

## • 临床研究 •

开放科学  
(资源服务)  
标识码(OSID)

# 基于机器学习的急性冠脉综合征患者急性肾损伤的预测模型

黄莹<sup>1,2,3</sup> 木胡牙提·乌拉斯汗<sup>4</sup> 陆晨<sup>1,2,3</sup> 李欣赛<sup>1,2,3</sup> 王正业<sup>5</sup> 王润泽<sup>1,2,3</sup>  
褚雪倩<sup>1,2,3</sup> 蒋绪燕<sup>1,2,3</sup> 李素华<sup>1,2,3</sup>

<sup>1</sup>新疆医科大学第一附属医院肾脏疾病中心, 乌鲁木齐 830054; <sup>2</sup>新疆维吾尔自治区肾脏病研究所, 乌鲁木齐 830054; <sup>3</sup>新疆医科大学第一附属医院省部共建中亚高发病成因与防治国家重点实验室, 乌鲁木齐 830054; <sup>4</sup>新疆医科大学第一附属医院心脏疾病中心, 乌鲁木齐 830054; <sup>5</sup>新疆医科大学医学工程技术学院, 乌鲁木齐 830017

通信作者: 李素华, Email: [lisuhuanh@sina.com](mailto:lisuhuanh@sina.com)

**【摘要】** 目的 基于机器学习(machine learning, ML)的方法构建急性冠脉综合征(acute coronary syndrome, ACS)患者在急诊入院后发生急性肾损伤(acute kidney injury, AKI)的临床预测模型。方法 收集 2018 年 1 月至 2020 年 12 月在新疆医科大学第一附属医院住院确诊 ACS 的患者临床资料。(1)使用 sklearn 工具包中的 train\_test\_split 函数将数据集分为训练组(70%)和测试组(30%), 在训练组中使用 logistic 回归、决策树、随机森林、极端梯度提升、支持向量机和轻量级梯度提升(light gradient boosting machine, LightGBM)方法构建 ACS-AKI 的预测模型;(2)通过模型可视化工具 SHAP 可解释模型与递归特征消除(recursive feature elimination, REF)对最佳 ML 预测模型变量进行筛选, 并压缩为特征减少的紧凑预测模型;(3)最后采用测试组数据从区分度, 校准和临床效益三个方面对筛选后的预测模型进行评价。结果 共纳入 499 例 ACS 患者临床资料, 其中 ACS-AKI 发生率为 29.5%; LightGBM 模型在训练集中表现出最佳预测性能(受试者工作特征曲线的曲线下面积=0.78)。对特征变量筛选后发现, 导致 ACS 并发 AKI 的预测因子有入院肌酐、射血分数、三酰甘油、B 型钠尿肽、血尿素氮、胱抑素 C、D-二聚体、年龄、袝利尿剂、心肌梗死范围。最后使用测试集数据对 ACS-AKI 的 ML 模型进行验证, 得到 LightGBM 模型在区分度(受试者工作特征曲线=0.79)、校准方面、临床效益方面都具有最好的表现。结论 通过不同的 ML 算法构建并验证了 ACS 患者并发 AKI 的临床预测模型, 并且得到了较好的效果, 但此模型仍需要更多外部数据进行验证。

**【关键词】** 急性冠脉综合征; 急性肾损伤; 机器学习; 预测模型

**基金项目:** 国家自然科学基金地区基金项目(81860125); 新疆医科大学研究生创新创业项目-实践创新项目(CXCY2021002); 省部共建中亚高发病成因与防治国家重点实验室开放课题项目面上项目(SK-LHIDCA-2021-8)

DOI: [10.3969/j.issn.1671-2390.2023.08.007](https://doi.org/10.3969/j.issn.1671-2390.2023.08.007)

## Prediction model of acute kidney injury in emergency patients with acute coronary syndrome based on machine learning

Huang Xuan<sup>1,2,3</sup>, Muhuyati·Wulasihan<sup>4</sup>, Lu Chen<sup>1,2,3</sup>, Li Xin-sai<sup>1,2,3</sup>, Wang Zheng-ye<sup>5</sup>, Wang Run-ze<sup>1,2,3</sup>, Chu Xue-qian<sup>1,2,3</sup>, Jiang Xu-yan<sup>1,2,3</sup>, Li Su-hua<sup>1,2,3</sup>

<sup>1</sup> Kidney Disease Center of the first affiliated Hospital of Xinjiang Medical University, Urumqi 830054, China; <sup>2</sup> Institute of Nephrology, Xinjiang Uygur Autonomous Region, Urumqi 830054, China; <sup>3</sup> State Key Laboratory of Causes and Prevention of High Morbidity in Central Asia, The First Affiliated Hospital of Xinjiang Medical University, Urumqi 830054, China; <sup>4</sup> Center for Heart Diseases, The First Affiliated Hos-

pital of Xinjiang Medical University, Urumqi 830054, China; <sup>5</sup> College of Medical Engineering Technology, Xinjiang Medical University, Urumqi 830017, China

Corresponding author: Li Su-hua, Email: [lisuhuanh@sina.com](mailto:lisuhuanh@sina.com)

**【Abstract】 Objective** Clinical prediction models of acute kidney injury (AKI) in patients with ACS after emergency admission was constructed based on the method of Machine Learning (ML).

**Methods** Clinical data were collected from January 2018 to December 2020 from patients hospitalized with confirmed ACS in the First Affiliated Hospital of Jiangxi Medical University. (1) The data set was randomly divided into training group (70%) and testing group (30%), and logistic regression, decision tree, random forest, XGboost, support vector machine (Support Vector Classification, SVC) and LightGBM methods were used in the training group to construct the ACS-AKI prediction model. (2) The best machine learning prediction model variables are screened by SHapley Additive exPlanations (SHAP) and Recursive feature elimination (REF), a model visualization tool, and compressed into a compact prediction model with reduced features. (3) Finally, the screened prediction models are evaluated in terms of discrimination, calibration and clinical effectiveness using test set data.

**Results** A total of 499 clinical data of ACS patients were included, of which the incidence of ACS-AKI was 29.5% (147/499); the LightGBM model showed the best predictive performance in the training set (AUC = 0.78). Screening of the characteristic variables revealed that the predictors leading to ACS complicated by AKI were admission creatinine, ejection fraction, triglycerides, B-type natriuretic peptide (BNP), blood urea nitrogen, cystatin C, D-dimer, age, loop diuretics, and extent of myocardial infarction. Finally, we validated the ML model of ACS-AKI using test set data and obtained that the LightGBM model had the best performance in terms of discrimination (AUC: 0.79), calibration aspects, and clinical benefit. **Conclusion** We have constructed and validated a clinical prediction model for concurrent AKI in ACS patients by different ML algorithms and obtained better results. However, this model still needs more external data for validation.

**【Key words】** Acute coronary syndrome; Acute kidney injury; Machine learning; Forecasting model

**Fund program:** National Natural Science Foundation of China Regional Fund Project (81860125); Xinjiang Medical University Postgraduate Innovation and Entrepreneurship Project - Practical Innovation Project (CXCY2021002); The open project of the State Key Laboratory of Causes and Prevention of High Morbidity in Central Asia, jointly constructed by the Ministry and the Province (SKL-HIDCA-2021-8)

DOI: [10.3969/j.issn.1671-2390.2023.08.007](https://doi.org/10.3969/j.issn.1671-2390.2023.08.007)

急性冠脉综合征(acute coronary syndrome, ACS)是由冠状动脉粥样硬化后形成闭塞性血栓导致心肌急性缺血引起的临床综合征。其发病率逐年上升,该病具有起病急骤、病情变化快、病死率高的特点<sup>[1]</sup>,给患者家庭和社会带来沉重的经济负担。急性肾损伤(acute kidney injury, AKI)是 ACS 患者常见的并发症,AKI 是一种以肾功能迅速恶化(在数小时或数天内)为特征的复杂的临床综合征<sup>[2]</sup>。其发生是增加短期、长期并发症发病率<sup>[1-2]</sup>和病死率以及延长住院时间的一个重要因素,很大程度上影响了患者的长期预后,已成为临床关注的焦点问题<sup>[3-5]</sup>。

近年来计算技术的进步促进了机器学习(machine learning, ML)和人工智能在 AKI 预测中的应用,使我们能够通过电子病历系统开发更多的方法

来预测和估计 AKI 风险<sup>[6]</sup>。故本研究旨在基于 ML 的方法上建立急诊 ACS 患者发生 AKI(ACS-AKI)的预测模型,以期今后临床工作提供早期预防及干预 AKI 的依据。开展早期识别 AKI 高危患者的研究对改善 ACS 患者的预后非常重要。

## 对象与方法

### 一、研究对象

搜集 2018 年 1 月至 2020 年 12 月新疆医科大学第一附属医院急诊转入冠心病重症监护室并确诊为 ACS 的 537 例患者的资料。通过网络病历系统,回顾性分析依据本次研究的纳入、排除标准,共纳入患者 499 例,其中 AKI 患者 147 例(29.5%),非 AKI 患者 352 例(70.5%)。本研究为回顾性研究,不涉及敏感个人信息或商业秘密,可以免除伦

理审查。

1. 纳入标准 (1) 我院行急诊入院明确诊断为 ACS 的患者; (2) 年龄  $\geq 18$  岁, 有完整的入、出院记录和手术记录, 住院时间  $> 24$  h。

2. 排除标准 (1) 检查结果缺失者; (2) 病史缺失者; (3) 患有慢性肾脏病的患者; (4) 住院期间肾功能检测少于 2 次者; (5) 入院血压、脉搏消失者。

## 二、资料收集

1. 一般情况 年龄、性别、民族、既往病史(高血压、糖尿病、慢性肾脏病、心脑血管疾病)、吸烟史、入院日期、入院时的基本生命体征。

2. 收入病房时指标 收缩压、舒张压、白细胞、血红蛋白、血小板、血钙、血钾、尿酸、尿素氮(blood urea nitrogen, BUN)、乳酸浓度、D-二聚体、心脏彩色多普勒超声左心室射血分数值、活化部分血浆凝血酶原时间、丙氨酸转氨酶、天冬氨酸转氨酶、白细胞介素 6、降钙素原、肌钙蛋白 I、B 型钠尿肽(type B natriuretic peptide, BNP)、氧合指数、入院首次血肌酐(serum creatinine, Scr)、抗生素、利尿剂、心脏 B 型超声等。

3. AKI 的诊断标准 依据 2012 年改善全球肾脏病预后组织(Kidney Disease: Improving Global Outcomes, KDIGO)<sup>[7]</sup> 发布的《急性肾损伤临床实践指南》, 即 Scr 48 h 内升高  $26.5 \mu\text{mol/L}$ , 或 7 d 内 Scr 水平较基线值增加 50%, 或尿量  $< 0.5 \text{ mL} \cdot \text{kg}^{-1} \cdot \text{h}^{-1}$  持续超过 6 h。

## 三、ML

此研究尝试了以下最流行的监督 ML 方法来开发分类结局的预测模型: 决策树(decision tree, DT)、随机森林(random forest, RF)、极端梯度提升(extreme gradient boosting, XGboost)和轻量级梯度提升(light gradient boosting machine, LightGBM), 并与传统的逻辑回归方法进行比较。这些 ML 模型更加接近人的思维, 具有可解释性强以及准确率高的特点。DT 方法是最简单的树状模型<sup>[8]</sup>, 构成它的元素是节点和边, 节点会根据样本的变量特征进行判断, 边则是指下一步的分类方向。DT 采用自上而下递归的思想, 构造信息熵下降最快的树, 并使用基尼不纯度作为分类结果判断的指标。基尼指数是数据集中的样本随机分类错误的概率。但是简单 DT 具有过度拟合和泛化能力弱的缺点, 因此在此基础上衍生出 RF。RF 的基本思路是通过多次构建随机样本的 DT, 采用袋装法以投票的方式输出结果, 构建最终的预测模型<sup>[9]</sup>。XG-

boost 与 LightGBM 则是对梯度提升决策树(gradient boosting decision tree, GBDT)的优化。GBDT 是一种使用弱分类器进行迭代的 decision trees 算法, 通过不断减少模型的残差, 使得 decision trees 累加的预测值更加逼近真实值。而 XGboost 主要在 GBDT 基础上增加正则化构建惩罚函数减少过拟合<sup>[10]</sup>。LightGBM 使用基于直方图的 decision trees 算法, 并且采用分裂增益最大的一个叶子生长策略, 在训练速度方面更加高效, 尤其在处理大数据方面具有明显优势<sup>[11]</sup>。以上这些方法, 在不同的数据集环境中各有利弊, 并没有绝对的优劣之分。

## 四、统计学方法

本研究使用 Python 软件 version 3.7.13(<https://www.python.org>)进行统计处理及分析, 主要包括画图工具包 Matplotlib(version 3.3.4), ML 框架 scikit-learn(version 0.24.2 and 1.0.2)。本次研究对于连续性变量, 在进行特征描述时使用 Kolmogorov-Smirnov 检验评估数据的正态分布, 对于服从正态分布的计量资料用  $\bar{x} \pm s$  表示, 并使用  $t$  检验在训练集和测试集进行组间比较; 不服从正态分布的数据用  $M(Q_1, Q_3)$  表示, 使用两独立样本的 Wilcoxon 秩和检验进行组间比较。对于分类变量, 在进行描述时采用例(%)表示, 使用  $\chi^2$  检验进行组间比较。设置双尾检验,  $P < 0.05$  认为差异有统计学意义。

将数据集随机(使用 sklearn 工具包中的 train\_test\_split 函数)分为训练集(70%)和测试集(30%)。在训练集中使用逻辑回归和多种 ML 方法, 采用 Bootstrap 重采样技术(1000 次), 每次随机在训练集中抽取随机数量的样本进行初始建模, 未被抽到的部分作内部验证, 并通过绘制训练集各个模型的受试者工作特征曲线(receiver operating characteristic, ROC)以及计算 ROC 的曲线下面积(area under the curve, AUC)大小评估预测准确性, 其中表现最好的模型用于后续研究。

为了突出本次研究纳入的特征对模型做出的具体贡献, 我们采用了特征可视化工具包 SHAP 可解释模型(SHapley additive exPlanations, SHAP)(version 0.40.0)用于观察特征在模型中的具体表现, 该方法基于博弈论的思想, 属于事后解释框架, 可以提供具体夏普利值来评估每个样本中每个特征相对于目标变量的重要性<sup>[12]</sup>。此外, 我们通过基于 Sklearn 库的递归特征消除(recursive feature elimination, RFE)的方法对变量进行筛选, 这是一种寻找最优特征子集的贪心算法, 相比既往使用较

多的最小绝对收缩和选择操作回归在处理多重共线性方面更加稳定。在变量筛选时使用五折交叉验证输出稳定可靠的变量,通过 SHAP 与 RFE 的结果确定最终纳入紧凑预测模型的变量。此外使用 Sklearn 库中的网格搜索方法对筛出的预测模型进行超参数优化搜索及优化。

最后利用测试集数据我们从准确性,校准度和临床效益三个维度对最终纳入的紧凑机器模型进行评价。区分度方面,我们通过绘制测试集中 ROC 曲线,对比紧凑模型与临床中广泛使用的指标 Scr、BUN 在诊断 AKI 方面的性能差异。此外为了评估各个 ML 模型对 AKI 发生的预测结果与真实结果之间的差异,我们在测试集中绘制了校正曲线,该曲线 X 轴表示预测 AKI 的概率, Y 轴表示 AKI 发生的实际概率。虚线代表理想曲线,越贴近虚线,一致性越高。此外用布里尔分数定量评估一

致性,布里尔分数对于一组预测值越低,预测校准越好。为了展现模型在临床中的实际效用,我们还绘制了决策曲线分析法满足临床决策者的实际需要,该曲线的 X 轴为阈概率, Y 轴为每个阈概率对应的净获益。

## 结 果

### 一、基线特征

我们最终将来自新疆医科大学第一附属医院急诊入院确诊为 ACS 的患者 499 例纳入分析并且使用 60 个特征,进一步将总数据集按 7 : 3 随机拆分后,训练集和验证集 60 个变量差异均无统计学意义( $P > 0.05$ ),表明两个数据集具有同质性和可比性。(表 1)

### 二、模型开发

我们使用以下 ML 方法,将 ACS 患者训练集

表 1 急性冠脉综合征患者临床特征和变量

变量	总患者(n=499)	验证集(n=150)	训练集(n=349)	t/ $\chi^2$ /Z 值	P 值
年龄(岁, $\bar{x} \pm s$ )	59.71 $\pm$ 12.99	58.88 $\pm$ 12.43	60.07 $\pm$ 13.22	0.939	0.348
入院收缩压(mmHg, $\bar{x} \pm s$ )	125.99 $\pm$ 20.66	125.73 $\pm$ 18.57	126.09 $\pm$ 21.51	0.180	0.857
入院时舒张压(mmHg, $\bar{x} \pm s$ )	77.42 $\pm$ 13.62	77.36 $\pm$ 12.00	77.44 $\pm$ 14.28	0.062	0.950
心率(次/min, $\bar{x} \pm s$ )	82.37 $\pm$ 15.63	82.48 $\pm$ 15.16	82.33 $\pm$ 15.85	-0.098	0.921
入院白细胞( $10^9/L$ , $\bar{x} \pm s$ )	9.91 $\pm$ 3.47	9.62 $\pm$ 3.52	9.96 $\pm$ 3.50	0.474	0.635
入院中性粒细胞 [ $10^9/L$ , $M(Q_1, Q_3)$ ]	6.56(4.97, 9.14)	6.48(4.79, 8.84)	6.61(4.99, 9.45)	0.354	0.723
血红蛋白(g/L, $\bar{x} \pm s$ )	147.13 $\pm$ 19.96	148.80 $\pm$ 19.23	146.41 $\pm$ 20.24	-1.225	0.221
入院红细胞计数 [ $10^{12}/L$ , $M(Q_1, Q_3)$ ]	4.87(4.48, 5.21)	4.88(4.48, 5.29)	4.50(4.89, 5.25)	-1.238	0.216
血细胞比容(% , $\bar{x} \pm s$ )	43.12 $\pm$ 5.34	43.19 $\pm$ 6.00	43.08 $\pm$ 5.03	-0.205	0.837
红细胞分布宽度(% , $\bar{x} \pm s$ )	12.76 $\pm$ 0.95	12.77 $\pm$ 0.88	12.75 $\pm$ 0.98	-0.189	0.849
血小板( $10^9/L$ , $\bar{x} \pm s$ )	232.34 $\pm$ 71.10	232.44 $\pm$ 68.97	232.72 $\pm$ 72.09	0.183	0.854
血钾(mmol/L, $\bar{x} \pm s$ )	3.68 $\pm$ 0.44	3.73 $\pm$ 0.43	3.66 $\pm$ 0.41	-1.647	0.100
血钠(mmol/L, $\bar{x} \pm s$ )	137.33 $\pm$ 3.48	137.18 $\pm$ 3.33	137.39 $\pm$ 3.55	0.614	0.539
血钙(mmol/L, $\bar{x} \pm s$ )	2.24 $\pm$ 0.13	2.23 $\pm$ 0.15	2.24 $\pm$ 0.12	0.986	0.147
尿素氮(mmol/L, $\bar{x} \pm s$ )	5.69 $\pm$ 1.80	5.51 $\pm$ 1.47	5.56 $\pm$ 1.93	1.451	0.325
入院肌酐( $\mu\text{mol/L}$ , $\bar{x} \pm s$ )	75.55 $\pm$ 20.91	73.66 $\pm$ 17.07	76.35 $\pm$ 22.33	1.318	0.188
尿酸(mmol/L, $\bar{x} \pm s$ )	347.89 $\pm$ 98.50	346.52 $\pm$ 100.09	348.48 $\pm$ 97.94	0.203	0.839
空腹血糖(mmol/L, $\bar{x} \pm s$ )	9.24 $\pm$ 4.33	9.67 $\pm$ 5.22	9.06 $\pm$ 3.87	-1.434	0.152
总胆固醇 [mmol/L, $M(Q_1, Q_3)$ ]	4.50(3.68, 5.08)	4.39(3.56, 5.04)	4.50(3.72, 5.09)	-0.591	0.555
三酰甘油 [mmol/L, $M(Q_1, Q_3)$ ]	1.88(1.22, 2.59)	1.91(1.27, 2.63)	1.88(1.22, 2.58)	-0.137	0.891
总胆红素( $\mu\text{mol/L}$ , $\bar{x} \pm s$ )	14.29 $\pm$ 7.44	14.47 $\pm$ 6.71	14.20 $\pm$ 7.74	-0.375	0.708
白蛋白(g/L, $\bar{x} \pm s$ )	41.54 $\pm$ 4.68	41.95 $\pm$ 4.97	41.37 $\pm$ 4.54	-1.282	0.200
丙氨酸转氨酶 [U/L, $M(Q_1, Q_3)$ ]	27.00(18.68, 41.85)	25.96(18.08, 40.33)	27.00(19.00, 43.20)	0.678	0.498
天冬氨酸转氨酶 [U/L, $M(Q_1, Q_3)$ ]	33.00(22.25, 69.31)	31.10(21.54, 57.49)	33.51(22.90, 74.52)	1.411	0.158
胱抑素 C(mg/L, $\bar{x} \pm s$ )	0.81 $\pm$ 0.29	0.78 $\pm$ 0.27	0.82 $\pm$ 0.30	1.549	0.122
肌钙蛋白 I [ $\mu\text{g/L}$ , $M(Q_1, Q_3)$ ]	0.110(0.020, 1.180)	0.071(0.014, 1.160)	0.121(0.018, 1.460)	1.693	0.090
肌酸激酶 [U/L, $M(Q_1, Q_3)$ ]	151.80(87.83, 528.59)	135.50(84.05, 500.81)	158.80(90.65, 536.00)	0.511	0.609
肌酸激酶同工酶 [ $\mu\text{g/L}$ , $M(Q_1, Q_3)$ ]	15.27(2.21, 39.70)	15.085(1.665, 50.025)	15.27(2.80, 37.80)	0.174	0.862
PT[s, $M(Q_1, Q_3)$ ]	11.4(10.9, 12.0)	11.5(10.1, 12.1)	11.4(10.9, 12.0)	-0.594	0.552

续表 1

变量	总患者(n=499)	验证集(n=150)	训练集(n=349)	$t/\chi^2/Z$ 值	P 值
PT 正常对照组 [s, $M(Q_1, Q_3)$ ]	11.5(11.5, 11.6)	11.5(11.5, 11.6)	11.5(11.5, 11.6)	-0.431	0.667
D-二聚体 [ $\mu\text{g/L}$ , $M(Q_1, Q_3)$ ]	122(79, 199)	122(81, 186)	122(77, 206)	0.632	0.527
C 反应蛋白 [ $\mu\text{g/L}$ , $M(Q_1, Q_3)$ ]	4.00(1.57, 8.46)	4.00(1.89, 7.60)	4.00(1.31, 9.17)	-0.158	0.875
白细胞介素 6[ng/L, $M(Q_1, Q_3)$ ]	6.28(3.44, 10.84)	6.22(3.50, 9.72)	6.46(3.32, 10.94)	0.422	0.673
降钙素原 [ $\mu\text{g/L}$ , $M(Q_1, Q_3)$ ]	0.04(0.02, 0.06)	0.04(0.03, 0.06)	0.04(0.02, 0.06)	-0.749	0.454
B 型钠尿肽 [ng/L, $M(Q_1, Q_3)$ ]	114.20(32.03, 387.12)	100.05(27.05, 247.30)	114.20(36.93, 418.70)	1.875	0.061
心脏射血分数(%, $\bar{x} \pm s$ )	57.95 $\pm$ 6.94	58.62 $\pm$ 6.65	57.66 $\pm$ 7.05	-1.413	0.158
舒张期左室后壁厚度(mm, $\bar{x} \pm s$ )	9.28 $\pm$ 0.83	9.19 $\pm$ 0.69	9.31 $\pm$ 0.88	1.464	0.144
舒张期室间隔厚度(mm, $\bar{x} \pm s$ )	9.44 $\pm$ 1.05	9.32 $\pm$ 0.86	9.49 $\pm$ 1.11	1.748	0.081
左房内径(mm, $\bar{x} \pm s$ )	34.35 $\pm$ 4.06	34.38 $\pm$ 4.26	34.33 $\pm$ 3.98	-0.137	0.891
性别 [例(%)]				0.379	0.538
女性	93(18.64)	25(16.67)	68(19.49)		
男性	406(81.36)	125(83.33)	281(80.51)		
所在地区 [例(%)]				0.826	0.363
新疆外	49(9.82)	18(8.05)	31(8.88)		
新疆内	450(90.18)	132(91.95)	318(91.12)		
吸烟史 [例(%)]				0.084	0.771
否	213(42.69)	66(44.00)	147(42.12)		
是	286(57.31)	84(56.00)	202(57.88)		
饮酒史 [例(%)]				0	1
否	279(55.91)	80(56.00)	195(55.87)		
是	220(44.09)	69(44.00)	154(44.13)		
心功能分级 [例(%)]				1.751	0.625
I 级	64(12.83)	21(14.09)	43(12.29)		
II 级	286(57.31)	84(56.38)	202(57.71)		
III 级	108(21.64)	28(18.79)	80(22.86)		
IV 级	41(8.22)	16(10.74)	25(7.14)		
高血压病史 [例(%)]				0.754	0.385
否	246(49.30)	69(46.00)	177(50.72)		
是	253(50.70)	81(54.00)	172(49.28)		
糖尿病史 [例(%)]				0.131	0.718
否	352(70.54)	108(72.00)	244(69.91)		
是	147(29.46)	42(28.00)	105(30.08)		
冠心病史 [例(%)]				0.023	0.880
否	9(1.80)	2(1.33)	7(2.00)		
是	490(98.2)	148(98.67)	342(98.00)		
脑血管病史 [例(%)]				0.017	0.896
否	475(95.19)	142(94.66)	333(95.41)		
是	24(4.81)	8(5.33)	16(4.49)		
慢性肾脏病史 [例(%)]				0.785	0.376
否	462(92.59)	136(90.66)	326(93.41)		
是	37(7.41)	14(9.33)	22(6.59)		
冠状动脉介入术 [例(%)]				0.759	0.383
否	107(21.44)	28(18.66)	79(22.63)		
是	392(78.56)	122(81.33)	270(77.37)		
支架位置(左主干)[例(%)]				0	1
否	489(98.00)	147(98.00)	342(98.00)		
是	10(2.00)	3(2.00)	7(2.00)		

续表 1

变量	总患者(n=499)	验证集(n=150)	训练集(n=349)	$t/\chi^2$ 值	P 值
支架位置(前降支)[例(%)]				0.005	0.952
否	279(55.91)	83(55.33)	196(56.00)		
是	220(44.09)	66(44.67)	153(44.00)		
支架位置(回旋支)[例(%)]				0.181	0.670
否	413(82.77)	122(81.33)	291(83.38)		
是	86(17.23)	28(18.66)	58(16.62)		
支架位置(右冠状动脉)[例(%)]				0.742	0.389
否	364(72.95)	105(70.00)	252(72.00)		
是	135(27.05)	45(30.00)	98(28.00)		
尿蛋白 [例(%)]				0.011	0.914
否	399(79.96)	119(79.33)	280(80.23)		
是	100(20.04)	31(20.67)	69(19.77)		
单纯性肾囊肿 [例(%)]				0.457	0.499
否	440(88.18)	135(90.00)	305(87.39)		
是	59(11.82)	15(10.00)	44(12.61)		
ACEI 和 ARB 使用 [例(%)]				1.581	0.208
否	95(19.04)	23(15.33)	72(20.63)		
是	404(80.96)	127(84.67)	277(79.37)		
利尿剂使用 [例(%)]				0.002	0.737
否	300(60.12)	88(58.67)	212(60.74)		
是	199(39.88)	62(41.33)	137(39.25)		
抗生素使用 [例(%)]				0	1
否	455(91.18)	137(91.33)	318(91.12)		
是	44(8.82)	13(8.67)	31(8.88)		
沙库巴曲缬沙坦钠片使用 [例(%)]				0.012	0.914
否	480(96.19)	145(96.67)	335(95.99)		
是	19(3.81)	5(6.33)	14(4.01)		

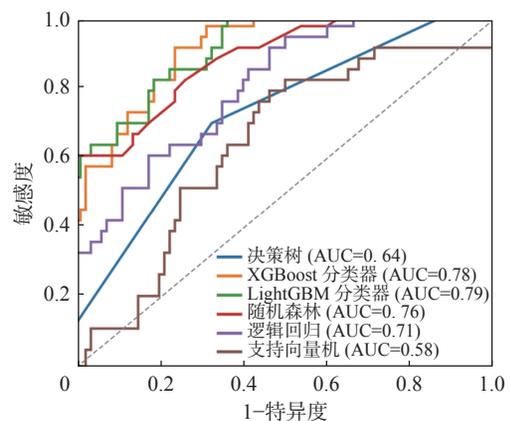
注: 1 mmHg=0.133 kPa; ARB 为血管紧张素受体拮抗剂; ACEI 为血管紧张素转化酶抑制剂; PT 为凝血酶原时间。

中所有特征分别作为输入变量,包括 Logistic 回归、决策树、RF、XGboost 和 LightGBM,支持向量机(support vector machine, SVC)来预测 ACS 患者发生 AKI。由下图 1 可知,对于 ACS-AKI, LightGBM 集成学习模型表现出最佳预测性能(AUC=0.78),随机森林模型也表现出相近的预测性能(AUC=0.77)。下图 2 可视化呈现了简单决策树由最初根节点到最终叶节点递归判断患者有无术后 AKI 的过程。由图 2 可知,在 ACS-AKI 决策树模型中,14 个叶节点中有 7 个基尼(gini)指数超过 0.2,提示准确度欠佳。

### 三、变量筛选

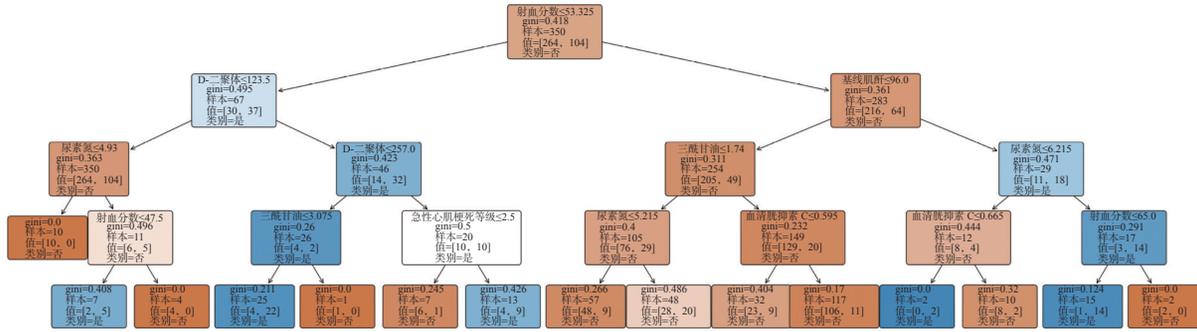
基于 ACS-AKI 最佳模型,使用 SHAP 包进行 ML 模型解释性分析,特征的 SHAP 值越高,ACS 合并 AKI 的可能性就越大。图 3 为训练集中纳入 60 个变量后,根据 SHAP 值降序的排布图。为了防止过拟合,增加临床可控性,我们通过递归特征

消除的方法对变量筛选。递归特征消除后选出 10 个重要特征。图 4 为筛选后再次通过 SHAP 解释的紧凑模型分布效果和转化后的预测变量重要性排序。此次研究发现,入院 Scr 是 ACS 患者完整



注: XGBoost 为极端梯度提升; LightGBM 为轻量级梯度提升。

图 1 机器学习模型之间曲线下面积的比较



注: 对患者是否急性肾损伤进行递归判断, 蓝色代表急性肾损伤, 橙色代表非急性肾损伤组; 每一个方框代表节点, 从上到下依次为根节点、中间节点和叶子节点; 带箭头的直线代表边, 左侧均为正确, 右侧均为错误; 方框所包含的内容主要有对当前节点进行切分所使用的特征, gini 代表当前节点的可能的犯错误率, 在本图中 gini 指数越小, 该节点的颜色密度越大; 值为当前节点实际包含的非急性肾损伤与非急性肾损伤患者人数, 类别代表当前节点预测的患者类别(类别=否为非急性肾损伤患者, 类别=是为急性肾损伤患者)。

图 2 决策树模型的可视化图

型和紧凑型模型中均排名第一的预测因素。在 ACS-AKI 中, 特异性的预测因子有: 入院白细胞、降钙素原、D-二聚体、术中血浆输注量和体外循环时间。ACS-AKI 预测模型中, 特异性的预测因子有: BNP、三酰甘油、BUN、胱抑素 C、D-二聚体、年龄、利尿剂、心肌梗死程度。

四、模型验证与评价

我们将验证集中的特征放到训练集构建的 ML 预测模型中进一步验证。图 5 为 ACS-AKI 预测模型中性能最佳的 LightGBM 集成学习模型与 AKI 临床敏感测量指标 Scr、BUN 预测性能的比较。发现 ML 的综合预测水准要高于单一预测因子。图 6 展示了 ACS-AKI 的不同 ML 模型的校准曲线, 发现 LightGBM 的校准曲线与理想曲线(对角线)贴近程度最高。也就是说, 与其他模型相比, LightGBM 预测值与真实值差异最小。图 7 显示了不同 ML 模型的 DCA 曲线。同样的, 在验证集中 LightGBM 在 ACS-AKI 中获益范围最广, 也就是说模型的临床实用价值较高。

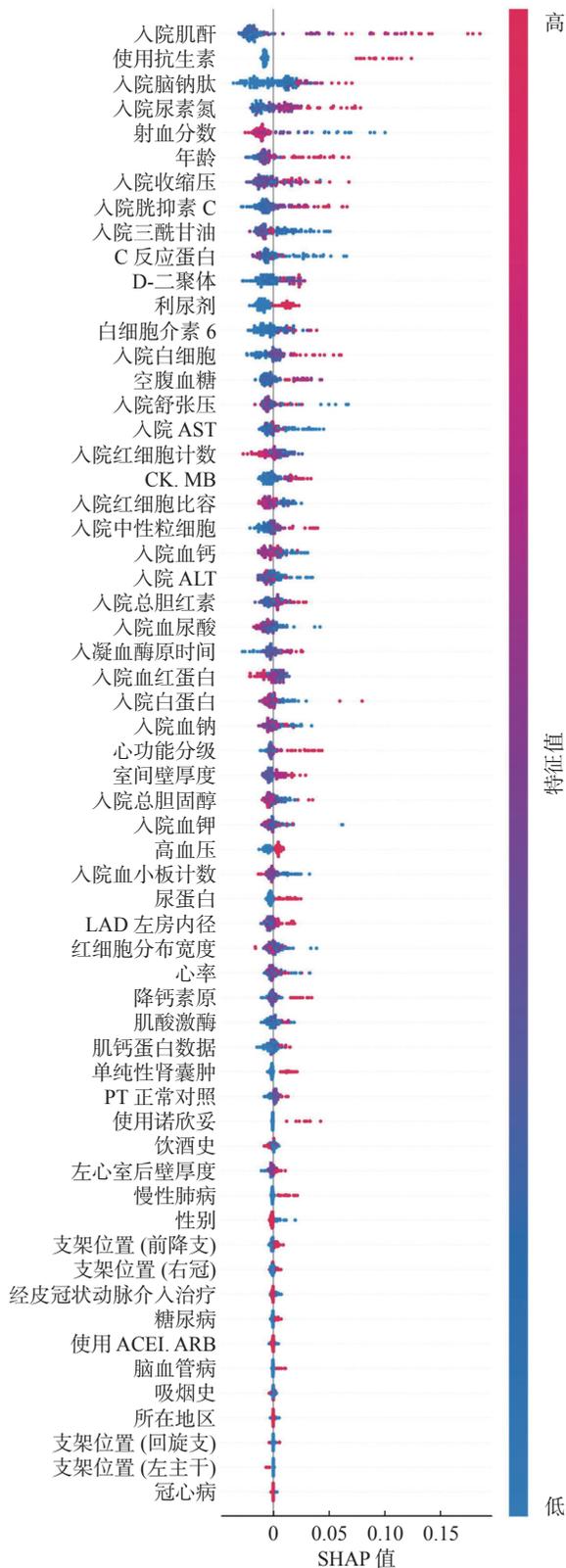
讨论

ACS 是心血管内科常见的急症之一, 而 AKI 是 ACS 患者常见的并发症。有研究报道即使消除了 AKI 对院内病死率的短期影响, AKI 的严重程度与患者的长期病死率之间仍存在着很强的关系<sup>[13]</sup>。近年来国内外研究陆续建立了许多 ACS 合并 AKI 的预测模型, 如 Ma 等<sup>[14]</sup>利用列线图建立了非 ST 段抬高型 ACS 患者对比剂导致的 AKI 模型; 王聪等<sup>[15]</sup>构建了急性 ST 段抬高性心肌梗死患者发生 AKI 的预测模型; Fan 等<sup>[16]</sup>开发了 ACS 患者冠状动脉介入术后发生 AKI 的风险预测模型等。故在此次回顾性研究中基于 ML 方法构建并

验证了 ACS-AKI 的预测模型, ML 方法可以分析具有非线性和复杂关系的众多变量, 比传统的 logistic 回归构建模型更加精准, 已经用于提高临床预测能力的研究<sup>[17]</sup>。如蓝潞杭等<sup>[18]</sup>构建了随机森林模型来预测急性心肌梗死患者 AKI 发生, 其 AUC 为 0.893, 有较好的预测能力。

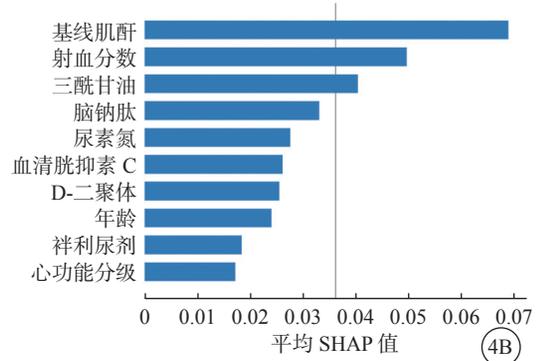
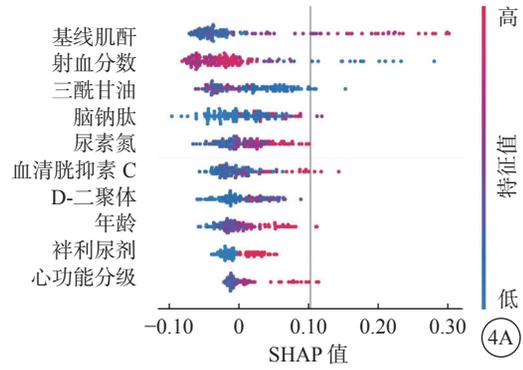
本研究使用 Logistic 回归、决策树、随机森林、XGboost、SVC 和 LightGBM 方法构建 ACS-AKI 的预测模型, 且更进一步的通过模型可视化工具 SHAP 与 RFE 的方法对最佳 ML 预测模型变量进行筛选, 并压缩为特征减少的紧凑预测模型。将变量按照重要性排序, 发现入院 Scr 是排名第一的预测因子。Scr、BUN 是评价肾脏功能最常见的重要指标, 以及 2012 年 KDIGO AKI 指南<sup>[7]</sup>对 AKI 的诊断也是通过 Scr 和尿量的变化来规定的。BUN 是人体氮代谢的主要最终产物, 常与 Scr 一起提供肾脏损伤的信息<sup>[19]</sup>。近年来发现胱抑素 C 是一种参与细胞内蛋白和肽分解代谢的内源性半胱氨酸蛋白酶抑制剂, 其半衰期比 Scr 更短, 因此胱抑素 C 水平升高可以合理地视为肾功能迅速恶化为特征的 AKI 的指标<sup>[20-22]</sup>。有项研究使用结合胱抑素 C 的列线图来预测颅脑损伤患者的 AKI(AUC=0.804)<sup>[22]</sup>。

ACS 患者血流动力学紊乱是引起 AKI 的原因, ACS 患者可出现心肌收缩力下降、心输出量减少, 导致肾血流量和肾小球滤过率随之下降<sup>[23]</sup>。而射血分数、BNP、心肌梗死范围这些指标与 ACS 患者的血流动力学变化密切相关, 因此重点关注这些指标变化也在提示临床医生要关注肾脏保护。本次研究同样发现年龄也是 ACS-AKI 的预测因子, 随着年龄的增长, 特定的结构、功能、血液动力学和细胞变化使肾脏在压力状态下容易受到损伤<sup>[24]</sup>。



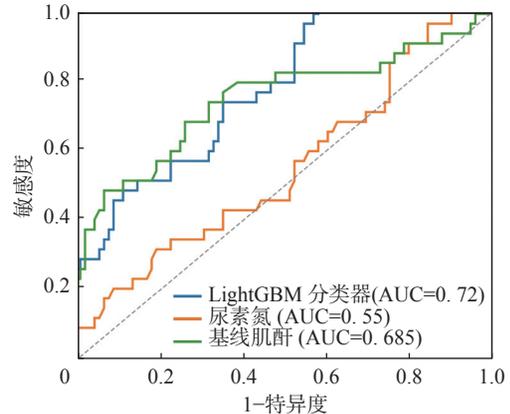
注: SHAP 在图中为患者每一个特征都分配了点, 特征由高低重要性依次下降, 颜色代表特征值的大小(红色高, 蓝色低); X 轴用于衡量特征对急性肾损伤的影响(右侧为正, 左侧为负, 数值越大, 影响越强); AST 为天冬氨酸转氨酶; CK.MB 为肌酸激酶同工酶; ALT 为丙氨酸转氨酶; LAD 为左房内径; PT 为血浆血酶原时间; ARB 为血管紧张素受体拮抗剂; ACEI 为血管紧张素转化酶抑制剂; SHAP 为 SHAP 可解释模型; SHAP 为 SHAP 可解释模型。

图 3 使用 SHAP 值解释变量重要性



注: SHAP 在图中为患者每一个特征都分配了点, 特征由高低重要性依次下降, 颜色代表特征值的大小(红色高, 蓝色低); X 轴用于衡量特征对急性肾损伤的影响(右侧为正, 左侧为负, 数值越大, 影响越强); SHAP 为 SHAP 可解释模型。

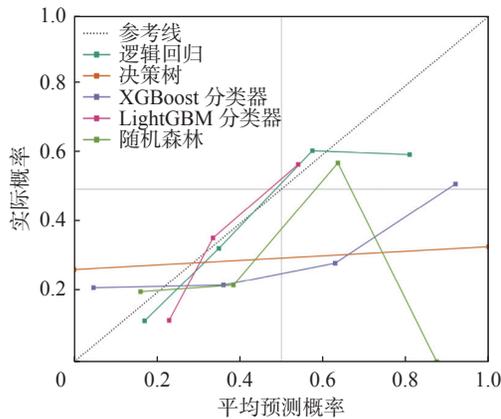
图 4 可视化展示各变量重要性 4A. 使用 SHAP 值解释变量重要性; 4B. 各个变量对模型的影响程度



注: LightGBM 为轻量级梯度提升; ROC 为受试者工作特征曲线; AUC 为 ROC 的曲线下面积。

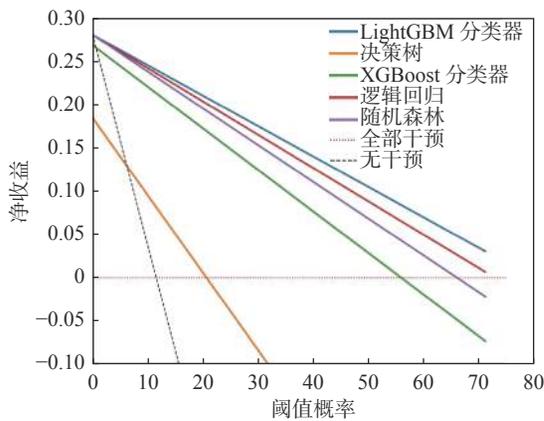
图 5 特征选择后紧凑模型与基线血肌酐或尿素氮预测急性肾损伤性能的比较

肾小球滤过率和肾血流量均随年龄增长而下降, 再加之患者往往合并有更多的慢性疾病、药物解毒能力下降, 使肾组织更容易受伤, 难以修复<sup>[25]</sup>。此次研究和国外 Pocock 等<sup>[26]</sup>研究结果相似, 他们研究表明大约有 5.5% 的 ACS 患者在出院后 2 年内死亡, 并确定了独立的死亡预测因素有年龄、低射



注: X 轴表示预测急性肾损伤的概率; Y 轴表示急性肾损伤发生的实际概率; 虚线代表理想曲线, 越贴近虚线, 一致性越高; XG-Boost 为极端梯度提升; LightGBM 为轻量级梯度提升。

图 6 不同机器学习模型的校准曲线



注: 决策曲线的 X 轴为阈概率; Y 轴为每个阈概率对应的净获益; 灰色虚线为所有患者都接受干预的获益率; 粉色虚横线为所有患者都不接受干预无获益; 取与 All 的交点为起点, 与 None 的交点为结束, 在此范围内为所对应的净收益总情况; XGBoost 为极端梯度提升; LightGBM 为轻量级梯度提升。

图 7 不同机器学习模型的临床决策曲线

血分数、Scr 升高、出院时服用利尿剂或醛固酮抑制剂等。除此之外, 肾毒性药物一直也是 AKI 的最大危险因素。本研究发现袢利尿剂是 ACS 患者发生 AKI 的预测因子, 如 Zhang 等<sup>[27]</sup> 研究发现有 14.26% 的患者发生 AKI 是由于利尿剂的使用导致的, 并且 90.00% 以上的患者未及时发现和诊断 AKI。袢利尿剂可以增加尿量, 但是可以通过影响肾小管球反馈的机制直接或间接地引起肾血灌注和肾小球滤过率的显著降低, 从而导致肾缺血和缺氧<sup>[28-29]</sup>。我们建议在患者接受利尿剂治疗之前, 需要充分评估 AKI 的危险因素, 以避免 AKI 的发生。

在人工智能等技术飞速发展的今天, ML 带来了对 AKI 进行更加精确预测的可能性。本研究基于 ML 的方法成功建立了 ACS-AKI 的预测模型, 为临床医生早期重视 ACS 患者合并 AKI 风险, 以

期进行早期识别、预防、干预 AKI 提供帮助。但此研究也有一定局限性, 本研究为单中心回顾性研究, 并且其结果缺乏外部验证。

利益冲突 所有作者均声明没有利益冲突

### 参 考 文 献

- [1] 孟洁, 苏立贞, 晋婕. 急性冠脉综合征患者血清 HMGB1、LP-PLA2、IMA 的表达水平及其与预后的关系[J]. 海南医学, 2023, 34(11): 1560-1564.  
Meng J, Su LZ, Jin J. Expression of serum HMGB1, LP-PLA2, and IMA in patients with acute coronary syndrome and their relationship with prognosis[J]. Hainan Med J, 2023, 34(11): 1560-1564.
- [2] Kellum JA, Romagnani P, Ashuntantang G, et al. Acute kidney injury[J]. *Nat Rev Dis Primers*, 2021, 7(1): 52. DOI: 10.1038/s41572-021-00284-z.
- [3] Oweis AO, Alshelleh SA, Saadeh N, et al. Long-term follow-up of contrast-induced acute kidney injury: a study from a developing country[J]. *Int J Vasc Med*, 2020, 2020: 1-6. DOI: 10.1155/2020/8864056.
- [4] Macedo E, Mehta RL. Preventing acute kidney injury[J]. *Crit Care Clin*, 2015, 31(4): 773-784. DOI: 10.1016/j.ccc.2015.06.011.
- [5] Perico N, Remuzzi G. Prevention programs for chronic kidney disease in low-income countries[J]. *Intern Emerg Med*, 2016, 11(3): 385-389. DOI: 10.1007/s11739-016-1425-7.
- [6] Gameiro J, Branco T, Lopes JA. Artificial intelligence in acute kidney injury risk prediction[J]. *J Clin Med*, 2020, 9(3): 678. DOI: 10.3390/jcm9030678.
- [7] Kellum J, Lameire N, Aspelin P, et al. Kidney disease: improving global outcomes (KDIGO) acute kidney injury work group. KDIGO clinical practice guideline for acute kidney injury[J]. *Kidney Int*, 2012, 2: 1-138. DOI: 10.1038/KISUP.2012.1.
- [8] Loh WY. Classification and regression trees[J]. *WIREs Data Mining Knowl Discov*, 2011, 1(1): 14-23. DOI: 10.1002/widm.8.
- [9] Segev N, Harel M, Mannor S, et al. Learn on source, refine on target: a model transfer learning framework with random forests[J]. *IEEE Trans Pattern Anal Mach Intell*, 2017, 39(9): 1811-1824. DOI: 10.1109/TPAMI.2016.2618118.
- [10] Ramraj S, Uzir N, Sunil R, et al. Experimenting XGBoost Algorithm for Prediction and Classification of Different Datasets[C] NCRISSECT, 2017.
- [11] Ke GL, Meng Q, Finley T, et al. LightGBM: a highly efficient gradient boosting decision tree[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, 2017: 3149-3157. DOI: 10.5555/3294996.3295074.
- [12] Lundberg SM, Erion GG, Lee SI. Consistent individualized feature attribution for tree ensembles[EB/OL]. 2018: arXiv: 1802.03888. <https://arxiv.org/abs/1802.03888>.
- [13] Parikh CR, Coca SG, Wang YF, et al. Long-term prognosis of acute kidney injury after acute myocardial infarction[J]. *Arch Intern Med*, 2008, 168(9): 987. DOI: 10.1001/archinte.168.9.987.
- [14] Ma K, Li J, Shen GQ, et al. Development and validation of a risk nomogram model for predicting contrast-induced acute kidney injury in patients with non-ST-elevation acute coronary syndrome undergoing primary percutaneous coronary intervention[J]. *Clin In-*

- terv Aging, 2022, 17: 65-77. DOI: [10.2147/cia.s349159](https://doi.org/10.2147/cia.s349159).
- [15] 王聪, 裴源源, 朱继红. 急性ST段抬高性心肌梗死患者发生急性肾损伤的危险因素[J]. 中国急救医学, 2021, 41(6): 461-464. DOI: [10.3969/j.issn.1002-1949.2021.06.001](https://doi.org/10.3969/j.issn.1002-1949.2021.06.001).
- Wang C, Pei YY, Zhu JH. Risk factors of acute kidney injury in the patients with acute ST segment elevation myocardial infarction[J]. Chin J Crit Care Med, 2021, 41(6): 461-464. DOI: [10.3969/j.issn.1002-1949.2021.06.001](https://doi.org/10.3969/j.issn.1002-1949.2021.06.001).
- [16] Fan PC, Chen TH, Lee CC, et al. ADVANCIS score predicts acute kidney injury after percutaneous coronary intervention for acute coronary syndrome[J]. Int J Med Sci, 2018, 15(5): 528-535. DOI: [10.7150/ijms.23064](https://doi.org/10.7150/ijms.23064).
- [17] Rajkomar A, Dean J, Kohane I. Machine learning in medicine[J]. N Engl J Med, 2019, 380(14): 1347-1358. DOI: [10.1056/nejmra1814259](https://doi.org/10.1056/nejmra1814259).
- [18] 蓝潞杭, 蒋炫东, 王茂峰, 等. 随机森林模型预测急性心肌梗死后急性肾损伤[J]. 中华急诊医学杂志, 2021, 30(4): 491-495. DOI: [10.3760/cma.j.issn.1671-0282.2021.04.021](https://doi.org/10.3760/cma.j.issn.1671-0282.2021.04.021).
- Lan LH, Jiang XD, Wang MF, et al. A random forest model to predict acute kidney injury after acute myocardial infarction[J]. Chin J Emerg Med, 2021, 30(4): 491-495. DOI: [10.3760/cma.j.issn.1671-0282.2021.04.021](https://doi.org/10.3760/cma.j.issn.1671-0282.2021.04.021).
- [19] Pundir CS, Jakhar S, Narwal V. Determination of urea with special emphasis on biosensors: a review[J]. Biosens Bioelectron, 2019, 123: 36-50. DOI: [10.1016/j.bios.2018.09.067](https://doi.org/10.1016/j.bios.2018.09.067).
- [20] Tan KR, Sethi SK. Biomarkers in cardiorenal syndromes[J]. Transl Res, 2014, 164(2): 122-134. DOI: [10.1016/j.trsl.2014.04.011](https://doi.org/10.1016/j.trsl.2014.04.011).
- [21] Shlipak MG, Mattes MD, Peralta CA. Update on cystatin C: incorporation into clinical practice[J]. Am J Kidney Dis, 2013, 62(3): 595-603. DOI: [10.1053/j.ajkd.2013.03.027](https://doi.org/10.1053/j.ajkd.2013.03.027).
- [22] Wang RR, He M, Gui XY, et al. A nomogram based on serum cystatin C for predicting acute kidney injury in patients with traumatic brain injury[J]. Ren Fail, 2021, 43(1): 206-215. DOI: [10.1080/0886022x.2021.1871919](https://doi.org/10.1080/0886022x.2021.1871919).
- [23] Lullo L, Bellasi A, Barbera V, et al. Cardioneurology and cardiorenal disease in Italy: state of the art[J]. Rev Cardiovasc Med, 2021, 22(3): 563. DOI: [10.31083/j.rcm2203068](https://doi.org/10.31083/j.rcm2203068).
- [24] Rosner M. The pathogenesis of susceptibility to acute kidney injury in the elderly[J]. Curr Aging Sci, 2009, 2(2): 158-164. DOI: [10.2174/1874609810902020158](https://doi.org/10.2174/1874609810902020158).
- [25] Formica M, Politano P, Marazzi F, et al. Acute kidney injury and chronic kidney disease in the elderly and polypharmacy[J]. Blood Purif, 2018, 46(4): 332-336. DOI: [10.1159/000492149](https://doi.org/10.1159/000492149).
- [26] Pocock S, Huo Y, Van de Werf F, et al. Predicting two-year mortality from discharge after acute coronary syndrome: an internationally-based risk score[J]. Eur Heart J, 2019, 8: 727-737. DOI: [10.1177/2048872617719638](https://doi.org/10.1177/2048872617719638).
- [27] Zhang RQ, Liu YX, Cao JA, et al. The incidence and risk factors analysis of acute kidney injury in hospitalized patients received diuretics: a single-center retrospective study[J]. Front Pharmacol, 2022, 13: 924173. DOI: [10.3389/fphar.2022.924173](https://doi.org/10.3389/fphar.2022.924173).
- [28] Liu C, Yan SY, Wang YQ, et al. Drug-induced hospital-acquired acute kidney injury in China: a multicenter cross-sectional survey[J]. Kidney Dis, 2021, 7(2): 143-155. DOI: [10.1159/000510455](https://doi.org/10.1159/000510455).
- [29] Perazella MA, Rosner MH. Drug-induced acute kidney injury[J]. Clin J Am Soc Nephrol, 2022, 17(8): 1220-1233. DOI: [10.2215/cjn.11290821](https://doi.org/10.2215/cjn.11290821).

(收稿日期: 2022-08-22)