Journal of
Applied Sciences

Research Article
Prosthetic Arm Control with Statistical Features of EEG Signals using K-star Algorithm

[1,2]V.V. Ramalingam and [3]S. Mohan

[1]Research and Development Centre, Bharathiar University, Coimbatore, India
[2]Department of Computer Science, SRM University, Kattankulathur, Chennai, Tamil Nadu, 603 203, India
[3]AL Yamamah University, Riyadh, Kingdom of Saudi Arabia

Abstract

This study presents an approach with a set of statistical features for the classification of various limb movements using K-star algorithm. The way of representing the characteristics of electroencephalogram (EEG) signals through the features are the most prominent one and it plays a vital role in the classification systems. The classification will be perfect when the sample is simplified through the feature extraction and feature selection process. In the present study, there are four categories of EEG signals recorded from healthy volunteers with finger open, finger close, wrist clockwise and wrist counter clockwise movements. The prominent statistical features were obtained from the EEG signals. The K-star algorithm was used to detect the changes in the EEG signals. The output of the classifier confirmed that the proposed K-star algorithm has potential to classify the EEG signals of the different hand movements.

Corresponding Author: V.V. Ramalingam, Research and Development Centre, Bharathiar University, Coimbatore, India

Competing Interest: The authors have declared that no competing interest exists.

Data Availability: All relevant data are within the paper and its supporting information files.

INTRODUCTION

The proposed work is mainly used to replace the natural arms with an artificial arm. There is a great demand for the replacement of the damaged arms of the human beings. The loss of human fingers is a major issue that intensely limits the everyday capabilities and interaction of the persons. The ability to interact with the real-world could be restored using electroencephalogram (EEG) signals. The EEG signals are the traces electrical activity of the human brain. The EEG signals control most parts of the human body. It is a highly complex signal and plays a vital role in the study of brain functions and neurological disorders. The EEG signal is an important general source of information used for the classification and identification of various finger movements. Generating a large amount of data by EEG monitoring systems for electroencephalographic changes and their visual analysis is a tedious process. In order to overcome this problem, computer system has been proposed. The automated systems for recognizing electroencephalographic changes have been used in the study for the past many years. In the field of medical sciences, there is great requirement for the development of such automated systems. Such automated system can detect the neurological disorders and help prevent the misinterpretation of EEG signals by Adeli *et al.* (2003), Agarwal *et al.* (1998) and Hazarika *et al.* (1997). Ubeyli (2010) proposed that the Least Square Support Vector Machines (LS-SVMs). To classify normal and epilepsy patients during epileptic seizures. For feature extraction, spectral analysis of the EEG signal was carried out with three model-based methods namely, Burg autoregressive-AR, moving average-MA, Yule-walker autoregressive moving average-ARMA methods. The author has proved that the Burg AR coefficients were the best features to represent the characteristics of EEG signals. Multi Layer Perceptron Neural Network (MLPNN) architecture was employed to classify the EEG signals. Ubeyli (2008) presented a novel approach on the implementation of multiclass Support Vector Machine (SVM) with the Error Correcting Output Codes (ECOC) trained on the extracted eigenvector features for classification of EEG signals. Ubeyli (2009) documented three sets; set-A, set-D and set-E of EEG signals were used for classification. Lyapunov exponents were extracted from the EEG signals given as the inputs to the MLPNNs. Finally, the features were trained with Levenberg-Marquadt algorithm and achieved good classification accuracy. Ubeyli and Guler (2007) used a diverse and composite input feature of EEG signals obtained by the eigenvectors. The classification was performed on diverse features (modified mixture of expert-MME) and composite feature (mixture of experts-ME) with five data set (set-A, set-B, set-C, set-D and set-E). The result demonstrated that the MME trained on diverse features have achieved a higher level of classification accuracy than ME.

Ubeyli (2008) presented to classify the EEG signals. Features were extracted using fourth-level Wavelet Packet Decomposition (WPD). Genetic Algorithm (GA) was used for feature selection and identified the best performing features to form the optimal feature subset. The approximate entropy values were derived as the feature vector and a Learning Vector Quantization (LVQ) was used as the classifier to attain the best classification accuracy for normal and epileptic epoch. Adeli *et al.* (2003) reported that the wavelet was an effective time-frequency analysis tool for analyzing EEG signals and most capable technique to extract features from EEG signals. Automated systems have been proposed to solve these problems by Bhattacharyya *et al.* (2015a, b), Rechy-Ramirez and Hu (2015) and Hortal *et al.* (2015). Most of the research works focused on the interpretation of EEG signals that are very complex and thus demand a lot of expertise. Automated systems have been proposed to solve these problems. Thus, it is clear that there are only a few literatures available for identification of limb movements using EEG signals. Decoding the information in EEG signal for various limb movements is a highly challenging one. Hence, there is more scope for designing a suitable automated system for recognizing limb movements using EEG signals.

In the present study, four different limb movements, namely finger open (Fopen), finger close (Fclose), wrist clockwise (WCW) and wrist counter clockwise (WCCW) are considered. To control the human limb movements the machine learning approach has been proposed. The majority of the people are right handlers around the world, thus, EEG signals of only right hand limb movements are considered in the present study. The decision tree clearly shows the structural information contained in the data (features) and explains the characteristics of features through the decision making 'if-then' rules. The 'if then' rules can be easily implemented in embedded systems to control the limb movements. Otherwise, one has to implement the classifier itself in embedded systems. Hence, decision tree classifier has been chosen for the present study to build a predictive model for right hand limb movements. The classifier performance largely depends on the information present in the features extracted from the EEG signals. Here statistical features are extracted from the EEG signals of four different limb movements and used as the input feature set for the classifier. The features will be useful for classification only if the signals are acquired properly.

MATERIALS AND METHODS

Most of the researchers used the datasets (http://www.meb.uni-bonn.de/epileptologie/science/physik/eegdata.html) available in public domains. In the present study, an effort is taken for data collection. The EEG signals are recorded from 27 healthy volunteers while performing the 4 different limb movements viz., finger open, finger close, wrist clockwise and wrist counter clockwise at the relaxed state. These four limb movements are the predominant physiological measures. Thus, the four limb moments viz., finger open, finger close, wrist clockwise and wrist counter clockwise are considered for the study. The complete architecture of the present study is shown in Fig. 1.

The EEG signals are recorded for each limb movements, respectively. The complete dataset consist of four different classes (Fopen, Fclose, WCW and WCCW) each signal containing 27 single-channel EEG signals of 60 sec duration is used for the present analysis. Each and every signal has been selected subsequently by visual inspection to remove the artifacts.

The EEG signals were recorded for the four classes using the standardized electrode placement method. The
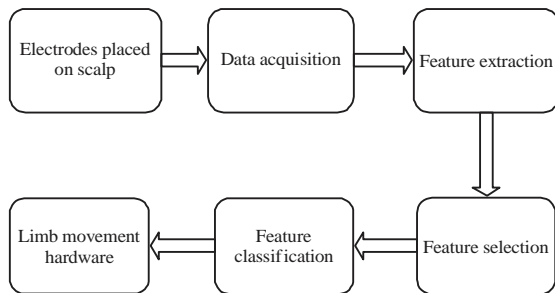

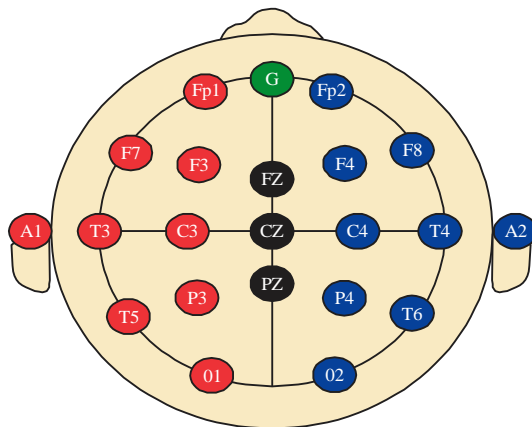Fig. 1: System architecture of prosthetic arm

Fig. 2: EEG electrode placement 10-20 system

experimental setup containing the RMS kit (Ranges of Transmission: 100 feet from the sight). Signals from electrodes C3, C4, CZ, FZ and PZ contains the information related to right hand movements investigated by Dornhege *et al.* (2007) and Bhattacharyya *et al.* (2014) and thus are selected in this study to identify the different limb movements. Conductive gel, cotton ball, wipes, computer unit connected with RMS kit and software tools. The required features were extracted through the MATLAB and processed with excel application. These lab instruments are used with the following procedure to conduct the experiments successfully. Initially, the primary experimental setup was confirmed to retrieve the EEG signals. The volunteers with short and oil free hair was employed to record the EEG signals. In order to stick the electrodes on the scalp region, conductive gel medium was used and the electrodes were cleaned properly with alcohol wipes. The electrodes such as C3, C4, CZ, FZ and PZ were placed on the scalp region according to the order shown in Fig. 2. The electrodes were fixed from the starting position at O1. The first two electrodes were placed in the frontal region, the other two electrodes were place on the occipital region and the last one was placed in the middle of the forehead. The safety measures were taken properly before recording the EEG signals by Johannesmeyer and Lindeman (2007).

Further, the instructions were given to the volunteers to commence the activity of finger open and monitored the changes in the EEG signals. The 10 sec of EEG data was saved under a file name "Fopen". To avoid the muscle fatigue, five minutes of rest was allowed to volunteers between each sequence of activity. Instructed the volunteers to perform the activity finger close and monitored the changes in the EEG signals through the five channels and the data were saved under the file name "Fclose". The same set of procedure was followed for the rest of the activities viz., wrist clockwise and wrist counter clockwise. The data points were written continuously onto the disk space of a computer unit which was connected with RMS kit at a sampling rate of 8-13 Hz. The length of EEG signal was fixed at 1024 (samples).

Feature extraction: Feature extraction is the process of determining the characteristic of the signal using some mathematical measure. In this study, the EEG signals of various conditions are consider as the input to the classifier. Classification algorithm is employed to map the attributes from the input space to the output space. The output space contains four regions representing the four limb movements. A set of 1024 data points of digitized EEG signals was taken as the input data to the classifier. Generally, classifiers find it difficult to manage the huge number of input variables. To

Table 1: Description of important statistical features

Name of the features	Description
Mean	It refers the average value of the signal point values
Median	Denoting or relating to a value or quantity lying at the midpoint of a frequency distribution of observed values or quantities, such that there is an equal probability of falling above or below it
Standard deviation	It is the Root Mean Square (RMS) deviation of the signal values from their arithmetic mean
Standard error	It is a measure of the amount of error in the prediction of y for an individual x in the regression, where and are the sample means and 'n' is the sample size
Sample variance	$$\text{Sample variance} = \frac{n\sum x^2 - (\sum x)^2}{n(n-1)} = \sigma^2$$
Skewness	Skewness characterizes the degree of asymmetry of a distribution around its mean
Kurtosis	$$\text{Kurtosis} = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_i - x^-}{S} \right)^S \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$$
Range	It refers to the difference between maximum and minimum signal point values for a given signal
Maximum value	It refers to the maximum signal point value in a given signal
Minimum value	It refers to the minimum signal point value in a given signal
Sum	It is the sum of all signal point values in a given signal

Table 2: Classification accuracy of statistical features from channel C4

Precision variable	Classification accuracy (%)
C4 median	51.85
C4 maximum	78.70
C4 mean	80.55
C4 skewness	79.62
C4 kurtosis	77.77
C4 range	78.70
C4 standard error	79.62
C4 minimum	80.55

minimize the number of input variables, researchers follow a few measures of the data points instead of the data themselves. These measures are called as 'features'. The definition of features is application dependent. The way of computing, such a measure is referred to as feature extraction. The dataset described in materials and methods section was taken and the statistical parameters were computed for each channel. Then C4.5 algorithm was used to perform the dimensionality reduction and classification. From the conducted experiments, amongst the channels C3, C4, CZ, FZ and PZ, the best classification accuracy was achieved by C4 channel (discussed in results section). Hence, only the C4 channel has been considered for the rest of the study. From C4 channel 12 statistical features were extracted. However, all the features extracted from the EEG signal may not contribute well in classification. The statistical features were given one by one to algorithm. Table 1 gives a description of the essential features.

The classification accuracy of each statistical feature was noted down. Then feature with highest accuracy "C4 median" were clubbed with next highest "C4 maximum" and accuracy was noted again. Since the accuracy increased third highest feature was also clubbed. This procedure was repeated until accuracy kept on increasing. Once the feature selection process is done, well performing features were identified. Feature selection is done in such a way that extracted features that do not contribute substantially to the classification accuracy are removed. To select the important features, C4.5 decision tree algorithm was utilized. Finally, the good features in ascending order of their contribution which contributes highest to classification accuracy are "C4 median", "C4 maximum" and "C4 mean" is shown in Table 2.

K-star algorithm: Lazy relates to a set of procedures which holds-up edifice in classifiers till the time taken for classification. A few variants of this algorithm are the IBK, K-star and Locally Weighted Learning (LWL). The K-star algorithm also referred to as an instance based classifier, utilizes entropic measures in light of the likelihood of changing a case into another by haphazardly picking between every conceivable change. Entropy is a measure of information and it can be used for classification of EEG signals. It is a reliable and important approach in real values, symbols and missing value attributes. An event based algorithm made for symbolic attributes fail in features of real value thus lacking an incorporated hypothetical base by Aha *et al.* (1991). Approaches which are effective in features of real values and are thus in an *ad-hoc* fashion are made to handle symbolic attributes. Treatment of missing values by classifiers forces a comparative issue. By and large, missing qualities are dealt with as a different esteem and are thought to be maximally diverse. These missing values are to be substituted for average value, because they are simply ignored. These issues can be best resolved by the entropy based classifier. Data hypothesis helps in figuring separation between instances. The unpredictability of a change of one instance into another is really the separation between instances.

The reduction of complexity can be achieved in two ways:

- Characterizing a limited arrangement of changes which will outline instance to another
- Change one instance (a) to another (b) with the assistance of ''program'' infinite sequences of transformations which initiates itself at a and terminate at b

Entropy as a distance measure: Data hypothesis helps in figuring separation between instances. The intricacy of a change of one instance into another is really the separation between instances. This is accomplished in two stages. In the first place, characterize a limited arrangement of changes they will delineate one instance to another. Then changing (a) one instance to (b) another with the assistance of ''program'' in a limited grouping of changes beginning at a and ending at b. To map instances with itself 'r' is used in T (σ (a) = a). The σ terminates the all set of prefix codes P from T*. Members in set T are defined at one to one unique transformation on I:

$$t\,(a) = t_n\,(t_n\text{-}1)\,(\ldots\,(t_1\,(a)\ldots)) \qquad (1)$$

where, $t = t_1,\ldots,t_n$, Let P is the probability function on T*. Thus, the given properties are practically satisfied.

Property 1: Entirety of aggregate likelihood of all ways from a to b. P* in relationship with definition of the likelihood of all ways from instance a to instance b:

$$P \times (\frac{b}{a}) = \sum\nolimits_{t \in p:t(a)=b} P(t) \qquad (2)$$

The function of K* is given as follows:

$$K \times (\frac{b}{a}) = \log_2 P \times (\frac{b}{a}) \qquad (3)$$

Note that K* algorithm is not precisely a distance measure function. As emphasized by the | notation, K* is a non-zero and not symmetric function.

Real numbers: The probability of the real numbers is found in

$$P \times (X) = \frac{1}{2x_0} e^{-\frac{x}{x_0}} \Delta x . x_0 \qquad (4)$$

Property 2: Category prediction computing entropy is the most initiative way to classification prediction. It is through by adding the probabilities of a to every single occurrence

which constitutes C. The likelihood of each and individual instance is processed and their relative probabilities are assessed. The biggest likelihood of the chose set is considered as the grouping of the new instance.

$$P \times (\frac{C}{a}) = \sum\nolimits_{b \in c} P \times (\frac{b}{a}) \qquad (5)$$

RESULTS

The four limb movements of EEG signals, namely Fopen, Fclose, WCW and WCCW are taken for the study. The machine learning technique has been used with the statistical features. The classification performance of K-star algorithm is reported.

The data set described in section materials and methods was taken and statistical features were computed for each channel from EEG signals as explained in section feature extraction. Out of five channel (C3, C4, CZ, FZ and PZ), C4 has better information for limb movement classification with statistical features. This is evident from Fig. 3.

Then the C4.5 algorithm was used to perform the dimensionality reduction. A decision tree has been developed as shown in Fig. 4. Therefrom well contributing features were selected for classification by following footsteps of Sugumaran *et al.* (2007). For EEG signals, out of 12 features, the selected statistical features are "C4 median", "C4 maximum" and "C4 mean". With these features the classification accuracy is computed using K-star algorithm. The best classification accuracy is 82.40% using EEG signals in this study. Here the objective is to find the best parameter out of statistical parameters with which the highest classification accuracy can be achieved. With the above feature set, one has to tune the classifier to increase the classification accuracy.

In the present study, K-star algorithm is employed as the classifier to classify the four limb movements. Here, the global blend is an important parameter that affects the classification accuracy. Hence, a series of experiments have been conducted
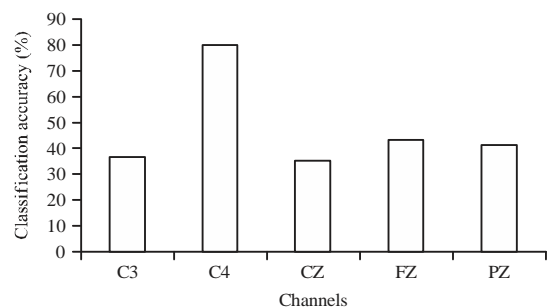
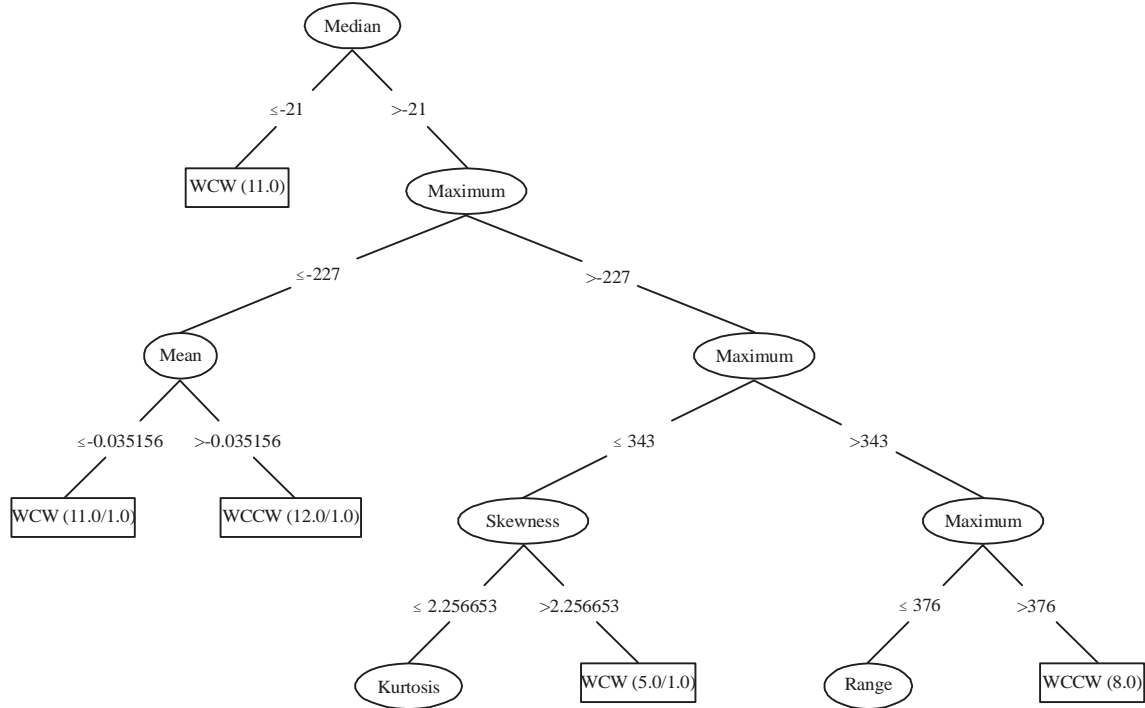Fig. 3: Classification accuracy vs channels
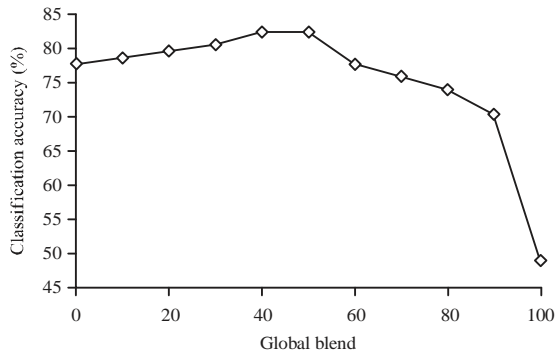
Fig. 4: Decision tree with statistical features

Fig. 5: Classification accuracy vs global blend

Table 3: Confusion matrix of K-star alogorithm with statistical features

	Fclose	Fopen	WCCW	WCW
Fclose	23	3	0	1
Fopen	1	23	2	1
WCCW	4	2	20	1
WCW	0	2	2	23

to investigate the effect of global blend parameter by varying from 0-100. The variation of classification accuracy with respect to global blend is shown in the Fig. 5.

The misclassification information can be expressed with the help of confusion matrix as shown in Table 3. The confusion matrix explains well about the actual class and the predicted class classified by the classifier. The first row represents data with "finger close" conditions. The second row

represents "finger open" condition. The third and fourth row gives the information about the "wrist counter clockwise" and "wrist clockwise" conditions, respectively. The diagonal elements of the confusion matrix had shown the correctly classified instances.

Classification errors analysis: The sensitivity, specificity and total classification accuracy were computed to test the performance of the classifier. The sensitivity, specificity and total classification accuracy are defined as:

$$\text{Sensitivity} = \frac{\text{Number of true positive}}{\text{Number of true positive} + \text{Number of fase negative}} \quad (6)$$

$$\text{Sensitivity} = \frac{\text{Number of true negative}}{\text{Number of true negative} + \text{Number of fase positive}} \quad (7)$$

$$\text{Total classification accuracy} = \frac{\text{Number of correct instances}}{\text{Total number of instances}} \quad (8)$$

In order to determine the performance of K-star classifier used for classification of the EEG signals, the classification accuracies (specificity, sensitivity and total classification accuracy) on the test data sets are presented in Table 4. In this case, the sensitivity shows the less percentage than the

Table 4: Values of the statistical parameters of the classifier

Classifier	Classes	Statistical parameters (%)		
		Sensitivity	Specificity	Total classification accuracy
K-star	FCLOSE	85.18	93.82	82.40
	FOPEN	85.18	91.35	
	WCCW	74.07	95.06	
	WCW	85.18	96.29	

Table 5: Summary of classifier's result

Parameters	Values
Correctly classified instances	82.40
Incorrectly classified Instances	17.59
Kappa statistic	0.76
Mean absolute error	0.16
Root mean squared error	0.27
Relative absolute error	44.73
Root relative squared error	62.48
Total number of Instances	108.00

Table 6: Detailed accuracy by class-K-star algorithm

TP rate	FP rate	Precision	Recall	F-measure	ROC area	Class
0.852	0.062	0.821	0.852	0.836	0.949	Fclose
0.852	0.086	0.767	0.852	0.807	0.936	Fopen
0.741	0.049	0.833	0.741	0.784	0.890	WCCW
0.852	0.037	0.885	0.852	0.868	0.940	WCW

specificity measure. However, the specificity is 95.06 and 96.29% for classes 'WCCW' and 'WCW', respectively. One can confidently say that the non 'WCCW' and non 'WCW' classes were identified correctly by the classifier. However, the sensitivity for these classes is 74.07 and 85.18%, respectively. This indicates that the classifier is not able to identify these classes ('WCCW' and 'WCW') correctly. Similarly, the sensitivity and specificity measures are high for the classes 'Fclose' and 'Fopen'. This indicates the ability of the classifier to classify these classes from others easily.

Table 5 shows the realistic picture of performance of the classifier. The achieved classification accuracy is 82.40%. The kappa statistics is a measure of the agreement between predicted and observed categories of data set. In this case, the kappa statistics happens to be 0.76.

The following could be considered for judging the quality of the model; TP rate and FP rate. TP rate represents a rate of true positive that has a value nearer to 1 (0.8). Likewise, the FP rate or false positive has a value close to 0 (0.03) is shown in Table 6. The slight variation is due to some misclassification of four different classes like Fclose, Fopen, WCCW and WCW.

DISCUSSION

The controlling process of prosthetic limb movements based on surface EMG signals extracted from remnant muscles are the promising ones in the analysis of EMG signals. There were three feature extraction techniques, namely autoregressive coefficients, mean frequency and EMG histogram used in the study. The combined features of mean frequency and EMG histogram were given as the input to neural networks classifier. Hence, it is noticed from this study that the EMG histogram feature vector performed well for the classification of prosthetic limb movements stated by Aishwarya *et al*. (2013). Fadzal *et al*. (2012) reported that the Power Spectral Density (PSD) was the well suited method to distinguish right and left hand writing movements using EEG signals. The finding of the study support to great extent the results obtained by Dornhege *et al*. (2007) and Bhattacharyya *et al*. (2014). The proposed work considered four classes of right hand limb movements Viz., finger open (Fopen), finger close (Fclose), wrist clockwise (WCW) and wrist counter clockwise (WCCW). The statistical features were extracted from the EEG signals. The set of well contributed statistical features were classified using K-star algorithm and achieved the best classification accuracy of 82.40%. From the above result and discussion one can conclude that EEG signals can be used for prosthetic limb movements with K-star classifier. The signal processing of EEG signals is a complex task and requires sophisticated techniques to yield better classification accuracy. A better feature extraction technique may be used in order to improve the classification accuracy.

REFERENCES

Adeli, H., Z. Zhou and N. Dadmehr, 2003. Analysis of EEG records in an epileptic patient using wavelet transform. J. Neurosci. Methods, 123: 69-87.

Agarwal, R., J. Gotman, D. Flanagan and B. Rosenblatt, 1998. Automatic EEG analysis during long-term monitoring in the ICU. Electroencephalogr. Clin. Neurophysiol., 107: 44-58.

Aha, D.W., D. Kibler and M.K. Albert, 1991. Instance-based learning algorithms. Mach. Learn., 6: 37-66.

Aishwarya, R., M. Prabhu, G. Sumithra and M. Anusiya, 2013. Feature extraction for emg based prostheses control. ICTACT J. Soft Comput., 3: 472-477.

Bhattacharyya, S., A. Konar and D.N. Tibarewala, 2014. A differential evolution based energy trajectory planner for artificial limb control using motor imagery EEG signal. Biomed. Signal Process. Control, 11: 107-113.

Bhattacharyya, S., D. Basu, A. Konar and D.N. Tibarewala, 2015. Interval type-2 fuzzy logic based multiclass ANFIS algorithm for real-time EEG based movement control of a robot arm. Robotics Auton. Syst., 68: 104-115.

Bhattacharyya, S., M. Pal, A. Konar and D.N. Tibarewala, 2015. An interval type-2 fuzzy approach for real-time EEG-based control of wrist and finger movement. Biomed. Signal Process. Control, 21: 90-98.

Dornhege, G., J.R. Millan, T. Hinterberger, D.J. McFarland and K.R. Muller, 2007. Toward Brain-Computer Interfacing. 1st Edn., MIT Press, Massachusetts, ISBN: 9780262042444, Pages: 507.

Fadzal, C., W. Mansor and L.Y. Khuan, 2012. Analysis of EEG signal from right and left hand writing movements. Proceeding of the IEEE Control and System Graduate Research Colloquium, July 16-17, 2012, Shah Alam, Selangor, pp: 354-358.

Hazarika, N., J.Z. Chen, A.C. Tsoi and A. Sergejew, 1997. Classification of EEG signals using the wavelet transform. Signal Process., 59: 61-72.

Hortal, E., E. Ianez, A. Ubeda, C. Perez-Vidal and J.M. Azorin, 2015. Combining a Brain-machine interface and an electrooculography interface to perform pick and place tasks with a robotic arm. Robotics Auton. Syst., 72: 181-188.

Johannesmeyer, S. and K. Lindeman, 2007. Electroencephalography (EEG) and Electromyography (EMG) measurement of action potentials in the brain and muscle. CleveMed (Cleveland Medical Devices Inc.), Cleveland, OH., USA.

Rechy-Ramirez, E.J. and H. Hu, 2015. Bio-signal based control in assistive robots: A survey. Digital Commun. Networks, 1: 85-101.

Sugumaran, V., V. Muralidharan and K.I. Ramachandran, 2007. Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing. Mech. Syst. Signal Process., 21: 930-942.

Ubeyli, E. and I. Guler, 2007. Features extracted by eigenvector methods for detecting variability of EEG signals. Pattern Recognit. Lett., 28: 592-603.

Ubeyli, E.D., 2008. Analysis of EEG signals by combining eigenvector methods and multiclass support vector machines. Comput. Biol. Med., 38: 14-22.

Ubeyli, E.D., 2009. Statistics over features: EEG signals analysis. Comput. Biol. Med., 39: 733-741.

Ubeyli, E.D., 2010. Least squares support vector machine employing model-based methods coefficients for analysis of EEG signals. Expert Syst. Applic., 37: 233-239.