Abstract ID: P203

Clustering Based Seizure Detection in EEG

Hammad Y Khan[1], Asadullah Khan[2], Atif Belal[1], Y U Khan[2]
Department of Computer Engineering
[1]Z. H College of Engineering & Technology, A.M.U, Aligarh
Email*: hammadi99@gmail.com, , imatif17@gmail.com*

Department of Electrical Engineering
[2]Z. H College of Engineering & Technology, A.M.U, Aligarh
Email: *asadkhan150899@gmail.com*, *yusufkhan.ee@amu.ac.in*

1. ABSTRACT

A system which utilizes K-means clustering to identify the presence of epileptic seizures in EEG is proposed. The signal is first wavelet filtered, then energy and entropy is calculated on each selected level. The features extracted are first used to train the clusters. The test data comprising of seizure as well as normal patterns from the EEG signal is finally used to assess the performance of the proposed methodology. Different combinations of normal and seizure EEG are used to test the proposed classifier. The data used for this work was taken from Andrzejak et al. [1]. It consists of set A, B, C, D, and E. Sets A and B consist of normal EEG recording. Sets C and D are an Inter-ictal EEG recordings and set E consists seizure EEG recordings. Dataset combinations of dataset (A and E), (AB and E), (C and E) and (ABCD and E) were used to compare EEG recordings. The results obtained showed that the trained K-means cluster had an accuracy of 100%. It clearly suggests that the features selected were able to discriminate between the seizure and normal activity.

2. KEYWORDS:

EEG, epileptic seizures, energy, entropy. *K-means* clustering, wavelet transform

3. INTRODUCTION

Electroencephalogram (EEG) is the recording of fluctuations in the electrical activity of large group of neurons in the brain. [2]. The detection of epileptic seizures in the EEG is of particular importance in the diagnosis of epilepsy. Different automated methods of detection have been proposed by the different research groups. The results on both real (offline and online) and surrogate data have been reported. Each group have observed different results under varied environment. EEG being complex in nature, 100% accuracy is difficult to obtain in an online environment.

Many automated EEG signal classification and seizure detection systems using different approaches have emerged in the past. Among them were, Gotman and Gloor's [3] method of recognition and quantification of inter-ictal epileptic activity (spikes and sharp waves) in human scalp EEG. Gotman [4] proposed the first widely applicable method for automatic seizure detection featuring elementary half waves acquired from decomposition of EEG signals that were extracted from each epoch. The accuracy achieved for detection was 73%. They successfully reported a low false positive rate of 0.5/hour whist maintaining a high sensitivity of 90%. It was later modified after extensive evaluation in 1990 [5]. While Qu and Gotman [6] proposed the use of the nearest-neighbor classifier on EEG features extracted in both time and frequency, Khan and Gotman [7] performed the validation of the traditional Gotman [4] method. They proposed the wavelet based method to capture rhythmic nature of seizure discharge using intracranial recordings. The detection sensitivity was 86%. Gigola et al [8] applied a method based on the evolution of accumulated energy using wavelet analysis for the prediction of epileptic seizure onset from intracranial epileptic EEG recordings. D. Gajic et al [9] has a limitation of using large number of features, extracted after wavelet transform, to make an objective decision about the type of the EEG data processed and thus the brain state of a patient. Khan and Rafiuddin [10] used a wavelet technique to extract features, relative energy and normalized coefficient of variation (NCOV) and reported an accuracy of 92% on database described in [1]. Also, Fatima et al [11] used statistical features like variance, skewness, and coefficient of variation as parameters and reported an accuracy of 97%. Many algorithms have been proposed since then to detect the seizures using Recurrent Neural Networks, autoregressive model using Artificial Neural Network and Discrete Wavelet Transform (DWT) followed by Probabilistic Neural Network.

In the proposed work, two level discrete wavelet transform is used to extract features, energy and approximate entropy, on the raw signal. By using the wavelet transform, a signal can be broken into several components enabling to extract discriminatory features. Finally, the clustering of the data is performed using K-means clusters to discriminate between seizure and normal EEG.

4. METHODOLOGY

A. Data Used

In the work, a publicly available EEG database [1] is used. The dataset contains five sets each containing 100 channel EEG segments. These segments were selected and cut out from continuous multi-channel EEG recordings after visual inspection for artefacts, e.g., due to muscle activity or eye movements. Sets A and B consisted of segments taken from surface EEG recordings that were carried out on five healthy volunteers using a standardized electrode placement scheme. Volunteers were relaxed in an awake state with eyes open (A) and eyes closed (B), respectively. Sets C, D, and E originated from EEG archive of pre surgical diagnosis. EEGs from five patients were selected, all of whom had achieved complete seizure control after resection of one of the hippocampal formations, which was therefore correctly diagnosed to be the epileptogenic zone. Segments in set D were recorded from within the epileptogenic zone, and those in set C from the hippocampal formation of the opposite hemisphere of the brain. While sets C and D contained only activity measured during seizure free intervals, set E only contained seizure activity.
EEG was recorded with the same 128 channel amplifier system and digitalized at 173.61 Hz sampling rate and 12 bit A/D resolution. A band-pass filter having a passband of 0.53-40 Hz (12 dB/oct) was used to select the EEG signal of the desired band. The data used has already gone through the pre-processing steps.

B. Wavelet Transformation

In general, features are extracted from the raw EEG data using time domain, frequency domain or time–frequency domain methods. Since, EEG is not stationary in general, it is appropriate to use time–frequency domain methods like the wavelet transform (WT) as means for feature extraction. It provides both time and frequency views of a signal simultaneously, which makes it possible to accurately capture and localize transient features in the data like the epileptic spikes. Features extracted from the EEG are used for training clusters that differentiate between normal and epileptic EEG.
In the proposed work, using the wavelet transform, the given raw signal is broken into several components enabling to extract discriminatory features. The sampling rate of the signal is 173 Hz, implying the maximum countenanced frequency is 86 Hz.
Furthermore, wavelets allow decimation in time and frequency, simultaneously. In this transform we filter out portion of signal by passing it through half band low pass and half band high pass filters. It results in a frequency range sets of (0-43 Hz) and (43-86) Hz at level one. Filtering out the low frequency range set i.e. (0-43) Hz and further passing it to half band low pass and half band high pass filter, we procured (0-21) Hz and (22-43) Hz frequency sets. The second level approximation and detail coefficients are used for feature evaluation.

B.1 Discrete wavelet transform:

The discrete wavelet transform (DWT) has become a powerful technique in biomedical signal processing. DWT uses scale and position values based on powers of two. Mathematically, it is described as follows:

$$\mathrm{DWT}_{j,k}(t) = \frac{1}{\sqrt{|2^j|}} \int_{-\infty}^{\infty} x(t)\Psi \frac{t - 2^j * k}{2^j}\, dt \tag{1}$$

Wavelet transform analysis also known as multi resolution decomposition was done at level two which is schematically represented as shown in Fig.1. It is also known as wavelet tree.

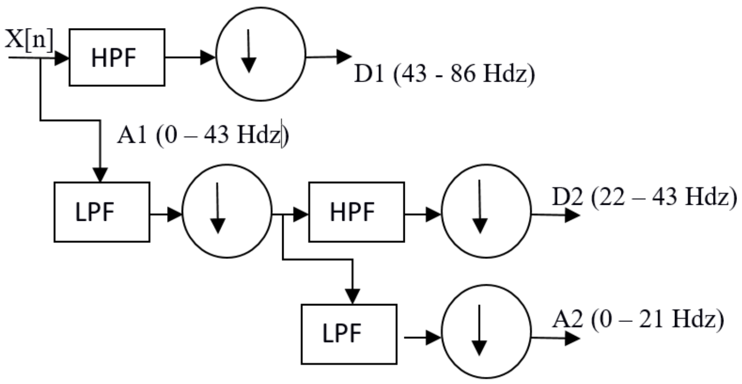

Figure 1: Discrete wavelet transform on the original EEG signal.

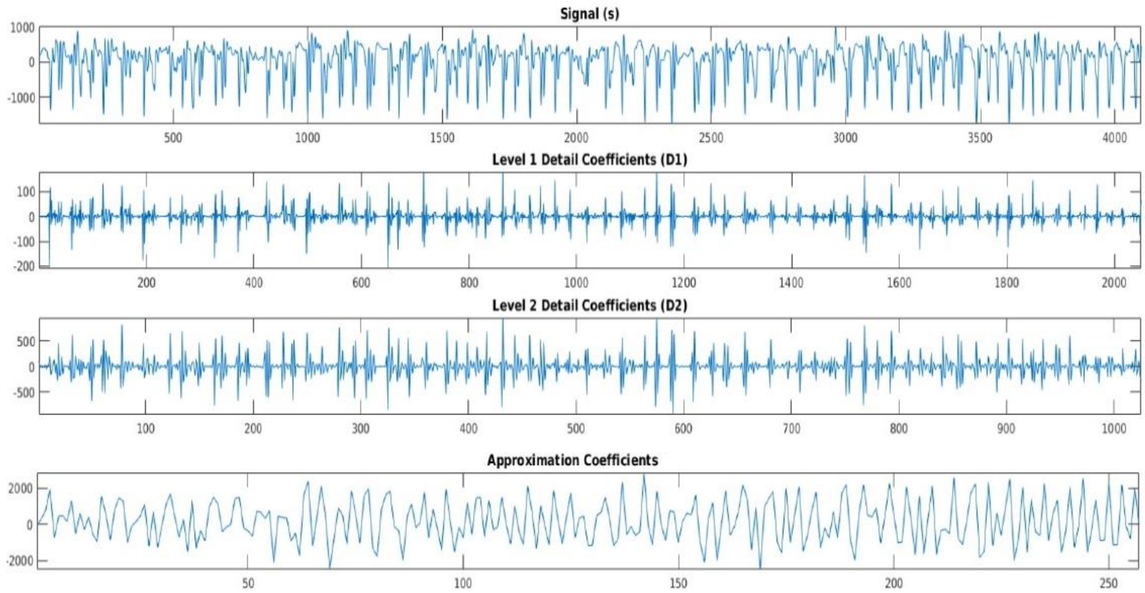

Figure 2: Two level wavelet decomposition of the EEG signal from set E. The signal ‘s’ has been decomposed into two levels of detail coefficients: D1, D2, and the residue signal is given as approximation coefficients.’

A2 and D2 are frequency range sets of (0-21) Hz and (22-43) Hz, used for feature extraction described below.

C. Feature Extraction

The performance of a classifier depends significantly on the features selected for classification. Therefore, the selection of optimal features is essential. The aim of extracting features is to choose minimum sets of parameters that can discriminate between seizure and non-seizure class.

1.1 Energy

Depending on the type of seizure the EEG undergoes a change in amplitude and frequency during the seizure. Hence, the feature that reflects the change in amplitude can be utilized. Energy is one such feature and is used in this work.

The energy, E for ‘N’ samples can be calculated by squaring and summing the amplitude as:

$$\mathrm{E} = \sum_{n=1}^{N} x_n^2 \qquad (2)$$

1.2 Approximate Entropy(A_pE_n)

An approximate entropy is a technique used to quantify the amount of regularity and the unpredictability of fluctuations over time-series data. It is known that the value of the approximate entropy falls abruptly during an epileptic seizure. As the seizures are commonly rhythmic, approximate entropy should be a good feature in the classification of seizures.

D. K-Means Clustering

K-means is a simple unsupervised machine learning algorithm to group similar data points and discover underlying patterns. To achieve this objective, K-means looks for a fixed number (*k*) of clusters in a dataset and a centroid for each cluster. During the test, every data point is allocated to a cluster by reducing the in-cluster sum of squares. In the proposed work two clusters are used one for seizure and other for normal signals.

E. Performance Measures

The performance of the clusters was assessed using standard statistical measures, Sensitivity, Specificity, and Accuracy. These are defined as follows:

$$Sensitivity = TP/ (TP + FN)$$
$$Specificity = TN/ (FP+TN)$$
$$Accuracy = (TP+TN)/ (TP+FN+FP+TN)$$

where, TP = True positives (actual seizures), TN = True negatives (actual normal), FP = False positives (normal classified as seizure) and FN = False negatives (seizure frames misclassified as normal).

5. RESULTS AND DISCUSSION

Entropy and energy were calculated on the original and wavelet transformed signals. The feature values were averaged to give row vector. The procedure was repeated for all 23 frames. Four-fold cross validation technique was used where at a time 5 out of 23 segments represented the test set.
To visualize the features ability to discriminate between the two classes, the feature set was reduced to 2-D pattern using principal component analysis. The two-dimensional dataset has been plotted in Figure 3. It clearly differentiates the feature data points between normal (circle) and seizure (star) EEG segments.

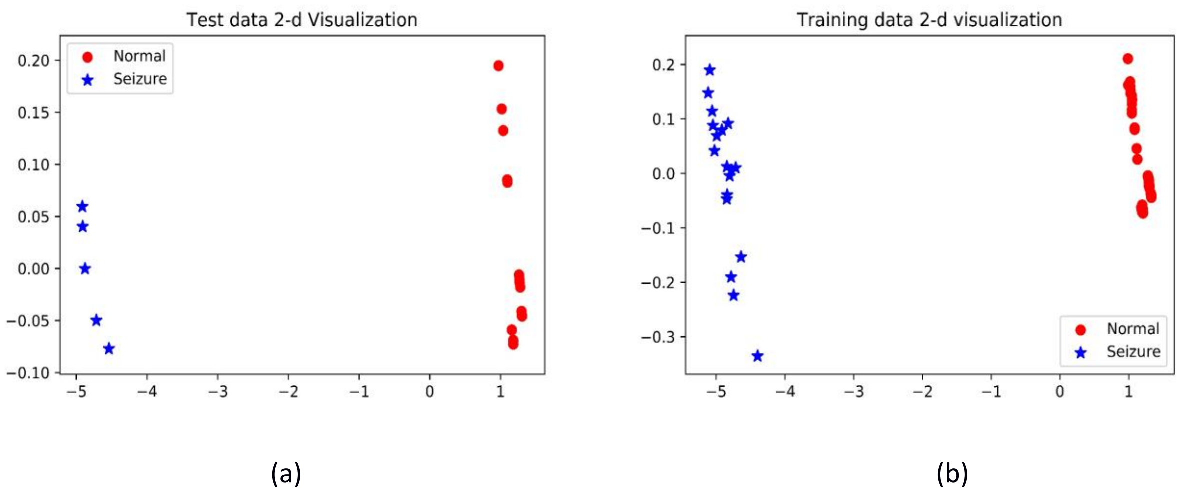

Figure 3: Two dimensional representation of the features on (a) Test set (b) Training set.

Therefore, the data was discriminated between normal and seizure segments. For this, different combinations of datasets (A and E), (AB and E), (C and E) and (ABCD and E) were utilized to compare EEG recordings.
Table 1 lists the results obtained for these different combinations. It shows 100% classification performed by the proposed algorithm on all the combinations of data sets.

Table 1: Classification result using quadratic classifier

Combinations Combination	**TP** | **FN** **FP** | **TN**	**Sensitivity**	**Specificity**	**Accuracy**
A vs E	5 | 0 0 | 5	100	100	100
AB vs E	10 | 0 0 | 5	100	100	100
C vs E	5 | 0 0 | 5	100	100	100
ABCD vs E	20 | 0 0 | 5	100	100	100

The obtained results were compared with the earlier reported work done on the same data sets. D. Gajic at el [9] used energy, entropy, and standard deviation on the wavelet coefficients and reported an accuracy of 99%. Ling Han at el [12] also proposed a wavelet transform based method and reported an accuracy of 96.7%. Standard statistical features were used by Badeeuzzaman at el [11] and obtained 85.2% accuracy for a two class problem. Fatima at el [13] took up dispersion measures and recorded accuracy of 96.9%. Ubeyli [14] used wavelet coefficients and came up with an accuracy of 94.8%. When considering the minimum computations involved in feature extraction and the use of rather less complex linear classifier for classification, the obtained accuracy of 100% can be considered to perform better with the other compared methods. The comparative result is shown in Table 2.

Table 2: Comparison of results from other researchers on the same database.

Author	Features	Sensitivity	Specificity	Accuracy
D. Gajic [9]	Energy, entropy, and standard deviation on wavelet coefficients	98.0%	99.0%	99.0%
Ling Han [12]	wavelet transform based on sample entropy	95.7%	96.1%	96.7%
Badeeuzzaman [11]	Statistical Features	85.3%	85.2%	85.2%
Fatima [13]	Dispersion Measures	94.0%	99.8%	96.9%
Ubeyli [14]	Wavelet coefficients	94.5%	96.0%	94.8%
This work	Approx. Entropy & Energy	**100%**	**100%**	**100%**

6. CONCLUSION

The study exhibits that a minimum set of features extracted from the original and wavelet transformed signal can be used to detect seizures. The extracted features were classified by K-means clustering technique.

However, the data used was small and preprocessed. The results suggested are therefore, preliminary and satisfactory. The same methodology will be extended on a bigger database without any pre-processing for checking the robustness of the proposed methodology.

7. REFERENCES

[1] R. Andrzejak, L. K, M. F, R. C, D. P and E. C, "Indication of non- linear deterministic and finite-dimensional structures in time series of brain elecrical activity: Dependence on recording region and brain state.," Physical Review, vol-64, 2011.

[2] Haykin and V. Veen, ""Signals and System"," John Wiley & Sons Ltd, 2002.

[3] Gotman J and Gloor P, "Automatic recognition and quantification of interictal epileptic activity in the human scalp EEG.," Electroencephelography and Clinical Neurophysiology, 1976.

[4] Gotman J. , "Automatic recognition of epileptic seizure in the EEG," Electroencephelograpy and Clinical Neurophysiology, 1982.

[5] Gotman J and Khan, "Automatic seizure detection: Improvements and evaluation," Electroencephelography and Clinical Neurophsiology, 1990.

[6] Que. H and Gotman J, "A patient-specific algorithm for the detection of seizure onset in long-term," Possible use as a warning device, IEEE Trans Biomed Eng 44(2):115, 1997.

[7] Khan Y. U. and Gotman J., "Wavelet based automatic seizure detection in intracerebral electroencephelogram," Electroencephelography and Clinical neurophsiology vol, 114, pp 898-908, 2003.

[8] Gigola. S., Ortiza. F, Attellisa C. E. and Silva. W., "Prediction of epileptic seizures using accumulated energy in a multiresolution framework," J Neurosci Methods 38:107, 2004.

[9] D. Gajic, Z. Djurovic, S. Di Gennaro and Fredrik G, "Classification of EEG signals for detection of epileptic seizures based on wavelets and statistical pattern recognition," World Scientific Publishing Co. , 2014.

[10] Khan and Rafiuddin, "Feature extraction and Classification of EEG for Automatic seizure detection," International conference on Multimedia Signal processing and communicatioin Technologies, Aligarh, 2011.

[11] Fathima, Badeeuzzaman M and Farooq O, "Wavelet based features for epeliptic Seizure Detection," MES journal of technologies and management, 2011.

[12] L. Han, H. Wang, L. Cong and L. Chunsheng , "Epileptic Seizure Detection Using Wavelet Transform Based Sample Entropy and Support Vector Machine," Proceeding of the IEEE , 2012.

[13] F. T, Khan Y. U, Badeeuzzaman K and Farooq M, "Discriminant analysis for epileptic seizure detetion," Proc. International conference on devices and communication, 2011.

[14] U. E and D, "Combined neural network model employing wavelet coefficients for EEG signals classifcation.," Digital signal processing, 2009.

[15] A. R, Lehnertz K, Mormann F, Rieke C, David P and Elger C, "Indication of non- linear deterministic and finite-dimensional structures in time series of brain elecrical activity: Dependence on recording region and brain state.," Physical Review, vol-64, 2011.