This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2585661, IEEE Access

DWT based Epileptic Seizure Detection from EEG Signals using Naïve Bayes/k-NN Classifiers

A.Sharmila, *Member, IEEE* and P. Geethanjali, *Member, IEEE*

***Abstract*—Electroencephalogram (EEG) comprises valuable details related to the different physiological state of the brain. In this study, a framework is offered for detecting an epileptic seizure from EEG data recorded from normal subjects and an epileptic patient. This framework is based on discrete wavelet transform (DWT) analysis of EEG signals using linear and non-linear classifiers. The performance of fourteen different combinations of two-class epilepsy detection is studied using naïve Bayes (NB) and k-nearest neighbor (k-NN) classifiers for the derived statistical features from DWT. It has been found that the NB classifier performs better and shows an accuracy of 100% for the individual and combined statistical features derived from DWT values of normal eyes open and epileptic EEG data provided by University of Bonn, Germany. It has been found that computation time of NB classifier is lesser than k-NN to provide better accuracy. So, the detection of an epileptic seizure based on DWT statistical features using NB classifiers are more suitable in real-time for a reliable, automatic epileptic seizure detection system to enhance the patient's care and quality of life.**

***Index Terms*— Discrete Wavelet Transform (DWT), Electroencephalograms (EEG), Epilepsy, k-nearest neighbor (k-NN), and naïve Bayes (NB).**

I. Introduction

Electroencephalogram (EEG) is an effective, low-cost, non-invasive technique used in clinical studies to examine the electrical activity of the brain. EEG is one of the techniques to identify an abnormality of the brain. One of the chronic, non-communicable neurological disorders that can be studied from EEG is epilepsy. The neurological condition of the epilepsy is characterized by recurrent seizures, a momentary electrical disruption in the brain. These seizures may cause a disturbance in movement, control of bowel or bladder function, loss of consciousness or other disturbances in cognitive functions.

Traditionally, seizures are examined from 20 minutes recording of pre-seizure periods. However, long-term EEG recording is necessary in the case of infrequent epileptic seizure detection and it is time-consuming. Visual inspection of the EEG for seizure detection varies with the human expertise. Therefore, an automatic diagnosis of an epileptic seizure is crucial in the clinical environments. Pattern recognition is one of the techniques in detecting epileptic seizure from EEG signals by extracting hidden patterns from the EEG. There are many feature extraction techniques, such as time-domain [1-4], frequency-domain [5-8], time-frequency domain [9-15], which researchers are attempting to extract the hidden patterns in the EEG signals. Researchers have attempted to use multi-fractional analysis based on generalized fractal dimensions(GFD) and DWT on classification of epileptic EEG signals [49][50]. Researchers have attempted various classifiers namely artificial neural network [16-19], support vector machines [8][20-23],k-nearest neighbor (k-NN)[24,25],quadratic analysis[26], logistic regression[6,13], Naïve Bayes (NB) [13], decision tree [13, 27], Gaussian mixture model [2, 25], adaptive neuro-fuzzy inference systems [20, 31], mixture of expert model [28-30], surrogate data analysis [32, 33], learning vector quantization [34], Markov modeling [35] to classify the epileptic seizure abnormality from the EEG data.

All the above pattern recognition approaches focus on improving the classification accuracy with the various combinations of feature extraction and classification technique in the detection of an epileptic seizure. Therefore, the pattern recognition classification accuracy depends upon the type of features, a number of features and the classifier [36]. The objective of this study is to identify the most efficient pattern recognition method for reliable seizure detection.

In this paper, we make an attempt to obtain improved classification accuracy with less number of features for fourteen different combinations of data sets using linear naïve Bayes and non-linear k-NN classifier. This study examines publicly available five EEG datasets A, B, C, D & E provided by Department of Epileptology at University of Bonn, Germany [37]. Three statistical features i.e., mean absolute value (MAV), Standard deviation (SD) and Average power (AVP) were derived from D3-D5 and A5 of DWT coefficients. The derived features are studied using NB and k-NN classifiers to identify the output as an epileptic seizure or not. It is found that the NB classifier performs better in terms of all the individual and combined statistical features when

This paragraph of the first footnote will contain the date on which you

A. Sharmila is with the School of Electrical Engineering, VIT University, Vellore 632014, Tamil Nadu, India (e-mail: asharmila@vit.ac.in).

Corresponding Author: P. Geethanjali is with the School of Electrical Engineering, VIT University,Vellore 632014, Tamil Nadu, India (e-mail: geethanjali@vit.ac.in).

compared to the k-NN classifier for the normal eyes open and epilepsy data sets and gives an accuracy of 100%. The k-NN performs better with the SD and SD&MAV for the normal eyes open and epilepsy data sets.

II. EEG DATA SEGMENTATION

The open source epileptic data of the University of Bonn, Department of Epileptology, Germany, consists of five EEG data sets A to E for the duration of 23.6 sec from 100 single channels [37]. The data sets A and B correspond to EEG recordings of five normal subjects with their eyes in an open state and closed state. The data set C recorded before the epileptic attack at hemisphere hippocampal formation, data set D recorded from an epileptogenic zone and the data set E recorded during an occurrence of epilepsy within an epileptogenic zone. The data were recorded with 128-channel amplifier system, digitized with a sampling rate of 173.61 Hz. In this work, each channel data consist of 4097 samples is segmented into 8 equal data segments of size 512 and discarding the last data. Therefore, a total of 800 data segments is obtained for each data set from 100 single channels. Statistical features are derived from discrete wavelet transform coefficients of each data segment for identification of epileptic seizure using pattern recognition techniques. Fig.1. shows sample EEG signals for set A to E.

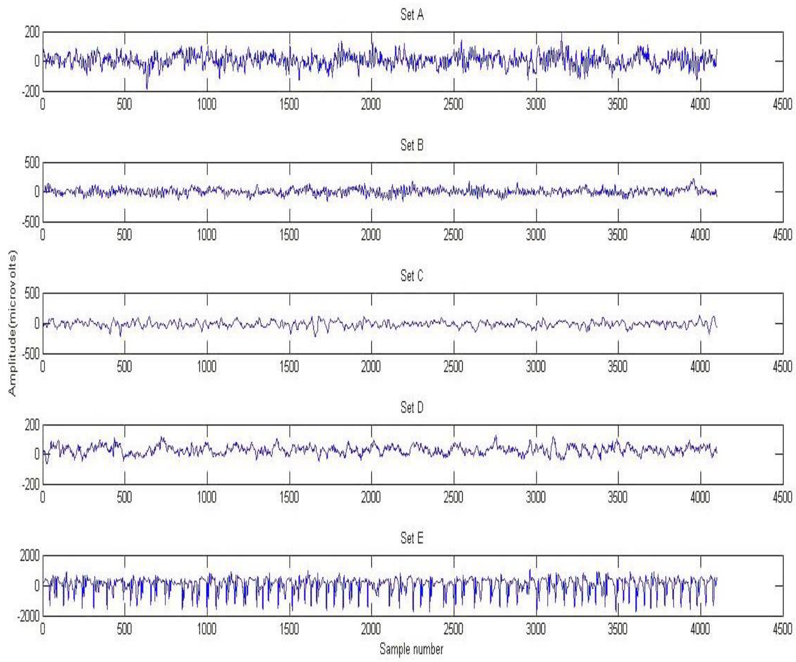

Fig. 1. Sample EEG signal for A to E

III. STATISTICAL FEATURES FROM DISCRETE WAVELET TRANSFORM (DWT) CO-EFFICIENT

EEG time series signals are non-stationary and may comprise more quantity of information apart from the high frequency oscillations due to electromagnetic interference, very low frequency artifacts from eye blinks, muscle movements,etc.,.Signal analysis based on fast Fourier transform (FFT) is not capable of capturing frequency information with the time of events [38].The time-frequency representation of time series signals is an attractive method to capture relevant frequency information at low frequencies along with relevant time information at the high-frequencies [39]. The wavelet transform (WT) is one such technique based on the multi-resolution analysis; decompose the signals into different frequency bands. This WT characteristic is useful in analyzing the epileptic seizure signal because the EEG signals contain low-frequency information with long time periods and high-frequency information with short time periods [39]. The WT could be continuous wavelet transform (CWT)/ discrete wavelet transform (DWT). The drawback of CWT is the high redundancy. However, with DWT, it is easier to decompose the signal into different levels using filter bank consisting of a group of filters. The wavelet decomposition of the signal x[n], using five level can be decomposed using filter bank as shown in Fig.2. Each level consists of filters with a down-sampler by 2.

In the initial stage, the signal x[n] is passed through high-pass filter h[n] and low-pass filter g[n] and the outputs of filters are regarded as first level detailed co-efficient, D1 and approximation(A1) correspondingly. D1, A1 represents the frequency content of an original signal. The approximation coefficient at every level is decomposed and the process has been repeated to get the subsequent levels of coefficient such as D2, A2, D3, A3, D4, A4, D5 and A5. At each stage of the decomposition, the filtering doubles the frequency resolution and down- sampling halves the time resolution.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2585661, IEEE Access

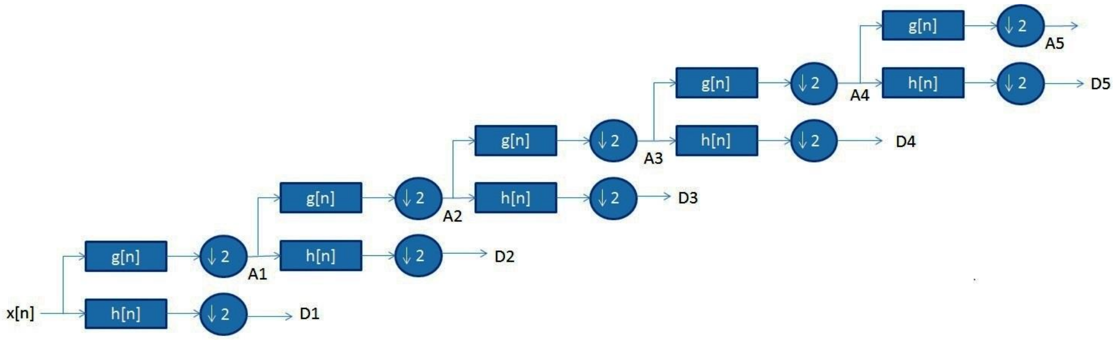

Fig. 2. Five level wavelet decomposition of the EEG signal

Daubechies 4(db4) wavelet function is chosen for five-level decomposition of epileptic EEG data into different frequency bands [30]. Figures 3 and 4 show the wavelet output for seizure-free and epilepsy signal. Most of the useful frequency component required to identify seizure from EEG signals are in the decomposition levels D3, D4, D5, and A5 [14]. So, the statistical features derived from these sub-bands alone can be applied to the NB and k-NN classifier to identify the epileptic seizure signal and the remaining coefficients (D1&D2) are not taken into account for obtaining the statistical features.

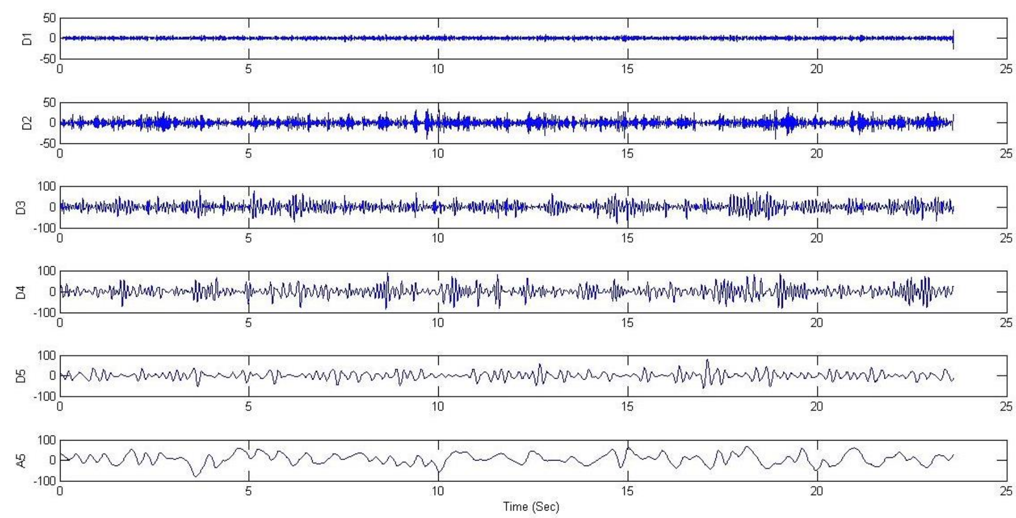

Fig. 3. Wavelet output for normal EEG signal

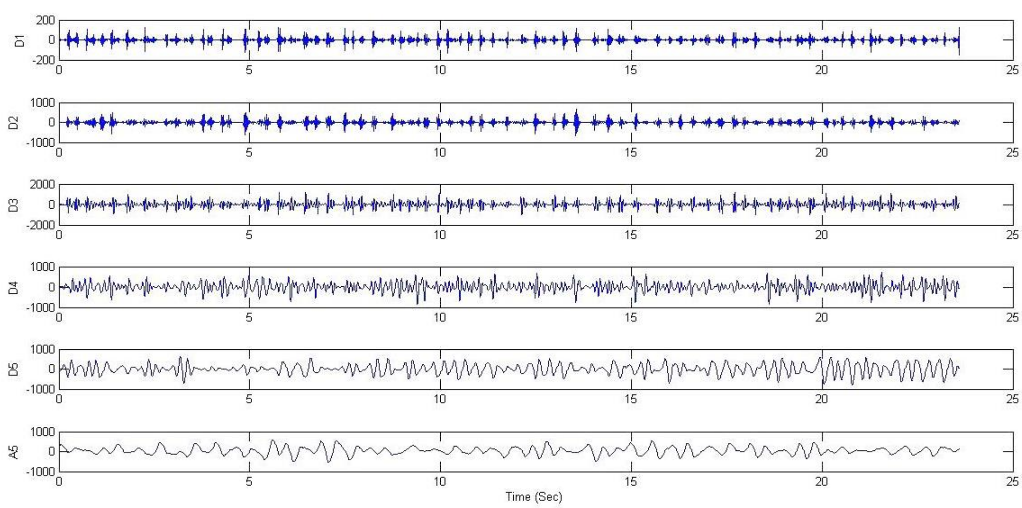

Fig. 4. Wavelet output for epilepsy EEG signal

Four statistical features were derived from D3-D5 and A5 [14].So, we also consider third to fifth level detailed coefficients (D3-D5) and the approximation (A5) in this work. Further, only three statistical features namely mean absolute value, average power, and standard deviations are derived from the coefficients of DWT. The following statistical features are derived from the coefficients of the DWT using the mathematical equation (1)-(3).

(i)Mean absolute value (MAV):

Mean absolute value is a measure of frequency information of the signal. This can be calculated using equation (1):

$$MAV = \frac{1}{N}\sum_{i=1}^{N}|x_i| \quad (1)$$

(ii) Average power (AVP):

This is another feature provides information about the frequency content of the signal and the mathematical expression is given below:

$$AVP = \frac{1}{N}\sum_{i=1}^{N}|x_i|^2 \quad (2)$$

(iii) Standard deviation (SD):

Standard deviation represents the amount of change in the frequency of the signal and calculated using the equation (3)

$$\mathrm{SD} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\mu)^2 \quad (3)$$

Where,

x_i – an ith sample of EEG data in a segment.

μ - mean of the segment.

N – Length of segment

IV. CLASSIFICATION

The statistical features derived from DWT are applied to the classifiers. The purpose of the classifier is to identify, epilepsy abnormality in EEG data by linear/non-linear mathematical approach. In this work, the classifiers used are naïve Bayes and k-nearest neighbor, to identify epileptic seizure EEG data for the individual and combined statistical features derived from DWT with a different combination of data sets A to D with E. The performance of NB and k-NN classifiers is assessed with the DWT based statistical features to detection the epileptic seizure abnormality.

A. Naive Bayes classifier

A naive Bayes is a probabilistic classifier which is based on Bayesian theory with assumptions of each feature of a particular class is independent of any other feature. Occurrence/particular absence estimation for NB model is based on maximum likelihood [40].The NB classifier requires less training data for classification and the classification is performed as given below :

Let D be a training set for n-classes with attribute vector Y and associated class labels. The attribute Y belongs to the class with highest posterior probability and is given using Eq. (4)

$$P(C_i|\,Y) > P(C_j|\,Y) \quad \text{for} \quad 1 \le j \le n\,, j \neq i \quad (4)$$

where

$$P(C_i|\,Y) = \frac{P(\,Y|C_i)P(C_i)}{P(Y)} \quad (5)$$

By Bayes theorem.

Where

$P(C_i)$ are the class prior probabilities.

$P(Y)$ is the prior probability of Y.

$P(C_i|\,Y)$ is the posterior probability.

$P(\,Y|C_i)$ is the posterior probability of Y conditioned on C_i .

As, $P(\,Y)$ is constant for all classes, only the numerator of $P(C_i|\,Y)$ need to be maximized. If the class prior probabilities are unknown, then $\mathrm{P}(C_1) = P(C_2) = \cdots = P(C_n)$ and $P(Y|C_i)$ is maximized. Otherwise, the class prior probabilities can be calculated by $P(C_i) = |C_{i,D}|/|D|$, where $|C_{i,D}|$ is the number of a training set of the class C_i in D.

To reduce the computation in an estimation of $P(Y|C_i)$, the classifier adopts the attributes that are independent conditionally of each other. Thus,

$$P(Y|C_i) = \prod_{k=1}^{n} P((y_k|C_i) \quad (6)$$

$$= P(y_1|C_i) \times P(y_2|C_i) \times \dots P(y_n|C_i) \quad (7)$$

The probabilities $P(y_1|C_i), P(y_2|C_i), \dots P(y_n|C_i)$ are determined from the training set and $'y_k'$ represent the value of an attribute for the data set Y.

To estimate the class label of Y , $P(Y|C_i)P(C_i)$ is evaluated for each class C_i.The classifier identifies the class label of attribute Y is C_i based on the condition given below:

$$P(Y|C_i)P(C_i) > P(Y|C_j)P(C_j) \quad \text{for} \quad 1 \le j \le n, j \neq i. \quad (8)$$

B. k-nearest neighbor Classifier

The k-nearest neighbor classifier is a nonparametric, nonlinear and relatively simple classifier [41].This method works intensively for larger training sets. It is based on similarity measure among the training and test set. The 'n' attributes are categorized by the data sets. Each set is a point in n-dimensional space and the training sets form the n-dimensional pattern space. A test/unknown data set is assigned to the class based on nearby 'k' data sets of training. The data sets 'nearness' is measured using Eq. (9)

$$(\mathrm{ED}) = \sqrt{\sum_{i=1}^{n}(Y_{1i} - Y_{2i})^2} \quad (9)$$

Where

$Y_{1i} = (y_{11}, y_{12} \dots y_{1n})$ and $Y_{2i} = (y_{21}, y_{22} \dots y_{2n})$

The values of each attribute can be normalized before doing the calculation on Euclidean distance(ED). As an alternative to taking the single nearest data set, the classifier normally takes a majority vote from the k-nearest neighbors. The value of k, the number of neighbors that gives the minimum error rate has been chosen as 2.

V. RESULTS AND DISCUSSION

In this study, fourteen different classification combinations, such as A-E, B-E, C-E, and D-E, AB-E, AC-E, AD-E, BC-E, CD-E, ABC-E, ACD-E, BCD-E and ABCD-E were used in order to identify the epileptic signal from EEG signal. This study aims to investigate comparison of the classifier for the identification of epilepsy from the EEG signals. The purpose of this study is to improve the classification accuracy with less number of features to identify the epilepsy abnormality from EEG data. EEG pattern recognition has been studied based on MAV, AVP, SD derived from wavelet sub-band frequencies D3-D5 and A5.The performance of k-NN & NB is studied with equal training and testing data sets.

For two-class classification, most of the researchers studied set A and set E and obtained accuracy in the range of 90 to 100% [5] [42-44] [13-15] [25-27] [31-35]. Researchers also classified set E with set B and achieved accuracy of 82.88% [44] & 92.5% [15]. Similarly set E is classified with set C and found the accuracy of 88.5%[44] & 100%[15] and with set D, the accuracy range of 79.94 % to 97% [15][44-46] is achieved. Researchers also studied two-class identification by considering different data sets combinations from A to D with E. The summary of accuracy with the different data sets combination for two-class identification is shown in Table 1.

TABLE I ACCURACY OF TWO-CLASS IDENTIFICATION FOR DIFFERENT COMBINATION OF DATA SETS A TO D WITH E

Combinations	Accuracy (%)	References
AB - E	100	[48]
CD- E	97.67	[45]
BCD - E	94	[15]
	98.22	[47]
ACD - E	96.65	[12]
	98	[34]
ABCD – E	94-100	[11,13,15,34,42,48]
AC- E,AD - E,BC - E, BD - E,ABC - E	Not reported	Reported in this work

The feature table shows the range of mean and standard deviation for 12 statistical features derived from D3, D4, D5 and A5 coefficients is presented in Table 2. The mean and standard deviation of epileptic seizure data set are considerably higher than other data sets such as normal(eyes open),normal (eyes closed),inter-ictal for entire 12 statistical features. The one-way ANOVA test is done with these 12 features for 14 different data sets combination to confirm that these features are significant for classification. From the test, it is observed that variance of between-class is higher than the variance of within-class and p-value <0.0001 indicates that these are significant features for identifying epileptic abnormality from the EEG signals.

Table 3 shows the performance of NB classifier for individual and combined statistical features such as MAV, SD and AVP. For the data set A-E, the highest accuracy is 100% for individual and combined statistical features. In using data set B-E, the highest accuracy of 99.25% is obtained with individual features such as SD and AVP and with combined features such as MAV+SD and MAV+SD+AVP. For the data set C-E, the highest accuracy of 99.62% is attained in using individual AVP feature. For the data set D-E, the highest accuracy of 95.12% is obtained with individual MAV feature and with combined SD+AVP features. For the data set AB-E, the highest accuracy of 99.16% is obtained in using SD and with MAV+SD and MAV+SD+AVP. In data set AC-E, the highest accuracy of 99.58% is attained with AVP and MAV+AVP features. In the data set AD-E, the highest accuracy of 96.66% is obtained in using individual MAV feature. For the data set BC-E, the highest accuracy of 98.25% is obtained in using SD and in data set BD-E, the highest accuracy of 96.5% is achieved in using MAV+SD and MAV+AVP. In the data set CD-E, the highest accuracy of 98.75% is obtained in using MAV+SD and in data set ABC-E, the highest accuracy of 98.68% is obtained with only individual features such as MAV and SD. In the data set ACD-E, the highest accuracy of 97.31 %is obtained with only individual MAV and SD features. For the data set BCD-E, the highest accuracy of 95.1% is attained in using MAV+SD and finally in the data set ABCD-E, the highest of 95.85% is achieved in using individual MAV feature. So from the Table 3, it is clear that the improved result can be attained in NB classifier in using either individual or combined statistical features derived from DWT coefficients.

Table 4 shows the performance of k-NN classifier for individual and combined statistical features such as MAV, SD and AVP. For the data set A-E, the highest accuracy is 100% for individual SD and combined MAV+SD features. In using data set B-E, the highest accuracy of 98.25% is obtained with individual MAV feature and with combined SD+AVP features. For the data set C-E, the highest accuracy of 97.25% is attained in using individual SD feature and combined MAV+SD. For the data set D-E, the highest accuracy of 95.62% is obtained with individual MAV feature and with combined SD+AVP features. For the data set AB-E, the highest accuracy of 98.83% is obtained in using MAV and in data set AC-E, the highest accuracy of 99.33% is attained MAV+SD features. In the data set AD-E, the highest accuracy of 97.08% is obtained in using individual MAV feature. For the data set BC-E, the highest accuracy of 97.33% is obtained in using MAV+SD and in data set BD-E, the highest accuracy of 96.33% is achieved in using MAV feature. In the data set CD-E, the highest accuracy of 96.08% is obtained in using SD feature and in data set ABC-E, the highest accuracy of 98% is obtained with MAV+ SD feature combination. In the data set ACD-E, the highest accuracy of 97.06 %is obtained with only individual SD feature. For the data set BCD-E, the highest accuracy of 96.37% is attained in using SD feature and finally in the data set ABCD-E, the highest of 97.1% is achieved in using individual SD feature. So from the Table 4, it is clear that the improved result can be attained in k-NN classifier in using either individual or combined statistical features derived from DWT coefficients.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2585661, IEEE Access

TABLE 2 FEATURE TABLE WITH RANGE OF MEAN±STANDARD DEVIATION FOR 12 FEATURES

Feature No.	**Features**	**Normal (eyes open)**	**Normal (eyes close)**	**Inter-ictal**	**Inter-ictal**	**Epileptic seizure**	**p-value**
1	MAV_d3	13.85±3.75	27.22±11.31	8.77±5.37	9.92±6.045	102.5±71.88	p<0.0001 for 14 data sets with 12 features
2	MAV_d4	13.58±3.43	24.98±11.12	14.041±6.28	17.63±11.02	127.87±75.21	
3	MAV_d5	10.83±2.87	12.79±4.1	17.42± 7.52	21.55±18	115.45±68.62	
4	MAV_a5	28.12±15.04	32.77±18.59	36.409±16.46	44.22± 38.43	86.579±56	
5	SD_d3	18.3±4.99	35.95±14.65	11.64±7.085	15.03±12.27	142.01±97.47	
6	SD_d4	17.89±4.61	33.33±14.89	18.6± 8.34	24.81±18.33	164.63±92.09	
7	SD_d5	14.13±3.78	16.65±5.28	22.89±10.05	29.5± 27.72	144.99± 83.87	
8	SD_a5	24.08±8.52	24.08±8.46±	36.54±16.87	46.53± 47.56	102.47±68.31	
9	AVP_d3	358.91±189.54	1504.6±1336.9	185.34± 279.7	375.88±992.43	29602±4007.7	
10	AVP_d4	340.97±173.22	1330±1324.7	414.75±411.6	949.65±2187	35508±3824.9	
11	AVP_d5	213.85±114.33	304.72± 197.48	624.01± 565.78	1635.6±5571.1	27998±3118	
12	AVP_a5	1325±1274	1762.6±1693.1	2292±1905.1	5100.4±1458.6	16015±2182	

TABLE 3 PERFORMANCE OF NB CLASSIFIER FOR INDIVIDUAL AND COMBINED STATISTICAL FEATURES

Data sets		MAV+ SD +AVP	MAV+ SD	MAV+ AVP	SD +AVP	MAV	SD	AVP
A-E	Accuracy (%)	100	100	100	100	100	100	100
	Sensitivity (%)	100	100	100	100	100	100	100
	Specificity (%)	100	100	100	100	100	100	100
B-E	Accuracy (%)	99.25	99.12	99.25	99	99	99.25	99.25
	Sensitivity (%)	99.49	99.74	99.49	99.49	99.49	100	99.49
	Specificity (%)	99	98.51	99	98.51	98.5	98.52	99
C-E	Accuracy (%)	99.5	99.5	99.5	99.12	99.12	99.5	99.62
	Sensitivity (%)	99.25	99.25	99.2	99.24	99.24	99.25	99.5
	Specificity (%)	99.74	99.74	99.74	99	99	99.74	99.74
D-E	Accuracy (%)	91.37	93.75	92	95.12	95.12	90	87.12
	Sensitivity (%)	95.59	94.41	96.67	95.23	95.23	93.95	97.14

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2585661, IEEE Access

	Specificity (%)	87.87	93.10	88.18	95.01	95.01	86.69	80.61
AB-E	Accuracy (%)	99.16	99.16	99.1	99.08	98.75	99.16	99.08
	Sensitivity (%)	98.02	98.02	98.02	98.02	97.5	98.02	98.02
	Specificity (%)	99.74	99.74	99.74	99.62	99.37	99.74	99.62
AC-E	Accuracy (%)	99.5	98.16	99.58	99.5	99.25	99.41	99.58
	Sensitivity (%)	98.76	97.25	99	98.76	98.99	98.51	99
	Specificity (%)	99.87	98.62	99.87	99.87	99.37	99.87	99.87
AD-E	Accuracy (%)	95.75	96.5	95.66	93.66	96.66	95.58	92.58
	Sensitivity (%)	94.17	93.65	94.16	94.26	94.11	93.92	96.14
	Specificity (%)	96.52	97.97	96.4	93.4	97.97	96.39	91.19
	Accuracy (%)	97.75	97.91	97.66	97.66	97.58	98.25	97.16
BC -E	Sensitivity (%)	98.69	98.7	98.43	98.43	98.68	99.47	98.15
	Specificity (%)	97.3	97.54	97.3	97.3	97.06	97.67	96.7
	Accuracy (%)	91.66	96.5	96.5	90.75	91.91	91.67	90.83
BD-E	Sensitivity (%)	92.98	93.87	93.87	93.65	92.2	93.35	96.47
	Specificity (%)	90.44	97.85	97.85	89.64	91.79	90.98	88.85
	Accuracy (%)	95.75	98.75	98.62	93.58	96.75	95.41	92.58
CD-E	Sensitivity (%)	94.4	97.5	97.48	94.73	94.13	93.89	96.41
	Specificity (%)	96.44	99.16	99	93.08	98.1	96.15	91.09
	Accuracy (%)	98.62	97.56	97	98.56	98.68	98.68	98.25
ABC-E	Sensitivity (%)	97.25	93.28	94.22	97	98.21	97.25	96.96
	Specificity (%)	99.08	99.07	97.92	99.08	98.84	99.16	98.67
ACD-E	Accuracy (%)	97.18	94.18	93.31	95.93	97.31	97.31	94.75

	Sensitivity (%)	94.04	91.37	91.26	94.19	93.01	93.64	94.88
	Specificity (%)	98.24	95.03	93.89	96.47	98.81	98.57	94.71
BCD-E	Accuracy (%)	93.56	95.1	95	93.43	94.43	93.87	93.18
	Sensitivity (%)	91.36	90.12	90.98	92.5	91.68	91.94	94.49
	Specificity (%)	94.19	97.25	95.89	93.69	95.27	94.43	92.85
ABCD-E	Accuracy (%)	95.25	91.8	91.8	94.9	95.85	95.6	94.4
	Sensitivity (%)	90.88	92.41	93.94	91.38	89.92	90.83	92.6
	Specificity (%)	96.25	91.58	89.96	95.67	97.31	96.72	94.76

TABLE 4 PERFORMANCE OF k-NN CLASSIFIER FOR INDIVIDUAL AND COMBINED STATISTICAL FEATURES

Data sets		MAV+ SD +AVP	MAV+ SD	MAV+ AVP	SD +AVP	MAV	SD	AVP
A-E	Accuracy (%)	99.87	100	99.87	99.87	99.87	100	99.87
	Sensitivity (%)	100	100	100	100	100	100	100
	Specificity (%)	99.75	100	99.75	99.75	99.75	100	99.75
B-E	Accuracy (%)	97.87	97.87	97.87	98.25	98.25	97.62	97.87
	Sensitivity (%)	100	100	100	99.48	99.48	100	100
	Specificity (%)	95.92	95.92	95.92	97.07	97.07	95.46	95.92
C-E	Accuracy (%)	95.5	97.25	95.5	95.25	95.25	97.25	95.5
	Sensitivity (%)	96.42	97.25	96.42	95.25	95.25	97.48	96.42
	Specificity (%)	94.6	97.25	94.6	95.25	95.25	97.01	94.6
D-E	Accuracy (%)	93.87	94.5	93.87	95.62	95.62	94.75	93.87
	Sensitivity (%)	93.54	93.41	93.54	94.4	94.4	94.3	93.54
	Specificity (%)	94.2	95.6	94.2	96.91	96.91	95.2	94.2

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2585661, IEEE Access

AB-E	Accuracy (%)	98.58	98.53	98.58	98.58	98.83	98.41	98.58
	Sensitivity (%)	100	100	100	100	99.48	100	100
	Specificity (%)	97.91	97.91	97.91	97.91	98.51	97.67	97.91
AC-E	Accuracy (%)	97	99.33	97	97	96.83	98.16	97
	Sensitivity (%)	96.42	98.51	96.42	96.42	95.25	97.48	96.42
	Specificity (%)	97.27	99.74	97.27	97.27	97.62	98.5	97.27
AD-E	Accuracy (%)	95.91	96.33	95.91	95.91	97.08	96.5	95.91
	Sensitivity (%)	93.54	93.41	93.54	93.54	94.4	94.3	93.54
	Specificity (%)	97.11	97.84	97.11	97.11	98.47	97.61	97.11
BC -E	Accuracy (%)	96.5	97.33	96.5	96.5	96.25	97.16	96.5
	Sensitivity (%)	96.85	97.42	96.85	96.85	94.71	97.65	96.85
	Specificity (%)	96.33	97.29	96.33	96.33	97.01	96.93	96.33
BD-E	Accuracy (%)	94.91	95.5	94.91	94.91	96.33	95.5	94.91
	Sensitivity (%)	93.35	92.25	92.4	93.35	94.05	94.13	93.35
	Specificity (%)	95.67	97.58	95.65	96.67	97.48	96.16	95.67
CD-E	Accuracy (%)	94.58	95.75	94.58	94.58	95.33	96.08	94.58
	Sensitivity (%)	92.4	97.42	96.85	92.4	92.15	93.58	92.4
	Specificity (%)	95.65	98.18	97.53	95.65	96.96	97.35	95.65
ABC-E	Accuracy (%)	97.37	98	97.37	97.37	97.18	97.87	97.37
	Sensitivity (%)	96.85	92.25	92.4	96.85	94.71	97.65	96.85
	Specificity (%)	97.53	98.39	97.09	97.53	98	97.94	97.53
ACD-E	Accuracy (%)	95.93	96.8	95.98	95.93	96.5	97.06	95.93
	Sensitivity (%)	92.4	92.28	92.5	92.4	92.17	93.58	92.4

	Specificity (%)	97.09	97.57	96.53	97.09	97.98	98.24	97.09
BCD-E	Accuracy (%)	95.56	96.25	95.56	95.56	96.12	96.37	95.56
	Sensitivity (%)	92.5	92.28	92.5	92.5	91.83	93.62	92.5
	Specificity (%)	96.53	98.18	97.39	96.53	97.57	97.26	96.53
ABCD-E	Accuracy (%)	96.45	97	96.45	96.45	96.9	97.1	96.45
	Sensitivity (%)	92.5	93.2	93.35	92.5	91.83	93.62	92.5
	Specificity (%)	97.39	96.62	95.67	97.39	98.18	97.94	97.39

It is observed from Table 5 that average accuracy of k-NN is 97.45% and it is 97.83% in using NB for the data sets combination. For the data set A-E, both NB and k-NN is best presented and attains 100% which is comparable with researchers[14][15][47][48].

TABLE 5 HIGHEST ACCURACY ACHIEVED BY NB AND k-NN FOR THE DERIVED STATISTICAL FEATURES FROM DWT

Data sets	NB	k-NN
	Accuracy (%)	Accuracy (%)
A-E	**100**	**100**
B-E	**99.25**	98.25
C-E	**99.62**	97.25
D-E	95.12	**95.62**
AB-E	**99.16**	98.83
AC-E	**99.5**	99.33
AD-E	96.66	**97.08**
BC-E	**98.25**	97.33
BD-E	**96.5**	96.33
CD-E	**98.75**	96.08
ABC-E	**98.68**	98
ACD-E	**97.31**	96.8
BCD-E	95.1	**96.37**
ABCD-E	95.85	**97.1**

It is evident from Table 3 and Table 4 that epileptic detection accuracy in case of NB with individual and combined statistical features are ranging from 95.1% to 100% whereas, in the case of k-NN, it is from 95.62 % to 100% for different data sets combinations.

Using NB, the highest accuracy is attained for 9 data sets and k-NN provides highest accuracy for 4 data sets. The highest accuracies obtained for the data sets combinations AC - E, BC - E, ABC- E is 99.5%, 97.75%, 98.62% using NB classifier. Further, the highest accuracies for the datasets AD – E, BD – E using k-NN classifier is 95.91% and 93.35%, respectively. The program was written in MATLAB software package R2014b environment and run on 1.6 GHz HP CPU processor machine with 8GB of memory.

We found that the accuracy obtained for normal eyes open and epileptic seizure EEG data sets using NB classifier is 100% for all statistical feature combinations. Also, it has been observed that in the k-NN classifier, for the data sets (A-E) gives 100 % accuracy only with SD and MAV & SD features of DWT frequency sub-bands D3-D5 and A5.The results are compared with the outperforming other pattern recognition methods as shown in Table 1.

VI. Conclusion

Pattern recognition approaches, in medical diagnosis systems, necessitate that the medical data to be inspected in lesser time with good accuracy. In this study, only three statistical features derived from EEG signals are vital for outstanding epileptic seizure detection. The NB and k-NN classifiers are used for identification of epileptic seizure from EEG signals. The results have been shown that the proposed pattern recognition technique could attain a higher accuracy of 100% using NB classifier for normal eyes open and epileptic seizure EEG data sets for all individual and combined statistical features derived from DWT for the detection of an epileptic seizure. The study of the proposed technique is evident from the other pattern recognition approaches considered by the researchers for fourteen different combinations of data sets A to D with E and we confirm that the NB classifier achieves better accuracy for 9 data set combinations with less computation time and k-NN attains better accuracy for 4 data sets for the detection of epileptic seizure abnormality.

ACKNOWLEDGMENT

We would thank the University of Bonn, Germany for their database. We would like to thank our VIT University for providing all help and support for our contribution.

REFERENCES

[1] N. Prada, P. K. Sadasivan, and G. R. Arunodaya, "Detection of seizure activity in EEG by an artificial neural network: A preliminary study," Computers and Biomedical Research, vol. 29, no.4, pp. 303–313, Aug.1996.

[2] V. P. Nigam and D. Graupe,"A neural-network-based detection of epilepsy," Neurological Research, vol. 26, no.1, pp. 55–60, Jan.2004.

[3] U.R. Acharya, F. Molinari, S.V. Sree, S. Chattopadhyay, Kwan-Hoong Ng and J.S. Suri, "Automated diagnosis of epileptic EEG using entropies," Biomedical Signal Processing and Control ,vol.7,no.4,pp.401–408,Jul.2012.

[4] V. Srinivasan, C. Eswaran and N. Sriraam,"Approximate entropy-based epileptic EEG detection using artificial neural networks," IEEE Transaction on Information Technology in Biomedicine,vol.11,no.3,pp. 288–295, May.2007.

[5] K. Polat and S.Gunes, "Artificial immune recognition system with fuzzy resource allocation mechanism classifier, principal component analysis and FFT method based new hybrid automated identification system for classification of EEG signals," Expert Systems with Applications, vol.34,no.3, pp. 2039-2048, April 2008.

[6] A. Alkan, E. Koklukaya and A. Subasi, " Automatic seizure detection in EEG using logistic regression and artificial neural network," Journal of Neuroscience Methods, vol.148,no.2, pp.167-176,Oct. 2005.

[7] Ubeyli, E. D. (2010a). Least squares support vector machine employing model-based methods coefficients for analysis of EEG signals. Expert Systems with Applications, 37, 233–239.

[8] S.R.Mousavi, M. Niknazar and B.V. Vahdat,"Epileptic seizure detection using AR model on EEG signals," in Proceedings of Cairo International Biom edical Engineering Conference , Cairo, Egypt, December 2008, pp. 1–4.

[9] A.T. Tzallas , M.G. Tsipouras and D.I. Fotiadis," Epileptic seizure detection in EEGs using time-frequency analysis," IEEE Transaction on Information Technology in Biomedicine, vol.13, no.5, pp. 703-710,Sep.2009.

[10] 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 Transaction on Biomedical Engineering, vol.54, no.2, pp. 205-211,Feb. 2007.

[11] L. Guo, D. Rivero, and A. Pazos," Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks," Journal of Neuroscience Methods, vol.193, no.1, pp.156-163, Oct.2010.

[12] H. Ocak," Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy," Expert Systems with Applications, vol.36, no.2, pp.2027–2036, Mar.2009.

[13] U. Orhan, M. Hekim, and M. Ozer, "EEG signals classification using the K-means clustering and a multilayer perceptron neural network model," Expert Systems with Applications, vol.38, no.10, pp.13475–13481, Sep.2011.

[14] A. Subasi and I. Gursoy," EEG signal classification using PCA, ICA, LDA and support vector machines," Expert Systems with Applications,vol. 37,no. 12,pp.8659–8666,Dec.2010.

[15] Kumar. Yatindra, M. L. Dewal, and R. S. Anand. "Epileptic seizures detection in EEG using DWT-based ApEn and artificial neural network."Signal, Image and Video Processing 8.7 (2014): 1323-1334.

[16]S. Ghosh-Dastidar, H. Adeli, and N. Dadmehr," Mixed-band wavelet-chaos-neural network methodology for epilepsy and epileptic seizure detection," IEEE Transaction on Biomedical Engineering, vol.54, no.9, pp.1545-1551, Sep.2007.

[17] E. D.Ubeyli, "Analysis of EEG signals using Luapunov exponents," Neural Network World, vol.16, no.3, pp.257-273, May 2006.

[18] E.D.Ubeyli," Probabilistic neural networks combined with wavelet coefficients for analysis of EEG signals," Expert systems, vol.26, no.2, pp.147-159, May 2009.

[19] E.D. Ubeyli," Lyapunov exponents/probabilistic neural networks for analysis of EEG signals," Expert Systems with Applications, vol.37, no.2 ,pp. 985–992,Mar.2010.

[20] I. Guler, and E.D. Ubeyli," Adaptive neuro-fuzzy inference system for classification of EEG signals using wavelet coefficients," Journal of Neuroscience Methods, vol.148, no.2, pp.113-121, Oct.2005.

[21] Y.Liu, W. Zhou, Q.Yuan and S.Chen," Automatic seizure detection using wavelet transform and SVM in long-term intracranial EEG," IEEE Transactions on Neural Systems and Rehabilitation Engineering, vol.20, no.6, pp.749–55, Nov.2012.

[22] L.Chisci , A. Mavino , G. Perferi , M. Sciandrone, C.Anile and G. Colicchio ," Real-time epileptic seizure prediction using AR models and support vector machines," IEEE Transactions on Biomedical Engineering, vol.57, no.5, pp.1124–1132,May.2010.

[23] Y. Tang, D. Durand," A tunable support vector machine assembly classifier for epileptic seizure detection," Expert System with Applications, vol.39, no.4, pp.3925–3938, Mar.2012.

[24] L.Guo, D. Rivero, J. Dorado, C.R. Munteanu, and A. Pazos," Automatic feature extraction using genetic programming: An application to epileptic EEG classification," Expert Systems with Applications, vol.38, no.8, pp.10425-10436, Aug.2011.

[25] C.A.Lima and A.L.Coelho," Kernel machines for epilepsy diagnosis via EEG signal classification: A comparative study," Artificial Intelligence in Medicine, vol.53, no. 2, pp.83-95, Oct.2011.

[26] D. Gajic, Z. Djurovic, S. DiGennaro and F. Gustafsson," Classification of EEG signals for detection of epileptic seizures based on wavelets and statistical pattern recognition," Biomedical Engineering: Applications, Basis, Communications, vol.26, no.2, Apr. 2014.

[27] K. Polat, and S. Gunes," Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform," Applied Mathematics and Computation, vol.187,no.2, pp.1017–1026,Apr.2007.

[28] E.D. Ubeyli, " Modified mixture of experts for analysis of EEG signals," in Proceeding of 29th Annual International IEEE conference on Engineering in Medicine and Biology Society,Lyon,France,2007, pp. 1546-1549.

[29] E.D. Ubeyli, "Wavelet/mixture of experts network structure for EEG classification," Expert Systems with Applications, vol.34, no.3, pp.1954-1962, Apr.2008.

[30] A.Subasi, "Signal Classification using wavelet feature extraction and a mixture of expert model," Expert Systems with Applications, vol.32, no.4, pp.1084-1093, May 2007.

[31] N. Kannathal, M.L. Choo, U.R. Acharya, and P.K. Sadasivan," Entropies for detection of epilepsy in EEG," Computer Methods and Programs in Biomedicine, vol. 80, no.3, pp.187-194, Dec.2005.

[32] N.Kannathal, U.R.Acharya, C.M.Lim and P.K.Sadasivan,"Characterization of EEG - A comparative study," Computer Methods and Programs in Biomedicine, vol.80, no.1, pp.17–23, Oct.2005.

[33] H. Ocak," Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy," Expert Systems with Applications, vol.36, no.2, pp.2027–2036, Mar.2009.

[34] H. Ocak, "Optimal classification of epileptic seizures in EEG using wavelet analysis and genetic algorithm," Signal Processing, vol.88, no.7, pp.1858–1867, Jul. 2008.

[35] B.Direito, C. Teixeira, B. Ribeiro, M.Castelo-Branco M, F.Sales and A. Dourado," Modeling epileptic brain states using EEG spectral analysis and topographic mapping," Journal of neuroscience methods, vol. 210, no.2, pp.220–229, Sep.2012.

[36] P. Geethanjali , K. K.Ray," A Low-Cost Real-Time Research Platform for EMG Pattern Recognition-Based Prosthetic Hand," IEEE/ASME Transactions on mechatronics, vol. 20, no.4, pp-1948-1955, Aug.2015.

[37] EEG database from University of Bonn [Online]. Available: http://www.epileptologiebonn.de.

[38] M. K. Kiymi, I. Güler, A. Dizibüyük and M.Akin," Comparison of STFT and wavelet transform methods in determining epileptic seizure activity in EEG signals for real-time application," Computers in Biology and Medicine, vol.35, no.7, pp. 603–616, Oct 2005.

[39] H. Adeli, Z. Zhou, N. Dadmehr," Analysis of EEG records in an epileptic patient using wavelet transform," Journal of neuroscience methods," vol.123,no.1,pp.69–87, Feb.2003.

[40] A.H.Fielding, Cluster and classification techniques for the biosciences, Cambridge, UK: Cambridge University Press, 2007.

[41] T.Cover and P.Hart," Nearest neighbor pattern classification," IEEE Transactions on Information Theory, vol.13, no.1, pp.21–27, Jan.1967.

[42] A.T. Tzallas, M.G. Tsipouras and D.I. Fotiadis," Automatic seizure detection based on time-frequency analysis and artificial neural networks," Computational Intelligence and Neuroscience, Dec.2007.

[43] A.T. Tzallas, M.G. Tsipouras and D.I.Fotiadis, " The use of time-frequency distributions for epileptic seizure detection in EEG recordings," in Proceeding of 29th Annual International IEEE conference on Engineering in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2016.2585661, IEEE Access

Medicine and Biology Society,Lyon,France,2007, pp.3-6 .

[44]N. Nicolaou and J. Georgiou,” Detection of epileptic electroencephalogram based on permutation entropy and support vector machine,” Expert Systems with Applications, vol.39, no.1, pp.202–209, Jan.2012.

[45] R.G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David and C.E. E lger,” 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, no.6, p.061907, Nov.2001.

[46]Y. Kaya, M. Uyar, R. Tekin, and S. Yildirim, “1D-local binary pattern bas ed feature extraction for classification of epileptic EEG signals,” Applied Mathematics and Computation , vol. 243, pp. 209 – 219, Sep.2014.

[47] Y. Kumar, M.L.Dewal and R.S.Anand,”Epileptic seizure detection using DWT based fuzzy approximate entropy and support vector machine,” Neurocomputing, vol.133, pp. 271–279, Jun.2014.

[48] G.Chen ,”Automatic EEG seizure detection using dual-tree complex wavelet-Fourier features,” Expert Systems with Applications, vol. 41.no.5,pp.2391-2394, Apr.2014.

[49] Easwaramoorthy, D., and R. Uthayakumar. "Analysis of biomedical EEG signals using Wavelet Transforms and Multifractal Analysis. “Communication Control and Computing Technologies (ICCCCT), 2010 IEEE International Conference on. IEEE, 2010.

[50] Uthayakumar, R., and D. Easwaramoorthy. "Epileptic seizure detection in EEG signals using multifractal analysis and wavelet transform." Fractals21.02 (2013): 1350011.