

OFFLINE LINEAR DISCRIMINANT ANALYSIS CLASSIFICATION OF TWO CLASS EEG SIGNALS

Mohammed Z. Al-Faiz ¹, Haider A. Abdulkareem ²

¹ Al-Farahidi University, Baghdad, Iraq

² Department of Electrical and Communications Engineering, URUK University, Iraq
mzalfaiz12@gmail.com ¹, haiderasadi@daad-alumni.de ²

Abstract- This paper investigates the use of LDA algorithm in the EEG classification. EEG feature extraction is implemented to reduce the dimensionality of data. The Sliding Window Technique is used to calculate the mean within each window samples. Then, classification is done based on hyperplane technique. The LDA algorithm is described in details with all the implementation issues. The LDA regularization is also discussed and its effects on the classification accuracy is given. In addition, both window size and channel selection effect on the accuracy is illustrated. Results show that a window size of 150 samples, channel 3 and regularization parameter of 0.9 gives an accuracy of 90%.

I. INTRODUCTION

In the last two decades, the research focus on Brain- Computer Interface (BCI) applications has been increased. A BCI application is defined as the action of recording the Electroencephalography (EEG) activity of the brain. Then, the recorded data is filtered from noise, features are extracted and classified to perform a predefined action (such as opening or closing an artificial arm) [1]. Since the EEG data is large (1 sec of EEG data can have up to 256 time samples), feature extraction is needed, which is the first phase of the EEG classifier. Feature extraction is mandatory because most classifiers include matrix operations, and when a matrix is large in size, it is called an ill-conditioned matrix. The inverse of such matrices includes large numerical errors. Therefore, reducing the data size is needed. Usually, common EEG features are extracted in time or frequency bands. The second step of the classifier includes two phases: training and recognition. In the training phase, EEG data sets of known classes are used to train the classification algorithm offline. Then, the unclassified EEG data is fed to the classifier and a decision is made (determining which class the EEG samples belong to). The classification decision is then fed to the implemented hardware and the required action is done (moving an artificial arm or driving a wheelchair). Many feature reduction and classification techniques have been considered with EEG based Motor Imagery (MI) applications. Several feature extraction techniques have been considered in the literature [2] :

- 1) time features.
- 2) spectrum energy features.
- 3) statistical features.

Time features can be extracted by many approaches, most commonly used are: Eigen Value Decomposition (EVD), Independent Component Analysis (ICA), Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA). Because BCI applications requires real- time or semi real- time approaches, time features are the best candidate regarding computation time and complexity, compared to other feature extraction approaches. Many researchers investigated the use of different feature extraction and they showed promising results. [3] used PCA as feature reduction and Support Vector Machine (SVM) as a classifier, to reach an accuracy of 97.5%. [4] discussed the small size problem of the feature matrix,

for which the number of time features is largely higher than the number of channels. It reaches an accuracy of 84% for two-class problem. [5] used Common Spatial Pattern (CSP) filtering along with LDA to reach a maximum accuracy of 99%. [6] showed that xDAWN algorithm accuracy outperforms both ICA and PCA. [7] and [8] demonstrated an accuracy comparison between different spectral feature extraction techniques, including Power Spectral Density (PSD), time-frequency energy distributions, periodogram, Spectrogram and Morlet Wavelet features. In this work, an offline EEG-MI classifier is built based on Linear Discriminant Analysis (LDA). It includes both steps mentioned above: feature extraction and classification. This work focuses on all implementation issues of the LDA classifier as well as the related theoretical background. The paper is organized as follows. In Section II, the LDA feature extraction is described. Then, in Section III the limitations of LDA are discussed with proposed solutions. The LDA classification algorithm is described in details. In Section IV, the setup procedure as well as the recording process is described. The algorithm parameters are varied and the corresponding accuracy are shown in figures.

II. LDA FEATURE EXTRACTION

Since the EEG signal time samples show no obvious difference between different MI commands, it is not possible to use classification directly on time samples. Fig. 1 depicts EEG time samples for two classes. It can be seen that there is no obvious classification pattern between the two classes. Therefore, feature extraction is needed, on which the classification accuracy relies on.

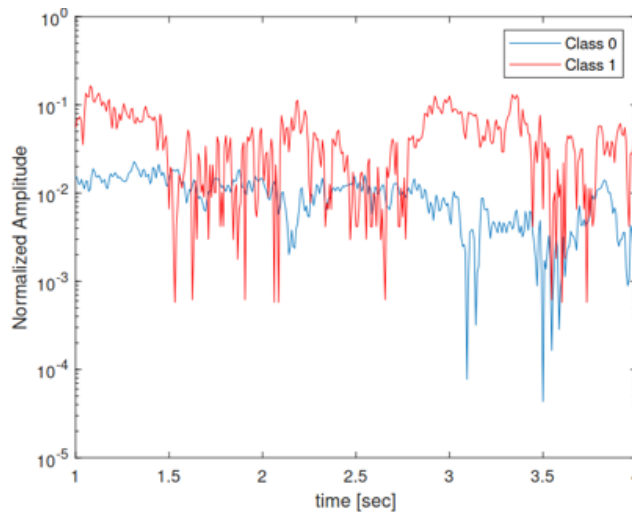


Figure 1: Logarithmic scale plot of EEG time samples of single channel for two classes

In this paper, LDA feature extraction is used because it introduces good separability between different EEG classes and hence simple linear classifier can be used. Before being able to use time features in LDA, it is mandatory to reduce the feature size. This is attained using sliding window technique (SWT). That is, for a single trial vector, the mean is calculated for 300 sample and a single mean value is stored. The operation is repeated for the next group of samples.

As a result, a single command trial of 640 samples (5 seconds) can be represented by just 5 features (instead of 640), resulting in more accurate feature representation and non-singular matrix operations. The concept of the SWT is described as follows: Since the EEG data is stationary within the window, therefore, the single mean value of the window is sufficient to express all time features within that window. Sometimes, Root Mean Square (RMS) value is used instead. In order to be able to extract features from EEG time samples, dimensionality reduction is needed. It removes redundant features by transforming from higher order to lower order space, allowing matrix manipulation operations to be performed more easily. Many dimensionality reduction techniques exist, such as Independent Component Analysis (ICA), Non-negative Matrix Factorization (NMF), Mixture Discriminant Analysis (MDA), and LDA. The latter is the most used approach [2]. This technique maximizes the between-class variance (dissimilarity between classes) and minimizes the within-class variance (dissimilarity within the same class). As a result, maximum separation between two classes is achieved. This separation is performed on the EEG features from both classes (Class 0 and Class 1) and saved as Training Features. New extracted features (unclassified) are projected against the training features. The classification is done by assigning the class that gives the minimum distance (least error) to either Class 0 or Class 1 features. The schematic flow chart of LDA is illustrated in Fig. 2.

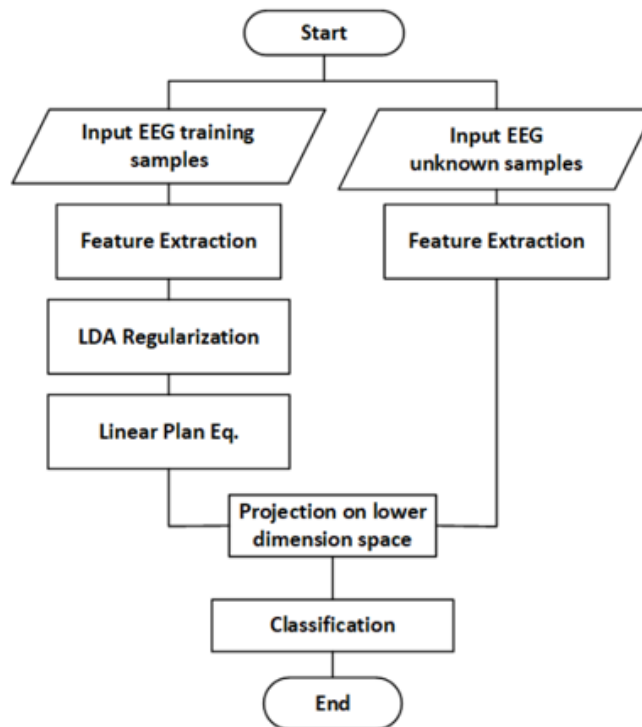


Figure 2: LDA algorithm flowchart

III. LINEAR DISCRIMINANT ANALYSIS ALGORITHM

LDA can be described as a projection of the EEG window time samples onto a lower dimensional space [9]. This projection maximizes the distance between the two classes and therefore a correct classification can be made. In order to describe LDA algorithm, the following assumptions about data samples are made. Each trial EEG data samples are stored as a matrix $X(ch, t) = \{x_1, x_2, \dots, x_T\}$, where the rows M are the channels and the columns T are the time samples:

$$X(ch, t) = \begin{bmatrix} x_{1,1} & x_{1,2} & \dots & x_{1,T} \\ x_{2,1} & x_{2,2} & \dots & x_{2,T} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ x_{M,1} & x_{M,2} & \dots & x_{M,T} \end{bmatrix} \quad (1)$$

As mentioned in Section II, data reduction must be conducted on the time samples. Using a sliding window of size t, the mean of each t samples is calculated and stored in a new matrix $Y(ch, t) = \{y_1, y_2, \dots, y_I\}$. Therefore, a total of I features for each channel is extracted, where $I < T$:

$$y(ch, t) = \begin{bmatrix} y_{1,1} & y_{1,2} & \dots & y_{1,I} \\ y_{2,1} & y_{2,2} & \dots & y_{2,I} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ y_{M,1} & y_{M,2} & \dots & y_{M,I} \end{bmatrix} \quad (2)$$

for a single channel LDA approach, the matrix $Y(ch, t)$ is reduced to a vector $y(t) = y_1, y_2, \dots, y_I$.

The next step of feature extraction is to stack the row features of each EEG trial into a column vector. This step results in matrix $F_i(t, n) = \{f_1, f_2, \dots, f_N\}$, where N is the number of trials:

$$F_i(t, n) = \begin{bmatrix} F_{1,1} & F_{1,2} & \dots & F_{1,N} \\ F_{2,1} & F_{2,2} & \dots & F_{2,N} \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ F_{I,1} & F_{I,2} & \dots & F_{I,N} \end{bmatrix} \quad (3)$$

After feature extraction, the LDA algorithm is applied on the extracted feature matrices of both classes. LDA can be summarized in three basic steps:

- 1) Between- Class Variance, S_{Bi} . It represents the distance between the projection of the mean of $class_i$, μ_i , and the total mean, μ_T . The mean μ_i of all trails of class is calculated as follows:

$$\mu_i = \frac{1}{N} \sum_n F_i(t, n) \quad (4)$$

and the total mean:

$$\mu_T = \frac{1}{2} \sum_{i=1}^2 \mu_i \quad (5)$$

where $\mu_i, \mu_T \in \mathcal{R}^I$, Therefore, S_{B_i} is calculated as:

$$S_{B_i} = N_i (\mu_i - \mu_T) (\mu_i - \mu_T)^T \quad (6)$$

where N_i is the total number of trials in $class_i$. The total Between- Class Variance, S_{BT} , is then calculated as:

$$S_{BT} = \sum_{i=1}^2 S_{B_i} \quad (7)$$

2) Within- Class Variance, S_{W_i} . It represents the difference between the features of $class_i$, F_i , and the mean of $class_i$, μ_i . S_{W_i} is calculated as:

$$S_{w_i} = (F_i - \mu_i) (F_i - \mu_i)^T \quad (8)$$

and the total Within- Class Variance, S_{WT} is calculated as:

$$S_{WT} = \sum S_{W_i} \quad (9)$$

3) Lower dimensional space projection. LDA algorithm searches for a solution that maximizes the total Between-Class Variance, S_{BT} . Hence, the separation distance between classes is increased. In addition, it tries to minimize the Within- Class Variance S_{WT} . This optimization problem is called Fisher's criterion, described in Eq. 10

$$J(W) = \arg \max_W \frac{W^T S_{BT} W}{W^T S_{WT} W} \quad (10)$$

A solution to maximize the objective function $J(W)$ is formulated as:

$$S_{WT} W = \lambda S_{BT} W \quad (11)$$

where λ are the Eigen values of the projection matrix W . Then:

$$W = S_{WT}^{-1} S_{BT} \quad (12)$$

After sorting the Eigen vectors in a decreasing order, according to λ , the largest Eigen vector, V is selected. The LDA hyperplane, used to classify EEG features is formulated as [10]:

$$a = V \cdot \mu_T \quad (13)$$

$$b = -a^T \mu_T \quad (14)$$

where a is a vector and b is a scalar. The classification step is simply done by evaluating the sign of the scalar variable Class in Eq. 15 for each of the input feature vectors F_j . If the sign is positive, it is assigned to class- 1, otherwise, class- 0 is chosen.

$$Class = \text{sign} (a^T F_j + b) \quad (15)$$

LDA suffers from two problems, the first one is called Small Sample Size (SSS), in which the number of samples is smaller than the dimensions. The second problem arises when the two classes are not linearly separable, that is, the mean of the two classes is approximately equal. The solution to the nonlinearity problem is done by mapping the samples into a newer space using a Radial Basis Function (RBF). Example is Gaussian RBF, defined as:

$$K(x_i, x_j) = \exp\left(-\|x_i - x_j\|^2 / 2\sigma^2\right) \quad (16)$$

For the SSS problem, the identity matrix is multiplied by a regularization parameter η . The result is then added to the within-class matrix, S_{WT} , to become non-singular:

$$S_{WT} = (1 - \eta) S_{WT} + \eta \frac{Tr(S_{WT})}{M} \Gamma_M \quad (17)$$

where M is the size of the squared S_{WT} matrix.

LDA algorithm is summarized as follows:

- 1) Read EEG time samples ($10 \text{ trials} \times 640 \text{ samples}$) (Eq. 1)
- 2) Feature extraction: for each trial(i):
 - Calculate mean of samples in window (Eq. 4)
 - Store in new matrix (Eq. 2), Each Class- 0 and Class- 1 are matrices ($10 \text{ trials} \times 6 \text{ time features}$)
- 3) Apply steps (2- 3) to test trials (test trials are $10 \text{ trials} \times 6 \text{ features}$)
- 4) **Transpose:** stack each trail matrix into one column vector (Eq. 3)
- 5) **Mean:** for each class matrix, calculate mean (column- wise) (Eq. 4) and the total mean (Eq. 5)
- 6) **Between Class Variance:**
 - S_{Bi} (Eq. 6)
 - S_{BT} (Eq. 7)
- 7) **Within Class Variance:**
 - Class- 0 Covariance Matrix ($6 * 6$) (Eq. 8)
 - Class- 1 Covariance Matrix ($6 * 6$) (Eq. 8)
 - Total Covariance Matrix ($6 * 6$) (Eq. 9)
- 8) **LDA Regularization** (Eq. 17)
- 9) Calculate separating plan (Eq. 13)
- 10) **Classification:** For each trial(i)
 - Test = trial(i) (6×1)
 - $polarity = sign$ (Eq. 15)
 - if $polarity > 0$: $class(i) = 1$
 else $class(i) = 0$
- 11) Calculate Accuracy.

IV. IMPLEMENTATION AND RESULTS

The EEG dataset was collected using EMOTIV EPOC headset, which provides 14 channels. The subject was asked to sit at rest state for 10 to 20 minutes prior to recording. This allows the relaxation of the brain activities in general and the accuracy of recorded EEG data. The subject is then asked put both of his hands rest on a table in front of him, while his eyes are opened. Then, a cue is shown to him to imagine the movement of his right hand (Close and Opening). While he closes and opens his right hand repeatedly, the EEG data is recorded for 5 seconds at a rate of 128 samples per second. Those trails are imported to MATLAB. Each trail consists of 640 samples. Trial recording is repeated until 10 trials are recorded and stored. These trials are named as Class Right training. The same procedure is repeated for Class Left training samples. MATLAB 2018a environment is used to record, store and implement the LDA algorithm. After the training data is acquired, the features representing each class are extracted and stored as matrices. The subject then performed either left- or right- hand simple movement (closing and opening of each hand) repeatedly for five seconds. The choice of which hand to move is random. During the subject session, the EEG signal is recorded at a sampling rate of 128 samples per second. This step is repeated for 20 trails with different random choices for either left or right hand of the subject. For each trial, the LDA algorithm then extracts the features from the EEG data set and those features are compared against the training features extracted during the training phase. According to the sign of Eq. 15, the winning class result is recorded. The accuracy of the algorithm is calculated as:

$$Accuracy = \frac{no.ofcorrectclassification}{totalnumberoftrials} \quad (18)$$

As it can be seen from Fig. 3, the best channel accuracy occurs at channel 3, which complies with F3 sensor. This sensor is the nearest to the MI region in the brain, related to EMOTIV EPOC headset (see Fig. 4). The regularization parameter (η) selection plays a role in increasing the accuracy of the inverse of the matrix. Several values were tested for η against accuracy, as shown in Fig. 5. The best accuracy occurs between [0.85, 0.9]. Fig. 6 shows the effect of windows size (in samples) on the classification accuracy. It can be seen that a window size of approximately 1.2 seconds gives the best accuracy. In another word, the mean value for every 150 samples represents certain EEG signal. On the other hand, increasing the window size beyond 175 samples will decrease the accuracy. The reason for this accuracy degradation is that larger window size will overlap with adjacent EEG samples, therefore, the unique time feature characteristics will be destroyed. As a result, an EEG recording of 1.5 seconds is sufficient for LDA classification. Therefore, the user does not have to repeat EEG emotion within a single trial.

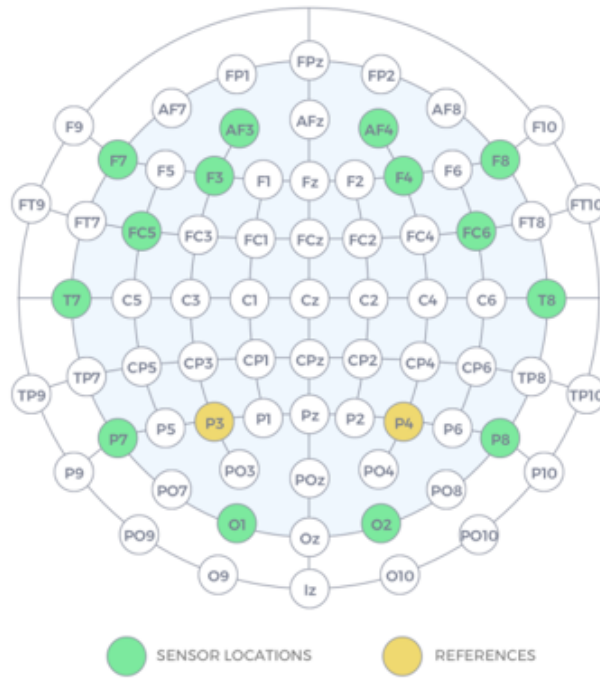


Figure 3: EMOTIV *EPOC+* channels locations [14]

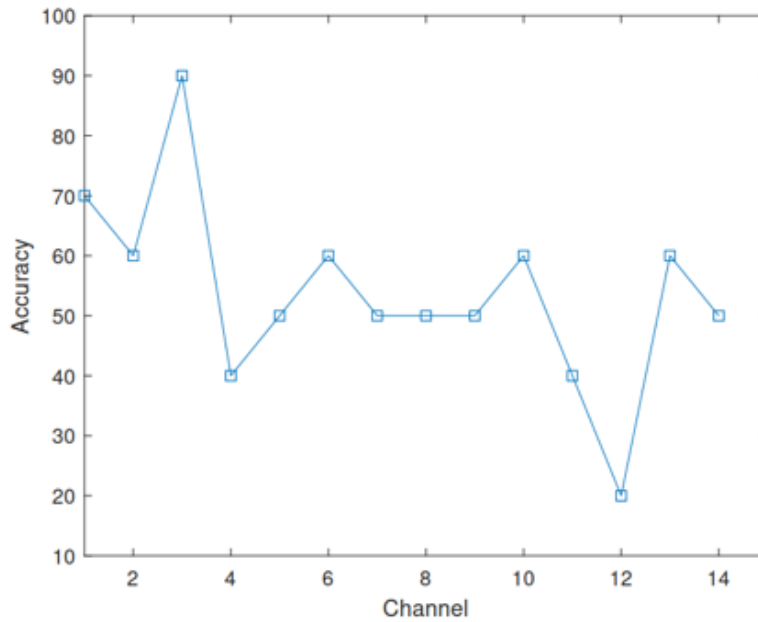


Figure 4: Channel selection vs.accuracy $\eta = 0.9$, window- size=150 samples

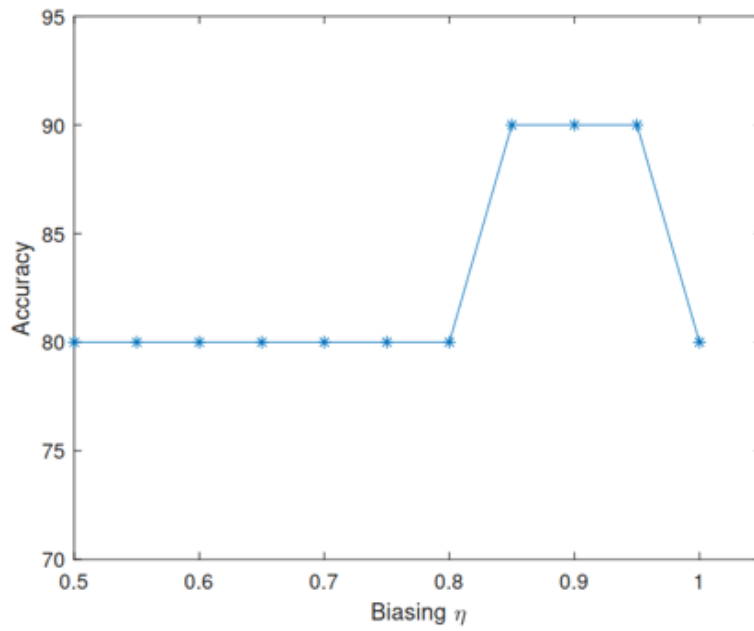


Figure 5: η vs.accuracy , ch=3, window- size=150 samples

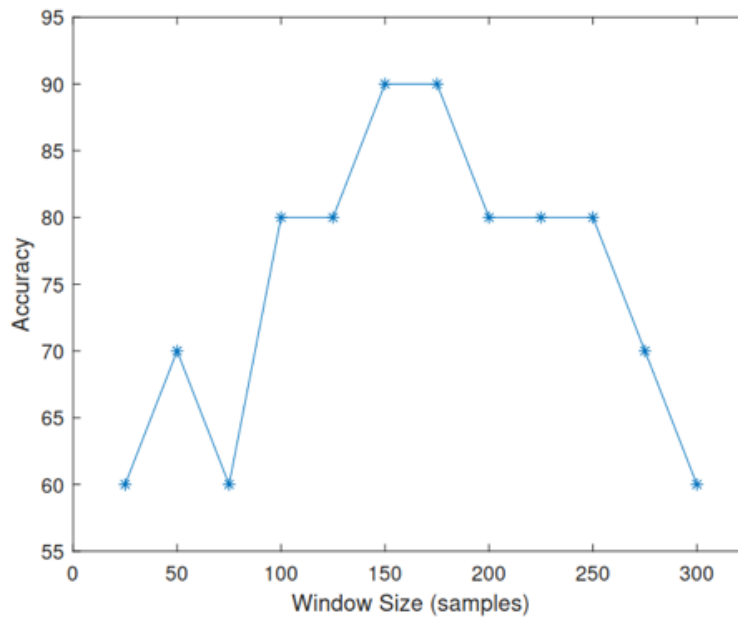


Figure 6: Window- size vs.accuracy , $\eta = 0.9$,ch=3

V. CONCLUSIONS

In this paper, EMOTIV EPOC headset is used to capture and store EEG data from brain. The EEG feature extraction procedure is described. Then, LDA algorithm used in EEG classification is depicted in low- level details. This work covered all the mathematical and programming implementation details of LDA, in both stages of extraction and clarification. In addition, the ill-conditioned matrix problem and solution is described. It has been found that the classification accuracy is maximum when channel three is selected. Moreover, the algorithm attains its best classification when the window size is 150 samples. Furthermore, the biasing parameter η , used to solve the ill- condition problem, is found to score best result when it is selected between 0.85 and 0.95.

REFERENCES

- [1] F. Lotte, "A Tutorial on EEG Signal Processing Techniques for Mental State Recognition in Brain- Computer Interfaces. Eduardo Reck Mirandam Julien Castet. Guide to Brain- Computer Music Interfacing" , 2014.
- [2] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy and F. Yger, " A review of classification algorithms for EEG- based brain- computer interfaces: A 10 year update" , J. Neural Eng. , vol. 15, no. 3, pp. aab2f2, 2018.
- [3] M. Z. A. Faiz and A. A. Al- Hamadani, " Online Brain Computer Interface Based Five Classes EEG To Control Humanoid Robotic Hand" , 2019 42nd International Conference on Telecommunications and Signal Processing (TSP), 2019.
- [4] B. Blankertz, S. Lemm, M. Treder, S. Haufe, and K. - R. Muller, " Single- trial analysis and classification of ERP components - A tutorial" , NeuroImage, vol. 56, no. 2, pp. 814- 825, 2011.
- [5] B. Rivet, A. Souloumiac, V. Attina, and G. Gibert, " xDAWN Algorithm to Enhance Evoked Potentials: Application to Brain-Computer Interface" , IEEE Trans. Biomed. Eng. , vol. 56, no. 8, pp. 2035- 2043, 2009.
- [6] N. Brodu, F. Lotte, and A. Lecuyer, " Comparative study of band-power extraction techniques for Motor Imagery classification" , 2011 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), 2011.
- [7] N. Brodu, F. Lotte, and A. Lecuyer, "Comparative study of band-power extraction techniques for Motor Imagery classification" , 2011 IEEE Symposium on Computational Intelligence, Cognitive Algorithms, Mind, and Brain (CCMB), 2011.
- [8] P. Herman, G. Prasad, T. McGinnity, and D. Coyle, " Comparative Analysis of Spectral Approaches to Feature Extraction for EEG- Based Motor Imagery Classification" , IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol. 16, no. 4, pp. 317- 326, 2008.
- [9] A. Tharwat, T. Gaber, A. Ibrahim, and A. E. Hassanien " Linear discriminant analysis: A detailed tutorial" , AI Communications, vol. 30, no. 2, pp. 169- 190, 2017.
- [10] F. Lotte, " Signal processing approaches to minimize or suppress calibration time in oscillatory activity-based brain-computer interfaces" , Proc. IEEE, vol. 103, no. 6, pp. 871- 890, 2015.
- [11] H. Ramoser, J. Muller- Gerking, and G. Pfurtscheller , " Optimal spatial filtering of single trial EEG during imagined hand movement" , IEEE Transactions on Rehabilitation Engineering, vol. 8, no. 4, pp. 441- 446, 2000.
- [12] C. Vidaurre, M. Kawanabe, P. V. Bunau, B. Blankertz, and K. R. Muller, " Toward Unsupervised Adaptation of LDA for Brain- Computer Interfaces" , IEEE Transactions on Biomedical Engineering, vol. 58, no. 3, pp. 587- 597, 2011.
- [13] L. C. Parra, C. D. Spence, A. D. Gerson, and P. Sajda, " Recipes for the linear analysis of EEG" , Neuroimage, vol. 28, no. 2, pp. 326- 341, 2005.
- [14] EMOTIV EPOC User Manual, <https://www.192-168-1-1-ip.co/manuals/13886.pdf>.
- [15] E. D. U. A. R. D. O. R. E. C. K. MIRANDA, Guide to brain-computer music interfacing. Place of publication not identified: SPRINGER LONDON LTD, 2016
- [16] BRAIN-COMPUTER INTERFACES HANDBOOK: technological and theoretical advances. S.I.: CRC PRESS, 2018.