

肝脏术后非手术部位并发症风险预测机器学习模型开发

徐立宁^{1*}, 王炜煜², 杨波³

1. 中国人民解放军总医院第二医学中心综合外科, 国家老年疾病临床医学研究中心, 北京 100853

2. 武汉大学肝胆疾病研究院, 武汉大学中南医院, 武汉 430071

3. 华中科技大学同济医学院附属协和医院放射科, 武汉 430022

* 联系人, E-mail: xu_lining@aliyun.com

2025-04-05 收稿, 2025-07-06 修回, 2025-07-11 接受, 2025-08-04 网络版发表

摘要 术后并发症分为手术部位并发症和非手术部位并发症, 之前的研究多聚焦于手术部位并发症, 目前尚无采用术前因素来综合评估肝脏术后非手术部位并发症发生风险的研究. 本研究的目的是建立一个可解释的机器学习模型, 将术前因素(包括手术预案相关因素)与肝脏术后非手术部位并发症的发生风险预测联系起来. 共有2053名参与者入组这项研究. 使用最小绝对值收敛和选择算子(least absolute shrinkage and selection operator, LASSO)回归筛选可能的预测因子组合成模型的预测因子, 建立了7个机器学习模型, 通过评估术前因素来预测肝脏术后非手术部位并发症的发生风险. 采用曲线下面积(area under curve, AUC)作为主要评价指标, 绘制各个模型的受试者工作特征(receiver operating characteristic, ROC)曲线, 校准曲线, 决策曲线分析(decision curve analysis, DCA)曲线, 选出最优模型. 使用沙普利加性解释法(shapley additive explanations, SHAP)来解释特征. 最终选择了基于10种术前因素的具有最佳表现的极度梯度提升(extreme gradient boosting, XGBoost)模型来预测肝脏术后非手术部位并发症发生风险(验证集AUC: 0.758, 95%置信区间: 0.699~0.817, 准确率: 0.859; 校准曲线显示预测概率与实际概率具有良好的 consistency; DCA曲线整体表现最佳). 体重指数(SHAP: 0.0532)、病变位置(SHAP: 0.0272)、异常的碱性磷酸酶值(SHAP: 0.0170)、非肝脏手术的腹部手术史(SHAP: 0.0125)、超过5种并存病(SHAP: 0.0114)、黄疸(SHAP: 0.0100)、肝切除术史(SHAP: 0.0030)、毗邻脏器侵犯(SHAP: 0.0013)对模型预测结果有正向作用, 已婚(SHAP: 0.0028)和正常的白蛋白值(SHAP: 0.0132)对模型预测结果有负向作用. 结果表明, XGBoost模型可准确预测术前因素与肝脏术后非手术部位并发症之间的关联.

关键词 肝切除术, 非手术部位并发症, 预测模型, 机器学习

肝脏外科是一门非常困难的学科, 在全球范围内造成了严重的医疗负担, 需要不断艰苦的探索和实践^[1]. 肝切除术具有高的术后并发症的发生率. 术后并发症分为手术部位并发症和非手术部位并发症. 虽然随着肝脏手术技术的革新, 手术部位并发症发生率逐渐降低^[2,3], 发生的手术部位并发症的严重程度也逐渐减小, 但是由于年龄、并存病等因素^[4], 非手术部位并发症(即手术部位之外的全身其他系统的并发症)仍是

一个严重的问题^[5-8]. 如何在术前评估非手术部位并发症的发生风险并给予早期预警和干预以减少术后非手术部位并发症的发生是一个临床问题^[9-11]. 尽管根据围术期因素已经建立了一些风险评估系统, 但由于其多聚焦于手术部位并发症和其他单个脏器的并发症, 很少有专门针对非手术部位并发症发生风险进行综合评估的研究^[12,13]. 且在现有文献中, 均未见有正式明确提出非手术部位并发症的概念.

引用格式: 徐立宁, 王炜煜, 杨波. 肝脏术后非手术部位并发症风险预测机器学习模型开发. 科学通报, 2025, 70: 4978–4989

Xu L, Wang W, Yang B. Development of a risk prediction model for non-surgical-site complication after hepatectomy based on interpretable machine learning (in Chinese). Chin Sci Bull, 2025, 70: 4978–4989, doi: [10.1360/CSB-2025-0439](https://doi.org/10.1360/CSB-2025-0439)

本研究首次明确提出了非手术部位并发症的概念,并对此进行针对性的研究。而且,既往的大多数研究采用的是传统的统计分析方法,预测效果并不理想,导致其在指导临床实践方面存在诸多问题,使这些系统的应用受到限制,尚未得到广泛认可^[14]。因此,需要对非手术部位并发症这个崭新范畴的发生风险,采用新的分析方法来进行更准确地预测。

在现有的研究中,采用传统的统计方法来预测临床结局时,需要满足许多数据标准和数据准备条件^[15],这为临床决策带来了诸多挑战。近年来,机器学习技术在预测临床结局方面的应用日益增多,其核心任务是利用患者初期的数据预测其对特定治疗方案的响应程度^[16,17]。这种预测能够帮助医生及早了解治疗可能的结果,并选择最佳的治疗策略,旨在最大化治疗效果并减少副作用的发生^[18]。机器学习不依赖于严格的统计假设,通过算法和模型的训练,使计算机能够从数据中学习,并基于学习的经验做出预测或决策。机器学习更加注重自动化和自适应性的数据处理,包括数据预处理、特征提取和特征选择^[19]。机器学习方法还可以通过降维技术来减少数据维度,以便更好地处理高维数据。因此,机器学习适用于多种的数据分析需求,可以通过对大量数据的筛选、分类和分析来探索隐藏在多个数据集背后的含义,这些信息可用于支持疾病诊断、风险识别和健康决策^[20]。

本研究使用了来自多中心的数据来回顾性研究术前因素与肝脏术后非手术部位并发症之间的关联,建立了7种可用于通过术前因素来识别肝脏术后非手术部位并发症的机器学习模型,并比较了这些模型的性能特征,筛选出最佳模型,确定了术前因素对肝脏术后非手术部位并发症识别的贡献,从而为术前干预提供依据,以降低术后非手术部位并发症发生的概率。

1 方法

1.1 研究对象

本研究目的是采用机器学习的方法建立肝脏术后非手术部位并发症发生风险预测模型。在2017年1月1日至2023年12月31日期间,共有3家医院的2053例接受肝切除术的患者最终入组本项回顾性研究。所有这些医院都是中国的国家级医院或地区中心医院,其病人来自中国各地。所有这些医院都根据当地医疗管理部门的要求获得了各自的批准。在涉及人类参与者的研

究中所进行的所有程序都符合中国人民解放军总医院伦理委员会的伦理标准(S2022-664-01)和1975年《赫尔辛基宣言》及其后来的修正案或类似的伦理标准。由于本研究是一项对参与者无风险的非干预性研究,且是回顾性研究,针对参与者的个人知情同意是不可行的,因此伦理委员会同意放弃知情同意。这些数据在中国人民解放军总医院进行了集中整理和分析。分析在2024年7月1日至12月31日之间进行。

所有研究的入选者都有明确的手术指征。对于恶性肿瘤患者,术前没有进行新辅助治疗等药物治疗。恶性病例主要包括肝细胞癌和胆管细胞癌。本研究中的良性病例主要包括肝胆管结石和肝血管瘤。在中国,肝胆管结石是最常见的需要手术的良性肝病。肝胆管结石患者由于其特殊的病理特征,常发生感染和肝功能异常,可导致局部肝损伤,这些病例需要施行部分肝切除术。其他良性病例,包括肝血管瘤,如果由于病变体积大、压迫周围器官、引起腹胀和其他临床症状,则需要手术。只有符合手术指征的患者才接受手术。

1.2 纳入标准

纳入标准如下:(1)符合肝切除术的手术适应证,接受选择性肝切除术;(2)具有完整的医疗记录;(3)非近期二次手术;(4)并发症发生的起始时间位于术后30日内;(5)并存病控制良好。

1.3 排除标准

排除标准如下:(1)术前行新辅助治疗等全身药物治疗;(2)有肝外远处脏器转移肿瘤病灶;(3)合并血管重建;(4)术前预定的手术方案与实际手术方案不一致。

1.4 变量资料

所有入选患者都有包含详细病史和体检结果信息的完整医疗记录。根据常用的手术风险评分系统和我们团队之前的临床研究经验,分析的围手术期术前因素和手术预案相关因素包括患者的基本信息、诊断、实验室检查、手术方案、相关疾病和用药史等。

社会人口特征变量包括:年龄、性别、婚姻(仅按手术时的婚姻状态是否为已婚分为两类,即已婚和非已婚,未婚、离异均按非已婚计算)、学历等。术前患者一般状况及病变特征变量包括:体重指数(body mass index, BMI)、虚弱情况、体重变化情况、患并存病超过5种、睡眠异常、食欲不振、患转移性肝癌、患黄

疸、患肝硬化、非肝切除术的腹部手术史、多发性病变、术前介入栓塞、再次肝切除术、Child肝功能分级、病变部位、病变大小、淋巴侵犯、毗邻脏器侵犯、肝脏外伤、肝结石、肝脏寄生虫、恶性肿瘤、术前住院天数。术前实验室检查变量包括：碱性磷酸酶(alkaline phosphatase, ALP)值、白蛋白(albumin, ALB)值、丙肝抗体(hepatitis C virus antibody, HCVAb)阳性、乙肝表面抗原(hepatitis B surface antigen, HBsAg)阳性。手术预期方案变量包括：机器人手术、腹腔镜手术、邻近器官切除、胆管成形术、切除方式。研究结局：是否发生非手术部位并发症。

1.5 定义

并发症是指疾病在发展过程中引起的另一种疾病或症状的发生,后者是前者的复杂化。本研究中术后并发症定义为术后30天内发生的并发症。本团队明确将手术并发症分为手术部位并发症(surgical-site complication, SSC)和非手术部位并发症(non-surgical-site complication, NSSC),并对其概念进行了阐述。手术部位是指包括手术路径在内的需要进行外科手术操作的所有组织、身体区域或器官,包括切口和/或深部器官和/或腔隙。手术部位并发症是指术后仅涉及手术部位的并发症,肝脏术后的手术部位并发症包括肝脏出血、肝脓肿、切口感染等。肝脏术后非手术部位并发症即全身其他系统的并发症,包括肺部感染、心功能不全、下肢静脉血栓等。由于手术部位并发症的发生受手术技术、手术器械、手术者熟练程度等因素的影响较大,非手术部位并发症多与病人身体状况、并存病情相关而受手术操作者本身的影响较小,同时由于手术技术的改进,手术部位并发症的发生率越来越低,而随着人口老龄化,具有多种疾病的身体状况欠佳的患者越来越多,故而本研究为了明确术前因素对术后非手术部位并发症的影响,仅统计术后非手术部位并发症,而不是术后所有并发症。

由于简单并发症对于术后康复无明显影响,故而本研究排除简单并发症,本研究的并发症仅包括Clavien-Dindo手术并发症分类 $\geq$ II级别的并发症,即患者的结局分组按照是否发生了Clavien-Dindo $\geq$ II级别的并发症分为相对应的两组。死亡亦指由术后非手术部位并发症所导致的死亡,而不包括手术部位并发症导致的死亡,例如术后手术部位大出血所导致的死亡应排除。

1.6 数据处理

标签定义中: No赋值0, Yes赋值1; 性别: 男性赋值0, 女性赋值1; ALB和ALP: 根据临床检验标准, 数值异常($ALB < 35$ g/L; $ALP > 130$ U/L)赋值0, 数值正常($ALB \geq 35$ g/L; $ALP \leq 130$ U/L)赋值1; Child肝功能分级: 本组病例中, Child肝功能级别均为A级或B级, A级赋值为1, B级赋值为2; 病变按大小分组: 参照中华医学会外科学分会肝脏外科学组对肝脏大小的分类方法, 无法计算大小(例如肝胆管狭窄、结石等)赋值0, ≤ 2 cm赋值1, > 2 且 ≤ 5 cm赋值2, > 5 且 ≤ 10 cm赋值3, > 10 cm赋值4; 切除方式: 依据切除时的不同解剖方式, 非规则性肝切除赋值1, 规则性肝切除赋值2, 非规则性肝切除+规则性肝切除赋值3; 病变部位: 依据Couinaud分段法, 病变位于1~8段分别赋值1~8、左外叶赋值9、左内叶赋值10、左半肝赋值11、左三叶赋值12、右前叶赋值13、右后叶赋值14、右半肝赋值15、右三叶赋值16、中肝赋值17、其他多个散在分布肝段(非规则肝段分布)赋值18。

1.7 统计分析方法

1.7.1 人群描述

以是否发生非手术部位并发症作为分组变量, 两组正态分布的连续数据采用平均值 $\pm$ 标准差(Mean $\pm$ SD)表示, 采用两独立样本的 t 检验; 偏态分布的变量采用 $M(Q_1, Q_3)$ 表示, 采用两独立样本的秩和检验; 分类的数据采用 $n(\%)$ 表示, 采用卡方检验或Fisher确切概率法。

1.7.2 数据拆分

设置随机种子, 对R对数据以7:3进行随机拆分, 70%为训练集(train), 剩下30%为验证集(test), 并进行两组人群的均衡性检验。

1.7.3 筛选可能的预测因子

在训练集数据中, 以曲线下面积(area under curve, AUC)为主要评价指标, 采用10折交叉验证的形式, 使用最小绝对值收敛和选择算子(least absolute shrinkage and selection operator, LASSO)回归筛选可能的预测因子组合成模型的预测因子。

1.7.4 模型构建

设置随机种子, 对数据进行随机拆分, 70%为训练集, 剩余30%为验证集。在训练集中针对数据采用LASSO回归筛选出特征变量, 然后构建逻辑回归(logistic regression, LR)、决策树(decision tree, DT)、极度梯度

提升(extreme gradient boosting, XGBoost)、随机森林(random forest, RF)、K-最近邻(K-nearest neighbors, KNN)、轻量的梯度提升机(light gradient boosting machine, LightGBM)和人工神经网络(artificial neural network, ANN)的预测模型。其中机器学习采用以训练集的AUC作为评价指标进行网格搜索方式调参。

1.7.5 最终的模型

采用AUC (95%置信区间(confidence interval, CI))作为主要的模型评价指标, 并计算每个模型的准确性、灵敏度、特异性、阴性预测值和阳性预测值, 绘制各个模型的受试者工作特征(receiver operating characteristic, ROC)曲线, 校准曲线, 决策曲线分析(decision curve analysis, DCA)曲线, 以AUC作为评价指标, 选出各个阶段的最优模型。软件: 采用R语言(版本4.3.3)进行所有的统计分析, 以 $P<0.05$ 认为差异有统计学意义。本研究的整体流程见图S1。

2 结果

2.1 总人群基本特征

本次研究共纳入2053例, 其中314例发生术后非手术部位并发症。总人群中, 年龄为: 48.20 ± 12.42 岁, 平均BMI为: $21.98\pm2.15\text{ kg/m}^2$ 。发生术后非手术部位并发症的人群和未发生术后非手术部位并发症的人群在BMI、切除肝段数目等方面存在显著差异。具体资料详见表S1。

2.2 均衡性检验

对数据进行随机拆分, 70%为训练集(train), 剩下30%为验证集(test), 并进行两组人群的均衡性检验。两组数据的变量均无统计学差异, 证明均衡可比, 具体见表S2。

表 1 验证集预测模型特征

Table 1 Prediction model characteristics of the test set

Feature ^{a)}	AUC(95%CI)	Accuracy	Specificity	Sensitivity	PPV	NPV
LR	0.732 (0.674, 0.79)	0.853	0.992	0.075	0.636	0.858
KNN	0.673 (0.608, 0.738)	0.847	0.960	0.215	0.488	0.873
DT	0.626 (0.578, 0.674)	0.852	0.996	0.043	0.666	0.854
RF	0.729 (0.666, 0.792)	0.862	0.992	0.129	0.750	0.865
ANN	0.731 (0.671, 0.79)	0.862	0.988	0.151	0.700	0.867
XGBoost	0.758 (0.699, 0.817)	0.859	0.994	0.100	0.750	0.861
LightGBM	0.752 (0.693, 0.812)	0.862	0.992	0.129	0.750	0.865

a) PPV: 阳性预测值(positive predictive value); NPV: 阴性预测值(negative predictive value)

2.3 训练集基本特征

训练集共1437人, 其中237例为发生术后并发症。总训练集人群中, 平均年龄为: 48.21 ± 12.51 岁, 平均BMI为: $21.96\pm2.20\text{ kg/m}^2$ 。训练集中, 发生术后非手术部位并发症的人群和未发生术后非手术部位并发症的人群在BMI、切除肝段数目等方面存在显著差异。其他资料详见表S3。

2.4 利用LASSO回归筛选预测变量

在训练集中进行LASSO回归, 从34个自变量中, 筛选出“已婚”“超过5种并存病”“既往曾行肝切除术”“非肝脏手术的腹部手术史”“BMI”“黄疸”“ALP”“ALB”“毗邻脏器侵犯”“病变位置”共10个可以影响术后非手术部位并发症结局的变量(具体见图S2), 将这些变量用于开发机器学习预测模型。

2.5 预测模型特征

7个预测模型的两个阶段(训练与验证)的最高AUC和置信区间等具体特征见表S4和表1。7个模型的训练集和验证集ROC曲线分别见图S3和图1。在验证阶段, 我们将训练好的模型应用于验证集, XGBoost模型具有最佳的性能, AUC: 0.758; 95%CI: 0.699~0.817; 准确率: 0.859。

2.6 模型校准曲线

绘制模型的校准曲线(图2和图S4), 其中XGBoost校准曲线最佳, 该校准曲线接近 $y=x$ 的对角线, 预测概率与实际概率匹配度高, 表明该模型的预测效果良好, 模型中变量适用于预测。

2.7 模型DCA曲线

绘制7个模型的DCA曲线(图3), 其中XGBoost模型

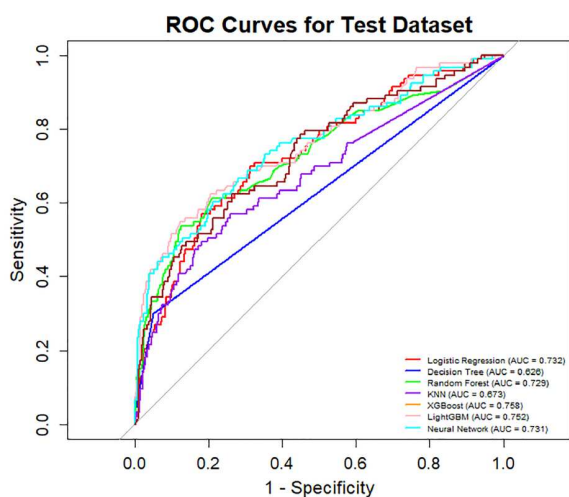


图 1 验证集ROC曲线

Figure 1 ROC curve for the test set

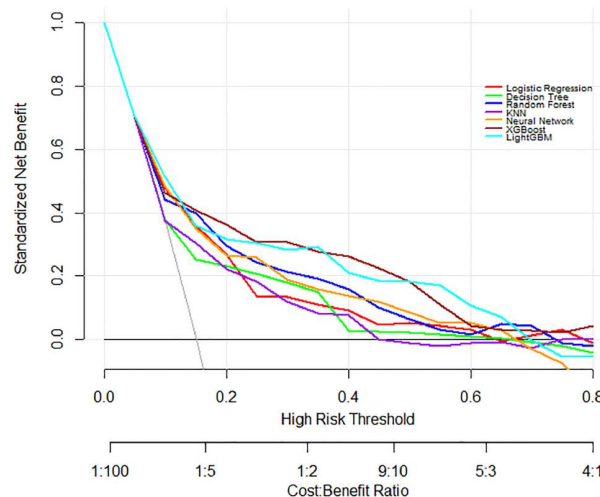


图 3 DCA曲线

Figure 3 DCA curve

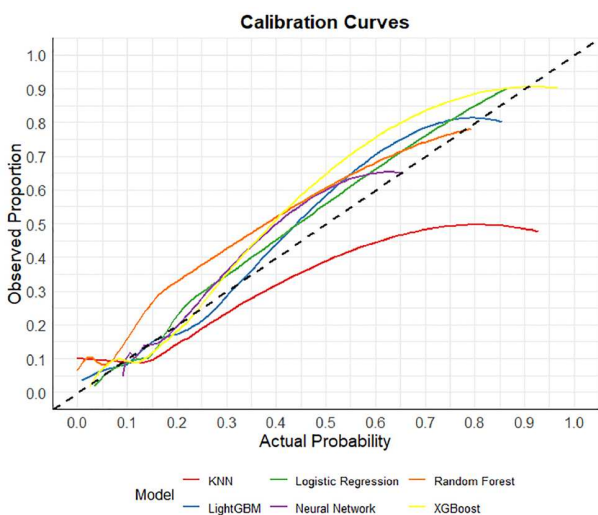


图 2 模型的校准曲线

Figure 2 Calibration curve of the model

在大多数高风险阈值下表现出最高的净收益，尤其是在中高阈值范围内，显示出很好的临床效用。

2.8 可解释性机器学习

SHAP被用于展示指定特征对XGBoost模型中肝切除术后非手术部位并发症的影响。图S5和表2展示了在XGBoost模型中各个特征的平均SHAP值，反映各个特征对模型预测的平均贡献。SHAP值表明BMI是最重要的特征，其次是病变位置、ALP、ALB等。这些特征在

预测肝切除术后非手术部位并发症是否发生时起着关键作用。

图4中，每个点代表一个样本，颜色表示特征值的高低(High或Low)，可以显示影响水平。图中可以看到，BMI、病变位置、ALP值、ALB值等特征的SHAP值分布，这些特征对模型预测有显著影响。这些特征因素中，BMI为连续变量，为自给数据，不进行额外处理。病变位置的赋值按病变分布范围赋值。ALB ($\geq 35\text{g/L}$ 为正常范围， $< 35\text{g/L}$ 为超出正常范围)和ALP ($\leq 130\text{ U/L}$ 为正常范围， $> 130\text{ U/L}$ 为超出正常范围)，数值超出正常范围赋值0，数值在正常范围赋值1。其余非肝脏手术的腹部手术史、既往曾行肝切除术、超过5种并存病、黄疸、已婚、毗邻脏器侵犯标签定义中：No赋值0，Yes赋值1。具体赋值方法见前述数据处理部分。这些特征因素对模型预测的贡献水平依次为BMI (平均SHAP: 0.0532)、病变位置(平均SHAP: 0.0272)、ALP (平均SHAP: 0.0170)、ALB (平均SHAP: 0.0132)、非肝脏手术的腹部手术史(平均SHAP: 0.0125)、超过5种并存病(平均SHAP: 0.0114)、黄疸(平均SHAP: 0.0100)、既往曾行肝切除术(平均SHAP: 0.0030)、已婚(平均SHAP: 0.0028)、毗邻脏器侵犯(平均SHAP: 0.0013)。具体见表2。

图5展示了BMI和SHAP值之间的关系。每个点代表一个样本，颜色表示病变位置。图中显示，随着BMI的增加，SHAP值也呈现上升趋势，意味着BMI对预测并发症的发生有正向影响。根据病变位置赋值规

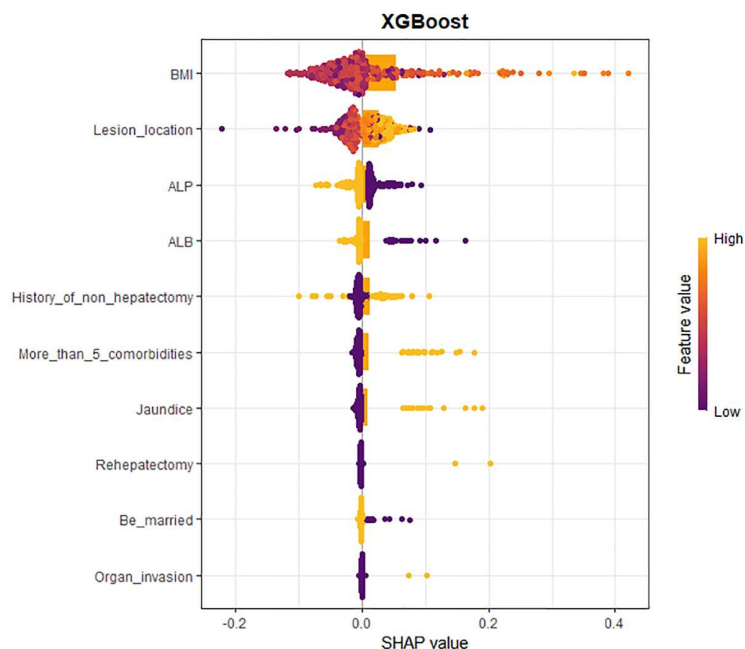


图 4 特征重要性叠加图
Figure 4 Overlay diagram of feature importance

表 2 特征的SHAP值
Table 2 SHAP values of the features

特征	SHAP平均值	SHAP最小值	SHAP最大值
BMI	0.0532	-0.1174	0.4208
Lesion_location	0.0272	-0.2217	0.1084
ALP	0.0170	-0.0723	0.0934
ALB	0.0132	-0.0361	0.1636
History_of_non_hepatectomy	0.0125	-0.0991	0.1067
More than 5 comorbidities	0.0114	-0.0164	0.1769
Jaundice	0.0100	-0.0132	0.1900
Rehepatectomy	0.0030	-0.0052	0.2018
Be married	0.0028	-0.0075	0.0758
Organ invasion	0.0013	-0.0043	0.1015

则，随着数字的增加，病变位置的范围越大。图中显示，随着病变位置的赋值增加，SHAP值也呈现上升趋势，意味着病变位置对预测并发症的发生也具有正向影响。

从图6结合图4中可以看出，患者自身因素中的BMI、病变位置、异常的ALP值、非肝脏手术的腹部手术史、既往曾行肝切除术、超过5种并存病、黄疸、毗邻脏器侵犯对模型预测结果有正向作用，而患者自身因素中的已婚和正常的ALB值对模型预测结果则有

负向作用。根据本研究中二分类数据的赋值规则，特征因素中的8个二分类特征因素(ALP值、ALB值、非肝脏手术的腹部手术史、既往曾行肝切除术、超过5种并存病、黄疸、毗邻脏器侵犯、已婚)的对立面对模型预测结果有相反的作用。

图6展示了模型对一个样本的预测过程，其中每个矩形代表一个特征的贡献。特征如ALP=0和病变位置=15对预测有正面影响，增加了并发症发生的概率。BMI=22.5和其他一些特征如超过5种并存病=0、

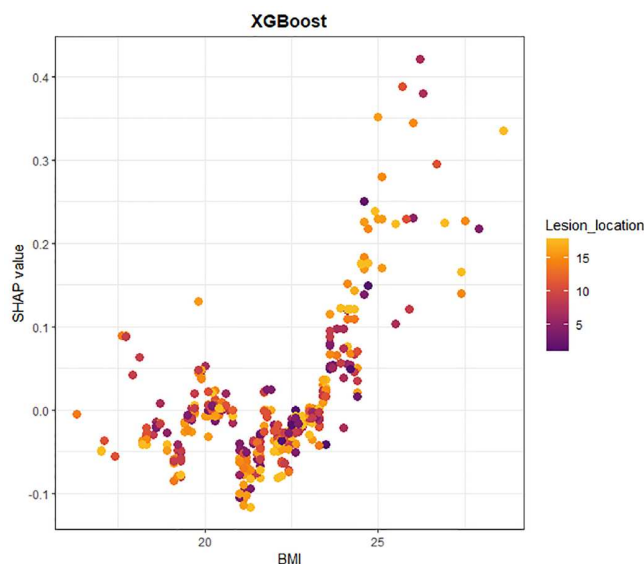


图 5 病变位置和BMI的偏相关依赖图

Figure 5 Partial correlation dependence plot of lesion location and BMI

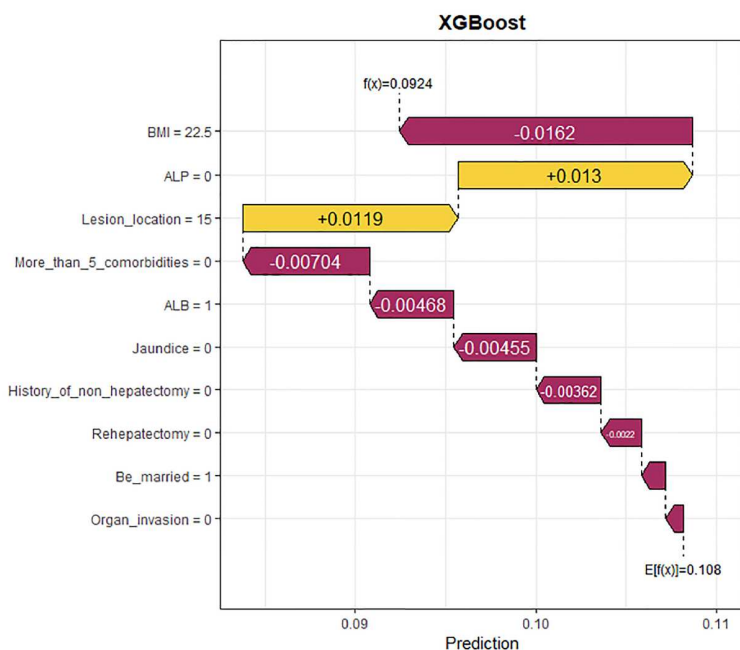


图 6 单样本瀑布图

Figure 6 Single-sample waterfall plot

ALB=1、黄疸=0、非肝脏手术的腹部手术史=0、既往曾行肝切除术=0、已婚=1、毗邻脏器侵犯=0则对预测有负面影响,降低了概率。最终,这些特征的贡献累加得到该样本的预测概率为0.108。

图S6以线性方式展示特征对单个样本的预测概率的边际贡献。病变位置和ALP对提高预测概率的贡献最大。BMI对降低预测概率的影响最显著。其他特征也有一定影响,但相对较小。

3 讨论

本机器学习模型聚焦于肝切除术后非手术部位并发症发生风险的预测。随着手术技术和手术器械的进步,肝脏手术已无禁区,手术部位并发症正逐年减少,非手术部位并发症在术后并发症中的比例逐年升高,如何降低非手术部位并发症是目前亟需解决的问题。我们运用大量数据训练建立的机器学习模型,重点关注个体的术前因素和非手术部位并发症,使模型更加具备实用价值。在我们的研究中,我们通过术前因素应用了7种机器学习方法来预测肝切除术后非手术部位并发症,这些方法在其他机器学习研究中已用于解决多种疾病。其中部分机器学习模型稳健性强,可应用于原始数据。尤其,数据越原始,算法的预测就越精确。此外,我们对机器学习模型的预测潜力进行了多层级评估。使用每个模型的判别验证数据集,采用AUC作为主要评价指标,绘制各个模型的ROC曲线、校准曲线、DCA曲线。由于在实践中正确理解机器学习方法并直观展示识别结果存在困难,我们将SHAP值应用于所选定的模型,以实现最佳的识别效果和可解释性。正的SHAP值意味着与该特征相关的值导致更高的肝切除术后非手术部位并发症发生风险,而负的SHAP值则意味着更低的风险。

在本研究中,我们通过收集的2053例肝切除术患者的术前因素(包括手术预案相关因素)建立了与术后非手术部位并发症相关的可解释机器学习模型,用于预测肝切除术后非手术部位并发症的发生风险。在考虑的7种机器学习模型中,XGBoost模型的性能最好。对于选择XGBoost模型作为最佳的预测模型,原因有:(1)本研究选用AUC作为判定的最佳指标,AUC在该模型中是为最高(XGBoost验证集AUC: 0.758, 95%CI: 0.699~0.817, 准确率: 0.859, XGBoost算法参数的选择细节见表1),对非并发症患者的排除能力非常强,能够有效减少误诊的风险,准确识别并发症群体;(2)校准曲线以及DCA曲线的构建,说明XGBoost在临床中的应用之于其他模型更好(校准曲线表明预测概率与实际观察到的比例具有良好的一致性,XGBoost算法参数的选择细节见图1、图2和图S4; DCA曲线整体表现最佳,XGBoost算法参数的选择细节见图3)。在观察群体样本大的时候,其预测概率更高,说明模型中变量适用于预测。XGBoost在大多数高风险阈值下表现出较高的净收益,尤其是在中高阈值范围内,显示出很好的临床

效用。另外补充了AUC高的几个模型(RF、ANN、LightGBM)的特征重要性条图(图S7~S9),从这些图中可以看出,综合所有模型对预测影响最大的均为BMI这一临床特征变量,其同质性表明该变量对于所预测的结局影响的稳定性良好,证明本研究是可靠的。因此,选用XGBoost恰可以说明模型的依据无人干预,均是基于研究内容所指向的最佳选择,故而根据客观指标选用XGBoost进行预测是可靠的结论。结果表明,XGBoost模型表现最佳,全面地体现了机器学习模型的性能。最终,我们选择了这种模型来预测肝切除术后非手术部位并发症的发生风险。通过SHAP方法阐释了XGBoost模型中每个所选特征的重要性,展示了XGBoost模型的准确性和稳健性。由XGBoost模型得出的SHAP值能够辅助临床医生直观化模型的决策过程,表明XGBoost在预测肝切除术后非手术部位并发症的发生风险方面具有广阔的应用前景。

本研究立足于将机器学习算法应用于肝切除术后非手术部位并发症预测。本研究发现,运用机器学习算法能够提高预测的准确性。机器学习是人工智能的一个分支,它借助数学算法来发现和分类异构数据中的结构,从而做出决策^[21,22]。在考虑机器学习算法时,需清楚是否会得出某些较为复杂的结论^[23,24]。与此同时,机器学习算法的解读难度阻碍了其在辅助医疗决策方面的应用。SHAP法作为预测模型解释工具,提供条形图、蜂群图、散点图、瀑布图等可视化图表,帮助理解特征对预测结果的影响。本研究使用了SHAP工具来解释特征,从而可视化所选模型的决策能力。

本模型中,“已婚”“超过5种并存病”“既往曾行肝切除术”“非肝脏手术的腹部手术史”“BMI”“黄疸”“ALP”“ALB”“毗邻脏器侵犯”“病变位置”共10个可以影响肝切除术后非手术部位并发症结局的术前特征因素。SHAP的结果明确了术前特征因素与肝切除术后非手术部位并发症的关系。患者自身因素中的体重指数(SHAP: 0.0532)、病变位置(SHAP: 0.0272)、异常的碱性磷酸酶值(SHAP: 0.0170)、非肝脏手术的腹部手术史(SHAP: 0.0125)、超过5种并存病(SHAP: 0.0114)、黄疸(SHAP: 0.0100)、肝切除术史(SHAP: 0.0030)、毗邻脏器侵犯(SHAP: 0.0013)对模型预测结果有正向作用,已婚(SHAP: 0.0028)和正常的白蛋白值(SHAP: 0.0132)对模型预测结果有负向作用。另外,需要说明的是,根据本研究中二分类数据的赋值规则,特征因素中的8个二分类特征因素(ALP、ALB、非肝脏手术的腹

部手术史、既往曾行肝切除术、超过5种并存病、黄疸、毗邻脏器侵犯、已婚)的对立面对模型预测结果有相反的作用。

本研究的机器学习模型具有若干显著特征。首先,它采用术前因素来预测肝切除术后非手术部位并发症的发生风险,可以在术前对肝切除术后非手术部位并发症发生风险进行预测,提前发现术后非手术部位并发症发生风险高的患者,并通过机器学习模型找出的高危因素,能够对术后非手术部位并发症预测风险高的患者在术前采取针对性的干预措施,使高危因素转变为低危因素,从而降低术后非手术部位并发症、改善围术期预后。在这些特征因素里,婚史、非肝脏手术的腹部手术史、肝切除术史、病变位置、毗邻脏器侵犯为固定因素,术前无法改变。超过5种并存病为相对固定因素,如果风险过高,可以通过调整治疗,使并存病控制稳定后再手术。BMI、黄疸、ALB值和ALP值为非固定因素,术前可以通过多种途径改变,例如可以采取降低胆红素水平来改善黄疸,控制体重使BMI保持在最佳范围,药物治疗来提高ALB值和降低ALP值。其次,本研究收集的相关变量均是临床中的常用变量,易于收集,所以本预测模型便于推广。

本研究表明,已婚是可以影响肝切除术后非手术部位并发症结局的术前特征因素,其对模型预测结果有负向作用,提示已婚可以减少术后并发症的发生。经查阅文献,一项队列研究发现,与已婚者相比,未婚状态以及未婚状态下的各类细分情况均与总体死亡率和特定病因死亡率呈正相关^[25]。已有研究证实已婚的手术患者(包括实施全膝关节置换术、因膀胱癌接受膀胱切除术、接受手术的头颈部腺样囊性癌的已婚患者等)具有更好的外科术后结局^[26-28]。分析原因,结婚后人们的生活方式会发生变化。已婚者可能获得更多家庭照顾和更佳的精神状态,同时由于群体生活增多导致体力活动增多,家庭协作也会减轻生活压力。简而言之,婚姻会使人们的社会支持度增加,可带来一系列良性的生活方式、心理状态等变化。虽然术前短期内无法改变婚姻状态,但是本研究提示婚姻带来的良好社会支持度使肝切除术具有更佳的结局。

本研究采用术前因素筛选特征因素,术前因素包括患者自身因素和手术预案相关因素^[29],最终筛选出的特征因素仅仅包括患者自身因素(包括病变特征相关因素),表明影响肝切除术后非手术部位并发症的特征因素仅仅与患者自身因素有关,与手术方案无关。该结果提示非手术部位并发症的发生主要受患者自身的综合状况影响,术前对患者进行综合评估是必要的。

未来,对所选术前因素的持续分析和解释追踪将为临床医生预测肝切除术后非手术部位并发症提供有价值的依据,帮助他们得出合理的结论,而非简单地接受算法的预测。我们还可以专注于通过扩展数据库以及增强临床医生-机器学习模型界面的可解释性来进一步验证模型的性能。

本研究亦存在若干局限。其一,由于受到病例数量的限制,未对年龄或任何其他可能在这些亚组分层分析中呈现动态关联的特征进行细分。例如随着人口老龄化,手术患者的范畴也逐渐扩大,且随着年龄的增长,人体的机能逐渐退化,并存病随之增多,对年龄进行分层分析更加有助于增强预测的准确性。其二,虽然本模型通过验证,准确性良好。但由于术后处理也会对并发症的发生产生影响,虽然本研究的各家参与医院在诊疗总体水平上存在一致性,但是仍会在具体术后处理细节上存在一定的区别,由于本研究的目的是术前因素对术后并发症的影响,故而未对术后处理的差异进行规避,任何由此产生的对风险因素的筛查偏差可能在一定程度上影响了机器学习模型对术后并发症的识别。

4 结论

本研究发现,XGBoost模型能够准确地采用术前因素预测肝切除术后非手术部位并发症的发生风险。患者自身因素中的BMI、病变位置、异常的ALP值、非肝脏手术的腹部手术史、肝切除术史、毗邻脏器侵犯、超过5种并存病、黄疸与肝脏术后非手术部位并发症呈正相关,而患者自身因素中的已婚和正常的ALB值与肝脏术后非手术部位并发症呈负相关。手术预案相关因素和术后非手术部位并发症无关。

参考文献

- Huber T, Huettl F, Hanke L I, et al. Leberchirurgie 4.0—OP-planung, volumetrie, navigation und virtuelle realität. *Zentralbl Chir*, 2022, 147: 361–368

- 2 Huang Z Q, Xu L N, Yang T, et al. Hepatic resection: an analysis of the impact of operative and perioperative factors on morbidity and mortality rates in 2008 consecutive hepatectomy cases. *Chin Med J (Engl)*, 2009, 122: 2268–2277
- 3 Gaspari R, Ardito F, Pafundi P C, et al. Development and validation of a comprehensive model to predict complications after hepatectomy. *Eur Rev Med Pharmacol Sci*, 2024, 28: 2509–2521
- 4 Clements N A, Gaskins J T, Martin Ii R C G. Predictive ability of comorbidity indices for surgical morbidity and mortality: a systematic review and meta-analysis. *J Gastrointestinal Surg*, 2023, 27: 1971–1987
- 5 Lee J S, Park D A, Ryoo S, et al. Efficacy and safety of surgical resection in elderly patients with hepatocellular carcinoma: a systematic review and meta-analysis. *Gut Liver*, 2024, 18: 695–708
- 6 Xu L, Wang W, Xu Y. A new risk calculation model for complications of hepatectomy in adults over 75. *Perioper Med*, 2024, 13: 10
- 7 Lu Q, Lu J W, Wu Z, et al. Perioperative outcome of elderly versus younger patients undergoing major hepatic or pancreatic surgery. *Clin Interv Aging*, 2018, 13: 133–141
- 8 Okinaga H, Yasunaga H, Hasegawa K, et al. Short-term outcomes following hepatectomy in elderly patients with hepatocellular carcinoma: an analysis of 10,805 septuagenarians and 2,381 octo- and nonagenarians in Japan. *Liver Cancer*, 2018, 7: 55–64
- 9 Duceppe E, Parlow J, MacDonald P, et al. Canadian cardiovascular society guidelines on perioperative cardiac risk assessment and management for patients who undergo noncardiac surgery. *Can J Cardiol*, 2017, 33: 17–32
- 10 Rajakumar A, Rammohan A, Rela M. Perioperative screening and management in elective complex hepatobiliary surgery. *Curr Opin Crit Care*, 2022, 28: 221–228
- 11 Wolfe J D, Wolfe N K, Rich M W. Perioperative care of the geriatric patient for noncardiac surgery. *Clin Cardiol*, 2020, 43: 127–136
- 12 Tomita K, Koganezawa I, Nakagawa M, et al. A new preoperative risk score for predicting postoperative complications in elderly patients undergoing hepatectomy. *World j surg*, 2021, 45: 1868–1876
- 13 Wang J, Zheng T, Liao Y, et al. Machine learning prediction model for post- hepatectomy liver failure in hepatocellular carcinoma: a multicenter study. *Front Oncol*, 2022, 12: 986867
- 14 Lema G. Risk prediction models. *Mayo Clin Proc*, 2021, 96: 1095
- 15 Bierle D M, Raslau D, Regan D W, et al. Preoperative evaluation before noncardiac surgery. *Mayo Clin Proc*, 2020, 95: 807–822
- 16 Khosravi M, Mojtabaiean S M, Demiray E K D, et al. A systematic review of the outcomes of utilization of artificial intelligence within the healthcare systems of the middle east: a thematic analysis of findings. *Health Sci Rep*, 2024, 7: e70300
- 17 Ennab M, Mcheick H. Enhancing interpretability and accuracy of AI models in healthcare: a comprehensive review on challenges and future directions. *Front Robot AI*, 2024, 11: 1444763
- 18 Ocampo Osorio F, Alzate-Ricaurte S, Mejia Vallecilla T E, et al. The anesthesiologist's guide to critically assessing machine learning research: a narrative review. *BMC Anesthesiol*, 2024, 24: 452
- 19 Chen M, Decary M. Artificial intelligence in healthcare: an essential guide for health leaders. *Health Manage Forum*, 2020, 33: 10–18
- 20 Wang L, Alexander C A. Big data analytics in medical engineering and healthcare: methods, advances and challenges. *J Med Eng Tech*, 2020, 44: 267–283
- 21 Hirani R, Noruzi K, Khuram H, et al. Artificial intelligence and healthcare: a journey through history, present innovations, and future possibilities. *Life*, 2024, 14: 557
- 22 Dodig S, Čepelak I, Dodig M. Are we ready to integrate advanced artificial intelligence models in clinical laboratory? *Biochem Med (Zagreb)*, 2025, 35: 010501
- 23 Stam W T, Goedknecht L K, Ingwersen E W, et al. The prediction of surgical complications using artificial intelligence in patients undergoing major abdominal surgery: a systematic review. *Surgery*, 2022, 171: 1014–1021
- 24 Kenig N, Monton Echeverria J, Muntaner Vives A. Artificial intelligence in surgery: a systematic review of use and validation. *J Clin Med*, 2024, 13: 7108
- 25 Leung C Y, Huang H L, Abe S K, et al. Association of marital status with total and cause-specific mortality in Asia. *JAMA Netw Open*, 2022, 5: e2214181
- 26 Rahman T M, Hennekes M, Mehaidli A, et al. Marital status, race, insurance type, and socioeconomic status—Assessment of social predictors for outcomes after total knee arthroplasty. *J Am Acad Orthop Surg*, 2024, 32: 169–177
- 27 Pruthi R S, Lentz A C, Sand M, et al. Impact of marital status in patients undergoing radical cystectomy for bladder cancer. *World J Urol*, 2009, 27: 573–576
- 28 Huang K, Przeslawski C, Ramirez C A. What risk factors are associated with poorer quality of life in patients with head and neck cancer? *J Oral Maxillofac Surg*, 2023, 81: 648–653
- 29 Gairola S, Solanki S L, Patkar S, et al. Artificial intelligence in perioperative planning and management of liver resection. *Ind J Surg Oncol*, 2024, 15: 186–195

补充材料

- 表S1 总人群基本特征
- 表S2 均衡性检验
- 表S3 训练集基本特征
- 表S4 训练集预测模型特征
- 图S1 流程图
- 图S2 LASSO筛选变量图
- 图S3 训练集ROC曲线
- 图S4 XGBoost校准曲线
- 图S5 XGBoost特征重要性条图
- 图S6 单样本特征图
- 图S7 RF特征重要性条图
- 图S8 ANN特征重要性条图
- 图S9 LightGBM特征重要性条图

本文以上补充材料见网络版csb.scichina.com. 补充材料为作者提供的原始数据, 作者对其学术质量和内容负责.

Summary for “肝脏术后非手术部位并发症风险预测机器学习模型开发”

Development of a risk prediction model for non-surgical-site complication after hepatectomy based on interpretable machine learning

Lining Xu^{1*}, Weiyu Wang² & Bo Yang³

¹ Department of General Surgery, The Second Medical Center & National Clinical Research Center for Geriatric Diseases, Chinese PLA General Hospital, Beijing 100853, China

² Institute of Hepatobiliary Diseases of Wuhan University, Zhongnan Hospital of Wuhan University, Wuhan 430071, China

³ Department of Radiology, Affiliated Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430022, China

* Corresponding author, E-mail: xu_lining@aliyun.com

Postoperative complications are divided into surgical-site complication (SSC) and non-surgical-site complication (NSSC). Previous studies have mostly focused on SSC, and no study has yet used preoperative factors to comprehensively assess the risk of NSSC after hepatectomy. Hepatectomies are associated with a high incidence of postoperative complications. With the innovation of hepatectomy technology, the incidence of SSC is gradually decreasing, and the severity of SSC is also gradually decreasing. However, owing to age, comorbidities, and other factors, NSSC remain a serious problem. Evaluating the risk of NSSC before surgery and providing early warning and intervention to reduce the occurrence of postoperative NSSC is a clinical problem. Existing surgical risk assessment systems focus on SSC and other single-organ complications, and there is no systematic model that can integrate preoperative factors and accurately predict the overall postoperative NSSC. This study aims to develop an efficient and interpretable machine learning model to correlate preoperative factors with the risk of NSSC after hepatectomy. The dataset for the retrospective investigation and machine learning of the association between preoperative factors and NSSC after hepatectomy was obtained from 3 national or regional central hospitals in China (2017–2023). A total of 2053 eligible participants were included in this study. Seven machine learning models were developed to predict the risk of NSSC after hepatectomy by assessing preoperative factors. With the area under the curve (AUC) as the main evaluation index, receiver operating characteristic (ROC), calibration, and decision curve analysis (DCA) curves for each model were drawn to select the best model. Finally, the shapley additive explanations (SHAP) tool was used to explain features to visualize the decision-making abilities of the selected models. The extreme gradient boosting (XGBoost) model showing the best performance based on 10 preoperative factors (all being factors of patients' own characteristics) was selected to predict the risk of NSSC after hepatectomy (AUC of test set: 0.758, 95% CI: 0.699–0.817, accuracy: 0.859). The calibration curve indicated that the predicted probability had good consistency with the actual probability. Overall, the DCA curve performed the best. Patient characteristics such as body mass index (BMI)(average SHAP: 0.0532), lesion location (average SHAP: 0.0272), abnormal alkaline phosphatase (ALP) levels (average SHAP: 0.0170), history of non-hepatic abdominal surgery (average SHAP: 0.0125), more than five comorbidities (average SHAP: 0.0114), jaundice (average SHAP: 0.0100), history of hepatectomy (average SHAP: 0.0030), and adjacent organ invasion (average SHAP: 0.0013) were positively correlated with model prediction, whereas married status (average SHAP: 0.0028) and normal albumin (ALB) levels (average SHAP: 0.0132) had a negative effect on the prediction result of the model. The XGBoost model accurately predicted the association between preoperative factors and NSSC after hepatectomy. BMI, lesion location, abnormal ALP levels, history of non-hepatic abdominal surgery, more than five comorbidities, jaundice, re-hepatectomy, and adjacent organ invasion were positively correlated with postoperative NSSC. Married status and normal ALB levels were negatively correlated with postoperative NSSC. In this study, for the first time, we explicitly present the concept of complications at non-surgical-site, and this is the first machine learning-based model developed to predict postoperative NSSC in patients undergoing hepatectomy using preoperative factors. We validated its performance in a test set and selected XGBoost as the best model. Ultimately, we developed an XGBoost model that can predict the risk of NSSC in patients undergoing hepatectomy.

hepatectomy, non-surgical-site complication, prediction model, machine learning

doi: [10.1360/CSB-2025-0439](https://doi.org/10.1360/CSB-2025-0439)