



EEG-based person identification through Binary Flower Pollination Algorithm



Douglas Rodrigues^b, Gabriel F.A. Silva^a, João P. Papa^{a,*}, Aparecido N. Marana^a, Xin-She Yang^c

^aSão Paulo State University, Department of Computing, Bauru, Brazil

^bFederal University of São Carlos, Department of Computing, São Carlos, Brazil

^cMiddlesex University, School of Science and Technology, London, UK

ARTICLE INFO

Article history:

Received 21 August 2015

Revised 3 June 2016

Accepted 4 June 2016

Available online 14 June 2016

Keywords:

Meta-heuristic

Pattern classification

Biometrics

Electroencephalogram

Optimum-path forest

ABSTRACT

Electroencephalogram (EEG) signal presents a great potential for highly secure biometric systems due to its characteristics of universality, uniqueness, and natural robustness to spoofing attacks. EEG signals are measured by sensors placed in various positions of a person's head (channels). In this work, we address the problem of reducing the number of required sensors while maintaining a comparable performance. We evaluated a binary version of the Flower Pollination Algorithm under different transfer functions to select the best subset of channels that maximizes the accuracy, which is measured by means of the Optimum-Path Forest classifier. The experimental results show the proposed approach can make use of less than a half of the number of sensors while maintaining recognition rates up to 87%, which is crucial towards the effective use of EEG in biometric applications.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

In modern life, we constantly make use of passwords to access our bank accounts, e-mail boxes, and social networks, just to name a few. As passwords can be easily circumvented, the use of biometrics has been proposed for safe person identification (Jain, Ross, & Nandakumar, 2011). Over the years, the use of biometric systems has increased, and systems based on several biometric modalities such as fingerprint, face and iris, have been successfully deployed. This successful and widespread deployment of biometric systems brings on a new challenge: spoofing. Spoofing methods are developed to breach the security of biometric systems so that unauthorized users can gain access to places and/or information (e.g., an artificial finger made from silicone is placed on the fingerprint scanner).

In this scenario, the EEG (electroencephalogram) signal presents a great potential for highly secure biometric-based person identification, due to its characteristics of universality, uniqueness, and robustness to spoofing attacks (Beijsterveldt & Boomsma, 1994). It is well-known the importance of EEG signals in several ar-

reas, since one can find a number of works that deal with such a source of data (Guo, Rivero, Dorado, Munteanu, & Pazos, 2011; Nunes, Coelho, Lima, Papa, & Albuquerque, 2014; Ocak, 2009; Subasi, 2007). In high security environments, EEG sensors can be integrated in order to contribute to the robustness of the system, and the person can be continuously authenticated. Although the idea of using EEG as a biometric trait is not new, there are a few works that address such kind of signal only. One possible explanation for that is the difficulty in obtaining such signals, and also because the biometric characteristics of the EEG signal may be held only for short periods of time (Pollock, Schneider, & Lyness, 1991).

With the emergence of new mobile devices that capture brain signals driven by the most keenly studies in the brain computer interface, the EEG as a biometric trait can now be used in some other scenarios, such as: (i) distance-based education environments, in which the continuous authentication of a student becomes increasingly necessary; (ii) with the increase in life expectancy worldwide, health monitoring systems may become popular along with home automation and smart homes, thus making the EEG-based identification very useful in this scenario; (iii) with the popularization of biometric systems for the validation of financial transactions, mobile EEG sensors become a viable alternative in the future.

Basically, an EEG-based biometric approach aims at placing a set of sensors in the person's head in order to capture the output signals for further feature extraction and analysis using signal

* Corresponding author.

E-mail addresses: douglasrodrigues.dr@gmail.com (D. Rodrigues), ec.gabrielalvarez@gmail.com (G.F.A. Silva), papa@fc.unesp.br, papa.joaopaulo@gmail.com (J.P. Papa), nilceu@fc.unesp.br (A.N. Marana), x.yang@mdx.ac.uk (X.-S. Yang).

processing techniques. The signal acquisition session is then repeated over time to make the system more discriminative and robust to errors. In a recent paper, (Campisi & La Rocca, 2014) presented a review on the state-of-the-art of EEG-based automatic recognition systems, as well as an overview of the neurophysiological basis that constitutes the foundations on which EEG biometric systems can be built. The authors also discussed about the major obstacles towards the deployment of EEG based biometric systems in everyday life.

One of the main problems of EEG-based person identification is the acquisition, which may be too invasive to the user. The process of putting a considerable amount of sensors up on a person's head might be a bit uncomfortable, and it also requires a previous knowledge by the person in charge of the sensors placement in order to put them in their correct positions. In light of this context, some questions may rise: "Is it really necessary to put all these sensors on a persons' head? If not, can we identify the most relevant channels for person identification and then use a smaller number of sensors in order to measure them?"

These questions motivated our work in modelling the task of channel selection as an evolutionary-based optimization problem. The idea is to propose a wrapper approach composed by an optimization technique and a pattern classifier, in which the accuracy of the latter is used to guide the evolutionary agents in the search space looking for the best solutions, i.e., the subset of channels that maximize the accuracy of the classifier in the validation set. Any optimization technique and classifier could be used.

In our work, we propose an optimum channel selection by means of a binary constrained version of the recently proposed optimization technique Flower Pollination Algorithm (BFPA) (Yang, 2012), and the Optimum-Path Forest (OPF) (Papa, Falcão, Albuquerque, & Tavares, 2012; Papa, Falcão, & Suzuki, 2009) classifier, which is a supervised pattern recognition technique that has the advantage of providing a faster training phase compared to other state-of-the-art classifiers. This characteristic of fast training is very important in the context of this paper, since a training procedure followed by a classification of a validation set need to be performed for each evolutionary agent (sometimes we may have several of them). Additionally, this version of OPF is parameterless, which is another advantage over other classifiers.

The main contributions of this paper are three-fold: (i) to evaluate a recent binary version of the Flower Pollination Algorithm (BFPA) proposed by Rodrigues, Yang, Souza, and Papa (2015) under different transfer functions¹; (ii) to model the problem of EEG channel selection as an evolutionary-based optimization task; and (iii) to introduce the OPF classifier for EEG-based biometric person identification. The use of evolutionary optimization algorithms for the EEG channel selection is due to their elegant and simple solutions to solve optimization problems, similar to the way nature does.

This paper is organized as follows: Section 2 presents a brief theoretical background about EEG, and Section 3 discusses previous works related to this paper. Section 4 presents the proposed approach for person identification using a reduced number of EEG channels, and Section 5 presents a description of the dataset and the experimental setup. Sections 6 and 7 discuss the experiments and conclusions, respectively.

2. The EEG signal

The human central nervous system consists of the encephalous (brain), which is inside the cranium, and the spinal cord contained in the spine. The nerve tissue is a complex network formed mostly

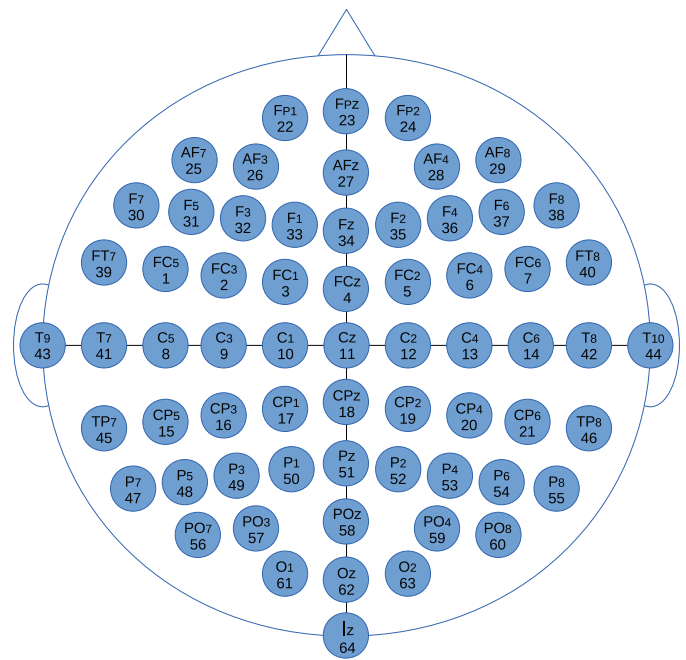


Fig. 1. International 10-10 System standards for sensor positioning. Just for the sake of clarification, sensor T9 is placed close to the left ear, as well as sensor #23 is placed close to the nose.

by millions of nerve cells (glial cells and neurons), whose primary function is the transmission of electrical impulses that run through this intrinsic and huge network, thus propagating information among cells (Sanei & Chambers, 2007; Tau & Peterson, 2009). These small electrical impulses emitted by the huge amount of neurons create an electric field that can be measured on the surface of the human skull, with the help of sensors or electrodes. The measurement of this complex electrical signal from our nervous system is what is known as electroencephalogram (EEG). In the literature, it is common among authors to directly refer to those brain waves as EEG.

The neural activity of the human being begins between the 17th and 23rd week of gestation. It is believed that, since this stage, and throughout the life, the signals from the brain activity represent not only the functioning of the brain, but also of the whole body. Published studies also show that even if a variation in amplitude of EEG signals during the development of a normal person exists, over the years, their functional connections remain largely unchanged (Gasser, Jennen-Steinmetz, Sroka, Verleger, & Macks, 1988; Tau & Peterson, 2009).

Fig. 1 shows an example of a map of sensors located at a person's head. This map describes the head surface locations via relational distances, also called as International 10-10 System (Jurcak, Tsuzuki, & Dan, 2007; Nuwer et al., 1998). The nomenclature of the electrodes is associated to the human brain areas as follows: Frontal (F), Central (C), Temporal (T), Parietal (P) and Occipital (O) lobes. Electrodes named with two letters refer to a location between areas, for example: CP electrode is in a position between central and parietal lobes. The sub-index indicates the side of the brain hemisphere (odd numbers are located on the left side and even numbers on the right side), and the sub-index "z" indicates that the electrode is located in the main vertical axis.

3. Related work

One of the first studies regarding EEG as a biometric trait was conducted by Poulos, Rangoussi, Chrissikopoulos, and Evangelou (1999), which described the EEG signal by means of an autoregressive (AR) model as the basis for a person identification method. In

¹ A transfer function, in this context, aims at mapping a real-valued solution to a binary-valued one.

their work, the correct classification rates reached 91% in experiments using data obtained from 45 EEG recordings of 75 subjects, who were at rest and with the eyes closed during the test. Another study by Poulos, Rangoussi, and Alexandris (1999) employed spectral features extracted from the EEG signals followed by the use of neural networks as classifiers to identify a person. The authors have achieved correct classification rates ranging from 80% to 100%, reaffirming the great potential of using EEG as a biometric feature. Abdullah, Subari, Loong, and Ahmad (2010) implemented a practical system that uses four (sometimes fewer) channels and two types of EEG signals (one with the eyes open and another one with the eyes closed), which were used in ten male subjects at rest in five different sessions conducted over the course of two weeks. The feature extraction was performed using AR models, and the classification was performed using a multilayer neural network. The authors observed classification rates from 70% to 97%, depending on the amount of channels and EEG type.

Palaniappan (2004) used the gamma-band spectral power ratio as features and a Multilayer Perceptron Neural Network to recognize a person based on the EEG signal. Later on, Palaniappan and Mandic (2007) proposed to use 61 channels for feature extraction followed by classification using Elman Neural Network. Kostílek and Šťastný (2012) focused on the importance of the repeatability and the influence of movements during the EEG signal acquisition session. In their work, an autoregressive model and a Mahalanobis distance-based classifier for person identification were applied to evaluate the robustness of the proposed approach. Safont, Salazar, Soriano, and Vergara (2012) used a set of classifiers and multiple features to perform EEG-based person identification. In their work, all possible combinations of features and classifiers have been addressed in order to improve the person recognition results.

More recently, La Rocca et al. (2014) proposed a novel approach that fuses spectral coherence-based connectivity between different brain regions as a possibly viable biometric feature. The proposed approach was tested on a dataset of 108 subjects with eyes-closed (EC) and eyes-open (EO) resting state conditions. Their results show that using brain connectivity leads to higher distinctiveness when compared with the traditional power-spectrum measurements, reaching 100% of recognition accuracy in EC and EO conditions when integrating functional connectivity between regions in the frontal lobe.

4. Proposed method

In this section, we present our proposed method for person identification based on features from EEG signals, as well as we briefly review some of the main concepts regarding the techniques employed in this paper.

4.1. Autoregressive model

An autoregressive model can be described by a linear difference equation in the time domain as follows:

$$x(k) = P + \sum_{i=1}^p a(i)x(t-i) + e(t), \quad (1)$$

where P is a constant, p stands for the number of parameters of the model and $e(t)$ denotes a white noise input (Jain & Deshpande, 2004). Notice In this work, we used the Yule–Walker method to estimate the coefficients of the AR model by employing the least square method criterion.

4.2. EEG channel selection

In order to select the best subset of channels, we evaluate a recent proposed binary version of the Flower Pollination Algo-

rithm (Rodrigues et al., 2015) under different transfer functions, and we also show we can obtain distinct results for each one. Firstly, we present the theoretical basis about FPA, and then its binary version.

4.2.1. Flower pollination algorithm

The Flower Pollination Algorithm proposed by Yang (2012) is inspired by the flow pollination process of flowering plants. The FPA is governed by four basic rules:

1. Biotic cross-pollination can be considered as a process of global pollination, and pollen-carrying pollinators move in a way that obeys Lévy flights;
2. For local pollination, abiotic pollination and self-pollination are used;
3. Pollinators such as insects can develop flower constancy, which is equivalent to a reproduction probability that is proportional to the similarity of two flowers involved; and
4. The interaction or switching of local pollination and global pollination can be controlled by a switch probability $p \in [0, 1]$, slightly biased towards local pollination.

In order to model the updating formulas, the above rules have to be converted into proper updating equations. For example, in the global pollination step, flower pollen gametes are carried by pollinators such as insects, and pollen can travel over a long distance because insects can often fly and move over a much longer range. Therefore, Rules 1 and 3 can be represented mathematically as follows:

$$x_i^{(t+1)} = x_i^t + \alpha L(\lambda)(g_* - x_i^t), \quad (2)$$

where

$$L(\lambda) = \frac{\lambda \cdot \Gamma(\lambda) \cdot \sin(\lambda)}{\pi} \cdot \frac{1}{s^{1+\lambda}}, \quad s \gg s_0 > 0 \quad (3)$$

where x_i^t is the pollen i (solution vector) at iteration t , g_* is the current best solution among all solutions at the current generation, and α is a scaling factor to control the step size. $L(\lambda)$ is the Lévy-flights step size, that corresponds to the strength of the pollination, $\Gamma(\lambda)$ stands for the gamma function and s is the step size. Since insects may move over a long distance with various distance steps, a Lévy flight can be used to mimic this characteristic efficiently.

For local pollination, both Rules 2 and 3 can be represented as:

$$x_i^{(t+1)} = x_i^t + \epsilon(x_j^t - x_k^t), \quad (4)$$

where x_j^t and x_k^t are pollen from different flowers j and k of the same plant species at time step t . This mimics flower constancy in a limited neighbourhood. Mathematically, if x_j^t and x_k^t come from the same species or are selected from the same population, it equivalently becomes a local random walk if ϵ is drawn from a uniform distribution in $[0,1]$. In order to mimic the local and global flower pollination, a switch probability (Rule 4) or proximity probability p is used.

4.2.2. Binary flower pollination algorithm

In the standard FPA, the solutions are updated in the search space towards continuous-valued positions. However, in the proposed Binary Flower Pollination Algorithm the search space is modelled as an n -dimensional boolean lattice, in which the solutions are updated across the corners of a hypercube. In addition, as the problem is to select or not a given feature, a solution binary vector is employed, where 1 corresponds to a feature being selected to compose the new set, and 0 otherwise. In order to build this binary vector, (Rodrigues et al., 2015) employed Eqs. (5) and (6), which can restrict the new solutions to only binary values:

$$S(x_i^j(t)) = \frac{1}{1 + e^{-x_i^j(t)}}, \quad (5)$$

$$x_i^j(t) = \begin{cases} 1 & \text{if } S(x_i^j(t)) > \sigma, \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

in which $\sigma \sim U(0, 1)$. Algorithm 1 presents the proposed approach that employs BFPA for EEG-channel selection using the OPF classifier as the objective function and Eqs. (5) and (6) as the transfer function. Note that the proposed approach can be used with any other supervised classification technique.

Algorithm 1: BFPA - Binary Flower Pollination Algorithm

```

input      : Training set  $Z_1$  and evaluating set  $Z_2$ ,  $\alpha$ , number
              of flowers  $m$ , dimension  $d$  and iterations  $T$ .
output    : Global best position  $\hat{g}$ .
auxiliaries: Fitness vector  $f$  with size  $m$  and variables  $acc$ ,
               $maxfit$ ,  $globalfit \leftarrow -\infty$  and  $maxindex$ .
1 for each flower  $i$  ( $\forall i = 1, \dots, m$ ) do
2   for each dimension  $j$  ( $\forall j = 1, \dots, d$ ) do
3      $x_i^j(0) \leftarrow \text{Random}\{0, 1\}$ ;
4      $f_i \leftarrow -\infty$ ;
5 for each iteration  $t$  ( $t = 1, \dots, T$ ) do
6   for each flower  $i$  ( $\forall i = 1, \dots, m$ ) do
7     Create  $Z'_1$  and  $Z'_2$  from  $Z_1$  and  $Z_2$ , respectively, such
      that both contains only features such that  $x_i^j(t) \neq 0$ ,
       $\forall j = 1, \dots, d$ ;
8     Train OPF over  $Z'_1$ , evaluate its over  $Z'_2$  and stores the
      accuracy in  $acc$ ;
9     if ( $acc > f_i$ ) then
10       $f_i \leftarrow acc$ ;
11      for each dimension  $j$  ( $\forall j = 1, \dots, d$ ) do
12         $\hat{x}_i^j \leftarrow x_i^j(t)$ ;
13   [ $maxfit$ ,  $maxindex$ ]  $\leftarrow \max(f)$ ;
14   if ( $maxfit > globalfit$ ) then
15      $globalfit \leftarrow maxfit$ ;
16     for each dimension  $j$  ( $\forall j = 1, \dots, d$ ) do
17        $\hat{g}^j \leftarrow x_{maxindex}^j(t)$ ;
18 for each flower  $i$  ( $\forall i = 1, \dots, m$ ) do
19   for each dimension  $j$  ( $\forall j = 1, \dots, d$ ) do
20      $rand \leftarrow \text{Random}\{0, 1\}$ ;
21     if  $rand < p$  then
22        $x_i^j(t) \leftarrow x_i^j(t-1) + \alpha \oplus \text{Lévy}(\lambda)$ ; else
23        $x_i^j(t) \leftarrow x_i^j(t-1) + \epsilon(x_i^j(t-1) - x_i^k(t-1))$ ;
24     if ( $\sigma < \frac{1}{1+e^{x_i^j(t)}}$ ) then
25        $x_i^j(t) \leftarrow 1$ ; else
26        $x_i^j(t) \leftarrow 0$ ;
  
```

Lines 1–4 initialize each pollen's position as being a binary string with random values, as well as the fitness value f_i of each individual i . The main loop in Lines 6–27 is the core of the proposed algorithm, in which the inner loop in Lines 7–13 is responsible for creating the new training Z'_1 and evaluating sets Z'_2 , and then OPF is trained over Z'_1 and it is used to classify Z'_2 . The recognition accuracy over Z'_2 is stored in acc and then compared with the fitness value f_i (accuracy) of individual i : if the later is worse than acc , the old fitness value is kept; in the opposite case, the fitness value is then updated. Lines 12–13 update the best local position of the current pollen. Lines 14–18 update the global optimum,

and the last loop (Lines 19–27) moves each pollen to a new binary position restricted by Eqs. (5) and (6) (Lines 25–27).

4.3. Optimum-path forest classifier

We used the Optimum-Path Forest classifier (Papa et al., 2012; Papa et al., 2009) applied to the features learned from the AR model to classify a person based on the EEG signal. The OPF works by modelling the samples as graph nodes, whose arcs are defined by an adjacency relation and weighted by a distance function. Further, a role competition process between some key nodes (prototypes) is carried out in order to partition the graph into optimum-path trees (OPTs) according to a path-cost function. In fact, each OPT is rooted at one prototype, which means a sample that belongs to a given tree is more strongly connected to its root than to any other in the forest.

5. Methodology

In this section, we present the proposed approach for channel selection in EEG-based signal acquisition, as well as we briefly describe the employed dataset, the nature-inspired meta-heuristic algorithms, and the experimental setup.

5.1. Dataset

The EEG signals used in this work were obtained from the EEG Motor Movement/Imagery dataset² (Goldberger et al., 2000). The data was collected from 109 healthy volunteers using the BCI2000 System (Schalk, McFarland, Hinterberger, Birbaumer, & Wolpaw, 2004), which makes use of 64 channels (sensors) and provides a separated EDF (European Data Format) file for each of them. The subjects performed different motor/imagery tasks: such tasks are mainly used in BCI (Brain-Computer Interface) applications and neurological rehabilitation, and consists of imagining or simulating a given action, like open and close the eyes, for example.

Each subject performed four tasks according to the position of a target that appears on the screen placed in front of the volunteers (if the target appears on the right or left side, the subject opens and closes the corresponding fist; if the target appears on the top or bottom side, the subject opens and closes both fists or both feet, respectively). In short, the four experimental tasks were:

1. To open and close left or right fist;
2. To imagine opening and closing left or right fist;
3. To open and close both fists or both feet; and
4. To imagine opening and closing both fists or both feet.

Each of these tasks were performed three times, thus generating 12 recordings for each subject of a two-minutes run, and the 64 channels were sampled at 160 samples per second.

The features of the twelve recordings are extracted by means of an AR model with three output configurations for each EEG-channel: 5, 10 and 20 features. Further, the average of each configuration is then been computed in order to obtain just one feature per EEG-channel (sensor). In short, for each sensor, we have extracted three different numbers of AR-based features, being the output of each sensor the average of their values. Henceforth, we have adopted the following notation for each of the dataset configurations: AR₅ for 5 autoregression coefficients extracted, and AR₁₀ and AR₂₀ for 10 and 20 autoregression coefficients, respectively.

² <http://physionet.org/pn4/eegmiddb>.

5.2. Nature-inspired meta-heuristic algorithms

In this work, we have compared our proposed method with other meta-heuristic-based optimization methods described below:

Genetic algorithm (GA): The Genetic algorithm was proposed by Holland (1975), and its main concept is to emulate the biological evolution to solve optimization problems. It is composed of an initial population (or a set of unique elements) and a set of operators inspired by the nature. These operators can change the elements, and according to the evolutionary theory, only the most capable individuals are able to survive and transmit their biological heredity to the next generations.

Particle swarm optimization (PSO): This method is inspired on the social behaviour of a bird flocking or a fish schooling (Kennedy & Eberhart, 2001). The fundamental idea is that each particle represents a potential solution which is updated according to its own experience and from its neighbours' knowledge. The motion of an individual particle for the optimal solution is governed through its position and velocity interactions, and also by its own previous best performance and the best performance of their neighbours.

Firefly algorithm (FA): This method was proposed by Yang (2010), being derived from the flash attractiveness of fireflies for mating partners (communication) and attracting potential preys. The brightness of a firefly at a given position is determined by the value of the objective function in that position. Each firefly is attracted by a brighter firefly through the attraction factor.

Harmony search (HS): This method is a meta-heuristic algorithm inspired in the improvisation process of music players (Geem, 2009). Musicians often improvise the pitches of their instruments searching for a perfect state of harmony. The main idea is to use the same process adopted by musicians to create new songs to obtain a near-optimal solution according to some fitness function. Each possible solution is modelled as a harmony, and each musical note corresponds to one decision variable.

Charged system search (CSS): This method, based on the governing Coulomb's law (a physics law used to describe the interactions between electrically charged particles), was proposed by Kaveh and Talatahari (2010). In this method, named CSS, each Charged Particle (CP) in the system is affected by the electrical fields of the others, generating a resultant force over each CP, which is determined by using the electrostatic laws. The CP interaction movement is determined by Newtonian mechanics laws.

We have used the binary optimization version of each aforementioned method, as proposed in: Binary GA (BGA) (Holland, 1975), Binary PSO (BPSO) (Firpi & Goodman, 2004), Binary HS (BHS) (Ramos, Souza, Chiachia, Falcão, & Papa, 2011), Binary Firefly (BFA) (Falcon, Almeida, & Nayak, 2011; Palit, Sinha, Molla, Khanra, & Kule, 2011), and Binary CSS (Rodrigues et al., 2013). The optimization algorithms were implemented in C language following the guidelines provided by their references. Notice the transfer function defined by Eqs. (5) and (6) were the very same for all techniques compared in this work.

5.3. Experimental setup

We partitioned our fully labeled dataset into $Z = Z_1 \cup Z_2 \cup Z_3$ subsets, in which Z_1 , Z_2 and Z_3 stand for training, validation, and test sets, respectively. The training dataset contains 50% of the original dataset, followed by 30% and 20% concerning the validation and test sets, respectively. The idea is to employ Z_1 and Z_2

Table 1

Parameters used for each meta-heuristic optimization technique. Notice the inertia weight w for PSO was linearly decreased from 0.9 to 0.4.

Technique	Parameters
BGA	$mutation = 0.1$
BPSO	$c_1 = c_2 = 2$
BFA	$\gamma = 0.8, \beta_0 = 1.0, \alpha = 0.01$
BCSS	–
BHS	HMCR= 0.9
BFPA	$\alpha = 1.0, p = 0.8$

to find the subset of features that maximize the accuracy over Z_2 , with the accuracy being the fitness function.

Each agent is initialized with random binary positions and the original dataset is mapped to a new one that contains the features that were selected in this first sampling. In addition, the fitness function of each agent is set to the OPF recognition rate over Z_2 after training in Z_1 . The final subset will be the one that maximizes the curve over the range of values, i.e., the features that maximize the accuracy over Z_2 . The accuracy over the test set Z_3 is then assessed by using the final subset of the selected features. Notice the fitness function employed in this paper is the accuracy measure proposed by Papa et al. (2009), which is capable of handling unbalanced classes. Fig. 2 presents the methodology used to evaluate the proposed approach.

Table 1 shows the parameters used for each optimization technique employed in this work³. The c_1 and c_2 parameters of PSO control the pace during the particles movement, and the “Harmony Memory Considering Rate” (HMCR) of BHS stands for the amount of information that will be used from the artist's memory (songs that have been already composed) in order to compose a new harmony. In regard to BFA, α and β_0 are related to the step size of a firefly, and γ stands for the light absorption coefficient.

6. Experimental results

The experimental results stand for the mean accuracy and standard deviation over 25 rounds using the methodology presented in Section 5.3. Since the meta-heuristic algorithms are non-deterministic, we adopt this protocol to avoid biased results. The experiments were executed in a computer with a Pentium Intel Core i7® 1.73Ghz processor, 6 GB of RAM and Linux Ubuntu Desktop LTS 13.04 as the operational system.

Figs. 3 and 4 present the mean OPF accuracy over the three different feature sets (AR₅, AR₁₀ and AR₂₀), as well as the average number of selected channels, respectively. Notice the “yellow” bar stands for the standard OPF, i.e., without channel selection. From Fig. 3, one can observe there is not a relevant difference in terms of accuracy considering the different number of autoregression coefficients. As the coefficients are averaged at the output of each channel, such non-linear operation may have alleviated the influence of each approach. However, this operation seems to work well, since a recognition rate of around 86% is very competitive when compared to other works in the literature (Section 3).

Table 2 presents the percentage of selected EEG-channels. From the data, it is possible to observe three important points: (i) BGA and BHS have selected the lowest number of channels for all dataset configurations; (ii) considering the accuracy results shown in Fig. 3, we can conclude that we can achieve similar performance of that obtained using all the 64 channels by using less than a half

³ We have used the same variable notation for different methods because we believe it makes it easier to understand since it is the same notation used in the respective original papers.

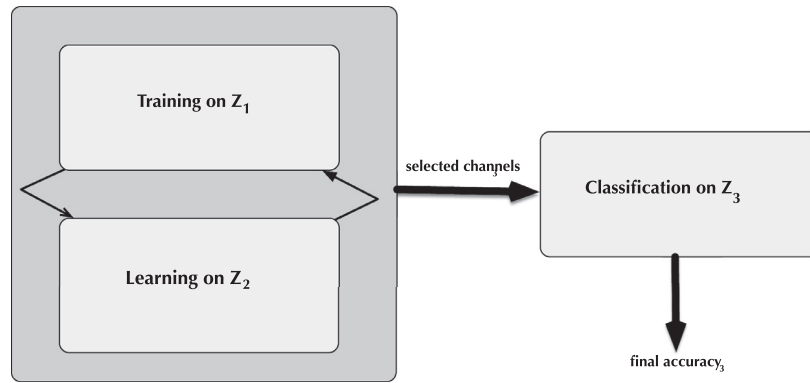


Fig. 2. Block diagram of the proposed approach.

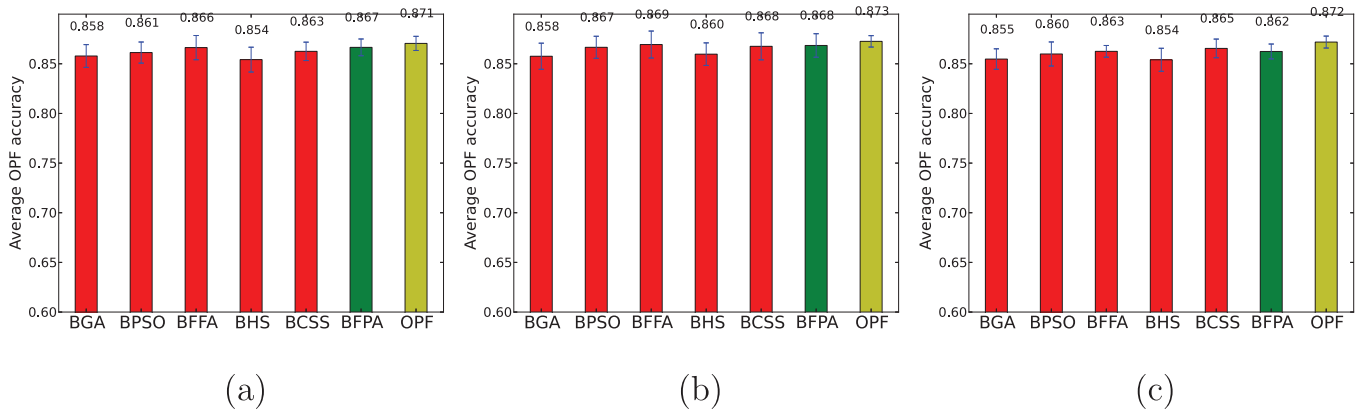


Fig. 3. Average OPF accuracy over (a) AR₅, (b) AR₁₀ and (c) AR₂₀ configurations.

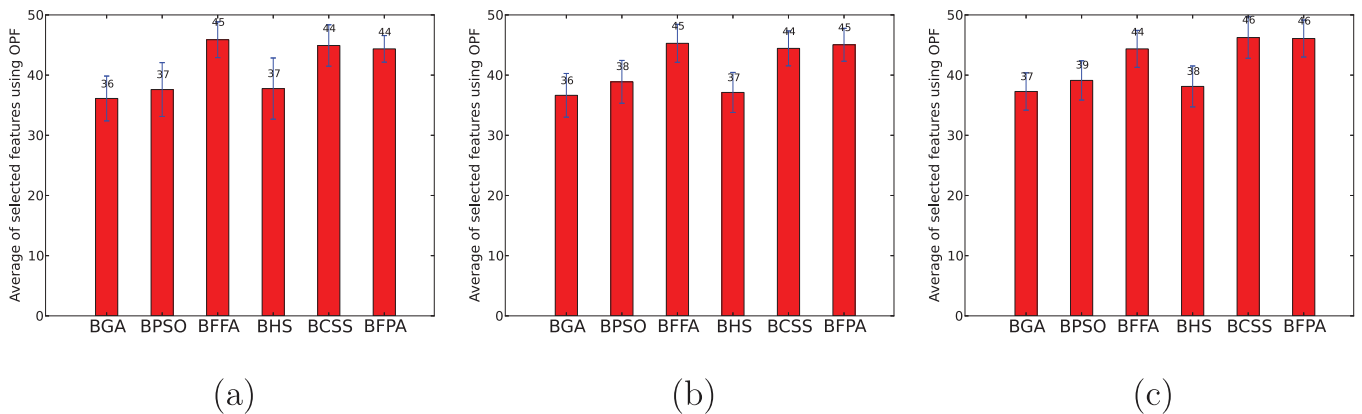


Fig. 4. Average number of selected channels of all techniques over (a) AR₅, (b) AR₁₀ and (c) AR₂₀ configurations. These values have been truncated for sake of simplicity.

Table 2
Percentual of selected EEG-channels.

Dataset	BGA	BPSO	BFA	BHS	BCSS	BFPA
AR ₅	36%	38%	45%	38%	44%	46%
AR ₁₀	36%	39%	44%	36%	45%	45%
AR ₂₀	37%	40%	44%	36%	44%	45%

of them; and (iii) the proposed BFPA has been very competitive in terms of binary-constrained optimization tasks when compared to the techniques addressed in this work.

Fig. 5 depicts the mean computational load (in seconds) for all optimization techniques regarding the learning step (dark gray module in Fig. 2). As we did not consider the feature extraction

procedure, i.e., the autoregression coefficients computation, the execution time over all dataset configurations are quite similar for each specific optimization technique. It is possible to observe BHS has been the fastest technique in all situations, since it only updates one agent per iteration. Although it may be a drawback in terms of convergence, it is still the fastest approach.

Finally, we performed the Wilcoxon signed-rank statistical test (Wilcoxon, 1945) to verify whether there is a significant difference between BFPA and the other techniques used in this work (considering the OPF recognition rate). Table 3 displays a pair-wise comparison against all techniques and BFPA, showing whether two techniques are considered similar ('=') or not ('≠') to each other. The only technique that has been considered similar to BFPA in all situations is BFA, followed by BPSO. An interesting point is related

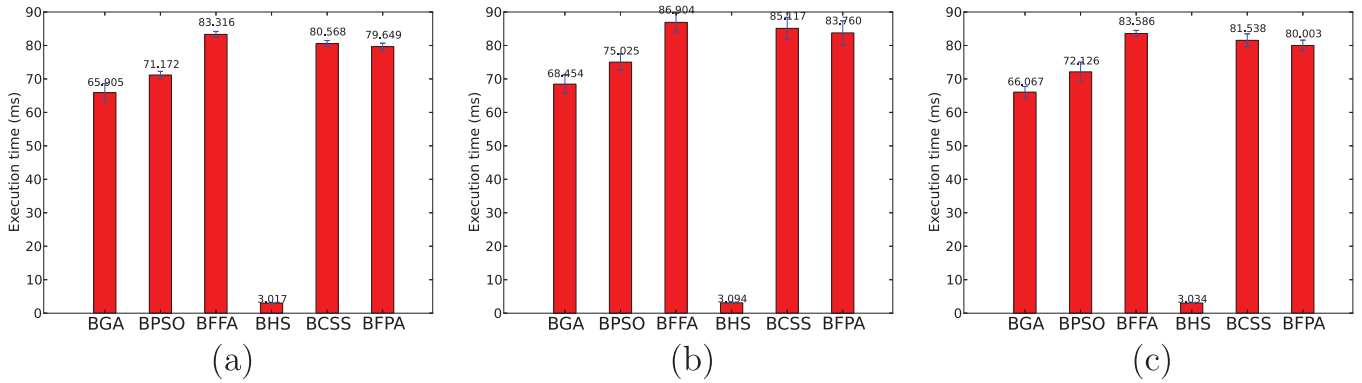


Fig. 5. Mean execution times of all techniques over (a) AR₅, (b) AR₁₀ and (c) AR₂₀ configurations.

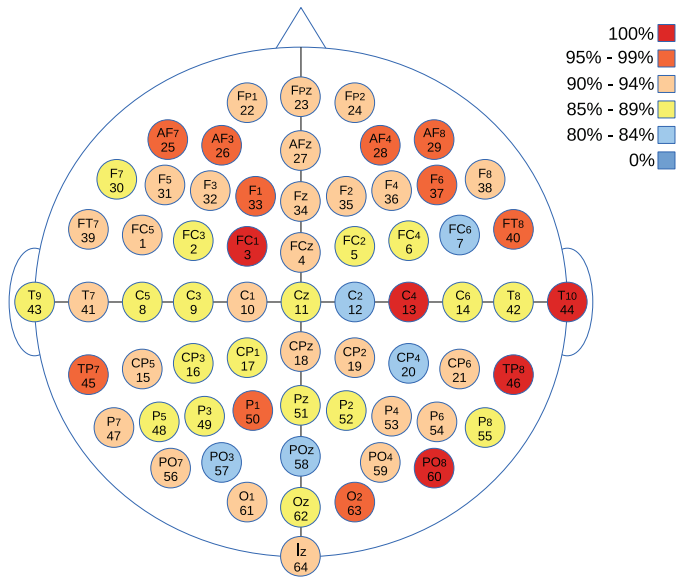


Fig. 6. Frequency of selected sensors during the experimental evaluation using AR₅ and BFFA.

Table 3
Wilcoxon signed-rank test evaluation.

Dataset	BGA	BPSO	BFA	BHS	BCSS
AR ₅	≠	=	=	≠	=
AR ₁₀	≠	=	=	≠	=
AR ₂₀	≠	=	=	≠	=

to the number of parameters, since BFFA requires only two, meanwhile BFA needs three parameters.

Since the nature of the proposed task in the EEG recording session has a close relation with different brain areas, like the movements of the hands and feet that mainly activates the central region of the brain (Wang, Gao, & Gao, 2005; Yang, Kyrgyzov, Wiart, & Bloch, 2013), it is important to figure out whether the expected channels are actually included in the subset selected by the optimization techniques. Therefore, since we executed a cross-validation procedure with 25 runnings, and due to the stochastic behaviour of the meta-heuristic techniques, this means a certain feature may not be selected at a given execution, and may be at another. In order to cope with this challenge, we opted to display the frequency of occurrence concerning each sensor, as displayed in Fig. 6. In this case, we considered BFFA with feature extraction by model AR₅.

Some interesting conclusions can be drawn if we consider the different range of frequencies modelled by distinct colours. It seems the frontal sensors are slightly more important than the back ones, since we can find more “yellow” and “blue” sensors right below the horizontal line (i.e., the one that goes from the left ear to the right one) than above that line. Another observation is that the “yellow” sensors are place everywhere, i.e., they correspond to the sensors that have been selected in between the range [85%, 89%], which is a considerable frequency. This means BFFA tried to select sensors placed at different positions of the brain in order to capture different information.

6.1. Transfer function analysis

In order to map the possible solutions (i.e., a position in the search space) from a continuous-valued space to a binary one, a transfer function needs to be employed (Mirjalili & Mohd Hashim, 2011; Rashedi, Nezamabadi-pour, & Saryazdi, 2010). A transfer function defines the probability of changing the position of a possible solution from 0 to 1 and vice-versa forcing the agents to move onto a binary space. Mirjalili and Lewis (2013) introduced a study of two families of transfer functions on binary-based PSO. Since the binary version of FPA makes use of a transfer function either, we also investigated these two different families of transfer functions (S-shaped and V-shaped) on Binary FPA. In short, we evaluated 8 transfer functions, as follows:

- S-shaped: S1, S2, S3 and S4; and
- V-shaped: V1, V2, V3 and V4.

Notice the transfer function S2 is the same one used in the experiments conducted in the previous section (Eqs. (5) and (6)). In this section, we just reproduced the results obtained with S2. For a more detailed explanation about the functions employed in this section, the reader can refer to the work by Mirjalili and Mohd Hashim (2011); Rashedi et al. (2010).

First of all, we evaluated the convergence of all transfer functions considering the AR models used in this work. Fig. 7 displays this experiment, in which transfer function S1 obtained the best results in all AR models, followed by S2 and V1. According to Mirjalili and Lewis (2013), the larger the velocity of a given particle, the highest it should be the probability to change its position from 1 to 0 and vice-versa, since this particle probably is far away from the best global solution. In this context, the “most abrupt” transfer functions are S1 and V1, i.e., they are more prone to switch the binary values.

Following a similar behaviour to the ones obtained in the convergence-driven experiment, functions S1 and V1 provided very good recognition rates over the test set, as displayed in Fig. 8. Such behaviour can be observed for all AR models. Additionally,

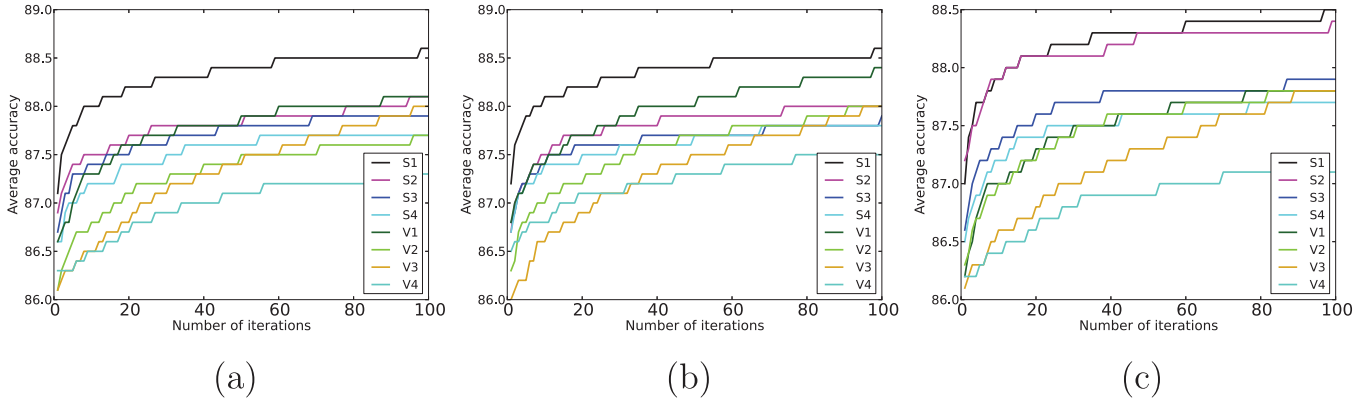


Fig. 7. Convergence evaluation of the transfer functions considering all AR models.

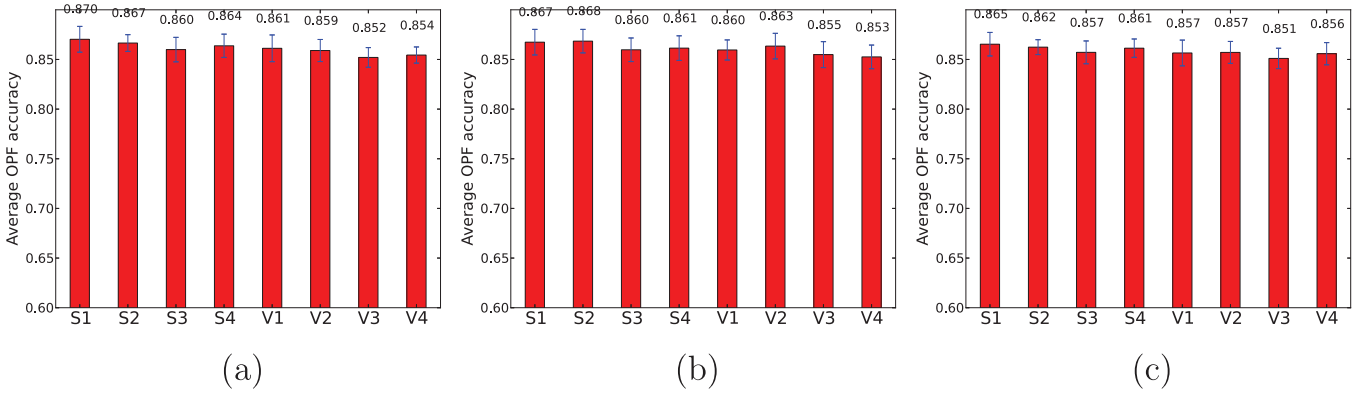


Fig. 8. Average OPF accuracy over (a) AR₅, (b) AR₁₀ and (c) AR₂₀ configurations considering different transfer functions.

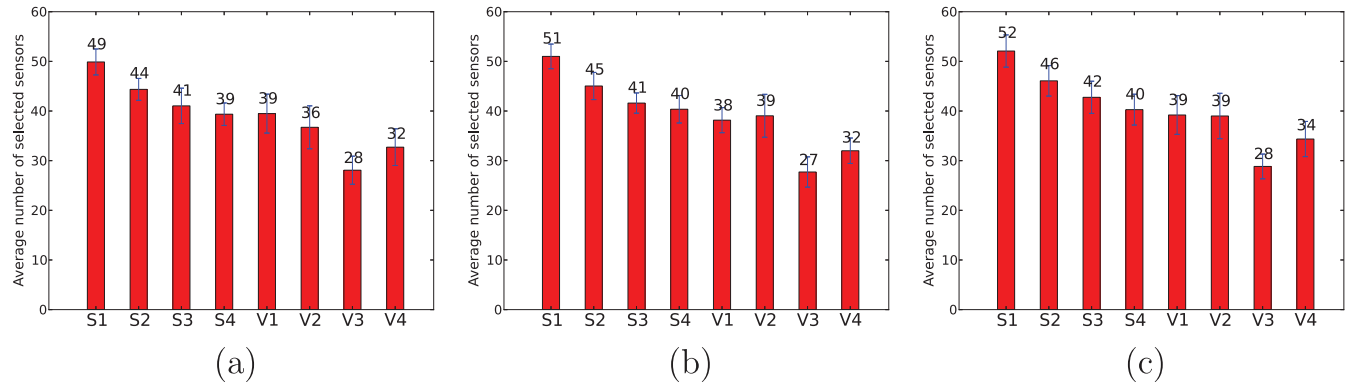


Fig. 9. Average number of selected channels of all techniques over (a) AR₅, (b) AR₁₀ and (c) AR₂₀ configurations considering different transfer functions. These values have been truncated for sake of simplicity.

the number of selected features can influence the recognition rates, as one can observe in Fig. 9. Although transfer function V3 has selected less features, it obtained the lowest recognition rates (Fig. 8), which is somehow expected. In regard to the computational load, Fig. 10 presents the mean execution time to learn the most representative subset of features. Since transfer function S1 has selected more features, it is expected a higher computational burden when compared to the others.

Table 4 displays the Wilcoxon signed-rank test considering the experiment with different transfer functions. Considering model AR₅, the most accurate techniques were S1, S2 and S4, and with respect to AR₁₀ we can highlight S1, S4 and V2 as the top-3 techniques. Finally, S1 and S4 obtained the best results considering the model AR₂₀.

6.2. Discussion

Roughly speaking, all techniques achieved similar recognition rates considering all AR models, with an advantage to BFPA and BFA, which are swarm-oriented. It is important to highlight one might obtain better recognition rates using a different feature extraction, but the main goal of this work is to evaluate BFPA in the context of sensor selection, as well as to show the importance of selecting sensors in order to make such approach less prone to errors and probably cheaper.

Using AR models with different number of coefficients seemed to does not provide different recognition rates, since the output of each AR model is given by the average of the coefficients. This could be a plausible explanation for that case. Such assump-

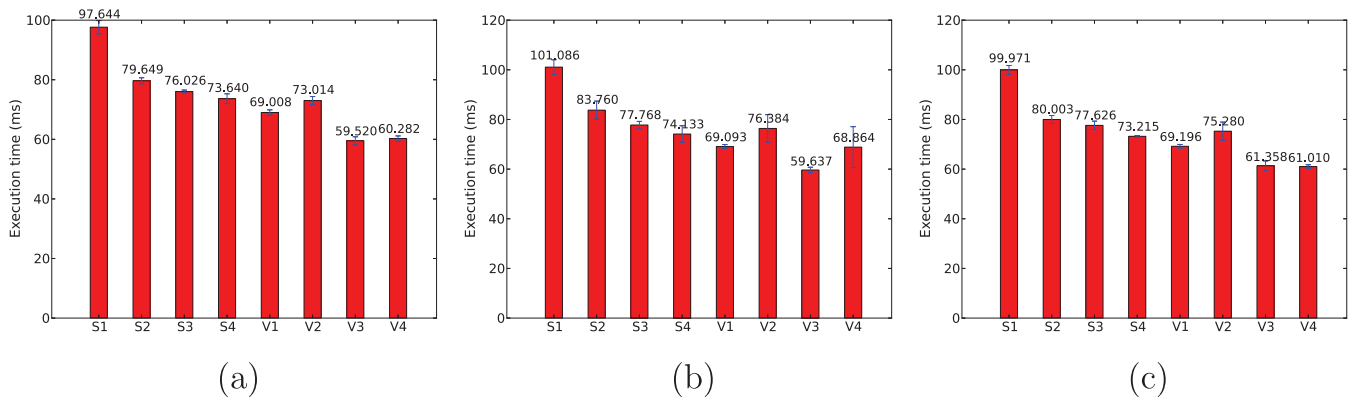


Fig. 10. Mean execution times of all techniques considering different transfer functions over (a) AR₅, (b) AR₁₀ and (c) AR₂₀ configurations.

Table 4

Wilcoxon signed-rank test computed between the transfer functions.

AR ₅	S1	S2	S3	S4	V1	V2	V3	V4
S1	–	=	≠	=	≠	≠	≠	≠
S2	=	–	≠	=	≠	≠	≠	≠
S3	≠	≠	–	=	=	=	≠	=
S4	=	=	=	–	=	=	≠	≠
V1	≠	=	=	=	–	=	≠	≠
V2	≠	≠	=	=	=	–	≠	=
V3	≠	≠	≠	≠	≠	≠	–	=
V4	≠	≠	=	≠	≠	=	=	–
AR ₁₀	S1	S2	S3	S4	V1	V2	V3	V4
S1	–	=	=	=	≠	=	≠	≠
S2	=	–	≠	=	≠	=	≠	≠
S3	=	≠	–	=	=	=	=	≠
S4	=	=	=	–	=	=	=	≠
V1	≠	≠	=	=	–	=	=	=
V2	=	=	=	=	=	–	≠	≠
V3	≠	≠	=	=	=	≠	–	=
V4	≠	≠	≠	≠	≠	≠	=	–
AR ₂₀	S1	S2	S3	S4	V1	V2	V3	V4
S1	–	=	≠	=	≠	≠	≠	≠
S2	=	–	=	=	=	≠	≠	=
S3	≠	=	–	=	=	=	=	=
S4	=	=	=	–	=	=	≠	=
V1	≠	=	=	=	–	=	=	=
V2	≠	≠	=	=	=	–	≠	=
V3	≠	≠	=	≠	=	≠	–	=
V4	≠	=	=	=	=	=	=	–

tion can be applied to all meta-heuristic techniques used in this paper.

Another important point concerns with the sensors selected by BFPA. A more detailed study showed the most frequent sensors are located in the front of the head, though they are also spread along the head. That is an interesting observation, which means BFPA tried to select sensors that are not so close to each other in order to capture relevant information from all places of the head.

Finally, an additional study with different transfer functions showed we can obtain different results, being the number of selected features strongly related to the final recognition rates. It seems the more features one has, the most accurate the transfer function. However, we still need to deal with a trade-off between the number of features and the computational efficiency. Using all sensors does not give us too much different results, which supports the idea of this work, that is to emphasize one can find out the subset of sensors that can obtain reasonable results.

7. Conclusions and future work

We have addressed the problem of channel selection in EEG-based biometric person identification. The goal of this work to highlight we may not need to employ all EEG channels available in order to obtain high identification rates. Therefore, we proposed to model the problem of channel selection as a meta-heuristic-based optimization task, in which the subset of channels that maximize the recognition rate over a validation set is used as the fitness function.

For the identification (classification) task, we have used the Optimum-Path Forest classifier, which has demonstrated to be similar to the state-of-the-art supervised pattern recognition techniques, but faster for training. In regard to the meta-heuristics, we have introduced a binary-constrained optimization version of the recently proposed Flower Pollination Algorithm, which seemed to be very competitive to other state-of-the-art optimization techniques employed in this paper: Binary Genetic Algorithm, Binary Particle Swarm Optimization, Binary Firefly Algorithm, Binary Harmony Search, and Binary Charged System Search.

The experimental results showed the BFPA outperformed many of the other methods, obtaining very good person identification rates using much less channels. It is important to emphasize that reducing EEG channels while keeping high identification rates is crucial towards the effective use of EEG in biometric applications. In addition, the selected sensors seemed to cover all the person's head, mainly in the front. Moreover, the number of coefficients in the AR model does not seem to impact in the final results, although we are taking the average of the coefficients as the final feature. Finally, different transfer functions were also analyzed, which allowed slightly better results.

Although using EEG data for biometric purposes seems to be a little bit far from reality in non-controlled environments, we would like to shed light over the importance in keep going with such studies, since good recognition rates can be obtained, being such sort of biometric approaches much less prone to spoofing attacks. Probably, in the future when mobile devices can be used to easily capture EEG signals, such techniques can be widely employed for biometric purposes as well.

Our future work will involve using modified versions of FPA to perform channel selection aiming at improving the overall identification performance while selecting fewer channels.

Acknowledgments

The authors are grateful to FAPESP grant #2014/16250-9, CNPq grants #470571/2013-6 and #306166/2014-3, and Capes grant.

References

- Abdullah, M. K., Subari, K. S., Loong, J. L. C., & Ahmad, N. N. (2010). Analysis of effective channel placement for an eeg-based biometric system. In *Proceedings of the IEEE conference on biomedical engineering and sciences* (pp. 303–306).
- Beijsterveldt, C., & Boomsma, D. (1994). Genetics of the human electroencephalogram (EEG) and event-related brain potentials (ERPs): A review. *Human Genetics*, 94(4), 319–330.
- Campisi, P., & La Rocca, D. (2014). Brain waves for automatic biometric-based user recognition. *Information Forensics and Security, IEEE Transactions on*, 9(5), 782–800. doi:10.1109/TIFS.2014.2308640.
- Falcon, R., Almeida, M., & Nayak, A. (2011). Fault identification with binary adaptive fireflies in parallel and distributed systems. In *Evolutionary computation (CEC), 2011 IEEE congress on* (pp. 1359–1366).
- Firpi, H. A., & Goodman, E. (2004). Swarmed feature selection. In *Proceedings of the 33rd applied imagery pattern recognition workshop* (pp. 112–118). Washington, DC, USA: IEEE Computer Society.
- Gasser, T., Jennen-Steinmetz, C., Sroka, L., Verleger, R., & Macks, J. (1988). Development of the eeg of school-age children and adolescents ii. topography. *Electroencephalography and clinical neurophysiology*, 69(2), 100–109.
- Geem, Z. W. (2009). *Music-inspired harmony search algorithm: Theory and applications* (1st ed.). Springer Publishing Company, Incorporated.
- Goldberger, A. L., Amaral, L. A. N., Glass, L., Hausdorff, J. M., Ivanov, P. C., Mark, R. G., et al. (2000). Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals. *Circulation*, 101(23), e215–e220.
- Guo, L., Rivero, D., Dorado, J., Munteanu, C. R., & Pazos, A. (2011). Automatic feature extraction using genetic programming: An application to epileptic EEG classification. *Expert Systems with Applications*, 38(8), 10425–10436.
- Holland, J. H. (1975). *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control and artificial intelligence*. Oxford, England: U Michigan Press.
- Jain, A. K., Ross, A. A., & Nandakumar, K. (2011). *Introduction to biometrics*. Springer.
- Jain, S., & Deshpande, G. (2004). Parametric modeling of brain signals. In *Biotechnology and bioinformatics, 2004. proceedings. technology for life: North carolina symposium on* (pp. 85–91).
- Jurcak, V., Tsuzuki, D., & Dan, I. (2007). 10/20, 10/10, and 10/5 systems revisited: Their validity as relative head-surface-based positioning systems. *NeuroImage*, 34(4), 1600–1611.
- Kaveh, A., & Talatahari, S. (2010). A novel heuristic optimization method: charged system search. *Acta Mechanica*, 213(3), 267–289.
- Kennedy, J., & Eberhart, R. C. (2001). *Swarm Intelligence*. M. Kaufman.
- Kostílek, M., & Št'astný, J. (2012). EEG biometric identification: Repeatability and influence of movement-related EEG. In *Proceedings of the international conference on applied electronics* (pp. 147–150).
- La Rocca, D., Campisi, P., Vegso, B., Cserti, P., Kozmann, G., Babiloni, F., et al. (2014). Human brain distinctiveness based on EEG spectral coherence connectivity. *Biomedical Engineering, IEEE Transactions on*, 61(9), 2406–2412. doi:10.1109/TBME.2014.2317881.
- Mirjalili, S., & Lewis, A. (2013). S-shaped versus v-shaped transfer functions for binary particle swarm optimization. *Swarm and Evolutionary Computation*, 9, 1–14.
- Mirjalili, S., & Mohd Hashim, S. Z. (2011). BMOA: Binary magnetic optimization algorithm. In *3rd IEEE international conference on machine learning and computing: vol. 1* (pp. 201–206). Singapore.
- Nunes, T. M., Coelho, A. L., Lima, C. A. M., Papa, J. P., & Albuquerque, V. H. C. (2014). EEG signal classification for epilepsy diagnosis via optimum path forest – a systematic assessment. *Neurocomputing*, 136, 103–123.
- Nuwer, M. R., Comi, G., Emerson, R., Fuglsang-Frederiksen, A., Guérit, J. M., Hinrichs, H., et al. (1998). IFCN standards for digital recording of clinical EEG. *Electroencephalography and Clinical Neurophysiology*, 106(3), 259–261.
- Ocak, H. (2009). Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy. *Expert Systems with Applications*, 36(2, Part 1), 2027–2036.
- Palaniappan, R. (2004). Method of identifying individuals using VEP signals and neural network. *IEE Proceedings - Science, Measurement and Technology*, 151(1), 16–20.
- Palaniappan, R., & Mandic, D. (2007). EEG based biometric framework for automatic identity verification. *The Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology*, 49(2), 243–250.
- Palit, S., Sinha, S. N., Molla, M. A., Khanra, A., & Kule, M. (2011). A cryptanalytic attack on the knapsack cryptosystem using binary firefly algorithm. In *Computer and communication technology (ICCCCT), 2011 2nd international conference on* (pp. 428–432).
- Papa, J. P., Falcão, A. X., Albuquerque, V. H. C., & Tavares, J. M. R. S. (2012). Efficient supervised optimum-path forest classification for large datasets. *Pattern Recognition*, 45(1), 512–520.
- Papa, J. P., Falcão, A. X., & Suzuki, C. T. N. (2009). Supervised pattern classification based on optimum-path forest. *Intl. Journal of Imaging Systems and Tech.*, 19(2), 120–131.
- Pollock, V., Schneider, L., & Lyness, S. (1991). Reliability of topographic quantitative EEG amplitude in healthy late-middle-aged and elderly subjects. *Electroencephalography and Clinical Neurophysiology*, 79(1), 20–26.
- Poulos, M., Rangoussi, M., & Alexandris, N. (1999). Neural network based person identification using EEG features. In *Proceedings of the IEEE international conference on acoustics, speech, and signal processing* (pp. 1117–1120).
- Poulos, M., Rangoussi, M., Chrissikopoulos, V., & Evangelou, A. (1999). Parametric person identification from the EEG using computational geometry. In *Proceedings of IEEE international conference on electronics, circuits and systems, 1999: vol. 2* (pp. 1005–1008).
- Ramos, C. C. O., Souza, A. N., Chiachia, G., Falcão, A. X., & Papa, J. P. (2011). A novel algorithm for feature selection using harmony search and its application for non-technical losses detection. *Computers & Electrical Engineering*, 37(6), 886–894.
- Rashedi, E., Nezamabadi-pour, H., & Saryazdi, S. (2010). BGSa: Binary gravitational search algorithm. *Natural Computing*, 9(3), 727–745.
- Rodrigues, D., Pereira, L. A. M., Papa, J. P., Ramos, C. C. O., Souza, A. N., & Papa, L. P. (2013). Optimizing feature selection through binary charged system search. In *Proceedings of 15th international conference on computer analysis of images and patterns* (pp. 377–384).
- Rodrigues, D., Yang, X.-S., Souza, A. N., & Papa, J. P. (2015). Binary flower pollination algorithm and its application to feature selection. In X.-S. Yang (Ed.), *Recent advances in swarm intelligence and evolutionary computation*. In *Studies in Computational Intelligence: vol. 585* (pp. 85–100). Springer International Publishing. doi:10.1007/978-3-319-13826-8_5.
- Safont, G., Salazar, A., Soriano, A., & Vergara, L. (2012). Combination of multiple detectors for eeg based biometric identification/authentication. In *Proceedings of the IEEE international Carnahan conference on security technology* (pp. 230–236).
- Sanei, S., & Chambers, J. (2007). *EEG signal processing*. Chichester, England; Hoboken, NJ: John Wiley & Sons.
- Schalk, G., McFarland, D., Hinterberger, T., Birbaumer, N., & Wolpaw, J. (2004). BCI2000: A general-purpose brain-computer interface (BCI) system. *IEEE Transactions on Biomedical Engineering*, 51(6), 1034–1043.
- Subasi, A. (2007). EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications*, 32(4), 1084–1093.
- Tau, G. Z., & Peterson, B. S. (2009). Normal development of brain circuits. *Neuropsychopharmacology*, 35(1), 147–168.
- Wang, Y., Gao, S., & Gao, X. (2005). Common spatial pattern method for channel selection in motor imagery based brain-computer interface. In *Engineering in medicine and biology 27th annual IEEE conference* (pp. 5392–5395).
- Wilcoxon, F. (1945). Individual comparisons by ranking methods. *Biometrics Bulletin*, 1(6), 80–83.
- Yang, X.-S. (2010). Firefly algorithm, stochastic test functions and design optimisation. *International Journal Bio-Inspired Computing*, 2(2), 78–84.
- Yang, X.-S. (2012). Flower pollination algorithm for global optimization. In *Proceedings of the 11th international conference on unconventional computation and natural computation*. In *UCNC'12* (pp. 240–249). Berlin, Heidelberg: Springer-Verlag. doi:10.1007/978-3-642-32894-7_27.
- Yang, Y., Kyrgyzov, O., Wiart, J., & Bloch, I. (2013). Subject-specific channel selection for classification of motor imagery electroencephalographic data. In *Acoustics, speech and signal processing (ICASP), 2013 IEEE international conference* (pp. 1159–1169).