

基于深度学习的胸部 X 射线图像识别及分类模型研究*

张晓熙^{1,2} 王远涵^{1,2} 杨顷落^{1,2} 黄雪^{1,2} 余红梅^{1,2,Δ} 武淑琴^{2,3,Δ}

【摘要】目的 以深度学习中的卷积神经网络为基础搭建胸部 X 射线 (chest X-ray, CXR) 图像分类模型,为肺部疾病提供可靠的辅助诊断技术。**方法** 经 KAGGLE 数据库收集新冠肺炎、轻度肺部感染、病毒性肺炎及正常的四种胸部 X 射线图片,按 3 : 1 : 1 的比例将数据随机划分成训练集,测试集和验证集;基于卷积神经网络架构搭建 CXR 图像分类模型,调节超参数对模型进行加强和优化;后通过混淆矩阵、准确率、灵敏度、K 折交叉验证结果等指标对模型进行验证及评价。**结果** 本研究模型对肺部影像图片的分类准确率为 0.81、灵敏度为 0.80、测试集和验证集损失值能够稳定在一个较低的水平。与相同迁移算法的模型相比,在测试数据集上的精确率、准确率、灵敏度、F1 分值分别提高了 1.7%、1.7%、1.3%、2.9%。**结论** 此模型对于 CXR 图像的识别和分类的性能更强,可以更有效地应用于肺部疾病的辅助分析和判断。

【关键词】 肺部疾病 模型构建 深度学习 影像识别

【中图分类号】 R195.1 **【文献标识码】** A **DOI** 10.11783/j.issn.1002-3674.2024.03.008

Research on Chest X-ray Image Recognition and Classification Model based on Deep Learning

Zhang Xiaoxi, Wang Yuanhan, Yang Qingluo, et al (Department of Health Statistics, School of Public Health, Shanxi Medical University (030000), Taiyuan)

【Abstract】 Objectives Building a chest X-ray (CXR) image classification model based on convolutional neural networks in deep learning, providing reliable auxiliary diagnostic techniques for lung diseases. **Methods** Four kinds of chest X-ray pictures of COVID-19, mild pulmonary infection, viral pneumonia and normal were collected through KAGGLE database, and the data were randomly divided into training set, test set and verification set according to 3 : 1 : 1 ratio. Building a CXR image classification model based on convolutional neural network architecture, adjusting hyperparameters to strengthen and optimize the model. Subsequently, the model was validated and evaluated using metrics such as confusion matrix, accuracy, sensitivity, and K-fold cross validation results. **Results** The classification accuracy of this research model for lung imaging images is 0.81, the sensitivity is 0.80, and the loss values of the test and validation sets can be stable at a relatively low level. Compared with models with the same migration algorithm, the accuracy, sensitivity, and F1 score on the test dataset were improved by 1.7%, 1.7%, 1.3%, and 2.9%, respectively. **Conclusion** This model has stronger recognition and classification performance for CXR images, and can be more effectively applied to auxiliary analysis and judgment of lung diseases.

【Key words】 Lung disease; Model building; Deep learning; Image recognition

肺是人体呼吸系统的主要器官,实现机体与外界环境之间的气体交换,所以容易受到感染并引发多种肺部疾病^[1]。近年来,随着空气污染和呼吸道传染病问题的加剧,人群肺部疾病发病率逐年增加。研究显示,早期诊断治疗可使肺部疾病患者生存率超过 90%,而延误诊治可能导致病情恶化,肺脏区域出现多发小斑片阴影及间质改变,严重情况下甚至可能导致肺实变、白肺甚至死亡^[2-3]。因此,在临床诊疗过程中,对疑似患者进行早期诊断排查具有重要的价值和意义。

目前,计算机断层扫描与胸部 X 射线是诊断肺部疾病的金标准^[4]。然而,对患者影像学图片的人工分析依赖于医生的经验技能和主观推断,且存在确认感染严重程度困难、耗费时间、人力物力等问题,进而影

响肺部疾病类型和病情进展的诊断。针对日益增长的肺部疾病高效诊断需求,本研究基于深度学习算法构建神经网络模型,利用成像技术探索患者胸部 X 射线图像中持续的视觉异常,通过定量化和自动化的方式对肺部疾病进行准确分类,以更好的估算方法协助医学工作者分析肺部影像并完成最终疾病诊断,进而节省看诊时间和医疗资源、提高工作效率、有效分担医疗系统的检测压力。

资料与方法

1. 资料来源

本研究的 CXR 影像数据来自公共数据库 (www.kaggle.com/datasets/covid19-radiography-database),共包括样本 CXR 图片 21164 张及相关病理信息,包含新冠肺炎图像 3615 张、病毒性肺炎图像 1345 张、健康肺部图像 10192 张以及轻型肺部感染图像 6012 张。所有图像均为便携式网络图形 (portable network graphics, PNG) 文件格式,分辨率为 299×299 像素,如图 1。

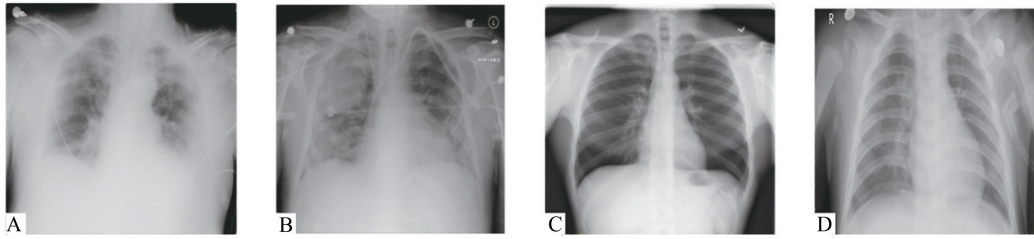
* 基金项目:国家自然科学基金面上项目(82273742)

1.山西医科大学公共卫生学院卫生统计学教研室(030000)

2.重大疾病风险评估山西省重点实验室

3.山西医科大学基础医学院数学教研室

Δ通信作者:武淑琴,E-mail:wsq13.wu@163.com;余红梅,E-mail:yu@sxmu.edu.cn



* :A 为新冠肺炎 CXR 图片,B 为轻度肺部感染 CXR 图片,C 为健康 CXR 图片,D 为病毒性肺炎 CXR 图片。

图 1 样本 CXR 分类图片

2. 图像划分指标特征定义

本文以肺部阴影、密度、毛玻璃浑浊等特征作为影像学划分依据^[5]。其中,肺部阴影指通过 X 光及核磁共振检查,以阴影区域大小、部位和状态作为标准,判断肺部密度增高或是降低的指标。毛玻璃浑浊指肺部纹理模糊、密度增加,使得肺组织看起来像被毛玻璃覆盖,呈现模糊、不清晰的状态。肺实变则表示肺组织密度增加、区域颜色较深且与周围组织对比明显、肺部斑纹因白化而减少,呈现实质化的状态^[6]。

不同肺部疾病在影像学上会呈现不同的特征性表现。新冠肺炎表现为双肺出现斑点状浸润影或磨玻璃影、密度增加、出现肺实变。病毒性肺炎双肺病变范围较小、肺部纹理增多,呈现小片状或广泛浸润影、密度增高但相较于新冠肺炎低。轻度肺部感染表现为双肺纹理增粗或模糊、阴影加深、肺实质密度增高、病变范围较小或限于一侧肺部。健康肺部为气管居中、纵隔无移位、胸廓对称、纹理清晰、无异常密度影像^[7-9]。

3. 模型构建

本研究基于 Python 软件的 Tensorflow 框架,通过构建一个新的前馈卷积神经网络模型用于图像分类,过程中模拟不同的网络层数、神经元个数、批量大小、学习率等指标的不同结构组合,最终选定在训练和验证阶段综合性能最优的模型。

此模型第一层为卷积层,包含 32 个 3×3 大小的卷积核,激活函数为 ReLU,图像输入形状为 (299, 299, 1)。第一层后为批归一化层 (batch normalization)。接下来是一个卷积层,使用 64 个 3×3 大小的卷积核及 ReLU 激活函数。后添加批归一化层,平均池化层 (average pooling) 和 Dropout 层,帮助提取图像特征并减少过拟合。以上模式被重复多次,每次包括两个卷积层、批归一化层、平均池化层和 Dropout 层,以增加模型复杂度和提取特征信息。后添加一个扁平层用于将之前卷积层输出的特征图转换成一维向量。之后依次添加批归一化层、全连接层 (具有 128 个神经元和 ReLU 激活函数)、Activation 层和 Dropout 层。最后一层是分类标签的概率分布输出层。

模型绩效评估结果显示:此模型的图片识别能力较强、训练时间较少且所占内存较低、收敛速度和精度较高,具体结构如图 2。

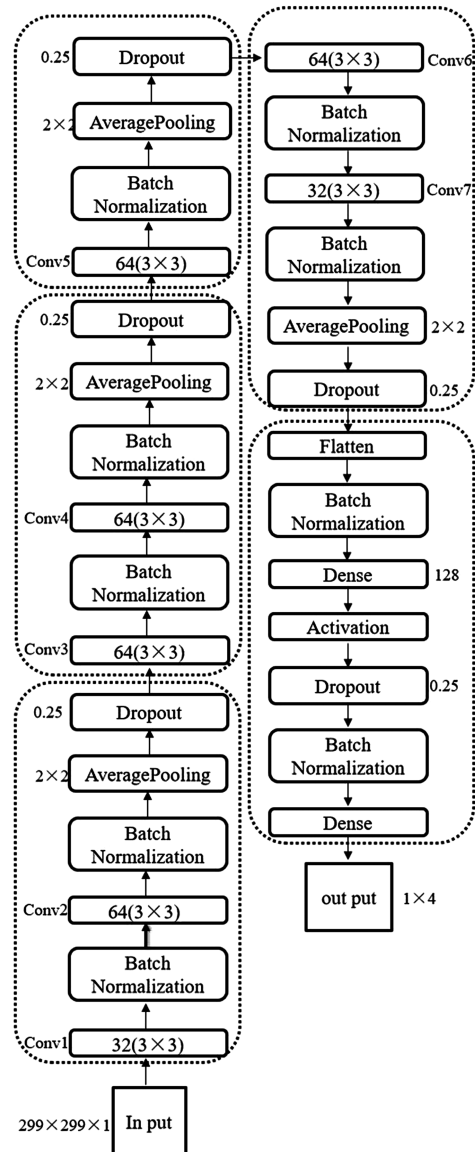


图 2 卷积神经网络模型图

4. 数据集划分

本研究所用的 21164 张胸部 X 射线图像分为四类: COVID (新冠肺炎)、Viral Pneumonia (病毒性肺炎)、Lung _ Opacity (轻度肺部感染) 和 Normal (正常)。以固定随机种子的随机化分配原则,将数据集划分为 60% 的训练集和 40% 的测试集。为进一步准备模型分类测试集,采用分层重采样方法将测试集分为两个子集:20% 验证集和 20% 测试集,20% 的验证集用于防止模型过度拟合和增强模型评估过程,划分后

各数据集用于后续模型预训练及开发,图像数量的划分结果如表 1 所示。

表 1 分割后数据集划分统计表

分类	训练集	测试集	验证集	总计
COVID	2169	723	723	3615
Viral Pneumonia	807	269	269	1345
Lung Opacity	3608	1202	1202	6012
Normal	6116	2038	2038	10192
Total	12700	4232	4232	21164

5. 数据预处理

研究结果显示,数据增强可以有效提升图像分类模型的性能^[10]。因此,在数据预处理阶段,对图像数据应用了采样、特征对齐、灰度变换、中心裁剪等方法。数据增强过程中,将所有图像分辨率大小调整统

一,以保持图像的一致性和兼容性。对于存在位置偏差的图像,以-100到+100之间的不同角度将图像随机翻转,并将图像缩放和裁剪,然后利用特征对齐算法(spatial transformer networks, STN)完成图片对齐^[11];为了降低背景噪声以及增强焦距位置,应用核大小为(3,3)的高斯过滤器以去除图像的高频分量,并进一步采用中心裁剪机制将所有的图像样本分割成不重叠的高分辨率子图像,裁剪后的每个切片定义一个感染区域,若有多个切片则选择最大区域;将图片中的每个像素分别表示为三个八位数字,对应于三个取值范围不同的数字通道,最终将每张图像转换为数字数组形式。将输出特征设置为0(Covid)、1(Lung Opacity)、2(Viral Pneumonia)、3(Normal)四类,将增强后的图片读入模型进行训练及评估。如图 3 所示。

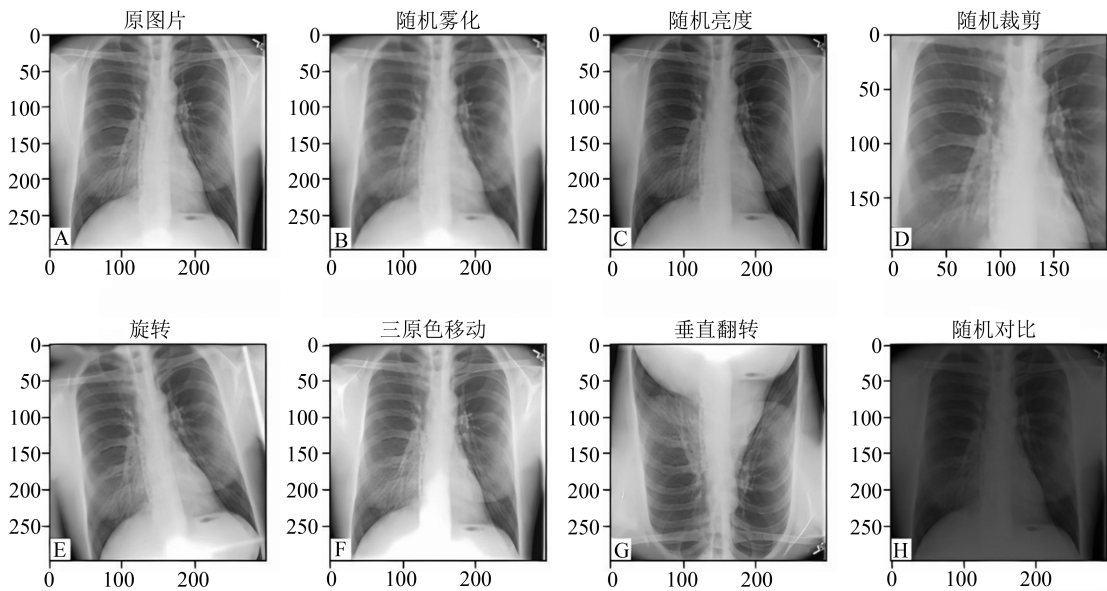


图 3 不同类型的图像增强

6. 模型训练及优化

在模型训练阶段,基于卷积神经网络架构对 CXR 图片进行建模,选取交叉熵损失函数减少错误标记程度^[12],初始特征设定为 4 并对模型进行预训练。然后使用验证集选择模型超参数并确定最优模型。最后通过验证集和测试集对最优模型的性能进行评估。

在模型优化阶段,采用人工干预和动态调节的方法将模型超参数进行微调测试,以确定最优参数组合,最大限度的减少模型在训练和验证阶段的损失。结果表明:学习率稳定在 0.0001、批量值为 16、迭代次数为 100、以自适应矩估计(adam)作为主要优化器时,模型表现出了最优的验证精度。训练结果如图 4 所示。

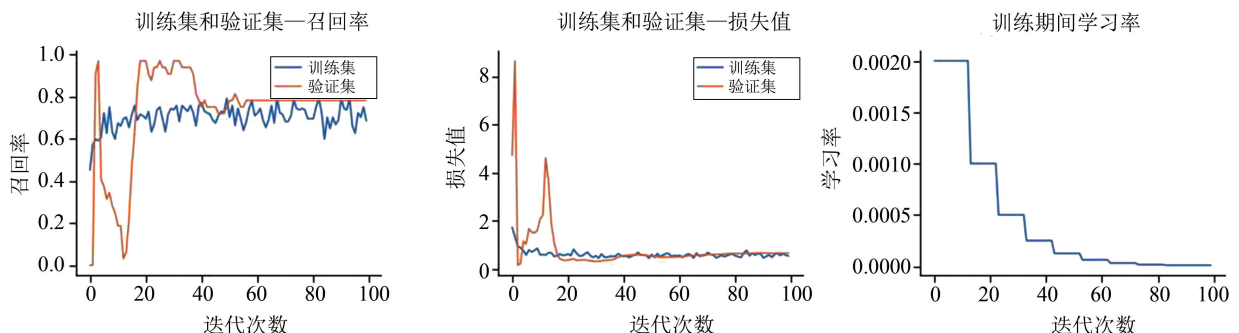


图 4 模型训练结果图

7.模型的评价

由于数据集不平衡,本文采用模型训练时间、损失值、准确率、灵敏度、精确率、F1 分值、混淆矩阵(confusion matrix)^[13-14]以及相关图表对模型的性能进行评估(表2)。模型性能评价指标计算公式如下:

$$\text{准确率} = \frac{N_{TN} + N_{TP}}{N_{TN} + N_{FN} + N_{TP} + N_{FP}}$$

$$\text{灵敏度} = \frac{N_{TP}}{N_{TP} + N_{FN}}$$

$$\text{精确率} = \frac{N_{TP}}{N_{FP} + N_{TP}}$$

$$\text{F1 分值} = 2 \times \frac{\text{精确率} \times \text{灵敏度}}{\text{精确率} + \text{灵敏度}}$$

表2 混淆矩阵表格形式

真实	预测			
	COVID	Lung_Opacity	Normal	Viral Pneumonia
COVID	N_{TP}	N_{FP}	N_{FP}	N_{FP}
Lung_Opacity	N_{FP}	N_{TP}	N_{FP}	N_{FP}
Normal	N_{FP}	N_{FP}	N_{TP}	N_{FP}
Viral Pneumonia	N_{FP}	N_{FP}	N_{FP}	N_{TP}

其中, N_{TP} 和 N_{TN} 分别表示正确诊断肺部疾病和非肺部疾病(健康)图像的数量, N_{FP} 和 N_{FN} 分别表示错误诊断肺部疾病和非肺部疾病(健康)图像的数量。此外,本研究还进行了 K 折交叉验证,以支持初始绩效指标评估。

结果

1.模型验证结果

本研究将 VGG16、VGG19、AlexNet 等模型作为预训练模型^[15],通过迁移学习方法将肺病分类模型(lung disease classification model, LDC-Model)与预训练模型的性能进行比较,后模拟并统计了基于 20% 验证数据集的各模型的评估结果,如表 3。本研究认为准确率得分是比较性能的最佳统计指标,首先使用 60% 训练数据集训练肺病分类模型和多个预训练模型,然后使用 20% 的验证数据集对每个模型进行验证,最后选择表现最优的模型,通过 20% 的测试数据集进行精度检验。以“训练-验证-测试”的方式,探讨肺病分类模型与其他模型相比的性能优劣程度。

表3 不同模型图像分类指标初步统计表

模型	精确率	准确率	灵敏度	F1 分值
VGG16	0.7273	0.7017	0.7668	0.7067
VGG19	0.7322	0.7109	0.7671	0.7155
AlexNet	0.7391	0.7544	0.7869	0.7449
InceptionV5	0.7484	0.7632	0.7533	0.7281
ResNet50	0.7548	0.7701	0.7731	0.7579
MobileNetV2	0.7580	0.7843	0.7896	0.7698
LDC-Model	0.7694	0.8187	0.8015	0.7963

初步验证结果显示:①所有预训练模型的准确率得分均在 0.7~0.8 之间且小于 LDC-Model(表3);② LDC-Model 具有更简单的结构设计,在 20% 的验证数据集中对四种肺部图像的分类性能更强,各项指标均优于其他预训练模型。

本研究进行了 K 折交叉验证程序,不仅可以对表 2 所示的模型评估结果提供支持,而且可以降低模型的方差、提高模型的稳健性和泛化能力。为将此模型与其他预训练模型进行精度得分比较,本研究选定了常用的折叠变体,即 $K=4$,将其中的一个子集作为测试数据集,其余的子集作为训练数据集,在训练集上对模型进行拟合,在测试集上进行验证,四折交叉验证的描述如表 4 所示;而后对每个子集获得的准确度分数、精确度分数、召回率和 F1 分数等关键性能指标进行平均,以产生模型的最终性能度量。

表4 四折交叉验证示意表

折叠	数据集			
折叠 1	测试	训练	训练	训练
折叠 2	训练	测试	训练	训练
折叠 3	训练	训练	测试	训练
折叠 4	训练	训练	训练	测试

在训练阶段,各个神经网络预训练模型的参数如下:学习率=0.0001,批量尺寸=16,训练轮数=100,使用 Adam 优化策略;模型的权重值在每次验证开始时被重置,以避免数据过度拟合的问题。根据先前的验证结果(表3),选择表现最好的 ResNet50 和 MobileNetV2 与 LDC-Model 进行交叉验证。结果显示, LDC-Model 在单次测试数据集上获得了 80% 的最高平均准确率得分,表现出最佳的性能(表5)。

表5 各模型四折交叉验证结果统计表(%)

模型	折叠	准确率	精确率	灵敏度	F1 分值
ResNet50	折叠 1	78	74	79	76
	折叠 2	80	75	78	74
	折叠 3	77	75	79	75
	折叠 4	78	76	77	74
	Mean	78	75	78	75
MobileNetV2	折叠 1	80	75	78	77
	折叠 2	77	76	80	78
	折叠 3	79	77	75	77
	折叠 4	80	75	78	77
	Mean	79	76	78	77
LDC-Model	折叠 1	80	78	81	79
	折叠 2	79	76	78	77
	折叠 3	80	77	80	80
	折叠 4	80	76	79	78
	Mean	80	77	80	79

2. 模型精度测试

与其他预训练模型相比,本研究模型对于 20% 验证数据集中的四种图像的分类准确率最高,在 K 折交叉验证过程中同样取得了相似的性能,进一步排除了数据集本身的随机误差影响,证明了该模型对 CXR 图像分类的稳健性。因此,本研究决定利用 20% 的测试数据集对该模型进行最终验证,并以 4×4 混淆矩阵的形式报告模型分类准确程度,如图 5 所示。其中每行和每列表示一个图像类,矩阵的主对角线以外的单元格表示被评估为假阳性和假阴性图像的数量。在肺病分类模型的 4232 张图像测试结果中,仅有少数图像被错误分类,且大多数发生在轻度肺部感染和正常 X 光图像类别之间。

综上所述,本研究提出的基于深度学习的肺部影像识别及分类模型表现较好,对 CXR 图像的分类效率较高,可应用于肺肿瘤、细菌性肺炎、病毒性肺炎以及其他类型肺部疾病的诊断和预测;此外,该模型开发成本较小,可通过扩展将其作为建立肺部疾病识别系统的基础,在疾病辅助诊断和预防方面具有很好的应用前景。

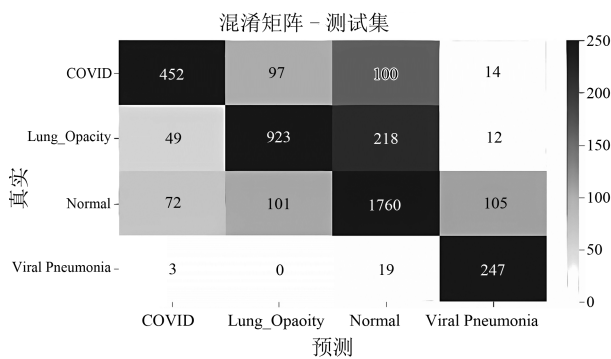


图 5 模型评估混淆矩阵图

讨论

本文以卷积神经网络为基础构建肺病分类模型,充分纳入图像特征,搭建通道筛选模块,增强重要特征信息,以肺部磨玻璃样阴影、纹理、病灶密度及肺间质形态等特征作为划分指标,利用特征识别和提取的方式将肺部影像图片持续分割,将四类不同图像高精度的区分开来,进而完成对不同疾病的分类识别。本研究模型具有唯一性、可靠性和自主开发的特点,识别准确率较高且各项指标较优,其分类结果可为肺部疾病提供基层临床辅助检测手段,协助医学工作者避免因肉眼的局限性而产生的错误判断,很大程度提高了诊断效率,在肺病分类预测方面表现出巨大的潜力。

在肺部疾病频发的条件下,通过深度学习算法构建肺病分类模型对疾病预防和临床救治具有十分重要的意义,及时准确发现疾病进展是遏制病情发展的突出前提,随着机器学习的不断发展,人工智能技术正在

进一步开发此类疾病的自动化检测方案。目前有多种适用于图像分类的深度学习方法,如 ResNet-101、VGG-16、VGG-19、AlexNet 等^[16],此类模型通过把 2D 成像图片或 3D 扫描图片以矩阵的方式读入模型,获取相关图像特征并进行分类,以准确度、特异度、灵敏度、宏观率、召回率等指标作为最终评估标准。然而不同模型在相同数据集上的准确度差别较大,其余指标优劣程度水平不一,出现这种现象是由于模型架构不同、图像特征提取方式不同等诸多因素^[17]。

本研究模型具有相对简单的架构,可以通过添加额外的卷积层或调整其他模型组件将模型扩展,然而在最终的精度测试中,有少数图像被肺病分类模型错误分类,这可能是由于图像分辨率降低和数据数量有限,导致模型对图像准确预测受到影响。因此在科学研究中要纳入不同方法,比较不同模型区分度、准确度、复杂程度的差异,通过计算机视觉技术(如拉伸、切片和直方图均衡化)提高分类图像的质量以对模型进一步改进^[18],过程中同样需要探讨模型与算法的结合,致力于开发出相关人工智能系统,使得医务工作者可以在临床环境中轻松测试 AI 模型并捕获相应结果,通过 CXR 图像准确预测肺部相关疾病,以准确高效的诊断方式改善临床治疗环境。

此外,本模型的研究过程中存在一些局限性,如训练图像数据时计算量较大、对电脑配置要求较高且时间可能较长、缺少足够数量的影像数据进行模拟、数据的有限可用性会产生类不平衡约束的情况等。虽然本研究所用数据集较大,且通过 K 折交叉验证对数据进行了增强,但纳入更大特征值或样本量时模型表现还未可知^[19]。因此,在日后的工作中,应注意提高模型的准确性、可重复性及代码可用性的问题,更深层次的探讨不同模型对于临床辅助诊断的表现。

参考文献

- [1] Liu S, Liu G, Zhou H. A robust parallel object tracking method for illumination variations. *Mob Netw Appl*, 2019, 24(1): 5-17.
- [2] 郭艺,杜秋晨,吴朦朦,等.基于轻量级神经网络的新肺炎 CT 新型识别技术. *中国医学物理学杂志*, 2022, 39(10): 1263-1269.
- [3] 沈昊钧. 基于 CT 图像的新肺炎检测研究. 电子科技大学, 2022.
- [4] Lee S, Negishi M, Urakubo H, et al. Mu-Net: multi-scale U-net for two-photon microscopy image denoising and restoration. *Neural Netw*, 2020, 125(26): 92-103.
- [5] 王光宇. 基于深度学习的新肺炎 CT 图像检测方法研究. 东华大学, 2022.
- [6] Li Q, Guan X, Wu P, et al. Early Transmission Dynamics in Wuhan, China, of NoveCoronavirus-Infected Pneumonia. *N Engl J Med*, 2020, 382(13): 1199-1207.
- [7] Anisotropic Diffusion Filtering Operation and Limitations - Magnetic Resonance Imaging Evaluation. *IFAC Proc*, 2014, 47(3): 3887-3892.

(下转第 375 页)