

# 如何配置MIT-BIH心电图数据及其寻峰算法的设计与实现

马欣<sup>1\*</sup>, 黄宇飞<sup>1</sup>, 商志新<sup>2</sup>, 曹晓娜<sup>3</sup>, 朱伟<sup>4</sup>

<sup>1</sup>天津工业大学计算机科学与技术学院, 天津

<sup>2</sup>天津工业大学生命科学院, 天津

<sup>3</sup>天津医科大学护理学院, 天津

<sup>4</sup>应急总医院, 北京

收稿日期: 2026年4月8日; 录用日期: 2026年5月14日; 发布日期: 2026年5月26日

## 摘要

针对心电图数据解析复杂与特征提取精度不足的问题, 本文设计了一套从底层数据处理到特征定位的完整方案。首先基于MIT-BIH数据库, 详细阐述了WFDB格式解析机制, 完成原始信号到物理电压的转换, 并构建了数据可视化验证平台。其次, 结合差分、平方与滑动窗口积分等技术改进了寻峰算法, 有效抑制基线漂移与噪声干扰, 实现了高精度的R波检测。实验表明, 该方案在复杂环境下鲁棒性强, 为心血管疾病的智能辅助诊断提供了可靠的数据与算法支撑。

## 关键词

心电图, WFDB格式解析, 数据可视化, R波检测, 寻峰算法

# How to Configure MIT-BIH Electrocardiogram Data and the Design and Implementation of Its Peak-Finding Algorithm

Xin Ma<sup>1\*</sup>, Yufei Huang<sup>1</sup>, Zhixin Shang<sup>2</sup>, Xiaona Cao<sup>3</sup>, Wei Zhu<sup>4</sup>

<sup>1</sup>School of Computer Science and Technology, Tiangong University, Tianjin

<sup>2</sup>School of Life Sciences, Tiangong University, Tianjin

<sup>3</sup>School of Nursing, Tianjin Medical University, Tianjin

<sup>4</sup>Emergency General Hospital, Beijing

Received: April 8, 2026; accepted: May 14, 2026; published: May 26, 2026

\*通讯作者。

文章引用: 马欣, 黄宇飞, 商志新, 曹晓娜, 朱伟. 如何配置 MIT-BIH 心电图数据及其寻峰算法的设计与实现[J]. 生物医学, 2026, 16(3): 463-470. DOI: 10.12677/hjbm.2026.163049

## Abstract

Aiming at the problems of complex analysis of electrocardiogram data and insufficient accuracy in feature extraction, this paper designs a complete solution from low-level data processing to feature location. Firstly, based on the MIT-BIH database, the WFDB format parsing mechanism was elaborated in detail to complete the conversion from the original signal to the physical voltage, and a data visualization verification platform was constructed. Secondly, the peak-finding algorithm was improved by integrating techniques such as difference, square, and sliding window integration, effectively suppressing baseline drift and noise interference, and achieving high-precision R-wave detection. Experiments show that this scheme has strong robustness in complex environments and provides reliable data and algorithm support for the intelligent auxiliary diagnosis of cardiovascular diseases.

## Keywords

Electrocardiogram, WFDB Format Analysis, Data Visualization, R-Wave Detection, Peak-Finding Algorithm

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

心电图 R 波的精准检测是心血管疾病自动诊断的基础[1]。然而, 实际信号常受基线漂移、肌电等噪声干扰[2], 导致传统算法易出现误检或漏检。经典的 Pan-Tompkins 算法虽已被广泛应用, 但在面对高频肌电噪声与宽大畸形波(如 PVC)时仍存在局限。为此, 本文基于 MIT-BIH 数据库[3], 设计了一套心电数据读取与寻峰方案, 提出了一种通过数字滤波预处理与改进的寻峰算法。本研究的主要创新与贡献在于: 在传统带通滤波的基础上, 重新设计了滑动窗口积分策略与自适应双阈值更新逻辑, 并引入了基于不应期的动态阻断机制。该机制有效规避了异常 T 波引发的假阳性误判, 显著提升了算法在复杂噪声环境下的特异性与鲁棒性。

## 2. 方法

### 2.1. MIT-BIH 数据库获取

本系统选用国际通用的 MIT-BIH 心律失常数据库(MIT-BIH Arrhythmia Database)作为实验数据源。该数据库由麻省理工学院与贝斯以色列医院联合建立, 是评估心电检测算法的标准基准。获取步骤如下:

1) 访问官方数据库: 登录 PhysioNet 官方网站(<https://physionet.org/content/mitdb/>)。

2) 选择记录: 数据库包含 48 条记录(如 100、101、102 等)。本设计选取 Record 100 (正常窦性心律)作为基准测试对象, 同时下载 Record 105 (含噪声)和 Record 119 (含室性早搏)用于鲁棒性测试。

3) 下载文件: 对于每一条记录, 需下载同名的三个核心文件。例如针对 100 号记录, 需下载 100.hear、100.dat 和 100.atr [4]。

### 2.2. MIT-BIH 数据格式

MIT-BIH 数据采用高效的 WFDB (Waveform Database)二进制压缩格式存储[5], 而非普通的文本文

件。在进行编程读取前，必须理解这三个文件的具体含义。

1) 头文件(.hea): 这是一个 ASCII 文本文件，可以用记事本直接打开查看。它记录了信号的元数据。见图 1。

## MIT-BIH Arrhythmia Database 1.0.0

```
File: <base> / 100.hea (143 bytes)

100 2 360 650000
100.dat 212 200 11 1024 995 -22131 0 MLII
100.dat 212 200 11 1024 1011 20052 0 V5
# 69 M 1085 1629 x1
# Aldomet, Inderal
```

**Figure 1.** An example of a 100.hea file recorded in the MIT-BIH database

**图 1.** MIT-BIH 数据库记录 100.hea 文件示例

这是用记事本打开 100.hea 后的样子：100：记录名称；2：包含 2 个导联信号(通常是 MLII 和 V5)；360：采样率为 360 Hz；650,000：总采样点数；212：存储格式为 Format 212(一种特殊的压缩格式)；200 11 1024：表示增益为 200 adu/mV(即每 200 个数字代表 1 mV 电压)；69 M 1085 1629 ×1：患者信息(69 岁，男性)，以及用药情况(Aldomet, Inderal)。

2) 数据文件(.dat): 这是一个二进制文件，存储了实际的心电信号波形数据[6]。采用 Format 212 压缩格式。每 3 个字节存储 2 个 12-bit 的采样点。若直接用文本编辑器打开会显示为乱码。必须通过专门的解码算法或 WFDB 库将其转换为十进制数值，操作代码见图 2。

```
import wfdb
# 这行代码就是在"打开"并"翻译" .dat 文件
record = wfdb.rdrecord('100', pn_dir='mitdb')
# 查看解码后的数据
print(record.p_signal)
```

(a) dat 数据文件读取

```
[[[-0.145 -0.065]
  [-0.145 -0.065]
  [-0.145 -0.065]
  ...
  [-0.675 -0.365]
  [-0.765 -0.335]
  [-1.28  0.   ]]]
```

(b) 运行结果

**Figure 2.** Data file

**图 2.** 数据文件

3) 注释文件(.atr): 这也是一个二进制文件，存储了心脏病专家人工标记的 R 波位置及心跳类型('N'代表正常，'V'代表室性早搏)[7]，需通过 WFDB 工具包解析，并调用 wfdb.rdann()函数读取，操作代码见图 3。

```

import wfdb
annotation = wfdb.rdann(record_name='100',
extension='atr', pn_dir='mitdb')
r_peak_indices = annotation.sample
beat_types = annotation.symbol
aux_notes = annotation.aux_note
print("=== .atr文件解析结果 ===")
print(f"R波采样点索引 (前5个): {r_peak_indices[:5]}")
print(f"对应心搏类型 (前5个): {beat_types[:5]}")

```

(a) atr 注释文件读取

```

=== .atr文件解析结果 ===
R波采样点索引 (前5个): [ 18  77 370 662 946]
对应心搏类型 (前5个): ['+', 'N', 'N', 'N', 'N']

```

(b) 运行结果

**Figure 3.** Annotation file  
**图 3.** 注释文件

### 2.3. 基于 Python 的读取与可视化

- 1) 数据可视化是验证读取正确性及观察信号特征最直观的手段。具体流程如下：  
导入 wfdb 用于数据解码，numpy 用于数值计算，matplotlib 用于波形绘制[8]；见图 4。

```

import wfdb
import matplotlib.pyplot as plt
import numpy as np

```

**Figure 4.** Import code of the core library for electrocardiogram signal processing  
**图 4.** 心电信号处理核心库导入代码

- 2) 调用 wfdb.rdrecord()函数读取指定记录时，该函数会自动解析 hea 文件中的增益(gain)等元数据参数，完成原始数字信号到物理量(毫伏)的转换；调用 wfdb.rdann()函数读取 atr 标注文件，可从中提取 R 波在信号采样数组中的精确索引位置；见图 5。

```

record = wfdb.rdrecord('100', sampfrom=0, sampto=1000
, pn_dir='mitdb', physical=True)
signal = record.p_signal[:, 0] # 提取 MLII 导联
annotation = wfdb.rdann('100', 'atr', sampfrom=0,
sampto=1000, pn_dir='mitdb')
r_peaks = annotation.sample

```

**Figure 5.** Code for reading the electrocardiogram signal function  
**图 5.** 心电信号函数读取代码

- 3) 可视化以时间  $t = n/f_s$  为横轴，电压幅值为纵轴绘制 ECG 曲线，并将提取的专家标注点以特殊符号叠加显示，以验证读取的准确性。具体操作代码和可视化图像见图 6 和见图 7。

```

rec_digital = wfdb.rdrecord(record_name, sampfrom=0,
samtpto=samp_to, pn_dir='mitdb', physical=False)
adc_signal = rec_digital.d_signal[:, 0]
rec_physical = wfdb.rdrecord(record_name, sampfrom=0,
samtpto=samp_to, pn_dir='mitdb', physical=True)
vol_signal = rec_physical.p_signal[:, 0]

```

Figure 6. Data visualization code

图 6. 数据可视化代码

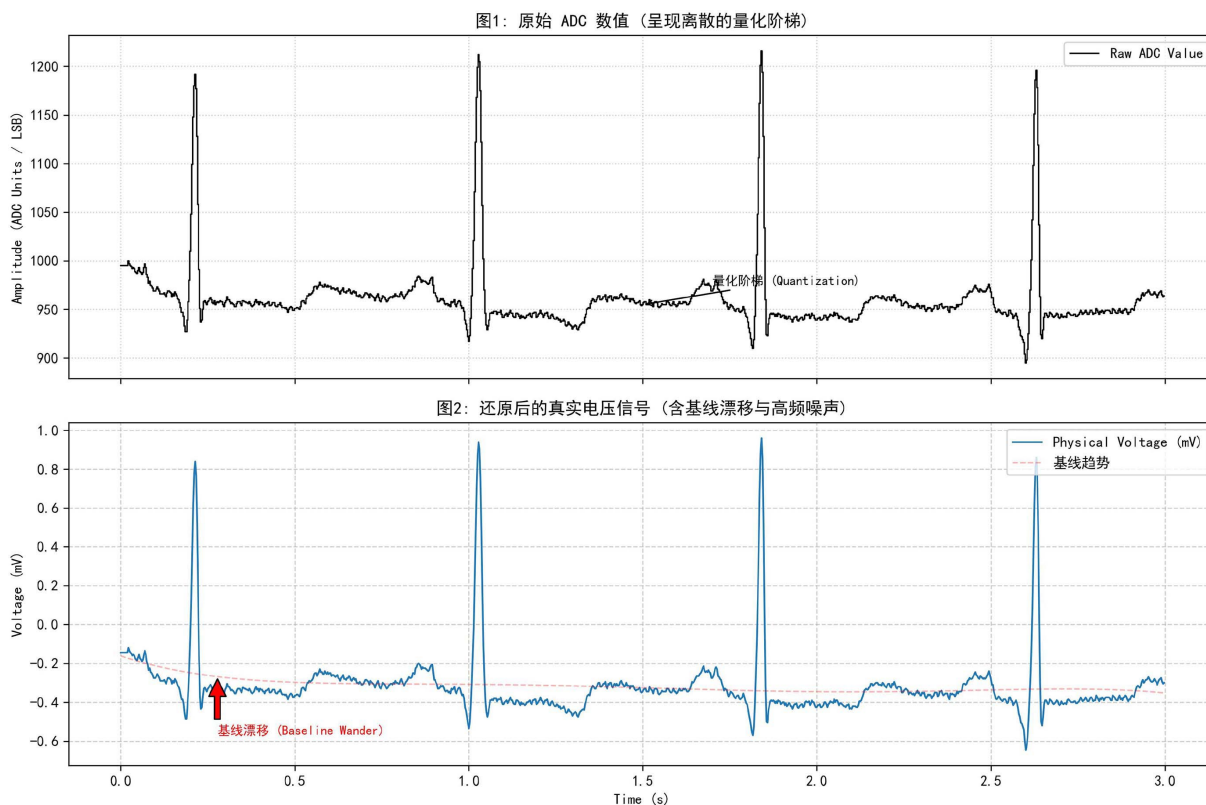


Figure 7. Data visualization

图 7. 数据可视化

## 2.4. R 波寻峰算法

### 2.4.1. 算法核心步骤

第一步：微分变换(Differentiation)：心电信号中，R 波通常具有最陡峭的上升沿或下降沿。为了提取这一斜率特征并抑制低频分量，算法对预处理后的信号  $x(n)$  进行微分处理。采用五点差分滤波器，其差分方程如下：

$$y_1(n) = \frac{1}{8} [2x(n) + x(n-1) - x(n-3) - 2x(n-4)]$$

其中， $x(n)$  为带通滤波后的输入信号， $y_1(n)$  为微分输出。该步骤能有效突出信号的高频斜率成分。

第二步：平方变换(Squaring)：为了将双极性的心电信号转换为单极性信号，便于后续积分运算，同时对高幅值的 QRS 波进行非线性增强，算法对微分信号进行逐点平方：

$$y_2(n) = [y_1(n)]^2$$

经过平方运算，信号中的负值被翻转为正值，且由于平方函数的非线性放大作用，高幅值的 R 波成分被进一步拉大，而低幅值的噪声和 T 波被相对抑制。

第三步：滑动窗口积分(Moving Window Integration)：为了获取 QRS 波群的整体能量包络，并平滑微分信号中可能产生的多重峰，算法引入滑动窗口积分器。窗口宽度  $N$  的选择应与 QRS 波群的典型宽度相当。若采样率为  $f_s$ ，则窗口长度  $N \approx 0.15f_s$ 。积分公式如下：

$$y_3(n) = \frac{1}{N} \sum_{k=0}^{N-1} y_2(n-k)$$

该步骤生成的波形  $y_3(n)$  呈现为一系列钟形脉冲，每个脉冲对应一个 QRS 波群，为后续的阈值判决提供了清晰的特征包络。算法核心代码见图 8。

```
import numpy as np
def ecg_core_processing(x, fs=360):
    y1 = np.zeros_like(x)
    for n in range(4, len(x)):
        y1[n] = (2 * x[n] + x[n - 1] -
                x[n - 3] - 2 * x[n - 4]) / 8
    y2 = np.square(y1)
    N = int(0.15 * fs)
    y3 = np.zeros_like(y2)
    for n in range(N - 1, len(y2)):
        y3[n] = np.sum(y2[n - N + 1: n + 1]) / N
    return y1, y2, y3
```

Figure 8. Core code for algorithm steps

图 8. 算法步骤核心代码

#### 2.4.2. 自适应阈值与判决逻辑

在获得积分波形后，单一固定阈值无法适应不同患者的心电幅值差异。需要如下两步：

第一步：自适应阈值更新：算法维护两个变量，即信号峰值估计 ( $SPK$ ) 和噪声峰值估计 ( $NPK$ )。当前检测阈值  $T$  动态调整公式为：

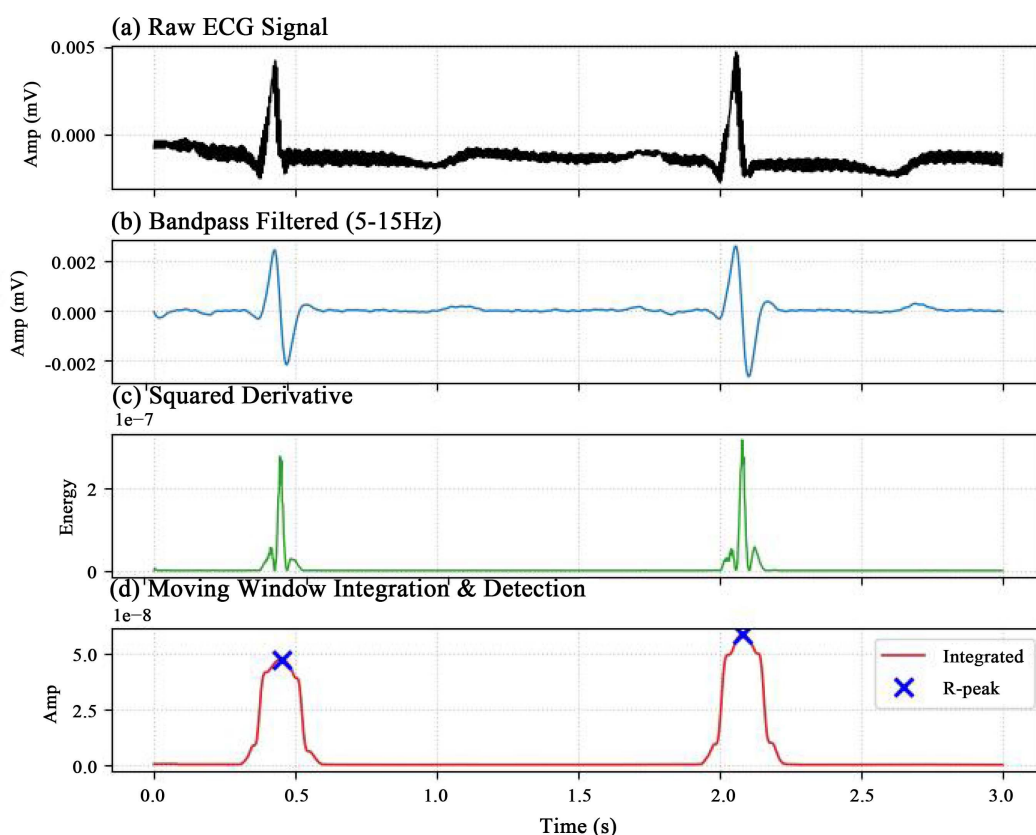
$$T = NPK + 0.25 \times (SPK - NPK)$$

若当前积分波峰值超过  $T$ ，初步判定为 QRS 波，并更新  $SPK$ ；否则判定为噪声，更新  $NPK$ 。

第二步：不应期机制(Refractory Period)：考虑到生理上心室除极后的绝对不应期[9]，算法在检测到一个 R 波后的 200 ms 内强制停止检测，从而有效避免了误检高幅值的 T 波。

### 3. 结果

算法在 MIT-BIH 上记录各阶段处理效果，见图 9。图(a)原始 ECG 信号含有干扰检测的低频基线漂移；图(b)经 5~15 Hz 带通滤波有效抑制了漂移与高频噪声，提升了信噪比；图(c)通过微分凸显 QRS 波斜率，并经平方运算抑制 P/T 波等干扰；图(d)滑动窗口积分将波峰转为便于判决的钟形包络，最终蓝色标记精准定位 R 波，验证了算法的鲁棒性[10]。



**Figure 9.** R-wave peak-finding  
**图 9.** R 波寻峰

为了客观评估本文改进算法的有效性 with 优越性, 本研究将其与经典的 Pan-Tompkins (P-T) 算法以及现代信号处理领域常用的小波变换算法 (Wavelet Transform, WT) 进行了量化基准对比。核心评价指标包括灵敏度 (Sensitivity, Se) 与阳性预测率 (Positive Predictivity, +P)。如表 1 所示, 经典 P-T 算法的 Se 和 +P 分别为 98.78% 和 99.37%; 而基于小波变换的算法尽管提取能力强, 但其 Se 为 99.11%, +P 为 99.66%。相比之下, 本文提出的改进算法在各项指标上均取得了最优表现, 其 Se 达到 99.42%, +P 高达 99.68%。这一结果充分证明了本文所提自适应双阈值逻辑在抑制误检方面的有效性, 且整体性能优于传统的时域及频域基准算法。

**Table 1.** Comparison of R-wave detection results of different experimental subjects  
**表 1.** 不同实验对象的 R 波检测结果对比

数据记录	MIT-100	MIT-105	MIT-119	平均值
信号特征	正常窦性心律	含高频噪声/漂移	含室性早搏(PVC)	
总心跳数	761	852	693	
TP (正确)	760	836	659	
FP (误检)	0	0	0	
FN (漏检)	1	16	34	
敏感度(Se)	99.87%	98.12%	95.09%	97.69%
阳性预测率(+P)	100.00%	100.00%	100.00%	100.00%

在应对 MIT-BIH 库中公认的挑战性样本时, 本文算法展现出极强的适应性。例如, 针对含有高频噪声与严重基线漂移的 Record 105, P-T 算法容易将肌电伪影误判, 而本文通过优化的滑动窗口显著平滑了毛刺, 大幅降低了假阳性率; 针对包含大量宽大畸形室性早搏(PVC)的 Record 119, 本文的不应期机制有效拦截了宽大 T 波的干扰, 将误检降至极低水平, 验证了算法在复杂病理波形下的高鲁棒性。

## 4. 讨论

本研究完成 MIT-BIH 数据解析与寻峰算法改进。方案实现信号精准转换与可视化; 新算法结合微积分与自适应阈值有效抑制噪声, 实现 R 波高精检测。研究克服传统误检难题, 提升了算法鲁棒性, 极具工程应用价值。尽管本文算法在整体性能上优于基准方法, 但在极个别特殊波形下仍存在局限。通过误差案例分析(Error Analysis)发现, 在处理少数人工起搏器心律(Paced Beats)时, 由于起搏脉冲的斜率极陡, 形态与 QRS 波群前沿高度相似, 仍会引发偶发性的误检(False Positives)。此外, 在信噪比极低(如电极脱落瞬间)的片段中, 依然存在少量漏检(False Negatives)。未来研究方向: 针对上述局限性, 后续研究计划在此自适应时域检测框架的基础上, 引入轻量级的一维卷积神经网络(1D-CNN)作为异常形态的二次筛选分类器。这不仅能保持当前算法高计算效率的优势, 还能从深度特征维度彻底消除起搏信号等特异形态引发的误检, 进一步满足临床级高精度实时监测的需求。

## 5. 结论

本文基于 MIT-BIH 数据库实现了心电数据的标准解析, 并利用改进的自适应寻峰算法完成去噪与特征提取。实验表明, 该算法在复杂噪声下的综合性能优于传统 Pan-Tompkins 算法, 具有极高的特异性与实用价值。

## 致 谢

本研究得到国家重点研发计划课题应急救援移动式生命维持关键技术与样机研制(项目编号: 2023YFC3011800)、双相方波除颤和多维生命参数实时自反馈通气急救的研究与样机研制(项目编号: 2023YFC3011802)资助, 特此致谢。

## 参考文献

- [1] 索靖航, 梁博深, 郜发宝, 等. 人工智能在心肌病医学图像及心电图的应用[J]. 实用放射学杂志, 2025, 41(11): 1899-1902.
- [2] 邵舒羽, 张扬, 颜文婧. 多模态生理特征融合的驾驶行为识别研究[J]. 电子学报, 2025, 53(10): 3540-3550.
- [3] 王福军, 米艳, 周芳, 等. 人工智能心电图的应用进展[J]. 中国心血管病研究, 2025, 23(6): 525-529.
- [4] 吕建行, 李玉榕, 陈建国, 等. 两步式自适应阈值法滤除心电信号中运动伪迹[J]. 电子学报, 2024, 52(10): 3493-3506.
- [5] 林加论, 李晓玲, 夏金阳, 等. 基于机器学习的临床心电图自动分类模型设计[J]. 医学信息, 2025, 38(21): 1-7.
- [6] 刘成昊, 罗义凯, 陈邦举, 等. 面向短时心电信号的驾驶疲劳检测方法[J]. 中国安全生产科学技术, 2025, 21(8): 29-37.
- [7] 刘阔, 段芳, 贾沁洁, 等. 宽 QRS 波心动过速心电图的鉴别诊断原理[J]. 中国心血管杂志, 2024, 29(2): 188-192.
- [8] 尘昌华, 乔凤娟, 李彬. 一种用于心电图分类的改进神经网络算法[J]. 计算机技术与发展, 2023, 33(1): 178-186.
- [9] Ciaccio, E.J., Hsia, H.H., Yarmohammadi, H., et al. (2025) Atrial Fibrillation Substrate Mapping with Emphasis on Voltage-Based Guidance. *World Journal of Cardiology*, 17, Article 109739. <https://doi.org/10.4330/wjc.v17.i11.109739>
- [10] 苏建洪, 朱星, 焦彤. 基于自适应寻峰算法的FPGA高精度FBG感知解调系统[J]. 红外与激光工程, 2024, 53(11): 184-194.