DOI:10.16016/j.2097-0927.202312018

基于机器学习算法建立非肝脏手术术后肝损伤风险预测模型

孙义竹,李雨捷,梁 浩,刘 祥,黄家号,舒 欣,宋艾璘,杨智勇,易 斌 400038 重庆,陆军军医大学(第三军医大学)第一附属医院麻醉科

[摘要] **目的** 基于术前及术中用药指标建立一个预测非肝脏手术患者术后肝损伤的机器学习预测模型。**方法** 采用病例-对照研究设计方案,选择国内3家大型综合医院开发的数据库中2014年1月至2022年9月行非肝脏手术术后出现肝损伤患者315例,按阳性∶阴性为1∶3的比例随机匹配同期非肝脏手术术后未发生肝损伤的928例病例作为阴性对照病例。将这1 243例患者根据R语言设置的代码以7∶3的比例随机拆分为建模组(n=869)和验证组(n=374)。利用患者术前临床指标(基本信息、病史、相关评分表、手术相关信息和实验室检查)及术中用药情况基于k-近邻(k-nearest neighbor,KNN)、支持向量机(support vector machine linear,SVM)、逻辑回归(logistic regression,LR)、极端梯度提升(extreme gradient boosting,XGBoost)4种机器学习算法建立预测非肝脏手术术后肝损伤的模型,并在验证组通过受试者工作特征曲线(receiver operating characteristic, ROC)、精准-召回曲线(percision-recall curve,P-R)、临床决策分析曲线(decision curve analysis, DCA)、Kappa值、灵敏度、特异度、Brier分数、F_1分数对模型效能进行评估。**结果** 4种机器学习算法建立的预测非肝脏手术术后肝损伤的模型以XGBoost算法最优,验证组中受试者工作特征曲线下面积(area under the receiver operating characteristic curve,AUROC)为0.916(95%*CI*:0.883~0.949),精准-召回曲线下面积(area under the precision-recall curve,AUPRC)为0.841,Brier评分0.097,灵敏度和特异度分别为78.95%和87.10%。**结论** 基于XGBoost算法建立的非肝脏手术术后肝损伤预测模型对术后肝损伤的发生有良好的预测能力。

[关键词] 机器学习;预测模型;术后肝损伤
[中图法分类号] R181.32; R319; R575 [文献标志码] A

Establishment of risk prediction model for postoperative liver injury after non-liver surgery based on different machine learning algorithms

SUN Yizhu, LI Yujie, LIANG Hao, LIU Xiang, HUANG Jiahao, SHU Xin, SONG Ailin, YANG Zhiyong, YI Bin
Department of Anesthesiology, First Affiliated Hospital, Army Medical University (Third Military Medical University), Chongqing, 400038, China

[**Abstract**] **Objective** To construct a machine learning prediction model for postoperative liver injury in patients with non-liver surgery based on preoperative and intraoperative medication indicators. **Methods** A case-control study was conducted on 315 patients with liver injury after non-liver surgery selected from the databases developed by 3 large general hospitals from January 2014 to September 2022. With the positive/negative ratio of 1∶3, 928 cases in corresponding period with non-liver surgery and without liver injury were randomly matched as negative control cases. These 1 243 patients were randomly divided into the modeling group (n=869) and the validation group (n=374) in a ratio of 7∶3 using the R language setting code. Preoperative clinical indicators (basic information, medical history, relevant scale score, surgical information and results of laboratory tests) and intraoperative medication were used to construct the prediction model for liver injury after non-liver surgery based on 4 machine learning algorithms, k-nearest neighbor (KNN), support vector machine linear (SVM), logic regression (LR) and extreme gradient boosting (XGBoost). In

[基金项目] 国家重点研发计划(2018YFC0116702)
[通信作者] 易 斌,E-mail:yibin1974@163.com

2097-0927/Copyright © 2024, 陆军军医大学(第三军医大学)。这篇开放获取文章遵循 CC BY 许可协议(https://creativecommons.org/licenses/by/4.0/)

the validation group, receiver operating characteristic (ROC) curve, precision-recall curve (P-R), decision curve analysis (DCA) curve, Kappa value, sensitivity, specificity, Brier score, and F_1 score were applied to evaluate the efficacy of model. **Results** The model established by 4 machine learning algorithms to predict postoperative liver injury after non-liver surgery was optimal using the XGBoost algorithm. The area under the receiver operating characteristic curve(AUROC) was 0.916 (95%*CI*: 0.883~0.949), area under the precision-recall curve (AUPRC) was 0.841, Brier score was 0.097, and sensitivity and specificity was 78.95% and 87.10%, respectively. **Conclusion** The postoperative liver injury prediction model for non-liver surgery based on the XGBoost algorithm has effective prediction for the occurrence of postoperative liver injury.

[**Key words**] machine learning;predicting model;postoperative liver injury

Supported by the National Key Research and Development Plan(2018YFC0116702). Corresponding author: YI Bin,E-mail:yibin1974@163.com

术后肝损伤是手术患者在围术期可能发生的严重并发症之一[1],研究显示术后肝损伤发病率因手术类型而异,约为3.2%~72.22%[2-3]。围术期肝功能障碍病人易出现肾功能障碍,产生肝肾综合征、肾前性氮质血症或急性肾小管坏死,一旦肝功能持续恶化,常会出现黄疸进行性加重、严重凝血机制障碍、腹水、肝性脑病、肾功能不全加重等肝功能衰竭的表现[1],而急性肝衰竭虽发生率为1%~3%,其病死率高达60%~90%[4],会增加住院时间和病死率,影响术后长期存活率,因此围术期肝脏功能的监测和保护尤其重要。因受人为因素的干扰,使用传统评分工具建立的预测模型的临床实用性及评估效率降低,而随着近年来机器学习被广泛应用于疾病预测[5],医护人员利用机器学习方法在肝病学方面开展了大量研究。WANG 等[6]纳入875例多中心患者数据进行回顾性队列研究,采用LightGBM算法对术后肝衰竭进行预测,在验证集得到较高模型效能(AUC:0.82),其相对于传统评分工具展现了更大的优势。然而,目前很少有研究使用机器学习来预测非肝脏手术患者术后并发症的风险。本研究将机器学习运用于术后肝损伤预测,进行早期风险评估,为术中麻醉管理和术后护理路径提供预警,以期提高围术期管理水平,降低各类手术术后肝损伤发生风险。

1 资料与方法

1.1 一般材料

1.1.1 数据来源 数据来源于国内3家大型综合医院开发的数据库中2014年1月至2022年9月行非肝脏手术的患者,并通过3家医院伦理委员会批准:陆军军医大学第一附属医院(2014年1月至2022年9月,KY201936),四川大学附属华西医院(2019年5月至2020年1月,2021-349),中山大学附属第一医院(2019年6月至2019年12月,2019-385)。

1.1.2 研究对象 阳性患者纳入标准:①年龄≥18岁;②接受除肝胆手术外任何类型的手术;③术后14 d内天门冬氨基酸转移酶(aspartate aminotransferase,AST)>200 U/L,丙氨酸氨基转移酶(alanine aminotransferase,ALT)>200 U/L。排除标准:①年龄>75岁;②术前最近1次检验AST>200 U/L,ALT>200 U/L;③术前诊断为肝衰竭、肝损伤、肝功能不全者;④重要病例资料不全者。按照纳入、排除标准,共筛选出315例阳性患者。

阴性患者纳入排除标准:采用病例对照研究设计方案,纳入、排除标准同阳性病例,按统计学要求,按阳性:阴性≈1:3比例随机匹配同期术后未发生肝损伤的928例患者纳入非肝损伤组。

本研究将术后肝损伤定义为:在本次住院期间手术后14 d内AST、ALT均大于5倍正常值(200 U/L)[7]。患者收集流程图见图1。

1.2 方法

1.2.1 建模选择指标 研究主要收集术前及术中临床指标,主要包括基础信息、病史、相关评分表、手术相关信息和实验室检查(术前最近1次结果)。基础信息包括年龄、性别、体质量指数(body mass idex,BMI);病史包括吸烟史、饮酒史、是否患有乙肝、合并症(高血压,糖尿病,心脏病,肺部疾病)、Child-Pugh评分;手术相关信息包括手术类型(上腹部手术、下腹部手术,心胸手术、脊柱手术、四肢手术、头面颈部手术、其他手术)、麻醉分级、急诊、手术时间、麻醉方式;实验室检查包括血红蛋白(Hb)、血小板计数/中性粒细胞计数(PLT/neu)、白细胞计数(WBC)、血小板计数(PLT)、凝血酶原时间(PT)、活化部分凝血活酶时间(APTT)、肌酐(Crea)、尿素(Urea)、钠(Na)、钙(Ca)、白蛋白(ALB)、白球比、AST、ALT、胆红素(BIL)、直接胆红素(DBIL)、间接胆红素(IBIL)及术中用药指标,包括红细胞使用量、血浆使用量、右美托嘧啶使用量、七氟烷使用量、丙泊酚使用量、舒芬太尼使用量、地佐辛使用量、利尿剂使用量。

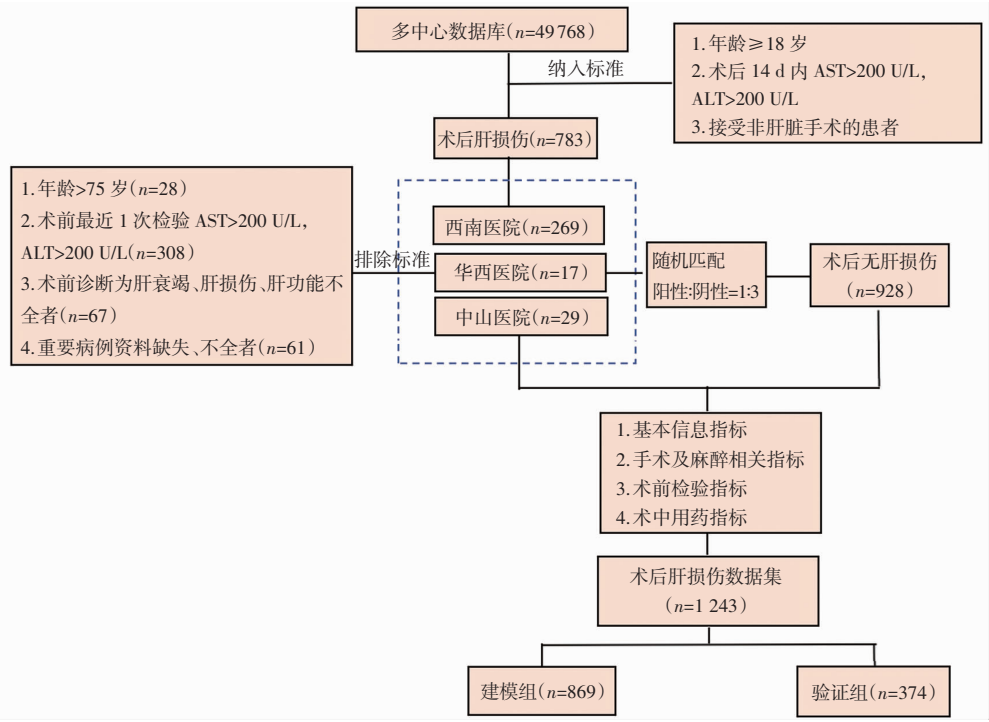


图 1 多中心数据库患者收集数量及模型集合划分

1.2.2 数据预处理 首先处理缺失数据,去掉纳入指标缺失超过 30%的患者数据,去掉单一指标缺失超过 30%的指标,剩余缺失数据进行 k-近邻(k-nearest neighbor,KNN)填充,完成数据收集和填充后对所有指标行数据预处理,以便于提高机器学习计算效率。对于连续变量剔除近零方差变量,归一化标准化数据,处理后数据的均值为零,标准差为 1。原始数据经过预处理后,各指标处于同一数量级,更加适合进行综合对比评价[8],针对分类变量(如性别),将其转化为因子。将纳入统计的数据,根据 R 语言设置的代码以 7∶3的比例被随机拆分为建模组(n=869)和验证组(n=374)。由于大部分机器学习算法在不平衡数据集上表现不佳,为了数据的平衡,在建模组中采用 Smote 采样将其修正为平衡数据,使用 LASSO 方法进行特征选择并且进行十折交叉验证,筛选和确定独立且有效的危险因素。

1.2.3 预测模型构建与评估 机器学习过程在 R 软件上进行。在 R 软件上运行 tidymodels 包、themis 包,选择 KNN、支持向量机(support vector machine linear,SVM)、逻辑回归(logistic regression,LR)、极端梯度提升(extreme gradient boosting,XGBoost)这 4 种机器学习建立模型。在建模组中,输入 LASSO 回归分析选择出来的预测指标,分别运行相应算法,进行监督学习,通过贝叶斯超化确定最优参数。在验证组中构建受试者工作特征曲线(receiver operating characteristic,ROC)和精准-召回曲线(percision-recall curve,P-R),通过曲线下面积(area under curve,AUC)来评价模型的区分度,判断模型鉴别能力,其中 0.5≤AUC<0.7 提示预测效果较差,0.7≤AUC<0.9 提示预测效果中等,AUC≥0.9 提示预测效果非常好;通过构建校准曲线来评价模型的准确性;通过构建临床决策曲线来评价模型的临床效用。为了评估模型综合性能,模型的 Kappa 值、灵敏度、特异度、Brier 评分、F_1 评分也被分析计算出来。预测模型预处理、开发和评估流程见图 2。

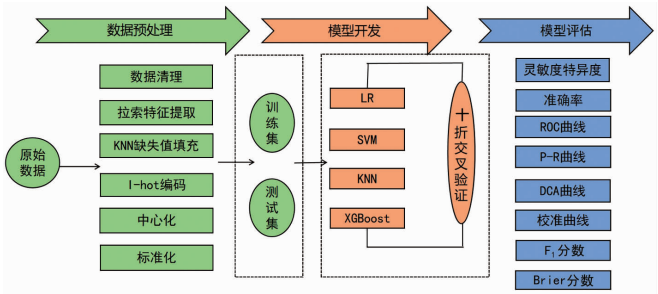


图 2 机器学习数据预处理、模型开发和模型评估流程图

1.3 统计学分析

采用 SPSS 26.0 软件进行统计分析。根据数据分布,正态分布的计量资料以 $\bar{x}\pm s$ 来表示,组间比较采用两独立样本 t 检验,而偏态分布的计量资料以 $M(P_{25},P_{75})$ 表示,比较采用 Mann-Whitney U 检验。对于计数资料,组间比较采用 χ^2 检验或 *Fisher* 确切概率。本研究统计检验均是双侧的,P<0.05 表示差异具有统计学意义。

2 结果

2.1 组间患者一般资料比较

最终纳入数据分析的患者有 1 243 例(肝损伤组 315 例,非肝损伤组 928 例),肝损伤组与非肝损伤组组间基本资料比较发现,手术部位、急诊、麻醉方式、年龄、麻醉分级、Child-Pugh 评分、性别、合并症(高血压,心脏病)差异有统计学意义($P<0.05$,表 1)。对于术前检验指标及术中指标 2 组间比较发现 Hb、WBC、PLT、PT、K、Ca、ALB、A/G、AST、ALT、BIL、IBIL、红细胞使用量、血浆使用量、七氟烷使用量、丙泊酚使用量、舒芬太尼使用量、地佐辛使用量、利尿剂使用量差异有统计学意义($P<0.05$,表 2)。

2.2 特征变量筛选结果

在建模组中,采用 LASSO 回归分析方法对术后肝损伤相关术前指标进行筛选,选择 1 个具备优良性能但是自变量个数最少的模型。最终以 λ = 1se 时筛选出的指标为建模指标,结果如表 3 所示(λ = 1se、λ = min 均为 0 的指标未被归纳在表格中)。

表 1 肝损伤组与非肝损伤组的基本资料比较[$M(P_{25},P_{75})$,例(%)]

指标	肝损伤组(n=315)	非肝损伤组(n=928)	P
手术部位			<0.001
上腹部	32(10.15)	65(7.00)	
下腹部	80(25.39)	349(37.61)	
心胸	79(25.08)	106(11.42)	
头面颈	59(18.73)	178(19.18)	
脊柱	20(6.34)	89(9.59)	
四肢	39(12.38)	125(13.47)	
其他	6(1.9)	16(1.73)	
急诊			<0.001
是	52(16.51)	57(6.14)	
否	263(83.49)	871(93.86)	
麻醉方式			0.005
全麻	292(92.69)	806(86.85)	
非全麻	23(7.31)	122(13.15)	
年龄/岁	53(46,64)	53(45,59)	0.035
BMI/(kg/m^2)	23.41(21.25,25.35)	23.67(21.51,25.71)	0.235
麻醉分级			<0.001
Ⅰ	2(0.63)	44(4.74)	
Ⅱ	210(66.68)	832(89.65)	
Ⅲ	73(23.17)	45(4.85)	
Ⅳ	30(9.52)	7(0.76)	
Child-Pugh 评分	2(2.00,3.00)	2(2.00,2.00)	<0.001
吸烟	108(34.29)	307(33.08)	0.695
饮酒	98(31.11)	268(28.88)	0.453
乙肝	25(7.93)	88(9.48)	0.418
女性	125(39.68)	457(49.24)	0.003
合并症			
高血压	50(15.87)	89(9.59)	0.002
糖尿病	21(6.67)	71(7.67)	0.564
心脏病	50(15.87)	55(5.92)	<0.001
肺部疾病	4(1.27)	11(1.18)	1

表 2 肝损伤组与非肝损伤组的实验室资料比较[$M(P_{25},P_{75})$,例(%)]

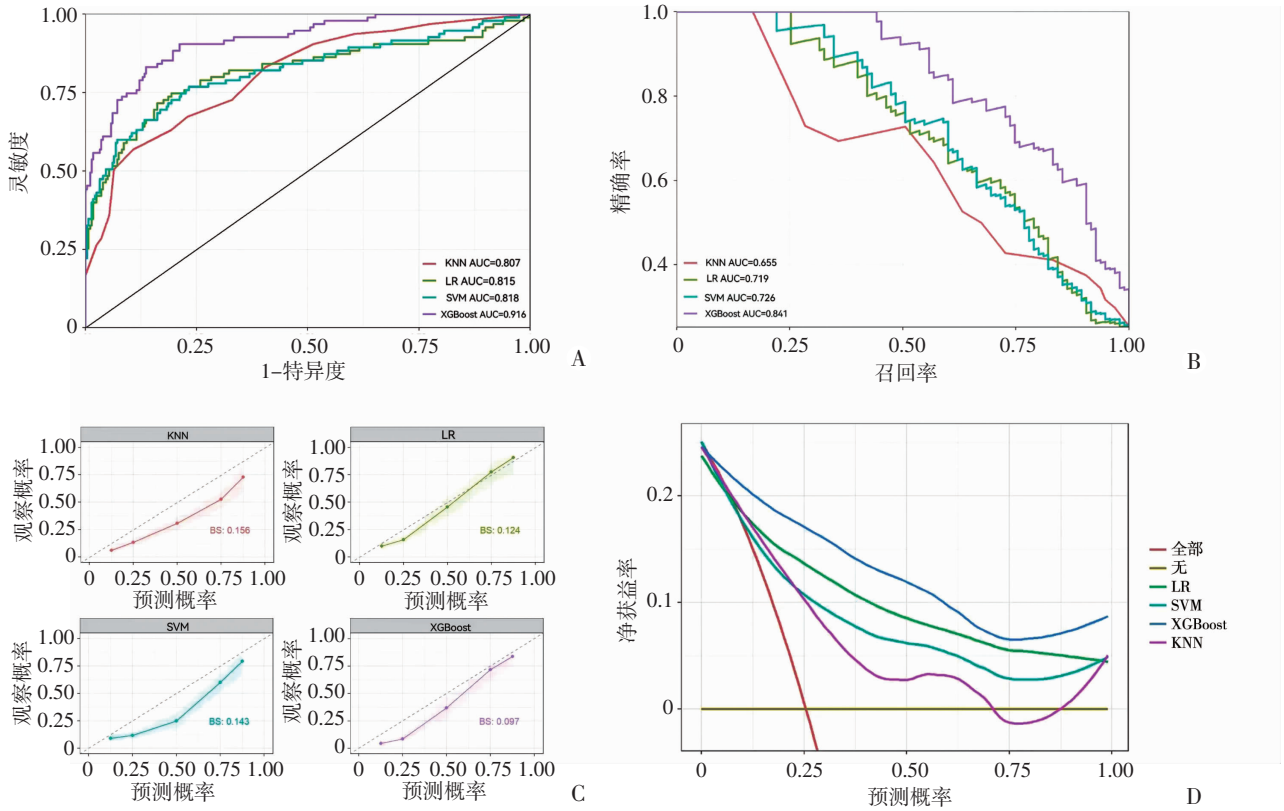
指标	肝损伤组(n=315)	非肝损伤组(n=928)	P
Hb/(g/L)	130.00(116.00,143.00)	133.00(120.00,144.00)	0.028
WBC/($\times10^9$/L)	7.22(5.28,9.03)	6.27(5.20,7.62)	<0.001
PLT/neu	44.69(23.87,63.59)	56.02(40.52,71.59)	<0.001
PLT/($\times10^{12}$/L)	189.00(150.00,245.00)	210.50(170.00,249.75)	<0.001
PT/s	11.00(10.40,11.50)	10.80(10.20,11.30)	<0.001
APTT/s	27.70(25.70,29.40)	27.45(25.80,28.90)	0.463
Crea /(μmol/L)	67.70(57.90,77.80)	66.10(57.20,76.70)	0.290
Urea/(mmol/L)	5.50(4.40,6.62)	5.30(4.30,6.30)	0.134
K^+/(mmol/L)	3.93(3.75,4.15)	4.00(3.78,4.21)	0.004
Na^+/(mmol/L)	139.97(138.30,141.80)	140.19(138.80,141.70)	0.285
Ca^+/(mmol/L)	2.25(2.25,2.33)	2.31(2.24,2.39)	<0.001
ALB/(g/L)	39.70(36.00,43.40)	41.10(38.20,44.10)	<0.001
A/G	1.40(1.21,1.58)	1.48(1.31,1.65)	<0.001
AST/(U/L)	26.20(20.10,36.70)	19.95(17.00,23.80)	<0.001
ALT/(U/L)	22.60(15.6,34.0)	17.00(12.30,23.70)	<0.001
BIL/(μmol/L)	13.20(9.60,17.20)	11.54(8.80,14.50)	<0.001
DBIL/(μmol/L)	2.50(1.30,3.38)	2.30(1.80,3.10)	0.304
IBIL/(μmol/L)	10.69(7.60,13.90)	11.60(8.80,15.00)	0.001
术中指标			
手术时间/min	206.00(130.85,296.00)	158.5(93.25,229.75)	<0.001
血浆/mL	0.00(0.00,0.00)	0.00(0.00,0.00)	0.002
红细胞/mL	0.00(0.00,0.00)	0.00(0.00,0.00)	0.011
右美托嘧啶/μg	50.00(20.00,200.00)	60.00(0.00,120.00)	0.135
七氟烷/mL	10.00(0.00,30.00)	0.00(0.00,20.00)	<0.001
丙泊酚/mg	1 000.00(500.00,1500.00)	900.00(300.00,1200.00)	0.004
舒芬太尼/μg	50.00(50.00,100.00)	50.00(40.00,50.00)	<0.001
地佐辛/mg	0.00(0.00,10.00)	5.00(0.00,10.00)	0.001
利尿剂/mg	0.00(0.00,0.00)	0.00(0.00,0.00)	<0.001

表 3 LASSO 特征选择指标结果

指标	λ=1se	λ=min	指标	λ=1se	λ=min
手术时间	0.176 696 653	0.263 577 507	糖尿病	0.000	7.12E-04
七氟烷使用量	0.019 493 42	0.112 599 294	PLT/neu	0.000	0.079 201 455
丙泊酚使用量	0.128 380 177	0.275 007 82	WBC	0.000	0.017 873 356
地佐辛使用量	−0.074 005 547	−0.257 476 97	Crea	0.000	0.165 643 849
急诊	0.000	−0.098 048 028	K	−0.015 824 943	−0.159 965 583
麻醉分级	0.519 049 477	0.759 328 022	Na	0.000	−0.113 613 818
年龄	0.000	0.058 555 839	Ca	−0.120 546 355	−0.242 606 667
BMI	0.000	−0.163 418 616	ALB	0.000	−0.013 845 208
Child 评分	0.074 847 974	0.098 753 818	AST	0.717 852 922	1.000 296 079
麻醉方式	0.000	0.074 515 368	ALT	0.150 488 978	0.324 876 549
吸烟史	0.000	−0.153 421 658	BIL	0.142 650 682	0.500 096 103
高血压	0.000	0.035 435 646	IBIL	−0.204 759 453	−0.670 462 206

2.3 预测模型的建立与验证

在 R 软件上,输入 LASSO 方法选出的特征指标,基于 LR、SVM、KNN 、XGBoost 4 种不同算法构建肝损伤预测模型,对其模型效能进行评估。基于 XGBoost 算法开发的模型在验证组获得了最佳的鉴别能力(AUROC:0.916,AUPRC:0.841)、最佳的校准能力(Brier 分数为 0.097)和准确率(85.03%)、最高的灵敏度(78.95%)和相对较好的特异度(87.10%);在临床决策曲线中可看出使用 XGBoost 算法开发的模型能够获得最大的临床效益(图 3),模型综合性能比较见表 4。



A:各模型在验证组的 ROC 曲线图;B:各模型在验证组的 P-R 曲线图;C:各模型在验证组的校准曲线图;D:各模型在验证组的临床决策曲线图

图 3 机器学习各预测模型性能曲线对比

表 4 机器学习各预测模型综合性能比较

算法	组别	准确率/%	Kappa	灵敏度/%	特异度/%	Brier 分数	F_1 分数	AUPRC	AUROC(95%*CI*)
KNN	建模组	81.93	0.59	88.64	79.66	0.112	0.713	0.834	0.925 (0.906~0.943)
	验证组	74.60	0.39	67.37	77.06	0.156	0.574	0.655	0.807 (0.756~0.859)
SVM	建模组	91.02	0.76	81.36	94.30	0.092	0.821	0.879	0.925 (0.899~0.949)
	验证组	82.62	0.53	62.11	89.61	0.143	0.645	0.726	0.818 (0.762~0.875)
XGBoost	建模组	97.70	0.94	99.55	97.07	0.025	0.956	0.991	0.997 (0.995~0.999)
	验证组	85.03	0.63	78.95	87.10	0.097	0.728	0.841	0.916 (0.883~0.949)
LR	建模组	82.05	0.56	75.91	84.13	0.114	0.682	0.755	0.845 (0.811~0.879)
	验证组	79.95	0.516	72.63	82.44	0.124	0.648	0.719	0.815 (0.756~0.873)

3 讨论

3.1 肝损伤的危险因素

术后肝损伤(肝功能障碍)是一种严重的并发症,但目前可供选择的治疗措施仍相对有限[9],一旦发生易导致早期病死率和发病率增加[10]。在一项心脏手术术后肝损伤的研究中,肝损伤患者发生严重不良结局的比例远高于同期无肝损伤的患者(导致了更高的气管切开率、ECMO 使用率、CRRT 治疗和更高的住院死亡率)[11]。术后肝损伤的发生与术中出血、低血压、肝脏灌注不足、体外循环(cardiopulmonary bypas, CPB)引起的内脏血流减少、栓子以及全身炎症反应,围术期的肝毒性药物和血制品输注等多因素有关[12]。当 CPB 时间>180 min 时,会发生肝脏代谢功能异常和肝细胞微结构的破坏[10],较长的 CPB 时间会导致更多的溶血,导致游离血红蛋白的增加,加速游离氰化物的立即释放,以及内源性物质的产生[13]。与此同时,在手术麻醉过程中使用的挥发性麻醉剂(如七氟烷),是公认的药物性肝损伤的原因[14],可引起药物性肝损伤,通常导致术后 2~14 d ALT 升高,并可表现为一系列疾病,如从无症状的肝脏生物化学紊乱[15]到急性严重肝炎。手术时间也是短期不良结局的独立预测因子,此外,反映肝脏代谢的血清指标,如 AST、ALT 和 GGT,也常常被用于预测术后肝癌患者的生存率[16]。这些研究结果与本研究中经过 LASSO 回归筛选出的术后肝损伤的危险因素基本相符。尽管由于患者和手术的危险因素。术后并发症是不可避免的[17],但通过早期识别患者的危险因素和使用循证治疗方法(如在术后护理环境中及时使用抗生素[18]),并发症是可以改变和可预防的[19]。

3.2 机器学习对肝损伤的预测作用

本研究在选取阴性患者时以 1∶3的比例进行选取,既可以避免 1∶1时可能会导致的检验效能不佳的问题,同时也可以减少大于 1∶4的数据收集时可能增加的统计分析难度。尽管如此,阳性与阴性病例数仍相差较大,易导致训练集灵敏度较低,直接使用不平衡数据集运用 KNN、SVM、XGBoost、LR4 种算法构建模型进行分析,结果是不理想的,所以本研究采用了 Smote 算法来改善训练数据集中的数据不平衡以达到更好的结果。Smote 算法就是对少数类别样本进行分析和模拟,并将人工模拟的新样本添加到数据集中,进而使原始数据中的类别不再严重失衡。

机器学习作为人工智能的重要分支,近年来已被广泛应用于医疗健康领域,在挖掘、处理医疗数据时具有极大潜力,不仅能够弥补线性模型的不足[20],并且可以为高危患者提供个性化的、智能的诊断和治疗方案。机器学习在肝病学中的应用主要包括:预测病毒性肝炎或非酒精性脂肪性肝病患者的纤维化;确定肝硬化患者是否存在食管静脉曲张;确定终末期肝病患者的预后[21],如 WU 等[22]采用多种机器学习方法(随机森林、朴素贝叶斯、人工神经网络和 Logistic 回归)对脂肪肝高危患者进行分类。经过十倍交叉验证的 RF、NB、ANN 和 LR 的受试者工作特征下面积分别为 0.925、0.888、0.895 和 0.854;YANG 等[23]应用 7 种 ML 方法预测肝移植后的急性肾损伤(acute kidney injury, AKI),梯度增压机模型的性能最佳,AUROC 为 0.90。这些结果表明,机器学习方法在预测特定的肝病方面具有较大的潜力,但这些研究多为单中心研究,研究人群主要针对单一移植或特定疾病的患者,对于手术患者及不同手术类型患者运用可能效果欠佳。本研究所采用的 XGBoost 是一种基于决策树的算法,是由 Kaggle.com 主办的机器学习和预测竞赛的最佳算法,它被认为是回归分析的一种有竞争力的替代方案,并成功用于预测临床不良结果[24]。结合本研究结果,

综合模型灵敏度、特异度、F_1 分数、Brier 分数、AUROC、AUPRC 的比较,发现基于 XGBoost 算法建立预测非肝脏手术术后肝损伤风险模型为最适模型,在建模组和验证组中,基于 XGBoost 算法建立的模型的 AUROC>0.9,AUPRC 为 0.841,灵敏度、特异度和 F1 分数皆比其他算法高,预测效能好。临床决策曲线分析用于确定在不同风险水平下执行各种不同 ML 模型的净获益率,以评估模型的决策效用[25],本研究中,基于 XGBoost 算法开发的预测模型表现出了最大的净获益率,即患者在使用该模型决策干预下能获得最大的收益。

本研究存在一定的局限性。首先,本研究的数据集虽然是多中心数据,但是均属于回顾性信息,纳入患者数量较少,且缺少外部验证;其次,本研究只对非肝脏手术这一大类手术进行了预测,未对具体手术类型进行细化研究;再次,由于病例时间跨度较大,部分纸质档病例存在字迹模糊,数据缺失,故术中血压等生命体征指标,在被选择为建模指标的过程中被舍弃,导致对术中出血和术中的生命体征变化这类肝损伤重要因素的关注度不够;最后,虽然机器学习可以分析大量的变量和数据,但由于数据量不足,可能会导致结果无效。

综上所述,本研究利用多中心数据基于多个机器学习算法开发了术后肝损伤预测模型,综合比较了各模型的预测性能和临床效益,最终筛选出了基于 XGBoost 算法开发的最佳预测模型,为术前肝损伤风险评估、指导围术期肝损伤预防性治疗提供可能。

参考文献:

[1] 栾正刚, 马晓春. 围手术期肝功能障碍类型及处理[J]. 中国实用外科杂志, 2014, 34(2): 123-126.
LUAN Z G, MA X C. Perioperative liver dysfunction: types and management[J]. Chin J Pract Surg, 2014, 34(2): 123-126.

[2] CHETTY G, SHARPE D A, NANDI J, *et al*. Liver blood flow during cardiac surgery[J]. Perfusion, 2004, 19(3): 153-156. DOI: 10.1191/0267659104pf735oa.

[3] ALSHAHRANI A S, GONG G S, YOO M W. Comparison of long-term survival and immediate postoperative liver function after laparoscopic and open distal gastrectomy for early gastric cancer patients with liver cirrhosis[J]. Gastric Cancer, 2017, 20(4): 744-751. DOI: 10.1007/s10120-016-0675-4.

[4] WANG H J, BAO Q Q, CAO D H, *et al*. Characteristics of computed tomography images for patients with acute liver injury caused by sepsis under deep learning algorithm[J]. Contrast Media Mol Imaging, 2022, 2022: 9322196. DOI: 10.1155/2022/9322196.

[5] ESMAILY H, TAYEFI M, DOOSTI H, *et al*. A comparison between decision tree and random forest in determining the risk factors associated with type 2 diabetes[J]. J Res Health Sci, 2018, 18(2): e00412. [PubMed]

[6] WANG J T, ZHENG T L, LIAO Y, *et al*. Machine learning prediction model for post-hepatectomy liver failure in hepatocellular carcinoma: a multicenter study[J]. Front Oncol, 2022, 12: 986867. DOI: 10.3389/fonc.2022.986867.

[7] DIAB M, SPONHOLZ C, VON LOEFFELHOLZ C, *et al*. Impact of perioperative liver dysfunction on in-hospital mortality and long-term survival in infective endocarditis patients[J]. Infection, 2017, 45(6): 857-866. DOI: 10.1007/s15010-017-1064-6.

[8] GARCÍA-ORDÁS M T, BENAVIDES C, BENÍTEZ-ANDRADES J A, *et al*. Diabetes detection using deep learning techniques with oversampling and feature augmentation[J]. Comput Methods Programs Biomed, 2021, 202: 105968. DOI: 10.1016/j.cmpb.2021.105968.

[9] ÇAGLı K, BAŞAR F N, TOK D, *et al*. How to interpret liver function tests in heart failure patients?[J]. Turk J Gastroenterol, 2015, 26(3): 197-203. DOI: 10.5152/tjg.2015.0086.

[10] LIANG S H, LIU Y X, ZHANG B W, *et al*. Postoperative hepatic dysfunction after frozen elephant trunk for type A aortic dissection[J]. Front Cardiovasc Med, 2021, 8: 739606. DOI: 10.3389/fcvm.2021.739606.

[11] LIN X F, XIE L F, JIANG D B, *et al*. Hepatic dysfunction and adverse outcomes after total arch repair of acute type A aortic dissection: application of the MELD-XI score[J]. BMC Cardiovasc Disord, 2022, 22(1): 491. DOI: 10.1186/s12872-022-02934-w.

[12] 黄海嵘, 景华. 心肺转流术后肝损伤机制的研究进展[J]. 医学研究生学报, 2007, 20(8): 885-889. DOI: 10.16571/j.cnki.1008-8199.2007.08.011.
HUANG H R, JING H. Research progress of the mechanism of hepatic injury after cardiopulmonary bypass[J]. J Med Postgrad, 2007, 20(8): 885-889. DOI: 10.16571/j.cnki.1008-8199.2007.08.011.

[13] CHEUNG A T, CRUZ-SHIAVONE G E, MENG Q C, *et al*. Cardiopulmonary bypass, hemolysis, and nitroprusside-induced cyanide production[J]. Anesth Analg, 2007, 105(1): 29-33. DOI: 10.1213/01.ane.0000264078.34514.32.

[14] LIN J, MOORE D, HOCKEY B, *et al*. Drug-induced hepatotoxicity: incidence of abnormal liver function tests

consistent with volatile anaesthetic hepatitis in trauma patients[J]. Liver Int, 2014, 34(4): 576-582. DOI: 10.1111/liv.12278.

[15] NICOLL A, MOORE D, NJOKU D, *et al*. Repeated exposure to modern volatile anaesthetics may cause chronic hepatitis as well as acute liver injury[J]. BMJ Case Rep, 2012, 2012: bcr2012006543. DOI: 10.1136/bcr-2012-006543.

[16] WANG Y, JI C P, WANG Y, *et al*. Predicting postoperative liver cancer death outcomes with machine learning[J]. Curr Med Res Opin, 2021, 37(4): 629-634. DOI: 10.1080/03007995.2021.1885361.

[17] HOLLINGER A, SIEGEMUND M, GOETTEL N, *et al*. Postoperative delirium in cardiac surgery: an unavoidable menace? [J]. J Cardiothorac Vasc Anesth, 2015, 29(6): 1677-1687. DOI: 10.1053/j.jvca.2014.08.021.

[18] YOUNG P Y, KHADAROO R G. Surgical site infections [J]. Surg Clin N Am, 2014, 94(6): 1245-1264. DOI: 10.1016/j.suc.2014.08.008.

[19] BRATZLER D W, HOUCK P M. Antimicrobial prophylaxis for surgery: an advisory statement from the National Surgical Infection Prevention Project[J]. Am J Surg, 2005, 189(4): 395-404. DOI: 10.1016/j.amjsurg.2005.01.015.

[20] BENKE K, BENKE G. Artificial intelligence and big data in public health[J]. Int J Environ Res Public Health, 2018, 15(12): E2796. DOI: 10.3390/ijerph15122796.

[21] FERRARESE A, SARTORI G, ORRÙ G, *et al*. Machine learning in liver transplantation: a tool for some unsolved questions? [J]. Transpl Int, 2021, 34(3): 398-411. DOI: 10.1111/tri.13818.

[22] WU C C, YEH W C, HSU W D, *et al*. Prediction of fatty liver disease using machine learning algorithms[J]. Comput Methods Programs Biomed, 2019, 170: 23-29. DOI: 10.1016/j.cmpb.2018.12.032.

[23] YANG M, PENG B, ZHUANG Q, *et al*. Models to predict the short-term survival of acute-on-chronic liver failure patients following liver transplantation [J]. BMC Gastroenterol, 2022, 22(1): 80. DOI: 10.1186/s12876-022-02164-6.

[24] CHEN T Q, GUESTRIN C. XGBoost: a scalable tree boosting system, proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2016[C]. California: ACM, 2016: 785-794. DOI: 10.1145/2939672.2939785.

[25] VICKERS A J, HOLLAND F. Decision curve analysis to evaluate the clinical benefit of prediction models[J]. Spine J, 2021, 21(10): 1643-1648. DOI: 10.1016/j.spinee.2021.02.024.

(收稿:2023-12-04;修回:2024-02-09)

(编辑 邓强庭)