

基于多域特征融合与通道注意力的异常心音检测

和含雷, 李子阳, 郑钟月, 朱加飞, 张丽艳*

大连交通大学轨道智能工程学院电子与通信工程系, 辽宁 大连

收稿日期: 2026年2月27日; 录用日期: 2026年3月19日; 发布日期: 2026年3月30日

摘要

心音信号蕴含丰富的生理与病理信息, 是辅助心血管疾病筛查与早期诊断的重要生物特征。传统心音分析方法通常依赖单一特征或手工规则, 难以充分刻画时频结构的复杂变化, 导致在多噪声环境和不同个体差异下识别准确率受限。为此, 本文提出一种基于多域特征融合与通道注意力机制的心音异常检测卷积网络(TTFI-CNN)。该网络综合利用MFCC (Mel频率倒谱系数)、STFT (短时傅里叶变换)与DWT (离散小波)三类互补特征, 构建“时域-频域-时频域”的多特征表示, 并设计轻量化SE (Squeeze-and-Excitation)通道注意力模块增强特征选择能力。此外, 模型还结合EarlyStopping与交叉验证设计, 以提升泛化能力。基于2016 PhysioNet/CinC心音挑战赛数据集进行验证, 所提出的模型在异常心音检测任务上取得97.8%的准确率。其中, 消融实验证明: 去除DWT分支会导致准确率下降至96.4%, 去除SE注意力则下降至97.09%, 充分说明三分支融合与通道注意力在提升分类性能方面的有效性。实验结果表明, 本方法能够稳健地捕获多尺度时频特征, 对心音异常检测具有良好的应用潜力。

关键词

心音分析, 异常检测, 多特征融合, 卷积神经网络, 注意力机制

Anomaly Heart Sound Detection Based on Multi-Domain Feature Fusion and Channel Attention

Hanlei He, Ziyang Li, Zhongyue Zheng, Jiafei Zhu, Liyan Zhang*

Department of Electronic Communications, School of Railway Intelligent Engineering, Dalian Jiaotong University, Dalian Liaoning

Received: February 27, 2026; accepted: March 19, 2026; published: March 30, 2026

*通讯作者。

文章引用: 和含雷, 李子阳, 郑钟月, 朱加飞, 张丽艳. 基于多域特征融合与通道注意力的异常心音检测[J]. 图像与信号处理, 2026, 15(2): 155-164. DOI: 10.12677/jisp.2026.152013

Abstract

Heart sound signals contain rich physiological and pathological information, serving as crucial biomarkers for assisting in the screening and early diagnosis of cardiovascular diseases. Traditional heart sound analysis methods typically rely on single features or manually designed rules, making it difficult to fully characterize the complex variations in time-frequency structures. This often results in limited recognition accuracy in environments with high noise levels or significant individual differences. To address this, this paper proposes a time-frequency feature fusion and channel attention-based convolutional neural network (TTFI-CNN) for abnormal heart sound detection. The network comprehensively utilizes three complementary features—Mel-Frequency Cepstral Coefficients (MFCC), Short-Time Fourier Transform (STFT), and Discrete Wavelet Transform (DWT)—to construct a multi-feature representation spanning the “time-domain, frequency-domain, and time-frequency domain”. Additionally, a lightweight Squeeze-and-Excitation (SE) channel attention module is designed to enhance feature selection capabilities. Furthermore, the model incorporates EarlyStopping and cross-validation strategies to improve generalization. Validated on the 2016 PhysioNet/CinC heart sound challenge dataset, the proposed model achieves an accuracy of 97.8% in the task of abnormal heart sound detection. Ablation studies demonstrate that removing the DWT branch reduces accuracy to 96.4%, while removing the SE attention module lowers it to 97.09%, fully confirming the effectiveness of the three-branch fusion and channel attention in enhancing classification performance. The experimental results indicate that this method can robustly capture multi-scale time-frequency features, demonstrating promising potential for application in abnormal heart sound detection.

Keywords

Heart Sound Analysis, Anomaly Detection, Multi-Feature Fusion, Convolutional Neural Network, Attention Mechanism

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

心血管疾病是全球范围内主要的致死与致残原因之一[1],其早期筛查与准确诊断对于降低疾病负担、改善患者预后具有重要意义。心音信号(Phonocardiogram, PCG)作为一种无创、低成本且易于采集的生理信号,能够反映心脏瓣膜及血流状态,在临床辅助诊断与可穿戴健康监测设备中展现出良好的应用前景。随着智能感知与计算能力的提升,基于心音信号的自动分析方法逐渐成为传统听诊的重要补充。

近年来,深度学习技术在医学信号处理领域取得显著进展,在心音分类任务中的表现尤为突出,性能已超越传统方法[2]-[4]。特别是自 PhysioNet/CinC Challenge 2016 [5]发布后,大量基于机器学习与深度学习的心音分析方法相继提出,有力推动了该领域的快速发展。Posts 等人[6]将卷积神经网络(Convolutional Neural Network, CNN)与自适应增强算法(Adaptive Boosting, AdaBoost)相结合,从心音信号中提取多维时频特征,在随机测试集上取得了较高的未加权平均召回率。Yaseen 等人[7]则利用梅尔频率倒谱系数(Mel Frequency Cepstral Coefficients, MFCC)与离散小波变换(Discrete Wavelet Transform, DWT)特征,结合支持向量机(Support Vector Machine, SVM)、K 近邻算法(K-Nearest Neighbors, KNN)及深度神经网络(Deep Neural Network, DNN)开展分类研究,在公开数据集上获得了较为理想的识别精度。Mubarak 等人[9]基于同

态包络分析与径向基函数核 SVM 实现心音分类; Li 等人[10]则通过强化多域特征融合与引入全局平均池化结构, 进一步提升了模型的泛化性能。Zhang 等人[11]提出一种基于心音循环谱图与偏最小二乘回归(Partial Least Squares Regression, PLSR)的心音分类方法, 从频谱结构角度对心音信号实现了有效建模。尽管上述方法在特定数据集上取得了一定效果, 但由于心音信号本身具有非平稳性强、个体差异显著、噪声干扰复杂等特点, 现有方法在实际应用中仍面临性能退化的问题。为此, 本文提出一种基于多特征融合与注意力机制的心音异常检测模型——TTFI-CNN (Time-Time-Frequency Integrated Convolutional Neural Network)。该模型通过融合 MFCC、短时傅里叶变换(Short-Time Fourier Transform, STFT)以及离散小波变换(DWT)特征, 从时域、频域与时频域多个层面构建心音信号的多尺度表征, 有效增强了特征表达能力。同时, 在多分支特征融合阶段引入轻量化的 SE (Squeeze-and-Excitation)通道注意力机制, 对不同特征分支的重要性进行自适应加权, 从而 selectively 强化与异常心音相关的关键信息。为验证所提出方法的有效性, 本文在 PhysioNet/CinC Challenge 2016 心音数据集上进行了系统实验, 并通过消融实验分析多分支融合结构及 SE 注意力模块对模型性能的影响。

2. 本文算法原理

2.1. 整体框架

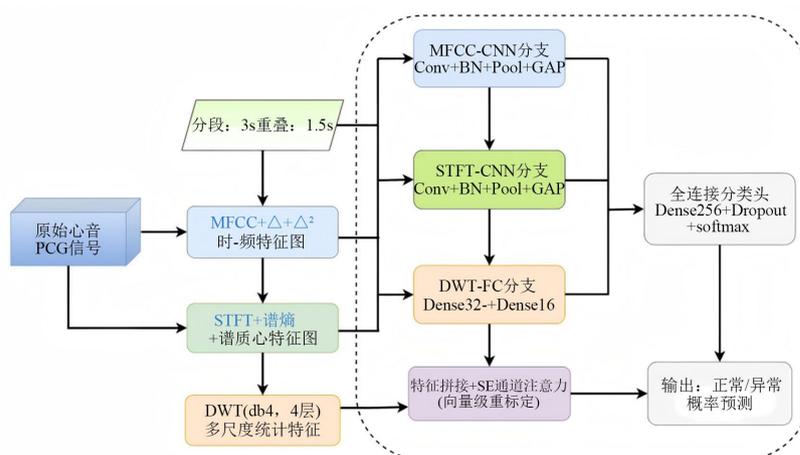


Figure 1. Overall framework of the abnormal heart sound detection algorithm
图 1. 异常心音检测算法整体框架

本文提出的 TTFI-CNN 异常心音检测算法整体框架如图 1 所示。该算法包括预处理、特征提取, 首先, 对原始 PCG 信号以 4 kHz 采样率读取, 并采用 3 s 窗长、1.5 s 重叠的方式划分为若干短时段; 随后利用 4 阶 Butterworth 带通滤波器(25~1000 Hz)抑制背景噪声并进行幅值归一化处理。在此基础上, 面向每个片段同时构建三类互补特征: 一是基于 Mel 频谱的 MFCC 及其一、二阶差分系数, 形成 $120 \times T$ 维的时-频倒谱特征图; 二是由 STFT 幅值谱、谱质心与谱带宽构成的 66×47 维时-频能量与谱形特征; 三是通过 db4 小波在 4 个尺度上的分解系数统计得到的 10 维时域纹理特征。MFCC 和 STFT 特征分别输入两个浅层卷积子网络, 经卷积、批归一化、LeakyReLU、池化和全局平均池化提取高维判别表征; DWT 特征则通过两层全连接网络提取紧凑的多尺度统计描述。随后, 将三分支特征在向量级进行拼接, 并引入 SE 风格的通道注意力模块, 对融合特征进行自适应重标定, 以突出对异常心音敏感的分支与维度, 抑制冗余或噪声信息。最后, 融合后的高层表征输入两层全连接分类头, 通过带标签平滑的交叉熵损失进行端到端训练, 从而实现正常与异常心音的自动识别。需要说明的是, PhysioNet/CinC Challenge 2016 数据集

在官方评测框架下将心音信号标注为正常与异常两类，构成二分类任务。然而，从临床角度看，异常心音样本来源于不同类型的心脏病理，其声学表现具有较大的异质性。由于数据集中未提供明确的病理子类别标签，本文遵循官方任务定义，将异常心音统一建模为一类，重点关注异常心音的自动检测能力。

2.2. 预处理

由于心音分类仅需判断“正常/异常”而非定位具体病灶，本研究无需精确分割心动周期。原始心音经数字听诊器采集后时长 5~120 s 不等，采样率重采样至 4 kHz。为消除设备启停瞬态，首先剔除每条记录首、尾各 1 s；随后采用 5 阶 Butterworth 带通滤波器(20 Hz~1 kHz, zero-phase filt-filt)去除基线漂移与高频噪声，保留心音主要能量频段。滤波后执行滑动切片：窗口 3 s，步长 1.5 s (50%重叠)，确保每段至少包含一个完整心动周期($\approx 0.8\sim 1.0$ s)。不足 3 s 的尾部自动舍去。切片过程加汉明窗抑制频谱泄漏，最终输出固定长度 12,000 点(3 s \times 4 kHz)的浮点序列，归一化至零均值、单位方差，其公式如(1)所示：

$$x_{\text{norm}}(t) = \frac{x(t) - \mu}{\sigma} \quad (1)$$

经上述处理，数据集共生成 26,410 段正常样本、8080 段异常样本，后续特征提取均以该等长、去噪、归一化片段为输入。

2.3. 特征提取

心音信号具有显著的非平稳性和多尺度特征，其异常模式在时域、频域及多尺度能量分布上均存在差异。为全面表征心音信号的判别信息，本文从时频分析与多尺度统计两个层面出发，分别提取梅尔频率倒谱系数(MFCC)、短时傅里叶变换(STFT)特征以及离散小波变换(DWT)特征，并将其作为多分支网络的输入。

首先，为刻画心音信号的局部时频特性，本文对每个心音片段计算梅尔频谱，并在此基础上提取 40 维 MFCC 特征。同时引入 MFCC 的一阶和二阶差分，以描述心音信号在时间维度上的动态变化信息。通过该方式获得的 MFCC 特征能够有效反映心音在不同频带上的能量分布及其随时间的变化趋势，在心音分类任务中具有较强的判别能力。

其次，为弥补 MFCC 在频谱细节刻画方面的不足，本文进一步引入 STFT 特征对心音信号进行时频分析。通过对心音信号进行短时分帧并计算其幅度谱，获得随时间变化的频谱表示。为增强特征的数值稳定性，对 STFT 幅度谱取对数形式，并对频率和时间维度进行统一裁剪，从而形成固定尺寸的时频特征图。该特征能够直观反映异常心音在不同时间区间内的频谱变化特征。

此外，考虑到异常心音在多尺度结构上的显著差异，本文采用离散小波变换对心音信号进行多尺度分解。选用 db4 小波基函数并分解至四层，分别对各尺度小波系数计算其统计特征(均值与标准差)，以获取反映心音整体结构和能量分布变化的低维特征表示。DWT 特征能够有效弥补时频特征在全局尺度信息建模方面的不足。

综合而言，MFCC 和 STFT 特征侧重于刻画心音信号的局部时频结构，而 DWT 特征则从多尺度角度描述信号的整体统计特性。三类特征在信息表达上具有良好的互补性，为后续 TTFI-CNN 网络的多分支融合学习提供了更加丰富和稳定的特征基础。

2.4. 基于特征融合的通道注意力机制

在多分支特征融合场景下，不同特征分支及其内部通道对异常心音的判别贡献存在显著差异。为充分挖掘多域特征之间的互补信息，并自适应突出对异常心音更为敏感的特征通道，本文在特征拼接之后

引入一种轻量化的通道注意力机制(Channel Attention Mechanism), 对融合特征进行动态重标定。

2.4.1. 特征融合

设经过 MFCC-CNN 分支、STFT-CNN 分支及 DWT-MLP 分支提取后的高层特征分别表示为:

$$f_{\text{MFCC}} \in R^{128}, f_{\text{STFT}} \in R^{128}, f_{\text{DWT}} \in R^{16} \quad (2)$$

三路特征在向量级进行拼接, 得到融合特征表示:

$$f = [f_{\text{MFCC}}; f_{\text{STFT}}; f_{\text{DWT}}] \in R^C \quad (3)$$

其中, $C = 272$ 表示融合后的通道维度。

该融合向量同时包含了心音信号的时 - 频倒谱信息、时 - 频能量结构以及多尺度时域纹理统计特征, 但不同通道中包含的信息冗余程度和判别能力并不一致, 因此有必要对其进行自适应加权。

2.4.2. 通道注意力建模

本文采用 SE (Squeeze-and-Excitation) 风格的通道注意力结构, 对融合特征 f 进行通道重标定。由于特征已在全局池化后呈现为一维向量形式, 通道压缩(Squeeze)阶段可视为对融合特征的全局统计描述, 直接以 f 作为通道描述向量:

$$z = f \quad (4)$$

在激励(Excitation)阶段, 通过两层全连接网络对通道间的相关性进行建模, 生成通道权重向量:

$$s = \sigma(W_2 \delta(W_1 z)) \quad (5)$$

其中, $W_1 \in R^{r \times C}$ 、 $W_2 \in R^{C \times r}$ 为可学习参数矩阵; r 为通道压缩比, 本文设置为 $r = 4$; $\delta(\cdot)$ 表示 ReLU 激活函数; $\sigma(\cdot)$ 表示 Sigmoid 激活函数, 用于将权重归一化至 $[0, 1]$ 。最终, 通过逐通道乘法对融合特征进行重标定, 得到加权特征表示:

$$\tilde{f} = s \odot f \quad (6)$$

其中, \odot 表示 Hadamard 逐元素乘法。

上述通道注意力机制能够在融合特征空间内自适应学习不同通道的重要性权重, 从而增强与异常心音相关的判别特征响应, 并抑制冗余或噪声通道的干扰。与直接拼接特征相比, 该机制在保持模型参数量较低的前提下, 提高了多分支特征融合的有效性与判别能力, 为后续分类层提供更加紧凑且具有判别性的高层表征。

2.5. TTFI-CNN 网络结构

本研究提出了一种基于多特征融合与通道注意力机制的卷积神经网络模型——TTFI-CNN (Time-Time-Frequency Integrated CNN), 用于异常心音信号的自动检测。模型以心音信号为输入, 通过多种时域与频域特征提取方法构建多分支特征表示。具体而言, 首先从原始心音信号中分别提取 MFCC 特征图、STFT 时频谱图以及 DWT 多尺度统计特征, 其中 MFCC 特征维度为 120×47 , STFT 特征维度为 66×47 , DWT 特征为 10 维向量。三类特征分别输入对应的网络分支进行并行建模。

如图 2 所示, MFCC 分支与 STFT 分支均采用基于卷积神经网络的特征提取结构。每个分支依次包含两层二维卷积操作, 卷积核大小均为 3×3 , 卷积通道数分别设置为 64 和 128, 卷积后均接批归一化 (Batch Normalization) 层与 LeakyReLU 激活函数, 以提升网络训练稳定性与非线性表达能力。在第一层卷积后引入最大池化操作以降低特征维度, 同时在池化后加入 Dropout 层 (丢弃率为 0.2) 以缓解过拟合问题。第二层卷积后通过全局平均池化 (Global Average Pooling, GAP) 将二维特征映射压缩为定长一维特征向量,

最终分别获得 128 维的 MFCC 特征表示和 STFT 特征表示。

针对 DWT 提取的低维统计特征，设计独立的多层感知机(MLP)分支进行建模。该分支首先通过一个包含 32 个神经元的全连接层进行非线性映射，随后引入批归一化与 Dropout 层以增强泛化能力，最后通过 16 单元的全连接层输出 DWT 特征表示。

在特征融合阶段，将来自 MFCC、STFT 和 DWT 三个分支的特征向量在通道维度进行拼接，形成 272 维融合特征。为进一步建模不同特征通道对异常心音判别的重要性差异，在融合特征后引入 SE(Squeeze-and-Excitation)通道注意力机制。该模块通过两层全连接结构对融合特征进行压缩与激励，生成通道权重向量，并对原始融合特征进行自适应重标定，从而增强模型对关键异常特征的关注能力。

最后，经过通道重标定后的融合特征输入分类模块，分类模块由一个包含 256 个神经元的全连接层组成，激活函数为 ReLU，并在其后引入 Dropout 层以降低过拟合风险。最终通过 Softmax 输出层实现对心音信号“正常”与“异常”两类的概率预测。

模型训练过程中采用 Adam 优化器对网络参数进行更新，初始学习率设置为 8×10^{-4} ，损失函数采用带标签平滑的分类交叉熵损失函数。训练过程中引入 EarlyStopping 与 ModelCheckpoint 策略，根据验证集准确率动态保存最优模型权重，以进一步提升模型的泛化性能与训练稳定性。

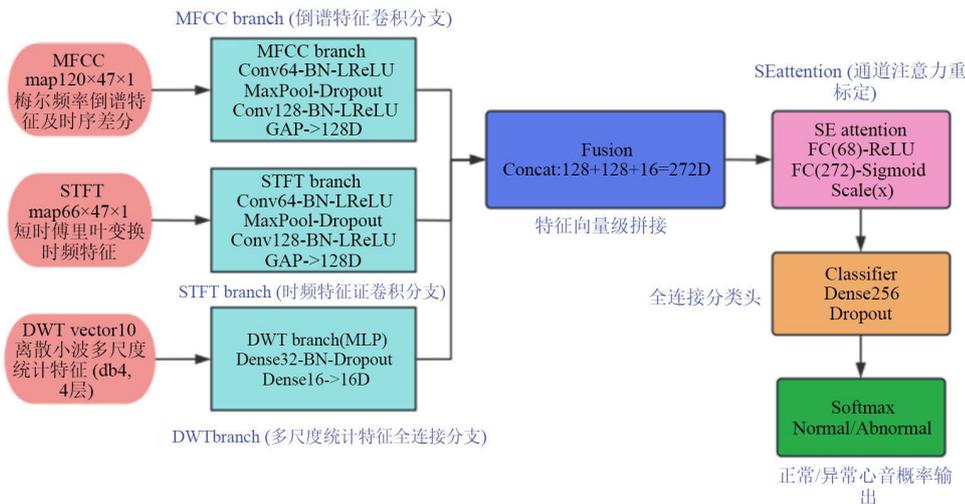


Figure 2. TTFI-CNN network architecture diagram
图 2. TTFI-CNN 网络结构图

3. 实验与结果

3.1. 实验数据

实验中所使用的心音信号来自 PhysioNet/CinC Challenge 2016 数据库，可在 Physio 网站上公开获取。该数据库来自于七个研究小组贡献的九个独立数据库，包括从临床和非临床环境下健康受试者和各种心脏病患者中收集的共 3240 份原始心音信号记录。这些心音数据被各种电子听诊器在 4 个主要听诊部位所采集：主动脉瓣听诊区、肺动脉瓣听诊区、三尖瓣听诊区、二尖瓣听诊区，长度从 5 s 到 120 s 不等，采样频率为 2000 Hz。其中异常的心音信号来自于心脏瓣膜缺损(包括二尖瓣脱垂、二尖瓣关闭不全、主动脉瓣反流和主动脉瓣狭窄等)和冠状动脉疾病等各类心脏病患者。该数据库共发布心音信号 3240 段，其中共有 2575 段心音信号被标记为正常，665 段心音信号被标记为异常。实验中将其中 20%作为测试集，80%作为训练集。具体数据集划分见表 1。在数据划分过程中，严格按照原始心音记录级别进行训练集与

测试集划分, 确保来自同一原始记录的不同切片不会同时出现在训练集与测试集中, 从而有效避免潜在的数据泄露问题。

Table 1. Data division

表 1. 数据划分

数据集	正常/段	异常/段
训练集	2060	532
测试集	515	133

3.2. 实验环境

实验中所使用的编程软件为 Pycharm 2021, 编程语言为 Python 3.8, 所使用并行卷积递归神经网络均在 TensorFlow 2.5 框架中实现。硬件环境: 中央处理器(14 vCPU Intel(R) Xeon(R) Gold 6330 CPU @ 2.00 GHz, RAM 为 90 GB), 独立显卡 NVIDIA Corporation GP104 [GeForce GTX 3090]。

本文的网络模型配置如下: 采用 Adam 优化器对目标函数进行迭代更新, 初始学习率设为 8×10^{-4} , 并结合 EarlyStopping 与 ModelCheckpoint 保证训练稳定性; 损失函数为带标签平滑($\epsilon = 0.05$)的分类交叉熵损失函数; 在各特征分支与融合全连接层中引入 Dropout (保持率 0.4~0.6, 经 Optuna 自动搜索)用于缓解过拟合, 并在融合阶段加入 SE 通道注意力机制以实现特征权重自适应重标定。

3.3. 评价指标

为全面评估所提出的 TTFI-CNN 模型在异常心音检测任务中的分类性能, 本文采用多种常用的二分类评价指标, 包括准确率(Accuracy)、召回率(Recall)、精确率(Precision)、F1 值(F1-score)、特异度(Specificity)和灵敏度(Sensitivity)。上述指标从不同角度反映模型在异常与正常心音识别中的综合判别能力, 其定义如下:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (7)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (8)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (9)$$

$$\text{F1} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (10)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \quad (11)$$

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (12)$$

其中, TP 表示被正确分类为异常心音的样本数量, TN 表示被正确分类为正常心音的样本数量, FP 表示被误判为异常心音的正常样本数量, FN 表示被误判为正常心音的异常样本数量。

3.4. 结果分析

3.4.1. TTFI-CNN 整体性能分析

为验证所提出的 TTFI-CNN 模型在异常心音检测任务中的有效性, 本文基于 PhysioNet/CinC Challenge 2016 数据集开展了系统的实验评估。图 3 展示了模型在测试集上的整体性能表现, 分别从训练收敛

过程、分类结果分布以及判别能力等方面进行分析。

如图 3(a)所示,模型在训练初期准确率迅速提升、损失函数显著下降,并在中后期逐渐趋于平稳;训练集与验证集曲线变化趋势一致,未出现明显振荡或性能退化现象,表明在所采用的数据划分策略与正则化机制下,模型能够实现稳定收敛并具备良好的泛化能力。图 3(b)给出了测试集上的混淆矩阵结果,可以看出绝大多数正常心音与异常心音样本均被正确识别,仅有少量样本发生误判,说明模型在复杂心音信号条件下具有较强的区分能力。进一步分析误分类样本发现,部分漏检(FN)样本对应的异常心音杂音幅度较弱,且易受背景噪声或记录质量影响,其时频结构与正常心音存在一定相似性;而少量误报(FP)样本多源于正常心音中出现的短时高频扰动或非病理性噪声成分,导致模型在多域特征空间中产生异常响应。尽管存在上述情况,得益于多域特征融合与通道注意力机制的引入,模型整体仍能够有效抑制噪声干扰并保持较低的误判比例,体现出较好的鲁棒性。

进一步地,图 3(c)和图 3(d)分别给出了模型的 ROC 曲线与 PR 曲线。其中,ROC 曲线下面积(AUC)超过 0.99,PR 曲线在高召回区间内仍保持较高的精确率,表明模型在不同判决阈值下均具有稳定的异常心音识别能力,尤其在类别分布不均衡的情况下仍能取得良好的检测性能。综合以上分析结果,所提出的 TTFI-CNN 模型在异常心音检测任务中展现出优越的整体性能。

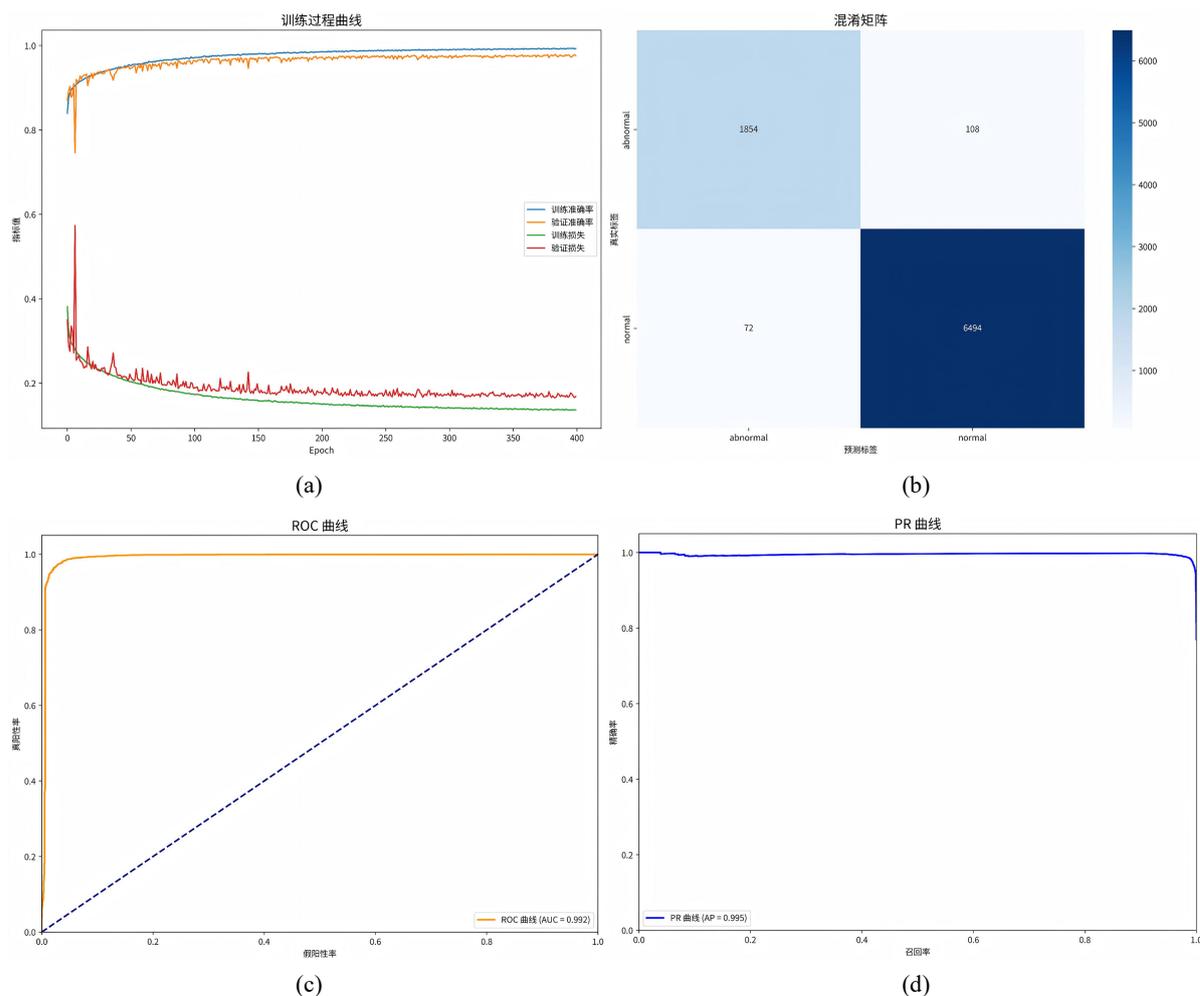


Figure 3. (a) Training process curve; (b) Confusion matrix; (c) ROC curve; (d) PR curve

图 3. (a) 训练过程曲线; (b) 混淆矩阵; (c) ROC 曲线; (d) PR 曲线

为进一步评估所提出方法，本文将 TTFI-CNN 与近年来基于 PhysioNet/CinC Challenge 2016 数据集的代表性深度学习方法进行了对比分析，如表 2 所示。

Table 2. Performance comparison of different methods on the PhysioNet/CinC 2016 dataset
表 2. 不同方法在 PhysioNet/CinC 2016 数据集上的性能对比

方法	特征类型	模型结构	Accuracy (%)
Potes <i>et al.</i> [6] (2016)	手工时频特征	CNN + AdaBoost	86.0
Li <i>et al.</i> [8] (2017)	多域手工特征	CNN + GAP	86.8
Yaseen <i>et al.</i> [7] (2018)	MFCC + DWT	SVM/DNN	97.0
Ren <i>et al.</i> [12] (2021)	Mel/STFT	CNN	95.6
Kumar <i>et al.</i> [13] (2022)	Mel-spectrogram	CNN + Attention	96.3
本文方法	MFCC + STFT + DWT	TTFI-CNN + SE	97.8

表 2 对比了本文方法与近年来基于 PhysioNet/CinC Challenge 2016 数据集的代表性心音分类方法。可以看出，随着特征表示和模型结构的不断改进，整体分类性能逐步提升。本文提出的 TTFI-CNN 通过融合 MFCC、STFT 与 DWT 多域特征，并引入通道注意力机制，在同一数据集上取得了 97.8% 的分类准确率，优于多数已有方法，表明所提方法在异常心音检测任务中具有较好的有效性。

3.4.2. 消融实验

为验证 TTFI-CNN 各组成模块对模型性能的贡献，本文设计了消融实验，对不同结构配置下的模型性能进行了对比分析。实验在相同数据集、相同训练策略与参数设置条件下进行，仅对模型结构进行调整，以保证实验结果的可比性。相关实验结果汇总如表 3 所示。

Table 3. Comparison table of ablation experiments
表 3. 消融实验对比表

模型配置	MFCC	STFT	DWT	SE	Accuracy (%)
二分支模型	√	√	×	√	96.40
三分支(无 SE)	√	√	√	×	97.09
TTFI-CNN (完整)	√	√	√	√	97.80

首先，为分析多尺度统计特征的作用，本文移除了 DWT 分支，仅保留 MFCC 与 STFT 两个特征分支构建二分支融合模型。实验结果表明，该模型的分类准确率为 96.4%，这表明仅依赖时频特征难以充分表征心音信号中的多尺度结构信息，而 DWT 特征能够有效弥补心音信号在不同时间尺度上的统计特性，有助于提升模型对异常心音的识别能力。在此基础上，进一步引入 DWT 分支构建三分支融合模型，但不使用通道注意力机制。该模型的分类准确率提升至 97.09%，说明多域特征融合能够显著增强模型的特征表达能力。然而，与完整模型相比仍存在性能差距，表明简单的特征拼接难以充分刻画各分支特征之间的重要性差异。最后，在三分支融合结构上引入通道注意力机制(SE)，对不同特征分支进行自适应加权，形成最终的 TTFI-CNN 模型。实验结果显示，模型准确率进一步提升至 97.8%。这说明通道注意力机制能够有效突出对异常心音判别更为敏感的特征分量，从而进一步提升模型的整体性能。

4. 结论

本文提出了一种基于多特征融合与通道注意力机制的卷积神经网络心音异常检测方法(TTFI-CNN)。

首先对原始心音信号进行重采样、滤波、归一化及固定窗口重叠切片等预处理操作,以提高信号质量并缓解样本长度不一致与数据不足问题;随后分别从时频域、频域及时域统计角度提取 MFCC、STFT 与 DWT 特征,构建多尺度、多域的心音特征表示;在此基础上,设计了由多分支卷积网络与轻量化 SE 通道注意力模块组成的 TTFI-CNN 模型,实现对不同特征分支的自适应加权融合;最后通过全连接分类器完成正常心音与异常心音的自动判别。实验结果表明,所提出的方法在 PhysioNet/CinC 2016 心音数据集上取得了 97.8% 的分类准确率,在准确率、灵敏度和综合性能方面均优于部分已有方法。消融实验进一步验证了 DWT 多尺度特征与 SE 通道注意力机制在提升模型性能方面的有效性,说明多域特征融合与特征重要性建模对于复杂心音信号分析具有重要意义。该方法无需显式心音分割,具有结构简单、鲁棒性强和泛化能力较好的优点,为心音信号的自动异常检测提供了一种有效且可行的技术方案。

参考文献

- [1] 国家心血管病中心,中国心血管健康与疾病报告编写组.中国心血管健康与疾病报告 2024 概要[J].中国循环杂志,2025,40(6):521-559.
- [2] Zeinali, Y. and Niaki, S.T.A. (2022) Heart Sound Classification Using Signal Processing and Machine Learning Algorithms. *Machine Learning with Applications*, 7, Article ID: 100206. <https://doi.org/10.1016/j.mlwa.2021.100206>
- [3] Krishnan, P.T., Balasubramanian, P. and Umapathy, S. (2020) Automated Heart Sound Classification System from Unsegmented Phonocardiogram (PCG) Using Deep Neural Network. *Physical and Engineering Sciences in Medicine*, 43, 505-515. <https://doi.org/10.1007/s13246-020-00851-w>
- [4] Nguyen, M.T., Lin, W.W. and Huang, J.H. (2022) Heart Sound Classification Using Deep Learning Techniques Based on Log-Mel Spectrogram. *Circuits, Systems, and Signal Processing*, 42, 344-360. <https://doi.org/10.1007/s00034-022-02124-1>
- [5] Liu, C., Springer, D., Li, Q., Moody, B., Juan, R.A., Chorro, F.J., et al. (2016) An Open Access Database for the Evaluation of Heart Sound Algorithms. *Physiological Measurement*, 37, 2181-2213. <https://doi.org/10.1088/0967-3334/37/12/2181>
- [6] Potes, C., Parvaneh, S., Rahman, A. and Conroy, B. (2016). Ensemble of Feature: Based and Deep Learning: Based Classifiers for Detection of Abnormal Heart Sounds. 2016 *Computing in Cardiology Conference (CinC)*, Vancouver, 11-14 September 2016, 621-624. <https://doi.org/10.22489/cinc.2016.182-399>
- [7] Yaseen, Son, G. and Kwon, S. (2018) Classification of Heart Sound Signal Using Multiple Features. *Applied Sciences*, 8, Article No. 2344. <https://doi.org/10.3390/app8122344>
- [8] Li, F., Liu, C., Wang, X. and Liu, C. (2017) Classification of Heart Sounds Using Convolutional Neural Networks Based on Time-Frequency Representations. *Computers in Biology and Medicine*, 84, 67-76.
- [9] Mubarak, Q., Akram, M.U., Shaukat, A., Hussain, F., Khawaja, S.G. and Butt, W.H. (2018) Analysis of PCG Signals Using Quality Assessment and Homomorphic Filters for Localization and Classification of Heart Sounds. *Computer Methods and Programs in Biomedicine*, 164, 143-157. <https://doi.org/10.1016/j.cmpb.2018.07.006>
- [10] Li, F., Tang, H., Shang, S., et al. (2020) Classification of Heart Sounds Using Convolutional Neural Network. *Applied Sciences*, 10, Article No. 3956.
- [11] Zhang, W., Han, J. and Deng, S. (2017) Heart Sound Classification Based on Scaled Spectrogram and Partial Least Squares Regression. *Biomedical Signal Processing and Control*, 32, 20-28. <https://doi.org/10.1016/j.bspc.2016.10.004>
- [12] Ren, J., Liu, C., Zhang, H. and Wang, X. (2021) Heart Sound Classification Using Deep Convolutional Neural Networks and Mel-Frequency Spectral Features. *Biomedical Signal Processing and Control*, 68, Article ID: 102720.
- [13] Kumar, A., Sharma, L.D. and Sunkaria, R.K. (2022) Attention-Based Deep Learning Model for Heart Sound Classification. *Biomedical Signal Processing and Control*, 71, Article ID: 103173.