

Article

Using the Information Provided by Forbidden Ordinal Patterns in Permutation Entropy to Reinforce Time Series Discrimination Capabilities

David Cuesta-Frau 

Technological Institute of Informatics, Universitat Politècnica de València, 03801 Alcoi Campus, Spain; dcuesta@disca.upv.es; Tel.: +34-966528505

Received: 13 March 2020; Accepted: 20 April 2020; Published: 25 April 2020



Abstract: Despite its widely tested and proven usefulness, there is still room for improvement in the basic permutation entropy (PE) algorithm, as several subsequent studies have demonstrated in recent years. Some of these new methods try to address the well-known PE weaknesses, such as its focus only on ordinal and not on amplitude information, and the possible detrimental impact of equal values found in subsequences. Other new methods address less specific weaknesses, such as the PE results' dependence on input parameter values, a common problem found in many entropy calculation methods. The lack of discriminating power among classes in some cases is also a generic problem when entropy measures are used for data series classification. This last problem is the one specifically addressed in the present study. Toward that purpose, the classification performance of the standard PE method was first assessed by conducting several time series classification tests over a varied and diverse set of data. Then, this performance was reassessed using a new Shannon Entropy normalisation scheme proposed in this paper: divide the relative frequencies in PE by the number of different ordinal patterns actually found in the time series, instead of by the theoretically expected number. According to the classification accuracy obtained, this last approach exhibited a higher class discriminating power. It was capable of finding significant differences in six out of seven experimental datasets—whereas the standard PE method only did in four—and it also had better classification accuracy. It can be concluded that using the additional information provided by the number of forbidden/found patterns, it is possible to achieve a higher discriminating power than using the classical PE normalisation method. The resulting algorithm is also very similar to that of PE and very easy to implement.

Keywords: permutation entropy; ordinal patterns; forbidden patterns; signal classification

1. Introduction

Despite its relatively young age in comparison with other entropy statistics, permutation entropy (PE) has already become one of the most utilised time series entropy-related measures. It was proposed in the well known paper by Bandt and Pompe [1] in 2002, and since then, it has given rise to a number of applications and further algorithm developments. This number is growing exponentially [2], which confirms the goodness of the PE approach.

Regarding PE applications, this measure has been used in many fields, such as medicine, engineering, seismology, and economics. In medicine, it has been frequently used as a diagnostic aid to disclose information hidden in physiological time series. The most common physiological time series processed using PE are probably electroencephalograms (EEG) and electrocardiograms, but in the form of R-wave interval series (RR records). For example, in [3], PE was applied to EEG records to find seizure-free, pre-seizure, and seizure phases. This a recurrent application in many similar

studies. The work described in [4] assessed two classifiers based on logistic regression and quadratic discriminant analysis to detect obstructive sleep apnea using PE as one of the features extracted from RR records. Other biomedical signals successfully processed using PE are body temperature time series [5] and glucose records [6]. These are just a few examples of the huge number of medical applications based on PE. A more detailed survey is described in [2].

Engineering is another field that has exploited the PE capabilities. It has been mainly used in mechanical engineering for fault detection. That is the case in [7], where the authors described a mathematical tool based on PE for bearings fault type diagnosis. Many similar applications can be found in the scientific literature [7,8]. There are other novel and interesting implementations in mechanical engineering, such as [9], where PE was used to assess road flatness.

Seismology and economics are emergent fields of PE utilisation. The work in [10] used noise randomness assessed by PE to quantify its changes as a possible precursor of volcanic activity. An example of PE used in economics is [11]. This work described a method based on PE to detect changes in stock market data.

The basic PE algorithm has also been improved or customised since its initial version. PE is based on the relative frequency of ordinal patterns that can be drawn from a time series, but it overlooks the information embedded in the amplitude. As a consequence, some PE-derived methods have been devised to address this problem, such as the weighted permutation entropy [12] (WPE), the fine grained permutation entropy [13] (FGPE), and the amplitude aware permutation entropy [14] (AAPE). A comparative analysis of these measures has been recently reported in [15]. Other problems, such as order ambiguities due to equal value samples in subsequences (ties), have been addressed in [14,16], with practical analyses described in [17,18]. Other PE algorithm tweaks can be found in the scientific literature; for instance, the bubble entropy (BE) method [19], aimed at improving the robustness of PE in terms of input parameters.

However, there is another topic related to PE that has not been very much exploited in practical term—*forbidden patterns*. When analysing the order (either ascending or descending) of a numeric subsequence of length m , theoretically, any of the $m!$ possible permutation patterns could be found. This is true for random sequences where the probability of occurrence of each ordinal pattern is not zero, provided the total length suffices to allocate even the least probable pattern. Nevertheless, if the sequence exhibits a certain degree of determinism, some order relations will never occur—the so-called *forbidden patterns* [20]. Actually, the numbers of admissible/*forbidden patterns* have been demonstrated to provide information with system distinguishing power [21], and a few studies have exploited this new feature independently of PE for signal classification [22]. Based on all these previous studies, we hypothesised that a new scheme that could combine standard PE and a metric related to the number of forbidden patterns, could better portray the dynamics of the underlying time series, and therefore result in a more powerful measure for signal classification.

Thus, this paper describes a modified PE method by means of integrating the number of ordinal patterns found (the complementary of the number of forbidden/missing patterns) in the normalisation step. With this simple customisation, the resulting new method, referred to in this paper as PE2, in contrast to the standard method, PE1, is clearly more robust and more powerful in terms of signal classification than the PE1 method, as the experiment results, using a varied and diverse set of time series, demonstrate in Section 3.

The structure of the paper is as follows. Section 2.1 describes the PE1 algorithm, followed by Section 2.2, describing the new variant proposed, PE2. Section 2.3 details the time series datasets used in the experiments. The metrics employed to quantify and assess the performance of the approaches are outlined in Section 2.4. The quantitative results of all the cases analysed are presented in Section 3. The Discussion section (Section 4) provides an interpretation of the results. Finally, the conclusions are described in Section 5.

2. Methods

2.1. Permutation Entropy

The present study is based on the original PE algorithm described in [1]. This method computes a normalised histogram of ordinal patterns found in the subsequences drawn from a time series, when sorted in ascending order, from which the Shannon entropy is calculated. The length of these subsequences is defined by an input parameter, the embedded dimension m .

Formally, the input time series under analysis is defined as a vector of N components $\mathbf{x} = \{x_0, x_1, \dots, x_{N-1}\}$. A generic subsequence extracted commencing at sample x_j of \mathbf{x} is defined as a vector of m components $\mathbf{x}_j^m = \{x_j, x_{j+1}, \dots, x_{j+m-1}\}$. In its original state, the samples in \mathbf{x}_j^m can be assigned a default growing set of indices given by $\pi^m = \{0, 1, \dots, m-1\}$. The subsequence \mathbf{x}_j^m undergoes, then, an ascending sorting process, and the sample order changes in it are mirrored in the vector of indices π^m . The resulting new version of this vector, $\pi_j^m = \{\pi_0, \pi_1, \dots, \pi_{m-1}\}$, with $x_{j+\pi_0} \leq x_{j+\pi_1} \leq x_{j+\pi_2} \dots \leq x_{j+\pi_{m-1}}$, is compared, in principle, with all the possible $m!$ ordinal patterns of length m . When a coincidence is found, a specific associated counter to that pattern, $c_i \in \mathbf{c}$, is increased. This process is repeated with all the possible $N - (m - 1)$ subsequences ($0 \leq j < N - m + 1$) until the complete histogram is obtained. Each bin of the histogram is finally normalised by $N - (m - 1)$ in order to get an estimation of the probability of each ordinal pattern: $\mathbf{p} = \{p_0, p_1, \dots, p_{m!-1}\} \mid p_i = \frac{c_i}{N - (m - 1)}$. This vector of probabilities is used to calculate PE as:

$$\text{PE}(\mathbf{x}, m, N) = - \sum_{k=0}^{m!-1} p_k \log p_k, \forall p_k > 0 \quad (1)$$

There is another input parameter for PE, the embedded delay τ . This parameter, when $\tau > 1$, defines the time scale at which PE is computed, and it can contribute to gain a deeper insight into the temporal correlations of the time series [23]. However, since this parameter is almost equivalent to a downsampling process [14], and given that the present study is a comparative study in relative terms, we took $\tau = 1$ in all the experiments, as in many other works [1,14,19,22].

Some numerical examples of PE computation can be found in the literature. For examples, see [15,18,24].

For comparative purposes, another improved version of PE will be included in some experiments, WPE [12]. The difference is to use a weight factor applied to the relative PE frequencies that quantifies amplitude values. The weighting factor for each relative frequency is given by $w_j = \frac{1}{m} \sum_{k=0}^{m-1} (x_{j+k} - \bar{X}_j^m)^2$, \bar{X}_j^m being the arithmetic mean of \mathbf{x}_j^m . The new value W becomes the new

denominator instead of $N - m - 1$, with $W = \sum_{j=0}^{N-m} w_j$. Further WPE details are described in [15].

2.2. Permutation Entropy Using the Number of Patterns Found

Forbidden patterns, in the sense of patterns that will never occur in a sequence regardless of its length, have been demonstrated to provide additional information about the determinism degree of the underlying time series [21]. These forbidden patterns can even be considered as a new dynamical property [21,25], and have already been used successfully as a quantifier to assess stock market inefficiency [22]. In cases of unobserved patterns due to low probabilities of occurrence in a relatively short time series, namely, not truly forbidden but missing patterns [25], they can also be considered potential distinguishing features if all the records are balanced with regard to length [26]. The numbers of forbidden and admissible patterns are two sides of the same coin, since they are complementary and totalise the theoretical number of possible patterns $m!$.

In a more formal way, if the probability of an ordinal pattern π_i is P_{π_i} , for an admissible pattern $P_{\pi_i} > 0$, whereas $P_{\pi_i} = 0$ for a forbidden pattern (these probabilities can be thought of as relative

frequencies for finite length time series). A forbidden pattern implies there is no \mathbf{x}_i^m in \mathbf{x} for which the ordinal pattern is π_i . The presence of a forbidden pattern is heuristically linked to determinism [25], and its presence causes an even higher number of forbidden patterns for longer subsequences, which can also be exploited from a classification point of view, as will be described in the experiments later. For example, if $\pi_i^3 = \{2, 1, 0\}$ is a forbidden pattern of \mathbf{x} , for $m = 4$, $\{3, 2, 1, 0\}$, $\{2, 3, 1, 0\}$, $\{2, 1, 3, 0\}$, and $\{2, 1, 0, 3\}$ will be also forbidden patterns [20], and so on.

Periodic sequences excellently illustrate the concept of forbidden patterns (a purely deterministic time series). Let $\mathbf{x} = \{1, 2, 3, 1, 2, 3, 1, 2, 3, 1, 2, 3, \dots, 1, 2, 3\}$ be a periodic sequence of length N and period 3. All the subsequences of length $m = 3$ that can be extracted from \mathbf{x} are $\mathbf{x}_0^3 = \{1, 2, 3\}$, $\mathbf{x}_1^3 = \{2, 3, 1\}$, $\mathbf{x}_2^3 = \{3, 1, 2\}$, $\mathbf{x}_3^3 = \{1, 2, 3\}$, $\mathbf{x}_4^3 = \{2, 3, 1\}$, $\mathbf{x}_5^3 = \{3, 1, 2\}$, $\mathbf{x}_6^3 = \{1, 2, 3\}$, $\mathbf{x}_7^3 = \{2, 3, 1\}$, $\mathbf{x}_8^3 = \{3, 1, 2\}$, $\mathbf{x}_9^3 = \{1, 2, 3\}$, $\mathbf{x}_{10}^3 = \{2, 3, 1\}$, $\mathbf{x}_{11}^3 = \{3, 1, 2\}$, \dots , $\mathbf{x}_{N-m}^3 = \{1, 2, 3\}$.

When the previous subsequences are ordered, the results are: $\pi_0^3 = \{0, 1, 2\}$, $\pi_1^3 = \{2, 0, 1\}$, $\pi_2^3 = \{1, 2, 0\}$, and these three motifs keep repeating indefinitely. It can be easily concluded that motifs $\{0, 2, 1\}$, $\{1, 0, 2\}$, and $\{2, 1, 0\}$, will never be found in \mathbf{x} , even for $N \rightarrow \infty$. These three motifs could therefore be considered forbidden patterns.

Given that PE looks at the dynamics of a time series in terms of relative frequency of ordinal patterns, but overlooks the additional information provided by the number of forbidden/admissible patterns (which is also related to the randomness of the time series [20]), we hypothesised that there could exist a potential synergy between the two sources. After studying several integration possibilities, a straightforward and simple solution was to change the PE probability normalisation factor $N - m + 1$, which accounts for the number of subsequences that can be drawn from the time series, by the actual number of different ordinal patterns found, termed T .

In principle, PE becomes non-normalised this way, since $\sum_{\forall p} p_k \neq 1$ in most cases. The T value can be considered a rescaling or weighting factor that embeds the forbidden patterns information in the modified PE measure, PE2, and its additional class discriminating power would be lost if an individual normalisation took place on a signal by signal basis (intrinsic). In order to keep the PE2 results within a reasonable interval, as with other similar measures, including PE, it is recommended to adopt a global feature normalisation scheme (extrinsic) after PE2 computation, if normalisation is desired.

There are many feature normalisation methods reported in the scientific literature: Z-Score, Min-Max, and other linear and non-linear scaling methods [27]. In case it is necessary, we propose using a linear proportionate normalisation scheme [28]. Each PE2 value can be divided by the sum of all the PE2 values. Therefore, each result accounts for its relative contribution within the entire set of results; it does not destroy proportionality (namely, discriminating power), and the newly computed value can be easily related to the original value. In other words, order and differences are not lost or modified. Moreover, it is not based on arbitrary choices, and its implementation is straightforward. An even more convenient variation of this method would be to divide the PE2 values by the maximum PE2 value obtained in order to keep the range of possible results between 0 and 1 [29], more easily interpretable.

For example, in [15] the sequence $\mathbf{x} = \{-0.45, 1.9, 0.87, -0.91, 2.3, 1.1, 0.75, 1.3, -1.6, 0.47, -0.15, 0.65, 0.55, -1.1, 0.3\}$ was analysed from a PE perspective ($N = 15, m = 3$). As a result, the counter of the frequencies for each ordinal pattern obtained was $\mathbf{c} = \{0, 4, 3, 1, 2, 3\}$, with $\mathbf{p} = \{0, 0.31, 0.23, 0.08, 0.15, 0.23\}$, using $N - m + 1 = 13$ as the normalisation factor. Since the ordinal pattern $\{0, 1, 2\}$ was not found (a missing pattern, because \mathbf{x} is a random sequence and $\{0, 1, 2\}$ would occur if given more samples), applying PE2, \mathbf{p} would have instead given $\mathbf{p} = \{0, 0.8, 0.6, 0.2, 0.4, 0.6\}$ ($T = 5$), from which $\text{PE2}(\mathbf{x}, 3, 15) = 2.135$ whereas PE1 was $\text{PE1}(\mathbf{x}, 3, 15) = 2.20$ [15]. Once the PE2 values from all the time series under analysis were computed, the normalisation scheme described above could be applied, although it would not be necessary for classification purposes because the differences would already be in the resulting PE2 values.

As for the PE1 algorithm, a PE2 algorithm can be implemented in many ways, but we have chosen to use a bubble sort approach to obtain the ordered indices, and dynamically update the

list of different ordinal patterns found, termed Π^m (initially empty), instead of assuming a set with all the possible $m!$ permutations. This implementation is less computationally efficient due to the list operations (searching and appending), but it is better suited to implementing the improvements devised in the PE2 algorithm. Besides, it can be more memory-efficient, since, in case of forbidden patterns, as in many chaotic time series [22,30,31], there is no need to store all the theoretically possible $m!$ ordinal patterns, only those really found in the data. This could entail significant memory savings when m is relatively large and/or forbidden patterns are frequent [26]. Last but not least, a linked-list facilitates the implementation and even integration of other PE variants based on a dynamic generation of patterns, such as FGPE [13]. Based on this approach, a suggested implementation is shown in Algorithm 1.

Algorithm 1 Permutation entropy scaled by the number of patters found (PE2) algorithm.

```

Input:  $x, m > 2, N > m + 1$ 
Initialisation:  $PE2 \leftarrow 0, c \leftarrow \{\emptyset\}, p \leftarrow \{\emptyset\}, \Pi^m \leftarrow \{\emptyset\}$ 
for  $j \leftarrow 0, \dots, N - m$  do
   $x_j^m \leftarrow \{x_j, x_{j+1}, \dots, x_{j+m-1}\}$ 
   $\pi^m \leftarrow \{0, 1, \dots, m - 1\}$ 
  bSorted  $\leftarrow$  false
  while (bSorted=false) do ▷ Bubble sort
    bSorted  $\leftarrow$  true
    for  $i \leftarrow j, \dots, j + m - 2$  do
      if ( $x_i > x_{i+1}$ ) then
        swap( $x_i, x_{i+1}$ )
        swap( $\pi_i, \pi_{i+1}$ )
      bSorted  $\leftarrow$  false
    end if
  end for
end while
  bFound  $\leftarrow$  false
  for  $i \leftarrow 0, \dots, \text{sizeof}(\Pi^m)$  do ▷ List search
    if  $\pi_j^m = \Pi_i^m$  then ▷ Ordinal pattern found
       $c_i \leftarrow c_i + 1$  ▷ Update counter
      bFound  $\leftarrow$  true
      break ▷ Pattern found, exit loop
    end if
  end for
  if not bFound then ▷ Ordinal pattern not found
     $\Pi^m \leftarrow \pi_j^m$  ▷ Append pattern to list
     $c \leftarrow 1$  ▷ Append and initialise pattern count
  end if
end for
 $T = \text{sizeof}(\Pi^m)$ 
for  $i \leftarrow 0, \dots, T$  do
   $p \leftarrow p_i \leftarrow \frac{c_i}{T}$  ▷ DESNORMALISATION
   $PE2 \leftarrow PE2 + (-p_i \log p_i)$ 
end for
Output:  $PE2(x, m, N)$ 

```

The PE2 algorithm can become equivalent to PE just by replacing T by $N - m + 1$ at the line labelled DESNORMALISATION in Algorithm 1. If the records are very short, it is not just a question about forbidden patterns but about the low probabilities of certain ordinal patterns. If all the classes

exhibit the same behaviour in terms of length, this should not be an influencing factor; otherwise, a length normalisation scheme should be devised.

2.3. Experimental Dataset

First, the addition of synthetic databases was considered, since this kind of records is also very useful for characterising the performances of methods under more controlled conditions. In a very recent paper [32], we proposed to use a hidden Markov model to create synthetic records based on transition probabilities of their ordinal patterns of length $m = 3$. This is a very suitable tool to create a synthetic dataset for the present study, since the main difference between PE1 and PE2 is the use of the number of actual ordinal patterns found. Assigning a 0 probability to some ordinal pattern transitions, the frequency of the destination pattern can be minimised, and for $m > 3$, the probability of derived patterns is likely to reach 0, since the number of forbidden patterns grows superexponentially [20].

a_{ij} being the transition probability between consecutive states q_i and q_j at time t , $a_{ij} = p[\text{State}_{t+1} = q_j | \text{State}_t = q_i]$, and the following correspondence between model states and ordinal patterns: $q_0 \leftarrow (0, 1, 2)$, $q_1 \leftarrow (0, 2, 1)$, $q_2 \leftarrow (1, 0, 2)$, $q_3 \leftarrow (1, 2, 0)$, $q_4 \leftarrow (2, 1, 0)$, and $q_5 \leftarrow (2, 0, 1)$, 100 records of two synthetic classes were generated using this model. For one class, the transition probabilities were $a_{ij} = \{0.5, 0.5, 0.0\}$, and for the second class $a_{ij} = \{0.5, 0.0, 0.5\}$, probabilities defined as in [32]. Therefore, each class penalised a different transition, and that would impact the number of patterns found at $m = 4$ and beyond in a different way for each class, since the model is not symmetric (see details in [32]). An example of the resulting signals is shown in Figure 1. The experiments on these records used 10 random realisations in each test.

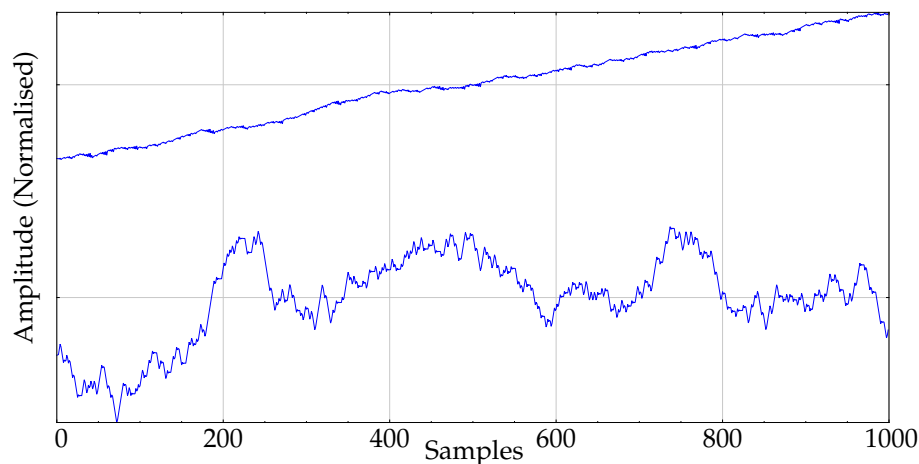


Figure 1. Example of synthetic signals of the experimental database.

In addition, a real experimental dataset was chosen with the primary goal of assembling a publicly available collection of widely representative data of the most common time series entropy applications. This would enable other researchers to replicate the experiments and draw sound conclusions about the most likely performance under a disparity of conditions, not just for a single dataset/case, as occurs in many studies. Figure 2 depicts an example of one signal from each dataset used, described next:

- BONN. This database was collected at the Department of Epileptology, University of Bonn [33], and is a frequently used dataset found in many similar research studies [17,34–38]. The length of the records is 4097, with two classes of 100 time series each, corresponding to seizure-free and seizure-included electroencephalograms (EEGs). This dataset was chosen due to its popularity among the scientific community, and because EEGs are the focus of many entropy-related studies.
- GAIT. The records included in this dataset were drawn from Physionet gait in the ageing and disease database [39]. Although this is a small collection of gait data, with only five subjects per class, we found it very representative of another group of physiological data, not as common as

EEGs, and useful to explore algorithms' performances. The 15 records correspond to five healthy young adults, five healthy old adults, and five old adults with Parkinson's disease. The data are stride intervals [40]. The length of the records is around 800 samples for healthy subjects, and it is 200 for pathological ones, which suffices for a representative classification analysis using PE, according to recent studies [26]. Anyway, a variation of this subset, termed GAIT2, where all the records were cut short to 200 samples, was included in the experiments for comparative purposes.

- FANT. The fantasia dataset contains 120 min of electrocardiographic and respiration data from 20 young and 20 elderly healthy subjects, and it is also available at [39]. Only the RR-interval time series were used in the experiments in this paper. RR records are also a field of intensive research [4,41–44]. A detailed description of this database can be found in [43].
- RATS. Records of blood pressure readings from Dahl SS rats on high and low salt diets [45]. The database contains nine records of each class, sampled at 100 Hz, with a total length of 12,000 samples [39].
- WORMS. This database corresponds to the recorded 2D movement of genetically-modified worms [46–48], and is publicly available at www.timeseriesclassification.com. It was included to have a dataset not related to physiological records, and to widen the scope of the analysis. Specifically, there were 181 records of two classes (76 wild-type and 105 mutant type) of length 900 in this subset used in the experiments.
- HOUSE. The records in this database are also publicly available at www.timeseriesclassification.com, and they correspond to non-physiological data too. There are two classes of 20 records each, with 1022 samples per record [49].
- PAF. This dataset contains paroxysmal atrial fibrillation (PAF) records [50]. There are two classes of 25 records, each one (PAF and PAF-free episodes) five minutes in duration. These records were also drawn from [39].

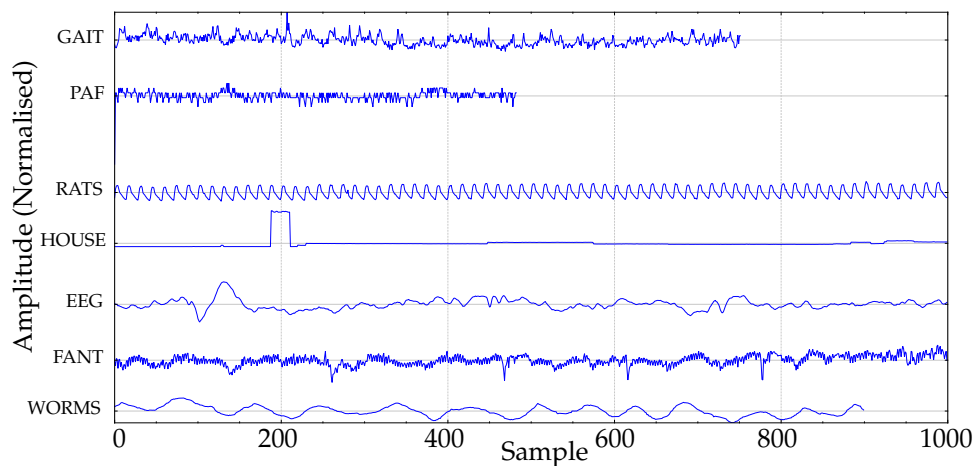


Figure 2. Example of real signals of the experimental database.

2.4. Performance Analysis

The performance of each approach under analysis was quantified using the classification accuracy: ratio of correctly classified records. Significance of this classification was qualitatively assessed by means of sensitivity (Se) and specificity (Sp), since very unbalanced results (for example, 0.7 accuracy with $Se = 0.4$ and $Sp = 1$ is not considered significant, at least a minimum 0.6 result is required), reflect an underlying poor discriminating power, regardless of the global classification accuracy. The classification threshold was taken as the ROC (receiver operating characteristic) curve point closest to the (0,1) coordinates [51]. It is important to note that the goal of the study was not to design an optimal classifier, but to carry out a fair comparison between the performances of the two measures tested under the same conditions.

The quantitative significance of the classification accuracy was assessed by means of an unpaired Wilcoxon–Mann–Whitney test. This is a very robust test that does not require data normality [52]. The significance threshold was set at $\alpha = 0.05$.

3. Results

The experiments were first carried out using the synthetic dataset described before. Since the transition probabilities were quite different, both PE1 and PE2 were capable of finding significant differences between the two classes generated. These results are shown in Table 1. It is important to note that as m increased, and the differences between classes in terms of patterns found became greater, and so did the PE2 accuracy, whereas PE1 performance was lower.

Table 1. Performances achieved using the synthetic records, including the number of patterns found in each case.

Method		$m = 3$	4	5	6	7	8
PE1	Accuracy	1 ± 0	0.65 ± 0.04	0.86 ± 0.07	0.97 ± 0.01	0.66 ± 0.09	0.71 ± 0.08
PE2	Accuracy	0.83 ± 0.01	0.87 ± 0.09	0.91 ± 0.03	1 ± 0	1 ± 0	0.99 ± 0.01
Patterns found	Class 1	4.32 ± 0.46	8.30 ± 0.45	19.51 ± 0.63	40.85 ± 1.63	93.14 ± 2.99	182.82 ± 5.79
	Class 2	4.33 ± 0.47	10.01 ± 0.76	22.71 ± 1.50	53.35 ± 1.91	117.08 ± 5.79	209.71 ± 8.65

Table 2 shows the classification performances and statistical significance of the results achieved by the standard PE1 on real records. These experiments also included a variation of the parameter m , from 3 to 8, since input parameter influence is another topic of intense debate and research in the scientific literature.

Table 2. Classification results achieved using PE1 applied to the real experimental dataset. Classification performance is quantified with three parameters: sensitivity, specificity, and accuracy.

	$m = 3$	4	5	6	7	8
BONN	0.93, 0.90, 0.91 $p < 0.0001$	0.93, 0.89, 0.91 $p < 0.0001$	0.92, 0.89, 0.90 $p < 0.0001$	0.91, 0.89, 0.90 $p < 0.0001$	0.93, 0.85, 0.89 $p < 0.0001$	0.90, 0.85, 0.87 $p < 0.0001$
	0.40, 1, 0.70 $p = 0.9168$	0.60, 1, 0.80 $p = 0.2505$	1, 1, 1 $p = 0.0090$	1, 1, 1 $p = 0.0090$	1, 1, 1 $p = 0.0090$	1, 1, 1 $p = 0.0090$
GAIT	0.60, 0.80, 0.70 $p = 0.4647$	0.40, 1, 0.70 $p = 0.4647$	0.80, 0.80, 0.80 $p = 0.1171$	1, 0.80, 0.90 $p = 0.0282$	1, 0.80, 0.90 $p = 0.0282$	1, 0.60, 0.80 $p = 0.1171$
	0.80, 0.60, 0.70 $p = 0.6015$	0.60, 0.80, 0.70 $p = 0.6015$	1, 1, 1 $p = 0.0090$	1, 1, 1 $p = 0.0090$	1, 1, 1 $p = 0.0090$	1, 1, 1 $p = 0.0090$
	0.60, 0.80, 0.70 $p = 0.4647$	0.80, 0.80, 0.80 $p = 0.0758$	0.80, 0.60, 0.70 $p = 0.7540$	1, 0.60, 0.80 $p = 0.4647$	1, 0.40, 0.70 $p = 0.7540$	0.60, 0.60, 0.60 $p = 0.7540$
GAIT2	0.60, 0.60, 0.60 $p = 0.9168$	1, 0.60, 0.80 $p = 0.2505$	0.80, 0.80, 0.80 $p = 0.2505$	1, 0.60, 0.80 $p = 0.3472$	0.80, 0.40, 0.60 $p = 0.8335$	0.40, 0.80, 0.60 $p = 1$
	0.80, 0.60, 0.70 $p = 0.4647$	1, 0.40, 0.70 $p = 0.6015$	0.60, 0.60, 0.60 $p = 0.9168$	0.60, 0.60, 0.60 $p = 0.9168$	0.60, 0.60, 0.60 $p = 0.5970$	0.60, 0.60, 0.60 $p = 0.9157$
FANT	0.66, 0.66, 0.66 $p = 0.4015$	0.88, 0.55, 0.72 $p = 0.2696$	0.77, 0.66, 0.72 $p = 0.1451$	0.77, 0.66, 0.72 $p = 0.1451$	0.77, 0.66, 0.72 $p = 0.1222$	0.66, 0.77, 0.72 $p = 0.1023$
RATS	0.75, 0.55, 0.65 $p = 0.2792$	0.7, 0.65, 0.675 $p = 0.2133$	0.55, 0.65, 0.60 $p = 0.2674$	0.75, 0.50, 0.625 $p = 0.4488$	0.70, 0.50, 0.60 $p = 0.7867$	0.45, 0.75, 0.60 $p = 0.4651$
WORMS	0.61, 0.67, 0.65 $p = 0.0405$	0.60, 0.71, 0.66 $p = 0.0402$	0.65, 0.68, 0.67 $p = 0.0340$	0.64, 0.70, 0.67 $p = 0.0271$	0.65, 0.70, 0.68 $p = 0.0177$	0.65, 0.70, 0.68 $p = 0.0096$
HOUSE	0.70, 0.60, 0.65 $p = 0.1941$	0.75, 0.55, 0.65 $p = 0.1762$	0.65, 0.60, 0.62 $p = 0.1941$	0.70, 0.60, 0.65 $p = 0.1440$	0.75, 0.50, 0.62 $p = 0.1677$	0.70, 0.55, 0.62 $p = 0.1198$
PAF	0.76, 0.88, 0.82 $p < 0.0001$	0.80, 0.84, 0.82 $p < 0.0001$	0.80, 0.80, 0.80 $p < 0.0001$	0.92, 0.72, 0.82 $p < 0.0001$	0.96, 0.68, 0.82 $p < 0.0001$	0.92, 0.68, 0.80 $p = 0.0002$

The previous experiments for PE were repeated using the new approach: PE2. These additional results are shown in Table 3.

Table 3. Classification results achieved using the new approach, PE2, applied to all the time series databases. Classification performance is quantified with three parameters: sensitivity, specificity, and accuracy.

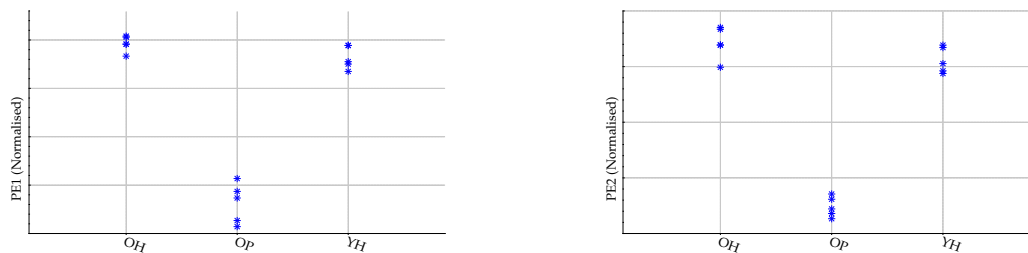
Dataset	$m = 3$	4	5	6	7	8
BONN	0.93, 0.89, 0.91 $p < 0.0001$	0.93, 0.89, 0.91 $p < 0.0001$	0.90, 0.97, 0.93 $p < 0.0001$	0.91, 0.90, 0.90 $p < 0.0001$	0.46, 0.79, 0.62 $p = 0.1782$	0.91, 0.81, 0.86 $p < 0.0001$
	1, 1, 1 $p = 0.0090$	1, 1, 1 $p = 0.0090$	1, 1, 1 $p = 0.0090$	1, 1, 1 $p = 0.0090$	1, 1, 1 $p = 0.0090$	1, 1, 1 $p = 0.0090$
GAIT	0.60, 0.80, 0.70 $p = 0.2505$	0.60, 0.80, 0.70 $p = 0.2505$	1, 0.80, 0.90 $p = 0.0472$	0.60, 0.80, 0.70 $p = 0.9168$	1, 0.40, 0.70 $p = 0.9168$	0.80, 0.80, 0.80 $p = 0.2505$
	1, 1, 1 $p = 0.0090$	1, 1, 1 $p = 0.0090$	1, 1, 1 $p = 0.0090$	1, 1, 1 $p = 0.0090$	1, 1, 1 $p = 0.0090$	1, 1, 1 $p = 0.0090$
	0.60, 0.80, 0.70 $p = 0.4647$	0.80, 0.80, 0.80 $p = 0.0758$	0.80, 0.60, 0.70 $p = 0.6015$	0.60, 1, 0.80 $p = 0.6015$	0.40, 1, 0.70 $p = 0.6015$	0.60, 0.60, 0.60 $p = 0.7373$
GAIT2	0.60, 0.60, 0.60 $p = 0.9168$	1, 0.60, 0.80 $p = 0.2505$	0.60, 0.60, 0.60 $p = 0.9168$	0.60, 1, 0.80 $p = 0.4647$	0.40, 0.80, 0.70 $p = 1$	0.60, 0.60, 0.60 $p = 1$
	0.80, 0.60, 0.70 $p = 0.4647$	1, 0.40, 0.70 $p = 0.6015$	0.60, 0.80, 0.70 $p = 0.7540$	0.60, 0.60, 0.60 $p = 0.9168$	0.60, 0.60, 0.60 $p = 0.5970$	0.40, 0.60, 0.50 $p = 0.9112$
FANT	0.65, 0.60, 0.62 $p = 0.0514$	0.9, 0.50, 0.70 $p = 0.0619$	0.70, 0.60, 0.65 $p = 0.0883$	0.85, 0.50, 0.67 $p = 0.0483$	0.70, 0.80, 0.75 $p = 0.0024$	0.45, 0.75, 0.60 $p = 0.1595$
RATS	0.66, 0.66, 0.66 $p = 0.4015$	0.77, 0.77, 0.77 $p = 0.0469$	0.77, 0.77, 0.77 $p = 0.0379$	0.77, 0.77, 0.77 $p = 0.0379$	0.77, 0.77, 0.77 $p = 0.151$	0.77, 1, 0.88 $p = 0.0091$
WORMS	0.61, 0.67, 0.65 $p = 0.0405$	0.60, 0.71, 0.66 $p = 0.0511$	0.65, 0.68, 0.67 $p = 0.0386$	0.73, 0.68, 0.70 $p < 0.0001$	0.65, 0.52, 0.58 $p = 0.0874$	0.55, 0.51, 0.52 $p = 0.6569$
HOUSE	0.70, 0.60, 0.65 $p = 0.1941$	0.75, 0.55, 0.65 $p = 0.1762$	0.75, 0.50, 0.62 $p = 0.2339$	0.55, 0.70, 0.62 $p = 0.1677$	0.60, 0.65, 0.62 $p = 0.2335$	0.55, 0.70, 0.62 $p = 0.1231$
PAF	0.60, 0.76, 0.68 $p = 0.0141$	0.60, 0.48, 0.54 $p = 0.7341$	0.80, 0.68, 0.74 $p = 0.0064$	0.60, 0.56, 0.58 $p = 0.2815$	0.72, 0.52, 0.62 $p = 0.3669$	0.56, 0.60, 0.58 $p = 0.2815$

In order to better support the addition of the number of admissible/forbidden patterns in the PE method, this number was computed for each dataset. The results of this experiment are shown in Table 4. A great difference between the number of patterns found for the classes under comparison, would suggest that the addition of this number could make a significant contribution to the discriminating power of PE.

Table 4. Actual number of patterns found in each dataset in terms of mean value \pm standard deviation.

	$m = 3$	4	5	6	7	8
BONN	6 \pm 0/6 6 \pm 0/6	24 \pm 0/24 23.03 \pm 1.54/24	113.10 \pm 5.37/120 79.87 \pm 17.00/120	374 \pm 57.17/720 192.42 \pm 61.17/720	780.43 \pm 154.06/5040 369.93 \pm 127.66/5040	1288.65 \pm 269.85/40,320 624.79 \pm 211.50/40,320
GAIT	6 \pm 0/6 6 \pm 0/6 6 \pm 0/6	24 \pm 0/24 24 \pm 0/24 24 \pm 0/24	100.60 \pm 5.88/120 119.60 \pm 0.48/120 119.20 \pm 0.74/120	199 \pm 25.17/720 466.80 \pm 15.52/720 429.80 \pm 22.43/720	230 \pm 29.27/5040 734.40 \pm 41.62/5040 669.60 \pm 45.23/5040	235.60 \pm 30.18/40,320 803.20 \pm 57.72/40,320 739.20 \pm 43.06/40,320
FANT	6 \pm 0/6 6 \pm 0/6	24 \pm 0/24 24 \pm 0/24	119.80 \pm 0.50/120 119.65 \pm 1.10/120	616.35 \pm 40.47/720 649.65 \pm 59.37/720	2259.95 \pm 282.22/5040 2428.60 \pm 486.31/5040	4470.40 \pm 572.96/40,320 4592.40 \pm 914.47/40,320
RATS	6 \pm 0/6 6 \pm 0/6	15.77 \pm 2.89/24 13.22 \pm 1.81/24	31.55 \pm 10.41/120 25 \pm 7.33/120	54.44 \pm 21.48/720 42.22 \pm 16.55/720	86.78 \pm 38.89/5040 65.44 \pm 29.84/5040	132 \pm 68.39/40,320 95.11 \pm 46.98/40,320
WORMS	5.93 \pm 0.56/6 5.95 \pm 0.48/6	23.59 \pm 2.64/24 23.45 \pm 2.38/24	96.77 \pm 21.11/120 99.80 \pm 24.40/120	251.07 \pm 114.37/720 273.28 \pm 108.67/720	378.43 \pm 188.75/5040 425.19 \pm 117.21/5040	456.52 \pm 208.39/40,320 524.13 \pm 203.89/40,320
HOUSE	6 \pm 0/6 6 \pm 0/6	24 \pm 0/24 23.95 \pm 0.21/24	106.15 \pm 14.75/120 110.40 \pm 11.40/120	358.30 \pm 94.82/720 396.90 \pm 92.47/720	664.90 \pm 155.76/5040 722.90 \pm 142.14/5040	854.30 \pm 121.47/40,320 899.45 \pm 101.64/40,320
PAF	6 \pm 0/6 6 \pm 0/6	24 \pm 0/24 23.52 \pm 0.94/24	108.44 \pm 12.88/120 87.24 \pm 15.42/120	300.48 \pm 68.53/720 199.48 \pm 45.72/720	406.92 \pm 94.72/5040 286.88 \pm 51.14/5040	437.52 \pm 92.95/40,320 335.88 \pm 46.67/40,320

Despite the results shown in Tables 2 and 3, some of them could be debatable due to the low number of subjects, especially in the case of the GAIT database. For this reason, we have included two plots of the PE1 and PE2 values obtained to visually check the validity of the classification results for the GAIT database and for the RATS database, with nine members in each class. These plots are depicted in Figures 3 and 4 respectively. Anyway, it is important to note that the analysis should be viewed in comparative terms. We are not proposing a classifier for this GAIT dataset; we are assessing whether PE2 performs better than PE1 or not.



(a) Range of values for PE1 in one experiment of the GAIT database using three groups of five members each. (b) Range of values for PE2 in one experiment of the GAIT database using three groups of five members each.

Figure 3. Range of values for PE1 and PE2 using the optimal m value to achieve the maximum possible classification accuracy in each case, using the three classes in the GAIT database (OH: old, healthy; OP: old, Parkinson's; YH: young, healthy). The OP class is clearly distinguished from the other two in both cases. The other two groups overlap in one or two class members.



(a) Range of values for PE1 in one experiment of the RATS database using two groups of nine members each. (b) Range of values for PE2 in one experiment of the RATS database using two groups of nine members each.

Figure 4. Range of values for PE1 and PE2 using the optimal m value to achieve the maximum possible classification accuracy using the two classes in the RATS database (High: high salt diet; Low: low salt diet). There is an overlapping of three or four members in PE1 for both classes, and that is surely why PE1 did not achieve statistical significance. For PE2, all the members in class High can be classified correctly, with an overlapping of the two top members of the Low class. This corresponds to the numerical results 0.77, 1, and 0.88 in Table 3.

A summary of the best statistically significant performance achieved using the two approaches assessed (PE1, PE2) is shown in Table 5 for real records. As hypothesised, the combination of PE and the actual number of patterns, the newly proposed scheme PE2, achieved the highest accuracy, and was more robust (only one dataset was not significant), than PE alone. Since in a previous study [15], WPE exhibited the best performance of a group of PE algorithm improvements, Table 5 also includes the results of applying this method, along with a denormalised version (WPE + number of patterns found) using the same approach as for PE2, since WPE also enables the computation of the measure without including the number of expected patterns. WPE has also outperformed PE in other studies, such as in [53,54].

Table 5. Summary of the performances achieved by PE1 and PE2, including comparison with the same approach applied to WPE. Only the best significant case is reported, including the corresponding m value, in terms of classification accuracy only. If statistical significance was not achieved, the performance was labelled NS.

Dataset	PE1	PE2	WPE	WPE+Patterns
BONN	$m = 3$ 0.91	$m = 5$ 0.93	$m = 3$ 0.80	$m = 3$ 0.96
GAIT	$m = 7$ 1 0.90 1	$m = 5$ 1 0.90 1	$m = 5$ 0.90 NS 0.90	$m = 4$ 0.90 0.80 1
FANT	NS	$m = 7$ 0.75	$m = 8$ 0.82	$m = 6$ 0.77
RATS	NS	$m = 8$ 0.88	$m = 3$ 0.77	$m = 3$ 0.94
WORMS	$m = 8$ 0.68	$m = 6$ 0.70	$m = 6$ NS	$m = 6$ NS
HOUSE	NS	NS	$m = 8$ 0.72	$m = 8$ 0.90
PAF	$m = 6$ 0.82	$m = 5$ 0.74	$m = 4$ 0.80	$m = 4$ 0.86

Other factors that are taken into account to assess the utility of an entropy measure are its dependence on parameters or artefacts, such as time series length and robustness against noise. The results of a length influence analysis are shown in Table 6. In this case, PE1 and PE2 were compared using m values for which their performances were most similar and significant, and datasets with at least 750 samples in their time series.

Table 6. Accuracy achieved for different lengths N . Values tested were 50, 250, 500, and 750. Datasets had to be at least 750 samples long. Embedded dimension was chosen in order to get the most similar performances for PE1 and PE2, provided the classification had statistical significance.

Dataset	Method	$N = 50$	250	500	750
BONN ($m = 6$)	PE1	0.67	0.78	0.83	0.83
	PE2	0.62	0.65	0.73	0.74
GAIT ($m = 5$)	PE1	(0.60, 0.60, 0.70)	(0.70, 0.70, 0.70)	(1, 0.70, 1)	(1, 0.70, 1)
	PE2	(0.50, 0.60, 0.70)	(0.80, 0.70, 0.70)	(1, 0.70, 1)	(1, 0.80, 1)
WORMS ($m = 5$)	PE1	0.61	0.61	0.63	0.65
	PE2	0.59	0.62	0.68	0.71
Synthetic ($m = 6$)	PE1	0.56 ± 0.01	0.77 ± 0.03	0.92 ± 0.04	0.96 ± 0.02
	PE2	0.60 ± 0.01	0.90 ± 0.01	0.91 ± 0.02	0.99 ± 0.01

The results of the noise robustness test are shown in Table 7. It is important to note that the time series were probably already affected by noise, whose level was not known. Therefore, this test considered the signals completely free of noise when computing the resulting SNR, but in reality it should be considered lower in practical terms. The levels tested were 40, 30, 25, 20, and 15 dB SNR. The noise was a random uniform time series added to the experimental datasets, with 10 realisations for each test.

Table 7. Accuracy achieved for different levels of synthetic noise. Values tested were 40, 30, 25, 20, and 15dB SNR. Embedded dimension was chosen in order to get the most similar performances of PE1 and PE2, provided the classification had statistical significance. GAIT results only correspond to a single case (between healthy young and old adults), since the others were 1 for all the tests.

Dataset	Method	SNR = 40 dB	30 dB	25 dB	20 dB	15 dB
BONN ($m = 6$)	PE1	0.87 ± 0.00	0.81 ± 0.00	0.78 ± 0.01	0.78 ± 0.01	0.80 ± 0.00
	PE2	0.87 ± 0.00	0.78 ± 0.00	0.78 ± 0.01	0.83 ± 0.01	0.83 ± 0.01
GAIT ($m = 5$)	PE1	0.80 ± 0.10	0.63 ± 0.05	0.80 ± 0.10	0.66 ± 0.11	0.63 ± 0.05
	PE2	0.80 ± 0.00	0.76 ± 0.05	0.80 ± 0.00	0.73 ± 0.05	0.70 ± 0.00
WORMS ($m = 5$)	PE1	0.66 ± 0.01	0.65 ± 0.01	NS	NS	NS
	PE2	0.66 ± 0.01	0.65 ± 0.01	NS	NS	NS
PAF ($m = 5$)	PE1	NS	NS	NS	NS	NS
	PE2	0.72 ± 0.00	0.70 ± 0.00	0.70 ± 0.02	0.70 ± 0.01	0.71 ± 0.01
Synthetic ($m = 6$)	PE1	1 ± 0.00	0.85 ± 0.22	0.89 ± 0.01	0.78 ± 0.03	NS
	PE2	0.99 ± 0.00	0.83 ± 0.23	0.83 ± 0.02	0.74 ± 0.06	NS

4. Discussion

There was a clear correlation in Table 1 between differences in patterns found and accuracy achieved using PE2 in synthetic records. Since the ordinal patterns were generated using a length of 3, for $m = 3$, and its multiple $m = 6$, the performance of PE1 was high, but far lower for other m values used in the computation. Those patterns scarcely or did not at all, at $m = 3$, give rise to a set of forbidden patterns for $m > 3$. The number of forbidden patterns grows superexponentially [20], and that is why the differences between classes became more apparent for higher m values, and only PE2 was able to take advantage of this effect.

PE2 yielded a higher classification accuracy than PE1 for all the experimental real datasets except PAF, and it was equal in the GAIT case (Table 5). Moreover, PE1 was unable to find differences in three out of the seven datasets, whereas PE2 only failed in one. Therefore, it can arguably be concluded that PE2 has a greater discriminating power than PE1. The experimental dataset was reasonably very varied and diverse, chosen from publicly available repositories, and used in other similar works.

It is important to note that the PE1 and PE2 results for low m values were very similar. This is probably due to the fact that there is no significant difference in terms of number of patterns found between classes for such low values. In fact, these differences become more apparent beyond $m = 5$ (Table 4).

Specifically, results for the BONN database were $Se = 0.93$, $Sp = 0.90$, and $Acc = 0.91$ for PE with $m = 3$, and 0.90 , 0.97 , and 0.93 respectively, for PE2 with $m = 5$. This is not a great difference, probably because the PE1 performance was already very good, despite the clear differences in terms of patterns found. The situation was the same for the GAIT dataset results. For the FANT dataset, PE1 did not find significant differences for all the m values tested, but PE2 did for $m = 6, 7$, precisely where the relative differences between the two classes in terms of patterns found was the highest. This was also the case for the RATS dataset: PE1 failed, but PE2 found differences, with higher accuracy correlated with higher relative differences in the number of patterns found; the maximum performance being reached at $m = 8$. The separability of the classes in the WORMS dataset was slightly improved with PE2. This could be due to the high variability in the number of patterns found, quantified in terms of standard deviation (Table 4). The records in the HOUSE dataset were not distinguished by any of the measures tested. This was the only case. These records, as can be seen in Figure 2, are very dichotomic, and it is very likely that just a few ordinal patterns monopolise most of the matches, making it difficult to correctly capture the differences between records [26]. In other words, it is not only the number of patterns found, as with such a regular time series, the motif histogram will be extremely biased.

The results for WPE also confirmed that the number of actual patterns found carries significant discriminant information. This method has already demonstrated its performance in previous studies [12,15], and the replacement of the weights as the normalisation factor by the number of different patterns found, as for PE (WPE + patterns in Table 5), further improved the WPE performance in most of the experiments. This fact suggests that the PE2 approach can be a transversal improvement that could be applied to many methods at the normalisation stage in order to enhance their discriminating power, not just as a method on its own.

A case that deserved further investigation was the PAF dataset, since this was the only case with better PE1 performance, despite significant differences and low variability in number of patterns between the two classes (Table 4). The hypothesis was that, somehow, the information provided by ordinal patterns and number of patterns cancelled each other out. A straightforward approach to solve this problem was to change the arithmetic operation in which T was involved, use a multiplication instead of a division. The PAF experiments were repeated with this PE2 algorithm tweak, whose results are shown in Table 8. PE2 performance improved significantly and again was better than PE1.

Table 8. Classification results achieved using the new approach, PE2, applied to the PAF dataset, and $\mathbf{p} \leftarrow p_i \leftarrow c_i * T$ instead of $\mathbf{p} \leftarrow p_i \leftarrow \frac{c_i}{T}$.

Dataset	$m = 3$	4	5	6	7	8
PAF	0.68, 0.72, 0.70 $p = 0.0080$	0.64, 0.76, 0.70 $p = 0.0067$	0.72, 0.96, 0.84 $p < 0.0001$	0.72, 0.88, 0.80 $p < 0.0001$	0.72, 0.92, 0.82 $p = 0.0003$	0.68, 0.84, 0.76 $p = 0.0009$

The results in Table 8 confirm that the patterns found provide discriminant information, but maybe in some specific cases the approach should be slightly different in terms of integration with ordinal patterns. This experiment was repeated for some records that already exhibited good performances using the initial approach described in Algorithm 1. These additional results, shown in Table 9, confirm PE2 is a robust approach in general, and only a few specific cases, such as that of PAF, need a more customised scheme. However, even with both approaches, it was impossible to find significant differences for the HOUSE database.

Although the desnormalisation can be detrimental for low m values when not all the records have the same length, for greater m values the forbidden patterns become more apparent and the additional information provided should make PE2 outperform PE1. In other words, in the worst case PE2 and PE1 should be equivalent in terms of classification accuracy, but for real time series, which usually exhibit some degree of determinism, PE2 will yield the best performance. With the two approaches proposed, PE2 is a clear winner over PE1.

Table 9. Classification results achieved using the new approach PE2, applied to other datasets, and $\mathbf{p} \leftarrow p_i \leftarrow c_i * T$ instead of $\mathbf{p} \leftarrow p_i \leftarrow \frac{c_i}{T}$.

Dataset	$m = 3$	4	5	6	7	8
BONN	0.93, 0.90, 0.91 $p < 0.0001$	0.49, 0.84, 0.66 $p = 0.0137$	0.92, 0.92, 0.92 $p < 0.0001$	0.93, 0.93, 0.93 $p < 0.0001$	0.92, 0.94, 0.93 $p < 0.0001$	0.92, 0.94, 0.93 $p < 0.0001$
FANT	0.65, 0.60, 0.62 $p = 0.0547$	0.55, 0.75, 0.65 $p = 0.0547$	0.75, 0.60, 0.67 $p = 0.0398$	0.80, 0.45, 0.62 $p = 0.4488$	0.60, 0.55, 0.57 $p = 0.7660$	0.85, 0.45, 0.65 $p = 0.1440$
WORMS	0.60, 0.68, 0.65 $p = 0.0494$	0.63, 0.70, 0.67 $p = 0.0015$	0.68, 0.51, 0.58 $p = 0.1698$	0.67, 0.59, 0.62 $p = 0.1301$	0.68, 0.64, 0.66 $p = 0.0564$	0.70, 0.65, 0.67 $p = 0.0300$

Despite the $\mathbf{p} \leftarrow p_i \leftarrow c_i * T$ change applied to the PAF case, the new method proposed is still that described initially in Algorithm 1. This final adjustment hints that there might not be a single way to integrate the information of the relative frequency of motifs with its absolute number, and other studies will be necessary to find a better combination and a more optimised algorithm. We tried to keep a very similar algorithm to PE to facilitate adoption and implementation, but there is room for

PE2 improvement with regard to generalisation, performance, and even normalisation that will have to be addressed in further studies.

The results for the GAIT2 database were not significant (all GAIT records cut short to 200 samples), and therefore, were far worse than those for GAIT. It could be argued that differences in the GAIT case were mainly due to differences in lengths, and this could be the case when comparing old–healthy with old–Parkinson’s, and young–healthy with old–Parkinson’s, since the lengths were 800–200 originally in both cases. However, the comparison of old–healthy with young–healthy became not possible as well, and this pair had the same length in the GAIT database. Therefore, it can be arguably stated that the lack of significance was due to the insufficient length, since 200 samples are borderline [26], not the lack of differences in length, another factor that requires further research. Anyway, the important point is the comparative analysis between PE1 and PE2.

It could also be hypothesised that the differences in number of patterns found would suffice to classify the records. In order to assess this point, a few experiments were repeated using the complete PE2 method and only the number of actual patterns found as single classification features. Although in some cases the performances of both approaches were very similar (for example, for the BONN database, with $m = 6$, the performances were PE2 = 0.935 vs. 0.930), in others, they were quite different (For the RATS dataset, with $m = 6$, the performances were PE2 = 0.78 vs. 0.72). Since the computational cost of both approaches is almost the same (the algorithm should be run almost completely; only the final Shannon calculation could be omitted), such a simplification is not advisable because not all the differences only lie on the number of admissible patterns.

The parameter and noise influence analysis provided an additional insight into the PE2 approach capabilities. The comparative length influence analysis between PE1 and PE2 in Table 6 showed that both metrics exhibit a similar robustness against N variations and short lengths. For extremely short lengths, 50 samples, it was not possible to find differences between classes, which is understandable since 50 samples do not suffice to provide a reasonable pattern frequency estimation. At 250 samples, most of the results became significant, but it was at 500 samples where both methods provided significant classifications in all cases, except PE1 for the WORMS database. With 750 samples, performances were really close to those achieved with the entire records.

The performances of PE1 and PE2 with regard to noise interference seemed very similar, except for the PAF dataset (Table 7). This could be due to the fact that noise impact can also be considered in terms of new ordinal patterns; random time series usually contain all the possible $m!$ patterns as a sign of non-determinism. Therefore, since the difference between PE1 and PE2 lies on the number of patterns found, it can arguably be assumed that, in general, as the noise level increases, PE2 will converge to the PE1 performance, since the differences in terms of number of patterns will become blurred.

PE2 is derived from PE1, and without specific studies yet, it would be sensitive to assume that some PE1 drawbacks can be inherited by PE2. Thus, PE1 could be influenced by amplitude differences in the ordinal patterns [15], by ties [17], or by the record length [26]. A similar set of characterisation studies would be necessary specifically for PE2 to shed some light on these possible issues.

5. Conclusions

Every year quite a number of new tools to quantify the dynamical features of time series are described in the scientific literature. These new tools are claimed to be more efficient in algorithmic terms, more sensitive, more robust, or less dependent on input parameters, among many other possible benefits.

As such, PE2 was introduced as an improvement over PE by taking advantage of the differences between time series classes in terms of numbers of different patterns actually found. The present study assessed this discriminant power using several real-life datasets, and we could conclude that the discriminating capabilities of ordinal patterns’ relative frequencies and their counts are clearly complementary and synergistic. This led us to try to combine both measures in a single method to take advantage of their strengths and simultaneously minimise their possible weaknesses. The scheme used for PE using dynamic lists provided the algorithmic template to merge ordinal pattern and pattern

number information together by changing the way histogram bins were normalised, and keeping the algorithm's implementation simple and similar to its ancestor PE.

According to the results obtained, the PE2 approach can be considered a very promising tool in the field of symbolic dynamics. It should not be claimed to be a *cure-all* method, but the classification performance confirmed the hypothesis, and PE2 seems to be able to seamlessly exploit the synergy between PE and the number of patterns found in most cases. It was clearly more robust, since statistical significance was reached in six out of the seven datasets, two more than with PE1 (PE). It also achieved the maximum performance of the two methods tested in five cases, or six if the final PE2 algorithm tweaking is considered (Section 4).

The PE2 algorithm is just a little bit more complex than that of PE, but more memory efficient, since it only needs to store the patterns found, not all the possible $m!$ ones. Moreover, the algorithm introduced can be easily further optimised. In addition to implementation issues to save memory requirements or computational cost, the algorithm could be improved in terms of ordinal and pattern number influences on the final calculations by using a normalisation or weighting scheme based on an additional parameter, such as $A/1 - A$, as in [14]. Additionally, other factors could be included as additional symbols in the motifs. This way, other properties such as sequence amplitude, would become part of the comparisons. In fact, we included the q parameter in the PE2 method, as described in [13], and the classification performance of PE2 increased in some cases. The experiments with WPE and WPE+Patterns also confirmed this point.

This new approach will need further studies using other databases and other integration schemes. The influences of equal values in the subsequences [17], and time delay τ , could also be characterised. Further integration with other PE improvements could be worth exploring [6,12–14,16] in future.

Acknowledgments: No funding was received to support this research.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Bandt, C.; Pompe, B. Permutation Entropy: A Natural Complexity Measure for Time Series. *Phys. Rev. Lett.* **2002**, *88*, 174102. [[CrossRef](#)] [[PubMed](#)]
2. Zanin, M.; Zunino, L.; Rosso, O.A.; Papo, D. Permutation Entropy and Its Main Biomedical and Econophysics Applications: A Review. *Entropy* **2012**, *14*, 1553–1577. [[CrossRef](#)]
3. Li, J.; Yan, J.; Liu, X.; Ouyang, G. Using Permutation Entropy to Measure the Changes in EEG Signals During Absence Seizures. *Entropy* **2014**, *16*, 3049–3061. [[CrossRef](#)]
4. Ravelo-García, A.G.; Navarro-Mesa, J.L.; Casanova-Blancas, U.; Martín-González, S.; Quintana-Morales, P.; Guerra-Moreno, I.; Canino-Rodríguez, J.M.; Hernández-Pérez, E. Application of the Permutation Entropy over the Heart Rate Variability for the Improvement of Electrocardiogram-based Sleep Breathing Pause Detection. *Entropy* **2015**, *17*, 914–927. [[CrossRef](#)]
5. Cuesta-Frau, D.; Miró-Martínez, P.; Ultra-Crespo, S.; Jordán-Núñez, J.; Vargas, B.; González, P.; Varela-Entrecanales, M. Model Selection for Body Temperature Signal Classification Using Both Amplitude and Ordinality-Based Entropy Measures. *Entropy* **2018**, *20*. [[CrossRef](#)]
6. Cuesta-Frau, D.; Miró-Martínez, P.; Ultra-Crespo, S.; Jordán-Núñez, J.; Vargas, B.; Vigil, L. Classification of glucose records from patients at diabetes risk using a combined permutation entropy algorithm. *Comput. Methods Programs Biomed.* **2018**, *165*, 197–204. [[CrossRef](#)]
7. Gao, Y.; Vilecco, F.; Li, M.; Song, W. Multi-Scale Permutation Entropy Based on Improved LMD and HMM for Rolling Bearing Diagnosis. *Entropy* **2017**, *19*. [[CrossRef](#)]
8. Wang, X.; Si, S.; Wei, Y.; Li, Y. The Optimized Multi-Scale Permutation Entropy and Its Application in Compound Fault Diagnosis of Rotating Machinery. *Entropy* **2019**, *21*. [[CrossRef](#)]
9. Wang, Q.C.; Song, W.Q.; Liang, J.K. Road Flatness Detection Using Permutation Entropy (PE). In *Applied Mechanics and Materials*; Trans Tech Publications Ltd.: Zürich, Switzerland, 2015; Volume 721, pp. 420–423.
10. Glynn, C.C.; Konstantinou, K.I. Reduction of randomness in seismic noise as a short-term precursor to a volcanic eruption. *Nat. Sci. Rep.* **2016**, *6*, 37733. [[CrossRef](#)]

11. Zhang, Y.; Shang, P. Permutation entropy analysis of financial time series based on Hill's diversity number. *Commun. Nonlinear Sci. Numer. Simul.* **2017**, *53*, 288–298. [[CrossRef](#)]
12. Fadlallah, B.; Chen, B.; Keil, A.; Príncipe, J. Weighted-permutation entropy: A complexity measure for time series incorporating amplitude information. *Phys. Rev. E* **2013**, *87*, 022911. [[CrossRef](#)] [[PubMed](#)]
13. Xiao-Feng, L.; Yue, W. Fine-grained permutation entropy as a measure of natural complexity for time series. *Chin. Phys. B* **2009**, *18*, 2690. [[CrossRef](#)]
14. Azami, H.; Escudero, J. Amplitude-aware permutation entropy: Illustration in spike detection and signal segmentation. *Comput. Methods Programs Biomed.* **2016**, *128*, 40–51. [[CrossRef](#)] [[PubMed](#)]
15. Cuesta-Frau, D. Permutation Entropy: Influence of amplitude information on time series classification performance. *Math. Biosci. Eng.* **2019**, *5*, 1–16. [[CrossRef](#)]
16. Bian, C.; Qin, C.; Ma, Q.D.Y.; Shen, Q. Modified Permutation-entropy analysis of heartbeat dynamics. *Phys. Rev. E* **2012**, *85*, 021906. [[CrossRef](#)] [[PubMed](#)]
17. Cuesta-Frau, D.; Varela-Entrecanales, M.; Molina-Picó, A.; Vargas, B. Patterns with Equal Values in Permutation Entropy: Do They Really Matter for Biosignal Classification? *Complexity* **2018**, *2018*, 1–15. [[CrossRef](#)]
18. Zunino, L.; Olivares, F.; Scholkmann, F.; Rosso, O.A. Permutation entropy based time series analysis: Equalities in the input signal can lead to false conclusions. *Phys. Lett. A* **2017**, *381*, 1883–1892. [[CrossRef](#)]
19. Manis, G.; Aktaruzzaman, M.; Sassi, R. Bubble Entropy: An Entropy Almost Free of Parameters. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 2711–2718. [[CrossRef](#)]
20. Amigó, J. *Permutation Complexity in Dynamical Systems*; Springer Science & Business Media: Berlin, Germany, 2010. [[CrossRef](#)]
21. Amigó, J.M.; Kocarev, L.; Szczepanski, J. Order patterns and chaos. *Phys. Lett. A* **2006**, *355*, 27–31. [[CrossRef](#)]
22. Zunino, L.; Zanin, M.; Tabak, B.M.; Pérez, D.G.; Rosso, O.A. Forbidden patterns, permutation entropy and stock market inefficiency. *Phys. A Stat. Mech. Its Appl.* **2009**, *388*, 2854–2864. [[CrossRef](#)]
23. Little, D.J.; Kane, D.M. Permutation entropy with vector embedding delays. *Phys. Rev. E* **2017**, *96*, 062205. [[CrossRef](#)] [[PubMed](#)]
24. Riedl, M.; Müller, A.; Wessel, N. Practical considerations of permutation entropy. *Eur. Phys. J. Spec. Top.* **2013**, *222*, 249–262. [[CrossRef](#)]
25. Rosso, O.A.; Carpi, L.C.; Saco, P.M.; Ravetti, M.G.; Plastino, A.; Larrondo, H.A. Causality and the entropy–complexity plane: Robustness and missing ordinal patterns. *Phys. A Stat. Mech. Its Appl.* **2012**, *391*, 42–55. [[CrossRef](#)]
26. Cuesta-Frau, D.; Murillo-Escobar, J.P.; Orrego, D.A.; Delgado-Trejos, E. Embedded Dimension and Time Series Length. Practical Influence on Permutation Entropy and Its Applications. *Entropy* **2019**, *21*. [[CrossRef](#)]
27. KumarSingh, B.; Verma, K.; Thoke, A. Investigations on Impact of Feature Normalization Techniques on Classifier's Performance in Breast Tumor Classification. *Int. J. Comput. Appl.* **2015**, *116*, 11–15. [[CrossRef](#)]
28. Talukder, B.; W. Hipel, K.; W. vanLoon, G. Developing Composite Indicators for Agricultural Sustainability Assessment: Effect of Normalization and Aggregation Techniques. *Resources* **2017**, *6*. [[CrossRef](#)]
29. Tofallis, C. Add or Multiply? A Tutorial on Ranking and Choosing with Multiple Criteria. *INFORMS Trans. Educ.* **2014**, *14*, 109–119. [[CrossRef](#)]
30. Henry, M.; Judge, G. Permutation Entropy and Information Recovery in Nonlinear Dynamic Economic Time Series. *Econometrics* **2019**, *7*. [[CrossRef](#)]
31. Zanin, M. Forbidden patterns in financial time series. *Chaos Interdiscip. J. Nonlinear Sci.* **2008**, *18*, 013119. [[CrossRef](#)]
32. Cuesta-Frau, D.; Molina-Picó, A.; Vargas, B.; González, P. Permutation Entropy: Enhancing Discriminating Power by Using Relative Frequencies Vector of Ordinal Patterns Instead of Their Shannon Entropy. *Entropy* **2019**, *21*. [[CrossRef](#)]
33. Andrzejak, R.G.; Lehnertz, K.; Mormann, F.; Rieke, C.; David, P.; Elger, C.E. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Phys. Rev. E* **2001**, *64*, 061907. [[CrossRef](#)]
34. Arafat, S.; Bellegdi, S. Automatic Detection of Epilepsy Using EEG Energy and Frequency Bands. *Int. J. Appl. Math. Electron. Comput.* **2017**, *1*, 36–41. [[CrossRef](#)]
35. Hussain, L.; Aziz, W.; Alowibdi, J.S.; Habib, N.; Rafique, M.; Saeed, S.; Kazmi, S.Z.H. Symbolic time series analysis of electroencephalographic (EEG) epileptic seizure and brain dynamics with eye-open and eye-closed subjects during resting states. *J. Physiol. Anthropol.* **2017**, *36*, 21. [[CrossRef](#)] [[PubMed](#)]

36. Cuesta-Frau, D.; Miró-Martínez, P.; Jordán-Núñez, J.; Oltra-Crespo, S.; Molina-Picó, A. Noisy EEG signals classification based on entropy metrics. Performance assessment using first and second generation statistics. *Comput. Biol. Med.* **2017**, *87*, 141–151. [[CrossRef](#)]
37. Ashok, S. Epilepsy detection from EEG signals: A review. *J. Med Eng. Technol.* **2018**, *42*, 1–13. [[CrossRef](#)]
38. Molina-Picó, A.; Cuesta-Frau, D.; Aboy, M.; Crespo, C.; Miró-Martínez, P.; Oltra-Crespo, S. Comparative Study of Approximate Entropy and Sample Entropy Robustness to Spikes. *Artif. Intell. Med.* **2011**, *53*, 97–106. [[CrossRef](#)] [[PubMed](#)]
39. Goldberger, A.L.; Amaral, L.A.N.; 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*, 215–220. [[CrossRef](#)] [[PubMed](#)]
40. Hausdorff, J.; Purdon, P.; Peng, C.K.; Ladin, Z.; Wei, J.; Goldberger, A. Fractal dynamics of human gait: Stability of long-range correlations in stride interval fluctuations. *J. Appl. Physiol. (Bethesda, Md. 1985)* **1996**, *80*, 1448–1457. [[CrossRef](#)]
41. Baumert, M.; Czippelova, B.; Ganesan, A.; Schmidt, M.; Zaunseder, S.; Javorka, M. Entropy Analysis of RR and QT Interval Variability during Orthostatic and Mental Stress in Healthy Subjects. *Entropy* **2014**, *16*, 6384–6393. [[CrossRef](#)]
42. Xia, Y.; Yang, L.; Zunino, L.; Shi, H.; Zhuang, Y.; Liu, C. Application of Permutation Entropy and Permutation Min-Entropy in Multiple Emotional States Analysis of RRI Time Series. *Entropy* **2018**, *20*, 148. [[CrossRef](#)]
43. Iyengar, N.; Peng, C.K.; Morin, R.; Goldberger, A.L.; Lipsitz, L.A. Age-related alterations in the fractal scaling of cardiac interbeat interval dynamics. *Am. J. -Physiol.-Regul. Integr. Comp. Physiol.* **1996**, *271*, R1078–R1084. [[CrossRef](#)] [[PubMed](#)]
44. Liu, C.; Li, K.; Zhao, L.; Liu, F.; Zheng, D.; Liu, C.; Liu, S. Analysis of heart rate variability using fuzzy measure entropy. *Comput. Biol. Med.* **2013**, *43*, 100–108. [[CrossRef](#)] [[PubMed](#)]
45. Bugenhagen, S.M.; Cowley, A.W.; Beard, D.A. Identifying physiological origins of baroreflex dysfunction in salt-sensitive hypertension in the Dahl SS rat. *Physiol. Genom.* **2010**, *42*, 23–41. [[CrossRef](#)] [[PubMed](#)]
46. Bagnall, A.; Lines, J.; Bostrom, A.; Large, J.; Keogh, E. The Great Time Series Classification Bake Off: A Review and Experimental Evaluation of Recent Algorithmic Advances. *Data Min. Knowl. Discov.* **2017**, *31*, 606–660. [[CrossRef](#)]
47. Yemini, E.; Jucikas, T.; Grundy, L.J.; Brown, A.E.X.; Schafer, W.R. A database of *Caenorhabditis elegans* behavioral phenotypes. *Nat. Methods* **2013**, *10*, 877–879. [[CrossRef](#)]
48. Brown, A.E.X.; Yemini, E.I.; Grundy, L.J.; Jucikas, T.; Schafer, W.R. A dictionary of behavioral motifs reveals clusters of genes affecting *Caenorhabditis elegans* locomotion. *Proc. Natl. Acad. Sci. USA* **2013**, *110*, 791–796. [[CrossRef](#)]
49. Murray, D.; Liao, J.; Stankovic, L.; Stankovic, V.; Hauxwell-Baldwin, R.; Wilson, C.; Coleman, M.; Kane, T.; Firth, S. A data management platform for personalised real-time energy feedback. In Proceedings of the 8th International Conference on Energy Efficiency in Domestic Appliances and Lighting, Horw-Lucerne, Switzerland, 26–28 August 2015.
50. Moody, G.B.; Goldberger, A.L.; McClennen, S.; Swiryn, S. Predicting the Onset of Paroxysmal Atrial Fibrillation: The Computers in Cardiology Challenge 2001. In Proceedings of the Computers in Cardiology, Rotterdam, The Netherlands, 23–26 September 2001; Volume 28, pp. 113–116. [[CrossRef](#)]
51. Cuesta-Frau, D.; Novák, D.; Burda, V.; Molina-Picó, A.; Vargás, B.; Mraz, M.; Kavalkova, P.; Benes, M.; Haluzik, M. Characterization of Artifact Influence on the Classification of Glucose Time Series Using Sample Entropy Statistics. *Entropy* **2018**, *20*. [[CrossRef](#)]
52. Bürkner, P.C.; Doebler, P.; Holling, H. Optimal design of the Wilcoxon–Mann–Whitney-test. *Biom. J.* **2017**, *59*, 25–40. [[CrossRef](#)]
53. Bian, Z.; Ouyang, G.; Li, Z.; Li, Q.; Wang, L.; Li, X. Weighted-Permutation Entropy Analysis of Resting State EEG from Diabetics with Amnesic Mild Cognitive Impairment. *Entropy* **2016**, *18*. [[CrossRef](#)]
54. Deng, B.; Liang, L.; Li, S.; Wang, R.; Yu, H.; Wang, J.; Wei, X. Complexity extraction of electroencephalograms in Alzheimer’s disease with weighted-permutation entropy. *Chaos Interdiscip. J. Nonlinear Sci.* **2015**, *25*, 043105. [[CrossRef](#)]

