

Classification of Interbeat Interval Time-series Using Attention Entropy

Jiawei Yang, Gulraiz I. Choudhary,
Susanto Rahardja, Fellow, *IEEE*, and Pasi Fränti, Senior Member, *IEEE*

Abstract—Classification of interbeat interval time-series which fluctuates in an irregular and complex manner is very challenging. Typically, entropy methods are employed to quantify the complexity of the time-series for classifying. Traditional entropy methods focus on the frequency distribution of all the observations in a time-series. This requires a relatively long time-series with at least a couple of thousands of data points, which limits their usages in practical applications. The methods are also sensitive to the parameter settings. In this paper, we propose a conceptually new approach called *attention entropy*, which pays attention only to the key observations. Instead of counting the frequency of all observations, it analyzes the frequency distribution of the intervals between the key observations in a time-series. The advantages of the attention entropy are that it does not need any parameter to tune, is robust to the time-series length, and requires only linear time to compute. Experiments show that it outperforms fourteen state-of-the-art entropy methods evaluated by real-world datasets. It achieves average classification accuracy of $AUC=0.71$ while the second-best method, multiscale entropy, achieves $AUC=0.62$ when classifying four groups of people with a time-series length of 100.

Index Terms—Attention entropy, the complexity-loss, peak points, heart rate variability, HRV, RR interval, interbeat interval, time-series, synthetic signals.

1 INTRODUCTION

BIOLOGICAL signals are space, time, or space-time records of biological events such as the heart beating or a muscle contracting [1]. Biological signals including electroencephalogram (EEG) [2], electrocardiogram (ECG) [3], [4], electro-oculography (EOG) [5], surface electromyogram (sEMG) [6], [7], galvanic skin response (GSR) [8], [9] and respiration, are widely used in fields such as clinical disease diagnosis.

Living systems exhibit self-regulating mechanisms that process inputs with a broad range of characteristics [10], [11]. Many biological time-series such as heart rate variability (HRV) also called interbeat intervals extracted from ECG are extremely inhomogeneous, non-stationary, and fluctuate in an irregular and complex manner [12]. Fig. 1 shows four time-series of interbeat intervals from different subjects. We can see that they vary in an irregular manner. HRV is used to physiological analysis, such as depressive disorder analysis [2], stress recognition [13], [14], [15], and affective states analysis [16]. There also has been considerable interest in quantifying the complexity of HRV to uncover hidden information, such as heart failures [17], [18], [19] and coronary artery disease [20]. Typical methods such as multiscale entropy (MSE) [21] and grouped horizontal visibility graph entropy (GHVE) [22] analyze complexity by segmenting the signals into equal-length sub-series and

calculating the entropy based on how frequently the artificial patterns occur extracted from the sub-series. The process of a typical method is illustrated in Fig. 2 (top). Given a time-series X , the method segments it into over-lapping sub-series of equal length, extracts patterns from the sub-series, and then calculates the entropy based on the frequencies of the patterns. The result depends on the length of the sub-series and the definition of the artificial patterns.

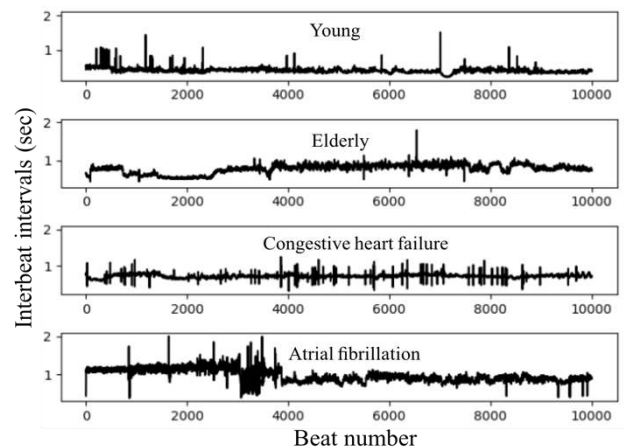


Fig. 1. Interbeat interval time-series from a young subject with age ≤ 55 , an elderly subject with age > 55 , a subject with congestive heart failure (CHF), and a subject with atrial fibrillation (AF).

- Jiawei Yang and Susanto Rahardja are with the School of Marine Science and Technology, Northwestern Polytechnical University, 127 West Youyi Road, Xi'an, Shaanxi 710072, China. ({susantorahardja, jiaweyang}@ieee.org).
- Pasi Fränti is with the College of Big Data and Internet, Shenzhen Technical University, Shenzhen.
- Pasi Fränti, Jiawei Yang, and Gulraiz I. Choudhary are with the School of Computing, University of Eastern Finland, Joensuu 80101, Finland. ({franti, jiawei, gulraiz}@cs.uef.fi).

bins to obtain a dense histogram. Typical entropy methods need a time-series length of at least 30,000 samples to model the data [21]. This takes more than 30 minutes to collect which induces high cost in the clinical diagnosis and therefore limits their usage in the real-world applications.

The second challenge is that it can take considerable time to extract the patterns. Most methods require $O(m^a n^b)$ time, where m is the dimension of the vector (see Section 2), n is the time-series length, and $a, b > 1$. This is a dilemma as the methods need a lot of data to calculate reliable entropy value, but having more data means also more time required. This prevents the use of the methods from large-scale data.

The third challenge is that the artificial patterns also lack clear intuitive interpretation. As a result, the patterns have no direct analytical capability which limits its contribution to the medical analysis of different diseases.

To overcome these challenges, we propose a conceptually new method called *attention entropy*, which pays attention only to the key observations and focuses on how regularly they repeat in the time-series. Fig. 2 (bottom) illustrates the process of computing the attention entropy. Given a time-series X , attention entropy extracts the key patterns and uses the intervals between the key patterns to calculate the entropy value.

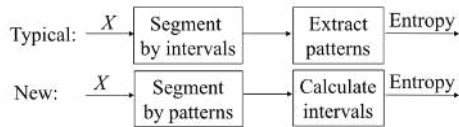


Fig. 2. Main process of calculating the entropy of series.

2 ENTROPY METHODS

Entropy is a quantitative measure of the randomness and disorder of a system. Rudolf Clausius [23] was the first to introduce a mathematical version of the concept to measure the proportion of heat energy transferred from a body to another. Boltzmann and Gibbs [24], [25] extended the concept into statistical mechanics to model the molecular disorder and chaos. Shannon later defined the entropy as the smallest size that a message can be encoded without loss [26]. In this section, we review the entropy measures that are most relevant to our study.

2.1 Existing methods

The process of typical entropy methods has four components as summarized in Fig. 3: (1) convert the original series into another series; (2) construct the sub-series; (3) extract the patterns from the sub-series; (4) analyze the frequency distribution of the patterns. Different entropy methods are based on the different combinations of these four components, as summarized in Table 1.

From Table 1, we can see some entropy methods convert the original series into another series and then segment the converted series into sub-series to extract patterns. For example, spectral entropy [27], average entropy [28], and MSE [21] convert the series using the discrete Fourier

transform [27], the grid [28], and the coarse-graining function [21], respectively.

We can also see that there are three typical methods to construct the sub-series: single value, template vector, and delay vector. They can be formed as $z_i^{m,\tau} = [x_i, x_{i+\tau}, \dots, x_{i+(m-1)\tau}]$ for $1 \leq i \leq n-(m-1)\tau$, where τ is the time delay, and m is the dimension of the vector, given a finite time-series $X = x_1, \dots, x_n$ with the length n . Single value is the case of $z_i^{m,\tau}$ with $m = 1$ and $\tau = 0$. Template vector is the case of $z_i^{m,\tau}$ with $m > 1$ and $\tau = 1$. Delay vector is the case of $z_i^{m,\tau}$ with $m > 1$ and $\tau > 0$.

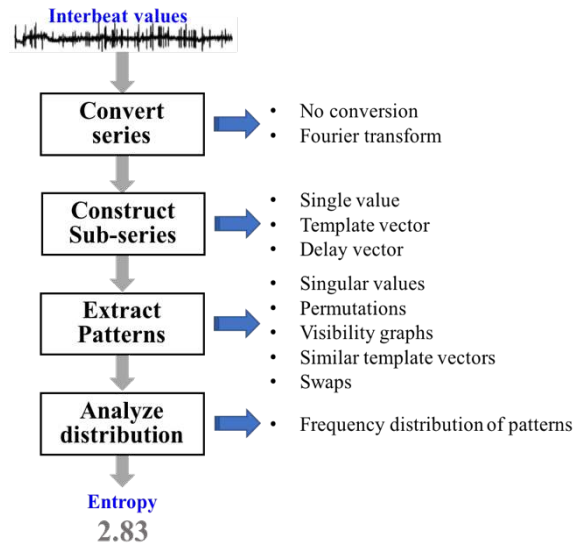


Fig. 3. Typical components of entropy methods.

TABLE 1 Summary of Entropy Methods

Entropy methods	Convert series	Sub-series	Patterns	Frequency of...
Shannon [26]	NO	Single value	Single value	Patterns
Rényi [29]			Single value	
Tsallis [30]		Template vector	Permutation	
Permutation [31]			Similar template vectors	
Approximate [33]			Swaps	
Sample [34]			Visibility graph	
Bubble [35]			Grouped visibility graph	
Horizontal visibility [36]			Delay vector	
Grouped horizontal visibility [22]		Permutation		
SVDE [38]		YES	Single value	
Edge permutation [32]	Similar template vectors			
Spectral [27]	Template vector		Similar template vectors	
Average [28]	NO	-	Peak points	Intervals
Multiscale [21]		-	Peak points	Intervals

Different entropy methods have major difference in the way they extract the patterns from the sub-series. Shannon

entropy [26], Rényi entropy [29], Tsallis entropy [30], spectral entropy [27], and average entropy [28] use the values directly. Permutation entropy [31] and edge permutation entropy (EPE) [32] use the permutations of the rankings of each value in the template vectors as the patterns. Approximate entropy [33], sample entropy [34], and multiscale entropy [21] use similar template vectors as the patterns. Bubble entropy [35] uses the swaps of sorting sub-series with bubble sort algorithm as the patterns. Horizontal visibility entropy (HVE) [36] uses visibility graphs [37] and GHVE [22] uses grouped visibility graphs as the patterns. The singular value decomposition entropy (SVDE) [38] uses the singular values obtained by performing singular value decomposition on the embedding space spanned by the delay vectors as the patterns.

Once the patterns are defined, the entropy values will be calculated by analyzing the frequency distribution of these patterns. Approximate entropy [33] and sample entropy [34] analyze the frequency distribution of the patterns defined with m and $m+1$ dimensional template vector, respectively. They calculate the entropy value from the difference of these two distributions.

From Table 1, we can also see that the proposed attention entropy does not need to convert the series. It uses *peak points* in the series as the patterns. It analyzes the frequency distribution of patterns' intervals, which will be discussed in Section 3.

2.2 Discussion

Each method introduced above has its advantages and disadvantages. Shannon entropy [26], Rényi entropy [29], and average entropy [28] can be applied globally to all data, or locally only to points around specific points [39]. However, they ignore the temporal order of the patterns in the signal [40].

Permutation entropy [31] and edge permutation entropy [32] use the temporal information [39], but they rely on the occurrence of equal values in the sub-series [41]. Approximate entropy [33] has the advantage of lower computational demand and less effect from noise, but it strongly depends on the time-series length and therefore lacks consistency [40]. Sample entropy [34] is invariant to the time-series length and it performs more consistently under various conditions. However, it has a strong dependency on the input parameters [39].

Bubble entropy [35] and GHVE [22] are not sensitive to the parameter settings. However, they have high computational costs, and therefore, they are not practical for large-scale data [35], [36].

MSE [21] is capable of discovering the multiscale feature of data but it requires long time-series to work. SVDE [38] allows analyzing even very short and non-stationary data, but it has high computational costs when applied to large-scale data [38]. Spectral entropy [27] has the advantage of simplicity, but it is sensitive to noise and relies on the assumption that the data error is independent of time [27].

3 ATTENTION ENTROPY

To overcome the shortcomings of the typical entropy methods, we propose *attention entropy*. We first introduce the general principle and then give a suggestion of how to select the key patterns.

3.1 The general principle of attention entropy

Attention entropy is calculated in three main steps: (1) define the key patterns; (2) calculate the intervals between two adjacent key patterns; (3) calculate Shannon entropy of intervals. The difference between classical entropy methods and attention entropy is demonstrated in Fig. 4. Classical frequency-based entropy methods cannot separate Series 1 and 2, as both have the same frequency distribution of the patterns. Attention entropy can do it because the distribution of the intervals of the key patterns (*Apple*) in the series are different.

Formally, given a finite series X , we first define the key pattern Ω . Second, we calculate the intervals $I^\Omega = \{v \mid v = j-i\}$ for any given sub-series u_i, u_k , and u_j of X which satisfy that u_i and u_j match in the pattern Ω , but u_k does not match in Ω for any $i < k < j$. We finally calculate Shannon entropy over I^Ω as the attention entropy.

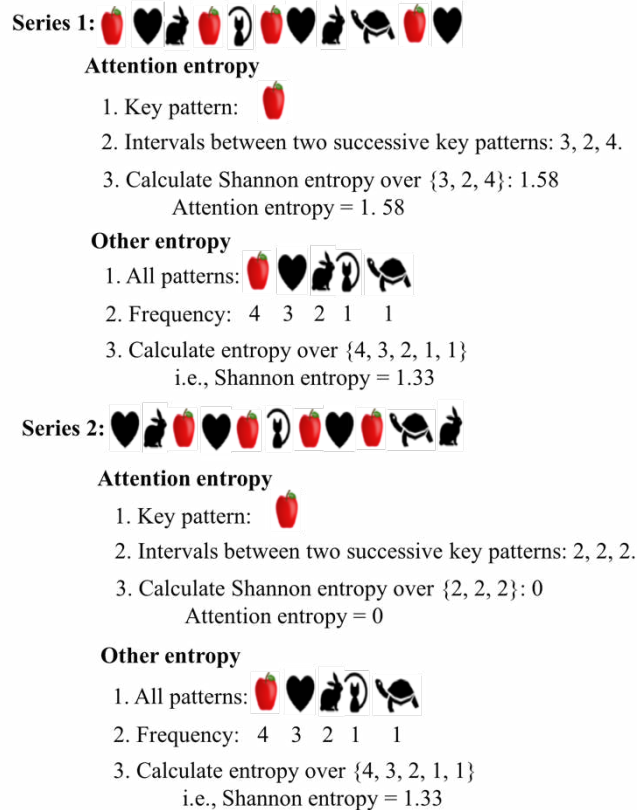


Fig. 4 Difference between attention entropy and other entropies.

3.2 Peak points as the key patterns

We define a point x_i as a *peak point*, including *local maxima* and *local minima*, if it satisfies one of the conditions below:

- $x_{i-1} < x_i$ and $x_i > x_{i+1}$ (x_i is defined as *local maxima*)
- $x_i < x_{i-1}$ and $x_i < x_{i+1}$ (x_i is defined as *local minima*)

If each point in a time-series is considered as one state of a

system, the change of the state can then be seen as the system's adjustment to the environment. A complex system is expected to have a complex process of the state changes when adapting to the environment. The peak points represent the local upper and lower bounds of the state changes. This makes them as the potential key patterns.

A time-series then can be represented by the series of the peak points. We then calculate the intervals between two successive peak points. We consider four cases:

- Intervals of local maxima to local maxima (Max-Max)
- Intervals of local minima to local minima (Min-Min)
- Intervals of local maxima to local minima (Max-Min)
- Intervals of local minima to local maxima (Min-Max)

We can use any one of these four cases individually by calculating the entropy of the respective interval distribution. We can also merge the results by analyzing the four distributions separately and then taking the average of the four individual entropy values. In the rest of the paper, we use this merging strategy as our recommended method and denote it as *Average-4*.

Fig. 5 shows an example of how to calculate the attention entropy when defining peak points as the key patterns. In general, the individual entropy values are not expected to differ much from each other. In most cases, the result is about the same regardless which of the four cases we use. However, using all the four cases brings two additional benefits. First, it can smooth possible abnormalities in the data. Second, we have four times more data. This can potentially make the method work with shorter time-series.

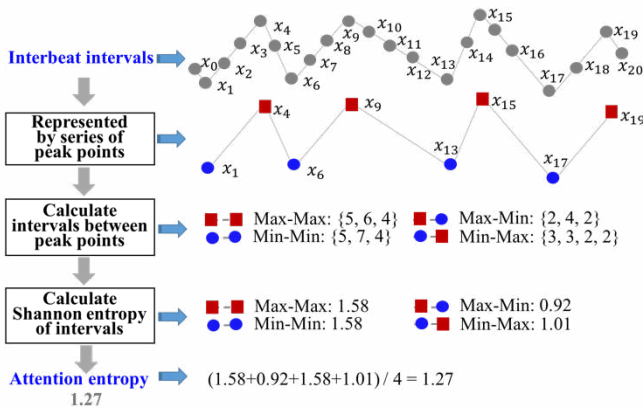


Fig. 5 A simulated time-series $\{x_0, \dots, x_{20}\}$ to illustrate the procedure of calculating attention entropy. First, we find local minima (circle points): $\{x_1, x_6, x_{13}, x_{17}\}$ and local maxima (squared points): $\{x_4, x_9, x_{15}, x_{19}\}$. Second, we calculate the intervals of Max-Max, Min-Min, Max-Min, Min-Max: $\{5, 6, 4\}$, $\{5, 7, 4\}$, $\{2, 4, 2\}$, and $\{3, 3, 2, 2\}$. Third, we calculate their Shannon entropies: 1.58, 1.58, 0.92, and 1.01. Finally, the attention entropy is calculated as the average of these four

entropy values: 1.27.

Fig. 6 shows the expected behavior of the attention entropy; it increases with increasing the randomness of peak points. Fig. 7 shows sample distributions of the intervals among peak points of the four different subjects from Fig. 1. We can see that all intervals of AF are smaller than 10 and the distribution of AF always concentrates on the lower values, leading to low entropy. Some intervals of CHF are bigger than 10 but all of that are smaller than 20, and the distribution of CHF drops faster than young and elderly. The difference between the distributions of young and elderly is less visible from the graphs, but the average of the four entropy values, however, makes the distraction clear (young=2.68, elderly=2.25).

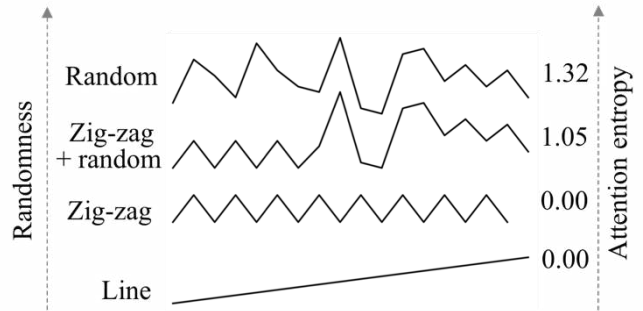


Fig. 6 The more randomly the peaks “^” and “v” appear, the greater is the attention entropy.

Algorithm 1: AttentionEntropy(X, Ω)

Input: X : Time-series of length n , Ω : key patterns

Output: E : Entropy value

FOR $i = 1$ TO n :

IF **matchKeyPatterns**(x_i, Ω) THEN:

interval = i - previous

$F_{interval} = F_{interval} + 1$

previous = i

$E = \text{calculateShannonEntropy}(F)$

3.3 Implementation

Implementation of attention entropy is shown in Algorithm 1. It requires $O(n)$ time, where n is the length of time-series X . The algorithm contains the following steps:

- (1) Detect whether the point is a key pattern;
- (2) Calculate the interval between two key patterns;
- (3) Count the frequencies of all intervals;
- (4) Calculate Shannon entropy over frequencies of all intervals.

When a point x_i is detected as a key pattern, we calculate the *interval* as $i - j$, where x_j is the previous key pattern before x_i . We store the counts of the *interval* values to $F_{interval}$. Once the algorithm has analyzed all the points, it then calculates the Shannon entropy over the frequency distribution of the intervals.

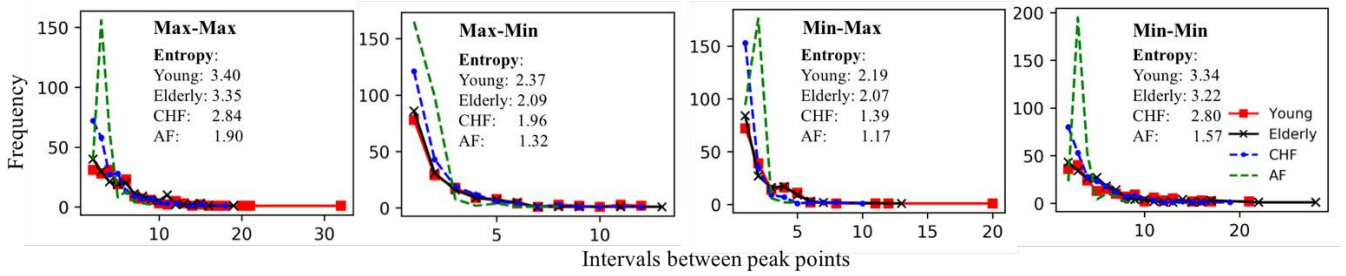


Fig. 7 Frequency distributions of the intervals between points of four subjects from Fig. 1. The numbers are the attention entropy values. The average entropies (Average-4) are: Young=2.82, Elderly=2.68, CHF=2.25, AF=1.49.

Similar to Algorithm 1, Average-4 is implemented by checking the peak point type (local maxima or local minima) and updating the respective frequency histogram. Thus, the original $O(n)$ time complexity still remains. Both Algorithm 1 and Average-4 require extra space for storing the frequency histograms, which is upper limited by $O(n)$.

4 EXPERIMENTAL SETUP

Datasets: we first tested with simulated Gaussian distributed white and $1/f$ noises [42]-[45], and then tested with real-world data of healthy and pathological subjects: the interbeat intervals dataset which is downloaded from PhysioNet [46]. There are 72 healthy subjects divided into two groups: subjects with age ≤ 55 (*young*) and subjects with age >55 (*elderly*). There are also 44 subjects with *congestive heart failure* (CHF), and 24 subjects with *atrial fibrillation* (AF). The information about the dataset is shown in Table 2, and the selected sub-series of different subjects are shown in Fig. 1. The length is the number of samples in the time-series with a sampling frequency of 128 Hz for young, elderly, and part of CHF and 250 Hz for AF and part of CHF [46].

TABLE 2. Dataset Information

Group	Instance	Min. length	Avg. length	Max. length
Young	26	75,100	101,277	126,945
Elderly	46	76,926	106,234	136,527
CHF	44	74,985	111,144	147,879
AF	24	34,837	48,701	61,915

Define key patterns: we used the peak points introduced in Section 3 as key patterns. The same attention entropy (Average-4) calculation illustrated in Fig. 5 was applied to the experiments.

Baseline methods: we compared the proposed method to all the entropy methods in Table 1. We used the parameters suggested from the original paper of each method.

Measurements: we used the analysis of variance (ANOVA) [47] and the area under the receiver operating characteristic curve (ROC AUC) [48] as the measurements. ANOVA can determine if the means of groups of data are significantly different from each other. ANOVA outcomes a p -value, and if the p -value is below the threshold chosen for statistical significance (usually 0.1, 0.05, or 0.01), there are significant differences among the groups. The idea of

receiver operating characteristic (ROC) curve is to plot the true-positive rate against the false-positive rate over the ranked entropy values at various threshold values. The *area under the ROC curve* (ROC AUC) serves as the accuracy evaluation ranging from 0 to 1. The value 1 corresponds to a perfect classification result.

5 RESULTS

5.1 Simulated white and $1/f$ noises

We applied the attention entropy method to the simulated Gaussian distributed white and $1/f$ noises, and the results are shown in Fig. 8. We can see that the attention entropy values of $1/f$ noise are significantly higher (p -value < 0.01) than white noise. This result is consistent with the fact that, unlike white noise, $1/f$ noise contains complex structures [42], [43].

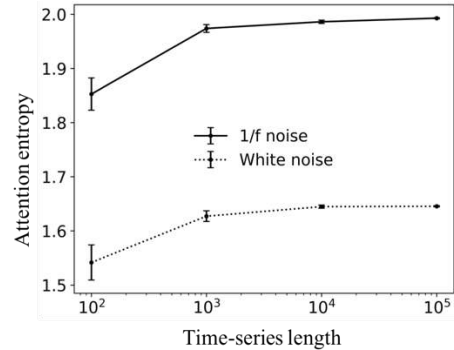


Fig. 8 Attention entropy analysis of 30 simulated Gaussian distributed (mean zero, variance one) white and $1/f$ noise time-series. Symbols represent the mean values of entropy, and bars represent the standard error (SE = standard deviation / \sqrt{n} , where n is the number of subjects).

5.2 Real-world heart-rate data

We next tested the interbeat interval time-series dataset with time-series length=100. The p -value results are shown in Table 3. We used the star symbol (*) to mark the results that are statistically significant (p -values < 0.01). We can see that the results of attention entropy are statistically significant in all the important cases of separating healthy and non-healthy subjects. The differences of the entropy values in case of young-vs-elderly and CHF-vs-AF are as we expected but not statistically significant. The possible reason

TABLE 3 P-value Results

Time-series length	Entropy methods	Shannon [26]	Rényi [29]	Tsallis [30]	Per. [31]	App. [33]	Sample [34]	Bub. [35]	HVE [36]	GHVE [22]	SVDE [38]	EPE [32]	Spe. [27]	Ave. [28]	MSE [21]	Attention
100	Young, Elderly	0.115	0.159	0.213	0.484	0.376	0.570	0.516	0.922	0.060	0.335	0.482	0.212	0.040	0.463	0.021
	Young, CHF	0.001*	0.001*	0.002*	0.352	0.318	0.029	0.139	0.006*	<0.001*	0.037	0.343	0.003*	0.002*	0.017	<0.001*
	Young, AF	0.019	0.025	0.109	<0.001*	0.856	0.025	0.754	0.020	0.015	<0.001*	<0.001*	0.011	0.064	<0.001*	<0.001*
	Elderly, CHF	0.012	0.006*	0.002*	0.675	0.705	0.060	0.307	0.030	0.018	0.002*	0.663	<0.001*	0.143	0.025	0.007*
	Elderly, AF	<0.001*	<0.001*	0.003*	0.002*	0.572	0.059	0.334	0.094	0.347	<0.001*	0.002*	<0.001*	<0.001*	<0.001*	0.001*
	CHF, AF	<0.001*	<0.001*	<0.001*	0.021	0.463	0.785	0.086	0.970	0.250	0.026	0.022	0.967	<0.001*	0.116	0.116
	AVG.	0.025	0.032	0.055	0.256	0.548	0.255	0.356	0.34	0.115	0.067	0.252	0.199	0.042	0.104	0.024
	Stars	3	4	4	2	0	0	0	1	1	3	2	3	3	2	4
1000	Young, Elderly	0.226	0.151	0.212	0.087	0.477	0.546	0.463	0.716	0.018	0.422	0.086	0.575	0.151	0.211	<0.001*
	Young, CHF	<0.001*	<0.001*	<0.001*	0.175	0.798	0.902	0.096	0.014	<0.001*	0.077	0.165	<0.001*	0.012	<0.001*	<0.001*
	Young, AF	0.104	0.233	0.771	<0.001*	0.361	0.149	0.909	0.315	<0.001*	<0.001*	<0.001*	<0.001*	0.039	<0.001*	<0.001*
	Elderly, CHF	<0.001*	<0.001*	<0.001*	0.827	0.714	0.679	0.235	0.004*	<0.001*	0.006*	0.797	<0.001*	0.098	<0.001*	<0.001*
	Elderly, AF	0.005*	0.014	0.174	0.005*	0.101	0.022	0.393	0.304	<0.001*	<0.001*	0.004*	<0.001*	0.001*	<0.001*	<0.001*
	CHF, AF	<0.001*	<0.001*	0.001*	0.042	0.239	0.093	0.080	0.344	0.591	0.020	0.043	0.014	<0.001*	0.148	0.004*
	AVG.	0.056	0.066	0.193	0.19	0.448	0.398	0.362	0.283	0.101	0.087	0.183	0.098	0.05	0.06	0.001
	Stars	4	3	3	2	0	0	0	1	4	3	2	4	2	4	6
10000	Young, Elderly	0.145	0.105	0.187	0.009*	<0.001*	<0.001*	0.953	0.857	<0.001*	0.345	0.008*	0.830	<0.001*	<0.001*	<0.001*
	Young, CHF	<0.001*	<0.001*	<0.001*	0.128	0.018	0.013	0.130	0.098	<0.001*	0.042	0.120	<0.001*	0.004*	<0.001*	<0.001*
	Young, AF	<0.001*	0.008*	0.052	<0.001*	0.812	0.383	0.545	0.071	<0.001*	<0.001*	<0.001*	<0.001*	0.013	<0.001*	<0.001*
	Elderly, CHF	<0.001*	<0.001*	<0.001*	0.564	0.075	0.223	0.090	0.065	0.001*	0.001*	0.583	<0.001*	0.883	0.249	0.038
	Elderly, AF	<0.001*	<0.001*	0.001*	<0.001*	0.004*	0.082	0.572	0.040	0.377	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*
	CHF, AF	<0.001*	<0.001*	<0.001*	0.001*	0.124	0.401	0.046	0.572	0.127	<0.001*	0.001*	0.019	<0.001*	0.014	0.004*
	AVG.	0.024	0.019	0.040	0.117	0.172	0.184	0.389	0.284	0.084	0.065	0.119	0.141	0.15	0.044	0.007
	Stars	5	5	4	4	2	1	0	0	4	4	4	4	4	4	5

The "*" symbol means the p -values <0.01.

TABLE 4 ROC AUC Results

Time-series length	Entropy methods	Shannon [26]	Rényi [29]	Tsallis [30]	Per. [31]	App. [33]	Sample [34]	Bub. [35]	HVE [36]	GHVE [22]	SVDE [38]	EPE [32]	Spe. [27]	Ave. [28]	MSE [21]	Attention
100	Young, Elderly	0.63	0.63	0.63	0.43	0.58	0.56	0.56	0.45	0.62	0.73	0.43	0.58	0.65	0.52	0.68
	Young, CHF	0.75	0.76	0.76	0.36	0.56	0.67	0.61	0.32	0.74	0.40	0.36	0.29	0.73	0.62	0.78
	Young, AF	0.36	0.36	0.36	0.17	0.55	0.71	0.54	0.34	0.68	0.20	0.17	0.26	0.39	0.71	0.85
	Elderly, CHF	0.65	0.65	0.66	0.43	0.53	0.63	0.55	0.38	0.65	0.30	0.43	0.24	0.60	0.59	0.67
	Elderly, AF	0.27	0.27	0.27	0.22	0.47	0.66	0.51	0.41	0.56	0.12	0.22	0.25	0.29	0.68	0.72
	CHF, AF	0.18	0.18	0.19	0.29	0.45	0.50	0.48	0.53	0.41	0.34	0.29	0.50	0.23	0.60	0.56
	AVG.	0.47	0.47	0.48	0.32	0.52	0.62	0.54	0.41	0.61	0.35	0.31	0.35	0.48	0.62	0.71
	SD	0.23	1.16	0.16	0.11	0.05	0.08	0.04	0.08	0.11	0.21	0.11	0.15	0.21	0.07	0.10
1000	Young, Elderly	0.59	0.61	0.60	0.35	0.55	0.52	0.55	0.53	0.68	0.61	0.35	0.54	0.63	0.60	0.77
	Young, CHF	0.75	0.77	0.77	0.33	0.51	0.52	0.61	0.30	0.88	0.46	0.33	0.22	0.73	0.80	0.90
	Young, AF	0.38	0.42	0.42	0.15	0.43	0.40	0.46	0.27	0.85	0.19	0.15	0.06	0.41	0.87	0.96
	Elderly, CHF	0.70	0.72	0.72	0.43	0.48	0.49	0.57	0.27	0.80	0.39	0.43	0.21	0.66	0.77	0.71
	Elderly, AF	0.33	0.35	0.35	0.22	0.41	0.38	0.40	0.25	0.81	0.16	0.22	0.08	0.33	0.88	0.82
	CHF, AF	0.21	0.21	0.21	0.30	0.42	0.41	0.34	0.44	0.57	0.30	0.30	0.33	0.24	0.62	0.67
	AVG.	0.49	0.56	0.56	0.30	0.47	0.45	0.49	0.34	0.77	0.35	0.30	0.24	0.50	0.76	0.81
	SD	0.22	0.18	0.18	0.10	0.06	0.06	0.11	0.11	0.12	0.17	0.10	0.18	0.20	0.12	0.11
10000	Young, Elderly	0.59	0.61	0.60	0.30	0.79	0.77	0.49	0.52	0.85	0.60	0.30	0.53	0.79	0.82	0.80
	Young, CHF	0.75	0.80	0.79	0.32	0.69	0.69	0.55	0.35	0.92	0.40	0.32	0.16	0.76	0.83	0.87
	Young, AF	0.22	0.30	0.28	0.10	0.62	0.67	0.48	0.27	0.79	0.05	0.10	0.03	0.32	0.95	0.94
	Elderly, CHF	0.71	0.74	0.74	0.49	0.42	0.45	0.56	0.35	0.70	0.37	0.48	0.17	0.54	0.60	0.61
	Elderly, AF	0.18	0.26	0.24	0.19	0.35	0.45	0.49	0.27	0.58	0.04	0.19	0.05	0.18	0.77	0.80
	CHF, AF	0.12	0.15	0.14	0.20	0.42	0.49	0.43	0.39	0.40	0.21	0.20	0.31	0.19	0.67	0.71
	AVG.	0.43	0.54	0.54	0.27	0.55	0.59	0.50	0.36	0.71	0.28	0.26	0.21	0.46	0.77	0.79
	SD	0.29	0.16	0.16	0.14	0.18	0.14	0.05	0.09	0.19	0.22	0.13	0.19	0.27	0.12	0.12

is that the number of samples in the data is too small for this. Attention entropy is the only method capable to separate all six groups when time-series length = 1000 so that the result is statistically significant (p -values < 0.01).

Another measurement of the results of classifying binary groups is shown in the AUC in Table 4. We can see that attention entropy outperforms other entropy methods on average (AUC: 0.71 vs. 0.62 with time-series length = 100, 0.81 vs. 0.77 with time-series length = 1000, 0.79 vs. 0.77 with time-series length = 10000). This indicates that the attention entropy is more powerful to separate the groups than the other methods are. It gives evidence that analyzing the frequencies of the intervals between patterns is more beneficial than analyzing the frequencies of patterns, especially when the time-series length is short, for example, 100.

5.3 Effect of the time-series length

We studied the effect of the time-series length and the results are summarized in Table 3, Table 4 and Fig. 9. From Table 3 and Table 4, we can see that, except for the case of CHF-and-AF with the times-series length of 100 (p -value = 0.116) and the case of elderly-and-CHF with the time-series length of 10000 (p -value = 0.038), the attention entropy values between each group are always statistically and significantly different (p -values < 0.01). This indicates that the attention entropy can differentiate the groups at the same time very well and is robust to the time-series length. However, all the other methods are sensitive to the length of the time-series.

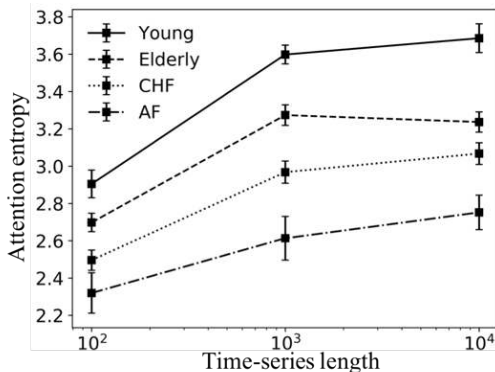


Fig. 9 Attention entropy analysis of interbeat intervals time-series derived from healthy subjects with age ≤ 55 (young), healthy subjects with age > 55 (elderly), subjects with congestive heart failure (CHF), and subjects with atrial fibrillation (AF). Symbols represent the mean values of entropies, and bars represent the standard error (SE = standard deviation / \sqrt{n} , where n is the number of subjects).

From Fig. 9, we found that, regardless of the time-series length, the attention entropy values decrease by following the order: entropy (young) $>$ entropy (elderly) $>$ entropy (CHF) $>$ entropy (AF). These results are consistent with the concept that the cardiac dynamics of healthy young subjects are the most complex [43] and provide stronger support for the hypothesized *complexity-loss of aging and disease theory* [49] than multiscale entropy. The attention entropy method reflects the regularity of repeating patterns of signals and plays more critical roles behind the complexity-

loss of aging and disease. The regularity-loss ignored by conventional entropy methods is explicitly addressed by the attention entropy.

5.4 Intervals between peak points

To study the intervals among peak points further, we tested the intervals between local maxima and local maxima (Max-Max intervals), the intervals between local minima and local minima (Min-Min intervals), the intervals between local maxima and local minima (Max-Min intervals and Min-Max intervals). We calculated Shannon entropy of these four intervals and the average of Shannon entropy of these four intervals (Average-4). The AUC results are summarized in Table 5. We can see that the choice of the interval does not matter regardless of which time-series length is used. To simplify the choice, we recommend using Average-4 by default.

TABLE 5 Average AUC of Binary Groups

Time-series length	100	1000	10000
Max-Max	0.68	0.81	0.80
Min-Min	0.71	0.80	0.78
Max-Min	0.64	0.75	0.75
Min-Max	0.68	0.75	0.76
Average-4	0.72	0.81	0.79

5.5 Compared with basic statistics

Basic statistics such as mean, standard deviation, root mean square, and the number of pairs of successive inter-beat intervals that differ by more than 50 ms (NN50 defined in [50]) are also used to analyze the interbeat time-series [50]. We make comparison with attention entropy and the results are summarized in Table 6. We can see attention entropy outperforms all basic statistics regardless of the time-series length.

TABLE 6 Average AUC of Binary Groups

Time-series length	100	1000	10000
Attention entropy	0.72	0.79	0.78
Mean	0.57	0.56	0.54
Root mean square	0.58	0.56	0.54
Standard deviation	0.61	0.48	0.55
NN50	0.57	0.63	0.58

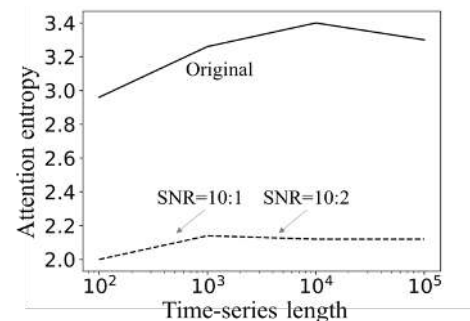


Fig. 10 Effects of different amounts of Gaussian distributed white noise on attention entropy curves. SNR corresponds to a single-noise-ratio. The attention entropy curve labeled *original* corresponds to the attention entropy results for the interbeat intervals from a healthy subject.

5.6 Effect of noise and outliers

The result of an experiment may be affected by the type of noise. Here, we discuss the effects of superimposing uncorrelated (Gaussian distributed white) noise on a physiologic time-series. Fig. 10 shows that the attention entropy method is sensitive to the noise. The same observation in Fig. 11 holds for the effects of outliers. This is because noise and outliers affect the key patterns, namely the peak points.

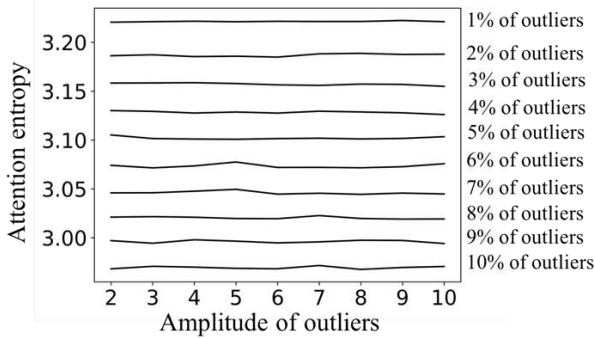


Fig. 11 Contour plot showing how the percentage of outliers and their amplitude (relative to the mean value of the time-series) affect the attention entropy.

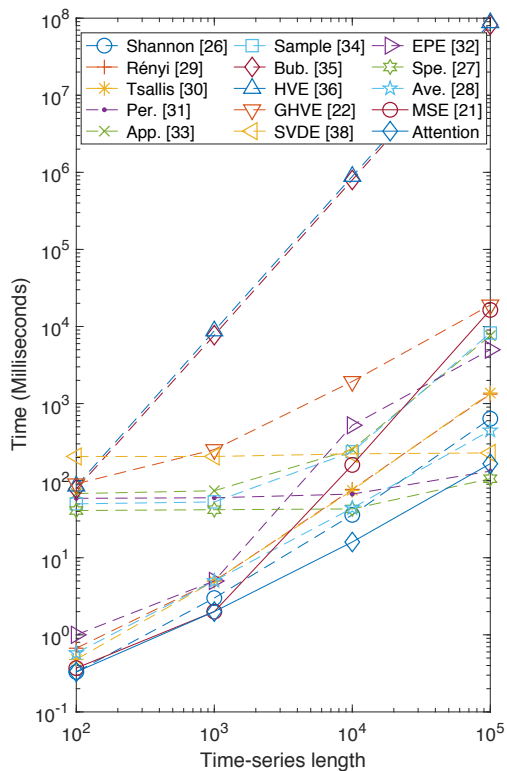


Fig. 12 Log-log plot of running time over time-series in Table 7.

5.7 Computational complexity

Attention entropy takes $O(n)$ time, where n is the time-series length. To measure the actual processing time of the algorithm, the algorithm was implemented in Python 3.7,

http://cs.uef.fi/sipu/soft/attention_entropy.py

which can be found from the web and tested using PC with CPU Intel Core i7, 16GB RAM, and clock frequency 2.3 GHz. Fig. 12 and Table 7 show the relationship between the running time and the time-series length of one young subject. We can see that with the increase of the time-series length, attention entropy requires much less computing time than most of the competing entropy methods, including the competitive MSE [21].

5.8 Discussion

In this section, we discuss the potential usability of the method in affective computing, its limitations, and the threats to the validity of the results of the proposed method. Many methods based on HRV have been developed for affective state analysis. This is because plenty of affective computing researches consider that specific emotional states can elicit changes in the autonomic nervous system, which can be exactly monitored by HRV analyses as shown by studies over the decades. However, quantifying HRV with entropy-based methods has been rarely used for affective analysis although it has been widely adopted in many tasks such as disease detection and classification. This may be because conventional entropy methods were proposed for long-term HRV analysis as introduced in Section 2, therefore, they were not applicable to short duration HRV analysis. This obstacle is expected to be removed by attention entropy, which can work well with short duration HRV signals and, therefore, can be potentially applied to affective state analysis. Moreover, attention entropy may be able to capture the change of affective states in a timely manner considering its advantage of requiring linear time complexity.

One limitation of the proposed method is that it needs to define key patterns in advance. The limitation of using peak points as key patterns is that it is sensitive to outliers and noise. The key patterns may be application-specific, which may be a threat to the validity of the results. However, these threats may be overcome by defining different key patterns and combining the results from multiple key patterns; future work could explore this strategy. The mechanisms behind the key patterns such as peak points could also be explored in future work.

6. CONCLUSION

A novel complexity analysis method called attention entropy is proposed, which does not need any parameter tuning when using peak points as key patterns. It has linear time complexity and is robust to the time-series length. We compared it to fourteen state-of-the-art complexity analysis methods with real-world datasets. The results show that attention entropy outperforms all the compared methods and is the only method to be able to separate all groups with statistical significance using time-series length of 1000. This shows attention entropy has higher discrimination power in short duration HRV signals and has potential in other tasks such as affective computing. Future work could uncover more key patterns and the hidden mechanisms behind them.

TABLE 7 Running Time over Time-Series Length (Milliseconds).

Time-series length	Shannon [26]	Rényi [29]	Tsallis [30]	Per. [31]	App. [33]	Sample [34]	Bub. [35]	HVE [36]	GHVE [22]	SVDE [38]	EPE [32]	Spe. [27]	Ave. [28]	MSE [21]	Attention
100	<1	<1	<1	59	68	50	78	85	91	206	1	41	<1	<1	<1
1000	3	5	5	60	74	53	7714	8763	253	206	5	42	5	2	2
10000	36	77	75	67	254	237	795821	885471	1935	223	523	43	44	160	16
100000	636	1328	1374	133	7599	8236	>1h	>1h	19019	229	5001	106	448	16404	166

REFERENCES

[1] J. Enderle and J. Bronzino, *Introduction to biomedical engineering*. Academic press, 2012.

[2] W. W. Butkevičiūtė E, Bikulčienė L, Sidekerskienė T, Blažauskas T, Maskeliūnas R, Damaševičius R, "Removal of movement artefact for mobile EEG analysis in sports exercises", *IEEE Access*, vol. 7, pp. 7206–7217, 2019.

[3] M. Nardelli, A. Lanata, G. Valenza, M. Felici, P. Baragli, and E. P. Scilingo, "A tool for the real-time evaluation of ECG signal quality and activity: Application to submaximal treadmill test in horses", *Biomed. Signal Process. Control*, vol. 56, p. 101666, 2020.

[4] R. D. Labati, E. Muñoz, V. Piuri, R. Sassi, and F. Scotti, "Deep-ECG: Convolutional neural networks for ECG biometric recognition", *Pattern Recognit. Lett.*, vol. 126, pp. 78–85, 2019.

[5] F. Ishida and K. Wakata, "Numerical study on adjusting parameters to improve gaze estimation using planar approximations from electro-oculogram signal voltage ratios", *Biomed. Eng. Lett.*, vol. 9, no. 2, pp. 181–187, 2019.

[6] M. D. Josephson and C. A. Knight, "Comparison of neural excitation measures from the surface electromyogram during rate-dependent muscle contractions", *J. Electromyogr. Kinesiol.*, vol. 44, pp. 15–20, 2019.

[7] B. Afsharipour, G. Li, S. Chandra, W. Z. Rymer, and N. L. Suresh, "Variations of Tendon Tap Force Threshold needed to Evoke Surface Electromyogram Responses after Botulinum Toxin Injection in Chronic Stroke Survivors", in *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 385–388, 2019.

[8] R. Martinez, A. Salazar-Ramirez, A. Arruti, E. Irigoyen, J. I. Martin, and J. Mugerza, "A Self-Paced Relaxation Response Detection System Based on Galvanic Skin Response Analysis", *IEEE Access*, vol. 7, pp. 43730–43741, 2019.

[9] T. Hong, X. Sun, F. Tian, and F. Ren, "Sentiment Classification And Personality Detection Via Galvanic Skin Response Based on Deep Learning Models", in *2019 5th International Conference on Big Data Computing and Communications (BIGCOM)*, pp. 313–317, 2019.

[10] M. F. Shlesinger, "Fractal time and 1/f noise in complex systems", *Ann. N. Y. Acad. Sci.*, vol. 504, no. 1, pp. 214–228, 1987.

[11] M. Malik and A. J. Camm, *Heart rate variability*, Futura, Armonk, NY, 1995.

[12] P. Indic, E. B. Salisbury, D. Paydarfar, E. N. Brown, and R. Barbieri, "Interaction between heart rate variability and respiration in preterm infants", in *2008 Computers in Cardiology*, pp. 57–60, 2008.

[13] W. S. Liew, M. Seera, C. K. Loo, E. Lim and N. Kubota, "Classifying Stress From Heart Rate Variability Using Salivary Biomarkers as Reference", in *IEEE Transactions on Neural Networks and Learning Systems*, vol. 27, no. 10, pp. 2035–2046, 2016.

[14] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis and M. Tsiknakis, "Review on psychological stress detection using biosignals", *IEEE Transactions on Affective Computing*, 2019.

[15] M. Nardelli, A. Greco, M. Bianchi, E. P. Scilingo and G. Valenza, "Classifying Affective Haptic Stimuli through Gender-specific Heart Rate Variability Nonlinear Analysis", *IEEE Transactions on Affective Computing*, 2018.

[16] L. A. Bugnon, R. A. Calvo and D. H. Milone, "Dimensional Affect Recognition from HRV: An Approach Based on Supervised SOM and ELM", *IEEE Transactions on Affective Computing*, vol. 11, no. 1, pp. 32–44, 2020.

[17] Y. Isler, A. Narin, M. Ozer, and M. Perc, "Multi-stage classification of congestive heart failure based on short-term heart rate variability", *Chaos, Solitons & Fractals*, vol. 118, pp. 145–151, 2019.

[18] H.H. Chao, C.-W. Yeh, C. F. Hsu, L. Hsu, and S. Chi, "Multiscale Entropy Analysis with Low-Dimensional Exhaustive Search for Detecting Heart Failure", *Appl. Sci.*, vol. 9, no. 17, p. 3496, 2019.

[19] W. Pan, A. He, K. Feng, Y. Li, D. Wu, and G. Liu, "Multi-frequency components entropy as novel heart rate variability indices in congestive heart failure assessment", *IEEE Access*, vol. 7, pp. 37708–37717, 2019.

[20] M. Shi, C. Zhan, et al., "Renyi distribution entropy analysis of short-term heart rate variability signals and its application in coronary artery disease detection", *Front. Physiol.*, vol. 10, 2019.

[21] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of complex physiologic time series", *Phys. Rev. Lett.*, vol. 89, no. 6, p. 68102, 2002.

[22] G. I. Choudhary, W. Aziz, I. R. Khan, S. Rahardja, and P. Fränti, "Analysing the Dynamics of Interbeat Interval Time Series Using Grouped Horizontal Visibility Graph", *IEEE Access*, vol. 7, pp. 9926–9934, 2019.

[23] R. Clausius, *The mechanical theory of heat: with its applications to the steam-engine and to the physical properties of bodies*. J. van Voorst, 1867.

[24] L. Boltzmann, "Weitere studien über das wärme-gleichgewicht unter gasmolekülen", in *Kinetische Theorie II*, Springer, pp. 115–225, 1970.

[25] J. W. Gibbs, *Elementary principles in statistical mechanics developed with especial reference to the rational foundation of thermodynamics* by J. Willard Gibbs. New York:: C. Scribner, 1902.

[26] C. E. Shannon, "A mathematical theory of communication", *The Bell System Technical Journal*, vol. 27, no. 3, pp. 379 - 423, 1948.

[27] T. Inouye, K. Shinosaki, et al., "Quantification of EEG irregularity by use of the entropy of the power spectrum", *Electroencephalogr. Clin. Neurophysiol.*, vol. 79, no. 3, pp. 204–210, 1991.

[28] C. F. Hsu, P.-Y. Lin, H.-H. Chao, L. Hsu, and S. Chi, "Average Entropy: Measurement of disorder for cardiac RR interval signals", *Phys. A Stat. Mech. its Appl.*, vol. 529, p. 121533, 2019.

[29] A. Rényi, "On measures of entropy and information", in *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics*, 1961.

[30] C. Tsallis, "Possible generalization of boltzmann-gibbs statistics", *Journal of Statistical Physics*, vol. 52, no. 1-2, pp. 479–487, 1988.

[31] C. Bandt and B. Pompe, "Permutation entropy: a natural complexity measure for time series", *Phys. Rev. Lett.*, vol. 88, no. 17, p. 174102, 2002.

[32] Z. Huo, Y. Zhang, L. Shu, and X. Liao, "Edge Permutation Entropy: An Improved Entropy Measure for Time-Series Analysis", in *IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society*, vol. 1, pp. 5998–6003, 2019.

[33] S. M. Pincus, "Approximate entropy as a measure of system complexity", *Proc. Natl. Acad. Sci.*, vol. 88, no. 6, pp. 2297–2301, 1991.

[34] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy", *Am. J. Physiol. Circ. Physiol.*, vol. 278, no. 6, pp. H2039–H2049, 2000.

[35] G. Manis, M. D. Aktaruzzaman, and R. Sassi, "Bubble entropy: an entropy almost free of parameters", *IEEE Trans. Biomed. Eng.*, vol. 64, no. 11, pp. 2711–2718, 2017.

[36] J. Iacovacci and L. Lacasa, "Sequential motif profile of natural visibility graphs", *Phys. Rev. E*, vol. 94, no. 5, p. 52309, 2016.

[37] L. Lacasa, B. Luque, F. Ballesteros, J. Luque, and J. C. Nuno, "From time series to complex networks: The visibility graph", *Proc. Natl. Acad. Sci.*, vol. 105, no. 13, pp. 4972–4975, 2008.

[38] O. Alter, P. O. Brown, and D. Botstein, "Singular value

decomposition for genome-wide expression data processing and modeling", *Proc. Natl. Acad. Sci.*, vol. 97, no. 18, pp. 10101–10106, 2000.

- [39] M. Borowska, "Entropy-based algorithms in the analysis of biomedical signals", *Stud. Logic Grammar Rhetoric* 2015, 43, pp. 21–32, 2015.
- [40] L.C. Amarantidis, D. Abásolo, "Interpretation of entropy algorithms in the context of biomedical signal analysis and their application to EEG analysis in epilepsy", *Entropy*, 21, 840, 2019.
- [41] S. M. Pincus, "Approximate entropy as a measure of system complexity", *Proc. Natl. Acad. Sci.*, vol. 88, no. 6, pp. 2297–2301, 1991.
- [42] E. Sejdić and L. A. Lipsitz, "Necessity of noise in physiology and medicine", *Comput. Methods Programs Biomed.*, vol. 111, no. 2, pp. 459–470, 2013.
- [43] M. Costa, A. L. Goldberger, and C.-K. Peng, "Multiscale entropy analysis of biological signals", *Phys. Rev. E*, vol. 71, no. 2, p. 21906, 2005.
- [44] Y. C. Zhang, "Complexity and 1/f noise. A phase space approach", *J. Phys. I*, vol. 1, no. 7, pp. 971–977, 1991.
- [45] H. C. Fogedby, "On the phase space approach to complexity", *J. Stat. Phys.*, vol. 69, no. 1–2, pp. 411–425, 1992.
- [46] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, et al., "Physiobank physiotoolkit and physionet: components of a new research resource for complex physiologic signals", *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.
- [47] R. Lowry. *Concepts and applications of inferential statistics*. Poughkeepsie, NY: Vassar College, 1999.
- [48] D.K. McClish. *Analyzing a portion of the ROC curve*. Med Decis Making. 9:190-5, 1989.
- [49] L. A. Lipsitz and A. L. Goldberger, "Loss of 'complexity' and aging: potential applications of fractals and chaos theory to senescence", *Jama*, vol. 267, no. 13, pp. 1806–1809, 1992.
- [50] L. Wang, W. Zhou, Q. Chang, J. Chen and X. Zhou, "Deep Ensemble Detection of Congestive Heart Failure Using Short-Term RR Intervals", in *IEEE Access*, vol. 7, pp. 69559–69574, 2019.



Jiawei Yang received the B.Eng. degree in Electronic Engineering from Beihang University, China in 2013, the M.Eng. and Ph.D. degrees in Computer Science from University of Eastern Finland, Finland, in 2019 and 2020, respectively.



Gulraiz Iqbal Choudhary received his master's degree in Software Technology from the Linnaeus University, Växjö, Sweden. Currently, he is pursuing Ph.D. in Computer Science in University of Eastern Finland, Joensuu, Finland. His focused area of research is heart rate variability analysis of normal and pathological subjects using nonlinear complexity measures. His other research interests include complex networks, disease progression network, data analytics. He has collaborated actively with researchers in several other disciplines of computing and information technology.



Susanto Rahardja (F'11) received the B.Eng. degree from National University of Singapore in 1991, the M.Eng. and Ph.D. degrees all in Electronic Engineering from Nanyang Technological University, Singapore, in 1993 and 1997, respectively. He is currently a Chair Professor at the Northwestern Polytechnical University (NPU) under the Thousand Talent Plan of People's Republic of China. His research interests are in multimedia, signal processing, wireless communications, discrete transforms and signal processing algorithms, implementation and optimization. Dr. Rahardja was the recipients of numerous awards, including the IEE Hartree Premium Award, the Tan Kah Kee Young Inventors' Open Category Gold award, the Singapore National Technology Award, A*STAR Most Inspiring Mentor Award, Finalist of the 2010 World Technology & Summit Award, the Nokia Foundation Visiting Professor Award and the ACM Recognition of Service Award.



PASI FRÄNTTI (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in computer science from the University of Turku, in 1991 and 1994, respectively. Since 2000, he has been a Professor of computer science with the University of Eastern Finland. He is currently a Visiting Professor with Shenzhen Technology University, China. He has published 89 journals and 173 peer review conference papers, including 15 IEEE transaction articles. His main research interests include machine learning, data mining, and pattern recognition, including clustering algorithms and intelligent location-aware systems.