2019 IEEE Bombay Section Signature Conference (IBSSC)

Detection and Prediction of the Preictal State of an Epileptic Seizure using Machine Learning Techniques on EEG Data

Manasvi Bhat K[1], Pratiksha P Anchalia[2], Yashashree S[3], Sanjeetha R, Anita Kanavalli
M S Ramaiah Institute of Technology, Affiliated to VTU, Bangalore, India
[1]kmanasvibhat@gmail.com, [2]pratikshapanchalia@gmail.com, [3]yashashreesuresh@gmail.com

***Abstract*—Epilepsy, a disorder that leads to abnormal activities in the brain is primarily caused by excessive neuronal activity. Patients diagnosed with epilepsy frequently suffer from seizures, the impact of which may vary from abnormal body movements to alterations in the levels of consciousness. An appropriate dosage of medication provided at the right time can help prevent an impending seizure. In this paper, real data obtained from Epilepsy Ecosystem is used for analysis. After preprocessing this data, several signal processing algorithms and mathematical computations are used for feature extraction. Two sets of features are identified viz. lasting features and transitory features. Several combinations of these features along with Machine Learning algorithms such as Extra Trees Classifier and XGBoost are used to train generalized models as well as a patient-specific models, both of which are immune to noise. It is observed that the XGBoost based generalized model which is trained using lasting features gives a relatively better accuracy of 90.41%.**

***Keywords*—epilepsy, epileptic seizure, detection, prediction, machine learning.**

I. Introduction

A seizure is described as a disturbance in the brain that is uncontrolled and sudden and is primarily caused by excessive neuronal activity in the brain. Closely related to the occurence of a seizure is epilepsy, which is a neurological disorder characterized by a seizure. The impact of a seizure on a patient may vary from abnormal body movements to loss of consciousness. Although high dosage of medication is usually effective in preventing an impending seizure, patients are often found to suffer from the side effects of these medicines. Additionally, for about 20-40% of patients, medication is found to be ineffective. For a considerable proportion of patients, spontaneous seizures continue to occur even after being surgically removed. Despite the infrequency of seizures, the likelihood of an impending seizure instills an elevated sense of anxiousness in the patients.

An epileptic seizure is primarily characterized by four stages, namely Preictal, Ictal, Interictal and Postictal. While the Preictal state appears prior to the commencement of the seizure, the Ictal state marks the beginning of the seizure. Postictal state, as the name suggests, commences post the Ictal state. Finally, the Interictal state begins post the Postictal state of the first seizure and terminates before the beginning of Preictal state of the next seizure. Out of the above mentioned four states, the Ictal and the Preictal states are of utmost importance due to the fact that the Ictal state assists in classifying seizure and non-seizure EEG signals, and the detection of the Preictal state that occurs several minutes before the commencement of the seizure predicts the seizure.

II. Literature Survey

Falco-Walter et al. discuss the upgraded classifications of epilepsies and seizures. These new classifications include some previously unclassifiable seizure types. In their proposed method they include two main seizure types namely basic and expanded. Doctors and Neurologists can make use of the basic classification whereas the expanded classification will be preferred by researchers and epileptologists. These new classifications have to adopted and used regularly to maximize their benefits [1].

Ihsan et al. in their research work establish that the existing methods do not perform well when classification of the ternary scenario is concerned. The known methods give a maximum accuracy of 97±1% for this case by the is. They overcame this problem by proposing a system which has deep learning as its foundation. The system consists of an ensemble constructed out of convolutional one dimensional neural networks (P-1D-CNN). The performance of the system is commendable when used on a benchmark dataset. To clinically validate this model is an area of future work. Incorporation of the proposed system on a wearable device can also be considered as the future scope [2].

Hussein et al. identify the challenges encountered by EEG based seizure detection systems in their research work. EEG signals are not stationary and are prone to several noise types. To overcome these challenges they introduce a deep learning based approach that automatically learns the distinct EEG features of epileptic seizures. The EEG data that is in time series format is first transformed into segments of non overlapping epochs. Secondly the Long Short-Term Memory (LSTM) network learns seizure EEG and normal patterns. Thirdly these representations are given as input to a Softmax function for classification and training. The results are tested against benchmark clinical datasets and it is found that the proposed approach is superior to the existing approaches. Compared to current methods the proposed method withholds its performance in the presence of white noise [3].

Usman et al. identify feature extraction and noise removal as the two major issues that have a negative impact on prediction rate and anticipation time. Their model provides methods that are reliable for both feature extraction and preprocessing. It detects and predicts epileptic seizures in time before the commencement of the seizure. The application of the empirical mode decomposition for extracting time and frequency related features has made the model robust. Thus, their model predicts the occurrence of the Preictal state with a better accuracy than the existing methods [4].

Mengni Zhou et al. state that the use of Convolutional Neural Networks rather than manual feature extraction can be of immense help to distinguish between the Preictal, interictal, and Ictal states of an epileptic seizure. The proposed CNN model was simple and avoided the complicated feature extraction procedure. They intend to design a forecasting model in the future [5].

Cao et al. address the issue that the existing method of epileptic seizure detection could not solve, the procedure to manually select the features. To overcome this problem they propose a system to detect an epileptic seizure using Convolutional Neural Networks and Short Time Fourier Transform. The proposed system is efficient while dealing with multi channel data [6].

Tsiouris et al. make use of the Long Short Term Memory (LSTM) networks to detect and predict epileptic seizures. The proposed model takes into consideration time and frequency features. The clinical validation of this technique is pending [7].

Han-Tai et al. in their research work propose a hybrid model consisting of Particle Swarm Optimization (PSO) and Genetic Algorithms (GA's). It decomposes the EEG signals into time and frequency sub bands and makes use of differing time scales for testing and training. The proposed system uses subject specific modelling. More work is required to decrease the run time of hybrid models [8].

III. Materials and Methods

A. EEG Database

The dataset that has been made use of for the purpose of this study is obtained from Epilepsy Ecosystem [9]. An implanted device consisting of 16-electrodes recorded the brain activity of a number of patients in the form of intracranial EEG (iEEG) by positioning the electrodes on the cerebral cortex, wherein the sampling was performed at 400 Hz. Since the purpose of the dataset is to pose as an effective evidence to support the premise that deployment of Machine Learning models to the prediction of occurrence of epileptic seizure is a feasible task, the recorded brain activity of three patients for whom the prediction of the seizure proved to be most challenging is taken into consideration. The potential of a trained model to accurately predict the seizure occurrence for these patients is indicative that the technique can be applied to a much wider range of patients.

The database of refractory epilepsy patients has been classified into a series of data clips of Preictal as well as Interictal states, each spaced by 10 seconds and each being 10 minutes in length. Only the seizures which are separated by intervals of at least 4 hours, termed as lead seizures, have been taken into account. Additionally, a five-minute seizure horizon has been provided for the data segments of Preictal state with the purpose of providing a reasonable amount of time between the stages of prediction of seizure and the occurrence of it.

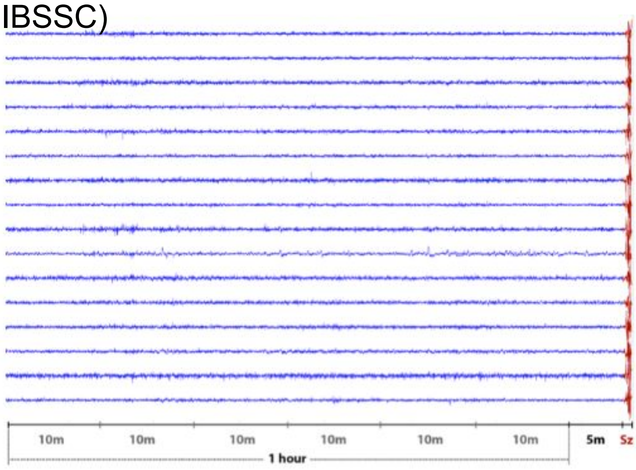

Figure 1. A sample of EEG data belonging to the Preictal state collected via 16 electrodes

Patient	Recorded time (in days)	Seizures	Lead seizures
1	559	390	231
2	393	204	186
3	374	545	216

Table 1. Characteristics of the dataset

B. Machine Learning Algorithms

Machine Learning refers to the capability of a system to automatically draw inferences from data without the requirement of the system being programmed to do so. The level of precision that can be achieved through the use of these techniques is remarkable in comparison with what is achievable merely through manual effort for the same. This is of paramount importance especially in the fields such as that of medicine, wherein the predictions of a model have a significant impact on the health of a patient and may even be a matter of life and death.

The focus of this work is on Supervised Learning, which is a type of Machine Learning wherein there exists a mapping between the input set of features and the output labels, which are automatically learnt by a model so as to be able to make predictions regarding the output given any previously unseen set of inputs, while ensuring that the accuracy obtained in such a process is maximized.

IV. Seizure Prediction

An outline of the seizure prediction system implemented in this study has been depicted in the Figure 2.

A. Data Preprocessing

The raw data measured by the implanted device, which is representative of the brain activity of an individual, is often prone to several errors, the approximation in the magnitude thus measured merely being one of them. Consequently, in order to diminish the influence of these errors on the results of a Machine Learning model, the application of various data transformation techniques is indispensable.

In case of classification problems, when the proportion of data points belonging to a particular class varies significantly with respect to the proportion of those belonging to another class, the predictions obtained using a model which has been trained on such a dataset is likely to

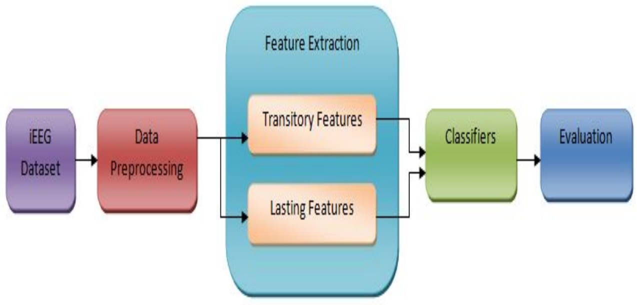

Figure 2. Outline of Seizure Prediction System

be biased towards the class with a greater proportion of data points. The existence of such an imbalance, termed as the class imbalance problem, in the dataset under consideration necessitates selection of a subset of the dataset such that the proportion of the data points belonging to both, the Preictal and interictal states, are approximately equal.

B. Feature extraction

As a result of the organization of the dataset into a series of data clips, each being 10 minutes in length, a number of statistical as well as temporal parameters are extracted from each of these data clips through the use of several signal processing algorithms and mathematical computations. For the purpose of this study, the features thus generated have been classified as the lasting features and transitory features.

The computation of the eigenvalues and correlation coefficients at complete temporal resolution for each of the 16 channels in each 10-minute long data clip results in the generation of the two lasting features, following which downsampling is performed by a factor of five. At the downsampled temporal resolution thus obtained, on a per-channel basis, several measures such as the standard deviation, number of zero crossings, kurtosis and skew as well as the first and second order derivative standard deviations are obtained on a channel by channel basis. After successful transformation of the features to the spectral domain, additional metrics such as the normalized summed energy, total summed energy, maximum frequency and entropy in bands of [0.1, 4, 6, 12, 30, 40] Hz are computed.

For the purpose of generation of the transitory features, setting the overlap time as 30 seconds, a series of bands are derived in the spectral domain on a minute by minute basis, following which the computation of the summed energy bands at the frequencies of [0.1, 4, 8, 12, 30, 70, 180] Hz is performed. The transitory features essentially represent the band energies wherein within every temporal segment, for each channel and for each band number, a corresponding feature is generated.

In addition to the above, processing of the data is performed to ensure that the mean value is zero and the standard deviation is unity on a per-channel basis. The Welch method is made use of in the generation of the power-spectrum features in order to set the noise contribution to minimal.

C. Classification

In this study, a variety of combinations of the generated features, i.e, the lasting features as well as the transitory features, and various Supervised Learning algorithms have been made use of to compare and contrast the performance of each of the models thus developed. The classification algorithms that have been implemented include:

- Support Vector Classifier
- Random Forest Classifier
- Extra Trees Classifier
- AdaBoost
- XGBoost

Using each of the above-mentioned classifiers, not only a generalized model has been trained, but also a patient-specific model, which learns the patterns in the brain activity of a specific patient during the Preictal and interictal states in order to predict an impending seizure for the particular patient.

V. Results

A. Prediction using a generalized model

The recorded brain activity of all the three patients is made use of to train a generalized model in order to make predictions pertaining to an impending seizure by making use of both, lasting as well as transitory features. The generated model is capable of determining the Preictal state for a test set that could belong to either of the three patients, indicating that apart from being able to predict seizures for the patients based on whose data the model was initially trained, it may additionally be used to make predictions for a relatively wider audience.

Among the various classifiers taken into consideration, it is found that XGBoost, which is a popular form of a gradient boosting algorithm operating by penalizing complex models through both L1 and L2 regularization, results in the highest accuracy of 88.09%. The resulting accuracy obtained through this algorithm may be attributed to the capability of the classifier to handle sparse data, which may be generated due to the presence of missing values as well as the pre-processing steps, such as one hot encoding.

B. Prediction using a generalized model trained using transitory features

In an attempt to determine the contribution of transitory as well as lasting features in case of a generalized model developed using both types of features, only the transitory features obtained through the application of the feature generation techniques for all the three patients are taken into consideration to train another generalized model. A comparative study between the accuracy obtained using such a generalized model with that obtained using a generalized model developed using both types of features indicates that the latter results in a relatively higher accuracy. Based on this observation, it may be hypothesized

978-1-5386-7401-7/19/$31.00 ©2019 IEEE

that the higher accuracy obtained using the generalized model developed using both types of features may be attributed to the contribution of the lasting features to the prediction outcome.

C. Prediction using a generalized model trained using lasting features

In order to validate the above hypothesis, only the lasting features obtained through the application of the feature generation techniques for all the three patients are taken into consideration to train a generalized model. It is observed that the accuracy obtained through such a model is indeed higher than that obtained using a generalized model trained using both types of features as well as that trained using only the transitory features, thus proving the validity of the hypothesis.

It is also observed that in case of the generalized model developed using using only lasting features, the highest accuracy of about 90% is obtained using XGBoost Classifier as well as the Random Forest Classifier, which is in accordance with the observation that in case of the generalized model developed using both types of features, the highest accuracy is obtained using the XGBoost Classifier, closely followed by the Random Forest Classifier.

Machine Learning Model	Accuracy		
	All Features	Transitory Features	Lasting Features
XG Boost Classifier	88.09%	84.80%	90.41%
SVM Classifier	86.02%	82.46%	86.52%
Extra Trees Classifier	85.92%	85.55%	88.74%
Random Forest Classifier	86.30%	85.27%	90.23%
AdaBoost Classifier	80.78%	79.92%	82.28%

Table 2. Recorded accuracies of the generalised model trained using all, transitory and lasting features respectively

D. Prediction using a patient-specific model

In the final approach, a patient-specific model is developed wherein each of the above-mentioned classifiers are trained using the recorded brain activity of only a particular patient while making use of both, the transitory and lasting features. Prior to the development of such a model, it was hypothesized that the uniqueness in the characteristics of any given patient being a distinguishing factor between several patients may result in a relatively higher accuracy through the use of patient-specific models. However, the observations of the accuracy obtained through the various models developed for each of the three patients indicated that the accuracy of such specific models per patient resulted in a relatively lower accuracy when compared with that obtained for generalized models developed using lasting features, thus disproving the hypothesis.

The results obtained using various generalized models have been tabulated and presented in table 2 and those obtained using the patient-specific model have been presented in Table 3.

Machine Learning Model	Accuracy		
	Patient 1	Patient 2	Patient 3
XG Boost Classifier	88.27%	89.44%	92.28%
SVM Classifier	88.27%	88.50%	88.73%
Extra Trees Classifier	85.80%	88.26%	88.94%
Random Forest Classifier	85.80%	87.79%	88.94%
AdaBoost Classifier	80.88%	84.42%	86.86%

Table 3. Recorded accuracies of patient specific models trained using all features .

VI. Conclusion

Recent advancements in technology have enabled the discovery of various revolutionary techniques such as Machine Learning, the application of which leads to the development of several unexplored models that could perform with an accuracy which was previously considered impossible to achieve. On the basis of this study, there are three primary conclusions that can be drawn for the prediction of an impending seizure. Firstly, it is found that a generalized model constructed using lasting features results in a better accuracy than a patient-specific model. Among the various Supervised Learning classification models taken into consideration for this study, the model developed using the XGBoost algorithm is found to result in the maximum accuracy. Finally, after comparing and contrasting the different generalized models constructed by taking several combination of the generated features into consideration, it can be concluded that the model developed using only the lasting features results in the highest accuracy. The future scope of this area of work would be to delve deeper into the analysis of several generalized models to identify the existence of similarity between the seizure occurrence patterns in groups of patients based on various characteristics, including but not limited to the current age and the age of diagnosis of epilepsy.

VII. Acknowledgement

We would sincerely like to express our gratitude towards our college M S Ramaiah Institute of Technology for providing us with the facilities that were required for the successful execution of our research work. We would also like to thank Mr. Levin Kuhlmann, representative of Epilepsy Ecosystem, for guiding us through the process of downloading and accessing the dataset.

References

[1] Falco-Walter, Jessica J., Ingrid E. Scheffer, and Robert S. Fisher. "The new definition and classification of seizures and epilepsy." *Epilepsy Research* 139 (2018): 73-79.

[2] Ullah, Ihsan, Muhammad Hussain, and Hatim Aboalsamh. "An automated system for epilepsy detection using EEG brain signals based on deep learning approach." *Expert Systems with Applications* 107 (2018): 61-71.

[3] Hussein, Ramy, Hamid Palangi, Rabab Ward, and Z. Jane Wang. "Epileptic seizure detection: a deep learning approach." *arXiv preprint arXiv:1803.09848* (2018).

[4] Usman, Syed Muhammad, Muhammad Usman, and Simon Fong. "Epileptic seizures prediction using machine learning methods." *Computational and mathematical methods in medicine* 2017 (2017).

[5] Zhou, Mengni, Cheng Tian, Rui Cao, Bin Wang, Yan Niu, Ting Hu, Hao Guo, and Jie Xiang. "Epileptic Seizure Detection Based on EEG Signals and CNN." *Frontiers in neuroinformatics* 12 (2018): 95.

[6] Cao, Yuzhen, Yixiang Guo, Hui Yu, and Xuyao Yu. "Epileptic seizure auto-detection using deep learning method." In *2017 4th*

International Conference on Systems and Informatics (ICSAI), pp. 1076-1081. IEEE, 2017.

[7] Tsiouris, Kostas M., Vasileios C. Pezoulas, Michalis Zervakis, Spiros Konitsiotis, Dimitrios D. Koutsouris, and Dimitrios I. Fotiadis. "A Long Short-Term Memory deep learning network for the prediction of epileptic seizures using EEG signals." *Computers in biology and medicine* 99 (2018): 24-37.

[8] Shiao, Han-Tai, Vladimir Cherkassky, Jieun Lee, Brandon Veber, Edward E. Patterson, Benjamin H. Brinkmann, and Gregory A. Worrell. "SVM-based system for prediction of epileptic seizures from iEEG signal." *IEEE Transactions on Biomedical Engineering* 64, no. 5 (2017): 1011-1022.*in biology and medicine* 99 (2018): 24-37.

[9] Epilepsy Ecosystem. (2019). *Epilepsy Ecosystem*. [online] Available at: https://www.epilepsyecosystem.org/ .

[10] Carrera, Enrique V., and Francisco Quinga. "Analysis of epileptic seizure predictions based on intracranial EEG records." In *2018 IEEE Colombian Conference on Communications and Computing (COLCOM)*, pp. 1-5. IEEE, 2018.