Epileptic Seizure Detection using PCA on wavelet subbands

Poonam Sheoran
Department of Biomedical Engineering,
Deenbandhu Chhotu Ram University of Science & Technology, Murthal, Sonepat, Haryana, India.
poonam_bme@yahoo.co.in

J.S. Saini
Department of Electrical Engineering,
Deenbandhu Chhotu Ram University of Science & Technology, Murthal, Sonepat, Haryana, India.
jssain@rediffmail.com

***Abstract* — The detection and classification of epileptic seizures using the Electroencephalography (EEG) signal has been an active field of research from past few decades. EEG is a non-stationary signal that represents electrical activity along the scalp containing very useful information about normal or epileptic brain states. In this paper, principal component analysis is performed on the wavelets subbands of normal & epileptic signals using discrete wavelet transform. This method is applied to two different groups of EEG signals, i.e., (1) Healthy states (2) Epileptic states, during a seizure (ictal EEG). The features extracted from the principal components that are evaluated from the wavelet subbands, differentiate between these two states. Further, t-student statistical distribution is applied to determine the measure of distinguishing between different subjects. The method of principal component analysis on wavelet subbands can discriminate between ictal & non-ictal subjects with 99.99% p-value (eye open) and 99.96% p-value (eye closed) using the delta subband. The results presented here are much better than the results of previous researches.**

Keywords— Epileptic Seizures; Discrete Wavelet Transform (DWT); t-student statistical distribution, Probability value (P-value); Wavelet Multiscale Principal Component Analysis (WMSPCA).

I. INTRODUCTION

The brain is a very complex system and epileptic seizures are the result of transient and abnormal electrical activities of the brain [5]. As per a recent survey, more than 10 % (50 million people) of the world's population is affected by epilepsy [18]. Eighty percent of the epileptic seizure activity can be controlled or treated affectively, if properly detected and diagnosed [21]. The uncontrolled firing of neurons in the brain is the root cause of the epileptic seizures that leads to change of awareness and the most appropriate way to analyse it is to depict the seizure activity on the background of EEG data. Early and accurate seizure detection and timely treatment are essential to optimize the acute care in the critically ill [3]. The analysis of seizure activity cannot be done just by visual scanning of EEG as it is a very tedious & time consuming task. Moreover, biomedical signals like EEG differ from person to person and also from time to time in the same record. Hence, computer analysis is necessary and highly useful [1]. Earlier, Fourier transform and Fast Fourier transform were used for feature extraction but the limitation of these methods is that they are highly sensitive to noise [1]. Furthermore, the nature of EEG signals is non-stationary and to analyse them, we need time-frequency analysis technique. The Discrete Wavelet Transform (DWT) is highly appropriate in this regard, providing the time-frequency representation of the signal. It performs multi-rate filtering and scale space analysis too [1]. The DWT also helps us in decomposing the whole EEG signal into various subbands; in fact, the subband analysis can give more accurate and useful information than the whole EEG about constituent neuronal activities [1].

Various techniques have been proposed in the literature for epileptic seizure detection. The use of genetic algorithm for optimal feature selection [10], the performance of Boolean CNN with linear weight functions as a feature extractor [17], the variance-based methods [14], the use of Blind Source Separation [9], the wavelet-based nonlinear similarity index (WNSI) [6] and Singular Spectrum Analysis are some of the techniques for extracting the valuable information out of EEG time-series data for epileptic seizure detection. In addition to the above, it has reviewed in [12] that the correlation function time domain analysis, frequency domain analysis, time-frequency domain analysis, artificial neural network based analysis and fuzzy logic based analysis are also some of the methods that are used by the researchers in this domain. All the above mentioned methods are either very complex to perform or give less classification rate as compared to the method. The technique discussed here is accurate as well as easier to implement.

Usually, two steps are involved in almost all seizure detection algorithms which comprise the feature extraction first and then designing a system for the classification purpose. Features have been calculated in this paper by using wavelet multiscale PCA which involves the wavelet transform that not only represents the signal in time-frequency domain but also accurately captures & localizes the transient features of the epileptic signal on the EEG background.

In this paper, the wavelet sub bands have been obtained from both normal as well as epileptic EEG signals using Discrete Wavelet transform. The Principal Component Analysis (PCA) is then performed on the wavelet subbands. PCA is often referred to as a technique for reducing the number of variables in a dataset without loss of information. PCA is also known as a possible process for identifying new variables with greater meaning [13]. PCA operates by transforming a set of correlated variables into a new set of uncorrelated variables

978-1-4799-4236-7/14/$31.00 ©2014 IEEE

that are known as the principal components. The features extracted from the principal components are the variance vector of principal components and the eigen values of the principal components of delta sub bands for both normal as well as epileptic EEG signals. The extracted features are analyzed using t-student test to obtain p-value. The whole process can be represented diagrammatically as in Fig. 1:

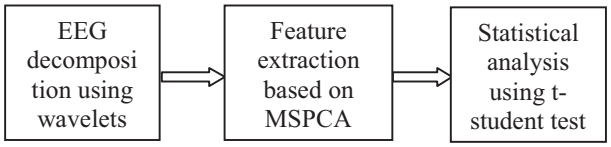


Fig.1. The block diagrammatic representation of the overall process of epileptic seizure detection.

II. Materials and Methods

A. Data selection and recording techniques

For the present work, data has been taken from University of Bonn, Germany and Dr. Andrzejak [15] has made it available publically for research use. This data comprises of three sets (denoted A, B & E) each containing 100 single channel EEG segments of 23.6-sec duration. The artifacts are taken care of while selecting the segments from continuous multichannel EEG recordings by visual inspection. The data is taken from two different groups of EEG signals namely (1) Healthy states – sets A and B consist of segments taken from surface EEG recordings, carried out on five healthy volunteers, using standardized electrode placement scheme shown in Fig.2. Volunteers were relaxed in an awake state with (A) eyes open and (B) eyes closed, respectively. (2) Epileptic states during a seizure (ictal EEG) – Set E.

Set E originated from the EEG archive of presurgical diagnosis and it contained only seizure activity. In this case, segments are selected from all recording sites exhibiting ictal activity. All EEG signals are recorded with the same 128-channel amplifier system, using an average common reference (omitting electrodes containing pathological activity (C, D, and E) or strong eye movement artifacts (A and B)) [1]. After 12 bit analog-to-digital conversion, the data is written continuously onto the disk of a data acquisition computer system at a sampling rate of 173.61 Hz. Band-pass filter settings used are 0.53–40 Hz (12 dB/oct.) [1].

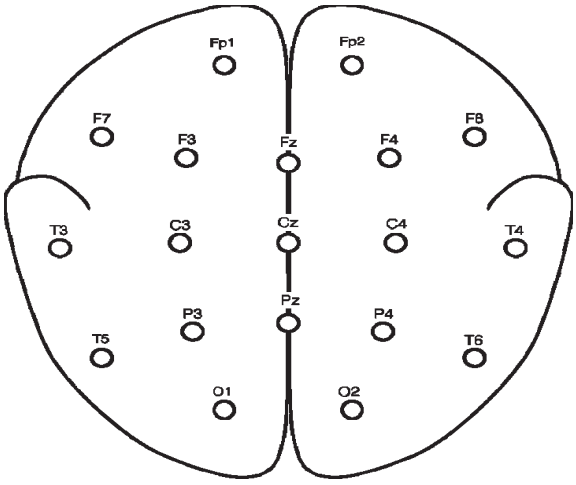


Fig. 2. The 10-20 international system for electrode placement. Segments of sets A and B were taken from all depicted electrodes [1].

B. Wavelet-Based EEG analysis

The data used for the present work has sampling frequency of 173.61 Hz. Using Nyquist sampling theorem, the useful frequency is calculated as 0-86.81 Hz [4]. DWT (comprising of different low-pass and high-pass filters) is used to extract out the signal from noise. The physiological aspects, say that of frequencies greater than 60 Hz, are considered as noise and can be neglected. The DWT that is applied to band-limited EEG (0-60 Hz) in this work is fifth-order Daubechies (DB5) discrete wavelet transform. After four levels of decomposition, the EEG sub bands obtained are delta (0-4Hz), theta (4-8Hz), alpha (8-13Hz), beta (13-30Hz), gamma (30-60Hz). In this process of decomposition, the original signal X(n) is first passed through a halfband highpass filter with impulse response g[n] and a lowpass filter with impulse response h[n]. The signal can, therefore, be subsampled by 2, simply by discarding every other sample. Mathematically, this first level of decomposition can be expressed as:

$$Y_{high}[k] = \sum_{n} x[n] \bullet g[2k - n] \quad (1)$$

$$Y_{low}[k] = \sum_{n} x[n] \bullet h[2k - n] \quad (2)$$

where yhigh[k] and ylow[k] are the outputs of the highpass and lowpass filters, respectively, after subsampling by 2 (S. G. Mallat 1989).

The DWT decomposition can be explained through the schematic diagram shown in Fig.3.

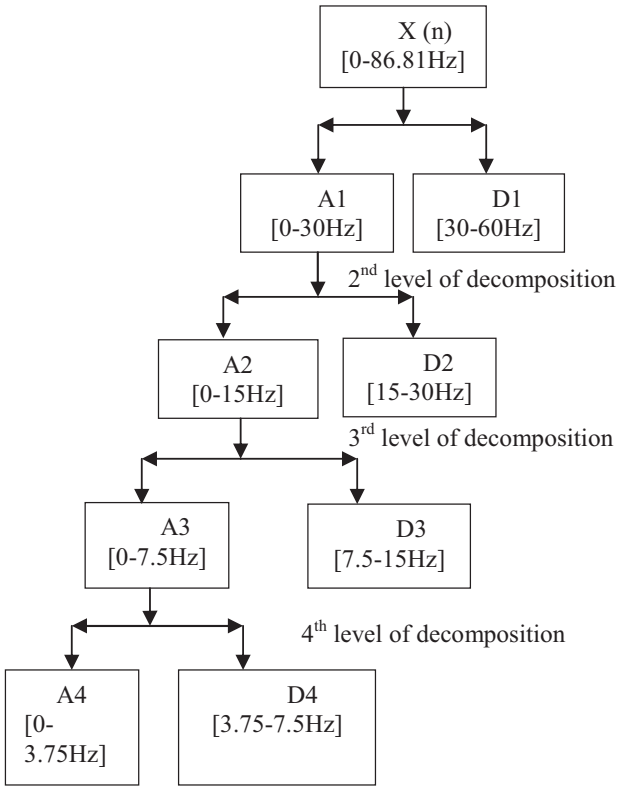


Fig. 3 The block diagrammatic representation of wavelet decomposition.

C. Multiscale Principal Component Analysis

The Principal Component Analysis is one of the oldest methods in multivariate signal analysis domain. A multiscale version of PCA has been proposed by Bakshi for considering the dimensionality reduction issues and selection of relevant factors across scales as well [2].The MSPCA scheme runs as follows:

Step-1: First of all, wavelet transform is performed at level K for each column of the input matrix X. K is the level of decomposition.

Step-2: Then, perform the PCA on the matrix Dk for $1 \leqslant k \leqslant K$, and select an appropriate number pk of useful principal components or supress the detail Dk (matrix of detailed coefficients).

Step-3: Similarly, perform the PCA on the matrix AK (approximate coefficients) and select pK+1 principal component.

Step-4: Reconstruct a new matrix X* from the simplified detail and approximation matrices, containing the main features of the original matrix, X, by inverting the wavelet transform.

Step-5: Finally perform the PCA on the matrix X* and build the requisite statistics for epileptic seizure detection.

The last step is very important for dimensionality reduction since rank(X*) = rank(X).

Steps 2 and 3 provide a compact representation of the p signals of length n in the form of wavelet coefficients [11].

D. T-Students Distribution

The statistical analysis tool used in this paper is T-student test. T-student distribution is used to measure the probability that serves as an important criterion of finding out the extent of differentiating the two samples named p-value (probability value). When we have two samples with equal or unequal sample sizes, and unequal variances, the t-student distribution value is calculated using the hypothesis that the sample means are different by the following [20]:

$$t = \frac{\overline{X_1}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} - \frac{\overline{X_2}}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} \quad (3)$$

For obtaining the p-value, the t-student degrees of freedom, v, should be calculated by:

$$v = \frac{\left(s_1^2 / n_1 + s_2^2 / n_2 \right)^2}{\left(\frac{s_1^2}{n_1} \right)^2 / (n_1 - 1) + \left(\frac{s_2^2}{n_2} \right)^2 / (n_2 - 1)} \quad (4)$$

Using v, the p-value can be calculated as follows:

$$p - value = \int_{-\infty}^{\infty} \frac{\Gamma\left(\frac{v+1}{2} \right)}{\sqrt{v\pi}\,\Gamma\frac{v}{2}} \left(\frac{1+t^2}{v} \right)^{-\frac{(v+1)}{2}} dt \quad (5)$$

where, and are the sample means-1 and -2 respectively, and & are sample sizes (taken as 100 each in this work) and & are sample variances. The p-value calculated from the t-test as shown in equation-5 is of significance as it is the measure of distinguishing between the two samples of the signals.

III. Results and Discussion

Using WMSPCA, the principal components variances matrix is calculated using MATLAB toolbox. After retaining some selective number of principal components (the selection is based on Kaiser's rule which says that only those components are to be kept that are associated with eigenvalues greater the mean of all eigenvalues), we get the A4 (delta) and D4 (theta) sub band matrices. The eigen values for both of these matrices are calculated which act as important features for classification as shown in Table-2 & Table-3 for epileptic and normal subjects (only 50) respectively. The average of the variance matrices of all the three subjects (healthy eye open & eye closed and epileptic) has been calculated and it is observed that there is a considerable difference between the ictal data and the healthy data as shown in Table-1. So, it is an important feature for statistical analysis. The next important feature is the eigen values of the matrix of A4 (delta) sub band. The scree plot of the variance explained% (percentage

of variance explained by each principal component) of data segment is shown in Fig.4 and the simplified matrices for the first three signals of the E-segment with normal distribution curve of the eigen values are shown in Fig.5 that are obtained after WMSPCA. The histogram plots of eigen values of the delta segment of ictal state and healthy states (with eyes open and eyes closed) are shown in Fig.6.

TABLE-1 AVERAGE VALUES OF PC VARIANCES OF ALL THE THREE SUBJECTS

	A	B	E
Average values of PC Variances	24.0228	84.0631	463.8536

TABLE-2 EIGEN VALUES OF NORMAL SUBJECTS

Sr. No.	Eigen values of 4^{th} Approximate coeff.	Eigen values of 4^{th} Detailed coeff.
1	0.5822	0.1691
2	-39.9215	-1.913
3	-124.9845	-5.5436
4	-82.2506	-37.8258
5	-23.3108	13.6804
6	-163.7324	24.4034
7	357.0631	-5.2187
8	-271.245	-30.9161
9	26.1414	-36.4981
10	245.0799	-63.507
11	44.7474	181.9379
12	115.9541	-28.255
13	-11.6976	-57.1819
14	-239.8727	-113.3288
15	-175.0738	-402.5934
16	130.0116	-439.4924
17	234.5124	55.1889
18	-96.7397	195.6074
19	81.7044	-110.7924
20	-212.7588	42.798
21	77.8096	155.6329
22	-81.3456	360.6212
23	-92.4766	22.4642
24	-38.7232	-31.4343
25	-280.4279	-14.2856
26	-59.8934	-44.7811
27	-201.8849	120.2396
28	-244.1691	91.2834
29	198.5379	70.8822
30	3.1007	8.966
31	-172.5585	253.9968
32	-99.9371	-38.7906
33	-4.2724	-30.5289
34	75.237	-0.6109
35	234.0941	-32.6196
36	86.2183	-5.2862
37	42.0292	-48.0975
38	227.1485	-193.8058
39	-296.4237	236.4208
40	50.2966	293.0992
41	-16.0445	162.1844
42	-16.3742	47.7459
43	-322.6753	59.1254
44	3.3182	-66.3202
45	-27.3011	33.7567
46	-230.537	92.542
47	-43.7329	-208.9803
48	192.571	-7.3667
49	-70.1471	-49.6826
50	-9.3352	140.7942

TABLE-3 EIGEN VALUES OF EPILEPTIC SUBJECTS

Sr. No.	Eigen values of 4th Approximate coeff.	Eigen values of 4th Detailed coeff.
1	0.6175	-0.0458
2	1.9014	-0.4879
3	1.549	-0.5137
4	-0.2281	0.1395
5	1.007	0.5588
6	-0.4392	-0.2081
7	0.8914	0.0503
8	-0.8322	-1.7964
9	0.0794	-0.292
10	-0.8564	0.5859
11	0.6675	0.9234
12	-0.3406	-0.9005
13	0.2531	1.8637
14	-0.1682	0.0996
15	-0.413	-0.1163
16	0.1882	-0.026
17	-0.0532	0.4451
18	0.1104	0.043
19	-1.1804	1.2175
20	0.206	0.5345
21	-0.0469	0.1225
22	-0.0528	0.0469
23	-0.5377	0.1677
24	0.2431	0.0512
25	-0.7978	-0.2729
26	0.0236	-0.6
27	-0.6353	-0.18
28	1.9021	-1.7508
29	0.3089	-0.0459
30	0.4562	-0.5312
31	-0.5617	-0.8022
32	-0.3724	-0.1524
33	-0.6789	0.1558
34	-0.2249	-0.1206
35	1.389	-1.352
36	0.0167	-0.3797
37	1.1105	0.5259
38	-0.0664	0.0299
39	-0.2984	0.0541
40	-0.1243	0.0131
41	0.5599	0.7923
42	0.3969	1.0563
43	-0.044	-0.1228

44	1.2819	-0.5272
45	-0.2879	0.0498
46	1.0075	-0.499
47	-0.1767	1.286
48	0.039	-1.0747
49	0.4578	-2.1034
50	-0.3036	-0.1

Table-4 shows the p-values for delta subband and we can easily see and distinguish between the ictal state and other states. The best distinction is obtained between ictal state (E) and the healthy state with eye open in delta subband with 99.99% p-value.

TABLE-4 PROBABILITY MEASURES OF DISTINGUISHING BETWEEN SUBBANDS

	p-value (A-E)	p-value (B-E)
A4 DELTA (0-4Hz)	99.99%	99.96%

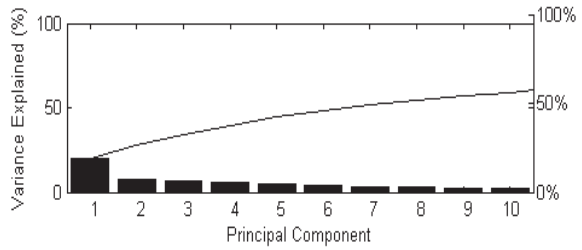
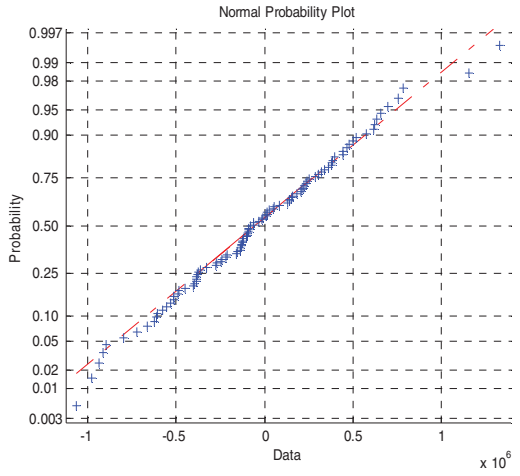
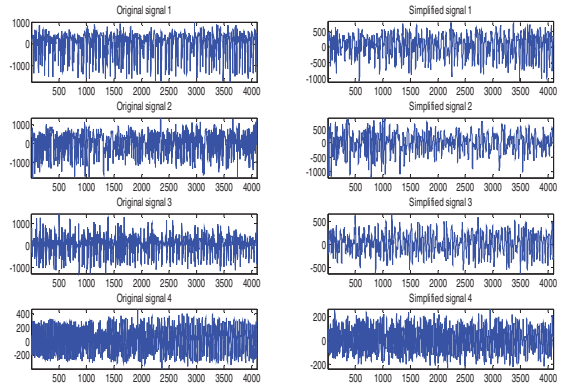


Fig. 4 Scree plot of the variance explained% with the principal component of E-segment.

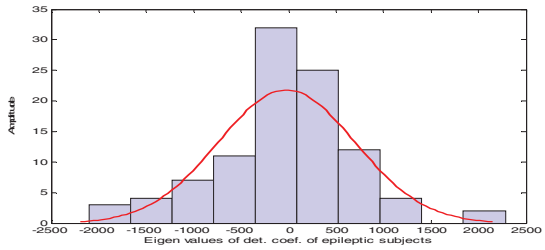


(a)

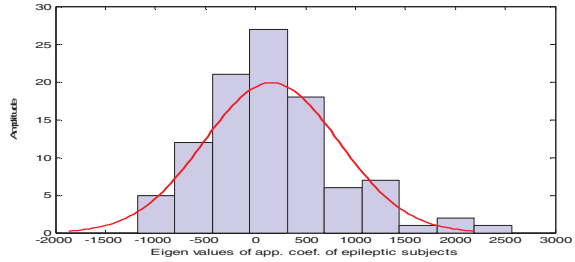


(b)

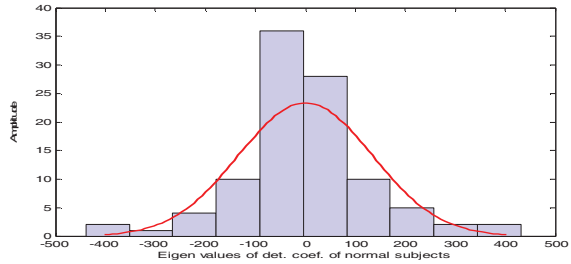
Fig. 5 (a) Normal probability distribution of the eigen values of delta subband of E-segment. (b) Three different original epileptic signals along with their simplified versions obtained after wavelet multiscale PCA.



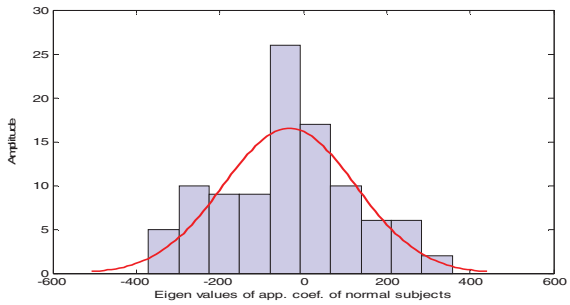
(a)



(b)



(c)



(d)

Fig. 6. Histogram plots of (a) the eigen values of the det. coef. of the epileptic subjects. (b) the eigen values of the app.coef. of the epileptic subjects. (c) the eigen values of det. coef. of the normal subjects. (d) the eigen values of app. coef. the normal subjects.

The presented method of epileptic seizure detection is quite efficient as it gives high p-values that signify the clear distinction between the seizure data and the normal data. As bourne out by our survey of related paper, this kind of application of wavelet multiscale principal component analysis for seizure detection is altogether new in this field of epileptic seizure detection and highly accurate too.

IV. Conclusion

In this paper, EEG and its subbands with their frequency ranges are discussed shown in Fig.3 and their eigen values are calculated from multiscale principal component analysis using discrete wavelet decomposition. The eigen values of delta sub-bands have shown considerable data reduction (by retaining selected number of principal components) carrying valuable information without noise. Based on statistical analysis by tstudent distribution, in A4 subband, the maximum probability of dissimilarity between ictal and non-ictal subjects is 99.99% p-value (eye open) and 99.96% p-value (eyes closed) which is a substantially valuable result.

REFERENCES

[1] Ahmad Mirzaei, Ahmad Ayatollahi, Parisa Gifani, Leili Salehi, “EEG Analysis Based on Wavelet-Spectral Entropy for Epileptic Seizures Detection”, 2010 3rd International Conference on Biomedical Engineering and Informatics.

[2] B.Bakshi, Multiscale PCA with application to MSPC monitoring, AIChE J. 44(1998), pp. 1596-1610.

[3] Cigdem l. Akman, Vesna Micic, Anita Thompson and James J. Riviello Jr., “Seizure detection using digital trend analysis: Factor affecting utility”, Epilepsy Research, Volume 93, Issue 1, January 2011, Pages 66-72.

[4] Daubechies, Ten Lectures on Wavelets. Philadelphia, PA: SIAM, 1992.

[5] H. R. Mohseni, A. Maghsoudi, and M. B. Shamsollahi, "Seizure Detection in EEG signals: A Comparison of Different Approaches", IEEE Proceedings EMBS Annual International Conference, New York City, USA, Aug. 30-Sept. 3, 2006.

[6] Hojjat Adeli, Samanwoy Ghosh-Dastidar, and Nahid Dadmehr, “A Wavelet-Chaos Methodology for Analysis of EEGs and EEG Subbands to Detect Seizure and Epilepsy”, IEEE Transactions on Biomedical Engineering, Vol. 54, No. 2, Feb. 2007

[7]. Iasemidis, L. D., Shiau, D. S., Chaovalitwongsa, W. (2003) “Adaptive epileptic seizure predication system”, IEEE Transaction on Biomedical Engineering, 50(5), 616–627.

[8] <http://www.who.int/topics/epilepsy/en/>.

[9] Javier Corsini, Leor Shoker, Saeid Sanei and Gonzalo Alarcón, “Epileptic Seizure Predictability From Scalp EEG Incorporating Constrained Blind Source Separation”, IEEE Transactions on Biomedical Engineering, Vol. 53, No. 5, May 2006.

[10] M. D'Alessandro', G. Vachtsevanos', A. Hinson' J, Echauz, B. Lie, “A Genetic Approach to selecting the optimal feature for Epileptic Seizure Prediction”, Proceedings of the 23rd Annual EMBS Int. Con., Oct. 25-28, 2001 Istanbul, Turkey.

[11] Mina Aminghafari, Nathalie Cheze, Jean-Michel Poggi, “Multivariate denoising using wavelets and principal component analysis”, Computational Statistics and Data Analysis, Vol 50, Issue 9, 1st May 2006, Pages 2381-2398.

[12] Ocak, H. (2008b)., “Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy”, Expert Systems with Applications, Volume 36, Issue 2, March, 2009.

[13] John L. Semmlow, “Biosignal and Biomedical Image Processing Matlab-based Application”, Marcel Dekker Press US, 2004.

[13] Patrick E. McSharry, Leonard A. Smith, and Lionel Tarassenko, “Comparison of Predictability of Epileptic Seizures by a Linear and a Nonlinear Method”, IEEE Transactions on Biomedical Engineering, Vol. 50, No. 5, May 2003.

[15] R. G. Andrzejak, “Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state”, Physical Review E, vol. 64, 061907, Nov. 2001.

[16]http://www.meb.unibonn.de/epileptologie/science/physik/eegdata.html

[17] R. Kunz, C. Niederhofer and R. Tetzlafe “Prediction of Epilepic Seizurs by CNN with Linear Weight Fuvctions”, Proceedings of IEEE Int. Workshop on Cellular Neural Networks and Their, Applications, pp. 129-134, Catania, ISBN 0-7803-63442, 2002.

[18] Saraceno, B., Avanzini, G., Lee, P. (2005). Atlas; Epilepsy care in the world. Geneva.

[19] S. G. Mallat, “A theory for multiresolution signal decomposition: the wavelet representation”, IEEE Trans. Pattern Anal. Mach. Intell., vol.11, no. 7, pp. 674–693, Jul. 1989.

[20] W. Mendenhall, R. J. Beaver, B. M. Beaver, “Introduction to Probability and Statistics”, 12th edition. Chapter 10, Cengage Learning Inc., pp. 404-406.

[21] http:// www.who.int/ topics/ epilepsy/ en/