Epileptic Seizure Classification for Each of Five Classes Using Time-Domain and Wavelet Decomposition Based Features

Mosab A. A. YOUSIF
Department of Biomedical Engineering
Institute of Graduate Studies
Istanbul University – Cerrahpasa
Istanbul, Turkey
mosab_mahala@yahoo.com

Mahmut OZTURK
Department of Electrical and Electronics Engineering
Engineering Faculty
Istanbul University – Cerrahpasa
Istanbul, Turkey
mahmutoz@iuc.edu.tr

***Abstract*— With the rapid developments in medical technologies and computers, it is getting easier to detect neurological disorders. However, the early detection of Epilepsy which is a common and important neurological disorder is still a crucial and challenging procedure. Like the other disorders, early detection gives some chances for controlling and smoothing of the effects of the disease. Developing early detection methods for epileptic seizures from Electroencephalography (EEG) signals using signal processing and machine learning techniques has been a popular research area in last decades. In this work, we propose to use some time-domain features together with same features extracted for sub-signals of the original signal. For obtaining sub-signals from EEG signal, we used classical discrete-time wavelet decomposition. All EEG signals have been decomposed to five important sub-bands. After extracting features from EEG signals and all sub-signals, we have applied the feature matrices to some machine learning algorithms. Most of the researches in this area distinguish only two or three states of epileptic seizures. In contrary to majority, we preferred to classify all five classes from EEG signals. Classification results of our method show an important success comparing with the other researches. Using statistical, auto-correlation based and wavelet transform based features of EEG signals together to predict and detect epilepsy makes our method robust. In this work, we present the comparison of machine learning techniques for the detection of epilepsy using electroencephalography (EEG) signals. The five states of epilepsy are classified using Naive Bayes (NB), Decision Tree (DT), Random Forest (RF), Multilayer Perceptrons (MLP), k-Nearest Neighbors (k-NN), and Support Vector Machines (SVM) algorithms. Results show that Random Forest, Multilayer Perceptrons, k-Nearest Neighbors (k-NN) with 10-fold cross-validation are the most successful algorithms with 92% accuracy.**

Keywords— Epilepsy, Electroencephalography, Machine Learning, Time-Domain Features, Wavelet Decomposition

I. INTRODUCTION

Epilepsy is a crucial neurological disorder which affects more than 50 million people around the world. [1]. Detection of epilepsy is a challenging problem because it is determined by recurrent seizures, and it is not easy to detect its symptoms without monitoring brain activity through EEG. Epileptic seizure which is described as a short event of hyperactivity of neurons can be pre-ictal, inter-ictal, and post-ictal [2-4].

This neurological disorder can be diagnosed by the non-invasive electroencephalograms (EEG) recordings. The electrodes are placed in the scalp and then brain's electrical activity are recorded for a while to detect the spikes formed in the EEG signal. These spikes are the most important characteristics in the detection of epilepsy [5,6].

In traditional investigation, the neurologists detect epileptic seizures with their eyes and experiences. Nowadays, new methods are started to use in EEG monitoring to detect seizures on EEG data without manually searching for them and wasting hours for inspection. The new methods have used spectral, statistical and/or time–frequency features of EEG signals [7]. Some of them use artificial neural networks (ANN) in order to detect pre-ictal events in EEG. The ANN performs the highest accuracy of 95.09% as presented at reference [8]. In another research support vector machines (SVM) algorithm has been used to distinguish ictal and seizure-free EEG as presented in [9]. The accuracy of this model has reached 95.33%. As presented at [10] kNN has been used to detect sleep apnea, and they performs the average accuracy of 87.64%. In the algorithm presented at [11], authors have used DWT and SVM to distinguish normal, inter-ictal, and ictal states. Their method has performed an average accuracy of 99.6%. The authors of [12] have used DT, K-NN, Discriminant Analysis, and SVM models in order to detect the absence of seizures. Their results show that the proposed method performed the highest accuracy of 76.7%. In the research presented at [13], the authors have proposed to use SVM to detect epilepsy. Their algorithm performs the classification accuracy of 90.1%. In the method proposed at [14], K-NN, SVM, DT and MLP methods have been used to detect the epileptic seizure on EEG signals. In their research, the MLP has achieved an average rating accuracy of 99.82%. An automatic generation of medical report method (AGMedRep) has been proposed in order to process electroencephalogram (EEG) segments using machine learning (ML) to generate textual reports for epilepsy detection in the research presented at [15]. They have showed that the ANN classifier with spectrogram and bispectrogram features has been the most accurate method with 86.00% of classification accuracy.

In this work, we propose a new and robust method using machine learning techniques to classify five classes of EEG segments. We have used Naive Bayes, J48 Decision Tree, Random Forest, Multilayer Perceptrons, k-nearest neighbors (k-NN) and support vector machines (SVM) algorithms with the help of Weka Data Mining software.

II. Materials and Methods

A. Dataset

In this research, we used the EEG database from the University of Bonn, due to the availability of the highest number of EEG classes related to epilepsy in this database. The database is given in five groups, as shown in figure1. Each of them contains 100 segments of single-channel EEG with a length of 23.6 seconds, which obtained from different people using a 128-channel amplification system, with a sampling rate of 173.61 Hz [16].

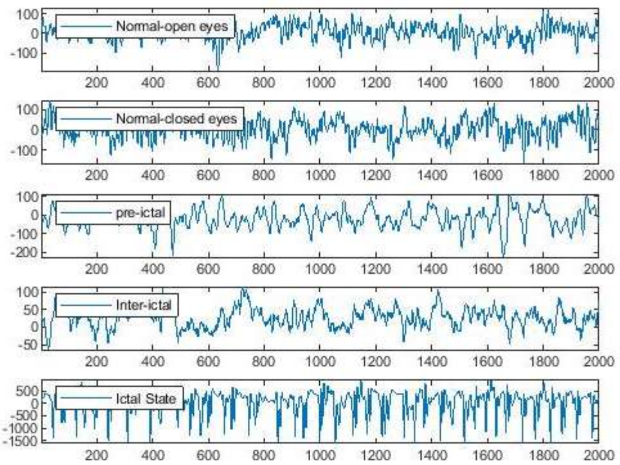

Figure 1. Five states of Epileptic seizure.

B. The EEG sets description:

A: Normal - open eyes;
B: Normal - closed eyes;
C: Epileptic, pre-ictal (epileptic focus);
D: Epileptic, Inter-ictal (Hippocampal region);
E: Epileptic, Ictal state.

90% of EEG signals from each group were randomly selected to build the model, while the 10% used to test the models [16].

C. Feature Extraction and Machine Learning

In time-domain, some statistical features of the original EEG signal have been extracted. These features are mean, median, mean absolute, root mean square, root mean square difference, second-order root mean square difference, standard deviation, variance, kurtosis, and skewness values. Thus, a total of 10 features can be extracted from each EEG signal.

By using auto-correlation of EEG signal, some more statistical features can be extracted. These are minimum, mean, standard deviation, variance, median, summation, sum of absolute values. A total of 7 features are obtained from each EEG signal.

The EEG signal has five clinically important sub-bands: delta (0–4 Hz), theta (4–8 Hz), alpha (8–15 Hz), beta (15–30 Hz), and gamma (30–60 Hz). A standard low pass filter has been used to restrict the EEG between 0–60 Hz.

A DWT, with fourth-order Daubechies (db4) wavelet, has been used for signal decomposition. The advantage of using the wavelet transform is its excellent multi-frequency representation, which includes time-frequency localization and scale-space analysis [17,18].

In the first level, the band-limited EEG (0–60 Hz) has been decomposed into detail and approximate coefficients of Gamma (30–60 Hz) and (0–30 Hz) respectively. The second level takes approximate coefficients resulted from the first level decomposition and further decomposed it to get high-resolution Beta (15–30 Hz) and low-resolution (0–15 Hz). The similar decomposition has been used in the third level of these low frequencies, Alpha (8–15 Hz) has been extracted along with the approximate coefficients (0–8 Hz), which is decomposed further in the fourth level for high frequencies Theta (4–8 Hz) and low frequencies Delta (0–4 Hz) [11].

From each sub-signals which are obtained using wavelet decomposition, statistical features such as mean, mean absolute, standard deviation, variance, median, root mean square, root mean square difference, second-order root mean square difference, kurtosis, and skewness values are extracted according to delta, theta, alpha, beta, and gamma bands. Thus, results in 50 features. Also we have taken the wavelet percentage of energy corresponding to the approximation, and we calculate 3-dB bandwidth in terms of the sampling rate. Thus, two features are extracted for every EEG signal.

Another features combination is the auto-correlation of delta, theta, alpha, beta, and gamma bands, and minimum, mean, standard deviation, variance, median, summation of auto-correlation vector, summation of absolutes of auto-correlation vector values are extracted for each sub-bands. Thus, results in 35 features.

In this work, states of epilepsy are classified using Naive Bayes (NB), J48 Decision Tree (DT), Random Forest (RF), Multilayer Perceptrons (MLP), k-Nearest Neighbors (k-NN), and Support Vector Machines (SVM) methods. Machine learning methods are applied to 500 EEG records in EEG database for epilepsy of the University of Bonn to classify them as normal - open eyes, normal - closed eyes, pre-ictal, inter-ictal, and ictal states. All the classification algorithms have been run in the Weka program.

III. Results

In this work, Weka (Waikato Environment for Knowledge Analysis) version 3.8.4 is used to apply the machine learning algorithms. Weka is a free licensed machine learning software (under the GNU General Public License) written in Java, developed at the University of Waikato, New Zealand. Some of the classification algorithms had been simulated by using Weka over 569 human beings with 32 attributes and the resulsts are presented in terms of speed and accuracy.

Accuracy (ACC) is a measure for correct prediction of the classifier, and it provides general information about how many samples are misclassified. It is defined as:

$$ACC = \frac{TP+TN}{TP+TN+FP+F} \quad (1)$$

where TP, FP, TN, and FN are the number of true positives, false positives, true negatives, and false negatives, respectively, when the classifier is predicted.

In this work, the metrics used to measure the success of the proposed method were Correctly Classified Instances, Incorrectly Classified Instances, Cohen Kappa Statistic (k), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE), and Root Relative Squared Error (RRSE). These metrics are defined as follow:

$$Correctly\ Classified\ Instances = \frac{Number\ of\ data\ classified\ correctly}{Number\ of\ total\ data} \quad (2)$$

$$k = \frac{(P_0-P_e)}{1-P_e} \quad (3)$$

$$MAE = \frac{1}{n}\sum_1^n |y_i - \hat{y_i}| \quad (4)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (y_i-\hat{y_i})^2}{n}} \quad (5)$$

$$RAE = \sum_{i=1}^n \frac{|e_i|}{y_i-\hat{y_i}} \quad (6)$$

$$RRSE = \sqrt{\sum_{i=1}^n \frac{e_i^2}{(y_i-\hat{y_i})^2}} \quad (7)$$

where y_i $(i = 1,2,\dots,n)$ is the real output, $\hat{y_i}$ $(i = 1,2,\dots,n)$ is the estimated output, n is the number of outputs, P_o is the probability of observed values, and P_e is the probability of values by chance.

In our experiments, as presented in Table 1, we found that Random Forest, Multilayer Perceptron and K-NN algorithms are considered as highly accurate with of 92% in comparison to all other 6 models that were used in this research. Table 1 represents the main model output characteristics according to the comparison criteria given above.

Table 1. Comparison for Accuracy of Machine Learning Techniques

Summary	Naive Bayes	J48	Random Forest	Multilayer Perceptron	k-NN	SVM
Correctly Classified	76 %	86 %	92 %	92 %	92 %	90 %
Incorrectly classified	24 %	14 %	8 %	8 %	8 %	10 %
Cohen Kappa Statistic	0.6964	0.8233	0.8991	0.8992	0.8989	0.8739
Mean Absolute Error	0.0976	0.059	0.0911	0.0422	0.0352	0.244
Root Mean Squared Error	0.3097	0.2353	0.1837	0.1721	0.178	0.3216
Relative Absolute Error	30.4609 %	0.235 %	28.443%	13.1879 %	10.9769 %	76.1663 %
Root Relative Squared Error	77.3266 %	58.7586 %	45.8615 %	42.9792 %	44.437 %	80.3122 %

Table 2 shows the comparison of our proposed method with some of the other researches in epileptic seizure detection area. This comparison has been made considering the choosed number of classes in the study, number of classifiers used and the highest accuracy.

Table 2. Comparison of our method with other epilepsy researches

Paper	Number of classes	Number of classifiers	Highest accuracy (%)
Saha et al. [10]	3	4	95.09
Sikdar et al. [11]	2	3	95.33
Zeng et al. [12]	3	1	99.60
Shen et al. [13]	3	4	76.70
Jaiswal et al. [14]	4	1	90.10
Oliva et al. [15]	3	32	99.90
Andrzejak et al. [16]	5	75	86.00
Our Method	5	6	92.00

Although the classification accuracy of our proposed algorithm has not been higher compared to the relevant works presented in Table 2, our results can be considered good because we have been used EEG signals from five different classes, which is more than those used in the related work and a hard procedure. As like the other non-stationary signals, the processing of the EEG signals is also difficult because of the characteristics of them, such as low signal-to-noise ratio, high complexity, instability, and non-linearity. At [16], the authors construct 15 datasets using four feature extraction methods and then apply it in five machine learning algorithms. They also try a total of 75 classifiers to obtain results. Whereas, we use only two feature extraction techniques and six classifiers.

The main reason behind the error rates of our model is the difficulty of classification between the pre-ictal and inter-ictal classes, and the normal (open eyes and closed eyes) as they generally have similar patterns, which further complicates the classification.

IV. Conclusions

In this research, we aimed to design a new and robust method to solve the automatic detection problem of epilepsy. The proposed method and the preferred machine learning algorithms have been applied to the EEG database of the University of Bonn.

The advantage of our research is that the accuracy results obtained is higher than the currently reported with same number of classes in literature. Random Forest, Multilayer Perceptron and K-NN algorithms are the most accurate methods with low error rates.

The disadvantages of our method are that we used a large number of features, which would slow down the machine learning algorithms and reduce the accuracy. This can be avoided by using criteria to select the features that give high classification accuracy.

This research could be extended to investigate another feature extraction techniques and machine learning methods. For the future studies, frequency domain and time-frequency domain features will be investigated and their detection accuracies will be compared. The accuracy rate and detection reliabilty of our proposed algorithm will probably increase using more robust feature selection algorithms and reducing the representation space.

Acknowledgment

Mosab A. A. Yousif would like to thank the turkish Presidency for Turks Abroad and University of Gezira – Sudan for financial support.

References

[1] WHO, Epilepsy, https://www.who.int/news-room/fact-sheets/detail/epilepsy. [Accessed Sep. 10, 2020].
[2] R.S. Fisher , W.E. Boas , W. Blume , C. Elger , P. Genton , P. Lee , J. Engel , Epileptic seizures and epilepsy: definitions proposed by the international league against epilepsy (ILAE) and the international bureau for epilepsy (IBE), Epilepsia 46 (4) (2005) 470–472.
[3] R.S. Fisher , C. Acevedo , A. Arzimanoglou , A. Bogacz , J.H. Cross , C.E. Elger , J. En- gel , L. Forsgren , J.A. French , M. Glynn , D.C. Hesdorffer , B.I. Lee , G.W. Mathern , S.L. Moshé, E. Perucca , I.E. Scheffer , T. Tomson , M. Watanabe , S. Wiebe , ILAE official report: a practical clinical definition of epilepsy, Epilepsia 55 (4) (2014) 475–482.
[4] G. Alarcón, A. Valentín, Introduction to Epilepsy, Cambridge University Press, 2012 .
[5] W.J. Freeman , R.Q. Quiroga , Imaging Brain Function With EEG: Advanced Tem- poral and Spatial Analysis of Electroencephalographic Signals, Springer, 2013.
[6] Ropper, A., & Brown, R. H., Principles of neurology (8th ed.). Boston, USA: McGraw-Hill, 2005.
[7] Mohseni, H. R., Maghsoudi, A., & Shamsollahi, M. B. (2006). Seizure detection in EEG signals: A comparison of different approaches. IEEE-EMBC.
[8] J. Kevric , A. Subasi , Classification of EEG signals for epileptic seizure predic- tion using ANN, in: Proceedings of the International Symposium Sustainable Development, IEEE, 2012.
[9] V. Joshi , R.B. Pachori , A. Vijesh ,Classification of ictal and seizure-free EEG sig- nals using fractional linear prediction, Biomed. Signal Process Control 9 (2014) 1–5.
[10] S. Saha , A. Bhattacharjee , M.A .A . Ansary , S.A . Fattah , An approach for automatic sleep apnea detection based on entropy of multi-band EEG signal, in: Proceed- ings of the Region 10 Conferernce, IEEE, 2016, pp. 420–423.
[11] D. Sikdar, R. Roy, M. Mahadevappa, Epilepsy and seizure characterisation by multifractal analysis of EEG subbands, Biomed. Signal Process Control 41 (2018) 264–270.
[12] K. Zeng , J. Yan , Y. Wang , A. Sik , G. Ouyang , X. Li , Automatic detection of absence seizures with compressive sensing EEG, Neurocomputing 171 (2016) 497–502.
[13] C.P. Shen , J.W. Lin , F.S. Lin , A.Y.Y. Lam , W. Chen , W. Zhou , H.Y. Sung , Y.H. Kao , M.J. Chiu , F.Y. Leu , GA-SVM modeling of multiclass seizure detector in epilepsy analysis system using cloud computing, Soft Comput. 21 (8) (2017) 2139–2149.
[14] A.K. Jaiswal , H. Banka , Local pattern transformation based feature extraction techniques for classification of epileptic EEG signals, Biomed. Signal Process Control 34 (2017) 81–92.
[15] Oliva, J.T. and Rosa, J.L.G., 2019. Classification for EEG report generation and epilepsy detection. Neurocomputing, 335, pp.81-95.
[16] R.G. Andrzejak , K. Lehnertz , F. Mormann , C. Rieke , P. David , C.E. Elger , Indi- cations of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state, Phys. Rev. E 64 (6) (2001) 061907.
[17] I. Daubechies, Ten lectures on wavelets CBMS-NSF Regional Conference Seriesin Applied Mathematics, vol. 61, 1992.
[18] S.G. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, IEEE Trans. Pattern Anal. Mach. Intell. 11 (7) (1989) 674–693.
[19] Han, J., Pei, J. and Kamber, M., Data mining: concepts and techniques. Elsevier, 2011.
[20] Tan, P.N., Steinbach, M. and Kumar, V., Introduction to data mining. Pearson Education India, 2016.