

Epilepsy Prediction using Entropies

Ashwini Holla V R,
Canara Engineering College
Benjanapadavu, India.

Akshatha Kamath
Canara Engineering College
Benjanapadavu, India.

Sandeep Prabhu
Canara Engineering College
Benjanapadavu, India.

ABSTRACT

A person suffering from Epilepsy experiences or exhibits spontaneous seizures during which his behavior and perceptions are altered. Prediction of seizure onsets would help the affected and the bystanders to take prudent measures. Nonlinear features of Electro EncephaloGram (EEG) are used to isolate a class of background epileptic EEG, by training Support Vector Machine (SVM) classifier. Very good accuracy results have been seen in the results.

General Terms

Support Vector Machine, Entropies, Pattern Recognition, Epilepsy prediction

Keywords

Electro EncephaloGram (EEG), Support Vector Machine (SVM), Wavelets, non linear features.

1. INTRODUCTION

The most disabling aspect of epilepsy is the unpredictability of seizures. A scheme for predicting seizures would help the involved people from potentially embarrassing, risky situations. Characteristic features from the continuous EEG, predictive of impending seizures need to be taken and an interictal or background epileptic state or the aura phase indicative of seizures needs to be segregated from the lot of epileptiform and normal EEG, with better accuracies. The study proposes a scheme of using an SVM classifier with features taken from entropies estimated from wavelet coefficients of the data, and isolate the interictal state or the aura phase, so as to automate the decision process and objectively judge the patient condition.

1.1 Epileptiform EEG

In general, the EEG signals are the projection of neural activities that are attenuated by *leptomeninges*, cerebrospinal fluid, *dura mater*, bone, *galea*, and the scalp. Amplitudes of 0.5-1.5 mV and up to several millivolts for spikes are typical. However, on the scalp the amplitudes commonly lie within 10-100 μ V [22]. Ictal wave patterns appear with the onset of seizures. Pre-ictal or interictal phase occurs before it and is often called the aura. Identifying this phase correctly indicates the likelihood of a seizure. Given in Fig.1 are plots of a typical Normal, Pre-Ictal and Ictal EEG for a duration of 23.6

seconds.

1.2 EEG classification

The major methods followed for EEG classification are: **Spectral analysis** Peaks in the power spectrum identify epileptic seizures. Power spectra may be plotted Nonparametrically from the estimate of the autocorrelation of a given data set or Parametrically using a model [2], [1]. **Local variance:** Variance of the signal is calculated in consecutive non overlapping windows [6], the variance in each segment decides whether the segment under consideration is epileptiform or not.

Transform domain Analysis: It can be by any transform but wavelets have been shown to provide good results reported in literature, Subasi et al [8] decomposed the EEG into 5 levels of Daubechies 4 wavelet filter and classified using Neural Networks, Indiradevi et al [7], used variances and standard deviation of Daubechies 4 Wavelet detail coefficients. **Nonlinear measures:** Nonlinear measures like dimensions, Lyapunov exponents, entropies, when used with EEG, help to understand the EEG dynamics and underlying chaos in the brain[9]. In [10] Andrzejak et al. used correlation dimension to characterize the interictal EEG for seizure prediction. These methods are said to be far more superior to the traditional linear methods such as the Fourier transforms and power spectral analysis. Seizure detection performance of various entropy measures tested by Kannathal et. al [12] and entropy values computed for the epileptic EEG were found to be lower compared to the values computed for the normal EEG. It has also been reported that EEG data during seizure activity has significant non linearity but is seen to be more predictable than the EEG data during seizure free intervals which resembles a Gaussian linear stochastic process, as expressed by Hasan Ocak in [13].

2. PROPOSED METHOD

The data is decomposed using a suitable DWT, after which non linear features are extracted from it and used to train the classifier. The EEG data used were courtesy of Dr. Ralph Andrzejak's recordings at the Epilepsy Center at the University of Bonn, Germany, made available at [14] site with details at [10]

2.1 Discrete Wavelet transform

The multi-scale feature of the wavelet transform allows the

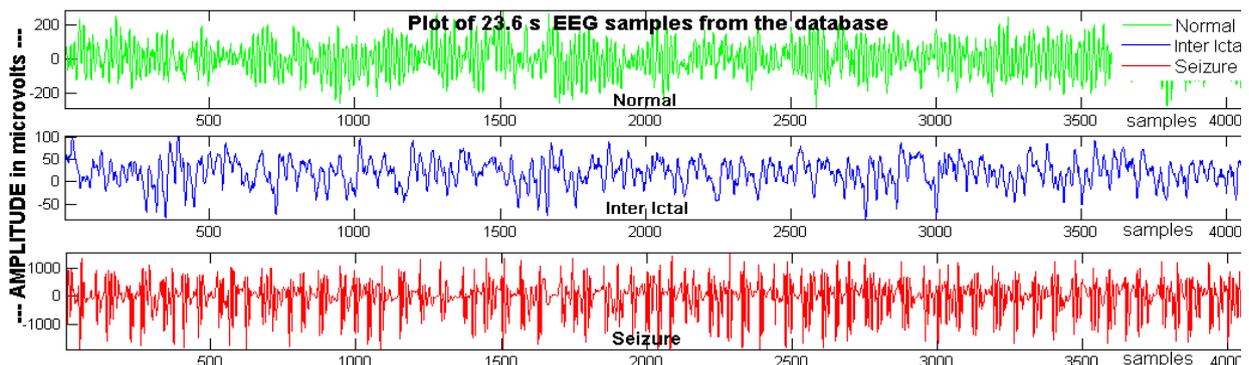


Fig 1: Typical EEG plots of Normal Inter Ictal and Ictal signals in Micro volt scale plotted for 23.6 s duration

decomposition of a signal into a number of scales, each scale representing a particular coarseness of the signal under study. The continuous wavelet transform, $W_\psi f(b, a)$, of a signal, $f(t)$, requires that the analyzing wavelet, be convolved with the signal as given as

$$W_\psi f(b, a) = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) dt \quad (1)$$

$\psi(t)$ is the mother wavelet basis function, a the scale coefficient, b the shift coefficient and $a, b \in \mathcal{R}, a \neq 0$. For large values of a , the basis function becomes a stretched version of the prototype wavelet, and for small values of a , the basis function becomes a contracted wavelet, hence high frequency components can be analyzed with a sharper time resolution than low frequency components. Using wavelets one can discriminate two signals having same frequency components occurring at different times. The baseline here being that the wavelet transform consists of translates and dilates of a basis function - the mother wavelet, at scale a and translation .

When the scale and translation variables are sampled on a dyadic grid, it results in DWT, whose wavelet equation is given by [5]:

$$(\mathbf{T}f)_{n,k} = \langle f, \psi_{n,k} \rangle = \int_{\mathcal{R}} f(x) \overline{\psi_{n,k}(x)} dx \quad (2)$$

$$\text{Where } \psi_{n,k}(x) = a^{-\frac{n}{2}} (a^{-n}x - kb) \quad (3)$$

ψ - the mother wavelet satisfies $\int_{\mathcal{R}} \psi(x) dx = 0$

For DWT we choose $n = 1; a = 2$. Among the numerous wavelets available, 7 families used for the study were Coiflets, Symlets, Biorthogonal, Reverse Biorthogonal, Discrete approximation of Meyer Wavelet, Haar, and Daubechies Wavelets.

2.2 Feature Extraction

The choice of entropies for feature extraction stems from the fact that epileptic EEG exhibit a high rate of periodicity, decreasing randomness hence, the measure of information during epilepsy[4]. The entropy Estimators used are as follows:
Approximate Entropy Estimator: $ApEn(m, r, N)$ of a series is considered given: Run length m , Tolerance window r , Number of sample points N . Briefly, $ApEn$ measures the logarithmic likelihood that runs of patterns that are close (within r) for m contiguous observations remain close (within the same tolerance width r) on subsequent incremental comparisons.[15] $ApEn$ measures the log likelihood that series of patterns close (within tolerance) for a given number of consequent observations are close on incremental comparisons that follow[15]:

$$ApEn(m, r, N) = \phi^m(r) - \phi^{m+1}(r) \quad (4)$$

$$\phi^m(r) = \frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \ln(C_r^m(i)) \text{ where } C_r^m(i) = \frac{N^m(i)}{N - m + 1}$$

Sample entropy Estimator: The $SampEn(m, r, N)$ is the negative logarithm of the conditional probability that two sequences with similar m points will remain similar at the next consecutive point. $SampEn(m, r, N)$ is the negative logarithm of the conditional probability that two template sequences with similar m points will remain similar at the next point. $SampEn$ is fairly independent of record length and displays relative consistency under circumstances where $ApEn$ does not [16],[17].

$$SampEn(m, r, N) = -\ln \left[\frac{A^m(r)}{B^m(r)} \right] \quad (5)$$

$$A^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} A_i^m(r) \text{ and } B^m(r) = \frac{1}{N-m} \sum_{i=1}^{N-m} B_i^m(r)$$

Renyi Entropy Estimator: Renyi entropy of the order α

given $\alpha \geq 0$ and $\alpha \neq 1$, for a discrete random variable $X_N = x_1, x_2, \dots, x_N$ with p_i the probability of occurrence of the event $X = x_i$, is modified from [16] and [12] as:

$$H_\alpha(X_N) = \frac{1}{1-\alpha} \sum_{i=1}^N \log(p_i^\alpha) \quad (6)$$

Renyi Entropy of 200 samples of Normal, Interictal Ictal EEG are plotted clearly indicating lower ictal values (crosses in red) as compared to the normal ones (crosses in blue).

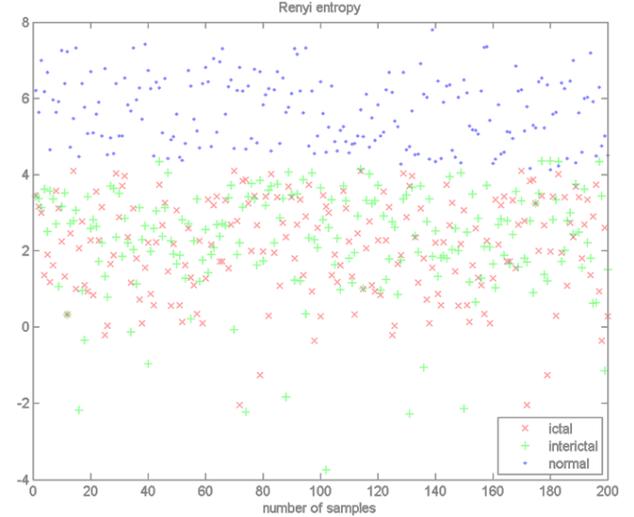


Fig. 2: Renyi Entropies of 200 samples of Normal, Interictal Ictal EEG.

Higher Order Spectrum (HOS) Entropy estimators:

These are normalised entropy estimators from polyspectra and are representations of higher-order moments or cumulants of a signal. The **bispectrum** of a signal is the Fourier transform of the third-order correlation of the signal. It can be estimated using the averaged biperiodogram given by [18],[19]

$$B(f_1, f_2) = E[X(f_1)X(f_2)X^*(f_1 + f_2)] \quad (7)$$

Where $X(f)$ is the Fourier transform of the signal $x(nT)$, $*$ denotes complex conjugation and $E[\cdot]$ denotes the expectation operation. As the Fourier transform of a real-valued signal shows conjugate symmetry, the power spectrum is redundant in the negative frequency region. Likewise the bispectrum being a product of three Fourier coefficients, exhibits symmetry and is computed in the non-redundant region Ω , as indicated in Fig.3.

Formulae for these bispectral entropy estimators taken from

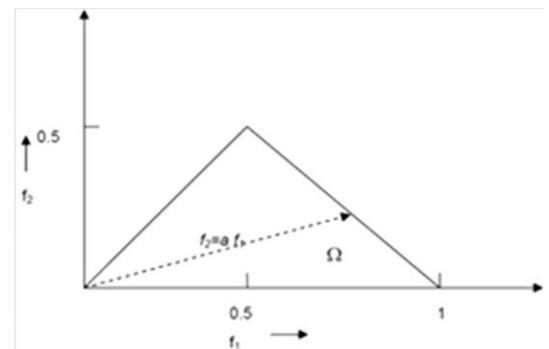


Fig. 3: Nonredundant region in PSD of the bispectrum

[19] are given: Normalized bispectral entropy 1 $P_1 = \sum_n p_i \log(p_i)$; $p_i = \frac{|B(f_1, f_2)|}{\sum_{\Omega} |B(f_1, f_2)|}$ (8)

Similarly Normalised Bispectral entropy 2, P_2 is given by:

$$P_2 = \sum_n p_n \log(p_n); p_n = \frac{|B(f_1, f_2)|^2}{\sum_{\Omega} |B(f_1, f_2)|^2} \quad (9)$$

The normalization in the equations above ensures that entropy is calculated for a parameter that lies between 0 and 1, as required of a probability, hence the Entropy estimators P and P computed are also between 0 and 1. HOS analysis helps detect non-linearity and phase relationships between harmonic components and characterises regularity of physiological signals much better than its peers [18]

2.3 Classifier

A classifier is a routine that takes in a set of data with assigned labels called classes and tries to group another new similar data set under the given classes based on some decision rule that considers some key features of the data sets. Since their inception, SVMs have continuously been outperforming many prior learning algorithms in both classification, and regression applications requiring relatively small to medium sized databases. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. The method is named after the support vectors, which are vectors lying closest to the decision boundary separating the classes. They have the greatest impact on the location of the decision boundary. If they are to be removed the decision boundary location could vary by a large range. A SVM constructs a set of hyperplanes in a high or infinite-dimensional space. Best separation is possible when the distance from the hyperplane to the nearest training data point of any class maintained is largest. Larger margin implies lower generalization error of the classifier, where margin is akin to the width of the largest tube not containing samples that can be drawn around the decision boundary. If the separating hyperplane is defined as [22]

$$x \cdot w + b = 0 \quad (10)$$

where w is its normal. for linearly separable data labeled $\{x_i, y_i\}, x_i \in \mathbf{R}^N d, y_i = \{-1, 1\}, i = 1, 2, \dots, N$ the optimum boundary chosen with maximal margin criterion is found by minimizing the objective function $E = \|w\|^2$ subject to $(x_i \cdot w + b)y_i \geq 1$ for all i (11)

The solution for the optimum boundary w_0 is a linear combination of a subset of the training data, $S \in \{1 \dots N\}$ the support vectors. These support vectors define the margin edges and satisfy the equality $(x_s \cdot w_0 + b)y_s = 1$. Data may be classified by computing the sign of $x_s \cdot w_0 + b$. Generally, the data are not separable and the inequality cannot be satisfied. A *slack* variable ξ_i that represents the amount by which each point is misclassified is hence introduced. The objective function is hence reformulated as

$$E = \frac{1}{2} \|w\|^2 + C \sum_i L(\xi) \quad (12)$$

Subject to $(x_i \cdot w + b)y_i \geq 1 - \xi_i$ for all i .

While the second term on the right-hand side denotes empirical risk associated with points misclassified or lying within the margin, L is a cost function and C a hyper-parameter that trades-off the effects of minimizing the empirical risk against maximizing the margin. Though the above equations are easy to comprehend there is a dual formulation of the above equation that is easier to solve with $L\{\xi\} = \xi$ given by:

$$\alpha^* = \max \left(\sum_i \alpha_i + \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i \cdot x_j \right) \quad (13)$$

Subject to $0 \leq \alpha_i \leq C$ and $\sum_i \alpha_i y_i x_i = 0$; $\alpha = \{\alpha_1 \dots \alpha_i\}$

is the set of Lagrange multipliers of the constraints in the primal optimization problem. The solution can be solved with quadratic programming techniques. The optimum decision boundary w_0 is given by

$$w_0 = \sum_i \alpha_i y_i x_i \quad (14)$$

and is a linear combination of all vectors that have $\alpha \neq 0$.

Kernel functions define a nonlinear mapping from the input space that is the observed data to a manifold in higher dimensional feature space, defined implicitly by the kernel functions. The hyperplane is constructed in the feature space and intersects with the manifold creating a nonlinear boundary in the input space. The value of the dot products between two vectors in the input space is replaced by the value that results when the same dot product is carried out in the feature space. The dot product in the feature space is expressed by the kernels that is the dot product of two vectors in input space. Linear polynomial and radial basis function (RBF) are commonly used and listed below:

- Linear : $K(X_i, X_j) = X_i^T X_j$
- Polynomial : $K(x_i, x_j) = (x_i \cdot x_j + 1)^n$
- Radialbasis : $K(x_i, x_j) = \exp \left[-\frac{1}{2} \left(\frac{\|x_i - x_j\|}{\sigma} \right)^2 \right]$

where n is the order of the polynomial and σ is the width of the RBF.

Extending the basic binary SVM classifier to be a multiclass classifier can be done using any of the following algorithms one-against-one, one-against-all, half against half and Directed Acyclic Graph (DAG) and Bias SVM. In this experiment, we use the one against all multi-class SVM.

3. DATA

The test subjects were 5 healthy and 5 epileptic patients diagnosed with temporal lobe epilepsy. The database is as summarised in [10] and available from [14] with normal EEG samples obtained using an Internationally standardized 10 20 surface electrode placement scheme while the ictal and interictal samples have been recorded intracranially. Each of the 23.6 s duration, single channel artifact free recording was got by sampling at 173.61 Hz, and then digitized using a 12-bit Analog to Digital Converter and encoded in ASCII. The typical samples of Epileptic Normal and Interictal EEG are as plotted in Fig 3.

4. RESULTS

Initially the 52 wavelets from the following 7 families Haar, Meyer, Daubechies(2-10), Coiflets(1-5), Symlets(2-7), Biorthogonal and Reverse Biorthogonal(1.1, 1.3, 1.5, 2.2, 2.4, 2.6, 2.8, 3.1, 3.3, 3.5, 3.7, 3.9, 4.4, 5.5, and 6.8), were used to decompose the data to six levels, but after the statistical significance test, only first level of detail coefficients were selected. The data was segmented with a 5 second window to improve upon the stationarity. *ApEn* values were estimated with m value 5 and r assigned 0.2 [20],[23]. The same values were used for estimating *SampEn* while a was assigned a value 2 for Renyi's entropy.

4.1 Test Vector Generation and Cross validation

420 samples were used for training and 360 samples for testing using three fold validation. The scheme was tested on Matlab 2006a with the Statistical and Pattern recognition

Table I Mean and Standard deviations (SD) for the features taken from Biorthogonal 2.8 wavelet

	Renyi		ApEn		SampEn		P ₁		P ₂	
	Mean	±SD	Mean	±SD	Mean	±SD	Mean	±SD	Mean	±SD
Normal	7.526	1.7336	2.279961	0.107371	2.161175	0.134143	0.954724	0.013707	0.8624	0.060305
Inter Ictal	5.225	2.674	2.070426	0.284491	1.90083	0.369976	0.961116	0.0097	0.828632	0.043006
Ictal	6.209	2.476	1.940931	0.221041	1.192673	0.408307	0.921564	0.032736	0.748621	0.129912

toolbox explained in [11]. A value of 10 was taken for the Box constraint *C* and radial basis function used for kernel. From the results, it was seen that Biorthogonal 2.8 filter yielded the best results of which the Mean, standard deviations for all of the 5 features are tabulated in Table.I.

Table II performance metrics and their intuitive meanings

Precision	TP/(TP + FP)Percentage of correct +ve predictions.
Recall/Sensitivity	TP/(TP +FN) Percentage of +ve labeled instances predicted as +ve.
Specificity	TN/(TN + FP) Percentage of -ve labeled instances predicted as -ve
Accuracy	(TP + TN)/(TP + TN + FP + FN) Percentage of correct predictions

The performance metrics usually selected to assess classifiers are as in Table II.Where True positives (TP) are seizures identified by the classifier and EEG experts and False positives (FP) are seizures identified by the classifier but not experts. False negatives (FN):Seizures missed by the classifier system.True Negatives (TN): Non Seizures identified by both parties.The performance of the proposed method is as documented in TableIII

4.2 Discussion

Ideally the best basis would be the one that has the highest correlation with the data. Based on the shapes of the mother wavelet functions intuitively one would expect the discrete approximation of Meyer to be the best fit. But it was found that the Biorthogonal 2.8 gave the best results as shown in Table III. The three factors for the good results have been the use of nonlinear features and the Biorthogonality of the wavelet basis along with the use of smaller database as elaborated in [23].

Table III: Performance Metrics along with time for computation of biorthogonal 2.8 features using SVM classifier

Accuracy with non DWT features	Accuracy with DWT features	Precision	Sensitivity	Specificity	Avg CPU time in seconds
85.6	96.3	96.9	97.4	97.6	0.54

5. CONCLUSION AND FUTURE WORK

Presented was the use of Support Vector Machine classifiers to test the efficacy of nonlinear features of detail coefficients of DWT of EEG samples in epileptic seizure detection and prediction. Since the features have been taken from only the

first level of Biorthogonal 2.8 wavelet, the complexity of the system is greatly reduced coupled with the fact that the features are simple to implement, the potential use of the system in clinical application after due modifications is very promising.

6. ACKNOWLEDGMENTS

The Authors wish to acknowledge the guidance of Prof. U Rajendra Acharya, Ngee Ann Polytechnic, Singapore. The authors would also like to acknowledge the work of Prof Andrzejak, Bonn University, Neurology Department, Centre for Epilepsy in recording the data and making it available for the public.

7. REFERENCES

- [1] Faust,O., U. Rajendra Acharya, Lim C.M, Bernhard H.Spath, 2010, "Automatic Identification of Epileptic and Background EEG Signals Using Frequency Domain Parameters", vol. 20.
- [2] Faust, O., U. Rajendra Acharya, Alen, A., Lim, C. M., 2008, "Analysis of EEG signals during epileptic and alcoholic states using AR modeling techniques", Innovations and Technology in Biology and Medicine,(ITBM-RBM). 29(1):4452, 2008.
- [3] D. Puthankattil Subha, Paul K. Joseph, U. Rajendra Acharya, Lim,C.M., 2010, "EEG Signal Analysis: A Survey", J. Medical Systems, vol. 34,no. 2, pg.no. 195-212, (2010).
- [4] N. Kannathal, Lim C. M., U. Rajendra Acharya, P.K. Sadasivan, 2005, "Entropies for detection of epilepsy in EEG", J.Computer Methods and Programs in Biomedicine, December 2005, Vol. 80, Issue 3, pp. 187-194.
- [5] Stéphane Mallat, 1999, "A Wavelet Tour of Signal Processing", San Diego:Academic Press,
- [6] Tarassenko L, Khan Y. U, Holt M. R. G,1998, "Identification of inter-ictal spikes in the EEG using neural network analysis," Inst. Elect.Eng.Proc.Sci. Meas. Technol, Nov.1998 vol. 145, no. 6, pp. 270278.
- [7] Indiradevi K.P, Elizabeth Elias, Sathidevi P.S, Nayak S. Dinesh, Radhakrishnan K.,2008, "A multi-level wavelet approach for automatic detection of epileptic spikes in the electroencephalogram" Computers in Biology and Medicine Vol. 38, 7, 2008, pp.805 - 816 .

- [8] Subasi A, 2007, "Application of adaptive neuro-fuzzy inference system for epileptic seizure detection using wavelet feature extraction", *Computers in Biology and Medicine*, 2007, 37(2), 227244 .
- [9] Kannathal N, U. Rajendra Acharya, Lim C.M, Sadasivan P.K, 2005, "Characterization of EEG - A comparative study ", *J.Computer Methods and Programs in Biomedicine*, Oct 2005 , Vol. 80, Issue 1, pp 17-23. Andrzejak R.G, 2001, "Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on region and brain state", *Physical Review E*, 2001 vol.64(6).
- [10] Franc, Vojt'ech and Hlav'ac, V'aclav, 2004, *Statistical Pattern Recognition Toolbox for MatlabR*, Research Report, pub: Center for Machine Perception, Czech Technical University, available at <ftp://cmp.felk.cvut.cz/pub/cmp/articles/franc/Franc-TR-2004-08.pdf> .
- [11] Kannathal N, Lim C.M, U. Rajendra Acharya, Sadasivan P.K, 2005, "Entropies for detection of epilepsy in EEG", *J.Computer Methods and Programs in Biomedicine*, Dec. 2005, Vol. 80, Issue 3, pp. 187-194.
- [12] Hasan Ocak, 2009, "Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy", *J Expert Systems with Applications* 36, 20272036, Elsevier Publications.
- [13] The data from Bonn University is available at <http://www.meb.uni-bonn.de/epileptologie/science/physik/eegdata.html>.
- [14] Pincus, S. M., 2000, "Approximate entropy as a measure of system complexity", *Proc. Natl. Acad. Sci.USA.*, vol. 88, pp. 2297-2301.
- [15] Costa M, Goldberger A.L, Peng C.K, 2005, "Multiscale entropy analysis of biological signals", *Phys Rev E* 2005,71:021906.
- [16] Richman, Joshua S, Moorman, J. Randall, 2000, "Physiological time-series analysis using approximate entropy and sample entropy", *Am J Physiol Heart Circ Physiol*, 2000, vol. 278, pp. H2039-2049.
- [17] David R. Brillinger, 1994, "Some basic aspects and uses of higher-order spectra", *Signal Processing*, vol. 36, 3, pp. 239-249.
- [18] U. Rajendra Acharya, Eric Chern-P.C, Chua K.C, Lim C.M, Toshiyo Tamura, 2010, "Analysis and Automatic Identification of Sleep Stages Using Higher Order Spectra", *Int. J. Neural Syst*, 2010, vol. 20, no. 6, pp 509-521.
- [19] Srinivasan V, Eswaran C, Sriram N, 2007, "Approximate Entropy-Based Epileptic EEG Detection Using Artificial Neural Networks", *IEEE Transactions on Information Technology In Biomedicine*, May 2007, Vol. 11, 3, pp.288-295.
- [20] Misra, U.K., Kalitha J, 2005, *Clinical Electroencephalography*, Elsevier India.
- [21] Hastie T, Tibshirani R, Friedman J.H, 2009, *The elements of statistical learning : data mining, inference, and prediction*, Springer.
- [22] Ashwini V R, Aparna Dinesh, 2012, "A Nearest Neighbor Based Approach for Classifying Epileptiform EEG Using NonLinear DWT Features", *IEEE conference on Signal Processing and Communication SPCOM 2012*, IISc.