iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** II **Month of publication:** February 2020

DOI: http://doi.org/10.22214/ijraset.2020.2062

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.177
Volume 8 Issue II Feb 2020- Available at www.ijraset.com

Performance Analysis of Machine Learning Algorithms for Classification of Electroencephalogram Signals for Detection of Seizures

Pranita Garde[1], Anuradha Thakare[2], Aishwarya Biradar[3], Neha Pawar[4]
[1, 2]*Computer Department, PCCOE, India*

Abstract: ***Over fifty million people in the world are affected by a neurological disorder namely Seizures. The seizure occurs due to uncontrolled electrical activities inside the brain. It may cause changes in someone's behaviour, movement, or emotions. Many seizures last for about 30 seconds to two minutes. Electroencephalogram (EEG) signals are used to detect the existence of seizures. The detection of seizures is very important from the patient's point of view. Therefore, there is a need to study and explore machine learning algorithms which may learn various patterns in the EEG datasets and perform accurate classification. In this work, the algorithms like K-nearest neighbour (KNN), Naïve Bayes (NB), Decision tree (DT) and Random forest (RF) are compared for classification of EEG dataset. These algorithms are applied on extracted features which result in the classification of EEG signals and are classified as seizures and non-seizures. After experimentation, it is observed that Random Forest gives best accuracy.***

Keywords: Electroencephalogram (EEG) signals, feature extraction, seizure detection, K-nearest neighbour (KNN), Naïve Bayes (NB), Decision tree (DT) and Random forest (RF).

I. INTRODUCTION

The human brain consists of many neurons that act as a significant role in controlling the behavior of the human body. The electroencephalography is a test that involves recording electrical signals of the brain [1]. Seizures, also known as epileptic seizure are symptoms of a brain problem. They result due to unexpected, uncommon condition occurring in the brain. Around, 1-2% of the population suffers from seizures. Mostly seizure lasts from 30 seconds to 2 minutes. But if seizures last more than 5 minutes then emergency medical aid is required. There are two types of seizures. Focal seizures are in one section of the brain. Generalized seizures tend to occur all over the brain. The symptoms usually differ based on the kind of seizure and may involve loss of awareness, convulsive movement of muscles, disturbances in eyesight or neural acoustic or also staring blankly without awareness. In this paper, we have performed analysis of various machine learning approaches to find out the best approach for classifying EEG signals for seizure detection.

II. RELATED PREVIOUS WORK

Various methods have been proposed for the detection of seizures. The authors proposed an algorithm for seizure detection that uses the Harmonic wavelet packet transform (HWPT) and Fractal dimension (FD) for feature extraction. Relevance vector machine [4] is used for classification and high accuracy rates are obtained. The proposed method first extracts the brain rhythms from EEG signals and then they are modelled based on statistical properties. Features such as Hurst component and autoregressive moving average (ARMA) parameters were extracted for building Support Vector Machine (SVM) classifier. This is only suitable for small database [5]. The proposed method extracts features using Fourier transform and then Deep learning technique based on multilayer perceptron is applied for classification. The highest accuracy is obtained by using multilayer perceptron with two hidden layers. If layers are increased it leads to over fitting problem [6]. Discrete Wavelet transform is used for feature extraction. Multilayer Perceptron and Random Forest are the two algorithms used for classification and their accuracies are compared. The results showed that the Random Forest classifier gives the highest accuracy [7]. The authors proposed a seizure detection technique which uses Discrete Wavelength Transform for fragmenting EEG signals into sub bands and Hjorth parameters are extracted. KNN is used for classification purpose. Improved classification accuracy is obtained as compared to existing models [8]. The authors conducted an experiment to compare performances of Frequency Domain and Time Domain signals for feature selection along with Convolution Neural Network as a classification algorithm. And as a result, Time Domain signals performance was better than Frequency Domain signals [9]. In the proposed work, authors have described an automated classification method for the epileptic seizure detection using wavelet transform and statistical pattern recognition. Quadratic Classifier gave an overall 99% accuracy [10]. A matrix of multi feature dimension is obtained that contains Brownian motion, Mean Absolute Value and Root Mean Square as key features. These features obtained from artefact free EEG signal are fed to various classifiers like KNN, Ensemble bagged tree and Support

Vector Machine (SVM). Their work demonstrated that the ensemble bagged tree is more efficient classifier [11]. The proposed work compared performance of optimized S-transform based method for seizure detection with standard S-transform, Short time Fourier transform (STFT) and other quadratic time-frequency distributions. Optimized S-transform based detection method gives better accurate results [12]. The authors have used t-SNE algorithm for pre-processing of EEG signal dataset and then three different algorithms (SVM, KNN and Random Forest) were applied on the dataset to classify EEG signals to distinguish between normal signals and signals with seizures. As a result, the best accurate classification was provided by the Random Forest classifier [13].

III. METHODOLOGY

A. *Classifiers*

1) K-nearest Neighbours (KNN): The K nearest neighbour algorithm assumes that similar things are close to each other. Hence it uses distance formulae for calculating dissimilarity between two instances.

Algorithm

a) Store all the training records in an array.

b) For each training sample in data, calculate Euclidean distance between test data and training data.

Euclidean distance is as follows:

$$d(X,Y)=\sqrt{\sum_{i=1}^{n}(X_i-Y_i)^2} \qquad (1)$$

c) Sort the calculated distances from smallest to largest (ascending order).

d) Pick first k rows from the sorted collection and find the class label of these rows.

e) Give back the predicted class label.

2) Naïve Bayes (NB): Naïve Bayes is a classification technique based on Bayes' probability theorem. The implementation of Naïve Bayes' classifier is easy and can execute effectively even without previous knowledge of data. This algorithm is widely used in the classification of text documents. It is quick to create models and make predictions with this algorithm.

$$P(a\,|\,b)=\frac{P(b\,|\,a)P(a)}{P(b)} \qquad (2)$$

$$P(a\,|\,b)=P(b_1\,|\,a)\times P(b_2\,|\,a)\times\ldots\times P(b_n\,|\,a)\times P(a) \qquad (3)$$

Above,

a) *P(a|b):* The posterior probability of class c given predictor or attribute x.

b) *P(a):* The prior probability of class c.

c) *P(b|a):* The likelihood probability of predictor class.

d) *P(a):* The prior probability of predictor or attribute.

3) *Decision Tree (DT):* The decision tree algorithm is based on tree structure which comprises nodes as splitting points of dataset or decisions and branches as results of decisions.

Algorithm

a) Spot the best attribute from dataset and place at the root of the tree. The best attribute can be selected based on two selection measures:

i) *Information Gain*

The attribute is selected as best which has highest information gain.

Following is the formula for information gain:

$$Gain(S,D)=H(S)-\sum_{V\in D}\frac{|V|}{|S|}H(V) \qquad (4)$$

Where,

H(S) is Base Entropy

H(V) is Conditional Entropy

Entropy is calculated as follows:

$$Entropy=\sum_{i=1}^{C}-p_i*\log_2(p_i) \qquad (5)$$

ii) Gini Index

Attribute with the lowest Gini index is chosen as the best attribute.

Gini index formula is as follows:

$$Gini = 1 - \sum_{i=1}^{C} (p_i)^2 \qquad (6)$$

b) Split the training dataset into subsets using the best attribute as a splitting point.

c) Repeat all above steps on each subgroup till the leaf nodes are found.

4) Random Forest (RF): The random forest algorithm generates the forest with a variety of trees. In this classifier, if the number of trees is higher then it gives the best accuracy results. It can overcome the over-fitting problem in decision trees. As it is a collection of trees it can handle abundant data.

Algorithm

a) Generate a Bootstrapped dataset from the given dataset.

b) Generate decision trees based on this bootstrapped dataset.

c) Repeat step 1.

d) Select the class label with maximum votes as a class label for the query/test sample.

B. K – fold Cross Validation

Cross Validation is a method, which is used to cross check how accurately the system predicts the output when the dataset is differed. In short, this method checks the consistency of the system. Here the number of folds applied are k=4. According to the value of k, the dataset is divided into k parts and for 'k' different iterations, 'k-1' part(s) are used for training and one part is used for testing.

IV. PERFORMANCE ANALYSIS OF MACHINE LEARNING APPROACHES

A. Dataset

The EEG dataset used is in numeric format. Fig. 1. mentioned below represents the dataset structure. This dataset was taken from Kaggle.

	Unnamed	X1	X2	X3	X4	X5	...	X176	X177	X178	y
2	X21.V1.79	135	190	229	223	192	...	-116	-83	-51	4
3	X15.V1.92	386	382	356	331	320	...	154	143	129	1
4	X8.V1.1	-32	-39	-47	-37	-32	...	-35	-35	-36	5
5	X16.V1.60	-105	-101	-96	-92	-89	...	-72	-69	-65	5
6	X20.V1.54	-9	-65	-98	-102	-78	...	-83	-89	-73	5
7	X14.V1.56	55	28	18	16	16	...	-60	-56	-55	5
8	X3.V1.191	-55	-9	52	111	135	...	11	67	128	4
9	X11.V1.27	1	-2	-8	-11	-12	...	-66	-57	-39	2
10	X19.V1.87	-278	-246	-215	-191	-177	...	-125	-79	-40	1
11	X3.V1.491	8	15	13	3	-6	...	-15	-15	-11	4
12	X3.V1.6	-5	15	28	28	9	...	-101	-89	-49	5
13	X21.V1.72	-167	-230	-280	-315	-338	...	215	165	103	1
14	X7.V1.162	92	49	0	-32	-51	...	-1	-7	-44	3
15	X1.V1.211	15	12	0	-17	-28	...	13	44	68	4

Fig. 1.EEG Dataset

The original database consists of 5 files and each file consisting of 100 files. These files consisted of brain activity of each patient for 23.6 seconds. 4097 data points were splitted and rearranged into 23 blocks and each block consists of 178 data points for 1 second. It contains one patient ID column and last column with label y with values 1-5. If y is 1 it means EEG recording of seizure activity, 2 means EEG recording from area of tumour, 3 means EEG recording of healthy brain area, 4 means recording when patient closed their eyes and 5 means when patient opened their eyes.

B. Proposed Model

We have proposed on model which will give more promising results for detection of seizures. Fig. 2. depicts the proposed model for classification of EEG data. The proposed system includes three steps: pre-processing, split the dataset and classification.

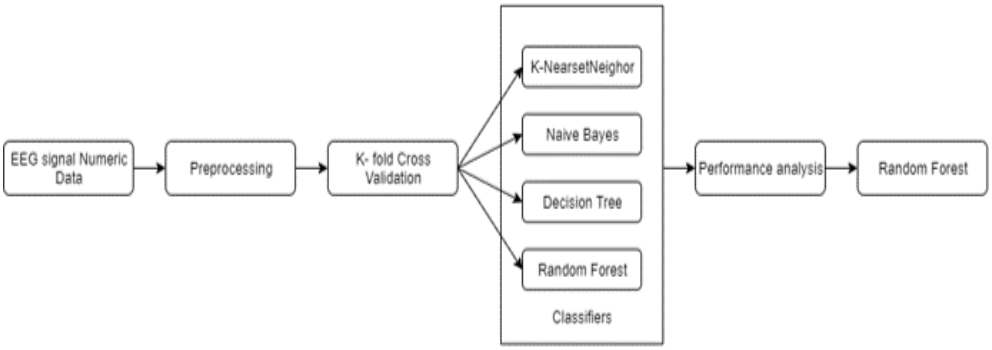


Fig. 2. Proposed Model

First pre-processing is performed on the numeric EEG dataset. The pre-processed data is splitted into training and testing datasets. When we split the dataset, most of the data is used for training and remaining for testing. Various classification algorithms such as K-Nearest Neighbour, Naïve Bayes, Decision Tree and Random Forest are applied and performances of all four classifiers are compared. Among all four algorithms, Random forest's performance is better than other algorithms.

V. RESULT AND DISCUSSION

The seizure detection from EEG data is a computationally complex task and it requires each and every parameter to be monitored during analysis phase. Machine learning algorithms, discussed in above section are used for the said work for classification of EEG data of patients. The results of all the four algorithms are computed for as training and testing phase. The experimentation is performed on 11500 samples and found promising results during training and testing phase. These are as depicted in Table I and Table II.

TABLE I
EXPERIMENTAL RESULTS - TRAINING PHASE

Name of Classifiers	Training				
	AUC	Accuracy	Recall	Specificity	Precision
KNN	99.2	62	24.1	99.7	99.9
NB	98.6	93.6	89.5	97.5	97.7
DT	98.3	97.5	95.6	99.4	94.4
RF	94.4	95.5	92.5	98.4	98.5

Table II
Experimental Results - Testing Phase

Name of Classifiers	Testing				
	AUC	Accuracy	Recall	Specificity	Precision
KNN	96.9	84.5	25.9	98.9	99.9
NB	98.5	96.2	90.3	91.1	97.7
DT	89.2	90.8	88.6	72.5	91.4
RF	98.5	95.4	90	87.5	96.7

As mentioned in the above table, K- nearest algorithm gives the least accuracy because it takes more computation time to calculate distances between instances. Naïve Bayes being a probabilistic approach provides good accuracy for limited data. Hence gives less accuracy for this dataset. Decision Tree gives good accuracy for the training dataset but gives less accuracy for the testing dataset as it is prone to outliers and over fitting. Whereas Random Forest gives the best accuracy in training as well as testing phase as it overcomes the limitation of Decision Tree by considering more than one decision trees at a time and choosing the best among them.
We have performed k-fold cross validation on given dataset before applying any classifier to improve the accuracy of seizure detection.

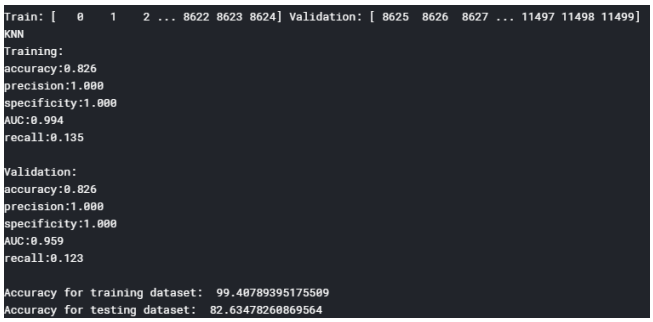


Fig. 3. Results of cross validation for K-Nearest Neighbors

Fig. 3. shows the cross-validation results for KNN algorithm. Before applying the 4-fold cross validation on given dataset, the accuracies for training and testing dataset was 62% and 84.5% respectively. After the 4-fold cross validation the mean accuracies for training and testing datasets became 99.4% and 82.63% respectively.

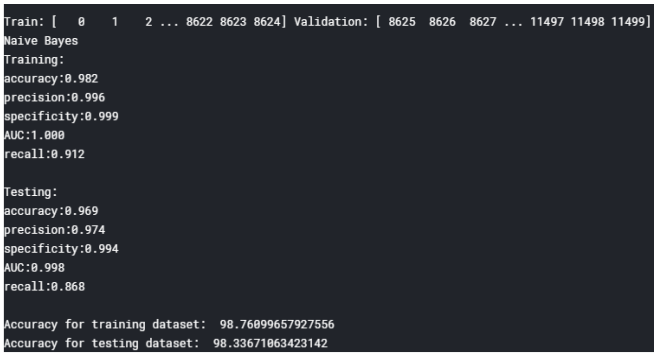


Fig. 4. Results of cross validation for Naïve Bayes

Fig. 4. shows the results of seizure detection after performing 4-fold cross validation on dataset using Naïve Bayes algorithm. The accuracy of Naïve Bayes algorithm before cross validation was 93.6% for training and 96.2% for testing. After performing 4-fold cross validation the accuracy increased to 98.76% and 98.33% for training dataset and testing dataset respectively.

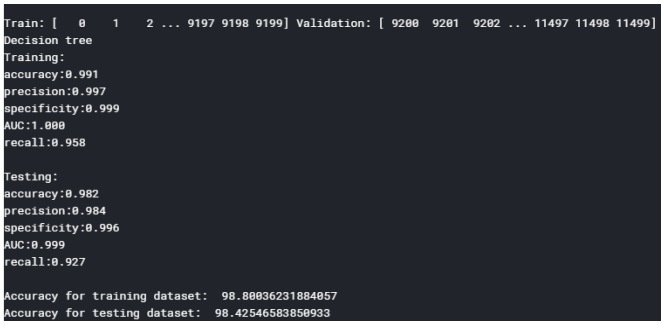


Fig. 5. Results of cross validation for Decision tree

Fig. 5. depicts the performance of decision tree after implementing cross validation. The accuracies for training and testing datasets was 97.5% and 90.8% respectively for decision tree algorithm. But after cross validation better accuracies were obtained. For training dataset 98.8% accuracy and for testing dataset 98.42% accuracy.

Fig. 6. Results of cross validation for Random Forest

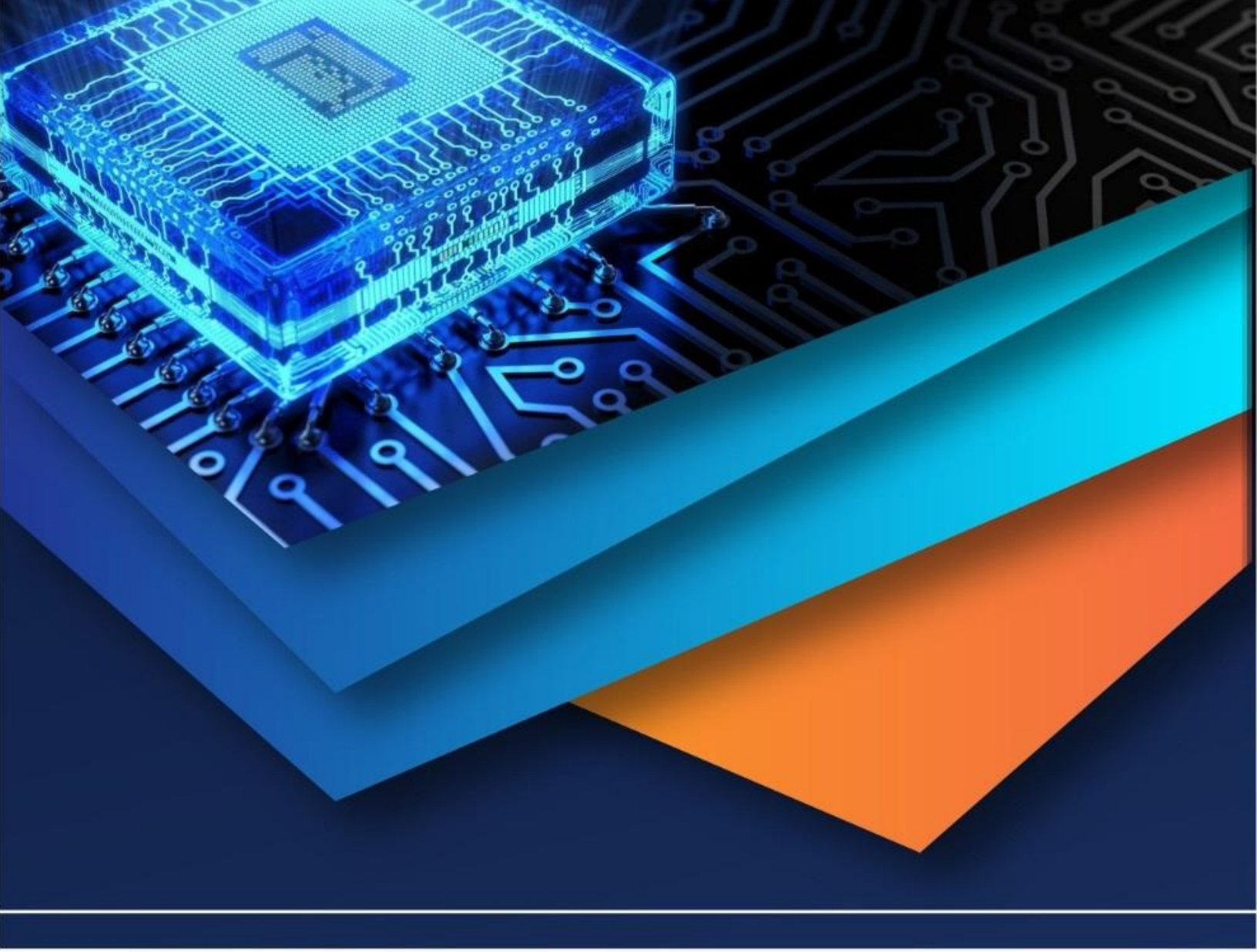
Fig. 6. represents outcome of Random Forest after implementation of cross validation. The accuracy of Random Forest before implementation of cross validation was 95.5 and 95.4 for training and testing respectively. After application of cross validation, accuracy of random forest has been improved to 98.2 and 96.9 for training and testing respectively.

VI. CONCLUSION

EEG signals are used to identify sudden uncontrollable electrical activities inside the brain. Hence, in this work, we used the EEG dataset from Kaggle for the detection of seizures. We have applied four machine learning algorithms namely K-nearest neighbor, Naïve Bayes, Decision Tree, Random Forest on this numeric dataset of EEG signals. We analyzed performances of these algorithms and observed that Random Forest showed promising results in terms of AUC, accuracy, recall, specificity and precision for both training and testing datasets respectively. After that we have performed k-fold cross validation (here k=4) on the given dataset before applying all four algorithms inorder to get more promising results and compared their accuracies. Random forest algorithm gave better results than other three algorithms.

REFERENCES

[1] Manish Sharmaa , Ram Bilas Pachorib , U. Rajendra Acharya " A new approach to characterize epileptic seizures using analytic time-frequency flexible wavelet transform and fractal dimension", 2011.
[2] Abdulhamit Subasi , Jasmin Kevric, M. Abdullah Canbaz "Epileptic seizure detection using hybrid machine learning methods", 2017.
[3] H. Stefan et al., "Objective quantification of seizure frequency and treatment success via long-term outpatient video-EEG monitoring: A feasibility study Seizure", 2011.
[4] Lasitha S. Vidyaratne, Khan M. Iftekharuddin, "Real-Time Epileptic Seizure Detection Using EEG", IEEE 2017.
[5] Anubha Gupta, Pushpendra Singh, Mandar Karlekar, "A Novel Signal Modeling Approach for Classification of Seizure and Seizure-free EEG Signals", IEEE 2018.
[6] J. Birjandtalab, M. Heydarzadeh, M. Nourani, "Automated EEG-Based Epileptic Seizure Detection Using Deep Neural Networks", IEEE 2017.
[7] Suvadeep Bose, V Rama, and Dr. C.B.Rama Rao, "EEG signal analysis for Seizure detection using Discrete Wavelet Transform and Random Forest", IEEE 2017.
[8] Md Abu Sayeed, Saraju P. Mohanty, Hitten Zaveri, "A Fast and Accurate Approach for Real-Time Seizure Detection" IEEE 2018.
[9] Mengni Zhou, Cheng Tian, Rui Cao, Bin Wang, Yan Niu, Ting Hu, Hao Guo1 and Jie Xiang, "Epileptic seizure detection based on EEG signals and CNN", IEEE 2017.
[10] D. Gajic, Z. Djurovic, S. Di Gennaro and Fredrik Gustafsson, "Classification of EEG signals for detection of epileptic seizures based on wavelets and statistical pattern recognition", IEEE 2015.
[11] Tarak Das, Sayanti Guha, Arijit Ghosh, Piyali Basak, "Classification of EEG Signals for Prediction of Seizure using Multi-Feature Extraction", IEEE 2017.
[12] Narendra Kumar Ambulkar, S N. Sharma "Detection of Epileptic Seizure in EEG Signals Using Window Width Optimized S-transform and Artificial Neural Networks", IEEE 2015.
[13] Seferkurnaz, Ahmed Ayad Saleh, "Comparative and Analysis Study of normal and epileptic seizure EEG signals by using various classification Algorithms", IEEE 2018.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY