A STUDY OF THE CLASSIFICATION OF MOTOR IMAGERY SIGNALS USING MACHINE LEARNING TOOLS

Anam Hashmi, Bilal Alam Khan and Omar Farooq

Department of Electronics Engineering,
Aligarh Muslim University, Aligarh, India

ABSTRACT

In this paper, we propose a system for the purpose of classifying Electroencephalography (EEG) signals associated with imagined movement of right hand and relaxation state using machine learning algorithm namely Random Forest Algorithm. The EEG dataset used in this research was created by the University of Tubingen, Germany. EEG signals associated with the imagined movement of right hand and relaxation state were processed using wavelet transform analysis with Daubechies orthogonal wavelet as the mother wavelet. After the wavelet transform analysis, eight features were extracted. Subsequently, a feature selection method based on Random Forest Algorithm was employed giving us the best features out of the eight proposed features. The feature selection stage was followed by classification stage in which eight different models combining the different features based on their importance were constructed. The optimum classification performance of 85.41% was achieved with the Random Forest classifier. This research shows that this system of classification of motor movements can be used in a Brain Computer Interface system (BCI) to mentally control a robotic device or an exoskeleton.

KEYWORDS

EEG. Machine learning. BCI. Motor Imagery signals. Random Forest.

1. INTRODUCTION

The last decade has witnessed some tremendous advancements particularly in the field of medicine and technology. The inter-dependence of the two said fields is becoming more and more pronounced day by day; virtual surgical theatre, robotic surgery, Brain-controlled wheelchair are the name of the few recent developments. Nowadays, the study of biomedical signals has caught the attention of researchers as it provides the avenue for efficient disease diagnosis, development of assistive technologies, health monitoring of the elderly and aiding humanity in general [1]. This study explores further this very dimension by analyzing different methodologies used in studying Brain-Computer Interface (BCI). Electroencephalogram or EEG is one of the most common non- invasive methodologies of BCI to record brain signals. It measures the electrical activity of the brain using electrodes that are placed over the scalp. EEG is preferred because of its ease of portability and capturing high temporal brain information, however, it fails in capturing high spatial information [2]. BCI uses these EEG signals associated with the user's activity and then apply different signal processing algorithms for translating the recorded signals into control commands for different applications. In an EEG there are five types of oscillatory waves that are commonly used for analysis, which are:

David C. Wyld et al. (Eds): VLSIA, INWES, CSIP, BMLI - 2021
pp. 23-33, 2021. CS & IT - CSCP 2021 DOI: 10.5121/csit.2021.112104

(a) delta (0.5–4 Hz);
(b) theta (4–7 Hz);
(c) alphaormu (7–13 Hz);
(d) beta (13–25 Hz);
(e) gamma (25–50 Hz).

Motor imagery (MI) is a process in which an individual rehearses or stimulates an action. It is a very popular paradigm in the analysis of an EEG based BCI system. MI activity usually lies in alpha (or mu) and beta bands [3].

In the past few years, significant advances have been made in the BCI systems and they have revolutionized rehabilitation engineering by providing the differently-abled individuals with a new avenue to communicate with the external environment. According to many works of literature, the strength of a BCI system depends upon the methods in which the brain signals are translated into control commands of machines. A novel method namely an arc detection algorithm to find an optimal channel was proposed by Erdem Ekran and Ismail Kurnaz [4]. For feature extraction DWT was used and a number of machine learning algorithms were used for classification purposes, which were SVM, K- nearest neighbor, and Linear Discriminant Analysis. The best accuracy achieved by their methodology was 95% in classifying ECoG signals (BCI competition III, dataset I). Jun Wang and Yan Zhao proposed feature selection based on one dimension real-valued particle swarm optimization, extracted nonlinear features such as Approximate entropy and Wavelet packet decomposition, and achieved the best accuracy of 100% [5]. Aswinseshadri. K et al. used the wavelet packet tree for feature extraction. They used genetic algorithm, applied information gain, and mutual information to find the best feature set and for classification K-NN and Naïve Bayes were employed [6]. Chea-Yau Kee et al. proposed a novel feature known as Renyi entropy that has been employed for feature extraction and BLDA for classification [7]. K. Venkatachalam et al. proposed the use of the Hybrid-KELM (Kernel Extreme Learning Machine) method based on PCA (Principal Component Analysis) and FLD (Ficher's Linear Discriminant) analysis for MI BCI classification of EEG signals. The best accuracy reported was 96.54% [8]. Rajdeep Chatterjee et al. used the AAR (Auto Adaptive Regressive) algorithm for feature extraction, proposed a novel feature selection method based on IoMT (Internet of Medical Things), and classified EEG signals using SVM and ensemble variants of classifiers. The best accuracy reported was 80% [9]. The authors of [10] employed a combination of common spatial patterns (CSP) and local characteristic- scale decomposition (LCD) algorithm for feature extraction, a combination of firefly algorithm and learning automata (LA) to optimize feature selection, and spectral regression discriminant analysis (SRDA) classifier for classifying MI-EEG signals. They have used this method for a real- time brain-computer interface in order to show their method's efficiency.

Most of these studies have worked on the classification of right vs left-hand movement, or hand vs tongue movements, or hands vs legs movements. There are very few works that have studied and classified intricate hand movements such as opening and closing of a hand, or movements of different fingers, or classification of different hand gestures using neural signals, and those who have worked on these subjects either did not achieve high enough accuracy or failed to work in a real-world setup. This paper probed this very aspect of studying intricate human motions and worked on the classification of imagining of opening and relaxing of a hand using MI-EEG signals.

The contributions of this paper are following:

(i) Accurate classification of the motor imagery signals using a very simple algorithm design which is general in nature and hence can be used for other physiological signal classification.

(ii) Identification and ranking of most important features, from which it can be observed that although using a lesser number of features may not seem intuitive but in reality it has improved the classification accuracy.

The organization of this paper is as follows: the first part is the Introduction stage, where a brief introduction was provided and related works were reported, followed by Materials and Methodology stage. In this part, the materials or data that was used in the paper is described and the methodology of this work was elucidated. The third stage involves the results of the study with detailed discussion followed by conclusion.

2. MATERIALS AND METHODOLOGY

2.1. Data Used

The data used in this study was taken from [11]. The data consist of EEG recordings of a single subject. The subject was connected with a high spinal cord lesion and was controlling an exoskeleton (Brain-Neural computer interface) attached to his paralysed limb. The cue-based BNCI paradigm consisted of two different tasks, namely the 'imagination of movement' of the right hand (Class 1) and 'relaxation/no movement' (Class 2).

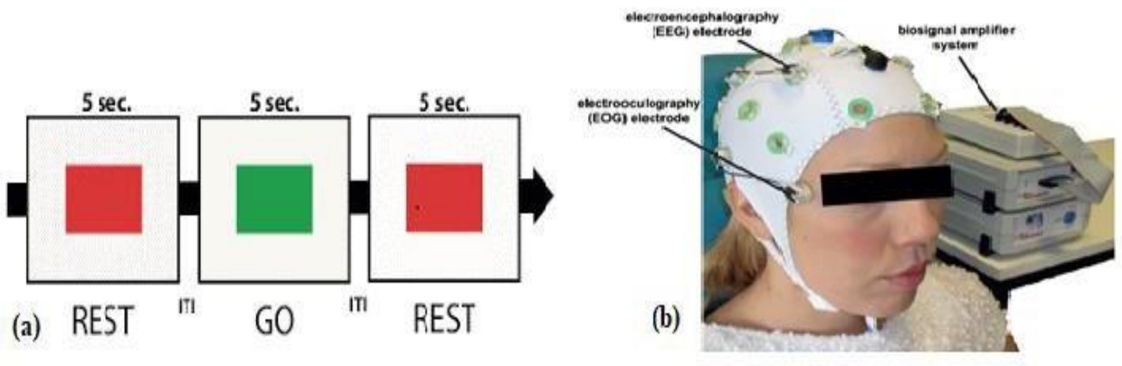

Figure 1. Timing Scheme of the trials used (a) and a subject during the EEG recording of the Dataset (b).

A randomly shown visual cue is used to indicate to the user when to open (for Green square) and when to close (for Red Square). These two indications were given 24 times each in total separated by inter-trial intervals (ITIs) of 4-6 seconds. Each indication was displayed for 5 seconds after which the device was driven back to open position. Re-setting the exoskeleton into op3en position required one second.

EEG was recorded from 5 conventional EEG recording sites F4, T8, C4, Cz, and P4 according to the international 10/20 system using an active electrode EEG system (Acti-cap® and BrainAmp®, BrainProducts GmbH, Gilching, Germany) with a reference electrode placed at FCz and ground electrode at AFz. EEG was recorded at a sampling rate of 200 Hz, bandpass filtered at 0.4-70Hz and pre-processed using a small Laplacian filter.

2.2. Pre-Processing

At this stage, the data was processed or filtered to capture information related to Motor Imagery. For that, the wavelet decomposition was done to obtain four-level details. The wavelet transformation of the EEG record at four levels resulted in four detail coefficients and one approximate coefficient with the frequency ranges listed in Table 1. Many electrophysiological features are associated with the brain's normal motor output channels [12]. Some of these important features are the mu (8-12 Hz) and beta (13-30 Hz) rhythms [13]. We concluded from Table 1 that the details cD2, cD3 and cD4 provide proper representation for the mu and beta rhythms and we decided to extract the vectors of features from these details.

Table 1. Frequency range for the decomposed details and approximation

Signal Component	**Frequency Range(Hz)**
cD1	50-100
cD2	25-50
cD3	13-25
cD4	7-13
cA4	0-7

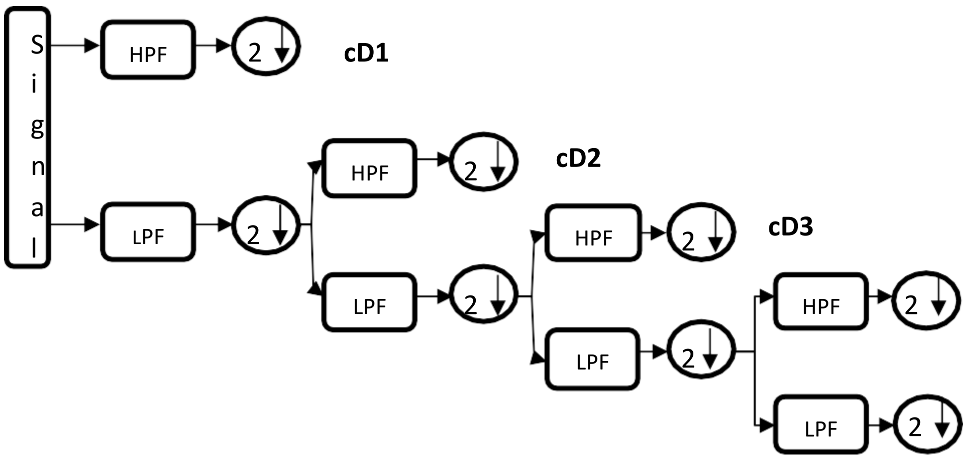

Figure 2. Showing decomposition of EEG signal into different wavelet coefficients corresponding to various frequency bands using wavelet transform.

2.3. Feature Extraction

A feature is a measurable property or characteristic of an observed signal. It should be informative, discriminative and orthogonal to other features. Feature extraction is the method of extracting these features. It can be defined as the process of transforming original data into a dataset with a reduced number of variables but with the most discriminative information.

After the wavelet transforms, the channels F4, T8, C4, Cz and P4 of each EEG record was analyzed using the Daubechies wavelet. Then the features namely Interquartile Range (IQR), Median Absolute Deviation (MAD), Variance, Skewness, Kurtosis, Energy, Mean Absolute Value (MAV) and Standard Deviation were calculated. The choice of these particular features can be understood as these two classes are different. Particularly, the two classes – which corresponds to imagining of opening of hand as 'class 1' and relaxation or no movement as 'class

2' – differ in dispersion. The same can be observed from the histogram of class 1 and class 2 appears, where the class 1 histogram appears to be skewed from the normal distribution. Thereby justifying the choice of IQR, MAD, Variance, Standard Deviation, Skewness and Kurtosis. Energy and MAV were chosen because it has been reported in many works that mu rhythm has a lower amplitude than that of the alpha wave [14].

The following are the mathematical equations of the extracted features:

2.3.1. IQR

It is a measure of statistical dispersion, being equal to the difference between 75th and 25th percentiles, or between upper and lower quartiles. Mathematically, it is defined as:

$$IQR = Q3 - Q1 \tag{i}$$

Where, Q3 and Q1 represents the 75^{th} and 25^{th} percentiles of the distribution.

2.3.2. Median Absolute Deviation

It is defined, as the name suggests, median value of the absolute deviations from the data median value.

$$MAD = median(|Xi - median(X)|) \tag{ii}$$

Where, Xi is the ith value of the data X.

2.3.3. Variance

It is defined as the expectations of the squared deviation of a random variable from its mean.

$$Var(X) = E[(X - \mu)^2] \tag{iii}$$

Where, Var(X) computes of variance of data X, 'μ' represents the average value, 'E' represents the expectation.

2.3.4. Skew

It is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean.

$$\gamma = E\left[\left(\frac{X - \mathrm{u}}{\sigma}\right)^3\right] \tag{iv}$$

Where, 'ɣ' represents the skewness of data X.

2.3.5. Kurtosis

It is a measure of 'tailedness' of the probability distribution of a real-valued random variable.

$$Kurt(X) = E\left[\left(\frac{X - \mathrm{u}}{\sigma}\right)^4\right] \quad \text{(v)}$$

2.3.6. Energy

It is the area under the squared magnitude of the considered signal. Mathematically,

$$Es = \sum_{n=-\infty}^{\infty} |X(n)|^2 \quad \text{(vi)}$$

2.3.7. Mean Absolute Value

It is defined as the mean value of the absolute values of the data. Mathematically,

$$MAV = \frac{1}{N}\sum_{i=1}^{N} |X_i(n)| \quad \text{(vii)}$$

2.3.8. Standard Deviation

It is a measure that is used to quantify the amount of variation or dispersion of a set of data values. It can be defined as,

$$SD = \sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(\left(x[n] - \frac{1}{N}\sum_{n=1}^{N} X[n]\right)^2\right)} \quad \text{(viii)}$$

2.4. Methodology

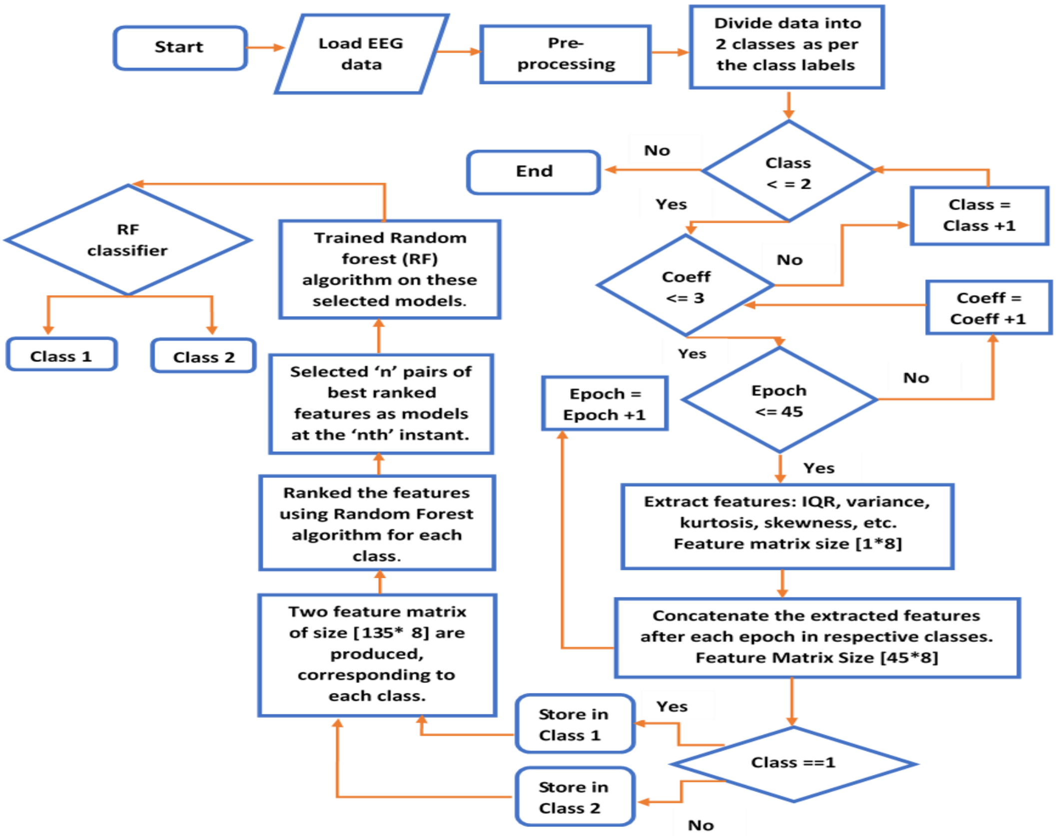

Figure 3. Showing the flowchart of the used algorithm for classification

In this study, there are two classes corresponding to the Motor Imagery (MI) tasks; hand opens and hand relaxes. The total duration of 'Class 1' is 45 seconds and the sample rate is 200 Hz, which produces 9000 data points. The data was pre-processed and decomposed into multiple bands using wavelet transform. Those bands were considered which corresponds to the Mu rhythm (8 – 13 Hz) and Beta rhythm (13 – 40 Hz) as these bands contain motor imagery related brain activity. Therefore, cD2, cD3, cD4 are chosen. A one-second sliding window is considered for the analysis of the signal and for each second 200 samples are considered, for which 8 features were extracted. This process was performed until the end of the recording, thereby producing a feature matrix of the size of [45×8]. This was done for a single coefficient of the wavelet-decomposed signal. Therefore, for all three coefficients, the produced feature matrix size was [135×8]. A similar analysis was performed for the 'Class 2' as well which also produced the feature matrix of size [135×8]. After the feature extraction, these features were made to pass through the feature selection stage where each feature is ranked (or given importance). The Figure 3. shows the flowchart of the used algorithm.

2.5. Feature Selection

This section describes the feature selection stage for the classification of MI-based EEG. Before the classification of signals is done, there are many features that do not provide any extra information than the currently selected features and are known as redundant features. Feature selection ranks the extracted feature based on information content that each feature adds to classify the two classes. As a result, it removes redundant features and improves the computational cost of the system. In this study, the Random forest based feature selection method is used. This selection method consists of a few hundred decision trees and each decision tree is built using a random extraction of the observations from the dataset and a random extraction of the features. The trees are de- correlated and less prone to over-fitting as every tree does not see all the features or all the observations. Based on a single or combination of features each tree represents a sequence of yes- no questions. At every node, the dataset gets divided into 2 groups, each of them consisting of observations that are more similar among themselves and different from the ones in the other group. Therefore, the importance of each feature is derived from how "pure" each of the group is. When a tree is trained, it is possible to measure the decrease in an impurity by each feature and consequently, the more important feature. In random forests, the final importance of the variable can be determined by aggregating (majority vote) the impurity decrease from each feature across several tees. For classification, the measure of impurity is either Gini impurity or entropy (Information gain). The ranked features using random forest are shown in Fig 4. It can be seen that Mean Absolute Value (MAV) has the highest feature importance while Kurtosis has the lowest. This suggests that MAV and Energy should be the best features while Skewness and Kurtosis depict the redundant features.

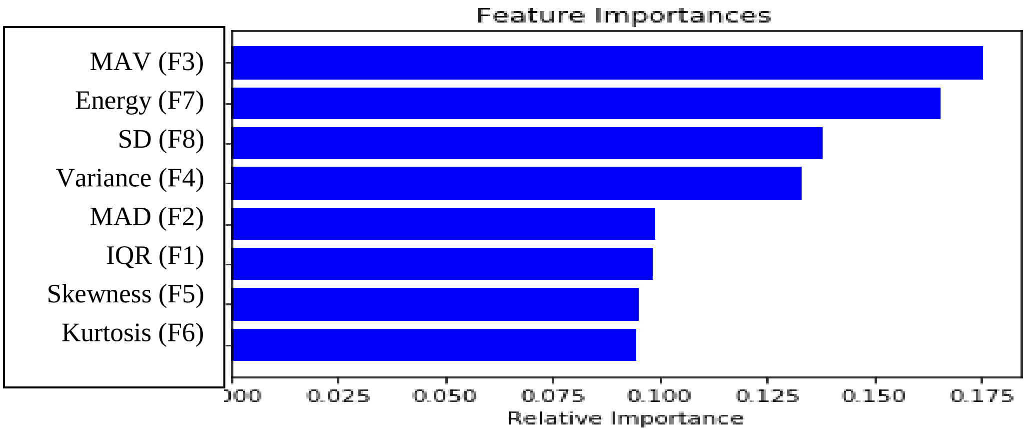

Figure 4. Showing relative feature importance

3. CLASSIFICATION

The classification of EEG signals plays a vital role in biomedical research. According to [15], there are mainly 5 types of classifiers used in BCI research such as linear classifiers, nonlinear classifiers, neural networks, nearest neighbour classifiers and a combination of these.

In this work, the Random Forest algorithm is employed for classifying EEG signals. Random forest is an ensemble learning algorithm. In this algorithm, at each node of the tree, we randomly select some subset of the features $s \subseteq S$, where 'S' is the set of features. The node then splits on the best feature in's' rather than 'S'. To decide on which feature to split is oftentimes the most computationally expensive aspect of decision tree learning. By narrowing the set of features, we drastically speed up the learning of the tree. The majority voting of the classification trees that have been formed obtains the prediction of the classification.

4. RESULTS

The wavelet transform analysis was performed on the dataset and the feature importance of the different extracted features was calculated using the Random Forest Algorithm in Scikit-learn followed by the classification of the selected features into two classes. We have used classification accuracy in order to evaluate the effectiveness of our method.

$$Classification\ accuracy\ (\%) = \frac{TP + TN}{TP + FP + FN + TN} \quad \text{(i)}$$

Where,

TP is True Positive;
TN is True Negative;
FP is False Positive;
FN is False Negative

Table 2 shows the combination of different features according to their relative importance, and with the classification accuracy of those combined features according to the relative importance. Therefore, we constructed eight different combined feature models with a different number of features to obtain the classification. In Table 2, F3 is shown to have the highest feature importance, hence selected first. After which F3 an F4 are combined, then F3, F4 and F8 and so on. From Table 2, it can be seen that though the feature F3 has the highest relative importance among the extracted features, it did not capture the significant distinctive information. However, combining F3 and F4 achieved the highest classification accuracy of this method. This suggests that feature F4 compliments the information captured by F3 and enhanced accuracy. Similarly, when 5 features are selected it achieved a similar accuracy of 85.41 as reported for 2 features.

Table 2. Showing different combination of features with their ranks according to feature selection method and the corresponding accuracies

No. of selected features	Feature name	Classification Accuracy (%)
1	F3	62.5
2	**F3, F4**	**85.41**
3	F3, F4,F8	76.14
4	F3, F4, F8,F2	63.85
5	**F3, F4, F8, F2, F7**	**85.41**
6	F3, F4, F8, F2, F7, F1	76.97
7	F3, F4, F8, F2, F7, F1, F6	78.64
8	F3, F4, F8, F2, F7, F1, F5	72.5

5. DISCUSSION

In this study, statistical features such as Mean Absolute Value, Median Absolute Deviation, Skewness, Kurtosis, Interquartile Range, Standard Deviation, Variance, and Energy were used to extract the underlying information from a dynamic EEG. This study has shown that the proposed features were successful in capturing the relevant distinguishing information. Also, it can be seen that the different combined features #2 and different combined features #5 have the same accuracy while #5 uses 5 features and #2 uses 2 features. This shows that good accuracy can be observed by using a lesser number of features and as a result improving the computational cost of the method. On the other hand, it could very well be true that using two features to represent the signal class may provide good classification accuracy but it could be at the cost of some vital information. Therefore, a future study may look into the choice of selecting lower number of features and the cost of crucial information lose. So, a follow up of this study would be selecting two features gave us good enough accuracy but whether crucial information was lost or not.

The choice of Random forest as a classifier has added robustness to the method. The proposed method can handle low dimensional as well as high dimensional data. However, it was observed that for high dimensional data, the method works slower but gives the same accuracy. Therefore, to reduce computational complexity, features were selected before the classification.

The application of this method in the near future is that it can be used to control an external device i.e. Neuro-prosthetics. The translated commands will be used as input to the external device via a computer (or micro-controller). This will, in turn, provide basic operations of the device. This study could also be used in the supervision of a trained physiotherapist to provide functional restoration to patients with spinal cord injury. In addition to that, this method can also be used in sports Biomechanics.

REFERENCES

[1] S. R. Guntur, R. R. Gorrepati, and V. R. Dirisala, "Robotics in Healthcare : An Internet of Medical Robotic Things (IoMRT)" Perspective. Elsevier Inc., 2016.
[2] M. Radman and A. Chaibakhsh, “Generalized Sequential Forward Selection Method for Channel Selection in EEG Signals for Classification of Left or Right Hand Movement in BCI,” ICCKE, pp. 137–142, 2019.
[3] N. E. Isa, A. Amir, M. Z. Ilyas, and M. S. Razalli, “Motor imagery classification in Brain- computer interface (BCI) based on EEG signal by using machine learning technique”, Bulletin of Electrical Engineering and Informatics, vol. 8, no. 1, pp. 269–275, 2019.
[4] E. Erkan and I. Kurnaz, “A study on the effect of psychophysiological signal features on classification methods,” Measurement, vol. 101, pp. 45–52, 2017.
[5] J. Wang, “EEG signal classification with feature selection based on one-dimension real- valued particle swarm optimization,” International Conference on Mechatronics, Control and Electronic Engineering, pp. 310–314, 2014.
[6] K. Aswinseshadri and V. T. Bai, “Feature Selection In Brain-Computer Interface Using Genetics Method,” 2015 IEEE Int. Conf. Comput. Inf. Technol. Ubiquitous Comput. Commun. Dependable, Auton. Secur. Comput. Pervasive Intell. Comput., pp. 270–275, 2015.
[7] C. K. S. G. Ponnambalam, “Binary and multi-class motor imagery using Renyi entropy for feature extraction,” Neural Comput. Appl., vol. 28, no. 8, pp. 2051–2062, 2017.
[8] K. Venkatachalam, A. Devipriya, J. Maniraj, M. Sivaram, A. Ambikapathy, and I. S. Amiri, “A novel method of motor imagery classification using EEG signal”, Artificial Intelligence In Medicine, vol. 103, no. December 2019, 2020.
[9] R. Chatterjee, T. Maitra, S. K. Hafizul, and M. Mehedi, “A novel machine learning-based feature selection for motor imagery EEG signal classification on the Internet of medical things environment,” Futur. Gener. Comput. Syst., vol. 98, pp. 419–434, 2019.
[10] Liu, A., Chen, K., Liu, Q., Ai, Q., Xie, Y., Chen, A.: Feature Selection for Motor Imagery EEG Classification Based on Firefly Algorithm and Learning Automata. Sensors. 17, 2576 (2017).
[11] M. Witkowski, M. Cortese, M. Cempini, J. Mellinger, and N. Vitiello, “Enhancing brain- machine interface (BMI) control of a hand exoskeleton using electrooculography (EOG)”, Journal of NeuroEngineering and Rehabilitation, pp. 1–6, 2014.
[12] J. R. Wolpaw, N. Birbaumer, D. J. McFarland, G. Pfurtscheller, and T. M. Vaughan, “Brain-computer interfaces for communication and control,” Clin. Neurophysiol., vol. 113, no. 6, pp. 767–791, 2002.

[13] D. J. McFarland, L. a Miner, T. M. Vaughan, and J. R. Wolpaw, “Mu and beta rhythm topographies during motor imagery and actual movements.,” Brain Topogr., vol. 12, no. 3, pp. 177–186, 2000.

[14] G. Pfurtscheller, A. Stancák, and C. Neuper, “Event-related synchronization (ERS) in the alpha band — an electrophysiological correlate of cortical idling: A review,” Int. J. Psychophysiol., vol. 24, no. 1, pp. 39–46, 1996.

[15] H. Bashashati, R. K. Ward, G. E. Birch, and A. Bashashati, “Comparing Different Classifiers in Sensory Motor Brain-Computer Interfaces”, PLOS, pp. 1–17, 2015.