Volume 4, Issue 8, August 2014 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Generation, Filtering, Feature Extraction, Classification and Forecasting of Electroencephalogram (EEG) Signals

Archana P
Dept. of Electronics & communication
PG Scholar, Oxford college of Engg.,
Bangalore, VTU, Belgaum, India

Shivaramaraju C. H
Dept. of Computer Science & Engg.
PG Scholar
Bangalore, India

Sandya. H B
Dept. of Telecommunication & Engg.
Asst. Prof., RV college of Engg.,
Bangalore, India

Abstract— ***Electroencephalogram (EEG) signals Generation, Pre-processing, Feature Extraction, Classification and Forecasting of EEG signals are core issues in EEG signal analysis. This research focuses on presenting a novel approach to classify EEG signals into diseased EEG signals and Normal EEG signals based on the extracted features. The Classification of EEG signals is of great medical importance. The EEG signals are generated using an Event Related Potentials (ERP) of phase-resetting method. The pre-processing of EEG is carried using Savitzky – Golay (S-G) filter. Feature extraction is done based on mathematical calculations. The final step of classification of EEG signals is carried using machine learning and pattern recognition technique such as Neural Network (NN) based techniques. The Forecasting of EEG signal focus on forecasting future samples of EEG signals in advance using Neural Network based time series forecasting technique***.

Keywords: EEG, ERP, S-G Filter, Feature Extraction, Forecasting and NN

I. INTRODUCTION

Electroencephalography (EEG) is the recording of electrical activity along the scalp. EEG measures voltage fluctuations resulting from ionic current flows within the neurons of the brain. In clinical contexts, EEG refers to the recording of the brain's spontaneous electrical activity over a short period of time. Time series measured in real world is frequently non-stationary and to extract important information from the measured time series it is significant to utilize a filter or smoother as a pre-processing step, In order to reduce the effect of noise on the original signal Savizky-Golay (S-G) Filter is used. This low pass filter is suitable for smoothing data. Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. Transforming the input data into the set of features is called Feature Extraction. Classification is a hierarchical arrangement of kinds of things (classes) or groups of kinds of things. Forecasting is a process of predicting or estimating the future based on past and present data. Forecasting provides information about the potential future events and their consequences for the organization.

II. METHODOLOGY

In this system initially generated Electroencephalogram (EEG) signal data set in mat lab using phase-reset method, secondly generated signal is filtered using Savitzky-Golay filter in order to smooth-out the noisy data, thirdly Features are extracted from the burst data set by using Average method, Maximum method, Minimum method, Standard Deviation method and Variance method, fourthly Extracted Feature data set and diseased EEG data set is applied as an input to Neural Network for the purpose of classification, and at last is the forecasting of future samples that is carried out using Neural Network.

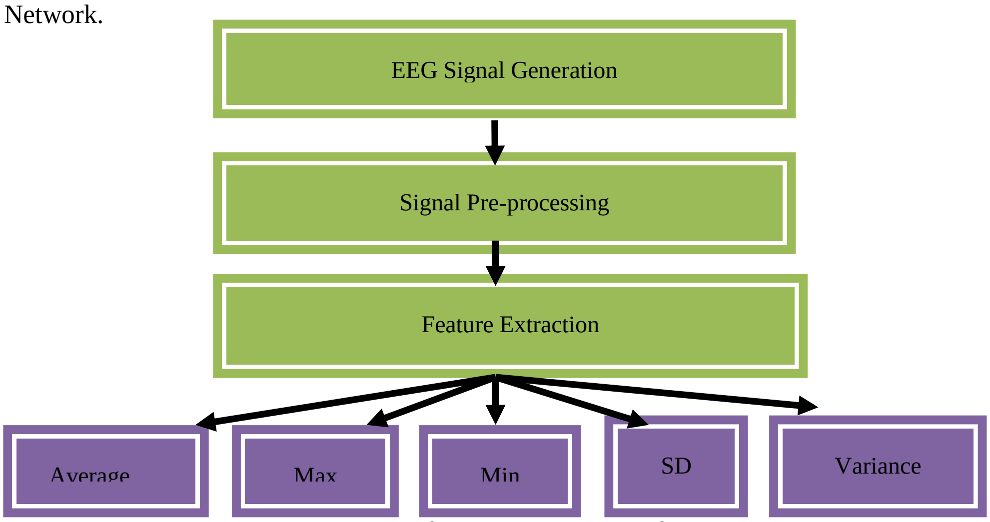

Fig. 1 Proposed Methodology for Feature Extraction of EEG Signal

A. Normal Electroencephalogram (EEG) Signal Generation:

EEG data is generated by an Event Related Potentials (ERP) of phase-resetting method. Frames, epochs, sampling-rate, frequency range and position are the parameters used for EEG signal generation. The parameter position defines the frame in which the reset should occur. EEG Signals are generated using Mat lab. Phase-reset allows generating a sinusoid whose phase is being reset. 10 different Normal EEG signals are generated; one such EEG signal generated has inputs defined as: frames-300, epoch-1, sampling rate-100Hz and position of phase reset-150, with frequencies chosen randomly from range 4Hz-16Hz, since EEG spectrum contains some characteristic waveforms that fall primarily in frequency range - within four frequency bands—delta (1–4 Hz), theta (4–8 Hz), alpha (8–13 Hz), and beta (above 13 Hz).EEG signal generated using this data is shown in Fig.2.

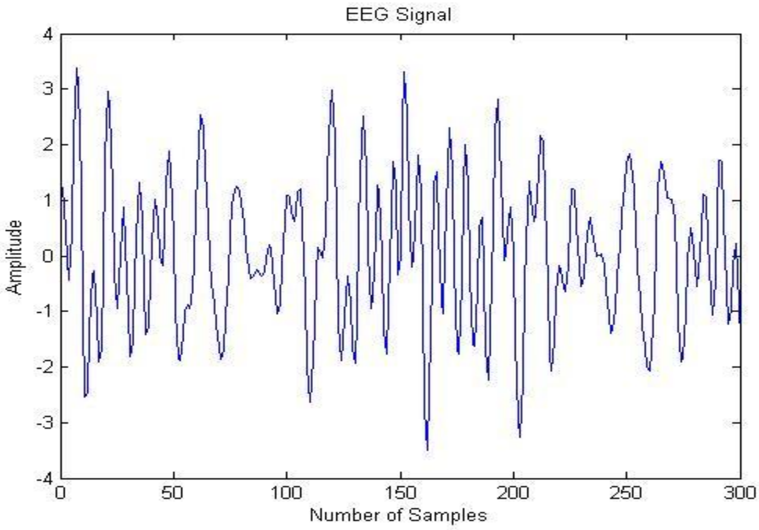

Fig. 2 Normal EEG signal generated using phase-reset method

B Filtering using Savitzky-Golay Filter:

EEG signal is pre-processed using Savizky-Golay Filter, sagolayfilt (X, k,F) smoothest the signal X using a Savitzky-Golay smoothing filter. The polynomial of order k must be less than the frame size, F (F must be odd). The length of the input X must be greater than or equal to F. If the polynomial order k equals f-1 then no smoothing occurs. Signal is filtered using the parameters: generated signal-X, polynomial order-2 and frame size-9. EEG signal in Fig.2 is filtered using S-G filter is shown in Fig.3.

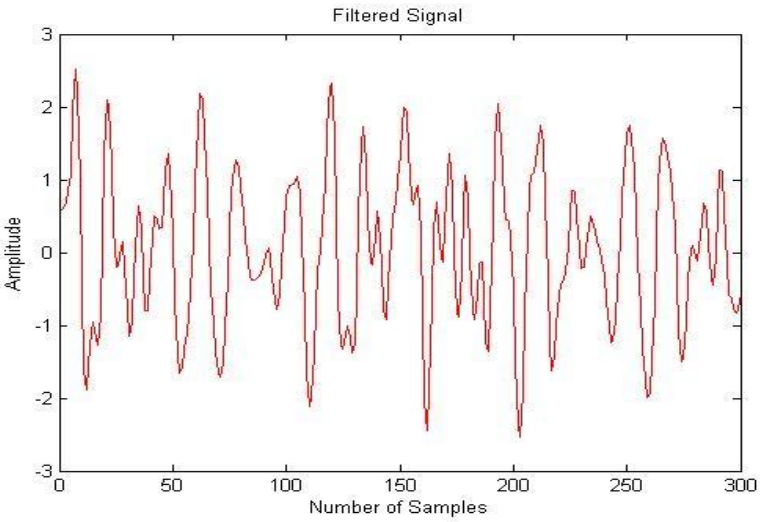

Fig. 3 Filtered Signals

C. Diseased Electroencephalogram (EEG) signal Generation:

Diseased EEG signals are generated using random phase reset method. Frames, epochs, sampling-rate are the parameters used for diseased EEG signal generation, the first three parameter of this function are the frame which indicates number of frames used for signal generation, epochs represents number of trials and third parameter indicates sampling rate of the signal, here the parameter: position, defining in which frame the reset should occur is absent, phase reset occur randomly. Diseased EEG signal is shown in Fig.4.

Diseased EEG Signals are generated using Mat lab, 10 different diseased EEG signals are generated; one such EEG signal generated has inputs defined as: frames-30, epoch-1, sampling rate-65Hz.

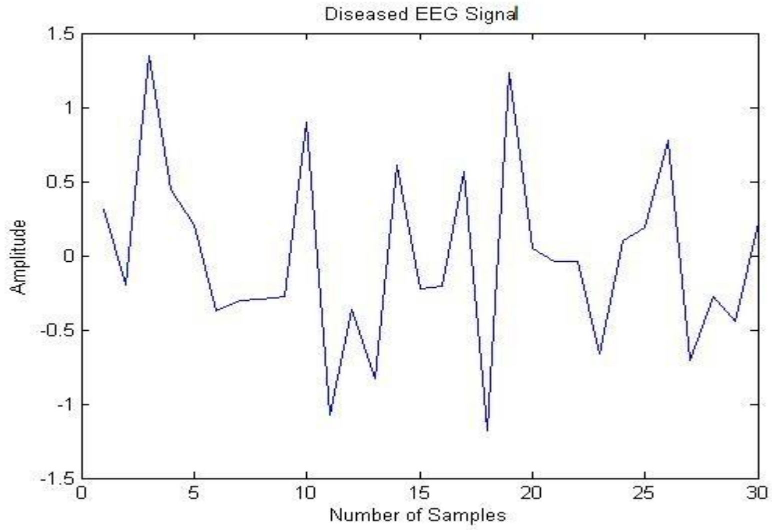

Fig.4. Diseased Electroencephalogram (EEG) signal Generated using phase-reset method

D. Feature Extraction:

Features are extracted based on basic statistics like minimum, mean, maximum, standard deviation and variance. The input signal size is 300 samples, the features are extracted for an interval of every 10 samples, so at the end there will be 30 features extracted for each of the parameter, since there are 5 parameters (min, mean, max, SD and variable) totally there will be 30 * 5= 150 features extracted for each signal. There are 10 Normal EEG signal, so totally there will be 1500 features extracted from 3000 samples. The extracted feature from the filtered signal using 5 parameters is shown in table 1

Table 1 Feature Extracted Result

Feature Extracted Result					
No. of Signals	**Minimum**	**Average**	**Maximum**	**Std. Dev.**	**Variance**
1	-0.50866	1.098614	2.516638	0.900551	0.810991
2	-1.89255	-0.88416	1.357794	0.963812	0.928935
3	-0.77844	0.453169	2.101164	0.988866	0.977856
4	-1.14132	-0.33252	0.64419	0.612614	0.375296
5	0.202152	0.639003	1.364992	0.412028	0.169767
6	-1.65681	-0.79449	0.97369	0.826637	0.683328
. . .	. . .	. . .	. . .	. . .	. . .
25	-1.23113	-0.15631	1.646502	1.069009	1.142779
26	-1.98372	-0.28527	1.744887	1.459619	2.130487
27	-1.47315	0.672297	1.564352	1.024889	1.050398
28	-1.49974	-0.52773	0.336958	0.669129	0.447733
29	-0.45591	0.152124	0.678548	0.406766	0.165459
30	-0.82654	-0.14637	1.131242	0.786705	0.618905

E. Classification using Neural Network:

The generated 10 Normal EEG signals have 1500 features.1500 Extracted feature dataset is arranged in terms of 50 columns and 30 rows, where each of the columns represents the parameters min, mean, max, SD and var..Target assigned for Normal EEG signals are ‘1’, is shown in Table 2 and Table 3.

Diseased EEG signal has 300 samples, arranged in terms of 10 columns and 30 rows. Target assigned for Diseased EEG signals are ‘0’, is shown in Table 5 & 6.Normal EEG and diseased EEG signal been classified will have exact output as the targets assigned, we find that the Normal EEG signal results in output that are very close to ‘1’, shown in Table 4 and that Diseased EEG signal results in output that are very close to ‘0’, shown in Table 7. Input data set: first 50 columns with 30 rows represent Normal EEG and next 10 columns (i.e 51 to 60) with 30 rows represent Diseased EEG signal, totally 1800 samples are fed as an input to Neural Network classification. How accurately Neural Network classification is performed can be checked using confusion matrix in Fig.7 and validation performance plot is shown in Fig. 8.

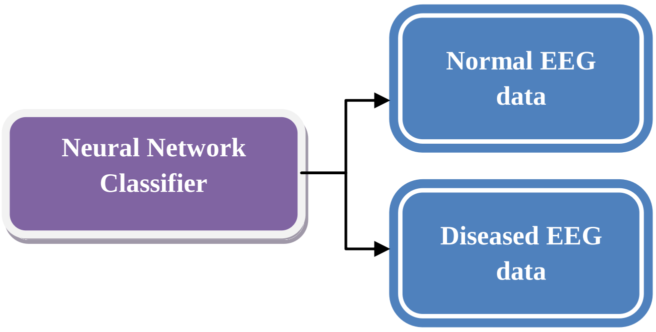

Fig. 5 Proposed Methodology for Classification of an EEG Signal

Table 2 Normal Eeg Signal Input With 1500 Features

	1	2	3	....	50
1	1.00742	1.00742	1.00742	………..	0.930079
2	0.95805	-0.09107	0.846662	………..	0.369834
3	-1.0896	-0.07918	1.03919	………..	1.44796
4	1.02081	0.082596	0.986409	………	0.448014
.	………	…….	……..	………	……..
30	1.66462	0.12981	0.743723	………	0.794226

Table 3: Targets Assigned To Normal Eeg

1	2	3	….	50
1	1	1	….	1

Output obtained when Normal EEG signals are fed to Neural Network, find that outputs are very close to'1'.Output obtained when Diseased EEG signals are fed to Neural Network, find that outputs are very close to'0'

Table 4 Output of Normal Eeg Signal

1	2	3	..	50
0.998069	0.993062	0.999912	..	0.999939

Table 5 Diseased Eeg Signal Input With 300 Features

	51	**52**	**53**	**….**	**60**
1	0.569389	1.084726	0.625307	**….**	1.81411
2	1.282916	0.95338	1.151854	**….**	1.015869
3	0.980156	-0.81077	0.33249	**….**	1.334216
4	2.536717	0.377776	0.432442	**….**	1.165002
….	**….**	**….**	**….**	**….**	
30	0.436988	-0.53094	0.755814	**….**	-1.75619

Table 6 Targets Assigned To Diseased Eeg Signal

51	52	53	..	60
0	0	0	..	0

Table 7 Output Of Normal Eeg Signal

51	52	53	..	60
0.001578	4.40E-05	0.003488	..	0.001549

Table 8 Neural Network Classification Output

	Input					
	1	2	....	48	49	50
1	-1.00742	-1.00742	....	1.153376	0.754579	0.569389
2	-0.95805	-0.09107	....	1.957212	1.132659	1.282916
3	-1.0896	-0.07918	....	2.053261	0.990028	0.980156
4	-1.02081	0.082596	....	1.665297	1.59277	2.536917
....	....	....	....	1.802217	1.794851	3.22149
28	-1.01658	-0.07297	....	2.014362	1.209262	1.462315
29	-0.9801	-0.10555	....	1.991284	1.441726	2.078575
30	-1.66462	-0.12981	....	2.33285	1.361981	1.854991
Target Assigned						
	1	1	....	1	1	1
Output						
	0.998069	0.993062	....	0.999995	0.999874	0.999939

	Input					
	51	52	....	58	59	60
1	1.084726	0.625307	....	1.084726	0.625307	1.084726
2	0.95338	1.151854	....	0.95338	1.151854	0.95338
3	-0.81077	0.33249	....	-0.81077	0.33249	-0.81077
4	0.377776	0.432442	....	0.377776	0.432442	0.377776
28	-0.5582	-1.65096	....	-1.19966	-0.20999	-0.3744
29	-0.42361	-1.18953	....	0.278345	-0.40009	-0.65478
30	-0.36847	-1.0403	....	-0.18012	-0.23396	-1.75619
Target Assigned						
	0	0	....	0	0	0
Output						
	0.001578	4.40E-05	....	0.003624	0.030545	0.001549

F. Forecasting of EEG signal:

The set of inputs are fed to the Neural Network for forecasting of future samples. For the input feed, the first predicted output sample is resulted. Now the resulting predicted output set is feed as an input to Neural Network, for forecasting next set of samples. The above process is repeated for forecasting future samples. Here there is 1:2 delay introduced in the system during the forecasting process, so we find that first two samples are missing during each forecasted output. The Forecasted output is shown in Table 9.

Fig 6 Proposed Methodologies for Forecasting EEG Signal

G. Confusion Plot:

In the field of artificial intelligence, a confusion matrix is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one. One measure of how well the neural network has fit the data is the confusion plot. Here the confusion matrix is plotted across all samples. The confusion matrix shows the percentages of correct and incorrect classifications. Correct classifications are the green squares on the matrices diagonal. Incorrect classifications form the red squares. If the network has learned to classify properly, the percentages in the red squares should be very small, indicating few misclassifications. If this is not the case then further training, or training a network with more hidden neurons, would be advisable. A plot of confusion matrix is shown in Fig.7.

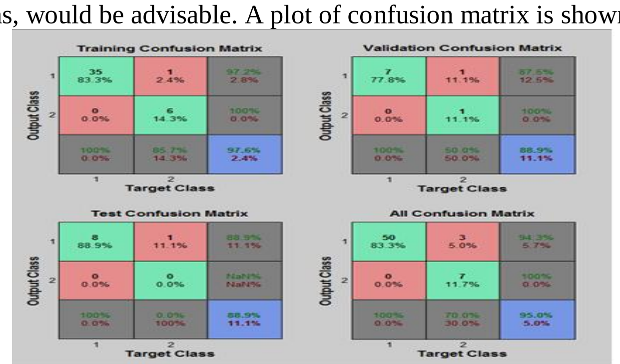

Fig. 7 Confusion Plot

H. Mean square error (MSE):

MSE is a network performance function. It measures the network's performance according to the mean of squared errors. Mean squared error (MSE) of an estimator is one of many ways to quantify the difference between values implied by an estimator and the true values of the quantity being estimated. MSE is a risk function, corresponding to the expected value of the squared error loss or quadratic loss. MSE measures the average of the squares of the errors. The error is the amount by which the value implied by the estimator differs from the quantity to be estimated. The difference occurs because of randomness or because the estimator doesn't account for information that could produce a more accurate estimate. The performance plot is shown in Fig.8.

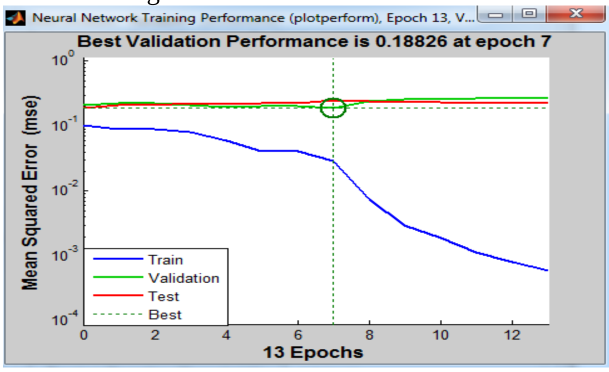

Fig.8.Validation performance plot

Table 9: Forecasted Output

Sl No.	Input signal	1st NN Forecasting	2nd NN Forecasting
1	25		
2	25.25		
3	25.5	25.50001496	
4	25.75	25.74999095	
5	26	25.99999253	26.00018505
6	26.25	26.25000173	26.25009087
7	26.5	26.50000784	26.49991164
8	26.75	26.75000673	26.74977045
9	27	26.99999959	26.99974239
. . .	. . .	. . .	. . .
95	48.5	48.5002647	48.50000284
96	48.75	48.75035626	48.74999025
97	49	49.00038248	48.99997108
98	49.25	49.25030628	49.24994473
99	49.5	49.5000894	49.49991922
100	49.75	49.7496933	49.74991144
101	50	49.99907993	49.99993854
Ist Forecasted Value		50.24821203	50.25000047
2nd Forecasted Value			50.50005722

III. RESULT

Features are extracted based on mathematical calculations like minimum, mean, maximum, standard deviation and variance. The input signal size is 300 samples, the features are extracted for an interval of every 10 samples so at the end there will be 30 features extracted from each signal. Totally 30 * 5=150 features are extracted. For filtered signal in figure 3, Features are extracted where the Feature Extraction Result containing 150 features is shown table 1. Using neural network classifies the signals as normal EEG and diseased EEG signal and assigned target as 1 for normal EEG signal and assigned 0 for diseased EEG signal as shown in table 3 and 5 and resulted table as shown in table 4 & 6. Using Mean square error method future signals are forecasted as shown in table 9.

IV. CONCLUSION

The research can be summarised as designed and implementation of signal generation, pre-processing, Feature Extraction and Classification of EEG signals. As EEG signal were generated artificially using phase resetting methods, the generated EEG signals and real EEG signals are very much similar in their frequency, amplitude and their phases. In the pre-processing step a Savitzky-Golay filter is applied to remove the noise and smoothen the EEG signals. The important and distinct features are extracted using 5 mathematical parameters such as mean, maximum, minimum, standard deviation and variance. This extracted feature represents the unique features of EEG signals which are used for the Classification purpose. The normal EEG signals and diseased EEG signals are classified using Neural Network based techniques. The future samples are successfully forecasted using Neural Network based on time series predictions.

REFERENCES

[1] Nandish.M, Stafford Michahial, Hemanth Kumar P, Faizan Ahmed, " Feature Extraction and Classification of EEG Signal Using Neural Network Based Techniques", International Journal of Engineering and Innovative Technology (IJEIT) Volume 2, Issue 4, October 2012

[2] Makinen, VT, Tiitinen, H, May, PJC, "Auditory Evoked Responses Are Additive to Brain Oscillations", Neurology and Clinical Neurophysiology 2004:45 (November 30, 2004)

[3] Bahareh-Atoufi, Caro-Lucas, Ali – Zakerolhosseini, "A Survey of Multi-Channel Prediction of EEG Signal in Different EEG States: Normal, Pre-Seizure, and Seizure". University of Tehran, University of ShahidBeheshti Tehran, Iran.

[4] AbdulhamitSubasi, Ergun Ercelebi, "Classification of EEG signals using neural network and logistic regression", Computer Methods and Programs in Biomedicine (2005) 78, 87—99

[5] Ayan Banerjee, KanadBasu and Aruna Chakraborty, "Prediction of EEG Signal by Digital Filtering",University of Florida, branch of Computer & Information Science and Engineering.

[6] Ronald W. Schafer, "What Is A Savitzky-GolayFilter?", IEEE Signal Processing Magazine July 2011. Digital Object Identifier 10.1109/Msp.2011.941097

[7] Neural networks III: The delta learning rulewithsemilinear activation function

[8] Martin F. Moller, "A Scaled Conjugate Gradient Algorithm for Supervised learning", Computer science Department Aarhus University NyMunkegade, Denmark. DAIMI P B-339

[9] Bogdan M. Wilamowski, Yixin Chen, "Effecient Algorithm for Training Neural Networks with one hidden layer", 0-7803-5529-6/99,1999 IEEE.

[10] Vadim V. Nikulin, Klaus Linkenkaer-Hansen, Guido Nolte, Steven Lemm, Klaus R. Muller, Risto J. Ilmoniemi and Gabriel Curio, "A novel mechanism for evoked responses in The human brain", European Journal of Neuroscience, Vol. 25, pp. 3146–3154, 2007

[11] HU Jian-feng, "Multifeature analysis in motor imagery EEG classification", Proc. IEEE, 2010. Third International Symposium on Electronic Commerce and Security, pp.114-117, 2010.

[12] Damien Coyle, Member, IEEE, Girijesh Prasad, Member, IEEE, and Thomas Martin Mcginnity, Member, IEEE, "A Time-Series Prediction Approach For Feature Extraction in A Brain–Computer Interface", IEEE Transactions on Neural Systems and Rehabilitation Engineering, Vol. 13, No. 4, December 2005.

