

A comparison approach toward finding the best feature and classifier in cue-based BCI

R. Boostani · B. Graimann · M. H. Moradi ·
G. Pfurtscheller

Received: 26 June 2006 / Accepted: 11 January 2007 / Published online: 23 February 2007
© International Federation for Medical and Biological Engineering 2007

Abstract In this paper, a comparative evaluation of state-of-the art feature extraction and classification methods is presented for five subjects in order to increase the performance of a cue-based Brain–Computer interface (BCI) system for imagery tasks (left and right hand movements). To select an informative feature with a reliable classifier features containing standard bandpower, AAR coefficients, and fractal dimension along with support vector machine (SVM), Adaboost and Fisher linear discriminant analysis (FLDA) classifiers have been assessed. In the single feature-classifier combinations, bandpower with FLDA gave the best results for three subjects, and fractal dimension and FLDA and SVM classifiers lead to the best results for two other subjects. A genetic algorithm has been used to find the best combination of the features with the aforementioned classifiers and led to dramatic reduction of the classification error and also best results in the four subjects. Genetic feature combination results have been compared with the simple feature combination to show the performance of the Genetic algorithm.

Keywords Brain–Computer interface (BCI) · Adaboost · Support vector machine (SVM) · Adaptive auto regressive (AAR) · Fisher linear discriminate analysis (FLDA) · Fractal dimension (FD) · Bandpower (BP) · Genetic algorithm

1 Introduction

In order to assist disabled people who suffer from motor impairment like amyotrophic lateral sclerosis (ALS) patients, an alternative communication channel can be provided by a brain computer interface (BCI) system [20]. BCI systems can enable the patients to move the cursor on a screen [20, 23], or grasp a glass by sending the signal commands to their orthosis, prosthesis, or functional electrical stimulation unit [14]. An optimistic view for the future of this research is that a BCI system might someday help tetraplegic patients with spinal cord injuries to move their limbs by functional electrical stimulation (FES) controlled by thought. This interesting application is in its primary stages since the classification rate even between two imagery tasks is still insufficient in some subjects. To increase the performance of BCI systems, much research has been done by BCI research groups. In this way, the Graz-BCI research group has employed discriminative features based on second order statistics such as bandpower [14], adaptive autoregressive coefficients (AAR) [18], and wavelet coefficients [19], and also combination of features by distinction sensitive learning vector quantization (DSLVO) [8] with well-known classifiers containing Fisher's linear discriminant analysis (FLDA) [13], finite impulse response multi layer perceptrons (FIRMLP) [2], linear vector quantization (LVQ) [17],

R. Boostani (✉)
Department of Computer Science and Engineering,
School of Engineering, Shiraz University, Shiraz, Iran
e-mail: boostani@shirazu.ac.ir

M. H. Moradi
Department of Biomedical Engineering
Amirkabir University of Technology (Tehran Polytechnic),
University of Technology, Tehran, Iran

B. Graimann · G. Pfurtscheller
Laboratory of Brain–Computer Interfaces,
Institute for Knowledge Discovery at the Graz,
University of Technology, Graz, Austria

hidden Markov models (HMM) [8], and minimum distance classifiers [18] to improve the classification rate between the various movement imagery tasks. Deriche and Al-Ani [3] selected the best feature combination among the variance, AR coefficients, wavelet coefficients, fractal dimension by modified mutual information feature selection (MMIFS) method. He showed that a combination of the aforementioned features is more efficient than each of them individually [3]. To accumulate the scattered results, a comprehensive assessment for well-known features and classifiers on the same data set is deemed necessary. The aim of this paper is to compare different feature extraction and feature classification methods based on electroencephalogram (EEG) data from two motor imagery tasks in the synchronous mode. It should be considered that most BCIs operate in a cue-based (synchronous or system-driven) mode where the subject can change his or her mental task only in certain intervals that are determined by the system. Ideally, a BCI system should be designed to operate in a self-paced (asynchronous or user-driven) mode. Such systems, however, are much more difficult to realize, which is the reason that most present-day BCI systems employ a synchronous paradigm. The data investigated in this study was also recorded according to a synchronous paradigm.

Brain activity associated with motor imagery can result in characteristic EEG-patterns as e.g. ERD/ERS [19] and these patterns should be best captured by the extracted features. It has been shown [19] that in the pre-motor-sensory area (the area that we catch the EEG signals), EEG of specific imagery tasks has similar patterns with the EEG of real movement of that specific task. Therefore, there is no difference between imagination of a normal subject and a disabled subject. In order to find an informative feature along with a reliable classifier, a comprehensive assessment has been done in this paper. Features such as bandpower (BP), AAR coefficients, and fractal dimension (FD) along with classifiers including support vector machine (SVM), Adaboost and FLDA are evaluated. Furthermore, in order to find the best combination of the features for the mentioned classifiers, a genetic algorithm (GA) has been used. The remainder of this paper is a short description of the aforementioned features, then an introduction of the classifiers, and finally a presentation of the results followed by a discussion.

2 Subjects and data acquisition

Five subjects (L1, o3, o8, g8, and f8), familiar with the Graz-BCI, participated in this study. Subjects are

ranged from 25 to 35 years old. Each subject sat in a armchair about 1.5 m in front of the computer screen. Three bipolar EEG-channels were recorded from six Ag/AgCl electrodes placed 2.5 cm anterior and 2.5 cm posterior to the standardized positions C3, Cz and C4 (international 10–20 system). The EEG was filtered between 0.5 and 50 Hz and recorded with a sample frequency of 128 Hz.

Graz-BCI training paradigm The training consisted of a repetitive process of triggered movement imagery trials. Each trial lasted 8 s and started with the presentation of a blank screen. A short acoustical warning tone was presented at second 2 and a fixation cross appeared in the middle of the screen. At the same time, the trigger was set from 0 to 1 for 500 ms. From second 3 to second 7, the subjects performed left or right hand motor imagery according to an arrow (cue) on the screen. An arrow pointing either to the left or to the right indicated the imagination of a left hand or right hand movement. The order of appearance of the arrows was randomized and at second 7 the screen content was erased. The trial finished with the presentation of a randomly selected inter-trial period (up to 2 s) beginning at second 8. Figure 1 shows the timing scheme. Three sessions were recorded for each subject on three different days. Each session consisted of three runs with 40 trials each.

3 Feature extraction

The goal of feature extraction is to find a suitable representation (signal features) of the data that simplifies the subsequent classification or detection of brain patterns. The signal features should encode the commands sent by the user, but should not contain or at least reduce noise or other patterns that can impede the classification process. To extract discriminative features, bandpower, AAR coefficients, and fractal dimension are selected in this paper. These are described briefly below.

Bandpower The EEG contains different specific frequency bands, that is standard alpha (10–12 Hz) and beta (16–24 Hz) bands, which are particularly

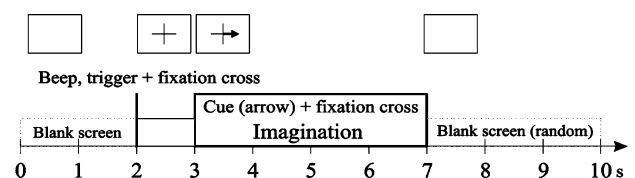


Fig. 1 Training paradigm

important to classify the different brain states, especially for discriminating of imagery tasks. For this study bandpower features were calculated by applying a butterworth filter (order 5), squaring of the samples and then averaging of subsequent samples (1 s average with 250 ms overlap) [18].

Adaptive autoregressive coefficients AAR coefficients are extracted from a signal and reflect the whole variation of the signal. In contrast to the AR model, AAR coefficients are adapted sample by sample and therefore [18]. Update coefficient and model order are two important factors that should be selected regarding the minimum error of model. An AAR model describes a signal Y_k as

$$Y_k = a_{1,k}Y_{k-1} + a_{2,k}Y_{k-2} + \dots + a_{p,k}Y_{k-p} + X_k$$

where X_k is assumed to be a white noise process, $a_{1,k}, \dots, a_{p,k}$ are the time dependent autoregressive parameters, and Y_{k-1}, \dots, Y_{k-p} are the past p samples of the time series, i.e. the model order. Writing the past samples and the estimated AAR parameters as vectors

$$\mathbf{Y}_k = [Y_{k-1} \ Y_{k-2} \ \dots \ Y_{k-p}]^T$$

$$\hat{\mathbf{a}}_k = [\hat{a}_{1,k} \ \hat{a}_{2,k} \ \dots \ \hat{a}_{p,k}]^T$$

where the T denotes vector transpose and the ‘hat’ notation indicates estimated values, the adaptive estimation procedure can be summarized by the following equations:

$$e_k = y_k - \hat{\mathbf{a}}_{k-1}^T \mathbf{Y}_{k-1}$$

$$\mathbf{k}_k = \frac{\mathbf{A}_{k-1} \mathbf{Y}_{k-1}}{\mathbf{Y}_{k-1}^T \mathbf{A}_{k-1} \mathbf{Y}_{k-1} + 1}$$

$$\mathbf{X}_k = \mathbf{A}_{k-1} - \mathbf{k}_k^T \mathbf{Y}_{k-1}^T \mathbf{A}_{k-1}$$

$$\mathbf{A}_k = \mathbf{X}_k + \frac{uc \cdot \mathbf{I} \cdot \text{trace}(\mathbf{A}_{k-1})}{p}$$

$$\hat{\mathbf{a}}_k = \hat{\mathbf{a}}_{k-1} + \mathbf{k}_k^T e_k$$

e_k is the one-step prediction error, K_k is the Kalman gain vector, and \mathbf{I} is the identity matrix. For this study a sixth order model and an update-coefficient uc of 0.0085 were used.

Fractal dimension (FD) BP and AAR features are based on the second order statistic of the signal and thus they describe the spectral information in the data. FD, however, captures nonlinear dynamics in the signal. Although all features here try to capture the underlying neurophysiological patterns in the signal,

fractal dimension (FD) has a direct relationship with the entropy of the signal, and in turn the entropy is related to the amount of information inside a signal. FD determines the amount of complexity of a signal. More fluctuation in the attractor shape is reflected by a higher value of fractal dimension. There are several methods to calculate the FD [5]. In this study we employed Higuchi’s method, which can be calculated as follows: Consider a signal containing N samples ($x(1), x(2), \dots, x(N)$). Construct k new time series x_m^k (embedded space) as:

$$x_m^k = \left\{ x(m), x(m+k), x(m+2k), \dots, x\left(m + \left\lceil \frac{N-m}{k} \right\rceil k\right) \right\}$$

for $m = 1, 2, \dots, k$.

where m indicates the initial time value, and k indicates the discrete time interval between points. For each of the k time series x_m^k , the length $L_m(k)$ is computed by:

$$L_m(k) = \frac{\sum_{i=1}^{\lceil (N-m)/k \rceil} |x(m+ik) - x(m+(i-1)k)| (N-1)}{[(N-m)/k]k}$$

for $m = 1, \dots, k$.

where N is the total length of the data sequence x and $(N-1)/[(N-m)/k]k$, is a normalization factor. An average length of every sub-sequence is computed as the mean of the k lengths $L_m(k)$. This procedure is repeated for the different values of k ($k = 1, 2, \dots, k_{\max}$), that k_{\max} varies for each k . There is no analytical formula for determining the number k , therefore, it has to be found experimentally. An average length for each k is obtained which may be expressed as proportional to k^{-D} , where D is the signal’s fractal dimension. In order to find the best value of k , from the log–log plot of $\ln(L(k))$ versus $\ln(1/k)$ one obtains the slope of the least-squares linear best fit. The fractal dimension of the signal, D , is then calculated as:

$$D = \log[L(k)] / \log(1/k)$$

4 Classification

Classification is a very important component in brain computer communication systems. In this study, support vector machine, linear discriminant analysis, and Adaboost were employed to discriminate between left and right movement imagery. A short description of these methods is given in the following:

Support vector machine (SVM) SVMs were proposed by Vapnik [22] and have been used successfully in many applications [1]. There are linear and non-linear SVMs that can classify the input patterns based on minimum error risk criterion. To increase SVM capability, a nonlinear kernel is used to map the data to a high dimensional space which is more separable according to the cover's theorem. Among different kernels, Gaussian kernels have a strong generalization property, and therefore were selected for our research. The variance of Gaussian kernels is a very important factor in the SVM performance and was determined by cross validation on the training set. To optimize the decision boundary, the following relation should be optimized.

$$Q(\alpha) = \sum_{i=1}^N \alpha_i - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N \alpha_i \alpha_j d_i d_j k(x_i, x_j)$$

$$k(x, x_i) = \varphi^T(x) \varphi(x_i)$$

where $\varphi(x)$ is the kernel function and α_i denotes the Lagrange coefficients which are bounded by the limitation parameter c . Limitation parameter c plays an important role in the performance of SVM because it is related inversely to the total error. After finding the value of alpha, the weight vector which classifies two classes is determined by

$$w = \sum_{i=1}^N \alpha_i d_i \varphi(x_i).$$

In order to perform the required calculations, the freely available SVM3 toolbox [11] was used in this study.

Fisher linear discriminant analysis (FLDA) The goal of FLDA is to find a direction in the feature space along which the distance of the means relative to the within-class scatter, described by the within-class scatter matrix S_w , reaches a maximum, thereby maximizing the class separability. This goal can be achieved by maximizing the following criterion with S_B the between scatter matrix:

$$J(W) = \frac{W^T S_B W}{W^T S_W W}.$$

The direction w that maximizes this criterion can be shown to be [4]

$$W = S_W^{-1}(m_1 - m_2)$$

where m_1 and m_2 are the means for the two classes. Fisher discriminant linear analysis (FLDA) is a sub-optimal classifier for two classes when their distribu-

tions are Gaussian (<http://www.ece.osu.edu>). FLDA has been found to produce very good and robust classification performance in a number of BCI related publications [20, 23].

Adaptive boosting (Adaboost) The Adaboost method was proposed by Schapire (<http://www.boosting.org>) in 1996. The principal of the Adaboost is that a committee machine can adaptively adjust to the errors of its components—the so-called weak learners. The classification rate of each weak learner should exceed more than 50% [15]. Neural networks with one hidden layer neuron (with the sigmoid activation function) are selected as weak learners. First, the first neural network (weak learner) trains with the equal weight vector of input samples:

$$D_i(i) = 1/N \quad (i = 1, \dots, N),$$

where N is the number of input samples and D is the weight vector of the i th sample in the first iteration. Then the next weak learner trains, but the weight vector of the samples which were misclassified by the previous weak learner is changed by the following relation:

$$\alpha_i = 0.5 \log \frac{1 - \text{error}(i)}{\text{error}(i)}$$

$$D_{t+1}(i) = \frac{D_t(i) e^{-\alpha_i F_t(x(i)) y(i)}}{z_t}$$

where z_t is a normalization factor chosen so that D_{t+1} will be a distribution. α_i is the weight of the i th weak learner output, $y(i)$ is the label of the i th sample, and $F_t(x)$ is the function of the t th weak learner. For each time point, we have a boosting process which increases the weight vector $D(i)$ of those samples that have a bigger error. These samples are considered in the next iteration with more sensitivity to learn the next weak learner. We expect that the next weak learner is trained so that the errors for the previous misclassified samples decrease in this stage. This iterative procedure repeats till the time that n reaches T (the maximum value considered for the number of weak learners). Finally, the output $\varphi(x)$ is a weighted sum of the weak learners:

$$\phi(x) = \text{Sign}(f(x)) = \text{Sign} \left(\sum_{t=1}^T \alpha_t F_t(x) \right)$$

The t th weak learner (classifier) has a vote α_t , and the final outcome of the classifier is a weighted vote of the weak learners. For this classifier, all of the feature values must be normalized in the interval $[-1, 1]$. Thus

the features were normalized by their Euclidian norm. A schematic diagram of the mentioned Adaboost is shown in the Fig. 2.

5 Evaluation of classification performance

Features were extracted from five subjects and applied to the FLDA, SVM and Adaboost classifiers. The classifiers were evaluated by ten times, ten folds cross validation. Best classifiers from the evaluation phase were selected and applied to the test data. It is important to note that the parameters of all methods for feature extraction and classification were adjusted solely on the training data. Totally 360 trials were available for each subject that 240 trials were selected as train features for cross validation process and the rest (120 trials) are selected for testing the methods.

6 Genetic algorithm

Genetic algorithms are widely used in the different fields for optimization [9]. In our case, the best combination of the features was found by a genetic algorithm. For combining the mentioned features (BP with two dimensions, AAR with six and FD with one for each channel) were gathering in one feature matrix. Then, for three channels our feature vector has a dimension of 27. The weights encoded in the chromosomes of the GA were used to linearly combine the features. The cross-validated classification rate calculated from the training set was used as fitness function. The characteristics of the standard genetic algorithm

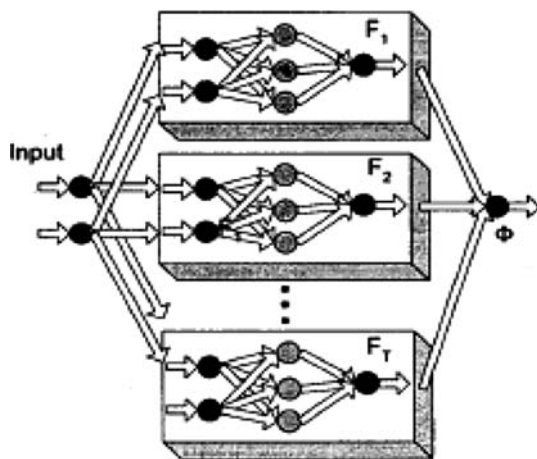


Fig. 2 A committee of the neural networks generated using Adaboost. F_1 , F_2 , and F_T are the output of weak learners and the output $\varphi(x)$ is the total output of the Adaboost [15]

[9] were: mutation rate 3%, roulette wheel selection, one point cross over, 500 generations, and the number of chromosomes for the initial population was 50 [11].

7 Results

The classification results are generally produced in two different ways. In the first manner, one of the features (like BP or FD or AAR) is applied to one of the classifiers (such as FLDA or SVM or Adaboost) referred to here as the combinations of single feature-classifier. In the second way, all of the mentioned features by two methods are applied to the classifiers. In the first method, all of the mentioned features are arranged in one feature vector without any weight (simple combination of the features) while in the second method, feature vectors are weighted by genetic algorithm (genetic combination of the features) and all of these two types of combination of the feature vectors are applied to the classifiers. In the single feature-classifier combinations, bandpower with SVM and FLDA led to the best results for all of our subjects in the cross validation phase. But in the test phase when the best classifiers, determined by the cross validation phase, were applied to the corresponding features of the unseen test data, the results were slightly changed. All of the test results with the whole of mentioned combinations for all of our subjects are shown in Table 1. The results include the minimum classification error and the latency of minimum error through our paradigm. Also, the classification results through our paradigm for all of our subjects in the single feature-classifier and combination of the features are shown, respectively in Figs. 3 and 4. In the test phase, the results of single feature-classifier despite the similarity of FLDA and SVM results in the cross validation phase, FLDA shows better and more reliable performance than SVM for four subjects but for the fifth case (f8), SVM with fractal dimension led to zero percent error. Table 1 indicates that the combination of BP and FLDA gives better results than other combinations for the subjects o3, o8 and g8. FD with FLDA shows a better result than BP and FLDA combination for the subject L1. In the second approach, simple feature combination without any weight (FC) and feature combination with genetic algorithm (FCGA) methods are evaluated. Genetic combination of the mentioned features showed better results for four cases (o3, o8, g8, and f8). In the case of o3, feature combination results showed a considerable reduction of classification error in comparison with the individual features with the SVM and Adaboost classifiers, but the combination of

Table 1 Minimum error rates and the corresponding latencies for the subjects L1, o3, o8, g8, and f8

Subject	Feature	FLDA		SVM		Adaboost	
		Error (%)	Latency (s)	Error (%)	Latency (s)	Error (%)	Latency (s)
L1	BP	28.57	4.75	27.38	4.25	33.33	4.25
L1	AAR	26.19	4.75	26.19	4.75	22.62	4.00
L1	FD	20.24	4.00	23.81	4.00	25.00	4.00
L1	FC	36.90	4.00	36.90	4.25	29.76	4.25
L1	FCGA	14.33	4.75	25.00	4.25	15.21	4.75
O3	BP	9.59	5.75	19.8	4.50	24.66	4.50
O3	AAR	28.77	4.50	30.14	6.00	30.14	6.00
O3	FD	21.92	4.25	26.03	4.25	30.14	5.50
O3	FC	38.10	4.75	28.75	5.00	30.50	4.75
O3	FCGA	13.69	5.00	15.07	5.50	12.32	5.00
O8	BP	17.14	6.25	21.90	4.00	20.00	4.00
O8	AAR	23.81	4.25	27.62	4.50	25.71	6.00
O8	FD	22.86	5.00	21.90	5.00	27.38	3.75
O8	FC	14.29	4.50	18.10	5.75	12.38	4.75
O8	FCGA	13.33	5.50	12.38	5.50	13.33	5.50
G8	BP	16.43	6.25	16.43	5.00	20.71	5.00
G8	AAR	22.14	6.25	30.00	5.75	27.14	6.00
G8	FD	22.86	5.50	21.43	5.50	27.86	5.50
G8	FC	23.57	5.25	31.43	6.25	27.86	4.75
G8	FCGA	11.14	5.00	12.86	4.75	18.70	6.25
F8	BP	10.10	5.75	6.12	5.75	10.20	5.75
F8	AAR	15.00	6.50	15.00	8.00	18.00	5.75
F8	FD	12.24	6.25	0.00	5.25	4.00	6.25
F8	FC	15.60	5.50	14.60	6.50	19.00	6.50
F8	FCGA	09.61	5.50	0.00	5.75	2.50	6.00

FLDA and BP gave the best results in case of o3. In the case of L1, feature combination by the genetic approach showed a drastic error reduction with FLDA and Adaboost classifiers. For the case o8, a remarkable decrease in the error rates was found by applying the classifiers to the genetic combination of the features. Also, for the case o8, simple combination of the features with Adaboost led to a remarkable improvement. In the case g8, BP with both FLDA and SVM led to the best results for the single feature-classifier combinations. In the case g8, when genetic combination of the features is applied to the different classifiers, classification error was decreased drastically. For the case f8, fractal dimension along with the SVM showed the dramatic decreasing in the error rate (zero) for the single feature-classifier and in the case of combination of the features, genetic combination showed the same results (zero percent error).

8 Discussion

Several informative features along with well-known classifiers have been evaluated in this research. Also, the combination of these features in two ways has been considered to find the best result for each classifier. In the first step, combination of single feature-classifier is analyzed and in the second step, results of simple and

genetic combination of the features with the mentioned classifiers are assessed.

Table 1 indicates that for the three cases of o3, o8, and g8, the best single combination for recognizing the mentioned imagery tasks is BP and FLDA, but this is not the best combination for the case L1 and f8. The reason is that the standard alpha (10–12 Hz) and beta (16–24 Hz) bands are not necessarily optimal for all of the subjects. In fact, these bands should be optimized for each subject. BP in alpha and beta bands is directly associated to the event-related desynchronization neural networks in the brain [6] and therefore BP is a promising method for the BCI based on oscillatory activity. Moreover, both BP and AAR are based on second-order statistics. FD with FLDA for the case L1 and FD with SVM for the subject f8 have shown better results than all the other single combinations. Many scientific publications have shown that the EEG signal shows a chaotic behavior, therefore, FD is another useful feature for the classification of brain patterns, which has been rarely considered for BCI applications. FD is a promising method in other fields of EEG, such as the detection of epileptic seizures in the EEG [7] and sleep analysis of EEG data [10]. This classification result for L1 and f8 may lead to the assumption that the EEG of these subjects is more complex and chaotic than three other subjects; therefore FD is more informative than the two other features.

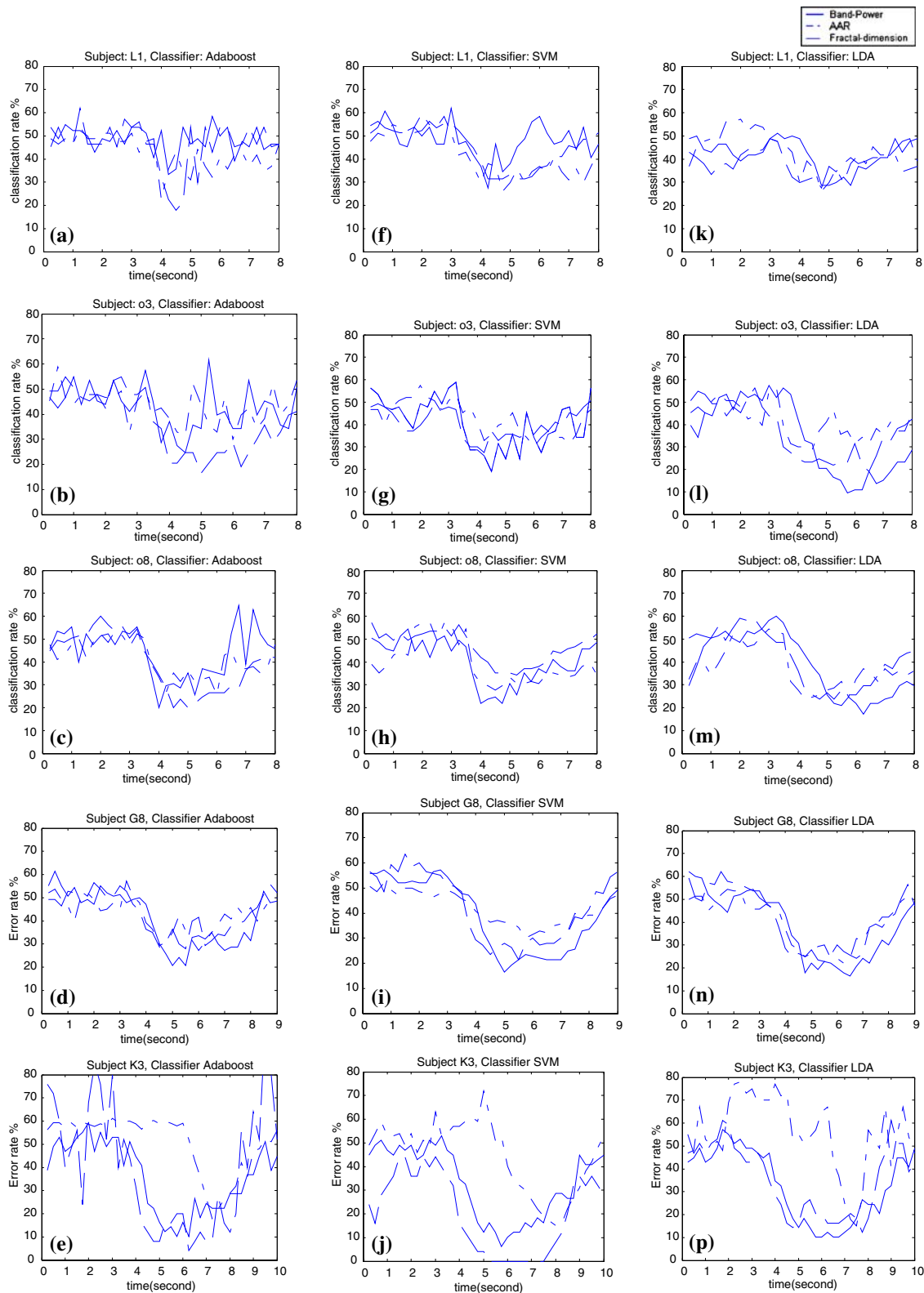


Fig. 3 These figures show the test results (classification results on unseen test data) for the whole time frame of the paradigm and for the subjects L1, o3, and o8. These results are derived by applying FLDA, SVM, and Adaboost classifiers to the band-

power, AAR coefficients, and fractal dimension features. **a–e** show the results of Adaboost, **f–j** show the results of SVM and **k–p** show the results of FLDA on the mentioned features for the cases L1, o3, o8, g8 and f8, respectively

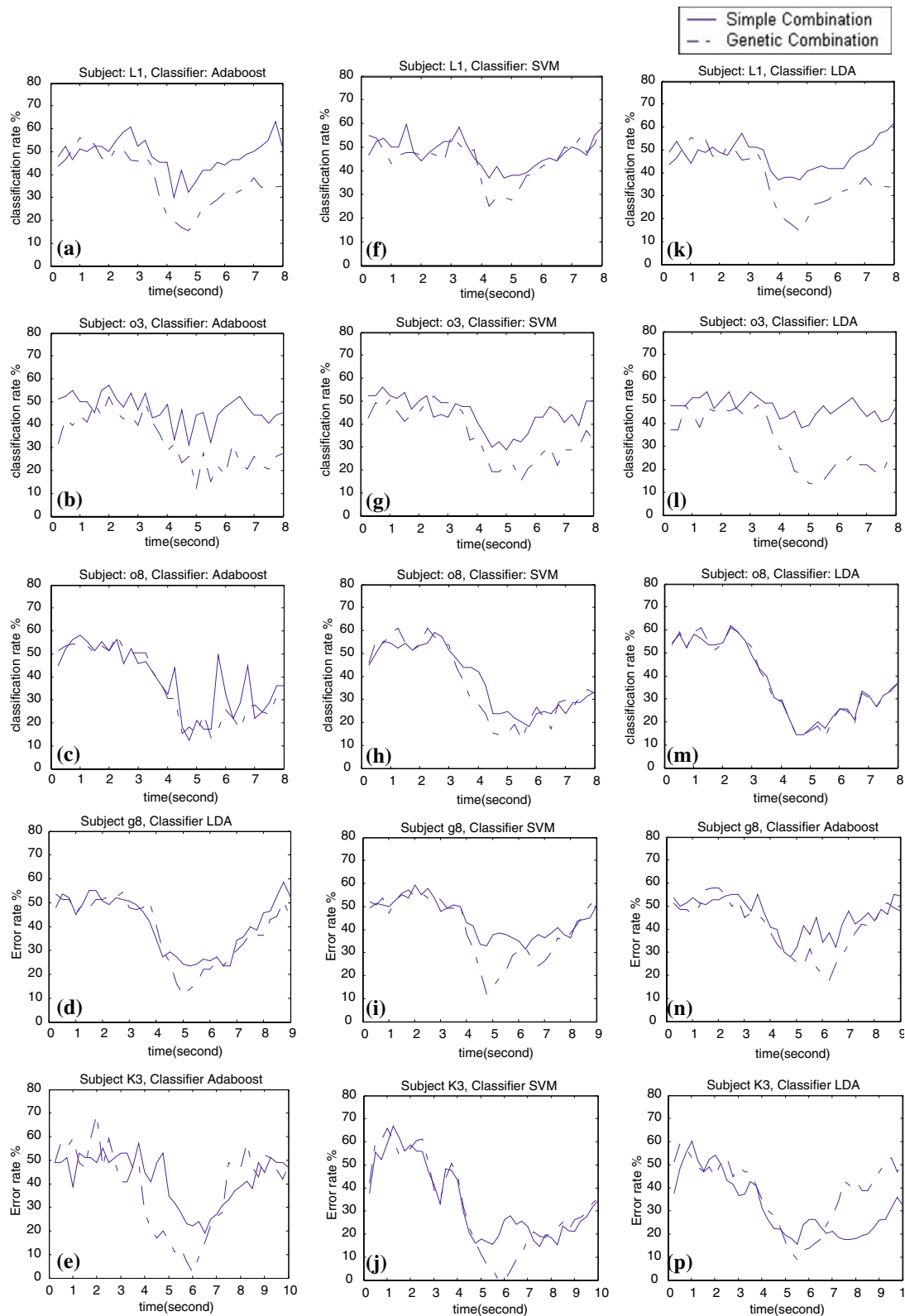


Fig. 4 These figures show the test results (classification results on unseen test data) for the whole time frame of the paradigm and for the subjects L1, o3, and o8. These results are derived by applying FLDA, SVM, and Adaboost classifiers to the band-power, AAR coefficients, and fractal dimension features.

a–e show the results of Adaboost, **f–j** show the results of SVM and **k–p** show the results of FLDA on the simple and genetic feature combinations for the subjects L1, o3, o8, g8, and f8, respectively

Another crucial factor is the latency of the minimum error in the cue-based BCI. For instance, in on-line BCIs, the subjects should be provided with a reliable and stable feedback of their performance as soon as possible, because such a feedback can improve the overall performance of a BCI system considerably. Therefore, not only the classification performance but also how fast the minimum error rate can be achieved (latency) is important. If we consider these two criteria together, FLDA and BP in the case of o3 do not yield the best performance, because the minimum error rate occurs at second 5.75. The significant property of FD is that it reports the changes in the mental state fast enough with an acceptable classification rate. It is important to note, that all subjects of this study were already experienced in BCI training. In fact, they participated in previous BCI feedback training sessions based on BP features. It might be possible, that a specific feedback training based on FD would also change the classification results in favor of FD.

AAR coefficients are often used in BCI applications. Sample by sample adaptation of the coefficients makes this feature especially suitable for online realizations, because no further parameters such as window size and form of the window have to be determined. The combination of AAR and FLDA in all cases gives a fast report in the EEG mental changes, but the error rates are high compared to other combinations of feature and classifier.

In the second step when the features are combined in the two different ways (simple and genetic combinations), the results were dramatically improved except in the case o3. Also, in the case f8, genetic combination has the same result with FD and SVM. Therefore, it can be claimed that genetic algorithm can combine and improve (by giving the weight) the positive discriminative points of the features in order to increase the classification rate. This concludes our interpretation of the features behavior.

The classifiers capabilities based on the results of this study can be summarized as follows: FLDA is faster than the two other classifiers. In the test phase, SVM showed a better result than FLDA and for the case f8, SVM and FLDA results are the same. Compared to FLDA, SVM is more complicated to train. SVM has some parameters like C (that limits the Lagrange coefficient) and α (variance of Gaussian kernel), which have to be optimized for the specific classification task. This selection of the SVM parameters was done by a simple bisectional algorithm, which found the nearly optimum value in an interval. Therefore, SVM is more time consuming than FLDA.

SVM with the genetic feature combination showed the best result for the cases o8 and f8.

In some specific cases when the boundary space between the classes in the feature space is complicated, Adaboost can act more efficiently than the two other classifiers. Adaboost in the cross validation phase is much more time consuming than FLDA and SVM because it requires considerable time to train its weak learners (neural networks) and to find the best value of T (number of weak learners). We also considered SVM as weak learner instead of a neural network in the Adaboost construction, but the results did not change compared to the normal SVM. A reason may be that Adaboost can improve the performance of weak classifiers and it is not suitable for strong classifiers like SVM.

In the test phase, for the individual combination of features and classifiers, in the subject o8, Adaboost behaves similarly to FLDA and SVM, and even in the case f8, Adaboost showed better results than FLDA. On the whole, Adaboost showed higher capabilities when features are combined by genetic algorithm.

Guger et al. [12] has stated that BCI cannot work for all people and shows that about 10% of all users cannot achieve control (only 50% accuracy in a second-class task, i.e. same performance than random). In many articles [11, 17], results show that the classification rate is not a constant value for all cases and it depends on many factors such as learning ability of the subject, intrinsic characteristics of subject's EEG, etc. For instance, some subjects have weak alpha and beta bands amplitude although they can have a good learning ability but the intrinsic characteristics of their EEG do not let them to be considered as good cases for BCI application. No article could prove that one feature is good for all BCI subjects. Some features work very well with some classifiers in some subjects but in some other subjects, other features might work better. Therefore if we add some other subjects, the results of features and classifiers in the different subjects will not definitely confirm each other.

To summarize, it can be stated that the combination of features produced by different feature extraction methods can lead to improved classification results. Moreover, the results suggest that incorporating complexity measures such as fractal dimension in the pool of features suitable for BCI is useful because of the improved achievable classification latency. Therefore, fractal dimension is a good alternative for those cases that their EEG signals have strong chaotic behaviors and their alpha and beta band changes are not significant. Generally, combination of the features is suggested because this reflects more complete information

from different aspects and this combination can be optimized by some methods such as genetic algorithm or simulated annealing for the classification tasks. Although SVM and Adaboost have shown remarkable performance in various applications, for a two class cue-based BCI, Fisher's LDA is preferable. This is because of its robustness and simplicity.

In the future researches, combination of the classifiers via boosting method is suggested in order to have more generalization and robustness and also improve the classification rate between the imagery tasks. Because, when a classifier cannot distinguish the features of different classes, that classifier for that database can be considered as a weak learner. Even efficient classifiers like SVM or LDA for some complex databases (like EEG features of a hard case) are considered as weak learners. Boosting technique can improve the performance and generalization of a weak learner (classifier) and it has been statistically proved that by adding the number of weak learners, error rate exponentially decreases [16, 22]. Nevertheless, one of the efficient factors, which play an important role in the generalization of the results, is to have more recording from each subject that is a very time consuming and expensive task.

References

1. Andrew Webb (1999) Statistical pattern recognition, Oxford University Press, New York
2. Bozorgzadeh Z, Birch GE, Mason SG (2000) The LF-ASD brain computer interface: on-line identification of imagined finger flexions in the spontaneous EEG of able-bodied subjects. *IEEE Int Conf Acoustic Speech Proc* 6:2385–2388
3. Deriche M, Al-Ani A (2001) A new algorithm for EEG feature selection using mutual information. *IEEE international conference on acoustics, speech, and signal processing* 2001. *Proceedings. ICASSP* 2:1057–1060
4. Duda RO, Hart PE, Stok DG (2001) *Pattern classification*. Wiley, New York
5. Esteller R, Vachtsevanos G, Echaz J, Lilt B (1999) A comparison of fractal dimension algorithms using synthetic and experimental data. *ISCAS Proceedings, IEEE international symposium on circuits and systems* 3:199–202
6. Esteller R, Vachtsevanos G, Echaz J, Henry T, Pennell P, Epstein C, Bakay R, Bowen C, Litt B (1999) Fractal dimension characterizes seizure onset in epileptic patients. *Proc IEEE Int Conf Acoust Speech Signal process* 4:2343–2346
7. Esteller R (2000) Detection of seizure onset in epileptic patients from intracranial EEG signals. Ph. D. thesis, School of Electrical and Computer Engineering, Georgia Institute of Technology
8. Flotzinger D, Pregoner M, Pfurtscheller G (1994) Feature selection with distinction sensitive learning vector quantisation and genetic algorithms. *IEEE Int Conf Comput Intell* 6:3448–3451
9. Goldberg (1989) *Genetic algorithm*, Prentice-Hall, New York
10. Goldberger AL, Amaral LA, Hausdorff JM, Ivanov P, Peng CK, Stanley HE (2002) Fractal dynamics in physiology: alterations with disease and aging. *Proc Natl Acad Sci USA* 99(Suppl 1):2466–2472
11. Graimann B, Huggins JE, Levine SP, Pfurtscheller G (2004) Toward a direct brain interface based on human subdural recordings and wavelet-packet analysis. *IEEE Trans Biomed Eng* 51(6):954–962
12. Guger C, Edlinger G, Harkam W, Niedermayer I, Pfurtscheller G (2003) How many people are able to operate an EEG-based Brain–Computer interface (BCI)? *IEEE Trans Neural Syst Rehab Eng* 11(2):145–147
13. Haselsteiner E, Pfurtscheller G (2000) Using time-dependent neural networks for EEG classification. *IEEE Trans on Rehab Eng* 8:457–463
14. Kalcher J, Flotzinger D, Pfurtscheller G (1992) A new approach to a Brain–Computer-interface (BCI) based on learning vector quantization (LVQ3). *Proc Annu Int Conf IEEE* 4:1658–1659
15. Murphey YL, Zhihang Chen, Hong Guo (2001) Neural learning using AdaBoost. *Proceedings IJCNN, International joint conference on neural networks* 2:1037–1042
16. Murua A (2002) Upper bounds for error rates of linear combination of classifiers. *IEEE Trans Pattern Anal Mach intell* 24(5):591–602
17. Obermaier B, Neuper C, Guger C, Pfurtscheller G (2001) Information transfer rate in a five-classes Brain–Computer interface. *IEEE Rehab Eng* 9:283–288
18. 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 on Rehab Eng* 6:316–328
19. Pfurtscheller G; Lopes da Silva (1999) Event related desynchronization. *Hand book of electroenceph. and clinical Neurophysiology*. vol 6 revised edition. Elsevier, Amsterdam
20. Pfurtscheller G, Neuper C (2001) Motor imagery and direct Brain–Computer communication. *Proceedings of the IEEE* 89:1123–1134
21. Schlögl A, Flotzinger D, Pfurtscheller G (1997) Adaptive autoregressive modeling used for single-trial EEG classification. *Biomed Tech* 42:162–167
22. Vapnic VN (1998) *Statistical learning theory*. Wiley, New York
23. Wolpaw JR, McFarland DJ, Vaughan TM (2000) Brain–Computer interface research at the Wadsworth Center. *IEEE Trans Rehab Eng* 8:222–226