Software Engineering and Applications 软件工程与应用, 2023, 12(3), 495-505
Published Online June 2023 in Hans. https://www.hanspub.org/journal/sea
https://doi.org/10.12677/sea.2023.123049

Hans汉斯

面向不平衡甲状腺眼病数据集的分类算法对比研究及应用

张天凤[1]，赵　廉[1]，邓浩然[1]，朱沁汶[1]，陈　钱[1]，宋诗淳[1]，宋雪霏[2,3]，周　雷[1*]

[1]上海理工大学，健康科学与工程学院，上海
[2]上海交通大学医学院附属第九人民医院眼科，上海
[3]上海市眼眶病眼肿瘤重点实验室，上海

收稿日期：2023年5月11日；录用日期：2023年6月9日；发布日期：2023年6月25日

摘　要

对不同数据进行分类是机器学习的研究热点，然而在各大领域，数据不平衡现象是普遍存在的。现有的许多机器学习算法虽然取得了良好的效果，但他们都是在默认数据集分布均衡的前提下进行的，并且认为不同类别的误分代价一致，这导致它们在不平衡数据集上表现很差。本文针对用于甲状腺眼病诊断的数据集出现的正负样本不平衡现象，选择了WCE loss，LDAM-loss，Focal-loss，Minimax四种面向不平衡数据的优化方法进行了对比实验。实验结果表明，用不平衡优化方法训练的分类模型相对于原始模型具有更好的分类性能。实验还发现随正负样本比例的不同，各方法对结果的提升存在一定差异，在重度不平衡条件下，LDAM loss和Minimax表现出更好的鲁棒性，尤其是Minimax方法，它对于少数类的分类性能更好。总结而言，本论文所展示的对比实验能在不平衡甲状腺眼病诊断数据的条件下，对分类算法的选取提供指导。

关键词

不平衡数据，机器学习，甲状腺相关眼病，分类

Comparative Study and Application of Classification Algorithms for Unbalanced Thyroid Eye Disease Datasets

Tianfeng Zhang[1], Lian Zhao[1], Haoran Deng[1], Qinwen Zhu[1], Qian Chen[1], Shichun Song[1], Xuefei Song[2], Lei Zhou[1*]

[1]School of Health Science and Engineering, University of Shanghai for Science and Technology, Shanghai
[2]Department of Ophthalmology, Ninth People's Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai

[*]通讯作者。

文章引用: 张天凤, 赵廉, 邓浩然, 朱沁汶, 陈钱, 宋诗淳, 宋雪霏, 周雷. 面向不平衡甲状腺眼病数据集的分类算法对比研究及应用[J]. 软件工程与应用, 2023, 12(3): 495-505. DOI: 10.12677/sea.2023.123049

[3]Shanghai Key Laboratory of Orbital Diseases and Ocular Oncology, Shanghai

Received: May 11th, 2023; accepted: Jun. 9th, 2023; published: Jun. 25th, 2023

Abstract

Data classification is a prominent area of machine learning, but data imbalance is a common issue across major fields. Although many machine learning algorithms have produced favorable outcomes, they rely on the assumption that the default dataset is uniformly distributed and the cost of false separation for different categories is consistent. Consequently, they exhibit poor performance on unbalanced datasets. In this study, four optimization methods were selected to address the issue of unbalanced data in the diagnosis of thyroid eye disease. These methods, including WCE loss, LDAM-loss, Focal-loss, and Minimax, were used to compare the positive and negative sample imbalance in the dataset. The experimental results demonstrate that the unbalanced optimization method produced a better classification performance than the original model. Furthermore, the experiments revealed that the improvement of results varied with the proportion of positive and negative samples, with LDAM loss and Minimax exhibiting better robustness under severe imbalance conditions. The Minimax method, in particular, demonstrated superior classification performance for minority classes. In conclusion, the comparative experiment presented in this study can offer valuable insights for the selection of classification algorithms under the condition of unbalanced thyroid eye disease diagnostic data.

Keywords

Unbalanced Data, Machine Learning, Thyroid-Associated Ophthalmopathy, Classification

1. 引言

在自然界或是医学领域，数据不平衡现象是普遍存在的。在一个数据集中，如果其中一个或多个类别包含大量的样本数据，而另一类别或另外一些类别只包含少量样本数据，那么这个数据集就存在类别不均衡问题，其中样本较多的类称为多数类，样本较少的类称为少数类。而在实际应用中，我们往往关注的是少数类的情况，比如在医院的影像检查中，患病者往往是少数，但一旦误将病人诊断成正常人，会导致疾病没有及时得到发现，错过最佳治疗时间，造成严重的后果。

不平衡数据分类问题的特性主要体现在数据本身的非均衡性以及传统分类算法的局限性两方面[1] [2]。首先是数据非均衡性问题，由于少数类样本数量少，训练器无法从少数类样本获得足够信息，使得模型学习到的都是多数类特征，导致少数类样本无法识别；其次是传统分类算法的局限性，传统分类器通常以最小化损失函数为训练目标，以达到最低误分率或最大类间间隔的效果[1]。当数据分布不平衡时，由于类间分布不均，算法往往会将少数类样本错误归为多数类，从而导致分类偏差。

目前对不平衡数据分类算法的研究主要集中在特征选择、数据重采样、改进分类算法几个方面。特征选择就是从所有特征中选择更有区分度的子集，从而提高少数类样本的分类性能；数据重采样包括对多数类数据进行欠采样，对少数类数据进行过采样以及过采样与欠采样相结合的方法，以达到让数据平

衡的目的，常见的有 Tomek Links 欠采样、SMOTE (Synthetic Minority Oversampling Technique) [3]过采样等等；改进分类算法主要包括代价敏感学习、集成学习等方法，代价敏感学习通常让少数类数据误分代价更高从而提高少数类样本的分类精度，比如 focal loss 方法[4]，集成学习则是通过将多个弱分类器组合成一个更强大的分类器，以提高分类性能，常见的集成学习方法有 AdaBoost [5]、Random Forest [6]等等。

甲状腺相关眼病(Thyroid-Associated Ophthalmopathy, TAO)是一种特异性自身免疫性疾病，患者会出现眼球凸出、视力模糊、斜视等症状，影响患者的外貌和视力，甚至导致失明[7]。对于 TAO 诊断，根据影像组学特征进行判断是常用方法。杨等人提取眼外肌横截面积和凸眼度进行 TAO 诊断[8]；丛等人结合眼球突出度和眼球影像组学特征进行 TAO 预测[9]。他们的方法虽取得了良好效果，但是他们均是在平衡数据集上进行分类预测，并未研究不平衡数据集下的情况，而在实际运用中，不平衡数据集才是常态。

本文对于两个不平衡甲状腺眼病数据集分别应用了 4 种不同的不平衡数据集分类优化算法，实验结果表明，相对于原始分类算法，改进的分类算法具有更好的鲁棒性和分类性能。

2. 研究方法

本文研究方法具体来说，就是将特征传入 BP (Back Propagation)神经网络[10]，获得分类结果。4 种不平衡优化方法都是在模型损失函数上做了改进，详见 2.3 至 2.6 节。

2.1. BP 分类网络

出于公平比较的目的，本文所有实验均采用同一个 BP 神经网络，网络结构如图 1 所示。网络共有 3 层，分别为输入层，隐藏层和输出层。输入层的维度即为原始特征维度，隐藏层的维度为输入特征维度的二分之一，由于是二分类，所以输出层的维度是 2，最后经过 softmax 函数获得最终分类结果。为了防止过拟合，BP 网络在训练阶段的隐藏层使用了丢弃法(Dropout) [11]，概率为 0.2，每次训练每个神经元有 0.2 概率不被激活。

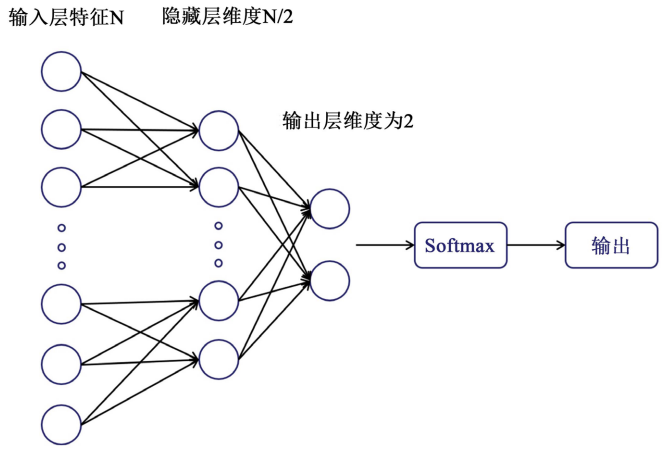

Figure 1. Structure of BP network
图 1. BP 网络结构图

2.2. 特征提取和降维

本文实验选用的数据集有两个。第一个数据集是极重度患者数据集，其特征均为医生手工计算得到，无需提取。第二个是用于甲状腺眼病诊断的 MRI 数据集，每一例都有对应的 MRI 图像和分割标签。分割标签由专业眼科医生进行标注，包含 8 类，分别为上直肌、下直肌、内直肌、外直肌、视神经、脂肪、

泪腺、上斜肌。本文使用 pyradiomics 库对 8 类分割标签提取组学特征，每一类提取到 1648 维特征，其特征类别及数量如表 1 所示，其中 shape 描述眼球眼肌的形状特征，firstorder 为一阶统计量，描述眼球眼肌区域的像素灰度分布，其余均为纹理特征。8 类标签共 13,184 维组学特征。

Table 1. Abstract for different types of features
表 1. MRI 数据集的特征

特征类别	特征数量
shape	14
firstorder	342
glrlm	304
glszm	304
gldm	266
glcm	418

由于 MRI 影像组学数据集特征维度太高，所以还需要对它进行特征降维。本文采用特征选择进行降维。一个好的特征应当符合以下 2 点要求：

- 该特征对于活跃期患者与非活跃期患者应有显著差异；
- 该特征不应该与除它本身以外的任何特征强相关。

本文采用双侧 t 检验来确定每个特征对于活跃期患者和非活跃期患者的差异性，当 p 值小于 0.05 时，本文认为该特征对于活跃期患者和非活跃期患者具有显著性差异。通过双侧 t 检验，本文将 13,184 维特征降维到 5129。然后通过计算两两特征间的泊松相关系数来确定两个特征的线性相关性，当相关系数的绝对值大于 0.9 时，本文认为这两个特征强相关，会删掉其中一个特征以减少特征冗余。通过泊松相关性分析，本文进一步将 5129 维特征进一步降到 1163 维。

接下来本文将对 4 种不平衡优化方法进行介绍。

2.3. WCE Loss

重加权是代价敏感学习的常用方法，它是从损失函数角度出发，通过增大少数类样本的权重来改善样本不平衡问题。加权交叉熵损失(WCE Loss, Weighted Cross Entropy)是一种经典的重加权方法，对于二分类问题，其公式如下所示：

$$L_{wce} = -wy\log(p_t) - (1-w)(1-y)\log(1-p_t) \tag{1}$$

在式(1)中，y 是样本标签(正样本为 1，负样本为 0)，p_t 是预测为正样本的概率，w 是 0 到 1 之间的常数，通过控制 w 可以调整正负样本对损失函数的贡献，假设数据集中正样本数量为 m，负样本数量为 n，那么当满足 $w = \frac{n}{n+m}$ 时，可以保证一次训练中正负样本对损失函数的影响相同，该损失函数也可以称为平衡交叉熵损失。

2.4. Focalloss

加权交叉熵损失虽然通过设置w平衡了样本间数量对梯度的贡献，但没有考虑到数据本身存在难分和易分样本的情况，针对这一问题，Lin 等人提出了 focal loss [4]。对于二分类问题，其公式如下所示：

$$L_{fl} = -(1-p_t)^{\gamma}\log(p_t) \tag{2}$$

式中，γ 为可调节的超参数，用于调整难易样本的权重。p_t 是分类器的输出概率，如果输入为正类且 p_t

接近 1，说明该样本分类正确且容易训练，则对应的$(1-p_t)^{\gamma}$值就会减小，从而抑制易分样本的贡献；当输入为正类且p_t接近 0 时，表示样本分类错误且难以训练，则对应的$(1-p_t)^{\gamma}$的值就会增大，从而增加难分样本的权重。

当$\gamma=0$时，focal loss 便会退化成一般的交叉熵损失，γ的值越大，模型越关注难分样本。

2.5. LDAM loss

LDAM Loss (Label-Distribution-Aware Margin Loss) [12]最小化边缘泛化边界损失，由 Cao 等人提出，其核心思想是将分类边界向多数类偏移，从而增大少数类样本的分类容错率。如下图 2 所示，三角形是多数类样本，圆是少数类样本，虚线为数据平衡前提下的理想分类边界，粗实线是不平衡数据下的实际分类边界，γ_1和γ_2分别为类别 1 和类别 2 到分类边界的最小距离。一般的，在数据平衡的情况下，两个类到达分类边界的最小距离是一样的，但在数据不平衡情况下，由于少数类样本过少，分类器无法从少数类样本学到足够特征从而导致少数类样本分类性能很差，因此 LDAM Loss 就考虑对少数类样本降低要求，也就是将分类边界向多数类样本偏移，从而增大少数类样本的容错率。

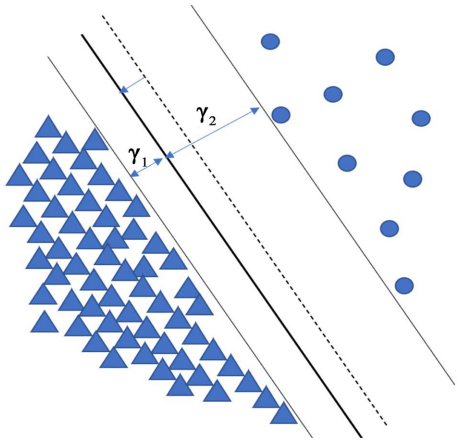

Figure 2. Binary linear separable classifier
图 2. 二分类线性可分分类器

γ与样本数量有关，计算方式如式(3)所示，式中j是样本的某一类别，n_j是样本中该类的样本数，C是人为设置的超参数，目的是使少数类样本的边界移动更大。

$$\gamma=\frac{C}{n_j^{\frac{1}{4}}} \tag{3}$$

$$L_{LDAM}=-\log\frac{e^{z_y-\gamma}}{e^{z_y-\gamma}+\sum_{j\neq y}e^{z_j}} \tag{4}$$

式(4)表示了多类 LDAMloss 的计算方法，式中，j和y都为类别标签，z_j和z_y即为对应类别样本的模型输出。

2.6. 非线性优化方法

优化法最初用于解决非线性回归问题，由于非线性回归原理与神经网络类似，因此优化法也可以用于解决当训练数据不满足独立同分布或不满足样本均衡的条件时，机器学习的损失计算问题。

假设样本训练集为$\{(x_1,y_1),(x_2,y_2),\cdots,(x_n,y_n)\}$，其中$(x_1,x_2,\cdots,x_n)$为自变量，$(y_1,y_2,\cdots,y_n)$为类别标签，$n$为样本数量，$f$是一种非线性变换。

$$\theta = f(x_i) - y_i \quad 0 \le i \le n \tag{5}$$

式(5)中的θ表示变换结果与真实值的差，在回归问题中，通常的目的是使θ尽可能的小。

$$\theta = \frac{1}{n}\sum_{i=1}^{n}\left(f(x_i)-y_i\right)^2 \tag{6}$$

$$\theta = \max_{1\le i\le n}\left(f(x_i)-y_i\right)^2 \tag{7}$$

式(6)就是回归问题中常用的均方误差损失，式(7)是最大损失。通常来说使用平均损失的情况要优于最大损失，平均损失会结合一次训练中每个数据的情况进行计算，但这种方式会在数据不平衡时出现问题，Xu 等人提出了极小极大优化法(Minimax) [13]来解决这一问题，它的思路是将数据分为几组，每组组内数据分布一致，先在组内计算每组的均方误差损失，再计算各组间最大损失作为最终损失函数。

$$\theta = \min\max_{1\le j\le n}\frac{1}{k_j}\sum_{i=1}^{k_j}\left(f(x_i)-y_i\right)^2 \tag{8}$$

式(8)就是最终的损失函数，将数据分为j组，每组有k个数据，k_j表示第j组中的所有数据。该损失函数的目的是使组间最大损失达到最小。

将回归问题推广到分类问题上，也可以采用类似的训练方法，假设一个不平衡数据集，正例 100，反例 200，那么按照此方法可以将数据分为 3 组，每组数据 100 个，其损失如下式所示：

$$loss = \max_{1\le j\le 3}\frac{1}{100}\sum_{i=1}^{100}\left(-y\log(p_t)-(1-y)\log(1-p_t)\right) \tag{9}$$

式中，y是类别标签(0 或 1)，p_t是分类器预测的概率。首先计算出每个组各自的损失，再用组间最大损失作为整体损失，通过让 loss 尽可能的小来优化分类器，由于每个组内部只包含一个类别，因此数据不平衡不会对组内损失产生干扰。

3. 实验

3.1. 数据集

本文实验选用的数据集有两个。第一个是来自上海第九人民医院的 300 例眼部 MRI 扫描，每一例都有对应的分割标签，由专业眼科医生进行标注。该数据集包含 200 例甲状腺眼病活跃期患者，100 例非活跃期患者，所有数据均经过脱敏操作。该数据集采用影像组学特征进行分类，组学特征提取和降维见 2.2 节。第二个数据集来自上海第九人民医院的 163 例 TAO 患者，其中 58 名极重度患者，105 名非极重度患者，特征全部由医生手工计算得到。

数据集具体信息如表 2 所示。

本文对 MRI 影像组学数据集和极重度患者数据集都进行了 MinMax 归一化，归一化公式如式(10)所示。

Table 2. Experimental datasets
表 2. 实验数据集

数据集	正类数量	反类数量	比值	特征数目
MRI 影像组学数据集	100	200	0.5	1163
极重度患者数据集	58	105	0.5524	26

$$x_i' = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad i = 1, 2, \cdots, k \tag{10}$$

式中，x_i 为原始特征，x_i' 为归一化后的新特征，$x_{\min}$ 是最小特征，$x_{\max}$ 是最大特征，k 是样本总数。

3.2. 实验细节

对于两个数据集，本文均随机选 80%数据作为训练集，20%数据作为测试集，并进行五折交叉验证，实验结果为五折交叉验证的平均值。

本文的各个方法参数设置如下：

1) 对于基线模型，采用 pytorch 框架下不加参数的 Crossentropy 函数；

2) 对于 WCE loss，在 Crossentropy 损失中增加了权重参数，参考其正负样本比例，为[0.66, 0.34]；

3) 对于 Focal loss，采用文献[4]中的实现，超参数 $\gamma = 1.5$；

4) 对于 LDAM loss，采用文献[12]中的实现，超参数 $C = 30$；

5) 对于 Minimax 方法，每个 bacth 传入一个只有正类的数据组和只有反类的数据组，分别计算组内损失，使组间损失最小化。

模型初始学习率为 3×10^{-4}，采用余弦退火衰减[14]来动态调整学习率，使用 adam 优化器，batch-size 为 36，每个方法各训练 50 epoch。

由于数据不平衡，使用常规指标会导致结果偏向某一类，因此本文采用平均 F1 分数(Average-F1)、平均精确率(Average-Precision)、平均召回率(Average-Recall)和准确率(Accuracy)作为评价指标。

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{11}$$

$$\text{Average-Precision} = 0.5\frac{TP}{TP + FP}(label = 0) + 0.5\frac{TP}{TP + FP}(label = 1) \tag{12}$$

$$\text{Average-Recall} = 0.5\frac{TP}{TP + FN}(label = 0) + 0.5\frac{TP}{TP + FN}(label = 1) \tag{13}$$

$$\text{Average-F1} = 2\frac{\text{Average-Precision} \times \text{Average-Recall}}{\text{Average-Precision} + \text{Average-Recall}} \tag{14}$$

除此以外，还引入了受试者工作特征曲线(Receiver Operating Characteristic Curve, ROC)进行评估，ROC 曲线下面积称为 AUC，AUC 越大分类器性能越好。

3.3. 实验结果与分析

3.3.1. MRI 影像组学数据集实验结果

表 3 为 MRI 影像组学数据集实验的五折交叉验证平均结果，图 3 为各模型的 ROC 曲线。从表中可以看出，相较于基线模型(CE)，其余 4 种方法的性能均有一定提升，其中 WCE loss 方法的平均精确率最高，达到了 0.71，LDAM loss 方法 AUC 最高，为 0.803。综合来看，Focal loss 方法的综合表现最好，在准确率、平均召回率、平均 F1 分数三项性能指标上达到了最大值，它的准确率为 0.747，AUC 为 0.759，平均精确率为 0.693，平均召回率为 0.742，平均 F1 分数为 0.696。

3.3.2. 极重度患者数据集实验结果

极重度患者数据集实验结果如表 4 所示，图 4 为各模型在极重度患者数据集上的 ROC 曲线。从表中可以看出，相较于基线模型(CE)，其余 4 种方法在所有指标上性能都有提升，其中表现最好的是 Minimax 方法，它的准确率为 0.788，平均 F1 分数为 0.785，平均精确率为 0.792，平均召回率为 0.816，AUC 为 0.859，均达到了最优性能。

Table 3. Prediction accuracy of MRI testset (positive and negative sample ratio 1:2)
表 3. 各模型在 MRI 数据集测试集(正负样本比例 1:2)上的预测准确率

分类器	准确率	平均精确率	平均召回率	平均 F1 分数	AUC
CE	0.700	0.610	0.697	0.592	0.738
WCE loss	0.710	**0.710**	0.715	0.680	0.783
LDAM loss	0.733	0.670	0.648	0.648	**0.803**
Focal loss	**0.747**	0.693	**0.742**	**0.696**	0.759
Minimax	0.727	0.690	0.706	0.681	0.758

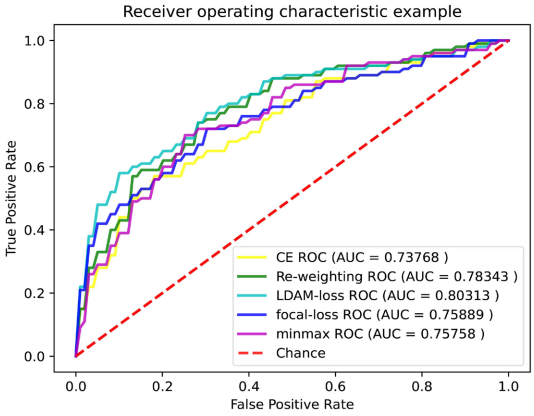

Figure 3. ROC curves of each model on the test set of MRI dataset (1:2)
图 3. 各模型在 MRI 数据集测试集(1:2)上的 ROC 曲线

Table 4. Prediction accuracy of extremely severe patient testset (positive and negative sample ratio 1:2)
表 4. 各模型在极重度患者测试集(正负样本比例 1:2)上的预测准确率

分类器	准确率	平均精确率	平均召回率	平均 F1 分数	AUC
CE	0.727	0.732	0.750	0.723	0.820
WCE loss	0.758	0.754	0.774	0.752	0.824
LDAM loss	0.758	0.772	0.792	0.756	0.848
Focal loss	0.758	0.754	0.774	0.752	0.832
Minimax	**0.788**	**0.792**	**0.816**	**0.785**	**0.859**

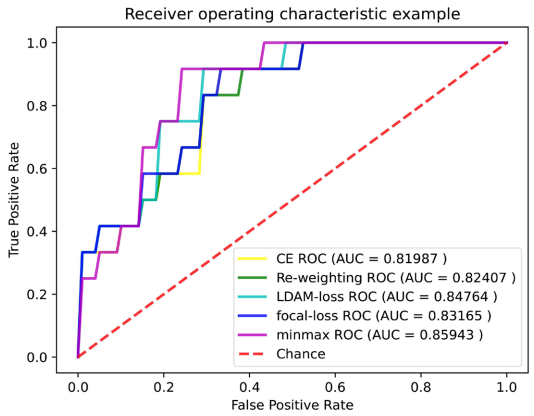

Figure 4. ROC curves of each model on the test set of extremely severe patient dataset
图 4. 各模型在极重度患者数据集测试集上的 ROC 曲线

3.3.3. 改变正负样本比例后的实验结果

以上两个数据集的正负样本比例均在 1:2 左右，为了测试数据不平衡更严重时的情况，本文将 MRI 影像组学数据进行了调整，取 50 例正样本，200 例负样本作为实验数据集，仍选取 20%数据作为测试，除了 WCE loss 的权重参数变为[0.8, 0.2]，其余超参数设置和训练方法与上文保持一致。实验结果如表 5 所示，图 5 为各模型的 ROC 曲线。相较于基线模型(CE)，其余 4 种方法的性能依旧有一定提升，其中 LDAM loss 方法在 AUC 和准确率上最高，分别为 0.748 和 0.836，而 Minimax 方法在剩余三项评价指标中达到了最优性能，其平均精确率为 0.653，平均召回率为 0.707，平均 F1 分数为 0.667。

Table 5. Prediction accuracy of MRI testset (positive and negative sample ratio 1:4)
表 5. 各模型在 MRI 数据集测试集(正负样本比例 1:4)上的预测准确率

分类器	准确率	平均精确率	平均召回率	平均 F1 分数	AUC
CE	0.780	0.645	0.623	0.625	0.642
WCE loss	0.796	0.550	0.458	0.497	0.701
LDAM loss	**0.836**	0.643	0.640	0.634	**0.748**
Focal loss	0.820	0.602	0.694	0.594	0.708
Minimax	0.816	**0.653**	**0.707**	**0.667**	0.725

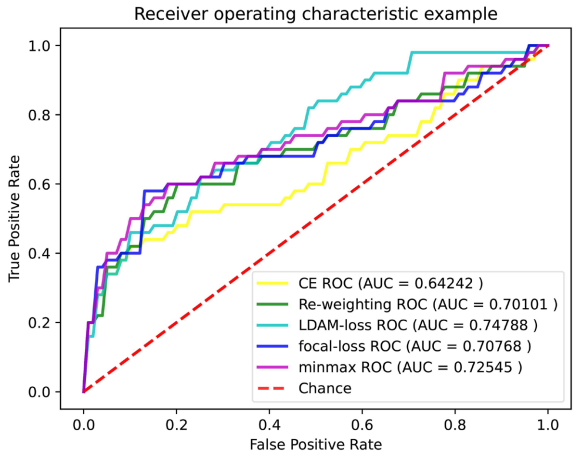

Figure 5. ROC curves of each model on the testset of MRI dataset(1:4)
图 5. 各模型在 MRI 数据集测试集(1:4)上的 ROC 曲线

3.3.4. 结果分析

结合表 3~5，可以看出，相较于使用原始交叉熵，WCE loss，LDAM loss，Focal loss，Minimax 四种优化方法训练的模型，其准确率和 AUC 均有一定提升，表明了这四种方法对不平衡数据集的有效性。

其中，在 MRI 影像组学数据集(正负样本比例 1:2)实验中，对结果综合提升最多的是 Focal loss 方法，其次是 LDAM loss 和 Minimax，提升最小的是 WCE loss 方法，仅提升了 1.0%准确率。

在极重度患者数据集(正负样本比例 1:2)实验中，Minimax 方法表现最好，在所有指标上均达到了最优性能。WCE loss，LDAM loss，Focal loss 三种方法准确率一致，但 LDAM loss 的 F1 分数略高，说明它对少数类的预测准确率高于另外两种方法，同时，五种方法的 AUC 均超过 0.8，属于有一定应用价值的分类器，通过合理设置阈值可以发挥其预测价值。

在 MRI 影像组学数据集(正负样本比例 1:4)上，LDAM loss 方法和 Minimax 方法依旧有良好的性能表现，两者在所有性能评价指标上都有提升，其中 Minimax 的平均 F1 分数最高，提升了 4.2%，这说明它对少数类的分类性能更好。另外两种重加权方法虽然在准确率和 AUC 上都有提升，但在其它性能指标上相较于原始交叉熵损失有所下降。这说明，随着数据集不平衡的严重程度提高，只凭借样本比例对损失函数进行简单的重加权依旧会使分类器的分类结果更倾向于多数类。

分析三次实验可以发现，当不平衡现象比较轻微(正负样本 1:2)时，4 种方法都能在所有性能指标上有所改进，但当不平衡现象增大后，LDAM loss 和 Minimax 表现出更好的鲁棒性，尤其是 Minimax 方法，它对于少数类的分类性能更好，这在我们医学诊断中十分重要，因为患者往往是少数类，能更好地诊断出患者是计算机辅助诊断的第一要务。因此在不平衡现象严重时，更推荐使用 Minimax 方法。

4. 讨论与结论

本文针对用于甲状腺眼病诊断的 MRI 影像组学数据集和极重度患者数据集出现的正负样本不平衡现象，选择了 WCE loss，LDAM-loss，Focalloss，Minimax 四种不平衡优化方法进行了实验。在使用相同的网络模型情况下，用不平衡优化方法训练的分类模型各项指标均超过使用原始交叉熵损失训练的模型，表明了不平衡优化方法的有效性。实验还发现随正负样本比例的不同，各方法对结果的提升存在一定差异。在轻度不平衡条件下，样本分类的难易程度对模型选择的影响更大，如果样本分类的难易程度较高，推荐使用 focal loss 方法，因为它更注重于提高难分样本的权重；如果样本分类的难易程度较低，则模型选择上需要更多考虑数据不平衡的影响，此时 Minimax 方法会成为更好的选择。在重度不平衡条件下，类别不平衡已成为我们模型选择的首要考虑因素，此时更推荐使用 Minimax 方法，它对于少数类的分类性能更好。本文所展示的对比实验能在不平衡甲状腺眼病诊断数据的条件下，对分类算法的选取提供指导。

基金项目

本研究获得国家自然科学基金项目资助，项目编号为 61906121。

参考文献

[1] 李艳霞, 柴毅, 胡友强, 等. 不平衡数据分类方法综述[J]. 控制与决策, 2019, 34(4): 673-688.

[2] 赵楠, 张小芳, 张利军. 不平衡数据分类研究综述[J].计算机科学, 2018, 45(z1): 22-27, 57.

[3] Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P. (2002) SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research*, **16**, 321-357. https://doi.org/10.1613/jair.953

[4] Lin, T.Y., Goyal, P., Girshick, R., He, K.M. and Dollár, P. (2017) Focal Loss for Dense Object Detection. 2017 *IEEE International Conference on Computer Vision*, Venice, 22-29 October 2017, 2999-3007. https://doi.org/10.1109/ICCV.2017.324

[5] Freund, Y. and Schapire, R.E. (1997) A Desicion-Theoretic Generalization of on-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, **55**, 119-139. https://doi.org/10.1006/jcss.1997.1504

[6] Ho, T.K. (1995) Random Decision Forests. *Proceedings of 3rd International Conference on Document Analysis and Recognition*, Montreal, 14-16 August 1995, 278-282.

[7] 陈欢欢, 杨涛. 甲状腺相关眼病发病机制研究进展[J]. 中国实用内科杂志, 2015, 35(7): 561-565.

[8] 杨科, 丰玲玲, 黄海新. 眼眶 CT 定量分析在甲状腺相关眼病诊治中的应用效果[J]. 影像研究与医学应用, 2022, 6(1): 58-60.

[9] 从志洋, 鲁婷, 范璟源, 宋雪霏, 王慧, 周雷. 基于眼球特征提取的甲状腺眼病预测方法研究[J]. 软件工程与应用, 2022, 11(6): 1288-1296.

[10] Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986) Learning Representations by Back Propagating Errors. *Nature*, **323**, 533-536. https://doi.org/10.1038/323533a0

[11] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R. (2014) Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, **15**, 1929-1958.

[12] Cao, K., Wei, C., Gaidon, A., Arechiga, N. and Ma, T.Y. (2019) Learning Imbalanced Datasets with Label-Distribution-Aware Margin Loss. 33*rd Conference on Neural Information Processing Systems*, Vancouver, 8-14 December 2019, 1567-1578.

[13] Xu, Q. and Xuan, X. (2018) Nonlinear Regression without i.i.d. Assumption. *Probability*, *Uncertainty and Quantitative Risk*, **4**, Article No. 8. https://doi.org/10.1186/s41546-019-0042-6

[14] Loshchilov, I. and Hutter, F. (2016) SGDR: Stochastic Gradient Descent with Warm Restarts. https://arxiv.org/abs/1608.03983