Neural Comput & Applic
DOI 10.1007/s00521-017-3003-y

ORIGINAL ARTICLE

# Epileptic seizure detection using hybrid machine learning methods

**Abdulhamit Subasi**[1] · **Jasmin Kevric**[2] · **M. Abdullah Canbaz**[3]

Received: 9 January 2016 / Accepted: 6 April 2017


**Abstract** The aim of this study is to establish a hybrid model for epileptic seizure detection with genetic algorithm (GA) and particle swarm optimization (PSO) to determine the optimum parameters of support vector machines (SVMs) for classification of EEG data. SVMs are one of the robust machine learning techniques and have been extensively used in many application areas. The kernel parameter's setting for SVMs in training process effects the classification accuracy. We used GA- and PSO-based approach to optimize the SVM parameters. Compared to the GA algorithm, the PSO-based approach significantly improves the classification accuracy. It is shown that the proposed Hybrid SVM can reach a classification accuracy of up to 99.38% for the EEG datasets. Hence, the proposed Hybrid SVM is an efficient tool for neuroscientists to detect epileptic seizure in EEG.

✉ Abdulhamit Subasi
absubasi@effatuniversity.edu.sa

Jasmin Kevric
jasmin.kevric@ibu.edu.ba

M. Abdullah Canbaz
mcanbaz@unr.edu

1 College of Engineering, Computer Science Department, Effat University, Jeddah 21478, Saudi Arabia

2 Faculty of Engineering and Information Technologies, Department of Electrical and Electronics Engineering, International Burch University, Francuske Revolucije bb. Ilidza, 71000 Sarajevo, Bosnia and Herzegovina

3 College of Engineering, Department of Computer Science and Engineering, University of Nevada, Reno, 1664 N. Virginia Street, Reno, NV 89557, USA

**Keywords** Discrete wavelet transform (DWT) · Electroencephalogram (EEG) · Epileptic seizure · Genetic algorithm (GA) · Particle swarm optimization (PSO) · Support vector machines (SVMs)

## 1 Introduction

The electroencephalogram (EEG) signal represents the overall electrical activity of the brain in a waveform arising from many neuronal activities. Non-invasive electrodes positioned on the scalp can record the electrical activity of the brain known as EEG. These signals keep significant information about the neurological conditions and other mental defectiveness [1]. Rhythmic sinusoidal activities can be recognized from the EEG signal. There are five frequency bands of the EEG signals which are usually used in analysis [2]:

- Delta (up to 4 Hz)
- Theta (4–8 Hz)
- Alpha (8–12 Hz)
- Beta (12–26 Hz)
- Gamma (26–100 Hz)

Over 50 million people in the world are affected by the neurological disorder known as epilepsy. Epilepsy is considered to be the second most common neurological disorder after stroke [3]. The brain cells' excessive electrical discharge resulting in abnormal movements and seizures are the signs of epilepsy. Health professionals separate non-seizure (normal) EEG signals from seizure (abnormal) EEG signals by characteristic patterns specified as "repetitive high-amplitude activities" with a combination of spikes and slow waves (SSW) [4].

A non-stationarity in the waveforms and semi-stationary time-dependent states also characterize the EEG signal.

Published online: 27 April 2017

Thus, detection of these characteristics is a difficult task. Power spectral analysis offers a quantitative measure of the frequency distribution of the EEG signals. Therefore, time-frequency algorithms such as discrete wavelet transform (DWT) are essential to find out different behavior of signals to describe it in the time and frequency domain [5–8]. Furthermore, the DWT is applied in the analysis of non-stationary signals such as EEG for feature extraction in particular [8]. Therefore, the DWT is suitable for localizing transient events which can occur as spikes during epileptic seizures [3, 9–12].

Since there is no apparent difference in EEG activity between epileptic and non-epileptic seizures, the visual detection of epileptic seizures from EEG signals is difficult [8]. Thus, automated detection techniques were tested to enable quicker and more precise detection of EEG recordings related to epileptic seizures. An automated analysis and detection of seizures using wavelet techniques to extract features from EEG signals were developed in [8]. The epileptic EEG segments showed more frequency changes than the non-epileptic EEG segments. There are researchers who applied a number of hybrid approaches to improve the detection of epileptic seizures. A fuzzy logic with genetic algorithms (GA) was applied in classification of epileptic and non-epileptic signals [13]. They calculated the risk factor by introducing various measures into a continuous and binary GA in seizure detection to provide accurate risk assessment. Another approach was a hybrid computational intelligence based on genetic algorithm for detection of electrode sites and features to predict optimum epileptic seizures [3]. Features extracted from pre-ictal and baseline data were used to train the genetic search algorithm, which was then validated on the remaining data sets. Time domain and frequency domain information was used as features. A DWT and a mixture of expert model were used for EEG signal classification in [11]. Peker et al. [14] proposed a new method for the diagnosis of epilepsy from electroencephalography (EEG) signals based on complex classifiers. They extracted the features of EEG data using a dual-tree complex wavelet transformation at different levels of granularity to obtain size reduction. Islam et al. [15] developed an automated seizure detection algorithm based on the stationary wavelet transform (SWT). Sharmila and Geethanjali [16] offered a framework for detecting the epileptic seizures from EEG data recorded from normal subjects and epileptic patients. This framework is based on a discrete wavelet transform (DWT) analysis of EEG signals using linear and nonlinear classifiers. The performance of the 14 different combinations of two-class epilepsy detection is studied using naïve Bayes (NB) and k-nearest neighbor (k-NN) classifiers for the derived statistical features from DWT. Wang et al. [17] used the partial directed coherence (PDC) analysis to extract the direction and intensity of information flow between brain regions. Hassan et al. [18] proposed a new automated epilepsy diagnosis scheme based on Tunable-Q factor wavelet transform (TQWT) and bootstrap aggregating (Bagging). In [19] a fusion method of variational mode decomposition (VMD)- and autoregression (AR)-based quadratic feature extraction was proposed for feature extraction and the random forest classifier was employed to hand with a three-classification task. Mursalin et al. [20] proposed a novel method for detecting epileptic seizure using Improved Correlation-based Feature Selection method (ICFS) with Random Forest (RF) classifier. Dhiman and Saini [21] decomposed EEG signal by using wavelet packet transform (WPT) and optimal selection of features by using biogeography based optimization (BBO). In [22] a novel method based on the dual-tree complex wavelet transform (DT-CWT) with the support vector machine (SVM) classifier is proposed for epileptic seizure detection.

In [12], DWT was used for feature extraction, principal component analysis (PCA), independent component analysis (ICA), and linear discriminant analysis (LDA) for dimension reduction and support vector machine (SVM) for EEG signal classification. Although the SVM is an efficient method for EEG signal classification, the parameter selection is crucial. Genetic algorithms (GAs) and particle swarm optimization (PSO) can be used to determine the optimum parameters of the SVM in EEG signal classification. The SVM kernel parameter selection in training process affects the classification accuracy. We applied the proposed GA- and PSO-based approaches to optimize the SVM parameters.

The rest of the paper is organized as follows: an introduction to an EEG dataset and signal processing methods used in this paper is given in Section 2. This section also describes the concept of GA, PSO, SVM, and Hybrid SVM. Section 3 presents the experimental results from using the proposed methods to classify EEG dataset. Section 4 summarizes the results and draws a general conclusion.

## 2 Materials and methods

### 2.1 Subjects and data recording

Publicly available data [23] were used in this paper. There are five sets (A–E), each containing 100 single-channel EEG segments, in the complete data set.[1] Sets A and B contain segments extracted from surface EEG recordings of five healthy volunteers in an awake state with eyes open (A) and eyes closed (B). The remaining three sets (sets C, D, and E) are taken from EEG archive of pre-surgical diagnosis from five different patients. Sets C and D contained only seizure-free activity, while set E was the only one that contained seizure activity. EEG signals were digitized at 173.61 Hz using 12-bit A/D converter. Since in the EEG, the important information is between 0 and 40 Hz

[1] EEG time series are available under (http://www.meb.unibonn.de/epileptologie/science/physik/eegdata.html).

frequency bands for epileptic seizure detection, the band-pass filter settings 0.53–40 Hz (12 dB/oct) were applied in the original dataset. In this paper, only two datasets (A and E) of the complete data were used, like in [11, 12].

### 2.2 Analysis using discrete wavelet transform

Time-frequency techniques like wavelet transform should be used in the characterization of EEG signals because of transitory and non-stationary characteristics of these signals. Continuous wavelet transform (CWT) is defined by the following equation:

$$W(a,b) = \int_{-\infty}^{\infty} x(t)\frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right)dt, \quad (1)$$

where $b$ translates the function across $x(t)$, while $a$ varies the time scale of the probing function $\psi(t)$. Analysis based on discrete wavelet transform (DWT) can be explained in terms of filter banks. *Sub-band filtering* defines the use of a group of filters to separate a signal into different signal components, and the procedure is called multi-resolution decomposition of a signal $x[n]$. Every stage of this procedure consists of two digital filters and two down-samplers by 2. The first filter $h[.]$ is a high-pass in nature. The second filter $g[.]$ is a low pass in nature. Outputs of the first high-pass and low-pass filters provide the detail D1 and the approximation A1 coefficients, respectively as seen in Fig. 1 [10–12, 24, 25].

In DWT analysis it is very important to choose the right number of decomposition levels and proper wavelet selection. The components of the dominant frequency of the signal are the main base for choosing the number of decomposition levels and correlate well with the frequencies required for classification of the signal.

DWT has been applied to EEG signals in time and frequency in such a way that EEG signals were decomposed into frequency sub-bands. Since sampling frequency is equal to 173.61 Hz, we chose the number of selected levels to be 5. Thus, we will have one final approximation A5 and details D1–D5. Daubechies 4 (DB4) wavelet filter was used to generate the detail and approximation records [10]. Using statistics over the discrete wavelet coefficient sets helped in decreasing the dimensionality of the extracted feature vectors [26]. The statistical features used in this study are the following:

1. Mean of the absolute of each sub-band coefficients values
2. Average power of the each sub-band of wavelet coefficients
3. Standard deviation each sub-band of the coefficients
4. Ratio of the absolute mean values of adjacent sub-bands

The first two statistical values represent the frequency distribution of the signal while the amount of changes in frequency distribution is represented by the last two statistical features. These features were derived for the approximation A5 and details D3–D5, and the resulting feature vectors were used to classify EEG signals.

### 2.3 Support vector machines

In 1992, Vapnik introduced SVMs as a new supervised machine learning method which could be applied to the famous function estimation problems [27]. SVMs are applied to many problems regarding text categorization [28], pattern recognition [29], and bio-informatics [30], and expanded to general nonlinear problems.

The main objective of classification is to train a computer to learn the nonlinear relationship between the features and their corresponding labels. In SVMs, the main objective is to separate the data into two classes by constructing a linear hyper-plane. The distance between the boundaries of the classes to this hyper-plane is called the margin. The maximization of the margin, which is the SVM's principal idea, gains better performance for a classifier. However, in nearly all real-life applications, datasets are not linearly separable. In this case, a nonlinear mapping from input space to higher dimensional feature space is defined in order to make the non-separable classes linearly separable as seen in Fig. 2.

This nonlinear mapping should be carried out by a nonlinear function which is defined by the use of a kernel function. Gaussian, radial basis function (RBF), anova, and polynomial kernels are some kernels that are suggested in the literature. Many machine learning algorithms require proper selection of their parameters to achieve optimal results. In the same manner, the SVM design requires the proper kernel parameters to accomplish the better classification accuracy. For example, the kernel function parameter like the gamma ($\gamma$) should be optimized for the radial basis function (RBF) kernel. The empirical trial-and-error approach to search for these values is not practical. SVM parameters can be automatically tuned by optimization algorithms similar to PSO or GAs. The values of parameters to be optimized are stored in one vector which is denoted as an individual. The aim of this research is to establish a hybrid model for epileptic seizure detection with GA and PSO finding the optimum parameters of the SVM. Detailed information related to the usage of kernels was given in [31, 32].

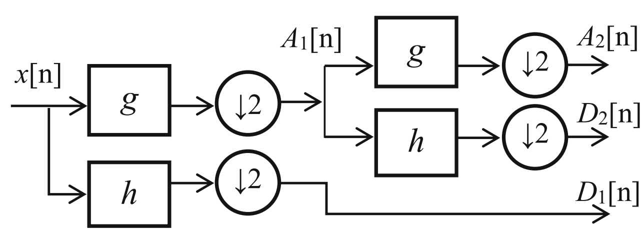


**Fig. 1** A filter bank at scale level 2

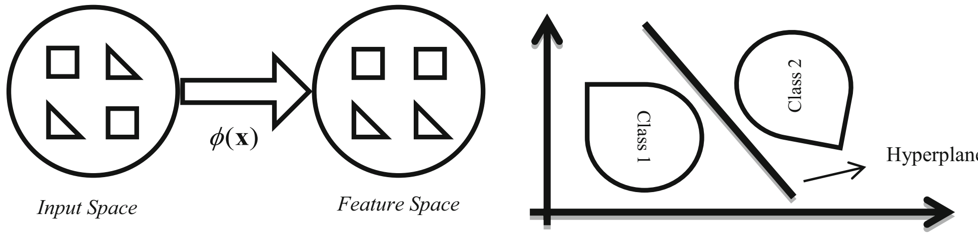


**Fig. 2** The nonlinear mapping $\varphi(\mathbf{x})$ makes inseparable points in input space separable in feature space

The closest points to the hyper-plane are referred to as non-bound support vectors. The SVM suffers from soft-margin classifier problem [33]. The trade-off between the number of misclassifications and generalization ability is controlled by the parameter $C$. Large values of $C$ decrease the number of training misclassifications but give inferior generalization capabilities. Thus, small values of $C$ result in more training errors.

### 2.4 Genetic algorithms

Genetic algorithm [34–36], modeled on the natural genetic system principles, represents adaptive, robust, and efficient optimization method. Pattern recognition, image processing, and machine learning are scientific fields where GA has a wide application area. GA starts with a set of individuals (population) which is usually generated randomly. Each individual in a population is represented as a single chromosome, whereas each chromosome comprises of a vector of elements called genes. In this study, parameters of the problem are SVM parameters which are encoded as a binary string of 0 and 1.

GA solves problems by optimizing a single criterion, known as a fitness function. The value of the fitness function represents the level of goodness of each (which is the vector of values of parameters to be optimized). To achieve better individuals (new population out of an old one), operations such as mutation, crossover, and selection are applied to the chromosomes.

Parent chromosomes responsible for reproduction are chosen within the selection procedure. Crossover is a method of exchanging data among parent chromosomes by recombining parts of their genetic materials. This operation joins parts of two parent chromosomes generating offspring for the next generation. The mutation process is described as an arbitrary modification in the genetic structure of a chromosome, establishing genetic diversity inside the population. More detailed information can be found in [37].

### 2.5 Particle swarm optimization

The PSO algorithm was introduced by Eberhart and Kennedy. It mimics the social behaviors of birds within a flock searching for food in an area. Only the individual that is nearest to the food knows where the food is located. Therefore, following that bird is the good strategy for other birds to find the food. Later, this concept evolved into a simple and efficient optimization algorithm [38].

Individuals, known as particles in PSO, fly through the multidimensional search space with a certain velocity (the dimension is determined by the number of parameters to be optimized) following the current optimum particle, which has the best fitness value calculated from the fitness function. As a result, the particles are anticipated to move towards better solution areas, thus discovering the optimal region [38, 39].

Each particle keeps a record of its previous best position called pbest, whereas the gbest (global best) represents the best position obtained so far by any particle in the entire population. Every particle may share information about the search space so that they can move in the direction of their pbest and the gbest by calculating their own velocity and updating their position in each iteration [40].

$$v_{id}^{t+1} = \chi\left(v_{id}^{t} + c_1 \cdot \psi_1 \cdot \left(p_{id}^{t} - x_{id}^{t}\right) + c_2 \cdot \psi_2 \cdot \left(p_{gd}^{t} - x_{id}^{t}\right)\right) \quad (2)$$

$$x_{id}^{t+1} = x_{id}^{t} + v_{id}^{t+1} \quad (3)$$

where.

| | |
|---|---|
| $v_{id}^{t}$ | $d$-component of the velocity of $i^{\text{th}}$ particle in iteration $t$ |
| $x_{id}^{t}$ | $d$-component of the position of $i^{\text{th}}$ particle in iteration $t$ |
| $c_1$, $c_2$ | constant weight factors |
| $p_i$ | best position of particle $i$ |
| $p_g$ | best position achieved by the neighbors of particle $i$ |
| $\psi_1$, $\psi_2$ | random factors in the [0, 1] interval |
| $\chi$ | constriction factor |

In PSO algorithm, each particle flying through the problem space with a certain velocity represents a potential solution. In analogy with GAs, a particle is similar to a chromosome and a swarm is similar to a population [39].

### 2.6 Hybrid SVM

The SVM has the regularization constant $C$ and the kernel-type parameters that influence its performance with respective parameters, and either default values are used or the values of parameters are defined manually by trial and error. The best approach is to adjust the values using optimization techniques such as GA or PSO, which results in GA-SVM or PSO-SVM algorithms.

PSO/GA searches for the best individual from $n$-randomly generated individuals. Each individual represents a possible solution in the form of an $m$-dimensional vector ($m$ is the number of parameters to be optimized). The individuals with high fitness values influence the next generation. Therefore, the fitness value should be properly defined as it will assure that every consecutive population of candidate individuals is usually better than its predecessor. On the other hand, the SVM classifier is built on some train set and its performance is evaluated on another dataset for each potential solution (individual). Hence, a single objective fitness function with classification accuracy representing the fitness is usually defined [40], so we used it in our study as well:

$$\text{Fitness} = \frac{\text{Good Classification}}{\text{Number of patterns}} \cdot 100 \tag{4}$$

This process repeats until the maximum number of iterations is reached.

We used PSO/GA to find optimal subset of parameters by detecting the best combinations of parameters. The Hybrid SVM approach is better explained using the following steps:

1. Data preparation: $T_r$, $V_a$ and $T_e$ represent training, validation, and test sets, respectively.
2. Initialization of population (generating initial individuals) and GA/PSO default parameters setting: setting iteration $t = 0$.
3. Setting iteration $t = t + 1$.
4. SVM model:

- Training and validation sets are pre-processed by extracting statistical features of DWT sub-bands.
- Training: conducting 10-fold cross validation (CV) on the training set $T_r$, and calculating the average CV accuracy for every (C, $\gamma$) for PSO. In GA, generating trained models based on the whole training set $T_r$ and every combination (C, $\gamma$).
- Evaluating the classification accuracy on validation set $V_a$ using the trained model based on the whole training set $T_r$ and every combination (C, $\gamma$).

5. Evaluating the fitness.
6. Checking the stop condition: If maximum iterations predefined are not reached, proceed with steps 7–8 and then go to step 3. Otherwise, go to step 9.
7. Updating the best individuals: Recording the average training CV accuracy and validation accuracy for the personal and global best (in PSO). In GA, this step indicates the selection of parents for the reproduction.
8. Particle movements (Eqs. 2 and 3) in PSO or application of genetic operators between the parents (crossover and mutation) in GA.
9. The validation accuracy curve is observed in order to avoid overtraining and the training is stopped when the iteration has the maximum validation accuracy.
10. Retraining and building the SVM classifier on the larger set $T_r + V_a$ based on the nominated feature subset and SVM model parameters. As a result, testing accuracy is measured on test set $T_e$ via the trained SVM classifier.
11. The training and testing procedure is terminated.

Advantages of the hybrid-SVM lay in combining the advantages of SVM's minimum structural risk and PSO's or GA's quick global optimizing ability [41]. The flowchart for the Hybrid-SVM, which accompanies the aforementioned steps, is shown in Fig. 3.

In the abovementioned steps and flowchart, a 10-fold cross-validation (CV) is mentioned under the SVM classifier training when PSO optimization is used. CV divides the training dataset into 10 independent subsets of equal size, where each subset will serve as the test dataset only once. During each run, nine subsets are used for training, while the remaining one-tenth is used for testing. The procedure repeats 10 times in a way that a different fold is used for testing in each run. The classifier's error rate is given by the medium of the error rates taken in each test fold and is called the cross-validation accuracy:

$$CVA = \frac{1}{k} \sum_{j=1}^{k} A_i \tag{5}$$

where $k$ is the number of folds used (10 in this case), and $A_i$ is the accuracy measure of each fold, $i = 1, \ldots, k$.

Further details about certain steps of the Hybrid SVM are presented in Section 3.

## 3 Experimental results

In this study, hybrid algorithms for epileptic seizure detection in EEG were proposed. We used DWT to extract feature vectors, and Hybrid SVM classifiers were used to classify EEG signals. A total set of features that characterize the EEG signals was created. Hybrid SVM detection was developed to accommodate these individual differences.

DWT has been applied to EEG signals in time and frequency in such a way that EEG signals were decomposed into frequency sub-bands. We have chosen the number of selected levels to be 5. Thus, we have one final approximation A5 and details D1–D5. Daubechies 4 (DB4) wavelet filter was used to generate the detail and approximation records [10]. Using statistics over the discrete wavelet coefficient sets helped in decreasing the dimensionality of the extracted feature vectors [26].

All input variables used with SVMs are normalized during the data pre-processing stage to avoid the domination of attributes in higher numerical ranges over those in lower ranges.

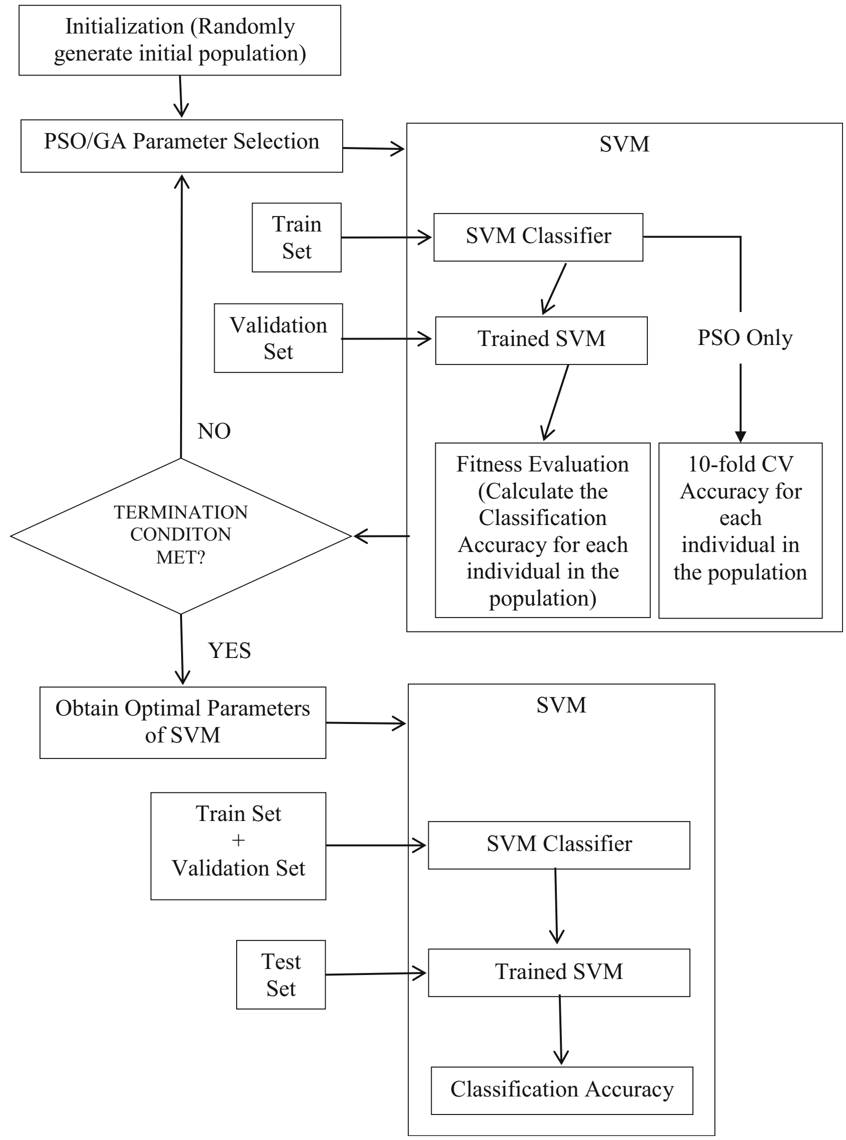


**Fig. 3** Flowchart of a hybrid SVM

Numerical difficulties during the calculation are prevented as well. Normalization of feature values can enhance the accuracy of the SVM classifier [42].

Before optimizing the SVM parameters for classification of EEG signals, the EEG dataset was additionally divided into train, validation, and test sets in the following ratio 50/25/25. For building the SVM model, the training set was used. To avoid an over-fitting to training data, the fitness of the classifiers is based on performance information measured in the validation sets to find the proper training iteration. For evaluating the classification accuracy of the model, the test set was used.

We applied a hybrid process of SVM parameter optimization through PSO/GA to classify EEG signals. The ANOVA kernel function for the SVM classifier is used for implementation of this proposed hybrid approach because it is suitable for high-dimensional data once $\gamma$ parameter is properly defined. Therefore, the values of the regularization constants $C$ and $\gamma$ are encoded in the particles of PSO and chromosomes of GA, so that their optimal values could be found. The initial particles of PSO and individuals of GA are generated randomly. The parameters were in the following intervals: $C$ [0.1, 1000] and $\gamma$ [0.001, 1].

In GA, these two parameters represent two genes which are encoded as binary strings within the chromosome. The SVM classifier is trained using the train set with the parameters specified in the chromosomes. Since the population size was set to 300, this stage results in 300 different trained SVM classifiers, one for each combination of parameters (chromosome). Once these trained SVM classifiers are tested using the validation set, the resulting classification accuracies will be used for fitness evaluation. More precisely, the best chromosomes in the population are the ones for which the highest classification accuracy has been achieved. The two best chromosomes (elitism strategy) are always preserved for the next generation, whereas two parents for reproduction (application of crossover and mutation) are chosen by roulette wheel method.

In PSO, $C$ and $\gamma$ can be considered as $x$- and $y$-coordinates of a position of a particle in 2D space represented via Eq. (3). However, a procedure of population update is slightly different than in GA. Namely, every individual ($C$, $\gamma$) stores its own personal and global best in terms of position which are determined according to fitness function (classification accuracy). Personal best for each individual is decided from the 10-fold cross-validation accuracy on the train set, whereas its global best is determined based on the classification accuracy on validation set (the SVM is trained on the train set, just like in GA). These stored sets of values for personal and global best of an individual are used in Eqs. (2) and (3) to update the velocity and position of that individual and hence to update the whole population. The population size for PSO is selected to be 70, almost 5 times lower than in GA.

Until the maximum number of iterations is reached (400 for PSO, and 30 for GA), the abovementioned procedure is repeated for every new population (updated set of SVM parameters). At the end, the individual that resulted in the highest classification accuracy on validation set is selected. In PSO, this means that the winner is the particle with the best global best of all particles. Having selected the optimal parameters, the SVM is now trained on *train* + *validation* set before being tested on the test set. Since the results of population-based optimization techniques, they were performed 10 times for each *training-validation-test* set partition of a dataset. The resulting classification accuracy values are represented in Table 1.

The number of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) are used to evaluate the performance of a classifier. The *sensitivity* and *specificity* are widely used in performance evaluation. Sensitivity is the percentage of people with disease, who have a positive test result,

$$\text{Sensitivity} = \frac{TP}{TP + FN} \text{ x } 100\% \tag{6}$$

Specificity is the percentage of people without disease who have a negative test result and defined as:

$$\text{Specificity} = \frac{TN}{TN + FP} \text{ x } 100\% \tag{7}$$

Accuracy is an overall measure for performance evaluation and defined as:

$$\text{Accuracy} = \frac{TP + TN}{TP + FN + TN + FP} \text{ x } 100\% \tag{8}$$

The performance of the SVM classifier and hybrid PSO-SVM and GA-SVM is verified on EEG datasets. Table 1 shows the performance of these classifiers using the statistical features of DWT as input, whereas Fig. 4 graphically represents the performance of the classifiers.

**Table 1** The values of statistical parameters of the SVM, GA-SVM, and PSO-SVM models for EEG signal classification

| Classification method | Accuracy | Specificity | Sensitivity |
|---|---|---|---|
| SVM (%) | 97.87 | 100 | 95.75 |
| GA-SVM (%) | 98.75 | 99.75 | 97.75 |
| PSO-SVM (%) | 99.38 | 99.25 | 99.5 |

## 4 Discussion

Even though the performance accuracy of the SVM and its hybrid approaches is comparable, hybrid classifiers require more time for execution than SVMs, at least for the training part. Once the optimal parameters are selected, hybrid approaches are as effective as the sole SVM when classifying test samples. The possible reason for this is the fact that hybrid classifiers perform parameter selection as an additional step prior to classification. When the range of parameters is high (in our study $C$ [0.1, 1000] and $\gamma$ [0.001, 1].), the hybrid approaches prove to be effective in optimizing the SVM parameters.

In terms of performance, PSO is more effective and computationally efficient than GA and converges at a faster rate. Implementation of basic PSO is quite simple since genetic operators like crossover or mutation are not necessary. On the other hand, PSO has weak local search abilities. It means that PSO can easily find where the optimum is located, but settling down at the optimum might take longer because of inappropriate particle's location update. That is why we selected 400 iterations for PSO (compared to 30 for GA).

The quality of attributes derived from the EEG depends upon the signal processing applied since they will be used as inputs to the classification model. DWT is selected because of its ability to cope with the non-stationary property of EEG signals. The developed Hybrid SVM models achieve an accuracy of up to 99.38% when classifying EEG signals using DWT statistical features, which makes them comparable to best existing tools in classification of EEG signals.

The following can be emphasized based on the results of this study:

- The high classification accuracy of the hybrid classifiers shows once again that statistical features of DWT coefficients the EEG signal give a good differentiation between the classes.
- Even though the original EEG signal features hold enough information for good classification, a proper nonlinear mapping to an arbitrarily high-dimensional feature space

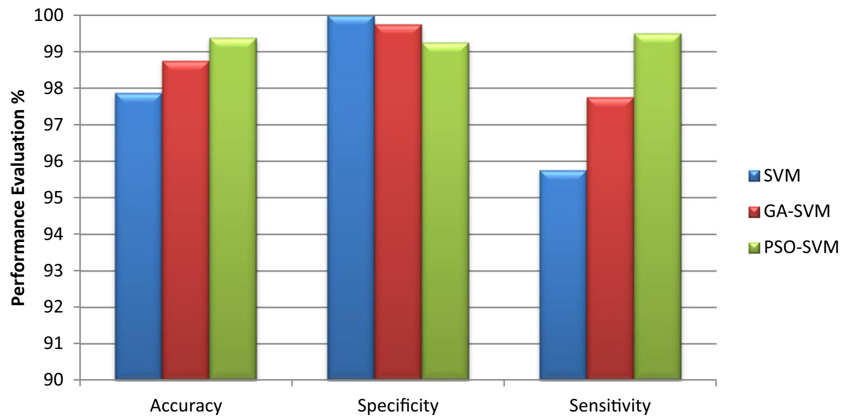


**Fig. 4** Graphical representation of evaluation performance of classifiers

by choosing appropriate kernel parameters can produce better discriminatory evidence that does not exist in the original feature space. We might choose whatever kernel function to use or this choice can be influenced by our knowledge of the problem data. The optimal values for kernel parameters are determined by using PSO and GA to automate this process.

- The classification results show that hybrid methods are comparable to the best existing tools in classification of EEG signals and can be used in classification of other non-stationary biomedical signals.
- PSO-SVM and GA-SVM based diagnostic system can be used in clinical studies as a real-time expert diagnostic system due to its automated nature.

## 5 Conclusions

In this study, we proposed two Hybrid SVM techniques that aim to optimize the performance of SVMs in the classification of EEG signals. The classification performance is increased by SVM parameter optimization carried out by PSO or GA. The $C$ and $\gamma$ parameters of SVM are automatically determined in this hybrid process prior to classification of new EEG signals. We decomposed EEG signals into time-frequency sub-bands using DWT and statistical features extracted from these DWT sub-bands. Experimental results demonstrate that the performance of PSO-SVM and GA-SVM are comparable, and they both outperform simple SVM. Although Hybrid SVM achieves high accuracy, the execution time is longer due to parameter selection process. Among the proposed hybrid classifiers, PSO-SVM slightly outperformed GA-SVM. The total classification accuracy of GA-SVM is 98.75%, whereas it is 99.38% for PSO-SVM. It was proved that the performance of SVM can be improved by GA's and PSO's parameter optimization, although more work may be required for decreasing the run time of hybrid classifiers.

**Compliance with ethical standards**

**Conflict of interest** The authors declare that they have no conflict of interest.

## References


1. Geva AB, Kerem DH (1999) Brain state identification and forecasting of acute pathology using unsupervised fuzzy clustering of EEG temporal patterns. In: Teodorescu HN, Kandel A, Jain LC (eds) Fuzzy and neuro-fuzzy Systems in Medicine. CRC Press, Boca Raton, FL
2. Bylsma FW, Peyser CE, Folstein SE, Ross C, Brandt J (1994) EEG power spectra in Huntington's disease: clinical and neuropsychological correlates. Neuropsychologia 32:137–150
3. D'Alessandro M, Esteller R, Vachtsevanos G, Hinson A, Echauz J, Litt B (2003) Epileptic seizure prediction using hybrid feature selection over multiple intracranial EEG electrode contacts: a report of four patients. IEEE Trans Biomed Eng 50:603–615
4. Geva AB, Kerem DH (1998) Forecasting generalized epileptic seizures from the EEG signal by wavelet analysis and dynamic unsupervised fuzzy clustering. IEEE Trans Biomed Eng 45:1205–1216
5. Preda RO, Vizireanu DN (2011) Robust wavelet-based video watermarking scheme for copyright protection using the human visual system. Journal of Electronic Imaging 20(1):013022–013022
6. Preda RO, Vizireanu DN (Dec. 2010) A robust digital watermarking scheme for video copyright protection in the wavelet domain. Measurement 43(10):1720–1726
7. Preda RO, Vizireanu DN (2011) Quantisation-based video watermarking in the wavelet domain with spatial and temporal redundancy. Int J Electron 98(3):393–405
8. Bigan C (1998) A recursive time-frequency processing method for neural networks recognition of EEG seizures. In: Ifeachor EC, Sperduti A, Starita A (eds) Neural networks and expert systems in medicine and healthcare. World Scientific, Singapore
9. Bronzino JD (2000) Principles of electroencephalography, *The biomedical engineering handbook*, Second edn. CRC Press LLC, Boca Raton
10. Adeli H, Zhou Z, Dadmehr N (2003) Analysis of EEG records in an epileptic patient using wavelet transform. J Neurosci Methods 123: 69–87
11. Subasi A (2007) EEG signal classification using wavelet feature extraction and a mixture of expert model. Expert Syst Appl 32: 1084–1093

12. Subasi A, Gürsoy MI (2010) Comparison of PCA, ICA and LDA in EEG signal classification using DWT and SVM. Expert Syst Appl 37:8659–8666
13. Harikumar, H., Narayanan, B. S (2003) Fuzzy techniques for classification of epilepsy risk level from EEG signals. Conference on Convergent Technologies for Asia-Pacific Region TENCON, p 209–213
14. Peker M, Sen B, Delen D (2016) A novel method for automated diagnosis of epilepsy using complex-valued classifiers. IEEE Journal of Biomedical and Health Informatics 20(1):108–118
15. Islam MK, Rastegarnia A, Yang Z (2016) A wavelet-based artifact reduction from scalp eeg for epileptic seizure detection. IEEE Journal of Biomedical and Health Informatics 20(5):1321–1332
16. Sharmila A, Geethanjali P (2016) DWT based detection of epileptic seizure from EEG signals using naive Bayes and k-NN classifiers. IEEE Access 4:7716–7727
17. Wang G, Sun Z, Tao R, Li K, Bao G, Yan X (2016) Epileptic seizure detection based on partial directed coherence analysis. IEEE J Biomed Health Inform 20(3):873–879
18. Hassan AR, Siuly S, Zhang Y (2016) Epileptic seizure detection in EEG signals using tunable-Q factor wavelet transform and bootstrap aggregating. Comput Methods Prog Biomed 137:247–259
19. Zhang T, Chen W, Li M (2017) AR based quadratic feature extraction in the VMD domain for the automated seizure detection of EEG using random forest classifier. Biomed. Signal Process. Control 31:550–559
20. Mursalin M, Zhang Y, Chen Y, Chawla NV (2017) Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier. Neurocomputing 241:204–214
21. Dhiman R, Saini J (2017) Biogeography based hybrid scheme for automatic detection of epileptic seizures from EEG signatures. Appl Soft Comput 51:116–129
22. Li M, Chen W, Zhang T (2017) Automatic epileptic EEG detection using DT-CWT-based non-linear features. Biomed Signal Process Control 34:114–125
23. Andrzejak, R.G., Lehnertz, K., Mormann, F., Rieke, C., David, P., Elger, C.E. (2001) Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. Physical Review E 64
24. Marchant BP (2003) Time–frequency analysis for biosystem engineering. Biosyst Eng 85:261–281
25. Semmlow JL (2004) Biosignal and biomedical image processing: MATLAB-based applications. Marcel Dekker, Inc., New York
26. Kandaswamy A, Kumar CS, Ramanathan RP, Jayaraman S, Malmurugan N (2004) Neural classification of lung sounds using wavelet coefficients. Comput Biol Med 34:523–537
27. Vapnik VN (2000) Nature of statistical learning theory, Second edn. Springer, New York
28. Joachims, T (1998) Text categorization with support vector machines. Proceedings of European conference on machine learning (ECML). Chemintz, DE, p 137–142
29. Pontil M, Verri A (1998) Support vector machines for 3D object recognition. IEEE Trans Pattern Anal Mach Intell 20:637–646
30. Yu, G. X., Ostrouchov, G., Geist, A., Samatova, N. F (2003) An SVM-based algorithm for identification of photosynthesis-specific genome features. Second IEEE computer society bioinformatics conference, CA, USA, pp. 235–243
31. Schlkopf B, Burges C, Smola A (1999) Advances in kernel methods: support vector learning. MIT Press, Cambridge, MA
32. Schlkopf B, Smola A (2002) Learning with kernels: support vector machines, regularization, optimization and beyond (adaptive computation and machine learning). MIT Press, Cambridge, MA
33. Cristianini N, Shawe-Taylor J (2000) An introduction to support vector machines: and other kernel-based learning methods. Cambridge University Press, New York
34. Davis L (ed) (1991) Handbook of genetic algotithms. Van Nostrand Reinhold, New York
35. Goldberg DE (1989) Genetic algorithms in search, optimization and machine learning. Addison-Wesley, New York
36. Michalewicz Z (1992) Genetic algorithms + data structures = evolution programs. Springer, Berlin Heideberg New York
37. Bandyopadhyay, S., Pal, S. K (2007) Classification and learning using genetic algorithms. Springer
38. Guo XC, Yang JH, Wu CG, Wang CY, Liang YC (2008) A novel LS-SVMs hyper-parameter selection based on particle swarm optimization. Neurocomputing 71:3211–3215
39. Engelbrecht, A. P. (2007) Computational Intelligence. Wiley
40. Huang C-L, Dun J-F (2008) A distributed PSO–SVM hybrid system with feature selection and parameter optimization. Appl Soft Comput 8:1381–1391
41. Ardjani, F., Sadouni, K (2010) Optimization of SVM multiclass by particle swarm (PSO-SVM). I J Modern Education and Computer Science, p 32–38
42. Hsu, C.-W., Chang, C.-C., Lin, C.-J. (2003) A practical guide to support vector classification