

Detection of healthy and pathological heartbeat dynamics in ECG signals using multivariate recurrence networks with multiple scale factors

Lu Ma(马璐), Meihui Chen(陈梅辉), Aijun He(何爱军),Deqiang Cheng(程德强), and Xiaodong Yang(杨小冬)
Citation: Chin. Phys. B, 2023, 32 (**10**): 100501. DOI: 10.1088/1674-1056/acb422
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Semi-analytical steady-state response prediction for multi-dimensional quasi-Hamiltonian systems

Wen-Wei Ye(叶文伟), Lin-Cong Chen(陈林聪), Zi Yuan(原子), Jia-Min Qian(钱佳敏), and Jian-Qiao Sun(孙建桥)
Chin. Phys. B, 2023, 32 (**6**): 060506. DOI: 10.1088/1674-1056/acae7c

Detecting physical laws from data of stochastic dynamical systems perturbed by non-Gaussian α-stable Lévy noise

Linghongzhi Lu(陆凌弘志), Yang Li(李扬), and Xianbin Liu(刘先斌)
Chin. Phys. B, 2023, 32 (**5**): 050501. DOI: 10.1088/1674-1056/aca7ee

Data-driven modeling of a four-dimensional stochastic projectile system

Yong Huang(黄勇) and Yang Li(李扬)
Chin. Phys. B, 2022, 31 (**7**): 070501. DOI: 10.1088/1674-1056/ac4e0d

Multifractal analysis of the software evolution in software networks

Meili Liu(刘美丽), Xiaogang Qi(齐小刚), and Hao Pan(潘浩)
Chin. Phys. B, 2022, 31 (**3**): 030501. DOI: 10.1088/1674-1056/ac1b8a

Identification of unstable individuals in dynamic networks

Dongli Duan(段东立), Tao Chai(柴涛), Xixi Wu(武茜茜), Chengxing Wu(吴成星), Shubin Si(司书宾), and Genqing Bian(边根庆)
Chin. Phys. B, 2021, 30 (**9**): 090501. DOI: 10.1088/1674-1056/abe92f

---

Detection of healthy and pathological heartbeat dynamics in ECG signals using multivariate recurrence networks with multiple scale factors

Lu Ma(马璐)[1,2], Meihui Chen(陈梅辉)[3], Aijun He(何爱军)[4], Deqiang Cheng(程德强)[1], and Xiaodong Yang(杨小冬)[3,†]

[1] *School of Information and Control Engineering, China University of Mining and Technology, Xuzhou* 221116, *China*
[2] *Suzhou Vocational and Technical College, Suzhou* 234000, *China*
[3] *School of Computer Science and Technology, China University of Mining and Technology, Xuzhou* 221116, *China*
[4] *School of Electronic Science and Engineering, Nanjing University, Nanjing* 210023, *China*

(Received 14 September 2022; revised manuscript received 4 January 2023; accepted manuscript online 18 January 2023)

The electrocardiogram (ECG) is one of the physiological signals applied in medical clinics to determine health status. The physiological complexity of the cardiac system is related to age, disease, etc. For the investigation of the effects of age and cardiovascular disease on the cardiac system, we then construct multivariate recurrence networks with multiple scale factors from multivariate time series. We propose a new concept of cross-clustering coefficient entropy to construct a weighted network, and calculate the average weighted path length and the graph energy of the weighted network to quantitatively probe the topological properties. The obtained results suggest that these two network measures show distinct changes between different subjects. This is because, with aging or cardiovascular disease, a reduction in the conductivity or structural changes in the myocardium of the heart contributes to a reduction in the complexity of the cardiac system. Consequently, the complexity of the cardiac system is reduced. After that, the support vector machine (SVM) classifier is adopted to evaluate the performance of the proposed approach. Accuracy of 94.1% and 95.58% between healthy and myocardial infarction is achieved on two datasets. Therefore, this method can be adopted for the development of a noninvasive and low-cost clinical prognostic system to identify heart-related diseases and detect hidden state changes in the cardiac system.

Keywords: electrocardiogram signals, multivariate recurrence networks, cross-clustering coefficient entropy, multiscale analysis

PACS: 05.10.–a, 87.18.–h, 87.85.Ng, 87.19.X– **DOI:** 10.1088/1674-1056/acb422

1. Introduction

The human cardiac system, as a typical complex system, has non-stationary and nonlinear characteristics. The electrocardiogram (ECG) is a non-invasive and efficient approach that has been adopted clinically to diagnose heart-related diseases. The analysis of traditional nonlinear time series has made great progress since the 1980s,[1] such as Lyapunov exponents,[2–4] entropy,[5–7] fractal dimensions,[8–10] etc. The transformation of a time series into a complex network has drawn much attention to its topological structure and the dynamic behavior of the system.[11] For instance, Zhang and Small[12] mapped the pseudo-periodic time series into a complex network, where each cycle represented a node of the network and the connectivity was determined by related parameters between periods, such as phase space distance, linear correlation coefficient,[13] etc. Xu *et al.*[14] proposed to apply the theory of phase space reconstruction to map the points in the phase space as nodes in the complex network, and the edge depended on the phase space distance of points. Lacasa *et al.*[15] declared that they combined visibility graphs with complex networks. On the basis of the recurrence plot,[16] Donner *et al.*[17] introduced the concept of the recurrence network. During the construction of the transition network taken by Sun *et al.*,[18] a class represented one node, and the weighted edge was determined by the transition probability between any two classes.

The recurrence plot is a versatile and reliable method that has been extensively used to graphically represent the recurrence dynamics in time series data, which characterizes the proximity of two states in the time-delay reconstructed state space of a dynamical system. Because the recurrence of trajectories in phase space is one of the most important fundamental features of dynamical systems, consequently, the recurrence network method has been successfully applied to evaluate the dynamic behavior of real systems, such as climate,[19,20] the cardiac system,[21,22] electroencephalogram signals,[23,24] etc. However, when the number of time series is greater than 2, the relationship between the time series cannot be portrayed. In order to overcome this issue, Gao *et al.*[25] constructed a multivariate recurrence network to evaluate the dynamic characteristics of the two-phase flow. Gao *et al.*[26] also mapped the multivariate time series to the multivariate multiscale recurrence network in order to further study the internal char-

†Corresponding author. E-mail: xyang@cumt.edu.cn
© 2023 Chinese Physical Society and IOP Publishing Ltd

acteristics of the two-phase flow. In recent years, multivariate recurrence networks[27–29] have been adopted to explore the complex dynamic behavior of multivariate time series. Eroglu *et al.*[30] put forward a different method to construct multivariate recurrence networks, which compare the similarities and differences between time series. Xu *et al.*[31] applied multivariate recurrence networks to analyze the information linkages between carbon and energy markets.

Coronary heart disease is a common heart system disease. Due to the spatial inhomogeneity of the heart tissue and the anisotropy of the electrical signal propagation in the heart, a single lead ECG signal cannot adequately characterize the entire cardiac system. Therefore, to fully investigate the effects of age and pathological heartbeat dynamics on the entire cardiac system, we propose a novel method to analyze multivariate time series based on multivariate recurrence networks. In the experiment, the multivariate time series, namely the synchronous 12-lead time series of healthy subjects and individuals diagnosed with coronary artery disease, are coarsely granularized and constructed as the corresponding multivariate recurrence network. Next, we calculate the cross-clustering coefficient entropy of multivariate recurrence networks and obtain the optimal scale by comparing the difference between different groups. The method of one-way analysis of variance (ANOVA) is adopted to explore the distinction between different groups at each lead on the optimal scale. Finally, to further investigate the structural changes and internal characteristics of ECG signals in healthy and coronary artery disease subjects, we construct weighted networks and calculate the parameters. Through setting the parameters as features of the SVM, the classification accuracy of the ECG signals recorded from healthy subjects and myocardial infarction is 94.1% and 95.58% on two datasets, respectively.

2. Theory

2.1. Recurrence networks and cross recurrence networks

Recurrence networks have been proposed by Donner *et al.*[17] to map recurrence plots to complex networks. The basic construction steps of a recurrence plot are as follows: First, take the time series $x(t)$ $(t = 1, 2, \ldots, n)$ and transform it into a phase-space vector $\boldsymbol{x}(t)$ using the embedding theory[32]

$$\boldsymbol{x}(t) = (x(t), x(t+\tau), \ldots, x(t+(m-1)\tau)), \tag{1}$$

where m represents the embedding dimension and τ denotes the delay time. Empirically, m and τ are chosen as 2 and 1 in this paper.

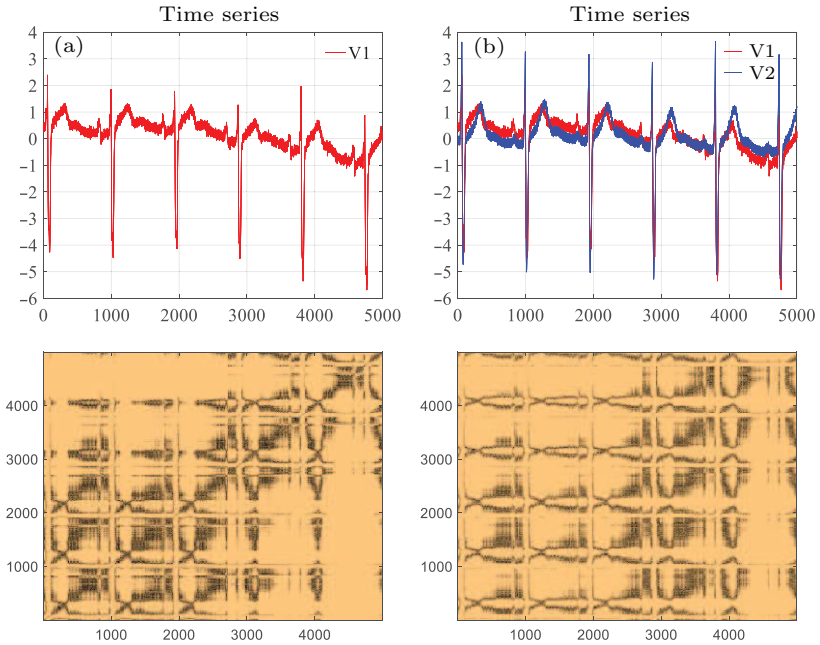


Fig. 1. Construction of a recurrence plot and a cross recurrence plot using a healthy elderly person's ECG time series as an example: (a) the V1 lead time series, and the subgraph below (a) is the recurrence plot constructed by the V1 lead time series; (b) the V1 and the V2 lead time series, and the subgraph below (b) is the cross recurrence plot constructed by the V1 and V2 lead time series, where $m = 2$, $\tau = 1$, and ε is set as 15% of the standard deviation of the time series. The black dots and yellow dots in the recurrence plot and cross recurrence plot represent recurrence and non-recurrence at the corresponding positions, respectively.

The phase-space distance of any two phase-space vectors is calculated in terms of the maximum norm. Compared to the Euclidean norm, the maximum norm is computationally faster and allows the study of some features of recurrence plots analytically.[16]

$$d[\boldsymbol{x}(a), \boldsymbol{x}(b)] = \max_{h=1,\ldots,m}\{|\boldsymbol{x}(a+(h-1)\tau) - \boldsymbol{x}(b+(h-1)\tau)|\}. \tag{2}$$

The recurrence plot is denoted by the recurrence matrix as fol-

lows:

$$R_{\text{auto}}(a,b) = \Theta(\varepsilon - d[\boldsymbol{x}(a) - \boldsymbol{x}(b)]), \tag{3}$$

where $\Theta(\cdot)$ is the Heaviside function, $\Theta(x) = \{1|x > 0;\ 0|x \leq 0\}$, and ε is the threshold. The selection of the threshold is a critical operation that can easily affect the results. The core criterion for choosing a threshold is the power in signal detection that threshold yields.[33] Therefore, several threshold selection techniques have been proposed.[34–38] For the sake of consistency, we always use the standard deviation of the time series in this work.[16,37] It is worth pointing out that we set the threshold at 15% of the standard deviation of the time series in our study.

In contrast to recurrence plots, cross recurrence plots[39] measure the recurrence relationship between two different time series, and the cross recurrence matrix is represented as

$$R_{\text{cross}}(a,b) = \Theta(\varepsilon - d[\boldsymbol{x}(a) - \boldsymbol{y}(b)]), \tag{4}$$

where $\boldsymbol{y}(b)$ represents the phase-space vector constructed by other time series $y(t)$ $(t = 1,2,\ldots,n)$. We use the methods provided in the tool[40] to obtain the recurrence plot and cross recurrence plot, which are shown in Fig. 1. The recurrence plot is constructed from the V1 lead time series of a healthy elderly person and the cross recurrence plot is obtained from the V1 and V2 leads of the same person in Fig. 1.

The adjacency matrix A of a recurrence network is obtained from the recurrence matrix R_{auto}:

$$A_{\text{auto}}(a,b) = R_{\text{auto}}(a,b) - \delta_{a,b}, \tag{5}$$

where $\delta_{a,b}$ is the Kronecker delta function that is introduced to avoid artificial self-loops in the recurrence network ($\delta_{a,b} = 1$ if $a = b$, otherwise $\delta_{a,b} = 0$). The adjacency matrix A of the cross recurrence network is the cross recurrence matrix R_{cross} of the corresponding cross recurrence plot.

2.2. Multivariate recurrence networks

In this study, an M-layer multivariate recurrence network is constructed by recurrence networks and cross recurrence networks. For M-dimensional multivariate time series, we can create M different recurrence networks and M^* $(M-1)$ different cross recurrence networks. We denote the adjacency matrix of the pth layer as A^p_{auto} and $A^p_{\text{auto}}(i,j) = 1$ if node i and node j are connected in the pth layer, $A^p_{\text{auto}}(i,j) = 0$ otherwise. We denote the adjacency matrix $R^{p,q}_{\text{cross}}$ to describe the connection between the pth and qth layers and $R^{p,q}_{\text{cross}}(i,j) = 1$ if node $i$in the pth layer and node j in the qth layer are connected, $R^{p,q}_{\text{cross}}(i,j) = 0$ otherwise. The giant adjacency matrix R_M describing the entire multivariate recurrence network is denoted by

$$R_M = \begin{pmatrix} A^1_{\text{auto}} & R^{1,2}_{\text{cross}} & \cdots & R^{1,M}_{\text{coss}} \\ R^{2,1}_{\text{cross}} & A^2_{\text{auto}} & \cdots & R^{2,M}_{\text{cross}} \\ \vdots & \vdots & \ddots & \vdots \\ R^{M,1}_{\text{cross}} & R^{M,2}_{\text{cross}} & \cdots & A^M_{\text{auto}} \end{pmatrix}, \tag{6}$$

where M is the number of time series, and A^p_{auto}, $R^{p,q}_{\text{cross}}$ $(1 \leq p \leq M, 1 \leq q \leq M, p \neq q)$ denote the adjacency matrix of a recurrence network formed by time series p and the adjacency matrix of a cross recurrence network formed by two different time series p and q, respectively.

By computing the global cross-clustering coefficients, we quantify the degree distribution of the nodes in the recurrence network and explore the correlation between the time series corresponding to the multivariate recurrence network. First, we calculate the local cross-clustering coefficient $C^{p,q}_{\text{cross}}(i)$ of the nodes i in recurrence network p. The equation is described as

$$C^{p,q}_{\text{cross}}(i) = \begin{cases} \dfrac{\sum\limits_{j \neq z} R^{p,q}_{\text{cross}}(i,j) R^{p,q}_{\text{cross}}(i,z) A^q_{\text{auto}}(j,z)}{k^{p,q}_i (k^{p,q}_i - 1)}, & k^{p,q}_i \geq 2, \\ 0, & \text{else}, \end{cases} \tag{7}$$

where $R^{p,q}_{\text{cross}}(ij)$ denotes whether the adjacency matrix of the cross recurrence network is recurrence between node i and node j. If it is recurrence, its value is 1, otherwise it is 0. A^q_{auto} is the adjacency matrix of the recurrence network q. $k^{p,q}_i$ represents the cross degree of node i in the cross recurrence network. It is the number of nodes connected to node i in the recurrence network q. The global cross-clustering coefficient of the recurrence network p is the average of the local cross-clustering coefficient of all nodes. The equation is

$$C^{p,q}_{\text{global}} = \frac{1}{N_p} \sum_i C^{p,q}_{\text{cross}}(i), \tag{8}$$

where N_p denotes the number of nodes in the recurrence network p.

In order to explore the intrinsic characteristics of the cardiac system, according to information entropy, we propose the novel concept of cross-clustering coefficient entropy, which is calculated by

$$E^{p,q}_{\text{cross}} = -\sum_{i=1}^{N} P_c(i) \log P_c(i), \tag{9}$$

$$P_c(i) = \frac{C^{p,q}_{\text{cross}}(i)}{\sum_{i=1}^{N} C^{p,q}_{\text{cross}}(i)}, \tag{10}$$

where $C^{p,q}_{\text{cross}}(i)$ denotes the local cross-clustering coefficient of node i, and $E^{p,q}_{\text{cross}}$ represents the cross-clustering coefficient entropy of recurrence network p about recurrence network q $(E^{p,q}_{\text{cross}} \neq E^{q,p}_{\text{cross}})$. By calculating the cross-clustering coefficient entropy of the recurrence network in multivariate recurrence

networks with different scale factors, the variation of the nonlinear dynamic characteristic of ECG signals between healthy subjects and patients is analyzed.

2.3. Weighted networks

The multivariate recurrence network not only quantifies the variation within time series but also explores the correlation between time series through adjacency matrices R_M (M denotes the number of time series). In our study, we consider each time series as a node, and the weight of the edge between two nodes is the mean of the cross-clustering coefficient entropy of the recurrence network constructed by relevant time series. The formula is as follows:

$$\bar{E}_{p,q} = (E^{p,q}_{\text{cross}} + E^{q,p}_{\text{cross}})/2, \tag{11}$$

where $\bar{E}_{p,q}$ denotes the weight of edge between nodes p, q ($E^{p,q}_{\text{cross}} \neq E^{q,p}_{\text{cross}}$). Therefore, we get a fully connected and weighted network. The introduction of a weighted network[41] simplifies the description of the adjacency matrix R_M into $M \times M$, which reduces the size of the matrix. Figure 2 shows the process of converting multivariate time series to a multivariate recurrence network and weighted network when the number of time series is 3.

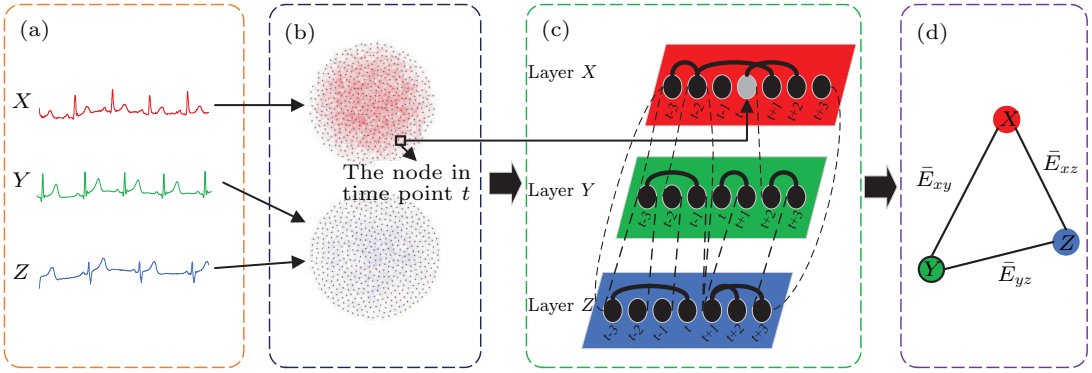


Fig. 2. Illustration of the steps from a multivariate time series to a weighted network (the number of time series is 3, $M = 3$). (a) Multivariate time series X, Y and Z, (b) recurrence network constructed by the time series X (upper) and cross recurrence network constructed by the time series Y and Z (lower), (c) multivariate recurrence network (take the nodes corresponding to time points near time point t), and (d) weighted network. In subplot (c), the solid lines indicate the intra-layer links and the dotted line represents the inter-layer connections. The node whose color is gray in layer X is the same node in time point t in the recurrence network in subplot (b). The connections between nodes near the time point t in the cross recurrence network in subplot (b) correspond to the connections between layer Y and layer Z in subplot (c).

In addition, the quantification of weighted networks is a well-established method.[42] In this paper, we adopt the average weighted path length (AWPL) and graph energy (GE)[43] to characterize the weighted networks. The equation for the AWPL is as follows:

$$L_{\text{w}} = \sum_{\alpha,\beta \in M} \frac{d(\alpha,\beta)}{n(n-1)}, \tag{12}$$

where M is the set of nodes (layers of the multivariate recurrence network) in the weighted network, n represents the number of nodes and $d(\alpha,\beta)$ represents the weighted shortest path length, defined as the minimum value of the sum of the weights among all the paths from node α to node β.[44] The equation of the GE is as follows:

$$E = \sum_{i=1}^{n} |\lambda_i|, \tag{13}$$

where λ_i denotes the eigenvalues of the adjacency matrix of the weighted network, and n represents the number of eigenvalues.

3. Experiments

3.1. Dataset 1

The ECG data applied in this experiment were carefully obtained from the Cardiovascular Medicine Department of Shanghai Changzheng Hospital.[45] The experimental subjects include four groups, i.e., healthy young (HY), healthy old (HO), myocardial ischemia (IO), and myocardial infarction (MI). In order to exclude the effect of other diseases on the experimental results, patients only suffer from one single disease. Healthy old people have normal blood pressure and no history of heart disease, and healthy young people have sinus rhythm. Among them, there are 40 cases of healthy young people (mean age ± standard deviation: 22.00±0.00 years), 23 cases of healthy old people (mean age ± standard deviation: 67.13±17.13 years), 42 cases of myocardial ischemia (mean age ± standard deviation: 68.03±15.03 years), and 11 cases of myocardial infarction (mean age ± standard deviation: 59.00±18.21 years).

The NHE-2000 multi-functional electrocardio-workstation developed by the Biomedical Electronics Research Institute of Nanjing University is adopted. This device is able to detect and display frequency components higher than 100 Hz (i.e., 0.1–2 kHz). Figure 3 depicts the 12-lead ECG waveform of a healthy elderly person captured by the NHE-2000 multi-functional electrocardio-workstation. As has been reported in some literature,[46,47] coronary artery disease is accurately identified by high-frequency ECG (HFECG, whose

frequency components are 0.02 Hz–2 kHz compared to routine ECG (0.05–100 Hz)). The standard ECG displays waveforms within the 0.05–100 Hz portion of the spectrum.[48] The standard 12-lead ECG records potential differences between prescribed sites on the body surface that reflect differences in transmembrane voltages in myocardial cells that occur during depolarization and repolarization within each cycle.[49] ANSI/AAMI affirmed that the low-frequency cutoff is 0.05 Hz for standard 12-lead ECGs.[50] A high-frequency cutoff of 100 Hz was considered adequate by the AHA in 1975 to maintain diagnostic accuracy during visual inspection of direct-writing tracings by electrocardiographers.[51] In our equipment, the lowest frequency is 0.02 Hz and the highest frequency is 2 kHz. The similarity is that a 12-lead approach to signal acquisition is taken.

The detection system has features such as wide frequency response and high sensitivity. Each single datum is denoted by 2 bytes (the highest 4 bits are set to zero) and the resolution of the data is 12 bits. Before the data are collected, the subjects lie down in a completely relaxed body position for several minutes. The next step is to first amplify the ECG signals, which are then transmitted to a special AD sample card located in the workstation with 1 kHz sampling rate per channel. In a time duration of about 1 min, as with routine ECG, the data are acquired at 12 different positions of the subject's body, with each lead corresponding to a different location and containing about 60000 data points. The first 1000 points of the data are deleted in order to reduce the influence of errors, and each lead selects 5000 points for subsequent calculations.

3.2. Dataset 2

On the other hand, we use the PTB Diagnostic ECG Database from Physionet[52] to validate the effectiveness of the method. In total, data from 231 subjects are applied in our analysis, 52 of which are healthy (HC), and rest of which are from one of three diseases: bundle branch block (BBB, 14), myocardial infarction (MI, 148), or cardiomyopathy/heart failure (CF, 17). Each record consists of mostly 60 s duration, sampled at 1000 Hz (60000 points), and 15 lead ECG signals. In our study, the remaining standard 12 leads of each record are selected except for the V_x, V_y and V_z leads. Similarly, starting at the 1001st point, we choose 5000 points to analyze the difference.

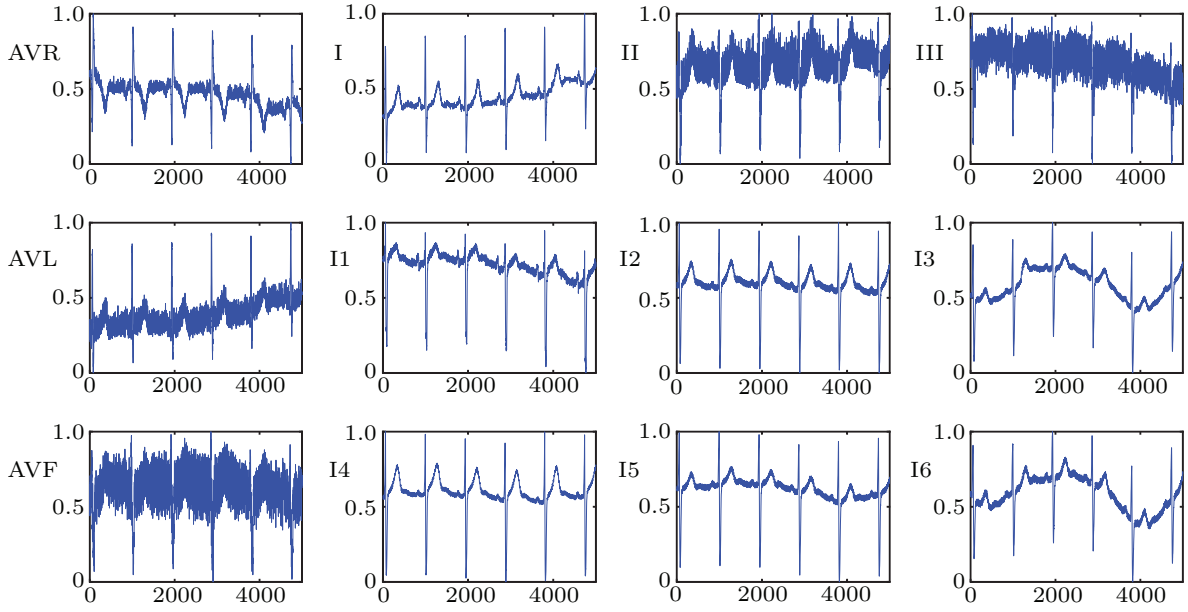


Fig. 3. The 12-lead signals collected by a high-frequency ECG from a healthy elderly person from dataset 1, where the number of every lead is 5000. The data have been normalized in this figure.

4. Analysis and results

4.1. Multiscale analysis

Coarse-grain processing of time series was proposed by Costa *et al.*[53] The original time series can be constructed as a shorter time series, and redundant information can be removed while extracting the basic time series features. For the given time series: $x(t)$ $(t = 1, 2, \ldots, N)$ (N denotes the number of elements in the time series), new time series are obtained by coarse granulation at different scale factors e

$$y(j)^e = \frac{1}{e} \sum_{t=(j-1)e+1}^{je} x(t), \tag{14}$$

where e represents the scale factor, $1 \le j \le N/e$.

To obtain new coarse-grained time series, the scale factors selected in this paper are from 1 to 20. Then multivariate recurrence networks are constituted and the cross-clustering coefficient entropy of the multivariate recurrence networks is cal-

culated. Figure 4 shows the results, in which the D-value represents the difference in ACCE (average cross-clustering coefficient entropy) from two groups at the same scale. In Fig. 4, we observe two different types of behaviors: (1) The ACCE measure for time series derived from HY and HO monotonically decreases. (2) The value of ACCE for time series derived from IO and MI decreases when the time scale varies from 1 to 10, then stabilizes at a constant value. Therefore, the optimal scale is obtained at scale 10. It is also found through experiments that the length of data has no effect on the selection of scale. From Fig. 4(a), we find that for all time scales, a higher value of ACCE is assigned to time series from young subjects. From a physiological point of view, the nonlinear dynamic characteristics of the human heartbeat signal may decrease as people get older. The conclusion that age affects the dynamical characteristics of complex systems is verified.[54,55] The complexity of the cardiac system is at its lowest in MI patients because of coronary artery occlusion and interruption of blood flow in patients with coronary heart diseases, leading to necrosis of part of the myocardium due to severe and persistent ischemia. Therefore, we obtain the result that cardiovascular disease can have an impact on complexity dynamics.[56]

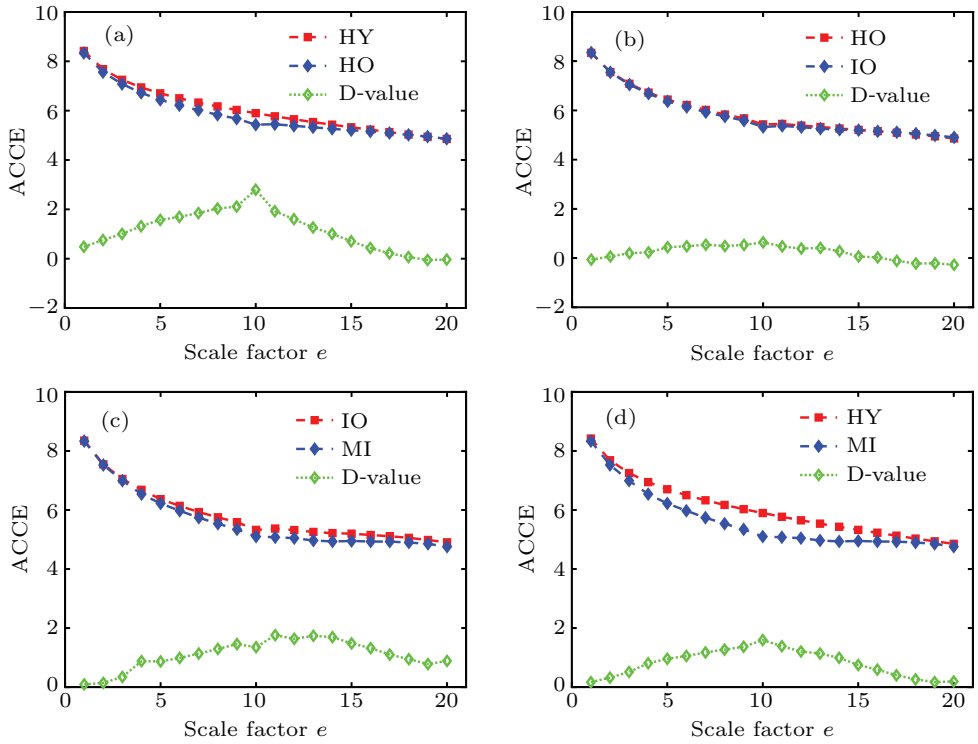


Fig. 4. ACCE of multivariate recurrence networks constructed from ECG data of different groups. Distributions of ACCE with a change in scale factor are presented. Comparisons of the differences between HY and HO, HO and IO, IO and MI, as well as HY and MI are described in subplots (a), (b), (c) and (d) respectively. In each subplot, the green line represents the difference between two groups under the same scale factors.

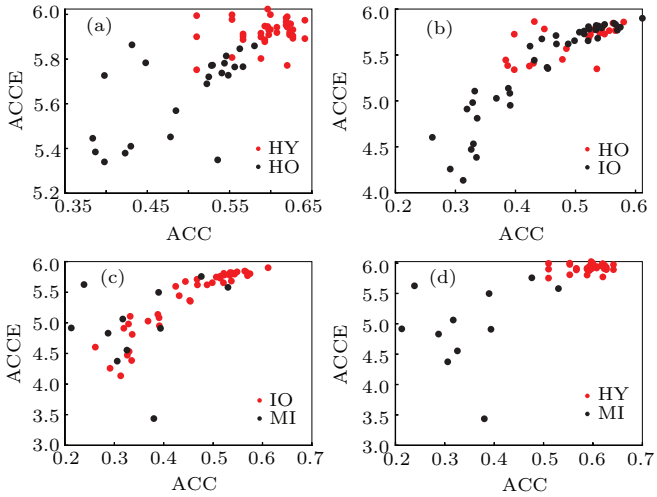


Fig. 5. Joint distribution of ACC and ACCE at optimal scale factor 10. (a) Difference in ACCE of HY and HO subjects. (b) Difference in ACCE of HO and IO subjects. (c) Difference in ACCE of IO and MI individuals. (d) Difference between the ACCE of HY and MI individuals.

For further verification, we calculate the cross-clustering coefficient for four different groups in scale 10. As we can see, Fig. 5 shows the joint distribution of the average cross-clustering coefficient (ACC) and the ACCE with a scale factor of 10. IO cannot be well distinguished because their hearts are less damaged and are similar to those of healthy elderly people. At the same time, it can be seen that the cross-clustering coefficient entropy is indeed capable of distinguishing between the HY and MI.

4.2. Multivariate recurrence networks

The 12-lead ECG signals of subjects are coarsely granulated at the optimal scale factor of 10 and first constructed to a multivariate recurrence network. In order to better demonstrate the changes of each lead under different physiological states, the cross-clustering coefficient entropy is calculated.

Similarly, the above operations are performed for all samples in the four categories.

Tables 1 and 2 show the mean value and standard deviation of the cross-clustering coefficient entropy of the same channel between different persons. Statistically, the ACCE of all leads in patients with MI, a life-threatening condition, are at the lowest magnitudes. At the same time, the ACCE for each lead is slightly higher in HO than in patients with IO. This possibly suggests that, although myocardial ischemia may cause some functional degradation to the heartbeat system and reduce the nonlinear complexity of the ECG signal to some degree, it is still in an early stage and not serious enough to cause damage to the dynamical system or even change its structure physically.[57] HY has the highest values of all leads.

Table 1. Cross-clustering coefficient entropy for 12-lead of different groups (limb leads).

Leads	I	II	III	AVR	AVL	AVF
HY	5.96±0.07	5.95±0.04	5.90±0.06	5.86±0.08	5.89±0.10	5.94±0.05
HO	5.71±0.25	5.61±0.23	5.48±0.52	5.59±0.33	5.57±0.48	5.62±0.39
IO	5.46±0.60	5.32±0.82	5.26±.87	5.52±0.44	5.54±0.45	5.14±0.98
MI	4.81±0.87	5.05±0.80	4.79±0.75	5.08±0.61	4.58±1.04	4.94±0.80
P-value	$\mathbf{9.57\times10^{-9}}$	$\mathbf{1.02\times10^{-5}}$	$\mathbf{9.05\times10^{-6}}$	$\mathbf{2.28\times10^{-6}}$	$\mathbf{2.18\times10^{-8}}$	$\mathbf{2.66\times10^{-6}}$
P, HY vs HO	0.19	0.12	0.07	0.03	0.07	0.26
P, HY vs IO	$\mathbf{1.01\times10^{-4}}$	$\mathbf{4.33\times10^{-5}}$	$\mathbf{2.44\times10^{-4}}$	$\mathbf{9.41\times10^{-4}}$	0.01	$\mathbf{8.09\times10^{-6}}$
P, HY vs MI	$\mathbf{1.71\times10^{-8}}$	$\mathbf{4.84\times10^{-4}}$	$\mathbf{6.09\times10^{-5}}$	$\mathbf{3.56\times10^{-6}}$	$\mathbf{4.51\times10^{-9}}$	$\mathbf{6.62\times10^{-4}}$
P, HO vs IO	0.20	0.21	0.55	0.89	0.99	0.03
P, HO vs MI	5.57×10^{-5}	0.12	0.05	0.01	$\mathbf{6.22\times10^{-5}}$	0.07
P, IO vs MI	$\mathbf{4.80\times10^{-3}}$	0.82	0.32	0.04	$\mathbf{3.73\times10^{-5}}$	0.96

Table 2. Cross-clustering coefficient entropy for 12-lead of different groups (chest leads).

Leads	V1	V2	V3	V4	V5	V6
HY	5.81±0.16	5.95±0.05	5.93±0.05	5.92±0.07	5.93±0.05	5.93±0.06
HO	5.65±0.34	5.55±0.24	5.59±0.35	5.63±0.27	5.57±0.32	5.49±0.54
IO	5.18±0.92	5.36±0.46	5.34±0.54	5.34±0.49	5.33±0.49	5.22±0.69
MI	4.81±0.72	4.94±0.40	4.95±0.79	4.70±1.12	4.50±.95	4.73±0.67
P-value	$\mathbf{1.81\times10^{-5}}$	$\mathbf{1.25\times10^{-12}}$	$\mathbf{2.46\times10^{-8}}$	$\mathbf{3.70\times10^{-10}}$	$\mathbf{6.89\times10^{-12}}$	$\mathbf{4.38\times10^{-8}}$
P, HY vs HO	0.74	$\mathbf{1.55\times10^{-4}}$	0.02	0.10	0.02	0.01
P, HY vs IO	$\mathbf{2.22\times10^{-4}}$	$\mathbf{2.50\times10^{-10}}$	$\mathbf{7.69\times10^{-7}}$	$\mathbf{6.00\times10^{-6}}$	$\mathbf{1.73\times10^{-6}}$	$\mathbf{1.06\times10^{-6}}$
P, HY vs MI	$\mathbf{7.94\times10^{-4}}$	$\mathbf{8.03\times10^{-10}}$	$\mathbf{1.89\times10^{-6}}$	$\mathbf{1.46\times10^{-9}}$	$\mathbf{1.37\times10^{-11}}$	$\mathbf{3.29\times10^{-6}}$
P, HO vs IO	0.02	0.14	0.14	0.10	0.19	0.22
P, HO vs MI	0.01	$\mathbf{3.50\times10^{-3}}$	0.01	$\mathbf{1.68\times10^{-5}}$	$\mathbf{2.31\times10^{-6}}$	0.03
P, IO vs MI	0.74	0.16	0.39	$\mathbf{4.20\times10^{-3}}$	$\mathbf{3.06\times10^{-4}}$	0.44

Moreover, ANOVA and post-hoc are applied to assess the statistical group differences in the topological characteristics of multivariate recurrence networks for four groups. The *P* value is considered statistically significant if it is less than 8.33×10^{-3} after Bonferroni correction. As we can see, the *P* value of ANOVA for all leads is far less than 8.33×10^{-3}. This indicates that the cross-clustering coefficient entropy can effectively differentiate between healthy individuals and patients. From the post hoc results, compared to HY, the V2 chest leads in HO are affected. At the same time, we find that the distinction between HO and IO is not very effective. Moreover, compared with IO and HO, the I and AVL leads in the limb leads and the V4 and V5 leads in the chest leads of the MI produce relatively large changes.

After analyzing the adjacency matrix of the multivariate recurrence network, the weighted network is constructed. In Fig. 6, we present the weighted networks of the HY subjects and MI subjects. In order to describe more clearly the changes in the status of the cardiac system in these four groups, the network parameters of the weighted network constituted by the four groups are calculated: the AWPL and GE. The results are shown in Fig. 7. As we can see from Fig. 7, the AWPL and GE of HY are at the highest values, while the HO group and patients with IO are in the middle due to age and cardiovascular disease, respectively. Patients with MI have the smallest values due to the severe damage to the myocardium and the possible alteration of the heart's structure.

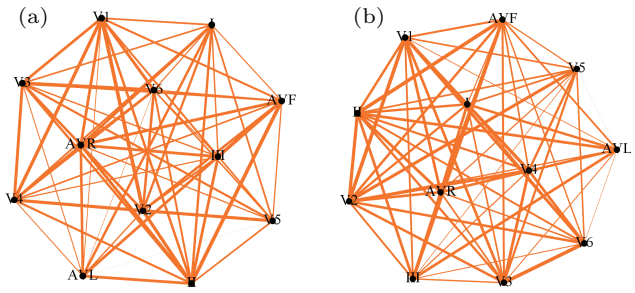


Fig. 6. Fully connected, undirected, and weighted network of (a) HY subjects and (b) MI subjects. Each black circle represents a node, and each orange line indicates an edge. The width of the edge is proportional to the size of the weight of the edge. The I-AVF are the names of limb leads, and V1–V6 are the names of chest leads.

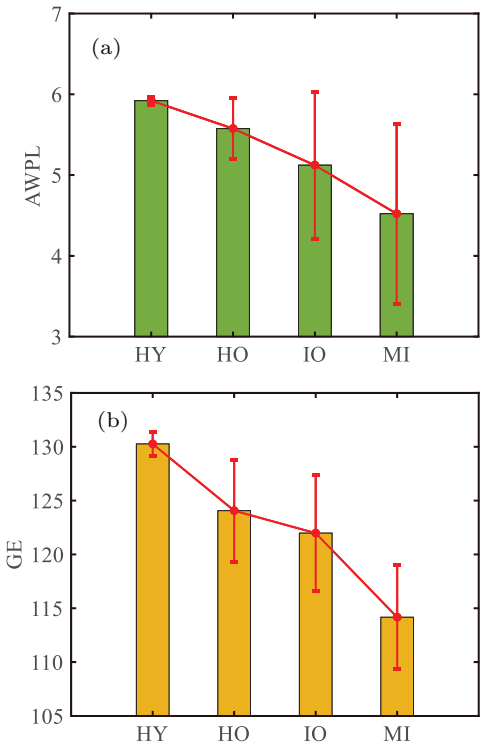


Fig. 7. Bar charts and error bars represent the parameters for four different people. (a) AWPL of four different groups. (b) GE of four different people, where HY, HO, IO, and MI indicate four categories of people, respectively.

4.3. Comparison with other studies

The same steps are applied to dataset 2 to verify the validity of the method in this paper. We first need to calculate the cross-clustering coefficient entropy for multivariate recurrence networks with scales between 1 and 20, and compare the differences between the average cross-clustering coefficient entropy for any different pair of groups at each scale. If all the differences are the largest on the same scale, then the definite scale is the optimal scale. If the differences reach the maximum at different scales, the scale with the larger difference is chosen as much as possible. We calculate and compare to find the optimal scale 2 for dataset 2. Figure 8 shows the results of the AWPL and GE at a scale factor of 2. As we can see, two parameters, namely the AWPL and GE, are at their maximum for HC. Compared to healthy individuals, the subjects of BBB have lower AWPL and GE. The reason for this is that the information capacity of BBB ECG signals reduces, the complexity is reduced, and more physiological adaptations are lost to the lesser adaptive capacity inside and outside the heart.[58]

On the one hand, patients with BBB do not have other associated heart diseases. Patients with BBB, on the other hand, have slightly smaller parameters than healthy people. The type of BBB may be RBBB, which is generally considered benign.[59] As we can see, two parameters of patients diagnosed with MI or CF are in the lowest position. We consider that it is possible that alterations in cardiac structure have led to a reduction in the complexity and nonlinear dynamic characteristics of the cardiac system.

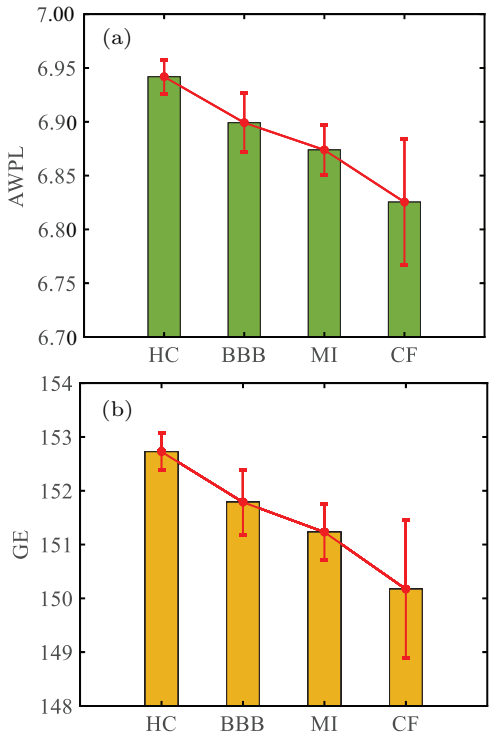


Fig. 8. Bar charts and error bars representing the parameters from four different people. (a) AWPL of four different groups, (b) GE of four different people, where HC, BBB, MI, and CF indicate four categories of people respectively.

Table 3. Comparison between the accuracies of this method and the other published results between healthy and myocardial infarction.

Author/Year	Methods	Accuracy
Dohare *et al.*[60] 2018	14 feature vectors + SVM	96.66%
Sadhukhan *et al.*[61] 2018	Phase of DFT + LR	95.6%
Liu *et al.*[62] 2019	dual-Q TQWT + KNN	86.3%
Kayikcioglu *et al.*[63] 2020	Choi–Williams feature + weighted KNN	94.23%
Ma *et al.*[64] 2021	AWD/AWCC + SVM	93.3%
Our proposed (dataset 1[57])	AWPL + GE + SVM	94.1%
Our proposed (dataset 2[52])	AWPL + GE + SVM	95.58%

The parameters of the weighted network serve as the input for the features of the SVM,[65] which are the AWPL and GE. It is a linear classification method. To solve the issues of overfitting and sample imbalance, we adopt stratified ten-fold cross-validation. This splits the data into ten folds, where the original ratio of samples is preserved. Each fold acts once as

the test split, while the remaining data act as the training data. This allows the models to retain reliable estimates on unseen data. The accuracy of the classification of healthy people and MI is illustrated in Table 3. Compared with the method in the literature,[60] only two parameters are applied in our study. In contrast to other methods, our approach achieves high accuracy.

5. Conclusion and perspectives

The investigation of typical nonlinear complex systems like the heart system has been a hot topic of research. Combining recurrence phenomena in heartbeat dynamics with complex networks provides an idea for analyzing the nonlinear dynamical characteristics of the cardiac system. In this paper, we construct multivariate recurrence networks and perform multiscale analysis on them. Meanwhile, the concept of cross-clustering coefficient entropy is innovatively proposed to achieve a distinction between healthy subjects and coronary heart disease patients. The results show that the complexity of ECG signals decreases in elderly people compared with young people due to the weakening of cardiovascular functions with age. The conclusion that age affects the dynamical characteristics of complex systems is verified.[54,55] On the other hand, the ECG signals of subjects with myocardial ischemia and myocardial infarction have lower complexity due to the fact that patients may have narrowed or blocked coronary vessels. This is consistent with the literature.[57] At the same time, we get 94.1% and 95.58% accuracy in two datasets between healthy subjects and persons diagnosed as myocardial infarction. Therefore, this method can effectively detect changes in cardiac disease status, which can provide ideas for the diagnosis of heart-related diseases.

The multivariate time series describes the internal characteristics and state changes of the cardiac system more adequately than a single time series. By constructing multivariate recurrence networks, we can explore the nonlinear dynamic characteristics of the complex system, which can be applied to different fields in the real world, such as paleoclimatology, neuroscience, economics, etc., and provide new ideas to investigate the dynamical properties and different state transitions of the system. However, when the number of time series in the complex system being applied becomes large, the adjacency matrix corresponding to the multivariate recurrence network becomes too large and the computational complexity increases. In addition, gender may also be a factor affecting the cardiac system, and we will also explore the differences in the cardiac system between different genders in our future studies.

Acknowledgements

Project supported by the Xuzhou Key Research and Development Program (Social Development) (Grant No. KC21304) and the National Natural Science Foundation of China (Grant No. 61876186).

References

[1] Zou Y, Donner R V, Marwan N, Donges J F and Kurths J 2019 *Physics Reports* **787** 1
[2] Wolf A, Swift J B, Swinney H L and Vastano J A 1985 *Physica D* **16** 285
[3] Rosenstein M T, Collins J J and Luca C 1993 *Phys D* **65** 117
[4] Deryaübeyli E 2010 *Expert Syst. Appl.* **37** 1192
[5] Richman J S and Randall M J 2000 *Am. J. Physiol.: Heart Circ. Physiol.* **278** H2039
[6] Bandt C and Pompe B 2002 *Phys. Rev. Lett.* **88** 174102
[7] Shannon and C E 2014 *Bell Syst. Tech. J.* **30** 50
[8] Mishra A K and Raghav S 2010 *Biomed. Signal Process. Control* **5** 114
[9] Burrough and P. A 1981 *Nature* **294** 240
[10] Paumgartner D, Losa G A and Weibel E R 2011 *J. Microsc.* **121** 51
[11] Albert R and Barabási A L 2002 *Rev. Mod. Phys.* **74** 47
[12] Zhang J and Small M 2006 *Phys. Rev. Lett.* **96** 238701
[13] Zhang J, Luo X, Nakamura T, Sun J and Small M 2007 *Phys. Rev. E* **75** 016218
[14] Xu X, Jie Z and Small M 2008 *Proc. Natl. Acad. Sci. USA* **105** 19601
[15] Lacasa L, Luque B, Ballesteros F, Luque J and Nuno J C 2008 *Proc. Natil. Acad. Sci. USA* **105** 4972
[16] Marwan N, Romano M C, Thiel M and Kurths J 2007 *Phys. Rep.* **438** 237
[17] Marwan N, Donges J F, Zou Y, Donner R V and Kurths J 2009 *Phys. Lett. A* **373** 4246
[18] Sun X R, Small M, Zhao Y and Xue X P 2014 *Chaos* **24** 024402
[19] Marwan N and Kurths J 2015 *Chaos* **25** 097609
[20] Lekscha J and Donner R V 2020 *NPGeo* **27** 261
[21] Kachhara S and Ambika G 2019 *Europhys. Lett.* **127** 60004
[22] Ramírezávila G M, Gapelyuk A, Marwan N, Walther T, Stepan H, Kurths J and Wessel N 2013 *Philos Trans. A Math. Phys. Eng.* **371** 20110623
[23] Subramaniyam N P and Hyttinen J 2014 *Phys. Lett. A* **378** 3464
[24] Subramaniyam N P and Hyttinen J 2013 *International IEEE/EMBS Conference on Neural Engineering* 605
[25] Gao Z K, Zhang X W, Jin N D, Donner R V, Marwan N and Kurths J 2013 *Europhys. Lett.* **103** 50004
[26] Gao Z K, Dang W D, Yang Y X and Cai Q 2017 *Chaos* **27** 035809
[27] Feldhoff J H, Donner R V, Donges J F, Marwan N and Kurths J 2012 EGU General Assembly 14 EGU2012-5450
[28] Zhang N, Wei N and Li K 2020 *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2020** 3788
[29] Kachhara S and Ambika G 2020 *Chaos* **30** 123106
[30] Eroglu D, Marwan N, Stebich M and Kurths J 2018 *Phys. Rev. E* **97** 012312
[31] Xu H, Wang M and Yang W 2020 *Complexity* **2020** 5841609
[32] Takens F 1981 *Detecting strange attractors in turbulence* (Berlin Heidelberg: Springer)
[33] Marwan N 2011 *Int. J. Bifurc. Chaos* **21** 1003
[34] Matassini L, Kantz H, Hołyst J and Hegger R 2002 *Phys. Rev. E* **65** 021102
[35] Thiel M, Romano M C, Kurths J, Meucci R, Allaria E and Arecchi F T 2002 *Physica D* **171** 138
[36] Marwan N, Carmen Romano M, Thiel M and Kurths J 2007 *Phys. Rep.* **438** 237
[37] Schinkel S, Dimigen O and Marwan N 2008 *Eur. Phys. J. Spec. Top.* **164** 45
[38] Kraemer K H, Donner R V, Heitzig J and Marwan N 2018 *Chaos* **28** 085720
[39] Marwan N and Kurths J 2002 *Phys. Lett. A* **302** 299
[40] Kraemer K H and Marwan N 2019 *Phys. Lett. A* **383** 125977

[41] Barrat A, Barthelemy M, Pastor-Satorras R and Vespignani A 2004 *Proc. Natl. Acad. Sci. USA* **101** 3747
[42] Boccaletti S, Latora V, Moreno Y, Chavez M and Hwang D U 2006 *Phys. Rep.* **424** 175
[43] Balakrishnan R 2004 *Linear Algebra Appl.* **387** 287
[44] Xu X J, Wu Z X and Wang Y 2005 *Int. J. Mod. Phys. C* **17** 521
[45] Wang J, Ning X, Ma Q, Bian C and Chen Y 2005 *Phys. Rev. E* **71** 062902
[46] Abboud S, Belhassen B, Miller H I, Sadeh D and Laniado S 1986 *J. Electrocardiol.* **19** 371
[47] Leinveber P, Halamek J and Jurak P 2016 *J. Electrocardiol.* **49** 902
[48] Garvey J L 2006 *Emerg. Med. Clin. North Am.* **24** 209
[49] Kligfield P, Gettes L, Bailey J, Childers R, Deal B, Hancock E, Herpen G, Kors J, Macfarlane P, Mirvis D, Pahlm O, Rautaharju P and Wagner G 2007 *Heart Rhythm* **4** 394
[50] Bailey J J, Berson A S, Garson A, Jr., Horan L G, Macfarlane P W, Mortara D W and Zywietz C 1990 *Circulation* **81** 730
[51] Kossmann C E, Brody D A, Burch G E, Hecht H H, Johnston F D, Kay C, Lepeschkin E, Pipberger H V, Pipberger H V, Baule G, Berson A S, Briller S A, Geselowitz D B, Horan L G and Schmitt O H 1967 *Circulation* **35** 583
[52] Goldberger A L, Amaral L A, Glass L, Hausdorff J M, Ivanov P C, Mark R G, Mietus J E, Moody G B, Peng C K and Stanley H E 2000 *Circulation* **101** E215
[53] Costa M, Goldberger A L and Peng C K 2002 *Phys. Rev. Lett.* **89** 068102
[54] Lipsitz L A and Goldberger A L 1992 *JAMA* **267** 1806
[55] Vaillancourt D E, Sosnoff J J and Newell K M 2004 *J. Appl. Physiol.* **97** 454
[56] Valenza G, Citi L and Barbieri R 2014 *PLoS One* **9** e105622
[57] Yang X, Wang Z, He A and Wang J 2020 *Physica A* **559** 125021
[58] Yang M J 2005 *Conf. Proc. IEEE Eng. Med. Biol. Soc.* **2005** 4630
[59] Eriksson P, Wilhelmsen L and Rosengren A 2005 *Eur. Heart J.* **26** 2300
[60] Dohare A K, Kumar V and Kumar R 2018 *Appl. Soft Comput.* **64** 138
[61] Sadhukhan D, Pal S and Mitra M 2018 *IEEE Trans. Instrum. Meas.* **67** 2303
[62] Liu J, Zhang C, Ristaniemi T and Cong F 2019 *Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.* **2019** 1496
[63] Kayikcioglu İ, Akdeniz F, Köse C and Kayikcioglu T 2020 *Comput. Electr. Eng.* **84** 106621
[64] Ma Z Y, Yang X D, He A J, Ma L and Wang J 2022 *Acta Phys. Sin.* **71** 050501 (in Chinese)
[65] Hammad M and Wang K 2017 *Proceedings of the* 2017 *International Conference on Biometrics Engineering and Application* pp. 39–44