Australasian Physical & Engineering Sciences in Medicine
https://doi.org/10.1007/s13246-019-00793-y

SCIENTIFIC PAPER

Multi optimized SVM classifiers for motor imagery left and right hand movement identification

Kamel Mebarkia[1] · **Aicha Reffad**[2]

Received: 16 November 2018 / Accepted: 21 August 2019

Abstract
EEG signal can be a good alternative for disabled persons who cannot perform actions or perform them improperly. Brain computer interface (BCI) is an attractive technology which permits control and interaction with a computer or a machine using EEG signals. Brain task identification based on EEG signals is very difficult task and is still challenging researchers. In this paper, the motor imagery of left and right hand actions are identified using new features which are fed to a set of optimized SVM classifiers. Multi classifiers based classification showed having high faculty to improve the classification accuracy when using different kind or diversified features. Features selection was performed by genetic algorithm optimization. In single optimized SVM classifier, a mean classification accuracy of 89.8% was reached. To further improve the rate of classification, three SVMs classifiers have been suggested and optimized in order to find suitable features for each classifier. The three SVMs classifiers were optimized and achieved a performance mean of 94.11%. The achieved performance is a significant improvement comparing to the existing methods which does not exceed 81% while using the same database. Here, combining multi classifiers with selecting suitable features by optimization can be a good alternative for BCI applications.

Keywords Electroencephalography EEG signals · Motor imagery BCI (MI) · Features extraction · BCI system · SVM classifier · Optimization

Introduction

EEG signals reflect the neurons electrical activities within the brain and changes with respect to these brain activities and states. Depending on the brain states and activities, EEG signals behave differently in terms of spectral content. In fact, several rhythms (waves) of EEG signals are distinguishable and therefore can be used to distinguish brain activities [1]. These waves are referred to as delta (δ: 0–4 Hz), theta (θ: 4–8 Hz), alpha (α: 8–13 Hz), beta (β: 13–30 Hz), and gamma (γ: above 30 Hz) waves. Each band covers a range of certain tasks and mental states of the brain [2, 3]. In addition to this EEG characteristic, it is known that brain activities can be localized in specific regions of the brain referring to spatial information. Such information can be gained by the known international 10–20 system [4]. These two characteristics of EEG signals are principally used to recognize brain activities. As an example, features from power spectral and common spatial pattern (CSP) features have been used to recognize brain activities as consequence of these EEG characteristics [5–7]. This recognition plays an important role in brain computer interface (BCI) systems in which a disable person can perform actions just by thinking [2, 8, 9].

In the last two decades, BCI technology has become more and more attractive by researchers due to its important applications in different domains especially in the medical domain [10, 11]. There are principally three types of BCI systems, which use different phenomena: the steady state visual evoked potential (SSVEP), motor imagery (MI) and P300 event-related potential [9, 12, 13]. Motor imagery, in which subjects use imagined movement, has been widely used as a major approach in BCI studies. MI-based BCI, which is the subject of this study, is particularly suitable for disabled people who do not have control over their limbs to give them a certain degree of freedom. Among the factors that restrict the spread use of BCIs are: the EEG signals acquisition problems, the low speed of information conveyed

✉ Kamel Mebarkia
kamel.mebarkia@rwth-aachen.de

1 LIS Laboratory, Electronics Department, Faculty of Technology, Sétif 1 University, Sétif, Algeria

2 LAS Laboratory, Electrotechnics Department, Faculty of Technology, Sétif 1 University, Sétif, Algeria

Published online: 30 August 2019

by brain-computer technology and the inter-subject variability [1, 10].

BCI is an artificial intelligence system which uses EEG signals as input and produces control signals or commands at its output for computer so disabled people can use PC without mouse or keyboard [8, 14]. In a BCI system there are five consecutive stages: signal acquisition, signal preprocessing or enhancement, feature extraction, classification, and a control interface stage [2, 15].

The recognition of a person intention from his/her EEG signals is the core of the BCI system. This is really difficult task that challenges researchers who aim to see the day the BCI technology became more manageable and usable. In most existing BCI systems, the identification/recognition of information relies on classification algorithms that rely themselves on features extraction [15–17]. So far, different features have been proposed to evaluate different classifiers in order to identify different brain tasks [2, 7, 9, 14, 16, 18].

EEG signals are of non-linearity and non-stationarity nature. Due to this nature, finding discriminant features to interface commands device for MI-based BCI is still challenging researchers [9]. To this end, time–frequency representation (TFR), short Fourier transforms (SFT) [19], power spectral density (PSD), band-power (BP) [5, 20–22], wavelet transform analysis (WT) [5, 18] and autoregressive (AR) [23] features, to not name all, have been used to identify brain activities in MI-based BCI systems. It has been found that relevant features for MI tasks should be extracted in specific frequency bands [21]. Some spatial and frequency domain features such as CSP, PSD and WT have been found to be informative in MI-classification [9, 24]. However, robust features are generally accompanied by the dimensionality curse especially in small training data which is the case of the BCI studies [10]. Moreover, a large number of features slows down the classification process speed [20]. To tackle these problems, principal component analysis (PCA) as well as the independent component analysis (ICA) have been used to reduce features vector dimension [25–28]. Optimization approaches like genetic algorithms (GA) have also been used to select relevant features for a specific classification task [29, 30].

Using many kinds of features, a large number of classifiers have been employed to identify brain tasks especially in noninvasive MI-based BCI system [18, 21, 24–26, 31–33]. In these works, multilayer perceptron (MLP), support vector machine (SVM) with different kernel functions, linear discriminant analysis (LDA), k-nearest neighbor (KNN), naïve Bayes, hidden Markov model (HMM) and Fuzzy classifiers have extensively been used to identify different MI tasks. For the same features, the classifiers perform more a less the same way, i.e. they give almost the same classification accuracy [24]. It has been shown from these works that features are very significant and play an important role in the classification. It has also been shown that combined classifiers perform generally better than single classifier [34, 35].

As reported in many papers, in which researchers used the database "BCI competition IIIb" to identify the left and right hand actions, the classification accuracy (CA) did not exceed in the best case 81% [5, 20, 22, 24, 36]. Wavelet transforms is a very effective technique to extract features from an EEG signal for BCI applications [5, 20, 24, 37–44].

Bashashati et al. [24] used logistic regression (LR) classier and wavelet features with 'morl' mother wavelet and reached 81.47%. Brodu et al. [5] has also used 'morl' wavelet and reached classification accuracy (CA) of 80.95% by using linear discriminant analysis (LDA) classifier. Mother wavelets that are used for wavelet features in BCI applications were used in a coarse manner but according to existing works [5, 24, 37, 38], the 'morl' mother wavelet seems to have a good analysis for EEG signal.

Using the same database, this paper is an attempt to find new features so the binary classification accuracy for identifying left/right hand movement in MI-based noninvasive BCI system might be improved and exceed 81%. With this objective in mind, different kind of features based on PSD, CWT, fractal dimension and energy of EEG signals are extracted in specific frequency bands.

In this work, multi classifiers classification principle was used to recognize MI left and right hand. This principle of classification is based on many SVMs classifiers by which the class is indicated by the majority of classifiers by voting. To improve the classification performance, suitable features for different SVMs classifiers were selected by optimization using genetic algorithm (GA).

Methodology

EEG data set

The data set is provided by the Institute for Human- Computer Interfaces, University of Technology Graz—BCI Lab. This data set (IIIb) contained 2-class EEG data from three subjects (O3, S4 X11). Each data set contains recordings from consecutive sessions during a BCI experiment [45, 46]. The experiment consists of three sessions for each subject. Each session consists of four to nine runs. The recordings were made with a bipolar EEG amplifier from g.tec. EEG signal was sampled with 125 Hz, it was filtered between 0.5 and 30 Hz with Notch filter. In this experiment, subjects were asked to perform imagery actions for left or right hand when they heard a beep sound. To keep such action a feedback system with monitor display is used. For each try, there is a signal and corresponding labelled class i.e. left or right imagery hand action. This database will constitute training examples as well as test examples for the training classifiers.

Features extraction

The extraction phase is the most critical operation in any classification problem. In fact, features should be discriminate to carry information about the different classes. After having examined the EEG signals and their PSD in the two classes, it has been noted that the power frequency distribution seeing by the PSD of each class are different. This is the reason for why the PSD is used in some proposed features.

The PSD of a signal $x(t)$ is calculated by the expression:

$$P_x(f) = |fft(x)|^2 \tag{1}$$

where *fft* represents the fast Fourier transform and f is the frequency.

The difference between the corresponding PSD can be measured by the *kurtosis* which is given by the following equation:

$$Kurtosis(x) = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] \tag{2}$$

where μ and σ represent the mean and the standard derivation of the random variable X respectively. E represents the expected value operator.

Due to the non stationary of EEG signals, it is important to look for features that handle the aspect of time–frequency domain. Consequently, the use of the wavelet transform becomes a logical option. The continuous wavelet transform (CWT) of a signal $x(t)$ is given by:

$$C_x(a,b) = \frac{1}{a}\int_{-\infty}^{\infty} x(t)\Psi\left(\frac{t-b}{a}\right)dt \tag{3}$$

where a is a positive scale parameter and b is a parameter for the position of the mother wavelet Ψ in time. To increase the possibility of finding discriminant features, signals s3 and s4 from both channels C3 and C4 respectively are filtered with Butterworth filter (order 1) in two frequency bands B1 ([23–30] Hz) and B2 ([9.5–19] Hz). Thereby, for each try we get four different signals named s31, s32, s41 and s42. The signal s31 is the filtered signal of the channel C3 within the first band. Signals are centred to have zero mean and are of 1 s duration beginning at different time positions. Figure 1 shows a tailed flowchart of the signals processing and extracted features.

The proposed features are defined by the following expressions:

$$f_1 = kurtosis({P_{s31}}^2)/var({P_{s41}}^2) \tag{4}$$

$$f_2 = kurtosis({P_{s32}}^2)/median({P_{s42}}^2) \tag{5}$$

$$f_3 = var(P_{s31})/var(P_{s31} + P_{s41}) \tag{6}$$

$$f_4 = var(P_{s32} - P_{s42})/var(P_{s32} + P_{s42}) \tag{7}$$

$$f_5 = E_3/E_4 \tag{8}$$

$$f_6 = FD_{cvs3}/FD_{cvs4} \tag{9}$$

$$f_i = E_{(i-6)31} - E_{(i-6)41} \quad i = 7 \ldots 11 \tag{10}$$

$$f_k = E_{(k-10)32}/E_{(k-11)42} \quad k = 12 \ldots 16 \tag{11}$$

where *var* represents the variance, E_{i31} represents the ith energy corresponding to the ith scale (a) in the decomposed signal s31 via wavelet and is calculated as below:

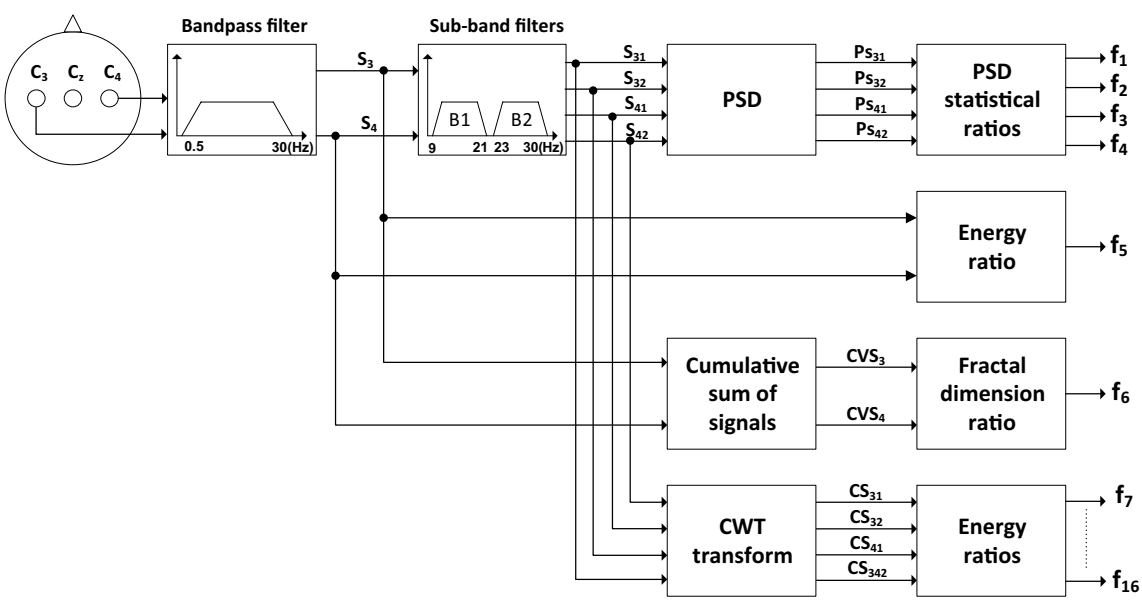

Fig. 1 Flowchart of the signal processing leading to features extraction

$$E_{i31} = \sum_{(b)} C_{s31}{}^{2}(i, :) \tag{12}$$

FD_{cs3} represents the fractal dimension of the cumulative sum of the signal s3 computed by the Higuchi method [47]. It was found that the Higuchi's method for FD computation in EEG signals is robust and gives accurate estimation results [48].

The cumulative sum of a signal x is given by:

$$cum\_sum(x) = cvx(n) = \sum_{k=0}^{n} x(k) \tag{13}$$

SVM classifier

SVM is a supervised learning algorithm for solving binary classification problems which was proposed by Vapnik [49]. To solve an SVM classification problem, an optimal hyperplane should be found to maximize the margin between two classes. Maximizing the margin leads to increase the generalization capabilities and the classification accuracy [50]. If two classes are not linearly separable, kernel functions are used to transform features vector to high dimension so the problem will be linearly separable. In case of using the radial base function (RBF), the parameter σ was adjusted to find the best fitting in term of classification performance.

Multi classifiers based classification principle

There are different ways to combine classifiers to perform classification: Bagging, Boosting, Voting and Stacking [16, 17, 51]. In case of voting method, the class of a given object, represented by its features, is given by the majority of several classifiers.

Figure 2 illustrates this principle of voting. It represents three classifiers CFi ($i = 1, 2, 3$) that had performed a 2D classification of imagined binary problem with seven objects each class (C1: circle, C2: square). Note that some objects are classified differently while using different classifiers with different features Fi.

The idea behind this method i.e. combining different classifiers with different features is to represent an object by different features; so it can be seen more correctly by all classifiers. Therefore, each kind of features can uncover another face of the object. For example, objects 1 and 4 (dashed circles) from the class C1 are misrecognized by the classifier CF1 while using their features vector F1, whereas; they are correctly classified by the others classifiers CF2 and CF3 when using features F2 and F3 respectively. The same thing can be said about objects 1 and 2 (dashed squares) from the class C2. The seventh object of both classes is misclassified by all classifiers. In this work, each classifier among three classifiers was trained by all training data i.e. not like in case of multi-classifiers Bagging technique where each classifier should be trained on a random subset from the whole training data [52].

Results

Classification accuracies in all results are performed by tenfold cross validation evaluation.

At the beginning, the proposed features were grouped in seven groups to feed seven separate SVMs classifiers. For these classifiers, RBF kernel function was used with the best sigma. The *sigma* parameter has been varied from 1 to 10 step 0.125 to get best performance. To see the effect of the features stationarity on the classification accuracy (CA), features were calculated at different time positions using the 'morl' mother wavelet (Table 1). Table 1 includes performed SVM classification with respect to time positions for the three subjects. Bold numbers indicate the maximum CA with respect to time positions. For sake of clarification, results of Table 1 are depicted in Fig. 3. According to these results, it is clear that all subjects performed their actions around 3.75 s (gray zone in Table 1) at which the CAs are maximum. It is also clear that all features group (16 features) performed better that other subgroups of features with best performance of 77.28%.

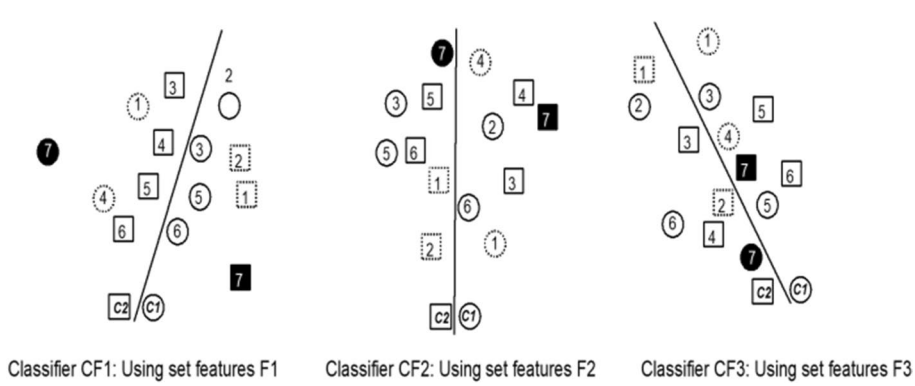

Fig. 2 Multi classifiers based classification principle

Table 1 Best SVM classification accuracy (CA) in percent (%) using different features groups, calculated for 1 s duration, with respect to time positions for all three subjects using tenfold cross validation (2.5 s corresponds to the interval [2.5–3.5] s of signal where the features are calculated; best SVM classifier was selected by varying sigma as 1:0.125:10 for each k in the tenfold cross validation)

Features (Number)	2.5s	2.75s	3s	3.25s	3.5s	3.75s	4s	4.25s	4.5s	4.75s	5s
f_1-$f6$ (6)	58.8	63.47	65.93	68.81	71.94	71.77	**72.88**	66.01	62.45	58.38	57.62
f_7-f_{11} (5)	61.77	62.45	66.27	71.61	72.37	**73.13**	72.79	70.16	68.13	65.42	66.10
$F_{12}f_{16}$ (5)	58.72	62.11	64.06	69.4	69.66	**70.59**	70.16	66.86	64.32	62.96	63.05
f_1-f_{11} (11)	59.83	64.57	67.71	70.93	74.15	**76.01**	74.15	71.61	68.64	67.28	67.62
F_7-f_{16} (10)	59.49	61.86	65.25	71.77	73.13	**73.47**	72.62	71.35	68.55	67.03	67.03
f_1-$f_6 \cup f_{12}$-f_{16} (11)	59.74	65.25	67.79	70.16	73.13	**75.67**	73.22	68.13	66.27	65.08	64.23
f_1-f_{16} (16)	60.84	64.06	67.88	70.67	73.81	**77.28**	74.32	71.77	68.55	67.54	69.4

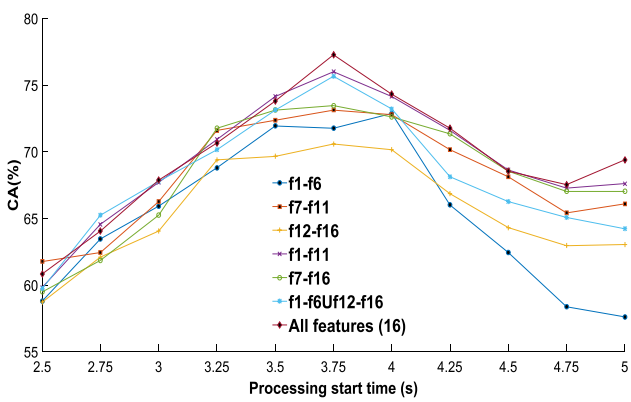

Fig. 3 Classification accuracy for three subjects using indicated features and SVM classifier in tenfold cross validation. Best sigma was selected in the interval [1–10] step 0.125

Practically, subjects cannot perform actions at the same time (synchronization problem). To explore time position in thorough manner, classification was performed with respect to time position in the interval [3.5–4] s with a step of 0.05 s. The effect of mother wavelet was also investigated in this case. Table 2 summarizes SVM CA using all features (the group that gives best CA in previous results) in all subjects with respect to time positions and indicated mother wavelets.

Bold numbers indicate the best CA with respect to time position and those underlined indicate the best CA with respect to mother wavelet. As shown by Table 2, the best CA, either for time position or mother wavelet, is localized in the interval [3.75–3.85] s (gray zone). These results localized more the subjects' actions and indicated that probable best mother wavelets are 'morl', 'sym6', 'db8', and 'sym10' in which the 'morl' is the best one with CA of 77.28%.

Table 2 Best SVM classification accuracy (CA) in percent (%) using all features (16) with respect to time positions and different mother wavelets for 1 s processing duration and tenfold cross validation for three subjects (Best SVM classifier was selected by varying sigma as: 1:0.125:10)

	3.5s	3.55s	3.6s	3.65s	3.7s	3.75s	3.8s	3.85s	3.9s	3.95s	4s
morl	73.81	72.45	73.55	73.98	75.67	**77.28**	76.18	75.59	75.5	74.74	74.32
mexh	74.40	72.88	72.71	73.89	75.76	74.83	75.59	**76.10**	74.83	74.49	75.25
coif4	74.06	73.13	75.59	73.98	75.42	**76.44**	76.44	76.35	75.76	74.15	74.32
coif5	74.49	73.3	75.16	74.4	75.33	**76.35**	76.27	**76.35**	75.84	74.4	74.15
db6	73.81	73.55	74.32	73.55	75.25	75.67	76.27	**76.52**	76.52	75	75.42
db8	73.89	72.96	75	73.98	75.33	76.01	76.44	**76.61**	75.5	74.83	74.83
db10	73.22	73.05	74.66	74.23	75.5	75.5	76.27	**76.44**	76.1	75.33	75.84
sym6	73.89	73.38	75	74.15	75.5	76.69	**76.94**	76.35	76.1	74.57	74.66
sym8	73.81	73.38	75.42	74.15	75.5	76.44	**76.61**	76.52	75.59	74.23	74.49
sym10	74.23	73.64	75.25	74.4	75.25	76.27	76.44	**76.61**	75.76	74.49	74.49

Previous results were evaluated for all subjects together; remaining results will be subject- dependent i.e. performing classification for each subject separately.

Table 3 shows the best SVM CA using all features at indicated time positions and mother wavelets.

These results indicate that best CA for subject S1 is 81% (at 3.85 s using 'sym6' mother wavelet), for subject S2 is 81.08% (at 3.85 s using the 'morl' mother wavelet) and for subject S3 is 80.96% (at 3.75 s using the 'morl' mother wavelet).

To improve the CA, a single SVM and multiple SVMs classifiers (three SVMs) were optimized for each subject. The optimization was performed by genetic algorithm (GA) in order to select suitable features for each classifier (either single or multiple classifiers) so the misclassification error would be minimal. This is done for all tenfolds while using tenfold cross validation evaluation. Table 4 represents CA for optimized single SVM and optimized three SVMs for each subject. In these optimizations, the time positions and mother wavelets for each subject are those that give the best performance in the previous results (Table 3). Features were calculated using the mother wavelet 'morl' for all subjects.

Although the best results for subject S1 was found by using 'sym6' mother wavelet (Table 3), but optimization results in this case (subject S1 with 'sym6') showed that using 'morl' wavelet is the best. Here are optimization results for subject S1:

1SVM ('sym6' at 3.85 s): [95 90 95 85 80 100 85 90 85 100] with a mean of 90.5%.

3SVMs ('sym6' at 3.85 s): [100 100 95 90 85 100 90 100 95 100] with a mean of 95.5%.

For this reason, all optimization results in Table 4 were performed by 'morl' mother wavelet.

Figures 4, 5 and 6 show examples of GA optimizations (italic values in Table 4) to select suitable features for single SVM and three SVMs for subjects S1 (k = 2), S2 (k = 4) and S3 (k = 10) respectively.

Figure 4a shows the GA optimization for selecting suitable features for the single SVM classifier in case of k = 2. The CA reaches 85% (100–15%) with the selected features f_2, f_5, f_8.

We have to mention that multi classifiers should have at least one linear SVM classifier to get a best CA. Figure 4b shows an example of three SVMs optimization for subject S1 in case of k = 2. The GA should find the best combination

Table 3 Best SVM classification accuracy (CA) in percent using all features (16) for subject S1 (O3), S2 (S4) and S3(X11) with respect to indicated time positions and mother wavelets in tenfold cross validation evaluation

Time position	3.75 s				3.8 s				3.85 s			
Subjects	S1	S2	S3	Mean	S1	S2	S3	Mean	S1	S2	S3	Mean
morl	73.5	**77.39**	**80.96**	77.28	78.5	**78.47**	79.42	78.79	79	**81.08**	**78.26**	79.44
sym6	75.5	76.52	80.19	77.4	**79**	78.26	**79.61**	78.95	**81**	80.86	77.5	79.78
db8	**76.5**	76.52	80.19	77.73	77.5	78.04	78.65	78.06	80	79.13	77.5	78.87
sym10	75.5	77.17	80	77.55	78	78.26	78.26	78.17	79.5	80.43	77.69	79.2

Best SVM classifier was selected by varying sigma as: 1:0.125:10)

Bold values indicate the maximum CA with respect to indicated variables (mother wavelet, time position and subject)

Table 4 Classification accuracy (CA) in percent by single optimized SVM classifier and 3 optimized SVM classifiers for three subjects with respect to k (fold or partition number) in tenfold cross validation

Subjects	S1 (3.85 s)		S2 (3.85 s)		S3 (3.75 s)	
k	1 SVM	3 SVMs	1 SVM	3 SVMs	1 SVM	3 SVMs
1	90	100	89.14	95.66	91.39	92.31
2	*85*	*95*	82.61	89.14	82.7	86.54
3	90	100	84.79	89.14	86.54	88.47
4	95	100	*95.66*	*97.83*	86.54	91.39
5	85	90	89.14	91.31	91.39	92.31
6	100	100	86.96	91.31	88.47	91.39
7	90	90	84.79	86.96	88.47	90.39
8	90	100	95.66	100	88.47	90.39
9	90	100	91.31	95.66	92.31	94.24
10	95	100	95.66	97.83	*92.31*	*96.16*
Mean	91	97.5	89.57	93.48	88.85	91.35

For each case (1 SVM or 3 SVM), suitable features were selected by GA. Indicated times represent the start times for 1 s processing duration. All features were calculated using 'morl' mother wavelet

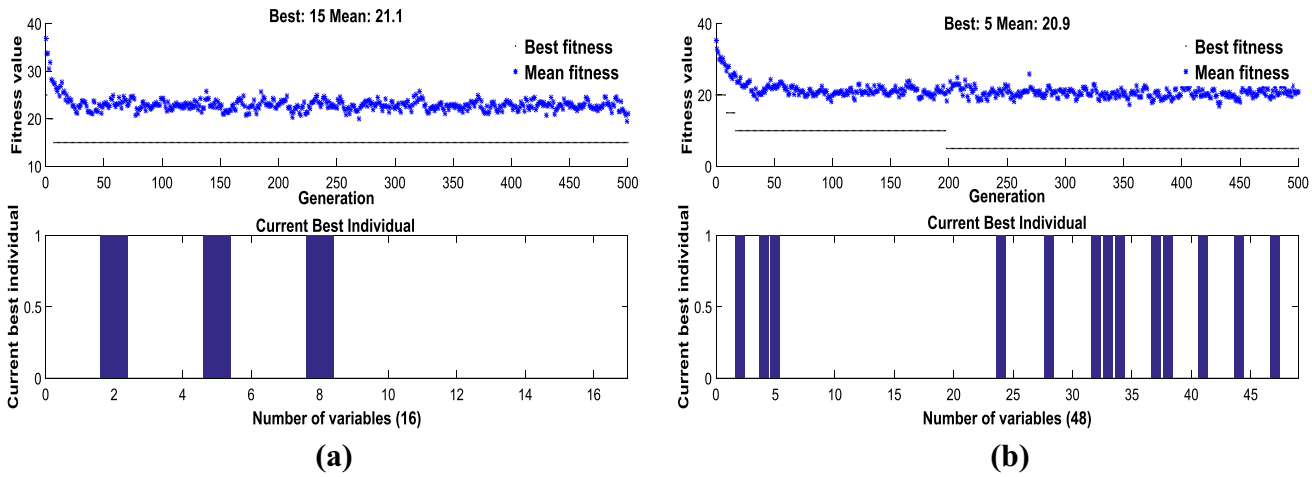

Fig. 4 Genetic algorithm (GA) optimization for subject S1 and in fold k = 2. **a** Misclassification accuracy and selected features for single SVM (σ = 2). **b** Misclassification accuracy and selected features for three SVMs (σ1 = 1, σ2 = 3.5)

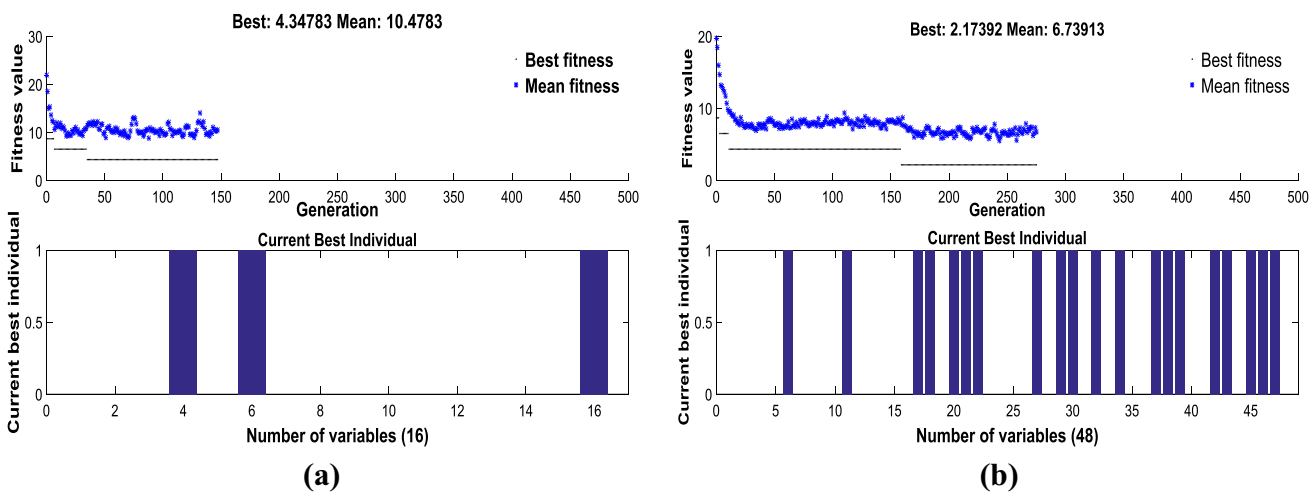

Fig. 5 Genetic algorithm (GA) optimization for subject S2 and in fold k = 4. **a** Misclassification accuracy and selected features for single SVM (σ = 3.5). **b** Misclassification accuracy and selected features for three SVMs (σ1 = 2, σ2 = 3.5)

among 2^{48} cases i.e. find an appropriate set of features among the 16 features for each classifiers. Therefore, GA has to select features coded in 48 (16 + 16 + 16) binary variables (feature is either used or not used for a given classifier). In this case, the CA was improved from 85 to 95% with the selected features: f_2, f_4, f_5 for the first classifier, f_8, f_{12}, f_{16} for the second classifier and $f_1, f_2, f_5, f_6, f_9, f_{12}, f_{15}$ for the third classifier.

In the same way, we have tried to find four SVMs classifiers, but we could not optimize them to find a mean CA more than 94.11%.

Let's now compare our methods with the existing methods in which the same database 'BCI competition IIIb' have been used. Table 3 summarizes the classification accuracies of different works to identify left and right hand movement based on motor imagery EEG signals. In these works, different features and different classifiers were used. Reported features and classifiers are those which gave the best performance as a mean for the three studied subjects.

According to Table 5, classification accuracies did not exceed in best case 81.47%. It is clear that the proposed features along with the multi SVMs classifiers used in the present work have remarkably improved the classification performance and achieved 94.11%.

Discussion and conclusion

In this work, imagery left and right hand movements are classified using multi SVMs classifiers. In fact, sixteen features, which are extracted from the PSD, wavelet transform

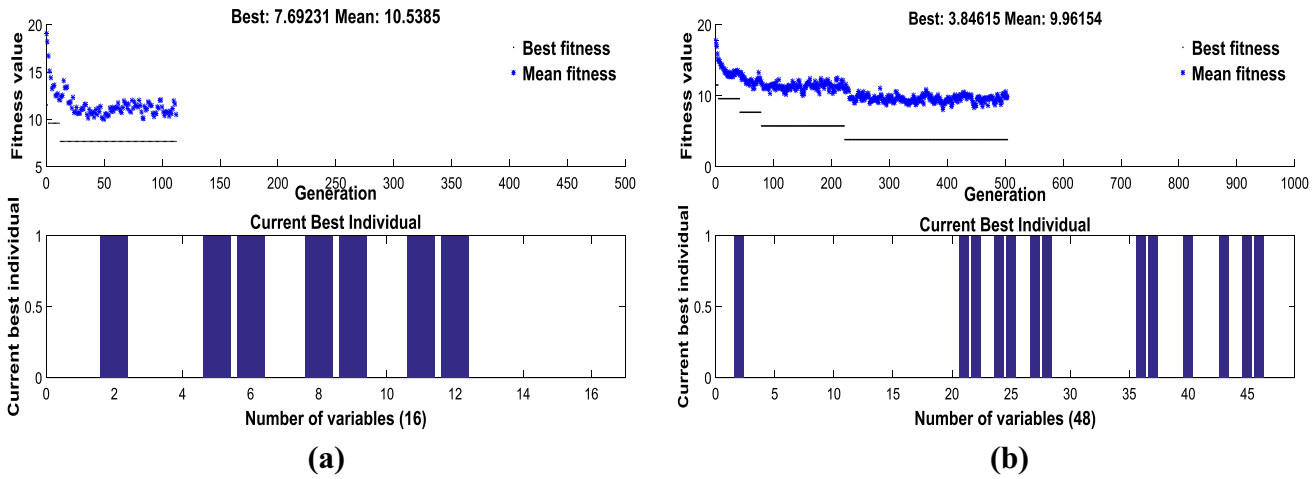

Fig. 6 Genetic algorithm (GA) optimization for subject S3 and in fold k = 10. **a** Misclassification accuracy and selected features for single SVM (σ = 1.25). **b** Misclassification accuracy and selected features for 3 SVMs (σ1 = 1.25, σ2 = 2.5)

Table 5 Summary of classification accuracies (CA %) for the indicated classification methods which used the database 'BCI competition IIIb' to identity left/right hand tasks

Works	Features	Classifiers	S1	S2	S3	Mean
Lotte et al. [36]	BP values	SVM	86.8	75.9	75.4	79.36
Zhong et al. [22]	BP values	VB	89.3	72.96	75.18	78.96
Brodu et al. [5]	Wavelet features (Morlet)	LDA	–	81.5	80.4	80.95
Brodu et al. [20]	BP, Multifractal, Complexity	LDA	–	81.7	80.9	81.3
Bashashati et al. [24]	Wavelet features (Morlet)	LR	82.39	83.89	78.15	81.47
This work	16 Diversified features	Three optimized SVMs	97.5	93.48	91.35	**94.11**

Bold values indicate the maximum CA among all classification works

and fractal dimension of EEG signals, are used to feed different SVMs classifiers. From two EEG signals of the channels C3 and C4, four signals were derived using filtering of two sub-bands in order to diversify the features for better distinction of motor imagery actions. To take the spatial information in account, all proposed features are ratios or subtractions of intermediate features calculated from the two channels.

This study reveals the difficulty of the identification of the hand movement from its motor imagery EEG signals. This is due principally to the EEG signals nature as well as their generation mechanism inside the brain. Furthermore, different subjects cannot perform actions simultaneously. Even for the same subject when performing actions in different trials, he/her cannot behave in the same way neither perform action in the same time position. This is a really non stationary classification problem.

Using the proposed features in coarse manner for all subjects did not lead to a good performance (Table 1). In fact, all features together give the maximum CA of 77.28%. When these sixteen features were used for different subjects separately, the maximum CA has exceeded 80% (Table 3). This means that subjects are not actions synchronized and they cannot be. It seems from this study, that the proposed features give the best performance for all studied subjects in the interval [3.5–4] s (Fig. 3). The start time position from this interval can vary from subject to subject. Time positions of 3.85 s, 3.85 s and 3.75 s were the best start positions for subjects S1, S2 and S3 respectively. The solution to the start time position problem is either finding stationary features which follow the action class whatever the action time or finding the correct start time position. The first solution is very difficult due to the variability between subjects while performing actions. The second solution (our suggestion) could be a supplement binary classification for the correct time position (to start, not yet). Once we get the correct start time decision, we can start action identification or classification. In this case, actions classification would be performed efficiently.

Most of proposed features are derived from continuous wavelet transform (CWT). This is why the effect of the mother wavelet on the CA was investigated. Used mother wavelets (Table 2, gray zone) seem having the same effect (there is no big difference) on the CA but the 'morl' mother wavelet is the best one.

To not use features randomly, genetic algorithm (GA) was used to select suitable features for single SVM and multi SVMs classifiers. A CA in optimized single SVM was reached a CA mean of 89.8%. Optimization of three SVMs classifier has improved the CA mean to 94.11%. It is clear that using multi classifiers classification improves remarkably the classification performance. Getting a CA mean of 94.11% is very interesting result comparing to the existing method for the same data base which did not exceed 81%.

We have tried many times to find four SVMs classifiers; unfortunately we could not reach a performance greater than already found. This is due to the nature of features which could not add discriminate information more than already given for the three classifiers (saturation).

The proposed features, in this work, with the combined optimized SVMs classifiers have clearly improved the classification accuracy comparing to the existing works. We think that the diversity of the proposed features (different kinds) has better contributed to improve the motor imagery actions identification when multi classifiers based classification is used. In diversified features, classes or objects to be classified are represented differently. This is why, as explained in 'Multi classifiers based classification principle' section, multi classifiers can find suitable features in a way to improve the CA more easier due the their diversity.

In this work, multi classifier based classification was investigated using the SVM classifier as an example due to its robustness. It would be preferable to compare other kind of classifiers for the same database to see which classifier would perform the best.

Acknowledgements The authors would like to acknowledge the Institute for Knowledge Discovery (Laboratory of Brain-Computer Interfaces), Graz University of Technology, for providing the dataset online (https://www.bbci.de/competition/iii).

Compliance with ethical standards

Conflict of interest No conflicts of interest are declared by the authors.

Ethical approval This article does not contain any studies with human participants performed by any of the authors.

References

1. Teplan M (2002) Fundamental of EEG measurement. Meas Sci Rev 2(2):1
2. Nicolas-Alonso LF, Gomez-Gil J (2012) Brain computer interfaces, a review. Sensors 12(2):1211–1279
3. Sanei S, Chambers JA (2007) EEG signal processing. Wiley, Chichester
4. The Ten Twenty Electrode System: International Federation of Societies for Electroencephalography and Clinical Neurophysiology. Am J EEG Technol 1961; 1:13–19
5. Brodu N, Lotte F, Lécuyer A (2011) Comparative study of band-power extraction techniques for motor imagery classification. In: IEEE symposium on computational intelligence, cognitive algorithms, mind, and brain (CCMB), pp 1–6
6. Ramoser H, Muller-Gerking J, Pfurtscheller G (2000) Optimal spatial filtering of single trial EEG during imagined hand movement. IEEE Trans Rehabil Eng 8(4):441–446
7. Blankertz B, Tomioka R, Lemm S, Kawanabe M, Muller KR (2008) Optimizing spatial filters for robust EEG single-trial analysis. Signal Process Mag IEEE 25(1):41–56
8. Ebrahini T, Vesin JM, Garcia G (2003) Brain-computer interface in multimedia communication. IEEE Signal Process Mag 20(1):14–24
9. Chaudhari R, Galiyawala HJ (2017) A review on motor imagery signal classification for BCI. Signal Process Int J (SPIJ) 11(2):16–34
10. Abdulkader SN, Atia A, Mostafa MSM (2015) Brain computer interfacing: applications and challenges. Egyptian Inform J 16:213–230
11. Malouin F, Richards CL (2013) Clinical applications of motor imagery in rehabilitation Multisensory imagery. Springer, New York, pp 397–419
12. Middendorf M, McMillan G, Calhoun G, Jones KS (2000) Brain–computer interfaces based on steady-state visual evoked response. IEEE Trans Rehabil Eng 8:211–213
13. Kundu S, Ari S (2017) P300 detection with brain-computer interface application using PCA and ensemble of weighted SVMs. IETE J Res 64(3):406–414
14. Bashashati A, Fatourechi M, Ward RK, Birch GE (2007) A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals. J Neural Eng 4(2):32–57
15. Wolpaw JR, Birbaumer N, McFarland DJ, Pfurtschellere G, Vaughan TM (2002) Brain–computer interfaces for communication and control. Clin Neurophysiol 113(6):767–791
16. Lotte F, Congedo M, Léecuyer A, Lamarche F, Arnaldi B (2007) A review of classification algorithms for EEG-based brain computer interfaces. J Neural Eng 4(2):R1–R13
17. Lotte F, Bougrain L, Cichocki A, Clerc M, Congedo M, Rakotomamonjy A, Yger F (2018) A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update. J Neural Eng 15(3):031005
18. Pattnaik PK, Sarraf J (2016) Brain computer interface issues on hand movement. J King Saud Univ Comput Inform Sci 30(1):18–24
19. Coyle D, Prasad G, McGinnity TM (2005) A time-frequency approach to feature extraction for a brain-computer interface with a comparative analysis of performance measures. EURASIP J Appl Signal Process 19:3141–3151
20. Brodu N, Lotte F, Lécuyer A (2012) Exploring two novel features for EEG-based brain–computer interfaces: multifractal cumulants and predictive complexity. Neurocomputing 79:87–94
21. Pfurtscheller G, Neuper C (2001) Motor imagery and direct brain computer communication. Proc IEEE 89(7):1123–1134
22. Zhong M, Lotte F, Girolami M, Lécuyer A (2008) Classifying EEG for brain computer interfaces using Gaussian processes. Pattern Recogn Lett 29(3):354–359
23. Pfurtscheller G, Neuper C, Schlogl A, Lugger K (1998) Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters. IEEE Trans Rehabil Eng 6(3):316–325
24. Bashashati H, Ward RK, Birch GE, Bashashati A (2015) Comparing different classifiers in sensory motor brain computer interfaces. PLoS ONE 10(6):e0129435
25. Lee H, Choi S (2003) PCA+HMM+SVM for EEG pattern classification. In: Proceedings of the IEEE seventh international symposium on signal processing and its applications, pp 541–544

26. Lin CJ, Hsieh MH (2009) Classification of mental task from EEG data using neural networks based on particle swarm optimization. Neurocomputing 72(4–6):1121–1130
27. Wang Y, Jung T-P (2012) Improving brain-computer interfaces using independent component analysis. In: Allison B, Dunne S, Leeb R, Del R, Millán J, Nijholt A (eds) Towards practical brain–computer interfaces, vol 67. Springer, Berlin, p 83
28. Chiappa S, Barber D (2006) EEG classification using generative independent component analysis. Neurocomputing 69(7–9):769–777
29. Garrett D, Peterson DA, Anderson CW, Thaut MH (2003) Comparison of linear nonlinear and feature selection methods for EEG signal classification. IEEE Trans Neural Syst Rehabil Eng 11(2):141–144
30. Amarasinghe K, Sivils P, Manic M (2016) EEG feature selection for thought driven robots using evolutionary algorithms. In: 9th IEEE international conference on human system interactions (HSI), pp 355–361
31. Herman P, Prasad G, McGinnity TM, Coyle D (2008) Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification. IEEE Trans Neural Syst Rehabil Eng 16(4):317–326
32. Obermaier B, Guger C, Pfurtscheller G (1999) Hidden Markov models used for off line classification of EEG data. Biomed Technik 44(6):158–162
33. Cai Y, Zhang L, Wenjuan J, Changsheng L (2015) A hybrid SVM/HMM classification method for motor imagery based BCI. J Comput Inform Syst 11(4):1259–1267
34. Jain AK, Duin RPW, Mao J (2000) Statistical pattern recognition: a review. IEEE Trans Pattern Anal Mach Intell 22(1):4–37
35. Rakotomamonjy A, Guigue V, Mallet G, Alvarado V (2005) Ensemble of svms for improving brain computer interface p300 speller performances. In: International conference on artificial neural networks, pp 45–50
36. Lotte F, Lécuyer A, Lamarche F, Arnaldi B (2007) Studying the use of fuzzy inference systems for motor imagery classification. IEEE Trans Neural Syst Rehabil Eng 15(2):322–332
37. Lemm S, Schafer C, Curio G (2004) BCI competition 2003–data set III: probabilistic modeling of sensorimotor mu rhythms for classification of imaginary hand movements. IEEE Trans Biomed Eng 51:1077–1080
38. Ende FM, Louis AK, Maass P, Mayer-Kress G (1998) EEG signal analysis by continuous wavelet transform techniques. In: Kantz H, Kurths J, Mayer-Kress G (eds) nonlinear analysis of physiological data. Springer, Berlin, pp 213–219
39. Samar VJ, Bopardikar A, Rao R, Swartz K (1999) Wavelet analysis of neuroelectric waveforms: a conceptual tutorial. Brain Lang 66:7–60
40. Akay M (ed) (1997) Time frequency and wavelets in biomedical signal processing, series in biomedical engineering. IEEE Press, Piscataway
41. Hsu WY, Sun YN (2009) EEG-based motor imagery analysis using weighted wavelet transform features. J Neurosci Methods 167(2):310–318
42. Yang BH, Yan GZ, Yan RG, Wu T (2007) Adaptive subject-based feature extraction in brain-computer interfaces using wavelet packet best basis decomposition. Med Eng Phys 29(1):48–53
43. Song A, Xu B (2010) Features extraction of motor imagery EEG based on wavelet transform and higher-order statistics. Int J Wavelets Multiresolut Inf Process 8(3):373–384
44. Xu Q, Zhou H, Wang Y, Huang J (2009) Fuzzy support vector machine for classification of EEG signals using wavelet-based features. Med Eng Phys 31(7):858–865
45. Blankertz B, Müller KR, Curio G, Vaughan TM, Schalk G, Wolpaw JR, Schlögl A, Neuper C, Pfurtscheller G, Hinterberger T, Schröder M, Birbaumer N (2004) The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials. IEEE Trans Biomed Eng 51(6):1044–1051
46. Blankertz B, Muller KR, Krusienski DJ, Schalk G, Wolpaw JR, Schlogl A, Pfurtscheller G, Millan JDR, Schroder M, Birbaumer N (2006) The BCI competition III: validating alternative approaches to actual BCI problems. IEEE Trans Neural Syst Rehabil Eng 14(2):153–159
47. Higuchi T (1988) Approach to an irregular time series on the basis of the fractal theory. Physica D 31:277–283
48. Accardo A, Affinito M, Carrozzi M, Bouquet F (1997) Use of the fractal dimension for the analysis of electroencephalographic time series. Biol Cyber 77:339–350
49. Vapnik VN (1995) The nature of statistical learning theory. ISBN 0-387-94559-8, Springer-Verlag
50. Burges CJC (1998) A tutorial on support vector machines for pattern recognition. Data Min Knowl Disc 2(2):121–167
51. Breiman L (1996) Bagging predictors. Mach Learn 24(2):123–140
52. Opitz D, Maclin R (1999) Popular ensemble methods: an empirical study. J Artif Intell Res 11:169–198