

## 纸质心电图的数字化及特征提取

常博文

(武汉邮电科学研究院,湖北 武汉 430074)

**摘要:** 心电图(ECG)在心脏相关疾病研究方面具有重要的参考价值。但是到目前为止,纸质心电图依然是最普遍的心电记录方式。因此,该文针对纸质心电图难以长期存储的问题,采用灰度阈值、边界扫描等方法将纸质心电图转化成数字的电压信号。另外,为了解决心电信号难以自动提取特征的问题,同时检验心电图数字化的效果,针对心电信号的不同波形采用窗口扫描、DBSCAN等不同方法提取出波形的幅值特征。通过与同一医师读取的幅值结果进行相关性分析,发现幅值的相关性约在0.98,时间相关参数相关性也在0.89以上。证明该方法得到的数字化信号可以很好地还原纸质ECG信息。

**关键词:** 数字化; 心电图; 特征提取; DBScan

**中图分类号:** TP391

**文献标识码:** A

**文章编号:** 1674-6236(2020)22-0083-06

**DOI:** 10.14022/j.issn1674-6236.2020.22.019

### Digitization and feature extraction of picture ECG images

CHANG Bowen

(Wuhan Research Institute of Posts and Telecommunications, Wuhan 430074, China)

**Abstract:** Electrocardiogram (ECG) has important value in the research of heart related diseases. But so far, paper recording is still the most common ECG recording method. The paper ECG is difficult to be stored for a long time. Therefore, in this paper, the paper ECG is converted into a digital voltage signal by using gray-scale thresholds and boundary scan. In addition, in order to solve the problem that it is difficult to automatically extract features of the ECG signal, and at the same time verify the effect of digitized ECG, different methods of window scan, DBSCAN and other methods are used to extract the amplitude characteristics of the waveform for different waveforms of the ECG signal. Correlation analysis was performed with the amplitude results read by the same physician, and it was found that the correlation of the amplitude was about 0.98, and the correlation of the time-related parameters was also above 0.89. It is proved that the digitized signal obtained by this method can well restore the paper ECG information.

**Keywords:** digitizing; ECG; feature extraction; DBScan

心电图(ECG)是利用心电图机从体表记录心脏每一心动周期产生的电活动变化的可见图形记录<sup>[1]</sup>。纸质心电图已存在超过一个世纪,目前电子心电图并没有被广泛的应用,纸质心电图依然是普及的心电记录方式。随着当代大数据业务的不断扩展,不论是纸质文本或者图像都最终会被电子化。数字心电图提供了快速分析心电图属性的方法,例如通过

观察矢量角度和波形间隔的方法,可以利用心电图进行疾病的关联研究和疾病预测,通过回顾性评估,学习QRS波形特征或者Tangle等特征可以找到病患的一些常见机理,快速将心电图特征与患者结果相关联,进而达到疾病预测的目的。纸质心电图记录着大量病人的心电数据,如果可以将这些已有的纸质心电图进行电子化,将会对我们的心脏疾病研究带来深远的影响。如何从纸质心电图准确地提取出

**收稿日期:** 2019-12-25 **稿件编号:** 201912216

**作者简介:** 常博文(1992—),男,河南郑州人,硕士研究生。研究方向:人工智能、机器学习。

所需的特征信号,是基于机器学习的纸质心电图分析模型的基础,所以研究纸质心电图电子化过程显得尤为重要。

该文提出了一种基于 Matlab 的纸质心电图数据数字化工具。该方法将纸质 ECG 扫描到图像,阈值去除图形网格并检测 ECG 信号,然后进行数字化。在数字化的心电信号的基础上运用聚类、窗口扫描等方法确定心电信号上关键点的位置。该工具可以快速数字化大量纸质 ECG 并提取关键特征点的读数。

## 1 总体流程

临床上常见的心电图是同步记录 12 导联的信息。为了便于常见的心律失常的分析,下部一般为一个长时间段的 II 导联记录。通过分析提供的心电图样本发现,需要提取的心电图波形为黑色,与橙红色坐标纸背景存在较大差异,因此考虑通过二进制阈值处理操作删除图形网格并将灰度图像转换为二进制图像。心电图数字化过程和特征提取主要基于 II 导联来研究,其他导联的信号可以用同样的方法来提取。II 导联图像识别主要流程如图 1 所示。

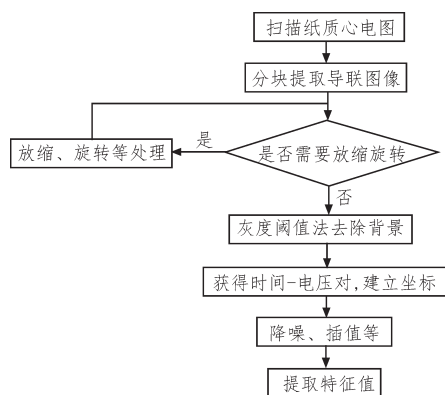


图1 II导联图像识别总流程框图

### 1.1 心电信号的数字化

首先,通过扫描仪将纸质的 ECG 扫描为电子图像,如图 2 所示,即为典型的纸质心电图的扫描图像。然后把扫描得到的彩色图像转换为 8 位灰度,接下来将图像分辨率缩小到 300 dpi。另外,需要检测图像是否需要歪斜校正,即校正 ECG 纸的角度取向。方法是通过在同一条网格线上选取两个点并使用获得的角度旋转扫描图像来计算网格线的斜率。

导入图像之后首先对图像进行切割,可以看到纸质心电图上每一个导联开始前的标准门波旁边有这个导联的罗马数字标志,将这个标志剪裁出来作

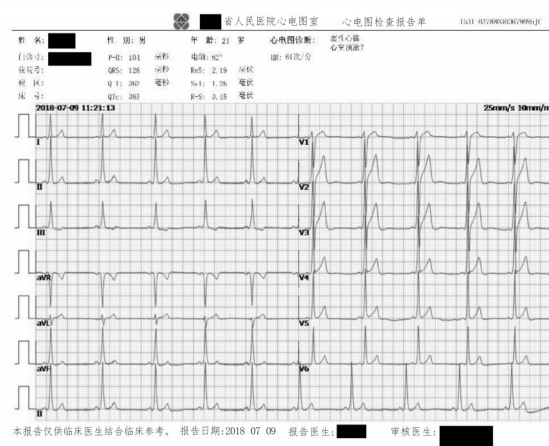


图2 典型的纸质心电图扫描图像

为模板,使用目标识别来定位 II 导联的坐标原点,然后根据固定的像素值作为长宽来剪裁,最终得到 II 导联心电图图像。图 3 是某患者的心电图报告切割后的示例。



图3 切割得到II导联心电图图像

下一步是阈值处理操作,以删除图形网格并将灰度图像转换为二进制图像<sup>[2]</sup>。首先得到 ECG 的灰度直方图,如图 4 所示。

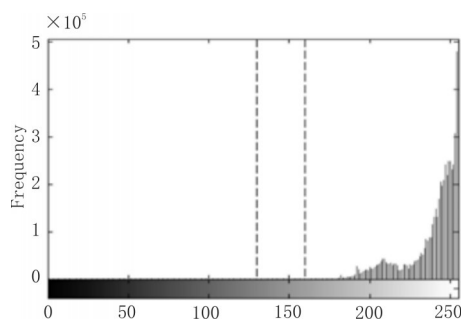


图4 心电图的灰度直方图

由于图形网格的像素强度值通常大于打印的 ECG 信号的像素强度值<sup>[3]</sup>,因此使用图像的直方图确定阈值,将大于阈值的图像像素点过滤即为所求心电图 II 导联波形曲线。从图 4 中可以看出,所需分离的信号分量(较低的像素强度<sup>[4]</sup>)与在 130~160 范围内观察到图形网格(较高的像素强度)有较大区别。可以确定在此示例中,范围 130~160 仅由网格线贡献。根据经验,可以确定该区域的阈值将确保从 ECG 图像中有效删除图形网格。

在此二进制图像上,执行基于列的像素扫描,以

根据零位置(黑色)找到ECG波形的轮廓。由于所打印的ECG信号轨迹具有一定的宽度,因此对于每一列都获得了对应于上下轮廓的一对行索引,对同一时刻的电压坐标做均值处理,根据导联前标准门波的高和宽可以确定像素点坐标和横坐标(时间)以及纵坐标(电压)的关系。从而可以获得ECG波形轮廓的均值的时间-电压对。另外从图像的分辨率中可以得出此数字信号的采样率。至此,纸质心电图的电子化工作已基本完成。

然而,二进制阈值处理可能在背景中引入“盐和胡椒”噪声,这通常是扫描纸质ECG过程引入的。此外,还可能由于去除主网格线而导致的ECG信号的不连续或间隙问题。在阈值处理之后执行滤波和插值可以消除这些问题。

## 1.2 心电信号滤波

通过图像识别提取的心电图曲线存在两个方向上的干扰:1)图像像素点不清晰,造成连接而成的心电曲线存在大量的“毛刺”;2)心电信号具有微弱、低频、高阻抗等特性,极易受到干扰<sup>[5]</sup>。所以分析噪声干扰的来源,针对不同干扰采取相应的滤除措施,使得心电曲线变得光滑、易于分析,是心电图特征参数提取所需要重点考虑的一个问题。

### 1) 滤除高频噪声干扰

心电信号属于肌电信号的一种。通常来说,肌肉产生的电信号的频率在20~5 000 Hz区间,不同类型的肌肉产生的肌电信号频率不同,一般主要分布在30~300 Hz,而心电信号频率主要集中在5~20 Hz的低频区域,而心电测量中常见的工频干扰(由交流电产生)在50 Hz,而测量心电时心电测量仪本身和周围其他仪器产生的干扰主要在1 000 Hz以上。巴特沃斯低通滤波器的特点是通带内的频率响应曲线平坦,而阻带内则逐渐下降为零。巴特沃斯滤波器的振幅对角频率单调下降,并且滤波器的阶数越高,在阻带内幅度衰减速度越快,其他滤波器高阶的振幅对角频率图和低阶数的振幅对角频率有不同的形状。

从图5来看,巴特沃斯滤波器的衰减速度相对其他类型滤波器来说较为缓慢,且十分平坦,没有幅度变化<sup>[6]</sup>。

### 2) 小波变换滤波法

在使用低通滤波器对高频杂波进行滤除之后,通过放大仍然可以看到各种小波之间仍然存在许多

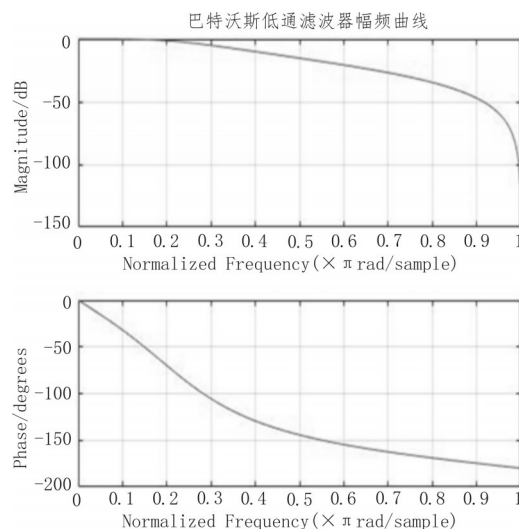


图5 巴特沃斯低通滤波器幅频曲线

“毛刺”,这对定位极值点,比较低小的Q波、S波的起始点,定位发生病变之后变得较为低小的P波、T波都存在一定的干扰。因此,该文又采用常用的去除心电信号高频干扰噪声的小波变换法来进一步滤除心电信号波形中的高频噪声干扰。首先对信号进行小波变换分解,采用合适的阈值对高频小波系数进行处理,来达到去除高频噪声的目的<sup>[7]</sup>。然后,对处理后的小波分量进行重构得到去除噪声后的信号<sup>[8]</sup>。

## 1.3 波形特征点定位

通常的Ⅱ导联心电图波形的组成包括P波、Q波、R波、S波、T波以及P-R间期、S-T间期、Q-T间期几个波段<sup>[9]</sup>。图6为一个标准的心拍周期,各波段形状及含义在图中标出说明。

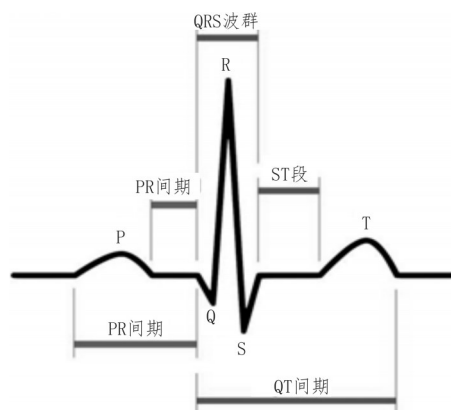


图6 标准心电图的各波段形状及含义

以T波终点到P波起点之间心电信号的平均值,作为基线位置<sup>[10]</sup>。心电临床诊断中对疾病诊断有参考意义的参数,如:P-P间期(两个相邻P波波峰之间



的时间间隔)、P波幅值、R-R间期(两个相邻R波波峰之间的间隔)、Q波幅值、S波幅值、R波幅值等参数可以自动化定位,提取坐标<sup>[11]</sup>,为分析海量纸质心电图提供数据基础。

### 1) 定位R波波峰

R波波峰是QRS波段中的最大的正向电压,根据图7可以看出R波的特点是整个心拍周期的最大值也是极大值。所以使用差分法先找出心电图曲线中的极值点,然后在所有的极大值点中找到在一个心拍周期中的极大值点即为R波的波峰<sup>[12]</sup>。由于提取得到的心电信号是一系列点的集合,是离散信号,所以考虑使用差分的方法来求极值点。各波峰值和峰谷点对应的是一阶差分的零点,二阶差分的极小值/极大值点<sup>[13]</sup>,如图7所示。

一阶差分:  $d(n) = x(n+1) - x(n)$

二阶差分:  $t(n) = d(n+1) - d(n)$

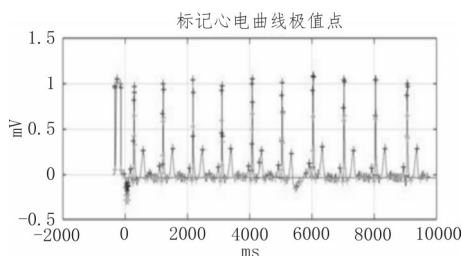


图7 心电图曲线中的极值点

接下来需要对得到的极值点进行最大值的定位,对所得到的极值点进行DBSCAN聚类定位可以确定R峰的位置。考虑到心电图R波的幅值最高,一般振幅超过0.4 mV,因此选择纵坐标在0.4 mV上的极值点进行DBSCAN聚类。DBSCAN是一个基于密度的聚类算法,由密度可达关系导出的最大密度相连的样本集合<sup>[14]</sup>。在将像素点聚成多个簇之后,分别提取出每个簇中y值最大的像素点作为R波峰值点,并在心电图曲线中进行标记。需要指出的是,在对原始信号进行巴特沃斯滤波之后,各个R波的波峰和T波的波峰会有所降低,这是低通滤波器的特性所决定的。所以为了保证R波波峰读数的准确,需要在提取到的原始电压-时间数据集上进行上述步骤。

### 2) Q波与S波波谷定位

窗口扫描法是一种基于心电图的时域特性,在限定范围的窗口内不断搜索满足条件的点,并据此确定波形特征点的方法。典型的QRS波群包括3个

紧密相连的波,在对R波进行定位之后,以R波位置为基准,根据心电图最大间期特征,分别向前和向后以一定步长进行搜索,即可得到Q波与S波的波谷。

Q波波谷:在R峰处向左40~110 ms内搜索纵坐标绝对值最大的极值点<sup>[15]</sup>。

S波波谷:在R峰处向右40~110 ms内搜索纵坐标绝对值最大的极值点。

### 3) 定位P波与T波波峰/谷

根据对心率的研究可知,如果要通过时域上的搜索来辅助确定特征波段的位置,必须要考虑心率对其他波段和期间的伸缩影响,其中心率可以通过之前求出的R峰值来得出。表1给出了正常心电图各特征波段的时间长度对照。

表1 正常心电图各特征波段时间长度对照表

心电图波段	标准值/s	与一个心动周期的比值 $\rho$
T波	0.05~0.25	0.2
P波	0.08~0.12	0.1
R波	0.01~0.05	0.06
Q波	0.01~0.04	0.05
P-P间期	0.6~1	1
R-R间期	0.6~1	1
P-R间期	0.12~0.2	0.2
Q-T间期	0.32~0.44	0.5
QRS波群	0.06~0.11	0.1

对照表中的比值,按照公式就可以推算出不同心率下各波段所需要的窗口扫描位置:

$$|R_{\text{time}} - M_{\text{time}}| = \sum \rho_i \times \text{Mean}(RR) \quad (1)$$

其中,  $\text{Mean}(RR)$  是上一次测试周期的心电R峰到R峰时间间隔差的1分钟平均值;  $R_{\text{time}}$  和  $M_{\text{time}}$  分别为R波峰值的时间位置和待提取的某个波段的时间位置,对 $\rho$ 的求和方式取决于从R波峰到该处要经过哪几个区间,利用窗口扫描法并结合心电图特性来定位P波和T波<sup>[16]</sup>。

P波搜索范围如式(2)所示,其中  $R_{\text{peak}}(\text{time})$  是R波顶点的时间,  $\text{Mean}(RR)$  是R-R期间的平均值:

$$\begin{aligned} & [R_{\text{peak}}(t) - 0.21\text{Mean}(RR), \\ & R_{\text{peak}}(t) - 0.06\text{Mean}(RR)] \end{aligned} \quad (2)$$

T波搜索范围如式(3)所示。

$$\begin{aligned} & [R_{\text{peak}}(t) + 0.15\text{Mean}(RR), \\ & R_{\text{peak}}(t) + 0.51\text{Mean}(RR)] \end{aligned} \quad (3)$$

## 2 实验结果分析

### 2.1 实验结果

在心电图数字化过程中具有代表性的中间步骤结果将在下面详细给出。

图8是由心电图切割出的II导联图像(图3)去除背景并且二值化之后得到的图像(0表示黑色,255表示白色)。图9是使用低通滤波器滤除高频杂波干扰后的心电曲线与原始曲线的对比图。

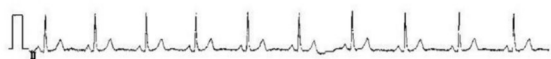


图8 去除背景并且二值化后的图像

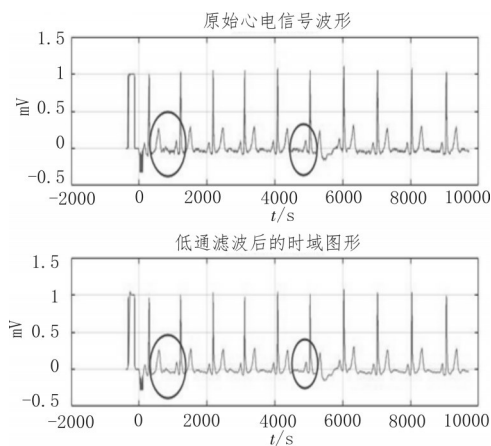


图9 低通滤波器滤除高频噪声效果图

从图9曲线可以看出,原始心电信号波形上存在许多“毛刺”,即存在多个高频波形噪声,部分波形受到的干扰比较严重,在采用巴特沃斯低通滤波器滤波之后,滤除了大量的“毛刺”,图10是经过小波变换滤波后的心电信号对比图。

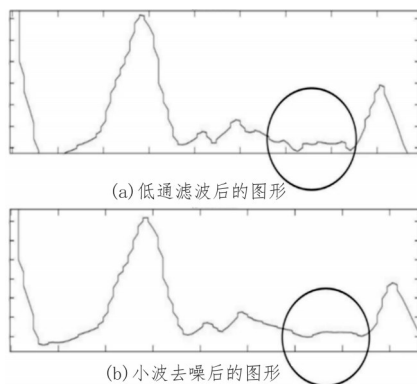


图10 小波变换滤波效果对比图

从圆圈标注的波形中可以看出,经过巴特沃斯

低通滤波器滤除高频噪声干扰的波形仍然存在许多小毛刺,且许多经过零点,这对定位波形起始点造成相当大的干扰。在经过小波变换去噪之后,波形则变得较为平坦。

以上是图像的直观表现,为了进一步对信号波形滤波效果进行说明,对信号求取其频谱图可知,小波变换滤波方法滤除了许多高频低幅杂波,能有效滤除心电信号中的高频噪声。

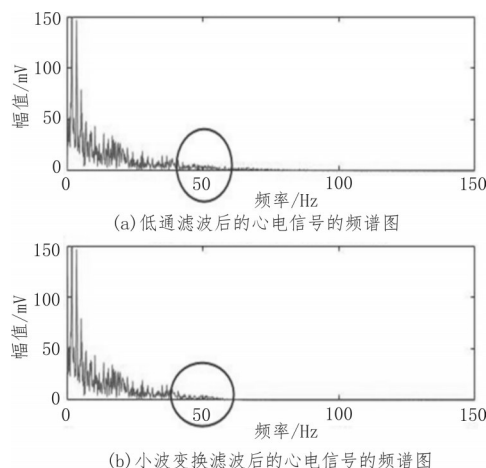


图11 低通滤波与小波变换处理后的信号频谱图

接下来是特征值的提取,首先是利用DBSCAN定位R波波峰。从图12中来看,在不存在高大P波和T波时,聚类效果非常好。

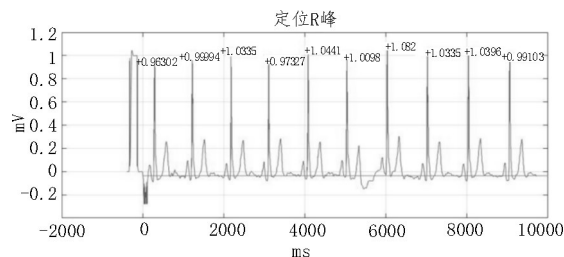


图12 DBSCAN聚类定位R峰

接下来定位Q波与S波的波谷,效果如图13所示。途中Q波波谷与S波波谷分别用星号与圆圈标记。

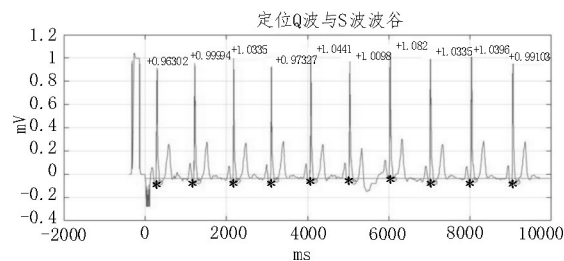


图13 Q波与S波波谷定位

最后定位P波与T波的波峰,得到P波与T波效果如图14所示,P波与T波的波峰分别用黑色三角与方块标记。

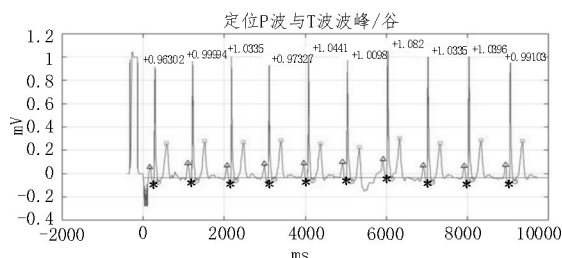


图14 P波与T波的定位效果

## 2.2 结果验证

为了确定心电图图像识别的精度,将100张心电图电子化以后提取出的5个特征值与心电图报告中的对应参数指标相对比算出绝对误差和相对误差。并且将电子化的心电图与同一医师读到的对应特征值比较计算组内相关性系数,所得结果如表2所示,其中Related代表组内相关性系数。

表2 5个心电图特征值的组内组间相关性

参数	最大绝对误差	最大相对误差	Related	p值
R波幅值	0.03 mV	0.08%	0.987	<0.001
R-R间期	200 ms	0.7%	0.934	<0.001
QRS间期	300 ms	1.9%	0.895	<0.001
T波幅值	0.01 mV	0.05%	0.979	<0.001

经过对比,发现组内相关性基本处于0.85~1之间,验证了电子化的心电信号基本保存了纸质心电信号的主要特征。心电图的特征值门诊结果与电子心电图读取的特征值的幅值误差最大不超过0.03 mV,时间维度,误差最大不超过27 ms,电压和时间的相对误差分别不超过 $\pm 1\%$ 和 $\pm 2\%$ ,证明特征提取的准确性。

## 3 结论

针对纸质心电图难存储、难分析、不利于大数据处理平台处理的问题,提出一种基于Matlab的工具,用于将纸质心电图电子化,并且自动提取出波形的部分特征值。经过实验发现,原始纸质图像和重建信号中提取的临床重要特征的测量值之间的差异是微不足道的,因此证明了该方法的准确性和有效性。

使用这种可以将纸质ECG数字化的工具,将允许更方便研究新出现的ECG特征。另外,通过批量

的数字化纸质心电图扩大心电图数据库,可以为基于机器学习的智能心电分析提供有效的数据保障。

### 参考文献:

- [1] 邵小艳. 碎裂QRS波在冠心病心绞痛患者中的价值分析[D]. 泸州: 西南医科大学, 2016.
- [2] 高莹. 基于图割和水平集的肾磁共振图像分割方法研究[D]. 大连: 大连海事大学, 2014.
- [3] 雷玉堂. 安防视频监控图像处理的技术与方法[J]. 中国公共安全, 2014(8):138-145.
- [4] 刘歌, 汪洪艳, 张国毅. 基于时频图像处理的多分量信号分离[J]. 电子信息对抗技术, 2017, 32(2):13-18, 57.
- [5] 王润. 心电信号的预处理算法分析[J]. 现代计算机(专业版), 2018(7):33-36.
- [6] 庄晨. 圆度仪中数字滤波器的研究与应用[D]. 西安: 西安电子科技大学, 2013.
- [7] 崔少华, 方振国, 王江涛, 等. 基于小波变换的地震数据去噪的研究[J]. 曲阜师范大学学报(自然科学版), 2018, 44(3):54-58.
- [8] 王晓燕, 鲁华祥, 金敏, 等. 基于相关性的小波降噪心电信号去噪算法[J]. 智能系统学报, 2016(6): 827-834.
- [9] 陈辰. 用于身份识别的心电信号关键特征研究[D]. 天津: 天津理工大学, 2015.
- [10] 叶继伦, 孙纪光, 吴跃胜, 等. 一种新型、快速心电图P波检测算法[J]. 中国医疗器械杂志, 2017, 41(1):13-16.
- [11] 冯玉蓉, 陈玮, 蔡光跃. 基于心电信号的生物特征提取和识别[J]. 计算机与数字工程, 2018, 46(6): 1099-1103.
- [12] 刘佳昱, 黄亦翔, 李炳初, 等. 基于双正交小波变换QRS波检测方法研究与改进[J]. 机电一体化, 2014, 20(5):21-25, 61.
- [13] 刘朋. 智能心电监测系统的研究与设计[D]. 哈尔滨: 哈尔滨理工大学, 2018.
- [14] 李朋. 聚类分析中新聚类有效性指标的研究[D]. 合肥: 安徽大学, 2018.
- [15] 吴士云. 基于人工神经网络的电梯导轨校直算法的研究[D]. 南京: 南京理工大学, 2014.
- [16] 张清丽. 心梗特征提取与辅助诊断模型研究[D]. 郑州: 郑州大学, 2017.