Epileptic Seizure Prediction with Stacked Auto-encoders: Lessons from the Evaluation on a Large and Collaborative Database

R. Barata, B. Ribeiro, A. Dourado, and C. A. Teixeira

Abstract

The seizure prediction performance of algorithms based in stacked auto-encoders deep-learning technique has been evaluated. The study is established on long-term electroencephalography (EEG) recordings of 103 patients suffering from drug-resistant epilepsy. The proposed patient-specific methodology consists of feature extraction, classification by machine learning techniques, post-classification alarm generation, and performance evaluation using long-term recordings in a quasi-prospective way. Multiple quantitative features were extracted from EEG recordings. The classifiers were trained to discriminate preictal and non-preictal states. The first part of the feature time series was considered for training, a second part for selection of the "optimal" predictors of each patient, while the remaining data was used for prospective out-of-sample validation. The performance was assessed based on sensitivity and false prediction rate per hour (FPR/h). The prediction performance was statistically evaluated using an analytical random predictor. The validation data consisted of approximately 1664 h of interictal data and 151 seizures, for the invasive patients, and approximately 4446 h of interictal data and 406 seizures for the scalp patients. For the patients with intracranial electrodes 18% of the seizures were correctly predicted (27), leading to an average sensitivity of 16.05% and average FPR/h of 0.27/h. For the patients with scalp electrodes 20.69% of the seizures (84) on the validation set were correctly predicted, leading to an average sensitivity of 17.49% and an average FPR/h of 0.88/h. The observed performances were considered statistically significant for 4/19 invasive patients ($\approx 21\%$) and for 5/84 scalp patients ($\approx 6\%$). The observed results evidence the fact that, when applied in realistic conditions, the auto-encoder based classifier shows limited performance for a larger number of patients. However, the results obtained for some patients point that, in some specific situations seizure prediction is possible, providing a "proof-of-principle" of the feasibility of a prospective alarming system.

Keywords

EPILEPSIAE database • Seizure prediction • Machine learning • Stacked auto-encoders • Deep learning

Introduction

Despite available drug and surgical treatment options, more than 30% of patients with epilepsy continue to experience seizures [1]. In these patients with pharmaco-resistant epilepsy, the apparent unpredictability of seizure occurrence

R. Barata · B. Ribeiro · A. Dourado · C. A. Teixeira (✉)
Centro de Informática e Sistemas (CISUC), Departamento de Engenharia Informática, Universidade de Coimbra, Coimbra, Portugal
e-mail: cteixei@dei.uc.pt

imposes a considerable limitation of their daily lives and results in a high risk of unforeseen endangering situations [2, 3]. A system that could warn the patient of an impending seizure or trigger an antiepileptic device to prevent seizure occurrence would dramatically improve the quality of life for especially these patients.

Although some promising results have been reported, suggesting the existence of a pre-seizure state and an consequently the capability to predict seizures [4], until recently prediction performances have not been prospectively evaluated on large, collaborative databases [3]. Thus, the question to which extent seizure prediction is possible remains unanswered [5–7]. In particular, the weaknesses of most studies are (1) that they optimize parameters retrospectively, (2) that they rely on short, selected data sets, and (3) that they lack rigorous statistical evaluation [4]. In [8, 9] report for the first time studies that make use of an appropriate database developed on behalf of the EPILEPSIAE project [10, 11]. The EPILEPSIAE database is used in our study.

The aim of this study is to investigate for the first time whether a deep-learning technique based on Stacked Auto-encoders (SA), and having as input EEG-based features are able to predict seizures quasi-prospectively on a large multiple-center database of long-term unselected recordings. Quasi-prospectively here refers to the fact that the data have actually been recorded, which renders a prospective study impossible. But as we only evaluate our results on unevaluated validation data, we are as close to a prospective study as possible.

It was referred by several authors that, to date, no single feature has shown predictive power [4, 12, 13]. Machine-learning methods can face at the same time several features computed from raw data, and collected from different cerebral sites. These methods try to classify the brain state based not only on a single feature, but also on the general behavior of the set of features, exploring their linear and non-linear interactions. This paper as others already published, such as [13], hypothesize that a patient-specific approach based on multiple EEG measures will achieve high sensitivity and specificity.

Data & Methods

In this section, the database used as well as the methodology employed are described. In this paper, we distinguish "classifier" from "predictor". Classifier is the first part of the prediction system that discriminates the feature samples in four brain states (classes). A predictor is the full system composed by the classifiers plus the alarm generation procedure.

Patient Characteristics and EEG Database

Long-term EEG recordings from 103 epilepsy patients (50 males; age range, 10–65 years; mean age: 35 years) suffering from medically intractable partial epilepsy were analyzed in this study. Data had been recorded in three different epilepsy units (Hospitais da Universidade de Coimbra, Portugal; Unité d'Épilepsie of the Pitié-Salpêtrière Hospital, Paris, France; Epilepsy Center, University Medical Centre of Freiburg, Germany) resulting in a total of almost 707 days (16,963 h) of EEG including 1062 seizures, and is part of the 275 patients containing in the EPILEPSIAE database [10, 11]. These 103 patients are those that have more than 5 seizures and enable the evaluation strategy implemented in this paper.

43% of the patients had temporal lobe epilepsy, lateralized to the right in 57%, to the left in 30% and bilateral in 13%. In 84 patients, EEG was recorded using 22–37 scalp electrodes; the average recording period was 149 h. In 19 patients, intracranial EEG with 14–121 recording sites was recorded using stereotactically implanted depth electrodes, subdural grids and/or strips; the average recording period was 232 h. EEG data were recorded using a Nicolet, Micromed, Compumedics, or Neurofile NT digital video EEG system at sampling rates of 256, 400, 512, 1024 or 2500 Hz.

Feature Extraction

For the recorded EEG channels, 22 univariate features were extracted every five seconds of EEG with no overlap, i.e., consecutive five-seconds windows were considered for all the analyzed patients. The computed features are listed in Table 1. More details can be obtained in [14]. Features have to be computationally efficient, i.e., with potential for online implementation in low computational power environments. This is why we restricted ourselves to univariate linear features in this study.

The Approach to a Quasi-prospective Study

The same feature was computed for all the electrodes and for all the patients. The evaluation methodology implemented in this paper encompasses the splitting of these time series into three parts. A first part containing the first three seizures, was used for optimizing the classifiers, i.e., for training. A second part containing the next two seizures was used for the selection of an "optimal" predictor among all the tested alternatives. The third and last part, containing at least one

Table 1 EEG features

Time domain	Mean
	Variance
	Skewness
	Kurtosis
	Energy
Frequency domain	MSE of estimated AR models
	Delta band rel. power (0.1–4 Hz)
	Theta band rel. power (4–8 Hz)
	Alpha band rel. power (8–15 Hz)
	Beta band rel. power (15–30 Hz)
	Gamma band rel. power (> 30 Hz)
	Spectral edge frequency (90%)
	Spectral edge power (90%)
	Decorrelation time
	Hjorth mobility
	Hjorth complexity
Time frequency	Energy of DB4 wavelet coefficients (6 decomposition levels)

seizure, was used for quasi-prospective long-term out-of-sample evaluation, i.e., for validation. The average interictal duration of the validation data was approximately 88 h, and 54 h for the invasive, and scalp recordings, respectively. In total, 6111 h of interictal out-of-sample data were considered containing a total of 557 seizures.

Classification Methodology

In this paper, the brain state is classified over time into one of two states: preictal and non-preictal; and the number of classifier's inputs are 22 times the number of channels, originating a very high dimensional space. The windows that are just before the seizure onset times are nominated as preictal windows. The non-preictal class encompass the ictal, postictal and interictal periods. The ictal period is the time frame where seizure occurs. The postictal period refers to epochs that are just after the seizure offset time. The interictal period relates to seizure-free epochs. The neurologists clinically defined the seizure onset and offset times. The exact time where the preictal state starts is unknown and can probably be patient-dependent. Thus, in a first approach, the search for appropriate predictors should include the consideration of a range of preictal times, or seizure occurrence periods (SOPs). In this work, four SOPs were considered: 10, 20, 30 and 40 min.

In this paper, we use machine-learning techniques to define decision boundaries between classes. The machine learning technique used was SA [15]. SA can be considered a deep machine learning algorithm, since there are based on the construction of a new representation of the data for posterior classification, a technique known as representation leaning. The neural networks developed contained between three and six layers, and the number of neuron varies between 1000 and 10 per layer, diminishing from the input to the output of the network. Their auto-encoders were trained according to a two-phase protocol: Greedy Layer-Wise Training procedure and that were then fine-tuned [16]. The first stage of the training is used to optimize each auto-encoder, individually, to compress and restructure the data in the best way possible, by introducing a bottleneck on the network [17]. The second stage of fine-tuning, is meant to lead the network to be more discriminative regarding the classes [16].

By making use of such a sophisticated technic we hope to overcome the obstacles faced when using shallow or classic classifiers. Moreover, due to the data dimensionality a system capable of automatically reduce the number of dimensions while likely retaining only the important information justify the use of SA.

Seizure Prediction Method

If we consider that a single positive classifier output represents a prediction, we can have a prediction at each 5 s. As in practice a classifier will most likely not classify all of the samples correctly, a lot of false alarms could be issued. To reduce the number of false alarms the output of the classifier was smoothed using a method described in [18, 19]. This method is based on a sliding window with a size equal to the

considered preictal time, or SOP. Alarms are raised whenever the number of epochs classified as preictal in the sliding window is larger than a predefined threshold. The thresholds considered for each patient were 0.2, 0.4, 0.6 and 0.8, meaning that 20%, 40%, 60%, and 80% of the samples were classified as preictal inside the window, respectively. Once an alarm is raised, another one can only occur after a dead-time equal to the SOP. The main advantage of the algorithm implemented is that it is exactly known during which time period a seizure is to be expected. In the case of approaches that allow re-triggering, such as the one presented in [20], the alarms may prolong for long times and the patients do not know definitely when the seizure is to be expected.

Quasi-prospective Performance Evaluation

The performance of the seizure prediction algorithm is evaluated using the seizure prediction characteristics, which characterizes the sensitivity of the seizure prediction algorithm, given the maximum rate of false alarms and two-time windows: the SOP and the time needed to perform any intervention, called the intervention time (IT). A correctly predicted seizure requires that the seizure onset to occur in the time window occurrence period. In this paper, SOP assumes one of the values defined for the preictal time, while the IT was fixed as 10 s. As specificity, we used the false prediction rate defined as the number of false alarms divided by the duration during which false alarms could be triggered, which is obtained by subtracting the time under false warning from the total interictal duration. Mathematically the FPR is given by [4]:

$$FPR = \frac{\#False\,Alarms}{Interictal\,Duration - (\#False\,Alarms \times SOP)}.$$
(1)

To statistically evaluate the results we use the analytic random predictor based on the binomial distribution [3, 21]. It quantifies critical sensitivities that could be obtained by chance given the time windows and the false prediction rate. No other information of the data is provided to the random predictor. Because we implemented a quasi-prospective evaluation protocol that resulted in the selection of a best predictor, we only tested a single predictor, i.e., a single degree of freedom is attained, leaving out the need for multiple testing corrections.

Results and Discussion

For each patient four different preictal times were tested, as well as four different thresholds used on the alarm generation system, making 16 different predictors per patient. The predictor applied on the validation data was selected based on the performance on the testing set, and was the one with performance closest to 100% sensitivity, and 0/h FPR. This selection was based on a period containing two seizures that were not used for the training of the classifiers nor for any performance evaluation.

For the patients with intracranial electrodes the validation data consisted of 1664.42 h of recording and 151 seizures. For the patients with scalp electrodes the validation data consisted in 4445.57 h and 406 seizures.

We found out that for the patients with intracranial electrodes 18% of the seizures were correctly predicted (27), leading to an average sensitivity of 16.05% and average false positive rate of 0.27/h. For the patients with scalp electrodes 20.69% of the seizures (84) on the validation set were correctly predicted, leading to an average sensitivity of 17.49% and an average false prediction rate of 0.88/h.

We have compared these performances to those achieved by a random predictor [21]. We found that preictal changes can be identified above chance level in four out of 19 intracranially monitored patients ($\approx$ 21%) and in five out of 84 scalp monitored patients ($\approx$ 6%).

Regarding the influence of certain variables in the results, several important conclusions can be taken. For the patients with intracranial electrodes, the sensitivity of the female subjects was less than half of the male patients, 9.70% and 24.71%, respectively. Regarding the focal character of the seizures, the sensibility is considerably higher on patients with well-defined focalization. Moreover, three of the four patients with results above chance level have well defined focalization character, and all of these have focus on the frontal right lobe. Regarding the patients with scalp electrodes, the general panorama is that sleep stage influences sensitivity (paired t-test p-value < 0.05). For the scalp population, the preictal period also significantly influences sensitivity (paired t-test p-value = 0.01). Of the different preictal period durations tested, 30 min was the value that leaded to better results.

Comparing our results with the ones published in [9] that employed a realistic validation schema similar to the one implemented in this paper, similar results were obtained. The advantage of our approach relies on the automatic feature and channel selection accomplished by the deep-learning technique used.

Conclusions

We approached seizure prediction to the best of our knowledge for the first time based on SA, applied in multicentre, long-term, unselected data. The result argues for the possibility to predict seizures, at least for some patients. Future work will be devoted to a true online assessment and evaluation of seizure prediction performance.

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Kwan P, Brodie MJ (2000) Early identification of refractory epilepsy. N Engl J Med 342:314–319

2. Cockerell OC, Johnson AL, Sander JW et al (1994) Mortality from epilepsy: results from a prospective population-based study. Lancet 344:918–921. https://doi.org/10.1016/S0140-6736(94)92270-5

3. Schulze-Bonhage A, Sales F, Wagner K et al (2010) Views of patients with epilepsy on seizure prediction devices. Epilepsy Behav 18:388–396. https://doi.org/10.1016/j.yebeh.2010.05.008

4. Mormann F, Andrzejak RG, Elger CE, Lehnertz K (2007) Seizure prediction: the long and winding road. Brain 130:314–333. https://doi.org/10.1093/brain/awl241

5. Lehnertz K, Litt B (2005) The first international collaborative workshop on seizure prediction: summary and data description. Clin Neurophysiol 116:493–505

6. Lehnertz K, Le Van Quyen M, Litt B (2007) Seizure prediction. In: Engel J, Pedley TA, Aicardi J (eds) Epilepsy A Compr. Lippincott Williams & Wilkins, Textb, pp 1011–1024

7. Stacey W, Le Van Quyen M, Mormann F, Schulze-Bonhage A (2011) What is the present-day EEG evidence for a preictal state? Epilepsy Res 97:243–251. https://doi.org/10.1016/j.eplepsyres.2011.07.012

8. Teixeira CA, Direito B, Alexandre Teixeira C et al (2014) Epileptic seizure predictors based on computational intelligence techniques: a comparative study with 278 patients. Comput Methods Programs Biomed 114:324–336. https://doi.org/10.1016/j.cmpb.2014.02.007

9. Direito B, Teixeira CA, Sales F et al (2017) A realistic seizure prediction study based on multiclass SVM. Int J Neural Syst 27:1750006. https://doi.org/10.1142/S012906571750006X

10. Klatt J, Feldwitch-Drentrup H, Ihle M et al (2012) The EPILEPSIAE database—an extensive electroencephalography database of epilepsy patients. Epilepsia 53:1669–1676. https://doi.org/10.1111/j.1528-1167.2012.03564.x

11. Ihle M, Feldwisch-Drentrup H, Teixeira CA et al (2012) EPILEPSIAE—a European epilepsy database. Comput Methods Programs Biomed 106:127–138

12. Feldwisch-Drentrup H, Schelter B, Jachan M et al (2010) Joining the benefits: combining epileptic seizure prediction methods. Epilepsia 51:1598–1606

13. Park Y, Netoff T, Parhi K (2010) Seizure prediction with spectral power of eeg using cost-sensitive support vector machines. J Med Device 4:27542

14. Teixeira CA, Direito B, Feldwisch-Drentrup H et al (2011) EPILAB: a software package for studies on the prediction of epileptic seizures. J Neurosci Methods 200:257–271. https://doi.org/10.1016/j.jneumeth.2011.07.002

15. Hinton GE, Salakhutdinov RR (2006) Reducing the dimensionality of data with neural networks. Science (80-) 313:504–507

16. Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press

17. Erhan D, Bengio Y, Courville A et al (2010) Why does unsupervised pre-training help deep learning? J Mach Learn Res 11:625–660

18. Teixeira CA, Direito B, Feldwisch-Drentrup H et al (2011) EPILAB: a software package for studies on the prediction of epileptic seizures. J Neurosci Methods 200:257–271. https://doi.org/10.1016/j.jneumeth.2011.07.002

19. Teixeira C, Direito B, Bandarabadi M, Dourado A (2012) Output regularization of svm seizure predictors: Kalman filter versus the "firing power" method. In: 2012 annual international conference on IEEE engineering medicine biology society (EMBC), pp 6530–6533

20. Snyder DE, Echauz J, Grimes DB, Litt B (2008) The statistics of a practical seizure warning system. J Neural Eng 5:392–401. https://doi.org/10.1088/1741-2560/5/4/004

21. Schelter B, Winterhalder M, Maiwald T et al (2006) Testing statistical significance of multivariate time series analysis techniques for epileptic seizure prediction. Chaos 16:13108. https://doi.org/10.1063/1.2137623