Pediatric epilepsy assessment based on EEG analysis

Yannis Misirlis*, Katerina D. Tzimourta*,†, Pantelis Angelidis*,
Nikolaos Giannakeas†, Alexandros T. Tzallas†, Markos G. Tsipouras*
*Department of Electrical and Computer Engineering, University of Western Macedonia, GR50100, Kozani, Greece
†Department of Informatics and Telecommunications, University of Ioannina, GR47100, Kostakioi, Arta, Greece
e-mail: yanismisirlis@gmail.com, ktzimourta@uowm.gr, paggelidis@uowm.gr, giannakeas@uoi.gr, tzallas@uoi.gr, mtsipouras@uowm.gr

***Abstract* — Epilepsy is one of the most common chronic neurological diseases, observed in all age groups and in all countries regardless of the level of the health system and the standard of living of the population. Epilepsy is caused by outbursts of electrical activity. The inability to detect seizures before they occur lead to the development of algorithms for detecting seizures. In this paper, an automated method of detection is presented. Filters are used to analyze the EEG signal into its frequency bands and then, features are extracted from the frequency and time domain. The extracted features are filtered to reweight the instances in the data and then the feature vector is used as input to train a Random Forests classifier, reaching a detection accuracy of 94.04%.**

Keywords — CHB-MIT; electroencephalogram; EEG; epilepsy; Random Forests; seizure detection

I. Introduction

Epilepsy is the fourth most common neurological disease which affects around 50 million people, while is calculated that around 70% of the people with epilepsy can live a normal life if the epilepsy is diagnosed early [1]. An epileptic seizure is a result of electrical discharges in the nerve cells of the brain and it is presented with different symptoms depending on the part of cerebral cortex from which it begins.

A seizure could last a few seconds and up to a few minutes. In short-term seizures the symptoms may be almost unnoticed, such as muscle twitching or loss of attention. In long-term seizures there are many symptoms such as loss of consciousness, movement disorders, loss of senses and other cognitive functions. The scientific community has managed to distinguish several of the causes of epilepsy, dividing them into six categories, namely structural, genetic, infectious, metabolic, immune and unknown [2]. Although some mechanisms that cause epilepsy have been identified, 50% of epileptic seizures worldwide have unknown causes.

The electroencephalogram (EEG) is a record of the brain activity and the main tool for epilepsy diagnosis. During the clinical examination, sensors are placed in the patient's scalp to record the electrical signals that neurons send to each other. Another type of recording is the intracranial EEG during which the electrodes are placed invasively inside the patient's brain and it is commonly used in presurgical evaluation of epileptic patients with refractory epilepsy. The placement of the electrodes is of crucial importance and the most common electrode placement system is the International System 10-20. After the signal is recorded, an experienced neurologist reviews the EEG that is usually captured between two seizures (interictal period) or during a seizure (ictal period), attempting to find epileptiform or non-epileptiform patterns that characterize the disorder. This is a time-consuming and difficult process because of the complexity of the seizures. Seizures are not always clearly distinguished due to extreme varied morphologies (i.e., spikes, sharp waves, spike-wave complexes, to name just few). As a result, seizure detecting algorithms have accumulated a great scientific interest and have proven to be an important clinical tool.

Research into automatic seizure detection has been around since 1970 and there have been many different approaches to tackling the problem [3, 4]. Initially, signal processing techniques were implemented to detect the peak waveforms that were created during the crisis. Later the researchers focused on finding the pre-ictal period and in crisis forecasting. Recently, many research teams have applied several different techniques and novel machine learning algorithms, such as the very promising neural networks, aiming to detect epilepsy before seizure onset and discriminate the different states of epileptic patients. Birjandtalab et al. [5] use power spectral analysis to extract features per subject for each time window, while Hu et al. [6] propose the partition of the preictal interval and distinguish the preictal subintervals from each other, in order to detect the preictal state. Kaleem et al. [7] presented a patient-specific seizure detection method based on features extracted from Discrete Wavelet Transformation (DWT), and compared different classifiers reaching accuracy values of 99.8%. In another research study, Zhou et al. [8] used Convolutional Neural Networks (CNN) to extract features from the raw EEG signal as an attempt to make the classification faster, and also compared the results to the ones extracted from the frequency domain. In another comprehensive work, Alickovic et al. [9] compared the performance of the Empirical Mode Decomposition, DWT and Wavelet Packet Decomposition, and also evaluated 4 different classifiers, namely k-Nearest Neighbors (k-NN), Multilayer Perceptron (MLP), Support Vector Machines (SVM), and Random Forests in order to develop a segment-based model for the classification of EEG signals. Khan et al. [10] investigated the hypothesis that the focal seizure can be predicted using scalp EEG data. The researchers used Convolutional filters on the

This work has been financed by the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH – CREATE – INNOVATE: "Intelli–WheelChair" (Project Code: T2EΔK-02438).

wavelet transformation to extract features for the interictal, preictal and ictal periods. Ahammad et al. [11] proposed a method for onset seizure detection based on wavelet features, interquartile range (IQR) and mean absolute deviation (MAD), utilizing a linear classifier that gave a sensitivity of 98.5%.

In this paper, a filter-based method is presented, which uses filters to divide the EEG recordings in 5 different subbands corresponding to the 5 brain waves. Then, the Energy of each subband along with the standard deviation and the interquartile range, which are being extracted from the time domain, are used as an input to train the Random Forests classifier.

II. MATERIALS AND METHOD

In the proposed method, FIR filters are applied in the EEG recordings aiming to extract the Energy from each subband of interest and create a feature vector. The feature vector is then used as an input in the Random Forest classifier to differentiate between ictal and interictal period. In Figure 1, a diagram of the proposed method is presented.

A. Database Description

The dataset used in this experiment is the CHB-MIT Scalp EEG Database collected from the Children's Hospital of Boston, which consists of EEGs from 23 patients [12]. The first and twenty-third subject are the same person in follow-up examination, recorded one and a half years apart. Also, the 24th subject is from the Beth Israel Deaconess in Boston and was added later in the database. In total, the dataset contains 956 hours of interictal EEG records and more than 3 hours of ictal EEG records. The EEG sampling frequency is 256 Hz with 16-bit resolution. The international 10-20 EEG system was followed for recording the EEG signals.

B. Preprocessing

The subjects did not have the same electrode arrangement; thus, only 18 channels that are common in all subjects are used. The 18 channels are: C3-P3, C4-P4, CZ-PZ, F3-C3, F4-C4, F7-T7, F8-T8, FP1-F3, FP1-F7, FP2-F4, FP2-F8, FZ-CZ, P3-O1, P4-O2, P7-O1, P8-O2, T7-P7 and T8-P8. Patients 4, 15, 18 and 19 were excluded from this experiment because they do not meet the required age criteria to be categorized as pediatric cases. Furthermore the 24th subject was excluded because no enough information about the patient is provided and the 16th subject was also excluded since the montage applied is different from the other subjects. Table I summarizes the data used in this study.

In this experiment, only EEG files that have at least one seizure included are used. The signals are segmented in 10-second epochs and the whole seizure is labelled as "ictal" period. The EEG data contained ten minutes before the seizure onset and up to two minutes before the seizure begins is considered as "interictal" period.

Afterwards, a FIR Band-Stop filter at 50 Hz is applied to all EEG recordings to avoid the power line noise interference. Then, 5 different Band-Pass filters are designed (0.5–4 Hz, 4–8 Hz, 8–13 Hz and 13–30HZ, 30–50 Hz) and applied, corresponding to the *delta (δ)*, *theta (θ)*, *alpha (α)*, *beta (β)* and *gamma (γ)* EEG rhythms.

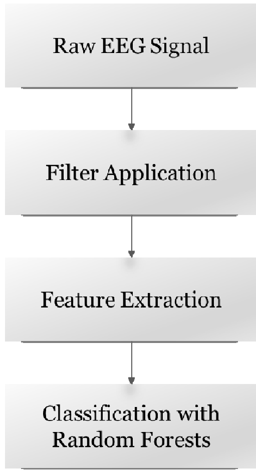

Fig.1 A diagram of the proposed method

C. Feature extraction

Initially, five statistical features from the 10-second epochs of the raw signal are extracted and then the Energy of each subband is calculated. The calculated features are

- Standard deviation: a statistical measurement of how dispersed the data is in relation to the mean value.
- Interquartile range (IQR): a statistical measurement that represents the spread of the middle of the dataset.
- Skewness: a measure showing the lack of symmetry of a dataset.
- Kurtosis: a statistical quantity that measures the complexity of the electroencephalogram.
- Mean: a statistical quantity that represents the average value of the dataset.
- Energy of δ, θ, α, β, γ: a measurement that represents the amount of electrical activity in a certain frequency band of the signal.

The Energy is the square value of the signal calculated for each band and the statistical features Standard deviation, Interquartile range, Skewness, Kurtosis and Mean were calculated using MATLAB functions. The extracted feature vector is used to train a Random Forests classifier.

D. Classification

Random Forests is an ensemble method in which many decorrelated decision trees are constructed during training phase [13]. The decision trees are trained and grown in different, randomly selected data, resulting in the reduction of variance. Each individual decision tree is responsible for its prediction and after a large number of trees created, they vote for the most popular class. Since the trees constructed in Random Forests are not correlated, the classifier outperforms Decision Trees and shows great classification performance.

Random Forests has shown great discrimination performance regarding binary classification problems [14], such as the one described in this study, which consists of two situations the "interictal" and the "ictal". In order to avoid an unbalanced dataset and get more accurate results, the ClassBalancer filter is applied, which is built-in in the Weka Knowledge Explorer [15] and it reweights the instances in the data so that each class has the same total weight.

TABLE I. STATISTICS OF CHB-MIT SCALP EEG DATABASE USED IN THIS EXPERIMENT.

No.	*Subject id*	*Age (years)*	*Interictal duration (sec)*	*Ictal duration (sec)*
1	Chb01	11	4.200	442
2	Chb02	11	1.800	199
3	Chb03	14	4.200	402
4	Chb05	7	3.000	618
5	Chb06	1.5	4.200	153
6	Chb07	14.5	1.800	325
7	Chb08	3.5	3.000	919
8	Chb09	10	1.800	276
9	Chb10	12	4.200	447
10	Chb11	12	1.800	806
11	Chb12	2	3.600	1475
12	Chb13	3	4.800	535
13	Chb14	9	4.200	169
14	Chb17	12	1.800	293
15	Chb20	6	3.600	294
16	Chb21	13	2.400	199
17	Chb22	9	1.800	204
18	Chb23	6	1.800	424
Total:	**18 cases**	**2-14.5**	**15 hours**	**2.2 hours**

The classifier was trained and tested according to the 10-fold cross-validation. During this learning technique the dataset is split into ten equal pieces of data, of which nine are used in training and one is used for accuracy testing. In the next iteration, another piece of data is used for testing and the rest 9 pieces for training the classifier. This procedure continues until all pieces of data have been used at least once for training and testing.

III. RESULTS

In this experiment the metrics calculated to evaluate the prediction of the proposed method are Accuracy, Sensitivity and Precision. These metrics are calculated by the statistical measures TP (True Positive), FP (False Positive), FN (False Negative) and finally TN (True Negative). These values are taken by the confusion matrix after the classification is completed.

The classifier performed well as it achieved 94.04% accuracy, 89.5% sensitivity and 98.4% precision. Results are obtained using the feature vector consisted of the standard deviation, the IQR and the Energy of each subband. The results of the classifier are presented in Table II.

TABLE II. RESULTS OF THE RANDOM FORESTS CLASSIFIER USED IN THE CHB-MIT SCALP EEG DATABASE FOR THE CLASSIFICATION PROBLEM "ICTAL/INTERICTAL".

Features	*Accuracy (%)*	*Sensitivity (%)*	*Precision (%)*
Standard deviation, Interquartile range, Energy of δ, θ, α, β and γ	94.04	89.50	98.40

IV. DISCUSSION

In this study, the publicly available CHB-MIT scalp epileptic EEG database is used and only the EEG recordings from the pediatric cases are included (i.e., until 16 years old) attempting to analyze, and classify the two major epileptic periods, ictal and interictal.

During the experimental procedure, several statistical features and different combinations with the extracted Energy from both all subbands and from certain subbands of interests were examined, in order to find the more appropriate features regarding seizure detection for children epilepsy. Results showed that the best feature vector was the one including the Energy of δ, θ, α, β and γ, Standard deviation and IQR.

Table III presents a comparison between this method and other proposed methods in the literature used CHB-MIT scalp epileptic EEG database. Even though classification accuracy is above 90% as obtained by other research studies, a direct comparison cannot be performed since the total duration of the database used in its study is not reported. The proposed methodology uses FIR filters for the analysis of the EEG recordings that are less complex than Short time Fourier Transform [17] and DWT [18] and shows good classification accuracy, regarding the performance of other studies proposed in the literature [16]-[19].

A strong advantage of this work is that the medical aspect is taken in consideration, as we eliminated the non-pediatric subjects from the database according to the instructions of the experts. The CHB-MIT database consists of mainly pediatric but also a few adult epileptic patients and most of the studies do not take into consideration this aspect, forming groups of EEG recordings from both pediatric and adult cases. In this case, the differences of the disease shown between these two groups are disregarded, resulting in misleading classification results. To the best of our knowledge this is the first study highlighting the severity of this aspect from the medical point of view.

TABLE III. COMPARISON TABLE BETWEEN THE PERFORMANCE OF THE PROPOSED METHOD AND OTHER METHODS PRESENTED IN THE LITERATURE FOR THE DISCRIMINATION BETWEEN ICTAL AND INTERICTAL PERIOD WITH EEG RECORDINGS FROM THE CHB-MIT DATABASE.

Authors	***Method***	***Dataset***	***Accuracy (%)***
Wei *et al.*[15]	Merger of the increasing and decreasing sequences - Convolutional Neural Network	70 seizure events	90.57
Choi *et al.*[16]	Short Time Fourier Transform, Convolutional Neural Network	184 seizure events	99.40
Rafiammal *et al.*[17]	Discrete Wavalet Transformation, Cluster K-Nearest Neighbors	All seizures	90.00
Zeng *et al.*[18]	Empirical Wavelet Transformation, K-Nearest Neighbors	All seizures	99.88
Proposed method	**Filters, Random Forests**	**Only pediatric seizures**	**94.04**

V. CONCLUSION

In this work, a filter-based method for automated detection of epileptic seizures is presented. The filters were used to extract the five EEG rhythms and then the Energy in each subband of interest. Statistical features have also been extracted from the raw signal from which the Standard deviation and the Interquartile range showed the best accuracy in combination with the Energy from each rhythm.

The extracted feature vector was reweighted so that both classes, interictal and ictal, have the same total weight and was used to train a Random Forests classifier. Classification accuracy above 94% was obtained for the discrimination between ictal and interictal period.

Epilepsy is a chronic brain disorder that affects almost 1% of the worldwide population. Scientific research has been focused in finding the way to detect the seizures before they happen to improve the patient's quality of life. The main finding concerning other studies utilizing the CHB-MIT database is that no other researchers filter out the non-pediatric subjects so that a potential patient can get medically accurate results with regards to pediatric patients, as the CHB-MIT is mainly a pediatric dataset. The proposed method is a tool, trying to address this issue, aiming to assist the experts in seizure detection. In the future, this method can be evaluated in adult databases to compare the different characteristics of the adult and the pediatric epileptic seizure. Alternative approaches, regarding the classifier and the feature vector must be tested to improve the method's statistical results.

REFERENCES

[1] World Health Organization. Epilepsy. Feb. 9, 2022. Accessed: May 5, 2022. [Online]. Available: https://www.who.int/news-room/fact-sheets/detail/epilepsy .

[2] R. S. Fisher, J.H. Cross, J.A. French, N. Higurashi, E. Hirsch, F. E. Jansen, L. Lagae, S. L. Moshé, J. Peltola, E. Roulet Perez, I. E. Scheffer, S. M. Zuberi, "Operational classification of seizure types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology" *Epilepsia*, vol 58, no. 4, 2017, pp. 522–530. doi: 10.1111/epi.13670

[3] K. D. Tzimourta, A. T. Tzallas, N. Giannakeas, L. G. Astrakas, D. G. Tsalikakis, P. Angelidis, M. G. Tsipouras, "A robust methodology for classification of epileptic seizures in EEG signals.", *Health and Technology*, vol 9, no. 2, 2019, pp 135-142. doi: 10.1007/s12553-018-0265-z

[4] A. T. Tzallas, M. G. Tsipouras, D. G. Tsalikakis, E. C. Karvounis, L. Astrakas, S. Konitsiotis, M. Tzaphlidou, "Automated epileptic seizure detection methods: a review study.", *Epilepsy–Histological, Electroencephalographic and Psychological Aspects*, 2012, pp. 2027-2036. doi: 10.5772/31597

[5] J. Birjandtalab, M. B. Pouyan, D. Cogan, M. Nourani and J. Harvey, "Automated seizure detection using limitedchannel EEG and non-linear dimension reduction", *Computers in Biology and Medicine*, vol. 82, 2017, pp. 49-58. doi: 10.1016/j.compbiomed.2017.01.011.

[6] W. Hu, J. Cao, X. Lai et al., "Mean amplitude spectrum based epileptic state classification for seizure prediction using convolutional neural networks", *Ambient Intell Human Comput*, 2019. doi: 10.1007/s12652-019-01220-6.

[7] M. Kaleem, A. Guergachi, S. Krishnan, "Patient-specific seizure detection in long-term EEG using wavelet decomposition", *Biomedical Signal Processing and Control*, vol. 46, 2018, pp 157-165. doi: 10.1016/j.bspc.2018.07.006.

[8] M. Zhou, C. Tian, R Cao et al., "Epileptic Seizure Detection Based on EEG Signals and CNN.", *Front Neuroinform*, vol. 12, no.95, 2018. doi: 10.3389/fninf.2018.00095.

[9] E. Alickovic, J. Kevric, A. Subasi, "Performance evaluation of empirical mode decomposition, discrete wavelet transform, and wavelet packed decomposition for automated epileptic seizure detection and prediction", *Biomedical Signal Processing and Control*, vol. 39, 2018, pp 94-102. doi: 10.1016/j.bspc.2017.07.022.

[10] H. Khan, L. Marcuse, M. Fields, K. Swann, B. Yener, "Focal Onset Seizure Prediction Using Convolutional Networks", *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 9, 2018, pp. 2109- 2118. doi: 10.1109/TBME.2017.2785401.

[11] N. Ahammad, T. Fathima, P. Joseph, "Detection of Epileptic Seizure Event and Onset Using EEG", *BioMed Research International*, vol. 2014, 2014. doi: 10.1155/2014/450573

[12] A. Shoeb. "Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment." PhD Thesis, Massachusetts Institute of Technology, September 2009.

[13] L. Breiman. "Random Forests". *Machine Learning*, vol. 45, 2001, pp. 5–32. doi: 10.1023/A:1010933404324

[14] A. Miltiadous, K. D. Tzimourta, N. Giannakeas, M. G. Tsipouras, T. Afrantou, P. Ioannidis, A. T. Tzallas. "Alzheimer's disease and frontotemporal dementia: A robust classification method of EEG signals and a comparison of validation methods." *Diagnostics*, vol *11*, no. 8, 2021, pp 1437. doi: 10.3390/diagnostics11081437

[15] E. Frank, M. A. Hall, and I. H. Witten. "The WEKA Workbench". Online Appendix for "Data Mining: Practical Machine Learning Tools and Techniques", Morgan Kaufmann, Fourth Edition, 2016.

[16] Z. Wei, J. Zou, J. Zhang, J. Xu, "Automatic epileptic EEG detection using convolutional neural network with improvements in time-domain.", *Biomedical Signal Processing and Control*, vol. 53, 2019, pp. 101551. doi: 10.1016/j.bspc.2019.04.028

[17] G. Choi *et al*., "A Novel Multi-scale 3D CNN with Deep Neural Network for Epileptic Seizure Detection," *2019 IEEE International Conference on Consumer Electronics (ICCE)*, 2019, pp. 1-2, doi: 10.1109/ICCE.2019.8661969.

[18] S. S. Rafiammal, D. N. Jamal, S. K Mohideen, "Detection of Epilepsy Seizure in Adults Using Discrete Wavelet Transform and Cluster Nearest Neighborhood Classifier", Iranian Journal of Science and Technology, vol. 45, 2021, pp. 1103–1115. doi: 10.1007/s40998-021-00437-6

[19] J. Zeng, T. Xiao-dan, A. Z. Chang'an, "Automatic detection of epileptic seizure events using the time-frequency features and machine learning." *Biomedical Signal Processing and Control*, vol 69, 2021, pp. 102916. doi: 10.1016/j.bspc.2021.102916.