

CT特征联合临床因素的机器学习模型预测部分实性结节肺腺癌气腔播散的价值

崔维征^{1*}, 刘润所¹, 张伟伟¹, 于美艳¹, 于晓杰¹, 林吉征²

¹乳山市人民医院放射科, 山东 威海

²青岛大学附属医院放射科, 山东 青岛

收稿日期: 2025年12月7日; 录用日期: 2026年1月1日; 发布日期: 2026年1月9日

摘要

目的: 探究CT特征联合临床因素的机器学习模型预测部分实性结节肺腺癌气腔播散(STAS)的价值。方法: 回顾性收集152例经病理证实的部分实性结节肺腺癌患者, 将其随机分为训练组(106例)及验证组(46例)。采用单因素及多因素logistic回归分析临床资料、CT特征, 确定STAS状态的独立预测因素。采用逻辑回归、多层感知器、随机森林(RF)及朴素贝叶斯算法建立机器学习模型, 受试者工作特征曲线下面积(AUC)评估模型的预测效能。结果: 单因素及多因素logistic回归分析显示, 癌胚抗原、肿瘤最大径、cT分期、毛刺征是STAS状态的独立预测因素。以上述变量构建的机器学习模型中, RF模型展现了良好的预测效能, 在训练组及验证组AUC分别为0.920和0.859。结论: CT特征联合临床因素的机器学习模型对部分实性结节肺腺癌STAS具有较好的预测价值。

关键词

肺腺癌, 气腔播散, 体层摄影术, X线计算机, 机器学习

Value of the Machine Learning Model That Combines CT Features with Clinical Factors in Predicting Spread through Air Space in Partially Solid Nodules of Lung Adenocarcinoma

Weizheng Cui^{1*}, Runsuo Liu¹, Weiwei Zhang¹, Meiyang Yu¹, Xiaojie Yu¹, Jizheng Lin²

¹Department of Radiology, The People's Hospital of Rushan, Weihai Shandong

²Department of Radiology, The Affiliated Hospital of Qingdao University, Qingdao Shandong

*通讯作者。

文章引用: 崔维征, 刘润所, 张伟伟, 于美艳, 于晓杰, 林吉征. CT特征联合临床因素的机器学习模型预测部分实性结节肺腺癌气腔播散的价值[J]. 临床医学进展, 2026, 16(1): 818-826. DOI: 10.12677/acm.2026.161109

Abstract

Objective: To investigate the value of the machine learning model incorporating CT features and clinical factors in predicting spread through air spaces (STAS) in lung adenocarcinoma presenting as part-solid nodules. **Methods:** A total of 152 patients with pathologically confirmed lung adenocarcinoma manifesting as part-solid nodules were retrospectively enrolled and randomly divided into a training cohort ($n = 106$) and a validation cohort ($n = 46$). Univariate and multivariate logistic regression analyses were performed on clinical data and CT features to identify independent predictors of STAS status. Machine learning models were constructed using logistic regression, multilayer perceptron, random forest (RF), and naive bayes algorithms. The predictive performance of each model was evaluated using the area under the receiver operating characteristic curve (AUC). **Results:** Univariate and multivariate logistic regression analysis identified carcinoembryonic antigen level, maximum tumor diameter, cT stage, and spiculation as independent predictors of STAS status. Among the machine learning models built with these variables, RF model demonstrated favorable predictive performance, with AUCs of 0.920 in the training cohort and 0.859 in the validation cohort. **Conclusion:** The machine learning model combining CT features and clinical factors show good predictive value for STAS in lung adenocarcinoma presenting as part-solid nodules.

Keywords

Lung Adenocarcinoma, Spread through Air Spaces, Tomography, X-Ray Computed, Machine Learning

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. 前言

随着低剂量计算机断层扫描(computed tomography, CT)的普及,早期肺癌的检出率明显提高,但肺癌仍是癌症相关死亡的主要原因[1]。腺癌是肺癌的主要组织学类型[2]。气腔播散(spread through air spaces, STAS)是肺癌一种新的侵袭模式,其特征为肿瘤细胞超出肿瘤边缘在邻近肺泡腔内扩散[3]。研究表明,STAS 不仅与肺癌淋巴血管浸润、高级别组织学类型相关,还是影响患者术后预后的重要危险因素[4][5]。对于 STAS 阳性肺腺癌,接受亚肺叶切除的患者比接受肺叶切除的患者复发风险明显增加、生存率显著降低[6][7]。亚实性结节是早期肺腺癌的常见表现类型,纯磨玻璃结节几乎无 STAS,而部分实性结节可出现 STAS [8]。STAS 主要依靠术后病理诊断,术前准确预测 STAS 状态将有助于指导临床选择适宜的手术方式。既往研究发现一些影像学特征与 STAS 存在相关性[9][10]。本研究旨在通过分析部分实性结节肺腺癌患者的临床及影像学资料,建立机器学习模型以术前识别 STAS 状态。

2. 资料与方法

2.1 研究对象

本研究回顾性收集 2021 年 2 月至 2023 年 6 月于青岛大学附属医院手术切除的 I 期肺腺癌患者。纳入标准:(1) 术前胸部 CT 图像表现为部分实性结节;(2) 经病理证实为浸润性肺腺癌,且病灶最大直径

≤3 cm; (3) 病理资料完整, 能满足评估 STAS 要求。排除标准: (1) 术前接受任何抗肿瘤治疗; (2) 临床资料不完整; (3) 图像质量不佳, 影响对肿瘤的评估。最终纳入 152 例患者, 其中女 90 例, 男 62 例, 年龄 29~81 岁, 平均(58.4 ± 10.4)岁, 按照 7:3 比例将患者随机分为训练组(106 例)和验证组(46 例)。

2.2. CT 方法

采用 Siemens Somatom Sensation Cardiac 64 及 GE Discover 750 HD 设备进行胸部 CT 平扫。扫描范围从肺尖至肺底, 扫描参数: 管电压 120 kV, 自动管电流, 层厚 5 mm, 重建层厚 0.625~1.25 mm。

2.3. 临床资料收集及图像分析

收集患者的临床资料, 包括年龄、性别、吸烟史、cT 分期、癌胚抗原(carcinoembryonic antigen level, CEA)水平。由 2 名具有 7 年以上诊断经验的影像科医师分别评估病灶 CT 特征, 包括结节位置、边界、肿瘤最大径、实性成分最大径、分叶征、毛刺征、胸膜凹陷征、空泡征, 并计算实性成分比例 (consolidation/tumor ratio, CTR), CTR = 实性成分最大径/肿瘤最大径。

2.4. 统计分析

采用 R 软件(v4.3.3)及 Python 软件(v3.9.7)进行统计分析。Kolmogorov-Smirnov 检验用于评估连续性数据的正态性。符合正态分布的数据以 $\bar{x} \pm s$ 表示, 否则以中位数(上下四分位数)表示。采用独立样本 *t* 检验或 Mann-Whitney *U* 检验分析计量资料, χ^2 检验分析定性资料。单因素 logistic 回归分析用于筛选与 SATS 状态相关的变量, 将差异有统计学意义的变量纳入多因素 logistic 回归分析, 向后逐步选择独立预测因素。使用逻辑回归(logistic regression, LR)、多层感知器(multilayer perceptron, MLP)、随机森林(random forest, RF)及朴素贝叶斯(naive bayes, NB)算法建立预测肺腺癌 STAS 状态的机器学习模型, 以受试者工作特征曲线下面积(area under the curve, AUC)评估模型的预测效能。10 折交叉验证用于训练各机器学习算法。采用校准曲线判断模型的拟合优度, 决策曲线分析显示模型的临床收益。*P* < 0.05 为差异有统计学意义。

3. 结果

3.1. 临床资料、CT 特征比较

训练组中 STAS 阳性 49 例, 阴性 57 例; 验证组中 STAS 阳性 20 例, 阴性 26 例。胸膜凹陷征在训练组和验证组间差异有统计学意义(*P* = 0.004), 其余临床及 CT 特征组间差异均无统计学意义(*P* > 0.05) (表 1)。

Table 1. Comparison of clinical data and CT features between the training group and the validation group
表 1. 训练组和验证组临床资料、CT 特征比较

	训练组(n = 106)	验证组(n = 46)	统计值	<i>P</i> 值
性别			0.075	0.784
女	62	28		
男	44	18		
年龄	58.37 ± 10.85	58.59 ± 9.31	0.119	0.905
吸烟史			0.091	0.763
无	83	35		
有	23	11		

续表

STAS			0.098	0.755
无	57	26		
有	49	20		
CEA			0.188	0.664
正常	84	35		
升高	22	11		
位置			1.383	0.847
右肺上叶	35	13		
右肺中叶	6	4		
右肺下叶	17	10		
左肺上叶	23	9		
左肺下叶	25	10		
肿瘤最大径	17.00 (15.00, 23.75)	17.00 (14.00, 23.75)	-0.040	0.968
实性成分最大径	11.00 (7.25, 13.00)	11.00 (8.00, 16.00)	-0.803	0.422
CTR	0.60 ± 0.18	0.65 ± 0.21	1.422	0.157
cT 分期			0.809	0.667
T1a	49	20		
T1b	50	21		
T1c	7	5		
边界			0.002	0.967
不清	48	21		
清	58	25		
分叶征			2.243	0.134
无	13	10		
有	93	36		
毛刺征			0.148	0.700
无	38	18		
有	68	28		
胸膜凹陷征			8.116	0.004
无	35	5		
有	71	41		
空泡征			2.646	0.104
无	47	27		
有	59	19		

注：STAS 为气腔播散；CEA 为癌胚抗原；CTR 为实性成分比例。

3.2 临床资料、CT 特征单因素及多因素 logistic 回归分析

单因素分析显示，CEA、肿瘤最大径、实性成分最大径、CTR、cT 分期、边界、分叶征、毛刺征与 STAS 表达状态显著相关(P 均 <0.05)，将上述特征纳入多因素 logistic 回归分析，结果表明 CEA、肿瘤最大径、cT 分期以及毛刺征是 STAS 阳性的独立预测因素(P 均 >0.05) (表 2)。图 1 展示了 1 例 STAS 阳性患者的 CT 图像。

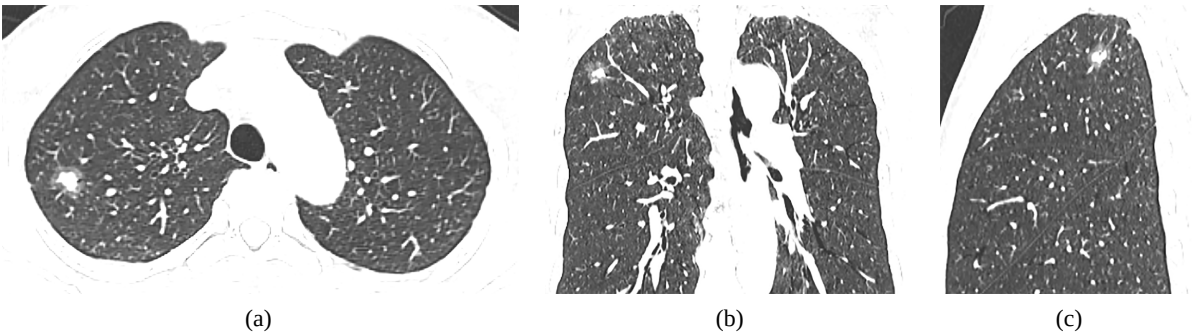


Figure 1. A 57-year-old male patient with STAS-positive lung adenocarcinoma. Axial (a), coronal (b), and sagittal (c) CT images show a part-solid nodule in the right upper lobe, with a well-defined boundary, containing a solid component and peripheral lobulation

图 1. STAS 阳性肺腺癌患者，男，57 岁。横轴位(a)、冠状位(b)、及矢状位(c) CT 图像示右肺上叶部分实性结节，边界清晰，内可见实性成分，周边可见分叶征

Table 2. Univariate and multivariate logistic regression analyses of clinical data and CT features

表 2. 临床资料、CT 特征单因素及多因素 logistic 回归分析

	单因素分析		多因素分析	
	OR (95%CI)	P 值	OR (95%CI)	P 值
性别	1.776 (0.813~3.877)	0.149		
年龄	1.035 (0.997~1.074)	0.070		
吸烟史	1.356 (0.537~3.422)	0.519		
CEA	7.694 (2.387~24.802)	<0.001	5.182 (1.158~23.189)	0.031
位置	0.933 (0.734~1.187)	0.573		
肿瘤最大径	1.276 (1.159~1.405)	<0.001	1.168 (1.029~1.326)	0.017
实性成分最大径	1.371 (1.201~1.566)	<0.001		
CTR	16.882 (1.596~178.584)	0.019		
cT 分期	14.981 (5.677~39.539)	<0.001	4.650 (1.291~16.744)	0.019
边界	2.244 (1.023~4.925)	0.044		
分叶征	12.800 (1.599~102.471)	0.016		
毛刺征	9.173 (3.369~24.977)	<0.001	10.045 (2.476~40.757)	0.001
胸膜凹陷征	1.225 (0.542~2.769)	0.625		
空泡征	1.525 (0.703~3.307)	0.286		

注：CEA 为癌胚抗原；CTR 为实性成分比例。

3.3. 模型构建与评估

基于 STAS 阳性的独立预测因素(CEA、肿瘤最大径、cT 分期、毛刺征)，使用 LR、RF、MLP 及 NB 算法构建机器学习模型。RF 模型在训练组与验证组中预测 STAS 的 AUC 值最高，分别为 0.920 (95%CI: 0.868~0.973)和 0.859 (95%: 0.733~0.985)(表 3)。校准曲线表明 RF 模型预测结果与实际结果具有较好的一致性(图 2)，DCA 显示 RF 模型在广泛的阈值范围内展现出更高的净收益(图 3)。因此，RF 模型是预测肺腺癌 STAS 状态的最佳机器学习模型。

Table 3. The efficiency of different machine learning models in predicting STAS status of lung adenocarcinoma
表 3. 不同机器学习模型预测肺腺癌 STAS 状态的效能

模型	训练组			验证组		
	AUC (95%CI)	敏感度	特异度	AUC (95%CI)	敏感度	特异度
LR	0.909 (0.853~0.965)	0.878	0.807	0.829 (0.698~0.960)	0.750	0.846
RF	0.920 (0.868~0.973)	0.837	0.877	0.859 (0.733~0.985)	0.700	0.962
MLP	0.904 (0.844~0.964)	0.816	0.895	0.842 (0.712~0.973)	0.750	0.885
NB	0.907 (0.851~0.963)	0.816	0.860	0.783 (0.641~0.924)	0.700	0.808

注：LR，逻辑回归；RF，随机森林；MLP，多层感知器；NB，朴素贝叶斯。

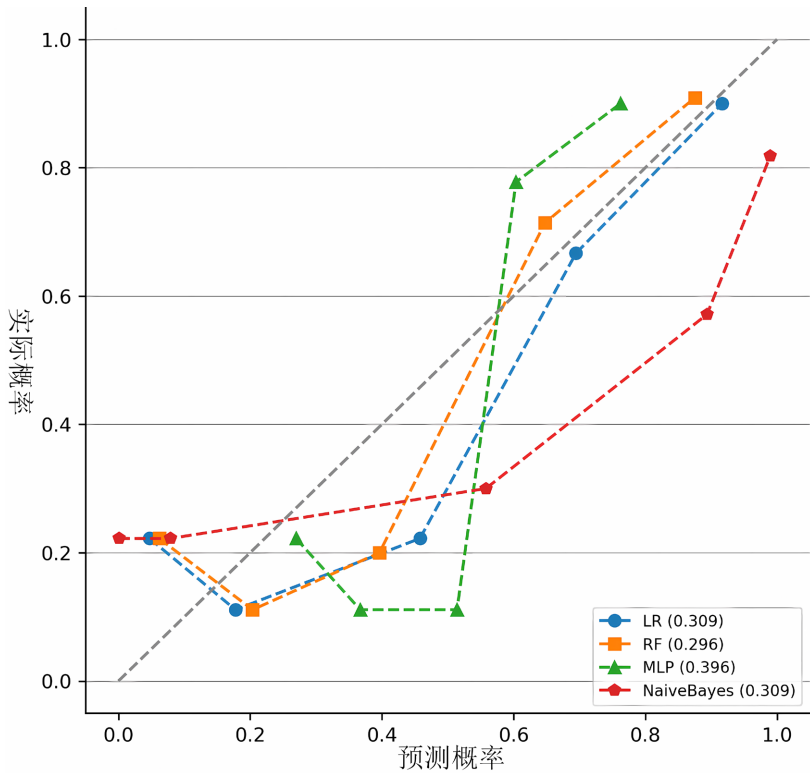


Figure 2. Calibration curves of different machine learning models in the validation group. LR, logistic regression; RF, random forest; MLP, multilayer perceptron
图 2. 验证组中不同机器学习模型的校准曲线。LR，逻辑回归；RF，随机森林；MLP，多层感知器

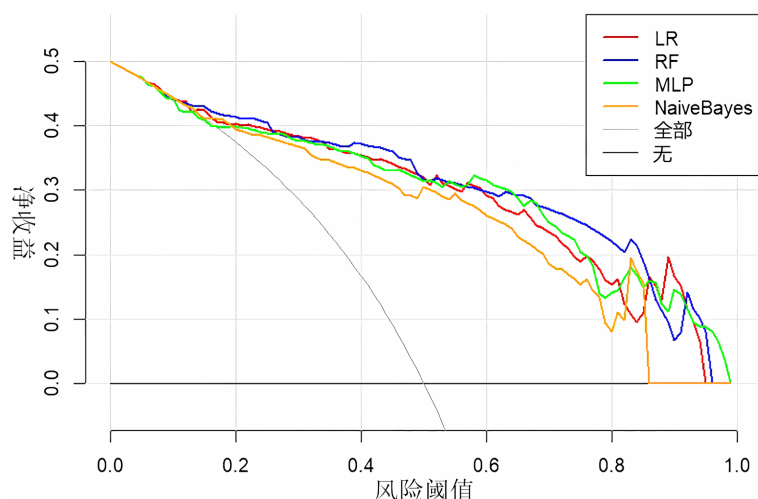


Figure 3. Decision curve analysis of different machine learning models. LR, logistic regression; RF, random forest; MLP, multilayer perceptron

图 3. 不同机器学习模型的决策曲线分析。LR, 逻辑回归; RF, 随机森林; MLP, 多层感知器

4. 讨论

STAS 作为肺癌的一种侵袭模式, 与患者不良预后显著相关, 术前准确评估 STAS 状态有助于临床选择最佳治疗方案, 改善患者预后。本研究基于 CT 特征及临床因素建立了 4 种机器学习模型以预测部分实性结节肺腺癌 STAS 状态, 其中 RF 模型预测效能最佳, 有利于临床术前准确诊断 STAS 状态。

既往研究表明, STAS 可见于 14.8%~56.4%肺腺癌患者[11], 本研究中 STAS 发生率为 45.4%, 在文献报道范围内。部分实性结节 STAS 发生率高, 恶性程度大, 预后较差, 需要重点关注。多项研究表明[10][12], STAS 与肿瘤实性成分相关, 本研究中单因素分析结果显示 STAS 阳性与实性成分最大径、CTR 显著相关, 与既往研究结果相符。本研究通过多因素 logistic 回归分析, 确定了 4 个部分实性结节肺腺癌 STAS 状态的独立危险因素: CEA、肿瘤最大径、cT 分期及毛刺征。CEA 是细胞粘附分子家族中的一种糖蛋白, 具有一定的粘附能力。然而, CEA 过度表达会破坏其他粘附分子的功能, 降低细胞粘附, 促使肿瘤细胞从原发部位脱落, 这是形成 STAS 的潜在因素[13][14]。本研究显示 STAS 阳性与肿瘤直径、T 分期显著相关, 与既往文献报道一致[9][15], 提示 STAS 阳性肿瘤更具侵袭性。毛刺征是肺肿瘤的典型恶性征象之一, 其主要形成因素是肿瘤沿肺间质、血管向外浸润性生长[16], 本研究中毛刺征是肺腺癌 STAS 状态的独立预测因子, 有助于评估 STAS 状态。

机器学习作为人工智能的重要分支, 已被广泛用于医疗领域, 其在肿瘤诊断、分类及预后等方面具有重要价值。既往已有多项研究建立预测模型诊断肺腺癌 STAS。徐凤等[17]联合临床及 CT 特征的 logistic 回归分析建立预测模型, 其 AUC 达 0.807, 但该研究未设立验证集进一步评估模型性能。同样, Li 等[18]建立基于 CT 的 logistic 回归模型预测肺腺癌 STAS, 其在验证集和外部测试集 AUC 分别为 0.801 和 0.692。另外, 江长思等[19]将年龄及 12 个 CT 征象纳入 RF 算法构建机器学习模型, 并取得了较好的预测效能(验证集 AUC 为 0.77), 表明机器学习可用于术前预测肺腺癌 STAS 状态。本研究利用临床因素及 CT 特征建立的 4 个机器学习模型中, RF 模型表现出良好的诊断效能(AUC = 0.859), 优于上述研究结果。RF 是一种使用多棵决策树来识别、分类和预测目标数据的机器学习算法, 可以提供更高的分类准确性[20]。

本研究尚存在局限性: 首先, 这是一项单中心回顾性研究, 可能存在选择偏倚; 其次, 样本量较小, 未来需扩大样本量, 进一步验证模型性能; 最后, 增强 CT 可能会提供更多信息, 后续有必要增加对增强

CT 图像的分析。

5. 结论

综上所述, CT 特征联合临床因素的机器学习模型可以有效识别部分实性结节肺腺癌 STAS 状态。

利益冲突

所有作者均声明不存在利益冲突。

声 明

本研究获得青岛大学附属医院伦理委员会批准(审批号: QYFY WZLL 29455)。

参考文献

- [1] Siegel, R.L., Giaquinto, A.N. and Jemal, A. (2024) Cancer Statistics, 2024. *CA: A Cancer Journal for Clinicians*, **74**, 12-49. <https://doi.org/10.3322/caac.21820>
- [2] Zhang, Y., Vaccarella, S., Morgan, E., Li, M., Etxeberria, J., Chokunonga, E., *et al.* (2023) Global Variations in Lung Cancer Incidence by Histological Subtype in 2020: A Population-Based Study. *The Lancet Oncology*, **24**, 1206-1218. [https://doi.org/10.1016/s1470-2045\(23\)00444-8](https://doi.org/10.1016/s1470-2045(23)00444-8)
- [3] Travis, W.D., Brambilla, E., Nicholson, A.G., Yatabe, Y., Austin, J.H.M., Beasley, M.B., *et al.* (2015) The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances since the 2004 Classification. *Journal of Thoracic Oncology*, **10**, 1243-1260. <https://doi.org/10.1097/jto.0000000000000630>
- [4] Yokoyama, S., Murakami, T., Tao, H., Onoda, H., Hara, A., Miyazaki, R., *et al.* (2018) Tumor Spread through Air Spaces Identifies a Distinct Subgroup with Poor Prognosis in Surgically Resected Lung Pleomorphic Carcinoma. *Chest*, **154**, 838-847. <https://doi.org/10.1016/j.chest.2018.06.007>
- [5] Kadota, K., Nitadori, J., Sima, C.S., Ujiie, H., Rizk, N.P., Jones, D.R., *et al.* (2015) Tumor Spread through Air Spaces Is an Important Pattern of Invasion and Impacts the Frequency and Location of Recurrences after Limited Resection for Small Stage I Lung Adenocarcinomas. *Journal of Thoracic Oncology*, **10**, 806-814. <https://doi.org/10.1097/jto.0000000000000486>
- [6] Eguchi, T., Kameda, K., Lu, S., Bott, M.J., Tan, K.S., Montecalvo, J., *et al.* (2019) Lobectomy Is Associated with Better Outcomes than Sublobar Resection in Spread through Air Spaces (STAS)-Positive T1 Lung Adenocarcinoma: A Propensity Score-Matched Analysis. *Journal of Thoracic Oncology*, **14**, 87-98. <https://doi.org/10.1016/j.jtho.2018.09.005>
- [7] Chae, M., Jeon, J.H., Chung, J., Lee, S.Y., Hwang, W.J., Jung, W., *et al.* (2021) Prognostic Significance of Tumor Spread through Air Spaces in Patients with Stage IA Part-Solid Lung Adenocarcinoma after Sublobar Resection. *Lung Cancer*, **152**, 21-26. <https://doi.org/10.1016/j.lungcan.2020.12.001>
- [8] Kim, S.K., Kim, T.J., Chung, M.J., Kim, T.S., Lee, K.S., Zo, J.I., *et al.* (2018) Lung Adenocarcinoma: CT Features Associated with Spread through Air Spaces. *Radiology*, **289**, 831-840. <https://doi.org/10.1148/radiol.2018180431>
- [9] Qin, L., Sun, Y., Zhu, R., Hu, B. and Wu, J. (2022) Clinicopathological and CT Features of Tumor Spread through Air Space in Invasive Lung Adenocarcinoma. *Frontiers in Oncology*, **12**, Article 959113. <https://doi.org/10.3389/fonc.2022.959113>
- [10] Qi, L., Xue, K., Cai, Y., Lu, J., Li, X. and Li, M. (2021) Predictors of CT Morphologic Features to Identify Spread through Air Spaces Preoperatively in Small-Sized Lung Adenocarcinoma. *Frontiers in Oncology*, **10**, Article 548430. <https://doi.org/10.3389/fonc.2020.548430>
- [11] Jiang, C., Luo, Y., Yuan, J., You, S., Chen, Z., Wu, M., *et al.* (2020) CT-Based Radiomics and Machine Learning to Predict Spread through Air Space in Lung Adenocarcinoma. *European Radiology*, **30**, 4050-4057. <https://doi.org/10.1007/s00330-020-06694-z>
- [12] 郭炜, 和清源, 王晓华, 等. CT 定量特征预测部分实性结节肺癌气腔播散的价值[J]. 国际医学放射学杂志, 2025, 48(1): 6-9, 69.
- [13] Zhang, Z., Zhao, Y., Ma, Y., Chen, C., Li, Z., Wang, Y., *et al.* (2025) Prediction of STAS in Lung Adenocarcinoma with Nodules ≤ 2 cm Using Machine Learning: A Multicenter Retrospective Study. *BMC Cancer*, **25**, Article No. 417. <https://doi.org/10.1186/s12885-025-13783-z>
- [14] Beauchemin, N. and Arabzadeh, A. (2013) Carcinoembryonic Antigen-Related Cell Adhesion Molecules (CEACAMs)

- in Cancer Progression and Metastasis. *Cancer and Metastasis Reviews*, **32**, 643-671.
<https://doi.org/10.1007/s10555-013-9444-6>
- [15] Liao, G., Huang, L., Wu, S., Zhang, P., Xie, D., Yao, L., *et al.* (2022) Preoperative CT-Based Peritumoral and Tumoral Radiomic Features Prediction for Tumor Spread through Air Spaces in Clinical Stage I Lung Adenocarcinoma. *Lung Cancer*, **163**, 87-95. <https://doi.org/10.1016/j.lungcan.2021.11.017>
- [16] 范丽, 望云, 周秀秀, 等. 孤立性肺结节的影像诊断思路及处理策略[J]. 中华放射学杂志, 2023(2): 235-238.
- [17] 徐凤, 李琦, 李娴, 等. 早期浸润性肺腺癌气腔播散的临床、病理及 CT 特征分析[J]. 放射学实践, 2025, 40(7): 852-859.
- [18] Li, C., Jiang, C., Gong, J., Wu, X., Luo, Y. and Sun, G. (2020) A CT-Based Logistic Regression Model to Predict Spread through Air Space in Lung Adenocarcinoma. *Quantitative Imaging in Medicine and Surgery*, **10**, 1984-1993.
<https://doi.org/10.21037/qims-20-724>
- [19] 江长思, 罗燕, 唐雪, 等. 基于 CT 机器学习模型预测肺腺癌气腔播散[J]. 中国医学影像技术, 2020, 36(12): 1834-1838.
- [20] Jin, Y., Lan, A., Dai, Y., Jiang, L. and Liu, S. (2023) Development and Testing of a Random Forest-Based Machine Learning Model for Predicting Events among Breast Cancer Patients with a Poor Response to Neoadjuvant Chemotherapy. *European Journal of Medical Research*, **28**, Article No. 394.
<https://doi.org/10.1186/s40001-023-01361-7>