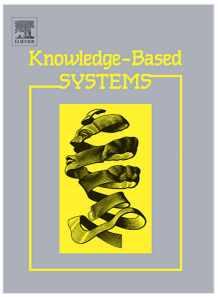
Journal Pre-proof



Ensemble entropy: A low bias approach for data analysis

Hamed Azami, Saeid Sanei, Tarek K. Rajji

PII: S0950-7051(22)00969-8
DOI: https://doi.org/10.1016/j.knosys.2022.109876
Reference: KNOSYS 109876

To appear in: *Knowledge-Based Systems*

Received date : 18 May 2022
Revised date : 27 August 2022
Accepted date : 4 September 2022

Please cite this article as: H. Azami, S. Sanei and T.K. Rajji, Ensemble entropy: A low bias approach for data analysis, *Knowledge-Based Systems* (2022), doi: https://doi.org/10.1016/j.knosys.2022.109876.

Ensemble Entropy: A Low Bias Approach for Data Analysis

Hamed Azami[a], Saeid Sanei[b], Tarek K. Rajji[a]

[a]*Centre for Addiction and Mental Health, Toronto Dementia Research Alliance, University of Toronto, Toronto, ON, Canada., Toronto, ON, Canada.*
[b]*School of Science and Technology, Nottingham Trent University, Nottingham, UK.*

Abstract

To quantify the data irregularity of data, there are a number of entropy measures each with its own advantages and disadvantages. In this pilot study, a new concept, namely ensemble entropy, is introduced and used to generate more stable and low bias signal patterns for entropy estimation. We propose ensemble versions of sample entropy (SampEn), permutation entropy, dispersion entropy (DispEn), fluctuation DispEn (FDispEn) based on the combination of different parameters initialization for a original entropy method. Also, ensemble Shannon and conditional entropy methods based on the entropy values obtained by different entropy algorithms. We applied the techniques to different synthetic and three biomedical datasets to investigate the behaviour of the ensemble methods on the changes in the data dynamics. The results suggest that ensemble approaches are able to distinguish different kinds of noises and the degrees of randomness in our generated MIX process. Ensemble SampEn, unlike SampEn, does not result in undefined values for short signals. Ensemble DispEn needs a smaller number of samples for distinguishing different kinds of noise. The majority of ensemble methods result in larger differences between younger and older subjects using their RR intervals as well as healthy young vs. elderly children using their walking stride interval data based on Hedges' g effect size. The ensemble algorithms lead to more stable results (lower coefficients of variations) for the synthetic data (different kinds of noises and mixed processes) and discriminated different types of physiological signals better than their corresponding original entropy approaches. The Matlab code used in this paper will be available at https://github.com/HamedAzami/ upon publication.

Keywords: Nonlinear analysis, Ensemble technique, Sample entropy,

Permutation entropy, Dispersion entropy.
PACS: 0000, 1111
2000 MSC: 0000, 1111

1. Introduction

Entropy techniques rooted in information theory are of great interest for evaluation of irregularity and uncertainty in the data [1]. Shannon entropy (ShEn) and conditional entropy (CondEn) are the most common concepts used in the context of entropy analysis in particular for physiological signals [2, 3, 4, 1]. ShEn and CondEn respectively show the amount of information learned and the rate of information production in a system [4, 3]. Based on these two approaches, various methods, such as approximate entropy (ApEn) [5], sample entropy (SampEn) [6], corrected conditional entropy [7], fuzzy entropy (FuzEn) [8], permutation entropy (PerEn) [9], dispersion entropy (DispEn) [10], and fluctuation DispEn (FDispEn) [11] have been proposed.

These entropy metrics have been used in a wide variety of physiological and non-physiological signal analysis applications to characterize various pathological or disordered states. For example, ApEn and SampEn have been used to analyze electroencephalogram (EEG) and magnetoencephalogram (MEG) in Alzheimer's disease (AD) [12]. SampEn has been used to study the heart rate variability during episodes of mechanical ventilation and acute anoxia in rats [13] as well as for diseases and aging [14]. FuzEn has been applied to a gait maturation database to distinguish the effect of age on intrinsic stride-to-stride dynamics and also the Fantasia database to distinguish short RR interval signals recorded from healthy young vs. elderly subjects [15]. PerEn has been applied in epilepsy research [16, 17] and in anesthesiology [18, 19] using EEGs. PerEn was also used to study financial time series [20]. DispEn was used to diagnose breathing and movement-related sleep disorders using ECGs and electromyograms (EMGs) [21]. DisEn is also used for the detection of different gear faults, fault diagnosis of rolling element bearings, and characterization of bearing degradation [22]. DispEn and FDispEn have been used to help clinicians in diagnosing AD and mild cognitive impairment using MEG signals [23].

Each of these methods has its own advantages and disadvantages based on their algorithms and applications [24]. For example, SampEn is either undefined or unreliable for short signals and computationally expensive for

real-time applications. SampEn is also sensitive to the changes in its parameters, especially to the tolerance factor r. FuzEn alleviates the problem of undefined values of SampEn but is still unreliable for short signals. Additionally, FuzEn, compared with SampEn, is less sensitive to the changes in its parameters and even data length. PerEn can be used for both short and long signals. However, when a time series is symbolized based on the permutation patterns (Bandt-Pompe procedure), only the order of amplitudes is considered and some information related to the actual amplitude values may be ignored [25, 11]. Additionally, PerEn is sensitive to noise (even when the data signal-to-noise ratio of is high), because a small change in amplitude value may vary the order relations among the amplitudes [11]. Although DispEn and FDispEn are not sensitive to noise and also do not result in undefined values, both, like PerEn, are based on symbolic dynamics or patterns originated from a coarse-graining of the measurements. That is, the data are transformed into a new signal with only a few different elements. Therefore, the study of time series dynamics is simplified to the study of distribution of symbol sequences. By doing this, although some of the invariant properties of the dynamics are maintained, some details may be lost [26, 27, 28]. Therefore, there is no entropy method capable of correctly finding the underlying structure for all datasets.

In general, different entropy values obtained by different entropy approaches (models) can be equally plausible, if there is no previous knowledge about the best way to evaluate the data dynamics. Therefore, the idea of combining different entropy approaches (ensemble or aggregation) can be considered as an alternative approach for enhancing the quality of an entropy estimation obtained by entropy algorithms.

An ensemble technique is a machine learning system that is constructed with a set of models working in parallel, whose outputs are combined with a decision fusion strategy to provide a better and unbiased result [29]. The models can be for classification (supervised), clustering (unsupervised), regression (supervised), prediction (supervised), or feature extraction (unsupervised or supervised) approaches, depending on the type of task [30]. The reason behind using an ensemble method is that no original model can be perfectly developed for solving non-trivial real-world problems. This rationale is similar to the reason we use such mechanisms in our daily lives (to enhance our confidence that we are making the right decision, by weighing various opinions, and combining them through some thought process to reach a final decision) [31]. Ensemble approaches have been successfully used in recent

decades and used widely in various supervised and unsupervised real-world applications such as object detection and tracking, signal analysis, image recognition, information retrieval, bioinformatics, and data mining [32, 31].

In this study, we develop unsupervised ensemble SampEn, PerEn, DispEn, FDispEn as well as ensemble ShEn and CondEn methods. The proposed approaches are evaluated using synthetic datasets and real-world physiological signals: focal and non-focal EEGs [33], RR intervals in Fantasia database [34], and walking stride measurements [35].

2. Ensemble Entropy

Given a signal, an entropy ensemble method includes two main steps: 1) Generation: creation of a set of patterns of a time series; and 2) Consensus Function: integration or combination of all the patterns or values obtained in the generation step. In the generation step, we can create or detect different kinds of patterns based on different entropy algorithms or different parameters initialization of an algorithm. The consensus function can be done based on maximum possibility or averaging,

The goal of ensemble entropy framework is having a nonlinear approach with the following characteristics: 1) Robustness: better average performance compared to original entropy algorithms. 2) Novelty: finding a new combined solution unattainable by any original ensemble algorithm. 3) Stability: ensemble models with lower sensitivity to noise and outlier.

In this section, we propose ensemble SampEn, PerEn, DispEn, and FDispEn based on the combination of different parameters initialization for a original entropy method and ensemble ShEn and CondEn based on the entropy values obtained by different entropy algorithms.

2.1. Ensemble Sample Entropy

One of the most important parameters of SampEn is the tolerance factor r. The parameter r is chosen to balance the quality of the logarithmic likelihood estimates with the loss of data's information. When r is too small (e.g., smaller than 0.1 of the standard deviation of a data), poor conditional probability estimates are achieved. This may also lead to undefined or unreliable SampEn values, especially for short signals. Furthermore, to avoid the effect of noise on data, larger r is recommended. In contrast, for a large r value (e.g., 0.3) of the standard deviation, very detailed data information is

lost. The ensemble SampEn (EnsSE) algorithm is based on using different r values so we can benefit from small and large r values at the same time.

The computation of EnsSE for a univariate signal of length N: $\mathbf{x} = \{x_1, x_2, \ldots, x_N\}$, defined on a domain $\mathbb{R}$, starts with the creation of template vectors $\mathbf{x}_i^m$ $(i = 1, 2, ..., N - (m-1))$ as:

$$\mathbf{x}_i^m = \{x_i, x_{i+1}, \ldots, x_{i+m-1}\}, \tag{1}$$

where m is the embedding dimension [6].

Then, for each r_j $(j = 1, 2, ..., NM)$, where NM refers to the number of models, the vector pairs in the template vectors of length m having $d[\mathbf{x}_i^m, \mathbf{x}_a^m] \leq r_j$ $(1 \leq a \leq N - m, a \neq i)$ are averaged as:

$$\phi_i^m(r_j) = \frac{[\#\ of\ \mathbf{x}_a^m \mid d[\ \mathbf{x}_i^m, \mathbf{x}_a^m] \leq r_j]}{N - m - 1}, \tag{2}$$

where # refers to cardinality and $d[\mathbf{x}_i^m, \mathbf{x}_a^m]$ denotes the greatest element of the absolute differences between $\mathbf{x}_i^m$ and $\mathbf{x}_a^m$ [6]. Then, $\phi^m(r_j)$ is estimated as follows:

$$\phi^m(r_j) = \frac{1}{N - m} \sum_{i=1}^{N-m} \phi_i^m(r_j). \tag{3}$$

Next, the average of $\phi^m(r_j), (j = 1, 2, ..., NM)$ is estimated as $\phi^m(\mathbf{r})$. After this, the dimension is increased to $m+1$ and $\phi^{m+1}(\mathbf{r})$ is estimated like $\phi^m(r)$ [6].

Finally, EnsSE is defined as:

$$EnsSE(\mathbf{x}, m, \mathbf{r}) = -\ln \frac{\phi^{m+1}(\mathbf{r})}{\phi^m(\mathbf{r})}. \tag{4}$$

In SampEn, only the number of matches whose differences are smaller than the tolerance r is counted and therefore, SampEn is sensitive to this parameter to a great extent. However, in EnsSE, the tolerance value of r is substituted by the tolerance vector $\mathbf{r}$. That is, when the number of matches whose differences are smaller than each of r_j $(j = 1, 2, ..., NM)$ is considered. In this way, small changes in amplitude (larger r - corresponding to lower frequency components in a signal) and large changse in amplitude (smaller r - corresponding to higher frequency components in a signal) are taken into account in EnsSE.

2.2. Ensemble Permutation Entropy

The value of embedding dimension plays the main role in characterizing data based on PerEn. In order to work with reliable statistics for PerEn, it is highly recommended to have $(m+1)! \leq N$ [25]. Thus, all of the m values (providing $(m+1)! \leq N$) can be chosen. To be less sensitive to the value of embedding dimension, Ensemble PerEn (EnsPE) is defined based on all of the m values (providing $(m+1)! \leq N$). For the time series $\mathbf{x} = \{x_1, \ldots, x_N\}$, the algorithm of EnsPE is defined as follows.

First, m_{max} is calculated as the maximum value of m where $(m_{max}+1)! \leq N$. For each m_j in the embedding dimension vector $\mathbf{m} = \{2, 3, \ldots, m_j \ldots, m_{max}\}$, we embed the signal $\mathbf{x}$ in an m_j-dimensional space to obtain the reconstruction vectors $\mathbf{x}_\Lambda^{m_j} = \{x_\Lambda, x_{\Lambda+1}, \ldots, x_{\Lambda+m_j-2}, x_{\Lambda+m_j-1}\}$ for $\Lambda = 1, 2, \ldots, N - m_j - 1$, where m_j denotes the embedding dimension. Next, the elements of $\mathbf{x}_\Lambda^{m_j}$ are arranged in an increasing order, with integer indices from 0 to $m_j - 1$, as follows:

$$\{x_{\Lambda+(\aleph_1-1)}, x_{\Lambda+(\aleph_2-1)}, \ldots, x_{\Lambda+(\aleph_{m_j-1}-1)}, x_{\Lambda+(\aleph_{m_j}-1}\} \tag{5}$$

where $\aleph_*$ is the (time) index of the element in the reconstruction vector. There are $m_j!$ potential ordinal patterns or symbol sequences η_t $(1 \leq t \leq m_j!)$, termed "motifs". Then, the occurrence of each of the order patterns η_t denoted as $f(\eta_t)$ is counted. For each η_t, the relative frequency $Pr(\eta_t)$ is estimated as follows:

$$Pr(\eta_t) = \frac{f(\eta_t)}{N - (m_j - 1)}. \tag{6}$$

For each m_j, the normalized PerEn (NPerEn) value is computed as follows [9]:

$$NPerEn(\mathbf{x}, m_j) = \frac{-\sum_{t=1}^{m_j!} Pr(\eta_t) \cdot \ln Pr(\eta_t)}{\ln(m_j!)}, \tag{7}$$

where ln denotes the natural logarithm [25, 9]. Finally, the EnsPE is estimated as follows:

$$EnsPE(\mathbf{x}, m) = \frac{\sum_{j=2}^{m_{max}} NPerEn(\mathbf{x}, m_j)}{m_{max} - 1}. \tag{8}$$

2.3. Ensemble (fluctuation) Dispersion Entropy

In the first form of DispEn, normalized cumulative distribution function (NCDF) is used to map the original signal to discrete classes [10]. Nevertheless, there are four other methods (linear, tansig, logsig, and sorting mapping approaches) to map the original signal to c discrete classes each with its own advantages and disadvantages [11]. To benefit from the advantages of each of them and have a low bias entropy approach, we propose to use all of the mapping approaches in ensemble DispEn (EnsDE).

Ensemble DispEn (EnsDE) is computed as follows:

1. The samples in $\mathbf{x}$ are mapped to c discrete classes, which can be denoted with integers ranging from 1 to c. In this step, we use all the five mapping approaches (linear, NCDF, tansig, logsig, and sorting mapping approaches). The transformed values are assigned into c bins of equal size depending on their amplitude levels after the transformation based on these five mapping approaches. This results in a temporal sequence of symbols $\eta^c = \{\eta_1^c, \ldots, \eta_N^c\}$, $\xi^c = \{\xi_1^c, \ldots, \xi_N^c\}$, $\kappa^c = \{\kappa_1^c, \ldots, \kappa_N^c\}$, $\lambda^c = \{\lambda_1^c, \ldots, \lambda_N^c\}$, and $\omega^c = \{\omega_1^c, \ldots, \omega_N^c\}$ for the linear, NCDF, tansig, logsig, and sorting mapping approaches, respectively [11].
2. Using the coarse-grained sequence η^c in the embedding dimension m and number of classes c, the *dispersion patterns* $\eta_i^{m,c}$ are formed as:
$$\eta_i^{m,c} = \{\eta_i^c, \eta_{i+1}^c, \ldots, \eta_{i+(m-1)}^c\}, i = 1, 2, \ldots, N-(m-1). \quad (9)$$
Similarly, $\xi_i^{m,c}, \kappa_i^{m,c}, \lambda_i^{m,c}$, and $\omega_i^{m,c}$ are estimated.
3. Each vector $\eta_i^{m,c}$ is mapped to a dispersion pattern $\pi_{\alpha_0\alpha_1\ldots\alpha_{m-1}}$, where $\eta_i^c = \alpha_0$, $\eta_{i+1}^c = \alpha_1$,..., $\eta_{i+(m-1)}^c = \alpha_{m-1}$. The number of possible dispersion patterns assigned to each vector $\eta_i^{m,c}$ is equal to c^m, since the vector has m elements and each member can be one of the integers from 1 to c [10]. Similarly, each vector $\xi_i^{m,c}$, $\kappa_i^{m,c}$, $\lambda_i^{m,c}$, and $\omega_i^{m,c}$, is mapped to a dispersion pattern $\pi_{\alpha_0\alpha_1\ldots\alpha_{m-1}}$.
4. For each potential dispersion pattern $\pi_{\alpha_0\alpha_1\ldots\alpha_{m-1}}$, its relative frequency of appearance, $p(\pi_{\alpha_0\alpha_1\ldots\alpha_{m-1}})$, is obtained by counting the number of sequences with that pattern in all η, ξ, κ, λ, and ω and dividing it by the total number of patterns extracted from the signal. If $p(\phi_{y_0\ldots y_m})$ denotes the relative frequency of dispersion pattern $\pi_{\alpha_0\alpha_1\ldots\alpha_{m-1}}$, we have:
$$p(\pi_{\alpha_0\alpha_1\ldots\alpha_{m-1}}) = \frac{\#\text{ of } i,\ \text{such that } Q_i^{m,c} \text{ has type } \pi_{\alpha_0\alpha_1\ldots\alpha_{m-1}}}{5(N-(m-1))} \quad (10)$$

where $Q_i^{m,c} = \{\alpha_i^{m,c}, \xi_i^{m,c}, \kappa_i^{m,c}, \lambda_i^{m,c}, \omega_i^{m,c}\}$. Note that $5(N-(m-1))$ shows the total number of dispersion patterns in $Q_i^{m,c}$.

5. Finally, based on Shannon entropy definition, the EnsDE value of $\mathbf{x}$ is estimated as follows:

$$EnsDE(\mathbf{x}, m, c) = -\frac{1}{\log(c^m)} \sum_{=1}^{c^m} p(\pi_{\alpha_0\alpha_1\ldots\alpha_{m-1}}) \cdot \log p(\pi_{\alpha_0\alpha_1\ldots\alpha_{m-1}})), \tag{11}$$

where the factor $\frac{1}{\log(c^m)}$ simply normalises the output to be in the range $[0, 1]$.

Like ensemble DispEn, the algorithm of ensemble FDispEn can also be defined to propose fluctuation DispEn (FDispEn), which disregards the absolute levels of amplitude in a time series [11]. In this variant, only the differences between adjacent elements of dispersion patterns are considered. The patterns obtained in this way are called *fluctuation dispersion patterns*, which have length $m-1$ and elements ranging from $-c+1$ to $+c-1$. The rest of the algorithm is the same as that of the original DispEn, with the difference of having $(2c-1)^{m-1}$ potential frequency-based dispersion patterns[11].

2.4. Ensemble Shannon and Conditional Entropy

As mentioned before, ShEn and CondEn are two main entropy approaches for processing biomedical signals. SampEn and FuzEn are based on CondEn while PerEn, DispEn, and FDispEn are based on ShEn. To have a low bias entropy for each main entropy technique, we develop the ensemble ShEn (EnsShE) as the average of normalized PerEn, normalized DispEn, and normalized FDispEn values for a signal. Similarly, ensemble CondEn (EnCE) is defined as the average of SampEn and FuzEn values. FuzEn here includes both the fuzzy local and the fuzzy global entropy mneasures to reflect the local and global characteristics of the time series [15].

2.5. Parameters of the Original and Ensemble Entropy Approaches

In the calculation of EnSE, the parameter m and vector of values $\mathbf{r}$ are chosen as 2 and [0.1,0.15,0.2,0.25,0.3] of the standard deviation of $\mathbf{x}$, respectively. Nevertheless, lower or higher values for r_j can be chosen as well. For the SampEn and EnsCE methods, we set $m = 2$, and $r = 0.2$ of the standard deviation (SD) of the original signal [6]. For all the DispEn- and FDispEn-based approaches, we set $c = 6$ and $m = 2$. For more information about the selection of c, please see [10, 11]. For PerEn, we followed $(m+1)! \leq N$ [25].

2.6. Statistical Analysis

In this study, the non-parametric Mann-Whitney U-test is employed to evaluate the differences between two sample means that come from the same population, and used to evaluate if two sample means are equal or not. The Hedges' g effect size [36] is also used to quantify the differences between the results for two groups.

3. Evaluation Data

In this section, we briefly explain the synthetic and real signals used in this study to evaluate the behaviour of ensemble entropy approaches.

3.1. Synthetic Signals

3.1.1. Noise Signals

It has been evidenced that noise is an essential ingredient of the systems [37, 38]. White, pink, and brown noise are three well-known noise types [37, 38]. White noise is a random signal having equal energy across all frequencies. The power spectral density of white noise is as $S(f) = C_w$, where C_w is a constant [38]. Pink and brown noise are random processes suitable for modelling evolutionary or developmental systems [39]. The power spectral density $S(f)$ of pink and brown noises are as $\frac{C_p}{f}$ and $\frac{C_b}{f^2}$, respectively, where C_p and C_b are constants [38].

3.1.2. Logistic Map

Data created by real-world systems, specifically physiological ones, have typically linear and nonlinear components and likely consist of deterministic and stochastic components [40, 41, 42]. To this end, logistic map has been used in many applications to assess a method in detecting periodicity and non-periodic nonlinearity in a signal [40, 41, 42, 11]. The model is dependent on its parameter α as: $x_j = \alpha x_{j-1}(1-x_{j-1})$, where the signal $\mathbf{x}$ was generated with the different values α (e.g., 3.5, 3.6, 3.7, 3.8, 3.9, and 4). In fact, the behavior of the logistic map changes from periodicity to non-periodic nonlinearity when α changes from 3.5 to 4 [43].

The length and sampling frequency of the signal are, respectively, 15,000 sample points (100 s) and 150 Hz. If α equals 3.5, the time series oscillates among four values. For $3.57 \leq \alpha \leq 4$, the series is chaotic, although it has segments with periodic behaviour (e.g., $\alpha \approx 3.8$) [43, 44].

3.1.3. MIX Process

Data created by biological and mechanical systems most likely include deterministic and stochastic components [45]. Hence, to inspect how ensemble entropy methods change when a stochastic sequence progressively turns into a periodic deterministic time series, we generated a MIX process employed by [5, 46]. It is defined as follows:

$$MIX_j = (1 - z_j)x_j + z_j y_j, \quad 1 \leq j \leq N \tag{12}$$

where N is the length of the signal vectors $\mathbf{z} = \{z_j\}$, $\mathbf{MIX} = \{MIX_j\}$, and $\mathbf{y} = \{y_j\}$. $\mathbf{z}$ denotes a random variable which equals 1 with probability p and equals 0 with probability $1 - p$. $\mathbf{x}$ shows a periodic time-series created by $x_j = \sqrt{2}\sin(2\pi j/12)$, and $\mathbf{y}$ is a uniformly distributed series on $\left[-\sqrt{3}, \sqrt{3}\right]$ [47, 46].

3.2. Real Biomedical Datasets

Entropy methods are widely used to characterize physiological signals, such as EEG, ECG, and walking stride interval signals [5, 48, 11, 25]. To this end, three non-invasive datasets including, EEGs [33], RR intervals [34], and walking stride interval signals [35] are used in this study.

3.2.1. Dataset of Focal and Non-focal EEGs

Epilepsy is a common neurological condition. EEG signals are used to identify areas that generate or propagate by seizures [33, 49]. Generally, focal EEG signals are recorded from the epileptic part of the brain, whereas non-focal EEGs correspond to brain regions unaffected by epilepsy [49]. The ability of ensemble entropy techniques to discriminate focal from non-focal signals is evaluated by the use of an EEG dataset (publicly-available at [50]) [33].

The dataset includes 5 patients and, for each patient, there are 750 focal and 750 non-focal bivariate time series. The length of each signal is 20s with sampling frequency of 512 Hz (10240 samples). For more information, please, refer to [33]. All subjects gave written informed consent that their signals from long-term EEG might be used for research purposes [33]. Before applying the complexity methods, the time series were digitally filtered using a Hamming window FIR band-pass filter of order 200 and cut-off frequencies 0.5 Hz and 40 Hz, a band typically used in the analysis of brain activity.

3.2.2. RR Intervals in Fantasia Dataset

To investigate the ability of entropy-based methods to characterize RR intervals, the Fantasia dataset (publicly-available at http://www.physionet.org) is used to distinguish elderly from young subjects [34]. The dataset includes 20 young (21-34 years old) and 20 old (68-85 years old) rigorously-screened healthy individuals who underwent about 120 minutes of continuous supine resting while uncalibrated non-invasive ECG signals were recorded. Each group consisted of 5 women and 5 men [34]. All 40 individuals remained in an inactive state in sinus rhythm when watching the movie Fantasia (Disney, 1940) to help to maintain wakefulness. For each subject, the time series were digitized at 250 Hz. Detailed information can be found in [34].

All 40 individuals provided written informed agreement and underwent a screening history, physical examination, routine blood count and biochemical analysis, ECG, and exercise tolerance test. Only healthy, nonsmoking individuals with normal exercise tolerance tests, no medical problems, and taking no medications were included to the research [34]. RR intervals in ECGs were extracted using jqrs based on the PT's algorithm [51].

3.2.3. Dataset of Walking Stride Intervals

The dataset includes the gait cycle duration on a stride-by-stride basis in 34 young healthy (3-8 years old - 6.03±1.47) vs. 16 elderly healthy (8-14 years old - mean ± standard deviation (SD) 11.58±1.19) children [35]. Their parents were happy with their child participate in this study. If the child and parent were willing to participate, parents were asked to provide informed written consent and to fill out a questionnaire describing the child's medical history. Children were excluded if they had any disorders likely to affect gait, if they were unable to walk independently for 8 minutes, or if they were born prematurely. The participants walked at their self-determined, normal pace for 8 min around a 400-m running track. Two force-sensitive switches were placed inside the subject's right shoe: one underneath the heel of the foot and the other underneath the ball of the foot. The output of these foot switches, which provides a measure of the force applied to the floor, was sampled at 300 Hz. The time series recorded from the subjects walking at their normal pace have the lengths of about 400–500 samples. For more information, please see [35].

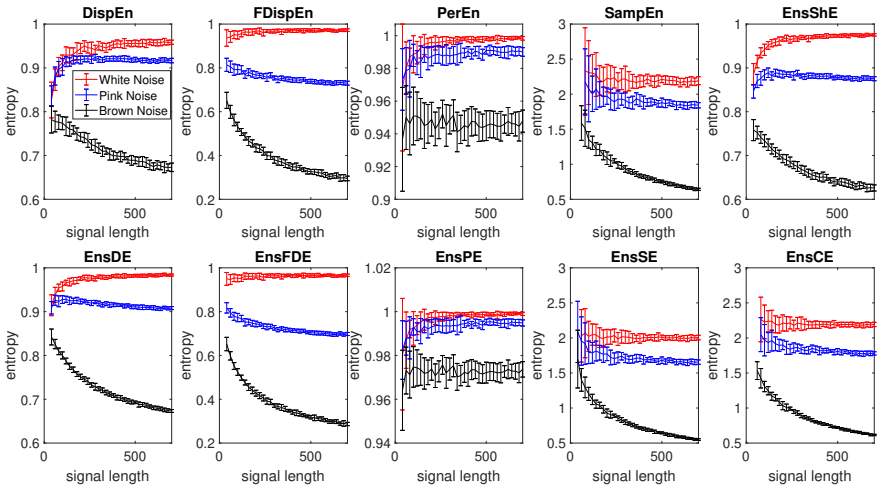


Figure 1: Mean and SD of the results obtained by original and ensemble entropy approaches for white, pink, and brown noises with different lengths changing from 40 to 700 samples.

4. Results and Discussion

4.1. Synthetic Data

4.1.1. Noise Signals

To evaluate the ability of ensemble entropy approaches, compared with their original entropy counterparts, to distinguish the dynamics of different noise signals, we created 40 independent realizations of white, 40 independent realizations of brown, and 40 independent realizations of pink noise signals with different lengths changing from 40 to 700 samples. Figure 1 shows that the profiles for the ensemble entropy techniques are similar to those obtained by their corresponding entropy algorithm. The results are in agreement with the fact that white noise is the most irregular signal, followed by pink and brown noise, in that order, based on the power spectral density of white, pink, and brown noises [37, 38].

EnsDE, compared with DispEn, needs a lower number of samples to discriminate pink from white noise demonstrating an advantage of proposed ensemble DispEn for short signals in detecting dynamical patterns of the

Table 1: Sum of the CV values obtained by ensemble and original entropy techniques for forty independent realizations of white, brown, and pink noises with length changing from 100 to 700 samples.

Methods →	DispEn	EnsDE	FDispEn	EnsFDE	PerEn	EnsPE	SampEn	EnsSE	EnsShE	EnsCE
white noise	0.42	0.19	0.37	0.35	0.14	0.09	NaN	1.59	0.21	NaN
brown noise	0.29	0.20	0.57	0.49	0.19	0.10	NaN	1.66	0.26	NaN
pink noise	0.58	0.26	1.38	1.00	0.39	0.20	NaN	1.57	0.48	NaN

noises. The SampEn and EnsCE, unlike EnSE, values for short signals (40 and 60 samples) are undefined showing the advantage of EnsSE over SampEn and EnsCE for short signals.

To evaluate the stability of the ensemble methods using the different noise types, we examine the sum of their coefficients of variation (CV; ratio of the standard deviation to the mean) for white, brown, and pink noises with lengths of 20, 50, and 700 samples in Table 1. The results show that the ensemble technique noticeably increase the stability of results (lower CV values).

4.1.2. Logistic Map

We employed a sliding window of 150 samples (1 s) with 50% overlap moves along the logistic map signal (Figure 2). The results obtained by all the original and ensemble entropy method are shown in Figure 2. For all the approaches, when $3.5 < \alpha < 3.57$ (periodic part), the entropy values are smaller than those for $3.57 < \alpha < 3.99$ (chaotic part), except those segments including periodic components (e.g., $\alpha \approx 3.8$) [43, 11, 44]. As expected, the entropy values, obtained by the ensemble entropy techniques, in addition to their original forms, increase along the signal, except for the downward spikes in the windows of periodic behavior ($\alpha \approx 3.8$). This behaviour is in agreement with Figure 4.10 (page 87 in [43]) and the other past studies [11, 9, 44].

4.1.3. MIX Process

The entropy approaches are applied to 40 realizations of the MIX process with lengths 100, 300, and 1,000 samples and $p = 0$, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, and 1. The mean and SD values of the results are depicted in Fig. 3. All the profiles show an increase in the irregularity of signals with an increase in the value of p for the MIX process. It is in agreement with the

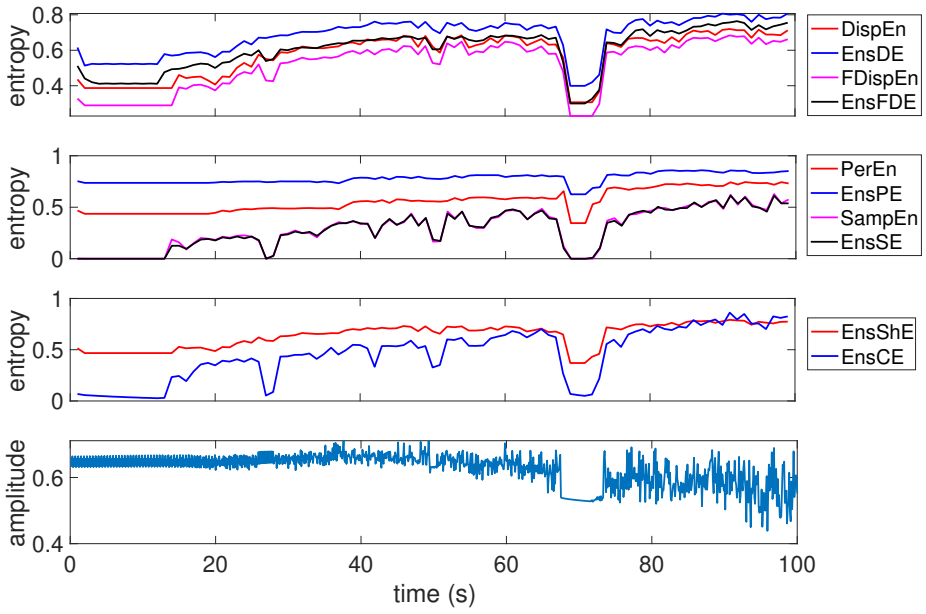


Figure 2: Logistic map with parameter α changing from 3.5 to 3.99 together with entropy values of the logistic map to batter understand original and ensemble entropy approaches.

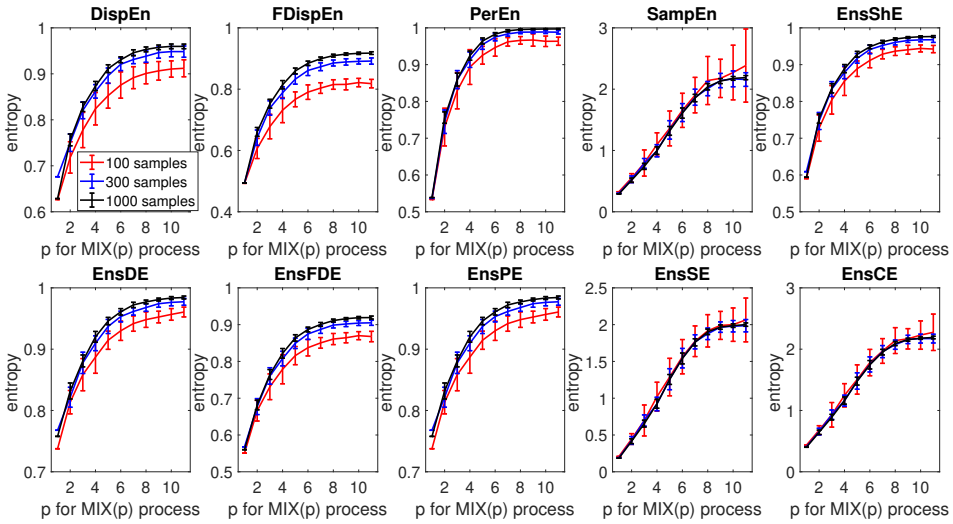


Figure 3: Mean and SD of the results obtained by original and ensemble entropy approaches for MIX(p) ($0 \leq p \leq 1$) with length 100, 300, and 1000 samples.

Table 2: Sum of the CV values obtained by ensemble and original entropy techniques for forty independent realizations of MIX process with length 100, 300, and 1,000 samples.

Methods →	DispEn	EnsDE	FDispEn	EnsFDE	PerEn	EnsPE	SampEn	EnsSE	EnsShE	EnsCE
100 samples	0.32	0.17	0.32	0.27	0.30	0.14	1.87	1.59	0.24	1.26
300 samples	0.25	0.10	0.19	0.16	0.14	0.09	0.80	0.73	0.13	0.58
1000 samples	0.08	0.06	0.10	0.09	0.07	0.05	0.40	0.40	0.07	0.30

fact that the higher the value of p for a MIX process, the more irregular the signal [47, 5].

To compare the results obtained by the original and ensemble entropy approaches, we used the sum of CV values for the MIX process with length 100, 300, and 1,000 samples (Table 2). It is found that the larger the length of signals, the more stable the results. It is also found that ensemble approaches lower the sum of CV values, especially for short signals.

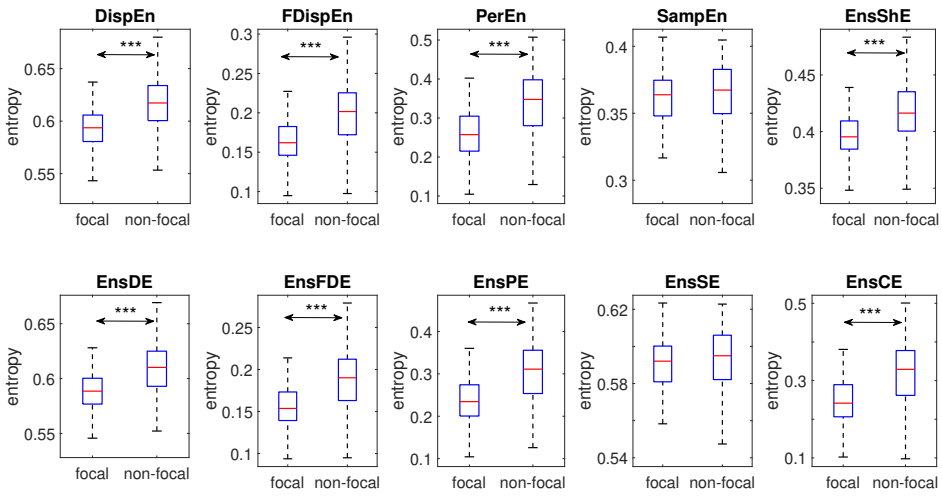


Figure 4: Boxplots for the original and ensemble DispEn, FDispEn, PerEn, and SampEn as well as ensemble ShEn and CondEn obtained from the focal and non-focal EEGs. p-values between 0.01 and 0.05, smaller than 0.01, and smaller than 0.001 are respectively shown with *, **, and ***.

4.2. Real Signals

4.2.1. Dataset of Focal and Non-focal EEGs

The ability of ensemble methods to distinguish the focal from non-focal signals is evaluated here. The results, depicted in Fig. 4, show that the non-focal signals are more irregular than the focal ones. This fact is in agreement with previous studies [33, 52]. Note that, because the entropy-based methods are used for stationary signals [6, 11], we separated each signal into segments of length 2s (1024 samples) and applied the algorithms to each of them and their averaged values are reported. The results demonstrate that all the techniques, except SampEn and EnsSE, lead to similar findings. It should be mentioned that the average entropy values over 2 channels for these bivariate EEG signals are reported for these univariate complexity techniques.

The non-parametric Mann-Whitney U-test was employed to evaluate the differences between the results for focal vs. non-focal signals at each scale factor. The differences for the elderly vs. young children based on Hedges' g effect size are illustrated in Table 3. The results show that the ensemble

Table 3: Differences between results for 1) focal vs. non-focal EEGs; 2) RR interval data for healthy young vs. healthy elderly subjects; and 3) stride interval fluctuations for 3-8 vs. 8-14 years old children (gait maturation) obtained by original and ensemble DispEn, FDispEn, PerEn, and SampEn as well as ensemble ShEn and CondEn based on the Hedges' g effect size.

Methods →	DispEn	EnsDE	FDispEn	EnsFDE	PerEn	EnsPE	SampEn	EnsSE	EnsShE	EnsCE
Focal vs. non-focal	0.94	0.97	0.99	1.01	0.99	1.01	0.11	0.14	0.82	1.00
young vs. elderly (RR intervals)	0.79	1.05	1.07	1.11	0.66	0.69	0.56	0.79	0.80	0.74
young vs. elderly (stride intervals)	0.49	0.58	0.57	0.62	0.20	0.24	0.44	0.62	0.54	0.51

methods, compared with their corresponding original approaches, lead to slightly higher effect sizes.

4.2.2. RR Intervals in Fantasia Dataset

We compare the performance of original and ensemble entropy approaches in analyzing RR intervals. The results are depicted in Fig. 5. The results with significant differences illustrate that young subjects' RR intervals are more irregular than those of elderly people. This finding is in agreement with the fact that aging is associated with irregularity decrease in heart rate [53, 15].

The p-values illustrate that the difference between the groups based on EnSE, unlike SampEn, is statistically significant. The differences for the 20 elderly vs. 20 young subjects based on Hedges' g effect size are illustrated in Table 3. The effect sizes show that the EnsDE and EnsSE, respectively compared with DispEn and SampEn, lead to larger differences for RR interval data in healthy young vs. healthy elderly subjects.

4.2.3. Dataset of Walking Stride Intervals

The results, depicted in Figure 6, show that the median entropy values obtained by the ensemble entropy approaches, except EnsPE, for the elderly children are larger than those for the young children, in agreement with previous studies [54, 55, 11]. The differences for the elderly vs. young children based on Hedges' g effect size (Table 3) demonstrate that ensemble approaches outperform their corresponding original entropy methods to distinguish various dynamics of the stride-to-stride recordings.

In spite of the promising findings, we suggest selecting a subset of entropy models to distinguish various features for an ensemble approach. Such a selection strategy aims to select better entropy models among base entropy

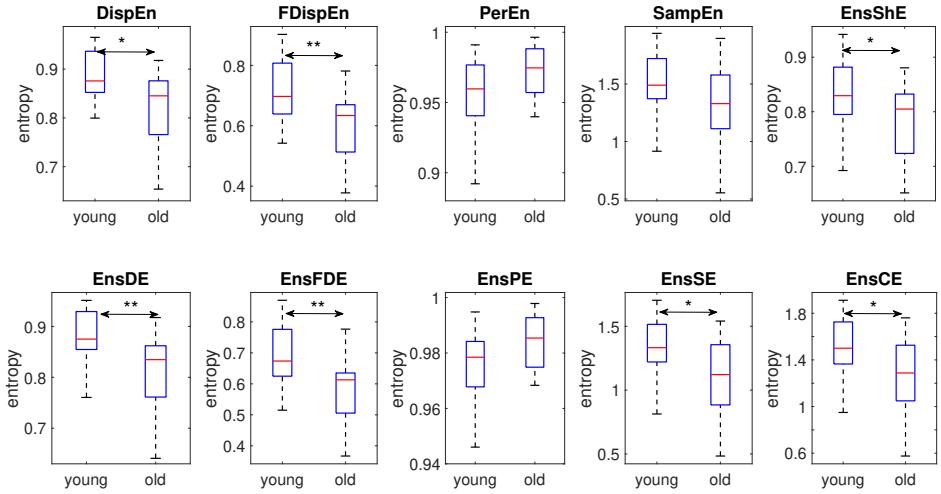


Figure 5: Boxplots for the original and ensemble DispEn, FDispEn, PerEn, and SampEn as well as ensemble ShEn and CondEn computed from RR interval data for healthy young vs. healthy elderly subjects. p-values between 0.01 and 0.05, smaller than 0.01, and smaller than 0.001 are respectively shown with *, **, and ***.

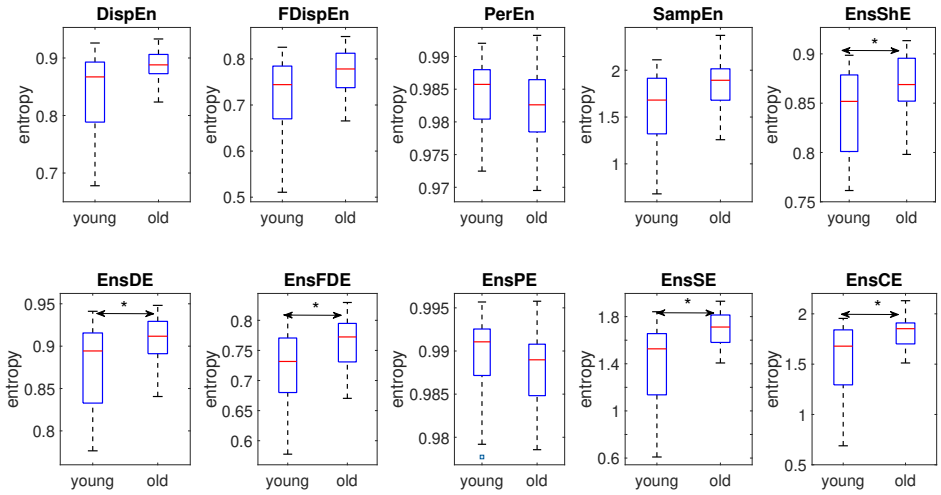


Figure 6: Boxplots for the original and ensemble DispEn, FDispEn, PerEn, and SampEn as well as ensemble ShEn and CondEn computed from the young and elderly healthy children's stride interval recordings. p-values between 0.01 and 0.05 are shown with *.

methods. In fact, the main objective of the entropy ensemble selection is the selection of an appropriate subset of base entropy methods and forms a smaller entropy ensemble that performs better than the set of all of the base entropy techniques.

It is also recommended to work on the interpretability of ensemble models as such models are usually more complex than original entropy approaches. Another issue is that ensemble entropy methods are computationally more expensive (time and space) than original entropy algorithms and also may not be better than all the original entropy approaches in distinguishing various states of physiological data. Therefore, there is a need to find applications or data we want to apply an ensemble entropy approach to it. The ensemble forms of multiscale and multivariate ensemble entropy in addition to 2D entropy approaches also can also be developed [56, 15, 57].

5. Conclusions

The aim of this pilot study is to introduce the concept of ensemble entropy, as a low bias approach, to quantify the data irregularity or uncertainty. The ensemble algorithms benefit from the advantages of different entropy methods (models) or different parameters initialization of a original entropy approach. We evaluated the ensemble approaches for synthetic and real datasets. The study has the following implications. First, ensemble SampEn, unlike SampEn, does not result in undefined values for short signals (100 samples), and a lower number of samples is needed for discrimination between white, pink, and brown noises based on ensemble DispEn, compared with the original DispEn algorithm. Second, all the ensemble techniques lead to lower CV values (more stable results) for different degrees of randomness, especially for short time series. Finally, the results for the real data suggest that the ensemble techniques may lead to larger differences (based on Hedges' g effect size) between physiological conditions known to alter the irregularity or uncertainty of the physiological recordings. Overall, thanks to its ability to detect different levels of signal dynamics and its successful performance, the concept of ensemble entropy opens up a new way to analyze data.

References

[1] W. Xiong, L. Faes, P. C. Ivanov, Entropy measures, entropy estimators, and their performance in quantifying complex dynamics: Effects of ar-

tifacts, nonstationarity, and long-range correlations, Physical Review E 95 (6) (2017) 062114.

[2] C. E. Shannon, Communication theory of secrecy systems, Bell Labs Technical Journal 28 (4) (1949) 656–715.

[3] M. Costa, A. L. Goldberger, C.-K. Peng, Multiscale entropy analysis of complex physiologic time series, Phys. Rev. Lett. 89 (6) (2002) 068102.

[4] L. Faes, A. Porta, G. Nollo, Information decomposition in bivariate systems: theory and application to cardiorespiratory dynamics, Entropy 17 (1) (2015) 277–303.

[5] S. M. Pincus, Approximate entropy as a measure of system complexity, Proceedings of the National Academy of Sciences of the United States of America 88 (6) (1991) 2297–2301.

[6] J. S. Richman, J. R. Moorman, Physiological time-series analysis using approximate entropy and sample entropy, Am J Physiol Heart Circ Physiol 278 (6) (2000) H2039–2049.

[7] A. Porta, G. Baselli, D. Liberati, N. Montano, C. Cogliati, T. Gnecchi-Ruscone, A. Malliani, S. Cerutti, Measuring regularity by means of a corrected conditional entropy in sympathetic outflow, Biological cybernetics 78 (1) (1998) 71–78.

[8] W. Chen, Z. Wang, H. Xie, W. Yu, Characterization of surface emg signal based on fuzzy entropy, IEEE Transactions on neural systems and rehabilitation engineering 15 (2) (2007) 266–272.

[9] C. Bandt, B. Pompe, Permutation entropy: a natural complexity measure for time series, Phys. Rev. Lett. 88 (17) (2002) 174102.

[10] M. Rostaghi, H. Azami, Dispersion entropy: A measure for time-series analysis, IEEE Signal Processing Letters 23 (5) (2016) 610–614.

[11] H. Azami, J. Escudero, Amplitude-and fluctuation-based dispersion entropy, Entropy 20 (3) (2018) 210.

[12] R. Hornero, D. Abásolo, J. Escudero, C. Gómez, Nonlinear analysis of electroencephalogram and magnetoencephalogram recordings in patients with alzheimer's disease, Philosophical Transactions of the Royal

Society A: Mathematical, Physical and Engineering Sciences 367 (1887) (2009) 317–336.

[13] H. Gonçalves, T. Henriques-Coelho, J. Bernardes, A. P. Rocha, A. Nogueira, A. Leite-Moreira, Linear and nonlinear heart-rate analysis in a rat model of acute anoxia, Physiological Measurement 29 (9) (2008) 1133.

[14] M. M. Platisa, V. Gal, Dependence of heart rate variability on heart period in disease and aging, Physiological Measurement 27 (10) (2006) 989.

[15] H. Azami, P. Li, S. E. Arnold, J. Escudero, A. Humeau-Heurtier, Fuzzy entropy metrics for the analysis of biomedical signals: assessment and comparison, IEEE Access 7 (2019) 104833–104847.

[16] X. Li, G. Ouyang, D. A. Richards, Predictability analysis of absence seizures with permutation entropy, Epilepsy research 77 (1) (2007) 70–74.

[17] A. A. Bruzzo, B. Gesierich, M. Santi, C. A. Tassinari, N. Birbaumer, G. Rubboli, Permutation entropy to detect vigilance changes and preictal states from scalp eeg in epileptic patients. a preliminary study, Neurological sciences 29 (1) (2008) 3–9.

[18] X. Li, S. Cui, L. J. Voss, Using permutation entropy to measure the electroencephalographic effects of sevoflurane, The Journal of the American Society of Anesthesiologists 109 (3) (2008) 448–456.

[19] E. Olofsen, J. Sleigh, A. Dahan, Permutation entropy of the electroencephalogram: a measure of anaesthetic drug effect, British journal of anaesthesia 101 (6) (2008) 810–821.

[20] Y. Zhang, P. Shang, Permutation entropy analysis of financial time series based on Hill's diversity number, Communications in Nonlinear Science and Numerical Simulation 53 (2017) 288–298.

[21] D. Jarchi, J. Andreu-Perez, M. Kiani, O. Vysata, J. Kuchynka, A. Prochazka, S. Sanei, Recognition of patient groups with sleep related disorders using bio-signal processing and deep learning, Sensors 20 (9) (2020) 2594.

[22] M. Rostaghi, M. R. Ashory, H. Azami, Application of dispersion entropy to status characterization of rotary machines, Journal of Sound and Vibration 438 (2019) 291–308.

[23] H. Azami, J. Escudero, A. Fernandez, S. E. Arnold, Amplitude-and fluctuation-based dispersion entropy for the analysis of resting-state magnetoencephalogram irregularity in MCI and Alzheimer's disease patients, Alzheimer's & Dementia 15 (2019) 762–762.

[24] H. Azami, L. Faes, J. Escudero, A. Humeau-Heurtier, L. E. Silva, Entropy analysis of univariate biomedical signals: Review and comparison of methods (2020).

[25] M. Zanin, L. Zunino, O. A. Rosso, D. Papo, Permutation entropy and its main biomedical and econophysics applications: a review, Entropy 14 (8) (2012) 1553–1577.

[26] J. Kurths, A. Voss, P. Saparin, A. Witt, H. Kleiner, N. Wessel, Quantitative analysis of heart rate variability, Chaos: An Interdisciplinary Journal of Nonlinear Science 5 (1) (1995) 88–94.

[27] B.-l. Hao, Symbolic dynamics and characterization of complexity, Physica D: Nonlinear Phenomena 51 (1-3) (1991) 161–176.

[28] A. Voss, J. Kurths, H. Kleiner, A. Witt, N. Wessel, Improved analysis of heart rate variability by methods of nonlinear dynamics, Journal of Electrocardiology 28 (1995) 81–88.

[29] W. Wang, Some fundamental issues in ensemble methods, in: 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), IEEE, 2008, pp. 2243–2250.

[30] T. Alqurashi, W. Wang, Clustering ensemble method, International Journal of Machine Learning and Cybernetics 10 (6) (2019) 1227–1246.

[31] C. Zhang, Y. Ma, Ensemble machine learning: methods and applications, Springer, 2012.

[32] O. Okun, Supervised and Unsupervised ensemble methods and their applications, Vol. 126, Springer, 2008.

[33] R. G. Andrzejak, K. Schindler, C. Rummel, Nonrandomness, nonlinear dependence, and nonstationarity of electroencephalographic recordings from epilepsy patients, Physical Review E 86 (4) (2012) 046206.

[34] N. Iyengar, C. Peng, R. Morin, A. L. Goldberger, L. A. Lipsitz, Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics, American Journal of Physiology-Regulatory, Integrative and Comparative Physiology 271 (4) (1996) R1078–R1084.

[35] J. Hausdorff, L. Zemany, C.-K. Peng, A. Goldberger, Maturation of gait dynamics: stride-to-stride variability and its temporal organization in children, Journal of applied physiology 86 (3) (1999) 1040–1047.

[36] R. Rosenthal, H. Cooper, L. Hedges, et al., Parametric measures of effect size, The handbook of research synthesis 621 (2) (1994) 231–244.

[37] L. Cohen, The history of noise [on the 100th anniversary of its birth], IEEE Signal Processing Magazine 22 (6) (2005) 20–45.

[38] E. Sejdić, L. A. Lipsitz, Necessity of noise in physiology and medicine, Computer Methods and Programs in Biomedicine 111 (2) (2013) 459–470.

[39] M. S. Keshner, 1/f noise, Proceedings of the IEEE 70 (3) (1982) 212–218.

[40] M. Costa, A. L. Goldberger, C.-K. Peng, Multiscale entropy analysis of biological signals, Physical review E 71 (2) (2005) 021906.

[41] S. M. Pincus, A. L. Goldberger, Physiological time-series analysis: what does regularity quantify?, American Journal of Physiology-Heart and Circulatory Physiology 266 (4) (1994) H1643–H1656.

[42] A. L. Goldberger, C.-K. Peng, L. A. Lipsitz, What is physiologic complexity and how does it change with aging and disease?, Neurobiology of aging 23 (1) (2002) 23–26.

[43] G. L. Baker, J. P. Gollub, Chaotic dynamics: an introduction, Cambridge University Press, 1996.

[44] J. Escudero, R. Hornero, D. Abásolo, Interpretation of the auto-mutual information rate of decrease in the context of biomedical signal analysis.

application to electroencephalogram recordings, Physiological measurement 30 (2) (2009) 187.

[45] M. Costa, A. L. Goldberger, C.-K. Peng, Multiscale entropy analysis of biological signals, Physical Review E 71 (2) (2005) 021906.

[46] J. Escudero, D. Abásolo, R. Hornero, P. Espino, M. López, Analysis of electroencephalograms in Alzheimer's disease patients with multiscale entropy, Physiological Measurement 27 (11) (2006) 1091.

[47] M. Ferrario, M. G. Signorini, G. Magenes, S. Cerutti, Comparison of entropy-based regularity estimators: application to the fetal heart rate signal for the identification of fetal distress, Biomedical Engineering, IEEE Transactions on 53 (1) (2006) 119–125.

[48] S. M. Pincus, A. L. Goldberger, Physiological time-series analysis: what does regularity quantify?, Am J Physiol Heart Circ Physiol 266 (4) (1994) H1643–1656.

[49] U. R. Acharya, H. Fujita, V. K. Sudarshan, S. Bhat, J. E. Koh, Application of entropies for automated diagnosis of epilepsy using EEG signals: a review, Knowledge-Based Systems 88 (2015) 85–96.

[50] http://ntsa.upf.edu/.

[51] A. N. Vest, G. Da Poian, Q. Li, C. Liu, S. Nemati, A. J. Shah, G. D. Clifford, An open source benchmarked toolbox for cardiovascular waveform and interval analysis, Physiological measurement 39 (10) (2018) 105004.

[52] R. Sharma, R. B. Pachori, U. R. Acharya, Application of entropy measures on intrinsic mode functions for the automated identification of focal electroencephalogram signals, Entropy 17 (2) (2015) 669–691.

[53] A. L. Goldberger, C.-K. Peng, L. A. Lipsitz, What is physiologic complexity and how does it change with aging and disease?, Neurobiology of Aging 23 (1) (2002) 23–26.

[54] S. L. Hong, E. G. James, K. M. Newell, Age-related complexity and coupling of children's sitting posture, Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology 50 (5) (2008) 502–510.

[55] M. Bisi, R. Stagni, Complexity of human gait pattern at different ages assessed using multiscale entropy: from development to decline, Gait & posture 47 (2016) 37–42.

[56] H. Azami, A. Fernández, J. Escudero, Multivariate multiscale dispersion entropy of biomedical times series, Entropy 21 (9) (2019) 913. doi:10.3390/e21090913.
URL http://dx.doi.org/10.3390/e21090913

[57] H. Azami, M. Rostaghi, D. Abásolo, J. Escudero, Refined composite multiscale dispersion entropy and its application to biomedical signals, IEEE Transactions on Biomedical Engineering 64 (12) (2017) 2872–2879.

Credit Author Statement

Ensemble Entropy: A Low Bias Approach for Data Analysis

Hamed Azami[a], Saeid Sanei[b], Tarek K. Rajji[a]
[a]Centre for Addiction and Mental Health, Toronto Dementia Research Alliance, University of Toronto, Toronto, ON, Canada., Toronto, ON, Canada.
[b]School of Science and Technology, Nottingham Trent University, Nottingham, UK.

Hamed Azami: Conceptualization, Methodology, Writing- Original draft preparation **Saeid Sanei**: Methodology and Editing : **Tarek K. Rajji:** Supervision and Editing,

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: