

## ICU 患者急性肾损伤发生风险的 LightGBM 预测模型

张 渊<sup>1</sup>, 冯 聪<sup>2</sup>, 李开源<sup>1</sup>, 张政波<sup>3,4,5</sup>, 曹德森<sup>3,5</sup>, 黎檀实<sup>2</sup>

<sup>1</sup>解放军医学院, 北京 100853; <sup>2</sup>解放军总医院第一医学中心 急诊科, 北京 100853; <sup>3</sup>解放军总医院医学工程保障中心, 北京 100853; <sup>4</sup>解放军总医院 医疗大数据中心, 北京 100853; <sup>5</sup>解放军总医院 医疗器械研发与评价中心, 北京 100853

**摘要:** **目的** 基于机器学习模型 LightGBM 构建 ICU 患者发生急性肾损伤 (acute kidney injury, AKI) 的预测模型, 为临床医护人员提供辅助决策支持。**方法** 采用公开的大型 ICU 数据库重症监护医学信息数据库 (MIMIC- III) 作为数据集, 提取 1 166 例患者, 其中男性 513 例 (44.00%), 中位年龄 70.93 岁, 75.8%(884 例) 的患者发展为 AKI。以患者入 ICU 时的生理生化指标为预测变量, 预测患者 24 h 后是否发展为 AKI。采用 LightGBM 构建预测模型, 并与 logistic 回归及随机森林模型进行对比, 采用五折交叉验证评价模型性能。**结果** 结果显示, LightGBM 模型对 AKI 预测的准确率为 0.89, AUC 为 0.92; logistic 回归模型和随机森林模型的 AUC 分别为 0.75 和 0.89, 准确率为 0.84 和 0.86。**结论** LightGBM 在 AKI 预测模型中表现最优, 采用患者入 ICU 时的生理生化指标, 预测模型准确率和 AUC 可达 0.89 和 0.92。

**关键词:** 医疗大数据; MIMIC- III 数据库; 急性肾损伤; LightGBM; 预测模型

中图分类号: R 692.5 文献标志码: A 文章编号: 2095-5227(2019)04-0316-05 DOI: 10.3969/j.issn.2095-5227.2019.04.004

网络出版时间: 2019-04-11 10:56 网络出版地址: <http://kns.cnki.net/kcms/detail/10.1117.R.20190411.1056.002.html>

引用本文: 张渊, 冯聪, 李开源, 等. ICU 患者急性肾损伤发生风险的 LightGBM 预测模型 [J]. 解放军医学院学报, 2019, 40 (4): 316-320.

### LightGBM model for predicting acute kidney injury risk in ICU patients

ZHANG Yuan<sup>1</sup>, FENG Cong<sup>2</sup>, LI Kaiyuan<sup>1</sup>, ZHANG Zhengbo<sup>3,4,5</sup>, CAO Desen<sup>3,5</sup>, LI Tanshi<sup>2</sup>

<sup>1</sup>Chinese PLA Medical School, Beijing 100853, China; <sup>2</sup>Emergency Department, the First Medical Center, Chinese PLA General Hospital, Beijing 100853, China; <sup>3</sup>Department of Biomedical Engineering and Maintenance Center, Chinese PLA General Hospital, Beijing 100853, China; <sup>4</sup>Medical Information Center, Chinese PLA General Hospital, Beijing 100853, China; <sup>5</sup>Department of Medical Device R&D and Evaluation Center, Chinese PLA General Hospital, Beijing 100853, China

Corresponding author: ZHANG Zhengbo. Email: zhengbozhang@126.com; LI Tanshi. Email: lts301@163.com

**Abstract: Objective** To develop a predictive model for acute kidney injury (AKI) in critically ill patients based on LightGBM and provide support for clinical decision. **Methods** An open source critical care database, MIMIC- III was used in this study. A total of 1 166 patients were included in this study with median age of 70.93 years, of which 513 cases (44.00%) were male, and 884 of them had developed AKI (75.8%). To predict the occurrence of AKI after 24 hours, the model was built utilizing predictors including laboratory tests and vital signs measured at ICU admission. LightGBM, logistic regression, and random forest models were established to predict the risk of AKI, and their predictive performances were evaluated using five-fold cross-validation. **Results** The accuracy of LightGBM was 0.89, and the AUC was 0.92. However, the accuracy of logistic regression and random forest were 0.84 and 0.86, and their AUC were 0.75 and 0.89, respectively. **Conclusion** LightGBM performs well in predicting AKI at 24 hours after ICU admission. The accuracy and AUC of LightGBM model are up to 0.89 and 0.92.

**Keywords:** medical big data; MIMIC- III database; acute kidney injury; light gradient boosting machine; prediction model

**Cited as:** Zhang Y, Feng C, Li KY, et al. LightGBM model for predicting acute kidney injury risk in ICU patients [J]. Acad J Chin PLA Med Sch, 2019, 40 (4): 316-320.

收稿日期: 2019-01-09

基金项目: 国家自然科学基金面上项目 (61471398); 北京市科委医药协同科技创新研究项目 (Z181100001918023); 首长保健专项 (16BJZ23); 解放军总医院医疗大数据研发项目 (2017MBD-30); 解放军总医院医疗大数据研发项目 (2016MBD-027)

Supported by the National Natural Science Foundation of China (61471398); Beijing Municipal Science and Technology Project (Z181100001918023); Special Grant for Healthcare (16BJZ23); Big Data Research and Development Project of Chinese PLA general hospital (2017MBD-30); Big Data Research and Development Project of Chinese PLA general hospital (2016MBD-027)

作者简介: 张渊, 女, 在读硕士。研究方向: 医疗数据分析。Email: yuanzhang22@126.com

通信作者: 张政波, 男, 博士, 高级工程师, 生物医学工程教研室主任, 医学工程保障中心副主任。Email: zhengbozhang@126.com; 黎檀实, 男, 博士, 主任医师, 急诊科主任。Email: lts301@163.com

急性肾损伤 (acute kidney injury, AKI) 是指突发或持续的肾功能突然下降, 临床上表现为氮质血症、水电解质和酸碱平衡紊乱及全身各系统症状, 可伴有少尿或无尿。在重症监护室 (intensive care unit, ICU) 中, 急性肾损伤有较高的发病率和死亡率, 医疗成本非常高<sup>[1-2]</sup>。对 AKI 进行早期预测一直是热点研究领域, 识别患者发生 AKI 的风险, 进行早期 AKI 预警能够提示医护人员在患者恶化早期进行干预, 进而改善患者预后, 降低死亡率, 减少医疗花费。目前临床中最常采用国际改善全球肾脏病预后组织 (Kidney Disease: Improving Global Outcomes, KDIGO) 指南对 AKI 进行判断, 指南中采用肌酐与尿量标准对 AKI 进行分级, 为临床研究提供了重要的参考。随着电子健康档案 (electronic health records, EHR) 的发展, 我们得以用大数据的研究方法探究疾病的演化规律, 这也为我们深入研究 AKI 提供了机会。Kristovic 等<sup>[3]</sup>用 logistic 回归模型建立了心脏手术后患者发生 AKI 的预测模型, 基于 KDIGO 中的肌酐标准判断 AKI 不同分期, 对不同分期进行了预测, 模型 AUC 为 0.78。Basu 等<sup>[4]</sup>基于多元回归模型提出了肾性心绞痛指数预测 AKI (KDIGO 中肌酐与尿量标准), 在不同医院的小样本人群中进行了验证, AUC 为 0.74 ~ 0.81。Flechet 等<sup>[2]</sup>开发了在线的 AKI 预测系统, 采用随机森林模型对 AKI (KDIGO 中的肌酐标准) 进行预测, 入 ICU 24 h 后发展为 AKI 的 AUC 为 0.75 (95% CI, 0.75 ~ 0.75)。Mohamadlou 等<sup>[5]</sup>2018 年基于机器学习模型发展了 AKI 预测模型, 基于肌酐判断 AKI 分期, 对 AKI 的不同分期进行预测并与传统的评分做对比, 模型 AUC 最高可达 0.872 (95% CI, 0.867 ~ 0.878), 但提前 24 h 的预测模型 AUC 仅为 0.795 (95% CI, 0.785 ~ 0.804)。数据挖掘与机器学习技术的发展为我们探索医疗大数据、挖掘新的知识提供了技术手段, 为发展性能更好的预测模型提供了可能。本文拟采用重症监护医学信息数据库 (Medical Information Mart for Intensive Care, MIMIC-III) 的数据构建 AKI 预测模型<sup>[6]</sup>, 提前 24 h 预测患者是否会发展为 AKI, 为临床决策提供辅助支持。

## 资料与方法

**1 资料来源** 本研究采用 MIMIC-III 数据, MIMIC-III 数据库是由麻省理工学院计算生理学实验室、哈佛医学院贝斯以色列女执事医学中心和飞利浦医疗联合开发的多参数重症监护数据库, 包含了

2001 - 2012 年来自贝斯以色列女执事医学中心超过 4 万例 ICU 病人的近 6 万份临床数据, 包括患者详细人口学信息、生理生化检查、诊断信息、医疗干预记录等。MIMIC-III 数据库经过了整合、高度结构化和去隐私化处理, 具有样本量大、质量高、医疗记录详细的特点。纳入标准: 1) 年龄 > 18 岁; 2) 入院肌酐测量值 < 4; 3) ICU 治疗时长 > 24 h。排除标准: 1) 有终末期肾病; 2) 在 ICU 期间呼吸心脏骤停; 3) 尿路梗阻; 4) 生理生化测量值缺失超过 70%。

**2 AKI 定义** 参考 KDIGO 指南标准: AKI 定义为 48 h 内血肌酐增高  $\geq 0.3$  mg/dl ( $> 26.5 \mu\text{mol/L}$ ), 或血肌酐增高至  $\geq$  基础值的 1.5 倍, 且是已知或经推断发生在 7 d 之内; 或持续 6 h 尿量 < 0.5 ml/(kg·h)<sup>[7-8]</sup>。由于 MIMIC-III 数据库仅包含患者入 ICU 之后的数据, 患者肌酐测量值的基线数据并不可知, 以此时的肌酐来表征基线值显然不准确。本文采用 KDIGO 中的尿量标准作为 AKI 的判定标准, 即持续 6 h 尿量 < 0.5 ml/(kg·h)。Kellum 等<sup>[9]</sup>对基于尿量标准判断 AKI 进行了研究, 并与基于肌酐标准进行了对比, 研究结果表明根据尿量标准纳入的病人相对客观。

**3 预测变量** 1) 人口学信息, 包括年龄、性别; 2) 生理指标: 心率、呼吸率、平均动脉压、平均收缩压、平均舒张压、体温、血氧饱和度、动脉氧分压、动脉二氧化碳分压; 3) 常用的实验室检查: 动脉血 pH 值、红细胞比容、白细胞计数、中性粒细胞百分比、乳酸、血清肌酐、血红蛋白、碳酸氢盐、白蛋白、血小板计数、总胆红素、心肌肌钙蛋白、肌酸激酶、尿素氮、活化部分凝血活酶时间、凝血酶原时间、尿比重; 4) 相关干预: 是否进行机械通气, 是否使用血管升压药物; 5) 液体入量、出量。

**4 预测模型** 采用 LightGBM 构建 AKI 预测模型, 与 logistic 回归及随机森林预测模型进行对比, 找到最佳的预测模型。1) logistic 回归模型: 一种对数线性模型。多元 logistic 回归中一个事件发生的对数概率为:

$$\log[p(x)/1-p(x)] = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p,$$

则事件发生的概率为:

$$p(X) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}$$

2) 随机森林模型: 随机森林是基于决策树发展的机器学习模型。基于给定的特征, 决策树算法将特征空间划分为互不交叉的单元, X 表示特征的集合, Y 表示类别的集合, 那么给定一个观测, 它被判断为某

一类的条件概率为： $P(y_i | x_i), x_i=(x_1, x_2, \dots, x_n), y_i \in \{1, 2, \dots, K\}^{[10]}$ 。其属于哪一个类的概率较大，分类时则将该节点划分为概率大的一类中。随机森林是通过有放回的抽样，从原数据中随机选取数据子集与特征，构建多个决策树，以多个决策树分类结果的众数作为最终输出。3) LightGBM 模型：对基于决策树发展出的梯度提升决策树算法 (gradient boosting decision tree, GBDT) 进行了优化，支持高效率的并行训练，可以快速处理大量数据。其思想是利用弱分类器迭代训练得到最优模型，并且采用梯度提升算法减少结构分数的计算量，在此基础上，采用基于直方图的算法选择分割点，并且使用 leaf-wise 策略替代 level-wise 分裂子节点，可简化计算并提高准确率。预测模型采用 Python 中的 LightGBM 及 Scikit-learn 软件包进行建模，协变量为患者发生少尿前 24 h 的生理生化指标及其他测量值，预测变量为患者是否发展为 AKI。随机选取 75% 的数据作为训练集，25% 作为测试集，采用五折交叉验证评价模型性能，评价指标选择受试者工作特征曲线下的面积。AUC 越大，模型的预测性能越好。

**5 统计学分析** 本研究使用 PostgreSQL 数据库对数据进行提取和清洗，使用 R 3.4.3 对数据进行统计分析。采用 Kolmogorov-Smirnov 检验计量数据正态性，符合正态分布的计量资料，以  $\bar{x} \pm s$  来表示，组间比较采用成组  $t$  检验。偏态分布的计量资料以中位数 (四分位数) 表示，组间比较采用 Wilcoxon 秩和检验。计数资料用例数 (百分比) 表示，用  $\chi^2$  检验比较组间差异。预测模型建立及评价见 4。 $P < 0.05$  为差异有统计学意义。

## 结 果

**1 纳入人群基线特征** MIMIC-III 数据库中共有 61 532 例次 ICU 诊疗记录，去掉 14 623 例重复住院记录，剩余 46 428 例患者记录。其中 38 468 例成年患者，排除 759 例患有终末期肾病、1 117 例入院时肌酐  $> 4$ 、1 445 例有呼吸心脏骤停的患者，排除 13 291 例在入 ICU 24 h 内已经发展为 AKI 的患者，排除 11 例患有尿路梗阻的患者，排除 20 680 例生理生化测量值缺失 70% 的患者，最终有 1 166 例患者入选。在 1 166 例患者中，AKI 患者 884 例，非 AKI 患者 282 例。相比非 AKI

组，AKI 组中位年龄较小 (70.30 岁 vs 73.37 岁， $P=0.023$ )，ICU 治疗时间更长 (9.61 d vs 2.47 d， $P < 0.001$ )，住院时间更长 (14.10 d vs 6.66 d， $P < 0.001$ )，死亡率更高 (33.1% vs 22.7%， $P < 0.001$ )。两组患者的心率、血压、体温、血氧饱和度等指标有统计学差异 (表 1)。

**2 三种预测模型性能比较** 纳入患者 33 项生理生化指标及相关干预构建预测模型，LightGBM 模型参数设置：feature\_fraction=1, bagging\_fraction=0.8, leaves=300, learning\_rate=0.4。采用五折交叉验证，测试集 AUC 为 0.92。LightGBM 特征重要性排名前 10 位分别是液体摄入量、红细胞比容、患者进行了机械通气、动脉氧分压、乳酸、体温、动脉血 pH、心肌肌钙蛋白、血小板计数、凝血酶原时间 (图 1)。LightGBM 的预测准确率为 0.89，AUC 为 0.92；logistic 回归预测的准确性为 0.84，AUC 为 0.75；随机森林的准确率与 AUC 分别是 0.86 和 0.89 (图 2)。LightGBM 的准确率与 AUC 均优于 logistic 回归与随机森林。对比结果见表 2。

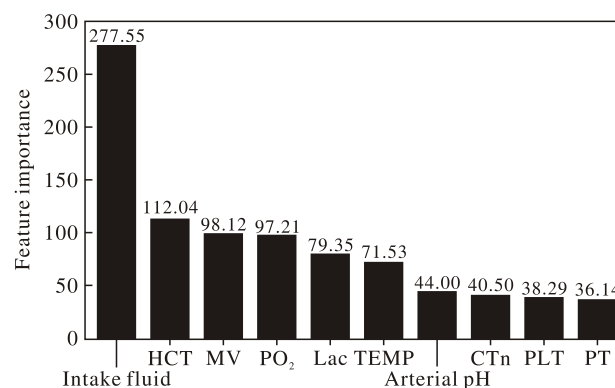


图 1 LightGBM 特征重要性排名

Fig.1 Top 10 important variables in LightGBM

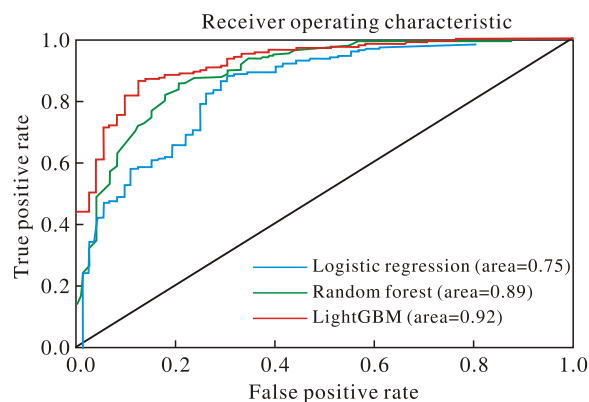


图 2 三种方法的 ROC 曲线对比

Fig.2 ROC curves for three AKI predict models

表 1 AKI 组与非 AKI 组基线特征比较

Tab. 1 Baseline characteristics in AKI and non-AKI patients [ $Med(P_{25}, P_{75})$ ]

Characteristics	Non-AKI (n=282)	AKI (n=884)	P
Age (yrs)	73.37(60.94,83.19)	70.30(58.17,80.84)	0.023
Female (n, %)	134(47.5)	379(42.9)	0.194
LOS ICU (d)	2.47(1.81,4.06)	9.61(4.95,17.40)	< 0.001
LOS hospital (d)	6.66(4.11,10.33)	14.10(8.07,23.91)	< 0.001
Hospital expire (n, %)	64(22.7)	293(33.1)	0.001
Heart rate (/min)	94.00(79.00,112.00)	86.00(73.00,100.00)	< 0.001
SBP (mmHg)	120.00(103.00,137.75)	114.00(100.00,130.00)	0.001
DBP (mmHg)	64.00(52.25,77.00)	57.00(49.00,66.00)	< 0.001
MBP (mmHg)	78.00(69.00,90.00)	75.00(66.00,84.33)	0.001
Respiratory rate (times/min)	20.00(16.00,24.00)	21.00(17.00,25.00)	0.163
Temperature (°C)	36.61(35.89,37.26)	36.89(36.39,37.39)	< 0.001
SpO <sub>2</sub> (%)	98.00(95.00,100.00)	97.00(95.00,99.00)	0.012
Arterial pH	7.33(7.27,7.41)	7.40(7.35,7.45)	< 0.001
Bicarbonate (mEq/L)	21.50(18.00,25.00)	24.00(21.00,27.25)	< 0.001
Hemoglobin (mg/dl)	11.40(9.90,13.00)	10.10(9.20,11.20)	< 0.001
Arterial WBC ( $\times 10^9/L$ )	12.50(8.30,17.38)	12.00(8.70,16.42)	0.757
Arterial RBC ( $\times 10^{12}/L$ )	3.75(3.24,4.38)	3.39(3.01,3.72)	< 0.001
Arterial neutrophils (%)	83.00(75.93,89.38)	83.05(74.97,89.00)	0.948
Hematocrit (%)	34.20(29.50,39.00)	30.20(27.40,33.23)	< 0.001
PLT ( $\times 10^9/L$ )	228.50(154.00,306.75)	193.50(118.00,297.00)	0.001
Lactate (mg/dl)	2.40(1.50,4.30)	1.50(1.10,2.10)	< 0.001
ALB (g/dl)	3.20(2.80,3.60)	2.90(2.40,3.40)	< 0.001
Total-bilirubin (mg/dl)	0.70(0.40,1.20)	0.70(0.40,1.50)	0.114
Troponin T (ng/ml)	0.07(0.03,0.22)	0.11(0.04,0.52)	< 0.001
CK (U/L)	173.00(86.25,421.00)	186.00(79.00,551.50)	0.798
Creatinine (mg/dl)	1.40(1.00,2.10)	1.20(0.80,2.00)	0.001
BUN (mg/dl)	28.00(18.00,48.00)	30.00(18.00,48.25)	0.992
APTT (s)	30.30(25.80,36.70)	33.10(27.90,45.10)	< 0.001
PT (s)	14.50(13.10,17.10)	14.30(13.20,16.50)	0.442
Fluid intake (ml)	3 387.87(2 094.18,5 433.14)	1 319.00(739.88,2 429.68)	< 0.001
Fluid output (ml)	1 791.50(1 085.75,2 888.25)	1 640.00(950.00,2 481.00)	0.020
Mechanical ventilation = Yes (n, %)	150(53.2)	171(19.3)	< 0.001
Vasopressor use = Yes (n, %)	134(47.5)	194(21.9)	< 0.001

表 2 三种模型预测性能比较

Tab. 2 Predictive performance of three models

Item	Accuracy	AUC	TPR	TNR
LightGBM	0.89	0.92	0.95	0.68
Random forest	0.86	0.89	0.97	0.56
Logistic regression	0.84	0.75	0.97	0.54

TPR: true positive rate; TNR: true negative rate

## 讨论

AKI 是临床中常见的危重病症, 其早期诊断和治疗一直是研究者们关注的重点。本研究基于 MIMIC-III 数据库, 采用 LightGBM 模型对重症患者入院 24 h 后发展为 AKI 进行预测分析, 并与 logistic 回归、随机森林模型预测效果进行对比。LightGBM 的预测准确率可达 0.89, AUC 可达 0.92,

优于传统的预测模型。

临床研究中最常用的分类模型是 logistic 回归模型。传统的回归模型可解释性强, 然而在数据量大、数据维度高的情况下的处理能力较差。随机森林是经典的机器学习模型, 近年来基于随机森林的预测模型在医疗领域应用广泛, 它适合于处理体量大、维度高的数据。在随机森林的基础上, 机器学习领域又发展出了一些提升算法, 比如 Xgboost、adaboost 等。LightGBM 是近年来新发展出的一种基于决策树算法的梯度提升算法框架<sup>[1]</sup>, 采用 left-wise 的分裂方式对树模型进行了进一步优化, 从理论上实践上都达到了最优性能。此模型基于重症患者的多项生理生化指标, 能够在重症患者人群中为临床提供决策支持。

LightGBM 预测模型中纳入了临床中常用的生理生化指标, 模型特征重要性排名中前 5 项重要的指标为液体入量、红细胞比容、机械通气、动脉氧分压以及乳酸。研究结果显示, 液体入量是预测 AKI 最重要的特征参数, 这提示医护人员在临床工作中要密切关注患者的液体出入量信息, 在维持必要的循环血量同时, 防止液体过量。已有研究表明, 过多或过少的液体负荷都有可能加重 AKI<sup>[12]</sup>。而血气分析指标, 如血 pH 值、乳酸值可间接提示患者的组织及器官灌注情况, 对 AKI 的发生具有预测价值<sup>[13-14]</sup>。而进行机械通气以及发生贫血, 常提示患者疾病严重程加重、氧输送不足以及多脏器衰竭发生的可能, 已有研究显示机械通气和贫血是 AKI 发生的独立危险因素<sup>[15]</sup>。值得注意的是, 肌酐在本研究人群中预测模型中的重要性并不高, 对比两组的肌酐值, AKI 组肌酐为 1.2(0.8, 2.0) mg/dl, 对照组为 1.4(1.00, 2.10) mg/dl,  $P=0.001$ , 差异并不大。这提示肌酐是判断 AKI 一项比较滞后的指标<sup>[16]</sup>。因此肌酐预测 24 h 后 AKI 的发生并不是很理想。

本研究也存在不足之处: 首先, 数据集采用了 MIMIC-III 数据库, MIMIC-III 数据库是单中心的 ICU 数据库, 研究人群来源于单中心的危重症患者, 模型对其他数据集、数据库的普适性需要进一步验证, 如使用 eICU 协作研究数据库 (eICU collaborative research database)<sup>[17]</sup>。其次, 由于患者入院的肌酐基线值难以确定, 本研究仅以尿量来判断患者 AKI 的发生, 这种判断方式可能会混入一些非典型的急性肾损伤患者, 不可避免地带来了选择偏倚。

下一步研究将着眼于将模型推广到更广泛的人群中进行验证, 发展更普适的模型。同时, 解放军总医院建立了基于中国人群的高度结构化的急诊数据库<sup>[18]</sup>, 下一步将基于此数据库对 AKI 预测模型进行验证和优化, 探索将预测模型移植到急诊业务系统的可行性, 为急诊医护人员提供更好的决策支持。

综上所述, 本研究采用 MIMIC-III 数据库发展了基于 LightGBM 算法的 AKI 预测模型, 对患者进入 ICU 24 h 后是否发展为 AKI 进行了预测分析, 相比于传统回归模型与常用的机器学习模型, LightGBM 模型的性能更优, 能够更好地辅助 ICU 医护人员监测患者的恶化风险, 及时对患者进行

干预, 进而提高医疗质量。

#### 参考文献

- 1 Alshaikh HN, Katz NM, Gani F, et al. Financial Impact of Acute Kidney Injury After Cardiac Operations in the United States [J]. *Ann Thorac Surg*, 2018, 105 (2): 469-475.
- 2 Flechet M, Guiza F, Schetz M, et al. AKIpredictor, an online prognostic calculator for acute kidney injury in adult critically ill patients: development, validation and comparison to serum neutrophil gelatinase-associated lipocalin [J]. *Intensive Care Med*, 2017, 43 (6): 764-773.
- 3 Kristovic D, Horvatic I, Husedzinovic I, et al. Cardiac surgery-associated acute kidney injury: risk factors analysis and comparison of prediction models [J]. *Interact Cardiovasc Thorac Surg*, 2015, 21 (3): 366-373.
- 4 Basu RK, Zappitelli M, Brunner L, et al. Derivation and validation of the renal angina index to improve the prediction of acute kidney injury in critically ill children [J]. *Kidney Int*, 2014, 85 (3): 659-667.
- 5 Mohamadlou H, Lynn-Palevsky A, Barton C, et al. Prediction of Acute Kidney Injury With a Machine Learning Algorithm Using Electronic Health Record Data [J/OL]. <https://journals.sagepub.com/doi/full/10.1177/2054358118776326>.
- 6 Johnson AE, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database [J]. *Sci Data*, 2016, 3: 160035.
- 7 Acute Kidney Injury Work Group. KDIGO Clinical Practice Guideline for Acute Kidney Injury [J/OL]. *Kidney Int*, 2012, 2 (Suppl 1): 1-138.
- 8 Ke GL, Meng Q, Finley T, et al. LightGBM: A Highly Efficient Gradient Boosting Decision Tree [C]. Long Beach: 31st Conference on Neural Information Processing Systems, 2017.
- 9 Kellum JA, Sileanu FE, Murugan R, et al. Classifying AKI by Urine Output versus Serum Creatinine Level [J]. *J Am Soc Nephrol*, 2015, 26 (9): 2231-2238.
- 10 李航. 统计学习方法 [M]. 北京: 清华大学出版社, 2012.
- 11 Gurm HS, Seth M, Kooiman J, et al. A novel tool for reliable and accurate prediction of renal complications in patients undergoing percutaneous coronary intervention [J]. *J Am Coll Cardiol*, 2013, 61 (22): 2242-2248.
- 12 Lee J, de Louw E, Niemi M, et al. Association between fluid balance and survival in critically ill patients [J]. *J Intern Med*, 2015, 277 (4): 468-477.
- 13 Bresolin N, Silva C, Hallal A, et al. Prognosis for children with acute kidney injury in the intensive care unit [J]. *Pediatr Nephrol*, 2009, 24 (3): 537-544.
- 14 Lee SW, Yu MY, Lee H, et al. Risk Factors for Acute Kidney Injury and In-Hospital Mortality in Patients Receiving Extracorporeal Membrane Oxygenation [J]. *PLoS One*, 2015, 10 (10): e0140674.
- 15 Malhotra R, Kashani KB, Macedo E, et al. A risk prediction score for acute kidney injury in the intensive care unit [J]. *Nephrol Dial Transplant*, 2017, 32 (5): 814-822.
- 16 Lachance P, Villeneuve PM, Rewa OG, et al. Association between e-alert implementation for detection of acute kidney injury and outcomes: a systematic review [J]. *Nephrol Dial Transplant*, 2017, 32 (2): 265-272.
- 17 Pollard TJ, Johnson AEW, Raffa JD, et al. The eICU Collaborative Research Database, a freely available multi-center database for critical care research [J]. *Sci Data*, 2018, 5: 180178.
- 18 赵宇卓, 王俊梅, 潘菲, 等. 急救数据库建设初探 [J]. *中华危重病急救医学*, 2018, 30 (6): 609-612.