

基于多任务学习构建尘肺病鉴别诊断及分期模型

彭姗姗¹, 韩梦茹², 陈晴¹, 刘丽芳¹, 周家青¹, 杜文¹, 周丁子¹, 伏代刚¹, 周敏¹, 史莹¹, 张勤¹, 周颖杰²,
张凌¹, 彭莉君¹, 姚于勤¹, 沈江¹, 张本¹, 伍东升¹

1. 四川大学华西公共卫生学院 / 华西第四医院, 华西职业性尘肺病队列工作组, 四川 成都 610041;

2. 四川大学计算机学院, 四川 成都 610041

摘要:目的 基于多任务学习, 构建深度学习模型辅助临床医师进行尘肺病鉴别诊断及分期。方法 收集 2011—2022 年四川省某职业病医院 3 600 名就诊者的数字化摄影后前位胸片, 使用全卷积神经网络(UNet)进行肺野分割。基于多任务学习, 利用任务间相关性, 以 DenseNet121 为骨干网络, 添加两个分类器构建多任务模型, 并在 ChestX-ray14 数据集上预训练。采用配对 t 检验比较单任务模型与多任务模型之间准确率、精确度、灵敏度、F1 分数的差异。结果 测试集结果显示, 单任务模型鉴别诊断和诊断分期性能约 90% 和 77%。多任务模型鉴别诊断和诊断分期性能约 94% 和 86%, 分别高于单任务模型约 4% 和 9%, 评价指标间差异具有统计学意义($P < 0.05$)。结论 多任务模型较单任务模型更具优势, 可有效实现尘肺病与肺结核的鉴别诊断和尘肺病精准分期。

关键词:尘肺; 肺结核; 鉴别诊断; 深度学习

中图分类号: R135.2 文献标志码: A 文章编号: 1003-8507(2024)07-1187-07

DOI: 10.20043/j.cnki.MPM.202310099

Construction of differential diagnosis and staging model of pneumoconiosis based on multi-task learning

PENG Shan-shan*, HAN Meng-ru, CHEN Qing, LIU Li-fang, ZHOU Jia-qing, DU Wen, ZHOU Ding-zi, FU Dai-gang,
ZHOU Min, SHI Ying, ZHANG Qin, ZHOU Ying-jie, ZHANG Ling, PENG Li-jun, YAO Yu-qin, SHEN Jiang,
ZHANG Ben, WU Dong-sheng

*West China School of Public Health/West China Fourth Hospital, Sichuan University, Chengdu, Sichuan 610041, China

Abstract: Objective To construct a deep learning model based on multi-task learning to assist clinicians in differential diagnosis and staging of pneumoconiosis. **Methods** The digital chest radiographs of 3 600 patients from an occupational disease hospital in Sichuan Province from 2011 to 2022 were collected, and the full convolution neural network (UNet) was used to segment the lung field. Based on multi-task learning, the multi-task model was constructed using the correlation between tasks. The multi-task model was pre-trained on the ChestX-ray14 dataset, whose backbone network was DenseNet121, and two classifiers were added behind the backbone network. Paired t -test was used to compare the differences in accuracy, precision, sensitivity, and F1 scores between single-task model and multi-task model. **Results** The test set results showed that the differential diagnosis and diagnostic staging performance of the single-task model was about 90% and 77%, respectively. The differential diagnosis and diagnostic staging performance of the multi-task model was about 94% and 86%, which was higher than that of the single-task model about 4% and 9%, respectively. The difference between the evaluation indexes was statistically significant ($P < 0.05$). **Conclusion** The multi-task model has more advantages than the single-task model and can effectively realize the differential diagnosis and accurate staging of pneumoconiosis and pulmonary tuberculosis.

Keywords: Pneumoconiosis; Pulmonary tuberculosis; Differential diagnosis; Deep learning

尘肺病是由于职业活动中长期吸入生产性矿物性粉尘并在肺内滞留而引起的以肺组织弥漫性纤维

基金项目: 国家自然科学基金项目(U22A20359, 82373548); 四川省科技厅自然科学基金项目(2023NSFSC0647, 2023NSFSC1729, 2023NSFSC0649, 2023NSFSC1965)

作者简介: 彭姗姗(1997—), 女, 硕士在读, 研究方向: 公共卫生

通信作者: 伍东升, E-mail: wdsxysy@163.com

化为主的疾病, 目前仍无法治愈。据全球疾病负担(2019年)研究数据显示, 我国职业性尘肺病死亡数达 1 302 例, 因尘肺病造成的残疾调整生命年(DALYs)为 41 359.25 人年^[1]。尘肺病分期及是否合并肺结核等因素, 将影响患者享受工伤保险待遇的水平和临床诊疗方案的选择。接尘工人的高千伏后前位胸片表现是尘肺病分期的主要依据。然而, 人工阅片较

为主观,费时费力,且诊断分期稳定性和准确性较差。据报道,尘肺病的个人两次阅片符合率仅为 15%~64%^[2]。吉林省 800 名诊断医师考核结果显示,期别符合率平均(71.5±10.0)%^[3]。广西省 8 个市级诊断机构的抽样调查发现,尘肺病漏诊率为 58.3%,误诊率为 16.7%^[4]。其中,壹期、贰期、叁期漏诊率分别为 61.5%、33.3%、75.0%。无尘肺和尘肺壹期最易产生诊断误差,诊断准确性主要受阅片者的经验影响。肺结核和尘肺病的影像学特征相似,其“异病同影”的现象导致尘肺病易误诊为肺结核。据统计,在尘肺病的误诊疾病中,约 58%被误诊为肺结核,其中壹期尘肺和肺结核的误诊率最高,最常见的误诊原因是阅片经验不足。因此,尘肺病的准确诊断与分期是我国尘肺病防治面临的一个巨大挑战。

目前,数字化胸片(digital chest radiography, DR)已成为尘肺病与肺结核病筛查和诊断的重要工具。计算机辅助诊断(computer-aided diagnosis, CAD)可减少人工阅片的主观差异,提高诊断效率和诊断准确性。大量研究表明,基于胸片的 CAD 表现优于人类专家^[5-8]。既往相关研究倾向于尘肺病筛查^[9-10]而非分期^[11]。Zhang 等人^[12]针对六个肺区分别构建卷积神经网络分类模型,综合六个亚区的预测结果确定最终诊断,得到目前最佳的分期诊断效能。该模型的最高分期准确率为 92.70%,但贰期和叁期的 F1 分数与灵敏度在 54%~67%,有待提升。国内外学者开发深度学习模型进行结核病检测^[7]、病灶定位和亚型鉴别诊断^[13]等,尚未有研究关注尘肺病和肺结核的鉴别诊断。

综上所述,尘肺病误诊为肺结核的误诊率居高不下,其诊断分期结论的稳定性和准确性较差。CAD 在两者鉴别诊断中的研究仍处于空白状态,在诊断分期中的性能有待提升。多任务学习通过使用相关任务间的领域知识来改善模型的泛化性能^[14],已被广泛应用于肺部疾病的分割^[15]、分类^[16-17]及预测^[18-19]等任务中。大量研究表明,在医学影像领域,多任务模型较单任务模型表现更好^[20]。为此,本研究基于多任务学习,创新地开发多任务模型,同时完成以下两个任务:(1)尘肺病与肺结核的鉴别诊断;(2)尘肺病的诊断分期。利用任务间相关性促进学习,提高模型的鲁棒性和准确性。辅助临床医师诊疗决策,实现尘肺病精确诊断。

1 资料与方法

1.1 数据来源 研究对象来源于四川省某职业病医院 2011 年 10 月—2022 年 12 月申请尘肺病诊断的接尘工人、门诊与住院患者及健康检查人群,共 3 600

例。规范化采集所有研究对象的数字化摄影后前位胸片,严格进行质量控制。成像设备采用德国 Siemens (AXIOM Aristos VX Plus)、FUJIFILM 数字化 X 线摄影装置。接尘工人的 DR 严格按照 GBZ 70-2009、GBZ 70-2015 附录 F 的技术要求完成。纳入标准:(1)健康对照:后前位胸片显示心肺未见确切异常且无粉尘接触史;(2)肺结核患者:无粉尘接触史的肺结核患者;(3)接尘工人:根据可靠的生产性矿物性粉尘接触史,以技术质量合格的后前位胸片表现为主要依据,结合工作场所情况和职业健康监护资料,参考临床表现和实验室检查,排除其它类似肺部疾病后,对照尘肺病诊断标准片,诊断为无尘肺、壹期、贰期和叁期尘肺。同时,未合并肺结核。此外,本研究未排除肺心病、肺气肿、气胸等其他肺部疾病。接尘工人的胸片由三位获得国家职业病诊断资质的尘肺病诊断专家(从事职业病诊断 15 年以上)讨论后确定尘肺病期别,并出具职业病诊断证明书。肺结核的最终诊断结果由临床病史、症状与体格检查、实验室检查、影像学检查结果共同确定。为获得高质量的数据,正常人和肺结核患者的胸片由两名高年资主治医师(均有 8 年以上诊断经验)标注后,再由一名高级职称胸部影像学专家核查标记准确性。未满足上述胸片标注条件且未达到标准要求的三、四级片未纳入本研究。所有数据已脱敏。本研究获得四川大学华西第四医院伦理委员会审核批准(批准号 HXSJ-EC-2023042)。

1.2 数据预处理 研究数据以医学数字成像和通信(DICOM)格式存储,使用 Unet 进行肺野分割。为避免过拟合,训练集数据采用如下仿射变换:在 98%~102%之间随机裁剪,然后缩放至 512 像素×512 像素,在 -15~15 度之间随机旋转,随机水平和垂直平移 2%。最后,所有图像的像素值归一化至(-1,1)。验证集和测试集仅采用像素归一化,未进行其他仿射增强操作。

1.3 模型构建 Dvsnath^[21]等人比较七种用于尘肺病检测的深度学习模型(包括 VGG16,VGG19,InceptionV3,ResNet50,Xception,DenseNet121^[22]和 CheXNet),得出与其他研究相似的结论,即 CheXNet 模型在尘肺检测任务上具有较好的表现。CheXNet 是由斯坦福大学开发的 121 层稠密卷积网络(DenseNet121),已在 ChestX-ray14 数据集上预训练。因此,为证明多任务模型的有效性,单任务模型和多任务模型均采用预训练的 DenseNet121 作为骨干网络提取图像特征。与其他研究^[23-24]类似,本研究使用全局平均池化层和两个全连接层替换 DenseNet121 末端全连接层构建新的分类层。两个全连接层中间添加丢弃层(dropout layer),分别以 0.25、0.50、0.75、0.90 的概率随机丢弃神

经元。与单任务模型不同,多任务模型在骨干网络后添加两个分支,经 Softmax 激活函数将模型输出转换为相应类别的概率分数。模型训练时,使用交叉熵损失函数,通过反向传播更新参数。根据输出及对应的标签,计算两个任务的损失(Loss1 和 Loss2)。Loss1、Loss2 分别优化分类器一和分类器二的参数,两者的加权和更新多任务模型骨干网络的参数,加权和损失

计算如下。

$$\text{Loss}_{\text{加权和}} = (1 - \omega) \times \text{Loss1} + \omega \times \text{Loss2}$$

$$\omega = \frac{N2}{N1 + N2}$$

其中,N1 和 N2 分别表示每次迭代时任务一和任务二的样本数,权重 ω 为任务二样本数所占比例。研究总体流程见图 1。

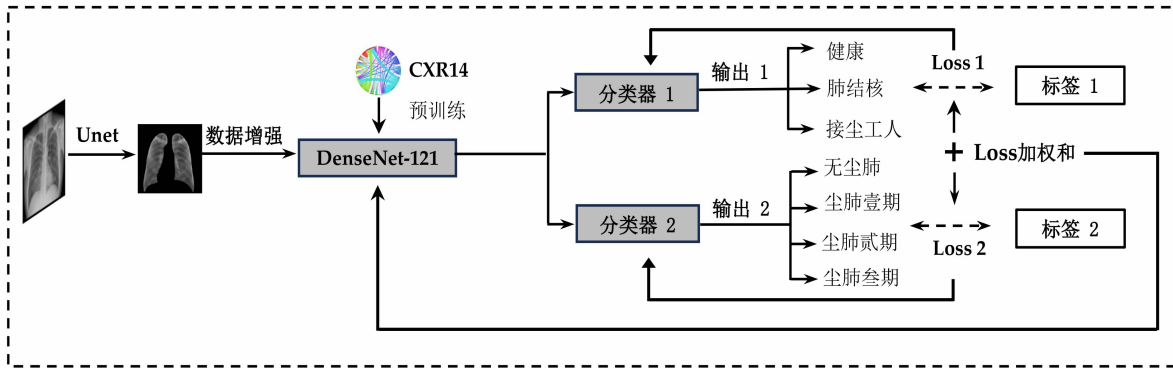


图 1 研究流程

Figure 1 Study workflow

1.4 实验平台与参数设置 本研究编程语言为 Python3.10.9,所有实验基于 Pytorch 深度学习框架,实验平台为 NVIDIA Quadro RTX 6000 GPU (4 608 流核,24 576 MB 内存)。模型初始学习率为 1×10^{-5} ,批处理大小为 16,epoch 为 50,使用 ReduceLROnPlateau 学习率调度器和 Adam 优化器。

1.5 统计分析 采用 Python3.10 中的 Scipy 库进行统计学分析,模型间的性能比较采用配对 *t* 检验,检验水准 $\alpha=0.05$ (双侧)。数据集随机划分为训练集 (3 240 例,90%)和测试集 (360 例,10%)。为准确、可靠的评估模型,调整超参数时采用十折交叉验证训练并验证模型,每折类别均衡。最终模型在训练集上重新训练,并在测试集上测试。使用准确率、精确率、灵敏度、特异度、F1 分数、受试者工作特征曲线(receiver operating characteristic curve, ROC) 及曲线下面积 (area under curve, AUC) 评价模型效果,数据可视化处理采用 Python3.10 的 Matplotlib 软件包。各指标计算公式如下:

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

$$\text{精确率} = \frac{TP}{TP+FP}$$

$$\text{灵敏度} = \frac{TP}{TP+FN}$$

$$\text{特异度} = \frac{TN}{TN+FP}$$

$$\text{F1 分数} = \frac{2(\text{precision} \times \text{recall})}{\text{precision} + \text{recall}}$$

2 结果

2.1 一般资料 本研究共纳入 3 600 名研究对象,健康对照、肺结核和接尘工人各 1 200 例。其中,接尘工人的男性占比达 99%,年龄范围,接尘类型等人口统计数据见表 1。

表 1 接尘工人的特征

Table 1 Characteristics of dust-exposed workers

特征	尘肺(n=900)	无尘肺(n=300)
男性[n (%)]	894 (99.33)	298 (99.33)
年龄(岁)	32 ~ 89	33 ~ 89
接尘类型[n (%)]		
煤尘	542 (60.22)	176 (58.67)
矽尘	347 (38.56)	57 (19.00)
其他	11 (1.22)	67 (22.33)
接尘年限(年)	1 ~ 50	2 ~ 49
类别		
非尘肺		300
壹期	300	
贰期	300	
叁期	300	

2.2 不同丢弃率下单任务模型与多任务模型的预测结果 由于数据和代码未公开,无法对既往研究进行准确评价。为证明多任务模型的有效性,在相同的训练集上训练单任务模型,分别实现任务一和任务二。表 2 为不同丢弃率下单任务模型十折交叉验证的结果。当丢弃率为 0.75 时,单任务模型在两个任务中分类性能较好。任务一各评价指标均高于任务二 10% 左右,表明深度学习模型更容易识别健康人、肺结核患者和接尘工人的影像学特征。

表 2 不同丢弃率下单任务模型的性能

Table 2 Performance of single-task model with different dropout rates

任务	丢弃率	准确率 (%) (95%CI)	精确度 κ (%) (95%CI)	灵敏度 κ (%) (95%CI)	F1* 分数 (%) (95%CI)
任务一	0.25	89.23 (89.09 ~ 89.37)	89.58 (89.48 ~ 89.69)	89.25 (89.13 ~ 89.37)	89.19 (89.08 ~ 89.30)
	0.50	87.41 (86.22 ~ 88.60)	89.19 (88.56 ~ 89.83)	88.50 (87.65 ~ 89.35)	88.60 (87.75 ~ 89.44)
	0.75 ^a	89.88 ^a (89.67 ~ 90.08) ^a	90.25 ^a (90.04 ~ 90.46) ^a	89.89 ^a (89.68 ~ 90.10) ^a	89.85 ^a (89.62 ~ 90.09) ^a
	0.90	89.59 (89.44 ~ 89.74)	89.87 (89.72 ~ 90.02)	89.54 (89.39 ~ 89.69)	89.44 (89.28 ~ 89.60)
	0.25	78.02 (77.47 ~ 78.58)	78.22 (77.62 ~ 78.82)	78.13 (77.59 ~ 78.68)	78.03 (77.45 ~ 78.62)
任务二	0.50	77.84 (77.41 ~ 78.28)	78.00 (77.52 ~ 78.48)	77.63 (77.23 ~ 78.03)	77.72 (77.24 ~ 78.20)
	0.75 ^a	78.44 ^a (77.59 ~ 79.30) ^a	79.34 ^a (78.46 ~ 80.22) ^a	79.16 ^a (78.32 ~ 80.00) ^a	78.76 ^a (77.90 ~ 79.62) ^a
	0.90	77.94 (77.47 ~ 78.42)	78.48 (78.04 ~ 78.93)	77.91 (77.44 ~ 78.39)	77.84 (77.37 ~ 78.30)

注: * 为宏平均值; ^a 为最优结果。

多任务模型两个丢弃层使用相同的丢弃率进行训练,表 3 为十折交叉验证的结果。与单任务模型相同,多任务模型在丢弃率为 0.75 时分类性能最佳,绘

制此丢弃率下多任务模型十折交叉验证的平均 ROC 曲线及 95% 置信区间,见图 2、3。

表 3 不同丢弃率下多任务模型的性能

Table 3 Performance of multi-task model with different dropout rates

任务	丢弃率	准确率 (%) (95%CI)	精确度 κ (%) (95%CI)	灵敏度 κ (%) (95%CI)	F1 分数 κ (%) (95%CI)
任务一	0.25	93.38 (93.23 ~ 93.54)	93.53 (93.35 ~ 93.71)	93.40 (93.15 ~ 93.64)	93.33 (93.13 ~ 93.50)
	0.50	93.63 (93.44 ~ 93.82)	93.73 (93.57 ~ 93.90)	93.63 (93.45 ~ 93.80)	93.60 (93.41 ~ 93.78)
	0.75 ^a	94.14 ^a (94.14 ~ 94.20) ^a	94.21 ^a (94.13 ~ 95.30) ^a	94.18 ^a (94.10 ~ 94.25) ^a	94.09 ^a (94.00 ~ 94.19) ^a
	0.90	93.61 (93.50 ~ 93.73)	93.80 (93.70 ~ 93.91)	93.61 (93.49 ~ 93.73)	93.57 (93.45 ~ 93.69)
	0.25	87.64 (87.33 ~ 87.94)	88.76 (88.32 ~ 89.20)	88.17 (87.53 ~ 88.80)	88.18 (87.56 ~ 88.80)
任务二	0.50	89.06 (88.92 ~ 89.21)	90.07 (89.74 ~ 90.41)	89.43 (89.06 ~ 89.80)	89.53 (89.20 ~ 89.86)
	0.75 ^a	89.25 ^a (88.90 ~ 89.60) ^a	90.33 ^a (89.99 ~ 90.68) ^a	89.72 ^a (89.34 ~ 90.11) ^a	89.83 ^a (89.46 ~ 90.20) ^a
	0.90	86.76 (86.55 ~ 86.97)	87.94 (87.70 ~ 88.18)	87.22 (87.00 ~ 87.45)	87.47 (87.24 ~ 87.69)

注: * 为宏平均值; ^a 为最优结果。

2.3 单任务模型和多任务模型的性能比较 设置丢弃率为 0.75,采用不同的随机种子在训练集上重复训练单任务模型和多任务模型 40 次,在测试集上测试,结果见表 4、5。使用配对 t 检验比较模型性能,结果

显示多任务模型对叁期尘肺 F1 分数无提升外,其他评价指标间差异均具有统计学意义,且多任务模型诊断性能优于单任务模型。

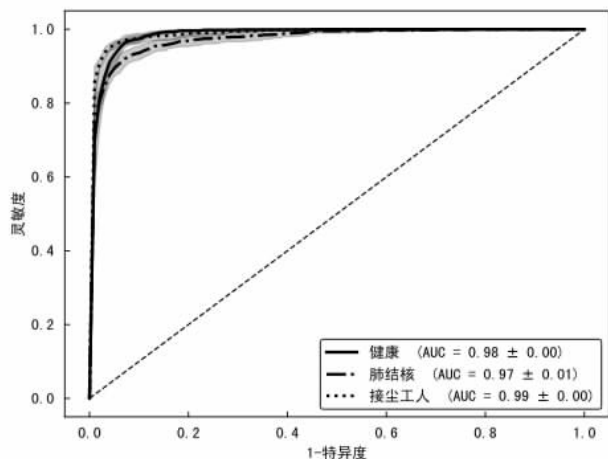


图 2 任务一:多任务模型 ROC 曲线

Figure 2 Task 1: ROC curves of multi-task model

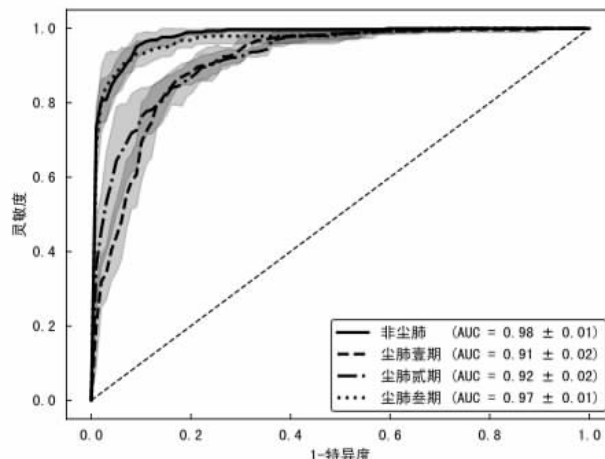


图 3 任务二:多任务模型 ROC 曲线

Figure 3 Task 2: ROC curves of multi-task model

表 4 单任务模型与多任务模型对任务一的性能比较($\bar{x} \pm s$)

Table 4 Performance comparison of Task 1 between single-task model and multi-task model ($\bar{x} \pm s$)

评价指标	模型	健康人	肺结核	接尘工人	平均值
准确率 (%)	单任务模型	—	—	—	90.03 ± 0.67
	多任务模型	—	—	—	94.13 ± 0.80
<i>t</i> 值		—	—	—	22.54
<i>P</i> 值		—	—	—	<i>P</i> <0.001
精确度 (%)	单任务模型	88.56 ± 2.42	93.11 ± 1.89	89.25 ± 3.01	90.31 ± 0.55
	多任务模型	93.34 ± 1.04	97.16 ± 1.08	93.34 ± 1.04	94.30 ± 0.73
<i>t</i> 值		10.10	10.03	4.59	24.78
<i>P</i> 值		<i>P</i> <0.001	<i>P</i> <0.001	<i>P</i> <0.001	<i>P</i> <0.001
灵敏度 (%)	单任务模型	92.13 ± 3.41	83.56 ± 1.94	94.42 ± 2.77	90.03 ± 0.67
	多任务模型	97.83 ± 1.03	87.85 ± 2.28	96.69 ± 0.83	94.13 ± 0.80
<i>t</i> 值		9.62	8.83	5.01	22.54
<i>P</i> 值		<i>P</i> <0.001	<i>P</i> <0.001	<i>P</i> <0.001	<i>P</i> <0.001
F1 分数 (%)	单任务模型	90.22 ± 1.01	88.04 ± 0.64	91.68 ± 0.83	89.98 ± 0.68
	多任务模型	95.53 ± 0.36	92.25 ± 1.30	94.47 ± 1.12	94.08 ± 0.82
<i>t</i> 值		33.71	18.76	17.73	32.35
<i>P</i> 值		<i>P</i> <0.001	<i>P</i> <0.001	<i>P</i> <0.001	<i>P</i> <0.001

表 5 单任务模型与多任务模型对任务二的性能比较($\bar{x} \pm s$)

Table 5 Performance comparison of Task 2 between single-task model and multi-task model ($\bar{x} \pm s$)

评价指标	模型	非尘肺 (%)	壹期 (%)	贰期 (%)	叁期 (%)	平均值 (%)
准确率 (%)	单任务模型	—	—	—	—	77.48 ± 1.54
	多任务模型	—	—	—	—	86.33 ± 2.42
<i>t</i> 值		—	—	—	—	20.30
<i>P</i> 值		—	—	—	—	<i>P</i> <0.001
精确度 (%)	单任务模型	77.20 ± 4.76	70.90 ± 3.32	77.06	85.69 ± 3.71	77.71 ± 1.47
	多任务模型	95.37 ± 4.05	81.64 ± 2.39	80.01 ± 4.34	89.12 ± 4.94	86.54 ± 2.76
<i>t</i> 值		21.86	19.94	3.26	3.30	17.84
<i>P</i> 值		<i>P</i> <0.001	<i>P</i> <0.001	<i>P</i> =0.002	<i>P</i> =0.002	<i>P</i> <0.001
灵敏度 (%)	单任务模型	88.00 ± 6.22	70.17 ± 5.06	63.08 ± 6.55	88.67 ± 1.65	77.48 ± 1.54
	多任务模型	94.92 ± 2.61	87.58 ± 6.13	76.585 ± 9.67	86.25 ± 1.72	86.33 ± 2.42
<i>t</i> 值		6.50	11.79	10.50	6.75	20.30
<i>P</i> 值		<i>P</i> <0.001	<i>P</i> <0.001	<i>P</i> <0.001	<i>P</i> <0.001	<i>P</i> <0.001
F1 分数 (%)	单任务模型	81.98 ± 2.62	70.34 ± 2.36	69.04 ± 3.32	87.11 ± 2.14	77.12 ± 1.70
	多任务模型	94.90 ± 2.00	84.15 ± 2.69	77.90 ± 4.48	87.32 ± 2.24	86.07 ± 2.37
<i>t</i> 值		28.85	22.34	8.58	3.59	17.02
<i>P</i> 值		<i>P</i> <0.001	<i>P</i> <0.001	<i>P</i> <0.001	<i>P</i> =0.722	<i>P</i> <0.001

3 讨论

尘肺病的及时发现与准确诊断,能显著地改变其转归和预后。与既往研究不同,本研究从临床实际问题出发,首次将多任务学习应用到尘肺病诊断中,结合鉴别诊断和诊断分期两个任务构建多任务模型。基于任务间较强的相关性,共享特征表示提高模型准确性和泛化性。

研究结果显示,卷积神经网络在尘肺病和肺结核鉴别诊断中表现较好,其准确率、精确度、灵敏度及 F1 分数均在 90% 左右,远高于尘肺病分期诊断性能,说明模型更易识别肺结核和接尘工人的胸部影像学特征。Dong 等人^[25]开发深度学习模型将接尘工人的胸部 X 线临床影像学特征分为三类,达到 98% 的平均准确率,揭示了接尘工人胸部 X 线影像特征的独特性,与本研究在尘肺病与肺结核鉴别诊断任务中取得较好性能的结论相符说明模型,比人类专家更容易识别两者的影像学鉴别要点。

通过对研究结果对比分析发现,多任务模型鉴别诊断和分期诊断的准确率、精确度、灵敏度和 F1 分数分别高于单任务模型约 4% 和 9%。肺结核常表现为模糊结节、斑片与条索影,好发于肺尖与下叶背段。而尘肺患者多为双肺弥漫性的圆形小阴影,胸片改变具有一定特征性。基于尘肺病和肺结核影像特征的相似性,通过一些共享表示将两个任务联系在一起。改善模型泛化能力,提升模型预测性能。Wu 等人^[26]基于多任务学习,联合分割和分类任务构建 JCS 系统用于 COVID-19 的诊断,证明分类特征和分割特征相结合,可以得到更好的预测结果。Goncharov 等人^[27]构建多任务模型,同时实现新冠肺炎的诊断和严重性评估,证明多任务框架可促进两个相关任务的学习,与本研究结论一致。本研究数据集类别均衡,模型可充分学习到各类别特征,各类别 AUC 均在 90% 以上,表明所构建的多任务模型具有良好的诊断性能。

综上所述,多任务模型可有效实现尘肺病与肺结核的鉴别诊断和尘肺病精准分期。其中,鉴别诊断性能达 94%,尘肺病诊断分期性能达 86%,值得进一步推广应用。但本研究仍存在一定局限性,首先,数据来自同一机构,模型泛化能力尚需进一步研究。其次,对于深度学习来说,样本量仍需扩大。最后,本研究仅参考数字化摄影后前位胸片诊断标准,而中国卫生监督协会于 2023 年 2 月 27 日发布《尘肺胸部 CT 辅助诊断技术指南》,故未来研究应考虑结合 CT,纳入人口学特征、接尘类型、接尘年限等不同数据类型建模。

利益冲突声明 本研究不存在任何利益冲突

参考文献

- [1] Anonym. A systematic analysis for the global burden of disease study 2019[J]. *Lancet*, 2020, 396(10258): 1204-1222.
- [2] 张镛琢,梁实,李智民. 1990-2018 年我国尘肺病诊断读片差异的中文文献研究 [J]. *职业卫生与应急救援*, 2019, 37(5): 420-424, 452.
Zhang LZ, Liang S, Li ZM. Deviation of X-ray diagnosis of pneumoconiosis in China: a literature review from 1990 to 2018[J]. *Occupational Health and Emergency Rescue*, 2019, 37(5): 420-424, 452.
- [3] 邢军,李岩,孙大为,等. 吉林省尘肺病诊断医师读片考核结果分析[J]. *海峡预防医学杂志*, 2015, 21(6): 77-78.
Xing J, Li Y, Sun DW, et al. Analysis on the examination results of pneumoconiosis diagnosis doctors reading films in Jilin Province[J]. *Strait Journal of Preventive Medicine*, 2015, 21(6): 77-78.
- [4] 秦克江,叶绍色,王力珩,等. 广西市级尘肺病诊断机构诊断质量调查研究 [J]. *中国煤炭工业医学杂志*, 2014, 17(8): 1324-1326.
Qin KJ, Ye SS, Wang LH, et al. Investigation and study on diagnostic quality of pneumoconiosis diagnosis institutions in Guangxi [J]. *Chinese Journal of Coal Industry Medicine*, 2014, 17(8): 1324-1326.
- [5] Khan FA, Majidulla A, Tavaziva G, et al. Chest x-ray analysis with deep learning-based software as a triage test for pulmonary tuberculosis: a prospective study of diagnostic accuracy for culture-confirmed disease[J]. *Lancet Digital Health*, 2020, 2(11): e573-e581.
- [6] Sugauna N, Yoshida SNH, Takeuchi Y, et al. Artificial intelligence in quantitative chest imaging analysis for occupational lung disease [J]. *Seminars in Respiratory and Critical Care Medicine*, 2023, 44(3): 362-369.
- [7] Lee S, Yim JJ, Kwak N, et al. Deep learning to determine the activity of pulmonary tuberculosis on chest radiographs [J]. *Radiology*, 2021, 301(2): 435-442.
- [8] Schwalbe N, Wahl B. Artificial intelligence and the future of global health[J]. *The Lancet*, 2020, 395(10236): 1579-1586.
- [9] Wang Y, Cui FT, Ding XP, et al. Automated identification of the pre-clinical stage of coal workers' pneumoconiosis from digital chest radiography using three-stage cascaded deep learning model [J]. *Biomedical Signal Processing and Control*, 2023, 83: 104607.
- [10] Devnath L, Luo SH, Summons P, et al. Automated detection of pneumoconiosis with multilevel deep features learned from chest X-Ray radiographs [J]. *Computers in Biology and Medicine*, 2021, 129: 104125.
- [11] Sun WJ, Wu DS, Luo Y, et al. A fully deep learning paradigm for pneumoconiosis staging on chest radiographs [J]. *IEEE Journal of Biomedical and Health Informatics*, 2022, 26(10): 5154-5164.
- [12] Zhang LZ, Rong RC, Li QW, et al. A deep learning-based model for screening and staging pneumoconiosis [J]. *Scientific Reports*, 2021, 11(1): 2201.
- [13] Wang L, Ding WL, Mo Y, et al. Distinguishing nontuberculous mycobacteria from *Mycobacterium tuberculosis* lung disease from CT images using a deep learning framework[J]. *European Journal of Nuclear Medicine and Molecular Imaging*, 2021, 48(13): 4293-4306.
- [14] Caruana R. Multitask learning [J]. *Machine Learning*, 1997, 28(1): 41-75.
- [15] Graham S, Vu QD, Jahanifar M, et al. One model is all you need: Multi-task learning enables simultaneous histology image segmenta

- The case of Turkish hospitals [J]. International Journal of Disaster Risk Reduction, 2020, 49: 101748.
- [8] Li ZG, Wei H. A comprehensive evaluation of China's TCM medical service system: an empirical research by integrated factor analysis and TOPSIS[J]. Front Public Health, 2020, 8: 532420.
- [9] Chen PY. Effects of normalization on the entropy-based TOPSIS method[J]. Expert Systems With Applications, 2019, 136: 33-41.
- [10] 周兆菲,何蕾,柴琪,等. 四川省基层医疗卫生服务提供水平综合评价[J]. 现代预防医学, 2021, 48(17): 3146-3149, 3161. Zhou ZF, He L, Chai Q, et al. Comprehensive evaluation on the level of primary medical and health services, Sichuan [J]. Modern Preventive Medicine, 2021, 48(17): 3146-3149, 3161.
- [11] 田凤调. RSR 法中的分档问题[J]. 中国卫生统计, 1993, (2): 26-28. Tian FD. Classification problems in RSR method[J]. Chinese Journal of Health Statistics, 1993, (2): 26-28.
- [12] 邱乐平,秦莹,阴旅宁,等. 2021 年四川省职业卫生技术服务机构服务能力现状调查[J]. 预防医学情报杂志, 2022, 38(4): 533-537. Qiu LP, Qin Y, Yin LN, et al. Survey on the status of the service capacity of occupational health technical service institutions in Sichuan Province in 2021 [J]. Journal of Preventive Medicine Information, 2022, 38(4): 533-537.
- [13] 朱少芳,冯玉超,张晋蔚,等. 广州市区级疾病预防控制中心职业健康技术支持能力现状调查 [C]//广州市卫生健康宣传教育中心. 广州市第十三届健康教育与健康促进学术交流活动稿集. 广州, 2022, 广州:广州市卫生健康宣传教育中心, 2022: 7. Zhu SF, Feng YC, Zhang JW, et al. Investigation on occupational health technical support ability of district-level centers for disease control and prevention in Guangzhou City [C]//Guangzhou Health Education Center. Collection of the 13th Guangzhou Health Education and health promotion academic exchange activities. Guangzhou, 2022, Guangzhou: Guangzhou Health Education Center, 2022: 7.
- [14] 邱乐平,阴旅宁,张成云. 四川省市级疾病预防控制中心职业卫生技术服务能力现状调查及对策研究[J]. 预防医学情报杂志, 2019, 35(11): 1245-1249. Qiu LP, Yin LN, Zhang CY. Investigation on occupational health service ability of city-level centers for disease control and prevention and its countermeasures in Sichuan Province [J]. Journal of Preventive Medicine Information, 2019, 35(11): 1245-1249.
- [15] 李亚娟,邢漪,康世娟,等. 云南省州(市)级疾病预防控制中心职业卫生服务能力现状调查 [J]. 职业与健康, 2017, 33(14): 1995-1999. Li YJ, Xing Y, Kang SJ, et al. Investigation on occupational health service ability of prefecture/city-level centers for disease control and prevention in Yunnan Province[J]. Occupation and Health, 2017, 33(14): 1995-1999.
- [16] 胡奎,毕珊,樊文力,等. 贵州省疾病预防控制中心职业卫生技术服务现状及对策研究 [J]. 微量元素与健康研究, 2017, 34(3): 52-54. Hu K, Bi S, Fan WL, et al. Current situation and countermeasures of occupational health service disease prevention and control in Guizhou province[J]. Studies of Trace Elements and Health, 2017, 34(3): 52-54.

收稿日期: 2023-09-02

(上接第 1192 页)

- tion and classification[J]. Medical Image Analysis, 2023, 83: 102685.
- [16] Jiang YM, Zhang ZC, Yuan QY, et al. Predicting peritoneal recurrence and disease-free survival from CT images in gastric cancer with multitask deep learning: a retrospective study [J]. Lancet Digital Health, 2022, 4(5): e340-e350.
- [17] Wang JX, Zheng YJ, Ma J, et al. Information bottleneck-based interpretable multitask network for breast cancer classification and segmentation[J]. Medical Image Analysis, 2023, 83: 102687.
- [18] Vandenhende S, Georgoulis S, Van Gansbeke W, et al. Multi-Task learning for dense prediction tasks: a survey [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44 (7): 3614-3633.
- [19] Liu YB, Li HN, Luo T, et al. Structural attention graph neural network for diagnosis and prediction of COVID-19 severity [J]. IEEE Transactions on Medical Imaging, 2023, 42(2): 557-567.
- [20] Zhao Y, Wang XY, Che TT, et al. Multi-task deep learning for medical image computing and analysis: A review[J]. Computers in Biology and Medicine, 2023, 153: 106496.
- [21] Devnath L, Luo SH, Summons P, et al. Performance comparison of deep learning models for black lung detection on chest x-ray radiographs [C]//Proceedings of the 3rd International Conference on Software Engineering and Information Management. Sydney, Australia, New York: Association for Computing Machinery, 2020.
- [22] Huang G, Liu Z, Maaten L, et al. Densely connected convolutional networks [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition. Honolulu: Institute of Electrical and Electronics Engineers, 2017.
- [23] Wang DD, Arzhaeva Y, Devnath L, et al. Automated pneumoconiosis detection on chest X-Rays using cascaded learning with real and synthetic radiographs [C]//Digital Image Computing: Techniques and Applications. Proceedings of the 2020 Digital Image Computing. Melbourne, Australia: Institute of Electrical and Electronics Engineers, 2020.
- [24] Wang XH, Yu JZ, Zhu Q, et al. Potential of deep learning in assessing pneumoconiosis depicted on digital chest radiography [J]. Occupational and Environmental Medicine, 2020, 77(9): 597-602.
- [25] Dong HT, Zhu BK, Zhang XR, et al. Use data augmentation for a deep learning classification model with chest X-ray clinical imaging featuring coal workers' pneumoconiosis [J]. BMC Pulmonary Medicine, 2022, 22(1): 271.
- [26] Wu YH, Gao SH, Mei J, et al. JCS: an explainable COVID-19 diagnosis system by joint classification and segmentation[J]. IEEE Transactions on Image Processing: a Publication of the IEEE Signal Processing Society, 2021, 30: 3113-3126.
- [27] Goncharov M, Pisov M, Shevtsov A, et al. CT-Based COVID-19 triage: deep multitask learning improves joint identification and severity quantification [J]. Medical Image Analysis, 2021, 71: 102054.

收稿日期: 2023-10-09