2016 International Conference on Computational Intelligence and Networks

EEG based Motor Imagery Classification using SVM and MLP

Rajdeep Chatterjee[1], Tathagata Bandyopadhyay[2]
School of Computer Engineering
KIIT University
Bhubaneswar, India
[1]cse.rajdeep@gmail.com [2]tathagatabanerjee15@rocketmail.com

Abstract— **This paper focuses on the classification of motor imagery of the left-right hand movements from a healthy subject. Elliptic Bandpass filters are used to discard the unwanted signals. Our study was on C3 and C4 electrodes particularly for the left-right limb movements. We deployed various feature extraction techniques on the EEG data. Statistical-based, wavelet-based energy-entropy & RMS, PSD based average power and bad power were performed to form the desired feature vectors. Variants of Support Vector Machines (SVM) were employed for classification and the results were also compared with Multi-layered Perceptron (MLP). Empirical results show that both SVM and MLP were suitable for such motor imagery classifications with the accuracy of 85% and 85.71% respectively. Among all employed feature extraction techniques wavelet-based methods specifically the energy-entropy feature set, gave promising results for both the classifiers.**

Keywords-EEG; BCI; motor imagery; SVM; MLP; classification

I. INTRODUCTION

A non-profit organization named "2045 Initiative", stated on their site that their aim was, "to create technologies enabling the transfer of an individual's personality to a more advanced non-biological carrier, and extending life, including to the point of immortality. We devote particular attention to enabling the fullest possible dialogue between the world's major spiritual traditions, science and society". Their quest for life extension, featured a project called "Avatar", which has four levels. The very first level, Avatar *A* aims at a humanoid robot remotely connected and controlled by human mind. It should able to interpret instructions directly from the brain using the brain–computer interface (BCI) [1]. Brain activities sometime called "rhythms" is nothing but micro volt electric signals. (i.e. Electroencephalogram or EEG) generated in our brain while we are performing a task [2,3].

It concentrates about specific brain activities and tries to use them to instruct a computer or a robotic device to do a particular task. These activities are measured by various means (such as Electrophysiological & Hemodynamic BCI techniques) in which Electroencephalography (EEG) is widely used because it is non-invasive and cheap [4]. As per Donoghue et al.: "A major goal of a BMI (brain-machine interface) is to provide a command signal from the cortex. This command serves as a new functional output to control disabled BCIs provide an alternative way of communication

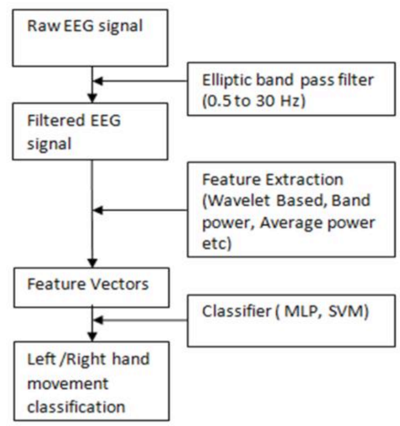

Figure 1. Flow of our study in this paper.

body parts or physical devices, such as computers or robotic limbs" [5]. BCIs provide an alternative way of communication i.e. "non-muscular" communication. This is one of the most important aspects the current BCI systems. Its main stress is to provide assistive devices for people with severe disabilities (people unable to perform physical movements). It is the primary aim of the BCI researchers to determine the right intention from the brain activities and reflect them into the desired movement accordingly [6,7]. This particular area in BCI is known as Motor Imagery (MI) movement [8-10]. In this paper we have worked on a data-set contained imagery of left-right hand movements.

We have applied support vector machines (SVM) and multi-layered perceptron to classify the data into its associative classes i.e. left/right hand movement [11-13]. We have used four variants of SVMs based on their kernels – linear, polynomial, radial basis and sigmoidal. Similarly after completing rigorous variations in learning rate & momentum for MLP, the best tuned variant is selected and compared with the SVMs. Before features are estimated from the raw EEG signal, filtering technique has been used to de-noise the signal. In simple words, de-noising, feature extraction and classification are performed. This paper tells about which variant of SVM and MLP provide the best results considering all the types of feature sets separately as well as combined. This paper is segmented in five sections- Section I is about

2375-5822/16 $31.00 © 2016 IEEE
DOI 10.1109/CINE.2016.22

introduction to BCI and our work, Section II tells about the experimental data-set, Section III briefly explains various features extraction techniques, Section IV shows the empirical results obtained from our study and finally Section V gives us the observation & the concluding remarks.

II. Experimental Preparation

In our study, we have taken the experimental data-set from BCI competition II 2003 provided by the Department of Medical Informatics, Institute for Biomedical Engineering, University of Technology Graz. According to their data description, it was collected during a feedback session where a normal subject (i.e. a female of 25 years old) was made to sit in a relaxing chair with armrests. The task was to control a feedback bar by means of imagery of left-right hand movement in which the order of left and right cues were random. The EEG data was sampled at 128Hz. Only C3 and C4 are considered for extracting information on left-right hand movement. We have used 140 trials with ten folds cross validation technique for each of the classifiers. As our topic is concentrated on two frequency bands namely the alpha (8-12Hz) and the beta band (18-25Hz), so the frequencies higher than this range can be treated as noise. Therefore an elliptic bandpass filter is used to filter in the frequency band of 0.5-30 Hz.

III. Feature Extraction

In literatures, we find few types of feature vectors were used for left-right imagery task discrimination. For comparison, we have used the most popularly used five types of feature extraction techniques. These methods are Statistical values (F_S), the wavelet based energy-entropy (F_W) & RMS values (F_{RMS}), Power Spectral Density (PSD) based average power (F_{AP}) and band power based (F_{BP}) feature vectors, and are briefly explained below [14-18].

A. Used feature extrection techniques

1) Statistical Features: Initially feature vector based on statistical method, was proposed only for physiological signals. Later it was extended to EEG signals as well [15, 16]. The formation of statistical features are defined as (*X(n)*, where *n = 1, ..., N.* is the raw *N*-sample EEG signal) given in the following.

- The mean of the raw signal-

$$\mu_X = \frac{1}{N}\sum_{n=1}^{N} X(n) \tag{1}$$

- The standard deviation of the raw signal-

$$\sigma_X = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(X(n)-\mu_X)^2} \tag{2}$$

- The mean of the absolute values of the first differences of the raw signal-

$$\delta_X = \frac{1}{N-1}\sum_{n=1}^{N-1}|X(n+1)-X(n)| \tag{3}$$

- The mean of the absolute values of the first differences of the standardized signal-

$$\bar{\delta}_X = \frac{1}{N-1}\sum_{n=1}^{N-1}|\bar{X}(n+1)-\bar{X}(n)| = \frac{\delta_X}{\sigma_X} \tag{4}$$

- The mean of the absolute values of the second differences of the raw signal-

$$\gamma_X = \frac{1}{N-2}\sum_{n=1}^{N-2}|X(n+2)-X(n)| \tag{5}$$

- The mean of the absolute values of the second differences of the standardized signal-

$$\bar{\gamma}_X = \frac{1}{N-2}\sum_{n=1}^{N-2}|\bar{X}(n+2)-\bar{X}(n)| = \frac{\gamma_X}{\sigma_X} \tag{6}$$

where $\bar{X}(n)$ is the normalized signal, i.e. $\bar{X}(n) = \frac{(X(n)-\mu_X)}{\sigma_X}$. The corresponding FV is defined as $F_S = [\mu_X, \sigma_X, \delta_X, \bar{\delta}_X, \gamma_X, \bar{\gamma}_X]$.

2) Wavelet-Based Energy & Entropy: In [17] Murugappan et al. introduced a new technique, wavlet analysis, for feature extraction from EEG signals. Unlike Short Time Fourier Transform (STFT), as this method does not have time-frequency tradeoff for features of non-stationary signal like EEG, it can be an important tool in signal processing. The discrete wavelet transforms the signals by decomposing it into coarse approximation and detail information. It uses hierarchical pyramid algorithm and in each step it halves the actual input signal by 2. The Detail D1 and approximation A1 are obtained from the first high-pass and low-pass filters after transforming the input signal i.e. the first level decomposition. The first approximation is further decomposed and the process is repeated for another two times. In this paper, we used Daubechies (db) mother wavelet of order 4 on filtered signal and the D3 features i.e. the detailed coefficients obtained after level three decomposition for the respective electrodes, were used to estimate the wavelet energy and entropy [19],

$$ENG_l^{C3} = \sum_{n=1}^{2^{S-l}-1}|C_X(l,n)|^2 \tag{7}$$

$$ENT_l^{C3} = -\sum_{n=1}^{2^{S-l}-1}|C_X(l,n)|^2 \log(|C_X(l,n)|^2) \tag{8}$$

$$N = 2^S, \qquad 1 < l < S$$

The parameters of (7) and (8) were used as a feature vector, i.e., $F_W = [ENG_l^{C3}, ENT_l^{C3}, ENG_l^{C4}, ENT_l^{C4}]$.

3) Wavelet-based RMS: In discrete wavelet transform $C_X(l,n)$ is the n^{th} sample of level *l* decomposition. Root Mean Square (RMS) values are obtaied as another feature, by using Daubechies (db) wavelet of order 4 and the D3 features, already stated the earlier. The feature vector is $F_{RMS} = [RMS_l^{C3}, RMS_l^{C4}]$.

$$RMS_l^{C3} = \sqrt{\frac{1}{N}\sum_{n=1}^{N}|C_X(l,n)|^2} \tag{9}$$

4) Average power: Power Spectral density methods are used to extract information from a signal and it explains how its power is varying in the frequency domain i.e. its power distribution. We have taken, the Welch approach along with a Hamming window of length 64. The used method divides the times series data into overlapping segments and then computing a modified periodogram of the each segment. Finally the PSD estimates is averaged to obtain the average power. The PSD estimation were done on two frequency bands, namely the alpha band (8-12Hz) and the beta band (18-25Hz) for each respective electrode. We have used (12) and (13) for estimating the average power instead of considering the exact values as features [20].

$$AVP_{\alpha}^{C3} = AveragePower(\sum_{f=a}^{b} PSD_{C3}(f)) \tag{10}$$

$$AVP_{\alpha}^{C4} = AveragePower(\sum_{f=a}^{b} PSD_{C4}(f)) \tag{11}$$

$$POW_{\alpha}^{1} = AVP_{\alpha}^{C4} - AVP_{\alpha}^{C3} \tag{12}$$

$$POW_{\alpha}^{2} = \frac{(AVP_{\alpha}^{C4} + AVP_{\alpha}^{C3})}{2} \tag{13}$$

Where, $PSD_{C3/C4}$ is the PSD estimations of the respective electrodes in the given band range $[a, \; b]$, where a & b are the frequency ranges for alpha band (8-12Hz) and beta band (18-25Hz), POW_{α}^{1} & POW_{α}^{2} are the average powers for alpha band and similarly the average powers for the beta band are also calculated. The obtained feature vector is $F_{AP} = [POW_{\alpha}^{1}, POW_{\alpha}^{2}, POW_{\beta}^{1}, POW_{\beta}^{2}]$.

5) Average Band power: It determines the percentage of the total power in a specified frequency interval. So we have computed band powers for alpha and beta frequency intervals. Again we use (16) and (17) for estimating the average band powers. The feature vector is $F_{BP} = [BP_{\alpha}^{1}, BP_{\alpha}^{2}, BP_{\beta}^{1}, BP_{\beta}^{2}]$.

$$BAND_{\alpha}^{C3} = \frac{\sum_{f=a}^{b} X(f)}{\sum_{n=1}^{N} X(n)} * 100 \tag{14}$$

$$BAND_{\alpha}^{C4} = \frac{\sum_{f=a}^{b} X(f)}{\sum_{n=1}^{N} X(n)} * 100 \tag{15}$$

$$BP_{\alpha}^{1} = BAND_{\alpha}^{C4} - BAND_{\alpha}^{C3} \tag{16}$$

$$BP_{\alpha}^{2} = \frac{(BAND_{\alpha}^{C4} + BAND_{\alpha}^{C3})}{2} \tag{17}$$

where *a & b* is indicate the frequency ranges for alpha band (8-12Hz) and also for the beta band (18-25Hz), $BAND_{\alpha}^{C3}$ & $BAND_{\alpha}^{C4}$ are the average band powers for alpha band and similarly we repeat the process for the beta band.

Statistical feature set, wavelet-based energy-entropy, wavelet-based RMS, PSD based average power and average band power, have 12, 4, 2, 4 and 4 features respectively. These numbers are also important because we have applied classifiers on them separately as well as we have combined all the five feature vectors as one and formed a combined feature vector $F_{ALL} = [F_S, F_W, F_{RMS}, F_{AP}, F_{BP}]$.

TABLE I. DETAILS OF FEATURE VECTORS USED

Feature vectors	Size (No. of trials x No. of Features per trial)
Statistical-based	140 x 12
Wavelet-based Energy & Entropy	140 x 04
Wavelet-based RMS	140 x 02
Average Power (PSD-based)	140 x 04
Band power	140 x 04
Total	140 x 26

IV. CLASSIFICATION AND EMPIRICAL RESULTS

One of the very important applications of BCI based on EEG signal is to drive a prosthetic or assistive device. It requires classification of the EEG signals, where a class corresponds to the movement of a particular limb. In our survey prior to this piece of work, we found that researchers commonly used Linear Discriminant Analysis (LDA), Quadratic Analysis (QDA) and K-Nearest Neighbor (KNN) as classifiers in BCI research. Although there are other classifiers that can be used in motor imagery tasks as well. Here our focus is on rigorous study of the various types of SVMs and MLP in the MI (motor imagery) task and also kept the feature sets as minimum as possible. We have used MATLAB R2014a for feature extraction and WEKA 3.6 classification tool to classify the obtained feature vectors in an Intel Core i5 computer.

Table II contains the types and variety of classifiers used in this research work. In this paper, eight variants of support vector machines (SVM) on the basis of kernel type and SVM type and customized Multi-layered Perceptron (MLP) are applied on each of the feature vectors and also on the total feature set. ROC area or the area under Receiver operating characteristic curve is a well-known metric for performance measurement of binary classifiers used in medical diagnosis and machine learning [21, 22]. Table III indicates the quality of prediction based on the estimated ROC area. Though we have little scope to discuss about it in this paper, ROC curve is a plot of false positive rate in x-axis and true positive rate in y-axis [23]. Table III signifies that if the area fall under 0.5 then there is some serious problem in the learning model and the predictions have gone wrong. It may happen when a data set contains mis-information or when a relatively very small data set is used for training whereas quite a large data set is used for testing. The area lies within 0.8 to 0.95 is indicating that the learning model is trained perfectly and also give quiet excellent predictions. In simple word, the higher the area means the higher the quality of prediction that means both the classes are clearly separable. Here we consider left hand movement as positive test whereas right hand movement as negative test.

In Table IV, we have forty eight different variants of SVM i.e. eight variants for each of the six feature vectors. In Table V, we obtain results from our tuned MLP. It shows the classifiers' accuracies and their ROC areas. The bold numbers signify the best results for the individual feature sets. For better understanding, the best ROC areas and the best FP rates (weighted average) for each classifier are represented in bar-

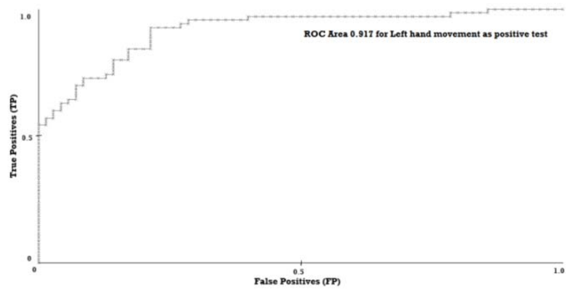

Figure 2. Left hand MI ROC Area plot from F_W for MLP.

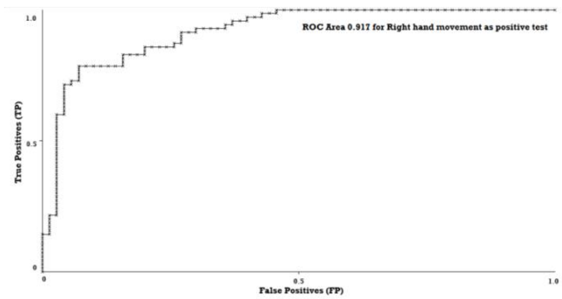

Figure 3. Right hand MI ROC Area plot from F_W for MLP.

charts in Fig. 4 and Fig. 5 respectively. The best result for MLP is 85.71% and for SVM it is 85%. In observation we can state that SVM gives mixed results, some are good but some are worst. However, for each feature vectors at least one of the variant of SVM provides very satisfactory results along with a descent ROC area.

TABLE II. TYPES AND VARIATIONS OF CLASSIFIERS USED

Classifier Name	**Variant type**	**Kernel functions used**			
SVM	C-SVC (C)	Linear (L)	Polynomial (P)	Radial Basis (R)	Sigmoidal (S)
	Nu-SVC (N)	Linear (L)	Polynomial (P)	Radial Basis (R)	Sigmoidal (S)
MLP	**Learning rate**		**Momentum**		
	0.7		0.29		

TABLE III. ROC AREA SIGNIFICANCE CHART

ROC Area	**Significance**
1.0	Perfect Prediction
0.9	Excellent Prediction
0.8	Good Prediction
0.7	Mediocre Prediction
0.6	Poor Prediction
0.5	Random prediction
<0.5	Something wrong!

Courtesy: www.cs.cornell.edu [23]

V. CONCLUSION

The best and second best results in SVM come from wavelet-based energy-entropy and wavelet-based RMS. The combined feature set F_{ALL} gives third best result. One of the reasons could be the presence of more overlapping features

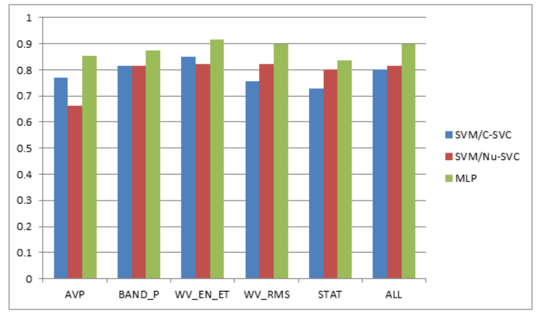

Figure 4. Best Mean ROC Area Chart for Best Cases.

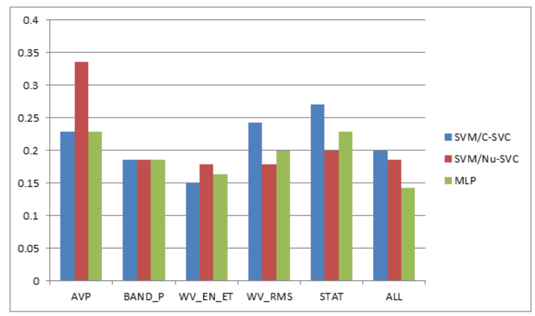

Figure 5. Weighted Average False-Positive Rate Chart for Best Cases.

which dilute its meaningful relevance. Although its correctness is third best, its ROC area (more than 0.8) is still impressive. Again we find that the ROC area of the wavelet-based energy-entropy is ranked first among all 48 variants. If we closely analyze the results obtained from MLP, all the feature vectors give accuracy higher than 75 percent and ROC areas above 0.8, which is significant in the EEG based motor imagery classifications. It validates that all of them contain relevant information for MI discrimination. It is also noticeable that we have used only 26 numbers of features whereas literatures involved quite large number of features to attain similar outputs. Even if we analyze them separately; wavelet-based energy-entropy gives accuracy as good as all the feature sets combined. Besides the average power and statistical features, MLP gives accuracy more than 80 percentages for all other sets. Although both the average power and the statistical features provide same level of correctness for MLP, the average power gives better ROC area than statistical feature set. So in conclusion average power has more meaningful information over statistical features. In another case F_{BP} has higher accuracy than F_{RMS}, however it has a lower ROC area.

In our study MLP (Table V) gives far better results than all the results combined from SVM (Table IV). In comparison, F_{ALL} from MLP has the highest accuracy along with the largest ROC area over SVMs. F_W , F_{RMS} and F_{ALL} , for these cases in MLP estimate ROC areas more than or equals to 0.9 i.e. excellent for any machine learning model. Out of all our experiments F_W provides highest ROC area i.e. 0.917 from MLP represented in Fig. 2 and Fig. 3. Finally we conclude by stating two observations from our extensive study, first-wavelet-based energy-entropy is very informative feature set among all other popular feature extraction techniques, second-based on the results, MLP classifier is more suitable for EEG-based motor imagery left/right hand movement.

Future study includes exploration of more feature extraction techniques to achieve higher classification performance with less number of features. As higher dimensionality being a matter of concern for EEG data classification, various data preprocessing techniques can also be examined in this regard.

ACKNOWLEDGMENT

Our special thanks to Dr. Debarshi Kumar Sanyal, Associate Professor, School of Computer Engineering, KIIT University for his supervision and continuous support.

REFERENCES

[1] Avatar Program –A 2045 Initiative. http://www.2045.com.

[2] Anderson R.A., Musallam S., Pesaran B., "Selecting the signals for a brain-machine interface", Curr Opin Neurobiol Vol.l4 (6), pp.720-726, December 2004.

[3] Saeid Sanei and J.A. Chambers, EEG SIGNAL PROCESSING. Centre of Digital Signal Processing Cardiff University, UK, 2007.

[4] Bernhard Graimann, Brendan Allison, and Gert Pfurtscheller, "Brain–Computer Interfaces: A Gentle Introduction," Brain Computer Interfaces, The Frontiers Collection, Springer-Verlag Berlin Heidelberg 2010.

[5] J.P. Donoghue, Connecting cortex to machines: recent advances in brain interfaces. Nat Neurosci. 5 (Suppl), 1085–1088, November 2002.

[6] Schwartz A.B., Cui X.T., Weber DJ., Moran D.W. "Brain Controlled Interfaces: Movement Restoration using Neural Prosthetics." Neuron vol. 52, pp. 205-220, October 2006.

[7] Lebedev M.A., Nicoleis, "Brain-machine interface: Past, present and future", Trends Neurosci. vol. 29(9), pp.536-546, September 2006.

[8] Xu Huaiyu, Lou Jian, Su Ruidan, Zhang Erpang "Feature Extraction and Classification of EEG for imaging left-right hands movement,"

[9] Pfurtscheller G, Neuper C, Schlögl A, Lugger K., "Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters," IEEE Trans Rehabil Eng. vol. 6(3), pp. 316-25, 1998.

[10] G. Pfurtscheller and C. Neuper, "Motor imagery and direct brain-computer communication," Proceedings of the IEEE, Vol. 89 , Issue. 7, pp. 1123 – 1134, July 2001.

[11] Martin T. Hagan, Howard B. Demuth and Mark Beale, Neural Network Design. PWS Publishing Company, 1996.

[12] R.Shantha Selva Kumari, J.Prabin Jose, "Seizure Detection In EEG Using Time Frequency Analysis and SVM ," International Conference on Emerging Trends in Electrical and Computer Technology (ICETECT), pp. 626 – 630, March 2011.

[13] Nello Cristianini and John Shawe-Taylor, An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press, 2000.

[14] R. W. Picard, E. Vyzas, and J. Healey, "Toward machine emotional intelligence: Analysis of affective physiological state," IEEE Trans. Pattern Anal. Mach., vol. 23, no. 10, pp. 1175–1191, October 2001.

[15] K. Takahashi, "Remarks on emotion recognition from bio-potential signals," in Proc. of 2nd Int.Conf. Autonomous Robots Agents, pp. 186–191, 2004.

[16] Panagiotis C. Petrantonakis and Leontios J. Hadjileontiadis, "Emotion Recognition From EEG Using Higher Order Crossings", IEEE TRANSACTIONS ON INFORMATION TECHNOLOGY IN BIOMEDICINE, Vol. 14, No. 2, March 2010.

[17] M. Murugappan, M. Rizon, R. Nagarajan, S. Yaacob, I. Zunaidi, and D. Hazry, "EEG feature extraction for classifying emotions using FCM and FKM," J. Comput. Commun., vol. 1, pp. 21–25, 2007.

[18] I. Daubechies, "Orthonormal bases of compactly supported wavelets," Commun. Pure Appl. Math., vol. 41, pp. 909–996, 1988.

[19] Dan Xiao, Zhengdong Mu and Jianfeng Hu, "Classification of Motor Imagery EEG Signals Based on Energy Entropy," International Symposium on Intelligent Ubiquitous Computing and Education, pp. 61 – 64, May 2009.

[20] Saugata Bhattacharya, Anwesha Khasnobish, Somsirsa Chatterjee, Amit Koner and D.N. Timberwala, "Perormance Analysis of LDA, QDA and KNN Algorithms in Left-Right Limb Movement Classification from EEG Data," International conference on Systems in Medicine and Biology, December 2010.

[21] Fawcett, Tom, "An Introduction to ROC Analysis," Pattern Recognition Letters, vol. 27(8), pp. 861 – 874, 2006.

[22] Powers, David M W, "Evaluation: From Precision, Recall and F-Measure to ROC, Informedness, Markedness & Correlation," Journal of Machine Learning Technologies, vol. 2 (1), pp. 37–63, 2011.

[23] Performance Measures for Machine Learning. (Cornell University) http://www.cs.cornell.edu/courses/cs578/2003fa/performance_measures.pdf

TABLE IV. CLASSIFICATION RESULTS FOR SVM

		CL	CP	CR	CS	NL	NP	NR	NS
F_{AP}	**Accuracy**	**77.1429**	59.2857	**77.1429**	**77.1429**	66.4236	56.4286	66.4236	66.4236
	ROC Area	**0.771**	0.593	**0.771**	**0.771**	0.664	0.564	0.664	0.664
F_{BP}	**Accuracy**	81.4286	48.5714	62.8571	50	80.7143	**81.4286**	65.7143	50
	ROC Area	0.814	0.486	0.629	0.5	0.807	**0.814**	0.657	0.5
F_W	**Accuracy**	**85**	**85**	78.5714	47.1429	82.1429	75.7143	78.5714	42.8571
	ROC Area	**0.85**	**0.85**	0.786	0.471	0.821	0.757	0.786	0.429
F_{RMS}	**Accuracy**	75.7143	65	75.7143	75.7143	**82.1429**	51.4286	81.4286	**82.1429**
	ROC Area	0.757	0.65	0.757	0.757	**0.821**	0.514	0.814	**0.821**
F_S	**Accuracy**	72.8571	63.5714	72.8571	72.8571	80	50.7143	**78.5714**	77.1429
	ROC Area	0.729	0.636	0.729	0.729	0.8	0.507	**0.786**	0.771
F_{ALL}	**Accuracy**	80	78.5714	77.1429	50.7143	80.7143	**81.4286**	75	50
	ROC Area	0.8	0.786	0.771	0.507	0.807	**0.814**	0.75	0.5

TABLE V. CLASSIFICATION RESULTS FOR MLP

	F_{AP}	F_{BP}	F_W	F_{RMS}	F_S	F_{ALL}
Accuracy	77.1429	81.4286	**83.5714**	80	77.1429	**85.7143**
ROC Area	0.854	0.874	**0.917**	0.899	0.836	**0.9**