Classifier Combination Approach in Motion Imagery Signals Processing for Brain Computer Interface

Homayoon Zarshenas, Mahdi Bamdad, Hadi Grailu, and Akbar A. Shakoori

***Abstract*—**In this study we focus on improvement performance of a cue based Motor Imagery Brain Computer Interface (BCI). For this purpose, data fusion approach is used on results of different classifiers to make the best decision. At first step Distinction Sensitive Learning Vector Quantization method is used as a feature selection method to determine most informative frequencies in recorded signals and its performance is evaluated by frequency search method. Then informative features are extracted by packet wavelet transform. In next step 5 different types of classification methods are applied. The methodologies are tested on BCI Competition II dataset III, the best obtained accuracy is 85% and the best kappa value is 0.8. At final step ordered weighted averaging (OWA) method is used to provide a proper aggregation classifiers outputs. Using OWA enhanced system accuracy to 95% and kappa value to 0.9. Applying OWA just uses 50 milliseconds for performing calculation**.**

***Keywords*—**BCI, EEG, Classifier, Fuzzy operator, OWA.

I. Introduction

MANY disorders like Spinal cord injury or stroke and Amyotrophic Lateral Sclerosis (ALS), can affect or even completely damage the usual communication channels that a person needs to communicate and interact with his or her environment. These kinds of disorders result in partial or complete loss of voluntary muscle activities including speech. In such disabilities, for less severe levels of affection, a brain–computer interface (BCI) can improve the quality of life of partially paralyzed patients or persons with severe motor disabilities during rehabilitation to have effective control over devices such as computers, speech synthesizers, assistive appliances and neural prostheses. Such an interface would increase an individual's independence, leading to an improved quality of life and reduced social costs [1].

H. Zarshenas is with the Mechanical Engineering Department, Shahrood university of technology, Shahrood, Iran (e-mail: homi_zarshenas@yahoo.com).

M. Bamdad is with the Mechanical Engineering Department, Shahrood university of technology, Shahrood, Iran (corresponding author to provide phone: ++98 273 3300240; fax: ++98 273 3300258; e-mail: Bamdad@shahroodut.ac.ir).

H. Grailu is with the Electrical and electronic Engineering Department, Shahrood university of technology, Shahrood, Iran (e-mail: grailu@shahroodut.ac.ir).

A. A. Shakoori is with the Mechanical Engineering Department, Shahrood university of technology, Shahrood, Iran (e-mail: shakoori_amir@yahoo.com).

Research on BCIs began in the 1970s at the University of California Los Angeles (UCLA) [2], [3]. BCI Researches started by experiments on animals in order to create a new direct communication path between brain and environment. In 1969 and 1970 cursor motion controlling was tested on monkeys. Following years of animal experimentation, the first neuroprosthetic devices implanted in humans appeared in the mid-1990s.There has been rapid development in BCIs since the 1990s. In 2002 a complete definition of BCI was introduced by Wolpaw and up to now competitions and congress are held in the field of BCI [4].

BCI is connected directly to the brain and receives electrical pulses from brain without the involvement of nervous system in order to recognize human decisions [5], [6]. Brain signals can be detected and measured in many ways; these include invasive methods like receive electrical activity of brain from the cortical surface (electrocorticographic [ECoG] activity) and noninvasive ones like the use of methods for recording electrical or magnetic fields, such as functional MRI, PET, electro electroencephelography (EEG), and Near Infrared Spectroscopy (NIRS) [7].

As EEG doesn't need surgery and has a simple manner and portable equipment for signal recording it becomes most widespread recording modality [8]. However, as the signals have to cross the scalp, skull, and many other layers the quality of the signals is very poor. In addition, EEG recordings are susceptible to contamination from electro oculographic or electro- myographic activity of limbs [9].

Variation in brain signals have different sources and appear by changes in amplitude and frequency of brain waves. These variations, according to the cause of generation, can be categorized as P300, VEP, SCP and sensorimotor signals include variation in mu and beta band of brain waves and ERP and MRP [10].

As Fig. 1 shows, a noninvasive BCI system captures brain signals through the EEG, then the preprocessing phase is done and artifacts that known as undesirable potentials with non-cerebral origins that contaminate the EEG signals are removed from raw signals [11]. After preprocessing phase, a feature vector generated for each input signal according to their time and frequency parameters, then classification is done in order to distinct input signals according to their feature vectors.

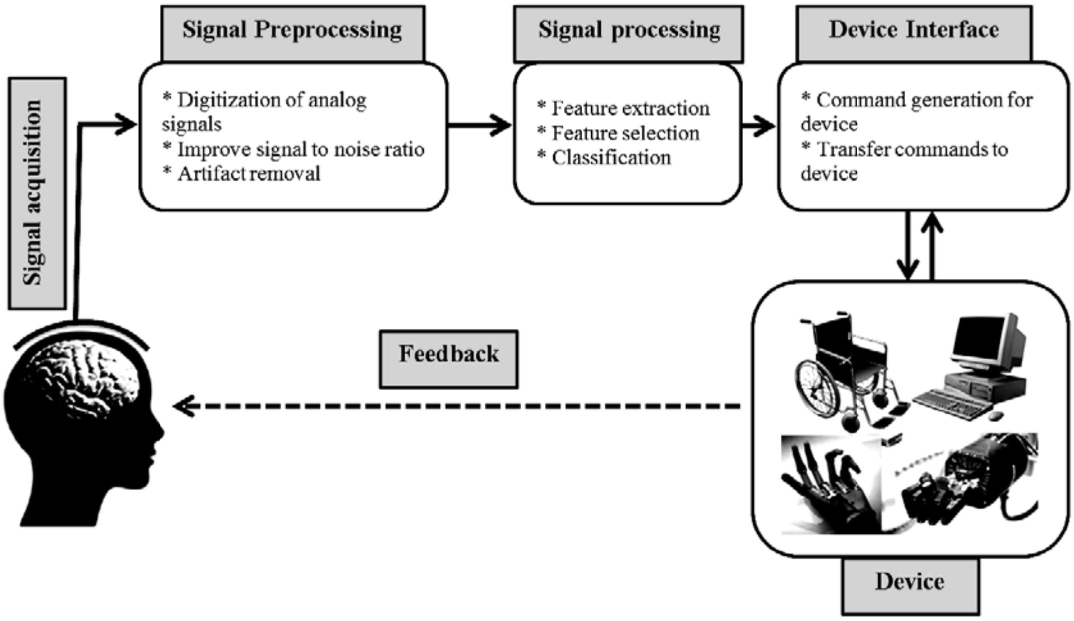

Fig. 1 Brain computer interface system structure

In next step, the related command was transmitted to the device by a central processing unit [12], [13]. In order to have a real time communication between brain and device, signal processing should be done in minimum time [14].

In this article a time- frequency method which is wavelet packet transform is used for feature extraction. In order to classify input signals 5 different classification method is used and at final part ordered weighted averaging method is applied as a classifier combination method for improving the performance of system by providing the best aggregation among classifiers outputs. In next sections signal processing methods that used in this article are explained.

II. Signal Processing Procedure Review Stage

Signal processing procedure that is used in this study is include these parts:

- Determine informative frequencies by DSLVQ and frequency search methods.
- Extract features of informative frequencies by packet wavelet method.
- Classify signals according to extracted features by 5 different classification methods include SVM, k-NN, Bayesian, Parzen, MLP.
- Apply classifier combination method (OWA) to result of classifiers.
- Compare the results and discus about impact of classifier combination on performance of system.

In next section every step and methods that are used in each step is explained.

A. Frequency Selection

The variation in brain signals occur in specific frequencies according to type of events. So it is important to determine informative frequencies for each variation in brain waves. In this article two frequency selection methods are used that are DSLVQ [15] and frequency search [16]. This work helps us to select the features of most informative frequencies and reduce the dimension of feature vector that cause to improve the performance of system.

1. Distinction Sensitive Learning Vector Quantization (DSLVQ)

DSLVQ finds the most discriminative features for the given classes [17]. It is applied to the frequency domain features to obtain effective frequencies of brain waves. It is modified from Learning Vector Quantization (LVQ). DSLVQ employs supervised learning while weighting the features.Weight vector that applied to features are updated through (1).

$$w(t+1) = norm(w(t) + \beta(t)(norm(nw(t) - w(t)))) \quad (1)$$

Here$w(t+1)$ is the consequent iteration weight vector,$w(t)$ is the present weight vector and $nw(t)$ is the new weight vector during iteration t that calculated by (2) and $\beta(t)$is the learning rate.

$$nw(t) = \frac{d_{o_n}(t) - d_{c_n}(t)}{\max(d_{o_n}, d_{c_n})} \quad (2)$$

That $d_{c_n}(t)$ denotes the distance between the nth feature of training example$(x(t))$, and closest reference vector from the correct class and $d_{o_n}(t)$ denotes the distance between the nth feature of$x(t)$ and closest reference vector among the other classes.

After all iterations features that have higher weight are more important and selected to create feature vector while low weighted features can be discarded since they are less important [18].

2. Frequency Search Method

This method is used to evaluate the results of DSLVQ method. At first STFT is applied to determine the fundamental frequency of signal. Then the signal energy according to each frequency is considered as a feature and used for classification input signals by a SVM classifier. At the end each frequency that causes better result in classification is more informative so features that related to these frequencies are used to create feature vector.

B. Feature Extraction

There are different types of feature extraction methods that based on frequency or time features of signals one of them is furrier transform. A weak point of the furrier transform is that it is not possible to demonstrate effect of high frequency changes in brain wave. In order to solve this problem short time furrier transform is introduced that divided signal to time periods by applying a window with specific length to signal in time domain. But using STFT causes to lose some information about very high or low frequencies.

As brain waves are time varying signals and contain frequency and time information so time- frequency methods like wavelet are more appropriate. The main advantage of wavelet transform method is that all information about fundamental frequencies of signal is extracted because in this method length of applied window to signal adjusted according to desired frequency. As a result packet wavelet is used for feature extraction in this article.

1. Packet Wavelet Transform (PWT)

PWT is an extension of discrete wavelet transform (DWT) [19]. The generic step in DWT splits the approximation coefficients into two parts. After splitting we obtain a vector of approximation coefficients and a vector of detail coefficients. The information lost between two successive approximations is captured in the detail coefficients. Then the next step consists of splitting the new approximation coefficient vector; extracted details are never reanalyzed. In the corresponding wavelet packet situation, each detail coefficient vector is also decomposed into two parts using the same approach as in approximation vector splitting. This offers the richest analysis.

C. Classification

Supervised classifiers are widely used in BCI applications. These classifiers are divided to parametric and nonparametric types. In this article 5 different classification methods are used that a brief explanation about each one presented here.

2. Support Vector Machine (SVM)

Linear classifiers define a linear function for separating 2 classes of data. These methods because of accurate performance beside simple calculation are popular in BCI applications. One of these methods is support vector machine (SVM).

SVM is a supervised machine learning algorithm which is known as a powerful tool for pattern classification. In this algorithm, the goal is to separate the classes with a hyper plane which is constructed by the observed examples [20]. The hyper plane should have a good generalization property in order to work well for the unseen samples. In this manner, SVM tries to find the separating hyper plane with the maximum margin which is known as the optimal separator plane. In other words, SVM maximizes the distance between the hyper plane and the nearest samples (support vectors) of each class [21].

3. K Nearest Neighbor Method (K-NN)

Nearest neighbor is a supervised learning algorithm where the classification of new coming instance is based on nearest neighbor class. For BCI, these nearest neighbors are usually obtained using a metric distance. With a sufficiently high value of k and enough training samples, k-NN can approximate any function which enables it to produce nonlinear decision boundaries. In BCI application distance between input data vector and each reference vector of each class is calculated by the Euclidean distance function [22]. Number on neighbor is defined according to structure of classes, in this article the optimum value for k is 1.

4. Bayesian Method

Bayesian classifier is a probabilistic classifier based on the Bayes' theorem. This method is based on an inductive algorithm and aim to categorize each input feature vector in the class that has the most similarity to it [23]. So through a training process number of index like mean value or covariance matrix are determined as features vectors of each class then according to similarity of these index, class of new features vectors clarified [24].

5. Parzen Method

The Parzen classifier provides an estimate of the class-conditional probability density function (PDF) by applying a kernel density estimator to the labeled feature vectors in the training set. The Parzen classifier estimates the class densities and has a built-in optimization for the smoothing parameter [25]. In Parzen method the radius of neighborhood is important parameter so it needs to determine the best distance of neighborhood in a training algorithm according to result of classification.

6. Multilayer Perceptron (MLP)

MLP neural network is a powerful classifier for discriminating mental tasks. For training the network, back propagation approach with steepest decent optimization algorithm has been used [26], [27]. In this study we used a MLP with one hidden layer and 8 neurons. Its Output layer had 2 neurons and it was equal to the number of classes or mental tasks.

D. Classifier Combination

The combination of classifiers helps reducing the variance component of the classification error which generally makes combinations of classifiers more efficient than their single counterparts. There are different approaches for classifier combination but fuzzy based aggregation methods are more popular [28]. Order weighted averaging is one of the fuzzy integration methods with few calculation time and proper performance that are important factors in BCI applications. We have applied OWA operators for aggregating multidecisions to form an overall decision function considered as the fuzzy majority based voting strategy.

1. OWA Method

The notion of OWA operators was first introduced by Yager in regarding the problem of aggregating multi-criteria to form an overall decision function [29]. A mapping$F: [0,1]^n \rightarrow [0,1]$ is called an OWA operator of dimension n if it is associated with a weighting vector $W = [\omega_1, \ldots, \omega_n]$ while $\omega_i \in [0,1]$ and $\sum \omega_i = 1$ and $F(a_1, \ldots, a_n) = \sum_{i=1}^{n} \omega_i b_i$ where b_i is the $i - th$ largest element in the collection $a_1, \ldots, a_n$.

OWA coefficients (ω_i) determined by different methods, one of them that is used in this article is optimistic approach [30]. By this approach we have:

$$w_1 = a; w_2 = a(1-a); w_3 = a(1-a)^2; \ldots; w_{n-1} = a(1-a)^{n-2}; w_n = (1-a)^{n-1} \quad (3)$$

That $a \in [0,1]$.

In order to obtain the best coefficients through an iterative learning algorithm optimum value of $\propto$ is determined by minimizing the error value (e).

$$e = \frac{1}{2}\left(\sum_{j=1}^{n} w_j x_{\sigma(j)} - X_d\right)^2 \quad (4)$$

That X_d is the desired value of classification.

III. Results

In order to evaluate methods that presented in previous sections dataset III of BCI competition II is selected. This dataset include brain waves obtained from channels C3 and C4 for two class of sensorimotor activity that related to left hand and right hand cue based motion imagination. More description about dataset and data acquisition condition is presented in [31].

First of all DSLVQ is used to determine more informative frequencies of signals. Result of DSLVQ is presented in Fig. 2.

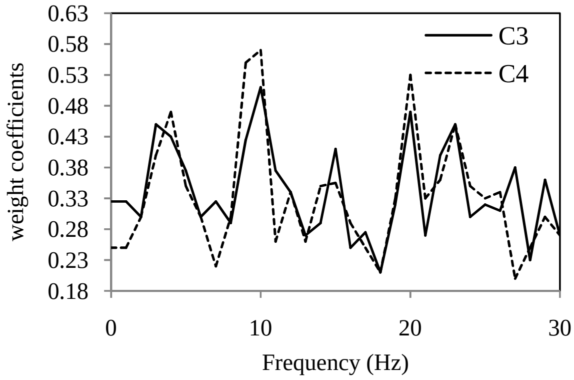

Fig. 2 DLSVQ coefficients for the frequencies 1-30 Hz for C3 and C4

It can be seen that weights of 11 and 21Hz for channel C3 and 10 and 20Hz for channel C4 have largest values so features of these frequencies are more informative and used in feature vector. For evaluation results of DSLVQ, frequency search method is used and its results demonstrated in Fig. 3.

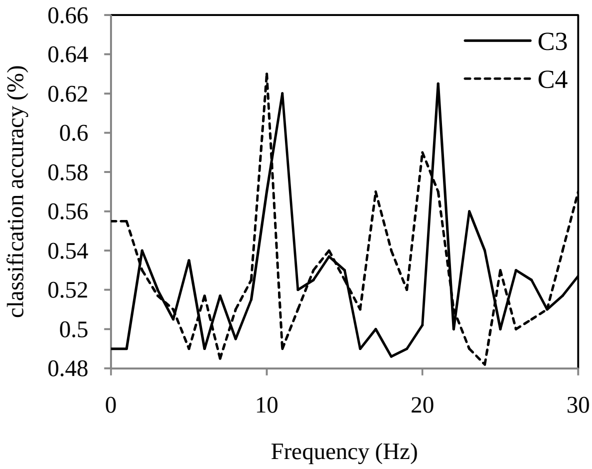

Fig. 3 Frequency search classification results for the frequencies 1-30 Hz for C3 and C4

According to the frequency search method results, 11 and 21Hz for C3 and 10 and 20Hz for C4 get best result in classification. It means that these frequencies have more disparity between two classes. These results approve results of DSLVQ method. Finally features of 10 to 11Hz and 20 to 21Hz are used for feature vector creation.

In next step feature extraction is done in specified frequencies. As explained before packet wavelet transform from model db2 is used in four step and then wavelet coefficient of desired frequencies are selected as feature vector. After feature extraction classification methods are applied to feature vectors and result demonstrated in Table I. The performance of different methods is compared based on different criteria. These criteria include classifier accuracy kappa value which is a presentation of persuasion of classifier and is cited in BCI studies. Another important factor is time consumption for each signal classification that is important in real time BCI applications. In all classification methods 70% of data is used for training and 30% for testing.

TABLE I
PERFORMANCE OF CLASSIFICATION METHODS

Classification method	Accuracy (%)	Kappa value	Training time (s)	Classification time (s)
MLP	82.5	0.65	0.27	0.072
SVM	80	0.6	0.72	0.012
k-NN	90	0.8	0.5	0.015
Parzen	85	0.7	0.68	0.02
Bayesian	67	0.34	0.71	0.03

According to classifiers results, by applying PWT for feature extraction from determined frequencies and then classify feature vectors by k-NN, better accuracy is achieved in comparison to winner of BCI competition II for this dataset.

In BCI applications it is important to obtain maximum accuracy in signal classification in minimum time delay. So we introduce new classifier combination method which is OWA. Coefficients of OWA are determined based on optimistic approach. So in order to identify optimum value for a in this method an iterative classification process is done by applying different value for a from 0.01 to 1. As it can be seen in Fig. 4 the best performance is achieved for $a = 0.79$ so OWA coefficients are determined based on this amount of a.

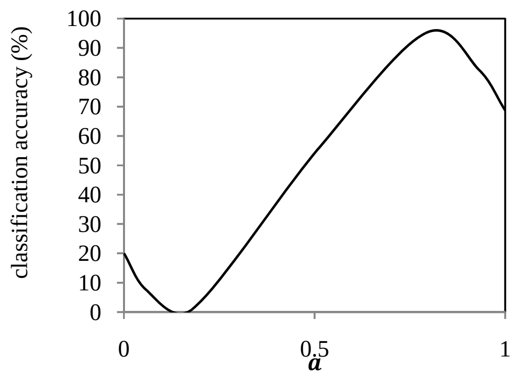

Fig. 4 OWA performance per different value of a

By applying classifier combination (OWA), accuracy of system improves to 95% and Kappa value rise to 0.9 that shows improvement in results. Combining the results of 5 classifier by OWA method just increase the time of calculation about 50 millisecond that is an advantage for online application of BCI such as smart home control.

IV. CONCLUSION

In this article packet wavelet method is used for feature extraction that is a time- frequency signal processing method. In classification step by applying 5 different types of classifiers their performance are compared with same feature vector in equal situation. In final part OWA is used for combining classifiers that lead to 5% improvement in system accuracy.

It must be mentioned that informative frequencies are not same for different individuals so it is a weak point of BCI systems that make it exclusive for each user. As a result in order to obtain a proper BCI system it is important to determine informative features for each user and retrain system by these new features.

REFERENCES

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

[2] J. J. Vidal, "Toward Direct Brain-Computer Communication", Annual Review of Biophysics and Bioengineering vol. 2, pp. 157-180, 1973

[3] J. J. Vidal, "Real-time detection of brain events in EEG" Proceedings of the IEEEvol. 65, no. 5, pp. 633–641, 1977.

[4] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G.Pfurtscheller,and T. M. Vaughan, "Brain–computer interfaces for communication and control", Clinical Neurophysiology, vol. 113, pp. 767–791, 2002.

[5] C. J. Bell, P. PradeepShenoy, R. RawichoteChalodhorn, and P. N. Rao, "Control of a humanoid robot by a noninvasive brain–computer interface in humans", Journal of Neural Engineeringvol. 5, no. 2, 2008.

[6] M. Zhong, F.Lotte, M. Girolami, and A. Lécuyer, "Classifying EEG for brain computer interfaces using Gaussian processes", Pattern Recognition Lettersvol. 29, no. 3, pp. 354–359, 2008.

[7] M. A. Lebedev, and M. A. Nicolelis, "Brain–machine interfaces: past, present and future", TRENDS in Neurosciences, vol. 29, no. 9, pp. 536-546, 2006.

[8] F. Mason, M. I. Norton, J. D. Van Horn, D. M. Wegner, S. T. Grafton, and C. Macrae, "Wandering Minds: The Default Network and Stimulus-Independent Thought", Science Magazine, vol. 315, no. 5810, pp. 393-395, 2007.

[9] B. H. Dobkin, "Brain–computer interface technology as a tool to augment plasticity and outcomes for neurological rehabilitation", The Journal of Physiology vol. 579, Issue 3, pp. 637–642, 2007

[10] T. Al-ani,and D.Trad, *Intelligent and Biosensors*, Publisher InTech, 2010, pp. 25-66.

[11] G. Dornhege, R. Millán, T. Hinterberger, D. McFarland,and K. R. Müller, *Towards Brain-Computer Interfacing*, MIT Press, Cambridge publishing, 2007, pp. 31-42.

[12] E. Pasqualotto, S. Federici,and M. O.Belardinelli, "Toward functioning and usable brain–computer interfaces (BCIs): A literature review", Disability and Rehabilitation: Assistive Technology, vol. 7, no. 2, pp. 89–103, 2012.

[13] N. Birbaumer, "Breaking the silence: Brain–computer interfaces (BCI) for communication and motor control", Psychophysiology, vol. 43, pp. 517–532, 2006.

[14] M. M. Moore, "Real-World Applications for Brain Computer Interface Technology", neural system and rehabilitation engineering, vol. 11, no. 2, pp. 162-165, 2003.

[15] D. FlotzFlotzinger, M. Pregenzer, and G. Pfurtscheller, "Feature selection with distinction sensitive learning vector quantisation and genetic algorithms Process", IEEE International Conference on Neural Networks, Orlando, FL, 1994, pp. 3448–3451.

[16] B. Akinci, *Realization of a cue based motor imagery brain computer interface with its potential application to a wheelchair*, Master of Science thesis, school of natural and applied sciences of middle-east technical university, Turkey, 2010.

[17] M. Pregenzer, *Distinction sensitive learning vector quantization*, Doctoral dissertation, University of Technology, Graz, Austria, 1997, pp. 43-60.

[18] T. Kohonen, "The Self-Orginizing Map", Proceedings of the IEEE, vol. 78, no. 9, 1990

[19] W. Tinga, Y. G.Zheng, Y. B. Hua, and S. Hong, "EEG feature extraction based on wavelet packet decomposition for brain computer interface", Measurement, vol. 41, pp. 618–625, 2008

[20] C. J. C. Burges, "A tutorial on support vector machines for pattern recognition", Knowledge Discovery and Data Mining, vol. 2, pp. 121-167, 1998.

[21] K. P. Bennett,and C. Campbell, "Support vector machines: hype or hallelujah?", ACM SIGKDD Explorations Newsletter, vol. 2, no. 2, pp. 1-13, 2000.

[22] B. Blankertz, G. Curio,and K. R. Muller, “Classifying single trial EEG: Towards brain computer interfacing”, Advances in Neural Information Processing Systems, vol. 14, pp.157-164, 2002

[23] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*, second edition. Wiley pree, Wiley-Interscience publisher, 2001, pp. 37-53.

[24] K. Fukunaga, *introduction to Statistical Pattern Recognition*, seconde edition. Academic press, 1990, pp. 25-42.

[25] C. Anderson’s website: http:// www. cs. colostate. edu /~anderson /res /eeg/#Contents

[26] A. Hiraiwa, K.Shimohara,and Y. Tokunaga, “EEG topography recognition by neural networks”, IEEE Engineering in Medicine and Biology Magazine, vol. 9, no. 3, pp. 39-42, 1990.

[27] C. W. Anderson,and Z.Sijercic, “Classification of EEG signals from four subjects during five mental tasks” Proceedings of the International Conference on Engineering Applications of Neural Networks, Turkey, 1996, pp. 407-414.

[28] S. Tulyakov, S.Jaeger, V.Govindaraju,and D.Doermann, “Review of Classifier Combination Methods”, Machine Learning in Document Analysis and RecognitionStudies in Computational Intelligence vol. 90, pp. 361-386, 2008.

[29] R. R. Yager,“On ordered weighted averaging aggregation operators in multi criteria decision making”, Man and Cybernetics, vol. 18, pp. 183-190, 1988.

[30] R. R. Yager,and D. P.Filev,“Induced ordered weighted averaging operators. Systems, Man, and Cybernetics, Part B: Cybernetics”, IEEE Transactions on. vol. 29, no. 2, pp. 141-50. 1999.

[31] BCI Competition. http:// ida. first. fraunhofer. de/ projects/ bci/ competition ii/, 2003.