Analysis of EEG Signals for Detection of Epileptic Seizure Using Hybrid Feature Set

Ammama Furrukh Gill, Syeda Alishbah Fatima, M. Usman Akram, Sajid Gul Khawaja and Saqib Ejaz Awan

Abstract Epileptic Seizures occur as a result of certain electrical action in the brain. This makes the patient behave abnormally for a limited amount of time. The electrical activity can be measured with the help electrodes attached to different areas of the scalp to capture the EEG signals. Usually, the signals from the aforementioned device are interpreted by the specialists who specialize in this very thing but their detection is susceptible to errors which prove fatal in some cases. This paper provides an automated system which will detect epileptic seizure without involving an expert opinion. The proposed system goes through a four step process i.e. pre-processing, where the data is organized to suit the system processing and noise is removed. Then temporal and spectral feature extraction is performed. The system then applies the feature selection procedure to extract best set of features which are finally passed to the next phase for classification of EEG signals as normal or abnormal. The suggested system is established on a publicly open dataset and provides an average accuracy of 86.93 %.

A.F. Gill (✉) · S.A. Fatima · M. Usman Akram · S.G. Khawaja · S.E. Awan
College of Electrical and Mechanical Engineering, National University of Sciences and Technology, Rawalpindi, Pakistan
e-mail: ammamagill@live.com

S.A. Fatima
e-mail: syeda.alishbah@gmail.com

M. Usman Akram
e-mail: usmakram@gmail.com

S.G. Khawaja
e-mail: sajid.gul.2009@gmail.com

S.E. Awan
e-mail: saqib.ejaz@yahoo.com

H.A. Sulaiman et al. (eds.), *Theory and Applications of Applied Electromagnetics*, Lecture Notes in Electrical Engineering 344,
DOI 10.1007/978-3-319-17269-9_6

1 Introduction

Human brain continuously generates electrical signals which are transmitted afterwards to different parts of the body in order to achieve several physical, semi-physical and non-physical activities [1]. These electrical signals are prone to multiple disorders which then leads to various human impairments. Epilepsy is ranked as the second most widespread neurological disorder of the human brain after stroke, nearly affecting 1 % of the world's total population [1]. Many irregular discharges of neurons are observed in the brain structures due to this disorder. These discharges arise either during the seizures (referred as the ictal periods) or between the two consecutive seizures (referred as the inter-ictal periods) [2]. The electroencephalogram (EEG) signals are recorded with the help of the electrodes pasted at dissimilar but definite positions over the patient's scalp. This is done to capture information regarding a number of physiological states of the brain as well as body. These signals are first recorded and then scrutinized for any anomaly that may take place in the brain. Seizures either appear as spikes or sharp waves in the EEG readings. The epileptic seizures are detected as partial seizures or widespread seizures. Partial seizures are the seizures where the disorder is detected in only a small number of channels of the EEG recordings while in widespread seizures, the disorder is observed in all the channels. [3] Conventionally epileptic seizures were determined by regularly checking the EEG recordings, which was both tiresome and consumed a lot of time. Also it was exposed to a lot of errors. Diagnostic systems, these days, are equipped with capabilities of automatic recognition of epileptic seizure since they seek the help of computer hence really helpful. Figure 1 shows an EEG signal for all the 23 channel inputs.

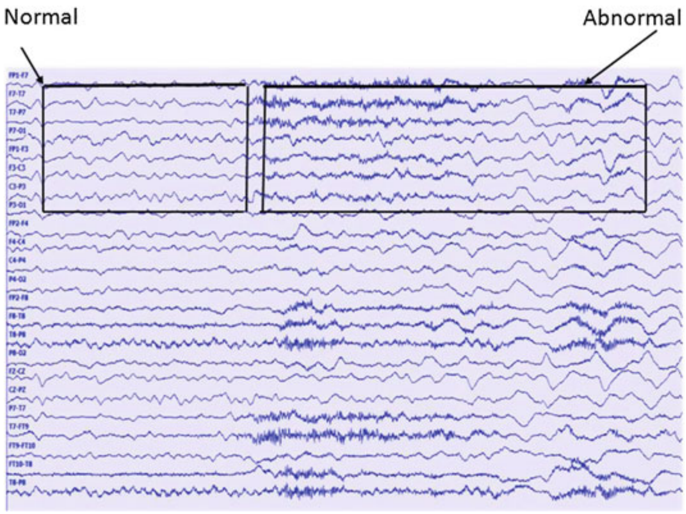

Fig. 1 EEG signal showing part of normal and abnormal waveforms

A variety of methods have been projected for recognition of epileptic seizures. In [4], the authors proposed a multiple stage fuzzy algorithm for the recognition of epileptic seizures. Average amplitude, entropy, rhythmicity (coefficient of variation of amplitude) and dominant frequency are the features used to implement the proposed technique. To combine all the features, a flexible fuzzy sub-system was formulated, the output of this sub-system was used in a threshold procedure to control the final result. iEEG datasets collected out of Freiburg Seizure Prediction EEG (FSPEEG) database were used to test the planned system, which returned an average detection latency of 15.8 s and a false recognition rate of 0.26 per h.

Sharanreddy et al. [5], presented a novel technique, combining multi-wavelet transform, feed forward artificial neural network and enhanced approximate entropy, to facilitate signal classification for epilepsy seizure detection. EEG signal was first converted from the EDF (European Data Format) to ASCII (American Standard Code for Information Interchange) format, which was then provided as input to MWT (Multi-Wavelet Transform). Irregularities were then determined in the brain signal, using IApE (Improved Approximate Entropy Method). IApE was then trained using a feed forward neural network (FFNN). CHB-MIT Scalp EEG Database was used as the testing database. Approximate accuracy of 90 % was achieved. Yet another technique for the automated detection of seizures was given in [6]. It covered two statistical features: skewness and kurtosis, and normalized coefficient of variation (NCOV): a feature from the wavelet background and a meek linear classifier. The proposed scheme was tested, using data for 10 patients, from the CHB-MIT scalp EEG database. An average latency of 3.2 s was detected, with a mean false recognition rate of 1.1 (per hour).

In [7], Ali Shoeb et al. assessed a machine learning approach to produce classifiers with reference to the patients. The features in this study include rhythmic activity, channel identity and a non-EEG feature (such as ECG) to establish the final output. The approach was tested on the CHB-MIT database and yielded a median false recognition rate of 2 false detections per 24 h.

This article consists of four sections. Section 2 describes an overview and details regarding all the steps of the recommended system. The results are presented in Sect. 3. Section 4 carries the conclusions.

2 Proposed System

The recommended system comprises of 4 steps including the preprocessing, feature extraction followed by feature selection and classification. In the preprocessing phase, the data from the CHB–MIT Scalp EEG database is arranged in a manner suitable for the further processing. This arranged data is then divided into multiple blocks with 1000 samples in each block. Various temporal and spectral domain features are extracted from the data to form a hybrid feature set. Rank sum tests are then applied to figure out the best and most discriminating features. Then these

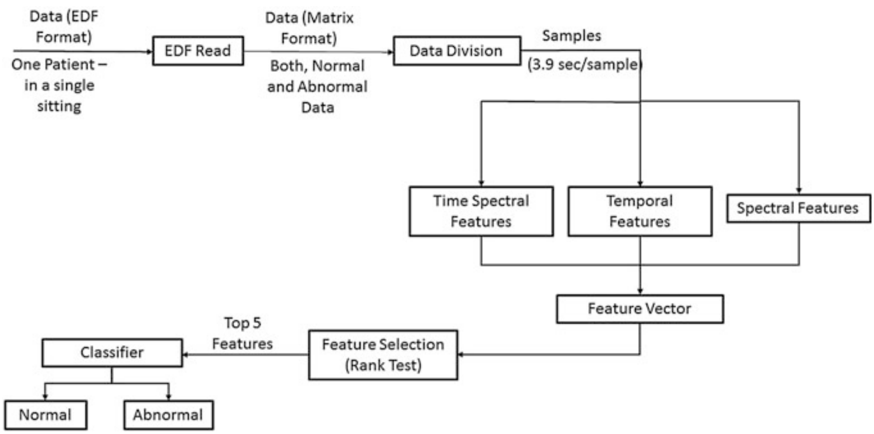

Fig. 2 Flow diagram of proposed system

features are fed to a classifier for the recognition of epileptic seizures. Figure 2 shows the flow diagram representing the recommended system.

2.1 Preprocessing

The dataset contained the CHB—MIT Scalp EEG data in European Data Format (EDF) [8]. This format is often used for multi-channel physical and biological signals. The data is collected from multiple channels for a lengthy timed durations, may be over 60 min, for most of the cases of a single sitting. The preprocessing step removes the noise present in the signal and arranges the data to further processing. The de-noised data is then divided into chunks one thousand samples, 3.9 s long in time. Then the patient summary is reviewed les that contain information regarding the events occurred during that particular EEG sitting of the respective patient. The blocks of data containing seizures are labelled abnormal while the ones which do not possess seizure information are labelled as normal.

2.2 Feature Extraction

The proposed system consists of a hybrid feature set which is a combination of multiple time domain, spectral domain and time spectral features. The details of features are

- Temporal -Domain Features

    1. Entropy (f_1) is calculated a first feature which is calculated in time domain to capture the variation in signal due to presence of seizure [9].

2. Mean (f_2) represents the average value of all samples present in selected window.
3. Harmonic Mean (f_3) is third feature calculated in time domain to suppress the effect of large values [10].
4. Range (f_4) is calculated to have an idea about the difference between maximum and minimum values.
5. Inter Quartile Range (f_5) [11] is calculated as a feature to have good idea about the range even in the presence of outliers as it only considers the difference between first and third quintiles.
6. Mean Absolute Deviation (f_6) [12] is the average distance of all samples from data mean present in the window.
7. Moment (f_7) calculates the central moment of the signals under study.
8. Skewness (f_8) [13] is calculated to have an idea about the distribution of sample along overall mean.
9. Kurtosis (f_9) [14] determines if the data, under study, is tall and thin or short and squat when compared to the normal distribution.
10. Percentile (f_{10}) indicates the value below which a given percentage of values in a group of values fall.
11. Gradient (f_{11}) points towards the greatest rate of increase in the under study data points.

- Time—Spectral Features

1. Wavelet Transform [15] is the type of a time-frequency transform which follows the basic idea of changes in time-extension without altering its overall shape. The wavelet decomposition vector (f_{12}) and the book keeping vectors (f_{13}) are formed by choosing the appropriate basis functions. A multi-level one dimensional wavelet decomposition is performed using 'db1' as the shape of the wave to work upon.

- Spectral—Domain Features

1. Wavelet Energy [15] calculates the energy for a 1-D wavelet packet decomposition. The percentage of energy which corresponds to the approximation (f_{14}) and the percentage of energy corresponding to the details (f_{15}) of the outputs of the wavelet transform feature are calculated to comprehend the energy details of the wavelet decomposition, calculated prior to this. Pseudo spectrum [16] refers to a set that contains the spectrum of the operators and the numbers which are almost Eigen values.
2. The pseudo spectrum estimate (f_{16}) of the input data and the normalized frequencies (f_{17}), where the estimate is calculated, are found. These values are calculated by using the correlation matrix of the data under study and utilizing its Eigen vector estimates.
3. Fast Approximate Entropy (f_{18}) [9] calculates the amount of irregularity over the time-series of a given data set.

Table 1 Feature selection based on rank tests

Wilcoxon Test			Ansari Bradley Test		
Features	P Value	Score	Features	P Value	Score
f18	$<10^6$	5.47	f7	$<10^6$	5.02
f16	$<10^6$	6.67	f3	$<10^6$	5.86
f12	$<10^6$	9.27	f9	$<10^6$	6.13
f1	$<10^6$	9.4	f18	$<10^6$	7.05
f3	$<10^6$	10.04	f14	$<10^6$	11.8
f10	$<10^6$	10.38	f15	$<10^6$	11.8
f11	$<10^6$	12.06	f4	$<10^6$	22.58
f6	$<10^6$	12.24	f16	$<10^6$	23.02
f5	$<10^6$	15.4	f1	$<10^6$	24.9
f7	$<10^6$	18.56	f12	$<10^6$	26.21
f2	0.0000261	3.94	f5	$<10^6$	27.15
f4	0.0000804	4.2	f8	0.0000472	4.07
f8	0.0000864	3.33	f10	0.011	2.54
f9	0.0238	2.26	f6	0.025	2.24
f14	0.62	0.5	f2	0.67	0.43
f15	0.62	0.5	f11	0.8747	0.16
f17	0.9	0.2	f13	0.89	0.11
f13	0.95	0.15	f17	0.91	0.06

2.3 *Feature Selection*

The feature selection process extracts those features which possess the maximum discriminating power out of the set of available features. This process increases the quality of the proposed mechanism. In this workout, we carried out two types of statistical tests: Wilcoxon signed—rank test [17] and the Ansari-Bradley test [18]. Table 1 depicts the results of Rank Sum tests over all features. Best five features based on lowest P value and best scores in both rank tests are selected. Final feature vector consists of entropy (f_1), central moment (f_7), wavelet de-composition vector (f_{12}), pseudo spectrum estimate (f_{16}) and fast approximate entropy (f_{18}).

2.4 *Classification*

Gaussian mixture models (GMM) were used to classify the data in this paper. This technique is based on Bayesian decision rule [19]. Seventy percent of the whole dataset constitutes the training data, with two class labels X_1 = Abnormal and X_2 = Normal. The GMM parameters are optimized using Expectation Maximization (EM). It uses an iterative procedure to find optimal values for each parameter [21].

3 Experimental Results

A publicly available EEG signal database for seizure detection is used for proper evaluation of proposed system. The CHB–MIT Scalp EEG database [7] served as the dataset for all the experimentation purposes and the data is sampled at a frequency of 256 Hz. The dataset consists of data from 12 patients suffering from different levels of seizure. In order to perform experiments properly, the data is broken into small segments of 1000 samples which is roughly of 3.90 s length. To avoid biasness in training and cross validation phase, equal number of normal and seizure samples are used for training.

Different confusion matrix based performance parameters are calculated for proper evaluation of proposed system. Equations 1–3 show the relations for these parameters such as sensitivity (TPR), specificity (TNR) and overall accuracy (Acc). These parameters were calculated using Eqs. 1–3 respectively.

$$\mathrm{TPR} = \frac{\mathrm{TP}}{(\mathrm{TP} + \mathrm{FN})} \tag{1}$$

$$\mathrm{TNR} = \frac{\mathrm{TP}}{(\mathrm{TN} + \mathrm{FP})} \tag{2}$$

$$\mathrm{Acc} = \frac{\mathrm{TP} + \mathrm{TN}}{(\mathrm{TP} + \mathrm{TN} + \mathrm{FN} + \mathrm{FP})} \tag{3}$$

where

- TP are correctly classified seizure segments out of all segments;
- TN are correctly classified normal segments out of all segments;

Table 2 Performance evaluation results of proposed system

Patient	Accuracy	Sensitivity	Specificity
1	95.3	94.1	95.7
2	93.4	93.9	93.1
3	90.1	88.3	92.3
4	80.1	79.9	80.7
5	99.6	99.9	99.3
6	87.3	86.8	88.9
7	90.4	91.3	89.3
8	79.5	77.2	82.3
9	95.7	94.3	95.9
10	75.9	75.3	76.3
11	79.2	78.7	80.1
12	76.7	75.4	77.1
Averaged	86.93	86.26	87.58

- FP are those segments which are wrongly classified as seizure however they actually belong to normal class;
- FN are those segments which are wrongly classified as normal however they actually belong to seizure class;

The evaluation results of our system for each patient are given in Table 2. It shows that proposed system detects seizure with good accuracy but with exception for those patients who have mild level of seizure.

4 Conclusion

Digital Signal Processing techniques have always proven to be very crucial to the automation of seizure detection. Research institutes have worked exhaustively to come up with methodologies targeting this domain. The proposed system presents a novel technique that focuses on automating seizure detection, under-going four distinct steps: Preprocessing, feature extraction, feature selection and classification. Average values of sensitivity, specificity and accuracy are 86.26, 87.58 and 86.93 % respectively as obtained with this technique. The proposed system has been shown valid since it has automatically detected a number of seizures in the dataset.

References

1. Alexandros, T.T.: Epileptic seizure detection in EEGs using time-frequency analysis. IEEE Trans. Inf. Technol. Biomed. **13**(5), 703–710 (2009)
2. Exarchos, T.P., Tzallas, A.T., Fotiadis, D.I., Konitsiotis, S., Giannopoulos, S.: EEG transient event detection and classification using association rules. IEEE Trans. Inf. Technol. Biomed. **10**(3), 451–457 (2006)
3. Gotman, J.: Automatic detection of seizures and spikes. J. Clin. Neurophysiol. **16**(2), 130–140 (1999)
4. Rabbi, A.F., Fazel-Rezai, R.: A fuzzy logic system for seizure onset detection in intracranial EEG. Comput. Intell. Neurosc. **2012**, Article ID 705140, 12 pp (2012). doi:10.1155/2012/705140
5. Sharanreddy, M.A., Kulkarni, P.K.: EEG Signal classification for epilepsy seizure detection using improved approximate entropy. Int. J. Public Health Sci. (IJPHS) **2**(1), 23–32 (2013). ISSN: 2252–8806
6. Khan, Y.U., Farooq, O., Sharma, P.: Automatic detection of seizure onset in pediatric EEG. Int. J. Embed. Syst. Appl. (IJESA) **2**(3), 81–89 16 pp (2012)
7. Shoeb, A.H., Guttag, J.V.: Application of machine learning to epileptic seizure detection. In: Proceedings of the 27th International Conference on Machine Learning (ICML-10), pp. 975–982 (2010)
8. Kemp, B., Varri[b], A., Rosa[c], A.C., Nielsen[d], K.D., Gade, J.: A simple format for exchange of digitized polygraphic recordings, Electroencephalogr. Clin. Neurophysiol. **82**, 391–393 (1992)
9. Kannathal, N., Choo, M.L., Acharya, U.R., Sadasivan, P.K.: Entropiesfor detection of epilepsy in EEG. Comput. Methods Prog. Biomed. **80**, 187–194 (2005)

10. Mitchell, Douglas W.: More on spreads and non-arithmetic means. The Mathematical Gazette **88**, 142–144 (2004)
11. Upton, G., Cook, I.: Understanding Statistics. Oxford University Press, Oxford, p. 55 (1996)
12. Kader, G.: Means and MADS. Math. Teach. Middle Sch. **4**(6), 398–403 (March 1999). Retrieved 20 Feb 2013
13. Dean, S., Illowsky, B.: Descriptive Statistics: Skewness and the Mean, Median, and Mode
14. Ballanda, Kevin P., MacGillivray, H.L.: Kurtosis: A Critical Review. Am. Stat. (American Statistical Association) **42**(2), 111–119 (1988)
15. Adeli, H., Zhou, Z., Dadmehr, N.: Analysis of EEG records inan epileptic patient using wavelet transform. J. Neurosci. Methods **123**, 69–87 (2003)
16. Marple, S.L.: Digital Spectral Analysis. Prentice-Hall, Englewood Cliffs, pp. 373–378 (1987)
17. Lowry, R.: Concepts and Applications of Inferential Statistics. Retrieved 24 March 2011
18. Bauer, D.F.: Constructing con dence sets using rank statistics. J. Am. Stat. Assoc. **67**, 687–690 (1972)
19. Akram, M.U., Tariq, A., Anjum, M.A., Javed, Y.: Automated detection of exudates in colored retinalimages for diagnosis of diabetic retinopathy. Appl. Opt. **51**(20)
20. Theodoridis, S., Koutroumbas, K.: Pattern Recognition, 1st edn. Academic, New York (1999)
21. Duda, R.O., Hart, P.E., Stork, D.G.: Pattern Classification. Wiley, New York (2001)