2021 18th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)
Mexico City, Mexico November 10-12, 2021

EEG motor/imagery signal classification comparative using machine learning algorithms

1st Alicia Guadalupe Lazcano-Herrera
School of Engineering and Sciences
Tecnologico de Monterrey
Mexico City, Mexico
alicia_gpe_hg@hotmail.com

2nd Rita.Q. Fuentes-Aguilar
School of Engineering and Sciences
Tecnologico de Monterrey
Mexico City, Mexico
rita.fuentes@tec.mx

3rd Mariel Alfaro-Ponce
School of Engineering and Sciences
Tecnologico de Monterrey
Mexico City, Mexico
marielalfa@gmail.com

***Abstract*—Electroencephalography (EEG) study allows the recording of brain activity associated with different mental tasks through electrodes placed on the scalp that amplifies the electricity changes at neurons activity. Because of the nature of EEG signals, their interpretation and classification require an expert. Recently, machine learning algorithms for EEG analysis have gained popularity and are applied in various activities such as brain-computer interfaces (BCI), diagnostic of brain disorders, etc. In this work, an EEG classification was performed with different machine learning algorithms. For this, Support Vector machines (SVM), K-nearest neighbor (KNN), Quadratic Discriminant Analysis (QDA), Linear Discriminant Analysis (LDA), Naive Bayes (NB) and Ensemble were implemented and the performance of different algorithms when distinguishing between two classes: one of movement and one of inactivity. The movement class was composed of Motor Imagery(MI) data and actual movement and inactivity class of a baseline. From the proposed techniques for EEG classification, the QDA and NB achieve the highest accuracy.**

***Index Terms*—Classification, EEG, Machine Learning.**

I. Introduction

EEG is the recording of the brain electrical activity, and it is obtained by placing electrodes on the surface of the scalp [1]. The state of brain activity can be assessed by examining the energy in frequency bands [2]. These frequency bands are called Alpha, Theta, Delta, Gamma and Mu, and they work in the order of 0.5 Hz to above 50 Hz [3] .

In EEG, Movement Imagery (MI) [4] is the mental execution of a movement (such as grasping or stomping) without actually performing the movement, as means without the execution of the muscular activity. That MI has been a fundamental part of the development of assistive rehabilitation technologies such as BCI [5]. These assistive technologies are composed of devices, products, or pieces of equipment used to increase, maintain, or improve the independence of people with different levels of disability.

To achieve a classification, exist different techniques distinguished by their domain.Exist the techniques related to the evoked activity, these are focused on the extraction of information from the frequency bands, such as the average band power, the relative band power, and the event-related potentials. One of the techniques related to spontaneous activity is Power Spectral Density (PSD) . This technique is defined as the rate at which motor units are triggered. You can also find techniques associated with frequency, such as the Welch method [6]. Time-frequency techniques, among which Fourier Transforms (STFT) [6], Wavelet Transforms (WT) [7] and other methods that represent signal information can be observed, such as Entropy, a measure of uncertainty and represents the number of bits necessary to represent certain information of the signal [8].

BCI systems are popular as support for people with disabilities, but it should be noted that these systems can also be used for non-biomedical applications such as industry, transportation, and entertainment. Many of these systems rely on machine learning algorithms to be able to process the EEG signals. Machine Learning algorithms can be classified [9] as supervised learning [10], unsupervised [11], and reinforcement of learning [9]. Each of these types of learning has different learning objectives; for example, some of the supervised learning algorithms are better for classification and regressions for data analysis. These learning algorithms have tools such as Neural Networks (NN) [12], Deep Learning (DL) [9] , Bayesian Networks (BN) [12] and Decision Trees [13], among others. Many works have approached the implementation of different classifiers based on machine learning algorithms to solve the problem of EEG signals classification. The work of [12] mentions some EEG algorithms which can perform the EEG classification task. Machine learning algorithms can also be classified through a ”family”; the differences between each family rely on their mathematical method for classifying or how they compute the data.

The family of Linear Models contains the LDA, QDA, and Logistic Regression. It’s a widely used [13] family of machine learning algorithms for dimension reduction and feature selection. Bayesian classifiers (BC) [12] is a family of simple probabilistic classifiers based on applying Bayes’ theorem, with strong independence assumptions between the features. For these assumptions, it’s called ”naive” bayesian. Support Vector Machines is an algorithm with a kernel approach [13]. A kernel is a support vector that helps to separate classes in a hyperplane. This type of algorithm has probed good accuracy for tasks that involves classification and regression. KNN family [12] classifies new values based on similarity. These algorithms search the nearest similar data to the previously trained ones and work better with a small size of datasets.

978-1-6654-0029-9/21/31.00©2021$IEEE$

Ensemble classifiers [14] are used in large datasets to increase the classification accuracy, stacking different classifiers in order to minimize the errors of classification committed by the previous classifier.

The present work compares the accuracy of different machine learning algorithms to classify EEG signals by using different features extracted from the MI dataset [15]. The input of the different machine learning algorithms (SVM, LDA, QDA, KNN, etc.) is 10 features extracted from the original signal consisting of 6 power bands and 4 time-frequency domain characteristics. These features proceed from ten subjects-12 trial sample of MI Physionet data [15] set. A dimensional reduction is then performed as proposed in [16] to lower the computational cost and increase the performance of the algorithms. The paper organization is as follows: In Section II a description of the proposed EEG classification method based on machine learning algorithms is established; Section III displays the results of the comparison between the proposed algorithms, and finally, in Section IV conclusions are presented.

II. Materials and Methods

The methodology followed in this work consists in : Data recollection, pre-processing, feature extraction, feature selection and signal classification, as can be seen in Figure 1 .

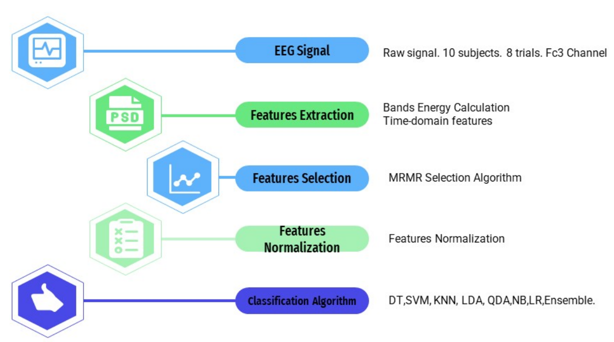

Fig. 1. Methodology employed

The hardware section employed a Dell Inspiron 15 3000 with a 10th generation Intel Dual Core i3-1005G1 processor (1.2 GHz, up to 3.4 GHz, 4 MB cache, two cores). 8 GB DDR4 SDRAM, 128 GB solid-state drive, 1 TB HDD, Intel UHD graphics. A Matlab R2020a version of the software was used, with the corresponding toolboxes. The Classification Learner application included with the mentioned version was used for the implementation of Machine Learning algorithms. It was chosen together with the programming environment, the EEGLAB toolbox [17], 14_1b version, which is an interactive Matlab toolbox to process continuous EEG data related to events as well as other electrophysiological data, and their extension of pre-pipeline that it's a tool focused in the pre-processing process.

A. Database

The database selected for their support information and size was the Physionet Motor-Imagery dataset [15].This database was used in the previous classification works [9] with different features extracted from the raw information. This allows comparing the performance of the present work with them. The complete database consisting of 1500 recordings of 1-2 minutes was obtained from 109 volunteers on 64 EEG channels using the BCI2000 system. Each subject performed 14 experimental runs, with the following characteristics: two one-minute baseline runs (one with eyes open, one with eyes closed) and three experimental runs of two minutes of each of the following four tasks:

1) Open and close the left or right fist
2) Imagine opening or closing the left or right fist
3) Open and close both fists or both feet
4) Imagine opening and closing both fists or both feet

For the present work, not the full database was employed, only a sub-set of 10 experimental subjects. The selection of the size was made taking into consideration a previous study [18]that achieved a classification using a small sample and to lower the computational cost. A 10% subject size can represent the information. Also, the sub-set is composed of information corresponding to the tasks and trials of opening and closing the upper limbs and imagery of the movement and the baseline. This information could lead to an upper limb rehabilitation system in the future. These ten experimental subjects contain 12 trials each. Three trials correspond to the MI, 3 to the performance of the movement, and six trials to a baseline (3 for eyes open and 3 for eyes closed), giving a total of 120 trials. These trials will compound the training set for the two classes Movement(For MI and Movement) and Inactivity.

B. Pre-processing

Before starting with the feature extraction, it is important to remove the artefacts that may contaminate the EEG signals to facilitate its reading, interpretation, and analysis. Various processes can accomplish this, including filtering, re-referencing, segmentation, channel removal, signal damage testing, and more. Filtering is the most precise procedure for removing artefacts that interfere with the signal. Using the EEGLab [17] plugin for Matlab, the pre-processing procedure of the new data set was performed.

The baseline was removed to eliminate signal compensation which provided a new reference. At the same time, a 60Hz line noise filter was applied to remove the noise caused by connecting the system to the electrical supply. This frequency corresponds to the frequency used in the United States and North America territory, which is the territory of origin of the database. Later, a 0.5 High-Pass filter was applied to correct the baseline.

C. Feature Extraction

The methodology employed for feature extraction and its selection is suggested in [16]. It is composed of data acquisition, feature extraction, and feature selection previous to a classification process, as can be seen in Figure 1.

Feature extraction is the first step to classification an EEG signal because it reflects the signal information that will allow us to distinguish between the different proposed classes. In previous works of EEG classification [9] is mentioned that correlation between imagery and motor signals can contribute to the performance of a BCI [19]. In Figure, 2 has portrayed the PSD feature of 3 tasks: Baseline, Movement of upper limbs, and Imagery of upper limbs, and is observable that Movement and Imagery could integrate the movement class without compromising it and could lead to a binary classification reducing the computational cost in the classification step and with a good performing in a BCI.

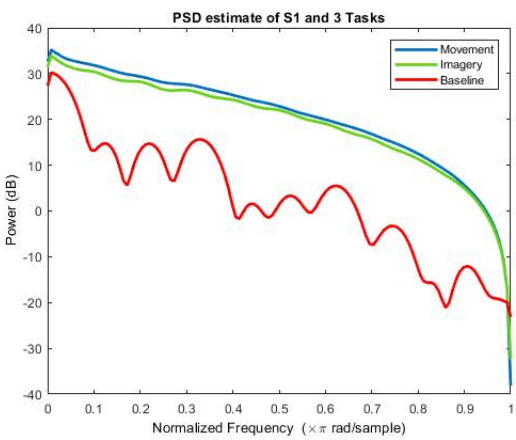

Fig. 2. PSD for S1 and 3 tasks

In this work, features were extracted and selected based on their association with the cerebral motor region to differentiate between the two proposed classes and consider the previously mentioned works. At first, the signals' extraction of the power bands Delta, Theta, Alpha, Beta, Gama, and Mu (an alphoid associated with motor events) was established.

In terms of features that reflected the randomness of the signals and could be helpful to distinguish between the two classes, the following were taken into account: Spectral Power Density (PSD), Shannon Entropy, Spectral Entropy and Discrete Wavelet Transform (DWT). These features were extracted for Fc3. This channel was considered due to its location in the 20-10 electrode positioning system and also its in a region associated with motor activity. This information can be found condensed in Figure 3.

D. Feature Selection

One of the goals of feature selection is to reduce computational time and increase system accuracy. In a previous review, [9], the systems that included the selection procedure for the output of characteristics achieved greater precision than the BCIs that did not involve that process.

There are different types of approaches to feature selection: manual selection, statistical selection, filter selection, probabilistic, and metaheuristics. In filter selection, for example,

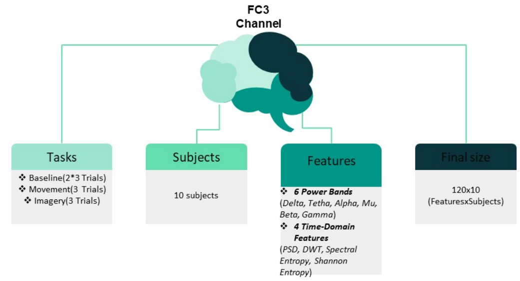

Fig. 3. Composition of Sample of Physionet MI [15]

exist an algorithm called MRMR (Minimum Redundancy Maximum Relevance) that reduces redundant data within the database and maximizes the relevant information to classify [20]. Metaheuristics is a procedure in which one tries to respond to an optimization model through heuristics. Heuristics are search algorithms used to select features in large databases because they can handle the dimensionality problems that such databases imply. Into these metaheuristic algorithms are genetic algorithms and differential computing.

Due to the size of the database used in the present study, an MRMR algorithm was employed to select the data and classify it. The feature selection through algorithms that can adapt according to the type and content of the databases has as an example the algorithm called Linear Discriminant Analysis (LDA). This algorithm [21] is the evolution of a statistical method designed to distinguish between classes of plants. It is defined by the probability of the joint discrete/categorical variables on their distribution and is expressed in (1).

$$I(x, y) = \sum_{i,j} p\left(x_i, yj\right) \log \frac{p\left(x_i, y_j\right)}{p\left(xi\right) p\left(yj\right)} \tag{1}$$

Then, the level of "similarity" or minimum redundancy is expressed by (2) where "S" represents the desired features and $|S|$ the number of features in the features subset.

$$\min W_1, W_1 = \frac{1}{|S|^2} \sum_{i,j \varepsilon S} I(i, j) \tag{2}$$

And (3) quantifies the relevance of certain characteristics for the classification task.

$$\max V_1, V_1 = \frac{1}{|S|} \sum_{i \varepsilon S} I\left(h, i\right) \tag{3}$$

E. Classification Methods

Once the feature selection finishes, the results (the best features) become the input of the collection of classification algorithms. In the present, the classification process was developed to distinguish between two classes: Inactivity and Movement. First, all the features available (10 features) and then the best-ranked pair of features for the system in the feature selection process. To classify signals adequately, algorithms should contain the following characteristics: Be able to deal with variant patterns, prevent overfitting, and adequate fast [12].The algorithms that are portrayed are: Decision Tree (Fine, Medium and Coarse), Linear Discriminant, Quadratic Discriminant, Logistic Regression, Naive Bayes (Gaussian and Kernel), Support Vector Machines(Linear, Quadratic, Cubic, Fine Gaussian, Medium Gaussian, Medium Gaussian, Coarse Gaussian) K-Nearest Neighbor (Fine, Medium, Coarse, Cosine, Cubic, Weighted) and a series of Ensemble algorithms like Boosted Trees, Bagged Trees, Subspace Discriminant, Sub-spaced KNN, RuBoosted Trees. The Figure 4 displays a representation of the Classification process.

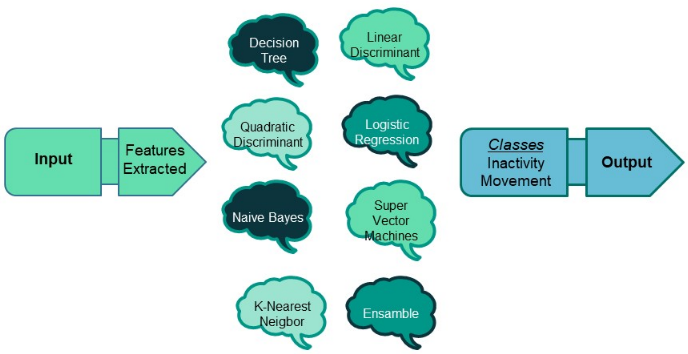

Fig. 4. Classification process

III. Results

To compare the impact and relevance of diverse features in the accuracy of the classification of the database using the proposed machine learning algorithms, the feature reduction approach was taken, making a feature selection. With all of the previously extracted features, a classification analysis was carried out using the Matlab classification learner toolbox to find out which would be the best algorithm to achieve an accurate classification. To carry out this analysis, the MRMR showed that the best characteristics to feed the selection algorithms for this data set are those corresponding to the Shannon Entropy and the Discrete Wavelet (DWT) for FC3 channel information. In Fig 5 the importance of these characteristics within the set can be appreciated.

Table I compares the accuracy for machine learning algorithms of all the available features and pairs of features chosen by their importance for the system. The first column portrays the best percentage of accuracy and corresponds to all the available features. The last column shows the features (Shannon Entropy and Discrete Wavelet Transform pair) that compete with the first column performance. A significant difference in accuracy can be observed depending on the conforming of the selected pairs. The pairs of features which algorithms accuracy compete with the use of all the extracted features are the ones with major relevance for the system.

In Table II it can be found the confusion matrix that shows the performance of the training data corresponding to the training for a QDA for all the features and the pair corresponding to Shannon/DWT. These features were selected by the results of MRMR that highlights them as the features with maximum relevance for the system. The QDA algorithm has a perfect score for the database sample; this is because of the size of the database, QDA performance is better in small-size databases. Also, QDA is outstanding for feature selection and dimension reduction. The performance of this pair of features is similar to the one displayed for the QDA training model for all the features.

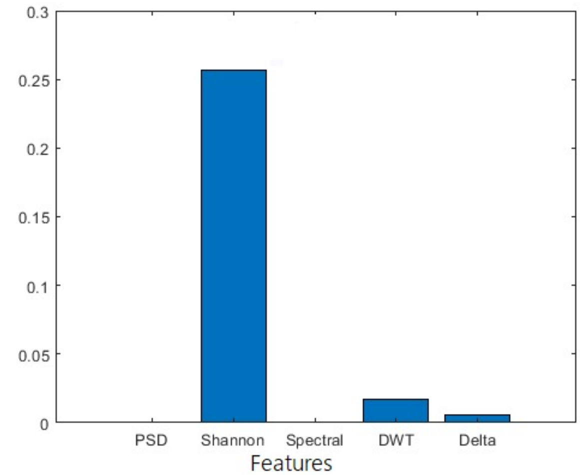

Fig. 5. Ranking of features according to their relevance to the system

Table III is a comparison between the machine learning algorithm training for all the features and the previously mentioned pair. It is focused on their training time and prediction speed because this parameter portrays the computational cost.

IV. Conclusions and Discussion

Observing the table of comparisons of the features and algorithms, the machine learning algorithms with the highest accuracy are the QDA and the BN algorithms; these algorithms work better with small data sizes. All the machine learning proposed algorithms have similar performance, but the independence between features increases the distance in a hyperplane between data. Consequently, the prediction and classification task became difficult for SVM and KNN algorithms, which significantly impacts accuracy and feature selection.

The results confirmed that machine learning algorithms could achieve a high classification accuracy for EEG data. The facts suggest that the classification accuracy depends on the features of the system. High accuracy can be achieved using just a pair of features instead of all the features; also, the computational cost reduces, but these features need to be the more relevant pair for the system. The use of redundant pairs of features for the system lows the accuracy and highs the misclassification cost. In some EEG classification, previous studies [9] it is mentioned the feature selection or dimension

TABLE I
MI CLASSIFICATION WITH MACHINE LEARNING ALGORITHM COMPARISON

		Accuracy (%)				
Type	***Classifier***	***All****	***PSD/Alfa***	***Shannon/Mu***	***Delta/Alfa***	***Shannon/ DWT***
Decision Tree	Fine tree	96.7	82.5	85.8	65	96.7
	Medium tree	96.7	82.5	85.8	65	96.7
	Coarse tree	96.7	85	80	65	96.7
Linear Discriminant		80	77.5	78.3	60	75
Quadratic Discriminant		100	76.7	79.2	60.8	99.2
Logistic Regression		90.8	78.3	80.8	58.3	77.5
Naïve Bayes	Gaussian Naive Bayes	100	75.8	79.2	60.8	99.2
	Kernel Naive Bayes	100	79.2	78.3	61.7	100
Support Vector Machines	Linear SVM	81.7	75	75.8	62.5	73.3
	Quadratic SVM	94.2	82.5	82.5	63.3	91.7
	Cubic SVM	95	85.8	84.2	61.7	73.3
	Fine Gaussian SVM	87.5	85	89.2	65	91.7
	Medium Gaussian SVM	89.2	80	78.3	64.2	95.8
	Coarse Gaussian SVM	80	75	75	65.8	88.3
K-Nearest Neighborgh	Fine KNN	50	83.3	85	65.8	75
	Medium KNN	83.3	86.7	83.3	66.7	97.5
	Coarse KNN	50	50	50	50	93.3
	Cosine KNN	83.3	82.5	75.8	60	50
	Cubic KNN	85	85.8	83.3	65.8	89.2
	Weighted KNN	94.2	82.5	86.7	70	93.3
Ensemble	Boosted Trees	50	50	50	65	94.2
	Bagged Trees	86.7	78.3	88.3	74.2	50
	Subspace Discriminant	82.5	77.5	78.3	60.8	96.7
	Subspace KNN	76.7	68.3	69.2	59.2	75
	RusBoosted Trees	50	50	50	61.7	50

*Features: PSD, Shannon Entropy , Spectral Entropy, DWT and Power Bands (Delta, Theta, Beta, Alpha, Gamma and Mu).

TABLE II
QUADRATIC DISCRIMINANT ANALYSIS CONFUSION MATRIX FOR COMPARISON

	Quadratic Discriminant Analysis			
True Class	All		Shannon	
Ina	100%		98.30%	1.7%
Mov		100%		100%
	Predicted Class			

TABLE III
QUADRATIC DISCRIMINANT ALGORITHM COMPARISON

Quadratic Discriminant Analysis		
	All	Shannon/DWT
Prediction Speed	˜300 obs/sec	˜450 obs/sec
Training Time	62.907	28.931 sec

reduction as a relevant process previous to classification because of the size of the data set and the features available to classify.

Another conclusion obtained in the present study is that machine learning algorithms have a good performance in EEG classification tasks, but this is related to the dataset and the feature extraction process and dimension reduction. The present classification can be integrated as the mentioned dimension reduction step in a further investigation using machine learning algorithms like NN or DNN.

The contribution of the present works is the comparative of performance of diverse Machine Learning algorithms for EEG MI classification. The performance of them is impacted by the feature extraction and selection process, so the feature extraction techniques and features like Shannon entropy or Spectral entropy were not reported in previous studies [22] The limitations of the study are the ones focused on the computational cost because the present work doesn't use all the available Physionet database or tasks corresponding to upper limbs. Also, there is the usage of a database that is not new or with data recollected by us. At least the classification algorithm accuracy was not probed by a testing set.

REFERENCES

[1] G. Y. Wiederschain, "Biomedical engineering principles," *Biochemistry*, vol. 71, no. 5, p. 581, 2006.

[2] A. B. Ritter, V. Hazelwood, A. Valdevit, and A. N. Ascione, *Biomedical engineering principles*. Crc Press, 2011.

[3] J. J. Carr and J. M. Brown, *Introduction to biomedical equipment technology*. Prentice hall, 2001.

[4] T. Mulder, "Motor imagery and action observation: cognitive tools for rehabilitation," *Journal of neural transmission*, vol. 114, no. 10, pp. 1265–1278, 2007.

[5] I. Lazarou, S. Nikolopoulos, P. C. Petrantonakis, I. Kompatsiaris, and M. Tsolaki, "Eeg-based brain–computer interfaces for communication and rehabilitation of people with motor impairment: a novel approach of the 21st century," *Frontiers in human neuroscience*, vol. 12, p. 14, 2018.

[6] L. Hu and Z. Zhang, *EEG Signal Processing and Feature Extraction*. Springer, 2019.

[7] E. Da Silva and D. Sampson, "Successive approximation wavelet vector quantization for image and video coding," in *Advances in Imaging and Electron Physics*, 1996, vol. 97, pp. 191–255.

[8] H. Misra, S. Ikbal, H. Bourlard, and H. Hermansky, "Spectral entropy based feature for robust asr," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1. IEEE, 2004, pp. I–193.

[9] A. Al-Saegh, S. A. Dawwd, and J. M. Abdul-Jabbar, "Deep learning for motor imagery eeg-based classification: A review," *Biomedical Signal Processing and Control*, vol. 63, p. 102172, 2021.

[10] P. P. Cruz, *Inteligencia artificial con aplicaciones a la ingeniería*. Alfaomega, 2011.

[11] I. N. Da Silva, D. H. Spatti, R. A. Flauzino, L. H. B. Liboni, and S. F. dos Reis Alves, "Artificial neural networks," *Cham: Springer International Publishing*, p. 39, 2017.

[12] E. J. Rechy-Ramirez and H. Hu, "Bio-signal based control in assistive robots: a survey," *Digital Communications and networks*, vol. 1, no. 2, pp. 85–101, 2015.

[13] N. Padfield, J. Zabalza, H. Zhao, V. Masero, and J. Ren, "Eeg-based brain-computer interfaces using motor-imagery: Techniques and challenges," *Sensors*, vol. 19, no. 6, p. 1423, 2019.

[14] W. Yi, S. Qiu, H. Qi, L. Zhang, B. Wan, and D. Ming, "Eeg feature comparison and classification of simple and compound limb motor imagery," *Journal of neuroengineering and rehabilitation*, vol. 10, no. 1, p. 106, 2013.

[15] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals," *circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[16] H. U. Amin, W. Mumtaz, A. R. Subhani, M. N. M. Saad, and A. S. Malik, "Classification of eeg signals based on pattern recognition approach," *Frontiers in computational neuroscience*, vol. 11, p. 103, 2017.

[17] A. Delorme, T. Mullen, C. Kothe, Z. Akalin Acar, N. Bigdely-Shamlo, A. Vankov, and S. Makeig, "Eeglab, sift, nft, bcilab, and erica: new tools for advanced eeg processing," *Computational intelligence and neuroscience*, vol. 2011, 2011.

[18] M. H. Alomari, A. Samaha, and K. AlKamha, "Automated classification of l/r hand movement eeg signals using advanced feature extraction and machine learning," *arXiv preprint arXiv:1312.2877*, 2013.

[19] H. Sugata, M. Hirata, T. Yanagisawa, K. Matsushita, S. Yorifuji, and T. Yoshimine, "Common neural correlates of real and imagined movements contributing to the performance of brain–machine interfaces," *Scientific reports*, vol. 6, no. 1, pp. 1–11, 2016.

[20] C. Ding and H. Peng, "Minimum redundancy feature selection from microarray gene expression data," *Journal of bioinformatics and computational biology*, vol. 3, no. 02, pp. 185–205, 2005.

[21] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Annals of eugenics*, vol. 7, no. 2, pp. 179–188, 1936.

[22] S. M. Khan, A. A. Khan, and O. Farooq, "Selection of features and classifiers for emg-eeg-based upper limb assistive devices—a review," *IEEE reviews in biomedical engineering*, vol. 13, pp. 248–260, 2019.