Patient-Aware Adaptive Ngram-based Algorithm for Epileptic Seizure Prediction using EEG Signals

Hussein Alawieh, Hussein Hammoud, Mortada Haidar, Mohamad H. Nassralla, Ahmad M. El-Hajj, and Zaher Dawy

Department of Electrical and Computer Engineering, American University of Beirut, Beirut, Lebanon
Email: {hma93, hah49, mmh98, mfn12, ame34, zaher.dawy}@aub.edu.lb

***Abstract*—This work proposes a novel patient-aware approach that utilizes an n-gram based pattern recognition algorithm to analyze scalp electroencephalogram (EEG) data and predict epileptic seizures. The method addresses the major challenge of extracting distinctive features from EEG signals through a detection of spatio-temporal signatures related to neurological events. By counting the number of occurrences of amplitude patterns with predefined lengths, the algorithm generates a probabilistic measure (anomalies ratio) that is used as a prediction marker. These extracted ratios are classified using state of the art machine learning algorithms into seizure and non-seizure windows. The efficacy of the prediction model is tested on patient records from the Freiburg database with more than 100 hours of recordings per patient and for a total of 145 seizures. The proposed algorithm is further optimized to obtain the n-gram parameters for enhanced feature extraction. Results demonstrate an average accuracy of 93.83%, sensitivity of 96.12%, and false alarm rate of 8.44%.**

I. Introduction

Epilepsy is one of the globally common neurological diseases with an estimate of 50 million affected people according to the World Health Organization [1]. Epilepsy is a disorder of the brain characterized by an enduring predisposition to generate recurrent seizures [1]; these are a result of excessive electrical discharges in a group of brain cells. Different parts of the brain can be the site of such discharges, and seizures can vary from the briefest lapses of attention or muscle jerks to severe and prolonged convulsions [1], [2]. Epilepsy can lead to significant medical consequences in addition to serious socioeconomic ramifications [3].

The potentially harmful consequences of epileptic seizures and the relatively high number of epileptic patients who do not respond to treatment or suffer from possible Sudden Unexpected Death from Epilepsy (SUDEP) give an incentive for developing alternative anti-epileptic measures. These include precautionary strategies that employ signal processing and computational intelligence to predict seizures before their onsets and give the necessary time to take safety procedures.

Epilepsy can be assessed by acquiring brain signals through electroencephalogram (EEG) measurements. Typical signals detected on the scalp are in the range of 20 to 150 V peak to peak over a 0.5 to 60 Hz bandwidth [4]. Studies on seizure prediction through EEG distinguish three major periods of an epileptic cycle as illustrated in Figure 1: a pre-ictal period before the seizure onset, an ictal period that contains the seizure, and an inter-ictal period of other intervals.

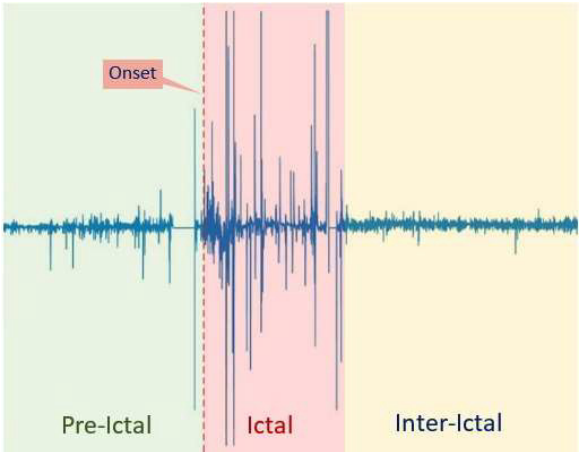

Fig. 1: Stages of an EEG recorded epileptic cycle.

Many studies have been performed to manifest the characteristics of seizures in the pre-ictal stage and differentiate them from those of other stages. This is one of the most challenging tasks in the seizure prediction process as it entails extracting distinctive features with high correlation to seizures from EEG signals. Prediction models try to find reliable features using different signal processing tools and techniques that can be classified into: univariate (i.e., computed on each EEG channel separately), and bivariate or multivariate which quantify some relationship such as correlation between two or more EEG channels.

In [5], authors used the Largest Lyapunov Exponent (LLE) to measure the chaotic behavior of scalp EEG signals. They used the difference between baselines of normal and pre-ictal stages as well as the LLEs of the EEG signals as features for classification. The chaos theory is revisited in [6] where the Discrete Wavelet Transform method was used to decompose EEG signals into a set of wavelets whose coefficients were used as features. In an attempt to extract new information about the complex dynamics of the brain neural networks, several studies tried applying non-linear time series analysis to EEG through different measures such as Lyapunov exponents, correlation dimension, or entropy [5], [7]. A hybrid approach

978-1-5090-3370-6/16/$31.00 ©2016 IEEE

was adopted in [8], where two synchrony measures were used: correlation from time domain and coherence from frequency domain. In [9], the Hilbert-Huang transform was applied through empirical mode decomposition of the nonlinear EEG data into basic components, after which a four-layer artificial neural network was employed for classification. A rather linear method was used in [10] with an autoregressive model applied on very short periods of EEG recordings to guarantee a quasi-stationary nature. Several statistical measures were used in [11] including the mean, variance, skewness, and kurtosis in order to extract information about the amplitude distribution of time series. Work in [11] incorporated frequency measures such as: relative power of spectral bands, the spectral edge frequency, the decorrelation time, as well as the Hjorth mobility and Hjorth complexity. In [12], the statistical behavior of local extrema in ranges of amplitude defined relative to the signals mean and standard deviation was used to feed features into a fuzzy logic system for prediction. Efforts in [13] were focused on calculating the relative change between consecutive segments of EEG data using phase correlation. Authors in [14] made use of the normalized spectral power of the different EEG bands as features. While much of the work had been based on univariate methods, [15] proposes a bivariate approach that uses the mean phase coherence to quantify the phase synchronization between EEG signals recorded from two different channels.

The majority of the conducted work in the field of epileptic seizure prediction tried complex linear and nonlinear methods of feature extraction to guarantee high accuracies while maintaining an acceptable low false alarm rate. On the contrary, this work presents a new low-complexity seizure prediction algorithm based on the n-gram counting process which was first used for seizure detection/prediction in [16], [17]. However, in comparison to these two seminal articles, this work employs a patient-specific optimization framework which fine-tunes the parameters of the counting process and utilizes advanced machine learning techniques to ensure optimized seizure prediction accuracy levels. As the work is intended to be part of a general framework that aims at incorporating an alarming mechanism to wearable EEG sensing headsets, it had to be less computationally expensive, fast, and more compatible with real time implementations. For the sake of such adaptability, the algorithm does not have any pre-processing steps, is established on patient-specific basis, uses selective channel inputs according to the patient's clinically identified seizure locations, and employs state of the art machine learning algorithms for classification. Moreover, in order not to compromise efficacy, the fine-tuning of the n-gram parameters is done through an offline exhaustive search using the subject's EEG records to maximize the accuracy in the prediction process.

The rest of the paper is organized as follows. The methodologies used throughout this work are presented in Section II. The n-gram based optimized algorithm is presented in Section III. Supporting simulation results are presented and analyzed in Section IV. Concluding remarks follow in Section V.

II. Datasets and Methodologies

A. Freiburg EEG Database

The Freiburg EEG database is part of EPILESIAE project which is an extensive EEG database for epilepsy patients containing recordings and metadata along with standardized annotations for 275 patients from several epilepsy centers [18]. During recording, each patient had at least three clinically manifest seizures with inter-ictal intervals greater than 4 hours.

This paper uses 17 patients from the Freiburg database to evaluate the performance of its proposed method for seizure prediction. EEG data for these patients were collected noninvasively with a sampling rate of 256 Hz. The average number of recorded seizures per patient ranged between 5 and 22 seizures.

B. N-gram based Feature Extraction

The n-gram model is mainly based on the assumption that the probability of a future event depends on a limited history of n previous events. Traditionally, the method is used for text and speech recognition applications [19]. The basic approach to find the conditional probability of obtaining a word sequence $(w_1, w_2, ..., w_m)$ of length m is shown in (1) as the product of the conditional probabilities of obtaining each of the words w_i of the sequence using the subsequence $(w_{i-n+1}, ..., w_{i-1})$ consisting of the n words that precede it. These probabilities are calculated in (2) as the count $C(w_{i-n+1}, ..., w_{i-1}, w_i)$ of the occurrences of each word after its relative preceding subsequence of length n to the total count $C(w_{i-n+1}, ..., w_{i-1})$ of occurrences of that subsequence.

$$P(w_1, w_2, ..., w_m) = \prod_{k=1}^{m} P(w_k | w_{k-n+1}, ..., w_{k-1}) \quad (1)$$

$$P(w_k | w_{k-n+1}, ..., w_{k-1}) = \frac{C(w_{k-n+1}, ..., w_{k-1}, w_k)}{C(w_{k-n+1}, ..., w_{k-1})} \quad (2)$$

The model is subject to application-dependent smoothing to account for zero probabilities and other constraints [19]. However, in this work, the n-gram sequence counts are used as an indicator for the variability of the EEG signal. The application of the n-gram algorithm to EEG signals was first described in [16] for seizure detection purposes. The variation of the number of occurrences of amplitude patterns in sequences of defined lengths give evidence of certain brain activity. The abrupt change in the electrical activity of the brain during a seizure would increase the variability of the EEG signal amplitudes and thus decrease the counts of previously recurrent and rhythmic amplitude patterns. Testing results showed a decrease in pattern counts during pre-ictal stages and an increase in post-ictal stages. Figure 2 shows the variation

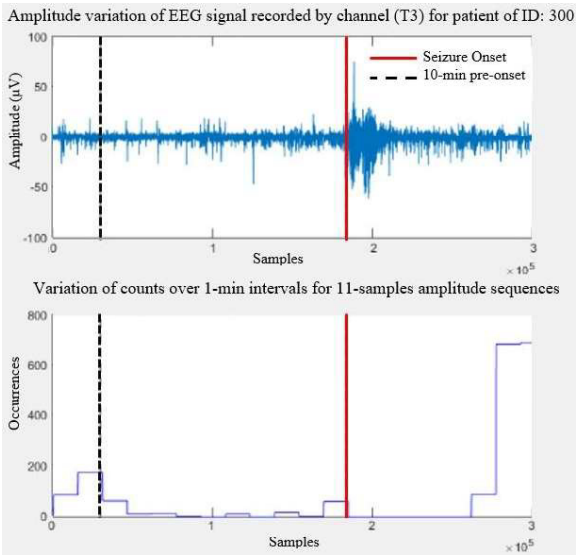

Fig. 2: Pattern count variation of amplitude sequences during 1 minute intervals of an epileptic cycle.

of counts for patterns (pattern length equals 11 samples) taken over 1-minute sub-intervals from a 27-minute recording of an epileptic cycle. The gradual decrease in counts or increase in variability of the EEG signal amplitudes in pre-ictal state is considered a prediction marker. Thus the variability, referred to as the anomalies ratio, is adopted as a distinctive feature for seizure prediction. Details for calculating the anomalies ratio are presented in Section III.

C. Machine Learning: Random Forest Algorithm

Learning models in this paper were built using the Random Forest machine learning algorithm which is significantly efficient in attaining high accuracies for classification purposes [20]. The algorithm undertakes dimensional reduction methods, treatment of missing values and outliers, and other essential data exploration steps. It comprises a decision forest which is an ensemble of decision trees. Each individual decision tree acts as the base classifier and the classification is performed by taking a vote based on the predictions made by each decision tree of the decision forest [21].

D. Prediction Model and Performance Metrics

All of the learning models were built using unbiased data sets; i.e., an equal number of seizure and non-seizure recording windows were used to extract features and train the prediction models. The evaluation metrics of accuracy, sensitivity, and false alarm rate (FAR) are calculated as follows:

$$\text{Accuracy} = \frac{\sum TP + \sum TN}{\sum TP + \sum FN + \sum TN + \sum FP} \tag{3}$$

$$\text{Sensitivity} = \frac{\sum TP}{\sum TP + \sum FN} \tag{4}$$

$$\text{FAR} = \frac{\sum FP}{\sum TN + \sum FP} \tag{5}$$

The variables used are defined as follows:

- **True Positives** (TP): Raising a seizure alarm knowing that a seizure occurred.
- **False Negative** (FN): Not raising a seizure alarm knowing that a seizure occurred.
- **True Negative** (TN): Not raising a seizure alarm knowing that no seizure occurred.
- **False Positive** (FP): Raising a seizure alarm knowing that no seizure occurred.

III. Proposed Method and Optimization

A. Feature Extraction

As previously indicated in Section II, the variability of the amplitude signal was adopted as a prediction feature, and it was referred to as the anomalies ratio (AR). The proposed n-gram algorithm calculates ARs by analyzing a window of time as follows:

1) Parametric initialization: Input parameters for the proposed method are:

- **Window size** (WS): It is the pre-ictal time period that is analyzed for prediction.
- **Interval Length** (L): The size of the sub-interval divisions of the window size.
- **Pattern Length** (PL): The length, in samples, of amplitude patterns to be analyzed. In our experiment, it is expressed as a set of lengths (e.g. [11, 9, 7, 5]).
- **Offset**: It is the time taken before the seizure and not included in the pre-ictal period under consideration. This is an important parameter as it allows to fine-tune the delimitation of the pre-ictal period. The offset period can also be seen as the grace period given to the patient to take precautionary measures before the seizure onset. This parameter gives an idea of the time between the seizure prediction outcome and the actual seizure onset.
- **Count Threshold** (CT): Number of occurrences of a pattern to be considered significant.
- **Scheme**: An overlapping scheme was used for counting pattern occurrences.
- **Weight**: Factor used to reduce the resolution of the data.

Figure 3 shows the parameters WS, L, and the Offset on a sample of an EEG recording.

2) Segmentation: After reducing the data by the Weight factor, the selected window of time WS is divided into equal sub-intervals of desirable length L. Each of these sub-intervals would result in an AR value. Feature vectors for classification are aggregates of N_{AR} (N_{AR} = WS/L) anomalies ratios that is used to decide whether the analyzed window refers to a pre-ictal (seizure) or an inter-ictal (non-seizure) duration.

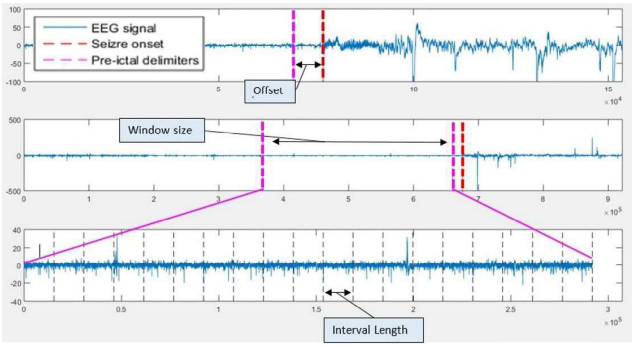

Fig. 3: Illustration of n-gram parameters (WS, L, and Offset) on a sample EEG recording.

3) AR *calculation:* The algorithm finds the set of significant patterns in each sub-interval according to CT. These would include all patterns of lengths specified by the set PL. The sample values that were not included in any of the significant patterns are referred to as unsequenced samples. The count of these samples in an interval of interest is used to calculate the AR of that interval as follows:

$$\mathrm{AR}_{\mathrm{interval}} = \frac{C(\mathrm{unsequenced\ samples})_{\mathrm{interval}}}{C(\mathrm{all\ samples})_{\mathrm{interval}}} \qquad (6)$$

4) Channel selection and feature vector: The classification decision is determined by the anomalies ratios that are extracted only from the K channels that detected the seizure out of all 24 channels recording EEG data. This is mainly motivated by the fact that partial seizures are localized in patient-specific parts of the brain. The overall anomalies matrix that is used for the analysis of a time window WS would have a size of (K x N_{AR}) as follows:

$$\begin{array}{c} \text{Anomalies Matrix} \\ \mathrm{Channel}_1 \\ \vdots \\ \mathrm{Channel}_{\mathrm{K}} \end{array} \begin{array}{c} \begin{array}{ccc} I_1 & \dots & I_{N_{\mathrm{AR}}} \end{array} \\ \begin{bmatrix} AR_{11} & \dots & AR_{1N_{\mathrm{AR}}} \\ \vdots & \ddots & \vdots \\ AR_{K1} & \dots & AR_{KN_{\mathrm{AR}}} \end{bmatrix} \end{array}$$

where I_i represents the i^{th} interval in the window.
For instance, if a patient had seizures that were detected on five frontal channels, then only these channels are considered in the process of predicting possible future seizures. For each channel, the algorithm would extract a feature matrix of size (5 x N_{AR}). If a window of 20 minutes is used with an interval length of 30 seconds then 40 ARs are extracted per channel. Thus the anomalies matrix is of size (5 x 40).

B. Classification

The resultant anomalies matrix is fed to the Random Forest algorithm for classification row by row. Each of the K selected channels contribute to a decision on the analyzed window WS through its corresponding feature vector. The algorithm would classify WS as a pre-ictal window if at least one of the channels' feature vectors led to such classification.

C. Optimization

Initialization of n-gram parameters was first done according to the visual observation of EEG signals and the study of pre-ictal period amplitude variations. However, using exhaustive search, the parameters discussed in Section III were extensively varied using test data to obtain the set of parameters that maximize prediction accuracy. The optimization was done on patient-specific basis. After preliminary simulations, parametric variations were limited to the following ranges:

- WS: 1 to 20 minutes before seizure
- L: 10 to 60 seconds
- PL: 3 to 16 samples
- Offset: 0 to 9 minutes

IV. Results and Analysis

A. Testing Model

The performance of the proposed method was evaluated using the scheme presented in Figure 4. For each patient with a set of N seizures, a dataset of N pre-ictal (seizure) windows and N random inter-ictal (non-seizure) windows was analyzed. An unbiased subset of N-1 (seizure, non-seizure) pairs was used to train the prediction model while the remaining pair was used to test it. The testing phase iterates through all N possible choices of the test pair and evaluates the performance metrics of accuracy, sensitivity, and FAR according to these N predictions. The whole procedure was repeated for different parametric combinations of the n-gram algorithm to optimize its performance.

B. Simulation Results

The above testing procedure was evaluated for 17 patients from the Freiburg database. An example of the variation of evaluation metrics upon optimization is illustrated in Figure 5 and Figure 6. In both figures, WS and the Offset were fixed at their optimized values from the exhaustive search procedure for patient (ID-300), which resulted in values of WS= 1min, Offset= 0, L= 30s, and PL= [11,9,7,5]. In Figure 5, L was fixed at its optimized value while varying PL and in Figure 6, PL was fixed at its optimized value while varying L. In Table I, the overall improvement after optimization was evident in all performance metrics. Table II details the optimized results for all patients while mentioning the number of recorded seizures for each one as well as a patient-specific combination of parameters WS, L, PL, and Offset that lead to the presented optimized results.

C. Analysis

The findings indicated that the adopted n-gram algorithm could achieve high accuracies even prior to optimization. However, a comparative analysis in Table I showed that the used optimization could make a significant increase of 4.66% in overall accuracy with both an increase in sensitivity (5.15%) and a decrease in the false alarm rate (4.17%). With an average

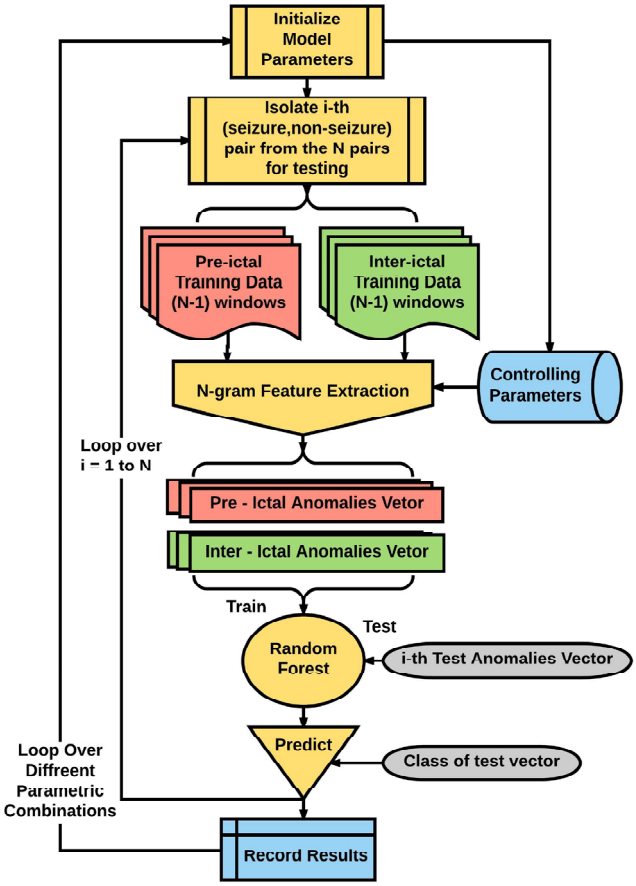

Fig. 4: Block diagram of the testing model.

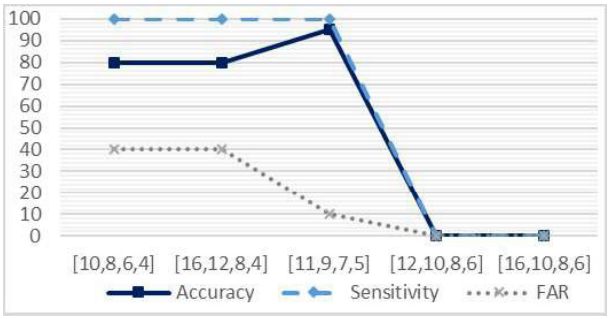

Fig. 5: Variation of performance metrics relative to PL for patient (ID-300) with all other parameters optimized.

accuracy of 93.83%, sensitivity of 96.12%, and FAR of 8.44%, the proposed method could correctly predict 137 out of the 145 tested seizures.

As demonstrated in Table II, 8 of the patients had 100% accuracy and 11 out of the 17 had 100 % sensitivity. It is also remarkable that the optimized parameters were not unique for each patient as several parametric combinations led to maximized accuracies. Figures 5 and 6 were clear examples of how accuracy went from 0 to above 90% while having

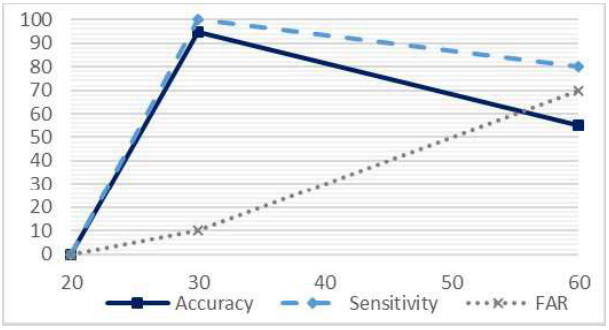

Fig. 6: Variation of performance metrics relative to L for patient (ID-300) with all other parameters optimized.

TABLE I: Comparison of performance metrics before and after optimization.

	Accuracy	Sensitivity	FAR
Pre_optimization	89.17%	90.97%	12.61%
Post_optimization	93.83%	96.12%	8.44%

all parameters at optimal values except for a single one. This implies that the overall performance is greatly dependent on all used parameters; this emphasizes the effectiveness of the used patient-aware exhaustive search optimization in improving the performance of the proposed algorithm. Columns 6, 7, 8 and 9 of Table II show a relatively wide diversity of optimized n-gram parameters that reflect variability among patients. This extensive analysis over large datasets along with the optimization performed justifies the patient-aware approach undertaken in this paper. In practice, custom-made routines could be developed per patient to increase the prediction accuracy.

V. Conclusions

With growing prevalence, escalating dangers, and lagging pace of medical treatment, epileptic seizure prediction has become an attractive solution for epilepsy patients. Performing valid and timely prediction requires characterizing the pre-ictal period with distinctive seizure-related features. This paper focused on building a compact, efficient, and accurate perdition model using features extracted by an n-gram based pattern recognition algorithm, classified with a state of the art machine learning algorithm, and optimized using patient-specific exhaustive search optimization that accounts for possible personal variability. Through experimental evaluation on a standard EEG dataset composed of 17 patients with a total of 145 seizures, the proposed method led to an average accuracy of 93.83 %, a sensitivity of 96.12 %, and a false alarm rate of 8.44 % which demonstrates the effectiveness of the proposed seizure prediction algorithm. The proposed algorithm does an offline optimization of the algorithm's parameters and then uses those to extract features based on basic counting procedures with no need for any pre-processing techniques.

TABLE II: Detailed prediction results with patient-specific optimized n-gram parameters.

Patient ID	Accuracy	Sensitivity	FAR	Seizure Count	WS (min)	L (s)	PL (samples)	Offset (min)
1	90.91%	90.91%	9.09%	11	6	30	10,8,6,4	0
73	92.86%	85.71%	0%	7	17	60	16,12,8,4	3
89	100%	100%	0%	5	10	60	11,9,7,5	5
110	100%	100%	0%	6	14	60	10,8,6,4	1
162	75%	83.33%	33.33%	6	16	20	11,9,7,5	4
219	100%	100%	0%	6	8	60	12,10,8,6	0
226	88.89%	88.89%	11.11%	18	18	60	11,9,7,5	0
239	100%	100%	0%	5	3	20	10,8,6,4	0
261	87.50%	100%	25%	8	6	60	12,10,8,6	0
300	95%	100%	10%	10	1	30	11,9,7,5	0
308	100%	100%	0%	9	10	30	12,10,8,6	2
325	87.50%	100%	25%	8	4	60	10,8,6,4	0
327	100%	100%	0%	6	8	30	10,8,6,4	2
454	100%	100%	0%	4	7	30	16,12,8,4	3
467	100%	100%	0%	6	10	30	16,12,8,4	2
568	87.5%	95%	20%	20	20	60	11,9,7,5	0
852	90%	90%	10%	10	8	60	10,8,6,4	0
Averages	93.83%	96.12%	8.44%	9-10	9.18	44.71	-	1.29

ACKNOWLEDGMENTS

This work was made possible by a research grant from NeuroPro AG. The statements made herein are solely the responsibilities of the authors.

REFERENCES

[1] W. H. Organization, *Epilepsy*. WHO Media centre, 2016.

[2] R. Fisher, C. Acevedo, A. Arzimanoglou, A. Bogacz, J. Cross, C. Elger, J. Engel, L. Forsgren, J. French, M. Glynn, D. Hesdorffer, B. Lee, G. Mathern, S. Mosh, E. Perucca, I. Scheffer, T. Tomson, M. Watanabe, and S. Wiebe, "ILAE Official Report: A Practical Clinical Definition of Epilepsy ," *Epilepsia*, vol. 55, no. 4, pp. 475–482, April 2014.

[3] M. England, C. Liverman, A. Schultz, and L. Strawbridge, "Epilepsy across the Spectrum: Promoting Health and Understanding," *Epilepsy Behavior*, vol. 25, no. 2, pp. 266–276, October 2012.

[4] A. Casson, D. Yates, S. Smith, J. Duncan, and E. Rodriguez-Villegas, "Wearable EEG: What Is It, Why Is It Needed and What Does It Entail?," *IEEE Engineering in Medicine and Biology Magazine*, vol. 29, no. 3, pp. 44–56, May 2010.

[5] M. Khalid, S. Aldosari, S. Alshebeili, T. Alotaiby, and F. El-Samie, "Online Adaptive Seizure Prediction Algorithm for Scalp EEG," *International Conference on Information and Communication Technology Research (ICTRC)*, May 2015.

[6] P. Rao, S. Akilan, S. Arumugam, V. Dhivakar, and D. Karthikeyan, "Epilepsy Seizure Detection using EEG - Curvelet Feature Selection and SVM Classification," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 5, no. 3, pp. 384–395, March 2010.

[7] S. Iqbal, Y. Khan, and O. Farooq, "Nonlinear Analysis of EEG for Seizure Prediction," *Annual IEEE India Conference (INDICON)*, December 2015.

[8] A. Sharma, J. Rai, and R. Tewari, "Epileptic Seizure Prediction and Identification of Epileptogenic Region using EEG Signal," *International Conference on Green Computing and Internet of Things (ICGCIoT)*, October 2015.

[9] T. Zhu, L. Huang, and X. Tian, "Epileptic Seizure Prediction by using Empirical Mode Decomposition and Complexity Analysis of Single-Channel Scalp Electroencephalogram," *International Conference on Biomedical Engineering and Informatics (BMIE)*, October 2009.

[10] L. Chisci, A. Mavino, G. Perferi, M. Sciandrone, C. Anile, G. Colicchio, and F. Fuggetta, "Real-Time Epileptic Seizure Prediction using AR Models and Support Vector Machines," *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 5, pp. 1124–1132, May 2010.

[11] C. R. R. A. A. K. P. D. C. E. F. Mormann, T. Kreuz and K. Lehnertz, "On the Predictability of Epileptic Seizures," *Clinical Neurophysiology*, vol. 116, no. 3, pp. 569–587, March 2005.

[12] H. Niknazar, K. Maghooli, and A. M. Nasrabadi, "Epileptic Seizure Prediction using Statistical Behavior of Local Extrema and Fuzzy Logic System," *International Journal of Computer Applications*, vol. 113, no. 2, pp. 24–30, March 2015.

[13] M. Parvez and M. Paul, "Seizure Prediction by Analyzing EEG Signal Based on Phase Correlation," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, August 2015.

[14] M. Bandarabadi, A. Dourado, C. Teixeira, T. Netoff, and K. Parhi, "Seizure Prediction with Bipolar Spectral Power Features using Adaboost and SVM Classifiers," *Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, July 2013.

[15] Q. Yuan and D. Wei, "A Seizure Prediction Method based on Efficient Features and BLDA," *IEEE International Conference on Digital Signal Processing (DSP)*, July 2015.

[16] W. Juffali, J. El-Imad, A. Eftekhar, and C. Toumazou, "The WiNAM Project: Neural Data Analysis with Applications to Epilespy," *IEEE Biomedical Circuits and Systems Conference (BioCAS)*, November 2010.

[17] A. Eftekhar, W. Juffali, J. El-Imad, T. Constandinou, and C. Toumazou, "Ngram-Derived Pattern Recognition for the Detection and Prediction of Epileptic Seizures," *PLOS One*, vol. 9, no. 6, pp. 1–15, June 2014.

[18] J. Klatt, H. Feldwisch-Drentrup, M. Ihle, V. Navarro, M. Neufang, C. Teixeira, C. Adam, M. Valderrama, C. Alvarado-Rojas, A. Witon, M. L. V. Quyen, F. Sales, A. Dourado, J. Timmer, A. Schulze-Bonhage, and B. Schelter, "The EPILEPSIAE Database: An Extensive Electroencephalography Database of Epilepsy Patients," *Epilepsia*, vol. 53, no. 9, pp. 1669–1676, June 2012.

[19] Y. Guo, H. Wang, , and J. V. Genabith, "Dependency-Based N-Gram Models for General Purpose Sentence Realisation," *Natural Language Engineering*, vol. 17, no. 4, pp. 455–483, October 2011.

[20] B. Liu, M. Ma, and J. Chang, *Information computing and applications*. Berlin: Springer, 2012.

[21] M. Adnan, "Improving the Random Forest Algorithm by Randomly Varying the Size of the Bootstrap Samples," *IEEE International Conference on Information Reuse and Integration (IRI)*, August 2014.