EEG Signal Clustering for Motor and Imaginary Motor Tasks on Hands and Feet

Víctor Asanza[1], Enrique Pelaez[2], Francis Loayza[3]
Escuela Superior Politécnica del Litoral, ESPOL
Facultad de Ingeniería en Electricidad y Computación,
Campus Gustavo Galindo Km 30.5 Vía Perimetral, P.O. Box 09-01-5863, Guayaquil, Ecuador
{vasanza[1], epelaez[2], floayza[3]}@espol.edu.ec

Abstract —. Modern technologies use Brain Computer Interfaces (BCI) to control devices or prosthesis for people with physical impairments. In some cases, EEG data are used to determine the intentionality of the subject when performing motor and imaginary motor tasks. However, EEG signals are very susceptible to noise due to the lower voltage levels that are acquired. We used a data set of 64 EEG recordings of 25 healthy subjects while they were doing motor and imaginary motor movements of hands and feet. Data were preprocessing, including the design of a filter for noise reduction outside the expected frequency spectral that operate the EEG signals. Then, we used features extraction based on spectral density. Finally, the application of five clustering algorithms to detect motor and imaginary motor tasks. Results showed that the k-means, k-medoids and Hierarchical clustering algorithms were better detecting motor activity, and hierarchical clustering for imaginary tasks of hands.

Keywords — Electroencephalography, Brain Computer Interface, Fast Fourier Transform, Butterworth Filter, direct current artifacts, Power Spectral Density, kmeans, k-medoids, dbscan, spectral clustering, hierarchical-clustering.

I. Introduction

The biomedical signals of Electroencephalography (EEG) allow us to measure the potential difference generated in the scalp; these signals represent neuronal activities corresponding to each area of the brain [1]. One of the most well-known applications of EEG signals are the Brain Computer Interfaces (BCI), BCI allows subjects with some type of motor disability to interact with some devices or prostheses [2].

During a motor or imaginary motor activity, electrical signals, in the order of micro volts (uV), are generated in the motor cortex of the brain [3, 4]. There are two methods for measuring these EEG signals: Invasive and superficial [5]. The invasive measurement is done via electrodes inserted directly on the surface of the brain; this method is considered risky and unnecessary in some cases. The superficial method allows us to detect the activity of neurons using electrodes placed on the scalp; it is the most common method for measuring because it can be performed with minimal risk to the subject [4].

978-1-5386-3894-1/17/$31.00 ©2017 IEEE

Acquisition EEG data has interference problems caused by the electrodes used in the skin to collect the signals coming from, either the electric network, or the relative movements between electrodes, blinking, muscular activity, heart rate, breathing, etc. [6, 7].

The development of technologies for enhancing the rehabilitation process in subjects with motor disabilities is needed, in particular for those who do not have the capacity to control their movements due to diseases, such as: birth defects, cerebro-vascular accidents, trauma, and others; the measurement of bioelectrical signals, that subjects generate while performing imaginary motor activities, could allow us to determine the subject's intentionality. Also, given the susceptibility of these signals, to the noise generated by the measurement techniques and technologies used, some methodologies of analysis and preprocessing of the EEG signals have been proposed as in Guger, C. et al. [7].

In this work, a method for preprocessing EEG signals was used to eliminate the noise components and a feature extraction mechanism that could later be used to determine the subject's motor intention; the processed data set will then be used as an input for the clustering algorithms that we will evaluate. Section II describes the related work about detecting motor intentions; section III details the data set used; section IV explains the methodology used for preprocessing data, extraction of features and the clustering algorithms evaluated; section V shows the results of the clustering algorithms used with the preprocessed data set. Finally, section VI presents a discussion of the results and conclusions.

II. Related Work

To the best of our knowledge, techniques for detecting cortical motor activity corresponding to the lower limbs, have not been well studied because they are difficult to detect, they are recorded in the central motor gyro located on the inner side of the longitudinal fissure of the brain [8]. For this reason, some topographic visualization techniques, based on EEG data of motor cortical activity, are mainly focused on the upper limbs [9-11].

The motor cortical activity measured with EEG-BCI systems is most evident in the frequency band of 13-30 Hz or β band and in the frequency band of 8-12 Hz or μ band [6]. It is common to use Power Spectral Density (PSD) to measure features for determining motor intention in the frequency range of β and μ, it is common to use Power Spectral Density (PSD) measurement features [12-14].

Some algorithms for automated learning (supervised and unsupervised), that have been used for detecting motor intentions in the upper limbs are: Support Vector Machines (SVM) [13, 15, 18-20], Artificial Neural Networks (ANN) [13, 14, 17], Linear Discriminant Analysis (LDA) [13, 16], and clustering algorithms [1, 3, 21]. SVM have better performance in detecting motor intentions in the upper limbs [11, 13, 20]. In this work, we performed a comparison between some clustering algorithms, to determine its performance in the detection of motor tasks.

III. DATA SET

For this work, we used a data set of EEG signal registers acquired from healthy subjects using a BCI-2000 system, available on the Physio Net website: Https://www.physionet.org/physiobank/database/eegmmidb/.

The data set consists of 14 files that follows the European Data Format (EDF) [12] with sampling frequency of 160Hz, with the recordings of brain activity from 25 healthy subjects, who were asked to perform different imaginary and motor tasks while the EEG signals were recorded using 64 electrodes placed in the surface of the scalp.

Each one of the 14 EDF files of each subject contains the acquired signals while performing the following tasks: base line with open eyes, base line with closed eyes, task 1 (open and close left or right hand), task 2 (imagine opening and closing left or right hand), task 3 (open and close both hands or both feet), task 4 (imagine opening and closing both hands or both feet). Tasks 1 and 2 have a duration of 1 minute each one; tasks 2, 3 and 4 have a duration of two minutes each one. The EDF file corresponding to each task has 30 actions organized in random order and identified by three annotations (T0, T1, and T2) that indicated the type of activity performed: T0 corresponds to rest. T1 corresponds to onset of motion (real or imagined) of the left hand (stored in files: 3, 4, 7, 8, 11, and 12), and both hands (stored in files: 5, 6, 9, 10, 13, and 14). T2 corresponds to the onset of motion (real or imagined) of the right hand (stored in files 3, 4, 7, 8, 11, and 12) and for both feet (stored in files: 5, 6, 9, 10, 13, and 14) [12].

Since our objective is to allow later to determine motor intention, was have focused on detecting and differentiating imaginary motor activity from the actual motor activity, so tasks 3 and 4 we used, which correspond to the EDF files: 5, 6, 9, 10, 13 and 14 for each subject. For each one of the 25 subjects, 6 EDF files were obtained with 30 actions. To facilitate the analysis each EDF file was grouped with the annotation T1 and T2, and saved in two Matlab files. All these 300 (25 x 6 x 2) Mat files have 4599 rows with EEG signal samples x 64 columns (electrodes) which will be used to extract the features and for the clustering algorithms. Matlab V9.2.0.556344 (The Matworks, Natick, MA) was used for the complete analysis.

Fig. 1 shows some samples of the EEG signal, corresponding to the imaginary motor task of both hands; this graph contains 4599 samples of the 64 surface electrodes captured simultaneously. In addition, it can be seen that the whole set of 64 signals contains DC components, trends or DC artifacts.

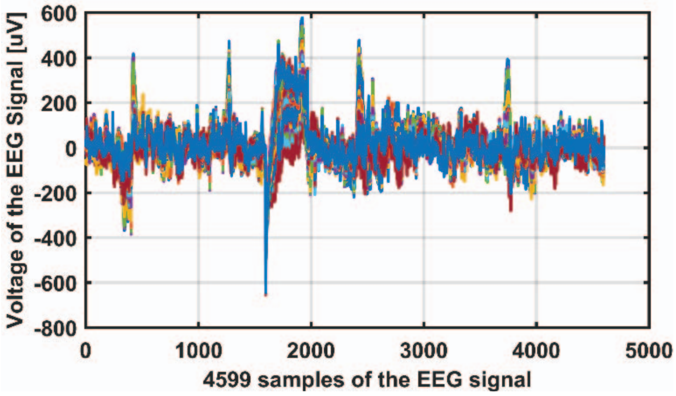

Fig. 1. DC artifact present on the 64 electrodes of the EEG signal of imaginary motor task

The proposed methodology in IV, considers performing an analysis of the frequency domain and determining the presence of noise generated by the power distribution network. For this analysis, the Fast Fourier Transform (FFT) was applied to each of the 300 Mat files. All electrodes have a strong activity, due to the presence of DC artifacts for values close to 0 Hz frequency.

IV. METHODOLOGY AND RESULTS

Once the evidence of the presence and type of noise in the EEG signals were identified, a Butterworth filter was designed, to produce a frequency response, as flat as possible, and to avoid distortion in the original signal in the frequency domain [7,14]. The bandpass filter of constant coefficients (Buttherworth-IIR or Wiener-FIR) was designed for the frequency range between 7 and 30 Hz [6, 22].

Each of the 300 Mat files with filtred data were randomly distributed in a single Data.Mat file, where the rows represent the events of motor activity (EDF files 5, 9, and 13) for both hands (T1) and both feet (T2) and imaginary motor activity (EDF files 6, 10 and 14) of both hands (T1) and both feet (T2).

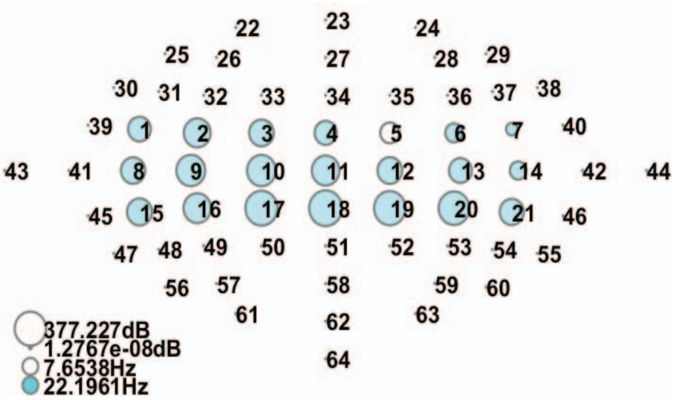

Fig. 2. Distribution of 64 surface electrodes

Fig. 2 shows the maximum PSD value and frequency occur in the 21 electrodes located in the motor cortex. Electrodes from

1 to 21 have neuronal electrical activity corresponding to motor or motor imaging tasks [3, 6].

A periodogram or Welch PSD was used to extract the Power Spectral Density (PSD) features, which will allow us to have the power distribution of the signal as a function of frequency [17, 20]. Columns in the Data.Mat file contain PSD features such as: The maximum PSD value, frequency, arithmetic mean and variance. All these 84 (21 x 4) features were added as columns in a new Features.Mat file, where rows correspond to the 300 events as indicated above.

In addition, a column with the indicator of the event to which each row belongs is added, in order to validate the results of the clustering algorithms. Fig. 3 shows the features norm values of an imaginary motor task both hands, the first 42 features are 21 maximum PSD values and 21 frequency values, respectively. The remaining 42 features are 21 values of arithmetic mean and 21 variance of PSD value, respectively.

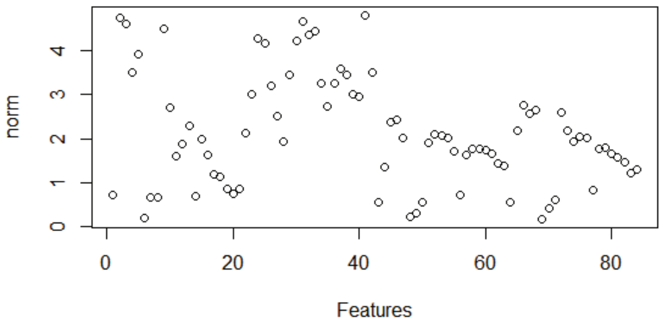

Fig. 3. 84 features in the imaginary motor task both hands

The first algorithm with which Features.mat data is tested was the k-means clustering, this algorithm, based on the Euclidean distance to the centroid of the cluster, will be used with 4 centroids (k = 4) [23]. Fig. 4 shows the data clustered using convex types, which allows all features to have a better clustering.

Fig. 4. K-means algorithm, with nine centroids

Another clustering algorithm tested was the k-medoids clustering, which is based on the Manhattan Norm distance to the centroid of a cluster. The difference with k-means is that this algorithm chooses data points as centroids. Fig. 5 shows the clustering features using convex type.

Fig. 5. K-medoids algorithm, with nine centroids

A third algorithm used was the Density-Based Clustering (DBSCAN), this algorithm needs the minimum distance between points as a parameter to be considered in the same cluster (EPS), Fig. 6 shows the optima calculation distance between points to be considered in the same cluster (EPS) [1].

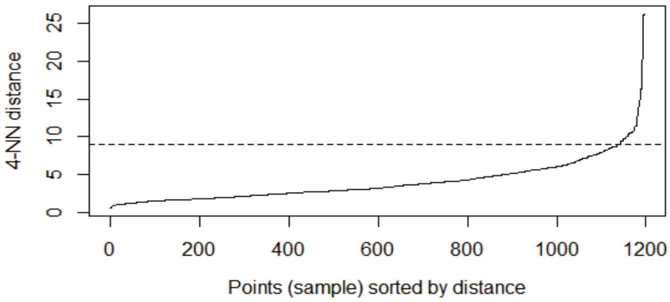

Fig. 6. Optimal EPS distance calculation for DBSCAN with minimum distance = 9

DBSCAN is an allotment based on density and was evaluated with a value of minpoint = 4. The result of clustering using convex type is shown in Fig. 7, as seen in the picture, this algorithm does not allow the detection of the 4 motor tasks of the experiment.

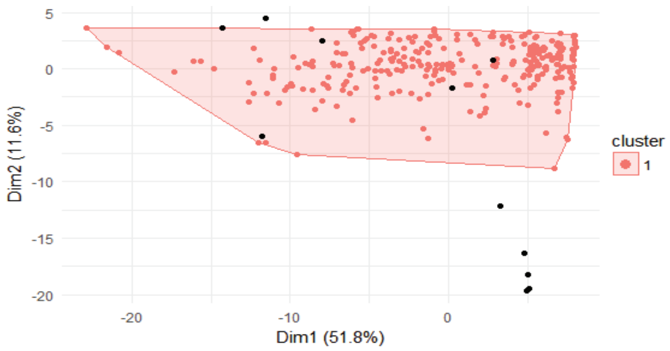

Fig. 7. Clustering results with DBSCAN

The spectral density clustering was also used with the dataset, this algorithm uses 4 centroids (k = 4). Fig. 8 shows the data clustered using convex type, unlike k-means and k-medoids algorithms, the spectral clustering has more features in cluster 4.

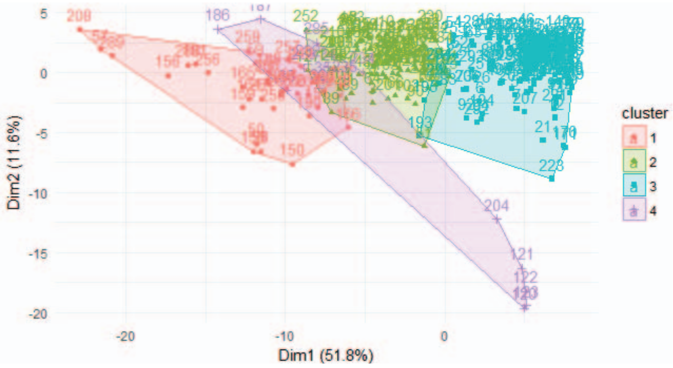

Fig. 8. Spectral Clustering results

The hierarchical-clustering algorithm construct a cluster tree, the results of using this algorithm with the dataset are shown in Fig. 9. Hierarchical-clustering presents a behavior like the k-means and k-medoids algorithms, with a small cluster containing few elements [13].

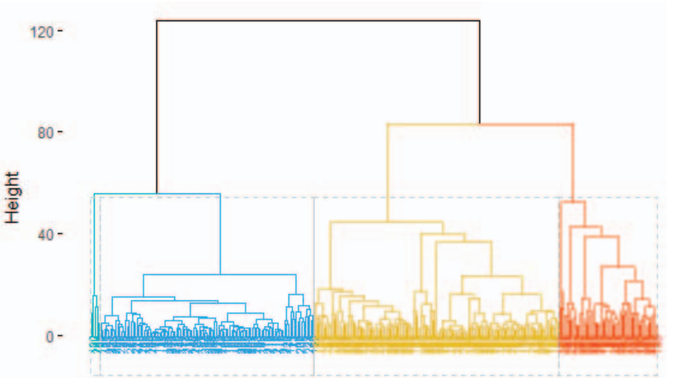

Fig. 9. Results of Hierarchical Clustering

V. Analysis of Results

After identifying the characteristics of the noise, the pre-processing aims to reduce the Signal-to-Noise Ratio (SNR) of the EEG data in the range of 7 - 30 Hz (μ and β rhythms), applying the designed filter [4]. The 4599 samples of the 64 surface electrodes, captured simultaneously in the time domain, are shown in Fig. 10, it can also be observed the effect of eliminating the noise, caused by the presence of DC artifacts, also a signal is observed without offsets or trends. The frequency components of the filtered signals are also shown in Fig. 10, where the activity is shown in the range of 7-30 Hz.

Fig. 11 shows the optimal cluster number values for each clustering algorithm used in this work (k-means, k-medoids, DBSCAN, Spectral-Clustering and Hierarchical-Clustering). In order to identify the four tasks or events, the cluster number defined was 4, these tasks are: Motor activity of both hands (T1) and of both feet (T2); imaginary motor activity of both hands (T1) and imaginary motor activity of both feet (T2).

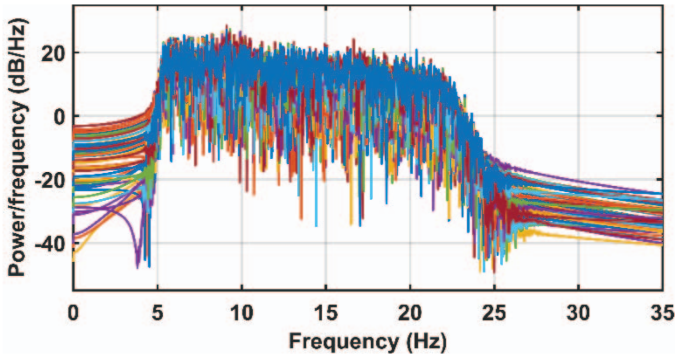

Fig. 10. Frequency analysis with the FFT of the filtered EEG signals

Fig. 12 shows the results of the correct percentage of each clustering algorithms; the percentage was calculated comparing the cluster assigned to each task by the algorithm vs the annotation provided in the dataset. The results correspond to, Cluster 1 Motor activity of both hands (T1), cluster 2 motor activity of both feet (T2), cluster 3 imaginary motor activity of both hands (T1) and cluster 4 imaginary motor activity of both feet (T2).

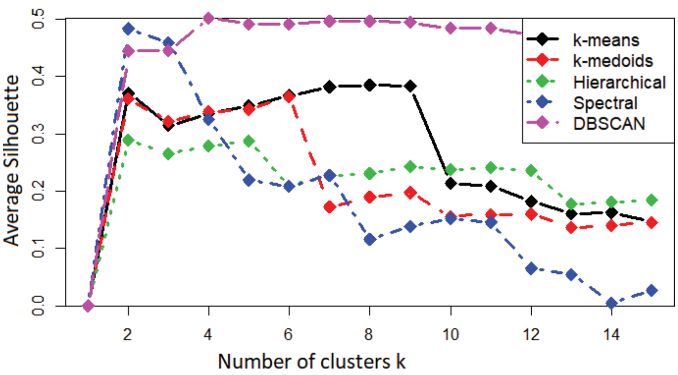

Fig. 11. Explore the optimal number of cluster for all cluster algorithm

The results show that none of the clustering algorithms could detect motor activity of both feet; the Hierarchical-clustering algorithm detects better motor activity of both hands; the k-medoids algorithm detects better imaginary motor activity of both hands; and, the Spectral clustering detects better imaginary motor activity of both feet.

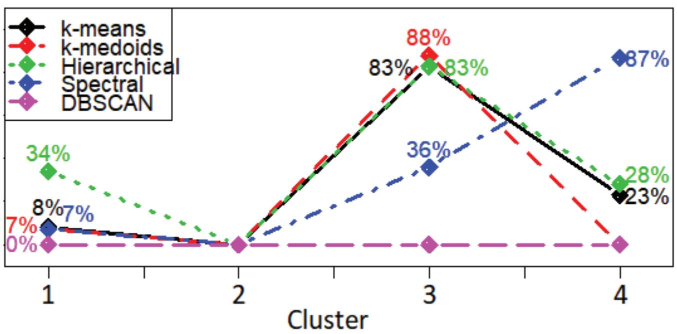

Fig. 12. Percent success of all clustering algorithms

VI. Discussion and Conclusions

The objective of this research was to compare clustering algorithms on EEG signals to detect different tasks that include motor and imaginary motor activity. It was necessary to use the FFT of the EEG signals to eliminate the strong activity of spectral density near 0 Hz. The use of the Butterworth filter allows us to define the behavior in power spectral density in the frequency range of 7-30 Hz of the signals acquired from the brain, which correspond to the motor cortex.

The analysis of the clustering obtained from each of the different algorithms evaluated, allows us to determine that the k-means, k-medoids and Hierarchical clustering algorithms have a better performance in detecting motor activity, specifically they are better with the motor tasks of both hands (T3) with a success greater than 80%. The results also show that the Hierarchical clustering algorithm detects better the imaginary motor tasks of both hands (T1) with a 34% success rate, but the spectral clustering algorithm is the one that has a better percentage of success in the detection of motor tasks of both feet (T4). In our experiments, none of the algorithms evaluated could perform a detection of both feet (T2) motor imaginary tasks.

In future work new features will be added for analysis for clustering, such as the Event Related Desynchronization (ERD) and Event Related Synchronization (ERS), also called (ERD / ERS); which is a characteristic related to SMR sensorimotor rhythms [11, 24], to evaluate the success rate in detection of motor activities of both hands and of both feet.

References

[1] Reaz, M. B., Hussain, M. S., & Mohd-Yasin, F. (2006). Techniques of EMG signal analysis: detection, processing, classification and applications. Biological procedures online, 8(1), 11-35.

[2] Pfurtscheller, G., Neuper, C., Guger, C., Harkam, W. A. H. W., Ramoser, H., Schlogl, A., ... & Pregenzer, M. A. P. M. (2000). Current trends in Graz brain-computer interface (BCI) research. IEEE Transactions on Rehabilitation Engineering, 8(2), 216-219.

[3] Mulroy, S., Gronley, J., Weiss, W., Newsam, C., and Perry, J. Use of cluster analysis for gait pattern classication of subjects in the early and late recovery phases following stroke. Gait & posture 18, 1(2003), 114-125.

[4] G. Schalk and E. C. Leuthardt, “Brain-computer interfaces using electrocorticographic signals,” IEEE reviews in biomedical engineering, vol. 4, pp. 140–154, 2011.

[5] Wolpaw, J. R. (2007). Brain–computer interfaces as new brain output pathways. The Journal of physiology, 579(3), 613-619.

[6] Minguillon, J., Lopez-Gordo, M. A., & Pelayo, F. (2017). Trends in EEG-BCI for daily-life: Requirements for artifact removal. Biomedical Signal Processing and Control, 31, 407-418.

[7] Guger, C., Schlogl, A., Neuper, C., Walterspacher, D., Strein, T., & Pfurtscheller, G. (2001). Rapid prototyping of an EEG-based brain-computer interface (BCI). IEEE Transactions on Neural Systems and Rehabilitation Engineering, 9(1), 49-58.

[8] Lee, J. J., Lee, D. R., Kyum Shin, Y., Lee, N. G., Han, B. S., & You, S. J. H. (2013). Comparative neuroimaging in children with cerebral palsy using fMRI and a novel EEG-based brain mapping during a motor task–a preliminary investigation. NeuroRehabilitation, 32(2), 279-285.

[9] Shin, Y. K., Lee, D. R., Hwang, H. J., You, S. J. H., & Im, C. H. (2012). A novel EEG-based brain mapping to determine cortical activation patterns in normal children and children with cerebral palsy during motor imagery tasks. NeuroRehabilitation, 31(4), 349-355.

[10] Müller-Putz, G. R., Scherer, R., Pfurtscheller, G., & Rupp, R. (2005). EEG-based neuroprosthesis control: a step towards clinical practice. Neuroscience letters, 382(1), 169-174.

[11] Gao, L., Wang, J., & Chen, L. (2013). Event-related desynchronization and synchronization quantification in motor-related EEG by Kolmogorov entropy. Journal of neural engineering, 10(3), 036023.

[12] Schalk, G., McFarland, D.J., Hinterberger, T., Birbaumer, N., Wolpaw, J.R. BCI2000: A General-Purpose Brain-Computer Interface (BCI) System. IEEE Transactions on Biomedical Engineering 51(6):1034-1043, 2004. [In 2008, this paper received the Best Paper Award from IEEE TBME.]

[13] Bascil, M. S., Tesneli, A. Y., & Temurtas, F. (2016). Spectral feature extraction of EEG signals and pattern recognition during mental tasks of 2-D cursor movements for BCI using SVM and ANN. Australasian physical & engineering sciences in medicine, 39(3), 665-676.

[14] Bascil, M. S., Tesneli, A. Y., & Temurtas, F. (2015). Multi-channel EEG signal feature extraction and pattern recognition on horizontal mental imagination task of 1-D cursor movement for brain computer interface. Australasian Physical & Engineering Sciences in Medicine, 38(2), 229-239.

[15] Rambabu, C., & Murthy, B. R. (2014). EEG signal with feature extraction using SVM and ICA classifiers. International Journal of Computer Applications, 85(3).

[16] Chai, R., Naik, G., Nguyen, T. N., Ling, S., Tran, Y., Craig, A., & Nguyen, H. (2016). Driver fatigue classification with independent component by entropy rate bound minimization analysis in an EEG-based system. IEEE journal of biomedical and health informatics.

[17] Hazrati, M. K., & Erfanian, A. (2010). An online EEG-based brain–computer interface for controlling hand grasp using an adaptive probabilistic neural network. Medical engineering & physics, 32(7), 730-739.

[18] Gunn, S. R. (1998). Support vector machines for classification and regression. ISIS technical report, 14, 85-86.

[19] Suykens, J. A., & Vandewalle, J. (1999). Least squares support vector machine classifiers. Neural processing letters, 9(3), 293-300.

[20] Singla, R., & Haseena, B. A. (2014). Comparison of ssvep signal classification techniques using svm and ann models for bci applications. International Journal of Information and Electronics Engineering, 4(1), 6.

[21] Özgen, C. (2011). Classification of motor imagery tasks in EEG signal and its application to a brain-computer interface for controlling assistive environmental devices (Doctoral dissertation, Middle East Technical University).

[22] Stoica, P., & Moses, R. L. (2005). Spectral Analysis of Signals: Prentice Hall. Upper Saddle River, NJ.

[23] Hartigan, J. A., and Wong, M. A. Algorithm as 136: A k-means clustering algorithm. Journal of the Royal Statistical Society. Series C (Applied Statistics) 28, 1 (1979), 100-108.

[24] Gomez-Pilar, J., Corralejo, R., Nicolas-Alonso, L. F., Álvarez, D., & Hornero, R. (2016). Neurofeedback training with a motor imagery-based BCI: neurocognitive improvements and EEG changes in the elderly. Medical & biological engineering & computing, 54(11), 1655-1666.