

基于 CT 临床放射组学列线图与深度学习鉴别非典型肺错构瘤和肺腺癌

王传彬^{1,2}, 李翠平¹, 曹 锋², 郜言坤¹, 钱宝鑫³, 董江宁², 吴兴旺¹

摘要 目的 探讨临床放射组学列线图(CRN)和深度学习卷积神经网络(DCNN)对非典型肺错构瘤(APH)和非典型肺腺癌(ALA)的鉴别诊断价值。方法 从两家医疗机构回顾性收集307例患者。机构1的患者按照7:3的比例随机分为训练集($n=184$;APH=97,ALA=87)和内部验证集($n=79$;APH=41,ALA=38),机构2的患者作为外部验证集($n=44$;APH=23,ALA=21)。分别建立CRN模型和DCNN模型,并采用德隆检验和受试者工作特性曲线(ROC)对两种模型的性能进行比较。通过人-机竞赛评估人工智能(AI)在肺结节Lung-RADS分类中的价值。结果 DCNN模型在训练集和内、外部验证集中的曲线下面积(AUC)均高于CRN模型(0.983 vs 0.968, 0.973 vs 0.953, 0.942 vs 0.932),但差异无统计学意义($P=0.23, 0.31, 0.34$)。在放射科医师-AI竞争实验中,AI倾向于下调APH组中更多的Lung-RADS类别,并肯定ALA中更多的Lung-RADS类别。结论 DCNN及CRN在区分APH和ALA方面具有较高价值,前者表现更优;AI在评价肺结节的Lung-RADS分类方面优于放射科医师。

关键词 错构瘤;肺腺癌;列线图;深度学习;人工智能;计算机断层扫描

中图分类号 R 734.2;R 447

文献标志码 A **文章编号** 1000-1492(2024)02-0344-07
doi:10.19405/j.cnki.issn1000-1492.2024.02.026

肺错构瘤(pulmonary hamartoma, PH)和肺腺癌(lung adenocarcinoma, LA)分别是肺部最常见的良性和恶性肿瘤,二者的临床处理方式不同,前者以随访为主,后者则主张手术切除^[1-2]。高分辨率CT(high resolution CT, HRCT)是术前诊断PH和LA

的最常用方法^[3]，“爆米花”样钙化或脂肪密度是PH的典型征象,而LA则表现出“毛刺征”、“空泡征”等典型征象。当二者缺乏上述征象,分别表现为非典型肺错构瘤(atypical pulmonary hamartoma, APH)和非典型肺腺癌(atypical lung adenocarcinoma, ALA)时,极易混淆。放射组学和深度学习卷积神经网络(deep convolutional neural network, DCNN)通过将具有生物学信息的医学图像转化为可进一步分析的高维数据提高非侵入性诊断能力,目前已被广泛应用于多种临床实践中^[4-7],该文拟建立基于临床-影像学特征和放射组学特征的临床放射组学列线图(clinical radiomic nomogram, CRN)及DCNN的两种预测模型,提高APH和ALA的鉴别诊断能力,探讨基于DCNN的人工智能(artificial intelligence, AI)评价肺结节Lung-RADS分类水平的能力。

1 材料与方法

1.1 病例资料 回顾性收集2012年10月至2022年10月来自安徽医科大学第一附属医院和中国科学技术大学附属第一医院的经病理确诊的PH和LA患者共2283例,根据CT征象,将APH定义为CT图像上缺乏可见钙化和可测量脂肪($CT值 \leq -20 HU$)的PH,将ALA定义为缺乏空洞或空泡征、缺乏毛刺征、胸膜凹陷征、血管束束征等恶性征象的LA。其他纳入标准为:①直径为 $\geq 1 cm$ 和 $\leq 3 cm$ 的实性结节(文献^[8]认为 $< 1.0 cm$ 的病变通常无法获得有效的放射组学特征, $> 3.0 cm$ 的结节通常具有典型的CT特征);②具有标准算法重建出的薄层胸部CT图像(层厚/间距为1.25 mm);③图像质量高,不影响肺结节的轮廓勾画和分割。最终纳入患者共307例,其中机构1的患者($n=263$)按照7:3的比例随机分为训练集($n=184$;APH=97,ALA=87)和内部验证集($n=79$;APH=41,ALA=38),机构2的患者作为外部验证集($n=44$;APH=23,ALA=21)。实验设计流程见图1。

1.2 图像采集及特征提取 两家机构的扫描设备

2023-12-05 接收

基金项目:安徽省自然科学基金面上项目(编号:2308085MH241)

作者单位:¹安徽医科大学第一附属医院放射科,合肥 230022

²中国科学技术大学附属第一医院(安徽省立医院)影像科,合肥 230031

³北京汇医慧影医疗科技有限公司,北京 100192

作者简介:王传彬,男,硕士研究生;

吴兴旺,男,教授,主任医师,博士生导师,责任作者, E-mail: duobi2004@126.com

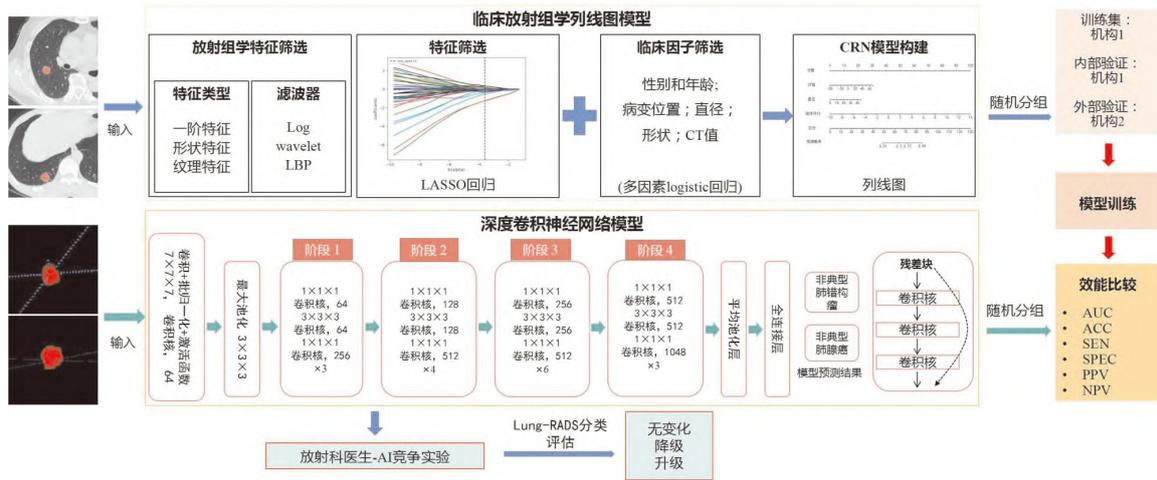


图1 CRN模型和DCNN模型建立过程

均为美国 GE 公司 16 排以上 CT, 包括 Light speed VCT、Optima CT 660 及 Discovery CT750 HD。主要参数如下:管电压 120 kV,管电流 150 ~ 300 mA;扫描层厚及层距均为 5 mm,薄层重建图像为标准算法,层厚及间距为 1.25 mm。由两名高年资放射科主治医师(W. C. B. 和 C. F)共同完成临床及影像学特征的统计,包括患者的性别、年龄、病变的位置、平均直径(取结节最大层面长径与短径之和的 1/2,单位 mm)、形状(分叶征)和 CT 值(在标准算法的薄层纵隔窗图像上进行测量,感兴趣区(region of interest, ROI)放置于病灶中心并尽量避开容积效应的影响,面积约占病灶的 1/2 ~ 2/3,单位 HU)。采用 ITK-SNAP(version 3.8, www.itksnap.org)软件包,由上述两名医师在薄层肺窗图像上,沿靶病灶边缘逐层勾画出分割线,分割线一般不超过病变边缘 1 ~ 2 mm,同时对靶病灶上下缘轮廓显示不完整的层面予以舍弃,最后保存为三维容积感兴趣区(volume of interest, VOI),利用人工智能科研平台(汇医慧影,http://radcloud.cn),完成放射组学特征的提取(共 1 316 个),包括:一阶特征、形状特征、二阶特征[灰度共生矩阵(gray level co-occurrence matrix, GLCM)、灰度行程矩阵(grey-level run-length matrix, GLRLM)、灰度区域大小矩阵(grey-level size zone matrix, GLSZM)、附近灰度色调差矩阵(neighbourhood grey-tone difference matrix, NGTDM)、灰阶依赖性矩阵(grey-level dependence matrix, GLDM)],以及小波变化(wavelet)、拉普拉斯(log)和局部二值模式滤波器(LBP)变化特征。

1.3 特征筛选及 CRN 模型的建立 对训练集中的临床及影像学特征采用先单因素分析后多因素分

析的方式筛选出差异有统计学意义的特征。对于放射组学特征的筛选,首先采用组间相关系数(intra-class correlation coefficient, ICC)(图2)评估特征的鲁棒性(ICC ≥ 0.75 表示一致性较好);针对训练集,本研究选择鲁棒性较好的特征,采用最小绝对收缩和选择算子(least absolute shrinkage and selection operator, LASSO)^[5]方法继续接下来的特征筛选;最后利用公式: $Radscore = Intercept + \sum_{i=1}^n coefficients [i] Feature [i]$ 计算出每个患者的放射组学评分(Radscore)。本文将 Radscore 作为独立的生物标志物与筛选出的临床及影像学特征相结合,建立多元逻辑回归模型,然后绘制个性化预测列线图将模型可视化,并采用决策曲线分析的方法,通过量化不同阈值下的净效益评价 CRN 模型的实用性。

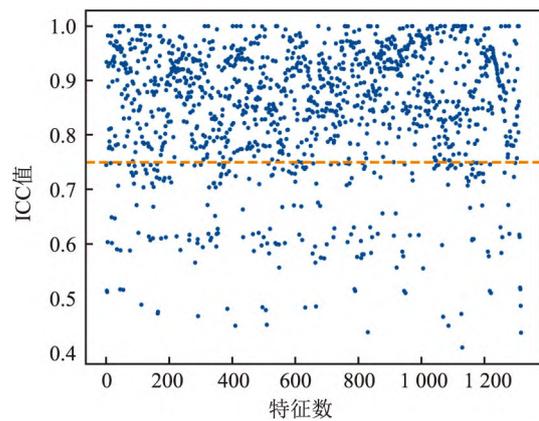


图2 组间相关系数图

观察者一致性检验(ICC 值为 0.89),保留 ICC 值 ≥ 0.75 的放射组学特征值(黄线以上)

1.4 DCNN 模型的建立 利用经典的深度学习 3D Resnet-50^[6]建立 APH 和 ALA 的分类模型,模型结构如图 1 所示,在残差模块中,具有可学习参数的层

用矩形框表示,每个矩形框代表一个卷积核,分别对应 $1 \times 1 \times 1$ 、 $3 \times 3 \times 3$ 以及 $1 \times 1 \times 1$ 卷积核,黑色实线箭头对应卷积操作,虚线箭头表示特征图之间的密集连接。Resnet-50 网络结构可以分成 7 个部分,第 1 部分不包含残差块,主要对输入进行卷积、批归一化、激活函数、最大池化的计算。第 2、3、4、5 部分结构都包含了残差块(阶段 1 ~ 阶段 4)。在 Resnet-50 网络结构中,残差块都有 3 层卷积,网络总共有 $1 + 3 \times (3 + 4 + 6 + 3) = 49$ 个卷积层,加上最后的全连接层总共是 50 层。本研究将所有感兴趣区域调整到 $64 \times 64 \times 32$ 像素并采用最大最小值的方式对图像进行归一化处理。将训练集中的感兴趣区域影像输入 3D Resnet-50 网络,经过前 5 部分的卷积计算,池化层会将其转化成 1 个特征向量,通过反向传播更新模型参数,以网络的输出作为分类结果,计算输出和标签的二进制交叉熵作为损失函数。使用数据增强策略减轻过度拟合,即每个训练图像被随机旋转 30° 、 60° 、 90° 和 180° (在验证集中没有数据增强技术),以及使用了策略包括 L2 正则化(权重衰减为 0.000 5)和早期停止。将批处理大小设置为 8,学习速率设置为 1×10^{-4} ,并应用 Adam 优化器进行优化。课题组使用的神经网络代码是运行于 Ubuntu16.04 系统,神经网络训练过程使用到的程序包包括 Python3.6.8、Pytorch 1.6.0、Opencv、Numpy、SimpleITK 等,同时在 2 块 NVIDIA 的 1080Ti 图形处理单元上训练模型多达 1 000 次迭代。

1.5 AI 辅助评价 Lung-RADS 分类 重新招募 1 名有 15 年诊断经验的放射科医师对机构 1 的 263

个病例(APH = 138, ALA = 125)进行 Lung-RADS 分类,根据 Lung-RADS 1.1 版本^[3]和本研究的入组标准,上述 263 例均被定义为 4A ~ 4X 类,表明这些结节均有一定的恶性概率(可疑恶性)。利用 AI 以概率法重新评估上述病例中结节的性质,并与放射科医师评估的 Lung-RADS 分类做对比,进一步判别在 AI 的协助下 Lung-RADS 分类的调整情况,包括无变化、升级或降级。

1.6 统计学处理 采用 SPSS 22.0 统计软件包(IBM SPSS, Armonk, NY, USA)进行分析,使用独立样本 *t* 检验比较年龄与 CT 值的差异,并以 $\bar{x} \pm s$ 表示;采用卡方检验分析性别、病灶分布、位置和形状的差异,以 % 表示;采用 Mann-Whitney U 检验分析病灶直径,以中位数 $M(P_{25}, P_{75})$ 表示。采用 Spearman 相关分析比较所选放射组学特征的共线性(系数 < 0.5 弱相关, $0.5 \sim 0.8$ 中等相关,系数 > 0.8 强相关)。采用 sklearn.metrics 统计软件包,使用 95% 置信区间(confidence interval, CI)、准确性(accuracy, ACC)、敏感性(sensitivity, SEN)、特异性(specificity, SPEC)、阳性预测值(positive predictive value, PPV)和阴性预测值(negative predictive value, NPV)的 AUC 值来评价基于上述模型的性能。采用 Delong 检验和 ROC 曲线比较两组模型的性能。 $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 研究对象及临床 - 影像学特征筛选 一般资料见表 1。训练集中的 ALA 组和 APH 组在性别、年

表 1 训练集、内部和外部验证集中患者的临床特征

项目	训练集 (n=184)			内部验证集 (n=79)			外部验证集 (n=44)		
	APH (n=97)	ALA (n=87)	P 值	APH (n=41)	ALA (n=38)	P 值	APH (n=23)	ALA (n=21)	P 值
年龄(岁, $\bar{x} \pm s$)	54.6 ± 1.0	61.8 ± 1.1	0.514	56.2 ± 1.4	63.5 ± 1.7	0.001	53.4 ± 2.9	63.0 ± 1.7	0.008
性别[n(%)]			0.737			0.538			0.016
女性	47 (48.5)	40 (46.0)		19 (46.3)	15 (39.5)		7 (30.4)	14 (66.7)	
男性	50 (51.5)	47 (54.0)		22 (53.7)	23 (60.5)		16 (69.6)	7 (33.3)	
CT 值($\bar{x} \pm s$)	18.1 ± 1.9	32.0 ± 1.5	0.001	16.8 ± 2.8	28.9 ± 2.0	0.004	19.5 ± 2.9	30.9 ± 3.0	0.009
直径[M(P_{25}, P_{75})]	12.6 (10.9, 18.1)	21.1 (17.4, 24.7)	0.000	12.6 (10.9, 14.6)	22.4 (18.2, 25.7)	0.000	16.4 (15.5, 18.4)	19.2 (16.8, 24.9)	0.012
肿瘤位置[n(%)]			0.196			0.152			0.129
左肺上叶	27 (27.8)	24 (27.6)		12 (29.3)	8 (21.1)		4 (17.4)	7 (33.3)	
左肺下叶	18 (18.6)	17 (19.5)		9 (22.0)	5 (13.2)		4 (17.4)	6 (28.6)	
右肺上叶	18 (18.6)	27 (31.0)		6 (14.6)	11 (28.9)		1 (4.3)	2 (9.5)	
右肺中叶	5 (5.2)	4 (4.6)		1 (2.4)	5 (13.2)		3 (13.0)	0 (0)	
右肺下叶	29 (29.9)	15 (17.2)		13 (31.7)	9 (23.7)		11 (47.8)	6 (28.6)	
分叶征[n(%)]			0.000			0.345			0.112
无	46 (47.4)	18 (20.7)		16 (39.0)	11 (28.9)		12 (52.2)	6 (28.6)	
有	51 (52.6)	69 (79.3)		25 (61.0)	27 (71.1)		11 (47.8)	15 (71.4)	

龄和病变位置的差异性无统计学意义 ($P = 0.737$ 、 0.514 、 0.196), 在 CT 值、直径和分叶征的差异性有统计学意义 ($P = 0.001$ 、 0.000 、 0.000), 其中 APH 组的 CT 值、直径和分叶征比例均表现更小; 进一步的多因素分析显示, 训练集中的 ALA 组和 APH 组仅在 CT 值和直径的差异有统计学意义 (P 值均为 0.000) (表 2)。

表 2 训练集中临床特征的多变量分析

	P 值	OR 值	95% CI for OR
CT 值	0.000	1.043	1.019 ~ 1.068
直径	0.000	1.177	1.101 ~ 1.258
分叶征	0.214	0.614	0.285 ~ 1.324

2.2 放射组学特征筛选及 CRN 模型建立 两名观察者间的 ICC 为 0.89, 将其中 1 203 个 ICC ≥ 0.75 的放射组学特征进行 LASSO 降维后, 最终筛选出 3 个关键特征: 特征 1 (original-glcM-MCC, 最大相关系数), 特征 2 (original-shape-flatness, 平面度) 和特征 3 (original-shape-sphericity, 球形度), 每个患者

对应的 Radscore 值 = $-0.0759 - 2.8236 \times \text{original-glcM-MCC} - 0.4388 \times \text{original-shape-flatness} - 2.4037 \times \text{original-shape-sphericity}$ 。Spearman 相关分析及热图显示 3 个特征的组内相关性较弱 ($r < 0.5$) (图 3)。最后, 基于 CT 值、直径和 Radscore 评分建立了 CRN 模型, 并使用列线图 (图 4) 将模型可视化 ($\text{Nomogram} = -2.5208 + 0.8661 \times \text{Radscore} + 0.0984 \times \text{直径} + 0.0269 \times \text{CT 值}$)。在 CRN 预测模型中, 预测概率范围为 0.01 ~ 0.09, Radscore 的风险参数最高, 其次是平均直径和 CT 衰减值。决策曲线 (图 5) 评估列线图的获益程度显示, 对所有患者使用 CRN 模型的预测比干预或不干预更有益。

2.3 CRN 及 DCNN 模型的性能评价及比较

ROC 曲线 (图 6) 结果显示, 各组模型中 DCNN 模型的 AUC 值均高于 CRN 模型。CRN 模型在训练集、内部和外部验证集中的 AUC 值分别为 0.968、0.953 和 0.932 (表 3), 分别低于 DCNN 模型在各验证集的 AUC 值 (0.983、0.973 和 0.942) (表 3), 但德隆检验 (表 4) 结果显示上述差异性均无统计学意义

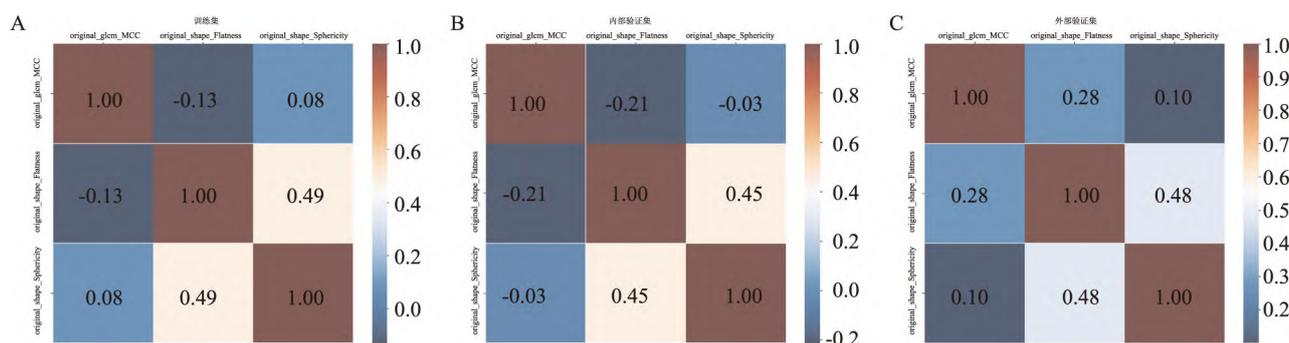


图 3 训练集 (A)、内部 (B) 和外部验证集 (C) 中 3 个放射组学特征的组内相关性的热图

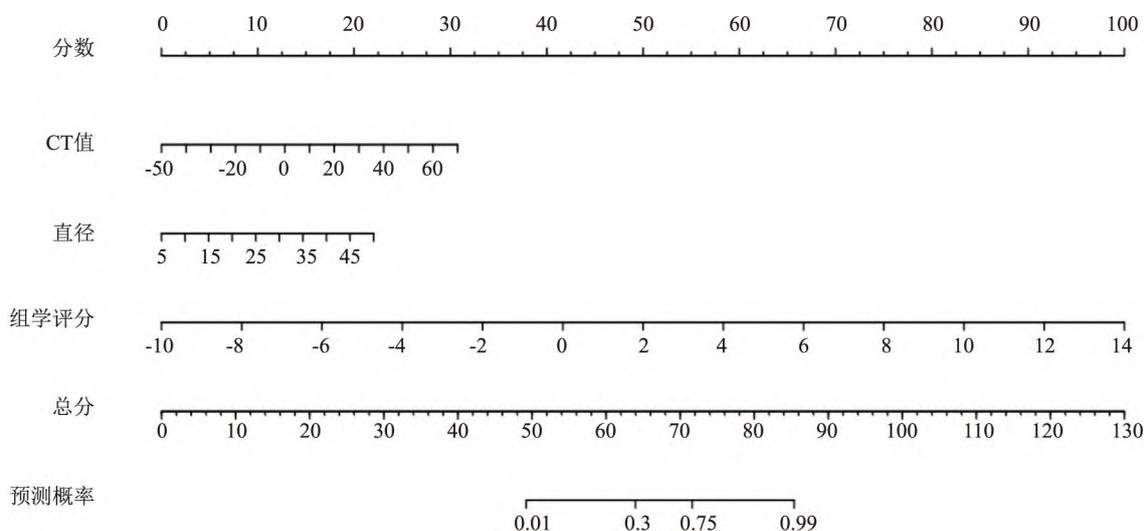


图 4 临床放射组学列线图

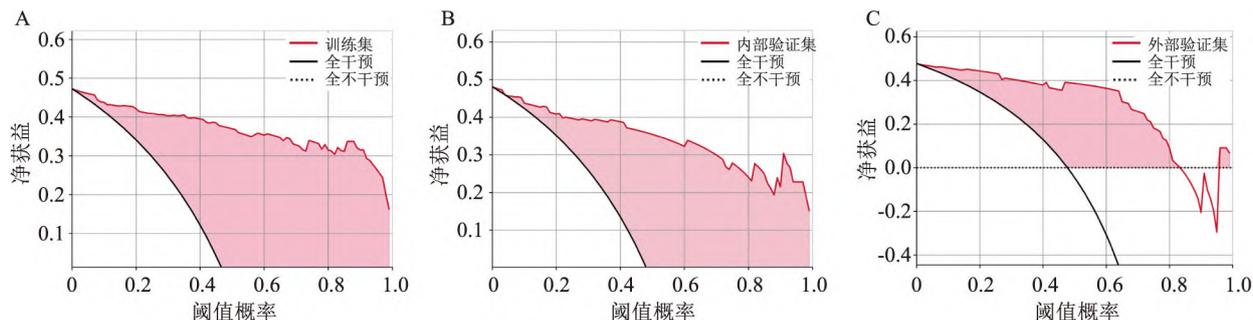


图5 训练集(A)、内部(B)和外部验证集(C)中评价CRN模型的实用性的决策曲线

表3 基于DCNN模型和CRN模型的训练集、内部和外部验证集的性能

项目	训练集		内部验证集		外部验证集	
	DCNN 模型	CRN 模型	DCNN 模型	CRN 模型	DCNN 模型	CRN 模型
AUC (95% CI)	0.983 (0.964, 1.000)	0.968 (0.949, 0.984)	0.973 (0.945, 1.000)	0.953 (0.916, 0.983)	0.942 (0.875, 1.000)	0.932 (0.858, 0.992)
ACC	0.918	0.902	0.861	0.886	0.886	0.886
SEN	0.874	0.874	0.789	0.868	0.905	0.857
SPEC	0.959	0.928	0.927	0.902	0.870	0.913
PPV	0.950	0.916	0.909	0.892	0.864	0.900
NPV	0.894	0.891	0.826	0.881	0.909	0.975

粗体数字代表了各队列中的最佳性能。

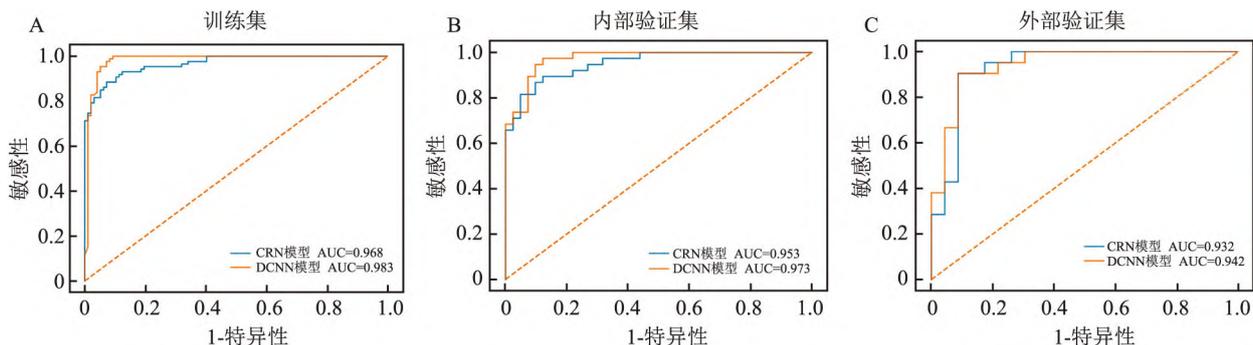


图6 训练集(A)、内部(B)和外部验证集(C)中比较CRN和DCNN模型性能的ROC曲线

表4 德隆检验比较CRN模型和DCNN模型的性能

	CRN 模型	DCNN 模型	P 值
	AUC (95% CI)	AUC (95% CI)	
训练集	0.968 (0.949, 0.984)	0.983 (0.964, 1.000)	0.23
内部验证集	0.953 (0.916, 0.983)	0.973 (0.945, 1.000)	0.31
外部验证集	0.932 (0.858, 0.992)	0.942 (0.875, 1.000)	0.34

($P=0.23, 0.31, 0.34$)。

2.4 AI 辅助评价 Lung-RADS 分类 在 AI 的重新评估下(表5), 138 例 APH 中有 7 例被 AI 错误判断为恶性结节(错误率 5%, 7/138), 其相应的 Lung-RADS 分类不变, 余下 131 例被 AI 正确判断为良性结节(准确率 95%, 131/138), 其相应的 Lung-RADS 分类降级; 125 例 ALA 中有 106 例被 AI 正确判断为恶性结节(准确率 85%, 106/125), 其相应的 Lung-

RADS 分类不变, 余下 19 例被 AI 错误判断为良性结节(错误率 15%, 19/125), 其相应的 Lung-RADS 分类降级。根据调整后的 Lung-RADS 分类结果, AI 判断 APH 的准确率明显高于放射科医师, 判断 ALA 的准确率与放射科医师基本一致。

表5 AI 辅助下对放射医生评估的 Lung-RADS 分类调整情况(n)

调整情况	无变化	降级	升级
APH (n = 138)	7	131	0
ALA (n = 125)	106	19	0

3 讨论

提高 APH 和 ALA 的无创性诊断能力有助于患者在临床决策中受益, 然而这对于依赖诊断经验的

放射科医师来说较为困难,本研究建立了基于 CT 图像的放射组学列线图 and DCNN 的两种预测模型,在解决上述问题方面取得了较为理想的结果。

CT 值和直径是本研究中构建 CRN 模型的重要影像学特征,本研究中 APH 组的 CT 值及直径均低于 ALA 组,可能原因如下:在组织学上,PH 主要是由 CT 值较低的黏液和脂肪组织构成,因此相对的“低密度”是区分 PH 与其他肺部病变的重要特征^[9];在生物学行为方面,PH 则具有良性肿瘤生长缓慢的特点,与肺癌快速生长明显不同,Hansen et al^[10]报道了 14 例 PH 的平均直径在平均 4.1 年内仅增加了(3.2 ± 2.6)mm。

基于 CT 图像的放射组学特征能够间接反映出病灶内部的衰减分布,这是一般影像学特征无法实现的^[11-13]。本研究筛选出的 3 个放射组学特征中,APH 组的 original-glcM-MCC, original-shape-flatness 均高于 ALA 组,这两个特征分别是纹理复杂度的测量和 ROI 中最大和最小主成分之间的关系^[13-14],这一结果表明,APH 的内部衰减分布比 ALA 更不均匀,APH 的组织成分的差异大于 ALA,这可能是由 APH 组织成分的多样性引起的^[9],放射组学的定量特性可以帮助医师识别这种差异性。此外,这也提示肿瘤的异质性不仅仅代表恶性病变,也有可能复杂的组织成分有关。APH 组的另一个关键特征 original-sham-sphericity 亦高于 ALA 组,该特征有可能反映了病变的扩张性和相似程度^[13-14],表明 APH 比 ALA 更有可能显示出球形的形态,这也符合良性肿瘤的膨胀性生长的特点^[7]。

最后,课题组利用上述影像学特征及放射组学特征所建立的 CRN 模型获得了较好的诊断性能,实现了在训练集、内部验证集和外部验证集依次为 0.968、0.953 和 0.932 的 AUC 值。个性化预测的列线图显示,Radscore 在区分 ALA 和 APH 方面优于影像学特征,进一步揭示了放射组学在鉴别诊断方面的能力可能优于放射科医师^[10-12]。

DCNN 是一种深度人工神经网络,它可以捕获影像图像在初始卷积层中存在的纹理信息^[4-7],近年来取得了一些比放射组学更先进的成就^[13-15]。在一项预测肺腺癌 EGFR 突变的研究中,Wang et al^[13-14]采用 DCNN 方法获得的结果明显优于放射组学方法。Li et al^[15]在另一项研究中采用了 DCNN 和放射组学方法预测了弥漫性胶质瘤的分子亚型,前者在大多数情况下表现更好。本研究中,在区分 ALA 和 APH 方面,DCNN 模型取得了一个相对于

CRN 模型更好的结果(训练集、内部和外部验证集的 AUC 分别为 0.983 vs 0.968、0.973 vs 0.953 和 0.942 vs 0.932),表明 DCNN 可能是更有前途的医学应用。

肺结节的 Lung-RADS 分类是否准确直接影响到患者的临床决策。在临床工作中,放射科医师对肺结节 Lung-RADS 分类的判断是以结节的直径和影像学特征作为参考的,并受到一定的主观判断性的影响,这常常导致分类结果不能真实反映出病灶的性质。根据 Lung-RADS 1.1 版本^[3]和入组标准,本研究的 263 例病灶在术前均被放射科医师定义为 4A~4X 类(可疑恶性),意味着 138 例良性的 APH 在术前均被错误定性。而在人机竞赛实验中,AI 将 131 例 APH(94.9%,131/138)判断为良性,相应的 Lung-RADS 分类被降级,表明 AI 在评估 APH 病变性质方面的表现明显优于放射科医生,可以显著降低 APH 的误诊率;同时 AI 将 106 例 ALA(84.8%,106/125)判断为恶性,相应的 Lung-RADS 分级保持不变,表明 AI 在评估 ALA 病变的性质方面倾向于与放射科医生的判断一致,有助于帮助放射科医师增加对 ALA 诊断的信心。此外,研究^[15]表明,DCNN 方法在研究队列较大时将表现得更好,在本研究中,仍有 7 例 APH 和 19 例 ALA 被 AI 误判,可能需要更大的样本去验证。

综上所述,DCNN 及 CRN 在区分 APH 和 ALA 方面均具有较高价值,前者表现更优;AI 在评价肺结节的 Lung-RADS 分类方面亦表现出优于放射科医生的能力。DCNN 在肺结节的临床实践中有可能具有更广阔的应用前景。

本研究仍有一些局限性。首先,需要一项包含更多中心成像数据的前瞻性研究来验证研究结果。其次,本研究仅选择了 APH 和 ALA 作为研究对象,需要纳入其他性质不同但影像表现相似的肺部结节进行研究。第三,不同 CT 设备采集的图像和不同人工分割的图像可能会影响结果。

参考文献

- [1] 阳 明, 杨 敏, 杨有优. 肺错构瘤的 CT 征象:着重强调低密度低强化征和强化分界征[J]. 临床放射学杂志, 2021, (11): 2110-3.
- [2] 陈苏闽. 肺腺癌和肺鳞癌与 CT 影像学表现的相关性研究[J]. 影像研究与医学应用, 2021, (19): 30-1.
- [3] Kastner J, Hossain R, Jeudy J, et al. Lung-RADS version 1.0 versus lung-RADS version 1.1: comparison of categories using nodules from the national lung screening trial[J]. Radiology,

- 2021, 300(1): 199–206.
- [4] 刘梦秋, 曾飞雁, 丁 虎, 等. 基于MRI影像组学模型鉴别鼻腔鼻窦内翻乳头状瘤与息肉[J]. 安徽医科大学学报, 2023, 58(3): 505–9.
- [5] Jiang M, Zhang D, Tang S C, et al. Deep learning with convolutional neural network in the assessment of breast cancer molecular subtypes based on US images: a multicenter retrospective study [J]. *Eur Radiol*, 2021, 31(6): 3673–82.
- [6] 孙 瑶, 王 祥, 萧 毅. 深度学习技术在肺癌影像学诊断中的应用进展[J]. 肿瘤影像学, 2021, 30(6): 525–31.
- [7] Heuvelmans M A, van Ooijen P M A, Ather S, et al. Lung cancer prediction by deep learning to identify benign lung nodules[J]. *Lung Cancer*, 2021, 154: 1–4.
- [8] 罗成龙, 宋一曼, 岳松伟, 等. 增强CT联合纹理分析鉴别无钙化肺错构瘤与原发肺腺癌的价值[J]. 中国医学影像学杂志, 2023, 31(8): 838–43.
- [9] Takumi K, Nagano H, Harasawa T, et al. Pulmonary hamartoma: feasibility of dual-energy CT detection of intranodular fat[J]. *Radiol Case Rep*, 2021, 16(5): 1032–6.
- [10] Hansen C P, Holtveg H, Francis D, et al. Pulmonary hamartoma [J]. *J Thorac Cardiovasc Surg*, 1992, 104(3): 674–8.
- [11] Yang X, Li C, Hou J, et al. Differentiating peripherally located pulmonary noncalcified hamartoma from carcinoid using CT radiomics approaches[J]. *J Comput Assist Tomogr*, 2023, 47(3): 402–11.
- [12] Zhong Y F, She Y L, Deng J J, et al. 深度学习预测临床I期非小细胞肺癌的N2转移与生存[J]. 国际医学放射学杂志, 2022, 45(2): 232.
- [13] Wang S, Yu H, Gan Y, et al. Mining whole-lung information by artificial intelligence for predicting EGFR genotype and targeted therapy response in lung cancer: a multicohort study[J]. *Lancet Digit Health*, 2022, 4(5): e309–19.
- [14] Wang S, Shi J, Ye Z, et al. Predicting EGFR mutation status in lung adenocarcinoma on computed tomography image using deep learning[J]. *Eur Respir J*, 2019, 53(3): 1800986.
- [15] Li Y, Wei D, Liu X, et al. Molecular subtyping of diffuse gliomas using magnetic resonance imaging: comparison and correlation between radiomics and deep learning[J]. *Eur Radiol*, 2022, 32(2): 747–58.

Clinical radiomics nomogram and deep learning based on CT in discriminating atypical pulmonary hamartoma from lung adenocarcinoma

Wang Chuanbin^{1,2}, Li Cuiping¹, Cao Feng², Gao Yankun¹, Qian Baoxin³, Dong Jiangning², Wu Xingwang¹
 (¹*Dept of Radiology, The First Affiliated Hospital of Anhui Medical University, Hefei 230022*; ²*Dept of Imaging, The First Affiliated Hospital of USTC, Division of Life Sciences and Medicine, University of Science and Technology of China, Hefei 230031*; ³*Huiying Medical Technology (Beijing), Beijing 100192*)

Abstract Objective To discuss the value of clinical radiomic nomogram (CRN) and deep convolutional neural network (DCNN) in distinguishing atypical pulmonary hamartoma (APH) from atypical lung adenocarcinoma (ALA). **Methods** A total of 307 patients were retrospectively recruited from two institutions. Patients in institution 1 were randomly divided into the training ($n = 184$: APH = 97, ALA = 87) and internal validation sets ($n = 79$: APH = 41, ALA = 38) in a ratio of 7:3, and patients in institution 2 were assigned as the external validation set ($n = 44$: APH = 23, ALA = 21). A CRN model and a DCNN model were established, respectively, and the performances of two models were compared by delong test and receiver operating characteristic (ROC) curves. A human-machine competition was conducted to evaluate the value of AI in the Lung-RADS classification. **Results** The areas under the curve (AUCs) of DCNN model were higher than those of CRN model in the training, internal and external validation sets (0.983 vs 0.968, 0.973 vs 0.953, and 0.942 vs 0.932, respectively), however, the differences were not statistically significant ($p = 0.23, 0.31$ and 0.34 , respectively). With a radiologist-AI competition experiment, AI tended to downgrade more Lung-RADS categories in APH and affirm more Lung-RADS categories in ALA than radiologists. **Conclusion** Both DCNN and CRN have higher value in distinguishing APH from ALA, with the former performing better. AI is superior to radiologists in evaluating the Lung-RADS classification of pulmonary nodules.

Key words hamartoma; lung adenocarcinoma; nomogram; deep learning; artificial intelligence; computed tomography