

DOI: 10.14188/j.1671-8852.2024.0850

文章编号 1671-8852(2026)05-0626-10

多种机器学习模型预测膀胱癌肺转移预后

徐冉 陈思安 吴金海 曾熙恒 雷东 陈彦飞 朱学进 王斌
广州医科大学附属肿瘤医院泌尿肿瘤外科 广东 广州 510120

摘要 目的:探讨膀胱癌肺转移(BCLM)患者的预后因素,并构建基于机器学习的生存预测模型,为临床提供个性化管理建议。方法:利用SEER数据库(2010—2021年)筛选出2 190例BCLM患者并采用单变量和多变量Cox回归及竞争风险Cox回归分析,识别影响总体生存(OS)和癌症特异性生存(BCSS)的因素,后通过Lasso回归与逻辑回归筛选5个关键变量,构建9种机器学习模型,并优化其参数以提升预测性能。结果:随机森林(RF)模型在9种模型中表现最佳,其AUC分别为6个月0.938 0、1年0.925 8、3年0.877 7。RF模型的整体AUC为0.814 4,高于XGBoost(0.807 1)、多层感知器(0.806 3)、CatBoost(0.804 1)、决策树(0.796 8)等模型。多部位转移、更高N分期(N2、N3)、鳞状细胞癌(SCC)组织类型、未手术患者为独立危险因素。空白组即未明确肿瘤分级的患者死亡风险显著增加($HR=1.511$, 95%CI:1.343~1.700, $P<0.001$)。结论:基于BCLM患者数据构建的随机森林模型,为BCLM患者生存预测及个性化管理提供了有力工具。

关键词 膀胱癌;肺转移;预后预测;机器学习;生存分析

中图分类号 R737.14

文献标识码 A

Prediction of prognosis for lung metastases in bladder cancer using multiple machine learning models

XU Ran, CHEN Si'an, WU Jinhai, ZENG Xiheng, LEI Dong, CHEN Yanfei, ZHU Xuejin, WANG Bin

Department of Urologic Oncology, Cancer Hospital of Guangzhou Medical University, Guangzhou 510120, Guangdong, China

Abstract

Objective: To explore the prognostic factors for bladder cancer with lung metastasis (BCLM) and develop a machine learning-based survival prediction model to provide personalized management recommendations for clinical practice. **Methods:** A total of 2 190 BCLM patients from the SEER database (2010 - 2021) were selected for this study. Univariate and multivariate Cox regression analyses and competing risk Cox regression were used to identify factors affecting overall survival (OS) and cancer-specific survival (BCSS). Five key variables were selected through Lasso regression and logistic regression. Nine machine learning models were constructed, and their parameters were optimized to improve prediction performance. **Results:** Among the nine models, the Random Forest (RF) model performed the best, with AUC values of 0.938 0 for 6-month, 0.925 8 for 1-year, and 0.877 7 for 3-year survival predictions. The overall AUC of the RF model was 0.814 4, which was higher than XGBoost (0.807 1), Multilayer Perceptron (0.806 3), CatBoost (0.804 1), and Decision Tree

课题来源:广州市卫生健康科技项目(编号:20251A011091);广州医科大学附属肿瘤医院GCP委员会5555项目[IIT-2021-003(MN1)]。

作者简介:徐冉,男,1998—,医学硕士生,主要从事泌尿系统恶性肿瘤的研究,E-mail:xuran6888@163.com。

通信作者:王斌,E-mail:972364766@qq.com。

(0.796 8). Multiple metastasis sites, higher N-stage (N2, N3), squamous cell carcinoma (SCC) histology, and non-surgical treatment were identified as poor prognostic factors. Patients with unspecified tumor grade had a significantly increased mortality risk ($HR=1.511$, 95%CI: 1.343-1.700, $P<0.001$). **Conclusion:** The Random Forest model developed based on BCLM patient data provides a powerful tool for predicting survival and supporting personalized management.

Key Words Bladder Cancer; Lung Metastasis; Prognostic Prediction; Machine Learning; Survival Analysis

膀胱癌(bladder cancer, BC)是男性第四大常见癌症,全球每年新诊断病例约573 000例,死亡人数超过21万,预计到2024年将大幅增加^[1-2]。约10%~15%的膀胱癌患者初诊时已出现转移,最常见的转移部位为肺、骨和肝^[3-5]。转移性膀胱癌患者生存率较低,特别是肺转移患者的5年生存率仅为5%^[6-7]。根据SEER数据库分析,膀胱癌肺转移(bladder cancer with lung metastasis, BCLM)患者的平均生存期为7.19个月,中位生存期仅为3个月,且目前治疗方法有限,主要集中在病情控制和症状缓解等方面^[8-9]。

然而,关于BCLM患者生存预后的预测模型较少。现有的早期预后模型(如nomograms)虽然有应用,但其精确度仍不理想,受试者工作特征(ROC)曲线下面积(AUC)或一致性指数(C-index)约为0.7^[9]。因此,开发更精准的模型显得尤为必要。随着机器学习(machine learning, ML)技术在临床肿瘤学中的广泛应用,ML不仅能帮助诊断癌症,还能预测疾病进程并为个性化治疗提供支持^[10-11]。本研究采用9种机器学习算法构建BCLM预后预测模型,最终发现随机森林算法(Random Forest, RF)表现最佳^[12]。通过SEER数据库分析,本研究识别了影响BCLM患者预后的关键因素,并利用优化后的随机森林算法预测了其在6个月、1年和3年的生存概率。这为推动了临床人工智能模型的发展以及选择更合适的治疗策略提供了有力支持。

1 资料与方法

1.1 数据资料收集 本研究通过回顾性提取SEER数据库(www.seer.cancer.gov)中的BCLM患者信息,选择2010—2021年间的3 095例。SEER数据库由美国综合国立癌症研究所收集,涵盖全美14个州18个基于人口的癌症登记机构的癌症数据,记录了癌症的发病率、患者人口统计学和死亡率,包括原发部位、组织学诊断、分级、形态学、病理分

期和生存状态等信息。本研究采用SEER*Stat软件(版本:8.4.3),在“Incidence. SEER Research Plus Data, 17 Registries, Nov 2023 Sub (2000—2021)”处,通过查询“Site code ICD-O-3/WHO 2008”并将“膀胱”作为发病部位,确定本研究的患者队列并提取患者相关信息,包括年龄、性别、种族、肿瘤原发部位(定义编码为C67.0-C67.9)、肿瘤大小、病理分级、TNM分期(基于美国癌症联合委员会第7版)、淋巴结转移、区域淋巴结清扫、化疗和放疗及患者的生存数据。纳入标准:(1)肿瘤位于膀胱(编码C67.0-C67.9),符合国际肿瘤疾病分类第3版(ICD-O-3);(2)诊断为BCLM的患者。排除标准:(1)同时患有两种或以上原发癌症的患者;(2)TNM分期不明确或未知的病例;本研究符合《赫尔辛基宣言》的要求,且SEER是公开数据库,纳入本研究中患者的记录和信息在分析前是匿名的。因此,本研究不需要获得机构伦理委员会的批准。

1.2 统计学方法 采用R 4.0.2软件分析数据,所用R包:survival、cmprsk、glmnet、stats、randomForest、xgboost、nnet、catboost、rpart、gbm、adabag、e1071、class、pROC、care等。首先以患者的生存状态(生存或死亡)为终点,进行基线描述性统计。随后,通过单变量和多变量Cox回归及多变量竞争风险模型筛选出具有统计学意义的特征,包括种族、病理类型、分级、N分期、骨转移、脑转移、肝转移、化疗、手术及淋巴结手术。进一步的,使用LASSO回归和逻辑回归进一步筛选出关键变量(病理类型、分级、骨转移、化疗、手术),并将这些变量作为构建机器学习模型的核心变量。

模型构建前,所有患者数据按7:3比例随机分为训练集和测试集。接着,采用9种机器学习算法(RF、XGB、MLP、CB、DT、GBDT、AB、NB、KNN)构建预后预测模型,通过AUC评估各模型的性能。最后,使用Shapley Additive Explanations(SHAP)法量化核心变量在模型中的贡献度,评估核心变量重要性。以 $P<0.05$ 为差异有统计学意义。

2 结果

2.1 纳入研究的患者特征 2 190 例 BCLM 患者数据被纳入研究。其中生存患者为 215 例,死亡患者为 1 975 例,总体生存时间为 3.00(1.00,9.00)个月,生存患者的生存时间为 8.00(2.00,21.00)个月,死亡患者为 3.00(1.00,8.00)个月。移行细胞癌(TCC)最为常见(74.57%),且该类型患者的生

存率较高。肿瘤分级以高级别(Ⅲ~Ⅳ级)为主(41.01%)。骨转移患者死亡率较高(30.03%)。在治疗方面,生存组中 63.72% 的患者接受了化疗,而死亡组中未接受化疗患者死亡率较高(61.77%)。手术治疗中,大多数患者未进行完全膀胱切除术(62.83%)。放疗的疗效较为有限,接受放疗患者仅 10.23%(见表 1)。

表 1 SEER 数据队列中 BC 肺转移(BCLM)患者的基线特征 [例(%);M(P₂₅,P₇₅)]

变量		总体(n=2 190)	存活(n=215)	死亡(n=1 975)	P
生存期/月		3.00 (1.00, 9.00)	8.00 (2.00, 21.00)	3.00 (1.00, 8.00)	<0.001
年龄	>80 岁	492 (22.46)	37 (17.21)	455 (23.04)	0.142
	65~80 岁	962 (43.93)	103 (47.91)	859 (43.49)	
	<65 岁	736 (33.61)	75 (34.88)	661 (33.47)	
性别	女性	698 (31.87)	79(36.74)	619 (31.34)	0.124
	男性	1 492 (68.13)	136(63.26)	1 356 (68.66)	
种族	黑种人	219 (10.00)	16 (7.44)	203 (10.28)	0.006
	其他	140 (6.39)	24 (11.16)	116 (5.87)	
	白种人	1 831 (83.61)	175 (81.40)	1 656 (83.85)	
组织学类型	腺癌	70 (3.20)	12 (5.58)	58 (2.94)	0.033
	鳞状细胞癌(SCC)	87 (3.97)	4 (1.86)	83 (4.20)	
	移行细胞癌(TCC)	1 633 (74.57)	167 (77.67)	1 466 (74.23)	
	其他	400 (18.26)	32 (14.88)	368(18.63)	
婚姻状况	已婚	1 006 (45.94)	117 (54.42)	889 (45.01)	0.016
	丧偶/离婚/其他	1 184(54.06)	98 (45.58)	1 086 (54.99)	
家庭收入中位数	<54 999 美元	267 (12.19)	20 (9.30)	247 (12.51)	0.075
	55 000~74 999 美元	719 (32.83)	65 (30.23)	654 (33.11)	
	75 000~94 999 美元	671 (30.64)	82 (38.14)	589 (29.82)	
	>95 000 美元	533 (24.34)	48 (22.33)	485 (24.56)	
分级	高分化;Ⅰ级	13 (0.59)	1 (0.47)	12 (0.61)	<0.001
	中分化;Ⅱ级	57 (2.60)	3 (1.40)	54 (2.73)	
	差分化;Ⅲ级	261 (11.92)	7 (3.26)	254 (12.86)	
	未分化;Ⅳ级	637 (29.09)	20 (9.30)	617 (31.24)	
	未知	1 222 (55.80)	184 (85.58)	1 038 (52.56)	
T 分期	T0	32 (1.46)	3 (1.40)	29 (1.47)	0.158
	T1	311 (14.20)	35 (16.28)	276 (13.97)	
	T2	766 (34.98)	86 (40.00)	680 (34.43)	
	T3	139 (6.35)	13 (6.05)	126 (6.38)	
	T4	358 (16.35)	22 (10.23)	336 (17.01)	
	TX	584 (26.67)	56 (26.05)	528(26.73)	
N 分期	N0	1 114 (50.87)	116 (53.95)	998 (50.53)	0.762
	N1	213 (9.73)	22 (10.23)	191 (9.67)	
	N2	321 (14.66)	27 (12.56)	294 (14.89)	
	N3	131 (5.98)	14 (6.51)	117 (5.92)	
	NX	411(18.77)	36 (16.74)	375 (18.99)	
骨转移	无/未知	1 549 (70.73)	167 (77.67)	1 382 (69.97)	0.023
	有	641 (29.27)	48 (22.33)	593 (30.03)	
脑转移	无/未知	2 099 (95.84)	208 (96.74)	1 891 (95.75)	0.606

续表

变量		总体(<i>n</i> =2 190)	存活(<i>n</i> =215)	死亡(<i>n</i> =1 975)	<i>P</i>
肝转移	有	91 (4.16)	7 (3.26)	84 (4.25)	0.001
	无/未知	1 609 (73.47)	178 (82.79)	1 431 (72.46)	
化疗	有	581 (26.53)	37 (17.21)	544 (27.54)	<0.001
	无/未知	1 298 (59.27)	78 (36.28)	1 220 (61.77)	
放疗	有	892 (40.73)	137 (63.72)	755 (38.23)	0.439
	无/未知	1 925 (87.90)	193 (89.77)	1 732 (87.70)	
淋巴结手术	有	265 (12.10)	22 (10.23)	243 (12.30)	>0.999
	无/未知	2 187 (99.86)	215 (100.00)	1 972 (99.85)	
原发部位手术	有	3 (0.14)	0 (0.00)	3 (0.15)	<0.001
	根治性膀胱切除术	76 (3.47)	9 (4.19)	67 (3.39)	
	无/未知	738 (33.70)	44(20.47)	694 (35.14)	
	非根治性膀胱切除术	1 376 (62.83)	162 (75.35)	1 214 (61.47)	

2.2 影响 BCLM 患者预后独立危险因素 本研究首先采用单因素 Cox 回归分析,以识别显著影响 BCLM 患者总体生存期(OS)的临床和病理特征。随后,我们进一步对单因素分析中显著的变量进行多因素 Cox 回归分析,最终筛选出以下独立预测因素:骨转移、肝转移、脑转移、化疗、组织学类型、N 分期、种族、手术方式以及淋巴结手术。其中,种族变量中的“其他”类别(包括美洲原住民、土著、亚洲人等)为保护因素($HR=0.771$, 95% CI : 0.611~0.972, $P=0.028$)。此外,鳞状细胞癌(SCC)组织学类型($HR=1.761$, 95% CI : 1.247~2.486, $P<0.001$)、N 分期中的 N2 期($HR=1.168$, 95% CI : 1.021~1.338, $P=0.024$)和 N3 期($HR=1.238$, 95% CI : 1.018~1.506, $P=0.033$),手术方式中的“无/未知”($HR=1.906$, 95% CI : 1.459~2.491, $P<0.001$),“非根治性膀胱切除术”($HR=1.396$, 95% CI : 1.075~1.812, $P=0.010$),肝转移($HR=1.492$, 95% CI : 1.343~1.656), $P<0.001$),脑转移($HR=1.379$, 95% CI : 1.099~1.730), $P<0.001$)为独立危险因素均为独立危险因素。此外,基于多因素竞争风险回归分析,我们发现:骨转移、化疗、肿瘤分级以及 N 分期是影响 BCLM 患者肿瘤特异性生存期 BCSS 的独立因素(见表 2)。这些变量与多因素 Cox 回归分析中影响 OS 的变量高度一致。此外,肿瘤分级在竞争风险模型中也表现出显著性($HR=0.938$, 95% CI : 0.887~0.991, $P=0.023$)。

2.3 BCLM 预后机器学习模型构建因素 本研究通过分析,识别出一系列显著影响 BCLM 患者 OS 和 BCSS 的特征变量,包括骨转移、肝转移、脑转移、化疗、组织学类型、肿瘤分级、N 分期、种族、手术方

式以及淋巴结手术治疗(见表 2)。为了提升所构建预后模型的稳定性、鲁棒性及其解释性和预测准确性,本研究引入了 LASSO 回归和逻辑回归进行多层次的特征筛选。在 LASSO 回归分析中,选择了 Lambda 值为 -5.379,通过筛选出 9 个具有显著影响力的特征变量(图 1)。随后,基于这些特征,进行了逻辑回归分析,以评估这些变量与膀胱癌患者 OS 的关系。最终筛选出了 5 个与膀胱癌肺转移患者预后显著相关的特征:组织学类型、肿瘤分级、骨转移、化疗及手术方式,作为构建机器学习预后模型的关键变量。

2.4 多种机器学习模型的构建与验证 通过关键变量我们构建了包括随机森林(RF)、XGBoost(XGB)、多层感知器(MLP)、CatBoost(CB)、决策树(DT)、梯度提升决策树(GBDT)、AdaBoost(AB)、朴素贝叶斯(NB)和 K 最近邻(KNN)九种机器学习模型 ROC 曲线见图 2。我们将患者数据按 7:3 比例分为训练组和测试组,使用训练组构建模型,测试组用于评估模型泛化能力。所有模型均采用 10 折交叉验证以提高评估的可靠性。各模型的预测性能通过 AUC 评估。结果显示,RF 模型表现最佳,AUC 为 0.814 4,召回率为 1.0, F1-Score 为 0.952 2(表 3)。

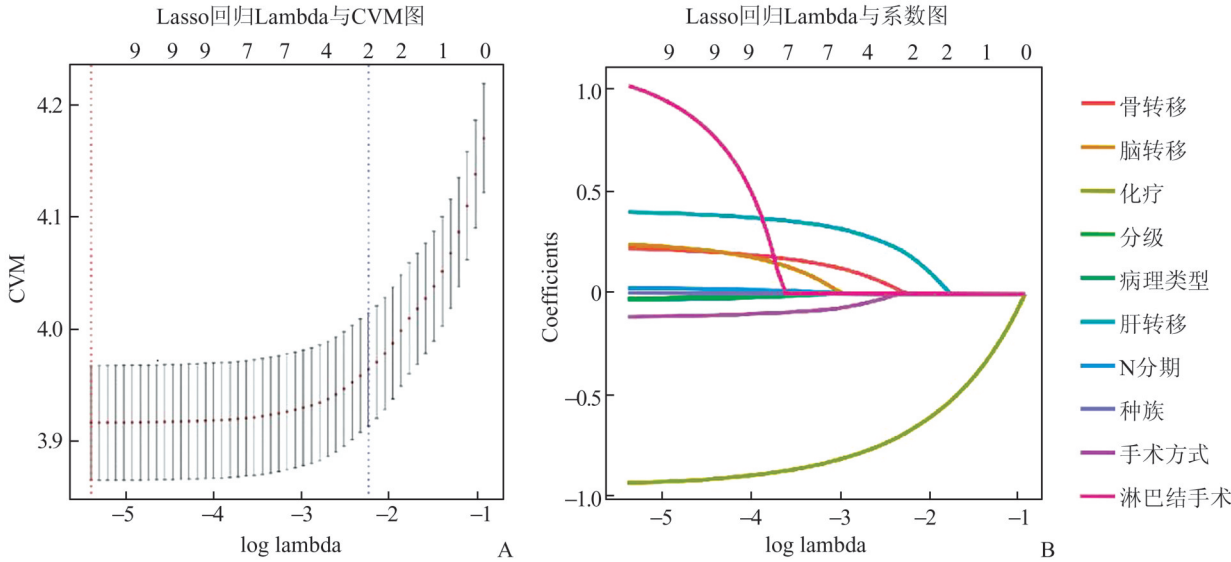
2.5 随机森林模型的构建与验证 为优化随机森林模型的性能,我们通过网格搜索(Grid Search)确定了随机森林最佳超参数(补充表 1,扫描本文 OSID 可见)。最终,我们基于优化后的 RF 模型构建预测系统,并引入 SMOTE(Synthetic Minority Oversampling Technique)采样技术,以平衡 6 个月、1 年和 3 年不同生存期的数据,用于预测 BCLM 患者在 6 个月、1 年和 3 年内的 OS,并绘制训练集和测

试集的 ROC 曲线,计算 AUC 值以评估预测性能(图 3):6 个月生存预测(测试集: $AUC=0.938$;训练集: $AUC=0.932$);1 年生存预测(测试集: $AUC=0.926$;训练集: $AUC=0.924$);3 年生存预测(测试集: $AUC=0.878$;训练集: $AUC=0.905$)。

表 2 基于临床病理特征数据的膀胱癌患者肺转移 BCLM 的单因素和多因素 Cox 回归分析及多因素风险竞争分析

变量	单因素 Cox 分析		多因素 Cox 分析		多因素竞争性 Cox 分析	
	HR (95%CI)	P	HR (95%CI)	P	HR (95%CI)	P
年龄					0.986 (0.936~1.037)	0.580
>80 岁						
65~80 岁	0.770(0.687~0.863)	<0.001	0.990(0.879~1.115)	0.869		
≤65 岁	0.635(0.564~0.717)	<0.001	0.960(0.841~1.095)	0.542		
骨转移					1.216 (1.094~1.351)	<0.001
无/未知						
有	1.329(1.207~1.464)	<0.001	1.269(1.147~1.403)	<0.001		
脑转移						
无/未知						
有	1.540(1.237~1.917)	<0.001	1.379(1.099~1.73)	<0.001		
化疗					0.478 (0.437~0.523)	<0.001
无/未知						
有	0.365(0.333~0.401)	<0.001	0.379(0.342~0.419)	<0.001		
性别						
女性						
男性	0.933(0.848~1.026)	0.154				
分级					0.938 (0.887~0.991)	0.023
高分化:Ⅰ级	1.035(0.554~1.935)	0.915				
中分化:Ⅱ级						
差分化:Ⅲ级	1.156(0.862~1.550)	0.334				
未分化:Ⅳ级	0.982(0.744~1.297)	0.900				
未知	1.205(0.917~1.585)	0.181				
组织学类型					0.991 (0.935~1.050)	0.760
腺癌						
其他	1.681(1.274~2.217)	<0.001	1.278(0.958~1.704)	0.095		
鳞状细胞癌(SCC)	1.896(1.354~2.654)	<0.001	1.761(1.247~2.486)	<0.001		
移行细胞癌(TCC)	1.235(0.950~1.606)	0.115	1.154(0.880~1.515)	0.300		
肝转移						
未知						
有	1.691(1.531~1.868)	<0.001	1.492(1.343~1.656)	<0.001		
婚姻状况					0.983 (0.914~1.057)	0.640
已婚	0.747(0.645~0.865)	<0.001	0.882(0.759~1.025)	0.102		
丧偶/离婚/其他	0.951(0.821~1.102)	0.507	0.965(0.831~1.121)	0.639		
家庭收入中位数						
<54 999 美元					1.016 (0.971~1.063)	0.490
55 000~74 999 美元	0.975(0.842~1.129)	0.733				
75 000~94 999 美元	0.925(0.797~1.073)	0.305				
>95 000 美元	0.999(0.857~1.164)	0.988				
N 分期					1.050 (1.017~1.083)	<0.001
N0						
N1	1.045(0.895~1.221)	0.574	1.097(0.936~1.286)	0.251		
N2	1.037(0.910~1.181)	0.585	1.168(1.021~1.338)	0.024		
N3	1.044(0.862~1.265)	0.658	1.238(1.018~1.506)	0.033		

续表						
变量	单因素 Cox 分析		多因素 Cox 分析		多因素竞争性 Cox 分析	
	HR (95%CI)	P	HR (95%CI)	P	HR (95%CI)	P
NX	1.195(1.061~1.346)	<0.001	1.011(0.885~1.155)	0.869		
种族					1.047 (0.971~1.129)	0.230
黑种人						
其他	0.730(0.581~0.917)	<0.001	0.771(0.611~0.972)	0.028		
白种人	0.937(0.810~1.085)	0.385	0.945(0.813~1.098)	0.457		
放疗						
无/未知						
有	0.799(0.699~0.914)	<0.001	0.917(0.795~1.057)	0.231		
手术方式					0.953 (0.878~1.035)	0.250
根治性膀胱切除术						
无/未知	2.545(1.979~3.273)	<0.001	1.906(1.459~2.491)	<0.001		
非根治性膀胱切除术	1.403(1.097~1.795)	<0.001	1.396(1.075~1.812)	0.010		
淋巴结手术					0.161 (0.017~1.497)	0.110
无/未知						
有	3.227(1.039~10.02)	0.040	4.265(1.349~13.483)	0.010		
T 分期					1.020 (0.983~1.060)	0.290
T0						
T1	0.800(0.546~1.173)	0.254	1.109(0.741~1.660)	0.616		
T2	0.684(0.471~0.991)	0.040	1.074(0.723~1.597)	0.724		
T3	0.644(0.430~0.965)	0.030	1.044(0.686~1.589)	0.840		
T4	0.857(0.586~1.253)	0.426	1.234(0.831~1.832)	0.298		
TX	1.041(0.717~1.514)	0.831	1.095(0.744~1.610)	0.646		



A.Lasso 回归中 lambda 与交叉验证误差 (CVM) 的关系图;B.Lasso 回归中不同 lambda 下各变量系数路径图。

图1 LASSO 回归筛选特征变量

2.6 随机森林模型特征重要性评分的 SHAP 图构建 我们使用 SHAP 模型量化评估各特征对 BCLM 患者预后预测的重要性,结果显示,在6个月的短期生存预测中,肿瘤分级是最具影响力的因

素,其次是化疗。对于1年和3年生存预测,肿瘤分级依然是主导因素,但次要影响因素有所变化:1年生存预测中,骨转移成为主要因素,而3年生存预测中,化疗的影响更为显著(图4)。

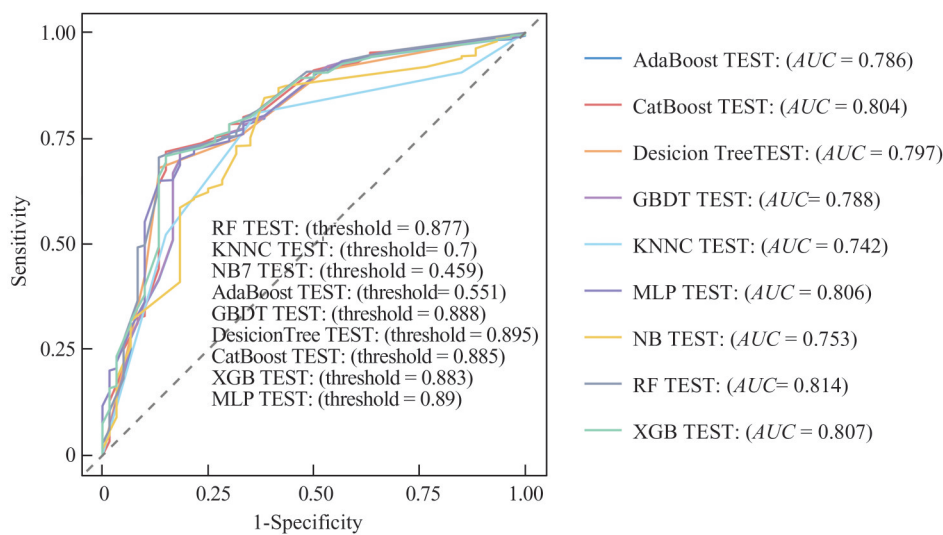


图2 9种机器学习模型ROC曲线

注:AUC.受试者工作特征曲线下面积;MLP.多层感知器;XGB.极端梯度提升;CatBoost.类别提升;DecisionTree.决策树;GBDT.梯度提升决策树;AdaBoost.自适应提升;NB.朴素贝叶斯;KNNCT.K近邻分类器;RF.随机森林。

表3 机器学习算法建立的预测模型在测试数据上的表现

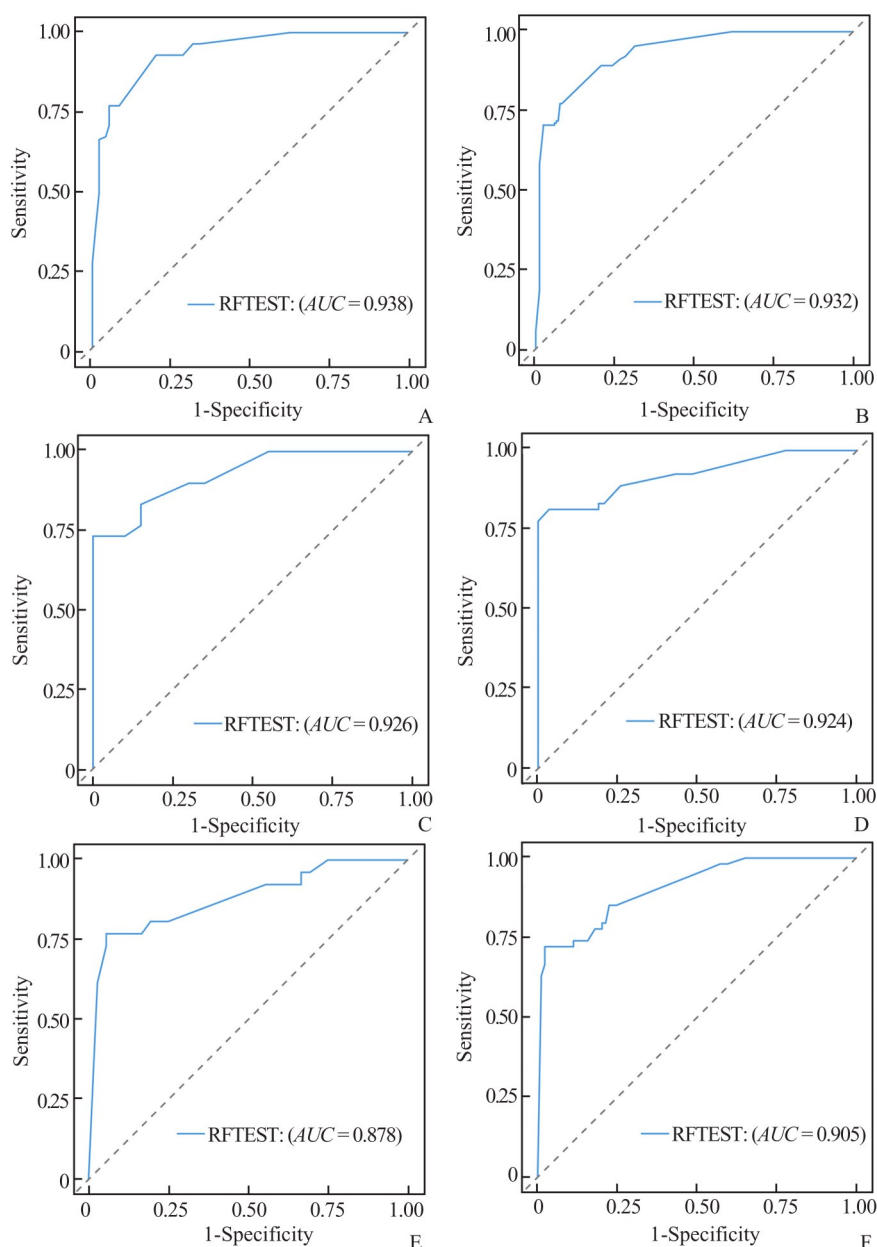
模型	准确率	发生率	召回率	F1-Score	马修斯相关系数	AUC	精确率	特异性	漏报率
MLP	0.899 5	0.908 7	0.986 6	0.946 9	0.046 9	0.806 3	0.910 4	0.033 3	0.013 4
XGB	0.904 1	0.908 7	0.995 0	0.949 6	-0.021 5	0.807 1	0.908 3	0.000 0	0.005 0
CatBoost	0.908 7	0.908 7	1.000 0	0.952 2	NA	0.804 1	0.908 7	0.000 0	0.000 0
DesicionTree	0.908 7	0.908 7	1.000 0	0.952 2	NA	0.796 8	0.908 7	0.000 0	0.000 0
GBDT	0.904 1	0.908 7	0.995 0	0.949 6	-0.021 5	0.788 3	0.908 3	0.000 0	0.005 0
AdaBoost	0.898 0	0.908 7	0.984 9	0.946 1	0.041 0	0.786 5	0.910 2	0.033 3	0.015 1
NB	0.622 5	0.908 7	0.606 4	0.744 9	0.226 8	0.752 7	0.965 3	0.783 3	0.393 6
KNNCT	0.837 1	0.908 7	0.906 2	0.910 0	0.054 2	0.741 5	0.913 9	0.150 0	0.093 8
RF	0.908 7	0.908 7	1.000 0	0.952 2	NA	0.814 4	0.908 7	0.000 0	0.000 0
均值	0.865 7	0.908 7	0.941 6	0.922 6	0.054 3	0.788 6	0.915 8	0.111 1	0.058 4

注:AUC.受试者工作特征曲线下面积;F1-Score.精确率与召回率的调和均值;MLP.多层感知器;XGB.极端梯度提升;CatBoost.类别提升;DecisionTree.决策树;GBDT.梯度提升决策树;AdaBoost.自适应提升;NB.朴素贝叶斯;KNNCT.K近邻分类器;RF.随机森林。

3 讨论

膀胱癌(BC)常见的转移部位包括骨、肺、脑和肝,其中肺转移尤为常见。尽管已有基于SEER数据的研究^[9]构建了生存预测列线图模型,但其预测准确率普遍低于70%。本研究是迄今为止针对膀胱癌肺转移(BCLM)患者的最大规模临床特征和预后分析,首次使用人工智能模型实现了高精度生存预测。研究结果表明,BCLM患者的生存预后受多种因素影响,6个月、1年和3年生存率的主要预测因素包括肿瘤分级、化疗、原发肿瘤手术等。在6个月生存预测中,AUC达到0.938(测试集)和0.932(训练集);1年生存预测中,AUC为0.925 8(测试集)和0.942 1(训练集);3年生存预测中,AUC为0.877 7(测试集)和0.905 4(训练集),这些结果显

示出随机森林模型在生存预测中的高性能。研究还表明,相关因素如“其他”种族、较低的N分期、接受全膀胱切除术和化疗与较长的生存期相关,早期干预和积极治疗对改善预后至关重要。图4显示,在短期生存预测(6个月和1年)中,化疗和原发部位手术对患者生存有显著影响,显示出在早期干预中这两项措施的关键作用。化疗通过抑制癌细胞增殖和诱导免疫反应,有效控制肿瘤生长并防止转移。原发部位手术如根治性膀胱切除术则通过去除主要肿瘤,减少癌细胞扩散机会,改善免疫监视,进而提高患者生存期。这些干预措施的生物学机制表明,早期治疗能有效抑制肿瘤扩散,改善短期生存率。相反,多部位转移(如骨、肝、脑)、鳞状细胞癌(SCC)组织学类型、更高的N分期(N2、N3)及未进行手术的患者生存预期显著降低。



A. 6个月生存期测试集随机森林预测模型图;B. 6个月生存期训练集随机森林预测模型图;C. 1年生存期测试集随机森林预测模型图;
D. 1年生存期训练集随机森林预测模型图;E. 3年生存期测试集随机森林预测模型图;F. 3年生存期训练集随机森林预测模型图。

图3 测试集与训练集6个月、1年、3年随机森林预测模型图

骨、肝、脑转移被确定为独立危险因素,与以往研究^[13-16]一致。多部位转移表明肿瘤细胞已经突破局部免疫屏障,通过血液或淋巴系统扩散到多个器官,导致更难控制的广泛转移。骨转移通过破骨细胞的活化,促使肿瘤细胞在骨骼中生长,肝转移则利用肝脏的血供和特有微环境促进癌细胞的存活和增殖,脑转移则通过突破血脑屏障增加了治疗的难度。鳞状细胞癌(SCC)通常伴随更强的侵袭性和更高的转移潜力,因为SCC细胞在转移过程中更易发生上皮-间质转化(EMT),增强了其迁移和侵

袭能力。此外,较高的N分期反映癌细胞已广泛侵袭淋巴结并有较高的远处转移风险。未进行手术的患者则因未能及时去除原发肿瘤,导致残留癌细胞的扩散和转移,使预后更差。多因素竞争风险分析显示,肿瘤分级未知与结局事件存在统计学关联($HR=0.938, 95\%CI:0.887\sim0.991$)。但未知分级并非明确的病理分级,而是肿瘤分级信息未明确记录或不可获得的集合类别,根据SEER数据库中GRADE PATHOLOGICAL(2018+)与GRADE RECODE(THRU 2017)划分病理分级的标准,肿

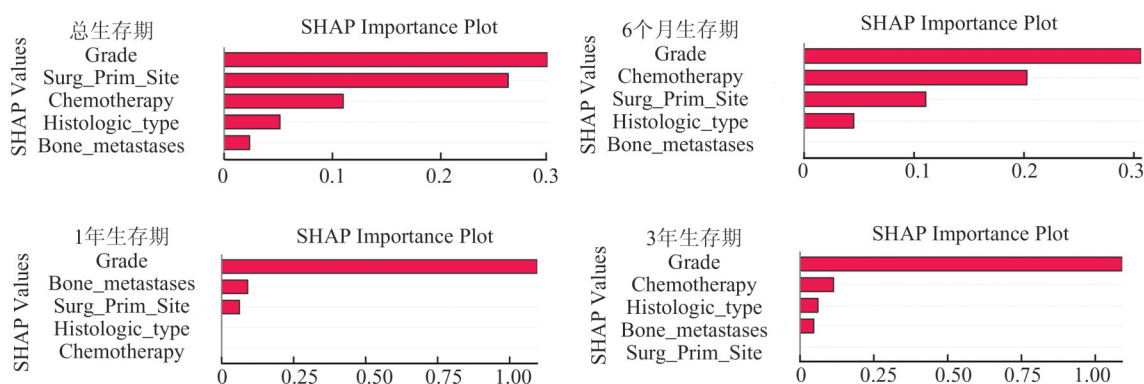


图 4 随机森林模型特征重要性评分的 SHAP 图

瘤分级(Grade)“未知”组包括:①编码为9用于无法获取或确定肿瘤分级的情况,例如原发肿瘤分级缺失、未进行切除或非标准手术(如TURB),以及治疗后病理分级不明确。此外,当多个肿瘤使用不同分级系统,或仅有一个分级但无法判断类型(临床、病理、术后)时,也应使用代码9;②空白(未作记录的);③Grad L/H(尿路上皮癌(Urothelial cancers)使用WHO/ISUP系统进行分级,其中L(低级)表示肿瘤分化良好,通常预后较好;H(高级)则指肿瘤分化差,预后较差)。因此,该结果不应被解释为未知分级具有真实的生物学保护作用。我们对肿瘤分级(Grade)为“未知”的患者进行了多因素Cox分层分析,旨在探讨三个亚组与膀胱癌肺转移患者OS之间的关系。在Cox分层分析中,空白组即未明确肿瘤分级的患者死亡风险显著增加($HR=1.511$, $95\%CI: 1.343\sim1.700$, $P<0.001$) (见补充表2,扫描本文OSID可见)。需要说明的是,Blank代表SEER数据库肿瘤分级变量中的空白或不可获得类别,即肿瘤分级信息未明确记录或不可获得,这反映出治疗过程中的混乱和其他潜在的不利因素。这种不确定性可能导致临床决策的延误或错误,从而增加了患者的OS风险。在膀胱癌肺转移患者的治疗过程,肿瘤分级的准确性至关重要,肿瘤分级反映了肿瘤细胞的分化程度。高分级(差分化)的肿瘤往往具有更高的增殖能力和转移潜力,高分级肿瘤可能通过激活特定的信号通路促进癌细胞的增殖和转移。临床医生需清晰了解患者的病理分级,以确保恰当的管理,避免对肿瘤侵袭性作出错误判断或实施不当治疗^[17-18]。

此外,本研究证实了原发肿瘤手术对BCLM患者OS的显著影响。既往研究^[19-23]表明,肺转移灶切除后,BCLM患者的中位生存期可延长至29~42个月,尤其是肺转移灶小于3 cm的患者,其无进展生

存率高达65%,显示出显著的临床获益。在化疗方面,顺铂仍然是膀胱癌最有效的单药治疗方案,而根治性膀胱切除术(RC)在延长患者生存期方面优于单独放疗(RT),但对健康状况不佳的患者,外部束放疗(EBRT)仍是有效的替代方案^[24],这一结论与本研究一致,放疗对BCLM患者的生存影响较为有限。

尽管本研究发现了一些有意义的结果,但仍存在一定的局限性。首先,SEER数据库仅包含2010年后肿瘤亚型和转移部位的数据,样本量较为有限;其次,尽管SEER数据具有代表性,但由于种族差异,结果可能不完全适用于亚洲人群;并且数据库缺乏关于疾病复发和进一步转移的信息,这可能影响部分分析结果的准确性。尽管如此,这些局限性并未削弱本研究在BCLM患者管理中的潜在价值,反而为未来的研究指明了方向。

本研究构建了高精度的BCLM生存预测模型(6个月 $AUC=0.938$),为精准医学在肿瘤治疗中的应用提供支持。整合肿瘤分级、远处转移、化疗和手术方式等关键变量,该模型可量化死亡风险,指导个性化治疗,提高短期及长期生存率。模型具备良好的适用性和推广潜力,跨医疗背景提供标准化评估,助力早期干预和精准治疗。研究不仅推动了人工智能在肿瘤治疗中的应用,也为未来开发更广泛的预后模型奠定基础,提升临床决策的科学性与效率。



开放科学(资源服务)标识码(OSID):

参考文献

- [1] SUNG H, FERLAY J, SIEGEL R L, et al. Global cancer statistics 2020: GLOBOCAN estimates of inci-

- dence and mortality worldwide for 36 cancers in 185 countries[J]. *CA Cancer J Clin*, 2021, 71(3): 209-249.
- [2] ANTONI S, FERLAY J, SOERJOMATARAM I, et al. Bladder cancer incidence and mortality: A global overview and recent trends[J]. *Eur Urol*, 2017, 71(1): 96-108.
- [3] SHINAGARE A B, RAMAIYA N H, JAGANNA-THAN J P, et al. Metastatic pattern of bladder cancer: correlation with the characteristics of the primary tumor [J]. *AJR Am J Roentgenol*, 2011, 196(1): 117-122.
- [4] GORDON ROBERTSON A, KIM J, AL-AHMADIE H, et al. Comprehensive molecular characterization of muscle-invasive bladder cancer[J]. *Cell*, 2017, 171(3): 540-556. e25.
- [5] KAMOUN A, DE REYNIÈS A, ALLORY Y, et al. A consensus molecular classification of muscle-invasive bladder cancer[J]. *Eur Urol*, 2020, 77(4): 420-433.
- [6] GALSKEY M D, MOSHIER E, KREGES S, et al. Nomogram for predicting survival in patients with unresectable and/or metastatic urothelial cancer who are treated with cisplatin-based chemotherapy[J]. *Cancer*, 2013, 119(16): 3012-3019.
- [7] ZAGHLOUL M S, CHRISTODOULEAS J P, SMITH A, et al. Adjuvant sandwich chemotherapy plus radiotherapy *vs* adjuvant chemotherapy alone for locally advanced bladder cancer after radical cystectomy: A randomized phase 2 trial[J]. *JAMA Surg*, 2018, 153(1): e174591.
- [8] LENIS A T, LEC P M, CHAMIE K. Bladder cancer [J]. *JAMA*, 2020, 324(19): 2006.
- [9] LIU L, SUN F Z, ZHANG P Y, et al. Development and validation a model for predicting overall survival of bladder cancer with lung metastasis: A population-based study[J]. *Eur J Med Res*, 2023, 28(1): 279.
- [10] SWANSON K, WU E, ZHANG A, et al. From patterns to patients: Advances in clinical machine learning for cancer diagnosis, prognosis, and treatment[J]. *Cell*, 2023, 186(8): 1772-1791.
- [11] TERRANOVA N, VENKATAKRISHNAN K. Machine learning in modeling disease trajectory and treatment outcomes: An emerging enabler for model-informed precision medicine[J]. *Clin Pharmacol Ther*, 2024, 115(4): 720-726.
- [12] IGNATENKO V, SURKOV A, KOLTCOV S. Random forests with parametric entropy-based information gains for classification and regression problems[J]. *PeerJ Comput Sci*, 2024, 10: e1775.
- [13] SHI S B, PENG G B, LUO L H, et al. Predictive nomograms for risk and prognostic factors in metastatic bladder cancer: A population-based study[J]. *Transl Cancer Res*, 2023, 12(12): 3284-3302.
- [14] CHEN C, HU L K, CHEN Y, et al. The prognostic value of histological subtype in patients with metastatic bladder cancer[J]. *Oncotarget*, 2017, 8(17): 28408-28417.
- [15] FAN Z Y, HUANG Z H, HUANG X H. Bone metastasis in renal cell carcinoma patients: Risk and prognostic factors and nomograms[J]. *J Oncol*, 2021, 2021: 5575295.
- [16] HU C, YANG J X, HUANG Z H, et al. Diagnostic and prognostic nomograms for bone metastasis in hepatocellular carcinoma[J]. *BMC Cancer*, 2020, 20(1): 494.
- [17] COMPÉRAT E M, BURGER M, GONTERO P, et al. Grading of urothelial carcinoma and the new "world health organisation classification of tumours of the urinary system and male genital organs 2016"[J]. *Eur Urol Focus*, 2019, 5(3): 457-466.
- [18] LOKESHWAR S D, RUIZ-CORDERO R, HUPE M C, et al. Impact of 2004 ISUP/WHO classification on bladder cancer grading[J]. *World J Urol*, 2015, 33(12): 1929-1936.
- [19] STENZL A, COWAN N C, DE SANTIS M, et al. Treatment of muscle-invasive and metastatic bladder cancer: Update of the EAU guidelines[J]. *Eur Urol*, 2011, 59(6): 1009-1018.
- [20] MATSUGUMA H, YOSHINO I, ITO H, et al. Is there a role for pulmonary metastasectomy with a curative intent in patients with metastatic urinary transitional cell carcinoma? [J]. *Ann Thorac Surg*, 2011, 92(2): 449-453.
- [21] SIEFKER-RADTKE A O, WALSH G L, PISTERS L L, et al. Is there a role for surgery in the management of metastatic urothelial cancer? The M. D. Anderson experience[J]. *J Urol*, 2004, 171(1): 145-148.
- [22] ABE T, KITAMURA H, OBARA W, et al. Outcome of metastasectomy for urothelial carcinoma: A multi-institutional retrospective study in Japan[J]. *J Urol*, 2014, 191(4): 932-936.
- [23] ABE T, SHINOHARA N, HARABAYASHI T, et al. Impact of multimodal treatment on survival in patients with metastatic urothelial cancer[J]. *Eur Urol*, 2007, 52(4): 1106-1113.
- [24] SHELLEY M D, BARBER J, WILT T, et al. Surgery versus radiotherapy for muscle invasive bladder cancer[J]. *Cochrane Database Syst Rev*, 2002(1): CD002079.

(2024-11-06 收稿)

编辑 沈建国