2020 IEEE Region 10 Symposium (TENSYMP), 5-7 June 2020, Dhaka, Bangladesh

Prediction of Epileptic Seizures using Support Vector Machine and Regularization

Shaikh Rezwan Rafid Ahmad, Samee Mohammad Sayeed, Zaziba Ahmed
Nusayer Masud Siddique, Mohammad Zavid Parvez
Software Engineering and HCI Research Group, Department of Computer Science and Engineering,
Brac University, Bangladesh
E-MAIL: rafid179@gmail.com, sameesayeed570@gmail.com, ahmedzaziba@gmail.com,
nusayermasud@gmail.com, zavid.parvez@bracu.ac.bd

***Abstract*—Epilepsy is a neurological disorder that causes abnormal behavior and recurrent seizures due to unusual brain activity. This study has attempted to predict seizures in epileptic patients through the process of feature extraction from EEG signals during preictal/ictal and interictal periods, classification and regularization. EEG signals from various parts of the brain from 10 epileptic patients are considered. Fast Fourier Transform (FFT) is used to determine the three features-the phase angle, the amplitude and the power spectral density of the signals. To classify the signals, these features are then used along with Support Vector Machine (SVM) as the classifier. Furthermore, regularization is used to make better predictions i.e. increase prediction accuracy and decrease the rate of false alarm. Finally, the proposed approach is tested on CHB-MIT Scalp EEG data set and it is able to predict epileptic seizures 25 minutes on average before the onset of the seizure with 100% accuracy and a low false-alarm rate of 0.46 per hour. This study intends to contribute to the development of better and advanced seizure predicting devices in the medical field.**

***Index Terms*—epilepsy, seizure, phase angle, power spectral density, support vector machine**

I. Introduction

Seizures occur due to electrical activities in the brain that are atypical or unusual. In other words, when neurons fail to communicate with each other due to a disruption. Effects of seizures range from uncontrollable shaking to loss of awareness or consciousness. Epilepsy is a chronic disorder that causes recurrent seizures. It is one of the most common neurological disorders which affects around 50 million people worldwide . [1] Seizures can affect the entire body since it stems from the central nervous system. The unpredictable nature of seizures can lead to numerous life-threatening situations for the patient and may cause great emotional stress. These situations can be averted by correctly predicting the approaching seizures before they occur. Prediction of oncoming seizures can vastly improve the quality of life for epileptic patients. Hazardous life-threatening situations can be completely avoided with a seizure predicting tool. EEG (electroencephalogram) helps to assess electrical activities in the brain and is crucial for studying seizures. Wave patterns are tracked by an EEG through numerous electrodes placed on the scalp.

A notable amount of research and study has been done in the field of seizure prediction. Shoeb *et al.* proposed a machine learning approach to detect epileptic seizures using *support vector machine* (SVM) [2]. This experiment resulted in 96% accuracy using 23 pediatric patients with a *false alarm rate* (FAR) of 2 per 24 hours. Kostas *et al.* proposed an unsupervised seizure detection method [3]. This technique achieved 81.15% to 91.35% accuracy with a FAR between 5.33/h-12.15/h. Alotaiby *et al.* [4] used a patient-specific model for seizure prediction using the CHB-MIT database [2]. This method used LDA classifier with features obtained from CSP. This achieved an accuracy of 89% and a FAR of 0.39/h. Another model proposed by Ahammed *et al.* acquired a detection accuracy of 84.2% [5]. This model performed classification using numerous features such as energy, entropy, standard deviation, etc. through the linear classifier.Williamson *et al.* proposed a seizure prediction algorithm using EEG features from 19 patients in conjunction with patient-specific machine learning [6]. This process also used a SVM and achieved an accuracy of 85% and FAR 0.03/h.

Based on literature [2]-[6], maintaining a prediction accuracy of 100% with a very low FAR is a challenging task due to the over-fitting nature of the model. The brain's electrical activity is different for different locations of the brain and has characteristics that overlap each other. The lack of consistency of the signals in terms of the patient's age and patient's sex also made it difficult to make accurate predictions with a low FAR. Therefore, the main concern of this study is to prioritize high accuracy and FAR.

Time domain and frequency domain is most commonly used in extracting features from EEG signals. Since seizure patterns are very much patient-specific, this study focuses on building a patient-specific model. *Fast Fourier transform* (FFT) is used in extracting features to differentiate two periods: preictal/ictal and interictal. 60 minutes of interictal signals and 30 min of preictal/ictal signals (as 30 min preceding a seizure onset) for each seizure of a patient are concatenated together. In this paper, an approach of method with SVM and regularization for predicting elliptic seizures with high accuracy and low FAR is proposed.

This paper is organized as follows: the data construction, feature extraction, classification, and post-processing are described in Section II. Results and Discussion in Section III, and Conclusion in Section IV.

978-1-7281-7366-5/20/$31.00 ©2020 IEEE

II. PROPOSED PLAN

The main goal of this paper is to predict the event of an epileptic seizure successfully with high accuracy and a low FAR in an automated way. A generic block diagram of the entire process is shown in Figure 1. In general, pre-processing is used to eliminate all the undesirable components from a signal. However, this paper's proposed plan allows a certain level of artifact tolerance with no filtering techniques being used. The phase angle, amplitude and power spectral density of the signals are the features that will be used to classify if they belong to the preictal/ictal or the interictal period. The classifier that is being used is SVM. Furthermore, regularization (i.e., windowing) technique is then used to make the final prediction.

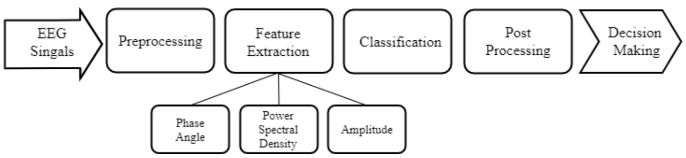

Fig. 1. Generic block diagram of our proposed model.

A. Data Construction

The data set that is being used is the CHB-MIT database collected from the Children's Hospital Boston [7]. The data set is publicly available on the internet and is a cited resource in many studies of detecting and predicting epilepsy [8]. It consists of EEG recordings from pediatric subjects with intractable epilepsy. In this study 10 patients are considered among which 6 are females and 4 are males. It is to be noted that each patient had readings from 23 different channels. The channels are - FP1-F7, F7-T7, T7-P7, P7-01, FP1-F3, F3-C3, C3-P3, P3-01, FP2-F4, F4-C4, C4-P4, P4-02, FP2-F8, F8-T8, T8-P8, P8-02, FZ-CZ, CZ-PZ, P7-T7, T7-FT9, FT9-FT10, FT10-T8, T8-P8

All signals are sampled at 16-bit resolution, at a rate of 256 samples per second. There are several files of EEG readings for all the patients, where each file contains hours of data, but for this research only the ones where seizures have taken place have been used. For each seizure that has occurred, 90 minutes of readings before the end of the seizure is noted. The first 60 minutes are classified as the interictal period and the remaining 30 minutes as the preictal/ictal period, where it is given a class value of 0 and 1 respectively. It is to be noted that only one seizure in a period of 90 minutes is considered, meaning that if there are two or more seizures, it is not considered.

B. Feature Extraction

The EEG signal after the construction of the data set contains preictal/ictal and interictal periods, where each seizure file is divided into epochs of 10s. FFT is then used to convert the signal from its time domain to its frequency domain. FFT is an optimized version of *discrete Fourier transform* (DFT). It transforms the signal into its spectral components and provides the signal frequency information. Shift function is then used which rearranges a Fourier transform by shifting the zero-frequency component to the center of the array. Both doted as F and T as follows:

$$F = \gamma(r) \tag{1}$$

$$T = \delta(F) \tag{2}$$

where γ is the FFT function and δ is the FFTshift function and represents each 1/256 s of the signal.

The three features to be used are the phase angle of the signal, the amplitude and the power spectral density of the signal. The phase is the position of a specific point in a particular instant on a waveform cycle. Phase angle refers to the phase difference. Using python's numpy library, the phase angle for the signal is then calculated using the shifted signal and numpy's angle function. A = numpy.angle(T), where A is the phase angle. The average of all the phase angle values, 2560 values (i.e., samples) for every 10s is considered and each 10s epoch is assigned the single corresponding mean phase angle value

In physics, the amplitude is the maximum displacement of a point on a vibrating body or wave measured from its equilibrium position. The amplitude of a signal measures the value of the signal at any point in time. The EEG signal amplitude is the frequency of the pattern in terms of electric energy micro volts. There are four common EEG frequency patterns: Beta (13-30 Hz), Alpha (7-12 Hz), Theta (4-7 Hz), and Delta (0.5-3 Hz), respectively. As the frequency increases the amplitude of the EEG decreases. The amplitude of the signal at a point in time is calculated using the formula below.

$$X = \left(\frac{2}{n * F}\right) \tag{3}$$

where X is the amplitude and the n is the epoch size

The average of all the amplitudes, 2560 values (i.e. samples) for every 10s, is considered and each 10s epoch is assigned the single corresponding mean amplitude value.

For a given signal, the power spectrum explains how power is transmitted into the frequency components that form the signal. The signal's *power spectral density* (PSD) represents the strength of the variations(i.e.,energy) as a function of frequency. It gives an idea of the frequencies at which the variations are strong and at which frequencies the variations are weak. The PSD of the signal at a point is calculated using the formula below.

$$P = 2\left(|\frac{f}{n}|\right)^2 \tag{4}$$

where P is the PSD and n is the epoch size.

The average of all the power spectral densities,2560 values(i.e. samples) for every 10s is considered and each 10s epoch is assigned the single corresponding mean PSD value. For each of the seizures that is considered, these three features are extracted for each of the 23 channels, producing a total of 69 (23 x3) features for each of the 540 rows of data.

C. Classification

After feature extraction, the data set (540 x 70) has 69 columns of input data i.e. 3 features for each of the 23 channels and one column of output data i.e. the class value for each row of data. The entire dataset is then randomly shuffled. To classify whether the epoch size of 10s belongs to preictal/ictal period or interictal period, SVM is being used as it is optimal for nonstationary EEG signals [9]–[11]. SVM is a supervised machine learning algorithm which categorizes data by drawing a line of separation between the two sets of data. The training data also known as the support vectors are represented by lines in space and are classified into different groups by drawing a separation between them also known as the hyperplane. SVM here assigns each row either to the interictal or preictal/ictal class. The hyperplane is placed in such a way that the gap between the two groups is as wide as possible.

Cross validation is used here to predict the class value for each row of data. The data set is first divided into five segments of equal size. Four of these segments are used as the training data and the remaining one as the test data, whose class values are to be predicted. Each of these segments are used as the test data once, thus predicting the class values each segment and therefore all the segments.

D. Post-processing

Many artifacts such as eye blinking, muscle movement, etc. may lead to misclassification of preictal/ictal and interictal signals. Thus, post-processing is required to make accurate predictions on the SVM classified signals. Regularization (i.e windowing) technique is used as done by Parvez *et al.* [12]. This is a two-step process, where x of v analysis is performed to make the predictions. It is to be noted that a window size of 5 minutes is used. The interictal period is assigned a class value of 0 and the preictal/ictal period is assigned a class value of 1. In the first step, five 10s (50s) epochs are analyzed and in the second step six of these 50s windows are analyzed. The first step is a 3 of 5 analysis i.e. if three or more 10s epochs have a value of 0 the entire 50s window is considered an interictal period and is assigned a class value of 0 and if not, it is considered a preictal/ictal period and is assigned a class value of 1. The second step is a 2 of 6 analysis i.e. if two or more 50s windows out of the six windows have the value 1 the entire 5-minute window is considered a preictal/ictal period and is assigned a class value of 1 and if not the 5-minute window is considered an interictal period and is assigned the value 0. This analysis made it very easy to calculate the false alarm i.e. the prediction of an epileptic seizure where the seizure had not occurred and also calculate when the seizure was predicted.

III. Results and Discussion

The features that are being used in this proposed model are - phase angles, power spectral density and amplitude. SVM classifier and regularization are used to accurately predict whether the signals belonged to the preictal/ictal or interictal period.

TABLE I
CALCULATED EARLY PREDICTION TIME PER SEIZURE PER PATIENT USING SVM

PN	TS	S1	S2	S3	S4	S5	S6	S7	S8	S9	Average
01	5	30	30	30	25	30	-	-	-	-	29
02	2	30	20	-	-	-	-	-	-	-	25
03	1	30	-	-	-	-	-	-	-	-	30
04	2	10	10	-	-	-	-	-	-	-	10
05	4	25	25	20	30	-	-	-	-	-	25
06	9	30	30	30	30	30	15	15	10	30	24.4
07	3	30	30	30	-	-	-	-	-	-	30
08	4	30	30	25	30	-	-	-	-	-	28.75
09	3	30	5	30	-	-	-	-	-	-	21.67
10	6	30	30	30	30	30	30	-	-	-	30
Average	-	-	-	-	-	-	-	-	-	-	25.382

PN = Patient Number, TS = Total Number of Seizures, S1-S9= Seizure Number

TABLE II
CALCULATED EARLY PREDICTION TIME PER SEIZURE PER PATIENT USING LOGISTIC REGRESSION

PN	TS	S1	S2	S3	S4	S5	S6	S7	S8	S9	Average
01	5	0	30	30	25	25	-	-	-	-	22
02	2	20	15	-	-	-	-	-	-	-	17.5
03	1	30	-	-	-	-	-	-	-	-	30
04	2	10	10	-	-	-	-	-	-	-	10
05	4	25	10	20	15	-	-	-	-	-	17.5
06	9	10	30	30	30	30	15	20	05	30	23.2
07	3	30	30	30	-	-	-	-	-	-	30
08	4	30	30	20	30	-	-	-	-	-	27.5
09	3	30	5	30	-	-	-	-	-	-	21.67
10	6	30	30	30	25	30	30	-	-	-	29.17
Average	-	-	-	-	-	-	-	-	-	-	22.85

PN = Patient Number, TS = Total Number of Seizures, S1-S9= Seizure Number

Table I shows that using SVM all the 39 seizures have been predicted numerous minutes before they had occurred, with a prediction accuracy of 100%. Table II shows that using the *logistic regression classifier* (LRC), 38 of the 39 seizures have been predicted, having a prediction accuracy of 97.4%. The tables also exhibit how SVM is more effective at predicting the seizures earlier compared to LRC. The average early prediction time using SVM is 25.38 minutes in comparison to 22.85 minutes using LRC.

TABLE III
TOTAL NUMBER OF FALSE ALARMS PER PATIENT AND FALSE ALARM RATE (HOUR) USING SVM

PN	TS	Total FA	False Alarm / hour
01	5	4	0.53
02	2	1	0.33
03	1	1	0.67
04	2	1	0.33
05	4	3	0.50
06	9	7	0.52
07	3	3	0.67
08	4	2	0.33
09	3	2	0.44
10	6	3	0.33
Total	39	27	0.46

PN = Patient Number, TS = Total Number of Seizures

TABLE IV
TOTAL NUMBER OF FALSE ALARMS PER PATIENT AND FALSE ALARM RATE (HOUR) USING LOGISTIC REGRESSION

PN	TS	Total FA	False Alarm / hour
01	5	5	0.67
02	2	1	0.33
03	1	1	0.67
04	2	0	0.00
05	4	3	0.50
06	9	9	0.67
07	3	5	1.11
08	4	1	0.17
09	3	2	0.44
10	6	3	0.33
Total	39	30	0.51

PN = Patient Number, TS = Total Number of Seizures

False alarm refers to the situation where a seizure is predicted but it has not occurred i.e. after regularization 1's are predicted as class values where 0's should have been predicted. This will give a false indication to a patient that a seizure might occur, whereas it will not. These false alarms might have occurred due to the over-fitting of this model.Table III and Table IV show the number of false alarms using LRC is greater than the one using SVM.Table III shows that using SVM, there are a total number of 27 false alarms with a FAR of 0.46/h compared to a total of 30 false alarms with a FAR of 0.51/h in Table IV where LRC is used as the classifier. SVM, therefore, is preferred for classification since it has a better prediction accuracy, lower FAR and also could predict the seizures earlier i.e. have a greater average early prediction time value.

Table V shows the comparison between our proposed method and the existing methods mentioned earlier in the paper.It is to be noted that except the studies conducted by Williamson *et al.* [6] and Parvez *et al.* [12], all the other studies here are using the CHB-MIT dataset [7]. Compared to other studies that had a similar approach, the proposed method succeeded in achieving a 100% prediction accuracy.However, FAR is relatively high in comparison with the other studies.

TABLE V
RESULT COMPARISON BETWEEN PROPOSED AND EXISTING METHODS

Related Studies	PA	FAR
Shoeb *et al.* [2]	96.0	0.13/h
Alotaiby *et al.* [4]	89.0	0.39/h
Williamson *et al.* [6]	85.0	0.03/h
Parvez *et al.* [12]	95.4	0.36/h
Proposed Method	100.0	0.46/h

PA = Prediction Accuracy, FAR = False Alarm Rate

IV. Conclusion

A seizure prediction model is proposed in this study that involves feature extraction, classification using SVM and regularization. This study focused on predicting the event of an epileptic seizure successfully with high accuracy. EEG signals from the data are constructed to have both preictal/ictal and interictal periods during which the features are extracted. Extraction of features (phase angle, amplitude and the power spectral density) is done using FFT and the SVM classifier is used to classify the signals. To refine the SVM classified signals, post-processing has to be done for optimal results. This paper manages to produce a perfect accuracy from a difficult and challenging dataset from the Children's Hospital Boston-Massachusetts Institute of Technology (CHB-MIT) database. This database provides EEG data from epileptic children with intractable epilepsy who stopped treatment 1 week before data acquisition. Compared to other studies that had a similar approach, this succeeded in achieving a 100% prediction accuracy. A lot of impressive work has been done on seizure detection but this approach aimed for a seizure prediction model, a field that is still lacking significant research. However, the number of observations from the CHB-MIT database is a concerning factor. This model succeeded in getting a 100% prediction accuracy but faltered in achieving an extremely low *false alarm rate* (FAR).This study hopes to expand this research in the future by working with more patients and applying our proposed model to a dataset with increased observations.

References

[1] J. R. Hughes, "A review of sudden unexpected death in epilepsy: Prediction of patients at risk," *Epilepsy Behavior*, vol. 14, no. 2, pp. 280–287, 2009.

[2] A. H. Shoeb and J. V. Guttag, "Application of machine learning to epileptic seizure detection," *Proceedings of the 27th International Conference on Machine Learning (ICML-10)*, pp. 975-982, 2010

[3] K. M. Tsiouris, S. Konitsiotis, D. D. Koutsouris, and D. I. Fotiadis, "Unsupervised Seizure Detection based on Rhythmical Activity and Spike Detection in EEG Signals," *2019 IEEE EMBS International Conference on Biomedical Health Informatics (BHI)*, 2019.

[4] T. N. Alotaiby, S. A. Alshebeili, F. M. Alotaibi, and S. R. Alrshoud, "Epileptic Seizure Prediction Using CSP and LDA for Scalp EEG Signals," *Computational Intelligence and Neuroscience*, vol. 2017, pp. 1–11, 2017.

[5] N. Ahammad, T. Fathima, and P. Joseph, "Detection of Epileptic Seizure Event and Onset Using EEG," *BioMed Research International*, vol. 2014, pp. 1–7, 2014.

[6] J. R. Williamson, D. W. Bliss, D. W. Browne, and J. T. Narayanan, "Seizure prediction using EEG spatiotemporal correlation structure," *Epilepsy & Behavior*, vol. 25, no. 2, pp. 230–238, 2012.

[7] CHB-MIT Scalp EEG Database, Jun. 2010. [Online]. Available: https://physionet.org/content/chbmit/1.0.0.

[8] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng, and H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet" *Circulation*, vol. 101, no. 23, 2000.

[9] H. J. Xing, M. H. Ha, B. G. Hu, and D. Z. Tian, "Linear Feature-weighted Support Vector Machine," *Fuzzy Information and Engineering*, vol. 1, no. 3, pp. 289–305, 2009.

[10] M. Z. Parvez and M. Paul, "Epileptic seizure detection by analyzing EEG signals using different transformation techniques,"*Neurocomputing*, vol. 145, pp. 190-200, 2014.

[11] S. Afroz, Z. H. Shimanto, R. Jahan, and, M. Z. Parvez, "Exploring the cognitive learning process by measuring cognitive load and emotional states," *Biomedical Engineering: Applications, Basis and Communications*, vol. 31, no, 04, pp. 1950032, 2019.

[12] M. Z. Parvez and M. Paul, "Seizure Prediction Using Undulated Global and Local Features," *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 1, pp. 208–217, 2017.