• 胸部影像学 •

可视化影像决策模型在评估肺结节浸润程度中的价值

张榕,蔡宏杰,梁演婷,洪敏萍,刘子蔚,杨少民,王华峰,胡秋根

【摘要】 目的:探讨基于临床资料、影像征象和影像组学特征构建的联合模型在术前对肺结节浸润 程度的预测价值,并通过决策热图及 Shaplev 算法对模型进行可视化分析。方法:回顾性搜集 2018 年 1 月-2022 年 3 月在本院经病理确诊的 179 例肺结节患者的临床资料和术前 CT 图像(肺窗平扫)。根据 肺肿瘤新分类,分为腺体前驱病变组(78 例)和浸润性肺腺癌组(101 例)。采用 Deepwise 软件,分别提 取瘤灶、瘤周3mm 和5mm 区域的影像组学特征。使用单因素分析、相关性分析、Boruta 算法和逐步 logistic 回归分析等特征筛选算法确定各区域的最佳组学特征,然后采用 logistics 方法分别构建 3 个单 区域及2个多区域(肿瘤+瘤周3mm及肿瘤+瘤周5mm)共5个影像组学模型,分析各模型的预测效 能并计算其影像组学评分(Radsocre)。通过单因素和多因素 logistic 回归方法筛选相关临床指标和结 节的主要 CT 征象,并采用 XGBoost 算法将筛选出的高危因素结合瘤灶+瘤周 3 mm 联合模型的影像 组学得分构建临床影像联合模型。额外收集浙江省嘉兴市中医医院经病理证实的 69 例肺结节患者的 临床和 CT 资料来完成联合模型的泛化性验证。利用决策热图和 Shapley 算法对模型分别进行可视化 和特征贡献度分析。结果:相比单区域影像组学模型(训练集:AUC=0.740、0753、0.768;验证集: AUC=0.841、0.856、0.809),多区域影像组学模型在两个数据集中均显示出更高的预测效能(AUC= 0.878 和 0.834)。XGBoost 联合模型的预测效能得到进一步地提高(AUC=0.948 和 0.886)。Shaplev 分析显示影像组学得分、CT 值和结节长度为预测肺结节浸润程度的最重要的 3 个特征。决策热图 算法实现了对浸润性预测推演过程的可视化。结论:XGBoost 模型对肺结节浸润性的评估具有较高的 准确性和泛化性。决策热图实现了可解释机器学习算法的可视化从而保障了模型的实用性,为肺结节 的临床处理及管理提供了一种无创性的辅助诊断工具。

【关键词】 肺结节:浸润程度;瘤周;影像组学;Boruta 算法;XGBoost 算法;Shaplev 算法 【中图分类号】R814.42;R734.2 【文献标志码】A 【文章编号】1000-0313(2024)08-1032-08

DOI: 10. 13609/j. cnki. 1000-0313. 2024. 08. 008

开放科学(资源服务)标识码(OSID):

Value of decision model based on visualized clinical imaging in evaluating the infiltration of pulmonary nodule ZHANG Rong, CAI Hong-jie, LIANG Yan-ting, et al. Department of Radiology, Shunde Hospital, Southern Medical University (the First People's Hospital of Shunde, Foshan), Guangdong 528308, China

[Abstract] Objective: To explore the predictive value of the combined model based on clinical data, imaging signs and radiomics features in the infiltration degree of preoperative pulmonary nodule, and to perform visualized analysis of the model by decision heat map and Shapley algorithm. Methods: Clinical data and preoperative CT images (lung window plain scan) of 179 patients with pulmonary nodules diagnosed pathologically in our hospital from January 2018 to March 2022 were retrospectively collected. According to the new classification of lung tumor, all subjects were divided into glandular prodromal disease group (n=78) and invasive lung adenocarcinoma group (n=101). Deepwise software was used to extract the imaging features of tumor foci, peritumor-3mm and -5mm regions. Unifactor analysis, correlation analysis, boruta algorithm and stepwise logistic regression analysis feature

作者单位:528308 广东佛山,南方医科大学顺德医院(佛山市顺德区第一人民医院)放射科(张榕、刘子蔚、王华峰、胡秋根); 51 浙江杭州,浙江中医药大学第一临床医学院(蔡宏杰);510080 广东广州,广东省人民医院(广东省医学科学院)放射科 310051 浙江嘉兴,浙江省嘉兴市中医医院放射科(洪敏萍);528315 广东佛山,佛山市顺德区乐从医院放射科(杨少 民)

作者简介: 张榕(1990-),男,湖南永州人,硕士研究生,主治医师,主要从事胸部 CT 影像诊断工作。

通讯作者:胡秋根,E-mail:hu6009@163.com

基金项目:佛山市科技计划项目(2220001005383);南方医科大学顺德医院科研启动计划项目(SRSP2021021);佛山市高层次医 学人才培养(医学骨干人才)

screening algorithm were used to determine the best radiomics features of each region. Then, three single-region models and two multi-regions (tumor+peritumor-3mm, and tumor+peritumor-5mm) models were constructed by logistics method. A total of five radiomics models were used to analyze the predictive efficacy of each model and calculate its radiomics score (Radsocre). The relevant clinical indicators and major CT signs of nodules were screened by univariate and multivariate logistic regression methods, and the combined clinical image model was constructed by combining the selected high risk factors and Radscore of tumor+peritumor-3mm using XGBoost algorithm. Additional clinical and CT data of 69 patients with pulmonary nodules confirmed by pathology in Jiaxing Hospital of Traditional Chinese Medicine of Zhejiang Province were collected to complete the generalization verification of the combined model. The decision heatmap and Shapley algorithm were used to visualize the model and explain the feature's contribution, respectively. Results: Compared with the three single-region radiomics models (training set: AUC=0.740,0753 and 0.768; validation set: AUC=0.841,0.856 and 0.809), the multi-region radiomics models showed better performance in the two datasets (AUC=0.878 and 0.834). The performance of the XGBoost combined model was further improved (AUC=0.948 and 0.886). Shapley algorithm analysis showed that Radscore, CT value and nodule length were the three most important characteristics for predicting pulmonary nodule invasiveness. The decision heatmap algorithm realizes the visualization prediction and deduction process of tumor infiltration. Conclusion: The clinical-radiomic combined XGBoost model has high accuracy and generalization in the evaluation of pulmonary nodule invasiveness. The decision heatmap can realize the visualization of interpretable machine learning algorithms ensures the practicability of the model and provides a non-invasive tool for clinical treatment and management of pulmonary nodules.

[Key words] Pulmonary nodules; Infiltration; Peritumor; Radiomic; Boruta algorithm; Extreme gradient boosting algorithm; Shapley additive explanations

根据 2020 年全球癌症数据统计结果,肺癌发病率 仅次于乳腺癌,位于第二位,但其致死率仍居于第一 位[1-2],故其已经成为了人们关注的一个重大公共卫生 问题。2021年5月国际癌症研究机构提出了肺肿瘤 最新分类标准,将肺腺癌分为腺体前驱病变和浸润性 肺腺癌,前者包括不典型腺瘤样增生(atypical adenomatous hyperplasia, AAH)和原位癌(adenocarcinoma in situ, AIS), 后者包括微浸润性肺腺癌 (minimally invasive adenocarcinoma, MIA) 和浸润性肺腺癌(invasive adenocarcinoma, IAC)[3]。多项研究证实, AAH/AIS 的术后 5 年无病生存率为 100.0%, MIA 为 95.0%~100.0%,IAC 为 38.0%~87.6%^[4]。对 于腺体前驱病变积极随访观察是主要的处理方式,而 浸润性肺腺癌则需要及时手术干预。由于 AIS 和 MIA 术后复发率均较低,因此有部分学者建议将两者 归并为一个类别进行研究[5]。然而,有国际癌症研究 机构指出[3],目前的数据不足以支持这样的建议。病 理分型是一种用于评估肿瘤侵袭或扩散程度的方法, 由于术后病理检查具有滞后性,无法指导临床在术前 制订精准的手术治疗方案,因此,在术前准确区分 AAH/AIS与 MIA/IAC 有助于临床医师选择合适的 治疗手段及随访计划。

肿瘤微环境的异质性在肺结节浸润性发展过程中起着至关重要的作用^[6]。近年来,已有研究证实基于瘤周影像的影像组学分析技术可有效捕获结节周围微环境的异质性信息^[4]。然而,肺结节浸润性预测模型的内在决策机制和推演过程尚不清晰,这限制了模型的推广应用。本研究通过联合瘤灶及不同瘤周范围的CT影像组学特征及传统临床和影像征象,开发和验证一种可解释且泛化性强的临床影像决策模型,旨在为肺结节浸润程度的术前精准评估及个体化治疗提供一种无创性方法。

材料与方法

1. 临床资料

本研究纳入了广东省南方医科大学顺德医院及浙江省嘉兴市中医医院 2018 年 1 月 - 2022 年 3 月共 248 例肺结节患者。根据肺腺癌最新分类标准,将其分为两组: AAH/AIS 组和 MIA/IAC 组。纳入标准:①肺结节最大直径≤30 mm(横轴面 CT 图像上),且经病理证实为腺癌;②CT 平扫图像、病理(包括免疫组织化学检测)及临床资料(性别、年龄、病史等)完整;③既往未接受肺结节相关抗肿瘤治疗;④无其它恶性肿瘤病史。排除标准:①CT 图像质量欠佳或有明显

移动伪影等;②病理证实肺结节内存在其它病理成分 (如鳞状细胞癌或小细胞肺癌等)。

将广东省南方医科大学顺德医院的 179 患者作为训练集(AAH/AIS 78 例, MIA/IAC 101 例),浙江省嘉兴市中医医院的 69 患者作为外部验证集(AAH/AIS 患者 23 例, MIA/IAC 患者 46 例)。

本研究获得了上述两家医院伦理委员会的批准, 系回顾性研究,无需患者签署知情同意书。

2. 检查方法

使用 Siemens Somatom Definition AS 128(南方医科大学顺德医院)或 GE Light speed 16(浙江省嘉兴市中医医院)螺旋 CT 机。所有患者取仰卧位,双臂上举,头先进,扫描时嘱患者屏住呼吸。扫描范围自胸廓入口至肺底水平,扫描仪器及参数见表 1。

表 1 两家医院 CT 机的扫描参数

扫描参数	Siemens CT 机	GE CT 机
管电压/kV	120	120
管电流/mA	300	250
扫描野/mm²	$360 \sim 400$	$360 \sim 400$
矩阵	512×512	512×512
扫描层厚/mm	7.0	2.5
重建层厚/ mm	0.6	0.6
准直宽/mm	128×0.60	16×1.25

3. 图像分析和影像组学特征提取

分析结节的常规影像学特征,主要包括病灶的平 均直径、CT 值、成分(实性/混合磨玻璃/纯磨玻璃)、 位置(左肺上叶/下叶、右肺上/中/下叶)、空泡征、空气 支气管征、月牙征、分叶征、毛刺征、微血管穿行征、胸 膜牵拉征、支气管充气征、月牙征、微血管征、空泡征和 瘤-肺边界(清晰/模糊)。将图像导入 Deepwise 软件 (https://keyan. deepwise. com/login),由两位具有 10 年以上胸部影像诊断经验的医师分别采用自动/半 自动分割方法在肺窗图像上(窗宽 1500 HU,窗位 -500 HU)沿着病灶边缘逐层勾画 ROI,由软件自动 生成容积 ROI(volume of interest, VOI)。完成肿瘤 勾画后,使用软件自动对所勾画的瘤灶边界分别向外 扩展 3 mm 和 5 mm,此时由医师对外扩的区域进行核 查,手动去除外扩 ROI 时包含的非肺部结构(如骨、胸 膜等),随后也分别生成 ROI 外扩 3 和 5 mm 相应的容 积 VOI(VOI_{3 mm}、VOI_{5 mm})。

间隔 2~3 个月后由其中一位医师从总样本中随机抽取 30 例患者,再次对病灶进行 ROI 的勾画。

采用 Pyradiomics 程序包对每个病灶的 3 种 VOI $(VOI \ all \ xVOI_{3 \ mm}, VOI_{5 \ mm})$ 分别进行特征提取,包括直方图(-阶统计特征)、形状特征(=维和形状),灰度区域大小矩阵、灰度共生矩阵、灰度依赖矩阵等灰度特征、squareroot、wavelet 滤波特征。采用组间和组内

相关系数(ICC)对提取的特征进行一致性检验,ICC>0.75认为一致性好。为了消除源于不同中心的不同扫描仪及成像参数的差异导致的潜在批次效应,采用ComBat 批次校正分析和 Z-score 标准化方法对影像组学数据进行标准化。

4. 影像组学特征筛选及模型构建

因提取特征的数据维度过高,为了避免模型过拟 合,提高模型的泛化性,我们对不同范围的 ROI 所提 取的影像组学特征分别进行特征筛选,具体步骤如下。 ①单因素分析:采用独立样本 t 检验或秩和检验进行 组间比较,筛选出有显著差异的特征。②相关性分析: 对所有特征两两进行相关性分析,若两特征的 Pearson 或 Spearman 相关性大于 0.6,则删除与其它特征 平均相关性更高的组学特征。③特征重要性分析:利 用 Boruta 算法随机引入影子变量与原始特征进行效 能比较,通过算法的多次迭代逐批删除低效能和冗余 的组学特征。④逐步 logistic 回归分析:对特征进行 逐步回归分析,保留对于肺结节浸润性评估最重要(优 势比)的特征,计算特征的权重系数并进行线性组合来 构建影像组学模型,并计算影像组学得分(Radscore); ④最后,再次基于逐步 logistic 回归方法对三种 ROI 区域筛选出的最优特征进行联合分析,构建联合模型 并计算影像组学得分。对基于单区域和多区域 ROI 构建的模型进行 ROC 曲线分析,获得最佳影像组学 模型。

5. 临床-影像及综合模型的构建

对临床因素和影像征象进行单因素 logistic 回归分析,筛选出 P<0.05 的显著因素;随后,进行多因素 logistic 分析,筛选出肺结节浸润性相关的高危因素并完成临床-影像模型的构建。最后,采用 XGBoost 机器学习算法构建临床、影像和组学的综合模型,并利用 Shapley 分析技术可视化模型中各特征的贡献情况。

6. 统计分析方法

组内相关系数分析、单因素分析、相关性分析、Boruta 特征选择算法、逐步 logistic 回归分析、ROC 曲线等统计学方法均使用 R 语言(版本 4.1.3)。通过 pycaret 软件包完成 XGBoost 机器学习模型的构建和 Shapley 分析。以 P < 0.05 为差异具有统计学意义。

结 果

1. 临床资料

两个医院的肺癌患者中 AAH/AIS 组和 MIA/IAC 组基线临床资料和 CT 征像的比较结果详见表2。顺德医院的 179 例中,肺结节直径、CT 值、成分、月牙征、分叶征、毛刺征、微血管穿行征及胸膜牵拉征在 AAH/AIS 组与 MIA/IAC 组之间的差异均有统计

学意义(P<0.05);中医院的69例中,肺结节直径、CT值、成分、月牙征、分叶征、毛刺征及胸膜牵拉征在AAH/AIS组与MIA/IAC组之间的差异均有统计学意义(P<0.05)。

利用 Python 软件对每例患者的 VOI 瘤灶、 $VOI_{3\,mm}$ 和 $VOI_{5\,mm}$ 分别进行影像组学特征的提取,每种 VOI 提取了 1015 个特征,其中一致性好的特征分别有 774、675 和 689 个。经过各步骤的特征筛选分析,最终基于 3 种 VOI 分别保留了 2、4 和 4 个组学特征(表 3)。

基于 3 个单 VOI 及联合 3 个 VOI 的组学特征构建预测结节浸润性的组学模型,各模型的预测效能指标值见表 4。在单 VOI 组学模型中,瘤周 5 mm 模型在训练集表现出了最好的预测效能(AUC=0.768,截断值=0.681),瘤周 3 mm 模型在验证集中表现出了最好的预测效能(AUC=0.856,截断值=0.367);而联合三区域影像组学特征基于逐步 logistic 回归方法构建的多区域组学模型(包含 1 个瘤灶特征和 3 个瘤

周 3 mm 特征)在验证集中表现出了更高的预测效能 (AUC=0.787,截断值=0.049),其影像组学得分计算公式:影像组学得分=-1.313×wavelet.HHL_gl-cm_Idm(瘤灶)-0.72×wavelet.HHH_ngtdm_Coarseness(瘤周 3 mm)-0.666×wavelet.HHL_ngtdm_Strength(瘤周 3 mm)+0.376×wavelet.HLL_ngtdm_Busyness(瘤周 3 mm)+0.441。

表 3 各种特征筛选方法筛选出的影像组学特征数

VOI 类型	ICC	单因素 分析	相关性 分析	Boruta	逐步 logistic 回归
VOI 瘤灶	774	651	52	28	2
$\mathrm{VOI}_{3\mathrm{mm}}$	675	622	50	8	4
$\mathrm{VOI}_{5\mathrm{mm}}$	689	648	64	14	4

3. 临床资料和影像特征筛选及模型建立

对所有临床资料和结节的常规影像特征进行单因素分析,选取 P<0.05 的特征进行多因素分析。经单因素分析发现,月牙征、CT 值、直径、微血管穿行征和胸膜牵拉征为结节浸润性的独立影响因素(表 5)。使用多因素 logistic 回归分析进一步筛选,结果显示月

	广东省	首南方医科大学顺德医	. 院	浙江省嘉兴市中医医院			
指标/征象	AAH/AIS (n=78)	MIA/IAC (n=101)	P值 χ ² 值	AAH/AIS (n=23)	MIA/IAC (n=46)	P 值	χ² 值
性别/例			0.172 1.862			0.855	0.033
女	54 (69.2%)	79 (78.2%)		16 (69.6%)	31 (67.4%)		
男	24 (30.8%)	22 (21.8%)		7 (30.4%)	15 (32.6%)		
空泡征/例	10 (12.8%)	18 (17.8%)	0.361 0.834	5 (21.7%)	10 (21.7%)	0.255	_
空气支气管征/例	4 (5.1%)	6 (5.9%)	>0.999 0.055	1 (4.3%)	8 (17.4%)	0.255	1.294
月牙征/例	11 (14.1%)	41 (40.6%)	<0.001 14.986	1 (4.3%)	13 (28.3%)	0.044	4.044
瘤-肺边界模糊/例	55 (70.5%)	73 (72.3%)	0.795 0.067	22 (95.7%)	36 (78.3%)	0.131	2.285
分叶征/例	7 (9.0%)	26 (25.7%)	0.004 8.230	0 (0.0%)	22 (47.8%)	< 0.001	16.149
毛刺征/例	0 (0.0%)	21 (20.8%)	<0.001 18.383	0 (0.0%)	22 (47.8%)	< 0.001	16.149
微血管穿行征/例	53 (67.9%)	87 (86.1%)	0.003 8.545	22 (95.7%)	45 (97.8%)	>0.999	0.243
胸膜牵拉征/例	4 (5.1%)	22 (21.8%)	0.002 11.366	3 (13.0%)	19 (41.3%)	0.017	5.639
成份/例			<0.001 13.983			0.017	8.092
实性	0 (0.0%)	1 (1.0%)		8 (34.8%)	21 (45.7%)		
混合磨玻璃	15 (19.2%)	45 (44.6%)		15 (65.2%)	16 (34.8%)		
纯磨玻璃	63 (80.8%)	55 (54.5%)		0 (0.0%)	9 (19.6%)		
位置/例			0.237 5.531			0.923	0.911
左肺下叶	4 (5.1%)	10 (9.9%)		6 (26.1%)	9 (19.6%)		
左肺上叶	19 (24.4%)	20 (19.8%)		7 (30.4%)	14 (30.4%)		
右肺下叶	14 (17.9%)	26 (25.7%)		3 (13.0%)	8 (17.4%)		
右肺中叶	8 (10.3%)	4 (4.0%)		1 (4.3%)	4 (8.7%)		
右肺上叶	33 (42.3%)	41 (40.6%)		6 (26.1%)	11 (23.9%)		
年龄/岁	54.1 ± 9.9	53.1 \pm 12.7	0.566 0.575	52.7 \pm 11.6	55.6 ± 10.9	0.312	1.020
直径/mm	6.0(5.0,8.0)	8.0(7.0,10.0)	<0.001 5.289	8.0(7.5,10.0)	11.0(8.0,14.0)	0.005	2.820
CT 值/HU	-588.1(-648.1,-478.3)) -480.0(-572.0,-335.4	4) < 0.001 4.525	-580.1(-665.3,-465.5)	-342.1(-476.0, -84.2)	< 0.001	4.481

表 2 两家医院两组患者基线临床资料和 CT 征像的比较结果

表 4 单 VOI 和多区域组学模型在训练集及验证集中的预测效能

模型类型	截断值	AUC (95 %CI)	符合率	敏感度	特异度
训练集					
VOI瘤灶模型	0.370	$0.740(0.664 \sim 0.814)$	0.707	0.786	0.592
VOI₃mm 模型	0.367	$0.753(0.676 \sim 0.821)$	0.707	0.767	0.620
VOI₅mm 模型	0.681	$0.768(0.690 \sim 0.841)$	0.724	0.748	0.690
多区域联合模型	0.049	$0.787(0.741 \sim 0.851)$	0.759	0.874	0.592
验证集					
VOI 瘤灶模型	0.370	$0.841(0.726\sim 0.937)$	0.757	0.727	0.800
VOI₃mm 模型	0.367	$0.856(0.757 \sim 0.943)$	0.730	0.659	0.833
VOI₅mm 模型	0.681	$0.809(0.700\sim 0.909)$	0.703	0.591	0.867
多区域联合模型	0.049	$0.834(0.728 \sim 0.926)$	0.797	0.841	0.733

牙征、CT 值、直径和微血管穿行征为肺结节浸润性的高危因素(表 6),构建的临床-影像模型的表达式为: Logit(P) = 月牙征×1.02675818490013+CT 值×0.00598207167754747+直径×0.193920692637306+微血管穿行征×1.22465739551877+胸膜牵拉征×0.715886667529435。

表 5 临床资料和影像特征的单因素 logistic 回归分析结果

变量	OR(95%CI)	P 值
年龄	1.001(0.975~1.029)	0.921
性别	0.846(0.432~1.668)	0.626
空泡征	2.264(0.983~5.711)	0.066
空气支气管征	$1.741(0.613 \sim 5.688)$	0.319
月牙征	$3.278(1.547 \sim 7.488)$	0.003
瘤肺边界	$1.292(0.699 \sim 2.389)$	0.412
分叶征	$1.902(1.004 \sim 3.684)$	0.052
微血管穿行征	$2.181(1.151 \sim 4.174)$	0.017
胸膜牵拉征	2.376(1.242~4.681)	0.010
直径	$1.342(1.185 \sim 1.546)$	<0.001
CT 值	1.006(1.004~1.009)	<0.001

表 6 临床影像特征多因素 logistic 回归分析结果

变量	OR(95%CI)	P 值
月牙征	2.792(1.100~7.588)	0.036
微血管穿行征	$3.403(1.32\sim9.714)$	0.015
胸膜牵拉征	2.046(0.689~6.494)	0.207
直径	1.214(1.048~1.429)	0.013
CT 值	1.006(1.003~1.009)	<0.001

4. 临床影像组学综合模型的建立及效能评估 联合临床资料和影像征象中筛选出的高危因素

(月牙征、CT 值、长度、微血管穿行征)和多区域组学模型的影像组学得分,采用 XGBoost 机器学习方法构建临床-影像-组学综合模型(XGBoost 模型)。采用ROC 曲线分析模型的预测效能,结果见表 7。在训练集及外部验证集中,XGBoost 模型对区分肺结节浸润性具有更强的预测效能,AUC 分别为 0.948 和0.886,对应的截断值为 0.449。

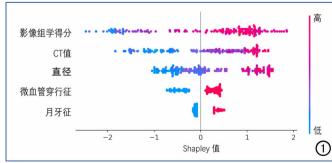
5. 临床影像组学综合模型的解释和可视化

Shapley 分析技术可以为模型提供定量解释,可视化展现各个特征对不同患者的贡献大小。Shapley 汇总图(图1)描述了影像组学得分、CT 值、长度、微血管穿行征和月牙征五个因素对不同肺结节患者浸润预测的贡献情况。其中,Shapley 值越高(颜色越接近红色),提示倾向于 MIA/IAC;反之(越接近蓝色),则倾向于 AAH/AIS。Shapley 值的绝对分布范围越大,提示特征对肺结节浸润评估的重要性更大,由结果可知提示影像组学得分、CT 值、长度是预测肺结节浸润性最重要的特征。

将 XGBoost 模型的第一棵树模型进行决策热图可视化(图 2)。若影像组学得分≪ - 0.31,病变倾向于为 AAH/AIS;若得分> - 0.31,则 CT 值是预测肺结节浸润性的关键特征,若 CT 值≪ -583.8 HU,病变倾向于为 AAH/AIS;若 CT 值> -583.8 HU,此时肺结节长度是预测肺结节浸润性的关键特征,若长度

表 7 训练集和外部验证集模型性能分析

模型	截断值	AUC (95%CI)	符合率	敏感度	特异度
训练集					
临床-影像模型	0.654	$0.837(0.781 \sim 0.890)$	0.759	0.699	0.845
多区域组学模型	0.049	$0.787(0.714 \sim 0.851)$	0.759	0.874	0.592
XGBoost 综合模型	0.449	$0.948(0.916 \sim 0.973)$	0.885	0.893	0.873
外部验证集					
临床-影像模型	0.654	$0.801(0.691 \sim 0.896)$	0.635	0.500	0.833
多区域组学模型	0.049	0.834(0.728~0.926)	0.797	0.841	0.733
XGBoost 综合模型	0.449	$0.886(0.802 \sim 0.955)$	0.824	0.864	0.767



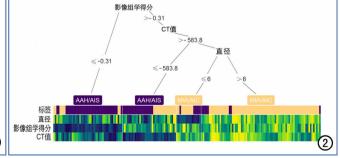


图 1 XGBoost 模型 Shapley 分析汇总图。横坐标为 Shapley 值,纵坐标为特征,每个点代表一例患者某个因素的得分,其颜色越红提示倾向于 MIA/IAC 的贡献值越大,越蓝则代表倾向于 MIA/IAC 的贡献值越小,即更倾向于为 AAH/AIS。 图 2 XGBoost 模型决策热图。紫色代表预测倾向于为 AAH/AIS,橙色代表预测倾向于为 MIA/IAC。

≪6 mm,病变倾向于为 AAH/AIS,若长度>6 mm,则 倾向于为 MIA/IAC。

7. 模型应用案例分析

Shapley 力图,可以用于解释对单个患者的评估。它将特征的 Shapley 值可视化为增加或减少评估的力,并且每个预测都以基础值开始,该基础值(base value)是所有预测特征的平均 Shapley 值。箭头的长度表示特征对 Shapley 值的贡献(以百分比表示),红色箭头为正值,蓝色为负值。图 3~6 列出了 4 例肺结节患者的 Shapley 力图。

讨论

在本研究中,我们联合结节内及结节周围的关键 影像组学特征和重要临床影像征象,开发了一种可解 释的肺结节浸润性评估模型。本研究中使用的 Combat 批次校正方法可消除不同中心、不同仪器扫描时可能导致的数据差异,在外部验证队列上实现了模型效能的有效验证。决策热图算法将模型的推演过程进行了可视化,有助于直观了解模型的分析过程,Shapley 分析算法实现了各特征对肺结节浸润性预测贡献情况的可视化,为临床提供了一种准确、便捷的肺结节浸润性评估手段。

近几年的相关研究结果均表明肿瘤周围微环境在 预测肿瘤侵袭性和治疗反应等方面具有重要作用^[7-8]; 瘤周影像组学特征对区分肺结节良恶性具有较好的预 测效能^[9-11]。然而,不同瘤周范围的影像组学特征的 作用具有差异性:有研究结果表明,基于瘤周 5 mm 区 域的影像组学特征在区分浸润性肺腺癌和良性肉芽肿



图 3 MIA/IAC 组患者的模型分析。女性患者,31 岁,CT 显示右肺中叶有一混合磨玻璃结节,直径为 $10~\mathrm{mm}$,CT 值偏高($-240~\mathrm{HU}$),结节内可见微血管穿行征,影像组学得分(1.885)较高,预测结果为 MIA/IAC,病理结果为 IAC,与之一致。 图 4 AAH/AIS 组患者的模型分析。女性患者,45 岁,CT 显示左肺上叶舌段有一纯磨玻璃肺结节,直径为 6 mm,CT 值($-775~\mathrm{HU}$)较低,未见微血管穿行征,影像组学得分(-0.311)较低,模型预测为 AAH/AIS,病理结果为 AAH,结果一致。 图 5 临床假阴性诊断的误诊分析。女性患者,37 岁,CT 显示右肺上叶有一混合肺磨玻璃结节,直径偏小($5~\mathrm{mm}$),CT 值较高($-346~\mathrm{HU}$),未见微血管穿行征,影像诊断为 AAH/AIS。但模型分析显示影像组学得分为 1.919,数值偏高,提示为 MIA/AIS。经病理确诊该结节为 MIA,与模型分析结果一致,表明模型可辅助识别影像诊断假阴性结节。

图 6 临床假阳性诊断的误诊分析。女性患者,47 岁,CT 显示右肺上叶后段有一纯磨玻璃结节,CT 值稍高 (-484 HU),可见微血管穿行征,影像诊断为 MIA/IAC。然而,模型分析发现结节的影像组学得分数值偏小(-0.557),提示结节倾向于为 AAH/AIS。最终病理确诊为 AIS,与模型分析结果一致,提示模型可辅助临床医师注意对影像假阳性结节的判断。

方面具有较好的预测效能[12-13];也有研究结果显示,与 瘤周2mm区域相比,瘤周5mm区域的影像组学模型 区分 IAC 与 MIA/AIS 的 AUC 的差异无统计学意 义[11]。在本研究中,我们对基于肺结节的不同区域所 构建的影像组学模型对 AAH/AIS 与 MIA/IAC 的鉴 别效能进行了对比分析,发现单区域(VOI 瘤灶、 VOI_{3 mm}和 VOI_{5 mm})组学模型在训练集中的 AUC 分别 为 0.740、0.753 和 0.768, 对 应 的 截 断 值 分 别 为 0.370、0.367 和 0.681,表明瘤周 5 mm 区域影像组学 特征的预测效能最佳。将多区域影像组学特征联合构 建的多区域组学模型的性能得到了改进, AUC = 0.787。但是需要指出,在多区域联合建模过程中,最终 仅纳入了1个瘤灶和3个瘤周3mm影像组学特征, 并未纳入瘤周 5 mm 的影像组学特征,分析其原因可 能是由于本研究中纳入的肺结节均较小,其微环境信 息尚主要集中于瘤周 3 mm 区域。

经过临床多因素分析,筛选出月牙征、CT 值、直 径和微血管穿行征为评估肺结节浸润性的重要危险因 素,与既往的研究结果基本一致[14-16]。在此基础上联 合影像组学得分,利用 XGBoost 机器学习算法完成了 临床影像综合模型的构建,预测效能得到了进一步地 提高(截断值为 0.449,在训练集和验证集中的 AUC 分别为 0.948 和 0.886)。然而,尽管采用影像组学方 法对肺结节浸润性进行评估的可行性已经被证实,如: Wu 等[11] 回顾性分析了 121 例磨玻璃结节样肺腺癌的 瘤周影像组学特征在预测肿瘤病理侵袭性中的价值, 结果显示,基于肺结节瘤周2mm和5mm的影像组学 特征可以有效预测 IAC 与 AIS/MIA; Huang 等[17]的 研究表明,临床-影像学特征和影像组学特征构建的列 线图可以预测孤立性肺结节的病理侵袭性。但以往的 研究中的模型仍缺乏可解释性,很难应用于临床实 践[4]。为了直观理解肺结节浸润预测模型的推演过 程,我们绘制了决策热图来实现模型的可视化(图 2)。 相比 Nomogram,决策热图能更为直观地显示出各特 征对肺结节浸润性预测的重要程度,并按照每个特征 的贡献力进行分层,可显示特征之间的交互作用,能综 合分析特征与结节浸润性的复杂关系[18-19]。本研究中 最终将影像组学得分、CT 值和直径纳入决策热图,分 析其原因可能是这3个特征属于连续变量,而月牙征 和微血管穿行征属于临床分类变量,评估结果的主观 性较强,且本研究中病例数较少,因此影像组学得分、 CT值和直径更适合纳入模型中来预测肺结节的浸润 性。

此外,我们还利用 Shapley 可解释技术分析了不同特征对不同患者肺结节浸润性进展的预测贡献度。 作为一种模型可解释技术,通过 Shapley 图形对模型 进行解释及可视化,目前已得到了初步的应用[19-22]: Wang 等^[23]采用 Shapley 技术建立并验证了基于 MRI 的临床-放射学模型,评估了全脑放疗的治疗效果,可 指导临床医师制订个性化的诊疗策略; Manikis 等[24] 的研究结果表明,利用 Shapley 技术能直观显示异柠 檬酸脱氢酶突变状态与放射组学特征之间的关系,增 加了影像组学的潜在临床应用价值;Pan 等[25]的研究 表明,利用 Shapley 可解释技术有助于识别骨肿瘤分 类中的重要特征,为临床提供更精确的诊断依据。在 本研究中,我们通过 Shapley 汇总图(图 1)可了解影像 组学得分具有最大的 Shapley 分布区间,其 Shapley 值越大则越倾向于提示病变为 MIA/IAC, 反之则倾 向于为 AAH/AIS。Shapley 力图(图 3)实现了患者 特征的个性化解释及可视化,临床医师可以通过 Shaplev力图箭头的颜色及长度去直观地了解每个特征如 何影响对个体患者肺结节浸润性评估的价值[26]。总 而言之,本研究中构建的临床影像组学综合模型不仅 能预测肺结节浸润性,而且还为模型预测的机制提供 了合理的解释及可视化,这将提高临床医师对模型预 测结果的信心,促进影像组学技术的推广和应用。

本研究存在一定的局限性:第一,本研究中的样本来自两个医疗中心,扫描仪和扫描参数的不同对成像数据可造成潜在的影响;第二,本研究为回顾性研究,病例数较少,病例筛选可能存在一定的选择偏倚,需进一步扩大样本进行前瞻性研究;第三,本研究中的影像组学特征是从 CT 图像上提取,未来需要将影像组学与病理免疫组化、基因学联合起来进行更深入地研究。

综上所述,本研究构建了临床影像组学综合模型,可用于鉴别肺结节的浸润性(AAH/AIS 和 MIA/IAC),且具有较好的预测效能;进一步采用了决策热图和 Shapley 分析方法来解释和将模型进行可视化,具有可解释性的机器学习算法可以帮助临床医师优化对肺结节的诊断及管理,可作为术前评估肺结节浸润性的无创性方法。

参考文献:

- [1] Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2021, 71(3): 209-249.
- [2] 刘梦雯,张丽. 肺亚实性结节生长的临床影像研究进展[J]. 放射 学实践,2023,38(2):235-240.
- [3] Nicholson AG, Tsao MS, Beasley MB, et al. The 2021 WHO classification of lung tumors; impact of advances since 2015[J]. J Thorac Oncol, 2022, 17(3):362-387.
- [4] Yu Z,Xu C,Zhang Y,et al. A triple-classification for the evaluation of lung nodules manifesting as pure ground-glass sign; a CT-based radiomic analysis[J/OL]. BMC Med Imaging, 2022, 22(1); e133[2022 Jul 27]. DOI:10.1186/s12880-022-00862-x.

- [5] Yoshida M, Yuasa M, Ogawa H, et al. Can computed tomography differentiate adenocarcinoma in situ from minimally invasive adenocarcinoma[J]. Thorac Cancer, 2021, 12(7):1023-1032.
- [6] Xing X, Yang F, Huang Q, et al. Decoding the multicellular ecosystem of lung adenocarcinoma manifested as pulmonary subsolid nodules by single-cell RNA sequencing [J/OL]. Sci Adv, 2021, 7 (5):e9738 [2021 Jan 29]. DOI:10.1126/sciadv. abd9738.
- [7] Nishino M. Perinodular radiomic features to assess nodule microenvironment; does it help to distinguish malignant versus benign lung nodules[J]. Radiology, 2019, 290(3); 793-795.
- [8] Wang T,She Y,Yang Y,et al. Radiomics for survival risk stratification of clinical and pathologic stage IA pure-solid non-small cell lung cancer[J]. Radiology, 2022, 302(2):425-434.
- [9] Altorki NK, Markowitz GJ, Gao D, et al. The lung microenvironment: an important regulator of tumour growth and metastasis [J]. Nat Rev Cancer, 2019, 19(1):9-31.
- [10] Mittal V,El RT, Narula N, et al. The microenvironment of lung cancer and therapeutic implications[J/OL]. Adv Exp Med Biol, 2016,890; e75-e110 [2015 Dec 25]. DOI: 10. 1007/978-3-319-24932-2 5.
- [11] Wu L, Gao C, Ye J, et al. The value of various peritumoral radiomic features in differentiating the invasiveness of adenocarcinoma manifesting as ground-glass nodules[J]. Eur Radiol, 2021, 31(12):9030-9037.
- [12] Beig N, Khorrami M, Alilou M, et al. Perinodular and intranodular radiomic features on lung CT images distinguish adenocarcinomas from granulomas[J]. Radiology, 2019, 290(3):783-792.
- [13] Uthoff J, Stephens MJ, Newell JJ, et al. Machine learning approach for distinguishing malignant and benign lung nodules utilizing standardized perinodular parenchymal features from CT [J]. Med Phys, 2019, 46(7): 3207-3216.
- [14] Fu BJ, Lv FJ, Li WJ, et al. Significance of intra-nodular vessel sign in differentiating benign and malignant pulmonary ground-glass nodules[J/OL]. Insights Imaging, 2021, 12(1): e65 [2021 May 26]. DOI: 10.1186/s13244-021-01012-7.
- [15] Zhan Y, Peng X, Shan F, et al. Attenuation and morphologic characteristics distinguishing a ground-glass nodule measuring 5~10mm in diameter as invasive lung adenocarcinoma on thin-slice CT[J]. AJR, 2019, 213(4):162-170.
- [16] 俞慧波,陈中港,李琼,等. 月牙征预测纯磨玻璃结节肺腺癌浸润性的价值[J]. 中华放射学杂志,2021,55(4):403-408.
- [17] Huang L, Lin W, Xie D, et al. Development and validation of a

- preoperative CT-based radiomic nomogram to predict pathology invasiveness in patients with a solitary pulmonary nodule; a machine learning approach, multicenter, diagnostic study[J]. Eur Radiol, 2022, 32(3):1983-1996.
- [18] Smayra T, Charara Z, Sleilaty G, et al. Classification and regression tree (CART) model of sonographic signs in predicting thyroid nodules malignancy [J/OL]. Eur J Radiol Open, 2019, 6 (11):e343-e349 [2019 Nov 28]. DOI:10.1016/j. ejro. 2019. 11. 003.
- [19] 赵帅,孟令思,郭君武.基于 CT 靶扫描构建多模态联合模型 No-mogram 评估孤立性肺结节 [J]. 放射学实践,2022,37(10): 1232-1237.
- [20] Giraud P, Giraud P, Nicolas E, et al. Interpretable machine learning model for locoregional relapse prediction in oropharyngeal cancers[J/OL]. Cancers (Basel), 2020, 13(1): e57 [2020 Dec 28]. DOI:10.3390/cancers13010057.
- [21] Li R, Shinde A, Liu A, et al. Machine learning-based interpretation and visualization of nonlinear interactions in prostate cancer survival[J]. JCO Clin Cancer Inform, 2020, 4(7):637-646. DOI: 10.1200/CCI. 20.00002.
- [22] Rodriguez-Perez R, Bajorath J. Interpretation of compound activity predictions from complex machine learning models using local approximations and shapley values [J]. J Med Chem, 2020, 63 (16):8761-8777.
- [23] Wang Y, Lang J, Zuo JZ, et al. The radiomic-clinical model using the SHAP method for assessing the treatment response of whole-brain radiotherapy; a multicentric study[J]. Eur Radiol, 2022, 32 (12):8737-8747.
- [24] Manikis GC, Ioannidis GS, Siakallis L, et al. Multicenter DSC-MRI-based radiomics predict IDH mutation in gliomas[J/OL]. Cancers (Basel), 2021, 13(16); e3965 [2021 Aug 5]. DOI: 10. 3390/cancers13163965.
- [25] Pan D, Liu R, Zheng B, et al. Using machine learning to unravel the value of radiographic features for the classification of bone tumors[J/OL]. Biomed Res Int, 2021, 11; 2021; e8811056[2021 Mar 11]. DOI: 10.1155/2021/8811056.
- [26] Zhang J, Jin J, Ai Y, et al. Computer tomography radiomics-based nomogram in the survival prediction for brain metastases from non-small cell lung cancer underwent whole brain radiotherapy [J/OL]. Front Oncol, 2021, 11 (10): e610691 [2021 Feb 11]. DOI: 10.3389/fonc. 2020. 610691.

(收稿日期:2023-08-15 修回日期:2024-01-22)