

Spike-and-wave epileptiform discharge pattern detection based on Kendall's Tau-b coefficient

Antonio QUINTERO-RINCÓN^{1,*}, Catalina CARENZO^{1,+}, Joaquín EMS^{1,+},
Lourdes HIRSCHSON^{1,+}, Valeria MURO² and Carlos D'GIANO²

¹ Department of Bioengineering, Instituto Tecnológico de Buenos Aires, Avenida Eduardo Madero 399, C1106ACD CABA, Buenos Aires, Argentina.

² Centro Integral de Epilepsia, Fundación para la Lucha contra las Enfermedades Neurológicas de la Infancia (FLENI), Montañeses 2325, C1428 AQB, Buenos Aires.

E-mail(s): ccarenzo@itba.edu.ar; jems@itba.edu.ar; lhirschson@itba.edu.ar; vmuro@fleni.org.ar; cdgiano@fleni.org.ar

* Author to whom correspondence should be addressed; Tel.:+54011 2150-4800.

+ All authors contributed equally to this work.

Received: November 22, 2018 / Accepted: March 30, 2019 / March 31, 2019

Abstract

Epilepsy is an important public health issue. An appropriate epileptiform discharge pattern detection of this neurological disease is a typical problem in biomedical engineering. In this paper, a new method is proposed for spike-and-wave discharge pattern detection based on Kendall's Tau-b coefficient. The proposed approach is demonstrated on a real dataset containing spike-and-wave discharge signals, where our performance is evaluated in terms of high Specificity, rule in (SpPI_n) with 94% for patient-specific spike-and-wave discharge detection and 83% for a general spike-and-wave discharge detection.

Keywords: Spike-and-wave discharge; Kendall's Tau-b coefficient; Electroencephalography (EEG); Epilepsy; high Specificity, rule in (SpPI_n)

Introduction

Electroencephalography (EEG) is widely used to record the electrical activity of the brain in neurological health centers. EEGs help study and diagnose several different types of brain disorders, such as epilepsy. Epilepsy is a neurological disorder caused by intense activity of nerve cells in the brain, causing seizures. This uncontrolled electrical disturbance can cause changes in the levels of consciousness, behavior and body movements. A spike-and-wave discharge (SWD) is an epileptiform discharge with a regular and symmetric morphology, which typically starts and ends abruptly [1], see Figure 1. In current literature, epileptiform pattern recognition has extensive signal processing methods for accurate detection. Some approaches can be found using spectrogram with harmonic analysis [2], discrete cosine transform coupled with Daubechies wavelet [3], time-frequency and non-linear analysis [4], spatio-temporal analysis combine with autocovariance [5], Hilbert-Huang transform [6], Pseudo-Wigner-Ville and Choi-William distributions followed by Renyi's entropy [7], convolutional neural networks [8], statistical modeling [9, 10], bootstrap resampling [11], complex network of neuronal oscillators [12], or by using cross-approximate entropy [13]. See [14-16] for some studies in morphological similarity or concordance between

signals and [17] for a complete state-of-the-art about methods of automated absence seizure detection.

Kendall's Tau-b coefficient is a nonparametric correlation analysis used to measure the ordinal association or concordance between two measured quantities. In our case this was done with two morphological waveforms signals. EEG studies with this coefficient are very diverse, in [18] with correlational neuronal activity in sleep; in [19] to compare the quality of life in epileptic patients; in [20] using time series or envelopes EEG's pairs to quantify dependence between channels were investigated, or in [21] where connectivity assessed by intracranial electrical stimulation was used to quantify and detect the association among multichannel biosignals.

In this work we study Kendall's Tau-b coefficient between two morphological waveforms of the same length. The goal is to estimate the statistical relationship between a spike-and-wave epileptiform discharge (SWD) against each EEG segment by channel, in order to quantify and detect the morphological similarity or concordance between signals.

Material and Method

Database

A database with 780 monopolar 256 Hz signals was created measured from ten patients from Fundación Lucha contra las Enfermedades Neurológicas Infantiles (FLENI). 390 spike-and-wave signals had different time-length and waveform but their morphology is preserved, while 390 non-spike-and-wave signals had normal waveforms, see [9,10] for more details. In this work, only 300 spike-and-wave signals were used, see Figure 1 and Figure 3.

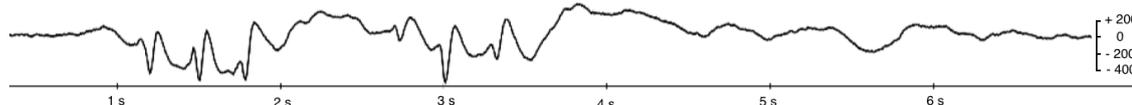


Figure 1. Spike-and-wave discharge example in an EEG channel. We can see the SWD epileptiform pattern between 1- 4 seconds

SWD are restricted to a narrow frequency band between 1-3 Hz. Each EEG was acquired with a 22-channel array using the standard 10-20 system through the following channels: Fp1, Fp2, F7, F3, Fz, F4, F8, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6, O1, O2, Oz, FT10 and FT9, see Figure 2.

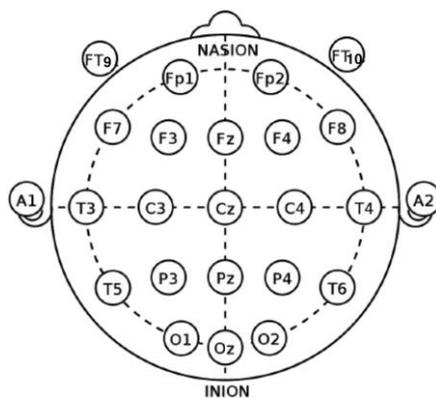


Figure 2. Electrodes position used in this work.

Methodology

Let $\mathbf{X}_r \in \mathbb{R}^{M \times N}$ denote the matrix assembly M EEG signals $x_m \in \mathbb{R}^{1 \times N}$ measured simultaneously on different channels and at N discrete time instants.

Let $\mathbf{Y}_p \in \mathbb{R}^{1 \times L}$ denote the spike-and-wave vector with length L . The proposed methodology was composed of four stages. The first stage splits each channel of the original signal \mathbf{X}_r into a set of N/L segments of the same sampling size of each spike-and-wave discharge without overlap. This segmentation was done using a rectangular sliding window, such that

$$\Omega_s = \frac{-(N/L) - 1}{2} \leq s \leq \frac{(N/L) - 1}{2} \quad (1)$$

so that $\mathbf{X}_s = \Omega_s \mathbf{X}_r$. The second stage consisted of filtering the segments \mathbf{X}_s and \mathbf{Y}_p with a k -point moving average filter, so that

$$\Gamma[n] = \sum_{k=0}^{K-1} \frac{1}{b[k]} \Gamma_i[n-k] \quad (2)$$

where $\Gamma[n]=\mathbf{X}_f[n]$ for input $\Gamma_i = \mathbf{X}_s$; $\Gamma[n]=\mathbf{Y}_f[n]$ for input $\Gamma_i = \mathbf{Y}_p$ and empirical value $k=5$, this value was chosen in order to not to change the original waveform signal characteristics. Next, in third stage $\mathbf{X}_f[n]$ and $\mathbf{Y}_f[n]$ were scaled between lower limit $\alpha = -1$ and upper limit $\beta = +1$ using

$$\mathbf{Y}[n] = \frac{(v[n] - \min(v[n]))(\beta - \alpha)}{\max(v[n]) - \min(v[n])} + \alpha \quad (3)$$

where $\mathbf{Y}[n]=\mathbf{X}[n]$ for input $v[n] = \mathbf{X}_f[n]$ and $\mathbf{Y}[n]=\mathbf{Y}[n]$ for input $v[n] = \mathbf{Y}_f[n]$.

Finally in the fourth stage a Kendall's Tau-b coefficient was applied between the SWD pattern $\mathbf{Y}[n]$ and each $\mathbf{X}[n]$ segment in order to detect a spike-and-wave epileptiform discharge pattern into EEG signals. We now introduce the Kendall's Tau-b coefficient used in this paper.

Kendall's Tau-b Coefficient

Let \mathbf{X} be each segment filtered and scaled for each EEG channel and \mathbf{Y} a SWD pattern filtered and scaled waveform, both through $\Gamma[n]$ and $\mathbf{Y}[n]$, see equations (2) and (3). Then the bivariate couples $(x_1, y_1), \dots, (x_n, y_n)$ are a sample of observations of the combined random variables \mathbf{X} and \mathbf{Y} , such that all the values of x_i and y_i are unique. Two bivariate observations (x_i, y_i) and (x_j, y_j) are:

- Concordant: when $x_i > x_j$ and $y_i > y_j$, or $x_i < x_j$ and $y_i < y_j$
- Discordant: when $x_i > x_j$ and $y_i < y_j$, or $x_i < x_j$ and $y_i > y_j$
- Neither concordant: when $x_i = x_j$, or $y_i = y_j$

Kendall's Tau-b coefficient can be estimated in two ways, a) according the numbers of concordant (n_c) and discordant (n_d) pairs.

$$\tau = 2 \frac{n_c - n_d}{n(n-1)} \quad (4)$$

where the denominator is the total number of pair combinations, $-1 \leq \tau \leq 1$. Or b) through the following expression

$$\tau = \frac{2}{n(n-1)} \sum_{i < j}^{N/L} \text{sgn}(x_i - x_j) \text{sgn}(y_i - y_j) \quad (5)$$

where

- $\tau = +1$: Implies perfect agreement between two rankings, when \mathbf{X} and \mathbf{Y} are the same.
- $\tau = -1$: Implies perfect disagreement between two rankings, \mathbf{X} ranking is the reverse of \mathbf{Y} .

- $\tau = 0$: Implies that \mathbf{X} and \mathbf{Y} are independent.

The statistical Kendall's Tau-b coefficient significance given by p -value is

$$p = \frac{3\tau\sqrt{n(n-1)}}{\sqrt{2(2n+5)}} \quad (6)$$

We refer the reader to [22,23] for a comprehensive treatment of the statistical and mathematical properties of Kendall's Tau-b coefficient.

Classifiers

Linear discriminant, quadratic discriminant and support vector machine (SVM) are machine learning techniques used in pattern recognition problems in order to make predictions or decisions about the input data. The process of predicting the class of given input data is according to the targets, labels or categories during the classification stage. In our case, these learning techniques use the input vector $[\tau, p] \in \mathbb{R}^2$ in order to predict between two classes namely: spike-and-wave and non-spike-and-wave, both training and classification stage. In these techniques the sensibility and the specificity are estimated by using the classical ROC analysis [24]. See [25] for deeper information about these techniques.

Results and Discussion

We evaluated the proposed methodology using 22 EEG monopolar 256 Hz channels of one patient of study during a long-time recording, and 300 spike-and-wave epileptiform discharges from the database presented above in subsection *Database*. Note that, the long-time signals were recorded during a sleep period of 8 hours. All the epochs of the study were selected by the physician. Therefore, each epoch had the beginning and end of each spike-and-wave epileptiform discharge, which we used as ground truth.

For illustration, Figure 3. shows some examples of the annotated medical data (a, c, e) and candidates for our method (b, d, f) in blue colors with respect to the SWD pattern in red color. The annotated medical data was chosen in a medical context, looking at the signal in the time domain, see Figure 1. In (b) example, we can see the similar morphology. In (d) example a wave is detected, but not the spike. It is a false detection, and in (f) example a false detection is found.

In total 3080 τ and p -values were estimated in 140 segments by each channel. The obtained τ and p -values for each segment per channel were corroborated with the medical annotations by visual inspection. Based on this analysis, a threshold= 0.5 and p -value= 0.05 were selected in order to perform a ROC analysis [24].

The threshold value was selected because values that were higher than this value were good candidates to be spike-and-waves, considering that the Kendall's Tau-b coefficient is a nonparametric correlation analysis; and the p -value was estimated using the equation (6). Therefore, the Tau-b statistic value was assigned to the annotated data. Next, these were compared with the threshold value; and later added to the confusion matrix in order to assign a specificity and sensitivity metric value.

The percentage of correct classifications was analyzed only in terms of high Specificity, rule in (SpPI_n) [26-28], because it is known that, the patient of study has spike-and-wave discharges and in this preliminary study it was important to know if the proposed method detects the SWD pattern. The values obtained were: 14% sensitivity (True positive rate), 83% specificity (True negative rate) evaluated with all 300 spike-and-waves from database in 3080 segments. Specificity values range from 0.81 to 0.84 while sensitivity values range from 0.12 to 0.16 following a 95% confidence interval.

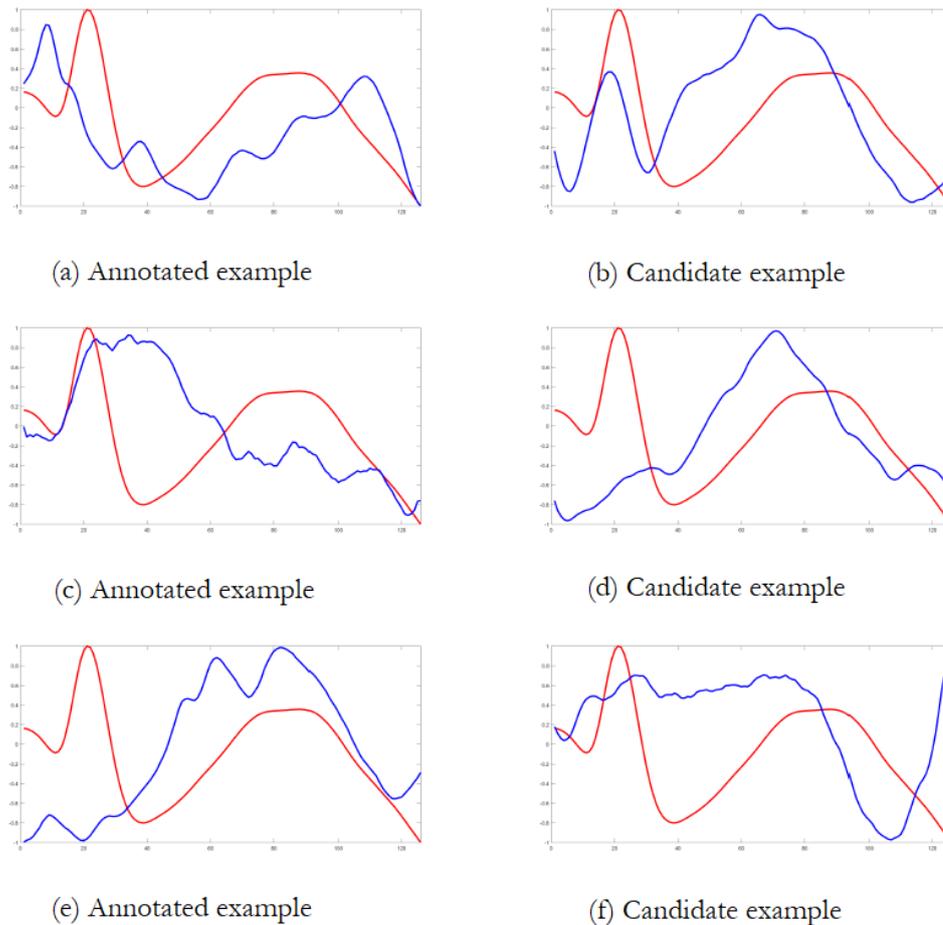


Figure 3. Left column are annotated medical data (a), (c) and (e). Right column are candidates for our method (b), (d) and (f). We can see how the EGG waveforms (blue color) have similar morphology with the SWD pattern (red color). In (b) example, we can see the similar morphology. (d) and (f) examples, are false detections. In (d) a wave is detected, but not the spike and in (f) the waveform is different.

If only 10 extracted patterns from each new patient are analyzed and included into the database, SpPI_n results improve by 11%; with 94% specificity with values ranges from 0.93 to 0.94 for the 95% confidence interval. This presented us with our next hypothesis: spike-and-wave epileptiform discharge detection must be focused on patient-specific spike-and-wave detection. With this hypothesis, three supervised classifiers were tested with leave-one-out cross-validation technique from vector $[\tau, p] \in \mathbb{R}^2$, obtaining the same SpPI_n results (94% specificity) only with the 10 SWD patterns: linear discriminant, quadratic discriminant and lineal Support Vector Machine (SVM). This suggests that with our method is possible train a classifier and detect with a small dataset, similar to [29]. Our results are promising in comparison to other methods of the state-of-the-art that reports their SpPI_n[2] by using harmonic analysis with SpPI_n of 97%, or mixing the discrete cosine transform with Daubechies wavelets with SpPI_n 90.7% [3], or applying the Morlet wavelet energy with SpPI_n of 98.7% [30].

The main limitation was defining the sliding time-window and the perfect overlap of epochs because epileptic signals have a high dynamic. The main advantage is that, it is possible to train a classifier and detect with a small dataset. But this remains an open issue.

Future work will focus on a deep evaluation of the proposed approach with a large group of patients: to determine the amplitude difference between signals in parallel brain regions, to estimate the l_1 -norm coupled with a filtering stage to outliers control [31], to apply an adaptive signed

correlation index [14], to improve the sensitivity in new patients with regularization techniques [32] and to tune parameters with information theory [33] or ensembles from multiple models [34].

Conclusions

This work presented a new spike-and-wave detection method to detect epileptiform discharge patterns in EEG signals. The method is based on Kendall's Tau-b coefficient which captures the statistical relationship between two waveform signals involved in the bivariate structure in order to quantify and detect the morphological similarity or concordance between signals. The proposed methodology was demonstrated on 3080 EEG segments of 300 monopolar 256 Hz spike-and-wave discharges database from Fundación Lucha contra las Enfermedades Neurológicas Infantiles (FLENI), which suggests that the proposed algorithm is a powerful tool for detecting seizures in epileptic signals in terms of high Specificity, rule in (SpPIn), 94% for patient-specific SWD detection and 83% for a general SWD detection.

List of abbreviations

EEG: Electroencephalography.
SpPIn: high Specificity, rule in.
SWD: spike-and-wave discharge.

Conflict of Interest

Authors declare that they have no conflict of interest.

Authors' Contributions

AQR carried out the aim of research and the design of experiment. AQR, CC, JE and LH carried out the experiments and participated in the design of the study and performed the statistical analysis. VM, CD and AQR designed the database with the medical annotations. All authors coordinated and helped to draft the manuscript. All authors read and approved the final manuscript.

References

1. Niedermeyer E, Lopes da-Silva F. *Electroencephalography Basic Principles, Clinical Applications, and Related Fields*: Lippincott Williams and Wilkins; 2010.
2. Hese PV, Martens J, Waterschoot L, Boon P, Lemahieu I. Automatic detection of spike and wave discharges in the EEG of genetic absence epilepsy rats from strasbourg. *IEEE Trans Biomed Eng.* 2009;56(3):706-717.
3. Ma D, Zheng X, Wang Z. Detection of epileptic spike wave in EEG signals based on morphological component analysis. *Journal of Biomedical Engineering.* 2013;30(4):710-713.
4. Gajic D, Djurovic Z, Jovan G, Di Gennaro S, Savic-Gajic I. Detection of epileptiform activity in EEG signals based on time-frequency and non-linear analysis. *Front Comput Neurosci.* 2015;9:38.
5. Rodgers K, Dudek F, DS B. Progressive, Seizure-Like, Spike-Wave Discharges Are Common in Both Injured and Uninjured Sprague-Dawley Rats: Implications for the Fluid Percussion Injury Model of Post-Traumatic Epilepsy. *J Neurosci.* 2015;35(24):9194-9204.

6. Zhu J, Lin C, Chang S, Wang J, Peng T, YY C. Analysis of spike waves in epilepsy using Hilbert-Huang transform. *J Med Syst.* 2015;39(170):1-13.
7. Khoshnoud S, Shamsi S. Epilepsy Classification Framework Utilizing Joint TimeFrequency Signal Analysis and Processing. *Appl Med Inform.* 2015;36(2):1-8.
8. Johansen A, Jin J, Maszczyk T, Dauwels J, Cash S, Westover M. Epileptiform spike detection via convolutional neural networks. In *Proc IEEE Int Conf Acoust Speech Signal Process 2016*, pp. 754-758.
9. Quintero-Rincón A, Prendes J, Muro V, D'Giano C. Study on spike-and-wave detection in epileptic signals using t-location-scale distribution and the k-nearest neighbors classifier. In *IEEE URUCON Conferences 2017*, pp. 1-4.
10. Quintero-Rincón A, Risk M, Muro V, D'Giano C, Prendes J, Batatia H. A novel spike-and-wave automatic detection in EEG signals. *International Journal of Signal and Imaging Systems Engineering.* 2018 (in press).
11. Joshi C, Chapman K, Bear J, Wilson S, Walleigh D, Scheuer M. Semiautomated Spike Detection Software Persyst 13 Is Noninferior to Human Readers When Calculating the Spike-Wave Index in Electrical Status Epilepticus in Sleep. *J Clin Neurophysiol.* 2018;35(5):370-374.
12. Medvedeva T, Sysoeva M, van Luijtelaaar G, Sysoev I. Modeling spike-wave discharges by a complex network of neuronal oscillators. *Neural Netw.* 2018;98:271-282.
13. Liu X, Zhao C, Sun C, Chen Z, Hayat T, Alsaedi A. A new closed-loop strategy for detection and modulation of epileptiform spikes based on cross approximate entropy. *Journal of Integrative Neuroscience.* 2018;17(3):271-279.
14. Lian J, Garner G, Muessig D, V. L. A simple method to quantify the morphological similarity between signals. *Signal Processing.* 2010;90(2):684-688.
15. Olenko A, Wong K, Mir H, Al-Nashash H. A generalized correlation index for quantifying signal morphological similarity. *Electronics Letters.* 2016;52(22):1832-1834.
16. Jas M, Dupré T, La Tour D, Simsekli U, Gramfort A. Learning the Morphology of Brain Signals Using Alpha-Stable Convolutional Sparse Coding. *31st Conference on Neural Information Processing Systems*, 2017.
17. Orem J. Augmenting expiratory neuronal activity in sleep and wakefulness and in relation to duration of expiration. *J Appl Physiol.* 1998;85(4):1260-1266.
18. Lowe A, David E, Kilpatrick C, Matkovic Z, Cook M, Kaye A, et al. Epilepsy surgery for pathologically proven hippocampal sclerosis provides long-term seizure control and improved quality of life. *Epilepsia.* 2004;3(45):237-242.
19. Fadlallah B, Brockmeier A, Seth S, Li L, Keil A, Príncipe J. An association framework to analyze dependence structure in time series. In *Conf Proc IEEE Eng Med Biol Soc.* 2012, pp. 6176-6179.
20. Xu W, Chen Z, Zhang Y, Cheng L. Order statistics concordance coefficient with applications to multi-channel biosignal analysis. *IEEE J. Biomedical and Health Informatics.* 2017;21(5):1206-1215.
21. Kendall M. Rank correlation methods: Hafner Publishing Co.; 1955.
22. Shevlyakov G, Oja H. Robust Correlation Theory and Applications: Wiley Series in Probability and Statistics, 2016.
23. Fawcett T. An introduction to ROC analysis. *Pattern Recognition Letters.* 2006;27:861-874.
24. Pewsner D, Battaglia M, Minder C, Marx A, Bucher H, Egger M. Ruling a diagnosis in or out with "SpPIn" and "SnNOut": a note of caution. *BMJ Publishing Group Ltd.* 2004;7459(329):p. 209-213.
25. Power M, Fell G, Wright M. Principles for high-quality, highvalue testing. *BMJ Evidence-Based Medicine.* 2013;18(1):5-10.
26. Cohen J, Korevaar D, Altman D, Bruns D, Gatsonis C, Hooft L, et al. Stard 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. *BMJ Open.* 2016;6:e012799.

27. Quintero-Rincón A, Pereyra M, D'Giano C, Risk M, Batatia H. Fast statistical model-based classification of epileptic EEG signals. *Biocybernetics and Biomedical Engineering*. 2018;4(38):877-889.
28. Cha S. Comprehensive survey on distance/similarity measures between probability density functions. *International Journal of Mathematical Models and Methods in Applied Sciences*. 2007;4:300-307.
29. Flynn C, Hurvich C, Simonof J. On the sensitivity of the lasso to the number of predictor variables. *Statistical Science*. 2017;32(1):88-105.
30. Majda A, Gershgorin B. Improving model fidelity and sensitivity for complex systems through empirical information theory. *PNAS*. 2011;108(25):10044-10049.
31. Abuassba A, Zhang D, Luo X, Shaheryar A, Ali H. Improving classification performance through an advanced ensemble based heterogeneous extreme learning machines. *Computational Intelligence and Neuroscience*. 2017;2017:3405463.