Biomedical Signal Processing and Control 31 (2017) 398–406

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

Comparison of signal decomposition methods in classification of EEG signals for motor-imagery BCI system

Jasmin Kevric[a], Abdulhamit Subasi[b,*]

[a] International Burch University, Faculty of Engineering and Information Technologies, Francuske Revolucije bb. Ilidza, Sarajevo, 71000, Bosnia and Herzegovina
[b] Effat University, College of Engineering, Department of Computer Science, Jeddah, 21478, Saudi Arabia

ARTICLE INFO

Article history:
Received 5 December 2015
Received in revised form 9 August 2016
Accepted 7 September 2016
Available online 19 September 2016

Keywords:
Empirical mode decomposition (EMD)
Discrete wavelet transform (DWT)
Wavelet packet decomposition (WPD)
Motor imagery (MI)
Brain computer interface (BCI)
Higher order statistics (HOS)
BCI competition III dataset IVa

ABSTRACT

In this study, three popular signal processing techniques (Empirical Mode Decomposition, Discrete Wavelet Transform, and Wavelet Packet Decomposition) were investigated for the decomposition of Electroencephalography (EEG) Signals in Brain Computer Interface (BCI) system for a classification task. Publicly available BCI competition III dataset IVa, a multichannel 2-class motor-imagery dataset, was used for this purpose. Multiscale Principal Component Analysis method was applied for the purpose of noise removal. In addition, different sets of features were formed to examine the effect of a particular group of features. The parameter selection process for signal decomposition methods was thoroughly explained as well. Our results show that the combination of Multiscale Principal Component Analysis de-noising and higher order statistics features extracted from wavelet packet decomposition sub-bands resulted in highest average classification accuracy of 92.8%. Our study is one among very few that provides a comprehensive comparison between signal decomposition methods in combination with higher order statistics in classification of BCI signals. In addition, we stressed the importance of higher frequency ranges in improving the classification task for EEG signals in Brain Computer Interface Systems. Obtained results indicate that the proposed model has the potential to obtain a reliable classification of motor imagery EEG signals, and can thus be used as a practical system for controlling a wheelchair. It can also further enhance the current rehabilitation therapies where appropriate feedback is delivered once the individual executes the correct movement. In that way, motor rehabilitation outcomes may improve over time.

1. Introduction

Brain-Computer Interface (BCI) represents a communication and control mechanism between the human brain and computers envisioned to support disabled people using their electrical activity of the brain, which is usually recorded using electroencephalogram (EEG). There are two types of BCI: spontaneous EEG, and evoked EEG based BCI systems. The former is generated as a result of specific mental activity, whereas the latter is a result of some kind of neural stimulation. A number of paradigms have recently been assessed to test the possibility of detecting few mental tasks from the EEG signals [1]. The aim is to generate distinct EEG signals by imagining certain motor activities, which are then translated into external actions [2].

In addition to having non-stationary nature, EEG signals are also susceptible to various other factors such as physical state, mood, posture, external noise, etc. Processing of EEG signals, which directly affects the classification accuracy, still represents a crucial challenge. Signal processing techniques for the motor imagery EEG mainly belong to some of the following:

- Fourier transform (FT), which can be used to calculate power spectrum (or power spectral density), whose main disadvantage is the lack of time domain information [3].
- Autoregressive (AR) model, which can be used to calculate the AR spectrum or the AR model coefficients [3]. This method is computationally efficient (being appropriate for online systems), but is also sensitive to the artifact [4].
- Common spatial patterns (CSP), a suitable spatial filtering technique for oscillatory EEG components during motor imagery aiming at multi-channel EEG data [5], and its many variations.

* Corresponding author.
E-mail address: absubasi@effatuniversity.edu.sa (A. Subasi).

http://dx.doi.org/10.1016/j.bspc.2016.09.007
1746-8094/

- Coefficients of discrete wavelet transform (DWT) at useful frequency bands. In addition, wavelet packet decomposition (WPD) algorithm decomposes the signal into the low-frequency and high-frequency components and then the coefficients at a certain band can be extracted [6].

Gajic et al. [7] described an automated classification of EEG signals for the detection of epileptic seizures by quadratic classifiers designed in the reduced two-dimensional feature space. The original feature set consisted of statistical features in time, frequency, time-frequency domain non-linear features extracted from a few frequency sub-bands of clinical interest. The overall classification accuracy was around 99% [8].

Considering noisy signals like EEG, linear de-noising methods considerably smooth out the rapid jumps in the signal. On the other hand, the nonlinear filtering techniques like Multiscale Principal Component Analysis (MSPCA) may eliminate the noise without sacrificing substantial amount of fast changes in the signal [9]. MSPCA de-noising has been successfully applied in classifying EEG signals for epileptic seizure detection in [10]. Apart from EEG, MSPCA has been exploited for diagnosis of neuromuscular disorders using Electromyography (EMG) signals [11]. In addition, MSPCA showed its strength for diagnosis of cardiovascular diseases by classifying ECG beats [12]. The effect of MSPCA in real-time wireless BCI system was also investigated in [13]. The study involved practical motor imagery BCI system based on Mindwave Mobile Headset from one subject including two motor imagery tasks: right hand and left hand. The model combining MSPCA de-noising and statistical wavelet features showed promising results [13].

One important part of signal processing is the extraction of distinguishable features from available EEG signals, because these techniques need to adjust to dynamic characteristics of EEG signals. In essence, EEG signals are non-stationary, non-linear, and non-Gaussian. The first and second order statistics (mean and the standard deviation), despite being very limiting in analysis of signal non-linearity, have attracted large attention in biomedical signal processing [14]. Moreover, different variations of wavelet transforms and decompositions proved to be suitable because of their ability to derive dynamical features from the signals and enhanced resolution. However, even the wavelets suffer from deriving nonlinear relationships within the signal 15]. On the other hand, higher order statistics (HOS) can discover irregularities from stationarity, linearity, or Gaussianity in the signal which is a clear advantage over traditional time and/or frequency domain methods in biomedical signals analysis. In addition, HOS features can be made insensitive to shift or amplification [14]. Therefore, HOS can be used as an effective technique for the analysis of non-linear nature of weak and noisy biomedical signals like EEG.

The review of signal processing and classification methods is not very appropriate unless they are applied on the same open database. Therefore, we provide a short review of the most notable recent studies done on Dataset IVa from BCI Competition III. In [16], a novel algorithm called iterative spatio-spectral patterns learning (ISSPL) is presented. It accomplishes automatic learning of spatio-spectral filters using statistical learning theory to accomplish good generalization performance. The authors claim that the suggested method accurately determines the significant frequency bands, making it superior to other classification paradigms [16]. Modified cross-correlation based logistic regression (CC-LR) approach tries to discover a feature set to appropriately describe the distribution of MI tasks from EEG data. The study shows that the proposed method has potential to enhance the classification performance of MI tasks in BCI systems [17]. In [18], a new heuristic method for the optimal channels selection for CSP is presented. The CSP is executed to training set prior to assigning a channel score to each channel based on l_1 norm. At the end, channels having higher scores are reserved for further CSP processing to extract features. The study indicates that this technique can successfully execute the job of selecting the optimal channels [18]. Lu et al. [19] suggested a regularization and aggregation method for CSP in a small sample setting. Since Conventional CSP performance in EEG classification worsens for small number of samples, authors suggested a regularized CSP (R-CSP) technique in which they regularize the covariance-matrix to lower the estimation bias and variance. In addition, the study suggests R-CSP with aggregation (R-CSP-A), in which a few R-CSPs are aggregated providing ensemble-based solution. The results indicate that R-CSP-A considerably beats the other methods regarding overall classification accuracy [19]. Li and Lu [20] enhanced the Common Spatial Subspace Decomposition (CSSD) method, a way of an adaptive feature extraction method. EEG signals have been classified by Improved-CSSD and SVM. Improved-CSSD improved the classification accuracy for about 8.26% over the traditional CSSD. The experiments indicate that the method has a low time loss and a decent adaptability [20]. As it can be seen, no signal decomposition techniques combined with HOS features have been studied before on Dataset IVa from BCI competition III, which is an important contribution of our study.

In this paper, we compare three different signal decomposition methods for MI BCI systems. Empirical Mode Decomposition (EMD), discrete wavelet transform (DWT), and wavelet packet decomposition (WPD) are used to generate several sub-band signals from which six different statistical features (including higher order statistics) are extracted. Finally, k nearest neighbor (k-NN) algorithm was responsible for the classification task. Our approach is partly motivated by the fact that EEG frequency bands hold different energies during different imaging tasks [1,2]. In addition, statistical features calculated from these sub-band signals (coefficients) will express additional evidence for discrimination between various imaging tasks. The performance of the proposed paradigm has been evaluated on the dataset IVa of BCI competition III. Our paper represents one out of a very few studies providing a thorough comparison of signal decomposition methods together with higher order statistics in BCI signal analysis [15,21].

The rest of this paper is organized as follows: Section 2 provides more information about the dataset being used. Section 3 explains the methods: a brief description of MSPCA-based denoising method and the descriptions of signal decomposition methods. The experimental results and analysis are presented in Section 4, whereas the Section 5 provides the throughout discussion, whereas the last sections concludes the paper.

2. Dataset

Dataset IVa from BCI competition III [22] was recorded from five healthy subjects sitting in a comfy chair with arms resting on armrests. Signals from 118 EEG channels of the extended international 10/20-system were captured and then band-pass filtered between 0.05 and 200 Hz. Although the sampling frequency used was 1000 Hz, EEG signals that are down-sampled at 100 Hz were also provided and used in this paper.

Visual cues showed the type of motor imagery the subject is to execute for 3.5 s: (R) right hand, or (F) foot. Periods of length around 2 s were introduced so that the subjects could take a short break. For each of five subjects, continuous signals having 118 EEG channels and markers showing the time points of 280 cues are available. Some markers contain NaN values as target class representing the instance belonging to evaluation/test data. Table 1 shows the number of training (labelled) trials and test (unlabelled) trials for all subjects.

Table 1
The number of training and test trials for each subject.

Subject	Training Trials	Test Trials
"aa"	168	112
"al"	224	56
"av"	84	196
"aw"	56	224
"ay"	28	252

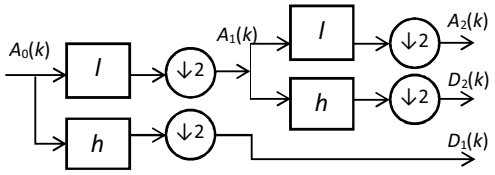

Fig. 1. DWT for scale level 2.

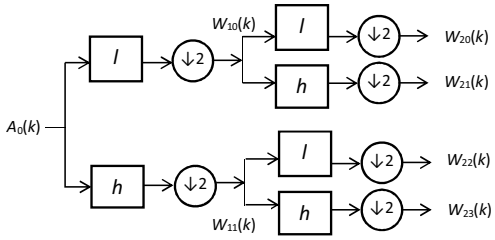

Fig. 2. WPD for scale level 2.

3. Methods

3.1. Signal decomposition methods: DWT, WPD, and EMD

Wavelet-based methods serve as suitable techniques for the analysis of different non-stationary signals like EEG. Discrete Wavelet Transform (DWT), for example, decomposes a discrete-time signal$x[k]$ into a set of signals (or wavelet coefficients) by scaling and shifting the mother wavelet. The first step in the DWT decomposition lies in picking the appropriate number of wavelet decomposition levels (or scale levels), j_m. For the initial scale level $j = 1$, signal $x[k]$ passes concurrently through both the high-pass and low-pass filter, $h[\cdot]$ and $l[\cdot]$ respectively, followed by the process of downsampling by 2. The output of each level j is represented in the form of two signals: Detail (D_j) and Approximation (A_j), described as:

$$D_j[i] = \sum_k x[k] \cdot h[2 \cdot i - k] \tag{1}$$

$$A_j[i] = \sum_k x[k] \cdot l[2 \cdot i - k] \tag{2}$$

Progressing to the next level, the approximation A_j is set as $x[k]$ and j is increases by 1. The above-mentioned procedure of generating D_j and A_j repeats as long as j does not surpass j_m [23], as shown in Fig. 1.

Wavelet Packet Decomposition (WPD), on the other hand, represents an extension of DWT in a way that the detail coefficients D_j are decomposed in addition to the decomposition ofA_j. This slight difference produces a different number of wavelet coefficient sets for both methods. In case of j-levels, DWT produces $j + 1$ sets of wavelet coefficients, whereas WPD generates 2^j sets of wavelet coefficients [6]. WPD achieves better frequency resolution for the decomposed signal than DWT as the later may miss important information in higher frequency components (Fig. 2).

Empirical Mode Decomposition (EMD) represents another method for the decomposition of non-linear and non-stationary signals like EEG. However, the result of such decomposition are not sets of coefficients but oscillatory modes better known as intrinsic mode functions (IMFs). A signal must meet certain requirements to be considered as IMF, such as [24]:

- The overall number of extrema and zero-crossings must be the same or differ by one;
- The envelope formed by the local minima and local maxima has the mean value very close to zero.

The EMD procedure for an input signal x is described below (sifting process) [24]:

1. Set $h = x \text{and} h_0 = h$
2. Every local maxima and local minima ofh_0 are found.
3. Lower envelope e_L, which connects all local minima with cubic spline, is determined
4. Upper envelope e_U, which connects all local maxima with cubic spline, is determined
5. The mean μ of e_U ande_L is calculated, i.e. $\mu = \frac{e_U + e_L}{2}$.
6. The first IMF candidate is generated, $h_1 = h_0 - \mu$

The sifting process is performed repeatedly on h_1 until the first IMF c_1 satisfying the IMF conditions is acquired. The first IMF is the highest frequency component of$x$, producing a residue$r_1 = x - c_1$ when subtracted from$x$. A residue is then used to find remaining IMFs by abiding to the above mentioned sifting process. This whole procedure terminates once the final residue r_n becomes constant, monotonic slope or a function with only one extremum. Eventually, signal x is decomposed into j_m number of IMFs where c_j is j^{th} IMF and r_A is the final residue [24]:

$$x = \sum_{j=1}^{j_m} c_j + r_A \tag{3}$$

3.2. MSPCA de-noising

Considering noisy signals like EEG, linear de-noising methods considerably smooth out the rapid jumps in the signal. On the other hand, the nonlinear filtering techniques like Multiscale Principal Component Analysis (MSPCA) may eliminate the noise without sacrificing substantial amount of fast changes in the signal [9].

Let us assume that we have the input signal matrix $X_{n\times p}$, where n is the number of measurements (samples) and p is the number of signals. Principal Component Analysis (PCA) converts this input matrix into two new matrices, L and S, such that the following holds [25]:

$$X_{n\times p} = LS^T \tag{4}$$

where L represents the matrix of principal component loadings, and S is the matrix of principal component scores. In order to perform de-noising, an appropriate number of loadings is usually chosen using several rules. Either loadings with eigenvalues greater than the average of all eigenvalues are preserved (Kaiser Criterion), or we keep the loadings which eigenvalues are greater than 5% of the sum of all eigenvalues (heuristic criteria). Since this parameter did not affect performance significantly, we used the default and more popular Kaiser rule to obtain the reduced matrix L', which will in return give us the de-noised signals in the input space:

$$X'_{n\times p} = L'S^T \tag{5}$$

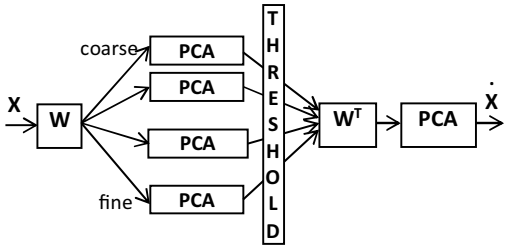

Fig. 3. The procedure for MSPCA.

The PCA algorithm is capable of splitting linear relationship between the variables in the form of matrix of loadings L. On the other hand, wavelets are used to split deterministic features from stochastic ones, resulting in approximate decorrelation of autocorrelation between the signals. The MSPCA algorithm is a combination of PCA and wavelets built on the strengths of both. By referring to Fig. 1, convolution with a filter "l" denotes the projection on the scaling function, whereas convolution with a filter "h" denotes the projection on a wavelet. If we assume $L(z)$, $H(z)$, $Cm(z)$ and $Dm(z)$ to be z-transforms of "l", "h", A_m and D_m respectively, we can write the following:

$$A_m(z) = L(z)A_{m-1}(z) \tag{6}$$

$$D_m(z) = H(z)A_{m-1}(z) \tag{7}$$

The overall filter that produces the detail coefficients $D_m(k)$ represents a cascade of the highpass filter "h" and the increasing number of lowpass filters "l". On the other hand, the coefficients $A_m(k)$ are acquired by cascaded lowpass filters only. In addition, the original waveform can be viewed as the vector of scaling function coefficients at the finest scale $m = 0$, $x(k) = A_0(k)$. Therefore, Eqs. (6) and (7) can be represented using the original signal $X(z)$ as:

(8) $A_m(z) = L_m(z)X(z)$,

(9) $D_m(z) = H_m(z)X(z)$,where $L_m(z)$ is acquired by applying the $L(z)$ filter m times, whereas $H_m(z)$ is acquired by applying the $H(z)$ filter once and the $L(z)$ filter $(m-1)$ times [26].

Let us now go back to the input signal matrix $X_{n\times p}$, containing p signals (waveforms) of length n. Each variable related to a signal (column) in X is decomposed to the wavelet coefficients using the orthonormal matrix W, representing the wavelet transformation operator. The matrix W holds the matrices of the filter coefficients H_m and L_J [26]:

$$\mathbf{W} = \begin{bmatrix} \mathbf{L}_J & \mathbf{H}_J & \mathbf{H}_{J-1} & \cdots & \mathbf{H}_m & \cdots & \mathbf{H}_1 \end{bmatrix}^T. \tag{10}$$

The procedure for MSPCA, as explained in [27], can be summarized in three steps. The wavelet decomposition of all signals from $X_{n\times m}$ is executed in the initial step. Then, PCA de-noising algorithm is performed for each wavelet decomposition level separately (for all signals belonging to a certain sub-band). If stated, wavelet coefficients which are greater than certain a priori threshold value can be preserved as well. The third step considers the application of PCA for all levels combined, resulting in the de-noised input signal matrix $\hat{X}_{n\times m}$ [27]. These steps of the MSPCA are illustrated in Fig. 3, and provide improved de-noising performance than sole PCA algorithm.

In our previous studies, MSPCA has been effectively exploited in biomedical signal processing such as classifying EEG signals for epileptic seizure detection [10]. The positive effects of noise removal provided by MSPCA for EEG signals are thoroughly explained [10]. The feature extraction techniques were applied to the EEG signals without MSPCA de-noising process to show the necessity and the impact of such step in the EEG signal analysis. A step-by-step illustration of the data analysis with exemplars was used to depict how much the MSPCA de-noising contributed to the final performance [10] .

Apart from EEG, MSPCA has been applied for diagnosis of neuromuscular disorders using Electromyography (EMG) signals [11]. In addition, MSPCA showed its strength for diagnosis of cardiovascular diseases by classifying ECG beats [12]. The effect of MSPCA in real-time wireless BCI system was investigated in [13]. The study involved practical motor imagery BCI system based on Mindwave Mobile Headset from one subject including two motor imagery tasks: right hand and left hand. The model combining MSPCA de-noising and statistical wavelet features showed promising results [13].

3.3. Feature extraction

DWT and WPD decomposition techniques will represent the EEG signals as a few sets of wavelet coefficients. EMD method, however, generates IMFs which length (duration) is the same as the original input signal. The ability to convert this data (set of coefficients or signals) into a reduced set of features represents very important step in any classification task. Since these features characterize the behavior of the BCI EEG signals, their selection is of crucial importance. The following six statistical features are chosen for EEG classification:

1. Mean of coefficients' (signals') absolute values in every sub-band, $\mu = \frac{1}{M}\sum_{j=1}^{M}|y_j|$
2. Average power of the coefficients (signal) in every sub-band, $\lambda = \sqrt{\frac{1}{M}\sum_{j=1}^{M}y_j^2}$;
3. Standard deviation of the coefficients (signal) in every sub-band, $\sigma = \sqrt{\frac{1}{M}\sum_{j=1}^{M}(y_j-\mu)^2}$;
4. Ratio of the absolute mean values of coefficients (signal values) of adjacent sub-bands, $\chi = \frac{\sum_{j=1}^{M}|y_j|}{\sum_{j=1}^{M}|z_j|}$
5. Skewness of the coefficients (signal) in every sub-band, $\phi = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\frac{(y_j-\mu)^3}{\sigma^3}}$;
6. Kurtosis of the coefficients (signal) in every sub-band, $\phi = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\frac{(y_j-\mu)^4}{\sigma^4}}$.

If $Y\{y_1, y_2, ..., y_M\}$ and$Z\{z_1, z_2, ..., z_M\}$ are two adjacent sub-bands (or IMFs), where M represents the length of a sub-band (or the number of signal samples).

Features 5 and 6, known as the third and fourth cumulants skewness and kurtosis, are considered in the literature as higher order statistics (HOS) features. In this paper, three manually selected sets of features were explored as well:

- Features 1–4 (Experiment 1)
- Features 1–2 and 5–6 (Experiment 2)

- All six features 1–6 (Experiment 3)

Note that HOS features are only extracted in Experiment 2 and 3, being absent in Experiment 1.

3.4. K-nearest neighbor classification

One of the simplest classification algorithms, *k*-nearest neighbor (*k*-NN), has received significant consideration making it very popular among researchers [28]. This type of classifier does not require training and provides robust performance, especially in case of two-class problems like in Dataset IVa [14]. In addition, *NN*'s decision boundary can undertake any kind of form since *k*-NN is non-parametric algorithm making no assumption about the distribution of data. This is of utmost importance when distinction between two noisy EEG signals needs to be made. Furthermore, the dataset that we use in this study does not consist of a very large number of data points (instances), so the usage of *k*-NN is quite justified.

If we assume that $P = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\}$ is a set of m training signals with corresponding class labels, then there exists $\mathbf{x}_i \in P\,(1 \leq i \leq m)$ such that it represents the nearest signal (neighbor) to a test signal $\mathbf{x}$. In order to locate the nearest neighbor (signal) for any test signal $\mathbf{x}$, we need to perform m distance calculations, after which $\mathbf{x}$ is given the class label of its nearest neighbor $\mathbf{x}_i$ [29]. There is a choice of a few distance measures to be used, but Euclidean distance is usually selected. The attributes (features) of a signal are usually scaled to give equal treatment to all dimensions (or to make the *k*-NN independent of measurement units). Weka [30] implementation of *k*-NN was used in this paper, with *k* parameter set to 7.

4. Experimental setup

4.1. Database preparation

The EEG segments accounting for only the motor imagery part were extracted from the database. For that purpose, we used the position of markers that indicated the beginning of 280 cues and the fact that every movement was 3.5 s long. At the end of this process, we had 280 EEG segments (with 118 channels) belonging to two classes (right hand and foot) for each patient.

The next stage was responsible for removing the noise from these EEG segments. The MSPCA algorithm was applied onto 118 channels of every motor imagery separately. However, further processing was not carried out on all 118 channels, but on the following three channels: C3, Cz, and C4. These three channels were selected because they hold the most discriminative information regarding motor imagery activities involving hands and foot. Noteworthy variations related to imagined right hand movement are usually perceived above the contralateral left motor cortex around C3 electrode. In the same way, foot movements should be seen around Cz electrode [31].

4.2. Parameter and feature selection

The first step in the proposed signal processing procedure was devoted to de-noising of EEG signals by MSPCA algorithm. The Kaiser principle, which keeps only those principal components whose eigenvalues were bigger than the average of all eigenvalues, was selected. For the wavelet segment of the MSPCA algorithm, the number of wavelet decomposition levels (or scale levels) was selected as 5. Considering the sampling rate of 100 Hz for BCI database IVa, the frequency range of approximation signal A5 was 0–1.56 Hz. The detail and approximation signals were acquired by Sym4 wavelet. The Sym4 wavelet function produced the best overall results in our testing. In fact, other wavelet functions did not result in drastically worse performance either. The choice of a proper wavelet function is usually performed by trial-and-error approach.

Table 2
Sub-band signals for 4 levels using WPD.

# Signals	Decomposed Signal	Frequency Range (Hz)	5-level DWT
1	WP40	0.00–3.13	A5, D5
2	WP41	3.13–6.25	D4
3	WP42	6.25–9.38	D3
4	WP43	9.38–12.50	
5	WP44	12.50–15.63	D2
6	WP45	15.63–18.75	
7	WP46	18.75–21.88	
8	WP47	21.88–25.00	
9	WP48	25.00–28.13	D1
10	WP49	28.13–31.25	
11	WP4A	31.25–34.38	
12	WP4B	34.38–37.50	
13	WP4C	37.50–40.63	
14	WP4D	40.63–43.75	
15	WP4E	43.75–46.88	
16	WP4F	46.88–50.00	

Table 3
Number of features in feature vector for every decomposition method.

Dec. Method	# Sub-bands	# Features in vector	
		EXP 3	EXP 1–2
DWT	6	36	24
WPD	16	96	64
EMD	6	36	24

After MSPCA de-noising and prior to process of extracting useful features, de-noised EEG signals were decomposed into sub-band signals using DWT, WPD, and EMD. Using EMD, the minimum of 5 IMFs could have been extracted for all EEG segments from all subjects. Hence, each de-noised EEG signals was decomposed into 6 signals if we take the residue signal into consideration.

This number of generated IMFs affected the number of wavelet decomposition levels used in DWT. In order to generate six sub-band signals using DWT, the number of scale levels was selected to be 5, the same as in the wavelet part of the previously explained MSPCA de-noising algorithm.

Apart from the effort to match the number of IMFs and sub-bands (which will result in the same number of features afterwards), the option of 5 scale levels is also justified by the frequency range of the resulting approximation signal, 0–1.56 Hz. There is certainly no purpose to go for more scale levels, making the range of the resulting approximation even lower.

However, the number of scale levels for WPD was selected to be 4. This combination of parameters resulted in $2^j = 2^4 = 16$ sub-bands represented in Table 2 with the corresponding relations to DWT sub-bands.

It can immediately be noticed that such selection makes WPD generating 10 more sub-bands than DWT or EMD, which will result in great unbalance in the number of overall features.

Six different statistical features (Experiment 1) and their combinations (Experiment 2 and 3) explained in Section 3.3 were extracted from sub-band signals of one de-noised EEG signal. The total number of features in a feature vector for each decomposition method within our comparison is given in Table 3.

5. Results and discussion

Although every subject contained separate train and test sets, they were combined into one dataset due to the low number of trials. This approach also influenced the performance evaluation criteria for *k*-NN classifier, with *k* parameter set to 7. For that pur-

Table 4
Classification Accuracy using EMD.

	AA	AL	AV	AW	AY	ALL
EXP1	54.0	56.1	49.4	55.1	75.3	58.2
EXP2	59.8	59.7	57.9	54.1	88.4	62.4
EXP3	56.6	60.7	55.9	55.1	90.1	62.8

Table 5
Classification Accuracy using DWT.

	AA	AL	AV	AW	AY	ALL
EXP1	62.7	65.7	61.1	59.8	67.0	64.7
EXP2	80.7	76.1	76.3	86.3	86.0	82.0
EXP3	77.1	72.2	75.2	85.6	86.0	81.1

Table 6
Classification Accuracy using 4 level WPD.

	AA	AL	AV	AW	AY	ALL
EXP1	83.0	81.4	78.4	85.3	75.4	82.6
EXP2	96.0	92.0	88.3	95.4	91.4	93.7
EXP3	94.8	92.3	88.9	94.7	91.1	94.5

pose, 10-fold cross-validation (CV) approach is employed. Dataset is arbitrarily divided into 10 mutually exclusive folds (subsets) of practically the identical size. Nine (9) folds are used for training and remaining one (1) fold is used for testing so the process repeats 10 times. The average of accuracies of each iteration is called the CV accuracy (or Overall Classification Accuracy).

Tables 4–6 show the Overall Classification Accuracies (in%) for all five subjects and each decomposition method. Results for all three feature combinations are presented as well. The last column (ALL) represents CV accuracies when feature vectors of all subjects were combined together. Tables 4–6 provide performance evaluation measures for clear comparison between three signal decomposition techniques, namely EMD, DWT, and EMD.

The numbers in these tables show that the combination of features in Experiment 2 outperforms Experiment 3, whereas the Experiment 1 produced the worst results by a huge margin. On the other hand, WPD was superior to DWT and especially EMD. Overall Classification Accuracies between the subjects were rather comparable.

The superiority of WPD may be explained by Table 2. Namely, the highest frequency sub-band of DWT (the detail D1) is split among eight WPD sub-bands (WP48-4F). This will produce much more features in the 25–50 Hz range for WPD, which resulted in better discrimination between the classes.

It means that WPD provides finer decomposition of highest frequencies using lower number of scale levels when compared to DWT. In regards to features, WPD generates a number of very discriminative features for reasonable scale level (Table 3 and 6). Producing the comparable number of features by DWT or EMD would require using very high and practically unjustified number of scale levels or IMFs.

In order to obtain further insight into the performance of the superior WPD method, the statistical performances such as True Positive Rates for each class together with total accuracies of the *k*-NN classifier are presented in Table 7. Overall, it can be concluded that the detection accuracies for "right hand" were slightly better than those for the "foot" class. However, that difference was never more than 5% in favor to "right hand" class. Moreover, there were a few cases when detection of "foot" class was slightly higher (Table 7).

The importance of the high frequency range 25–50 Hz may be further highlighted by connecting the poor performance of EMD technique presented in Table 4with power spectral densities (PSDs) showed in Figs. 4–6 . These figures show the PSDs of the highest frequency component generated from EEG signals using EMD (Figs. 4 and 5 and DWT (Fig. 6). Fig. 4 (Fig. 5) represents PSDs of IMF1 for three MSPCA de-noised motor-imagery EEG signals from subject AA belonging to right hand (foot) class. The three EEG signals have been selected based on the maximum number of generated IMFs. Fig. 6, however, represents PSDs of the detail 1 (D1) for the same three EEG signals as in Fig. 4.

We also show Figs. 4 and 5 in order to investigate the differences in PSDs of IMF1 for EEG signals that generate different number of IMFs, and to see if EEG signals of different class have different frequency distribution.

Figs. 4 and 5 clearly show that the frequency range 35–45 Hz is rarely covered by IMF1, whereas the detail 1 (D1) perfectly covers the whole range 25–50 Hz (Fig. 6). One exception for EMD can be seen in Fig. 4, where the PSD of IMF1 for right hand EEG signal generating 9 IMFs nicely covers frequencies above 25 Hz (green line). The other exception can be seen in Fig. 5 where the PSD of IMF1 for foot EEG signal generating 7 IMFs (red line) better covers higher frequencies than the other two EEG signals. Apart from these two exceptions, all other PSDs in Figs. 4 and 5 generally cover the frequency range 15–35 Hz. Moreover, the PSDs for these two exceptions are much worse than the ones presented in Fig. 6 for DWT in terms of covering highest frequencies. Therefore, much better classification accuracy has been achieved when the features were extracted from DWT sub-bands (Tables 4 and 5).

We can conclude that IMFs generated from EEG signals using EMD generally do not hold highest frequencies and this information is usually lost during decomposition. On the other hand, the majority of information lost by EMD is contained in the detail 1 (D1) generated by DWT. The features extracted from this detail are the ones that highly affect the classification accuracy since IMFs other than IMF1 cover the same lower frequencies as A5 and D2-D5 (in DWT).

The confusion about the different number of generated IMFs and the whole analysis approach may be solved by using Enseble EMD (EEMD) method, which produces the same number of IMFs for signals of the same duration (length). The additional advantage of EEMD method may be the short duration of BCI signals which would in turn produce fewer number of IMFs to analyze.

Another notable comparison presented in Tables 4–6 is related to three sets of features: EXP1, EXP2, and EXP3. In general, classification accuracy was poorest in EXP1 where features 1 – 4 were used. The superior performance of EXP2 and EXP3 may be explained by the inclusion of features 5–6 (skewness and kurtosis) which belong to higher order statistics (HOS). EEG signals are non-linear, non-stationary and non-Gaussian in nature. The first and second order statistics have attracted substantial significance in biomedical signal processing although being very restrictive in analyzing non-linearity of signals such as EEG [14]. On the other hand, different types of wavelet transforms and decompositions are appropriate for the biomedical signal analysis due to their superior resolution and ability to derive dynamical features from the signals. Nonetheless, when it comes to extracting non-linear relationships within the signal, even the powerful wavelets fail [15]. HOS can be used as an influential tool for the analysis of non-linear behavior of weak and noisy biomedical signals such as EEG. HOS is able to discover abnormalities from linearity, stationarity or Gaussianity in the signal which makes it a superior methodology than traditional time and/or frequency domain approaches in the analysis of biomedical signals. In addition, features extracted from HOS can be made non-susceptible to shift or amplification [14].

The classification accuracy of the suggested paradigm, as the main standard for evaluating different BCI approaches, is taken as a quantitative performance indicator. Table 8 offers a comparative study of the most recent BCI approaches developed for dataset IVa

Table 7
True Positive Rates for "right hand" and "foot" classes together with the accuracies using 4 level WPD.

	AA		AL		AV		AW		AY		ALL	
EXP1	83.0		81.4		78.4		85.3		75.4		82.6	
	84.3	81.9	83.1	79.8	78.6	78.3	84.8	86.0	77.4	73.6	84.0	81.1
EXP2	96.0		92.0		88.3		95.4		91.4		93.7	
	96.4	95.7	94.0	90.0	88.3	88.3	94.8	96.2	91.9	91.0	94.0	93.4
EXP3	94.8		92.3		88.9		94.7		91.1		94.5	
	94.5	95.2	94.0	90.7	89.0	88.8	94.0	95.5	91.3	91.0	95.3	93.9

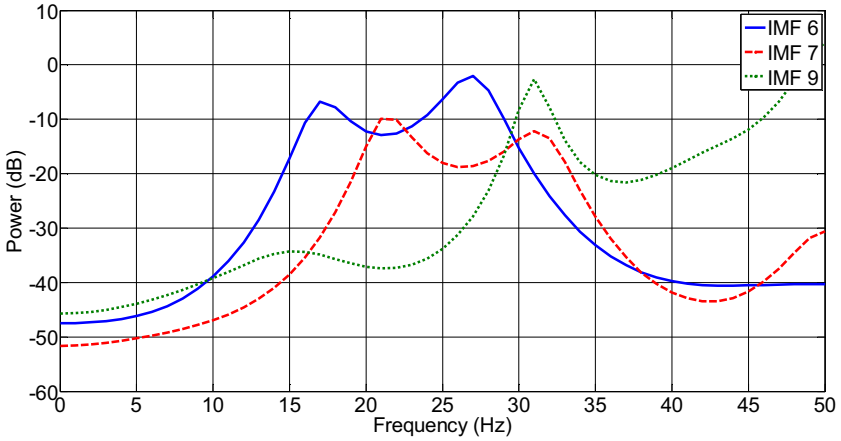

Fig. 4. PSDs of IMF1 for three MSPCA de-noised motor-imagery EEG signals belonging to right hand class.

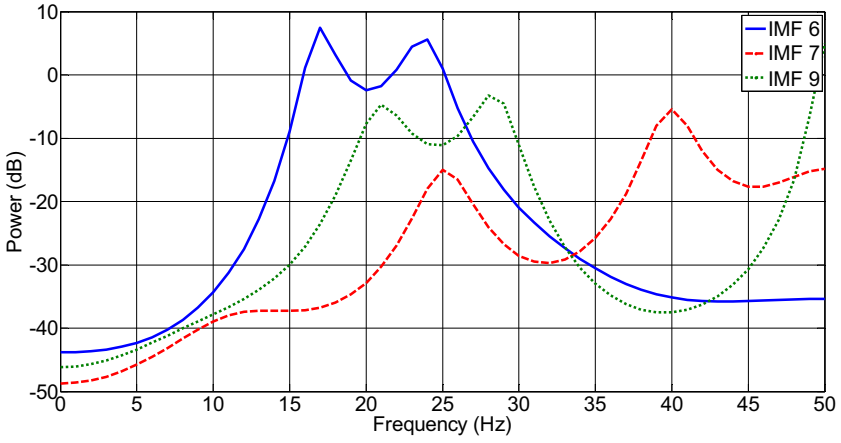

Fig. 5. PSDs of IMF1 for three MSPCA de-noised motor-imagery EEG signals belonging to foot class.

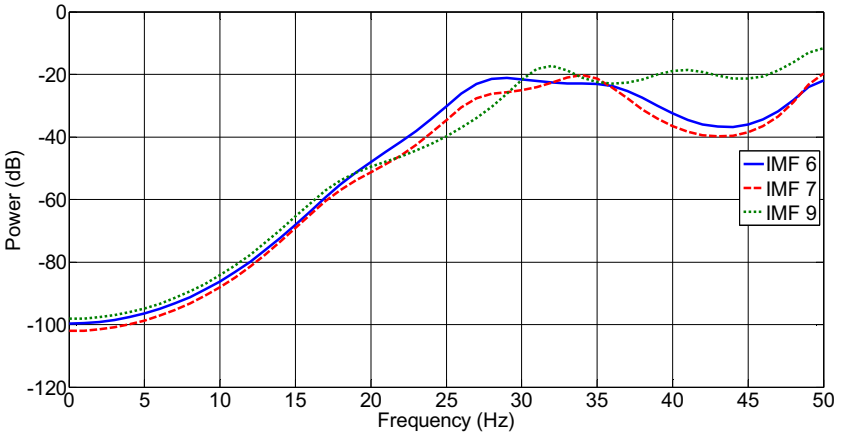

Fig. 6. PSDs of D1 for three MSPCA de-noised motor-imagery EEG signals belonging to right hand class.

Table 8
Comparison of the classification performance for dataset IVa in BCI competition III.

Authors	Method	*Accuracy (%)*					
		aa	*al*	*av*	*aw*	*ay*	*Average*
Wu et al. [16]	ISSPL	93.57	**100**	79.29	99.64	**98.57**	**94.21**
Siuly et al. [17]	CC-LR + statistics	**100**	94.23	**100**	**100**	75.33	93.91
Meng et al. [18]	CSP + channel selection + SVM	82.4	98.6	76.8	94	96.6	89.68
Hu et al. [19]	R-CSP	76.8	98.2	74.5	92.9	77.0	83.9
Li and Lu [20]	CSSD + SVM	78.6	99.6	67.5	75.11	–	80.20
This work	*MSPCA + WPD + HOS + k-NN*	*96*	*92.3*	*88.9*	*95.4*	*91.4*	*92.8*

with classification accuracies above 80% across all subjects (average accuracy). For every subject and their averages, the highest values of classification accuracy are emphasized in bold.

As it can be seen from Table 8, the proposed algorithm does not offer the best classification accuracy for any of the five subjects. At least one of the algorithms in Table 8 achieves 100% accuracy for the first 4 subjects, whereas three algorithms reported accuracies above 96% for subject "ay". On the other hand, the highest classification accuracy achieved by our proposed approach is 96% for subject "aa".

Despite that, the proposed method has produced the steadiest performance across the subjects than any other paradigm in Table 8. As shown, the worst classification accuracy achieved by our approach is 88.9% for subject "av" (other approaches in Table 8 also struggled the most with this subject), whereas the accuracy is above 90% for all other subjects. Every approach reported in Table 8 struggles with at least one subject in a way that the classification accuracy for that particular subject could not surpass 81%.

The robustness of our approach across the subjects may have resulted in lower classification accuracy for subject "al". However, our approach provided the second best classification accuracy for two subjects: "aa" and "av". As a result, the average classification accuracy of the proposed algorithm is 92.8% for the dataset IVa. Furthermore, the accuracy when feature vectors of all subjects were combined together is reported to be 94.5% (Table 6). These results show that the performance of our method is very close and comparable to the best results reported in the literature.

6. Conclusion

In this study, three popular signal decomposition methods were investigated for the classification task of EEG signals in BCI system. Database IVa, a multichannel 2-class motor-imagery dataset, was used for this purpose. MSPCA method was applied for the purpose of noise removal. In addition, different sets of features were formed to examine the effect of particular group of features. Our results show that the combination of MSPCA de-noising and higher order statistics (HOS) features extracted from wavelet packet decomposition (WPD) sub-bands resulted in highest classification accuracy. The proposed method uses WPD coefficients that have higher frequency resolution than DWT coefficients, whereas the shortcomings of wavelets approaches have been compensated by HOS. Our study is one among very few approaches delivering a detailed comparison of signal decomposition methods combined with higher order statistics in classification of BCI signals. These results show that the proposed model has the potential to obtain a reliable classification of MI EEG signals, and can be used practically in controlling a wheelchair for example. As far as the medical applications are concerned, appropriate feedback is delivered once the individual executes the correct movement in stroke rehabilitation. Therefore, the proposed model may enhance the current rehabilitation therapies, or improve motor rehabilitation outcomes.

Acknowledgement

This research has been supported by International Burch University, Sarajevo, Bosnia and Herzegovina.

References

[1] H.-J. Hwang, S. Kim, S. Choi, C.-H. Im, EEG-Based brain-Computer interfaces: a thorough literature survey, Int. J. Hum. Comp. Interact. 29 (2013) 814–826.
[2] R.J. Rak, M. Kołodziej, A. Majkowski, Brain-computer interface as measurements and control system. The review Paper, Metrol. Meas. Syst. 19 (3) (2012) 427–444.
[3] G. Rodríguez-Bermúdez, P.J. García-Laencina, Automatic and adaptive classification of electroencephalographic signals for brain computer interfaces, J. Med. Stems 36 (2012) S51–S63.
[4] D.J. Krusienski, D.J. McFarland, J.R. Wolpaw, An evaluation of autoregressive spectral estimation model order for brain-Computer interface applications, in: Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biologyociety, New York, USA, September, 2006.
[5] H. Ramoser, J. Müller-Gerking, G. Pfurtscheller, Optimal spatial filtering of single trial EEG during imagined hand movement, IEEE Trans. Rehabil. Eng. (2000) 441–446.
[6] M. Unser, A. Aldroubi, A review owavelets in biomedical applications, Proc. IEEE 84 (4) (1996) 626–638.
[7] D. Gajic, Z. Djurovic, S.D. Gennaro, F. Gustafsson, Classification of EEG signals for detection of epileptic seizures based on wavelets and statistical pattern recognition, Biomed. Eng. 26 (2) (2014).
[8] D. Gajic, Z. Djurovic, J. Gligorijevic, S.D. Gennaro, I. Savic-Gajic, Detection of epileptiform activity in EEG signals based on time-frequency and non-linear analysis, Front. Comput. Neurosci. 9 (2015).
[9] L. Sornmo, P. Laguna, Bioelectrical Signal Processing in Cardiac and Neurological Applications, Elsevier Academic Press, 2005, 2016.
[10] A. Kevric, A. Subasi, The effect of multiscale PCA de-noising in epileptic seizure detection, J. Med. Syst. 38 (August (10)) (2014).
[11] E. Gokgoz, A. Subasi, Effectof multiscale PCA de-noising on EMG signal classification for diagnosis of neuromuscular disorders, J. Med. Syst. 38 (April (4)) (2014) 1–10.
[12] E. Alickovic, A. Subasi, Effectof Multiscale PCA de-noising in ECG beat classification for diagnosis ofcardiovascular diseases, Circuits Syst. Signal Process. 34 (2) (2015) 513–533.
[13] J. Kevric, A. Subasi, The impact of mspca signal de-Noising In real-Time wireless brain ComputerInterface system, Southeast Eur. J. Soft Comput. 4 (1) (2015) 43–47.
[14] Y. Kutlu, D. Kuntalp, Featureextraction for ECG heartbeats using higher order statistics of WPDcoefficients, Comput. Methods Programs Biomed. (2012) 257–267.
[15] K.C. Chua, V. Chandran, R. Acharya, C.M. Lim, Application of higher order spectra toIdentify epileptic EEG, J. Med. Syst. 35 (2011) 1563–1571.
[16] W. Wu, X. Gao, B. Hong, S. Gao, Classifying single-tal EEG during motor imagery by iterativespatio-spectral patterns learning (ISSPL), IEEE Trans. Biomed. Eng. (2008).
[17] Siuly, Y. Liand, P. Wen, Modified CC-LR algorithm with three diverse feature sets for motorimagery tasks classification in EEG based brain computer interface, Comput. Methods Programs Biomed. 113 (3) (2014) 767–780.
[18] J. Meng, G. Liu, G. Huang, X. Zhu, Automated selecting subset of channels based on CSP in motor imagery brain-Computer interface system, in: IEEE International Conference OnRobotics and Biomimetics, Gulin, China, December 19–23, 2009.
[19] H. Lu, H.L. Eng, C. Guan, K.N. Plataniotis, A.N. Venetsanopoulos, Regularizedcommon spatial patterns with aggregation for EEG classification in small-samplesetting, IEEE Trans. Biomed. Eng. 57 (2010) 2936–2945.
[20] M. Li, C.C. Lu, The recognition of EEG with CSSD and SVM, in: 10th World Congress OnIntelligent Control and Automation, Beijing, China, July 6–8, 2012.
[21] M. Kołodziej, A. Majkowski, R.J. Rak, A new method of EEG classification for BCI with feature extraction based on higher order statiics of wavelet components and selection withGenetic Algorithms,inAdaptive and natural computing algorithms, in: 10th International Conference, ICANNGA 2011, Ljubljana, Slovenia, April 14–16, 2011, Proceedings, Part I, Springer Berlin Heidelberg, 2011, pp. 280–289.

[22] B. Blankertz, K.R. Müller, D.J. Krusienski, G. Schalk, J.R. Wolpaw, A. Schlögl, G. Pfurtscheller, R.M. Jdel, M. Schröder, N. Birbaumer, The BCI competition III:Validating alternative approaches to actual BCI problems, IEEE Tansa. Neural Syst. Rehabil. Eng. 14 (2) (2006) 153–159.
[23] P. Ghorbanian, D.M. Devilbiss, A.J. Son, A. Bernstein, T. Hess, H. Ashrafiuon, Wavelet transform EEG features of alzheimer's disease in activated states, in: 34thAnnual International Conference of the IEEE EMBS, San Diego, California USA, 2012.
[24] N.E. Huang, Z. Shen, S.R. Long, M.C. Wu, H.H. Shih, Q. Zheng, N.C. Yen, C.C. Tung, H.H. Liu, Theempirical mode decomposition and the Hilbert spectrum fornonlinear andnon-stationartime series analysis, Proc. R. Soc. Lond A 74 (1971) (1998) 903–995.
[25] I.T. Jolliffe, Principal Component Analysis, Springer, New York, USA, 2002.
[26] B.R. Bakshi, Multiscale PCA with apication to multivariate statistical process monitoring, AlChE 44 (7) (1998) 1596–1610.
[27] B.R. Bakshi, Multiscale PCA with application to multivarie statistical process monitoring, AlChE J. 44 (7) (July 1998) 1596–1610.
[28] B.V. Dasarathy, NN concepts and techniques. An introductory survey, in: B.V. Dasarathy (Ed.), Nearest Neighbour Norm: NN Pattern Classification Techniques, IEEE Computer Society Press, Los Alamitos, CA, 1991, pp. 1–30.
[29] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, John Wiley and Sons, New York, 2000.
[30] G. Holmes, A. Donkin, I.H. Witten, WEKA: a machine learning workbench, in: Proceedings of the 1994 Second Australian and New Zealand Conference on Intelligent Information Systems, Brisbane, Australia, 1994.
[31] J.A. Wilson, G. Schalk, L.M. Walton, J.C. Williams, Using an EEG-Based brain-computer interface for virtual cursor movement with BCI2000, J. Visualized Exp. 29 (2009).