IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.-6, Issue-9, Sept. 2018 E-ISSN: 2347-2693

Epileptic Electroencephalogram Classification Using Machine Learning Algorithms

T. Perumal Rani[1*], Heren Chellam G.[2]

[1]Rani Anna Government College for Women, Affiliated by MS University, Tirunelveli, India
[2]Dept. of Computer Science, Rani Anna Government College for Women, Tirunelveli, India

[*]*Corresponding Author: perumalrani81@gmail.com, Ph: 9994733853*

Available online at: www.ijcseonline.org

Accepted: 14/Sept/2018, Published: 30/Sept/2018

***Abstract*—**Epilepsy is disease which is caused due to neurological disorder of a brain. It may cause recurrent seizures. It can be detected with the EEG signals and records the activity of brain electrically. In this paper K-Nearest Neighbor, Random Forest and Naive Bayes algorithms are used for classification of Electroencephalogram (EEG) signal as epilepsy or normal signal. These Machine learning algorithms learn directly from the data by experience which is not interrupted manually. Supervised learning uses labeled data for training which maps the input to the corresponding output. It classified into two types such as classification and regression. Classification means prediction of output from the input to which class it relies on, such as boy or girl. Whereas Regression means prediction of output from the input but output is predicted as a real value like measurement of rainfall etc. Here Random Forest method performs the best classification other than that of KNN and Naïve Bayes.

***Keywords*—**Classification, Electroencephalogram, Epilepsy, Machine Learning, Regression

I. INTRODUCTION

Machine learning is used to learn from the past experience that is programs which access data for learning themselves. It is categorized as supervised, unsupervised and semi supervised. Some methods like Decision Tree, Random forest, KNN and Naïve Bayes are iterative process used for classification [1].

In this project Epilepsy EEG signals are classified by using the machine learning algorithms. Electroencephalography (EEG) is used for recording electrical activity of the brain by placing electrodes along the scalp. Within neuron, it finds the voltage fluctuation and measures it from the ionic current over a period of time [2].

EEG is mostly used for diagnosing epilepsy. Electrical discharge suddenly occurs during epilepsy. Epilepsy means neurological disorder symptoms are no fever, confused memory, extreme tiredness. When the person gets seizure the patient suddenly becomes stiff and falls down and cannot communicate. Jerking movements of arms and legs will appear. Symptom of epilepsy is seizure and it may vary from people to people [3].

Machine learning is also one of the subparts of artificial intelligence (AI). The primary goal of here is to understand the structure and fitness of data into models. Machine learning algorithms are used to train the data inputs. By using statistical analysis they categorize the output values to the specified range and automatically create decision- making process for the given data inputs [4]. The major machine learning tasks are Regression, Classification, and Clustering. Regression is a supervised method to model and predict continuous, numeric variables. Classification means supervised learning method to model and predict categorical variables. Random forest is an ensample learning method which is used for classification, regression and other tasks. KNN is a statistical classification method and Naïve Bayes method is used for probabilistic distribution. Here K-Nearest Neighbor, Naïve Bayes are used for classification and Random Forest algorithm is only used for finding regression of EEG signals [5].

The following sections display the existing work of the above said method, its methodology, results and conclude Random Forest gives the best for the given problem based on the performance analysis.

II. RELATED WORK

Khalid Alkhatib, Hassan Najadat, Ismail Hmeidi, Mohammed K. Ali Shatnawi 2013presented their paper author use KNN and nonlinear regression for prediction of stock prices. Here they find out KNN have strong and low error rate and also predicted result are same like the original stock prices [6]. Aman Kataria1, M. D. Singh, 2013 Euclidean distance formula is used here for finding the distance for efficiency than Bayes method. It is used for pattern recognition. It has a high complexity which is based on the training samples. To avoid this genetic algorithm is combined with KNN and achieve the efficiency of 100% for small set of data [7]. Chih-Min Ma, Wei-Shui Yang and Bor-Wen Cheng 2014 presented method first we have to find out the value of K. Next to find out the distance metric and it has to be normalized. It achieved 96.73+5.97% by the use of normalization technique and Euclidian distance formula for distance calculation [8].

Weiting Chen, Yu Wang, Guitao Cao, Guoqiang Chen and QiufangGu presented in their paper the author compare the experimented result of Random Forest (RF) with the other classifier like SVM-linear, SVM-RBF, ANN, Decision tree, Logistic Regression (LR), ML and LDA. Basic statistical and segmented features are compared with the combined one. It can be better the separate one. They achieve the result with combined one as accuracy of 92.52%. The F1 score as 95.26% [9]. Sawthivaid, preetisingh and chamandeepkaur presented their paper to draw out the statistical characteristic from the EEG signal. Emotion means the person's inner state which is used for analyzing the condition of mind. They used a new technique Multi-Wavelet Transform (MWT) and RF to classify emotions in the EEG signals Finally the results of multi-Wavelet feature set are classified using MLP, KNN, MC-SVM and RF (ensample). They yield 98.1% for the emotions like happiness, sadness, and excitement and hatred on the whole [10]. Differentiation of dementia with levy bodies (DLB) from the disease which is called Alzheimer, RF is used for this purpose. Totally 66 DLB, 66AD patients and 66 controls are used for experimenting. Quantitative EEG (qEEG) is used with combination of clinical, neuropsychological, visual EEG, neuro imaging and cerebrospinal fluid data. Finally, they yield the result with accuracy 87%. Here they identified discriminating variable as a Beta power. qEEG accuracy with 10% more than other multi-model one [11].

[Juliano Machado and Alexandre Balbinot 2014] Here they compare the Linear Discriminant Analysis (LDA) and the Naïve Bayes (NB) classification algorithm. These algorithms are applied on EEG signals. The input features are the band pass filter signal energy, spectral energy's components and common spatial pattern filter. 70% was obtained as hit rates by using the database of this experiment. Hand movement behavior of EEG signals are classified and physiological effect also verified. Performance of NB classifier was similar to LDA even though it has a higher rate by the ANOVA analysis. Statistical behavior is different from one another [12]. [Beata Szuflitowska, Przemyslaw Orlowski 2017] Short Time Fourier Transform is used here for classification. The spectrum contains some features. It was extracted Linear Discriminant Analysis, NB Classifier and Gaussian NB Classifier are used for EEG signal classification. This method having the minimum size of training samples less than testing sample, the performance of classification was done using the STFT feature extraction method. NB classifier gains 95% which was better than 84% and 81% gained for LDA and GNBC methods. Mean value gains 83% for NB which was better than 78.5% for LDA and 79% for GNBC [13]. [Ali Akbar Hossinezadeh, AzraYaghoobi Karimoi, Reza Yaghoobi Karimoi, Mohammad Ali Khalilzadeh] Continuous Wavelet Transform (CWT) split and sales dominantly and extract some statistical features from it. Sequential Backward Search (SBS) and input of NB and Bayes Methods are used for reading dimensionality. In these methods contain normal, inter-ictal and ictal outputs. Thus, they obtained the best result with the selected features SFS and SBS 100% [14].

III. METHODOLOGY

EEG signals are classified using the algorithms like KNN, Random Forest and Naïve Bayes respectively. The Block Diagram of this system is shown in figure 1.

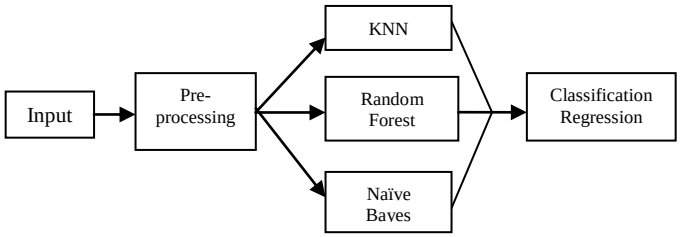

Figure 1: Block Diagram preprocessing is performed on EEG data

A. K-NEAREST NEIGHBOR

It is one of the slowest methods of learning. This Method is not build internal model and stores training data samples. It performs classification by voting the closest neighbor of each sample by voting the closest neighbor of each sample value. Main advantage of this method is simple, robust and effective for large data set. The disadvantage of this method is need to calculate the value of k and it is also has to compute the distance between each sample. So the computation cost is also increased. It is based on supervised learning [15]. It needs high memory requirement, very effective for noisy data. It is time-consuming and has very efficient memory indexing. It's accuracy also high which is

sensitive for outliers. KNN is used for pattern recognition and to estimate statistical report. If case k=1 it simply assigns to the closest one. Distance function is calculated using the Euclidean Distance formula for continues variables.

Predictions are done using the testing samples. It searches the training samples to the related K most samples for predicting the new variable and finally it summaries it as a mean variable for regression and class value for classification

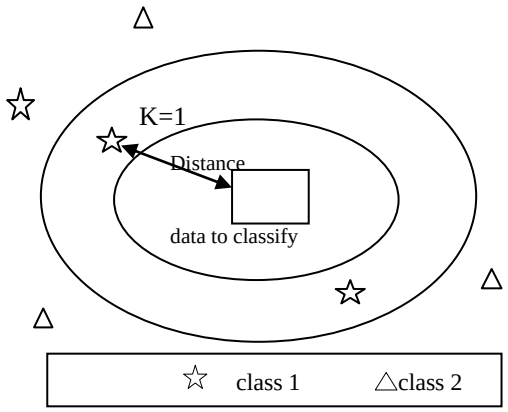

Figure 2: KNN Structure Diagram

The above figure2 shows how new data classified according to the nearest class value. Distance was calculated by the Euclidean distance formula.

The following are the features of KNN

1. All instances are samples relate to values in an n-dimensional Euclidean space.
2. When a new sample arrives it was classified by comparing the features vector of different samples.
3. Target function may be real valued.

The KNN Algorithm contains the following steps.

1. Take the training samples x1,x2,x3…xn
2. Find the Euclidean distance D (xi,xj)=sqrt(sum(xi-xj)^2) [i= 1 to n; j=1 to n]
3. Training sample's class labels are predicted with testing samples by voting majority of them to the K nearest neighbor [7].

The Nearest neighbor class label are assigned to the testing samples. The performance of the KNN algorithm predictions can be done using the cross table and the accuracy will be calculated by adding the true positive, true negative values and divide it by 100.

B. RANDOM FOREST

The base for this Random Forest method is decision tree. The decision tree classifiers limitations are overcome by a method which is proposed by Ho in 1995. Amit and Geman in 1997 proposed a method for space recognition. It was based on joint induction of space features with tree classifiers. In 1998 Ho produce a method with high accuracy by developing a decision tree classifier which increases the complexity.

In a controlled variation, construction of decision tree collection develops a random forest method. This method can do both classification and regression, which was in implemented by Breimin in 2001. It is combination of random selection features (Ho in1995; 1998 and Amit Geman in1997) and Breiman's sampling technique of voting system is used for classification and prediction.

RF is an efficient method of classification which introduces a bagging system and selection is made by randomly to build a tree. It is used to a group of classification built by the training data. The variables are randomly selected on each split and bagging is also used. By these two combinations it achieves the low correlation in each tree.

The main Advantages of this RF method are high accuracy. It finds the important features and can handle a thousand features without missing. It can be used for clustering and outlier detection.

RF Algorithm:

Ts : Training Sample (x1, x2, x3…xn)

NO_{tree}: The number of trees which is to be built.

Max_{try}: The amount of variable for each split.

For each i=1 to NO_{tree}

Draw sample with all features using a bootstrap sampling technique.

Max $_{try}$ Variable are randomly selected.

For each node of tree

find out the best split

The above algorithm displays the processing steps of random forest method.

Classification:

Maximum vote for N trees [12].

Favg(X) = (p1(X)…Pk(X)= $\frac{1}{N}\sum_{1}^{N} f_i(x)$

$f_{RF}(X)=argmax_k\{p1(X)\ldots Pk(X)\}$

The parameters for optimization was displayed by the following steps,

Parameters for Optimization

Max_{try} depends on data.

NO_{tree} depends on *Max_{try}* candidates.

These are main two parameters used for optimization of random forest algorithm.

C. NAÏVE BAYES

It is a simple, supervised and effective technique which is based on conditional probability like Bayesian theorem [16].

Previous knowledge is used for making predictor in Naïve Bayes method and it can be changed based on collection of evidence. The formula used for naïve bayes theorem [17]. Naive Bayes is best for categorical data but can also be used for continuous data

$$P\left(\frac{X}{Y}\right) = \left[P\left(\frac{Y}{X}\right) * P(X)\right] / P(Y) \quad (4)$$

Where X-Categorical outcome events

Y- Series of Predictors

P(X)-Probability of X

P(Y)-Probability of Y

P(X/Y)-Probability of X conditional on Y

P(Y/X)-Probability of Y conditional on X

If some attributes depend on each other than it is possible to gain output as good classification.

$$P(x1, x2, x3, \ldots xn/Y) = \pi_i P(xi/Y) \quad \square\square\square$$

The most Posterior probability classifier output

$$= \arg\max\left\{P(Y) * \pi_i P\left(\frac{xi}{Y}\right)\right\} \quad \square\square\square$$

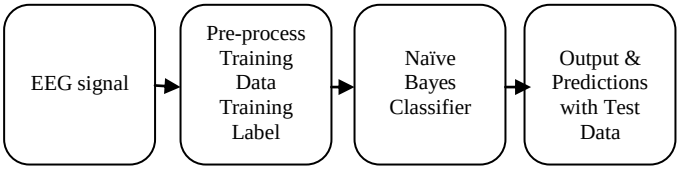

Figure 3: Block Diagram of Naïve Bayes.

The above figure 3 displays the block diagram of Naïve Bayes Method. It shows that pre-processed data of training and testing samples are classified by the Naïve Bayes algorithm. The output is predicted with the testing data.

IV. RESULTS AND DISCUSSION

The experiment will be performed by using the epileptology department data sets Bonn University. The Datasets A, B, C, D, E contain 100 single-channel segments with 23.6 s duration, recorded by 128-channel amplifier which will be digitized by 173.61Hz sampling rate. Data sets A, B of five healthy people EEG signal will be taken with eyes open and closed form respectively. Five Epileptic patients EEG signals were represented as C, D and E. Here C, D is collected before the surgery and E is recorded during the epileptic zone.

The results are analyzed with the above said data sets, using three algorithms KNN, RF and NB for classification. The summarized results are tabulated below for every method.Total accuracy for the KNN of all data set pair are displayed in table 1

Table 1: Total Accuracy for KNN

Dataset Pair Name	Total Accuracy for the dataset pair %
AE	69%
BE	68%
AD	82%
DE	74%

The result of Random Forest algorithm before and after fine tuning process is summarized with the best Out-of-Bag Error values are displayed in table 2

Table 2: Result summary of Random Forest

Name of the Data set pair	% Var explained before mtry	Best Mtry OOB Error	% Var explained after mtry
AE	98.82 %	0.0004213453	99.84 %
AD	98.94 %	0.0003967476	99.85 %
BE	98.84 %	0.0004237137	99.84 %
DE	98.8 %	0.0004534876	99.84 %

The Out-of-Bag (OOB) error is plotted in figure 4

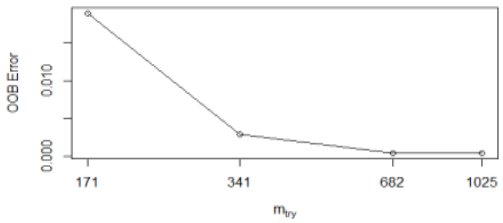

Figure 4: RF's OOB error plot

The result of Naïve Bayes method are summarized with MSE and RMSE in table 3

Table 3: Result summary of Naïve Bayes

Data set pair	Accuracy	MSE	RMSE
AE	0.579399	0.4124453	0.6422191
BE	0.573677	0.3986572	0.6313931
AD	0.669399	0.3639988	0.6033231
DE	0.5554939	0.4473157	0.6688166

The overall performance is analyzed by a chart for the above three algorithms is shown in figure 4

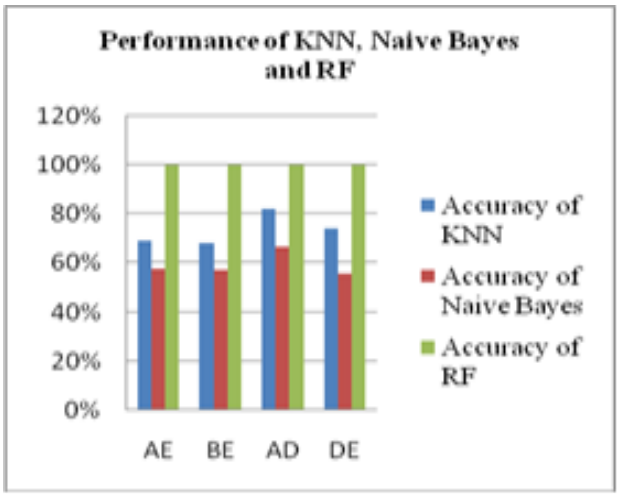

Figure 4: Performance of KNN, Naive Bayes and Random Forest

V. CONCLUSION

In this paper Electroencephalogram (EEG) signal data sets were classified like healthy person's or Epilepsy patient's signal using the algorithms KNN, Random Forest and Naïve Bayes. It was classified by using the KNN algorithm with accuracy as 82%. Random Forest algorithm too is used here for finding regression and the results are fine-tuned by the best Out-of-Bag error result yields result as 99.85%. Naïve Bayes algorithm is used here for probability distribution of these EEG signals. The maximum accuracy from this method is 66%. Finally, according to the results, Random forest algorithm performs better results than the other algorithms with an accuracy of 99.85%.

REFERENCES

[1] J.V.N. Lakshmi, Ananthi Sheshasaayee, "A Big Data Analytical Approach for Analyzing Temperature Dataset using Machine Learning Techniques", International Journal of scientific Research in Computer Science and Engineering, vol 5, issue 3, pp 92-97, June (2017), E-ISSN : 2320-7639

[2] Md. Khayrul Bashar, Ishio Chiaki, Hiroaki Yoshida Human Identification from Brain EEG signals Using Advanced Machine Learning Method, 978-1-4673-7791-1/16/$31.00 ©2016 IEEE

[3] Sabrina Ammar, Mohamed Senouci, "Seizure Detection with Single-Channel EEG using Extreme Learning Machine" 978-1-5090-3407-9/16/$31.00 ©2016 IEEE

[4] Vinay K, "Machine learning approach via an ensemble of classifiers for computer aided lung nodule diagnosis", Shodhganga :a reservoir of Indian theses @ INFLIBNET, Mar2015

[5] Rishi Das Roy, "Development and application of machine learning tool in deciphering biological information", Shodhganga :a reservoir of Indian theses @ INFLIBNET, June 2016

[6] Khalid Alkhatib, Hassan Najadat, Ismail Hmeidi, Mohammed K. Ali Shatnawi, 'Stock Price Prediction Using K-Nearest Neighbor (kNN) Algorithm' International Journal of Business, Humanities and Technology Vol. 3 No. 3; March 2013 32

[7] Aman Kataria1, M. D. Singh, International Journal of Emerging Technology and Advanced Engineering Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 3, Issue 6, June 2013)

[8] Chin-Min Ma, Wei-Shui Yang and Bor-Wen Cheng, 'How the parameters of K-Nearest Neighbor Algorithm Impact on the Classification Accuracy: In Case of Parkinson Dataset' Journal of Applied Sciences 14 (2): 171-176,2014 ISSN 1812-5654 / DOI: 10.3923/ jas 2014.171.176 ©2014 Asian Network for Science Information

[9] Weiting Chen, Yu Wang, Guitao Cao, Guoqiang Chen, Qiufang Gu, December 2013 'A random forest model based classification scheme for neonatal amplitude-integrated EEG' IEEE International Conference on Bioinformatics and Biomedicine (BIBM 2013) Shanghai, China. 18-21

[10] Swati Vaid, Preeti Singh and Chamandeep Kaur, "Classification of Human Emotions using Multiwavelet Transform based Features and Random Forest Technique"', Indian Journal of Science and Technology, Vol 8(28), DOI: 10.17485/ijst/2015/v8i28/70797, ISSN (Print) : 0974-6846 ISSN (Online) : 0974-5645 October 2015

[11] Meenakshi Dauwan, Jessica J. van der Zande, Edwin van Dellen, Iris E. C. Sommer, Philip Schelten, Afina W. Lemstra, Cornelis J. Stam, "Random Forest to differentiate dementia with lewy bodies from alzheimers disease", 2352-8729/Ó2016 Published by Elsevier Inc

[12] Juliano Machado, Alexandre Balbinot, Adalberto schuck "A study of Naïve Bayes Classifier for analyzing imaginary movement EEG signal using the periodogram as spectral estimation", 2013

[13] Beata szuflitowska, przemyslaw, orlowski, "Comparision of the EEG signal classifier LDA, NBC and GNBC based on time frequency features", 2017

[14] Ali Akbar Hossinezadeh, Azra Yoghoobi Karimoi, Reza Yaghoobi, Mohammad Ali Khalizadeh, "EEG signal classification using Bayes and Naïve Bayes classifier and extracted features of continues wavelet transform"

[15] Khalid Alkhatib, Hassan Najadat, Ismail Hmeidi, Mohammed K. Ali Shatnawi, "Stock Price Prediction Using K-Nearest Neighbor (kNN) Algorithm" International Journal of Business, Humanities and Technology Vol. 3 No. 3; March 2013 32

[16] Deepika Mallampati, "An Efficient Spam Filtering using Supervised Machine Learning Techniques", International Journal of scientific Research in Computer Science and Engineering, vol 6, issue 2, pp 33-37, April (2018), E-ISSN : 2320-7639

[17] Zhongheng Zhang, Submitted, " Naïve Bayes classification in R", Accepted for publication Feb 24, 2016. doi: 10.21037/atm.2016.03.38 View this article at: http://dx.doi.org/10.21037/atm.2016.03.38, Jan 25, 2016

Authors Profile

Mrs. T. Perumal Rani is currently pursing research [RegNo.18121172162004], Department of Computer Science, Rani Anna Government College for Women, Affiliated by Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli-627 012, Tamil Nadu, India. She is working as guest lecturer in that college. She pursued her Bachelor of Computer Science from Rose Mary College, Aft.Manonmaniam Sundaranar, Tirunelveli and Master of Information Technology from Allahabad Agricultural Institute-Deemed University. She received her M.Phil Degree in Madurai Kamraj University, Madurai. She has presented a paper in one day National Conference published a paper in 'Sadakath: A Research Bulletin with ISSN 2347-7644'. Her main research work focuses on Deep Learning Algorithms and Network Security. She has 9 years of teaching experience and 6 months of Research Experience.

Dr. G. Heren Chellam is currently working as an Assistant Professor in the Department of Computer Science, Rani Anna Government College for Women, Gandhi Nagar, Tirunelveli-627 008, Tamil Nadu, India. She received her MCA in the Dept. of Computer Science and Engineering, Annamalai University, Chidambaram and M.Phil Degree in Mother Teresa Women's University, Kodaikanal. She received her Doctorate in Computer Applications from Manonmanium Sundaranar University. She has published papers in many National and International level Journals and Conferences. Her Research interest are in the field of Neural Networks, Digital Image Processing, Pattern Recognition. She has 26 years of teaching experience and 13 years of Research Experience.

