

# 基于改进交叉熵损失函数与Transformer的心电信号高风险分类研究

李嘉言

北京工商大学数学与统计学院, 北京

收稿日期: 2026年5月18日; 录用日期: 2026年6月22日; 发布日期: 2026年6月30日

## 摘要

在心血管疾病早期筛查过程中, 能够有效检测出心电信号中高风险部分对于防止患者发生心跳骤停有重大作用, 在此背景下, 我们利用PhysioNet带有高风险标签的心电小片段数据集(EGC Fragment Database with High-Risk Labels, v1.0.0)进行研究, 提出了一种基于随机森林特征选择以及 Transformer 分类器的心电图高风险分类算法同时提出一种新的交叉熵损失函数ECGLoss。此损失函数包括焦点损失以及置信度惩罚两部分: 焦点损失通过对权重放大难以区分样本梯度影响, 置信度惩罚给低置信度正确分类增加约束, 使网络更关注边缘样本分类。使用5折分层交叉验证评价此模型性能, 在ECGLoss与普通交叉熵损失之间比较两者分类准确率及收敛速度。实验结果显示, 提出的ECGLoss比普通的交叉熵损失有更高的分类准确率, 说明此损失函数对于心电信号高危分类是有效的。

## 关键词

心电信号, 高风险分类, Transformer, 交叉熵损失, 焦点损失, 随机森林

# Transformer-Based High-Risk ECG Signal Classification with an Enhanced Cross-Entropy Loss

Jiayan Li

School of Mathematics and Statistics, Beijing Technology and Business University, Beijing

Received: May 18, 2026; accepted: June 22, 2026; published: June 30, 2026

## Abstract

In the early screening of cardiovascular diseases, the effective detection of high-risk segments

within electrocardiogram (ECG) signals plays a crucial role in preventing sudden cardiac arrest in patients. Against this backdrop, we conduct our study using the PhysioNet ECG Fragment Database with High-Risk Labels (v1.0.0), a dataset comprising short ECG segments annotated with high-risk labels. We propose an ECG high-risk classification algorithm that integrates random forest-based feature selection with Transformer classifier, and simultaneously introduce a novel cross-entropy loss function termed ECGLoss. This loss function consists of two components: focal loss and a confidence penalty. The focal loss component amplifies the gradient impact of difficult-to-classify samples through weight scaling, while the confidence penalty imposes an additional constraint on correctly classified samples with low confidence, thereby compelling the network to focus more on borderline sample classification. The model performance is evaluated using five-fold stratified cross-validation, with ECGLoss and the standard cross-entropy loss compared in terms of both classification accuracy and convergence speed. Experimental results demonstrate that the proposed ECGLoss achieves higher classification accuracy than the conventional cross-entropy loss, indicating that this loss function is effective for high-risk classification of ECG signals.

## Keywords

ECG Signal, High-Risk Classification, Transformer, Cross-Entropy Loss, Focal Loss, Random Forest

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

### 1.1. 研究背景与意义

心血管疾病是危害人类生命健康的常见病、多发病,在全世界范围内都是一个严重的公共卫生问题,患病人数及死亡率较高。据世界卫生组织报道,每年有大约 1790 万人死于心血管疾病,占全部死亡人数的 32%。而心电图(Electrocardiogram, ECG) [1]是用于监测心脏电生理活动的方法,也是检测心血管疾病的重要工具。

伴随着医学信息化的发展,大量 ECG 数据的产生也给人工阅片带来极大负担。传统的电生理检查需要依靠医师的经验以及临床知识进行判读,易受到人为因素的影响并且耗时较长。因此采用计算机辅助的方法对 ECG 信号进行分类及识别有助于提升工作效率,减少病人就医费用。

PhysioNet [2]提供的带有高风险标注的心电图片段数据库(ECG Fragment Database with High-Risk Labels, v1.0.0)是建立自动高风险心电图检测模型的良好基础数据集。正确地检测出高风险心电图片段可以尽早发现心脏问题,在临床上起到积极的作用。

### 1.2. 国内外研究现状

对于心电信号自动分类,在心电信号自动分类中,目前的研究工作主要是针对特征选择以及分类器的研究工作[3]。在特征选择上,传统的方法使用时间域特征(例如 R-R 间距, QRS 宽度等)、频率域特征(例如功率谱密度)以及时频域特征(例如小波变换系数等),而近年来基于深度学习的方法[4]得到越来越多的关注。而在分类器的设计上,传统的支持向量机(SVM)、随机森林等方法由于其良好的可解释性仍然被广泛应用于 ECG 分类中;一维卷积神经网络(1D-CNN) [5]、长短期记忆网络(LSTM)以及 Transformer 等深度学习的方法[6]应用于时序信号领域[7],也表现出良好效果。

从损失函数来看，通常使用标准交叉熵作为分类问题的损失函数，在心电图分类等医疗相关问题上由于存在类别不平衡以及边界样本难以区分等问题导致其效果不佳，而焦点损失(Focal Loss)通过引入一个因子来减少容易被正确分类的样本的影响从而更关注那些不易被正确分类的样本，但是目前研究较少结合心电图自身特点进行损失函数的设计。

### 1.3. 本文主要工作

针对目前的研究缺陷，本文所做的主要贡献为：

- (1) 提出一种利用随机森林进行特征选择以及基于 Transformer 实现 ECG 高风险检测的方法；
- (2) 提出一种改进的交叉熵损失函数 ECGLoss，包含焦点损失部分以及置信度惩罚部分；
- (3) 在 Physionet ECG 高风险标签数据集上进行消融实验，在分类准确率及收敛速度方面对比 ECGLoss 与其他常用损失函数的效果。

## 2. 相关理论与技术基础

### 2.1. 数据集介绍

本文采用的数据集是来自 PhysioNet 的带有高风险标签的心电图片段数据(ECG Fragment Database with High-Risk Labels, v1.0.0)。此数据集共有 537 段心电信号，其中高风险样本有 337 个(占 62.8%)，低风险样本有 200 个(占 37.2%)，每个样本有 21 个数值型特征以及一个二分类标签。见表 1。

Table 1. Dataset class distribution statistics

表 1. 数据集类别分布统计

类别	样本数量	占比
高风险	337	62.8%
低风险	200	37.2%
合计	537	100%

### 2.2. 随机森林特征选择

随机森林(Random Forest, RF)是 Bagging 的思想而设计的一种集成方法，它建立许多棵决策树，然后取所有决策树投票结果中得票最多的一个作为最终分类结果。随机森林可以给出每个特征的重要性评分(Gini Importance)，这个分数是根据特征对于决策树节点划分所起到的作用大小决定的。我们用这个分数对原特征进行排序选择，选取排名最高的  $K$  个特征作为 Transformer 分类器的输入，从而减少维度以及消除冗余。

### 2.3. Transformer 模型

Transformer 是 Vaswani 等人提出的[8]，其主要思想是自注意力机制[9]，在给定输入序列的情况下，利用三个不同的可学习矩阵将它们分别映射成查询  $Q$ 、键  $K$  和值  $V$  后进行计算得到注意力输出：

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (1)$$

为了捕获不同的子空间信息，Transformer 使用了多头注意机制，将  $Q$ 、 $K$ 、 $V$  分别投影到  $h$  个不同的子空间中并行计算后再进行拼接。编码器有  $L$  层，每一层都是相同的结构，即有多头自注意力子层以

及前馈网络子层，它们之间均有残差连接并且都使用了层归一化：

$$Z^{(l)} = \text{LayerNorm}\left(Z^{(l-1)} + \text{MultiHead}\left(Z^{(l-1)}, Z^{(l-1)}, Z^{(l-1)}\right)\right) \quad (2)$$

本论文对选取  $K$  维特征进行线性投射降维至维表示，送入 Transformer 编码器学习深层次信息，再通过池化得到结果交予线性分类器完成分类任务[10]。

## 2.4. 损失函数理论基础

交叉熵损失是常见的分类损失函数，在二分类问题中，假设我们得到模型对于正类预测的概率值，则交叉熵损失可以表示为：

$$L_{\text{CE}} = -\log(p_i) \quad (3)$$

焦点损失(Focal Loss)是在普通的交叉熵基础上增加一个调节因子，弱化容易分类样本的影响，让网络更多地关注难以区分样本。

$$L_{\text{FL}} = -\alpha(1-p_i)^\gamma \log(p_i) \quad (4)$$

其中， $\alpha$  是类别平衡因子， $\gamma$  是焦点参数。当  $\gamma=0$  时，Focal Loss 变成普通的交叉熵损失。

## 3. 改进的交叉熵损失函数设计

### 3.1. 标准交叉熵损失的局限性分析

标准交叉熵损失应用于心电信号高风险分类存在如下问题：(1) 类别不平衡——低风险样本(200个，占 37.2%)远远小于高风险样本(337个，占 62.8%)，而标准交叉熵对待两类同样重要，因此容易造成模型偏好多数类，对于稀有类无法很好区分；(2) 难易样本同等对待——只需要让正确的类别的概率大于等于 0.5 就可以完成分类，没有必要让其更有把握；(3) 边界样本难以区分——高风险和低风险之间有一个过渡地带，在这个地带的标准交叉熵并不能很好地关注到这些重要的样本。

### 3.2. ECGLoss 损失函数设计

为了解决以上问题，我们提出了 ECGLoss，由焦点损失项以及置信度惩罚项构成：

$$L_{\text{ECG}} = \alpha(1-p_i)^\gamma (-\log p_i) + \lambda \max(\tau - p_i, 0)^2 \quad (5)$$

第一项是焦点损失项，用一个因子  $(1-p_i)^\gamma$  来调整每个样本的重要性，对于  $p_i$  接近 1 容易分类的样本大幅度降低它们的损失贡献，而对于  $p_i$  远小于 1 难以分类的样本保留原来的梯度，从而使网络关注边缘模糊的样本学习。第二项是置信度惩罚项，在正确类别的概率低于一定阈值  $\tau$  此时会产生二次惩罚  $\lambda(\tau - p_i)^2$ ，促使模型提高“勉强正确”样本的预测置信度；而当  $p_i \geq \tau$  时该项为零。两者相辅相成，焦点损失项解决样本间难易不平衡问题，置信度惩罚项解决样本内置信度问题。

### 3.3. 超参数分析

ECGLoss 具有 4 个超参数，每个参数的作用以及推荐设置见表 2。

## 4. 实验设计与方法

### 4.1. 数据预处理与特征筛选

实验先对待使用的数据集进行预处理：

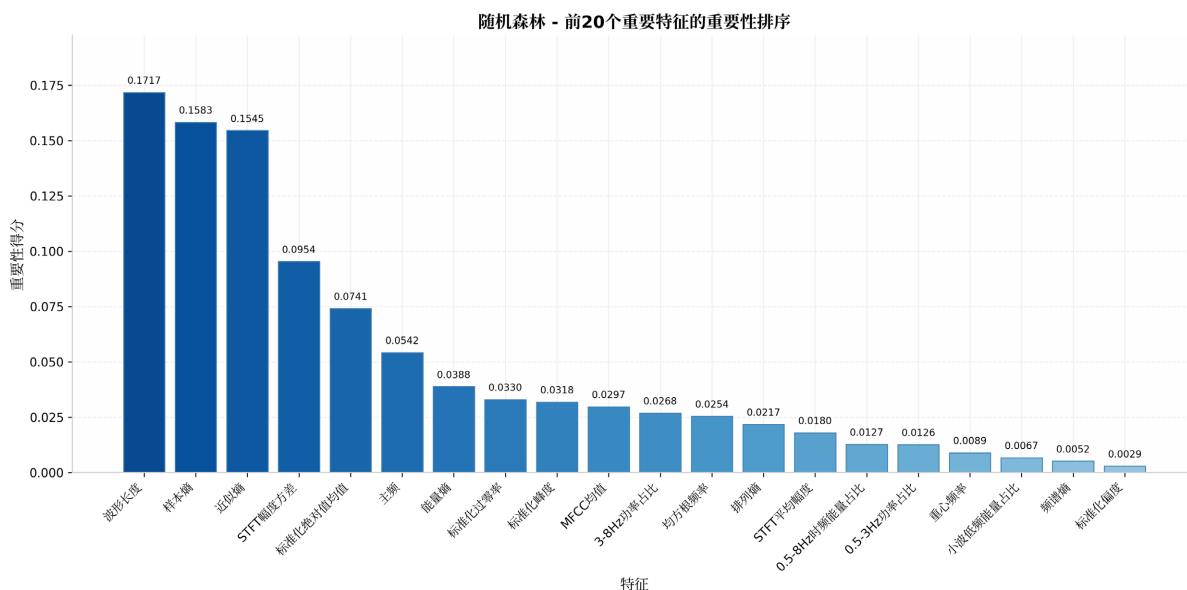
对于每一个特征都进行了 Z-score 归一化，使得它们的均值为 0，方差为 1。

**Table 2.** ECGLoss hyperparameter description  
**表 2.** ECGLoss 超参数说明

参数	符号	说明
焦点参数	$\gamma$	控制易分类样本衰减, 建议值 2
平衡因子	$\alpha$	调和正负样本权重, 取值 0.25
置信度阈值	$\tau$	期望最低置信度, 取值 0.70
惩罚权重	$\lambda$	置信度惩罚项权重, 取值 0.10

$$x' = \frac{x - \mu}{\sigma} \tag{6}$$

采用 5 折分层交叉验证划分数据集保证每一折内各类别比例与整体数据集相同; 基于全部特征进行随机森林(n\_estimators = 100)训练之后根据 Gini 重要性由大到小选取最前面的 20 个特征。前五类重要指标如图 1 所示:



**Figure 1.** Random forest feature importance ranking (Top 20)

**图 1.** 随机森林特征重要性排序(Top-20)

1. 波形长度表示的是一个信号在一定时间内所包含“量”。脑电信号的波形长度越大, 则说明在这个时间段内有较多神经元同时进行兴奋, 而且兴奋程度较高。例如, 在癫痫发作之前, 大量神经元在短时间内大量同步放电, 因此其波形长度也会相应增长, 所以它能够位居第一是合乎情理的。

2. 样本熵是用于描述一个信号的复杂性和随机性的指标, 在正常情况下人的脑电一般都具有一定的复杂性, 而在疾病状态下(如脑损伤或者癫痫等)脑电波可能会过为整齐(异常同步)或者过于杂乱无章, 样本熵就可以检测出这样的差异; 同时相比于近似熵而言, 样本熵对于数据量的需求较低并且更加稳定。

3. 近似熵与样本熵一样都是用来表征信号复杂性的量, 但是计算方式不同。它是表示脑电信号在不同时间尺度下的自相似性。正常情况下, 各个脑区之间电位变化是“有规律但是无法预见”, 即有序但不可预测的状态, 此时近似熵较大; 而某些病理情况下脑电波出现单调、简单重复的现象, 则近似熵

较小。

4. STFT 幅度方差表示的是脑电信号频谱能量变化情况。人处于不同的意识水平下(例如静息闭眼、注意集中、睡眠各个时相中), 各个频率段的能量并不是固定不变的, 而是在时刻发生变化, 在 STFT 幅度方差较大时, 说明脑电的频谱特性变化较快, 可能是由于认知任务的变化或者疾病导致的。

5. 标准化绝对值均值即为信号“平均强度”。它是与整个脑电信号的能量有关, 在很大程度上受到大脑皮层兴奋性的影响。该指标较高一般表示皮层神经元兴奋性较高; 较低一般表示皮层神经元兴奋性受到抑制, 如在睡眠或者服用一些镇静剂情况下会出现低的标准差绝对值均值。这五个参数可以从不同角度反映脑电活动的能量、时间以及频率的一致性等方面的信息, 结合起来就可以很好地对各种不同的情况加以区分。

## 4.2. 模型结构与训练设置

本文所使用的 Transformer 分类器结构见表 3, 训练参数详见表 4。模型把筛选出的 20 维特征向量通过一个线性变换映射到  $= 32$  维, 作为单序列 token; 在训练过程中向输入添加均值为 0 标准差为 0.01 的高斯噪声防止过拟合, 然后让这些特征经过一层 Transformer 编码器学习深层次信息之后再通过一个线性层进行二分类。

**Table 3.** Transformer classifier structure

**表 3.** Transformer 分类器结构

层	参数/操作	说明
Input Noise	$\sigma = 0.01$	训练时对输入添加高斯噪声
Linear Embedding	$20 \rightarrow 32$	特征向量线性投影为 token
Transformer Encoder x1	$d_{\text{model}} = 32$ , nhead = 2, dropout = 0.5, norm_first = True	多头自注意力 + FFN + 残差连接 + 层归一化
Mean Pooling	—	序列维度取平均
Classifier	$32 \rightarrow 2$	线性层输出类别 logits

**Table 4.** Model training hyperparameter settings

**表 4.** 模型训练超参数设置

超参数	设置值	说明
优化器	Adam	—
学习率	$5 \times 10^{-4}$	—
批大小	16	—
最大训练轮数	80	—
早停耐心值	10	训练损失连续 10 轮未改善则停止
$d_{\text{model}}$	32	模型隐层维度
nhead	2	多头注意力头数
num_layers	1	Transformer 编码器层数
dropout	0.5	Dropout 比例
输入噪声	$\sigma = 0.01$	仅训练时生效

### 4.3. 消融实验设计

为了检验 ECGLoss 是否有效,在此基础上进行消融实验比较使用 ECGLoss 还是传统的交叉熵损失 (CrossEntropyLoss)。除了损失函数不同之外,其余所有条件都相同,例如网络结构、超参数、随机种子以及数据集划分等。评估指标有:(1) 准确率 Accuracy 和 F1;(2) 收敛速度即网络获得最优的测试结果所需要的迭代次数。

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

$$\text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (8)$$

## 5. 实验结果与分析

### 5.1. 分类精度对比

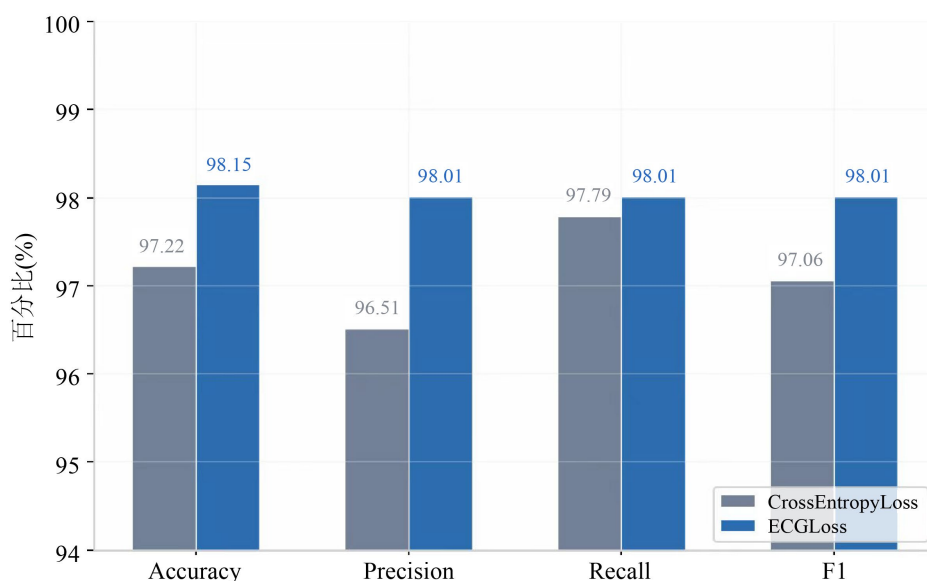
表 5 给出了 ECGLoss 和标准交叉熵损失在测试集上分类效果比较。

**Table 5.** Comparison of classification performance in ablation experiments

**表 5.** 消融实验分类性能对比

损失函数	Accuracy	Precision	Recall	F1
CrossEntropyLoss	97.22%	96.51%	97.79%	97.06%
ECGLoss	98.15%	98.01%	98.01%	98.01%

从表 5 可以看出, ECGLoss 测试准确率为 98.15%, 比标准交叉熵损失 97.22% 提高 0.93 个百分点; F1 得分是 98.01%, 提高 0.95 个百分点。而且 ECGLoss 精确率以及召回率都高于标准交叉熵损失。说明 ECGLoss 中的焦点损失和置信度惩罚共同起到积极作用, 有利于提高模型分类效果。图 2 是两类损失函数的分类指标对比。



**Figure 2.** Comparison of classification metrics between ECGLoss and CrossEntropyLoss

**图 2.** ECGLoss 与 CrossEntropyLoss 分类指标对比

### 5.2. 混淆矩阵分析

图 3 给出了 ECGLoss 在测试集上混淆矩阵。ECGLoss 混淆矩阵中 TN=39、FP=1、FN=1、TP=67，只有 1 个假负例。

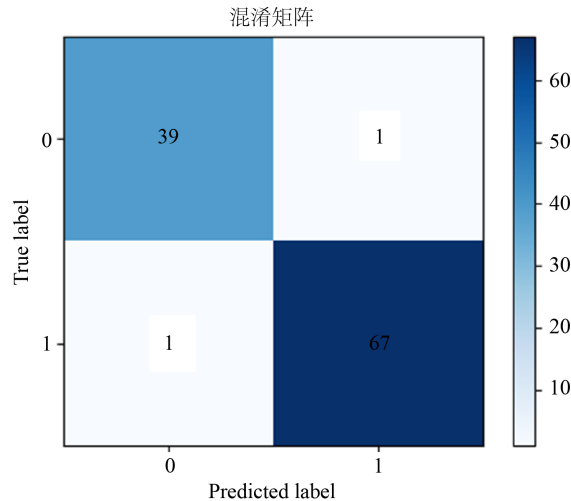


Figure 3. ECGLoss confusion matrix  
图 3. ECGLoss 混淆矩阵

ECGLoss 将假负例从标准交叉熵损失中的 3 个减少为 1 个，从而减少了误判高危患者的可能性。尽管 FP 数量由原来的 0 变为 1，但是总体而言 ECGLoss 对两种类型的样本有更好的区分效果。

### 5.3. 收敛速度对比

图 4 是两种损失函数在训练过程中所对应的损失的变化情况，在第 11 轮时标准交叉熵损失在测试集上的准确率为 97.22%，训练损失下降到一定程度后不再下降并且触发早停，在第 38 轮时 ECGLoss 在测试集上准确率为 98.15%。

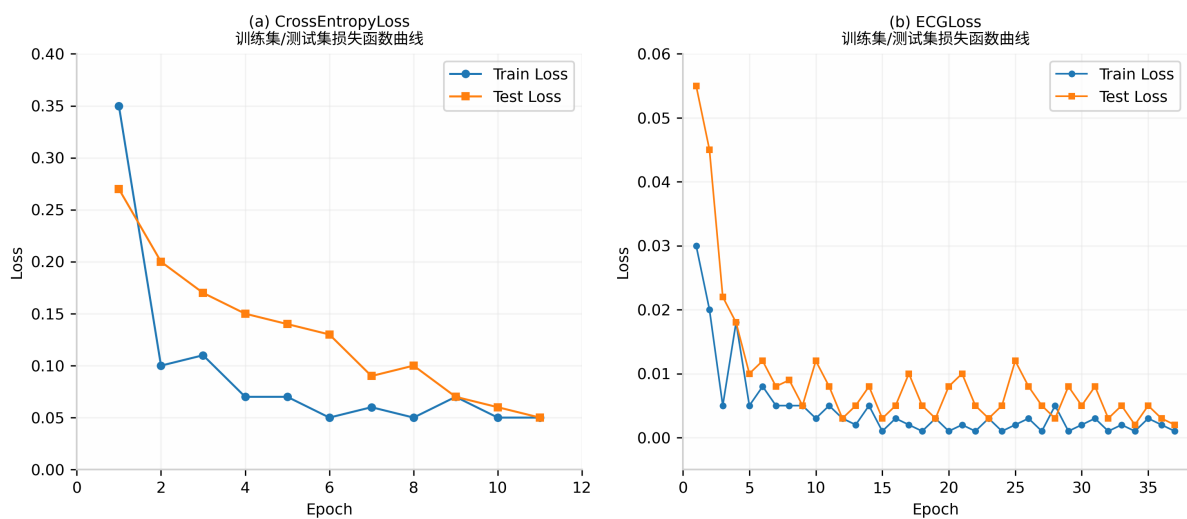


Figure 4. Comparison of different loss functions during training (left: CrossEntropyLoss, right: ECGLoss)  
图 4. 训练过程中不同损失函数比较(左: CrossEntropyLoss, 右: ECGLoss)

ECGLoss 需要更多的训练迭代次数才能收敛，这是由于置信度惩罚项给低置信度样本施加了更大的限制——模型需要更多时间去优化这些样本的预测置信度。但是最终 ECGLoss 达到的效果更好，并且其训练过程更加平滑，在测试损失上也没有大的波动，说明 ECGLoss 对训练具有一定的鲁棒性改善作用。

#### 5.4. 参数敏感性分析

为了考察 ECGLoss 的鲁棒性，我们在两个重要的超参数即置信度阈值  $\tau$  以及惩罚权重  $\lambda$  上做了敏感性实验。在实验中我们通过改变一个参数同时保持另一个参数不变来进行实验，然后记录下每次实验得到的平均 Validation Loss、Accuracy 和 F1-score 以及它们的标准差来判断模型性能的稳定性。

**Table 6.** Parameter  $\tau$  analysis

**表 6.** 参数  $\tau$  分析

$\tau$	Loss Mean	Loss Std	Accuracy Mean	Accuracy Std	F1 Mean	F1 Std
0.05	0.0014	0.0006	0.9846	0.0044	0.9846	0.0044
0.10	0.0013	0.0006	0.9877	0.0044	0.9877	0.0044
0.20	0.0014	0.0007	0.9877	0.0044	0.9877	0.0044
0.30	0.0014	0.0007	0.9877	0.0044	0.9877	0.0044
0.40	0.0013	0.0007	0.9846	0.0044	0.9846	0.0044
0.50	0.0013	0.0006	0.9877	0.0044	0.9877	0.0044
0.60	0.0013	0.0007	0.9877	0.0044	0.9877	0.0044
0.70	0.0014	0.0007	0.9877	0.0044	0.9877	0.0044
0.80	0.0015	0.0008	0.9846	0.0044	0.9846	0.0044
0.90	0.0015	0.0008	0.9877	0.0044	0.9877	0.0044

由表 6 所示的结果我们可以看出，ECGLoss 对于这两个超参数而言具有良好的鲁棒性。当  $\tau$  在一个较大的范围内变化时，模型的 Accuracy 以及 F1-score 的变化很小并且都维持在一个较高的水平，这说明这个损失函数对于置信度阈值具有较强的鲁棒性。但是当  $\tau$  较大时，模型的 Validation Loss 略有上升，这是因为太高的置信度阈值会降低模型对于一些难分样本的学习能力而导致其泛化能力减弱。综合考虑模型的效果以及稳定性，我们选择  $\tau = 0.6$  作为实验所用的默认参数。

**Table 7.** Parameter  $\lambda$  analysis

**表 7.** 参数  $\lambda$  分析

$\lambda$	Loss Mean	Loss Std	Accuracy Mean	Accuracy Std	F1 Mean	F1 Std
0.01	0.0014	0.0006	0.9877	0.0044	0.9877	0.0044
0.03	0.0013	0.0006	0.9877	0.0044	0.9877	0.0044
0.05	0.0014	0.0007	0.9877	0.0044	0.9877	0.0044
0.07	0.0013	0.0007	0.9877	0.0044	0.9877	0.0044
0.09	0.0013	0.0006	0.9907	0.0000	0.9908	0.0000
0.11	0.0013	0.0006	0.9907	0.0000	0.9908	0.0000
0.13	0.0013	0.0006	0.9907	0.0000	0.9908	0.0000
0.15	0.0013	0.0006	0.9877	0.0044	0.9877	0.0044
0.17	0.0013	0.0007	0.9877	0.0044	0.9877	0.0044
0.20	0.0014	0.0006	0.9877	0.0044	0.9877	0.0044

从表 7 来看, 对于惩罚权重  $\lambda$  来说, 我们发现它也存在一个比较合适的范围, 在 0.09 到 0.13 之间, 模型得到最好的 Accuracy 以及 F1-score, 这是由于适当的 margin penalty 可以提高模型对于那些低置信度样本的约束力进而改善分类的效果。若  $\lambda$  小, 则惩罚项不起作用无法起到应有的约束作用; 反之若  $\lambda$  过大, 则会造成过度拟合使得模型效果变差。此外从多个实验得到的标准差来看, 各个指标变化很小, 这也进一步证明了我们提出的 ECGLoss 在训练中具有很好的稳定性和可复现性, 在不同的超参数下都能得到基本一致的分类结果。

## 5.5. 结果讨论

从实验效果上看, ECGLoss 相对于普通的交叉熵损失有两个方面的提升。一个是对分类准确率提高, 准确率以及 F1 都提高了大约 1%, 而且假负例也显著下降; 另一个是训练更加稳定, 在测试集上的损失曲线比较平稳并且没有明显的过拟合现象。

ECGLoss 缺陷是收敛速度慢, 需要接近四倍的时间才能获得最优效果, 这是由于置信度惩罚增加优化难度造成的, 在使用时可以调节  $\lambda$  和  $\tau$  以取得较好效果。

## 6. 总结与展望

本论文针对心电信号高风险分类问题提出一种基于随机森林特征选择以及 Transformer 分类器方法并设计一种新的交叉熵损失函数 ECGLoss。该模型对选出的特征进行线性变换后送入一个单层 Transformer 编码器中再通过平均池化送入线性分类器进行预测。ECGLoss 包含焦点损失项以及置信度惩罚项, 实验结果显示该方法可以取得 98.15% 的测试精度比普通的交叉熵损失提高 0.93 个百分点并且将假负例从 3 个降低到 1 个。

未来的工作可以考虑以下几个方向进行:

- (1) 将 ECGLoss 应用到更多的多类心律失常分类问题中;
- (2) 在更大的公开数据集上对提出的算法进行测试;
- (3) 研究如何使网络更轻量化以适应嵌入式的实时心电图仪;
- (4) 研究如何在保证准确性的前提下使 ECGLoss 更快收敛等。

## 致 谢

本论文的研究工作得到了导师的悉心指导和同学们的帮助, 在此表示衷心的感谢。同时感谢 PhysioNet 提供的高风险心电图片段数据集对本研究的支持。

## 参考文献

- [1] 宋喜国, 邓亲恺. MIT-BIH 心率失常数据库的识读及应用[J]. 中国医学物理学杂志, 2004, 21(4): 230-232.
- [2] Goldberger, A.L., Amaral, L.A.N., Glass, L., Hausdorff, J.M., Ivanov, P.C., Mark, R.G., *et al.* (2000) PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation*, **101**, e215-e220. <https://doi.org/10.1161/01.cir.101.23.e215>
- [3] 张丹, 隋文涛, 梁钊, 等. 基于深度残差网络的心律失常自动分类算法研究[J]. 电子测量与仪器学报, 2019, 33(4): 140-145.
- [4] Dosovitskiy, A., Beyer, L., Kolesnikov, A., *et al.* (2021) An Image Is Worth  $16 \times 16$  Words: Transformers for Image Recognition at Scale. *International Conference on Learning Representations (ICLR)*, Vienna, 3-7 May 2021.
- [5] 刘光达, 周葛, 董梦坤, 等. 基于 FFNN 和 1D-CNN 的实时心律失常诊断系统与算法[J]. 电子测量与仪器学报, 2021, 35(3): 35-42.
- [6] 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>
- [7] Wang, Y., Yang, G., Li, S., Li, Y., He, L. and Liu, D. (2023) Arrhythmia Classification Algorithm Based on Multi-Head Self-Attention Mechanism. *Biomedical Signal Processing and Control*, **79**, Article ID: 104206. <https://doi.org/10.1016/j.bspc.2022.104206>
- [8] Vaswani, A., *et al.* (2017) Attention Is All You Need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6000-6010.
- [9] 李明阳, 王磊, 张伟, 等. 基于多尺度特征融合与注意力机制的心律失常检测[J]. 计算机研究与发展, 2023, 60(8): 1789-1801.
- [10] 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>