·论著·一次研究·

不同机器学习算法在直肠鳞状细胞癌预后模型构建中的应用：基于SEER数据库

黄 诗，林丹丹

武汉大学人民医院胃肠肿瘤II科（武汉 430060）

【摘要】**目的** 探讨不同机器学习模型在预测直肠鳞状细胞癌（rSCC）患者 5 年生存状态中的应用价值。**方法** 利用 SEER*Stat 软件，收集 SEER 数据库中 2004 年至 2015 年确诊的 rSCC 患者资料，按 7∶3 的比例将患者随机分为训练集和验证集。分别使用极端梯度提升、随机森林、支持向量机和 k 近邻算法在训练集上构建模型，使用受试者工作特征曲线的曲线下面积（AUC），校准曲线、决策曲线评估模型的预测能力。对表现最佳的模型，使用 SHAP 算法明确变量对模型的贡献。**结果** 共纳入 833 例 rSCC 患者，其中训练集 584 例，验证集 249 例。基于年龄、性别、种族、婚姻状态、T 分期、N 分期、M 分期、手术、化疗和放疗 10 个变量，构建了多个机器学习模型。在验证集中，极端梯度提升模型表现最佳[AUC=0.758，95%CI（0.696，0.820）]，具有一定预测能力，其校准度良好，决策曲线显示较高的临床应用价值。SHAP 分析显示，T 分期对极端梯度提升模型的决策贡献最大，放疗对模型预测贡献最小。**结论** 本文构建了一种可解释的极端梯度提升模型，可辅助医生判断 rSCC 患者的 5 年生存状态及治疗效果，有助于制定个性化的诊疗方案。

【关键词】直肠鳞状细胞癌；机器学习；预测模型；预后

【中图分类号】R 735.3+7 【文献标识码】A

Application of different machine learning algorithms in the construction of prognostic models for rectal squamous cell carcinoma: based on SEER database

HUANG Shi, LIN Dandan

Department II of Gastrointestinal Oncology, Renmin Hospital of Wuhan University, Wuhan 430060, China
Corresponding author: LIN Dandan, Email: lindandan@whu.edu.cn

【Abstract】**Objective** To investigate the application value of different machine learning models in predicting 5-year survival in patients with rectal squamous cell carcinoma (rSCC). **Methods** Data from patients diagnosed with rSCC between 2004 and 2015 were collected using SEER*Stat software in SEER database. Patients were randomly divided into a training set and a validation set in a 7 : 3 ratio. Models were constructed using extreme gradient boosting (XGBoost), random forest, support vector machine, and k-nearest neighbor algorithms on the training set. The predictive ability of the models was evaluated using the area under the receiver operating characteristic curve (AUC), calibration curves, and decision curves. The SHAP algorithm was used to identify the contribution of variables to the model for the best-performing model. **Results** A total of 833 patients with rSCC were included, including 584 in the training set and 249 in the

DOI: 10.12173/j.issn.1004-5511.202410124
基金项目：国家自然科学基金面上项目（32270951）
通信作者：林丹丹，博士，主任医师，博士研究生导师，Email：lindandan@whu.edu.cn

validation set. Multiple machine learning models were constructed based on 10 variables: age, sex, race, marital status, T stage, N stage, M stage, surgery, chemotherapy, and radiotherapy. In the validation set, the XGBoost model performed best [AUC=0.758, 95%CI (0.696, 0.820)], demonstrating moderate predictive ability and good calibration. The decision curves demonstrated high clinical value. SHAP analysis showed that T stage contributed most to the XGBoost model's decision-making, while radiotherapy contributed least. Conclusion This study constructed an interpretable XGBoost model that can assist physicians in assessing the 5-year survival and treatment efficacy of rSCC patients and in developing personalized treatment plans.

【Keywords】Rectal squamous cell carcinoma; Machine learning; Prediction model; Prognosis

数据显示，结直肠癌在所有恶性肿瘤中发病率居全球第三位，死亡率居全球第二位[1]。直肠癌作为结直肠癌的一部分，包括腺癌、腺鳞癌、杯状细胞腺癌、淋巴瘤、髓样癌和鳞癌等多种组织学类型，其中腺癌最为常见。而直肠鳞状细胞癌（rectal squamous cell carcinoma，rSCC）则是一种罕见的亚型，仅占直肠癌的 0.1%~0.3%[2]。rSCC 的发病机制目前尚未完全明确，可能与多种因素相关，包括吸烟、盆腔放疗史、慢性直肠炎症、HPV 感染等[3-4]。目前关于 rSCC 的研究多为个案报道[5-6]，缺乏大样本临床研究，且尚未形成针对 rSCC 的统一治疗指南。部分研究认为，手术在 rSCC 患者的治疗中至关重要，是提高患者生存率的主要手段[7-8]。然而，部分研究则主张 rSCC 的治疗应参照肛门鳞状细胞癌的治疗模式，将放化疗作为首选的治疗方法，并建议手术作为局部复发或放化疗无效患者的辅助或补救措施[9-11]。与直肠腺癌相比，rSCC 的预后更差[12]。因此，深入探讨影响 rSCC 患者预后的因素，尤其是对提高 rSCC 患者长期预后具有重要的临床意义。

美国国家癌症研究所的监测、流行病学和最终结果（Surveillance，Epidemiology，and End Results，SEER）数据库提供了大量癌症患者的相关信息，已经被广泛应用于乳腺癌[13]、肺癌[14]、直肠癌[15]等多种癌症的预后研究。鉴于 5 年生存率是肿瘤预后评估的重要指标，本研究利用 SEER 数据库的 rSCC 患者数据，使用多种机器学习模型构建 rSCC 患者 5 年生存率的预测模型，旨在评估其长期生存情况并探索相关预后因素，为 rSCC 患者的临床诊疗决策提供参考。

1 资料与方法

1.1 研究对象

本研究基于 SEER 数据库（8.4.2 版本）回顾性收集 2004—2015 年经病理明确诊断为 rSCC 患者的临床资料。纳入标准：①经病理学确诊为 rSCC，组织学编码为 C20.9，形态学编码为 8052、8070–8076、8083 及 8084（依据 ICD–O–3 标准）；②生存时间≥ 1 个月（生存时间定义为从诊断日期至全因死亡或末次随访的时间间隔）；③年龄≥ 20 岁。排除标准：①未经病理学确诊的原发性 rSCC 病例；②生存时间未记录；③缺失关键临床数据（如婚姻状态、种族、TNM 分期、治疗信息等）；④多发肿瘤或合并其他系统恶性肿瘤。本研究基于开放获取的 SEER 数据库数据，无需伦理审批。

初步筛选出 2 322 名患者。根据纳入和排除标准进一步筛选，最终纳入 833 名患者用于分析。

1.2 数据收集

使用 SEER*Stat 8.4.2 软件，选取 Incidence–SEER Research Data，17 Registries，Nov 2023 Sub（2000—2021）数据集，末次随访时间为 2021 年 12 月 31 日。纳入患者的信息包括性别、年龄、种族、婚姻状态、是否接受手术、是否接受化疗、是否接受放疗、生存月数、生存状态以及肿瘤分化程度等临床变量。为了评估患者的预后，定义总生存期（overall survival，OS）为从确诊日期起至因任何原因死亡的时间。种族分为白种人、黑种人和其他人种（包括美洲印第安人、阿拉斯加原住民、亚裔或太平洋岛民）。分化程度分为高分化（Grade Ⅰ）、中分化（Grade Ⅱ）、低分化（Grade Ⅲ）和未分化（Grade Ⅳ）。肿瘤的 TNM 分期则依据美国癌症联合委员会（American Joint Committee on Cancer，AJCC）第六版的标准进行分期。

1.3 统计学分析

采用 R 4.3.3 软件进行统计学分析。分类变量采用例数和百分比（n，%）表示，组间比较

采用 χ^2 检验。将纳入的患者按 7∶3 的比例随机分为训练集与验证集。训练集用于模型的构建和训练，验证集用于评估和解释模型的性能。本研究的主要终点为 rSCC 患者在确诊后 5 年的生存状态，即是否在 5 年时存活。为筛选进入模型的候选变量，采用 Kaplan–Meier 方法分别计算训练集中每个候选变量（如种族、性别、年龄等）与 5 年生存状态的相关性，并通过 Log–rank 检验比较不同分组的生存差异。为控制多重比较引起的假阳性风险，本研究对 Log–rank 检验 P 值进行了 Benjamini–Hochberg 校正，设定错误发现率阈值为 0.05，校正后具有统计显著性的变量用于后续模型构建。

使用 4 种机器学习的方法进行模型构建，包括极端梯度提升（extreme gradient boosting，XGBoost）、随机森林（random forest，RF）、支持向量机（support vector machines，SVM）、k 近邻算法（k–nearest neighbor，KNN）。通过受试者工作特征（receiver operating characteristic，ROC）曲线，并计算曲线下面积（area under the curve，AUC）评估模型的区分度。使用校准曲线评估模型的校准度，使用决策曲线分析评估模型的临床适用性。为进一步解释模型的预测结果，采用 SHAP（SHapley Additive exPlanations）算法明确各特征对 5 年生存状态预测结果的贡献。以 $P < 0.05$ 为差异具有统计学意义。

2 结果

2.1 一般情况

共纳入 833 例 rSCC 患者，训练集 584 例（70.1%），验证集 249 例（29.9%）。性别以女性为主（71.5%），种族以白种人为主（87.0%）。在肿瘤学特征方面，T 分期以 T1 为主（37.7%），N 分期以 N0 为主（68.3%），M 分期以 M0 为主（91.5%）；肿瘤分化程度以 III 期为主（49.1%）。治疗情况方面，手术治疗率为 31.2%，而接受放疗和化疗的患者分别占 83.4% 及 81.0%。训练集与验证集各项基线特征均无统计学差异（$P > 0.05$），见表 1。

2.2 生存分析

生存曲线显示，随着时间推移，患者的生存概率显著下降，尤其在早期阶段下降较为明显，此后曲线趋于平稳，5 年生存率为 63.01%（图 1）。

在训练集中，Kaplan–Meier 生存曲线分析显示年龄、性别、种族、婚姻状态、T 分期、N 分期、M 分期、手术、化疗和放疗是患者预后的影响因素（表 2）。经过 Benjamini–Hochberg 校正后，这些变量与患者 5 年生存状态仍显著相关。

2.3 模型构建与评估

基于生存分析中的 10 个显著性变量和训练集样本构建了 4 个机器学习模型，并比较了模型在训练集和验证集中的表现。结果显示，XGBoost

表1 rSCC患者的基线特征情况（n，%）

Table 1. Baseline characteristics of rSCC patients (n, %)

变量	总体（n=833）	训练集（n=584）	验证集（n=249）	χ^2值	P值
性别				0.05	0.822
男性	237（28.5）	168（28.8）	69（27.7）		
女性	596（71.5）	416（71.2）	180（72.3）		
年龄（岁）				0.63	0.482
＜60	399（47.9）	274（46.9）	125（50.2）		
≥60	434（52.1）	310（53.1）	124（49.8）		
种族				0.22	0.894
白种人	725（87.0）	509（87.2）	216（86.7）		
黑种人	97（11.7）	68（11.6）	29（11.6）		
其他人种	11（1.3）	7（1.2）	4（1.6）		
婚姻状态				1.02	0.796
结婚	384（46.1）	267（45.7）	117（47.0）		
离婚	138（16.6）	101（17.3）	37（14.9）		
丧偶	108（12.9）	73（12.5）	35（14.0）		
未婚	203（24.4）	143（24.5）	60（24.1）		

续表1

变量	总体（n=833）	训练集（n=584）	验证集（n=249）	χ^2值	*P*值
T分期				0.61	0.894
T1	314（37.7）	225（38.5）	89（35.7）		
T2	140（16.8）	96（16.4）	44（17.7）		
T3	261（31.3）	181（31.0）	80（32.1）		
T4	118（14.2）	82（14.0）	36（14.5）		
N分期				2.87	0.239
N0	569（68.3）	393（67.3）	176（70.7）		
N1	209（25.1）	147（25.2）	62（24.9）		
N2	55（6.6）	44（7.5）	11（4.4）		
M分期				0.54	0.460
M0	762（91.5）	531（90.9）	231（92.8）		
M1	71（8.5）	53（9.1）	18（7.2）		
分化程度				1.57	0.666
Grade I	63（7.6）	43（7.4）	20（8.0）		
Grade II	337（40.4）	243（41.6）	94（37.8）		
Grade III	409（49.1）	283（48.5）	126（50.6）		
Grade IV	24（2.9）	15（2.6）	9（3.6）		
手术				0.47	0.491
否	573（68.8）	397（68.0）	176（70.7）		
是	260（31.2）	187（32.0）	73（29.3）		
放疗				0.75	0.387
否	138（16.6）	92（15.8）	46（18.5）		
是	695（83.4）	492（84.2）	203（81.5）		
化疗				1.47	0.226
否	158（19.0）	104（17.8）	54（21.7）		
是	675（81.0）	480（82.2）	195（78.3）		

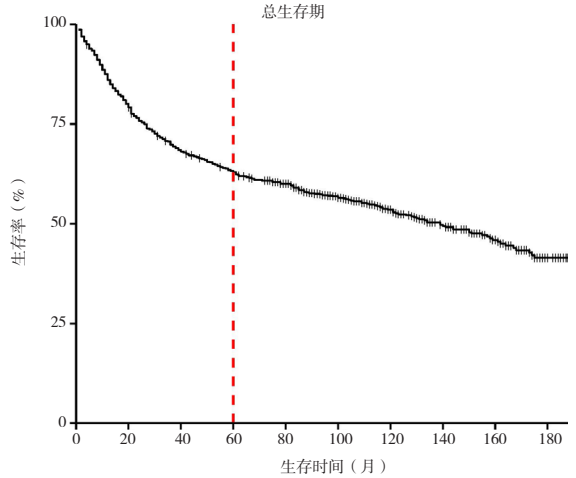

图1 833例rSCC患者的总体生存曲线

Figure 1. Overall survival curve for the 833 rSCC patients

模型在训练集和验证集中均表现出最佳的预测效果（图 2 至图 4）。

ROC 曲线显示，在训练集中，XGBoost 模型[AUC=0.792，95%CI（0.755，0.830）]和RF 模型[AUC=0.792，95%CI（0.754，0.831）]的 AUC 值高于 SVM 模型[（AUC=0.757，95%CI（0.717，0.797）]和KNN模型[（AUC=0.727，95%CI（0.685，0.768）]；在验证集中，XGBoost 模型[AUC=0.758，95%CI（0.696，0.820）]分别优于 RF 模型[AUC=0.750，95%CI（0.688，0.811）]、SVM 模型[AUC=0.707，95%CI（0.640，0.772）]和KNN模型[(AUC=0.717,95%CI(0.652,0.873)]，表明 XGBoost 模型在预测 rSCC 患者 5 年生存状态方面具有较好的区分度（图 2）。

校准曲线分析结果显示，无论在训练集还是验证集中，XGBoost 模型的校准曲线最接近理想的 45 度对角线，表明其预测概率与实际观测值之间具有良好的一致性。RF 模型的校准曲线虽表现较为稳定，但整体略逊于 XGBoost。表明 XGBoost 模型在预测 rSCC 患者 5 年生存状态方面具有更好的校准度（图 3）。

表2 不同临床特征分组的rSCC患者Kaplan–Meier生存分析结果表

Table 2. Kaplan–Meier survival analysis results of rSCC patients grouped by different clinical characteristics

变量	HR（95%Cl）	*P*值
性别		
男性	Ref.	
女性	0.72（0.55，0.93）	0.007
年龄（岁）		
＜60	Ref.	
≥60	1.76（1.40，2.22）	＜0.001
种族		
白色	Ref.	
黑色	1.51（1.02，2.24）	0.014
其他	1.32（0.43，4.08）	0.581
婚姻状态		
结婚	Ref.	
离婚	1.35（0.95，1.93）	0.073
丧偶	2.44（1.60，3.71）	＜0.001
未婚	1.58（1.16，2.16）	0.002
T分期		
T1	Ref.	
T2	0.90（0.62，1.31）	0.593
T3	1.57（1.18，2.09）	0.001
T4	2.40（1.61，3.57）	＜0.001
N分期		
N0	Ref.	
N1	1.77（1.33，2.36）	＜0.001
N2	1.55（1.04，2.55）	0.039
M分期		
M0	Ref.	
M1	3.39（1.97，5.81）	＜0.001
分化程度		
I	Ref.	
II	1.51（0.96，2.37）	0.116
III	1.68（1.10，2.56）	0.046
IV	2.09（0.80，5.45）	0.068
手术		
否	Ref.	
是	0.66（0.52，0.84）	0.002
放疗		
否	Ref.	
是	0.59（0.42，0.83）	＜0.001
化疗		
否	Ref.	
是	0.64（0.47，0.88）	0.001

临床决策曲线结果显示，在训练集中，同一风险阈值下 XGBoost 模型和 RF 模型的净收益更高；在验证集中，同一风险阈值下 XGBoost 模型的净收益更高。表明 XGBoost 模型在预测 rSCC 患者 5 年生存状态方面具有更好的临床适用性（图4）。

2.4 XGBoost模型的特征重要性

通过 SHAP 值对 XGBoost 模型中的预测变量进行可视化排序，结果显示，T 分期的 SHAP 值最高，表明其对模型预测结果的影响最为显著；放疗的 SHAP 值最低，其对模型预测的贡献最小（图 5）。

3 讨论

rSCC 是一种罕见的肿瘤亚型，其发病机制尚不明确，与更常见的直肠腺癌在临床特征和预后方面存在显著差异。与直肠腺癌相比，rSCC 在相同分期下的总体生存率更低，其 5 年总体生存率为 48.9%[16]。本研究的生存分析发现，rSCC 患者在前 5 年内的生存率急剧下降，随后趋于平缓。这一趋势表明，rSCC 患者在诊断初期面临较高的死亡风险。本研究以 5 年生存状态作为预测终点，通过 SEER 数据库分析，构建了基于机器学习的 rSCC 患者的预后预测模型，为临床决策提供了新的视角和工具。

本研究显示性别、年龄、种族、婚姻状态、T 分期等是 rSCC 患者预后的影响因素，与既往研究一致[17–19]。本研究进一步确认了 N 分期也与 rSCC 患者生存具有显著相关性，这表明 N 分期可能在预后评估中同样具有重要作用。

现有研究主要依赖传统统计模型，如 Cox 回归模型，在捕捉预后因素之间复杂的交互效应方面存在一定局限。此外，Chiu 等[18]的研究侧重于比较 rSCC 和直肠腺癌的病理学特征及预后因素，并未构建预后模型；Diao 等[19]虽然构建了列线图，但未采用机器学习算法进行复杂关系的识别与预测。因此本研究首次引入 4 种机器学习模型来预测 rSCC 患者的 5 年生存状态，能够更精确地识别和预测这些复杂的交互关系[20]。结果显示在所有模型中，XGBoost 模型表现最佳。XGBoost 模型通过梯度提升决策树框架，优化每棵树的残差[21]，从而更有效地学习复杂的非线性关系和特征之间的交互效应。此外，它还具有强大的正则化机制

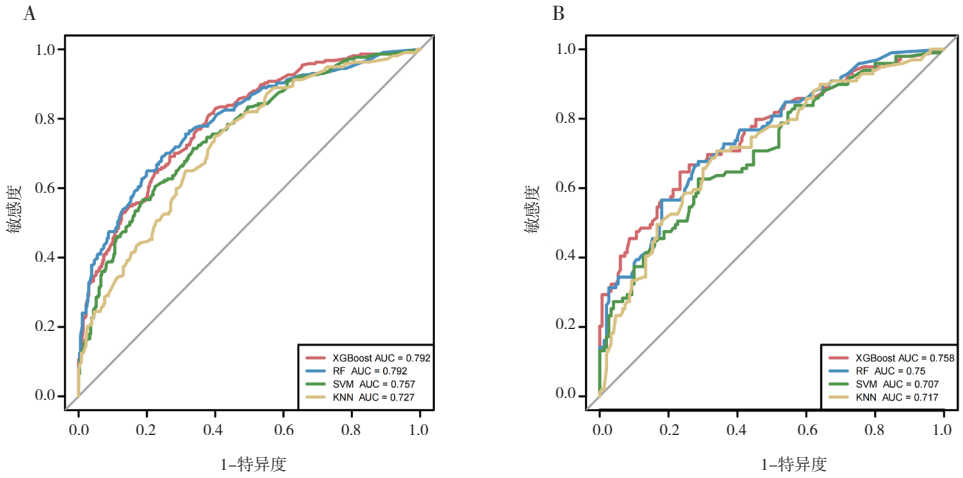

图2 4个模型的ROC曲线

Figure 2. ROC curves of the 4 models

注：A.训练集；B.验证集；XGBoost.极端梯度提升；RF.随机森林；SVM.支持向量机；KNN.k近邻算法。

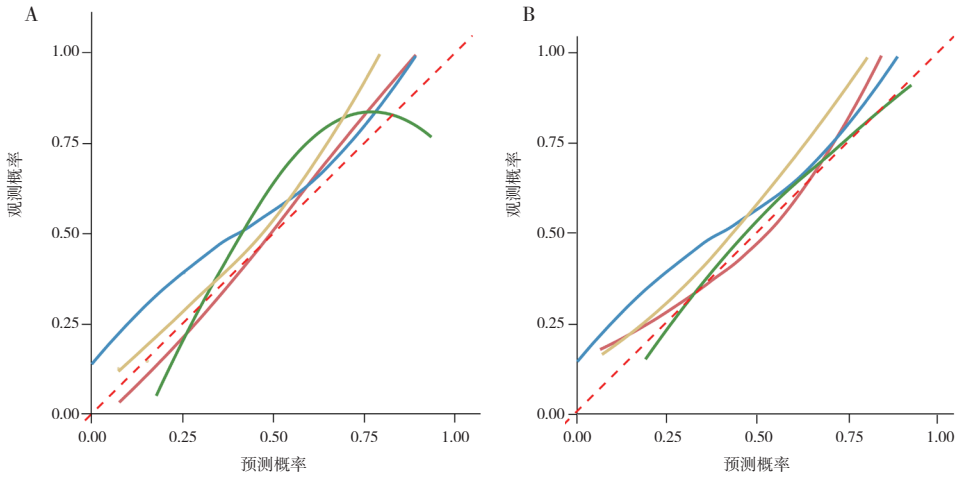

图3 4个模型的校准曲线

Figure 3. Decision curves of the 4 models

注：A.训练集；B.验证集；XGBoost.极端梯度提升；RF.随机森林；SVM.支持向量机；KNN.k近邻算法。

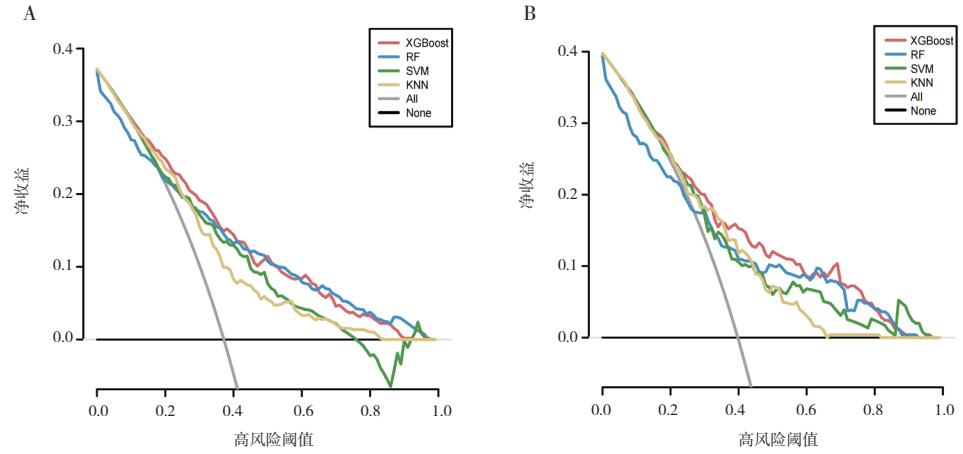

图4 4个模型的临床决策曲线

Figure 4. Decision curves of the 4 models

注：A.训练集；B.验证集；XGBoost.极端梯度提升；RF.随机森林；SVM.支持向量机；KNN.k近邻算法。

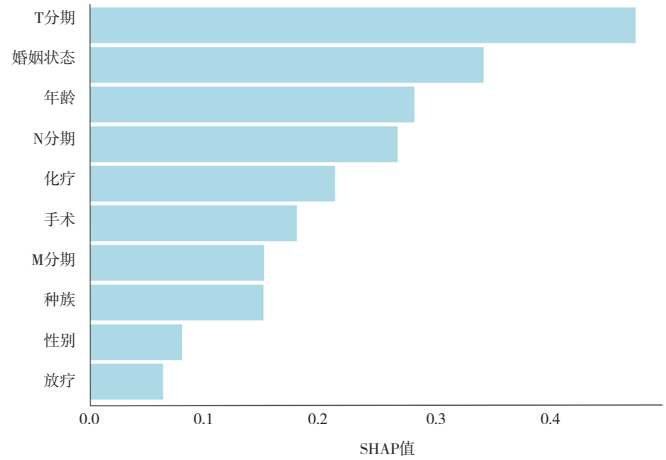

图5 XGBoost模型变量重要性排序

Figure 5. Importance ranking of XGBoost model variables

和丰富的超参数调优选项，能够有效避免过拟合，进一步提高泛化能力。因此，XGBoost 模型已被应用于多种疾病的预后分析[22-25]。在本研究中，XGBoost 模型在训练集和验证集中均具有最高的 AUC 值，且都在 0.75 以上，表现出良好的稳定性和较高的预测准确性。临床决策曲线及校准曲线进一步验证了其在实际应用中的潜在价值。相比之下，RF 模型在训练集和验证集中的表现稳定，但在特征交互效应的捕捉方面略逊于 XGBoost 模型。SVM 模型和 KNN 模型的表现则相对较弱，尤其是在高风险患者的预测中，校准曲线偏差较大，限制了它们在临床中的实际应用价值。

本研究仍存在一些局限性。首先，样本量相对有限，可能对模型的泛化能力和精度造成一定影响，未来研究应扩大样本规模以进一步优化模型性能。其次，本研究未纳入肿瘤大小、癌胚抗原水平、淋巴结清扫数目、基因突变等可能影响预后的生物标志物，未来研究应引入更多潜在的预后因素，以构建更全面的预测模型。最后，由于本研究为回顾性研究，可能存在选择偏倚，仅在内部数据中验证了模型的有效性，尚未通过多中心或外部数据进行验证，未来可结合多中心数据进行外部验证以提高模型的适用性。

综上所述，本研究基于 SEER 数据库中常见的临床特征，构建了预测 rSCC 患者 5 年生存状态的机器学习模型。其中，XGBoost 模型表现优异，具有较高的临床应用潜力，可帮助临床医生更好地管理 rSCC 患者，推动个体化诊治的发展。但未来仍需通过多中心前瞻性研究进一步验证模型的可靠性与稳定性，以促进其在临床中的广泛应用。

伦理声明：不适用

作者贡献：研究设计：林丹丹、黄诗；数据采集及分析、论文撰写：黄诗；论文审定：林丹丹

数据获取：本研究中使用和（或）分析的数据可在SEER数据库获取（https://seer.cancer.gov/）

利益冲突声明：无

致谢：不适用

参考文献

1 Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries[J]. CA Cancer J Clin, 2024, 74(3): 229-263. DOI: 10.3322/caac.21834.

2 Kulaylat AS, Holleneak CS, Stewart DS. Squamous cancers of the rectum demonstrate poorer survival and increased need for salvage surgery compared with squamous cancers of the anus[J]. Dis Colon Rectum, 2017, 60(9): 922-927. DOI: 10.1097/DCR.0000000000000881.

3 Guerra GR, Kong CH, Warrier SK, et al. Primary squamous cell carcinoma of the rectum: an update and implications for treatment[J]. World J Gastrointest Surg, 2016, 8(3): 252-265. DOI: 10.4240/wjgs.v8.i3.252.

4 Astaras C, Devito C, Chaskar P, et al. The first comprehensive genomic characterization of rectal squamous cell carcinoma[J]. J Gastroenterol, 2023, 58(2): 125-134. DOI: 10.1007/s00535-022-01937-w.

5 Pigott JP, Williams GB. Primary squamous cell carcinoma of the colorectum: case report and literature review of a rare entity[J]. J Surg Oncol, 1987, 35(2): 117-119. DOI: 10.1002/jso.2930350211.

6 Liz-Pimenta J, Ferreira C, Araujo A, et al. Comprehensive look

at rectal squamous cell carcinoma[J]. BMJ Case Rep, 2024, 17(1): e255284. DOI: 10.1136/bcr-2023-255284.

7 Schizas D, Katsaros I, Mastoraki A, et al. Primary squamous cell carcinoma of colon and rectum: a systematic review of the literature[J]. J Invest Surg, 2022, 35(1): 151-156. DOI: 10.1080/08941939.2020.1824044.

8 Steinemann DC, Muller PC, Billeter AT, et al. Surgery is essential in squamous cell cancer of the rectum[J]. Langenbecks Arch Surg, 2017, 402(7): 1055-1062. DOI: 10.1007/s00423-017-1614-5.

9 Loganadane G, Servagi-Vernat S, Schernberg A, et al. Chemoradiation in rectal squamous cell carcinoma: bi-institutional case series[J]. Eur J Cancer, 2016, 58: 83-89. DOI: 10.1016/j.ejca.2016.02.005.

10 Song EJ, Jacobs CD, Palta M, et al. Evaluating treatment protocols for rectal squamous cell carcinomas: the duke experience and literature[J]. J Gastrointest Oncol, 2020, 11(2): 242-249. DOI: 10.21037/jgo.2018.11.02.

11 Kommalapati A, Tella SH, Yadav S, et al. Survival and prognostic factors in patients with rectal squamous cell carcinoma[J]. Eur J Surg Oncol, 2020, 46(6): 1111-1117. DOI: 10.1016/j.ejso.2020.02.039.

12 Ozuner G, Aytac E, Gorgun E, et al. Colorectal squamous cell carcinoma: a rare tumor with poor prognosis[J]. Int J Colorectal Dis, 2015, 30(1): 127-130. DOI: 10.1007/s00384-014-2058-9.

13 Giannakeas V, Lim DW, Narod SA. Bilateral mastectomy and breast cancer mortality[J]. JAMA Oncol, 2024, 10(9): 1228-1236. DOI: 10.1001/jamaoncol.2024.2212.

14 Ganti AK, Klein AB, Cotarla I, et al. Update of incidence, prevalence, survival, and initial treatment in patients with non-small cell lung cancer in the US[J]. JAMA Oncol, 2021, 7(12): 1824-1832. DOI: 10.1001/jamaoncol.2021.4932.

15 Guan X, Jiao S, Wen R, et al. Optimal examined lymph node number for accurate staging and long-term survival in rectal cancer: a population-based study[J]. Int J Surg, 2023, 109(8): 2241-2248. DOI: 10.1097/JS9.0000000000000320.

16 Astaras C, Bornand A, Koessler T. Squamous rectal carcinoma: a rare malignancy, literature review and management recommendations[J]. ESMO Open, 2021, 6(4): 100180. DOI: 10.1016/j.esmoop.2021.100180.

17 Liu R, Zhang J, Zhang Y, et al. Treatment paradigm and prognostic factor analyses of rectal squamous cell carcinoma[J]. Front Oncol, 2023, 13: 1160159. DOI: 10.3389/fonc.2023.1160159.

18 Chiu MS, Verma V, Bennion NR, et al. Comparison of outcomes between rectal squamous cell carcinoma and adenocarcinoma[J]. Cancer Med, 2016, 5(12): 3394-3402. DOI: 10.1002/cam4.927.

19 Diao JD, Wu CJ, Cui HX, et al. Nomogram predicting overall survival of rectal squamous cell carcinomas patients based on the SEER database: a population-based STROBE cohort study[J]. Medicine (Baltimore), 2019, 98(46): e17916. DOI: 10.1097/MD.0000000000017916.

20 Miller DD, Brown EW. Artificial intelligence in medical practice: the question to the answer?[J]. Am J Med, 2018, 131(2): 129-133. DOI: 10.1016/j.amjmed.2017.10.035.

21 Ma B, Meng F, Yan G, et al. Diagnostic classification of cancers using extreme gradient boosting algorithm and multi-omics data[J]. Comput Biol Med, 2020, 121: 103761. DOI: 10.1016/j.compbiomed.2020.103761.

22 Yu W, Lu Y, Shou H, et al. A 5-year survival status prognosis of nonmetastatic cervical cancer patients through machine learning algorithms[J]. Cancer Med, 2023, 12(6): 6867-6876. DOI: 10.1002/cam4.5477.

23 Jiang J, Pan H, Li M, et al. Predictive model for the 5-year survival status of osteosarcoma patients based on the SEER database and XGBoost algorithm[J]. Sci Rep, 2021, 11(1): 5542. DOI: 10.1038/s41598-021-85223-4.

24 孟祥勇，秦嘉怡，陈文生. 基于机器学习的早期胃癌淋巴结转移预测模型构建与验证[J]. 陆军军医大学学报，2024, 46(21): 2432-2442. [Meng XY, Qin JY, Chen WS. Construction and validation of a prediction model for lymph node metastasis in early gastric cancer based on machine learning[J]. Journal of Army Medical University, 2024, 46(21): 2432-2442.] DOI: 10.16016/j.2097-0927.202403126.

25 田园，林志浩，李瑞，等. 基于组合优化的机器学习模型预测胃癌术后感染性并发症的诊断性研究[J]. 中国循证医学杂志，2024, 24(9): 993-1003. [Tian Y, Lin ZH, Li R, et al. Diagnostic study of machine learning model based on combinatorial optimization to predict postoperative infectious complications of gastric cancer[J]. Chinese Journal of Evidence-Based Medicine, 2024, 24(9): 993-1003.] DOI: 10.7507/1672-2531.202310069.

收稿日期：2024 年 10 月 25 日 修回日期：2025 年 03 月 07 日
本文编辑：李绪辉 曹 越

引用本文：黄诗, 林丹丹. 不同机器学习算法在直肠鳞状细胞癌预后模型构建中的应用: 基于SEER数据库[J]. 医学新知, 2025, 35(8): 870-877. DOI: 10.12173/j.issn.1004-5511.202410124.
Huang S, Lin DD. Application of different machine learning algorithms in the construction of prognostic models for rectal squamous cell carcinoma: based on SEER database[J]. Yixue Xinzhi Zazhi, 2025, 35(8): 870-877. DOI: 10.12173/j.issn.1004-5511.202410124.