

基于机器学习的脓毒症相关心肌损伤患者死亡风险预测模型构建

章新元^{1,2} 庄晓欢^{1,2} 尹晓鹏^{1,2} 刘彦珩^{1,2} 王大成^{1,3} 陈秋华^{1,3} 庄燕^{1,3}

¹南京中医药大学附属医院重症医学科, 江苏南京 210029; ²南京中医药大学第一临床医学院, 江苏南京 210023; ³江苏省中医院重症医学科, 江苏南京 210029

通信作者: 庄燕, Email: athena2004112@163.com

【摘要】目的 探讨基于机器学习脓毒症相关心肌损伤 (SRMI) 患者死亡风险预测模型的构建与验证, 实现 SRMI 患者的早期风险识别。**方法** 基于美国重症监护医学信息数据库 (MIMIC-III v1.4 和 MIMIC-IV v2.2) 筛选 SRMI 患者, 以 MIMIC-IV 为训练集。采用两阶段随机森林 (RF) 特征筛选方法筛选模型变量, 包括递进式特征累加 (RFA) 和递减式容差剔除 (SEIFE) 实现变量精简, 构建多种机器学习模型, 并通过五折交叉验证及外部验证评估模型性能。采用 SHapley 加和解释 (SHAP) 方法分析模型的可解释性。**结果** 从 MIMIC-IV 中共纳入训练集患者 1 789 例, 验证集患者 1 117 例。RFA-SEIFE 框架最终筛选出 10 个核心特征, 包括血乳酸 (Lac)、尿量、体温、活化部分凝血活酶时间 (APTT)、碱性磷酸酶 (ALP)、平均动脉压 (MAP)、血细胞比容 (HCT)、年龄、有创机械通气及使用抗菌药物。基于上述特征构建 7 种机器学习模型, 在五折交叉验证中, RF 模型表现最佳, 受试者工作特征曲线 (ROC 曲线) 下面积 (AUC) 为 0.758 [95% 可信区间 (95%CI) 为 0.736~0.779], 精确率-召回率 (PR) AUC 为 0.693 (95%CI 为 0.676~0.710)。外部验证结果显示, RF 模型具有良好的泛化能力, AUC 为 0.837, PR AUC 为 0.757, F1 值为 0.688, 准确率为 0.800, 平衡准确率为 0.768。校准曲线提示预测概率与实际观察值一致性良好; 决策曲线分析 (DCA) 表明, 在临床可接受阈值范围内模型具有较高净获益; Kolmogorov-Smirnov (KS) 曲线显示模型区分度良好 (KS 评分为 0.518); 风险分层分析表明, 随着预测风险十分位增加, 死亡事件发生率呈递增趋势, 提示模型具备稳定的分层能力。SHAP 分析显示, 尿量、有创机械通气及 Lac 为影响模型预测的主要特征, 其中尿量减少及 Lac 升高与死亡风险显著增加相关。**结论** 基于机器学习的 RF 模型可高效、准确预测 SRMI 患者的死亡风险。该模型简洁、可解释性强, 具有良好的临床应用潜力。

【关键词】 脓毒症; 心肌损伤; 机器学习; 随机森林; 风险预测; 可解释性模型

基金项目: 江苏省中医重点专科建设项目 (2023-5); 江苏省研究生科研与实践创新计划和学位与研究生教育改革课题 (SJX25-1001); 江苏省中医药管理局科技发展计划项目 (MS2021008); 南京中医药大学自然科学基金 (XZR2023002); 江苏省中医院院级课题 (Y25092)

DOI: 10.3969/j.issn.1008-9691.2026.02.004

Based machine learning construction of predictive model for mortality risk in patients with sepsis-related myocardial injury

Zhang Xinyuan^{1,2}, Zhuang Xiaohuan^{1,2}, Yin Xiaopeng^{1,2}, Liu Yanpin^{1,2}, Wang Dacheng^{1,3}, Chen Qiu Hua^{1,3}, Zhuang Yan^{1,3}

¹Department of Critical Care Medicine, Affiliated Hospital of Nanjing University of Chinese Medicine, Nanjing 210029, Jiangsu, China; ²The First Clinical Medical College, Nanjing University of Chinese Medicine, Nanjing 210029, Jiangsu, China; ³Department of Critical Care Medicine, Jiangsu Province Hospital of Chinese Medicine, Nanjing 210029, Jiangsu, China

Corresponding author: Zhuang Yan, Email: athena2004112@163.com

【Abstract】Objective To develop and validate a machine learning-based predictive model for mortality risk in patients with sepsis-related myocardial injury (SRMI), aiming to achieve early risk stratification. **Methods** Patients with SRMI were identified from the American Medical Information Mart for Intensive Care (MIMIC-III v1.4 and MIMIC-IV v2.2), with MIMIC-IV serving as the training set. The model variables were selected using a two-stage random forest (RF) feature selection method, which included recursive feature accumulation (RFA) and sequential elimination with incremental feature evaluation (SEIFE) to achieve variable reduction. Multiple machine learning models were developed and evaluated using five-fold cross-validation and external validation. Model interpretability was assessed using SHapley Additive exPlanations (SHAP). **Results** A total of 1 789 patients from MIMIC-IV were included as the training cohort, and 1 117 patients from MIMIC-III served as the external validation cohort. Using the two-stage RFA-SEIFE, 10 key features were ultimately identified, including lactate acid (Lac), urine output, temperature, activated partial thromboplastin time (APTT), alkaline phosphatase (ALP), mean arterial pressure (MAP), hematocrit (HCT), age, invasive mechanical ventilation, and antibiotic use. Based on these features, seven machine learning models were developed. In five-fold cross-validation, the RF model demonstrated the best performance, with an area under the receiver operator characteristic curve (AUC) of 0.758 [95% confidence interval (95%CI) was 0.736-0.779] and a precision-recall (PR) AUC of 0.693 (95%CI was 0.676-0.710). In the external validation cohort, the RF model exhibited good generalizability, achieving an AUC of 0.837, a PR AUC of 0.757, an F1 score of 0.688, an accuracy of 0.800, and a balanced accuracy of

0.768. Further multidimensional evaluation showed good calibration, with predicted probabilities closely aligned with observed outcomes. Decision curve analysis (DCA) indicated a higher net clinical benefit across a range of clinically relevant threshold probabilities. The Kolmogorov-Smirnov (KS) curve demonstrated good discriminative ability (KS = 0.518). Risk stratification analysis revealed a progressive increase in mortality across higher predicted risk strata, indicating robust stratification performance. SHAP interpretability analysis identified urine output, invasive mechanical ventilation, and Lac as the most influential predictors. Specifically, decreased urine output and elevated Lac levels were significantly associated with increased mortality risk. **Conclusions** The machine learning-based RF model can accurately predict mortality risk in SRMI patients. The model is simple, interpretable, and shows good potential for clinical application.

【Key words】 Sepsis; Myocardial injury; Machine learning; Random forest; Risk prediction; Interpretable model

Fund program: Key Specialty Construction Project of Traditional Chinese Medicine in Jiangsu Province (2023-5); Postgraduate Research & Practice Innovation Program of Jiangsu Province (SJX25-1001); Science and Technology Program of Bureau of Traditional Chinese Medicine of Jiangsu Province (MS2021008); Natural Science Foundation of Nanjing University of Chinese Medicine (XZR2023002); Department-Level Project of Jiangsu Provincial Hospital of Traditional Chinese Medicine (Y25092)

DOI: 10.3969/j.issn.1008-9691.2026.02.004

脓毒症是由感染引发的宿主免疫反应失衡所导致的危及生命的器官功能障碍,是全球重症医学领域的重要公共卫生挑战^[1]。流行病学数据显示,全球每年约发生脓毒症4 900万例,相关死亡患者数超过1 100万例,占全球总死亡患者数的19.7%^[2]。脓毒症相关心肌损伤(sepsis-related myocardial injury, SRMI)是脓毒症常见且严重的并发症之一,其发生率可高达50%~70%,主要表现为心肌酶学指标〔如肌钙蛋白、N末端脑钠肽前体(N-terminal pro-brain natriuretic peptide, NT-proBNP)〕升高及心功能障碍^[3-5]。已有研究证实,SRMI的发生与脓毒症患者的病死率密切相关,可使28 d病死率升高至30%~40%^[6]。从病理生理学角度分析,SRMI的发生机制涉及炎症因子风暴、微循环障碍、氧化应激和线粒体功能损伤等多重途径^[7-8]。

尽管早期识别SRMI对改善预后至关重要,但目前临床常用的风险评估工具〔如急性生理学及慢性健康状况评分II(acute physiology and chronic health evaluation II, APACHE II)、序贯器官衰竭评分(sequential organ failure assessment, SOFA)及生物标志物的预测效能均有限^[9-10]。单一生化指标〔如肌钙蛋白或血乳酸(lactic acid, Lac)〕虽可提示心肌损伤程度,但对患者短期死亡的预测准确性不足〔受试者工作特征曲线(receiver operator characteristic curve, ROC曲线)下面积(area under the curve, AUC)通常<0.750〕^[11]。因此,建立一种基于多维临床数据、能动态整合多因素影响的风险预测模型,对于SRMI患者的早期风险识别具有重要意义。近年来,机器学习技术在重症医学领域的应用快速发展^[12-13]。已有研究表明,基于随机森林

(random forest, RF)、极端梯度提升(extreme gradient boosting, XGBoost)或神经网络等算法构建的脓毒症死亡预测模型,其表现优于传统逻辑回归(Logistic regression, LR)方法^[14-15]。然而,目前仍鲜见针对SRMI人群的机器学习研究,现有模型普遍存在特征冗余、过拟合及可解释性不足等问题,限制了其在临床决策中的推广与应用。

本研究利用美国重症监护医学信息数据库(Medical Information Mart for Intensive Care, MIMIC-III和MIMIC-IV)筛选符合SRMI诊断标准的患者,采用两阶段RF特征筛选框架,包括递进式特征累加(recursive feature accumulation, RFA)和递减式容差剔除(sequential elimination with incremental feature evaluation, SEIFE)实现变量精简,并结合Shapley加和解释(SHapley Additive exPlanations, SHAP)算法解析模型的决策机制。通过多种机器学习算法构建并比较预测模型的性能,旨在建立一种基于常规临床数据的SRMI死亡风险预测模型,为临床早期识别高危患者,并制定个体化治疗策略提供参考。

1 资料与方法

1.1 研究对象与数据来源:采用回顾性队列研究方法。从MIMIC-III v1.4与MIMIC-IV v2.2中筛选SRMI患者,以MIMIC-IV为训练集^[16]。

1.1.1 纳入标准:依据国际疾病分类编码(international classification of diseases, ICD-9, ICD-10)临床修订版筛选符合脓毒症3.0定义的首次入住重症监护病房(intensive care unit, ICU)的脓毒症患者,且在入住ICU后24 h内至少存在一项心肌损伤标志物升高〔脑钠肽(brain natriuretic peptide,

BNP) > 100 ng/L、NT-proBNP > 300 ng/L、肌钙蛋白 T (troponin T, TnT) > 0.01 μg/L 或肌酸激酶同工酶 (MB isoenzyme of creatine kinase, CK-MB) > 5 μg/L^[1,3,5,17]。

1.1.2 排除标准: ① 既往存在慢性心血管疾病(充血性心力衰竭、心肌病、缺血性心脏病、瓣膜性或先天性心脏病)或心脏术后直接转入 ICU 者; ② 合并转移性实体瘤、活动性恶性肿瘤或获得性免疫缺陷综合征 (acquired immunodeficiency syndrome, AIDS) 者; ③ 关键临床指标 (SOFA 评分、心肌标志物、生命体征) 资料缺失率 > 20%; ④ 多次入住 ICU 仅保留首次记录者; ⑤ 整体基线变量缺失超过 30% 者。

1.2 数据预处理与变量提取: 提取患者入住 ICU 后 24 h 内的基线数据, 包括 ① 人口学特征: 性别、年龄; ② 疾病严重程度评分: SOFA 评分、格拉斯哥昏迷评分 (Glasgow coma scale, GCS) 及 Charlson 合并症指数; ③ 治疗相关变量: 使用抗菌药物和血管加压药及行有创机械通气; ④ 既往合并症: 脑血管疾病、慢性肺部疾病、轻度肝脏疾病、重度肝脏疾病、糖尿病 (无并发症)、糖尿病 (伴并发症) 及肾脏疾病; ⑤ 生命体征: 心率 (heart rate, HR)、收缩压 (systolic blood pressure, SBP)、舒张压 (diastolic blood pressure, DBP)、平均动脉压 (mean arterial pressure, MAP)、呼吸频率 (respiratory rate, RR)、体温、尿量; ⑥ 血气分析指标: 动脉血 pH 值、动脉血氧分压 (partial pressure of oxygen, PaO₂)、动脉二氧化碳分压 (partial pressure of carbon dioxide, PaCO₂) 及脉搏血氧饱和度 (oxygen saturation, SpO₂); ⑦ 血液学指标: 血红蛋白 (hemoglobin, Hb)、血细胞比容 (hematocrit, HCT)、血小板计数 (platelet count, PLT) 及白细胞计数 (white blood cell count, WBC); ⑧ 代谢及肾功能指标: 血乳酸 (lactate acid, Lac)、剩余碱 (base excess, BE)、总二氧化碳、阴离子隙、尿素氮 (blood urea nitrogen, BUN)、肌酐 (creatinine, Cr) 及血糖; ⑨ 凝血功能指标: 凝血酶原时间 (prothrombin time, PT)、国际标准化比值 (international normalized ratio, INR) 及活化部分凝血活酶时间 (activated partial thromboplastin time, APTT); ⑩ 肝功能指标: 丙氨酸转氨酶 (alanine aminotransferase, ALT)、天冬氨酸转氨酶 (aspartate aminotransferase, AST)、碱性磷酸酶 (alkaline phosphatase, ALP) 及总胆红素 (total bilirubin, TBil); 结局指标为 28 d 死亡。对缺失率 ≤

30% 的变量采用 K 近邻算法 (k-nearest neighbor, KNN, k=5) 进行插补以最小化偏倚; 连续变量经稳健标准化处理后用于建模。分类变量编码为二元或多类形式。

1.3 特征选择: 为优化特征维度并提升模型泛化能力, 本研究构建两阶段特征选择框架 RFA-SEIFE, 包括 RFA 和 SEIFE^[18]。在 RFA 中, 使用 RF 模型 (n_estimators=200 表示基学习器数量; random_state=42 表示随机种子, 以保证结果可重复性) 计算变量重要度并按降序排列^[19]。利用袋外验证 (out-of-bag, OOB) 计算平均 AUC。当连续 5 次 AUC 提升幅度 < 0.001 时, 视为性能趋于稳定, 选取 AUC 最高的特征组合作为 RFA 最优子集。随后进入 SEIFE 阶段, 按特征重要性升序依次尝试剔除变量, 每次剔除后重新训练模型并计算 AUC。当 AUC 下降幅度不超过容差阈值 (最大下降幅度为 0.005) 时则剔除变量, 否则终止迭代, 以实现有效特征捕捉同时减少过拟合风险^[20]。

1.4 模型构建与性能验证: 基于 RFA-SEIFE 筛选的核心特征, 构建 LR、RF、XGBoost、LightGBM、多层感知机 (multilayer perceptron, MLP)、支持向量机 (support vector machine, SVM) 及 KNN 7 种机器学习模型。所有模型在五折分层交叉验证下进行训练与超参数优化。为缓解类别不平衡问题, 训练阶段应用合成少数类过采样技术 (synthetic minority oversampling technique, SMOTE) 并采用网格搜索优化超参数。外部验证采用独立数据集评估模型的泛化性能。为解析模型决策机制, 在外部验证集中使用 SHAP 算法计算每个变量的边际贡献。

1.5 统计学分析: 使用 Python 3.9 (Scikit-learn 1.3.0、XGBoost 1.7.5、SHAP 0.42.1、Statsmodels 0.14.0 等) 及 R 4.3.1 统计软件进行分析。连续变量符合正态分布以均数 ± 标准差 ($\bar{x} \pm s$) 表示, 两组间比较采用独立样本 *t* 检验; 连续变量不符合正态分布以中位数 (四分位数) [$M(Q_L, Q_U)$] 表示, 两组间比较采用 Mann-Whitney *U* 检验。分类变量以例 (率) 表示, 组间比较采用 χ^2 检验。双侧 $P < 0.05$ 为差异有统计学意义。模型性能评估指标包括 ROC 曲线下面积 AUC、精确率 - 召回率 (precision-recall, PR) AUC、F1 评分、敏感度、特异度及准确率。

2 结果

2.1 两组基线特征比较 (表 1): 本研究共纳入训练集 SRMI 患者 1 789 例, 验证集 1 117 例。总体人

群男性占 54.71% (1 590/2 960), 中位年龄为 62.00 (51.00, 75.00) 岁。在人口学及疾病严重程度方面, 与验证集比较, 训练集中男性比例明显降低, 年龄明显增大, SOFA 评分和 GCS 评分均明显升高 (均 $P < 0.05$), 而验证集和训练集 Charlson 合并症指数比较差异无统计学意义 ($P > 0.05$)。在治疗措施方面, 训练集抗菌药物使用率、血管加压药使用率及有创机械通气率均明显高于验证集 (均 $P < 0.05$)。在合并症方面, 训练集脑血管疾病和糖尿病 (伴并发症) 患者比例均明显高于验证集 (均 $P < 0.05$), 而轻度及重度肝脏疾病患者比例均明显低于验证集 (均 $P < 0.05$); 验证集和训练集合并慢性肺部疾病、糖尿病 (无并发症) 及肾脏疾病患者比例比较差异均无统计学意义 (均 $P > 0.05$)。在实验室指标方面, 训练集 BE、总二氧化碳、HCT、Hb、阴离子隙、Cr 及血糖水平均明显高于验证集 (均 $P < 0.05$), 而验证集 INR、PT、APTT、ALP 和 TBil 水平均明显高于训练集 (均 $P < 0.05$)。两组 Lac、PLT、WBC 及 BUN 水平比较差异均无统计学意义 (均 $P > 0.05$)。在生命体

征及血气指标方面, 验证集心率、体温、动脉血 pH 值及 PaO₂ 水平均明显高于训练集, 而训练集 DBP、PaCO₂ 及尿量明显高于验证集 (均 $P < 0.05$); 验证集和训练集 SBP、MAP、RR、SpO₂ 比较差异均无统计学意义 (均 $P > 0.05$)。在结局方面, 训练集病死率明显高于验证集 ($P < 0.05$)。

2.2 特征筛选结果 (图 1): 在 RFA 中, 随着特征数量逐步增加, 模型判别性能持续提升, 当纳入 26 个特征时 AUC 达到 0.762, 随后性能提升趋于平稳。在此基础上进入 SEIFE, 在容差阈值 (0.005) 范围内逐步剔除低贡献变量, 最终筛选得到 10 个核心特征, 对应 AUC=0.760。与 RFA 阶段相比, 特征数量减少 61.5%, 而模型判别性能仅轻微下降 ($\Delta AUC = 0.002$), 表明该筛选策略在有效降低特征维度的同时, 能保持模型稳定的区分能力。最终核心特征包括: Lac、尿量、体温、APTT、ALP、MAP、HCT、年龄、有创机械通气及使用抗菌药物。

2.3 模型性能评估: 基于上述 10 个核心特征, 分别构建 7 种机器学习模型, 模型在五折交叉验证下

表 1 基于机器学习的脓毒症相关心肌损伤患者基线特征

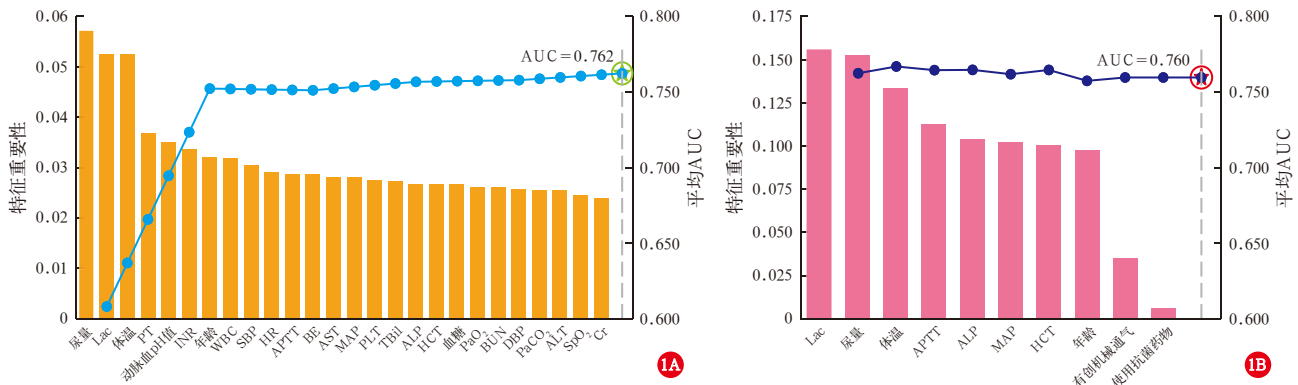
变量	训练集 (1 789 例)	验证集 (1 117 例)	$\chi^2/Z/t$ 值	P 值	变量	训练集 (1 789 例)	验证集 (1 117 例)	Z/t/ χ^2 值	P 值
性别 [例 (%)]			8.854	0.003	阴离子隙 [mmol/L, $M(Q_L, Q_U)$]	18.00 (15.00, 22.00)	17.00 (14.00, 20.00)	8.090	<0.001
男性	940 (52.54)	650 (58.19)			BUN [mmol/L, $M(Q_L, Q_U)$]	29.00 (19.00, 47.00)	29.00 (17.00, 48.00)	1.346	0.178
女性	849 (47.46)	647 (41.81)			Cr [$\mu\text{mol/L}, M(Q_L, Q_U)$]	132.6 (88.6, 238.7)	123.8 (79.6, 221.0)	2.216	0.027
年龄 [岁, $M(Q_L, Q_U)$]	64.00 (52.00, 76.00)	60.53 (50.47, 70.91)	4.102	<0.001	INR [$M(Q_L, Q_U)$]	1.40 (1.20, 1.80)	1.50 (1.30, 2.00)	-5.987	<0.001
SOFA 评分 [分, $M(Q_L, Q_U)$]	3.00 (2.00, 5.00)	6.00 (4.00, 10.00)	-20.807	<0.001	PT [s, $M(Q_L, Q_U)$]	15.10 (13.10, 19.30)	16.40 (14.30, 21.10)	-8.082	<0.001
GCS 评分 (分, $\bar{x} \pm s$)	12.92 \pm 3.41	13.37 \pm 2.87	-3.835	<0.001	APTT [s, $M(Q_L, Q_U)$]	34.80 (29.20, 48.40)	36.60 (29.90, 49.10)	-2.433	0.015
Charlson 合并症指数 [$M(Q_L, Q_U)$]	4.00 (2.00, 5.00)	4.00 (2.00, 5.00)	1.785	0.074	ALT [U/L, $M(Q_L, Q_U)$]	44.28 (22.00, 105.00)	48.38 (26.00, 102.00)	-1.332	0.183
使用抗菌药物 [例 (%)]	1 710 (95.58)	909 (81.38)	155.902	<0.001	ALP [U/L, $M(Q_L, Q_U)$]	91.90 (67.00, 132.00)	102.00 (71.00, 150.00)	-4.417	<0.001
使用血管加压药 [例 (%)]	904 (50.53)	496 (44.40)	10.337	0.001	AST [U/L, $M(Q_L, Q_U)$]	69.37 (35.00, 176.65)	69.00 (37.55, 138.00)	0.979	0.327
有创机械通气 [例 (%)]	1 108 (61.93)	544 (48.70)	49.081	<0.001	TBil [$\mu\text{mol/L}, M(Q_L, Q_U)$]	13.68 (6.84, 28.98)	29.07 (11.97, 58.85)	-13.811	<0.001
脑血管疾病 [例 (%)]	280 (15.65)	100 (8.95)	27.147	<0.001	心率 (次/min, $\bar{x} \pm s$)	90.34 \pm 17.54	94.52 \pm 17.61	-6.243	<0.001
慢性肺部疾病 [例 (%)]	364 (20.35)	209 (18.71)	1.162	0.281	SBP (mmHg, $\bar{x} \pm s$)	85.42 \pm 19.07	84.88 \pm 16.72	0.795	0.427
轻度肝脏疾病 [例 (%)]	363 (20.29)	338 (30.26)	37.336	<0.001	DBP (mmHg, $\bar{x} \pm s$)	43.673 \pm 12.519	42.710 \pm 11.531	2.100	0.036
糖尿病 (无并发症) [例 (%)]	390 (21.80)	265 (23.72)	1.459	0.227	MAP (mmHg, $\bar{x} \pm s$)	54.06 \pm 15.92	52.94 \pm 14.79	1.925	0.054
糖尿病 (伴并发症) [例 (%)]	141 (7.88)	61 (5.46)	6.229	0.013	RR (次/min, $\bar{x} \pm s$)	13.24 \pm 4.26	13.27 \pm 4.20	-0.183	0.855
肾脏疾病 [例 (%)]	374 (20.91)	205 (18.35)	2.809	0.094	体温 ($^{\circ}\text{C}, \bar{x} \pm s$)	37.51 \pm 1.02	37.74 \pm 0.99	-5.945	<0.001
重度肝脏疾病 [例 (%)]	161 (9.00)	178 (15.94)	32.104	<0.001	动脉血 pH 值 ($\bar{x} \pm s$)	7.26 \pm 0.14	7.30 \pm 0.12	-6.058	<0.001
Lac [mmol/L, $M(Q_L, Q_U)$]	2.60 (1.70, 4.60)	2.50 (1.70, 4.20)	1.816	0.069	PaO ₂ [mmHg, $M(Q_L, Q_U)$]	60.00 (41.94, 87.00)	73.00 (56.42, 92.00)	-9.644	<0.001
BE [mmol/L, $M(Q_L, Q_U)$]	0.00 (-3.00, 1.06)	-1.00 (-4.00, 0.77)	3.952	<0.001	PaCO ₂ (mmHg, $\bar{x} \pm s$)	49.99 \pm 15.94	45.39 \pm 13.73	8.259	<0.001
总二氧化碳 (mmol/L, $\bar{x} \pm s$)	25.48 \pm 5.50	24.39 \pm 5.01	5.357	<0.001	SpO ₂ ($\bar{x} \pm s$)	0.901 \pm 0.094	0.902 \pm 0.095	-0.472	0.637
HCT ($\bar{x} \pm s$)	0.300 \pm 0.069	0.288 \pm 0.060	5.044	<0.001	血糖 [mmol/L, $M(Q_L, Q_U)$]	9.77 (7.50, 13.6)	8.89 (6.90, 11.89)	0.013	<0.001
Hb (g/L, $\bar{x} \pm s$)	98.91 \pm 23.46	95.93 \pm 20.04	3.651	<0.001	尿量 [mL, $M(Q_L, Q_U)$]	1 225.00 (599.00, 2 155.00)	1 377.00 (738.00, 2 295.00)	-2.985	0.003
PLT [$\times 10^9/L$, $M(Q_L, Q_U)$]	164.00 (102.00, 233.00)	169.00 (101.00, 262.00)	-1.599	0.110	死亡 [例 (%)]	702 (39.24)	366 (32.77)	12.178	<0.001
WBC [$\times 10^9/L$, $M(Q_L, Q_U)$]	14.80 (10.40, 20.70)	15.00 (9.70, 21.10)	0.642	0.521					

注: 1 mmHg \approx 0.133 kPa

的内部验证结果见表 2。在交叉验证中, RF 模型 AUC=0.758, 95% 可信区间(95% confidence interval, 95%CI)为 0.736 ~ 0.779, PR AUC=0.693, 95%CI 为 0.676 ~ 0.710; LightGBM 模型表现次之。外部验证结果显示, RF 模型的整体性能最优, AUC 达 0.837, PR AUC 为 0.757, F1 评分为 0.688, 均优于其余模型(图 2A 和图 2B)。决策曲线分析(decision curve analysis, DCA)显示(图 2C), 在临床可操作的阈值概率范围内, 模型的净获益明显高于“全部干预”或“无干预”方案。总体混淆矩阵显示(图 2D), 模型对死亡与存活患者均具有良好的识别能力。校准曲线表明(图 2E), 预测概率与实际事件发生率高度一致, 模型校准度良好。风险分层分析显示(图 2F), 事件率随预测风险十分位递增, 且内部与外部验证结果趋势一致, 提示模型具备稳定的分层判别能力。Kolmogorov-Smirnov (KS) 曲线分析显示(图 2G), 内

部验证 KS 值为 0.373, 外部验证 KS 值为 0.518, 说明模型区分度较好。预测概率分布图(图 2H)呈现正负样本明显的双峰分离特征, 表明模型输出概率有良好的可解释性。临床影响曲线(clinical impact curve, CIC)分析进一步表明(图 2I), 模型在不同风险阈值下均能稳定识别高风险人群, 有良好的临床可用性。综合模型的判别力、校准性与临床应用价值, 最终确定 RF 模型作为本研究的最优预测模型。

2.4 SHAP 分析(图 3): SHAP 特征重要性分析显示(图 3A), 尿量和有创机械通气对模型输出的影响最大, 其次为年龄、Lac、体温、APTT、ALP、MAP、HCT 及使用抗菌药物。SHAP 蜂群图显示(图 3B), 尿量和有创机械通气的 SHAP 值分布范围最广, 提示其对个体预测结果的影响程度及变异性均较大。决策路径图显示(图 3C), 高风险个体的 SHAP 累积值整体向正向偏移; 瀑布图进一步展示典型个体的



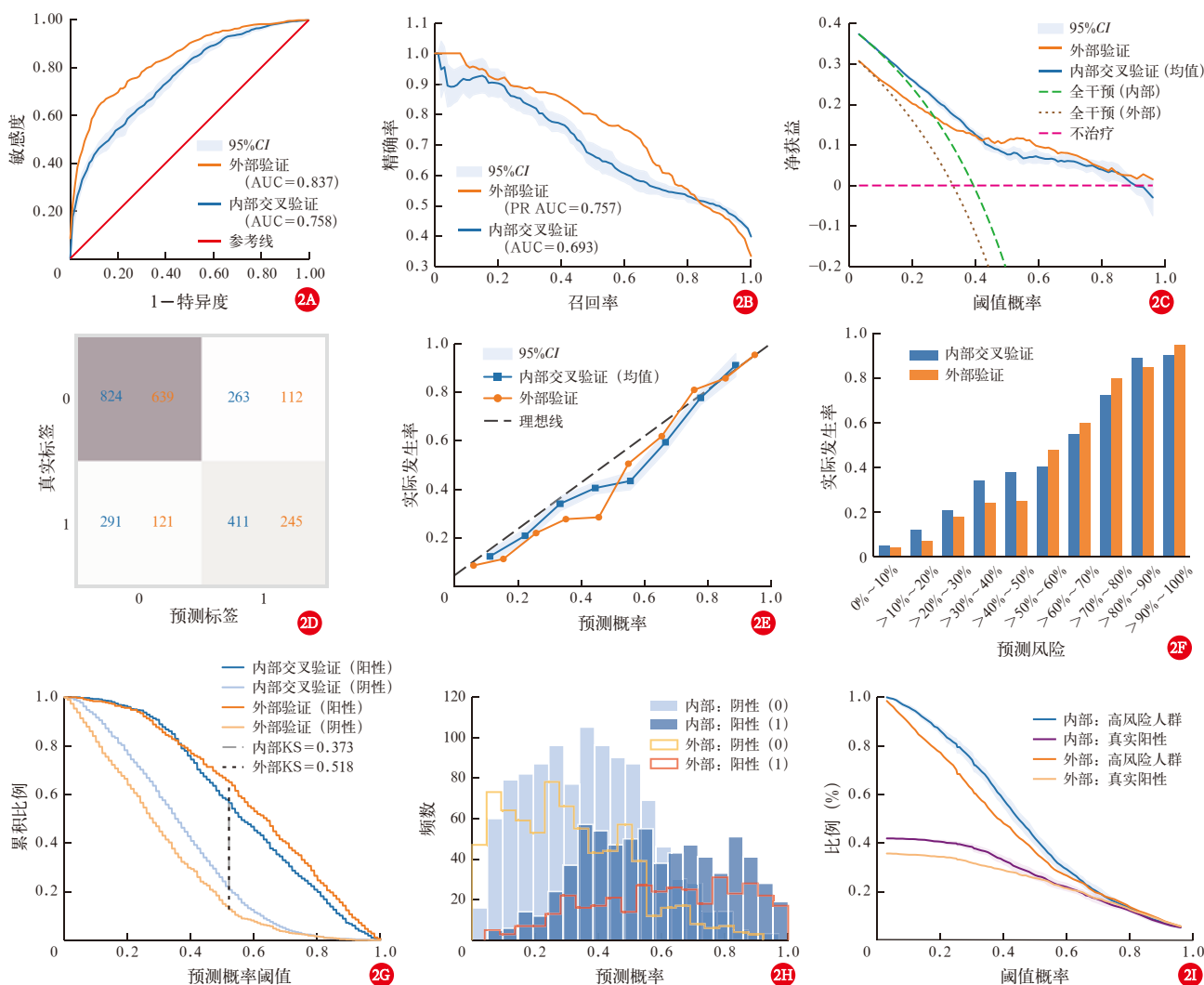
注: A 为 RFA, B 为 SEIFE; 柱状图表示特征重要性, 折线表示不同特征数量下模型 AUC 变化, 虚线表示最优特征数量

图 1 RF 变量筛选框架及性能变化趋势

表 2 不同机器学习模型在交叉验证集与外部验证集中的性能比较

模型	数据集	准确率 [均值(95%CI)]	平衡准确率 [均值(95%CI)]	AUC [均值(95%CI)]	PR AUC [均值(95%CI)]	F1 值 [均值(95%CI)]	召回率 [均值(95%CI)]	精确率 [均值(95%CI)]
LR	交叉验证集	0.674 (0.660 ~ 0.689)	0.669 (0.649 ~ 0.688)	0.733 (0.711 ~ 0.755)	0.662 (0.630 ~ 0.693)	0.607 (0.580 ~ 0.633)	0.642 (0.598 ~ 0.687)	0.576 (0.562 ~ 0.589)
	外部验证集	0.724	0.701	0.770	0.597	0.601	0.634	0.571
RF	交叉验证集	0.692 (0.683 ~ 0.700)	0.671 (0.658 ~ 0.684)	0.758 (0.736 ~ 0.779)	0.693 (0.676 ~ 0.710)	0.594 (0.573 ~ 0.615)	0.577 (0.536 ~ 0.618)	0.614 (0.602 ~ 0.626)
	外部验证集	0.800	0.768	0.837	0.757	0.688	0.672	0.705
XGBoost	交叉验证集	0.680 (0.667 ~ 0.693)	0.664 (0.648 ~ 0.680)	0.735 (0.719 ~ 0.751)	0.671 (0.644 ~ 0.697)	0.591 (0.568 ~ 0.615)	0.591 (0.558 ~ 0.624)	0.592 (0.576 ~ 0.607)
	外部验证集	0.755	0.730	0.818	0.729	0.638	0.659	0.618
LightGBM	交叉验证集	0.680 (0.664 ~ 0.696)	0.664 (0.646 ~ 0.683)	0.745 (0.731 ~ 0.759)	0.677 (0.657 ~ 0.697)	0.592 (0.567 ~ 0.616)	0.593 (0.557 ~ 0.627)	0.592 (0.572 ~ 0.611)
	外部验证集	0.764	0.740	0.817	0.736	0.651	0.672	0.631
MLP	交叉验证集	0.670 (0.647 ~ 0.694)	0.659 (0.628 ~ 0.690)	0.734 (0.710 ~ 0.757)	0.664 (0.649 ~ 0.680)	0.589 (0.546 ~ 0.632)	0.607 (0.537 ~ 0.676)	0.574 (0.552 ~ 0.597)
	外部验证集	0.714	0.693	0.778	0.619	0.592	0.631	0.557
SVM	交叉验证集	0.679 (0.674 ~ 0.683)	0.673 (0.662 ~ 0.683)	0.744 (0.724 ~ 0.763)	0.660 (0.638 ~ 0.682)	0.611 (0.590 ~ 0.631)	0.645 (0.594 ~ 0.696)	0.582 (0.573 ~ 0.591)
	外部验证集	0.725	0.712	0.778	0.634	0.617	0.675	0.568
KNN	交叉验证集	0.639 (0.621 ~ 0.657)	0.634 (0.608 ~ 0.660)	0.681 (0.662 ~ 0.700)	0.532 (0.515 ~ 0.549)	0.568 (0.527 ~ 0.609)	0.610 (0.540 ~ 0.679)	0.534 (0.514 ~ 0.553)
	外部验证集	0.699	0.684	0.739	0.533	0.582	0.639	0.534

注: 交叉验证结果为五折交叉验证的均值(95%CI); 外部验证结果来源于独立验证集的性能表现。F1 值为精确率与召回率的调和平均值



注: A 为 ROC 曲线用于评估模型判别能力,内部交叉验证 AUC 为 0.758(95%CI 为 0.736~0.779),外部验证 AUC 为 0.837; B 为 PR 曲线内部交叉验证 PR AUC 为 0.693(95%CI 为 0.679~0.710),外部验证 PR AUC 为 0.757; C 为 DCA 用于评估不同阈值概率下模型的临床净获益; D 为混淆矩阵:蓝色表示内部交叉验证结果,橙色表示外部验证结果; E 为校准曲线用于评估模型预测概率与实际发生率的一致性; F 为风险分层分析:不同预测风险分位区间下事件发生率呈递增趋势; G 为 KS 曲线,内部交叉验证 KS 值为 0.373,外部验证 KS 值为 0.518; H 为预测概率分布:阳性与阴性样本预测概率分布呈明显分离; I 为 CIC 用于评估不同阈值下高风险人群识别及真实事件发生情况

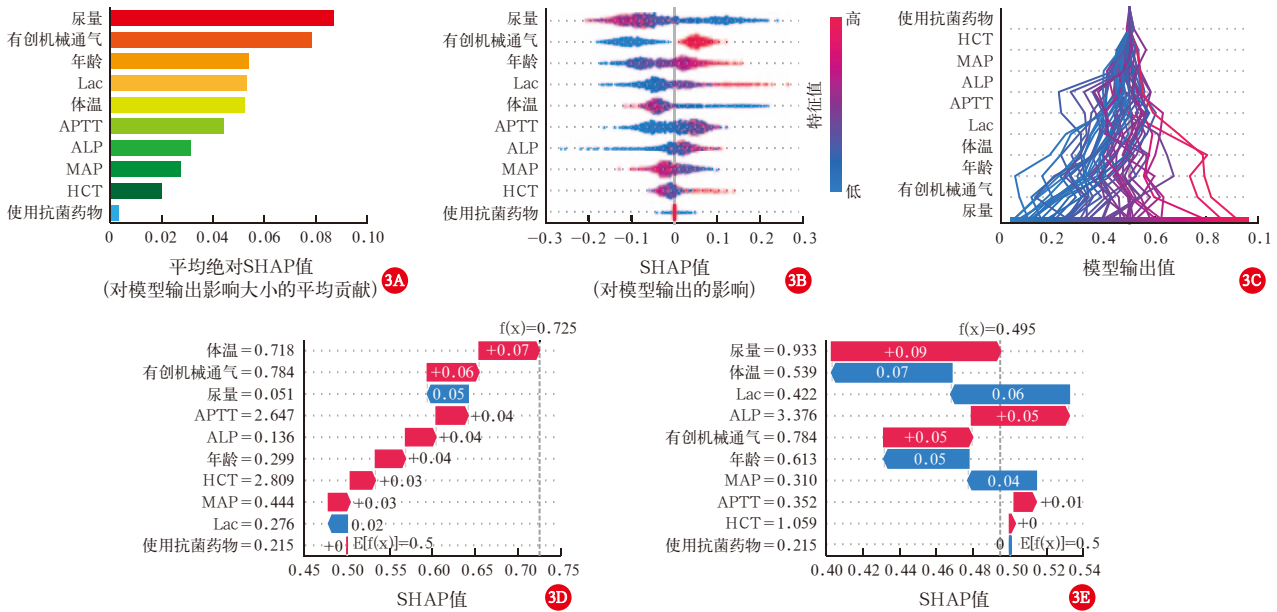
图 2 RF 模型的性能评估与验证结果

特征贡献情况(图 3D 和图 3E):在生存患者中,尿量增加对预测结果产生负向贡献(降低死亡风险);而在死亡患者中,尿量、ALP 及有创机械通气等变量对预测结果主要产生正向贡献,而体温及 Lac 等变量则表现为负向贡献。综上,尿量减少、Lac 升高及有创机械通气与死亡风险增加密切相关。

3 讨论

本研究基于 MIMIC-III v1.4 与 MIMIC-IV v2.2 两个公开数据库,构建并验证了针对 SRMI 患者的短期死亡风险预测模型。结果显示,RF 模型在交叉与外部验证中均表现出较好的判别力与校准性,其 AUC 分别为 0.758 与 0.837, PR AUC 分别为 0.693 与 0.757, F1 值为 0.688,性能稳定、可解释性良好。

模型仅依赖入住 ICU 24 h 常规指标即可实现有效预后分层,为 SRMI 患者的早期风险分层提供新的工具。SRMI 的发生涉及炎症反应、循环灌注不足及代谢紊乱等多重病理过程^[21-23]。本研究基于 SHAP 可解释性分析表明,模型筛选出的 10 个核心特征在临床与生理层面均有明确的病理学意义,其变化方向与 SRMI 的病程演变高度一致^[5]。其中,尿量与 Lac 的平均绝对 SHAP 值最高,是死亡风险最主要的判别因子。尿量减少提示有效循环血容量不足与肾脏低灌注,是反映微循环障碍与组织氧供不足的早期信号^[24];Lac 升高反映氧利用障碍与线粒体功能损伤,与心肌缺血、代谢应激和能量耗竭密切相关^[25-27]。两者均异常往往标志着循环衰竭的持



注：A 为 SHAP 特征重要性排序（显示各变量对模型预测结果的平均绝对贡献）；B 为 SHAP 蜂群图（展示各特征在不同取值下对模型输出的影响方向及分布范围）；C 为决策路径图（反映高风险个体中各特征对预测结果的累积贡献）；D 和 E 为瀑布图（展示典型生存与死亡个体中各变量对预测结果的正负贡献）

图 3 SHAP 可解释性分析图

续存在,与脓毒性休克和多器官功能障碍综合征的临床演变一致^[28]。APTT 延长与 MAP 降低则提示凝血功能障碍与血流动力学不稳定之间存在协同放大效应:脓毒症患者凝血系统激活可导致微血栓形成与内皮损伤,进一步加剧组织灌注不足,而持续低灌注又可促进凝血因子消耗与纤溶亢进,形成恶性循环^[29-30]。ALP 升高反映肝细胞功能障碍及胆汁淤积状态,提示脓毒症相关多器官损伤的肝脏受累,是全身炎症反应与肝微循环障碍的重要信号^[31-32]。HCT 降低在脓毒症中常见,可反映稀释性贫血、红细胞破坏或骨髓抑制等综合效应,影响氧输送能力与组织灌注稳定性^[5]。体温的双向作用亦符合临床观察:适度升高提示免疫反应活跃,低体温则多与免疫抑制和预后不良相关^[33]。年龄作为传统预后指标之一,与免疫功能、心肌储备及器官耐受性密切相关,年龄越大,SRMI 患者的死亡风险越高^[23]。在治疗相关变量中,有创机械通气与使用抗菌药物同样被模型识别为重要特征。有创通气既是病情严重的标志,也可能通过改善氧合与减少心脏负荷发挥短期保护作用,但同时伴随呼吸机相关感染及循环抑制等风险^[34]。使用抗菌药物则反映感染控制强度与治疗时机的差异,其在模型中的重要性提示“早期识别—及时抗感染”仍是降低死亡风险的关键环节^[35]。

本研究 RF 外部验证集的模型性能略高于交

叉验证结果(AUC:0.837 比 0.758)。这一现象在机器学习研究中并不罕见,可能源于外部验证样本分布更集中、事件率及特征变异更有利于风险区分^[36]。RF 通过随机特征抽样与袋外验证,可有效抑制过拟合;因此外部性能更优并不意味着“学习了噪声”,反而体现出模型对不同样本分布的稳健泛化能力^[37]。此外,五折交叉验证包含更多异质性样本组合,其 AUC 略低反映的是模型在复杂情境下的平均性能,而非模型不稳定^[38]。

需要注意的是,训练集与外部验证集之间部分基线特征差异有统计学意义,这与 MIMIC-III 与 MIMIC-IV 在时间跨度、病例选择及数据结构上的差异密切相关。尽管基线不平衡可能引入选择偏倚,但模型在验证集上仍保持良好的判别力与校准度,提示其具备一定的稳健性与迁移能力^[39]。未来研究可通过多中心外部验证、倾向评分加权(inverse probability of treatment weighting, IPTW)或再平衡采样等方法进一步检验模型的稳定性与普适性^[40]。

本研究仍存在若干局限。首先,数据均来自于单中心的公开数据库,尽管进行了跨版本外部验证,但尚缺乏不同地区、不同医疗体系人群的多中心验证。其次,研究仅纳入入住 ICU 24 h 内的静态基线数据,未能捕捉关键变量的动态变化,可能低估病情演变对风险的影响。第三,部分变量存在缺失,虽采用 KNN 插补以减轻偏倚,但仍可能影响模型稳定

性。最后,模型的临床应用效果尚未通过前瞻性试验验证,未来研究应结合动态时间序列建模、再校准及真实世界干预试验,以进一步评估其在临床流程中的决策增益^[41]。

利益冲突 所有作者均声明不存在利益冲突

参考文献

- [1] Singer M, Deutschman CS, Seymour CW, et al. The Third International Consensus Definitions for sepsis and septic shock (Sepsis-3) [J]. *JAMA*, 2016, 315 (8): 801-810. DOI: 10.1001/jama.2016.0287.
- [2] Rudd KE, Johnson SC, Agesa KM, et al. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the Global Burden of Disease Study [J]. *Lancet*, 2020, 395 (10219): 200-211. DOI: 10.1016/S0140-6736(19)32989-7.
- [3] Wang RZ, Xu YR, Fang YX, et al. Pathogenetic mechanisms of septic cardiomyopathy [J]. *J Cell Physiol*, 2022, 237 (1): 49-58. DOI: 10.1002/jcp.30527.
- [4] 中国医疗保健国际交流促进会急诊医学分会, 中华医学会急诊医学分会, 国家老年医学中心(北京医院), 等. 成人脓毒症相关心肌损伤和(或)心功能障碍: 急诊专家共识 [J]. *中国急救医学*, 2025, 45 (9): 737-747. DOI: 10.3969/j.issn.1002-1949.2025.09.001.
- [5] 王琳, 孙媛, 李筱妍. 脓毒症心肌病生物标志物的研究进展 [J]. *临床检验杂志*, 2023, 41 (12): 923-926. DOI: 10.13602/j.cnki.jcls.2023.12.10.
- [6] Xu KZ, Xu P, Li JJ, et al. Predictors and nomogram of in-hospital mortality in sepsis-induced myocardial injury: a retrospective cohort study [J]. *BMC Anesthesiol*, 2023, 23 (1): 230. DOI: 10.1186/s12871-023-02189-8.
- [7] Cecconi M, Evans L, Levy M, et al. Sepsis and septic shock [J]. *Lancet*, 2018, 392 (10141): 75-87. DOI: 10.1016/S0140-6736(18)30696-2.
- [8] Gajardo A, Ferrière-Steinert S, Valenzuela Jiménez J, et al. Early high-sensitivity troponin elevation and short-term mortality in sepsis: a systematic review with meta-analysis [J]. *Crit Care*, 2025, 29 (1): 76. DOI: 10.1186/s13054-025-05249-2.
- [9] Raith EP, Udy AA, Bailey M, et al. Prognostic accuracy of the SOFA score, SIRS criteria, and qSOFA score for in-hospital mortality among adults with suspected infection admitted to the intensive care unit [J]. *JAMA*, 2017, 317 (3): 290-300. DOI: 10.1001/jama.2016.20328.
- [10] Gursel G, Demirtas S. Value of APACHE II, SOFA and CPIS scores in predicting prognosis in patients with ventilator-associated pneumonia [J]. *Respiration*, 2006, 73 (4): 503-508. DOI: 10.1159/000088708.
- [11] Fleuren LM, Klausch T, Zwager CL, et al. Machine learning for the prediction of sepsis: a systematic review and meta-analysis of diagnostic test accuracy [J]. *Intensive Care Med*, 2020, 46 (3): 383-400. DOI: 10.1007/s00134-019-05872-y.
- [12] Bomrah S, Uddin M, Upadhyay U, et al. A scoping review of machine learning for sepsis prediction-feature engineering strategies and model performance: a step towards explainability [J]. *Crit Care*, 2024, 28 (1): 180. DOI: 10.1186/s13054-024-04948-6.
- [13] Mao QQ, Jay M, Hoffman JL, et al. Multicentre validation of a sepsis prediction algorithm using only vital sign data in the emergency department, general ward and ICU [J]. *BMJ Open*, 2018, 8 (1): e017833. DOI: 10.1136/bmjopen-2017-017833.
- [14] Gao JY, Lu YY, Ashrafi N, et al. Prediction of sepsis mortality in ICU patients using machine learning methods [J]. *BMC Med Inform Decis Mak*, 2024, 24 (1): 11. DOI: 10.1186/s12911-024-02630-z.
- [15] Wang SR, Liu XY, Yuan SH, et al. Artificial intelligence based multispecialty mortality prediction models for septic shock in a multicenter retrospective study [J]. *NPJ Digit Med*, 2025, 8 (1): 228. DOI: 10.1038/s41746-025-01643-w.
- [16] Johnson AE, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database [J]. *Sci Data*, 2016, 3: 160035. DOI: 10.1038/sdata.2016.35.
- [17] Balk RA. Severe sepsis and septic shock. Definitions, epidemiology, and clinical manifestations [J]. *Crit Care Clin*, 2000, 16 (2): 179-192. DOI: 10.1016/s0749-0704(05)70106-8.
- [18] Guyon I, Elisseeff A. An introduction to variable and feature selection [J]. *J Mach Learn Res*, 2003, 3: 1157-1182.
- [19] Breiman L. Random forests [J]. *Mach Learn*, 2001, 45 (1): 5-32. DOI: 10.1023/A:1010933404324.
- [20] Tao QL, Yang XT, Bao LK, et al. Transforming machine learning model knowledge into material insights for multi-principal-element superalloy phase design [J]. *npj Comput. Mater*, 2025, 11: 99. DOI: 10.1038/s41524-025-01578-6.
- [21] 庄燕, 张海东, 戴林峰, 等. 脓毒症心肌功能障碍危险因素及临床预测模型的构建 [J]. *中国中西医结合急救杂志*, 2022, 29 (3): 262-267. DOI: 10.3969/j.issn.1008-9691.2022.03.002.
- [22] 甄军海, 李莉, 严静. 脓毒症心肌损伤生物标志物的研究进展 [J]. *中华危重病急救医学*, 2018, 30 (7): 699-702. DOI: 10.3760/cma.j.issn.2095-4352.2018.07.017.
- [23] 周智恩, 严善秀, 卢萍, 等. 影响脓毒症患者发生心肌功能障碍的危险因素及临床特征研究 [J]. *中国中西医结合急救杂志*, 2024, 31 (4): 428-432. DOI: 10.3969/j.issn.1008-9691.2024.04.008.
- [24] Vincent JL, De Backer D. Circulatory shock [J]. *N Engl J Med*, 2013, 369 (18): 1726-1734. DOI: 10.1056/NEJMra1208943.
- [25] Oliva PB. Lactic acidosis [J]. *Am J Med*, 1970, 48 (2): 209-225. DOI: 10.1016/0002-9343(70)90117-8.
- [26] 温良鹤, 权震, 李云龙, 等. 早期输注艾司洛尔通过降低血乳酸抑制心肌细胞焦亡改善脓毒症心肌病 [J]. *内科急危重症杂志*, 2024, 30 (6): 501-507. DOI: 10.11768/nkjwzzzz20240604.
- [27] 尹言, 于君, 王征, 等. 脓毒症心肌病短期预后及其影响因素研究 [J]. *中国循证心血管医学杂志*, 2022, 14 (4): 451-453, 464. DOI: 10.3969/j.issn.1674-4055.2022.04.16.
- [28] Angus DC, van der Poll T. Severe sepsis and septic shock [J]. *N Engl J Med*, 2013, 369 (21): 2063. DOI: 10.1056/NEJMc1312359.
- [29] Levi M. Coagulation and sepsis: a winding road ahead [J]. *Crit Care Med*, 2012, 40 (9): 2733-2734. DOI: 10.1097/CCM.0b013e31825ce4e8.
- [30] Iba T, Levy JH. Sepsis-induced coagulopathy and disseminated intravascular coagulation [J]. *Anesthesiology*, 2020, 132 (5): 1238-1245. DOI: 10.1097/ALN.0000000000003122.
- [31] 管建慧, 易钊泉. 脓毒症肝损伤的致病机制及治疗新进展 [J]. *中国急救医学*, 2019, 39 (6): 601-604. DOI: 10.3969/j.issn.1002-1949.2019.06.017.
- [32] Zhang X, Liu H, Hashimoto K, et al. The gut-liver axis in sepsis: interaction mechanisms and therapeutic potential [J]. *Crit Care*, 2022, 26 (1): 213. DOI: 10.1186/s13054-022-04090-1.
- [33] Thomas-Rüddel DO, Hoffmann P, Schwarzkopf D, et al. Fever and hypothermia represent two populations of sepsis patients and are associated with outside temperature [J]. *Crit Care*, 2021, 25 (1): 368. DOI: 10.1186/s13054-021-03776-2.
- [34] 李雪珍, 甘伟妮, 周开焕, 等. 脓毒症心肌抑制患者血流动力学、器官功能及临床预后观察研究 [J]. *岭南急诊医学杂志*, 2024, 29 (3): 215-217, 222. DOI: 10.3969/j.issn.1671-301X.2024.03.002.
- [35] Kollef MH, Shorr AF, Bassetti M, et al. Timing of antibiotic therapy in the ICU [J]. *Crit Care*, 2021, 25 (1): 360. DOI: 10.1186/s13054-021-03787-z.
- [36] Moons K, Wolff RF, Riley RD, et al. PROBAST: a tool to assess risk of bias and applicability of prediction model studies: explanation and elaboration [J]. *Ann Intern Med*, 2019, 170 (1): W1-W33. DOI: 10.7326/M18-1377.
- [37] Cutler A, Cutler DR, Stevens JR. Random forests [M]/Zhang C, Ma Y. Ensemble machine learning: methods and applications. New York: Springer, 2012: 157-175.
- [38] Steyerberg EW. Clinical prediction models: a practical approach to development, validation, and updating [M]. 2nd ed. New York: Springer, 2019.
- [39] Guo LL, Pfohl SR, Fries J, et al. Systematic review of approaches to preserve machine learning performance in the presence of temporal dataset shift in clinical medicine [J]. *Appl Clin Inform*, 2021, 12 (4): 808-815. DOI: 10.1055/s-0041-1735184.
- [40] Topol EJ. High-performance medicine: the convergence of human and artificial intelligence [J]. *Nat Med*, 2019, 25 (1): 44-56. DOI: 10.1038/s41591-018-0300-7.
- [41] Futoma J, Simons M, Panch T, et al. The myth of generalisability in clinical research and machine learning in health care [J]. *Lancet Digit Health*, 2020, 2 (9): e489-e492. DOI: 10.1016/S2589-7500(20)30186-2.

(收稿日期: 2026-01-15)

(责任编辑: 邸美仙)