

膀胱癌患者术后影响因素研究

冯帅宁

大连交通大学理学院, 辽宁 大连

收稿日期: 2025年6月11日; 录用日期: 2025年7月2日; 发布日期: 2025年7月15日

摘要

目的: 探讨膀胱癌患者术后预后的关键影响因素, 为临床诊疗提供参考。方法: 基于大连医科大学第二医院和南方医科大学南方医院714例膀胱癌术后患者的临床数据, 采用机器学习算法XGBoost结合SHAP可解释性分析进行特征重要性排序, 结合生存分析(Kaplan-Meier曲线及Cox回归)验证核心影响因素。结果: 年龄增长、高T分期(肿瘤侵袭程度深)及淋巴结转移(N分期升高)显著增加术后死亡风险, 而白蛋白水平升高可改善预后。生存分析进一步表明, 高龄、晚期肿瘤分期及低白蛋白患者生存率显著降低。结论: 年龄、肿瘤分期、淋巴结转移及白蛋白水平是膀胱癌术后预后的核心影响因素, 临床需针对高危患者优化干预策略。

关键词

膀胱癌, 机器学习, 术后预后, 生存分析

Study on Postoperative Influencing Factors in Patients with Bladder Cancer

Shuaining Feng

College of Science, Dalian Jiaotong University, Dalian Liaoning

Received: Jun. 11th, 2025; accepted: Jul. 2nd, 2025; published: Jul. 15th, 2025

Abstract

Objective: To investigate the key influencing factors of postoperative prognosis of patients with bladder cancer, and to provide reference for clinical diagnosis and treatment. **Methods:** Based on the clinical data of 714 postoperative bladder cancer patients in the Second Hospital of Dalian Medical University and Nanfang Hospital of Southern Medical University, the machine learning algorithm XGBoost combined with SHAP interpretability analysis was used to rank the importance of

文章引用: 冯帅宁. 膀胱癌患者术后影响因素研究[J]. 世界肿瘤研究, 2025, 15(3): 140-150.

DOI: 10.12677/wjcr.2025.153018

features, and the core influencing factors were verified by survival analysis (Kaplan-Meier curve and Cox regression). Results: Increasing age, high T stage (deep tumor invasion) and lymph node metastasis (increased N stage) significantly increased the risk of postoperative mortality, while increased albumin level improved the prognosis. Survival analysis further showed that the survival rate of patients with advanced age, advanced tumor stage and low albumin was significantly reduced. Conclusion: Age, tumor stage, lymph node metastasis and albumin level are the core influencing factors for the prognosis of bladder cancer after surgery, and it is necessary to optimize the intervention strategy for high-risk patients.

Keywords

Bladder Cancer, Machine Learning, Postoperative Prognosis, Survival Analysis

Copyright © 2025 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. 引言

膀胱癌是全球范围内严重威胁人类健康的常见泌尿系统恶性肿瘤[1][2],其发病率呈现持续上升趋势。根据国际癌症研究机构(IARC)最新统计数据显示,膀胱癌在所有恶性肿瘤中发病率位居第十位[3],具有显著的疾病负担。尽管近年来诊疗技术不断进步,但该肿瘤仍表现出较高的复发倾向和异质性预后特征,这使得精准预后评估成为临床研究的重要课题。目前临床实践中,预后评估主要依赖于组织病理学分级和 TNM 分期系统[4]。然而,这些传统评估方法存在明显的局限性:首先,它们主要反映肿瘤的局部解剖学特征,而忽略了宿主全身状态对疾病转归的影响;其次,在临床应用中经常出现某些指标预测效能不足的情况;更重要的是,这些指标难以全面体现肿瘤生物学行为的复杂性。随着精准医学理念的深入发展,基于多维度临床数据的预后评估新模式正在形成。人工智能和机器学习技术在处理复杂临床数据方面展现出独特优势[5],能够识别传统统计方法可能忽略的重要预后因素。特别是集成学习方法,通过特征重要性分析和变量交互作用检测,为建立更精准的预后预测模型提供了新的技术路径[6]-[18]。本研究创新性地采用 XGBoost 算法[19]结合生存分析方法[20],系统评估了 714 例膀胱癌患者的临床资料。研究不仅验证了传统预后指标的价值,更发现了包括患者基础特征和实验室指标在内的新型预测因素。这些研究成果为完善现有的预后评估体系提供了重要依据,对推动膀胱癌精准治疗具有重要的临床价值。

2. 材料与方法

2.1. 数据准备

本研究共纳入 2007 年 1 月至 2019 年 1 月期间在大连医科大学第二医院和南方医科大学南方医院接受手术治疗的 832 例患者。纳入标准:初始患者接受手术治疗。术后病理诊断为膀胱尿路上皮癌。可获得完整的临床和随访信息。术后随访至少 3 年。膀胱滞留患者在手术后接受了全疗程的基于表柔比星的膀胱输注化疗。排除标准:围手术期接受新辅助或辅助放疗或化疗。无法获得临床或随访信息。术后随访时间少于 3 年。收集每位参与者的临床信息和随访资料。对数据进行预处理将数据中未访问满三年以及失访的数据进行剔除共 118 例,对剩下的 714 例数据进行异常值、空白值检测均未发现异常。即在后续方法分析中使用的数据为 714 例。

本研究使用的数据集由大连医科大学第二医院提供,已完全匿名化处理(删除姓名、ID、住址等直接标识符及可能简介标识的信息)。该数据仅用于数学建模分析,不作其他用途。

数据描述

针对离散数据：性别(0: 女, 1: 男)、高血压或心脑血管疾病(1: 有, 0: 无)、糖尿病(1: 有, 0 无)、吸烟(1: 有, 0: 无)、单发/多发(0: 单发, 1 多发)、病理分级(1: 低度恶性潜能, 2: 低级别, 3: 高级别)、T分期(0: T0期, 1: T1期, 2: T2期, 3: T3期, 4: T4期)、N分期(0: N0期, 1: N1期, 2: N2期)、M(0: M0期)、术式(1: 经尿道膀胱肿瘤电切术, 2: 膀胱部分切除, 3: 膀胱全切除)、status (0: 存活, 1: 死亡)这些离散数据的基础描述详见表 1。

Table 1. Underlying discrete data information

表 1. 基础的离散数据信息

| 名称 | 类别 | 统计 | 计数 |
|------------|----|-------|-----|
| 性别 | 0 | 19.2% | 137 |
| | 1 | 80.8% | 577 |
| 高血压或心脑血管疾病 | 0 | 68.2% | 487 |
| | 1 | 31.8% | 227 |
| 糖尿病 | 0 | 88.7% | 633 |
| | 1 | 11.3% | 81 |
| 吸烟 | 0 | 63.7% | 455 |
| | 1 | 36.3% | 259 |
| 单发/多发 | 0 | 58% | 414 |
| | 1 | 42% | 300 |
| 病理分级 | 1 | 12% | 86 |
| | 2 | 36.3% | 259 |
| | 3 | 51.7% | 369 |
| T 分期 | 0 | 0.1% | 1 |
| | 1 | 74.9% | 535 |
| | 2 | 16.9% | 121 |
| | 3 | 5.6% | 40 |
| N 分期 | 4 | 2.4% | 17 |
| | 0 | 98.2% | 701 |
| | 1 | 1% | 7 |
| | 2 | 0.8% | 6 |
| M 分期 | 0 | 100% | 714 |
| | 1 | 69.1% | 493 |
| 术式 | 2 | 7% | 50 |
| | 3 | 23.9% | 171 |
| | 0 | 88.4% | 631 |
| status | 1 | 11.6% | 83 |

通过对表 1 的分析发现，在 N 分期中 N0 占比过大，而 N1, N2 占比过小，这样很有可能导致模型低估 N1、N2 的风险因素；在 M 分期中数据全部为 M0，为一个常数项这样的话会降低模型的性能，于是在后续研究中去除掉 M 分期。status 为生死状态，包含两种结果存活和死亡。

针对连续数据，本次研究中纳入的因素包含：年龄、血红蛋白、白细胞、粒/淋巴(中性粒细胞/淋巴细胞比)、血小板、肌酐清除率、白蛋白、白/球比(白蛋白/球蛋白)、谷草转氨酶、谷丙转氨酶、乳酸脱氢酶、血钙、time。连续数据的基础信息描述详见表 2。

Table 2. Underlying continuous data information**表 2.** 基础的连续数据信息

| 名称 | 均值 | 标准差 | 最小值 | 25% | 50% | 75% | 最大值 | 共计 | 缺失值 |
|-------|--------|-------|-------|-------|--------|--------|--------|-----|-----|
| 年龄 | 63.74 | 13.31 | 19 | 56 | 64 | 73 | 95 | 714 | 0 |
| 血红蛋白 | 136.82 | 20.27 | 47 | 127 | 139 | 151 | 242 | 714 | 0 |
| 白细胞 | 6.61 | 1.93 | 0.56 | 5.29 | 6.32 | 7.5 | 18.08 | 714 | 0 |
| 粒/淋比 | 2.41 | 2.15 | 0.37 | 1.47 | 1.91 | 2.73 | 30.65 | 714 | 0 |
| 血小板 | 219.66 | 69.88 | 1.58 | 178 | 212 | 250.75 | 978 | 714 | 0 |
| 肌酐清除率 | 78.82 | 21.97 | 1.25 | 65.99 | 87.04 | 90 | 185.34 | 714 | 0 |
| 白蛋白 | 41.44 | 4.61 | 13.2 | 38.71 | 41.77 | 44.5 | 53.7 | 714 | 0 |
| 白/球比 | 1.63 | 0.36 | 0.55 | 1.39 | 1.61 | 1.83 | 3.01 | 714 | 0 |
| 谷草转氨酶 | 22.82 | 11.77 | 1 | 17 | 20.78 | 25 | 172.01 | 714 | 0 |
| 谷丙转氨酶 | 22.07 | 15.68 | 2.6 | 13 | 18 | 26 | 209.57 | 714 | 0 |
| 乳酸脱氢酶 | 185.21 | 50.3 | 15.14 | 158 | 182.69 | 212.56 | 472 | 714 | 0 |
| 血钙 | 2.28 | 0.86 | 0 | 2.18 | 2.28 | 2.37 | 22.26 | 714 | 0 |
| time | 70.1 | 31.89 | 2 | 52 | 70 | 86 | 186 | 714 | 0 |

通过分析表 2 可以看出数据中存在个别极端值，其中血钙最大值为 22.26 mmol/L, 远超其值，谷丙转氨酶和谷草转氨酶的最大值也是显著升高，粒/淋比的最大值也是远高于常见范围(1~3)。针对这些个别极端值，在本次研究中没有舍去而是纳入了研究。time 为生存时间其单位为月。

2.2. 方法

本研究采用基于 Python 的集成分析方法来讨论那些因素对生存状态和生存时间的影响，结合机器学习特征选择与生存分析技术，系统评估影响膀胱癌患者术后预后的关键因素。所有分析均在 Python 3.12 环境下完成，主要依赖以下关键库：XGBoost (2.1.1 版本)用于特征重要性分析，SHAP (0.46.0 版本)提供模型解释性支持，lifelines (0.29.0 版本)进行生存分析，scikit-learn (1.5.1 版本)处理数据预处理和模型验证。

在特征选择阶段，我们采用 XGBoost 算法计算特征重要性评分，同时运用 SHAP 方法[21]评估各特征对预测结果的贡献度。这两种评估指标经过标准化处理后，按照 50%:50%的权重比例进行加权融合，生成综合特征重要性评分。为验证模型的稳健性，我们特别设计了三重敏感性分析策略：(1) 基于临床指南阈值(血钙>3.0 mmol/L、白蛋白<25 g/L 等)剔除异常值；(2) 采用 Winsorization 缩尾处理极端值；(3) 保留异常值但使用 Huber 回归降低其权重。分析结果显示，核心预后因素(年龄、白蛋白)的参数估计在不同处理方法下保持稳定(SHAP 值波动<2%，HR 变化<5%)，证实主要结论的可靠性。

在生存分析方面，本研究采用 Kaplan-Meier 法估计生存函数，组间差异比较使用双侧 log-rank 检验(检验水准 $\alpha = 0.05$)。多因素 Cox 比例风险回归模型采用 Breslow 法处理结数据，通过 Schoenfeld 残差检验(显著性水平 $\alpha = 0.05$)评估比例风险假设。模型拟合优度通过 Akaike 信息准则(AIC)评价，变量显著性采用 Wald 检验判定。

3. 结果

3.1. 机器学习特征筛选分析结果

采用 XGBoos 模型评估预后特征的重要性如图 1，结果显示：N 分期(F score = 4.34)对模型预测的贡

献度最高；其次为年龄(F score = 3.91)、白蛋白(F score = 2.86)、病理分级(F score = 2.83)合 T 分期(F score = 2.72)，其重要性依次递减。

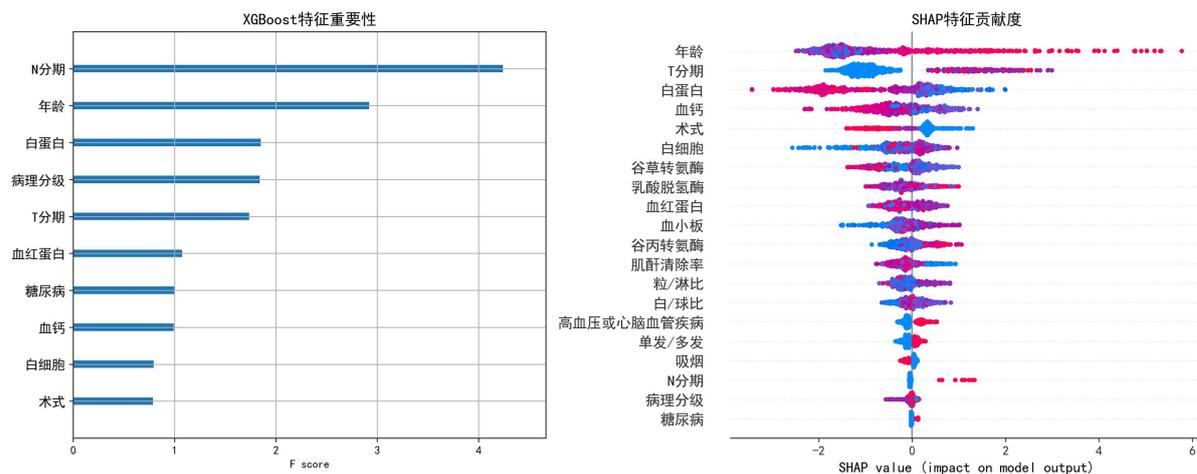


Figure 1. XGBoost feature importance and SHAP contribution combined chart
图 1. XGBoost 特征重要性与 SHAP 贡献度合并图

SHAP 分析进一步揭示了特征对预测结果的影响机制，结果如图 1：年龄：高值(红色散点)与正向 SHAP 值有关，提示年龄增长增加预后风险；T 分期：高分期(红色)与风险正相关，低分期(蓝色)与保护性效应相关；术式：SHAP 值分布分散，表明不同手术方式对预后的影响存在异质性。

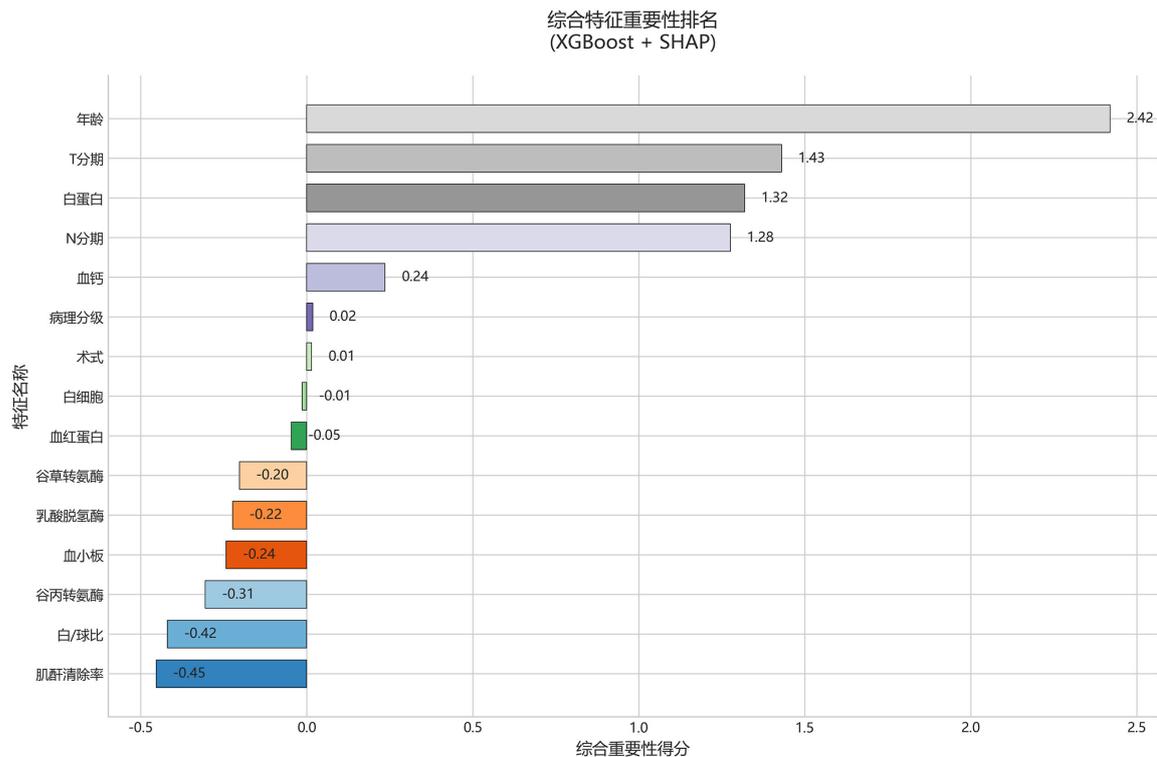


Figure 2. Overall feature importance ranking (XGBoost + SHAP)
图 2. 综合特征重要性排名(XGBoost + SHAP)

通过 XGBoost 结合 SHAP，我们筛选出 4 个关键预后影响因素。综合特征重要性结果如图 2，核心影响因素(综合得分 > 1.0)：年龄(2.42)综合得分最高，是影响预后的最核心特征、T 分期(1.43)分期越高，预后风险越大、白蛋白(1.32)保护性因素，水平越高预后越好、N 分期(1.28)淋巴结转移情况显著影响预后；次要影响因素($0 < \text{综合得分} < 1$)：血钙(0.24)影响相对有限、病理分级(0.02)贡献度较低、术式(0.01)不同手术方式影响差异不显著。

图 2 展示了基于 XGBoost 特征重要性(F score)和 SHAP 值(贡献度)的综合评分结果。条形图按综合得分降序排列，正值表示特征对模型预测起关键作用，负值表示影响较弱。颜色深浅区分不同重要性层级，深色标注的 7 个特征为筛选出的关键因素，而对于白细胞(-0.01)、血红蛋白(-0.05)至肌酐清除率(-0.45)等特征综合得分负向，且绝对值越大(如肌酐清除率)，说明其对模型的贡献度最低，在综合分析中重要性几乎可忽略。

该分析为后续生存分析提供了经过严格筛选的特征集合，既保留了最具预测价值的临床指标，又避免了过多变量导致的模型过拟合问题。筛选结果与临床经验相符，特别是年龄和肿瘤分期作为核心预后因素得到数据支持。

3.2. 关键预后因素的生存分析验证

3.2.1. 生存分析结果

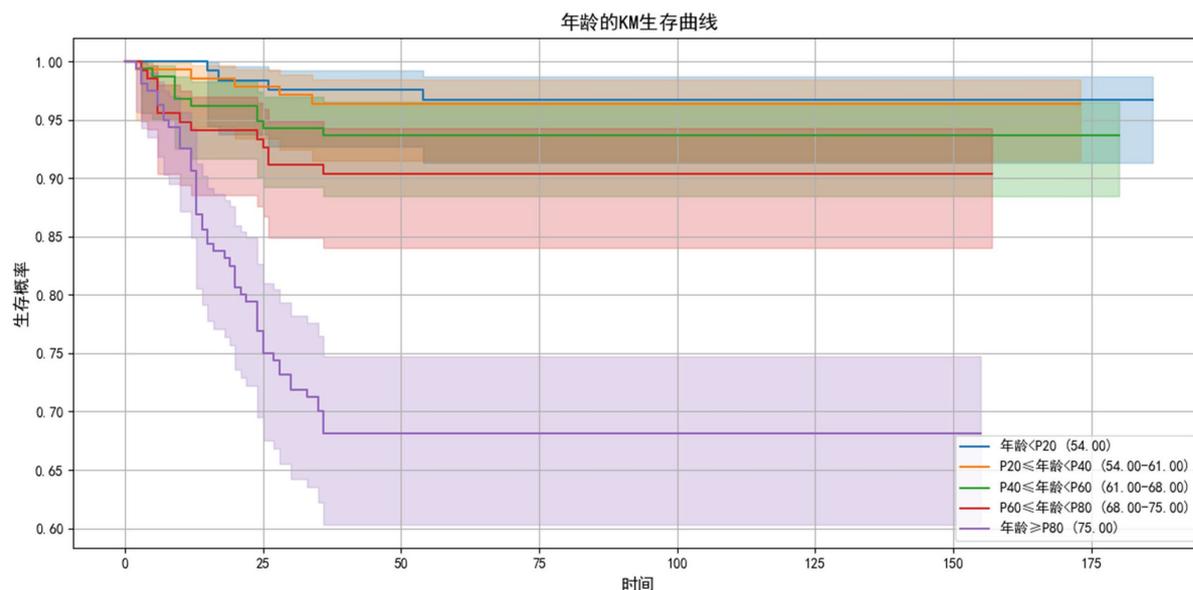


Figure 3. KM survival curve chart for age

图 3. 年龄的 KM 生存曲线图

本研究基于年龄连续变量的五等分分层(Q1 < 54 岁, Q2: 54~61 岁, Q3: 61~68 岁, Q4: 68~75 岁, Q5 ≥ 75 岁)进行生存分析(时间单位: 月)。KM 曲线显示如图 3，随着分位等级升高，患者预后呈阶梯式恶化(log-rank $P < 0.001$)：Q1 组(<54 岁) 150 个月生存率 > 95%，生存曲线近乎水平；而 Q5 组(≥75 岁)生存率快速下降，150 个月时仅 75%。中间三组(Q2~Q4)呈现明显的剂量效应，每上升一个分位(约 7 岁)生存率下降 10%~15%。该结果证实年龄是预后的连续性危险因素，且 ≥75 岁患者需要特别关注。需注意 Q5 组在随访后期置信区间变宽，可能与样本量减少有关。

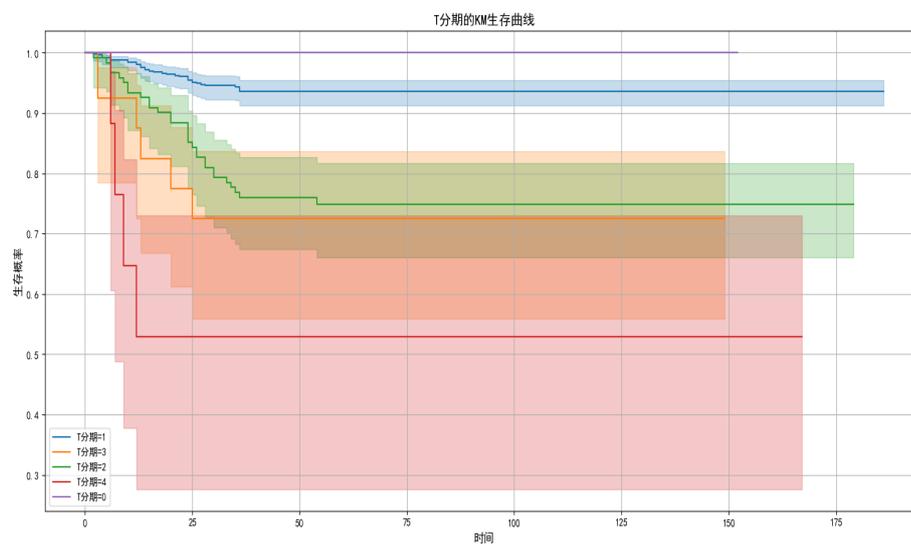


Figure 4. KM survival curve chart for T stage
图 4. T 分期的 KM 生存曲线图

本生存分析通过 KM 曲线如图 4 展示了不同 T 分期患者的生存差异。结果显示，生存结局与 T 分期呈显著负相关(log-rank $P < 0.001$): T0 期患者生存概率始终维持在接近 1.0 的最高水平，曲线近乎水平; T1 期生存表现次之，整体保持平稳; T2~T3 期呈现渐进式恶化，其中 T3 期下降趋势更显著(3 年生存率较 T2 期低约 15%); 而 T4 期预后最差，生存概率早期即快速下滑至 0.5 左右，且长期维持低位。该结果证实 T 分期是评估肿瘤侵袭性和预后的关键指标，分期越高(尤其 T4 期)患者生存压力越大，需采取更积极的治疗和随访策略。

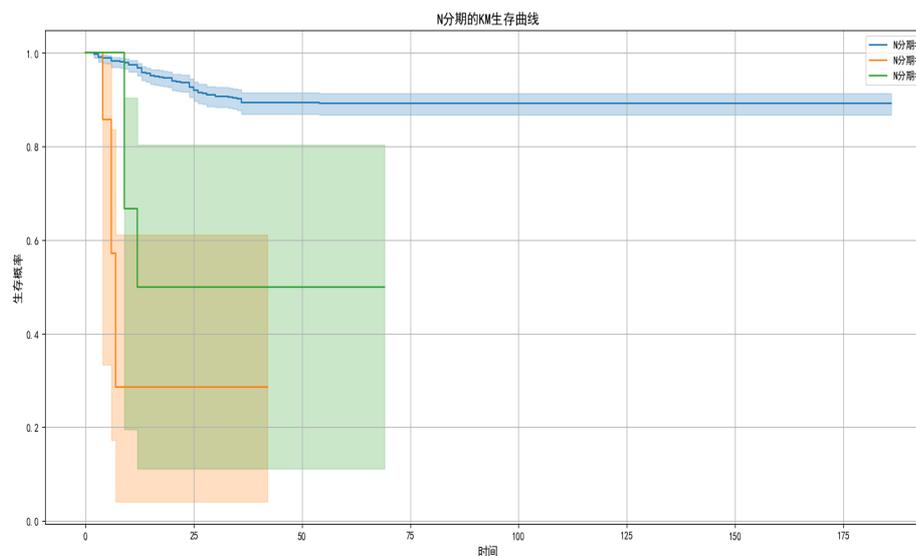


Figure 5. KM survival curve chart for N stage
图 5. N 分期的 KM 生存曲线图

图 5 是 N 分期的 KM 生存曲线，用于呈现不同淋巴结转移分期(N = 0, 1, 2)患者的生存概率随时间变

化趋势。横轴为观察时间，纵轴为生存概率，结果显示：N分期 = 0(无淋巴结转移)的生存概率最高，下降最缓慢；N分期 = 1 和 N分期 = 2 的生存概率较低，且 N分期 = 2 的下降幅度更显著。整体反映出 N分期与生存概率呈负相关，即淋巴结转移越严重(N分期越高)，患者生存概率越低、预后越差。

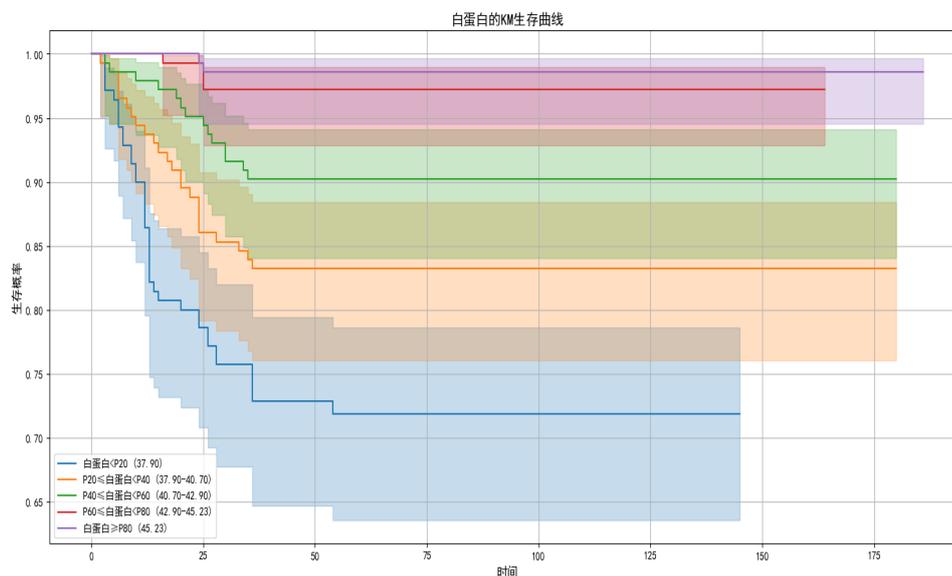


Figure 6. KM survival plot of albumin

图 6. 白蛋白的 KM 生存曲线图

通过图 6 白蛋白的 KM 生存曲线分析揭示了膀胱癌患者白蛋白水平与预后的显著相关性(log-rank $P < 0.001$)。结果显示，白蛋白水平与生存结局呈明确的正向剂量 - 效应关系：高白蛋白组(≥ 45.23 g/L)表现出最优生存预后(3 年生存率 $> 95\%$)，曲线平缓；低白蛋白组(< 37.90 g/L)预后最差，生存曲线早期即快速下降(3 年生存率降幅达 $30\% \sim 40\%$)；中间各组($37.90 \sim 45.23$ g/L)随白蛋白水平升高呈现 $15\% \sim 20\%$ 的生存率梯度改善。这一发现证实了白蛋白作为膀胱癌患者独立预后指标的重要价值，提示临床应特别关注低白蛋白患者的营养支持和综合治疗，同时为预后评估提供了可靠的量化依据。

3.2.2. 多因素 Cox 回归分析

本研究采用带 L2 正则化($\lambda = 0.5$)的 Cox 比例风险模型，重点分析年龄、白蛋白及 T/N 分期等关键预后因素。分类变量(T/N 分期)经独热编码处理，连续变量(年龄、白蛋白)标准化后纳入模型。最终模型基于 714 例患者(事件数 83 例)构建，拟合优度良好(Concordance = 0.84，似然比检验 $P < 0.001$)。cox 回归结果详见表 3。

Table 3. Cox regression results for key prognostic factors

表 3. 关键预后因素的 Cox 回归结果

| 变量 | 回归系数 | HR (95% CI) | Wald χ^2 | P 值 | 临床转化建议 |
|------|-------|------------------|---------------|-----------|---------------------|
| 年龄 | 0.01 | 1.01 (1.01~1.02) | 12.04 | < 0.005 | 高龄患者缩短复查间隔 |
| 白蛋白 | -0.03 | 0.97 (0.95~0.99) | 6.97 | 0.01 | 治疗期间维持 > 40.7 g/L |
| T1 期 | -0.28 | 0.75 (0.6~0.93) | 6.3 | 0.01 | 可以考虑保守治疗策略 |
| N1 期 | 0.94 | 2.55 (1.05~6.16) | 4.33 | 0.04 | 需强化辅助治疗 + 密切随访 |

3.3. 模型稳健性验证

敏感性分析显示表 4, 在不同异常值处理策略下: 年龄的特征重要性(SHAP 值)保持稳定(波动幅度 $\leq 0.3\%$); 白蛋白的保护效应: 剔除异常值后 HR 从 0.935 降至 0.915 (强化 2.2%)、缩尾处理后 HR 为 0.929 (变化 0.7%); 样本量变化: 剔除异常值后保留 684 例(损失 4.2%), 缩尾处理保留全部样本。结果表明, 年龄作为预后因素的可靠性最高, 而白蛋白的 HR 值虽有小幅波动, 但其保护效应的统计学意义和临床方向性始终保持不变。

Table 4. Impact of different data processing methods on analysis results

表 4. 不同数据处理方法对分析结果

| 分析方法 | 年龄(SHAP) | 白蛋白(HR) | 样本量 |
|-------|----------|---------|-----|
| 原始数据 | 0.355 | 0.935 | 714 |
| 剔除异常值 | 0.356 | 0.915 | 684 |
| 缩尾处理 | 0.355 | 0.929 | 714 |

虽然缩尾处理保留了全部样本量, 但两种异常值处理方法得出的结论一致, 这显著增强了研究结果的可信度。对于临床实践中遇到的极端化验值, 建议优先参考缩尾处理结果(HR = 0.929), 因其在数据完整性和临床合理性之间达到了较好的平衡。本研究在方法学上遵循了合理原则, 但由于样本量的限制, N1/N2 期和 T4 期样本较少, 异常值处理后样本量进一步减少, 导致部分变量无法进行深入分析。根据统计学原则, COX 模型要求每个变量至少有 10~15 个事件, 而小样本量会导致 HR 估计的不稳定性, 这进一步限制了对部分变量的分析深度。在临床特征方面, T 分期的预后价值已通过 Kaplan-Meier 曲线验证, 但由于高阶分期样本量有限, 不适合进行子组分析。此外, N 分期因淋巴结转移是强预后因素, 且 N + 患者的治疗方案相对固定, 因此无需重复分析。基于上述考虑, 在模型稳健性验证中, 主要关注了年龄和白蛋白这两个变量。

4. 讨论

本研究通过整合机器学习算法(XGBoost)与经典生存分析方法, 系统探讨了膀胱癌患者术后预后的关键影响因素, 为临床预后评估提供了新的思路和证据支持。研究发现年龄、T 分期、N 分期和白蛋白水平是影响膀胱癌患者术后生存的核心因素, 这一结果不仅验证了传统临床认知, 还通过量化分析进一步明确了各因素的相对重要性。

在临床意义方面, 本研究有以下几个重要发现: 首先, 年龄作为连续变量对预后的影响呈现明显的剂量 - 效应关系, 每增加约 7 岁生存率下降 10%~15%, 这提示临床医生对高龄患者(特别是 ≥ 75 岁)需要制定更积极的随访策略。其次, T 分期和 N 分期作为肿瘤侵袭性和转移程度的指标, 其预后价值在本研究中得到充分验证, 尤其是 T4 期患者 3 年生存率较 T2 期显著降低 15%, 这一量化结果为临床分期系统的应用提供了更精确的参考。特别值得注意的是, 本研究通过机器学习方法明确了白蛋白水平的独立预后价值。生存曲线分析表明, 白蛋白水平越高, 患者的生存率越高(图 6)。Cox 回归分析进一步验证了这一关系, 结果显示白蛋白水平每增加 1 g/L, 患者的死亡风险降低约 3% (HR = 0.97, 95% CI: 0.95~0.99, P = 0.01)。低白蛋白组(<37.9 g/L)患者 3 年生存率降低 30%~40%, 这一发现为临床营养支持治疗提供了重要依据。基于上述分析结果, 建议在术后膀胱癌患者的管理中, 将白蛋白水平维持在 > 40.7 g/L, 以显著降低死亡风险并提高生存率。具体措施包括通过营养支持(如补充蛋白质、氨基酸)或药物干预(如白蛋白输注)来改善白蛋白水平, 并定期监测以确保其稳定。同时, 建议关注患者的炎症状态, 并结合抗炎治疗

优化整体管理策略。推动营养科、肿瘤科和内科的多学科协作，共同管理白蛋白水平异常的患者，可以进一步提升治疗效果。在方法学层面，本研究创新性地结合了机器学习特征选择与传统生存分析的优势。XGBoost 算法的高维数据处理能力帮助我们众多临床变量中筛选出最具预测价值的特征，而 SHAP 分析则提供了直观的特征贡献度解释。敏感性分析进一步增强了研究结果的稳健性。结果显示，在不同异常值处理策略下，年龄的特征重要性保持稳定，白蛋白的保护效应虽有小幅波动，但其统计学意义和临床方向性始终不变。具体而言，剔除异常值后白蛋白的 HR 从 0.935 降至 0.915，缩尾处理后 HR 为 0.929，表明两种方法得出的结论具有一致性。然而，本研究也存在一些局限性。数据的地域局限性可能影响结果的普适性，尤其是 N1/N2 期和 T4 期样本量较少，可能导致对这些高危人群的风险评估不够准确。此外，研究未纳入分子标志物[22]，限制了模型在分子水平的解释深度。在临床转化方面，白蛋白干预阈值的确定需要更大样本的验证，营养支持等干预措施的具体影响也有待前瞻性研究进一步证实。未来研究可从以下几个方向深入：一是开展多中心合作研究，扩大样本量，特别是高危人群的样本数量，提高模型的稳定性和泛化能力；二是整合多组学数据，构建更全面的预后预测模型[23]-[26]，探索分子机制与临床特征的交互作用；三是开发动态预后监测系统，将术后指标变化趋势纳入分析，实现更精准的个体化预测；四是设计前瞻性干预研究，验证个体化治疗策略的临床效果，推动研究成果向临床实践转化。

综上所述，本研究通过多方法融合的策略，为膀胱癌术后预后评估提供了新的视角和工具。研究结果对完善现有预后评估体系、优化临床决策具有重要价值，未来需要通过更大规模、更深入的研究进一步验证和拓展这些发现，以实现膀胱癌患者的精准化、个体化诊疗。

参考文献

- [1] Siegel, R.L., Miller, K.D., Fuchs, H.E. and Jemal, A. (2022) Cancer Statistics, 2022. *CA: A Cancer Journal for Clinicians*, **72**, 7-33. <https://doi.org/10.3322/caac.21708>
- [2] Xia, C., Dong, X., Li, H., Cao, M., Sun, D., He, S., et al. (2022) Cancer Statistics in China and United States, 2022: Profiles, Trends, and Determinants. *Chinese Medical Journal*, **135**, 584-590. <https://doi.org/10.1097/cm9.0000000000002108>
- [3] Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A. and Jemal, A. (2018) Global Cancer Statistics 2018: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA: A Cancer Journal for Clinicians*, **68**, 394-424. <https://doi.org/10.3322/caac.21492>
- [4] 陈锦元, 蒋伟. 宫颈癌 TNM 分期与预后相关性分析[J]. 世界肿瘤研究, 2025, 15(1): 45-51.
- [5] Cruz, J.A. and Wishart, D.S. (2006) Applications of Machine Learning in Cancer Prediction and Prognosis. *Cancer Informatics*, **2**. <https://doi.org/10.1177/117693510600200030>
- [6] 侯钊, 双卫兵. 机器学习算法的概述及其在生物医学中的应用[J]. 中国医学工程, 2025, 33(3): 72-78.
- [7] 刘霞. 机器学习助力多种癌症早期筛查[N]. 科技日报, 2025-01-15(004).
- [8] 叶璟威, 刘卓炜. 人工智能在膀胱癌诊断中的研究进展[J]. 临床泌尿外科杂志, 2024, 39(12): 1071-1075.
- [9] 方昱衡, 李泽伟, 许迎盈, 等. 基于机器学习的膀胱癌患者生存预测模型研究[J]. 现代信息科技, 2024, 8(16): 83-87.
- [10] 杜龙廷. 基于多组学分析的膀胱癌分子机制研究[D]: [硕士学位论文]. 西安: 西安电子科技大学, 2023.
- [11] 何俊翔, 张海燕, 李海, 等. 基于机器学习筛选膀胱癌预后标志物[J]. 中国老年学杂志, 2022, 42(24): 5957-5961.
- [12] 徐文浩, 田熙, 艾合太木江·安外尔, 等. 人工智能在泌尿系统肿瘤中的应用研究进展[J]. 中国癌症杂志, 2022, 32(1): 68-74.
- [13] 朱韦宇, 樊纪山, 江阳, 户彩凤, 孙巧榆. 基于机器学习的新冠疑似人员预测[J]. 计算机科学与应用, 2023, 13(7): 1420-1427.
- [14] 杨兴俊, 杨兴华. 机器学习在医疗数据发展中的应用思考[J]. 应用数学进展, 2022, 11(6): 3496-3501.
- [15] 郭东斌, 苏鑫洪, 黄俊龙. 机器学习预测脑膜瘤术后静脉血栓高危因素[J]. 福建医药杂志, 2024, 46(5): 20-26, 122.
- [16] 刘玥, 刘启玲, 苏海霞, 等. 机器学习法在生存分析中的应用研究[J]. 预防医学, 2024, 36(6): 496-500, 505.

-
- [17] 吴娇. 基于可解释机器学习的晚期结直肠癌中西医生存预测模型的构建及验证[D]: [博士学位论文]. 北京: 中国中医科学院, 2024.
- [18] 王佩佩, 双卫兵. 生存分析概述及模型应用[J]. 中国医学工程, 2023, 31(11): 62-68.
- [19] Chen, T. and Guestrin, C. (2016) XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, 13-17 August 2016, 785-794. <https://doi.org/10.1145/2939672.2939785>
- [20] 骆杨. 生存分析在客户流失预测模型中的应用[J]. 统计学与应用, 2021, 10(1): 10-20.
- [21] 万雨勤, 仝纪龙, 刘永乐, 等. 基于 XGBoost 算法量化气象要素对兰州市主城区夏季臭氧浓度的影响[J/OL]. 环境科学: 1-16. <https://doi.org/10.13227/j.hjlx.202412038>, 2025-04-03.
- [22] Robert, C. (2020) A Decade of Immune-Checkpoint Inhibitors in Cancer Therapy. *Nature Communications*, **11**, Article No. 3801. <https://doi.org/10.1038/s41467-020-17670-y>
- [23] 鲁鹤臻. 早期膀胱癌多组学分析及 APOA2 对膀胱癌生物学功能的调控[D]: [博士学位论文]. 长春: 吉林大学, 2023.
- [24] 文英美, 姚颐. 代谢组学在恶性肿瘤研究中的应用进展[J]. 肿瘤学杂志, 2024, 30(1): 57-62.
- [25] 邓细蒙. 基于生物信息学方法构建膀胱癌基底膜相关基因预后模型[D]: [硕士学位论文]. 扬州: 扬州大学, 2023.
- [26] 王铁, 许建成. 膀胱癌的代谢组学研究进展[J]. 检验医学与临床, 2023, 20(12): 1771-1774.