Feature extraction schemes for BCI systems

L. Vega-Escobar, A.E. Castro-Ospina
Research Center of the Instituto Tecnológico Metropolitano
Calle 73 No 76A-354, Medellín, Colombia
{lauravega,andrescastro}@itm.edu.co

L. Duque-Muñoz
Universidad de Antioquia
Calle 70 No. 52 - 21, Medellín, Colombia
leonardo.duquem@udea.edu.co

Abstract

Brain-computer Interfaces (BCIs) are control and communication systems based on acquisition and processing of brain signals to control a computer or an external device. Usually, BCI is focused in recognizing acquired events by different neuroimage methods, but the most used is the electroencephalography (EEG). Feature extraction over EEG signals for BCI systems is crucial to the classification performance. In this paper a comparison between various methods for feature extraction of EEG signals in BCI systems is presented. Different methodologies were taken into account, both at extracting frequency information from each electrode as for extracting shared information between electrodes.

1. Introduction

A brain-computer interface (BCI) is a hardware and software communication system that allows cerebral activity to control computers or external devices. The main aim is to help people with extreme disabilities, such as amyotrophic lateral sclerosis, cerebral palsy, ischemic stroke or spinal cord injury disorders [1]. BCI system translates the brain activity into commands for a computer application or other devices, such a spelling system or a robotic limb. To meet this aim, many aspects of BCI systems have been investigated. Research areas include evaluation of invasive and non-invasive technologies to measure brain-activity, development of new BCI applications, evaluation of control-signals (*i.e*. patterns of brain-activity that can be used for communication), development of algorithms for translation of brain-signals into computer commands, and the development and evaluation of BCI systems specifically for disabled subjects [2]. Most common non-invasive BCI systems are based on neuroimaging technologies, such as Magnetoencephalography (MEG) and Magnetic resonance imaging (MRI). However, electroencephalography (EEG) is the most used, mainly due to its fine temporal resolution and lower cost [3].

One of the most important components of a BCI system is the EEG signal feature extraction procedure, due to its role on the proper performance of the classification stage at discriminating mental states.

In [4] a classification of imagined right/left hand movements was presented. Features were extracted by means of the wavelet transform and classified with support vector machines (SVM). Tests results showed that the proposed method could accurately extract EEG substantial features and provide an effective means to classify the motor mental tasks. Similar work was done in [5], where a new method for classifying the EEG signals from the BCI Competition 2003 was presented. Wavelet coefficients and power spectral density (PSD) are combined as future vector for further classification step. It concludes the importance of using a feature vector.

In [6] was proposed a feature extraction method to discriminate hand movements, based on the processing of EEG signals recorded. from two subjects with 10 electrodes. Features are extracted from the raw data by means of PSD, and alpha and beta band power. Even though results shows the discrimination ability given by such features, it is worth to note the dependent variability of each subject, also it is important to determine the most prominent electrodes for each task. Another feature extraction methodology was proposed in [7]. The rhythmic component extraction (RCE) was applied to characterize left/right hand motor imagery. The EEG signals was recorded from two subjects with 14 electrodes. The experiment shows that the combination of RCE and fisher discriminant analysis on the 12-15 Hz frequency

978-1-4673-9461-1/15/$31.00

band performed slightly better than other methods.

Although the results found in the State of the art are appropriate, in most cases the shared information between different electrodes is not taken into account. This information is useful because allows to extract new relevant features with small amount of electrodes, and comparable classification results.

Due to the high importance of a suitable feature extraction stage in any classification system, in this paper a comparison between different characterization methodologies of EEG signals for BCI systems is presented. By the extraction of frequency information by different approaches in each electrode as well as shared information among electrodes. Methodologies are tested in the data-set BCI competition 2003, which has motor imagination data of left and right hand with 140 recordings with 3 electrodes and 1152 samples.

2. Materials and Methods

2.1. Discrete Wavelet Transform

Discrete Wavelet Transform (DWT) decomposes a signal into approach and detail components, allowing to analyze the signal in different scale bands (frequency bands), with different resolutions. Decomposition is made by employing two sets of functions, called scaling and wavelet functions, which are associated with low-pass filters and high-pass filters, respectively [8].

2.2. Phase Synchronization Measure

Let $x_i(t)$ and $x_j(t)$ be two time signals, *e.g.* two different electrodes from an EEG measure. Among the different methods to measure synchronization between two EEG signals, the most commonly used is the classical coherence $Coh_{ij}(f)$. Nonetheless, such measure is amplitude-dependent, leading to... Another measure termed the Phase Locking Value (PLV) has recently emerged which only takes into account the phases of the two signals and is defined as:

$$PLV = \frac{1}{N}\left|\sum_{t=1}^{N}\exp\left(j\left\{\Phi_i\left(t\right)-\Phi_j\left(t\right)\right\}\right)\right| \tag{1}$$

where N is the total number of time samples and $\Phi_i\left(t\right)$ is the instantaneous phase of electrode i at time instant t calculated as:

$$\Phi_i\left(t\right) = \arctan\frac{\tilde{x}_i\left(t\right)}{x_i\left(t\right)} \tag{2}$$

with $\tilde{x}_i\left(t\right)$ representing the Hilbert transform:

$$\tilde{x}_i\left(t\right) = \frac{1}{\pi}PV\int_{-\infty}^{\infty}\frac{x_{i(\tau)}}{t-\tau}d\tau \tag{3}$$

here, PV denotes that the integral is taken in the sense of Cauchy principal value, and the instantaneous phase is calculated as follows:

PLV measure in Eq. (1) ranges between 0 and 1. If the phase differences are randomly distributed over $[0, 2\pi]$ the PLV value will be 0. On the other hand, if there is a constant phase difference then the PLV value will be 1, meaning a perfect phase synchronization.

2.3. Power spectral density(PSD)

Power spectral density (PSD), describes how to the power of a signal is distributed in frequency. Since signal with nonzero average power is not square integrable, the Fourier transforms do not exist in this case. The PSD is the Fourier transform of the autocorrelation function of the signal.

The power of a signal in a given frequency band is calculated by integrating over positive and negative frequencies. The definition of power spectral density generalizes in a straight manner to finite time series with $1 \leq n \leq N$,such as signal sampled at discrete times $xn = x\left(n\Delta t\right)$ for a total measurement period $T = N\Delta t$ [9].

$$S\left(e^{jw}\right) = \frac{1}{2\pi N}\left|\sum_{n=1}^{N}x_n e^{-jwn}\right|^2 \tag{4}$$

2.4. Support Vector Machine (SVM)

Support vector machines (SVMs) are binary classifiers. Ideally, the SVM should produce a hyperplane which discriminates in a proper manner the studied data into its classes. However, perfect separation is not always possible and, if so, the outcome of the model can not be generalized to other data. This is known as overfitting. To allow flexibility, the SVM handle a C parameter that controls the trade-off between training errors and rigid margins, creating a soft margin which allows some misclassification Eq. (5).

$$\tau\left(\boldsymbol{w}\right) = \text{minimize}\frac{1}{2}\|\boldsymbol{w}\|^2 + C\sum\nolimits_{i=1}^{n}\varepsilon_i \tag{5}$$

where $\boldsymbol{w}$ is a normal vector to the hyperplane and ε_i refers to a slack variable and represents the penalty for i-th example not satisfying the margin constraint. More details on Soft Margin SVM can be found in [10].

Some data could be distributed in such a way that there is no linear separation between classes, therefore, a mapping of the data to a higher-dimensional space is needed, where a linear separation is possible. By using the *kernel trick* an explicit mapping to the higher-dimensional space not need

to be done, since it returns a measure of similarity $(k(x, y))$ in such space. The Gaussian kernel is an example of radial basis function kernel, expressed as:

$$k(x, y) = \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) \quad (6)$$

where σ is the so-called Gaussian band-width parameter, which must be carefully adjusted for the performance of the kernel, if overrated, exponential and behave almost linearly higher-dimensional projection will start to lose its nonlinear power. On the other hand, if understated, lack of regularization function and the decision boundary will be very sensitive to noise in the training data.

To calculate the performance of classifiers overall accuracy, sensitivity and specificity are used, implementing confusion matrices, which can distinguish between true class and the class predicted by the classifier using tags. You have four possible cases: If an instance is positive and is classified as positive, will be called true positive (TP), if it is rated as negative, it will be counted as false positive (FP). If the instance in negative and is classified as negative, it will be counted as true negative (TN) and if it is classified as positive, it will be counted as a false negative (TN).

	True	False	
Positive	TP	FP	Positive Predicted Values
Negative	FN	TN	Negative Predicted Values
	Sensitivity	Specificity	Accuracy

Table 1. Construction of an array of confusion for binary classification

According to Table 1, we can obtain the following equations;

$$Sensitivity = \frac{TP}{(TP + FN)} \quad (7)$$

$$Specificity = \frac{TN}{(TN + FP)} \quad (8)$$

$$Accuracy = \frac{(TP + TN)}{(TP + FN + TN + FP)} \quad (9)$$

The accuracy will be the standard measure used to evaluate the performance of classifiers, as it allows to know that the measured value is so close to the real value. In more practical terms, having an accuracy of 100% means that the original tags are exactly equal to those obtained by the algorithm of classification.

3. Experimental Framework

The methodology used in this work is depicted in Figure 1 and consists of three stages, namely data preprocessing and decomposition; feature extraction, dimension reduction and classification. A brief description of the considered dataset is presented.

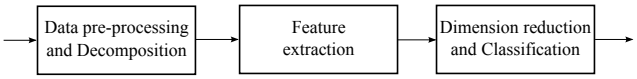


Figure 1. Block diagram of experimental framework

3.1. Dataset

The used Dataset was provided by Department of Medical Informatics, Institute for Biomedical Engineering, University of Technology Graz [11]. It was recorded from a normal subject *(female, 25 yrs)* during a feedback session. The subject sat in a relaxing chair with armrests. The task was to control a feedback bar by means of imagery left or right hand movements. The order of left and right cues was random. Figure 2 shows the timing of the experiment. The first $2s$ was quite, at $t = 2s$ an acoustic stimulus indicates the beginning of the trial, the trigger channel $(\#4)$ went from low to high, and a cross ("+") was displayed for $1s$; then at $t = 3s$, an arrow (left or right) was displayed as cue. At the same time the subject was asked to move a bar into the direction of the cue. The feedback was based on AAR parameters of channel $\#1$ $(C3)$ and $\#3$ $(C4)$, the AAR parameters were combined with a discriminant analysis into one output parameter.

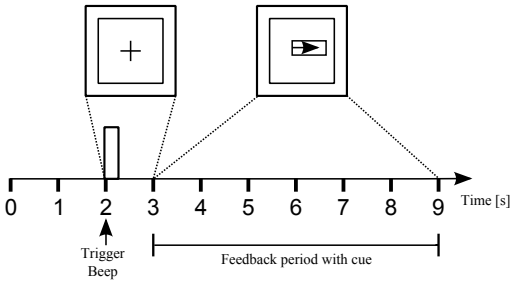


Figure 2. Timing of the experiment

The recording was made using a G.tec amplifier and a Ag/AgC1 electrodes. Three bipolar EEG channels channels (anterior '+ ', posterior '-') were measured over $C3$, Cz and $C4$ as depicted in Figure 3. The EEG was sampled with 128Hz, it was filtered between 0.5 and 30Hz. The experiment consists of 7 runs with 40 trials each. All runs were conducted on the same day with several minutes break in between. Since one half of the dataset are provided for training there are 140 trials of $9s$ length.

3.2. Data pre-processing and decomposition

Given that the first $3s$ are irrelevant (quiet state and cross displaying), EEG signals are used from second 3rd to 9th. Moreover, to decompose EEG signals given by each electrode into frequency bands, discrete Wavelet transform is

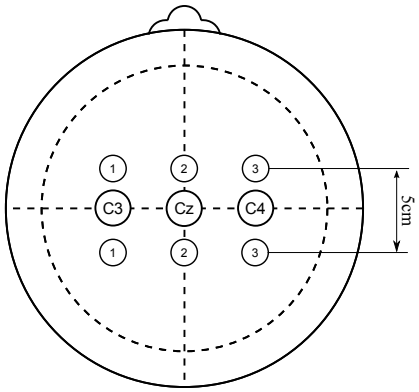


Figure 3. Electrode placement

employed. The number of levels of decomposition were chosen on the basis of the dominant frequency components of the signal. According to imagined right/left hand movements. The mother wavelet Daubechies 4 with 5 leveles was chosen [12]. As a result, the EEG signal is decomposed into the details D2-D5 and approximation A5 for each of the three electrodes. The ranges of different frequency band are shown in Table 2.

Decomposed signal	Frequency range (Hz)	Rhythms
A5	0 – 2	Delta (δ)
D5	2 – 4	Delta (δ)
D4	4 – 8	Theta (θ)
D3	8 – 16	Alpha (α)
D2	16 – 32	Beta (β)

Table 2. Frequencies corresponding to different levels of decomposition for Daubechies order 4

3.3. Feature Extraction

To compute a feature vector from each of the EEG signals recorded by electrodes located at $C3$, $C4$ and Cz, the following strategies are used. It is worth to note that the features for each electrode are concatenated into a single vector, for each trial, leading to a feature matrix of dimension $140 \times p$, where p is the corresponding feature dimension.

- **Statistics:** From each wavelet decomposition level, the three first moments are calculated, *i.e*. mean, variance and kurtosis.
- **Phase Locking Value:** To relate the information and extract shared features between adjacent electrodes we used phase locking value (PLV) measure. PLV method measures the synchronization between the EEG signals from a pair of sensors at a frequency of interest. A Windowing of 1 second is used to relate each pair of electrodes. There are only three electrode pair combination, since there are only three electrodes.
- **Power Spectral Density over raw data:** PSD features are computed over the entire EEG recording for each electrode, *i.e*. without decomposition, as usually done in the state of the art methods.
- **Power Spectral Density:** Finally, PSD features are calculated from each band previously computed by means of the DWT, we aim to demonstrate that such decomposition and further characterization enhances the classification accuracy.

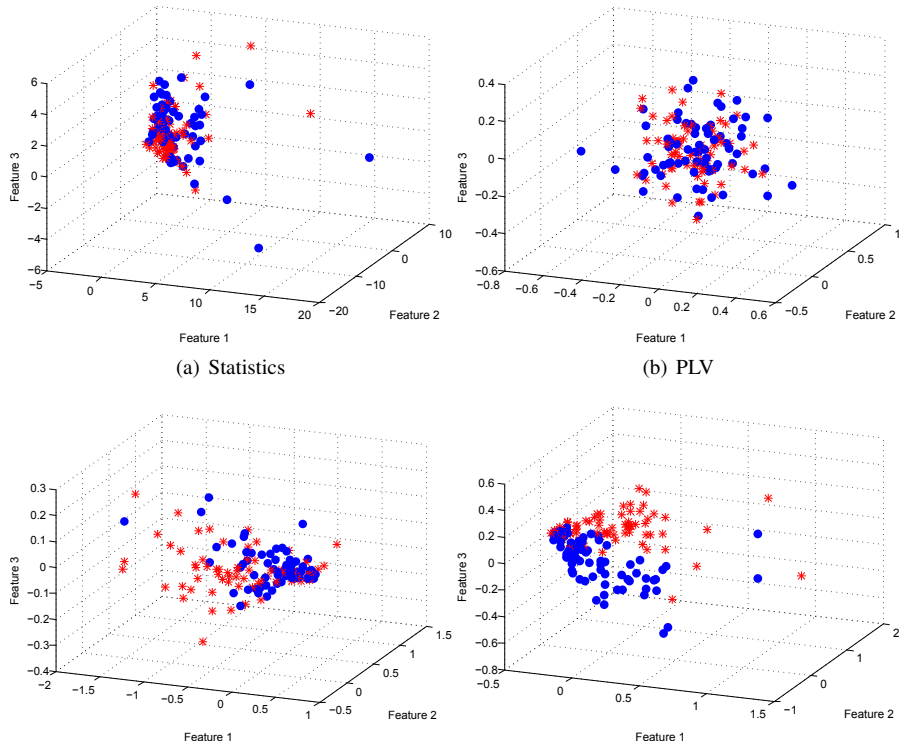
3.4. Classification

Due to the high amount of features, principal component analysis (PCA) technique is used to reduce data dimension. This is done for the purpose of provide better classification accuracy and reduce the computational cost. PCA is set to retain 99% of data variance. It is worth to note that dimension reduction is done for all the characterization techniques described before. Besides, an SVM is used for the classification stage. Since it has several parameters, a suitable tuning must be done. Thus, a two-dimensional exploration for all the possible values of the SVM trade-off constant C and the kernel band-width σ is carried out by means of a Particle Swarm Optimization (PSO) meta-heuristic [13]. To avoid over-training of the models, a cross-validation of ten folds is performed. The limits of the search space were defined as (10^{-3},10^{4}) for σ and (1,10^{6}) for C. Additionally, the number of particles for the search was set to 10, while the maximum number of iterations was set to 20.

4. Results and Discussion

Figure 4 shows the scatterplots for each of the considered feature extraction methods. As depicted in Figures 4(a) to 4(b) both classes are strongly overlapped for the cases of statistics and PLV features, yielding to low accuracy results. Such behavior demonstrates that a more complex characterization than merely statistics over the wavelet decomposition signals is required. For the case of PLV results are inconclusive as it is shown how the shared information among the electrodes is not enough, this could be explained by the low number of electrodes used for this test, i.e. three, which could not generate an activation map. Moreover, feature extraction made by PSD over raw data gives a better representation of both classes, however, there are still some overlapping areas as shown in Figure 4(c). Finally, for the case of PSD features over DWT decomposition levels, in Figure 4(d) it can be seen how both classes has fewer overlapping regions, which is suitable for the classification stage.

Table 3 shows the classification accuracy results for the considered feature extraction techniques. Classification of the features computed by statistics values over the DWT decomposition levels presents high deviation and a low accuracy. The features that uses the time dynamics of the EEG



(a) Statistics (b) PLV

(c) PSD for raw EEG (d) PSD

Figure 4. Scatterplots for considered feature extraction methods. • right hand ∗ left hand.

signal measures the temporal variations of the signals; these are particularly adapted to describe the neurophysiological signals with a precise compass and specific time. The compromise between the computational complexity and speed of calculation, in addition to the growing demand of BCI systems in real time, exposes the need for methods with less possible complexity [14]. In this sense statistical measures over time are a good choice, because its low computational complexity and its possible evaluation online. However, it is not possible to capture the complex brain dynamics measured with statistical features in different time windows, resulting in low classification rates. For the PLV values among pair of electrodes, a poor accuracy is reached, such behavior can be explained because of the closeness of used electrodes, which are measuring the cortex area, therefore, similar activity. Table 4 summarizes the PLV values for EEG segments computed at different windowing values, such comparison is done between electrode 1 and 2. It can be seen how, despite of the window length, for both classes the mean and standard deviation of the PLV values are almost equal, avoiding the discrimination of classes. Therefore, it is needed a higher amount of electrodes in order to have a set from which it can be selected the most important ones or use a different electrode location which measures different brain activity.

Finally, as shown in Table 3, it is worth to note how, for the PSD-based features classification, the PSD values over the raw data performs poorly and more unstable that the PSD values computed over the DWT decomposition levels. Features based upon a combination of different feature types are becoming increasingly popular in BCI research. It is likely that the recent development of BCIs that combine two or more features, *e.g.*, wavelets and PSD, performs a better and more stable classification accuracy, sensitivity and specificity.

	Accuracy	Sensitivity	Specificity
Statistics	0.65 ± 0.125	0.7 ± 0.184	0.614 ± 0.203
PLV	0.54 ± 0.096	0.5 ± 0.205	0.59 ± 0.218
PSD over raw EEG	0.7 ± 0.174	0.6 ± 0.188	0.8 ± 0.225
PSD with DWT	$\mathbf{0.85 \pm 0.103}$	$\mathbf{0.9 \pm 0.118}$	$\mathbf{0.8 \pm 0.154}$

Table 3. Mean values and standard deviation for classification performance

5. Conclusions

A feature extraction comparison for motor imaginery tasks was presented. Features extracted from statistical moments of DWT coefficients are not enough for this type of task. Moreover, it was found that shared information between electrodes located close to each other is almost identically, therefore such features were not useful to enhance

Window										
Segment	0.5s		1s		2s		3s		6s	
	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2	Class 1	Class 2
1	0.4672 ± 0.1248	$0,4537 \pm 0.1388$	0.4776 ± 0.1606	0.4499 ± 0.1526	0.4978 ± 0.1939	0.4922 ± 0.1738	0.4727 ± 0.1822	0.4537 ± 0.2060	0.5154 ± 0.1478	0.4913 ± 0.1494
2	0.4571 ± 0.1399	0.4534 ± 0.1563	0.4825 ± 0.1882	0.4599 ± 0.1646	0.4954 ± 0.1919	0.4995 ± 0.1825	0.4974 ± 0.1939	0.4961 ± 0.2050		
3	0.4440 ± 0.1456	0.4533 ± 0.1374	0.4945 ± 0.1581	0.4659 ± 0.1460	0.4934 ± 0.2016	0.4853 ± 0.1990				
4	0.4634 ± 0.1422	0.4614 ± 0.1415	0.4939 ± 0.1847	0.4629 ± 0.1686						
5	0.4748 ± 0.1245	0.4626 ± 0.1321	0.4902 ± 0.1870	0.4520 ± 0.1648						
6	0.4590 ± 0.1446	0.4365 ± 0.1306	0.4879 ± 0.1785	0.4694 ± 0.1580						
7	0.4715 ± 0.1530	0.4516 ± 0.1333								
8	0.4622 ± 0.1387	0.4411 ± 0.1426								
9	0.4541 ± 0.1481	0.4377 ± 0.1358								
10	0.4651 ± 0.1428	0.4427 ± 0.1458								
11	0.4631 ± 0.1465	0.4460 ± 0.1353								
12	0.4643 ± 0.1447	0.4542 ± 0.1421								

Table 4. Mean and standard deviation for windowed PLV values between electrode 1 and 2

the classification performance. For the case of PSD-based feature techniques, it was shown how its use over each electrode raw data does not allow a suitable discrimination of classes, however, with a previous frequency band decomposition, by means of DWT, a proper discrimination among classes is achieved, which were reflected in terms of accuracy.

As future work, it would be interesting to test synchronization measures over datasets with more electrodes and other feature extraction techniques.

Acknowledgments

This work is carried out under grants provided by *Programa Nacional de Jóvenes Investigadores e Innovadores - 2014 (COLCIENCIAS)*, *Convocatoria de Doctorados nacionales 647, colciencias 2014* and the support of *Instituto Tecnológico Metropolitano (ITM)*.

References

[1] N. Birbaumer, "Breaking the silence: brain–computer interfaces (bci) for communication and motor control," *Psychophysiology*, vol. 43, no. 6, pp. 517–532, 2006.

[2] U. Hoffmann, J.-M. Vesin, and T. Ebrahimi, "Recent advances in brain-computer interfaces," in *IEEE International Workshop on Multimedia Signal Processing (MMSP07)*, no. LTS-CONF-2007-063, 2007.

[3] B. He, S. Gao, H. Yuan, and J. R. Wolpaw, "Brain–computer interfaces," in *Neural Engineering*. Springer, 2013, pp. 87–151.

[4] X. Qiao, Y. Wang, D. Li, and L. Tian, "Feature extraction and classifier evaluation of eeg for imaginary hand movements," in *Natural Computation (ICNC), 2010 Sixth International Conference on*, vol. 4. IEEE, 2010, pp. 2112–2116.

[5] H. Xu, J. Lou, R. Su, and E. Zhang, "Feature extraction and classification of eeg for imaging left-right hands movement," in *Computer Science and Information Technology, 2009. ICCSIT 2009. 2nd IEEE International Conference on*. IEEE, 2009, pp. 56–59.

[6] N. G. Ozmen and L. Gumusel, "Classification of real and imaginary hand movements for a bci design," in *Telecommunications and Signal Processing (TSP), 2013 36th International Conference on*. IEEE, 2013, pp. 607–611.

[7] H. Higashi, T. Tanaka, and A. Funase, "Classification of single trial eeg during imagined hand movement by rhythmic component extraction," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. IEEE, 2009, pp. 2482–2485.

[8] C. Parameswariah and M. Cox, "Frequency characteristics of wavelets," *Power Delivery, IEEE Transactions on*, vol. 17, no. 3, pp. 800–804, 2002.

[9] S. Unde, R. Shriram *et al.*, "Coherence analysis of eeg signal using power spectral density," in *Communication Systems and Network Technologies (CSNT), 2014 Fourth International Conference on*. IEEE, 2014, pp. 871–874.

[10] H. Tjandrasa, R. E. Putra, A. Y. Wijaya, and I. Arieshanti, "Classification of non-proliferative diabetic retinopathy based on hard exudates using soft margin svm," in *Control System, Computing and Engineering (ICCSCE), 2013 IEEE International Conference on*. IEEE, 2013, pp. 376–380.

[11] B. Blankertz, K.-R. Müller, G. Curio, T. M. Vaughan, G. Schalk, J. R. Wolpaw, A. Schlögl, C. Neuper, G. Pfurtscheller, T. Hinterberger *et al.*, "The bci competition 2003: progress and perspectives in detection and discrimination of eeg single trials," *Biomedical Engineering, IEEE Transactions on*, vol. 51, no. 6, pp. 1044–1051, 2004.

[12] T. Kalayci and Ö. Özdamar, "Wavelet preprocessing for automated neural network detection of eeg spikes," *Engineering in Medicine and Biology Magazine, IEEE*, vol. 14, no. 2, pp. 160–166, 1995.

[13] C. Bendtsen, "pso: Particle swarm optimization," *R package version*, vol. 1, no. 3, 2012.

[14] A. Khorshidtalab, M. J. E. Salami, and M. Hamedi, "Robust classification of motor imagery eeg signals using statistical time–domain features," *Physiological measurement*, vol. 34, no. 11, p. 1563, 2013.