



**HAL**  
open science

## A quadratic linear-parabolic model-based EEG classification to detect epileptic seizures

Antonio Quintero Rincón, Carlos d’Giano, Hadj Batatia

► **To cite this version:**

Antonio Quintero Rincón, Carlos d’Giano, Hadj Batatia. A quadratic linear-parabolic model-based EEG classification to detect epileptic seizures. *The Journal of Biomedical Research*, 2020, 34 (3), pp.203-210. 10.7555/JBR.33.20190012 . hal-02875045

**HAL Id: hal-02875045**

**<https://hal.science/hal-02875045>**

Submitted on 19 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Open Archive Toulouse Archive Ouverte



OATAO is an open access repository that collects the work of Toulouse researchers and makes it freely available over the web where possible

This is a publisher's version published in: <https://oatao.univ-toulouse.fr/26050>

### Official URL:

<https://doi.org/10.7555/JBR.33.20190012>

### To cite this version:

Quintero Rincon, Antonio  and D'Giano, Carlos and Batatia, Hadj  *A quadratic linear-parabolic model-based EEG classification to detect epileptic seizures.* (2020) *The Journal of Biomedical Research*, 34 (3). 203-210. ISSN 1674-8301 .

Any correspondence concerning this service should be sent to the repository administrator: [tech-oatao@listes-diff.inp-toulouse.fr](mailto:tech-oatao@listes-diff.inp-toulouse.fr)



# A quadratic linear-parabolic model-based EEG classification to detect epileptic seizures

Antonio Quintero-Rincón<sup>1,2,✉</sup>, Carlos D'Giano<sup>1</sup>, Hadj Batatia<sup>2</sup>

<sup>1</sup>Epilepsy and Telemetry Integral Center, Foundation for the Fight against Pediatric Neurological Disease, Montañeses 2325, Buenos Aires C1428AOK, Argentina;

<sup>2</sup>Computer Science Research Institute of Toulouse–National Polytechnic Institute of Toulouse, University of Toulouse, Toulouse, Cedex 7 B.P. 7122-31071, France.

## Abstract

The two-point central difference is a common algorithm in biological signal processing and is particularly useful in analyzing physiological signals. In this paper, we develop a model-based classification method to detect epileptic seizures that relies on this algorithm to filter electroencephalogram (EEG) signals. The underlying idea was to design an EEG filter that enhances the waveform of epileptic signals. The filtered signal was fitted to a quadratic linear-parabolic model using the curve fitting technique. The model fitting was assessed using four statistical parameters, which were used as classification features with a random forest algorithm to discriminate seizure and non-seizure events. The proposed method was applied to 66 epochs from the Children Hospital Boston database. Results showed that the method achieved fast and accurate detection of epileptic seizures, with a 92% sensitivity, 96% specificity, and 94.1% accuracy.

**Keywords:** two-point central difference, curve fitting, parabolic curves, epilepsy, electroencephalogram, random forest

## Introduction

Epilepsy is a neurological disease that affects people of all ages. It is characterized by unpredictable seizures resulting from the hyperexcitability of neurons. The electroencephalogram (EEG) is the predominant modality to study cerebral activity. The analysis of EEG signals is nearly completely dependent on visual inspection by the physician to quantify or qualify the morphology of waves. Their goal is the identification and classification of abnormal patterns in order to provide aid for an

epilepsy diagnosis.

This study proposes a simple, fast, and adaptable method, implementable in real-time, to help physicians visually inspect EEG signals for epileptic seizure detection. The proposed method fits within the framework of model-based classification. It is based on the statistical parameters obtained from fitting a quadratic parabolic model to specifically filtered and transformed EEG signals. Precisely, the two-point central difference algorithm is used to build a filter for the EEG signals. The filtered signal is subsequently represented using a quadratic linear-parabolic model,

✉ Corresponding author: Antonio Quintero-Rincón, Epilepsy and Telemetry Integral Center, Fight Against Child Neurological Diseases Foundation, Montañeses 2325, Buenos Aires C1428AOK, Argentina. E-mail: [tonioquintero@ieee.org](mailto:tonioquintero@ieee.org).

Received 16 January 2019, Revised 12 April 2019, Accepted 03 June 2019, Epub 28 August 2019

CLC number: R742.1, Document code: A

The authors reported no conflict of interests.

This is an open access article under the Creative Commons Attribution (CC BY 4.0) license, which permits others to distribute, remix, adapt and build upon this work, for commercial use, provided the original work is properly cited.

using the curve fitting approach. Four statistical parameters are then calculated to characterize the model fitting. These parameters are considered as a feature-vector in the classification stage.

The literature abounds with research work dealing with the detection of epileptic seizure onset. Various methods and techniques have been proposed for this purpose. Many existing methods fit within the large framework of the feature-based machine learning approach.

The two-point central difference algorithm has proven useful in analyzing physiological signals, with various medical applications reported in the literature. A typical work has been reported by Frei *et al* to detect epileptiform discharges in muscles in EEG signals<sup>[1]</sup>. This algorithm is usually used to estimate the derivative of a function, giving an estimate valid only over a limited frequency range. Here, our idea was to use this algorithm to design a filter that enhances the frequency range of EEG signals corresponding to brain activity characteristic of epileptic seizures. To single out the related waveform, we transform the filtered signal into a quadratic form. A quadratic linear-parabolic model is then fitted to the transformed signal, using the curve fitting technique<sup>[2-3]</sup>. The fitting is assessed using four statistical parameters (weighted sum of squared residuals, R-square, adjusted R-square, and root mean squared error). In order to show that our quadratic linear-parabolic model is pertinent to characterize epileptic seizures, we develop a classification method where the model-fitting parameters are used as features. We show that these features are good biomarkers of epileptic seizures.

The curve fitting technique adopted here is widely used in signal processing in general, and biomedical applications in particular. It has also been applied a few times in the domain of EEG and epilepsy. Orhan *et al* used the polynomial curve fitting method to fit a probability density function (PDF) of EEG signals discretized using equal frequency binning. The method has been applied to epileptic seizure detection using the comparison of PDFs of seizures and non-seizures. Second-order Fourier curve fitting has been used to smooth EEG signals in a pre-processing stage of a model-based method to predict epileptic seizures<sup>[4]</sup>. A spline curve fitting to EEG data was proposed to detect phase cone patterns to characterize epileptic activity<sup>[5]</sup>. Linear curve fitting was applied to interictal heartbeat intervals in order to combine EEG and ECG signals with the purpose of characterizing modulation patterns in patients with drug-resistant epilepsy<sup>[6]</sup>. More recently, a first-order exponential function has been fit to voltage-gated sodium channel type 2 to model the conductance-voltage curves in

neonatal epileptic subjects<sup>[7]</sup>.

Supervised machine learning methods have been widely used to detect seizures in EEG, with support vector machine (SVM) and K-Nearest Neighbor (KNN) being the most popular. SVM classifiers have been used with various features extracted from spectral and entropy analysis<sup>[8]</sup>, matching pursuit algorithm<sup>[9]</sup>, tensor discriminant analysis<sup>[10]</sup>, multifractal detrended fluctuation analysis<sup>[11]</sup>, and cross-bispectrum analysis<sup>[12]</sup>. While KNN classifiers have been developed, among many others, with features from the fractal dimension<sup>[13]</sup>, and non-linear dimension reduction of frequency domain parameters<sup>[14]</sup>. In addition to these two popular approaches, Acharya *et al* have recently been the first to develop a seizure detection method using a convolutional neural network<sup>[15]</sup>.

In general, these methods have good performance thanks to the advanced signal processing techniques they rely on. However, they consequently have high computational costs<sup>[16]</sup>. In this work, we combine for the first time the two-point difference algorithm and the quadratic linear-parabolic model to design an original model-based statistical classification method to detect epileptic seizures in EEG. The proposed feature-vector is simple and fast to calculate from the quadratic model. Our classification method is based on bootstrap-aggregated (bagged) decision trees. This approach combines results from several decision trees to overcome the overfitting effect due to the variability in the training data<sup>[17]</sup>. Using the different subset of features (subspace sampling) to build the weak-tree classifiers reduces the training time<sup>[18]</sup>, and there are some recent works in seizure classification using EEG<sup>[19-22]</sup>.

## Materials and methods

### Data

For the experimentation purpose, we considered a dataset from the Children's Hospital Boston database<sup>[23-24]</sup> which consists of 22 EEG recordings from pediatric subjects with intractable seizures. All signals were sampled at 256 Hz with a 16-bit resolution by using the International 10-20 system. The set of recordings lasted on average 35 minutes for 30 subjects in total, 2 hours for 4 subjects, and 12 hours for 2 other subjects. Taken together the recordings account for 60 hours of EEG recordings and 139 seizures. No distinctions regarding the types of seizure onsets were considered. The data contains focal, lateral, and generalized seizure onsets. Furthermore, the recordings were made in a routine clinical environment, so non-seizure activity and

artifacts such as head/body movement, chewing, blinking, early stages of sleep, and electrode pops/movements were present in the data. In this work, we have used 66 epochs from 9 different subjects (*Table 1*). Each recording contains a seizure event, whose onset time has been labeled by an expert neurologist. The duration has an average of 79.4 seconds. Moreover, for each seizure segment, the neurologist also selected one adjacent non-seizure signal segment before and after the seizure, of the same length to represent healthy brain activity. In total, we considered 33 seizures and 33 non-seizure events. Non-seizure control events have been chosen just before seizures. This choice is justified for two reasons. First, we are interested in identifying seizure onset, therefore it is the most important to distinguish its signal from preceding normal activity. Second, the EEG activity following a seizure remains chaotic for a long time, and comparing it to seizure onset is not pertinent. The selected signals had the same montage by using 23 scalp EEG channels.

## Methodology

Sea  $X \in \mathbb{R}^{N \times M}$  denotes the matrix of  $M$  EEG signals  $x_m \in \mathbb{R}^{N \times 1}$ , measured simultaneously on different channels and at  $N$  discrete-time instants. The proposed methodology is composed of five stages. The first stage divides the original signal  $X$  into a set of non-overlapping 1-second segments using a rectangular sliding window  $\Omega = \Omega_0(\omega - (W - 1)/2)$  with  $0 \leq \omega \leq W - 1$ , so that  $X[n] = \Omega[n]X$ . In the second stage, the two-point central difference algorithm is used to estimate the coefficients  $b[k]$  of the filter for each 1-second segment  $X[n]$ . The signal is then convolved by the resulting filter  $\tilde{X}[n] = X[n] \times b[k]$ . The third stage fits a quadratic linear-parabolic model to the filtered signal and estimates the associated parameters. The fourth stage calculates four statistical parameters for each EEG segment, namely the weighted sum of squared residuals ( $\zeta$ ) the R-square value ( $\phi$ ), the adjusted R-square value ( $\sigma$ ) and the root means squared error ( $\psi$ ), to assess the quadratic model fitting. Finally, in stage five, the feature-vector  $\rho = [\zeta, \phi, \sigma, \psi]$  associated with each segment is classified using a random forest classifier to discriminate between seizure and non-seizure.

### Two-point central difference algorithm

The principle of the two-point central difference algorithm consists in subtracting non-adjacent, regularly spaced, and pairs of points. The objective is to extract a slope from  $X[n]$ :

$$X'[n] = \frac{X[n+L] - X[n-L]}{2LT_s} \quad (1)$$

where  $L$  is called the skip factor that defines the distance between points, and  $T_s$  is the sample interval scaling. Taking the Z-transform of both sides, we obtain

$$X'(z) = \frac{z^L - z^{-L}}{2LT_s} X(z) \quad (2)$$

The transfer function of (2) is given by

$$Y(z, L) = \frac{X'(z)}{X(z)} = \frac{z^L - z^{-L}}{2LT_s} \quad (3)$$

Thus, the frequency response is obtained by replacing  $z$  by  $e^{j\Omega}$

$$Y(e^{j\Omega}, L) = \frac{e^{jL\Omega} - e^{-jL\Omega}}{2LT_s} = \frac{j \sin(L\Omega)}{LT_s} \quad (4)$$

Using a symmetric FIR (finite impulse response) filter, the two-point central difference algorithm is based on an impulse function containing two coefficients of equal but opposite sign spaced  $L$  points apart with the following coefficients

$$b[k] = \begin{cases} \frac{1}{2LT_s}, & k = -L \\ -\frac{1}{2LT_s}, & k = +L \\ 0, & k \neq \pm L \end{cases} \quad (5)$$

Note that, FIR filters are free from stability problems and they cause no time delay and no phase distortion within the pass-band. A convolution operation is estimated between each 1-second segment  $X[n]$  and the coefficients  $b[k]$  to obtain a new filtered signal in a bandwidth frequency until 50 Hz:

$$\tilde{X}[n] = X[n] \times b[k] \quad (6)$$

Note that, this is the effective bandwidth of the filter where the EEG activity has an important clinical relevance<sup>[25]</sup>. This allows us to automatically reject the high line-noise artifacts greater than 50 Hz as  $L = F_s/5$ , as the sampling rate  $F_s$  is the 256 Hz. Note that  $L$  is rounded. A comprehensive treatment of the mathematical properties of the two-point central difference algorithm is referenced<sup>[26-28]</sup>.

### Quadratic linear-parabolic model

We propose to fit a quadratic linear-parabolic model to the filtered signal. Precisely, the model is fitted to the couple  $(\tilde{X}, \tilde{X}^2)$ , where  $\tilde{X}$  is the complete filtered signal, obtained by concatenating the segments given in equation (6).

The idea is to fit the signal with a curve of the form:

$$y = a \sin(x - \pi) + b(x - 10)^2 + c \quad (7)$$

The fitting is performed using the least-squares method to estimate the three parameters  $a$ ,  $b$ ,  $c$ <sup>[29-31]</sup>.

*Model-fitting statistical parameters*

To assess the fitting of the curve given in equation (7), the following four statistical parameters are estimated. Note that  $y$  is the observed data,  $\hat{y}$  is the predicted value using the quadratic linear-parabolic model, and  $\bar{y}$  is the mean of the observed data.

Weighted sum of squared residuals ( $\zeta$ ): It is used to measure the total deviation of the response values from the predicted values, and is defined as

$$\zeta = \sum_{i=1}^n \omega_i (y_i - \hat{y}_i)^2 \tag{8}$$

The weights allow taking into consideration the different uncertainties of the measurements and are calculated as follows

$$\omega_i = 1/\sigma_i^2 \tag{9}$$

where  $\sigma^2$  is the variance.

R-square value ( $\phi$ ): It is the square of the correlation between the response values and the predicted response values, and is defined as

$$\phi = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{10}$$

Adjusted R-square value ( $\sigma$ ): It adjusts the R-square residual degrees of freedom, and is defined as

$$\sigma = 1 - \frac{(1 - \phi)(n - 1)}{n - m - 1} \tag{11}$$

where  $n$  is the number of response values and  $n$  is the estimated fitted coefficients from the response values. A value of  $\sigma$  closer to 1 indicates a better fit.

Root mean squared error ( $\psi$ ): It is an estimate of the standard deviation of the random component in the data and is defined as

$$\psi = \sqrt{\frac{\zeta}{n - m}} \tag{12}$$

*Random forest classification*

In this section, we present a classification method to identify seizures. The four statistical parameters presented above are used as classification features. A random forest classification technique is adopted. Random forest is an ensemble learning technique that combines the Bagging algorithm and the random subspace method using decision trees as the base classifier. The goal is to discriminate between seizure and non-seizure events, specifically the seizure onset through the feature predictor vector  $\rho = [\zeta, \phi, \sigma, \psi]$  associated with each EEG segment  $\tilde{X}$ . The idea is to design a prediction function  $f(\rho)$  to estimate the binary response for two classes,  $\Gamma = \{\gamma_1, \gamma_0\}$ , where  $\gamma_1 = 1$  for a seizure event and  $\gamma_0 = 0$  for a non-seizure event. A loss function  $L(\Gamma, f(\rho))$  determines the prediction function as follows:

$$E_{\rho, \Gamma} \{L[\Gamma, f(\rho)]\} \tag{13}$$

where  $E_{\rho, \Gamma}$  denotes the expectation with respect to an unknown joint distribution  $P_{\rho, \Gamma}(\rho, \Gamma)$  of  $\rho$  and  $\Gamma$ . For the classification, the next zero-one loss function is used

$$L[\Gamma, f(\rho)] = I[\Gamma \neq f(\rho)] = \begin{cases} 1 & \text{if } \Gamma = f(\rho) \text{ for } \gamma_1 \text{ or seizure,} \\ 0 & \text{if } \Gamma \neq f(\rho) \text{ for } \gamma_0 \text{ or non-seizure.} \end{cases} \tag{14}$$

The ensemble learning approach constructs  $f$  by using  $J$  decision-tree base-classifiers

$$\{h_1(\rho, \Theta_1), \dots, h_J(\rho, \Theta_J)\} \tag{15}$$

where  $\Theta_j, j=1, \dots, J$ , is an independent collection of random variables. These base classifiers are combined by majority vote, so that

$$f(\rho) = \operatorname{argmax}_{\gamma \in \Lambda} \sum_{j=1}^J I[\gamma = h_j(\rho)] \tag{16}$$

Consider a dataset  $\mathcal{D} = \{d_1, d_2, \dots, d_N\}$  with  $d_i = (\rho_i, c_i)$ , where  $c_i$  is a class label,  $\rho$  denotes the vector predictor and  $\gamma_i$  the response. For a particular realization  $\theta_j$  of  $\Theta_j$ , the fitted tree is denoted  $\hat{h}_j(\rho, \theta_j, \mathcal{D})$ . The random forest classifier uses random subspace in two ways. First, each tree is fitted to an independent bootstrap sample from the original data. The randomization involved in bootstrap sampling gives one part of  $\Theta_j$ . Second, the best split is retained for all  $\rho$  predictors according to the randomly selected subset of  $m$  predictor variables.

The randomization used to sample the predictors gives the remaining part of  $\Theta_j$ . The resulting class predicted is by the majority combination of unweighted voting of the trees (**Algorithm 1**). A comprehensive treatment of the properties of random forest classifier are referenced<sup>[17–18,32]</sup>.

**Results**

In this section, we evaluated the proposed methodology using the Children Hospital Boston database, which presented in the Material and methods section.

**Model fitting**

This section presents the results of the quadratic linear-parabolic model fitting. **Table 2** presents the three coefficients ( $a, b, c$ ) estimated according to equation (7). Coefficients for seizures and non-seizures are shown separately, for illustration purposes only. This allows us to note the significant difference between values corresponding to the two classes. The actual quadratic linear-parabolic model is fitted to the entire signal, without distinction of seizure and non-seizure.

<b>Algorithm 1 Random forest classifier algorithm</b>	
<b>Data:</b>	data set $\mathcal{D}$ and predictor vector $\rho$ .
<b>Result:</b>	ensemble of tree models whose predictions are to be combined by voting or averaging.
<b>for</b> $j=1$ to $J$ <b>do</b>	
1.	Take a bootstrap sample $\mathcal{D}_j$ of size $N$ from $\mathcal{D}$ .
2.	Using the bootstrap sample $\mathcal{D}_j$ as the training data, fit a tree using binary recursive partitioning:
a.	Start with all observations in a single node;
b.	Repeat the following steps recursively for each unsplit node until the stopping criterion is met:
i.	Select $m$ predictors at random from the $\rho$ available predictors;
ii.	Find the best binary split among all binary splits on the $m$ predictors from Step i;
iii.	Split the node into two descendant nodes using the split from Step ii;
<b>End</b>	
To make a prediction at a new point, use equation (16).	

*Fig. 1* shows the quadratic linear-parabolic curves obtained for seizure and non-seizure segments, respectively. One notices that the proposed model (7) provides a visually precise fit to the EEG waveform. In addition, the prediction bounds have a small uncertainty throughout the entire data range, hence new observations should be predicted with high accuracy.

**Classification**

The dataset described above has been used to train the classifier according to the proposed method. The capacity of the proposed classification scheme to discriminate between seizure and non-seizure events in order to detect the seizure onset in EEG signals has been assessed. In *Fig. 2*, we can observe the great difficulty in discriminating between seizure and non-seizure in EEG raw data. The start and end of the seizure in this EEG signal were labeled by the neurologist using two lines. The first line divides the EEG signal at 81 second (onset) and the second at 162 second (offset).

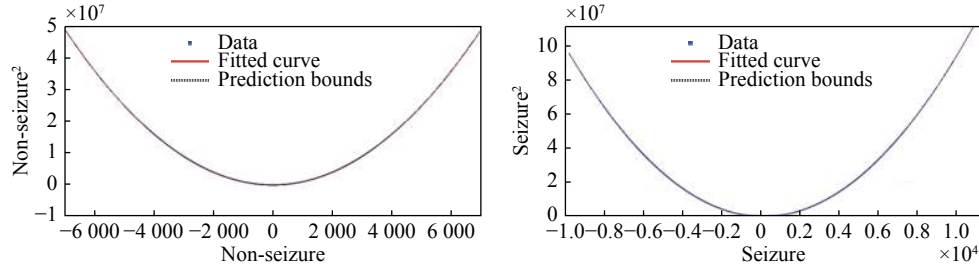
*Table 3* reports the four statistical parameters, calculated for all seizure and non-seizure segments. The weighted sum of squared residuals ( $\zeta$ ) and the root mean squared error ( $\psi$ ) show much larger values for seizure events with respect to non-seizure events, which suggests that these features can be used to discriminate between seizure and non-seizures in EEG signals. While R-square ( $\phi$ ) and adjusted R-square ( $\sigma$ ) have values close to one for both types of events, which suggests that the model has high accuracy in the fit. These four parameters show good potential to

**Table 1 Length of the 33 seizures used in this study**

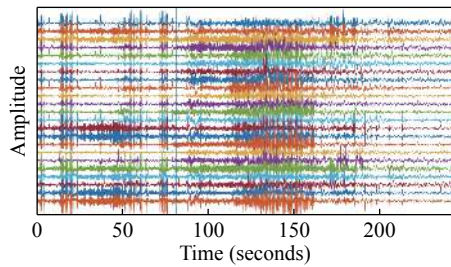
Epoch	Seizure	Duration (seconds)	Gender	Age (years)
01	01_03	40	F	11
02	01_04	27		
03	01_15	40		
04	01_18	90		
05	01_21	115		
06	01_26	101		
07	02_16	82	M	11
08	02_16	21		
09	02_19	9		
10	03_01	52	F	14
11	03_02	55		
12	03_03	69		
13	03_04	52		
14	03_34	47		
15	03_35	64		
16	03_36	67		
17	05_17	120	F	7
18	05_22	117		
19	06_18	12	F	1.5
20	06_24	16		
21	07_12	86	F	14.5
22	07_13	144		
23	07_19	83		
24	08_02	189	M	3.5
25	08_05	190		
26	08_21	338		
27	09_06	64	F	10
28	09_19	64		
29	10_20	10	M	3
30	10_27	65		
31	10_30	62		
32	10_31	76		
33	10_89	54		

**Table 2 The quadratic linear-parabolic model coefficients and the associated 95% confidence bounds (CB)**

Coefficients	Non-seizure		Seizure	
	Value	95% CB	Value	95% CB
a	-0.6855	(-4.095, 2.725)	1.368	(-6.717, 9.453)
b	0.9989	(0.9989, 0.9989)	0.9998	(0.9998, 0.9998)
c	-36.05	(-38.53, -33.57)	-35.45	(-41.47, -29.43)



**Fig. 1** Curves of the quadratic linear-parabolic model for seizure and non-seizure events. Note that the scale of the seizure curve (Right) is much larger than the one of the non-seizure curve (Left). The X-axis represents  $\bar{x}$  while the Y-axis represents  $\bar{x}^2$  (the unit on the X and Y axes correspond to the amplitude of the signal). The prediction bounds have a small uncertainty throughout the entire data range, therefore new observations can be predicted with high accuracy.



**Fig. 2** EEG raw example. The onset detection begins in 81 seconds according to the medical annotation but by non-expert visual inspection cannot reach the same conclusion. The y-axis scale for each channel is between  $\pm 200$  mv.

Events	$\zeta$	$\phi$	$\sigma$	$\psi$
Non-seizure	3.624e+14	0.9996	0.9996	4839
Seizure	2.03e+15	0.9999	0.9999	1.145e+04

discriminate between seizure and non-seizures in EEG signals. This can be visually corroborated in **Fig. 1**.

Estimating the four statistical parameters at each 1-second segment allows us to detect changes between epileptic EEG seizures. When we apply equation (7) to each 1-second segments, similar results are obtained as those applied to the signal as a whole. **Fig. 3** shows the similarity with **Fig. 1**. Both seizure and non-seizure events have an excellent fitting with the linear-parabolic model, suggesting that the scale changes and the different values of the estimated parameters can be used to differentiate seizure and non-seizure events in EEG signals.

To cope with the high variability of EEG signals, the features were normalized to be in the range  $[0,1]$ <sup>[33]</sup>. **Table 4** reports the mean values for the four statistical parameters  $\rho=[\zeta, \phi, \sigma, \psi]$  calculated for each 1-second segment from  $X[n]$ . The weighted sum of squared residuals ( $\zeta$ ) has values closer to zero while R-square ( $\phi$ ) and adjusted R-square ( $\sigma$ ) have values closer to one. This shows that the model has a high fit accuracy and can be used for seizure detection, as

corroborated by **Fig. 3**.

A value closer to zero in the weighted sum of squared residuals ( $\zeta$ ) and Root mean squared error ( $\psi$ ) and values closer to one in R-square ( $\phi$ ) and adjusted R-squared ( $\sigma$ ) suggest a good fit quality and therefore can be used as a predictor.

**Fig. 4** shows the main scatter plots of couples of the normalized statistical parameters ( $\zeta, \phi, \sigma, \psi$ ). It is interesting to note that in  $\zeta$  versus  $\phi$  (**Fig. 4A**) and  $\psi$  versus  $\phi$  (**Fig. 4C**) the seizure events (red dots) are concentrated closer to one with respect to non-seizure events (blue dots). In **Fig. 4B**,  $\zeta$  versus  $\psi$  has a big tendency to go to the right in the seizure events (red dots) with respect to non-seizure events (blue dots). In **Fig. 4D**,  $\phi$  versus  $\sigma$  shows a clear concentration of seizure events (red dots) with respect to non-seizure events (blue dots) on the upper right side. These scatter plots show that our predictor vector  $\rho=[\zeta, \phi, \sigma, \psi]$  is potentially useful to discriminate between seizures and non-seizures events in EEG signals.

To assess the performance of the proposed method, we adopted a supervised testing approach and used 1 518 events (1-second signals) to train and test the method with the 20-fold cross-validation technique of the predictor vector  $\rho=[\zeta, \phi, \sigma, \psi]$ . As explained above, the events were extracted from 33 seizure epochs and 33 non-seizure epochs, giving a total of 66 epochs from 9 different subjects. The method gives good classification performance, with 92% true positives rate (TPR) or sensitivity, 96% true negative rate (TNR) or specificity, 4% false positive rate, and 94.1% accuracy. Please note that segments for training and testing are drawn randomly without consideration of their epochs. This process was repeated 1 000 times to ensure not-bias in the partitioning.

### Discussion

This work presented a new method to detect epileptic seizures in EEG signals. The proposed model-based classification method relies on the design



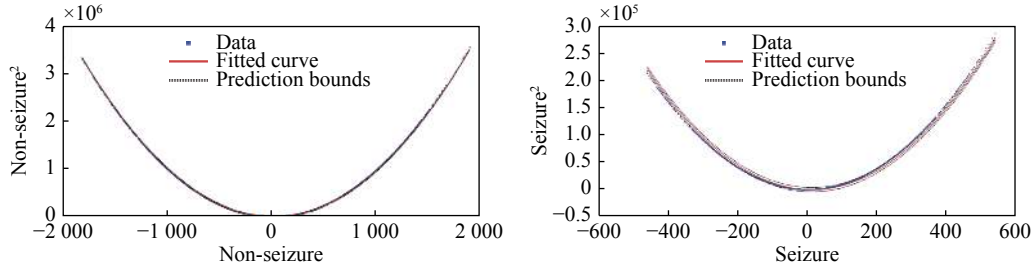


Fig. 3 Quadratic parabolic curve fitting example for one segment  $\tilde{X}[n]$ .

**Table 4** Means of the statistical parameters with values normalized between 0 and 1 for all each 1-second time segments

Events	$\zeta$	$\phi$	$\sigma$	$\psi$
Non-seizure	0.14	0.99	0.99	0.30
Seizure	0.11	1.00	1.00	0.25

of a specific filter using the two-point central difference algorithm. A linear-parabolic model is fitted to the filtered signal using mean squares. Four statistical parameters associated with the model fitting are used as classification features to discriminate seizure and non-seizure events. A random forest classifier has been adopted to evaluate the ability of these parameters to detect seizures. The proposed methodology was applied to 66 balanced events from the Children Hospital Boston database. Results suggest that the proposed algorithm is a powerful tool for detecting epileptic seizure events in EEG signals achieving a 94.1% accuracy.

The proposed method has two main advantages

compared to existing methods. First, its low computational complexity makes it implementable in quasi-real-time. Second, the method can be easily extended to work separately with different brain rhythms, by adapting the skip factor  $L$ . Despite these advantages, the method suffers some limitations, mainly noise and robustness.

The signal is affected by noise and artifacts due to the acquisition and pre-processing. This affects the precision of the results, especially when the method is generalized to complex epileptic forms. Outliers are not directly considered in the estimation procedure that needs more robustness.

Future work will focus on robust approaches to consider noise and artifacts. Optimization techniques will be investigated to remove outliers and improve the accuracy of the detection.

#### Acknowledgments

Part of the work was conducted by AQR during his Ph.D. work at Buenos Aires Institute of Technology (ITBA).

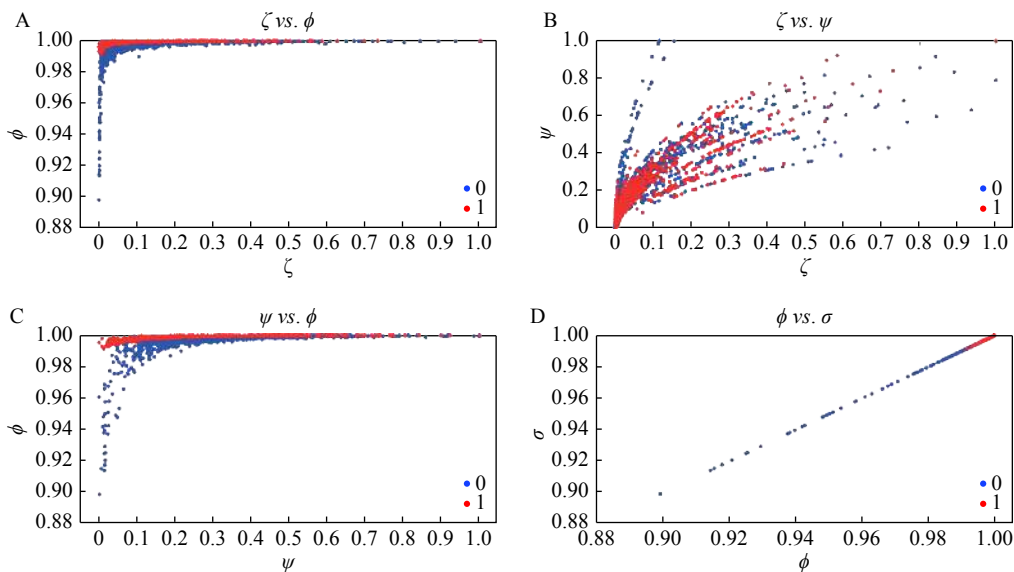


Fig. 4 Scatter plots for the four statistical parameters ( $\zeta$ ,  $\phi$ ,  $\sigma$ ,  $\psi$ ) observed through 1 518 events from 66 epochs, 33 seizures (red dots) and 33 on-seizures (blue dots) before the seizure. The weighted sum of squared residuals ( $\zeta$ ), R-square ( $\phi$ ), adjusted R-squared ( $\sigma$ ) and root mean squared error ( $\psi$ ) shows a clear concentration of seizure events (red dots) with respect to non-seizure events (blue dots).

## References

- [1] Frei MG, Davidchack RL, Osorio I. Least squares acceleration filtering for the estimation of signal derivatives and sharpness at extrema [and application to biological signals][J]. *IEEE Trans Biomed Eng*, 1999, 46(8): 971–977.
- [2] Chernov N, Lesort C. Statistical efficiency of curve fitting algorithms[J]. *Comput Stat Data Anal*, 2004, 47(4): 713–728.
- [3] López-Rubio E, Thurnhofer-Hemsi K, Blázquez-Parra EB, et al. A fast robust geometric fitting method for parabolic curves[J]. *Pattern Recognit*, 2018, 84: 301–316.
- [4] Zhang HH, Su JZ, Wang QY, et al. Predicting seizure by modeling synaptic plasticity based on EEG signals - a case study of inherited epilepsy[J]. *Commun Nonlinear Sci Numer Simul*, 2018, 56: 330–343.
- [5] Ramon C, Holmes MD, Wise MV, et al. Oscillatory patterns of phase cone formations near to epileptic spikes derived from 256-channel scalp EEG data[J]. *Comput Mathem Methods Med*, 2018, 2018: 9034543.
- [6] Liu HY, Yang Z, Meng FG, et al. Chronic vagus nerve stimulation reverses heart rhythm complexity in patients with drug-resistant epilepsy: an assessment with multiscale entropy analysis[J]. *Epilepsy Behav*, 2018, 83: 168–174.
- [7] Begemann A, Acuña MA, Zweier M, et al. Further corroboration of distinct functional features in SCN2A variants causing intellectual disability or epileptic phenotypes[J]. *Mol Med*, 2019, 25: 6.
- [8] Liang SF, Wang HC, Chang WL. Combination of EEG complexity and spectral analysis for epilepsy diagnosis and seizure detection[J]. *EURASIP J Adv Sign Processing*, 2010, 2010: 853434.
- [9] Sorensen TL, Olsen UL, Conradsen I, et al. Automatic epileptic seizure onset detection using matching pursuit: a case study[C]//Proceedings of 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology. Buenos Aires, Argentina: IEEE, 2010: 3277–3280.
- [10] Nasehi S, Pourghassem H. A novel fast epileptic seizure onset detection algorithm using general tensor discriminant analysis[J]. *J Clin Neurophysiol*, 2013, 30(4): 362–370.
- [11] Zhang ZN, Wen TX, Huang W, et al. Automatic epileptic seizure detection in EEGs using MF-DFA, SVM based on cloud computing[J]. *J X-Ray Sci Technol*, 2017, 25(2): 261–272.
- [12] Mahmoodian N, Boese A, Friebe M, et al. Epileptic seizure detection using cross-bispectrum of electroencephalogram signal[J]. *Seizure*, 2019, 66: 4–11.
- [13] Polychronaki G, Ktonas P, Gatzonis S, et al. Comparison of fractal dimension estimation algorithms for epileptic seizure onset detection[J]. *J Neural Eng*, 2014, 7(4): 046007.
- [14] Birjandtalab J, Pouyan MB, Cogan D, et al. Automated seizure detection using limited-channel EEG and non-linear dimension reduction[J]. *Comput Biol Med*, 2017, 82: 49–58.
- [15] Acharya UR, Oh SL, Hagiwara Y, et al. Deep convolutional neural network for the automated detection and diagnosis of seizure using EEG signals[J]. *Comp Biol Med*, 2018, 100: 270–278.
- [16] Quintero-Rincón A, Pereyra M, D'Giano C, et al. Fast statistical model-based classification of epileptic EEG signals[J]. *Biocybern Biomed Eng*, 2018, 38(4): 877–889.
- [17] Breiman L. Random forests[J]. *Mach Learn*, 2001, 45(1): 5–32.
- [18] Flach P. Machine learning[M]. Cambridge: Cambridge University Press, 2012.
- [19] Manzouri F, Heller S, Dümpelmann M, et al. A comparison of machine learning classifiers for energy-efficient implementation of seizure detection[J]. *Front Syst Neurosci*, 2018, 12: 43.
- [20] Le Douget JE, Fouad A, Maskani Filali M, et al. Surface and intracranial EEG spike detection based on discrete wavelet decomposition and random forest classification[C]//Proceedings of the 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Seogwipo, South Korea: IEEE, 2017: 475–478.
- [21] Donos C, Dümpelmann M, Schulze-Bonhage A. Early seizure detection algorithm based on intracranial EEG and random forest classification[J]. *Int J Neural Syst*, 2015, 25(5): 1550023.
- [22] López S, Suarez G, Jungreis D, et al. Automated identification of abnormal adult EEGs[C]//Proceedings of 2015 IEEE Signal Processing in Medicine and Biology Symposium. Philadelphia, PA, USA: IEEE, 2015.
- [23] Shoeb A, Edwards H, Connolly J, et al. Patient-specific seizure onset detection[J]. *Epilepsy Behav*, 2004, 5(4): 483–498.
- [24] Goldberger AL, Amaral LA, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals[J]. *Circulation*, 2000, 101(23): E215–E220.
- [25] Sanei S. Adaptive processing of brain signals[M]. United Kingdom: Wiley, 2013.
- [26] Bahill AT, McDonald JD. Frequency limitations and optimal step size for the two-point central difference derivative algorithm with applications to human eye movement data[J]. *IEEE Trans Biomed Eng*, 1983, 30(3): 191–194.
- [27] Tseng CC, Lee SL. Design of digital differentiator using difference formula and Richardson extrapolation[J]. *IET Sign Process*, 2008, 2(2): 177–188.
- [28] Semmlow JL, Griffel B. Biosignal and medical image processing[M]. 3rd ed. Boca Raton: CRC Press, 2014.
- [29] Hansen PC, Pereyra V, Scherer G. Least squares data fitting with applications[M]. Baltimore: Johns Hopkins University Press, 2013.
- [30] Neuhauser C. Calculus for biology and medicine[M]. 2<sup>nd</sup> ed. Upper Saddle River, NJ: Pearson, 2004.
- [31] Samarasinghe G, Sowmya A, Moses DA. A semi-quantitative analysis model with parabolic modelling for DCE-MRI sequences of prostate[C]//Proceedings of 2014 International Conference on Digital Image Computing: Techniques and Applications. Wollongong, NSW, Australia: IEEE, 2014.
- [32] Zhang C, Ma YQ. Ensemble machine learning: methods and applications[M]. Boston, MA: Springer, 2012.
- [33] Quintero-Rincón A, Careno C, Ems J, et al. Spike-and-wave epileptiform discharge pattern detection based on Kendall's Tau-b coefficient[J]. *Appl Med Inform*, 2019, 41(1): 1–8.