Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2016) 24: 38 – 49

doi:10.3906/elk-1307-17

Investigation of the most appropriate mother wavelet for characterizing imaginary EEG signals used in BCI systems

Önder AYDEMİR*, Temel KAYIKÇIOĞLU
Department of Electrical and Electronics Engineering, Karadeniz Technical University, Trabzon, Turkey

Received: 02.07.2013 • **Accepted/Published Online:** 22.09.2013 • **Final Version:** 01.01.2016

Abstract: Feature extraction is a very challenging task, since choosing discriminative features directly affects the recognition rate of the brain–computer interface (BCI) system. The objective of this paper is to investigate the effect of mother wavelets (MWs) on classification results. To this end, features were extracted from 3 different datasets using 12 MWs, and then the signals were classified using 3 classification algorithms, including k-nearest neighbor, support vector machine, and linear discriminant analysis. The experiments proved that Daubechies and Shannon were the most suitable wavelet families for extracting more discriminative features from imaginary EEG/ECoG signals.

Key words: Continuous wavelet transform, brain computer interface, feature extraction

1. Introduction

The key idea behind brain–computer interface (BCI) systems is to allow paralyzed subjects to interact more freely with society by driving external devices, such as robotic arms, wheelchairs, and computer cursors, simply with their thoughts. The input signal of such a system may consist of various motor imagery signals such as an electroencephalogram (EEG), which is noninvasively recorded from the scalp, or an electrocorticogram (ECoG), which is invasively recorded by subdural electrodes [1].

Operation of current EEG- and ECoG-based BCI systems is generally performed in 5 main steps:

1. Signal acquisition: BCI systems begin with signal acquisition. In this step, brain signals are captured and digitized for further analysis.
2. Preprocessing: Because EEG and ECoG signals are inevitably contaminated by various artifacts such as ocular and myoelectric signals, a preprocessing stage is applied to prepare suitable signals for further steps [2].
3. Feature extraction: Feature extraction is necessary to represent input signals in a reduced feature space and identify discriminating information for every kind of signal that has been recorded.
4. Classification: The main task of this step is categorization of the signals between a fixed set of classes by taking feature vectors into account.
5. Device control interface: This step translates the categorized signals into control commands that are used to control external devices.

*Correspondence: onderaydemir@ktu.edu.tr

Among these steps, feature extraction is the most challenging because choosing discriminating features directly affects the recognition rate of the BCI system. The relevant literature has proposed several feature extraction algorithms to represent BCI signals, which include wavelet transform [1,3], power spectral density [4], autoregressive (AR) parameters [5,6], amplitude of slow cortical potentials [7], and event-related (de)synchronization (ERD/ERS) features [8]. Among these techniques, continuous wavelet transform (CWT) has received the most attention from BCI researchers. CWT gives a highly redundant representation of BCI signals in the time-scale domain [9]. Therefore, it preserves time and frequency information.

Mother wavelets (MWs) belonging to different wavelet families (WFs) can be used for extracting features from imaginary EEG/ECoG signals. Although the fundamental attributes of CWT representations using different MWs are similar, in reality these signal representations indicate different amplitudes and locations with respect to time and frequency. Therefore, choosing the WF and MW significantly affects the success of the CWT algorithm and its results. Researchers in the field of BCI have generally reported their results with only one or a few MW(s). Hsu and Sun applied CWT together with Student's two-sample t-statistics (a commonly applied technique for assessing whether the means of 2 groups are statistically different from each other or not) for feature extraction and representation of the left and right motor imagery (MI) data [10]. Using Daubechies wavelets of order 4 in the second family [11] as the MW, they calculated the means and variances of CWTs to represent left and right MI signals. In another CWT-based study, Darvishi and Al-Ani calculated the average energy value of wavelet coefficients as features representing BCI competition II, dataset III signals [12]. In their paper, they used a Morlet wavelet as the MW. Aydemir and Kayikcioglu applied their proposed method to the BCI competition III dataset I by Morlet wavelet [1]. They calculated the means and standard deviations of CWT coefficients in order to extract the features. In another approach, Bostanov used Mexican Hat as the MW to extract features from BCI competition II datasets Ib and IIb [13].

No previous studies have examined the role of MWs in BCI signal processing using different classification algorithms. The objective of this paper is to investigate the effect of MWs on classification results and find the most appropriate MW. For this purpose, features were extracted from 3 different datasets using 12 MWs, including Morlet, Shannon1-1.5, (shan1-1.5), Shannon2-3 (shan2-3), Daubechies1 (db1), Daubechies4 (db4), Symlet2 (sym2), Symlet5 (sym5), Gaussian5 (gaus5), Gaussian6 (gaus6), Meyer, Coiflet3 (coif3), and Coiflet4 (coif4). Two MW types were randomly selected from the WFs. Then, in order to reach a stable conclusion and decisively determine the most appropriate MW, 3 different EEG/ECoG-based BCI datasets were classified using 3 classification algorithms, including k-nearest neighbor (k-NN), support vector machine (SVM), and linear discriminant analysis (LDA). WFs and classification algorithms that are commonly used by BCI researchers were taken into consideration in this study.

In Section 2 of this paper, 3 kinds of EEG/ECoG-based BCI datasets are introduced, and then the CWT technique and feature extraction procedures are mathematically explained. In the last part of this section the 3 different classifiers are briefly defined. The EEG/ECoG-based BCI datasets are classified, and the results are presented in the Section 3 in the form of tables and figures. A discussion and conclusion are given at the end of the paper.

2. Materials and methods

In the following subsections, the used datasets are described, followed by a detailed description of the parts of the proposed method.

2.1. Description of the datasets

The algorithm was applied to 3 datasets: BCI competition II dataset Ia (Dataset 1), BCI competition II dataset III (Dataset 2), and BCI competition III dataset I (Dataset 3). The following subsections describe these datasets in detail.

2.1.1. Dataset 1

Dataset 1 was recorded from a single healthy subject by Blankertz et al. at the University of Tübingen, Germany [14]. The subject was asked to move a cursor up and down a computer screen, while slow cortical potentials (SCPs) of the subject were acquired. The subject received visual feedback for his/her SCPs, which were corrected for vertical eye movements. It was observed that cortical positivity (negativity) led to downward (upward) movement of the cursor on the screen.

Brain activity was recorded from 6 different channels at a sampling frequency of 256 Hz. Six EEG electrodes were located according to the standard 10–20 system and were referenced to the vertex electrode Cz as follows: Channel 1, A1 (left mastoid); Channel 2, A2 (right mastoid); Channel 3, FC3 (2 cm frontal of C3); Channel 4, CP3 (2 cm parietal of C3); Channel 5, FC4 (2 cm frontal of C4); and Channel 6, CP4 (2 cm parietal of C4). Each trial lasted for 6 s. During every trial, the task was visually presented by a highlighted goal at the top or bottom of the screen to indicate up or down from 0.5 s until the end of the trial. The visual feedback was presented from 2 to 5.5 s. Only this 3.5-s interval for every trial was provided for training and testing.

All trials were separated into a training set (268 trials, 135 for cursor-up and 133 for cursor-down) and a testing set (293 trials), both of which contained EEG data only from the 3.5-s feedback phase of each trial. The purpose was to categorize the trials in the testing set into cursor-up or cursor-down groups.

2.1.2. Dataset 2

Dataset 2 was recorded from a single healthy female subject at the Institute for Biomedical Engineering, Graz University of Technology, Austria. Brain activity was acquired with 3 bipolar EEG channels (C3, Cz, C4) at a sampling frequency of 128 Hz and was filtered between 0.5 and 30 Hz. The recording length of the trial was set to 9 s. The first 2 s were quiet. At t = 2 s, an acoustic stimulus indicated the beginning of the trial and a cross ('+') was displayed for 1 s. Then, at t = 3 s, an arrow (left/right) was displayed as a cue. The subject was asked to imagine left or right hand movements and move the feedback bar toward the cue direction. The order of the left and right cues was random. During the experiment, the subject sat in a comfortable chair with armrests. It was decided to use only the last 6 s, because no event occurred in the first 3 s.

The experimental dataset consisted of 140 trials for the training set (70 trials for right hand movement, RHM, and 70 trials for left hand movement, LHM) and 140 trials for the testing set (70 trials for RHM and 70 trials for LHM). For further information about the dataset, please refer to [15] and [16]. The purpose was to categorize the trials in the testing set into RHM or LHM.

2.1.3. Dataset 3

Dataset 3 was recorded from an epileptic subject on 2 different days with approximately 1 week of delay by Lal et al. at the University of Tübingen, Germany [17]. In both sessions, the subject was asked to imagine either a left small finger or a tongue movement.

The datasets consisted of 278 trials (training data: 139 trials for finger movements and 139 trials for tongue movements), which were performed during the first session, and 100 trials (testing data) from the

second session. Each trial lasted for 3 s. To prevent visually evoked potentials from being reflected by the data, the recording intervals started 0.5 s after the end of the visual cue. Electrical brain activity was recorded by an 8 × 8 ECoG platinum electrode grid (from a total of 64 points) that was placed on the contralateral (right) motor cortex. All the recordings were performed at a sampling rate of 1 kHz (acquired 3000 samples per channel for every trial). For further information about the dataset, please refer to [17]. The purpose was to categorize the trials in the test set as finger or tongue movement imagery.

2.2. Continuous wavelet transform (CWT)

CWT is a powerful signal processing tool that is used for many BCI data analysis applications in order to extract feature(s) [1,10–13].

CWT is defined as the convolution of signal $x(t)$ with wavelet function $\psi_{\tau,s}(t)$,which is given by:

$$CWT_x^{\psi}(\tau,s) = \frac{1}{\sqrt{|s|}} \int x(t)\psi^* \left(\frac{t-\tau}{s}\right) dt, \tag{1}$$

where $\psi_{\tau,s}(t)$ is dilated and shifted versions of the wavelet function $\psi(t)$, and is defined as follows:

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{s}}\psi(\frac{t-\tau}{s}). \tag{2}$$

Here, t, τ, and s denote time parameter, shift parameter, and scale parameter, respectively [18]. The wavelet function $\psi_{\tau,s}(t)$ has a zero mean, as given in Eq. (3):

$$\int_{-\infty}^{+\infty} \psi_{\tau,s}(t)dt = 0. \tag{3}$$

An important point of wavelet analysis is choosing a particular wavelet function $\psi(t)$.

2.3. Feature extraction

Two types of feature extraction procedures were utilized to represent BCI signals. The first procedure was used for Datasets 1 and 3. The second was used for Dataset 2. These 2 procedures are described in the following subsections.

2.3.1. Feature extraction procedure 1

First, a variance normalization process was implemented in all trials, as given in Eq. (4) [1]:

$$T_N = \frac{T}{std(T)}, \tag{4}$$

where T_N and $std(T)$ denote normalized trial signal and standard deviation of the trial signal, respectively. The estimated standard deviation of a trial was calculated with the following formula:

$$std(T) = \sqrt{\frac{\sum (T-\bar{T})^2}{n-1}}, \tag{5}$$

where $\bar{T}$ is the mean value of all samples of the trial and n is the length of the trial.

After the normalization process, wavelet transform coefficients (WTCs) of the signal bands, including theta (4–8 Hz), alpha (8–12 Hz), beta (12–20 Hz), and total band (4–20 Hz), were calculated. The scale of the wavelet function was set to integer values of the corresponding MW with a step size of $\tau = 1$. Empirical feature analysis of the training sets demonstrated that the means and standard deviations of the absolute values of the WTCs of all 6 channels of Dataset 1, and the 12th and the 29th channels of Dataset 3, yielded discriminative feature sets. Hence, those features were selected to classify both tasks. Mean and standard deviations were calculated with the following formulas, respectively:

$$WTCs^{Avr} = \frac{\sum |WTCs|}{L_{WTCs}}, \tag{6}$$

$$WTCs^{std} = \sqrt{\frac{\sum (|WTCs| - WTCs^{Avr})^2}{L_{WTCs} - 1}}, \tag{7}$$

where L_{WTCs} is the length of WTCs. All computed WTCs of the selected channels were used to calculate features.

2.3.2. Feature extraction procedure II

Since feature extraction procedure I did not provide a discriminative feature set for Dataset 2, feature extraction procedure II was developed.

WTCs of the signal bands were obtained by applying the same procedure. Empirical feature analysis of the training set demonstrated that total energy (TE) of WTCs of all 3 channels yielded discriminative feature sets for Dataset 2. Hence, those 3-dimensional features were used to classify both tasks. Total energy of WTCs was calculated with the following formula:

$$TE = \sum |WTCs|^2. \tag{8}$$

2.4. Classification algorithms

In the present study, k-NN, SVM, and LDA classification algorithms were trained using training datasets. In the following subsections, the aspects of the 3 classifiers are briefly reviewed.

2.4.1. k-nearest neighbor (k-NN)

k-NN classifier is a common classification algorithm that determines a testing sample's class by the majority class of k closest training samples [18,19]. This issue is illustrated with a simple example in Figure 1, which shows each data record with 2 attributes that represent 2 classes of data (plus, '+', and circle, 'o'). In this case, k = 6. The unlabeled test trial would be labeled by category of class plus, because 4 of its 6 closest samples (neighbors) are plus.

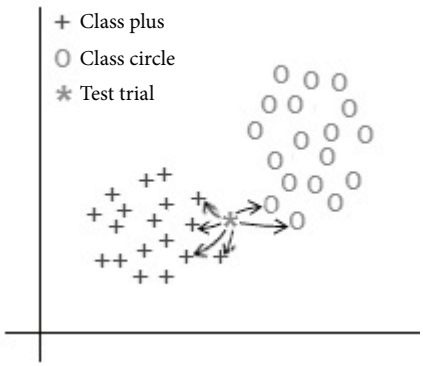

Figure 1. A simple example of the k-NN algorithm.

In the present work, the Euclidean distance metric and leave-one-out cross-validation technique were used to determine the best value of k for maximizing the classification performance. The k value was sought in the interval between 1 and 25 with a step size of 1.

2.4.2. Support vector machine (SVM)

The SVM performs classification tasks by constructing the best hyperplane in a multidimensional space through finding the maximum margin [20–22]. In this paper, the SVM classification framework was implemented by the following expression:

$$f(x) = sign(\sum_{i=1}^{T} \alpha_i y_i K(x, x_i) + b), \tag{9}$$

where $f(x)$ is the decision function, T is the number of trials, $\alpha_i \in R$ are Lagrangian multipliers obtained by solving a quadratic optimization problem, $y_i \in \{1, -1\}$ are class labels, b is bias, and $K(x, x_i)$ is kernel function. Although there are many alternatives for kernel function, the most common radial basis function kernel was used as:

$$K(x, x_i) = \exp(-\frac{\|x - x_i\|^2}{2\sigma^2}). \tag{10}$$

This kernel was chosen because the number of its hyperparameters was smaller than that of other kernels. This kernel function was specified by the scaling factor sigma, σ. The same validation procedure as in the k-NN classification algorithm was used to determine the optimum value for the scaling factor. The most appropriate σ value was sought in intervals of 0.1 to 2.5 (step size: 0.2) with the same validation procedure used in the k-NN classification algorithm.

2.4.3. Linear discriminant analysis (LDA)

LDA categorizes 2 classes based on the assumption that both classes are under normal distribution with equal covariance matrices. The discriminating hyperplane is obtained by finding the projection of the labeled training data that maximizes the distance between the means of 2 classes and minimizes the interclass variance. The main aim is to solve the following problem:

$$y = w^T x + w_0, \tag{11}$$

where x is feature vector. Vectors w and w_0 are determined by maximizing interclass means and minimizing interclass variance [23].

3. Results

The results of training and testing classification accuracy (CA) for Datasets 1, 2, and 3 are presented in Figures 2, 3, and 4, respectively. In these figures, TrCA represents training classification accuracy and TCA represents test classification accuracy. Classification accuracy is defined as the percentage of the number of trials classified correctly in the corresponding (training or testing) set over total trials. With respect to the CA results of the 3 classifiers, the following was found from Figures 2–4:

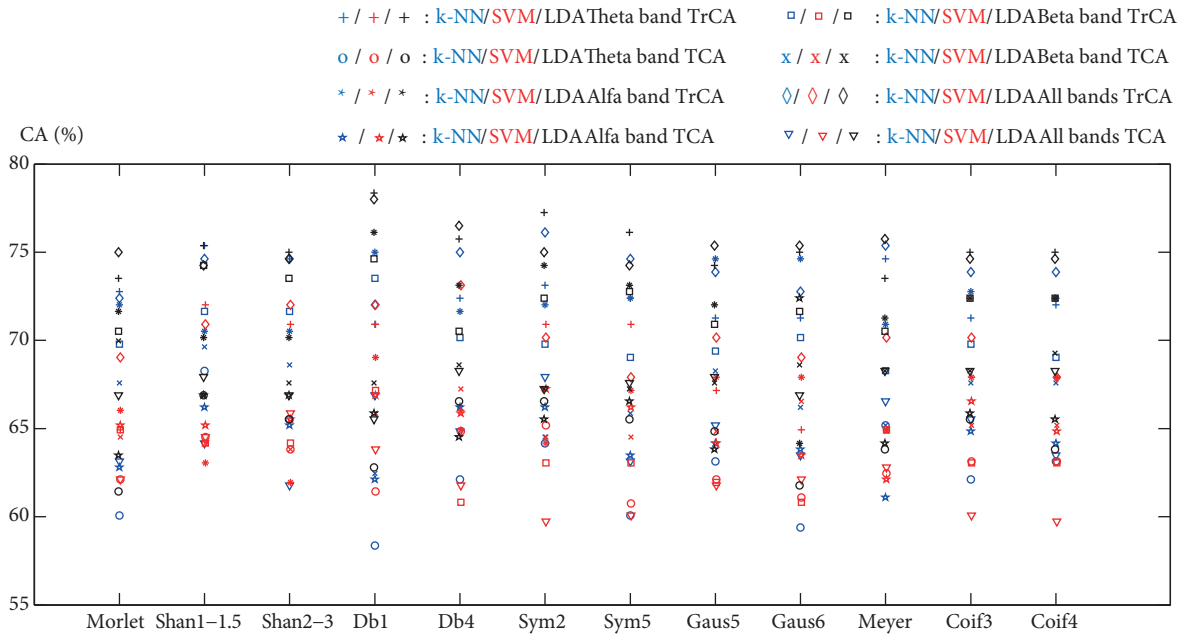

Figure 2. The training and test classification accuracy results of the classification algorithms for Dataset 1.

1) For Dataset 1, LDA provided the highest value of CA on the training dataset as 78.36% when theta band features were extracted by Morlet wavelet. On the contrary, the lowest value of CA was calculated by k-NN (k was calculated as 3) on the testing dataset as 58.36% when theta band features were extracted with the db1 wavelet.

2) For Dataset 2, k-NN (k was calculated as 21) had the highest value of CA on the testing dataset as 85.00% when theta band features were extracted by the shan2-3 wavelet. In addition, the lowest CA was obtained by k-NN (k was calculated as 4) on the testing dataset as 52.14% when theta band features were extracted using the db1 wavelet.

3) For Dataset 3, the highest value of CA was obtained by SVM (σ was calculated as 2.5) on the testing dataset as 93.00% when alpha band features were extracted by the sym2 wavelet. However, the lowest value of CA was calculated by k-NN (k was calculated as 21) on the testing dataset as 45.00% when theta band features were extracted by the Meyer wavelet.

4) There was no single MW that could perform the best CA with all the classifiers and on each dataset.

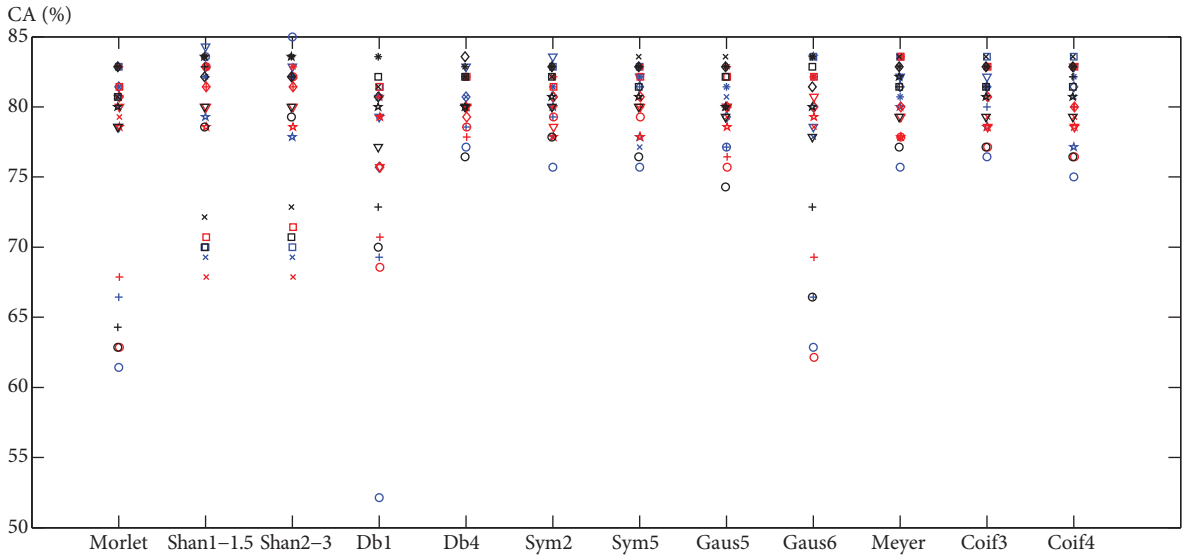

Figure 3. The training and test classification accuracy results of the classification algorithms for Dataset 2.

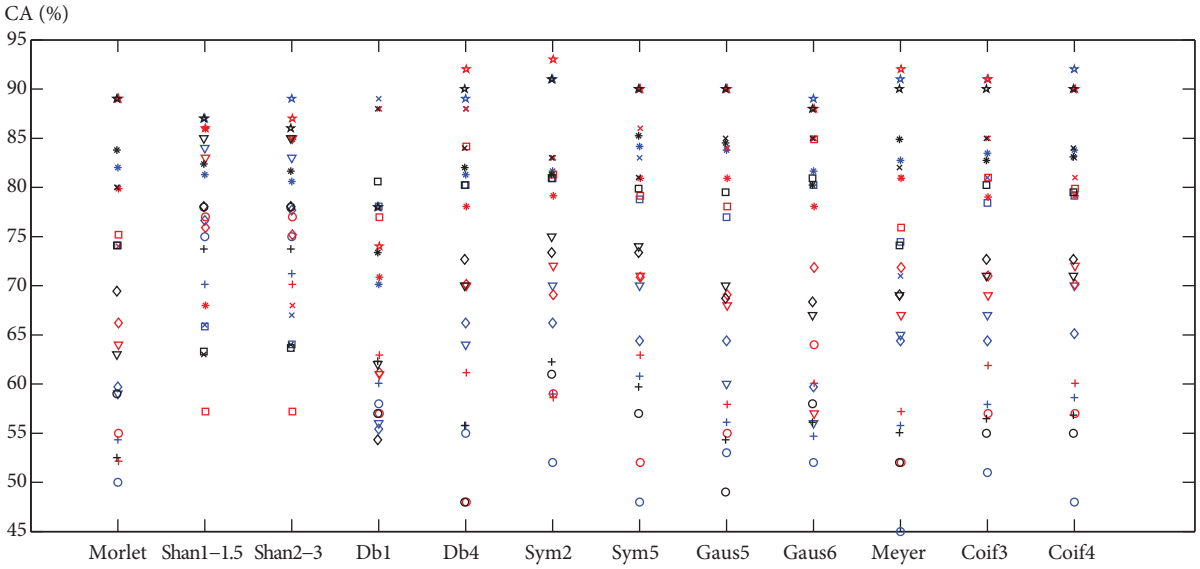

Figure 4. The training and test classification accuracy results of the classification algorithms for Dataset 3.

According to the performance results of the individual classifiers with the considered MWs, it is difficult to propose the most appropriate MW, because the best and worst classification accuracies obviously depend on the feature set, classifier, and MW. It is well known that if the results of different datasets are not comparable, their means and standard deviations are significantly different. Hence, the means and standard deviations of the CA results were calculated for a general comparison. The means and standard deviations of the CA of Datasets 1, 2, and 3 are given in Tables 1, 2, and 3, respectively. These tables demonstrate that although the average classification accuracy (ACA) results were very close, the classification algorithms reached their highest ACA using different MWs.

For Dataset 1 (Table 1), LDA had the highest mean value of CA on the training dataset as 76.78% when the features were extracted by the db1 wavelet. As seen in Table 2, LDA provided the highest mean values of

CA on the training dataset as 82.68% when the features were extracted by the db4 or sym2 wavelets. Finally, as seen in Table 3, SVM provided the highest mean value of CA on the testing dataset as 79.25% when the features were extracted by the shan2-3 wavelet.

Table 1. Averages and standard deviations of the classification accuracy of Dataset 1.

Mother	k-NN		SVM		LDA	
wavelet	Training	Test	Training	Test	Training	Test
Morlet	71.74 ± 1.34	63.40 ± 3.11	66.23 ± 1.94	63.49 ± 1.60	72.67 ± 1.99	65.44 ± 3.77
shan1-1.5	73.04 ± 2.33	67.07 ± 2.39	67.54 ± 4.57	64.59 ± 0.43	73.51 ± 2.30	67.15 ± 0.52
shan2-3	72.86 ± 2.10	65.28 ± 2.79	67.26 ± 4.96	64.76 ± 1.09	73.32 ± 2.21	66.72 ± 0.86
db1	72.86 ± 1.78	62.46 ± 3.49	69.78 ± 2.14	64.50 ± 2.41	76.78 ± 1.73	65.45 ± 1.98
db4	72.30 ± 2.03	64.85 ± 1.93	66.23 ± 5.12	64.94 ± 2.32	73.97 ± 2.72	66.98 ± 1.88
sym2	72.76 ± 2.64	65.70 ± 1.73	67.07 ± 4.02	64.17 ± 3.18	74.72 ± 2.01	66.64 ± 0.81
sym5	72.11 ± 2.31	63.14 ± 2.38	67.26 ± 3.23	62.89 ± 2.95	74.07 ± 1.51	66.73 ± 0.91
gaus5	72.30 ± 2.41	65.19 ± 2.21	66.79 ± 3.47	63.23 ± 1.51	73.14 ± 2.04	66.04 ± 2.02
gaus6	72.20 ± 1.94	63.23 ± 2.83	65.67 ± 3.67	63.31 ± 2.37	71.54 ± 5.20	67.42 ± 4.40
Meyer	72.30 ± 3.32	64.51 ± 2.37	66.24 ± 2.61	63.06 ± 1.23	72.76 ± 2.36	66.13 ± 2.47
coif3	71.92 ± 1.78	65.02 ± 2.26	67.26 ± 2.99	63.74 ± 2.82	73.60 ± 1.41	66.98 ± 1.49
coif4	71.83 ± 2.03	64.59 ± 2.04	66.70 ± 2.43	63.23 ± 2.50	73.60 ± 1.41	66.72 ± 2.50

Table 2. Averages and standard deviations of the classification accuracy of Dataset 2.

Mother	k-NN		SVM		LDA	
wavelet	Training	Test	Training	Test	Training	Test
Morlet	78.04 ± 7.77	75.18 ± 9.21	78.22 ± 6.96	75.18 ± 8.23	77.68 ± 8.98	75.54 ± 8.50
shan1-1.5	79.29 ± 6.20	79.11 ± 6.91	79.11 ± 5.64	77.32 ± 6.56	79.64 ± 6.46	78.57 ± 4.78
shan2-3	79.11 ± 6.07	78.75 ± 6.98	79.29 ± 5.28	77.14 ± 6.36	79.64 ± 5.99	78.93 ± 4.46
db1	76.79 ± 5.61	72.86 ± 13.83	77.14 ± 4.99	75.72 ± 5.05	79.82 ± 4.79	77.14 ± 5.09
db4	80.89 ± 1.69	80.18 ± 2.36	80.36 ± 2.14	79.64 ± 0.72	82.68 ± 0.69	79.64 ± 2.37
sym2	81.07 ±1.49	79.29 ± 3.35	81.07 ± 0.92	78.93 ± 0.92	82.68 ± 0.36	80.18 ± 1.78
sym5	81.97 ± 0.68	78.21 ± 2.77	81.43 ± 1.31	79.47 ± 1.22	82.50 ± 0.72	80.18 ± 2.94
gaus5	80.00 ± 2.26	79.46 ± 1.58	80.36 ± 2.89	78.39 ± 1.88	81.97 ± 1.35	79.29 ± 3.82
gaus6	78.04 ± 7.88	75.72 ± 8.94	78.39 ± 6.15	75.18 ± 8.74	80.18 ± 4.96	76.97 ± 7.41
Meyer	81.96 ± 1.22	79.64 ± 2.77	81.25 ± 2.82	78.75 ± 1.07	82.15 ± 0.83	80.54 ± 2.88
coif3	81.61 ± 1.47	79.64 ± 2.64	81.79 ± 1.24	78.39 ± 0.90	82.15 ± 0.83	80.18 ± 2.70
coif4	81.79 ± 1.49	78.93 ± 3.52	81.43 ± 1.65	78.22 ± 1.24	82.32 ± 0.69	80.00 ± 2.97

The simplest way to determine the overall performance of the MWs and choose one of them is the voting method. Table 4 indicates the MW with which the classification algorithms achieved the highest ACA. According to the voting method, both the Shannon and Daubechies families provided the highest ACA 6 times with different classifiers. On the other hand, Meyer and Coiflet provided the highest ACA only once, whereas the Morlet wavelet never provided the highest ACA.

4. Discussion and conclusion

This paper presented the use of CWT-based methodology for extracting features from imaginary EEG/ECoG signals and proposed the most suitable MW for this purpose. The proposed approach was based on evaluating classification accuracies between the calculated means, standard deviations, and energy of the wavelet coefficients for 12 different MWs.

Table 3. Averages and standard deviations of the classification accuracy of Dataset 3.

Mother wavelet	k-NN		SVM		LDA	
	Training	Test	Training	Test	Training	Test
Morlet	67.54 ± 12.76	69.50 ± 18.08	68.35 ± 12.19	70.50 ± 14.57	69.97 ± 13.08	72.75 ± 14.15
shan1-1.5	73.47 ± 6.85	78.00 ± 9.49	71.76 ± 12.19	78.50 ± 7.94	74.37 ± 8.17	78.25 ± 10.87
shan2-3	73.38 ± 7.36	78.50 ± 9.57	71.85 ± 11.53	79.25 ± 8.66	74.28 ± 7.78	78.25 ± 10.15
db1	65.92 ± 10.17	70.25 ± 15.97	67.99 ± 7.33	70.00 ± 14.02	67.65 ± 11.64	71.25 ± 14.32
db4	70.87 ± 12.20	74.00 ± 17.15	73.38 ± 9.97	74.50 ± 20.09	72.66 ± 11.98	73.00 ± 18.66
sym2	71.95 ± 11.20	74.00 ± 17.03	72.03 ± 10.40	76.75 ± 14.62	74.46 ± 8.94	77.50 ± 14.62
sym5	72.03 ± 11.22	72.75 ± 18.46	73.47 ± 8.28	74.751 ± 17.23	74.55 ± 11.02	75.50 ± 13.96
gaus5	70.33 ± 12.43	71.75 ± 18.01	71.49 ± 10.37	74.25 ± 15.84	71.77 ± 13.37	73.50 ± 18.41
gaus6	69.07 ± 13.88	70.50 ± 19.19	73.72 ± 10.54	73.50 ± 15.33	71.41 ± 11.71	74.50 ± 14.39
Meyer	69.34 ± 11.75	68.00 ± 18.94	71.47 ± 10.22	73.50 ± 17.34	70.78 ± 12.39	73.25 ± 16.60
coif3	71.04 ± 11.90	72.50 ± 17.39	73.22 ± 8.71	75.50 ± 15.44	73.02 ± 11.83	75.25 ± 15.71
coif4	71.67 ± 11.78	73.25 ± 19.10	72.30 ± 9.28	75.00 ± 14.07	73.02 ± 11.62	75.00 ± 15.51

Table 4. Performance of the MWs.

		Dataset 1	Dataset 2	Dataset 3
k-NN	TrCA	shan1-1.5	sym5	shan1-1.5
	TCA	shan1-1.5	db4	shan2-3
SVM	TrCA	db1	coif3	gaus6
	TCA	db4	db4	shan2-3
LDA	TrCA	db1	db4 & sym2	sym5
	TCA	gaus6	meyer	shan1-1.5 & shan2-3

The results showed that the EEG/ECoG signals can be represented using CWT coefficients. However, it is important to develop a suitable feature extraction procedure for each individual dataset. Hence, since feature extraction procedure I did not provide a discriminative feature set, feature extraction procedure II was developed for Dataset 2.

It was observed that the LDA algorithm achieved much better performance than the SVM and k-NN algorithms in terms of the obtained highest classification accuracy results. Moreover, the LDA classifier was more robust than the SVM and k-NN algorithms, because it had only limited flexibility (less free parameters to tune) and was less prone to overfitting.

According to the obtained statistical voting results, it could be generalized that Daubechies and Shannon were the most suitable WFs for extracting more discriminative features from the imaginary EEG/ECoG signals. The general characteristics of these WFs should be examined to understand and reveal the cause of their superior performance. In comparison to other tested wavelets, Daubechies wavelets are compactly supported with extreme phase and the highest number of vanishing moments, which are a necessary condition for the smoothness of the wavelet function and for a given support width. On the other hand, Shannon wavelets are analytically defined, infinitely differentiable, and sharply bounded in the frequency domain. These advantages provide a very good localization of energy in the frequency domain and make those WFs the best candidates to identify the EEG/ECoG-based BCI signals. On the contrary, it is worthwhile to mention that the Morlet wavelet presented the poorest performance due to the fact that it is a Fourier-based time-frequency transformation and thus suffers from many of the shortcomings of Fourier analysis.

It can be concluded that the proposed study could greatly contribute to the selection of a suitable MW through parameterization that leads to performance improvement of EEG/ECoG-based BCI signal classification in comparison to random selection of the MW.

The results also proved that the performance of the classifiers depended on the characteristics of the datasets. Therefore, it is obvious that for another kind of EEG/ECoG signals, researchers should specifically seek the most appropriate MW among the existing members of the WFs.

Acknowledgment

This work was supported by the Scientific and Technological Research Council of Turkey (TÜBİTAK), Project No: EEEAG-110E035.

References

[1] Aydemir O, Kayikcioglu T. Wavelet transform based classification of invasive brain computer interface data. Radioengineering 2011; 20: 31–38.

[2] Nicolas-Alonso LF, Gomez-Gil J. Brain computer interfaces, a review. Sensors 2012; 12: 1211–1279.

[3] Zalay OC, Kang EE, Cotic M, Carlen PL, Bardakjian BL. A wavelet packet-based algorithm for the extraction of neural rhythms. Ann Biomed Eng 2009; 37: 595–613.

[4] Herman P, Prasad G, McGinnity TM, Coyle D. Comparative analysis of spectral approaches to feature extraction for EEG-based motor imagery classification. IEEE T Neur Sys Reh 2008; 16: 317–326.

[5] Burke DP, Kelly SP, de Chazal P, Reilly RB, Finucane C. A parametric feature extraction and classification strategy for brain–computer interfacing. IEEE T Neur Sys Reh 2005; 13: 12–17.

[6] Kamrunnahar M, Dias NS, Schiff SJ. Toward a model-based predictive controller design in brain-computer interfaces. Ann Biomed Eng 2011; 39: 1482–1492.

[7] Birbaumer N, Kubler A, Ghanayim N, Hinterberger T, Perelmouter J, Kaiser J, Iversen I, Kotchoubey B, Neumann N, Flor H. The thought translation device (TTD) for completely paralyzed patients. IEEE T Rehabil Eng 2000; 8: 190–193.

[8] Pfurtscheller G, Brunner C, Schlögl A, Da Silva FHL. Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks. NeuroImage 2006; 31: 153–159.

[9] Samar VJ, Bopardikar A, Rao R, Swartz K. Wavelet analysis of neuroelectric waveforms: a conceptual tutorial. Brain Lang 1999; 66: 7–60.

[10] Hsu WY, Sun YN. EEG-based motor imagery analysis using weighted wavelet transform features. J Neurosci Meth 2009; 176: 310–318.

[11] Daubechies I. Orthonormal bases of compactly supported wavelets. Comm Pure Appl Math 1988; 41: 909–996.

[12] Darvishi S, Al-Ani A. Brain–computer interface analysis using continuous wavelet transform and adaptive neuro-fuzzy classifier. In: Proceedings of the 29th International Annual Conference of the IEEE EMBS; 22–26 August 2007; Lyon, France. New York, NY, USA: IEEE. pp. 3220–3223.

[13] Bostanov V. BCI competition 2003—data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram. IEEE T Biomed Eng 2004; 51: 1057–1061.

[14] Blankertz B, Müller KR, Curio G, Vaughan TM, Schalk G, Wolpaw GR, Schlögl A, Neuper C, Pfurtscheller G, Hinterberger T et al. The BCI competition 2003: progress and perspectives in detection and discrimination of EEG single trials. IEEE T Biomed Eng 2004; 51: 1044–1051.

[15] Lin SC, Chang YCI, Yang WN. Meta-learning for imbalanced data and classification ensemble in binary classification. Neurocomputing 2009; 73: 484–494.

[16] Neuper C, Schlögl A, Pfurtscheller G. Enhancement of left-right sensorimotor EEG differences during feedback-regulated motor imagery. J Clin Neurophysiol 1999; 16: 373–382.

[17] Lal TN, Hinterberger T, Widman G, Schröder M, Hill J, Rosenstiel W, Elger CE, Schölkpf B, Birbaumer N. Methods towards invasive human brain computer interfaces. Adv Neur In 2005; 17: 737–744.

[18] Duda RO, Hart PE, Stork DG. Pattern Classification. 2nd ed. New York, NY, USA: Wiley, 2001.

[19] Avendaño-Valencia LD, Godino-Lorente JI, Blanco-Velasco M, Castellanos-Dominguez G. Feature extraction from parametric time-frequency representations for heart murmur detection. Ann Biomed Eng 2010; 38: 2716–2732.

[20] Vapnik VN. Statistical Learning Theory. New York, NY, USA: Wiley, 1998.

[21] Tu W, Sun S. Semi-supervised feature extraction for EEG classification. Pattern Anal Appl 2013; 16: 213–222.

[22] Zadeh AE, Khazaee A. High efficient system for automatic classification of the electrocardiogram beats. Ann Biomed Eng 2011; 39: 996–1011.

[23] Müller KR, Anderson CW, Birch GE. Linear and nonlinear methods for brain-computer interfaces. IEEE T Neur Syst Reh 2003; 11: 165–169.