

# 基于未来引导学习的心电图心律失常分类研究

李玉昇

青岛大学数学与统计学院, 山东 青岛

收稿日期: 2026年3月8日; 录用日期: 2026年4月2日; 发布日期: 2026年4月9日

## 摘要

可穿戴设备在心电图实时监测中对计算资源和推理延迟要求极高, 通常只能截取极短的局部信号片段进行分析。然而, 短窗口轻量级模型因缺失心跳前后的全局上下文信息, 在心律失常分类中面临性能瓶颈。为解决极低延迟要求与局部信息缺失之间的矛盾, 本文提出了一种基于未来引导学习的心律失常分类方法。该方法构建了非对称的教师-学生网络架构, 在训练阶段, 教师模型利用包含心跳前后非对称扩展的宽视窗捕捉代偿间歇等长程演变机制; 随后通过跨时域知识蒸馏, 将结构化的暗知识传递给仅输入窄视窗的学生模型。此外, 针对数据集类别极度不平衡问题, 设计了基于有效样本数的类别平衡交叉熵损失函数以稳定训练过程。在MIT-BIH心律失常数据库上的实验结果表明, 该框架在不增加学生模型推理阶段计算成本的前提下, 将宏平均F1分数由基准模型的94.12%显著提升至95.60%。尤其是对于形态复杂且极易混淆的室上性异常心拍和融合波, 其F1分数分别实现了3.57%和2.15%的绝对提升。

## 关键词

心律失常分类, 知识蒸馏, 未来引导学习, 深度学习

# Research on ECG Arrhythmia Classification Based on Future-Guided Learning

Yusheng Li

School of Mathematics and Statistics, Qingdao University, Qingdao Shandong

Received: March 8, 2026; accepted: April 2, 2026; published: April 9, 2026

## Abstract

Wearable devices for real-time electrocardiogram (ECG) monitoring impose stringent requirements on computational resources and inference latency, typically necessitating the analysis of extremely short, localized signal segments. However, lightweight models operating on such short windows suffer from a performance bottleneck in arrhythmia classification due to the lack of global contextual information before and after the heartbeat. To resolve the conflict between the

**extremely low latency requirement and the absence of local information, this paper proposes an arrhythmia classification method based on future-guided learning. This method constructs an asymmetric teacher-student network architecture. During the training phase, the teacher model utilizes a wide window that asymmetrically extends before and after the heartbeat to capture long-term evolution mechanisms such as compensatory intervals. Subsequently, through cross-temporal knowledge distillation, it transfers structured dark knowledge to the student model, which only receives a narrow window as input. Furthermore, to address the severe class imbalance in the dataset, a class-balanced cross-entropy loss function based on the effective number of samples is designed to stabilize the training process. Experimental results on the MIT-BIH Arrhythmia Database demonstrate that this framework significantly increases the macro-averaged F1-score from 94.12% (baseline model) to 95.60% without adding computational cost during the student model's inference phase. Notably, for morphologically complex and easily confusable supraventricular ectopic beats and fusion beats, absolute improvements of 3.57% and 2.15% in F1-score were achieved, respectively.**

## Keywords

**Arrhythmia Classification, Knowledge Distillation, Future-Guided Learning, Deep Learning**

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

心血管疾病是目前全球范围内发病率和致死率最高的疾病之一，对人类的生命健康构成了极大威胁[1]，心电图作为一种无创、经济且高效的诊断工具，能够客观记录心脏电生理活动，是临床诊断心律失常的核心手段[2]。随着可穿戴设备的普及，开发高效的自动化 ECG 分类算法对实现心脏健康的早期预警与实时监测具有重大意义[3]。近年来，深度学习技术在 ECG 自动化分类领域取得了突破性进展[4]。卷积神经网络(CNN)凭借其强大的局部特征提取能力被广泛应用，但由于其感受野受限，难以有效捕捉长序列中的时序依赖关系[5]。循环神经网络(RNN)虽然擅长时间序列建模，但在处理长程心电信号时常面临梯度消失问题，且计算复杂度较高[6]。而基于自注意力机制的 Transformer 模型虽然能够卓越地建立全局上下文联系[7]，但其庞大的参数量和极高的计算资源消耗使其难以在边缘设备上直接部署[8]。同时，临床实时监测对计算资源和延迟要求极高，模型推断时通常只能截取极短的局部信号片段[9]，然而短窗口轻量级模型因先天缺失这种全局上下文信息，导致其分类性能面临瓶颈[5]。为了满足可穿戴设备的低延迟和低功耗要求，知识蒸馏(Knowledge Distillation, KD)技术被引入模型压缩领域[10]。它通过将复杂教师模型的暗知识转移给轻量级学生模型，在不增加推断延迟的前提下提升了模型性能[11]。然而，传统的知识蒸馏在处理时间序列时，教师与学生通常面对相同的时间窗口，未能充分挖掘时序数据固有的前后向演变信息[12]。为解决极低延迟要求与局部信息缺失之间的矛盾，本文使用了一种基于未来引导学习(Future-Guided Learning, FGL)的心律失常分类方法，突破传统特权信息学习仅利用未来信息的局限[13]，在训练阶段同时融合心跳前后的“过去”与“未来”信息指导学习，增强模型对心律失常演变机制的捕捉能力。

## 2. 数据描述与预处理

### 2.1. 数据集介绍与类别映射

在构建分类模型之前，获取高质量且标准化的输入数据是保证模型性能的基础。本研究采用国际公

认的 MIT-BIH Arrhythmia Database 作为实验数据源。该数据库包含了 48 条时长约为 30 分钟的双通道动态心电图记录, 采样率为 360 Hz。为了使研究结果具备广泛的临床参考价值, 本文遵循美国医疗器械促进协会推荐的分类标准, 将数据库中繁杂的心拍类型重新映射为四大类, 具体包括: 正常心拍及束支传导阻滞(N 类)、室上性异常心拍(S 类)、室性异常心拍(V 类)以及融合波(F 类)[9]。值得注意的是, 本研究最终选择聚焦于 N、S、V、F 四个类别进行分析, 而排除了 Q 类(未分类心跳), 主要原因是 Q 类样本极度稀缺, 排除包含起搏心搏的记录, Q 类在全数据库中仅有 15 次记录, 这种极端的数据稀缺性使得无法从训练集中学习到 Q 类的有效特征, 同时也会导致在测试集上对该类别的性能评估缺乏统计意义[14]。

## 2.2. 信号去噪与基线校正

心电信号作为微弱的生物电信号, 在采集过程中极易受到呼吸起伏、电极滑动导致的基线漂移干扰。基线的非平稳性会严重扭曲波形幅度的统计特性。为了在消除干扰的同时, 最大程度地防止平滑操作抹除可能对病理分类具有判别力的微细形态特征, 如 QRS 波群的切迹或微小的 ST 段改变, 本研究摒弃了可能引入吉布斯现象或相位失真的线性滤波方法, 且未执行低通滤波等高频去噪操作, 而是采用两级非线性中值滤波技术来估算并校正基线[15]。

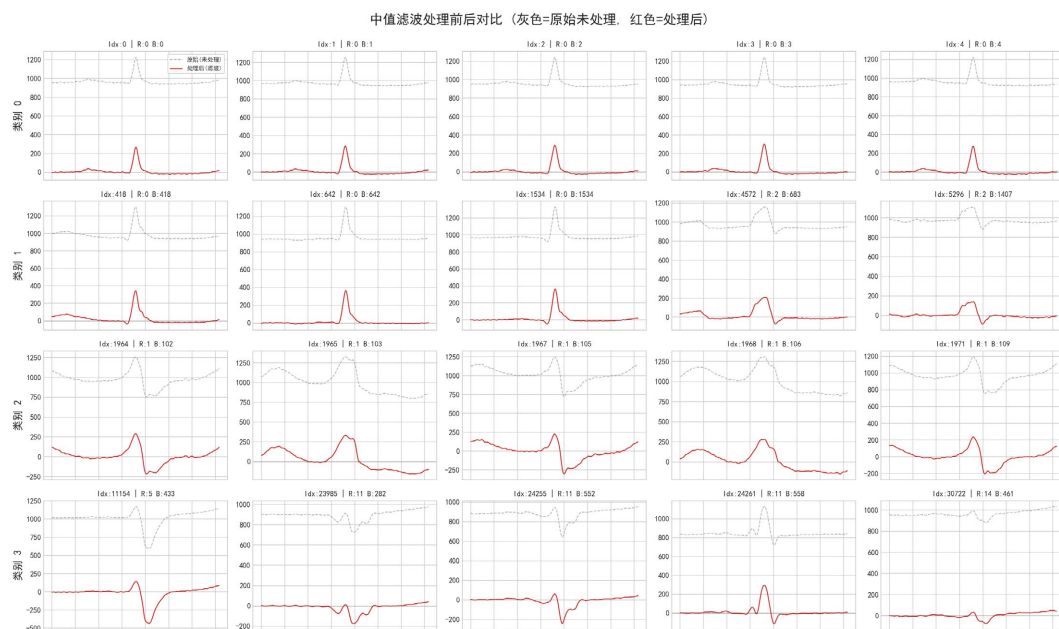
设原始连续 ECG 信号序列为  $X = [x_1, x_2, \dots, x_M]$ 。第一级滤波旨在去除信号中的 QRS 波群等高幅大振幅成分, 保留粗略的基线与低频波 P 波、T 波等。选取局部窗口长度为  $W_1 = 72$ , 对于任意采样点  $i$ , 第一级输出  $y_1(i)$  为局部窗口内数值的中位数:

$$y_1(i) = \text{Median}(x(i - W_1/2), \dots, x(i + W_1/2))$$

第二级滤波在第一级输出的基础上, 采用更宽的窗口  $W_2 = 216$  进一步平滑信号, 以获取准确的基线漂移曲线  $y_2(i)$ :

$$y_2(i) = \text{Median}(y_1(i - W_2/2), \dots, y_1(i + W_2/2))$$

最后从原始信号中减去估算出的基线分量, 得到去基线后的纯净信号  $\tilde{x}(i) = x(i) - y_2(i)$ 。



**Figure 1.** Comparative chart of baseline removal effects by median filtering in ECG signals  
**图 1.** 心电信号中值滤波去除基线效果对比图

图 1 展示了上述基线去除算法在不同类别心拍上的实际处理效果。图中灰色虚线表示原始 ECG 信号，存在显著且不规则的基线漂移，红色实线表示经过两级中值滤波处理后的信号。可以看出，该方法有效地将信号基线校正至零电平附近，消除了低频趋势项的影响，且未改变波形的主体结构。

值得注意的是，与大多数传统预处理流程不同，本研究并未执行高频噪声滤波。这一决策是基于保留信号原始细节的考量，旨在防止平滑操作抹除可能对病理分类具有判别力的微细形态特征。为了验证该策略在保留形态特征方面的有效性，我们将原始信号与处理后信号进行了对齐比较，如图 2 所示。图中灰色虚线为经过均值平移对齐后的原始信号，红色实线为本文预处理后的信号，两者在波形轮廓、峰值位置及局部细节上高度重合。这表明本文采用的中值滤波策略在有效去除基线干扰的同时，未造成明显的波形失真或高频信息丢失，保证了特征提取的真实性。

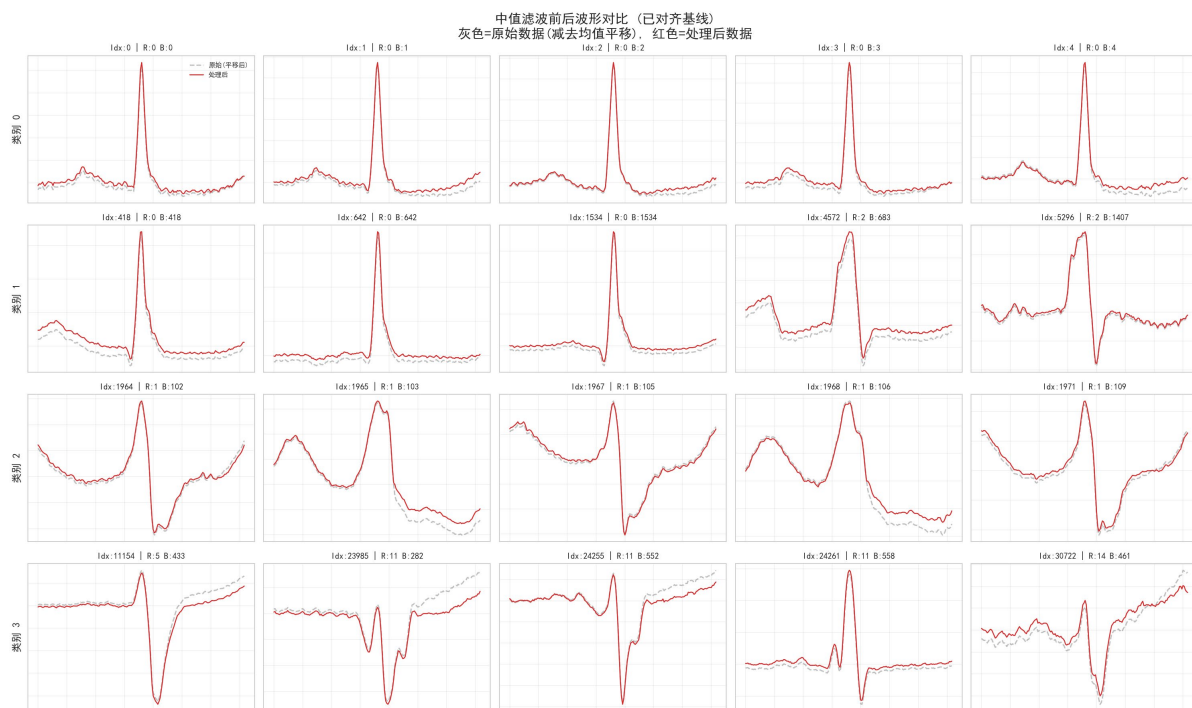


Figure 2. Comparison of waveform morphology preservation before and after preprocessing  
图 2. 预处理前后的波形形态保留情况对比

在完成信号的基线校正后，需将连续的时间序列分割为独立的心拍样本。在实际的心电图自动分析流程中，一个至关重要的预处理步骤是将连续的原始 ECG 信号准确分割为独立的单心搏片段。这一步骤通常依赖于高效的 QRS 波群检测算法，然而为了将研究焦点严格限定于后续分类模型的性能评估，排除因前端心搏分割算法(如传统的 Pan-Tompkins 算法)本身的检测误差而引入的干扰变量，本研究直接采用了 MIT-BIH 心律失常数据库中提供的 R 峰标注位置作为基准。基于这些专家标注信息，本研究采用固定窗口截取法对心拍进行标准化分割，以 R 峰为中心，左右各采集 90 个采样点，确保截取的片段能够完整涵盖一个典型心搏的关键形态学结构。经过预处理与分割后的数据集不仅消除了非平稳趋势的干扰，还最大程度保留了不同病理类型心拍在波形形态上的细微差异，为后续模型性能奠定了坚实的数据基础[16]。

### 3. 方法与模型设计

本章将详细阐述未来引导学习的核心设计，包括非对称教师 - 学生网络架构、双流时序对齐的非对

称窗口构建策略，以及保证知识有效转移的两阶段训练策略。

### 3.1. 非对称教师 - 学生网络架构

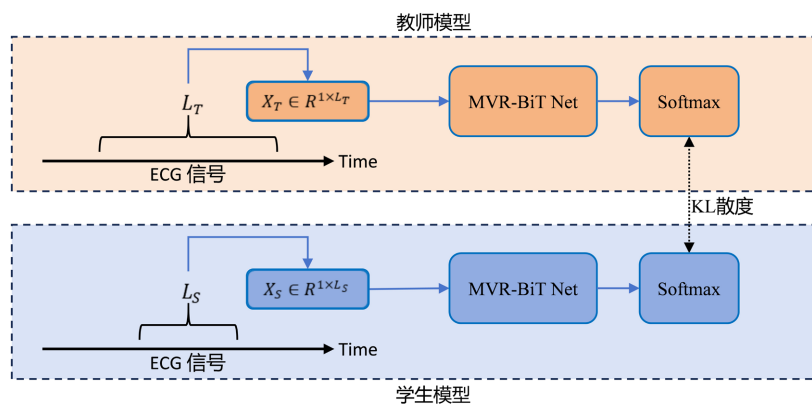
在 FGL 框架中，教师模型与学生模型均使用相同的深度学习网络架构，包含多尺度卷积、通道注意力残差连接及双向 Transformer 三个模块，该架构有效地融合了局部形态与全局节律，在心律失常分类中取得了出色的性能表现。为了保证实验的透明度与模型的可复现性，表 1 详细列出了 FGL 框架中骨干网络的核心超参数配置。教师模型与学生模型虽然共享相同的特征提取骨干网络，但在输入视窗维度与特征映射空间的超参数配置上存在显著差异。这种非对称设计旨在匹配两者所处理的信息熵差异，并构建从高维特征空间向低维流形的蒸馏路径。

**Table 1.** Detailed hyperparameter configuration for the backbone network

**表 1.** 骨干网络详细超参数配置

模块	参数说明	超参数值
多尺度卷积模块	卷积核大小(Kernel Size)	分支 1: 5, 分支 2: 9, 分支 3: 11
	步长(Stride)	1
	填充(Padding)	分支 1: 2, 分支 2: 4, 分支 3: 5
	最大池化(MaxPool1d)	Kernel Size = 2, Stride = 2, Padding = 1
通道注意力残差模块	注意力降维比例(Reduction)	4
	残差块堆叠数量(Blocks)	3
	模块特征通道数	128
双向 Transformer	编码器层数(Num Layers)	3
	嵌入维度	25
	多头注意力头数(Num Heads)	5

如图 3，教师模型被设计为拥有对数据的特权访问权。教师模型的输入序列长度设定为  $L_T$ ， $L_T$  的长度决定了样本充分统计量的丰富程度。 $L_T$  根据实验需求进行调整，覆盖不同的历史或未来范围，以探究最佳的信息边界。相较于传统的单心拍截取，教师视窗将可能完整覆盖了当前心拍及其前序和后继心拍，使其能够直接捕捉代偿间歇、R-R 间期变异等长程依赖特征。



**Figure 3.** Asymmetric teacher-student network architecture diagram

**图 3.** 非对称教师 - 学生网络架构图

其次是高维特征适配, 由于输入维度的增加, 经过多视图卷积层下采样后的特征序列长度显著增长。为了维持 Transformer 自注意力机制的表达力, 本研究相应地提升了教师网络的模型容量, 嵌入维度  $d_{model}^T$  和多头注意力机制的头数  $h^T$  均被设置为较高数值。学生模型模拟真实的临床部署环境, 受限于实时性与计算资源。学生模型的输入序列长度保持为  $L_S = 180$ , 仅聚焦于当前心拍的 QRS 波群形态及邻近波段。

### 3.2. 非对称窗口的构建策略

FGL 的核心假设是教师与学生观测的是同一生物事件的不同视图。因此, 构建严格对齐的数据对  $(x_S, x_T)$  是算法有效的前提。所以本研究重构了数据生成过程, 通过双流采样机制构建训练集。如图 4 所示, 设原始 ECG 记录为时间序列  $X = \{x_1, x_2, \dots, x_N\}$ , 其中包含了  $M$  个标注的心拍。对于第  $i$  个心拍, 其 R 峰位置为  $t_i$ , 真实标签为  $y_i$ 。定义采样函数  $\mathcal{S}(X, t, w)$  为从序列  $X$  中以  $t$  为中心截取长度为  $w$  的片段。则构建的数据对满足:  $x_S^{(i)} = \mathcal{S}(X, t_i, L_S), x_T^{(i)} = \mathcal{S}(X, t_i, L_T)$ 。

在这一机制下, 必然满足集合包含关系  $x_S^{(i)} \subset x_T^{(i)}$ 。在具体实现中, 每次迭代数据加载器会同步返回一个三元组张量  $(x_S^{(i)}, x_T^{(i)}, y_i)$ 。这种严格的逐样本对齐保证了在计算 KL 散度时, 我们是在最小化同一潜在分布在不同观测下的差异, 而非两个无关分布的差异。

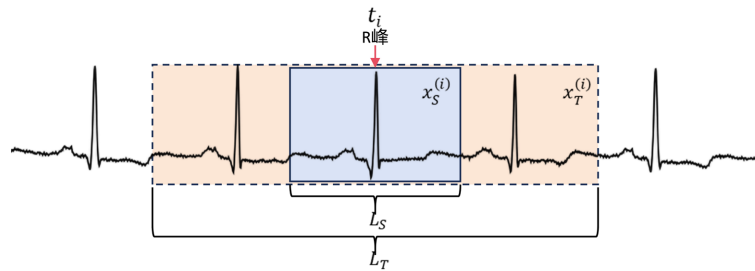


Figure 4. Two-stream temporal alignment sampling diagram  
图 4. 双流时序对齐采样示意图

### 3.3. 两阶段训练策略

为了有效地将教师模型在高维全局视野中捕捉的暗知识转移至低维局部的学生模型, 本研究采用了严格的两阶段训练策略。针对 MIT-BIH 数据库中各心律失常类别分布极度不均衡的难题, 本研究在训练流程中设计了基于有效样本数的类别平衡损失函数。

传统的交叉熵损失容易被多数类的梯度主导。本研究通过有效样本数重构损失权重, 定义类别  $i$  (包含样本数  $N_i$ ) 的原始权重与有效样本数成反比:

$$\alpha_i \propto \frac{1}{E_{N_i}} = \frac{1-\beta}{1-\beta^{N_i}}$$

为保持梯度的数值稳定, 避免学习率失效, 对权重进行归一化处理:

$$w_i = \left( \alpha_i / \sum_{j=1}^K \alpha_j \right) \cdot K$$

将其融入标准交叉熵, 得到类别平衡交叉熵损失:

$$L_{CB}(y, \hat{y}) = -\sum_{i=1}^K w_i y_i \log(\hat{y}_i)$$

在实验中超参数  $\beta$  设为 0.9999。该设计有效抑制了多数类的冗余梯度，合理放大了少数类的权重，在关注少数类与训练稳定性之间取得了最佳平衡。该损失函数将作为核心监督信号贯穿以下训练阶段。

如图 5 所示，训练共分为两个阶段：

### 阶段一：教师模型预训练

此阶段的目标是在全信息条件下获得最优的后验概率估计，将教师模型视为一个独立的分类器，仅使用宽视窗数据  $X_T$  进行监督训练。优化目标为最小化关于参数  $\phi$  的任务损失：

$$\phi^* = \arg \min_{\phi} \frac{1}{M} \sum_{i=1}^M L_{CB} \left( T_{\phi} \left( X_T^{(i)} \right), y_i \right)$$

训练结束后，选取验证集 F1 分数最高的模型权重  $\phi^*$  并将其冻结。此时，教师模型转化为一个确定性的知识生成器。

### 阶段二：学生模型蒸馏训练

此阶段是未来引导学习的核心。学生模型在真实标签和教师软标签的双重监督下进行参数更新。对于每个样本，计算其总体损失：

$$\mathcal{L}_{Total} = \alpha \cdot L_{CB}(Z_S, Y) + (1 - \alpha) \cdot \tau^2 \cdot D_{KL}(\sigma(Z_T/\tau) \parallel \sigma(Z_S/\tau))$$

其中， $Z_T$  与  $Z_S$  分别为教师和学生的 Logits 输出； $\tau$  为蒸馏温度，用于平滑概率分布； $D_{KL}$  为 KL 散度。在反向传播时，仅针对学生模型的参数  $\theta$  计算梯度并更新： $\theta \leftarrow \theta - \eta \cdot \nabla_{\theta} \mathcal{L}_{Total}$ 。通过此跨时域蒸馏，学生模型成功用低维流形近似了教师的高维特征表示。

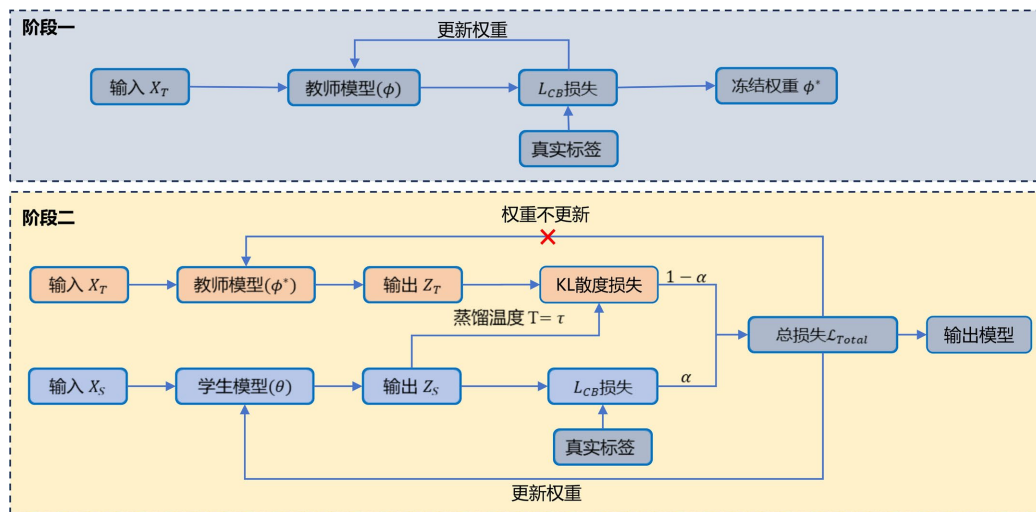


Figure 5. Two-stage training flow diagram  
图 5. 两阶段训练流程图

## 4. 实验结果与参数敏感性分析

### 4.1. 实验结果与模型性能评估

经过系统性的实验，当平衡因子设为  $\alpha = 0.5$ ，蒸馏温度设为  $\tau = 4$ ，且教师模型的视窗构型为 R 峰左侧 90 点，右侧 360 点时，模型取得了最高的宏平均 F1 分数。

表 2 分别展示了三种模型，一是原始的深度学习模型，输入为 180 采样点的单独心拍，二是拥有更大视窗的原始模型，作为未来引导学习的教师模型以及在教师模型指导下，输入为原始采样点的未来引

导学习模型在测试集上的总体性能对比。首先, FGL 取得了显著的性能增益。相较于仅使用硬标签训练的基准模型, 引入未来引导学习机制的 FGL 在所有核心指标上均实现了全面超越。其中, 最关键的评价指标宏平均 F1 分数从 94.12% 提升至 95.60%, 涨幅达 1.48%。考虑到基础模型已处于较高的性能区间, 这一提升幅度在统计显著性上是非常可观的, 证明了引入跨时域的特权信息能有效改善模型的判别边界。此外, 准确率、宏平均精确率和宏平均召回率也分别提升了 0.25%、1.60% 和 1.38%。这表明, FGL 策略在不增加推理阶段计算成本的前提下, 成功将模型的泛化能力推向了新的高度。

**Table 2.** Performance metrics comparison (%)

**表 2.** 性能指标对比(%)

模型	准确率(%)	宏平均精确率(%)	宏平均召回率(%)	宏平均 F1 分数(%)
CNN-Transformer	99.0284	94.2614	93.9993	94.1233
FGL-Teacher	99.1486	93.2176	<b>95.4526</b>	94.2958
FGL-ECG	<b>99.2788</b>	<b>95.8637</b>	95.3820	<b>95.6045</b>

值得深入探讨的是, 尽管 FGL 教师模型的宏平均 F1 分数(94.29%)相较于基准模型(94.12%)的提升幅度有限, 但其指导下的学生模型却实现了 95.60% 的显著飞跃。这一看似反直觉的弱教师指导强学生现象, 其内在机制源于非对称视窗下的信噪比差异与暗知识的正则化效应。

首先, 教师模型右侧高达 360 采样点的宽视窗虽然囊括了代偿间歇等关键的长程节律信息, 但也不可避免地引入了大量与当前心拍形态弱相关的远期节律波动。这种过剩的上下文信息在统计学上构成了背景噪声, 降低了输入信号的信噪比, 导致教师模型在最终决策时容易产生较高的方差, 限制了其硬标签分类性能。

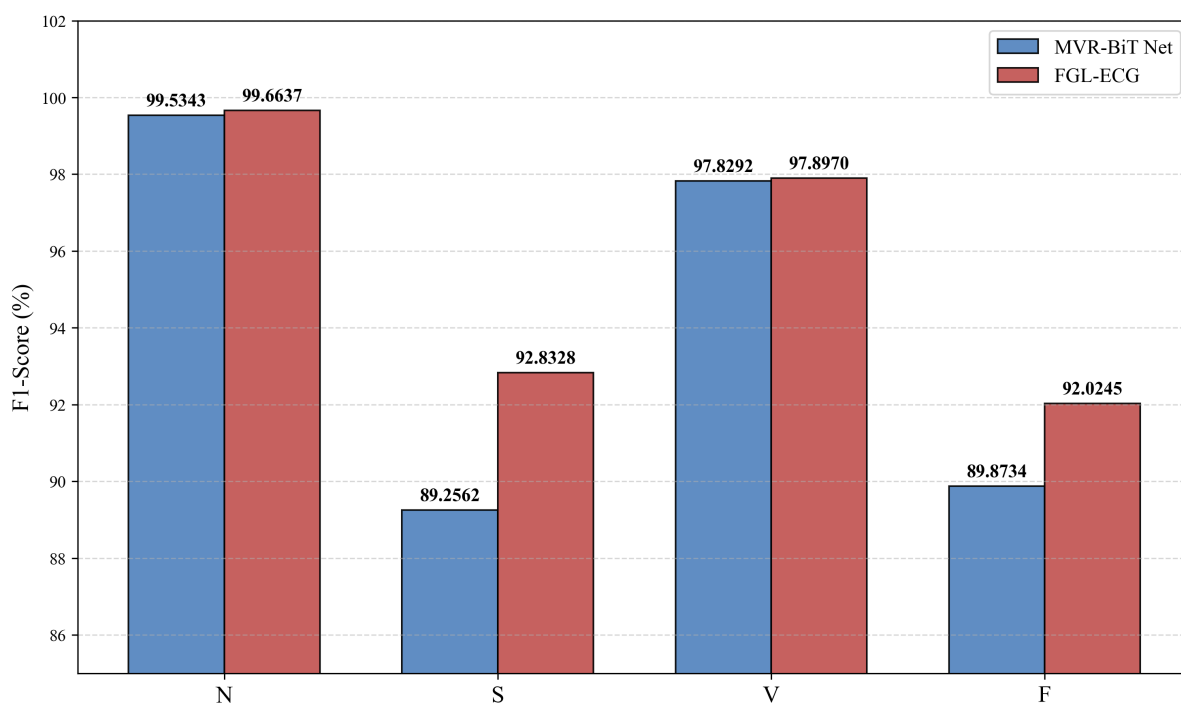
其次, 知识蒸馏的本质并非单纯复制教师的最终决策, 而是转移其输出的连续概率分布。即使教师模型在长程噪声干扰下得出了次优的预测结果, 其软标签依然编码了极具价值的结构化暗知识——即不同类别心律失常之间的拓扑距离与形态相似度。例如, 对于形态极易混淆的融合波, 教师模型提供的平滑分布能够指示其同时具备正常与室性搏动的混合特征。

最终, 学生模型在训练中实现了高信噪比局部特征与全局结构化先验的最优结合。学生模型仅接收窄视窗输入, 避免了远期背景噪声的干扰, 保证了对 QRS 波群局部形态提取的低方差性; 同时, 通过最小化 KL 散度, 学生模型将教师模型从未来视窗中提取的类别关联规律作为一种强大的归纳偏置。这种双重机制使得学生模型既保持了局部判断的精准, 又获得了长程视野的泛化能力, 从而突破了教师模型自身的性能瓶颈。

图 6 进一步展示了 CNN-Transformer 模型与 FGL 在四个具体心律失常类别上的 F1 分数对比。该图直观地揭示了 FGL 机制对不同病理类型的差异化贡献。

首先是对 S 类与 F 类的提升最为显著。S 类 F1 分数由 89.26% 跃升至 92.83%, 绝对提升幅度高达 3.57%。S 类心律失常的诊断高度依赖于 P 波的形态改变及其对后续心律的重置效应。短窗口模型往往难以完整观测到这些长程节律变化, 而 FGL 通过蒸馏教师的未来视野, 成功补全了这一缺失的上下文信息。F 类 F1 分数由 89.87% 提升至 92.02%, 提升了 2.15%。融合心跳兼具正常与室性搏动的特征, 形态极易混淆, FGL 提供的软标签包含了类别间的相似度信息, 帮助模型更精准地界定 F 类与 N/V 类之间的模糊边界。

对于样本量充足且形态特征较为显著的 N 类和 V 类, FGL 保持了与基准模型持平甚至微优的高性能。这表明, 引入未来引导机制并没有导致模型在简单样本上发生灾难性遗忘, 反而通过优化特征空间的整体结构, 微幅提升了对典型异常的识别信心。



**Figure 6.** Comparison of F1-scores across various classes  
**图 6.** 各类别的 F1 分数对比图

综上所述，实验数据有力地支撑了本章的核心假设：通过向拥有全时域视野的教师模型学习，受限视野的学生模型能够显著增强对长程依赖性强、形态复杂的少数类心律失常，特别是 S 类和 F 类的识别能力，从而实现模型整体诊断性能的质的飞跃。

## 4.2. 超参数定义与敏感性分析

为了确保实验的可复现性与分析的严谨性，除通用训练参数(如 Epochs = 60, Batch Size = 128, Learning Rate =  $1e-3$ , Optimizer = Adam)外，本研究引入了三个控制知识蒸馏行为的核心超参数。这些参数直接决定了学生模型在优化过程中的梯度方向与收敛流形。

### 4.2.1. 未来信息量

未来信息量即教师模型的输入序列长度  $L_T$ 。 $L_T$  决定了特权信息量的上界，即教师模型相对于学生模型的费雪信息增量。视窗过短，信息差不足以产生有效的梯度引导；视窗过长，可能引入无关的背景噪声，导致教师信噪比下降。为了系统性地探究教师模型中特权信息的来源及其对蒸馏效果的影响，本文并未简单地设定单一的教师窗口长度，而是设计了对称扩展与非对称未来扩展两组共六种教师视窗构型，以解耦历史信息与未来信息对模型性能的边际贡献。

表 3 展示了在固定学生模型输入左右各 90 点的前提下，六种不同教师视窗构型对最终蒸馏性能的影响，对比了全上下文对称扩展与非对称扩展的实验数据。

对比全上下文对称扩展数据实验的结果趋势，可以发现模型性能并未随着视窗长度的增加而单调上升，从双侧 180 采样点到 270 采样点，宏平均 F1 分数从 94.63% 提升至 95.07%，表明适度增加上下文有助于捕捉局部的节律模式。然而，当视窗进一步扩展至双侧 360 采样点时，性能反而显著下降至 94.29%。这一现象揭示了历史信息的有效性边界，过长的历史视窗虽然增加了数据量，但也引入了大量与当前心拍形态去相关的远期节律波动。在统计上，这些无关的远期历史对于当前分类任务构成了背景噪声，降

低了教师模型后验分布的信噪比，这也验证了为何单纯堆砌更长的输入序列并不总是有效。

**Table 3.** Comparison of performance metrics for different window sizes (%)  
**表 3.** 不同视窗大小性能指标对比(%)

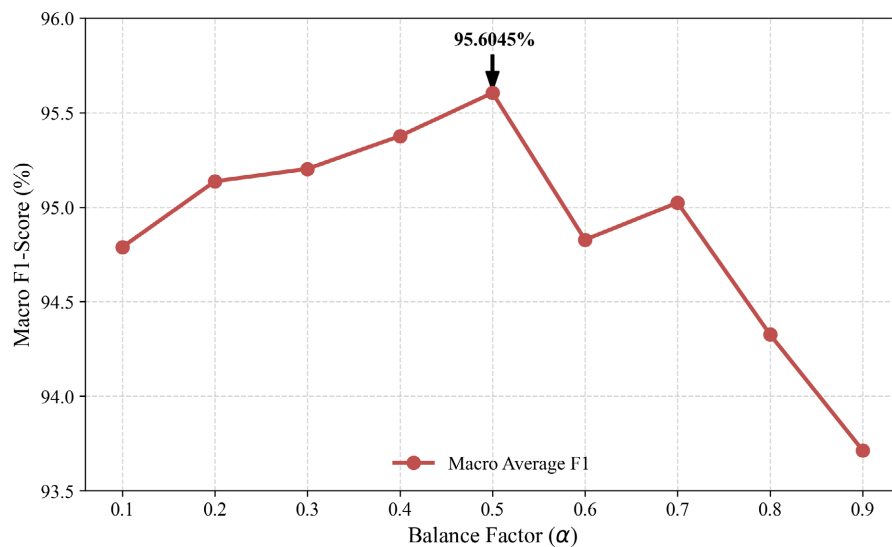
视窗选择	准确率	宏平均精确率	宏平均召回率	宏平均 F1 分数
$T_{sym\_180}$	99.1887	93.4855	<b>95.8608</b>	94.6288
$T_{sym\_270}$	99.2588	94.4725	95.7003	95.0716
$T_{sym\_360}$	99.1687	93.5069	95.2700	94.2904
$T_{asym\_180}$	99.1486	93.7328	95.1980	94.4129
$T_{asym\_270}$	99.1386	93.3855	95.4083	94.3340
$T_{asym\_360}$	<b>99.2788</b>	<b>95.8637</b>	95.3820	<b>95.6045</b>

相比之下，非对称未来引导的数据展示了截然不同的趋势，特别是右侧 360 采样点取得了最高的宏平均 F1 分数(95.60%)，显著优于对应的对称窗口的 94.29%。

在非对称未来引导中，锁定了左侧历史视窗为 90 点，与学生模型一致，仅单向扩展右侧未来视窗。这意味着性能增益完全源于其额外的未来视野。该结果有力地证明了未来信息相比历史信息具有更高的边际信息增益。心律失常的诊断具有显著的非因果性，例如室性早搏后的代偿间歇、或房性早搏后的节律重置，这些关键的判别性证据均位于未来的时间轴上。教师模型通过非对称的构型，精准地捕获了这些高价值的后验证据，并通过知识蒸馏将其转化为学生模型对当前波形的隐式先验知识。

#### 4.2.2. 平衡因子

平衡因子  $\alpha$  是损失函数中用于调节真实标签监督与未来引导权重的系数。较大的  $\alpha$  倾向于拟合当前数据的经验分布，即低偏差但可能过拟合；较小的  $\alpha$  则强迫学生模仿教师的平滑分布，引入归纳偏置以降低方差，但过度依赖可能导致对真实标签的欠拟合。本研究将在区间[0.1, 0.9]内进行网格搜索，步长为 0.1。



**Figure 7.** Balance factor performance comparison chart  
**图 7.** 平衡因子性能对比图

图 7 展示了宏平均 F1 分数随平衡因子  $\alpha$  变化的轨迹。实验结果表明，模型性能并未随  $\alpha$  单调递增或递减，而是呈现出显著的非单调倒 U 型特征，并在  $\alpha = 0.5$  处达到全局极大值。

当  $\alpha$  增大至 0.8 或 0.9 时，Macro-F1 分数出现急剧下滑。此时损失函数被交叉熵项主导，KL 散度项的权重趋近于零。这意味着学生模型逐渐忽略了教师提供的软标签指导，退化为仅依赖短窗口输入的基准模型，导致其在处理复杂样本时的泛化能力下降。当  $\alpha$  减小至 0.1 时，性能同样表现不佳，此时模型过度依赖教师的概率分布。虽然教师模型拥有更宽的视野，但其预测本身也包含贝叶斯误差和估计噪声。过强的蒸馏权重使得教师的预测不确定性被过度传递给学生，导致学生模型对真实标签的拟合程度不足，产生欠拟合现象。

在  $\alpha = 0.5$  时，模型实现了性能峰值。这表明，真实标签提供的硬约束与教师提供的软平滑在统计上构成了最佳的互补关系。硬标签确立了分类的决策边界骨架，而软标签填充了边界附近的概率密度细节，两者共同作用，实现了最佳权衡。

进一步观察各类别的详细性能曲线，图 8 各个子图，我们可以发现不同心律失常类型对  $\alpha$  的敏感度。

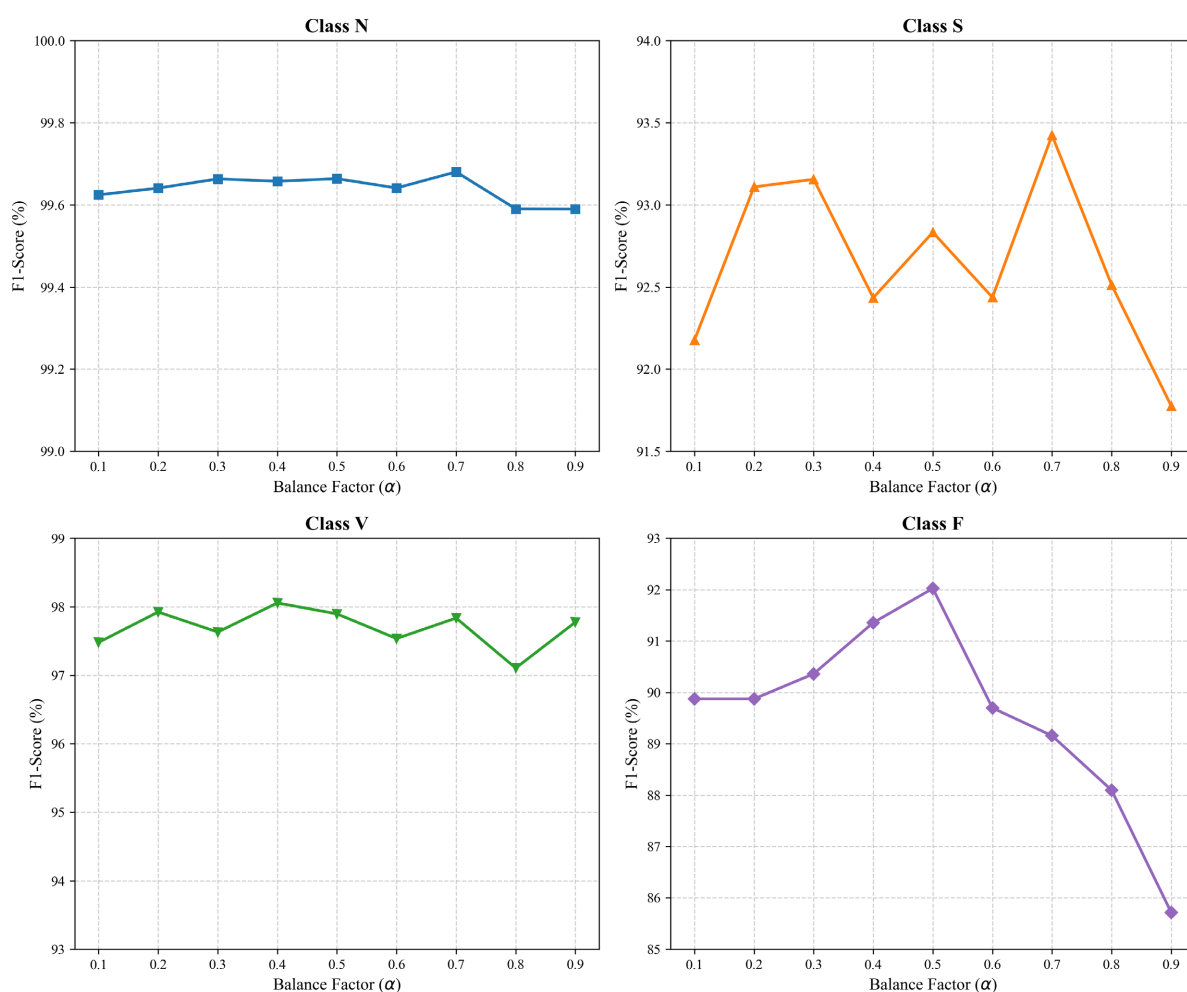


Figure 8. Balance factor performance comparison chart by category

图 8. 平衡因子各类别性能对比图

F 类的曲线表现出最剧烈的波动。在  $\alpha = 0.9$  时，F 类性能大幅崩塌至 86% 以下，而在  $\alpha = 0.5$  时达到

峰值。这源于 F 类是形态最复杂、样本最少的类别，且兼具 N 类与 V 类的特征。仅靠短窗口的硬标签监督，模型极易陷入局部极小值。教师模型通过软标签提供的暗知识，有效地为学生模型指明了优化方向。

### 4.2.3. 蒸馏温度

蒸馏温度  $\tau$  是 Softmax 函数中的缩放参数，用于调整概率分布的平滑程度。 $\tau$  决定了输出概率分布的香农熵。当  $\tau \rightarrow 0$  时，分布坍缩为狄拉克分布，仅保留最大概率类别信息。当  $\tau > 1$  时，分布被软化，使得非最大概率类别的概率值显现。这些次要概率携带了类别间的隐含相关性。本文在离散集合  $\{1, 2, 3, 4, 5, 6, 7, 8\}$  中进行寻优。

图 9 清晰地展示了模型宏平均 F1 分数随蒸馏温度  $\tau$  变化的轨迹。实验结果表明，蒸馏效率对温度参数高度敏感，呈现出先升后降的非单调趋势，并在  $\tau = 4$  处达到全局最优。这一现象揭示了软标签中的信息熵对知识迁移质量的决定性作用。

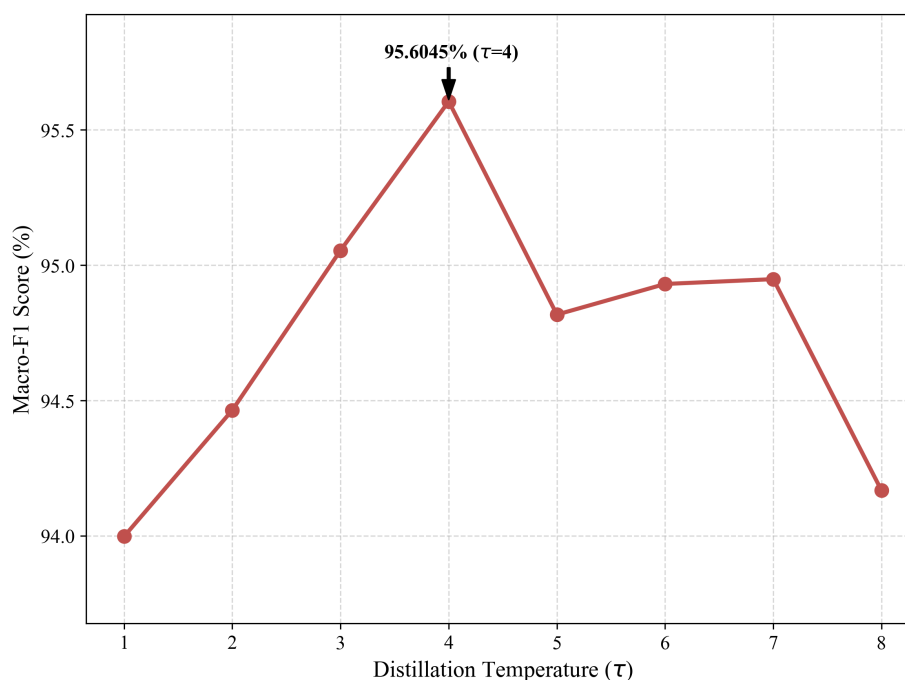


Figure 9. Distillation temperature performance comparison chart

图 9. 蒸馏温度性能对比图

当  $\tau = 1$  时，宏平均 F1 分数为 94.00%，此时模型的性能增益最小。在低温下，教师模型的 Softmax 输出分布极其尖锐，趋近于狄拉克分布。这意味着概率质量几乎全部集中在最大值对应的类别上，非目标类别的概率值接近于零。此时，软标签退化为硬标签，教师模型未能传递关于类别间相似度的额外信息，导致 KL 散度项的梯度贡献仅相当于重复的强监督，无法提供有效的正则化增益。当  $\tau$  增加至 8 时，性能出现显著下滑，随着温度过度升高，Softmax 分布逐渐被拉平，趋向于均匀分布。此时，不同负类之间的概率差异被抹平，教师输出的结构化信息被高熵噪声淹没。在  $\tau = 4$  时，模型性能达到峰值。此温度设定使得 Softmax 分布处于“软硬适中”的状态。一方面，它保留了目标类别的显著优势，确保了分类的准确性；另一方面，它充分放大了非目标类别的概率尾部。这种概率分布隐含了该样本虽为 S 类，但在形态上更接近 N 类而非 V 类的知识。 $\tau = 4$  正是最大化这种类别间相关性传递效率的最佳状态。

观察图 10 的各个子图，可以发现不同心律失常类型对温度  $\tau$  的响应模式存在显著差异，这进一步印

证了 FGL 机制的生效机理。

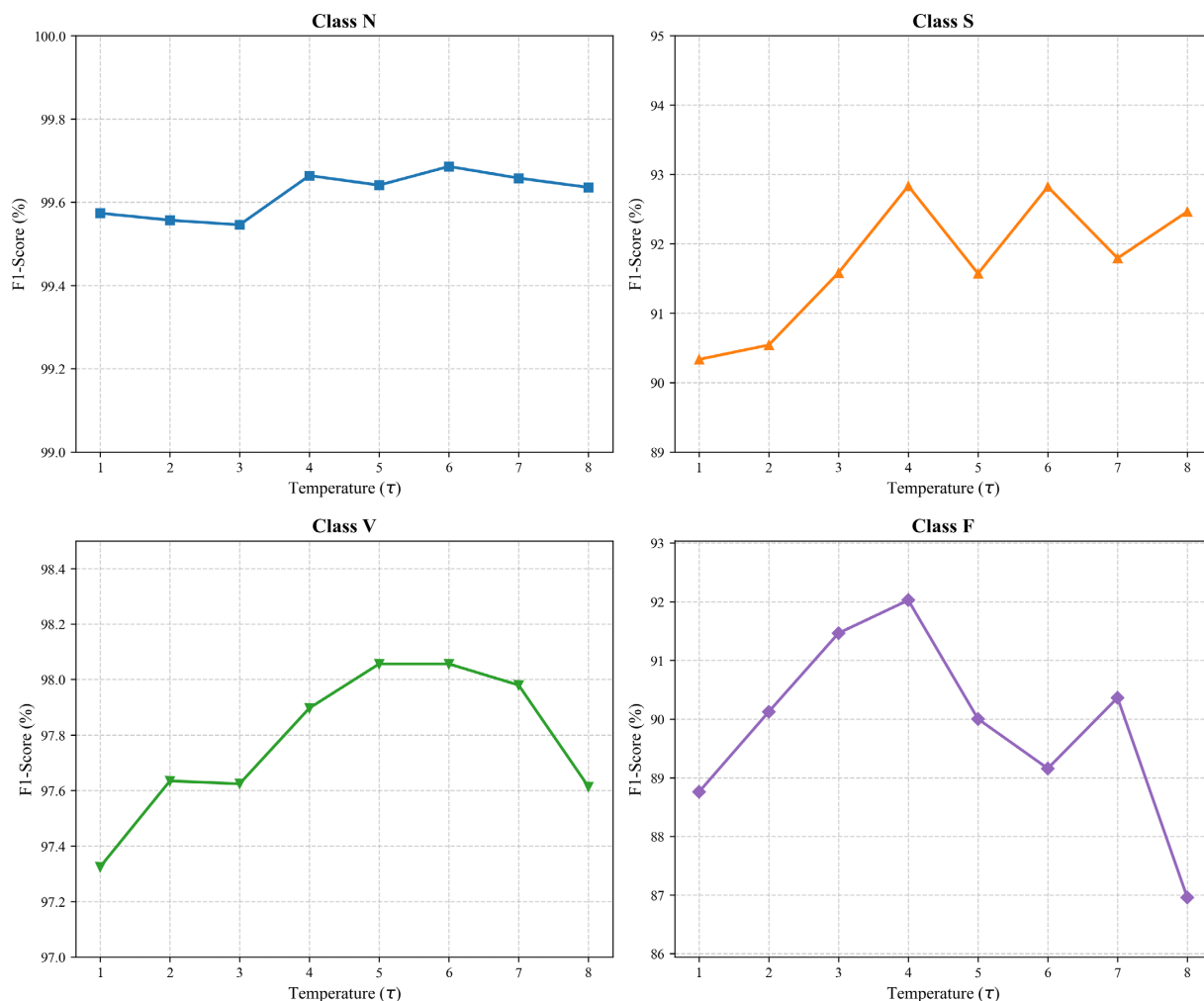


Figure 10. Distillation temperature performance comparison chart by category

图 10. 蒸馏温度各类别性能对比图

F 类对温度最为敏感，在  $\tau=1$  时，F1 分数仅为 88.76%；而在  $\tau=4$  时，飙升至 92.02%，随后在  $\tau=8$  时迅速回落至 86.96%。这源于融合心跳本质上是正常心跳与室性心跳的电生理叠加，其波形特征天然具有模糊性。在低温度下，教师模型给出的非黑即白判断无法描述这种中间态特征；而在  $\tau=4$  时，教师通过软标签准确传递了 F 类与 N、V 类之间的纠缠关系。

综上所述，选取  $\tau=4$  作为最终的蒸馏温度，是在保留判别力与揭示潜在结构之间取得的最优解。该设置使得教师模型能够最有效地将其在长视窗下观测到的、关于心律失常类别间内在联系的先验知识，通过 KL 散度无损地迁移至学生模型，从而显著提升了模型对复杂样本的泛化能力。

## 5. 总结

本文针对轻量级深度学习模型在极短推断窗口下因全局上下文缺失而导致的分类性能瓶颈，提出并实现了一种基于未来引导学习的心电图心律失常分类框架。研究证明，非对称时序视窗有效突破了局部信息壁垒，相比于对称的历史信息扩展，包含更多未来信息的非对称视窗能带来更高的边际信息增益，

能够精准捕获心律失常中至关重要的非因果性特征。同时，通过将优化过程解耦为特权特征学习与跨时域蒸馏两个子问题，学生模型能够有效利用教师模型输出的软标签。结合类别平衡交叉熵损失函数，FGL框架不仅在宏平均 F1 分数上实现了从 94.12%到 95.60%的提升，更在最具挑战性的 S 类和 F 类心拍上取得了突破性进展。

## 参考文献

- [1] Roth, G.A., Mensah, G.A., Johnson, C.O., *et al.* (2020) Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update from the GBD 2019 Study. *Journal of the American College of Cardiology*, **76**, 2982-3021.
- [2] Hannun, A.Y., Rajpurkar, P., Haghpanahi, M., Tison, G.H., Bourn, C., Turakhia, M.P., *et al.* (2019) Cardiologist-Level Arrhythmia Detection and Classification in Ambulatory Electrocardiograms Using a Deep Neural Network. *Nature Medicine*, **25**, 65-69. <https://doi.org/10.1038/s41591-018-0268-3>
- [3] Perez, M.V., Mahaffey, K.W., Hedlin, H., Rumsfeld, J.S., Garcia, A., Ferris, T., *et al.* (2019) Large-Scale Assessment of a Smartwatch to Identify Atrial Fibrillation. *New England Journal of Medicine*, **381**, 1909-1917. <https://doi.org/10.1056/nejmoa1901183>
- [4] Strodthoff, N., Wagner, P., Schaeffter, T. and Samek, W. (2021) Deep Learning for ECG Analysis: Benchmarks and Insights from PTB-XL. *IEEE Journal of Biomedical and Health Informatics*, **25**, 1519-1528. <https://doi.org/10.1109/jbhi.2020.3022989>
- [5] Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L. and Muller, P. (2019) Deep Learning for Time Series Classification: A Review. *Data Mining and Knowledge Discovery*, **33**, 917-963. <https://doi.org/10.1007/s10618-019-00619-1>
- [6] Ribeiro, A.H., Ribeiro, M.H., Paixão, G.M.M., Oliveira, D.M., Gomes, P.R., Canazart, J.A., *et al.* (2020) Automatic Diagnosis of the 12-Lead ECG Using a Deep Neural Network. *Nature Communications*, **11**, Article No. 1760. <https://doi.org/10.1038/s41467-020-15432-4>
- [7] Natarajan, A., Chang, Y., Mariani, S., Rahman, A., Boverman, G., Vij, S., *et al.* (2020). A Wide and Deep Transformer Neural Network for 12-Lead ECG Classification. *Computing in Cardiology Conference (CinC)*, Vol. 47, 1-4. <https://doi.org/10.22489/cinc.2020.107>
- [8] Lin, T., Wang, Y., Liu, X. and Qiu, X. (2022) A Survey of Transformers. *AI Open*, **3**, 111-132. <https://doi.org/10.1016/j.aiopen.2022.10.001>
- [9] 甘屹, 施俊丞, 高丽, 何伟铭. 基于深度学习并行网络模型的心律失常分类方法[J]. 南方医科大学学报, 2021, 41(9): 1296-1303.
- [10] 纪荣嵘, 林绍辉, 晁飞, 吴永坚, 黄飞跃. 深度神经网络压缩与加速综述[J]. 计算机研究与发展, 2018, 55(9): 1871-1888.
- [11] 邵仁荣, 刘宇昂, 张伟, 王骏. 深度学习中知识蒸馏研究综述[J]. 计算机学报, 2022, 45(8): 1638-1673.
- [12] Gou, J., Yu, B., Maybank, S.J. and Tao, D. (2021) Knowledge Distillation: A Survey. *International Journal of Computer Vision*, **129**, 1789-1819. <https://doi.org/10.1007/s11263-021-01453-z>
- [13] Vapnik, V. and Vashist, A. (2009) A New Learning Paradigm: Learning Using Privileged Information. *Neural Networks*, **22**, 544-557. <https://doi.org/10.1016/j.neunet.2009.06.042>
- [14] Sellami, A. and Hwang, H. (2019) A Robust Deep Convolutional Neural Network with Batch-Weighted Loss for Heartbeat Classification. *Expert Systems with Applications*, **122**, 75-84. <https://doi.org/10.1016/j.eswa.2018.12.037>
- [15] deChazal, P., O'Dwyer, M. and Reilly, R.B. (2004) Automatic Classification of Heartbeats Using ECG Morphology and Heartbeat Interval Features. *IEEE Transactions on Biomedical Engineering*, **51**, 1196-1206. <https://doi.org/10.1109/tbme.2004.827359>
- [16] Kiranyaz, S., Ince, T. and Gabbouj, M. (2016) Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks. *IEEE Transactions on Biomedical Engineering*, **63**, 664-675. <https://doi.org/10.1109/tbme.2015.2468589>