This full-text paper was peer-reviewed and accepted to be presented at the IEEE WiSPNET 2016 conference.

Performance Analysis of KNN Classifier and K-Means Clustering for Robust Classification of Epilepsy from EEG Signals

M.Manjusha
Department of ECE
Bannari Amman Institute of Technology
Sathyamangalam, India

R.Harikumar
Department of ECE
Bannari Amman Institute of Technology
Sathyamangalam, India

***Abstract*— Epilepsy is a neurological disorder which affects persons of all age. The brain waves are studied for epilepsy detection. The Electroencephalogram (EEG) is the simplest diagnostic technique available for brain wave analysis. In this paper, we investigate the performance of KNN classifier and K-means clustering for the classification of epilepsy risk level from EEG signals. To identify the non linearity present in the data, detrend analysis is done. An EEG record of twenty patients is analyzed. The power spectral density is determined which is further used for dimensionality reduction. The performance index achieved by KNN classifier and K-means clustering are 78.31% and 93.02% respectively. A high Quality value of 22.37 with K-means clustering and a low value of 18.02 are obtained with KNN classifier. The results show that K-means outperforms KNN classifier in epilepsy risk level classification.**

Keywords— EEG signal; Epilepsy; Power spectral density; Detrend fluctuation analysis, KNN classifier; K-means clustering.

I. INTRODUCTION

Epilepsy is a fatal neurological disorder which causes interruption of the normal neuronal activity [1][2]. During an epileptic attack, the nerve cell continues to fire causing seizures. The exact cause for epileptic seizures is not precisely known. Some of the conditions leads to seizures are brain damage, stroke, illness and brain tumors [3]. Seizures are broadly classified into two categories namely generalized seizures and partial seizures [4]. Generalized seizures involve the whole brain activity hammering while partial seizures involve only a specific part of brain [5]. The most common symptoms of seizures are convulsions, loss of consciousness and confusion. The normal rhythm of the EEG wave is altered when a seizure attack happens [6].

The visual inspection of the brain signals provides a valuable diagnostic technique for epilepsy detection. Brain wave analysis can be performed by Electroencephalogram (EEG) [7][8][9][10]. Other clinical tools available are Magneto encephalogram (MEG) and functional Magnetic Resonance Imaging (fMRI). EEG is the most employed technique for seizure diagnosis [11].During epileptic seizure the normal electrical activity of brain is disrupted or altered which can be easily extracted with the help of EEG. It uses a set of electrodes placed over scalp. This electrode arrangement follows standard 10-20 system, adopted by the American EEG society. The signals acquired contain enormous information which can be decoded to predict seizure [12].

The further organization of this paper is as follows. Section II reviews the main approaches applied for classification of epilepsy from EEG signals. Section III combines the result for comparison between two approaches and discussion is done. Section IV describes our conclusion.

II. MATERIALS AND METHODS

A. EEG Data Acquisitiom

In this work, we have used the EEG data set which has been acquired from twenty epileptic patients who were under examination and medication in the Neurological Department of Sri Ramakrishna Hospital, Coimbatore. A 16 channel clinical EEG monitoring system, employing 10-20 interactive electrode placement method is used for obtaining the paper record of EEG. This raw EEG signals recorded are contaminated by artifacts. Major sources of artifacts are biological artifacts, contact impedance from electrodes, and other electronic instruments in the room. With the assistance of Neurologist, the artifact free EEG signals are selected and scanned by Umax 6696 Scanner with a resolution of 600 dpi.

Using five pole Butterworth filter, the EEG signals are band passed between the frequency range of 0.5Hz and 50Hz. The recorded EEG signals are continuous value of thirty second, they are converted into epochs of two seconds. This two second epochs are enough for identify any neurological changes. The maximum frequency of EEG signals is 50 Hz hence each epoch is sampled at 200 Hz. Each sample provides instantaneous amplitude of the EEG signals recorded. There are 400 values for each epoch which contributes to large dimension of data value. To eradicate this, dimensionality reduction is done.

B. Dimensionality Reduction

The preprocessing step is to reduce the dimension of the EEG data. The most conventional dimensionality reduction methods are Singular value decomposition (SVD), Principal component analysis (PCA), and Independent component analysis (ICA) [2][13]. The drawback of these techniques is that they cannot handle complex and non linear data [14]. In this work, we have used power spectral density (PSD) for the dimensionality reduction purpose. The power spectral estimate of the entire EEG set is determined.

The power spectral density provides the information about the frequency distribution of EEG signals. It provides only the magnitude information which is available in the frequency domain and does not provide any knowledge about the phase values. The power spectral estimate of signal x(t) is given below and it is denoted by P(f)

$$P(f) = Re^2\,[X(f)] + Im^2\,[X(f)] \qquad (1)$$

where X(f) is the Fourier transform of signal x(t) [15].

C. Detrend Fluctuation Analysis

Most of the biological signals measured are non linear in nature due to complexity of the system being measured [16]. To examine the non-linearity present in the epileptic EEG data detrend fluctuation analysis (DFA) is performed. The EEG signal recorded from scalp of the patient arises from non linear interactions of neurons. [17].The scaling exponent α is calculated for the EEG dataset. The autocorrelation properties of a signal are given below [18]:

1. $\alpha < 0.5$ anti-correlated signal
2. $\alpha = 0.5$ uncorrelated signal (white noise)
3. $\alpha > 0.5$ positive correlation in the signal

The value obtained for the parameter α is tabulated below.

TABLE I SCALING EXPONENT VALUE FOR EEG SIGNALS

Patient number	Scaling exponent α
1	0.1780
2	0.1610
3	0.1585
4	0.1978
5	0.1745
6	0.2066
7	0.2608
8	0.1473
9	0.1908
10	0.1137
11	0.2133
12	0.1861
13	0.1265
14	0.1693
15	0.1127
16	0.1451
17	0.2021
18	0.2248
19	0.1664
20	0.1610

From the analysis it is evident that the EEG signals are anti-correlated in nature. The average value of scaling exponent achieved is 0.1748

D. KNN Classifier

The KNN classifier is a non parametric instance based classifier [19]. It is a lazy learning method which does not learn from training data, simply stores all the samples in the training data [20]. These stored values are needed during the training phase. This algorithm is based on the nearest neighborhood estimation. The new cases are classified on the basis of similarity measure which is the distance metric. Most commonly used is Euclidean distance. Drawback of KNN classifier is large time required to find nearest neighborhood in a large training set [21]. Hence dimensionality reduction step is done to overcome this.

In KNN classifier, the class of x is found by following procedure.

a) Determine the k instances which are nearest to the class x based on the distance measure.
b) The next step is to allow this k instances to vote to find the class of x.

Here the number of closest neighborhood instances selection is crucial step in this algorithm which affects the overall performance of classifier.

In this work, we applied the KNN classifier to the dimensionality reduced power spectral density values. The result obtained is compared with the target value for mean square error estimation. The various performance parameters are calculated to analyze the efficiency of KNN classifier in this study.

E. K-Means clustering

K-Means clustering is one of the most popular partitional clustering algorithms [22]. It is a non deterministic and unsupervised partitioning method [23]. This algorithm clusters the given set of data into different subsets based on some criteria. The objective of the algorithm is to maximize the similarity between the patterns in same cluster while to minimize between different clusters [24]. It is a fast and robust method of clustering. Choice of initial centriods and number of clusters greatly influence the performance of the K- means clustering [25]. The k value must be varied over a range to get optimum results.

The outline of basic K-means algorithm is given below.

a) Choose K points as the initial centriods.
b) Determine distance between each data point and the cluster centriod.
c) Reassign data point to new cluster with minimum distance.
d) Recomputed the new cluster centriods.
e) Repeat till convergence criteria is meet.

The sum of squared error is the objective function employed by K-means method. The mathematical representation of objective function is given below.

$$SSE = \sum_{i=1}^{c} \sum_{j=1}^{c_i} (\| X_i - Y_j \|)^2 \tag{2}$$

where,
'$\| X_i - Y_j \|$' is the Euclidean distance between X_i and Y_j; 'C_i' is the total number of pattern points in i[th] cluster; 'C' is the number of cluster centroids. The results of K-means clustering algorithm applied to the EEG data set are discussed below.

III. RESULTS AND DISCUSSION

The relative performance of KNN classifier and K-means clustering are studied through calculating various parameters namely Performance index, Sensitivity, Specificity, Average detection and Quality value [26]. With K-means clustering, we have obtained the performance index of 93.02%, sensitivity of 100% , specificity of 93.75% and quality value of 22.37.The Quality value obtained with KNN classifier is less when compared to that of K-means clustering. The results obtained are tabulated below.

The formulae for obtaining above specified values are given below

$$PI = \frac{PC - MC - FA}{PC} X\,100 \tag{3}$$

$$Se = \frac{PC}{PC + MC} X\,100 \tag{4}$$

$$Sp = \frac{PC}{PC + MC} X\,100 \tag{5}$$

where, PI - Performance index, PC - Perfect classification, MC - Missed Classification, FA - False alarm, Se - Sensitivity, Sp – Specificity.

To assess the overall efficiency of classifier quality value estimation is done. The Quality value Q_V is defined as below.

$$Qv = \frac{C}{(R_{fa} + 0.2) * (T_{dly} * P_{dct} * 6 * P_{msd})} \tag{6}$$

where, C is scaling constant, R_{fa} is the number of false alarm per set, T_{dly} is the average delay of the onset classification in seconds, P_{dct} is the percentage of perfect classification, P_{msd} is the percentage of perfect risk level missed[27]. The results obtained are tabulated below.

TABLE III PERFORMANCE COMPARISON OF KNN CLASSIFIER AND K-MEANS CLUSTERING

Performance Parameters	KNN Classifier	K-Means Clustering
Perfect Classification (%)	83.3325	93.7540
Missed Classification (%)	7.1485	6.1077
False Alarm(%)	9.5029	0
Performance Index (%)	78.3194	93.0242
Sensitivity (%)	90.4878	100
Specificity (%)	92.8475	93.7540
Time delay (seconds)	2.1097	2.2427
Quality Value	18.0206	22.3728

It is evident from above results that K-means clustering outperforms KNN classifier in epilepsy risk level classification.

The various performance plots are drawn below.

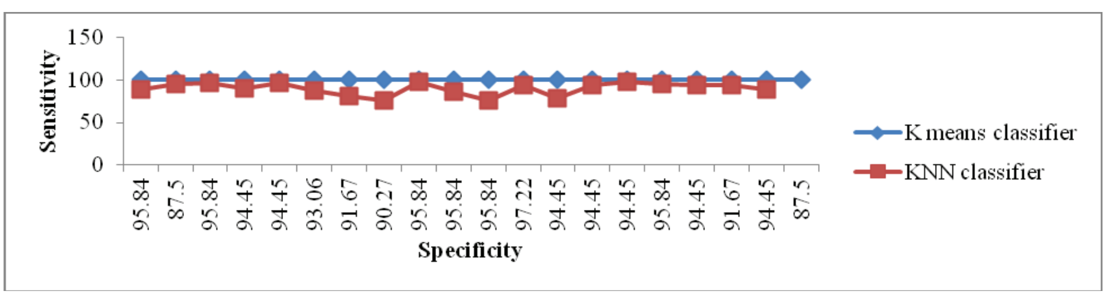

Fig. 1. Sensitivity and Specificity measures of KNN classifier and K means clustering

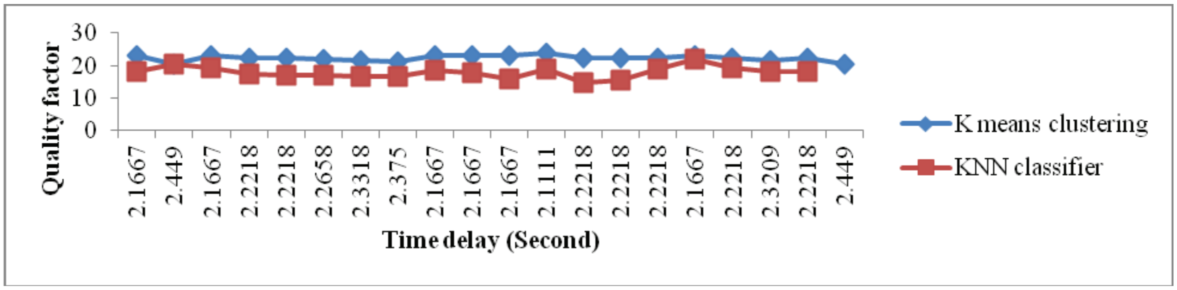

Fig. 2. Quality factor and time delay measures of KNN classifier and K means clustering

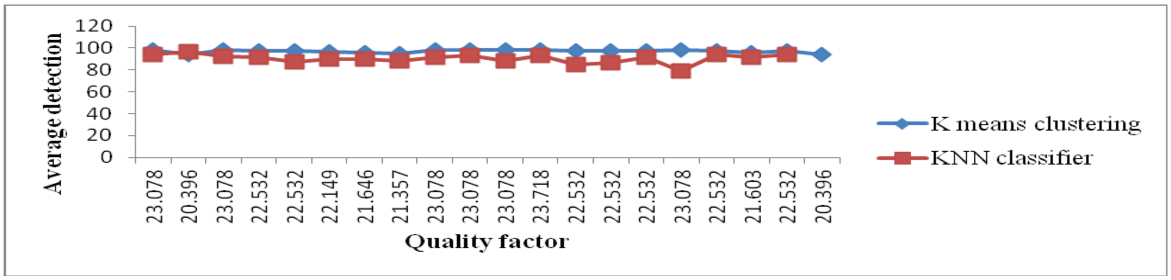

Fig. 3. Average Detection and Quality Value measures of KNN classifier and K means clustering

IV. CONCLUSION

This paper investigates the performance of KNN classifier and K- means clustering in classifying the epilepsy risk level of epileptic patients from EEG signals. The parameter estimated here is the power spectral density of the EEG signals and dimensionality reduction is carried out for these spectral values. Finally classification is done by applying KNN classifier and K-means clustering. Also the fluctuations present in the signals are studied. The aim was to design classification algorithm with high performance index and minimum false alarm rate and missed classification. We have obtained better performance with K- means clustering with false alarm rate of 0%. Thus applying this technique risk level of epilepsy can be identified and proper medication can be given to the patients.

REFERENCES

[1] Ye Yaun, “Detection of epileptic seizure based on EEG signals”, *2010 3rd International Congress on Image and Signal Processing*, vol. 9, pp. 4209 – 4211, October 2010.

[2] Dr.R.Harikumr, C. Ganeshbabu, M. Balasubramani, G.A. Nivhedhita, “Performance Analysis of Extreme Learning Machine for Robust Classification of Epilepsy from EEG Signals”, *IJCA Proceedings on International Conference on Innovations in Information, Embedded and Communication systems, 2014, International Journal of Computer Applications* (0975 – 8887).

[3] Ashwini D.Bhople, P.A.Tijare, “Performance Evaluation of Different Neural Networks Used for Seizure Detection” *International Journal of Computer and Organization Trends*. vol. 1, Issue. 3,2011.

[4] Sunil Kumar Prabhakar, Harikumar Rajaguru “ICA, LGE and FMI as Dimensionality Reduction Techniques followed by GMM as Post Classifier for the Classification of Epilepsy Risk Levels from EEG Signals”, *2015 IEEE European Modelling Symposium*.

[5] Amir B. Geva, Dan H. Kerem, “Forecasting Generalized Epileptic Seizures from the EEG Signal by Wavelet Analysis and Dynamic Unsupervised Fuzzy clustering”, *IEEE Transactions on Biomedical Engineering*, vol. 45, No. 10, October 1998.

[6] Alison A. Dingle, Richard D. Jones, Grant J. Carroll, W. Richard Fright, “A Multistage System to Detect Epileptifrom Activity in the EEG”, *IEEE Transactions On Biomedical Engineering*, vol. 40, No. 12, December 1993.

[7] Niederhauser JJ, Esteller R, Echauz J, Vachtsevanos G, Litt B. Detection of seizure precursors from depth-EEG using a sign periodogram transform. *IEEE Trans Biomed Eng 2003*, pp. 449-58.

[8] Balkis Solehah Zainuddin, Zakaria Hussain, Iza Sazanita Isa , “Alpha and Beta EEG Brainwave Signal Classification Technique: A Conceptual Study”, *2014 IEEE 10th International Colloquium on Signal Processing & its Applications (CSPA2014), 7 - 9 Mac.,* Kuala Lumpur, Malaysia,2014.

[9] Apeksha Rani H. M, Prathibha Kiran, “A Novel Method for Analysis of EEG Signals using Brain Wave Data Analyzer”, *International Journal of Soft Computing and Engineering (IJSCE)* , vol. 5 Issue. 2, May 2015.

[10] Anders Isaksson, Arne Wennberg, Lars H. Zetterberg, “Computer analysis of EEg sisgnals with Parametric Models”, *Proceedings Of The IEEE*, vol. 69, No. 4, April 1981.

[11] M. Heers, S. Rampp, M. Kaltenhauser, E. Pauli, C.Rauch, M.T. Dolken, H.Stefn, “Detection of Epileptic Spikes by Magnetoencephalography and Electroencephalography after sleep deprivation”, *Seizure*, vol. 19, Issue. 7, pp. 397-403, September 2010.

[12] A.S.Gevins, “Analysis of the electromagnetic signals of the human brain: Milestones, Obstacles, and Goals, “*IEEE Trans. Biomed Eng*, vol. 31, pp. 833 850, 1984.

[13] Arunasakthi. K, KamatchiPriya. L, “A Review On Linear and Non – Liner dimensionality Reduction techniques, Machine Learning an Applications: *An International Journal(MLAIJ)* vol. 1, No. 1, September 2014.

[14] Tian Siva Tian, Gareth M.James, “Interpretable Dimesion Reduction for Classifiying Function Data”,*omputional Statistics and Data Analysis*. vol. 57, Issue. 1, pp. 282 – 296, January 2013.

[15] N.Faud, M.N.Taib, R.Jailani, M.E.Marwan, " Brainwave classification for Brain Balancing Index (BBI) via 3D EEG Model using KNN technique, *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, vol. 8, No. 8, 2014.

[16] Gifani P, Rabiee H.R, Hashemi M.H, "Optimal fractal-scaling analysis of human EEG dynamic for depth of anesthesia quantication", *Journal of the Franklin Institute*, vol. 344, pp. 212-229, 2007.

[17] Linkenkaer-Hansen K, Nikouline.VV, Palva JM,Ilmoniemi RJ, "Long – range temporal correlations and scaling behaviour in human brain oscillations", *The journal of Neuroscience*, vol 21(4), pp.1370-1377, February 2001.

[18] Agnieszka Kitlas Golinska, "Detrended Fluctuation Analysis (DFA) in biomedical signal processing: selected examples", *Studies in Logic, Grammer and Rhetoric* , vol 29, pp. 107-115, 2012

[19] R. Shantha Selva Kumari , J.Prabin Jose, " Seizure detection in EEG using Biorthogonal wavelet and fuzzy KNN classifier", *Elixir Human Physio*. vol. 42, pp. 5766 – 5770,2011.

[20] Sadegh Bafandeh Imandoust, Mohammad Bolandraftar, " Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background", *International Journal of Engineering Research and Applications* vol. 3, Issue.5, pp.605-610, Sep-Oct 2013.

[21] Belur V. Dasarathy, " Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques", *Mc Graw-Hill Computer Science Series, IEEE Computer Society Press*, Las Alamitos, California, pp.217-224,1991.

[22] Ming-Chuan Hung, Jungpin Wu,Jin-Hua Chang, Don-Lin Yang, "An Efficient k-Means Clustering Algorithm using Simple Partioning", *Journal of Informtion Science and Engineering*, pp.1157-1177,2005.

[23] Jennifer G.Dy , Carla E. Brodley, " Feature Selection for Unsupervised lerning", *Journal of Machine Learning Research* vol.5 , pp. 845 – 889,2005.

[24] Umut Orhan , Mahmut Hekim , Mahmut Ozer, "EEG signals Classification using the K-means clustering and a multilayer perceptron neural network model", *Expert Systems with Applications* vol.38,pp.13475–13481,2011.

[25] Kalpana D. Joshi, P.S. Nalwade, " Modified K-Means for better initail cluster centres", *International Journal of Computer Science and Mobile Computing*, vol. 2, Issue. 7, pp.219 – 223, July 2013.

[26] Dr.R.Harikumar, T.Vijayakumar, M.G. Sreejith, "Performance Analysis of SVD and K-Means Clustering for Optimization of Fuzzy Outputs In Classification of Epilepsy Risk Level from EEG Signals", *Recent Advances In Intelligent Computational Systems (RAICS), 2011 IEEE*, pp. 718 – 723.

[27] A.Keerthi Vasan, R.Harikumar, M.Logesh Kumar, "Performance Analysis of Support Vector Machine (SVM) for Optimization of Fuzzy based Epilepsy Risk Level Classifications using different Types of kernel Functions from EEG Signal parameters", *Proceedings of the International MultiConference of Engineers and Computer Scientists 2009*,vol 1, IMECS 2009, March 18-20, 2009, Hong Kong.