Pattern Recognition 105 (2020) 107393

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Person identification using EEG channel selection with hybrid flower pollination algorithm

Zaid Abdi Alkareem Alyasseri[a,b,c,*], Ahamad Tajudin Khader[a], Mohammed Azmi Al-Betar[d], Osama Ahmad Alomari[e]

[a] School of Computer Sciences, Universiti Sains Malaysia, Pulau Pinang, Malaysia
[b] ECE Department-Faculty of Engineering, University of Kufa, Najaf P.O. Box 21, Iraq
[c] Braintech Sdn Bhd, Pulau Pinang, Malaysia
[d] Department of Information Technology, Al-Huson University College, Al-Balqa Applied University, Al-Huson, Irbid P.O. Box 50, Jordan
[e] Department of Computer Engineering, Faculty of Engineering and Architecture, Istanbul Gelisim University, Istanbul, Turkey

A R T I C L E   I N F O

Article history:
Received 6 July 2019
Revised 7 March 2020
Accepted 21 April 2020
Available online 26 April 2020

Keywords:
EEG
Biometric
Channel selection
Flower pollination algorithm
β-hill climbing

A B S T R A C T

Recently, electroencephalogram (EEG) signal presents a great potential for a new biometric system to deal with a cognitive task. Several studies defined the EEG with uniqueness features, universality, and natural robustness that can be used as a new track to prevent spoofing attacks. The EEG signals are the graphical recording of the brain electrical activities which can be measured by placing electrodes (channels) in various positions of the scalp. With a large number of channels, some channels have very important information for biometric system while others not. The channel selection problem has been recently formulated as an optimisation problem and solved by optimisation techniques. This paper proposes hybrid optimisation techniques based on binary flower pollination algorithm (FPA) and β-hill climbing (called FPAβ-hc) for selecting the most relative EEG channels (i.e., features) that come up with efficient accuracy rate of personal identification. Each EEG signals with three different groups of EEG channels have been utilized (i.e., time domain, frequency domain, and time-frequency domain). The FPAβ-hc is measured using a standard EEG signal dataset, namely, EEG motor movement/imagery dataset with a real world data taken from 109 persons each with 14 different cognitive tasks using 64 channels. To evaluate the performance of the FPAβ-hc, five measurement criteria are considered:accuracy (Acc), (ii) sensitivity (Sen), (iii) F-score (F_s), (v) specificity (Spe), and (iv) number of channels selected (No. Ch). The proposed method is able to identify the personals with high Acc, Sen., F_s, Spe, and less number of channels selected. Interestingly, the experimental results suggest that FPAβ-hc is able to reduce the number of channels with accuracy rate up to 96% using time-frequency domain features. For comparative evaluation, the proposed method is able to achieve results better than those produced by binary-FPA-OPF method using the same EEG motor movement/imagery datasets. In a nutshell, the proposed method can be very beneficial for effective use of EEG signals in biometric applications.

1. Introduction

Several decades ago, the world was transformed into a digital society where every individual has a unique digital identifier. Digital identifiers can be categorised into traditional identifiers, such as using passwords and ID cards. However, this kind of identifier can be easily circumvented [1]. Therefore, another type of identifier that is based on a person's behaviour or personal characteristics, which are called biometrics, was established; it includes face, voice, fingerprint and iris recognition [2]. Personal identification via biometric systems has recently attracted the attention of security research communities. Security systems are considered one of the most important challenges that any society seeks to resolve continually. One of the main tools for security systems is the use of personal identification systems. However, the widespread use of personal identification for biometric systems has resulted in a new challenge called spoofing [1,3,4]. The spoofing personal identification dilemma is the most dangerous challenge facing any security systems. In principle, spoofing methods are used to attack the security of biometric systems and allow unauthorised persons to en-

* Corresponding author at: ECE Department-Faculty of Engineering, University of Kufa, Najaf P.O. Box 21, Iraq.
E-mail addresses: zaid.alyasseri@uokufa.edu.iq (Z.A.A. Alyasseri), tajudin@usm.my (A.T. Khader), mohbetar@bau.edu.jo (M.A. Al-Betar), oalomari@gelisim.edu.tr (O.A. Alomari).

https://doi.org/10.1016/j.patcog.2020.107393
0031-3203/

ter the system [2]. Several spoofing attacks on biometric systems have already occurred [5]. Examples include the following: Face recognition systems have been spoofed using several attacks, such as *"printed photo to spoof face recognition systems on three laptops"* 2D face spoofing, and 3D mask attacks [6]. Fingerprint scanning has been attacked using *"gummy fingers"* [7]. A Finger-vein commercial system has been spoofed by using a piece of paper [8]. An iris recognition system has been spoofed by using an eyeball in front of an iris scanner [9]. Voice recognition has been spoofed by replaying a voice recording in front of a speaker recognition system [5]. Given these attacks, new biometric identification systems are required to identify persons based on *invisible characteristics* and thus eliminate *external threats*. These new biometric identification systems can be developed using an authentication method based on brain signal electroencephalogram (EEG) [10].

EEG signals have been recently captured and recorded accurately; they can be plugged to new biometric systems to enhance their defense strategies. Several studies has shown that EEG presents unique features [11], universality [4] and natural robustness to spoof attacks [1,12]. EEG signals represent a graphical recording of the brain's electrical activity, which can be measured by placing electrodes (channels) in various positions on the scalp [1,13].

Marcel and Jose in [14] proofed that the brain-wave has a pattern of every individual is unique which can be used as a new biometric for person identification. Also, they expected the EEG-based person identification technique will be an interesting area for new research directions and applications in the future. Also, Palaniappan and Mandic [12] proposed a method for person identification using Visual Evoked Potential (VEP) with energy features of the gamma band as a feature extracted for the EEG signal. The proposed method was tested on a large group of subjects and it achieved a high accuracy rate. The results showed that the analysis and simulations have clearly indicated the significant potential of brain electrical activity as biometrics. Rodrigues et al. [1] used the Binary Flower Pollination Algorithm (BFPA) [15] to obtain the best channels concerning EEG signals for person verification purposes. The authors used a standard EEG dataset focused on motor and movement and imagination [16] using autoregressive models with different orders for feature extraction. They authors were able to obtain recognition rates of around 86% using the Optimum-Path Forest (OPF) classifier with a reduction in the number of EEG channels to half. Alyasseri et al. [4] proposed a novel approach for user identification based on the EEG signals. The method used a multi-objective Flower Pollination Algorithm and the Wavelet Transform (MOFPA-WT) to extract EEG features, in which several variations of EEG energy information from the EEG sub-bands have been extracted. The MOFPA-WT method extracts several time-domain features. The performance results were evaluated using accuracy, sensitivity, specificity, false acceptance rate, and F-score. The MOFPA-WT method was compared with some state-of-the-art techniques using different criteria with promising results.

One of the main challenges in the EEG-based user identification technique is signal acquisition. The acquisition process is implemented by placing a number of electrodes (channels) on top of a person's head, as shown in Fig. 1. This process might be slightly uncomfortable. High proficiency is required to hang electrodes in their correct positions. Several problems should be carefully addressed in this case. For example, unnecessary electrodes hung on the top of a persons' head must be removed. Thus, only the most relevant EEG channels must be selected for user identification. The selection of EEG channels has been recently modeled as an optimisation problem and addressed by using several optimisation methods.

Recently, several researchers have utilised different methods to select EEG channels [1,17–19]. Rodrigues et al. [1] used the binary

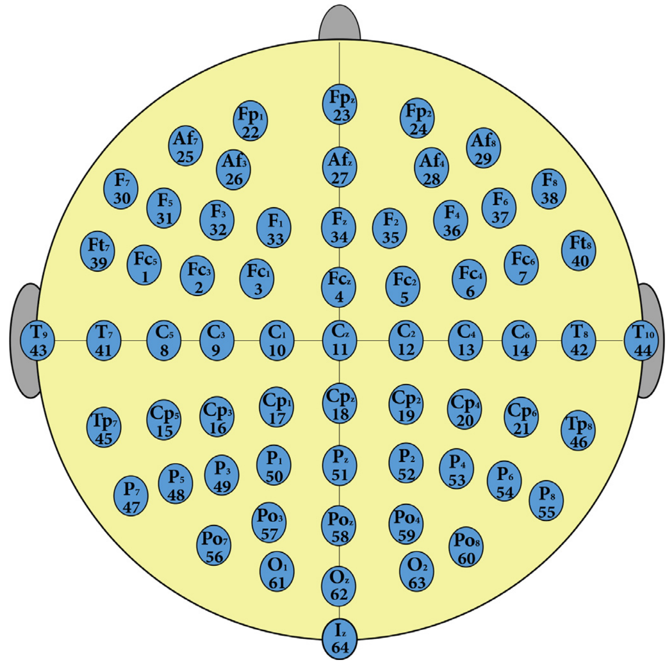


Fig. 1. Distribution of electrodes (channels) to 64 different positions.

flower pollination algorithm (BFPA) [15] to obtain the best channel for the EEG signal with the highest recognition rate for person identification. The authors tested the approach by using a standard EEG dataset that focused on motor and movement and imagination [16]. The BFPA method extracted the autoregressive feature in 5, 10 and 20 different orders. The authors obtained the highest recognition rate of 86% by using the optimum-path forest (OPF) classifier, and the number of EEG channels was reduced to half. The authors comparatively evaluated the method against five optimisation methods (binary genetic algorithm (BGA), binary particle swarm optimisation (BPSO), binary firefly algorithm (BFFA), binary harmony search (BHS) and binary charged system search (BCSS)), and the proposed method ranked first.

FPA is a recent optimisation swarm intelligence method proposed by Yang [15] and inspired by the mating process of flowering plants. It has several advantages over other optimisation methods. It does not require intensive configurations in the initial run nor required derivative data to begin. It has several positive features, such as simplicity, ease of use, extendability, adaptability, flexibility, soundness and completeness. Given its impressive features, it has been successfully utilised for several optimisation problems, such as identification systems [3,4].

Although FPA has been intensively mastered for simple optimisation problems, it exhibits challenges in dealing with nonlinear, non-convex optimisation problems with combinatorial features in nature, such as the EEG channel selection problem. Therefore, the theories of FPA have been improved either by hybridizing the method with other optimisation techniques or tweaking its current operators for the approach to become relevant in addressing the complexity of the optimisation problem on hand.

Given that EEG channel selection can be considered a complex optimisation problem [1], this study proposes an optimum EEG channel selection method by means of a binary constrained version of hybridizing FPA with β-hill climbing. The proposed approach is called FPAβ-hc, and it can determine the optimal subset of channels. The radial basis function-kernel support vector machine (RBF-SVM) classifier for personal identification is used to measure the accuracy of the channels selected. The proposed method (FPA-βhc) selects EEG channels from three different groups, namely, time domain, frequency domain and time-frequency domain features. FPAβ-hc is tested using a standard EEG signal dataset, namely, EEG motor movement/imagery dataset[1], with real-world data obtained from 109 persons each with 14 different cognitive tasks using 64 channels. The following five measures were used to evaluate the performance of FPAβ-hc:(i) accuracy (Acc), (ii) sensitivity (Sen), (iii) F-score (F_s), (v) specificity (Spe), and (iv) number of channels selected (No. Ch). For performance evaluation, the results of the proposed method are compared with those obtained in [1] by using the same EEG motor movement/imagery dataset. FPAβ-hc can reduce the number of channels and achieves an accuracy rate of up to 96% by using time-frequency domain channels.

The rest of this paper is organised as follows: Section 2 explains the EEG channels selection problem. Section 3 provides the background of FPA and β-hill climbing algorithm. The selection schemes are presented in Section 4. An analysis of the results obtained by the proposed method is provided in Section 5. Section 6 presents the conclusions and future work directions.

2. EEG Channel selection problem

EEG channel selection is formulated as an optimisation problem. Therefore, two main optimisation concepts, namely, solution formulation and objective function, are required to utilise any optimisation algorithm for EEG channel selection. This section provides information on the EEG channels selection problem and how this problem is modelled in terms of optimisation context.

2.1. Features for EEG channel selection

Extracting an effective feature (or EEG channel) is crucial in any authentication system [20,21]. The main purpose of the extracted feature is to find unique patterns from input EEG signals that allow for the achievement of a high classification rate. Feature extraction generally involves converting a raw EEG signal into a relevant data structure called a feature vector $\boldsymbol{x} = (x_1, x_2, \dots, x_N)$ by deleting noise and highlighting important data. It could also include "dimensionality reduction," which eliminates redundant and noisy features (repeated data) from the feature vector, to facilitate the classification process [22]. According to Phinyomark [23], Ang et al. [24], the features that can be extracted from any bio-signals, such as EEG, ECG and EMG, can be categorised into three types: time domain features (TDF), frequency domain features (FDF), and time-frequency domain features (T-FDF). These features are explained and formulated as follows.

- **TDF**: This type of feature is commonly used with bio-signals because of its easy and quick extraction from the original signals, given that it does not require a transformation. TDFs are extracted using the signal amplitude, and the resultant values provide a measure of frequency, waveform amplitude and duration within several limited parameters [22].
  The TDF type can be formulated as follows:
  1. Mean (EEG_{Mean})
  $$EEG_{Mean} = \frac{1}{N} * \sum_{j=1}^{N} D_{ij}, \quad i = 1, 2, 3, \dots, L, \tag{1}$$
  where D_{ij} is a time series and N is the number of EEG data points.
  2. Standard deviation (EEG_{Std})
  $$EEG_{Std} = \sqrt{\frac{1}{N} \sum_{j=i}^{N} (x_i - \bar{x})^2}, \quad i = 1, 2, 3, \dots, L, \tag{2}$$
  where $\bar{x}$ is the mean value.
  3. Entropy ($EEG_{Entropy}$)
  $$EEG_{Entropy} = -\sum p(x) \log p(x) \tag{3}$$
  4. Energy (EEG_{Energy})
  $$EEG_{Energy} = \sum_{j=1}^{N} | D_{ij} |^2, \quad i = 1, 2, 3, \dots, L \tag{4}$$
  5. Root mean square (EEG_{RMS})
  $$EEG_{RMS} = \sqrt{\frac{1}{N} * \sum_{j=1}^{N} x_i^2} \tag{5}$$
  6. Variance (EEG_{VAR})
  $$EEG_{VAR} = \frac{1}{N} * \sum_{j=1}^{N} (x_i - \bar{x})^2, \tag{6}$$
  where $\bar{x}$ is the mean value of the EEG signal.
  7. Maximum peak value (EEG_{MPV})
  $$EEG_{MPV} = max \mid x_i \mid \tag{7}$$
  8. Skewness ($EEG_{Skewness}$)
  $$EEG_{Skewness} = \frac{1}{N} * \sum_{j=1}^{N} D_{ij}, \quad i = 1, 2, 3, \dots, L, \tag{8}$$
  where *Skewness* is the moment coefficient of skewness.
  9. Kurtosis ($EEG_{Kurtosis}$)
  $$EEG_{Kurtosis} = \frac{1}{N} * \sum_{j=1}^{N} D_{ij}, \quad i = 1, 2, 3, \dots, L \tag{9}$$
  10. Cross correlation (EEG_{CCR})
  $$EEG_{CCR} = \frac{1}{N} * \sum_{j=1}^{N} D_{ij}, \quad i = 1, 2, 3, \dots, L \tag{10}$$
- **FDF**: This type of EEG feature requires more computational time than TDF. Usually, FDF is measured using the EEG estimated power spectrum density (PSD) or autoregressive coefficient features [22].
  The FDF type is formulated as follows:
  1. Autoregressive coefficients (AR)
  $$EEGAR_{seg} = -\sum_{i=1}^{N} a_i * x_{seg-i} + e * seg, \tag{11}$$
  where a_i is the AR coefficients for feature *i*, *e* is white noise or the error sequence and *N* is the order of the AR model.
  2. Power spectrum density (EEG_{PSD})
  $$EEG_{PSD} = \left| \sum_{i=0}^{N-1} x_i * e^{\frac{-j*2*\pi*seg_i}{N}} \right|^2, \tag{12}$$
  where $seg = 0; 1; 2,\dots N$ is the length of the EEG data.
- **T-FDF**: T-FDF can be represented by localizing the signal energy in terms of time and frequency, and it can provide an accurate description of the physical phenomenon. However, these features generally require a shift that may be computationally heavy. The most commonly used in T-FDF is the short time Fourier transform (STFT) feature [22]. The Fourier transform technique divides the input signal into segments; then, the signal in each window can be assumed to be stationary. The STFT can be formulated as follows:
  $$EEGSTFT_{x(t,w)} = \int W * (\tau - t) x(\tau) e^{jw\tau} d\tau, \tag{13}$$
  where $W(t)$ is the window function, τ represents time and w stands for frequency.

[1] https://www.physionet.org/physiobank/database/eegmmidb/.

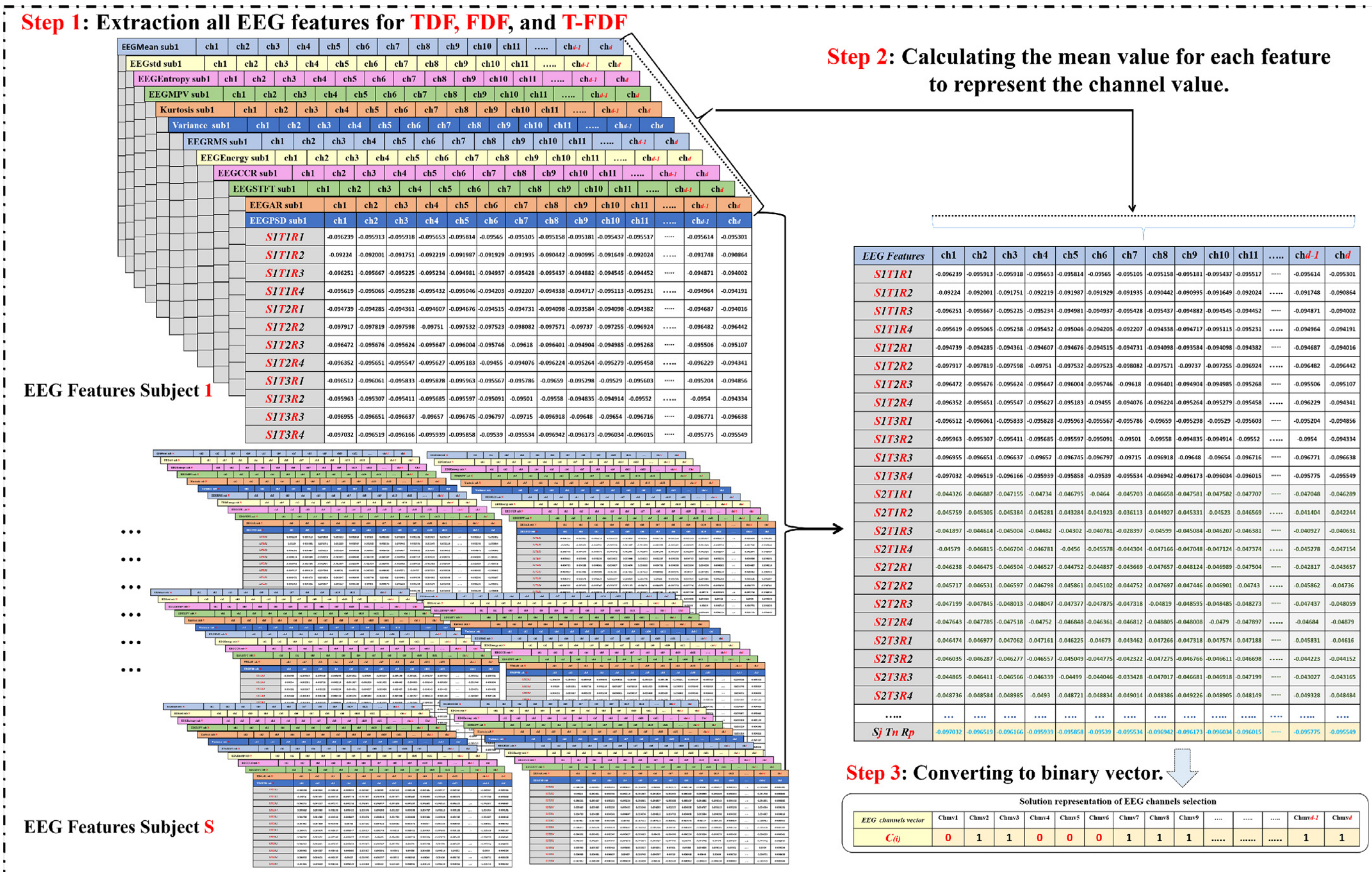


Fig. 2. EEG dataset representation.

2.2. Modelling of EEG channel selection features

To model the features of the EEG channel selection problem, we have to know how to represent the captured EEG signal inside a standard dataset. In general, the EEG dataset can be represented as a matrix of size $K \times d$, where K is calculated as $S \times R \times T$, S denotes the number of subjects, R denotes the number of trials and T denotes the number of tasks. Each EEG channel (sensor) can capture brain activity from the human scalp. Th activity is then presented as a single set of raw EEG data. The total number of EEG channels is presented as a vector of d channels, $\mathcal{C} = (ch_1, ch_2, \ldots, ch_d)$. Each of these channels is represented as a set of features that can be extracted from the original EEG (e.g. the features explained in Section 2.1). For instance, ch_i can be represented as a set of $\{EEG_{Mean}(i), EEG_{Std}(i), EEG_{Energy}(i), \ldots, EEGSTFT_x(i)\}$, where i refers to the channel number between within $(1, 2, \ldots, d)$. Frequently, the current EEG dataset cannot be modelled into the EEG channel selection problem because the high dimensionality of the current EEG dataset leads to a complex problem. For this case, the mean value (i.e. *Chmv*) is calculated for each feature to represent the channel value to be stored on the corresponding location of that channel in the final EEG dataset (i., e., $\mathcal{C} = (Chmv_1, Chmv_2, \ldots, Chmv_d)$).

$$Chmv_i = \frac{\sum_{i=1}^{d} (EEG_{Mean}(i) + EEG_{Std}(i) + \ldots + EEGSTFT(i))}{K}$$

Fig. 2 (step 1 and 2) shows the final EEG dataset representation of EEG data recorded from several subjects.

Notably, each subject can record several tasks and trials for the same task (see Eq. (14)). This represents the final EEG dataset with K records, where K refers to $S \times R \times T$, S denotes the number of subjects, R denotes the number of trials and T denotes the number of tasks.

$$EEG_{features} = \begin{bmatrix} Chmv_1^1 & Chmv_2^1 & \cdots & Chmv_d^1 \\ Chmv_1^2 & Chmv_2^2 & \cdots & Chmv_d^2 \\ \vdots & \vdots & \ldots & \vdots \\ Chmv_1^K & Chmv_2^K & \cdots & Chmv_d^K \end{bmatrix}. \quad (14)$$

Notably, not all of these features are useful for final decisions. Several of these features affect the efficiency of the results by increasing the misclassification rate (i.e. using all of these features affects the unique pattern of the EEG signal). Therefore, only useful EEG features with the highest accuracy rate must be used. One of the best ways to solve this problem is implementing a feature selection technique to select optimal EEG channels.

In short, the EEG channel selection solution can be represented as a binary vector $\mathcal{C} = (Chmv_1, Chmv_2, \ldots, Chmv_d)$ of d channels, where $Chmv_i = 1$ means that channel i is selected and 0 otherwise. The conversion of the mean value of the channel (*Chmv*) is performed based on the transfer function of sigmoid (Eq. (21)). Fig. 2 (step 3) shows an example of binary solution representation of EEG channel selection. Optimal channels are selected according to an objective function such as in Section 2.3 where the best channels that achieved the best results are selected.

2.3. Objective function

This section describes in details the objective function of EEG channels selection. However, we must first know the measures that directly affect the objective functions of EEG channel selection. These measures can be summarised as follows:

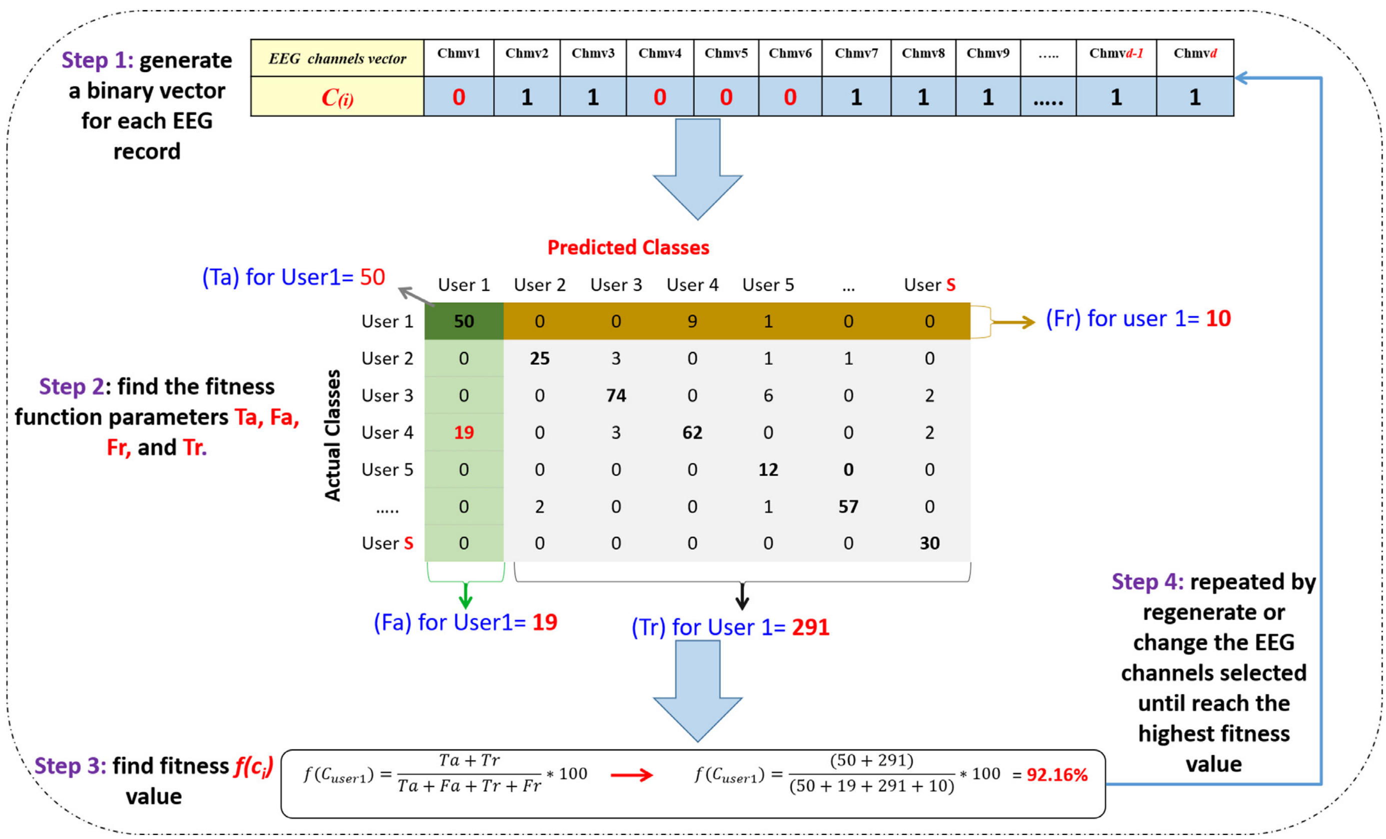


Fig. 3. Solution representation of EEG channel selection.

1. ***True accept*** *(T_a)* is the percentage measure of valid matches. It is the number of times (in percentage) the system recognises authorised users as genuine users.
2. ***True reject*** *(T_r)* is the measure of times (in percentage) the system recognises unauthorised users as **impostors**. It is the percentage measure of rejecting invalid users.
3. ***False accept*** *(F_a)* is the percentage measure of invalid matches. It is the number of times (in percentage) the system recognises unauthorised users as genuine users. For a robust biometric system, this error must be as low as possible.
4. ***False reject*** *(F_r)* is the measure of times (in percentage) the system recognises unauthorised users as **impostors**. It is the percentage measure of rejecting valid inputs. From the user's point of view, this number must be as low as possible.

The objective function used to evaluate the classification performance of EEG channel selection in this work is formulated in Eq. (15), as suggested by Xue et al. [25,26].

$$max \quad f(C) = \frac{T_a + T_r}{T_a + F_a + T_r + F_r}, \tag{15}$$

where $f(\boldsymbol{C})$ denotes the objective function and T_a, T_r, F_a and F_r represent the true acceptance, true reject, false acceptance and false reject, respectively.

EEG data are generally divided into training and testing datasets [1]. The main purpose of the training phase is to select the optimal EEG channel set that can achieve the highest accuracy rate. During the running time of the algorithm, the features of each single EEG row as visualised in Fig. 2 are converted into binary values and passed to a classifier technique to calculate the accuracy rate. This case is repeated within each iteration of the algorithm. After a certain number of iterations, the best EEG channel set (optimal set) is selected and represented as a binary vector, as shown in Fig. 3 (step 1), where 1 means that the channel is selected and 0 otherwise. With the selection of the selected optimal EEG channel set, the training phase is achieved. Notably, the final results on the accuracy rate are calculated according to these features of the selected channels in the testing dataset. For instance, Fig. 3 presents the procedure of calculating the final accuracy rate ($f(C)$) for one person. Step 1 shows how to generate the binary value for EEG channel selection. Then, this binary vector is passed to a classifier, such as SVM, KNN, to find objective function parameters T_a, T_r, F_a and F_r, as shown in Step 2. T_a represents the true acceptance percentage of person i and indicates how many times the classifier correctly classified the EEG features of person i. F_a represents the false acceptance percentage of person i and shows how many times the classifier classified the EEG features from other persons as those of person i. F_r represents the false reject percentage of person i and indicates how many times the classifier classified the EEG features of person i as those of other persons. In Step 3, the final accuracy rate ($f(C)$) is calculated by repeating these three steps until the highest accuracy rate is reached.

3. Background

This section explain in details the main concepts of the flower pollination algorithm and β-hill climbing algorithm. Section 3.1 describes the fundamentals of the flower pollination algorithm. Section 3.2 explains the fundamentals of β-hill climbing algorithm.

3.1. Fundamentals of the flower pollination algorithm

FPA is a nature-inspired algorithm which introduced by Yang in 2012 [15]. It is inspired from analogous to the pollination behavior of flowering plants. The main idea of the standard version of FPA can summarize by the following concepts:

Concept 1 Local pollination of FPA, which is represented the abiotic and self-pollination in nature.
Concept 2 Global pollination of FPA which is represented the biotic and cross-pollination in nature where pollinators carry the pollen-based on Levy flights law.
Concept 3 The probability of reproduction can be considered that the stability of the flower corresponds to the similarity between any two flowers.
Concept 4 External factors, such as wind or distance between flowers, which are affected on the global and local pollination. Therefore, the balancing between global and local pollination can be controlled by switch probability $p \in [0, 1]$.

In general, we can summarize FPA procedure in five steps which are shown as follows:

Step 1: *Initialization parameters*. Parameters for both FPA and the problem which we try to solved must be initialized within possible range parameters value $\boldsymbol{x}$. Therefor, the general formulation of the FPA initialization can be generalized as follows:

$$\min \text{ or } \max\{f(\boldsymbol{x})\,|\,\boldsymbol{x} \in \mathbf{X}\},$$

where $f(\boldsymbol{x})$ is the objective function; $\boldsymbol{x} = \{x_i\,|\,i = 1, \ldots, d\}$ is the set of decision variables. $\mathbf{x} = \{\boldsymbol{x}_i\,|\,i = 1, \ldots, d\}$ is the possible value range for each decision variable, where $\boldsymbol{C}_i \in [LowerB_i, UpperB_i]$, where $LowerB_i$ and $UpperB_i$ are the lower and upper bounds for the decision variable C_i respectively and d is the number of decision variables.
Also, other FPA parameters should be initialized as well, where these parameters can be summarized as follows:

- FPA_s: representing the population size (Number of flowers).
- G^*_{best}: representing the best current solution from the initialized population size.
- Switch probability P: Where the P value will determine to FPA to follow either global or local pollination.
- L_{dis}: Refers to a step size, is the strength of the pollination.

The next steps will provide a full explanation of these parameters.

Step 2: *Initialize FPA population memory*. The flower population memory (FPM) can be represented as a 2-dimensional matrix with size $FPA_s \times d$ which contains sets of flower location vectors as many as FPA_s (see Eq. (20)). Where these flowers are randomly generated as follows: $x_i^j = LowerB_i + (UpperB_i - LowerB_i) \times U(0, 1), \ \forall i = 1, 2, \ldots, z$ and $\forall j = 1, 2, \ldots, FPA_s$, and $U(0, 1)$ generates a uniform random number between 0 and 1. The generated solutions are stored in the FPM in ascending order according to their objective function values where $f(\boldsymbol{x}^1) \le f(\boldsymbol{x}^2) \le \ldots \le f(\boldsymbol{x}^{FPA_s})$.

$$\text{FPM} = \begin{bmatrix} x_1^1 & x_2^1 & \cdots & x_d^1 \\ x_1^2 & x_2^2 & \cdots & x_d^2 \\ \vdots & \vdots & \cdots & \vdots \\ x_1^{FPA_s} & x_2^{FPA_s} & \cdots & x_d^{FPA_s} \end{bmatrix}. \tag{16}$$

Also in this step, the global best flower location G^*_{best} is memorized where $G^*_{best} = x^1$.

Step 3: *Intensification of the current flower population* As we mentioned above the (p) value will determine to the pollinator which path will follow either global or local pollination as follows:

- Local Search of FPA (abiotic) The pollination of this type occurs without any pollinators. That means, it occurs based on the wind and diffusion to transfer the pollen. The local pollination and flower constancy represented as follows:

$$x_i^{t+1} = x_i^t + \epsilon(x_j^t - x_k^t) \tag{17}$$

where x_j^t and x_j^k are pollens from the different flowers of the same plant type. This essentially mimic the flower constancy in a limited neighborhood. Mathematically, if x_j^t and x_j^k comes from the same species or selected from the same population, this become a local random walk if we draw ϵ from a uniform distribution in [0,1].

- Global Search of FPA (biotic) In this type of pollination the flowers pollens are transferred by pollinators such as bees, birds, bats.etc. to long distances. This ensures the pollination and reproduction of the most fittest. Therefore, we can represent the procedure of biotic FPA as follows:

$$x_i^{t+1} = x_i^t + L_{dis} * (G^*_{best} - x_i^t) \tag{18}$$

Where x_i^{t+1} the pollen i or solution vector x_i at iteration t, and G^*_{best} is the current best solution found among all solutions at the current iteration. The parameter L_{dis} is the strength of the pollination, which essentially is a step size. Since insects may move over a long distance with various distance steps, we can use a Levy flight to mimic this characteristic efficiently [1,15,27]. That is, we draw $L_{dis} > 0$ from a Levy distribution

$$L_{dis} \sim \frac{\lambda\Gamma(\lambda) sin(\pi\lambda/2)}{\pi} \frac{1}{Q^{1+\lambda}}, (Q >> s_0 > 0) \tag{19}$$

$\Gamma(\lambda)$ denotes the standard gamma function, and this distribution is valid for large steps Q > 0. In all our simulations below, In this study the (λ) used equal (1.5).

Step 4: *Updating the best solution* (G^*_{best}). During for each iteration in FPA procedure, the global best flower location G^*_{best} will be updated if $f(\boldsymbol{x}'^j) < f(G^*_{best})$.

Step 5: *Stop condition*. FPA repeats step 3 and step 4 until the termination criterion is met. The termination criterion is normally met based on some criterion, such as the number of iterations or the quality of the final outcomes.

3.2. β-hill climbing algorithm

Hill climbing can be considered as one of the simplest optimisation technique to find the local optimal solution. In general, as other local search techniques, the iterative approach of the hill climb algorithm begins with the creation of an arbitrary solution to the problem and then proceeds with a trajectory search for a better solution than the previous solution. The previous process is repeated until you reach the local optima that the solution can no longer be improved [28,29].

However, the original hill climb algorithm suffers from several problems: the most important of which is that it only accepts the uphill movement, often leading to stuck in local optima [30]. Therefore, several extensions of the hill climb algorithm have been proposed to overcome such problem. β-hill climbing, an extension to hill climbing, was proposed by Al-Betar [28]. Where he proposed to add one operator called β-operator controlled by β parameter (i.e., $\beta \in [0, 1]$). This operator is used to achieve the appropriate balance between exploration and exploitation during the search process to eliminate the problem of falling into to stuck in local optima.

To elaborate, suppose the optimisation problem is formulated as follows:

$$\min\{f(\boldsymbol{x})\,|\,\boldsymbol{x} \in \mathbf{X}\},$$

where $f(\boldsymbol{x})$ refers to the objective function which will evaluate the new solution $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)$ where the new solution contains a set of decision variables. Each decision variable $x_i \in \boldsymbol{X}_i$ where $\mathbf{X} = \{\boldsymbol{X}_i \,|\, i = 1, \ldots, d\}$ is the possible value range for each decision variable. Note that the $\boldsymbol{X}_i \in [LowerB_i, UpperB_i]$, and $LowerB_i$ and $UpperB_i$ are the lower and upper bounds for the decision variable x_i respectively and d is the total number of decision variables.

As mentioned above, the β-hill climbing algorithm is a trajectory search technique that begins with single random solution, $x = (x_1, x_2, \ldots, x_d)$. During the running time, the new solution, $solnew = (x_1, x_2, \ldots, x_d)$, must be created by modifying the current solution using two operators namely: $\mathcal{N}$-operator and β-operator, which function as the main sources for exploitation and exploration, respectively. Specifically, the $\mathcal{N}$-operator works as neighbourhood search, while β-operator works similar to mutation operator. At each iteration, the new solution can be enhanced by $\mathcal{N}$-operator stage or β-operator stage until the optimal solution is reached.

When the algorithm begins to generate the solution randomly, then the solution is evaluated using the objective function $f(\boldsymbol{x})$. The solution is then modified using $\mathcal{N}$-operator, which employs the $improve(\mathcal{N}(\boldsymbol{x}))$ function within a random range of its neighbors. The solution $\boldsymbol{x}$ is as follows:

$$solnew_i = sol_i \pm U(0,1) \times bw \qquad \exists i \in [1, d]$$

Where i is randomly selected from the space range, $i \in [1, 2, \ldots, d]$. The parameter bw representees the bandwidth between the current value and the new value.

In β-operator, within the β range where $\beta \in [0, 1]$, variables of new solution will be assigned based on selected randomly from available range or from the existing values of the current solution as follows:

$$solnew_i \leftarrow \begin{cases} x_r & rnd \le \beta \\ x_i & otherwise. \end{cases}$$

Where rnd generates a uniform random number between 0 and 1 and $x_r \in \boldsymbol{X}_i$ is the possible range for the decision variable $solnew_i$.

4. EEG Channel selection using hybridizing FPAβhc with RBF-SVM classifier: proposed method

To select the optimal subset of EEG channels, this section provides in detail the full explanation of the proposed method for EEG channel selection based on hybridizing the FPA with the β-hill climbing algorithm (FPAβ-hc). Fig. 4 shows flowchart of the proposed method. The procedural steps of the proposed method are described in detail below.

Step 1: Initialization parameters. The parameters for FPA, β-hill climbing algorithm, and EEG channel selection problem must be initialised within a possible range of parameter values. The utilisation of FPA initialisation for channel selection can be given as follows:

$$\max\{f(\boldsymbol{C}) \,|\, \boldsymbol{C} \in \mathbf{X}\},$$

where $f(\boldsymbol{C})$ is the objective function and $\boldsymbol{C} = \{Chmv_i \,|\, i = 1, \ldots, d\}$ is the set of channels. ***Chmv_i*** is equal to the mean value of EEG features in position i, and d is the total number of EEG channels (Section 2.2). Other parameters for FPA, β-hc and the EEG channel selection problem should be initialised as well, and these parameters can be summarised as follows:

- FPA_s: represents the population size (number of flowers).
- G^*_{best}: represents the best current solution from the initialised population size that provides the highest accuracy rate.
- Switch probability P: Determines whether FPA will follow either global or local pollination for the selection of the optimal EEG channel set.
- L_{dis}: Refers to step size and is the strength of the pollination.
- d: Refers to the total number of EEG channels and represents the solution size.
- bw: Refers to the bandwidth between the current value and the new value.
- β-operator: $\beta \in [0, 1]$.

The next steps show how these parameters are used.

Step 2: Initializations of flower population memory (FPM). FPM can be represented as a 2D matrix with size $FPA_s \times d$ where FPA_s is calculated as $S \times R \times T$, (S denotes the number of subjects, R denotes the number of trials, T denotes the number of tasks, and d refers to the number of channels) (Eq. (20)). These flowers are created from the EEG recorded and stored in FPM in ascending order according to their objective function values, where $f(\boldsymbol{C}^1) \le f(\boldsymbol{C}^2) \le \ldots \le f(\boldsymbol{C}^{FPA_s})$.

$$\text{FPM} = \begin{bmatrix} Chmv_1^1 & Chmv_2^1 & \cdots & Chmv_d^1 \\ Chmv_1^2 & Chmv_2^2 & \cdots & Chmv_d^2 \\ \vdots & \vdots & \cdots & \vdots \\ Chmv_1^{FPA_s} & Chmv_2^{FPA_s} & \cdots & Chmv_d^{FPA_s} \end{bmatrix} \tag{20}$$

Step 3: Improvement Loop. According to the number of flowers N, FPA repeats the following procedure to find the optimal subset of the EEG channels to achieve the highest accuracy rate.
The (p) value helps the pollinator determine which path to follow (either global or local pollination) as follows:

Step 3.1: Local pollination FPA selects two solutions j and k randomly from FPM to manipulate them to generate a new solution C_i^{itr} (see 17).

Step 3.2: Global pollination The new solution is generated using the current solution with the current best solution G^*_{best} after manipulation with the strength parameter of pollination L_{dis} (see 18).

Step 3.3: Transform to binary by sigmoid The proposed method uses the standard version of FPA, which adopts continuous-valued positions to update the solution in the search space. However, the EEG channel selection problem is classified as a binary vector problem which means (0 and 1), where 1 refers to the selected channel and 0 refers to the non-selected channel [1,31]. Therefore, FPA is converted to the binary version to address the EEG channel selection problem; the solution can be represented as a binary vector $\mathcal{C} = (Chmv_1, Chmv_2, \ldots, Chmv_d)$ of d channels, where $Chmv_i = 1$ means that channel i is selected and 0 otherwise. For restricting binary solutions based on FPAβ-hc, two equations (Eqs. (21) and (22)) are used to build this binary vector.

$$sigmoid_{(C_i^{itr}(t))} = \frac{1}{1 + e^{-C_i^{itr}(t),}} \tag{21}$$

$$C_i^{itr}(t) = \begin{cases} 1 & C_i^{itr}(t) > \sigma \\ 0 & otherwise, \end{cases} \tag{22}$$

where σ is a random number between 0 and 1.

Step 3.4: β-hill climbing algorithm (β-hc) To improve the behaviour of standard FPA for EEG channel selection, this study proposes hybridizing the standard FPA with the β-hc algorithm. β-hc, a the local search technique, takes the

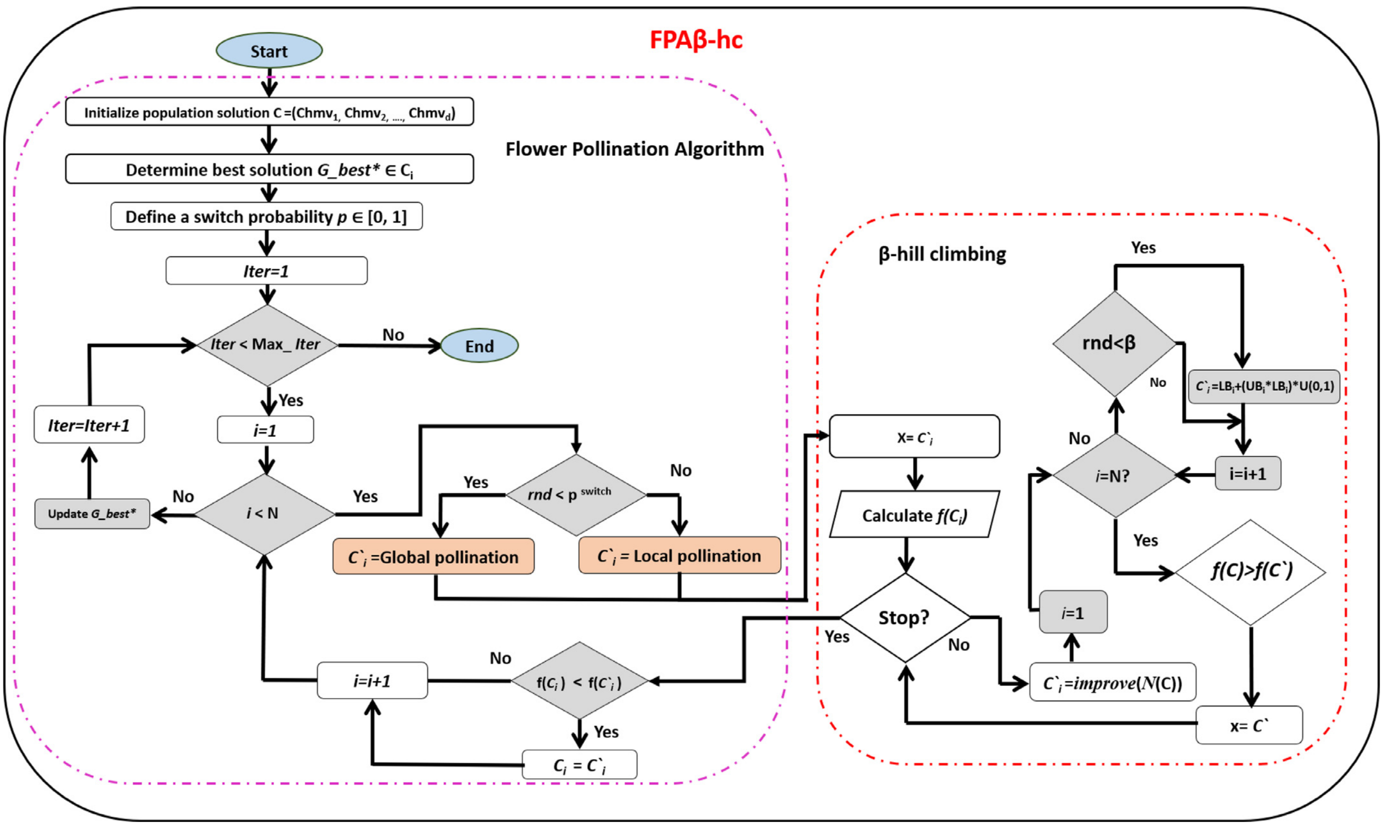


Fig. 4. Flowchart of hybridizing FPA with β-hill climbing.

current solution C_i^{itr} from FPA (either global or local pollination) and tries to improve it. If the current solution C_i^{itr} is improved, the new solution ($New - sol_i^{\prime\prime itr}$) will replace the previous solution C_i^{itr}.

Step 4: RBF-SVM classifier The improved solution by the β-hc algorithm ($New - sol_i^{\prime\prime itr}$) is evaluated using the RBF-SVM classifier to calculate its objective function of the accuracy rate of EEG channels selection (Eq. (15)). Then, if ($f(New - sol_i^{\prime itr})$ $>$ f(C_i^{itr}), the current best solution will be replaced by the new solution.

Step 5: Update the population The current best solution is replaced when improvement is achieved. Therefore, FPAβ-hc algorithm checks the current best solution with the global best flower location G^*_{best} during each iteration. The global best flower location G^*_{best} will be updated if $f(C_i^{itr})$ $>$ $f(G^*_{best})$.

Stop 6: Stop criteria FPA repeats steps 3 and 5 until the termination criterion is met. The termination criterion is normally met based on another criterion, such as the number of iterations or the quality of the final outcomes.

Step 7: Output Return the G^*_{best} **best channel subset with the highest accuracy rate.**

Algorithm 1 pseudo-codes the proposed method that employs BFPAβhc for EEG channel selection by using the RBF-SVM classifier as the objective function and Eqs. (21) and (22) as a transfer function.

5. Results and discussions

This section explains the performance of the proposed method (i.e. FPAβ-hc) for EEG channel selection. Section 5.1 describes the EEG dataset used in this work. The parameter setting and experimental setup are introduced in Section 5.2. Section 5.3 compares the results of standard FPA with RBF-SVM and hybridizing FPAβ-hc with RBF-SMV classifier. Section 5.4 presents the comparison results of the proposed method FPAβ-hc with state-of-the-art approaches for EEG channels selection.

5.1. EEG Dataset

EEG signal acquisition is performed over a standard EEG signal dataset [32]. The EEG signals are collected from 109 healthy volunteers using a brain-computer interface software called BCI2000 system [16]. The EEG signals are captured from 64 sensors (i.e. electrodes), and each subject performs 12 motor/imagery tasks that are mainly used in different fields, such as neurological rehabilitation and brain-computer interface applications. In general, the tasks involve imagining or motor movement, such as opening and closing of the eyes. The signals are recorded from each person by requiring them to perform four tasks according to the position of the target appearing on the screen in front of them, as follows:

- **Task(1):** A subject is asked to open and close his/her fist corresponding to the position of the target on the screen. If the target appears on the right or left side of the screen, then the subject relaxes.
- **Task(2):** A subject is asked to imagine opening and closing his/her fist corresponding to the position of the target on the screen. If the target appears on the right or left side of the screen, then the subject relaxes.
- **Task(3):** A subject is asked to open and close both fists or both feet. If the target appears on either the bottom or the top of the screen, then the subject relaxes.

Algorithm 1 Hybridizing Flower Pollination Algorithm with β-hill climbing (FPAβ-hc) for EEG Channel Selection.

Input:
Initialize the problem and FPA parameters
Initialize FPA population and select current best solution G^*_{best}
Channels= $\{ch_1, ch_2, \ldots, ch_d\}$
for $a = 1$ to N **do**
  Evaluate fitness value of $f(C)$ based on 10-fold-CSV RBF-SVM and accuracy rate of EEG channels selection [equation 15]
end for
Find G^*_{best}, where $G^*_{best} \in (1, 2, \ldots, N)$
$itr = 0$
while $itr < Total\_iterations$ **do**
  for $j = 1$ to N **do**
    for $i = 1$ to $number\,of\,channels(d)$ **do**
      if $rnd \le p$ **then**
        Global pollination via $C_i^{itr} = C_i^{itr-1} + L_{dis} * (G^*_{best} - C_i^{itr-1})$
        $sigmoid(C_i^{itr}) = \frac{1}{1+e^{-C_i^{itr}}}$
        if $sigmoid(C_i^{itr+1}) > U(0,1)$ **then**
          $C'^{itr}_{i,j} = 1$
        else
          $C'^{itr}_{i,j} = 0$
        end if
      else
        Do local pollination via $C_i^{itr} = C_i^{itr-1} + \epsilon\,(C_j^{itr} - C_k^{itr})$
        $sigmoid(C_i^{itr}) = \frac{1}{1+e^{-C_i^{itr}}}$
        if $sigmoid(C_i^{itr}) > U(0,1)$ **then**
          $C'^{itr}_{i,j} = 1$
        else
          $C'^{itr}_{i,j} = 0$
        end if
      end if
    end for
    Run β-hill climbing algorithm using $C'^{itr}_{i,j}$.
    while Stop criterion is not met **do**
      $New - sol'^{itr}_{i,j} = \mathcal{N} - Operator(C'^{itr}_{i,j})$
      $New - sol''^{itr}_{i,j} = \beta - Operator(New - sol'^{itr}_{i,j})$
      Calculate fitness value of $f(New - sol''^{itr}_{i,j})$ using RBF-SVM classifier for EEG channels selection [equation 15]
      if $f(New - sol'^{itr}_{i,j}) < f(New - sol''^{itr}_{i,j})$ **then**
        replace $(New - sol'^{itr}_{i,j})$ by $(New - sol''^{itr}_{i,j})$
      end if
    end while
    $sol'^{itr}_{i,j} = New - sol''^{itr}_{i,j}$
  end for
  Update G^*_{best}, *where* $G^*_{best} \in (1, 2, \ldots, N)$
  $itr = itr + 1$
end while
Output
Return G^*_{best}: **best channels subset with highest accuracy rate.**
End

- **Task(4):** A subject is asked to imagine opening and closing both fists or both feet. If the target appears on either the bottom or the top of the screen, then the subject relaxes.

Each person performs four tasks, which are repeated three times for two minutes per recording. The outcome of this phase is 12 records of EEG signals for each person. The EEG signals

Table 1
parameters setting.

Parameters and values	β-hc	FPA
Iterations number ($N - itr$)	100	100
Population size	1	20
Solution size	64	64
β-operator	0.5	–
Switch probability P	–	0.8

are recorded using 64 sensors with 160 samples per second. Then, the EEG features are extracted from these 12 recordings with three different categories, as mentioned in Section 2.1). To reduce the dispersion of the EEG pattern (obtain unique features) and achieve quick processing of the extracted EEG features, the mean value for each electrode is calculated and called ($Chmv_{(}i)$), where i refers to the channel number. This means each electrode is represented by one value. We use the following notations for each of the dataset configuration, such as time domain feature. TDF1 includes the features $\{EEG_{Mean}(i)$, $EEG_{Std}(i)$, $EEG_{Entrpy}(i)$, $EEG_{Energy}(i)$, $EEG_{RMS}(i)\}$. TDF2 includes the features $\{EEG_{VAR}(i)$, $EEG_{MPV}(i), EEG_{Skewness}(i), EEG_{Kurtosis}(i), EEG_{CCR}(i)\}$. TDF includes the combination of TDF1 and TDF2 features, such as $\{EEG_{Mean}(i)$, $EEG_{Std}(i)$, $EEG_{Entrpy}(i)$, $EEG_{Energy}(i)$, $EEG_{RMS}(i)$, $EEG_{VAR}(i), EEG_{MPV}(i), EEG_{Skewness}(i), EEG_{Kurtosis}(i), EEG_{CCR}(i)\}$, where i refers to the channel number. For FDF, FDF1 includes AR features with five orders. FDF2 includes PSD (EEG_{PSD}) features. FDF includes the combination of FDF1 and FDF2 features. T-FDF includes STFT (*EEGSTFT*) features.

5.2. Experimental setup

In the experimental test, the 10fold cross-validation approach is applied, this approach is widely used to validate machine learning algorithms due to its consistency and reduced results variability with regard to input data [33]. The main purpose of using the 10fold cross-validation approach in our dataset is to determine the optimal subset of features that can provide the maximum accuracy, with accuracy being the fitness function. The proposed method (FPAβ-hc) begins to create a mapping between the original EEG dataset and a new scalar feature (i. e. a binary value initialised randomly for each channel, where 1 refers to the selected channel and 0 refers to the non-selected channel). In addition, the fitness function of each row of features is set to RBF-SVM for the training data part, and the accuracy recognition rate is determined over the validation subset. Then, we select the optimal subset from the validation part that provides the highest accuracy rate. This subset is passed to the testing dataset for calculating the final accuracy rate. Table 1 shows the parameters used for FPA and the β-hc algorithm used in this work. $N - itr$ is a parameter to determine the number of iterations used in the experiments.

5.3. Comparing performance of standard FPA and hybridizing FPAβ-hc for EEG channel selection

Given that the proposed method (FPAβ-hc) belongs to metaheuristic algorithms that are non-deterministic, we determine the average of the accuracy rate over 25 rounds by using the proposed method, to avoid biased results. The experiment results are obtained using a LENOVO Ideapad 310 PC with, Intel Core i7 2.59 Ghz processor, 8 GB of RAM and Windows 10 operating system. To evaluate the performance of the proposed FPAβ-hc, we consider five measures, namely, (i) accuracy (EEG_{Acc}), (ii) sensitivity (EEG_{Sen}), (iii) F-score (EEG_{Fs}), (v) specificity (EEG_{Spe}), and (iv) number of channels selected (No. Ch). These measures can formulated

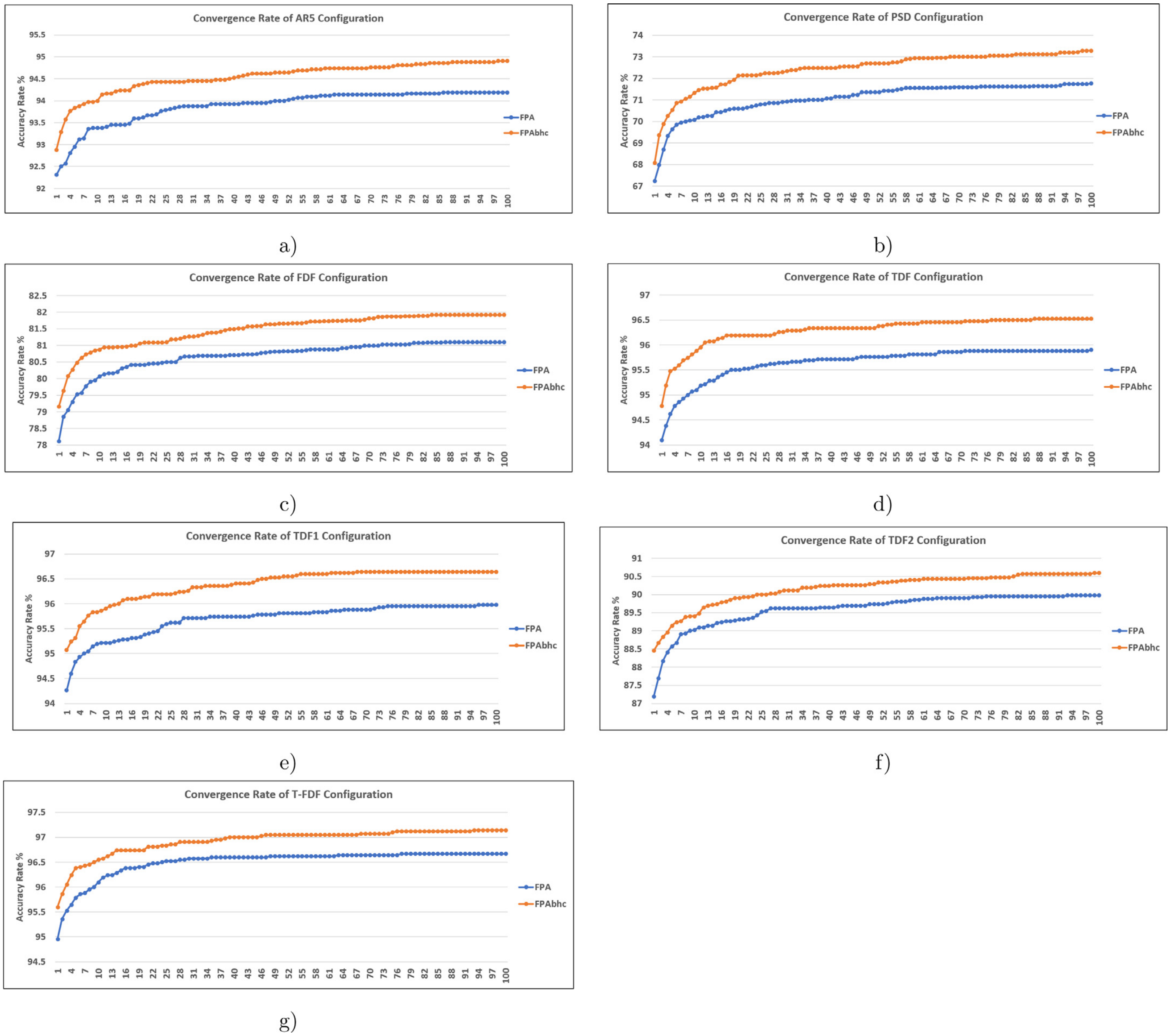


Fig. 5. Convergence rate of hybridizing FPAβ-hc compared with that of standard FPA. Where a) FPA_{AR5} and FPAβ-hc$_{AR5}$, b) FPA_{PSD} and FPAβ-hc$_{PSD}$, c) FPA_{FDF} and FPAβ-hc$_{FDF}$, d) FPA_{TDF} and FPAβ-hc$_{TDF}$, e) FPA_{TDF1} and FPAβ-hc$_{TDF1}$, f) FPA_{TDF2} and FPAβ-hc$_{TDF2}$, and g) FPA_{T-FDF} and FPAβ-hc$_{T-FDF}$.

as follows:

$$EEG_{Acc} = \frac{T_a + T_r}{T_a + F_a + T_r + F_r} \times 100 \tag{23}$$

$$EEG_{Sen}\ (Recall) = \frac{T_a}{T_a + F_r} \tag{24}$$

$$EEG_{Spe} = \frac{T_r}{T_r + F_r} \tag{25}$$

$$Precision(Pre) = \frac{T_a}{T_a + F_a} \tag{26}$$

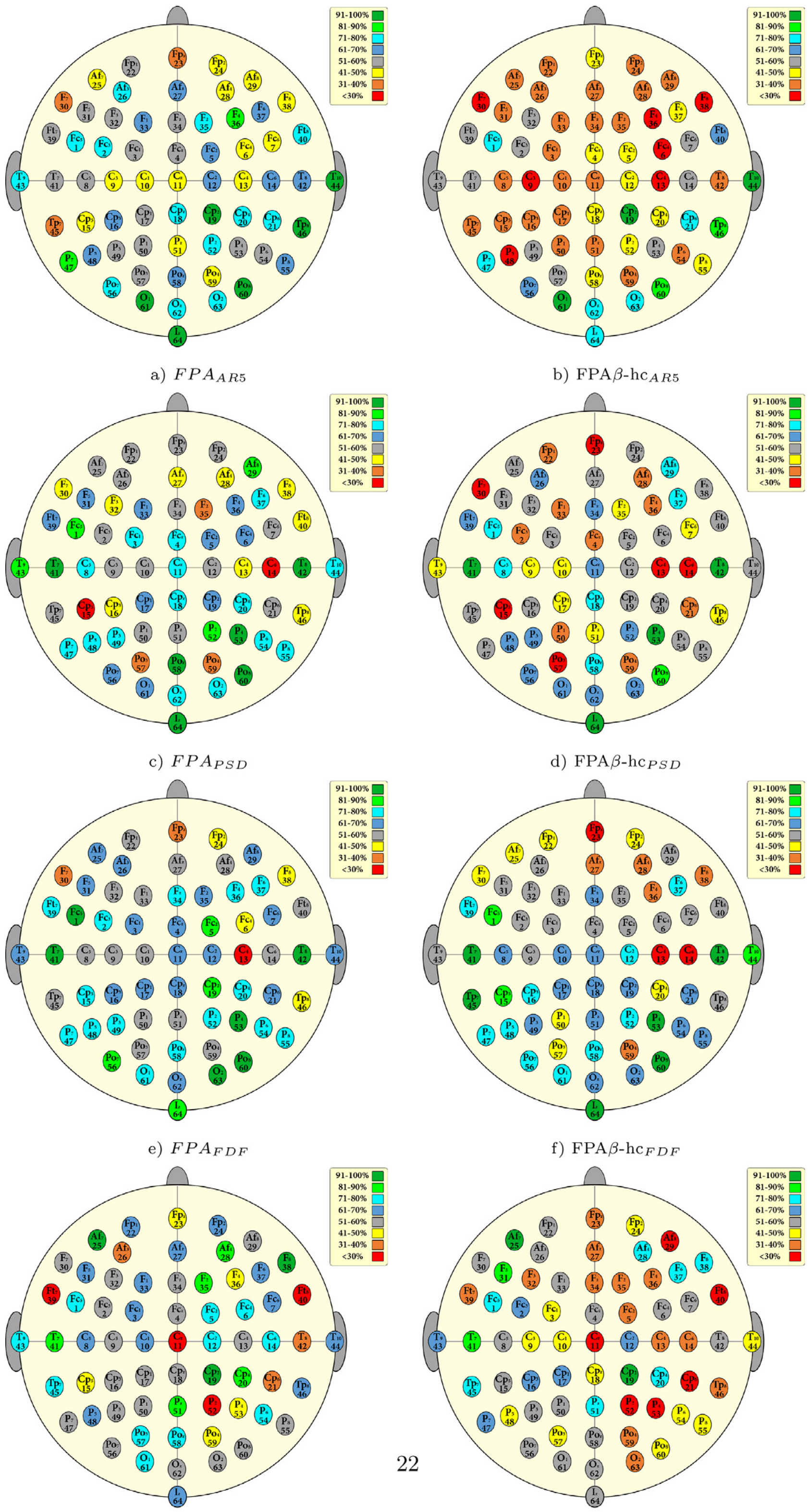
$$EEG_{Fs} = 2 \times \left(\frac{Pre\,.\,Recall}{Pre + Recall}\right), \tag{27}$$

where T_a, T_r, F_a and F_r represent true acceptance, true reject, false acceptance and false reject, respectively.

Figs. 5–7 show the convergence rate and frequency of selected electrodes over 25 runs for standard FPA and the proposed method (FPAβ-hc) during the experimental evaluation using FDF1, FDF2, FDF, TDF1, TDF2, TDF and T-FDF.

Table 2 shows the comparison results of the proposed methods (i. e., FPA and FPAβ-hc) with three method of feature selection which are LASSO [34], Information Gain [35], RelifF [36] for all EEG features extracted from the input EEG signal as follows:

1. TDF group 1, presented as (TDF1), contains features: mean, standard deviation, energy, entropy and root mean square (RMS).
2. TDF group 2, presented as (TDF2), has also five EEG features: variance (VAR), maximum peak value, skewness, kurtosis, and cross correlation.
3. The combination of all TDFs, which is presented as (TDF), merges the features of TDF1 and TDF2.
4. FDF group 1, presented as (FDF1), contains AR features with order 5 (AR5).



a) FPA_{AR5} b) FPAβ-hc$_{AR5}$

c) FPA_{PSD} d) FPAβ-hc$_{PSD}$

e) FPA_{FDF} f) FPAβ-hc$_{FDF}$

g) FPA_{TDF1} h) FPAβ-hc$_{TDF1}$

Fig. 6. Distribution of frequency of selected electrodes for FPA and FPAβ-hc.

5. FDF group 2, presented as (FDF2), contains five features: PSD of the EEG sub-signal of delta δ, theta θ, beta β, alpha α, and gamma γ.
6. The combination of all FDFs, which is presented as (FDF), merges the features of FDF1 and FDF2.
7. T-FDF presented as (T-FDF), includes STFT features.

The results show that the performance of the proposed method (FPAβ-hc) exhibits a significant improvement compared with the standard FPA algorithm based on all the comparison measures.

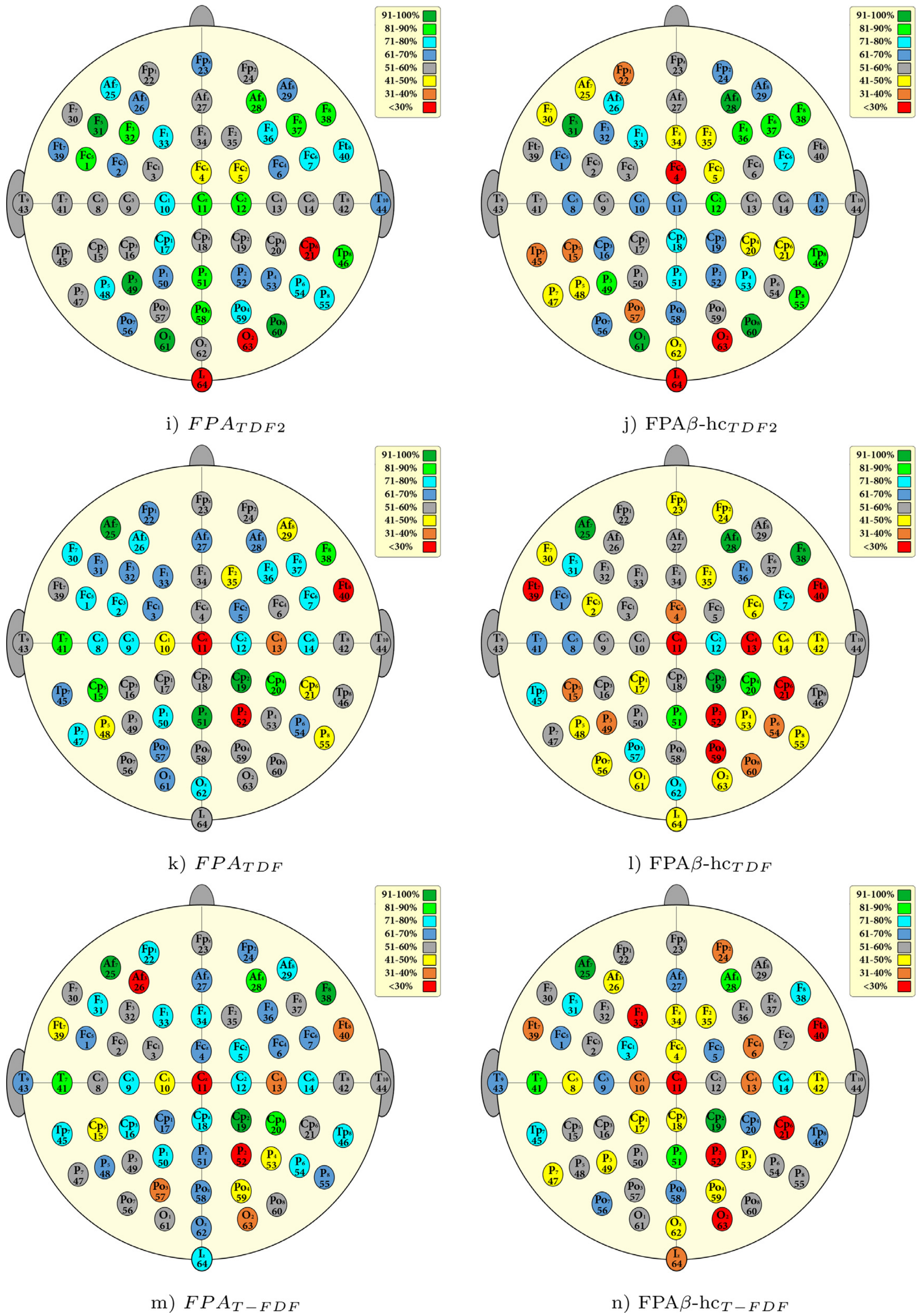


Fig. 7. Distribution of frequency of selected electrodes for FPA and FPAβ-hc.

FPAβ-hc achieves better results than the standard FPA algorithm with all the EEG features extracted. In the FDF1 group, the proposed method obtains accuracy rates of 93.7619, 32, 0.9376, 0.9943 and 0.9383 compared with standard FPA with accuracy rates of 92.9523, 41, 0.9295, 0.9935 and 0.93 the number of channels, sensitivity, specificity, and F1-score, respectively. In the FDF2 group, FPAβ-hc achieves 70.0476, 35, 0.7005, 0.9727 and 0.6916 accuracy rates compared with standard FPA with 68.2857, 42, 0.6828, 0.9711 and 0.6739 accuracy rates for the number channels, sensitivity, specificity, and F1-score, respectively. For the combination of FDF,

Table 2
Comparing performance of proposed method FPAβ-hc with feature selection methods.

Dataset	Measure	FPA	FPAβ-hc	LASSO	Information Gain	RelifF
FDF1(AR5)	Accuracy	92.9523	**93.7619**	86.3095	85.119	85.119
	No.of Channels	41	**32**	16	33	17
	Sensitivity	0.9295	**0.9376**	0.8630	0.85119	0.85119
	Specificity	0.9935	**0.9943**	0.9875	0.9864	0.98647
	F1-Score	0.93	**0.9383**	0.8653	0.8542	0.8545
FDF2(PSD)	Accuracy	68.2857	**70.0476**	51.1905	41.0714	40.4762
	No.of Channels	42	**35**	10	14	10
	Sensitivity	0.6828	**0.7005**	0.5119	0.4107	0.4048
	Specificity	0.9711	**0.9727**	0.9556	0.9464	0.9459
	F1-Score	0.6739	**0.6916**	0.5118	0.4161	0.4296
FDF	Accuracy	79.1666	**79.6428**	68.1548	67.8571	75.5952
	No.of Channels	43	**40**	16	37	45
	Sensitivity	0.7916	**0.7964**	0.6815	0.6786	0.7560
	Specificity	0.981	**0.9814**	0.9710	0.9708	0.9778
	F1-Score	0.7914	**0.7957**	0.6760	0.6829	0.7612
TDF1	Accuracy	95	**95.548**	82.1429	85.1190	81.5476
	No.of Channels	40	**33**	16	33	17
	Sensitivity	0.95	**0.95547**	0.8214	0.8512	0.8155
	Specificity	0.9954	**0.9959**	0.9838	0.9865	0.9832
	F1-Score	0.95	**0.956**	0.8232	0.8524	0.8152
TDF2	Accuracy	88	**88.642**	73.2143	63.6905	64.8810
	No.of Channels	42	**39**	16	39	15
	Sensitivity	0.88	**0.8864**	0.7321	0.6369	0.6488
	Specificity	0.989	**0.9896**	0.9756	0.9670	0.9681
	F1-Score	0.8819	**0.8882**	0.7461	0.6410	0.6601
TDF	Accuracy	94.833	**95.214**	88.0952	91.6667	92.2619
	No.of Channels	40	**34**	16	15	16
	Sensitivity	0.9483	**0.9521**	0.8810	0.9167	0.9226
	Specificity	0.9953	**0.9956**	0.9892	0.9924	0.9930
	F1-Score	0.9493	**0.9529**	0.8811	0.9166	0.9223
T-FDF	Accuracy	95.619	**96.0476**	88.6905	91.6667	91.0714
	No.of Channels	41	**35**	16	13	17
	Sensitivity	0.9561	**0.9605**	0.8869	0.9167	0.9107
	Specificity	0.996	**0.9964**	0.9897	0.9924	0.9919
	F1-Score	0.9569	**0.9611**	0.8870	0.9165	0.9111

Bold value indicates best results.

FPAβ-hc achieves 79.6428, 40, 0.7964, 0.9814 and 0.7957 accuracy rates compared with standard FPA with 79.1666, 43, 0.7916, 0.981, and 0.7914 accuracy rates for the number of channels, sensitivity, specificity, and F1-score, respectively.

With regard to TDF extraction, in the TDF1 group, FPAβ-hc obtains 95.548, 33, 0.9554, 0.9959 and 0.956 accuracy rates compared with standard FPA with 95, 40, 0.95, 0.9954 and 0.95 accuracy rates for the number of channels, sensitivity, specificity, and F1-score, respectively. In the TDF2 group, FPAβ-hc obtains 88.642, 39, 0.8864, 0.9896 and 0.8882 accuracy rates compared with standard FPA with 88, 42, 0.88, 0.989 and 0.8819 accuracy rates for the number of channels, sensitivity, specificity, and F1-score, respectively. For the combination all TDF, FPAβ-hc obtains 95.214, 34, 0.9521, 0.9956 and 0.9529 accuracy rates compared with standard FPA with 94.833, 40, 0.9483, 0.9953 and 0.9493 accuracy rates for the number of channels, sensitivity, specificity, and F1-score, respectively. For T-FDF, the proposed method (FPAβ-hc) achieves the best performance results, where it obtained 96.0476, 35, 0.9605, 0.9964 and 0.9611 accuracy rates compared with standard FPA with 95.619, 41, 0.9561, 0.996 and 0.9569 accuracy rates for the number of channels, sensitivity, specificity, and F1-score, respectively.

To further evaluate the performance of FPAβ-hc, the results are compared against well-known filter methods in the literature of feature selection methods such as ReliefF [36], Information Gain (IG) [35], and LASSO [34]. Conventionally, ReliefF and IG the most important feature ranking methods which evaluate each feature independently according to its relevance to class labels and the top K features are chosen as a final subset of features. In other hands, LASSO is also one of the most common types for embedded feature selection methods. It produces a subset of features and evaluates them using machine learning algorithms. The results of filter methods in Table 2 are reported based on multiple experiments using a various number of features (i.e., top K = 5, K = 10, K = 15, etc). It can be seen that FPAβ-hc outperforms other filter-based methods on almost all evaluation measures, except a number of channels selected. LASSO resulted in the smallest number of channels on most of the datasets; however, it produced less classification accuracy on all datasets when compared with FPAβ-hc. In classification systems, higher classification accuracy with a reasonable increase in number of channels is more desirable than lower classification accuracy with smaller number of channels. In a nutshell, the results prove that integration between FPA and β-hc promotes its local exploitation process in finding the most discriminative subset and, thus, produced more accurate and reliable identification system.

Figs. 8–10 show the performance of the proposed method compared standard FPA algorithm using accuracy rate, the number of channels, sensitivity, specificity, and F1-score.

Then, we perform the Wilcoxon signed-rank statistical test [37] to verify whether a significant difference exists between FPA and FPAβ-hc. Table 3 shows a comparison of all EEG features extracted using FPA and FPAβ-hc (Fig. 11).

5.4. Comparison with state-of-arts

The proposed method (FPAβ-hc) is compared with state-of-the-art approaches [1] by using the same dataset and feature extraction, namely, AR features with five-order coefficients called AR5. However, the other approaches used binary metahurstic algorithms with the OPF classifier. The performance of the proposed method is compared with five that of optimisation methods (binary ge-

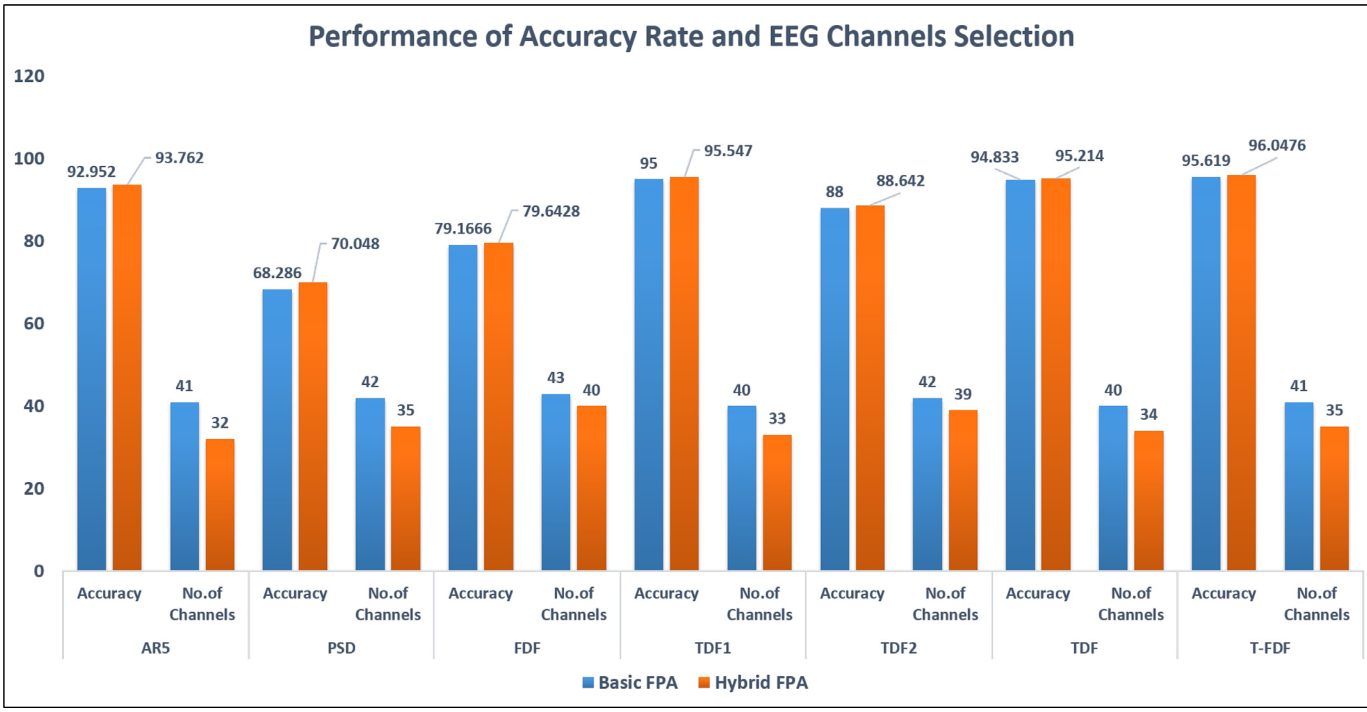


Fig. 8. Performance results of FPA and FPAβ-hc using accuracy rate and number of channels selected.

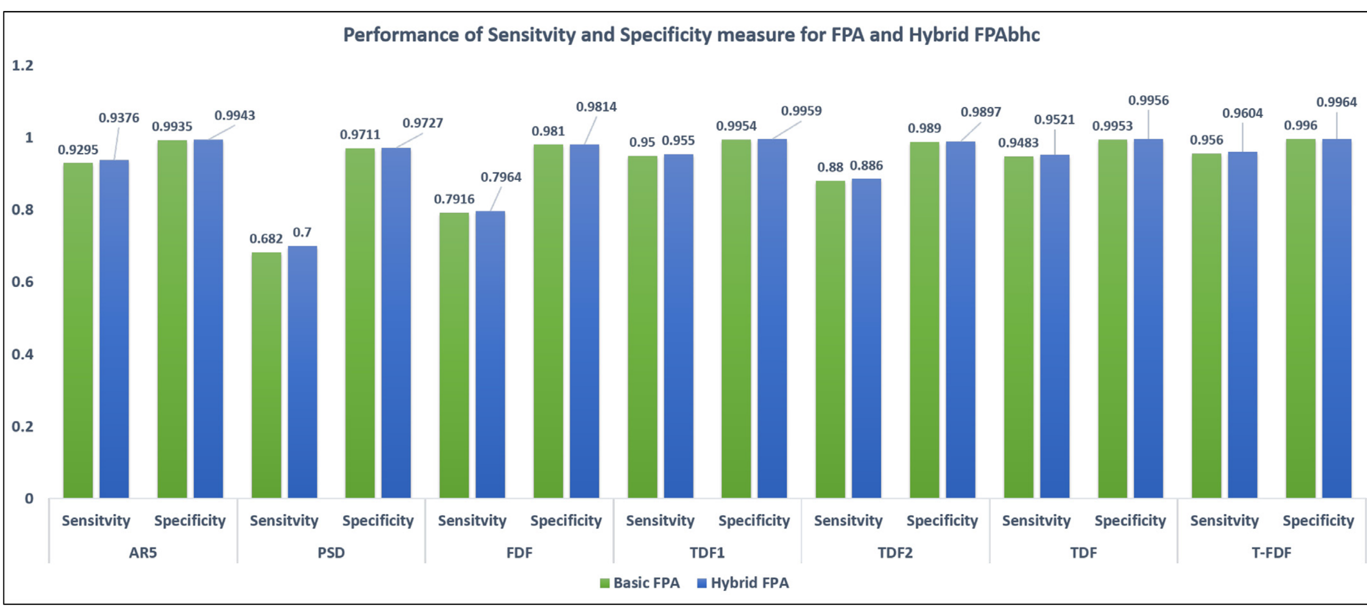


Fig. 9. Performance results of FPA and FPAβ-hc using sensitivity and specificity measures.

Table 3
Wilcoxon signed-rank test evaluation. of FPA and HyFPAβ-hc.

Dataset	*P*-Value	*W*-value	Mean Difference	Sum of pos. ranks	Sum of neg. ranks	*Z*-value	Mean (W)	Std(W)	T-Sig	FPAβ-hc
AR5	0.05	16.5	0.17	16.5	214.5	−3.441	115.5	28.77	0.00058	+
PSD	0.05	3	−4.81	3	273	−4.106	138	32.88	0	+
FDF	0.05	20	−0.78	20	280	−3.7143	150	35	0.0002	+
TDF1	0.05	0	−0.57	0	210	−3.9199	105	26.79	8.00E-05	+
TDF2	0.05	8	−1.44	8	182	−3.5011	95	24.85	0.00046	+
TDF	0.05	18	−0.57	18	258	−3.6498	138	32.88	0.00026	+
T-FDF	0.05	4	−1.49	4	132	−3.3094	68	19.34	0.00049	+

Std is Standard Deviation (W), + refers to Significant.

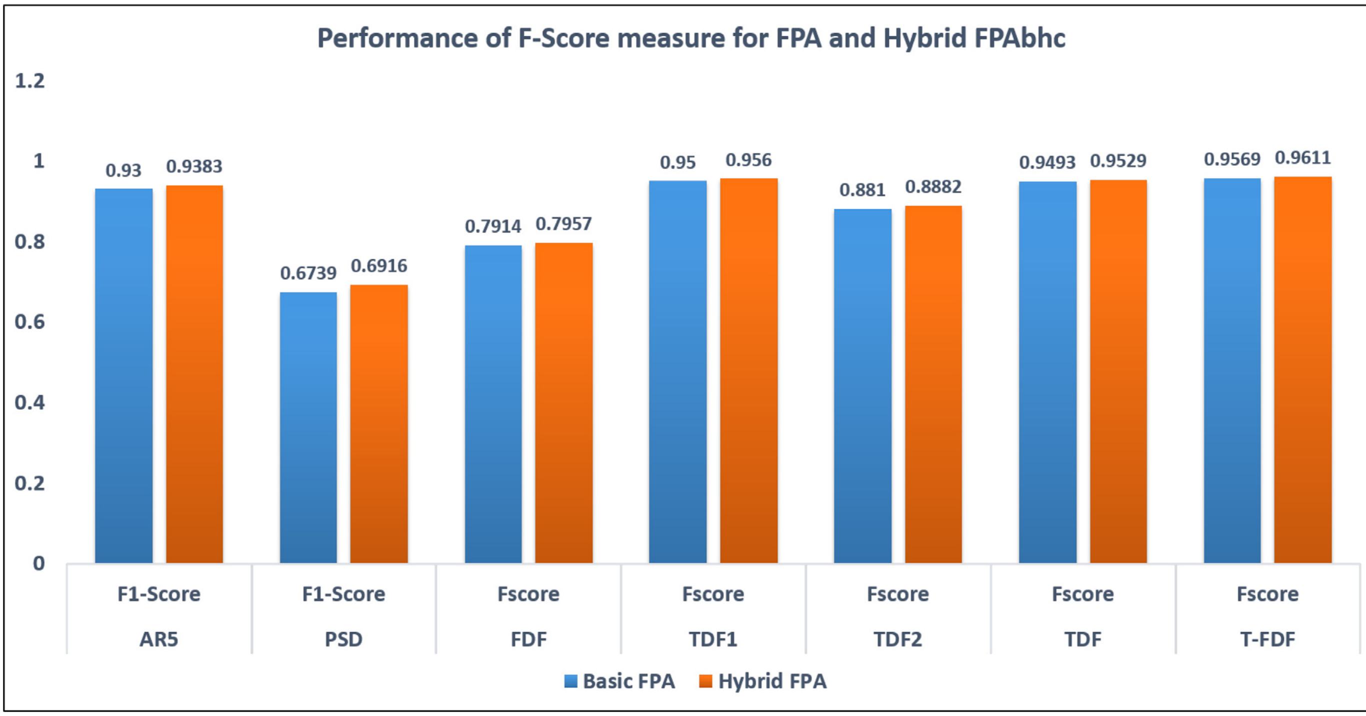


Fig. 10. Performance results of FPA and FPAβ-hc using the F1-score measure.

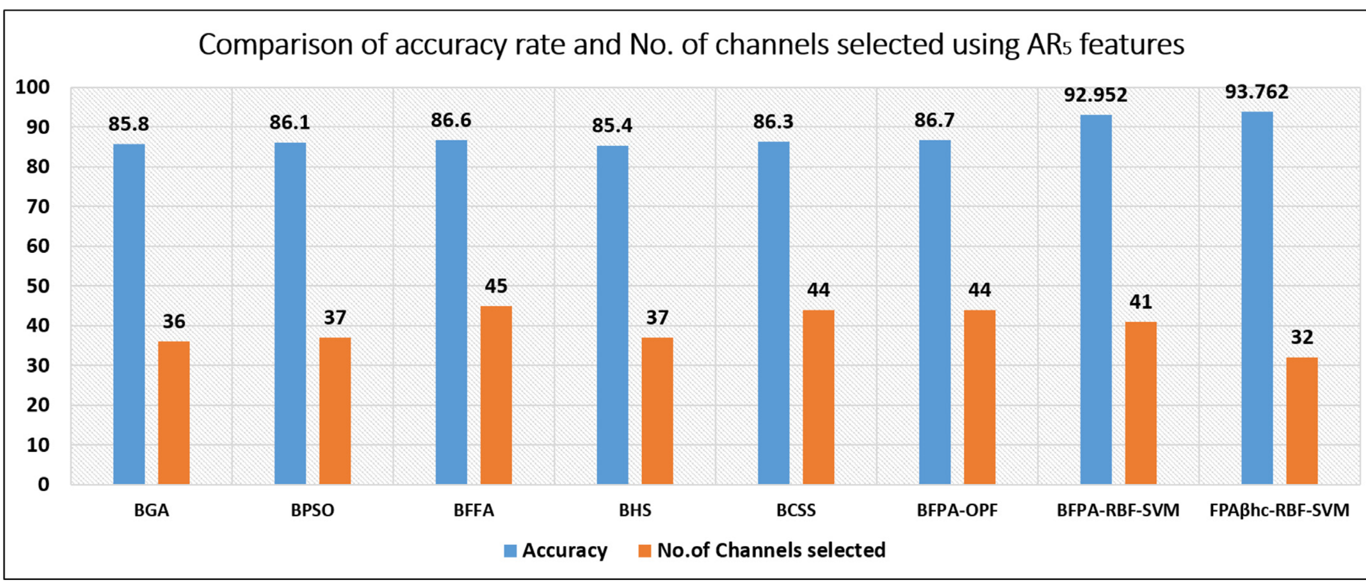


Fig. 11. Comparison of the accuracy rate and No. of EEG channels selected using AR5.

netic algorithm (BGA),binary particle swarm optimisation (BPSO), binary firefly algorithm (BFFA), binary harmony search (BHS), and binary charged system search (BCSS)), and the proposed method is ranked first. The comparison involves two criteria, which are accuracy rate and number of channels selected. FPAβ-hc exhibits significant superiority in both criteria. It has an accuracy rate of 93.7619 compared to 85.8, 86.1, 86.6, 85.4, 86.3, 86.7 and 92.952 for BGA, BPSO, BFFA, BHS, BCSS, BFPA-OPF, and FPA-RBF-SVM, respectively. Moreover, the FPAβ-hc has the minimum number EEG channels selected, where it achieves 32 compared to 36, 37, 45, 37, 44, 44 and 41 for BGA, BPSO, BFFA, BHS, BCSS, BFPA-OPF, and FPA-RBF-SVM, respectively. Fig. 11 shows a comparison of the accuracy rate and number of EEG channels selected using the proposed method with [1].

5.5. Discussion

The main objectives of this work are to evaluate the hybridizing version of FPA with β-hill climbing (FPAβ-hc) for EEG-based person identification, to model the problem of EEG channel selection as an evolutionary-based optimisation problem and to introduce the RBF-SVM classifier to EEG-based biometric person identification. These objectives are achieved successfully, and the results can be summarised as follows. Evidently, the proposed method has the best accuracy recognition rates using the RBF-SVM classifier with T-FDF, where FPAβ-hc achieves the highest accuracy rate of 96.0476%.

In the case of modelling EEG channel selection as an optimisation problem, the proposed method (FPAβ-hc) reduces the number of channels needed to obtain the best accuracy rate of less than half of the total number of sensors. FPAβ-hc achieves the best result of reducing the number of channels selected with FDF1 equal to 32 channels.

In the case of EEG-based person identification, the proposed method (FPAβ-hc) extracts three different groups of features, which are time domain, frequency domain, and time-frequency domain features. The extracted features can provide different accuracy rates.

Another important point relates to the electrodes selected by FPAβ-hc. A more detailed investigation shows that the most frequent electrodes are located on the front and back of the head, although they also spread along the head. This interesting observation means that FPAβ-hc attempts to identify electrodes that are not too close together to capture relevant information from all head locations.

6. Conclusions and future work

This paper proposes a hybrid method combining FPA and β-hill climbing algorithm (i.e. FPAβ-hc) to address the problem of EEG channel selection in EEG-based personal identification. The main objective of this work is to emphasise that it is not necessary to use all EEG channels available in order to obtain a high accuracy rate. This work proposes modelling the problem of channel selection as an optimisation problem, in which the subset of channels that maximise the recognition rate over a validation set is used as the fitness function. For the identification task, the RBF-SVM classifier is used, which achieves the best accuracy rate and a performance that is similar to that of other classifier techniques.

The experimental results show that the proposed FPAβ-hc method outperforms the standard FPA with RBF-SVM by obtaining excellent person identification rates using a few channels only. Notably, the number of channels is reduced to half while maintaining a high accuracy rate.

In addition, a positive correlation exists between the number of features extracted from the EEG signal and accuracy rate. A large number of extracted features leads to a high accuracy rate, as evidenced by T-FDF. This finding suggests that the proposed method can exclude duplicate and unwanted features and maintain unique features that provide the highest accuracy rate.

Furthermore, the proposed FPAβ-hc method is compared with five state-of-the-arts methods [1]by using the same dataset and feature extraction, namely, AR features with five-order coefficients called AR5. The comparison involves two criteria, which are accuracy rate and number of channels selected. The FPAβ-hc method exhibits superiority in both criteria.

With regard to the future work,

- With regard to the immediate future, the multi-objective technique is recommended to be applied using FPA to achieve the maximum accuracy rate and the minimum number of channels selected.
- the proposed method (FPAβ-hc) will be used with different feature selection approaches, such as filtering, wrapper and hybrid feature selection techniques, to perform channel selection with the aim of improving the overall identification performance while selecting fewer features of EEG signals.
- An unsupervised technique (EEG clustering) could be applied to determine the power of EEG as an identification technique.
- Another recommended direction is to perform the identification process according to several mental tasks in order to select the best task that can provide the highest accuracy rate. Subsequently, the channel selection method could be applied.
- In the future, the fMRI can be effectively fused in EEG to improve the results due to their complementary properties.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.patcog.2020.107393.

References

[1] D. Rodrigues, G.F. Silva, J.P. Papa, A.N. Marana, X.-S. Yang, EEG-Based person identification through binary flower pollination algorithm, Expert Syst. Appl. 62 (2016) 81–90.

[2] J. Unar, W.C. Seng, A. Abbasi, A review of biometric technology along with trends and prospects, Pattern Recognit. 47 (8) (2014) 2673–2688.

[3] Z.A.A. Alyasseri, A.T. Khader, M.A. Al-Betar, J.P. Papa, O. ahmad Alomari, EEG-based person authentication using multi-objective flower pollination algorithm, in: 2018 IEEE Congress on Evolutionary Computation (CEC), IEEE, 2018, pp. 1–8.

[4] Z.A.A. Alyasseri, A.T. Khader, M.A. Al-Betar, J.P. Papa, O.A. Alomari, EEG Feature extraction for person identification using wavelet decomposition and multi-objective flower pollination algorithm, IEEE Access (2018).

[5] S. Marcel, M.S. Nixon, S.Z. Li, Handbook of Biometric Anti-Spoofing, Springer, 2014.

[6] J. Galbally, S. Marcel, J. Fierrez, Biometric antispoofing methods: a survey in face recognition, IEEE Access 2 (2014) 1530–1552.

[7] T. Matsumoto, H. Matsumoto, K. Yamada, S. Hoshino, Impact of artificial gummy fingers on fingerprint systems, in: Electronic Imaging 2002, International Society for Optics and Photonics, 2002, pp. 275–289.

[8] P. Tome, M. Vanoni, S. Marcel, On the vulnerability of finger vein recognition to spoofing, in: Biometrics Special Interest Group (BIOSIG), 2014 International Conference of the, IEEE, 2014, pp. 1–10.

[9] P. Gupta, S. Behera, M. Vatsa, R. Singh, On iris spoofing using print attack, in: Pattern Recognition (ICPR), 2014 22nd International Conference on, IEEE, 2014, pp. 1681–1686.

[10] P. Campisi, D. La Rocca, Brain waves for automatic biometric-based user recognition, IEEE Trans. Inf. Forensics Secur. 9 (5) (2014) 782–800.

[11] S. Marcel, J.d.R. Millán, Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation, IEEE Trans. Pattern Anal. Mach. Intell. 29 (4) (2007) 743–752.

[12] R. Palaniappan, D.P. Mandic, Biometrics from brain electrical activity: a machine learning approach, IEEE Trans. Pattern Anal. Mach. Intell. 29 (4) (2007).

[13] A. Albasri, F. Abdali-Mohammadi, A. Fathi, Eeg electrode selection for person identification thru a genetic-algorithm method, J. Med. Syst. 43 (9) (2019) 297.

[14] M. Fraschini, S.M. Pani, L. Didaci, G.L. Marcialis, Robustness of functional connectivity metrics for eeg-based personal identification over task-induced intra–class and inter-class variations, Pattern Recognit. Lett. 125 (2019) 49–54.

[15] X.-S. Yang, Flower pollination algorithm for global optimization, in: International Conference on Unconventional Computing and Natural Computation, Springer, 2012, pp. 240–249.

[16] G. Schalk, D.J. McFarland, T. Hinterberger, N. Birbaumer, J.R. Wolpaw, BCI2000: a general-purpose brain-computer interface (BCI) system, IEEE Trans. Biomed. Eng. 51 (6) (2004) 1034–1043.

[17] A. Piryatinska, W.A. Woyczynski, M.S. Scher, K.A. Loparo, Optimal channel selection for analysis of eeg-sleep patterns of neonates, Comput. Methods Programs Biomed. 106 (1) (2012) 14–26.

[18] I.K. Al-Ani Ahmed, G. Naik, Dynamically identifying relevant eeg channels by utilizing channels classification behaviour, Expert Syst. Appl. 83 (2017) 273–282.

[19] R.S.-V. Garro, A. Beatriz, R.A. Vazquez, Eeg channel selection using fractal dimension and artificial bee colony algorithm, in: 2018 IEEE Symposium Series on Computational Intelligence (SSCI), IEEE, 2018, pp. 499–504.

[20] N.D. Sarier, Improving the accuracy and storage cost in biometric remote authentication schemes, J. Netw. Comput. Appl. 33 (3) (2010) 268–274.

[21] P.K. Sharma, A. Vaish, Individual identification based on neuro-signal using motor movement and imaginary cognitive process, Optik 127 (4) (2016) 2143–2148.

[22] E.J. Rechy-Ramirez, H. Hu, Stages for developing control systems using EMG and eeg signals: a survey, in: School of Computer Science and Electronic Engineering, University of Essex, 2011, pp. 1744–8050.

[23] A. Phinyomark, EMG Feature evaluation for improving myoelectric pattern recognition robustness, Expert Syst. Appl. 40 (12) (2013) 4832–4840.

[24] K.K. Ang, Z.Y. Chin, H. Zhang, C. Guan, Mutual information-based selection of optimal spatial–temporal patterns for single-trial EEG-based BCIs, Pattern Recognit. 45 (6) (2012) 2137–2144.

[25] B. Xue, M. Zhang, W.N. Browne, Particle swarm optimization for feature selection in classification: a multi-objective approach, IEEE Trans. Cybern. 43 (6) (2013) 1656–1671.

[26] B. Xue, M. Zhang, W.N. Browne, Particle swarm optimisation for feature selection in classification: novel initialisation and updating mechanisms, Appl. Soft Comput. 18 (2014) 261–276.
[27] Z.A.A. Alyasseri, A.T. Khader, M.A. Al-Betar, M.A. Awadallah, X.-S. Yang, Variants of the flower pollination algorithm: a review, in: Nature-Inspired Algorithms and Applied Optimization, Springer, 2018, pp. 91–118.
[28] M.A. Al-Betar, β -hill climbing: an exploratory local search, Neural Comput. Appl. 28 (1) (2017) 153–168.
[29] Z.A.A. Alyasseri, A.T. Khader, M.A. Al-Betar, M.A. Awadallah, Hybridizing β-hill climbing with wavelet transform for denoising ecg signals, Inf. Sci. 429 (2018) 229–246.
[30] S.S. Skiena, The Algorithm Design Manual: Text, vol. 1, Springer Science & Business Media, 1998.
[31] D. Rodrigues, X.-S. Yang, A.N. De Souza, J.P. Papa, Binary flower pollination algorithm and its application to feature selection, in: Recent Advances in Swarm Intelligence and Evolutionary Computation, Springer, 2015, pp. 85–100.
[32] A.L. Goldberger, L.A. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.-K. Peng, H.E. Stanley, Physiobank, physiotoolkit, and physionet, Circulation 101 (23) (2000) e215–e220.
[33] C. Ambroise, G.J. McLachlan, Selection bias in gene extraction on the basis of microarray gene-expression data, Proc. Natl. Acad. Sci. 99 (10) (2002) 6562–6566.
[34] B.S. Wade, S.H. Joshi, B.A. Gutman, P.M. Thompson, Machine learning on high dimensional shape data from subcortical brain surfaces: a comparison of feature selection and classification methods, Pattern Recognit. 63 (2017) 731–739.
[35] J. Hua, W.D. Tembe, E.R. Dougherty, Performance of feature-selection methods in the classification of high-dimension data, Pattern Recognit. 42 (3) (2009) 409–424.
[36] C. Hu, Y. Chen, L. Hu, X. Peng, A novel random forests based class incremental learning method for activity recognition, Pattern Recognit. 78 (2018) 277–290.
[37] F. Wilcoxon, Individual comparisons by ranking methods, Biom. Bull. 1 (6) (1945) 80–83.

Z.A.A. Alyasseri is a senior lecturer at the University of Kufa, Iraq. Received a BSc in computer science from Babylon University in 2007 and MSc in computer science from University Science Malaysia (USM) in 2013. Recently, in 2019 He received a Ph.D. from USM in the field of Computational Intelligence (Brain-Inspired Computing). In general, his research interests are Optimization, Pattern recognition, EEG, Brain-Computer Interface, Signal and Image Processing, Machine and Deep Learning.

A.T. Khader received the BSc. and M.Sc. in Mathematics from the University of Ohio, USA, in 1982 and 1983, respectively. He obtained his Ph.D. in Computer Science from University of Strathclyde, U.K., in 1993. He is a Professor in the School of Computer Sciences, Universiti Sains Malaysia. His research interest Mainly focuses on Optimization and Scheduling.

M.A. Al-Betar is an associate professor in the Dept of Computer Science at Al-Huson University College, Al-Balqa Applied Univ., Jourdan. He received a B.Sc and M.Sc from Computer Science Department at Yarmouk University, Jordan in 2001 and 2003 respectively. He obtained PhD from the School of Computer Sciences, USM in 2010. His research interests are mainly directed to metaheuristic optimization methods and hard combinatorial optimization problems.

O.A. Alomari received his M.Sc. in computer science from the Universiti Kebangsaan Malaysia (UKM), Malaysia in 2014. He is currently a Ph.D. Candidate in the school of Computer Sciences at Universiti Sains Malaysia. His research interests include evolutionary algorithms, nature inspired computation, and their applications to optimization problems.