Turkish Journal of Electrical Engineering & Computer Sciences

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

Research Article

Turk J Elec Eng & Comp Sci
(2023) 31: 1168 – 1186
© TÜBİTAK
doi:10.55730/1300-0632.4041

A practical low-dimensional feature vector generation method based on wavelet transform for psychophysiological signals

Erdem ERKAN[1], **Yasemin ERKAN**[2,*]

[1]Department of Computer Engineering, Faculty of Engineering, Architecture, and Design, Bartın University, Bartın, Turkiye

[2]Department of Electrical-Electronics Engineering, Faculty of Engineering, Architecture, and Design, Bartın University, Bartın, Turkiye

Received: 03.04.2023 • **Accepted/Published Online:** 28.07.2023 • **Final Version:** 30.11.2023

Abstract: High-dimensional feature vectors entail computational cost and computational complexity. However, a successful classification can be obtained with an optimally sized feature vector consisting of distinctive features. With the widespread use of the internet and mobile devices, the need for systems with low computational costs is increasing day by day. In this study, starting from the idea that each motor imagery is represented as a subject-specific pattern in the brain, we propose a new and practical method that can generate a low-dimensional feature vector based on wavelet transform. The feature vector is obtained from the correlation between each trial and each class average. To investigate the effect of possible temporal shifts in the trial signals, the proposed method is analyzed with signal segments with different starting points and lengths. The effect of these signal segments on classification is shown. The proposed feature extraction approach is tested on two different datasets and the classification results are presented in comparison with previous studies. With the method proposed in this study, much lower-dimensional feature vectors are obtained compared to previous studies and very satisfactory results are obtained. It is observed that EEG signals related to motor imagery in the brain have a subject-specific pattern, and this pattern is successfully classified with a feature vector consisting of only 1 feature per class.

Key words: Brain–computer interface, signal processing, feature extraction, wavelet transform

1. Introduction

Classification is a learning phenomenon performed by using distinctive features. Living organisms develop behaviors against internal and external stimuli through classification. In order to understand the nature of this learning adventure and to develop successful classification algorithms, researchers have been trying to identify the distinctive features that reveal the differences between psychophysiological signals such as electroencephalography (EEG) and magnetoencephalography. Feature extraction in EEG refers to the process of identifying relevant and informative characteristics or patterns from raw EEG signals. These extracted features are then used as inputs for further analysis or classification tasks. Effective feature extraction is crucial for capturing the discriminative information contained within the EEG data. Various common techniques are used for feature extraction in EEG analysis. For example, time-domain features are used to capture the properties of EEG signals in the time domain. Examples include mean, variance, skewness, kurtosis, and other statistical measures calculated over different time windows or segments of the signal. With frequency-domain features,

*Correspondence: yerkan@bartin.edu.tr

EEG signals can be transformed into the frequency domain using techniques such as fast Fourier transform (FFT) or discrete wavelet transform (DWT) [1]. Frequency-domain features include spectral power, dominant frequency, spectral entropy, and band power in specific frequency bands (e.g., delta, theta, alpha, beta, and gamma). Using time-frequency features, EEG signals exhibit dynamic changes in both time and frequency domains. Time-frequency analysis methods like the Fourier or wavelet transform can be applied to capture the signal's spectral changes over time. Time-frequency features include spectrograms, power spectral density estimates, and event-related spectral perturbation (ERSP). Common spatial patterns [2] and the Savitzky–Golay filter [3] are also used to find the most distinctive features in signal analysis. However, wavelet transform-based feature extraction methods have also been used in EEG analysis to decrease the size of the feature vector and determine the feature subset that maximizes classification success [4]. The classifier method and the feature vector used with this classifier are the most important factors affecting the success of the classification task. Common classification methods such as support vector machines (SVMs) [5], the K-nearest neighbors (KNN) algorithm [6], linear discriminant analysis (LDA) [7], and artificial neural networks [8] have been widely used in brain–computer interface (BCI) studies. In addition, convolutional neural networks (CNNs), as a type of artificial neural network that is very popular today, are also used to classify EEG signals and analyze EEG data [9]. In one such study, Turk compared traditional classification methods with a CNN in an EEG classification task and presented the results obtained from using hybrids of CNNs with traditional methods [10]. CNNs owe their popularity to the amount of data that feeds them. However, this machine learning method, which is frequently used in image processing, needs thousands of images for successful clustering and classification. This is a disadvantage in terms of computational cost. While traditional methods such as the KNN algorithm and SVMs have the advantage of working with less data, CNNs can handle the feature extraction step themselves. However, with the widespread use of the internet and mobile devices, the need for systems with fast speed and low computational costs is increasing day by day. The main purpose of signal classification is to achieve a successful classification result with an optimal number of features.

In this study, to reduce the computational cost, we propose a new feature extraction method that has an optimal number of features. The proposed method was tested on two different public datasets, which we present below. Datasets obtained by magnetoencephalography (MEG) and electrocorticography (ECoG) methods and given in BCI Competitions IV and III, respectively, were used to evaluate the proposed method. Sardouie and Shamsollahi achieved 46.9% (Subject 1 = 59.5%, Subject 2 = 34.3%) accuracy with the MEG dataset as the best classification success in the competition. They achieved this success with a feature vector obtained by reducing the frequency domain features of 12 channels and the wavelet coefficients with a supervised algorithm. Moreover, they used a linear SVM and LDA classification methods together and reached optimal classification accuracy using a genetic algorithm (GA) [11]. For second place in the competition, Li et al. performed a classification with average accuracy of 25.1% (Subject 1 = 31.1%, Subject 2 = 19.2%) using an 8-Hz low-pass filter. In that work, for generating the feature vector, they used the principal components of the amplitude and phase of the Fourier transform of each sample. They also reduced the size of frequency characteristics with Fisher discriminant analysis (FDA) [12]. In another study, the MEG dataset was classified with a linear SVM classifier of the feature vector consisting of wavelet coefficients. That study achieved 23.9% (Subject 1 = 16.2%, Subject 2 = 31.5%) accuracy [12]. When studies with the same MEG dataset other than those from the aforementioned competitions are examined, Goni et al. achieved 64% success for only Subject 1 using UNEQ class models with autoregressive features [13]. Hatipoğlu et al. first converted the signals into an angle-amplitude graphic image. They then obtained the features by combining image feature detection, feature transform, and the visual bag-of-words method. They achieved 78.09% and 79.88% classification success for both subjects with 500 features per channel using the KNN algorithm [14]. The ECoG dataset was obtained in an experiment in which a subject performed imaginary movements of the left little finger or tongue. For the ECoG dataset, the highest classification accuracy of 79% was obtained with 58% kappa using only six features per channel [15]. On the other hand, in other studies, higher classification successes have been achieved. In their study, Qi et al. performed channel and feature selection with a method based on particle swarm optimization [16].

In this study, unlike the previous literature, a practical feature extraction method based on wavelet transform, which reduces the computational complexity by producing very low-dimensional feature vectors, is proposed. In the proposed method, the feature vector is generated by calculating the correlation of the signal reduced in size by wavelet transform with the mean of each class. The obtained feature vectors are classified by SVM and classification results are presented in comparison with previous studies. In addition, we think that EEG signals with a complex structure should be classified with individual feature extraction methods, and there may be subject-specific information in different segments of EEG signals. In a previous study, it was hypothesized that the frequency feature of EEG signals would differ from individual to individual, and it was shown that the classification success was increased by 4.4% [17]. In this context, we also present the analysis of classification results obtained by using different segments of the EEG signal in this study.

The rest of this paper is organized as follows. Section 2 describes related work in the field of psychophysiological signal analysis. Section 3 introduces both the MEG and ECoG datasets and explains the signal preprocessing methods. Subsequently, the developed wavelet transform-based feature extraction and classification method is introduced. Section 4 presents the experimental results in comparison to the literature. Section 5 concludes the study by presenting improvements to be made to the proposed approach, as well as recommendations for future research.

2. Related works

Feature size in EEG signal analysis refers to the number of features extracted or selected from raw EEG data for subsequent analysis. Feature size directly affects the dimensionality of the data and can have an impact on computational complexity, interpretability, and the performance of machine learning algorithms. Feature extraction methods aim to transform raw EEG data into lower-dimensional representations by calculating specific features that capture relevant information. Commonly used features include spectral power in different frequency bands, statistical measures (e.g., mean, variance), wavelet transform (WT), and time-domain characteristics. Dimensionality reduction techniques aim to reduce the number of features while preserving important information. Techniques like principal component analysis (PCA) and LDA can be used to reduce dimensionality. Yu et al. analyzed the success of a motor imagery classification task in a 2014 study using PCA [18]. In another study in which epilepsy detection from EEG signals was performed, PCA, independent components analysis (ICA), and LDA were used to reduce the size of the data [19]. In the literature, there are other studies that have achieved successful classifications with the optimum number of features. For example, a very successful drowsiness detection system for drivers with a feature vector was obtained by Belakdar et al. with only five features from a single EEG channel [20]. In another study, spectral power-based effective feature extraction was proposed [21]. The feature vector size, or the number of features used in classification, plays a crucial role in determining the performance and efficiency of classification algorithms.

3. Materials and methods

The method proposed in this study was tested on two different datasets of psychophysiological signals, the details of which are presented below. All simulations were realized on a computer that had i7 2.8 GHz processors and 16 GB RAM using MATLAB 2022b.

3.1. Dataset

The presented method was tested on the two different datasets given in the following subsections.

3.1.1. MEG dataset

This dataset was obtained from two healthy subjects in total by experiments using 10 electrodes placed on the motor cortex of each subject. Each subject had to move his/her wrist in four different directions in the horizontal plane (left-right-forward-backward). All EEG signals were recorded at a sampling frequency of 625 13 Hz. Each trial was resampled at 400 Hz. Trials were cut to contain a total of 1-s data between 0.4 s before and 0.6 s after the movement. In the competition, participants were asked to classify the unlabeled evaluation data of 74 and 73 trials for Subject 1 and Subject 2 using labeled training data containing a total of 160 trials (40 trials per class) for each subject.

3.1.2. ECoG dataset

The ECoG dataset consisting of ECoG signals from BCI Competition III was recorded with a 1000-Hz sampling rate using 64 electrodes. In the experiment, the subject was asked to imagine the movement of her or his tongue or left little finger. Brain activity was picked up as a time series of 3 s during several trials using a platinum electrode grid (approximately 8 × 8 cm), which was located on the motor cortex invasively. The training dataset contains labeled data for the tongue (139) and finger (139) classes. Participants tried to predict correctly the labels of the test dataset with 100 trials [15, 22, 23].

3.2. Wavelet transform (WT)

WTs are mathematical analysis methods often used in mathematics and engineering to analyze data where signals vary at different scales. The continuous wavelet transform (CWT) provides continuous-time and continuous-frequency information in the time-frequency domain. It allows for precise localization of frequency components at each point in time. In contrast, the DWT provides discrete-time and discrete-frequency information. It achieves this by dividing the signal into successive time intervals and analyzing each interval at different scales. $W_{Coefficients}$ refers to the coefficients obtained from this calculation method. The first step of the proposed method is based on a WT that allows modeling the signal in the scaled time domain. The WT formulated as in Eq. (1) was used to reduce the size of the signal. The WT, which represents the signal with the different scale and position of the main wavelet used in the conversion, also preserves the time and frequency information of the signal. Here x represents n signal and ψ represents the main wavelet.

$$W_{Coefficients}(Scale, Pos) = \int_{-\infty}^{+\infty} S(t)\psi(Scale, Pos, t)dt \tag{1}$$

When the scales and shifts are chosen as powers of 2, the wavelet form is expressed as the DWT given in Eq. (2).

$$DWT(a,b) = \sum_{a}\sum_{b} x(b)2^{-a/2}\psi(2^{-a}n - b) \tag{2}$$

The DWT was calculated from 5th-order and 7th-order WTs using the Daubechies (db2) mother wavelet for the MEG and ECoG datasets, respectively. Table 1 and Table 2 show the signal decomposition levels and the frequency bands in each level for the MEG and ECoG datasets. The levels used in the classification are highlighted.

Table 1. DWT decomposition levels and frequency bands for the MEG dataset.

Level	**Frequency range (Hz)**	**EEG**
App_1	0–200	
App_2	0–100	
App_3	0–50	
App_4	0–25	
App_5	0–12.5	Delta (0–4)–Theta (4–8)
Det_1	200–400	
Det_2	100–200	
Det_3	50–100	Gamma (32–64)
Det_4	25–50	Beta (16–32)
Det_5	12.5–25	Alpha (8–16)

Table 2. DWT decomposition levels and frequency bands for the ECoG dataset.

Level	**Frequency range (Hz)**	**EEG**
App_1	0–500	
App_2	0–250	
App_3	0–125	
App_4	0–62.5	
App_5	0–31.5	
App_6	0–15.63	
App_7	0–7.81	Delta (0–4)–Theta (4–8)
Det_1	500–1000	
Det_2	250–500	
Det_3	125–250	
Det_4	62.5–125	
Det_5	31.25–62.5	Gamma (32–64)
Det_6	15.63–31.25	Beta (16–32)
Det_7	7.81–15.63	Alpha (8–16)

The obtained wavelet coefficients were used to calculate the correlation of the class averages with the wavelet outputs.

3.3. Pearson correlation coefficient

The Pearson correlation coefficient (PCC), which expresses the linear correlation between two variables such as A and B, is calculated as follows:

$$P(A,B) = \frac{COV(A,B)}{\sigma_A \sigma_B} \tag{3}$$

The function $COV(A,B)$ is the covariance of A and B. σ_A and σ_B are the deviations of A and B. $P(A,B)$ ranges from $+1$ to -1. A value of $+1$ implies that A is completely positively linearly correlated to B while a value of 0 indicates that A is not linearly correlated to B at all. A value of -1 implies that A is completely negatively linearly correlated to B. The correlation coefficient matrix of two random variables is the matrix of correlation coefficients given in Eq. (4) for each pairwise variable combination [24].

$$R = \begin{pmatrix} P(A,A) & P(A,B) \\ P(B,A) & P(B,B) \end{pmatrix} \tag{4}$$

Since A and B are always directly correlated to themselves, the diagonal entries are just 1, as follows:

$$R = \begin{pmatrix} 1 & P(A,B) \\ P(B,A) & 1 \end{pmatrix} \tag{5}$$

In this paper, the PCC is used to generate feature vectors by calculating the correlation of each trial with the average of the training classes separately.

3.4. Signal processing

We analyzed the average of the training data for both datasets to observe the pattern of each class in the time domain. The averages of the training data of each class that makes up the MEG and ECoG datasets are given in Figure 1 and Figure 2, respectively.

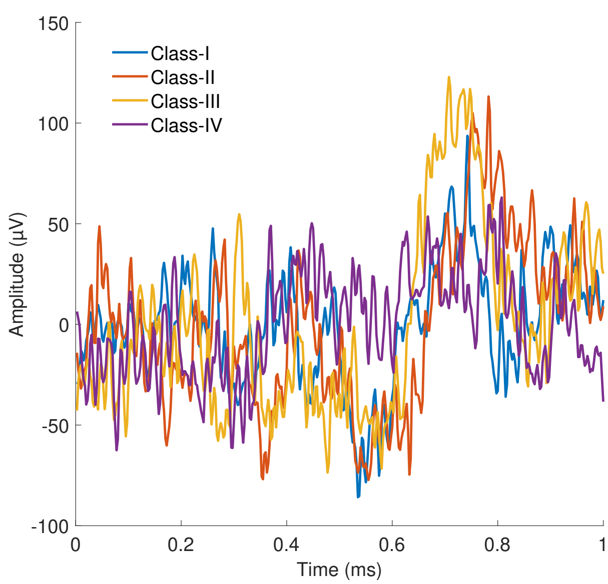

Figure 1. Average of training data for all classes (MEG dataset, Subject 1, Channel 3).

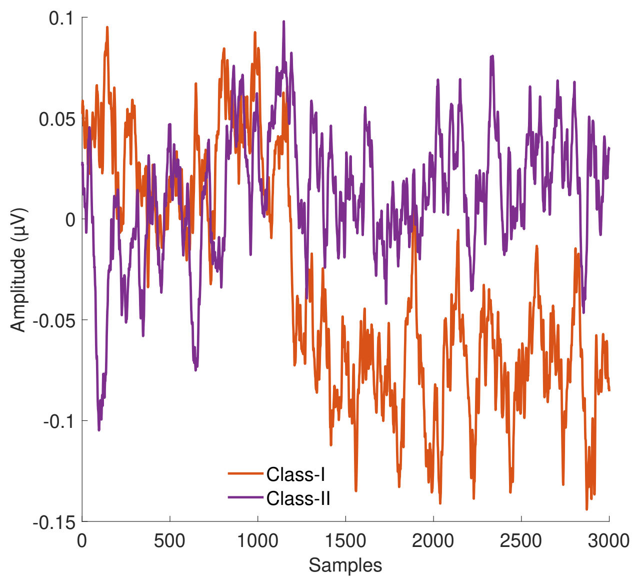

Figure 2. Average of training data (ECoG dataset, Channel 12).

Figure 1 and Figure 2 show the differences between the classes of the MEG and ECoG datasets. We applied the WT to class averages, both to reduce data size and to more clearly observe interclass differences. In Figure 3 and Figure 4, the class averages of the MEG and ECoG datasets and the approximate (App_n) components obtained as a result of the WT of these averages at different levels are given, respectively. It is seen that the wavelet approximation components of the 5th level for the MEG dataset and the 7th level for the ECoG dataset, which include the delta and theta frequencies, given in Figure 3f and Figure 4h, have a pattern that reflects the general characteristic of the signals in the time domain. Starting from this point, 5th-level WT was applied to the MEG dataset and 7th-level WT was applied to the ECoG dataset. In addition, there are studies in the literature in which alpha and theta bands of psychophysiological signals such as EEG are used [29]. For MEG dataset-Subject 1, all class means and WT approximation coefficient 5 outputs of these averages are given in Figure 5, and for the ECoG dataset, all class averages and WT approximation coefficient 7 outputs of these averages are given in Figure 6.

From Figure 5 and Figure 6, the differences between classes are clearly seen with both the class averages and the WT outputs of the class averages.

3.5. Feature extraction

As mentioned before, it was seen in the tests performed that wavelet approximation components containing the delta and theta frequencies of the EEG signal can exhibit better classification results. Therefore, for both datasets, feature vectors were created by calculating the PCCs, which determine the correlation of wavelet components that include delta and theta frequencies with each class average. The flow chart of the proposed feature extraction method is given in Figure 7. As seen in Figure 7, the correlation of the WT of each trial was scaled in the range of [–1,+1] and the WT of each class mean was calculated. In this way, a feature vector was created by calculating 1 PCC per class for each trial signal. In the training phase, the correlation of each trial in the training data with each training class average was calculated, and in the evolution phase, the feature vector

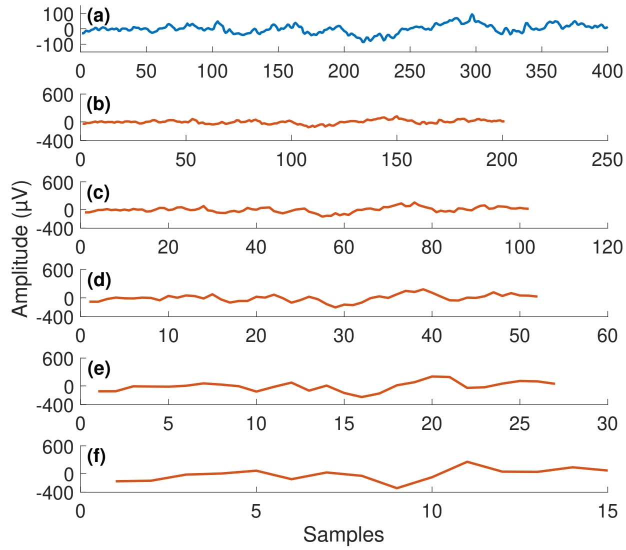

Figure 3. WT of the average training data of Class-I for Subject 1 of the MEG dataset (channel 3: (a) Class-I average; (b) App_1, Level 1 approximation coefficients of average of Class-I; (c) App_2, Level 2 approximation coefficients of average of Class-I; (d) App_3, Level 3 approximation coefficients of average of Class-I; (e) App_4, Level 4 approximation coefficients of average of Class-I; (f) App_5, Level 5 approximation coefficients of average of Class-I).

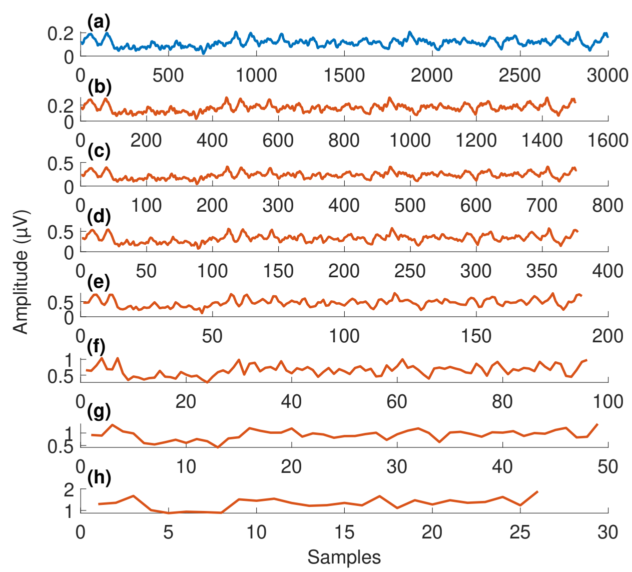

Figure 4. WT of the average training data of Class-I for the ECoG dataset (channel 38: (a) Class-I average; (b) App_1, Level 1 approximation coefficients of average of Class-I; (c) App_2, Level 2 approximation coefficients of average of Class-I; (d) App_3, Level 3 approximation coefficients of average of Class-I; (e) App_4, Level 4 approximation coefficients of average of Class-I; (f) App_5, Level 5 approximation coefficients of average of Class-I; (g) App_6, Level 6 approximation coefficients of average of Class-I; (h) App_7, Level 7 approximation coefficients of average of Class-I).

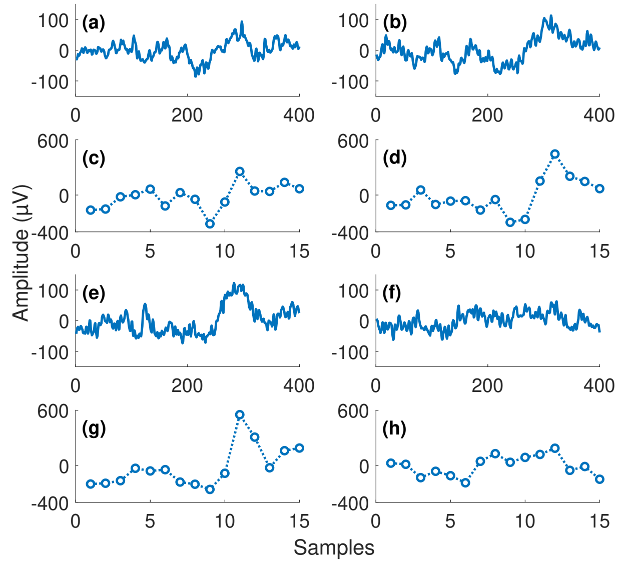

Figure 5. WT of the average training data of Subject 1 for each class in the MEG dataset for channel 7: (a) Class-I, (b) Class-II, (c) Class-I level 5 WT approximation coefficients, (d) Class-II level 5 WT approximation coefficients, (e) Class-III, (f) Class-IV, (g) Class-III level 5 WT approximation coefficients, (h) Class-IV level 5 WT approximation coefficients.

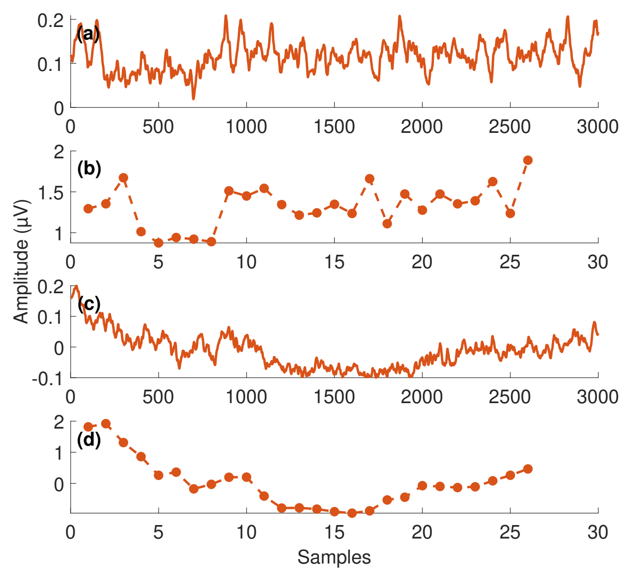

Figure 6. WT of the average training data for each class in the ECoG dataset for channel 38: (a) Class-I, (b) Class-II, (c) Class-I level 7 WT approximation coefficients, (d) Class-II level 7 WT approximation coefficients.

was produced by calculating the correlation of each trial in the evolution data with each training class average. In summary, the similarity of each trial with the training class averages constitutes feature vectors. In order to obtain the wavelet components containing the delta and theta frequencies of the signal, k (in Figure 7), which represents the WT level, was chosen as 5 and 7, respectively, for the MEG and ECoG datasets. Feature vectors were calculated with wavelet approximation components that only included delta and theta frequencies.

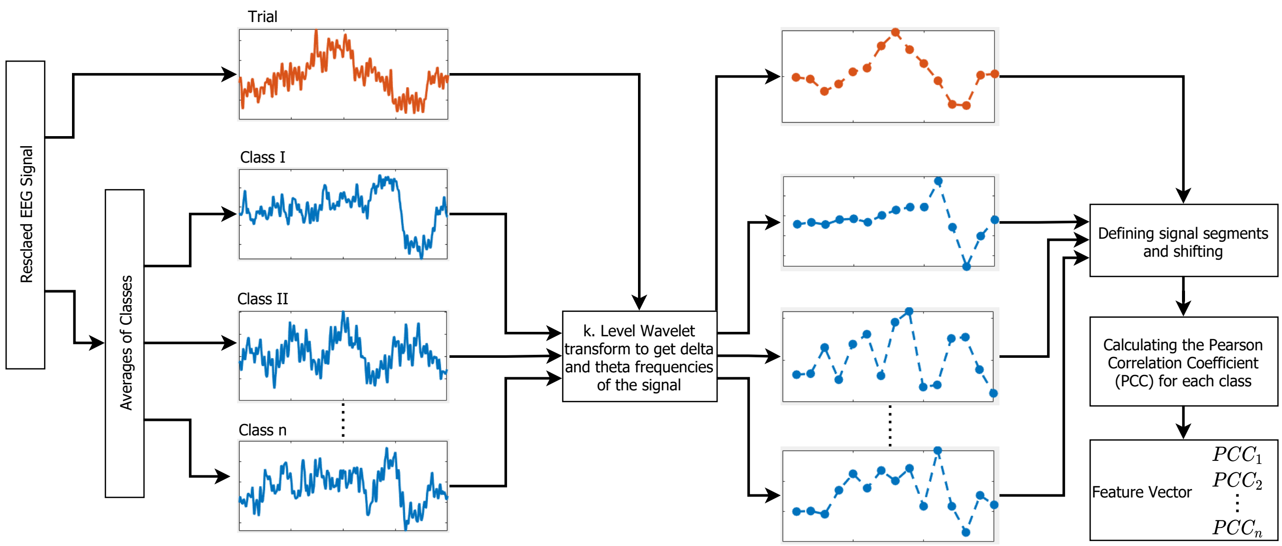

Figure 7. Generating the feature vector.

3.5.1. Analysis of time complexity

The pseudo-code of the proposed feature extraction approach that we suggest is also given in Algorithm 1. The time complexity of the proposed approach can be given as approximately O(kst), with s representing the number of classes, c the number of channels, and t the number of training data trials. The MEG dataset's training and test practice computation times were determined as 3.88 s and 2.02 s, respectively. In addition, we would like to note that the feature vector for space complexity changes in direct proportion to the number of classes.

4. Classifier

4.1. Support vector machine (SVM)

$x(i)\epsilon R^{dim}$ being the training data, $y_i\epsilon\{-1,+1\}$ and dim represent the class labels and dimension of the hyperplane, respectively, as follows:

$$(x_1, y_1), (x_2, y_2), \ldots (x_k, y_k) \tag{6}$$

The purpose of the SVM is to determine the optimal hyperplane ($w * x + b = 0$) that divides the training data into two classes, maximizing the margin. $F(t)$ determines which class t belongs to in $\{-1, +1\}$ with optimal values of w and b expressed by Eq. (7) [5, 25].

$$F(t) = sgn((\widehat{w}, t) + \widehat{b}) \tag{7}$$

Algorithm 1 Feature extraction method.

Input:
signal: EEG Signal(s,c,t);
s: Number of Classes;
c: Number of Channels;
t: Number of Trials;

Output:
FV: Feature Vector (s,c,t)
s: Number of Classes;
c: Number of Channels;
t: Number of Trials;

Description:

```
Define parameters: s, c, t
for ((s = 1; s < class_number; s++) do
    for ((c = 1; c < channel_number; c++) do
        class_average(s, c) = mean(signal(s, c))
        Wavelet_coeff_class(s, c) = WT(class_average(s, c))
    end for
end for
for ((c = 1; c < channel_number; c++) do
    for ((t = 1; s < trial_number; t++) do
        Wavelet_coeff_trial(s, c, t) = WT(signal(s, c, t))
        for ((s = 1; s < class_number; s++) do
            FV(c, t, s) = corrcoef(Wavelet_coeff_class(s, c), Wavelet_coeff_trial(s, c, t))
        end for
    end for
end for
```

SVM algorithms use mathematical functions such as linear, nonlinear, and polynomial, which are defined as the kernel. These functions transform input data to the required form. The kernel function of the SVM is chosen as linear in this study.

4.2. Evaluation parameters

Evaluation metrics with formulas given below were used in this study, including overall accuracy (Acc), recall (Re), precision (Pr), F1 score (F1), and Cohen's kappa coefficient (K). The accuracy of the multiclass problem was evaluated as follows:

$$Acc = \frac{T_P + T_N}{T_P + T_N + F_P + F_N} \tag{8}$$

Here, T_P, F_P, T_N, and F_N denote the true positives, the false positives, the true negatives, and the false negatives, respectively. Accuracy (Acc) represents the ratio of the correctly classified instances to the total number of instances [26]. The other performance metrics were calculated as follows:

$$Re = \frac{T_P}{T_P + F_N} \tag{9}$$

$$Pr = \frac{T_P}{T_P + F_P} \tag{10}$$

$$F1 = \frac{2 \times Pr \times Re}{Pr + Rr} \tag{11}$$

Cohen's kappa coefficient is usually utilized for consistency testing and measurement of classification accuracy.

$$k = \frac{p_o - p_e}{1 - p_e} \tag{12}$$

In Eq. (12), p_o represents the sum of the number of samples that are correctly classified within each category. This metric provides the overall classification accuracy obtained by dividing the total number of samples. p_e given in Eq. (12) is calculated by Eq. (13). $a_1, a_2, ..., a_c$ and $b_1, b_2, ..., b_C$ express the predicted number of samples in each category and the total number of samples (N), respectively.

$$p_e = \frac{a_1 x b_1 + a_2 x b_2 + ... + a_c x b_c}{N x N} \tag{13}$$

Kappa is a consistency measure that can be between 0 and 1. Values of kappa close to 1 indicate high consistency and accuracy [27].

5. Experimental results

In the time domain, trials of the MEG and ECoG datasets, expressed as 400 and 1000 points, respectively, were represented by 15 and 26 points as a result of WT. The most successful classification results, dataset, signal segment, channel subset, and classification accuracy with metrics supporting that accuracy such as kappa, precision, recall, and F1 score are given in Table 3.

Table 3. Best classification results with PCC feature vectors.

Dataset	Subject	Signal segment	Channel subset	Eval. Acc. %	Eval. Kappa	Eval. Precision	Eval. Recall	Eval. F1 score
MEG	Subject 1	1–15	2,3,4,5,6	68.92	0.58	0.69	0.69	0.68
MEG	Subject 1	1–15	1,2,3,4,5,6	68.92	0.58	0.72	0.69	0.68
MEG	Subject 2	1–15	1,4,8,9	45.21	0.26	0.46	0.45	0.45
MEG	Subject 2	1–15	6,8,9	45.21	0.24	0.46	0.43	0.43
ECoG	Subject	1–26	4,11,22	81	0.62	0.81	0.81	0.81
ECoG	Subject	1–26	4,7,12,31	82	0.64	0.82	0.82	0.82
ECoG	Subject	1–26	4,12,14,22,37	84	0.68	0.85	0.84	0.84

When the results given in Table 3 are evaluated, it is seen that quite satisfactory classification accuracies were obtained with the datasets used compared to previous studies. These results are 68.92%, 45.21%, and 84% for MEG dataset-Subject 1, MEG dataset-Subject 2, and the ECoG dataset, respectively. More importantly, these results were obtained with feature vectors of much smaller size than those previously used in studies with these datasets. The mentioned classification results were achieved with vectors consisting of only 1 feature per class. For example, while the feature vector size for the 4-class MEG dataset is channel $\times$ 4, the feature vector size for the 2-class ECoG dataset is channel $\times$ 2. The results are supported by four different evaluation metrics including kappa, precision, recall, and F1 score. In addition, distinctive features for classes may be hidden in different temporal segments of the signal. In a previous study it was shown that the features obtained by shifting the parts of the signal in the time domain increased the classification success [28]. With this approach, signal segments with all possible starting points and lengths on the approximation component of the WT of the signal are selected, and feature vectors created by calculating similarities of these segments with class averages are classified according to all possible channel combinations. In this way, the most effective part of the signal is tried to be determined.

We present the classification results of the MEG dataset obtained from the coefficients of the WT at different levels (A4, A6) in Table 4. When Table 4 is compared to Table 3, it is seen that more successful classification results were obtained from WT components corresponding to the delta and theta bands.

Table 4. Best classification results with PCC feature vectors using A4 and A6 coefficients of WT.

Dataset	Subject	Signal segment	Channel subset	Eval. Acc. %	Eval. Kappa	Eval. Precision	Eval. Recall	Eval. F1 score
A4	Subject 1	1–27	2,3,4,5,6	59.46	0.44	0.59	0.59	0.59
A4	Subject 1	1–27	1,2,3,4,5,6	58.11	0.44	0.61	0.58	0.58
A4	Subject 2	1–27	1,4,8,9	27.40	0.34	0.31	0.27	0.28
A4	Subject 2	1–27	6,8,9	38.36	0.18	0.39	0.38	0.38
A6	Subject 1	1–9	2,3,4,5,6	54.05	0.38	0.55	0.54	0.54
A6	Subject 1	1–9	1,2,3,4,5,6	50.00	0.33	0.52	0.50	0.50
A6	Subject 2	1–9	1,4,8,9	35.62	0.12	0.37	0.36	0.36
A6	Subject 2	1–9	6,8,9	35.62	0.14	0.38	0.36	0.36

The classification accuracies of Subject 1 and Subject 2 for all possible signal segment lengths and origins for the MEG evaluation dataset are given in Figure 8 and Figure 9, respectively. The x-axis represents the origin of the signal segment and the y-axis its length.

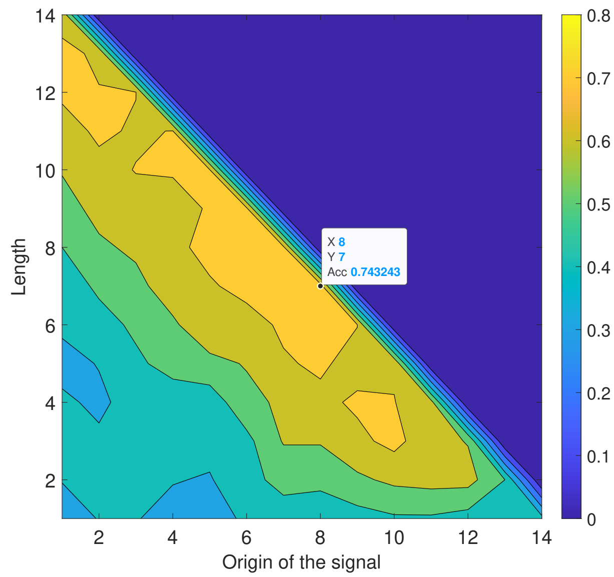

Figure 8. Classification successes according to different signal lengths and starting points for Subject 1 on evaluation dataset.

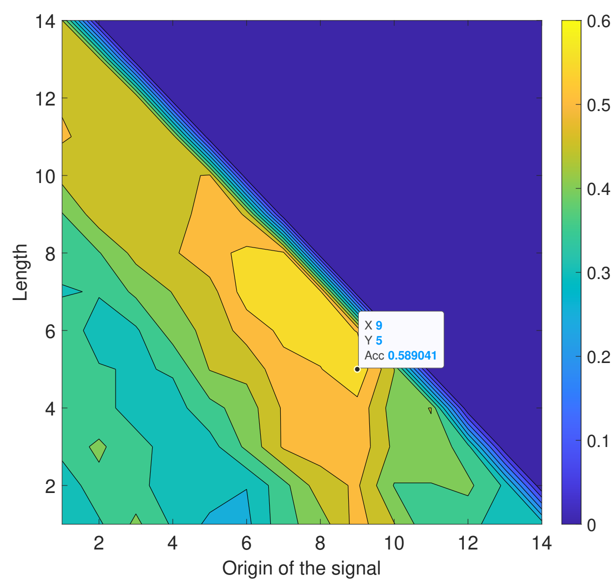

Figure 9. Classification successes according to different signal lengths and starting points for Subject 2 on evaluation dataset.

As seen from Figure 8, the highest classification accuracy for Subject 1, 74.32%, was obtained by using the part of the signal containing only points 8–15 of the 5th-level WT. Similarly, as can be seen in Figure 9, the highest classification accuracy for Subject 2 was obtained as 58.90% by using the part of the signal containing only points 9–14 of the 5th-level WT. It can be said that these signal segments are the signal segments that carry the most information for each subject. In Table 5, these classification results for Subject 1 and Subject 2 are given in more detail. The best classification results are highlighted. The dataset, signal segment, channel subset, and classification accuracy with metrics supporting this accuracy including kappa, precision, recall, and F1 score are presented in Table 5.

Table 5. Best classification results with PCC feature vectors.

Dataset	**Subject**	**Signal segment**	**Channel subset**	**Eval. Acc. %**	**Eval. Kappa**	**Eval. Precision**	**Eval. Recall**	**Eval. F1 score**
MEG	Subjec t1	8–15	1,3,4,7,8,9	74.32	0.65	0.76	0.74	0.74
MEG	Subject 1	6–14	2,3,4,5,9,10	72.97	0.63	0.76	0.73	0.73
MEG	Subject 1	5–13	2,4,5,7	72.97	0.63	0.74	0.73	0.73
MEG	Subject 2	9–14	1,2,5,6,7,8,10	58.90	0.44	0.59	0.59	0.59
MEG	Subject 2	7–14	6,7,8,9,10	57.53	0.42	0.59	0.58	0.57
MEG	Subject 2	6–13	2,3,5,6,7,8,9	57.34	0.43	0.62	0.58	0.57

Similarly, the classification successes of the training datasets are given in Figure 10 and Figure 11 for Subject 1 and Subject 2, respectively. It is seen that the signal segments from which the most successful classification results were obtained for the evaluation and training datasets given in Figure 8 and Figure 10 for Subject 1 are highly similar. In addition, for Subject 2, the signal segments from which the most successful classification results for the evaluation and training datasets in Figure 9 and Figure 11 were achieved appear to be highly similar. We predict that this similarity rate will increase in direct proportion to the number of trials in the dataset.

When the signal segments from which the maximum classification result was obtained are examined, it is seen that different signal segments are effective in classification for all three subjects. This result is evidence that motor imagery movements are encoded in a subject-specific way in the brain, unlike studies in which the same feature vectors are used for each subject. Table 6 summarizes the comparison of the results of previous studies and our study. NF represents the total number of features used in classification. As seen from Table 6, the highest classification success was achieved for both subjects when we compared the proposed practical method with those of the literature, except for the last study [14]. Moreover, this work also presented the highest classification successes achieved with the lowest dimensional feature vectors for both subjects. The results of our study in Table 6 were obtained with only 24 and 28 numerical features in total for Subject 1 and Subject 2, respectively. When we evaluate the findings on the basis of the subjects, it is also seen that the results achieved are in harmony with the previous studies, except for the last study [14] in the literature with this dataset. In the mentioned study, it is seen that the obtained results were achieved by transferring the signal in the time domain to the image space and selecting the image properties in an ambiguous manner. Our results are also supported by four different evaluation metrics as given in Table 5. Attention should be paid to the NF column, where the number of features used in classification is given.

ROC curves for the best classification results obtained with the MEG dataset for Subject 1 and Subject 2 are given in Figure 12.

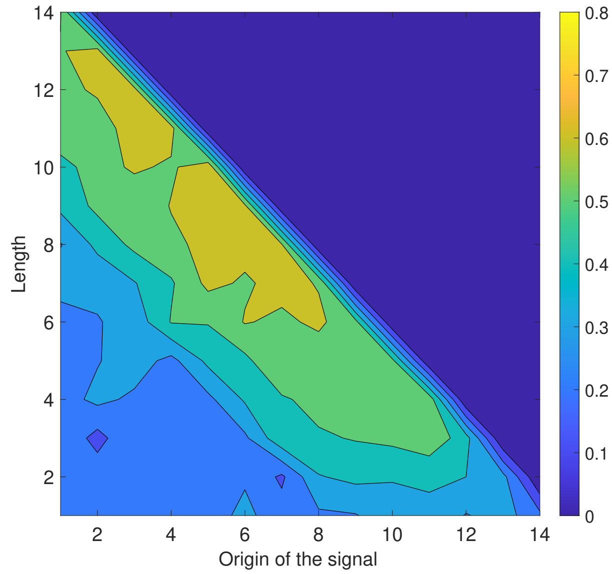

Figure 10. Classification successes according to different signal lengths and starting points for Subject 1 on training dataset.

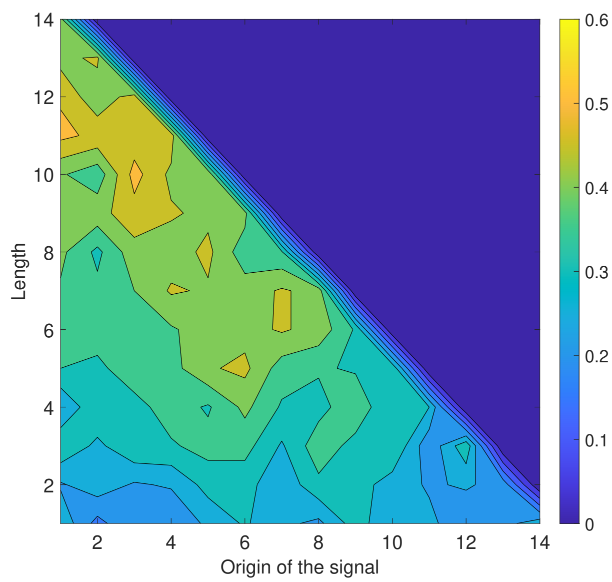

Figure 11. Classification successes according to different signal lengths and starting points for Subject 2 on training dataset.

6. Conclusion

In this work, a new and practical feature extraction method based on WT and PCC has been analyzed. With this method, which can produce low-dimensional feature vectors by WT, the most effective parts of the signal could be detected. With the proposed method, WT is first applied to the signal to use the approximation

Table 6. Comparison of the results of previous studies and our study.

Studies	**NF**	**Subject 1 Acc. %**	**Subject 1 Kappa**	**Subject 2 Acc. %**	**Subject 2 Kappa**	**Avr. Acc. %**
Sardouie et al.	4788	59.50	-	34.30	-	46.9
Li et al.	1024	31.10	-	19.20	-	25.1
Montazeri et al.	-	16.20	-	31.50	-	23.9
Wang et al.	1024	23.00	-	17.80	-	20.4
Goni et al.	270	64.00	-	-	-	-
Hatipoğlu et al.	500	78.09	-	79.88	-	78.99
Our study	24-28	74.32	0.74	58.90	0.59	66.61

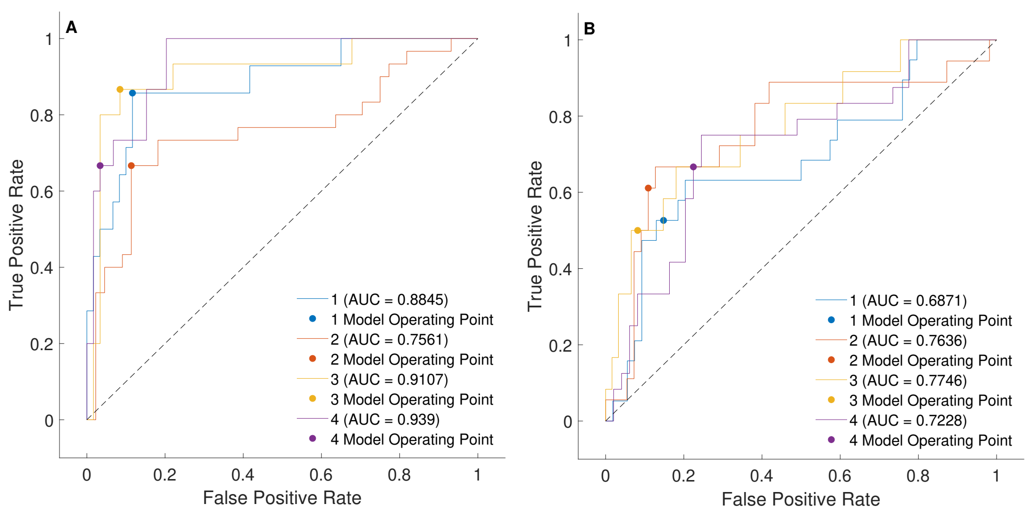

Figure 12. ROC curves for MEG dataset: A) Subject 1, B) Subject 2.

coefficients that have the overall pattern of the signal in the time domain. Thus, the size of the signal to be processed is reduced. Secondly, the segment of the signal that best represents the general character of the signal is determined by calculating all possible probabilities. Finally, the feature vector is created by calculating the correlation of this signal part with the corresponding signal part of the WT output of each class average. In this way, feature vectors created by calculating only one feature per class are classified by the SVM. The proposed analysis method was tested on two different datasets with three subjects. The best classification accuracies of 74.32% and 58.90% for the MEG dataset with Subject 1 and Subject 2 were respectively achieved. The classification successes for both subjects were supported by the highest evaluation metrics in the literature of 0.74 and 0.59, which represent high and medium classification harmony. The proposed approach produces low-dimensional feature vectors. These feature vectors are obtained with just one PCC calculated from each signal segment for a class. For example, classification success accuracy of 86% with kappa of 0.72 for the ECoG dataset was obtained by calculating only one PCC value of the signal per class. However, the best classification results of Subject 1 and Subject 2 for the MEG dataset were obtained with different channel subsets. This may be evidence that the regions in the brain where motor imagery tasks are encoded may differ from subject to subject. In addition, the presented study has also shown that the alpha and theta signal bands contain distinctive cognitive information in terms of classification [29, 31]. In this paper, it has been shown that there is a subject-specific signal pattern formed as a result of motor imagery tasks in the brain and this pattern can be successfully classified with a low-dimensional feature vector. Our study, in contrast to complex feature extraction methods, encourages more studies on the mentioned pattern. As future work, the suggested feature extraction approach can be improved for automatically detecting the subject-specific signal segments.

Acknowledgment

The authors thank the editor and anonymous reviewers for their valuable comments and suggestions.

References

[1] Lu N, Li T, Ren X, Miao H. A deep learning scheme for motor imagery classification based on restricted Boltzmann machines. IEEE Transactions on Neural Systems and Rehabilitation Engineering 2017; 2: 566-576. https://doi.org/10.1109/TNSRE.2016.2601240

[2] Talukdar U, Hazarika S, Gan J. Adaptation of common spatial patterns based on mental fatigue for motor-imagery BCI. Biomedical Signal Processing and Control 2020; 58: 101829. https://doi.org/10.1016/j.bspc.2019.101829

[3] Savitzky A, Golay M. Smoothing and differentiation of data by simplified least squares procedures. Analytical Chemistry 1964; 36: 1627-1639. https://doi.org/10.1021/ac60214a047

[4] Rahman M, Fattah S. An efficient feature extraction scheme for classification of mental tasks based on inter-channel correlation in wavelet domain utilizing EEG signal. Biomedical Signal Processing and Control 2020; 61: 102033. https://doi.org/10.1016/j.bspc.2020.102033

[5] Anuragi A, Sisodia D. Empirical wavelet transform based automated alcoholism detecting using EEG signal features. Biomedical Signal Processing and Control 2019; 57: 101777. https://doi.org/10.1016/j.bspc.2019.101777

[6] Choubey H, Pandey A. A combination of statistical parameters for the detection of epilepsy and EEG classification using ANN and KNN classifier. SIViP 2021; 15: 1-9. https://doi.org/10.1007/s11760-020-01767-4

[7] Das K, Pachori R. Schizophrenia detection technique using multivariate iterative filtering and multichannel EEG signals. Biomedical Signal Processing and Control 2021; 67: 102525. https://doi.org/10.1016/j.bspc.2021.102525

[8] Sakhavi S, Guan C, Yan S. Learning temporal information for brain-computer interface using convolutional neural networks. IEEE Transactions on Neural Networks and Learning Systems 2018; 29: 5619-5629. https://doi.org/10.1109/TNNLS.2018.2789927

[9] Zhou M, Tian C, Cao R, Wang B, Niu Y et al. Epileptic seizure detection based on EEG signals and CNN. Frontiers in Neuroinformatics 2018; 12: 95. https://doi.org/10.3389/fninf.2018.00095

[10] Türk Ö. Classification of electroencephalogram records related to cursor movements with a hybrid method based on deep learning. International Journal of Imaging Systems and Technology 2021; 31 (4): 2322-2333. https://doi.org/10.1002/ima.22643

[11] Sardouie S, Shamsollahi M. Selection of efficient features for discrimination of hand movements from MEG using a BCI competition IV data set. Frontiers in Neuroscience 2012; 6: 4. https://doi.org/10.3389/fnins.2012.00042

[12] Tangermann M, Müller K, Aertsen A, Birbaumer N, Brunner R et al. Review of the BCI Competition IV. Frontiers in Neuroscience 2012; 6: 55. https://doi.org/10.3389/fnins.2012.00055

[13] Goni M, Rahman T. Predictive modeling on MEG signal to classify hand and wrist movement using UNEQ and KNN. In: IEEE Region 10 Symposium (TENSYMP); Dhaka, Bangladesh; 2020. pp. 815-818. https://doi.org/10.1109/TENSYMP50017.2020.9230860

[14] Hatipoglu B, Yilmaz C, Kose C. A signal-to-image transformation approach for EEG and MEG signal classification. Signal, Image and Video Processing 2019; 13: 483-490. https://doi.org/10.1007/s11760-018-1373-y

[15] Lal T, Hinterberger G, Widman T, Schroder G, Hill J et al. Methods towards invasive human brain computer interfaces. In: Advances in Neural Information Processing Systems 17; Vancouver, BC, Canada; 2004.

[16] Qi Y, Ding F, Xu F, Yang J. Channel and feature selection for a motor imagery-based BCI system using multilevel particle swarm optimization. Computational Intelligence and Neuroscience 2020; 2020: 8890477. https://doi.org/10.1155/2020/8890477

[17] Oh SH, Lee YR, Kim HN. A novel EEG feature extraction method using Hjorth parameter. International Journal of Electronics and Electrical Engineering 2014; 2 (2): 106-110. https://doi: 10.12720/ijeee.2.2.106-110

[18] Yu X, Chum P, Sim KB. Analysis the effect of PCA for feature reduction in non-stationary EEG based motor imagery of BCI system. Optik 2014; 125 (3): 1498-1502. https://doi.org/10.1016/j.ijleo.2013.09.013

[19] Subasi A, Gursoy MI. EEG signal classification using PCA, ICA, LDA and support vector machines. Expert Systems with Applications 2010; 37 (12): 8659-8666. https://doi.org/10.1016/j.eswa.2010.06.065

[20] Belakhdar I, Kaaniche W, Djemal R, Ouni B. Single-channel-based automatic drowsiness detection architecture with a reduced number of EEG features. Microprocessors and Microsystems 2018; 58: 13-23. https://doi.org/10.1016/j.micpro.2018.02.004

[21] Kim C, Sun J, Liu D, Wang Q, Paek S. An effective feature extraction method by power spectral density of EEG signal for 2-class motor imagery-based BCI. Medical & Biological Engineering & Computing 2018; 56: 1645-1658. https://doi.org/10.1007/s11517-017-1761-4

[22] Alotaiby T, El-Samie F, Alshebeili S, Ahmad I. A review of channel selection algorithms for EEG signal processing. EURASIP Journal of Advanced Signal Processing 2015; 2015: 66. https://doi.org/10.1186/s13634-015-0251-9

[23] Erkan E. Kurnaz I. A study on the effect of psychophysiological signal features on classification methods. Measurements in Biology and Medicine 2017; 101: 45-52. https://doi.org/10.1016/j.measurement.2017.01.019

[24] Fisher R. Statistical methods for research workers. In: Kotz S, Johnson NL (editors). Breakthroughs in Statistics. Berlin, Germany: Springer, 1992, pp. 66-70. https://doi.org/10.1007/978-1-4612-4380-9

[25] Schüldt C, Laptev I, Caputo B. Recognizing human actions: a local SVM approach. In: Proceedings of the 17th International Conference on Pattern Recognition; Cambridge, UK; 2004. pp. 32-36. https://doi.org/10.1016/j.bspc.2019.101777

[26] Azar A, El-Said S. Performance analysis of support vector machines classifier in breast cancer mammography recognition. Neural Computing and Applications 2014; 24 (5): 1163-1177. https://doi.org/10.1007/S00521-012-1324-4

[27] Zhao X, Wang X, Yang T, Ji S, Wang H et al. Classification of sleep apnea based on EEG sub-band signal characteristics. Scientific Reports 2021; 11 (1): 5824. https://doi.org/10.1038/s41598-021-85138-0

[28] Hersche M, Rellstab T, Schiavone P, Cavigelli L, Benini L et al. Fast and accurate multiclass inference for MI-BCIs using large multiscale temporal and spectral features. In: 26th European Signal Processing Conference; Rome, Italy; 2018. pp. 1690-1694. https://doi.org/10.23919/EUSIPCO.2018.8553378

[29] Klimesch W. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Research Reviews 1999; 29 (2): 169-195. https://doi.org/10.1016/S0165-0173(98)00056-3

[30] Rossi C, Vidaurre D, Costers L, Akbarian F, Woolrich M et al. A novel description of the network dynamics underpinning working memory. BioRxiv Preprint, 2023. BioRxiv: 2023-01. https://doi.org/10.1101/2023.01.20.524895

[31] Murphy AC, Bertolero MA, Papadopoulos L, Lydon-Staley DM, Bassett DS. Multimodal network dynamics underpinning working memory. Nature Communications 2020; 11: 3035. https://doi.org/10.1038/s41467-020-15541-0