A Comparative Study of Wavelet and CSP Features Classified using LDA, SVM and ANN in EEG based Motor Imagery

Majid Aljalal and Ridha Djemal
Electrical Engineering Department
College of Engineering, King Saud University
P.O.Box 800 – 11421 Riyadh, KSA
algalalmajid@hotmail.com
rdjemal@ksu.edu.sa

***Abstract*— Brain-computer interface (BCI) can interchange messages and orders between the user's brain and the computer. The motor imagery (MI) is presented by specific signal features that reflect the user's intention to be extracted and interpreted as commands. This paper focuses on the classification of two types of MI tasks (Right Hand and Foot). We deployed various feature extraction techniques for EEG data using wavelet transform and common spatial pattern. For the wavelet features, statistical values, energy, entropy and band power were used to form the desired feature vectors. Before extracting wavelet coefficients, we performed two scenarios, with and without surface laplacian filter around the channels C3, C4 and Cz. Three types of classifiers were employed for classification, linear discriminant analysis (LDA), support vector machines (SVM) and artificial neural network (ANN). The aim of this work is to compare between them and to recommend the suitable combination for synchronous two-class motor-imagery-based brain-computer interface experiments. The data were recorded from five subjects, provided by BCI-Competition III. The results show that SVM is more suitable with the features than those extracted from wavelet coefficients and combination of entropy-energy-band power, and LDA is more suitable with common spatial pattern. Overall, the results from CSP-LDA are better than those obtained from WT-SVM with the average classification accuracy of 84.79% and 82.64%, respectively.**

Keywords— artificial neural network, brain computer interface, common spatial pattern, motor imagery, linear discriminant analysis, support vector machines, wavelet transform

I. INTRODUCTION

Brain Computer Interface (BCI) is a system that can identify what human think and interchange messages and commands between the user's brain and the computer with high accuracy. As per Donoghue et al.: Providing a possibility to command external device/machine directly from the brain is the major goal of brain-machine interface [1]. This will be very useful to help people with body disability.

BCI can be mostly classified into invasive and noninvasive models. Invasive BCI processes signals inside the brain, whereas non-invasive BCI processes signals outside the brain. Although the captured signals from invasive BCI are stronger, but a surgery will be requested [2]. Brain activity sometime called "rhythms". These rhythms are micro volt electric signals (i.e. Electroencephalogram or EEG) which are created in our brain while we are doing a task [3, 4].

The three main types for EEG based BCI system have been used based the three following paradigms: P300, SSVEP and ERD/ERS. P300 is a localized response to a joined visual, auditory, or tactile stimulus, and it is mainly red from the parietal lobe during 300 msec after starting of the stimulus [5, 6]. Steady-State Visual Evoked Potentials (SSVEP) is a response to a visual stimulus modulated at a frequency greater than 6 Hz, which can be red from the occipital area [7, 8]. The event-related desynchronization (ERD) and event-related synchronization (ERS), which are prompted by performing mental tasks, such as motor imagery, mental arithmetic, or mental rotation [9]. In the case of the motor imagery paradigm, the mu (8–13Hz) and beta (14–30Hz) rhythms of the sensorimotor cortex are used [10]. During physical and motor imagery of right and left hand movements, beta band brain activation β-ERD happens predominantly over the contralateral left and right motor areas and a β-ERS ipsilaterally. The post movement ERS related with ceasing to move can also be found over the contralateral motor areas. The appropriate analyzing and processing (filtering, features extraction and classification) of MI states can generate appropriate commands to control robots, rehabilitation devices or any application else.

To guarantee a higher level of accuracy for motor imagery, preprocessing methods, containing both spatial and band-pass filter are requested such as in [11]. The two algorithms frequently used for feature extraction of the data after preprocessing procedures are wavelet transform (WT) and common spatial patterns (CSP) [12-14]. A. Baziyad et al aggregated the features extracted from CSP with others extracted from WT and concluded the best results were obtained when using SVM as classifier with an average classification accuracy of 75% over three subjects [15]. R.

978-1-5386-2756-3/17/$31.00 ©2017 IEEE

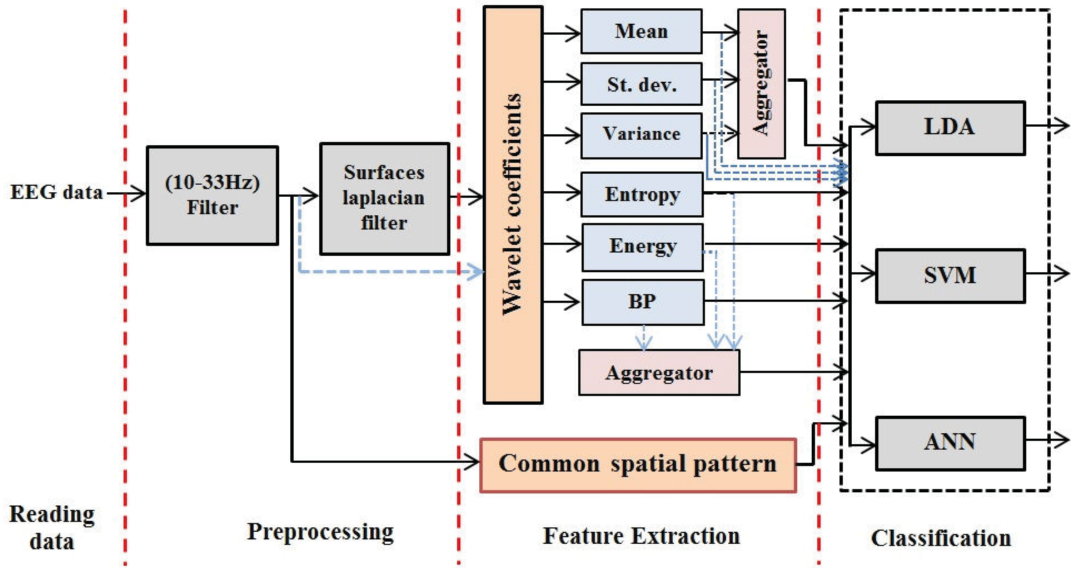

Fig. 1 Block diagram of the proposed methods.

Chatterjee et al employed wavelet-based energy-entropy & RMS with SVM and MLP classifiers and obtained accuracy of 85% and 85.71% respectively over only one subject [16]. In this paper, we deployed various feature extraction techniques on the EEG data using common spatial pattern and wavelet coefficients. In the case of wavelet features, statistical values, energy, entropy, band power and combination between them were performed to form the desired feature vectors. Three types of algorithms were employed for classification, linear discriminant analysis (LDA), support vector machines (SVM) and artificial neural network (ANN). The aim of this paper is to compare between them and to recommend a suitable combination for synchronous two-class (Right Hand and Foot) motor-imagery-based brain-computer interface experiments. Classification accuracy with cross-validation over five subjects was employed to evaluate performance of methodology.

The remainder of the paper is organized into three sections. Section II describes the EEG data set used in our work. This section also briefly explains various features extraction techniques, experimental results are presented and discussed in Section III and finally Section IV concludes the paper.

II. METHODS

In this section we describe the methods of feature extraction and classification techniques as well as their validation using Matlab simulation tools. Fig. 1 shows the block diagram of the proposed methods. First of all, EEG data was red and then segmented into time windows of 3 sec with offset of 0.5 sec. Then, the output of the segmentation process was fed to bandpass filter. In this work, we used frequency band of 10-30Hz because it gave good results. Before extract wavelet coefficients, we implemented two scenarios, first with surface laplacian filter around C3, C4 & Cz, and the other scenario is using all channels without surface laplacian filter. In the other hand, common spatial pattern algorithm was applied to the filtered signals. In order to extract the feature vectors, we used statistical values, entropy, energy and band-power with wavelet coefficient and used logarithmic variance with common spatial pattern. Linear discriminant analysis, support vector machines and artificial neural network techniques were then applied as classifiers. All possible combinations of the proposed approach were implemented and tested.

A. Datasets

In our work, we have taken the experimental data-set from BCI competition III. This dataset (III-IVa) [13] was provided by Berlin BCI group: Fraunhofer First, Intelligent Data Analysis Gro up (Klaus Robert Muller, Benjamin Blankertz) and campus Benjamin Franklin of the Charite University Medicine Berlin .The datasets are for two class motor imagery (right hand and foot) taken from 5 subjects. For each subject they conducted 280 (right hand 140 / foot 140) trials. Data was recorded from 118 EEG channels with 100Hz sampling rate for each trial.

B. Preprocessing

For analysis raw EEG data was arranged into E-matrices. Each E-matrix is of size (N x T), where “N” is the number of channels and “T”, time windows, is the number of EEG samples per channel in a specific interval of time “T”. In our work, for each trial, time windows of 3 sec with offsets of 0.5 sec were applied for all subjects.

After windowing raw signal the, EEG data was filtered using bandpass butterworth filter of order 4 to remove the artifacts coming from others frequencies except mu and beta rhythms [17]. Filtration greatly improved the accuracies obtained. In this work, 10-30Hz frequency band was applied for all subjects and all approaches.

There exist important channels and un-important channels for classification, so, in our work, we used surface laplacian filtering around C3, C4 and Cz channels with eight channels around each one. The signals from C3, C4 and Cz were multiplied by 8 and subtracted by all surrounding eight signals. Three channels were common between channels surrounding C3 and Cz and others three channels were common between channels surrounding the channels C4 and Cz. Thus the number of non-recurring channels used for surface laplacian filter was 21. Here, we use this filter for all approaches except when extracting common spatial patterns because CSP requires a large number of channels to get better results [18].

C. Feature Extraction

There are several types of feature vectors which used for two class imagery task discrimination. For comparison, we have used the most popularly used types of feature extraction techniques. These methods are wavelet based statistical values, wavelet based energy, wavelet based entropy, wavelet based entropy-energy-band power, and CSP based variance.

1. *Discrete wavelet transform*

Fourier Transform is not suitable to extract features from such signals because the EEG signals are non-stationary signals. Short Time Fourier Transform (STFT) can be used to analyze non-stationary signals, but it offers a constant resolution at all frequencies. For analyzing different frequencies with different resolutions, Wavelet Transform (WT) which uses multi-resolution technique is used. In addition, WT can offer a smaller number of features for the signal to be processed; this means that it may be suitable relatively for avoiding on the curse dimensionality problem. Basically, wavelet transform analyzes the characteristics of signal in time and frequency domain by decomposing such signal into a several functions by a single function to generate shifting and detailing [19] [20]. This function is called mother function and known by

$$\psi(t) = \frac{1}{\sqrt{2}}\psi\left(\frac{t-b}{a}\right) \quad a, b \in R, a > 0 \tag{1}$$

where '*a*' and '*b*' are the scaling parameter and shifting parameter, respectively, and R is the wavelet space. The wavelet transform is given by

$$F(a,b) = \frac{1}{\sqrt{a}}\int_R \psi\left(\frac{t-b}{a}\right)dt \tag{2}$$

In this work, we employed discrete wavelet transform (DWT) because it provides highly efficient wavelet representation [19]. In first level decomposition, low pass and high pass filters are frequently employed to obtain the representation of digital signal as approximation (A1) and detail (D1) coefficients. The equation that defines DWT decomposition is as follows;

$$f(t) = \sum_{k=-\infty}^{k=+\infty} c_{n,k}\phi(2^{-n}t-k) + \sum_{k=-\infty}^{k=+\infty}\sum_{k=-\infty}^{k=+\infty} d_{j,k}2^{-j/2}\psi(2^{-j}t-k) \tag{3}$$

where $d_{j,k}$ and $c_{j,k}$ represent the approximation and detail coefficients, respectively, n is the level, and ∅ is the function of scale. The first approximation is more decomposed and the process is repeated for another time and so on. In this work, we chose 'haar' wavelet function and level is 3 because this gave best properties of signals features which classified successfully.

After obtaining on all approximation and detail coefficients (D1, D2, D3 and A3), different combination between these coefficients were tried to get the best result. However, the highest classification accuracy was obtained with all the coefficients. The features vectors were formed using the following techniques;

Let $S(n)$, and $n = 1,2,.....,N$, is a discrete signal, and N is sample number of the signal, then

Statistical Features:

- The mean of the signal:

$$\mu_S = \frac{1}{N}\sum_{n=1}^{N} S(n) \tag{4}$$

- The variance of the signal:

$$V_S = \frac{1}{N}\sum_{n=1}^{N}(S(n) - \mu_S) \tag{5}$$

- The standard deviation of the signal:

$$\sigma_S = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(S(n) - \mu_S)} \tag{6}$$

For three level wavelet coefficients, twelve features (3× (3+1)) were obtained for each channel. As previously mentioned, the selected channels were spatially filtered to produce 3 channels hold the most features of motor imagery movements. Subsequently, 36 features were obtained as a result of wavelet decomposition.

Entropy Features:

None normalized Shannon entropy -

$$Ent = \sum_{n=1}^{N}|S(n)|^2 \log|S(n)|^2 \tag{7}$$

Four features were obtained for each channel. Thus, 12 features were obtained as a result of wavelet decomposition.

Energy Features:

The energy of the signal-

$$Eng = \sum_{n=1}^{N}|S(n)|^2 \tag{8}$$

The number of features was 12 as obtained as a result of wavelet decomposition.

Logarithmic band power (LBP) Features:

Logarithmic band power of the signal-

$$Eng = \log(\frac{1}{N}\sum_{n=1}^{N}|S(n)|^2) \tag{9}$$

As a result of wavelet decomposition, 12 features were obtained.

Entropy-Energy-LBP Features:

Combining entropy, energy and LBP provides twelve features obtained for each channel. Thus, 36 features were obtained as a result of wavelet decomposition.

2. *Common spatial pattern*

In this method, all channels were used without Laplacian filter. We used *Common Spatial Pattern* (CSP) algorithm as spatial filter that lead to peak variances for the discrimination of two classes of EEG related right hand and foot motor imagery and to reduce the number of channels used [22]. A set of CSP filters are constructed by computing the projection matrix;

$$W_{CSP} = (w_1 \quad w_2 \quad w_{ch-1} \quad w_{ch}) \in R^{ch\times ch} \tag{10}$$

where first CSP filter w_1 provides the maximum variance of class 1 and the last CSP filter w_{ch} provides the maximum variance of class 2. In our work we selected first and last m filters such as

$$W_{CSP} = (w_1 \quad . \quad .w_m \quad w_{ch-m+1}. \quad w_{ch}) \in R^{ch\times ch} \tag{11}$$

The filtered signal is given by

$$S = W^{'}E \quad or$$
$$s(t) = W\ e(t) = (s_1(t) \quad .... \quad s_d(t))' \tag{12}$$

where d=2m and it is the number of channels that desired to reduce to. Thus, for each class EEG sample matrix we are going to select only small number of signals m that are most important for discrimination between the two classes. The feature vectors $f = (f_1, f_2, f_3 .... f_{2m})'$ can be calculated by the following equation;

$$f_i = \log\left[\frac{\mathrm{var}[\ s_i(t)]}{\sum_{i=1}^{2m}\mathrm{var}[\ s_i(t)]}\right] \tag{13}$$

Thus, d features were obtained as a result of common spatial filtering. Table I includes size of features vectors for all proposed approaches.

TABLE I. SIZE AND NUMBER OF FEATURE VECTORS FOR ALL APPROACHES.

Feature vectors	**Size (No. of trials x No. of Features per channel x No. of channels)**	
Laplacian filter	**yes**	**No**
No. of channels	-----------	**21channels**
WT+Mean+Var+St	280×12×3	280×12×21
WT+Entropy	280×04×3	280×04×21
WT+Energy	280×04×3	280×04×21
WT+LBP	280×04×3	280×04×21
WT+Eng+Ent+LPB	280×12×3	280×12×21
CSP	280×2m	------------

WT: wavelet transform; Var: variance; St: standard deviation; LBP: logarithmic band power; CSP: common spatial pattern.

D. *Classification and cross-validation*

There are several algorithms used for classification. In this work, we have employed the most popularly used types of classification techniques to classify the feature, Linear Discriminant Analysis (LDA) [23], Support Vector Machine (SVM) [24], and Artificial Neural Network (ANN). The aim was to compare between them and to conclude which one provides best results for two motor imagery tasks classification. In order to make linear SVM convergence fast, we have decreased parameter ‘BoxConstraint’ from 0.5 to 0.1. For ANN, we have changed the transfer functions of layer1 and layer2 to ‘satlins’ and ‘tansig’ respectively, the train function to ‘trainbr’ , train learning to 0.01 and others parameters as well. All these modification were done to obtain better results.

In our experiments, we used k-fold cross-validation to estimate the classification accuracy. In k-fold cross validation the dataset is randomly divided into k equal parts (k subsets) [25]. All the subsets are used to the training except one for the test (validation). The cross-validation is repeated k times (folds). Then the results of k times are averaged to produce a single classification rate. In this work, we used 5-fold cross validation (20% for testing and 80% for training). The classification accuracy can be given by

$$accuracy = \left(\frac{N_{correct}}{N_{total}}\right)\times 100\% \tag{14}$$

where N_{total} is the overall number of vectors to be classified and $N_{correct}$ is the number of the correct vectors.

III. RESULTS AND DISSCUSSION

As already mentioned, dataset (III-IVa) has five subjects with 280 trials for each one divided for training (224 trials) and testing (56 trials) data according to 5-fold cross validation. This dataset includes two mental tasks, right hand and right foot. Before feature extraction, the signals were segmented to time window of 3 s with offset of 0.5 s and band-pass filtered using Butterworth filter with 10-30Hz frequency band. Table II shows classification accuracy of each subject for all proposed features and classified using LDA. Surface laplacian filer was employed and its output was applied as input for wavelet transform. As shown in the table, wavelet-based LBP and wavelet-based energy-entropy-LBP approaches provide the better results than those obtained by wavelet-based-entropy, wavelet-based energy and wavelet-statistical values ones, with average classification accuracy of 82.57% and 82.43% respectively. Feature extraction based common spatial pattern and dimension equal to 2 provided good results, as well, with average classification accuracy of 82.36%. In order to show the effect of surface laplacian filtering, Table III shows the average classification accuracy of WT features using LDA with and without surface laplacian filter. It can be noticed that the best results were obtained when using surface laplacian filtering. Similarly, Table IV shows the average classification accuracy of CSP features using LDA with different values of dimensions. It can be noticed that the results improved especially at d=10 and d=22 with average classification accuracy of 84.79% and 84.00%, respectively. Fig. 2 shows the average accuracy for all approaches classified by LDA, SVM and ANN classifiers. It can be noticed that SVM provided the best results with WT features whereas LDA provided the best results with CSP features. In addition to the results obtained using ANN classifier were the worst whether with discrete wavelet transform or from common spatial pattern. The highest average classification accuracy obtained from combination of ANN classifier and wavelet-based band power was 80.21%.

TABLE II. CLASSIFICATION ACCURACY OF EACH APPROACH USING

Features	aa	al	av	aw	ay	Avg.
WT+Mean+ Var+St	61.07	97.14	65.71	87.14	90.00	**80.21**
WT+Entropy	65.00	91.43	64.29	92.50	89.4	**80.57**
WT+Energy	66.79	91.43	65.00	93.21	91.07	**81.50**
WT+LBP	67.14	97.50	64.29	91.43	92.50	**82.57**
WT+Eng+ Ent+LPB	64.29	97.86	66.07	91.07	92.86	**82.43**
CSP (d=2)	82.14	94.64	55.71	90.36	88.93	**82.36**

TABLE III. AVERAGE CLASSIFICATION ACCURACY OF WT FEATURES USING LDA WITH/WITHOUT SURFACE LAPLACIAN FILTER.

Features	Surface laplacian filter	Non Laplacian filter		
		C3&C4& Cz +18 ch.	C3&Cz&C 4 (only)	All channels
WT+Mean +Var+St	80.21	61.14	74.57	74.43
WT+Entropy	80.57	76.71	74.07	72.43
WT+Energy	81.50	78.71	75.36	72.57
WT+LBP	**82.57**	78.36	76.64	70.36
WT+Eng+ Ent+LPB	**82.43**	60.79	75.07	73.93

TABLE IV. AVERAGE CLASSIFICATION ACCURACY OF CSP FEATURES USING LDA DIFFERENT VALUES OF DIMENSIONS.

	d=2	d=4	d=6	d=10	d=20	d=22	d=118
CSP	82.36	81.36	81.79	**84.79**	83.79	**84.00**	75.79

IV. CONCLUSION AND FUTURE STUDY

In our work, we have focused on the classification of two types of motor imagery tasks (Right Hand and Foot). We used various feature extraction techniques on the EEG data using common spatial pattern and wavelet coefficients. In the case of wavelet features, statistical values, energy, entropy and band power were used to form the desired feature vectors. Three types of algorithms were employed for classification, linear discriminant analysis (LDA), support vector machines (SVM) and artificial neural network (ANN). We have used unified bandpass butterworth filter for all subjects during all proposed approaches. The aim was to compare between them and to recommend an appropriate combination for synchronous two-class motor-imagery. We notify that SVM is more appropriate with the features extracted from wavelet coefficients and combination of entropy-energy-band power, and LDA is more appropriate with common spatial pattern.

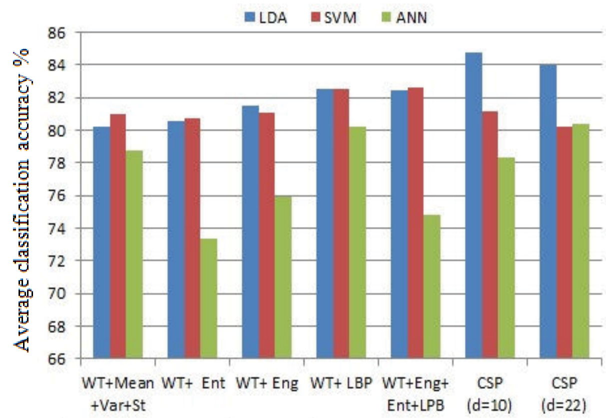

Fig. 2: Average accuracy of classification of all approaches using LDA, SVM and ANN classifiers

Overall, the results from CSP-LDA are the best from those obtained from WT-SVM. In other words, when using CSP for feature extraction, LDA is simpler and more suitable to use as classifier than SVM. In future work, a new data will be recorded from subjects and online training and testing will be studied.

ACKNOWLEDGMENT

The authors gratefully acknowledge the support from the King Saud University, Riyadh, Saudi Arabia.

REFERENCES

[1] J.P. Donoghue, "Connecting cortex to machines: recent advances in brain interfaces". Nat Neurosci. 5 (Suppl), 1085–1088, November 2002.

[2] Shih JJ, Krusienski DJ, Wolpaw JR. "Brain-Computer Interfaces in Medicine". Mayo Clinic Proceedings. 2012;87(3):268-279.

[3] Anderson R.A., Musallam S., Pesaran B., "Selecting the signals for a brain-machine interface", Curr Opin Neurobiol Vol.l4 (6), pp.720-726, December 2004.

[4] Saeid Sanei and J.A. Chambers, "EEG signal processing". Centre of Digital Signal Processing Cardiff University, UK, 2007.

[5] Wolpaw J. R., Birbaumer N., McFarland D. J., Pfurtscheller G., and Vaughan T. M. "Braincomputer interfaces for communication and control", Clinical Neurophysiology, 2002, 113: 767-791.

[6] Serby H., Yom-Tov E., and Inbar G. F., "An improved P300-based brain–computer interface", IEEE Trans. Neural Syst. Rehabil, Eng. 13, 89–98.

[7] Middendorf M., McMillan G., Calhoun G., and Jones K. S. (2000), "Brain–computer interfaces based on the steady-state visual-evoked response", IEEE Trans. Rehabil. Eng. 8, 211–214.

[8] Wang Y., Wang R., Gao X., Hong B., and Gao S. "A practical VEP-based braincomputer interface". IEEE Trans. Neural Syst. Rehabil, Eng. 14, 234–239 10.1109/tnsre.2006.875576.

[9] G. Pfurtscheller and F. H. Lopes da Silva, "Event-related EEG/MEG synchronization and desynchronization: Basic principles," Clin. Neurophysiol., vol. 110, no. 11, pp. 1842–1857, Nov. 1999.

[10] Wolpaw J.R. and McFarland D.J., "Control of a two-dimensional movement signal by a noninvasive brain-computer interface in humans", Proc. Natl. Acad. Sci. 2004;101:17849– 17854.

[11] Ridha Djemal, Ayad G. Bazyed, Kais Belwafi, Sofien Gannouni, and Walid Kaaniche, "Three-Class EEG-Based Motor Imagery

Classification Using Phase-Space Reconstruction Technique," *Brain Sci.* 2016, *6*(3), 36.

[12] Kais Belwafi1, Fakhreddine Ghaffari, Ridha Djemal, Olivier Romain, "A Hardware/Software Prototype of EEG-based BCI System for Home Device Control," J Sign Process Syst (2016).

[13] G. Pfurtscheller, J. Kalcher, C. Neuper, D. Flotzinger, and M. Pregenzer, "On-line EEG classification during externally-paced hand movements using a neural network-based classifier," Electroenceph. Clin. Neurophysiol., vol. 99, pp. 416–425, 1996.

[14] H. Ramoser, J. M. Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," IEEE Trans. Rehab. Eng., vol. 8, no.4, pp.441-446. 2000.

[15] A. G. Baziyad and R. Djemal, "A study and performance analysis of three paradigms of wavelet coefficients combinations in three-Class motor imagery based BCI," 2014 5th International Conference on Intelligent Systems, Modelling and Simulation, Langkawi, 2014, pp. 201-205.

[16] R. Chatterjee and T. Bandyopadhyay, "EEG based motor imagery classification using SVM and MLP," 2nd International Conference on Computational Intelligence and Networks (CINE), At Bhubaneswar, Volume: 2, 10.1109/CINE.2016.22.

[17] X. Perrin, "Semi-autonomous navigation of an assistive robot using low throughput interfaces," Ph.D. dissertation, ETHZ, Zurich, Switzerland, 2009.

[18] G. Townsend, B. Graimann, and G. Pfurtscheller, "A comparison of common spatial patterns with complex band power features in a four-class BCI experiment," Biomedical Engineering, IEEE Transactions on, vol. 53, pp. 642-651, 2006.

[19] V. Bostanov, "BCI competition 2003-data sets Ib and IIb: feature extraction from event-related brain potentials with the continuous wavelet transform and the t-value scalogram," Biomedical Engineering, IEEE Transactions on, vol. 51, pp. 1057-1061, 2004.

[20] Y. Yong, N. Hurley, and G. Silvestre, "Single-trial EEG classification for brain-computer interface using wavelet decomposition," in European Signal Processing Conference, EUSIPCO 2005, 2005.

[21] V. J. Samar, A. Bopardikar, R. Rao, and K. Swartz, "Wavelet analysis of neuroelectric waveforms: a conceptual tutorial," Brain and language, vol. 66, pp. 7-60, 1999.

[22] Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, and Klaus-Robert Müller," Optimizing spatial filters for robust EEG single-trial analysis" IEEE Signal Processing Magazine 25 (1): 41-56. 2008.

[23] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern classification: John Wiley & Sons", 2012.

[24] C. J. Burges, "A tutorial on support vector machines for pattern recognition," Data mining and knowledge discovery, vol. 2, pp. 121-167, 1998.

[25] P. Refaeilzadeh, L. Tang, and H. Liu, "Cross-validation," in Encyclopedia of database system, ed: Springer, 2009, pp. 532-538.