Plausibility Assessment of a Subject Independent Mental Task-Based BCI Using Electroencephalogram Signals

S.Hatamikia[*1]
[*1]Department of biomedical engineering
Science and Research branch, Islamic Azad University
Tehran, Iran
[*1]shatamikia11@gmail.com

A.M Nasrabadi[2]
[2]Department of biomedical engineering
Shahed University
Tehran, Iran
[2]nasrabadi@shahed.ac.ir

N.Shourie[3]
[3]Faculty of Technology and engineering
Central Tehran branch
Islamic Azad University
Tehran, Iran
[3]nas.shourie @iauctb.ac.ir

Abstract—In this research, we study the possibility of designing a mental-task based subject-independent Brain Computer Interface (BCI) using Electroencephalogram (EEG) signals. Due to major differences in the EEG signal of individuals during different mental tasks, designing a universal BCI seems impossible. Hence, almost all the previous studies concentrated on designing custom-based Brain Computer Interface systems (BCIs) which are appropriate to be used by only one particular subject. In order to overcome this limitation, this paper presents an efficient subject-independent procedure for EEG-based BCIs. The main aim of this research is to develop ready-to-use BCIs that can be applicable for all users. To achieve this goal, three feature extraction methods including Autoregressive modeling, Wavelet transform and Power spectral density were applied; then, a new method based on Genetic Algorithm (GA) wrapped Self Organization Map (SOM) feature selection was used to select the most related features with the use of leave-one-subject-out cross-validation strategy. According to the experimental results, the proposed algorithm based on GA wrapped SOM feature selection is an efficient method for designing subject-independent BCIs and is able to distinguished different cognitive tasks of different individuals, effectively.

Keywords: Brain Computer Interface (BCI; subject-independent; Genetic Algorithm Wrapper(GA-Wrapper); Self Organization Map(SOM)

I. INTRODUCTION

Brain computer interface systems (BCIs) are communication systems which allow to generate control orders with the only help of the thoughts [1]. The main goal for designing such systems is to help people with neuromuscular disabilities to have effective control over external appliances such as assistive devices or computers [2]. In the recent years, BCI based on Electroencephalogram (EEG) signals derived from the scalp electrodes, is getting interesting more and more because of its rather simple and inexpensive recording process. To date, numerous researches have been accomplished in order to design custom based or subject depended BCIs using EEG signals. A custom based BCI needs the days of training on the subject before using the system, and finally the provided system is usable only for the same subject. Although, for designing such a system a lot of time and cost is needed, but the prepared system is utilizable by just one particular person. A subject independent or general BCI system could overcome such inconvenience. In such a system, unlike the custom based BCI systems, it is not needed to design a particular BCI for each subject, but a universal system is designed which is usable for all the users. In this case, the designed general BCI can be used for each untrained subject; so, all the costs and time which are required for training of a new subject will be eliminated. Some previous studies have been done to develop ready-to-use BCIs for unseen subjects [3, 4]. In these studies, the proposed ready-to-use BCIs are efficient only for one or two untrained subjects; but the main goal of our suggested general BCI is to design ready-to-use BCIs which are utilizable for all untrained subjects. In a study, a subject independent BCI based on motor imagery was presented [5]. In this research, different motor imagery tasks were classified efficiently using other subjects database and an event related de-synchronization (ERD)-based BCI system was suggested as the first attempt to design a subject-independent BCI based on Common Spatial Pattern (CSP) method.

In our study, a subject independent BCI based on mental task is suggested. For this aim, we used the data of Keirn and Aunon which are consists of the EEG signals of five different mental tasks of 7 subjects. Since, several researches have employed this database, but all of them introduced a custom-based BCI for a particular participant; while, our aim is conducting a general BCI which is usable for all users and each untrained subject can use it. With this aim, the combination of three extracted features based on Wavelet transform, Autoregressive modeling and Power spectral density is applied and a novel feature selection algorithm based on Genetic Algorithm (GA) wrapped self-organization map (SOM) is used. This feature selection algorithm searches for the best feature subsets with a wrapper approach. GA, the search strategy for wrapper algorithm, tries to find the most informative subsets of feature which result the best SOM classification results. This technique has lower computational

load compared to the other search techniques such as exhaustive search, forward and backward searching. Furthermore, this random search algorithm is stable against sticking in local maxima [6]. In the proposed general BCI, the data of all 7 subjects are divided into training and test dataset and our model is simulated by eliminating each individual subject completely from the training dataset and by testing on that subject. The performance of SOM classifier is then evaluated by employing it on hold out subjects by leave-one-subject-out cross validation strategy.

The rest of this research is summarized as follows. Section II explains about data collection and subjects. Section III presents the research methodology include: feature extraction, feature selection and classification methods. Section IV presents the experimental results and discussion and conclusions of this current research are given in and section V.

II. DATA COLLECTION

In this study, the EEG data which were gathered by Keirn and Aunon in Colorado State University are used [7]. Six electrodes were placed on different position (O1, O2, P3, P4, C3 and C4 based on the International 10-20 System) and two reference electrodes (A1 and A2) were electrically linked mastoids. Fig.1 shows the positions of different electrodes. The sampling frequency was considered at 250 Hz. The maximum impedance of all electrodes during recording was set bellow 5 kΩ and the band pass filters (0.1-100 Hz) were applied to the signals. The EEG data of 7 subjects were recorded. In the experiments, 5 different mental tasks were considered include: multiplication (solving a simple multiplication problem, mentally), rotation (rotating a 3D object around an axis, mentally), letter composing (writing a letter to a special subject, mentally), counting (thinking of a sequence of numbers) and baseline (relaxing).

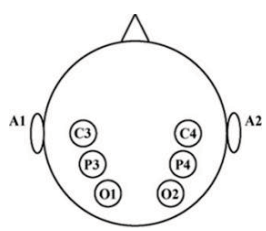

Figure 1. The position of electrodes according to International 10–20 System

TABLE 1. THE NUMBER OF TRIALS AND SESSIONS FOR EACH PARTICIPANT

Subject number	Number of trials For each Mental task	Number of sessions
1	10	2
2	5	1
3	10	2
4	10	2
5	15	3
6	10	2
7	5	1

Each individual session contains five trials, which each trial lasts 10-second and is recorded of five different mental tasks. Totally, in one session, 25 trials (5 for each task) were performed for each subject. The number of completed sessions differed in subjects and the sessions were done on different days. The complete information about subjects and the number of their recorded sessions and trials are presented in Table I.

III. METHODOLOGY

A. Feature extraction

1) Power spectral density

Normally, Welch algorithm and Fast Fourier Transform (FFT) are applied for calculating power spectral density (PSD) from EEG subbands. In this study PSD is estimated by Welch method. In this algorithm, signals are divided into L overlapping segments of length N. The window W is applied to each data segment which $x^{(k)}[n]$ is defined the k_{th} segment of signal $X[N]$ .Therefore, the periodogram is computed for each segment as bellow:

$$p_d{}^{\wedge}(f) = \frac{1}{NU}\left|\sum_{n=0}^{N-1} w[n]\ x^{(k)}[n]e^{-j2\Pi fn}\right|^2, \quad k = 1.....K \tag{1}$$

Where, U is a normalized factor which is defined as:

$$U = \frac{1}{N}\sum_{n=0}^{N-1}|w[n]| \tag{2}$$

The periodograms are then averaged and welch's Power spectrum of signal is defined as:

$$p_{welch}{}^{\wedge}(f) = \frac{1}{k}\sum_{k=1}^{K} P_d{}^{\wedge}(f) \tag{3}$$

In this study, PSD from different subbands: delta and theta (0.5-8 HZ), alpha (8-13 HZ), beta (13-30 HZ) and gamma (30-50 HZ) are extracted as features from different channels.

2) Autoregressive feature

The autoregressive (AR) model for signal $y[n]$ is determined with following equation:

$$y[n] = \sum_{k=1}^{p} a_k y[n-k] + e[n] \tag{4}$$

Where, a_k is the AR coefficients and p is the order of model. Furthermore, $e[n]$ is a random signal with a limited variance and a zero mean that is independent of past samples of the time series. The goal is to compute the a_k coefficients using the bounded values of the signal $y[n]$. In our research, the Burg algorithm is used for calculating the AR coefficients and then, these scalar AR coefficients are used as the features. This algorithm is the most common algorithm for computing

the AR coefficients. Previous studies using this database, have demonstrated that the model order 6 is the best model order for mental task classification [8]; so, we select the model order 6, and six features of six AR coefficients are computed for each EEG signal.

3) Wavelet transform

Discrete wavelet transform (DWT) is a suitable tool for the analysis of signals with the transitory or non-stationary properties. In this study, DWT is used for the time-frequency representation of EEG signals. In wavelet analysis, various functions may be applied by contracting, expanding and shifting a mother wavelet function which the mother wavelet function $\varphi_{a,b}(t)$ is specified as follow:

$$\varphi_{a,b}(t) = \frac{1}{\sqrt{a}}\varphi(\frac{t-b}{a}) \quad (5)$$

Where, b and a are shifting and scaling factor respectively. DWT performs a time- frequency decomposition of the signal using a group of filters that are called filter banks and divides them into a set of approximation (A) and detail (D) coefficients at different levels. In this study, we use wavelet function of Daubechies with order 4 to decompose the signals into 5 levels related to five different frequency bands: D2 decomposition (31.25-62.5 Hz), D3 decomposition (15.63-31.25 Hz), D4 decomposition (7.81-15.63 Hz), D5 decomposition (3.91-7.81 Hz) and A5 decomposition (0-3.91 Hz). In our study, we extract 7 statistical features from the components of 5 mentioned decompositions include:

1) Maximum of the coefficients values.
2) Minimum of the coefficients values.
3) Standard deviation of the coefficients values.
4) Mean of the coefficients values.
5) Power of coefficients values.
6) Ratio of energy of the decompositions to the total energy of all subbands, which is named Recoursing Energy Efficiency (REE).

$$REE_{sub-band} = \frac{E_{sub-band}}{E_{total}} \quad (6)$$

7) Absolute logarithm of REE, which is named Absolute Logarithmic REE (ALREE).

$$ALREE_{sub-band} = abs(\log_{10}[\frac{E_{sub-band}}{E_{total}}]) \quad (7)$$

These features are computed for each one of the five decomposition levels. Therefore, for each channel, 35(=7*5) wavelet features are used as the input of classifiers.

B. Feature Normalization

To avoid neglecting the effect of features with low numeric ranges and also to reduce numerical difficulties between participants during the classification, each feature is independently normalized by applying follow equation:

$$x_i^{norm} = \frac{x_i - \mu_x}{\sigma_x} \quad (8)$$

Where, μ_x and σ_x are the average and standard deviation over all samples, respectively. Furthermore, x_i^{norm} is a normalized vector of each independent feature.

C. Feature Selection

High-dimensional feature space makes the classification procedure less reliable and more complex. Since, some of features are irrelative with the classification process and lead to unsatisfactory classification accuracy. Therefore, there is the need of selecting and reducing the number of features to achieve better performances.

Feature selection algorithms are mainly divided into two groups: filter algorithms which select the best subsets of features before reaching to classification process and wrapped algorithms that perform the selection of feature subsets around the classification task [9]. In this research, GA-Wrapper feature selection is applied to find optimal subsets of features. GA Wrapper is a stochastic-based search strategy based on the classification accuracy [10]. GA wrapper feature selection has shown higher performance compared to the other feature selection methods based on wrapper approach [11]. In this study, we use SOM classifier for wrapper approach. We use leave-one-subject out cross validation to avoid of overfitting and also to evaluate the performance of classification procedure. At first, the data of the all subjects were divided into 7 folds (7 folds are chosen because of 7 subjects), which each fold includes all trials of a subject. The total data except the data of one fold were dedicated as training dataset and the remaining were dedicated as testing dataset. Afterward, the data specialized for training, were also divided into 6 folds and all this data except the data of one fold is dedicated as the train data and the remaining were dedicated as the validation dataset for Wrapper approach (see Fig.2). This procedure is repeated 7 times for 7 different training dataset and the average accuracies over cross validations and testing data were calculated. The steps of GA-SOM wrapper are shown in the following:

1. Divide data into 7 fold (6 fold for training and one fold for testing)
2. Generate an initial population of 200 chromosomes (subsets of features)
3. Compute the features associated with each chromosome from the training dataset
4. Assign 5 fold of the training dataset to train data and one fold to validation data
5. Repeat 6-folds cross validation 7 times for different training dataset
6. Compute the average accuracy of cross validations by SOM classifier as fitness of chromosomes
7. Select those chromosomes which have highest fitness

8. Perform cross over with probability of 0.8 on selected chromosomes at step7
9. Perform mutation with probability of 0.03 on selected chromosomes at step7
10. Go to step2 (up to 20 generations)
11. Select the best chromosomes with maximum fitness
12. Compute the selected features for testing dataset
13. Compute average accuracy on testing data for 7 folds

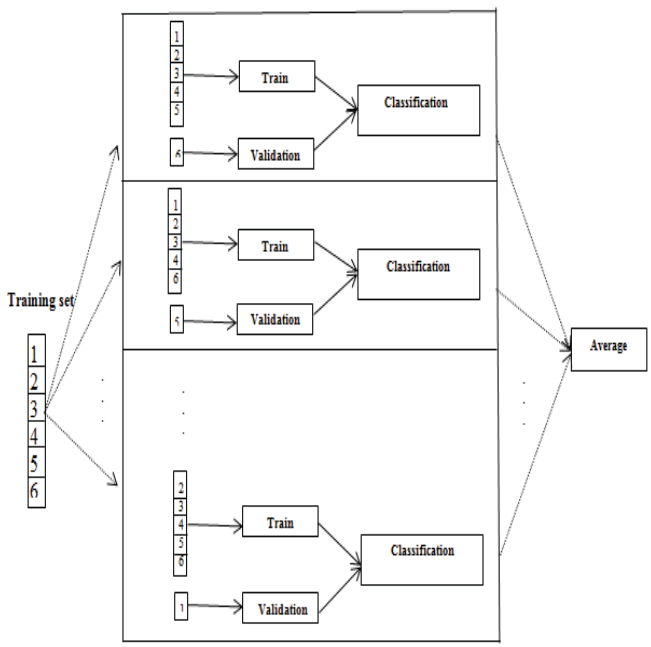

Figure 2. 6-fold cross validation for accuracy estimation of one subset of feature with wrapper approach

D. *Classification with Self-organization map*

The self-organization map (SOM) algorithm is one of the well-known unsupervised method which is widely applied in vector quantization, time series prediction and clustering. Furthermore, this network is used as a powerful classification tool in various fields. In SOM network, each neuron i is specified by a vector of weights w_i with a dimension equal to the dimension of the input data. First, the winner neuron $i^*(t)$ is chosen according to the input data $x(t)$ as shown in :

$$i^*(t) = \arg\min\{g_i(x(t))\}, \quad where, g_i(x(t)) = \|x(t) - w_i(t)\| \tag{9}$$

Then, the vectors of weight should be updated according to follow equation:

$$\Delta w_i(t) = \alpha(t) h(i*, i; t)[x(t) - w_i(t)] \tag{10}$$

Where, $h(i^*, i; t)$ is the neighborhood function which is defined:

$$h(i*, i; t) = \exp(-\|r_i(t) - r_{i^*}(t)\|^2 / \sigma^2(t) \tag{11}$$

In the output layer, the distance between the place of the neurons i and i^* is $\|r_i(t)-r_{i^*}(t)\|$. Furthermore, σ(t) and α (t) are the neighborhood radius parameters and the learning rate, respectively. For the classification based on SOM, the class labels of the output neurons have to be determined, which is done after the training process and is called the labeling phase. At first, the winner neuron is specified for each one of the training vectors; afterward, the class labels of the training vectors those chose neuron i as the winner neuron are determined. Finally, the label of i_{th} winner neuron is specified according to the most frequent class labels of these selected training vectors. In the testing stage, the class labels of each new sample is determined according to the class label of it's selected winner neuron.

In our study, the initial neighborhood size is considered 3 and after 100 steps it shrinks to 1. Furthermore, the initial α (t) is selected 0.9 and decreases 0.01 by each step. In order to ensure the reproducibility of classification results, SOM network was run 3 times and the average accuracies were considered.

We used three performance metrics to evaluate the results of suggested system which include: sensitivity, specificity and accuracy. These parameters are determined using the confusion matrix parameters including FP (false positives rate), TP (true positives rate), FN (false negatives rate) and TN (true negatives rate) as follow:

$$Sensitivity = \frac{TP}{TP + FN} \times 100 \tag{12}$$

$$Specificity = \frac{TN}{TN + FP} \times 100 \tag{13}$$

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100 \tag{14}$$

IV. EXPERIMENTAL RESULTS

In this research, the EEG signals of seven participants were used. For each particular subject, the EEG data is consists of a number between 5 to 15 trials. Totally, there are 65 trials of all subjects for each mental task. In our experiments we applied all 65 trials to design our proposed model. Our goal was to develop a general BCI which is able to recognize the mental states of different subjects. For this approach, all 5×65 signals (5 mental tasks × 65 trials) were divided into the train and test dataset. The training dataset was applied to train the classifier and also to determine the most informative feature subsets and the test dataset was applied to calculate the performance of the proposed model. Several features include: 4 power spectral features, 35 wavelet features and 6 AR coefficients (totally 45 features) were used. We extracted these feature from all (6) channels; so the constructed feature vectors were consist of 270(=45*6) features. The GA wrapped SOM classifier was used as feature selection algorithm to select the most discriminatory features. The test feature

vectors were then computed according to the best selected features. The performance of our suggested system was evaluated by eliminating all the trials of an individual subject, using the remaining trials as training dataset and using this subject's data as testing dataset. Finally, average accuracies were achieved using leave-one-subject-out cross validation

We considered 10 different pairs of mental tasks for classification of cognition tasks. The cross validation and testing results of proposed method based on GA wrapper coupled with SOM are shown in Table II and III, respectively. The results of the experiments were assessed with three statistical indices include: sensitivity, specificity and accuracy. According to the results of Tables II and III, the cross validation results are in line with the testing results; furthermore, almost in all cases, there is a little difference between the sensitivity and specificity results. These results indicate the robust performance of our proposed method. According to the results in Table III, testing accuracies higher than 65% are achieved for all mental task pairs. Also, the highest testing accuracy obtained for the rotation/letter mental task was 76.52% with the specificity and sensitivity of and 77.17% and 75.78%, respectively.

Table II THE RESULTS OF CROSS VALIDATION FOR DIFFERENT COMBINATION TASKS

Mental task pairs	SOM Classifier (%)		
	Sensitivity	Specificity	Accuracy
Baseline/Multiplication	61.65	69.49	65.57
Baseline/Rotation	70.25	74.47	72.36
Baseline/Count	68.38	63.86	66.12
Baseline/Letter	72.68	60.86	66.77
Multiplication/Rotation	53.24	74.76	64.84
Multiplication/Count	58.94	70.38	64.66
Multiplication/Letter	64.44	70.06	67.25
Rotation/Count	62.50	64.80	63.65
Rotation/Letter	76.48	74.80	75.64
Letter/Count	69.83	60.17	65.00
Average (Standard deviation)	65.84(±6.9)	68.36(±5.6)	67.18(±3.8)

Table III THE RESULTS OF TESTING FOR DIFFERENT COMBINATION TASKS

Mental task pairs	SOM Classifier (%)		
	Sensitivity	Specificity	Accuracy
Baseline/Multiplication	65.37	67.99	66.68
Baseline/ Rotation	72.45	77.89	75.17
Baseline/Count	71.56	68.32	69.94
Baseline/Letter	70.30	65.79	68.09
Multiplication/Rotation	62.28	79.14	70.71
Multiplication/Count	59.60	75.36	67.48
Multiplication/Letter	74.82	68.82	71.82
Rotation/Count	70.16	60.10	65.13
Rotation/Letter	75.78	77.17	76.52
Letter/Count	67.90	69.32	68.61
Average (Standard deviation)	69.02(±5.2)	70.9(±6.1)	70.02(±3.6)

V. DISCUSSION AND CONCLUSION

In current research, we examined the possibility of having a mental task based subject-independent BCI using EEG signals. We have shown that our suggested method is an efficient method for such systems which is able to distinguished different cognitive tasks of different individuals, effectively.

Our aim was to design ready-to-use BCIs that are utilizable for all untrained subjects. To evaluate the classification results, leave-one-subject-out cross validation strategy was used. Because the number of EEG dataset which is used in this study is small, this method was selected among different cross validation methods.

In the current study, the combination of three feature types together instead of using just one feature type was employed to form the feature vectors. According to the experimental results, the selected features with GA-SOM algorithm, were include of all these three feature types. These results show the significant role of using variant features in the obtained performances. Such consequence is resulted from this fact that variant features represent the signals in different domains and provide more useful information about discrimination of cognition tasks.

We used SOM classifier for classification of mental tasks. SOM classifier performs locally on the data space and can prepare a good approximation of the input space; so, it can be useful for distinguishing various data structures with various distributions.

Although suggested general BCI did not achieve very high performances, but the obtained accuracies are promising; since, these results illustrate that it is possible to have a mental task based subject independent BCI using EEG signals. As compared to the only previous study on subject independent BCI designs, our proposed system has shown better performances. The average testing accuracy of our proposed study is %70.02, whereas, the accuracy nearby %50 was achieved for the suggested subject independent BCI based on CSP method [5]. However, the classification performances must be increased higher for real situations. In future, there is an intention to develop our proposed system with a larger database; since, using a larger dataset can be useful to obtain more accurate system. Furthermore, we will try to examine other subject-independent features that are able to eliminate the inner-subjective variability and improve the performance of the general system.

REFRENCES

[1] Z. A. Keirn and J. I. Aunon, "A new mode of communication between man and his surroundings," IEEE Trans. Biomed. Eng., vol.37, pp.1209–1214, December 1990.

[2] B. Blankert, G. Dornhege, M. Krauledat., K-R. Müller, V. Kunzmann, F. Losch.,et al. The Berlin Brain-Computer Interface: EEG-based communication without subject training. IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 14, pp.147_152, 2006.

[3] B. Blankert, G. Dornhege, M. Krauledat., K-R. Müller. The Berlin Brain_Computer Interface: Accurate performance from first-session in BCI-naïve subjects. IEEE Transactions on Biomedical Engineering, vol.55, pp. 2452_2462, 2008.

[4] M. Krauledat, M. Tangermann, B. Blankertz , K. Muller Towards zero training for brain–computer interfacing. PLoS One, 3(8), e2967, 2008

[5] S. Fazli, F. Popescu, M Danóczya, B. Blankertz, K-R. Müller ,C. Grozea, Subject-independent mental state classification in single trials, Neural Networks, vol. 22, pp. 1305_131, 2009.

[6] K. Tavakolian, A.M. Nasrabadi, A.M. S. Rezaei, Selecting better EEG channels for classification of mental tasks, ISCAS, Proceedings of the International Symposium on Circuits andSystems, vol.3, pp.537-540, 2004.

[7] Z.A. Kerin and J. I. Aunon, A new mode of communication between man and his surroundings, IEEE Trans. Biomed. Eng , vol. 37, pp. 1209-1214, December 1990.

[8] R. Palaniappan, R.Paramesran, S.Nishida, N.A Saiwaki, new brain – computer interface design using fuzzy ARTMAP. IEEE Trans Neural Syst Rehabil Eng, vol. 3, pp. 40–48, 2002.

[9] R. Kohavi, G. John, Wrappers for feature subset selection, Artif Intell, vol.1, pp. 273–324, 1997.

[10] Yu. Enzhe, Cho. Sungzoon, Ensemble based on GA wrapper feature selection, Computers & Industrial Engineering, vol. 51, pp. 111–116, 2006.

[11] A. Ozcift and A. Gulten, Genetic algorithm wrapped Bayesian network feature selection applied to differential diagnosis of erythemato-squamous diseases, Digital Signal Processing, vol.23, pp.230-237, 2013.