EEG-based Motor Imagery Classification using Wavelet Coefficients and Ensemble Classifiers

Reza Ebrahimpour, Kioumars Babakhani, Morteza Mohammad-Noori

***Abstract*—brain-computer interface (BCI) is a system that captures and decodes electroencephalogram (EEG) signals and transforms human thoughts into actions. To achieve this goal, using classification algorithms are most popular approach. However, classification of EEG signals can be categorized in complex problems because of high nonlinearity, high dimensionality, poor signal to noise ratio and poor spatial resolution. Combining classifiers is an approach to improve the performance of complex problems. In this article we studied the application of combining classifiers based on wavelet features to improve the performance of EEG signal classification in BCI systems. Three normal subjects K3b, K6b and L1b were asked to perform imaginary movements of left hand, right hand, tongue and foot during predefined time interval. EEG signals were decomposed into wavelet coefficients by discrete wavelet transform and used as feature vectors, presenting them into classifiers. Four combining classifiers were used to evaluate the EEG signals. Experimental results show that wavelet transform is an appropriate tool for the analyzing EEG signals. Also, according to the results of the experiments, mixture of experts overcomes the other used combining methods.**

Keywords: ***Brain-computer interface (BCI), EEG signals, Wavelet transform, combining methods, Mixture of experts***

I. INTRODUCTION

A brain-computer interface (BCI) is a system which allows users to act on their environment by using only brain activity, without using peripheral nerves and muscles. The idea is to provide a new communication channel to people who suffer from amyotrophic lateral sclerosis, brain-stem stroke, brain or spinal cord injury, cerebral palsy, muscular dystrophies, multiple sclerosis, and other diseases. [1]. This new communication channel is obtained in the following way: the brain signals are acquired and then processed to extract some features of interest from them; these features are then classified and encoded into semantic symbols that get finally mapped into the output commands.

Reza Ebrahimpour, *Brain and Intelligent Systems Research Lab, Department of Electrical and Computer Engineering, Shahid Rajaee Teacher Training University, P.O.Box:16785-163, Tehran, Iran, ebrahimpour@ipm.ir,*

Kioumars Babakhani, *School of Mathematics, Statistics and Computer Science, University of Tehran, Iran, kbabakhani@ut.ac.ir*

Morteza Mohammad-Noori, *School of Mathematics, Statistics and Computer Science, University of Tehran, Iran, morteza@ipm.ir*

Since a BCI system is controlled directly by the brain, it needs a method to detect brain activity. Scalp-recorded electroencephalography (EEG) is a noninvasive convenient way eavesdropping brain signals from cortical areas. At present, the EEG is widely used to monitor brain activity in BCI research [2]. The research is based on recording and analyzing EEG brain activity and recognizing EEG patterns associated with mental states.

In the literature of motor imagery classification, several feature extraction methods were used to extract the most informative set of features, representing them as input patterns into classifiers. An autoregressive (AR) model was applied to short overlapping EEG segments. From the AR spectrum, band power is calculated in several frequency bands and the power sum is used as independent variables in a linear function that defines the control signals [3]. In a different approach, the most reactive frequency bands for each subject were selected by using the distinction sensitive learning vector quantization (DSLVQ). For the selected bands, the band power is calculated and classified by a nonlinear classifier (LVQ) [4]. Common spatial patterns (CSP) combined with simple linear or quadratic Bayes classifiers were another method [5]. Also, CSP and adaptive autoregressive parameters (AAR) were used as feature extraction methods in [6, 7]. Recently, algorithms using weighted time–frequency analysis [8] were developed.

Wavelet Transform (WT) was proposed to address the problem of poor temporal resolution in nonstationary EEG signals. Wavelet coefficients were used as feature vectors identifying characteristics of the signal that were not apparent from the original time domain signal [9].

Supervised classification methods are employed to recognize the patterns of EEG activities. Several methods have been proposed to BCI. As the classification methods, Linear discriminant analysis (LDA) and fisher's linear discriminant analysis (FLDA) [2, 10], support vector machine (SVM) [7, 11] lazy learning classifiers such as k-nearest neighbor [11], artificial neural networks (ANNs) were investigated to design EEG-based BCI systems [12, 13].

Due to having non-stationary signals, poor signal to noise ratio, highly overlapped classes, small sample size and high dimensional feature sets, EEG classification can be categorized into complex problems [14]. Combining classifiers is an approach to improve the performance in classification particularly for complex problems [15].

In the present study, we have developed a wavelet-based ensemble scheme to identify motor imagery tasks using EEG signals. EEG signals which are recorded from three subjects when imagine the left hand (L), right hand (R), foot (F) and tongue (T) movements were classified using ensemble methods.

The outline of the paper is as follows. Section 2 gives a brief description of EEG data set and classification methods used in the study. Section 3 presents the results of the experiments and discuses about them. Finally, section 4 concludes the study.

II. Materials and Methods

A. EEG data acquisition

We used the dataset IIIa from the BCI competition 2005 [16]. It contains data from three subjects: K3b, K6b and L1b and was collected as follows: Each subject, sitting in front of a computer, was asked to perform imaginary movements of the left hand, right hand, tongue or foot during a predetermined time interval. Sixty electrodes were placed according to the international 10⁄20 system on the scalp of the subject recording a signal sampled at 250 Hz and filtered between 1 and 50 Hz using a Notch filter. Figure 1(a) shows the position of EEG electrodes.

Each trial starts with a blank screen. Figure 1(b) depicts the timing of a trial: At time point $t = 2s$ a short acoustic stimulus and a cross "+" on the screen were given to advise the subject to pay attention. At $t = 3s$ the cross was overlapped with an arrow pointing either to the left, right, up or down for 1.25 seconds. According to the direction of the displayed arrow the subject was asked to imagine a left hand, right hand, and tongue or foot movement, respectively. The movement imagination had to be performed until the cross disappeared at $t = 7s$. After this, a short break with a randomly selected length of up to 2 seconds was given before the next trial.

Each of the four types of cues was displayed ten times within each run in a randomized order. No feedback was provided to the subject. The data set recorded from subject K3b consisted of 9 runs, whereas the data set of K6b and L1b consisted of 6 runs each.

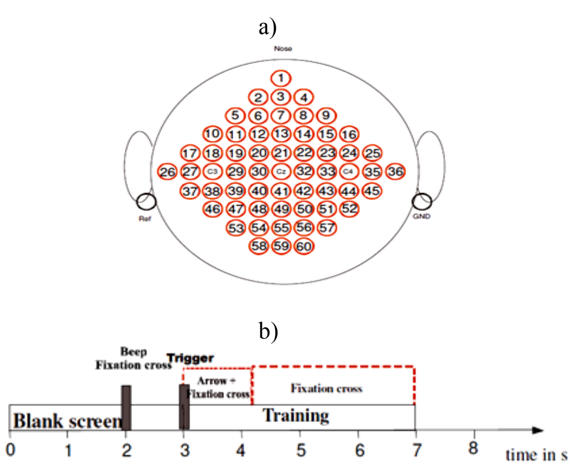

Figure 1 Position of EEG electrodes (a) and timing of the training paradigm (b)

B. Neural network ensemble methods

Combining classifiers have two major components; first, a method to create individual classifiers, second, a method to combine their outputs such that the combination improves upon the performance of a single classifier. This requires, however, that individual classifiers make errors on different instances. Such a set of classifiers is said to be *diverse* [15].

The second key component of any ensemble system is the strategy employed in combining classifiers. According to the existence of the training process in the combiner parts of ensemble, combining methods is categorized into two groups: trainable vs. non-trainable combiners [15]. In trainable strategy, combiners need training, after the base classifiers in the ensemble have been trained individually. In this category, the parameters of the combiner, usually called *weights*, are determined through a separate training algorithm.

In non-trainable group, the combiner has no extra parameters that need to be trained; that is, the ensemble is ready for operation as soon as the base classifiers are trained. The simple averaging (SA) and decision templates (DT) are two examples of this group. Also, the trainable combining methods, according the influence of input data on the combining process are classified into data-dependent and data-independent ensembles [15]. In the first group, the input has explicit influence in ensemble as the weights of combining are the functions of input. In the second ones, the input does not have influence on the combining process. Mixture of experts (ME) is the most famous method in the first group and stacked generalization (SG) is the member of second one. For further reading about these methods, please refer to [15]. In this application, we compared the performance of the mentioned ensemble methods on the classification of motor imagery EEG signals.

III. Experimental results

A. Feature extraction using Discrete Wavelet Transform

In the present study, EEG signals were recorded during motor imagery task from three subjects K3b, K6b and L1b. In each trial, subjects were asked to imagine one of predefined tasks (left hand, right hand, tongue and foot movement) in the training phase which is from t= 3s to t= 7s (see figure 1b).

In order to train and evaluate the classifiers, we divided the trials into small time segments with length 0.5 second of three channels C3, Cz and C4. Since these channels are placed in the motor cortex area, these channels can be used to recognize the changes of ongoing EEG pattern during motor imagery tasks better than other ones. Also, since that the sampling rate of the recording was set 250 Hz, each selected EEG segment is a 125–dimensional vector.

To obtain the informative set of features, all EEG segments were decomposed into wavelet coefficients using discrete wavelet transform. The number of decomposition levels was chosen to be 4. Also, by trial and error, the smoothing feature of the Daubechies wavelet of order 2 (db2) was used as wavelet function. For each EEG segment, the detail wavelet coefficients (D_k, k = 1, 2, 3, 4) at the first, second, third and fourth levels (64 + 33 + 18 + 10 coefficients) and the approximation wavelet coefficients (A_4) at the fourth level (10 coefficients) were computed. Then 135 wavelet coefficients were obtained for each EEG segment. In order to reduce the dimensionality of the feature vectors, the five statistical features *Maximum*, *Minimum, Median, Mean and Standard deviation* of the wavelet coefficients in each subband were computed.

Thus, each feature vector that is presented to classifier has 25 (five statistical features for each subband) dimensions. The number of patterns is 1080 (three channels × number of trials) patterns for subject K3b and 720 patterns for both subjects K6b and L1b. The classification accuracy was measured using a k-fold cross validation technique, with k=4. The data set for each subject was partitioned into four equally sized subsets. Each of the subsets was taken in turn as the test set for a total of four trials.

B. Experimental setup

The parameters of ensemble methods are as follows: Each of individual experts is a multilayer perceptron (MLP) with one hidden layer. All ensemble methods were evaluated with ensemble size of three, five and seven expert networks. In all ensemble methods, the number of hidden neurons of each expert was set 15. Also, the number of hidden neurons of gating network of ME was set 10. In SA, DT and SG methods, individual experts were constructed using bootstrapping technique. All methods were trained using back–propagation (BP) training algorithm with the learning rate value of $\eta_e=0.1$. Also, the learning rate value of gating network of ME was set $\eta_g=0.05$. The average classification accuracy of SA, DT, SG and ME methods for different size of ensemble for three subjects K3b, K6b and L1b are presented in tables 1-3, respectively. The classification accuracies are shown under two situations: with and without feature extraction. In the first one, the EEG segments were decomposed by discrete wavelet transform, and then statistical features of each subband were used as feature vectors. In the second condition the raw EEG segments without any feature extraction were used as input patterns.

Table 1: Average classification accuracies of ensemble classifiers for subject K3b.

Feature type	Number of experts	Ensemble classifiers			
		SA	DT	SG	ME
Wavelet coefficients	3	75.78	83.45	84.65	85.39
	5	78.36	84.67	86.15	90.83
	7	77.14	85.50	82.44	88.15
Raw EEGs	3	60.54	66.41	64.63	69.55
	5	61.63	69.14	68.00	73.75
	7	62.25	69.45	65.89	72.67

Table 2: Average classification accuracies of ensemble classifiers for subject K6b.

Feature type	Number of experts	Ensemble classifiers			
		SA	DT	SG	ME
Wavelet coefficients	3	57.33	61.50	61.09	63.45
	5	61.90	65.85	64.34	65.96
	7	60.84	66.15	62.87	64.33
Raw EEGs	3	49.20	50.78	52.10	55.25
	5	52.62	54.11	55.79	61.47
	7	52.53	53.13	50.00	56.79

Table 3: Average classification accuracies of ensemble classifiers for subject L1b.

Feature type	Number of experts	Ensemble classifiers			
		SA	DT	SG	ME
Wavelet coefficients	3	65.23	71.76	65.80	76.94
	5	68.39	73.81	71.26	75.85
	7	67.50	73.40	69.10	75.54
Raw EEGs	3	53.55	58.34	59.80	63.15
	5	58.62	62.45	60.67	66.93
	7	58.11	62.89	57.05	64.95

As it can be seen from tables 1-3, in all cases, the performance of ensemble classifiers which used wavelet coefficients is significantly better than the classifiers which had no feature extraction process. Also, as expected, the classification accuracy of the ME method is significantly better than DT and SG and is so much better than SA.

The effect of different ensemble size on the performance of the ensemble methods was investigated in this experiment, too. As observed in tables 2-4, in all cases, the performance of mixture of experts (ME) and stacked generalization (SG) methods with five experts, was better than the other cases. The highest classification accuracies for subjects K3b, K6b and L1b are 90.83%, 65.96% and 76.94%, respectively. These accuracies can be justified by the note that subjects have had different amounts of experience in BCI training; Subject K3b has had a lot of experience in BCI training, L1b has had only a little experience and K6b was a beginner [7].

The time course of the classification accuracy in case five experts for ensemble methods simple averaging (SA), decision templates (DT), stacked generalization (SG) and mixture of experts (ME) are presented in figure 2(a-f). As mentioned earlier, each trial was divided into small time segments with length 0.5 second. In this experiment, each of small time segments were evaluated using mentioned ensemble methods.

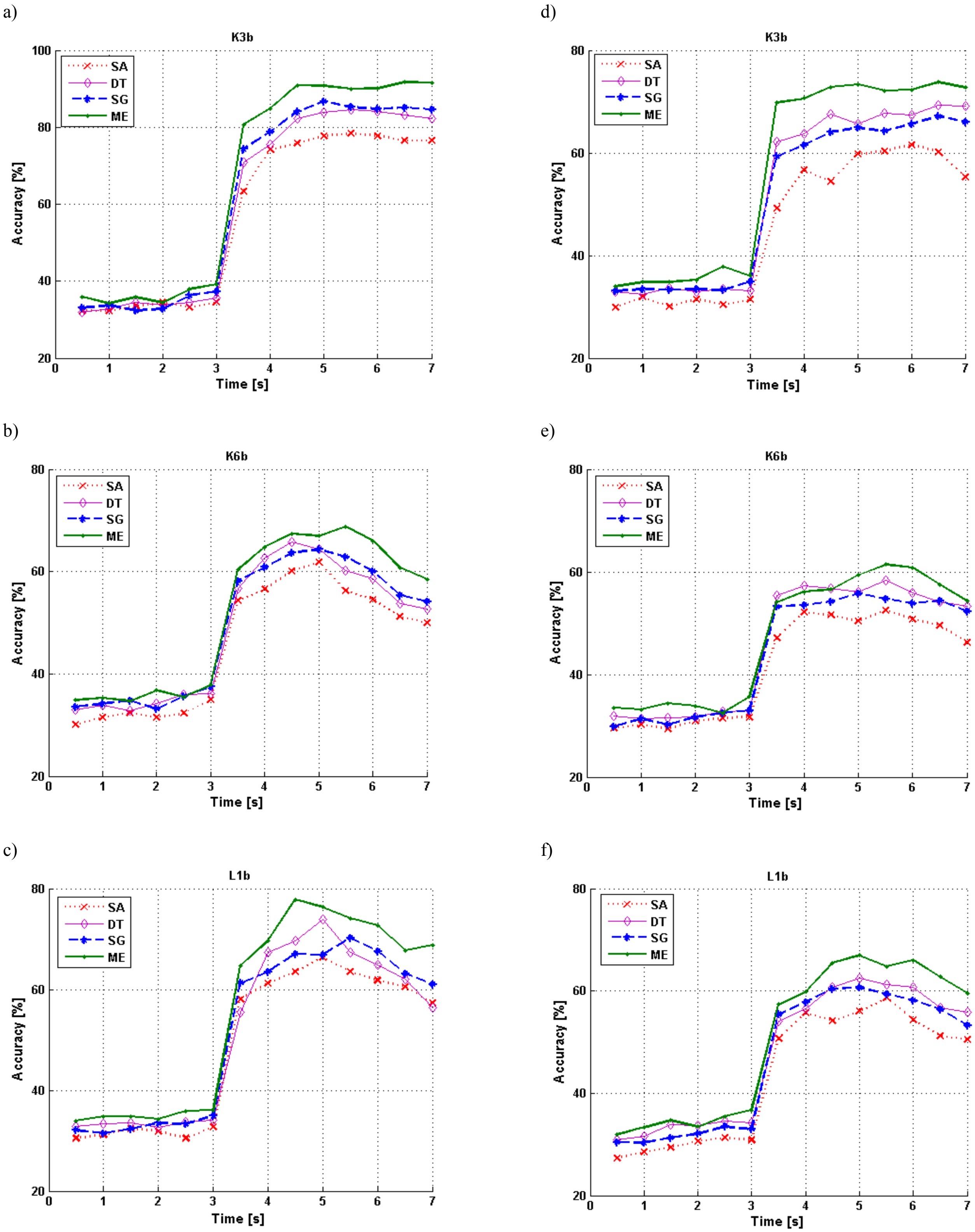

Figure 2: Classification accuracies of decision template (DT), stacked generalization (SG), simple averaging (SA) and mixture of experts (ME) for subjects K3b, K6b and L1b. Figures a-c): with DWT feature extraction method. Figures d-f): with raw EEG segments.

The figures show that, in all cases, the performance of ensemble methods appeared to oscillate in the first three seconds of the trial, when the training of the subjects has not yet started. Then simultaneous with the beginning of the training phase (at t=3s), the classification accuracies of ensemble methods increased dramatically and reached the best performance during t=4s to t= 7s.

The best classification accuracies of mixture of experts method are compared with some of the state of the art classification techniques such as k-NN, SVM and AdaBoost. To do this, k-NN with k=1 and k=3, both one-against-one (o-o) and one-against-all (o-a) approaches of SVM and AdaBoost.M1, as the multi-class version of AdaBoost, are used to classify EEG signals. The classification accuracies of 1-NN, 3-NN, SVM with two approaches, ME and AdaBoost.M1 methods for subjects K3b, K6b and L1b are shown in tables 2-4, respectively.

Table 4: The classification accuracies of classifiers for subject K3b

Methods	Feature type	
	Wavelet coefficients	Raw EEGs
1-NN	78.24	75.67
3-NN	79.56	75.98
SVM (o-o)	92.45	86.12
SVM (o-a)	90.78	87.67
ME	90.83	73.75
AdaBoost.M1	92.68	88.34

Table 5: The classification accuracies of classifiers for subject K6b

Methods	Feature type	
	Wavelet coefficients	Raw EEGs
1-NN	59.89	54.65
3-NN	60.10	56.34
SVM (o-o)	67.19	63.90
SVM (o-a)	65.23	62.45
ME	65.96	61.47
AdaBoost.M1	64.56	62.87

Table 6: The classification accuracies of classifiers for subject L1b

Methods	Feature type	
	Wavelet coefficients	Raw EEGs
1-NN	68.54	64.19
3-NN	69.56	65.98
SVM (o-o)	75.34	70.14
SVM (o-a)	76.78	71.36
ME	76.94	66.93
AdaBoost.M1	77.68	69.34

As it seen from tables 4-6, SVM and AdaBoost.M1, as the two powerful state of the art classification techniques overcome the ME method, although SVM and AdaBoost.M1 methods in comparison with ME are more time-consuming. Average classification accuracies of ensemble methods over the three subjects for two types of used features are presented in figure 3.

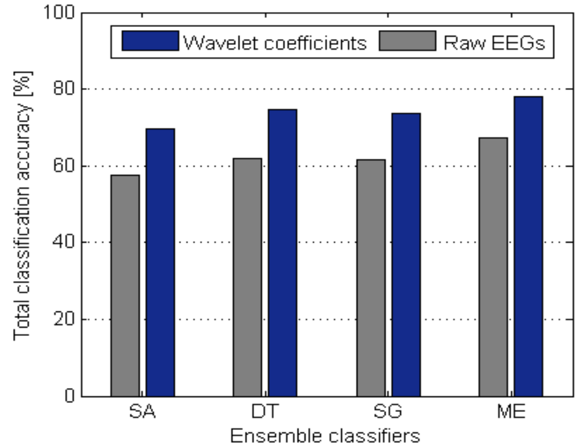

Figure 3: Average classification accuracies over three for used ensemble methods using wavelet coefficients and raw EEG segments.

The one against one approach of SVM with the classification accuracy of 78.32%, AdaBoost.M1 with the accuracy of 78.30% and ME with the accuracy of 77.91% were the best overall classifiers based on wavelet coefficients. The decision template (DT) with the accuracy of 74.77%, closely followed by stacked generalization (SG) with the accuracy of 73.58% and simple averaging (SA) with the accuracy of 69.55%.

Since there is no analytical method to achieve the best wavelet function for a particular data set, the optimal wavelet should be determined experimentally. So in our experiments apart from db2, Daubechies of order 6 (db6), Symmlet of order 4 (sym4), Symmlet of order 10 (sym10), Coiflet of order 2 (coif4) and Coiflet of order 4 (coif4) were also investigated. Total classification accuracy obtained for each wavelet function when the EEG segments were classified using mixture of experts with five expert networks is presented figure 4.

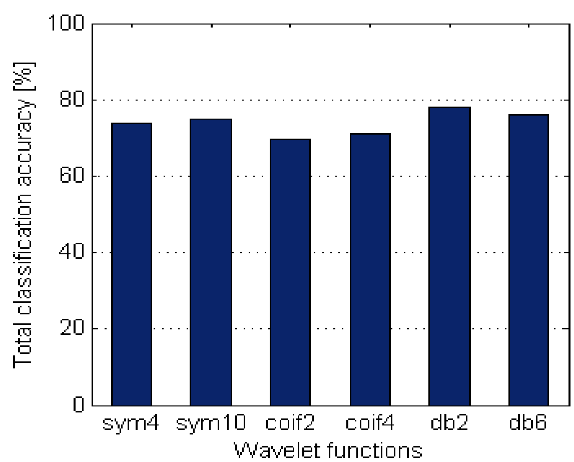

Figure 4: Total classification accuracy obtained for different wavelet functions when the EEG segments were classified using ME method with five expert networks.

As observed in figure 4, using the wavelet function of Daubechies of order 2 (db2) to decompose the EEG segments gives the highest classification accuracy of 77.91% which followed by Daubechies of order 6 (db6) with the accuracy of 76.10%.

IV. CONCLUSION

In this paper, we studied the classification of motor imagery tasks for EEG–based BCIs. Three subjects were asked to perform imaginary movements of left hand, right hand, tongue and foot during predefined time interval. EEG signals which are recorded from three subjects K3b, K6b and L1b were decomposed into wavelet coefficients by discrete wavelet transform. The statistical features of the wavelet coefficients were used as feature vectors, presenting them into classifiers. Four ensemble methods were used to evaluate the EEG signals; simple averaging (SA) and decision templates (DT) as two non-trainable combiners, stacked generalization (SG) as a data–independent trainable combiner and finally mixture of experts (ME) as a data-dependent trainable ensemble method. The classification accuracies of ensemble methods compared with the performance of k-NN, SVM and AdaBoost.M1. Experimental results showed that the ME method based on wavelet features overcome to other ensemble classifiers. Also, SVM and AdaBoost.M1, as the two state of the art classification techniques obtained the highest classification accuracy respect to the ME method.

References

[1] 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.

[2] C. Guger, A. Schlogl, C. Neuper, D. Walterspacher, T. Strein, and G. Pfurtscheller, "Rapid prototyping of an EEG-based brain-computer interface (BCI)," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on,* vol. 9, pp. 49-58, 2001.

[3] D. J. McFarland, L. A. Miner, T. M. Vaughan, and J. R. Wolpaw, "Mu and beta rhythm topographies during motor imagery and actual movements," *Brain Topography,* vol. 12, pp. 177-186, 2000.

[4] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer, "EEG-based discrimination between imagination of right and left hand movement," *Electroencephalography and clinical neurophysiology,* vol. 103, pp. 642-651, 1997.

[5] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," *Clinical Neurophysiology,* vol. 110, pp. 787-798, 1999.

[6] O. AlZoubi, I. Koprinska, and R. A. Calvo, "Classification of Brain-Computer Interface Data," 2008, pp. 123-132.

[7] F. Lee, R. Scherer, R. Leeb, C. Neuper, H. Bischof, and G. Pfurtscheller, "A comparative analysis of multi-class EEG classification for brain computer interface," 2005.

[8] T. Wang, J. Deng, and B. He, "Classifying EEG-based motor imagery tasks by means of time-frequency synthesized spatial patterns," *Clinical Neurophysiology,* vol. 115, pp. 2744-2753, 2004.

[9] N. Hazarika, J. Z. Chen, A. C. Tsoi, and A. Sergejew, "Classification of EEG signals using the wavelet transform," *Signal Processing,* vol. 59, pp. 61-72, 1997.

[10] D. Garrett, D. A. Peterson, C. W. Anderson, and M. H. Thaut, "Comparison of linear, nonlinear, and feature selection methods for EEG signal classification," *Neural Systems and Rehabilitation Engineering, IEEE Transactions on,* vol. 11, pp. 141-144, 2003.

[11] B. Blankertz, G. Curio, and K. R. Müller, "Classifying single trial EEG: Towards brain computer interfacing," 2002, p. 157.

[12] M. del R, J. Mouriño, M. Franzé, F. Cincotti, M. Varsta, J. Heikkonen, and F. Babiloni, "A local neural classifier for the recognition of EEG patterns associated to mental tasks," *Neural Networks, IEEE Transactions on,* vol. 13, pp. 678-686, 2002.

[13] C. W. Anderson and Z. Sijercic, "Classification of EEG signals from four subjects during five mental tasks," 1996, pp. 507-414.

[14] F. Lotte, M. Congedo, A. Lécuyer, F. Lamarche, and B. Arnaldi, "A review of classification algorithms for EEG-based brain–computer interfaces," *Journal of neural engineering,* vol. 4, p. R1, 2007.

[15] L. I. Kuncheva, *Combining pattern classifiers: methods and algorithms*: Wiley-Interscience, 2004.

[16] A. Schlögl, F. Lee, H. Bischof, and G. Pfurtscheller, "Characterization of four-class motor imagery EEG data for the BCI-competition 2005," *Journal of neural engineering,* vol. 2, p. L14, 2005.