Comparative Study of intelligent classifiers for EEG data

Dr. Anuradha Thakare[1], Tejas Chaudhari[2], Rushikesh Phalke[3] ,Shubham Chavan[4],Sopan Bembde[5]
[1,2,3,4,5](Department of Computer Engineering PCCOE, SPP University, Pune
Email: anuradha.thakare@pccoepune.org, tejasjc@gmail.com, rushikeshsp25@gmail.com,
shubhamchavan7177@gmail.com,sopanbembde@gmail.com)

ABSTRACT
This article presents comparison of four classifiers for EEG data. Algorithms are implemented, and results are compared with weka tool. It is observed that Linear regression results in less absolute mean error as compared to other classifier for EEG data. Further the classifier will be used to build computational model for high level applications

Keywords-*Decision Table, Electroencephalogram (EEG), Linear regression, Multiclass, Random Tree*

1. INTRODUCTION

Brain is best elaborated as interconnection of neurons which ultimately decides the human behavior [9]. Human course has been dealing with many brain related issues and this issue can be better solved by understanding the functional and cognitive behavior of human brain [10] [11]. Changes in brain activity can be determined by EEG that may be useful in diagnosing brain disorders. Alpha, beta, theta and delta waves are four major EEG waves [3]. The analytical study and occurrence of these waves helps in various purposes. In the raw EEG, there are contaminations called artefacts which are generally non-cerebral Signals. The minimization of these artefacts is very important as minute mistake in the interpretation of the signal may harm the results. Hence such EEG signals must be processed to become readable. To obtain actual signal various techniques are performed on EEG signals. The feature extraction is generally performed using statistical methods. And, after performing feature extraction, the signal is finally classified using intelligent classifiers.

An EEG may help for diagnosing or treating disorders like Epilepsy or other seizure disorder, Brain tumor, Head injury, Brain dysfunction that may have a variety of causes (encephalopathy), Inflammation of the brain (encephalitis), Stroke, Sleep disorders, Dementia.

This comparative study aims to determine which classification algorithm works best for given data set of EEGs. Also, a standard dataset of EEG is used for study and the results are analyzed using various intelligent classifiers.

2. RELATED RESEARCH

This section provides some of the existing works related to EEG dataset. *Giulia Fiscon et al* [2] analyzed the EEG signals using various supervised algorithms like Support Vector Machines (SVM), Decision Trees,Rule Based Classifiers and showed the advantages of Rule Based Classifier over SVM. *Saleha Khatun et al* [4] analyzed the MCI detection by using classification algorithms like SVM, Logistic Regression and Random Forest.

Norsiah Fauzana,et al [7] identified EEG features of MCI in comparison with the normal aging but the conclusion is drawn on the basis of results generated from manual comparison, no standard Intelligent Classifiers were used. *A.Nancy,et al* [1] Multi-level Pattern Learning (MPL) classification method to classify the abnormal category of EEG signal.

This Intelligent Classifiers used earlier possess few drawbacks like the performance of SVM classification is depends on the data that is used for both training and testing. The relative error obtained in each classifier was different. Aim of our work is to analyze the Results of Intelligent Classifiers by calculating the mean absolute error of Intelligent Classifiers.

3. INTELLIGENT CLASSIFIERS

This section introduces various intelligent classifiers used in classification of EEG dataset. One of the best ways of Classification of EEG dataset is done by using Weka Tool. Weka is a collection of machine learning algorithms. It contains tools used for pre-processing, classification and visualization. Different intelligent classifiers used in this work are Random Tree Classifier, Decision Table, linear Regression and Multiclass.

3.1 Random Tree Classifier

Random Tree is a supervised Classifier; it is an ensemble learning algorithm generating lots of

individual learners. It employs a bootstrap aggregating idea to construct a random set of data for constructing a decision tree. The random trees classifier gets the input feature vector, classifies it with every tree in the forest, and outputs the class label that received most "votes".

3.2 Decision Table Classifier

Decision tables are a concise visual representation for specifying which actions to perform depending on given conditions. They are algorithms whose output is in the form of a set of actions. The information expressed in decision tables could also be represented as decision trees or in a programming language as a series of if-then-else and switch-case statements.

3.3 Linear Regression Classifier

Linear regression is a linear approach for modelling the relationship between a scalar dependent variable y and one or more explanatory variables (or independent variables) denoted X. The case of one explanatory variable is called simple linear regression. For more than one explanatory variable, the process is called multiple linear Regression.

3.4 Multi Class Classifier

In Machine Learning, multinomial or multiclass classification is nothing but the problem of classification of instances into one of three or more classes. (Classifying instances into one of the two classes is called binary classification.) While some classification algorithms naturally permit the use of more than two classes, others are by nature binary algorithms; these can, however, be turned into multinomial classifiers by a different no. of strategies.

4. RESULTS AND ANALYSIS

For this comparative study TUH EEG Corpus Dataset is used. EEG contained in this dataset consist of 24 to 36 channels of signal data. It also contains an annotation channel containing markers identifying events of interest to the physicians and technicians. Signals sampling is at 250 Hz using 16 bits per sample. Around 20 Mbytes of data is present in typical EEG filet. These files are stored in a European Data Format (EDF+) file format. The intelligent classifiers are classified using Weka Tool. Weka is used for classification of various machine learning algorithms and data mining. The results obtained from weka are compared and based on its conclusion is made. As the training set is small 10-fold cross validation method is used to classify the data.

4.1 Results of Random Tree Classifier

Table 1 Results of Random Tree Classifier

Correlation coefficient	0.9577
Mean absolute error	7.1777
Root mean squared error	10.3006
Relative absolute error	28.3597 %
Root relative squared error	28.7803 %

The Random Tree Classifier implemented on the preprocessed dataset (section IV) took 1.49 seconds to build where the size of the tree was 83611. Based on the above table 1 we can observe that Relative absolute error for random tree classifiers is comes out to be 28.3597% and Root relative squared error is 28.7803 %.

4.2 Results of Decision Table Classifier

Table2 Results of Decision Table

Correlation coefficient	0.97
Mean absolute error	6.3209
Root mean squared error	8.76
Relative absolute error	24.9746 %
Root relative squared error	24.4758 %

The decision table classifier implemented on the preprocessed dataset (section IV) took 20.8 seconds to build which is highest among the classifiers taken into considerations in this comparative study. Based on the above table 2 we can observe that Relative absolute error for random tree classifiers is comes out to be 24.9746 % and Root relative squared error is 24.4758 %.

4.3 Results of Linear Regression

Table 3 Results of Linear Regression

Correlation coefficient	0.9747
Mean absolute error	5.728
Root mean squared error	8.0005
Relative absolute error	22.6317 %
Root relative squared error	22.3537%

Linear Regression classifier took 1.03 seconds to build model. The build time can be considered optimal among all as total instances were 75312.Based on the above table 3 it is observed that Relative absolute error for random tree classifiers is comes out to be 22.6317 % and Root relative squared error is 22.3537%.The obtained Relative absolute error and Root relative squared error depict that Linear Regression classifier gives lesser difference between the exact value and the approximation value among all the classifiers taken into considerations in this comparative study.

4.4 Results of Multiclass Classifiers

Table 4 Results of Multiclass

Correlation coefficient	-0.0041
Mean absolute error	21.4817
Root mean squared error	28.7975
Relative absolute error	100 %
Root relative squared error	100%

The Multiclass classifier implemented on the preprocessed dataset (section IV) took 0.53. Based on the above table 4 it is observed that Relative absolute error for random tree classifiers is comes out to be 100 % and Root relative squared error is 100 %.

Table 5 Performance comparison of classification algorithms

Intelligent Classifiers	Build time	Correlation coefficient	Root Mean Squared Error	Relative Absolute Error(%)
Random Tree	1.49	0.97	8.76	6.3209
Decision Table	20.8	0.9577	10.3006	7.1777
Linear Regression	1.03	0.9747	8.0005	5.728
Multischeme	0.53	-0.0041	28.7975	21.4817

The observation from above table are; Linear Regression has limited absolute error as compared to other intelligent classifiers.

5. CONCLUSION

This work presents comparison of classification methods for EEG data. Pre-processed dataset is used for experimentation. Four methods compared are Linear regression classifier, Decision table classifier, Random Tree Classifier and Multi class classifier. The results are tabulated and compared with 10-fold cross validations on weka tool.
It is observed that, performance of linear regression is better than other classifiers in terms of absolute mean error as shown in Table 5. The analysis given in results and analysis (section IV) will be used further for research in this area.

REFERENCES

[1] A. Nancy,Dr. M. Balamurugan and Vijaykumar SA BRAIN, (2017) ' EEG classification system for the mild cognitive impairment analysis ', 2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS).

[2] Basker E. (1998) 'Brain function and oscillations', Springer. Berlin Heidelberg,New York.

[3] Caton, Richard. (1985) ‘ Electrical currents of the Brain’, Chicago Journal of Nervous and Mental Disease.

[4]Giulia Fiscon,Emanuel Weitschek,Giovanni Felici,Paola Bertolazzi,Simona De Salvo ,Placido Bramanti and Maria Cristina De Cola, (2014) ‘Alzheimer’s disease patients classification through EEG signals processing", IEEE Symposium on Computational Intelligence and Data Mining (CIDM).

[5] Jasjeet Kaur and Amanpreet Kaur, (2015) ‘A Review on Analysis of EEG Signals",2 International Conference on Advances in Computer Engineering and Applications (ICACEA).

[6] Norsiah Fauzana and Nur Hurunain Amran, (2015) ‘Brain Dynamics of Mild Cognitive Impairment (MCI) from EEG Features’, Procedia - Social and Behavioral Sciences 165.

[6] N. Pradhan, D. N. Dutt, (1994) ‘Data compression by linear prediction for storage and transmission of EEG signals’, International Journal of Biomedical Computing, vol. 35, no. 3, pp. 207-217.

[7] Palaniappan R and Raveendran P , (2001) ‘ Cognitive task prediction using parametric spectral analysis of eeg signals’, Malaysian. Journal;1; of Computer Science. 14(1): 58-67.

[8] Saleha Khatun,Bashir I. Morshed and Gavin M. Bidelman, (2017) ‘Single Channel EEG Time-Frequency Features to Detect Mild Cognitive Impairment’,IEEE International Symposium on Medical Measurements and Applications (MeMeA)