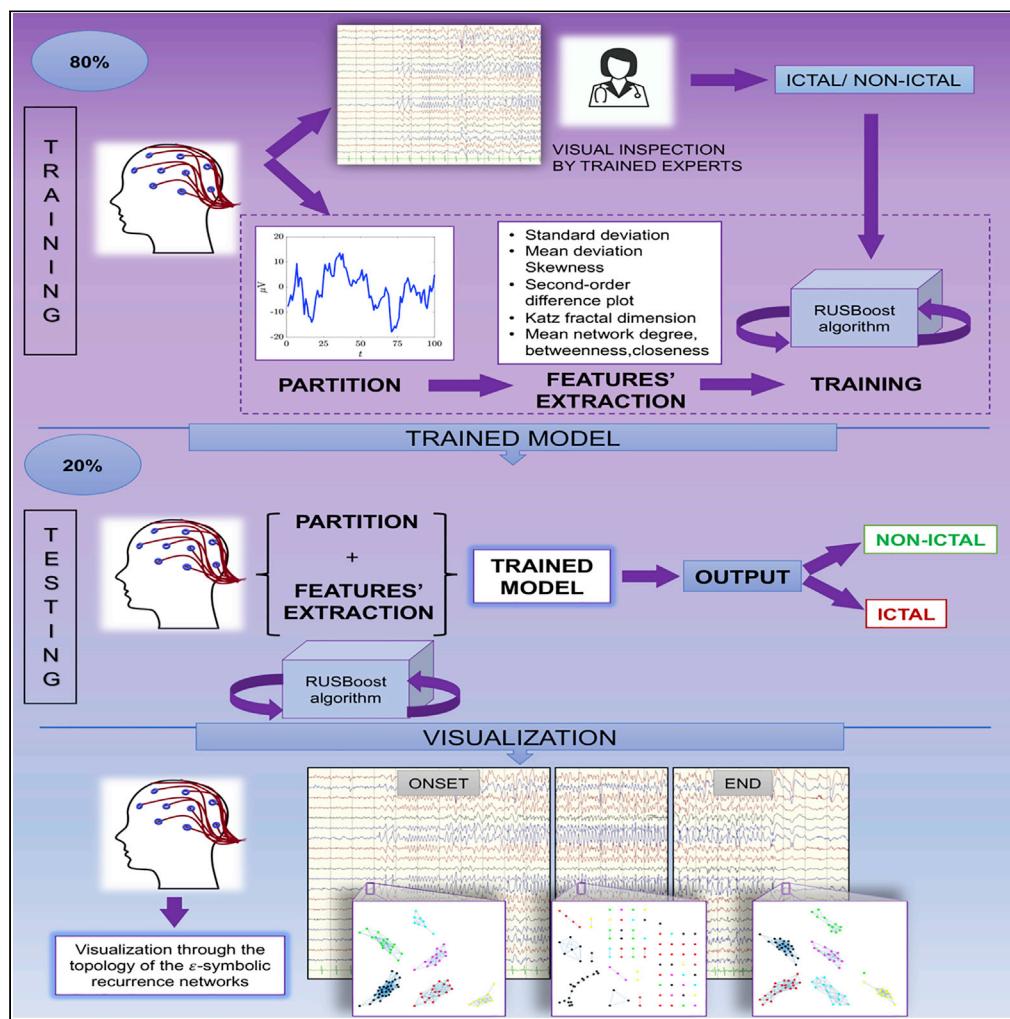


## Article

## Integrating old and new complexity measures toward automated seizure detection from long-term video EEG recordings



Manuel Ruiz  
Marín, Irene  
Villegas Martínez,  
Germán  
Rodríguez  
Bermúdez,  
Maurizio Porfiri

manuel.ruiz@upct.es (M.R.M.)  
irene.villegas@carm.es (I.V.M.)

## HIGHLIGHTS

Complexity measures are formulated to enhance classical time-domain statistics of EEG

The detection algorithm does not need ad-hoc data preprocessing to address artifacts

Focal seizures are detected 95% of the time with less than four false alarms per day

The approach offers a visual representation of a seizure as a time-evolving network



## Article

## Integrating old and new complexity measures toward automated seizure detection from long-term video EEG recordings

Manuel Ruiz Marín,<sup>1,5,6,\*</sup> Irene Villegas Martínez,<sup>2,5,6,7,\*</sup> Germán Rodríguez Bermúdez,<sup>3</sup> and Maurizio Porfiri<sup>1,4</sup>

## SUMMARY

**Automated seizure detection in long-term video-EEG recordings is far from being integrated into common clinical practice. Here, we leverage classical and state-of-the-art complexity measures to robustly and automatically detect seizures from scalp recordings. Brain activity is scored through eight features, encompassing traditional time domain and novel measures of recurrence. A binary classification algorithm tailored to treat unbalanced dataset is used to determine whether a time window is ictal or non-ictal from its features. The application of the algorithm on a cohort of ten adult patients with focal refractory epilepsy indicates sensitivity, specificity, and accuracy of 90%, along with a true alarm rate of 95% and less than four false alarms per day. The proposed approach emphasizes ictal patterns against noisy background without the need of data preprocessing. Finally, we benchmark our approach against previous studies on two publicly available datasets, demonstrating the good performance of our algorithm.**

## INTRODUCTION

Epilepsy is a common and chronic group of neurological disorders. Worldwide, more than 65 million people of all age have epilepsy, affecting individuals (Devinsky et al., 2018). Despite optimal medication management, about 30% of persons with epilepsy (PWE) will continue to have uncontrolled seizures and will eventually need a presurgical evaluation in an epilepsy monitoring unit (Kwan et al., 2010; Sisodiya, 2007).

Electroencephalography (EEG) is an essential tool in the evaluation, diagnosis, and management of epilepsy (Rao and Lowenstein, 2015). The International League Against Epilepsy recommends long-term video-EEG monitoring (LEM) in PWE, when there is diagnostic uncertainty in the classification of seizure type or epilepsy syndrome, quantification of seizures, and evaluation of electroclinical seizure characteristics prior to an epilepsy surgery (Velis et al., 2007). Whether measurements are conducted on the scalp or intracranially, the recordings of LEM can last from 24 h to 78 days (Ryvlin et al., 2014).

Visual scanning of LEM recordings by expert epileptologists is the conventional approach to seizure detection. EEG manifestations of epileptic events can take various forms, including desynchronization, decrease in amplitude, appearance of moderate or high amplitude rhythmic activity at frequencies ranging from 1 to 50 Hz, presence of high amplitude electromyogram (EMG) obscuring the EEG, and irregular paroxysmal activity (Fisher et al., 2014a). The complexity of these phenomena constitutes a significant hurdle to manual analysis of LEM recordings, which could often lead to errors even when conducted by highly trained physicians (Wilson et al., 2003). These errors are exacerbated in the analysis of noisy scalp data, which are, however, the most common, due to obvious challenges in implanting electrodes in the brain. Automatic seizure detection systems offer a promising alternative to efficiently and accurately analyze LEM recordings (Baumgartner and Koren, 2018).

Many attempts have been made to automatize LEM readings (Saini and Dutta, 2017; Sharmila and Geethanjali, 2019). Since the first algorithm for seizure detection proposed by Gotman (1982), several research teams have tried to reach the holy grail, that is, to detect every seizure in LEM recordings and minimize the number of false alarms that would slow down the physician, instead of easing his task.

<sup>1</sup>Department of Quantitative Methods, Law and Modern Languages, Technical University of Cartagena (UPCT), Cartagena, Murcia 30201, Spain

<sup>2</sup>Department of Projects and Innovation, Health Service of Murcia (SMS), Murcia, Spain

<sup>3</sup>University Centre of Defense at the Spanish Air Force Academy, Murcia, Spain

<sup>4</sup>Department of Mechanical and Aerospace Engineering, and Department of Biomedical Engineering New York University Tandon School of Engineering (NYU), Brooklyn, NY, USA

<sup>5</sup>Bio-Health Institute (IMIB-Arrixaca), Health Science Campus, Murcia, CP 30120, Spain

<sup>6</sup>These authors contributed equally

<sup>7</sup>Lead contact

\*Correspondence:  
manuel.ruiz@upct.es  
(M.R.M.),  
irene.villegas@carm.es  
(I.V.M.)

<https://doi.org/10.1016/j.isci.2020.101997>



The simplest class of detection algorithms uses intuitive time-domain features to create some discriminating statistics between seizure and “non-seizure” epochs. Mean, variance, mode, median, and skewness are some of the common statistics that are employed in seizure detection, along with amplitude difference and time separation between minima and maxima (Alotaiby et al., 2014). Although easy to interpret, these features alone are not sufficient to decipher brain activity. Measures of complexity from time-series analysis could complement classical, statistically based features toward improved detection of seizures (Kannathal et al., 2005). Information theory, fractal analysis, and symbolic representations have been leveraged to establish a range of powerful measures of complexity for LEM recordings (Saini and Dutta, 2017). For example, the theoretic notion of fractal dimension has been found to have a precise clinical application in the quantification of rhythmic patterns during an epileptic seizure (Wang et al., 2013).

Despite the multitude of automated computer-aided detection algorithms published in the past decades, why are epileptologists still relying on visual scanning when it comes to seizure detection?

(1) Existing algorithms are typically validated on the dataset by the University of Bonn (Andrzejak et al., 2001). These data are from intracranial recordings, where artifacts are virtually nonexistent compared with scalp data. In addition, the dataset only includes selected segments of ictal activity, excluding non-ictal recordings from the same patient. Validating an algorithm on this dataset prior to testing on real scalp data may lead to not only high sensitivity but also an unacceptably high number of false alarms (Hopfen-gärtner et al., 2014). As a result, prudence is commonly warranted in transitioning from the validation on the database of the University of Bonn to clinical practice.

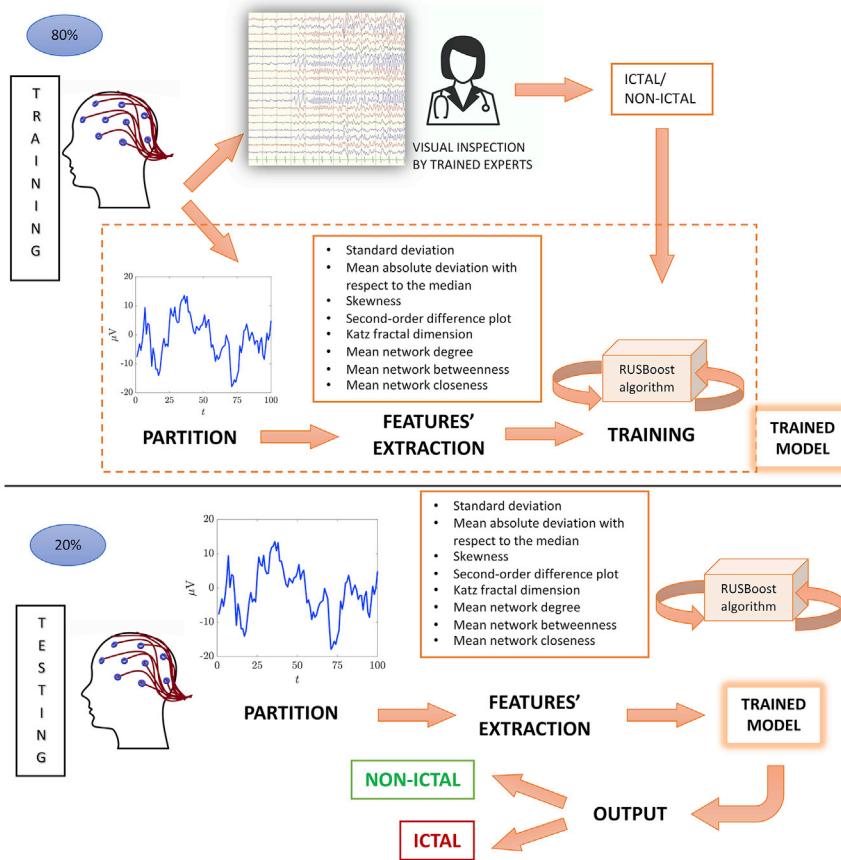
(2) EEG signals are highly complex, nonlinear, and non-stationary processes (Cohen, 2017). Most of the existing approaches utilize pre-processing techniques to remove artifacts, which may come at the expense of masking true brain activity and adding confounding effects (de Cheveigné and Nelken, 2019). For example, the method proposed by Fürbass et al. (2015) in the EpiScan study analyzes digital LEM recordings in intervals of a quarter-second and utilizes dedicated low-pass, high-pass, and notch filters to process their data. The EpiScan algorithm was lately implemented for commercial purposes into the Encevis software. Similar filtering techniques are applied in the Reveal algorithm by Wilson et al. (2004), which is behind the Persyst software.

(3) When confronted with everyday clinical practice, experimental models and commercialized software do not perform as well as promised. Gonzalez-Otalura et al. have recently reported a detection of only 53% of seizures by Persyst 12, Persyst 13, and Gotman in a sample of 1,478 ambulatory prolonged EEG studies (González-Otárla et al., 2019).

As a result, existing algorithms are far from replacing manual interpretation of LEM recordings by epileptologists.

Here, we seek to address this issue through a methodology that integrates a range of EEG-specific features, from traditional statistics in the time domain to state-of-the-art complexity measures. Along with the computation of several time-domain characteristics, we introduce an alternative approach toward recurrence quantification. Such an approach combines classical and symbolic recurrence to mitigate measurement noise that is known to plague classical recurrence analysis and avoid crude coarsening due to the limited alphabet of a symbolic representation (Caballero Pintado et al., 2018; Porfiri and Ruiz Marín, 2019). Using this notion of  $\epsilon$ -symbolic recurrence, we construct a recurrence network, whose topological characteristics are utilized to enhance the discrimination between ictal and non-ictal activity from scalp LEM.

Overall, brain activity from LEM recordings is scored through eight features, encompassing time-domain statistics (standard deviation, mean absolute deviation, skewness coefficient, fractal dimension, and area of the second-order difference plot) and measures of recurrence (mean degree, betweenness centrality, and closeness of  $\epsilon$ -symbolic recurrence networks). By applying an algorithm for binary classification (RUS-Boost algorithm (Seiffert et al., 2009)), on this range of features (evaluated on small time segments), we explore the possibility of automatic detection of seizures in scalp LEM recordings, without the need of any pre-processing. More specifically, the classification algorithm is first trained using a fraction of manually analyzed LEM recordings by clinical experts, and then it is tested on the remaining fraction of the dataset, without clinical supervision (Figure 1).



**Figure 1. Sketch of the proposed algorithm for automated seizure detection during training and testing**

Figure360▷

For a Figure360 author presentation of this figure, see <https://doi.org/10.1016/j.isci.2020.101997>.

The algorithm takes as input the LEM recording signal and partitions it into non-overlapping windows. For each window, it extracts eight features (descriptive statistics and complexity measures) that are used for classification by the RUSBoost algorithm. During training, clinical supervision is needed to determine the onset and ending of a seizure for 80% of the data. During testing, no clinical supervision is required, and the trained model is employed to classify the remaining 20% of the data, within 5-fold cross-validation.

## RESULTS

To demonstrate the performance of the detection algorithm, we examined a total of ten LEM recordings, from three men and seven women (age: 25–53 years, mean: 37.5 years; **Table 1**) with refractory focal epilepsy from our original database. Each patient in the cohort suffered from at least one epileptic seizure during the 24-h monitoring comprising the dataset.

Upon completing the study of our original dataset, we benchmarked our approach against published results, by focusing on two publicly available databases: the University of Bonn (UB) dataset and Temple University Hospital EEG Seizure Corpus (TUSZ). The former dataset offers the possibility to test our algorithm on intracranial data, whereas the second contributes to improved validation on surface recordings.

### The detection algorithm yields about 90% accuracy, specificity, and sensitivity

Results of the application of our detection algorithm to LEM recordings of the ten patients in **Table 1** using 5-fold cross-validation are summarized in **Table 2**. For each patient, we report sensitivity, specificity, accuracy, true alarm rate (TAR), and false alarm rate (FAR) for the clinically selected channel, according to the clinical profiles in **Table 1**. All these metrics were assessed from ground-truth data compiled by two clinical experts, on each of the 100-observation windows of 0.391 s that comprised the recording. Ground-truth

**Table 1. Clinical data and seizure description of LEM recordings**

Patient	Epileptic syndrome	Semiology of seizures during LEM	Duration (seconds)	Localization/ lateralization at onset	Seizure onset
1	Structural epilepsy. Left mesial hippocampal sclerosis	Seizure 1: focal unaware to bilateral TC	95	Mesial and anterior left temporal	Sharp rhythmic activity
		Seizure 2: focal unaware with non-motor onset (cognitive)	44	Mesial and anterior left temporal	Sharp rhythmic activity
2	Structural epilepsy. Right mesial hippocampal sclerosis	Seizure 1: focal unaware with non-motor onset (cognitive)	57	Right temporal	Low-voltage fast activity
		Seizure 2: focal unaware with motor onset (automatisms)	65	Right temporal	Low-voltage fast activity
		Seizure 3: focal unaware with motor onset (automatisms)	74	Right temporal	Low-voltage fast activity
		Seizure 4: focal unaware with non-motor onset (cognitive)	46	Right temporal	Low-voltage fast activity
3	Structural epilepsy. Right mesial hippocampal sclerosis	Focal unaware with motor onset (automatisms)	68	Anterior right temporal	Sharp rhythmic activity
4	Structural epilepsy. Left mesial hippocampal sclerosis	Seizure 1: focal unaware with motor onset (automatisms)	74	Anterior left temporal	Low-frequency high-amplitude rhythmic spikes
		Seizure 2: focal unaware with motor onset (automatisms)	64	Left temporal	Low-frequency high-amplitude rhythmic spikes
		Seizure 3: focal unaware with motor onset (automatisms)	51	Left temporal	Spike-and-wave activity
5	Structural epilepsy. Right mesial hippocampal sclerosis	Focal unaware to bilateral TC	487	Anterior and mesial right temporal	Sharp rhythmic activity
6	Structural epilepsy. Left mesial hippocampal sclerosis	Seizure 1: subclinical	33	Mesial left temporal	Sharp rhythmic activity
		Seizure 2: focal unaware with non-motor onset (cognitive)	47	Left temporal	Sharp rhythmic activity
7	Epilepsy of unknown origin	Seizure 1: focal unaware with non-motor onset (cognitive)	65	Left front-otemporal	Low-voltage fast activity
		Seizure 2: focal unaware to bilateral TC	88	Left fronto-temporal	Low-voltage fast activity

*(Continued on next page)*

**Table 1. Continued**

Patient	Epileptic syndrome	Semiology of seizures during LEM	Duration (seconds)	Localization/ lateralization at onset	Seizure onset
8	Structural epilepsy. Cortical dysplasia right temporal lobe	Seizure 1: focal unaware seizure with non-motor onset (behavior arrest)	78	Right temporal	Low-voltage fast activity
		Seizure 2: focal unaware to bilateral TC	121	Anterior and mesial right temporal	Low-voltage fast activity
9	Structural epilepsy. Left mesial hippocampal sclerosis	Seizure 1: focal unaware with motor onset (automatisms)	70	Left temporal	Low-voltage fast activity
		Seizure 2: focal unaware with motor onset (automatisms)	69	Left temporal	Low-voltage fast activity
		Seizure 3: focal unaware with motor onset (automatisms)	108	Mesial left temporal	Low-voltage fast activity
10	Epilepsy of unknown origin	Seizure 1: subclinical	24	Right temporal	Sharp rhythmic activity
		Seizure 2: focal aware with non-motor onset (behavior arrest)	36	Right temporal	Sharp rhythmic activity
		Seizure 3: focal aware with non-motor onset (behavior arrest)	37	Right temporal	Sharp rhythmic activity
		Seizure 4: focal unaware to bilateral TC	214	Posterior right temporal	Sharp rhythmic activity

Scalp EEG was recorded at a sampling rate of 256 Hz, with 19 electrodes placed according to the international 10–20 system, using a 64-channel system Nicolet™EEG NicOne. EEG was recorded from the following electrode positions: Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, and O2 and a reference electrode (Z). The classification of seizure onset was based on [Perucca et al. \(2014\)](#).

TC: tonic-clonic. LEM: long-term video EEG monitoring.

**Table 2. Sensitivity, specificity, accuracy, true alarm rate (TAR), and false alarm rate (FAR) per hour for 5-fold cross-validation analysis of 24-h LEM recordings of 10 different patients.**

Patient	Channel	Sensitivity	Specificity	Accuracy	TAR	FAR/h
1	T3	91.57%	90.61%	90.61%	100%	0.38
2	T4	92.57%	93.60%	93.59%	100%	0.42
3	F8	93.68%	88.00%	88.00%	100%	0.21
4	T3	92.96%	97.93%	97.92%	100%	0.00
5	F8	90.46%	95.29%	95.27%	100%	0.08
6	T3	88.78%	89.73%	89.73%	100%	0.04
7	F7	90.26%	97.00%	96.99%	100%	0.00
8	F8	66.01%	96.45%	96.36%	100%	0.25
9	T3	84.18%	88.40%	88.39%	100%	0.00
10	T4	85.96%	90.25%	90.24%	75%	0.04

The algorithm is implemented on a specific, single channel for each of the patient, based on clinical considerations in [Table 1](#).

data were provided by the experts in terms of the inception and the end of a seizure, such that a window was considered to be ictal if at least 50% of its observations pertained to a seizure.

By comparing the output of the classifier to ground-truth, we scored the number of true positives, false positives, true negatives, and false negatives. A true positive is an ictal window that is correctly classified as part of a seizure by the algorithm, whereas a false positive is a non-ictal window that is erroneously classified as part of a seizure. Likewise, a true negative is a non-ictal window that is classified to be outside of a seizure, and a false negative is an ictal window that is classified to be outside of a seizure.

Sensitivity measures the ratio between the number of true positives and the total number of positives in ground-truth (true positives and false negatives). A sensitivity of 100% indicates that the detection algorithm has no false negatives, such that it can correctly detect every ictal window to be part of a seizure. [Table 2](#) shows sensitivity values that are above 90% for most of the patients. For one patient (patient 8), we determined a modest sensitivity of 66%. Specificity quantifies the ratio between the number of true negatives and the total number of negatives in ground-truth (true negatives and false positives). A specificity of 100% identifies the case in which the algorithm has no false positives, whereby it never classifies a non-ictal window as part of a seizure. [Table 2](#) confirms the reliability of the algorithm, whose specificity was as high as 98% and never below 88%. Accuracy scores the overall predictive power of the algorithm, as the ratio between the total number of correctly classified windows (sum of the number of true positives and true negatives) and the total number of windows. Results from this metric are consistent with specificity and sensitivity, ranging from 88% to 98%.

### True alarm rate is about 95% and false alarms are less than four per day

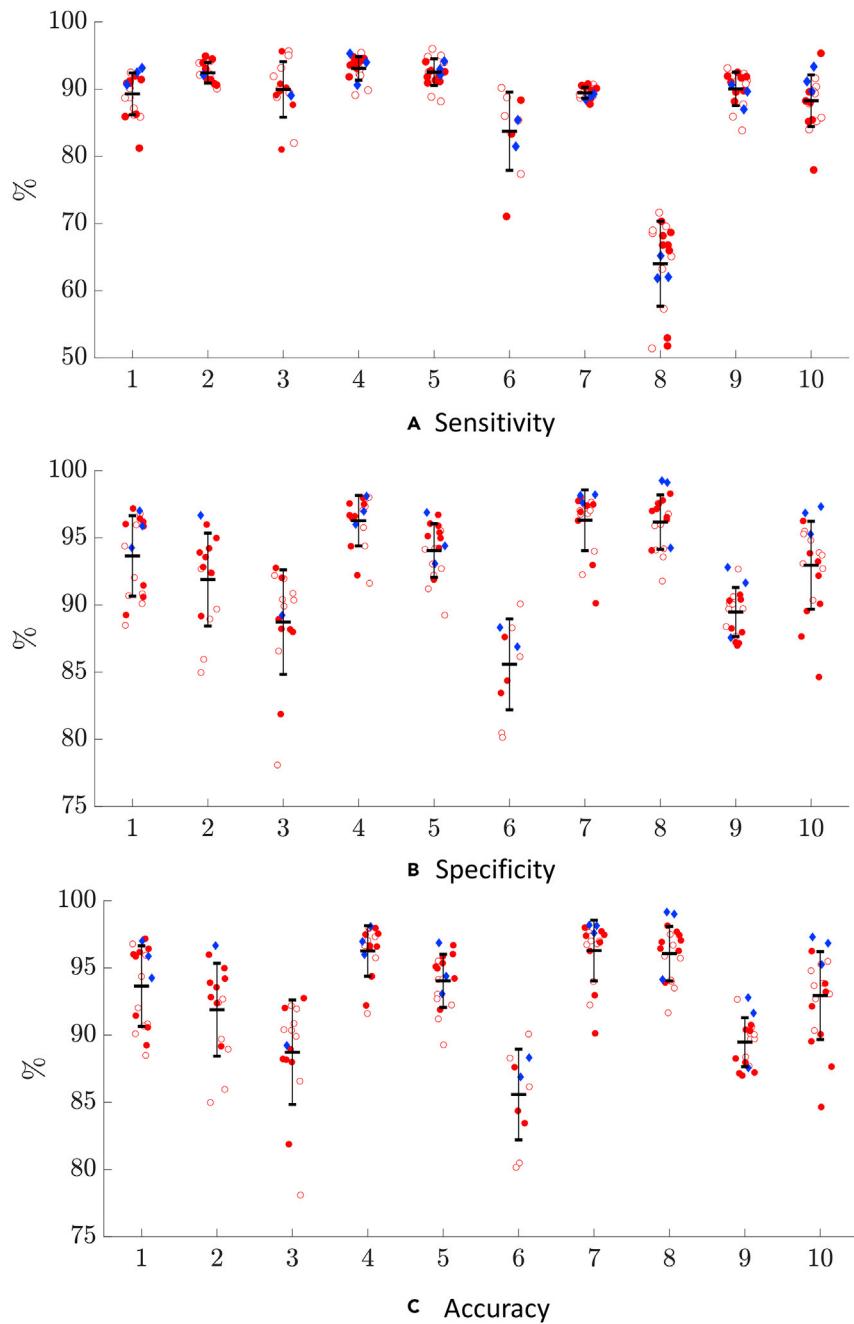
Specificity, sensitivity, and accuracy were all scored by examining each window independent of the other, without clinical consideration of the typical duration of a seizure that may be far larger than 0.391 s. As a result, these measures could be affected by isolated misclassification, thereby providing an overly conservative picture of the performance of the detection algorithm toward its clinical use. Consistent with previous literature ([Hopfengärtner et al., 2007](#); [Hunyadi et al., 2012](#)), we defined an alarm as a sequence of consecutive windows of at least 10 s in duration that are all classified as ictal. We joined instances of alarms, so that adjacent alarms are counted as the same one. TAR was calculated as the percentage of alarms that constitute a seizure in ground-truth, whereas FAR was assessed in terms of the number of events per hour to quantify the clinical burden of potentially verifying incorrect alarms.

TAR was 100% for all the patients, except for one patient (patient 10), thereby leading to an average of 95%. Notably, the patient for which TAR was not perfect does not correspond to the one who had the lowest sensitivity (patient 8). FAR was 0.14 events per hour on average, corresponding to at most four false alarms per day. For some patients, no false alarms were ever recorded during that 24-h period. In the worst case, FAR was 0.42 events per hour, corresponding to 10 false alarms in a day.

### The performance of the detection algorithm is marginally influenced by the selection of the channel and the patient's clinical characteristics

Although the results in [Table 2](#) are based on LEM recordings from a specific, single channel, the performance of the detection algorithm was robust with respect to the selection of the channel ([Figure 2](#)). Specifically, we run the algorithm on any of the channels where a seizure was detected by the clinical experts in average montage and performed 5-fold cross-validation to compute sensitivity, specificity, and accuracy.

Results in [Figure 2](#) demonstrate that the selection of the channel had a secondary influence on all the performance metrics, whereby the standard deviation was always about 5%. The channels with the lowest performance were those located in the frontopolar region, which are known to be more affected by artifacts ([Hopfengärtner et al., 2007](#)). The performance of the detection algorithm did not seem to be affected by the type of onset, whereby we recorded comparable performance metrics for different epileptic syndromes (mesial hippocampal sclerosis, cortical dysplasia, and unknown origin; [Table 1](#)), seizure semiology (focal aware, unaware, and bilateral tonic-clonic; [Table 1](#)), and type of onset (sharp rhythmic activity, low-voltage fast activity, and low-frequency high-amplitude rhythmic spikes; [Table 1](#)).

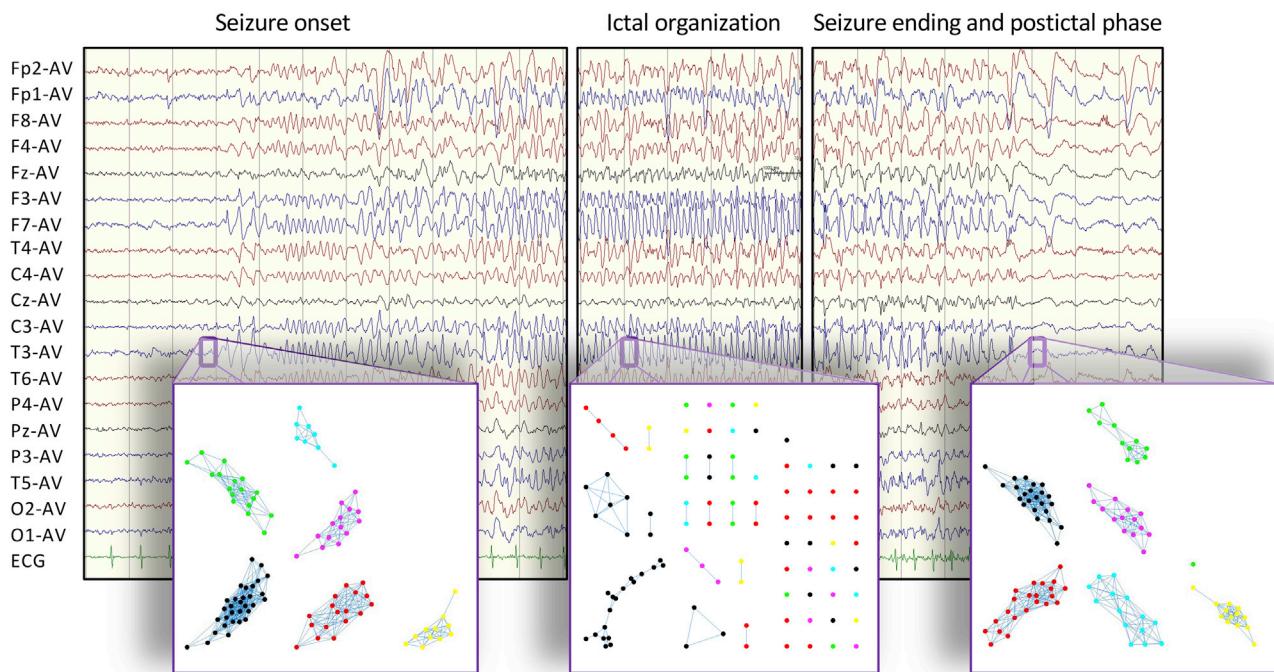


**Figure 2. Sensitivity, specificity, and accuracy (in percent) for 5-fold cross-validation analysis of 24-h LEM recordings of 10 different patients, from different channels of the LEM recordings**

For each patient and each metric, we report data from central (blue diamonds) electrodes and electrodes in the right (filled, red circles) or left (open, red circles) hemispheres, along with mean and standard deviation (black bars with whiskers).

#### **$\epsilon$ -symbolic recurrence networks can serve as a visualization aid for seizure detection**

Figure 3 illustrates the time-evolution of the topology of a  $\epsilon$ -symbolic recurrence network associated with a channel during a seizure. In the ictal window, the network contains many isolated nodes that contribute to low mean degree and closeness of the network, used as features by the RUSBoost algorithm. In the pre-ictal and post-ictal phases, the connectivity of the network increases and only one node remains isolated. Notably, these sharp changes in the topology of the network do not require a large number of



**Figure 3. Visualization of a seizure through the topology of the  $\varepsilon$ -symbolic recurrence, constructed from 100 observations (0.391 s) from a single channel (T3)**

For clarity, the network is overlaid with the EEG recordings to display the onset of the seizure, ictal organization, and seizure ending and post-ictal. The network is assembled using six symbols (embedding dimension  $m = 3$ ) and proximity parameter  $\varepsilon = 10 \mu\text{V}$ ; each color identifies symbolic recurrence to a different symbol. From the left to the right network, mean degree, betweenness centrality, and closeness are  $(11.04, 4.58, 1.37 \times 10^{-3})$ ,  $(1.22, 4.55, 0.16 \times 10^{-3})$ , and  $(9.77, 3.82, 1.21 \times 10^{-3})$ .

observations, whereby 100 datapoints are sufficient to visually discriminate different phases of the seizure.

### The performance of the detection algorithm compares well with existing methods that need pre-processing

Results of the application of our algorithm to the UB dataset are presented in Table 3; in the analysis, subsets Z, O, N, and F are regarded as interictal and subset S as ictal recording. Predictably, the use of intracranial recordings systematically improves sensitivity, specificity, accuracy, and FAR with respect to our original dataset in Table 2. Even though our approach does not require any form of data preprocessing, its performance is highly comparable to other methods that rely on Gaussian or band-pass filtering. In order to enhance the evaluation of the algorithm, we also tested its performance on surface LEM recordings from TUSZ. Results presented in Table 3 indicate that our algorithm outperforms other FAR and has sensitivity, specificity, and accuracy comparable to the best ones. Only one method (Raghu et al., 2020) leads to an appreciably higher sensitivity than our approach. However, in addition to filtering, the methodology presented therein requires an independent component analysis (multichannel-based), which further hinders its real-time use in seizure detection. Likewise, only one study reports a specificity superior to ours (Golmohammadi et al., 2017), but this comes at an unacceptable value of sensitivity that would hinder clinical use of the algorithm.

### DISCUSSION

In this work, we present a robust algorithm for epileptic seizure detection that yields high accuracy with non-filtered scalp LEM recordings. We tested the algorithm on a new dataset of ten adult patients with different epileptic syndromes, seizure semiology, and type of onset. For focal seizures, our results indicate sensitivity, specificity, and accuracy of about 90%, true alarm rate of 95%, and selectivity (measured as false alarm rate, FAR, per hour) of 0.14, in a 19-electrode common average montage. The selection of the channel has a marginal influence on the performance, which can vary by at most 5%. This could be explained by

**Table 3. Comparison of our detection algorithm against existing methods using the University of Bonn (UB) database and Temple University Hospital Seizure Corpus (TUSZ)**

Work		Patients/subsets	Window length	Preprocessing	Sensitivity	Specificity	FAR/ Accuracy h
<b>UB dataset</b>							
Tiwari et al. (2017)	ZONF-S	100 samples	Gaussian filters	93.10%	83.90%	88.50%	n. r.
Samiee et al. (2015)	ZONF-S	173 samples	n. r.	98.30%	91.60%	96.90%	n. r.
Diykh et al., (2017)	ZONF-S	384 samples	Band-pass filter 0.3–40 Hz	97%	98%	97.90%	0.04
Li et al., (2018)	ZONF-S	1 second	Band-pass filter 0.53–40 Hz	93%	90%	91%	n. r.
<i>This study</i>	ZONF-S	100 samples	None	94.48%	97.88%	97.20%	0
<b>TUSZ</b>							
Ayodele et al., (2020)	29	1 second	n. r.	78.35%	n. r.	n. r.	0.9
Golmohammadi et al., (2017)	246	21 second	n. r.	30.83%	91.49%	n. r.	0.25
Raghu et al., (2020)	316	1 second	Notch filter + band-pass filter (0.5–40 Hz) + ICA	95.50%	n. r.	n. r.	0.49
Tsiouris et al., (2018)	23	1 second	Band extraction (1–13 Hz)	84.92%	n. r.	n. r.	3.46
<i>This study</i>	13	100 samples	None	86.64%	87.04%	87.15%	0.14

Performance is presented in terms of sensitivity, specificity, accuracy, and false alarm rate (FAR) per hour. Results from our algorithm are displayed in italic to ease legibility.

n. r.: not reported. ICA: independent component analysis.

the type of seizures included in the analysis of our own dataset: unaware focal seizures or focal evolving to tonic-clonic. In both cases, changes on the EEG can be revealed in every channel, because of the generalized propagation of the seizure.

In addition, we benchmarked our algorithm on two publicly available datasets that have been used by other researchers to evaluate the performance of their algorithms. Whether the dataset comes from intracranial or surface electrodes, our algorithm offers a valuable alternative to existing methods that require data preprocessing, thereby opening the door to real-time seizure detection. Overall, the improved performance of our approach with respect to the state of knowledge is based on the integration of a range of EEG-specific features, from traditional statistics in the time-domain to unique  $\epsilon$ -symbolic recurrence measures.

### Sensitivity

Our detection algorithm is characterized by a high seizure detection performance, whereby we registered a sensitivity of more than 90% in at least 60% of the recordings from our dataset. Interestingly, the algorithm performed equally well across the entire spectrum of seizure patterns. Both focal and bilateral tonic-clonic seizures were successfully detected, independent of their type of onset. Even low-voltage fast-activity patterns, which have been found to be difficult to detect through other methods (Meier et al., 2008; Hopfen-gärtner et al., 2014; Bomela et al., 2020), did not challenge the application of our algorithm. Just as the algorithm performance was not affected by the seizure pattern, it did not vary across epileptic syndromes or etiologies. The accuracy of the detection was equivalent for temporal mesial sclerosis, cortical dysplasia, and unknown origin.

Specifically, we documented sensitivity values that were generally above 90% except for one patient, for whom it dropped to 66%. The reduced performance for this individual was likely due to the uniqueness of one of the seizures that was suffered by this patient (discontinuous discharge). Our method is trained to thoroughly recognize ictal patterns, characterized by the progressive reduction of the connectivity of the  $\varepsilon$ -symbolic recurrence network, along with sustained variations in the local growth of the time series. This specific seizure seemed to fade twice during the recording, and its pattern varied between its onsets. Although the algorithm was able to detect this seizure, some of the epochs during the event were not correctly classified as positive, thereby reducing the sensitivity.

The sensitivity of our algorithm compares very well with other studies on LEM recordings in adults. Following the lead of early studies by [Gotman \(1982\)](#), [Gotman \(1990\)](#), and [Gabor et al. \(1996\)](#), 15 years ago [Wilson et al. \(2004\)](#) introduced the Reveal algorithm, which was tested on a total of 1,049 h of EEG containing 672 seizures. The authors employed an eight-channel bipolar montage and a 60 s moving window to identify background, seizures, and offset sections, demonstrating a sensitivity of 76%. In the last 15 years, other groups, including [Hartmann et al. \(2011\)](#), [Kelly et al. \(2010\)](#), and [Hopfengärtner et al. \(2007\)](#), have put forward alternative algorithms, with sensitivity ranging from 79.5% (IdentEvent) to 90.9% (BESA).

### True alarm rate

Sensitivity alone cannot be used as a metric of the accuracy of the detection algorithm. The same value of sensitivity may correspond to vastly different scenarios, in which the algorithm perfectly detects all the windows of a seizure and miss entire seizures, or it captures almost all the windows in any seizure. The second scenario bears a higher practical relevance toward automated seizure detection. Although there is not an official minimum time duration to define a seizure, we chose a time duration of 10 s, which also allows to discriminate between sharp artifacts and meaningful subclinical epileptic discharges. This definition was based on the following grounding. First, several previous studies have used an analogous definition ([Hopfengärtner et al., 2007](#); [Hunyadi et al., 2012](#)), thereby facilitating comparisons between methods. Second, although a single generalized spike associated with a myoclonic jerk could be considered a very brief seizure, semiologically relevant events usually last more than 20 s ([Fisher et al., 2014b](#)). Finally, most artifacts last less than 10 s ([Schindler et al., 2001](#)).

Twenty-three of the 24 seizures in our original database were correctly detected by the algorithm in cross-validation analysis. The only event that was missed by the algorithm was very short in duration and limited to only two channels, wavering on the edge of temporal intermittent rhythmic delta activity and subclinical seizure. This finding is of great practical importance, whereby the usefulness of an automated seizure detection algorithm depends on its capacity to discriminate every significant seizure during LEM recordings, and, ideally, to not miss any of them.

In the study by [González-Otárla et al. \(2019\)](#), it is reported that almost 50% of seizures can be missed by automated detection algorithms (Persyst and Gotman Event Detection), thereby questioning the added value of automated approaches in comparison with visual analysis by clinical experts. Similar evidence regarding the possibility of missing several seizures using existing approaches has been widely documented in the technical literature. For example, [Kamitaki et al. \(2019\)](#) also found that the Persyst software detected 80 out of 105 seizures in a study with 38 patients, and [Rommens et al. \(2018\)](#) reported a 19.7% of missed seizures by Encevis EpiScan and BESA Epilepsy in a sample of 115 patients with 188 recorded seizures.

### Selectivity

The high sensitivity and excellent seizure detection rate were accompanied by a low false alarm rate. Working with our original dataset, the average FAR of our detection algorithm was of 0.14 events per hour (that is, less than four per day), which is in the range of the lowest FAR reported in the technical literature ([Baumgartner and Koren, 2018](#)). Just as high detection rate is required for a truly automated detection process, a low FAR is essential to minimize the clinical burden required for verifying the alarm and, potentially, act on it.

The integration of several features of the time series in the detection algorithm allows to faithfully classify windows of short duration, which is critical to recognize chewing and EMG artifacts without data manipulations or ad-hoc filtering. With only 100 observations (0.391 s) per window, our algorithm generates a comprehensive representation of brain activity, upon which to detect seizures. Chewing is a rhythmic artifact, which can be a potential confounder for seizures. By affording the analysis of small windows, we

successfully discriminated between a continuous rhythmic ictal pattern and a discontinuous, although repetitive, chewing pattern. The EMG artifact, instead, has a wide spectral distribution that perturbs all the classic EEG bands. In particular, EMG considerably overlaps with beta activity in the 15–30 Hz range but may be as low as 2 Hz (similar to chewing), making the widely used alpha band also vulnerable to muscle artifacts (de Cheveigné and Nelken, 2019). Previous automated seizure detection algorithms involve high and low band pass filters (Baumgartner and Koren 2018), which can lead to a loss of information. Our algorithm analyzes brain activity without succumbing to the need of artifact removal methods or, more importantly, filtering (Islam et al., 2016; Urigüen and Garcia-Zapirain, 2015)

### Benchmarking on publicly available datasets

The majority of automated seizure detection algorithms was tested on the intracranial dataset on adults by the University of Bonn (Andrzejak et al., 2001), scalp EEG data from children by the Massachusetts Institute of Technology (Fergus et al., 2015; Bomela et al., 2020), and, to a lesser extent, scalp data from adults from Temple University Hospital (Shah et al., 2018; Obeid and Picone, 2018). It is generally recognized that EEG signals differ with age (Sheth, 2019), therefore it can be difficult to extrapolate the results obtained in a children's dataset to an adult population. Hence, we benchmarked our approach against datasets from the University of Bonn and Temple University Hospital.

We demonstrated comparable or even superior performance to other methods on publicly available datasets, with the very same parameters utilized in the earlier implementation of the algorithm on our original dataset. Specifically, we analyzed five subsets of intracranial recordings from the University of Bonn dataset (Z, O, N, and F, against S) and scalp recordings from 13 patients from the Temple University Hospital Seizure Corpus. Predictably, for the intracranial dataset, our algorithm performs even better than on our original dataset, reaching a sensitivity of approximately 94%, a specificity of 98%, accuracy of 97%, and no false alarms. This performance compares very well with existing methodologies (Tiwari et al., 2017; Samiee et al., 2015; Diykh et al., 2017; Li et al., 2018), which require data preprocessing for artifact removal, in the form of Gaussian or band-pass filters.

With respect to the dataset by Temple University Hospital, we report sensitivity, specificity, and accuracy close to 87% and 0.14 false alarms per hour. These performance values are highly comparable to those attained during the analysis of our dataset, thereby supporting the robustness and reliability of our algorithm in the detection of seizures from LEM scalp recordings. Compared with other algorithms that were benchmarked against this dataset (Ayodele et al., 2020; Golmohammadi et al., 2017; Raghu et al., 2020) our approach offers the best false alarm rate and its sensitivity, specificity, and accuracy are in line with the best available methods. Once again, in contrast with the literature, our algorithm does not require data preprocessing, thereby favoring real-time applications.

Existing methods are typically tailored to a specific set of scalp or intracranial recordings, characterized by its own unique features. The ability to successfully detect seizures across three different datasets, spanning intracranial and scalp recordings, is a key strength of our algorithm. Importantly, this is achieved without any parameter tuning, whereby the same window length, proximity parameter, and embedding dimension are used across all datasets. In general, we recommend the validation of new methods to mirror a similar approach, without exclusively focusing on one database over another. Prudence is particularly warranted when working with the popular database from the University of Bonn, where artifacts are virtually nonexistent and only selected segments of ictal activity are included. It is tenable that validation on this single dataset may lead to high sensitivity, but it ultimately results in an unacceptably high number of false alarms in clinical practice that uses scalp data.

### A step forward in the state-of-the-art on automated seizure detection

The introduction of  $\epsilon$ -symbolic recurrence in combination with other complexity measures, associated with fractal dimension and Poincaré plot, holds promise in automated seizure detection. Although classical,  $\epsilon$ -recurrence analysis has been pursued in previous studies on epilepsy (Acharya et al., 2011), its application suffers from two main limitations. First, measurement noise can add arbitrariness to the analysis, by challenging the selection of the  $\epsilon$ -neighborhood that must be chosen to balance between the need to minimize the effect of measurement noise in LEM recordings and the need to detect the seizure. Second, all recurrences are equivalently treated, without bookkeeping details about the local pattern of the LEM recording, which can be informative of the seizure.

A symbolic analysis can address both these issues by examining each LEM recording as a sequence of symbols, which robustly encode information about the local pattern. However, a symbolic approach alone (Caballero-Pintado et al., 2018) can create shortcuts in the recurrence analysis that could mask the inference of brain activity from EEG data. Assessing recurrence with both the symbolic approach and traditional  $\epsilon$ -neighborhood obviates this shortcoming, leading to meaningful network representations of brain activity. The notion of  $\epsilon$ -symbolic recurrence is a contribution of this study. Other seizure detection algorithms, based on a continuous representation of the EEG recordings, need to renounce part of the data in the interest of gaining accuracy (Meier et al., 2008; Kelly et al., 2010; Hartmann et al., 2011). Anchoring our approach in asymbolic representation of the dataset mitigates this need, whereby our algorithm works on a complete dataset. Converting the huge amount of information encoded in an EEG epoch into a coarse grain representation is similar to what epileptologists routinely do through visual inspection (Petras et al., 2019). Human brain can recognize a rhythmic activity hiding below mild EMG or chewing artifact, due to our ability to extract important information from noisy environments (DiCarlo et al., 2012). Our algorithm mirrors this very step through sequence of symbols that encapsulate salient dynamics of EEG signals, without confounds from measurement noise.

Upon  $\epsilon$ -symbolic recurrence, we can construct colored network representations that could assist in the visualization of epileptic seizures by epileptologists. Nodes in the network are associated with the time instants in the recording, links correspond to recurrence, and the coloring labels the specific recurrent pattern. By monitoring the density of the links in the network and their color, the epileptologist could visualize the brain activity during a seizure, which will progressively destroy the links in the network. Ultimately, it is possible to witness how the seizure starts, organizes, and ends, only by inspecting images. To the best of our knowledge, none of the current systems offers such a transparent visual aid to the study of seizures. Although digital EEG trend analysis and quantitative EEG with color-coded graphs have been performed in intensive care monitoring for seizure and status epilepticus detection, they are not easy to interpret and have relatively low sensitivity (Haider et al., 2016). A visual aid, easier to “read” than EEG graphoelements, could be of great help to avoid misinterpretations and eventual disagreements between electroencephalographers.

In addition, complexity measures such as Katz’s fractal dimension contribute to seizure recognition by scoring the regularity and divergence of signals (Litt and Echauz, 2002), which are indicators of rhythmic patterns (Wang et al., 2013). Even though features based on Poincaré maps are not commonly part of the toolbox of seizure detection algorithms, they offer valuable insight into brain activity that was included in our approach. Recently, Kusmakar et al. (2017) reported on the successful use of Poincaré plots in seizure detection through accelerometry, but their successful use on LEM recordings was yet to be demonstrated. Sustained variations in the local growth of the time series are often seen as a hallmark of ictal activity, mostly in focal seizures, thereby providing a plausible explanation for the remarkable performance of features based on Poincaré maps in our method.

Overall, this array of dynamic features offered a rich representation of brain activity during seizures, on which the RUSBoost algorithm was successful in performing classification. Although previous studies have documented the successful use of black-box classifiers, such as support vector machine (SVM) or K-nearest neighbors (KNN) (Baumgartner and Koren, 2018), the unique nature of LEM recordings called for an alternative approach. Not only are these EEG epochs high dimensional but also they are characterized by an imbalanced distribution, with very few windows containing seizures and the vast majority pertaining to non-ictal activity. The RUSBoost algorithm is a hybrid approach that uses a combination of sampling and boosting, whereby it performs random undersampling of the majority class before building an ensemble of classifiers (Seiffert et al., 2009). RUSBoost algorithm has been successfully used for sleep apnea detection in polysomnography (Veauthier et al., 2019), and recent studies offer further backing to its use in seizure detection, through direct comparison with black-box classifiers (SVM or KNN) (Solaija et al., 2018).

### Limitations of the study

Our study has several limitations. First, we analyzed a small sample of patients, as the study was designed in a proof-of-concept fashion. Some measures and methods implemented in our algorithm had never been used in seizure detection, thereby calling for further testing before transition to a larger dataset—our next step. Although working with a small group of patients, we avoided overfitting by utilizing 5-fold cross-validation, a well-known method that overcomes small dataset limitations and provides accurate estimations regarding

performance (Abbasi and Goldenholz, 2019). Second, we only included focal seizures in training and testing. Adding generalized seizures in such a small database would have confounded the detection algorithm instead of reaching a satisfying training. Now that the algorithm performance has been validated in the study of focal seizures, we intend to broaden the spectrum to other types of ictal events, as well as a larger range of epileptic syndromes. Finally, although a single-channel-based approach similar to ours is widely accepted (Baumgartner and Koren 2018), it is prone to some unavoidable artifacts, especially in a common average montage, such as electrode artifacts. This class of artifacts was entirely responsible for the very small number of false alarms in our study; we believe that the combination of different montages would resolve this issue.

In forthcoming efforts, we intend to generate a universal multi-classifier to distinguish between focal and generalized seizures (developing phase). By adding interactions between channels to the analysis and including different montages, we could tackle propagation and synchronization in seizure evolution. By expanding on our  $\epsilon$ -symbolic approach, we foresee the possibility of creating global recurrence network of the epileptogenic zone, which could empower automated diagnosis and help elucidate how epileptic brain works.

### Resource availability

#### Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Irene Villegas ([irene.villegas@carm.es](mailto:irene.villegas@carm.es)).

#### Materials availability

This study did not generate new materials associated to this paper.

#### Data and code availability

The automated detection algorithm is available for download at <https://github.com/ManuelRuizMarin/Classification-Algorithm>. Datasets are stored at the I.V.M.'s home institution and could be provided upon request.

## METHODS

All methods can be found in the accompanying [Transparent Methods supplemental file](#).

## SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2020.101997>.

## ACKNOWLEDGMENTS

This study is part of the collaborative activities carried out under the programs of the region of Murcia (Spain): "Groups of Excellence of the region of Murcia, the Fundación Séneca, Science and Technology Agency" project 19884/GERM/15 and "Call for Fellowships for Guest Researcher Stays at Universities and OPIS" project 21144/IV/19. M.P. would like to express his gratitude to the Technical University of Cartagena for hosting him during a Sabbatical leave and acknowledge partial support from the National Science Foundation under grant number ECCS 1928614. The authors are also grateful to the Regional Health-care Service of Murcia (SMS) and Drs. Diego Tortosa and María López for providing LEM recordings.

## AUTHORS CONTRIBUTION

Conceptualization, M.R.M., I.V.M., G.R.B., and M.P.; Methodology, M.R.M., I.V.M., and M.P.; Software, M.R.M.; Investigation and Data Curation, I.V.M.; Visualization, M.R.M., I.V.M., and G.R.B.; Writing—Original Draft, M.R.M., I.V.M., G.R.B., and M.P.; Writing—review and editing, I.V.M. and M.P.; Supervision, M.P.

## DECLARATION OF INTERESTS

The authors declare no competing interests. We confirm that we have read the Journal position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.

Received: June 26, 2020

Revised: October 23, 2020

Accepted: December 23, 2020

Published: January 22, 2021

## REFERENCES

Abbas, B., and Goldenholz, D.M. (2019). Machine learning applications in epilepsy. *Epilepsia* 60, 2037–2047.

Acharya, U.R., Sree, S.V., Chattopadhyay, S., Yu, W., and Ang, P.C. (2011). Application of recurrence quantification analysis for the automated identification of epileptic EEG signals. *Int. J. Neural Syst.* 21, 199–211.

Alotaiby, T.N., Alshebeili, S.A., Alshawi, T., Ahmad, I., and Abd El-Samie, F.E. (2014). EEG seizure detection and prediction algorithms: a survey. *EURASIP J. Adv. Signal. Process.* 2014, 183.

Andrzejak, R.G., Lehnertz, K., Mormann, F., Rieke, C., David, P., and Elger, C.E. (2001). Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. *Phys. Rev.* 64, 061907.

Ayodele, K., Ikezogwo, W., Komolafe, M., and Ogunbona, P. (2020). Supervised domain generalization for integration of disparate scalp EEG datasets for automatic epileptic seizure detection. *Comput. Biol. Med.* 120, 103757.

Baumgartner, C., and Koren, J.P. (2018). Seizure detection using scalp-EEG. *Epilepsia* 59, 14–22.

Bomela, W., Wang, S., Chou, C.-A., and Li, J.-S. (2020). Real-time inference and detection of disruptive EEG networks for epileptic seizures. *Sci. Rep.* 10, 1–10.

Caballero-Pintado, M.V., Matilla-García, M., and Ruiz Marín, M. (2018). Symbolic recurrence plots to analyze dynamical systems. *Chaos* 28, 063112.

Cohen, M.X. (2017). Where does EEG come from and what does it mean? *Trends Neurosci.* 40, 208–218.

de Cheveigné, A., and Nelken, I. (2019). Filters: when, why, and how (not) to use them. *Neuron* 102, 280–293.

Devinsky, O., Vezzani, A., O'Brien, T.J., Jette, N., Scheffer, I.E., de Curtis, M., and Perucca, P. (2018). Epilepsy. *Nat. Rev. Dis. Primers* 4, 18024.

DiCarlo, J.J., Zoccolan, D., and Rust, N.C. (2012). How does the brain solve visual object recognition? *Neuron* 73, 415–434.

Diykh, M., Li, Y., and Wen, P. (2017). Classify epileptic EEG signals using weighted complex networks based community structure detection. *Expert Syst. Appl.* 90, 87–100.

Fergus, P., Hignett, D., Hussain, A., Al-Jumeily, D., and Abdel-Aziz, K. (2015). Automatic epileptic seizure detection using scalp EEG and advanced artificial intelligence techniques. *Biomed. Res. Int.* 2015, 986736.

Fisher, R.S., Acevedo, C., Arzimanoglou, A., Bogacz, A., Cross, J.H., Elger, C.E., Engel, J., Jr., Forsgren, L., French, J.A., and Glynn, M. (2014a). ILAE official report: a practical clinical definition of epilepsy. *Epilepsia* 55, 475–482.

Fisher, R.S., Scharfman, H.E., and DeCurtis, M. (2014b). How can we identify ictal and interictal abnormal activity? *Adv. Exp. Med. Biol.* 813, 3–23.

Fürbass, F., Ossenblok, P., Hartmann, M., Perko, H., Skupch, A.M., Lindinger, G., Elezi, L., Pataria, E., Colon, A.J., Baumgartner, C., and Kluge, T. (2015). Prospective multi-center study of an automatic online seizure detection system for epilepsy monitoring units. *Clin. Neurophysiol.* 126, 1124–1131.

Gabor, A.J., Leach, R.R., and Dowla, F.U. (1996). Automated seizure detection using a self-organizing neural network. *Electroencephalogr. Clin. Neurophysiol.* 99, 257–266.

Golmohammadi, M., Ziyabari, S., Shah, V., Von Weltin, E., Campbell, C., Obeid, I., and Picone, J. (2017). Gated recurrent networks for seizure detection. In 2017 IEEE Signal Processing in Medicine and Biology Symposium (SPMB) (IEEE), pp. 1–5.

González Otárla, K.A., Mikhæl-Demo, Y., Bachman, E.M., Balaguera, P., and Schuele, S. (2019). Automated seizure detection accuracy for ambulatory EEG recordings. *Neurology* 92, e1540–e1546.

Gotman, J. (1982). Automatic recognition of epileptic seizures in the EEG. *Electroencephalography Clin. Neurophysiol.* 54, 530–540.

Gotman, J. (1990). Automatic seizure detection: improvements and evaluation. *Electroencephalography Clin. Neurophysiol.* 76, 317–324.

Haider, H.A., Esteller, R., Hahn, C.D., Westover, M.B., Halford, J.J., Lee, J.W., Shafi, M.M., Gaspard, N., Herman, S.T., and Gerard, E.E. (2016). Sensitivity of quantitative EEG for seizure identification in the intensive care unit. *Neurology* 87, 935–944.

Hartmann, M.M., Fürbaß, F., Perko, H., Skupch, A., Lackmayer, K., Baumgartner, C., and Kluge, T. (2011). EpiScan: online seizure detection for epilepsy monitoring units. In 2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (IEEE), pp. 6096–6099.

Hopfengärtner, R., Kasper, B.S., Graf, W., Gollwitzer, S., Kreiselmeyer, G., Stefan, H., and Hamer, H. (2014). Automatic seizure detection in long-term scalp EEG using an adaptive thresholding technique: a validation study for clinical routine. *Clin. Neurophysiol.* 125, 1346–1352.

Hopfengärtner, R., Kerling, F., Bauer, V., and Stefan, H. (2007). An efficient, robust and fast method for the offline detection of epileptic seizures in long-term scalp EEG recordings. *Clin. Neurophysiol.* 118, 2332–2343.

Hunyadi, B., Signoretto, M., Van Paesschen, W., Suykens, J.A., Van Huffel, S., and De Vos, M. (2012). Incorporating structural information from the multichannel EEG improves patient-specific seizure detection. *Clin. Neurophysiol.* 123, 2352–2361.

Islam, M.K., Rastegarnia, A., and Yang, Z. (2016). Methods for artifact detection and removal from scalp EEG: a review. *Neurophysiol. Clin.* 46, 287–305.

Kamitaki, B.K., Yum, A., Lee, J., Rishty, S., Sivarapalan, K., Esfahanizadeh, A., Mani, R., and Wong, S. (2019). Yield of conventional and automated seizure detection methods in the epilepsy monitoring unit. *Seizure* 69, 290–295.

Kannathal, N., Acharya, U.R., Lim, C.M., and Sadasivan, P.K. (2005). Characterization of EEG—a comparative study. *Comput. Methods Programs Biomed.* 80, 17–23.

Kelly, K.M., Shiao, D.S., Kern, R.T., Chien, J.H., Yang, M.C., Yandora, K.A., Valeriano, J.P., Halford, J.J., and Sackellares, J.C. (2010). Assessment of a scalp EEG-based automated seizure detection system. *Clin. Neurophysiol.* 121, 1832–1843.

Kusmakar, S., Karmakar, C.K., Yan, B., O'Brien, T.J., Muthuganapathy, R., and Palaniswami, M. (2017). Detection of generalized tonic-clonic seizures using short length accelerometry signal. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2017, 4566–4569.

Kwan, P., Arzimanoglou, A., Berg, A.T., Brodie, M.J., Allen Hauser, W., Mathern, G., Moshé, S.L., Perucca, E., Wiebe, S., and French, J. (2010). Definition of drug resistant epilepsy: consensus proposal by the ad hoc Task Force of the ILAE Commission on Therapeutic Strategies. *Epilepsia* 51, 1069–1077.

Li, P., Karmakar, C., Yearwood, J., Venkatesh, S., Palaniswami, M., and Liu, C. (2018). Detection of epileptic seizure based on entropy analysis of short-term EEG. *PLoS One* 13, e0193691.

Litt, B., and Echauz, J. (2002). Prediction of epileptic seizures. *Lancet Neurol.* 1, 22–30.

Meier, R., Dittrich, H., Schulze-Bonhage, A., and Aertsen, A. (2008). Detecting epileptic seizures in long-term human EEG: a new approach to automatic online and real-time detection and classification of polymorphic seizure patterns. *J. Clin. Neurophysiol.* 25, 119–131.

Obeid, I., and Picone, J. (2018). The Temple university hospital EEG data Corpus.

Augmentation of brain function: facts, fiction and controversy. *Brain-Machine Interfaces* 1, 394–398.

Perucca, P., Dubeau, F., and Gotman, J. (2014). Intracranial electroencephalographic seizure-onset patterns: effect of underlying pathology. *Brain* 137, 183–196.

Petras, K., Ten Oever, S., Jacobs, C., and Goffaux, V. (2019). Coarse-to-fine information integration in human vision. *Neuroimage* 186, 103–112.

Porfiri, M., and Ruiz Marín, M. (2019). Transfer entropy on symbolic recurrences. *Chaos* 29, 063123.

Raghu, S., Sriram, N., Rao, S.V., Hegde, A.S., and Kubben, P.L. (2020). Automated detection of epileptic seizures using successive decomposition index and support vector machine classifier in long-term EEG. *Neural Comput. Applic.* 32, 8965–8984.

Rao, V.R., and Lowenstein, D.H. (2015). Epilepsy. *Curr. Biol.* 25, R742–R746.

Rommens, N., Geertsema, E., Jansen Holleboom, L., Cox, F., and Visser, G. (2018). Improving staff response to seizures on the epilepsy monitoring unit with online EEG seizure detection algorithms. *Epilepsy Behav.* 84, 99–104.

Ryvlin, P., Cross, J.H., and Rheims, S. (2014). Epilepsy surgery in children and adults. *Lancet Neurol.* 13, 1114–1126.

Saini, J., and Dutta, M. (2017). An extensive review on development of EEG-based computer-aided diagnosis systems for epilepsy detection. *Network* 28, 1–27.

Samiee, K., Kovács, P., and Gabbouj, M. (2015). Epileptic seizure classification of EEG time-series

using rational discrete short-time fourier transform. *IEEE Trans. Biomed. Eng.* 62, 541–552.

Schindler, K., Wiest, R., Kollar, M., and Donati, F. (2001). Using simulated neuronal cell models for detection of epileptic seizures in foramen ovale and scalp EEG. *Clin. Neurophysiol.* 112, 1006–1017.

Seiffert, C., Khoshgoftaar, T.M., Van Hulse, J., and Napolitano, A. (2009). RUSBoost: a hybrid approach to alleviating class imbalance. *IEEE Trans. Syst. Man Cybernetics A Syst. Humans* 40, 185–197.

Shah, V., von Weltin, E., Lopez, S., McHugh, J.R., Veloso, L., Golmohammadi, M., Obeid, I., and Picone, J. (2018). The Temple university hospital seizure detection Corpus. *Front. Neuroinform* 12, 83.

Sharmila, A., and Geethanjali, P. (2019). A review on the pattern detection methods for epilepsy seizure detection from EEG signals. *Biomed. Tech. (Berl)* 64, 507–517.

Sheth, R.D. (2019). Patterns specific to pediatric EEG. *J. Clin. Neurophysiol.* 36, 289–293.

Sisodiya, S. (2007). Etiology and management of refractory epilepsies. *Nat. Clin. Pract. Neurol.* 3, 320–330.

Solaija, M.S.J., Saleem, S., Khurshid, K., Hassan, S.A., and Kamboh, A.M. (2018). Dynamic mode decomposition based epileptic seizure detection from scalp EEG. *IEEE Access* 6, 38683–38692.

Tiwari, A.K., Pachori, R.B., Kanhangad, V., and Panigrahi, B.K. (2017). Automated diagnosis of epilepsy using key-point-based local binary pattern of EEG signals. *IEEE J. Biomed. Health Inform.* 21, 888–896.

Tsiouris, K.M., Konitsiotis, S., Markoula, S., Koutsouris, D.D., and Fotiadis, D.I. (2018). An unsupervised methodology for the detection of epileptic seizures using EEG signals: a multi-dataset evaluation. *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* 2018, 3390–3393.

Urigüen, J.A., and Garcia-Zapirain, B. (2015). EEG artifact removal-state-of-the-art and guidelines. *J. Neural Eng.* 12, 031001.

Veauthier, C., Ryczewski, J., Mansow-Model, S., Otte, K., Kayser, B., Glos, M., Schöbel, C., Paul, F., Brandt, A.U., and Penzel, T. (2019). Contactless recording of sleep apnea and periodic leg movements by nocturnal 3-D-video and subsequent visual perceptive computing. *Sci. Rep.* 9, 16812.

Velis, D., Plouin, P., Gotman, J., da Silva, F.L., and Neurophysiology, I.D.S.o. (2007). Recommendations regarding the requirements and applications for long-term recordings in epilepsy. *Epilepsia* 48, 379–384.

Wang, Y., Zhou, W., Yuan, Q., Li, X., Meng, Q., Zhao, X., and Wang, J. (2013). Comparison of ictal and interictal EEG signals using fractal features. *Int. J. Neural Syst.* 23, 1350028.

Wilson, S.B., Scheuer, M.L., Emerson, R.G., and Gabor, A.J. (2004). Seizure detection: evaluation of the Reveal algorithm. *Clin. Neurophysiol.* 115, 2280–2291.

Wilson, S.B., Scheuer, M.L., Plummer, C., Young, B., and Pacia, S. (2003). Seizure detection: correlation of human experts. *Clin. Neurophysiol.* 114, 2156–2164.

## **Supplemental Information**

# **Integrating old and new complexity measures toward automated seizure detection from long-term video EEG recordings**

**Manuel Ruiz Marín, Irene Villegas Martínez, Germán Rodríguez Bermúdez, and Maurizio Porfiri**

## SUPPLEMENTAL INFORMATION

### Transparent Methods

#### Dataset

The main database used in this study was obtained from the Epilepsy Unit in Hospital Clínico Universitario Virgen de la Arrixaca (Murcia, Spain), after approval from the Hospital's Ethics Committee. Adult patients with refractory focal epilepsy, who were admitted to the Epilepsy Unit for non-invasive LEM as part of their presurgical evaluation, were included if they suffered from at least one epileptic seizure with corresponding EEG ictal patterns during 24-hour monitoring.

Scalp LEM was recorded at a sampling rate of 256 Hz, with 19 electrodes placed according to the international 10-20 system, using a 64-channel system Nicolet™EEG NicOne. EEG was recorded from the following electrode positions: Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, and O2 and a reference electrode (Z). Two certified experienced epileptologists, with the same display setting, visually labeled the epileptic seizures in off-line LEM recordings, following the onset classification based on the work by Perucca *et al.* (2014). Ictal activities with a minimum duration of ten seconds and a maximum of any period were considered for the study. All raw LEM datasets in common average montage were converted to EDF format and analyzed without any filter settings, except for notch default filter. No epochs were excluded due to artifacts (movement, chewing, EMG, eye-blinks, deteriorated electrodes, etc.). Thus, a total of 241.5 hours of scalp LEM recordings was analyzed, including 24 identified seizures (Table 1).

In addition, we benchmarked our algorithm on two publicly available datasets (the UB dataset and TUSZ). The UB dataset was recorded from five different subjects undergoing pre-surgical evaluation (Andrzejak *et al.*, 2001). Recordings are divided into five subsets of 100 single-channel EEG segments of 23.6 s duration recorded at a sampling rate of 173.61 Hz. Subsets O and Z contain recordings obtained through external surface electrodes from healthy subjects and subsets N, F, and S were recorded intracranially from the epileptic patients. N and F are interictal recordings whereas subset S correspond to seizure activity. In our study, subsets O, Z, N, and F were considered as non-seizure and subset S as seizure activity.

The TUSZ is the largest open-source corpus of seizures and represents an accurate characterization of clinical conditions (Shah *et al.*, 2018; Obeid and Picone, 2018). EEG signals were recorded using the International 10-20 system electrode

placement at a sampling rate of 256 Hz. These EEG recordings were obtained from the following electrode placements: Fp1, Fp2, F7, F3, Fz, F4, F8, T1, T3, C3, Cz, C4, T2, T4, T5, P3, Pz, P4, T6, O1, and O2, in average reference montage (AR). In order to guarantee comparability with our own dataset, we selected all of the EEG recordings from version 1.5.2 (both development and training datasets) that matched the following criteria: long term monitoring labeling (LTM), Epilepsy Unit origin (EMU) and at least one seizure reported. A total of 20 patients fulfilled those criteria. From those, we excluded recordings that were diagnosed of status epilepticus, for clinical reasons, and recordings from patient 88 because they were unreadable on average montage. Therefore, 27 recordings from 13 patients were finally selected for analysis, containing a total of 77 seizures and 33.15 hours. No filtering or pre-processing was performed (not even a notch filter).

### ***Detection algorithm***

Our approach unfolds along the following steps (Figure 1):

1. Manually analyze with clinical supervision a given fraction of the available LEM recordings to determine the onset and termination of seizures.
2. Partition each LEM recording into  $w$  non-overlapping windows of length  $n_w$ .
3. Extract  $f$  a priori-selected features from each window to encapsulate brain activity and facilitate detection of seizures, such as standard deviation and fractal dimension.
4. Train an algorithm of choice for binary classification on the  $f$  features of the manually analyzed recordings.
5. Test the trained model on the remaining fraction of the available LEM recordings, without any supervision.

EEG is a non-stationary signal, which is typically segmented into shorter stationary epochs (Wilson *et al.*, 2004; Hopfengärtner *et al.*, 2007). The number of these non-overlapping windows is controlled by the length and resolution of the time-series. In our study, we use  $n_w = 100$ , which corresponds to approximately 0.391 s, given the sampling frequency of 256 Hz. The fraction of windows that should be manually classified to train the classification algorithm depends on practical considerations. Training on a small dataset may not be sufficient to identify model parameters, but training on an excessively large dataset could produce overfitting and hinder the predictive power of the model on new observation. Here, we opt for a  $K$ -folds cross-validation with  $K = 5$  to evaluate the model performance.

Step 3 of the detection algorithm contains the main element of novelty of our

approach, which combines descriptive and complexity measures of time-series. Specifically, given the time-series of EEG recordings in a given window  $x = \{x_1, x_2, \dots, x_{n_w}\}$ , we compute few classical statistical measures (D: descriptive) and a range of features that capture fractal dimensions, local growth, and recurrent dynamics (C: complexity).

While all the descriptive measures are computed on the raw time-series, some of the complexity measures addressing recurrent dynamics are based on a new scheme to perform recurrence quantification. This scheme integrates traditional recurrence quantification methods with recent advancements in symbolic dynamics (Caballero-Pintado *et al.*, 2018), whose application to the detection of atrial fibrillation and the study of behavioral patterns in humans and animals has been demonstrated by Pérez-Valero *et al.* (2019) and Boldini *et al.* (2019), respectively. The uniqueness of the application presented herein entails the study of multivariate time-series.

Classical recurrence is defined in terms of  $m$ -histories (Takens, 1981) that encapsulate local properties of the time-evolution of a times-series, such as their value and slope. These  $m$ -histories take the form  $\tilde{x}_t = (x_t, x_{t+1}, \dots, x_{t+m-1})$ , where  $m$  is the embedding dimension and  $t = 1, \dots, n_w - m + 1$ . As originally stated by Eckmann (1987), for a positive real number  $\varepsilon$ , two  $m$ -histories  $\tilde{x}_t$  and  $\tilde{x}_s$  are said to be  $\varepsilon$ -recurrent when  $\|\tilde{x}_t - \tilde{x}_s\| < \varepsilon$ , where  $\varepsilon$  (known as proximity parameter) is a threshold distance and  $\|\cdot\|$  is the Euclidean norm.

This definition can be extended to symbolic representations through ordinal patterns as illustrated by Caballero-Pintado *et al.* (2018). Ordinal patterns offer a potent coarse-grain tool to study dynamical systems. For example, the seminal work by Bandt and Pompe (2002) has demonstrated the rigorous use of Shannon entropy of ordinal patterns to probe the complexity of dynamical systems. Although the use of ordinal patterns is responsible for obvious information losses (such as the numeric values of the time-series), it offers a range of practical advantages that often offset this limitation. For instance, ordinal representations are more robust to the presence of noise and can naturally capture complexities and dynamical changes in time-series (Zanin *et al.*, 2012; Keller *et al.*, 2014).

Formally, the ordinal pattern associated with  $\tilde{x}_t$  is the permutation  $(i_1, i_2, \dots, i_m)$  satisfying the following conditions:

- (a)  $x_{t+i_1} \leq x_{t+i_2} \leq \dots \leq x_{t+i_m}$ , and
- (b)  $i_{s-1} < i_s$  if  $x_{t+i_{s-1}} = x_{t+i_s}$

Condition (b) guarantees the uniqueness of the permutation. In a symbolic sense, two  $m$ -histories  $\tilde{x}_t$  and  $\tilde{x}_s$  are said to be symbolic recurrent if both have the same ordinal

pattern. For example, if  $m = 2$ , we would require that the slope of the time-series be the same, either positive or negative.

In general,  $m$ -histories that are recurrent in the traditional sense may not be recurrent in a symbolic sense, and vice versa. By merging these two notions of recurrence, we identify  $\varepsilon$ -symbolic recurrent  $m$ -histories, as the subset of  $m$ -histories that are recurrent in both the classical and the symbolic sense. The combination of classical and symbolic recurrence is expected to beget a viable strategy to mitigate measurement noise that is known to plague classical recurrence analysis, while avoiding shortcuts that are associated with the limited alphabet of a symbolic representation (Caballero-Pintado *et al.*, 2018; Porfiri *et al.*, 2019). In our analysis, we use  $m = 3$  to create a rich and reliable symbolic representation of the time-series, from the  $n_w = 100$  instants of each window. The value of  $\varepsilon$  is chosen to be  $10 \mu\text{V}$ , which is on the order of magnitude of normal brain activity during wakefulness (Tatum *et al.*, 2018; Biasiucci *et al.*, 2019).

Below, we report the  $f = 8$  features we examine.

D1. Standard deviation:

$$\sigma = \frac{1}{\sqrt{n_w - 1}} \sqrt{\sum_{t=1}^{n_w} (x_t - \bar{x})^2} \quad (2)$$

where  $\bar{x}$  is the mean.

D2. Mean absolute deviation with respect to the median:

$$D_{\text{Me}} = \frac{1}{n_w} \sum_{t=1}^{n_w} |x_t - \text{Me}| \quad (3)$$

where  $\text{Me}$  is the median. This quantity differs from the standard deviation, whereby it ascertains the dispersion of the time-series with respect to the median, rather than the mean.

D3. Skewness coefficient:

$$\text{Skw} = \frac{1}{\sigma^3} \sum_{t=1}^{n_w} (x_t - \bar{x})^3 \quad (4)$$

This quantity provides an estimate of the symmetry of the time-series about its mean.

C1. Katz fractal dimension:

$$\text{KFD} = \frac{\log(n_w - 1)}{\log(n_w - 1) + \log(d/L)} \quad (5)$$

Here,  $d$  is the diameter of the curve in the  $t - x$  plane associated with the time-series, estimated as the Euclidean distance between the first and the farthest

point of the sequence

$$d = \max \left\{ \sqrt{(1-t)^2 + (x_1 - x_t)^2} \right\}_{t=1}^{n_w} \quad (6)$$

and  $L$  is the total length of the curve, estimated from the sum of the distance between successive points

$$L = \sum_{t=1}^{n_w-1} \sqrt{1 + (x_t - x_{t+1})^2} \quad (7)$$

The fractal dimension is typically bounded between one and two, with one representing the least complex time-series (a straight line) and higher values being associated with more complex time-series.

C2. Area of the second-order difference plot (Pachori *et al.*, 2014). To calculate this feature, we create a Poincaré plot of the increments in the time-series

$$Y_t = x_{t+1} - x_t \quad (8)$$

$$Z_t = x_{t+2} - x_{t+1} \quad (9)$$

We fit the Poincaré plot with an ellipse, whose 95% estimated area  $A$  is the feature utilized to estimate the local growth, according to

$$A = 3\pi \sqrt{(S_Y^2 + S_Z^2)^2 - D^2} \quad (10)$$

with

$$S_Y = \frac{1}{\sqrt{n_w - 2}} \sqrt{\sum_{t=1}^{n_w-2} Y_t^2} \quad (11)$$

$$S_Z = \frac{1}{\sqrt{n_w - 2}} \sqrt{\sum_{t=1}^{n_w-2} Z_t^2} \quad (12)$$

$$S_{YZ} = \frac{1}{n_w - 2} \sum_{t=1}^{n_w-2} Y_t Z_t \quad (13)$$

$$D = \sqrt{S_Y^2 + S_Z^2 - 4(S_Y^2 S_Z^2 - S_{YZ}^2)} \quad (14)$$

The larger is the area of the ellipse the more is variation in the local growth of the time-series.

C3. Mean degree of the  $\varepsilon$ -symbolic recurrent network, which is constructed from the recurrence analysis. Grounded in the proposed notion of recurrence, we put forward a network visualization of recurrent behavior, wherein each node is associated with a time instant and two nodes are connected if the corresponding  $m$ -histories are  $\varepsilon$ -symbolic recurrent. The vertex set of the network is  $\mathcal{V} =$

$\{1, \dots, n_w - m + 1\}$  and  $(t, s)$  is an element of the edge set  $\mathcal{E}$  if  $\tilde{x}_t$  and  $\tilde{x}_s$  are  $\varepsilon$ -symbolic recurrent. This definition extends the analysis to combined  $\varepsilon$ -symbolic recurrence (Zou *et al.*, 2019). The mean degree is indicative of the number of  $m$ -histories that are  $\varepsilon$ -symbolic recurrent.

C4. Mean betweenness centrality of the  $\varepsilon$ -symbolic recurrent network:

$$\overline{\text{bet}} = \frac{1}{2|\mathcal{V}|} \sum_{u=1}^{|\mathcal{V}|} \sum_{t,s=1}^{|\mathcal{V}|} \frac{n_{st}(u)}{N_{st}} \quad (15)$$

where  $n_{st}(u)$  is the number of shortest paths from  $s$  to  $t$  that pass through node  $u$  and  $N_{st}$  is the total number of shortest paths from  $s$  to  $t$ . This quantity scores the local connectivity in the network, which is a measure of  $\varepsilon$ -proximity of the  $m$ -histories that have the same ordinal pattern.

C5. Mean closeness of the  $\varepsilon$ -symbolic recurrent network:

$$\overline{\text{clo}} = \frac{1}{|\mathcal{V}|(|\mathcal{V}| - 1)^2} \sum_{t=1}^{|\mathcal{V}|} \frac{M_t^2}{C_t} \quad (16)$$

where  $M_t$  is the number of reachable nodes from  $\tilde{x}_t$  (not counting itself), and  $C_t$  is the sum of distances from node  $\tilde{x}_t$  to all reachable nodes. This quantity offers a further measure of centrality in the network.

Upon scoring features for all the windows in the dataset, we must train an algorithm for classification on the manually analyzed data. We implement the RUSBoost algorithm (Seiffert *et al.*, 2009), which is a hybrid sampling/boosting algorithm with decision trees as learners. This algorithm is optimal to solve classification problems, where one class in a training dataset vastly outnumbers the other class, as in the case of epileptic seizure classification. After training, the detection algorithm works directly on LEM recordings, without any clinical input.

## Supplemental references

Andrzejak, R. G., Lehnertz, K., Mormann, F., Rieke, C., David, P. and Elger, C. E. (2001) 'Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state', *Physical Review E*, 64(6), pp. 061907.

Bandt, C. and Pompe, B. (2002) 'Permutation entropy: a natural complexity measure for time series', *Physical review letters*, 88(17), pp. 174102.

Biasiucci, A., Franceschiello, B. and Murray, M. M. (2019) 'Electroencephalography', Current Biology, 29(3), pp. R80-R85.

Boldini, A., Karakaya, M., Ruiz Marín, M. and Porfiri, M. (2019) 'Application of symbolic recurrence to experimental data, from firearm prevalence to fish swimming', Chaos: An Interdisciplinary Journal of Nonlinear Science, 29(11), pp. 113128.

Caballero-Pintado, M. V., Matilla-García, M. and Ruiz Marín, M. (2018) 'Symbolic recurrence plots to analyze dynamical systems', Chaos: An Interdisciplinary Journal of Nonlinear Science, 28(6), pp. 063112.

Eckmann, J. P. (1987). 'Recurrence plots of dynamical systems', *Europhys Lett* 5, pp. 973-977.

Hopfengärtner, R., Kerling, F., Bauer, V. and Stefan, H. (2007) 'An efficient, robust and fast method for the offline detection of epileptic seizures in long-term scalp EEG recordings', *Clin Neurophysiol*, 118(11), pp. 2332-43.

Obeid, I., and Picone, J. (2018). The Temple University Hospital EEG Data Corpus. 'Augmentation of Brain Function: Facts, Fiction and Controversy', *Brain-Machine Interfaces* 1, pp. 394-398.

Pachori, R. B. and Patidar, S. (2014) 'Epileptic seizure classification in EEG signals using second-order difference plot of intrinsic mode functions', *Comput Methods Programs Biomed*, 113(2), pp. 494-502.

Pérez-Valero, J., Caballero Pintado, M. V., Melgarejo, F., García-Sánchez, A. J., García-Haro, J., García Córdoba, F., García Córdoba, J. A., Pinar, E., García Alberola, A., Matilla-García, M., Curtin, P., Arora, M. and Ruiz Marín, M. (2019) 'Symbolic Recurrence Analysis of RR Interval to Detect Atrial Fibrillation', *J Clin Med*, 8(11), 1840.

Perucca, P., Dubeau, F. and Gotman, J. (2014) 'Intracranial electroencephalographic seizure-onset patterns: effect of underlying pathology', *Brain*, 137(Pt 1), pp. 183-96.

Porfiri, M. and Ruiz Marín, M. (2019) 'Transfer entropy on symbolic recurrences', *Chaos*, 29(6), pp. 063123.

Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J. and Napolitano, A. (2009) 'RUSBoost: A hybrid approach to alleviating class imbalance', *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 40(1), pp. 185-197.

Takens, F. (1981) 'Detecting strange attractors in turbulence', *Dynamical systems and turbulence*, Warwick 1980: Springer, pp. 366-381.

Tatum, W. O., Rubboli, G., Kaplan, P. W., Mirsatari, S. M., Radhakrishnan, K., Gloss, D., Caboclo, L. O., Drislane, F. W., Koutroumanidis, M., Schomer, D. L., Kasteleijn-Nolst Trenite, D., Cook, M. and Beniczky, S. (2018) 'Clinical utility of EEG in diagnosing and monitoring epilepsy in adults', *Clin Neurophysiol*, 129(5), pp. 1056-1082.

Shah, V., von Weltin, E., Lopez, S., McHugh, J. R., Veloso, L., Golmohammadi, M., Obeid, I. and Picone, J. (2018) 'The Temple University Hospital Seizure Detection Corpus', *Front Neuroinform*, 12, pp. 83.

Wilson, S. B., Scheuer, M. L., Plummer, C., Young, B. and Pacia, S. (2003) 'Seizure detection: correlation of human experts', *Clin Neurophysiol*, 114(11), pp. 2156-64.

Zanin, M., Zunino, L., Rosso, O. A. and Papo, D. (2012) 'Permutation entropy and its main biomedical and econophysics applications: a review', *Entropy*, 14(8), pp. 1553-1577.

Zou, Y., Donner, R. V., Marwan, N., Donges, J. F. and Kurths, J. (2019) 'Complex network approaches to nonlinear time series analysis', *Physics Reports*, 787, pp. 1-97.