

文章编号: 1673-3193(2024)04-0428-11

基于改进的深度学习模型分类乳腺癌分子亚型

段美华, 焦雄

(太原理工大学 生物医学工程学院, 山西 晋中 030600)

摘要: 由于乳腺癌的高度异质性, 不同分子亚型的乳腺癌在治疗和诊断方面都有着很大的差异, 因此提高乳腺癌亚型诊断的准确性, 对于进一步降低乳腺癌的误诊率, 避免乳腺癌的过度治疗具有重要的意义。基于深度学习算法对 TCIA 数据库中的 MRI 乳腺癌医学图像进行了乳腺癌 Luminal A 和非 Luminal A 的分子亚型的分类研究。为了对比深度学习算法在分类乳腺癌分子亚型方面的优越性, 使用 TCGA 数据库基因表达数据进行了乳腺癌分子亚型分类研究; 在相同乳腺癌 MRI 图像数据库下, 基于传统机器学习算法进行了乳腺癌分子亚型分类研究。在利用深度学习算法处理 MRI 图像时, 对比了多种深度学习模型; 微调 VGG16 网络, 冻结卷积层; 添加 densenet 网络模块改进了 VGG16 网络模型, 即得到 VGG16+densenet(4) 模型。改进的 VGG16 网络模型的准确率达到 0.96, AUC 达到 0.97; 基因表达数据分类准确率为 0.73, AUC 为 0.79; 传统机器学习的分类准确率最高达到 0.80, AUC 达到 0.87。实验结果表明, 提出的 VGG16+densenet(4) 模型提高了乳腺癌分子亚型的准确率, 具有更好的分类效果。

关键词: 深度学习; 乳腺癌分子亚型; 传统机器学习; 数据增强; 基因表达数据

中图分类号: TP391.4; TP18 **文献标识码:** A **doi:** 10.3969/j.issn.1673-3193.2024.04.003

引用格式: 段美华, 焦雄. 基于改进的深度学习模型分类乳腺癌分子亚型[J]. 中北大学学报(自然科学版), 2024, 45(4): 428-438.

DUAN Mei-hua, JIAO Xiong. Improved deep learning models to classify breast cancer molecular subtypes[J]. Journal of North University of China(Natural Science Edition), 2024, 45(4): 428-438.

Improved Deep Learning Models to Classify Breast Cancer Molecular Subtypes

DUAN Meihua, JIAO Xiong

(College of Biomedical Engineering, Taiyuan University of Technology, Jinzhong 030600, China)

Abstract: Due to the high heterogeneity of breast cancer, different molecular subtypes of breast cancer have great differences in the treatment and diagnosis. Therefore, it is of important research significance to improve the accuracy of the diagnosis of breast cancer subtypes, thus to further reduce the misdiagnosis rate of breast cancer and avoid overtreatment of breast cancer. The molecular subtypes of Luminal A and non-Luminal A of breast cancer were classified based on deep learning algorithm for MRI breast cancer medical images in TCIA database. In order to compare the superiority of deep learning algorithm in classification of breast cancer molecular subtypes, gene expression data of TCGA database was used to classify breast cancer molecular subtypes. In the same breast cancer MRI image database, the classification of breast cancer molecular subtypes was studied based on traditional machine learning algorithm. Compara-

收稿日期: 2023-03-02

基金项目: 山西省自然科学基金资助项目(202203021221063)

作者简介: 段美华(1998-), 女, 硕士, 主要从事影像组学、图像处理的研究。

通信作者: 焦雄(1974-), 男, 教授, 博士, 主要从事生物信息学、影像基因组学的研究。E-mail: jiaoxiong@tyut.edu.cn。

tive studies between a variety of deep learning models were conducted in the processing of MRI images with deep learning algorithm. Fine-tuning VGG16 network and frozen convolution layer were implemented. An improved VGG16 network model was proposed, and a densenet network module was added, namely VGG16+densenet (4) model. The accuracy of the improved VGG16 network model was 0.96, and the AUC was 0.97. The classification accuracy of gene expression data was 0.73, and the AUC was 0.79. The classification accuracy and AUC of traditional machine learning reached 0.80 and 0.87 respectively. The experimental results show that the proposed VGG16+densenet (4) model improves the accuracy of the molecular subtypes of breast cancer and has better classification effect.

Key words: deep learning; breast cancer molecular subtypes; traditional machine learning; data augmentation; gene expression data

0 引 言

世界卫生组织国际癌症研究机构(IARC)发布的2020年全球最新癌症数据显示:2020年全球新发癌症病例1 929万例。其中一个最明显的变化就是乳腺癌的新发病例数快速增长,已达226万,首次正式取代肺癌(220万)成为全球第一大癌症,占所有新增癌症患者数量的11.7%^[1]。作为乳腺癌临床影像检查的常用手段,核磁共振成像(MRI)对软组织的分辨率高,能清晰地显示乳腺癌病变形态和内部结构等信息,这对了解乳腺的侵蚀和浸润的情况,术前手术方式的选择及术后的评价有较大应用价值。自动提取大量乳腺癌MRI的成像特征,能够有效地阐明乳腺癌图像中的细微特征,可以为疾病的诊断、预后和治疗反应提供信息^[2]。前瞻性研究证实,对具有已知或疑似基因突变的女性进行MRI筛查,可以获得比其他方法更高的癌症检测灵敏度,进而能够改善患者的局部治疗结果^[3]。

到目前为止,影像组学在乳腺癌医学影像的分类研究方面主要包括人工提取特征的传统机器学习和自动提取特征的深度学习这两种分类方法。但是,人工提取特征存在耗费时间长、流程复杂以及在专业知识能力方面要求较高等局限性。深度学习最初的发展架构比较基础、简单和直观^[4]。深度学习提取特征可以用于许多不同的任务^[5],有更好的模型泛化能力,相比机器学习的单任务模式能够挖掘更多的有用信息。

根据基因表达水平分类,乳腺癌可以分为4个亚型,常见的是Luminal A、Luminal B、人体表皮生长因子受体-2(HER-2)阳性和三阴性(Basal-like),如表1所示^[6-7]。不同分子亚型的乳腺癌,它们的疾病进程、治疗方式以及对治疗的

反应和预后都不尽相同。

表1 乳腺癌分子亚型标准及其治疗手段

Tab. 1 Breast cancer molecular subtypes standard and treatment methods

分子分型	标志物	治疗方法
Luminal A	ER 阳性、PR 阳性、Her-2 阴性 Ki-67 低表达	内分泌治疗
Luminal B	ER 阳性、PR 阳性、Her-2 阴性 Ki-67 高表达	内分泌治疗敏感型
HER-2	ER 阴性、PR 阴性、Her-2 阳性	专门针对基因的 靶向治疗
Basal-like	ER、PR、Her-2 阴性	化疗

影像组学就是从放射影像的图像中高通量地提取大量的影像特征,用特征算法将影像数据的感兴趣区域转化为可挖掘使用的影像特征数据。基于影像组学研究乳腺癌分子亚型的分类方面,国内外学者进行了大量的工作。袁成成等^[8]将202例浸润性导管癌患者乳腺DCE-MRI影像数据集随机分为训练集(122例)、测试集(80例),多序列影像作为输入,采用多任务学习方法分别对基于Ki-67表达与Luminal A型、基于Ki-67表达与组织学分级、基于Luminal A型与组织学分级进行预测,得到Ki-67表达、Luminal A型和组织学分级的最佳AUC分别为0.804, 0.757和0.724。马梦伟等^[9]将200例浸润性乳腺癌X线及超声影像,分为Luminal组与非Luminal组,组内按7:3比例随机分为训练组及测试组,建立了4种机器学习模型:随机森林(RF)、极端梯度提升(XGBoost)、逻辑回归(LR)及支持向量机(SVM)模型。乳腺癌不同分子亚型分类的准确率分别为0.68, 0.61, 0.67和0.65。Zhu等^[10]利用乳腺的MRI图像数据集,将乳腺癌的分子亚型分为两类,Luminal A型单独作为一类,其他3种亚型作为一类,使用预训练网络提取特征,再用提取的特征训练网络模

型(包括VGG16、GoogLeNet等多种经典模型,)并进行乳腺癌分子亚型分类,最终将所有的测试结果取平均值,得到的平均AUC为0.64。Sun等^[11]研究了266例腔内亚型(Luminal A和Luminal B)或非腔内亚型(HER-2和三阴性)乳腺癌患者的DCE-MRI影像信息和分子信息,开发了一种基于深度学习的集成CNN模型来用于乳腺癌分子亚型的分类,Luminal与非Luminal的测试数据集上的平均预测特异性、准确度、精确度和ROC曲线下面积分别为0.958,0.852,0.961和0.867。Liu等^[12]使用TCGA-BRCA数据集作为乳腺癌分子亚型预测的样本集,将患者的基因数据与病理图像数据相结合来构建多模态融合框架,提出了一种基于多模态数据的混合的深度学习模型,预测luminal A亚型二分类的AUC值为0.9316。

不同的乳腺癌分子亚型有不同的治疗方法,分子亚型的正确分类对于乳腺癌的诊断治疗及预后具有重要的意义。有关乳腺癌分子亚型分类的研究大多数都是从病理形态学和基因组学的方面开展,然而从影像组学的研究角度,采用非侵入式的研究方法大量提取和分析图像、训练数据特征,可以阐明

特征与乳腺癌状态之间存在的关系,能够提供更多乳腺癌潜在的数据信息,进而实现乳腺癌不同亚型的精准治疗,并制定临床的个性化治疗方案,防止乳腺癌患者的过度治疗。那么,如何提高乳腺癌分子亚型的准确率就变得十分重要。

针对以上问题,本文拟采用迁移学习结合深度学习的方法,分类乳腺癌的分子亚型,并改进深度学习模型使其达到更高的分类准确率,使得不同亚型的患者能得到个体化的治疗方案,避免过度治疗。

1 方法

针对乳腺癌MRI的医学图像数据,本文分别使用微调深度学习模型对乳腺癌的Luminal A和非Luminal A的分子亚型进行分类研究;在原有深度学习网络模型的基础上添加densenet网络模块,基于改进的深度学习模型分类乳腺癌分子亚型;利用传统机器学习算法对乳腺癌分子亚型进行分类研究。本文针对乳腺癌的基因组学数据,在相同病例下利用深度神经网络对乳腺癌的基因表达数据做了亚型研究,并对比分析了上述方法在分类乳腺癌亚型上的优越性能。实验流程如图1所示。

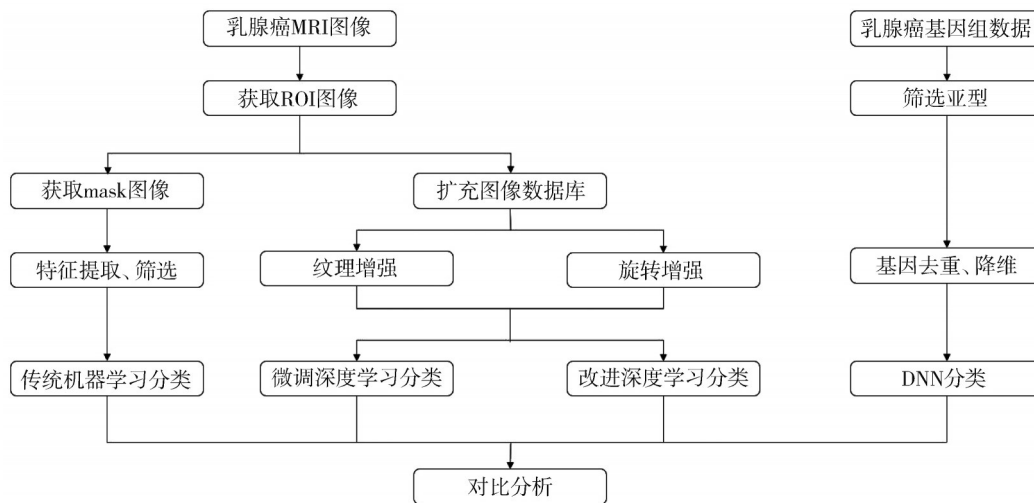


图1 实验流程

Fig. 1 Workflow of experiment

2 实验数据

2.1 患者图像信息数据

本文将公共数据集TCIA里的TCGA-BRCA中的MRI医学图像数据信息,用作乳腺癌分子亚型分类的研究样本集。TCGA-BRCA数据集共有139例病人的临床信息,但是仅有108例病人能根据医生

的标注信息找到其对应的分子分型,包括52例Luminal A型、12例Basal-like、5例HER-2、25例癌旁组织和14例Luminal B型。考虑到实验数据的平衡性,将108例患者的数据信息划分为Luminal A(52例)和非luminal A(56例)两类。原始数据集的dicom图像通过MicroDicom软件转换成png格式图像,dicom图像的像素分辨率为256*256,层厚为7mm,层间距为19mm,png格式图像的像素分辨率为512*

512, 转换后的部分图像如图 2 所示。在 108 例病人的医学图像中, 每例病人都能挑选出有医生标注肿瘤位置信息的 3 张 MRI 图像, 最后得到包含原始的 324 张乳腺癌图像的数据库。基因表达数据从公共数据集 TCGA 中乳腺癌基因库里转录组学获取, 共有 1 226 个样本数据, 60 660 个基因数据。

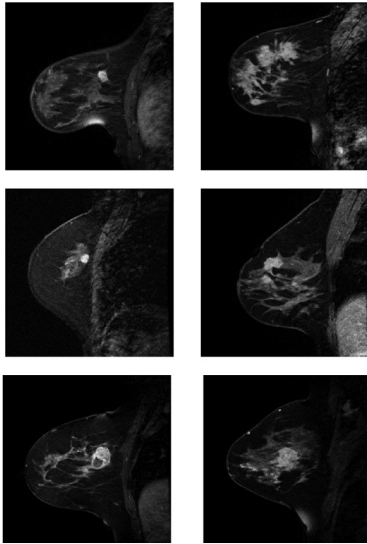


图 2 部分乳腺癌的原始 MRI 图像
Fig. 2 Initial MRI images of partial breast cancer

2.2 ROI 图像获取

根据数据库中现有医生标注的 ROI 位置信息可以划分出与原始图像对应的 ROI 图像, 为了保证深度学习和机器学习的可对比性, 将 ROI 图像长和宽的像素尺寸都限制在 80 以内, 部分 ROI 图像如图 3 所示。

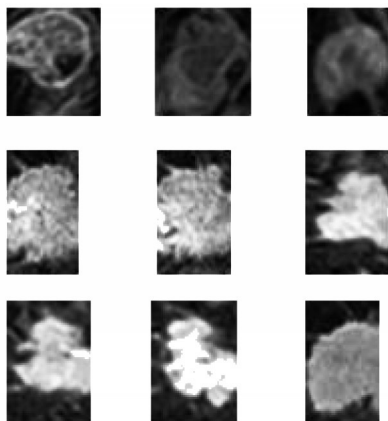


图 3 部分 ROI 图像
Fig. 3 ROI partial images

为了符合深度学习模型要求输入的数据集尺寸为正方形, 将 ROI 图像周围添加黑边, 统一图像的像素分辨率为 80*80, 其中部分统一尺寸

的图像如图 4 所示。

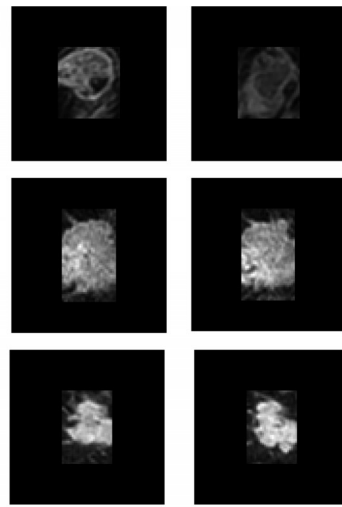


图 4 Luminal A 型 ROI 统一 80*80 像素尺寸后的图像
Fig. 4 ROI unified 80*80 images(Luminal A)

2.3 mask 图像获取

mask 图像的获取分为三个部分: 1) 将 ROI 图像经过形态学的线性变换得到增强图像数据; 2) 进行形态学开运算处理, 先腐蚀后膨胀, 得到更为连通的 mask 区域; 3) 使用大津算法分割图像。部分 mask 图像如图 5 所示。

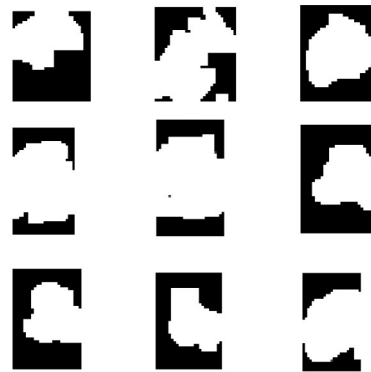


图 5 ROI 图像对应的 mask 图像
Fig. 5 The mask images corresponding to the ROI images

2.4 数据库扩充

医学图像数据集通常很小, 然而深度学习需要大量的训练样本数据才能得到较好的模型泛化能力。本文使用数据增强方法对原始数据集进行扩充, 增加训练样本数量。数据增强是解决图像分类训练样本数量不足的一种关键且不可或缺的方法, 其有用性已经在医学图像中反复报道^[13]。首先将医生给出的 ROI 坐标信息进行扩大 ROI 位置坐标边界的处理, 在原始坐标基础之上, 进行坐标加 4 和加 8 处理。将

原始图像数据量扩大到3倍,并在图像周围添加黑边,统一图像尺寸大小为80*80,数据量由324张扩充为972张。在此基础上应用两种数据增强方法进一步增强数据库,一种是对图像进行90°和180°旋转处理,获得旋转增强数据库,另一种是对图像进行变亮和变暗处理,获得纹理增强数据库。最后,这两种图像增强数据库均由972张扩充为2916张,这样可以满足深度学习需要大量训练样本数据的要求,同时也可对比旋转增强数据库和纹理增强数据库对乳腺癌分子亚型分类效果的影响。

3 基于深度学习方法分类

3.1 数据集划分

深度学习可以从图像中自动提取高层抽象特征^[14]。将扩充后的旋转增强数据库和纹理增强数据库分别按训练集:验证集:测试集为6:2:2的比例,随机划分数据库为训练集1752张、验证集579张、测试集585张。

3.2 迁移学习

迁移学习(Transfer Learning)通过预训练好的模型来学习新样本,以解决训练数据的不足所带来的过拟合问题^[15]。首先本文利用python软件训练深度学习网络模型,保存ImageNet预训练的权重参数,直接将已经训练好的模型参数迁移到本文实验模型中进行新的训练。这样可以在训练过程中得到更高的模型初始化性能,提升模型的运行速率,从而达到更好的收敛效果^[16]。

3.3 评价指标

使用准确率(Accuracy)、混淆矩阵(Confusion

Matrix)、精确率(Precision)、召回率(Recall)、F1-score和AUC(Area Under Curve)等评价指标来判断网络模型分类结果的优劣。

TP是被模型预测为正类的正样本,TN是被模型预测为负类的负样本,FP是被模型预测为正类的负样本,FN是被模型预测为负类的正样本。 $Accuracy = (TP+TN)/(TP+FP+FN+TN)$ 是分类问题中最常用的指标,计算了分类正确的预测数与总预测数的比值。 $Precision = TP/(TP+FP)$,着重评估在预测为正类的所有数据中,真正正类的数据所占比例。 $Recall = TP/(TP+FN)$,着重评估在所有的正类数据中,有多少数据被成功预测为正类。 $F1 = 2(Precision*Recall)/(Precision+Recall)$,是精确率和召回率的调和均值。ROC全称是“受试者工作特征”(Receiver Operating Characteristic)。AUC是ROC曲线下与坐标轴围成的面积。

3.4 模型微调

在迁移学习的基础上,搭建深度学习网络模型,并对深度学习模型进行微调以达到更高的分类准确率。在微调深度学习模型前,分别测试了8种深度学习网络对乳腺癌数据集的分类能力,其中VGG16网络对乳腺癌分子亚型的分类能力表现最好,准确率最高,曲线拟合效果最好。VGG16网络结构如图6所示。卷积层提取局部特征(如线条、形状、边缘和纹理),这些特征可以转移到类似的视觉识别任务中,如超声图像中的癌症识别^[17]。最大池化层可以实现下采样降维,去除冗余信息,对特征进行压缩,简化网络复杂度,减小计算量,减小内存消耗,具有平移不变性、旋转不变性和尺度不变性。

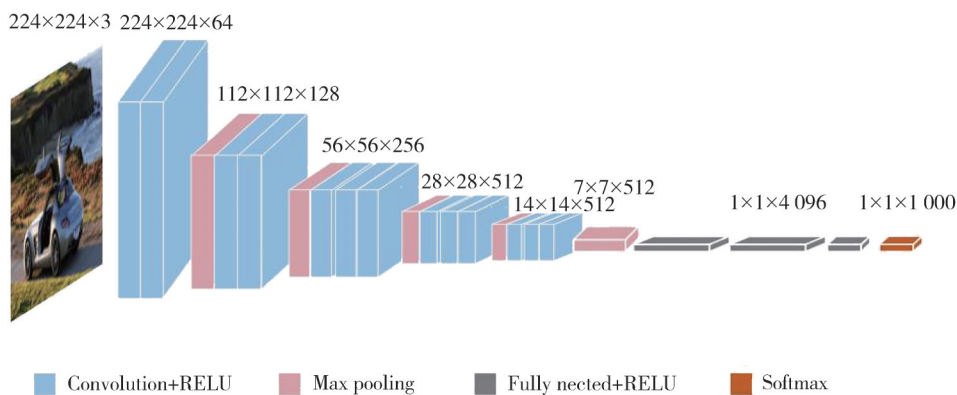


图6 VGG16网络结构

Fig. 6 VGG16 network structure

微调模型方面首先分别冻结了模型block5最后的3个卷积层,选择效果最优,然后分别在纹理

图像数据库和旋转图像数据库上对比8种网络模型的评价指标。

冻结层指的是该层将不参加网络训练,即该层的权重永远都不会更新。通常选择冻结网络底层提取的图像高层特征,如 block5_conv1、block5_conv2 和 block5_conv3 可视化结果如图 7 所示(单位:像素),可以看出层次越深,提取到的特征就越细节化。

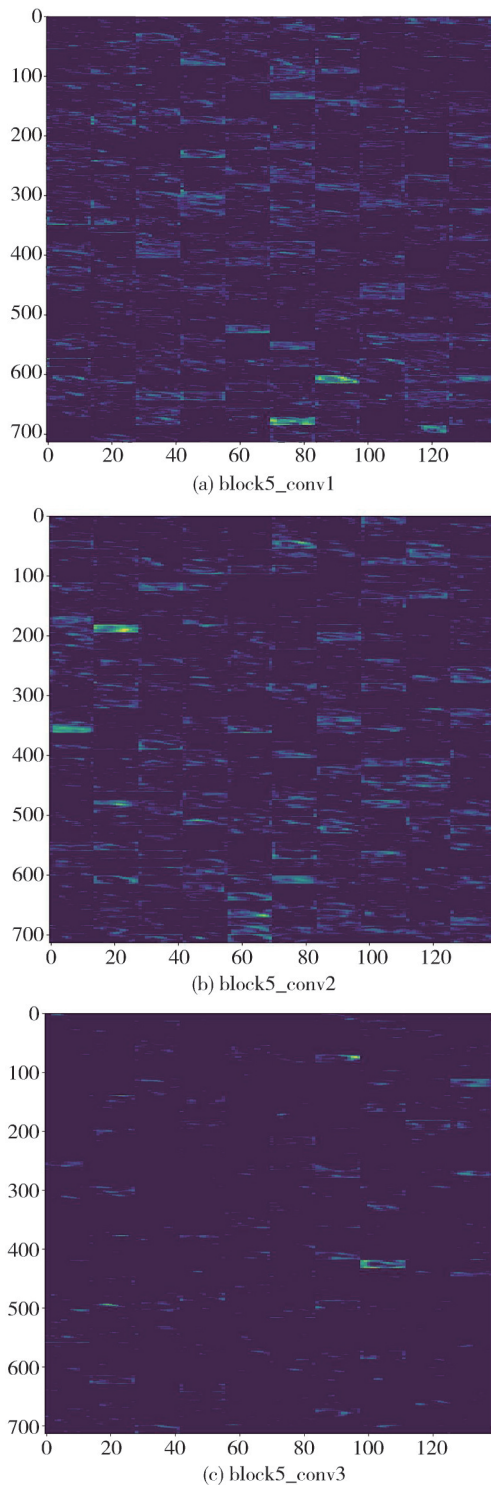


图 7 block5_conv1、block5_conv2 和 block5_conv3 可视化
Fig. 7 block5_conv1, block5_conv2 and block5_conv3 visualization

初始深度学习分类亚型使用纹理增强数据库,在迁移学习 VGG16^[18]网络模型的基础之上,设置初始 batch_size 为 32,学习率 Adam 为 0.000 1, dropout 为 0.5,测试冻结不同卷积层,如 block5_conv1、block5_conv2、block5_conv3 和 block5_conv(冻结 block5 最后 3 个卷积层),添加二分类的完全连接层。初始设置下冻结不同层的测试集评价指标如表 2 所示。

表 2 冻结不同层的测试集评价指标

Tab. 2 The testing sets evaluation indicators in freezing different layers

网络层	Accuracy	AUC	Recall	Precision	F1-score
block5_conv1	0.94	0.96	0.92	0.96	0.94
block5_conv2	0.92	0.93	0.91	0.94	0.92
block5_conv3	0.93	0.95	0.95	0.93	0.94
block5_conv	0.92	0.94	0.93	0.91	0.92

模型训练期间,对数据进行归一化处理。模型在训练集上的精度一直不断上升,但是验证集上的精度在第 15 轮就达到最高点并且基本维持稳定不变,这表明模型在训练数据上出现了过拟合^[19]的现象。在 VGG16 网络全连接层前面添加 dropout 层用于调节模型过拟合问题,随机丢弃概率值设为 0.5。

由表 2 可以看出初始设置下冻结 block5_conv1 的准确率最高为 0.94,但是训练预测数据所采用的优化方法、学习率、步长和迭代次数等参数只是一个初始的选择,可能并不是最适合模型的参数。因此,为了找到最优的参数设置,分别调整不同的参数选择,如 batch_size 取 30, 34 和 36, Adam 取 0.000 2, dropout 取 0.4 和 0.6,对比初始参数设置取得的测试集评价指标,最终得出冻结 VGG16 网络的 block5_conv1 层的最优参数如表 3 所示。

表 3 模型最优参数

Tab. 3 Optimal parameters of the model

VGG16 网络模型参数	结果
loss	binary_crossentropy
Adam	0.0001
dropout	0.5
batch_size	32
epochs	30

在模型最优参数基础之上,对比 VGG16、VGG19、ResNet50、ResNet152、MobileNet、Xception、DenseNet201 与 DenseNet169 共 8 种网络的测试集分类结果,评价指标如表 4 所示,可以看出,VGG16 网络在分类纹理增强数据库上取得的最高准确率为 0.94, AUC 为 0.96。

表4 8种网络对比的纹理增强数据库测试集评价指标结果

Tab. 4 Evaluation index results of eight kinds of network in texture enhancement database test sets

网络模型	Accuracy	AUC	Recall	Precision	F1-score
VGG16	0.94	0.96	0.92	0.96	0.94
VGG19	0.93	0.94	0.95	0.92	0.93
ResNet50	0.86	0.93	0.85	0.88	0.86
ResNet152	0.86	0.92	0.93	0.92	0.92
MobileNet	0.82	0.89	0.81	0.83	0.82
Xception	0.87	0.91	0.90	0.86	0.88
DenseNet169	0.90	0.96	0.94	0.87	0.90
DenseNet201	0.90	0.94	0.90	0.91	0.90

同样基于VGG16网络,使用旋转增强数据库,初始设置batch_size取32,Adam取0.0001,dropout取0.5,测试冻结不同层,结果如表5所示。

表5 冻结不同层的测试集评价指标

Tab. 5 The testing sets evaluation indicators in freezing different layers

网络层	Accuracy	AUC	Recall	Precision	F1-score
block5_conv1	0.91	0.97	0.92	0.90	0.91
block5_conv2	0.91	0.96	0.95	0.88	0.91
block5_conv3	0.51	0.50	1.00	0.51	0.67
block5_conv	0.91	0.97	0.88	0.93	0.91

由于冻结block5_conv层后训练集的准确率和损失曲线拟合效果差距较大,且测试集损失率为0.36,大于冻结block5_conv1层的损失率0.33。后续研究基于block5_conv1层进行参数调整,如batch_size取30,34和36,Adam取0.0002,dropout取0.4和0.6,对比初始参数设置取得的测试集评价指标,最终得出冻结VGG16网络的block5_conv1层的最优参数如表6所示。

表6 模型最优参数

Tab. 6 Optimal parameters of the model

VGG16网络模型参数	结果
loss	binary_crossentropy
Adam	0.0001
dropout	0.5
batch_size	36
epochs	30

在模型最优参数基础之上,对比VGG16、VGG19、ResNet50、ResNet152、MobileNet、Xception、DenseNet201与DenseNet169共8种网络的测试集分类结果,评价指标如表7所示。可以看出,VGG16网络在分类旋转增强数据库上取得的最高准确率为0.93,AUC为0.96。

对比两种数据库的测试集评价指标发现,VGG16网络在乳腺癌二分类上存在优越性,同时纹理增强数据库的各项评价指标均要高于旋转增强数据库,

纹理增强数据库的混淆矩阵如图8所示。但是,纹理增强数据库的验证集损失曲线在20轮之后开始上升,说明VGG16网络模型存在不稳定性。

表7 8种网络对比的旋转增强数据库测试集评价指标结果

Tab. 7 Evaluation index results of eight kinds of network in rotation enhancement database test sets

网络模型	Accuracy	AUC	Recall	Precision	F1-score
VGG16	0.93	0.96	0.93	0.93	0.93
VGG19	0.92	0.97	0.95	0.90	0.92
ResNet50	0.86	0.95	0.80	0.92	0.85
ResNet152	0.77	0.86	0.67	0.84	0.75
MobileNet	0.86	0.93	0.87	0.86	0.86
Xception	0.86	0.93	0.78	0.92	0.85
DenseNet169	0.87	0.92	0.87	0.87	0.87
DenseNet201	0.85	0.92	0.88	0.83	0.86

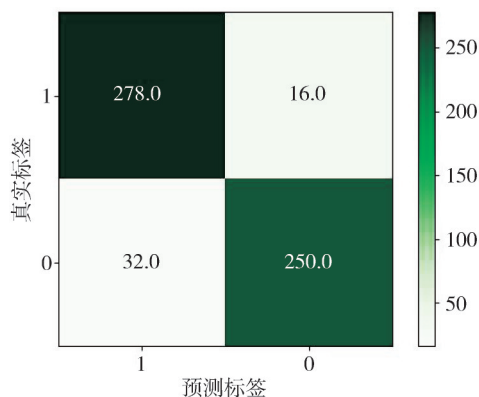


图8 纹理增强数据库的混淆矩阵

Fig. 8 Confusion matrix of texture enhancement database

3.5 densenet模块的添加

在上述微调模型的基础上,添加densenet的第4个模块,分别使用纹理增强数据库和旋转增强数据库对比8种网络模型的评价指标。

为了进一步提高模型分类准确率,使训练集验证集的准确率和损失曲线来拟合,实验在VGG16网络的完全连接层前添加densenet第4个模块,其网络结构如图9所示,即VGG16+densenet(4)。这样在加深网络深度的同时减少了参数量,优化了训练数据。使用纹理增强数据库,迁移学习VGG16网络模型,设置初始batch_size为32,Adam为0.0001,dropout为0.5,测试冻结不同卷积层,如block5_conv1、block5_conv2、block5_conv3和block5_conv,测试集评价指标如表8所示。

虽然冻结block5_conv3和block5_conv的评价指标很接近,但是block5_conv的训练集验证集曲线有几处尖锐,相反block5_conv3的训练集验证集曲线相对稳定,综合考虑,本文后续研究选择冻结

block5_conv3, 调整参数值, batch_size 取 30, 34 和 36, Adam 取 0.000 2, dropout 取 0.4 和 0.6, 对比初始参数设置取得的测试集评价指标, 最终得出冻结 block5_conv3 的 VGG16 网络模型最优参数如表 9 所示。

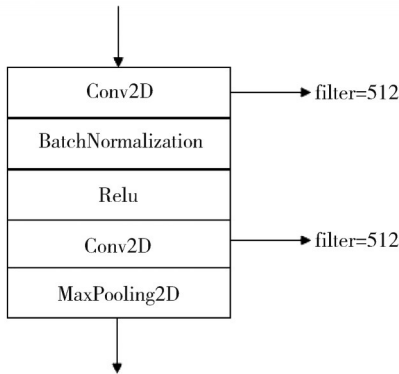


图 9 densenet(4)网络层

Fig. 9 densenet (4) network layers

表 8 冻结不同层的测试集评价指标

Tab. 8 Evaluation index of the testing sets in freezing different layers

网络层	Accuracy	AUC	Recall	Precision	F1-score
block5_conv1	0.94	0.96	0.93	0.94	0.93
block5_conv2	0.93	0.95	0.96	0.90	0.93
block5_conv3	0.94	0.96	0.95	0.94	0.94
block5_conv	0.94	0.97	0.95	0.94	0.94

表 9 模型最优参数

Tab. 9 Optimal parameters of the model

VGG16 网络模型参数	结果
loss	binary_crossentropy
Adam	0.0001
dropout	0.5
batch_size	34
epochs	30

使用纹理增强数据库, 在最优参数设置下对比 VGG16、VGG19、ResNet50、ResNet152、MobileNet、Xception、DenseNet201 与 DenseNet169 共 8 种网络的测试集分类结果, 评价指标如表 10 所示。可以看出, VGG16 网络在分类任务上取得的最高准确率为 0.96, AUC 为 0.97。

基于 VGG16+densenet(4) 网络模型并使用旋转增强数据库, 初始设置 batch_size 取 32, Adam 取 0.000 1, dropout 取 0.5, 测试冻结不同层, 结果如表 11 所示。

实际冻结 block5_conv1 层的准确率为 0.9323, AUC 为 0.9716, 冻结 block5_conv3 层的准确率为 0.928 8, AUC 为 0.971 5。后续实验基于冻结 block5_conv1 层进行参数调整, 如

batch_size 取 30, 34 和 36, Adam 取 0.000 2, dropout 取 0.4 和 0.6, 对比初始参数设置取得的测试集评价指标, 最终得出冻结 VGG16 网络的 block5_conv1 层的最优参数如表 12 所示。

表 10 8 种网络对比的纹理增强数据库测试集评价指标结果

Tab. 10 Evaluation index results of eight kinds of network in texture enhancement database test sets

网络模型	Accuracy	AUC	Recall	Precision	F1-score
VGG16	0.96	0.97	0.97	0.95	0.96
VGG19	0.90	0.93	0.89	0.90	0.89
ResNet50	0.83	0.94	0.97	0.76	0.85
ResNet152	0.84	0.90	0.81	0.87	0.84
MobileNet	0.89	0.94	0.92	0.88	0.90
Xception	0.87	0.93	0.89	0.86	0.87
DenseNet169	0.90	0.964	0.92	0.89	0.90
DenseNet201	0.89	0.94	0.88	0.90	0.89

表 11 冻结不同层的测试集评价指标

Tab. 11 Evaluation index of the testing sets in freezing different layers

网络层	Accuracy	AUC	Recall	Precision	F1-score
block5_conv1	0.93	0.97	0.90	0.96	0.93
block5_conv2	0.88	0.96	0.95	0.83	0.89
block5_conv3	0.93	0.97	0.92	0.94	0.93
block5_conv	0.91	0.96	0.87	0.95	0.91

表 12 模型最优参数

Tab. 12 Optimal parameters of the model

VGG16 网络模型参数	结果
loss	binary_crossentropy
Adam	0.0001
dropout	0.5
batch_size	30
epochs	30

在最优参数设置下对比 8 种网络的测试集分类结果, 测试集评价指标如表 13 所示。可以看出, VGG16 网络在分类任务上取得的最高准确率为 0.93, AUC 为 0.97。

表 13 8 种网络对比的旋转增强数据库测试集评价指标结果

Tab. 13 Evaluation index results of eight kinds of network in rotation enhancement database test sets

网络模型	Accuracy	AUC	Recall	Precision	F1-score
VGG16	0.93	0.97	0.92	0.95	0.93
VGG19	0.87	0.94	0.83	0.92	0.87
ResNet50	0.90	0.97	0.89	0.92	0.90
ResNet152	0.86	0.94	0.85	0.88	0.86
MobileNet	0.87	0.93	0.86	0.89	0.88
Xception	0.87	0.93	0.81	0.92	0.86
DenseNet169	0.82	0.92	0.76	0.88	0.81
DenseNet201	0.88	0.94	0.83	0.92	0.87

通过对比两种数据库的评价指标结果, 发现使用纹理增强数据库的各方面评价指标都要优于旋转增强数据库。纹理增强的图像数据集上准确率更高, 效果更好, 其混淆矩阵如图 10 所示, 说明纹理增强数据库在本文研究中确实存在优势,

这可能与VGG16网络中最大池化层对纹理识别有影响,而对图像旋转没有识别效果有关。

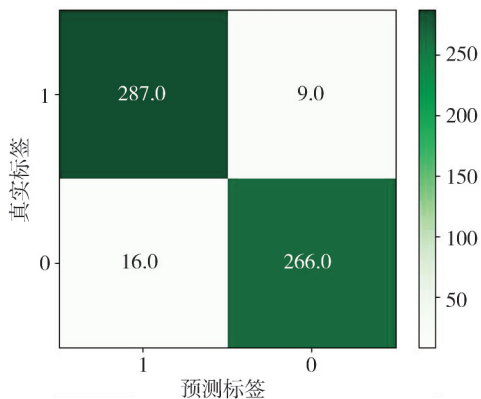


图 10 纹理增强数据库的混淆矩阵

Fig. 10 Confusion matrix of texture enhancement database

综上所述,本文使用TCGA-BRCA的MRI图像数据集,提出VGG16+densenet(4)的网络模型,在预测Luminal A亚型二分类上,准确率为0.96,AUC为0.97,高于Zhu等二分类乳腺癌亚型得到的0.65的AUC。同时,VGG16+densenet(4)网络模型的训练集和验证集的准确率和损失曲线都比微调VGG16网络稳定,证明模型可信,说明本文提出的乳腺癌分类方法有一定的价值和意义。

3.6 基于深度神经网络的基因表达数据处理

为了对比本文提出的VGG16+densenet(4)网络模型在分类乳腺癌分子亚型方面的优越性,在相同病例下,使用神经网络来处理基因表达数据。基因表达数据包含1226个样本数据,其中仅有135个样本有明确分型,再去除25个正常组织样本。将剩余110个样本数据和60660个基因进行去重,处理缺失值,log2减少差异基因,标准化基因Z-score,最终得到110个样本数据和39276个基因。进行PCA降维,将110个样本分为Luminal A型54例,非Luminal A型56例,训练集:验证集:测试集为6:2:2。搭建简单的深度卷积神经网络模型(DNN),如图11所示。

调整参数,如batch_size取5,6和7,Adam取0.0001,0.0003,0.0005,0.0004,0.0008,0.001,0.005和0.01,dropout取0.1,0.2和0.3,得到最优参数如表14所示。

测试集评价指标中,准确率为0.73,AUC为0.79,召回率为0.82,精确率为0.69,F1评分为0.75,其各项评价指标都远小于图像二分类数据指标。训练集和验证集的准确率损失曲线以及混

淆矩阵如图12所示。

表 14 模型最优参数

Tab. 14 Optimal parameters of the model

DNN网络模型参数	结果
loss	binary_crossentropy
Adam	0.0005
dropout	0.2
batch_size	6
epochs	30

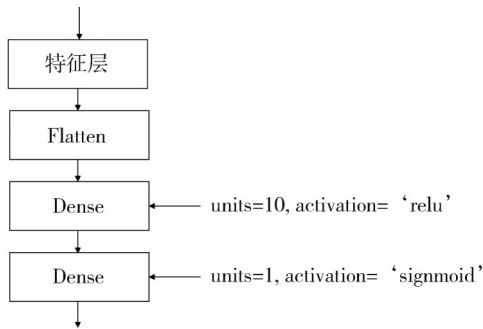


图 11 DNN网络层

Fig. 11 DNN network layers

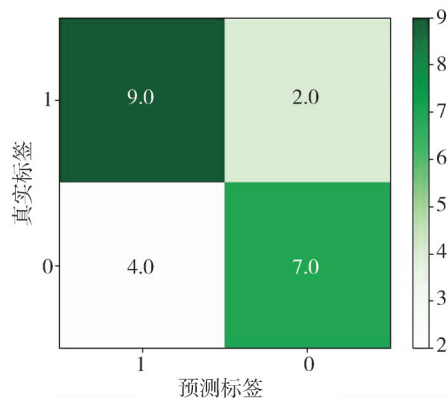


图 12 基因表达数据的混淆矩阵

Fig. 12 Confusion matrix of gene expression data

4 基于传统机器学习方法的分类

4.1 特征提取和筛选

使用与深度学习方法相同的乳腺癌MRI数据库,基于传统机器学习分类乳腺癌分子亚型。将统一尺寸的ROI图像及其对应的mask图像作为输入,利用python中的pyradiomics提取图像特征,pyradiomics提供的许多可以使用的滤波器,可以用于指定提取的特征类型。本文所要提取的特征类型包括First Order Features(共19个)、Shape Features(2D)(共10个)、Gray Level Co-occurrence Matrix(GLCM) Features灰度共生矩阵特征(共24个)、Gray Level Size Zone Matrix(GLSZM) Features灰度级大小区域矩阵特征(共

16个)、Gray Level Run Length Matrix (GLRLM) Features 灰度游程矩阵特征(共 16 个)、Neighbouring Gray Tone Difference Matrix (NGTDM) Features 相邻灰度差矩阵(共 5 个)、Gray Level Dependence Matrix (GLDM) Features 灰度依赖矩阵特征(共 14 个),总计提取到 1125 个特征。使用 Lasso 算法对 1125 个特征进行筛选,最终筛选出 15 个关键性特征。

4.2 分类结果

根据 Lasso 筛选得到的 15 个关键性特征,构建传统机器学习分类器模型,包括支持向量机(Support Vector Machines, SVM)、决策树、高斯贝叶斯、K-近邻(K-nearest neighbor, KNN)、线性回归和随机森林(Random Forest, RF),其中准确率最高为 0.80, AUC 为 0.87。传统机器学习分类模型的计算结果如表 15 所示。改进的深度学习网络模型的准确率最高可达 0.96, AUC 最高为 0.97, 由此表明了 VGG16+densenet(4) 在分类乳腺癌分子亚型上的优势。

表 15 传统机器学习分类结果

Tab. 15 Classification results of traditional machine learning

方法	Accuracy	AUC
SVM	0.70	0.75
决策树	0.72	0.75
高斯贝叶斯	0.61	0.62
KNN	0.67	0.73
线性回归	0.64	0.71
RF	0.80	0.87

5 对比分析

在本研究中,基于深度学习对乳腺癌 MRI 图像的乳腺癌分子亚型进行分类,通过微调 VGG16 网络模型并添加 densenet(4)网络模块,成功取得了较高且稳定的准确率和 AUC,两者分别为 0.96 和 0.97,见表 10。微调 VGG16 网络得到的分类的准确率和 AUC 分别为 0.94 和 0.96,见表 4;利用深度卷积神经网络分类相同病例的基因表达数据,得到的准确率为 0.73, AUC 为 0.79;基于相同乳腺癌 MRI 图像数据库进行了传统机器学习亚型分类,其中 RF 的准确率最高为 0.80, AUC 最高为 0.87,见表 15,证明本文提出的 VGG16+densenet(4)网络模型结构在区分乳腺癌 Luminal A 型和非 Luminal A 型上的优越性能。通过对比

不同学者在乳腺癌亚型分类方法上的结果,也能说明本文提出的改进深度学习模型在乳腺癌分子亚型分类研究上有极高的优势,详见表 16。

表 16 不同分类方法总结

Tab. 16 Different classification methods summary

分类方法	来源	Accuracy
VGG16+densenet(4)	本文	0.96
VGG16	袁成成 ^[8]	0.71
集成 CNN 模型	Sun 等 ^[11]	0.85
RF	马梦伟 ^[9]	0.68

本文研究也存在一定的局限性,首先是数据来源少,主要使用公开数据集并且病例信息十分有限。如果在乳腺癌分子亚型方面有更充足的多模态数据来源,例如 CT、超声等医学图像数据,针对乳腺癌 4 种亚型进行更细致具体的四分类研究,那么最终取得的分类效果会具有更高的研究价值,也能更好地帮助医生根据不同分子亚型做出精准的个性化治疗,对提高诊断准确率上也有着重要的意义。另外,本研究只是使用了影像信息,而乳腺癌 MRI 影像信息还可以融合乳腺癌相关文本信息、基因组学等多组学的信息,进而对乳腺癌分子亚型的认识会更加全面具体。

6 结论

本文基于改进的深度学习方法实现了乳腺癌 MRI 图像的分子亚型分类。采用的 VGG16+densenet(4)模型,具有更深层次的网络结构,并且在分类乳腺癌 Luminal A 和非 Luminal A 的问题上实现了更高的评价指标,其准确率达到 0.96, AUC 达到 0.97。采用的纹理增强数据效果和迁移学习方法也有效地避免了由于样本量不足而引发的过拟合问题,实验证明本文提出改进的深度学习模型提高了乳腺癌分子亚型的识别率,具有较好的鲁棒性和泛化性。后续研究可以更深层次地融合影像数据和基因组学数据,精炼模型的层次结构,使该方法具有更好的临床实用性。

参考文献:

[1] SUNG H, FERLAY J, SIEGEL R L, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries [J]. CA: Cancer Journal for Clinicians, 2021, 71(3): 209-249.

[2] TOMASZEWSKI M R, GILLIES R J. The biological meaning of radiomic features (2021) [J]. Radiol-

- ogy, 2021, 298(3): 505-516.
- [3] MORROW M, WATERS J, MORRIS E. MRI for breast cancer screening, diagnosis, and treatment[J]. *The Lancet*, 2011, 378(9805): 1804-1811.
- [4] PANG T, WONG J H D, WEI L N, et al. Deep learning radiomics in breast cancer with different modalities: overview and future[J]. *Expert Systems with Applications*, 2020, 158: 113501.
- [5] ZHE Z A, EA A, AS A, et al. Deep learning for identifying radiogenomic associations in breast cancer[J]. *Computers in Biology and Medicine*, 2019, 109: 85-90.
- [6] 中国抗癌协会乳腺癌专业委员会. 中国抗癌协会乳腺癌诊治指南与规范(2021年版)[J]. *中国癌症杂志*, 2021, 31(10): 954-1040.
Chinese Anti-Cancer Association, Committee of Breast Cancer Society. Chinese Anti-Cancer Association guidelines of breast cancer (2021) [J]. *China Oncology*, 2021, 31(10): 954-1040. (in Chinese)
- [7] BURSTEIN H J, CURIGLIANO G, THÜRLIMANN B, et al. Customizing local and systemic therapies for women with early breast cancer: the St. Gallen International Consensus Guidelines for treatment of early breast cancer 2021 [J]. *Annals of Oncology*, 2021, 32(10): 1216-1235.
- [8] 袁成成, 范明, 许茂盛, 等. 基于DCE-MRI影像的深度多任务学习联合预测乳腺癌病理信息研究[J]. *传感技术学报*, 2021, 34(3): 354-360.
YUAN Chengcheng, FAN Ming, XU Maosheng, et al. Prediction of breast cancer histological information by deep multi-task learning based on DCE-MRI images[J]. *Chinese Journal of Sensors and Actuators*, 2021, 34(3): 354-360. (in Chinese)
- [9] 马梦伟, 秦耿耿, 徐维敏, 等. 基于X线及超声乳腺影像报告和数据系统构建机器学习模型预测乳腺癌分子分型[J]. *中国医学影像技术*, 2020, 36(12): 1814-1819.
MA Mengwei, QIN Genggeng, XU Weimin, et al. Machine learning models for predicting molecular types of breast cancer based on X-ray and ultrasound breast imaging reporting and data system[J]. *Chinese Journal of Medical Imaging Technology*, 2020, 36(12): 1814-1819. (in Chinese)
- [10] ZHU Z, ALBADAWY E, SAHA A, et al. Breast cancer molecular subtype classification using deep features: preliminary results[C]//*Medical Imaging 2018: Computer-aided diagnosis*. SPIE, 2018, 10575: 651-656.
- [11] SUN R, MENG Z, HOU X, et al. Prediction of breast cancer molecular subtypes using DCE-MRI based on CNNs combined with ensemble learning[J]. *Physics in Medicine and Biology*, 2021, 66(17): 175009.
- [12] LIU T, HUANG J, LIAO T, et al. A hybrid deep learning model for predicting molecular subtypes of human breast cancer using multimodal data[J]. *Innovation and Research in Biomedical Engineering: IRBM*, 2022, 43(1): 62-74.
- [13] FUJIOKA T, YASHIMA Y, OYAMA J, et al. Deep-learning approach with convolutional neural network for classification of maximum intensity projections of dynamic contrast-enhanced breast magnetic resonance imaging[J]. *Magnetic Resonance Imaging*, 2021, 75: 1-8.
- [14] XIE J, LIU R, LUTTRELL J, et al. Deep learning based analysis of histopathological images of breast cancer[J]. *Frontiers in Genetics*, 2019, 10: 80.
- [15] SHAMSHIRI M A, KRZYŻAK A, KOWAL M, et al. Compatible-domain transfer learning for breast cancer classification with limited annotated data [J]. *Computers in Biology and Medicine*, 2023, 154: 106575.
- [16] HINTON B, LIN M, MAHMOUDZADEH A P, et al. Deep learning networks find unique mammographic differences in previous negative mammograms between interval and screen-detected cancers: a case-case study[J]. *Cancer Imaging*, 2019, 19(1): 41.
- [17] ZHU Y C, AIZOUBI A, JASSIM S, et al. A generic deep learning framework to classify thyroid and breast lesions in ultrasound images [J]. *Ultrasonics*, 2021, 110: 106300.
- [18] KONG L, CHENG J. Classification and detection of COVID-19 X-Ray images based on densenet and VGG16 feature fusion[J]. *Biomedical Signal Processing and Control*, 2022, 77: 103772.
- [19] WANG R Q, ZHU F, ZHANG X Y, et al. Training with scaled logits to alleviate class-level over-fitting in few-shot learning [J]. *Neurocomputing*, 2023, 522: 142-151.