ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645
Indian Journal of Science and Technology, *Vol 9(47), DOI: 10.17485/ijst/2016/v9i47/102100, December 2016*

Classification of EEG Signals for Prosthetic Limb Movements with ARMA Features Using C4.5 Decision Tree Algorithm

V. V. Ramalingam[1,2]*, S. Mohan[3], B. Rebecca Jeya Vadhanam[4] and V. Sugumaran[5]

[1]Research and Development Centre, Bharathiar University, Coimbatore – 641046,Tamil Nadu, India; ramabi1976@gmail.com
[2]Department of Computer Science and Engineering,Faculty of Engineering and Technology, S. R. M University, Kattankulathur – 603203; Tamil Nadu,India; ramabi1976@gmail.com
[3]CCIS, AL Yamamah University, Riyadh, Kingdom of Saudi Arabia, s.mohan77@gmail.com
[4]Department of Computer Applications, Faculty of Science and Humanities, S. R. M University, Kattankulathur – 603203, Tamil Nadu, India; beckyatsrm@gmail.com
[5]School of Mechanical and Building Sciences (SMBS), VIT University, Chennai Campus, Chennai-600127, Tamil Nadu, India; v_sugu@yahoo.comm

Abstract

Objectives: This paper presented a novel approach with a set of Auto Regressive Moving Average (ARMA) features for the best classification of different hand moments in Electroencephalogram (EEG) signals using C4.5 Decision tree algorithm. **Methods/Analysis:** The characteristics of EEG signals can be represented through the best features is the most prominent and significant role in the classification systems. The classification is more flawless when the specimen is streamlined through the feature extraction and feature selection process. **Findings**: In this study, there are four kinds of EEG signals recorded from strong volunteers with finger open, finger close, wrist clockwise and wrist counterclockwise. The well performing statistical features are acquired from the EEG signals. C4.5 Decision tree classifier is used to identify the changes in the EEG signals. The yield of the classifier confirmed that the proposed C4.5 Decision tree classifier has potential to classify the EEG signals of the specific hand movements. **Improvement**: The proposed work is contributed to manage the right hand movements through the EEG signals. The efficient techniques are required to process the complex EEG signals to achieve the better classification result. To improve the classification accuracy, an efficient feature extraction technique may be applied.

Keywords: ARMA Features, C4.5 Decision Tree, Classification, Electroencephalogram (EEG) Signals

1. Introduction

When humans lose a limb it extremely limits the day-to-day activities and interaction of the person and hence it is a worthy issue that needs to be addressed. There are two types of signals such as EMG and EEG they can be used to regulate the prosthetic limb activities. EMG signals are extracted from the muscles which contains a lot of required data to discriminate the limb activities. Unfortunates lead the great loss of the human's parts. Mostly, Extracting EMG signals from the affected area would not be of great potential. In addition, EMG signals are auxiliary signals, whereas the EEG signals are primary or essential signals. The EEG signals originates from the brain behavior and

*Author for correspondence

hence the characteristics remain almost same irrespective of the extent of amputation. Hence, it makes great sense that EEG signal is most suitable for controlling limb movements. It is partially true that due to the fact of EEG signals are produced by any thought process, decoding process of EEG signals are quite complex. The present challenge is to effectively decode and dig the information from the EEG signals. This work is a novel attempt to discriminate the EEG signals which are used to control the prosthetic limb movements . A review of the various techniques available identifies the merits and demerits. The scope and need of the present work can also evolve from this. In[1] the experiments were conducted to find the appropriate method to differentiate the right and left hand writing movements using Power Spectral Density (PSD) through EEG signals. In[2] two kinds of experiments were conducted on healthy volunteers with eyes open state and epilepsy patients during epileptic seizures. ARMA features were extracted from EEG signals and classified using Support Vector Machine. It has achieved good classification accuracy. In[3–5] presented the primary requirements for the developing a computerized framework. It helps to recognize the neurological issues and it is a good aid to detect the misinterpretation of EEG signals given by the analyst. In[6] experimented to discriminate the EEG signals utilizing eigenvector features and classified with support vector machine. In[7] the wavelet packet decomposition was used to extract the meaningful feature and Genetic Algorithm has utilized for the feature selection process. The Learning Vector Quantization (LVQ) classifier was employed to obtain high classification accuracy for both normal and epileptic volunteers. In[8] Multilayer Perceptron Neural Network (MLPNN) framework was utilized to identify the electroencephalographic changes. The MLPNN classifier was used to classify three sets such as, set-A, Set-D and set-E of EEG signals and formed as an inputs to the MLPNNs classifier. From the results obtained, the MLPNN has proved that it was a potential application for classifying EEG signals. In[9] the experiment that was performed for classification the EEG signals based on the diverse and composite features with five data sets. The result gained from the MME with diverse features was achieved best classification performance than ME. In[10] the surface EMG signals were extracted and utilized for controlling the prosthetic limb movements. These signals extricated from remnant muscles for investigate the EMG signals. In their study, EMG histogram, autoregressive coefficients and mean frequency feature extraction techniques were used to extract the useful features. The neural network classifier has employed to classify mean frequency and EMG histogram features. Therefore, the EMG histogram features out performed with prosthetic arm control system.

In particular, the main stream of the research work paid attention towards the elucidation of EEG signals which are extremely complex to understand and hence, an expertise system is required.

However, automated frameworks have been proposed to take care of the issues. The existing literature study is very limited in the field EEG signals to identify the limb movements. Therefore, scheming a reasonable mechanized system that classifies the movements of limb using EEG signals is an extensive task.

Generally, most of the population habitually right handlers in the world. Human right hand activities are very essential and hence it is considered in this study. Hence, the EEG signals have been extracted from the four different right hand limb movements. Machine learning approach has been utilized to identify these four different limb movements. Decision tree classifier has been employed to classify the four limb movements. The structural information of the drawn features are clearly clarified by applying the decision making 'if then' rules. This rules plays a major role to identify the four types of distinctive EEG signals can be too implemented in the embedded framework. Contrarily, the classifiers itself has to be implemented in the embedded frameworks. Performance of the classifier can be measured from the information that is present in the features derived from the EEG signals.

The ARMA features were extracted from EEG signals for four different limbs activities respectively. The data set were formed with ARMA features and applied to the classifier as input. These important features will be helpful for classification of the signals that are gained appropriately.

In section 2, the data acquisition processes and system architecture are explained in detail. The C4.5 Decision tree classifier with the ARMA features are projected as an appropriate classifier for regulating the prosthetic limb activities using the EEG signals.

2. Materials and Methods

Datasets are primary requirement in the research area of EEG signals. Much research work has been done with the datasets that are publically available (http://www.meb.uni-bonn.de/epileptologie/science/ physik/eegdata.html). No exclusive efforts have been made for data collection process. In this study, specific focus has been given in creating a dataset to classify the four different limb activities. EEG signals attributing to the four different limb movement's viz., finger close, finger open, wrist clockwise and wrist counterclockwise were extracted from 27 volunteers in their relaxed condition. In fundamental nature, these four types of limb movements are the physical events of the human. Thus, these four limb moments viz., finger close, finger open, wrist clockwise and wrist counterclockwise have been contemplated to develop the proposed work. The entire structural design of this research is shown in Figure 1. The EEG signals were recorded for each type of limb movement individually. The entire dataset comprises four distinctive classes (fclose, fopen, wcw and wccw) . For the experimental evaluation, the 60 seconds recording of EEG signals were considered and each signal containing 27 single-channels. These signals are envisioned to evacuate the curios.

Standardized electrode placement techniques have been used to record the EEG signals from the four different limb movements. The four classes such as fclose, fopen, wcw and wccw are assigned for the respective limb movements. Signals from electrodes C3, C4, CZ, FZ and PZ contains information identified with right hand[11,12] and accordingly chosen in this study to recognize the distinctive limb movements. Matrix Laboratory (MATLAB) is an efficient tool which is used to extract the features and then these feature were processed in the excel application. Figure 2 shows the structure of the electrodes such as C3, C4, CZ, FZ and PZ placement on the human scalp. These set of five electrodes were placed from the starting point O1 onwards. The first two electrodes are positioned in the frontal region, two other electrodes were placed in the

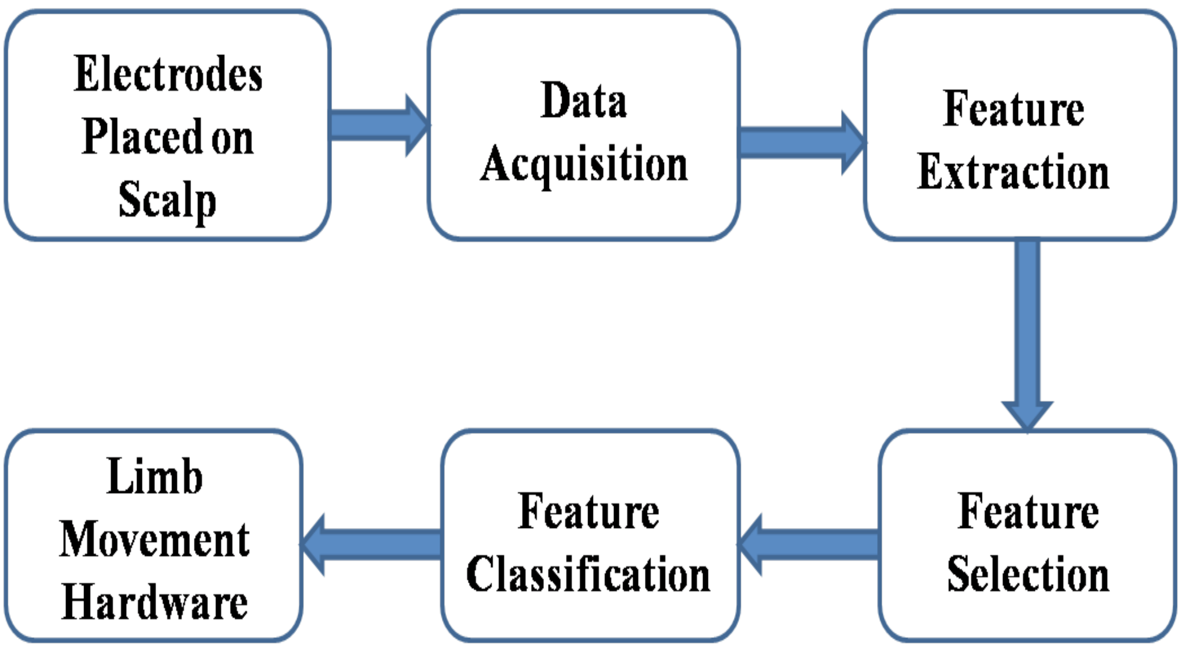

Figure 1. System architecture of prosthetic limb movement.

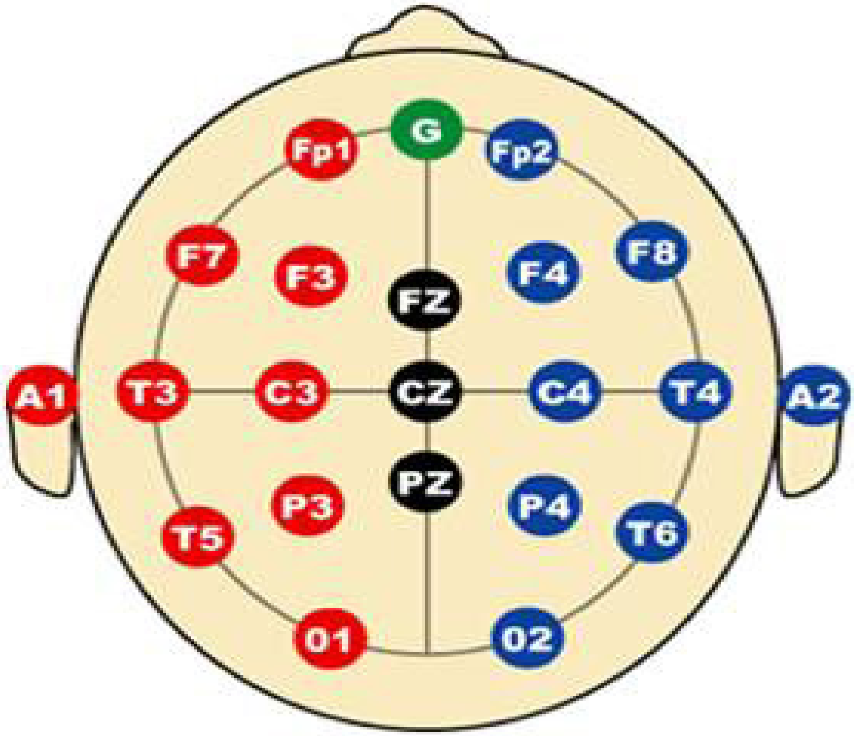

Figure 2. EEG electrode placement 10 – 20 system.

occipital region and the final electrode was positioned at the middle[13].

The volunteers were instructed to follow the rules to initiate the activities of finger open movement and scrutinized the variations in EEG signals. For each activity, the EEG data for a duration of 10 seconds has been updated on the hand movement's name such as finger close (fclose), finger open (fopen), wrist clockwise (wcw) and wrist counterclockwise (wccw). In order to avoid the muscle weakness, the volunteers were allowed to take five minutes rest in the equal intervals of each hand movement. An efficient approach for automatically updating the EEG data points into the integrated set up of computer system connected with RMS kit with the sampling rate ranging from 8 Hz to 13 Hz. In this case, EEG signal length was 1024 (samples).

2.1 Channel Selection

The statistical parameters were computed and the dataset were updated for each channel as described in section 2. However, it is essential to carry out the dimensionality reduction and classification and hence C4.5 algorithm was used for dimensionality reduction and classification was performed with the default value set to 0.25 and the number of objects was kept as 2. The obtained result shows the entire specified electrode C3, C4, CZ, FZ and PZ; C4 channel was attained the best classification result. The channel C4 is then useful to evaluate the classification performance and further considered for the study to achieve better results. Figure 3 shows the performance of the selected channels with respect to their classification accuracy.

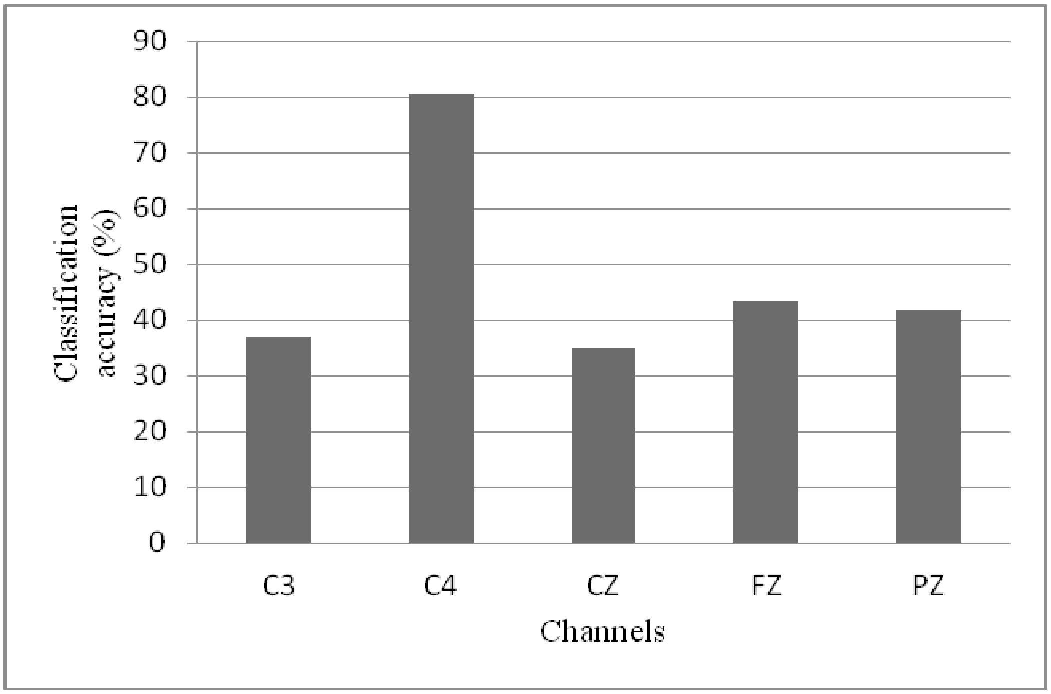

Figure 3. Channels vs. classification accuracy.

2.2 Feature Extraction

In machine learning approach, feature extraction is an essential process to determine the distinctiveness of the signals on the basis of statistical measures. In the present study, the EEG signals of four different classes were considered and given as the input to the classifier. ARMA features have been extracted to associate the attributes from input space to resultant space. Each input data set contains 1024 data points facilitating the EEG signals. These signals were supplied as the source to the algorithm. In general, algorithms find it is complex to deal with the large number of input features. In order to minimize the number of input variables, many researches provide a small number of measures of the data points rather than the data themselves. Thus, feature extraction process pays a particular attention to extract the meaningful information from the signals.

2.2.1 ARMA Features

ARMA models are numerical models of the auto correlation in a time series. ARMA models can be utilized to foresee the behavior of a time series of past values alone. Such an expectation can be utilized as a standard to assess the conceivable significance of different variables to the system. In[14] reported ARMA models are generally utilized for forecast of monetary and mechanical time arrangement. ARMA models can be delineated by a progression of conditions. For effortlessness, the time arrangement was decreased to zero-mean first by subtraction of the specimen mean.

$$y_t = Y_t - \overline{Y}, \quad t = 1, 2, \ldots N \qquad (1)$$

Where,

Y_t is the original time series

$\overline{Y}$ is its sample mean

y_t is the mean-adjusted series

Autoregressive models are a subset of ARMA models. In AR model a time series is represented as a linear function of its defined values. The number of lagged defined values included is shown by the order of the AR model. The first-order autoregressive model can be understood easily. The equation for this model is

$$y_t + a_1 y_{t-1} = e_t \tag{2}$$

Where,

y_t is the mean-adjusted series in year t,

y_t −1 is the series in the previous year,

a_t is the lag-1 autoregressive coefficient

e_t is the noise.

The occurred errors are defined as: the random-shock, and the residual. The residuals e_t is assumed to be random in time (not auto-correlated), and normally distributed. The equation for the AR (1) model can be rewritten as

$$y_t = -a_1 y_{t-1} + e_t \tag{3}$$

The AR (1) model takes a form of regression model that has y_t regressed on its past value, and that e_t is analogous to the regression residuals. It is termed as autoregressive due to the regression on self (auto). An Autoregressive model with Higher-order contains more lagged yt terms as predictors. For instance, the second order autoregressive model, AR (2), is given as

$$y_t + a_1 y_{t-1} + a_2 y_{t-2} = e_t \tag{4}$$

Where, a1 and a2 are the coefficients of autoregressive on lags 1 and 2. The pth order autoregressive model. AR (p) incorporates lagged terms on time t −1 to t−p .

The Moving Average (MA) model is a structure of ARMA model in which the time series is regarded as a MA (unevenly weighted) of a random shock series et . The first-order moving average (MA (1)) model is given by

$$y_t = e_t + c_1 e_{t-1} \tag{5}$$

Where

e_t , e_{t-1} are the residuals at times *t* and *t-1,*

c_1 is the first-order MA coefficient.

as with the AR models, MA models with higher-order include higher lagged terms. For instance, the second order moving average model, MA (2), is

$$y_t = e_t + c_1 e_{t-1} + c_2 e_{t-2} \tag{6}$$

The order of the MA model is denoted with the letter q. A second-order MA model is denoted by MA (q) with q = 2.

"It has been seen that the autoregressive model incorporates lagged terms of the time series itself, while that the MA model incorporates lagged terms on the error or residuals. By including both sorts of lagged terms, it can be distinguished, what are called autoregressive-moving-average, or ARMA, models. The order of the ARMA model is incorporated in brackets as ARMA (p,q), where p is the autoregressive order and q the moving-average order. The simplest, and most frequently utilized ARMA model is ARMA (1,1) model as given below"

$$y_t + a_1 y_{t-1} = e_t + c_1 e_{t-1} \quad (7)$$

2.3 Feature Selection

The ARMA feature extraction method is described in section 2.2.1. The extraction of feature is performed through three strategies, In particular Burg AR, Yule, Yule walker. Three components from every strategy were gotten and in this manner absolutely nine features are accessible. Following the footsteps of [15–17], one can play out the feature choice utilizing C4.5 algorithm to decrease the measurement of the dataset. In the present study, the order of ARMA model was varied from 1 to 20. For each order, the ARMA features were extracted and classified using decision tree to choose the right order. For each order, the statistical parameters were computed like mean, median, skewness and kurtosis etc. The mean is computed to serve as a best feature when the order is 9. Figure 4 shows the performance of various orders. Referring Figure 4, the classifier gives a best accuracy of the 9th order. Order 9 was used as input to C4.5 algorithm as shown in Figure 5. From the decision tree, just five contributors were recognized as prevailing among the accessible features. The contributing features are namely, "a1", "e1", "k1", "a2" and "e2". "The features were selected based on entropy reduction and information gain. The information gain is a measure of the discriminating capability of a feature of the given data set. The Figure 5 shows the decision tree generated using C4.5 decision tree algorithm".

2.4 Decision Tree (C4.5) Algorithm

The C4.5 algorithm has two phases mainly construction phase and pruning phase. "In the construction phase, C4.5 constructs a decision tree by using the concept of information theory. The tree has only one root node for the entire training set. A new node is added to the decision tree for every partition. For a set of samples in a partition *S*, a test attribute *M* is selected for further partitioning the set into $S_1, S_2, \ldots, S_L$. New nodes for S are created and these are added to the decision tree as a sibling. The building

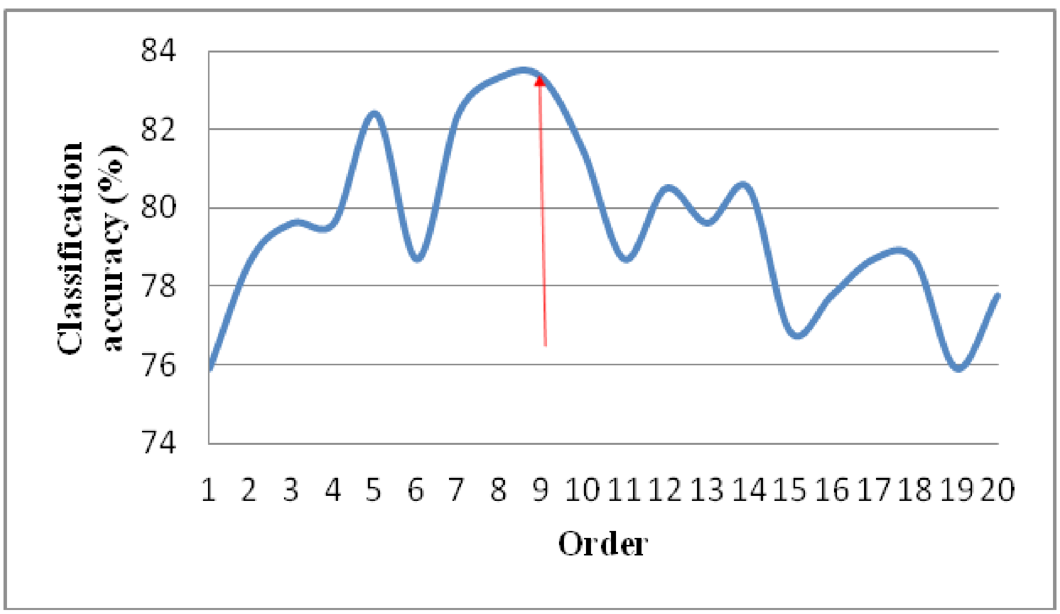

Figure 4. Order vs. classification accuracy.

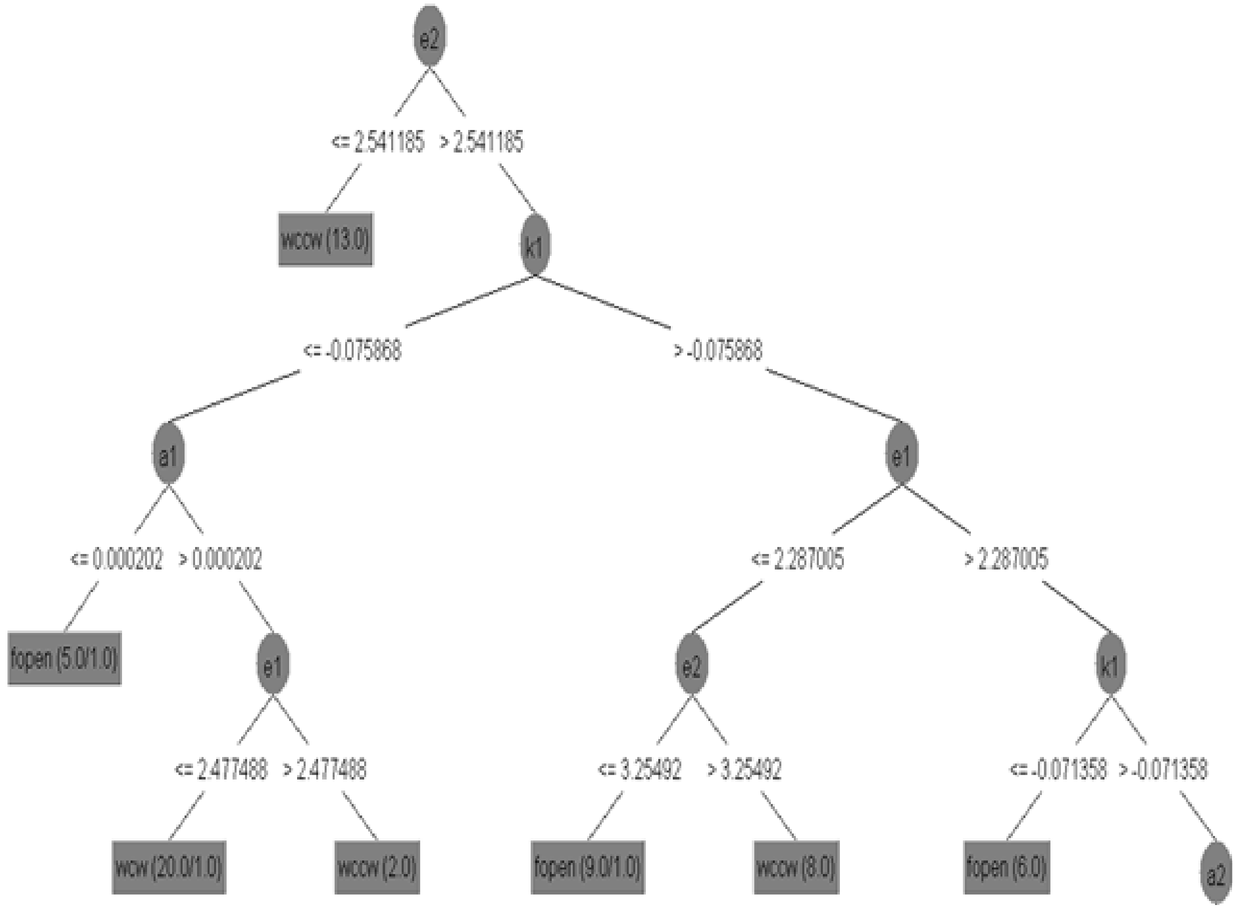

Figure 5. Decision tree with ARMA features.

of decision tree depends on a test attribute *M*. The C4.5 classification algorithm uses entropy based information gain as the selection criteria – discussed in the following subsection".

2.4.1 Information Gain and Entropy Reduction

"Information gain is the expected reduction in entropy caused by partitioning the examples according to the specified feature. It measures how well a given attribute separates the training examples according to their objective function. Information gain (*S, A*) of a feature A relative to a collection of examples *S*, is defined as"

$$Gain(S, A) = Entropy(S) - \sum_{v \in Value(A)} \frac{|S_v|}{|S|} Entropy(S_v) \tag{8}$$

Where, value (*A*) is the set of all possible information for attribute *A* and *Sv* is the subset of *S*.

The entropy is a measure of the homogeneity in random samples and it is given by

$$Entropy(S) = \sum_{i=1}^{c} - P_i \log_2 P_i \tag{9}$$

"Where, P_i is the proposition of '*S*' belonging to the class '*i*' and '*c*' is the number of classes from the training samples. The gain is just the entropy collection of *S*. The expected value of entropy after *S* is partitioned using function *A*. The expected entropy described by the second term is simply the sum of the entropies of each subset S_v, weighted by the fraction of examples $|S_v|/|S|$ that belongs to S_v. Therefore Gain (*S, A*) is the projected reduction in entropy caused by knowing the value of feature *A*".

3. Results and Discussion

The feature selection and classification was carried out with the help of C4.5 algorithm. Among the extracted 9 features, 5 most essential feature set were recognized and classified utilizing C4.5 algorithm. The classification performance was discovered utilizing the 10-fold cross validation. A big major favorable position of the 10-fold cross validation strategy is that all perceptions are utilized for both training and validation and each observation is utilized for validation precisely once. This prompts to a more precise approach to gauge the exactness in view of training data set. The confusion matrix in Table 1 shows the classification and misclassification details.

From the confusion matrix in Table 1, The first row first element shows that there are 23 instances classified correctly as 'fclose' class and 4 instances were misclassified as fopen, wccw, and wcw. Similarly, the second row, second element shows that there are 24 instances classified correctly as 'fopen' and 3 instances misclassified as rest of the class. Hence, all the diagonal elements of the confusion matrix table show the correctly classified instances of the class. Thus the classification accuracy was calculated. However, C4.5 algorithm achieves best performance (82.40) utilizing more prominent ARMA features. In short, C4.5 algorithm can be utilized adequately to select prominent features and classifying the features

Table 1. Confusion matrix – C4.5 decision tree algorithm

Class	fclose	fopen	wcw	wccw
fclose	23	2	1	1
fopen	3	24	0	0
wcw	2	1	22	2
wccw	1	4	2	20

Table 2. Detailed accuracy by class – C4.5 decision tree algorithm

TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
0.852	0.074	0.793	0.852	0.821	0.925	***fclose***
0.889	0.086	0.774	0.889	0.828	0.905	***fopen***
0.815	0.037	0.880	0.815	0.846	0.892	***wcw***
0.741	0.037	0.870	0.741	0.800	0.895	***wccw***
0.824	0.059	0.829	0.824	0.824	0.904	***Wt. AVG***

selected from the data set. The exhaustive accurateness by class shown in Table 2.

The classification accuracy can be confirmed utilizing the itemized exactness by class. For a perfect algorithm, the true positive rate (TP rate) should be 1, whereas the false positive rate (FP rate) should be 0. Referring Table 2, the true positive rate, precision, recall and F-measure values are close to 1. Due to some misclassification, the classification exactness was observed to be 82.40 %.

4. Conclusion

In this proposed work, there are four right hand limb movements incorporated with the four classes viz., finger close (fclose), finger open (fopen), wrist clockwise (wcw) and wrist counterclockwise (wccw). The ARMA features were obtained from EEG signals. Through Decision Tree, only five best features were selected as prevailing features. The contributing features are namely, “a1”, “e1”, “k1”, “a2” and “e2”. The obtained classification accuracy of 82.40% from the well contributing ARMA features using Decision tree classifier. From the above result and discussion one can conclude that EEG signals can be utilized to control the prosthetic limb movements. These EEG signals are trained by C4.5 Decision tree classifier and it produced good classification accuracy. Moreover, an efficient techniques are required to process the complex EEG signals to achieve the better classification results. To improve the classification accuracy, an efficient feature extraction technique may be applied.

5. References

1. Fadzal FCW, Mansor W, Khuan LY. Analysis of EEG signals from right and left hand writing movements. Institute of Electrical and Electronics Engineers (IEEE) Control and System Graduate Research Colloquium; 2012. p. 354–58.
2. Elif DÜ. Least squares support vector machine employing model-based methods coefficients for analysis of EEG signals. Expert Systems with Applications. 2010 Jan; 37(1):233–39.
3. Adeli H, Zhou Z, Dadmehr N. Analysis of EEG records in an epileptic patient using wavelet transformation. Journal of Neuro Science Methods. 2003 Feb; 123(1):69–87.
4. Agarwal R, Gotman J, Flanagan D, Rosenblatt B. Automatic EEG analysis during long-term monitoring in ICU. Electroencephalography and Clinical Neurophysiology. 1998 Jul; 107(1):44–58.

5. Hazarika N, Chen JZ, Tsoi AC, Sergejew A. Classification of EEG signals using wavelet transform. Signal Processing, Elsevier. 1997 May; 59(1):61–72.
6. Elif DU. Optimal classification of epileptic seizures in EEG using wavelet analysis and genetic algorithm. Computers in Biology and Medicine. 2008 Jan; 38(1):14–22.
7. Ocak H. Optimal classification of epileptic seizures in EEG using wavelet analysis and genetic algorithm. Signal Processing. 2008 Jul; 88(7):1858–67.
8. Ubeyli ED. Statistics over features: EEG signals analysis. Computers in Biology and Medicine. 2009 Aug; 39(1):733–41.
9. Ubeyli ED, Guler I. Features extracted by eigenvector methods for detecting variability of EEG signals. Pattern Recognition Letters. 2007 Apr; 28(5):592–603.
10. Aishwarya R, Prabhu M, Sumithra G, Anusiya M. Feature extraction for EMG based prosthesis control. ICTACT Journal on Soft Computing. 2013Jan; 3(2):472–7.
11. Dornhege G, Millan JR, Hinterberger T, McFarland DJ, Muller KR. Toward brain - computer interfacing. 1st edition, Massachusetts Institute of Technology (MIT) Press, Massachusetts; 2007.
12. Bhattacharyya S, Konar A, Tibarewala DN. A differential evolution based energy trajectory planner for artificial limb control using motor imagery EEG signal. Biomedical Signal Processing and Control. 2014 May; 11(1):107–13.
13. Electroencephalography (EEG) [Internet]. 2016 [cited 2016 Aug 2]. Available from: https://en.wikipedia.org/wiki/Electroencephalography.
14. Kashyap, Rangasami L. Optimal choice of AR and MA parts in autoregressive moving average models. Institute of Electrical and Electronics Engineers (IEEE) Transactions on Pattern Analysis and Machine Intelligence. 1982 Feb; 4(2):99–104.
15. Ramalingam VV, Mohan S, Sugumaran S. Prosthetic arm control using Clonal Selection Classification Algorithm (CSCA) - a statistical learning approach. Indian Journal of Science and Technology. 2016 Apr; 9(16):1–8.
16. Rani BJA, Umamakeswari A. Electroencephalogram-based brain controlled robotic wheel chair. Indian Journal of Science and Technology. 2015May; 8(S9):188–97.
17. Sugumaran V, Muralidharan V, Ramachandran KI. Feature selection using decision tree and classification through proximal support vector machine for fault diagnostics of roller bearing. Mechanical Systems and Signal Processing. 2007 Feb; 21(2):930–42.