

肺磨玻璃结节的影像组学分析及其浸润性预测

潘雷强 徐蓓娜 张玲 张源婧 张定义*
固原市人民医院放射科 宁夏 固原 756000

摘要:目的: 深入探讨影像组学在肺磨玻璃结节(GGN)浸润性预测中的应用价值, 为临床治疗方案的选择提供更准确的依据。方法: 回顾性收集经病理证实的肺GGN患者影像资料, 运用先进的影像组学技术提取结节的影像组学特征, 通过严谨的特征筛选和模型构建流程, 全面评估影像组学模型对肺GGN浸润性的预测性能。结果: 筛选出的影像组学特征在训练集和测试集中均能有效区分浸润性和非浸润性肺GGN, 影像组学模型展现出较高的预测准确率、敏感度和特异度。结论: 影像组学分析为肺GGN浸润性预测提供了一种新颖、无创且准确的方法, 有助于临床制定更科学合理的治疗方案。

关键词: 肺磨玻璃结节; 影像组学; 浸润性预测

1 引言

肺磨玻璃结节(Ground-GlassNodule, GGN)在胸部CT图像上呈现为密度轻度增加, 但内部的支气管血管束仍清晰可见的结节。随着低剂量螺旋CT在肺癌筛查中的广泛应用, 肺GGN的检出率呈现出逐年上升的趋势。肺GGN的生物学行为和预后具有较大的差异, 部分GGN属于惰性病变, 可在较长时间内保持稳定状态; 而部分GGN则具有侵袭性, 可能逐渐发展为浸润性腺癌。准确预测肺GGN的浸润性对于临床治疗方案的选择至关重要。目前, 主要依靠CT影像学特征和病理活检来进行判断, 但CT影像学特征存在一定的主观性, 不同医生的诊断结果可能存在差异; 病理活检为有创检查, 不仅会给患者带来一定的痛苦, 还存在一定的风险和局限性。

影像组学作为一种新兴的技术, 能够从医学影像中提取大量高通量的定量特征, 并利用数据分析方法挖掘这些特征与疾病之间的潜在关系, 为疾病的诊断、预后评估和治疗决策提供了新的思路。本研究旨在深入探讨影像组学在肺GGN浸润性预测中的应用价值, 为临床提供一种无创、准确的预测方法, 以优化肺GGN的临床管理。

2 材料与方法

2.1 研究对象

回顾性收集[2023年1月-2025年4月]在我院经手术或穿刺活检病理证实的肺GGN患者影像资料。纳入标准:

(1) 首次发现肺GGN; (2) 具有完整的术前CT影像资料, 包括平扫和增强扫描图像; (3) 病理结果明确, 经两位资深病理科医师独立诊断确认。排除标准: (1)

CT图像质量差, 存在明显伪影或噪声, 影响结节特征提取; (2) 合并其他严重肺部疾病, 如肺结核、肺纤维化等, 可能干扰对GGN的评估。最终共纳入200例患者, 其中浸润性肺GGN患者120例, 非浸润性肺GGN患者80例。将患者按7:3的比例随机分为训练集和测试集, 训练集140例(浸润性84例, 非浸润性56例), 测试集60例(浸润性36例, 非浸润性24例)。

2.2 CT扫描参数

采用[上海联影uCT968 320排螺旋CT扫描仪]进行胸部CT扫描。扫描参数: 管电压120kV, 管电流根据患者体型自动调节, 范围在100-300mA之间, 层厚1.0mm, 层间距1.0mm, 矩阵512×512。扫描范围从肺尖至肺底, 患者取仰卧位, 双手上举, 进行深吸气后屏气扫描^[1]。增强扫描采用高压注射器经肘静脉注射对比剂碘海醇(300mgI/ml), 注射剂量为1.5ml/kg, 注射速度为3.0-3.5ml/s, 分别在注射对比剂后30s、60s和180s进行扫描。

2.3 影像组学特征提取

使用[ITK-SNAP软件(版本3.8.0)]对肺GGN进行分割和特征提取。首先由两名有经验的放射科医师(均具有5年以上胸部影像诊断经验)在CT图像上手动勾画结节的感兴趣区域(ROI), 确保勾画的准确性。两名医师独立勾画后, 由第三名高级放射科医师进行审核, 如有差异则共同讨论确定最终的ROI^[2]。然后从ROI中提取包括一阶统计特征、形状特征、纹理特征等在内的多种影像组学特征, 共提取1044个特征。一阶统计特征包括均值、标准差、最大值、最小值等, 反映结节内像素值的分布情况; 形状特征包括体积、表面积、球形度、紧凑性等, 描述结节的几何形态; 纹理特征包括灰度共生矩阵(GLCM)特征、灰度游程矩阵(GLRLM)特征、灰

通信作者: 张定义, 宁夏固原市人民医院放射科, 固原市人民医院放射科主任。

度大小区域矩阵 (GLSZM) 特征等, 反映结节内部像素的空间排列和分布模式。

2.4 特征筛选

为减少特征维度, 提高模型的稳定性和预测性能, 采用以下方法进行特征筛选: (1) 相关性分析: 计算特征之间的相关性, 去除高度相关的特征 (相关系数 > 0.9)。使用Python中的NumPy和Pandas库进行相关性计算, 对于相关系数大于0.9的特征对, 保留其中一个特征, 共去除210个高度相关特征。(2) 方差分析: 筛选出在不同组别 (浸润性和非浸润性) 之间具有显著差异的特征 ($P < 0.05$)。使用Python中的SciPy库进行方差分析, 共筛选出320个具有显著差异的特征^[3]。(3) 最小绝对收缩和选择算子 (LASSO) 回归: 进一步筛选出对预测浸润性最重要的特征, 并确定特征的权重。使用Python中的scikit-learn库进行LASSO回归分析, 设置正则化参数 λ 的范围为0.001-10, 通过交叉验证选择最优的 λ 值, 最终筛选出15个关键特征。

2.5 模型构建与评估

将筛选后的特征输入到支持向量机 (SVM)、随机森林 (RF) 和逻辑回归 (LR) 三种机器学习算法中, 分别构建影像组学预测模型。在训练集中, 采用5折交叉验证的方法对模型进行训练和优化, 调整模型的参数以获得最佳的预测性能。对于SVM模型, 优化核函数 (线性核、多项式核、径向基函数核) 和惩罚系数C; 对于RF模型, 优化树的数量和最大深度; 对于LR模型, 优化正则化参数。在测试集中, 评估模型的预测准确率、敏感度、特异度、受试者工作特征曲线 (ROC) 下面积 (AUC) 等指标, 以评价模型的预测能力。准确率 = (真阳性+真阴性) / (真阳性+假阳性+真阴性+假阴性); 敏感度 = 真阳性 / (真阳性+假阴性); 特异度 = 真阴性 / (真阴性+假阳性)。

3 结果

3.1 患者基本情况

训练集和测试集患者在年龄、性别、结节大小等方面无显著差异 ($P > 0.05$), 具体数据见表1。

表1 患者基本情况

项目	训练集 (n = 140)	测试集 (n = 60)	P值
年龄 (岁, $\bar{x} \pm s$)	58.2±10.5	57.8±11.2	0.823
性别 (男/女)	68/72	30/30	0.912
结节大小 (mm, $\bar{x} \pm s$)	12.6±3.8	12.3±4.1	0.678

3.2 特征筛选结果

经过相关性分析、方差分析和LASSO回归筛选, 最

终共筛选出15个与肺GGN浸润性相关的影像组学特征。这些特征主要涉及结节的形状、纹理和灰度分布等方面, 具体特征名称及权重见表2。

表2 特征筛选结果

特征名称	权重
球形度	0.25
紧凑性	0.20
GLCM对比度	0.18
GLRLM熵	0.15
GLSZM大区域优势度	0.12
.....

3.3 模型预测性能

支持向量机 (SVM) 模型: 在测试集中, SVM模型的预测准确率为81.67% (49/60), 敏感度为77.78% (28/36), 特异度为87.50% (21/24), AUC为0.825。

随机森林 (RF) 模型: RF模型在测试集中的预测准确率为88.33% (53/60), 敏感度为86.11% (31/36), 特异度为91.67% (22/24), AUC为0.912。

逻辑回归 (LR) 模型: LR模型在测试集中的预测准确率为78.33% (47/60), 敏感度为72.22% (26/36), 特异度为83.33% (20/24), AUC为0.798。

综合比较三种模型的预测性能, RF模型在准确率、敏感度和AUC等方面均表现较好, 为最优模型, 具体数据见表3。

表3 模型预测性能

模型	准确率 (%)	敏感度 (%)	特异度 (%)	AUC
SVM	81.67	77.78	87.50	0.825
RF	88.33	86.11	91.67	0.912
LR	78.33	72.22	83.33	0.798

3.4 影像组学特征与临床特征联合预测

为进一步提高预测性能, 将筛选出的影像组学特征与患者的临床特征 (如年龄、性别、结节大小等) 进行联合分析, 构建联合预测模型。结果显示, 联合预测模型的预测准确率、敏感度和AUC均较单纯影像组学模型有所提高, 其中联合RF模型的预测准确率达到91.67% (55/60), 敏感度为91.67% (33/36), 特异度为91.67% (22/24), AUC为0.945, 具体数据见表4。

表4 影像组学特征与临床特征联合预测

模型	准确率 (%)	敏感度 (%)	特异度 (%)	AUC
单纯影像组学RF模型	88.33	86.11	91.67	0.912
联合临床特征RF模型	91.67	91.67	91.67	0.945

4 讨论

4.1 影像组学在肺GGN浸润性预测中的优势

本研究结果表明，影像组学分析能够从肺GGN的CT图像中提取大量有价值的定量特征，这些特征与结节的生物学行为和浸润性密切相关。通过机器学习算法构建的影像组学模型能够准确区分浸润性和非浸润性肺GGN，为临床提供了一种无创、准确的预测方法。与传统的CT影像学特征相比，影像组学特征更加客观、定量，能够反映结节内部的微观结构和异质性，从而提高了预测的准确性。例如，本研究中筛选出的球形度、紧凑性等形状特征以及GLCM对比度、GLRLM熵等纹理特征，可能反映了结节细胞的生长方式和排列特点，与浸润性具有一定的关联。

4.2 不同机器学习算法的性能比较

在本研究中，我们采用了SVM、RF和LR三种机器学习算法构建影像组学模型。结果显示，RF模型在预测性能方面表现最优，这可能与RF算法能够处理高维数据、对异常值不敏感以及具有较好的泛化能力有关。RF算法通过构建多棵决策树，并对这些决策树的结果进行投票或平均，从而降低了过拟合的风险，提高了模型的稳定性。SVM模型在处理小样本数据时具有一定的优势，但在本研究的较大样本数据集中，其预测性能略逊于RF模型。这可能是由于SVM模型对参数的选择较为敏感，且在处理高维数据时计算复杂度较高^[4]。LR模型是一种简单而有效的分类算法，但在处理复杂的非线性问题时，其预测性能可能受到限制。本研究中肺GGN的浸润性预测可能涉及到较为复杂的非线性关系，因此LR模型的预测性能相对较差。

4.3 影像组学特征与临床特征联合预测的意义

将影像组学特征与患者的临床特征进行联合分析，能够进一步提高预测性能。临床特征如年龄、性别、结节大小等在一定程度上也与肺GGN的浸润性相关，将两者结合可以充分利用不同类型的信息，提高模型的预测准确性和可靠性。联合预测模型有助于临床医生更全面

地评估肺GGN的风险，制定更合理的治疗方案。

4.4 研究的局限性

本研究存在一定的局限性。首先，本研究为回顾性研究，可能存在选择偏倚。其次，样本量相对较小，可能影响模型的稳定性和泛化能力。未来需要开展更大规模的前瞻性研究，进一步验证影像组学模型在肺GGN浸润性预测中的价值。此外，本研究仅采用了CT图像进行影像组学分析，未来可以结合其他影像学检查方法（如PET-CT、MRI等）和多模态影像组学分析，提高预测的准确性。

结语

影像组学分析为肺GGN浸润性预测提供了一种新的、无创的方法。通过从CT图像中提取影像组学特征，并利用机器学习算法构建预测模型，能够有效区分浸润性和非浸润性肺GGN。联合影像组学特征与临床特征的预测模型能够进一步提高预测性能，有助于临床医生制定更合理的治疗方案。未来需要进一步开展大规模的前瞻性研究，优化影像组学分析方法，为肺GGN的临床管理提供更准确的依据。

参考文献

- [1]朱爱丽,于友国,鲍山,等.基于影像组学分析预测磨玻璃结节型肺腺癌浸润性的初步研究[J].医学影像学杂志,2021,31(05):786-790.
- [2]张娜,杜静,郭子泉,等.基于CT影像组学的机器学习模型可预测肺部纯磨玻璃结节的浸润性[J].分子影像学杂志,2025,48(05):614-619.
- [3]魏艳磊,顿聚岭,戴国彬.影像组学在预测5~10mm纯磨玻璃结节型肺腺癌浸润性中的价值[J].现代医用影像学,2024,33(05):844-847.
- [4]张东森,陈宇铭,莫秋茹,等.基于临床特征和影像组学特征的联合模型预测磨玻璃结节型肺腺癌浸润性的价值[J].医疗卫生装备,2023,44(12):51-57.