Volume 4 No. 2, February 2014 ISSN 2223-4985

International Journal of Information and Communication Technology Research

http://www.esjournals.org

Mouse control by Independent Component Analysis and Differential Evolutions

Karim adinehvand[1], Abdolreza Asadi Ghanbari[2], Ehsan Heidari[3]
[1,2]Computer Department of Islamic Azad University, Boroujerd Branch, Iran
[3] Computer Department of Islamic Azad University, Doroud Branch, Iran

ABSTRACT

During the past 10 years, the research pace of brain computer interfaces (BCIs) has quickened greatly because of their potential application value. The goal of a BCI is to provide its users a communication and control channel that do not depend on brain's traditional output pathways of peripheral nerves and muscles. Its potential applications include restoring functions to those with motor disabilities, alarming paroxysmal diseases (e.g., epileptic seizure prediction), manipulating human's control in inhospitable or even dangerous environments, etc. Inherently, research on BCIs is an interdisciplinary field involving neuroscience, psychology, engineering, mathematics, clinical rehabilitation, and computer science.

Feature subset selection is the process of identifying and removing as much irrelevant and redundant information as possible. This reduces the dimensionality of the data and may allow learning algorithms to operate faster and more effectively. In some cases, accuracy on future classification can be improved; in others, the result is a more compact, easily interpreted representation of the target concept. In new BCI systems for increase accuracy, usually used increased number of electrodes. In this case the increased number of electrodes causes a non-linear increase in computational complexity (i.e. decrease transfer rate). In this paper, we attempt to enhance the single trial EEG patterns and Redundancy Reduction using the components obtained by independent component analysis (ICA) for reduction of artifacts and Differential evolution (DE) for Feature subset selection.

Keywords: *Brain Computer Interfaces, Redundancy Reduction, Differential Evolution Algorithm, Blind source separation (BSS).*

1. INTRODUCTION

Scientists have speculated for decades on the possibility of a direct interface between a brain and a machine. The basic ideas were put forward in the 1970s, and some initial experiments were carried out—basically analysing the brain's electrical activity generated in response to changes in gaze direction [1]. Our understanding of how the brain works has increased since, and the recent years have seen the development of prototypes based on these and other principles, showing the possibility to use the brain electrical activity to directly control the movement of robots. Such a kind of brain–computer interface is a natural way to augment human capabilities by providing a new interaction link—i.e. an additional communication channel—with the outside world and is particularly relevant as an aid for paralysed humans, although it also opens up new possibilities in human–robot interaction for able-bodied people.

There is a variety of methods to monitor brain activity, broadly classified as invasive and non-invasive. Most noninvasive BCI systems use electroencephalogram (EEG) signals, i.e., the electrical brain activity is directly recorded from the surface of the scalp, the main source of which is the synchronous activity of thousands of clustered cortical neurons. Although measuring the EEG is a relatively simple and non-invasive method, it does not provide detailed information on the activity of small brain areas and is characterised by noisy measurements and small amplitudes in the range of a few micro-volts.

Besides electrical signals, neural activity produces other types of signals that could be used in a BCI, such as magnetic and metabolic. Magnetic activity can be recorded with magneto-encephalography (MEG), while metabolic activity (reflected in changes in blood flow and the blood oxygenation level) can be observed with functional Magnetic Resonance Imaging (fMRI), Positron Emission Tomography (PET) and optical imaging. However, these alternative techniques require very sophisticated devices that can only be operated in special facilities, making impractical their use for prototyping and practical implementation.

Invasive BCI systems measure the activity of single neurons from microelectrodes directly implanted in the brain. Although these techniques can provide signals that are less noisy than EEG signals and with higher spatial resolution, they still require surgical operations. In some cases, scar tissue might form from the surgical operations, impacting the strength of the electrical signal (as it implies changes of electrical impedance) and requiring repeated surgical treatments to maintain the signal strength [2].

Although there are now a variety of non-invasive methods available for monitoring brain functions and recording electrophysiological signals, such as electroencephalography, magnetoencephalography, positron emission tomography, and functional magnetic resonance imaging (fMRI), electroencephalography is a relatively convenient and inexpensive one [3, 4]. Electroencephalograms (EEGs) are the electrophysiological signals recorded in terms of electroencephalography with electrodes mounted on the human scalp. They are essentially generated by the underlying neurons in the cortex. In this paper, the BCI we study is also based on EEG signals.

International Journal of Information and Communication Technology Research

http://www.esjournals.org

Present-day BCIs have maximum information transfer rates ≤25 bits/min. With this capacity, they can provide basic communication and control functions (e.g. environmental controls, simple word processing) to those with the most severe neuromuscular disabilities, such as those locked in by late-stage ALS or brainstem stroke. They might also control a neuroprosthesis that provides hand grasp to those with mid-level cervical spinal cord injuries. More complex applications useful to a larger population of users depend on achievement of greater speed and accuracy, that is, higher information transfer rates [5].

In new BCI systems for increase accuracy, increased number of electrodes. In this case the increased number of electrodes causes a non-linear increase in computational complexity (i.e. decrease transfer rate). In this paper, we attempt to enhance the single trial EEG patterns and Redundancy Reduction using the components obtained by independent component analysis for reduction of artifacts and Differential evolution for Feature subset selection.

2. MATERIALS AND METHODS

In this research, EEG signal used as the basic data for classification. The EEG data is from an open EEG database of University of Tuebingen. Two types of the EEG database are employed as [6].

2.1 Dataset I

The datasets were taken from a healthy subject. The subject was asked to move a cursor up and down on a computer screen, while his cortical potentials were taken. During the recording, the subject received visual feedback of his slow cortical potentials (Cz-Mastoids). Each trial lasted 6s. During every trial, the task was visually presented by a highlighted goal at either the top or bottom of the screen to indicate negativity or positivity from second 0.5 until the end of the trial. The visual feedback was presented from second 2 to second 5.5. Only this 3.5 second interval of every trial is provided for training and testing. The sampling rate of 256 Hz and the recording length of 3.5s results in 896 samples per channel for every trial. This dataset contain 266 trials that 70% of this dataset is considered as train dataset and the rest are considered as test.

2.2 Dataset II

The datasets were taken from an artificially respirated ALS patient. The subject was asked to move a cursor up and down on a computer screen, while his cortical potentials were taken. During the recording, the subject received auditory and visual feedback of his slow cortical potentials (Cz-Mastoids). Each trial lasted 8s. During every trial, the task was visually and auditorily presented by a highlighted goal at the top or bottom of the screen from second 0.5 until second 7.5 of every trial. In addition, the task ("up" or "down") was vocalised at second 0.5. The visual feedback was presented from second 2 to second 6.5. Only this 4.5 second interval of every trial is provided for training and testing. The sampling rate of 256 Hz and the recording length of 4.5s results in 1152 samples per channel for every trial. This dataset contain 200 trials that 70% of this dataset is considered as train dataset and the rest are considered as test.

2.3 Proposed Methods

The block diagram of the proposed method for EEG signal classification is depicted in Fig.1. The method is divided into sex steps: (1) EEG acquisition and sampling, (2) EEG preprocessing, (3) calculation of feature vector, (4) feature selection, (5) classification [6, 7, 8].

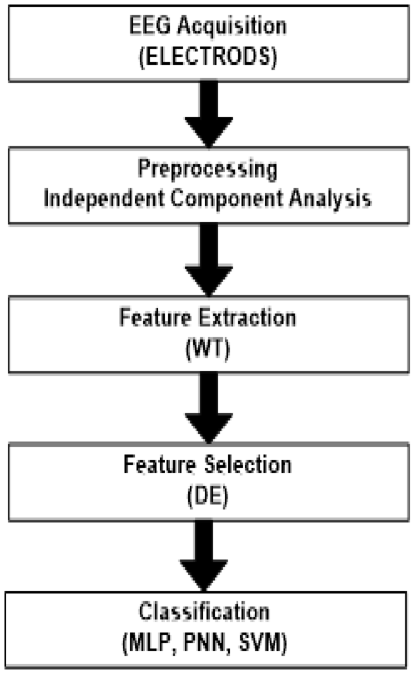

Fig. 1. Block diagram of the proposed method for EEG signal classification

3. PRE PROCESSING

3.1 Artifact Removal

Artifact removal is the process of identifying and removing artifacts from brain signals. An artifact-removal method should be able to remove the artifacts as well as keeping the related neurological phenomenon intact. Common methods for removing artifacts in EEG signals are as follows.

3.1.2. Blind source separation (BSS)

BSS techniques separate the EEG signals into components that ‘‘build’’ them. They identify the components that are attributed to artifacts and reconstruct the EEG signal without these components [10]. Among the BSS methods, Independent Component Analysis (ICA) is more widely used. ICA is a method that blindly separates mixtures of independent source signals, forcing the components to be independent. It has been

International Journal of Information and Communication Technology Research

http://www.esjournals.org

widely applied to remove ocular artifacts from EEG signals [11]. Preliminary studies have shown that ICA increases the strength of motor-related signal components in the Mu rhythms, and is thus useful for removing artifacts in BCI systems [13].

One advantage of using BSS methods such as ICA is that they do not rely on the availability of reference artifacts for separating the artifacts from the EOG signals [9]. One disadvantage of ICA, along with other BSS techniques, is that they usually need prior visual inspection to identify artifact components [12]. However, some automatic methods have been proposed [14].

4. FEATURE EXTRACTION

For features extraction from the raw EEG data many methods such as time domain, frequency domain, and time–frequency domain are used. In this article we used Entropy and Wavelet Transform for feature extraction.

4.1 Wavelet Transform

For features extraction from the raw EEG data many methods such as time domain, frequency domain, and time–frequency domain are used. Since the EEG is non-stationary in general, it is most appropriate to use time–frequency domain methods like wavelet transform as a mean for feature extraction [15]. The WT provides a more flexible way of time–frequency representation of a signal by allowing the use of variable sized windows. In WT long time windows are used to get a finer low-frequency resolution and short time windows are used to get high-frequency information. Thus, WT gives precise frequency information at low frequencies and precise time information at high frequencies. This makes the WT suitable for the analysis of irregular data patterns, such as impulses occurring at various time instances. The EEG recordings were decomposed into various frequency bands through fourth-level wavelet packet decomposition (WPD). The decomposition filters are usually constructed from the Daubechies or other sharp mother wavelets, when the data has discontinuities. In this research, based on the analysis of the data, Daubechies mother wavelet was used in the decomposition. The power spectrum, variance and mean of the signal (each channel) are extracted as features. So the feature set for each subject in each trial consisted of 3*number of channels. As a result, the feature matrix was 266*18 and 200*21 for subject A and B respectively. Finally the feature matrix is normalized.

5. FEATURE SELECTION

Feature selection is one of the major tasks in classification problems. The main purpose of feature selection is to select a number of features used in the classification and at the same time to maintain acceptable classification accuracy. Besides deciding which types of features to use, the weighting of features also plays an important role in classification. Emphasizing features that have better discriminative power will usually boost classification. Feature selection can be seen as a special case of feature weighting, in which features that are eliminated are assigned zero weight. Feature selection reduces the dimensionality of the feature space, which leads to a reduction in computational complexity. Furthermore, in some cases, classification can be more accurate in the reduced space. Various algorithms have been used for feature selection in the past decades. One of the best methods that can be used for features selection is Evolutionary Algorithms [16].

5.1. Differential Evolution

Differential evolution (DE) is a stochastic, population-based search strategy developed by Storn and Price [17, 18] in 1995. While DE shares similarities with other evolutionary algorithms (EA), it differs significantly in the sense that distance and direction information from the current population is used to guide the search process.

In evaluation algorithm e.g. genetic algorithm, variation from one generation to the next is achieved by applying crossover and/or mutation operators. If both these operators are used, crossover is usually applied first, after which the generated offspring are mutated. For these algorithms, mutation step sizes are sampled from some probability distribution function. DE differs from these evolutionary algorithms in that

1) mutation is applied first to generate a trial vector, which is then used within the crossover operator to produce one offspring, and
2) Mutation step sizes are not sampled from a prior known probability distribution function.

In DE, mutation step sizes are influenced by differences between individuals of the current population.

Difference Vectors

The positions of individuals provide valuable information about the fitness landscape. Provided that a good uniform random initialization method is used to construct the initial population, the initial individuals will provide a good representation of the entire search space, with relatively large distances between individuals. Over time, as the search progresses, the distances between individuals become smaller, with all individuals converging to the same solution. Keep in mind that the magnitude of the initial distances between individuals is influenced by the size of the population. The more individuals in a population, the smaller the magnitude of the distances.

Distances between individuals are a very good indication of the diversity of the current population, and of the order of magnitude of the step sizes that should be taken in order for the population to contract to one point. If there are large distances between individuals, it stands to reason that individuals should make large step sizes in order to explore as much of the search space as possible. On the other hand, if the distances between individuals are small, step sizes should be small to exploit local areas. It is this behaviour that is achieved by DE in calculating mutation step sizes as weighted differences between randomly selected individuals. The first step of mutation is therefore to

Volume 4 No. 2, February 2014 ISSN 2223-4985

International Journal of Information and Communication Technology Research

http://www.esjournals.org

first calculate one or more difference vectors, and then to use these difference vectors to determine the magnitude and direction of step sizes.

Using vector differentials to achieve variation has a number of advantages. Firstly, information about the fitness landscape, as represented by the current population, is used to direct the search. Secondly, due to the central limit theorem [19], mutation step sizes approaches a Gaussian (Normal) distribution, provided that the population is sufficiently large to allow for a good number of difference vectors [20].1 The mean of the distribution formed by the difference vectors are always zero, provided that individuals used to calculate difference vectors are selected uniformly from the population [21, 22]. Under the condition that individuals are uniformly selected, this characteristic follows from the fact that difference vectors $(x_{i_1} - x_{i_2})$ and $(x_{i2} - x_{i_1})$ occur with equal frequency, where xi1 and xi2 are two randomly selected individuals. The zero mean of the resulting step sizes ensures that the population will not suffer from genetic drift. It should also be noted that the deviation of this distribution is determined by the magnitude of the difference vectors. Eventually, differentials will become infinitesimal, resulting in very small mutations.

Mutation

The DE mutation operator produces a trial vector for each individual of the current population by mutating a target vector with a weighted differential. This trial vector will then be used by the crossover operator to produce offspring. For each parent $x_i(t)$, generate the trial vector, $u_i(t)$, as follows: Select a target vector, $x_{i1}(t)$, from the population, such that $i \neq i_1$. Then, randomly select two individuals, x_{i2} and x_{i3}, from the population such that $i \neq i_1 \neq i_2 \neq i_3$ and $i_2, i_3 \in u(1, n_s)$. Using these individuals, the trial vector is calculated by perturbing the target vector as follows:

$$u_i(t) = x_{i1}(t) + \beta(x_{i2}(t) - x_{i3}(t)) \qquad (1)$$

Where $\beta \in (0, \infty)$ is the scale factor, controlling the amplication of the differential variation.

Crossover

The DE crossover operator implements a discrete recombination of the trial vector, $u_i(t)$, and the parent vector, $x_i(t)$, to produce offspring, $x_i'(t)$. Crossover is implemented as follows:

$$x_{ij}'(t) = \begin{cases} u_{ij}(t) & if \;\; j \in J \\ x_{ij}(t) & otherwise \end{cases} \qquad (2)$$

Where $x_{ij}(t)$ refers to the $j-th$ element of the vector $x_i(t)$, and J is the set of element indices that will undergo perturbation (or in other words, the set of crossover points). Different methods can be used to determine the set , J .

Selection

Selection is applied to determine which individuals will take part in the mutation operation to produce a trial vector, and to determine which of the parent or the offspring will survive to the next generation. With reference to the mutation operator, a number of selection methods have been used. Random selection is usually used to select the individuals from which difference vectors are calculated. For most DE implementations the target vector is either randomly selected or the best individual selected.

To construct the population for the next generation, deterministic selection is used: the offspring replaces the parent if the fitness of the offspring is better than its parent; otherwise the parent survives to the next generation. This ensures that the average fitness of the population does not deteriorate.

The proposed approach to the use of GAs for Feature selection involves encoding a set of d, Feature s as a binary string of d elements, in which a 0 in the string indicates that the corresponding Feature is to be omitted, and a 1 that it is to be included. This coding scheme represents the presence or absence of a particular Feature from the Feature space (see Fig. 2). The length of chromosome equal to Feature space dimensions.

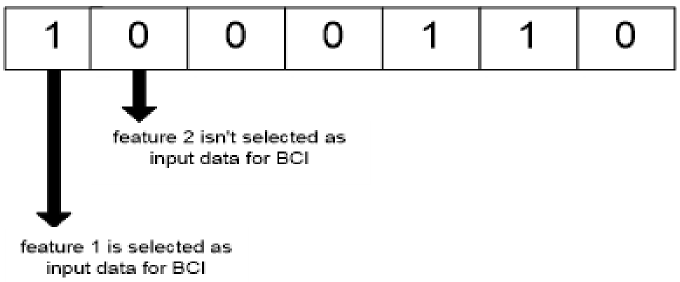

Fig. 2. Schema of the proposed DE-based feature selection approach

International Journal of Information and Communication Technology Research

http://www.esjournals.org

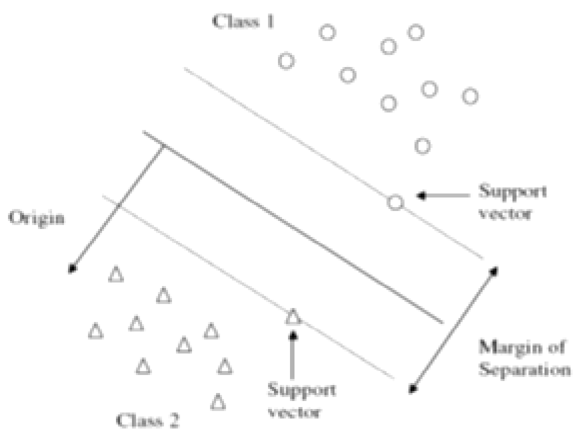

Fig. 3. SVM classification with a hyperplane that maximizes the separating margin between the two classes (indicated by data points marked by "Δ"s and "**O**"s). Support vectors are elements of the training set that lie on the boundary hyperplanes of the two classes.

6. CLASSIFICATION APPROACHES

An artificial neural network (ANN) is an interconnected group of artificial neurons simulating the thinking process of human brain. One can consider an ANN as a "magical" black box trained to achieve expected intelligent process, against the input and output information stream. ANN are useful in application areas such as pattern recognition, classification etc. [23].

6.1 Multilayered Perceptron Neural Networks

The decision making process of the ANN is holistic, based on the features of input patterns, and is suitable for classification of biomedical data. Typically, multilayer feed forward neural networks can be trained as non-linear classifiers using the generalized back-propagation (BP) algorithm.

Our network has one hidden layer with five neurons and output layer with one neuron. Generalized BP algorithm with momentum used as training procedure. Momentum is a standard training technique which is used to speed up convergence and maintain generalization performance [24]. For hidden and output layers, we used bipolar and unipolar sigmoid functions respectively as decision function on the other hand we normalized weights and inputs. With these methods we achieved a NN classifier that is the most suitable classifier for the task at hand. We determined the most effective set as well as the optimum vector length for high accuracy classification. This NN classifier was trained and tested by using the feature sets described above.

By means of minimizing error optimized the number of neurons in hidden layer to five with tansig functions and sigmoid function for output layer.

6.2 Probabilistic Neural Network

The probabilistic approach to neural networks has been developed in the framework of statistical pattern recognition. Probabilistic neural network (PNN) is derived from radial basis function (RBF) network which is an ANN using RBF. RBF is a bell shape function that scales the variable nonlinearly. PNN is adopted for it has many advantages [7]. Its training speed is many times faster than a BP network. PNN can approach a Bayes optimal result under certain easily met conditions. Additionally, it is robust to noise examples. We choose it also for its simple structure and training manner. The most important advantage of PNN is that training is easy and instantaneous. Weights are not "trained" but assigned. Existing weights will never be alternated but only new vectors are inserted into weight matrices when training. So it can be used in real-time. Since the training and running procedure can be implemented by matrix manipulation, the speed of PNN is very fast.

6.3 Support Vector Machine

The SVM is a relatively new classification technique developed by Vapnik [25] which has shown to perform strongly in a number of real-world problems, including BCI.

The invention of SVM was driven by underlying statistical learning theory, i.e., following the principle of structural risk minimization that is rooted in VC dimension theory, which makes its derivation even more profound. The SVMs have been a topic of extensive research with wide applications in machine learning and engineering. The SVM for the linearly separable case find optimal separating hyper plane, as shown in "Fig. 3".

7. SIMULATION RESULTS

In this paper, we proposed a scheme to combine ICA, DE and neural network classifiers for EEG signal classification.

To classify cursor movements two types of the EEG database are used, 70% of each dataset used for training and the rest for test classifiers. Both neural network classifiers and SVM demonstrated high classification accuracies with relatively small number of features. Between the three classifiers, SVM shows slightly better performance than MLP and PNN in terms of classification accuracy and robustness to different number of features. The results prove that the proposed scheme a promising model for the discrimination of clinical EEG signals. The performance of a classifier is not just measured as the accuracy achieved by the network, but aspects such as computational complexity and convergence characteristics are just as important. To reduce complexity, the DE used to select essential EEG channels. This approach to BCI helps to reduce the computational complexity of the Classification process, and helps to improve transfer rate in real-time BCI systems.

Generally, the classification accuracy over files, which were included in training, is higher than the accuracy for the testing set. Tables I and II indicate the results of classification accuracy during training and test stages for both datasets. In comparison with the neural network classifier, SVM has a better training and test accuracy rate of neural network classifier, because of the nature of SVM classifier, this classifier is more general than neural network and this specification is very important in the use of classifiers. The most important advantage of PNN is that training is easy and instantaneous in comparison with SVM and MLP classifiers.

Table I: Results of the Dataset Type I

Results obtained by the Different Feature Selection Methods for Dataset I

Feature selection method / Classifier		PCA	DE
MLP	Train	96.8	99.7
	Test	84.67	85.31
PNN	Train	100	100
	Test	86.35	87.15
SVM	Train	98.83	99.94
	Test	91.29	92.01

Table II: Results of the Dataset type II Results bbtained by the Different Feature Selection Methods for Dataset II

Feature selection method / Classifier		PCA	DE
MLP	Train	98.15	98.25
	Test	86.30	87.26
PNN	Train	100	100
	Test	84.18	83.47
SVM	Train	99.65	99.96
	Test	86.91	88.63

8. CONCLUSION

The goal of this paper is on one hand to reduce the redundancy and on the other hand to increase the BCI speed and making use of it in real time form. Hence, the ICA is applied to Artifact removal

In order to extract the most suitable features from the raw EEG data different methods in time or frequency domain can be used. Since the EEG is non-stationary in general, it is most appropriate to use time–frequency domain methods like wavelet transform (WT) as a mean for feature extraction. The simulation results confirm this fact.

The DE is applied in order to choose the best features from the feature space have been used. DE is an evolutionary algorithm which its optimality has been proved in other fields, the computation complexity is low and it is an appropriate method in real time problems.

One of the main privileges of the mixed method used in this paper is that, the redundant data are removed by the selection power of the DE. This fact reduces the data dimensions and reduced the time response of system significantly. Moreover, the accuracy of classifiers has not only reduced but also in most

Acknowledgment

This study was supported by Islamic Azad University, Borujerd Branch, Iran. The authors would like to acknowledge staff of university.

REFERENCES

[1]. Vidal J (1977) Real-time detection of brain events in EEG. IEEE Proc Special issue on Biological Signal Processing and Analysis 65:633–664

[2]. Chapin JK, Moxon KA, Markowitz RS, Nicolelis MAL (1999) Real-time control of a robot arm using simultaneously recorded neurons in the motor cortex. Nat Neurosci 2:664–670

[3]. Wolpaw JR, Birbaumer N, Heetderks WJ, McFarland DJ et al (2000) Brain-computer interface technology: a review of the first international meeting. IEEE Trans Neural Syst Rehabil Eng 8:164–173

[4]. Vaughan TM, Heetderks WJ, Trejo LJ, Rymer WZ et al (2003) Brain-computer interface technology: a review of the second international meeting. IEEE Trans Neural Syst Rehabil Eng 11:94–109

[5]. McFarland DJ, Wolpaw JR,“ Brain-computer interface operation of robotic and prosthetic devices”. Computer 41(10):52–56, 2008.

[6]. M. R. Nazari Kousarrizi, A. Asadi Ghanbari, A. Gharaviri, M. Teshnehlab, M. Aliyari, “Classification of Alcoholics and Non-Alcoholics via EEG Using SVM and Neural Networks,” IEEE International Conference on Bioinformatics and Biomedical Engineering, 2009.

[7]. D. K. Kim and S. K. Chang, “Advanced Probabilistic Neural Network for the prediction of concrete Strength”, ICCES, vol. 2, pp. 29-34, 2007.

[8]. S. Chandaka, A. Chatterjee, S. Munshi, “Cross-correlation aided support vector machine classifier for classification of EEG signals,” Expert Systems with Applications, 2008.

[9]. Zhou W, Gotman J. Removing eye-movement artifacts from the eeg during the intracarotid amobarbital procedure. Epilepsia 2005;46:409–14.

[10]. De Beer NA, van de Velde M, Cluitmans PJ. Clinical evaluation of a method for automatic detection and removal of artifacts in auditory evoked potential monitoring. J Clin Monit 1995;11:381–91.

[11]. Choi S, Cichocki A, Park HM, Lee SY. Blind source separation and independent component analysis: a review. Neural Inf Process-Lett Rev, 2005;6:1–57.

[12]. Jung TP, Makeig S, Westerfield M, Townsend J, Courchesne E, Sejnowski TJ. Analysis and visualization of single-trial event-related potentials. Hum Brain Mapp 2001;14:166–85.

[13]. Makeig S, Enghoff S, Jung TP, Sejnowski TJ. A natural basis for efficient brain-actuated control. IEEE Trans Rehabil Eng, 2000b;8:208–11.

[14]. Joyce CA, Gorodnitsky IF, Kutas M. Automatic removal of eye movement and blink artifacts from EEG data using blind component separation. Psychophysiology, 2004;41:313–25.

[15]. A. Asadi Ghanbari, M. R. Nazari Kousarrizi, M. Teshnehlab, M. Aliyari, A. Gharaviri, "Wavelet and Hilbert Transform-based Brain Computer Interface," IEEE International Conference on advances tools for engineering application. Notre dame university-Lebanon, 2009.

[16]. Te-Sheng Li, " Feature Selection For Classificatin By Using a GA-Based Neural Network Approach", Journal of the Chinese Institute of Industrial Engineers, Vol. 23, No. 1, pp. 55-64, 2006.

[17]. K.V. Price, R.M. Storn, and J.A. Lampinen. Differential Evolution: A Practical Approach to Global Optimization. Springer, 2005.

[18]. R. Storn and K. Price. Differential Evolution – A Simple and Efficient Heuristic for global Optimization over Continuous Spaces. Journal of Global Optimization, 11(4):431–359, 1997.

[19]. J. Davidson. Stochastic Limit Theory. Oxford Scholarship Online Monographs, 1994.

[20]. R. Storn. On the Usage of Differential Evolution for Function Optimization. In Proceedings of the Biennial Conference of the North American Fuzzy Information Processing Society, pages 519–523, 1996.

[21]. K.V. Price. Differential Evolution vs. The Functions of The 2nd ICEO. In Proceedings of the IEEE International Conference on Evolutionary Computation, pages 153–157, 1997.

[22]. I.L. L´opez Cruz, L.G. van Willigenburg, and G. van Straten. Efficient Differential Evolution algorithms for multimodal optimal control problems. Applied Soft Computing, 3(2):97–122, 2003.

[23]. Sun S, "Research on EEG signal classification for braincomputer interfaces based on machine learning methodologies". Ph.D. dissertation, Dept Automation, Tsinghua Univ, Beijing, 2006.

[24]. Duda RO, Hart PE, Stork DG, "Pattern classification". 2nd edn. Wiley, New York, 2000.

[25]. S. Avidan, "Support Vector Tracking," IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 26, no. 8, pp.1064-1072, Aug. 2004