A Wavelet -PCA based methodology for accurate EEG classification using Support Vector Machine and K-Nearest Neighbor

Conference Paper · February 2015

DOI: 10.13140/2.1.4841.0406

3 authors:

Tayyab Saeed
Aligarh Muslim University

Yusuf Uzzaman Khan
Aligarh Muslim University

M.F. Azeem
Aligarh Muslim University

Emerging Trends in Electrical And Electronics Engineering (ETEEE-2015)
Jamia Millia Islamia, New Delhi

A Wavelet -PCA based methodology for accurate EEG classification using Support Vector Machine and K-Nearest Neighbor

Mohd Tayyab Saeed
Dept. of Electrical Engineering,
AMU, Aligarh, India
m.tayyab.saeed@zhcet.ac.in

Yusuf Uzzaman Khan
Dept. of Electrical Engineering,
AMU, Aligarh, India
yusufkhan.ee@amu.ac.in

Mohammad Fazle Azeem
Dept. of Electrical Engineering,
AMU, Aligarh, India
mf_azeem@hotmail.com

***Abstract*—Diagnosis of seizure requires visual inspection of EEG, however for large volume of EEG it requires a lot of time and work. Thus, this work is aimed towards the automatic seizure detection with greater accuracy in less time and burden. As EEG signals are non-stationary, conventional method of frequency analysis is not highly successful in classification. This paper deals with a method of automatic analysis of EEG signals using wavelet transform, principle component analysis (PCA) and classification using support vector machine (SVM) technique. Its novelty lies in its simple statistical feature based formulae which makes the algorithm fast and accurate. This proposed technique can classify normal, interictal and ictal EEG with 100% accuracy, sensitivity & specificity, hence can be used to design expert system for seizure diagnosis purpose in various hospitals.**

***Keywords*—** *EEG; Seizure; Wavelet; Statistical Features; SVM; Machine Learning; PCA*

I. Introduction

Neuronal activity starts as early as between 17th and 23rd week of prenatal human embryo [3]. Since this stage, electrical signals are continuously generating and, is believed to be the reflection of whole body function. This induces motivation for EEG signal processing. Epilepsy is one of the most common neurological disorder of which about 50 million people worldwide are suffering and out of these 80 percent are in developing countries [2]. Epileptic people are two or three times more likely to die prematurely when compared to a normal person if not treated. Hence, study of epilepsy has always been of utmost importance in the biomedical field of research.

Seizures are known to be the result of sudden, usually brief, excessive electrical discharges in a group of brain cells (neurons). It is characterized by recurrent convulsions over a time-period [1]. Eighty percent of the epileptic seizure activity can be treated effectively, if detected and diagnosed properly. But visual inspection of EEG is very time consuming and there may be expert's disagreement over same EEG due to its subjective nature and variety of inter-ictal spikes morphology. Thus a reliable computer based expert system is needed to assist the clinicians.

II. Background

Work in field of seizure detection started in 1970s and first notable work was done by J. Gotman in 1982 [4]. After that various efforts have been made in this regard using correlation function, time domain analysis, frequency domain analysis, time-frequency domain analysis, AR modeling, and genetic programming etc. Generally, every epileptic seizure detection technique has three steps. First, step is the preprocessing of the EEG data i.e. removing artifacts and noises. Second step is to elicit inherent information that characterizes the different states of brain electrical activity using some feature extraction techniques, and subsequently, third step is to train a chosen expert system based on the obtained features. In some early works of spike detection [6, 7, 8] a number of parameters such as relative amplitude, sharpness, and duration of EEG waves were measured from the EEG signals and evaluated. These method were very sensitive to various artefacts. A multi-stage rule base system was given by Dingle *et al* [9]. It was able to detect 67% of the spikes, later both multichannel temporal and spatial information, and including the electrocardiogram and electromyogram information into a rule-based system [10, 11], a higher detection rate was achieved.

Since, EEG is non-linear and non-stationary real time signal, its information is not limited to any one particular domain. A good system need both time and frequency domain analysis. So most of the new methods are based on time-frequency analysis e.g. [12, 13, 14, 15]. Wavelet decomposition is one of the techniques for time-frequency domain analysis which is also used in this paper.

III. Methodology

Figure 1 illustrates the steps to be followed in this analysis. Initially, the pre-processed EEG is decomposed into different wavelet bands. Each band is then used to extract statistical features. The SVM classifier is then used for classifying the EEG into different classes.

A. Data Collection

Data used in this work is an open source database provided for research [16]. The complete data consists of five sets (A to E), each containing 100 channels of EEG signals. Each signal is of 23.6 s duration containing 4097 samples.

Recording was carried by 10-20 electrode placement scheme. Each of the five sets was recorded under different circumstances. Both sets A and B were recorded from healthy subjects, set A was recorded with their eyes open whereas set B with their eyes closed. On the other hand, sets C, D, E were obtained from epileptic patients. Set C and D were recorded during seizure free period, where set C was recorded from the hippocampal formation of the opposite hemisphere of the brain, whereas set D was obtained from within the epileptogenic zone. The last data set, set E, contains ictal data that were recorded when the patients were experiencing seizure [16]. The data were digitized at 173.61 samples per second using 12 bit resolution.

This database has been widely used for testing algorithms designed for epileptic seizure detection [5], [13], [17]. However, most of them are tested only on set A and E. Although, the classification was made with good accuracy but the within group variability is less explored. So, in this work we aim towards the classification of normal, inter-ictal and ictal EEG. Although this EEG record has no muscle artifacts but it may contain high frequency noises which can be suppressed using low pass filter with cut off frequency of 64 Hz. Then, a rectangle window formed by 512 discrete data is selected, so as it form a single EEG segment. Thus, each channel has 8 EEG segments. In total, now we have 800 EEG data segments in all five sets which can be used for the training and testing purpose.

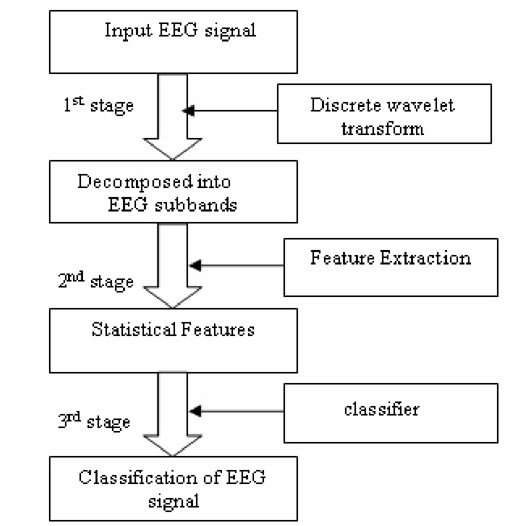


Fig 1

B. Wavelet Decomposition

In order to extract useful sub frequency signals by retaining time series information, wavelet decomposition technique is used. The EEG signal is decomposed into progressively finer details by means of multi-resolution analysis using complementary low pass (LPF) and high pass (HPF) filters (explained further in [18]). The high pass filter HPF is the discrete mother wavelet and the low pass filter LPF is its mirror version. In this work, fourth order daubechies wavelets (db4) is used as mother wavelet. Schematic representation, of wavelet decomposition used in this work is shown in Fig 2. Band limited EEG (0-64 Hz) is decomposed into 4 levels as illustrated with D1 (32-64 Hz), D2 (16-32 Hz) D3 (8-16 Hz), D4 (4-8Hz) and A4 (0-4 Hz). Decomposed signals of set A with their frequencies are shown in Fig 3.

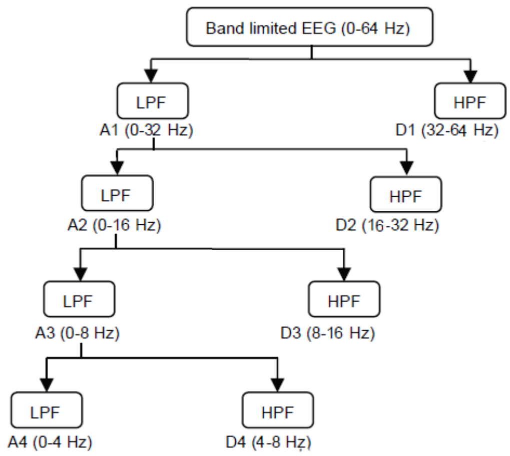
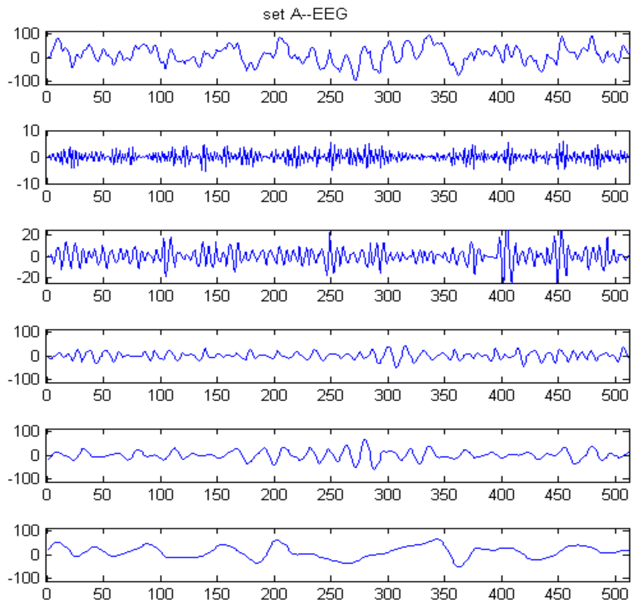


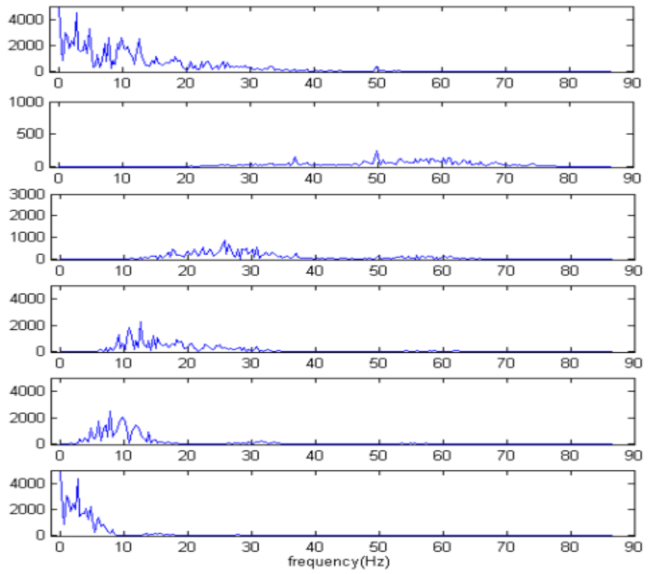
Fig 2 Wavelet Decomposition

(i)



SET	Features	D1	D2	D3	D4	A4
A	1st feature	1.0232	3.5030	7.80199	7.3948	10.8581
	2nd feature	49.2898	180.0091	415.48132	384.6801	476.9431
B	1st feature	1.28845	4.37075	12.0916	11.0823	14.8548
	2nd feature	71.6466	221.6876	569.7662	474.4267	594.9434
C	1st feature	0.6499	1.6230	4.70130	10.4370	21.010
	2nd feature	34.6897	82.512	209.014	479.9304	810.765
D	1st feature	0.77088	1.66968	4.4079	5.24653	13.5462
	2nd feature	43.0502	94.1481	217.7872	245.7715	488.8541
E	1st feature	9.51853	60.7096	146.2800	97.4307	101.0744
	2nd feature	441.2567	2977.3592	6987.3422	4673.4505	4493.2212

Table 1. Features obtained on first segment of each set



(ii)

Fig 3 (a) set A decomposed signal (b) Subband frequencies

C. *Feature Extraction*

After decomposition of each channel of windowed EEG into 5 sub-bands namely D1, D2, D3, D4, A4, features extraction is done. In this work, 2 features per band i.e. mean absolute deviation (MAD) of rectified EEG and mean absolute deviation of frequencies present, are calculated. Thus for 1 segment, 5 * 2 = 10 features are computed which results in 800 by 10 matrix for each set. Table 1 shows features extracted from first epoch.

MAD in rectified EEG has highest magnitude in ictal EEG than follows normal and inter-ictal. However in D4 and A4, inter-ictal has higher magnitude than normal. MAD of frequencies present, follow the same trend.

IV. RESULTS AND DISCUSSION

Feature matrix obtained is a ten dimensional data which is reduced to two dimensional data by using principal component analysis (PCA). Thus, we obtained a PCA reduced feature matrix of 800 by 2. Feature plots of sets B-E and sets A-C are shown in Fig 3. Similarly all other plots can be obtained. Out of 800, 500 samples are used for training SVM classifier with hyperplane kernel while remaining 300 are used for testing purpose.

Classification results are evaluated in terms of Accuracy, Sensitivity and Specificity defined as follows:

$$\text{Accuracy} = \frac{\text{Number of correct decisions}}{\text{Total number of cases}}$$

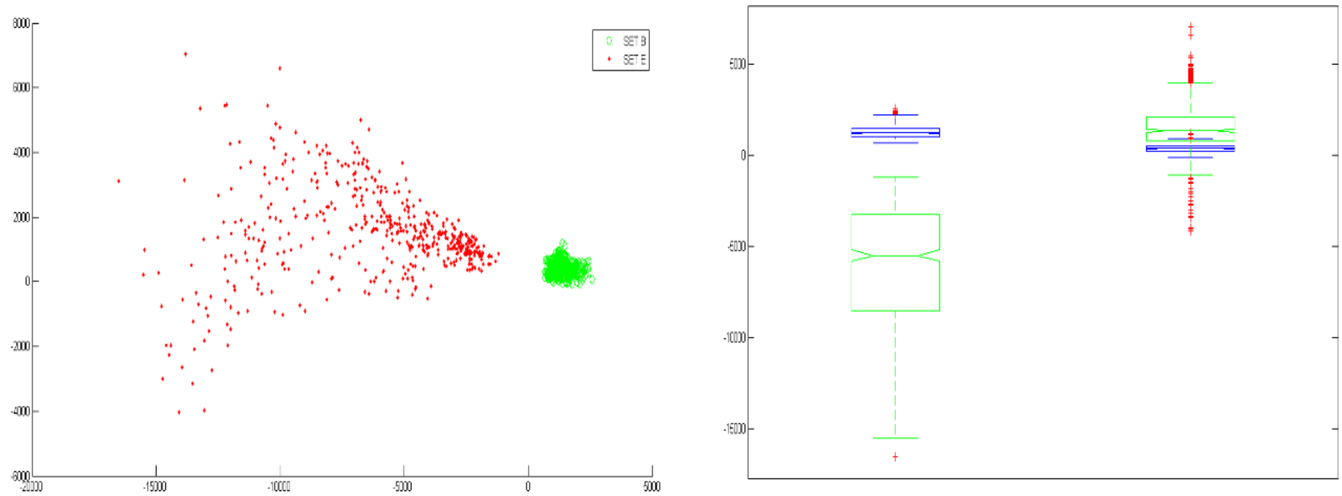
$$\text{Specificity} = \frac{\text{Number of true negative decisions}}{\text{Number of actually negative cases}}$$

$$\text{Sensitivity} = \frac{\text{Number of true positive decisions}}{\text{Number of actually positive cases}}$$

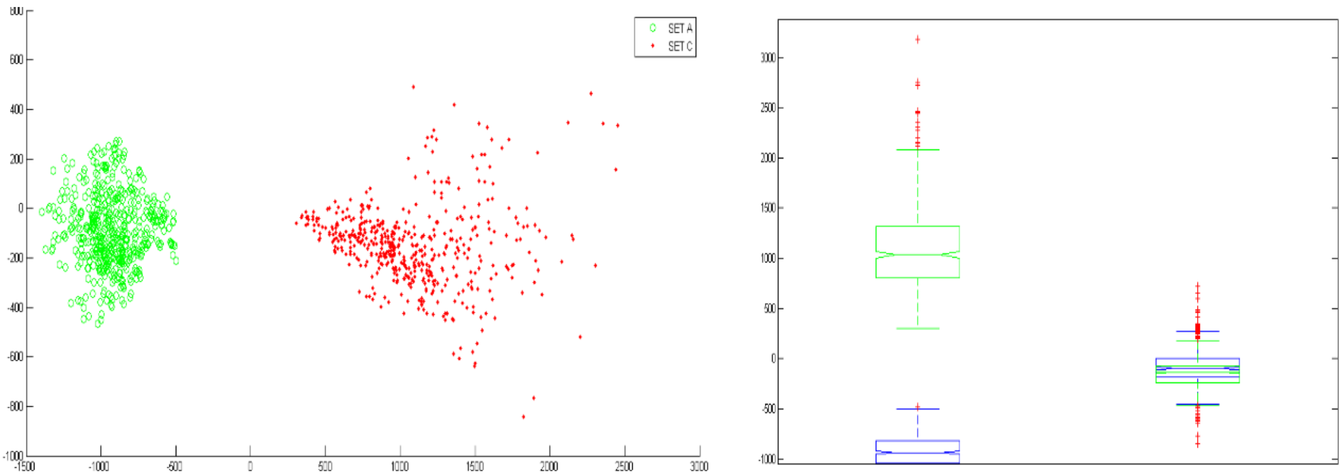
The classification accuracy which is defined as the percentage ratio of the number of segments correctly classified to the total number of segments considered for classification depends on the type of wavelet chosen for the application. Daubechies wavelet of order 4 (db4) was used and found to yield good results in classification of the EEG segments. The results obtained are summarized in Table 2.

Sets	Performance	Classes
A-E	100 % Accuracy, Sensitivity, Specificity	Normal-Ictal
B-E	100 % Accuracy, Sensitivity, Specificity	
C-E	100 % Accuracy, Sensitivity, Specificity	Interictal-Ictal
D-E	100 % Accuracy, Sensitivity, Specificity	
A-C	100 % Accuracy, Sensitivity, Specificity	Normal-Inter-ictal
A-D	100 % Accuracy, Sensitivity, Specificity	
B-C	100 % Accuracy, Sensitivity, Specificity	
B-D	100 % Accuracy, Sensitivity, Specificity	

Table 2



A. Set B-E Features Plot and Boxplot



B. Set A-C Features Plot and Boxplot

Fig 3. Feature space plots and box plots of various data sets

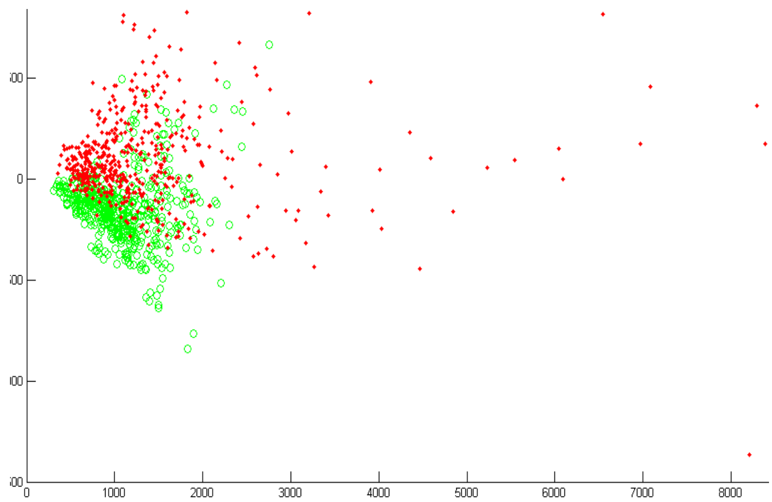


Fig 4. Set C-D Feature space plot

Figure 4 shows the feature plot of interictal EEG sampled from epileptogenic zone (set C) and interictal EEG sampled from opposite hemisphere (setD).Clearly they are not easy to get distinguished by using any classifier. Hence, these features are not sufficient for this case. So, entropy was calculated on D3 sub-band along with two previously defined features on all bands. Thus, we have 800 by 11 feature matrix which is then reduced to three dimension by PCA. Feature space so obtained, is shown in figure 5. 500 samples are given for training K-Nearest Neighbor classifier with K value chosen as 6 and 300 samples are used for testing purpose. Results obtained along with some other methods reported earlier for same database are summarized in table 3.

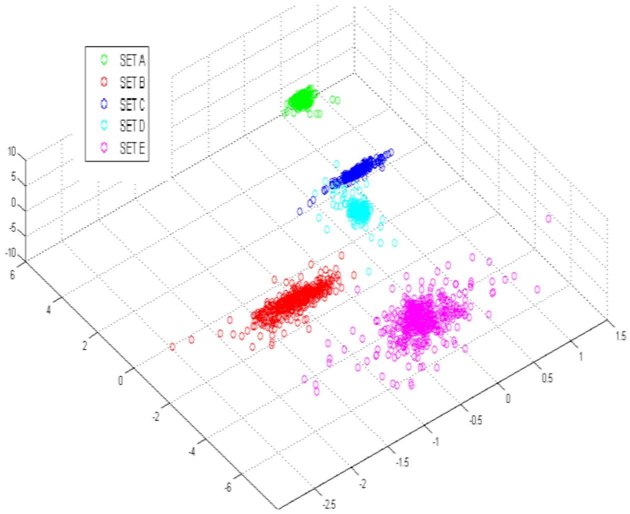


Fig 5. Features space

Author	Method	SETS	Accuracy %
Tazallas et al. [19]	Time frequency analysis, ANN	A-B-C-D-E AB-CD-E	89 98.60
Ubeyli & Guiler [20]	Eigen vector, Mixed models	A-B-C-D-E	98.60
Ubeyli [21]	Wavelet Transform, PNN	A-B-C-D-E	97.63
Ling et al. [22]	Time frequency analysis & App. Entropy, SVM	ABCD-E A-C-E A-B-C-D-E	98.51 98.67 85.9
This Work	Wavelet, SVM	AB-CD-E	100
	Wavelet, K-NN	A-B-C-D-E	100

Table 3

V. BACKGROUND

A method for the classification of EEG into different groups using wavelet based features has been presented here. As EEG is a non-stationary signal the wavelet transform gives good results. After wavelet decomposition at level 4 using daubechies wavelet of order 4, two statistical features i.e. mean absolute deviation of rectified EEG and mean absolute deviation of frequencies present were computed, over the wavelet decomposed signal at each level. Feature matrix obtained is then reduced to two dimension using PCA. These features are good enough to classify EEG as normal, interictal or ictal. However, these features are not able to classify between epileptogenic EEG and EEG from opposite hemisphere. But, if an additional feature entropy is also computed from D3 sub-band and PCA reduction is done to three dimensions. All five sets can be classified with 100 % accuracy. On Comparing this method to the earlier works done on the same dataset, this method is found to give better results.

REFERENCES

1. U. Rajendra Acharya, S. Vinitha Sree, G. Swapna, Roshan Joy Martis, Jasjit S. Suri, *"Automated EEG analysis of epilepsy: A review"*, Knowledge-Based Systems, vol. 45, 147–165, 2013.
2. World Health Organization, Epilepsy (*http://www.who.int/mental_health/neurology/epilepsy/en/index.html*) (last accessed 10.20.14).
3. Saeid Sanei, J. A Chambers, "*EEG signal processing*", John Wiley & Sons Ltd, pp. 1, 2007.
4. J. Gotman, "*Automatic recognition of epileptic seizures in the EEG*", Electroencephalography and Clinical Neurophysiology, vol. 54, no. 5, pp. 530–540, 1982.
5. Thasneem Fathima, M. Bedeeuzzaman, Omar Farooq and Yusuf U Khan, "*Wavelet Based Features for Epileptic Seizure Detection*", MES Journal of Technology and Management pp. 108-112.
6. Gotman, J., and Gloor, P., "*Automatic recognition and quantification of interictal epileptic activity in the human scalp EEG*", Electroencephalogram & Clinical Neurophysiology, vol. 41, 513–529, 1976.
7. Gotman, J. and Wang, L. Y., "*State-dependent spike detection: concepts and preliminary results*", Electroencephalogram & Clinical Neurophysiology, vol. 79, 11–19, 1991.
8. Glover Jr., J. R., Raghavan, N. Ktonas, P. Y. and Frost, J. D., "*Context-based automated detection of epileptogenic sharp transients in the EEG: elimination of false positives*", IEEE Transaction on Biomedical Engineering, vol. 36, 519–527, 1989.
9. Dingle, A. A., Jones, R. D., Caroll, G. J., and Fright, W. R., "*A multistage system to detect epileptiform activity*

in the EEG", IEEE Transactions on Biomedical Engineering, vol. 40, 1260–1268. 1993.

10. Kohonen, T., "*The Self-Organizing Maps*", 2nd edition, Springer, New York, 1997.
11. Kurth, C., Gilliam, F., and Steinhoff, B. J., "*EEG spike detection with a Kohonen feature map*", Ann. Biomedical Engineering, vol. 28, 1362–1369, 2000.
12. Mosquera CG, Trigueros AM, Franco JI, Va´zquez AN, "*New feature extraction approach for epileptic EEG signal detection using time-frequency distributions*", Medical & Biological Engineering & Computing, vol.48 (4), 321–330, 2010.
13. H. Adeli, S. Ghosh-Dastidar, and N. Dadmehr, "*A wavelet chaos methodology for analysis of EEGs and EEG sub-bands to detect seizure and epilepsy*," IEEE Transactions on Biomedical Engineering, vol. 54, no. 2, pp. 205–211, Feb. 2007.
14. H Ocak, "*Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy*", Expert Systems & Applications, vol. 36(2), 2027– 2036, 2009.
15. E. Juarez-Guerra, V. Alarcon-Aquino and P. Gomez-Gil., "*Epilepsy Seizure Detection in EEG Signals Using Wavelet Transforms and Neural Networks*", in the Proceedings of the Virtual International Joint Conferences on Computer, Information and Systems Sciences and Engineering (CISSE 2013). Dec. 12-14, 2013.
16. R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David and C. E. Elger, "*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, pp (061907)1-8, 2001.
17. V Srinivasan, C Eswaran, N Sriraam, "*Artificial neural network based epileptic detection using time-domain and frequency domain features*", Journel of Medical Systems, vo1. 29(6), 647–660, 2005.
18. H. Adeli, Z. Zhou, and N. Dadmehr, "*Analysis of EEG records in an epileptic patient using wavelet transform*," Journal of Neuroscience Methods, vol. 123, no. 1, pp. 69–87, 2003.
19. Tzallas, A. T., Tsipouas, M.G., & Fotiadis, D. I., "*Automatic seizure detection based on time frequency analysis and artificial neuron networks*", Computational intelligence and neuroscience, 2007.
20. Ubeyli, E. D., & Guler, I., "*Features extracted by eigenvector methods for detection variability of EEG signals*", Pattern Recognition Letters, 28(5), 592-603, 2007.
21. Ubeyli, E. D. "*Probabilistic neural networks combined with wavelet coefficients for analysis of EEG signals*", Expert systems, 26(2), 147-159, 2009.
22. Liang, S. F., Wang, H. C., & Chang, W. L., "*Combination of EEG Complexity and Spectral Analysis for Epilepsy Diagnosis and Seizure Detection*", EURASIP Journal on Advances in Signal Processing, 853-434, 2010.