

基于DR图像的多模型融合用于壹期尘肺病筛查的研究

房 钊^{1,2}, 蒋天姿³, 张 军³, 孙振博³, 张慧坤³, 段 峰^{1,3*}

¹青岛大学青岛医学院, 山东 青岛

²淄博市第一医院放射科, 山东 淄博

³青岛大学附属医院放射科, 山东 青岛

收稿日期: 2026年3月21日; 录用日期: 2026年4月16日; 发布日期: 2026年4月21日

摘 要

目的: 基于胸部X线片影像, 分别构建放射组学模型与2D深度学习模型, 通过多因素Logistic回归对两种模型独立预测评分进行决策级融合, 构建尘肺病早期筛查联合模型, 评估其临床价值, 为模型选择及临床转化提供依据。材料与方法: 回顾性收集193例壹期尘肺病患者及178例健康体检者胸部X线片及临床资料, 采用ITK-SNAP软件在胸部X线片上完成双侧肺野感兴趣区(ROI)的手动勾画, 提取高通量放射组学特征, 构建放射组学模型; 同时基于勾画ROI为输入构建2D深度学习模型。通过多因素Logistic回归对两种单一模型的独立预测风险评分进行决策级融合, 构建联合预测模型。以受试者工作特征曲线下面积、准确率、灵敏度和特异度为核心指标, 通过训练集(70%)与独立测试集(30%)分层验证模型性能。采用校准曲线、决策曲线分析DCA (Decision Curve Analysis)评估模型校准性能与临床净获益, 并通过DeLong检验比较模型间的区分能力。引入三位均具有职业性尘肺病诊断资质放射科医师(高、中、低年资各1名)进行独立读片诊断, 作为人工对照, 采用双盲设计进行独立读片诊断。结果: 在独立测试集中, 放射组学模型、2D深度学习模型及联合模型的准确率依次为72.1%、66.7%、75.7%, AUC值分别为0.800 (95% CI: 0.718~0.882)、0.803 (95% CI: 0.718~0.888)、0.813 (95% CI: 0.731~0.894)。联合模型灵敏度为75.4%, 特异度为75.9%, 鉴别效能显著优于各单一模型。校准曲线结果显示, 联合模型预测概率与实际进展的发生概率一致性良好。决策曲线分析(DCA)结果表明, 在多数阈值概率范围内, 联合模型的净获益均高于单一模型。DeLong检验证实, 放射组学模型、2D深度学习模型与联合模型的AUC值两两比较, 差异均无统计学意义。在人机对比中, 联合模型的准确率及AUC值均远高于中低年资医师, 略低于高年资医师。结论: 联合模型在尘肺病早期DR胸片筛查中展现出良好的诊断效能与临床净获益, 整体性能优于单一模型, 可作为职业健康首筛的核心工具。

关键词

壹期尘肺病, 胸部X线片, 放射组学, 深度学习, 联合模型

*通讯作者。

A Study on Multi-Model Fusion Based on DR Images for the Screening of Stage I Pneumoconiosis

Fang Fang^{1,2}, Tianzi Jiang³, Jun Zhang³, Zhenbo Sun³, Huikun Zhang³, Feng Duan^{1,3*}

¹Qingdao Medical College of Qingdao University, Qingdao Shandong

²Department of Radiology, Zibo First Hospital, Zibo Shandong

³Department of Radiology, The Affiliated Hospital of Qingdao University, Qingdao Shandong

Received: March 21, 2026; accepted: April 16, 2026; published: April 21, 2026

Abstract

Objective: Based on chest X-ray images, we constructed radiomics models and 2D deep learning models, respectively. We then performed decision-level fusion of the independent predictive scores from these two models using multivariate logistic regression to develop a combined model for the early screening of pneumoconiosis. We evaluated the clinical value of this model to provide a basis for model selection and clinical translation. **Materials and Methods:** We retrospectively collected chest X-ray images and clinical data from 193 patients with Stage I pneumoconiosis and 178 healthy individuals. Using ITK-SNAP software, we manually delineated regions of interest (ROIs) in both lung fields on the chest X-ray images, extracted high-throughput radiomic features, and constructed a radiomic model. Concurrently, we built a 2D deep learning model using the delineated ROIs as input. We performed decision-level fusion of the independent predictive risk scores from the two standalone models using multivariate logistic regression to construct a combined predictive model. Using the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, and specificity as core metrics, we validated model performance through stratified validation using a training set (70%) and an independent test set (30%). Calibration curves and Decision Curve Analysis (DCA) were used to evaluate model calibration performance and clinical net benefit, and the DeLong test was employed to compare the discriminatory capabilities of the models. Three radiologists, each qualified to diagnose occupational pneumoconiosis (one each with high, medium, and low years of experience), were recruited to perform independent film readings as a manual control, using a double-blind design. **Results:** In the independent test set, the accuracy rates of the radiomics model, the 2D deep learning model, and the combined model were 72.1%, 66.7%, and 75.7%, respectively, with AUC values of 0.800 (95% CI: 0.718~0.882), 0.803 (95% CI: 0.718~0.888), and 0.813 (95% CI: 0.731~0.894), respectively. The combined model had a sensitivity of 75.4% and a specificity of 75.9%, demonstrating significantly superior diagnostic performance compared to each individual model. Calibration curve results showed good agreement between the combined model's predicted probability and the actual probability of disease progression. Decision curve analysis (DCA) results indicated that, across most threshold probability ranges, the net benefit of the combined model was higher than that of any single model. The DeLong test confirmed that, when comparing AUC values between the radiomics model, the 2D deep learning model, and the combined model in pairs, the differences were not statistically significant. In a human-machine comparison, the accuracy and AUC values of the combined model were significantly higher than those of junior and mid-career physicians, and slightly lower than those of senior physicians. **Conclusion:** The combined model demonstrates good diagnostic performance and clinical net benefit in the early screening of pneumoconiosis using DR chest radiographs. Its overall performance is superior to that of single models, and it can serve as a core tool for initial screening in occupational health.

Keywords

Stage I Pneumoconiosis, Chest X-Ray, Radiomics, Deep Learning, Combined Model

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

尘肺病是我国职业病中最为常见且危害最为严重的疾病，其发病率居高不下，并且呈现出年轻化的趋势。目前，根据现行标准 GBZ 70-2015《职业性尘肺病的诊断》[1]，尘肺病诊断主要依据高千伏或数字化 X 线摄影(DR)后前位胸片表现，辅以粉尘暴露史和临床实验室检查进行诊断及分期。然而，胸片存在解剖结构重叠，分辨率有限等问题，对早期尘肺病细微病变识别能力不足。同时，传统人工阅片方法主观性强，耗时且效率低下，诊断的稳定性和准确性难以保证。随着人工智能技术的迅猛发展，为医生提供了更为高效和准确的辅助工具。放射组学可通过提取影像中隐藏的高维特征实现病变量化分析，深度学习则擅长自动挖掘图像深层信息，两者融合有望突破单一技术的局限性。本研究旨在系统探讨放射组学模型、2D 深度学习模型及二者决策级融合构建的联合模型在壹期尘肺病筛查中的诊断效能，并以人工诊断为参照，明确联合模型的临床应用价值，为尘肺病早期筛查的模型选择与临床转化提供依据。

2. 资料与方法

2.1. 一般资料

回顾性收集 2018 年 1 月至 2025 年 4 月在淄博市第一医院(淄博市职业病防治院) 371 例患者的胸部 X 线片，其中壹期尘肺病患者(实验组) 193 例，健康体检者(对照组) 178 例。按照 7:3 的比例将尘肺组及正常组随机分为训练集组 260 例及验证集组 111 例。

纳入标准：① 尘肺病组研究对象均经诊断专家组根据《职业性尘肺病的诊断》(GBZ 70-2015)明确诊断。② 对照组研究对象为同期健康体检人员，经临床及影像学检查排除肺部疾病、粉尘暴露史及其他可能影响肺部影像的全身性疾病。③ 行后前位胸部 X 线正位片检查且相关临床资料完整。

排除标准：① 图像质量为三级片[1]，无法用于尘肺病初诊。② 尘肺合并肺结核、肺癌、肺炎等。

2.2. 检查方法

DR 胸片检查采用 GE 公司 Definium6000 DR 设备，拍摄后前立位胸片，投照电压 120 kV，电流 320 mA，采用 AEC 自动控制曝光时间。所有 DR 胸片符合《职业性尘肺病的诊断》(GBZ70-2015)中附录 C 和 F 的质量要求。

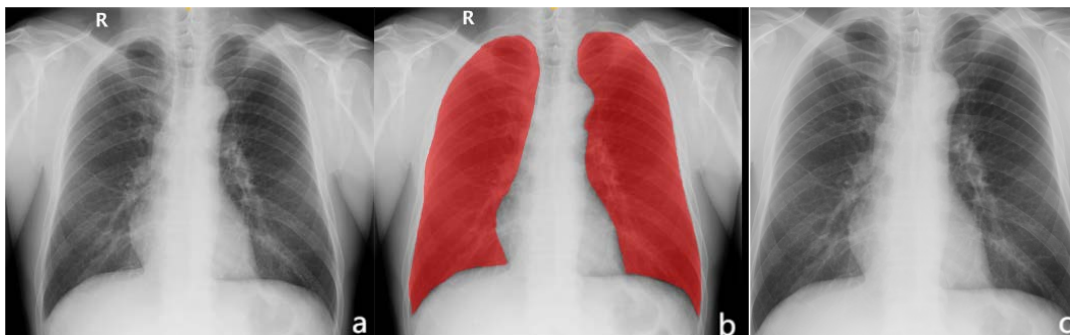
2.3. 研究方法

2.3.1. 感兴趣区域(ROI)分割与图像预处理

将训练集与测试集所有胸部 X 线片图像导入 ITK-SNAP 软件进行手动分割，由两名具有 10 年以上诊断经验的放射科医师在胸片图像上手动勾画肺部区域，剔除心脏、大血管及胸壁软组织等区域，所勾画的图像保存为 ROI(见图 1)。差异部分由两名医师通过共识解决，以确保准确性和一致性。所有图像均

调整为统一空间分辨率(1 mm × 1 mm), 并应用像素强度归一化处理, 保持灰度分布可比性。

随机纳入 30 例患者, 分别采用组内相关系数(interclass correlation coefficient, ICC)评价同一勾画者前后两次勾画及不同勾画者间勾画结果的一致性, ICC > 0.75 提示一致性良好。



注: a 为壹期尘肺原始图像; b 为对图像进行分割; c 为对图像进行裁剪。

Figure 1. Schematic diagram of ROI delineation

图 1. ROI 勾画示意图

2.3.2. 放射组学模型构建

根据影像生物标志物标准化倡议(IBSI)指南, 使用 PyRadiomics (v3.0.1)软件进行放射组学特征提取。特征选择为确保特征稳定性并防止过拟合, 采用分层选择流程。首先通过学生 t 检验、曼-惠特尼 U 检验或 χ^2 检验筛选出与尘肺病分类显著相关的特征($P < 0.05$), 然后将皮尔逊相关系数 $r > 0.9$ 的特征剔除, 随后应用最大相关性最小冗余(mRMR)方法, 最终确定 32 个信息量最大且相关性最低的特征。最后采用最小绝对收缩与选择算子(LASSO)回归结合 10 折交叉验证进行进一步降维后, 最终得到 17 个简洁且具有区分度的特征子集(图 2、图 3)。采用随机森林(RandomForest)算法构建放射组学模型, 并使用网格搜索法联合 5 折交叉验证完成模型超参数优化。

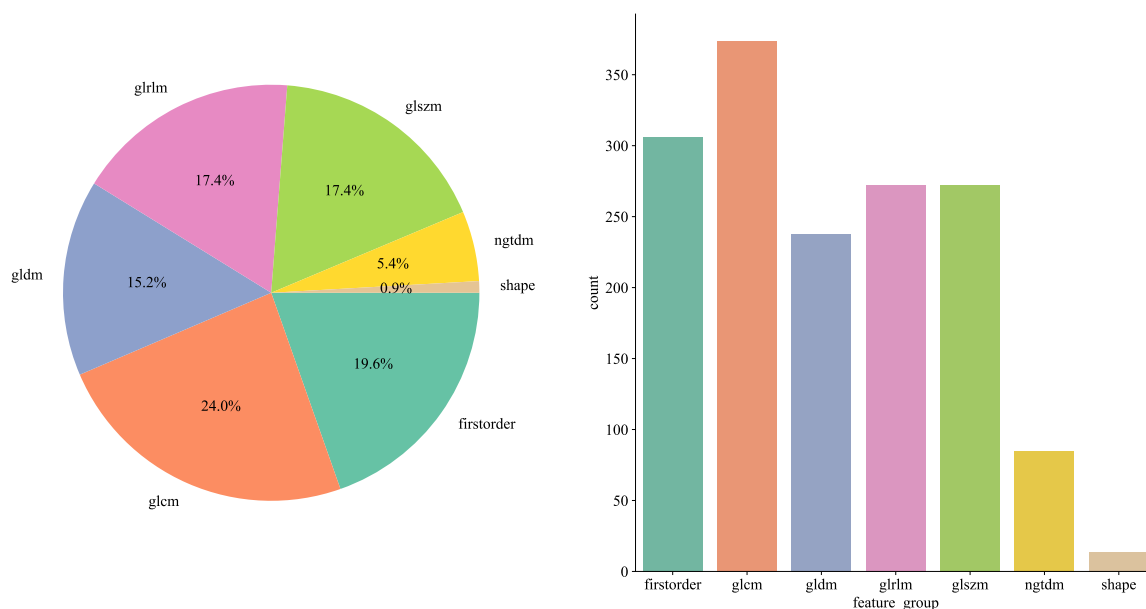


Figure 2. Statistical chart showing the distribution of manually extracted features in the radiomics group

图 2. 放射组学手工提取特征的分布统计图表

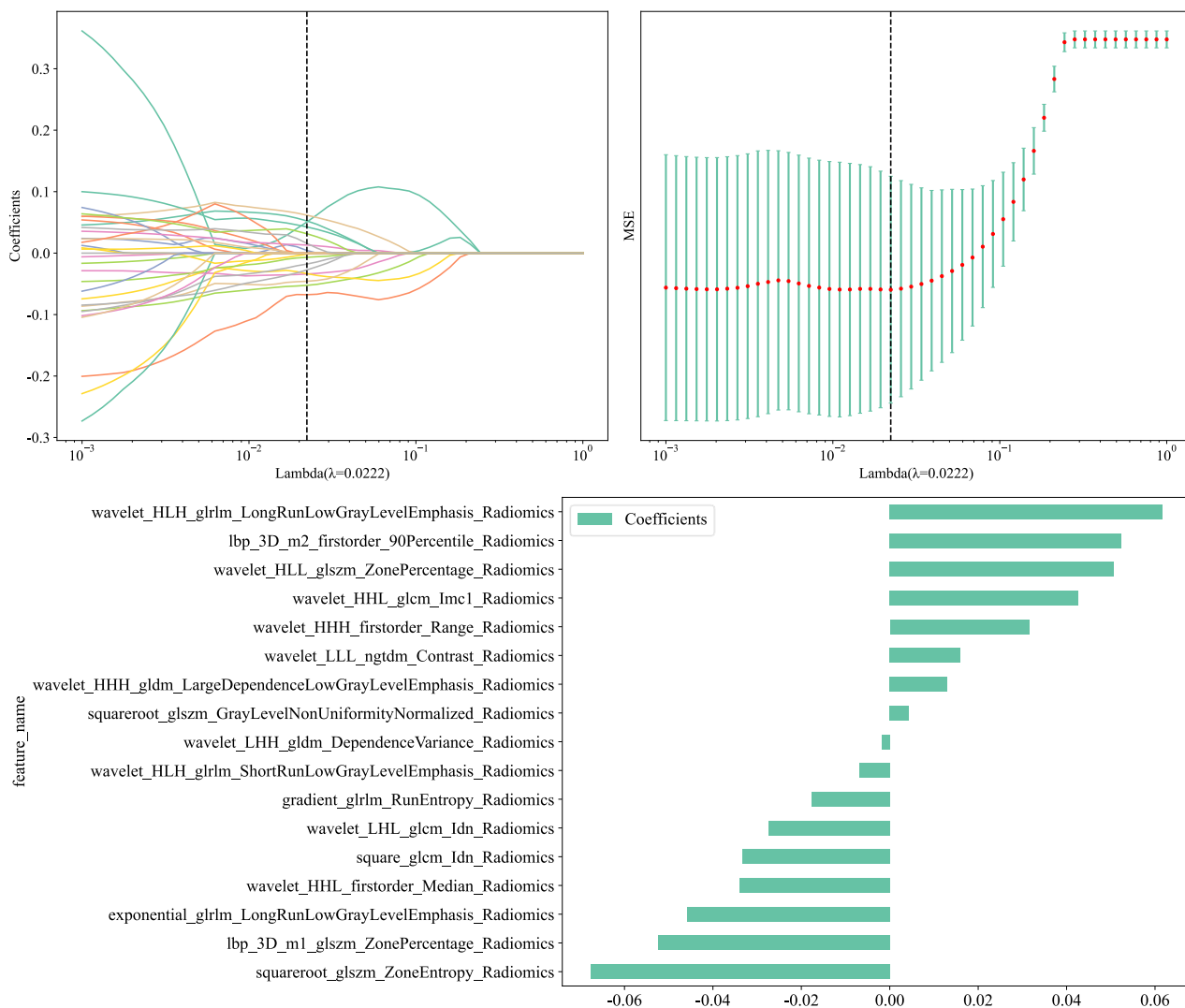


Figure 3. Distribution of core feature coefficients following parameter tuning and feature selection via LASSO regression of radiomics features

图 3. 放射组学特征的 LASSO 回归参数调优及筛选后核心特征系数分布

2.3.3. 深度学习模型构建

基于勾画感兴趣区域(ROI)裁剪最小外接矩形区域, 采用最近邻插值法将所有图像统一调整为 224 × 224 像素, 匹配神经网络输入规格。数据归一化方面, 对图像切片灰度值实施最小 - 最大归一化处理, 将像素值缩放至[-1, 1]区间, 同时对 RGB 通道强度分布采用 Z 分数标准化, 实现样本数据分布均一化, 提升模型收敛效率与特征学习稳定性。数据增强策略仅应用于训练集, 通过随机裁剪、水平翻转、垂直翻转的实时增强技术, 在保持图像结构完整性的前提下增加数据集方差, 提升模型的鲁棒性与泛化能力; 训练集及测试集仅执行归一化处理, 未进行数据增强操作, 确保模型评估框架的一致性与结果的可靠性。

本研究采用 DenseNet121 经典卷积神经网络(CNN)架构, 基于 ImageNet 数据集预训练权重开展迁移学习, 针对尘肺病与非尘肺病的二分类任务进行微调。通过迁移学习弥补医学影像样本量有限的问题, 实现模型高效稳定收敛。训练过程中采用余弦衰减策略调整学习率, 数学表达式如下:

$$\eta_i = \eta_{\min_i} + \frac{1}{2}(\eta_{\max_i} - \eta_{\min_i}) \left[1 + \cos\left(\frac{T_{cur}}{T_i} \pi\right) \right] \quad (1)$$

其中, η_{\min_i} 设为 0, η_{\max_i} 设为 0.01, T_i 为总训练迭代次数(设为 48), T_{cur} 为当前迭代次数。

同时辅以提前停止策略,有效规避模型过拟合风险,保障训练稳定性。优化器选用随机梯度下降(SGD),可有效优化网络参数更新过程,保障单模型训练梯度稳定、收敛效果可靠。

2.3.4. 联合模型构建

采用多因素 Logistic 回归构建融合模型,将放射组学模型与 2D 深度学习模型输出的独立预测评分纳入模型,开展决策级融合,最终构建尘肺病早期筛查联合预测模型。

2.3.5. 人工对照

招募 3 名均具备职业性尘肺病诊断资质的放射科医师(高年资: 20 年以上经验, doctor1; 中年资: 10~20 年经验, doctor2; 低年资: 5~10 年经验, doctor3), 采用双盲设计分别独立对入组病例进行读片诊断, 将人工诊断结果与联合模型结果进行对比分析。

2.4. 统计学分析

采用 Shapiro-Wilk 检验对连续临床变量进行正态性检验。正态分布数据采用 t 检验进行比较, 非正态分布变量用 Mann-Whitney U 检验分析。分类变量采用卡方检验(χ^2)进行比较。所有统计分析与建模均在 Python (v3.7.12)环境下的 OnekeyAI 平台(v3.9.1)完成。放射组学特征通过 PyRadiomics (v3.0.1)提取, 深度学习模型基于 PyTorch1.11.0 实现, 并采用 CUDA11.3.1 和 cuDNN8.2.1 加速。统计计算使用 Statsmodels0.13.2 完成。

通过绘制受试者工作特征(ROC)曲线,计算曲线下面积(AUC)及 95% 置信区间(95% CI),结合灵敏度、特异度等指标,评价各模型的诊断效能;采用校准曲线、决策曲线分析量化模型的校准可靠性及临床净获益。采用 DeLong 检验比较各模型间 AUC 值的统计学差异。

3. 结果

3.1. 一般资料比较

训练集和测试集一般资料比较显示性别及年龄均无统计学意义($P > 0.05$)。提示两组基线资料均衡,具有可比性,见表 1。

Table 1. Statistical summary of demographic characteristics of the study sample for pneumoconiosis

表 1. 尘肺病研究样本人口学特征统计表

特征名称	总样本(n = 371)	训练集(n = 260)	测试集(n = 111)	P 值
年龄(岁)	52.90 ± 6.85	53.37 ± 7.15	51.78 ± 5.96	0.076
性别				1.0
女性	28 (7.55)	20 (7.69)	8 (7.21)	—
男性	343 (92.45)	240 (92.31)	103 (92.79)	—

3.2. 放射组学模型、2D 深度学习模型及联合模型的效能评估

3.2.1. 基础效能评估

本研究构建并比较了放射组学模型(Radiomics)、2D 深度学习模型(DL2D)及联合模型(Combined)在早期尘肺病筛查中的效能,训练集与独立测试集的评估结果详见表 2 及图 4。在训练集中,联合模型的整体

性能最优, 其 AUC 为 0.886 (95% CI: 0.846~0.927), 高于放射组学模型(AUC = 0.873, 95% CI: 0.831~0.915) 与 2D 深度学习模型(AUC = 0.828, 95% CI: 0.778~0.878)。联合模型准确率为 81.9%, 灵敏度与特异度分别为 81.0%、82.9%, (PPV)与阴性预测值(NPV)分别为 84.1%、79.7%, 各项指标均优于单一模型。相比之下, 放射组学模型特异度高(92.7%), 但灵敏度较低(65.7%), 而 2D 深度学习模型灵敏度高(85.4%), 但特异度低(69.1%)。在独立测试集中, 联合模型仍具有最优区分能力, AUC 为 0.813 (95% CI: 0.731~0.894), 高于放射组学模型(AUC = 0.800, 95% CI: 0.718~0.882)与 2D 深度学习模型(AUC = 0.803, 95% CI: 0.718~0.888); 联合模型的准确率为 75.7%, 灵敏度、特异度分别为 75.4%、75.9%, PPV 与 NPV 分别为 76.8%、74.5%, 有效弥补了单一模型的短板, 实现了灵敏度与特异度的均衡提升。

Table 2. Detailed performance metrics of the radiomics model, 2D deep learning model, ensemble model, and manual ground-truth model on the training and test sets

表 2. 放射组学模型、2D 深度学习模型、联合模型及人工对照模型在训练集与测试集的详细性能指标

模型名称	性能						队列
	AUC (95% CI)	准确率(%)	灵敏度(%)	特异度(%)	PPV (%)	NPV (%)	
Radiomics	0.873 (0.831~0.915)	78.5	65.7	92.7	90.9	70.8	训练集
DL2D	0.828 (0.778~0.878)	77.7	85.4	69.1	75.5	81.0	
Combined	0.886 (0.846~0.927)	81.9	81.0	82.9	84.1	79.7	
Doctor1	0.893 (0.857~0.929)	88.8	81.0	97.6	97.4	82.2	
Doctor2	0.751 (0.699~0.802)	74.6	66.4	83.7	82.0	69.1	
Doctor3	0.731 (0.681~0.782)	72.3	57.7	88.6	84.9	65.3	
Radiomics	0.800 (0.718~0.882)	72.1	64.9	79.6	77.1	68.3	测试集
DL2D	0.803 (0.718~0.888)	66.7	80.7	51.9	63.9	71.8	
Combined	0.813 (0.731~0.894)	75.7	75.4	75.9	76.8	74.5	
Doctor1	0.825 (0.762~0.887)	82.0	64.9	100	100	73.0	
Doctor2	0.700 (0.628~0.772)	69.4	45.6	94.4	89.7	62.2	
Doctor3	0.689 (0.606~0.771)	68.5	54.4	83.3	77.5	63.4	

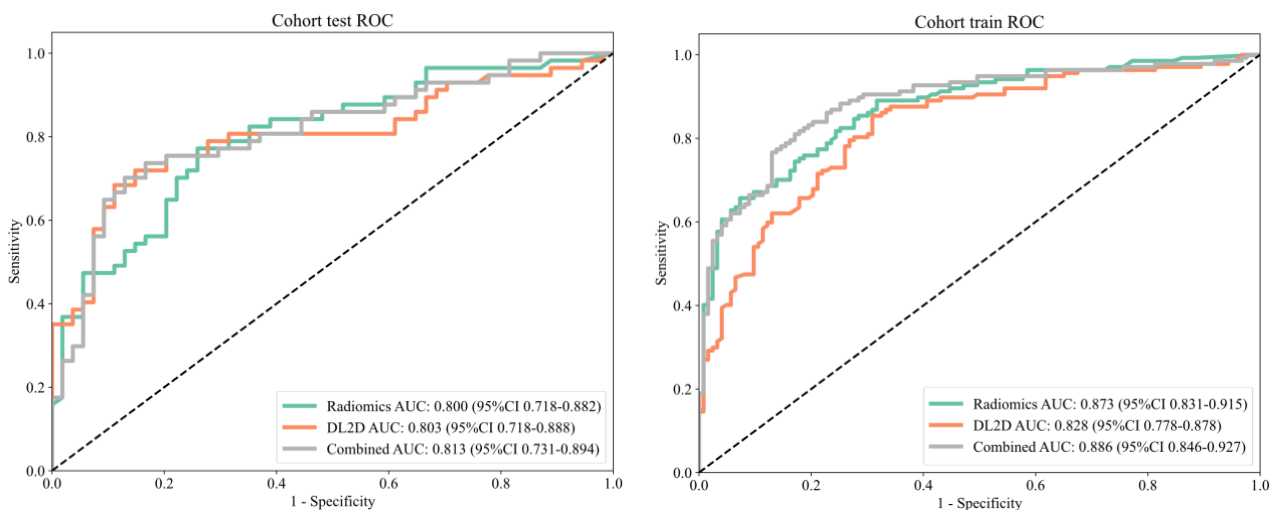


Figure 4. Receiver operating characteristic curves for the training and validation sets of the radiomics model, 2D deep learning model, combined model, and manual reference model

图 4. 放射组学模型、2D 深度学习模型、联合模型及人工对照模型训练集及验证集受试者工作特征曲线

3.2.2. 校准曲线性能分析

校准曲线是模型效能评估的重要补充，三种模型的校准曲线表现分析如图 5 所示。训练集中，放射组学、2D 深度学习模型及联合模型校准曲线均贴近理想校准线，其中联合模型贴合度最优，表明其预测概率与实际观测概率的一致性最佳。独立测试集中，各模型校准度均有下降，但联合模型仍最贴近理想校准线，整体表现优于单一模型。

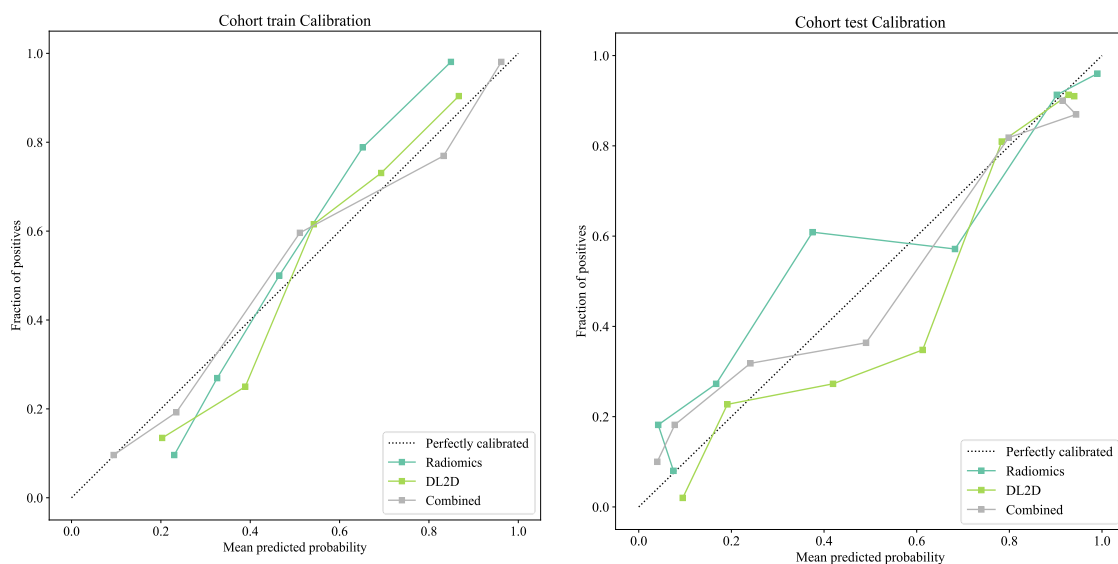


Figure 5. Calibration curves for the radiomics model, 2D deep learning model, and combined model

图 5. 放射组学模型、2D 深度学习模型、联合模型的校准曲线图

3.2.3. 临床净获益分析

决策曲线分析(DCA)结果显示(图 6)，放射组学模型、2D 深度学习模型及联合模型均具备尘肺病早期筛查的临床应用价值，其中联合模型在训练集与测试集中均展现出最优的临床净获益，临床应用潜力显著优于放射组学与 2D 深度学习模型。

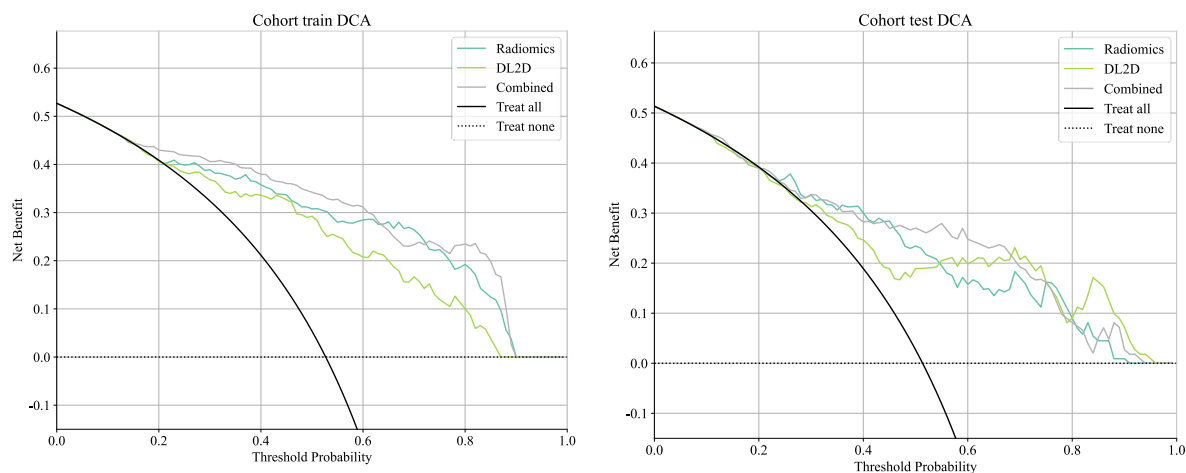


Figure 6. Clinical decision curves for the radiomics model, the 2D deep learning model, and the combined model

图 6. 放射组学模型、2D 深度学习模型及联合模型临床决策曲线

3.2.4. DeLong 检验结果

采用 DeLong 检验对放射组学模型、2D 深度学习模型及联合模型的 AUC 进行比较, 结果如图 7 所示。在训练集中, 联合模型与 2D 深度学习模型 AUC 差异具有显著统计学意义($P < 0.001$), 放射组学模型与 2D 深度学习模型 AUC 差异具有统计学意义($P = 0.044$); 放射组学模型与联合模型 AUC 差异无统计学意义($P = 0.110$)。在独立测试集中, 放射组学模型、2D 深度学习模型与联合模型三者间的 AUC 两两比较, 差异均无统计学意义。

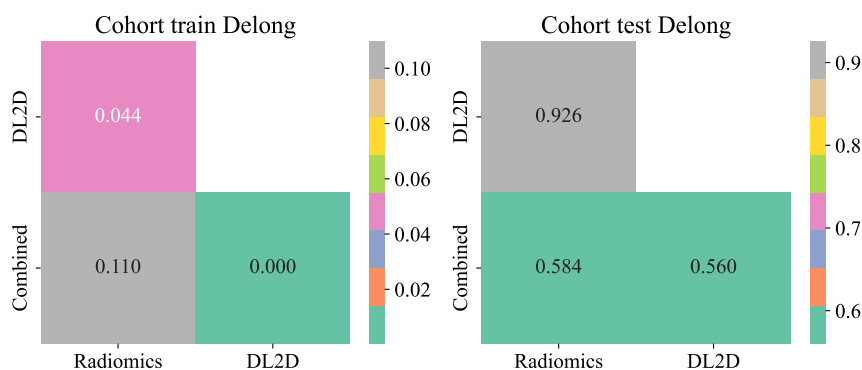


Figure 7. Results of the DeLong test for differences in AUC between the radiomics model, the 2D deep learning model, and the combined model

图 7. 放射组学模型、2D 深度学习模型、联合模型 AUC 差异的 DeLong 检验结果

3.3. 人机对比

在训练集中, 联合模型及高、中、低三位医师的准确率分别为 81.9%、88.8%、74.6%、72.3%, AUC 值分别为 0.886 (95% CI: 0.846~0.927)、0.893 (95% CI: 0.857~0.929)、0.751 (95% CI: 0.699~0.802)、0.731 (95% CI: 0.681~0.782)。在测试集中, 联合模型及高、中、低三位医师的准确率分别为 75.7%、82.0%、69.4%、68.5%, AUC 值分别为 0.813 (95% CI: 0.731~0.894)、0.825 (95% CI: 0.762~0.887)、0.700 (95% CI: 0.628~0.772)、0.689 (95% CI: 0.606~0.771)。因此, 在训练集及测试集中, 联合模型的准确率及 AUC 值均远高于中低年资医师, 略低于高年资医师, 灵敏度及特异度达到良好平衡(表 2)。

4. 讨论

尘肺病是以肺组织弥漫性纤维化为主要特征的一类职业性肺部疾病的统称[1]。目前尘肺病诊断主要依据高千伏或数字化 X 线摄影(DR)后前位胸片中小阴影的整体密集程度、肺区分布范围以及是否存在小阴影聚集、大阴影或胸膜斑等征象, 进行诊断与分期。因胸片空间分辨率有限, 且人工阅片主观性强等问题, 限制了早期尘肺病的筛查效能[2]-[4]。本研究以胸部 DR 片为基础, 选取壹期尘肺病患者及健康体检者为研究对象, 分别构建放射组学模型、2D 深度学习模型及二者决策级融合构建的联合模型, 系统评估各模型在壹期尘肺病筛查中的诊断效能, 结果表明联合模型在壹期尘肺病早期筛查中展现出良好的诊断效能、校准度与临床净获益, 整体性能优于单一模型, 且诊断水平接近高年资放射科医师, 可为基层职业健康首筛提供可靠的辅助工具。

本研究结果显示, 单一放射组学模型与 2D 深度学习模型在性能上呈现明显的互补特征。在独立测试集中, 放射组学模型特异度为 79.6%, 但灵敏度仅为 64.9%; 2D 深度学习模型灵敏度为 80.7%, 但特异度仅为 51.9%。造成这一结果的根本原因在于两类模型在特征提取逻辑存在本质差异。放射组学通过人工定义的方式提取高通量的纹理、形态、灰度特征, 可精准捕捉尘肺病特征性的肺间质纹理改变与结节

状小阴影,对符合诊断标准的明确影像征象识别特异度较高[5];但对于壹期尘肺病极细微、密度差异极小的早期病灶,手工特征的捕捉能力有限,导致模型灵敏度不足。2D深度学习模型通过卷积神经网络自动学习影像的隐层深度特征,对肺部整体密度的细微变化更为敏感,能够识别人工阅片容易遗漏的早期病灶,因此具备更高的灵敏度[6]-[8];但这类模型易受肋骨、锁骨等肺外组织及正常肺纹理的干扰,容易将正常解剖结构误判为阳性病灶,导致假阳性率升高、特异度显著下降。两类模型特征提取逻辑的差异,构成了二者性能互补的基础。

通过多因素 Logistic 回归对放射组学模型与 2D 深度学习模型的预测评分进行融合,构建诊断性能更强、稳定性更高且具备有一定生物学可解释性的联合模型[9]-[12]。这一优势已在多项研究中展现出优势[13]-[16]。本研究系统评估了放射组学模型、2D 深度学习模型及联合模型在壹期尘肺病 DR 胸片筛查中的诊断效能。结果显示,在训练集和测试集中,联合模型的 AUC 值和准确率均优于单一模型,灵敏度和特异度达到良好平衡,未出现单一模型中常见的灵敏度与特异度不均衡问题。校准曲线结果显示,联合模型的预测概率与实际患病概率具有良好的一致性,说明模型的预测结果稳定可靠,未出现明显的过度预测或预测不足问题。决策曲线分析进一步证实,在大多数临床阈值概率范围内,联合模型的临床净获益均高于单一模型,从临床决策角度为模型的潜在应用价值提供了参考依据。

本研究在系统评估模型效能时,除传统诊断效能指标与临床实用性指标外,还增设了人机对比实验。结果显示,联合模型的准确率及 AUC 值均明显高于中低年资医师,略低于高年资医师。这一发现与 Zhang 等[17]在 Legg-Calvé-Perthes 病影像学辅助诊断中的研究结果相一致,其构建的放射组学与深度学习联合模型在外部验证中的 AUC 达 0.878,显著优于低年资医师,与高年资医师诊断水平相当。这一结果提示,联合模型在不同疾病的影像学辅助诊断场景中均能弥补中低年资医师的经验差距,且其性能已接近高年资医师,不仅为模型诊断效能提供了直观的临床参照标准,也有助于提升研究结论的临床转化价值。

DeLong 检验结果显示,放射组学模型、2D 深度学习模型及其联合模型的 AUC 两两比较无统计学差异,但联合模型的 AUC 为最优,分析其原因主要有两点:一是本研究为单中心研究,测试集样本量有限,统计学检验效能不足,难以检出组间的细微性能差异;二是本研究聚焦的壹期尘肺病病灶细微,诊断难度大,单一模型本身已具备一定的区分效能,组间的性能差异未被充分放大,后续通过大样本、多中心研究可进一步验证联合模型的性能优势。

本研究仍存在一定局限性:其一,样本量有限且单中心来源,导致模型出现潜在过拟合风险,泛化能力受限。其二,研究队列仅纳入健康人群作为对照,未纳入长期粉尘暴露的影像学正常人群及其他易与尘肺病混淆的肺部疾病患者,易产生谱系偏倚,同时本研究构建的 2D 深度学习模型存在特异度偏低、诊断性能不均衡的问题;其三,仅基于影像特征构建模型,未纳入职业粉尘暴露史等尘肺病诊断核心临床信息,未能完全贴合临床综合诊断逻辑。针对上述局限,后续研究将联合多地区医疗机构、不同病程阶段及不同成像设备的壹期尘肺病 DR 影像数据,构建大样本、多中心的研究队列,以增强模型泛化能力[18][19]。完善队列设置以降低偏倚,采用代价敏感学习、Focal Loss 等策略优化模型性能均衡性,同时纳入关键临床信息构建多模态融合诊断模型,进一步提升模型的诊断效能与临床适配性。

综上,本研究构建的联合模型在尘肺病早期 DR 胸片筛查中展现出良好的诊断效能与临床净获益,整体性能优于单一模型。未来研究需通过构建多中心大数据集、整合临床多模态信息,并优化模型的鲁棒性与泛化能力,最终推动尘肺病早期筛查智能模型的规模化临床应用,为职业健康防护体系提供技术支撑。

声 明

本研究经淄博市第一医院(淄博市职业病防治院)医学伦理委员会批准(伦理批号:YXLL2025060542),知情同意已豁免。

参考文献

- [1] 职业性尘肺病的诊断: GBZ (卫生) 70-2015 [S]. 北京: 中国标准出版社, 2015.
- [2] 张镗琢, 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.
- [3] 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.
- [4] 李娜, 尹扬光, 王世松, 等. 2016-2020 年某职业病防治院尘肺病诊断情况的调查分析[J]. 重庆医学, 2023, 52(16): 2523-2527.
- [5] 王威, 谢武桃, 张洪, 等. CT 影像组学联合多种机器学习模型鉴别尘肺大阴影与周围型肺癌的价值比较[J]. 重庆医学, 2026, 55(2): 255-260, 266.
- [6] Xiao, L.I., Liu, C.F., Li, G., *et al.* (2021) Deep Learning in Chest Radiography: Detection of Pneumoconiosis. *Biomedical and Environmental Sciences*, **34**, 842-845.
- [7] El-Regaily, S.A., Salem, M.A.M., Abdel Aziz, M.H. and Roushdy, M.I. (2020) Multi-View Convolutional Neural Network for Lung Nodule False Positive Reduction. *Expert Systems with Applications*, **162**, Article ID: 113017. <https://doi.org/10.1016/j.eswa.2019.113017>
- [8] Suresh, S. and Mohan, S. (2020) Roi-Based Feature Learning for Efficient True Positive Prediction Using Convolutional Neural Network for Lung Cancer Diagnosis. *Neural Computing and Applications*, **32**, 15989-16009. <https://doi.org/10.1007/s00521-020-04787-w>
- [9] Li, J., Lv, D., Guo, Z., Zhou, H., Yao, X., Rong, Y., *et al.* (2025) Integration of Multi-Scale Radiomics and Deep Learning for Ki-67 Prediction in Clear Cell Renal Carcinoma. *NPJ Precision Oncology*, **10**, Article No. 12. <https://doi.org/10.1038/s41698-025-01214-y>
- [10] Sun, Q., Yu, L., Song, Z., Wang, C., Li, W., Chen, W., *et al.* (2025) Deep Learning and Radiomics Fusion for Predicting the Invasiveness of Lung Adenocarcinoma within Ground Glass Nodules. *Scientific Reports*, **15**, Article No. 29285. <https://doi.org/10.1038/s41598-025-13447-9>
- [11] Zhao, J., Wang, T., Wang, B., Satishkumar, B.M., Ding, L., Sun, X., *et al.* (2025) Deep Learning Radiomics Fusion Model to Predict Visceral Pleural Invasion of Clinical Stage IA Lung Adenocarcinoma: A Multicenter Study. *Journal of Cardiothoracic Surgery*, **20**, Article No. 246. <https://doi.org/10.1186/s13019-025-03488-6>
- [12] Wang, J., Song, M., Fan, D., Wang, X., Zhang, S., Yang, J., *et al.* (2025) Radiologist-Inspired Symmetric Local-Global Multi-Supervised Learning for Early Diagnosis of Pneumoconiosis. *Expert Systems with Applications*, **276**, Article ID: 127173. <https://doi.org/10.1016/j.eswa.2025.127173>
- [13] Zhang, B., Zhu, J., Xu, R., Zou, L., Lian, Y., Xie, X., *et al.* (2025) A Combined Model Integrating Radiomics and Deep Learning Based on Multiparametric Magnetic Resonance Imaging for Classification of Brain Metastases. *Acta Radiologica*, **66**, 24-34. <https://doi.org/10.1177/02841851241292528>
- [14] Niu, W., Yan, J., Hao, M., Zhang, Y., Li, T., Liu, C., *et al.* (2025) MRI Transformer Deep Learning and Radiomics for Predicting IDH Wild Type TERT Promoter Mutant Gliomas. *NPJ Precision Oncology*, **9**, Article No. 89. <https://doi.org/10.1038/s41698-025-00884-y>
- [15] Gan, Y., Hu, Q., Shen, Q., Lin, P., Qian, Q., Zhuo, M., *et al.* (2024) Comparison of Intratumoral and Peritumoral Deep Learning, Radiomics, and Fusion Models for Predicting KRAS Gene Mutations in Rectal Cancer Based on Endorectal Ultrasound Imaging. *Annals of Surgical Oncology*, **32**, 3019-3030. <https://doi.org/10.1245/s10434-024-16697-5>
- [16] Zhang, Z., Zhang, D., Yang, Y., Liu, Y. and Zhang, J. (2024) Value of Radiomics and Deep Learning Feature Fusion Models Based on DCE-MRI in Distinguishing Sinonasal Squamous Cell Carcinoma from Lymphoma. *Frontiers in Oncology*, **14**, Article ID: 1489973. <https://doi.org/10.3389/fonc.2024.1489973>
- [17] Zhang, D., Li, Y., Li, C. and Guo, W. (2025) Radiomics and Deep Learning Model Based on X-Ray Imaging for the Assisted Diagnosis of Early Legg-Calvé-Perthes Disease. *BMC Musculoskeletal Disorders*, **26**, Article No. 910. <https://doi.org/10.1186/s12891-025-09189-4>
- [18] 汪伟. 基于深度卷积神经网络的尘肺病 DR 影像诊断研究[D]: [硕士学位论文]. 秦皇岛: 华北理工大学, 2021.
- [19] Xie, S.M. (2024) Foundation Models from a Data-Distributional View. Ph.D. Thesis, Stanford University.