·论　著·

可解释机器学习模型预测心脏骤停患者院内死亡风险：基于 MIMIC-Ⅳ 2.0 数据库

龚欢欢，柯晓伟，王爱民，李湘民

中南大学湘雅医院急诊科，长沙 410008

通信作者：王爱民，E-mail：wangaimin@csu.edu.cn
李湘民，E-mail：lxm8229@csu.edu.cn

【摘要】目的　构建可预测心脏骤停患者住院期间死亡风险的机器学习模型，并对其进行解释。**方法**　提取美国重症监护医学信息数据库Ⅳ（Medical Information Mart for Intensive Care database Ⅳ，MIMIC-Ⅳ）2.0 中心脏骤停患者转入 ICU 24 h 内首次临床资料及住院期间转归，基于机器学习算法构建 6 种可预测心脏骤停患者院内死亡风险的模型，包括 XGBoost模型、轻量级梯度提升机（light gradient boosting machine，LGBM）模型、决策树（decision tree，DT）模型、K 近邻（K-nearest neighbor，KNN）模型、Logistic 回归模型、随机森林（random forest，RF）模型。采用受试者操作特征（receiver operator characteristic，ROC）曲线、临床决策曲线及校准曲线对模型进行评价，并采用 Shapley 加性解释（Shapley additive explanation，SHAP）算法评估不同临床特征对最优模型的影响，以增加模型的可解释性。**结果**　共 1465 例符合纳入与排除标准的心脏骤停患者入选本研究。其中住院期间存活 773 例、死亡 692 例。经筛选，共纳入 82 个临床特征用于机器学习模型构建。模型评价结果显示，相较于其余 5 种模型，LGBM 模型预测心脏骤停患者院内死亡的曲线下面积（area under the curve，AUC）更高［0.834（95% CI：0.688～0.894）］，且相对于 Logistic 回归模型、XGBoost 模型，其对死亡风险的预测准确性更高（校准度：0.166），临床决策性能更优，整体性能最佳。SHAP 算法分析显示，对 LGBM 模型输出结果影响最大的 3 个临床特征分别为格拉斯哥睁眼反应评分、碳酸氢盐水平、白细胞计数。**结论**　基于大型公共医疗卫生数据库建立的可预测心脏骤停患者住院期间死亡风险的机器学习模型中，LGBM 模型性能最优，其可辅助临床进行更高效的疾病管理和更精准的医疗干预。

【关键词】心脏骤停；预测模型；机器学习；SHAP 算法；美国重症监护医学信息数据库
【中图分类号】R541.7+8　　**【文献标志码】**A　　**【文章编号】**1674-9081(2023)03-0528-08
DOI：10.12290/xhyxzz.2022-0733

An Interpretable Machine Learning Model for Predicting In-hospital Death Risk in Patients with Cardiac Arrest: Based on US Medical Information Mart for Intensive Care Database Ⅳ 2.0

GONG Huanhuan, KE Xiaowei, WANG Aimin, LI Xiangmin

Department of Emergency, Xiangya Hospital, Central South University, Changsha 410008, China

Corresponding authors: WANG Aimin, E-mail: wangaimin@csu.edu.cn
LI Xiangmin, E-mail: lxm8229@csu.edu.cn

【Abstract】Objective　To develop and validate an interpretable machine learning model based on clinical

龚欢欢、柯晓伟对本文同等贡献

基金项目：湖南省自然科学基金（2022JJ30938，2022JJ70165）
引用本文：龚欢欢，柯晓伟，王爱民，等. 可解释机器学习模型预测心脏骤停患者院内死亡风险：基于 MIMIC-Ⅳ 2.0 数据库［J］. 协和医学杂志，2023，14（3）：528-535. doi：doi：10.12290/xhyxzz.2022-0733.

characteristics to predict the risk of in-hospital death in patients with cardiac arrest. **Methods** First clinical data of cardiac arrest patients admitted to ICU within 24 h and outcomes during hospitalization were extracted from Medical Information Mart for Intensive Care database Ⅳ (MIMIC-Ⅳ) 2.0. Six models predicting in-hospital death risk of cardiac arrest patients were constructed based on machine learning algorithm: XGBoost model, light gradient boosting machine (LGBM) model, decision tree (DT) model, K-nearest neighbor (KNN) model, Logistic regression model, and random forest (RF) model. Receiver operator characteristic (ROC) curve, clinical decision curve and calibration curve were used to evaluate the 6 models. Shapley additive explanation (SHAP) algorithm was used to explain and evaluate the effects of different clinical features on the optimal model to increase its interpretability. **Results** A total of 1465 patients with cardiac arrest who met inclusion and exclusion criteria were included in the study. Among them, 773 patients survived and 692 died during hospitalization. After screening, a total of 82 clinical features were included for machine learning model construction. Compared with the other five models, the LGBM model had a higher area under the curve for predicting in-hospital death in cardiac arrest patients [0.834 (95% CI: 0.688-0.894)], higher prediction accuracy for the risk of death than the Logistic regression model and XGBoost model (calibration degree: 0.166), better clinical decision performance, and displayed optimal overall performance. SHAP algorithm analysis showed that the three clinical features that had the greatest impact on the output of LGBM model were Glasgow eyes score, bicarbonate level and white blood cell count. **Conclusion** Based on a large public medical and health database, a machine learning model named LGBM has the best performance to predict the risk of in-hospital death in patients with cardiac arrest, which will be helpful to assist more efficient clinical disease management and more precise medical intervention.

【Key words】 cardiac arrest; prognostic prediction; machine learning; SHAP algorithms; Medical Information Mart for Intensive Care database

Funding: Natural Science Foundation of Hunan Province (2022JJ30938, 2022JJ70165)

Med J PUMCH, 2023,14(3):528-535

心脏骤停是成人死亡的主要原因之一，全球每年新增病例达 800 万至 900 万，而我国每年约 54 万人发生心脏骤停且该数据呈逐年上升趋势[1]。心脏骤停后机体血流循环中断，数分钟内即可导致脑缺血死亡，即使进行有效的心肺复苏，短期死亡风险仍较高，而住院期间转归是临床医生及患者家属最迫切关注的问题。准确预测心脏骤停患者院内死亡风险有助于治疗方案的优化、临床决策的制订以及和谐医患关系的建立。由于心脏骤停病因的多样性和病情的复杂性、危重性，传统指标如神经元特异性烯醇化酶（neuron-specific enolase，NSE）、S100β、脑电图、颅脑影像学表现、格拉斯哥昏迷评分（Glasgow coma score，GCS）、急性生理学评分系统Ⅲ（acute physiology score Ⅲ，APSⅢ）在此类患者死亡风险的预测中难度较大[2]。

近年来，伴随计算机性能的巨大突破，医工融合现象逐渐明显，机器学习算法逐步被引入医学领域。Ngiam 等[3]研究表明，对于病情严重、临床数据广泛且复杂患者的健康评估，机器学习算法的表现优于传统方法。基于其运行速度快、可高效处理大数据的优势[4]，机器学习算法已广泛用于多种危重症患者的预后评估[5-9]。在心脏骤停方面，既往大量研究已证实，基于机器学习算法构建的模型在此类人群神经功能评定、疾病复发风险预测方面表现出良好的性能[10-13]。但此类模型存在“黑盒子”问题，即缺乏临床易于理解的可解释性。2020 年 Lundberg 等[14]建立了 Shapley 加法解释（Shapley additive explanation，SHAP）算法用以解释任何机器学习模型的输出，其不仅可根据 SHAP 值正负性反映变量对模型的影响程度，并可通过 SHAP 值对模型中每个变量的贡献进行量化，突破了机器学习模型难以解释的“黑盒子”问题[15]。迄今为止，基于可解释机器学习模型对公共医疗卫生大数据库中心脏骤停患者住院期间死亡风险预测的相关研究仍较缺乏。本研究基于美国重症监护医学信息数据库Ⅳ（Medical Information Mart for Intensive Care database Ⅳ，MIMIC-Ⅳ）2.0 中的数据，开发 6 种预测心脏骤停患者住院期间死亡风险的机器学习模型，经筛选后采用 SHAP 算法对最优模型进行

解释，以期辅助心脏骤停患者临床决策的制订。

1 资料与方法

1.1 研究对象

本研究数据来源于 MIMIC-Ⅳ 2.0（https://mimic.physionet.org/）。该数据库包含 2008—2019 年贝斯以色列女执事医疗中心 4 万余例转入 ICU 的患者临床资料。纳入标准：（1）年龄≥18 岁；（2）根据国际疾病分类（international classification of diseases，ICD）诊断为心脏骤停，疾病编码为 ICD-9 中的“4275”，ICD-10 中的“I46”“I462”“I468”“I469”“I9712”“I97120”“I97121”“I9771”“I97710”“I97711”。排除标准：（1）多次（≥2 次）住院；（2）住院时间<24 h；（3）孕妇；（4）临床资料不完整者。

本研究以心脏骤停患者住院期间转归为结局指标，并据此将患者分为死亡组和存活组。

本研究人员已完成美国国立卫生研究院开设的“保护人类研究参与者”课程，并获得 MIMIC-Ⅳ2.0 数据库使用权限（认证号：10264242），可下载数据进行相关研究。

1.2 方法

1.2.1 数据提取与处理

采用 PostgreSQL 13 提取患者转入 ICU 后 24 h 内的临床资料（若多次检测，以首次数据为准），主要包括：（1）转入 ICU 首日记录：如年龄、性别、生理特征及实验室检查，并计算 GCS、器官功能障碍逻辑性评分（Logistic organ dysfunction score，LODS）、APSⅢ、牛津急性疾病严重程度评分（Oxford acute severity of illness score，OASIS）、序贯器官功能衰竭评价（sequential organ failure assessment，SOFA）评分、全身炎症反应综合征（systemic inflammatory response syndrome，SIRS）评分；（2）主要的基础疾病：包括高血压、糖尿病、心力衰竭、肾衰竭等；（3）使用的药物及特殊操作：抗感染药物、血管活性药物、抗凝药物、静脉补液量、尿量、是否机械通气/肾脏替代治疗等。（4）其他资料：住院时间、转入 ICU 时间等。变量缺失值的处理：若缺失值超过 40%，该变量予以删除；否则采用 K 近邻（K-nearest neighbor，KNN）插补法进行填补。KNN 插补法可根据已知的数据点之间的距离，选择 K 个距离最近的点作为邻近点，然后根据邻近点的属性值进行加权平均，得到缺失值的估计值。

1.2.2 模型构建与评估

基于机器学习算法，构建 6 种预测心脏骤停患者院内死亡风险的模型，分别为 XGBoost 模型、轻量级梯度提升机（light gradient boosting machine，LGBM）模型、决策树（decision tree，DT）模型、KNN 模型、Logistic 回归模型、随机森林（random forest，RF）模型。模型构建时，采用网格搜索法对超参数进行优化。随机将 80% 的数据划分为一个训练集，同时保留剩余 20% 的数据作为独立的测试集。采用十折交叉验证法进行模型训练。在训练集中，将训练集数据随机划分为 10 个小组，其中 9 个小组用于模型训练，1 个小组用于算法的性能评估。将训练集中所有可能的训练小组和测试小组进行折叠组合，重复该过程 10 次，然后在独立测试集中对得到的 10 个模型进行评估并计算评价指标均值。评价指标包括灵敏度、特异度、曲线下面积（area under the curve，AUC）、阳性似然比（positive likelihood ratio，PLR）、阴性似然比（negative likelihood ratio，NLR）。选取 AUC 居前 3 位的模型，绘制临床决策曲线和校准曲线，进一步评价模型的临床实用性（净收益）及准确性。

1.2.3 可解释性分析

SHAP 是一种机器学习解释方法，可用于解释模型预测结果的特征重要性。其基于合作博弈理论中的 Shapley 值概念，采用一种加性方法计算每个特征对模型预测结果的贡献。SHAP 算法可为每个特征提供一个解释值，表示该特征对于模型预测结果的影响程度，计算结果不仅可解释单个预测结果的特征重要性，还可用于解释整个数据集的特征重要性分布。同时，该方法可提供一种可视化工具，以直观展示每个特征对于每个数据点的影响程度，以及整个数据集的特征重要性分布结果。此外，SHAP 支持对多输出模型和时间序列数据进行解释，并能够处理缺失值和分类特征等常见问题。因此，该方法已成为机器学习领域中重要的解释方法之一，被广泛应用于数据科学、自然语言处理、计算机视觉等领域。本研究采用 Python 3.9 软件构建模型并通过 SHAP 算法对模型进行解释，采用代码“shap.summary_plot”导出汇总图，采用代码“shap.dependence_plot”导出依赖图。

1.3 统计学处理

采用 SPSS 25.0 软件进行统计学分析。年龄、心率、呼吸频率等符合正态分布的计量资料以均数±标准差表示，组间比较采用 t 检验；体温、住院时间、SIRS 评分等不符合正态分布的计量资料以中位数

（四分位数）表示，组间比较采用 Mann-Whitney U 检验。性别、合并的主要基础疾病等计数资料以频数（百分数）表示，组间比较采用卡方检验。采用受试者操作特征（receiver operator characteristic，ROC）曲线计算模型预测心脏骤停患者院内死亡的 AUC、灵敏度、特异度等指标。采用 Python 3.9 软件绘制临床决策曲线和校准曲线。以 $P<0.05$ 为差异具有统计学意义。

2 结果

2.1 一般临床资料

基于 MIMIC-Ⅳ 2.0 数据库共筛选 1996 例心脏骤停患者，排除多次转入 ICU 者 253 例、ICU 住院时间<24 h 者 233 例、临床资料不完整者 45 例，最终入选 1465 例符合纳入与排除标准的心脏骤停患者。其中存活组 773 例、死亡组 692 例。研究对象入选流程见图 1。

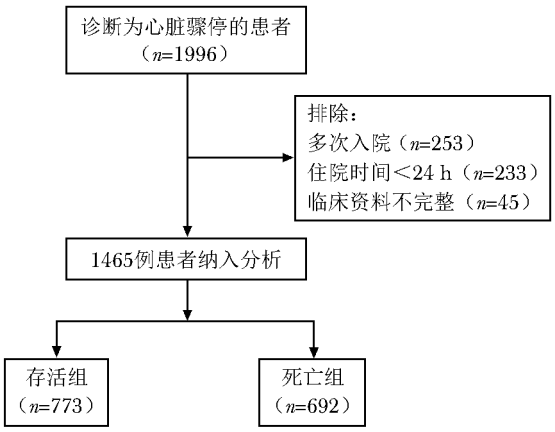

图 1 心脏骤停患者入选流程图

死亡组在年龄、心率、呼吸频率、药物治疗、合并基础疾病、多种系统评分以及住院时间等方面与存活组均有显著差异（P 均<0.05），详见表 1。

2.2 6 种预测模型的性能

经筛选，共纳入 82 个临床特征用于构建 6 种机器学习模型（每个模型均包括 82 个相同的临床特征），并基于测试集数据评价了模型性能。ROC 曲线显示，LGBM 模型预测心脏骤停患者院内死亡的 AUC 最高（AUC：0.834），Logistic 回归模型（AUC：0.809）、XGBoost 模型（AUC：0.827）次之，KNN 模型、DT 模型、RF 模型的 AUC 较低（AUC 均低于 0.8）。详见图 2，表 2。

校准度反映模型预测概率与实际概率之间的差异，该数据越小表示模型预测结果与实际结果越接近，即模型的准确性越高。校准曲线如图 3 所示，通过计算，LGBM 模型的校准度（0.166）较 Logistic 回归模型（0.178）、XGBoost 模型低（0.179）。临床决

表 1 1465 例心脏骤停患者基线主要临床资料

指标	死亡组（n=692）	存活组（n=773）	P 值
年龄（$\bar{x}\pm s$，岁）	67.57±16.35	65.57±16.15	0.019
女性［n（%）］	286（41.3）	290（37.5）	0.136
心率（$\bar{x}\pm s$，次/min）	93.66±22.36	87.44±22.19	<0.001
呼吸频率（$\bar{x}\pm s$，次/min）	21.25±6.50	20.01±6.22	<0.001
体温［M（P_{25}，P_{75}），℃］	36.50（35.80，36.89）	36.72（36.33，37.06）	<0.001
糖尿病［n（%）］	113（16.3）	141（18.2）	0.335
心力衰竭［n（%）］	248（35.8）	345（44.6）	0.001
肾衰竭［n（%）］	395（57.1）	340（44.0）	<0.001
肾脏替代治疗［n（%）］	97（14.0）	74（9.6）	0.008
使用多巴胺［n（%）］	125（18.1）	91（11.8）	0.001
使用肾上腺素［n（%）］	147（21.2）	113（14.6）	0.001
GCS 评分［M（P_{25}，P_{75}），分］	10（3，15）	13（9，15）	<0.001
SOFA 评分（$\bar{x}\pm s$，分）	9.82±4.32	7.60±4.38	<0.001
LODS 评分（$\bar{x}\pm s$，分）	9.24±3.80	6.59±3.75	<0.001
APSⅢ评分（$\bar{x}\pm s$，分）	81.63±29.39	59.34±27.76	<0.001
SIRS 评分［M（P_{25}，P_{75}），分］	3（3，4）	3（2，3）	<0.001
住院时间［M（P_{25}，P_{75}），d］	5.76（2.77，11.28）	11.88（6.97，22.05）	<0.001

GCS：格拉斯哥昏迷评分；SOFA：序贯器官功能衰竭评价；LODS：器官功能障碍逻辑性评分；APSⅢ：急性生理学评分系统Ⅲ；SIRS：全身炎症反应综合征

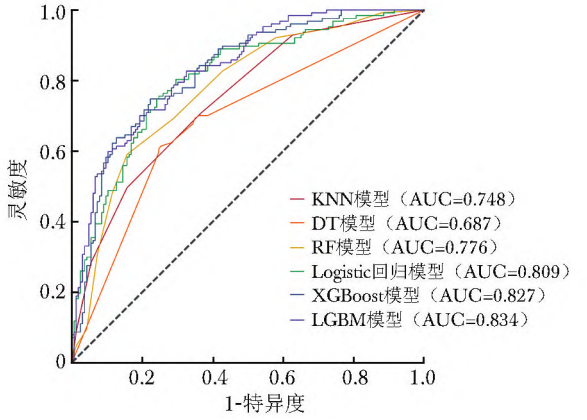

图 2 6 种机器学习模型预测心脏骤停患者院内死亡风险的 ROC 曲线图

KNN：K 近邻；DT：决策树；RF：随机森林；LGBM：轻量级梯度提升机；ROC：受试者操作特征

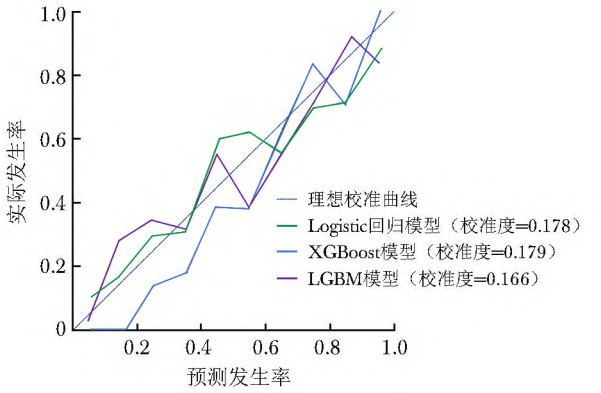

图 3 预测效能 Top 3 模型的校准曲线

LGBM：同图 2

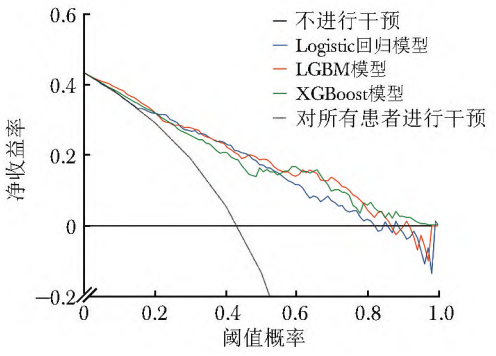

图 4 预测效能 Top 3 模型的临床决策曲线

LGBM：同图 2

策曲线可用于衡量机器学习模型在不同决策阈值下的性能表现。临床决策曲线显示，相较于 Logistic 回归模型、XGBoost 模型，若阈值概率（判断结局变量发生的概率）处于 5%～90%，则 LGBM 模型预测患者住院死亡风险时可增加更多的净收益，整体来看 LGBM 模型的临床应用价值更优（图 4）。

2.3 模型可解释性分析

采用 SHAP 算法对 LGBM 模型进行可解释性分析，并输出 SHAP 汇总图，汇总图可视化展示了临床特征对 LGBM 模型输出结果的影响。其中图 5A 展示了前 20 个临床特征 SHAP 值的分布情况：图中每个点表示一个特征，点的位置表示特征的 SHAP 值，其值代表该特征对模型输出的贡献大小。如果数值为正，则说明该特征对输出结果产生正面影响；如果数值为负，则说明该特征对输出结果产生负面影响。红色表示高值，蓝色表示低值。颜色越深表示该特征对目标变量的影响越强。条形图为按照特征的平均 SHAP 绝对值大小从高至低进行排列后形成，该排序表示每个特征对于整个模型的贡献程度，SHAP 绝对值越大表示该特征越重要，对模型输出结果的影响越大。影响性居前 10 位的重要临床特征依次为 GCS 睁眼反应评分、碳酸氢盐水平、白细胞计数、APS Ⅲ 评分、谷草转氨酶水平、GCS 运动评分、红细胞分布宽度、体温、钙离子含量、年龄（图 5B）。

基于 SHAP 汇总图，进一步导出影响性居前 3 位临床特征的 SHAP 依赖图，以解释临床特征对患者死亡风险的影响。SHAP 依赖图的纵轴为临床特征的

表 2 6 种机器学习模型预测心脏骤停患者院内住院死亡风险的性能比较

预测模型	AUC（95% CI）	灵敏度（95% CI,%）	特异度（95% CI,%）	PLR（95% CI）	NLR（95% CI）
KNN 模型	0.748（0.659～0.815）	71（67.4～82.2）	63（57.7～73.6）	1.93（1.662～2.215）	0.46（0.152～0.694）
DT 模型	0.687（0.581～0.778）	61（59.3～75.5）	74（66.9～83.9）	2.43（1.790～2.683）	0.52（0.263～0.821）
RF 模型	0.776（0.650～0.820）	59（55.1～72.9）	84（66.8～88.4）	3.77（2.825～4.556）	0.48（0.189～0.701）
Logistic 回归模型	0.809（0.661～0.853）	76（66.8～84.7）	75（52.2～81.4）	3.06（2.549～4.018）	0.32（0.125～0.682）
XGBoost 模型	0.827（0.679～0.875）	75（64.5～80.1）	78（63.8～82.0）	3.36（2.699～4.343）	0.32（0.119～0.668）
LGBM 模型	0.834（0.688～0.894）	70（63.2～81.0）	81（65.9～85.4）	3.64（2.776～4.626）	0.37（0.165～0.694）

KNN、DT、RF、LGBM：同图 2；AUC：曲线下面积；PLR：阳性似然比；NLR：阴性似然比

SHAP 值，横轴为该临床特征的变化范围，若 SHAP 值高于零，表示患者院内死亡风险增加，见图 6。

3 讨论

本研究基于 MIMIC-Ⅳ 2.0 数据库，构建了 6 种可预测心脏骤停患者院内死亡风险的机器学习模型，并尝试采用 SHAP 算法对最优模型进行可解释性分析。结果显示，LGBM 模型在心脏骤停患者院内死亡风险的预测中表现［AUC：0.834（95% CI：0.688～0.894）］优于其他模型，且临床实用性强、预测准确性高，综合性能最佳。可解释性分析显示，对 LGBM 模型输出结果影响性居前 10 位的临床特征依次为 GCS 睁眼反应评分、碳酸氢盐水平、白细胞计数、APSⅢ评分、谷草转氨酶水平、GCS 运动评分、红细胞分布宽度、体温、钙离子含量、年龄。

3.1 机器学习预测模型构建与评价

机器学习算法可对数据进行深度挖掘，以分析数据之间的内部联系，在大数据的处理中优势凸显。近年来，其在心脏骤停预警及心脏骤停患者神经功能预后预测方面取得了长足进步[12, 16-18]。Wu 等[16]研究表明，相较于传统预测模型，基于机器学习算法生成的 XGBoost 模型可提高急性冠脉综合症患者住院期间发生心脏骤停风险的预测准确性。系统评价显示，机器

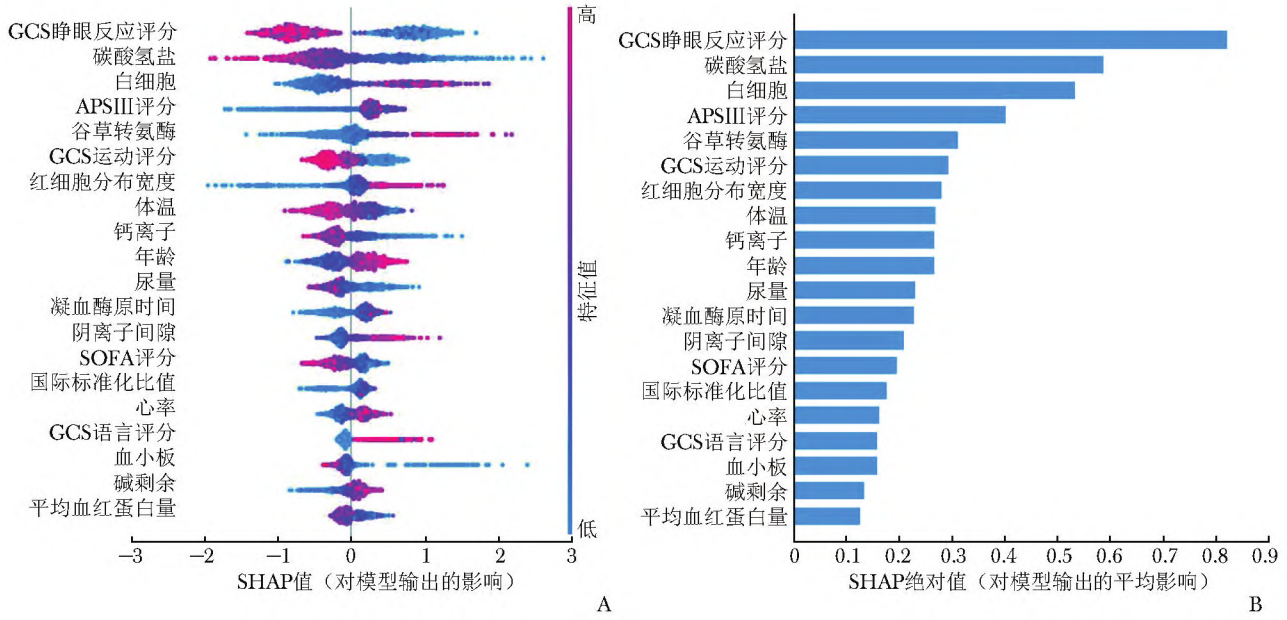

图 5 SHAP 汇总图

A. 不同临床特征对模型输出结果影响性的 SHAP 值；B. 各临床特征的平均 SHAP 绝对值

SHAP：Shapley 加法解释；GCS：APSⅢ、SOFA：同表 1

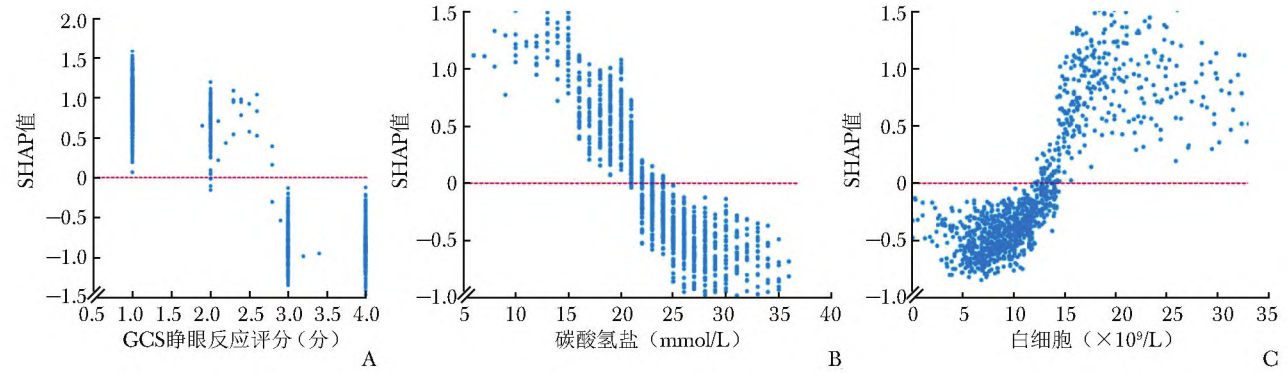

图 6 对模型输出结果影响性 Top 3 临床特征的 SHAP 依赖图

SHAP：同图 5；GCS：同表 1

学习模型可更准确地预测院外心脏骤停患者神经功能结局，且在某些特定情况下其预测效能优于传统统计学模型[17]。Mayampurath 等[19]基于117 674例院内心脏骤停患者的临床资料比较了不同机器学习模型在此类人群神经功能预后中的预测作用，发现梯度增强算法模型的预测准确性最高。本研究以 MIMIC-Ⅳ 2.0 数据库中心脏骤停患者的临床资料为基础，经筛选后保留82 个临床特征用于建立6 种可预测心脏骤停院内死亡风险的机器学习模型，包括 KNN 模型、DT 模型、RF 模型、Logistic 回归模型、XGBoost 模型、LGBM 模型。本研究首先通过 ROC 曲线评估了 6 种模型的区分度，即早期识别出心脏骤停院内死亡患者的能力，结果显示 Logistic 回归模型、XGBoost 模型、LGBM 模型具有较高的识别度，其中以 LGBM 模型的表现最佳。进一步对 3 种区分度较好模型的准确性及临床实用性进行评价。相较于 Logistic 回归模型、XGBoost 模型，LGBM 模型校准曲线的校准度最低，提示该模型的准确性较高；临床决策曲线示，LGBM 模型的整体净收益高于 Logistic 回归模型、XGBoost 模型，提示其临床实用性更佳；且综合灵敏度、特异度等指标后，LGBM 模型的整体表现亦更好，提示其在心脏骤停患者死亡风险的预测中更具优势。LGBM 模型是一种经过改进的梯度提升集成算法，主要用于分类和回归预测，其可利用决策树迭代训练以提升模型的性能[20]，具有准确度高、内存消耗低、训练速度快的优势[21]。既往 Rufo 等[22]在糖尿病的研究中证实，LGBM 模型凭借其训练速度快、预测性能高的优势在糖尿病诊断模型的构建中优势得到凸显。Ge 等[23]在一项纳入 12 460 例脓毒症患者的研究中亦发现，基于 LGBM 构建的脓毒症相关脑损伤预测模型显著优于 XGBoost、DT 等常见模型。由此可见，LGBM 模型训练速度快、支持大样本量运算的优势可满足心脏骤停患者住院期间死亡风险预测的全面性、广维度要求。综上可知，LGBM 模型预测心脏骤停患者院内死亡风险的总体性能较高，可辅助临床早期识别死亡高风险个体，有助于对患者进行个体化管理和精准诊疗的实施。

3.2 模型的可解释性

机器学习预测模型作为临床疾病诊断及患者预后评估的有效工具，由于其形成过程存在的“黑盒子”问题，导致临床医生难以理解模型的原理，进而限制了其临床应用。本研究基于 SHAP 算法对 LGBM 模型预测心脏骤停患者院内死亡风险的可解释性进行分析，结果显示对模型输出结果影响性较大的 3 个临床特征分别为 GCS 睁眼反应评分、碳酸氢盐水平、白细胞计数，可作为预测此类患者住院死亡率的重要指标。GCS 评分是神经系统检查的常用指标，可评估患者昏迷程度，具有简便、快捷、低成本的优势，既往研究证实入院时 GCS 评分超过 4 分可预测院外心脏骤停患者的院内生存率[24]。睁眼反应是 GCS 评分的重要组成部分，蔡兰兰等[25]研究发现，GCS 睁眼反应评分≤2 分的心脏骤停患者预后明显较≥3 分患者差。本研究 SHAP 依赖图显示，GCS 睁眼反应评分≤2 分时，心脏骤停患者住院死亡风险显著升高（图 6A），进一步证实了上述观点。血液中碳酸氢盐是调节机体酸碱度的重要成分，对心脏骤停时因缺血缺氧造成的酸中毒具有中和作用。Chen 等[26]研究发现，院外心脏骤停患者予以适当的碳酸氢钠干预有助于提高生存率。Celik 等[27]则研究认为，过高或过低的碳酸氢盐均可增加心脏骤停患者死亡风险。本研究结果符合既往研究结论，发现碳酸氢盐处于 20~40 mmol/L时，心脏骤停患者住院期间的死亡风险显著降低（图 6B），提示维持适宜的碳酸氢盐水平对改善患者预后至关重要。白细胞是反映机体炎症水平的重要因素，与多种患者住院期间死亡率密切相关。既往研究显示，根据白细胞中的中性粒细胞、淋巴细胞计算的比值与心脏骤停患者不良预后风险呈正相关[28]。本研究亦证实，白细胞>15×10^9/L 时心脏骤停患者院内死亡风险显著升高（图 6C），进一步验证了白细胞水平在心脏骤停患者住院期间死亡风险预测中的重要性。

本研究局限性：（1）临床特征相关信息为回顾性收集，可能存在信息偏倚；（2）由于数据库限制或变量存在严重缺失（如体质量指数、脑功能表现分级评分等），可能影响了预测模型性能的提高；（3）部分患者于入院 24~48 h 内死亡，其入院24 h内首次检查/检验结果可能极差，针对该部分人群，预测模型可能存在一定程度的标签泄露风险；（4）研究数据来源于 MIMIC-Ⅳ 2.0，模型的普适性仍有待验证。

综上，本研究基于大型公共医疗卫生数据库，建立了可预测心脏骤停患者住院期间死亡风险的可解释性机器学习模型。结果示 LGBM 模型在心脏骤停患者死亡风险的预测中更具优势，对该模型影响较大的 3 个临床特征分别为 GCS 睁眼反应评分、碳酸氢盐水平、白细胞计数，上述研究结果有助于增加临床医师对机器学习模型的理解度，促进了模型的临床应用，从而早期识别院内死亡高风险人群并优化治疗方案，作出符合患者最大利益的临床决策。

作者贡献：龚欢欢负责数据统计、图表绘制及论文撰写；柯晓伟负责数据整理及论文修订；李湘民、王爱民负责研究设计及写作指导。

利益冲突：所有作者均声明不存在利益冲突

参 考 文 献

[1] 陈红, 张重阳, 徐俊祥. 急诊院前、院内心脏骤停患者心肺复苏效果分析 [J]. 河北医药, 2017, 39: 2475-2477.

[2] Chen J, Mei Z, Wang Y, et al. A nomogram to predict in-hospital mortality in post-cardiac arrest patients: a retrospective cohort study [J]. Pol Arch Intern Med, 2023, 133: 16325

[3] Ngiam KY, Khor IW. Big data and machine learning algorithms for health-care delivery [J]. Lancet Oncol, 2019, 20: e262-e273.

[4] Rauschert S, Raubenheimer K, Melton PE, et al. Machine learning and clinical epigenetics: a review of challenges for diagnosis and classification [J]. Clin Epigenetics, 2020, 12: 51.

[5] Lee YW, Choi JW, Shin EH. Machine learning model for predicting malaria using clinical information [J]. Comput Biol Med, 2021, 129: 104151.

[6] Hou N, Li M, He L, et al. Predicting 30-days mortality for MIMIC-Ⅲ patients with sepsis-3: a machine learning approach using XGboost [J]. J Transl Med, 2020, 18: 462.

[7] Chu J, Leung KHB, Snobelen P, et al. Machine learning-based dispatch of drone-delivered defibrillators for out-of-hospital cardiac arrest [J]. Resuscitation, 2021, 162: 120-127.

[8] 苏枫, 张少衡, 陈楠楠, 等. 基于机器学习分类判断算法构建心力衰竭疾病分期模型 [J]. 中国组织工程研究, 2014, 49: 7938-7942.

[9] 张颖莹, 刘怡果, 赵丹, 等. 基于机器学习建立脓毒症心肾综合征患者早期死亡风险预测模型 [J]. 中华肾脏病杂志, 2022, 38: 785-793.

[10] Stevens RD. Machine Learning to Decode the Electroencephalography for Post Cardiac Arrest Neuroprognostication [J]. Crit Care Med, 2019, 47: 1474-1476.

[11] Jennings JB. Can machine learning predict recurrent cardiac arrest? [J]. Resuscitation, 2023, 184: 109704.

[12] Blomberg SN, Folke F, Ersbøll AK, et al. Machine learning as a supportive tool to recognize cardiac arrest in emergency calls [J]. Resuscitation, 2019, 138: 322-329.

[13] Kwon JM, Jeon Kh, Kim HM, et al. Deep-learning-based out-of-hospital cardiac arrest prognostic system to predict clinical outcomes [J]. Resuscitation, 2019, 139: 84-91.

[14] Lundberg SM, Erion G, Chen H, et al. From Local Explanations to Global Understanding with Explainable AI for Trees [J]. Nat Mach Intell, 2020, 2: 56-67.

[15] 王鑫, 廖彬, 李敏, 等. 融合 LightGBM 与 SHAP 的糖尿病预测及其特征分析方法 [J]. 小型微型计算机系统, 2022, 43: 1877-1885.

[16] Wu TT, Lin XQ, Mu Y, et al. Machine learning for early prediction of in-hospital cardiac arrest in patients with acute coronary syndromes [J]. Clin Cardiol, 2021, 44: 349-356.

[17] 郑萍, 刘宁. 机器学习应用于院外心脏骤停神经系统预后预测模型的系统评价 [J]. 中国胸心血管外科临床杂志, 2022, 29: 1172-1180.

[18] 吴秋硕, 陆宗庆, 刘瑜, 等. 机器学习应用于心脏骤停早期预测模型的系统评价 [J]. 中国循证医学杂志, 2021, 21: 942-952.

[19] Mayampurath A, Hagopian R, Venable L, et al. Comparison of Machine Learning Methods for Predicting Outcomes After In-Hospital Cardiac Arrest [J]. Crit Care Med, 2022, 50: e162-e172.

[20] Yan J, Xu Y, Cheng Q, et al. LightGBM: accelerated genomically designed crop breeding through ensemble learning [J]. Genome Biol, 2021, 22: 271.

[21] 张红莉, 李月琴, 韩磊, 等. 基于 LGBM 和深度神经网络的 HRRP 目标识别方法 [J]. 探测与控制学报, 2022, 44: 97-103, 114.

[22] Rufo DD, Debelee TG, Ibenthal A, et al. Diagnosis of Diabetes Mellitus Using Gradient Boosting Machine (LightGBM) [J]. Diagnostics (Basel), 2021, 11: 1714.

[23] Ge C, Deng F, Chen W, et al. Machine learning for early prediction of sepsis-associated acute brain injury [J]. Front Med (Lausanne), 2022, 9: 962027.

[24] Nadolny K, Bujak K, Obremska M, et al. Glasgow Coma Scale score of more than four on admission predicts in-hospital survival in patients after out-of-hospital cardiac arrest [J]. Am J Emerg Med, 2021, 42: 90-94.

[25] 蔡兰兰, 杨增强. 70 例心脏骤停后复苏患者格拉斯哥昏迷评分及预后分析 [J]. 内科急危重症杂志, 2018, 24: 431-433.

[26] Chen YC, Hung MS, Liu CY, et al. The association of emergency department administration of sodium bicarbonate after out of hospital cardiac arrest with outcomes [J]. Am J Emerg Med, 2018, 36: 1998-2004.

[27] Celik T, Ozturk C, Balta S, et al. Sodium bicarbonate dilemma in patients with out-of-hospital cardiac arrest: A double-edged sword [J]. Am J Emerg Med, 2016, 34: 1314-1315.

[28] Kim HJ, Park KN, Kim SH, et al. Association between the neutrophil-to-lymphocyte ratio and neurological outcomes in patients undergoing targeted temperature management after cardiac arrest [J]. J Crit Care, 2018, 47: 227-231.

（收稿：2022-12-30 录用：2023-02-20）

（本文编辑：董 哲）