

Automatic identification of epileptic EEG signals through binary magnetic optimization algorithms

Luís A. M. Pereira¹ · João P. Papa²  · André L. V. Coelho³ · Clodoaldo A. M. Lima⁴ · Danillo R. Pereira² · Victor Hugo C. de Albuquerque³

Received: 2 November 2016 / Accepted: 17 June 2017 / Published online: 28 June 2017
© The Natural Computing Applications Forum 2017

Abstract Epilepsy is a class of chronic neurological disorders characterized by transient and unexpected electrical disturbances of the brain. The automated analysis of the electroencephalogram (EEG) signal can be instrumental for the proper diagnosis of this mental condition. This work presents a systematic assessment of the performance of different variants of the binary magnetic optimization algorithm (BMOA), two of which are introduced here, while serving as feature selectors for epileptic EEG signal identification. In this context, the optimum-path forest classifier was adopted as a classification model, whereas different wavelet families were considered for EEG feature extraction. In order to

compare the performance of the improved BMOA variants against the traditional one, as well as other metaheuristic techniques, namely particle swarm optimization, binary bat algorithm, and genetic algorithm, we employed a well-known EEG benchmark dataset composed of five classes of EEG signals (two of which comprising normal patients with eyes open or closed, and the remaining comprising ill patients with different levels of epilepsy). Overall, the results evidenced the robustness of the proposed BMOA and its variants.

Keywords Feature selection · Epilepsy · EEG signal classification · Magnetic optimization algorithm · Metaheuristics · Optimum-path forest

✉ João P. Papa
papa@fc.unesp.br

Luís A. M. Pereira
luismartinspr@gmail.com

André L. V. Coelho
acoelho.albuquerque@unifor.br

Clodoaldo A. M. Lima
c.lima@usp.br

Danillo R. Pereira
danilopereira@unoeste.br

Victor Hugo C. de Albuquerque
victor.albuquerque@unifor.br

¹ Instituto de Computação, Universidade Estadual de Campinas, Campinas, SP, Brazil

² Departamento de Computação, UNESP - Univ Estadual Paulista, Bauru, SP, Brazil

³ Programa de Pós-Graduação em Informática Aplicada, Universidade de Fortaleza, Fortaleza, CE, Brazil

⁴ Escola de Artes, Ciências e Humanidades, Universidade de São Paulo, São Paulo, SP, Brazil

1 Introduction

Broadly speaking, epilepsy can be defined as a medical condition related to the occurrence of seizures, which affect a variety of mental and physical functions of an individual. In short, the term “epilepsy” encompasses a number of different neurological syndromes characterized by transient and unexpected electrical disturbances of the brain [4]. In epileptic patients, the brain’s normal electrical activity is disrupted by overactive electrical discharges, causing a temporary communication problem among nerve cells [3]. It is estimated that epilepsy is the third most common neurological disorder in the USA, being around 50–65 million people worldwide affected by such class of syndrome. Besides, the mortality rate is two to three times higher among people with epilepsy, which is fair enough for increasing the investments on novel methodologies and computational devices for the early and correct diagnosis of this medical condition.

One of the most reliable examinations for the proper diagnosis of seizures and epilepsy is the well-known electroencephalogram (EEG) [3], which records the brain's electrical activity as a series of traces, each of them corresponding to a different region of the brain. However, the visual inspection of the EEG signals for the detection of normal, interictal, and ictal activities in the patient's brain is usually a time-consuming and error-prone task due to the huge volumes of EEG segments that have to be analyzed. Therefore, the adoption of computer-based techniques for the purpose of tackling epilepsy diagnosis via EEG signal classification has been actively pursued in the last decades [8]. Besides, since EEG signals are nonlinear and dynamic in nature [1], there has been a growing interest in applying nonlinear signal analysis techniques, such as those based on wavelets, entropy, fractal, and chaos theory [37], for studying the behavior of these signals and also to extract relevant and condition-discriminatory information from them.

In order to assess the pros and cons of different machine learning approaches to cope with the epilepsy diagnosis problem, several prominent works have recently employed different configurations of the EEG dataset made available by Andrzejak et al. [1, 2]. This benchmark dataset is composed of five classes in total (two of which comprising normal patients with eyes open or closed, and the remaining comprising ill patients with different levels of epilepsy), whose full discrimination is very hard to achieve. In this context, Subasi [26–28] employed some variants of artificial neural networks (ANN) and also mixture-of-experts (ME) models aiming to discriminate between seizure and seizure-free profiles. In [27], in particular, the author reported 94.5% of accuracy rate achieved by ME models while discriminating solely classes A and E, which was a score better than that achieved by single multilayer perceptron (MLP) neural networks (93.2%). The specificity and sensitivity values reported for the ME and MLP models were, respectively, 94%/92.6% and 95%/93.6%. ME models induced with wavelet coefficients have also been considered by Übeyli [33], even though, in that work, the performance of the models was measured over three sets of the EEG dataset (namely sets A, D, and E). The total classification accuracy achieved by the ME network structures was 93.17% [33].

On the other hand, the paper of Tzallas et al. [31, 32] presents a methodology whereby selected segments of the EEG signals (maybe with different sizes) are analyzed using time–frequency methods, and then several features are extracted for each segment representing the energy distribution in the time–frequency plane. These features are used as input to a feedforward neural network, which provides the final classification. In order to evaluate the methodology, the authors generated four different

classification problems, none of which, however, involving the five classes at the same time, and the results achieved in terms of overall accuracy ranged from 97.72 to 100%. Nunes et al. [21] carried out a simple application of the optimum-path forest classifier [23, 24] to diagnose patients with epilepsy via EEG signal classification using four types of wavelet functions for feature extraction, being the Coiflets as the most accurate ones.

Lima et al. [14–16] evaluated the potentials of several kernel-based learning machines, such as support vector machines (SVM) and relevance vector machines (RVM), in the task of automatic discrimination of epileptic from non-epileptic EEG signals. The performance levels obtained by the kernel machines were contrasted in terms of predictive accuracy, sensitivity to the kernel function/parameter value, and sensitivity to the type of features extracted from the signal. For this purpose, several types of features extracted from the EEG signal, including statistical values derived from the discrete wavelet transform, Lyapunov exponents, and combinations thereof, were considered. Overall, the results evidenced that all considered kernel machines were competitive in terms of accuracy, and the choice of the kernel function and parameter value, as well as the choice of the feature extractor, are really critical decisions to be taken into account.

In this paper, we focus our attention on one specific step of the whole classification process that was not deeply investigated in the aforementioned works, i.e., the step of selecting the optimal subset of discriminatory features extracted from the EEG signal. In a nutshell, feature selection, also known as variable or attribute selection, is the task of selecting a subset of relevant features for inducing a classifier model [9, 10]. The central assumption when using a feature selection technique is that the data contain many redundant or irrelevant features. While redundant features are those which provide no more information than the currently selected features, irrelevant features provide no useful information at all. Even though the theme of feature selection has been much researched in the last years, it is noticeable that only a few works have given some attention to the study of the impact of this step in the context of EEG signal classification.

The paper of Ocaik [22], for instance, is an exception, where the use of a genetic algorithm-based (GA) EEG feature selector was investigated. In the proposed scheme, normal and epileptic EEG segments were decomposed into various frequency bands through a wavelet packet decomposition. Then, approximate entropy values of the wavelet coefficients at all nodes of the decomposition tree were used as candidate features to characterize the predictability of the EEG data within the corresponding frequency bands. Finally, the GA was used to find the subset of features that maximizes the classification performance

of an EEG classifier based on learning vector quantization (LVQ). It was particularly demonstrated in [22] that, if the GA was not used for the optimal feature selection, the good classification accuracies achieved by the LVQ classifier would drop noticeably.

In this paper, our emphasis is on the investigation of the potentials of a recently introduced population-based metaheuristic technique, named as magnetic optimization algorithm (MOA) [30], to serve as selector of optimal EEG features extracted by different wavelet families. Since the feature selection task is computationally intractable for even moderate sizes of feature sets [9, 10], the analysis of the performance of different metaheuristic algorithms for performing this task is readily justified [39]. Moreover, since the feature selection task can be regarded as a binary optimization problem, different variants of the binary magnetic optimization algorithm (BMOA) [18] have been considered in this study, two of which are introduced here. We compared MOA-based algorithms against GA, particle swarm optimization (PSO) [17, 38], and binary bat algorithm (BBA) [19], being the experiments conducted over the aforementioned EEG benchmark dataset.

The remainder of the paper is organized as follows. In Sect. 2, we outline the main steps behind the BMOA variants considered. Section 3 formalizes the steps of the proposed feature selection methodology, while Sect. 4 characterizes the EEG dataset and the wavelet basis used as feature extractors. Section 5 presents how the computational experiments were set up, while Sect. 6 is devoted to assess the performance of the techniques for EEG signal classification, taking into account the impact of the different feature selectors. Finally, Sect. 7 states conclusions.

2 Magnetic optimization algorithm

The electromagnetic force concept is one of the four fundamental interaction forces in nature. In this interaction force, the force intensity concerning two electromagnetic particles is inversely proportional to the distance between them, i.e., the greater the distance, the smaller the interaction force. Based on this definition, Tayarani and Akbarzadeh [30] proposed a new metaheuristic algorithm called magnetic optimization algorithm (MOA), which models a system of magnetic particles (agents) that seek for a solution in a search space using their magnetic fields, i.e., their fitness values, to interact with each other. The mathematical definitions of MOA are summarized as follows:

- Initially, MOA starts randomly placing all agents in the search space. Each agent is modeled as a solution

vector $\mathbf{x}_i \in \mathfrak{R}^d$, where i denotes the i -th agent, and x_i^d stands for its position at d -th dimension.

- At each iteration of the algorithm, the solution vectors are evaluated, and their respective fitness values are stored in B_i , which denotes the magnetic field value of the particle i .
- The mass M_i of each agent is given by:

$$M_i = \alpha + \rho B_i, \tag{1}$$

where α and ρ are constant parameter values.

- The interaction force between two particles i and j at dimension d is given as follows:

$$F_{ij}^d = B_i \frac{x_j^d - x_i^d}{D(\mathbf{x}_j, \mathbf{x}_i)}, \tag{2}$$

in which $D(\cdot, \cdot)$ is a distance function.

- The acceleration, velocity, and the position of each agent are updated, respectively, by:

$$a_i^d = \frac{F_i^d}{M_i}, \tag{3}$$

$$v_i^d(t + 1) = \theta v_i^d(t) + a_i^d \tag{4}$$

and

$$x_i^d(t + 1) = x_i^d(t) + v_i^d, \tag{5}$$

where t is the iteration step and $\theta \sim U(0, 1)$.

Tayarani and Akbarzadeh [30] also proposed a lattice where each agent can be influenced by the magnetic field from its neighborhood, being possible to determine the total force acting over each particle. However, this sort of lattice provides low and limited interactions, since an agent can interact with its immediate neighbors only (four neighborhoods) [18].

2.1 Binary MOA

Mirjalili and Hashim [18] proposed a binary version of the original MOA (BMOA) aiming to tackle binary optimization problems. In addition, they introduced a fully connected topology, in which all particles are connected and can interact to each other, thereby improving the shortcomings of the four-neighborhood lattice topology. Hereafter, we will refer to BMOA configured with a four-neighborhood lattice as BMOA₁, whereas BMOA₂ refers to the one with fully connected topology.

In order to restrict the new particle’s position to only binary values, the authors employed a hyperbolic tangent function [18]:

$$S(v_i^d(t)) = \lfloor \tanh(v_i^d(t)) \rfloor. \tag{6}$$

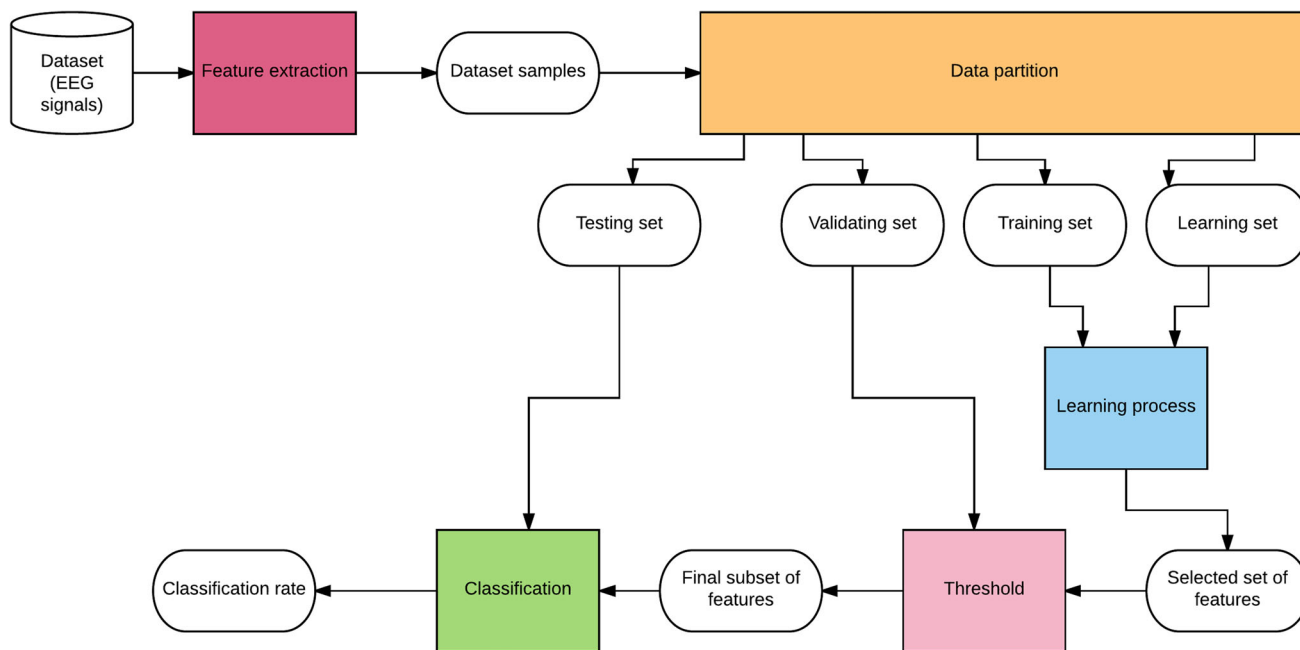


Fig. 1 Pipeline of the proposed feature selection methodology

Equation (6) can be rewritten as:

$$x_i^d(t + 1) = \begin{cases} \neg(x_i^d(t)) & \text{if } S(v_i^d(t + 1)) > \sigma, \\ x_i^d(t) & \text{otherwise} \end{cases} \quad (7)$$

in which $\neg(\cdot)$ means the binary complement operator and $\sigma \sim U(0, 1)$. To provide a good convergence rate, the velocity was limited to $|v_i^d(t + 1)| < v_{\max}$, where v_{\max} was set to 6.

2.2 Improving BMOA

Although BMOA₂ has demonstrated more interaction benefits than BMOA₁, the particles suffer from the attraction of bad ones, which may cause the loss of the previous good solution. As such, we propose here two new variants of the BMOA algorithm:

1. In the first variant, named as BMOA₃, we model the interaction between good and bad particles, so that only particles with a good magnetic field can attract particles with bad magnetic ones. Therefore, a particle i with a magnetic field B_i will attract a particle j only if $B_i > B_j$ (in case of maximization problems). Thus, the resultant force F_j over a bad particle j is formed by the attraction from particles with better magnetic fields than B_j .
2. The second variant, called BMOA₄, uses the same interaction strategy employed by BMOA₃; however, it allows that some good particles be attracted by some bad ones, as follows:

$$\frac{B_j - bestB}{B_j - B_i} > \sigma \text{ or } B_j > B_i. \quad (8)$$

However, these conditions may introduce the same problem as in BMOA₂, i.e., losing good solutions. In order to avoid this problem, we introduce a vector \mathbf{y} to store the best local position of each particle i . Thus, we update \mathbf{y}_i only if the new solution $\mathbf{x}_i(t + 1)$ achieves a better solution. These procedures are similar to those presented in [12], but for a different approach.

3 Feature selection methodology

In this section, we present the methodology used to evaluate the proposed variants of BMOA. The main idea is to allow a fair unbiased mean recognition rate computation together with a proper subset of suitable features. In order to accomplish with such deals, let us introduce some important definitions. Let \mathcal{Z} be a labeled dataset, such that $\mathcal{Z} = \mathcal{Z}_1 \cup \mathcal{Z}_2 \cup \mathcal{Z}_3 \cup \mathcal{Z}_4$, in which $\mathcal{Z}_1, \mathcal{Z}_2, \mathcal{Z}_3$, and \mathcal{Z}_4 stand for the training, learning, validating, and test sets, respectively. Roughly speaking, the main goal of a meta-heuristic-based feature selection approach is to employ some classifier’s recognition rate to be part of the fitness function (wrapper approaches). In this work, we use the training and learning sets to guide the search process onto the solution space (“Learning process” module in Fig. 1). Therefore, the idea is to train a classifier over \mathcal{Z}_1 for further

classification of Z_2 (for each search agent), being the recognition rate over the latter set used as the fitness function. As one can realize, we need a fast and effective classifier, since we need to perform the training step followed by the classification of the learning set every time an agent changes its position. Therefore, we opted to employ the supervised optimum-path forest (OPF) classifier [23, 24], which is a parameter-free technique that has been used for several applications.

The above procedure, which outputs the selected subset of features that maximizes the OPF accuracy over Z_2 , is then conducted 10 times with randomly generated training and learning sets. Thus, one has at the final of the process, 10 subsets of selected features, being now the main goal to choose the best one. Such step is conducted by the “Threshold” module in Fig. 1: we employed a threshold-based approach to find out the final subset of features, being such threshold value ranged from 10 to 90%, with steps of 10%. A threshold value of $T\%$, for instance, means we selected the features that appeared at least $T\%$ on that 10 subsets outputted over the 10 executions of the “Learning process” module in Fig. 1. The selected features for that threshold, i.e., $T\%$, are then used to train OPF for further classification of the validating set (Z_3). Therefore, if we perform the above procedure for each threshold within the range [10%, 90%], we obtain a curve that represents the recognition rate over Z_3 for each threshold value. The final subset of features is the one which maximizes the accuracy of such curve, being such subset used to train OPF for further classification of the unseen testing set

(Z_4). Notice the test set has not been used so far, i.e., it has been employed for assessing the effectiveness of the final subset of features only.

Figure 1 illustrates our proposed methodology to select the subset of features that best represents EEG signals. We used 30% of the original dataset for Z_1 , 20% for Z_2 , 20% for Z_3 , and 30% for Z_4 . These percentages were set up empirically.

We also compared the MOA-based approaches against a mutual information (MI) filter-based method [20]. The best MI model was chosen by selecting the features percentage among [10%, 20%, ..., 90%] with the highest mutual information. For this purpose, Z_1 was employed as the training set and $Z_2 \cup Z_3$ as the validating sets. Thereafter, the OPF classifier was trained on Z_1 to classify Z_4 .

4 Dataset description

In this work, we evaluate the performance of BMOA and its variants in the context of automatic epilepsy diagnosis. The complete dataset consists of five sets (denoted as A–E), which contains 100 single-channel EEG segments of 23.6s. These segments were selected and cut out from continuous multi-channel EEG recordings after visual inspection for artifacts due to muscle activity or eye movements. All EEG signals were recorded with the same 128-channel amplifier system using an average common reference. The data were digitized at 173.61 Hz sampling rate with 12 bit analog-to-digital resolution, and the band-pass filter settings were 0.5340 Hz (12 dB/oct). The data are made available by Andrzejak et al. [1, 2].

The signals from folds A and B were obtained extracranially from surface EEG recordings of five healthy individuals with eyes open and closed, respectively. Notice the sets C, D, and E were originated from an EEG archive of pre-surgical diagnosis. The EEG signals from five patients were selected, all of whom had achieved complete seizure control after resection of one of the hippocampal

Table 1 Parameter setting of the metaheuristic algorithms

Technique	Parameters
BBA	$\alpha = 0.9, \gamma = 0.9$
BGA	$pm = 0.1$
BMOA	$\alpha = 0.9, \rho = 4.8$
BPSO	$c_1 = 2.0, c_2 = 2.0, w = 0.9$

Table 2 Mean recognition rates considering OPF over the original (baseline) testing set

	Accuracy (%)	F-measure					Precision					Recall				
		A	B	C	D	E	A	B	C	D	E	A	B	C	D	E
		Coif2	64	63	70	55	39	92	55	74	57	42	93	73	67	53
Coif3	70	70	74	61	56	91	64	72	65	55	100	77	77	57	57	83
Coif4	68	58	74	67	45	95	56	78	59	52	97	60	70	77	40	93
Db2	61	69	60	43	41	85	65	70	42	46	76	73	53	43	37	97
Db3	60	65	51	46	48	87	56	71	48	45	84	77	40	43	50	90
Db4	60	72	51	59	42	71	64	62	47	56	86	83	43	80	33	60
Sym2	62	73	64	42	46	84	61	74	44	45	92	90	57	40	47	77
Sym3	59	66	58	41	48	81	65	64	41	44	83	67	53	40	53	80
Sym4	59	72	52	54	33	84	65	58	48	38	92	80	47	63	30	77

Table 3 Mean recognition rates and number of selected features considering BMOA variants over the testing set

	BMOA ₁		BMOA ₂		BMOA ₃		BMOA ₄	
	Acc (%)	#Selected features	Acc (%)	#Selected features	Acc (%)	#Selected features	Acc (%)	#Selected features
Coif2	71	16	71	24	69	39	69	20
Coif3	70	7	72	11	75	18	81	13
Coif4	76	31	71	4	76	34	75	24
Db2	67	24	61	22	73	8	67	10
Db3	69	11	73	11	69	12	69	15
Db4	69	16	72	16	71	26	72	14
Sym2	66	8	62	7	68	4	71	6
Sym3	67	14	58	20	61	22	69	10
Sym4	66	12	67	16	67	10	67	22

Bold values indicate the most accurate techniques

Table 4 Mean recognition rates and number of selected features considering BBA, BGA, BPSO, and MI on test set

	BBA		BGA		BPSO		MI	
	Acc (%)	#Selected features	Acc (%)	#Selected features	Acc (%)	#Selected features	Acc (%)	#Selected features
Coif2	69	20	72	16	66	18	72	32
Coif3	72	11	78	22	80	24	80	28
Coif4	78	10	76	33	75	31	72	16
Db2	69	33	67	6	67	15	64	8
Db3	64	25	75	16	66	15	68	8
Db4	64	4	68	15	68	8	60	8
Sym2	63	14	64	15	68	4	66	8
Sym3	65	14	54	21	69	11	62	16
Sym4	67	6	66	14	64	16	65	8

Bold values indicate the most accurate techniques

formations, which was therefore correctly diagnosed to be the epileptogenic zone. Signals in the folds C and D were sampled intracranially in seizure-free intervals from five patients. While the signals in fold C were captured from the hippocampal formation of the opposite hemisphere of the brain, those from fold D were extracted directly from the epileptogenic zone. Finally, fold E contains signals obtained intracranially and related to the seizure activity. These signals were selected from all recording sites of the brain exhibiting ictal activity. We consider here the whole dataset of 500 EEG segments, each one with 4096 samples, as employed in [34].

5 Experimental design

This section describes the main steps involved in our experimental procedures. We compared the MOA-based approaches against three other feature selection techniques: binary bat algorithm (BBA) [25], binary particle swarm

optimization algorithm (BPSO) [6], and binary genetic algorithm (BGA) [13].

- *Parameter setting*: Table 1 presents the parameters employed for each metaheuristic technique. Notice we used 30 agents with 100 iterations for all techniques. These parameters were set based on previous experiments.
- *Statistical evaluation*: In order to give more support for our conclusions, we carried out two round of statistical tests. Firstly, we performed the nonparametric Friedman test, which was used to rank the algorithms for each dataset separately. In case of Friedman test to provide meaningful results to reject the null-hypothesis (h_0 : all techniques are equivalent), then we can perform a post hoc test. For this purpose, we perform the Nemenyi test [20], which allows us to verify whether there is a critical difference (CD) among techniques. The results of the Nemenyi test can be represented in a simple diagram, in which the average ranks of the methods are plotted on an horizontal axis,

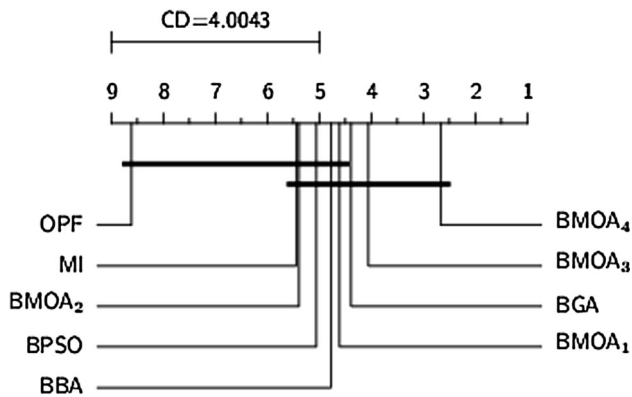


Fig. 2 Nemenyi statistical test considering the accuracy results

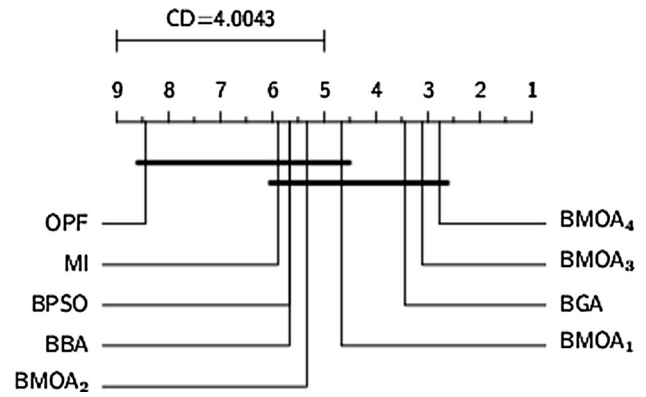


Fig. 4 Nemenyi statistical test considering the precision results

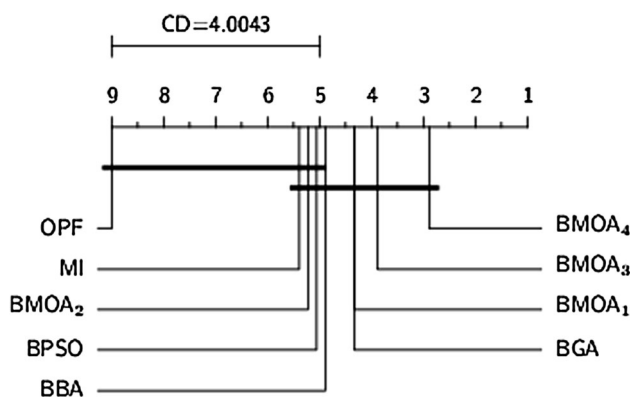


Fig. 3 Nemenyi statistical test considering the F-measure results

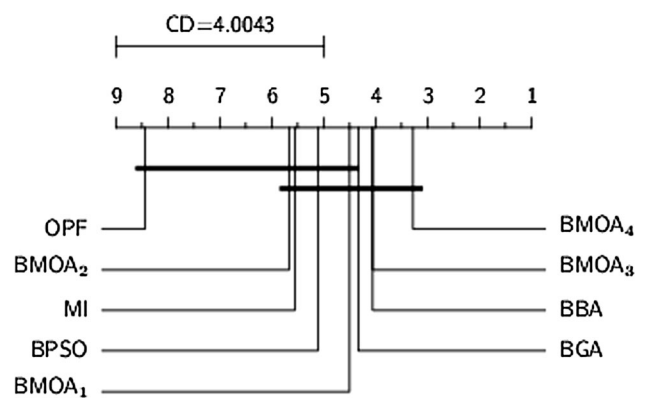


Fig. 5 Nemenyi statistical test considering the recall results

where the lower average rank is better. Furthermore, the groups with no significantly difference are then connected. More about these procedures can be found in Demšar [5].

- *Performance measures:* in order to assess the performance of the feature selection techniques, four well-known measures were employed: standard accuracy, F-measure, precision, and recall. Since we are dealing with a problem with multiple classes, the three latter measures were calculated for each class separately.
- *Fitness function:* as the reader may have noticed, our methodology (Sect. 3) requires a fast training and classification steps. In this fashion, we employed the OPF classifier, since it is a nonparametric and very robust classifier. Thus, for each iteration of the optimization techniques, the fitness function is calculated as the accuracy [24] of the OPF classifier on the learning set.
- *Platform:* it is important to highlight that all experiments were carried out on a PC Intel® Core i7 Q740 1.73GHz with 3GB RAM running Ubuntu 10.04 as operational system.

In order to extract discriminatory features from raw EEG data, the discrete wavelet transform (DWT) was employed in this work [29, 36]. The basic idea underlying wavelet analysis consists in expressing a signal as a linear combination of a set of localized functions, which are obtained by shifting, contracting, and dilating one particular prototype function, called a *mother wavelet* [11]. The decomposition of the signal leads to a set of values, referred to as wavelet coefficients.

While conducting the experiments for this paper, we have also considered different wavelet families with different orders and parametrization factors. However, due to the lack of space, we focus our analysis here on the Coiflets (Coif) order 2–4, the Symlet (Sym) order 2–4, and Daubechies (Db) order 2–4 [7, 14]. Therefore, 40 feature values were extracted from each of the 500 data patterns available in the dataset. The chosen features are related to the well-known statistics calculated over the wavelet coefficients in each or adjacent sub-bands, i.e., minimum, maximum, mean, standard deviation, power, absolute mean, and ratio of absolute mean [14, 27, 33, 35].

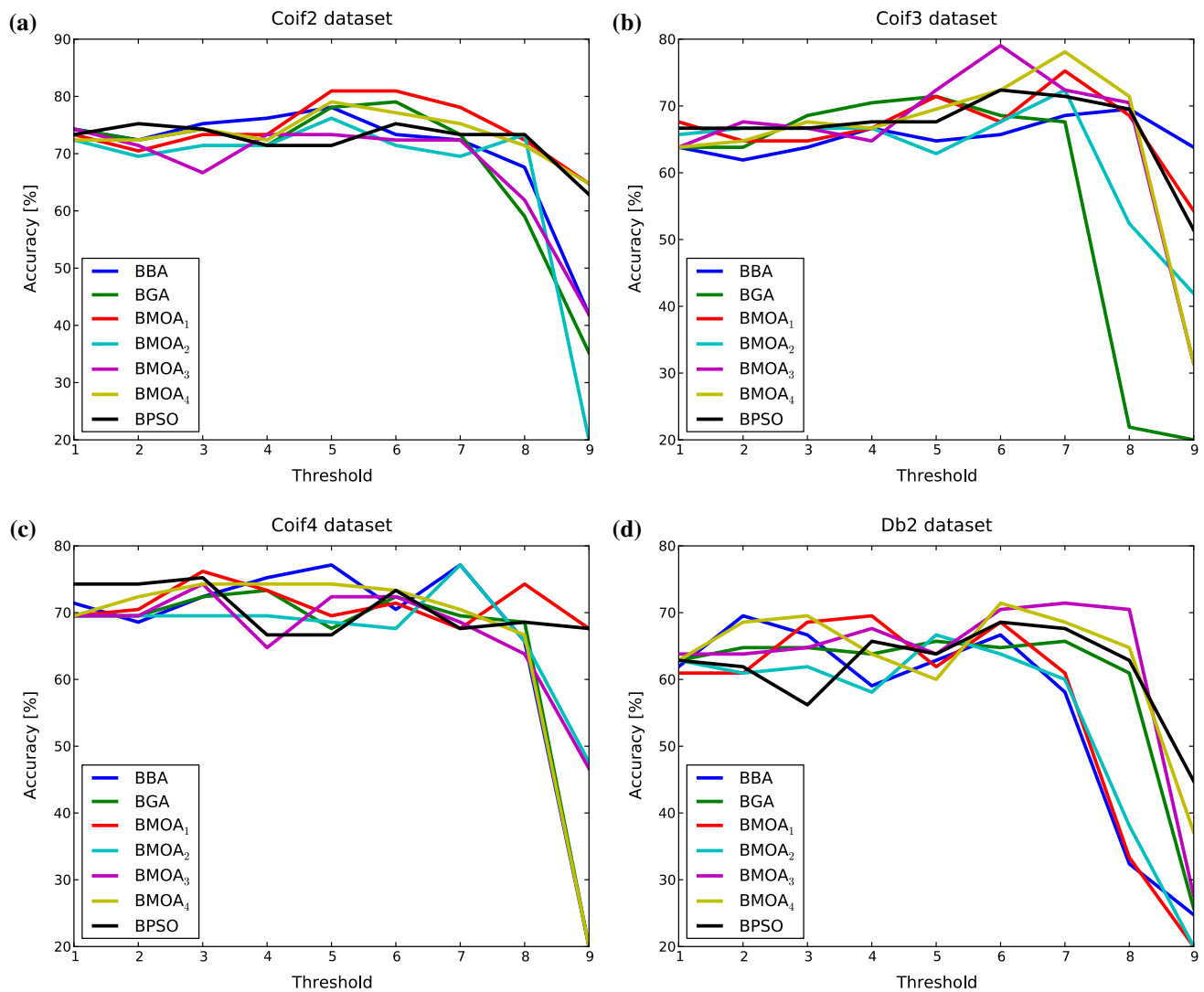


Fig. 6 Accuracy rates over the validating set considering Coif2, Coif3, Coif4, and Db2 datasets

6 Results and discussion

In this section, we present the results obtained using the proposed approaches. In order to provide a baseline for comparison purposes, we evaluated the performance of OPF classifier over the original datasets, i.e., without feature selection. For such experiment, we employed only the training and testing sets, since the learning and validating sets were used for feature selection purposes (Sect. 3). Notice the training and test sets were the same as the ones used in the feature learning process.

Table 2 shows the OPF classifier results over the original datasets (baseline), as well as Tables 3 and 4 display the recognition rates concerning the feature selection approaches. Notice that improvements on the accuracies after feature selection for all datasets can be observed. In regard to the Coif3 dataset, for instance, BMOA₁ achieved

the same results as the OPF classifier, but it has selected only seven features. The same behavior can be observed for Db2 and Sym2 datasets, in which BMOA₂ presented the same OPF results, but it has selected 22 and 7 features, respectively. If we consider the Sym3 dataset, BMOA₂ was the only technique that did not surpass the performance of the OPF classifier.

Figures 6 and 7 depict the curves generated by the “Threshold” module described in Fig. 1. Roughly speaking, one can observe that all techniques have presented similar behavior concerning variations on the threshold value. In addition, the great majority of the datasets have been better described with a threshold greater or equal than 50%, which means there might be an inferior bound for the feature selection problem. However, as the threshold increases, it does not imply the accuracy will also increase. Additionally, it is important to shed light over that BMOA

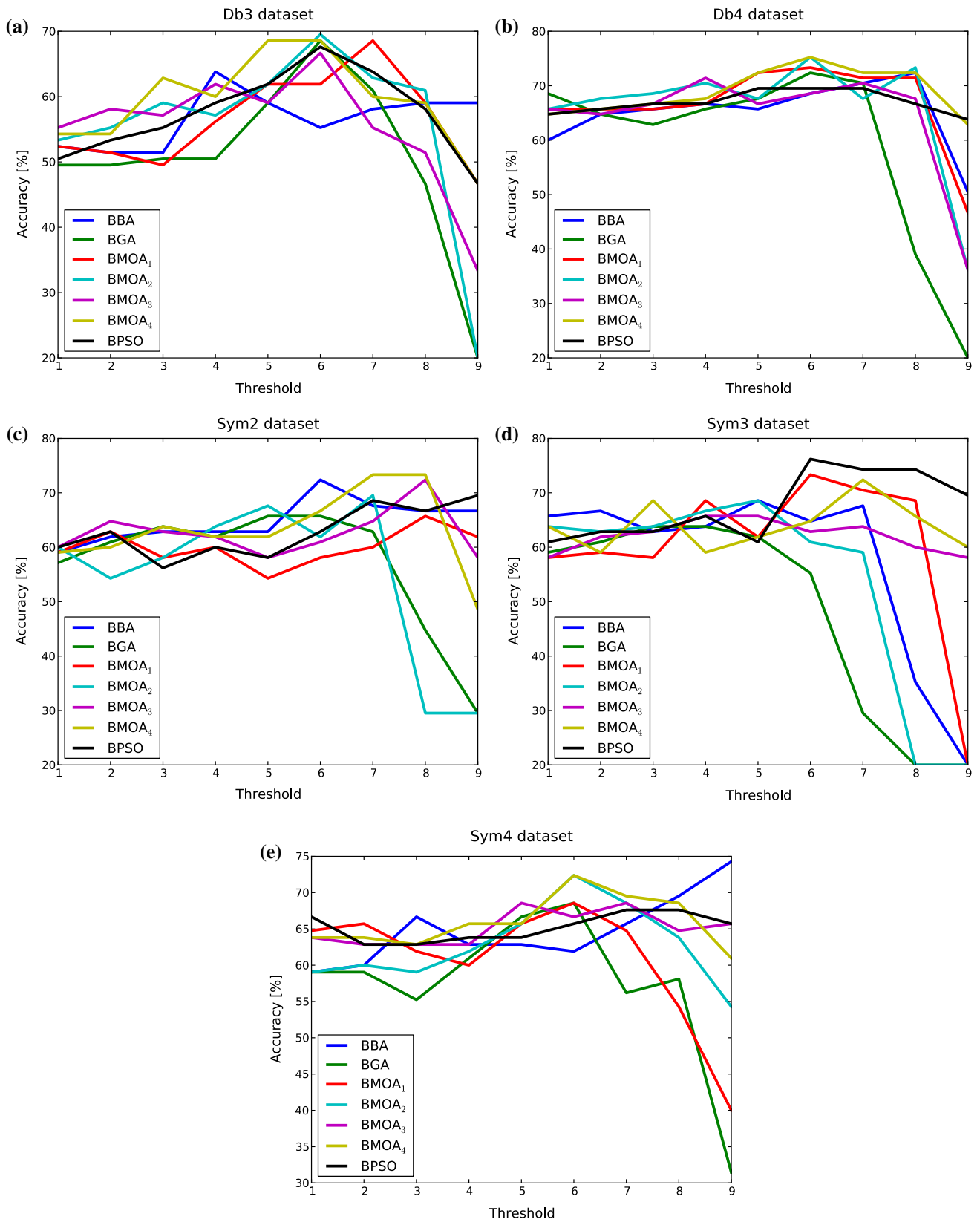


Fig. 7 Accuracy rates over the validating set considering Db3, Db4, Sym2, Sym3, and Sym4 datasets

Table 5 F-measure, precision and recall rates over the testing set considering BMOA variants

	BMOA ₁					BMOA ₂					BMOA ₃					BMOA ₄				
	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E
<i>F-measure</i>																				
Coif2	64	80	68	52	92	67	70	70	54	90	63	71	70	46	92	63	64	72	55	89
Coif3	79	77	58	50	88	68	69	70	63	90	71	75	65	67	95	83	84	75	67	95
Coif4	84	90	59	51	95	69	72	62	64	91	77	88	62	58	95	83	84	61	50	97
Db2	73	67	56	50	85	68	60	50	45	85	84	90	49	46	87	75	72	49	49	90
Db3	78	86	46	45	94	81	78	56	59	91	74	70	55	54	92	70	73	60	52	92
Db4	77	71	64	48	83	85	73	64	48	87	81	71	62	51	88	76	78	66	54	87
Sym2	73	66	51	48	90	73	59	49	40	87	81	77	46	49	86	77	80	54	56	85
Sym3	74	68	45	54	90	63	63	35	41	87	63	63	42	45	89	78	72	50	52	87
Sym4	67	71	59	50	84	68	68	65	53	84	67	72	57	46	93	65	66	65	57	85
<i>Precision</i>																				
Coif2	69	80	63	54	90	67	67	67	64	88	57	77	64	55	93	63	74	64	60	85
Coif3	76	81	59	47	93	66	76	70	61	90	69	76	72	64	94	77	89	76	67	97
Coif4	81	88	61	52	97	65	75	73	56	96	75	90	61	59	97	77	89	59	54	97
Db2	73	70	59	50	78	73	60	54	41	84	81	88	57	50	78	76	71	52	48	88
Db3	88	92	55	37	91	83	79	62	51	96	83	74	60	46	88	79	70	63	49	90
Db4	75	62	69	54	86	84	70	58	60	87	78	72	54	62	93	79	70	62	58	96
Sym2	70	62	56	50	90	70	61	52	40	84	83	75	50	46	89	71	74	64	59	84
Sym3	72	66	52	52	90	67	63	33	43	84	71	55	44	46	87	74	68	59	54	84
Sym4	67	69	58	50	89	69	66	60	56	89	70	71	51	48	96	62	68	58	65	86
<i>Recall</i>																				
Coif2	60	80	73	50	93	67	73	73	47	93	70	67	77	40	90	63	57	83	50	93
Coif3	83	73	57	53	83	70	63	70	67	90	73	73	60	70	97	90	80	73	67	93
Coif4	87	93	57	50	93	73	70	53	73	87	80	87	63	57	93	90	80	63	47	97
Db2	73	63	53	50	93	63	60	47	50	87	87	93	43	43	97	73	73	47	50	93
Db3	70	80	40	57	97	80	77	50	70	87	67	67	50	63	97	63	77	57	57	93
Db4	80	83	60	43	80	87	77	70	40	87	83	70	73	43	83	73	87	70	50	80
Sym2	77	70	47	47	90	77	57	47	40	90	80	80	43	53	83	83	87	47	53	87
Sym3	77	70	40	57	90	60	63	37	40	90	57	73	40	43	90	83	77	43	50	90
Sym4	67	73	60	50	80	67	70	70	50	80	63	73	63	43	90	67	63	73	50	83

Bold values indicate the most accurate techniques

variants obtained the best results in four out nine datasets, and also they achieved the same recognition rate as BPSO and BBA for Sym3 and Sym4 datasets, respectively.

If we consider Coif2 dataset, for instance, BMOA₁ was the best technique with 80.95% of accuracy (considering a threshold of 50%), as displayed in Fig. 6a. In addition, BMOA₃ selected the 60% of the features that maximized the classification rate for Coif3 dataset (Fig. 6b). Finally, for Coif4 dataset, BMOA₂ and BBA were the best performers with 77.14% of accuracy and a threshold equal to 70%. In case of Db datasets, for Db2, BMOA₃ and BMOA₄ achieved the same accuracy rates, but with a different threshold: 60 and 70%, respectively. For Db3 dataset, BMOA₂ achieved 69.52% of recognition rate considering a threshold of 60%. For Db4 dataset, BMOA₄ maximized the accuracy measure with a threshold of

60%, reaching 75.23%. Considering Sym2 dataset, BMOA₄ was the best performer achieving 73.33% of accuracy (threshold of 70%), and for Sym3 and Sym4, BPSO and BBA achieved the best accuracy rates of 76.19 and 74.28%, respectively.

From Tables 3 and 4, it is possible to observe the proposed BMOA₄ has been the most accurate technique in five out nine datasets, being them: Coif3, Db4, Sym2, Sym3 and Sym4. These results show us that the BMOA₄ interaction mechanism provides better convergence rates than the other BMOA variants. Among the other algorithms, BPSO was the best performer in four out of nine datasets. It also achieved a great result over the Coif3 dataset with accuracy equal to 80%, being slightly less accurate than BMOA₄. Nevertheless, BMOA₄ has selected less features than BPSO.

Table 6 F-measure, precision and recall rates over the testing set considering BBA, BGA, BPSO and MI techniques

	BBA					BGA					BPSO					MI				
	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E	A	B	C	D	E
<i>F-measure</i>																				
Coif2	66	73	63	50	90	64	76	72	56	90	64	67	64	45	88	62	88	62	67	93
Coif3	74	72	63	58	93	80	84	66	67	91	82	85	72	67	91	78	85	80	67	93
Coif4	73	82	68	68	98	73	84	73	57	94	80	90	60	50	94	74	86	65	49	97
Db2	70	68	58	51	95	72	77	58	40	86	65	74	54	47	90	60	70	60	50	78
Db4	68	60	55	50	90	85	88	48	57	95	70	71	52	41	94	68	85	59	38	91
Db3	75	79	51	27	81	88	73	57	38	79	80	77	57	39	83	65	56	45	50	91
Sym2	75	64	39	48	92	70	55	54	56	86	76	79	52	39	87	71	77	50	50	79
Sym3	75	74	47	38	85	62	50	36	46	78	78	70	55	55	83	58	62	50	50	88
Sym4	76	80	59	24	91	63	68	60	55	85	61	68	58	51	84	64	64	54	60	89
<i>Precision</i>																				
Coif2	70	73	58	56	87	73	76	61	65	87	62	73	62	48	84	87	70	70	47	90
Coif3	71	75	63	56	96	74	86	75	62	96	85	81	75	65	93	83	77	67	80	93
Coif4	73	78	71	71	96	69	86	69	61	96	78	90	60	51	96	87	80	37	67	93
Db2	74	69	53	56	91	75	71	59	40	89	72	69	52	52	88	70	70	50	40	93
Db3	73	60	53	47	93	84	93	54	51	94	74	66	54	43	91	77	73	57	37	97
Db4	74	77	48	33	76	82	80	50	50	76	80	75	51	48	83	87	47	67	30	70
Sym2	81	66	41	43	90	67	60	58	50	89	73	76	52	48	84	73	77	40	53	87
Sym3	71	69	47	45	86	64	45	40	43	88	74	67	60	57	83	60	77	50	33	93
Sym4	73	80	51	32	96	63	69	54	60	86	62	66	54	52	92	70	60	67	50	80
<i>Recall</i>																				
Coif2	63	73	70	46	93	56	76	86	50	93	66	63	66	43	93	72	78	66	55	92
Coif3	76	70	63	60	90	86	83	60	73	86	80	90	70	70	90	81	81	73	73	93
Coif4	73	86	66	66	100	76	83	76	53	93	83	90	60	50	93	80	83	47	56	95
Db2	67	67	63	47	100	70	83	57	40	83	60	80	57	43	93	65	70	55	44	85
Db3	63	60	57	53	87	87	83	43	63	97	67	77	50	40	97	72	79	58	37	94
Db4	77	80	53	23	87	93	67	67	30	83	80	80	63	33	83	74	51	54	38	79
Sym2	70	63	37	53	93	73	50	50	63	83	80	83	53	33	90	72	77	44	52	83
Sym3	80	80	47	33	83	60	57	33	50	70	83	73	50	53	83	59	69	50	40	90
Sym4	80	80	70	20	87	63	67	67	50	83	60	70	63	50	77	67	62	60	55	84

Bold values indicate the most accurate techniques

Figure 2 displays the statistical test concerning the accuracy results. Clearly, one can observe the proposed BMOA approaches (i.e., BMOA₄ and BMOA₃) have been placed as the two top best techniques (from right to left), though all techniques have been considered similar to each other, except the baseline provided by OPF (i.e., without feature selection).

Similarly, Figs. 3, 4, and 5 depict the statistical tests concerning the F-measure, precision and recall results. Notice all performance measures placed the proposed approaches as the best ones, though all being similar to each other concerning the statistical test, excepting the baseline provided by OPF. Roughly speaking, we can argue the proposed approaches are suitable for feature selection, and

the neighborhood information can really improve the results (Figs. 6, 7).

In regard to the *F – measure* results (Table 5), which is the harmonic average between precision and recall measures, BMOA variants have achieved the highest values for classes A, B, and C, since such classes are well separated by kernel machines in general (please, refer to [14]). The proposed BMOA₄ has been the one with the highest accuracy over class D, followed by BMOA₁ and BMOA₂ that achieved the best results over classes B (a tie with BPSO can be observed) and A, respectively. In addition, BBA has obtained the best accuracy considering class D. For the sake of comparison purposes, Table 6 displays the F-measure values concerning the techniques compared in this work.

7 Concluding remarks

In this work, we carried the problem of EEG signal classification by means of four variants of the magnetic optimization algorithm, being two of them proposed in this work. In addition, three well-known metaheuristic algorithms were considered in this study, namely particle swarm optimization, binary bat algorithm, and genetic algorithm.

The proposed BMOA₄ variant has prevailed in terms of effectiveness (accuracy, precision, recall, and F-measure) measures considering the great majority of datasets, as well as in terms of the number of selected features. In special, BMOA₄ recognition rate over the features extracted via Coif-3 wavelets has shown very satisfactory levels of performance (with accuracy equal to 81%). Besides, BMOA₄ has always prevailed over the other BMOA-based methods in terms of the discrimination power between classes C, D, and E.

It is also worth noting the main idea of this work is to show the importance in considering distinct neighborhood information when dealing with metaheuristic techniques. The proposed approaches were validated in the context of feature selection purposes concerning the task of epileptic identification by means of EEG signals. Although state-of-the-art results were not achieved, BMOA approaches seemed to be very much suitable to the problem, as well as they can also be applied to different other applications.

Acknowledgements LAMP and JPP are grateful to FAPESP Grants #2011/14094-1, #2009/16206-1, and #2014/16250-9, respectively, and also CNPq Grants #303182/2011-3, #470571/2013-6, and #306166/2014-3. The ALVC and CAML also acknowledge the sponsorship from CNPq via Grants #475406/2010-9, #304603/2012-0, 308816/2012-9, and #303182/2011-3. VHCA acknowledges CNPq for the Grants #470501/2013-8 and #301928/2014-2.

Compliance with ethical standards

Conflicts of interest The authors declare no conflict of interest.

References

- Andrzejak RG, Lehnertz K, Mormann F, Rieke C, David P, Elger CE (2001) Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. *Phys Rev E Stat Nonlinear Soft Matter Phys* 64:061907-1–061907-6
- Andrzejak RG, Widman G, Lehnertz K, Rieke C, David P, Elger CE (2001) The epileptic process as nonlinear deterministic dynamics in a stochastic environment: an evaluation on mesial temporal lobe epilepsy. *Epilepsy Res* 44:129–140
- Browne TR, Holmes GL (2003) *Handbook of Epilepsy*. Lippincott Williams & Wilkins, Philadelphia
- Chang BS, Lowenstein DH (2003) Epilepsy. *N Engl J Med* 349:1257–1266
- Demšar J (2006) Statistical comparisons of classifiers over multiple data sets. *J Mach Learn Res* 7:1–30
- Firpi HA, Goodman E (2004) Swarmed feature selection. In: *Proceedings of the 33rd applied imagery pattern recognition workshop*, IEEE Computer Society, Washington, DC, USA, pp 112–118
- Gandhi T, Panigrahi BK, Anan S (2011) A comparative study of wavelet families for EEG signal classification. *Neurocomputing* 74(17):3051–3057
- Gotman J (1982) Automatic recognition of epileptic seizures in the EEG. *Electroencephalogr Clin Neurophysiol* 54:530–540
- Guyon I, Elisseeff A (2003) An introduction to variable and feature selection. *J Mach Learn Res* 3:1157–1182
- Guyon I, Gunn S, Nikravesh M, Zadeh LA (2006) *Feature extraction: foundations and applications*. Springer, Berlin
- Hazarika N, Chen JZ, Tsoi AC, Sergejew A (1997) Classification of EEG signals using the wavelet transform. *Signal Process* 59:61–72
- Kaveh A, Talatahari S (2010) A novel heuristic optimization method: charged system search. *Acta Mech* 213(3):267–289
- Koza JR (1992) *Genetic programming: on the programming of computers by means of natural selection*. MIT Press, Cambridge
- Lima CAM, Coelho ALV (2011) Kernel machines for epilepsy diagnosis via EEG signal classification: a comparative study. *Artif Intell Med* 53:83–95
- Lima CAM, Coelho ALV, Chagas S (2009) Automatic EEG signal classification for epilepsy diagnosis with relevance vector machines. *Expert Syst Appl* 36:10054–10059
- Lima CAM, Coelho ALV, Eisencraft M (2010) Tackling EEG signal classification with least squares support vector machines: a sensitivity analysis study. *Comput Biol Med* 40:705–714
- Lin S-W, Ying K-C, Chen S-C, Lee Z-J (2008) Particle swarm optimization for parameter determination and feature selection of support vector machines. *Expert Syst Appl* 35(4):1817–1824
- Mirjalili S, Mohd Hashim SZ (2011) BMOA: binary magnetic optimization algorithm. In: *3rd IEEE international conference on machine learning and computing*, Singapore, vol 1, pp 201–206
- Nakamura RYM, Pereira LAM, Costa KA, Rodrigues D, Papa JP, Yang X-S (2012) BBA: a binary bat algorithm for feature selection. In: *Proceedings of the XXV SIBGRAPI—conference on graphics, patterns and images*, pp 291–297
- Nemenyi P (1963) *Distribution-free multiple comparisons*. Princeton University, Princeton
- Nunes TM, Coelho ALV, Lima CAM, Papa JP, de Albuquerque VHC (2014) EEG signal classification for epilepsy diagnosis via optimum path forest—a systematic assessment. *Neurocomputing* 136:103–123
- Ocak H (2008) Optimal classification of epileptic seizures in EEG using wavelet analysis and genetic algorithm. *Signal Process* 88:1858–1867
- Papa JP, Falcão AX, Albuquerque VHC, Tavares JMRS (2012) Efficient supervised optimum-path forest classification for large datasets. *Pattern Recognit* 45(1):512–520
- Papa JP, Falcão AX, Suzuki CTN (2009) Supervised pattern classification based on optimum-path forest. *Int J Imaging Syst Technol* 19(2):120–131
- Rodrigues D, Pereira LAM, Nakamura RYM, C KAP, Yang X-S, Souza AN, Papa JP (2014) A wrapper approach for feature selection based on bat algorithm and optimum-path forest. *Expert Syst Appl* 41(5):2250–2258
- Subasi A (2005) Epileptic seizure detection using dynamic wavelet network. *Expert Syst Appl* 28:701–711

27. Subasi A (2007) EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Syst Appl* 32:1084–1093
28. Subasi A, Ercelesi E (2005) Classification of EEG signals using neural network and logistic regression. *Comput Methods Programs Biomed* 78:87–99
29. Tang YY (2009) *Wavelet Theory Approach to Pattern Recognition*. World Scientific Publishing, Singapore
30. Tayarani MH, Akbarzadeh-Totonchi MR (2008) Magnetic optimization algorithms a new synthesis. In: *IEEE congress on evolutionary computation*, IEEE, pp 2659–2664
31. Tzallas AT, Tsipouras MG, Fotiadis DI (2007) Automatic seizure detection based on time-frequency analysis and artificial neural networks. *Comput Intell Neurosci* 2007:80510-1–80510-13
32. Tzallas AT, Tsipouras MG, Fotiadis DI (2009) Epileptic seizure detection in EEGs using time-frequency analysis. *IEEE Trans Inf Technol Biomed* 13:703–710
33. Übeyli ED (2008) Wavelet/mixture of experts network structure for EEG signals classification. *Expert Syst Appl* 34:1954–1962
34. Übeyli ED (2009) Combined neural network model employing wavelet coefficients for EEG signals classification. *Digit Signal Process* 19:297–308
35. Übeyli ED (2009) Statistics over features: EEG signals analysis. *Comput Biol Med* 39:733–741
36. Walnut DF (2004) *An introduction to wavelet analysis*. Birkhäuser, Basel
37. Willi-Hans S (2011) *The nonlinear workbook*, 5th edn. World Scientific, Singapore
38. Yang H, Du Q (2011) Particle swarm optimization-based dimensionality reduction for hyperspectral image classification. In: *IEEE international geoscience and remote sensing symposium*, pp 2357–2360
39. Yusta SC (2009) Different metaheuristic strategies to solve the feature selection problem. *Pattern Recognit Lett* 30:525–534