used to extract two discriminatory features, namely, IP and CEC. All the recorded channels were used for feature extraction, and finally, the extracted features were fed to ML classifiers, namely, KNN and SVM. The proposed method was studied on two publicly available standard human emotion analysis databases: AMIGOS and DREAMER. The dimensions (arousal, dominance, and valance) of the 3-D emotional space model were classified into a high or low category. Extensive experiments demonstrate that SM-SSA-based features with KNN classifier achieved the highest classification Acc of 92.38% in classifying high and low dominance classes. The performance of the proposed method has shown improvement compared to the existing state-of-art methods. The analysis indicates that the proposed method can discriminate human emotional states using physiological signals implemented in a software system.

References

[1] T. Sharot, M. R. Delgado, and E. A. Phelps, "How emotion enhances the feeling of remembering," *Nature Neurosci.*, vol. 7, no. 12, pp. 1376–1380, Dec. 2004.

[2] T. Brosch, G. Pourtois, and D. Sander, "The perception and categorisation of emotional stimuli: A review," *Cognition Emotion*, vol. 24, no. 3, pp. 377–400, Apr. 2010.

[3] T. Brosch, G. Pourtois, D. Sander, and P. Vuilleumier, "Additive effects of emotional, endogenous, and exogenous attention: Behavioral and electrophysiological evidence," *Neuropsychologia*, vol. 49, no. 7, pp. 1779–1787, Jun. 2011.

[4] R. Harper and J. Southern, "A Bayesian deep learning framework for end-to-end prediction of emotion from heartbeat," *IEEE Trans. Affect. Comput.*, early access, Mar. 20, 2020, doi: 10.1109/TAFFC.2020.2981610.

[5] A. V. Bocharov, G. G. Knyazev, and A. N. Savostyanov, "Depression and implicit emotion processing: An EEG study," *Neurophysiol. Clin.*, vol. 47, no. 1, pp. 225–230, 2017.

[6] A. Bhattacharyya, R. K. Tripathy, L. Garg, and R. B. Pachori, "A novel multivariate-multiscale approach for computing EEG spectral and temporal complexity for human emotion recognition," *IEEE Sensors J.*, vol. 21, no. 3, pp. 3579–3591, Feb. 2021.

[7] P. Song and W. Zheng, "Feature selection based transfer subspace learning for speech emotion recognition," *IEEE Trans. Affect. Comput.*, vol. 11, no. 3, pp. 373–382, Jul. 2020.

[8] W. Zheng, X. Zhou, C. Zou, and L. Zhao, "Facial expression recognition using kernel canonical correlation analysis (KCCA)," *IEEE Trans. Neural Netw.*, vol. 17, no. 1, pp. 233–238, Jan. 2006.

[9] T. Song, W. Zheng, C. Lu, Y. Zong, X. Zhang, and Z. Cui, "MPED: A multi-modal physiological emotion database for discrete emotion recognition," *IEEE Access*, vol. 7, pp. 12177–12191, 2019.

[10] L. Shu *et al.*, "A review of emotion recognition using physiological signals," *Sensors*, vol. 18, no. 7, p. 2074, Jun. 2018.

[11] H. Jebelli, S. Hwang, and S. Lee, "EEG-based workers' stress recognition at construction sites," *Autom. Construct.*, vol. 93, pp. 315–324, Sep. 2018.

[12] S. Katsigiannis and N. Ramzan, "DREAMER: A database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices," *IEEE J. Biomed. Health Inform.*, vol. 22, no. 1, pp. 98–107, Jan. 2018.

[13] C. P. Niemic, "Studies of emotion: A theoretical and empirical review of psychophysiological studies of emotion," *J. Undergraduate Res.*, vol. 1, pp. 15–18, Oct. 2002.

[14] V. Gupta, M. D. B. Chopda, and R. B. Pachori, "Cross-subject emotion recognition using flexible analytic wavelet transform from EEG signals," *IEEE Sensors J.*, vol. 19, no. 6, pp. 2266–2274, Mar. 2019.

[15] D. Maheshwari, S. K. Ghosh, R. K. Tripathy, M. Sharma, and U. R. Acharya, "Automated accurate emotion recognition system using rhythm-specific deep convolutional neural network technique with multi-channel EEG signals," *Comput. Biol. Med.*, vol. 134, Jul. 2021, Art. no. 104428.

[16] M. Soleymani, M. Pantic, and T. Pun, "Multimodal emotion recognition in response to videos," *IEEE Trans. Affect. Comput.*, vol. 3, no. 2, pp. 211–223, Apr. 2012.

[17] M. Li and B.-L. Lu, "Emotion classification based on gamma-band EEG," in *Proc. Annu. Int. Conf. Eng. Med. Biol. Soc.*, Sep. 2009, pp. 1223–1226.

[18] D. Iacoviello, A. Petracca, M. Spezialetti, and G. Placidi, "A classification algorithm for electroencephalography signals by self-induced emotional stimuli," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 3171–3180, Dec. 2016.

[19] S. K. Hadjidimitriou and L. J. Hadjileontiadis, "EEG-based classification of music appraisal responses using time-frequency analysis and familiarity ratings," *IEEE Trans. Affect. Comput.*, vol. 4, no. 2, pp. 161–172, 2013.

[20] H. J. Yoon and S. Y. Chung, "EEG-based emotion estimation using Bayesian weighted-log-posterior function and perceptron convergence algorithm," *Comput. Biol. Med.*, vol. 43, no. 12, pp. 2230–2237, 2013.

[21] R.-N. Duan, J.-Y. Zhu, and B.-L. Lu, "Differential entropy feature for EEG-based emotion classification," in *Proc. 6th Int. IEEE/EMBS Conf. Neural Eng. (NER)*, Oct. 2013, pp. 81–84.

[22] W.-L. Zheng and B.-L. Lu, "Investigating critical frequency bands and channels for EEG-based emotion recognition with deep neural networks," *IEEE Trans. Auton. Mental Develop.*, vol. 7, no. 3, pp. 162–175, Sep. 2015.

[23] T. Zhang, W. Zheng, Z. Cui, Y. Zong, and Y. Li, "Spatial–temporal recurrent neural network for emotion recognition," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 839–847, Mar. 2019.

[24] X. Li, D. Song, P. Zhang, Y. Zhang, Y. Hou, and B. Hu, "Exploring EEG features in cross-subject emotion recognition," *Frontiers Neurosci.*, vol. 12, p. 162, May 2018.

[25] A. Mert and A. Akan, "Emotion recognition based on time–frequency distribution of EEG signals using multivariate synchrosqueezing transform," *Digit. Signal Process.*, vol. 81, pp. 106–115, Oct. 2018.
[26] J. Harmouche, D. Fourer, F. Auger, P. Borgnat, and P. Flandrin, "The sliding singular spectrum analysis: A data-driven nonstationary signal decomposition tool," *IEEE Trans. Signal Process.*, vol. 66, no. 1, pp. 251–263, Jan. 2018.
[27] H. Singh, R. K. Tripathy, and R. B. Pachori, "Detection of sleep apnea from heart beat interval and ECG derived respiration signals using sliding mode singular spectrum analysis," *Digit. Signal Process.*, vol. 104, Sep. 2020, Art. no. 102796.
[28] T. Siddharth, R. K. Tripathy, and R. B. Pachori, "Discrimination of focal and non-focal seizures from EEG signals using sliding mode singular spectrum analysis," *IEEE Sensors J.*, vol. 19, no. 24, pp. 12286–12296, Jun. 2019.
[29] J. A. Miranda-Correa, M. K. Abadi, N. Sebe, and I. Patras, "AMIGOS: A dataset for affect, personality and mood research on individuals and groups," *IEEE Trans. Affect. Comput.*, vol. 12, no. 2, pp. 479–493, Apr. 2021.
[30] H. Hassani, "Singular spectrum analysis: Methodology and comparison," *J. Data Sci.*, vol. 5, pp. 239–257, Apr. 2007.
[31] A. M. Álvarez-Meza, C. D. Acosta-Medina, and G. Castellanos-Domínguez, "Automatic singular spectrum analysis for time-series decomposition," in *Proc. ESANN*, 2013, pp. 1–10.
[32] B. Tacer and P. J. Loughlin, "Instantaneous frequency and time-frequency distributions," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, vol. 2, Oct. 1995, pp. 1013–1016.
[33] O. Faust and M. G. Bairy, "Nonlinear analysis of physiological signals: A review," *J. Mech. Med. Biol.*, vol. 12, no. 4, 2012, Art. no. 1240015.
[34] Y. Lu, M. Wang, W. Wu, Y. Han, Q. Zhang, and S. Chen, "Dynamic entropy-based pattern learning to identify emotions from EEG signals across individuals," *Measurement*, vol. 150, Jan. 2020, Art. no. 107003.
[35] D. Xu and D. Erdogmuns, "Renyi's entropy, divergence and their non-parametric estimators," in *Information Theoretic Learning* (Information Science and Statistics). New York, NY, USA: Springer, 2010, doi: 10.1007/978-1-4419-1570-2_2.
[36] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: Properties and applications in non-Gaussian signal processing," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.
[37] D. W. Aha, D. Kibler, and M. K. Albert, "Instance-based learning algorithms," *Mach. Learn.*, vol. 6, no. 1, pp. 37–66, 1991.
[38] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[39] S. Ray, "An analysis of computational complexity and accuracy of two supervised machine learning algorithms—K-nearest neighbor and support vector machine," in *Data Management, Analytics and Innovation* (Advances in Intelligent Systems and Computing), vol. 1174, N. Sharma, A. Chakrabarti, V. Balas, and J. Martinovic, Eds. Singapore: Springer, 2021, doi: 10.1007/978-981-15-5616-6_24.
[40] A. Abdiansah and R. Wardoyo, "Time complexity analysis of support vector machines (SVM) in LibSVM," *Int. J. Comput. Appl.*, vol. 128, no. 3, pp. 28–34, Oct. 2015.
[41] E. F. Schisterman, N. J. Perkins, A. Liu, and H. Bondell, "Optimal cut-point and its corresponding youden index to discriminate individuals using pooled blood samples," *Epidemiology*, vol. 16, no. 1, pp. 73–81, Jan. 2005.
[42] L. Santamaria-Granados, M. Munoz-Organero, G. Ramirez-González, E. Abdulhay, and N. Arunkumar, "Using deep convolutional neural network for emotion detection on a physiological signals dataset (AMIGOS)," *IEEE Access*, vol. 7, pp. 57–67, 2018.
[43] T. Song, W. Zheng, P. Song, and Z. Cui, "EEG emotion recognition using dynamical graph convolutional neural networks," *IEEE Trans. Affect. Comput.*, vol. 11, no. 3, pp. 532–541, Jul./Sep. 2020.
[44] J. Cheng *et al.*, "Emotion recognition from multi-channel EEG via deep forest," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 2, pp. 453–464, Feb. 2021.
[45] C. Li, Z. Bao, L. Li, and Z. Zhao, "Exploring temporal representations by leveraging attention-based bidirectional LSTM-RNNs for multi-modal emotion recognition," *Inf. Process. Manage.*, vol. 57, no. 3, May 2020, Art. no. 102185.
[46] M. A. Klados *et al.*, "A framework combining delta event-related oscillations (EROs) and synchronisation effects (ERD/ERS) to study emotional processing," *Comput. Intell. Neurosci.*, vol. 2009, Jul. 2009, Art. no. 549419.
[47] R. J. Barry, A. R. Clarke, S. J. Johnstone, and C. R. Brown, "EEG differences in children between eyes-closed and eyes-open resting conditions," *Clin. Neurophysiol.*, vol. 120, no. 10, pp. 1806–1811, Oct. 2009.
[48] L. I. Aftanas, A. A. Varlamov, S. V. Pavlov, V. P. Makhnev, and N. V. Reva, "Time-dependent cortical asymmetries induced by emotional arousal: EEG analysis of event-related synchronization and desynchronization in individually defined frequency bands," *Int. J. Psychophysiol.*, vol. 44, no. 1, pp. 67–82, Apr. 2002.
[49] C. M. Krause, V. Viemerö, A. Rosenqvist, L. Sillanmäki, and T. Åström, "Relative electroencephalographic desynchronization and synchronization in humans to emotional film content: An analysis of the 4–6, 6–8, 8–10 and 10–12 Hz frequency bands," *Neurosci. Lett.*, vol. 286, no. 1, pp. 9–12, May 2000.
[50] B. Reuderink, C. Mühl, and M. Poel, "Valence, arousal and dominance in the eeg during game play," *Int. J. Auto. Adapt. Commun. Syst.*, vol. 6, pp. 45–62, Dec. 2013.
[51] D. Sammler, M. Grigutsch, T. Fritz, and S. Koelsch, "Music and emotion: Electrophysiological correlates of the processing of pleasant and unpleasant music," *Psychophysiology*, vol. 44, no. 2, pp. 293–304, 2007.
[52] P. C. Petrantonakis and L. J. Hadjileontiadis, "Adaptive emotional information retrieval from EEG signals in the time-frequency domain," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2604–2616, May 2012.
[53] S. K. Khare and V. Bajaj, "Time–frequency representation and convolutional neural network-based emotion recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 7, pp. 2901–2909, Jul. 2021.
[54] P. Zhong, D. Wang, and C. Miao, "EEG-based emotion recognition using regularized graph neural networks," *IEEE Trans. Affect. Comput.*, early access, May 11, 2020, doi: 10.1109/TAFFC.2020.2994159.
[55] W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, and A. Cichocki, "EmotionMeter: A multimodal framework for recognizing human emotions," *IEEE Trans. Cybern.*, vol. 49, no. 3, pp. 1110–1122, Mar. 2019.

Lakhan Dev Sharma (Member, IEEE) received the M.Tech. degree from Atal Bihari Vajpayee-Indian Institute of Information Technology and Management, Gwalior, India, in 2012, and the Ph.D. degree from Dr. B. R. Ambedkar National Institute of Technology Jalandhar, India, in 2018. He has been serving as a Sr. Assistant Professor with the School of Electronics Engineering, VIT-AP University, Amaravati, India. He has teaching experience at various technical institutes and university levels. His research interests include biomedical signal and image processing, machine learning, and deep learning.

Abhijit Bhattacharyya (Member, IEEE) received the B.E. (Hons.) degree in electronics and communication engineering (ECE) from the University of Burdwan, Burdwan, India, in 2010, the M.Tech. degree in ECE from Dr. B. R. Ambedkar National Institute of Technology (NIT) Jalandhar, India, in 2013, and the Ph.D. degree in electrical engineering from the Indian Institute of Technology Indore, India, in 2018. During his Ph.D. degree, he was awarded the Raman–Charpak Fellowship 2017 by the Indo-French Centre for the Promotion of Advanced Research (IFCPAR). Under this fellowship, he worked at the University of Lorraine–CNRS, Nancy, France. He worked as a Postdoctoral Researcher with the University of Michigan, USA. Presently, he has been serving as an Assistant Professor with the Department of ECE, NIT Hamirpur, India. His research interests include biomedical signal processing, medical image processing, time-frequency analysis, multivariate signal analysis, machine learning, and deep learning.