

基于BIRADs多任务学习模型的乳腺肿块分类

吴书裕¹, 周露¹, 王琳婧¹, 李慧君¹, 张书旭¹, 梅颖洁²

1. 广州医科大学附属肿瘤医院放疗科, 广东广州 510095; 2. 广东省人民医院放射科, 广东广州 510080

【摘要】目的:为解决乳腺图像肿块分类与深度学习应用的难题,提出一种基于乳腺影像报告与数据系统(BIRADs)多任务学习模型的肿块分类方法。**方法:**构建迁移学习的形态学和纹理特征提取器,并在此基础上引入多任务分类器,实现BIRADs诊断相关的边缘、形状、密度和微小性评估。研究通过训练策略、输入图像和模型架构系列实验和指标,分析评估模型性能。**结果:**在迁移学习策略下,Base模型和BIRADs模型性能均有显著提升。原始肿块图像作为输入的模型性能均优于掩模图像模型。在迁移学习和原始肿块输入下,BIRADs模型相较Base模型有更高的AUC值(0.830 vs 0.793)、准确率(0.747±0.024 vs 0.712±0.023)、精确率(0.643±0.032 vs 0.607±0.030)、召回率(0.774±0.037 vs 0.715±0.042)、F1-score(0.702±0.028 vs 0.656±0.029)。多任务学习模型在乳腺肿块分类中具有显著优势。**结论:**BIRADs多任务学习模型结合临床知识与数据驱动方法显著提高肿块分类准确性和模型鲁棒性,有望提高乳腺癌诊断准确性。

【关键词】乳腺癌; 肿块分类; 数据驱动; 迁移学习; 多任务学习

【中图分类号】R318; R445

【文献标志码】A

【文章编号】1005-202X(2023)10-1220-08

Breast mass classification based on BIRADs multi-task learning model

WU Shuyu¹, ZHOU Lu¹, WANG Linjing¹, LI Huijun¹, ZHANG Shuxu¹, MEI Yingjie²

1. Department of Radiotherapy, Affiliated Cancer Hospital and Institute of Guangzhou Medical University, Guangzhou 510095, China;

2. Department of Radiology, Guangdong Provincial People's Hospital, Guangzhou 510080, China

Abstract: Objective To propose a breast mass classification method based on multi-task learning model of breast imaging reporting and data system (BIRADs) for addressing the challenges in breast mass classification and deep learning applications. **Methods** A BIRADs multi-task learning model with a transfer learning-based morphological feature extractor, texture feature extractor, and multi-task classifier, which enables the assessment of BIRADs-related characteristics such as margin, shape, density and subtlety, was constructed. Model performance was analyzed through experiments involving training strategies, network architectures, and input types. **Results** With transfer learning as the training strategy, both Base and BIRADs models exhibit significant performance improvement. The models with original mass images as input outperformed those with masked images. With transfer learning as the training strategy and original mass images as input, the BIRADs model had higher AUC, accuracy, precision, recall rate, and F1-score as compared with Base model (0.830 vs 0.793, 0.747±0.024 vs 0.712±0.023, 0.643±0.032 vs 0.607±0.030, 0.774±0.037 vs 0.715±0.042, 0.702±0.028 vs 0.656±0.029, respectively). The multi-task learning model demonstrated significant advantages in breast mass classification. **Conclusion** The proposed BIRADs multi-task learning model combining clinical knowledge and data-driven methods enhances accuracy and robustness in breast mass classification, which is expected to improve the diagnostic accuracy of breast cancer.

Keywords: breast cancer; mass classification; data-driven; transfer learning; multi-task learning

前言

乳腺癌作为全球女性头号癌症,对女性的健康构成重大威胁,乳腺癌防治关键在于早期乳腺病变的发现和及时治疗^[1-2]。数字乳腺X射线摄影(Digital Mammography, DM)因其对早期乳腺癌的高敏感性,从而作为一种广泛应用于乳腺癌筛查和诊断的影像学技术^[3-4]。而肿块作为乳腺的主要病变之一,在乳腺癌的早期指征中起着重要作用^[5-6],其在

【收稿日期】2023-06-01

【基金项目】广东省自然科学基金(2020A151110577);广东省医学科研基金(A2023291)

【作者简介】吴书裕,博士,研究方向:医学图像处理与肿瘤放射物理, E-mail: wsyeasy@outlook.com

【通信作者】梅颖洁,博士,研究方向:医学成像技术与图像处理, E-mail: yingjie.mei@gmail.com

DM影像中可能呈现出形状、密度、边界和纹理等多种复杂特征的变化^[7]。因此,基于DM的肿块精准分类仍是一项具有挑战性的任务。DM的计算机辅助诊断系统对可疑乳腺病变提供自动化和客观评估,协助放射科医生做出准确的诊断决策^[8-11]。相关系统通常采用一系列图像处理技术、特征提取方法和分类算法,以分析DM图像及其异常特征,并区分良性肿块和恶性肿块等病变。随着深度学习数据驱动技术在自然图像领域的研究取得显著的进展,目前已应用于众多的医学图像CAD技术等研究中^[12-13]。然而,深度学习模型在医学图像分析中的有效性主要受几个方面的制约:(1)对于乳腺肿块及其他异常病变表现出广泛的变异性,增加了深度网络模型分类任务的复杂性和难度^[14];(2)深度学习是基于数据驱动的学习范式,在医学图像领域获取大规模、高质量的数据具有较大的挑战性,深度学习的性能很大程度上受制于标注数据的可用性^[15]。

对于乳腺肿块复杂的分类任务,引入可用的监督信息强化DM影像和肿块病变的理解能力,则有可能提高肿块的分类性能^[16]。目前较多研究如原始图像和掩模图像组合输入,可提高模型对异常病灶的特征学习能力^[17]。另外,也可通过多任务方式共同学习乳腺肿块相关诊断任务来提供额外的监督信息^[18],模型通过共享网络层学习任务间的内在关系,从而提高分类准确性和鲁棒性。此外,针对医学图像数据有限的问题,迁移学习被应用于乳腺癌病理学图像分类中^[19]。从源域学习可泛化特征并适应特定图像域,这种思路有望缓解数据驱动下乳腺肿块分类数据稀缺和模型泛化性的问题^[20-21]。若合理优化模型的架构、输入类型、学习策略,则有望改善其分类性能,为乳腺肿块计算机辅助分类的相关研究提供新的可能。基于此,本研究拟结合临床的乳腺影像报告与数据系统(Breast Imaging Reporting and Data System, BIRADs)的诊断评估,提出一种基于BIRADs多任务学习的乳腺肿块分类模型。该模型模拟临床医生观察和决策乳腺不同影像特征的过程,以提供有效的监督信息来加强肿块分类任务的性能。本研究还将探索不同的训练策略和图像输入类型组合,以构建通用的多任务训练框架。该研究有望为深度学习在其他乳腺病变的分类研究乃至其他医学辅助诊断场景提供有益的参考。

1 资料与方法

1.1 数据与预处理方法

本研究使用数字乳腺X射线摄影的精选乳腺成像子集(Curated Breast Imaging Subset of the Digital

Database for Screening Mammography, CBIS-DDSM)^[22],以保证研究结果的可靠性和可复现性。该数据集为本研究提供相对完整的组织病理学结果和BIRADs特征相关的标注信息,目前已广泛用于评估不同算法和方法在乳腺癌相关研究中。该数据集的采用可使本研究及分析结果更具代表性和可普适性。

图像预处理方面,对图像数据和标注进行必要的的数据预处理:首先,使用阈值化和形态学操作在相应的掩模上获取乳腺肿块的边界位置,将肿块感兴趣区域进行裁剪并调整至同一尺寸,以确保模型输入的一致性;然后,对裁剪后的图像进行像素值归一化,以解决来自不同设备和不同时段图像之间的强度差异;最后,采用如随机裁剪、翻转、旋转等数据增强技术,以增加数据集的多样性。通过上述步骤,获得肿块图像img。此外,将掩模图像进行同步后处理和数据增强,以获得对应肿块感兴趣区域(ROI)的msk,同时对肿块图像img利用掩模msk进行裁剪,以获得去除背景信息的乳腺肿块ROI图像imgmsk。

标注预处理方面,CBIS-DDSM数据集提供了乳腺肿块的良性或恶性分类和基于BIRADs特征的评估,但存在部分标注缺失或不规范等情况,由于边缘和形状等特征可能存在多个标签,使用以下准则对原始标注进行重新编码和处理:(1)Mass margin:使用值从0到5进行编码,分别为spiculated、ill defined、circumscribed、obscured、microlobulated和NAN(缺失值);(2)Mass shape:使用值从0到8进行编码,分别为lobulated、asymmetric breast tissue、round、irregular、architectural distortion、lymph node、oval、focal asymmetric density和NAN;(3)Breast density:使用值从1到4进行编码,表示乳腺密度的增加水平;(4)Subtlety:使用值从1到5进行编码,表示微小性的增加水平。

1.2 模型设计

本研究的模型架构如图1所示,主要由特征提取器和BIRADs多任务分类器组成。特征提取器采用了ResNet50^[23]深度神经网络作为迁移学习的干网络,该网络在自然图像分类任务中表现出色。为适应本研究,移除了ResNet的平均池化层和最终全连接层,但保留了相关卷积层作为特征提取器。该部分将加载ImageNet预训练的权重,并在乳腺肿块任务中进行Fine-tune,通过迁移学习的方法使其能够更好地适应乳腺图像的特定特征,进一步提高乳腺肿块的特征提取能力。对BIRADs肿块边缘、肿块形状、乳腺密度和微小性4个征象的评估进行分类。其中,肿块边缘和形状可以归类为肿块的形态学特征,将构建形态学特征提取器(Morphology Feature

Extrantor, MFE)进行病灶的相关特征提取,并在后续构建 margin task 和 shape task 子任务分支进行对应征象的评估;乳腺密度和微小性归类为肿块的纹理特

征,并将构建纹理特征提取器(Texture Feature Extrantor, TFE),再通过 density task 和 subtlety task 进行乳腺密度和微小性的BIRADs评分。

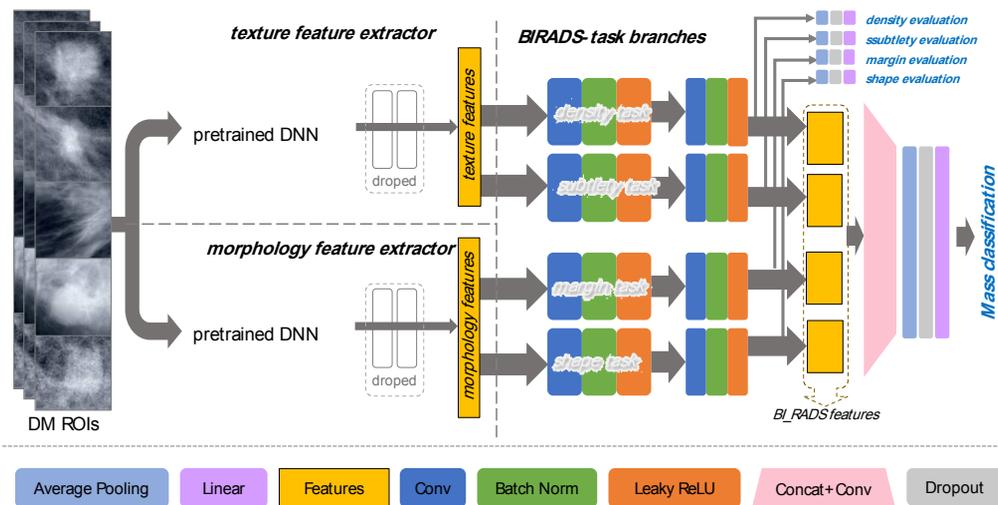


图1 基于BIRADs多任务学习的乳腺肿块分类模型框架示意图

Figure 1 Schematic diagram of breast mass classification framework based on BIRADs multi-task learning

上述的BIRADs任务分支由两个CBR block(卷积层、批归一化层和ReLU激活层)组成,提取对应的BIRADs子任务特征,任务分支后连接一个分类block(包括全局平均池化层、随机失活层、线性层),实现BIRADs子任务分类;与此同时,BIRADs子任务分支的对应特征,将通过concat操作组合成BIRADs features,并通过主分类器block实现乳腺肿块的良恶性分类。这种模型框架的目的是通过构建不同类型的特征提取器,并整合BIRADs相关征象的特征信息来增强主任务的分类性能。

1.3 损失函数设计

本研究BIRADs的4个子任务作为独立且非互斥的分类任务,使用二元交叉熵损失函数(BCELoss)作为BIRADs任务的损失函数。该损失函数适用于多标签分类,将每个BIRADs特征分类视为一个单独的二元分类任务,有助于有效学习各个特征的区别:

$$Loss^{BIRADs} = -\frac{1}{N} \sum_{i=1}^N [y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)] \quad (1)$$

其中, N 表示样本数, y_i 表示真实标签, \hat{y}_i 表示第*i*类的预测概率。

模型的主任务是乳腺肿块分类,作为一个独立且互斥的任务,采用交叉熵损失函数(Cross Entropy Loss)来衡量预测的类别概率与真实标签之间的差异:

$$Loss^{Main} = -\sum_{i=1}^N y_i \log(\hat{y}_i) \quad (2)$$

为了平衡主任务和BIRADs多任务对总体损失的贡献,适当地权衡了主任务损失 $Loss^{Main}$ 和辅助任务损失 $Loss^{BIRADs}_i$,得到了模型的总损失 $Loss^{Total}$ 。其计算公式如下:

$$Loss^{Total} = Loss^{Main} + \beta \sum_{i=1}^M Loss^{BIRADs}_i \quad (3)$$

其中, M 表示辅助BIRADs子任务的数量,在本研究中设定为4,而 β 表示辅助任务损失的权重。

1.4 实验设置

CBIS-DDSM数据集已划分了独立的训练集(1318个样本)和测试集(378个样本)。考虑本研究训练过程中的超参数调优和模型选择,从训练集中随机选择264个样本作为验证集,剩余的1054个样本作为新的训练集。测试集则保持不变,作为模型的性能评估。对网络特定层使用标准配置实现随机初始化。当训练策略为迁移学习时,特征提取器对应层面的初始权重是通过Imagenet预训练获得的^[23-24],该权重通过与主任务和BIRADs任务相关联的损失值进行权重更新和优化。所有模型均在NVIDIA RTX A4000 GPU上进行训练,为了平衡计算效率和模型收敛性,选择batch size为16,并使用RAdam优化器优化相关模型参数。初始学习率设置为0.00005,经过100个epochs训练,采用提前停止策略以防止过拟合。为了平衡主任务和BIRADs子任务损失的贡献,通过实验将Loss的权重系数 β 设置为0.5。

本研究将采用包括接收者操作特征(ROC)曲线,接收者操作特征曲线下面积(AUC),采用准确率(Accuracy)、精确率(Precision)、召回率(Recall)和F1-score指标评估模型的性能。其中,AUC表示模型区分正例和负例的能力。其他指标的公式如下所示:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \quad (4)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (5)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (6)$$

$$\text{F1 - score} = \frac{2\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (7)$$

其中,TP为真正例(True Positives),TN为真负例(True Negatives),FN为假正例(False Positives),FP为假负例(False Negatives)。

2 结果与分析

从训练策略、模型架构、输入类型3个方面分析和评估模型的肿块分类性能,相关实验和对应模型的配置如表1所示。表1中Base模型是基线网络,主要由ResNet50作为特征提取器加分类器block组成^[23],该模型主要以肿块良恶性鉴别作为分类任务。BIRADs为基于BIRADs多任务学习的乳腺肿块分类模型,该模型主要通过MFE和TFE进行形态学和纹理特征提取,其分类任务包括BIRADs多任务(margin、shape、density、subtlety评估)和良恶性分类主任务。训练策略方面,without TL表示模型未进行迁移学习(Transfer Learning, TL),对应层面采集默认初始化方式,TL and finetune则表示对应模型的特征提取层权重加载预训练权重,并在训练过程中微调。通过上述实验配置对6个模型进行训练和测试,得出结果如图2和表2所示。

表1 乳腺肿块分类模型相关实验及配置列表

Table 1 Experiments and corresponding configurations for breast mass classification models

| 模型名称 | 模型架构 | 任务类型 | 训练策略 | 输入类型 |
|----------|----------|------------------|-----------------|--------|
| Base_n | Base模型 | 肿块良恶性分类 | without TL | img |
| Base | Base模型 | 肿块良恶性分类 | TL and finetune | img |
| Base_m | Base模型 | 肿块良恶性分类 | TL and finetune | imgmsk |
| BIRADs_n | BIRADs模型 | BIRADs评估&肿块良恶性分类 | without TL | img |
| BIRADs | BIRADs模型 | BIRADs评估&肿块良恶性分类 | TL and finetune | img |
| BIRADs_m | BIRADs模型 | BIRADs评估&肿块良恶性分类 | TL and finetune | imgmsk |

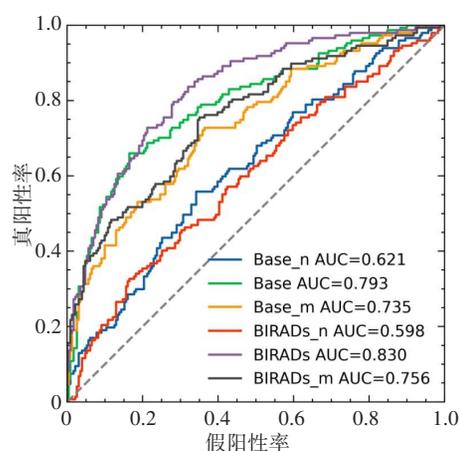


图2 不同训练策略、不同模型架构、不同输入类型对应模型的ROC曲线和AUC值比较

Figure 2 Comparison of ROC curves and AUC values for models with different training strategies, model architectures, and input types

表2 不同训练策略、不同模型架构、不同输入类型对应模型的评估指标值($\bar{x} \pm s$)

Table 2 Evaluation metrics for the models with different training strategies, model architectures, and input types ($Mean \pm SD$)

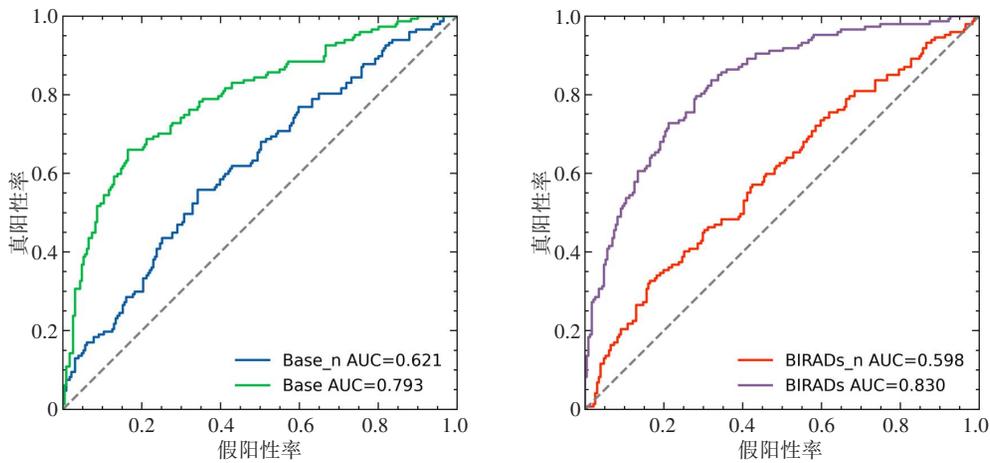
| 模型名称 | 准确率 | 精确率 | 召回率 | F1_score |
|----------|-------------|-------------|-------------|-------------|
| Base_n | 0.561±0.025 | 0.452±0.028 | 0.658±0.038 | 0.536±0.028 |
| Base | 0.712±0.023 | 0.607±0.030 | 0.715±0.042 | 0.656±0.029 |
| Base_m | 0.673±0.027 | 0.583±0.039 | 0.534±0.041 | 0.557±0.036 |
| BIRADs_n | 0.542±0.025 | 0.435±0.027 | 0.637±0.040 | 0.517±0.029 |
| BIRADs | 0.747±0.024 | 0.643±0.032 | 0.774±0.037 | 0.702±0.028 |
| BIRADs_m | 0.693±0.024 | 0.612±0.036 | 0.553±0.043 | 0.581±0.034 |

2.1 不同训练策略的结果分析

采用Base和BIRADs模型,以肿块图像img作为

输入,分别以不同训练策略(是否利用迁移学习)进行对应模型的训练和测试,以分析训练策略对模型性能的影响。通过未迁移学习Base模型(Base_n)和迁移学习的Base模型(Base)对应的ROC曲线,可以看出使用Base在乳腺肿块分类任务中取得了显著的

性能提升,其AUC值明显高于Base_n(0.793 vs 0.621),见图3a;分析图3b结果可知,经过迁移学习的BIRADs模型相比于BIRADs_n模型性能提升效果同样明显(0.830 vs 0.598)。



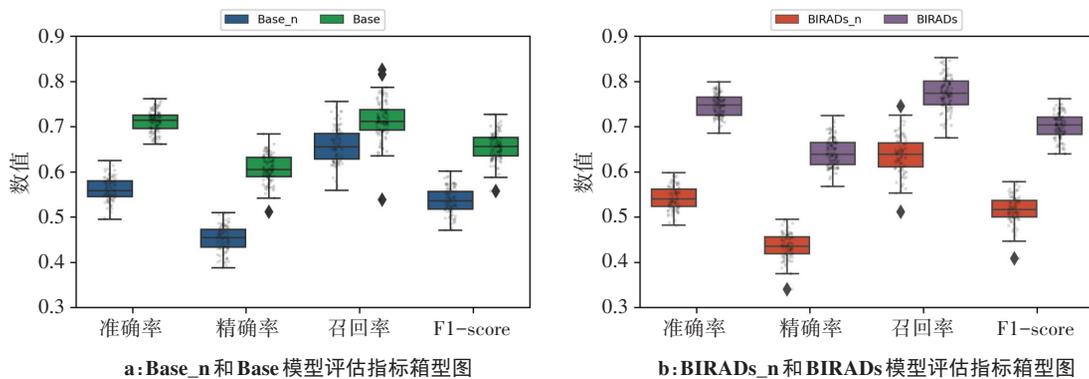
a: 未迁移学习和迁移学习的Base模型的ROC曲线 b: 未迁移学习和迁移学习的BIRADs模型的ROC曲线

图3 不同训练策略下对应模型的ROC曲线和AUC值比较

Figure 3 Comparison of ROC curves and AUC values for the same model with different training strategies

进一步计算和分析不同训练策略下不同模型的关键性能指标差异。如表2和图4a所示,迁移学习的Base模型对应准确率为0.712±0.023,精确率为0.607±0.030,召回率为0.715±0.042,F1-score为0.656±0.029,上述所有指标上,迁移学习的Base模型与未迁移学习的Base模型之间的差异均有统计学意

义($P<0.001$)。类似地,从图4b中可以看出,对于BIRADs模型对应准确率为0.747±0.024,精确率为0.643±0.032,召回率为0.774±0.037,F1-score为0.702±0.028。相比于BIRADs_n,BIRADs在所有性能指标上也取得了显著的提升($P<0.001$)。



a: Base_n和Base模型评估指标箱型图

b: BIRADs_n和BIRADs模型评估指标箱型图

图4 不同训练策略下对应的评估指标箱型图

Figure 4 Box plots of evaluation metrics for the same model with different training strategies

通过对Base和BIRADs模型在不同训练策略下的性能对比分析可知,使用大规模图像的预训练权重和迁移学习的训练策略可提高模型对乳腺肿块分

类的判别特征学习能力,显著提升模型分类能力。

2.2 不同网络模型的结果分析

为进一步了解不同网络模型对肿块分类性能的

影响,以Base和本研究提出的BIRADs模型进行如下实验:分别以肿块图像和肿块掩模图像作为输入时,BIRADs模型在不同阈值下的ROC曲线均优

于Base模型(图5a),且AUC值更高(0.830 vs 0.793),表明BIRADs模型在乳腺肿块分类任务中具有更好的区分能力和性能表现;以肿块掩模图像

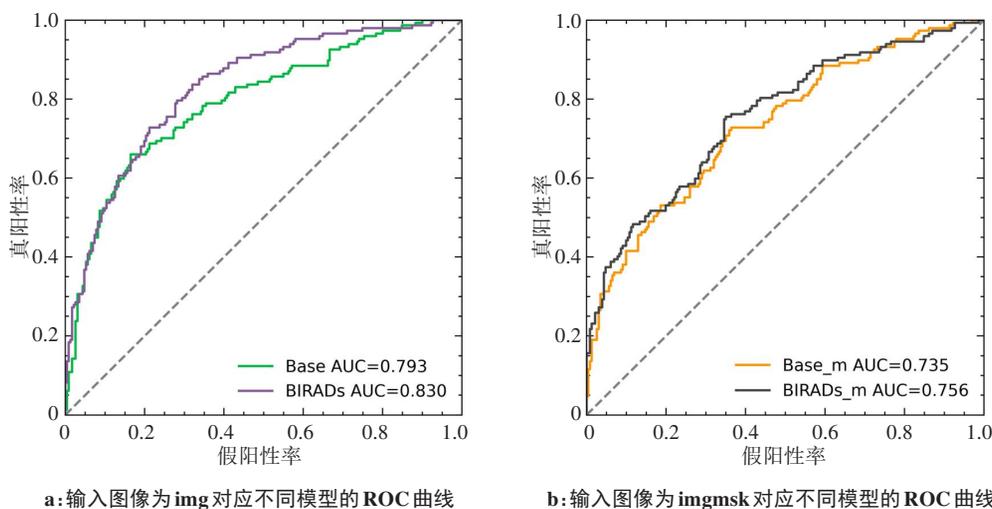


图5 不同网络模型下在同-输入图像类型下的ROC曲线和AUC值比较

Figure 5 Comparison of ROC curves and AUC values for different models with the same input type

通过量化指标对两个模型的性能进行综合对比。实验结果如图6a所示,以肿块图像作为输入时,BIRADs模型的准确率、精确率、召回率、F1-score均取得了较高的数值($P<0.001$);如图6b所示,以肿块掩模图像P<0.001)。上述结果表明与Base模型相比,BIRADs模型通过对BIRADs多任务进行

评估,在肿块分类任务中能够实现更准确的分类结果。相关结果进一步证实,BIRADs模型利用纹理特征提取器和形态学特征提取器,可以获得更符合肿块诊断的特征信息,结合BIRADs多任务评估,进一步提升了网络的分类性能,该网络在识别恶性乳腺肿块方面具有更好的准确率和召回率。

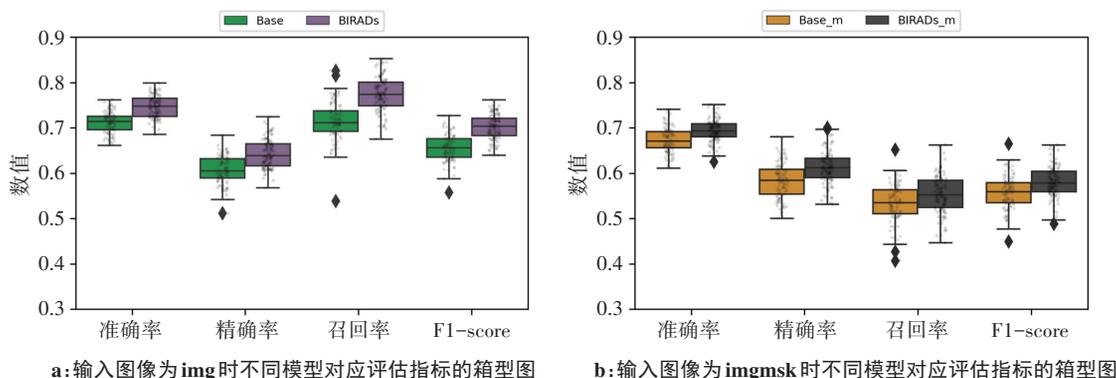


图6 不同网络模型下对应的评估指标的箱型图

Figure 6 Box plots of evaluation metrics for different models with the same input type

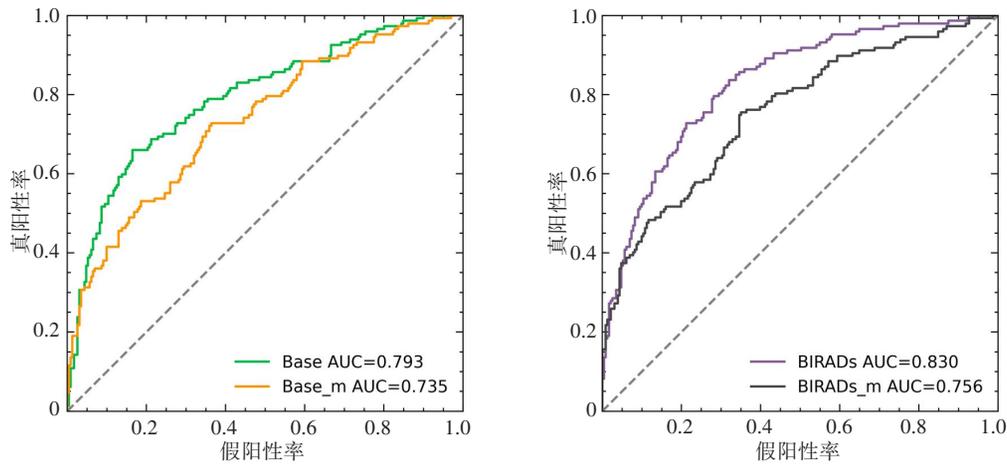
2.3 不同输入图像类型的性能比较

本研究分析不同输入图像类型对模型性能的影响,

并通过Base和BIRADs模型进行对比研究,旨在探究不同输入图像的优劣,从而为模型选择和优化提供更为

准确的指导。如图7a所示,以img和imgmsk作为模型输入比较,img的Base模型的AUC最优(0.793 vs 0.735);

从图7b可得,img作为输入的BIRADs模型的AUC最优(0.830 vs 0.756)。



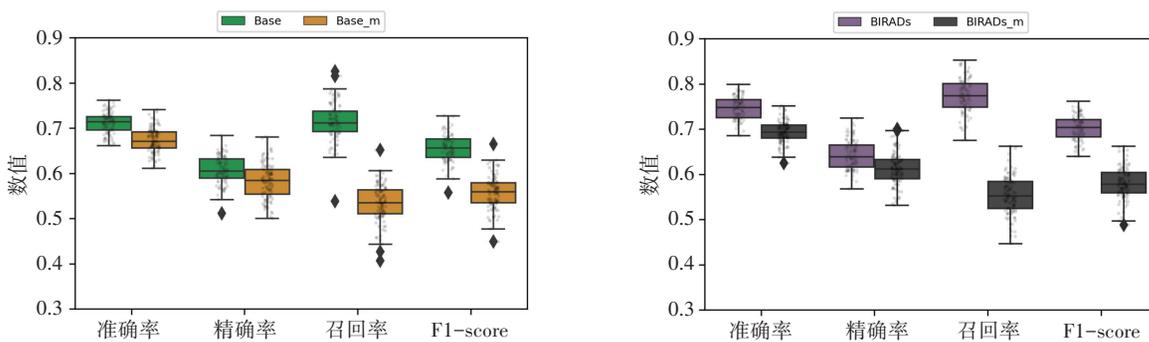
a:输入图像为img和imgmsk的Base模型对应的ROC曲线 b:输入图像为img和imgmsk的BIRADs模型对应的ROC曲线

图7 不同输入图像类型下对应模型的ROC曲线和AUC值比较

Figure 7 Comparison of ROC curves and AUC values for the same model with different input types

此外,通过比较准确率、精确率、召回率和F1-score等数值指标,用以比较不同输入类型对模型分类性能。如图8所示,结果表明,Base模型和BIRADs模型在所有指标上均取得了更好的结果,验

证了img作为输入类型的优越性。虽然imgmsk组合理论上可以提供更全面的病变信息,但本研究实验结果显示,在当前研究任务中以原始肿块图像(img)作为输入在乳腺肿块分类任务中表现更优。



a:Base模型在输入图像为img和imgmsk时对应评估指标的箱型图 b:BIRADs模型在输入图像为img和imgmsk时对应评估指标的箱型图

图8 不同输入图像类型下对应的评估指标的箱型图

Figure 8 Box plots of evaluation metrics for the same model model with different input types

上述结果表明原始肿块图像能够充分包含关键的乳腺病变信息,使得模型能够更准确地区分不同病变类型和良恶性。相比之下,使用imgmsk的Base模型和BIRADs模型性能相对较差,这可能是因为利用msk裁剪后的肿块掩模图像imgmsk虽然提供了更强的形状和边缘特征,但也引入了一定的噪音和冗余信息,影响模型分类准确性;另一潜在原因可能为CBIS-DDSM的掩模图像的精度不足,导致模型无法获得更精准的形态学信息。

3 讨论

本研究旨在应对乳腺肿块分类问题和深度学习的应用挑战,提出一种基于BIRADs的多任务分类模型,通过构建基于迁移学习的纹理特征提取器和形态学特征提取器,将自然图像域特征提取能力泛化至乳腺肿块图像域中,然后利用任务特定分支使网络专注于各种BIRADs特征,模拟临床医生的观察和决策过程。最后整合BIRADs相关特征层,提高乳腺肿块分类的准确

性和鲁棒性。本研究通过系列实验和指标分析表明:训练策略方面,利用迁移学习策略引入预训练权重并进行微调,可使Base/BIRADs模型的AUC值、准确率、精确率、召回率和F1-score指标均取得显著的提升;模型架构方面,BIRADs模型通过整合形态学和纹理特征提取器,结合多任务学习实现了BIRADs诊断指标评估,与基线的Base模型相比,BIRADs模型表现出更高的分类性能;输入类型方面,尽管肿块掩模图像提供可用的形态学信息,但原始肿块图像可保留病变边缘浸润情况,同时捕捉不同病变和良恶性的差异,该输入类型在本研究中分类任务表现更优。具体结果显示,在迁移学习和原始肿块输入情况下,BIRADs模型相较Base模型表现出更高的AUC值(0.830 vs 0.793)、准确率(0.747±0.024 vs 0.712±0.023)、精确率(0.643±0.032 vs 0.607±0.030)、召回率(0.774±0.037 vs 0.715±0.042)和F1-score(0.702±0.028 vs 0.656±0.029)。多任务学习模型在乳腺肿块分类中呈现出明显优势。

本研究仍存在一定局限性:首先,采用的公开数据集在真实世界中可能存在数据分布偏差,影响模型的泛化能力;其次,模型侧重于结合临床的评估特征,但忽略了如对侧乳腺特征等临床信息;最后,本研究还需探讨其他输入类型、分类模型^[25-27]、图像增强方法^[28]。在未来研究中,将致力于扩展数据源、进一步优化模型和图像输入及预处理等方面,以便更全面地解决深度学习在乳腺肿块分类问题。

综上所述,本研究针对乳腺图像肿块分类和深度学习应用的瓶颈问题,提出一种基于BIRADs多任务学习的乳腺肿块分类方法,采用迁移学习策略和融合临床评估的BIRADs多任务学习,在乳腺肿块分类任务中取得了显著的优势。此外,原始肿块图像在不同网络模型均表现出更好的分类性能。通过整合BIRADs中的临床信息,本研究模型在分类的准确性和鲁棒性均得到提升。本文探索了不同的输入组合和架构,提供一个综合性的可复用框架,既利用计算分析又借鉴临床专业知识,以提高诊断的准确性。本研究提出将临床知识和数据驱动方法结合起来的分类框架,有望提高乳腺癌诊断的准确性,对其他乳腺病变分类研究和其他医学图像分析领域具有重要参考意义,亦为该领域的发展开辟新的可能性。

【参考文献】

- [1] Fahad Ullah M. Breast cancer: current perspectives on the disease status [J]. *Adv Exp Med Biol*, 2019, 1152: 51-64.
- [2] 周星彤,沈松杰,孙强. 中国乳腺癌筛查现状及进展[J]. *中国医学前沿杂志(电子版)*, 2020, 12(3): 6-11.
- [3] Zhou XT, Shen SJ, Sun Q. Current situation and progress of breast cancer screening in China [J]. *Chinese Journal of the Frontiers of Medical Science (Electronic Version)*, 2020, 12(3): 6-11.
- [3] Ginsburg O, Yip CH, Brooks A, et al. Breast cancer early detection: a phased approach to implementation [J]. *Cancer*, 2020, 126: 2379-

- 2393.
- [4] 黄育北,佟仲生,陈可欣,等.《中国女性乳腺癌筛查指南》解读(精编版)[J]. *中国肿瘤临床*, 2019, 46(9): 432-440.
- [4] Huang YB, Tong ZS, Chen KX, et al. Interpretation of guideline for breast cancer screening in Chinese women [J]. *Chinese Journal of Clinical Oncology*, 2019, 46(9): 432-440.
- [5] Benson JR, Jatoi I, Keisch M, et al. Early breast cancer [J]. *Lancet*, 2009, 373(9673): 1463-1479.
- [6] Sun YS, Zhao Z, Yang ZN, et al. Risk factors and preventions of breast cancer [J]. *Int J Biol Sci*, 2017, 13(11): 1387.
- [7] 陈春华,马小敏,丁雅玲. 乳腺钼靶X线摄影在诊断早期乳腺癌中的应用价值[J]. *国际医药卫生导报*, 2019, 25(18): 3060-3062.
- [7] Chen CH, Ma XM, Ding YL. The application value of mammography in the diagnosis of early breast cancer [J]. *International Medicine and Health Guidance News*, 2019, 25(18): 3060-3062.
- [8] Dromain C, Boyer B, Ferre R, et al. Computed-aided diagnosis (CAD) in the detection of breast cancer [J]. *Eur J Radiol*, 2013, 82(3): 417-423.
- [9] Ragab DA, Sharkas M, Attallah O. Breast cancer diagnosis using an efficient CAD system based on multiple classifiers [J]. *Diagnostics*, 2019, 9(4): 165.
- [10] Rodriguez-Ruiz A, Lång K, Gubern-Merida A, et al. Stand-alone artificial intelligence for breast cancer detection in mammography: comparison with 101 radiologists [J]. *J Natl Cancer Inst*, 2019, 111(9): 916-922.
- [11] 陈智丽,高皓,潘以轩,等. 乳腺X线图像计算机辅助诊断技术综述 [J]. *计算机工程与应用*, 2022, 58(4): 1-21.
- [11] Chen ZL, Gao H, Pan YX, et al. Review of computer aided diagnosis technology in mammography [J]. *Computer Engineering and Applications*, 2022, 58(4): 1-21.
- [12] Shen D, Wu G, Suk HI. Deep learning in medical image analysis [J]. *Annu Rev Biomed Eng*, 2017, 19: 221-248.
- [13] Singhal A, Phogat M, Kumar D, et al. Study of deep learning techniques for medical image analysis: a review [J]. *Mater Today Proc*, 2022, 56: 209-214.
- [14] Mendelson EB. Artificial intelligence in breast imaging: potentials and limitations [J]. *AJR Am J Roentgenol*, 2019, 212(2): 293-299.
- [15] Shah SM, Khan RA, Arif S, et al. Artificial intelligence for breast cancer analysis: trends & directions [J]. *Comput Biol Med*, 2022, 142: 105221.
- [16] 陈远琴. 基于卷积神经网络的乳腺肿瘤良恶性分类方法研究 [D]. 郑州: 郑州大学, 2019.
- [16] Chen YQ. Study on classification of benign and malignant breast tumors based on convolutional neural network [D]. Zhengzhou: Zhengzhou University, 2019.
- [17] Salama WM, Aly MH. Deep learning in mammography images segmentation and classification: automated CNN approach [J]. *Alex Eng J*, 2021, 60(5): 4701-4709.
- [18] Shen R, Zhou K, Yan K, et al. Multicontext multitask learning networks for mass detection in mammogram [J]. *Med Phys*, 2020, 47(4): 1566-1578.
- [19] Bejnordi BE, Veta M, Van Diest PJ, et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer [J]. *JAMA*, 2017, 318(22): 2199-2210.
- [20] Niu S, Liu Y, Wang J, et al. A decade survey of transfer learning (2010-2020) [J]. *IEEE Trans Artif Intell*, 2020, 1(2): 151-166.
- [21] Kim HE, Cosa-Linan A, Santhanam N, et al. Transfer learning for medical image classification: a literature review [J]. *BMC Med Imaging*, 2022, 22(1): 69.
- [22] Lee RS, Gimenez F, Hoogi A, et al. A curated mammography data set for use in computer-aided detection and diagnosis research [J]. *Sci Data*, 2017, 4(1): 1-9.
- [23] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.
- [24] 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. IEEE, 2009: 248-255.
- [25] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning [C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2017.
- [26] Dai Y, Gao Y, Liu F. Transmed: transformers advance multi-modal medical image classification [J]. *Diagnostics*, 2021, 11(8): 1384.
- [27] Aly GH, Marey M, El-Sayed SA, et al. YOLO based breast masses detection and classification in full-field digital mammograms [J]. *Comput Methods Programs Biomed*, 2021, 200: 105823.
- [28] Chlap P, Min H, Vandenberg N, et al. A review of medical image data augmentation techniques for deep learning applications [J]. *J Med Imaging Radiat Oncol*, 2021, 65(5): 545-563.

(编辑:陈丽霞)