

Article

# Effects of Tau and Sampling Frequency on the Regularity Analysis of ECG and EEG Signals Using ApEn and SampEn Entropy Estimators

Ricardo Espinosa <sup>1,\*</sup> , Jesica Talero <sup>1</sup>  and Alejandro Weinstein <sup>2</sup>

<sup>1</sup> Department of Biomedical Engineering, Universidad ECCI, Biomedical Applications EMB-IEEE, Bogotá 111311, Colombia; jesica.talerom@ecci.edu.co

<sup>2</sup> Center of Research and Development in Health Engineering, Valparaíso University, Valparaíso 2362905, Chile; alejandro.weinstein@uv.cl

\* Correspondence: respinosam@ecci.edu.co

Received: 30 October 2020; Accepted: 10 November 2020; Published: 14 November 2020



**Abstract:** Electrocardiography (ECG) and electroencephalography (EEG) signals provide clinical information relevant to determine a patient's health status. The nonlinear analysis of ECG and EEG signals allows for discovering characteristics that could not be found with traditional methods based on amplitude and frequency. Approximate entropy (ApEn) and sampling entropy (SampEn) are nonlinear data analysis algorithms that measure the data's regularity, and these are used to classify different electrophysiological signals as normal or pathological. Entropy calculation requires setting the parameters  $r$  (tolerance threshold),  $m$  (immersion dimension), and  $\tau$  (time delay), with the last one being related to how the time series is downsampled. In this study, we showed the dependence of ApEn and SampEn on different values of  $\tau$ , for ECG and EEG signals with different sampling frequencies ( $F_s$ ), extracted from a digital repository. We considered four values of  $F_s$  (128, 256, 384, and 512 Hz for the ECG signals, and 160, 320, 480, and 640 Hz for the EEG signals) and five values of  $\tau$  (from 1 to 5). We performed parametric and nonparametric statistical tests to confirm that the groups of normal and pathological ECG and EEG signals were significantly different ( $p < 0.05$ ) for each  $F$  and  $\tau$  value. The separation between the entropy values of regular and irregular signals was variable, demonstrating the dependence of ApEn and SampEn with  $F_s$  and  $\tau$ . For ECG signals, the separation between the conditions was more robust when using SampEn, the lowest value of  $F_s$ , and  $\tau$  larger than 1. For EEG signals, the separation between the conditions was more robust when using SampEn with large values of  $F_s$  and  $\tau$  larger than 1. Therefore, adjusting  $\tau$  may be convenient for signals that were acquired with different  $F_s$  to ensure a reliable clinical classification. Furthermore, it is useful to set  $\tau$  to values larger than 1 to reduce the computational cost.

**Keywords:** electrophysiological signals; nonlinear signals; entropy; sampling frequencies; time delay

## 1. Introduction

The health status of a patient can be evaluated through the signals generated by the body. These signals are captured by different biomedical technologies and provide relevant information that is interpreted by a clinician. Based on these signals, the clinician can then propose a treatment to restore the patient's health. One of the most popular techniques to capture biological signals is the placement of electrodes on the skin's surface. These electrodes allow for the recording of the sum of action potentials generated by excitable cells (muscle and neuronal) in the body's specific area. These signals are amplified, filtered, processed, and finally visualized and recorded in a digital medium or printed. Several physiological systems can be evaluated through these records. The most common of these electrophysiological signals are electrocardiographic (ECG) and electroencephalographic (EEG).

ECG signals provide information about the activity and function of the heart. Under normal conditions, the ECG signal presents waves with specific characteristics that are fully reported in the scientific and clinical literature. Arrhythmias are generated by a deterioration of the patient's cardiac system or as a symptom of another underlying disease. The ECG signal can be described through its spectral characteristics, its morphology, and its regularity. Changes in these characteristics can help identify arrhythmias [1]. With the increase in deaths due to heart disease, it is paramount to develop techniques to detect heart pathologies with high precision during the clinical exploration [2].

EEG signals serve to analyze a patient neuronal activity using electrodes placed on the scalp. EEG recordings allow physicians to evaluate the patient while conscious or during sleep. The EEG allows for the evaluation of pathological states such as seizures, epilepsy, and other degenerative diseases. The EEG can be analyzed in the time and the frequency domains [3].

Traditional analysis methods provide superficial information about the biological system dynamics that generate the electrophysiological signals. Amplitude-based methods (such as mean, rectification, and root mean square) and spectrum analysis-based methods (such as Fourier transform and power spectral density) have been successfully used in the past. However, they have several disadvantages. For example, signals with very small amplitudes are more sensitive to noise contamination from surrounding radio transmissions, cables pickup, and artifacts due to patient movement [4]. Algorithms based on amplitude or frequency can be highly affected when the signal is contaminated with noise from various sources. Perhaps the main disadvantage of traditional algorithms based on amplitude or frequency is that they cannot widely describe the system's dynamics since they are based on a linear paradigm, where small changes in the input produce small changes in the output [5]. Many characteristics that best describe the system are hidden under this approach.

Current technologies and measurement systems have allowed us to better understand the elements and their way of interaction that make up the human body's systems. Likewise, the evolution of processing methods has made it possible to obtain more coherent information regarding the phenomena present in biological systems. However, there is still a long way to find the best tools to quantify complex phenomena effectively [6]. Physiological systems expose complex nonlinear dynamics. They consider inputs dependent on time and also inputs with random characteristics (described in this way because many details of biological interaction are still unknown) that produce irregular outputs, which have to be analyzed using nonlinear time series methods [5].

A significant number of methods have been developed to analyze nonlinear signals and characterize a biological system's dynamic behavior. Nonlinear approaches, such as effective correlation dimension, correlation density, Hurst exponent, Lyapunov exponents, approximate entropy (ApEn), and sampling entropy (SampEn), have been applied to electrophysiological signals such as ECG and EEG [7–9]. This study focuses on ApEn and SampEn because numerous studies that use these estimators have emerged in recent years, both at the research-theoretical level and at the level of technological development [10].

Statistical estimators, such as entropy, are useful to characterize and analyze ECG and EEG signals. ApEn, proposed by Pincus [11,12], quantifies a time series's regularity, particularly in nonlinear dynamic signals such as physiological signals. This method allows for distinguishing significant and subtle differences in the regularity of the data. Although ApEn is a widely used estimator, it strongly depends on the length of the recorded data. The ApEn algorithm takes segments of length  $m$  samples along a time series of length  $N$ . Each segment serves as a template for comparison with other segments of length  $m$  from the time series. The similarity or closeness between this template segment and the remaining segments is calculated as the conditional probability that the signal is the same under the same similarity criteria for both  $m$  and  $m + 1$  points. This result is used as a measure of regularity [12]. In the process, a self-comparison occurs because each template vector is compared with itself to avoid evaluating the logarithmic function at zero. For this reason, ApEn generates an inconsistency in the measure of regularity that may be more noticeable when the temporal signal is short [13]. Although this bias could be eliminated if self-coincidences are excluded, this is not an alternative in ApEn.

Richman and Moorman [13] proposed the SampEn estimator as an improvement over the ApEn estimator. This improvement is based on the fact that SampEn does not generate self-comparison bias. The SampEn algorithm shows greater consistency and less dependence on the length of the data [13]. It is also more precise than ApEn, and this is essential in the study of electrophysiological time series. The calculation is expressed as  $\ln(A/B)$ , where  $A$  is the total number of matches in the next sample  $m + 1$ , and  $B$  is the total number of matches in the pattern of length  $m$ . The expression  $A/B$  corresponds to the conditional probability that two sequences within the same similarity criterion (tolerance  $r$ ) for  $m$  points remain within  $r$ , at the next point ( $m + 1$ ) [13].

In general, entropy algorithms take segments of  $m$  samples of the signal, called patterns, and compare them with the remaining signal (taking segments of the same length  $m$ ). The comparison is made by measuring the Euclidean distance between the standard and the signal samples. The pattern and the segment are similar when the distances are less than a threshold  $r$  (or tolerance parameter). Finally, the similarity probability is calculated. Low entropy values indicate that several similarities were found, that is to say, the signal is very regular. High entropy values indicate a few similarities, that is, the signal is irregular [12,13]. Entropy algorithms are insensitive to the amplitude and frequencies of the signal. However, they show a dependence on the length of the data. When the signal length is too short, it is difficult to obtain reliable entropy values [14]. In this regard, the parameter  $\tau$ , or time delay, corresponds to a time delay for subsampling the time series. Based on this parameter, the computation of ApEn and SampEn is performed from the first sample, and subsequently, each  $\tau$ -th sample.

The right choice of  $\tau$  helps with the analysis of the data. However, to find the best choice for  $\tau$  is not easy. Suppose  $\tau$  is small, and the acquisition time is also minimal. In that case, neighboring data are similar, with little variation, which does not provide much information to understand the system under study. Moreover, they generate a large amount of data for each second of recording, increasing the processing time and computational cost. If  $\tau$  is very large, the successive elements are already almost independent, the risk of loss of relevant information is high, and the processed data does not reliably represent the system's response under study. The first zero crossing of the signal's autocorrelation function and the mutual information are processing methods that help estimate a value of  $\tau$  that allows for a balance between these two extremes [5,6]. Commonly, a unit delay ( $\tau = 1$ ) is used. Kaffashi et al. [15], in 2008, considered the effect of the delay time  $\tau$  in the estimation of ApEn and SampEn. They reported that a  $\tau = 1$  is sufficient to generate good estimates of ApEn and SampEn for a time series whose autocorrelation function decays rapidly. However, for a time series where the autocorrelation function decays slowly, the entropy estimate can be confusing [15]. Other researchers showed that a higher  $\tau$  value would quantify the regularity of a long-range nonlinear time series [16,17].

The methods applied in the nonlinear analysis of physiological signals show great potential in obtaining quantitative information for medical interpretation, which is relevant for the diagnosis of diseases. However, prior treatment of the data with traditional methods, such as Fourier analysis [18], should be guaranteed. To ensure that the digitally acquired data effectively represents the physical signal, the sampling theorem, also known as the Nyquist theorem, must be fulfilled. This theorem prescribes a condition for the sampling frequency  $F_s$ .

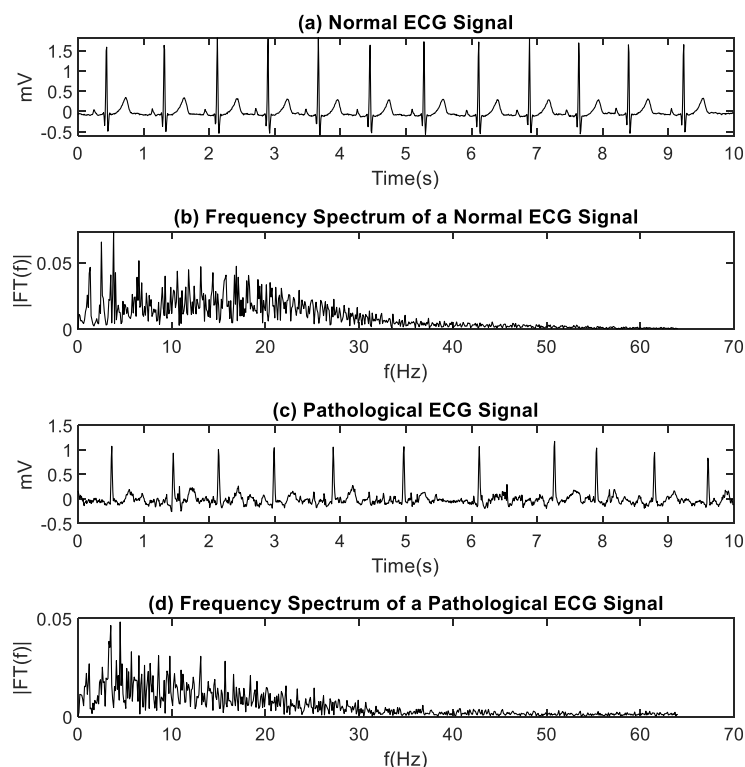
The parameters for calculating the ApEn and SampEn entropy (tolerance threshold  $r$  [19], immersion dimension ( $m$ ) [20], and time delay tau ( $\tau$ ) [15]) are described in the literature. However, little is said about the setting of these parameters for ECG and EEG signals acquired under different technologies. In particular, there are no recommendations to achieve more efficient estimations to save time, get stable entropy values, and reduce the computational cost while considering that not all technologies provide the same values of  $F_s$  and gain ( $G$ ). Entropy estimators with the same values of  $r$ ,  $m$ , and  $\tau$  (usually  $\tau = 1$ ) are often applied to signals from different sources and acquisition conditions. The following are some situations currently ignored in the literature: (i) Researchers can acquire ECG and EEG signals in the laboratory where the investigator can select  $F_s$  and the gain. (ii) Often, the researcher must use fixed parameters ( $F_s$  and  $G$ ) given by the laboratory's technology. (iii) The researcher must simply adapt to signals with different recording times,  $F_s$ , and also  $G$ , already recorded

or available in databases. It is necessary to find out whether the entropy value is the same for a signal with different acquisition characteristics (e.g., different  $F_s$ ) using the same parameters of  $r$ ,  $m$ , and  $\tau$ . Although the answer can be intuited, in this work, we propose a systematic study focused on the effects of  $\tau$  in the estimation of the regularity in normal and pathological ECG and EEG signals, using ApEn and SampEn. The study shows variations in entropy estimations and provides relevant information to researchers who use ECG or EEG entropy to classify patients as healthy or sick with a different  $F_s$ .

## 2. Materials and Methods

### 2.1. Databases for ECG and EEG

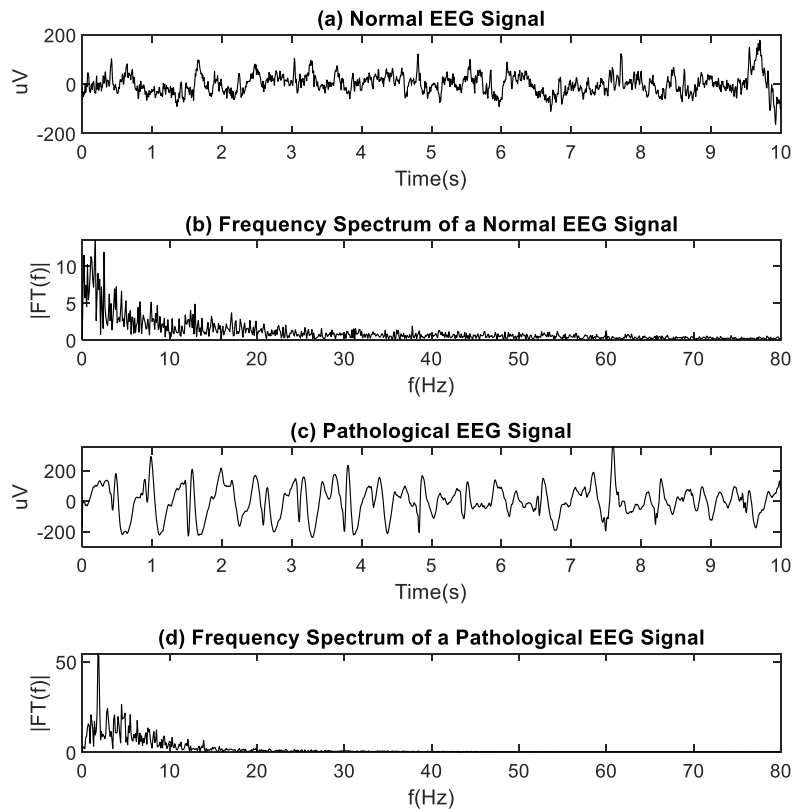
Normal and pathological human electrophysiological signals (ECG, EEG), obtained from PhysioNet [21], were used. Normal ECG signals (subjects corresponding to healthy controls) were obtained from the PAF Prediction Challenge Database [22], which contains sets of records. Each set stores two 30-min signals, sampled at 128 Hz. From there, 20 sets were downloaded, and the first 10 s of the first signal of the set were kept. Figure 1a shows a normal ECG signal with its respective frequency spectrum in Figure 1b, showing that the relevant frequency components are below 40 Hz. Pathological ECG signals were obtained from the AF Termination Challenge Database [23], which contains sets of records. Each set stores two ECG signals with atrial fibrillation. Each signal corresponds to a 1 min segment, sampled at 128 Hz. Twenty sets were downloaded, and 10 s of the first signal in the set were kept. Figure 1c,d shows a pathological ECG signal and its respective frequency spectrum.



**Figure 1.** (a) Normal electrocardiography (ECG) signal; (b) Normal ECG spectrum; (c) Pathological ECG. (d) Pathological ECG spectrum.

Normal EEG signals were obtained from the EEG Motor Movement/Imagery Dataset, which consists of more than 1500 EEG recordings between 1 and 2 min in duration. Signals were obtained from 109 volunteers, using a BCI2000 64-channel EEG system, with a sampling frequency of 160 Hz. EEG recordings correspond to several experimental conditions: 2 “baseline” or “baseline tests” of 1 min (eyes open, eyes closed) and 4 “tasks” or tests of 2 min (opening and closing the left or right

fist, imagining opening and closing the left or right fist, opening and closing both fists or both feet, imagining opening and closing both fists or both feet) in several repetitions [24]. From this database, 20 signals from healthy patients (eyes open) were downloaded, and 10 s of each bipolar signal generated by the difference between the Cz-Pz electrodes were kept. Figure 2a,b shows a normal EEG signal and its frequency spectrum, respectively.



**Figure 2.** (a) Normal electroencephalography (EEG); (b) Normal EEG spectrum; (c) Pathological EEG; (d) Pathological EEG spectrum.

Pathological EEG signals were obtained from the CHB-MIT Scalp EEG Database. The database contains electroencephalographic records of pediatric subjects with intractable seizures. Each case is made up of between 9 and 42 sets of records, sampled at 256 Hz. Each set contains 24–26 EEG signals [25]. Twenty recordings from different patients were downloaded, and 10 s of the bipolar Cz-Pz signal were kept. The selected electrodes (Cz-Pz) show epileptiform graphoelements present in epileptic seizures [26]. These electrodes were located on top of the brain area comprising the mid-parietal, frontal, and temporal-temporal regions. Figure 2c,d shows the pathological EEG signal and its spectrum, respectively. The characteristics of the ECG and EEG signals used in this study are summarized in Table 1.

**Table 1.** Characteristics of ECG and EEG signals used in this study.

Signals	No.	Databases	Time	$F_s$
Normal ECG	20	The PAF Prediction Challenge Database	10 s	128 Hz
Pathological ECG	20	The AF Termination Challenge Database	10 s	128 Hz
Normal EEG	20	EEG Motor Movement/Imagery Dataset	10 s	160 Hz
Pathological EEG	20	CHB-MIT Scalp EEG Database	10 s	256 Hz

## 2.2. Resampling of ECG and EEG Signals

The 40 ECG signals (20 normal and 20 pathological) initially had  $F_s = 128$  Hz. These signals were resampled to evaluate the effect of  $F_s$  in the estimation of entropy, using linear interpolation. Sampling frequency was increased to  $n * F_s$ , with  $n$  set to 1, 2, 3, and 4. The entropy values ApEn and SampEn were calculated by varying the parameter  $\tau$  from 1 to 5. Changes of  $F_s$  were made for each value of  $\tau$ . A total of 1600 datasets were processed, i.e., 2 (ApEn and SampEn)  $\times$  40 (normal ECG signals and pathological ECG signals)  $\times$  4 (different  $F_s$ )  $\times$  5 (different  $\tau$  values).

The 40 EEG signals (20 normal and 20 pathological) had different sampling frequencies. For this reason, before calculating ApEn and SampEn, pathological EEG signals with  $F_s = 256$  Hz (original) were downsampled to  $F_s = 160$  Hz to match the normal EEG recordings, which were sampled at 160 Hz [24]. Once the normal and pathological EEG signals had the same sampling frequency ( $F_s = 160$  Hz), the value of ApEn and SampEn was calculated, with the parameter  $\tau$  varied from 1 to 5. For each change in  $F_s$ , 5 changes were made for  $\tau$  for each EEG signal, without breaking Nyquist's theorem. A total of 1600 data were processed, i.e., 2 (ApEn and SampEn)  $\times$  40 (normal EEG signals and pathological EEG)  $\times$  4 (different  $F_s$ )  $\times$  5 (different  $\tau$  values).

## 2.3. Approximate Entropy (ApEn) and Sampled Entropy (SampEn)

ApEn and SampEn estimate the regularity of a time series. Their calculation depends on the conditional probability that two similar sequences, with a tolerance of  $r$  and of length  $m$ , remain similar in the next  $m + 1$  sample. ApEn and SampEn are smaller when there are many similar sequences and a few nonsimilar sequences. Low ApEn and SampEn values reflect high regularity. In this work, the following parameters were used to calculate the entropy: (immersion dimension)  $m = 2$  and (tolerance threshold)  $r = 0.2$  of the signal's standard deviation. These values coincide with those suggested by Pincus et al. [12] in their work on ApEn and Richman and Moorman [13] in their study on SampEn, among many other studies that used the same values. The ApEn and SampEn algorithms are shown below.

### 2.3.1. Algorithm ApEn

Step 1: For a time series of  $N$  samples  $\{u(n) : 1 \leq n \leq N\}$ , fix a value of  $m$  and  $\tau$  integers and  $r$ .

Step 2: Form the sequence of the vector  $X^m(1)$  to  $X^m(N - (m - 1)\tau)$  as

$$X^m(i) = \{u(n), u(n + 1\tau), \dots, u(n + (m - 1)\tau)\}, n = 1, \dots, N - (m - 1)\tau. \quad (1)$$

Step 3: Define the distance between  $X^m(i)$  and  $X^m(j)$  as:

$$d_{ij} = d[X^m(i), X^m(j)] = \max\{|u(i + k\tau) - u(j + k\tau)| : 0 \leq k \leq m - 1\} \quad (2)$$

Step 4: Estimate the probability  $C_i^m(r)$  as:

$$C_i^m(r) = (N - (m - 1)\tau)^{-1} \sum_{j=1}^{N-(m-1)\tau} \theta(d_{ij} - r), \quad (3)$$

where  $\theta$  is the Heaviside function defined as:  $\theta = \begin{cases} 1, & \text{si } d_{ij} \leq r \\ 0, & \text{si } d_{ij} > r \end{cases}$

Step 5: Calculate

$$\Phi^m(r) = (N - (m - 1)\tau)^{-1} \sum_{i=1}^{N-(m-1)\tau} \ln C_i^m(r) \quad (4)$$

Step 6: Increase  $m$  in 1, and find  $C_i^{m+1}(r)$  and  $\Phi^{m+1}(r)$

Step 7: For a finite sequence of length  $N$ , define

$$ApEn(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r). \quad (5)$$



### 2.3.2. Algorithm SampEn

Step 1: For a time series of  $N$  samples  $\{u(n) : 1 \leq n \leq N\}$ , fix a value of  $m$  and  $\tau$  integers and  $r$ .

Step 2: Form the sequence  $X^m(i) = \{u(n), u(n + \tau), \dots, u(n + (m - 1)\tau)\}$  for  $n = i = 1, \dots, N - (m - 1)\tau$ .

Step 3: Define the distance between  $X^m(i)$  and  $X^m(j)$  as

$$d_{ij} = d[X^m(i), X^m(j)] = \max\{|u(i + k) - u(j + k)| : 0 \leq k \leq (m - 1)\tau\} \quad (6)$$

Step 4: Define  $B_i^m$  as  $(N - (m + 1)\tau)^{-1}$  times the number of vectors  $X^m(j)$  within  $r$ , which remain close to  $X^m(i)$ , where  $j$  takes the rank of 1 to  $N - m\tau$ , and  $j \neq i$  to exclude the auto matches, then it is defined as

$$B^m = (N - m\tau)^{-1} \sum_{i=1}^{N-m\tau} B_i^m, \quad (7)$$

where  $B_i^m = (N - (m + 1)\tau)^{-1} \sum_{j=1, j \neq i}^{N-m\tau} \theta(r - d[X^m(i), X^m(j)])$

Step 5: Change the embedding dimension from  $m$  to  $m + 1$ , repeat from Step 2 to Step 4, then get  $A^m$ :

$$A^m = (N - m\tau)^{-1} \sum_{i=1}^{N-m\tau} A_i^m, \quad (8)$$

where  $A_i^m = (N - (m + 1)\tau)^{-1} \sum_{j=1, j \neq i}^{N-m\tau} \theta(r - d[X^{m+1}(i), X^{m+1}(j)])$

Step 6: Define

$$A^m = (N - m\tau)^{-1} \sum_{i=1}^{N-m\tau} A_i^m, \quad (9)$$

with  $B^m$  and  $A^m$  being the probability that 2 sequences coincide for  $m$  points and  $m + 1$  points, respectively.

### 2.4. Statistical Analysis of Entropy in ECG and EEG Signals

Results obtained with ApEn and SampEn were compared between normal and pathological signals (ECG, EEG) using parametric and nonparametric statistical tests. The t-student test was applied to compare the groups that presented normal distribution and equality of variances, and the Mann–Whitney U test for the groups that did not comply with the normality and variance test.

## 3. Results

In this section, we present graphs of the ApEn and SampEn entropy estimation for normal and pathological ECG signals and normal and pathological EEG signals for different values of  $F_s$  and  $\tau$ .  $F_s$  values start from a minimum value, complying with the Nyquist Theorem, and then increases in  $n * F_s$ , with  $n = 1, 2, 3$ , and 4. Values of  $\tau$  increased by one unit, from 1 to 5. Entropies of 40 ECG signals were estimated (20 normal and 20 pathological). Entropy values for normal ECG signals were concentrated in the low entropy values, confirming the ECG signal's regularity. Entropy values for pathological ECG signals were concentrated at high entropy values. The nonparametric Mann–Whitney U test was applied to compare normal and pathological ECG for each  $F_s$  and  $\tau$ . For the ApEn case, normality evaluation, using the Shapiro–Wilk test, showed that the group of ApEn values for normal-ECG signals did not have a normal distribution. In contrast, the group of ApEn values for pathological-ECG signals follows a normal distribution. Figure 3 shows the box plot for ApEn values for different  $F_s$  and  $\tau$ . The statistical test showed significant differences ( $p < 0.05$ ) between normal and pathological ECG for each  $F_s$  and  $\tau$ .

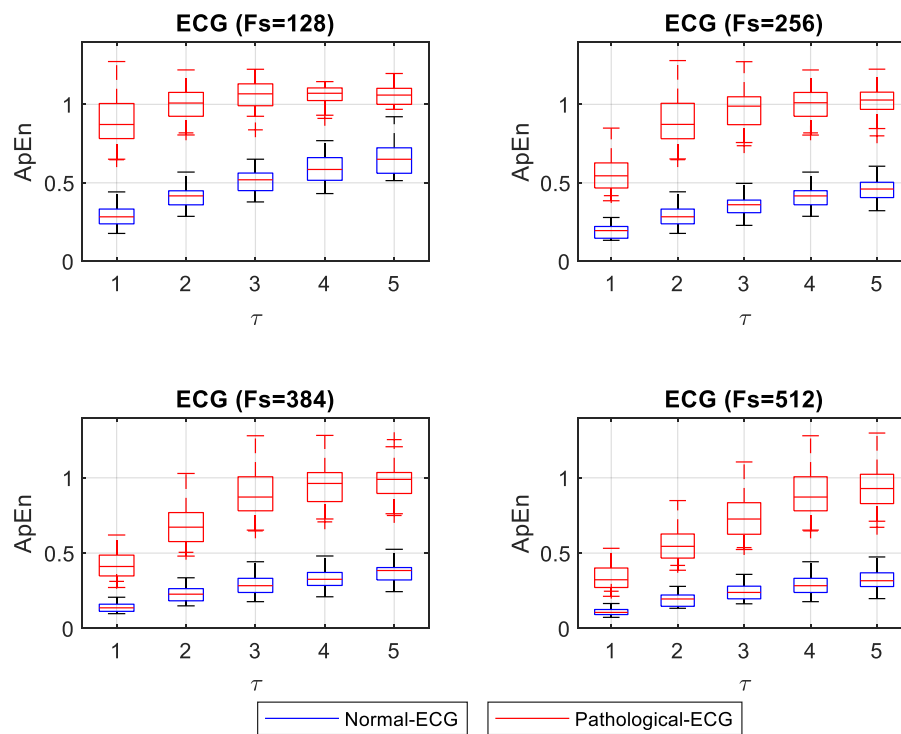


Figure 3. Box plot of ApEn for different  $F_s$  and  $\tau$  of normal and pathological ECG signals.

SampEn did not show a normal distribution for the normal ECG signals. In contrast, it did show a normal distribution for the pathological signals. Figure 4 shows the box plot of SampEn for different  $F_s$  and  $\tau$ . The statistical test showed significant differences ( $p < 0.05$ ) between normal and pathological ECG for each  $F_s$  and  $\tau$ .

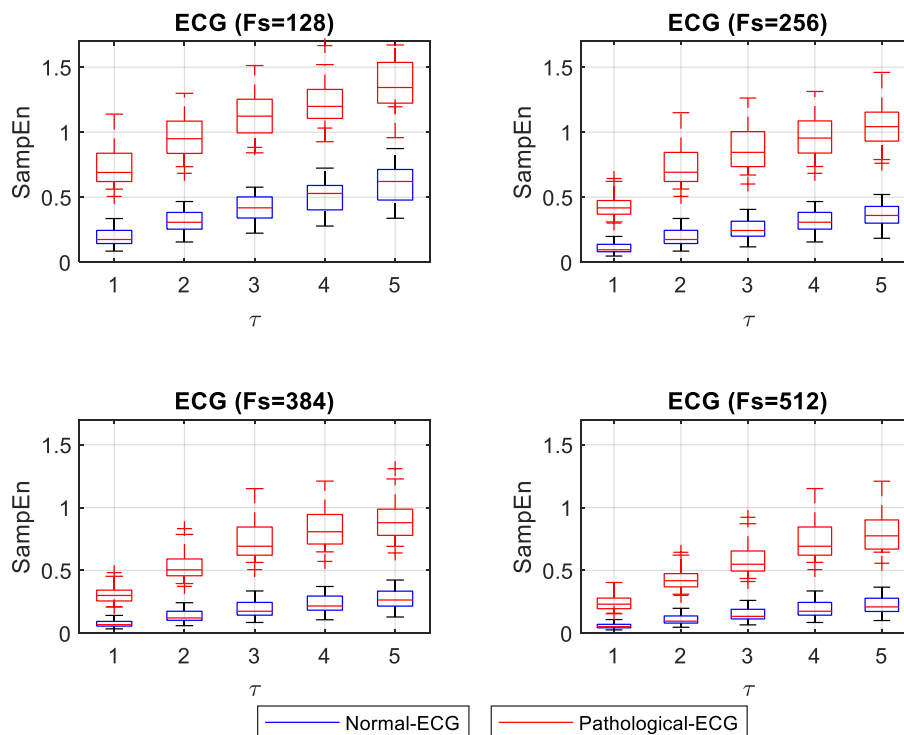
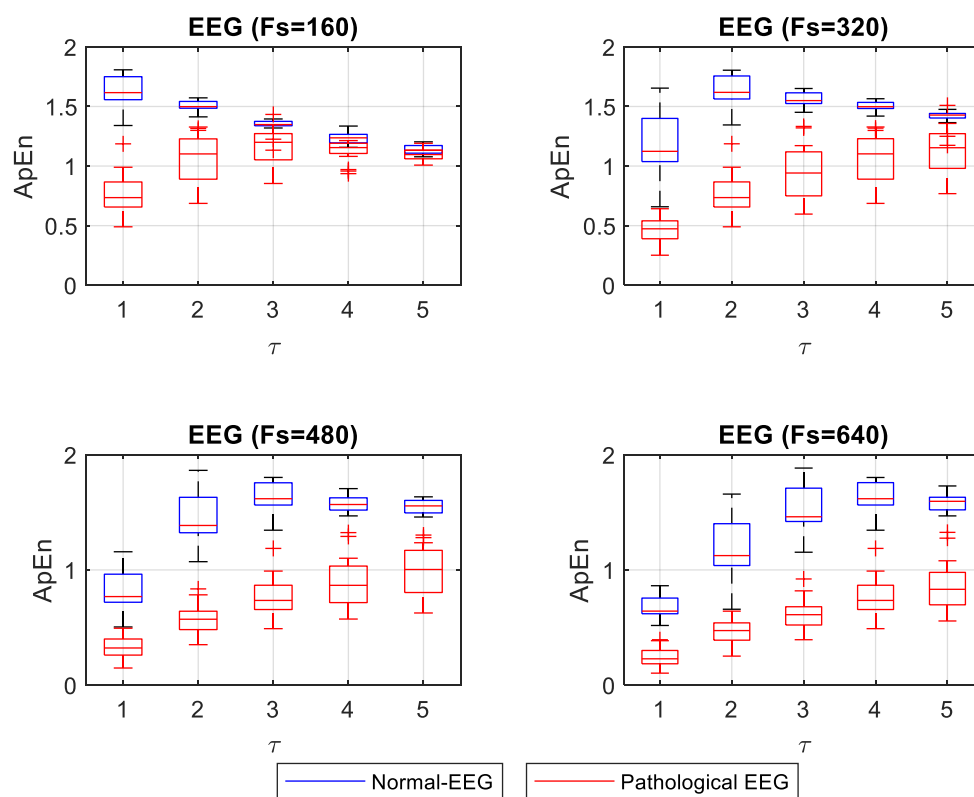


Figure 4. Box plot of SampEn for different  $F_s$  and  $\tau$  of normal and pathological ECG signals.



ApEn and SampEn entropy values were estimated for 40 EEG signals (20 normal and 20 pathological). Entropy values for normal EEG signals concentrated on high entropy values, showing that these signals are irregular. Entropy values for pathological EEG signals concentrated at low entropy values. A nonparametric Mann–Whitney U statistical test was applied to compare normal and pathological EEG for each  $F_s$  and  $\tau$ . For the ApEn case, normality evaluation, using the Shapiro–Wilk test, showed that 40% of the groups of ApEn values for normal-EEG signals followed a normal distribution. In comparison, 95% of the ApEn group of values for pathological-EEG signals showed a normal distribution. In the case where both the normal and pathological EEG ApEn value groups had a normal distribution, the Levene test was applied to evaluate the equality of the variances. The test showed differences in variances, confirming the result of the Mann–Whitney U test. Figure 5 shows the box plot for ApEn values for different  $F_s$  and  $\tau$ . The statistical test showed significant differences ( $p < 0.05$ ) between normal and pathological EEG for each  $F_s$  and  $\tau$ . However, for  $F_s = 160$  and  $\tau = 5$ , it had a  $p = 0.045$ , showing closeness between normal and pathological EEG.



**Figure 5.** Box plot of ApEn for different  $F_s$  and  $\tau$  of normal and pathological EEG signals.

For the case of SampEn for EEG signals, the Shapiro–Wilk test showed that 90% of the groups of SampEn values for normal-EEG signals followed a normal distribution, while all groups of SampEn values for pathological-EEG signals followed a normal distribution. For cases in which both the normal and pathological EEG SampEn values groups had a normal distribution, the Levene test was applied to evaluate the variances' equality. The test showed that 50% of the groups fulfilled the equality of variances. For these groups, the t-student test was applied. The remaining groups were tested with the Mann–Whitney U test. Figure 6 shows the box plot for SampEn values at different  $F_s$  and  $\tau$  values. The statistical test showed significant differences ( $p < 0.05$ ) between normal and pathological EEG in each  $F_s$  and  $\tau$ .

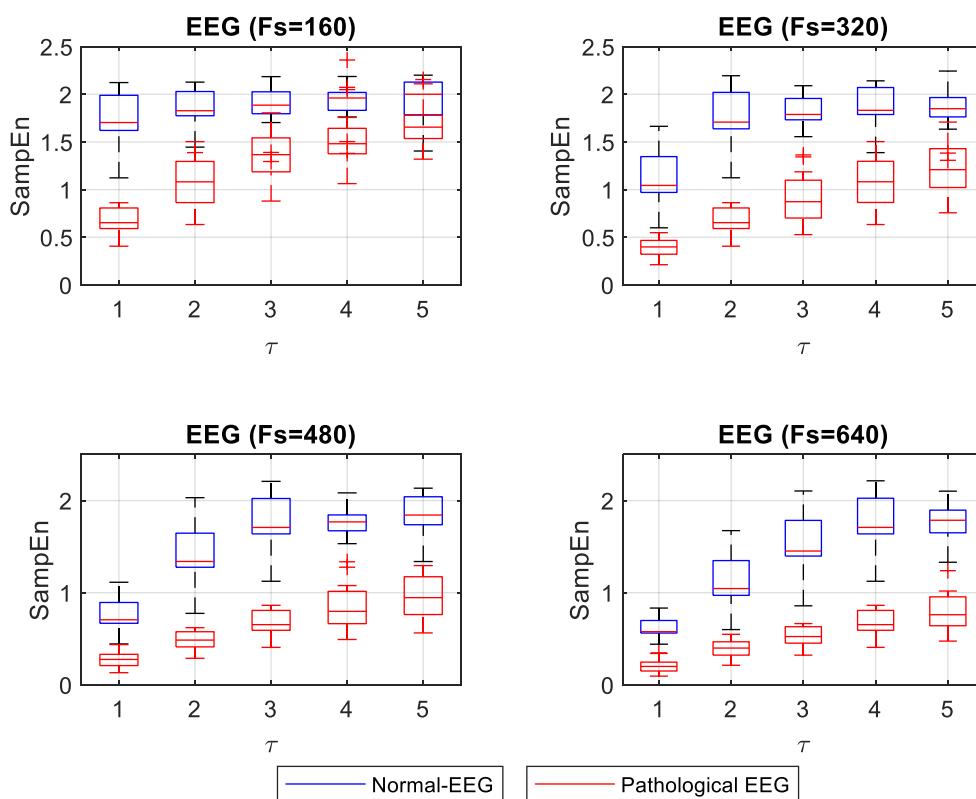


Figure 6. Box plot of SampEn for different  $F_s$  and  $\tau$  of normal and pathological EEG signals.

#### 4. Discussion

ECG and EEG signal analysis is prevalent in assessing the health condition of a patient. This prevalence explains why many researchers have focused on applying processing methods to extract characteristics that reveal information from the body's system under study. Processing methods for nonlinear signals, such as entropy, are useful to classify ECG and EEG signals as normal or pathological, based on their regularity. The algorithms to calculate ApEn and SampEn are widely used today, and the values for parameters  $r$ ,  $m$ , and  $\tau$  are widely discussed in the scientific literature. Different reports explain the appropriate values of  $r$ ,  $m$ , and  $\tau$ , assuming that an ECG or EEG signal's natural characteristics are maintained during the acquisition process. However, this is not entirely true because several factors may change during the signal acquisition, such as the gain, the sampling frequency, the noise level, among other factors. Sampling frequency  $F_s$  is a critical factor in obtaining appropriate signals representing the system's response under evaluation. A decreased  $F_s$  that is below the Nyquist Theorem's conditions can distort the signal and give false information from the source system. A high  $F_s$  may end in high-frequency noise induced by radio transmission interference, in addition to resulting in a large amount of data that could overwhelm the processing stage. Therefore, the values of the parameters  $r$ ,  $m$ , and  $\tau$  should be adjusted depending on the signal acquisition conditions. Subtle changes in the resulting entropy values between signals acquired from different sources could be of diagnostic value in the classification of these signals as normal or pathological.

A numerical experiment was designed to study changes in ApEn and SampEn values for ECG and EEG signals with different  $F_s$ , representing different acquisition systems. The experiment consisted of calculating ApEn and SampEn for normal and pathological ECG and EEG signals. Figures 3 and 4 show that normal ECG signals have low entropy values. This result means that they exhibit some regularity. In contrast, entropy values for pathological ECG signals have higher values. This result confirms the observations made in 2018 by Horie et al., where they examined SampEn values in ECG signals from patients with atrial fibrillation (AF) and patients without arrhythmia. In this work, they found that the SampEn values were higher for patients with AF than for patients without arrhythmia. Horie

used a data acquisition frequency of  $F_s = 200$  Hz, and the results of entropy SampEn were compared with  $\tau = 1$  and  $\tau = 5$ . In both cases, SampEn was different between patients with AF and without arrhythmia [27]. Based on these findings, it could be interpreted that the entropy values remained stable with changes in  $\tau$ . However, Figures 3 and 4 show that entropy values, both for normal and pathological ECG signals for  $F_s$  set to 256, slightly increased as  $\tau$  increased. Although normal and pathological conditions remain significantly different ( $p < 0.05$ ), with  $\tau = 1$  the entropy values between normal and pathological were closer than for other values of  $\tau$ . An adjustment of  $\tau$  in the calculation of entropy and  $F_s$  in data acquisition could be appropriate to guarantee a more stable distinction between signals from patients with AF and without arrhythmia.

The most substantial distinction between normal and pathological ECG signals was for the smallest  $F_s$  and with  $\tau = 1$ . With increasing values of  $\tau$ , ApEn values for normal ECG signals tend to rise, indicating that the signals are less regular. The opposite happens with higher  $F_s$ , where a  $\tau = 1$  shows that pathological ECG signals become more regular, approaching the normal ECG values. Values of  $\tau$  greater than 1 with higher  $F_s$  showed separation between the normal and pathological ECG. Moreover, they show more stable ApEn values. SampEn results showed a similar separation for all  $\tau$  values evaluated with the minimum  $F_s$ . SampEn values increased slightly with increasing values of  $\tau$ . However, for higher  $F_s$ , the SampEn values of normal and pathological signals were too close for low  $\tau$  values. As  $\tau$  increased, the distinction between normal and pathological ECG signals improved.

For the EEG signals, ApEn and SampEn entropy values were high for the normal condition and low for the pathological condition (see Figures 5 and 6). Having regular and irregular signals allows for observing the effect of different  $F_s$  and  $\tau$  values on the entropy capacity to distinguish the normal from pathological cases. These changes are not discussed in the literature thoroughly. Mesin in 2018 proposed to modify ApEn for estimating the regularity of EEG signals to correct the dependence on the length of the data; this work evidenced the convenience of increasing the  $F_s$  to obtain more reliable entropy values [17]. In this present study, we confirmed that increasing  $F_s$  improves entropy measurements when one wants to differentiate the normal from the pathological EEG signals. When ApEn is estimated with a lower  $F_s$ , it is convenient to use  $\tau = 1$ . The increases in  $\tau$  make the concentration of ApEn values (high and low) closer, thus making it very difficult to distinguish between normal and pathological signals ( $p = 0.045$  for a  $\tau = 5$  and an  $F_s = 160$ ). However, for higher  $F_s$  it may be more advisable to use a  $\tau$  between 1 and 3, where a better separation between the high and low values of ApEn is shown. In Figure 6, SampEn shows a more stable separation. The SampEn value of the pathological EEG signals increases slightly with increasing  $\tau$ . With a lower  $F_s$ , normal-EEG values remain almost stable, and a  $\tau = 1$  is advisable to make a significant separation between normal and pathological EEG. With increasing  $F_s$ , it is convenient to use a  $\tau$  greater than 1 if a more stable separation between regular and irregular signals is desired.

In summary, our results highlight the importance of appropriately selecting  $F_s$  and  $\tau$  to be able to differentiate robustly between the healthy and the pathological cases using the entropy of ECG and EEG signals. This entropy estimation allowed us to characterize signals with AF and epileptic patterns. These results suggest that this approach can also be applied to other cardiac and neurological pathologies.

## 5. Conclusions

The use of ApEn and SampEn in the analysis of nonlinear ECG and EEG signals have had an important reception in current research. According to their regularity, ECG or EEG signals can be classified as normal or pathological. This behavior makes entropy a quantity of interest for clinical uses. To compute ApEn and SampEn values, parameters  $r$  (tolerance threshold),  $m$  (immersion dimension), and  $\tau$  (time delay) are set beforehand. The first two ( $r$  and  $m$ ) are widely discussed in the literature, and their effects on the entropy value are known. However,  $\tau$  is also a factor that significantly affects the value of entropy. Many studies set  $\tau = 1$ , and in many others, its value is not discussed.

Parameters  $F_s$  and  $\tau$  are familiar. They are related to the electrophysiological signal's resampling, but they are taken into account at different times.  $F_s$  is considered during the acquisition of the signal. It is frequently not a value that the researcher can manipulate since it is generally fixed in the acquisition device. Furthermore, the value of  $F_s$  may be different from one device to the next. On the other hand,  $\tau$  is set during the ApEn or SampEn entropy calculation, and the researcher can purposely manipulate this parameter.

When the researcher sets the parameters  $r$ ,  $m$ , and  $\tau$ , it is common to standardize these parameters for signals from different sources (a laboratory or from a digital repository), assuming that the entropy value will not depend on the signal origin. For ECG and EEG signals, a stable entropy calculation must be ensured to provide reliable values. This is particularly important for clinical interpretation, where a patient is being classified as healthy or sick.

In this study, we show the effect of  $\tau$  and  $F_s$  on the computation of ApEn and SampEn. We show the convenience of adjusting values of  $\tau$  depending on the nature of the signal (ECG or EEG) and the acquisition characteristics, particularly the sampling frequency, to obtain more reliable regularity measures able to distinguish between normal and pathological signals.

This study's most important observations are as follows. SampEn and ApEn can be used to significantly differentiate between regular signals (normal-ECG and pathological-EEG) and irregular ones (pathological-ECG and normal-EEG) by varying  $\tau$  and  $F_s$ . However, for ApEn, with a minimum  $F_s$ , a value of  $\tau = 1$  is advisable when it is desirable to distinguish between normal and pathological ECG. In contrast, with a higher  $F_s$ , the use of  $\tau$  greater than 1 is advisable because it maintains a distinction between normal and pathological ECG while reducing the data processing burden. SampEn for the analysis of ECG signals showed a slight increase in entropy values with the increasing values of  $F_s$  and  $\tau$ . The distance between the SampEn values of normal and pathological ECG signals was similar to the minimum  $F_s$  for all values of  $\tau$ . The optimal parameters to achieve a large separation, observed between normal and pathological ECG entropy values, were obtained using SampEn with a low  $F_s$  (128 or 256 Hz) and a large value of  $\tau$  ( $\tau = 5$  or  $\tau = 4$ ).

In the analysis of ApEn and SampEn measurements for normal and pathological EEG, a value of  $\tau = 1$  is convenient when using a minimum  $F_s$ . An increase of  $\tau$  in this condition (low  $F_s$ ) degrades the entropy value, bringing the normal and pathological EEG measurements closer together. This makes the distinction difficult. With increasing values of  $F_s$ , it is advisable to use a value of  $\tau$  greater than 1 to obtain a clear separation of normal and pathological EEG. However, a high value of  $\tau$  (4 or 5) could corrupt the values, especially for the regular signals (pathological-EEG). The optimal parameters to achieve a large separation, observed between normal and pathological EEG entropy values, were obtained using SampEn with a high value of  $F_s$  (480 or 640 Hz) and a large value of  $\tau$  (3, 4, or 5).

It is of great value to recognize the appropriate adjustments of parameter  $\tau$  in calculating ECG and EEG signals' entropy to have consistent regularity measures while reducing processing cost, especially for long-duration recordings.

In this study, we analyzed the separation between entropy values of normal and pathological signals (AF for the ECG case and epileptic seizures for the EEG case). However, for other cardiac or neurological pathologies, the separation between normal and pathological may depend on the patient's conditions. This separation is a characteristic to be explored, which may improve the clinical identification of groups of patients with different conditions.

**Author Contributions:** Conceptualization, R.E. and J.T.; methodology, J.T.; software, R.E. and J.T.; validation, R.E., A.W. and J.T.; formal analysis, R.E., A.W. and J.T.; investigation, R.E. and J.T.; writing—original draft preparation, R.E., A.W. and J.T.; writing—review and editing, A.W.; visualization, R.E., A.W. and J.T.; supervision, R.E.; project administration, R.E.; funding acquisition, R.E. and A.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** A.W. was supported by grant BASAL FB0008 and R.E. was supported by Universidad ECCI.

**Acknowledgments:** A.W. was supported by grant BASAL FB0008 and R.E. was supported by Universidad ECCI.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

- Jianwei, Z.; Chu, H.; Daniele, S.; Zhang, J.; Magdi, Y.S.; El-Askary, H.; Guohua, F. Optimal Multi-Stage Arrhythmia Classification Approach. *Sci. Rep.* **2020**, *10*, 2898. [[CrossRef](#)]
- Anczykowski, J.; Willems, S.; Hoffmann, B.A.; Meinertz, T.; Blankenberg, S.; Patten, M. Early Detection of Symptomatic Paroxysmal Cardiac Arrhythmias by Trans-Telephonic ECG Monitoring: Impact on Diagnosis and Treatment of Atrial Fibrillation. *J. Cardiovasc. Electrophysiol.* **2016**, *27*, 1032–1037. [[CrossRef](#)] [[PubMed](#)]
- Niedermeyer, E.; Lopes da Silva, F.H. *Electroencephalography: Basic Principles, Clinical Applications, and Related Fields*, 5th ed.; Lippincott Williams & Wilkins: Philadelphia, PA, USA, 2005; ISBN 0-7817-5126-8.
- Zhang, X.; Zhou, P. Sample entropy analysis of surface EMG for improved muscle activity onset detection against spurious background spikes. *J. Electromyogr. Kinesiol.* **2012**, *22*, 901–907. [[CrossRef](#)] [[PubMed](#)]
- Kantz, H.; Schreiber, T. *Nonlinear Time Series Analysis*, 2nd ed.; Cambridge University Press: Cambridge, UK, 2003. [[CrossRef](#)]
- Abarbanel, H.D.I.; Gollub, J.P. Analysis of observed chaotic data. *Phys. Today* **1996**, *49*, 86. [[CrossRef](#)]
- Lehnertz, K. Epilepsy and non-linear dynamics. *J. Biol. Phys.* **2008**, *34*, 253–266. [[CrossRef](#)]
- Vijith, V.S.; Jacob, J.E.; Iype, T.; Gopakumar, K.; Yohannan, D.G. Epileptic seizure detection using non linear analysis of EEG. In Proceedings of the 2016 International Conference on Inventive Computation Technologies (ICICT), Coimbatore, India, 26–27 August 2016; pp. 1–6. [[CrossRef](#)]
- Owis, M.I.; Abou-Zied, A.H.; Youssef, A.-B.M.; Kadah, Y.M. Study of features based on non-linear dynamical modeling in ECG arrhythmia detection and classification. *IEEE Trans. Biomed. Eng.* **2002**, *49*, 733–736. [[CrossRef](#)]
- Gao, J.; Hu, J.; Tung, W. Entropy measures for biological signal analyses. *Nonlinear Dyn.* **2012**, *68*, 431–444. [[CrossRef](#)]
- Pincus, S.M. Approximate entropy as a measure of system complexity. *Proc. Natl. Acad. Sci. USA* **1991**, *88*, 2297–2301. [[CrossRef](#)]
- Pincus, S.M.; Gladstone, I.M.; Ehrenkranz, R.A. A regularity statistic for medical data analysis. *J. Clin. Monitor. Comput.* **1991**, *7*, 335–345. [[CrossRef](#)]
- Richman, J.S.; Moorman, J.R. Physiological time-series analysis using approximate entropy and sample entropy. *Am. J. Physiol. Heart Circ. Physiol.* **2000**, *278*, H2039–H2049. [[CrossRef](#)]
- Yentes, J.M.; Hunt, N.; Schmid, K.K.; Kaipust, J.P.; McGrath, D.; Stergiou, N. The appropriate use of approximate entropy and sample entropy with short data sets. *Ann. Biomed. Eng.* **2013**, *41*, 349–365. [[CrossRef](#)] [[PubMed](#)]
- Kaffashi, F.; Foglyano, R.; Wilson, C.G.; Loparo, K.A. The effect of time delay on Approximate & Sample Entropy calculations. *Phys. D* **2008**, *237*, 3069–3074. [[CrossRef](#)]
- Montesinos, L.; Castaldo, R.; Pecchia, L. On the use of approximate entropy and sample entropy with centre of pressure time-series. *J. Neuroeng. Rehabil.* **2018**, *15*, 1–15. [[CrossRef](#)] [[PubMed](#)]
- Mesin, L. Estimation of Complexity of Sampled Biomedical Continuous Time Signals Using Approximate Entropy. *Front. Physiol.* **2018**, *9*, 1–15. [[CrossRef](#)]
- Sarnthein, J.; Abarbanel, H.D.; Pockberger, H. Non-linear analysis of epileptic activity in rabbit neocortex. *Biol. Cybern* **1998**, *78*, 37–44. [[CrossRef](#)] [[PubMed](#)]
- Castiglioni, P.; Di Rienzo, M. How the Threshold “R” Influences Approximate Entropy Analysis of Heart-Rate Variability. *Comput. Cardiol.* **2008**, *35*, 561–564. [[CrossRef](#)]
- Cao, L. Practical method for determining the minimum embedding dimension of a scalar time series. *Phys. D* **1997**, *110*, 43–50. [[CrossRef](#)]
- Goldberger, A.L.; Amaral, L.A.; Glass, L.; Hausdorff, J.M.; Ivanov, P.C.; Mark, R.G.; Mietus, J.E.; Moody, G.B.; Peng, C.K.; Stanley, H.E. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. *Circulation* **2000**, *101*, e215–e220. [[CrossRef](#)]
- Moody, G.B.; Goldberger, A.L.; McClennen, S.; Swiryn, S.P. Predicting the Onset of Paroxysmal Atrial Fibrillation: The Computers in Cardiology Challenge 2001. *Comput. Cardiol.* **2001**, *28*, 113–116. [[CrossRef](#)]
- Moody, G.B. Spontaneous Termination of Atrial Fibrillation: A Challenge from PhysioNet and Computers in Cardiology 2004. *Comput. Cardiol.* **2004**, *31*, 101–104.
- Schalk, G.; McFarland, D.J.; Hinterberger, T.; Birbaumer, N.; Wolpaw, J.R. BCI2000: A general-purpose brain-computer interface (BCI) system. *IEEE Trans. Biomed. Eng.* **2004**, *51*, 1034–1043. [[CrossRef](#)] [[PubMed](#)]

25. Shoeb, A.H. Application of Machine Learning to Epileptic Seizure Onset Detection and Treatment. Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, USA, 2009.
26. Hampel, K.G.; Garcés-Sánchez, M.; Gómez-Ibáñez, A.; Palanca-Cámara, M.; Villanueva, V. Diagnostic challenges in epilepsy. *Rev. Neurol.* **2019**, *68*, 255–263. [[CrossRef](#)] [[PubMed](#)]
27. Horie, T.; Burioka, N.; Amisaki, T.; Shimizu, E. Sample entropy in electrocardiogram during atrial fibrillation. *Yonago Acta Med.* **2018**, *61*, 49–57. [[CrossRef](#)] [[PubMed](#)]

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).