IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 28, 2021, accepted July 16, 2021, date of publication July 27, 2021, date of current version August 11, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3100638

A Channel Selection Method for Emotion Recognition From EEG Based on Swarm-Intelligence Algorithms

ESEN YILDIRIM[1], YASİN KAYA[2], AND FATIH KILIÇ[2]

[1]Department of Electrical and Electronics Engineering, Adana Alparslan Türkeş Science and Technology University, 01250 Adana, Turkey
[2]Department of Computer Engineering, Adana Alparslan Türkeş Science and Technology University, 01250 Adana, Turkey

Corresponding author: Esen Yildirim (eyildirim@atu.edu.tr)

This work involved human subjects or animals in its research. The authors confirm that all human/animal subject research procedures and protocols are exempt from review board approval.

ABSTRACT Increasing demand for human-computer interaction applications has escalated the need for automatic emotion recognition as emotions are essential for natural communication. There are various information sources that can be used for recognizing emotions, such as speech, facial expressions, body movements, and physiological signals. Among those physiological signals are more reliable for better affective communication with machines since they are almost impossible to control. Therefore, automatic emotion recognition from EEG signals has been a topic intensely investigated. Emotions are experiences that arise various cognitive functions observed in different frequency bands involving multiple brain areas and recognition from EEG with high accuracies is only possible with a large number of features extracted from the whole brain in various bands. Emotion regulation also requires integration of cognitive functions and thus functional connectivity between regions should also be considered. In this paper, we extract 736 features based on spectral power and phase-locking values. We particularly focus on finding salient features for emotion recognition using swarm-intelligence (SI) algorithms. We applied well-known classification algorithms for recognizing positive and negative emotions using the feature sets that are selected by these algorithms. Besides, features that are selected by all of them commonly are used as a new feature set. We report accuracies between 56.27% and 60.29% on the average; noting that by decreasing the feature size by 87.17% (from 736 to 94.40) an average accuracy of 60.01±8.93 was obtained with the random forest classifier. We also highlight the efficient electrode locations for emotion recognition. As a result, we define 11 channels as dominant and promising classification results are obtained.

INDEX TERMS EEG, emotion classification, channel selection, feature selection, swarm-Intelligence algorithms.

I. INTRODUCTION

The increasing role of technology and machines in human life makes it necessary to strengthen the interaction between humans and machines. Since emotions are extremely important in social interactions in daily life, emotion recognition systems are essential for affective human-machine interaction. Therefore, emotion recognition has become immensely popular for various fields such as virtual reality, augmented reality, advanced driver assistant systems, neuromarketing, and specifically for human-computer interaction (HCI).

The associate editor coordinating the review of this manuscript and approving it for publication was Gustavo Olague.

Emotion is mostly defined as an experience that is associated with psychological phenomena. It can be very general such as joy, fear or disgust, or a very specific experience such as love or hate to a specific person or object, or embarrassment in a specific situation. Emotions might last for a long time or might emerge for a very short time period. They are mostly accompanied by physiological reactions and physical responses. There exists several studies focusing on emotion recognition systems based on speech [1]–[4], facial expressions [5]–[7] and physiological signals [8]–[15]. Developing a successful emotion recognition system requires a multi-layered architecture. Feature extraction, feature selection, and classification are the main

stages, and the success of each step affects the performance of the entire system. The main issue is to find emotional salient features from several sources, analyzing feature sets [16], [17] to eliminate the irrelevant/unnecessary features and developing new classification frameworks to improve accuracies of existing classifiers [3], [18]. This study focuses on emotion recognition from EEG using band powers and phase-locking values as features and sophisticated feature selection method based on swarm-intelligence (SI) algorithms and well-known classification algorithms such as k-nearest neighbour (k-NN), random forest, and support vector machines (SVM).

In recent studies, swarm-intelligence algorithms have been widely used for engineering optimization problems [4], [19], global radiation forecasting [20], multiple unmanned aerial vehicles (UAV) control [21], and health monitoring of civil engineering structures [22]. Particle Swarm Optimization (PSO) [23], [24], Cuckoo Search (CS) [25], Grey Wolf Optimizer (GWO) [26], and Dragonfly Algorithm (DA) [27], [28] are the most widely used SI algorithms. PSO has been attracted researchers' attention because of its easy implementation and the small number of parameters to be adjusted. CS, one of the relatively recent SI algorithms, has local and global search mechanisms to converge the global best. It uses Lévy flight to search different solution spaces. In addition to these algorithms, GWO and DA are other recent SI algorithms to solve engineering optimization problems. GWO is inspired by the hunting habits of wolves and DA is inspired dynamic and static swarming behaviours of dragonflies. These algorithms are introduced to solve continuous problems. However, the representation of some problems must be designed to search in the binary solution space because of its nature. Thus, most SI methods have binary versions to solve binary optimization problems such as feature selection. For classification problems in machine learning, some features increase the performance of the learning algorithm, while the others are ineffective or negatively affect the performance of these systems. Feature selection aims to select or score more relevant ones among these features. This process requires extremely high computational time especially when the number of features is high. This problem is considered an NP-hard problem [29]. SI algorithms such as cuckoo search [30], GWO [31], DA [27], genetic programming [32], artificial bee colony [33], ant colony optimization [34] are preferred to overcome this problem.

In this paper, we present emotion recognition (positive vs. negative) from EEG signals. We, particularly focus on finding salient features for emotion recognition using swarm-intelligence algorithms, namely GWO, PSO, CS, and DA. Feature selection is applied on a large set constructed by combining features calculated based on spectral power and phase-locking values, extracted from 1 second-long EEG segments without overlapping using sliding windows. As a contribution of the study, we also evaluate the effectiveness of a new feature set containing only the features that are selected by all SI algorithms we used. Performances of the feature selection algorithms are evaluated by 10 fold cross-validation designed in a similar way to one video out scheme; in particular, 4 out of 40 videos are randomly selected for testing while 36 videos are used for training. Another contribution of the study is investigating the dominant channels for emotion regulation. Dominant channels are marked by analyzing the features selected commonly by all selection methods independent from the subjects.

The organization of this paper as the following: Section II describes materials and methods. Experimental results and discussion is provided in Section III to confirm and verify the performances of GWO, PSO, CS, and DA algorithms. Finally, conclusions are stated in Section IV.

II. METHODS

The main focus of the study is to apply sophisticated feature selection algorithms to determine the best features for a successful emotion recognition system. For this purpose, we used a well-known EEG emotional database, Database for Emotion Analysis using Physiological Signals (DEAP) [35]. After extracting the features using Hilbert-Huang Transform (HHT), we applied various feature selection algorithms and completed the system by applying k-NN, random forest, and SVM for classification. We present our results using 10 fold cross-validation assuring that EEG segments in train and test sets are extracted from separate videos providing a more realistic scenario.

The overview of the proposed emotion recognition model is shown in figure 1. The details are given in the following subsections.

A. EMOTIONAL EEG DATABASES

In this paper, we conducted our classification experiments on a very popular emotional EEG database Database for Emotion Analysis using Physiological Signals (DEAP) [35].

DEAP dataset includes EEG and physiological signals collected from 32 subjects, 15 females and 17 males. Subjects aged between 19 and 37 (27.29 ∓ 4.48) and all are right-handed except one female subject. Emotions were elicited through one–minute long music clips. 40 video clips were used for the study. These clips were selected out of 120 video clips, which were initially chosen among tagged songs in a music enthusiast website, by using the subjective ratings from volunteers through a web-based interface. During the experiments physiological signals were recorded for 67 seconds; a 2-second progress report (a screen to inform the subject about the current video number), a 5-second baseline recording (display of a fixation cross for the subject to relax), and a one-minute long video playing, for each video clip. After each video clip, the subject was asked to fill out a self-assessment form for arousal, valence, liking, and dominance through Self-Assessment Manikins (SAM) [36].

In this study, we classify positive and negative emotions. A valence value that is less than or equal to 5 is evaluated as negative.

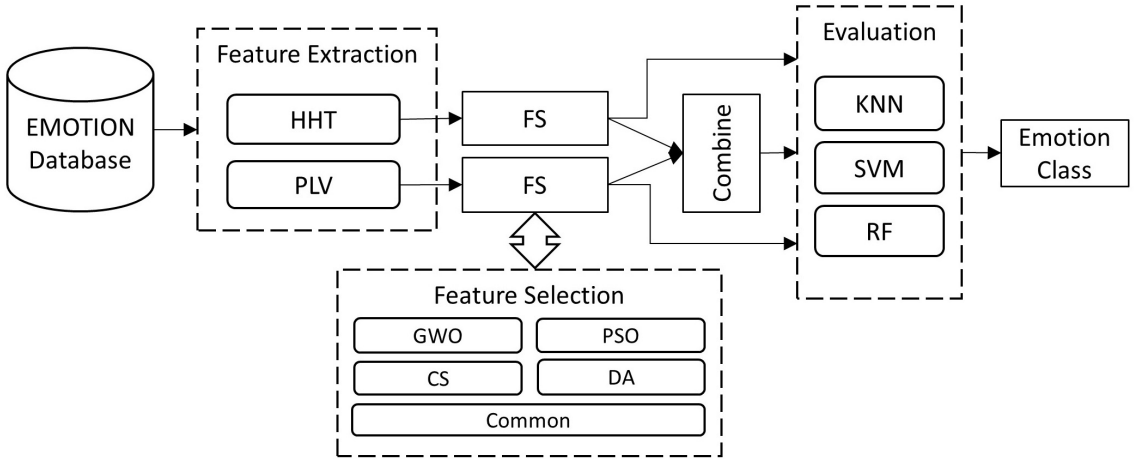

FIGURE 1. Overview of the proposed emotion recognition model.

B. FEATURE EXTRACTION FOR EEG SIGNAL REPRESENTATION

We extract features based on Hilbert Huang Transform (HHT) and Phase Locking Value (PLV) from all EEG channels. HHT is a transform method representing the time-frequency spectrum of EEG and PLV shows the phase synchronization between two narrow-band signals.

1) SPECTRAL FEATURES

HHT is an empirical method to obtain the time-frequency spectrum of the signal using data-driven basis functions. It has two stages; Mode Decomposition and Hilbert spectral analysis (HSA). Empirical Mode Decomposition (EMD) is the first introduced process for mode decomposition.

EMD process is used to extract a complete and nearly orthogonal basis from the data [37]. The basis for the decomposition process is the idea that every signal is composed of different internal oscillation modes. These internal modes, namely Intrinsic Mode Functions (IMFs), are extracted by a recursive process called sifting. IMFs, each of which is a monocomponent signal, i.e. having only one or a very narrow range of frequency components, are extracted through a recursive process called sifting [38]. Each IMF has zero-crossing and extrema numbers that are equal to each other or the difference is at most one. Besides upper and lower envelopes, defined by local maxima and minima respectively, have zero mean values. Decomposing a real-valued signal $x(t)$ into K IMFs and a residue, the signal through the EMD process can be written as:

$$x(t) = \sum_{k=1}^{K} IMF_k(t) + r_N \tag{1}$$

where IMF_k represents k^{th} IMF and r_N represents the residue. IMFs have narrow-band frequency contents and the first IMF has the largest frequency content, whereas the last one has the minimum frequency content. Details of IMF decomposition process can be found in [38].

Improvements on EMD process have been introduced [39]–[43]. Variational Mode Decomposition is introduced in [39] and it is a non-recursive model, in contrast to EMD. VMD uses a concurrent approach to extract IMFs in contrast to recursive EMD. VMD searches for IMFs considering the decomposition process as an optimization problem. Assuming the signal $x(t)$ has K modes, it can be written as:

$$x(t) = \sum_{k=1}^{K} u_k \tag{2}$$

Note that, each mode, u_k, is a narrow-band signal around the center frequency, w_k which is determined as the decomposition process progresses. The optimization problem to be solved is [39]:

$$\min_{u_k, w_k} \left\{ \sum_{k=1}^{K} \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t} \right) * u_k(t) \right] e^{-jw_k t} \right\|_2^2 \right\} \tag{3}$$

The solution to this optimization problem is given in [39] thoroughly.

In order to represent the signal in terms of time and frequency, instantaneous frequencies (IFs) are computed for each time point by differentiating the phase. Having a narrow frequency range, IMFs are convenient for this process. Time-frequency map is constructed by recording the amplitudes of the signal for respective frequency values along the time. Amplitudes and phases, or IFs in this context, are easy to extract for complex signals. Ambiguity for real signals is eliminated by defining an analytical signal by means of Hilbert Transform.

$$\begin{aligned} c(t) &= x(t) + j\tilde{x}(t) = A(t)e^{j\theta(t)} \\ f(t) &= \frac{1}{2\pi} \frac{d\theta(t)}{dt} \end{aligned} \tag{4}$$

where $\tilde{x}(t)$ is Hilbert transform of the original signal $x(t)$ and $c(t)$ is the analytic function to extract the instantaneous frequency values. $A(t)$ is the amplitude value and $f(t)$ is the frequency content at time t respectively. Finally, the signal is expressed as follows:

$$x(t) = Re\left(\sum_{k=1}^{K} A_k(t) e^{j\int f_k(t)dt}\right) \tag{5}$$

for K IMFs, where $A_k(t)$ and $f_k(t)$ are amplitude and frequency of k^{th} IMF, respectively.

In this work, we have used vmd and hht functions in MATLAB R2020a to obtain IMFs and related time-frequency maps. This map is, then used for feature extraction. Log band power values, namely differential entropy [44], for various bands, namely $\theta, \alpha, \beta, \gamma$ and *total* are used as spectral features. HHT based spectral features are extracted as:

$$F_b^{ch} = log\left\{\int_{f_b}\int_t H(\nu,\tau)d\tau d\nu\right\} \tag{6}$$

$$F_{std,b}^{ch} = log\left\{std\left\{\int_{f_b} H(\nu,\tau)d\nu\right\}\right\} \tag{7}$$

$$F_{max,b}^{ch} = log\left\{max\left\{\int_{f_b} H(\nu,\tau)d\nu\right\}\right\} \tag{8}$$

$H(f,t)$ is the Hilbert-Huang spectrum and F_b^{ch}, $F_{std,b}^{ch}$ and $F_{max,b}^{ch}$ are the log band power in band b and the logarithms of standard deviation and the maximum value of the spectrum integrated over the band b for channel ch respectively. f_b represents the frequency values in [4, 8), [8, 13), [13, 30), [30, 64) and [4, 64) Hz for $\theta, \alpha, \beta, \gamma$, and *total*, respectively. 480 (32 channels x 5 bands x 3) spectral features are calculated in total.

2) PHASE LOCKING VALUES

Phase synchronization is a method for understanding the underlying neuronal coordination between brain regions. Phase synchronization focuses on phase locking between the time series or activation maps and does not take the relation between amplitude values into account.

Phase locking value (PLV) was introduced by Lachaux *et al.* [45] and there are many studies that present the effectiveness of PLV in EEG signal analysis, both for normal [10], [46]–[48] and pathological [49]–[51] cases. Computing PLV between two signals, $s_x(t)$ and $s_y(t)$, requires instantaneous phase values of both signals. Assuming that signals are narrow band, and complex-valued, PLV is computed by averaging the phase difference over trials.

$$PLV_t = \frac{1}{N}\left|\sum_{n=1}^{N} exp(j\theta(t,n))\right| \tag{9}$$

where N is the number of trials, and $\theta(t,n) = \theta_x(t,n) - \theta_y(t,n)$ is the phase difference at time t for trial n. In case of a single trial, PLV can be estimated by measuring the phase difference across time steps with a predefined latency [52]:

$$sPLV_t = \frac{1}{\delta}\int_{t-\delta/2}^{t+\delta/2} exp(j\theta(\tau))d\tau \tag{10}$$

In the context of this work, we have calculated PLV values in θ, α, β, and γ bands for each channel-pair (a total of 496 values for each band). In equation 10, $\theta(t) = \theta_i(t) - \theta_j(t)$ is the phase difference between channels i and j, and δ is chosen as 1 second, i.e. one second long part of the 10 seconds segment. Two approaches are followed to decrease the number of possible features (496 values*4 bands); average PLV values for each channel and PLV coefficients. PLV values for each channel pair, i, j where $i \neq j$, is calculated as:

$$sPLV(i,j) = \frac{1}{T}\left|\int_t sPLV_t(i,j)dt\right| \tag{11}$$

where $sPLV(i,j)$ is the average single trial PLV values, $sPLV_t(i,j)$, along the 10 second segment.

Average PLV value for channel ch is computed as the average of $sPLV(ch,j)$ values for all channels, $j \neq ch$:

$$aPLV_b^{ch} = \frac{1}{N}\left|\sum_{j,j\neq ch} sPLV(ch,j)\right| \tag{12}$$

Calculating $aPLV_b^{ch}$ for $ch = 1, 2, .., 32$ in 4 bands, $b \in$ [4, 8), [8, 13), [13, 30), and [30, 64), we acquire $4*32 = 128$ average PLVs for each segment.

PLV calculation gives us the opportunity to observe the brain network created between every possible channel pair. Recently, graph theory has been used to examine the brain networks and we use graph theory utilized clustering coefficients for each channel as features along with average PLVs. The clustering coefficient, namely c-coefficient [53], can be calculated as follows [54]:

$$cCoef_b^i = \frac{\sum_{k\neq i}\sum_{\substack{l\neq i \\ l\neq k}} c_{ik}c_{il}c_{kl}}{\sum_{k\neq i}\sum_{\substack{l\neq i \\ l\neq k}} c_{ik}c_{il}} \tag{13}$$

where $c_{i,j}$ is $sPLV(i,j)$ and $cCoef_b^i$ is the c-coefficient for channel i in band b. 128 more features are obtained by calculating c-coefficients for 32 channels in 4 bands.

C. SWARM-INTELLIGENCE BASED FEATURE SELECTION

Feature selection is an optimization problem considering two major issues. One of them is to get higher classification accuracy and the other is to decrease the number of the selected features. Wrapper feature selection methods require an evaluation function to compare created solutions. In the literature, researchers generally preferred the classification error of a classifier as an evaluation function (fitness function) [30], [55]. Moreover, the number of the selected features was included in the evaluation function to get similar accuracy using a lower number of features. In this

study, the fitness function as shown in equation 14 is performed [23], [56]–[58].

$$Fitness = \alpha * \gamma_R(D) + \beta * \frac{|s|}{|d|} \tag{14}$$

where $\alpha \in [0, 1]$, $\beta = 1 - \alpha$, and $\gamma_R(D)$ represents the classification error of selected features R relative to decision D. $|d|$ shows the total number of features and $|s|$ indicates the number of selected features. In this study, the error rate of k-NN classifier was employed to compute the fitness function. The K parameter of k-NN was used as $K = 5$ and 5-fold cross-validation was employed to deal with overfitting problems [58].

1) PARTICLE SWARM OPTIMIZATION

The particle swarm optimization (PSO) is a well-known nature-inspired algorithm, which inspired by bird flocking, fish schooling, and bee swarming and established by Kennedy and Eberhart [59]. PSO is the population-based algorithm, and each individual in a population is called a particle that represents a possible solution in the search space. The population is called a swarm which consists of particles. Each particle uses the velocity function to moves from its current position to a new position. The velocity function is constantly updated, taking account of its past experience *pBest* and the swarm's experience *gBest*. The first version of PSO is designed to optimize real-valued problems. A binary version of the PSO proposed by Kennedy and Eberhart [24]. It uses the concept of velocity function using a probability concept. A position in the solution takes either 1 or 0. In other words, each particle has a position which is a form of a binary vector X_i. The positions and velocities formulated as equation (15,16);

$$X_i = x_i^1, x_i^2, ., x_i^j, ., x_i^d \tag{15}$$

$$V_i = v_i^1, v_i^2, ., v_i^j, ., v_i^d \tag{16}$$

where d is the number of features, i is the particle's index in a swarm, and j represents the index of the feature. Value of x_i^j is either 0 or 1 that shows whether the feature is unselected or selected.

In BPSO, particles are initialized randomly. In this study, equation 17 is performed to generate the binary position of the particles.

$$x_i^j = \begin{cases} 1 & rand \geq 0.5 \\ 0 & rand < 0.5 \end{cases} \tag{17}$$

The fitness function is used to calculate the fitness value of each particle. The position having the best fitness value in particle's past experience is *pBest*. The position having the best fitness value in swarm's past experience is *gBest*. After particles start from a random position, the velocities and positions of particles are updated according to the particle's best *pBest* and *gBest* at each iteration until the number of maximum iteration. The velocity of each particle is updated using equation (18).

$$v_i^{j,t+1} = w \cdot v_i^{j,t} + c_1 \cdot rand \cdot (pBest_i^{j,t} - x_i^{j,t}) + c_2 \cdot rand \cdot (gBest^{j,t} - x_i^{j,t}) \tag{18}$$

The particles update their binary position using velocity vector and transfer function for feature selection. V-shaped and S-shaped are well-known transfer function [60], [61]. These names are given in accordance with their shapes. In this study, S-shaped transfer function is used for binary conversion since S-shaped function gives better results than V-shaped function in [61]. Equation 20 defines the sigmoid transfer function.

$$Sigmoid_t(v_i^{j,t+1}) = \frac{1}{1 + exp^{-v_i^{j,t+1}}} \tag{19}$$

The particle changes its position using equation 20

$$x_i^{j,t+1} = \begin{cases} 1 & rand < Sigmoid_t(v_i^{j,t+1}) \\ 0 & rand \geq Sigmoid_t(v_i^{j,t+1}) \end{cases} \tag{20}$$

2) CUCKOO SEARCH

Cuckoo search (CS) is a nature-inspired meta-heuristic algorithm proposed by Yang and Deb [25]. The breeding parasitism behavior of the cuckoo birds inspires the algorithm. With this behavior, the female cuckoos lay their eggs in the nests of other host birds. The host birds unconsciously grow their offspring. If the host birds discover in the nest an alien egg, they will throw it out or desert the nest and build another nest elsewhere.

In the CS algorithm, each egg in the nest indicates a solution and each cuckoo egg indicates a new candidate solution. In [62], the authors describe the following three rules of CS:

1) Each cuckoo chooses a random nest and lays one egg (a candidate solution).
2) The next generation contains only high-quality (better solutions) of eggs transferred from the previous generation.
3) The host bird can discover the alien egg using a predefined probability rate and the number of usable host nests is a constant number.

In the CS, Levy flight is used to create new solution x_i^{new} as in equation (21-23) assuming $x_i = (x_{i1}, x_{i2}, \cdots, x_{iD})$ is the position value of i^{th} egg. The calculation of random numbers from a Levy distribution employs Mantegna's method in terms of two normally distributed random numbers (u and v) as in equation(22) [63].

$$x_i^{new} = x_i^{old} + \alpha(x_i - x_g) \oplus \text{Levy}(\beta) = x_i^{old} + \alpha \frac{u}{|v|^{1/\beta}}(x_i - x_g) \tag{21}$$

$$u \sim N(0, \sigma_u^2), v \sim N(0, \sigma_v^2) \tag{22}$$

$$\sigma_u = \left[\frac{\sin(\pi\beta/2) \cdot \Gamma(1+\beta)}{2^{(\beta-1)/2}\beta \cdot \Gamma(\frac{1+\beta}{2})} \right]^{1/\beta}, \quad \sigma_v = 1 \tag{23}$$

where $\oplus$ represents entry-wise multiplications, $\alpha > 0$ means cuckoo's step size scaling parameter relating to the scales

of the problem domain, β is Levy flight exponent. u and v are random numbers and calculated using equation (22-23). x_i represents position of i^{th} egg (solution) and x_g represents the best position in the current population. $\Gamma(\cdot)$ indicates Gamma function.

In the CS, a discovery operator, that changes the discovered nests with new nests using pre-defined *pa* probability, is also employed.

$$x_{ij}^{new} = \begin{cases} x_{ij}^{old} + rand1(x_{r1,j}(k) - x_{r2,j}(k)) & if\ rand2 > pa \\ x_{ij}^{old}(k) & else \end{cases} \tag{24}$$

where x_{ij}^{new} represents the j^{th} element of the i^{th} solution, $rand1$ and $rand2$ are real random numbers in $[1, 0]$, pa represents discovery probability of CS, $x_{r1,j}$ and $x_{r2,j}$ represent j^{th} element of solutions x_{r1} and x_{r2}, where $r1$ and $r2$ are two different numbers between 0 and N, N is the size of the population.

The CS is used for continuous and binary problems. To solve binary problems, a binary version of the CS can be used. A Binary Cuckoo Search Algorithm (BCSA) for feature selection was introduced in [64]. In this algorithm, continuous values transferred to binary values using a binarisation operator. Various binarisation operators have been used in the literature. But, the most popular one is the sigmoid function formulated in equation (26).

$$x_{ij}(t+1) = \begin{cases} 1 & if\ Sigmoid(x_{ij}(t)) > rand3 \\ 0 & otherwise \end{cases} \tag{25}$$

$$Sigmoid(a) = \frac{1}{1 + e^{-a}} \tag{26}$$

in which $rand3 \sim U(0, 1)$ is a uniform distributed random number, $x_{ij}(t)$ indicates the new egg's value at time step t.

3) GREY WOLF OPTIMIZER

The grey wolf optimizer, is another young bio-inspired heuristic search algorithm, employed to solve non-linear, computationally difficult, constrained optimization problems. The hierarchical and hunting behavior of grey wolves that perform their activities in a flock inspired the GWO algorithm [26]. The first version of GWO, just as PSO, is designed to optimize continuous valued problems. A binary version of the GWO was proposed by Emary *et al.* [57] for feature selection.

$$\overrightarrow{X}(t+1) = \overrightarrow{X}_p(t) + \overrightarrow{A} \cdot \overrightarrow{D} \tag{27}$$

where $\overrightarrow{D}$ is given in equation (28), t represents the iteration number, $\overrightarrow{X}_p$ is the prey position, $\overrightarrow{A}$ and $\overrightarrow{C}$ are defined as coefficient vectors, and $\overrightarrow{X}$ indicates the position of the grey wolf.

$$\overrightarrow{D} = |\overrightarrow{C} \cdot \overrightarrow{X}_p(t) - \overrightarrow{X}(t)| \tag{28}$$

Eqs. (29), (30) calculate $\overrightarrow{A}$ and $\overrightarrow{C}$ vectors, respectively.

$$\overrightarrow{A} = 2a \cdot \overrightarrow{r_1} - a \tag{29}$$

$$\overrightarrow{C} = 2\overrightarrow{r_2} \tag{30}$$

where a decreases linearly from 2 to 0 for each iteration, $\overrightarrow{r_1}$ and $\overrightarrow{r_2}$ are uniformly distributed random vectors in [0, 1]. The hunting habit is generally led by the alpha wolf. The beta and delta wolves might attend hunting activities occasionally. It is assumed that grey wolves, the alpha (best candidate solution), beta (the second-best candidate solution), and delta (the third-best candidate solution) have better knowledge about the potential location of prey to mimic the hunting behavior of wolves. All wolves, containing the first three best candidate solutions achieved so far and other search individuals, must adjust their positions according to the position of the best search agents using equation (31).

$$\overrightarrow{X}(t+1) = \frac{\overrightarrow{X_1} + \overrightarrow{X_2} + \overrightarrow{X_3}}{3} \tag{31}$$

where $\overrightarrow{X_1}$, $\overrightarrow{X_2}$, $\overrightarrow{X_2}$ are calculated using Eqs. (32), (33), and (34), respectively.

$$\overrightarrow{X_1} = |\overrightarrow{X_\alpha} - \overrightarrow{A_1} \cdot \overrightarrow{D_\alpha}| \tag{32}$$

$$\overrightarrow{X_2} = |\overrightarrow{X_\beta} - \overrightarrow{A_2} \cdot \overrightarrow{D_\beta}| \tag{33}$$

$$\overrightarrow{X_3} = |\overrightarrow{X_\delta} - \overrightarrow{A_3} \cdot \overrightarrow{D_\delta}| \tag{34}$$

The best three solutions are defined according to their fitness values. $\overrightarrow{X_\alpha}$, $\overrightarrow{X_\beta}$, and $\overrightarrow{X_\delta}$ are the three best solutions at t^{th} iteration in Eqs. (32), (33), and (34). $\overrightarrow{A_1}$, $\overrightarrow{A_2}$, and $\overrightarrow{A_3}$, are given as in equation (29), and are calculated using Eqs. (35), (36), and (37), respectively.

$$\overrightarrow{D_\alpha} = |\overrightarrow{C_1} \cdot \overrightarrow{X_\alpha} - \overrightarrow{X}| \tag{35}$$

$$\overrightarrow{D_\beta} = |\overrightarrow{C_2} \cdot \overrightarrow{X_\beta} - \overrightarrow{X}| \tag{36}$$

$$\overrightarrow{D_\delta} = |\overrightarrow{C_3} \cdot \overrightarrow{X_\delta} - \overrightarrow{X}| \tag{37}$$

where $\overrightarrow{C_1}$, $\overrightarrow{C_2}$, and $\overrightarrow{C_2}$ are given as in equation (30)

In this paper, we implemented binarization mechanism of [57].

$$X_i^{t+1} = Crossover(x_1, x_2, x_3) \tag{38}$$

where $Crossover(x_1, x_2, x_3)$ is a simple stochastic crossover process to crossover a, b, and c solutions given in equation (39). x_1, x_2, x_3 represent binary vectors which are updated considering the movement of wolves toward the α, β, δ grey wolves, respectively. x_1, x_2, and x_3 are calculated using equation (40).

$$x_d = \begin{cases} a_d & \text{if } rand < \frac{1}{3} \\ b_d & \frac{1}{3} \le rand < \frac{2}{3} \\ c_d & \text{otherwise} \end{cases} \tag{39}$$

Equation (40) is performed to calculate the position of The leading grey wolves. In equation (40), *wolfIndex* is 1, 2 and 3 representing α, β, δ grey wolves, respectively.

$$x_{wolfIndex}^d = \begin{cases} 1 & \text{if } (x_{wolf}^d + bstep_{wolf}^d) \ge 1 \\ 0 & \text{otherwise} \end{cases} \tag{40}$$

where *wolf* indicates one of α, β, δ, d is dimension in *wolf*, and $bstep^d_{wolf}$ indicates binary step at dimension d and calculated using equation (41).

$$bstep^d_{wolf} = \begin{cases} 1 & \text{if } cstep^d_{wolf} \geq rnd1 \\ 0 & \text{otherwise} \end{cases} \tag{41}$$

where $rnd1$ is a uniform distributed random number in [0, 1], $cstep^d_{wolf}$ is a sigmoid function.

$$cstep^d_{wolf} = \frac{1}{1 + e^{-10(A^d_1 D^d_{wolf} - 0.5)}} \tag{42}$$

where A^d_1 and D^d_{wolf} are calculated Eqs. (29) and (35) in the dimension d.

4) DRAGONFLY ALGORITHM

Another SI method, inspired by behaviors of dragonflies in nature and introduced by Mirjalili [27], is the dragonfly algorithm (DA). In the algorithm, there are two main processes namely exploration and exploitation that modeled statically or dynamically food searching or avoiding the enemy of dragonflies.

Dragonflies' swarming behavior and its mathematical expression of the algorithm implementation are given in the following steps;

(1) In order to avoid collision with other neighbor individuals, the algorithm uses a separation mechanism mathematically modeled as in equation (43).

$$S_i = -\sum_{j=1}^{N} X - X_j \tag{43}$$

where X is the position of the current individual, N represents the size of the neighborhood, and X_j is the j^{th} position of X.

(2) The second operator, named Alignment, represents the individuals' velocity matching taking into account other neighborhood individuals. This operator is given as in equation (44)

$$A_i = \frac{\sum_{j=1}^{N} V_j}{N} \tag{44}$$

where V_j is the velocity of j^{th} neighborhood individual.

(3) Another operator is Cohesion, which indicates the individuals' tendency toward the neighborhood's center of mass. This operator is given as in equation (45) mathematically.

$$C_i = \frac{\sum_{j=1}^{N} x_j}{N} - X \tag{45}$$

(4) In order to survive, each individual has two key behaviors, which are moving towards a food source and avoiding the enemy. Eqs. (46) and (47) represent the mathematical model of the attraction towards food.

$$F_i = X^+ - X \tag{46}$$

where X^+ is the position of the food source.

$$E_i = X^- + X \tag{47}$$

where X^- is the position of the enemy.

DA employs two vectors named step (ΔX) and position (X) vectors to adjust the position of dragonflies and mimic their movements. The step vector is calculated as in equation (48):

$$\Delta X_{t+1} = (sS_i + aA_i + cC_i + fF_i + eE_i) + w\Delta X_t \tag{48}$$

where t is the iteration number, w represents the inertia weight, s demonstrates the separation weight, S_i represents the separation of the i^{th} individual. a, c, f, and e are the alignment weight, the cohesion weight, the food factor, and the enemy factor, respectively. A_i represents the alignment of i^{th} individual, C_i shows the cohesion of the i^{th} individual, F_i represents the food source of the i^{th} individual, E_i indicates the position of enemy of the i^{th} individual.

Equation (49) calculates the position vector:

$$x_{t+1} = \begin{cases} \neg X_t, & x < T(\Delta x_{t+1}) \\ X_t, & x \geq T(\Delta x_{t+1}) \end{cases} \tag{49}$$

where $T(\Delta x_{t+1})$ is given as in equation (50).

$$T(\Delta x) = \left| \frac{\Delta x}{\sqrt{\Delta x^2 + 1}} \right| \tag{50}$$

D. CLASSIFICATION

To evaluate the effectiveness of the proposed features in terms of emotion classification, we performed several classification experiments using k nearest neighbours (k-NN) [65], random forest [66] classifiers and support vector machines (SVM) [67]. For k-NN classifier, k is selected as 1, and Euclidean distance is selected as the distance metric. All classification experiments were carried out in Matlab software with version R2020a.

III. RESULTS AND DISCUSSION

EEG recordings from 32 subjects are split into two categories: positive and negative emotions. Categories are defined using the individual ratings of each subject; a valence value less than or equal to 5 is considered as a negative emotion while the rest (a value between 5 and 9) is considered as positive. 60-second-long segments with no overlap led to 60 samples for each video clip resulted in 2400 samples in total for each subject. We evaluated 3 feature sets; HHT based spectral features (480 features), PLV based features (256 features), and a combination of these two sets (736 features).

A. FEATURE SELECTION

Four feature selection algorithms, namely PSO, CS, GWO, and DA are used for selecting the most salient features. The intersection of these four selected feature sets is also used. We employed 5-fold cross-validation to evaluate the performance of the proposed EEG feature selection approaches [56], [58]. α and β parameters are set to 0.99 and 0.01 respectively in equation (14) [57]. Table 1 represents the global

TABLE 1. Parameter settings for the experiments.

Parameter	Value(s)
Number of search individual	50
K for cross validation	5
k for k-NN	5
k-NN distance metric	Euclidean
Number of iterations T	100
Search dimension d	The feature sizes of datasets
Fitness	
α parameter of the fitness	0.99
β parameter of the fitness	0.01
PSO	
c_1	2
c_2	2
$Vmax$	6
$Wmax$	0.9
$Wmin$	0.4
CS	
pa	0.4
α	1

TABLE 2. Average Number of Selected Features for HHT Feature Set.

Sub.	GWO	PSO	CS	DA	Common
1	372.10	236.40	306.90	236.20	62.20
2	380.50	233.20	305.90	237.20	64.60
3	374.20	234.50	307.20	238.10	64.90
4	375.10	239.40	297.70	236.20	60.60
5	371.70	235.80	304.10	223.10	60.60
6	367.70	233.60	294.70	233.10	57.70
7	374.00	236.20	305.00	236.40	64.30
8	375.70	241.90	308.00	231.60	64.60
9	370.40	230.70	309.60	239.60	62.30
10	377.80	240.60	302.40	235.50	60.60
11	366.80	228.20	307.30	228.50	53.90
12	369.60	227.90	309.80	232.40	58.80
13	363.80	239.30	306.60	230.00	59.80
14	373.80	228.80	303.70	226.80	60.50
15	373.30	231.30	305.40	228.50	61.70
16	375.50	229.00	304.30	237.30	61.10
17	374.70	230.40	291.30	236.30	57.10
18	381.30	240.80	293.40	223.00	60.20
19	377.30	239.10	307.90	238.70	65.70
20	390.30	238.00	303.90	233.60	65.50
21	370.90	239.20	310.10	228.40	60.70
22	372.10	240.10	308.90	235.90	64.80
23	376.50	243.40	310.80	236.30	70.00
24	370.60	233.70	306.10	225.40	60.20
25	369.20	238.70	304.40	237.40	61.80
26	377.80	235.00	304.60	225.40	62.50
27	370.00	235.00	300.50	226.00	56.10
28	365.40	235.10	306.80	235.10	63.90
29	372.00	238.20	301.50	221.10	61.20
30	372.70	236.00	306.20	231.40	64.00
31	371.70	235.60	295.40	231.40	60.90
32	387.50	235.20	306.30	234.20	65.30
AVG	373.81	235.63	304.27	232.19	61.82

values and algorithm specific parameter values. Note that the GWO, and DA have no specific parameters. Figure 2 and figure 3 show the average convergence rates of 10 trials on HHT and PLV feature sets, respectively. It can be seen from the figures that the GWO and DA achieve similar global solutions. However, GWO converges fast in early iterations, while DA achieves convergence rates similar to GWO in late

TABLE 3. Average Number of Selected Features for PLV Feature Set.

Sub.	GWO	PSO	CS	DA	Common
1	198.20	122.70	163.80	122.40	30.60
2	196.80	123.50	159.60	122.10	33.20
3	196.90	127.40	162.60	121.70	36.70
4	193.00	120.60	164.30	123.60	33.90
5	197.70	121.00	162.40	116.40	32.30
6	189.20	124.10	160.20	117.00	29.40
7	199.60	122.50	164.60	127.20	35.90
8	193.80	124.00	162.40	126.80	35.50
9	198.50	125.50	162.40	127.40	35.10
10	190.10	119.00	158.50	117.20	29.90
11	194.30	122.80	160.00	116.30	30.20
12	192.80	127.50	160.90	120.90	32.60
13	187.50	126.40	157.60	116.60	30.40
14	199.50	123.40	164.70	121.70	34.70
15	198.90	122.80	163.80	120.30	34.10
16	196.80	126.90	161.10	127.90	33.30
17	199.00	125.50	163.40	118.90	29.80
18	194.00	123.40	160.20	120.10	32.80
19	188.30	121.70	159.20	118.60	29.20
20	198.00	121.80	163.20	122.60	32.80
21	196.80	120.40	160.30	122.40	31.80
22	190.10	120.20	159.60	123.30	29.60
23	193.60	127.30	161.30	124.50	33.00
24	196.50	120.60	158.20	118.60	29.50
25	195.10	121.10	162.60	125.30	35.00
26	194.70	125.00	164.80	122.50	33.80
27	194.10	125.40	165.60	122.70	34.10
28	190.40	122.00	162.80	116.20	32.50
29	197.70	118.50	163.00	120.20	29.50
30	195.60	126.10	156.70	125.70	35.10
31	198.70	122.20	161.80	120.20	31.70
32	199.20	121.70	161.10	124.80	34.70
AVG	195.17	123.22	161.65	121.63	32.58

iterations for most of the subjects. These figures show that newly introduced algorithms, DA and GWO, tend to utilize global solutions.

The average number of features selected from HHT and PLV feature sets are shown in tables 2 and 3 respectively. As can be seen from the tables, although GWO converges very fast, the average number of features is decreased by only about 22%. PSO and DA result in similar sizes, decreasing the numbers by 51.24% and 51.93%, respectively. We also applied the classification on the feature set which is an intersection of the features selected by all methods, named as Common in the tables. For HHT+PLV feature set, selected features for HHT and PLV feature sets are combined arising feature numbers 568.98, 358.85, 465.92, 353.82, and 94.40 for GWO, PSO, CS, DA, and common on the average, respectively. Note that, the number of features is decreased from 736 to 94.40, corresponding to an 87.17% decrease, on average if common features are selected.

B. CLASSIFICATION

k-NN, random forest, and SVM are used for the classification of the categories. Subject dependent cross-validation, where each subject is trained and tested independently, is employed for the evaluation of the performances of feature selection and classification methods. The use of cross-validation methods

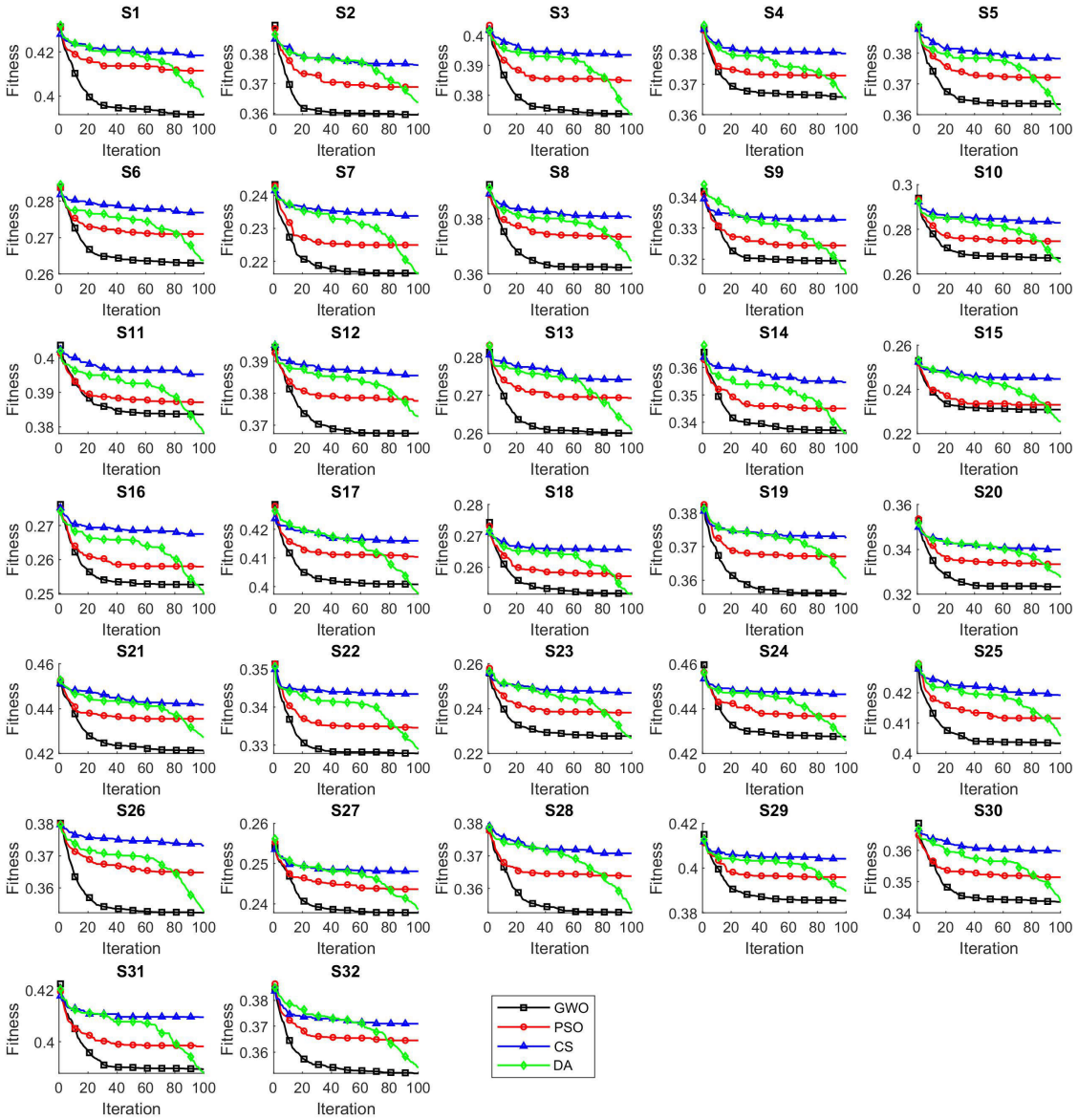

FIGURE 2. Convergence Curves for HHT feature selection process.

TABLE 4. Classification Results.

Classifier	Feat. Set	GWO	PSO	CS	DA	Common	ALL
RF	HHT	59.47±8.23	59.36±8.42	59.67±8.34	59.49±8.30	59.17±8.11	59.53±8.37
	PLV	60.10±8.85	59.91±8.90	59.91±8.67	60.10±8.58	59.44±9.03	60.02±8.82
	HHT+PLV	60.14±9.03	60.18±9.07	60.20±9.00	60.29±8.89	60.01±8.93	60.19±8.94
k-NN	HHT	56.88±6.51	56.85±6.63	57.04±6.59	57.05±6.68	56.97±6.78	56.92±6.48
	PLV	57.78±7.04	57.48±7.24	57.53±7.06	57.60±6.90	57.55±7.35	57.73±6.89
	HHT+PLV	57.71±7.24	57.58±7.16	57.54±7.26	57.57±7.33	57.73±7.55	57.71±7.16
SVM	HHT	56.60±6.16	56.65±5.93	56.84±6.08	56.68±6.25	56.27±5.82	56.75±6.02
	PLV	57.47±6.31	57.08±6.36	57.33±6.25	57.32±6.25	57.05±6.77	57.85±6.33
	HHT+PLV	57.70±6.68	57.68±6.61	57.63±6.67	57.74±6.53	57.09±6.64	57.81±6.50

on features collected using the standard sliding window technique is very common in EEG studies. In the cross-validation scheme, the feature matrix is divided into folds randomly and then each fold is tested after the system is trained on the rest

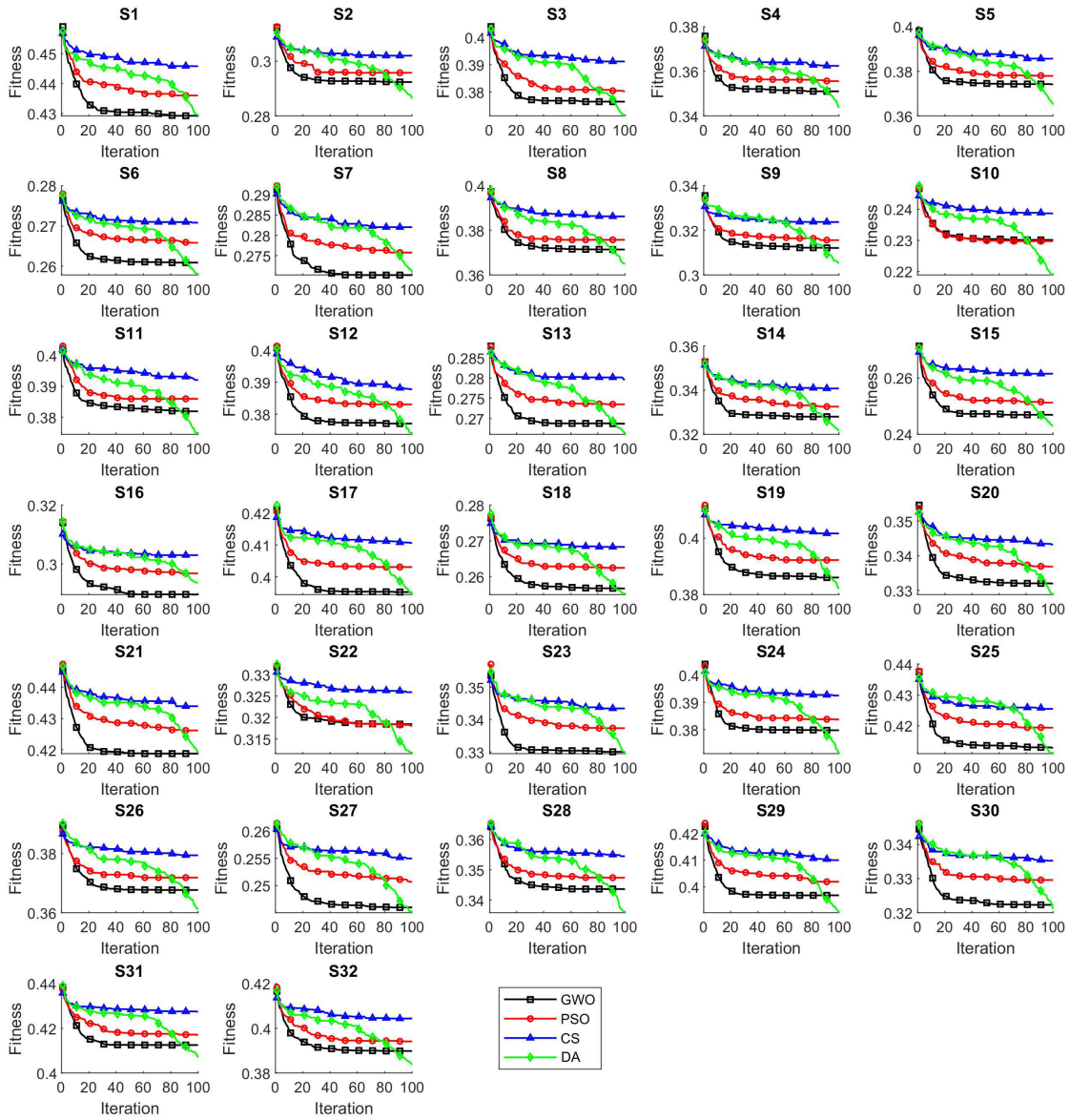

FIGURE 3. Convergence Curves for PLV feature selection process.

of the dataset. However, due to the sliding window analysis, there will be samples extracted from segments belonging to EEG signals collected during a particular video stimulus in both train and test sets with a very high probability. Therefore, the accuracy obtained might have high values due to the proximity between the segments belonging to the same stimuli. In a realistic scenario or a real-time experiment; in the training set, there will be no segments belonging to the stimuli that will be used for testing. Therefore, the performances reported using this scheme should be read carefully.

The ultimate goal of the emotion recognition system presented here is to see how well the system performs for detecting emotions for a subject watching a specific video, while the system is trained with EEG recordings collected from this subject as they were watching other videos. For this purpose, EEG signals recorded during a video that is used for testing are left completely out of the training dataset. More precisely, 10-fold cross-validation is applied as follows:

- Randomly sort 40 videos, divide them into 10 groups,
- Hold the EEG recordings during videos in i^{th} group, $i = 1, 2, \ldots, 10$, for test set and train the data by using the rest of the feature set,
- Repeat the process for all subsets,
- Calculate the average accuracy

A summary of classification results for each feature set and feature selection method is given in table 4. Average results for each subject using HHT+PLV feature set and random forest classifier for all feature selection methods are given

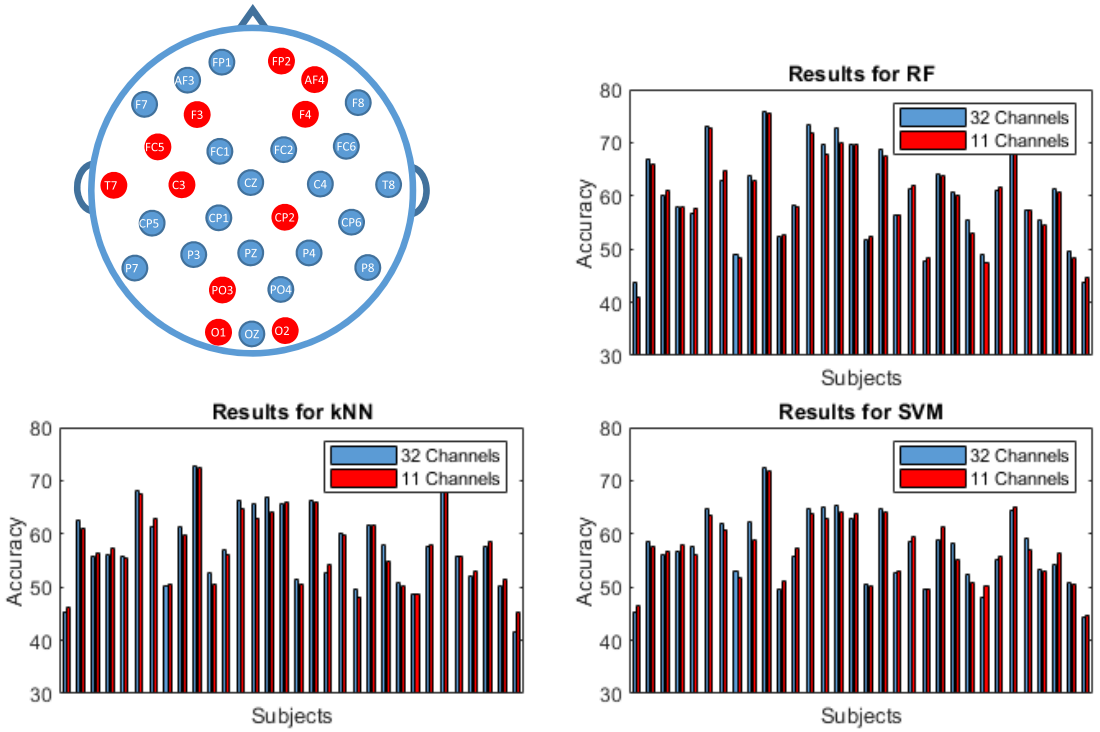

FIGURE 4. Comparison of accuracies for all and dominant channels.

in table 5 as an example. Experiment results show that the random forest classifier achieves the best performance for our problem. Before feature selection, classification accuracies are 59.53%, 60.02%, and 60.19% for HHT, PLV, and all feature sets, respectively. Applying selection algorithms, accuracies between 59.17% and 60.29% are attained. Features selected from PLV and HHT+PLV sets by DA algorithm exceeded the accuracy obtained by the whole feature set after reducing the number of features by 51.93%. Performance of commonly selected feature set is also remarkable as the number of features is deduced by more than 87%. Results show that the performances of the feature selection methods are very close to the accuracy rates obtained using the whole feature set. These results show that swarm-intelligence algorithms achieve comparable accuracy rates after reducing the feature set by almost 88%. Comparing the type of the feature sets, we see that better accuracies are attained by combining the HHT and PLV feature sets for all cases than using them solely. It is worth to mention that, higher accuracies are attained using PLV based features than that using spectral band power-based features. This result shows the existence of functional brain networks through phase-locking during emotion regulation.

C. ANALYSIS FOR DOMINANT CHANNELS

Analyzing the selected features from all feature selection methods we used, we also extracted the most dominant channels for emotion recognition. For that purpose, we selected the channels for which at least one spectral or PLV based

TABLE 5. Classification Results For HHT+PLV Feature Set and Classifier RF.

Sub	GWO	PSO	CS	DA	Common	ALL
1	41.54	43.08	42.33	43.96	43.75	42.00
2	67.58	67.04	67.75	67.29	67.00	66.92
3	59.46	60.62	60.17	60.83	60.08	60.00
4	59.21	59.08	59.50	59.17	57.79	59.71
5	57.46	56.46	56.29	57.00	56.58	56.00
6	72.58	73.21	72.92	72.58	73.00	72.79
7	62.75	64.00	63.75	63.21	62.92	63.92
8	49.63	50.75	49.29	49.33	48.96	50.88
9	63.79	64.50	63.17	63.58	63.67	63.71
10	75.83	75.67	75.42	75.92	75.96	75.42
11	52.92	53.08	52.67	52.75	52.42	53.17
12	57.00	59.00	59.00	59.42	58.21	58.54
13	73.38	73.83	74.29	73.63	73.42	73.63
14	69.54	69.54	69.33	70.17	69.63	69.75
15	71.79	72.75	71.79	72.13	72.83	71.79
16	70.58	70.46	70.50	70.13	69.58	70.25
17	52.21	52.75	53.33	53.13	51.88	52.63
18	69.42	68.25	69.38	68.96	68.75	69.29
19	55.50	55.71	56.33	55.63	56.25	56.25
20	63.00	62.25	62.17	62.17	61.29	62.25
21	49.63	49.17	48.42	49.50	47.71	48.46
22	65.04	64.83	64.63	64.67	64.00	64.96
23	60.50	61.04	60.67	60.58	60.62	61.33
24	55.50	54.21	55.08	55.29	55.58	54.42
25	48.25	47.21	49.67	49.13	49.08	49.92
26	61.46	61.17	60.92	61.58	61.04	60.54
27	71.29	71.42	70.92	71.17	70.92	71.38
28	56.50	56.54	57.71	57.50	57.38	57.54
29	56.38	55.58	55.29	55.92	55.42	55.63
30	61.50	60.88	61.71	61.38	61.38	60.92
31	49.88	48.58	48.54	48.71	49.67	48.63
32	43.50	43.25	43.54	43.00	43.71	43.63
AVG	60.14	60.18	60.20	60.29	60.01	60.19

TABLE 6. Classification Results for Dominant Channels.

Classifier	GWO	PSO	CS	DA	Common	ALL
RF	60.05±9.05	60.15±8.94	60.07±9.15	60.05±8.91	59.59±8.85	60.20±9.13
k-NN	57.38±7.24	57.50±7.14	57.48±7.00	57.57±7.02	57.48±6.99	57.45±7.13
SVM	57.26±6.69	57.21±6.57	57.18±6.59	57.19±6.58	56.93±6.17	57.39±6.61

feature is selected in the common feature set for at least 90% of the training experiments (9 out of 10) for at least 31 subjects. The analysis resulted in 11 dominant channels; F3, F4, FC5, AF4 in the frontal, T7 in the temporal, C3 and CP2 in the central, PO3 in the parietal, and O1 and O2 in occipital regions. This result is expected as emotion recognition is known to involve multiple brain areas [68]. Previous studies have shown the association of emotion with the regions our results reveal. The frontal lobe is reported as the main cortex area that is related to emotion in the brain [69]–[72]. The presence of pre-frontal, parietal, and temporal asymmetries in different bands for valence recognition is shown in [73]. Interestingly, the somatosensory cortex, which is included in the post-central gyrus, is also shown to have a significant impact on emotional processing [74]–[77]. An fMRI study involving images and videos for emotion elicitation shows the visual cortex activation for emotion encoding [78]–[80].

We repeated the classification experiments involving only dominant channels for validation. In these runs, we only used the selected features extracted from dominant channels, for each feature selection method. The average number of selected features (HHT+PLV) are 128.91+67.05, 81.29+42.39, 105.64+55.32, 80.41+42.25, and 21.93+11.32 for GWO, PSO, CS, DA, and common respectively. Note that, this corresponds to about a 65% decrease in the number of features used for classification with respect to the numbers selected for all channels. In summary, feature numbers are reduced by 73.38%, 83.20%, 78.13%, 83.33%, and 95.48% on average for GWO, PSO, CS, DA, and common respectively when the feature set before selection is considered.

Average classification results for dominant channels are shown in table 6. Figure 4 shows the location of the dominant channels and the comparison of classification results for dominant and all channels using the common feature set for all subjects. Table and figure show the accuracies for features selected from HHT+PLV sets for 11 channels. These results validate the effectiveness of the dominant channels, determined by the swarm intelligence based feature selection algorithms, as the classification results are comparable to the ones that were obtained by using all channels.

IV. CONCLUSION

Emotions are experiences that involve many functions such as cognition, memory, and motor functions and possibly integration of them. Considering the brain being a control mechanism for all actions we take, the use of electrophysiological signals for emotion recognition has become an emerging research area. This paper has focused on finding the salient features using swarm-intelligence algorithms for emotion recognition from EEG signals. We applied 4 algorithms, namely PSO, CS, GWO, and DA, and decreased the number of features by 22.69%, 51.24%, 36.70%, and 51.93%, respectively. Using the features that are commonly selected by all methods feature size is decreased by 87.17%. Classification results show that all of the selected feature sets, including the common set, achieve comparable accuracies with the whole set which contains 736 features, approving the eligibility of the SI algorithms for feature selection.

One of the biggest challenges in EEG signal processing studies is the inter- and intra-subject variabilities [81]. Intra-subject variability corresponds to the differences in the extracted features from the same subject in separate sessions, whereas the inter-subject variability points to the differences between subjects. Therefore, it is arduous to find EEG predictors, even for a specific task. In this study, we also undertook channel analysis for specifying the electrode locations that are commonly selected by all subjects for all sessions. Our analysis resulted in 11 channels distributed over all brain regions. As emotion is known to be a very complex process for our brain, that was an expected result supported by a number of previous studies. We obtained promising results as the classification accuracies for feature sets selected by SI algorithms, sets that are obtained by the intersection of them and also the sets obtained only considering the dominant channels, decreasing the number of features drastically, are comparable to the other experiments with all features. In future work, the proposed methods will be tested with different EEG feature sets and classifiers to achieve higher accuracies.

REFERENCES

[1] V. Petrushin, "Emotion in speech: Recognition and application to callcenters," in *Proc. Artif. Neural Netw. Eng.*, vol. 710. St. Louis, MO, USA: Citeseer, 1999, p. 22.
[2] F. Yu, E. Chang, Y.-Q. Xu, and H.-Y. Shum, "Emotion detection from-speech to enrich multimedia content," in *Advances in Multimedia Information Processing—PCM 2001*, H.-Y. Shum, M. Liao, and S.-F. Chang, Eds. Berlin, Germany: Springer, 2001, pp. 550–557.
[3] Q. Mao, M. Dong, Z. Huang, and Y. Zhan, "Learning salient features for speech emotion recognition using convolutional neural networks," *IEEE Trans. Multimedia*, vol. 16, no. 8, pp. 2203–2213, Dec. 2014.
[4] S. Yildirim, Y. Kaya, and F. Kılıç, "A modified feature selection method based on Metaheuristic algorithms for speech emotion recognition," *Appl. Acoust.*, vol. 173, Feb. 2021, Art. no. 107721.
[5] P. Ekman, "Facial expression and emotion," *Amer. Psychol.*, vol. 48, no. 4, p. 384, 1993.
[6] S. V. Ioannou, A. T. Raouzaiou, V. A. Tzouvaras, T. P. Mailis, K. C. Karpouzis, and S. D. Kollias, "Emotion recognition through facial expression analysis based on a neurofuzzy network," *Neural Netw.*, vol. 18, no. 4, pp. 423–435, May 2005.

[7] M. G. Calvo and L. Nummenmaa, "Detection of emotional faces: Salient physical features guide effective visual search," *J. Exp. Psychol., Gen.*, vol. 137, no. 3, p. 471, 2008.

[8] G. Valenza, L. Citi, A. Lanatá, E. P. Scilingo, and R. Barbieri, "Revealing real-time emotional responses: A personalized assessment based on heartbeat dynamics," *Sci. Rep.*, vol. 4, no. 1, pp. 1–13, May 2015.

[9] S. S. Uzun, S. Yildirim, and E. Yildirim, "Emotion primitives estimation from EEG signals using Hilbert Huang transform," in *Proc. IEEE-EMBS Int. Conf. Biomed. Health Informat.*, Jan. 2012, pp. 224–227.

[10] Y. Dasdemir, E. Yildirim, and S. Yildirim, "Analysis of functional brain connections for positive–negative emotions using phase locking value," *Cognit. Neurodyn.*, vol. 11, no. 6, pp. 487–500, Dec. 2017.

[11] M. S. Özerdem and H. Polat, "Emotion recognition based on EEG features in movie clips with channel selection," *Brain Inf.*, vol. 4, no. 4, pp. 241–252, 2017.

[12] P. Bruno, V. Melnyk, and F. Völckner, "Temperature and emotions: Effects of physical temperature on responses to emotional advertising," *Int. J. Res. Marketing*, vol. 34, no. 1, pp. 302–320, Mar. 2017.

[13] 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.

[14] H. Chao, L. Dong, Y. Liu, and B. Lu, "Emotion recognition from multiband EEG signals using CapsNet," *Sensors*, vol. 19, no. 9, p. 2212, May 2019.

[15] C. Xiefeng, Y. Wang, S. Dai, P. Zhao, and Q. Liu, "Heart sound signals can be used for emotion recognition," *Sci. Rep.*, vol. 9, no. 1, pp. 1–11, Dec. 2019.

[16] S. R. Bandela and T. K. Kumar, "Speech emotion recognition using semi-NMF feature optimization," *Turkish J. Electr. Eng. Comput. Sci.*, vol. 27, no. 5, pp. 3741–3757, Sep. 2019.

[17] S. B. Alex, L. Mary, and B. P. Babu, "Attention and feature selection for automatic speech emotion recognition using utterance and syllable-level prosodic features," *Circuits, Syst., Signal Process.*, vol. 39, no. 11, pp. 5681–5709, Nov. 2020.

[18] H. Chao, L. Dong, Y. Liu, and B. Lu, "Improved deep feature learning by synchronization measurements for multi-channel EEG emotion recognition," *Complexity*, vol. 2020, pp. 1–15, Mar. 2020.

[19] Y. Kaya, "A novel method for optic disc detection in retinal images using the cuckoo search algorithm and structural similarity index," *Multimedia Tools Appl.*, vol. 79, nos. 31–32, pp. 23387–23400, Aug. 2020.

[20] F. Kılıç, İ. H. Yılmaz, and Ö. Kaya, "Adaptive co-optimization of artificial neural networks using evolutionary algorithm for global radiation forecasting," *Renew. Energy*, vol. 171, pp. 176–190, Jun. 2021.

[21] F. Dai, M. Chen, X. Wei, and H. Wang, "Swarm intelligence-inspired autonomous flocking control in UAV networks," *IEEE Access*, vol. 7, pp. 61786–61796, 2019.

[22] S. Das and P. Saha, "Performance of swarm intelligence based chaotic meta-heuristic algorithms in civil structural health monitoring," *Measurement*, vol. 169, Feb. 2021, Art. no. 108533.

[23] F. Kılıç, Y. Kaya, and S. Yildirim, "A novel multi population based particle swarm optimization for feature selection," *Knowl.-Based Syst.*, vol. 219, May 2021, Art. no. 106894.

[24] J. Kennedy and R. C. Eberhart, "A discrete binary version of the particle swarm algorithm," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. Comput. Simul.*, vol. 5, Oct. 1997, pp. 4104–4108.

[25] X.-S. Yang and S. Deb, "Cuckoo search via Lévy flights," in *Proc. World Congr. Nature Biologically Inspired Comput. (NaBIC)*, Dec. 2009, pp. 210–214.

[26] S. Mirjalili, S. M. Mirjalili, and A. Lewis, "Grey wolf optimizer," *Adv. Eng. Softw.*, vol. 69, pp. 46–61, Mar. 2014.

[27] S. Mirjalili, "Dragonfly algorithm: A new meta-heuristic optimization technique for solving single-objective, discrete, and multi-objective problems," *Neural Comput. Appl.*, vol. 27, no. 4, pp. 1053–1073, 2016.

[28] M. M. Mafarja, D. Eleyan, I. Jaber, A. Hammouri, and S. Mirjalili, "Binary dragonfly algorithm for feature selection," in *Proc. Int. Conf. New Trends Comput. Sci. (ICTCS)*, Oct. 2017, pp. 12–17.

[29] G. Chandrashekar and F. Sahin, "A survey on feature selection methods," *Comput. Elect. Eng.*, vol. 40, no. 1, pp. 16–28, Jan. 2014.

[30] L. A. M. Pereira, D. Rodrigues, T. N. S. Almeida, C. C. O. Ramos, A. N. Souza, X.-S. Yang, and J. P. Papa, "A binary cuckoo search and its application for feature selection," in *Cuckoo Search and Firefly Algorithm* (Studies in Computational Intelligence). Cham, Switzerland: Springer, 2014, pp. 141–154.

[31] N. P. N. Sreedharan, B. Ganesan, R. Raveendran, P. Sarala, B. Dennis, and R. R. Boothalingam, "Grey wolf optimisation-based feature selection and classification for facial emotion recognition," *IET Biometrics*, vol. 7, no. 5, pp. 490–499, Aug. 2018.

[32] B. Tran, B. Xue, and M. Zhang, "A new representation in PSO for discretization-based feature selection," *IEEE Trans. Cybern.*, vol. 48, no. 6, pp. 1733–1746, Jun. 2018.

[33] H. Rao, X. Shi, A. K. Rodrigue, J. Feng, Y. Xia, M. Elhoseny, X. Yuan, and L. Gu, "Feature selection based on artificial bee colony and gradient boosting decision tree," *Appl. Soft Comput.*, vol. 74, pp. 634–642, Jan. 2019.

[34] M. Paniri, M. B. Dowlatshahi, and H. Nezamabadi-Pour, "MLACO: A multi-label feature selection algorithm based on ant colony optimization," *Knowl.-Based Syst.*, vol. 192, Mar. 2020, Art. no. 105285.

[35] S. Koelstra, C. Muhl, M. Soleymani, J. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, and I. Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE Trans. Affect. Comput.*, vol. 3, no. 1, pp. 18–31, Jan. 2012.

[36] J. D. Morris, "Observations: SAM: The self-assessment manikin; an efficient cross-cultural measurement of emotional response," *J. Advertising Res.*, vol. 35, no. 8, pp. 63–68, 1995.

[37] N. E. Huang and S. S. P. Shen, *Hilbert-Huang Transform and Its Applications*. Singapore: World Scientific, 2005.

[38] N. E. Huang, Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis," *Proc. Roy. Soc. London A, Math., Phys. Eng. Sci.*, vol. 454, no. 1971, pp. 903–995, Mar. 1998.

[39] K. Dragomiretskiy and D. Zosso, "Variational mode decomposition," *IEEE Trans. Signal Process.*, vol. 62, no. 3, pp. 531–544, Feb. 2014.

[40] S. Meignen and V. Perrier, "A new formulation for empirical mode decomposition based on constrained optimization," *IEEE Signal Process. Lett.*, vol. 14, no. 12, pp. 932–935, Dec. 2007.

[41] N. Pustelnik, P. Borgnat, and P. Flandrin, "A multicomponent proximal algorithm for empirical mode decomposition," in *Proc. 20th Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2012, pp. 1880–1884.

[42] T. Y. Hou and Z. Shi, "Adaptive data analysis via sparse time-frequency representation," *Adv. Adapt. Data Anal.*, vol. 3, nos. 1–2, pp. 1–28, Apr. 2011.

[43] M. Feldman, "Time-varying vibration decomposition and analysis based on the Hilbert transform," *J. Sound Vib.*, vol. 295, nos. 3–5, pp. 518–530, Aug. 2006.

[44] L.-C. Shi, Y.-Y. Jiao, and B.-L. Lu, "Differential entropy feature for EEG-based vigilance estimation," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 6627–6630.

[45] J.-P. Lachaux, E. Rodriguez, J. Martinerie, and F. J. Varela, "Measuring phase synchrony in brain signals," *Hum. Brain Mapping*, vol. 8, no. 4, pp. 194–208, 1999.

[46] W. Jian, M. Chen, and D. J. McFarland, "EEG based zero-phase phase-locking value (PLV) and effects of spatial filtering during actual movement," *Brain Res. Bull.*, vol. 130, pp. 156–164, Apr. 2017.

[47] Z. Yu, T. Ma, N. Fang, H. Wang, Z. Li, and H. Fan, "Local temporal common spatial patterns modulated with phase locking value," *Biomed. Signal Process. Control*, vol. 59, May 2020, Art. no. 101882.

[48] Z. Wang, Y. Tong, and X. Heng, "Phase-locking value based graph convolutional neural networks for emotion recognition," *IEEE Access*, vol. 7, pp. 93711–93722, 2019.

[49] N. E. Cámpora, C. J. Mininni, S. Kochen, and S. E. Lew, "Seizure localization using pre ictal phase-amplitude coupling in intracranial electroencephalography," *Sci. Rep.*, vol. 9, no. 1, p. 803, Dec. 2019.

[50] M. C. Gómez, H. Keijzer, J. Hofmeijer, B. J. Ruijter, R. B. Fernández, M. C. Tjepkema-Cloostermans, and M. J. A. M. van Putten, "O-46 EEG connectivity measures in prognostication of postanoxic coma patients," *Clin. Neurophysiol.*, vol. 130, no. 7, p. e36, Jul. 2019.

[51] D. Cho, B. Min, J. Kim, and B. Lee, "EEG-based prediction of epileptic seizures using phase synchronization elicited from noise-assisted multivariate empirical mode decomposition," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 8, pp. 1309–1318, Aug. 2017.

[52] J.-P. Lachaux, E. Rodriguez, M. L. Van Quyen, A. Lutz, J. Martinerie, and F. J. Varela, "Studying single-trials of phase synchronous activity in the brain," *Int. J. Bifurcation Chaos*, vol. 10, no. 10, pp. 2429–2439, Oct. 2000.

[53] P. Li, H. Liu, Y. Si, C. Li, F. Li, X. Zhu, X. Huang, Y. Zeng, D. Yao, Y. Zhang, and P. Xu, "EEG based emotion recognition by combining functional connectivity network and local activations," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 10, pp. 2869–2881, Oct. 2019.

[54] C. J. Stam, W. de Haan, A. Daffertshofer, B. F. Jones, I. Manshanden, A. M. van Cappellen van Walsum, T. Montez, J. P. A. Verbunt, J. C. de Munck, B. W. van Dijk, H. W. Berendse, and P. Scheltens, "Graph theoretical analysis of magnetoencephalographic functional connectivity in Alzheimer's disease," *Brain*, vol. 132, no. 1, pp. 213–224, Jan. 2008.
[55] Z. Zhu, Y.-S. Ong, and M. Dash, "Wrapper–filter feature selection algorithm using a memetic framework," *IEEE Trans. Syst., Man, B, Cybern.*, vol. 37, no. 1, pp. 70–76, Mar. 2007.
[56] M. M. Mafarja and S. Mirjalili, "Hybrid whale optimization algorithm with simulated annealing for feature selection," *Neuro Comput.*, vol. 260, pp. 302–312, Oct. 2017.
[57] E. Emary, H. M. Zawbaa, and A. E. Hassanien, "Binary grey wolf optimization approaches for feature selection," *Neurocomputing*, vol. 172, pp. 371–381, Jan. 2016.
[58] M. Abdel-Basset, D. El-Shahat, I. El-henawy, V. H. C. de Albuquerque, and S. Mirjalili, "A new fusion of grey wolf optimizer algorithm with a two-phase mutation for feature selection," *Expert Syst. Appl.*, vol. 139, Jan. 2020, Art. no. 112824.
[59] J. Kennedy and R. Eberhart, "Particle swarm optimization," in *Proc. Int. Conf. Neural Netw. (ICNN)*, vol. 4, 1995, pp. 1942–1948.
[60] S. Mirjalili and A. Lewis, "S-shaped versus V-shaped transfer functions for binary particle swarm optimization," *Swarm Evol. Comput.*, vol. 9, pp. 1–14, Apr. 2013.
[61] M. Mafarja, I. Aljarah, H. Faris, A. I. Hammouri, A. M. Al-Zoubi, and S. Mirjalili, "Binary grasshopper optimisation algorithm approaches for feature selection problems," *Expert Syst. Appl.*, vol. 117, pp. 267–286, Mar. 2019.
[62] X.-S. Yang and S. Deb, "Engineering optimisation by cuckoo search," *Int. J. Math. Model. Numer. Optim.*, vol. 1, no. 4, pp. 330–343, 2010.
[63] X.-S. He, N. Li, and X.-S. Yang, "Non-dominated sorting cuckoo search for multiobjective optimization," in *Proc. IEEE Symp. Swarm Intell.*, Dec. 2014, pp. 1–7.
[64] D. Rodrigues, L. A. Pereira, T. Almeida, J. P. Papa, A. Souza, C. C. Ramos, and X.-S. Yang, "BCS: A binary cuckoo search algorithm for feature selection," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, May 2013, pp. 465–468.
[65] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Trans. Inf. Theory*, vol. 13, no. 1, pp. 21–27, Sep. 2006.
[66] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, Oct. 2001.
[67] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[68] J. Pan, L. Zhan, C. Hu, J. Yang, C. Wang, L. Gu, S. Zhong, Y. Huang, Q. Wu, X. Xie, Q. Chen, H. Zhou, M. Huang, and X. Wu, "Emotion regulation and complex brain networks: Association between expressive suppression and efficiency in the fronto-parietal network and default-mode network," *Frontiers Hum. Neurosci.*, vol. 12, p. 70, Mar. 2018.
[69] R. J. Davidson, "What does the prefrontal cortex 'do' in affect: Perspectives on frontal EEG asymmetry research," *Biol. Psychol.*, vol. 67, nos. 1–2, pp. 219–234, Oct. 2004.
[70] P. R. Goldin, K. McRae, W. Ramel, and J. J. Gross, "The neural bases of emotion regulation: Reappraisal and suppression of negative emotion," *Biol. Psychiatry*, vol. 63, no. 6, pp. 577–586, Mar. 2008.
[71] T. Iwaki and M. Noshiro, "EEG activity over frontal regions during positive and negative emotional experience," in *Proc. ICME Int. Conf. Complex Med. Eng. (CME)*, Jul. 2012, pp. 418–422.
[72] S. M. Alarcão and M. J. Fonseca, "Emotions recognition using EEG signals: A survey," *IEEE Trans. Affective Comput.*, vol. 10, no. 3, pp. 374–393, Jul./Sep. 2019.
[73] D. Huang, C. Guan, K. K. Ang, H. Zhang, and Y. Pan, "Asymmetric spatial pattern for EEG-based emotion detection," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jun. 2012, pp. 1–7.
[74] I. Bufalari, T. Aprile, A. Avenanti, F. Di Russo, and S. M. Aglioti, "Empathy for pain and touch in the human somatosensory cortex," *Cerebral Cortex*, vol. 17, no. 11, pp. 2553–2561, Nov. 2007.
[75] D. Purves, R. Cabeza, S. A. Huettel, K. S. LaBar, M. L. Platt, M. G. Woldorff, and E. M. Brannon, *Cognitive Neuroscience*. Sunderland, MA, USA: Sinauer Associates, 2008.
[76] A. R. Damasio, T. J. Grabowski, A. Bechara, H. Damasio, L. L. B. Ponto, J. Parvizi, and R. D. Hichwa, "Subcortical and cortical brain activity during the feeling of self-generated emotions," *Nature Neurosci.*, vol. 3, no. 10, pp. 1049–1056, Oct. 2000.
[77] E. Kropf, S. K. Syan, L. Minuzzi, and B. N. Frey, "From anatomy to function: The role of the somatosensory cortex in emotional regulation," *Brazilian J. Psychiatry*, vol. 41, no. 3, pp. 261–269, May 2019.
[78] P. A. Kragel, M. C. Reddan, K. S. LaBar, and T. D. Wager, "Emotion schemas are embedded in the human visual system," *Sci. Adv.*, vol. 5, no. 7, Jul. 2019, Art. no. eaaw4358.
[79] K. A. Lindquist, A. B. Satpute, T. D. Wager, J. Weber, and L. F. Barrett, "The brain basis of positive and negative affect: Evidence from a meta-analysis of the human neuroimaging literature," *Cerebral Cortex*, vol. 26, no. 5, pp. 1910–1922, May 2016.
[80] M. Kuniecki, K. Wołoszyn, A. Domagalik, and J. Pilarczyk, "Disentangling brain activity related to the processing of emotional visual information and emotional arousal," *Brain Struct. Function*, vol. 223, no. 4, pp. 1589–1597, Nov. 2017.
[81] S. Saha and M. Baumert, "Intra- and inter-subject variability in EEG-based sensorimotor brain computer interface: A review," *Frontiers Comput. Neurosci.*, vol. 13, p. 87, Jan. 2020.

ESEN YILDIRIM received the B.S. degree in electrical and electronics engineering from Çukurova University, Adana, Turkey, in 1997, and the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California (USC), Los Angeles, in 2000 and 2006, respectively. She is currently an Associate Professor of electrical and electronics engineering with Adana Alparslan Türkeş Science and Technology University, Adana. Her general research interests include biomedical signal processing, epileptic seizure detection, functional connectivity, learning methods, and emotion recognition from physiological signals.

YASİN KAYA received the B.S. degree in statistics and computer science and the M.S. and Ph.D. degrees in computer engineering from Karadeniz Technical University, Trabzon, Turkey, in 1999, 2006, and 2017, respectively. From 1999 to 2018, he was a Lecturer with the Department of Informatics, Karadeniz Technical University. He has been an Assistant Professor with the Department of Computer Engineering, Adana Alparslan Türkeş Science and Technology University, Turkey, since 2018. His research interests include biomedical signal processing, arrhythmia detection and classification, biomedical image processing, pattern recognition, and nature inspired algorithms.

FATIH KILIÇ received the B.S. degree in computer engineering from Mersin University, Turkey, in 2002, and the M.S. and Ph.D. degrees in electrical and electronics engineering from Çukurova University, Adana, Turkey, in 2010 and 2015, respectively. He is currently an Assistant Professor with the Department of Computer Engineering, Adana Alparslan Türkeş Science and Technology University, Adana. His research interests include transit network design problem, optimization algorithms, and web application security.