ISSN 2228-9860 eISSN 1906-9642 CODEN: ITJEA8

International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies

http://TuEngr.com

Framework to Predict Epileptic Seizure Using EEG Signals

N Veena[1*], S. Mahalakshmi[1*], B E Abhijith[1], Adarsh Sadanand Shetty[1]

[1] *Department of Information Science and Engineering, BMS Institute of Technology & Management, Bengaluru, INDIA.*

**Corresponding Author (Tel:+91 8660180443, Email: veena @bmsit.in, maha.shanmugam @bmsit.in).*

Paper ID: 13A10G

Volume 13 Issue 10

Received 24 March 2022
Received in revised form 23 June 2022
Accepted 05 July 2022
Available online 14 July 2022

Keywords:
Electroencephalographic signals (EEG); Random Forest; Epilepsy; Seizure prediction; KNN; SVM; Hybrid algorithm; Detection accuracy of epileptic seizures.

Abstract

Epileptic seizures are neurological disorders seen in many people across the world. There are nearly 10 lakh cases recorded globally every year for this disease. People who are suffering from this disease may cry out, fall, shake or jerk, and become unaware of what is going on around them. Preventing such conditions is very important. We use soft computing methods to predict epileptic seizures from Electroencephalograms (EEG) signals, so that appropriate medication can be suggested. This paper deals with a software tool through which this condition can be predicted and identified, the software tool basically provides an interface for doctors to pass the EEG Signals in the overall seizure prediction process. This paper also deals with a comparative analysis of various algorithms such as Random Forest, KNN (K Nearest Neighbors), SVM (Support Vector Machine) to train the model.

Discipline: Brain Science & Neuroengineering (Electroencephalography).

Cite This Article:

Veena, N., Mahalakshmi, S., Abhijith, B E., Sadanand, A. S. (2022). Framework to Predict Epileptic Seizure Using EEG Signals. *International Transaction Journal of Engineering, Management, & Applied Sciences & Technologies*, *13*(10), 13A10G, 1-10. http://TUENGR.COM/V13/13A10G.pdf DOI: 10.14456/ITJEMAST.2022.196

1 Introduction

Epilepsy is a brain disease that causes uncommon and erratic behavior in the human body (Yuan *et. al.,* 2010). It occurs due to the change that occurs in the brain. Since the brain is the most important organ of our body that controls the entire human activity, any small change that occurs in the brain might cause a severe effect on our body. Each and every cell conducts some value of electric charge that moves around passing messages. Any irregularity can end up causing one or more deterioration in the condition of the human.

Epilepsy is known to cause random, uncontrollable and frequent convulsions either focused on a part or throughout the whole body. Presently, the diagnosis and analysis of epilepsy are mainly done through electroencephalogram (EEG) i.e., it measures the electrical potential of neurons in the cerebral cortex (Zubai *et. al.,* 2021). Manual identification and analysis are both difficult for the human eye and time-consuming even when done by an experienced neurologist. The availability of a specialist to do the above also affects the count of successful identifications and treatment of patients (Zhao, *et. al.,* 2020).

In order to identify this condition electroencephalographic signals are used. These signals are directly taken from the brain. Obtained signals are in the frequency format and depending on the frequency these signals have various classifications such as alpha (8-12Hz), beta (12-40Hz), gamma (40-100Hz), theta (4-8Hz), delta (0-4Hz). each of these are responsible for different activities that are produced in our body (Acharya, *et. al.,* 2018). However, if any of these waves are over-produced or under-produced then it leads to severe brain problems.

The signals that are obtained from the brain are used for training the model using various algorithms and once the model is trained it is embedded into the software thereby providing a software interface that predicts the epileptic seizure condition.

The tool is designed to identify epileptic wavelets from normal wavelets. By using ML algorithms and webapps, the tool works by taking in data and predicting the possible status conditions.

Even if epilepsy and seizures are sometimes used synonymously in common literature (Chaovalitwongse *et. al.,* 2018), it is worth noting that not all seizures are epileptic and convulsions and seizures may also be caused due to acute neurological reasons without necessarily reflecting a long-term predisposition to recurrent unprovoked seizures (Rasheed *et. al.,* 2020). The existing system for the diagnosis and identification are manual. There exists no tool that can simplify the process.

Different non-invasive devices EEG devices with invasive devices as discussed (Veena *et. al.,* 2020) can be used to collect the real-time EEG signals. Cloud-based remote diagnoses can be done after collecting the signals (Veena *et. al.,* 2018) after which medication can be provided based on the severity of the disease to the patient (Veena *et. al.,* 2018) to refer them to the appropriate hospital if the disease is serious.

EEG signals recorded, before and during a seizure, contains characteristics that can be used to identify the different stages of an epileptic seizure, and the pre- and post-seizure periods. These stages are briefly described below (Zhou *et. al.,* 2018).

- Pre-Ictal phase: This is the time before seizures. It can last from minutes to days.
- Pro-Ictal State: In this state, the seizures are more likely but not guaranteed to occur.
- Ictal and Interictal State: The period between the seizures or convulsions can be termed as Ictal/interictal state. The period or span of a seizure may vary based on the number of epileptogenic neurons in the cortical region of the brain.

• Post-Ictal State: The altered state of consciousness after an epileptic seizure is termed as post-ictal state. It usually lasts for 5-30 minutes.

2 Literature Review

The survey is done by collecting data on existing and proposed systems to deal with epilepsy. Going through multiple published papers and analyzing the suitable methods of detection.

Yuan *et. al.* (2010) detected epileptic seizure based on EEG signals which entails the details of epilepsy and its cause. It explains in detail how the raw data can be collected and with a 16-channel signals and preprocess. Using a support vector machine to build a model for detection and the results from the model and compares it with a probabilistic neural network to find the pros and cons.

Rasheed *et al.,* (2020) discussed and reviewed machine learning to predict epileptic seizures using EEG signals, which goes into a review study of multiple papers and different approaches that are given by various researchers and authors. Explaining the various features and methodologies like EEG signals, ES prediction, DL for EEG and feature extraction. Techniques like time domain, frequency and time-frequency domains are discussed.

Zubai, *et. al.* (2021) analyse the time prerequisite domain feature of EEG signals extracted through DWT and then reduce the dimensionality using SPPCA and SUBXPCA. Then these are fed to ML model to classify whether signals are epileptic or non-epileptic. The dataset is from the department of epileptology at the university of Bonn, Germany. The models used were SPPCA (spare principal component analysis) and SUBXPCA (sub-pattern principal component analysis) to reduce the dimensionality and increase the accuracy. Cat boost classifier and random forest classifier. Finally, the SPPCA gives 97% accuracy using "cat boost classifier" and SUBXPCA gives 98% accuracy using "Random Forest classifier". We see that the datasets are being reduced to smaller sets and no software tool developed.

Sharma *et al.* (2018) locates the epileptogenic region (and neurologist in decision making) and predict an upcoming seizure. The dataset is from physionet website recorded by placing 23 -electrodes as per 10-20 electrode placement system. Here the ANN (Artificial Neural Network) algorithm is used for training of data for epileptic seizure prediction. The processing of EEG signals is done by CAR (Common averaging reference), SL (Surface Laplacian) , ICA (Independent component analysis) . Resulting in an accuracy of 92.3% achieved using ANN. We see that a higher accuracy can be achieved through other means and the validation of data is not done.

Acharya *et al.* (2018) developed a computer-aided diagnosis system that can automatically categorise the class of EEG signals using ML based techniques. With the implementation of CNN algorithm to detect seizure classes and pre-processing that is done by using CNN and ANN. The dataset used is from Bonn university, Germany. The dataset was divided into two parts (30% training, 90% validation). The results indicated that 88% of the test data was classified without Errors but the 12% was classified incorrectly as safe. The gaps identified are that error in prediction is a sensitive issue as it deals with a medical condition.

Zhou *et al.* (2018) developed a convolutional neural network model to detect different phases in epilepsy, by using EEG signal data from 128-channel Neurofile NT EEG machine. The CNN model which was employed had 3 layers, via these layers all features/attributes were extracted. Results were found satisfactory. the gaps identified in this paper are that inconsistent EEG signals lead to false predictions.

Hussein *et al.* (2019) included deep neural spec for the detection of epilepsy. The authors obtained a dataset from the Bonn University database, which had 5 sets of data. The signals were amplified for clear identification. A sturdy deep neural spec i.e., RNN was used in this study. The authors explained how RNN was used to identify the epileptic condition. Results were not affected by the noise in the data. The drawback of this study is that the data initially was collected from a single electrode EEG, slight modification to accommodate a multi-channel EEG system was needed.

Amin *et al.* (2020) developed a robust and efficient model using computer-aided design technique (CAD) that's applied to the EEG signals to differentiate the seizure and normal activity, the author has used the dataset from the university of Bonn which contains 4097 samples with a length of 23.6 sec and with 5 labels, the model is developed using support vector machine (SVM), Multi-layer perception and other algorithms. But there are certain major gaps such as 1. The model does not use any Real-time data 2. The model does not discuss the future enhancements

Boonyakitanont *et al.* (2018) summarized various features and compare various algorithms to predict epileptic seizures using EEG signals and also review their performance. The paper uses a dataset from Bonn university which has 100 signal channel samples. the modelling is done using some of the major algorithms such as naïve bayes and others. CFS is used for performance analysis and redundancy reduction.

The major gaps that are identified in this study are

1. no proper validation
2. Major focus is only on feature selection not on model development

Zhao, *et al.* (2020) introduced the LSTM networks into a CNN to participate in the classification of Epilepsy. Used here is the data from the open CHB-MIT Scalp EEG database. The raw data input undergoes feature extraction and is then forwarded into the LSTM network which predicts the seizure with patient-specific particularities. The model is tested for multiple different durations of data and comes to a conclusion of a better predictive value in each window. The model also outperforms traditional ML algorithms like SVM, RIPPER and Decision tree by significant value and lesser false alarms. The model has a good response for the test dataset but needs to undergo testing with other datasets to validate the performance.

Zhao *et al.* (2020) suggest a novel CNN model to help in robust Detection. The dataset used for training and testing is the Bonn University Database, which has 5 sets. The proposed model has 6 layers, the three 5-Layered Convolutional blocks and 3 FC layers that take in raw EEG data and perform feature extraction and classification respectively. The performance of the model is compared with other classification models for certain dataset classes. A few gaps could be identified as the model ran well for the Bonn Dataset but with learning rate and training, time would increase highly with the increase in dataset size.

3 Method

3.1 Existing Algorithms

The data set we choose has to undergo all the preprocessing to make it suitable so that this can be fed as an input to the tool. The data set should be in the format of frequency data points, that is, the data from the wave format is to be converted into frequency values with each data point having a value.

The interface is designed with a simple and easy of use design for quick and easy understanding.

A support vector machine (SVM) constructs a model which decides into which category a given data should be mapped. This model usually has the ability to work with a huge number of predictors, this is applicable to biomedical models.

K-Nearest Neighbor (KNN) constructs a classification model that makes categories and maps the values onto them based on the vote of its neighbors. This has a good accuracy value due to its cross-validation from training.

Random Forest (RF) constructs a model similar to a random tree. Each data value is best matched with training data and results in an ensemble of random trees. The recognition accuracy is higher because the individual trees vote for the popular class.

3.2 Hybrid Algorithms

3.2.1 Hybrid Model#1

This builds a model with multiple algorithms combined and stacked into both sub and meta-classifiers. Each of the sub-classifying models is trained separately to give a running model which feeds its output into the meta classifier which is trained to predict the final result. The Hybrid1 model uses KNN and RF as the respective sub-classifiers which are directed into the meta classifier with a trained SVM model as in Figure 1.

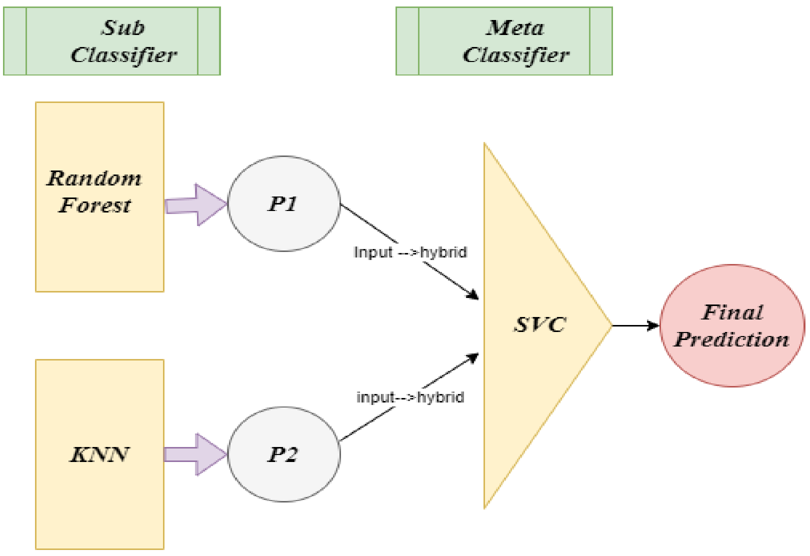

Figure 1: Hybrid Model#1

3.2.2 Hybrid Model#2

This model is constructed based on the logic of the ensemble voting technique. This technique involves taking in predictions from multiple models as individuals and letting the voting classifier predict the result by inferring the results of all the participating models. This hybrid model takes each input as a vote and decides the final result based on the majority weight as depicted in Figure 2.

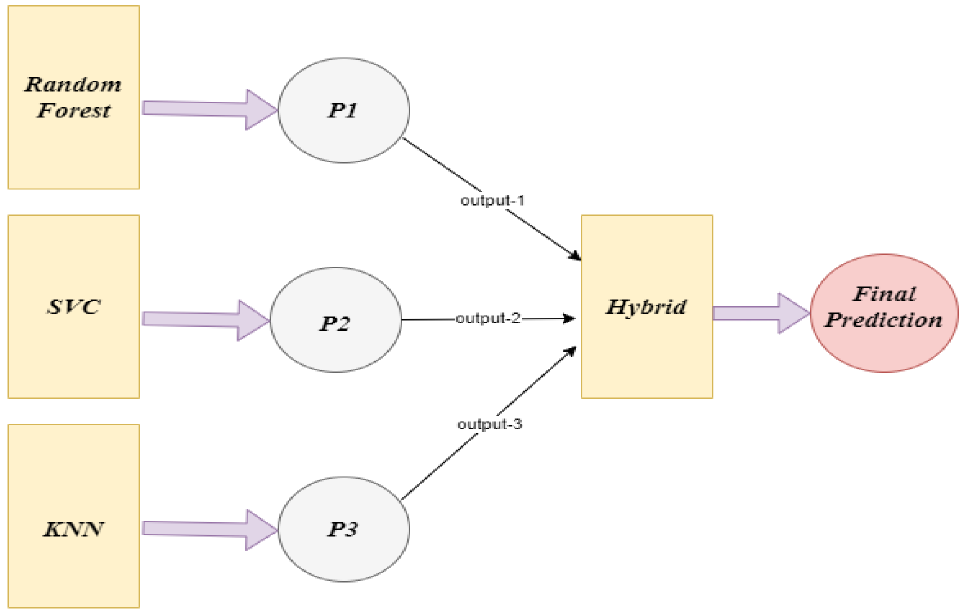

Figure 2: Hybrid Model#2

3.3 Accuracy of Algorithms

Support vector machine, K-Nearest Neighbor, Random Forest and the Hybrid models were the algorithms that were used to train models for supervised learning and the accuracy was noted. The training and execution time plays a role in deciding which algorithm to be used but accuracy has a vital role in the medical field.

All the models are trained with the training data and test data was fed to get the accuracy of each. SVM and KNN and Hybrid#1 had decent accuracies. Hybrid#2 (92.23%) outperformed the previous three with an accuracy of 95.14%. Random Forest with the highest accuracy amongst the above was a good fit with 98.53% and lesser runtime.

3.4 Machine Learning Models

EEG signals have been and are one of the most important tools for monitoring brain activity. Monitoring signals before the onset, during, or after an epileptic seizure and analyzing the data recorded can play a major role in helping control the disease and its treatment.

The data is to be converted into the preferred and understandable unit and needs to undergo cleaning. In the preprocessing phase, the selective noise and artifacts need to be identified and removed. This is a crucial part of raw data collection. The data is also to be normalized and corrupted parts are to be rectified. Out of the 4000 data points, not all will be acceptable and clear. The cleaning should be done so that the model does not face any error while predicting. Corrupt data may clutter itself and change the predictions.

The data is to be analyzed and undergo feature extraction and selection. This can be done using many methods like time or frequency domains, principal component analysis, or wavelet transform. The feature decides whether the mass heaps of data are needed and safe for analysis. The selected data blocks are recorded in a file that will be the input for the trained model. The model classifies the data points and predicts the verdict of the input data.

With Exploratory Data Analysis (EDA), the visualization and understanding of the data. The model deals with the data and its set parameters to get an accuracy score and prediction list that is stored in a data frame for further reference.

The model produces the output for the data points and predicts the status of each wavelet. The following Figure 3 shows how the data chunk looks when plotted as waves for visualization.

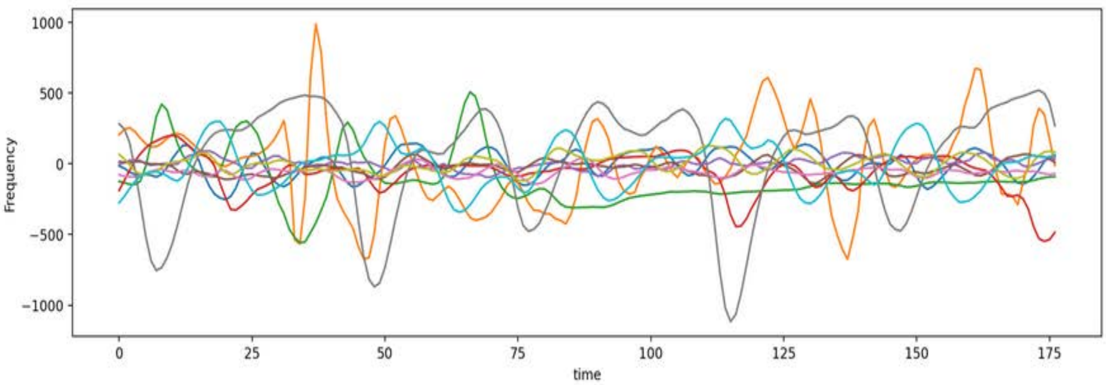

Figure 3: Data visualization

4 Result and Discussion

The data set, and the algorithm is combined together to build a model. The data is randomly divided into two parts i.e., training and testing datasets. The model is built using the Random

Forest algorithm and trained with the training dataset. This training enables the model to draw inferences and predict other given data using the previous knowledge learnt.

23 chunks of data are inputted to the model, which reads each second of the EEG signal. The model returns the accuracy and predictions. The result is then passed to the tool to be displayed and restructured into understandable bits.

The software tool has the trained model implemented into itself and acts as the interface between its intended end user, i.e., the physician and the prediction model. With a statistical output formatted in an understandable layout, the tool increases the efficiency of the detection process. The tool is designed statically with HTML & CSS and with flask as in Figure 4, which makes it a webapp capable of displaying the data into the app and presenting it in a cleaner order.

Table 1 shows the result score and accuracy of the model trained. The tool provides a simple login window for the user to authenticate themselves and attach the data file to be analyzed for prediction. By moving along the path, the user is greeted with the dashboard featuring the results of the model. The dashboard includes the total values of epileptic and non-epileptic wavelets and individual classification. The output is represented in a statistical layout for easier understanding and recognition. The user may choose to generate a report and save the results of the model.

Figure 5 shows the result after using Random Forest in training the model

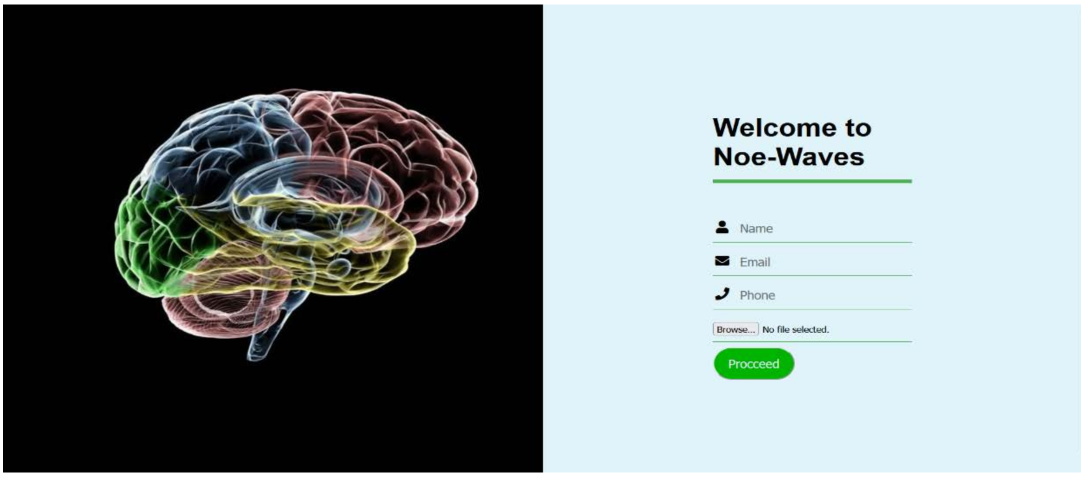

Figure 4: Login page to add the EEG signals

Table 1: Accuracy of Algorithms

Algorithms	Score %	Predictions
LogisticRegression	57.14	[0 1 1 … 0 0 1]
SVC	91.32	[1 1 1 … 0 1 0]
DecisionTreeClassifier	91.01	[1 1 1 … 0 1 1]
RandomForestClassiFier	98.53	[1 1 0 … 0 1 0]
KNeighborsClassifier	90.34	[1 1 0 … 0 1 1]
Hybrid#1Classifier	92.23	[1 1 0 … 0 1 0]
Hybrid#2Classifier	95.14	[1 1 0 … 0 1 0]

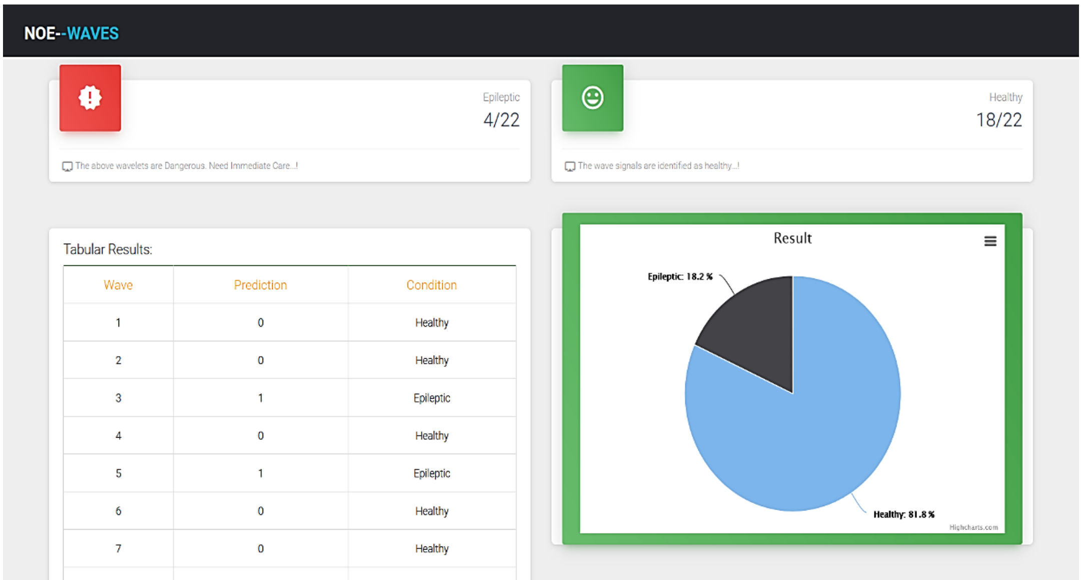

Figure 5: Prediction Result

5 Conclusion

This paper highlights the various algorithms that are used to solve the detection of Epileptic Seizures and the best models to work with for high accuracy results. This implementation of a tool makes the process universal in use. The software tool designed with the model and interface provides the end-user with a tool to run their collected data on the designed model and successfully give a result with high accuracy.

The accuracy of 98% is a very good rate by which the identification and classification can be taken as a final test. This will ease the examination and identification process leading to a simpler and more efficient mode of detection.

This model can further be enhanced to be used in the actual real-time system to provide the user with a quicker and easier use case experience. The model can be further trained and tested with real-time data which would help in increasing the efficiency further leading to even more accurate predictions.

There is always scope for increasing the accuracy of the model. In the field of medical science, accuracy matters the most. Even a small improvement can have a huge impact on the outcome. Implementation of Neural networks and better hybrid models can help in increasing the accuracy and the speed of prediction. reported and emphasized.

6 Availability of Data and Material

Public data set is been taken for the analysis from kaggle.

7 Acknowledgement

We thank the BMS Institute of Technology & Management for providing the facilities to complete the research work.

8 References

Acharya, U. R., Oh, S. L., Hagiwara, Y., Tan, J. H., & Adeli, H. (2018). Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals. *Computers in biology and medicine*, 100, 270-278.

Amin, H. U., Yusoff, M. Z., & Ahmad, R. F. (2020). A novel approach based on wavelet analysis and arithmetic coding for automated detection and diagnosis of epileptic seizure in EEG signals using machine learning techniques. *Biomedical Signal Processing and Control*, 56, 101707.

Boonyakitanont, P., Lek-Uthai, A., Chomtho, K., & Songsiri, J. (2020). A review of feature extraction and performance evaluation in epileptic seizure detection using EEG. *Biomedical Signal Processing and Control*, 57, 101702.

Hussein, R., Palangi, H., Ward, R. K., & Wang, Z. J. (2019). Optimized deep neural network architecture for robust detection of epileptic seizures using EEG signals. *Clinical Neurophysiology*, *130*(1), 25-37.

Rasheed, K., Qayyum, A., Qadir, J., Sivathamboo, S., Kwan, P., Kuhlmann, L., ... & Razi, A. (2020). Machine learning for predicting epileptic seizures using EEG signals: A review. *IEEE Reviews in Biomedical Engineering*, 14, 139-155.

Sharma, A., Rai, J. K., & Tewari, R. P. (2018). Epileptic seizure anticipation and localisation of epileptogenic region using EEG signals. *Journal of Medical Engineering & Technology*, 42(3), 203-216.

Yuan, Y. (2010). Detection of epileptic seizure based on EEG signals. In *2010 3rd international congress on image and signal processing* (Vol. 9, pp. 4209-4211). IEEE.

Zhao, W., Zhao, W., Wang, W., Jiang, X., Zhang, X., Peng, Y., ... & Zhang, G. (2020). A novel deep neural network for robust detection of seizures using EEG signals. *Computational and mathematical methods in medicine*.

Zhou, M., Tian, C., Cao, R., Wang, B., Niu, Y., Hu, T., ... & Xiang, J. (2018). Epileptic seizure detection based on EEG signals and CNN. *Frontiers in neuroinformatics*, 12, 95.

Zubair, M., Belykh, M. V., Naik, M. U. K., Gouher, M. F. M., Vishwakarma, S., Ahamed, S. R., & Kongara, R. (2021). Detection of epileptic seizures from EEG signals by combining dimensionality reduction algorithms with machine learning models. *IEEE Sensors Journal*, 21(15), 16861-16869.

Dr. Veena N is an Associate Professor at the Department of Information Science and Engineering, BMS Institute of Technology & Management (BMSIT&M), Bengaluru, India. She got her B E, Master's and Ph.D. degrees from Visvesvaraya Technological University (VTU), Belagavi, Karnataka, India. She secured 2nd rank in master's from VTU in 2008. Her areas of interest include Brain Computer Interface, Machine Learning and Artificial Intelligence.

S Mahalakshmi is pursuing her Ph.D. at Vellore Institute of Technology, Vellore, India. She got her B.E from Jayaram College of Engineering and M.E from Bannari Amman Institute of Technology. She is a rank holder and received gold medal in her PG from Anna University, Chennai and won Best Student 2008 Award. She got a certificate on CCNA Exploration from Cisco Networking Academy and a workshop on High impact teaching skills from Dale Carnegie. Her areas of interest include Soft Computing Techniques, Brain-Computer Interface and Image Processing.

B.E Abhijith is a Final-year student, in Information Science &Engineering, BMS Institute of Technology & Management, Bengaluru. His interests include python programming, flutter app development and mentoring. He works on projects and has organized several workshops and special sessions.

Adarsh Sadanand Shetty is a Final-year student in Information Science & Engineering, BMS Institute of Technology & Management, Bengaluru. He has worked on App Development, Python programming, and Arduino devices. His area of interest includes Game Development and Machine learning.