



· 论 著 ·

基于多种机器学习算法的MRI影像组学模型预测直肠癌微卫星不稳定状态

李 陆, 卜 文, 孙巧玉, 王 伟, 张玉文, 姜海东, 陈艾琪, 沈俊杰

蚌埠医科大学第一附属医院放射科, 安徽 蚌埠 233000

[摘要] 目的: 探究联合临床危险因素、多参数磁共振成像 (magnetic resonance imaging, MRI) 影像组学特征构建的多种影像组学模型预测直肠癌微卫星不稳定 (microsatellite instability, MSI) 状态的价值。方法: 纳入2020年12月—2023年11月蚌埠医科大学第一附属医院149例直肠癌患者, 其中34例为MSI, 115例为微卫星稳定 (microsatellite stability, MSS)。基于MRI检查的多序列图像勾画感兴趣区, 提取组学特征降维筛选出最佳特征, 然后使用逻辑回归 (logistic regression, LR)、随机森林 (random forest, RF)、支持向量机 (support vector machine, SVM)、K最近邻 (k-nearest neighbor, KNN), 朴素贝叶斯 (naive Bayes tree, NBT) 5种不同的机器学习算法将最优组学特征用于构建不同的影像组学模型, 并绘制出受试者工作特征 (receiver operating characteristic, ROC) 曲线评估不同模型的诊断效能。结果: RF模型表现最稳定, 且基于临床独立危险因素及影像组学构建的临床-影像组学联合模型列线图展现出对直肠癌MSI较高的诊断效能, 训练组和验证组曲线下面积 (area under curve, AUC) 分别为0.923、0.914, 评估直肠癌MSI效果最为显著。结论: 结合不同机器学习算法, 由临床危险因素和多参数MRI影像组学特征构建的临床-影像组学列线图可以有效地预测术前直肠癌MSI状态。

[关键词] 直肠癌; 微卫星不稳定性; 磁共振成像; 影像组学; 列线图

中图分类号: R735.3+7; R445.2 文献标志码: A DOI: 10.19732/j.cnki.2096-6210.2024.06.002

Prediction of microsatellite instability in rectal cancer using MRI radiomics models based on multiple machine learning algorithms LI Lu, BU Wen, SUN Qiaoyu, WANG Wei, ZHANG Yuwen, JIANG Haidong, CHEN Aiqi, SHEN Junjie

(Department of Radiology, The First Affiliated Hospital of Bengbu Medical University, Bengbu 233000, Anhui Province, China)

Correspondence to: SHEN Junjie E-mail: 395932108@qq.com

[Abstract] **Objective:** To investigate the value of multiple radiomics models constructed by combining clinical risk factors and multiparametric magnetic resonance imaging (MRI) radiomics features to predict microsatellite instability (MSI) in rectal cancer. **Methods:** A total of 149 rectal cancer patients were included in the First Affiliated Hospital of Bengbu Medical University from December 2020 to November 2023, including 34 patients with MSI and 115 patients with microsatellite stability (MSS). Based on MRI examination of multiple sequence images, 3D regions of interest were delineated, radiomics features were extracted and dimensionality was reduced to select the best features. Then, five different machine learning algorithms, including logistic regression (LR), random forest (RF), support vector machine (SVM), k-nearest neighbor (KNN), and naive Bayes tree (NBT) were used to construct different imaging radiomics models using the optimal radiomics features. And receiver operating characteristic (ROC) curves were drawn to evaluate the diagnostic performance of different models. **Results:** The RF model showed the most stable performance, and the clinical imaging radiomics joint model nomogram based on clinical independent risk factors and imaging radiomics showed high diagnostic efficiency for MSI in rectal cancer. The area under curve (AUC) of the training group and the validation group were 0.923 and 0.914, respectively, indicating the most significant evaluation of MSI in rectal cancer. **Conclusion:** Combining different machine learning algorithms, a clinical imaging radiomics nomogram constructed from clinical risk factors and multi-parameter MRI radiomics features can effectively predict the unstable state of preoperative rectal cancer.

[Key words] Rectal cancer; Microsatellite instability; Magnetic resonance imaging; Radiomics; Nomogram

通信作者: 沈俊杰 E-mail: 395932108@qq.com

直肠癌是常见的恶性肿瘤之一,其发病率和死亡率呈上升趋势,且发病人群趋于年轻化^[1-2]。而微卫星不稳定(microsatellite instability, MSI)则为直肠癌的发病机制之一。在直肠癌微卫星不同状态中,DNA错配修复蛋白完整则表示为微卫星稳定(microsatellite stability, MSS),而当DNA错配修复蛋白存在缺陷,且错误积累到某种程度时则表示为MSI。MSI是结直肠癌非常重要的病理学标志物,不仅与患者的治疗方案有一定的相关性,还可以用于评估患者的预后表现^[3]。目前,根据国内外多项共识及规范(包括《中国结直肠癌诊疗规范》等)均建议对新确诊为直肠癌的患者进行MSS状态检测^[1-2]。目前,评估直肠癌MSS状态一般通过肠镜活检或者对肿瘤术后的标本进行病理学检查,常用的检测方式包括免疫组织化学染色(immunohistochemical, IHC)、聚合酶链式反应(polymerase chain reaction, PCR)等。肠镜活检是一种有创性检查,不仅存在假阴性结果的风险,还可能给患者带来一定的损害,而手术取出的肿瘤组织标本也存在检测价格高、耗时长,同时有滞后性的缺点。目前已有研究^[4-5]表明,基于影像组学模型可用于无创性预测直肠癌PD-1表达状态、新辅助治疗后T分期、壁外血管侵犯等情况。因此,影像组学使术前预测直肠癌MSI成为可能。本研究基于多种机器学习方法,进一步探讨临床危险因素和多参数磁共振成像(magnetic resonance imaging, MRI)影像组学特征构建的联合模型及可视化列线图对直肠癌MSI状态术前预测的应用价值。

1 资料和方法

1.1 研究对象

回顾并收集2020年12月—2023年11月蚌埠医科大学第一附属医院直肠癌患者共149例,其中男性96例,女性53例,根据病理学检查结果,将所有患者分为MSI组(34例)和MSS组(115例)。另外再随机按7:3将患者分为训练组(104例)和验证组(45例),通过训练组调整

不同的参数构建模型,再利用验证组验证模型的泛化应用能力。

纳入标准:①经病理学检查证实为直肠癌;②术前行多参数MRI检查;③所有患者未进行任何临床治疗;④术后病理学检查结果包括MSS或MSI诊断结果。排除标准:①有MRI检查禁忌证的患者;②MRI图像质量不佳,肿瘤病灶显示不清者;③临床资料不完整者。

1.2 临床资料

收集患者相关临床资料,包括年龄、性别、术前T分期、组织学分级、淋巴结转移情况、肿瘤长径、肿瘤位置、术前癌胚抗原(carcinoembryonic antigen, CEA)水平、糖类抗原(carbohydrate antigen, CA)19-9水平、高密度脂蛋白水平、低密度脂蛋白水平、胆固醇水平、白细胞计数、红细胞计数、血小板计数(platelet count, PLT)。

1.3 检查与方法

所有患者均采用荷兰Philips公司的Achieva 3.0 T双梯度超导磁共振仪进行检查,检查开始前需使用开塞露进行肠道清洁,具体序列MRI扫描参数如下:

(1)轴位T2加权成像(T2-weighted imaging, T2WI):层厚5 mm;层间距1 mm;重复时间(repetition time, TR)3 000 ms;回波时间(echo time, TE)80 ms;视野(field of view, FOV)260 mm×260 mm;矩阵256×320。

(2)弥散加权成像(diffusion-weighted imaging, DWI):层厚5 mm;层间距1 mm;TR 3 500 ms, TE 55 ms;FOV 289 mm×289 mm;矩阵256×256。

(3)轴位增强T1WI(T1-weighted imaging, T1WI):层厚5 mm;层间距1 mm;TR 600 ms;TE 29 ms;FOV 225 mm×225 mm;矩阵640×432;MRI增强造影剂采用Gd-DTPA(钆喷酸葡胺)经静脉注射,速率为2.0 mL/s,剂量为0.1 mmol/kg,并用等剂量生理盐水冲洗。

1.4 MRI组学特征提取和筛选

将获取到的149例直肠癌患者MRI检查中T2WI、增强T1WI及DWI原始图像(DICOM格

式)导入ITK-SNAP软中,由具有2年腹部MRI影像诊断经验的医师勾画感兴趣区,随后融合成3D图像(图1),并计算组内相关系数(intraclass correlation coefficient, ICC)来评估观察者勾画图像的一致性,并保留 $ICC > 0.75$ (一致性较好)的影像组学特征。

使用pyradiomics对每个序列提取影像组学特

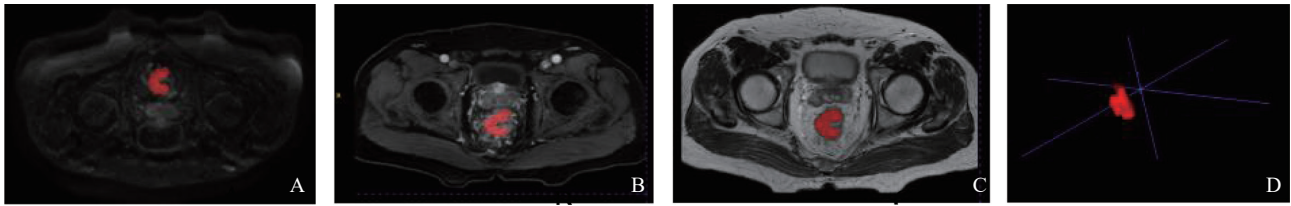


图1 不同序列MRI分割肿瘤区域

男性直肠癌MSI患者,红色区域内为感兴趣区。A:轴位DWI序列;B:轴位增强T1WI序列;C:轴位T2WI序列;D:单层感兴趣区融合得到的3D图像。

1.5 模型的构建与评估

对上述经最小绝对收缩与选择算子(least absolute shrinkage and selection operator, LASSO)降维筛选出的最优组学特征运用不同机器学习算法构建影像组学模型,其中包括逻辑回归(logistic regression, LR)、随机森林(random forest, RF)、支持向量机(support vector machine, SVM)、K最近邻(k-nearest neighbor, KNN),朴素贝叶斯(naive Bayes tree, NBT)5种不同的机器学习算法。根据受试者工作特征曲线的曲线下面积(area under curve, AUC)选出最优机器学习模型并构建影像组学标签,计算影像组学分数(radiomics score, Rad-score)构建影像组学模型。先后利用单因素和多因素logistic分析,筛选出临床独立危险因素,构建临床模型。然后根据临床独立危险因素及影像组学分数构建临床-影像组学联合模型,并绘制可视化列线图。采用AUC评估临床模型、影像组学模型及联合模型的诊断效能,并计算相应的95% CI、灵敏度和特异度等指标。采用DeLong检验评估3种模型的预测效能差异是否有统计学意义, $P < 0.05$ 为差异有统计学意义。绘制校准曲线和决策曲线,方便观察模型预测概率与真实概率之间的一致性,以及评估3种模型的

征,然后对特征进行Z-score标准化。利用单因素秩和检验对影像组学特征进行显著性检验, $P < 0.05$ 为阈值,再将得到的影像组学特征进行Spearman相关性分析,大于0.5定义为冗余特征,将其进行剔除处理。最后对特征进行Elastic-logistic分析,并筛选出评估直肠癌MSI的最优特征。

实际临床应用价值。

1.6 统计学处理

采用SPSS 23.0软件对数据进行统计学分析,符合正态分布的计量资料包括患者年龄、肿瘤长径等以 $\bar{x} \pm s$ 表示,对不符合正态分布的计量资料以 $M(P_{25}, P_{75})$ 表示,采用t检验或Mann-Whitney U检验进行组间的差异对比;采用 χ^2 检验或Fisher精确概率检验比较组间患者计数资料的差异, $P < 0.05$ 为差异有统计学意义。

2 结果

2.1 一般临床资料

训练组和验证组之间的临床资料比较显示,训练组中MSI患者的PLT高于MSS患者($P < 0.001$)。而患者的肿瘤长径、病理学分级、肿瘤T分期、淋巴结转移情况、肿瘤位置、术前CEA水平、CA19-9水平、高密度脂蛋白水平、低密度脂蛋白水平、胆固醇水平、白细胞计数、红细胞计数等临床资料在两组间比较,差异均无统计学意义($P > 0.05$)。

2.2 临床特征危险因素

对上述指标先后进行单因素及多因素logistic回归分析。单因素logistic回归分析结果

显示, 年龄 (OR=0.967, 95% CI 0.943~0.99, $P=0.006$)、性别 (OR=2.612, 95% CI 1.288~5.487, $P=0.009$) 及PLT (OR=1.005, 95% CI 1.001~1.01, $P=0.023$) 差异有统计学意义; 多因素logistic回归分析显示, 年龄 (OR=0.959, 95% CI 0.933~0.984, $P=0.002$)、性别 (OR=3.747, 95% CI 1.725~8.623, $P=0.001$) 及PLT (OR=1.005, 95% CI 1.00~1.01, $P=0.043$) 差异有统计学意义。最终筛选出年龄、性别、PLT为预测直肠癌MSI状态的独立危险因素, 并构建临床预测模型。

2.3 特征筛选和机器学习模型构建

使用pyradiomics对每个序列共提取出944个特征影像组学特征, 然后对全部特征进行Z-score标准化。利用单因素秩和检验对影像组学特征进行显著性分析, 得到240个单因素特征。再将影像组学特征进行Spearman相关性分析, 剩余影像组学特征20个。最后对影像组学特征进行Elastic-logistic分析, 最终得出DWI_wavelet.HHH_glrIm_HighGrayLevelRunEmphasis、T1CE_wavelet.HHL_firstorder_Mean、T2_wavelet.LHL_glcm_Correlation、T2_wavelet.HLH_glcm_ClusterShade共4个影像组学特征 (图2), 并计

算出影像组学标签分数Rad-score, 绘制出箱式图 (图3)。制作水落图展现出训练组和验证组的影像组学标签分数有着显著的分布差异 (图4)。使用LR、SVM、RF、K-NN、NBT 5种机器学习算法对组学特征进行建模, 其中RF表现出了最佳的稳定性及较好的预测性能, 训练组及验证组AUC分别为0.891、0.896 (图5)。最终选择RF算法构建的模型作为最佳影像组学模型。为了尝试提高预测模型的准确度, 进一步构建了临床-影像组学联合模型。

2.4 模型比较与验证

通过比较, 临床-影像组学联合模型预测效果最佳, AUC (训练组AUC=0.923; 验证组AUC=0.914) 均高于单一临床模型及影像组学模型 (图6, 表1、2)。根据DeLong检验结果比较各模型之间的预测效能, 差异均有统计学意义 ($P<0.05$)。联合临床独立危险因素及影像组学分数, 构建可视化列线图 (图7)。在训练组和验证组中, 列线图的校准曲线均展现出较好的拟合度, 表示临床-影像组学联合模型预测直肠癌MSI状态具有较高的稳定性 (图8)。绘制的决策曲线 (图9) 展现了不同模型预测直肠癌MSI状态的临床实际价值, 其中联合模型展现出了最高净收益, 提示其临床应用价值最大。

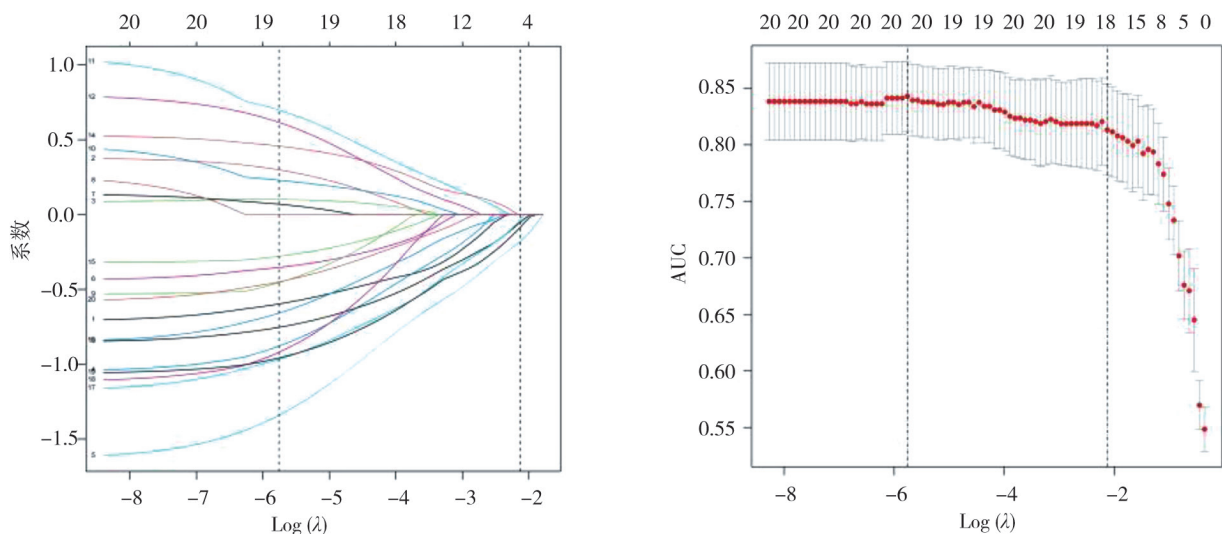


图2 LASSO筛选出4个最优影像组学特征

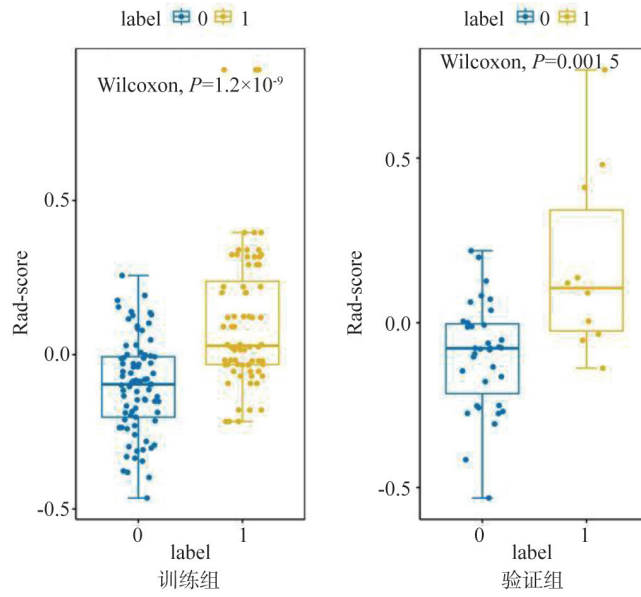


图3 训练组与验证组的Rad-score箱式图

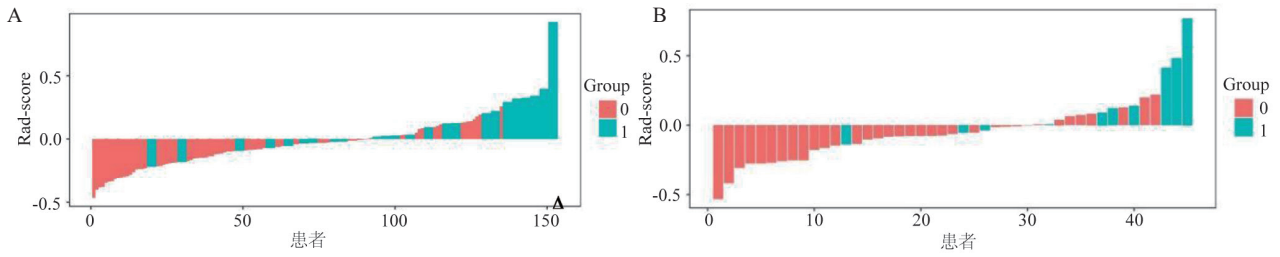


图4 MRI影像组学标签的影像组学评分分布图

A: 训练组; B: 验证组。

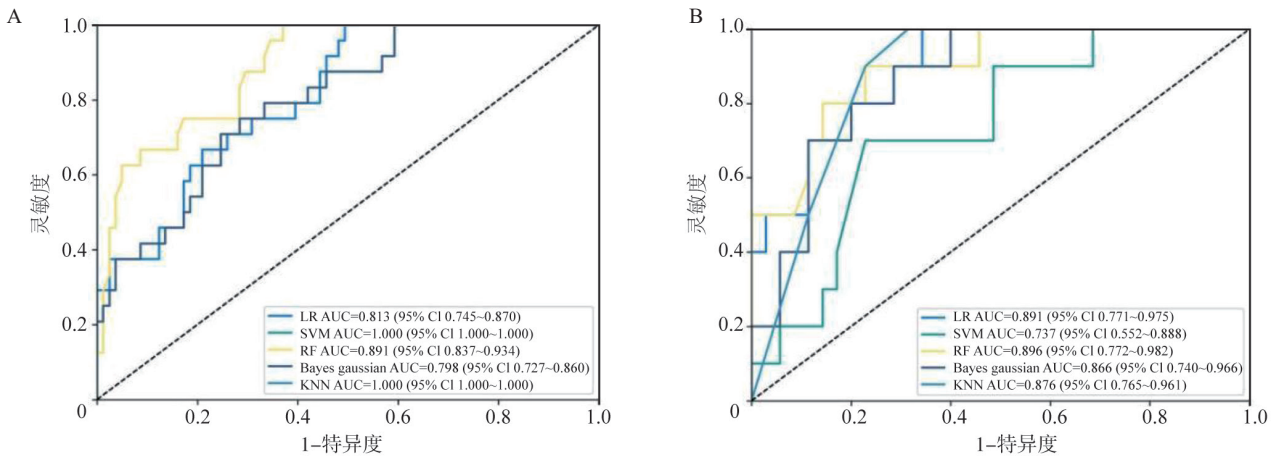


图5 5种机器学习算法的ROC曲线

A: 训练组; B: 验证组。

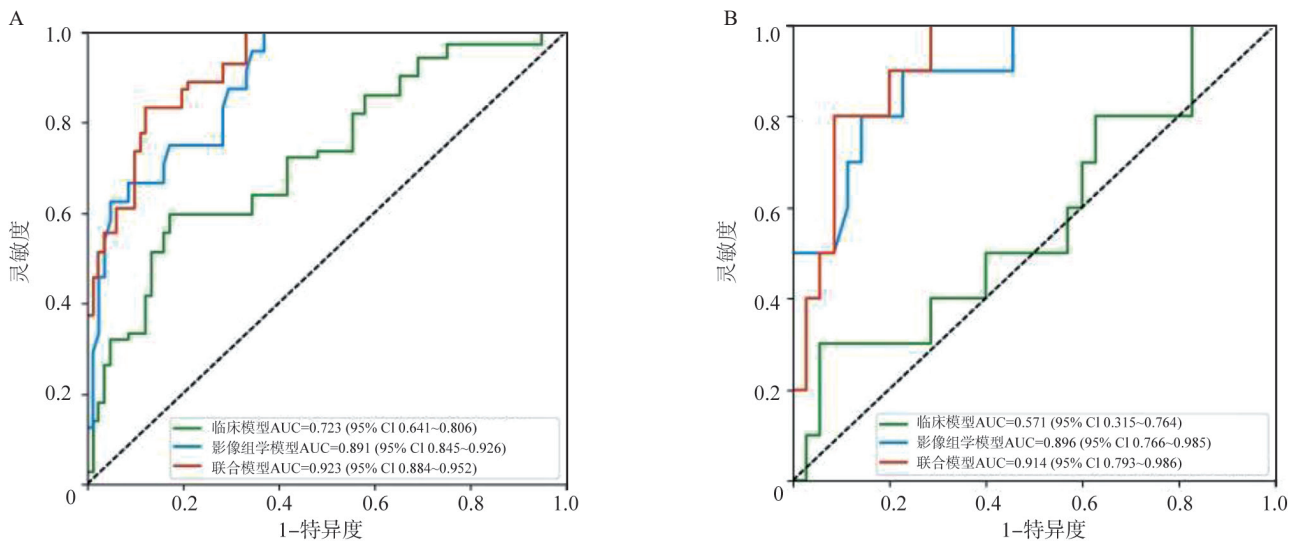


图6 临床、影像组学及临床-影像组学联合模型评估MSI状态的ROC曲线

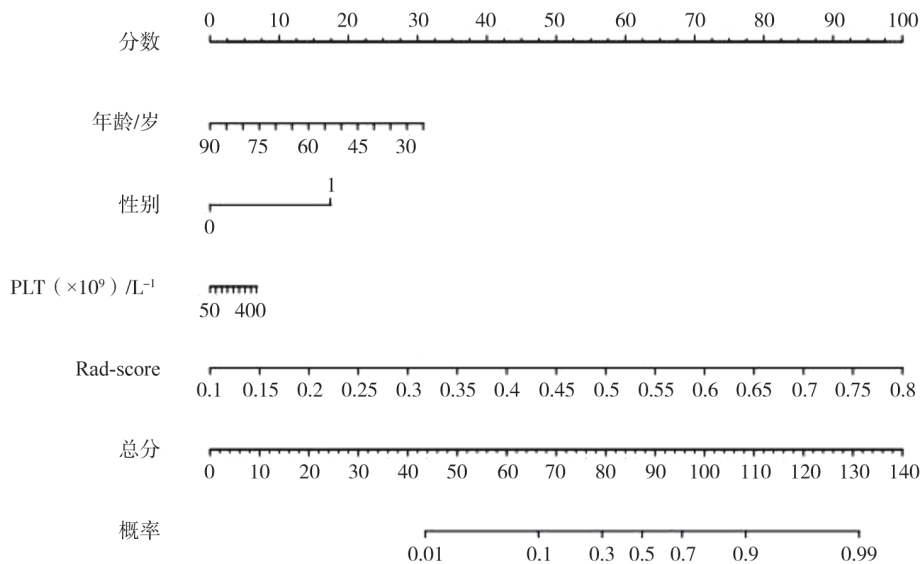


图7 预测直肠癌MSI状态的列线图

A: 训练组; B: 验证组。

表1 训练组模型比较

训练组	AUC	95% CI	灵敏度	特异度	准确度
临床模型	0.723	0.641~0.806	0.597	0.827	0.719
影像组学模型	0.849	0.845~0.926	1.000	0.630	0.804
联合模型	0.923	0.884~0.952	0.833	0.877	0.856

表2 验证组模型比较

验证组	AUC	95% CI	灵敏度	特异度	准确度
临床模型	0.571	0.315~0.764	0.400	0.686	0.622
影像组学模型	0.896	0.766~0.985	0.900	0.714	0.756
联合模型	0.914	0.793~0.986	0.800	0.914	0.889

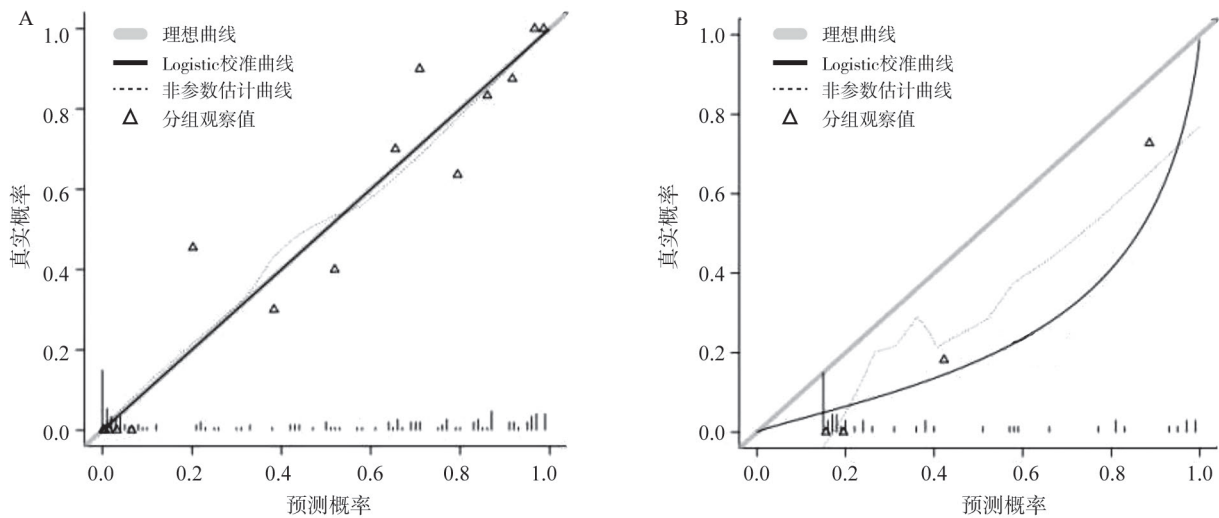


图8 校正曲线

A: 训练组; B: 验证组。

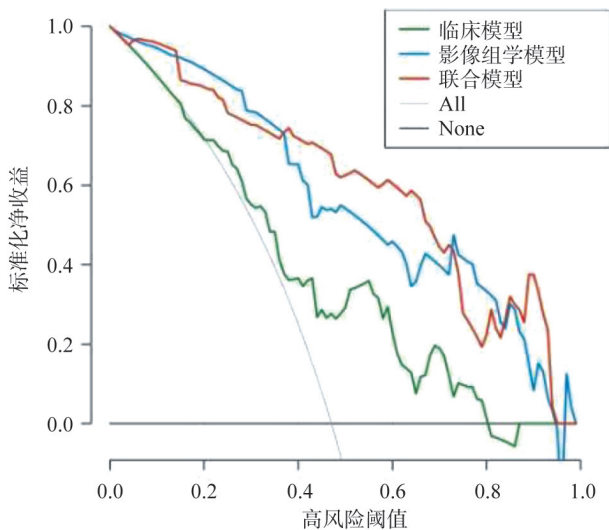


图9 决策曲线

3 讨论

直肠癌是结直肠癌中最常见的类型,在约15%的结直肠癌患者中可以检测到MSI^[6]。已有研究^[7]表明,Ⅱ期结直肠癌MSI患者容易对5-氟尿嘧啶产生一定的耐药性,导致以5-氟尿嘧啶为基础的新辅助化疗对其治疗效果不佳。但是,免疫治疗可能对部分MSI患者具有非常好的疗效,甚至无需新辅助化疗及手术治疗。因此,术前

预测直肠MSI状态为患者制订个性化治疗方案具有重要意义^[8]。

本研究中患者外周血的PLT为单因素、多因素logistic回归筛选出的临床独立危险因素,并发现MSI患者血液中PLT略高于MSS患者,这与既往其他学者研究^[8]结果一致。有相关研究^[9]表明,血小板不仅参与机体生理性止血作用,还具有一定的免疫功能,可以直接调节肿瘤微环境中的淋巴细胞效应功能,触发肿瘤细胞的特定信号通路,通过募集能够释放增加渗透性的细胞因子和生长因子的前转移性白细胞,间接支持肿瘤细胞的迁移。除此之外,年龄和性别也是临床独立危险因素之一,同样具有良好的预测价值。

根据以往的影像组学研究发现,结合临床危险因素、定性影像学数据和影像组学特征的临床-影像组学模型可以有效地预测结直肠癌的MSI状态。Ying等^[10]基于增强CT影像组学构建预测结直肠癌MSI状态模型,利用ICC、mRMR和LASSO进行特征筛选,共筛选出12个影像组学特征,3个临床危险因素以及CT报告的肿瘤炎症反应征象。结果,在训练集和验证集中联合模型的AUC最高,分别为0.87(95% CI 0.81~0.93)和0.90(95% CI, 0.83~0.96),校准曲线和决策曲线也表现出了良好的校准效能和临床实用性。

Pei等^[11]基于CT影像组学的结直肠癌研究发现,结合影像组学特征、肿瘤位置、患者年龄、高密度脂蛋白水平和PLT的影像组学列线图可以预测MSI状态。Li等^[7]通过基于MRI的影像组学预测结直肠癌MSI状态并分析得出,T2WI和表观弥散系数(apparent diffusion coefficient, ADC)特征联合模型效能高于单一序列模型,其原因可能是多个序列MRI的肿瘤图像信息更全面,可以进一步反映结直肠癌的异质性。这些结果均证实了基于影像组学的预测模型可以有效地捕捉更深层次的定量图像信息,及时获取肿瘤的生物特征。而目前仅针对直肠癌方面的研究较少,且MRI对软组织分辨率高,勾画3D感兴趣区准确度较高,因此,本研究基于多种机器学习,探究多参数MRI影像组学联合临床独立危险因素对直肠癌MSI的预测价值。

本研究共纳入了T2WI、DWI及增强T1WI序列勾画3D感兴趣区,综合多个序列构建的影像组学模型在训练组和验证组中AUC分别为0.891、0.896。加入临床独立危险因素后,临床-影像组学联合模型AUC分别为0.923、0.914,临床因素的纳入一定程度上提高了预测模型的诊断效能。其可能的原因是多个因素的纳入更加全面地展示了肿瘤的生物行为信息,在预测直肠癌患者的MSI状态中具有较高的可行性。这与很多学者^[12-13]研究结论相似。与此同时,为了促进临床应用,我们基于临床特征和影像组学特征绘制出可视化列线图,为临床术前预测直肠癌MSI状态提供了新思路。另外,我们在研究中同时比较了5种不同的常用机器学习算法,包括LR、SVM、RF、K-NN、NBT,结果表明RF算法的影像组学模型表现出了最佳的稳定性及较好的预测性能,其训练组和验证组的AUC分别为0.891和0.896,因此,稳定性较好的RF算法在预测直肠癌MSI状态方面具有很大的潜力。

本研究存在一定的局限性:首先,直肠癌MSI数据量较少,后续将继续增加样本量。其

次,本文通过手动勾画病灶感兴趣区难免存在偏差,随着自动分割软件的发展今后也许能更加精准地勾画感兴趣区。最后,本研究是一项单中心、回顾性研究,尚待加入多中心、前瞻性的数据验证研究结果的真实性及稳定性。

综上所述,本研究基于多种机器算法,构建了一个联合多参数MRI影像组学和临床独立危险因素的联合模型列线图,可作为一种术前无创性预测直肠癌MSI状态的可靠工具,为患者的个性化治疗方案提供更精准的帮助。

【利益冲突声明】所有作者均声明不存在利益冲突。

【参 考 文 献】

- [1] 国家卫生健康委员会医政司,中华医学会肿瘤学分会. 中国结直肠癌诊疗规范(2023版)[J]. 协和医学杂志, 2023, 14(4): 706-733.
- [2] BATTAGLIN F, NASEEM M, LENZ H J, et al. Microsatellite instability in colorectal cancer: overview of its clinical significance and novel perspectives [J]. Clin Adv Hematol Oncol, 2018, 16(11): 735-745.
- [3] NGUYEN L H, GOEL A, CHUNG D C. Pathways of colorectal carcinogenesis [J]. Gastroenterology, 2020, 158(2): 291-302.
- [4] 陈吉虎, 李 瑞, 刘 浩, 等. 影像组学术前预测结直肠癌淋巴管侵犯的价值[J]. 临床放射学杂志, 2022, 41(3): 495-499.
- [5] 贾建业, 刘紫妍, 丁 聪, 等. 基于多参数MRI影像组学列线图预测直肠癌PD-1表达状态[J]. 临床放射学杂志, 2024, 43(1): 90-95.
- [6] VANGALA D, NILIUS-ELILIWI V. Neue medikamentöse konzepte Bei patienten mit kolorektalen karzinomen und mikrosatelliteninstabilität [J]. Zentralbl Chir, 2023, 148(6): 475-482.
- [7] LI Z B, DAI H, LIU Y X, et al. Radiomics analysis of multi-sequence MR images for predicting microsatellite instability status preoperatively in rectal cancer [J]. Front Oncol, 2021, 11: 697497.
- [8] 刘 旭, 冯卫华, 王 涛, 等. 临床-影像组学联合模型对直肠癌微卫星不稳定性的预测价值[J]. 精准医学杂志, 2022, 37(4): 326-331.
- [9] GAERTNER F, MASSBERG S. Patrolling the vascular borders: platelets in immunity to infection and cancer [J]. Nat Rev Immunol, 2019, 19: 747-760.

- [10] YING M L, PAN J F, LU G H, et al. Development and validation of a radiomics-based nomogram for the preoperative prediction of microsatellite instability in colorectal cancer [J] . BMC Cancer, 2022, 22(1): 524.
- [11] PEI Q, YI X P, CHEN C, et al. Pre-treatment CT-based radiomics nomogram for predicting microsatellite instability status in colorectal cancer [J] . Eur Radiol, 2022, 32(1): 714-724.
- [12] ZHANG W, YIN H K, HUANG Z X, et al. Development and validation of MRI-based deep learning models for prediction of microsatellite instability in rectal cancer [J] . Cancer Med, 2021, 10(12): 4164-4173.
- [13] HUANG Z X, ZHANG W, HE D, et al. Development and validation of a radiomics model based on T2WI images for preoperative prediction of microsatellite instability status in rectal cancer: study protocol clinical trial (SPIRIT compliant) [J] . Medicine (Baltimore), 2020, 99(10): e19428.
- (收稿日期: 2024-09-06 修回日期: 2024-10-10)

《肿瘤影像学》2025年征订启事

《肿瘤影像学》自1992年创刊以来深受医学界欢迎, 1998年经中华人民共和国科学技术部、国家新闻出版署批准为国内外公开正式发行的期刊。杂志刊号: ISSN 2096-6210, CN 31-2087/R。采用优质铜版纸印制, A4开本, 正文64页/期, 双月刊。被中国学术期刊综合评价数据库、中国核心期刊(遴选)数据库、中国期刊全文数据库等收录, 是中国科技核心期刊。主要报道医学影像领域中科研成果、临床应用、综述、病例报告、讲座及与理工结合的有关论文等。

《肿瘤影像学》坚持学术性与科学性, 信息量大, 具有临床实用价值。是医院图书馆、影像科室及高等医药院校收存和使用的学术刊物, 是临床医学影像专业医务人员晋升中、高级职称的重要论文发表园地。欢迎各医学院校、医学图书馆、影像科室及个人向当地邮政局订阅。

本刊双月月末出版, 邮发代号4-653, 定价每期30元, 每年共180元整。

单位全称: 《肿瘤影像学》编辑部

通信地址: 上海市东安路270号复旦大学附属肿瘤医院10号楼415室

邮 编: 200032

电 话: (021)64188274

E - m a i l : zlyxx@zhongliuyingxiangxue.com

网 址 : www.zhongliuyingxiangxue.com

《肿瘤影像学》编辑部