Multichannel Feature Extraction and Classification of Epileptic States Using Higher Order Statistics and Complexity Measures

K. Palani Thanaraj[#1], K. Chitra [#2]

[#1] Department of Electronics and Instrumentation Engineering,
St. Joseph's College of Engineering, Anna University, Chennai- 600119, India.
[#2] Department of Electronics and Communication Engineering,
St. Joseph's College of Engineering, Anna University, Chennai-600119, India.
[1] palanithanaraj.k@hotmail.com
[2] chitra_kris@yahoo.com

Abstract—Epilepsy is a brain dysfunction that is characterized by recurrent seizures. An important analysing tool in detection of epilepsy is Electroencephalogram (EEG). The random and non-linear nature of the EEG imposes great difficulty in understanding the pathological process. In this work a multichannel epilepsy detection system is proposed. A feature vector is formed by performing Higher Order Statistics (HOS) and complexity analysis on the signal. Singular Value Decomposition is then used to reduce the dimension of the feature vector. A one-way ANOVA test was performed on the extracted feature vector to select statistically significant singular values $\left(p\,value < 0.001\right)$. The selected singular values are used to train the Support vector machine (SVM) based classifier. Here SVM is trained as a patient centric epilepsy classifier as the nature of epilepsy differs between patients. The classification performance of the proposed system is evaluated based on *K*-fold cross validation technique which showed noteworthy results.

Keyword-EEG signal, Poly Spectra, Higher Order Statistics, Singular Value Decomposition, Epilepsy, Complexity Analysis, ANOVA test

I. INTRODUCTION

Epilepsy is a serious brain disorder that is affecting nearly 50 million people of the total world population [1]. Recent studies on the prevalence of the epilepsy disease show a steady increase in developing nations like India and China [2]. Even with the advancement of medical technology the exact cause and nature of Epilepsy is not clearly known. However International League against Epilepsy (ILAE) has classified different types of this disease. Two main forms are the partial epileptic seizures and generalized seizures. A primary tool in studying epilepsy is Electroencephalogram (EEG). EEG machine records the brainwaves by placing electrodes on the human head region based on the golden standard of 10-20 electrode placement system.

EEG signal exhibits high non-stationarity and non-linearity which is difficult to analyse using conventional signal processing techniques [3]. Current research work is directed towards the processing of single channel information on a limited domain. As EEG signal is inherently multichannel the cross information available across various lead locations are not used. Prominent EEG analysis methods include time domain [4], frequency domain [5], time-frequency [6] and non-linear [7] techniques. These methods extract features related to Epilepsy disease and classify them into normal or ictal states which are commonly referred to as two-class problem. To efficiently analyse the available data a multichannel data processing system that extracts information from different signal sources is proposed here.

II. METHODOLOGY

A. Dataset

The proposed system of Epilepsy detection involves obtaining a high dimensional feature vector of form (MxN) that captures subtle characteristics of epileptic seizures. Here 'M' belongs to multichannel signal sources and 'N' belongs to the extracted features for each channel. Fig. 1 gives the schematic representation of the proposed method of automatic epilepsy detection. The patient data for this work is obtained from Physionet CHB database [8]. The patient records are downloaded in European Data Format (*.edf) for a record length of 1 hour. The data records are then divided into time epochs of 60 sec for further processing.

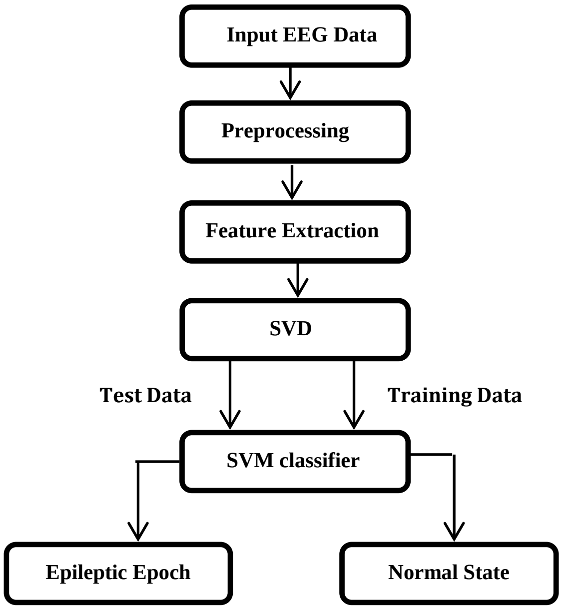

Fig. 1. Proposed system of Epilepsy Detection

A collection of 204 data sets from 3 epileptic patients are used for the evaluation of epilepsy detection system. TABLE I gives the specifications of the EEG recording system used for acquiring the brain signals from the patient.

TABLE I
EEG Recorder Specifications

Name	Value
Sampling Rate	256 Hz
ADC Resolution	11 bits
Gain	2.559375 adu/uV
File Type	European Data Format (*.edf)
Record Length	1 hour

B. Selection of Leads

Typical multichannel recording of EEG signal consist of 32 channels that are acquired by placing metal electrodes on the scalp region. The position of the electrodes is indicated in standard 10-20 electrode placement system. For this work 8 channels are selected from Frontal Pole to Occipital region covering both the hemispheres of the cerebral cortex. During the onset of epileptic seizures the cerebral functions of the human brain show less randomness. This is captured by the EEG electrodes and transmitted to the detection system for processing. Fig. 2 gives the 10-20 EEG lead placement system that is used for data acquisition.

C. Pre-processing

The pre-processing stage separates the input EEG data into different multichannel signals. The 8 signals corresponding to FP1-F3, F3-C3, C3-P3, P3-O1, FP2-F4, F4-C4, C4-P4 and P4-O2 are used for detection of epilepsy in this work. The selected lead locations are highlighted in the Fig. 2. The channels are sampled at 256 Hz and are noise free. The data sets used for training and testing the classifier have same time duration of 60 sec. Fig. 3 shows the EEG signal during the normal state with higher degree of randomness as many neurons are active due to normal cognitive process. On the contrary Fig. 4 shows the EEG signal of a patient during the onset of epileptic seizure. A close examination of the brain signal can reveal the reduction in randomness of the EEG signal as the number of neurons involving brain functions are reduced during epileptic seizures [9]. This characteristic nature of Epilepsy is used as an important classification rule in the proposed system.

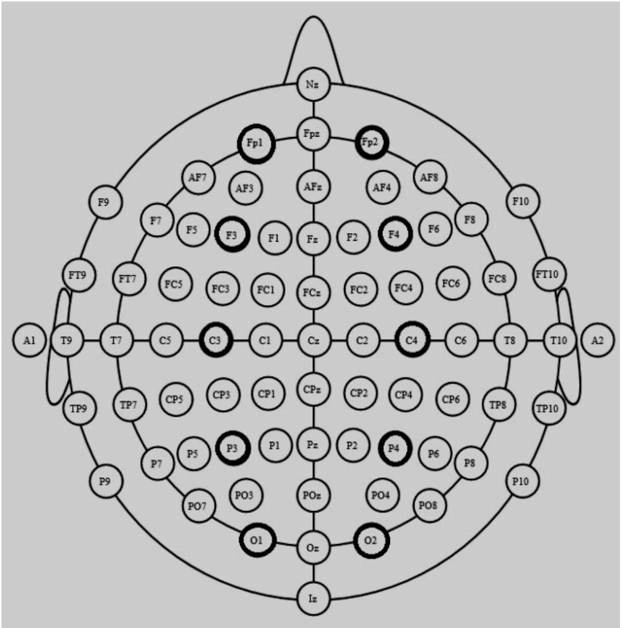

Fig. 2 EEG 10-20 Electrode Placement System

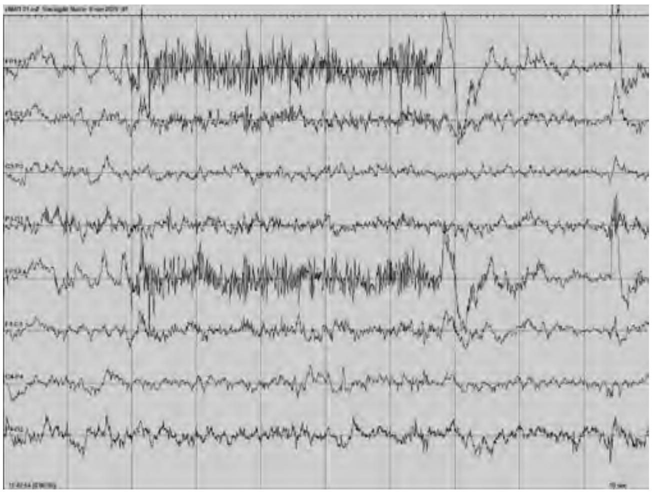

Fig. 3 EEG Signal during Normal State

Fig. 4 EEG Signal During Epileptic Seizure

D. Feature Extraction

The feature extraction is an important step in the design of epilepsy detection system. In this paper a multidimensional feature vector for the 8 channels of input given to this stage is considered. A combination of features that includes HOS and complexity dynamics are used. The reason for this approach is that non-linear parameters can capture the subtle activities of EEG during seizures [3]. The following sections explains the different features used in the work.

1) HOS features: Higher order spectra or Poly spectra are characterized by higher order moments and cumulants. Third and fourth order cumulants are used in addition to Skewness and Kurtosis in the proposed feature vector scheme. The equations of the poly spectra for the input signal s*(n)* are given below [10]:

The 2 order Cumulant (C_{2s}) function is also termed as Autocorrelation function defined as,

$$C_{2s}(k) = E\{s(n+k)s(n)\} \tag{1}$$

The 3 order Cumulant (C_{3s}) function is defined as follows:

$$C_{3s}(k,l) = E\{s(n)s(n+k)s(n+l)\} \tag{2}$$

The 4 order Cumulant (C_{4s}) function is given below:

$$C_{4s}(k,l,m) = E\{s(n)s(n+k)s(n+l)s(n+m)\} - C_{2s}(k)\,C_{2s}(l-m) - C_{2s}(l)\,C_{2s}(k-m) - C_{2s}(m)\,C_{2s}(k-l) \tag{3}$$

Skewness (*Skew*) is a derived parameter which is given as,

$$Skew = C_{3s}(0,0)/\sigma^4 \tag{4}$$

Kurtosis (*Kur*) is another higher order statistic measure which is given as,

$$Kur = C_{4s}(0,0,0)/\sigma^4 \tag{5}$$

Here σ is the standard deviation.

2) Detrended Fluctuation Analysis: Complexity measurement is necessary because it can detect the randomness and irregularity in the EEG signal. Studies shows that there is reduction in randomness of the physiological process during seizures as only few neurons are active performing cerebral functions. This can be efficiently detected by the complexity parameters. Firstly Detrended Fluctuation Analysis (DFA) for identifying the loss of randomness in the epileptic epochs is considered [11]. In this method the scaling exponent 'α' is determined by finding the slope corresponding to the line relating *log F(n)* to *log n* according to the formulae given below:

$$F(n) = \sqrt{\frac{1}{N}\sum_{1}^{N}\left[s(k) - s_n(k)\right]^2} \tag{6}$$

Where s(k) is the integrated time series and $s_n(k)$ is the local trend in the time series.

3) Approximate Entropy (ApEn): This is another complexity measurement method which is used in this work. This quantifies the amount of irregularity present in the EEG signal. ApEn for normal brain functions shows a higher value due the simultaneous firing of neuronal endings during cognitive process [12]. But during epileptic seizures the irregularity reduces and hence the ApEn value decreases. Approximate entropy of the time series s(n) is given by:

$$ApEn\,(m,r,N) = \Phi^m(r) - \Phi^{m+1}(r) \tag{7}$$

$$\Phi^m(r) = \frac{1}{N-m+1}\sum_{i=1}^{N-m+1}\log C_i^m(r) \tag{8}$$

$$C_i^m = \frac{2}{N_m(N_m-1)}\sum\sum\Theta(r - \|s_i - s_j\|) \tag{9}$$

4) Brain Wave Power: Power of the EEG signal has vital information about the nature of the physiological process. Normally there a four types of brain waves usually seen. These are Alpha waves in the frequency range of 8-12 Hz, Beta waves in the range of 13-30 Hz, Theta waves in the range of 4-8 Hz and Delta wave which are less than 4 Hz. During active thought process humans emit high frequency brain waves that are normal. The slow brain waves occur during deep sleep and unconscious state. In this work the power of brain waves in EEG signal is estimated. Higher value of power indicates abnormality because of excessive electrical discharges that takes place in different parts of the brain during epileptic seizures. Non-parametric based Welch method [5] of periodogram averaging is used for estimating the power present in frequency range of 0.5 to 50 Hz. The equations related to Welch method is given as:

$$P_w(e^{j\omega}) = \frac{1}{KLU}\sum_{i=0}^{K-1}\left|\sum_{n=0}^{L-1} w(n)s(n+iD)e^{-jn\omega}\right|^2 \tag{10}$$

where,
w(n) is the general data window of the modified periodogram with 'D' specifying the amount of overlapping.

$$U = \frac{1}{L}\sum_{n=0}^{L-1}\left|w(n)\right|^2 \qquad (11)$$

E. Dimensionality Reduction: The multidimensional feature vector produced from the feature extraction stage has to be ordered so that it contains a unique set of uncorrelated variables which can be used for epilepsy classification. Singular Value Decomposition (SVD) is a valuable tool in analysing high dimensional data. Here the feature vector $\Gamma_{M\times N}$ representing the HOS and complexity measures is decomposed into set of three matrices as given below:

$$\Gamma = U_{M\times M}\, S_{M\times N}\, V_{N\times N}^{T} \qquad (12)$$

where U and V are orthogonal matrices and S is a diagonal matrix consisting of singular values (ρ_{ii}). The reduced feature vector is formed from the singular values (SVs) and is used to characterize the epileptic state of the patient.

$$\Gamma_r = \{\rho_{11}, \rho_{22}, \ldots\ldots, \rho_{MM}\} \qquad (13)$$

The Fig. 5 shows the difference in SVs for the normal and epileptic states of a sick patient.

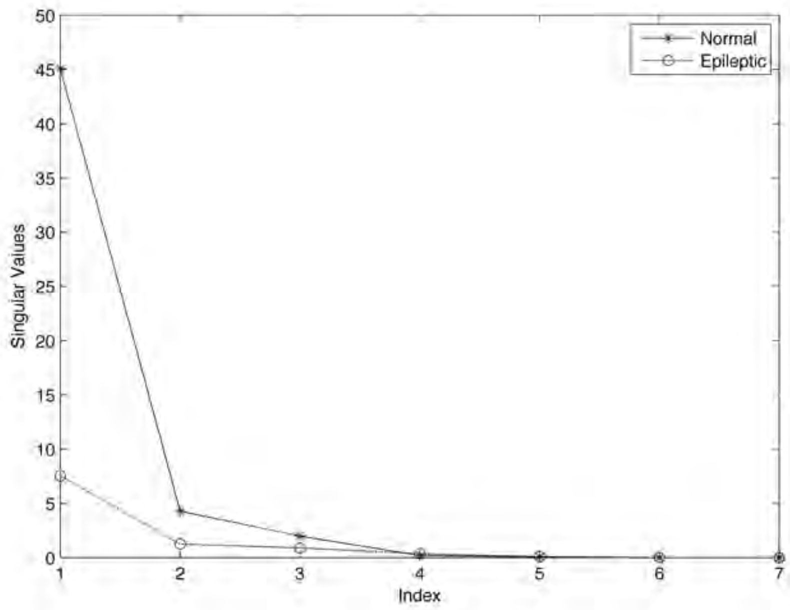

Fig. 5. Comparison of SVs of Normal and Epileptic States

F. Epilepsy State Classification

The reduced feature vector Γ_r comprising the SVs is taken as input to the classifier. At this point a one way ANOVA test is performed on the reduced feature set so that the significance level of the singular values is evaluated. Fig. 6 shows the box plot of the SVs. The First two significant SVs $(p\,value < 0.001)$ are selected to train the Support Vector Machine (SVM) which is used as a patient specific classifier. As SVM is supervised classification method it consists of training and test phases. Firstly the data set for training the SVM is obtained by accessing the annotated file of the records that are marked as epileptic seizures by epileptogenesists. The seizure time epochs are extracted and are used for SVM classification.

G. Performance Measurement

The efficiency of the proposed epilepsy detection system is expressed in terms of following parameters:

1) TP (True Positive): Number of trials a time epoch that is annotated as ictal state by human expert and the same is classified as epileptic by the proposed system.

2) TN (True Negative): Trials that are clinically marked as normal and the same are classified as normal state by the detection system.

3) FP (False Positive): Trials that are clinically marked as normal but classified as epileptic by the proposed system.

4) FN (False Negative): Time epochs that are marked as abnormal by human experts but are classified as normal state by the suggested method.

$$Test\ Accuracy = \frac{[TP+TN]}{N} \times 100\% \tag{14}$$

$$Test\ Sensitivity = \frac{TP}{[TP+FN]} \times 100\% \tag{15}$$

$$Test\ Specificity = \frac{TN}{[TN+FP]} \times 100\% \tag{16}$$

where, N is the number of trials.

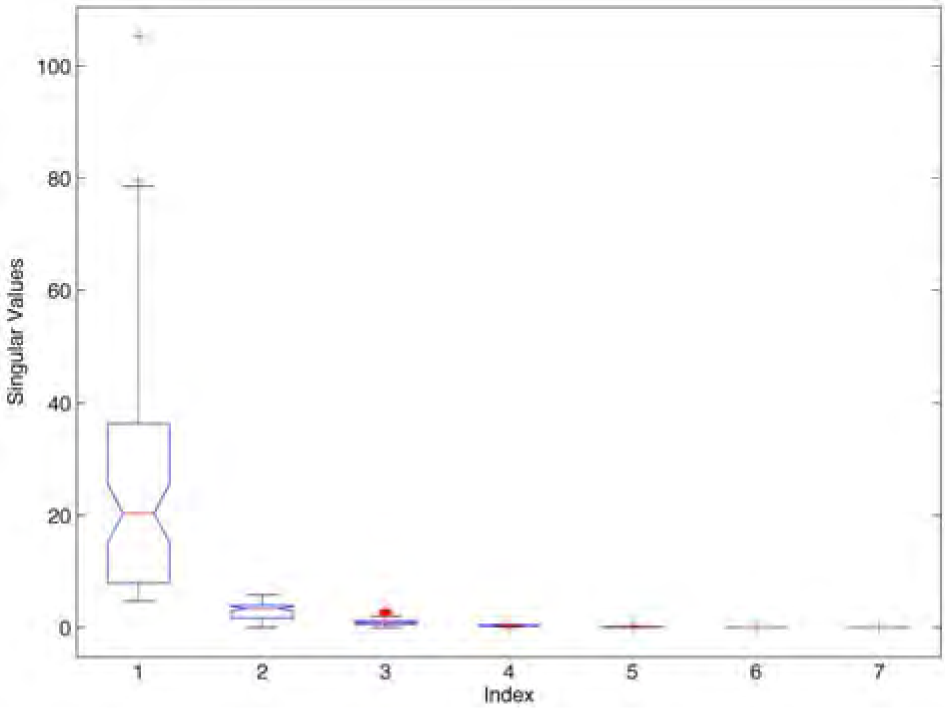

Fig. 6. Box plot representation of SVs

III. RESULTS

Epilepsy brain disorder is a phenomenon that differs between patients. . In this work a schema of patient specific classifier that can identify normal and ictal states pertaining to the subject of interest is tried. Much research work was done in this context and has reported significant results but based on certain bench mark datasets [13]. Here a publicly available CHB dataset is used to validate the proposed system. To evaluate the diagnostic performance measures of the proposed epilepsy detection system a *K*-fold cross validation is considered [14]. In this method the data set is divided into (*K=10)* groups. During each trial 90% of the data is selected in random and used for training the classifier and remaining 10% is used for testing the classifier. The EEG dataset is tested based on 10-fold scheme and the results are tabulated in TABLE II.

TABLE II
Classification performance of the proposed Epilepsy Detection System

Patient Record	Number of Dataset	Test Accuracy (%)	Error Rate (%)	Test Sensitivity (%)	Test Specificity (%)
CHB-01 Female-11years	74	79.73	20.27	100	77.61
CHB-02 Male-11years	63	98.41	1.59	66.67	100
CHB-03 Female-14years	67	100	0	100	100

The hyper plane of the trained SVM is shown in Fig. 7 . The confusion matrix of the classifier for the three datasets is given below:

1) CHB-01

$$\begin{bmatrix} 7 & 15 \\ 0 & 52 \end{bmatrix}$$

2) CHB-02

$$\begin{bmatrix} 2 & 0 \\ 1 & 60 \end{bmatrix}$$

3) CHB-03

$$\begin{bmatrix} 7 & 0 \\ 0 & 60 \end{bmatrix}$$

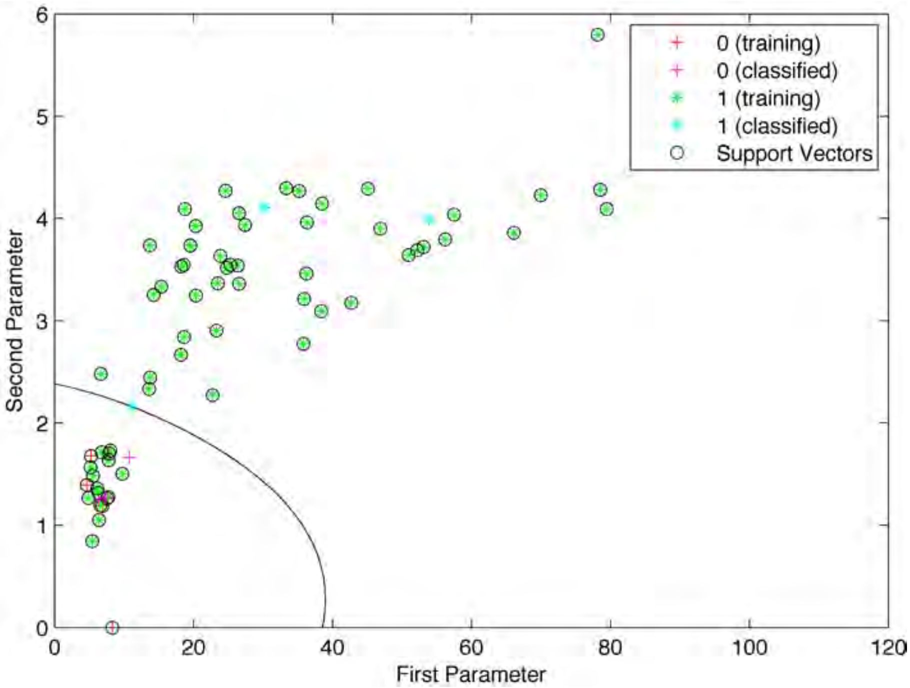

Fig. 7. SVM Hyperplane for the Two Class Epilepsy Detection System

IV. CONCLUSION

In this work a novel method for detection of epilepsy in multichannel EEG recordings is reported with notable results. A patient specific classifier for the two class problem of detecting normal and ictal states is tried in this paper. Fig. 5 shows the value of SVs of epileptic state smaller than normal state. This is in agreement to the results reported in recent studies that states that there is reduction in randomness of EEG signal due to the fact that only less neurons are active in cerebral functions [7, 9]. The complementary information available across EEG electrodes is used for analysis which is in contrast to traditional methods of processing only a single channel data [3, 10]. MATLAB software (HOSA Toolbox) is used for performing various processing of the input datasets [15]. Further investigations are required for evaluating the efficiency of the proposed system in using all the available channels for epilepsy detection. A promising direction in this field would be examining EEG signal for detection of normal, inter-ictal and epileptic states (Three class problems) which is left for further study.

REFERENCES

[1] WHO Media Center (2012) Epilepsy. [Online]. Available: http://www.who.int/mediacentre/factsheets/fs999/en/

[2] Tu Luong Mac, Duc-Si Tran, Fabrice Quet, Peter Odermatt, Pierre-Marie Preux, C T, "Epidemiology, aetiology, and clinical management of epilepsy in Asia: a systematic review," Lancet Neurol., vol. 6, pp. 533–543, 2007.

[3] Rajendra Acharya U, Vinitha Sree S, Swapna G, Roshan Joy Martis, Jasjit Suri, "Automated EEG analysis of epilepsy: A Review," Knowledge-Based Systems., vol. 45, pp. 147-165, 2013.

[4] Altunay S, Telatar Z, Erogul O, "Epileptic EEG detection using the linear prediction error energy," Expert Syst Appl., vol. 37 (8), pp. 5661-5665, 2010.

[5] Welch P D, "The use of fast fourier transform for the estimation of power spectra: a method based on time averaging over short, modified periodograms," IEEE Trans Audio Electroacoust., vol. AU-15, pp. 70-73, 1967.

[6] Subasi A, "EEG signal classification using wavelet feature extraction and a mixture of expert model," Expert Syst Appl., vol. 32 (4), pp**.** 1084-1093, 2007.

[7] Acharya U R, Chua K C, Lim T C, Dorithy, Suri J S, "Automatic identification of epileptic EEG signals using nonlinear parameters," J Mech Med Biol., vol. 9(4), pp.539-553, 2009.

[8] Ali Shoeb,"Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment." Ph D Thesis, Massachusetts Institute of Technology, September 2009 .

[9] Kannathal N, Puthusserypady S.K., Min L.C, " Complex dynamics of epileptic EEG," IEMBS '04. Annual International Conference of the IEEE, vol. 1, pp. 604-607, Sept. 2004.

[10] Chua K C, Chandran V, Acharya C M, Lim, "Automatic identification of epileptic EEG signals using higher order Spectra", J Eng Med., vol. 223 (4), pp. 485-495, 2009.

[11] Peng C K, Shlomo Havlin, Eugene Stanley, Ary Goldberger, "Quantification of scaling exponents and crossover Phenomena in nonstationary heartbeat time series," American Institute of Physics, Chaos., vol. 5(1), pp. 82-87, 1995.

[12] Pincus S M, "Approximate entropy as a measure of system complexity," Proc Nat Acad Sci (USA) , pp. 2297-2301, 1991.

[13] Andrzejak R G, Widman G, Lehnertz K, Rieke C, David P, Elger C E, "The epileptic process as nonlinear deterministic dynamics in a stochastic environment: an evaluation on mesial temporal lobe epilepsy," Epilepsy Res.,vol. 44 (2), pp.129-140, 2001.

[14] Kadir Tufan, "Noninvasive diagnosis of atherosclerosis by using empirical mode decomposition, singular spectral analysis, and support vector machines," Biomedical Research, vol. 24 (3), pp. 303-313, 2013.

[15] Swami A, Mendel C M and Nikias C L, "Higher-order spectral analysis (HOSA) Toolbox", Version 2.0.3, 2000.