



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

R.Harikumar, Ph.D

Department of ECE

Bannari Amman Institute of Technology
Sathyamangalam, India

M.Manjusha

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 acquisition. In this paper, we investigate the performance of KNN classifier and K-means clustering for the classification of epilepsy risk level from EEG signals. 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 is 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; 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)[8][9][10]. Other clinical tools available are Magnetoencephalogram (MEG) and functional Magnetic Resonance Imaging (fMRI). EEG is the most employed technique for seizure diagnosis [11]. EEG measures the electrical activity of brain by using a set of electrodes placed over the 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 Acquisition

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 help of Neurologist, the artifact free EEG signals are selected and scanned by Umax 6696 Scanner with a resolution of 600 dpi.

The EEG signals are band pass filtered between 0.5 Hz and 50 Hz using five pole Butterworth filter. 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 $P(f)$

$$P(f) = \text{Re}^2[X(f)] + \text{Im}^2[X(f)] \quad (1)$$

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

C. KNN Classifier

The KNN classifier is a non parametric instance based classifier [16]. It is a lazy learning method which does not learn from training data, simply stores all the samples in the training data [17]. 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 [18]. Hence dimensionality reduction step is done to overcome this.

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

- Determine the k instances which are nearest to the class x based on the distance measure.
- 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. Christo Ananth et al. [7] discussed about Improved Particle Swarm Optimization. The fuzzy filter based on particle swarm optimization is used to remove the high density image impulse noise, which occur during the transmission, data acquisition and processing. The proposed system has a fuzzy filter which has the parallel fuzzy inference mechanism, fuzzy mean process, and a fuzzy composition process.

D. K-Means clustering

K-Means clustering is the most popular partitionial clustering algorithm. It is a non deterministic and unsupervised partitioning method [19]. This algorithm classifies a given set of data into a certain number of clusters

which is predefined. The objective of the algorithm is to maximize the similarity between the patterns in same cluster while to minimize between different clusters [20]. It is a fast and robust method of clustering. Choice of initial centroids and number of clusters greatly influence the performance of the K-means clustering [21]. The k value must be varied over a range to get optimum results.

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

- Choose K points as the initial centroids.
- Determine distance between each data point and the cluster centroid.
- Reassign data point to new cluster with minimum distance.
- Recomputed the new cluster centroids.
- 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.

$$SSE = \sum_{i=1}^C \sum_{j=1}^{C_i} (\|X_i - Y_j\|)^2 \quad (2)$$

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 [22]. With K-means clustering, we have obtained the performance parameter of 93.75%, sensitivity of 93.75%, specificity of 100% and quality value of 22.37. The performance index 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

$$PI = \frac{PC - MC - FA}{PC} \times 100 \quad (3)$$

$$Se = \frac{PC}{PC + MC} \times 100 \quad (4)$$

$$Sp = \frac{PC}{PC + MC} \times 100 \quad (5)$$



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

The overall quality of the classifier is measured by quality value which depends on classification rate, classification delay and false alarm rate. The Quality value Q_v is defined as below.

$$Q_v = \frac{C}{(R_{fa} + 0.2) * (T_{dly} * P_{det} * 6 * P_{msd})} \quad (6)$$

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_{det} is the percentage of perfect classification, P_{msd} is the percentage of perfect risk level missed.

The results obtained are tabulated below

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

Performance Parameters	KNN Classifier	K-Means Clustering
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.

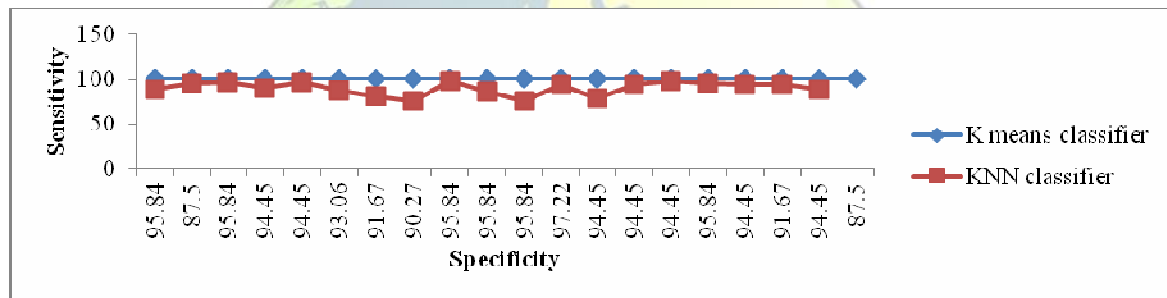


Fig. 1. Sensitivity and Specificity measures of KNN classifier and K means clustering

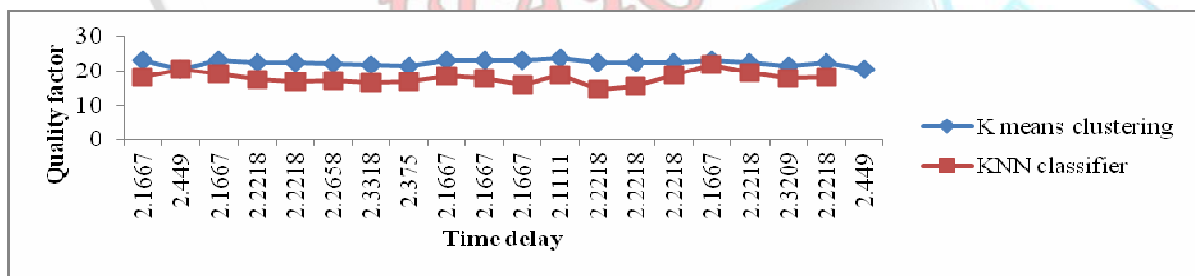


Fig. 2. Quality factor and time delay measures of KNN classifier and K means clustering

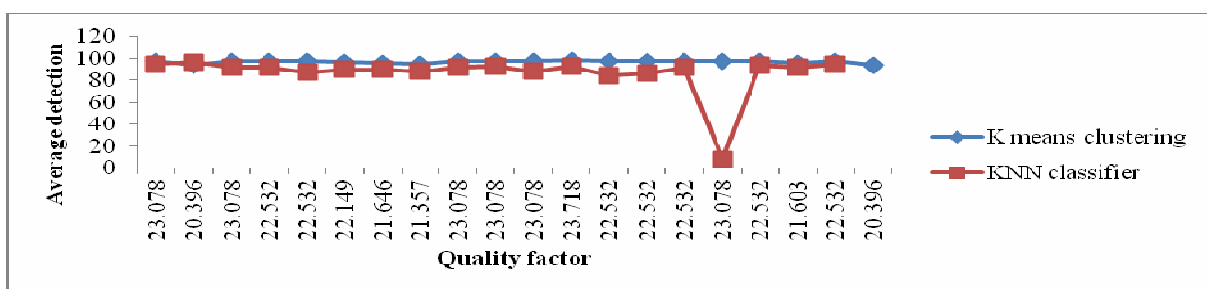


Fig. 3. Average Detection and Quality Value measures of KNN classifier and K means clustering



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 optimization is done by applying KNN classifier and K-means clustering. The objective was to classify risk level with high rate of classification, short delay from onset and a low false alarm rate. 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.Harikumar, 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] Dr.R.Harikumar, Sunil Kumar Prabhakar "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 Epileptic Activity in the EEG", IEEE Transactions On Biomedical Engineering, vol. 40, No. 12, December 1993.
- [7] Christo Ananth, Vivek.T, Selvakumar.S., Sakthi Kannan.S., Sankara Narayanan.D, "Impulse Noise Removal using Improved Particle Swarm Optimization", International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE), Volume 3, Issue 4, April 2014, pp 366-370
- [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 signals 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.Gevens, "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 Classifying Function Data", omputational 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] 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.
- [17] 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.
- [18] 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.
- [19] Jennifer G.Dy , Carla E. Brodley, " Feature Selection for Unsupervised lerning", Journal of Machine Learning Research vol.5 , pp. 845 – 889, 2005.
- [20] 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.
- [21] 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.
- [22] 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.