Artificial Intelligence Research and Development
T. Alsinet et al. (Eds.)

doi:10.3233/FAIA240410

Detection of Epileptic Seizures with EEG Sequential Patterns

Cristina MONTSERRAT, Jonah FERNANDEZ, Bianca INNOCENTI and Beatriz LÓPEZ [1]

eXiT research group, University of Girona

ORCiD ID: Jonah Fernandez https://orcid.org/ 0000-0002-1107-3163, Bianca Innocenti https://orcid.org/ 0000-0002-1544-9658, Beatriz López https://orcid.org/0000-0001-9210-0073

Abstract. The rise of wearable EEG devices has opened the opportunity to develop new tools for neurological disorder monitoring, particularly for conditions like epilepsy. Machine learning plays a key role in processing the EEG signal towards an assessment of the person's state, and eventually evaluating some condition risk. However, existing approaches often rely on raw EEG data, keeping a numerical representation of the information contained in the data. Conversely, in a previous work, we explored representing the EEG signals using sequential patterns. In this work, we analyze the potential of such a representation through several machine learning methods, including decision trees, support vector machines, k-nearest neighbors, and random forest. The experiments carried out with the CHB-MIT scalp EEG database of Physionet show the outperformance of random forest.

Keywords. Predictive ML, EEG data, sequential pattern mining, neurological disorders, mental disorders

1. Introduction

The burden of neurological diseases has promoted the development of powerful tools for neuroimaging [1]. In particular, the non-invasive techniques of functional magnetic resonance imaging (fMRI) and electroencephalography (EEG) are the most popular tools to understand epilepsy [2]. fMRI provides information about the spatial localization of the brain functions, while EEG evidence about the brain's temporal dynamics. Diagnosing tools, however, differ from monitoring devices outside of hospitals. In that regard, wearable sensor technology has evolved significantly in recent years, proving EEG gadgets. The combination of these smart EEG devices, together with software developments based on machine learning approaches that interpret the data provided by the wearable, supports the assessment of the person's health condition in her daily life [2].

Our research concerns the use of this kind of EEG wearable for epilepsy detection, with the aim to provide a decision support tool to advice persons when a seizure is detected. To that end, we explored in [3] sequential frequent patterns to represent EEG data. In this work, we exploit such representation throughout several machine learning

[1]Corresponding Author: beatriz.lopez@udg.edu

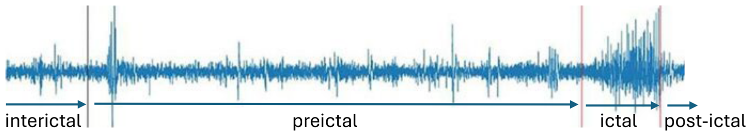

Figure 1. Typical brain states of epilepsy (adapted from [4]).

(ML) approaches, including decision trees (DT), support vector machines (SVM), k-nearest neighbors (k-NN), and random forest (RF).

2. Background and related work

In the EEG of a person diagnosed with epilepsy, four phases can be distinguished: interictal, preictal, ictal, and post-ictal (see Figure 1). In the interictal phase, the patient does not present symptoms of a seizure. In the preictal phase, it is possible to read abnormalities in the electrical activity of the brain that indicate the possibility of a future seizure. In the ictal phase, the seizure occurs. Finally, in the post-ictal phase, the brain may temporarily exhibit abnormal electrical activity due to the previous phase.

There are several previous works that aim to detect preictal states, so as to anticipate possible seizures and differentiate them from interictal states. The ictal state is discarded (too late for detecting a seizure), as well as the post-ictal state, as could have noisy data due to the recent seizure. The main current trend is about time-series classification techniques, including deep learning approaches [5] and XGBoost [6], based on extraction of some statistics or frequency features from the EEG signal. In this work, we explore the representation of EEG data with frequent pattern mining. To that end, in our previous work [3], we first used SAX [7] to transform the EEG signal into a symbolic representation. Next, we applied VEPRECO [8] to find sequential patterns separately for each kind of signal phase (preictal and interictal). Then, we represented each EEG recording as a collection of sequential patterns that were used in a case-based reasoning (CBR) framework for seizure detection. The results of this work were compared according to the k-NN outcomes.

3. Methodology

Figure 2 shows an overview of the methodology. Our starting point is the public database of the Physionet repository called CHB-MIT Scalp EEG Database [9], which includes data of 23 subjects. Each EEG file contains information from 20 channels; we work only with two channels, F7-T7 and T7-FT9, corresponding to the temporal lobe, which has been demonstrated to be related to epileptic attacks [10].

First, EEG data is segmented, so preictal and interictal periods are separated. The goal is to detect in future EEG signals segments corresponding to preictal patterns. We get an unbalanced dataset of 198 preictal segments, with a high variability among the different subjects. For example, patient chb03 had 33 records without a seizure period and only 7 with a seizure period. Secondly, each EEG recording has been transformed into a binary feature representation, where each feature represents the presence (value 1)

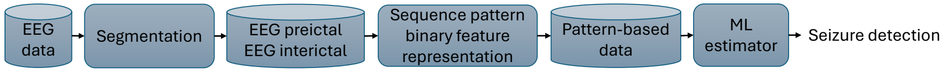

Figure 2. Flow chart of the methodology.

or absence (value 0) of a sequence frequent pattern. Patterns have been found using the VEPRECO algorithm as explained in section 2. Finally, different machine learning (ML) approaches have been used to assess the potential detection of a seizure including DT, SVM, k-NN and RF.

4. Results

Two experimental scenarios have been considered: 1) **intra-subject**: detection based on the data of each subject (75% of the data of the preictal and interictal segments of a subject were used for training and the remaining 25% for testing), and 2) **inter-subject**: detection based on the whole available data (75% of the subjects were used for training, and the remaining 25% for testing). Results are analyzed in terms of accuracy, precision, recall, and f1-score.

Table 1 shows the results achieved in both experimental scenarios. First of all, it is possible to observe that despite the models having a reasonable accuracy score, the precision and recall are very low. That means that the seizures are not detected appropriately. RF performs the best in the intra-subject scenario, outperforming the previous results published in [3] (corresponding to the k-NN method). In Figure 3 provides details on the performance of RF per each individual. A high variability can be observed, with some good results for some of the subjects (chb08 and chb024). Regarding the inter-subject results, the results are worse than in the intra-subject scenario, indicating the need to develop personal models for epileptic seizure detection. Overall, the results are far away from other representation and classification techniques, but the sequence representation we propose could be more interpretable than other state-of-the-art approaches. On the other hand, these poor results could be due to the scarcity of available data (23 subjects, with a high variability of epileptic events). Future research will involve using additional public datasets to validate the preliminary results presented in this paper.

5. Conclusions

EEG has huge potential to be used for developing wearable sensors that facilitate monitoring persons suffering from epilepsy. In this work, we analyze different machine learn-

	Intra-subject				Inter-subject			
	Acc	Prec	Rec	f1	Acc	Prec	Rec	f1
DT	0.78 (0.14)	0.18 (0.33)	0.26 (0.44)	0.21 (0.36)	0.76	0.25	0.03	0.05
RF	0.73 (0.23)	0.20 (0.32)	0.35 (0.48)	0.23 (0.35)	0.26	0.23	0.92	0.37
SVM	0.80 (0.12)	0.10 (0.27)	0.13 (0.34)	0.11 (0.29)	0.77	0.00	0.00	0.00
k-NN	0.75 (0.15)	0.11 (0.23)	0.22 (0.41)	0.14 (0.28)	0.75	0.00	0.00	0.00

Table 1. Experimental results. Mean values and standard deviation between brackets.

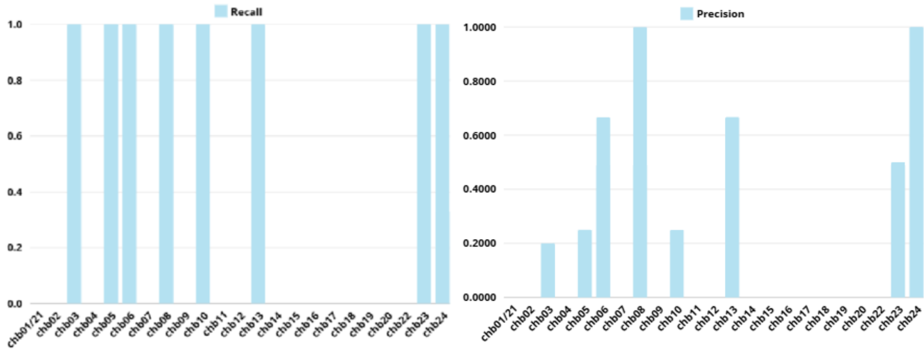

Figure 3. Results details per subject, for the Random Forest method.

ing estimators regarding an EEG representation based on sequence frequent patterns. The results show that managing individual data is a key issue in order to get good performance results; in particular, RF has exhibited to be the best technique among the different ones tested. Nevertheless, the results are not good enough to be translated to clinical practice. Future work involves testing the methodology with more individuals.

Acknowledgements. This study was conducted with the support of the Generalitat de Catalunya via the Consolidated Research group 2021 SGR 01125, and grant CPP2021-00831 funded by MICIU/AEI/10.13039/501100011033 and, by "European Union NextGenerationEU/PRTR".

References

[1] Perovnik M, Rus T, Schindlbeck KA, Eidelberg D. Functional brain networks in the evaluation of patients with neurodegenerative disorders. Nature Reviews Neurology 2022 19:2. 2022 12;19:73-90.

[2] Goodman AM, Szaflarski JP. Recent Advances in Neuroimaging of Epilepsy. Neurotherapeutics 2021 18:2. 2021 5;18:811-26.

[3] Fernandez J, et al. Examining the Potential of Sequence Patterns from EEG Data as Alternative Case Representation for Seizure Detection. In: ICCBR 2024. LNCS. vol. 14775. Springer; 2024. p. 258-72.

[4] Wei X, Zhou L, Chen Z, Zhang L, Zhou Y. Automatic seizure detection using three-dimensional CNN based on multi-channel EEG. BMC Medical Informatics and Decision Making. 2018 12;18:71-80.

[5] Chen W, Wang Y, Ren Y, Jiang H, Du G, Zhang J, et al. An automated detection of epileptic seizures EEG using CNN classifier based on feature fusion with high accuracy. BMC Medical Informatics and Decision Making. 2023 12;23:1-17.

[6] Gadda AAS, et al. Optimization of Pre-Ictal Interval Time Period for Epileptic Seizure Prediction Using Temporal and Frequency Features. Studies in Health Technology and Informatics. 2023 5;302:232-6.

[7] Lin J, Keogh E, Lonardi S, Chiu B. A symbolic representation of time series, with implications for streaming algorithms. Proc 8th ACM SIGMOD Workshop DMKD '03. 2003:2-11.

[8] Mordvanyuk N, Bifet A, López B. VEPRECO: Vertical databases with pre-pruning strategies and common candidate selection policies to fasten sequential pattern mining. Expert Systems with Applications. 2022 10;204:117517.

[9] Guttag J. CHB-MIT Scalp EEG Database v1.0.0; 2010. Available from: https://physionet.org/content/chbmit/1.0.0/.

[10] Eftekhar A, Juffali W, El-Imad J, Constandinou TG, Toumazou C. Ngram-Derived Pattern Recognition for the Detection and Prediction of Epileptic Seizures. PLOS ONE. 2014 6;9:e96235.