Expert Systems With Applications 90 (2017) 184–195

Contents lists available at ScienceDirect

Expert Systems With Applications

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

Differential evolution algorithm as a tool for optimal feature subset selection in motor imagery EEG

Muhammad Zeeshan Baig, Nauman Aslam*, Hubert P.H. Shum, Li Zhang

Department of Computer Science and Digital Technologies, Northumbria University, Newcastle upon tyne, UK, NE18ST, United Kingdom

A R T I C L E I N F O

Article history:
Received 7 December 2016
Revised 22 June 2017
Accepted 21 July 2017
Available online 22 July 2017

Keywords:
Differential evolution
EEG
Feature selection
BCI
Motor imagery
CSP

A B S T R A C T

One of the challenges in developing a Brain Computer Interface (BCI) is dealing with the high dimensionality of the data when extracting features from EEG signals. Different feature selection algorithms have been proposed to overcome this problem but most of them involve complex transformed features, which require high computation and also result in increasing size of the feature set. In this paper, we present a new hybrid method to select features that involves a Differential Evolution (DE) optimization algorithm for searching the feature space to generate the optimal feature subset, with performance evaluated by a classifier. We provide a comprehensive study of the significance of evolutionary algorithm in selecting the best features for EEG signals. The BCI competition III, dataset IVa has been used to evaluate the method. Experimental results demonstrate that the proposed method performs well with Support Vector Machine (SVM) classifier, with an average classification accuracy of above 95% with a minimum of just 10 features. We also present a comparison of Differential Evolution (DE) with other evolutionary algorithms, and the results show the superiority of DE which implies that, with the selection of a good searching algorithm, a simple Common Spatial Pattern filter features can produce good results.

1. Introduction

Brain computer interface (BCI) is a device that permits an alternative channel of communication by sending signals directly from brain to computer. The computer analyses brain activity and converts it into decision signals. In the last two decades, due to its numerous benefits and characteristics, BCI has been progressively significant among industries and scientific institutes. BCI has been categorized mainly into two types, namely invasive BCI and non-invasive BCI (Sitaram et al., 2007). In invasive BCI, signals are extracted by placing electrodes into the brain skin (requires surgery). In non-invasive BCI, placement of electrodes is on the surface of the scalp. Due to its wide range of application, the BCI system has been used to provide assistance to paralysis, quadriplegic and amyotrophic lateral sclerosis patients to control computers and machines, without physical movement using nerves and muscles, directly by brain signals. At the same time, it is equally useful for non-disabled individuals to control a hardware system in a convenient way. A BCI system can also be applied in different areas including robotics, biomedical technologies, surgery, etc. (Coyle, Ward, & Markham, 2007).

Multiple invasive and non-invasive sources are available to record brain activities. For invasive BCIs electrocorticography (ECoG), single micro-electrode (ME), micro-electrode array (MEA) and local field potentials (LFPs) have been used. For non-invasive BCIs, electroencephalography (EEG), magnetoencephalography (MEG), Functional magnetic resonance imaging (fMRI) and Near Infrared Spectroscopy (NIRS) have been utilized. One of the most popular choices of BCI system is considered to be EEG due to the non-invasive EEG electrodes, low hardware cost and transferability. EEG signals also exhibit high temporal resolution (Hill et al., 2006). They can be acquired through different ways over different positions on the brain. The method that requires a subject to imagine a motor movement is known as Motor Imagery (MI) EEG signals. MI-based EEG signals have been applied to many BCI applications where these signals have been controlled to open an interface with the external environment (Pineda, 2005).

In the last decade, the use of BCI has increased. Many researchers utilize single trial EEG signals for developing BCI applications (Hsu, 2012; Parra et al., 2002). Some studies are specific to MI data to discriminate between left and right hand movement using event related synchronization (ERS)/desynchronization (ERD) (Hsu, 2011). However, many studies have been done on multichannel data without any feature selection. Not applying feature selection

* Corresponding authors.
E-mail addresses: baig.mzeeshan@gmail.com (M.Z. Baig), nauman.aslam@northumbria.ac.uk (N. Aslam), hubert.shum@northumbria.ac.uk (H.P.H. Shum), li.zhang@northumbria.ac.uk (L. Zhang).

http://dx.doi.org/10.1016/j.eswa.2017.07.033
0957-4174/© 2017 Published by Elsevier Ltd.

results in a reduction of the performance and practicality of the system and also increases the overall computation (Hsu, 2012).

The fuzzy logic theory is also gaining importance in decision-making applications and put challenges to the feasibility of traditional techniques for EEG signals analysis, recognition, and classification (Nauck & Kruse, 1999). Yang, Wang, and Ouyang (2014) presented a study based on adaptive neuro-fuzzy inference system (ANFIS) to classify background EEG from electrical status epilepticus slow wave sleep (ESES) syndrome patients and controls using permutation and sample entropy as features. The average accuracy of 89% was recorded. In another study, Herman, Prasad, and McGinnity (2017) utilizes interval type-2 fuzzy logic system (IT2FLS) to handle the uncertainties of non-stationary EEG signals and presented a method suitable for online BCI development. The method surpasses other state-of-the-art classification methods and achieved a classification accuracy of 71.2% plus minus 8.4 with 5-fold cross validation. Jiang et al. (2017) applied Takagi–Sugeno–Kang fuzzy system (TSK-FS) to detect epileptic EEG signals. They introduced a multi-view learning framework and combine it with TSK-FS to get a better generalization and interpret-ability. The proposed method achieved a Friedman Rank of 3.65 with TSK-FS and 1 with multiview TSK-FS. In a study by Datta, Khasnobish, Konar, and Tibarewala (2015), classification of cognitive activities by IT2FS has been proposed. Hurst Exponents, Approximate Entropy, Adaptive Autoregressive and Hjorth Parameters are used as features. The algorithm generates a classification accuracy of 85.33%.

Feature selection techniques have been used in the literature to improve classification accuracy. Some of the most common feature selection and dimensionality reduction algorithms include principal component analysis (PCA) (Yu, Chum, & Sim, 2014), independent component analysis (ICA) (Guo, Wu, Gong, & Zhang, 2013), sequential forward and backward searches (Chandrashekar & Sahin, 2014). Recently, researchers are exploring the applications of evolutionary algorithms such as particle swarm optimization (Kennedy, 2011), differential evolution (DE) (Qin, Huang, & Suganthan, 2009), artificial bee colony (ABC) optimization (Karaboga, 2005) in BCI applications.

Typical feature selection techniques have few drawbacks: the classification is poor even if the variance is good, which may be because of the redundant features that simple feature selection algorithm failed to remove. Simple features extraction techniques usually transform features linearly to reduce dimensionality without considering the classifier stage. If the linear transformation of the original features reduces the dimensionality, we still need to consider the original features for transformation. Evolutionary algorithms. on the other hand, has shown some success in a task that has large search space of features. The optimal features subset is only employed for classification not the all features and this method also optimizes the classifier performance.

It is still a hard problem to find optimal feature set using features selection algorithms. Different studies have been performed that applied evolutionary algorithms to reduce the feature set. To the best of our knowledge, this is the first work that uses a differential evolution (DE) based technique to select discriminating features in MI-based EEG (Price, Storn, & Lampinen, 2006). To evaluate the proposed method, different potential feature selection algorithms have been implemented with a set of classifiers. Our method can effectively increase the accuracy, decrease the number of features and reduce the computation complexity.

In this paper, features are extracted through common spatial pattern (CSP) filters and, to generate the optimized feature subset for each subject, the DE optimization algorithm has been used, a type of evolutionary algorithm. The objective function for DE is classification accuracy. For establishing a comparison, other evolutionary algorithms like simulated annealing (SA), artificial bee colony (ABC), ant colony optimization (ACO) and particle swarm optimization (PSO) are also implemented and the results of these algorithms have been compared. For classifying data, five different classifiers have been used to generate balanced and generalized results.

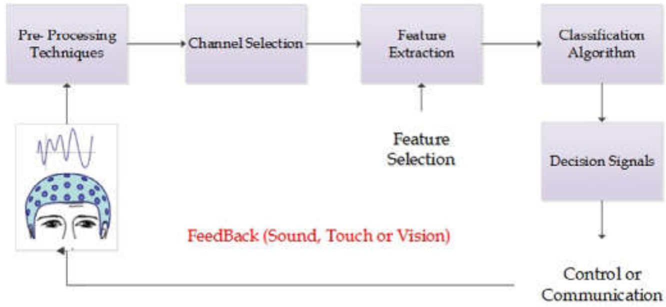

Fig. 1. Schematic diagram of brain computer interface with feedback.

The rest of the paper is organized as follows. Section 2 contains details about related methods and techniques used in BCIs along with brief details of evolutionary algorithms. Section 3 is about the results, analysis and a comparison with the results of techniques that have not used any feature selection algorithm. In Section 4, we conclude the paper and provide details about future work directions.

2. Materials and methods

A typical BCI system contains a signal acquisition part, followed by pre-processing, feature extraction and a selection algorithm; the last part is to classify the selected features and generate decision signals for control or communication as shown in Fig. 1. Different feature extracting techniques have been used in the literature to convert the raw EEG signal into useful features, e.g. common spatial pattern (CSP) (Ramoser, Muller-Gerking, & Pfurtscheller, 2000), AAR parameters (Guger, Edlinger, Harkam, Niedermayer, & Pfurtscheller, 2003), Wavelet coefficients (Baig et al., 2014; Güler & Übeyli, 2005), power spectral density (Trejo, Rosipal, & Matthews, 2006), etc. In this study, we have implemented common spatial pattern (CSP) filters for extracting features. For subset generation, a Differential Evolution based evolutionary algorithm has been used in off-line training, and different classifiers have been used to classify the features.

2.1. Data acquisition and pre-processing

The dataset used in this experiment is from BCI competition III, dataset IVa provided by the Fraunhofer FIRST, Intelligent Data Analysis Group (Dornhege, Blankertz, Curio, & Mller, 2004). The EEG signal was recorded from five healthy subjects sitting in a chair, with arms in the rest position. The data was captured without any feedback from the initial 4 sessions. The visual stimuli became active from 3.5 s. The task is to generate the motor imagery signals related to left and right hand movement or foot movement based on the clues shown. The data were acquired from 118 channels, and a total of 280 trials for each subject was recorded. A BrainAmp amplifier was used to record data from 128-channel Ag/AgCl electrodes placed using the extended international 10/20 position system. The signals were then filtered using a Band-pass filter of 0.05–200 Hz. The sampling rate of the data is 1000 Hz, which was down-sampled to 100 Hz for analysis purposes. EEG signal pre-processing is important to remove artefacts as well as to extract frequencies of interest. To get the μ and β bands of the EEG signals a fifth-order Butter worth band-pass filter of pass band 8–30 Hz was used. All 118 channels have been filtered and used for further processing.

2.2. Feature extraction

After processing the acquired EEG data, the next step is to extract features from that data. The purpose of feature extraction is to enhance the variance between classes and improve the classification. Feature extraction favorably affects the classification process, as better features directly contribute to classification accuracy. To improve classification, different feature selection algorithms are available in the literature; for extracting features from motor imagery EEG signals, the best algorithm is a CSP (Ramoser et al., 2000). In this paper, we use CSP to extract features from all 118 EEG channels. Details of the CSP algorithm are given below:

2.2.1. Common spatial pattern

Motor Imagery EEG signals are generated because motor movement, practical or imaginary, causes an increase or decrease in neural activities called event-related synchronization (ERS) or de-synchronization (ERD) (Pfurtscheller & Da Silva, 1999). To differentiate between ERS and ERD, CSP has been used widely because it has the ability to maximize the difference in variance between the two classes (Koles, Lazar, & Zhou, 1990).

Let $\overline{X_l}$ and $\overline{X_r}$ be the left and right hand MI EEG matrix with a dimension of NxM, where N is the total number of channels used and M is the number of samples per channel. The normalized spatial co-variance of left and right EEG signals can be calculated as:

$$C_l = \frac{\overline{X_l}\,\overline{X_l}^T}{trace\left(\overline{X_l}\,\overline{X_l}^T\right)} \qquad C_r = \frac{\overline{X_r}\,\overline{X_r}^T}{trace\left(\overline{X_r}\,\overline{X_r}^T\right)} \tag{1}$$

where X^T represents the transpose of X and trace is the sum of all the diagonal entries of a matrix. The average normalized co-variance matrices $\overline{C}_l$ and $\overline{C}_r$ can be obtained by averaging over all trials for each group. The combined average normalized co-variance matrix can be given as

$$C = \overline{C}_l + \overline{C}_r = V_0 \Sigma V_0^T \tag{2}$$

where V_0 is the eigenvector matrix and Σ is the corresponding diagonal matrix of eigenvalues. The whitening transformation matrix $P = \Sigma^{-\frac{1}{2}} V_0^T$ converts the average normalized covariance matrices as follows:

$$S_l = P\overline{C}_l P^T \tag{3}$$

$$S_r = P\overline{C}_r P^T \tag{4}$$

where S_l and S_r have common eigenvectors and the sum of their diagonal matrix is an identity matrix,

$$S_l = V\Sigma_l V^T \qquad S_r = V\Sigma_r V^T \qquad \Sigma_l + \Sigma_r = I \tag{5}$$

The eigenvector with the largest eigenvalue corresponds to one class and the eigenvector with smallest eigenvalue corresponds to the other class. The whitening transformation of EEG onto the eigenvectors of largest eigenvalues is optimal for separating the variance in the two signal matrices. The projection matrix can be computed as:

$$W = V^T P \tag{6}$$

where W is the projection matrix, and the original EEG can be transformed into uncorrelated components by multiplying EEG with the projection matrix W:

$$Z = WX \tag{7}$$

where Z is the EEG source component including common and specific component of different tasks. The original EEG can be reconstructed by multiplying the inverse of W with Z:

$$X = W^{-1} Z \tag{8}$$

The columns of W^{-1} are known as spatial patterns that are considered to be the EEG source distribution vector. The first column of W^{-1} shows the largest variance of one task and the last column shows the smallest variance of the other task. A time window of 3.5 s has been selected for CSP and all data samples were used. The window of 3.5 s is selected because the visual stimulus is active for 3.5 s and we have used all the samples in the active time window for CSP. As the sampling rate is 100 Hz, so a total of 350 samples were used to calculate CSP for each trial.

2.3. Feature selection

The features extracted from EEG signals are usually large in size for a single channel and increase proportionally for multiple channels. Classifying a large number of features requires more time and computation. To overcome this problem, there is a need for a feature selection algorithm to generate a subset of features that are more closely related to the mental task than other features.

The feature selection algorithms may generally be classified into three categories, Filtering, Wrapper and Hybrid techniques. In the filtering technique, an independent evaluation criterion such as distance or information measure has been used to rank features. Most of the filtering techniques are high speed, provide independence from the classifier and stability, but they are of low accuracy (Alotaiby, El-Samie, Alshebeili, & Ahmad, 2015). Wrapper Techniques use a classification algorithm to evaluate the feature subset. The Wrapper uses a predictor and its output as an objective function to evaluate the subset. The advantages of using wrapper techniques are the accuracy, as wrappers generally achieve better recognition rates than filters since they are tuned to the specific interactions between the classifier and the dataset.

Wrappers have a mechanism to avoid over-fitting, as they typically use cross-validation measures of predictive accuracy. The disadvantages are slow execution and the lack of generality: the solution lacks generality since it is tied to the bias of the classifier used in the evaluation function. The optimal feature subset will be specific to the classifier under consideration (Chandrashekar & Sahin, 2014). The hybrid feature selection method uses a combination of filtering and wrapper techniques to select features. Some of the algorithms used for EEG data are discussed in the next subsection.

2.3.1. Feature selection methods for EEG

Different feature selection algorithms have been used in literature to generate an optimal feature subset so that the overall computational complexity is reduced and the classification accuracy is improved. A genetic algorithm (GA) based feature selection algorithm along with SVM, neural network and LDA as classifiers has been suggested by Garrett, Peterson, Anderson, and Thaut (2003). The method used GA to search the feature space and generate an optimal feature subset, SVM, was used to evaluate the fitness of the generated feature subset. The results suggested that 72% of the test data had been correctly classified using SVM. Kohonen's learning vector quantization (LVQ), commonly known as distinction sensitive learning vector quantization (DSLVQ), was used to overcome the problem of feature dependency using a weighted distance function. The results demonstrated a classification accuracy of 80%. The major problem with LVQ is static preselection which cannot be updated with additional data (Pregenzer, Pfurtscheller, & Flotzinger, 1996).

A technique for feature selection of an EEG signal using an intelligent genetic search process had been presented for epileptic seizure prediction. The features used in this experiment were curve length, energy, spectral entropy, sixth power, non-linear energy, energy of wavelet packets, second and third-level features. For evaluating the subset performance, a probabilistic neural network

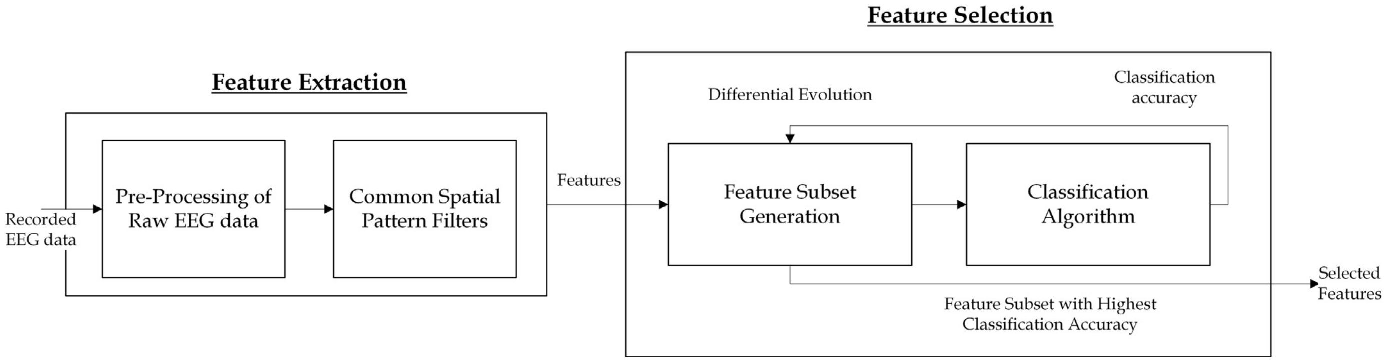

Fig. 2. System diagram for dividing the EEG feature space. The feature space is searched in a wrapper fashion directed by classifier performance.

(PNN) was used as a classifier and managed to restrict the classification error to 0.09 (D'Alessandro et al., 2003). Another wrapper method based on genetic algorithm and SVM was proposed for automated feature selection on two different datasets, recorded with the TTD system (Birbaumer et al., 1999) and neural information processing systems (NIPS) 2001 (Blankertz, Curio, & Muller, 2002). The results displayed an average improvement of 8.11% in classification accuracy for the TTD dataset and 3.15% for the NIPS dataset (Schroder, Bogdan, Hinterberger, & Birbaumer, 2003).

Gupta, Agrawal, and Kaur (2015) implemented a multivariate feature selection algorithm to increase the classification accuracy of the LDA and SVM for mental tasks using EEG signals. Atkinson and Campos (2016) proposed feature selection technique for emotion recognition from EEG signals. Their method utilizes mutual information for feature selection and SVM for classification. In another experiment, Lee, Anaraki, Ahn, and An (2015) proposed a feature selection algorithm based on fuzzy rough theory and multi-tree genetic programming to improve the classification accuracy of brain signals data measured by fNIRS. Adam et al. (2016) implemented angle modulated simulated Kalman filter to select features along with neural network as a classifier. They found the detection performance of the method is better than the other feature selection method with random weight neural network classifier. In another study, a wrapper technique based on multiobjective optimization using deep belief networks (DBN) classifier for feature selection application has been studied. The results showed that performance of DBN decreases with a large number of features and they suggest that it is necessary to determine the suitable number of hidden layer units for efficient classification.

In the literature, Genetic algorithms have been widely used for searching feature space. The effectiveness and searching power for feature selection of other evolutionary algorithms has not been studied in detail for EEG data. In this paper, we have implemented a novel hybrid method that uses CSP to extract feature space, then used Differential Evolution with a classifier (Wrapper) to discover the optimized feature subset. An architecture of the system is presented in Fig. 2. The feature selection part has two main modules, the first module is the feature subset generation using differential algorithm and the second module is the classification algorithm used to evaluate the credibility of a subset towards classification accuracy.

2.4. Evolutionary feature selection methods

Evolutionary algorithms gained a lot of interest for their applications in feature selection. Differential evolution (DE) is the main focus of this work, as well as a comparison with other evolutionary algorithms like simulated annealing (SA), particle swarm optimization (PSO), artificial bee colony (ABC) and ant colony optimization (ACO). We have implemented a modified version of DE based on a float optimizer, as no well-known binary version of DE is currently available that can yield results as sound as generated by the above mentioned nature-inspired algorithms. To give a comparative study, other evolutionary algorithms have been implemented in MATLAB and their results are compared. Details of these algorithms are given below:

2.4.1. Differential evolution

Differential evolution (DE) is a vector-based search algorithm similar to a pattern search or genetic algorithm and has a good convergence property. It is a stochastic search algorithm with an ability of self-organization without using derivative information. Like GA, DE also uses the crossover and mutation concepts as differential operators, but it has explicit updating equations (Qin et al., 2009). The process of DE is shown in Fig. 3. The algorithm starts with a randomly selected solution $\boldsymbol{x}_i$, where i=1, 2, 3, ... , n. and 3 randomly selected distinct vectors $\boldsymbol{x}_p$, $\boldsymbol{x}_q$, $\boldsymbol{x}_r$ to generate a new donor vector $\boldsymbol{v}$ by mutation

$$v_i^{t+1} = x_p^t + F\left(x_q^t - x_r^t\right) \tag{9}$$

where Fϵ[0, 2] is the differential weight and t denotes the current iteration. The crossover is controlled by Cr ϵ [0, 1]. The jth component of each v_i, update:

$$u_{j,i}^{t+1} = \begin{cases} v_{j,i} & if \quad r_i \le C_r \ or \ j = J_r, \\ x_{j,i}^t & r_i > C_r \ and \ j \ne J_r, \end{cases} \tag{10}$$

The solution is selected and updated by minimizing the objective function by the following equation:

$$x_i^{t+1} = \begin{cases} u_i^{t+1} & if \ f\left(u_i^{t+1}\right) \le f(x_i^t), \\ x_i^t & otherwise. \end{cases} \tag{11}$$

where f() is the objective function; in our case is the classification accuracy (Khushaba, Al-Ani, & Al-Jumaily, 2011). The Population size has been set to 50 and a total of 100 iterations has been used. The crossover probability Cr used in our experiment is 0.5 and the initial and final inertia weights are 0.95 and 0.35 respectively. The other population-based algorithms used in this study are discussed briefly below.

2.5. Particle swarm optimization

Particle swarm optimization (PSO) is one of the most popular evolutionary algorithms and has been applied to almost every field of research where there is a need for optimization. PSO algorithms regulate the velocities and trajectories of individual elements called particles in order to move through the space to find the objective function. The particle movement is based on a stochastic and a deterministic component. Each particle in a swarm is attracted to the position of the current global best and

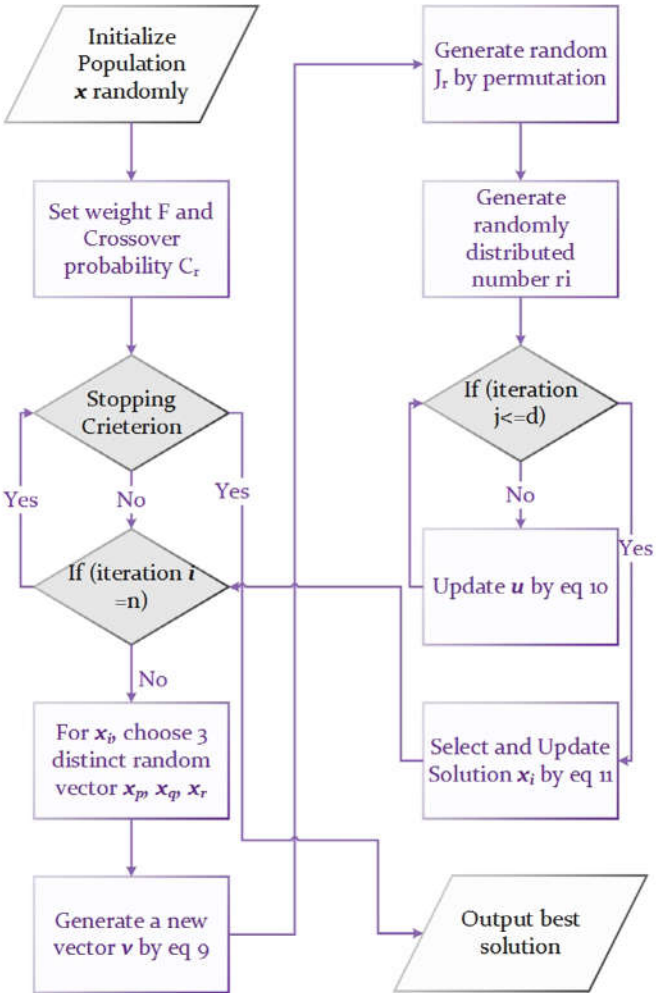

Fig. 3. Flowchart of differential evolution algorithm.

its own best location. When a particle finds a better location, it updates the current location and considers the current best for all n particles at any time. The main goal is to search for the global best among all current best solutions until the objective function does not improve further (Kennedy, 2011). The update in a particle's velocity can be calculated as

$$v_i^{t+1} = \theta v_i^t + \alpha\epsilon_1\left[g^* - s_i^t\right] + \beta\epsilon_2\left[x_i^{*(t)} - x_i^t\right], \tag{12}$$

where x_i and v_i are the position and velocity vectors for particle i, respectively. ϵ_1 and ϵ_2 are two random vectors with each entry between 0 and 1. α and β are the learning parameters, g^* is the current global best and $x_i^{*(t)}$ is the current best for particle i at time t. θ is the inertia function used to stabilize particle motion. The updated position can be calculated as

$$x_i^{t+1} = x_i^t + v_i^{t+1} \tag{13}$$

The flowchart for PSO is given in Fig. 4. In our experiment, the swarm size is 20, the inertia weight is 0.72, damping ratio is 1, individual learning coefficient and global learning coefficient have been set to 1.4962.

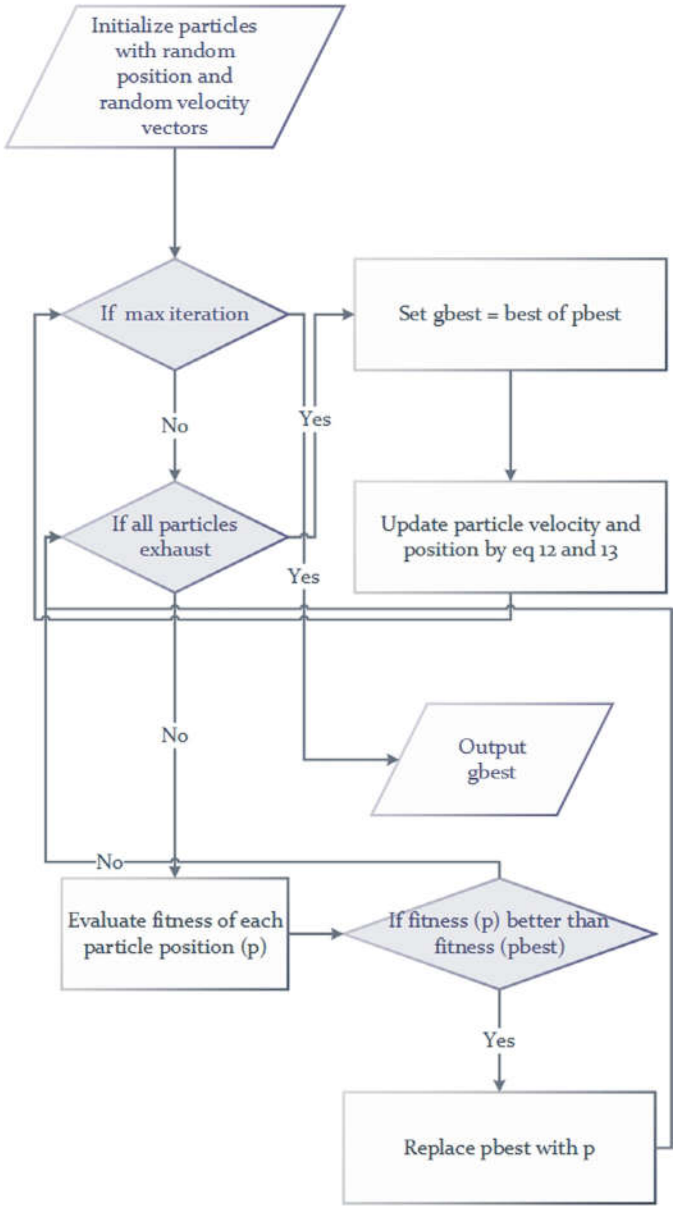

Fig. 4. Block diagram of particle swarm optimization.

2.5.1. Simulated annealing

Simulated annealing (SA) is a trajectory-based technique that uses a random search to find the global optimization, and is one of the earliest nature-inspired algorithms.

The basic theme of SA is to use random-search algorithms like a Markov chain and keep accepting those changes that improve the objective function (Szu & Hartley, 1987). The SA algorithm can be summarized as pseudo code in Algorithm 1(Yang, 2014) .

2.5.2. Ant colony optimization

Ant colony optimization (ACO) utilizes the behavior of ants to find the optimal solution. Ants live in colonies and communicate with the environment in groups or swarms. The optimization is achieved based on the way ants communicate indirectly the direction to each other while searching for food. The first algorithm was presented by Marco Dorigo in his PhD thesis (Dorigo, Birattari, & Stutzle, 2006) and after that many variants have been presented. The main point to follow in this algorithm is the probability of choosing the route and evaporation rate of a pheromone.

The probability of selection of a particular node $\boldsymbol{i}$ to choose the route from node $\boldsymbol{i}$ to $\boldsymbol{j}$ can be calculated as

$$p_{ij} = \frac{\phi_{ij}^{\alpha} d_{ij}^{\beta}}{\sum_{i,j=1}^{n} \phi_{ij}^{\alpha} d_{ij}^{\beta}} \tag{14}$$

Algorithm 1: Simulated annealing.

Data: Initial temperature T_0 and initial guess x_0
Result: Best guess x_* and best function f_*
Objective function $f(x)$Initialization of T_0 and x_0;
Set the final temperature T_f and max number of iterations N;
Define $\alpha \epsilon (0 < \alpha < 1)$;
while $T > T_f$ *and* $t < N$ **do**
  Select ϵ from Gaussian distribution;
  Randomly shift to a new location: $x_{t+1} = x_t + \epsilon$;
  Compute $\Delta f = f_{t+1}(x_{t+1}) - f_t(x_t)$;
  Select new solution if better;
  if *solution not improved* **then**
    Select a random number r;
    Select if $p = exp[-\Delta/T] > r$;
  Update best x_* and f_*;
  $t = t + 1$;

where α and β are the influence parameters and must be greater than zero, typically 2. ϕ_{ij} is the concentration of pheromone on route $\boldsymbol{i}$ to $\boldsymbol{j}$ and d_{ij} is the desirability of the route. The pheromone concentration usually varies with time as

$$\phi(t) = \phi_0 e^{-\gamma t} \tag{15}$$

where ϕ_0 is the initial concentration and γ is the constant rate of pheromone evaporation (Dorigo et al., 2008). The flowchart of a simple ACO is shown in Fig. 5. A total of 20 iterations has been used in this experiment with an ant population size of 10 and evaporation rate of 0.05. α, β and initial pheromone were set to 1.

2.5.3. Artificial bee colony

The artificial bee colony (ABC) algorithm was proposed by Karaboga in 2005 and is an optimization algorithm utilizing the foraging behavior of a honey bee swarm (Karaboga, 2005). In the ABC algorithm, the swarm or colonies are divided into three different categories of bees: employed, onlookers and scout bees. The assumption is that, for every food source there is only one artificial employed bee, i.e. the number of employed artificial bees in a colony is equal to the number of food sources around a hive. The area between hive and food source becomes the dancing area that can be explored by employed bees. A scout is an employed bee whose food source has been abandoned; they start searching for new food sources. Onlookers select the food source based on an employed bees dance. A simple explanation of the different steps involved in the ABC algorithm is given below:

1. Initialize food source for all employed bees.
2. Repeat Step 3 to Step 6.
3. Every employed bee calculates nectar of food source and dances in the hive after going to a food source and determining a neighbor source.
4. The onlooker watches the dance of employed bees to choose a source based on their dance and goes to that source and evaluates its nectar amount.
5. Scout bees determine and replace new food sources for abandoned food sources.
6. Output the best food source found so far.

In the ABC algorithm, the possible solutions are the positions of food sources, and the fitness of this solution is related to the nectar amount. The first step is to initialize the population using randomly distributed values. After initialization the search process of employed, onlooker and scout bees starts. The employed bees

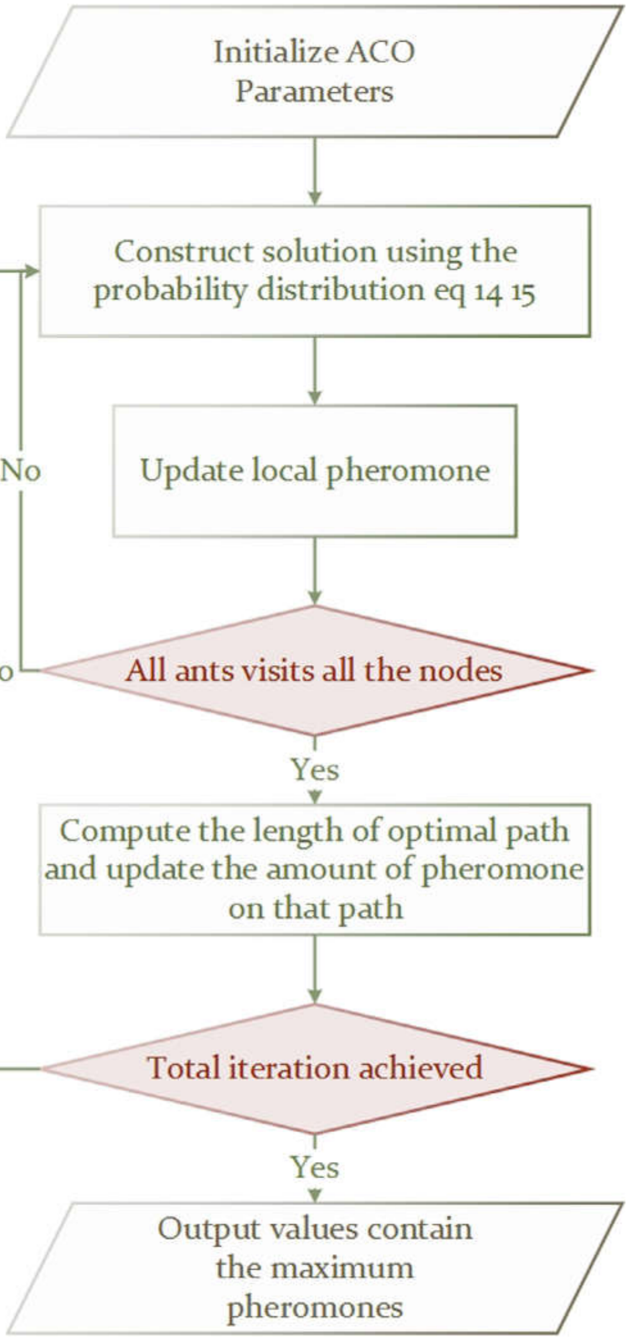

Fig. 5. Ant colony optimization algorithm.

generate a modification of a source in their memory and memorize it if the amount of nectar is greater than the current food source. After each iteration, the employed bees share the position with the onlookers on the dance area. The onlooker evaluates the amount of nectar at a new food source position and selects accordingly. The scout bees replace the abandoned sources with new sources produced randomly (Karaboga, 2010).

2.6. Classification

The selection of a classification algorithm is of vital importance in some scenarios. In this study, we utilize five different classifiers to get a generalized view of the hybrid method. The methods include linear discriminant analysis (LDA) that attempts

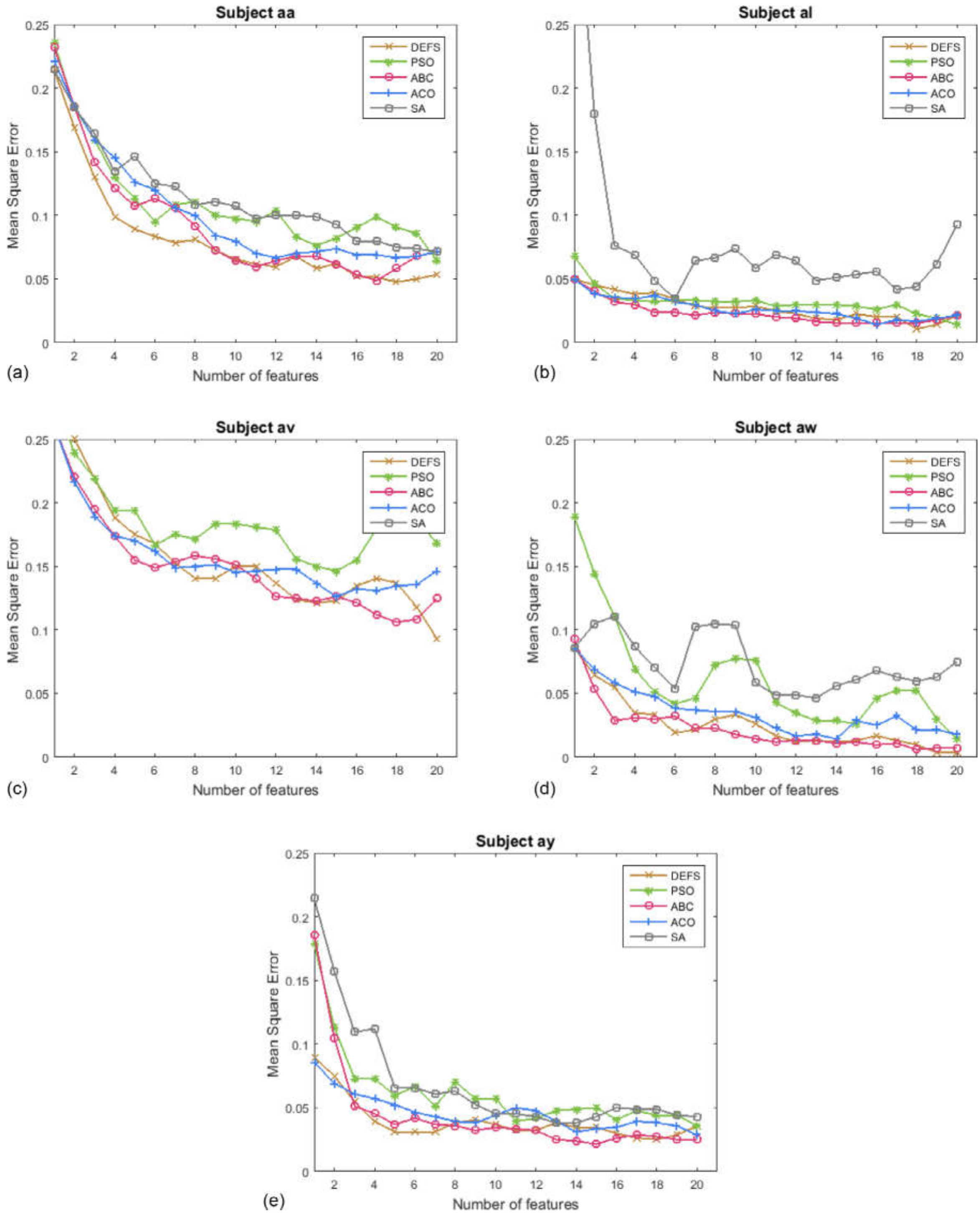

Fig. 6. Plot of mean square error against number of features selected by different evolutionary algorithms using SVM as a classifier for 5 subjects: (a) subject aa; (b) subject al; (c) subject av; (d) subject aw; (e) subject ay.

to transform one dependent variable as a linear combination of other features (Mika, Ratsch, Weston, Scholkopf, & Mullers, 1999), Support Vector Machine (SVM) which constructs a hyper-plane in a high-dimensional space to separate the classes (Hearst, Dumais, Osman, Platt, & Scholkopf, 1998), K-nearest neighbors (k-NN) which utilizes a voting-based mechanism based on neighbors with the sample being assigned to the class most appear in its k nearest neighbors (Fukunaga & Narendra, 1975), Naive Bayes (NB) classifiers, simple probabilistic classifiers, based on Bayes' theorem with an assumption of naive independence between features (Leung, 2007), Regression Tress which generates decisions using a tree-like graph or model and their possible consequences (Breiman, Friedman, Stone, & Olshen, 1984).

3. Results and discussion

To evaluate and compare the performance of the presented DE feature selection method, some well-known population-based feature selection algorithms were also implemented. Each of the algorithms was tested on the same dataset. In the experiment,

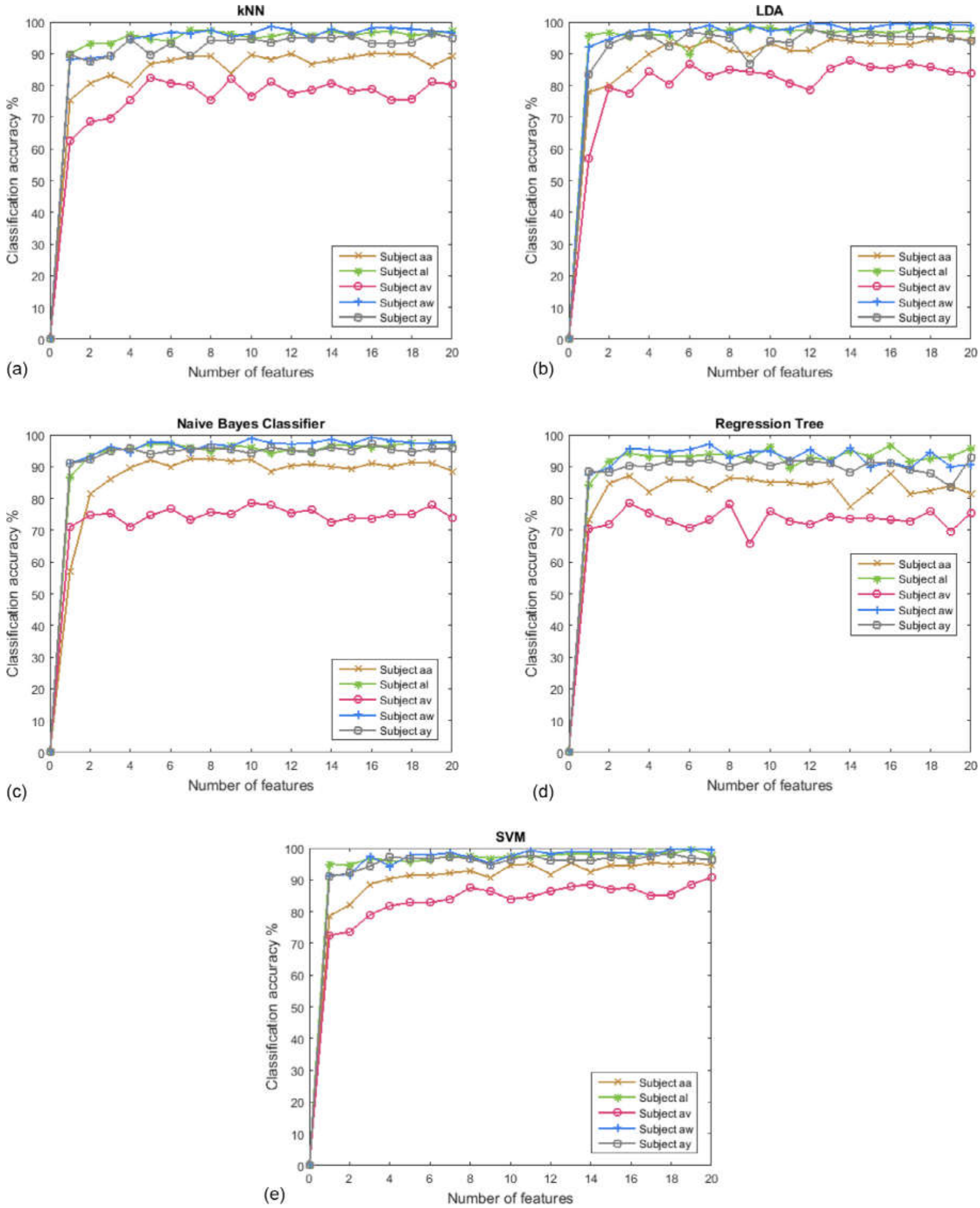

Fig. 7. Average Classification accuracy plot of 5 subjects using differential evolution algorithm for feature selection using different classifiers: (a) k-NN; (b) LDA; (c) Naive Bayes; (d) Regression tree; (e) SVM.

five different classifiers were used to evaluate the classification accuracy of DE feature selection method including LDA, SVM, k-NN, Naive Bayes and regression trees. For other feature selection algorithms, only SVM and LDA were implemented.

Data from all five subjects in BCI competition III dataset IVa have been used for training and testing purposes. The dataset contains a total of 280 samples for each subject with 118 channels. After applying CSP, 236 features were extracted from all five subjects. A 10-fold cross-validation has been used to get an average classification accuracy.

The results in Fig. 6 shows the mean square classification error of five subjects with different feature selection techniques with respect to the number of features selected, using SVM as a classifier. After a certain number of features, the mean square error (MSE) remains almost steady. For example, the subject aa maintains an average of 10% MSE with 9 features or more. Simulated annealing shows the worst classification compared to other algorithms, and differential evolution (DE) and artificial bee colony (ABC) generate maximum classification accuracy for all five subjects with DE beating ABC in some cases. The subject av shows a high mean

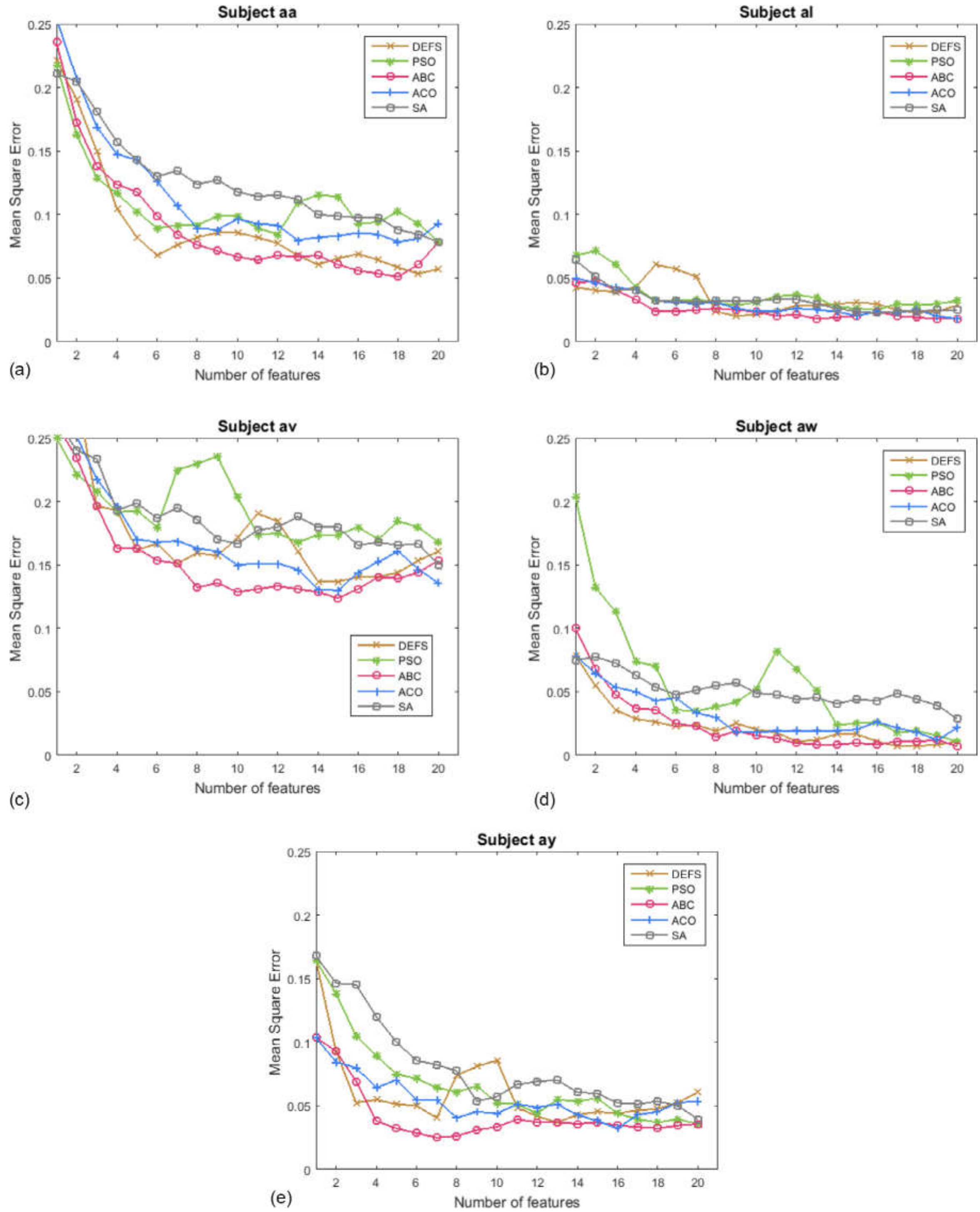

Fig. 8. Mean square error against number of features selected by different evolutionary algorithms using LDA as a classifier for 5 subjects: (a) subject aa; (b) subject al; (c) subject av; (d) subject aw; (e) subject ay.

square error compared to other subjects with an average 15 to 20%. The best MSE is of 0% with the number of features 18 and 19 for the subject aw.

Different classifiers have been used to evaluate the performance of the proposed method and the results are recorded in Fig. 7. SVM and LDA perform well compared to kNN, Naive Bayes and Regression Trees for all the five subjects, with an average classification of around 90%. The best results are obtained for feature subset sizes of 8 or more. LDA and SVM classifiers generate the maximum classification accuracy. The lowest classification rate is shown by subject av followed by subject aa and the maximum classification accuracy has been achieved by subject aw for all classifiers. For subject av, the possible reason not to generate a better classification can be due to weak brain signals or lack of interest in the experiment by the subject.

In Fig. 8, LDA has been used for classification. The graph shows that ABC results are stable compared to other classification algorithms for LDA, with a lowest mean square error for subjects al and ay. It can be seen that the second-best algorithm for feature selection is DE whose classification accuracy is approximately

Table 1
Number of features selected for 5 subjects and 5 evolutionary algorithms along with average classification accuracy of individual subjects using all features and a feature subset generated by evolutionary algorithms, reduction of features using DEFS and gain in accuracy.

Subject	aa	al	av	aw	ay	Subset selection technique
Number of features selected	13	19	20	18	18	DEFS
	14	19	14	12	6	PSO
	17	14	18	19	17	ABC
	12	19	18	13	15	ACO
	19	7	6	5	13	SA
Average accuracy using all features	82 ± 3.94%	94 ± 2.38%	70 ± 4.11%	87 ± 1.91%	87 ± 3.23%	–
Average accuracy using selected feature	95.8 ± 2.40	98.8 ± 0.79	89.8 ± 3.36	99.2 ± 0.72	96.5 ± 1.11	DEFS
	90.8 ± 2.74	96.9 ± 1.32	73.5 ± 5.22	96.7 ± 1.13	94.1 ± 2.27	PSO
	91.9 ± 1.77	97.2 ± 2.54	87.8 ± 2.91	98.9 ± 0.50	96.6 ± 1.47	ABC
	82.4 ± 3.43	91.5 ± 3.06	70.8 ± 3.52	94.3 ± 2.95	83.7 ± 2.23	ACO
	86.0 ± 2.82	94.9 ± 1.60	67.0 ± 4.21	94.8 ± 2.25	94.5 ± 1.63	SA
Features reduction ratio (DEFS)	94%	91%	91%	92%	92%	–
Increase in accuracy (DEFS)	13.8%	4.8%	19.8%	12.2%	9.5%	–

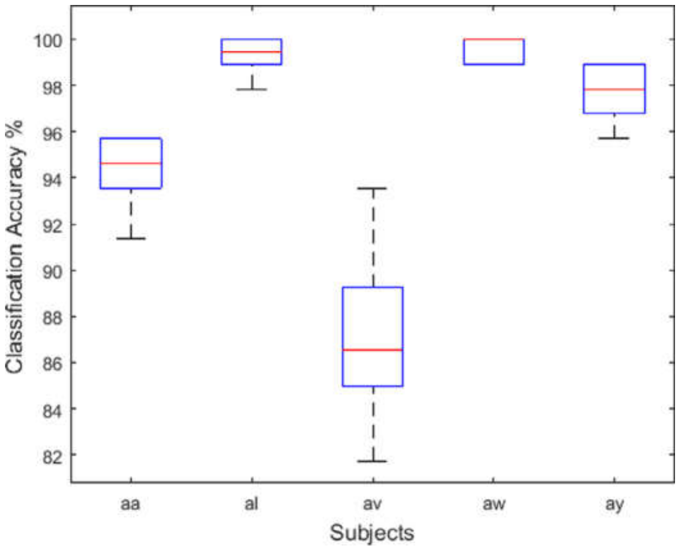

Fig. 9. Box plot of average classification accuracy for the 5 subjects using DEFS and SVM.

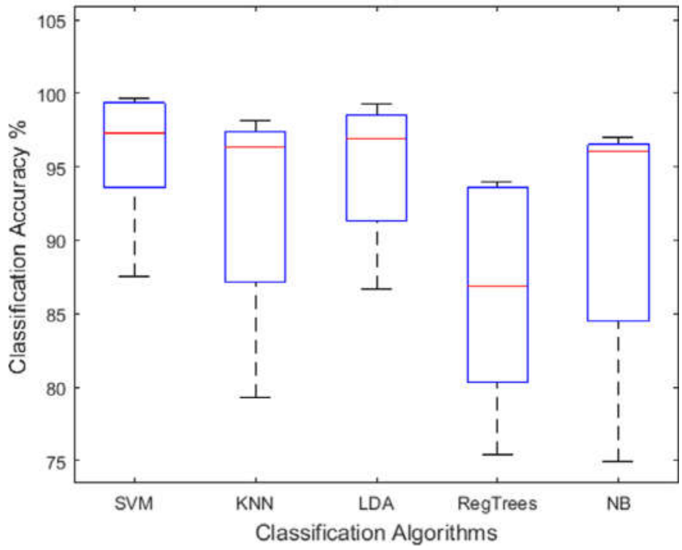

Fig. 10. Box plot of average classification accuracy by different classifiers using DEFS.

equal to that of ABC. The MSE becomes stable after the feature subset size becomes 12 or more for almost all cases.

Table 1 shows the number of features that generates the maximum classification accuracy along with the corresponding classification rate and subset generation algorithm. DE generates the best results with an average of 19 features and a classification accuracy of above 90%. The results show a reduction of 90% in the features with an improvement of 8 to 10% in classification results.

The box plot in Fig. 10 shows the average classification accuracy of all five subjects through different classifiers while using DE for feature subset generation. The results show that SVM is the best classifier for this problem with an average classification accuracy of 97%, and the variation is in between 88 and 100%. Regression tree shows a minimum accuracy of 86% with a swing between 75 and 95%. The LDA performs well with a performance measure approximately the same as SVM.

The individual subject performance has been shown in Fig. 9 with the help of a box plot. The plot shows that the subjects al and aw achieved the maximum classification accuracy of 100% for most of the cases. There is a slight reduction in accuracy for subject ay, which swings around 98%. Subject aa has the ability to generate an accuracy of 95% for most of the cross-validations, and the lowest performance has been recorded for subject av with an average of 85%.

Table 2 shows a comparison of previously published techniques for classification of motor imagery EEG signals. The references mentioned applied their proposed method on the same dataset used in our experiment. Wang et al. used CSP and Auto-regressive (AR) models for feature extraction to achieve an accuracy of 94.17%. The subjects al, aw and ay showed almost 100% of classification accuracy, but this method used a bootstrap aggregation to make the final decision which makes the method computationally expensive. CSP was applied for subjects al, aw and ay whereas CSP along with AR and LDA were used for subjects aa and av, which makes the algorithm complex and adaptive (Wang, Gao, & Gao, 2006). With the implementation of discriminant filter CSP for all channels, Thomas*, Guan, Lau, Vinod, and Ang (2009) managed to produce a classification accuracy of 91.57%. Ang, Chin, Zhang, and Guan (2008) applied filter-bank CSP for feature extraction and achieved an accuracy of 88%.

Lu et al. (2015) maximized the classification accuracy to 68.94% by applying structure constrained semi-non-negative matrix factorization. Zhang, Zhou, Jin, Wang, and Cichocki (2015) generated features using sparse filter-band CSP and got a classification accuracy of 92.05%. Das, Suresh, and Sundararajan (2016) proposed a method for subject specific filter selection using cognitive fuzzy inference system of type 2 (SS-CFIS). The method has gain an increase in performance by 15–18% compared to CSP algorithm. They have achieved a classification accuracy of 77.75% with a standard deviation of more than 16% with the SS CFIS approach. In another study, Das, Sundaram, and Sundararajan (2016) presented a robust CSP algorithm for motor imagery EEG signals. The algorithm uti-

Table 2
Comparison of feature selection method with proposed technique.

Contributor	Accuracy	Technique	aa	al	av	aw	ay
Wang et al. (2006)	94.17%	CSP and Auto-regressive models	95.5%	**100%**	80.6%	**100%**	**97.6%**
Thomas et al. (2009)	91.75%	Discriminant Filter CSP	90.21%	98.68%	77.80%	97.85%	94.23%
Ang et al. (2008)	88%	Filter bank CSP	94%	97%	86%	93%	93%
Lu et al. (2015)	68.94%	Structure constrained semi-non-negative matrix factorization	64.21%	92.67%	60.00%	72.58%	55.56%
Zhang et al. (2015)	92.05%	Sparse Filter bands CSP	91.64%	98.67%	77.43%	98.03%	94.69%
Das et al. (2016)	77.75%	Subject-specific CFIS	82.14%	100%	63.27%	83.04%	60.32%
Proposed method	**96.02%**	CSP with DE for feature selection	**95.8%**	98.8%	**89.8%**	99.2%	96.5%

Table 3
Summary of channel selection method for motor imagery EEG evaluated on BCI competition III dataset IVa.

Techniques	Channel selection strategy	Classifier	Performance metrics (Average)	No of channels selected/Total no. of channels (Average)
Yong, Ward, and Birch (2008)	Maximizing variance of CSP	Linear discriminant analysis (LDA)	73.5%	13/118
Meng, Liu, Huang, and Zhu (2009)	Heuristic algorithm based on *L1* norm	SVM with Gaussian RBF	89.68%	20/118
Arvaneh, Guan, Ang, and Quek (2011)	Recursive feature elimination using Sparse CSP	SVM	82.28% (SCSP1)	22.6/118
			79.28% (SCSP2)	7.6/118
Shenoy and Vinod (2014)	Minimum redundancy maximum relevancy (mRMR)	SVM	90.77%	10/118
Das and Suresh (2015)	Cohen's d effect size	SVM	85.85%	9.20/118

lizes a self regulated interval type 2 neuro fuzzy inference system in for handling EEG signals. The mean classification accuracy has increased with these proposed techniques, also the standard deviation which is more than 10% in each case. In comparison, the proposed method utilizes the DE algorithm to generate an optimal features subset and the subset is then made available to the classifier to achieve maximum results. The features have been extracted by applying CSP to EEG data from all the channels. The results showed that subjects al and aw have a maximum accuracy of almost 100%, almost the same as other algorithms. Subjects aa and ay displayed an accuracy of above 95% and most importantly the classification accuracy for subject av is averaged around 90% which is better than for all other algorithms.

Table 3 shows a summary of channel selection algorithm for motor imagery EEG signals. The classification accuracy after applying channel selection 90% (maximum). With the proposed feature selection method, the classification accuracy is of 96% with a variance of less than 3%.

Subject variability is observed to have a notable affect the performance as seen in Table 2. We observe that the classification accuracy of subject aa, al, aw and ay are much higher than av. This trend (effect) can also be seen in feature selection using evolutionary algorithm. Because we choose wrapper techniques, the effect has been minimized (as seen from the results). The difference in patterns of different subjects cannot be seen from the naked eye.

4. Conclusion

The benefit of applying a feature selection algorithm is that it reduces computational complexity and increases feature set effectiveness by selecting relevant features. Once a feature subset is selected, it requires a classifier to generate control signals. This paper presents a new hybrid method for feature selection of an EEG signal for the MI-related task. The proposed DE-based EEG feature selection successfully generates the optimal feature subset based on the maximum classification accuracy of a subset feature. The method uses a simple feature extraction technique i.e. CSP and produces promising results. The dataset used to evaluate the proposed method is dataset IVa from BCI competition III and the method successfully enhances the classification accuracy of the dataset. The results showed that the average of 96.02% classification accuracy was achieved with an increase of 2% in classification accuracy compared to other methods in the literature. SVM and LDA prove to be the best classifiers for DE-based feature selection; both classifiers showed an accuracy of 95% with a deviation of 0.1. Subjects al and av produced an average accuracy of almost 100%, aa and ay showed a mean classification accuracy of 95% and av showed an average of 88%. Preliminary results of the presented method show a simple feature extraction technique can produce remarkable results. The proposed algorithm is slow compared to the typical feature selection algorithms and the classifier of the wrapper technique make it even more slower. However, we argue that it has significant benefit that over-weigh the limitation of slow operation. A notable benefit of using the evolutionary algorithm is that it is applied only once to select the best features related to an application. Once the set of optimal features is obtained, then it is simply a classification problem which can be processed in much efficient manner. The results show the effectiveness of the presented feature selection method, and future work includes more extensive testing of evolutionary algorithms and its application to an on-line BCI system.

Acknowledgment

The project is supported in part by European Unions Erasmus Mundus Grant cLINK project (Grant Number 2012-2645) and the Engineering and Physical Sciences Research Council (EPSRC) (Ref: EP/M002632/1).

References

Adam, A., Ibrahim, Z., Mokhtar, N., Shapiai, M. I., Mubin, M., & Saad, I. (2016). Feature selection using angle modulated simulated Kalman filter for peak classification of eeg signals. *SpringerPlus, 5*(1), 1580.
Alotaiby, T., El-Samie, F. E. A., Alshebeili, S. A., & Ahmad, I. (2015). A review of channel selection algorithms for EEG signal processing. *EURASIP Journal on Advances in Signal Processing, 2015*(1), 66. doi:10.1186/s13634-015-0251-9.
Ang, K. K., Chin, Z. Y., Zhang, H., & Guan, C. (2008). Filter bank common spatial pattern (FBCSP) in brain-computer interface. In *2008 ieee international joint conference on neural networks (IEEE world congress on computational intelligence)* (pp. 2390–2397). IEEE.
Arvaneh, M., Guan, C., Ang, K. K., & Quek, C. (2011). Optimizing the channel selection and classification accuracy in EEG-based BCI. *IEEE Transactions on Biomedical Engineering, 58*(6), 1865–1873.
Atkinson, J., & Campos, D. (2016). Improving BCI-based emotion recognition by combining EEG feature selection and kernel classifiers. *Expert Systems with Applications, 47*, 35–41.
Baig, M. Z., Javed, E., Ayaz, Y., Afzal, W., Gillani, S. O., & Naveed, M. (2014). Classification of left/right hand movement from EEG signal by intelligent algorithms.

In *Computer applications and industrial electronics (ISCAIE), 2014 IEEE symposium on* (pp. 163–168). IEEE.
Birbaumer, N., Ghanayim, N., Hinterberger, T., Iversen, I., Kotchoubey, B., & Kübler, A. (1999). A spelling device for the paralysed. *Nature, 398*(6725), 297–298.
Blankertz, B., Curio, G., & Muller, K.-R. (2002). Classifying single trial EEG: Towards brain computer interfacing. *Advances in Neural Information Processing Systems, 1*, 157–164.
Breiman, L., Friedman, J., Stone, C. J., & Olshen, R. A. (1984). *Classification and regression trees*. CRC press.
Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. *Computers & Electrical Engineering, 40*(1), 16–28.
Coyle, S. M., Ward, T. E., & Markham, C. M. (2007). Brain-computer interface using a simplified functional near-infrared spectroscopy system. *Journal of Neural Engineering, 4*(3), 219.
D'Alessandro, M., Esteller, R., Vachtsevanos, G., Hinson, A., Echauz, J., & Litt, B. (2003). Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: A report of four patients. *IEEE Transactions on Biomedical Engineering, 50*(5), 603–615.
Das, A., & Suresh, S. (2015). An effect-size based channel selection algorithm for mental task classification in brain computer interface. In *Systems, man, and cybernetics (SMC), 2015 IEEE international conference on* (pp. 3140–3145). IEEE.
Das, A., Suresh, S., & Sundararajan, N. (2016). A discriminative subject-specific spatio-spectral filter selection approach for EEG based motor-imagery task classification. *Expert Systems with Applications, 64*, 375–384.
Das, A. K., Sundaram, S., & Sundararajan, N. (2016). A self-regulated interval type-2 neuro-fuzzy inference system for handling nonstationarities in EEG signals for BCI. *IEEE Transactions on Fuzzy Systems, 24*(6), 1565–1577.
Datta, S., Khasnobish, A., Konar, A., & Tibarewala, D. (2015). Cognitive activity classification from EEG signals with an interval type-2 fuzzy system. In *Advancements of medical electronics* (pp. 235–247). Springer.
Dorigo, M., Birattari, M., Blum, C., Clerc, M., Sttzle, T., & Winfield, A. (2008). *Ant colony optimization and swarm intelligence: 6th international conference, ANTS 2008, Brussels, Belgium, september 22–24, 2008, proceedings*: Vol. 5217. Springer.
Dorigo, M., Birattari, M., & Stutzle, T. (2006). Ant colony optimization. *IEEE Computational Intelligence Magazine, 1*(4), 28–39.
Dornhege, G., Blankertz, B., Curio, G., & Mller, K.-R. (2004). Boosting bit rates in noninvasive eeg single-trial classifications by feature combination and multi-class paradigms. *Biomedical Engineering, IEEE Transactions on, 51*(6), 993–1002.
Fukunaga, K., & Narendra, P. M. (1975). A branch and bound algorithm for computing k-nearest neighbors. *IEEE Transactions on Computers, 100*(7), 750–753.
Garrett, D., Peterson, D. A., Anderson, C. W., & Thaut, M. H. (2003). Comparison of linear, nonlinear, and feature selection methods for EEG signal classification. *Neural Systems and Rehabilitation Engineering, IEEE Transactions on, 11*(2), 141–144.
Guger, C., Edlinger, G., Harkam, W., Niedermayer, I., & Pfurtscheller, G. (2003). How many people are able to operate an EEG-based brain-computer interface (BCI)? *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 11*(2), 145–147.
Güler, I., & Übeyli, E. D. (2005). Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients. *Journal of Neuroscience Methods, 148*(2), 113–121.
Guo, X., Wu, X., Gong, X., & Zhang, L. (2013). Envelope detection based on online ICA algorithm and its application to motor imagery classification. In *Neural engineering (NER), 2013 6th international IEEE/EMBS conference on* (pp. 1058–1061). IEEE.
Gupta, A., Agrawal, R., & Kaur, B. (2015). Performance enhancement of mental task classification using EEG signal: A study of multivariate feature selection methods. *Soft Computing, 19*(10), 2799–2812.
Hearst, M. A., Dumais, S. T., Osman, E., Platt, J., & Scholkopf, B. (1998). Support vector machines. *IEEE Intelligent Systems and Their Applications, 13*(4), 18–28.
Herman, P. A., Prasad, G., & McGinnity, T. M. (2017). Designing an interval type-2 fuzzy logic system for handling uncertainty effects in brain–computer interface classification of motor imagery induced eeg patterns. *IEEE Transactions on Fuzzy Systems, 25*(1), 29–42.
Hill, N. J., Lal, T. N., Schrder, M., Hinterberger, T., Widman, G., & Elger, C. E. (2006). Classifying event-related desynchronization in EEG, ECOG and MEG signals. In *Pattern recognition* (pp. 404–413). Springer.
Hsu, W.-Y. (2011). Eeg-based motor imagery classification using enhanced active segment selection and adaptive classifier. *Computers in Biology and Medicine, 41*(8), 633–639.
Hsu, W.-Y. (2012). Application of competitive hopfield neural network to brain-computer interface systems. *International Journal of Neural Systems, 22*(01), 51–62.
Hsu, W.-Y. (2012). Enhanced active segment selection for single-trial EEG classification. *Clinical EEG and Neuroscience, 43*(2), 87–96.
Jiang, Y., Deng, Z., Chung, F.-L., Wang, G., Qian, P., & Choi, K.-S. (2017). Recognition of epileptic EEG signals using a novel multiview tsk fuzzy system. *IEEE Transactions on Fuzzy Systems, 25*(1), 3–20.
Karaboga, D. (2005). An idea based on honey bee swarm for numerical optimization. *Technical Report*. Technical report-tr06, Erciyes University, Engineering faculty, Computer Engineering department.
Karaboga, D. (2010). Artificial bee colony algorithm. *Scholarpedia, 5*(3), 6915.
Kennedy, J. (2011). Particle swarm optimization. In *Encyclopedia of machine learning* (pp. 760–766). Springer.
Khushaba, R. N., Al-Ani, A., & Al-Jumaily, A. (2011). Feature subset selection using differential evolution and a statistical repair mechanism. *Expert Systems with Applications, 38*(9), 11515–11526. doi:10.1016/j.eswa.2011.03.028.
Koles, Z. J., Lazar, M. S., & Zhou, S. Z. (1990). Spatial patterns underlying population differences in the background EEG. *Brain Topography, 2*(4), 275–284. doi:10.1007/bf01129656.
Lee, J.-H., Anaraki, J. R., Ahn, C. W., & An, J. (2015). Efficient classification system based on fuzzy–rough feature selection and multitree genetic programming for intension pattern recognition using brain signal. *Expert Systems with Applications, 42*(3), 1644–1651.
Leung, K. M. (2007). Naive bayesian classifier. *Polytechnic University Department of Computer Science/Finance and Risk Engineering*.
Lu, N., Li, T., Pan, J., Ren, X., Feng, Z., & Miao, H. (2015). Structure constrained semi-nonnegative matrix factorization for EEG-based motor imagery classification. *Computers in Biology and Medicine, 60*, 32–39. doi:10.1016/j.compbiomed.2015.02.010.
Meng, J., Liu, G., Huang, G., & Zhu, X. (2009). Automated selecting subset of channels based on CSP in motor imagery brain-computer interface system. In *Robotics and biomimetics (ROBIO), 2009 IEEE international conference on* (pp. 2290–2294). IEEE.
Mika, S., Ratsch, G., Weston, J., Scholkopf, B., & Mullers, K. R. (1999). Fisher discriminant analysis with kernels. In *Neural networks for signal processing IX: Proceedings of the 1999 IEEE signal processing society workshop (cat. no.98th8468)* (pp. 41–48). doi:10.1109/NNSP.1999.788121.
Nauck, D., & Kruse, R. (1999). Obtaining interpretable fuzzy classification rules from medical data. *Artificial Intelligence in Medicine, 16*(2), 149–169.
Parra, L., Alvino, C., Tang, A., Pearlmutter, B., Yeung, N., & Osman, A. (2002). Linear spatial integration for single-trial detection in encephalography. *Neuroimage, 17*(1), 223–230.
Pfurtscheller, G., & Da Silva, F. H. L. (1999). Event-related EEG/MEG synchronization and desynchronization: Basic principles. *Clinical Neurophysiology, 110*(11), 1842–1857.
Pineda, J. A. (2005). The functional significance of mu rhythms: Translating seeing and hearing into doing. *Brain Research Reviews, 50*(1), 57–68.
Pregenzer, M., Pfurtscheller, G., & Flotzinger, D. (1996). Automated feature selection with a distinction sensitive learning vector quantizer. *Neurocomputing, 11*(1), 19–29.
Price, K., Storn, R. M., & Lampinen, J. A. (2006). *Differential evolution: A practical approach to global optimization*. Springer Science & Business Media.
Qin, A. K., Huang, V. L., & Suganthan, P. N. (2009). Differential evolution algorithm with strategy adaptation for global numerical optimization. *IEEE Transactions on Evolutionary Computation, 13*(2), 398–417.
Ramoser, H., Muller-Gerking, J., & Pfurtscheller, G. (2000). Optimal spatial filtering of single trial EEG during imagined hand movement. *IEEE Transactions on Rehabilitation Engineering, 8*(4), 441–446.
Schroder, M., Bogdan, M., Hinterberger, T., & Birbaumer, N. (2003). Automated EEG feature selection for brain computer interfaces. In *Neural engineering, 2003. Conference proceedings. First international IEEE EMBS conference on* (pp. 626–629). IEEE.
Shenoy, H. V., & Vinod, A. P. (2014). An iterative optimization technique for robust channel selection in motor imagery based brain computer interface. In *Systems, man and cybernetics (SMC), 2014 IEEE international conference on* (pp. 1858–1863). IEEE.
Sitaram, R., Caria, A., Veit, R., Gaber, T., Rota, G., & Kuebler, A. (2007). Fmri brain–computer interface: A tool for neuroscientific research and treatment. *Computational Intelligence and Neuroscience, 2007*.
Szu, H., & Hartley, R. (1987). Fast simulated annealing. *Physics Letters A, 122*(3–4), 157–162.
Thomas, K. P., Guan, C., Lau, C. T., Vinod, A. P., & Ang, K. K. (2009). A new discriminative common spatial pattern method for motor imagery brain& x2013;computer interfaces. *IEEE Transactions on Biomedical Engineering, 56*(11), 2730–2733. doi:10.1109/TBME.2009.2026181.
Trejo, L. J., Rosipal, R., & Matthews, B. (2006). Brain-computer interfaces for 1-d and 2-d cursor control: Designs using volitional control of the EEG spectrum or steady-state visual evoked potentials. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 14*(2), 225–229.
Wang, Y., Gao, S., & Gao, X. (2006). Common spatial pattern method for channel selelction in motor imagery based brain-computer interface. In *Engineering in medicine and biology society, 2005. IEEE-EMBS 2005. 27th Annual international conference of the* (pp. 5392–5395). IEEE.
Yang, X.-S. (2014). *Nature-inspired optimization algorithms*. Elsevier.
Yang, Z., Wang, Y., & Ouyang, G. (2014). Adaptive neuro-fuzzy inference system for classification of background EEG signals from ESES patients and controls. *The Scientific World Journal, 2014*.
Yong, X., Ward, R. K., & Birch, G. E. (2008). Sparse spatial filter optimization for EEG channel reduction in brain-computer interface. In *Acoustics, speech and signal processing, 2008. ICASSP 2008. IEEE international conference on* (pp. 417–420). IEEE.
Yu, X., Chum, P., & Sim, K.-B. (2014). Analysis the effect of PCA for feature reduction in non-stationary EEG based motor imagery of BCI system. *Optik-International Journal for Light and Electron Optics, 125*(3), 1498–1502.
Zhang, Y., Zhou, G., Jin, J., Wang, X., & Cichocki, A. (2015). Optimizing spatial patterns with sparse filter bands for motor-imagery based braincomputer interface. *Journal of Neuroscience Methods, 255*, 85–91. doi:10.1016/j.jneumeth.2015.08.004.