基于混合对比的无偏半监督医学影像分类模型

袁 满，李宇健，袁靖舒

（东北石油大学 计算机与信息技术学院，黑龙江 大庆 163318）

摘 要：针对医学影像领域标注数据获取成本高、类间高度相似以及数据集不平衡等问题，提出了一种基于混合对比的无偏半监督分类模型，用于医学影像分类任务。通过引入半监督训练模式，减少了模型训练对标注数据的依赖。模型包含混合对比模块和自适应去偏模块两个关键组件。混合对比模块通过对标注数据与无标注数据进行类级和实例级对比学习，捕获其中丰富的隐含信息，从而提升模型对医学影像特征的提取能力。自适应去偏模块通过对导致分类偏差的因素进行解耦，生成样本的无偏预测，提升了模型对于不平衡医学影像数据的分类性能。在两个不同模态且类别不平衡的医学数据集上的实验结果表明，本文提出的模型在所有分类性能指标上均优于现有同类型的半监督医学分类模型。

关键词：深度学习；半监督学习；医学影像处理；图像分类；对比学习

中图分类号：TP391 **文献标志码**：A **文章编号**：1671-5497(2026)06-1718-09

DOI：10. 13229/j. cnki. jdxbgxb. 20241281

Unbiased semi-supervised medical image classification model based on hybrid contrastive learning

YUAN Man，LI Yu-jian，YUAN Jing-shu

（*School of Computer and Information Technology*，*Northeast Petroleum University*，*Daqing* 163318，*China*）

Abstract：To address the challenges of high annotation costs，high inter-class similarity，and data imbalance in medical imaging，an unbiased semi-supervised classification model was proposed based on hybrid contrastive learning for medical image classification tasks. By introducing a semi-supervised training paradigm，the model reduces reliance on annotated data during training. The proposed model features two key components：a hybrid contrastive module and an adaptive debiasing module. The hybrid contrastive module performs class-level and instance-level contrastive learning on both labeled and unlabeled data，capturing rich implicit information and enhancing the model′s feature extraction capabilities for medical images. The adaptive debiasing module decouples factors contributing to classification bias，enabling unbiased predictions and improving classification performance on imbalanced medical image datasets. Experimental results on two medical datasets with different modalities and imbalanced class distributions demonstrate that the proposed model outperforms existing state-of-the-art semi-supervised medical classification models across all evaluation metrics.

收稿日期：2025-01-12.

基金项目：黑龙江省科技创新基地奖励项目(JD24A009).

作者简介：袁满(1965-)，男，教授，博士. 研究方向：知识组织. E-mail：yuanman@nepu.edu.cn

Key words: deep learning; semi-supervised learning; medical image processing; image classification; contrastive learning

0　引　言

随着深度学习技术在各个领域的落地,极大地推动了各种产业的迅速发展。对医学影像进行分类是医学影像处理过程中的重要环节,使用深度学习模型对大量的医学影像进行分类可以在医院人员培训和临床实践过程中极大地减轻影像科室医生的工作负担,同时为后续的诊断提供可靠的辅助数据。然而,深度学习模型的训练依赖大量的标注数据支持。在医疗领域,对医学数据进行标注要求标注者具备高度专业化的知识且需注意力高度集中地进行标注,这导致获取标注数据的成本远高于其他大部分图像领域。近年来,半监督学习凭借对标注数据依赖性远小于传统监督学习的优势,在多种深度学习任务中得到了广泛应用,半监督学习通过将数据集划分为少量有标签数据与大量的无标签数据,通过设计不同的利用无标签数据的方式,如伪标签生成[1]、一致性正则化[2]和标签传播[3]等,在少量标注数据的引导下训练深度学习模型,这一方案可以极大地减少数据标注的成本,目前已有不少工作[4,5]将半监督学习引入到医学影像分类任务中。需要注意的是,与自然图像数据相比,医学影像通常具有类间高度相似和类别分布不均衡的特性。这种差异导致自然图像分类中的半监督方法在医学影像分类任务中表现欠佳。尽管如此,半监督学习在医学影像处理领域的潜力仍然得到了研究者们的广泛认可[6],同时针对医学影像的数据特性许多研究提出了针对性的改进方案。

针对类间高度相似的问题,许多研究提出了结合数据扰动与一致性正则化的方案。具体而言,这类方法通过对样本添加噪声或应用不同的数据增强方式,生成样本的不同扰动版本,并施加一致性损失以鼓励模型学习更深层的语义信息。例如,基于FixMatch的一系列改进工作[7-9],要求样本的强增强视图与弱增强视图的预测结果相似。这些方法在自然图像分类任务中表现出色,但其依赖的强增强策略不适用于医学影像数据。针对自然图像设计的强增强方式可能削弱甚至改变医学影像的语义,从而导致模型性能在医学数据集上显著下降。另一类基于MeanTeacher改进的工作[10-11]采用较弱的增强方式,并通过添加网络扰动生成不改变图像语义的版本,从而计算一致性损失。这类方法避免了强增强策略对医学影像语义的破坏。然而,随着模型的迭代更新,教师模型和学生模型可能逐渐趋于同质化[12],限制了方法的性能进一步提升。

在分类任务中,类别不平衡的数据集容易导致分类器在训练过程中产生预测偏差[13],从而倾向于多数类,无法充分学习少数类的特征。这一问题是深度学习算法在实际应用中必须解决的重要挑战。针对数据集不平衡问题,目前主要的解决方案包括重采样[14]、加权函数[15]和数据生成[16]等方法。重采样通过对少数类样本进行过采样,以增加少数类在训练数据中的比例,缓解数据不平衡的问题。事实上,由于无法提前准确获知样本分布,重采样的比例难以把握:过采样可能导致模型对噪声更为敏感,而欠采样则可能丢失重要信息;使用权重函数的方法为不同类别分配不同的权重,从而平衡模型在训练过程中对各类别的关注程度。这种方法需要人工设置参数,具体效果在很大程度上依赖于数据集的特性,因而泛化性能难以保证;数据生成方法使用生成对抗网络(GAN)[16]通过对抗训练生成少数类样本,从而提升这些类别在训练数据中的比例。尽管生成网络能够生成高质量的合成样本,并显著改善模型性能,但若生成数据与真实数据过于接近,可能引发模型的过拟合问题。尤其是在少数类样本本身较少的情况下,生成网络的有效性可能受到限制。

针对上述问题,本文提出了用于对医学影像分类的基于混合对比的无偏半监督分类模型(Hybrid contrastive unbiased semi-supervised classification model,HCUM)。该模型包含两个关键模块,混合对比模块与自适应去偏模块。混合对比模块通过对有标签数据和无标签数据进行类级特征对比与实例级特征对比,深度挖掘两类数据的潜在特征,从而提升模型对类间相似医学影像的分类精度。自适应去偏模块通过反事实推理,解耦样本预测中的直接影响和间接影响,有效消除类别不平衡对分类器造成的预测偏差,进一步

生成无偏的预测分数。与同领域的其他方法相比,HCUM在多个不同模态的医学影像数据集上的分类任务中表现出更优异且更加平衡的分类性能,充分验证了其有效性和适用性。

1 模型描述

本文提出的基于混合对比的无偏半监督分类模型,包含混合对比模块和自适应去偏模块。下面将详细介绍半监督学习基本框架与两个创新模块的设计。

1.1 半监督分类框架

在半监督学习的范式下,本文将数据集划分为两个部分,有标注数据集与无标注数据集,往往无标注数据集的数据量远大于有标注数据集,即$M \gg N$。本文提出的半监督学习框架由基于一致性正则化[2]机制的半监督基础框架改进而来。该基础框架通过对相同的输入施加各种扰动并鼓励模型对相同输入的不同扰动视图做出相似的预测,进而探索未标记数据中的隐含信息。本文提出的方案在该基础框架上加入了额外的针对有标注数据与无标注数据的对比损失,以进一步引导模型更好地训练。在损失构建时,定义有标签样本的监督学习损失为:

$$L_{\text{sup}}=\frac{1}{N}\sum_{i=1}^{N}H\left(y_i,P\left(y|\eta\left(x_i\right)\right)\right) \tag{1}$$

式中:$H(\cdot,\cdot)$表示交叉熵损失;η代表弱增强;$P(\cdot)$为模型输出的预测概率。

对于无标签样本的无监督损失L_{unsup},进行如下定义:

$$L_{\text{sup}}=\frac{1}{N}\sum_{j=1}^{M}MSE\left(P\left(\eta^1\left(xj\right)\right),P\left(\eta^2\left(xj\right)\right)\right) \tag{2}$$

式中:η^1和η^2分别表示两种不同的弱增强方式;$MSE(\cdot,\cdot)$表示均方误差(Mean squared error,MSE)。

除了上述常规的损失设置,本文额外设计了针对有标签数据和无标签数据的对比学习损失和,该损失由本文提出的混合对比模块生成。

在半监督框架下,输入样本经过两种增强后得到增强后的视图x_i^1和x_i^2,两个增强视图经过骨干网络的特征抽取后得到对应的样本表征$\tilde{x}_i^1$和$\tilde{x}_i^2$。之后,样本表征将分别进入分类分支和混合对比分支。在分类分支中,样本表征通过分类头得到对应的预测概率p_1和p_2,再由自适应去偏模块(Adaptive de-deviation module,ADM)进行去偏,得到无偏的样本预测概率$\hat{p}_1$和$\hat{p}_2$。若样本x_i为有标签样本时,计算其监督学习损失$L_{\text{sup}}=L_{\text{sup1}}+L_{\text{sup2}}$,若样本$x_i$为无标签样本,计算其无监督损失$L_{\text{unsup}}$。在混合对比分支中,$\tilde{x}_i^1$和$\tilde{x}_i^2$进入混合对比模块(Hybrid contrastive module,HCM),根据x_i是否为有标签样本,构造对比学习损失$L_{\text{ct_l}}$或$L_{\text{ct_u}}$。

综上,最终的损失L_{total}可以写成:

$$L_{\text{total}}=L_{\text{sup}}+\lambda\left(L_{\text{unsup}}+L_{\text{ct_l}}+L_{\text{ct_u}}\right) \tag{3}$$

式中:λ为一个上升权重因子,用于控制无监督损失与其他损失之间的权重比例,其具体数值由当前样本置信度通过指数移动平均(Exponential moving average,EMA)进行调整:

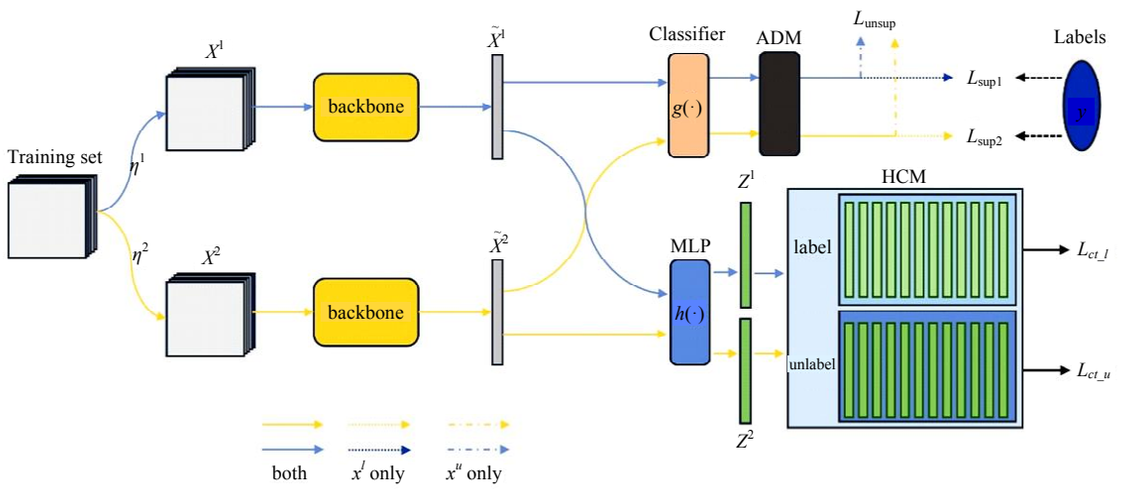


图1 基于混合对比的无偏半监督分类模型

Fig. 1 Framework diagram of hybrid contrastive unbiased semi-supervised classification model

$$\lambda_t=\begin{cases}\bar{p}, \text{ while } t=0\\ \alpha\lambda_{t-1}+(1-\alpha)\bar{p}_t, \text{ others}\end{cases} \tag{4}$$

式中：λ_t为训练第t个迭代时的λ取值；$\bar{p}_t$为训练第t个迭代时模型对于输入的一个批次无标签样本的平均置信度；α为动量系数，在本文实验中设定为0.99。

1.2　混合对比模块

为充分挖掘无标签数据中丰富的隐含信息，本文设计了混合对比模块，以增强模型对特征的学习能力。该模块通过投影头将样本映射到特征空间，采用对比学习策略对样本特征进行深度对齐和优化。针对有标签数据与无标签数据，该模块分别进行处理：有标签数据用于类级特征对比，以强化类别间的区分能力；无标签数据则通过实例级特征对比，捕捉样本间的细粒度特征关系。这种双重对比机制有效提升了模型对类间相似性和数据特征复杂性的适应能力。

对于输入模型的一批数据，其中包含批次大小为B的有标签数据与批次大小为μB的无标签数据，该模块将采用不同的处理逻辑。

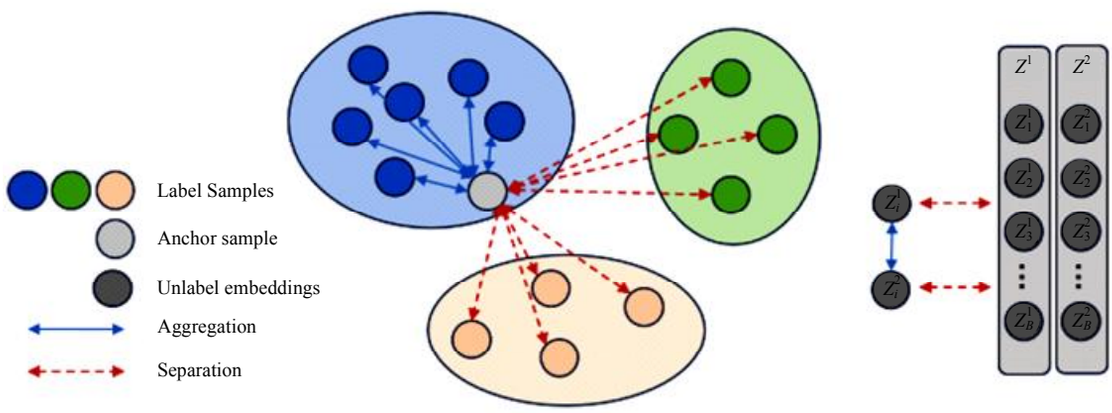


图2　混合对比模块

Fig. 2　Hybrid contrastive module

具体来说，用$X^l=\left\{(X_i^{(1,2)},y_i)\right\}_B$表示有标签样本经过增强后的视图，$Z=\{z_i^{(1,2)}\}_B=z_1^1,z_1^2,\cdots,z_B^1,z_1^2]$表示通过特征投影头$h(\cdot)$后对于某一个样本的特征嵌入$z_i^1$，定义其正样本为与其类别相同的样本的特征嵌入（包括样本另一视图z_i^2），即$z^+=\{z_j|y_j=y_i=c\}$，其中c为样本的类别。通过在特征空间中拉近正样本之间的距离并推离负样本，构造类级对比损失，鼓励模型在类级特征空间学习不同类别的特征。

对于有标签数据的类级对比损失定义如下：

$$L_{ct_1}=-\frac{1}{2B}\sum_{z_i\in Z}\frac{1}{2B_c-1}\times\sum_{z_j\in Z^+}\log\frac{\exp(z_i\cdot z_j/\tau)}{E^l} \tag{5}$$

$$E^l=\sum_{c\in C}\frac{1}{2B_c-1}\sum_{z_k\in Z^c}\exp(z_i\cdot z_j/\tau) \tag{6}$$

式中：B为批次的大小；B_c为一个批次中类别为c的样本数量；Z^+为样本z_j的所有正样本组成的集合；Z^c为这一批次中类别标签为c的样本集合；τ为温度参数。

对于增强后的无标签样本$X^u=\{x_i^{(1,2)}\}_{\mu B}$，使用$Z^{(1,2)}=\{z_i^{(1,2)}\}_B=\{z_i^1,z_i^2,\cdots,z_{\mu B}^1,z_{\mu B}^2\}$表示其特征嵌入，因为没有类别信息，选择另一种正/负样本构造方式，一批数据的每个样本x_i通过特征投影头获得的z_i^1、z_i^2互为正样本，同时与同批中其他样本的特征嵌入互为负样本，因此无标签样本的样本级对比损失定义如下：

$$L_{ct_u}=-\frac{1}{\mu B}\sum_{z_i\in Z}\log\frac{\exp(z_i^1\cdot z_i^2/\tau)}{E^u} \tag{7}$$

$$E^u=\sum_{z_i\in Z}\left(\exp(z_i^1\cdot z_i^2/\tau)+\exp(z_i^1\cdot z_i^1/\tau)\right) \tag{8}$$

最终混合对比模块的损失为：

$$L_{ct}=L_{ct_l}+L_{ct_u} \tag{9}$$

1.3　自适应去偏模块

医学数据集因疾病种类发生概率的差异和数据采集过程中的限制，常存在类别不平衡问题。为减轻这一问题对模型分类性能的影响，本文提出基于因果推理的自适应去偏模块，旨在无需依赖边缘类别分布的先验知识，动态减轻数据不平衡对模型训练的负面影响。因果推理通过基于事实前提推导反事实结论，为变量间的干预效应建

模提供理论基础。本文采用因果图模型表示输入样本对模型预测的直接和间接影响,并对导致分类器产生偏差的样本间关系进行建模。因果推理已被广泛应用于消除混淆效应、校正选择偏差及解耦直接和间接效应等场景[17,18]。本文在此基础上针对医学数据分类任务提出创新的改进方案。

对导致分类器产生偏差的样本间影响进行因果图建模,如图3所示,其中M表示模型偏差,D表示中介变量(预测模型),x表示输入样本,Y表示预测结果。图3中左侧因果图代表样本x_i在模型偏差M的影响下对于预测结果Y的受控直接效应($x_i \to Y$)与间接影响($D \to Y$)。在固定中介变量D(表示为$do(D)$)的条件下,根据CDE(Controlled direct effect)[19]的定义,当样本为x_i(表示为$do(x_i)$)时,预测为Y的受控直接效应可以计算为样本取x_i($do(x_i)$)表现出的反事实结果与样本取所有样本$X=\{x_1,\cdots x_N,\}$(即$do(x)$)表现出的反事实结果的差异。因此可以计算输入样本x_i对预测Y的受控直接影响CDE为:

$$CDE(Y_i)=[Y_i|\mathrm{do}(x_i),do(D)]-[Y_i|do(X),do(D)] \tag{10}$$

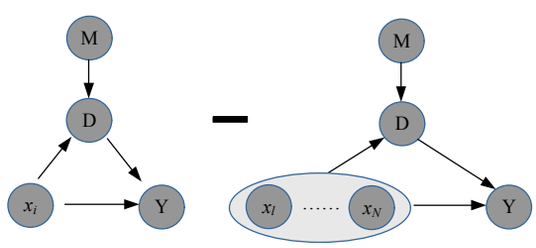


图3 反事实推理去偏因果图

Fig. 3 Causal graph of debiasing with counterfactual reasoning

在模型的训练过程中,单个样本的反事实推理结果$[Y_i|do(x_i),do(D)]$可以表示为样本x_i的预测概率分数,而所有样本的反事实结果$[Y_i|do(X),do(D)]$代表其余所有样本的预测概率分数的类均值。

然而,直接计算所有样本的反事实结果会显著增加模型训练的计算成本,因此本文采用了一种高效的近似方案。具体而言,通过计算每个批次样本的反事实结果,并在批次之间使用指数移动均值EMA动态更新,从而近似估计所有样本的反事实结果。该方法不仅降低了计算开销,还有效保留了模型校正预测分数中模型偏差的能力。因此,样本x_i无偏的预测分数$\hat{p}$可以由下面公式计算得到:

$$\hat{p}_i^{(t+1)}=f\left(\eta(x_i^{(t+1)})\right)-\alpha\log\bar{p}^{t} \tag{11}$$

$$\bar{p}_i^{t(t+1)}=m\bar{p}^{(t-1)}+(1-m)\frac{1}{B}\sum_{k=1}^{B}p_k \tag{12}$$

式中:$f\left(\eta(x_i^{(t+1)})\right)$代表样本$x_i^{t+1}$弱增强视图的logits;$\bar{p}^{t}$为在第$t$个迭代时,本文近似估计的所有样本反事实结果,然后使用一个log函数来调节$\bar{p}^{t}$,使之匹配logits的尺寸;α为去偏因子,用于控制间接因素的影响程度;$m\in[0,1)$为动量系数;p_k为样本x_k通过softmax函数获得的概率分布。

通过公式(11)将模型对每个样本的预测消除模型偏差,最终获得去偏的概率预测,进一步引导模型向更加准确的方向进行训练。

2 实验结果及分析

2.1 数据集

本文使用4个公开医学数据集来验证本算法的分类性能。数据集分辨涵盖了多种不同模态下的分类任务,包括ISIC 2018[20]数据集、Chest X-ray14[21]数据集和RSNA ICH[22]数据集。为了进一步验证本文提出的模型的性能,本文额外在通用图像分类数据集CIFAR-10/100[23]和STL-10[24]上进行性能对比,数据集的细节见表1。

表1 数据集参数

Table 1 Dataset parameters

数据集	图片数量	类别数量	图片大小	模态
ISIC 2018	10 015	7	尺寸变化	皮肤镜图片
ChestX-ray14	112 120	14	1 024×1 024	X光图片
RSNA ICH 2019	25 000	6	512×512	CT图片
CIFAR10	60 000	10	32×32	RGB图像
CIFAR-100	60 000	100	32×32	RGB图像
STL-10	113 000	10	96×96	RGB图像

2.2 评价指标

为了验证本文模型的分类性能,实验采用了多项评估指标进行综合评估,包括:召回率(Recall)、特异度(Specificity)、准确率(ACC)、ROC曲线下面积(AUC)、F_1分数。这些指标在分类模型投入到临床实践中具有非常重要的现实意义,对应的计算方法如下:

$$\mathrm{Recall}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} \tag{13}$$

$$\mathrm{Specificity}=\frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} \tag{14}$$

$$\mathrm{ACC}=\frac{\mathrm{TP}+\mathrm{TN}}{\mathrm{TP}+\mathrm{TN}+\mathrm{FP}+\mathrm{FN}} \tag{15}$$

$$\mathrm{AUC}=\sum(\mathrm{FPR}_{i+1}-\mathrm{FPR}_i)\times\frac{\mathrm{TPR}_i+\mathrm{TPR}_{i+1}}{2} \tag{16}$$

$$\mathrm{Precision}=\frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{F}P} \tag{17}$$

$$F_1=2\times\frac{\mathrm{Precision}\times\mathrm{Recall}}{\mathrm{Precision}+\mathrm{Recall}} \tag{18}$$

式中:TP(True Positive)为正例预测为正例的个数;TP(False Positive)为负例预测为正例的个数;TN(True Negative)为正例预测为负例的个数;FN(False Negative)为负例预测为负例的个数;FPR和TPR分别为假阳性率与真阳性率。

在医疗诊断的背景下召回率(Recall)用于衡量模型对某一类别样本的召回能力,更高的召回率代表疾病的漏诊率更低;特异度(Specificity)用于衡量模型正确识别某类以外类别的能力,更高的特异度可以降低误诊率;准确率(ACC)用于衡量模型整体分类的正确率,表示所有样本中被正确分类的比例;ROC曲线下面积(AUC)在多种疾病分类任务中可以反映模型对每个类别的整体区分能力,尤其适用于类别之间样本分布不均衡的场景;F_1分数因其能综合模型对少数类的精确率和召回率,特别适用于类别分布不平衡任务性能的测定。

2.3　实验设置

该模型设计基于PyTorch深度学习框架,使用8张NVIDIA GeForce RTX 3090显卡。使用DenseNet[25]系列网络作为骨干网络,特征嵌入维度设为128。使用SGD作为优化器,权重衰减设为1e-4。初始学习率设置为0.002,并通过余弦时间表衰减。对于ISIC 2018数据集批次大小设置为128,Chest X-ray14数据集设置为64,训练的总周期统一设置为100。超参数α、β和γ分别设为0.5、1和1。使用ResNet[26]针对ImageNet数据集[27]的默认数据增强策略作为分类分支的η^1。将η^1中的随机灰度、旋转和垂直翻转作为η^2来丰富数据表示,每组配置进行3次实验,取最优结果。

2.4　实验结果

2.4.1　ISIC 2018皮肤病灶分类数据集实验结果

ISIC 2018分类数据集一共包含10 015张变化尺寸的皮肤病变区域图片,共包含7类皮肤病。本文将数据集按70%训练集、10%验证集和20%测试集的比例划分。其中训练集进一步分为20%标注数据和80%无标注数据。数据加载时统一将图片尺寸调整为224×224。使用在ImageNet数据集上进行预训练的DenseNet121[25]作为骨干网络。表2列出了各算法在该数据集上的实验对比结果。

表2　不同半监督分类模型在ISIC 2018数据集上的实验结果

Table 2　Experimental results of different semi-supervised classification models on the ISIC 2018 dataset

方法	比例		性能指标/%				
	有标签/%	无标签/%	AUC	Recall	Specificity	Accuracy	F_1
监督学习	100	0	95.43	75.20	94.94	95.10	70.13
SRC-MT[11]	20	80	93.58	71.47	92.72	92.54	60.68
FedIRM[28]	20	80	92.46	—	93.29	92.89	55.81
AIFS[29]	20	80	94.02	69.94	91.78	92.61	62.08
HCUM(本文)	20	80	94.28	72.01	92.90	93.07	64.21

由表2结果可知,本文方案在所有指标上达到了最优性能,特别是召回率(Recall)的明显提升意味着在临床过程中漏诊情况减少,这对病情的早发现和早治疗非常重要。特异度(Specificity)的提升可以在辅助诊断阶段减少健康病例误诊的可能性。F_1分数达64.21%,相较于其他半监督模型的最佳性能提升了3.53%,证明本文方案在类别不平衡数据集的各类都表现出了更优异的性能。

2.4.2　Chest X-ray14 X光胸片数据集实验结果

Chest X-ray14数据集包含112 120张胸部X光图片,其中每个X光片都标有14种常见胸部疾病的一个或多个类别。同样对该数据集按照70%训练集、10%验证集与20%测试集的比例进行划分。将尺寸为1024×1024的原始图像调整为384×384。训练集同样分为20%标注数据与80%无标注数据。因为数据量的增大,针对该数据集采用在ImageNet数据集上进行预训练的DenseNet169[25]网络作为骨干网络。

由表3中的结果可以看出,本文提出的模型在14个类别中9个类别相对于其他算法实现了性能的提升,尤其是对一些较难学习类(如Infiltration、Pleural Thickening和Pneumonia)在AUC指数上都实现了提升,对于少样本的类别Hernia、Fibrosis和Pneumonia分别实现了明显的性能的提升。这证明了本文提出的混合对比无偏半监督医学分类模型在医学影像分类任务中的有效性。

表3 不同半监督分类模型在Chest X-ray14数据集上的AUC指标结果

Table 3 Experimental results of different semi-supervised classification models on the Chest X-ray14 dataset with AUC

方法	监督学习	SRC-MTS[8]	AIFS[27]	HCUM
有标签比例/%	100	20	20	20
无标签比例/%	0	80	80	80
Atelectasis	77.32	75.38	76.12	75.93
Cardiomegaly	88.85	87.70	88.06	88.11
Effusion	82.11	81.58	81.77	81.55
Infiltration	70.95	70.40	70.57	70.57
Mass	82.92	78.03	78.86	79.33
Nodule	77.00	73.64	74.23	73.68
Pneumonia	71.28	69.27	69.56	69.56
PneumothoraX	86.87	86.12	86.32	86.45
Consolidation	74.88	73.11	73.84	73.35
Edema	84.74	82.94	83.13	83.37
Emphysema	93.35	88.98	90.02	89.85
Fibrosis	84.46	79.22	80.43	81.05
PleuralThickening	77.34	75.63	75.61	75.83
Hernia	92.51	87.27	87.86	89.46
Average AUC	81.75	79.74	79.74	79.86

2.4.3 RSNA ICH CT数据集实验结果

使用的RSNA ICH检测数据集包含ICH疾病的5种亚型和正常的CT切片共超过25 000张。同样采用与ISIC 2018数据集相同的数据集划分方式对该数据集进行划分。

由表4中的结果对比可以发现,在RSNA ICH数据集上,HCUM模型在AUC、特异性、准确率和F_1分数等指标上均优于FedIRM[28]模型。通过F_1分数和AUC的提升可以看出,HCUM在不过度牺牲特异性(减少健康患者的误判)的前提下,显著优化了其他核心指标,这对实际部署时的医患信任度至关重要。

表4 不同半监督分类模型在RSNA ICH数据集上的实验结果

Table 4 Experimental results of different semi-supervised classification models on the RSNA ICH dataset

方法	AUC	Recall	Specificity	Accuracy	F_1
FedIRM	87.56	—	91.56	88.89	59.86
HCUM	88.79	65.76	87.58	89.33	65.28

2.4.4 通用图像分类数据集的实验结果

CIFAR-10、CIFAR-100和STL-10是通用图像分类任务中广泛使用的标准数据集,常用于评估和比较不同算法的性能。上述数据集同样为半监督学习算法的评估提供了标准化的基准。在该部分实验中,本文采用不同的数据集划分方式,针对每个数据集的训练集,只为每一类分配4个有标签数据,其余都作为无标签数据辅助半监督学习分类模型的训练。

表5的实验结果表明,本文提出的HCUM模型在通用图像数据集上展现出与主流半监督方法相当的分类性能。在CIFAR10数据集上,HCUM的准确率超过FixMatch和MPL,虽略低于FlexMatch和FreeMatch,但仍保持竞争力。在更复杂的CIFAR100任务中,HCUM相较表中其他方法提升显著,但与当前通用数据集上最优的半监督学习方法FreeMatch相比仍存在一些差距。值得注意的是,在STL-10数据集上HCUM取得70.75%的准确率,优于FixMatch和MPL,展现出对更高分辨率图像的适应性。这些结果验证了HCUM在跨领域图像分类任务中的泛化能力,其医学领域特化设计未对通用性能造成显著影响。

表5 不同半监督分类模型在通用数据集上的实验结果

Table 5 Experimental results of different semi-supervised classification models on the common dataset

方法	CIFAR10	CIFAR100	STL-10
FixMatch[7]	92.53	53.58	64.03
MPL[30]	93.38	53.74	64.24
FlexMatch[8]	95.03	60.06	70.85
FreeMatch[9]	95.1	62.02	84.44
HCUM	94.57	60.10	70.75

2.4.5 消融实验

为验证本文提出的混合对比模块与自适应去偏模块的有效性,针对混合对比模块(HCM)和自适应去偏模块(ADM)的有效性在ISIC2018数据集上进行消融实验,实验设置使用与表2相同的

实验设置,使用20%的标注数据进行消融实验。消融实验共设置4个组别。

(1)不包含本文提出的两个模块的基础模型(Base)。该组别将删除分类分支中的ADM模块和整个混合对比分支,损失仅包含基础的监督损失与无监督损失:

$$L_{total}=L_{sup}+\lambda L_{unsup} \tag{19}$$

(2)只添加混合对比模块的Base模型。该组别删除了ADM模块,但是保留了混合对比分支,其损失形式如下:

$$L_{total}=L_{sup}+\lambda(L_{unsup}+L_{ct_l}+L_{ct_u}) \tag{20}$$

(3)只添加自适应去偏模块的Base模型。该组别保留了ADM模块,但是删除了整个混合对比分支,因此其损失形式为:

$$L_{total}=L_{sup}+\lambda L_{unsup} \tag{21}$$

(4)本文提出的基于混合对比的无偏半监督分类模型(HCUM),其损失形式与公式(3)相同。消融实验结果如表6所示。

表6　不同对照组在ISIC 2018数据集上的实验结果

Table 6　Experimental results of different control groups on the ISIC 2018 dataset

方法	AUC	Recall	Specificity	Accuracy	F1
Base	90.20	67.05	92.11	92.25	53.04
Base with HCM	94.22	67.86	92.19	92.27	59.49
Base with ADM	91.28	71.20	92.43	92.42	60.01
HCUM (ADM&HCM)	94.28	72.01	92.90	93.07	64.21

由表6中结果可以看出,单独增加自适应去偏模块对召回率(Recall)和特异度(Specificity)分别提升了4.15%和0.32%,因为自适应模块通过消除模型的预测偏差获得了无偏的样本预测概率,提升了对各类样本学习的准确性。单独增加混合对比模块对AUC、F_1和准确率(Accuracy)提升明显,说明本文提出的混合对比模块对有标签和无标签数据的隐含特征提取有明显效果,使模型学习到了更多的隐含特征,从而提升了模型的整体分类性能。由此可见,本文提出的两个模块都能针对不同的方面对模型进行提升,证明了本文提出的方法的有效性。

3　结束语

本文针对医学影像数据提出了一个基于混合对比的无偏半监督分类模型。该模型包含的混合对比模块通过对标注数据与无标注数据分别进行特定的对比学习,对有标签数据与无标签数据的隐含特征进行抽取和学习,实现了对类间高度相似的医学影像数据的优异分类性能。针对医学影像数据集普遍存在的类间不平衡问题,提出了自适应去偏模块,通过反事实推理解耦不平衡数据集导致的类间间接影响,获取样本的无偏预测,从而引导模型更好地训练。实验结果表明,本文提出的模型在多个模态的医学数据集上表现出稳定且平衡的分类性能,较其他同类模型取得了显著提升,尤其在难分类和少数类别的识别上展现了明显优势。

本文提出的模型虽然在医学影像分类工作中表现出了较强的性能,但是目前只能针对静态图像数据进行分类,在实践过程中无法针对原始数据进行端对端的结果输出。后续工作将集中研究和开发模型对于多种模态图像数据(如视频和DICOM数据)的处理能力。

参考文献:

[1] Lee Dong-Hyun. Pseudo-label: the simple and efficient semi-supervised learning method for deep neural networks[C]//IICML 2013 Workshop on Challenges in Representation Learning (WREPL), Atlanta, Georgia, USA, 2013,3: 896.

[2] Laine Samuli, Timo Aila. Temporal ensembling for semi-supervised learning[J]. arXiv,2016,1610.02242.

[3] Berthelot David, Nicholas Carlini , Ian Goodfellow, et al. Mixmatch: a holistic approach to semi-supervised learning[C]//33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada,2019.

[4] Solatidehkordi Zahra, Imran Zualkernan. Survey on recent trends in medical image classification using semi-supervised learning[J]. Applied Sciences, 2022, 12(23): 12094.

[5] Amine Amyar, Modzelewski Romain, Li Hua, et al. Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: classification and segmentation[J]. Computers in Biology and Medicine, 2020, 126: 104037.

[6] Han Kai, Sheng Victor S, Song Yu-qing, et al. Deep semi-supervised learning for medical image segmentation: a review[J]. Expert Systems with Applications, 2024, 245: 123052.

[7] Sohn K, Berthelot D, Li C L,et al. Fixmatch: simplifying semi-supervised learning with consistency and confidence[C]//34th Conference on Neural Informa-

tion Processing Systems (NeurIPS 2020), Vancouver, Canada, 2020.

[8] Zhang B, Wang Y, Hou W, et al. FlexMatch: boosting semi-supervised learning with curriculum pseudo labeling[J]. Advances in Neural Information Processing Systems, 2021, 34: 18408-18419.

[9] Wang Y, Chen H, Heng Q, et al. FreeMatch: self-adaptive thresholding for semi-supervised learning[J]. arXiv preprint arXiv: 2205. 07246.

[10] Tarvainen A, Valpola H. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, California, USA, 2017: 1195-1204.

[11] Liu Q, Yu L, Luo L, et al. Semi-supervised medical image classification with relation-driven self-ensembling model[J]. IEEE Transactions on Medical Imaging, 2020, 39(11): 3429-3440.

[12] JWu J, Fan H, Zhang X, et al. Semi-supervised semantic segmentation via entropy minimization[J]. Advances in Neural Information Processing Systems, 2022, 35: 2803-2816.

[13] Cui Y, Jia M, Lin T Y, et al. Class-Balanced Loss Based on Effective Number of Samples[C]//In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 9268-9277.

[14] Fernandez A, Garcia S, Chawla N V, et al. SMOTE for learning from imbalanced data: Progress and challenges, Marking the 15-year anniversary[J]. Journal of Artificial Intelligence Research, 2018, 61: 863-905.

[15] Chen J, Liu D, Luo B, et al. Residual objectness for imbalance reduction[J]. Pattern Recognition, 2022, 130: 108781.

[16] Qu L, Zhu H S, Zheng R Q, et al. ImGAGN: imbalanced network embedding via generative adversarial graph networks[C]//Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining. Online:ACM, 2021:1390-1398.

[17] Bareinboim E, Pearl J. Controlling selection bias in causal inference[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2011, 25(1): 1754-1755.

[18] Zhang D, Zhang H, Tang J, et al. Causal intervention for weakly-supervised semantic segmentation[J]. Advances in Neural Information Processing Systems, 2020, 33: 655-666.

[19] Richiardi L, Bellocco R, Zugna D, et al. Mediation analysis in epidemiology: methods, interpretation and bias[J]. International Journal of Epidemiology, 2013, 42(5): 1511-1519.

[20] Codella N C F, Gutman D, Celebi E, et al. Skin lesion analysis toward melanoma detection: A challenge at the 2017 International symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)[C]//2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 2018, pp. 168-172,

[21] Wang X, Peng Y, Lu L, et al. ChestX-ray8: hospital-scale chest X-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases[J]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2097-2106.

[22] RSNA Intracranial Hemorrhage Detection Challenge [EB/OL]. 2019-07-26. https://www.kaggle.com/c/rsna-intracranial-hemorrhage-detection

[23] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images[J]. Handbook of Systemic Autoimmune Diseases, 2009, 1(4): 7.

[24] Coates A, Ng A Y, Lee H. An analysis of single-layer networks in unsupervised feature learning[J]. Journal of Machine Learning Research, 2011, 15: 215-223.

[25] Huang G, Liu Z, Laurens V D M, et al. Densely Connected Convolutional Networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 2017: 4700-4708.

[26] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 2016: 770-778.

[27] Deng J, Dong W, Socher R, et al. ImageNet: a large-scale hierarchical image database[C]//2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009: 248-255.

[28] Liu Q, Yang H, Dou Q, et al.. Federated semi-supervised medical image classification via inter-client relation matching[C]//Medical Image Computing and Computer Assisted Intervention-MICCAI 2021, Strasbourg, France, 2021: 325-335.

[29] Zhou Z, Lu C, Wang W, et al. Semi-supervised medical image classification based on attention and intrinsic features of samples[J]. Applied Sciences, 2022, 12(13): 2076-3417.

[30] Hieu Pham, Dai Zi-hang, Xie Qi-zhe, et al. Meta pseudo labels[C]//In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 2021: 11557-11568.