

Research Article

Open Access

Germán Rodríguez-Bermúdez*, Andrés Bueno-Crespo, and F.José Martínez-Albaladejo
Classifying BCI signals from novice users with extreme learning machine

DOI 10.1515/phys-2017-0056

Received Feb 14, 2017; accepted Apr 28, 2017

Abstract: Brain computer interface (BCI) allows to control external devices only with the electrical activity of the brain. In order to improve the system, several approaches have been proposed. However it is usual to test algorithms with standard BCI signals from experts users or from repositories available on Internet. In this work, extreme learning machine (ELM) has been tested with signals from 5 novel users to compare with standard classification algorithms. Experimental results show that ELM is a suitable method to classify electroencephalogram signals from novice users.

Keywords: Brain computer interface, motor imagery, extreme learning machine, novice users

PACS: 87.19.le, 87.85.dd, 87.85.dq

1 Introduction

The electroencephalogram (EEG) measures and records the electrical activity generated by brain structures. When brain cells (neurons) are activated, local current flows are produced therefore using electrodes placed on the scalp, the electrical activity of the brain can be collected and saved. In this way, electroencephalographic reading is a completely non-invasive procedure that can be applied repeatedly to patients, normal adults, and children with virtually no risk or limitation [14]. As it is widely accepted [5, 6], EEG signal is considered a chaotic dynamical sys-

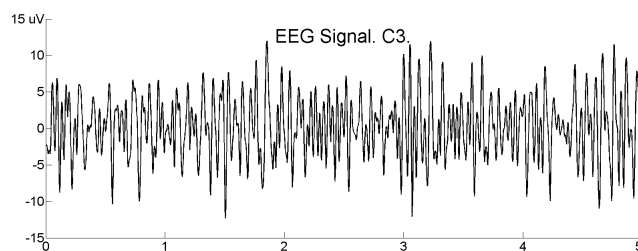


Figure 1: Five seconds of EEG signal in a Motor Imagery Brain Computer Interface experiment from channel C3. (X axis in seconds and Y axis in microvolts)

tem. EEG signal is chaotic because its amplitude changes randomly with respect to time, see Figure 1 [4, 12, 13].

The research on EEG signals offers an excellent opportunity to improve the understanding of the brain function. Thus the analysis of the EEG signal has been used to detect abnormal functions of the brain like disorders such as autism, dementia, schizophrenic, depression etc... [8, 19–21]. Besides medical applications, the information extracted in this type of analysis has been used for non-medical purposes such as mental fatigue, emotion recognition or BCI.

A BCI offers an alternative to natural communication and control. A BCI is an artificial system that bypasses the body's normal afferent pathways, which are the neuromuscular output channels [1]. This means that there is direct communication between the brain signal and the computer. A general scheme of a motor imagery (MI) BCI system could be represented by 5 main steps, as presented in Figure 2. All BCI systems start with a brain signal, thus it is crucial to acquire the brain activity and it is necessary to place electrodes on the scalp in the most suitable areas for each experiment. In a first stage, the user must develop a predefined mental task (motor imagery). The EEG signal changes according to the task and these signals are acquired by electrodes and transmitted to the next step. In step 2, the EEG signal is digitized, amplified and filtered in order to delete undesired signals called artifacts. The artifacts could be bio-signals such as heart beating or breath, or noise such as a power line. Then, the clean signal goes to step 3 where the system applies several techniques in order to compute features to be used as a descriptors of

*Corresponding Author: **Germán Rodríguez-Bermúdez:** Centro Universitario de la Defensa de San Javier (University Centre of Defence at the Spanish Air Force Academy), Spain; Email: german.rodriguez@tud.upct.es

Andrés Bueno-Crespo: Bioinformatics and High Performance Computing Research Group (BIO-HPC), Universidad Católica de Murcia, Spain; Email: abueno@ucam.edu

F.José Martínez-Albaladejo: Telecommunications Department, Universidad Católica de Murcia, Spain; Email: fjmartinez@ucam.edu



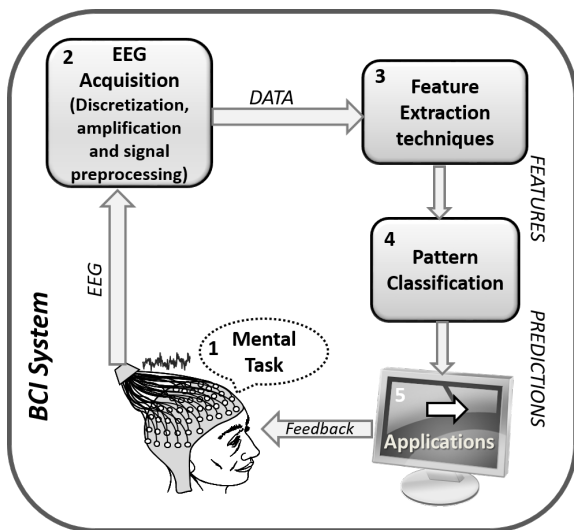


Figure 2: Generic Scheme of a BCI system compound by 5 main blocks: 1) Mental task; 2) EEG acquisition; 3) Feature Extraction techniques; 4) Pattern Classification; 5) Applications

the EEG signals for the pattern recognition machine. The Features are classified in step 4, Pattern Classification, and the predictions are used or presented in step 5. Finally, the user observing the application gets feedback from the system and can modify his EEG signals to improve the performance of the BCI system.

Although there are several studies about BCI systems it is usual that every one selects one or two feature extraction techniques for step 3, and one or more from step 4 due to the lack of a unique way to implement BCI systems. In addition, it is common that researchers use signals from repository data available on Internet such as BCI competitions I-IV or from expert users. It is well known that a BCI system requires the adaptation of the user to the system, so it is necessary to train the user in order to get good results [15–17]. In a MI BCI experiment an accuracy between 80% and 90% is expected after 6-9 training sessions of 20 minutes [18]. Nevertheless, and according to the state-of-the-art, certain subjects may face difficulties to use MI-based BCI systems and, in these cases, the classification performances are quite poor even using multiple training sessions [11]. Therefore, it is expected a previous selection of subjects with good classification performances in the experiments. In contrast, in this work the response of a system with EEG signals from 5 novel users without previous selection will be analyzed. We will test power spectrum density (PSD) as a feature extraction method combined with three classifiers: extreme learning machine (ELM); linear discriminant analysis (LDA) and support vector machine (SVM) in order to know which is the combination that reaches the highest classification accuracy.

Extreme learning machine (ELM) provides a fast and efficient multilayer perceptron (MLP) training [28]. Although Huang formalized the idea of ELM algorithm [25, 26], it was previously analyzed in other works [23, 24]. Huang demonstrated that the ELM is a universal approximator for a wide range of random computational nodes. ELM has been used for BCI systems, using it in its classic form, in voting optimized strategy, based on weighted probabilistic model, with adaptive extreme learning machine, and other variants [37–40]. However, it has not been tested with novice users.

The main contribution of this work is to test the ELM suitability to classify EEG signals from novice subjects for BCI systems, which until now had not been used with this type of users.

The paper is organized as follows: Section 2 describes the acquisition of the EEG signal, the Feature extraction and, ELM; Section 3 describes the experimental work; Section 4 presents and discuss the experimental results; and in Section 5, the main concluding remarks are stated. Finally, the acknowledgements to the funding organizations and late Master are stated in this article.

2 Material and methods

2.1 EEG signal generation and acquisition

For this work we have used a dataset generated in the University Centre of Defence at the Spanish Air Force Academy, Spain. The dataset is composed by EEG signals from 5 right-handed males volunteers (21, 30, 30, 33, 33 years old) with normal vision implementing a Motor Imagery task. They have imagined the right and left hand movement according to a predefined timing. It is important to remark that none of these subjects have used a BCI system before the current experiment. It entails that there has been no a previous selection of the users, therefore, low classification results are expected in comparison with expert users.

The EEG signals have been captured with a g.USBamp (g.tec Medical Engineering GmbH, Austria) and two bipolar channels (C3-C4) thus passive electrodes were located at FC3, CP3, FC4 and CP4 positions. In this case Ag/AgCl passive electrodes were used. The amplifier was configured to acquired the data at 256Hz with 8 bits. A band pass filter was also applied between 0.5 and 30 Hz to erase artifacts. The users were comfortably seated in front of a screen to do the experiments. They must be relaxed and imagine the movement of the right or left hand looking at

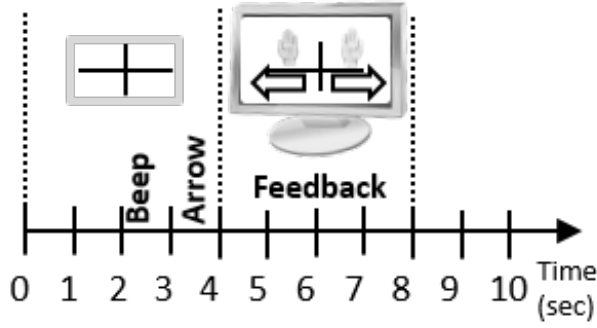


Figure 3: Scheme of BCI paradigm applied

the screen when the BCI system requires it. Each experiment was composed by 40 trials (20 left and 20 right hand).

At the beginning of each trial (see Figure 3) the screen is black, then a cross appears and in second 2 an acoustic *beep* signal sounds *beep* to attract the user's attention. At the third second of the trial the system presents an arrow pointing to the left or right for 1.25 s and the subject must imagine the movement of the right or left hand. Along 4 s the user must imagine the movement, therefore each trial needs 8 s. The system does not provide feedback to the users. Finally, it is important to remark that there is a random time between trials from 0.5 to 2.5 s to avoid adaptation. The system does not offer feedback to the users.

2.2 Feature extraction methods

In BCI, researchers have applied different techniques in step 3 [4, 10, 11]. Power spectrum density (PSD) has been computed as a feature extraction approach to test the ELM as a suitable method to be implemented in BCI systems to classify EEG signals from the first session of novice users. It is important to remark that PSD is a standard method for these systems [8]. Below there is a brief explanation about the algorithm used in this work.

2.2.1 Power spectral density

It is usual to decomposed EEG signals in four bands called α , β , θ and δ [2, 3]. However it is usual that researchers use only two bands in motor imagery BCI systems, α (8-13 Hz) and β (13-30Hz) [7, 10]. α waves are rhythmical waves and they are found in the EEGs of most adults when they are awake. When the awake person's attention is directed to some specific type of mental activity, the α waves are

replaced by higher frequency β waves. θ waves (4-8 Hz) normally occur in parietal and temporal regions in children, but they also appear during emotional stress in some adults [3].

It is important to note that the researchers only calculate α and β bands because they have the most discriminative information for the BCI experiments [7]. For computing these energies, the first step is to obtain the fast Fourier transform (FFT) of the EEG signals. Then, the corresponding coefficients of the α and β bands are added. Finally, the power of this addition is obtained by:

$$P_{\alpha} = \frac{FFT^*(EEG_{\alpha}) FFT(EEG_{\alpha})}{p^2}; \quad (1)$$

$$P_{\beta} = \frac{FFT^*(EEG_{\beta}) FFT(EEG_{\beta})}{p^2}. \quad (2)$$

where p is the number of points in the signal temporal window of the signal, $FFT(EEG_{\alpha})$ denotes the FFT of the EEG signal in α band and $FFT^*(EEG_{\alpha})$ denotes its complex conjugate. The same notation is used for the β band of the EEG signal, (EEG_{β}) .

2.3 Classifier

ELM is an algorithm used to train an MLP. It is based on the concept that if the MLP input weights are fixed to random values, it can be considered as a linear system, so the output weights can be easily obtained using the pseudo-inverse of the hidden neurons outputs matrix \mathbf{H} for a given training set [28]. Given a set of N input vectors, an MLP can approximate N cases with zero error, $\sum_{i=1}^N \|\mathbf{y}_i - \mathbf{t}_i\| = 0$, \mathbf{y}_i being the output network for the input vector \mathbf{x}_i with target vector \mathbf{t}_i . Thus, there are β_j , \mathbf{w}_j and b_j so that,

$$\mathbf{y}_i = \sum_{j=1}^M \beta_j f(\mathbf{w}_j \cdot \mathbf{x}_i + b_j) = \mathbf{t}_i, \quad i = 1, \dots, N. \quad (3)$$

where $\beta_j = [\beta_{j1}, \beta_{j2}, \dots, \beta_{jm}]^T$ is the weight vector connecting the j th hidden node with the output nodes, $\mathbf{w}_j = [w_{j1}, w_{j2}, \dots, w_{jn}]^T$ is the weight vector connecting the j th hidden node and the input nodes, and b_j is the bias of the j th hidden node.

The previous N equations can be expressed by:

$$\mathbf{H}\mathbf{B} = \mathbf{T}, \quad (4)$$

where

$$\mathbf{H}(\mathbf{w}_1, \dots, \mathbf{w}_M, b_1, \dots, b_M, \mathbf{x}_1, \dots, \mathbf{x}_N) =$$

$$= \begin{bmatrix} f(\mathbf{w}_1 \cdot \mathbf{x}_1 + b_1) & \dots & f(\mathbf{w}_M \cdot \mathbf{x}_1 + b_M) \\ \vdots & \dots & \vdots \\ f(\mathbf{w}_1 \cdot \mathbf{x}_N + b_1) & \dots & f(\mathbf{w}_M \cdot \mathbf{x}_N + b_M) \end{bmatrix}_{N \times M} \quad (5)$$

$$\mathbf{B} = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_M^T \end{bmatrix}_{M \times m} \quad \text{and} \quad \mathbf{T} = \begin{bmatrix} \mathbf{t}_1^T \\ \vdots \\ \mathbf{t}_N^T \end{bmatrix}_{N \times m} \quad (6)$$

where $\mathbf{H} \in \mathbb{R}^{N \times M}$ is the hidden layer output matrix of the MLP, $\mathbf{B} \in \mathbb{R}^{M \times m}$ is the output weight matrix, and $\mathbf{T} \in \mathbb{R}^{N \times m}$ is the target matrix of the N training cases. The MLP training is given by the solution of the least square problem of (4). The optimal output weight layer is $\hat{\mathbf{B}} = \mathbf{H}^\dagger \mathbf{T}$, where \mathbf{H}^\dagger is the Moore-Penrose pseudo-inverse [27].

ELM needs to fix the number of hidden neurons. To do this, several pruned methods have been proposed [29–34]. The most commonly used method to avoid the exhaustive search for the optimal value of M , is the ELM optimally pruned (OP-ELM) [34]. The OP-ELM sets a very high initial number of hidden neurons and classifies them according to their importance is solving the problem [35]. The pruning of neurons is done by choosing the combination of neurons that provides lower Leave-One-Out error [34].

3 Experimental work

The aim of the experiments was to test ELM as an algorithm to classify features extracted with a linear method (PSD) from novice subjects. In this study EEG signals have been used as described in Section 2.1. To get the features, in a first step, it has been computed from 2 windows of 1 second located in the center of each trial (5-6s and 6-7s), see Figure 3. Subsequently, each feature has been averaged in a single vector, thus there are two features (α and β) from each channel in each trial that go to the classifier. It is important to remark that we have done a previous evaluation of some of the most used linear feature extraction techniques such as PSD, Hjorth and AAR [8]. However preliminary results show that Hjorth and AAR present very low performance with all the classifiers tested in this work, while PSD could get acceptable results for these signals. Thus we have discarded others approaches and we have chosen PSD as the most suitable method to explore the ELM in BCI systems.

In order to get a reference, the performances of ELM have been compared with the standard LDA and SVM techniques. These methods are widely used by researchers in BCI environments due to the successful results obtained by

them [9, 10]. For SVM, a linear kernel has been used, since a standard SVM with linear kernel is the most used method for BCI [9, 22]. For the ELM, a Gaussian kernel has been used, since it provides a non-linear solution and better results. In order to make an accurate and fair performance evaluation of the different classification approaches, this study uses a leave-one-out cross validation (LOO-CV) procedure [36]. LOO-CV avoids undesirable shifts from the random selection of training and test sets. For the N total number of samples involved in the study, one is retained for testing, and the remaining $N-1$ are used for training the classifier using the ELM approach. This process is repeated N times (i.e. an iteration for each input vector) Note that all cases are used for training and testing purposes during the N iterations of the LOO-CV procedure and, also, the performance evaluation measures are computed at the end of this iterative procedure.

4 Results

Table 1 shows the accuracy results obtained under LOO-CV (in %) using PSD as feature and three classifiers: ELM, LDA and SVM. As may be expected, values were not especially high. This is considered within normality due to the inexperience of the users and the absence of feedback in the system.

For the three classifiers, LOO-CV has been used. LDA and SVM produce stable performance, however, ELM showed a random initialization of the weights, so 30 initializations were made, and the results have been shown in relation to mean and standard deviation.

According to the results ELM improves the LDA and SVM performance for User 2 with $71.36\% \pm 02.84\%$ vs 66.67% of LDA, for User 3 with $69.79\% \pm 02.40\%$ vs 56.41% of LDA and SVM, and for User 4 with $58.54\% \pm 04.70\%$ vs 53.81% of SVM. However for User 1 SVM showed the best result with 71.79% , and for User 5 LDA improved the other results with 71.79% . It is interesting to note that ELM gets results between $65.67\% \pm 3.21\%$ and $71.36\% \pm 4.02\%$ in contrast to LDA that showed a low value for User 4 (41.03%) and SVM that presents 51.28% for User 2. Therefore, in this study, ELM achieved acceptable results for all users. As is shown in Figure 3, from the average data of the three methods, ELM improves significantly the mean classification performance of the LDA and SVM approaches: 66.51% vs. 60.51% and 59.47% respectively.

To validate this assertion, a non parametric statistical test has been performed. Specifically, the Wilcoxon

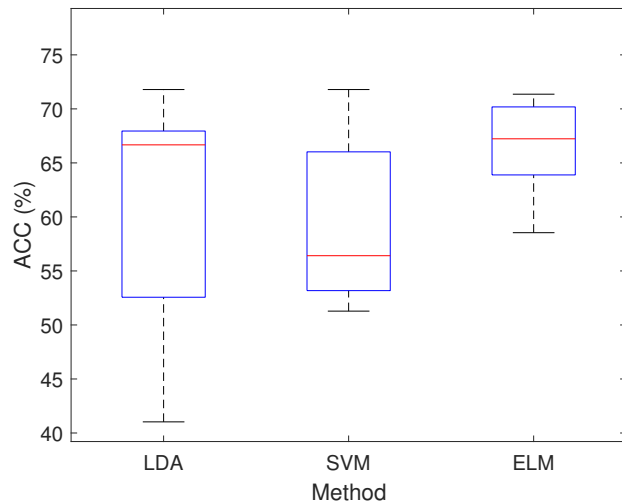


Figure 4: Comparison of the three methods means

Signed Ranks Test is used [41]. A peer review has been performed. Comparing ELM to LDA, the p -value obtained is 0.04, which indicates that there are significant differences to 96%, being the best ELM. Likewise, when applying the test with ELM against SVM, the p -value=0.01 indicates that there are significant differences to 99%, being ELM better than SVM. However, there are no significant differences between LDA and SVM, since the p -value = 0.477.

Table 1: Obtained PSD results (Accuracy -ACC-) by several classifier of five novel users. Performance evaluation has been done under a LOO-CV procedure.

Novel user	Method	ACC (%)
<i>User₁</i>	LDA	66.67
	SVM	71.79
	ELM	67.23±02.84
<i>User₂</i>	LDA	66.67
	SVM	51.28
	ELM	71.36±04.02
<i>User₃</i>	LDA	56.41
	SVM	56.41
	ELM	69.79±02.40
<i>User₄</i>	LDA	41.03
	SVM	53.81
	ELM	58.54±04.70
<i>User₅</i>	LDA	71.79
	SVM	64.10
	ELM	65.67±03.21

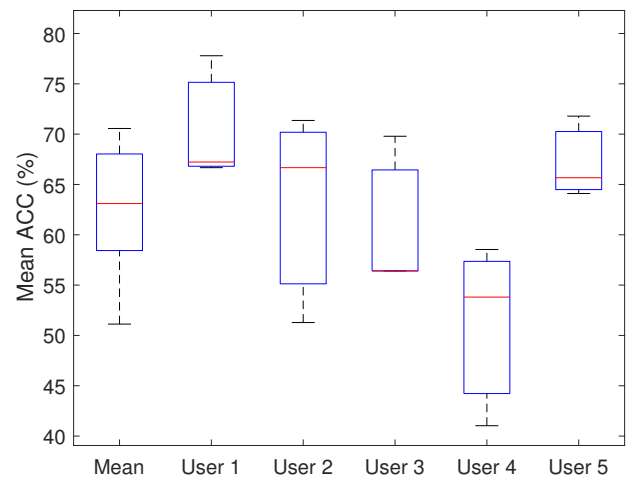


Figure 5: Comparison of the five users means

It is widely known each subject is different and the performance of the same experiment and system could be different between users and even for the same user in different trials or sessions. Figure 4 shows the mean accuracy values of all users taking into account the mean of the data of the three classifiers applied for each user. Thus, Users 1, 2 and 5 improved the mean value of the group in contrast with Users 2 and 3. Probably Users 1 and 2 have good skills to control BCI systems under motor imagery paradigms. Therefore, they could be suitable candidates to be trained to develop future experiments.

5 Conclusions and future work

This study tested an ELM with EEG signals from 5 novices users using a MI based BCI system. PSD features were computed and averaged from 2 central windows with 1 second each one. In order to compare the accuracy of the classification results obtained by ELM, two standard classifiers, LDA and SVM have been implemented. The results have been evaluated under LOO-CV and showed that, in contrast to LDA and SVM, ELM reached appropriate results in all the users, and besides that, it outperformed in 3 of 5 users standard methods. Therefore, the tested methodology showed a suitable performance for application in MI BCI systems.

Future research on this topic should be focused on testing ELM with wider data sets, as well as with signals from BCI expert users or from reference signals in BCI approaches such as BCI competitions. In addition, future studies could be accomplished to compute and combine features in three or four temporal windows to take into

account the time course of the signal. Other possibilities of the ELM will be evaluated, such as different variants of pruning and selection of the architecture. Finally, it might be interesting to implement more feature extraction techniques and classifiers.

Acknowledgement: This work has been partially supported by Spanish MINECO grant number MTM2014-51891-P and by the Spanish MINECO under grant TIN2016-78799-P (AEI/FEDER, UE). But the most important is that all the persons that have worked in this paper were fortunate to learn, to work and to be friends of PhD. Pedro José García Laencina. Rest in peace.

References

- [1] Wolpaw J.R., Birbaumer N., McFarland D.J., Pfurtscheller G., Vaughan T.M., Brain computer interfaces for communication and control, *Clin Neurophysiol*, 2002, 113, 767–791.
- [2] Sanei D., Chambers J., EEG Signal Processing. John Wiley & Sons., 2008.
- [3] Sabeti M., Boostani R., Katebi S., Price G., Selection of relevant features for EEG signal classification of schizophrenic patients, *Biomedical Signal Processing and Control*, 2007, 2(2), 122–134.
- [4] Rodríguez-Bermúdez G., García-Laencina P., Analysis of eeg signals using nonlinear dynamics and chaos: A review, *Applied Mathematics & Information Sciences*, 2015.
- [5] Tomasevic N., Neskovic A., Neskovic N., Neural network-based approach to EEG signal simulation, *International Journal of Neural Systems*, 2012, 22 (3), 1–16.
- [6] Tong S., Thankor N. V., Quantitative EEG Analysis Methods and Clinical Applications, Artech House.
- [7] Pfurtscheller G., Brunner C., Schlögl A., da Silva FL. Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks, *NeuroImage*, 2006, 31(1), 153–159.
- [8] Bashashati A., Fatourechhi M., Ward R.K., Birch G.E., A survey of signal processing algorithms in brain-computer interfaces based on electrical brain signals, *Journal of Neural Engineering*, 2007, 4(2), 32–57.
- [9] Lotte F., Congedo M., Lécuyer A., Lamarche F., Arnaldi B., A review of classification algorithms for EEG-based brain-computer interfaces, *Journal of neural engineering*, 2007, 4(2), 1–13.
- [10] Nicolas-Alonso L., Gomez-Gil J., Brain computer interfaces, a review, *Sensors*, 2012, 1211–1279.
- [11] Garca-Laencina P. J., Rodriguez-Bermudez G., Roca-Dorda J., Exploring dimensionality reduction of EEG features in motor imagery task classification, *Expert Systems with Applications*, 2014, 23 (04), 5285–5295.
- [12] Álvarez-Arenas A., Belmonte-Beitia J., Calvo G.F., Nonlinear waves in a simple model of high-grade glioma, *Applied Mathematics and Nonlinear Sciences*, 2016, 1 (2), 405–422.
- [13] Pérez-García V.M., Fitzpatrick S., Pérez-Romasanta L.A., Pesic M., Schucht P., Arana E., Sánchez-Gómez P., Applied mathematics and nonlinear sciences in the war on cancer, *Applied Mathematics and Nonlinear Sciences*, 2016, 1 (2), 423–436.
- [14] Taplan M. Fundamentals of EEG measurement. *Measurement Science Review*, 2002, 2, 1-11.
- [15] Nijboer F., Furdea A., Gunst I., Mellinger J., McFarland D.J., Birbaumer N., Kubler A., An auditory brain-computer interface (BCI), *Journal of Neuroscience Methods*, 2008, 167, 43-50.
- [16] Ahn M., Chan Jun S., Performance variation in motor imagery brain-computer interface: A brief review, *Journal of Neuroscience Methods*, 2015, 243, 103-110.
- [17] Vukelić M., Gharabaghi A., Oscillatory entrainment of the motor cortical network during motor imagery is modulated by the feedback modality, *NeuroImage*, 2015, 111, 1-11.
- [18] Guger C., Edlinger G., Harkam W., Niedermayer I., Pfurtscheller G.A., How Many People are Able to Operate an EEG-Based Brain-Computer Interface (BCI)?, *IEEE Transactions on neural systems and rehabilitation engineering*, 2003, 11(2), 145-147.
- [19] Stam C., Nonlinear dynamical analysis of EEG and MEG: Review of an emerging field, *Clinical Neurophysiology*, 2005, 116, 2266-2301.
- [20] Liu J., Zhang C., Zheng C., EEG-based estimation of mental fatigue by using KPCA-HMM and complexity parameters, *Biomedical Signal Processing and Control*, 2010, 5, 124-130.
- [21] Bajaj V., Pachori R., Classification of human emotions based on multiwavelet transform of EEG signals, 2013 AASRI Conference on Intelligent Systems and Control, Elsevier, Vancouver, 2013.
- [22] Krusienski D. J., Sellers E. W., Cabestaing F., Bayouth S., McFarland D. J., Vaughan T. M., and Wolpaw J. R., A comparison of classification techniques for the P300 speller, *J. Neural Eng*, 2006, 3(4), 299–305.
- [23] Pao Y.-H., Park G.-H., Sobajic D. J., Learning and generalization characteristics of the random vector functional-link net Neurocomputing, Elsevier, 1994, 6, 163-180.
- [24] Igel'nik B., Pao Y.-H., Stochastic choice of basis functions in adaptive function approximation and the functional-link net Neural Networks, *IEEE Transactions on*, 1995, 6, 1320-1329.
- [25] Huang G.-B., Chen L., Convex incremental extreme learning machine, *Neurocomputing*, 2007, 70 (16), 3056–3062.
- [26] Huang G.-B., Wang D. H., Lan Y. Extreme learning machines: a survey, *International Journal of Machine Learning and Cybernetics*, 2011, 2 (2), 107–122.
- [27] Serre D., *Matrices: Theory and Applications*, Springer, New York, 2002.
- [28] Huang G., Zhu Q., Siew C., Extreme learning machine: Theory and applications, *Neurocomputing*, 2006, 70 (1-3), 489–501.
- [29] Rong H.-J., Ong Y.-S., Tan A.-H., Zhu Z., A fast pruned-extreme learning machine for classification problem, *Neurocomputing*, 2008, 72 (1-3), 359–366.
- [30] Miche Y., Bas P., Jutten C., Simula O., Lendasse A., A methodology for building regression models using extreme learning machine: OP-ELM, In: *Proceedings of the European Symposium on Artificial Neural Networks (ESANN)*, 2008, 247–252.
- [31] Miche Y., Sorjamaa A., Lendasse A., OP-ELM: Theory, experiments and a toolbox, In: *Proceedings of the International Conference on Artificial Neural Networks (ICANN)*, LNCS, 2008, 5163, 145–154.
- [32] Mateo F., Lendasse A., A variable selection approach based on the delta test for extreme learning machine models, In: *Proceedings of the European Symposium on Time Series Prediction (ESTP)* 2008, 57–66.
- [33] Miche Y., Lendasse A., A faster model selection criterion for OPELM and OP-KNN: Hannan-quinn criterion, In: *Proceeding of*

- the European Symposium on Artificial Neural Networks (ESANN), 2009, 177–182.
- [34] Miche Y., Sorjamaa A., Bas P., Simula O., Jutten C., Lendasse A., OP-ELM: Optimally Pruned Extreme Learning Machine, *IEEE Transactions on Neural Networks*, 2009, 21 (1), 158–162.
- [35] Similä T., Tikka J., Multiresponse sparse regression with application to multidimensional scaling, In: *Proceedings of the 15th International Conference on Artificial Neural Networks: Formal Models and Their Applications - ICANN 2005, LNCS*, 2005, 3697, 97–102.
- [36] Alpaydin E., *Introduction to Machine Learning*, MIT Press. Cambridge, MA, USA, 2010.
- [37] Duan L. and Zhong H. and Miao J., Yang Z., Ma W., Zhang X., A voting optimized strategy based on ELM for improving classification of motor imagery BCI data, *Cognitive Computation*, 2014, 477–483.
- [38] Bamdadian A., Guan C., Ang K.K., Xu J., Improving session-to-session transfer performance of motor imagery-based BCI using adaptive extreme learning machine, *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, 2013, 2188–2191.
- [39] Tan P., Tan G.-Z., Cai Z.-X., Sa W.-P., Zou Y.-Q., Using ELM-based weighted probabilistic model in the classification of synchronous EEG BCI, *Medical & biological engineering & computing*, 2017, 55 (1), 33–43.
- [40] Tan P., Sa W., Yu L., Applying Extreme Learning Machine to classification of EEG BCI, *Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, 2016 IEEE International Conference on, 2016, 228–232.
- [41] Kruskal W. H. Historical notes on the wilcoxon unpaired two-sample test, *Journal of the American Statistical Association*, 1957, 52 (279), 356–360.