DOI:10.16016/j.2097-0927.202301053

基于术中指标建立非心胸手术术后呼吸衰竭梯度提升预测模型

黄家号,李雨捷,刘 祥,杨智勇,孙义竹,梁 浩,易 斌,鲁开智 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** 开发并验证一个基于术中指标在非心胸手术患者术后呼吸衰竭(postoperative respiratory failure,PRF)的机器学习预测模型。**方法** 纳入西南医院 2014 年 1 月至 2019 年 6 月行非心胸手术患者 705 例[训练集 565 例(PRF 128 例),测试集 140 例(PRF 35 例)]、华西医院 2019 年 5 月至 2020 年 1 月和中山医院 2019 年 6 月至 2019 年 12 月行非心胸手术患者 164 例[验证集 164 例(PRF 41 例)]。提取患者 19 项术中预测指标,通过 6 种机器学习算法:梯度提升模型(gradient boosting model,GBM)、广义线性模型(generalize linear model,GLM)、k-近邻(k-nearest neighbor,KNN)、朴素贝叶斯(naive bayes,NB)、神经网络(neural network,NNET),支持向量机(support vector machine linear,SVM)开发及测试模型,并在验证集进行验证,通过各模型间性能对比,筛选出最佳模型,最终建立网页预测模型。**结果** GBM 获得了最佳性能,准确性 76.2%(95%*CI*:69.0%~82.5%),受试者操作曲线下面积(area under the subject curve,AUC):0.794(95%*CI*:0.707~0.882),精准-召回曲线下面积(area under the precision-recall curve,AUPRC):0.641,Brier 评分:0.169。**结论** 基于 GBM 算法开发的模型具有更高的泛化性、准确性、临床实用性,并有助于避免过度拟合。建立的网页预测模型(http://150.158.55.139)可为患者 PRF 提供新的动态评估方法,量化手术风险。

[关键词] 术后呼吸衰竭;预测模型;临床实用性;量化手术风险
[中图法分类号] R319;R563.806;R619.9 [文献标志码] A

Establishment of a gradient boosting prediction model for respiratory failure after non-cardiothoracic surgery based on intraoperative indicators

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

[**Abstract**] **Objective** To develop and validate a machine learning prediction model for postoperative respiratory failure (PRF) in patients after non-cardiothoracic surgery based on intraoperative indicators. **Methods** A total of 705 patients undergoing non-cardiothoracic surgery in our hospital from January 2014 to June 2019 were enrolled, and then 565 patients of them were assigned in the training set (including 128 cases of PRF) and 140 patients into the test set (35 cases of PRF). Another 164 patients undergoing non-cardiothoracic surgery at West China Hospital from May 2019 to January 2020 and Zhongshan Hospital from June 2019 to December 2019 were assigned into the validation set (41 cases of PRF). Nineteen intraoperative indicators were extracted, and 6 machine learning algorithms, such as gradient boosting model (GBM), generalize linear model (GLM), k-nearest neighbor (KNN), naive bayes (NB), neural network (NNET), and support vector machine linear (SVM) were used to develop and test the models and were verified in the validation set. The best model was screened out by comparing the performance of each model, and finally, the web page prediction model was established. **Results** GBM obtained the best performance, with an accuracy

[基金项目] 国家重点研发计划(2018YFC0116702)
[通信作者] 鲁开智,E-mail:lukaizhi2010@163.com

2097-0927/Copyright © 2022,陆军军医大学(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/)

of 76.2%(95%*CI*: 69.0%~82.5%), an area under the subject curve (AUC) of 0.794 (95%*CI*: 0.707~0.882), an area under the precision-recall curve (AUPRC) of 0.641, and a Brier score of 0.169. **Conclusion** The model developed based on GBM algorithm is of higher generalization, accuracy, and clinical utility, and helps avoid overfitting. The developed web page prediction model (http://150.158.55.139) can provide a new dynamic evaluation method for PRF and quantify surgical risk.

[**Key words**] postoperative respiratory failure; prediction model; clinical utility; quantification of surgical risk

Supported by the National Key R&D Program of China (2018YFC0116702). Corresponding author: LU Kaizhi, E-mail: lukaizhi2010@163.com

术后呼吸衰竭(postoperative respiratory failure, PRF)是常见的术后肺部并发症[1],发病率为0.2%~7.5%,住院死亡率达25%~40%,使其成为最昂贵和致命的术后并发症之一[2],PRF比率被国外医疗研究质量机构和国家质量论坛等多个组织视为医疗质量的衡量标准[3]。麻醉学的临床研究越来越重视对围手术期各种并发症的预测[4],该领域产生了许多疾病严重程度计算器、早期预警评分等预测分析工具[5-7],包括相关呼吸衰竭的预测评分[8-9]。随着手术的进展,同时伴随着术中相关因素动态变化,患者患有PRF的风险逐渐升高,为了优化术后结局、预防术后并发症及确定术后转归,预测PRF高危患者至关重要。因此,本试验通过收集患者术中相关指标,开发并验证一个非心胸手术患者PRF的机器学习(machine learning, ML)动态预测模型,量化术后结局的风险。

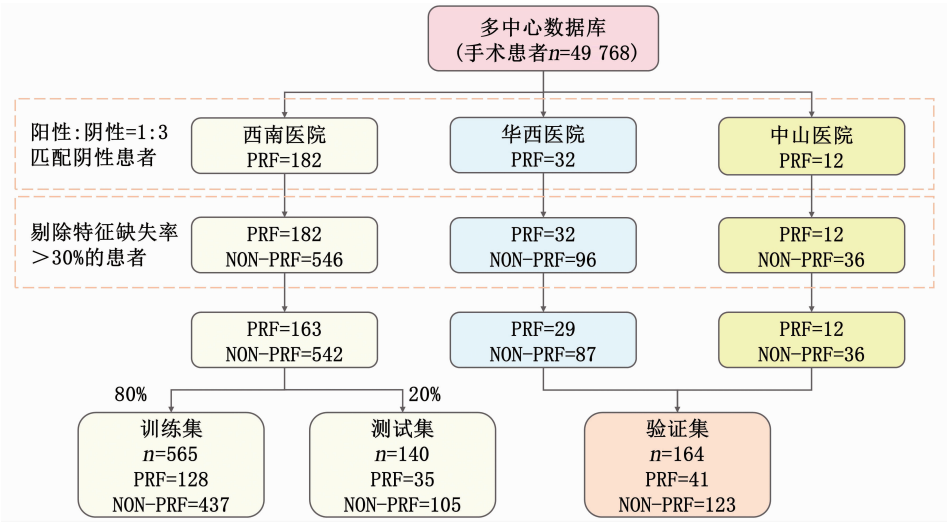
1 材料和方法

1.1 一般材料

1.1.1 数据来源 本研究数据来源于国内3家不同区域大型综合医院开发的多中心数据库,并通过3家医院伦理委员会批准,主要包括陆军军医大学附属西南医院(2014年1月至2019年6月,伦理审查批件号:KY201936)、四川大学附属华西医院(2019年5月至2020年1月,伦理审查批件号:2021-349)、复旦大学附属中山医院(2019年6月至2019年12月,伦理审查批件号:2019-385)行非心胸手术患者组成。

1.1.2 研究对象 阳性纳入标准:①年龄≥18岁;②行择期非心胸手术患者;③术后诊断包含:各类型呼吸衰竭。阳性排除标准:①术前诊断包含:各类型呼吸衰竭;②术前已气管插管患者;③二次及以上手术患者;④手术方式经胸腔入路;⑤住院期间中发生呼吸心跳骤停、脑死亡等危重症患者;⑥缺失数据>30%的患者。在数据库中,其余纳入、排除标准同阳性病例,按阳性:阴性=1:3比例倾向匹配同期非心胸手术术后未发生呼吸衰竭病例作为阴性对照病例。

本试验将PRF定义为:在本次住院期间手术后出现血气分析中动脉氧分压(PaO_2)与吸入氧分数(FiO_2)之比<300 mmHg、动脉血氧饱和度(SpO_2%)<90%且需要氧气治疗改善症状或手术后因呼吸相关非计划的再次气管插管,经有经验的医师明确诊断呼吸功能衰竭[1,7]。患者收集流程见图1。



PRF,术后呼吸衰竭;NON-PRF,无术后呼吸衰竭

图1 多中心数据库患者收集数量及模型集合划分

1.1.3　数据收集　　患者术中预测指标收集具体如下:①手术部位(其他部位,四肢手术,脊柱手术,头、面或颈部手术,下腹部手术,上腹部手术);②手术类型(开放手术、微创手术);③患者入手术室后第一次生命体征(收缩压、舒张压、脉搏、体温);④手术中输注液体量、血液量(晶体液、胶体液、红细胞、血浆);⑤手术出液量(失血量、尿量);⑥手术时间等共 19 项来开发模型。

1.2　方法

1.2.1　数据预处理　　本研究使用了多重插补方法(multiple imputation,MI)填补缺失程度小于总体 30%的缺失值;该方法是从单一插补的基础上衍生而来,每个数据集中的缺失数据用蒙特卡洛方法来填补,给每个缺失值都构造多个替代值,这将导致创建多个完整的数据集,其中已填充缺失的值以合理的值输入。然后,在每个完整数据集中分别进行科学价值分析,并将结果汇总到输入数据集中。通过这种方式,MI 能让缺失值的输入具有不确定性,并提升模型的阳性判断率,进而提升模型的预测性能[10]。

对于分类变量,本研究执行了 1-hot 编码技术,其定义为将分类值分成零值和非零值的过程,目的是将二分类变量转换为可用于分类算法的格式,使之符合机器学习计算过程,并提高计算效率[11]。

对于数值变量,本研究将数据进行了中心化和标准化,不同预测指标往往具有不同的量纲和量纲单位,为了消除指标之间的量纲影响,需要进行数据中心化、标准化处理,以解决预测指标之间的可比性。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[12]。将西南医院数据集随机分成训练集(80%)和测试集(20%),采用随机过采样(random over sampling examples,ROSE)来改善训练数据集中的数据不平衡,训练集用于模型的开发,确定超参数;测试集用于模型的测试,华西医院和中山医院数据集全部用于模型的外部验证。

1.2.2　模型开发　　为了将模型性能最大化,患者所有客观指标均被作为预测指标,在训练集中进行了多个模型的开发,并进行外部验证。使用了 R 软件“caret”包中包含的 6 种代表性模型算法来进行模型训练,分别是:梯度提升模型(gradient boosting model,GBM),广义线性模型(generalize linear model,GLM),k-近邻(k-nearest neighbor,KNN),朴素贝叶斯(naive bayes,NB),神经网络(neural network,NNET),支持向量机(support vector machine linear,SVM)。为了得到更准确可靠的预测模型,对于每种模型的建立都使用了 10 折交叉验证,迭代 5 次来增强模型性能。机器学习数据预处理、模型开发、测试和验证由 R 软件(4.2.1 版本)完成。

1.2.3　模型评估　　在外部验证集中构建受试者工作特征曲线(receiver operating characteristic curve,ROC 曲线)和精准-召回曲线(percision-recall curve,P-R 曲线),通过受试者操作曲线下面积(area under the subject curve,AUC)来评价模型的区分度,即模型鉴别能力;通过观察到的 PRF 概率和预测 PRF 概率构建校准曲线来评价模型的校准度,即模型的准确性;通过风险阈概率和患者净获益率构建临床决策曲线来评价模型的临床实用性。为了评估模型综合性能,同时也计算了所有模型的 Kappa 值、灵敏度、特异度、Brier 评分、F_1 评分筛选出最佳模型;最后,通过开发的最佳模型计算了每个预测特征的输入贡献,即特征重要性,从而使预测结果透明化。预测模型预处理、开发和评估流程见图 2。

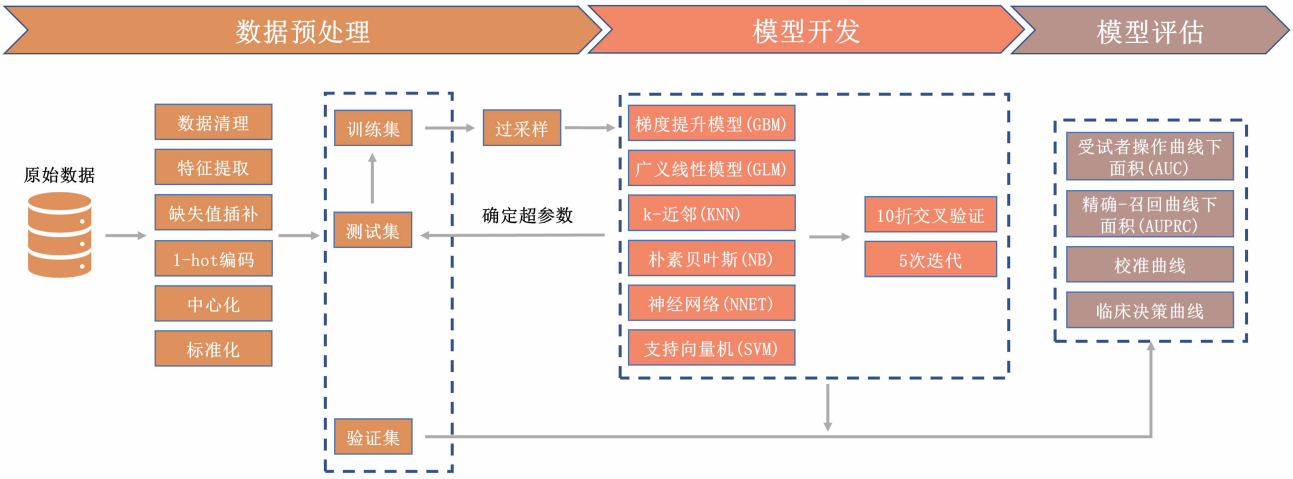


图 2　机器学习数据预处理、模型开发和模型评估流程图

1.2.4 模型网页建立 通过对多个模型的综合评估,确定适合 PRF 的最佳预测模型,并使用 Python 软件的“PyWebIO”包将模型映射到网页,通过输入患者预测指标信息即可出现患 PRF 的风险概率,从而量化结果风险。

1.3 统计学分析

PRF 组和 NON-PRF 组患者各项预测特征间进行统计学分析由 SPSS (26.0 版本)软件完成。正态分布的计量资料用 $\bar{x}\pm s$ 表示,组间比较采用两独立样本 t 检验;偏态分布的计量资料以 $M(P_{25},P_{75})$ 表示,组间比较采用 Mann-Whitney U 检验。计数资料采用计数值及百分比表示,组间比较采用 χ^2 检验或 *Fisher* 确切概率法;所有的统计检验都是双侧的,$P<0.05$ 为差异有统计学意义。

2 结果

2.1 一般情况

最终纳入患者数据分析 869 例(PRF 组 204 例,NON-PRF 组 665 例),模型拆分患者数据训练集 565 例(PRF 128 例),测试集 140 例(PRF 35 例),验证集 164 例(PRF 41 例)。与 NON-PRF 组患者相比,PRF 组患者在手术中输注的液体量更多(晶体液 1 000 mL *vs* 700 mL,$P<0.001$;胶体液 500 mL *vs* 0 mL,$P<0.001$),出液量更多(尿量 475 mL *vs* 300 mL,$P<0.05$;出血量 200 mL *vs* 150mL,$P<0.001$),手术时间更长(181.5 min *vs* 140.0 min,$P<0.001$,表 1)。

表 1 术后呼吸衰竭患者和无术后呼吸衰竭患者基线特征对比[例(%),$M(P_{25},P_{75})$]

变量	术后呼吸衰竭(n=204)	无术后呼吸衰竭(n=665)	P
手术部位			
上腹部	53(26.0)	92(13.8)	<0.001
下腹部	65(31.9)	205(30.8)	0.780
头、面或颈部	62(30.4)	116(17.4)	<0.001
脊柱	11(5.4)	84(12.6)	0.004
四肢	12(5.9)	163(24.5)	<0.001
其他	1(0.5)	2(0.3)	0.687
手术类型			
开放	171(83.8)	408(61.4)	<0.001
微创	33(16.2)	257(38.6)	<0.001
收缩压/mmHg	124(110,140)	125(115,137)	0.608
舒张压/mmHg	74(66,85)	78(70,87)	0.006
脉搏/次·min^{-1}	85(78,100)	80(75,89)	<0.001
体温/℃	36.6(36.5,37.0)	36.5(36.4,36.7)	<0.001
输注晶体液/mL	1 000(500,1 500)	700(500,1 100)	0.001
输注胶体液/mL	500(0,1000)	0(0,500)	<0.001
输注红细胞/mL	0(0,400)	0(0,0)	<0.001
输注血浆/mL	0(0,395)	0(0,0)	<0.001
尿量/mL	475(200,800)	300(200,600)	0.008
出血量/mL	200(100,500)	150(50,300)	<0.001
手术时间/min	181.5(101.3,298.8)	140.0(81.0,215.0)	<0.001

2.2 模型预测性能及验证结果

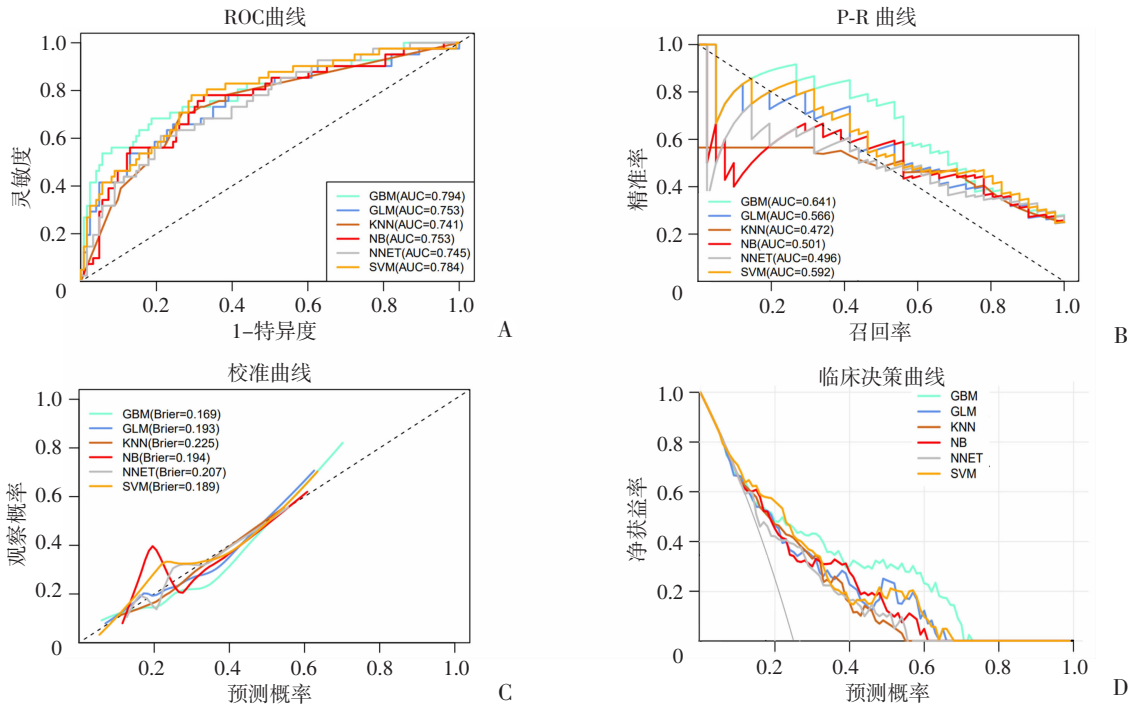
基于 GBM 算法开发的模型在验证集获得了最佳的鉴别能力(AUC:0.794,95% *CI*:0.707~0.882,AUPRC:0.641)、最佳的校准能力(Brier 评分=0.169)和准确性(76.2%);在临床决策曲线中,使用 GBM 算法开发的模型在整个阈概率范围内能够获得最大的净获益率(图 3),模型综合性能比较见表 2;GBM 模型超参数设定:树选择=150,树深度=3,收缩量=0.1,节点中开始分裂的最小训练集样本数=10。

2.3 特征重要性及网页模型建立

GBM 模型给出了特征重要性排序(图 4),观察特征重要性可以得知模型在预测患者结局时各特征对结局的贡献度,对 PRF 影响程度较大的术中特征为输注胶体量、收缩压、开放手术等。开发的预测模型被映射到网页(http://150.158.55.139),在预测网页界面输入患者术中特征信息后,点击预测即可出现该患者发生 PRF 的概率。

表 2 机器学习各预测模型综合性能比较

模型	准确性(%),95%*CI*	Kappa 值	灵敏度(%)	特异度(%)	Brier 评分	F_1 评分	AUPRC	AUC,95%*CI*
GBM	76.2,69.0~82.5	0.427	68.3	78.9	0.169	0.589	0.641	0.794,0.707~0.882
GLM	71.9,64.4~78.7	0.348	65.9	74.0	0.193	0.540	0.566	0.753,0.661~0.846
KNN	71.3,63.8~78.1	0.365	73.2	70.7	0.225	0.561	0.472	0.741,0.654~0.827
NB	76.2,69.0~82.5	0.381	56.1	82.9	0.194	0.541	0.501	0.753,0.663~0.843
NNET	73.2,65.7~79.8	0.328	56.1	78.9	0.207	0.511	0.496	0.745,0.659~0.831
SVM	71.3,63.8~78.1	0.356	70.7	71.5	0.189	0.552	0.592	0.784,0.700~0.868



A:各模型间 ROC 曲线比较;B:各模型间 P-R 曲线比较;C:各模型间校准曲线比较;D:各模型间临床决策曲线比较

图 3 机器学习各预测模型性能曲线对比

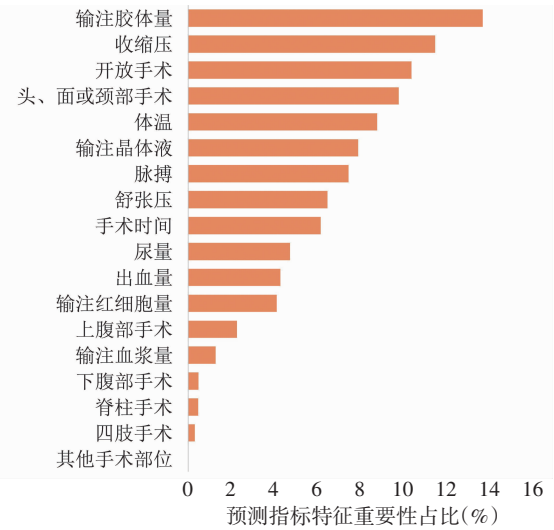


图 4 GBM 预测模型术中特征重要性占比排序

3 讨论

使用自动化工具评估术后呼吸衰竭的概率可以识别出有风险的患者进行早期干预,这将有助于有效分配资源并改善患者的预后[13]。我们提取了临床工作中最为常见的 19 项患者特征作为模型的预测指标,这些指标也是术中最容易获取的,通过这些指标建立了术后呼吸衰竭预测模型,并在外部数据集得到了很好的验证。

在各研究中,PRF 的发生率不同,主要是由于 PRF 发生时间的限定未达成共识,根据研究者情况不同而定义不同[14],对于 PRF 术后 2[15]、3[4]、5[1]、7[16]天及整个住院时间内[17]发生的呼吸衰竭均有研究,为了发掘住院期间手术患者该病症的发生与各项特征间的联系,因此,本研究对 PRF 的定义如前所述。研究证明病例对照超过 1∶4时,统计效率不会明显增加,但工作量却增大。如果病例与对照来源都充足,则以 1∶1匹配最合适;考虑到模型建立需要大量病例来进行训练,本研究选择了 1∶3比例进行匹配阴性对照病例。阳性患者相对于阴性患者较少,训练数据中阴性患者过多,会导致模型无法充分学习到阳性患者的信息,这时候模型的正确率往往较高,但特异度却很低,即模型识别阳性患者的能力很差,需要解决数据集不均衡问题带来的影响,所以我们采用了 ROSE 来改善训练数据集中的数据不平衡;ROSE 是一种统计技术,用于以平衡的方式增加数据集中的少数类案例数量,ROSE 使用自举方法从少数类创建人工实例,为复制的样本添加了一些噪声,并在合成的少数样本中创建了更多的多样性,与传统过采样相比,该技术更受青睐,并可减少过采样可能导致的过拟合。

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,并评估模型的决策效用[18-19];具有高决策曲线分析的模型可以帮助临床

医生筛选最终 PRF 患者的获益率,基于 GBM 算法开发的预测模型在广泛的阈概率范围内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。结合对模型的综合评估,可以得出结论,GBM 模型更适合提高泛化性、准确性、临床实用性,并有助于避免过度拟合。这一观察结果也与 KHAN、STIGLIC 等[20-21]开发的 ML 模型结果一致。

GBM 算法是一种集成学习技术,它结合了基于梯度的优化和增强技术。基于梯度的优化使用梯度计算来最小化模型在训练数据方面的损失函数,而增强技术应用了组装弱学习器的迭代过程,以获得适合回归和分类任务的预测模型。该算法还能进行自动特征选择,对重要特征进行优先级排序,并丢弃包含不相关或冗余信息的变量。

从 GBM 模型给出的特征重要性排序可以得知,输注胶体量在术中特征中占了最重要的位置,该特征是血浆胶体渗透压的重要影响因素,短时间内急剧增加血浆胶体渗透压可能会导致循环负荷过重,表现出呼吸窘迫、低氧血症、肺水肿等;我们更加考虑的重要原因是随着手术的进展,对于复杂性手术,手术难度大,出血量多,手术时间延长,患者术后结局更加不可控,均导致了术中输注胶体量增加;与下腹部手术相比,PRF 更容易在上腹部手术中发生,可能的解释是上腹部手术导致膈肌运动受限,而上腹部手术的切口往往会损伤膈肌和呼吸肌,术后疼痛会引起低通气,更多的是导致术后肺扩张不足的影响,从而导致肺不张,也可能导致术后患者排痰困难,逐渐发生肺部感染,最终导致呼吸衰竭。

本试验具有几个局限性:首先,本试验的数据集虽然来自 3 家国内大型综合医院,但是均属于回顾性信息,并且数量有限,所以我们下一步考虑继续使用更多病症数据集的数据库进行模型开发,同时使用预测模型到日常评估患者的工作中来收集患者数据,进行前瞻性验证。其次,本研究提取的术中预测指标较少,可能与术后结局相关的潮气量,呼吸末正压等呼吸机参数未能获取,在后续的研究中我们会更加关注此类特征的纳入。

综上所述,本研究通过多个机器学习算法进行模型的开发,综合比较了各模型的预测性能和临床实用性,最终筛选出了基于 GBM 算法开发的最佳预测模型,量化了 PRF 风险,提高循证治疗的依从性,为临床医生提供了较好的评估参考,并指导有关呼吸衰竭治疗的临床决策。

参考文献:

[1] CANET J, SABATÉ S, MAZO V, *et al*. Development and validation of a score to predict postoperative respiratory failure in a multicentre European cohort: a prospective, observational study[J]. Eur J Anaesthesiol, 2015, 32(7): 458-470. DOI: 10.1097/EJA.0000000000000223.

[2] JACQUELINE C S, GARTH H U, CHRISTIANA D, *et al*. Postoperative respiratory failure: an update on the validity of the Agency for Healthcare Research and Quality Patient Safety Indicator 11 in an era of clinical documentation improvement programs[J]. Am J Surg, 2020, 220(1): 222-228. DOI: 10.1016/j.amjsurg.2019.11.019.

[3] HADAYA J, VERMA A, SANAIHA Y, *et al*. Machine learning-based modeling of acute respiratory failure following emergency general surgery operations[J]. PLoS One, 2022, 17(4): e0267733. DOI: 10.1371/journal.pone.0267733.

[4] YOUNG A, RAMACHANDRAN S K. Clinical prediction of postoperative respiratory failure[J]. Anesthesiology, 2013, 118(6): 1247-1249. DOI: 10.1097/ALN.0b013e31829303c7.

[5] CHURPEK M M, CAREY K A, DELA MERCED N, *et al*. Validation of early warning scores at two long-term acute care hospitals[J]. Crit Care Med, 2019, 47(12): e962-e965. DOI: 10.1097/CCM.0000000000004026.

[6] DZIADZKO M A, NOVOTNY P J, SLOAN J, *et al*. Multicenter derivation and validation of an early warning score for acute respiratory failure or death in the hospital[J]. Crit Care, 2018, 22(1): 286. DOI: 10.1186/s13054-018-2194-7.

[7] YU S, LEUNG S, HEO M, *et al*. Comparison of risk prediction scoring systems for ward patients: a retrospective nested case-control study[J]. Crit Care, 2014, 18(3): R132. DOI: 10.1186/cc13947.

[8] CANET J, GALLART L, GOMAR C, *et al*. Prediction of postoperative pulmonary complications in a population-based surgical cohort[J]. Anesthesiology, 2010, 113(6): 1338-1350. DOI: 10.1097/ALN.0b013e3181fc6e0a.

[9] AROZULLAH A M, DALEY J, HENDERSON W G, *et al*. Multifactorial risk index for predicting postoperative respiratory failure in men after major noncardiac surgery. The National Veterans Administration Surgical Quality Improvement Program[J]. Ann Surg, 2000, 232(2): 242-253. DOI: 10.1097/00000658-200008000-00015.

[10] AUSTIN P C, WHITE I R, LEE D S, *et al*. Missing data in clinical research: a tutorial on multiple imputation[J]. Can J Cardiol, 2021, 37(9): 1322-1331. DOI: 10.1016/j.cjca.2020.11.010.

[11] AL'AREF S J, SINGH G, VAN ROSENDAEL A R, *et al*. Determinants of in-hospital mortality after percutaneous

coronary intervention: a machine learning approach[J]. J Am Heart Assoc, 2019, 8(5). DOI: 10.1161/jaha.118.011160.

[12] 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.

[13] BOLOURANI S, BRENNER M, WANG P, *et al*. A machine learning prediction model of respiratory failure within 48 hours of patient admission for COVID-19: model development and validation[J]. J Med Internet Res, 2021, 23(2): e24246. DOI: 10.2196/24246.

[14] RUAN H Y, LIU F C, GONG C F, *et al*. Risk factors associated with postoperative respiratory failure in tuberculous empyema patients[J]. Medicine, 2021, 100(23): e25754. DOI: 10.1097/MD.0000000000025754.

[15] WAKEFIELD C J, JOCHUM S B, HEJNA E, *et al*. Novel application of respiratory muscle index obtained from chest computed tomography to predict postoperative respiratory failure after major non-cardiothoracic surgery[J]. Am J Surg, 2021, 222(5): 1029-1033. DOI: 10.1016/j.amjsurg.2021.04.022.

[16] ZAYED Y, KHEIRI B, BARBARAWI M, *et al*. Effect of oxygenation modalities among patients with postoperative respiratory failure: a pairwise and network meta-analysis of randomized controlled trials[J]. J Intensive Care, 2020, 8(1). DOI: 10.1186/s40560-020-00468-x.

[17] CHEN J, TIAN Z, ZHANG H, *et al*. Risks of postoperative respiratory failure in elderly patients after hip surgery: a retrospective study[J]. J Orthop Surg Res, 2022, 17(1). DOI: 10.1186/s13018-022-02909-9.

[18] 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.

[19] HSU J F, YANG C, LIN C Y, *et al*. Machine learning algorithms to predict mortality of neonates on mechanical intubation for respiratory failure[J]. Biomedicines, 2021, 9(10): 1377. DOI: 10.3390/biomedicines9101377.

[20] KHAN O, BADHIWALA J H, WITIW C D, *et al*. Machine learning algorithms for prediction of health-related quality-of-life after surgery for mild degenerative cervical myelopathy[J]. Spine J, 2021, 21(10): 1659-1669. DOI: 10.1016/j.spinee.2020.02.003.

[21] STIGLIC G, PAJNKIHAR M. Evaluation of major online diabetes risk calculators and computerized predictive models[J]. PLoS One, 2015, 10(11): e0142827. DOI: 10.1371/journal.pone.0142827.

(收稿:2023-01-01;修回:2023-02-16)

(编辑 栾 嘉)