DOI:10.16016/j.1000-5404.202104100

肝硬化伴自发性细菌性腹膜炎的诊断模型研究及验证

向守书[1],谈军涛[1],文元久[2],谭 超[1],龚 军[1],赵文龙[1] 400016 重庆,重庆医科大学医学信息学院[1];400038 重庆,陆军军医大学(第三军医大学)全科医学系[2]

[**摘要**] **目的** 筛选肝硬化伴自发性细菌性腹膜炎(spontaneous bacterial peritonitis,SBP)的影响因素,建立肝硬化伴 SBP 的诊断模型。**方法** 收集重庆 A、B 两家三甲医院 2015 年 7 月至 2019 年 12 月肝硬化患者 7 461 例,根据患者住院期间是否发生 SBP 分为 SBP 组(1 173 例)和非 SBP 组(6 288 例)。随机抽取 A 院 70%(3 776 例)患者作为训练集,剩余 30%(1 619 例)作为内部验证集,B 院患者(2 066 例)作为外部验证集。通过单因素分析和 Logistic 回归分析筛选变量,建立 Logistic 回归、随机森林、决策树、XGBoost 和优化 Logistic 回归模型。在内部验证和外部验证中,比较并验证模型对肝硬化伴 SBP 的诊断价值。**结果** 机器学习算法均提示指标失代偿(OR=5.354,95%CI:3.770-7.803),淋巴细胞百分比(OR=0.951,95%CI:0.939-0.962)、总胆红素(OR=1.003,95%CI:1.002-1.004)、C 反应蛋白是否异常(OR=1.626,95%CI:1.310-2.017)、国际标准化比值(OR=1.346,95%CI:1.091-1.681)、前白蛋白(OR=0.990,95%CI:0.987-0.993)、终末期肝病模型评分(model for end stage liver disease,MELD)(OR=1.038,95%CI:1.015-1.063)为显著性因素,优化 Logistic 内部验证的 AUC 为 0.860,灵敏度 0.872,特异度 0.719;外部验证的 AUC 为 0.818;灵敏度为 0.662,特异度为 0.812。经过 delong test 检测优化 Logistic 回归模型和其余模型的 AUC 差异无统计学意义。**结论** 基于机器学习算法建立的肝硬化合并 SBP 风险预测模型具有较高的诊断价值,其中优化 Logistic 回归模型的诊断效能在内部验证和外部验证中均表现良好,可为临床诊断 SBP 提供参考。

[**关键词**] 肝硬化;自发性细菌性腹膜炎;机器学习;诊断模型

[**中图法分类号**] R195.4;R572.2;R575.2 [**文献标志码**] A

Establishment and verification of a diagnostic model of liver cirrhosis with spontaneous bacterial peritonitis

XIANG Shoushu[1], TAN Juntao[1], WEN Yuanjiu[2], TAN Chao[1], GONG Jun[1], ZHAO Wenlong[1] [1]School of Medical Information, Chongqing Medical University, Chongqing, 400016; [2]Department of General Medicine, Army Medical University (Third Military Medical University), Chongqing, 400038, China

[**Abstract**] **Objective** To screen the influencing factors of liver cirrhosis with spontaneous bacterial peritonitis (SBP) so as to establish and verify a diagnostic model of cirrhosis with SBP. **Methods** A total of 7 461 patients with liver cirrhosis were collected from 2 Grade-A hospitals in Chongqing (Hospital A and B) between July 2015 and December 2019. According to the occurrence of SBP during hospitalization, they were divided into SBP (n = 1 173) and non-SBP (n = 6 288) groups. A total of 3 776 patients (70%) from Hospital A were randomly selected as the training set, while the remaining 1 619 patients (30%) and the 2 066 patients from hospital B were subjected as the internal and external validation sets respectively. Univariate and logistic analyses were used to screen variables, and logistic regression, random forest (RF), decision tree (DT) and XGBoost models were subsequently established using the training set. Then, the

[**基金项目**] 重庆医科大学智慧医学项目(ZHYX2019001)

[**通信作者**] 赵文龙,E-mail:cqzhaowl@163.com

1000-5404/Copyright © 2021, 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/)

optimized logistic regression model was built on the basis of the above 4 models. Finally, the 5 models were applied in the internal and external verification sets for evaluating and comparing the diagnostic value of different models for cirrhosis with SBP. **Results** The machine learning algorithm suggested that the 7 common influencing factors with significance in all models were as follows: decompensation stage (*OR* = 5.354, 95%*CI*: 3.770-7.803), lymphocyte percentage (*OR* = 0.951, 95%*CI*: 0.939-0.962), total bilirubin (*OR* = 1.003, 95%*CI*: 1.002-1.004), abnormal C-reactive protein (*OR* = 1.626, 95%*CI*: 1.310-2.017), international normalized ratio (*OR* = 1.346, 95%*CI*: 1.091-1.681), prealbumin (*OR* = 0.990, 95%*CI*: 0.987-0.993) and model for end stage liver disease (MELD) score (*OR* = 1.038, 95%*CI*: 1.015-1.063). The AUC of the internal verification of the optimized logistic model was 0.860, with a sensitivity of 0.872 and a specificity of 0.719; while the corresponding values of the external verification were 0.818, 0.662 and 0.812, respectively. Delong test showed that there was no statistical difference in AUC between the optimized logistic regression model and the other models with favorable performance. **Conclusion** The risk prediction models established by machine learning algorithm for liver cirrhosis complicated with SBP have high diagnostic value. Among them, the optimized logistic regression model has performed well in both internal and external verification, and is able to provide reference for clinical diagnosis of SBP.

[**Key words**] liver cirrhosis; spontaneous bacterial peritonitis; machine learning; diagnostic model

Supported by the Intelligent Medicine Research Project of Chongqing Medical University (ZHYX2019001). Corresponding author: ZHAO Wenlong, E-mail: cqzhaowl@163.com

自发性细菌性腹膜炎(spontaneous bacterial peritonitis,SBP)是在腹腔、邻近组织无感染源的情况下发生的腹水感染,是晚期肝脏病变和慢性重症肝炎的严重并发症。研究发现,SBP 在肝硬化患者中发病率为 10%~30%[1-2];有 SBP 病史的肝硬化患者 1 年内 SBP 复发率高达 40%~70%[3]。与未发生感染性疾病的肝硬化患者相比,被感染的肝硬化患者病死可能性要高出 4 倍[4-5],其住院病死率为 30%~50%[6-7],是终末期肝病的重要死亡原因之一[8]。由于 SBP 临床表现不典型、症状少、或有其他疾病常见的症状和体征,使得临床早期诊断十分困难。目前虽然临床上关于 SBP 的发病机制、诊断和治疗已有较多研究,但对于其早期诊断和预测仍存在一定的困难。

随着"健康中国 2030"[9]的提出,疾病零级预防显得愈发重要,同时,居民对于个体健康管理服务的需求也在迅速增加,极大地催生了疾病风险预测的需求。收集和分析与医学诊疗过程和结果相关的大型数据集有望将医学转变为一门以数据为驱动、以结果为导向的学科,对疾病的检测、诊断、治疗具有深远影响[10]。机器学习(machine learning,ML)是一门以数学和计算机为基础的学科,通过一系列算法来构建数据驱动的推理和预测模型[11],因其方法学优势在疾病风险预测领域发挥着越来越重要的作用。临床医学常用监督学习算法,即通过学习一组包含输入对象和期望的输出值的标签数据集,产生具有推断功能的模型,然后将模型应用于未知数据集,达到分类或回归的目标[12]。本研究利用常见的监督学习算法,如 Logistic 回归、随机森林(random forest,RF)、决策树(decision tree,DT)和 XGBoost[13-14],建立了机器学习模型,探究各个模型对早期自发性腹膜炎的诊断价值,并对模型进行内部验证和外部验证,为临床诊断 SBP 提供参考。

1 资料与方法

1.1 研究对象

回顾性收集重庆医科大学医学数据研究院大数据平台中重庆医科大学附属第二医院(A 医院)、重庆医科大学附属永川医院(B 医院)两家三甲医院 2015 年 7 月至 2019 年 12 月 1 173 例肝硬化伴自发性腹膜炎患者作为病例组,未发生自发性腹膜炎的 6 288 例肝硬化患者作为对照组。

纳入标准:年龄≥18 岁;所有肝硬化患者通过肝脏活检、临床、生化和影像学资料或既往病案证实,且诊断符合《肝硬化诊治指南》[15];SBP 的诊断标准依据 2010 年欧洲肝病学会肝硬化腹水、SBP 指南[16],即具备下述指标中的 2 项:(1)临床表现发热、腹痛,查体呈腹部压痛、反跳痛等腹腔内感染的症状与体征;(2)腹水常规检查白细胞>0.5×10^9/L,中性粒细胞>0.25×10^9/L;(3)腹水细菌培养阳性。

排除标准:死亡;患有获得性免疫缺陷综合征;患有肿瘤;继发性腹膜炎;结核性腹膜炎;既往有自发性腹膜炎病史。

1.2 研究指标

本研究通过咨询临床专家和查询相关文献纳入肝硬化患者的常见临床指标。(1)根据病案首页诊断确定病因:乙肝、丙肝、酒精肝、胆汁性肝炎、自身免疫性肝病;(2)根据出院诊断纳入 SBP 的常见并发症:上消化道出血、肝性脑病;(3)实验室资料:血常规、肝功能、肾功能、电解质、凝血功能指标和终末期肝病模型评分(model for end-stage liver disease, MELD);(4)病历中个人史等提取一般资料:年龄、性别、吸烟史、饮酒史、是否患有高血压、是否患有糖尿病。删除缺失率>30%的指标。

1.3 研究方法

本研究为横断面研究,采用 SPSS 24.0、R4.0.2 软件对数据进行整理分析,检验标准设为 $P<0.05$。计数资料采用率和百分比表示,组间比较采用卡方检验。服从正态分布的计量数据采用 $\bar{x}\pm s$ 表示,组间比较采用 t 检验;不服从正态分布的采用中位数和四分位数间距[$M(Q_{25}\sim Q_{75})$]表示,组间比较采用 Mann-Whitney U 检验。

将 A 医院的数据按照 7∶3的比例分为训练集(3 776 例)和内部验证集(1 619 例)。B 医院数据作为外部验证集(2 066 例)。将训练集中单因素分析($P<0.05$)筛选输入变量,是否发生自发性腹膜炎作为结局变量,利用 step()函数进行基于 AIC 准则的逐步 Logistic 回归模型。再将 Logistic 模型中有意义的指标作为输入变量,在训练集中分别建立 Logistic 回归模型、随机森林模型、决策树模型、XGBoost 模型。提取各个模型中的重要指标建立优化 Logistic 模型。在内部验证集和外部验证中用 ROC 曲线下面积(area under the curve,AUC)、灵敏度、特异度、阳性预测率、阴性预测率和准确度验证模型的诊断价值。

2 结果

2.1 训练集基线特征比较

共纳入 44 项研究指标。两组在性别、住院天数、吸烟、饮酒、失代偿期、淋巴细胞百分比、钙、总胆红素、尿酸、前白蛋白、凝血酶原时间、国际标准化比值等 34 项指标的差异具有统计学意义($P<0.05$),见表 1、2。

2.2 机器学习模型建模

在 Logistic 回归模型中,利用 step()函数进行基于 AIC 准则的逐步 Logistic 回归,最终纳入具有统计学意义的指标变量 13 个和临床重要指标 MELD 评分,见表 3。

随机森林模型中利用 which.min()函数寻找最小误差,误差最小时,ntree 为 180;分类问题中,mtry 为[$\log_2(p)+1$]个,P 为指标个数(14),故 mtry 设置为 4。变量重要性柱状图如图 1。

决策树模型中,应用 rpart()函数构建决策树模型,提示分支数为 3 次时,模型的最小交叉验证误差为 0.89,最优树的 CP 值为 0.010,见表 4。最后利用 prune()函数对模型进行剪枝。

XGBoost 模型利用 xgb.train()函数构建 XGBoost 模型,其 train()函数调整模型参数为:eta=0.01,max_depth=5,subsample=0.5,colsample_bytree=1,gamma=0.5,nrounds=75。用 gain 来衡量某个指标的增益[17],gain 值越大,指标对模型的影响越大。MELD 的 gain 值最大,为 0.298。变量重要性柱状图如图 2。

表 1 训练集中 SBP 组和非 SBP 组二分类变量比较[例(%)]

指标	非 SBP 组(n=3 180)	SBP 组(n=596)	χ^2	P
性别(男)	2 024(63.64)	443(74.32)	25.283	<0.001
高血压	484(15.22)	61(10.23)	10.100	0.001
糖尿病	599(18.83)	101(16.94)	1.187	0.276
肝性脑病	158(4.96)	86(14.42)	74.330	<0.001
上消化道出血	774(24.33)	123(20.63)	3.798	0.051
失代偿	2 078(65.34)	556(93.28)	185.763	<0.001
乙肝	1 588(49.93)	380(63.75)	38.422	<0.001
酒精肝	514(16.16)	146(24.49)	24.165	<0.001
胆汁性	518(16.28)	46(7.72)	29.022	<0.001
丙肝	181(5.69)	25(4.19)	2.181	0.140
自身免疫性	269(8.45)	38(6.375)	2.916	0.088
吸烟	1 369(43.05)	314(52.68)	18.858	<0.001
饮酒	1 361(42.79)	310(52.01)	17.276	<0.001
C 反应蛋白是否异常	1 011(31.79)	340(57.04)	139.321	<0.001

表 2 训练集中 SBP 组和非 SBP 组定量变量比较 [$M(Q_{25},Q_{75})$,$\bar{x}\pm s$]

指标	非 SBP 组(n=3 180)	SBP 组(n=596)	t/Z	P
就诊时年龄/岁	53.00(46.00,63.00)	52.00(45.00,61.00)	−3.100	0.002
住院天数/d	10.00(6.00,15.00)	15.00(9.00,22.75)	−12.256	<0.001
中性粒细胞百分比(%)	65.80(56.50,74.21)	73.66(65.90,81.60)	−14.142	<0.001
中性粒细胞绝对数/$10^9\cdot L^{-1}$	2.69(1.83,4.08)	3.94(2.61,6.41)	−12.732	<0.001
平均红细胞血红蛋白含量/pg	30.70(28.40,32.50)	31.80(29.40,33.80)	−7.211	<0.001
平均红细胞血红蛋白浓度/$g\cdot L^{-1}$	335.00(323.00,345.00)	344.00(330.00,356.00)	−11.025	<0.001
淋巴细胞百分比(%)	23.20(16.32,31.30)	15.71(10.50,21.78)	−15.809	<0.001
淋巴细胞数/$10^9\cdot L^{-1}$	0.98(0.65,1.43)	0.83(0.55,1.25)	−5.588	<0.001
白细胞计数/$10^9\cdot L^{-1}$	4.37(3.16,6.07)	5.45(3.83,8.38)	−9.466	<0.001
红细胞计数/$10^9\cdot L^{-1}$	3.75(3.13,4.32)	3.41(2.83,3.97)	−8.098	<0.001
血小板计数/$10^9\cdot L^{-1}$	84.00(54.00,141.00)	75.00(48.00,120.00)	−3.622	<0.001
钙/$mmol\cdot L^{-1}$	2.15(2.03,2.27)	2.05(1.96,2.16)	−11.888	<0.001
镁/$mmol\cdot L^{-1}$	0.80(0.73,0.86)	0.78(0.71,0.86)	−3.005	0.003
钠/$mmol\cdot L^{-1}$	139.00(136.00,141.40)	138.40(135.53,141.00)	−2.812	0.005
总胆红素/$\mu mol\cdot L^{-1}$	24.00(14.70,48.00)	73.10(27.28,236.13)	−17.736	<0.001
总胆汁酸/$\mu mol\cdot L^{-1}$	26.40(9.43,69.18)	76.45(25.18,164.80)	−14.682	<0.001
尿素/$mmol\cdot L^{-1}$	5.12(4.02,6.71)	5.60(3.99,7.99)	−3.907	<0.001
尿酸/$\mu mol\cdot L^{-1}$	288.85(222.40,369.48)	253.10(169.73,367.95)	−5.782	<0.001
碱性磷酸酶/$IU\cdot L^{-1}$	114.00(82.00,166.00)	125.00(91.25,170.75)	−2.978	0.003
前白蛋白/$mg\cdot L^{-1}$	102.00(64.00,154.00)	52.50(36.00,81.00)	−20.456	<0.001
凝血酶原时间/s	15.50(14.00,17.70)	19.40(16.60,23.88)	−21.308	<0.001
国际标准化比值	1.24(1.08,1.46)	1.64(1.35,2.11)	−21.389	<0.001
MELD	9.64(6.52,13.80)	17.53(11.26,24.03)	−19.95	<0.001
平均红细胞体积/fL	91.10(85.70,96.00)	91.70(85.13,98.15)	−1.944	0.052
血红蛋白浓度/$g\cdot L^{-1}$	104.00(85.00,125.00)	105.00(84.00,122.00)	−0.953	0.341
钾/$mmol\cdot L^{-1}$	3.90(3.60,4.20)	3.88(3.56,4.28)	−0.322	0.748
白蛋白/$g\cdot L^{-1}$	31.50(28.33,35.80)	31.35(28.10,35.30)	−1.22	0.223
肌酐/$\mu mol\cdot L^{-1}$	66.80(54.90,82.88)	69.00(55.03,88.93)	−1.922	0.055
γ 谷氨酰基转移酶/$IU\cdot L^{-1}$	59.50(29.00,146.00)	62.00(33.00,117.75)	−0.158	0.875
总蛋白/$g\cdot L^{-1}$	68.30±9.83	62.55±9.10	416.73	<0.001

表 3 多因素 Logistic 回归模型结果

指标	β	SE	OR(95%CI)	P
住院天数	0.020	0.004	1.020(1.012−1.029)	<0.001
失代偿	1.702	0.196	5.485(3.780−8.156)	<0.001
乙肝	0.271	0.135	1.312(1.008−1.713)	0.045
C 反应蛋白是否异常	0.433	0.114	1.541(1.232−1.927)	<0.001
平均红细胞血红蛋白浓度	0.008	0.003	1.008(1.002−1.014)	0.008
淋巴细胞百分比	−0.061	0.011	0.941(0.922−0.961)	<0.001
淋巴细胞数	0.618	0.265	1.854(1.118−3.158)	0.020
白细胞计数	−0.312	0.156	0.732(0.529−0.971)	0.046
前白蛋白	−0.010	0.002	0.990(0.987−0.993)	<0.001
国际标准化比值	0.325	0.121	1.384(1.099−1.768)	0.007
总胆红素	0.003	0.001	1.003(1.002−1.004)	<0.001
总蛋白	−0.016	0.006	0.985(0.972−0.997)	0.015
尿酸	0.002	0.000	1.002(1.001−1.003)	<0.001
MELD	0.026	0.014	1.026(0.997−1.055)	0.075

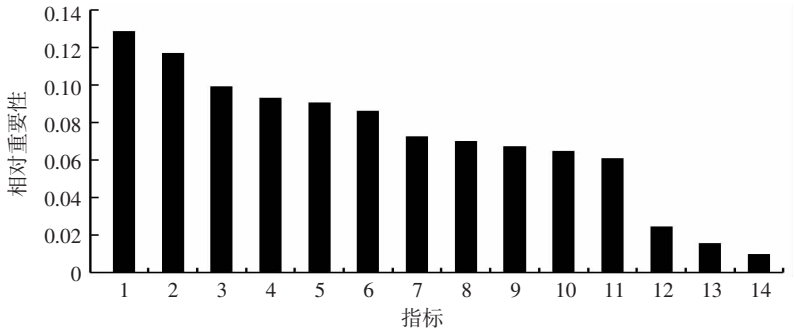
表 4 决策树模型参数分析

复杂度参数	分支数	预测估计误差	交叉验证误差	交叉验证误差的标准差
0.049	0	1.000	1.000	0.038
0.042	2	0.901	0.924	0.036
0.010	3	0.859	0.891	0.036
0.010	5	0.839	0.923	0.036

2.3 模型预测效能比较

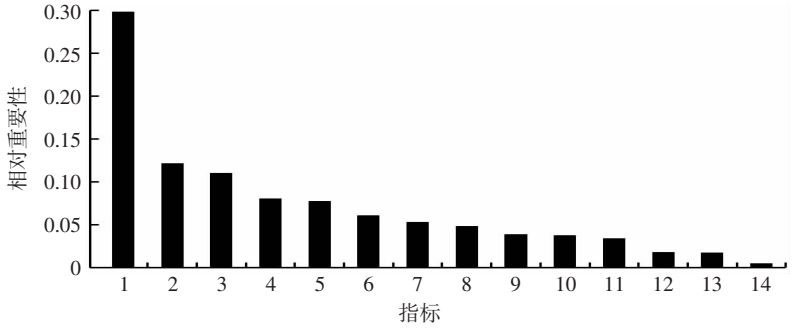
进一步将上述 4 个模型中均具有统计学差异的 7 项指标(是否是失代偿期肝硬化、C 反应蛋白是否异常、淋巴细胞百分比、前白蛋白、国际标准化比值、MELD、总胆红素)提取出来,构建优化 Logistic 回归模型,见表 5。并在内部验证集和外部验证集中对比优化 Logistic 回归模型与其余 4 个模型的性能。由于只进行内部验证会导致模型的外推性不足,因此利用 B 院的数据进行外部验证,以提高模型的可信度[18-20]。

将最终的 Logistic 回归模型、随机森林模型、决策树模型、XGBoost 模型以及优化 Logistic 进行内部验证和外部验证。内部验证结果显示 Logistic 回归模型、随机森林模型、XGBoost 模型和简化 Logistic 模型 AUC 较高,分别为 0.866、0.869、0.873、0.860;决策树模型 AUC 最低,为 0.727。外部验证结果较内部验证结果有一定程度的降低。其余指标情况见表 6、7。内部验证和外部验证的 ROC 曲线见图 3。经由 delong test 检验可知优化 Logistic 模型与其余 3 个性能较好模型没有统计学差异($P>0.05$),见表 8,表明优化 Logistic 模型仍然保持着较好的诊断效能。优化 Logistic 模型所较其他模型涉及的指标减少了一半,临床使用更加方便快捷。



1: MELD; 2: 国际标准化比值; 3: 总胆红素; 4: 淋巴细胞百分比; 5: 前白蛋白; 6: 尿酸; 7: 白细胞计数; 8: 总蛋白; 9: 红细胞血红蛋白浓度; 10: 住院天数; 11: 淋巴细胞数; 12: 失代偿; 13: C 反应蛋白是否异常; 14: 乙肝

图 1 随机森林变量重要性测度指标



1: MELD; 2: 国际标准化比值; 3: 淋巴细胞百分比; 4: 前白蛋白; 5: 总胆红素; 6: 尿酸; 7: 住院天数; 8: 失代偿; 9: 白细胞计数; 10: 红细胞血红蛋白浓度; 11: 总蛋白; 12: 淋巴细胞数; 13: C 反应蛋白是否异常; 14: 乙肝

图 2 XGBoost 模型指标重要性

表 5 优化 Logistic 回归模型结果

指标	β	SE	OR(95%CI)	P
失代偿	1.678	0.185	5.354(3.770-7.803)	<0.001
淋巴细胞百分比	-0.051	0.006	0.951(0.939-0.962)	<0.001
总胆红素	0.003	0.001	1.003(1.002-1.004)	<0.001
C 反应蛋白是否异常	0.486	0.110	1.626(1.310-2.017)	<0.001
国际标准化比值	0.297	0.110	1.346(1.091-1.681)	0.007
前白蛋白	-0.010	0.001	0.990(0.987-0.993)	<0.001
MELD	0.038	0.012	1.038(1.015-1.063)	0.001

表 6 5 种模型在内部验证集中的性能比较

模型	灵敏度	特异度	阳性预测率	阴性预测率	准确度	AUC(95%*CI*)
Logistic 模型	0.900	0.678	0.318	0.976	0.710	0.866(0.844-0.889)
随机森林模型	0.818	0.762	0.364	0.962	0.770	0.869(0.847-0.892)
决策树模型	0.580	0.837	0.372	0.923	0.800	0.727(0.692-0.762)
XGBoost 模型	0.887	0.710	0.337	0.974	0.735	0.873(0.852-0.895)
优化 Logistic 模型	0.827	0.719	0.329	0.961	0.734	0.860(0.837-0.884)

表 7 5 种模型在外部验证集中的性能比较

模型	灵敏度	特异度	阳性预测率	阴性预测率	准确度	AUC(95%*CI*)
Logistic 模型	0.760	0.751	0.380	0.940	0.752	0.831(0.809-0.852)
随机森林模型	0.777	0.680	0.328	0.938	0.696	0.809(0.785-0.832)
决策树模型	0.364	0.921	0.481	0.878	0.828	0.645(0.619-0.672)
XGBoost 模型	0.829	0.658	0.328	0.950	0.687	0.819(0.797-0.841)
优化 Logistic	0.662	0.812	0.414	0.923	0.787	0.818(0.795-0.841)

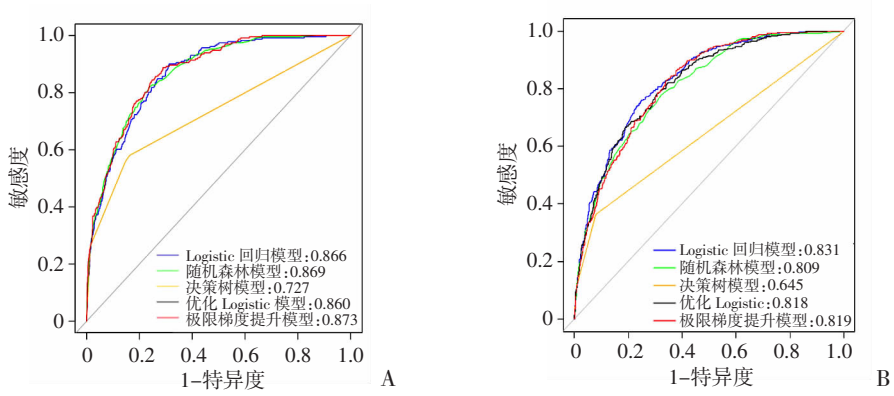


图 3 5 种模型内部验证(A)和外部验证(B)的 ROC 曲线

表 8 内部验证集中不同模型的 ROC 比较

统计值	ROC A	ROC B	ROC C	ROC D	A-D	B-D	C-D
AUC	0.866	0.869	0.873	0.860			
Z					1.550	1.083	1.897
P					0.121	0.279	0.058

A:Logistic 模型;B:随机森林模型;C:XGBoost 模型;D:优化 Logistic 模型

2.4 优化 Logistic 模型的校正度

采用 bootstrap 重抽样方法对优化 Logistic 模型进行 Calibration 曲线校正(图 4),45°对角线表示预测值与实际值完全重合参考线,Apparent 点线表示预测模型实际情况,Bias-corrected 实线表示校正后的预测模型实际情况。模型 C-index 为 0.853(95%*CI*:0.837-0.868),Brier 分数为 0.098(95%*CI*:0.092-0.105),提示诊断模型效能良好。

2.5 优化 Logistic 模型的列线图展示

以列线图来呈现优化 Logistic 模型(图 5)。列线图的透明度解决了机器学习模型结果不容易解释的问题。列线图根据模型中各个影响因素对结局变量的贡献程度(回归系数的大小),给每个影响因素的每个取值水平进行赋分。然后根据每个影响因素所在轴上的对应点作垂线对应至 Points 评分标尺,得到该影响因素得分,最后将各个影响因素对应得分相加得到 Total Points 值,最终 Total Points 值对应在肝硬化患者合并自发性腹膜炎风险轴(预测概率)上的点即为相对应的肝硬化患者合并自发性腹膜炎的发生风险。

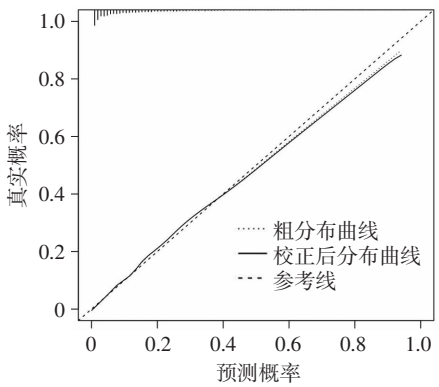


图 4 优化模型的校正曲线

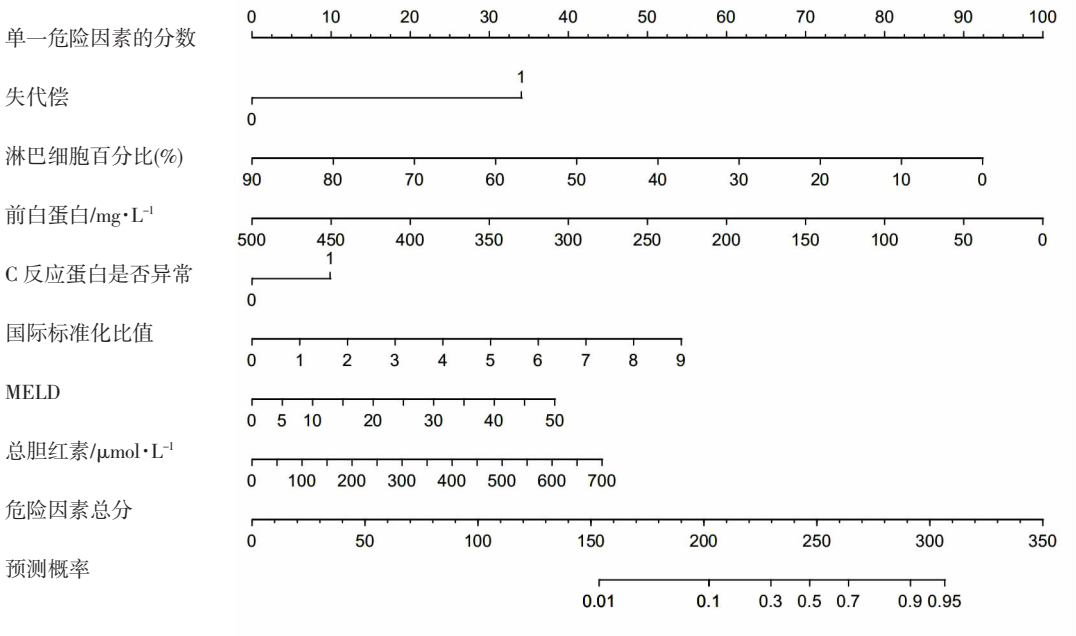


图5 优化 Logistic 模型列线图

3 讨论

肝硬化是慢性弥漫性肝损害的终末阶段,是各种致病的因子反复作用于肝脏,造成肝细胞变性、坏死、纤维组织增生及再生等病理变化[21],最终破坏肝脏的正常组织结构,导致肝组织弥漫性纤维化、假小叶及再生结节形成,导致肝脏形体异常、质地变硬。SBP 作为肝硬化的常见并发症,其早期诊断十分困难。临床上 SBP 诊断的金标准为腹水多核细胞数>250/mm^3,但腹水培养临床依从性差、检查费时(24~48 h)、阳性率低至 35%~65%[22],SBP 临床症状常常不典型,如不及时治疗,极易发展为感染性休克、肝肾综合征等严重并发症而危及生命[23]。系统的使用抗生素预防 SBP 会导致不合理的医源性或院内并发症,增加医疗费用,影响预后,因此早期识别或预测 SBP 的发生十分有必要。

指标失代偿(OR=5.354,95%CI: 3.770-7.803)、淋巴细胞百分比(OR=0.951,95%CI: 0.939-0.962)、总胆红素(OR=1.003,95%CI: 1.002-1.004)、C 反应蛋白是否异常(OR=1.626,95%CI: 1.310-2.017)、国际标准化比值(OR=1.346,95%CI: 1.091-1.681)、前白蛋白(OR=0.990,95%CI: 0.987-0.993)、MELD(OR=1.038,95%CI: 1.015-1.063)为显著影响因素。肝硬化患者失代偿期患病风险是普通的肝硬化患者的 5.354 倍。前白蛋白是一种相对分子量为 55×10^3 的急性相反应蛋白,参与机体免疫调节,起到维持酸碱平衡的作用。李维勤等[24]提出,在感染条件下,血清白蛋白从血管内到血管外的分布速率明显增加,白蛋白的分解速率也显著增加。王新颖等[25]提出,内毒素通过抑制细胞因子 IL-6 和 TNF-α 释放,间接抑制大鼠肝细胞白蛋白 mRNA 的表达。因此,炎症感染会影响到血清白蛋白的水平,低前白蛋白和白蛋白会增加肝硬化并发 SBP 的几率。生理状态下,前白蛋白主要存在于血清和脑脊液中,与白蛋白相比,其半衰期短(1.9 d)[26],短期升高和降低更为明显,检测血清前白蛋白水平对各种肝脏疾病包括失代偿期肝硬化的诊断、治疗和预后判断有重要临床价值[27-28],同样早期诊断预测肝硬化患者发生 SBP 具有更高的敏感性。多位学者曾报道总胆红素是肝硬化患者并发 SBP 的独立危险因素[11, 29-30],总胆红素升高反映了肝脏损伤及肝脏合成功能下降的情况,而且总胆红素瘀积可使 Kupffer 细胞功能受损,肝功能损害加重并伴有单核-巨噬细胞系统功能明显下降,容易发生细菌迁移及菌血症,可增加 SBP 发生的可能性。SBP 的发生进一步加重肝脏负担,两者之间相互影响、相互作用,因此肝硬化患者伴有明显的总胆红素水平升高时需要高度警惕 SBP 的发生。C 反应蛋白为急性期反应蛋白的一种,创伤及感染发生后,C 反应蛋白明显提升。将其应用到各类疾病的诊断中,能够辅助判断患者是否合并炎性反应,是否为细菌感染。肝硬化合并自发性腹膜炎发病后,患者通常可见急性炎性反应。发病期间,对 C 反应蛋白指标进行检验,可见该指标升高现象[31]。肝脏能合成除组织因子和钙离子外的所有凝血因子和纤溶因子,凝血因子是评估肝功能状态最为敏感的指

标,测定凝血因子的水平,能快速反映肝功能的状况。国际标准化比值延长提示肝功能降低,当出现国际标准化比值增大时应及时警惕 SBP 的发生。MELD 评分是 2001 年由 KAMATH 等[32]提出的新式评分体系,目前已被广泛应用于评价各种终末期肝病患者的病情严重程度及预后的判断,MELD 评分值越高患者病情越严重[33]。

综上所述,本研究采用了 4 种机器学习算法建立肝硬化伴自发性腹膜炎的风险预测模型,一定程度上减少了单一算法建模带来的偏移;优化 Logistic 模型使用较少指标,但较其他模型诊断性能差异无统计学意义,临床使用更加便捷;模型在内部验证和外部验证中表现良好,使其可推广到大量的患者群体中,建立的列线图可解释性更强,临床使用更加便捷。模型中包含的预测因子是肝硬化患者常见的临床可收集因素,使模型在日常实践中有用和适用。由于该研究为横断面研究,我们不能推断出这些重要影响因素与 SBP 发生的因果关系,需要后期开展前瞻性试验来验证这些因素与 SBP 发生的因果关系。

参考文献:

[1] BUNCHORNTAVAKUL C, CHAMROONKUL N, CHAVALITDHAMRONG D. Bacterial infections in cirrhosis: a critical review and practical guidance[J]. World J Hepatol, 2016, 8(6): 307-321. DOI: 10.4254/wjh.v8.i6.307.

[2] ALBILLOS A, LARIO M, ÁLVAREZ-MON M. Cirrhosis-associated immune dysfunction: distinctive features and clinical relevance[J]. J Hepatol, 2014, 61(6): 1385-1396. DOI: 10.1016/j.jhep.2014.08.010.

[3] 徐小元, 丁惠国, 李文刚, 等. 肝硬化腹水及相关并发症的诊疗指南[J]. 实用肝脏病杂志, 2018, 21(1): 21-31.
XU X Y, DING H G, LI W G, *et al*. Guidelines on the management of ascites and complications in cirrhosis[J]. J Pract Hepatol, 2018, 21(1): 21-31.

[4] BONNEL A R, BUNCHORNTAVAKUL C, REDDY K R. Immune dysfunction and infections in patients with cirrhosis[J]. Clin Gastroenterol Hepatol, 2011, 9(9): 727-738. DOI: 10.1016/j.cgh.2011.02.031.

[5] ARVANITI V, D'AMICO G, FEDE G, *et al*. Infections in patients with cirrhosis increase mortality four-fold and should be used in determining prognosis[J]. Gastroenterology, 2010, 139(4): 1246-1256, 1256.e1. DOI: 10.1053/j.gastro.2010.06.019.

[6] BAJAJ J S, O'LEARY J G, REDDY K R, *et al*. Second infections independently increase mortality in hospitalized patients with cirrhosis: the North American consortium for the study of end-stage liver disease (NACSELD) experience[J]. Hepatology, 2012, 56(6): 2328-2335. DOI: 10.1002/hep.25947.

[7] SINGAL A K, SALAMEH H, KAMATH P S. Prevalence and in-hospital mortality trends of infections among patients with cirrhosis: a nationwide study of hospitalised patients in the United States[J]. Aliment Pharmacol Ther, 2014, 40(1): 105-112. DOI: 10.1111/apt.12797.

[8] 李浩, 程璠, 郭东星, 等. 肝硬化患者并发自发性细菌性腹膜炎的高危特征与潜在风险分析[J]. 中国卫生统计, 2020, 37(5): 664-666.
LI H, CHENG F, GUO D X, *et al*. Latent class analysis of distribution characteristics in cirrhosis complicated by spontaneous bacterial peritonitis[J]. Chin J Heal Stat, 2020, 37(5): 664-666.

[9] 新华社. 中共中央国务院印发《"健康中国 2030"规划纲要》[EB/OL]. (2016-10-25)[2021-03-18]. http://www.gov.cn/zhengce/2016-10/25/content_5124174.htm.
Xinhua News Agency. State Council of China issued "Healthy China 2030" Planning Outline [EB/OL]. (2016-10-25)[2021-03-18]. http://www.gov.cn/zhengce/2016-10/25/content_5124174.htm.

[10] GOECKS J, JALILI V, HEISER L M, *et al*. How machine learning will transform biomedicine[J]. Cell, 2020, 181(1): 92-101. DOI: 10.1016/j.cell.2020.03.022.

[11] ZHU L C, XU L, HE W H, *et al*. A quick screening model for symptomatic bacterascites in cirrhosis[J]. Saudi J Gastroenterol, 2016, 22(4): 282-287. DOI: 10.4103/1319-3767.187600.

[12] 王苏淮, 李晶洁. 机器学习在心血管疾病中的临床应用进展[J]. 心血管病学进展, 2021, 42(2): 144-147. DOI: 10.16806/j.cnki.issn.1004-3934.2021.02.013.
WANG S H, LI J J. Clinical applications of machine learning in cardiovascular[J]. Adv Cardiovasc Dis, 2021, 42(2): 144-147. DOI: 10.16806/j.cnki.issn.1004-3934.2021.02.013.

[13] BZDOK D, KRZYWINSKI M, ALTMAN N. Machine learning: supervised methods[J]. Nat Methods, 2018, 15(1): 5-6. DOI: 10.1038/nmeth.4551.

[14] FELDMANN J, YOUNGBLOOD N, WRIGHT C D, *et al*. All-optical spiking neurosynaptic networks with self-learning capabilities[J]. Nature, 2019, 569(7755): 208-214. DOI: 10.1038/s41586-019-1157-8.

[15] 徐小元, 丁惠国, 李文刚, 等. 肝硬化诊治指南[J]. 临床肝胆病杂志, 2019, 35: 2408-2425.
XU X Y, DING H G, LI W G, *et al*. Chinese guidelines on the management of liver cirrhosis[J]. J Clin Hepatol, 2019, 35: 2408-2425.

[16] European Association for the Study of the Liver. EASL clinical practice guidelines on the management of ascites, sponta-

neous bacterial peritonitis, and hepatorenal syndrome in cirrhosis[J]. J Hepatol, 2010, 53(3): 397-417. DOI: 10.1016/j.jhep.2010.05.004.

[17] ZOPLUOGLU C. Detecting examinees with itempreknowledge in large-scale testing using extreme gradient boosting (XGBoost)[J]. Educ Psychol Meas, 2019, 79(5): 931-961. DOI: 10.1177/0013164419839439.

[18] 王俊峰, 章仲恒, 周支瑞, 等. 临床预测模型: 模型的验证[J]. 中国循证心血管医学杂志, 2019,11(2): 141-144.
WANG J F, ZHANG Z H, ZHOU Z R, *et al*. Clinical prediction models: Model validation[J]. Chin J Evid Based Cardiovasc Med, 2019, 11(2): 141-144.

[19] STEYERBERG E W, HARRELL F E. Prediction models need appropriate internal, internal-external, and external validation[J]. J Clin Epidemiol, 2016, 69: 245-247. DOI: 10.1016/j.jclinepi.2015.04.005.

[20] ZASTROW S, KRABBE L M, WOLFF I, *et al*. External validation of a postoperative nomogram for the prediction of disease-specific survival in patients with papillary renal cell carcinoma using a large multicenter database[J]. Int J Clin Oncol, 2020, 25(1): 145-150. DOI: 10.1007/s10147-019-01530-x.

[21] POCA M, CONCEPCIÓN M, CASAS M, *et al*. Role of albumin treatment in patients with spontaneous bacterial peritonitis[J]. Clin Gastroenterol Hepatol, 2012, 10(3): 309-315. DOI: 10.1016/j.cgh.2011.11.012.

[22] SHIZUMA T. Spontaneous bacterial and fungal peritonitis in patients with liver cirrhosis: A literature review[J]. World J Hepatology, 2018,10(2): 254-266. DOI: 10.4254/wjh.v10.i2.254.

[23] 张天洪, 陈婧, 万雪梅. 肝硬化并自发性腹膜炎96例临床诊治探讨[J]. 西南医科大学学报, 2018,41(5): 436-440.
ZHANG T H, CHEN J, WAN X M. Clinical diagnosis and treatment of liver cirrhosis complicated by spontaneous peritonitis: An analysis of 96 cases[J]. J Xinan Med Univ, 2018, 41(5): 436-440.

[24] 李维勤, 王新颖, 朱虹, 等. 严重感染患者血清白蛋白动力学研究[J]. 解放军医学杂志, 2005, 30(11): 978-980. DOI: 10.3321/j.issn: 0577-7402.2005.11.013.
LI W Q, WANG X Y, ZHU H, *et al*. Kinetics of serum albumin in patients with severe Sepsis[J]. Med J PLA, 2005, 30(11): 978-980. DOI: 10.3321/j.issn: 0577-7402.2005.11.013.

[25] 王新颖, 李维勤, 李宁, 等. 内毒素抑制肝细胞白蛋白表达的分子机制研究[J]. 中国危重病急救医学, 2004(10): 626-628. DOI: 10.3760/j.issn: 1003-0603.2004.10.015.
WANG X Y, LI W Q, LI N, *et al*. Study on molecular mechanisms of indirect suppression of albumin mRNA expression of hepatocytes by lipopolysaccharide in rat[J]. Chin Crit Care Med, 2004(10): 626-628. DOI: 10.3760/j.issn: 1003-0603.2004.10.015.

[26] LIU F, CAI L Y, ZHONG L, *et al*. Model for end-stage liver disease combined with serum prealbumin to predict the prognosis of patients with decompensated liver cirrhosis[J]. J Dig Dis, 2010, 11(6): 352-357. DOI: 10.1111/j.1751-2980.2010.00465.x.

[27] 杜小华. 血清前白蛋白检测在肝脏疾病诊断中的应用[J]. 中国继续医学教育, 2018, 10(18): 53-54.
DU X H. The application of prealbumin in the diagnosis of liver disease[J]. China Continu Med Educ, 2018, 10(18): 53-54.

[28] TAN L, MENG Y, ZENG T, *et al*. Clinical diagnostic significance of prealbumin, cholinesterase and retinol binding protein in liver cirrhosis combined with encephalopathy[J]. Br J Biomed Sci, 2019, 76(1): 24-28. DOI: 10.1080/09674845.2018.1523673.

[29] 申红霞, 娄小平, 苌新伟, 等. 肝硬化患者自发性细菌性腹膜炎危险因素的Meta分析[J]. 世界华人消化杂志, 2016, 24(12): 1903-1909.
SHEN H X, LOU X P, CHANG X W, *et al*. Risk factors for spontaneous bacterial peritonitis in cirrhotic patients: A meta analysis[J]. World Chin J Dig, 2016, 24(12): 1903-1909.

[30] PAUL K, KAUR J, KAZAL H L. To study the incidence, predictive factors and clinical outcome of spontaneous bacterial peritonitis in patients of cirrhosis with ascites[J]. J Clin Diagn Res, 2015, 9(7): OC09-OC12. DOI: 10.7860/jcdr/2015/14855.6191.

[31] 林太顺, 林清标, 孔金峰, 等. 降钙素原联合C反应蛋白检测在肝衰竭合并自发性腹膜炎疾病中的诊断价值[J]. 临床合理用药杂志, 2019, 12(14): 5-6. DOI: 10.15887/j.cnki.13-1389/r.2019.14.003.
LIN T S, LIN Q B, KONG J F, *et al*. Diagnostic value of procalcitonin combined with C-reactive protein in liver failure complicated with spontaneous peritonitis[J]. Chin J Clin Ration Drug Use, 2019,12(14): 5-6. DOI: 10.15887/j.cnki.13-1389/r.2019.14.003.

[32] KAMATH P S, WIESNER R H, MALINCHOC M, *et al*. A model to predict survival in patients with end-stage liver disease[J]. Hepatology, 2001, 33(2): 464-470. DOI: 10.1053/jhep.2001.22172.

[33] YANG W B, CHEN E Q, BI H X, *et al*. Different models in predicting the short-term prognosis of patients with hepatitis B virus-related acute-on-chronic liver failure[J]. Ann Hepatol, 2012,11(3): 311-319. DOI: 10.1016/S1665-2681(19)30925-1.

(**收稿**:2021-04-16;**修回**:2021-07-05)

(编辑 邓强庭)