XXXX DOI:

基于机器学习构建失代偿期乙型肝炎肝硬化并发ACLF患者的短期预后模型

任 静[1],柏承志[2],苏绍辉[2],汤善宏[1]

(1.川北医学院临床医学院,南充 637000;2.中国人民解放军西部战区总医院消化内科,成都 610083)

【摘 要】目的:基于机器学习构建失代偿期乙型肝炎肝硬化并发慢加急性肝衰竭(acute-on-chronic liver failure,ACLF)患者的90 d预后模型,并评估其预测效能。**方法:**回顾性收集2015年1月至2023年1月于西部战区总医院收治的218例失代偿期乙型肝炎肝硬化并发ACLF患者的临床资料,将患者以7∶3的比例随机分为训练集(153例)和验证集(65例),训练集用于构建模型,验证集用于验证模型;通过最小绝对收缩与选择算子(least absolute shrinkage and selection operator,LASSO)回归分析方法筛选与其预后相关的关键变量,采用Logistic回归、梯度提升机(Gradient Boosting Machine,GBM)、随机森林(Random Forest)、K近邻算法(K-nearest neighbor,KNN)、轻量级梯度提升机(Light Gradient Boosting Machine,LightGBM)等10种机器学习算法构建失代偿期乙型肝炎肝硬化并发ACLF患者90 d短期预测模型,绘制受试者工作特征曲线(receiver operating characteristic curve,ROC),比较ROC曲线下面积(area under the curve,AUC)、准确度、灵敏度、特异度和F1分数等指标评估不同模型的预测效能。通过沙普利可加性解释(Shapley additive explanations,SHAP)分析各关键变量对最优模型预测结果的贡献度并排序。**结果:**通过LASSO回归筛选出6个变量,基于此构建10种机器学习模型,其中LightGBM模型预测效能最佳,AUC为0.813,准确度为0.769,SHAP图显示重要性排序分别为总胆红素(total bilirubin,TBIL)、国际标准化比值(international normalized ratio,INR)、肝性脑病(hepatic encephalopathy,HE)、白细胞(white blood cell count,WBC)、年龄(age)和低密度脂蛋白胆固醇(low-density lipoprotein cholesterol,LDL-C)。**结论:**成功构建出预测失代偿期乙型肝炎肝硬化并发ACLF患者90 d预后的模型,其中LightGBM模型预测效能最佳,有助于识别高风险患者,为临床决策及干预提供依据。

【关键词】失代偿期肝硬化;慢加急性肝衰竭;预后;机器学习

【中图分类号】R575.3 **【文献标志码】**A

Construction of a short-term prognostic model for patients with decompensated hepatitis B cirrhosis complicated by acute-on-chronic liver failure based on machine learning

Ren Jing[1],*Bai Chengzhi*[2],*Su Shaohui*[2],*Tang Shanhong*[1]

(*1.School of Clinical Medicine,North Sichuan Medical College;2.Department of Gastroenterology,The General Hospital of Western Theater Command,PLA*)

【Abstract】*Objective*: To construct a predictive model for the 90-day prognosis of patients with decompensated hepatitis B cirrhosis complicated by acute-on-chronic liver failure (ACLF), and to assess its predictive efficacy. ***Methods*:** A retrospective analysis was performed for the clinical data of 218 patients with decompensated hepatitis B cirrhosis complicated by ACLF who were admitted to The General Hospital of Western Theater Command from January 2015 to January 2023, and these patients were randomly divided into a training set with 153 patients and a validation set with 65 patients at a ratio of 7∶3, with the training set used for model construction and the validation set for model validation. The least absolute shrinkage and selection operator (LASSO) regression analysis was used to identify the key variables associated with prognosis, and 10 machine learning algorithms were used to construct a predictive model for the 90-day short-term prognosis of patients with decompensated hepatitis B cirrhosis complicated by ACLF, including logistic regression model, gradient boosting machine (GBM), random forest, k-nearest neighbor (KNN), and light gradient boosting machine (LightGBM). The receiver operating characteristic (ROC) curve was plotted, and the area under the ROC curve (AUC), accuracy,

作者介绍:任 静,Email:renjing3036@163.com,
研究方向:肝衰竭临床与基础研究。

通信作者:汤善宏,Email:tangshanhong@swjtu.edu.cn。

基金项目:四川省中医药管理局课题资助项目(编号:2023MS458)。

sensitivity, specificity, and F1-score of different models were compared to assess their predictive efficacy. Shapley Additive exPlanations(SHAP) was used to analyze and rank the contribution of key variables to the prediction results of the optimal model. ***Results***: Six variables were identified by LASSO regression, and 10 machine learning models were constructed based on these variables, among which the LightGBM model showed the optimal predictive performance, with an AUC of 0.813 and an accuracy of 0.769. The SHAP plots showed that total bilirubin had the greatest importance, followed by international normalized ratio, hepatic encephalopathy, white blood cell count, age, and low-density lipoprotein cholesterol. ***Conclusion***: This study successfully constructed the models for predicting the 90-day prognosis of patients with decompensated hepatitis B cirrhosis complicated by ACLF, among which the LightGBM model has the best predictive performance and can help to identify high-risk patients and provide a basis for clinical decision-making and intervention.

【Key words】decompensated cirrhosis; acute-on-chronic liver failure; prognosis; machine learning

乙型肝炎病毒(hepatitis B virus, HBV)感染是一个全球性的健康问题。据估计,全球有近3亿人感染乙型肝炎病毒,约有33万人死于相关的肝硬化和慢性肝病等并发症。我国属于HBV感染高流行区,受人口基数大、既往感染防控水平有限等因素影响,HBV感染者约9 600万,约占全球总感染人数的1/3,其中慢性乙型肝炎患者约2 800万,患病率约为5%~6%[1]。慢性肝炎可逐渐进展为肝硬化,依据是否伴随腹水(ascites)、肝性脑病(hepatic encephalopathy, HE)和上消化道出血(upper gastrointestinal hemorrhage, UGIB)等并发症,可分为代偿期和失代偿期。其中,失代偿期肝硬化可在急性诱因驱使下并发乙型肝炎相关慢加急性肝衰竭(HBV-related acute-on-chronic liver failure, HBV-ACLF),这是一种因肝功能迅速恶化而导致多器官功能衰竭的严重临床综合征[2],常以肝细胞大量凋亡坏死、凝血功能异常和黄疸等为典型表现。国内外大量研究表明[3-6],失代偿期肝硬化并发ACLF患者的短期死亡率极高,因此,目前迫切需要更全面、更易获取、更实用的临床指标来预测失代偿期肝硬化并发ACLF患者的预后。机器学习(machine learning, ML)作为一种强大的计算工具,已经在医学的各个领域显示出显著优势[7-8],但在预测失代偿期乙型肝炎肝硬化并发ACLF的短期预后方面的应用价值有待验证。本研究回顾性分析西部战区总医院收治的失代偿期乙型肝炎肝硬化并发ACLF患者的临床资料,筛选出与其预后相关的关键变量,构建多种机器学习模型,并评估其预测效能,旨在提升对该类患者预后的预测效能,为临床提供一个更精准、更有效、更客观的决策工具。

1 资料与方法

1.1 研究对象

选取2015年1月至2023年1月在西部战区总医院住院的失代偿期乙型肝炎肝硬化并发ACLF患者共218例。

1.1.1 纳入标准 ①所有入组的患者均符合《肝衰竭诊治指南(2018年版)》定义的ACLF诊断标准[9];②年龄≥18岁,≤80岁;③乙肝表面抗原阳性,并且阳性持续时间>6个月。入组患者均签署知情同意书,本研究方案经西部战区总医院伦理委员会批准,批号:2020ky005。

所有患者根据病情需要针对性给予保肝、抗病毒治疗、补充白蛋白及血浆、输注凝血物质、防治并发症,必要时行人工肝治疗。

1.1.2 排除标准 ①同时感染其他病毒[如甲型、丁型、丙型、戊型肝炎病毒或人类免疫缺陷病毒(human immunodeficiency virus, HIV)];②导致肝功能衰竭的其他原因(如自身免疫性肝病、药物性肝炎、酒精性肝损伤等疾病);③入院时疑诊或确诊肝癌或其他恶性肿瘤;④合并严重的心、肺、肾等其他器官系统的疾病;⑤年龄<18岁或>80岁;⑥临床资料不完整或失访。

1.1.3 分组方法 主要研究终点为患者入院后90 d内的全因死亡率,并根据患者随访的预后情况分为生存组和死亡组。

1.1.4 观察指标 收集住院患者的性别、年龄等一般人口学资料、并发症情况及入院后基线的实验室检查资料,包括血脂指标、血常规、肝肾功能、凝血功能及电解质等指标。

1.2 模型构建及评估

1.2.1 模型构建 所有患者按照7∶3比例随机分成训练集(153例)和验证集(65例),在训练集中采用LASSO回归分析筛选与失代偿期乙型肝炎肝硬化并发ACLF患者90 d预后相关的变量,将回归系数不为0的变量纳入Logistic回归、支持向量机(Support Vector Machines, SVM)、梯度提升机(Gradient Boosting Machine, GBM)、神经网络(Neural Net-

work）、随机森林（Random Forest）、极端梯度提升（Xtreme Gradient Boosting，XGBoost）、K近邻算法（K-nearest neighbor，KNN）、自适应提升算法（Adaptive Boosting machine，AdaBoost）、轻量级梯度提升机（Light Gradient Boosting Machine，LightGBM）、分类提升（Categorical Boosting，CatBoost）10种机器学习算法，构建失代偿期乙型肝炎肝硬化并发ACLF患者短期预后预测模型，针对训练集数据进行5折交叉验证获得最优模型。

1.2.2 模型评估 使用验证集计算各模型的准确度、灵敏度、特异度、F1分数及受试者工作特征曲线（receiver operating characteristic，ROC）及其曲线下面积（area under the curve，AUC）评估模型预测效能。通过沙普利可加性解释（Shapley additive explanations，SHAP）量化关键变量在模型中的贡献度，评估变量的重要性。

1.3 统计方法

采用SPSS 27.0和R软件（R 4.4.2）进行统计分析。正态分布的连续型变量以均数±标准差（$\bar{x}\pm s$）表示，组间比较采用独立样本t检验；非正态分布的连续型变量用中位数（四分位间距）[$M_d(P_{25},P_{75})$]表示，用Mann-Whitney U检验进行组间比较。计数资料以频数和百分比表示，组间比较采用卡方检验。检验水准α=0.05。

2 结 果

2.1 构建训练集和验证集

根据纳入和排除标准，共收集218例失代偿期乙型肝炎肝硬化并发ACLF患者，死亡发生率为52.3%（114/218）。通过R语言将患者按7∶3随机分为训练集（153例）与验证集（65例），在训练集中有127例（83.0%）患者为男性，验证集中有54例（83.1%）患者为男性。两组间各项指标差异均无统计学意义（P值均>0.05）（表1）。

2.2 生存组和死亡组患者各指标的比较

在训练集中，根据患者90 d的预后情况分为2组，其中生存组73例，死亡组80例（表2）。研究发现死亡组年龄、TBIL、PT、INR、WBC、尿素氮、总胆汁酸及肌酐显著高于生存组（P值均<0.05）；而LDL-C、前白蛋白、纤维蛋白原及Na^+水

表1 训练集与验证集患者临床特征比较

参数	训练集（n=153）	验证集（n=65）	P值
性别（男）	127（83.0%）	54（83.1%）	0.990
年龄（岁）	52.00（45.00，61.00）	57.00（48.00，61.00）	0.152
HDL-C（mmol/L）	0.32（0.17，0.42）	0.32（0.20，0.45）	0.695
LDL-C（mmol/L）	1.46（1.03，1.72）	1.46（1.07，1.88）	0.343
白蛋白（g/L）	30.20（27.30，33.20）	31.70（27.90，33.50）	0.147
前白蛋白（g/L）	41.00（27.00，56.00）	41.00（27.00，51.00）	0.571
碱性磷酸酶（U/L）	164.00（120.30，201.50）	152.20（122.10，205.80）	0.147
TBIL（μmol/L）	344.10（242.71，480.50）	344.70（228.20，545.26）	0.440
ALT（U/L）	213.10（98.10，627.80）	350.60（97.40，676.10）	0.253
AST（U/L）	243.10（116.40，558.50）	290.60（124.00，713.90）	0.393
总胆汁酸（μmol/L）	220.80（169.90，272.10）	217.40（176.00，303.80）	0.517
胆碱酯酶（U/mL）	3.24（2.40，3.60）	3.24（2.71，3.80）	0.125
γ-谷氨酰转移酶（U/L）	82.90（56.00，124.30）	92.50（54.80，156.80）	0.356
视黄醇结合蛋白（mg/L）	11.98（10.00，13.49）	11.98（10.40，12.38）	0.911
PT（s）	20.90（18.40，25.80）	21.20（19.10，26.70）	0.420
INR	1.87（1.63，2.32）	1.89（1.68，2.41）	0.568
纤维蛋白原（g/L）	1.15（0.88，1.43）	1.15（0.93，1.38）	0.830
AFP（ng/mL）	49.69（14.22，141.20）	26.88（7.44，77.33）	0.095
WBC（$\times10^9$/L）	6.13（4.35，8.45）	6.80（5.16，8.55）	0.530
PLT（$\times10^9$/L）	70.00（42.00，101.00）	73.00（52.00，110.00）	0.296
LYM（$\times10^9$/L）	0.79（0.54，1.12）	0.82（0.57，1.36）	0.436
C反应蛋白（$\times10^9$/L）	16.50（9.24，21.65）	18.01（12.08，21.07）	0.307
K^+（mmol/L）	3.88（3.48，4.22）	3.86（3.46，4.25）	0.803
Na^+（mmol/L）	134.40（131.51，136.93）	134.80（131.10，138.10）	0.511
尿素氮（mmol/L）	5.00（3.74，7.82）	5.51（3.86，7.61）	0.475
肌酐（μmol/L）	78.00（64.60，94.00）	82.00（68.00，106.90）	0.159

注：HDL-C，高密度脂蛋白胆固醇；LDL-C，低密度脂蛋白胆固醇；TBIL，总胆红素；ALT，谷丙转氨酶；AST，谷草转氨酶；PT，凝血酶原时间；INR，国际标准化比值；AFP，甲胎蛋白；WBC，白细胞；PLT，血小板；LYM，淋巴细胞；K^+，血清K离子；Na^+，血清Na离子

平显著低于生存组（P值均<0.05）。

2.3 失代偿期乙型肝炎肝硬化并发ACLF患者预后不良预测模型的变量选择

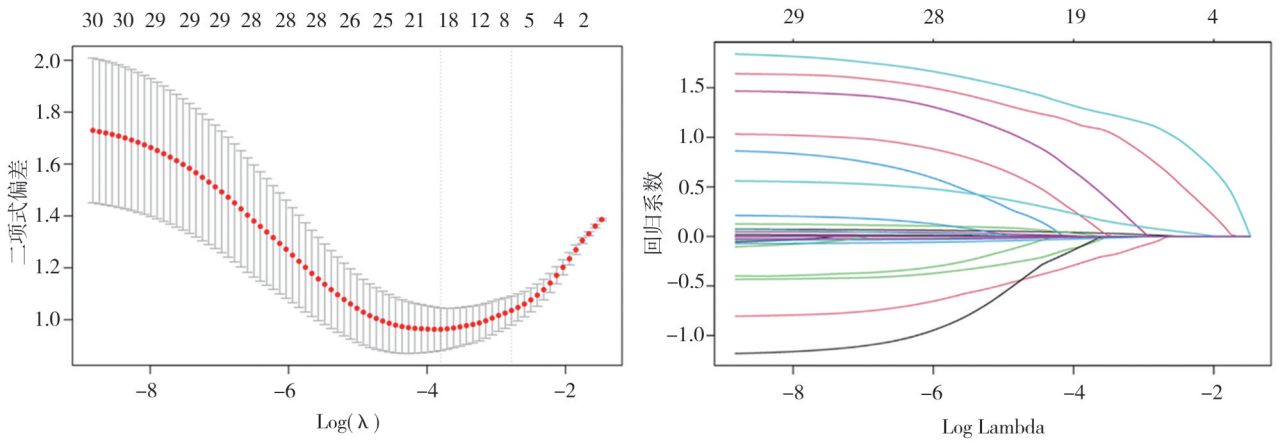
以失代偿期乙型肝炎肝硬化并发ACLF患者90 d是否发生死亡为因变量（死亡=1，生存=0），将30个变量纳入LASSO回归分析结果显示，设定最小标准误差的λ值为0.06，最终筛选出6个预测因素：HE、年龄、LDL-C、TBIL、INR和WBC（图1）。

2.4 不同机器学习模型的预测效能评估

基于上述特征构建10种机器学习模型，ROC曲线显示

表2 生存组和死亡组患者各指标的比较

参数	生存组（n=73）	死亡组（n=80）	χ^2/Z值	P值
性别（男）	60（82.2%）	67（83.8%）	0.066	0.798
年龄（岁）	50.00（44.00，56.00）	54.50（46.00，64.00）	−2.412	0.016
HDL-C（mmol/L）	0.31（0.16，0.48）	0.32（0.18，0.40）	−0.610	0.542
LDL-C（mmol/L）	1.47（1.25，1.86）	1.32（0.93，1.51）	−3.135	0.002
白蛋白（g/L）	29.80（26.67，32.51）	30.50（27.70，33.40）	−0.961	0.337
前白蛋白（g/L）	42.45（34.00，63.60）	36.80（21.00，48.86）	−2.134	0.033
碱性磷酸酶（U/L）	163.30（121.30，189.90）	167.65（118.60，215.20）	−1.302	0.193
TBIL（μmol/L）	280.30（210.40，402.30）	413.74（298.93，524.61）	−4.407	<0.001
ALT（U/L）	213.80（100.40，496.10）	207.10（85.25，686.10）	−0.544	0.586
AST（U/L）	223.30（114.50，457.90）	274.45（123.05，661.55）	−1.264	0.206
总胆汁酸（μmol/L）	203.60（159.10，250.93）	246.95（179.63，285.86）	−2.418	0.016
胆碱酯酶（U/mL）	3.24（2.40，3.54）	3.24（2.38，3.65）	−0.263	0.793
γ-谷氨酰转移酶（U/L）	85.90（57.80，119.80）	79.60（55.50，136.70）	−0.438	0.661
视黄醇结合蛋白（mg/L）	11.98（10.00，13.80）	11.98（10.05，13.39）	−0.601	0.548
PT（s）	18.90（17.50，21.60）	23.60（20.15，29.55）	−5.368	<0.001
INR	1.69（1.56，1.91）	2.12（1.77，2.64）	−5.213	<0.001
纤维蛋白原（g/L）	1.27（0.94，1.62）	1.08（0.82，1.31）	−1.984	0.047
AFP（ng/mL）	69.49（15.14，203.46）	26.98（13.99，101.56）	−1.552	0.121
WBC（×10^9/L）	5.46（3.28，7.19）	6.90（4.99，10.11）	−3.567	<0.001
PLT（×10^9/L）	79.00（42.00，107.00）	65.50（41.50，97.50）	−1.359	0.174
LYM（×10^9/L）	0.83（0.56，1.16）	0.74（0.54，1.09）	−0.888	0.375
C反应蛋白（×10^9/L）	14.57（8.93，20.58）	19.01（10.63，21.78）	−1.053	0.292
K^+（mmol/L）	3.90（3.47，4.14）	3.87（3.49，4.34）	−0.763	0.445
Na^+（mmol/L）	134.80（132.52，138.04）	134.00（130.26，136.14）	−2.398	0.016
尿素氮（mmol/L）	4.83（3.48，5.92）	5.82（3.95，9.26）	−2.321	0.020
肌酐（μmol/L）	74.00（63.00，89.00）	82.00（66.80，104.85）	−2.104	0.035



A. LASSO回归系数与变量对应图　　B. 依据Λ值筛选变量图

图1 基于LASSO回归的预测因子选择

LightGBM模型预测效能最佳，AUC为0.813[95%置信区间（confidence interval，CI）=0.708~0.918]，准确度为0.769，灵敏度为0.765（表3，图2）。校准图（图3）和决策曲线分析（图4）结果显示，LightGBM模型的性能良好，混淆矩阵见图5。

表3 不同机器学习模型在验证集中的效能

模型	AUC(95%CI)	准确度	灵敏度	特异度	F1分数
Logistic	0.778(0.664~0.892)	0.738	0.647	0.839	0.72
SVM	0.768(0.652~0.884)	0.754	0.824	0.677	0.78
GBM	0.792(0.684~0.901)	0.738	0.588	0.903	0.70
Neural Network	0.777(0.662~0.892)	0.738	0.676	0.806	0.73
Random Forest	0.777(0.660~0.893)	0.754	0.676	0.839	0.74
XGBoost	0.805(0.699~0.910)	0.754	0.794	0.710	0.77
KNN	0.737(0.609~0.864)	0.769	0.765	0.774	0.78
AdaBoost	0.796(0.688~0.904)	0.754	0.588	0.935	0.71
LightGBM	0.813(0.708~0.918)	0.769	0.765	0.774	0.78
CatBoost	0.792(0.681~0.904)	0.754	0.794	0.710	0.77

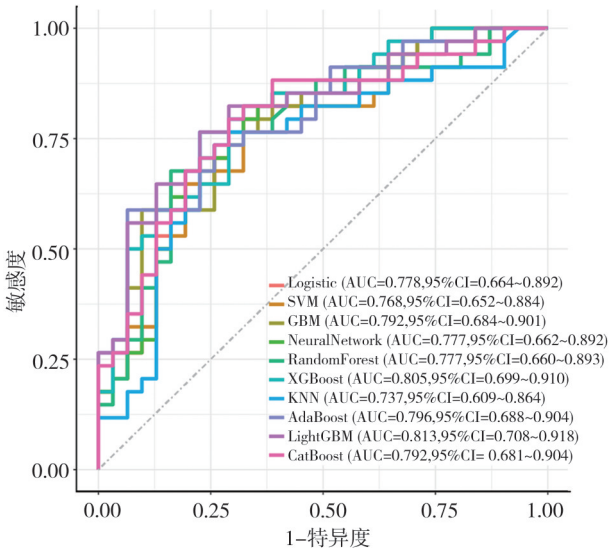


图2 模型在验证集中的ROC曲线

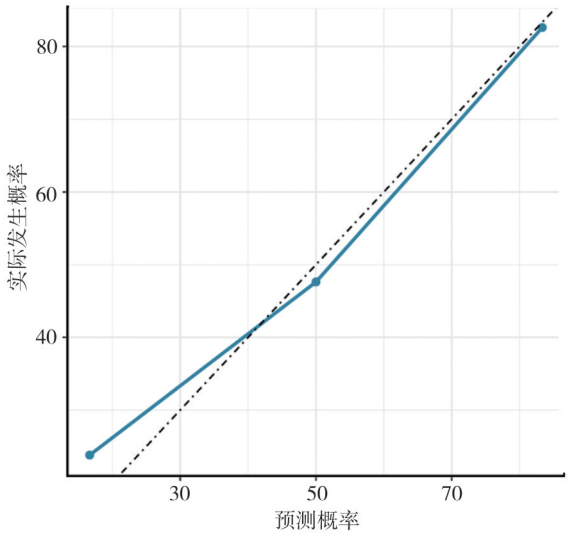


图3 LightGBM模型预测患者预后的校准图

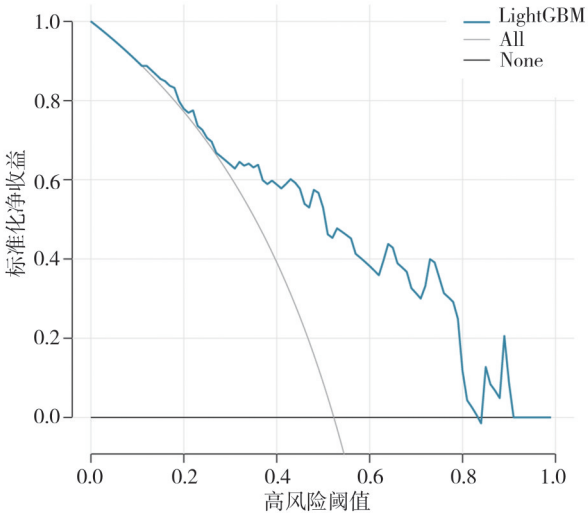


图4 LightGBM模型预测患者预后的决策曲线分析

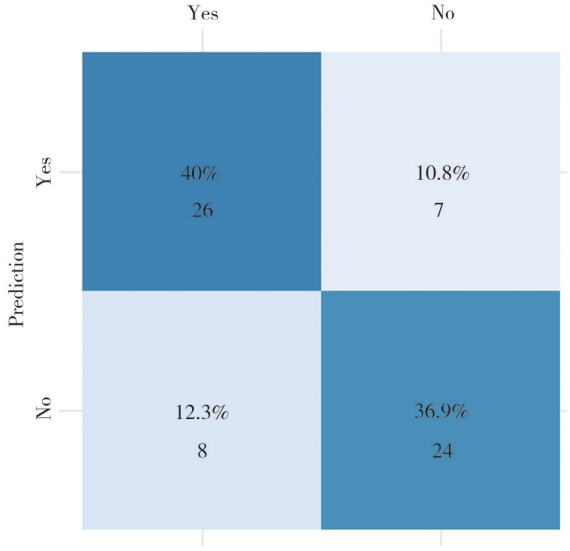


图5 LightGBM模型的混淆矩阵图

2.5 LightGBM模型的SHAP分析

通过SHAP分析量化评估各因素对失代偿期乙型肝炎肝硬化并发ACLF患者预后预测的重要性，结果显示，TBIL是最重要的因素，其次是INR、HE、WBC、年龄、LDL-C（图6）。图7展示了年龄对失代偿期乙型肝炎肝硬化并发ACLF患者90 d预后预测的贡献如何受到HE状态的调控。随着年龄增长，SHAP值由负转正并逐步升高，提示高龄是预后不良的独立危险因素。同时，散点颜色揭示了显著的交互效应：在几乎所有年龄段，伴有HE（HE=1，图中浅色点）的患者，其年龄对应的SHAP值普遍高于无HE（HE=0，图中深色点）的患者，说明HE状态可进一步放大年龄对不良预后的正向贡献，2者存在协同效应。

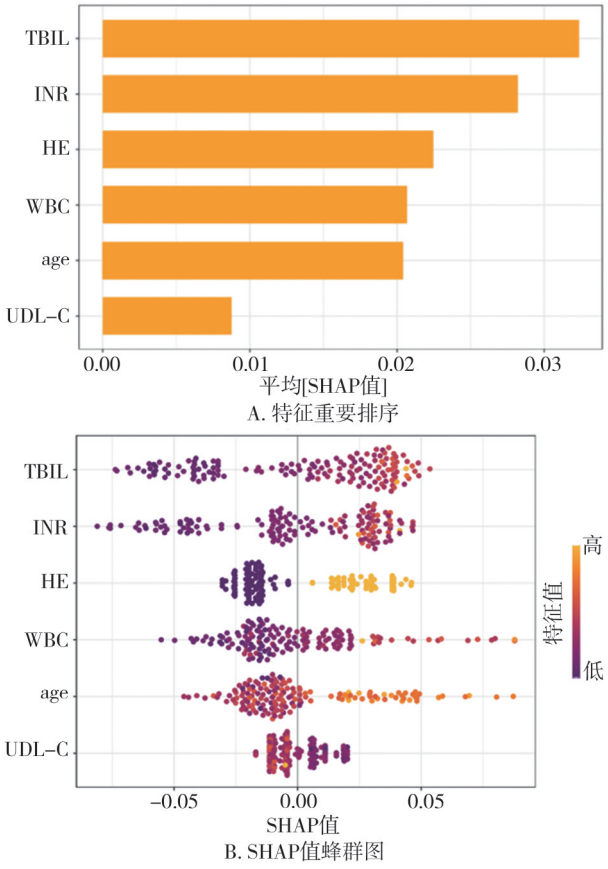


图 6 LightGBM模型的SHAP图解释

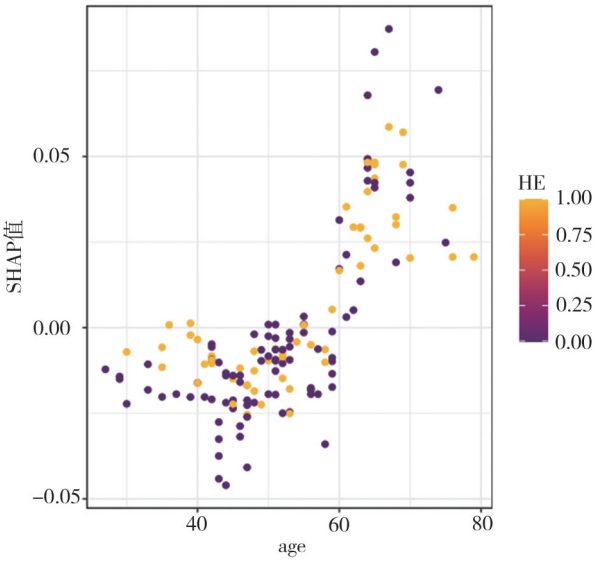


图 7 年龄与HE交互作用的SHAP依赖图

3 讨 论

失代偿期乙型肝炎肝硬化并发ACLF作为一种临床危重综合征，其病情进程以急剧恶化与多系统受累为显著特征，且短期病死率极高[3]。因此，目前迫切需要更全面、更易获取、更实用的临床指标来预测ACLF患者的预后。本研究基于失代偿期乙型肝炎肝硬化并发ACLF患者的临床资料，采用LASSO回归分析进行变量筛选，构建并比较了10种机器学习模型的预测效能。

HE是脑功能衰竭的表现，是ACLF患者的常见并发症及主要致死原因之一，其病理特征在于血氨升高导致的神经毒性[10]。血氨升高可触发一系列连锁反应，包括细胞肿胀、氧化应激、炎症反应、线粒体功能障碍以及谷氨酸系统受损。这些变化进而导致血脑屏障功能障碍、神经元凋亡与神经功能损伤，最终诱发并加重HE的临床症状[11]。目前，多种肝衰竭预后评估系统，如CTP、AARC、CLIF-C ACLFs、COSSCH-ACLF及COSSCH-ACLF Ⅱ等，均将HE列为重要的预后预测因子。本研究还发现年龄是失代偿期乙型肝炎肝硬化并发ACLF患者90 d预后的重要变量，这一结果与多数研究结论一致[12-13]，其核心原因在于老年患者的器官储备功能自然减退，肝脏再生能力减弱，对急性肝损伤的耐受程度明显降低。散点特征变量图提示HE的存在显著放大了年龄本身所带来的死亡风险，二者存在协同作用，共同构成了高龄且合并HE这一预后风险极高的患者亚组，因此，临床中对于老年HBV-ACLF患者应采取更为积极的治疗策略，加强器官功能支持，重点预防并发症，同时重视基础疾病的规范管理。此外，杜蔷等[14]通过回顾性纳入98例ACLF患者的研究发现，年龄、HE、INR是患者预后不良的危险因素，这一结果与本研究一致。另有研究表明[4]，年龄、TBIL、INR与ACLF患者的预后密切相关。TBIL作为胆红素在肝细胞内代谢的最终产物，其代谢全过程高度依赖肝脏功能，肝细胞需先后完成胆红素的摄取、结合、转化及排泄等连续步骤，任一环节的功能异常均会导致血清TBIL水平异常升高，因此该指标是反映肝细胞损伤程度的经典生化标志物。INR是反映外源性凝血功能的核心指标，当ACLF发生时，肝细胞广泛坏死会导致凝血因子合成锐减，同时炎症反应引发的凝血-纤溶紊乱，进一步加重凝血功能障碍，最终表现为INR升高。因此，在ACLF的预后评估领域，TBIL和INR的预测价值已得到广泛验证，无论是MELD、MELD-Na等传统预后评分系统，还是CLIF-C ACLFs、COSSH-

ACLF II等近年推出的新型预测模型，均将基线TBIL和INR纳入核心预测指标体系。

WBC是全身炎症反应的血液学标志物，其计数水平可直接、快速地反映全身炎症反应的程度，是临床判断机体是否存在炎症损伤的经典且便捷的标志物，是ACLF的关键危险因素[15]。邓珂欣和罗红春[16]通过纳入208例ACLF患者的研究发现，患者发生细菌感染的占68.8%，且感染患者的WBC水平显著高于未感染的患者。LDL-C主要由肝脏合成和代谢，其血清水平可直接反映肝细胞的合成功能储备，当肝细胞受损时，LDL-C水平可下降，在ACLF中常可观察到血清胆固醇水平低于正常水平[17]。1项ACLF前期研究发现，在ACLF前期进展到ACLF的过程中，LDL-C水平显著降低[18]。Xiao CC等[19]针对HBV-ACLF患者的回顾性研究显示，LDL-C降低是3个月死亡率的独立危险因素，且LDL≥1.0mmol/L的患者总体生存率显著高于低水平患者。因此，将LDL-C纳入预后评估模型，有助于早期识别高风险人群，为临床判断病情严重程度、制定干预策略提供客观依据。

机器学习作为一类高效的数据分析工具，已在医学诸多领域的预测与决策研究中展现出显著应用优势[20-22]，但其在失代偿期乙型肝炎肝硬化并发ACLF患者短期预后预测中的价值，目前尚未明确。本研究基于10种机器学习算法，结合患者基线临床指标，构建失代偿期乙型肝炎肝硬化并发ACLF患者的90 d不良预后预测模型。验证集结果显示，LightGBM模型整体性能最优，其AUC、准确度及F1分数分别为0.813、0.769、0.780。针对本研究样本量偏小的情况，本研究通过引入正则化技术、早停策略及类别平衡权重参数优化，有效缓解了复杂机器学习模型在小样本数据中易出现的过拟合问题，使得LightGBM模型既能精准捕捉临床指标间的非线性交互关系，又能维持良好的外部稳定性。此外，本研究采用SHAP可解释性分析方法，量化各临床特征对模型预测结果的贡献度并实现模型可视化，有效提高了模型的可解释性与临床可信度，为模型的临床实际应用提供了客观依据。SHAP分析结果显示，模型构建所得的特征风险关联模式与临床认知高度契合，其中TBIL、INR等经典核心指标对预后预测贡献度突出，HE、WBC等其他指标的风险贡献规律也符合临床诊疗经验。综上，本研究构建的LightGBM预测模型，在失代偿期乙型肝炎肝硬化并发ACLF患者90 d不良预后预测中表现优异，结合SHAP可解释性分析方法进一步凸显了其临床应用价值，可辅助临床医师快速识别高危患者，进而制定个体化诊疗方案与干预策略，改善患者短期预后。

本研究仍存在一些局限性。首先，样本量较小，可能限制模型的泛化性能，未来需扩大样本量，并采用多中心数据进行外部验证；其次，变量选择仅纳入了常规临床指标，未包含新型生物标志物，后续可结合多组学数据提升模型性能；最后，本研究未开展前瞻性验证，未来需通过临床试验评估模型对临床决策的实际影响。

综上所述，本研究构建出预测失代偿期乙型肝炎肝硬化并发ACLF患者90 d预后的LightGBM预测模型，该模型预测效能最佳，通过对不同特征在最优模型预测结果的贡献度进行排序，明确了TBIL、INR、HE等是该类患者90 d预后的重要预测因子，为临床风险分层提供了可操作的工具。

利益冲突 所有作者声明不存在利益冲突。

作者贡献声明 任静、汤善宏：研究方案设计；任静：数据统计分析、作图、文章撰写；柏承志：数据收集；苏绍辉：文章修改；汤善宏：研究指导、论文最终审阅。

参 考 文 献

[1] Polaris Observatory Collaborators. Global prevalence, cascade of care, and prophylaxis coverage of hepatitis B in 2022: a modelling study [J]. Lancet Gastroenterol Hepatol, 2023, 8(10): 879-907.

[2] 中华医学会感染病学分会肝衰竭与人工肝学组，中华医学会肝病学分会重型肝病与人工肝学组. 肝衰竭诊治指南(2024年版)[J]. 中华肝脏病杂志, 2025, 33(01): 18-33.

Liver Failure and Artificial Liver Group, Chinese Society of Infectious Diseases, Chinese Medical Association; Severe Liver Disease and Artificial Liver Group, Chinese Society of Hepatology, Chinese Medical Association. Guideline for diagnosis and treatment of liver failure[J]. Chin J Hepatol, 2025, 33(01): 18-33.

[3] Trebicka J, Praktiknjo M, Peiffer KH, et al. Acute on chronic liver failure[J]. Dtsch Ärztebl Int, 2025, 122(4): 96-102.

[4] Li CY, Hu H, Bai CZ, et al. Alpha-fetoprotein and APRI as predictive markers for patients with Type C hepatitis B-related acute-on-chronic liver failure: a retrospective study[J]. BMC Gastroenterol, 2024, 24(1): 191.

[5] Sarin SK, Choudhury A, Sharma MK, et al. Acute-on-chronic liver failure: consensus recommendations of the Asian Pacific association for the study of the liver (APASL): an update[J]. Hepatol Int, 2019, 13(4): 353-390.

[6] 孔雪杰,李 江,李昊天,等.不同肝病基础的慢加急性肝衰竭患者临床特征和预后评分模型预测效能比较[J].实用肝脏病杂志,2021,24(3):387-390.
Kong XJ, Li J, Li HT, et al. Clinical feature and efficacy of different prognostic scoring models in patients with acute-on-chronic liver failure with different underlying liver diseases[J]. J Prac Hepatol, 2021, 24(3): 387-390.
[7] 韩亚飞,汪 静,张添添,等.基于机器学习的肠球菌血流感染预后不良预测模型的构建与评估[J].中华医院感染学杂志,2026,36(5):790-795.
Han YF, Wang J, Zhang TT, et al. Establishment of prediction model for adverse prognosis of patients with *Enterococcus* bloodstream infection based on machine learning and its predictive efficiency[J]. Chin J Nosocomiol, 2026, 36(5): 790-795.
[8] Deng B, Bai CZ, Xu HQ, et al. Machine learning-based algorithms for the prediction of 90-day survival in patients with liver failure receiving artificial liver therapy[J]. Front Physiol, 2025, 16: 1687860.
[9] 中华医学会感染病学分会肝衰竭与人工肝学组,中华医学会肝病学分会重型肝病与人工肝学组.肝衰竭诊治指南(2018年版)[J].中华肝脏病杂志,2019,27(1):18-26.
Liver Failure and Artificial Liver Group, Chinese Society of Infectious Diseases, Chinese Medical Association; Severe Liver Disease and Artificial Liver Group, Chinese Society of Hepatology, Chinese Medical Association. Guideline for diagnosis and treatment of liver failure[J]. Chin J Hepatol, 2019, 27(1): 18-26.
[10] Li JQ, Liang X, You SL, et al. Development and validation of a new prognostic score for hepatitis B virus-related acute-on-chronic liver failure[J]. J Hepatol, 2021, 75(5): 1104-1115.
[11] Ferenci P, Lockwood A, Mullen K, et al. Hepatic encephalopathy: Definition, nomenclature, diagnosis, and quantification: Final report of the Working Party at the 11th World Congresses of Gastroenterology, Vienna, 1998[J]. Hepatology, 2002, 35(3): 716-721.
[12] 张仕华,柏承志,李春燕,等.老年HBV相关慢加急性肝衰竭预后的危险因素及风险预测列线图模型构建[J].临床肝胆病杂志,2024,40(10):1976-1984.
Zhang SH, Bai CZ, Li CY, et al. Risk factors for the prognosis of elderly patients with hepatitis B virus-related acute-on-chronic liver failure and construction of a nomogram model for risk prediction[J]. J Clin Hepatol, 2024, 40(10): 1976-1984.
[13] Xu XW, Hou ZH, Xu YY, et al. The dynamic of platelet count as a novel and valuable predictor for 90-day survival of hepatitis B virus-related acute-on-chronic liver failure patients[J]. Clin Res Hepatol Gastroenterol, 2021, 45(2): 101482.
[14] 杜 蔷,陆盛菊,卢大秀,等.HBV感染失代偿期肝硬化并发慢加急性肝衰竭患者的临床特点及预后[J].中国老年学杂志,2021,41(3):520-523.
Du Q, Lu SJ, Lu DX, et al. Clinical characteristics and prognosis of patients with decompensated cirrhosis complicated with acute-on-chronic liver failure caused by HBV infection[J]. Chin J Gerontol, 2021, 41(3): 520-523.
[15] 牛 宏,姚 佳,崔少波,等.预后营养指数对人工肝治疗乙型肝炎相关慢加急性肝衰竭的预后评估价值[J].传染病信息,2025,38(4):320-325.
Niu H, Yao J, Cui SB, et al. Prognostic value of the prognostic nutritional index in evaluating outcomes of artificial liver support system treatment for hepatitis B virus-related acute-on-chronic liver failure[J]. Infect Dis Inf, 2025, 38(4): 320-325.
[16] 邓珂欣,罗红春.常用血液学指标评定慢加急性肝衰竭患者并发细菌感染和预后价值研究[J].实用肝脏病杂志,2023,26(2):242-245.
Deng KX, Luo HC. Evaluation of bacterial infection by common hematological indices in patients with acute-on-chronic liver failure[J]. J Prac Hepatol, 2023, 26(2): 242-245.
[17] 王 航,孙长峰,何 川,等.血脂水平对HBV相关慢加急性肝衰竭患者短期预后的预测价值[J].山东医药,2022,62(8):74-77.
Wang H, Sun CF, He C, et al. Predictive value of blood lipid level on short-term prognosis in patients with HBV-related acute-on-chronic liver failure[J]. Shandong Med J, 2022, 62(8): 74-77.
[18] 官 敏,李嘉妮,高银杰,等.血脂指标对乙型肝炎相关慢加急性肝衰竭前期进展的预测价值[J].临床肝胆病杂志,2023,39(7):1564-1569.
Guan M, Li JN, Gao YJ, et al. Value of blood lipid parameters in predicting the progression of HBV-related acute-on-chronic pre-liver failure [J]. J Clin Hepatol, 2023, 39(7): 1564-1569.
[19] Xiao CC, Gong J, Zhu SG, et al. Nomogram based on blood lipoprotein for estimation of mortality in patients with hepatitis B virus-related acute-on-chronic liver failure[J]. BMC Gastroenterol, 2020, 20(1): 188.
[20] Song BL, Liu P, Fu KR, et al. Developing a predictive model for septic shock risk in acute pancreatitis patients using interpretable machine learning algorithms[J]. Digit Health, 2025, 11: 20552076251346361.
[21] Zhang JY, Lei YT, Liu R, et al. Development and validation of a machine learning-based explainable predictive model for long-term net adverse clinical events in patients with high bleeding risk undergoing percutaneous coronary intervention: results from a prospective cohort study[J]. Int J Surg, 2025, 111(9): 6082-6092.
[22] Lynam AL, Dennis JM, Owen KR, et al. Logistic regression has similar performance to optimised machine learning algorithms in a clinical setting: application to the discrimination between type 1 and type 2 diabetes in young adults[J]. Diagn Progn Res, 2020, 4(1): 6.

(**收稿**:2026-03-22;**修回**:2026-06-05;**录用**:2026-06-08)
(责任编辑:周一青)
本文引用格式:
任静,柏承志,苏绍辉,等.基于机器学习构建失代偿期乙型肝炎肝硬化并发ACLF患者的短期预后模型[J].重庆医科大学学报,XXXX,XX(XX):1-8.