

Review

Evolutionary computation algorithms for feature selection of EEG-based emotion recognition using mobile sensors

Bahareh Nakisa^{a,*}, Mohammad Naim Rastgoo^a, Dian Tjondronegoro^b, Vinod Chandran^a^a Science and Engineering Faculty, Queensland University of Technology, Brisbane, Qld, Australia^b School of Business and Tourism, Southern Cross University, Gold Coast, Qld, Australia

ARTICLE INFO

Article history:

Received 12 June 2017

Revised 8 September 2017

Accepted 29 September 2017

Available online 2 October 2017

Keywords:

Emotion classification

EEG signals

Feature selection

Evolutionary computation algorithms

ABSTRACT

There is currently no standard or widely accepted subset of features to effectively classify different emotions based on electroencephalogram (EEG) signals. While combining all possible EEG features may improve the classification performance, it can lead to high dimensionality and worse performance due to redundancy and inefficiency. To solve the high-dimensionality problem, this paper proposes a new framework to automatically search for the optimal subset of EEG features using evolutionary computation (EC) algorithms. The proposed framework has been extensively evaluated using two public datasets (MAHNOB, DEAP) and a new dataset acquired with a mobile EEG sensor. The results confirm that EC algorithms can effectively support feature selection to identify the best EEG features and the best channels to maximize performance over a four-quadrant emotion classification problem. These findings are significant for informing future development of EEG-based emotion classification because low-cost mobile EEG sensors with fewer electrodes are becoming popular for many new applications.

Crown Copyright © 2017 Published by Elsevier Ltd. All rights reserved.

1. Introduction

Recent advances in emotion recognition using physiological signals, particularly electroencephalogram (EEG) signals, have opened up a new era of human computer interaction (HCI) applications, such as intelligent tutoring (Calvo & D' Mello, 2010; Du Boulay, 2011), computer games (Mandryk & Atkins, 2007) and e-Health applications (Liu, Conn, Sarkar, & Stone, 2008; Luneski, Bamidis, & Hitoglou-Antoniadou, 2008). Automatic emotion recognition using sensor technologies such as wireless headbands and smart watches is increasingly the subject of research, with the development of new forms of human-centric and human-driven interaction with digital media. Many of these portable sensors are easy to set up and connect via Bluetooth to a smart phone or computer, where the data can be readily analyzed. These mobile sensors are able to support real-world applications, such as detecting driver drowsiness (Li, Lee, & Chung, 2015), and, more recently, assessing the cognitive load of office workers in a controlled environment (Zhang et al., 2017).

Emotion recognition supports automatic interpretation of human intentions and preferences, allowing HCI applications to better

respond to users' requirements and customize interactions based on affective responses. The strong correlation between different emotional states and EEG signals is most likely because these signals come directly from the central nervous system, providing information (features) about internal emotional states. EEG signals can thus be expected to provide more valuable than less direct or external indicators of emotion such as interpreting facial expressions.

Previous works on extraction of EEG features have demonstrated that there are many useful features from time, frequency and time–frequency domains, which have been shown to be effective for recognizing different emotions. A recent study (Jenke, Peer, & Buss, 2014) proposed the most comprehensive set of extractable features from EEG signals, noting that advanced features like higher order crossing can perform better than common features like power spectral bands for classifying basic emotions. However, their experiment did not include a publicly available dataset and cannot be directly compared with our proposed method. Moreover, there is no standardized set of features that have been generally agreed as the most suitable for emotion recognition. This leads to what is known as a high-dimensionality issue in EEG, as not all of these features would carry significant information regarding emotions. Irrelevant and redundant features increase the feature space, making patterns detection more difficult, and increasing the risk of overfitting. It is therefore important to identify salient features that have significant impact on the performance of the emotion

* Corresponding author.

E-mail addresses: bahareh.nakisa@qut.edu.au, bahareh.nakisa@hdr.qut.edu.au (B. Nakisa), mohammadnaim.rastgoo@qut.edu.au (M.N. Rastgoo), dian.tjondronegoro@scu.edu.au (D. Tjondronegoro), v.chandran@qut.edu.au (V. Chandran).

classification model. Feature selection methods have been shown to be effective in automatically decreasing high dimensionality by removing redundant and irrelevant features and maximizing the performance of classifiers.

Among the many methods which can be applied to feature selection problems, the simplest are filter methods which are based on ranking techniques. Filter methods select the features by scoring and ordering features based on their relevance, and then defining a threshold to filter out the irrelevant features. The methods aim to filter out the less relevant and noisy features from the feature list to improve classification performance. Filter methods that have been applied to emotion classification systems include Pearson Correlation (Kroupi, Yazdani, & Ebrahimi, 2011), correlation based feature reduction (Nie, Wang, Shi, & Lu, 2011; Schaaff & Schultz, 2009), and canonical correlation analysis (CCA). However, filtering methods have two potential disadvantages as a result of assuming that all features are independent of each other (Zhang & Zhao, 2008). The first disadvantage is the risk of discarding features that are irrelevant when considered individually, but that may become relevant when combined with other features. The second disadvantage is the potential for selecting individual relevant features that may lead to redundancies.

Evolutionary computation (EC) algorithms can help to overcome the limitations of individual feature selection by assessing the subset of variables based on their usefulness. The main advantage of using EC to solve optimization problems is the ability to search simultaneously within a set of possible solutions to find the optimal solution, by iteratively trying to improve the feature subset with regard to a given measure of quality. Five well-known EC algorithms – Ant Colony Optimization (ACO), Simulated Annealing (SA), Genetic Algorithm (GA), Particle Swarm Optimization (PSO) and Differential Evolution (DE) – are widely used for feature selection in various applications, including facial expression-based emotion recognition (Mistry, Zhang, Neoh, Lim, & Fielding, 2016) and classification of motor imagery EEG signals (Baig, Aslam, Shum, & Zhang, 2017). Baig's study, particularly, has achieved a very high (95%) accuracy using a DE algorithm, but was based on only 5 subjects.

Compared to previous work, this is the first study to identify the best EC-based feature selection method(s), which have not been previously tested on EEG-based emotion recognition. The proposed method is evaluated using two public datasets (DEAP and MAHNOB) and a newly collected dataset using wireless EEG sensors to give a comprehensive review on experimental results in different contexts. In all three datasets, video clips and music were used as stimuli to induce different emotions. In addition, this study investigates the most optimal subset of features within each dataset and identifies the most frequent set of selected channels using the principle of weighted majority voting. These findings are significant for informing future development of EEG-based emotion classification because low-cost mobile EEG sensors with fewer electrodes are becoming popular for many new applications.

This paper is organized as follows. Section 2 discuss the framework and adjustment of algorithms for the feature-selection problem. Section 3 presents the methodology, including the data collection. Section 4 discusses the experimental results, while Section 5 provides the conclusion and discusses future work.

2. System framework

A typical emotion classification system using EEG signals consists of four main tasks: *pre-processing*, *feature extraction*, *feature selection* and *classification* (see Fig. 1).

The first and most critical step is pre-processing, as EEG signals are typically noisy as a result of contamination by physiological artefacts caused by electrode movement, eye movement, mus-

cle activities, heartbeat and so on. The artefacts that are generated from eye movement, heartbeat, head movement and respiration are below the frequency of 4 Hz, while the artefacts caused by muscle movement are higher than 40 Hz. In addition, there are some non-physiological artefacts caused by power lines with frequencies of 50 Hz, which contaminate the EEG signal.

In order to remove artefacts while keeping the EEG signals within specific frequency bands, sixth-order (band-pass) *Butterworth* filtering was applied to obtain 4–64 Hz EEG signals to cover different emotion-related frequency bands. Notch filtering was applied to remove 50 Hz noise caused by power lines. In addition to these pre-processing methods, independent component analysis (ICA) was used to reduce the artefacts caused by heartbeat and to separate complex multichannel data into independent components (Jung et al., 2000), and provide a purer signal for feature extraction.

The purer EEG signals were then passed through a feature extraction step, in which several types of features from time, frequency and time–frequency domains were extracted to distinguish different emotions. Subsequently, all the extracted features from each channel were concatenated into a single vector representing a large feature set. To reduce the number of features used for the machine learning process, EC algorithms are applied iteratively to the different feature sets to find the optimal and most effective set. The classification and feature selection steps were integrated to iteratively evaluate the quality of the feature sets produced by the feature selection against the classification of specific emotions based on experimental results. To evaluate the performance of each EC feature selection algorithm, a probabilistic neural network (PNN) (Specht, 1990) was adopted, as it has been shown to be effective for emotion recognition using different modalities. A PNN is a feedforward network with three layers which is derived from Bayesian networks. In our framework, training and testing of each EC algorithm was conducted using 10-fold cross validation, which helps to avoid overfitting. This process is made possible thanks to PNN's faster training process compared to other classification methods, as the training is achieved using one pass of each training vector rather than several passes.

Of these tasks, feature extraction and the integrated feature selection and classification methods represent the most important parts of the framework. After reviewing and evaluating these methods, the key contribution of this paper is to find the optimal strategy for feature selection of high-dimensional EEG-based emotion recognition.

2.1. Feature extraction

EEG features are generally categorized into three main domains: *time-*, *frequency-* and *time–frequency*.

2.1.1. Time-domain features

Time-domain features have been shown to correlate with different emotional states. Statistical features – such as mean, maximum, minimum, power, standard deviation, 1st difference, normalized 1st difference, standard deviation of 1st difference, 2nd difference, standard deviation of 2nd difference, normalized 2nd difference, quartile 1, median, quartile 3, quartile 4 – are good at classifying different basic emotions such as joy, fear, sadness and so on (Chai, Woo, Rizon, & Tan, 2010; Takahashi, 2004). Other promising time-domain feature is Hjorth parameters: Activity, Mobility and Complexity (Ansari-Asl, Chanel, & Pun, 2007; Horlings, Datcu, & Rothkrantz, 2008). These parameters represent the mean power, mean frequency and the number of standard slopes from the signals, which have been used in EEG-based studies on sleep disorder and motor imagery (Oh, Lee, & Kim, 2014; Redmond & Heneghan, 2006; Rodriguez-Bermudez, Garcia-Laencina, & Roca-Dorda, 2013).

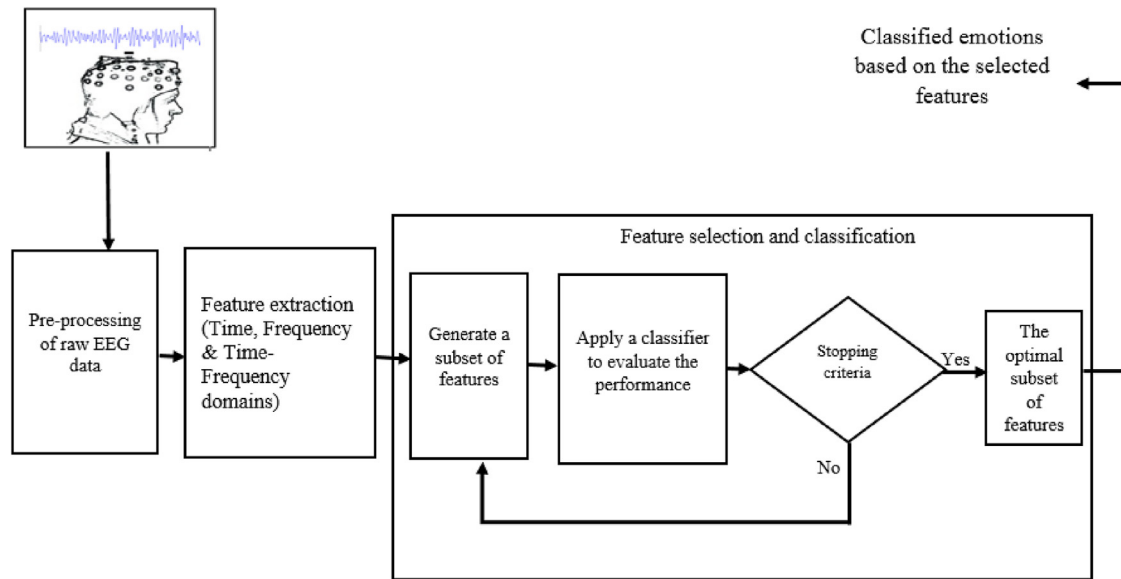


Fig. 1. The proposed emotion classification system using evolutionary computational (EC) algorithms for feature selection.

All abovementioned features were applied for real-time applications, as they have the least complexity compared with other methods (Khan, Ahamed, Rahman, & Smith, 2011).

In addition to these well-known features, we incorporated two newer time-domain features. The first one is fractal dimension to extract geometric complexity, which has been shown to be effective for detecting concentration levels of subjects (Aftanas, Lotova, Koshkarov, & Popov, 1998; Sourina & Liu, 2011; Sourina, Kulish, & Sourin, 2009; Sourina, Sourin, & Kulish, 2009; Wang, Sourina, & Nguyen, 2010). Among the several methods for computing fractal-dimension features, the Higuchi method has been shown to outperform other methods, such as box-counting and fractal brownian motion (Liu & Sourina, 2012). The second newer feature is Non-Stationary Index (NSI) (Kroupi et al., 2011), which segments EEG signals into smaller parts and estimates the variation of their local averages to capture the degree of the signals' non-stationarity. The performance of NSI features can be further improved by combining them with other features, such as higher order crossing features (Petrantonakis & Hadjileontiadis, 2010) that are based on the zero-crossing count to characterize the oscillation behavior.

2.1.2. Frequency-domain features

Compared to time-domain features, *frequency-domain* features have been shown to be more effective for automatic EEG-based emotion recognition. The power of the EEG signal among different frequency bands is a good indicator of different emotional states. Features such as power spectrum, logarithm of power spectrum, maximum, minimum and standard deviation should be extracted from different frequency bands, namely Gamma (30–64 Hz), Theta (13–30 Hz), Alpha (8–13 Hz) and Beta (4–8 Hz), as these features have been shown to change during different emotional states (Barry, Clarke, Johnstone, Magee, & Rushby, 2007; Davidson, 2003; Koelstra et al., 2012; Onton & Makeig, 2009).

2.1.3. Time-frequency domain features

The limitation of frequency-domain features is the lack of temporal descriptions. Therefore, *time-frequency-domain* features are suitable for capturing the non-stationary and time-varying signals, which can provide additional information to characterize different emotional states. The most recent and promising features are discrete wavelet transform (DWT) and Hilbert Huang spectrum (HHS). DWT decomposes the signal into different

frequency bands while concentrating in time, and it has been used to recognize different emotions using different modalities, such as speech (Shah et al., 2010), electromyography (Cheng & Liu, 2008) and EEG (Murugappan et al., 2008). HHS extracts amplitude, squared amplitude and instantaneous amplitude from decomposed signals obtained from intrinsic mode functions, and has been applied to investigate the connection between music preference and emotional arousal (Hadjidimitriou & Hadjileontiadis, 2012). From the Discrete Wavelet Transform, different features can be further extracted to distinguish basic emotions (Murugappan, Ramachandran, & Szali, 2010; Murugappan, Rizon, Nagarajan, & Yaacob, 2010), including power, recursive energy efficiency (REE), root mean square, and logarithmic REE.

This study adopts all of the abovementioned features to automatically select the most important subset of EEG features that can achieve the optimal classification performance. The features are summarized in Table 1.

2.2. Feature selection

Evolutionary computation (EC) algorithms can be used to select the most relevant subset of features from extracted EEG features. Our study used the five population-based heuristic search algorithms useful for global searching, mentioned above, namely: Ant Colony Optimization, Simulated Annealing, Genetic Algorithm, Particle Swarm Optimization and Differential Evolution. Although each of these algorithms is different, the common aim among them is to find the optimum solution by iteratively evaluating new solutions. All EC algorithms follow three steps: 1) initialization, where the population of solutions is initialized randomly; 2) evaluation of each solution in the population for fitness value; 3) iteratively generating a new population until the termination criteria are met. The termination criteria could be the maximum number of iterations or finding the optimal set of features that maximize classification accuracy.

2.2.1. Ant colony optimization (ACO)

First proposed by Dorigo and Gambardella (1997), ACO is inspired from the foraging behavior of ant species. It is based on the finding the shortest paths from food sources to the nest. In an ant colony system, ants leave a chemical, pheromone, on the

Table 1
Extracted features from EEG signal.

<i>Time domain</i>	
Minimum	Mean
Maximum	Power
Standard deviation	1st difference
Normalized 1st difference	2nd difference
Normalized second difference	Hjorth feature (Activity, mobility and complexity)
Non-stationary index	Fractal dimension (Higuchi algorithm)
Higher order crossing	Variance
Root mean square	Quartile 1, quartile median, quartile 3, quartile4
<i>Frequency domain</i>	
Power spectrum density (PSD) from Gamma, Theta, Alpha, Beta	Mean
<i>Time–frequency domain</i>	
Power of Discrete Wavelet Transform (DWT) from Gamma, Theta, Alpha, Beta	
Root Mean Square (RMS) of DWT from Gamma, Theta, Alpha, Beta	Recursive Energy Efficiency (REE) of DWT from Gamma, Theta, Alpha, Beta
Log (REE) of DWT from Gamma, Theta, Alpha, Beta	Abs (log (REE)) of DWT from Gamma, Theta, Alpha, Beta

ground while they are searching for food (Dorigo, Birattari, & Stutzle, 2006). Once an ant finds a food source, it evaluates the quantity and quality of the food and, during its return trip to the colony, leaves a quantity of pheromone based on its evaluation of that food. This pheromone trail guides the other ants to the food source. Ants detect pheromone by smell and choose the path with the strongest pheromone. This process is performed iteratively until the ants reach the food source.

ACO has been widely applied in many domains such as job scheduling (Blum & Sampels, 2002; Colorni, Dorigo, Maniezzo, & Trubian, 1994), sequential ordering (Dorigo & Gambardella, 1997), graph coloring (Costa & Hertz, 1997), shortest common super sequences (Michel & Middendorf, 1998) and connectionless network routing (Sim & Sun, 2003). Studies have shown the utility of the ACO algorithm for the feature-selection problem (Al-Ani, 2005; Sivagaminathan & Ramakrishnan, 2007).

To apply the ACO algorithm to the feature-selection problem, we need to include a path for feature selection algorithms. The path can be represented as a graph, where each node in the graph represents a feature and the edge shows the next feature to be selected. Based on this path, the ants are generated with a random set of features. From their initial positions, they start to construct the solution (set of features) using heuristic desirability, which denotes the probability of selecting feature i by ant r at time step t :

$$P_i^r(t) = \begin{cases} \frac{\tau(i)^\alpha \cdot n(i)^\beta}{\sum_{u \in J(r)} \tau(u)^\alpha \cdot n(u)^\beta} & \text{if } i \in J(r) \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Where, for the ant r , $n(i)$ and $\tau(i)$ are the heuristic information and the pheromone value of feature i , and α and β are the parameters which determine the importance of pheromone value and heuristic information respectively. The $n(i)$ and $\tau(i)$ parameters can create a balance between exploration and exploitations, influenced by α and β values. If $\alpha = 0$, then no pheromone information is used and the previous search is overlooked and if $\beta = 0$, then the exploration or global search is overlooked. After constructing the solutions (a set of features) for each ant, the fitness function is applied on each solution to evaluate performance. In this study, a PNN classifier was employed to evaluate the accuracy of each solution, which represents the set of features. Then the pheromone evaporation was applied as follows:

$$\tau(t+1) = (1 - \rho) * \tau(t) \quad (2)$$

Where $\rho \in (0, 1)$ is the pheromone decay coefficient. Finally, the process stops when the termination criteria are met – either the optimum set of features with highest accuracy or the maximum number of iterations is achieved.

2.2.2. Simulated annealing (SA)

First proposed by Kirkpatrick, Gelatt, and Vecchi (1983), SA is inspired from metallurgy processes. It is based on selecting the best sequence of temperatures to achieve the best result. The algorithm starts from the initial state of high temperature, then iteratively creates a new random solution and opens a search space widely to slowly decrease the temperature to a frozen ground state. SA is used for different domains such as job shop scheduling (Suresh & Mohanasundaram, 2006), clustering (Bandyopadhyay, Saha, Maulik, & Deb, 2008) and robot path planning (Zhu, Yan, & Xing, 2006).

For feature selection, SA iteratively generates new solutions from the neighborhood and then the fitness function of the new generated solution is calculated and compared with the current solutions. A neighbor of a solution is generated by selecting a random bit and inverting it; if the fitness function of the new solution is better than the current solution, then the new solution will be replaced for the next iteration. Otherwise, it will be accepted based on the Metropolis condition which states that, if the difference between the fitness function of the current solution and the new solution is equal or higher than zero, then a random number δ will be generated between [0,1]. And then if the Boltzmann's function (Eq. 4) value is higher than δ , the new generated solution will be accepted for the next iteration.

$$\exp(\Delta E/T) \geq \delta \quad (4)$$

After all iterations at each temperature are complete, the next temperature state is selected based on a temperature updating rule (Eq. 5)). This process continues iteratively until the termination criteria are reached, namely a fixed number of iterations or until no further improvement is observed.

$$T_n = \alpha^N T_i \quad (5)$$

where:

- T_d is the new decreasing temperature state
- α is the cooling ratio
- N is the number of iteration in each temperature state
- T_i is the initial temperature state

2.2.3. Genetic algorithm (GA)

First proposed by Goldberg and Holland (1988), GA is inspired by natural selection. The algorithm aims to find the (near) optimal solution for chromosomes to continue surviving, based on stochastic optimization. It has been applied to find the optimal solutions for job scheduling (Gonçalves, de Magalhães Mendes, & Resende, 2005; Yang & Honavar, 1998) problems.

The algorithm contains a set of chromosomes, which are represented in binary form, with operators for fitness function, breed-

ing or crossover, and mutation. Each of these binary chromosomes is represented as a solution and these solutions are used to generate new solutions. Initially, the chromosomes are created randomly to represent different points in the search space. The fitness of each chromosome is evaluated and the chromosomes with better fitness value are more likely to be kept for the next generation (as a parent). New chromosomes are then generated using a pair of the fittest current solutions through the combination of successive chromosomes and some crossover and mutation operators. The crossover operator replaces a segment of a parent chromosome to generate a new chromosome, while the mutation operator mutates a parent chromosome into a newly generated chromosome to make a very small change to the individual genome. This mutation process helps in introducing randomness into the population and maintaining diversity within it. Otherwise, the combination of the current population can cause the algorithm to become trapped in the local optima, unable to explore the other search space. Finally, the newly generated chromosomes are used for the next iterations. This process continues until some satisfactory criteria are met.

To apply the GA algorithm to the feature-selection problem, each chromosome is represented by a binary vector of dimension m , where m is the total number of features. If a bit is 1, then the corresponding feature is included, and if a bit is 0, the feature is not included. The process of GA for feature selection problem is the same as GA. The process terminates when it finds the subset of features with highest accuracy or reaches the maximum number of iterations.

2.2.4. Particle swarm optimization (PSO)

First proposed by Eberhart and Kennedy (1995), PSO is inspired by the social behaviors of bird flocking and fish schooling. The algorithm is a population-based search technique similar to ACO (see 2.4.1). PSO was originally applied to continuous problems and then extended to discrete problems (Kennedy & Eberhart, 1997). Due to its simplicity and effectiveness, this algorithm is used in different domains such as robotics (Couceiro, Rocha, & Ferreira, 2011; Nakisa, Nazri, Rastgoo, & Abdullah, 2014; Nakisa, Rastgoo, Nasrudin, & Nazri, 2014; Rastgoo, Nakisa, & Nazri, 2015) and job scheduling (Sha & Hsu, 2006; Zhang, Shao, Li, & Gao, 2009).

The algorithm is similar to GA, as it consists of a set of particles, which resemble the chromosome in GA, and a fitness function. Each particle in the population has a position in the search space, a velocity vector and a corresponding fitness value which is evaluated by the fitness function. However, unlike GA, PSO does not require sorting of fitness values of any solution in any process, which may be a computational advantage, particularly when the population size is large.

To apply PSO to feature-selection problems, the first step is initialization. At each iteration, the population of particles spread out in the search space with random position and velocity. The fitness value of each particle is evaluated using the fitness function. The particles iterate from one position to another position in the search space using the velocity vector. This velocity vector (Eq. (6)) is calculated using the particle's personal best position (P_{best}), the global best (g_{best}) and the previous velocity vector. The particle's personal best value is the best position that the particle has visited so far and the global best (g_{best}) is the best visited position by any particle in the population. These two values can be controlled by some learning factors. The next particle's position will be evaluated through the previous position and the calculated velocity vector (as described in Eqs. (6) and (7)).

$$v_i^{t+1} = \omega v_i^t + c_1 r_1 (P_i^t - x_i^t) + c_2 r_2 (G^t - x_i^t) \quad (6)$$

$$x_i^{t+1} = x_i^t + v_i^{t+1} \quad (7)$$

where v_i^t and x_i^t are the previous iteration's velocity vector and the previous particle's position respectively, ω is the inertia weight, c_1, c_2 are learning factors and r_1, r_2 are random numbers that are uniformly distributed between [0, 1].

2.2.5. Differential evolution (DE)

Another stochastic optimization method is Differential Evolution (DE), which has recently attracted increased attention for its application to continuous search problems (Price, Storn, & Lampinen, 2006). Although its process is similar to the PSO algorithm, for unknown reasons it is much slower than PSO. Recently, its strength has been shown in different applications such as strategy adaptation (Qin, Huang, & Suganthan, 2009) and job shop scheduling (Pan, Wang, & Qian, 2009). Most recently this algorithm has shown promising performance as a feature-selection algorithm for EEG signals in motor imagery applications (Baig et al., 2017).

DE algorithm represents a solution by a D -dimensional vector. A population size of N with a D -dimensional vector is generated randomly. Then a new solution is generated by combining several solutions with the candidate solution, and these solutions are evolved using three main operators: mutation, crossover and selection. Although the concept of solution generation is applied in the DE algorithm in the same way as it is applied in GA, the operators are not all the same as those with the same names in GA.

The key process in DE is the generation of a trial vector. Consider a candidate or a target vector in a population of size N of D -dimensional vectors. The generation of a trial vector is accomplished by the mutation and crossover operations and can be summarized as follows. 1) Create a mutant vector by mutation of three randomly selected vectors. 2) Create a trial vector by the crossover of the mutant vector and the target vector. When the trial vector is formed, the selection operation is performed to keep one of the two vectors, that is, either the target vector or the trial vector. The vector with better fitness value is kept, and is the one included for the selection of the next mutant vector. This is an important difference since any improvement may affect other solutions without having to wait for the whole population to complete the update.

3. Experimental method

We conducted extensive experiments to determine if evolutionary computation algorithms can be used as effective feature selection processes to improve the performance of EEG-based emotion classification and to find the most successful subset of features. Based on the experiments, we determined which features are generally better, across all stimuli and subjects.

The experiments were based on the goal of classifying four-class emotions based on EEG signals, using the most successful subset of features generated from five different EC algorithms (ACO, SA, GA, PSO and DE). The experiments used two public datasets (MAHNOB and DEAP), which contain EEG signals with 32 channels. In addition, a new dataset of EEG signals with only 5 channels was used for comparison purposes. In order to decrease the computation time, the experiments used data from 15 randomly-selected subjects. We simplified the dimensional (arousal-valence based) emotions into four quadrants: 1) High Arousal-Positive emotions (HA-P); 2) Low Arousal-Positive emotions (LA-P); 3) High Arousal-Negative emotions (HA-N); 4) Low Arousal-Negative emotions (LA-N).

As described in Section 2.1, some noise reduction techniques, including Butterworth, notch filtering and ICA, were applied to the EEG signals to remove artefacts and noise. After noise reduction, a variety of EEG features from time, frequency and time-frequency domains were extracted from a one-second window with 50% overlap (45 features in total from each window). For the experiments using the MAHNOB and DEAP datasets, we extracted features from

A black Emotiv EPOC+ EEG headset with the brand name 'EMOTIV' visible on the side.

Fig. 2. The Emotiv insight headset.

- For the ACO algorithm: number of ants=20, evaporation rate=0.05, initial pheromone and heuristic value = 1.
- For the SA algorithm: initial temperature=10, cooling ratio = 0.99, maximum number of iteration in each temperature state = 20.
- For the GA algorithm: crossover percentage=0.7, mutation percentage = 0.3, mutation rate = 0.1, selection pressure = 8.
- For the PSO algorithm: construction coefficient = 2.05, damping ratio = 0.9, particle size = 20. For the DE algorithm: population size = 20, crossover probability = 0.2, lower bound of scaling factor = 0.2, upper bound of scaling factor = 0.8.

Fig. 3. The location of five channels used in Emotiv sensor (represented by the black dots), while the white dots are the other channels normally used in other wired sensors).

experimental results, we will determine if the use of wireless sensor can become a viable option for future studies that require subjects to move around freely.

We used TestBach software for acquiring raw EEG data from the *Emotiv* headset while a participant was watching videos. Emotions were induced by video clips, used in MAHNOB dataset, and the participants' brain responses were collected while they were watching 9 video clips in succession. The participants were asked to report their emotional state after watching each video, using a keyword such as neutral, anxiety, amusement, sadness, joy or happiness, disgust, anger, surprise, and fear. Before the first video clip, the participants were asked to relax and close their eyes for one minute to allow their baseline EEG to be determined. Between each video clip stimulus, one minute's silence was given to prevent mixing up the previous emotion. The experimental protocol is shown in Fig. 4.

To ensure data quality, we manually analyzed the signal quality for each subject. Some EEG signals from the 5 channels were either lost or found to be too noisy due to the long study duration, which may have been caused by loose contact or shifting electrodes. As a result, only signal data from 11 (5 female and 6 male) out of 13 participants is included in this dataset. Despite this setback, the experiment with this new data allows an investigation into the feasibility of using the Emotiv Insight sensor for emotion classification purposes. The expected benefit of this sensor is due to its lightweight, and wireless nature, making it possibly the most suitable for free-living studies in natural settings.

3.1.3. New experiment dataset (video)

The third dataset was newly collected from 13 subjects, aged between 20 and 38 years old, while they watched video clips. To collecting the EEG data, the *Emotiv Insight* wireless headset was used (see Fig. 2). This Emotiv headset contains 5 channels (AF3, AF4, T7, T8, Pz) and 2 reference channels located and labeled according to the international 10–20 system (see Fig. 3). Compared to the EEG recording devices that were used in MAHNOB and DEAP, our data collection used Emotiv wireless sensor, which only captures EEG signals from five channels (instead of 32). Based on the

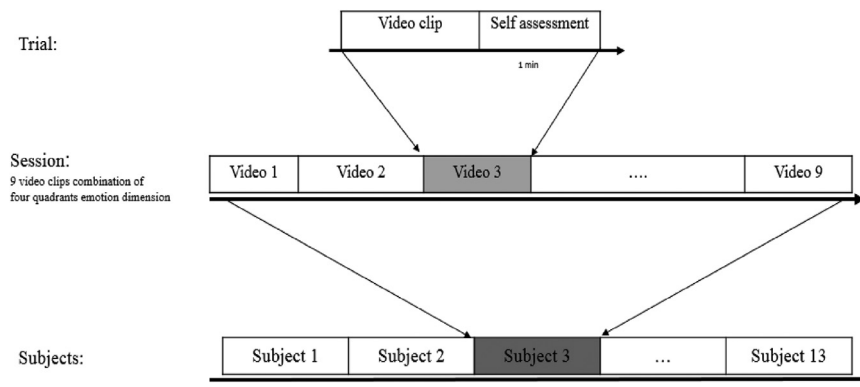


Fig. 4. Illustration of the experimental protocol for emotion elicitations.

Table 2

The average accuracy of EC algorithms over three datasets (MAHNOB, DEAP, our dataset).

FS method	No. iteration	(MAHNOB)		(DEAP)		(Our dataset)		Average time across datasets
		Time (h)	Accuracy (%)	Time (h)	Accuracy (%)	Time (h)	Accuracy (%)	
PSO	15	12.5	82.94263 ± 2.865	12.9	51.18436 ± 4.23412	11.7	59.16438 ± 4.07415	82.8 h
	25	20.9	85.5012 ± 2.5847	21.5	51.64421 ± 3.06233	19.5	60.84009 ± 5.77746	
	45	37.5	95.30953 ± 2.0627	38.7	54.07454 ± 2.36788	35.4	60.41786 ± 2.81327	
	100	83.9	96.58661 ± 1.83694	86.1	65.31437 ± 3.22760	78.4	61.48601 ± 6.59215	
ACO	15	16	89.63667 ± 1.54467	16.2	46.09645 ± 1.22374	15.3	53.58217 ± 1.99010	106.2 h
	25	26.6	89.96537 ± 1.45023	27.2	56.7229 ± 1.87711	25.5	54.4585 ± 1.57451	
	45	48	90.7626 ± 1.32020	48.8	58.23569 ± 1.54648	45.9	55.58698 ± 1.27219	
	100	106.6	91.97158 ± 1.14934	110	59.25536 ± 1.39521	102	55.58698 ± 5.94841	
GA	15	11.9	84.85624 ± 2.21359	12.3	50.12445 ± 2.98451	11.8	57.62701 ± 2.89420	77.6 h
	25	19.7	86.26149 ± 1.97594	20.6	55.91216 ± 2.63143	19.6	59.59106 ± 2.67428	
	45	35.6	94.99567 ± 1.93795	37.1	57.98351 ± 2.42171	35.4	57.58815 ± 3.60488	
	100	79	97.11983 ± 1.41752	82.5	63.63564 ± 3.17107	71.3	61.237212 ± 2.16085	
SA	15	10.9	83.11863 ± 3.11444	11.2	44.65506 ± 2.73391	10	49.04851 ± 1.94359	71.4 h
	25	18.1	84.54847 ± 5.37894	18.7	48.6937 ± 3.62394	16.6	51.30788 ± 2.31224	
	45	32.7	86.86808 ± 3.89385	33.7	52.20574 ± 5.76376	30	51.14426 ± 3.10572	
	100	72.7	89.01175 ± 2.31283	75	55.25314 ± 2.64210	66.6	52.04453 ± 2.95477	
DE	15	11.6	83.01443 ± 1.9758	11.9	60.92632 ± 2.92837	10.8	56.14765 ± 2.73955	76.4 h
	25	19.4	89.67716 ± 2.00478	19.8	63.59828 ± 2.17161	18	59.11505 ± 3.03241	
	45	35.1	92.99567 ± 1.85698	35.7	64.59933 ± 2.85288	32.4	63.04303 ± 2.19556	
	100	77.8	96.97023 ± 1.89385	79.3	67.47447 ± 3.38984	72.1	65.043028 ± 3.19556	

4. Experimental results and discussion

4.1. Benchmarking feature selection methods

The performance of EC algorithms was assessed based on the optimum accuracy that can be achieved within reasonable time frames for convergence. Each algorithm was tested based on its ability to achieve the best subset of features within a limited time. As mentioned in Section 3, the total number of features generated from 32 channels was 1440. In order to find the minimum subset of features to maximize the classification performance, we empirically limited the selected number of features to 30 out of 1440. Based on our findings, this number of features not only maximizes the performance of the proposed model, but can also keep the computational cost sufficiently low. In addition to feature size, the computational complexity of ECs algorithms is dependent on the number of iterations required for convergence, which may need to be obtained by a trial-and-error process. Despite EC algorithms are computationally expensive, they are less complex than full search and sequential algorithms. Moreover, the computational complexity of the proposed feature selection method is less important, as long as it is possible to achieve an acceptable result and complete the step in a reasonable time. This is due to the fact that the process of proposing the most optimized feature selection method is only conducted during development and training stages.

Table 2 presents the relative performance of each EC algorithm, providing average accuracy ± standard error and processing time

for 30 selected features over 10 runs on three datasets (MAHNOB, DEAP and our dataset). Processing time was determined by Intel Core i7 CPU, 16 GB RAM, running windows 7 on 64-bit architecture. Based on the average processing time across three datasets, ACO takes the most amount of processing time (106.2h), while not delivering the highest accuracy (92%, 60% and 60% for MAHNOB, DEAP and our dataset respectively). In contrast, DE gives the highest accuracy (96%, 67% and 65% for MAHNOB, DEAP and our dataset respectively), while the processing time (76.4h on average) is lower than others and only higher than that of SA (but SA gives the lowest accuracy). Based on the average accuracy across all datasets (representing the overall performance), the best result is achieved when DE is applied, followed by PSO and GA. From 25 to 100 iterations, the mean accuracy rate of PSO and GA improved by 11% over the MAHNOB dataset and 14% and 8% over the DEAP dataset respectively. However, the improvement rate for DE was lower (7% for MAHNOB and 4% for DEAP). Similarly, the improvement rate for ACO and SA was only 2% and 5% over MAHNOB respectively. This phenomenon is most likely due to PSO and GA's diversity property (i.e. its capability in searching for solutions more widely), while ACO and SA are more likely to be trapped in local optima from the early iterations and seem to converge faster than the other algorithms. This is one of the common problems among EC algorithms, known as the *premature convergence* problem.

Further analysis of the diversity property of all algorithms is provided in Fig. 5, showing that the proposed system reached an acceptable result based on the peak performance within 100 it-

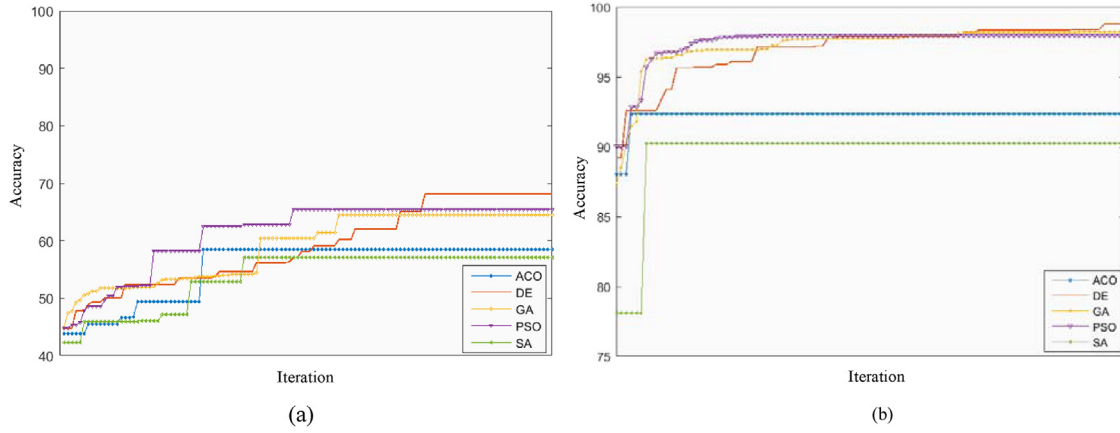


Fig. 5. Performance of different algorithms for DEAP (a) and MAHNOB (b) databases.

erations within an acceptable processing time period. This graph depicts the challenging part of EC algorithms, when these algorithms become trapped in local optima and fail to explore the other promising subsets of features. Generally, these algorithms converge to local optima/ global optima after some iterations and their performance remains steady without any improvement. In this study, it is shown that as the number of iterations increased, the performance of DE, PSO and GA increased and they found better solutions with higher accuracy, while ACO and SA converged in the early iterations and failed to improve their performance. DE had the best convergence speed and founds better solutions. All DE, PSO and GA algorithms came very close to the global optima in the MAHNOB dataset in all runs, but the other two algorithms (SA and ACO) usually failed to explore and stagnated into the local optima, due to less diversity in the nature of these algorithms.

In terms of comparing results across the different datasets, the results from MAHNOB are significantly better than DEAP. This can be explained by the fact that video is a more effective stimulus to induce different emotions, compared to music. This hypothesis is further justified by the results from our new dataset, which was obtained using the same video stimuli as MAHNOB. Using only 5 EEG channels (instead of 32), the overall performance appears to be very similar to that of the DEAP dataset (slightly lower), which confirms the feasibility of using the wireless and light-weight EEG sensor (albeit of lower accuracy).

4.2. Frequently-selected features

To investigate the general quality features for emotion recognition based on two shared datasets (MAHNOB and DEAP), i.e. which feature selected by 5 EC feature-selection methods are repeated more frequently than the others and performed more successfully in emotion classification, we computed the occurrence number of each of the 45 features in the best generated features by each EC algorithm. In this case, we put all the best generated subset of features using 5 EC algorithms tested against each dataset (MAHNOB, DEAP) together and then generated a set of features regardless of the channels (since each channel has the set of 45 features and in order to find the most repeated features we do not need to consider the channels). Then we provided figures to show the occurrence weight of each feature for each dataset.

The most frequent time-domain features from the MAHNOB dataset are maximum, 1st difference, 2nd difference, normalized 2nd difference, band power, HOC, mobility and complexity. Among the frequency-domain features, PSD from alpha, beta, and gamma are repeated more than the other frequency bands. Of the time-frequency domain features, *Rms_Theta*, *REE_Beta*, *power_Gamma*,

and *power_Theta* are selected more often than the others (see Fig. 6).

Similarly, as shown in Fig. 7, the most repeated features using the 5 EC algorithms tested against the DEAP dataset are mostly selected from the time domain such as: maximum, 1st difference, 2nd difference, normalized 2nd difference, median, mobility, complexity, HOC. Features from the frequency domain, such as PSD from Alpha and Gamma, are repeated more than the other frequency bands. Among the time-frequency features, *rms_Theta* and *rms_Gamma*, *power_Gamma*, *power_Theta* and *power_Alpha* are more frequent than the others.

For the two datasets (collectively), the most common frequently selected features are: maximum, 1st difference, 2nd difference, normalized 2nd difference, complexity, mobility, HOC, PSD from Gamma and Alpha, *rms_Theta*, *power_Gamma* and *power_Alpha*. However, results suggest that REE and LogREE from different bands from the time-frequency features are less efficient in emotion classification, since their relative occurrence is lower than the others. The relative occurrence of power features from DWT is higher than the other features in this domain, but not as high as PSD features from the frequency domain. It should be noted that the combination of time and frequency domain features is more efficient, since EC algorithms find the more successful and efficient subset of features by the combination of these features.

4.3. Channel selection

We investigated the most frequent set of selected channels (i.e. electrodes usage) by the combination of EC algorithms via the principle of weighted majority voting. Having the best subsets of features from each 10 runs of the EC algorithms, we then considered the task of building a vector representing the importance of channels. The importance of channels can be represented as follows:

$$w_c = \sum_{i=1}^k \sum_{j=1}^n \alpha_i * f_c, \quad 0 < \alpha_i < 1, \quad (8)$$

where κ , n are the number of algorithms and the number of runs in each algorithm respectively (10 runs for each algorithm is considered for this study). α_i represents the average weight of i_{th} algorithm over all runs which is dependent on the accuracy of classifiers over 10 runs. It means that at each run the performance of each EC algorithm is collected and then, based on the average performance over all 10 runs, an average weight (α_i) is allocated to each EC algorithm. This number is multiplied by f_c , which is the total number of selected features for c_{th} channel. w_c represents the weight of channel c .

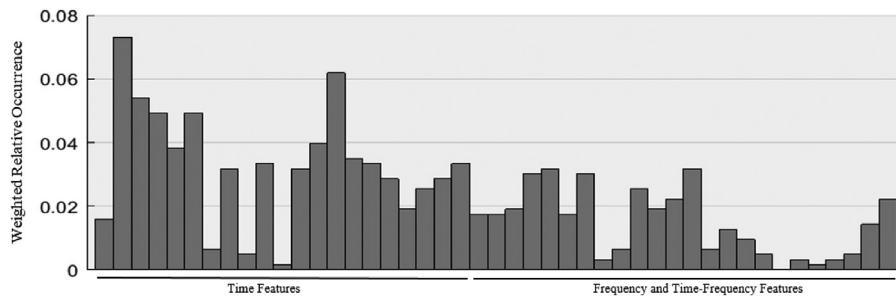


Fig. 6. The weighted relative occurrence of features over the MAHNOB dataset.

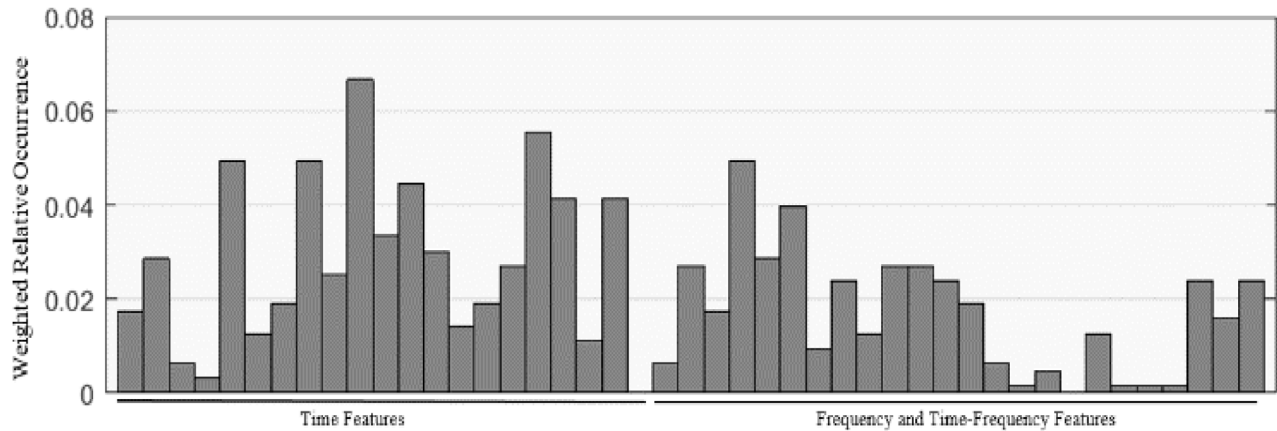


Fig. 7. The weighted relative occurrence of features over the DEAP dataset.

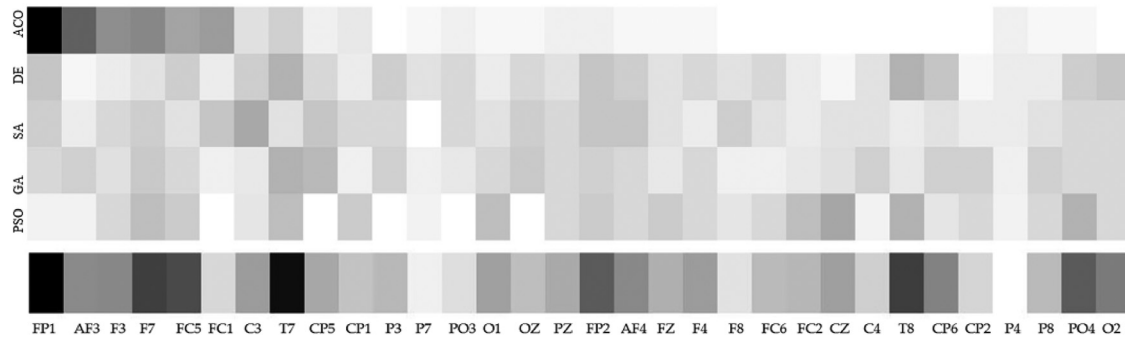


Fig. 8. Average electrode usage of each EC algorithm within 10 runs on the DEAP dataset is specified by darkness on each channel.

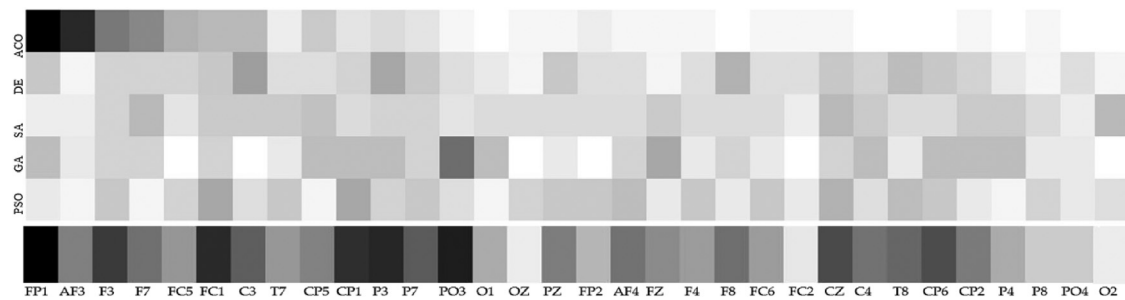


Fig. 9. Average electrode usage of each EC algorithm within 10 runs on the MAHNOB dataset is specified by darkness on each channel.

Figs. 8–10 show the plots of w_c based on the experiments using DEAP and MAHNOB (32 channels), and our dataset with 5 channels respectively. Darker boxes are channels with higher weight (w_c), which indicates the most repeated channels among the EC algorithms.

Based on the DEAP dataset, the average accuracy shows that the DE algorithm achieves better accuracy, followed by the PSO and

GA algorithms. Therefore, their average weight is higher than others. Based on the obtained results, FP1, F7, FC5, AF4, CP6, PO4, O2, T7 and T8 are the prominent electrodes among all other channels. Although the average weights of ACO and SA are lower, the number of selected features from the frontal lobe channels (FP1, F7, FC5) are big enough to compensate and increase the value of w_c . T7 and T8 are also salient, since these channels were frequently

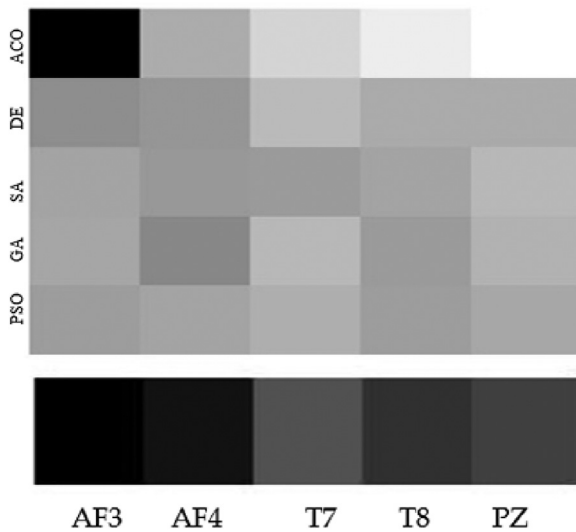


Fig. 10. Average electrode usage of each EC algorithm within 10 runs on our dataset is specified by darkness on each channel.

selected by PSO, GA and DE, which have bigger average weights. Moreover, features selected from PO4 and O2 are shown to give acceptable accuracy using these feature selection methods. Based on these findings, electrodes located over the frontal and parietal-partial lobes are generally favored over the occipital lobes (see Fig. 11a).

Based on the MAHNOB dataset, the most selected features using the ACO and GA algorithms are from the frontal lobes, which increase the w_c value of these channels (FP1, FC1, F3, F7, and AF4). In addition to those channels from the frontal lobe, the channels CP1, CZ, CP6, C3, T8, C4 and Cp2 are also highly selected by most of the feature-selection methods. We can conclude that most of the prominent channels using the 5 EC algorithms over the MAHNOB dataset are selected from the frontal and central lobes (Fig. 11a). Based on the most optimal subsets of features from each of the 10 runs of the EC algorithms, the most frequently selected channels are CZ, CP6, C3, T8, C4, CP2 from the frontal and central lobe.

Across DEAP and MAHNOB datasets, the electrodes located on the frontal and central lobes, such as FP1, AF4, CZ, T8, are found to be most salient. This is aligned with the results from our dataset, which found that the electrodes of frontal and central lobes were the most activated for four quadrant emotion classification (as

shown previously in Fig. 3). This confirms that emotion classification using the mobile and wireless *Emotiv Insight* sensor is feasible.

4.4. Comparison with other work over MAHNOB and DEAP datasets

A final experiment compared the best-tuned configuration of our system against some state-of-the-art methods. To this end, DE and PSO were used for feature selection and PNN as a classifier. Experimental results are shown in Table 3, indicating the classification accuracy for different emotion classification methods. While it only shows the DE results to represent the proposed method (as DE yields the highest performance based on the experiments). EC-based feature selection after 100 iterations is consistently better than not using any feature selection.

The comparisons show that although Yin, Wang, Liu, Zhang, and Zhang (2017) achieved promising accuracy (about 78%), their feature-selection method was Recursive Feature Elimination, which is computationally expensive. In addition, this method was used on two-class classification (Arousal and Valence), while our method was able to achieve similar accuracy (Maximum 71% ($67.474 \pm 3.389\%$) on DEAP dataset) on four-class classifications (HA-P, LA-P, HA-N, LA-N). The performance of their method was tested on one specific dataset (DEAP) with on mode of stimuli (music), whereas our proposed method was tested on two public datasets (DEAP and MAHNOB) plus a newly collected dataset using mobile sensors (Emotiv Insight), across two different modes of stimuli (music and video).

Ackermann, Kohlschein, Bitsch, Wehrle, and Jeschke (2016) classified three different emotions (anger, surprise and others) using Support Vector Machine (SVM) and Random Forest classification systems, which are trained by a smaller number of features. They only applied the Minimum Redundancy Maximum Relevance (mRMR) feature-selection method to eliminate less useful features. Their evaluation of the results on the DEAP dataset shows that the performance of the proposed method using the SVM classifier (average of 55%) is more robust and successful when the number of selected features is between 80 and 125.

Menezes et al. (2017) extracted some limited features from time and frequency domains and applied SVM method to classify each emotion dimension (Arousal and Valence), into two and three-class (Bipartition and Tripartition). This model has been tested on DEAP dataset. Although, the performance of their method in Bipartition is promising, the extracted features using SVM classifier could not classify Tripartition-class significantly.

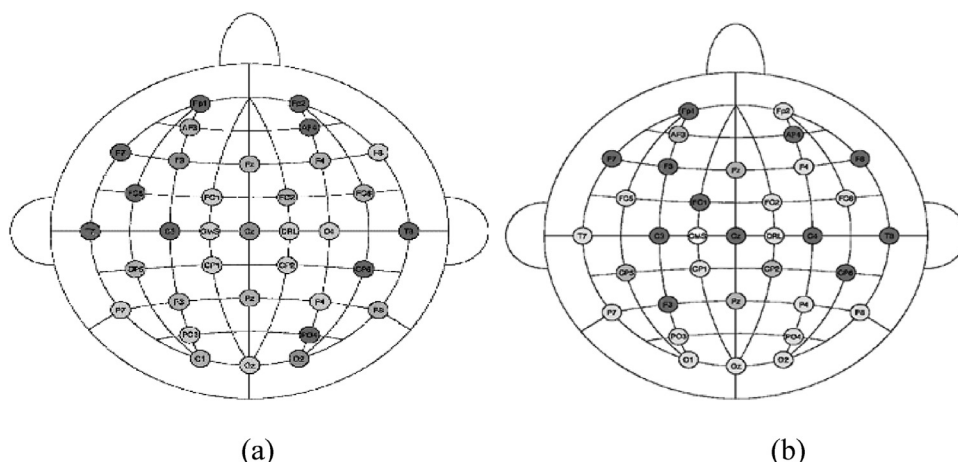


Fig. 11. The average electrode usage of each EC algorithm within 10 runs on (a) DEAP and (b) MAHNOB (a) dataset. On the greyscale, darkest nodes indicate the most frequently used channel.

Table 3

Comparison of our recognition approach with some state-of-the-art methods.

Method	Extracted features (No.)	Feature selection methods	Classifier	No. classes	Accuracy	Dataset
(Zhu, Wang, & Ji, 2014)	Statistical features (5)	–	SVM	2 Arousal Valence	Arousal: 60.23% Valence: 55.72%	MAHNOB
(Candra et al., 2015)	Time-frequency feature (2)	–	SVM	4 Sad Relaxed Angry Happy	60.9 ± 3.2%	DEAP
(Feradov & Ganchev, 2014)	Short term energy (2)	–	SVM	3 Negative Positive Neutral	62%	DEAP
(Ackermann et al., 2016)	Statistical features (number not specified)	mRMR	SVM & random forest	3 Anger, Surprise Others	Average accuracy: 55%	DEAP
(Kortelainen & Seppänen, 2013)	Frequency-domain features (number not specified)	Sequential feed-forward selection (SFFS)	KNN	2 Arousal Valence	Valence: 63% Arousal: 65%	MAHNOB
(Menezes et al., 2017)	Frequency domain and Time domain (11)	–	SVM	Bipartition Arousal & Valence 3 Triple on Arousal & Valence	Arousal = 69% Valence = 88% Arousal = 58% Valence = 63%	DEAP
(Yin et al., 2017)	Frequency and Time domain features (16)	Transfer Recursive Feature elimination (T-RFE)	LSSVM	2 Arousal Valence	Arousal: 78% Valence: 78%	DEAP
Our proposed method	Frequency, Time and Time–Frequency domain features (45)	EC algorithms	PNN	4 HA-P HA-N LA-P LA-N	DEAP: 67.474 ± 3.389% MAHNOB: 96.97 ± 1.893% Our Dataset: 65.043028 ± 3.195%	DEAP, MAHNOB & Our Dataset

In comparison with other latest studies, our proposed EEG-based emotion classification framework has shown state-of-the-art performance for both the DEAP and MAHNOB datasets, which confirms the value of integrating EC algorithms for feature selection to improve classification performance.

4.5. Towards the use of mobile EEG sensors for real world applications

The reliability and validity of mobile EEG sensors has been tested in earlier studies (Duvinage et al., 2013; Stytsenko, Jablonskis, & Prahm, 2011). These studies found that, while mobile sensors are not as accurate as a wired and full-scale EEG device, they can be used in noncritical applications. Some recent studies into the benefits of mobile EEG sensors on different domains have demonstrated an acceptable reliability (Leape, Fong, & Ratwani, 2016; Lushin & others, 2016; Wu, Wei, & Tudor, 2017; F. Zhang et al., 2017).

In our study, a mobile EEG sensor (Emotiv Insight) was used to recognize four-quadrant dimensional emotions while watching video clips. The proposed method uses different EC algorithms for feature selection applied to three different datasets (DEAP, MAHNOB and our collected dataset), and the result of 65% accuracy shows the validity of using mobile EEG sensors in this domain. In addition, we compared the salient channels in two datasets using full-scale EEG devices (32 channels) with the collected dataset using the mobile EEG sensor (5 channels). The result shows that the most frequent channels from the two public datasets (DEAP and MAHNOB) were from the frontal and central lobes, the same channels detected in our four-quadrant emotion recognition using five EC algorithms. This comparison confirms the feasibility of using mobile EEG sensor for emotion classification. However, the performance based on the MAHNOB dataset (96.97%) is still significantly higher than that of our new dataset (65.04%) despite using the same stimuli. This confirms that more progress is needed to improve the results for mobile EEG sensors. This paper paves a way for future sensor development in selecting the correct channels and features to focus on the most important electrodes.

5. Conclusion and future work

In this study, we propose the use of evolutionary computation (EC) algorithms (ACO, SA, GA, PSO and DE algorithms) for feature selection in an EEG-based emotion classification model for the classification of four-quadrant basic emotions (High Arousal-Positive emotions (HA-P), Low Arousal-Positive emotions (LA-P), High Arousal-Negative emotions (HA-N), and Low Arousal-Negative emotions (LA-N)). Our experiments have used two standard datasets, MAHNOB and DEAP, obtained using an EEG sensor with 32 channels, and a new dataset obtained using a mobile sensor with 5 channels. We have reported the performance of these algorithms with different time intervals (different iterations) – 10, 25, 45 and 100 iterations – to demonstrate the benefits of using EC algorithms to identify salient EEG signal features and improve the performance of classifiers.

Of the EC algorithms, DE and PSO showed better performance over every iteration. Moreover, the combination of time and frequency features consistently showed more efficient performance, compared to using only time or frequency features. The most frequently selected source of features (i.e. the EEG channels) were analyzed using the 5 EC algorithms and weighted majority voting. The electrodes in the frontal and central lobes were shown to be more activated during emotions based on DEAP and MAHNOB datasets, confirming the feasibility of using a lightweight and wireless EEG sensor (Emotiv Insight) for four-quadrant emotion classification.

Despite the promising results, this paper has identified an important limitation due to the premature convergence problem in EC algorithms, particularly DE, PSO and GA. Therefore, for future work, it is worthwhile exploring development of new EC algorithms or modifying the existing ones to overcome this problem and improve the classification performance accordingly.

Acknowledgements

We thank copy editing services provided by Golden Orb Creative, NSW, Australia. This work is supported by QUT International Postgraduate Research Scholarship (IPRS).

References

- Ackermann, P., Kohlschein, C., Bitsch, J. Á., Wehrle, K., & Jeschke, S. (2016). EEG-based automatic emotion recognition: Feature extraction, selection and classification methods. In *2016 IEEE 18th international conference on e-health networking, applications and services (Healthcom)* (pp. 1–6). <https://doi.org/10.1109/HealthCom.2016.7749447>.
- Aftanas, I. L., Lotova, N. V., Koshkarov, V. I., & Popov, S. A. (1998). Non-linear dynamical coupling between different brain areas during evoked emotions: An EEG investigation. *Biological Psychology*, 48(2), 121–138.
- Al-Ani, A. (2005). Feature subset selection using ant colony optimization. *International Journal of Computational Intelligence*. Retrieved from <https://opus.lib.uts.edu.au/handle/10453/6181>.
- Ansari-Asl, K., Chanel, G., & Pun, T. (2007). A channel selection method for EEG classification in emotion assessment based on synchronization likelihood. In *signal processing conference, 2007 15th European* (pp. 1241–1245). IEEE. Retrieved from <http://ieeexplore.ieee.org/abstract/document/7099003>.
- Baig, M. Z., Aslam, N., Shum, H. P., & Zhang, L. (2017). Differential evolution algorithm as a tool for optimal feature subset selection in motor imagery EEG. *Expert Systems with Applications*. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0957417417305109>.
- Bandyopadhyay, S., Saha, S., Maulik, U., & Deb, K. (2008). A simulated annealing-based multiobjective optimization algorithm: AMOSA. *IEEE Transactions on Evolutionary Computation*, 12(3), 269–283.
- Barry, R. J., Clarke, A. R., Johnstone, S. J., Magee, C. A., & Rushby, J. A. (2007). EEG differences between eyes-closed and eyes-open resting conditions. *Clinical Neurophysiology*, 118(12), 2765–2773.
- Blum, C., & Sampels, M. (2002). Ant colony optimization for FOP shop scheduling: A case study on different pheromone representations. In *Evolutionary computation, 2002. CEC'02. Proceedings of the 2002 congress on*: 2 (pp. 1558–1563). IEEE. Retrieved from <http://ieeexplore.ieee.org/abstract/document/1004474>.
- Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing*, 1(1), 18–37.
- Candra, H., Yuwono, M., Handojoseno, A., Chai, R., Su, S., & Nguyen, H. T. (2015). Recognizing emotions from EEG subbands using wavelet analysis. In *2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (pp. 6030–6033). <https://doi.org/10.1109/EMBC.2015.7319766>.
- Chai, T. Y., Woo, S. S., Rizon, M., & Tan, C. S. (2010). Classification of human emotions from EEG signals using statistical features and neural network. In *International: 1* (pp. 1–6). Penerbit UTHM. Retrieved from <http://eprints.uthm.edu.my/511/>.
- Cheng, B., & Liu, G.-Y. (2008). Emotion recognition from surface EMG signal using wavelet transform and neural network. In *Proceedings of the 2nd international conference on bioinformatics and biomedical engineering (ICBBE)* (pp. 1363–1366). Retrieved from http://www.paper.edu.cn/en_releasepaper/downPaper/200706-289.
- Colomi, A., Dorigo, M., Maniezzo, V., & Trubian, M. (1994). Ant system for job-shop scheduling. *Belgian Journal of Operations Research, Statistics and Computer Science*, 34(1), 39–53.
- Costa, D., & Hertz, A. (1997). Ants can colour graphs. *Journal of the Operational Research Society*, 48(3), 295–305.
- Couceiro, M. S., Rocha, R. P., & Ferreira, N. M. (2011). A novel multi-robot exploration approach based on particle swarm optimization algorithms. In *Safety, security, and rescue robotics (SSRR), 2011 IEEE international symposium on* (pp. 327–332). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6106751.
- Davidson, R. J. (2003). Affective neuroscience and psychophysiology: Toward a synthesis. *Psychophysiology*, 40(5), 655–665.
- Dorigo, M., Birattari, M., & Stutzle, T. (2006). Ant colony optimization. *IEEE Computational Intelligence Magazine*, 1(4), 28–39.
- Dorigo, M., & Gambardella, L. M. (1997). Ant colony system: A cooperative learning approach to the traveling salesman problem. *IEEE Transactions on Evolutionary Computation*, 1(1), 53–66.
- Du Boulay, B. (2011). Towards a motivationally intelligent pedagogy: How should an intelligent tutor respond to the unmotivated or the demotivated? In *New perspectives on affect and learning technologies* (pp. 41–52). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4419-9625-1_4.
- Duvinage, M., Castermans, T., Petieau, M., Hoellinger, T., Cheron, G., & Du-toit, T. (2013). Performance of the Emotiv Epoc headset for P300-based applications. *Biomedical Engineering Online*, 12(1), 56.
- Eberhart, R., & Kennedy, J. (1995). A new optimizer using particle swarm theory. In *Micro machine and human science, 1995. MHS'95, proceedings of the sixth international symposium on* (pp. 39–43). IEEE. Retrieved from <http://ieeexplore.ieee.org/abstract/document/494215/>.
- Feradov, F., & Ganchev, T. (2014). Detection of negative emotional states from electroencephalographic (EEG) signals. *Annual Journal of Electronics*, 8, 66–69.
- Goldberg, D. E., & Holland, J. H. (1988). Genetic algorithms and machine learning. *Machine Learning*, 3(2), 95–99.
- Gonçalves, J. F., de Magalhães Mendes, J. J., & Resende, M. G. (2005). A hybrid genetic algorithm for the job shop scheduling problem. *European Journal of Operational Research*, 167(1), 77–95.
- Hadjilimitriou, S. K., & Hadjileontiadis, L. J. (2012). Toward an EEG-based recognition of music liking using time-frequency analysis. *IEEE Transactions on Biomedical Engineering*, 59(12), 3498–3510.
- Horlings, R., Datu, D., & Rothkrantz, L. J. (2008). Emotion recognition using brain activity. In *Proceedings of the 9th international conference on computer systems and technologies and workshop for PhD students in computing* (p. 6). ACM. Retrieved from <http://dl.acm.org/citation.cfm?id=1500888>.
- Jenke, R., Peer, A., & Buss, M. (2014). Feature extraction and selection for emotion recognition from EEG. *IEEE Transactions on Affective Computing*, 5(3), 327–339. <https://doi.org/10.1109/TAFFC.2014.2339834>.
- Jung, T.-P., Makeig, S., Humphries, C., Lee, T.-W., Mckeown, M. J., Iragui, V., et al. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, 37(2), 163–178.
- Kennedy, J., & Eberhart, R. C. (1997). A discrete binary version of the particle swarm algorithm. In *Systems, man, and cybernetics, 1997. Computational cybernetics and simulation, 1997 IEEE international conference on*: 5 (pp. 4104–4108). IEEE. Retrieved from <http://ieeexplore.ieee.org/abstract/document/637339/>.
- Khan, M., Ahmed, S. I., Rahman, M., & Smith, R. O. (2011). A feature extraction method for realtime human activity recognition on cell phones. In *Proceedings of 3rd international symposium on quality of life technology (isQoLT 2011), Toronto, Canada*. Retrieved from <https://pdfs.semanticscholar.org/8aaa/9aa902b524992324fcb359ee4f01beeabdf.pdf>.
- Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598), 671–680.
- Koelstra, S., Muhl, C., Soleymani, M., Lee, J.-S., Yazdani, A., Ebrahimi, T., et al. (2012). Deap: A database for emotion analysis; using physiological signals. *IEEE Transactions on Affective Computing*, 3(1), 18–31.
- Kortelainen, J., & Seppänen, T. (2013). EEG-based recognition of video-induced emotions: Selecting subject-independent feature set. In *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)* (pp. 4287–4290). <https://doi.org/10.1109/EMBC.2013.6610493>.
- Kroupi, E., Yazdani, A., & Ebrahimi, T. (2011). EEG correlates of different emotional states elicited during watching music videos. In *Affective computing and intelligent interaction* (pp. 457–466). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-642-24571-8_58.
- Leape, C., Fong, A., & Ratwani, R. M. (2016). Heuristic usability evaluation of wearable mental state monitoring sensors for healthcare environments. In *Proceedings of the human factors and ergonomics society annual meeting*: 60 (pp. 583–587). CA: Los Angeles, CA: SAGE Publications Sage. Retrieved from <http://journals.sagepub.com/doi/abs/10.1177/1541931213601134>.
- Li, G., Lee, B.-L., & Chung, W.-Y. (2015). Smartwatch-based wearable EEG system for driver drowsiness detection. *IEEE Sensors Journal*, 15(12), 7169–7180.
- Liu, C., Conn, K., Sarkar, N., & Stone, W. (2008). Online affect detection and robot behavior adaptation for intervention of children with autism. *IEEE Transactions on Robotics*, 24(4), 883–896.
- Liu, Y., & Sourina, O. (2012). EEG-based dominance level recognition for emotion-enabled interaction. In *Multimedia and expo (ICME), 2012 IEEE international conference on* (pp. 1039–1044). IEEE. Retrieved from <http://ieeexplore.ieee.org/abstract/document/6298540/>.
- Luneski, A., Bamidis, P. D., & Hitoglou-Antoniadou, M. (2008). Affective computing and medical informatics: State of the art in emotion-aware medical applications. *Studies in Health Technology and Informatics*, 136, 517.
- Lushin, I. (2016). Detection of epilepsy with a commercial EEG headband. Retrieved from <http://www.doria.fi/handle/10024/123588>.
- Mandryk, R. L., & Atkins, M. S. (2007). A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International Journal of Human-Computer Studies*, 65(4), 329–347.
- Menezes, M. L. R., Samara, A., Galway, L., Sant'Anna, A., Verikas, A., Alonso-Fernandez, F., et al. (2017). Towards emotion recognition for virtual environments: An evaluation of eeg features on benchmark dataset. *Personal and Ubiquitous Computing*, 1–11.
- Michel, R., & Middendorf, M. (1998). An island model based ant system with look-ahead for the shortest supersequence problem. In *International conference on parallel problem solving from nature* (pp. 692–701). Springer. Retrieved from <http://link.springer.com/chapter/10.1007/BFb0056911>.
- Mistry, K., Zhang, L., Neoh, S. C., Lim, C. P., & Fielding, B. (2016). A micro-GA embedded PSO feature selection approach to intelligent facial emotion recognition. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=7456259.
- Murugappan, M., Rizon, M., Nagarajan, R., & Yaacob, S. (2010). Inferring of human emotional states using multichannel EEG. *European Journal of Scientific Research*, 48(2), 281–299.
- Murugappan, M., Ramachandran, N., & Szalai, Y. (2010). Classification of human emotion from EEG using discrete wavelet transform. *Journal of Biomedical Science and Engineering*, 3(04), 390.
- Murugappan, M., Muthusamy, Rizon, M., Nagarajan, R., Yaacob, S., Hazry, D., & Zunaiddi, I. (2008). Time-frequency analysis of EEG signals for human emotion detection. In *4th Kuala Lumpur international conference on biomedical engineering 2008* (pp. 262–265). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-69139-6_68.
- Nakisa, B., Nazri, M. Z. A., Rastgoo, M. N., & Abdullah, S. (2014). A survey: Particle swarm optimization based algorithms to solve premature convergence problem. *Journal of Computer Science*, 10(9), 1758–1765.
- Nakisa, B., Rastgoo, M. N., Nasrudin, M. F., & Nazri, M. Z. A. (2014). A multi-swarm particle swarm optimization with local search on multi-robot search system. *Journal of Theoretical and Applied Information Technology*, 71(1). Retrieved from <http://www.jatit.org/volumes/Vol71No1/15Vol71No1.pdf>.
- Nie, D., Wang, X.-W., Shi, L.-C., & Lu, B.-L. (2011). EEG-based emotion recognition during watching movies. In *Neural engineering (NER), 2011 5th international IEEE/EMBS conference on* (pp. 667–670). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5910636.

- Oh, S.-H., Lee, Y.-R., & Kim, H.-N. (2014). A novel EEG feature extraction method using Hjorth parameter. *International Journal of Electronics and Electrical Engineering*, 2(2), 106–110.
- Onton, J. A., & Makeig, S. (2009). High-frequency broadband modulation of electroencephalographic spectra. *Frontiers in Human Neuroscience*, 3, 61.
- Pan, Q.-K., Wang, L., & Qian, B. (2009). A novel differential evolution algorithm for bi-criteria no-wait flow shop scheduling problems. *Computers & Operations Research*, 36(8), 2498–2511.
- Petrantonakis, P. C., & Hadjileontiadis, L. J. (2010). Emotion recognition from brain signals using hybrid adaptive filtering and higher order crossings analysis. *IEEE Transactions on Affective Computing*, 1(2), 81–97.
- Price, K., Storn, R. M., & Lampinen, J. A. (2006). *Differential evolution: A practical approach to global optimization*. Springer Science & Business Media. Retrieved from https://books.google.com.au/books?hl=en&lr=&id=hakXI-dEhTkC&oi=fnd&pg=PR7&dq=Differential+Evolution:+A+Practical+Approach+to+Global+Optimization&ots=c_5DKPKf60&sig=bn7XPVZZGJqyG8Ko-8sdN9h0JD8.
- 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.
- Rastgoo, M. N., Nakisa, B., & Nazri, M. Z. A. (2015). A hybrid of modified PSO and local search on a multi-robot search system. *International Journal of Advanced Robotic Systems*, 11. Retrieved from <http://journals.sagepub.com/doi/abs/10.5772/60624>.
- Redmond, S. J., & Heneghan, C. (2006). Cardiorespiratory-based sleep staging in subjects with obstructive sleep apnea. *IEEE Transactions on Biomedical Engineering*, 53(3), 485–496.
- Rodriguez-Bermudez, G., Garcia-Laencina, P. J., & Roca-Dorda, J. (2013). Efficient automatic selection and combination of EEG features in least squares classifiers for motor imagery brain–computer interfaces. *International Journal of Neural Systems*, 23(04), 1350015.
- Schaaff, K., & Schultz, T. (2009). Towards emotion recognition from electroencephalographic signals. In *2009 3rd international conference on affective computing and intelligent interaction and workshops* (pp. 1–6). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5349316.
- Sha, D. Y., & Hsu, C.-Y. (2006). A hybrid particle swarm optimization for job shop scheduling problem. *Computers & Industrial Engineering*, 51(4), 791–808.
- Shah, F. (2010). Discrete wavelet transforms and artificial neural networks for speech emotion recognition. *International Journal of Computer Theory and Engineering*, 2(3), 319.
- Sim, K. M., & Sun, W. H. (2003). Ant colony optimization for routing and load-balancing: Survey and new directions. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 33(5), 560–572.
- Sivagaminathan, R. K., & Ramakrishnan, S. (2007). A hybrid approach for feature subset selection using neural networks and ant colony optimization. *Expert Systems with Applications*, 33(1), 49–60.
- Soleymani, M., Lichtenauer, J., Pun, T., & Pantic, M. (2012). A multimodal database for affect recognition and implicit tagging. *Affective Computing*, *IEEE Transactions on*, 3(1), 42–55.
- Sourina, O., Kulish, V. V., & Sourin, A. (2009). Novel tools for quantification of brain responses to music stimuli. In *13th international conference on biomedical engineering* (pp. 411–414). Springer. Retrieved from <http://www.springerlink.com/index/m8681026rr545831.pdf>.
- Sourina, O., & Liu, Y. (2011). A fractal-based algorithm of emotion recognition from EEG using arousal-Valence model. *BIO SIGNALS*, 209–214. Retrieved from http://ntu.edu.sg/home/eosourina/Papers/OSBIO SIGNALS_66_CR.pdf.
- Sourina, O., Sourin, A., & Kulish, V. (2009). EEG data driven animation and its application. *Computer Vision/Computer Graphics Collaboration Techniques*, 380–388.
- Specht, D. F. (1990). Probabilistic neural networks. *Neural Networks*, 3(1), 109–118.
- Stytsenko, K., Jablonskis, E., & Prahm, C. (2011). Evaluation of consumer EEG device emotiv EPOC. *MEI: CogSci conference 2011*. Ljubljana. Retrieved from <http://www.univie.ac.at/meicogsci/php/ocs/index.php/meicog/meicog2011/paper/view/210>.
- Suresh, R. K., & Mohanasundaram, K. M. (2006). Pareto archived simulated annealing for job shop scheduling with multiple objectives. *The International Journal of Advanced Manufacturing Technology*, 29(1), 184–196.
- Takahashi, K. (2004). Remarks on emotion recognition from multi-modal biopotential signals. In *2004 IEEE international conference on industrial technology, 2004. IEEE ICT '04*: 3 (pp. 1138–1143). Vol. 3. <https://doi.org/10.1109/ICIT.2004.1490720>.
- Wang, Q., Sourina, O., & Nguyen, M. K. (2010). Eeg-based “serious” games design for medical applications. In *Cyberworlds (CW), 2010 international conference on* (pp. 270–276). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5656236.
- Wu, Y., Wei, Y., & Tudor, J. (2017). A real-time wearable emotion detection headband based on EEG measurement. *Sensors and Actuators A: Physical*. Retrieved from <http://www.sciencedirect.com/science/article/pii/S092442471631192X>.
- Yang, J., & Honavar, V. (1998). Feature subset selection using a genetic algorithm. In *Feature extraction, construction and selection* (pp. 117–136). Springer. Retrieved from http://link.springer.com/chapter/10.1007/978-1-4615-5725-8_8.
- Yin, Z., Wang, Y., Liu, L., Zhang, W., & Zhang, J. (2017). Cross-subject EEG feature selection for emotion recognition using transfer recursive feature elimination. *Frontiers in Neuroinformatics*, 11. <https://doi.org/10.3389/fninf.2017.00019>.
- Zhang, F., Haddad, S., Nakisa, B., Rastgoo, M. N., Candido, C., Tjondronegoro, D., et al. (2017). The effects of higher temperature setpoints during summer on office workers' cognitive load and thermal comfort. *Building and Environment*. Retrieved from <http://www.sciencedirect.com/science/article/pii/S0360132317302834>.
- Zhang, G., Shao, X., Li, P., & Gao, L. (2009). An effective hybrid particle swarm optimization algorithm for multi-objective flexible job-shop scheduling problem. *Computers & Industrial Engineering*, 56(4), 1309–1318.
- Zhang, S., & Zhao, Z. (2008). Feature selection filtering methods for emotion recognition in Chinese speech signal. In *2008 9th international conference on signal processing* (pp. 1699–1702). IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4697464.
- Zhu, Q., Yan, Y., & Xing, Z. (2006). Robot path planning based on artificial potential field approach with simulated annealing. In *Intelligent systems design and applications, 2006. ISDA'06. Sixth international conference on*: 2 (pp. 622–627). IEEE. Retrieved from <http://ieeexplore.ieee.org/abstract/document/4021735/>.
- Zhu, Y., Wang, S., & Ji, Q. (2014). Emotion recognition from users' EEG signals with the help of stimulus VIDEOS. In *2014 IEEE international conference on multimedia and expo (ICME)* (pp. 1–6). <https://doi.org/10.1109/ICME.2014.6890161>.