

## 论 著

# 基于机器学习算法的肝硬化相关肝性脑病预测模型的构建

谈军涛<sup>1</sup>, 许晓梅<sup>2</sup>, 何雨芯<sup>3</sup>, 谭超<sup>1</sup>, 龚军<sup>1</sup>, 刘蕴宇<sup>1</sup>, 向守书<sup>1</sup>, 赵文龙<sup>1\*</sup>

<sup>1</sup>重庆医科大学医学信息学院, 重庆 400016; <sup>2</sup>成都市第五人民医院消化内科, 成都 611130; <sup>3</sup>重庆医科大学公共卫生与管理学院, 重庆 400016

**[摘要]** **目的** 构建基于机器学习算法的肝硬化相关肝性脑病预测模型。**方法** 采用横断面研究方法, 收集2019年6月—2020年6月于重庆市7家医疗机构就诊的1498例肝硬化患者的病例资料, 依据是否发生肝性脑病分为肝性脑病组( $n=285$ )与非肝性脑病组( $n=1213$ )。随机抽取70%(1048例)患者的病例资料作为训练集构建预测模型, 其余30%(450例)作为测试集进行内部验证。采用单因素logistic回归筛选输入指标, 应用基于机器学习的logistic回归、随机森林、决策树和XGBoost算法构建预测模型, 比较4种方法构建的模型对肝硬化相关肝性脑病的诊断价值。**结果** Logistic回归、随机森林、决策树和XGBoost模型均提示, 凝血酶原活动度( $OR=0.933$ , 95%CI 0.921~0.946)、年龄( $OR=1.045$ , 95%CI 1.029~1.061)、全血钠( $OR=0.964$ , 95%CI 0.928~1.000)及尿素氮( $OR=1.063$ , 95%CI 1.022~1.105)是肝硬化相关肝性脑病的重要影响因素。Logistic回归、随机森林、决策树和XGBoost模型诊断肝硬化相关肝性脑病的灵敏度分别为0.843、0.904、0.759、0.892, 特异度分别为0.785、0.695、0.717、0.706, AUC分别为0.875、0.883、0.767、0.847。**结论** 基于机器学习算法建立的肝硬化相关肝性脑病风险预测模型具有较高的诊断价值, 其中logistic回归模型和随机森林模型的诊断效能优于决策树模型和XGBoost模型。

**[关键词]** 肝硬化; 肝性脑病; 机器学习; 预测模型

**[中图分类号]** R575.2

**[文献标志码]** A

**[文章编号]** 0577-7402(2021)04-0354-07

**[DOI]** 10.11855/j.issn.0577-7402.2021.04.06

## Construction of prediction model of cirrhosis-related hepatic encephalopathy based on machine learning algorithm

Tan Jun-Tao<sup>1</sup>, Xu Xiao-Mei<sup>2</sup>, He Yu-Xin<sup>3</sup>, Tan Chao<sup>1</sup>, Gong Jun<sup>1</sup>, Liu Yun-Yu<sup>1</sup>, Xiang Shou-Shu<sup>1</sup>, Zhao Wen-Long<sup>1\*</sup>

<sup>1</sup>School of Medical Informatics, Chongqing Medical University, Chongqing 400016, China

<sup>2</sup>Department of Gastroenterology, Chengdu Fifth People's Hospital, Chengdu 611130, China

<sup>3</sup>School of Public Health and Management, Chongqing Medical University, Chongqing 400016, China

\*Corresponding author, E-mail: cqzhaowl@163.com

This work was supported by the Project of Smart Medicine of Chongqing Medical University (ZHYX2019001)

**[Abstract]** **Objective** To construct a prediction model based on machine learning algorithm for cirrhosis-related hepatic encephalopathy. **Methods** A cross-sectional survey was conducted in 1498 patients from 7 medical institutions in Chongqing from June 2019 to June 2020, who were divided into hepatic encephalopathy group ( $n=285$ ) and non-hepatic encephalopathy group ( $n=1213$ ) according to whether hepatic encephalopathy occurred. 70% (1048 in total) of the data collected from 1498 patients was randomly chosen as the training set for establishing the prediction model and the rest 30% (450 in total) was used for internal verification. Univariate logistic regression was used to filter input indicators. Logistic regression, random forest, decision tree and XGBoost algorithm based on machine learning were used to construct a diagnostic predictive model. The models constructed by the four methods were compared for predictive and diagnostic value of cirrhosis-related hepatic encephalopathy. **Results** Logistic regression, random forest, decision tree and XGBoost models all suggested prothrombin activity ( $OR=0.933$ , 95%CI 0.921-0.946), age ( $OR=1.045$ , 95%CI 1.029-1.061), blood sodium ( $OR=0.964$ , 95%CI 0.928-1.000) and urea nitrogen ( $OR=1.063$ , 95%CI 1.022-1.105) are important influencing factors of hepatic encephalopathy. The sensitivities of the four models were 0.843, 0.904, 0.759 and 0.892; the specificities were 0.785, 0.695, 0.717 and 0.706; the area under the curve (AUC) were 0.875, 0.883, 0.767 and 0.847 respectively. **Conclusions** The risk prediction model of cirrhosis-related hepatic encephalopathy established based on machine learning has high diagnostic value. The diagnostic effects of logistic and random forest models are better than those of decision tree

**[基金项目]** 重庆医科大学智慧医学项目(ZHYX2019001)

**[作者简介]** 谈军涛, 硕士研究生, 主要从事电子病历数据挖掘方面的研究

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

and XGBoost models.

**[Key words]** liver cirrhosis; hepatic encephalopathy; machine learning; prediction models

肝性脑病(hepatic encephalopathy)是严重肝脏疾病并发的一种以代谢紊乱为基础的中枢神经系统功能性失调的综合征,患者主要表现为意识障碍、行为失常甚至昏迷。美国肝病研究协会和欧洲肝病研究协会将肝性脑病定义为由肝损伤或门体分流引起的脑功能障碍<sup>[1]</sup>。临床上将肝性脑病分为A、B、C三型,其中A型与急性肝衰竭有关,B型与无肝病的门体分流有关,C型与肝硬化有关<sup>[2]</sup>。肝性脑病除增加患者、家庭和社会的各种负担外,也是肝硬化患者发生跌倒、骨折、再入院等不良事件的常见原因之一<sup>[3]</sup>。为减少肝性脑病带来的不良影响,有必要对肝硬化患者进行肝性脑病风险预测,以帮助医护人员进行评估并提前采取治疗和护理措施。机器学习作为一种新兴的统计分析方法,能够对大数据进行深度挖掘与分析,目前已在疾病发生、预后预测等方面广泛应用<sup>[4-5]</sup>。本研究采用logistic回归、随机森林、决策树和XGBoost等4种机器学习算法,构建并评价肝硬化相关肝性脑病风险预测模型,旨在为肝硬化患者肝性脑病风险评估奠定基础,并为临床防治工作提供参考。

## 1 资料与方法

**1.1 研究对象** 本研究为回顾性横断面研究。收集2019年6月—2020年6月就诊于重庆市7家医疗机构(其中5家为重庆医科大学附属第一医院,2家为教学指导医院)消化内科、肝胆外科、感染科等科室的1498例肝硬化患者。纳入标准:肝硬化失代偿期;年龄>18岁。排除标准:合并恶性肿瘤;合并肝衰竭;既往曾行经颈静脉肝内门-体静脉支架分流术(transjugular intrahepatic portosystemic shunt, TIPS);伴有精神疾病;临床资料不完整。

**1.2 分组** 参照《肝硬化肝性脑病诊疗指南》<sup>[6]</sup>,结合临床表现、实验室和辅助检查判断患者是否发生肝性脑病,并据此将1498例患者分为肝性脑病组( $n=285$ )与非肝性脑病组( $n=1213$ )。将肝性脑病组和非肝性脑病组按照7:3随机分为训练集( $n=1048$ )和测试集( $n=450$ ),进行内部验证。本研究通过重庆医科大学医学研究伦理委员会审批。

**1.3 研究指标** 患者的一般资料(年龄、性别、吸烟史、饮酒史、高血压、糖尿病)、病因(乙肝、丙肝、酒精肝、自身免疫性肝病)、并发症(消化道出血、腹膜炎、腹水)及实验室检查(血常规、肝功能、肾功能、电解质、凝血功能)等。

**1.4 统计学处理** 采用SPSS 22.0、R4.0.2、Excel

2013软件进行统计分析。符合正态分布的计量资料以 $\bar{x}\pm s$ 表示,组间比较采用 $t$ 检验;不符合正态分布的计量资料以 $M(Q_1, Q_3)$ 表示,组间比较采用Mann-Whitney  $U$ 检验。计数资料以率(%)表示,组间比较采用 $\chi^2$ 检验。内部验证采用单因素logistic回归分析,以 $P<0.05$ 为纳入多因素分析的标准,将筛选后的变量作为输入变量,以是否发生肝性脑病为结局变量,在训练集中分别建立logistic回归、随机森林(random forest)、决策树(decision tree)和XGBoost模型;然后在测试集中采用ROC曲线分析比较4种模型的预测价值。

采用R语言中的glmnet包、random forest包、rpart包及XGBoost包分别构建4种机器学习模型。随机森林模型主要包含ntree(树的数目)和mtry(随机选择特征的数目)两个重要参数,其中mtry在通常情况下为 $\lceil \log_2(p)+1 \rceil$ 个, $p$ 为纳入指标个数<sup>[7]</sup>,模型采用准确度平均下降量(mean decrease accuracy, MDA)衡量指标的重要度。决策树模型分裂属性的度量指标为信息熵<sup>[8]</sup>,采用后剪枝过程中的最小代价复杂度参数(CP)进行模型优化。XGBoost模型采用gain值衡量某个指标的增益<sup>[9]</sup>,gain值越大,表明指标对模型的影响越大。

## 2 结果

**2.1 基线资料** 两组患者年龄、吸烟史、饮酒史、高血压、糖尿病、乙肝、酒精肝、自身免疫性肝病、腹膜炎、总胆红素、中性粒细胞比值、血红蛋白总量、全血钠、全血钾、白蛋白、尿酸、红细胞计数、淋巴细胞比值、凝血酶原活动度、尿素氮、丙氨酸氨基转移酶及白细胞计数等差异有统计学意义( $P<0.05$ ,表1)。

**2.2 构建机器学习模型** Logistic回归分析结果显示,高血压、酒精肝、糖尿病、自身免疫性肝病、年龄、全血钠、凝血酶原活动度、尿素氮等10项指标为肝性脑病的影响因素( $P<0.05$ ,表2),其中淋巴细胞比值( $P=0.123$ )与红细胞计数( $P=0.100$ )通过咨询临床专家和查阅参考文献,也属于肝性脑病的影响因素,故纳入模型。如图1所示,当ntree超过500后,模型趋于稳定,最终随机森林模型参数设置为:ntree=500, mtry=5。各指标对应的MDA如表3所示。在分裂次数为4次时,决策树模型的交叉验证预测误差最小,为0.936(表4),此时对应的CP值为0.029,模型中各个分裂点对应的指标如图2所示,其中凝血酶原活动度为第一个分裂点,年龄为

表1 肝性脑病组与非肝性脑病组患者基线资料比较

Tab.1 Comparison of the baseline data between hepatic encephalopathy group and control group

指标	肝性脑病组(n=285)	非肝性脑病组(n=1213)	$\chi^2/t/Z$	P
年龄(岁, $\bar{x}\pm s$ )	63.0 ± 12.2	56.9 ± 12.2	7.579	<0.001
男性[例(%)]	201(70.5)	817(67.4)	0.926	0.336
吸烟史[例(%)]	149(52.3)	530(43.7)	6.525	0.011
饮酒史[例(%)]	151(53.0)	509(42.0)	10.929	<0.001
高血压[例(%)]	64(22.5)	142(11.7)	21.587	<0.001
糖尿病[例(%)]	67(23.5)	191(15.7)	9.218	0.002
乙肝[例(%)]	132(46.3)	729(60.1)	17.378	<0.001
丙肝[例(%)]	16(5.6)	59(4.9)	0.138	0.710
酒精肝[例(%)]	84(29.5)	121(10.0)	72.637	<0.001
自身免疫性肝病[例(%)]	37(13.0)	101(8.3)	5.438	0.020
消化道出血[例(%)]	61(21.4)	232(19.1)	0.622	0.430
腹膜炎[例(%)]	91(31.9)	189(15.6)	39.517	<0.001
腹水[例(%)]	13(4.6)	41(3.4)	0.618	0.432
总蛋白(g/L, $\bar{x}\pm s$ )	64.93 ± 10.33	66.22 ± 9.89	-1.906	0.057
总胆红素[ $\mu\text{mol/L}$ , $M(Q_1, Q_3)$ ]	43.60(25.50, 77.70)	25.00(15.90, 49.00)	-8.496	<0.001
中性粒细胞比值(% , $\bar{x}\pm s$ )	70.96 ± 12.63	65.87 ± 13.15	6.062	<0.001
血小板计数[ $\times 10^9/\text{L}$ , $M(Q_1, Q_3)$ ]	65.00(43.00, 112.00)	71.00(49.00, 109.00)	-1.575	0.115
血红蛋白总量(g/L, $\bar{x}\pm s$ )	95.88 ± 26.93	106.75 ± 28.54	-6.052	<0.001
天门冬氨酸氨基转移酶[U/L, $M(Q_1, Q_3)$ ]	48.00(33.00, 70.00)	50.00(31.00, 94.00)	-1.066	0.286
全血钠(mmol/L, $\bar{x}\pm s$ )	137.00 ± 6.29	139.27 ± 4.37	-5.773	<0.001
全血钾(mmol/L, $\bar{x}\pm s$ )	4.00 ± 0.78	3.89 ± 0.62	2.091	0.037
全血磷(mmol/L, $\bar{x}\pm s$ )	1.07 ± 0.38	1.07 ± 0.27	-1.802	0.071
全血氯(mmol/L, $\bar{x}\pm s$ )	103.94 ± 7.63	104.30 ± 5.16	0.756	0.450
全血钙(mmol/L, $\bar{x}\pm s$ )	2.09 ± 0.22	2.07 ± 0.23	-1.200	0.231
全血镁(mmol/L, $\bar{x}\pm s$ )	0.80 ± 0.14	0.79 ± 0.11	-1.263	0.207
白蛋白(g/L, $\bar{x}\pm s$ )	28.88 ± 4.63	32.52 ± 6.33	-11.042	<0.001
尿酸( $\mu\text{mol/L}$ , $\bar{x}\pm s$ )	345.65 ± 185.08	303.73 ± 128.68	-3.623	<0.001
红细胞计数( $\times 10^{12}/\text{L}$ , $\bar{x}\pm s$ )	3.11 ± 0.84	3.56 ± 0.86	8.046	<0.001
淋巴细胞比值(% , $M(Q_1, Q_3)$ ]	17.90(10.42, 23.54)	23.10(14.92, 30.30)	-7.229	<0.001
凝血酶原活性度(% , $\bar{x}\pm s$ )	51.47 ± 14.91	69.82 ± 17.68	-17.988	<0.001
尿素氮[mmol/L, $M(Q_1, Q_3)$ ]	6.96(4.89, 10.40)	5.32(4.17, 7.27)	-7.504	<0.001
碱性磷酸酶[U/L, $M(Q_1, Q_3)$ ]	115.46(87.00, 162.00)	109.00(80.00, 154.00)	-1.788	0.074
丙氨酸氨基转移酶[U/L, $M(Q_1, Q_3)$ ]	28.15(20.80, 44.00)	35.00(21.00, 70.00)	-4.007	<0.001
白细胞计数[ $\times 10^9/\text{L}$ , $M(Q_1, Q_3)$ ]	4.61(3.24, 7.38)	4.10(2.88, 5.88)	-3.845	<0.001

表2 Logistic回归分析肝性脑病的影响因素

Tab.2 Influencing factors of hepatic encephalopathy (logistic regression model)

指标	$\beta$	SE	OR	95%CI	P
高血压	0.850	0.263	2.340	1.397~3.918	0.001
酒精肝	1.173	0.232	3.232	2.051~5.092	<0.001
糖尿病	0.550	0.241	1.733	1.081~2.780	0.022
自身免疫性肝病	0.641	0.324	1.898	1.006~3.582	0.048
年龄	0.044	0.008	1.045	1.029~1.061	<0.001
全血钠	-0.037	0.019	0.964	0.928~1.000	0.045
凝血酶原活动度	-0.069	0.007	0.933	0.921~0.946	<0.001
尿素氮	0.061	0.020	1.063	1.022~1.105	0.003
淋巴细胞比值	-0.016	0.010	0.984	0.965~1.004	0.123
红细胞计数	-0.200	0.122	0.819	0.645~1.040	0.100

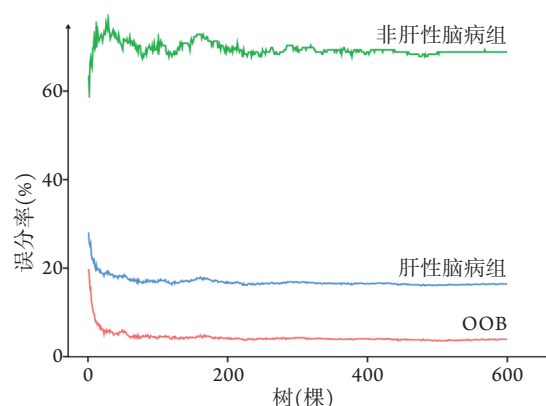


图1 两组肝病患者随机森林模型OOB趋势  
Fig.1 OOB trend of random forest model  
OOB. 袋外数据

表3 随机森林模型的指标重要性测度

**Tab.3** Index importance of random forest model

指标	MDA (%)
凝血酶原活动度	46.52
全血钠	26.76
年龄	24.01
白蛋白	20.76
尿酸	20.19
尿素氮	20.09
总胆红素	19.72
白细胞计数	18.17
全血钾	17.87
丙氨酸氨基转移酶	17.52
红细胞计数	17.47
淋巴细胞比值	17.30
中性粒细胞比值	16.07
血红蛋白总量	15.26
酒精肝	8.02
高血压	4.31
乙肝	3.15
饮酒史	2.96
腹膜炎	2.92
糖尿病	2.65
吸烟史	2.54
自身免疫性肝病	2.16

MDA. 模型准确度平均下降量

表4 决策树模型参数分析

**Tab.4** Analysis of decision tree model parameters

分裂次数	预测误差估计	预测误差标准误	交叉验证预测误差	CP值
0	1.000	0.063	1.000	0.054
2	0.891	0.064	1.030	0.043
4	0.822	0.062	0.936	0.029
6	0.772	0.062	0.970	0.024
9	0.703	0.063	0.975	0.019
14	0.629	0.063	0.990	0.012

CP. 最小代价复杂度参数

第二个，其余两个分裂点为白细胞计数和淋巴细胞比值。XGBoost模型中各指标对应的gain值如图3所示，凝血酶原活动度的gain值最大，其次为全血钠、年龄、全血钾、白细胞计数和尿素氮等。

**2.3 各模型的预测性能比较** 将构建的logistic回归模型、随机森林模型、决策树模型及XGBoost模型在测试集中进行内部验证，结果显示，各模型的AUC均较高，依次为0.875、0.883、0.767、0.847。Delong test检验结果显示，logistic回归模型和随机森林模型的预测性能优于决策树模型和XGBoost模型( $P < 0.05$ )。随机森林模型的灵敏度最高，为0.904，决策树模型最低，为0.759；logistic回归模

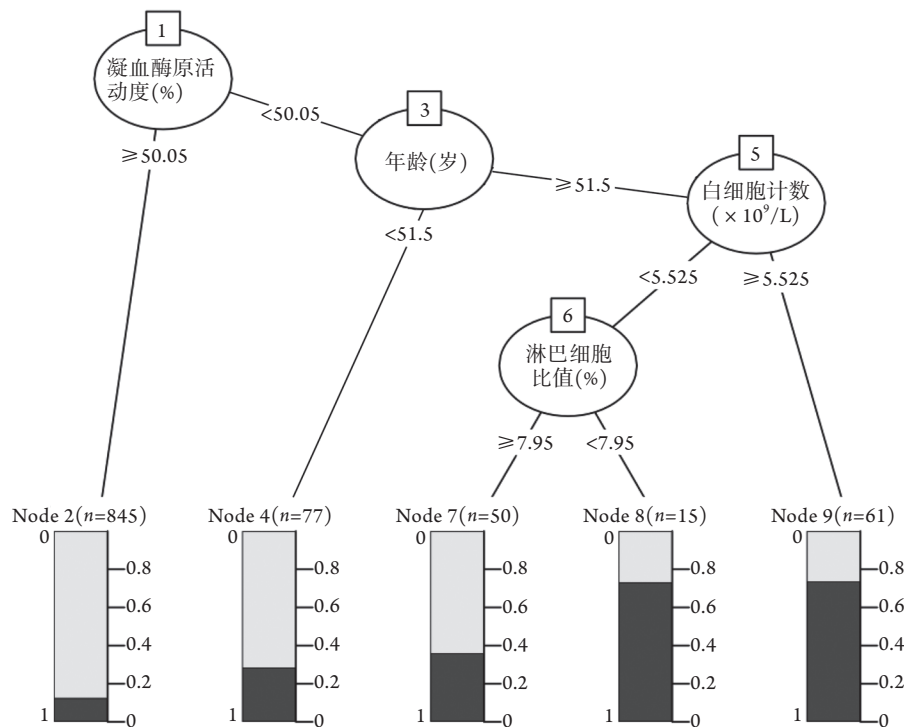


图2 决策树模型的可视化分析

Fig.2 Visual analysis of decision tree model

型的特异度最高，为0.785，随机森林模型最低，为0.695。随机森林模型的综合预测效能最优，其AUC最高，为0.883(图4、表5)。

**2.4 肝硬化相关肝性脑病重要影响因素分析** 对比4种模型，综合OR值、MDA值、Gain值发现，凝

血酶原活动度、年龄、全血钠及尿素氮在各自模型中均表现显著，提示这几个指标可作为肝硬化相关肝性脑病的重要影响因素。进一步绘制箱线图，并添加各指标显著性程度进行对比，结果显示，肝性脑病组年龄和尿素氮明显高于非肝性脑病组，++

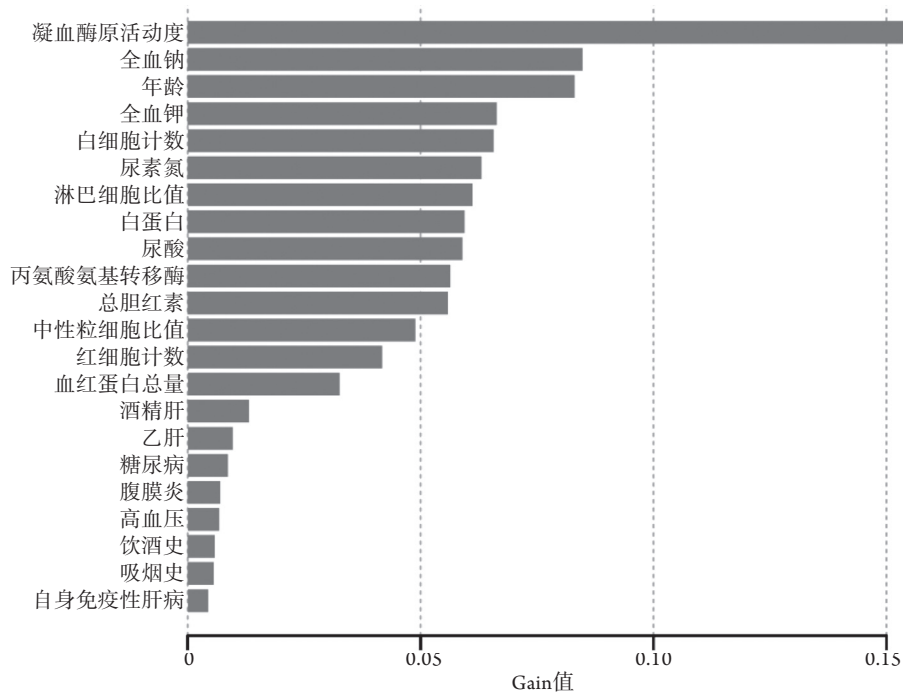


图3 XGBoost模型指标的重要性分析

Fig.3 Importance analysis of indexes in XGBoost model

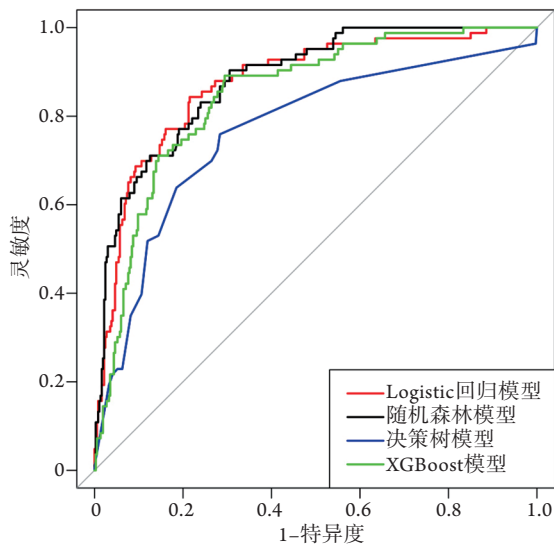


图4 各模型在测试集中的ROC曲线分析

Fig.4 ROC curves of four models in test set

表5 各模型在测试集中的预测性能比较

Tab.5 Comparison of the prediction performance of four models in test set

模型	灵敏度	特异度	阳性预测率	阴性预测率	准确率	AUC
Logistic回归模型	0.843	0.785	0.470	0.957	0.796	0.875
随机森林模型	0.904	0.695	0.401	0.970	0.733	0.883
决策树模型	0.759	0.717	0.378	0.929	0.724	0.767
XGBoost模型	0.892	0.706	0.407	0.966	0.740	0.847

AUC. 曲线下面积

全血钠和凝血酶原活动度明显低于非肝性脑病组 ( $P < 0.05$ , 图5)。

### 3 讨论

肝性脑病在肝硬化患者中具有发生率高、病死率高的特点，且其发病机制复杂，目前尚无标准的诊断方法<sup>[10]</sup>，因此，在肝性脑病发生前进行早期评估与预防具有积极意义。既往关于肝性脑病的风险研究局限于危险因素、预后以及神经生理、心理测试筛查方法的探究<sup>[11-12]</sup>，仅少数为基于危险因素的预测模型研究<sup>[13-14]</sup>。本研究结合患者的一般资料、病因、并发症和实验室检查结果，采用logistic回归、随机森林、决策树和XGBoost算法构建预测模型，所建模型在灵敏度、特异度和AUC等评价指标上表现良好，能够辅助临床医师预测肝硬化患者肝性脑病的发生风险。

机器学习可对大量数据的特征进行有效学习，为精准预测提供了新的研究思路和方法。机器学习算法包括常规算法(K-近邻、决策树和支持向量机等)和集成算法(随机森林、XGBoost和极限树等)。已有专家学者基于机器学习算法在医疗领域进行探索，取得了一定的成果<sup>[15-17]</sup>。本研究采用logistic回归、随机森林、决策树和XGBoost 4种机器学习算法分别构建预测模型以预测肝硬化患者并发肝性脑病的风险。Delong test检验显示，logistic回归和随机森林两种算法构建的预测模型优于决策树和XGBoost，其中综合预测效能最优的为随机森林模型，其AUC最高，为0.883。随机森林是多个决策树的集合，能弥补决策树泛化能力弱的缺点<sup>[18]</sup>，该

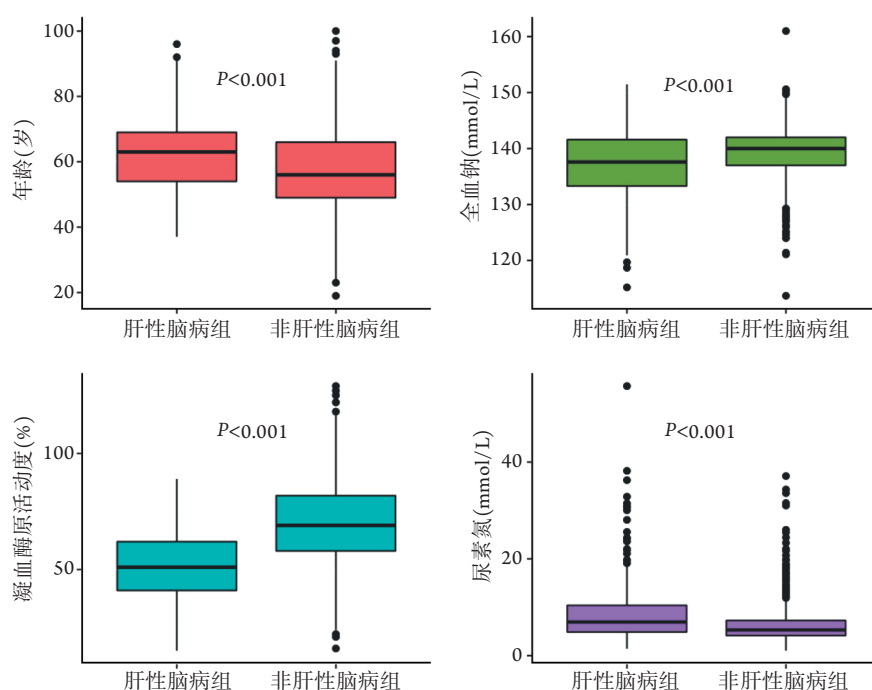


图5 显著性指标分组箱线图

Fig.5 Grouped box plot of significance index

方法依靠计算机,通过最小化观测结果和预测结果之间的误差来学习变量之间所有复杂的非线性相互作用<sup>[19]</sup>,使用bootstrap聚合和预测因子随机化来获得较高的疾病预测准确率<sup>[20-21]</sup>。

凝血酶原活动度是判断肝病严重程度最经典的指标,其水平降低常提示患者的肝功能被不同程度地破坏<sup>[22]</sup>。多项研究发现,凝血酶原活动度是肝性脑病的独立危险因素<sup>[23-24]</sup>。随着年龄增加,患者肝脏解毒能力变差,脑细胞更容易受到毒性作用,进而导致肝性脑病的发生风险增加。Routhu等<sup>[25]</sup>发现,高龄是肝硬化患者发生肝性脑病的重要影响因素。尿素氮水平升高是肾脏有效血容量不足的敏感指标。肝衰竭伴有肾功能受损时,会导致氮质血症,表现为血尿素氮水平升高,诱发脑水肿,从而导致肝性脑病发生风险升高。徐言等<sup>[26]</sup>在探究影响肝性脑病预后独立危险因素的研究中发现,除中性粒细胞/淋巴细胞比值外,尿素氮为第二影响因素。低钠血症是肝硬化患者常见的电解质紊乱之一,由于肝硬化患者存在肝功能衰退,因此容易引起心房钠尿肽、抗利尿激素、醛固酮等激素失调,导致钠、水潴留,并常合并低钠血症<sup>[27]</sup>。蒋汉梅等<sup>[28]</sup>在探讨失代偿期肝硬化患者血钠水平与病情、肝性脑病、肝肾综合征的关系时发现,低钠血症与失代偿期肝硬化患者的并发症和预后有关,血钠水平可作为患者病情和并发症的判断指标之一。本研究中的随机森林模型(AUC最高)提示,白蛋白是肝性脑病的重要影响因素。白蛋白是判断营养状态的良好

指标。最近研究发现,白蛋白是肝性脑病发生的独立危险因素<sup>[29]</sup>,另有研究发现,输注白蛋白有助于预防肝性脑病的发生,并可改善肝硬化患者肝性脑病的严重程度<sup>[30]</sup>。

综上所述,本研究构建了基于logistic回归、随机森林、决策树和XGBoost算法的肝性脑病预测模型,并结合灵敏度、特异度和AUC等评价指标对模型的效能进行对比,一定程度上减少了单一算法和单一评价指标带来的偏倚。同时本研究纳入的患者来自多家医疗机构,样本量大且具有良好的代表性,研究结果可为临床干预肝性脑病提供决策支持。但本研究仍存在一定的局限性:为横断面研究,无法确立肝性脑病风险与肝硬化之间的因果关系;模型只进行了内部验证,未进行外部验证,模型的准确性仍需今后通过更多的人群进行验证。

#### 【参考文献】

- [1] Vilstrup H, Amodio P, Bajaj J, *et al.* Hepatic encephalopathy in chronic liver disease: 2014 Practice Guideline by the American Association for the Study Of Liver Diseases and the European Association for the Study of the Liver[J]. *Hepatology*, 2014, 60(2): 715-735.
- [2] Davis BC, Bajaj JS. Effects of alcohol on the brain in cirrhosis: Beyond hepatic encephalopathy[J]. *Alcohol Clin Exp Res*, 2018, 42(4): 660-667.
- [3] Rassameehiran S, Mankongpaisarnrung C, Sutamtewagul G, *et al.* Predictor of 90-day readmission rate for hepatic encephalopathy[J]. *South Med J*, 2016, 109(6): 365-369.
- [4] Dong JY, Wei WQ, Wu K, *et al.* The application of machine learning in depression[J]. *Adv Psychol Sci*, 2020, 28(2): 266-

274. [董健宇, 韦文棋, 吴珂, 等. 机器学习在抑郁症领域的应用[J]. 心理科学进展, 2020, 28(2): 266-274.]
- [5] Miao FS, Li Y, Gao C, *et al.* Diabetes prediction method based on CatBoost algorithm[J]. *Comput Syst Appl*, 2019, 28(9): 215-218. [苗丰顺, 李岩, 高岑, 等. 基于CatBoost算法的糖尿病预测方法[J]. 计算机系统应用, 2019, 28(9): 215-218.]
- [6] Xu XY, Ding HG, Li WG, *et al.* Guidelines on the management of hepatic encephalopathy in cirrhosis[J]. *Chin J Liver Dis (Electr Version)*, 2018, 10(4): 17-32. [徐小元, 丁惠国, 李文刚, 等. 肝硬化肝性脑病诊疗指南[J]. 中国肝病杂志(电子版), 2018, 10(4): 17-32.]
- [7] Pudlo P, Marin JM, Estoup A, *et al.* Reliable ABC model choice via random forests[J]. *Bioinformatics*, 2016, 32(6): 859-866.
- [8] Xia XC, Wang XY. Improved C4.5 decision tree algorithm based on cosine similarity[J]. *Comput Eng Desig*, 2018, 39(1): 120-125. [夏修臣, 王秀英. 基于余弦相似度的改进C4.5决策树算法[J]. 计算机工程与设计, 2018, 39(1): 120-125.]
- [9] Zopluoglu C. Detecting examinees with item preknowledge in large-scale testing using extreme gradient boosting (XGBoost) [J]. *Educ Psychol Meas*, 2019, 79(5): 931-961.
- [10] Shawcross DL, Dunk AA, Jalan R, *et al.* How to diagnose and manage hepatic encephalopathy: a consensus statement on roles and responsibilities beyond the liver specialist[J]. *Eur J Gastroenterol Hepatol*, 2016, 28(2): 146-152.
- [11] Allampati S, Duarte-Rojo A, Thacker LR, *et al.* Diagnosis of minimal hepatic encephalopathy using stroop encephalApp: A multicenter US-based, norm-based study[J]. *Am J Gastroenterol*, 2016, 111(1): 78-86.
- [12] Yu BB, Jin KY. Analysis on risk factors of prognosis of hepatocirrhosis patients with hepatic encephalopathy[J]. *Chin J Hosp Stat*, 2019, 26(4): 246-248. [余卞卞, 金可吟. 影响肝硬化合并肝性脑病预后的危险因素分析[J]. 中国医院统计, 2019, 26(4): 246-248.]
- [13] Riggio O, Amodio P, Farcomeni A, *et al.* A model for predicting development of overt hepatic encephalopathy in patients with cirrhosis[J]. *Clin Gastroenterol Hepatol*, 2015, 13(7): 1346-1352.
- [14] Tapper EB, Parikh ND, Sengupta N, *et al.* A risk score to predict the development of hepatic encephalopathy in a population-based cohort of patients with cirrhosis[J]. *Hepatology*, 2018, 68(4): 1498-1507.
- [15] Ramroach S, Joshi A, John M. Optimisation of cancer classification by machine learning generates an enriched list of candidate drug targets and biomarkers[J]. *Mol Omics*, 2020, 16(2): 113-125.
- [16] Rao FY, Song YP, Lv XR, *et al.* Prediction of mortality among ischemic stroke patients one year after hospital discharge based on machine learning model[J]. *Chin J Publ Heal*, 2019, 35(9): 1187-1191. [饶夫阳, 宋艳平, 吕芯芮, 等. 基于机器学习模型缺血性脑卒中1年死亡预测效果评价[J]. 中国公共卫生, 2019, 35(9): 1187-1191.]
- [17] Wang M, Qin L, Wang CJ, *et al.* Machine learning-based models for prediction of intracerebral hemorrhage associated pneumonia[J]. *Chin J Stroke*, 2020, 15(3): 243-249. [王孟, 覃露, 王春娟, 等. 基于机器学习算法的脑出血相关肺炎预测模型研究[J]. 中国卒中杂志, 2020, 15(3): 243-249.]
- [18] Breiman L. Random forests[J]. *Mach Learn*, 2001, 45(1): 5-32.
- [19] Dreiseitl S, Ohno-Machado L. Logistic regression and artificial neural network classification models: a methodology review[J]. *J Biomed Inform*, 2002, 35(5): 352-359.
- [20] Xiao LH, Chen PR, Gou ZP, *et al.* Prostate cancer prediction using the random forest algorithm that takes into account transrectal ultrasound findings, age, and serum levels of prostate-specific antigen[J]. *Asian J Androl*, 2017, 19(5): 586-590.
- [21] Sarica A, Cerasa A, Quattrone A. Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: a systematic review[J]. *Front Aging Neurosci*, 2017, 9: 329.
- [22] Wang CX, Chen X, Liu FB, *et al.* An analysis of the effect of serum prothrombin activity (PTA) and related factors on prognosis of patients with severe chronic hepatitis with liver fibrosis[J]. *J North Sichuan Med Coll*, 2019, 34(3): 441-444. [王超秀, 陈妍, 刘发彬, 等. 血清凝血酶原活动度及相关因素对重度慢性肝炎肝纤维化患者预后的影响分析[J]. 川北医学院学报, 2019, 34(3): 441-444.]
- [23] Cui YP, Liu FH, Shi QF, *et al.* Logistic regression analysis of prognostic factors in 106 acute-on-chronic liver failure patients with hepatic encephalopathy[J]. *J Clin Hepatol*, 2014, 30(10): 992-995. [崔燕平, 刘凤华, 石庆凤, 等. 106例慢加急性肝衰竭合并肝性脑病患者预后影响因素的Logistic回归分析[J]. 临床肝胆病杂志, 2014, 30(10): 992-995.]
- [24] Lv RY, Li SX. Study on the risk factors influencing the prognosis of hepatic encephalopathy[J]. *Chin J Gen Pract*, 2012, 10(11): 1687-1689. [吕日英, 李仕雄. 肝性脑病患者预后的危险因素分析[J]. 中华全科医学, 2012, 10(11): 1687-1689.]
- [25] Routhu M, Safka V, Routhu SK, *et al.* Observational cohort study of hepatic encephalopathy after transjugular intrahepatic portosystemic shunt (TIPS)[J]. *Ann Hepatol*, 2017, 16(1): 140-148.
- [26] Xu Y, Hu YY, Cao L, *et al.* Prognostic value of neutrophil/lymphocyte ratio in patients with hepatic encephalopathy[J]. *Mode Digest Intervent*, 2019, 24(3): 300-303. [徐言, 胡园园, 曹亮, 等. 中性粒细胞/淋巴细胞比值对肝性脑病预后的预测价值[J]. 现代消化及介入诊疗, 2019, 24(3): 300-303.]
- [27] Liu XF. Discussing the method and effect of intravenous sodium supplement in decompensated cirrhosis complicated with hyponatremia[J]. *J Front Med*, 2018, 8(3): 101-102. [刘学芳. 失代偿期肝硬化并发低钠血症的静脉补钠方法及效果探讨[J]. 医药前沿, 2018, 8(3): 101-102.]
- [28] Jiang HM, Wang XH, Yang Y, *et al.* Relationship between hyponatremia and hepatic encephalopathy and renal function in patients with decompensated cirrhosis[J]. *World Chin J Digestol*, 2013, 21(31): 3418-3421. [蒋汉梅, 王学红, 杨勇, 等. 肝硬化失代偿期患者低钠血症与肝性脑病及肾功能的相关性[J]. 世界华人消化杂志, 2013, 21(31): 3418-3421.]
- [29] Bai Z, Guo X, Tacke F, *et al.* Association of serum albumin level with incidence and mortality of overt hepatic encephalopathy in cirrhosis during hospitalization[J]. *Therap Adv Gastroenterol*, 2019, 12: 1756284819881302.
- [30] Bai Z, Bernardi M, Yoshida EM, *et al.* Albumin infusion may decrease the incidence and severity of overt hepatic encephalopathy in liver cirrhosis[J]. *Aging (Albany NY)*, 2019, 11(19): 8502-8525.

(收稿日期: 2020-09-10; 修回日期: 2021-01-12)

(责任编辑: 熊晓然, 纪方方)