A Novel Approach for EEG Signal Classification using Wavelet Transform and Random Forests

Parth Sagar[1]*, Shivraj Kone[2], Munmun Bhagat[3]
[1,2,3]Assistant Professor, Department of Computer Engineering, SPPU, Pune, Maharashtra, India
*Email:parth.cool007@gmail.com
DOI:

Abstract
Unprovoked seizure is the symptoms of Epilepsy disorder. An electroencephalogram (EEG) is a test that perceives electrical activity in your mind using nearly nothing, level metal circles (anodes) added to your scalp. The EEG signals are used for diagnosis of the patient whether it is seizure or non-seizure which causes epilepsy. The essential objective of this paper is to build a classifier that can accurately recognize whether a subject is seizure patient or non-seizure patient. The proposed system classifies the EEG signals into two classes using different supervised learning algorithm. Features are extracted using discrete wavelet transform (DWT) and Genetic algorithm approach. Features dimensions have been reduced using Principal component analysis (PCA) before feature extraction. Different classification algorithms like support vector machine, logistic regression and decision tree classifier. Random forest based classification with regularization gives best results which is best from the existing system.

Keywords:*EEG, DWT, Neurons, Brain*

INTRODUCTION

An EEG records the electrical movement of neurons in the brain. Fluctuations in the electricity are estimated from the voltage distinction of electrodes associated with the brain network [1]. The indicative procedure utilizing an EEG is completed by checking patients constantly more than a few days. By the by, a large portion of the recorded information must be watched and examined outwardly by a specialist so as to identify epilepsy accurately[2, 3]. This discovery methodology is viewed as wasteful, as the time has come expending and exorbitant. Hence, devices or techniques that lead to the quick and precise determination of epilepsy are required. Actually, various techniques have been produced to recognize instances of epilepsy. Ongoing takes a shot at EEG arrangement are, among others, those of who utilize factual ghostly element extraction straightforwardly to characterize the EEG flag, which apply k-NN to the unearthly highlights of EEG records, and,

who play out the arrangement by straightforward arbitrary inspecting and highlight extraction[7]. A large portion of the strategies complete component extraction on the signs preceding order and would thus be able to be considered as two-advance methodology. In spite of the fact that it has points of interest, highlight extraction may prompt some essential data in the crude highlights being lost. Besides, utilizing separated highlights to foresee the class of new cases now and again prompts poor execution, which implies that the strategy performs well in preparing information however neglects to order the testing information well. Hira and Gillies (2015) thoroughly talk about the shortcomings and points of interest of highlight choice and extraction. This circumstance prompts the proposal that the full highlights ought to be utilized now and again.

On account of the idea of the EEG record, the RF and SVM strategies can't be

straightforwardly connected to the flag produced from an EEG and the flag in this manner must be changed. This examination changes the flag utilizing Discrete Wavelet Transform (DWT) as the preprocessing strategy. DWT is a generally prescribed approach for investigating information that has both time and frequency. By deteriorating the flag into restricted components (both time and frequency), Mallat and Hwang (1992) contend that DWT can describe the example well.

The base outcome required for this venture to be a full achievement is to have built up a classifier that is able to do precisely ordering scraps of EEG session information as being from the representation of either a commonplace or a new ability. Since this is a double characterization issue with adjusted classes, the base standard for accuracy is 0.5. Full achievement would mean having an accuracy of in any event 70% (in spite of the fact that this number is self-assertive). Best in class EEG

characterization systems as of now score extensively higher than this. Information handling and growth is required to be vital and various methodologies will be considered[8].

When a practical classifier has been created, the objectives are twofold. First to utilize present day profound learning procedures to expand the test accuracy. One proposed approach is sketched out in a paper that proposes anticipating the changed EEG frequencies into a 2D picture with a profundity measurement for frequency band. This organization makes for perfect contributions to a standard convolution neural network. This or another profound learning approach might be utilized to accomplish the most astounding conceivable accuracy[9]. The second arrangement of objectives base on giving knowledge into the basic systems in brain work. The techniques for achievement of these objectives will rely upon the subtleties of the machine learning calculations that can effectively group the EEG information[10].
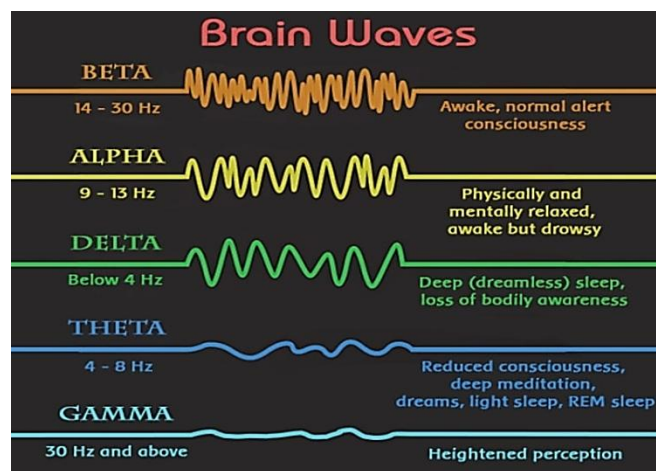
Figure 1:Brain Waves Type.

Delta: has a frequency of < 3 Hz. It will in general be the most noteworthy in force and the wave with biggest timeframe. Usually as the predominant musicality in babies up to 1 year and in stages 3 and 4 of

rest. It might happen centrally with subcortical sores and all in all appropriation with diffuse sores, metabolic encephalopathy hydrocephalus or profound midline sores. It is typically most

noticeable frontally in grown-ups (for example FIRDA - Frontal Intermittent Rhythmic Delta) and posteriorly in non grown-ups for example OIRDA - Occipital Intermittent Rhythmic Delta) [5].

Theta: has frequency of 3.5-7.5 Hz and is delegated "moderate" action. It is consummately regular in kids <13 years and in rest however not ordinary in conscious adults.It can be viewed as an appearance of central subcortical injuries; it can likewise be seen in summed up appropriation in diffuse scatters, for example, metabolic encephalopathy or a few occasions of hydrocephalus[4].

Alpha: has a frequency of 7.5 to 13 Hz. Is typically best seen in the back locales of the brain on each side, being higher in force on the prevailing side. It is seen when shutting the eyes and unwinding, and evaporates when opening the eyes or cautioning by components like reasoning, ascertaining, and so forth. It is the significant mood found in typical loosened up grown-ups. It is available for the greater part of the life especially after the thirteenth year[4].

Beta: beta action is "quick" action. It has a frequency of more noteworthy than 14Hz. It is typically seen on the two sides in symmetrical circulation and is most apparent frontally.It is quickened by narcotic entrancing medications especially the benzodiazepines and barbiturates. It might be not be available or less in zones of cortical harm. It is typically considered as an ordinary beat. It is the prevailing mood in patients who are alert, on edge or have their eyes opened. The noticeable brainwaves have been shown in Figure 1[4].

LITERATURE SURVEY
The universally institutionalized 10-20 frameworks are normally used to record unconstrained EEG. In this framework 21 electrodes are connected on the outside of

the scalp. An EEG test or electroencephalogram can take a gander at the electrical action of the brain and may help in deciding if a patient experiencing EEG test is encountering a seizure[13]. It might likewise help in anticipating whether more seizures will happen. Certain examples on the EEG are average of epilepsy. On the off chance that an individual's brain waves demonstrate those kinds of examples, the individual is about twice as liable to create epilepsy[14]. There are diverse mental phases of an epileptic patient as can be recognized from the adjustment in their brain waves and henceforth are pondered the EEG signals. These are: Ictal alludes to the physiologic state or occasion of a seizure. The word begins from the Latin ictus, which means a stroke or a blow. In (EEG), the account amid a seizure is designated "ictal"[15]. The accompanying definitions allude to the relationship of seizures with time. Pre-ictal alludes the condition just before the seizure, cerebral agony, or stroke, anyway it has starting late been discovered that a couple of qualities of this state, (for instance, Migraine air) demonstrate the start of the ictal state. Post-Ictal shows the state not long after the ictal state. Interictal alludes to the interval between seizures, that are portrays an epileptic issue. For most by far of the overall public of epileptic patients, the interictal on the scalp state exists for over 99% of their lifetime. This Interictal period EEG recording is broke down by numerous nervous system specialists to distinguish epilepsy in patients[16].

Epileptic patients experience the ill effects of different difficulties in their everyday life. They may experience different physical harms or mishaps at the season of the seizure in the event that they are occupied in specific undertakings which may include overwhelming machinery for example driving. Thusly discovery of seizure is critical for determining its motivation, root, and so forth with the goal

that legitimate treatment and prescription can be offered to those epileptic patients experiencing epilepsy and diminish the opportunity of physical harm or mishaps[18]. The customary method for identifying seizure in EEG signals is the manual skimming through 10s or even 100s of long stretches of recorded EEG motions alongside the requirement for a specialist for the undertaking. This is extremely a strenuous and tedious employment and furthermore inclined to human blunders. Subsequently auto location of seizure in EEG signals is essential. Numerous such strategies are accessible utilizing diverse methodologies for a similar issue. In this task we made a framework which dissects the EEG signals channel astute, and extricates highlights from them utilizing Discrete Wavelet Transform (DWT) trailed by nourishing those highlights to a neural network for discovery of seizure. The accessible EEG signals have been characterized into 3 classes-ordinary, interictal and ictal. Different classifiers have been tried against similar highlights extricated from the EEG signs to pass judgment on their relative execution.

PROPOSED WORK
Data Preprocessing
Data is preprocessed using standardization and normalization to get the proper accuracy after applying the algorithms. For the experimental setup the data label is divided into two classes that are classes with 5,4,3,2 are as non-seizure activity and 1 as seizure activity.

Dimension Reduction using Principle Component Analysis (PCA)
There are more than 170 features in the EEG dataset used for the experiment. For avoiding over fitting of the classification model reduction of size of the feature is necessary. PCA is used to reduce the dimensionality of a data set consisting of a large number of interrelated and dependent variables while retaining the variation present in the data set. This is achieved by transforming to uncorrelated features called as principal components, and which are ordered so that the first few retain most of the variation present in all of the original variables. First represent the independent features in terms of matrix. Find the mean of the matrix for each features results in a vector of features mean. Compute the covariance matrix of the whole dataset using the formula.

$$cov(X,Y) = \frac{1}{n-1}\sum_{i=1}^{n}(Xi - x')(Yi - y'),$$

Where, X & Y are the two features.

Computer eigen values and eigen vectors and sort the eigenvectors by decreasing eigen values and choose k eigenvectors with the largest eigen values to form a d × k dimensionalmatrix W. So W is the new dimension of the features.

Feature Extraction
Feature extraction is the process of selecting best features from the available features. Discrete Wavelet Transform (DWT) is used along with Genetic Algorithm (GA).

Discrete Wavelet Transform (DWT):
DWT is an implementation of the wavelet transform using a discrete set of the wavelet scales and translations obeying some defined rules. The discrete signal $x = \left[x[0],....,x[N-1]\right]T$, is a set of N samples taken from a continuous signal $x[m] = x+$ m Δt), (m = 0,1, …, N-1) for some initial time t_0 and sampling period Δt. The basic functions $\varphi = \left[\varphi[0],...,\varphi[N-1]\right]^T$ and $\omega = \left[\omega[0],...,\omega[N-1]\right]^T$ are also vectors containing *N* elements. Correspondingly the wavelet expansion becomes discrete wavelet transform (DWT). The discrete function is represented as a weighted sum in the

space spanned by the bases φ and ω.

Genetic Algorithm (GA)

GA is an optimization approach which is analogous to evolution process of Man and based on natural selection. At each step, it selects individuals at random from the current population to be root and uses them to produce the leaves for the next generation maintaining the parent child relationship. Over successive generations, the population "evolves" toward an optimal solution. Here the genetic algorithm is used as search problem for finding the best features from the data set. The initial population for EEG is a matrix of dimension population size. A fitness function is defined which selects the subset of the features extracted from DWT. The search for best features is carried out in all generations until the optimal solution is obtained or the maximum generation limit is reached.

Classification

The selected features and the label are given as input to the classifier with regularization for maintain the best fit model. Different classification algorithms are used for comparing the results. In this approach Random Forest classifier has been used which the best classifier is found out after experimentation. Random Forest is a decision tree algorithm. It is the group of different decision trees and the best one is selected. The nodes of the tree are calculated using different parameters like entropy and variance.

RandomForestClassifier(bootstrap=True, class_weight=None, criterion='gini', max_depth=None, max_features='auto', max_leaf_nodes=None, min_impurity_decrease=0.0, min_impurity_split=None, min_samples_leaf=1, min_samples_split=2,

min_weight_fraction_leaf=0.0, n_estimators=100, n_jobs=1, oob_score=False, random_state=None, verbose=0, warm_start=False)

EXPERIMENTAL SETUP AND RESULTS

Dataset

The first dataset from the reference comprises of 5 unique envelopes, each with 100 documents, with each record speaking to a solitary subject/individual. Each document is an account of mind movement for 23.6 seconds. The relating time-arrangement is inspected into 4097 information focuses. Every datum point is the estimation of the EEG recording at an alternate point in time. So we have all out 500 people with every ha 4097 information focuses for 23.5 seconds. The information is rearranged and each 4097 information focuses into 23 lumps, each piece contains 178 information focuses for 1 second, and every datum point is the estimation of the EEG recording at an alternate point in time. So now we have 23 x 500 = 11500 snippets of data (push), every data contains 178 information focuses for 1 second (segment), the last section speaks to the mark y{1,2,3,4,5}. The reaction variable is y in segment 179, the Explanatory factors X1, X2,..., X178, y contains the classification of the 178-dimensional information vector. Explicitly y in {1, 2, 3, 4, 5}. In the test setup we have isolated the information into two classes that is classes with 5,4,3,2 are as non-seizure action and 1 as seizure action.

Experimental Setup

Python numpy, pandas and scikit-learn packages have been used for experimenting on the database. For our experimental setup we have divided the data into two classes that is classes with 5,4,3,2 are as non-seizure activity and 1 as seizure activity. The data is divided into training data set (75%) and testing dataset (25%) with k-fold cross validation. The

database is then standardizes and feed to the feature extraction algorithm as per the above approach. The accuracy of the classification model is the calculated and compared.

Results

The results obtained are encouraging. Without even using a recurrent neural network, the Random Forest algorithm is able to correctly classify the EEG signals which detect the seizure and non-seizure activity. However, the best results were obtained when the network was trained on samples from the same recording session. While this may be practical for basic brain research, it would be less practical for use in BCI technology.This further recommends utilizing EEG will probably require an iterative methodology of preparing on a huge populace and after that adjusting on a particular person. It is consequently suggested that future research be done on the conceivable utilization of Transfer Learning methods to the order of EEG signals.

Different classification algorithms were compared like logistic regression, k nearest neighbor, SVM and random forest. Random forest will give the maximum accuracy when trained.

The performance is evaluated by confusion matrix and then finding accuracy.

 # Accuracy = TP + TN / TP + TN + FN + FP
accuracy = (cm[0][0] + cm[1][1]) / (cm[0][0] + cm [0][1] + cm [1][0] + cm[1][1])
 # Accuracy
*accuracy * 100*

CONCLUSION

EEG signal records were decomposed using multi-level DWT and further subjected to envelope analysis of the coefficients obtained from DWT. This preprocessing before feature extraction from the EEG signals particularly help in increasing the classification accuracy for detection of seizure. Finally the EEG records obtained from the database were classified into 2 classes: (i) Seizure, (ii) Non Seizure. Classifier used for classifying this data Random Forest showed promising results in terms of accuracy, sensitivity and specificity when verified using 75% of the available EEG data as training set and the remaining 25% as the test set. Different classification algorithms were compared like logistic regression, k nearest neighbor, SVM and random forest. Random forest will give the maximum accuracy when trained.

REFERENCES

1. Biomedical signal acquisition. https://www.medicine.mcgill.ca. Accessed: 2017-04-30.
2. The bonneeg database. http://epileptologiebonn.de/cms/frontcontent.php?idcat=193lang=3. Accessed: 2017−04−30.
3. Recognition of speech artifacts in eeg records. http://docplayer.com.br/18572674-Reconhecimento-de-artefactosrelativos-a-fala-em-registos-de-eeg.html. Accessed: 2017-06-20.
4. The-mystery-of-gamma-waves. http://q4lt.com/the-mystery-of-gammawaves. Accessed: 2017-06-20.
5. Lehnertz K. Rieke C. Mormann F. David P. Andrzejak, R. &C. Elger. Indications of nonlinear deterministic and finite dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. *Phys. Rev. E. v64. 061907.*
6. Bradley Lee Barnhart. The Hilbert-Huang Transform: theory, applications, development. *PhD thesis, University of Iowa*. 2011.
7. Dr Colin Tidy Dr Tim Kenny. Epilepsy with focal seizures. http://www.stillingtondoctors.co.uk. Accessed: 2017-04-30.

8. David Monaghan Edmond Mitchell &Noel E. O'Connor. Classification of sporting activities using smartphone accelerometers. *Sensors*. 2013; 13(4).

9. GoldbergerL., L. AmaralA. N., Glass L., HausdorffJ. M., IvanovP. Ch., Mark R. G., MietusJ. E., Moody G. B., PengC.-K., & Stanley H. E.. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation*.2000; June; 101(23): pp. e215–e220, Circulation Electronic Pages: http://circ.ahajournals.org/content/101/23/e215. full PMID: 1085218; Doi: 10.1161/01.CIR.101.23.e215.

10. Robert plonseyJaakkomalmivuo. Bio-electromagnetism: Principles and applications of bioelectric and biomagnetic fields. http://www.bem.fi/book/13/13.htm. 1995. Accessed: 2017-04-30.

11. Brain Y.,KabirInf., Zhang E.Siuly. World academy of science, engineering and technology.*International Journal of Medical, Health, Biomedical, Bioengineering and Pharmaceutical Engineering*. 2007; 1(3).

12. Brain Y. Inf. Kabir, Zhang E.Siuly. Epileptic seizure detection from eeg signals using logistic model trees. *Brain Informatics*.June 2016; 3(2): pp. 93–100.

13. Dr Tim Kenny. Epilepsy with focal seizures. https://patient.info/in/health/epilepsy-with-focal-seizures. 2015. Accessed: 2017-04-30.

14. Julian Dorado Cristian R. Munteanu Alejandro Pazos Ling Guo, ´Daniel Rivero. Automatic feature extraction using genetic programming: An application to epileptic eeg classification. Expert Systems with Applications: *An International Journal*.August 2011; 38(8): pp. 10425–10436.

15. S.N. Sharma Narendra Kumar Ambulkar. Detection of epileptic seizure in eeg signals using window width optimized s-transform and artificial neural networks. *In IEEE Transactions on Biomedical EngineeringBombay Section Symposium (IBSS)*. September 2015.

16. HasanOcak. Automatic detection of epileptic seizures in eeg using discrete wavelet transform and approximate entropy. Expert Systems with Applications: *An International Journal*.March 2009; 36(2): pp. 2027–2036.

17. NahidDadmehrSamanwoyGhosh-Dastidar, HojjatAdeli. Principal component analysis-enhanced cosine radial basis function neural network for robust epilepsy and seizure detection. *IEEE Transactions on Biomedical Engineering*.February 2008; 55(2): pp. 512–518.

18. Amit M. Shelat. Seizures. https://medlineplus.gov/ency/article/003200.htm. Accessed: 2017-04-30

Cite this article as: