International Journal of Internet, Broadcasting and Communication Vol.15 No.2 124-132 (2023)
http://dx.doi.org/10.7236/IJIBC.2023.15.2.124

IJIBC 23-2-14

Making Thoughts Real – a Machine Learning Approach for Brain-Computer Interface Systems

Tengis Tserendondog*, Uurstaikh Luvsansambuu, Munkhbayar Bat-Erdende, Batmunkh Amar

Associate professor, Dept. of Electronics, Mongolian University of Science and Technology, Mongolia
*tengis@must.edu.mn**
uurtsaikh@must.edu.mn
munkhbayar.b@must.edu.mn
abatmunkh@must.edu.mn

Abstract

In this paper, we present a simple classification model based on statistical features and demonstrate the successful implementation of a brain-computer interface (BCI) based light on/off control system. This research shows study and development of light on/off control system based on BCI technology, which allows the users to control switching a lamp using electroencephalogram (EEG) signals. The logistic regression algorithm is used for classification of the EEG signal to convert it into light on, light off control commands. Training data were collected using 14-channel BCI system which records the brain signals of participants watching a screen with flickering lights and saves the data into .csv file for future analysis. After extracting a number of features from the data and performing classification using logistic regression, we created commands to switch on a physical lamp and tested it in a real environment. Logistic regression allowed us to quite accurately classify the EEG signals based on the user's mental state and we were able to classify the EEG signals with 82.5% accuracy, producing reliable commands for turning on and off the light.

Keywords: *EEG, data processing, logistic regression, control system, brain signal.*

1. Introduction

In recent years, researches based on BCI have been advancing rapidly. People with disabilities in mobility, communication and writing face challenges in their daily lives. BCI technology helps them overcome these obstacles and lead a more independent life [1, 2]. The human brain is composed of approximately 86 billion neurons. Nowadays with the help of specialized technique electrical signals from 5 up to 96 different points on the scalp can be scanned. One of such systems is a device of Emotiv company, which is capable of processing data from 5 to 32 different points. The Emotiv Epoc X device we utilized in our study is a cost-effective system capable of accurately detecting signals from 14 points on the scalp. The device is equipped with a wireless headset having a special shape to wear on the scalp. Due to the 14 electrodes the dynamic changes of the electrical potentials of the brain activities are scanned, transferred and recorded into files. Locations of the electrodes as green circles and their preference points as orange circles are shown in Figure

Manuscript Received: March. 15, 2023 / Revised: March. 18, 2023 / Accepted: March. 22, 2023
Corresponding Author: tengis@must.edu.mn
Tel: +976 -99095618
Associate professor, Dept. of Electronics, Mongolian University of Science and Technology, Mongolia

1. Each electrode is named after the area where it is placed on the scalp with reference to where the electrode for each area is located. The frontal electrodes (AF3, F7, F3, F4, F8, AF4) generally pick up activity related to cognitive processes such as attention, decision-making, and ` memory. The temporal electrodes (T7, T8) are associated with auditory processing and the occipital electrodes (O1, O2) are associated with visual processing. The central electrodes (FC5, FC6) are involved in motor planning and execution, while the parietal electrodes (P7, P8) are associated with spatial processing and sensory integration.

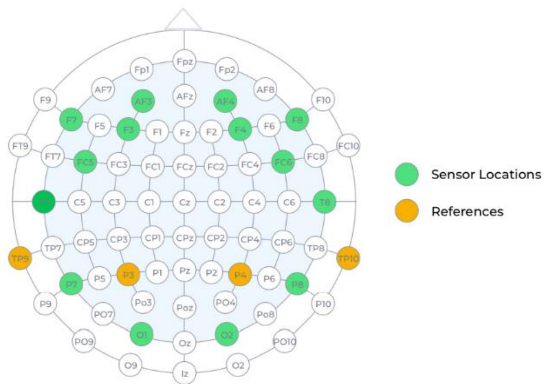


Figure 1. Electrode locations on the scalp

The human brain generates electrical activity, which can be measured using EEG. The EEG signal is recorded and analyzed using various frequency bands, which are used in subsequent data processing and analysis. These frequency bands reveal the corresponding brain states and related neural activity which can be filtered by band pass filters (BPF). Delta (0.5-4 Hz) brainwaves are the slowest brainwaves and are typically associated with deep sleep, unconsciousness, and physical healing. Theta (4-8 Hz) brainwaves are associated with a state of relaxation, creativity, insight, and light sleep. They can also be linked to certain types of meditative states accompanied by vivid mental imagery. Alpha (8-12 Hz) brainwaves are associated with a relaxed, calm state of mind and are related to certain types of meditative states characterized by mental imagery. They can also be linked to a state of "alert relaxation". Beta (12-30 Hz) brainwaves are associated with an alert, focused state of mind and are typically produced when you are actively engaged in mental activity such as problem-solving, decision-making, or critical thinking. Gamma (30-100 Hz) brainwaves are the fastest brainwaves and are associated with higher levels of consciousness, spiritual experiences, and moments of insight or inspiration.

BCI directly connects the brain and physical devices and can translate the states of different types of brain activities into commands for various types of operations [3-6, 15, 16]. A number of research works are being carried out using machine learning techniques to study how the brain controls various activities such as movement of the limbs, speech, attention, relaxation and other behaviors [7, 8].

This article covers the following topics: 1) general overview of brain-computer interface, 2) methods for collecting data for research studies and how to preprocess the data, 3) filtering and extracting the features of the collected EEG data, 4) explains the machine learning algorithm of logistic regression, which was optimized for performance and finally, the research findings and conclusions are presented.

2. Methods for Collecting Training Data

The main goal of our study is to control some physical object with commands received as a result of processing a logistic regression algorithm using data from BCI, which transforms the information received from the electrode attached to the scalp with a certain number of channels. The method tested in [2, 3, 8-10] was used to extract differences in human thoughts. The main aim of this method is to present two images to participants in a sequential order with an interval of 6 seconds between them. One image shows the lights turned off, while the other image is either black or blank. Subjects see a light that is off and think it as "on" and then shift their attention to a blank or black image to focus on a different task. During several experiments, three different hypotheses - such as "light off", "light on" and other "normal" variations - were recorded and saved in a file for a specific period of time. We tested 2 types of images: first, the image of a "light on" that is displayed continuously for 6 seconds, and second, a dynamic image of a "light on" that gradually increases the brightness of the light bulb. During the experiment, brain activity was measured on four separate days with the participation of five individuals, four men and a woman, aged 20-42 years. Each participant's recording is separated by a short "break" interval. Each record consists of 10 trials. Each trial consisted of 6 second of "light on" thought followed by 6 second of "normal" thought. Another record involved "light off" thought, followed by 6 seconds of "normal" thought, as shown in Figure 2.

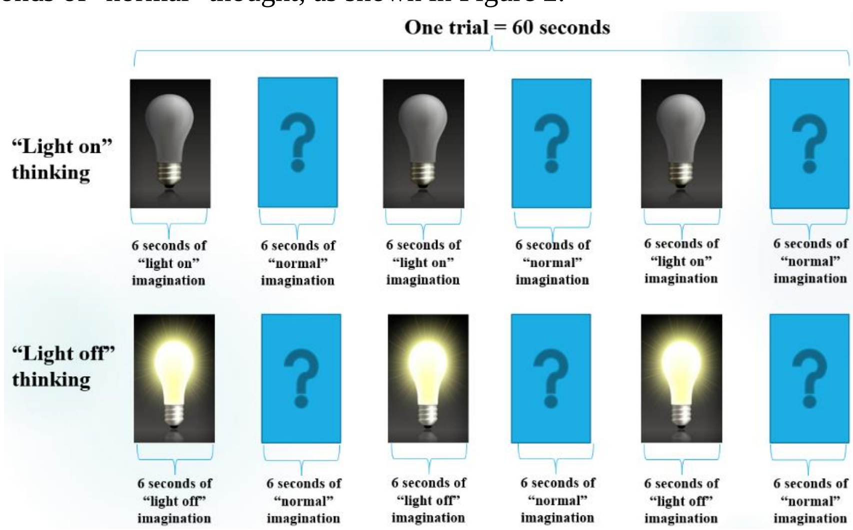


Figure 2. Collecting training data process

The data collected from the Emotiv device, consisting of 14 channels, is transmitted to a computer via a Bluetooth interface. The received data is then saved as a CSV file using Python. The connections of devices used in the test experiments are shown in Figure 3.

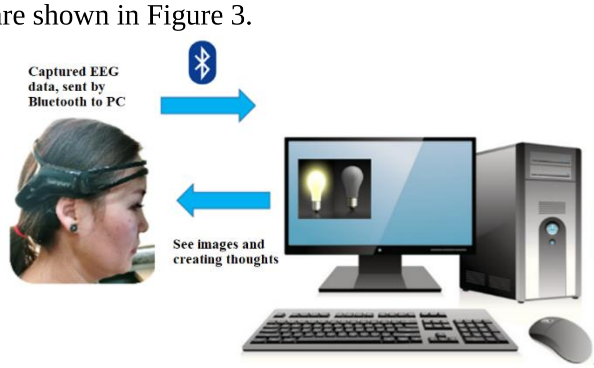


Figure 3. Collecting training data

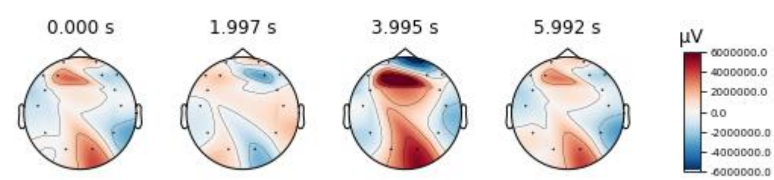
Durations of recorded data in the experiments for each of five participants are: B.B (1784 seconds), D.B (1784 seconds), L.Y (1781 seconds), L.O (1781 seconds), and Ts.T (5356 seconds). The total recording time for all participants was 12486 seconds. Afterward, the data in the file was preprocessed, and signal features were extracted. Finally, machine learning logistic regression was applied to classify the thoughts.

3. Signal processing and feature extraction

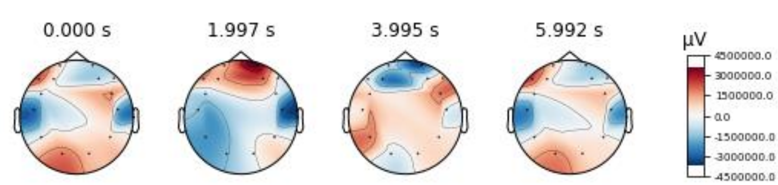
Since recorded EEG signals are noisy, preprocessing is done before analyzing the signal. EEG recordings are affected by the following factors: specific muscle contractions, eye movements or blinks, and external electromagnetic waves [11, 12, 16, 17]. These effects can cause unwanted noise in the EEG recording and distort the EEG results leading to false conclusions. The Emotiv Epoc device samples each channel at a rate of 128 Hz. Usually, BPF is used to remove unwanted signal noise. Therefore, the signals are filtered using BPF of range 0.1 - 40 Hz.

After filtering the signals, we need answers to the questions such as what signals the brain generates during "light on", "light off" or "normal" etc. The signals for thoughts "light on", "light off" and "normal" were recorded for a period of 6 seconds and then processed separately into 12 features and the "light on" and "light off" signals were distinguished from them.

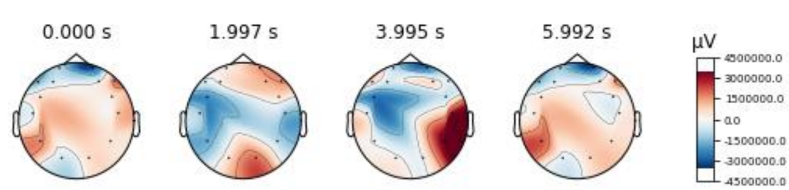
Figure 4 shows topographic map of data points of the brain electrical signals, where (a) displays the average values of 14 data points of the electrical signals recorded during the six-second "light on" thought, while (b) shows the average values of the brain signals recorded during the six-second "light off" thought and (c) displays the average brain imaging data recorded during 6 seconds period, when a black image “normal” was presented.



(a)



(b)



(c)

Figure 4. Topographic map of different thoughts

Based on the image above, it appears that the values of the 14 brain points vary depending on the thoughts at different moments in time and that different parts of the brain become more active when a person thinks about different things, such as "light on". The first step in the EEG classifier is the requirement to extract features from the signal. Usually EEG features are considered in the time or frequency domain. Several methods of feature extraction have been considered in the research. It includes:

Time characteristics

Spectral energy characteristics

Statistical properties

There are many time feature decomposition techniques, among which the most widely used are Eigen Value Decomposition (EVD), Independent Component Analysis (ICA), Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) [1, 2, 13-15].

In this study, we used statistical analysis calculated from each sample, extracting totally of 168 features produced from 12 features × 14 channels for each trial. These 12 features include:

1. mean(x) – average value. The method of finding mean(x) means to consider the average of n=768 values of 1 channel for 6 seconds of recording.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \tag{1}$$

2. std(x) – standard deviation of the mean. In statistics, the standard deviation is a measure of the amount of variability or spread of a set of values around its mean.

$$s_x = \sqrt{\frac{\sum_i^n (x_i - \bar{x}^2)}{n-1}} \tag{2}$$

3. ptp(x) – pick to pick signal amplitude or difference between positive and negative values.

$$V_{pp} = V_{max} - V_{min} \tag{3}$$

4. var(x) is the variance of the random variable x, defined as the expected value of the squared difference between x and its expected value

$$s^2 = \sqrt{\frac{\sum_i^n (x_i - \bar{x})^2}{n-1}} \tag{4}$$

5. minim(x) - minimum value

$$V_{min} = \min\,(x_1, x_2 .. x_n) \tag{5}$$

6. maxim(x) - maximum value

$$V_{max} = \max\,(x_1, x_2 .. x_n) \tag{6}$$

7. argminim(x) - the position of the minimum value
8. argmaxim(x) - the position of the maximum value
9. rms(x) - root mean square. RMS is also called root mean square and is a specific case of general average.

$$x_{rms} = \sqrt{\frac{1}{n}\left(x_1^2 + x_2^2 + \cdots . + x_n^2\right)} \tag{7}$$

10. the sum of the absolute difference

$$x_{abs} = \sum_{i=1}^{n} |x_i - x_{i-1}| \tag{8}$$

11. skewness - is a measure of the asymmetry or unevenness of the probability distribution of a real-valued random variable. The function "skewness(x)" calculates the skewness of the distribution represented by the data in variable x.

$$\tilde{\mu} = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] \tag{9}$$

12. kurtosis - measures how much a data set deviates from a normal or bell-shaped curve.

$$Kurt[x] = E\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] \tag{10}$$

Based on the above features, classification was done by logistic regression.

4. Classification algorithms

In the training phase, offline classification algorithms are trained using EEG data sets of known classification. Then, the unclassified EEG data is sent to the classifier, which predicts the class type. After preprocessing, we divided the data into epochs consisting of 6 seconds. When creating the epochs, experiments were performed with data without any time overlap.

Here, a total of 12 features were generated and the experiment was carried out with 2 methods of classification training. Sequences of realization of the two methods are shown in Figure 5. In the first method, 12 features were calculated for each channel and classified basing on total of 168 features, while in the second method, 12 features were processed for each channel on each of the brain's Delta, Theta, Alpha, Beta, and Gamma frequency ranges, and a total of 840 features were used for classification.

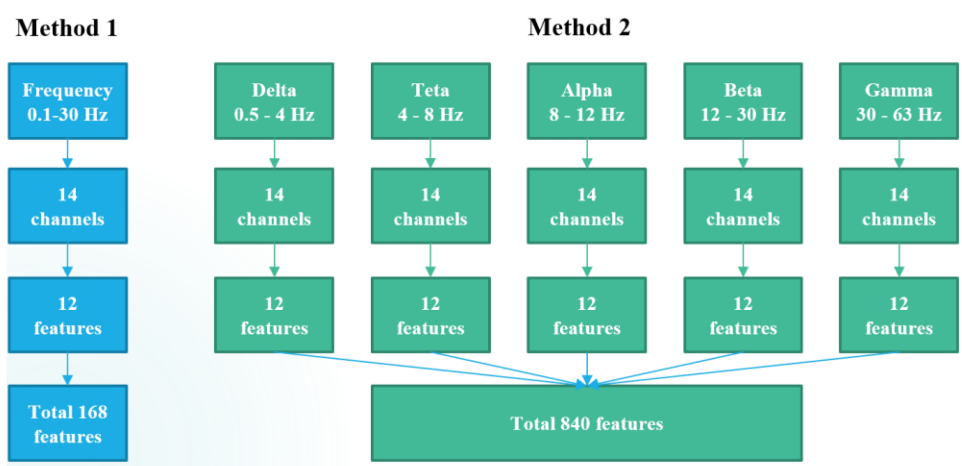


Figure 5. Classification methods

Logistic regression was used in this study to train classification. In our case, we need to predict the output value y, and for logistic regression, y ϵ {0, 1}, that is, we can get two types of output values: 0 or 1. Here, *h* is the hypothesis function, which maps *x* to *y*. The following conditions are taken into account.

$$0 \leq h_\theta(x) \leq 1 \tag{11}$$

A sigmoid function is applied to satisfy condition (11).

$$h_\theta(x) = g(\theta^T x) \tag{12}$$

$$z = \theta^T x = \theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \theta_4 x_4 + \theta_5 x_5 + \theta_6 x_6 + \theta_7 x_7 + \cdots + \theta_{168} x_{168} \tag{13}$$

$$h_\theta(x) = g(z) = \frac{1}{1+e^{-z}} = \frac{1}{1+e^{-\theta^T x}} \tag{14}$$

Here, x_1-x_{168} are the features we calculated.

The function g(z) transforms any real number into the interval (0, 1), making it a useful function for mapping to a probability in a given category. $h_\theta(x)$ gives the probability of our output being 1. To obtain a discrete 0 or 1 class, the output of the prediction function can be converted as follows:

$$h_\theta(x) \geq 0{,}5 \text{ when } y = 1$$
$$h_\theta(x) < 0{,}5 \text{ when } y = 0$$

Since there is a need to classify into three types of outcomes, the “One vs All” approach is used to create categories. We extracted “light on-normal thought”, “light off-normal thought” and “light on-light off” thoughts on each of the 5 participants, and the average training result on a total of 168 features was 82.5%. Below, in Table 1 and Table 2 shown results of Method 1.

Table 1. Classification accuracy result

Frequency range (Hz)	Static image based classification				Dynamic image based classification			
	0.5-30	0.5-40	4.-30	4-60	0.5-30	0.5-40	4.-30	4-60
Light on-off-norm	0.557	0.57	0.58	0.588	0.946	0.95	0.955	0.95
Ligh on-norm	0.71	0.71	0.73	0.725	0.99	0.99	0.979	0.976
Ligh on-off	0.71	0.74	0.71	0.71	0.98	98	0.99	0.99
Ligh off-norm	0.71	0.696	0.756	0.74	0.94	0.94	0.947	0.948

Table 2. Confusion matrix of Method 1

	Predicted "Light on"	Predicted "Light off"	Predicted "Normal"
Actual "Light on"	TP1 - 96	FN1 – 4	FN2 - 4
Actual "Light off"	FP1 - 2	TP2 – 126	FN3 -1
Actual "Normal"	FP2 - 9	FP3 – 2	TP3 -127

The average training results of Method 2 on a total of 840 features was 73.7% and the Table 3 shows confusion matrix.

Table 3. Confusion matrix of Method 2

	Predicted "Light on"	Predicted "Light off"	Predicted "Normal"
Actual "Light on"	TP1 – 62	FN1 – 11	FN2 – 31
Actual "Light off"	FP1 – 20	TP2 – 95	FN3 -14
Actual "Normal"	FP2 - 19	FP3 – 0	TP3 -109

5. Conclusion

In this paper, we presented the simple classification model based on statistical features of EEG signals, classified the data using logistic regression and demonstrated the successful implementation of the BCI based light on/off control system. During the study five individuals participated in the experiments using Emotiv EpocX device to collect brain activity signals from 14 points on the scalp during a cognitive task. Totally have been saved data of 12486 seconds in a .csv file. Each participant's record consists of 3 types of thoughts: “light on”, “light off” and “normal”.

Before conducting machine learning, we processed 12 features from the given data and trained a logistic regression model. By categorizing the input of each user into one of the 3 types mentioned above, namely “light on”, “light off” and “normal”, we were able to achieve an 82.5% accuracy on Method 1. For example, when machine learning was done using one person's thoughts of "turn on" and prediction or testing was done using the thoughts of a second person, the prediction was reduced to 60%, indicating that the perception of "turn on" was different for each participant. 840 feature-based classifications were conducted by computing features at each of the brain's Delta, Theta, Alpha, Beta, and Gamma frequency ranges, but the training performance dropped to 73.7% accuracy on Method 2. From the above it can be concluded that some frequency ranges are unnecessary or some of the 12 main features do not have a significant effect on the classification. When the image is dynamically changed, the classification accuracy increases.

In the future, by deepening this research based on 32 points of the brain, it is possible to control and operate objects aimed at specific goals. A simple example is the design and implementation of a system to solve the problems of people with disabilities.

References

[1] Samaa S. Abdulwahab, “EEG Motor-Imagery BCI System Based on Maximum Overlap Discrete Wavelet Transform (MODWT) and Machine learning algorithm”, Iraqi Journal for Electrical and Electronic Engineering, Vol. 17, December 2021. DOI: **10.1088/1742-6596/1973/1/012056**

[2] L. Bahatti1, A. El Magri1, A. Lekova2, O.Bouattane, “Developing Brain Computer Interface for Motor Imagery Mental Commands”, ISSN 2603-4697 (Online) Complex Control Systems Vol. 2, No 1, 2020, pp. 1-6, pp.

[3] Sandeep Bodda, Shyam Diwakar, “Exploring EEG spectral and temporal dynamics underlying a hand grasp movement”, PLoS One. 2022 Jun 23;17(6):e0270366., **DOI: 10.1371/journal.pone.0270366**

[4] J. Hurtado-Rincon, S. Rojas-Jaramillo, Y. Ricardo-Cespedes, A. M. Álvarez-Meza and G. Castellanos-Domínguez, “Motor imagery classification using feature relevance analysis: An Emotiv-based BCI system”, 2014 XIX Symposium on Image, Signal Processing and Artificial Vision, Armenia, Colombia, 2014, pp. 1-5, **DOI: 10.1109/STSIVA.2014.7010165**

[5] Rodríguez-Bermúdez G, García-Laencina PJ. “Automatic and adaptive classification of electroencephalographic signals for brain computer interfaces”. J Med Syst. 2012 Nov;36 Suppl 1:S51-63. **DOI: 10.1007/s10916-012-9893-4**. Epub 2012 Nov 2. PMID: 23117792.

[6] Reder, E.E., de Quadros Martins, A.R., Ferreira, V.R.T., Kalil, F. (2014). Neural Interface Emotiv EPOC and Arduino: Brain-Computer Interaction in a Proof of Concept. In: Kurosu, M. (eds) Human-Computer Interaction. Advanced Interaction Modalities and Techniques. HCI 2014. Lecture Notes in Computer Science, vol 8511. Springer, Cham. **https://doi.org/10.1007/978-3-319-07230-2_58**

[7] STRMISKA, Martin, Zuzana KOUDELKOVÁ a Martina ŽABČÍKOVÁ. “Measuring brain signals using emotiv devices”. WSEAS Transactions on Systems and Control. 2018, vol. 13, s. 537-542. ISSN 1991-8763

[8] O. Carrera-León, J. M. Ramirez, V. Alarcon-Aquino, M. Baker, D. D'Croz-Baron and P. Gomez-Gil, “A motor imagery BCI experiment using wavelet analysis and spatial patterns feature extraction”, 2012 Workshop on Engineering Applications, Bogota, Colombia, 2012, pp. 1-6, **DOI: 10.1109/WEA.2012.6220084**.

[9] Szczepan Paszkiel, “Brain–computer technology-based training system in the field of motor imagery”, IET Science, Measurement & Technology 14, December 2020, **DOI: 10.1049/iet-smt.2019.0522**

[10] Ekansh Sareen, Anubha Gupta, “Studying functional brain networks from dry electrode EEG set during music and resting states in neurodevelopment disorder”, **http://dx.doi.org/10.1101/759738doi**: posted online Sep. 8, 2019;

[11] V.Asanza, “SSVEP-EEG Signal classification based on Emotiv Epoc BCI and Raspberry PI”, IFACT conference paper, 388 -393 pp., 2021, **https://doi.org/10.1016/j.ifacol.2021.10.287**

[12] Tat’y Mwata-Velu, Jose Ruiz-Pinales “Motor Imagery Classification Based on a Recurrent-Convolutional Architecture to Control a Hexapod Robot”, Mathematics 2021, 9, 606. **DOI: 10.3390/math9060606**

[13] Marquos Zaki, Ali Alquraini, “Home Automation using EMOTIV: Controlling TV by Brainwaves”, Journal of Ubiquitous Systems & Pervasive Networks Volume 10, No. 1 (2018) pp. 27-32, **DOI: 10.5383/JUSPN.10.01.004**

[14] Craik A, He Y, Contreras-Vidal JL. Deep learning for electroencephalogram (EEG) classification tasks: a review. J Neural Eng. 2019 Jun;16(3):031001. **DOI: 10.1088/1741-2552/ab0ab5**. Epub 2019 Feb 26. PMID: 30808014.

[15] Luján, M.Á.; Jimeno, M.V.; Mateo Sotos, J.; Ricarte, J.J.; Borja, A.L. “A Survey on EEG Signal Processing Techniques and Machine Learning: Applications to the Neurofeedback of Autobiographical Memory Deficits in Schizophrenia”, Electronics 2021, 10, 3037. **https://doi.org/10.3390/electronics10233037**

[16] William O, “Handbook of EEG interpretation”, 2008 Demos Medical Publishing, LLC, ISBN-10: 0826147089

[17] Shoorangiz, Reza, Stephen J. Weddell, and Richard D. Jones. "EEG-Based Machine Learning: Theory and Applications", Handbook of Neuroengineering. Singapore: Springer Singapore, 2021. 1-39, **https://doi.org/10.1007/978-981-15-2848-4_70-1**

[18] Toshihisa Tanaka and Mahnaz Arvaneh, "Signal Processing and Machine Learning for Brain--Machine Interfaces", The Institution of Engineering and Technology is registered as a Charity in England, 2018, ISBN: 9781785613982, **DOI: 10.1049/PBCE114E**