

DOI:10.3969/j.issn.1005-202X.2024.05.008

医学信号处理与医学仪器

基于心率变异性的阵发性心房颤动预测方法

牛晓东^{1,2}, 柴国强³, 王大为³, 卢莉蓉¹, 韩玲娜⁴, 连亚军⁵

1. 长治医学院生物医学工程系, 山西 长治 046000; 2. 长治医学院山西省智能数据辅助诊疗工程研究中心, 山西 长治 046000; 3. 山西师范大学物理与信息工程学院, 山西 太原 030000; 4. 长治医学院生理学教研室, 山西 长治 046000; 5. 长治医学院附属和平医院全科医疗科, 山西 长治 046000

【摘要】基于心率变异性(HRV)的特征分析,提出一种患者阵发性房颤(PAF)发作的预测系统方法。首先,基于一种新的自适应滤波技术逐次平滑滤波并粗粒化HRV后,采用熵量化HRV在多个自适应尺度的复杂性特征;其次,特征经Min-Max归一化和序列前向选择特征子集,输入支持向量机识别HRV类型,预测PAF发作。经50例时长5 min HRV序列集的五折交叉验证,得到最优预测结果为:准确率98%,敏感性100%,特异性96%,性能表现优越。另外,实验表明远离和紧随PAF时的HRV复杂性特征值在不同频段内,分别具有不同的显著变化($P<0.05$),反映受试者神经系统调节心脏节律改变,以及调控机体、应激等适应外界环境变化能力的下降。

【关键词】阵发性房颤;心率变异性;熵;尺度;积分均值模式分解

【中图分类号】R318.04

【文献标志码】A

【文章编号】1005-202X(2024)05-0579-09

Prediction of paroxysmal atrial fibrillation based on heart rate variability analysis

NIU Xiaodong^{1,2}, CHAI Guoqiang³, WANG Dawei³, LU Lirong¹, HAN Lingna⁴, LIAN Yajun⁵

1. Department of Biomedical Engineering, Changzhi Medical College, Changzhi 046000, China; 2. Shanxi Engineering Research Center for Intelligent Data Assisted Diagnosis and Treatment, Changzhi Medical College, Changzhi 046000, China; 3. School of Physics and Information Engineering, Shanxi Normal University, Taiyuan 030000, China; 4. Department of Physiology, Changzhi Medical College, Changzhi 046000, China; 5. Department of General Practice, Heping Hospital Affiliated to Changzhi Medical College, Changzhi 046000, China

Abstract: Based on the analysis of heart rate variability (HRV), a prediction method for paroxysmal atrial fibrillation (PAF) attacks is proposed. A new adaptive filtering technique is used for smoothing and coarse graining of HRV, followed by entropy-based quantification of HRV complexity at multiple adaptive scales. After the features are normalized by Min-Max, feature subsets are selected by sequential forward selection method, and then input to support vector machine to identify HRV types and predict PAF attacks. Through 5-fold cross-validation on a set of 50 HRV sequences each lasting 5 minutes, the optimal prediction results are obtained: 98% accuracy, 100% sensitivity, 96% specificity, demonstrating excellent performance. In addition, the experiment shows significant changes ($P<0.05$) in the complexity eigenvalues of HRV far away from and close to PAF at different frequency bands, reflecting alterations in nervous system regulation of cardiac rhythm and a decline in the ability to adapt to external environmental changes such as stress regulation.

Keywords: paroxysmal atrial fibrillation; heart rate variability; entropy; scale; integral mean mode decomposition

前言

房颤 (Atrial Fibrillation, AF) 是一种严重威胁人

类健康的心血管疾病。在一个31 230例受试者分析中,35岁及以上人群的AF患病率为1.14%,并且34%的患者不知道自己的病情^[1]。另外一项1年期随访研究表明,确诊AF的患者死亡率可能超过11%^[2]。阵发性房颤 (Paroxysmal Atrial Fibrillation, PAF) 是一种无症状的早期AF^[3],长期PAF将转变为AF。如果能预测患者PAF发作,并及时采用射频消融等技术有效干预和治疗,可以防止心脏发生重构,PAF转变为AF。PAF预测一般通过特征分类方法实现,这些

【收稿日期】2023-12-13

【基金项目】国家自然科学基金(62201332, 62201333);山西省基础研究计划(自由探索类)(20210302124328);长治医学院博士科研启动基金(BS202123)

【作者简介】牛晓东,博士,副教授,研究方向:生物医学信号处理, E-mail: niuxd@czmc.edu.cn

特征来源于心电图信号(Electrocardiogram, ECG)或心率变异性(Heart Rate Variability, HRV)。在临床上,通常通过检测ECG信号中的P波变异确诊PAF。然而,ECG的非线性、非平稳性和高信噪比使得典型P波变异难以检测。对于5 min的ECG,经P波变异预测PAF的准确率(Accuracy, ACC)仅为75.6%^[4]。HRV是ECG信号中NN(RR)间期变化,比ECG更容易获得,并且噪声对HRV的影响要小于ECG。近些年来,通过HRV特征识别PAF被广泛应用^[5]。除此之外,一些文献研究表明HRV的特征包含调节心率的神经系统潜在信息^[6],这些特征的显著变化意味着与神经系统受损相关的心血管疾病的发生^[7-12]。因此,HRV也受到广泛关注。

用以识别PAF的HRV特征主要来自其时域、频域,以及多种状态空间域^[13]。一些基于非线性系统理论分析方法抽取的HRV特征,可以更深入分析心率动力学,可能揭示疾病导致心脏自主调节的细微异常变化^[13-15]。因此,HRV特征也被用于PAF识别^[16-18]。熵可以度量HRV复杂性,反映生物系统携带信息^[19],特别是近似熵(AE)^[20]、样本熵(SE)^[21]等适用于分析短且包含噪声的数据,因此,熵被广泛应用于对AF、猝死、心衰等心脑血管疾病的快速识别^[16-18, 22-25]。然而,AE、SE分析某些疾病赋予患病系统高熵值,而健康系统低熵值,这与患病系统相比健康系统的熵值应该降低是不一致的,原因是这些熵测度基于单尺度分析,没有考虑生理系统的多尺度特性^[26-27]。为了克服这个缺点,文献[26]提出在不同尺度上量化系统复杂性的多尺度熵(MSE),然而MSE中的多尺度化本质是对数据的线性平滑和抽取,容易导致高频分量丢失^[28]。其次,利用的均值滤波器的频率响应不能防止混叠^[29]。最终,导致抽取的特征不能准确量化HRV在多尺度上的复杂性,因此不能很好地识别PAF。

基于经验模态分解(Empirical Mode Decomposition, EMD)^[30],文献[31]提出一种信号的自适应正交滤波技术,有效地减少EMD方法的混叠问题,被称为IMMD方法^[32]。IMMD分解信号的完备性和正交性,克服线性滤波器处理非线性HRV的缺点。因此,将IMMD与熵结合可以得到HRV在多个自适应尺度上的复杂性,克服MSE缺点,更为准确地抽取HRV复杂性特征。这种在多个自适应尺度上的熵度量方法,本文称为自适应尺度熵(Adaptive-Scale Entropy, ASE)。

多种分类器已应用于PAF特征分类,实现PAF预测,包括贝叶斯线性分类器^[4, 33]、K最邻近(K-Nearest Neighbor, KNN)^[34]、随机森林(Random Forest, RF)^[5]、

极限学习机(Extreme Learning Machine, ELM)^[35]、卷积神经网络(Convolutional Neural Network, CNN)^[18]以及支持向量机(Support Vector Machine, SVM)^[16-17, 36]等。其中,SVM是基于统计学习理论中的结构风险最小化准则,有效解决小样本下的分类问题,它比其它分类器具有优异的泛化能力。因此,本文采用SVM进行特征分类,实现PAF预测。

1 方法

PAF预测流程图见图1。对于HRV数据:首先,采用ASE方法得到HRV在多个自适应尺度上的分量,并计算各个分量的熵,作为抽取的HRV特征;其次,采用Wilcoxon符号秩检验,与序列前向选择(Sequential Forward Selection, SFS)方法选取最优特征子集^[37];最后,将特征子集输入SVM,并采用五折交叉验证得到最优分类模型,实现PAF预测。

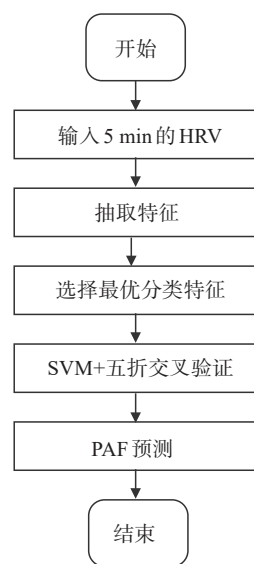


图1 PAF预测流程图

Figure 1 PAF prediction flowchart

1.1 HRV特征抽取

ASE方法中:IMMD自适应分解HRV得到一套它在不同尺度上的剩余分量集合;集合中每一元素被粗粒化;计算集合中每一元素尺度;计算集合中每一元素的熵,作为HRV特征。

1.1.1 分量集合 IMMD自适应分解HRV序列得到剩余分量集合方法^[31]:

(1)HRV序列记为 $x(n)$ 。定义 $X_{jk}(j, k=1, 2, \dots; k>j)$ 为数据序列 $x(n)=[x_1, x_2, \dots, x_i, \dots, x_n]$ 邻接两个极值 x_j 和 x_k 之间局部部分, X_{jk} 长度为 $\tau=j-k+1$ 。局部序列 X_{jk} 均值为:

$$m_{jk} = \frac{1}{\tau} \sum_{i=j}^k x_i \quad (1)$$

其中, m_{jk} 固定于局部中点 $t = j + \frac{\tau}{2}$ 。基于式(1), 求得数据序列所有局部均值点。

(2) 三次样条插值数据序列所有局部均值点构建数据均值序列 $m(n)$ 。

(3) 原型模式函数(Proto Mode Function, PMF)^[30-32] 即为:

$$\text{PMF}(n) = x(n) - m(n) \quad (2)$$

上述步骤(1)~(3)称为一次均值筛选过程, PMF 记为 PMF_1 。

(4) PMF_1 迭代重复均值筛选过程 $k-1$ ($k=2, 3, \dots$) 次, 得到 PMF_k 。当 PMF_k 满足柯西筛选停止准则^[30-32]:

$$\text{SD} = \sum_{i=0}^T \left[\left| \text{PMF}_{k-1}(n) - \text{PMF}_k(n) \right|^2 \right] / \left[\text{PMF}_{k-1}^2(n) \right] < 0.2 \sim 0.3 \quad (3)$$

筛选停止, PMF_k 即为 IMF_1 。

(5) 剩余分量 $r_1(n) = x(n) - \text{IMF}_1$ 作为新信号, 重复上述所有过程得到 IMF_2 及 $r_2(n)$ 。同理, 可得信号其余 IMF_k ($k=1, 2, 3, \dots, m$) 分量及剩余分量 $r_k(n)$ 。

(6) 每一个剩余分量的所有局部数据由其局部均值替代[式(1)], 最后, 所有剩余分量构成一个原时间序列的“趋势”集合 $R = \{r_1(n), r_2(n), \dots, r_m(n)\}$ 。R 即为 IMMD 逐次自适应滤除 HRV 序列中高频分量后, 剩余低频成分集合。

1.1.2 元素粗粒化 R 中每一个元素的所有局部数据由其局部均值替代[式(1)], 得到粗粒化的 R。元素粗粒化示意图见图2。

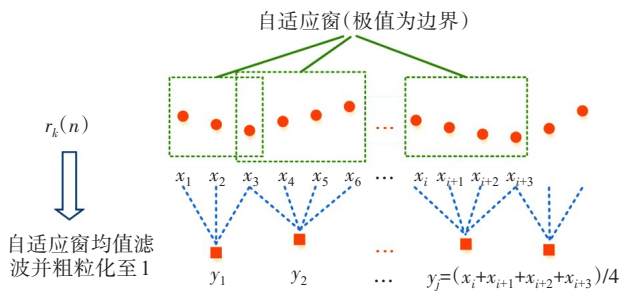


图2 元素粗粒化示意图

Figure 2 Schematic diagram of coarse graining of elements

1.1.3 计算元素尺度 定义 $\tau_i^{(k)}$ ($i=1, 2, \dots$) 为 IMF_k 中第 i 个局部 (IMF_k 中第 i 和 $i+1$ 个极大值之间局部部分) 尺度。因为 IMF_k 频谱具有窄带特性, 因此 IMF_k

所有局部尺度 $\tau_i^{(k)}$ 近似相等, 定义所有局部尺度 $\tau_i^{(k)}$ 的平均值为 IMF_k 的时间尺度, 公式为:

$$\tau^{(k)} = \frac{1}{\max(i)} \sum_{i=1}^{\max(i)} \tau_i^{(k)} \quad (4)$$

根据 IMMD 方法的分解过程, 这个尺度即是滤除 IMF_k 的自适应滤波窗的宽度, 即是 R 中第 k 个元素 $r_k(n)$ 尺度。

1.1.4 计算元素的熵 因为 AE、SE、排列熵(Permutation Entropy, PE)^[38]、模糊熵(Fuzzy Entropy, FE)^[39]、散布熵(Dispersion Entropy, DE)^[40] 5 种熵适合分析长度较短并且含有噪声数据的复杂性, 近些年来成为常见且流行的熵度量。本文分别采用这 5 种熵, 度量粗粒化集合 R 复杂性。对于一个长度为 N 的时间序列 $x(i)$, 5 种熵的度量方法如下所述:

(1) AE: 将序列 $\{x_1, x_2, \dots, x_N\}$ 重构得到 $N-m+1$ 个 m 维向量 $X_m(i) = [x_i, x_{i+1}, \dots, x_{i+m-1}]$, $i=1, 2, \dots, N-m+1$ 。任意两个 m 维向量之间的距离 $d[X_m(j) - X_m(i)]$ 定义为这两个向量内对应元素差值绝对值的最大值, 若距离 \leq 阈值 r , 称为两个向量模板匹配, 与 $X_m(i)$ 模板匹配的向量的概率记为 $C_i^m(r)$:

$$C_i^m(r) = \frac{\text{Number} \{j | d[X_m(j) - X_m(i)] \leq r\}}{N - m + 1} \quad (5)$$

定义 $\phi^m(r)$ 为:

$$\phi^m(r) = (N - m + 1)^{-1} \sum_{i=1}^{N-m+1} \ln[C_i^m(r)] \quad (6)$$

将重构向量的维数增加到 $m+1$, 重复上述步骤, 得到与 $m+1$ 维向量 $X_{m+1}(i)$ 模板匹配的概率 $C_i^{m+1}(r)$, 以及相应 $\phi^{m+1}(r)$ 。则条件概率 AE 定义为:

$$\text{AE}(m, r, N) = \phi^m(r) - \phi^{m+1}(r) \quad (7)$$

其中, m 一般取为 2, r 一般取为序列的标准差的 0.2 倍。

(2) SE: 将序列 $\{x_1, x_2, \dots, x_N\}$ 重构得到 $N-m+1$ 个 m 维向量 $X_m(i) = [x_i, x_{i+1}, \dots, x_{i+m-1}]$, $i=1, 2, \dots, N-m+1$ 。任意两个 m 维向量之间的距离 $d[X_m(j) - X_m(i)]$ 定义为两向量内对应元素差值绝对值的最大值, 若距离 \leq 阈值 r , 称为两个向量模板匹配, 与 $X_m(i)$ 模板匹配的概率记为 $B_i^m(r)$ (没有自匹配):

$$B_i^m(r) = \frac{\text{Number} \{j | d[X_m(j) - X_m(i)] \leq r, j \neq i\}}{N - m} \quad (8)$$

定义 $B^m(r)$ 为:

$$B^m(r) = (N - m)^{-1} \sum_{i=1}^{N-m} B_i^m(r) \quad (9)$$

将重构向量的维数增加到 $m+1$, 重复上述步骤, 得到 $m+1$ 维向量模板匹配的概率 $A_i^m(r)$ 以及相应 $A^m(r)$ 。则 SE 定义为:

$$SE(m, r, N) = -\ln \left[\frac{A^m(r)}{B^m(r)} \right] \quad (10)$$

m 一般取为2, r 一般取为序列的标准差的0.2倍。

(3) PE: 将序列 $\{x_1, x_2, \dots, x_N\}$ 重构得到 $N-m+1$ 个 m 维向量 $X_m(i) = [x_i, x_{i+1}, \dots, x_{i+m-1}]$, $i=1, 2, \dots, N-m+1$ 。对 $X_m(i)$ 序列中的各个元素进行数值大小的排序, 得到排序类型 π_i (对于 m 个元素, 其数值排序类型 π 共有 $m!$ 种)。则 $N-m+1$ 个 m 维向量具有相同的排序类型 π_i 的概率为:

$$p(\pi_i) = \frac{\text{Number} \{i | i \leq N - m + 1, (x_i, x_{i+1}, \dots, x_{i+m-1}) \text{ has type } \pi_i\}}{N - m + 1} \quad (11)$$

假设 $N-m+1$ 个 m 维向量共有不同的排序类型个数为 j , 则 PE 定义为:

$$PE = -\sum_{i=1}^j p(\pi_i) \ln p(\pi_i) \quad (12)$$

其中, m 一般取为3~7, 推荐为5~7^[41]。

(4) FE: 将序列 $\{x_1, x_2, \dots, x_N\}$ 重构得到 $N-m+1$ 个 m 维向量 $X_m(i) = [x_i, x_{i+1}, \dots, x_{i+m-1}] - \frac{1}{m} \sum_{j=i}^{i+m-1} x_j$, $i=1, 2, \dots, N-m+1$ 。任意两个 m 维向量之间的距离 d_{ij}^m 为两向量内对应元素差值绝对值的最大值, 则两个向量的相似度定义为:

$$D_{ij}^m(n, r) = \exp(-(d_{ij}^m)^n / r) \quad (13)$$

定义 $\phi^m(n, r)$ 为:

$$\phi^m(n, r) = \frac{1}{N - m} \sum_{i=1}^{N-m} \left(\frac{1}{N - m - 1} \sum_{j=1, j \neq i}^{N-m} D_{ij}^m \right) \quad (14)$$

则 FE 定义为:

$$FE(m, n, r) = \ln \phi^m(n, r) - \ln \phi^{m+1}(n, r) \quad (15)$$

其中, m 一般取为2, n 一般取为2, r 一般取为序列的标准差的0.2倍。

(5) DE: 将序列 $\{x_1, x_2, \dots, x_N\}$ 通过正态累积分布函数从0到1映射到 y 上:

$$y_i = \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{x(i)} e^{-\frac{(t-\mu)^2}{2\sigma^2}} dt \quad (16)$$

其中, μ 和 σ 代表序列的均值和标准偏差。将 y_i 通过 $z_i^c = \text{round}(c \cdot y_i + 0.5)$ 映射到整数1到 c 上, $\text{round}(\cdot)$ 表示取整函数。 z_i 重构得到 m 维向量 $z_j^{m,c} = \{z_j^c, z_{j+1}^c, \dots, z_{j+m-1}^c\}$, $j=1, 2, \dots, N-m+1$ 。每个时间序列 $z_j^{m,c}$ 被映射到色散模式 $\pi_{v_0 v_1 \dots v_{m-1}}$, 其中, $z_j^c = v_0$, $z_{j+1}^c = v_1, \dots, z_{j+m-1}^c = v_{m-1}$ 。模式的概率定义为:

$$p(\pi_{v_0 v_1 \dots v_{m-1}}) = \frac{\text{Number} \{j | j \leq N - m + 1, z_j^{m,c} \text{ has type } \pi_{v_0 v_1 \dots v_{m-1}}\}}{N - m + 1} \quad (17)$$

则 DE 定义为:

$$DE(m, c) = -\sum_{\pi=1}^{c^m} p(\pi_{v_0 v_1 \dots v_{m-1}}) \ln(p(\pi_{v_0 v_1 \dots v_{m-1}})) \quad (18)$$

其中, m 一般取为2, c 一般取为4~8。

1.2 HRV 特征子集选取

采用 Wilcoxon 符号秩检验方法检验两类 HRV 数据特征量之间差异性, 选取出具有统计显著差异的特征 ($P < 0.05$)。Wilcoxon 符号秩检验是一种非参数检验方法, 不需要对数据进行正态分布验证, 适用于小样本数据。

对选取出的特征进行 Min-Max 归一化处理, 特征 f 归一化后的数据 f_i' 为:

$$f_i' = \frac{f_i - \min(f)}{\max(f) - \min(f)} \quad (19)$$

其中, f_i 是 f 中第 i 个值, $\min(f)$ 是 f 中最小值, $\max(f)$ 是 f 中最大值。

采用 SFS 选择特征子集, SFS 是一种启发式搜索方法, 先把目标特征集定义为空集, 按照特征评估函数, 每次加入使评估函数更优的特征, 最后得到使评估函数最优的特征子集。本文将 SVM 分类器的 ACC 作为特征评估函数。具体步骤如下所述: (1) 将第 1.1~1.2 节选取出的 n 个特征分别输入 SVM 并在训练集内训练 n 个模型, 这些模型在测试集上得出最高 ACC, 记为 ACC0, ACC0 的模型对应特征设为 f_0 , 特征子集记为 $F_0 = \{f_0\}$; (2) 将剩余的 $n-1$ 个特征分别加入 F_0 内, 构成 $n-1$ 个新的特征子集, 将新的特征子集分别输入 SVM 并在训练集内训练 $n-1$ 模型, 这些模型在测试集上得出最高 ACC, 记为 ACC1, 如果 $\text{ACC1} > \text{ACC0}$, ACC1 的模型对应的新增特征记为 f_1 , 特征子集记为 $F_1 = \{f_0, f_1\}$; (3) 迭代重复步骤 (2), 直到剩余 $n-i$ 个特征分别加入特征子集, 构成 $n-i$ 个新的特征子集都不能提高 ACC, 即 $\text{ACC}(i+1) \leq \text{ACC}i$, 此时停止迭代。则最优 ACC 的特征子集即为 $F_i = \{f_0, \dots, f_i\}$ 。

1.3 特征分类

本文采用 LIBSVM 工具箱中的高斯核函数 (Radial Basis Function, RBF) 实现 SVM 分类, 因为它具有较少核函数参数, 较低训练复杂度和较好分类性能, 被广泛用于特征分类^[42]。另外, 数据采用 5 倍交叉验证对 SVM 模型训练和测试, 结果评价采用 ACC、敏感性 (Sensitivity, SN) 和特异性 (Specificity, SP) 3 个广泛使用的性能指标:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (20)$$

$$\text{SN} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (21)$$

$$SP = \frac{TN}{TN + FP} \tag{22}$$

式中, TP (True Positive) 代表真阳性; TN (True Negative) 代表真阴性; FP (False Positive) 代表假阳性; FN 代表 (False Negative) 假阴性。

2 结果

2.1 数据来源

本研究采用 PhysioNet 的 AFDDB 数据库^[43], 其中包含 50 个记录集标记为 p , 取自 PAF 患者。选取其中两种类型数据: 远离 PAF (记为 D-PAF, 来自 p 中的奇数记录前 5 min; 这些记录在结束后 45 min 或开始前 45 min 没有 PAF); 紧随 PAF (记为 F-PAF, 来自 p 中的偶数记录后 5 min, 之后 PAF 立即出现)。每种类型数据包含 25 个 ECG 记录。由 D-PAF 和 F-PAF 共 50 个 5 min 的 HRV 组成分类数据集^[44]。

2.2 特征抽取结果

2.2.1 IMMD 自适应滤波与粗粒化 R 结果 图 3 是 1 例典型 PAF 的 5 min HRV (p20); 图 4 是经 ASE 自适应滤波和粗粒化后得到的此例 HRV 分量集合 $R(n)=\{r_1(n), r_2(n), \cdots, r_{11}(n)\}$ 及其尺度值。这些尺度递减变化, 计算 $r_1 \sim r_{11}$ 的熵就可以得到这些尺度上 HRV 复杂性量化值。

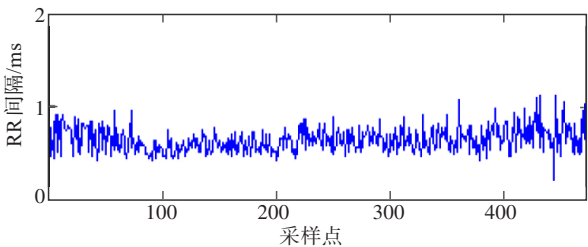


图3 1例典型 PAF 发作的 5 min HRV (p20)
Figure 3 Five-minute HRV of a typical PAF attack (p20)

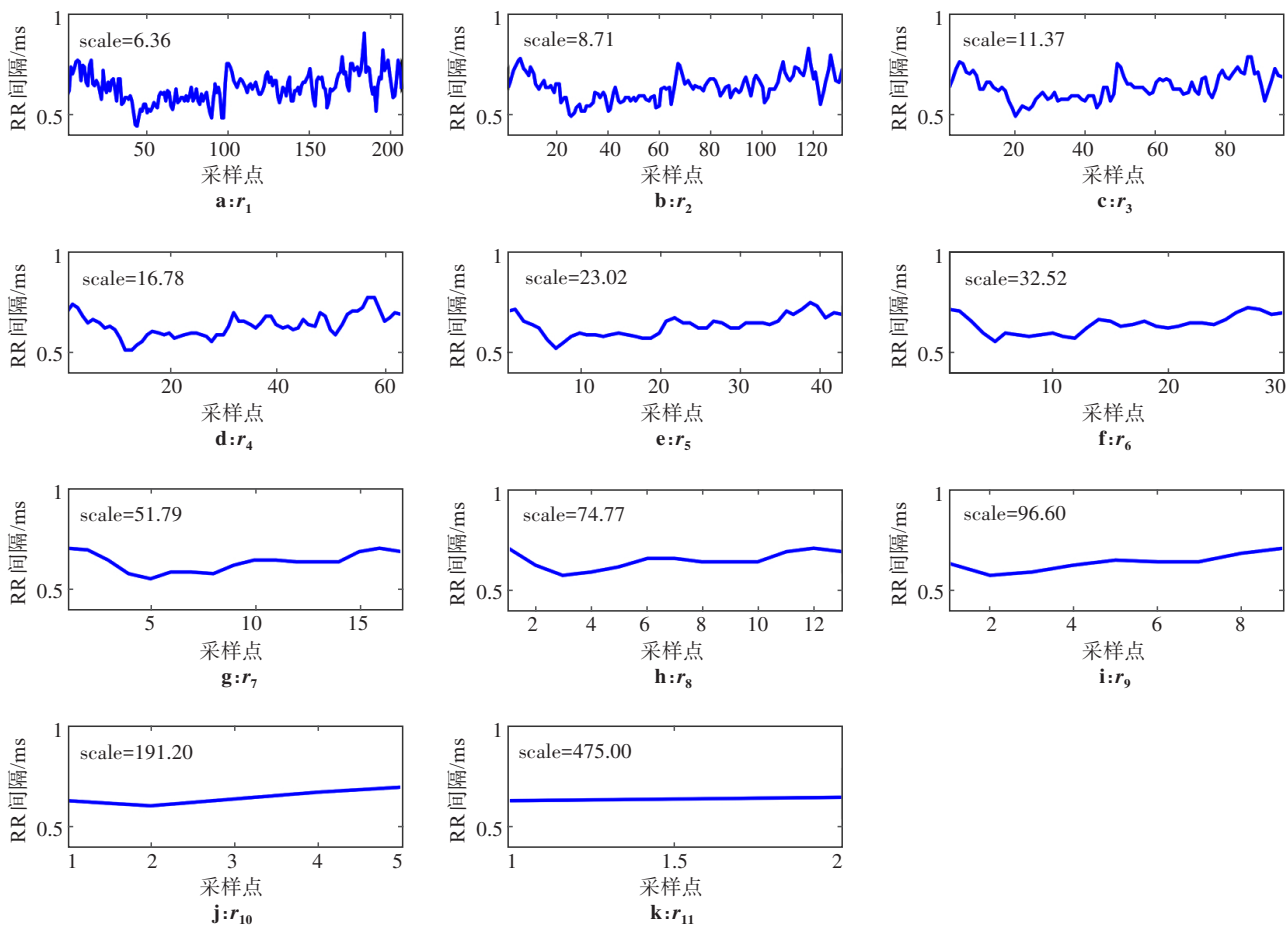


图4 经 ASE 自适应滤波和粗粒化后得到的 HRV (p20) 的分量及尺度
Figure 4 Components and scales of HRV (p20) obtained by adaptive filtering and coarse graining of ASE

2.2.2 抽取的 HRV 特征 所有样本数据经 IMMD 滤波和分量集合粗粒化后得到至少 10 个分量。计算这些分量的 AE、SE、PE、FE、DE 的平均值, 得到 HRV 复

杂性的 5 种不同特征集。两种不同类型 HRV 数据经 ASE 方法抽取特征的结果见图 5。在多个尺度上的 AE, D-PAF 值大于 F-PAF 值; 在所有尺度上的 SE, D-PAF

值大于 F-PAF 值;在多个尺度上的 PE, D-PAF 值小于 F-PAF 值;在所有尺度上的 FE, D-PAF 值小于 F-PAF 值。

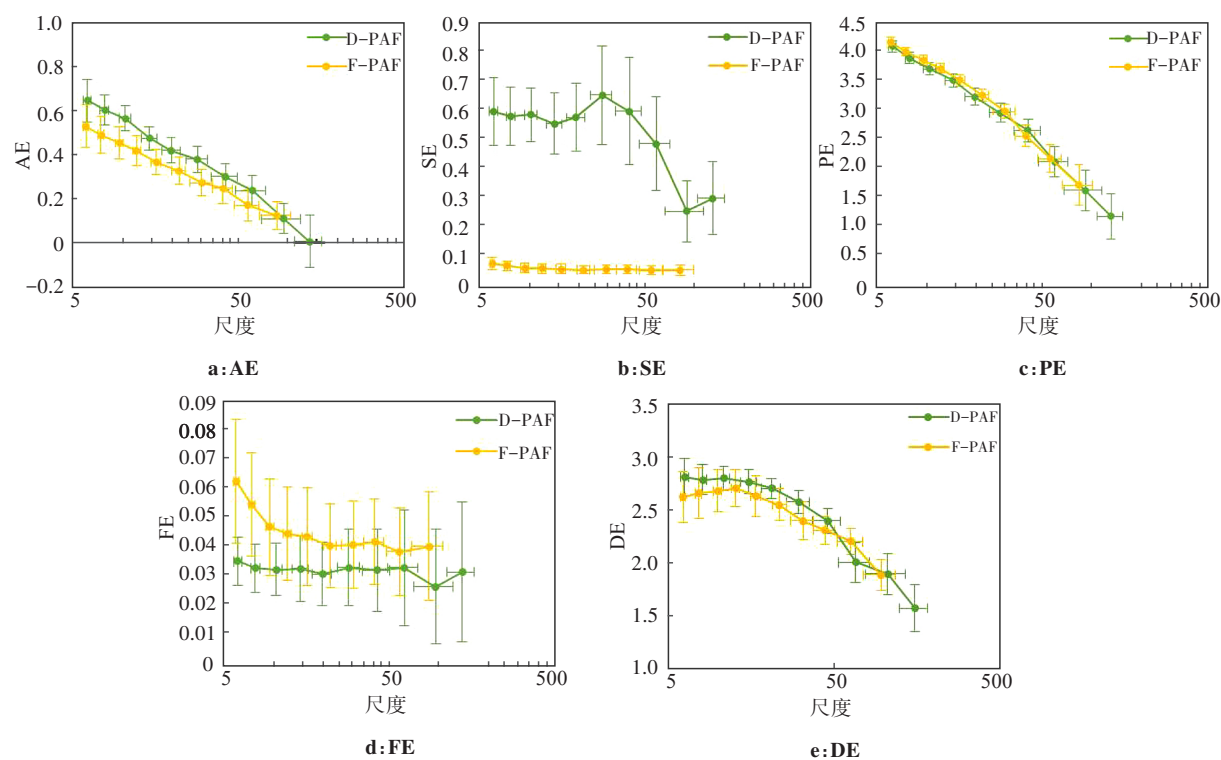


图5 两种不同类型HRV数据经ASE方法抽取特征的结果(均值±2×标准偏差/ \sqrt{n} , $n=25$)

Figure 5 Results of feature extraction from two different types of HRV data by ASE (mean±2×standard deviation/ \sqrt{n} , $n=25$)

2.3 特征选取结果

2.3.1 统计检验方法选取特征结果 Wilcoxon符号秩检验两种不同HRV类型数据的对应特征差异($P<0.05$), 结果见表1。为了比较结果,分别采用时域抽取的

SDNN、RMSSD、pNN50、SDSD, 频域抽取 ULF、VLF、LF、HF、TP、LF/HF^[45], 以及MSE方法抽取的20个HRV特征(一般地,滤波窗口 $\tau=1\sim20$)后,分别经Wilcoxon符号秩检验方法选取出的特征见表1。

表1 具有统计显著差异的特征集($P<0.05$)

Table 1 Feature sets with statistically significant differences ($P<0.05$)

线性方法		MSE	ASE				
时域特征	频域特征		AE	SE	PE	FE	DE
SDNN、RMSSD、pNN50、SDSD	LF、HF、TP	1~20	1~3	1~10	2、3、4~10	1、2	8~10

数字 $n(n=1\sim20)$ 代表第 n 个尺度上的特征量

2.3.2 SFS选取特征结果 本文将SVM分类器ACC作为特征评估函数,采用SFS选出特征,构成最优分类特征子集。表1中的特征经SFS得到最优特征子集见表2。

2.4 分类结果

将最优分类特征子集输入SVM进行分类,SVM分类模型采用RBF,以及五折交叉验证选择RBF的最佳参数。对应于表2中的最优特征子集,在 $[2^{-10}, 2^{10}]$

区域内选择出最优RBF核函数参数(γ, g)和惩罚函数参数(cost, c),见表3。

ACC、SN和SP性能指标评价方法结果见表4。ASE的第3尺度上的SE特征表现十分优异,ACC为98%,SN为100%,SP为96%,远高于表2中其他特征子集的分类性能;ASE的PE特征子集的ACC为78%,SN为76%,SP为80%,略优于时域、频域和MSE特征子集。

表 2 SFS 得到的最优特征子集
Table 2 Optimal feature subsets obtained by SFS

线性方法		MSE	ASE				
时域特征	频域特征		AE	SE	PE	FE	DE
SDNN, RMSSD, SDDSD	LF	3	3	3	3, 5, 8	1, 2	9

表 3 SVM 的最优参数
Table 3 Optimal parameters of SVM

线性方法		MSE	ASE				
时域特征	频域特征		AE	SE	PE	FE	DE
c=1.41	c=1.41	c=0.25	c=0.25	c=0.25	c=0.71	c=11.31	c=0.25
g=0.35	g=0.06	g=0.18	g=0.18	g=5.66	g=16	g=2	g=2.8

c 为惩罚参数, g 为核函数参数

表 4 SVM 分类性能(%)
Table 4 Classification performance of SVM (%)

指标	线性方法		MSE	ASE				
	时域特征	频域特征		AE	SE	PE	FE	DE
ACC	76	74	72	62	98	78	62	68
SN	84	56	52	48	100	76	40	84
SP	68	92	92	76	96	80	84	52

2.5 与其他方法比较

表 5 给出文献中相近 PAF 预测方法工作, 这些文献采用 HRV 的常用典型特征, 且这些工作与本文一样, 都是在 PhysioNet 的 AFDDB 数据库中进行。其中, 大部分文献选取同样的 5 min HRV 片段, 另外, 选取数据大于 5 min 的文献也包含本文所选 HRV 片段。文献[18]中 ACC、SN、SP 均可达到 100%, 但是该方法采用了 48 个特征, 实际应用中容易陷入“维度诅咒”, 并且方法的准确性仅通过数据的单折验证, 其泛化能力需要进一步评估。另外, 该方法采用 HRV 数据时长 30 min, 实时性能要比其他方法差。因此, 本文提出的 ASE(SE)方法系统分类最优。此外, 表 5 中绝大多数方法采用的特征数量至少为 4 个, 特别是性能表现越好的方法采用的特征数量越多。特征数量越多在实际应用中越容易陷入维度诅咒, 导致方法泛化能力差。大多数方法采用特征包括较多的时域和/或频域特征, 这些特征的医学含义难以解释^[46]。相比之下, 本文方法采用 1~3 个特征, 因此方法泛化能力好; 这些特征可以追溯到原始 HRV 序列的复杂性, 因此特征医学意义可以很容易跟踪研究。

3 讨论

生物系统的复杂性反映其在不断变化的环境中适应和发挥作用的能力; 生物系统的复杂性是多尺度的; 疾病状态和衰老都会降低个体的适应能力, 系统复杂性将降低甚至丧失^[27]。在提出的方法中, ASE(SE)表现出最优性能, 表明 ASE(SE)抽取 HRV 序列最典型特征, 本文仅对这些特征进行讨论。图 5b 中, 在所有时间尺度上的 SE 值, D-PAF 大于 F-PAF。这是因为受试者远离 PAF 时, 个体调节心脏节律能力要比紧随 PAF 时更强, 因此 D-PAF 复杂性要高于 F-PAF。F-PAF 紧随 PAF, HRV 值(RR 间期值)即将转为绝对不等(PAF 发作), 心脏动力系统即将从复杂转为混乱^[47-49]。

另一方面, 将自适应尺度换算成频率后, 得到两种不同类型 HRV 序列的频率-SE 关系曲线图, 如图 6 所示。其中, 黄色区域属于 VLF(0.015~0.040 Hz), 蓝色区域属于 LF(0.04~0.15 Hz), 绿色区域属于 HF(0.15~0.40 Hz)。大多数学者认为 HRV 的 LF 成分主要反映交感神经调节, HF 成分反映副交感神经调节^[50-51]。图 6 中, D-PAF 和 F-PAF 在所有对应特征之

表5 典型特征预测PAF方法结果

Table 5 Prediction results of different methods for PAF using typical features

文献	方法	特征(个数)	时长/min	交叉验证	ACC/%	SN/%	SP/%
文献[16]	遗传算法+U检验+SVM	NN50、pNN50、SE、SD2(庞加莱图)、HH-P1f、HH-P2g(6)	10	十折	69.8	58.5	81.1
文献[34]	遗传算法+KNN	SDNN、RMSSD、pNN20、小波总功率(4)	2.5~7.5	十折	80.0	72.0	88.0
文献[17]	U检验+SVM	时频域特征、双谱、庞加莱图特征、FFT特征、SE(7)	5	十折	87.7	86.8	88.7
文献[5]	后向选择+RF	AVNNc、NN20d、pNN50、SDNN、Z2ROI、rrTri(HRV指数)、TINNc、FFT_HF(8)	5	十折	86.0	85.6	86.4
文献[35]	ELM	时域特征、稀疏RR特征(16)	5	五折	97.2	100.0	94.4
文献[36]	mRMR、ILFS+SVM	庞加莱图特征(7)	5	十折	96.7	96.7	96.7
文献[18]	CNN	时频域特征、小波熵、庞加莱图特征(48)	30	单折	100.0	100.0	100.0

mRMR=最小冗余最大相关性;ILFS=无限潜在特征选择;AVNN = NN间隔平均值;pNN20=差值大于20 ms的连续HRV值的总数;TINN=NN间隔三角插值;HH-P1=归一化双谱熵;HH-P2=归一化双谱平方熵

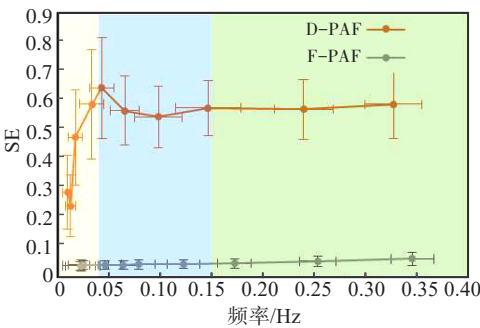


图6 两种不同类型HRV序列的频率-SE关系曲线图

Figure 6 Frequency-SE curves of two different types of HRV sequences

间的差异均有统计学意义(表1):这种差异在LF与HF上的不同,反映远离PAF和紧随PAF状态下,副交感神经和交感神经张力变化程度不同,导致二者平衡被破坏,这被认为是PAF发作的主要机制^[7];另外,相关实验表明VLF反映了传出交感神经调控机体(包括心脏)活动、应激等适应外界环境变化能力,是心脏事件的有力预测因子之一^[8]。因此,F-PAF比D-PAF在VLF上SE值的显著降低,反映这种适应能力的降低。

4 结 论

本文提出一种ASE方法抽取HRV复杂性特征预测PAF方法,经特征抽取、选取后,采用公开数据集的五折交叉验证可知,采用ASE的第3尺度的SE特征输入SVM进行HRV识别,预测PAF即将发作的性能优异。此外,提出的ASE方法建立在自适应滤波技术上,可以得到HRV在多个自主频率上的复杂性特征,反映受试者自主神经系统调节心脏节律能力。这些特征有助于识别不同状态下的PAF的HRV序列

之间的关键差异,从而帮助心脏病专家探索研究PAF患者发病机制。然而,提出的方法受到数据样本量较小的限制,因此本文提出的方法需要更多数据测试。

【参考文献】

[1] Wang ZW, Chen Z, Wang X, et al. The disease burden of atrial fibrillation in China from a national cross-sectional survey[J]. Am J Cardiol, 2018, 122(5): 793-798.

[2] Healey JS, Oldgren J, Ezekowitz M, et al. Occurrence of death and stroke in patients in 47 countries 1 year after presenting with atrial fibrillation: a cohort study[J]. Lancet, 2016, 388(10050): 1161-1169.

[3] Gladstone DJ, Spring M, Dorian P, et al. Atrial fibrillation in patients with cryptogenic stroke[J]. N Engl J Med, 2014, 370(26): 2467-2477.

[4] de Chazal P, Heneghan C. Automated assessment of atrial fibrillation [C]//Computers in Cardiology 2001. Vol.28 (Cat. No.01CH37287). Piscataway, NJ, USA: IEEE, 2001: 117-120.

[5] Castro H, Garcia-Racines JD, Bernal-Norena A. Methodology for the prediction of paroxysmal atrial fibrillation based on heart rate variability feature analysis[J]. Heliyon, 2021, 7(11): e08244.

[6] Seely AJ, Newman KD, Herry CL. Fractal structure and entropy production within the central nervous system[J]. Entropy (Basel), 2014, 16(8): 4497-4520.

[7] Zhong X, Jiao HC, Gao JC, et al. Autonomic tone variations prior to onset of paroxysmal atrial fibrillation[J]. Med Sci Monit, 2022, 28: e934028.

[8] Hadase M, Azuma A, Zen K, et al. Very low frequency power of heart rate variability is a powerful predictor of clinical prognosis in patients with congestive heart failure[J]. Circ J, 2004, 68(4): 343-347.

[9] Mohebbi M, Ghassemian H. Prediction of paroxysmal atrial fibrillation based on non-linear analysis and spectrum and bispectrum features of the heart rate variability signal[J]. Comput Methods Programs Biomed, 2012, 105(1): 40-49.

[10] McCraty R, Shaffer F. Heart rate variability: new perspectives on physiological mechanisms, assessment of self-regulatory capacity, and health risk[J]. Glob Adv Health Inform, 2015, 4(1): 46-61.

[11] Sessa F, Anna V, Messina G, et al. Heart rate variability as predictive factor for sudden cardiac death[J]. Aging (Albany NY), 2018, 10(2): 166-177.

[12] Li YF, Pan WF, Li KY, et al. Sliding trend fuzzy approximate entropy as a novel descriptor of heart rate variability in obstructive sleep apnea [J]. IEEE J Biomed Health Inform, 2019, 23(1): 175-183.

[13] Gupta V, Mittal M, Mittal V. Chaos theory and ARTFA: emerging tools for interpreting ECG signals to diagnose cardiac arrhythmias[J]. Wirel Pers Commun, 2021, 118(4): 3615-3646.

- [14] Henriques T, Ribeiro M, Teixeira A, et al. Nonlinear methods most applied to heart-rate time series: a review[J]. Entropy (Basel), 2020, 22(3): 309.
- [15] de Godoy MF. Nonlinear analysis of heart rate variability: a comprehensive review[J]. J Cardiol Ther, 2016, 3(3): 528-533.
- [16] Boon KH, Khalil-Hani M, Malarvili MB, et al. Paroxysmal atrial fibrillation prediction method with shorter HRV sequences[J]. Comput Methods Programs Biomed, 2016, 134: 187-196.
- [17] Boon KH, Khalil-Hani M, Malarvili MB. Paroxysmal atrial fibrillation prediction based on HRV analysis and non-dominated sorting genetic algorithm III[J]. Comput Methods Programs Biomed, 2018, 153: 171-184.
- [18] Surucu M, Isler Y, Perc M, et al. Convolutional neural networks predict the onset of paroxysmal atrial fibrillation: theory and applications[J]. Chaos, 2021, 31(11): 113119.
- [19] Seely AJ. Optimizing our patients' entropy production as therapy? Hypotheses originating from the physics of physiology[J]. Entropy (Basel), 2020, 22(10): 1095.
- [20] Pincus SM. Approximate entropy as a measure of system complexity [J]. Proc Natl Acad Sci U S A, 1991, 88(6): 2297-2301.
- [21] Richman JS, Moorman JR. Physiological time-series analysis using approximate entropy and sample entropy[J]. Am J Physiol Heart Circ Physiol, 2000, 278(6): H2039-H2049.
- [22] Pan WY, Su MC, Wu HT, et al. Multiscale entropy analysis of heart rate variability for assessing the severity of sleep disordered breathing [J]. Entropy (Basel), 2015, 17(1): 231-243.
- [23] Shi MH, He HX, Geng WC, et al. Early detection of sudden cardiac death by using ensemble empirical mode decomposition-based entropy and classical linear features from heart rate variability signals[J]. Front Physiol, 2020, 11: 118.
- [24] Deschodt-Arsac V, Blons E, Gilriche P, et al. Entropy in heart rate dynamics reflects how HRV-Biofeedback training improves neurovisceral complexity during stress-cognition interactions [J]. Entropy (Basel), 2020, 22(3): 317.
- [25] Selek MB, Yesilkaya B, Egeli SS, et al. The effect of principal component analysis in the diagnosis of congestive heart failure *via* heart rate variability analysis[J]. Proc Inst Mech Eng H, 2021, 235 (12): 1479-1488.
- [26] Costa M, Goldberger AL, Peng CK. Multiscale entropy analysis of complex physiologic time series[J]. Phys Rev Lett, 2002, 89(6): 068102.
- [27] Costa M, Goldberger AL, Peng CK. Multiscale entropy analysis of biological signals[J]. Phys Rev E Stat Nonlin Soft Matter Phys, 2005, 71(2 Pt 1): 021906.
- [28] Hu M, Liang HL. Adaptive multiscale entropy analysis of multivariate neural data[J]. IEEE Trans Biomed Eng, 2012, 59(1): 12-15.
- [29] Zhao DF, Liu SL, Cheng SG, et al. Parallel multi-scale entropy and its application in rolling bearing fault diagnosis[J]. Measurement, 2021, 168: 108333.
- [30] Huang NE, Shen Z, Long SR, et al. The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis[J]. Proc R Soc Lond A, 1998, 454(1971): 903-995.
- [31] 牛晓东, 卢莉蓉, 王鉴, 等. 基于改进经验模态分解域内心动物理特征识别模式分量的心电信号重建[J]. 物理学报, 2021, 70(3): 038702. Niu XD, Lu LR, Wang J, et al. Electrocardiogram signal reconstruction based on mode component identification by heartbeat physical feature in improved empirical mode decomposition domain[J]. Acta Physica Sinica, 2021, 70(3): 038702.
- [32] Niu XD, Lu LR, Wang J, et al. An improved empirical mode decomposition based on local integral mean and its application in signal processing[J]. Math Probl Eng, 2021, 2021: 8891217.
- [33] Hickey B, Heneghan C. Screening for paroxysmal atrial fibrillation using atrial premature contractions and spectral measures [C]// Computers in Cardiology. Piscataway, NJ, USA: IEEE, 2002: 217-220.
- [34] Narin A, Isler Y, Ozer M, et al. Early prediction of paroxysmal atrial fibrillation based on short-term heart rate variability[J]. Phys A, 2018, 509: 56-65.
- [35] Maghawry E, Ismail R, Gharib TF. An efficient approach for paroxysmal atrial fibrillation events prediction using extreme learning machine[J]. J Intell Fuzzy Syst, 2021, 40(3): 5087-5099.
- [36] Parsi A, Glavin M, Jones E, et al. Prediction of paroxysmal atrial fibrillation using new heart rate variability features[J]. Comput Biol Med, 2021, 133: 104367.
- [37] Whitney AW. A direct method of nonparametric measurement selection [J]. IEEE Trans Comput, 1971, C-20(9): 1100-1103.
- [38] Bandt C, Pompe B. Permutation entropy: a natural complexity measure for time series[J]. Phys Rev Lett, 2002, 88(17): 174102.
- [39] Chen WT, Wang ZZ, Xie HB, et al. Characterization of surface EMG signal based on fuzzy entropy[J]. IEEE Trans Neural Syst Rehabil Eng, 2007, 15(2): 266-272.
- [40] Rostaghi M, Azami H. Dispersion entropy: a measure for time-series analysis[J]. IEEE Signal Process Lett, 2016, 23(5): 610-614.
- [41] Cao YH, Tung WW, Gao JB, et al. Detecting dynamical changes in time series using the permutation entropy[J]. Phys Rev E Stat Nonlin Soft Matter Phys, 2004, 70(4 Pt 2): 046217.
- [42] Chang CC, Lin CJ. LIBSVM: a library for support vector machines[J]. ACM Trans Intell Syst Technol, 2011, 2(3): 27.
- [43] Goldberger AL, Amaral LA, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals[J]. Circulation, 2000, 101(23): E215-E220.
- [44] Moody GB, Goldberger AL, McClennen S, et al. Predicting the onset of paroxysmal atrial fibrillation: the computers in cardiology challenge 2001 [C]//Computers in Cardiology 2001. Vol. 28 (Cat. No.01CH37287). Piscataway, NJ, USA: IEEE, 2001: 113-116.
- [45] Camm AJ. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology[J]. Circulation, 1996, 93(5): 1043-1065.
- [46] Mendez MM, Hsu MC, Yuan JT, et al. A heart rate variability-based paroxysmal atrial fibrillation prediction system[J]. Appl Sci, 2022, 12 (5): 2387.
- [47] 樊永强, 汤宝鹏, 芦颜美. 深度学习智能心电图识别方法的临床应用进展[J]. 临床心电学杂志, 2022, 31(5): 372-376. Fan YQ, Tang BP, Lu YM. Clinical application progress of deep learning intelligent in electrocardiography recognition[J]. Journal of Clinical Electrocardiology, 2022, 31(5): 372-376.
- [48] 牛向东. 心电散点图诊断持续性心房颤动伴II度房室传导阻滞的形态特征分析[J]. 中西医结合心脑血管病杂志, 2019, 17(24): 4015-4021. Niu XD. Analysis of ECG scatter plot in the diagnosis of persistent atrial fibrillation with second degree atrioventricular block[J]. Chinese Journal of Integrative Medicine on Cardio-Cerebrovascular Disease, 2019, 17(24): 4015-4021.
- [49] 黄从新, 张澍, 黄德嘉, 等. 心房颤动: 目前的认识和治疗的建议-2018 [J]. 中国心脏起搏与心电生理杂志, 2018, 32(4): 315-368. Huang CX, Zhang S, Huang DJ, et al. Atrial fibrillation: current understanding and treatment suggestions-2018[J]. Chinese Journal of Cardiac Pacing and Electrophysiology, 2018, 32(4): 315-368.
- [50] Malliani A, Lombardi F, Pagani M. Power spectrum analysis of heart rate variability: a tool to explore neural regulatory mechanisms[J]. Br Heart J, 1994, 71(1): 1-2.
- [51] Lahiri MK, Kannankeril PJ, Goldberger JJ. Assessment of autonomic function in cardiovascular disease: physiological basis and prognostic implications[J]. J Am Coll Cardiol, 2008, 51(18): 1725-1733.

(编辑:陈丽霞)