·论著·一次研究·

机器学习模型辅助肿瘤手术高危患者筛查及预住院模式效果评价

弓玉红[1]，尚　楠[2]，贺娅莎[3]，张绍果[1]，王文天[4]，杨　壮[4]，杨露露[4]，刘艳梅[1]，高红梅[1]，赵婧媛[5]，周培斌[6]

1. 山西医科大学第一医院住院部（太原 030001）
2. 山西医科大学第一医院药学部（太原 030001）
3. 山西医科大学第一医院病案室（太原 030001）
4. 山西医科大学第一医院运营部（太原 030001）
5. 山西医科大学第一医院团委（太原 030001）
6. 山西医科大学第一医院院长办公室（太原 030001）

【摘要】目的 构建机器学习（ML）模型识别肿瘤患者住院资源消耗的风险因素，筛选高危患者实施预住院管理，评估其对住院时长与费用的影响。**方法** 回顾性收集 2021 年 4 月至 2022 年 10 月山西医科大学第一医院收治的肿瘤住院患者资料，采用 Lasso 回归进行特征筛选，构建 ML 模型识别影响住院时长与住院费用的关键因素，并利用 SHAP 值解释模型结果。基于建模结果，于 2022 年 11 月至 12 月构建风险因素筛查和预住院管理模式，于 2023 年 1 月至 6 月实施该管理模式并收集实施与未实施预住院管理的患者资料，采用倾向性评分匹配（PSM）控制组间基线差异，并使用因果森林（CF）评估预住院干预对住院费用与住院时间的影响。**结果** 共纳入 5 211 例肿瘤患者，构建的 9 种 ML 模型中 CatBoost 模型效果最优，肿瘤类型、术前等待天数和年龄为影响住院资源消耗的主要风险因素。基于构建模型共筛选了 698 例患者纳入预住院管理，PSM 后得到 563 对患者，CF 分析结果显示，与对照组相比，接受预住院干预的患者平均住院时长显著缩短 3.004 [95%CI(−3.334，−2.675)]d，住院总费用平均减少 1 473.124 [95%CI（−2,166.093，−780.155）] 元，表明预住院管理可有效改善住院资源利用效率。**结论** 对基于 ML 模型识别的高危人群开展预住院管理，能够显著缩短肿瘤患者住院时间、降低住院费用，对优化住院流程和资源配置具有积极意义。

【关键词】预住院模式；住院时长；住院费用；倾向性评分匹配；机器学习；高危肿瘤患者

【中图分类号】R 730.56；R 197.3　　**【文献标识码】**A

Research on machine learning model-assisted screening of high-risk tumor surgery patients and the effectiveness of pre-admission management

GONG Yuhong[1], SHANG Nan[2], HE Yasha[3], ZHANG Shaoguo[1], WANG Wentian[4], YANG Zhuang[4], YANG Lulu[4], LIU Yanmei[1], GAO Hongmei[1], ZHAO Jingyuan[5], ZHOU Peibin[6]

1. Inpatient Department, First Hospital of Shanxi Medical University, Taiyuan 030001, China

2. Department of Pharmacy, First Hospital of Shanxi Medical University, Taiyuan 030001, China

3. Medical Records Department, First Hospital of Shanxi Medical University, Taiyuan 030001, China

DOI: 10.12173/j.issn.1004-5511.202503188

基金项目：山西省人民政府重大决策咨询课题（WTD08）

通信作者：周培斌，高级工程师，硕士研究生导师，Email：zybgyh4939813@126.com

4. Operations Department, First Hospital of Shanxi Medical University, Taiyuan 030001, China
5. Youth League Committee, First Hospital of Shanxi Medical University, Taiyuan 030001, China
6. Office of the President, First Hospital of Shanxi Medical University, Taiyuan 030001, China
Corresponding author: ZHOU Peibin, Email: zybgyh4939813@126.com

【Abstract】Objective This study aimed to identify high-risk factors associated with high resource consumption during hospitalization among tumor patients using machine learning (ML) models, and to screen high-risk patients for pre-admission management to evaluate its impact on length of stay and hospitalization costs. Methods The date were retrospectively collected from inpatients treated at First Hospital of Shanxi Medical University between April 2021 and October 2022. Lasso regression was applied for feature selection, and multiple ML models were constructed to identify key factors influencing length of stay and hospitalization costs. SHAP values were used to interpret model results. Based on the modeling results, a risk factor screening and pre-admission management program was established between November and December 2022. The program was implemented from January to June 2023, and data were collected from patients who received or did not receive pre-admission management. Propensity score matching (PSM) was used to control baseline differences between groups, and causal forest (CF) analysis was performed to assess the effects of pre-admission intervention on hospitalization duration and costs. Results A total of 5,211 patients were included. The CatBoost model had the best results among the nine ML models constructed, and the prediction results showed that tumor type, preoperative waiting time, and age were the main high-risk factors affecting hospital resource utilization. Based on the model, 698 patients received pre-admission management. After PSM, 563 matched pairs were analyzed. CF analysis showed that the pre-admission group had a mean reduction in length of stay of 3.004 [95%CI (-3.334 to -2.675)] days compared with the control group and a mean reduction in hospitalization costs of 1,473.124 [95%CI (-2,166.09 to -780.16)] CNY, indicating that pre-admission management effectively improved hospital resource utilization efficiency. Conclusion Implementing pre-admission management for high-risk patients identified by ML models can significantly reduce length of stay and hospitalization costs for tumor patients, and may positively contribute to optimizing hospital workflows and resource allocation.

【Keywords】Pre-admission model; Length of stay; Hospitalization costs; Propensity score matching; Machine learning; High-risk tumor patients

公立医院高质量发展是落实健康中国战略规划的重要支撑，《国务院办公厅关于推动公立医院高质量发展的意见》提出推进医疗服务模式创新，加快优质医疗资源区域均衡布局。当前，肿瘤已成为威胁我国居民健康的主要公共卫生问题，据国家癌症中心数据显示，我国每年新发肿瘤病例数超 400 万 [1]，其中约 30% 为肿瘤高危人群（如合并多种基础疾病、体能状态较差等），这类患者治疗需求具有特殊性（如术前评估复杂、术后并发症风险高）、治疗过程具有复杂性，因此在保障该群体治疗效果的同时，优化住院流程、缩短非必要住院时间及减少医疗费用尤为关键，已成为医疗管理研究的重要课题。

为此，预住院模式应运而生，该模式核心价值在于显著缩短患者住院时间、优化床位使用率、降低治疗费用，同时减少等待时间与潜在院内感染风险，其实施不仅可提升医院服务流程与工作效率，还能有效减轻患者经济及心理压力，提高患者对医疗服务的总体满意度 [2–3]。但当前预住院患者筛选多依赖临床经验，缺乏数据驱动的高危人群精准识别工具，而机器学习（machine learning，ML）可通过挖掘多维度临床数据，精准识别影响住院资源消耗的关键因素，其分析复杂非线性关系的能力优于传统回归方法 [4]。据此，本研究拟通过构建 ML 模型 [5–6] 优化预住院管理流程，为探索预住院模式在实际医疗服务中的应用效果提供参考。

1 资料与方法

1.1 研究对象

以 2021 年 4 月至 2023 年 6 月期间山西医科

大学第一医院收治的 ICD 编码范围为 C00–D48 的良恶性肿瘤住院患者为研究对象。纳入标准：①病种符合《关于开展“预住院”费用医保支付试点工作》政策；②研究时间窗内入院并出院；③年龄≥ 18 岁。排除标准：①临床关键信息严重缺失，无法确定主要结局或暴露变量，例如缺失入院 / 出院时间或无法确认是否接受手术治疗；②住院期间未接受任何手术相关治疗。若同一患者在研究期间存在多次住院的情况，仅保留符合纳入 / 排除标准的住院数据。本研究经山西医科大学第一医院伦理委员会审核批准（批号：N0.KYLL–2024–017）。

1.2 研究实施

本研究分为三个阶段，详见图 1。

1.2.1 第一阶段

（1）数据收集与预处理

回顾性收集 2021 年 4 月至 2022 年 10 月于山西医科大学第一医院住院治疗的肿瘤患者的一般信息，包括患者性别、年龄、住址、住院次数、合并疾病、肿瘤类型、住院费用、住院天数等，研究所需数据由医院运营部门通过结构化数据库统一提取。

对收集的数据进行预处理，包括：①删除缺失比例超过 30% 的变量，以减少数据噪声对模型训练的影响；②对其余存在缺失值的变量，根据变量类型采用不同策略进行填补，如对连续变量使用中位数填补、对分类变量使用众数填补；③采用 Z–score 标准化方法，以消除不同变量量纲差异对模型性能的影响。

（2）风险因素遴选

以总住院天数和住院费用作为围手术期预后结局的短期替代指标，用于衡量手术治疗后资源消耗与恢复情况。使用 Lasso 回归进行风险因素筛选，分别构建总住院天数和住院费用的影响因素模型。

（3）机器学习模型构建

经数据清洗与预处理后，本研究构建了 9 种 ML 模型，用于总住院天数和费用的预测分析，包括 Logistic 回归（Logistic regression，LR）、支持向量机（support vector machine，SVM）、决策树（decision tree，DT）、随机森林（random forest，RF）、K 近邻算法（K–nearest neighbors，KNN）、梯度提升机（gradient boosting machine，GBM）、CatBoost 分类器（CatBoost）、朴素贝叶斯（naive Bayes，NB）以及极端随机树（extra trees，ET），涵盖从传统统计模型到集成学习算法的多种方法，以全面评估不同模型在预测任务中的表现。对纳入的患者按 7∶3 的比例随机分为训练集和验证集。训练集用于模型构建和交叉验证过程中的参数调优；验证集作为独立测试集，用于评估模型的泛化性能，并检测是否存在过拟合。

（4）模型评估与验证

采用受试者操作特征（receiver operating characteristic，ROC）曲线的曲线下面积（area under curve，AUC）、准确率、精确率、召回率和 F1 分数等指标评估模型区分度。通过绘制模型校准曲线评估模型预测概率与真实情况的一致性，并通过 SHAP 值（SHapley Additive exPlanations）进一步分析各个特征对预测结果的贡献。

1.2.2 第二阶段

基于第一阶段开发的模型，本阶段在实际临床场景中对择期行肿瘤手术的患者进行了风险筛查，并针对高风险患者建立了预住院管理模式。主要工作包括：①运用前期构建的模型在择期行手术的肿瘤患者中识别具有高风险因素的人群；②对高风险患者提前进行术前评估、完善术前各

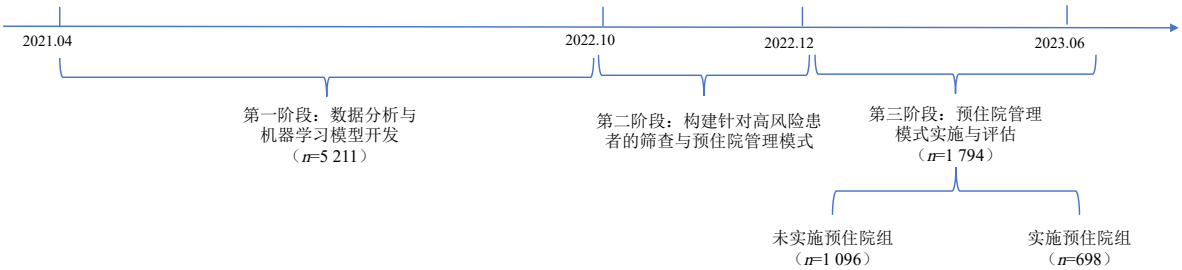

图1 研究流程图

Figure 1. Study flowchart

项检查、检验及相关准备工作等“预住院”操作；③建立“筛查—干预—再评估”闭环机制，确保干预措施的可执行性与记录完整性。

1.2.3 第三阶段

在完成前期筛查流程和模式优化后，对符合纳排标准的患者进行了风险筛查并实施了预住院管理。以是否实施预住院管理为依据，将患者分为干预组与对照组。干预组为经模型识别出的高风险患者并接受预住院管理；对照组则为同期在本院接受择期手术、病情稳定但未被识别为高风险且未接受预住院管理的患者。主要工作包括：①收集两组患者的住院天数、住院费用等关键结局指标；②采用倾向评分匹配（propensity score matching，PSM）方法按照 1∶1 最近邻匹配（卡钳值 0.05，随机种子数 42）选择对照个体，以减少混杂偏倚；③运用因果森林（causal forests，CF）算法评估预住院干预效果，在模型中除处理变量（是否实施预住院）和结局变量（住院天数、住院费用）外，将患者一般特征及合并疾病等均作为协变量，以控制混杂偏倚并提升处理效应估计的准确性。CF 在建模过程中通过样本分裂实现无偏估计，并提供效应估计的置信区间；统计推断采用双侧检验，显著性水平设定为 0.05。

1.3 统计学分析

采用 SPSS 26.0 和 Python 3.1.1 软件进行统计分析。住院天数和住院费用 2 项结局变量按照四分位间距（interquartile range，IQR）法以 Q3 为界整理为二分类变量，分别为长住院组（住院天数≥ Q3）和常规住院组（住院天数< Q3）、高费用组（住院费用≥ Q3）和常规费用组（住院费用< Q3）。符合正态分布的计量资料，采用均数和标准差（$\bar{x} \pm s$）表示，使用 t 检验进行组间比较；不符合正态分布计量资料，采用中位数和四分位数 [M（P_{25}，P_{75}）] 表示，使用 Mann-Whitney U 检验进行组间比较。计数资料采用例数和百分比（n，%）表示，使用卡方检验或 Fisher 精确概率检验进行组间比较。双侧检验 $P < 0.05$ 被认为差异具有统计学意义。

2 结果

2.1 一般情况

第一阶段回顾性收集 5 211 例住院肿瘤患者的临床数据，随机分为训练集（3 998 例）和验证集（1 213 例），两组在年龄、家庭住址、是否具有呼吸系统疾病、术前等待天数、实施预住院管理术前等待天数、总住院天数、总住院费用方面差异具有统计学意义（$P < 0.05$），见表 1。

2.2 风险因素的筛选

分别以住院天数与住院费用为结局变量，纳入表 1 所列的所有临床和人口学特征构建 Lasso 回归模型进行变量选择，见附件图 1、图 2。通过 10 折交叉验证确定的最优正则化参数（λ_min）

表1 住院肿瘤患者一般特征比较（n，%）

Table 1. Comparison of general characteristics of hospitalized cancer patients (n, %)

特征	训练集（n=3 998）	验证集（n=1 213）	$t/\chi^2/Z$值	P值
性别			0.82	0.37
女	2 998（74.99）	894（73.70）		
男	1 000（25.01）	319（26.30）		
年龄*	47.89 ± 9.54	49.15 ± 14.04	2.93	<0.01
家庭住址			10.27	<0.01
山西省内	3 388（84.74）	1 073（88.46）		
山西省外	610（15.26）	140（11.54）		
高血压			0.40	0.53
否	3 175（79.41）	951（78.40）		
是	823（20.59）	262（21.60）		
糖尿病			1.42	0.23
否	3 656（91.45）	1 122（92.50）		
是	342（8.55）	91（7.50）		
心脏病			0.07	0.79
否	3 838（96.00）	1 166（96.13）		
是	160（4.00）	47（3.87）		

续表1

特征	训练集（n=3 998）	验证集（n=1 213）	$t/\chi^2/Z$值	P值
呼吸系统疾病			4.17	0.04
否	3 660（91.55）	1 133（93.40）		
是	338（8.45）	80（6.60）		
肿瘤类型			0.38	0.54
良性	1 607（40.20）	475（39.16）		
恶性	2 391（59.80）	738（60.84）		
住院次数			0.46	0.50
首次入院	3 342（83.59）	1 004（82.77）		
多次入院	656（16.41）	2 09（17.23）		
术前等待天数#	4（2，6）	3（2，5）	4.97	<0.01
预住院管理术前等待天数#	0（0，0）	0（0，5）	9.80	<0.01
总住院天数#	9（6，14）	6（5，8）	18.68	<0.01
总住院费用（元）#	18 116.50（8 498.98，29 548.45）	21 036.74（19 236.60，23 378.10）	8.01	<0.01

注：*符合正态分布的计量资料，采用均数和标准差（$\bar{x} \pm s$）表示；#不符合正态分布的计量资料，采用中位数和四分位数[M（P_{25}，P_{75}）]表示。

对应的模型中，输入变量的回归系数均为非零，提示在当前数据特征与参数设置下，各变量对结局变量均具有一定的预测贡献。由于变量间共线性较弱、样本量充足，正则化未显著压缩系数，侧面说明所纳入的临床和人口学特征对住院时长及费用均具有较强解释力。

2.3 模型训练和验证

通过计算不同 ML 模型的关键性能指标后发现，在住院天数和费用模型中，CatBoost 模型表现最佳，其性能表现如表 2、附件图 3 所示。

2.4 特征重要性分析

住院费用模型中特征重要性排序分别为肿瘤类型、术前等待天数、年龄、性别、高血压、入院次数等（图 2–A），其中恶性肿瘤、术前等待天数较长、年龄较大、合并高血压、糖尿病等变量呈正向影响（提高住院费用）（图 2–B）。住院天数模型的预测结果主要术前等待天数、肿瘤类型、年龄、住院次数等特征的显著影响（图 2–C），其中术前等待天数、年龄大、入院次数多等体现为正向影响（延长住院天数）（图 2–D）。

表2 不同机器学习模型性能比较

Table 2. Comparison of performance of different machine learning models

模型	AUC	准确率	精确率	召回率	F1分数
总住院天数					
LR	0.834 9	0.815 9	0.763 6	0.646 2	0.700 0
SVM	0.860 1	0.823 5	0.806 5	0.617 3	0.699 3
DT	0.746 2	0.759 0	0.637 2	0.638 5	0.637 8
RF	0.832 0	0.784 5	0.680 5	0.663 5	0.671 9
KNN	0.814 7	0.791 6	0.705 5	0.640 4	0.671 4
GBM	0.872 0	0.825 4	0.790 6	0.646 2	0.711 1
CatBoost	0.876 1	0.828 6	0.801 4	0.644 2	0.714 3
NB	0.797 4	0.767 3	0.663 2	0.609 6	0.635 3
ET	0.792 4	0.773 7	0.670 8	0.626 9	0.648 1
总住院费用					
LR	0.785 0	0.771 7	0.653 1	0.370 4	0.472 7
SVM	0.801 4	0.769 8	0.653 8	0.354 2	0.459 5
DT	0.655 4	0.705 9	0.468 9	0.488 4	0.478 5
RF	0.782 2	0.753 8	0.558 3	0.520 8	0.538 9
KNN	0.773 0	0.757 7	0.573 4	0.479 2	0.522 1
GBM	0.826 7	0.781 3	0.635 5	0.488 4	0.552 4
CatBoost	0.832 9	0.786 4	0.655 1	0.479 2	0.553 5
NB	0.782 4	0.734 0	0.520 3	0.474 5	0.496 4
ET	0.750 6	0.739 1	0.530 6	0.481 5	0.504 9

2.5 预住院管理模式实施效果评价

2.5.1 实施与未实施组一般情况比较

2022 年 11 月至 12 月期间，基于最优模型 CatBoost 的预测结果，识别出部分风险较高的患者，并优先安排实施预住院管理。2023 年 1 月至 6 月期间，共收治住院患者 1 794 例，其中 698 例为接受模型辅助筛查后实施预住院的患者，1 096 例为未实施预住院的患者。采用 PSM 法进行匹配后共获得 563 对匹配样本，两组基线可比，P 值均> 0.05，见表 3。

2.5.2 实施效果评价

CF 结果显示，与对照组相比，接受预住院干预的患者平均住院时长显著缩短 3.004 [95%CI（−3.334，−2.675）]d，住院总费用平均减少 1 473.124 [95%CI（−2 166.093，−780.155）] 元，两者差异均具有统计学意义（P < 0.001），见表 4。结果表明，预住院干预在优化住院流程、缩短住院时间和控制住院成本方面具有显著效果。

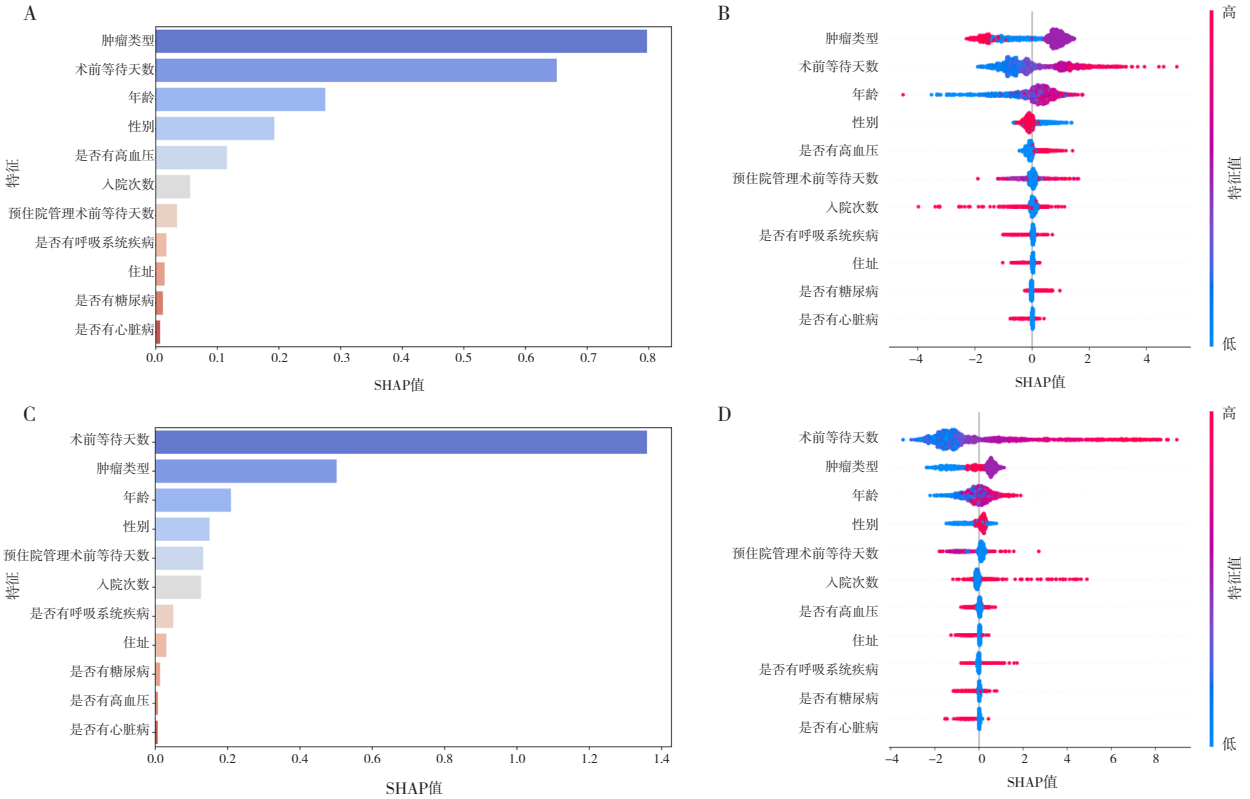

图2 住院费用与住院天数的特征重要性分析

Figure 2. Feature importance analysis for hospitalization costs and length of hospital stay

注：A.住院费用特征重要性排序；B.住院费用SHAP图；C.住院天数特征重要性排序；D.住院天数SHAP图。。

表3 PSM匹配前后两组一般资料比较（n，%）

Table 3. Comparison of general characteristics of two groups before and after PSM (n, %)

资料	PSM前				PSM后			
	非预住院组（n=1 096）	预住院组（n=698）	$t/\chi^2/Z$值	P值	非预住院组（n=563）	预住院组（n=563）	$t/\chi^2/Z$值	P值
性别			2.99	0.84			<0.01	0.94
女	765 (69.80)	460 (65.90)			421 (74.78)	422 (74.96)		
男	331 (30.20)	238 (34.10)			142 (25.22)	141 (25.04)		
年龄（岁）*	45.60 ± 14.50	48.30 ± 15.70	3.66	<0.01	47.60 ± 13.50	47.50 ± 13.30	0.13	0.90
住址			5.70	0.01			<0.01	0.99
省内	875 (79.84)	589 (84.38)			463 (82.24)	464 (82.42)		
省外	221 (20.16)	109 (15.62)			100 (17.76)	99 (17.58)		
高血压			1.70	0.19			0.39	0.53
无	920 (83.94)	570 (81.66)			468 (83.13)	460 (81.71)		
有	176 (16.06)	128 (18.34)			95 (16.87)	103 (18.29)		

续表3

资料	PSM前				PSM后			
	非预住院组 (*n*=1 096)	预住院组（*n*=698）	$t/\chi^2/Z$值	*P*值	非预住院组（*n*=563）	预住院组（*n*=563）	$t/\chi^2/Z$值	*P*值
糖尿病			1.16	0.28			0.07	0.80
无	1 040 (94.89)	654 (93.70)			532 (94.49)	530 (94.14)		
有	56 (5.11)	44 (6.30)			31 (5.51)	33 (5.86)		
心脏病			0.04	0.85			0.32	0.57
无	1 050 (95.80)	670 (95.99)			539 (95.74)	535 (95.03)		
有	46 (4.20)	28 (4.01)			24 (4.26)	28 (4.97)		
呼吸系统疾病			4.76	0.03			1.04	0.31
无	1 000 (91.24)	615 (88.11)			515 (91.47)	505 (89.70)		
有	96 (8.76)	83 (11.89)			48 (8.53)	58 (10.30)		
肿瘤类型			0.42	0.52			0.08	0.78
良性肿瘤	250 (22.81)	150 (21.49)			128 (22.74)	124 (22.02)		
恶性肿瘤	846 (77.19)	548 (78.51)			435 (77.26)	439 (77.98)		
住院次数			2.42	0.12			0.07	0.79
首次入院	930 (84.85)	610 (87.39)			490 (87.03)	493 (87.57)		
多次入院	166 (15.15)	88 (12.61)			73 (12.97)	70 (12.43)		
术前等待天数#	4 (3, 6)	1 (1, 3)	15.64	<0.001	3 (1, 4)	3 (2, 5)	0.53	0.58

注：*符合正态分布的计量资料，采用均数和标准差（$\bar{x} \pm s$）表示；#不符合正态分布的计量资料，采用中位数和四分位数[*M*（P_{25}，P_{75}）]表示。

表4 预住院管理实施后效果评价

Table 4. Outcome evaluation after implementation of pre-admission management

变量	效应值（95%CI	标准差	*Z*值	*P*值
住院时长（d）				
总体住院时长	−3.004（−3.334，−2.675）	0.168	−17.879	<0.001
模型开发期	−2.014（−2.075，−1.504）	0.306	−6.871	<0.001
预住院管理期	−2.413（−2.924，−1.901）	0.261	−9.243	<0.001
结果评估	−4.529（−5.207，−3.852）	0.346	−13.103	<0.001
住院费用（元）				
总体住院费用	−1 473.124（−2 166.093，−780.155）	355.562	−4.167	<0.001
模型开发期	−586.989（−1 863.741，689.762）	651.416	−0.901	0.368
预住院管理期	−1 178.865（−2 072.795，−284.936）	456.095	−2.585	0.010
结果评估期	−2 979.066（−4 688.197，−1 269.935）	872.022	−3.416	0.001

3 讨论

本研究通过机器学习、PSM 和 CF 分析方法[7-11]，揭示了预住院模式能显著缩短肿瘤患者住院时长并降低医疗费用。传统上，患者需在医院内完成所有的术前准备工作，这不仅增加了患者的等待时间，也降低了医院的服务效率[12]。预住院模式通过将部分术前准备工作前置，不仅减少了患者的实际住院时间，也使得医院能够更灵活地安排手术和床位使用，从而提升整体的医疗服务效率和质量。

本研究显示预住院组较对照组住院时长平均减少 3.004 d，住院费用总体降低 1 473.124 元，该结果与多项既往研究[13-16]结论一致。本研究首次结合 CF 分析量化预住院的干预效应，且纳入样本量显著大于既往研究，结果稳定性较强。CatBoost 对住院时长 / 费用的预测 AUC 值较高，显著优于传统回归模型，意味着通过系统的术前准备和优化流程可有效提高医院床位的利用效率，减轻患者经济负担，这在当前医疗资源紧张、医疗费用持续上升的背景下，具有重要的实践价值和应用前景。

本文采用了 SHAP 方法[17-19]对住院天数和住院费用的影响因素进行了分析，揭示了预住院模式的作用途径，术前天数是住院时长的首要预测因子，预住院通过将检查前移，使术前天数从

3 d 压缩至 1 d，直接减少无效住院日；肿瘤类型的重要性在费用模型中排名居首，但预住院通过缩短住院周期，显著降低了恶性肿瘤的高费用负担。该评估揭示了影响患者接受预住院以及治疗费用和住院时长的关键协变量，肿瘤类型、术前天数、年龄被确定为最主要的潜在因素。该结果为个性化的预住院决策提供了依据。在实施预住院策略时，特别应考虑患者的肿瘤类型和年龄等因素，以更好地发挥预住院模式的效果。

本研究结果为医疗政策制定者提供了重要的参考信息。通过促进预住院模式的实施，不仅能提高医疗服务的效率，还可作为医疗成本控制的一种有效手段。政策制定者应考虑在医疗政策中纳入对预住院模式的支持，如提供必要的财政和政策支持，鼓励医院改进服务流程，提高医疗服务的整体效率和质量[20]。

本研究也存在一定局限性，如未分析费用构成（检查费、药费、耗材费）的降幅差异，以及肿瘤本身的特性在本研究中对住院时间的主导影响可能部分掩盖或改变了慢性病（如呼吸系统、心脏病）的作用模式。未来需结合更精细的临床数据和深入分析（如按疾病严重程度分层）来验证这一关联并阐明其背后的确切机制。此外，也应该探索预住院模式在不同类型的疾病和患者群体中的适用性和效果，以及研究该模式对医院其他方面（如患者安全、医疗质量等）的影响。

综上所述，本研究通过构建 ML 模型结合 SHAP 归因分析，明确术前天数、肿瘤类型、年龄为资源消耗的核心因素，为优化住院流程奠定了基础。预住院模式不仅提升了医疗服务的效率与质量，还减轻了患者的经济负担，优化了医疗资源的利用效率，对患者、医院及整个医疗体系将产生积极影响。

附件见《医学新知》官网附录（https://yxxz.whuznhmedj.com/futureApi/storage/appendix/202503188.pdf）

伦理声明：本研究已获得山西医科大学第一医院伦理委员会审核批准（批号：N0.KYLL-2024-017）

作者贡献：研究设计：周培斌、弓玉红；数据采集：王文天、杨壮、杨露露；数据分析：尚楠、贺娅莎、张绍果、刘艳梅、高红梅、赵婧媛；论文撰写：弓玉红、尚楠；论文审定：周培斌

数据获取：本研究中使用和（或）分析的数据可联系通信作者获取

利益冲突声明：无

致谢：不适用

参考文献

1 郑荣寿，陈茹，韩冰峰，等. 2022 年中国恶性肿瘤流行情况分析 [J]. 中华肿瘤杂志，2024, 46(3): 221-231. [Zheng RS, Chen R, Han BF, et al. Cancer incidence and mortality in China, 2022[J]. Chinese Journal of Oncology, 2024, 46(3): 221-231.] DOI: 10.3760/cma.j.cn112152-20240119-00035.

2 张欣，李维，张晨，等. 基于数据中心的预住院模式效果评价[J]. 中国卫生标准管理，2020, 11(17): 18-21. [Zhang X, Li W, Zhang C, et al. Effect assessment of pre-registration hospitalization based on clinical data repository[J]. China Health Standard Management, 2020, 11(17): 18-21.] DOI: 10.3969/j.issn.1674-9316.2020.17.007.

3 李小芳，宋慧，蔡滨，等. 基于预住院开展多学科参与的麻醉术前访视应用思考 [J]. 现代医院，2021, 21(10): 1557-1560. [Li XF, Song H, Cai B, et al. Application of multidisciplinary participation in anesthesia preoperative visit based on pre-admission [J]. Modern Hospital, 2021, 21(10): 1557-1560.] DOI: 10.3969/j.issn.1671-332X.2021.10.024.

4 丁四清，陆晶，秦春香，等. 数据挖掘在护理不良事件管理中的应用进展 [J]. 中华护理杂志，2019, 54(6): 873-877. [Ding SQ, Lu J, Qin CX, et al. Application progress of data mining in the management of nursing adverse events[J]. Chinese Journal of Nursing, 2019, 54(6): 873-877.] DOI: 10.3761/j.issn.0254-1769.2019.06.013.

5 Chang TH, Stuart EA. Propensity score methods for observational studies with clustered data: a review[J]. Stat Med, 2022, 41(18): 3612-3626. DOI: 10.1002/sim.9437.

6 Bonander C, Svensson M. Using causal forests to assess heterogeneity in cost-effectiveness analysis[J]. Health economics, 2021, 30(8): 1818-1832. DOI: 10.1002/hec.4263.

7 Brinkman-Stoppelenburg A, Polinder S, Olij BF, et al. The association between palliative care team consultation and hospital costs for patients with advanced cancer: an observational study in 12 Dutch hospitals[J]. Eur J Cancer Care (Engl), 2020, 29(3): e13198. DOI: 10.1111/ecc.13198.

8 Mooney K, Titchener K, Haaland B, et al. Evaluation of oncology hospital at home: unplanned health care utilization and costs in the huntsman at home real-world trial[J]. J Clin Oncol, 2021, 39(23): 2586-2593. DOI: 10.1200/JCO.20.03609.

9 Numico G, Cristofano A, Mozzicafreddo A, et al. Hospital admission of cancer patients: avoidable practice or necessary care?[J]. Plos One, 2015, 10(3): e0120827. DOI: 10.1371/journal.pone.0120827.

10 May P, Garrido MM, Cassel JB, et al. Using length of stay to control for unobserved heterogeneity when estimating treatment

effect on hospital costs with observational data: issues of reliability, robustness, and usefulness[J]. Health Serv Res, 2016, 51(5): 2020–2043. DOI: 10.1111/1475–6773.12460.

11 刘同柱，赵昕昱，童贵显．我国医院预住院模式研究现状与对策[J]. 中国医院管理，2018, 38(7): 26–28. [Liu TZ, Zhao XY, Tong GX. Research status and improvement strategies of pre–registration hospitalization mode in Chinese hospitals[J]. Chinese Hospital Management, 2018, 38(7): 26–28.] DOI: CNKI:SUN:YYGL.0.2018–07–014.

12 刘亚楠，张瑞，弓玉红，等．基于 PSM 法的预住院模式在甲状腺外科手术中的应用效果评价[J]. 中国医院管理，2023, 43(7): 38–41, 52. [Liu YN, Zhang R, Gong YH, et al. Evaluation on the application effect of pre–hospitalization mode in thyroid surgery based on propensity score matching method[J]. Chinese Hospital Management, 2023, 43(7): 38–41, 52.] https://d.wanfangdata.com.cn/periodical/ChVQZXJpb2RpY2FsQ0hJMjAyNTA2MjISD3pneXlnbDIwMjMwNzAxMBoIZTdpMzY5NXE%3D

13 Chen Y, Huang L, Chen X. Application of a selective hospitalization model in the clinical practice of breast surgery[J]. Medicine (Baltimore), 2023, 102(26): e34209. DOI: 10.1097/MD.0000000000034209.

14 曾洪兰，姚嘉，周钰，等．预住院管理模式在乳腺外科手术中的应用效果[J]. 中国地方病防治杂志，2017, 32(4): 448. [Zeng HL, Yao J, Zhou Y, et al. Application effect of pre–admission management mode in breast surgery[J]. Chinese Journal of Control of Endemic Diseases, 2017, 32(4): 448.] DOI: CNKI:SUN:DYBF.0.2017–04–053.

15 鲍萍，申建民，查浩，等．某三甲医院预住院管理模式的实践与体会[J]. 江苏卫生事业管理，2016, 27(5): 36–37. [Bao P, Shen JM, Zha H, et al. Practice and experience of pre–admission management mode in a tertiary hospital[J]. Jiangsu Healthcare Administration, 2016, 27(5): 36–37.] DOI: CNKI:SUN:JWSG.0.2016–05–022.

16 Lundberg S, Lee SI. A unified approach to interpreting model predictions[J]. In Proceedings of the 31st International Conference on Neural Information Processing Systems (Long Beach, California, USA) (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, 4768–4777. 2017. DOI: 10.48550/arXiv.1705.07874.

17 徐良辰，郭崇慧．基于集成学习的胃癌生存预测模型研究[J]. 现代图书情报技术，2021, 5(8): 86–99. [Xu LC, Guo CH. Predicting survival rates for gastric cancer based on ensemble learning[J]. Data Analysis and Knowledge Discovery, 2021, 5(8): 86–99.] DOI: 10.11925/infotech.2096–3467.2021.0045.

18 王鑫，廖彬，李敏，等．融合 LightGBM 与 SHAP 的糖尿病预测及其特征分析方法[J]. 小型微型计算机系统，2022, 43(9): 1877–1885. [Wang X, Liao B, Li M, et al. Combination of light GBM and SHAP for diabetes prediction and feature analysis[J]. Journal of Chinese Computer Systems, 2022, 43(9): 1877–1885.] DOI: 10.20009/j.cnki.21–1106/TP.2022–09–016.

19 陈小昆，左航旭，廖彬，等．融合 XGBoost 与 SHAP 的冠心病预测及其特征分析模型[J]. 计算机应用研究，2022, 39(6): 1796–1804. [Chen XK, Zuo HX, Liao B, et al. Coronary artery disease prediction and feature analysis model based on XGBoost and SHAP[J]. Application Research of Computers, 2022, 39(6): 1796–1804.] DOI: 10.19734/j.issn.1001–3695.2021.11.0639.

20 张国杰，张红梅，柏庆花，等．公立医院预住院流程建设探索与思考[J/OL]. 协和医学杂志，2025: 1–14. [2025–09–15]. [Zhang GJ, Zhang HM, Bai QH, et al. Exploration and reflection on the construction of pre–admission process in public hospitals[J/OL]. Xiehe Medical Journal, 2025: 1–14. Published online September 15, 2025.] https://link.cnki.net/urlid/11.5882.R.20250711.1016.002

收稿日期：2025 年 03 月 24 日 修回日期：2025 年 09 月 15 日
本文编辑：桂裕亮 曹 越

引用本文：弓玉红, 尚楠, 贺娅莎, 等. 机器学习模型辅助肿瘤手术高危患者筛查及预住院模式效果评价[J]. 医学新知, 2025, 35(9): 987–995. DOI: 10.12173/j.issn.1004–5511.202503188.
Gong YH, Shang N, He YS, et al. Research on machine learning model–assisted screening of high–risk tumor surgery patients and the effectiveness of pre–admission management[J]. Yixue Xinzhi Zazhi, 2025, 35(9): 987–995. DOI: 10.12173/j.issn.1004–5511.202503188.