

# Predicting the risk factors of diabetic ketoacidosis-associated acute kidney injury: A machine learning approach using XGBoost

预测糖尿病酮症酸中毒相关急性肾损伤的危险因素：使用 XGBoost 的机器学习方法

1Department of Endocrinology, Second Affiliated Hospital of Jilin University, Changchun, China

1 中国长春吉林大学第二附属医院内分泌科

2Digestive Diseases Center, Department of Hepatopancreatobiliary Medicine, Second Affiliated Hospital of Jilin University, Changchun, China

2 中国长春吉林大学第二附属医院消化疾病中心，肝胆胰内科

Edited by: Debasis Samanta, Indian Institute of Technology Kharagpur, India

编辑：Debasis Samanta，印度 Kharagpur 印度理工学院

Reviewed by: Lakshmi Kannan, University of Pikeville Kentucky College of Osteopathic Medicine, United States;

Chao-Yu Guo, National Yang Ming Chiao Tung University, Taiwan

审核人：Lakshmi Kannan，皮克维尔肯塔基州大学骨病医学院，美国；Chao-Yu Guo，国立阳明交通大学，台湾

✉\*Correspondence: Chuan Zhang, [zhangchuan@jlu.edu.cn](mailto:zhangchuan@jlu.edu.cn)

- 通讯：Chuan Zhang, [zhangchuan@jlu.edu.cn](mailto:zhangchuan@jlu.edu.cn)

This article was submitted to Clinical Diabetes, a section of the journal Frontiers in Public Health

这篇文章被提交给临床糖尿病，公共卫生前沿杂志的一个部分

PMCID: PMC10117643 PMID: 37089510

## Abstract 摘要

### Objective 目的

The purpose of this study was to develop and validate a predictive model based on a machine learning (ML) approach to identify patients with DKA at increased risk of AKI within 1 week of hospitalization in the intensive care unit (ICU).

本研究的目的是开发和验证基于机器学习（ML）方法的预测模型，以识别在重症监护室（ICU）住院 1 周内阿基风险增加的 DKA 患者。

### Methods 方法

Patients diagnosed with DKA from the Medical Information Mart for Intensive Care IV (MIMIC-IV) database according to the International Classification of Diseases (ICD)-9/10 code were included. The patient's medical history is extracted, along with data on their demographics, vital signs, clinical characteristics, laboratory results, and therapeutic measures. The best-performing model is chosen by contrasting the 8 ML models. The area under the receiver operating characteristic curve (AUC), sensitivity, accuracy, and specificity were calculated to select the best-performing ML model.

纳入了根据国际疾病分类（ICD）-9/10 代码从重症监护医学信息市场 IV（MIMIC-IV）数据库中诊断为 DKA 的患者。提取患者的病史，沿着关于其人口统计学、生命体征、临床特征、实验室结果和治疗措施的数据。通过对比 8 个 ML 模型来选择性能最佳的模型。计算受试者工作特征曲线下面积（AUC）、灵敏度、准确度和特异性，以选择性能最佳的 ML 模型。

## Results 结果

The final study enrolled 1,322 patients with DKA in total, randomly split into training (1,124, 85%) and validation sets (198, 15%). 497 (37.5%) of them experienced AKI within a week of being admitted to the ICU. The eXtreme Gradient Boosting (XGBoost) model performed best of the 8 ML models, and the AUC of the training and validation sets were 0.835 and 0.800, respectively. According to the result of feature importance, the top 5 main features contributing to the XGBoost model were blood urea nitrogen (BUN), urine output, weight, age, and platelet count (PLT).

最终研究共招募了 1,322 名 DKA 患者，随机分为训练集（1,124, 85%）和验证集（198, 15%）。其中 497 人（37.5%）在入住 ICU 后一周内发生阿基。极限梯度增强（XGBoost）模型在 8 个 ML 模型中表现最好，并且训练集和验证集的 AUC 分别为 0.835 和 0.800。根据特征重要性的结果，对 XGBoost 模型贡献最大的 5 个主要特征是血尿素氮（BUN）、尿量、体重、年龄和血小板计数（PLT）。

## Conclusion 结论

An ML-based individual prediction model for DKA-associated AKI (DKA-AKI) was developed and validated. The model performs robustly, identifies high-risk patients early, can assist in clinical decision-making, and can improve the prognosis of DKA patients to some extent.

开发并验证了基于 ML 的 DKA 相关阿基（DKA-AKI）个体预测模型。该模型具有较强的鲁棒性，能够早期识别高危患者，辅助临床决策，并在一定程度上改善 DKA 患者的预后。

## Keywords:

diabetic ketosis, acute kidney injury, machine learning, XGBoost, outcome

关键词：糖尿病酮症，急性肾损伤，机器学习，XGBoost，结局

## 1. Introduction 1.介绍

Diabetic ketoacidosis (DKA) is a serious acute complication of diabetes mellitus that can be fatal if left untreated. DKA is characterized by uncontrolled blood glucose (BG) levels, acidosis, and ketosis. It can also cause imbalances in electrolytes and fluids, leading to complications such as cerebral edema, acute kidney injury (AKI), and even renal failure in severe cases (1, 2). AKI is a common complication, affecting 40 to 50 percent of DKA patients. Unfortunately, AKI can result in increased morbidity and mortality, prolonged intensive care unit (ICU) stays, a greater susceptibility to chronic kidney disease (CKD), and recurrent AKI episodes during ICU treatment. Therefore, it is critical to closely monitor DKA patients for signs of AKI and promptly intervene to mitigate its potential negative impact (3, 4). The diagnosis of AKI is typically based on the dynamic changes in serum creatinine (SCr) and urine output, following the clinical practice recommendations established by the Kidney Disease Improving Global Outcomes (KDIGO) organization (5). However, renal damage usually precedes the elevation of SCr levels, so that renal damage has already begun by the time AKI is diagnosed (6). DKA-associated AKI (DKA-AKI) usually occurs after hypoperfusion of the kidney due to hypovolemia. Renal function can be improved to some extent by effective prevention and treatment, such as applying vasoactive drugs, ensuring adequate renal perfusion, and avoiding nephrotoxic drugs (7). Consequently, it is necessary to explore predictors for AKI and monitor the population at risk for DKA-AKI. Clinicians can improve the prognosis of DKA patients with timely management.

糖尿病酮症酸中毒（DKA）是糖尿病的一种严重急性并发症，如果不及时治疗可能致命。DKA 的特征是血糖（BG）水平不受控制、酸中毒和酮症。它还可能导致电解质和液体失衡，导致并发症，如脑水肿，急性肾损伤（阿基），甚至在严重病例中出现肾衰竭（1，2）。阿基是一种常见的并发症，影响 40%至 50%的 DKA 患者。不幸的是，阿基可导致发病率和死亡率增加、重症监护室（ICU）住院时间延长、对慢性肾脏疾病（CKD）的易感性增加以及 ICU 治疗期间阿基复发。因此，密切监测 DKA 患者的阿基体征并及时干预以减轻其潜在的负面影响至关重要（3，4）。阿基的诊断通常基于血清肌酐（SCr）和尿量的动态变化，遵循肾脏疾病改善全球结局（KDIGO）组织（5）建立的临床实践建议。然而，肾损害通常先于 SCr 水平升高，因此在诊断阿基时肾损害已经开始（6）。DKA 相关阿基（DKA-AKI）通常发生在由于血容量不足导致的肾脏灌注不足之后。通过有效的预防和治疗，如应用血管活性药物，保证足够的肾脏灌注，避免肾毒性药物，可以在一定程度上改善肾功能（7）。因此，有必要探索阿基的预测因子并监测 DKA-AKI 风险人群。临床医生可以通过及时处理改善 DKA 患者的预后。

Several studies have demonstrated that various factors, including age, type of diabetes, comorbidities, respiratory rate (RR), blood pressure, baseline Scr, blood urea nitrogen (BUN), and urine output, can be used to predict the risk of DKA-AKI (3, 4, 8). In our previous work (9), a model for predicting DKA-AKI risk based on logistic regression was developed and a nomogram was drawn. Besides, numerous models for predicting AKI have been developed (10–13). However, few publications have identified the specific risk of AKI in DKA patients.

多项研究表明，各种因素，包括年龄、糖尿病类型、合并症、呼吸频率（RR）、血压、基线 Scr、血尿素氮（BUN）和尿量，可用于预测 DKA-AKI 的风险（3、4、8）。在我们之前的工作（9）中，开发了基于逻辑回归的 DKA-AKI 风险预测模型，并绘制了诺模图。此外，已经开发了许多用于预测阿基的模型（10 - 13）。然而，很少有出版物确定了 DKA 患者发生阿基的特定风险。

In recent years, machine learning (ML) algorithms have been found to have excellent predictive performance (14). ML has also demonstrated excellent performance in the administration of ICU patients, and combining it with electronic health record systems can increase the reliability of technological support for critical care.

近年来，机器学习（ML）算法被发现具有出色的预测性能（14）。ML 在 ICU 患者管理方面也表现出出色的性能，将其与电子健康记录系统相结合可以提高重症监护技术支持的可靠性。

In the administration of ICU patients, ML has demonstrated excellent performance and, when combined with electronic health record systems, can increase the reliability of technological support for critical care (15). Unfortunately, to our knowledge, there are no relevant studies that apply machine learning algorithms to build a model and identify risk factors for DKA-AKI. Therefore, in this research, we developed a model for predicting DKA-AKI in real-time using ML algorithms and validated its performance. This is an important step toward identifying and managing DKA-AKI, as it could help clinicians intervene early and prevent further complications.

在 ICU 患者的管理中，ML 表现出出色的性能，当与电子健康记录系统结合时，可以提高重症监护技术支持的可靠性（15）。不幸的是，据我们所知，目前还没有相关的研究应用机器学习算法来构建模型并识别 DKA-AKI 的风险因素。因此，在本研究中，我们开发了一种使用 ML 算法实时预测 DKA-AKI 的模型，并验证了其性能。这是识别和管理 DKA-AKI 的重要一步，因为它可以帮助临床医生早期干预并预防进一步的并发症。

## 2. Methods 2.方法

### 2.1. Database 2.1.数据库

Patients’ information for this retrospective investigation was obtained from the database called Marketplace for Medical Information in Intensive Care IV (MIMIC-IV). It is consisted of the medical history for the ill critical patients at Beth Israel Deaconess Medical Center (Boston, MA), including demographic information, disease diagnosis, vital signs, laboratory tests, treatment information, survival status, and other comprehensive clinical records. The inclusion of the MIMIC-IV database has been expanded from 2008–2012 to 2008–2019, compared to the previous version of the MIMIC-III (Certificate number: 9168028).

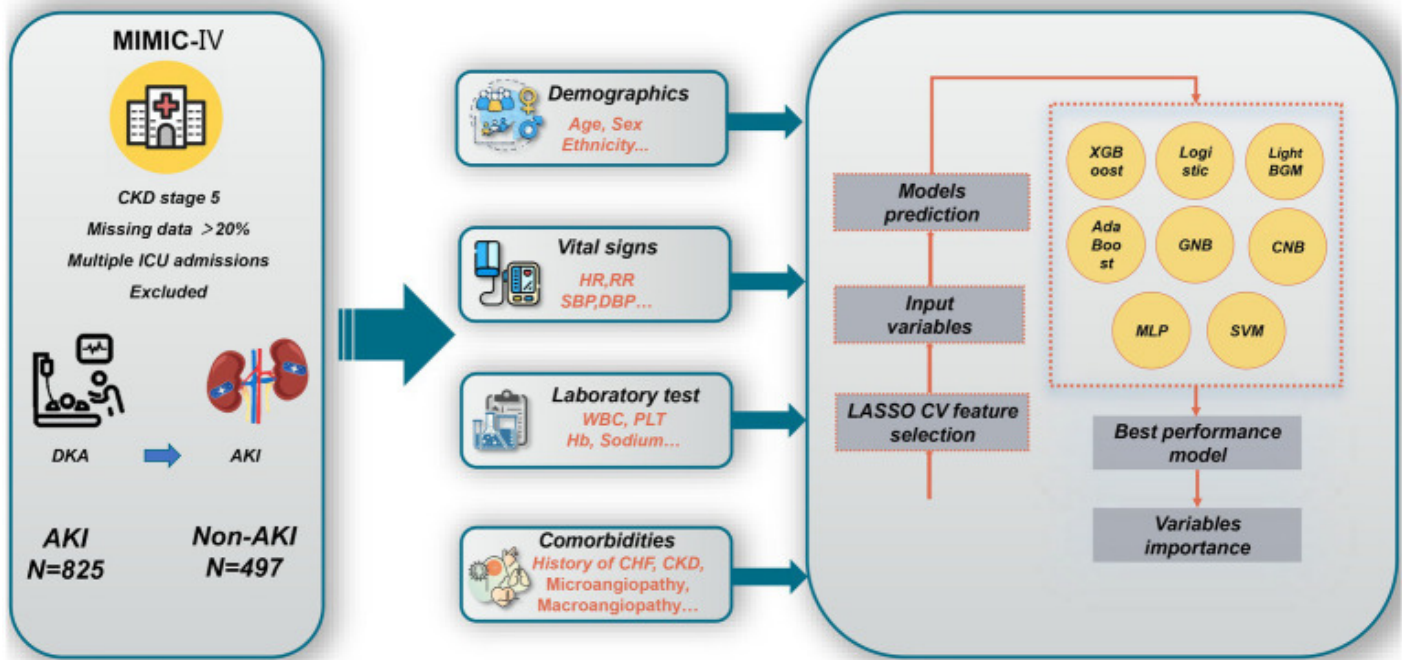
本回顾性研究的患者信息来自重症监护医学信息市场 IV（MIMIC-IV）数据库。它由 Beth Israel Deaconess Medical Center（Boston，MA）的危重症患者的病史组成，包括人口统计学信息、疾病诊断、生命体征、实验室检查、治疗信息、生存状态和其他综合临床记录。与 MIMIC-III（证书编号：9168028）之前的版本相比，MIMIC-IV 数据库的纳入范围已从 2008-2012 年扩大到 2008-2019 年。

### 2.2. Study population 2.2.研究人群

DKA patients in the MIMIC-IV databases identified with the International Classification of Diseases (ICD) -9/10 code were included. The excluded criteria were: (1) patients with a diagnosis of CKD stage 5; (2) if a patient was admitted to the ICU repeatedly during one hospitalization, only the first hospitalization information was retained; and (3) patients with more than 20% missing data. The overall flowchart is shown in Figure 1.

纳入 MIMIC-IV 数据库中使用国际疾病分类（ICD）-9/10 代码识别的 DKA 患者。排除标准为：（1）诊断为 CKD 5 期的患者；（2）如果患者在一次住院期间多次入住 ICU，则仅保留首次住院信息；（3）数据缺失超过 20% 的患者。总体流程图如 Figure 1 所示。

Figure 1.



Overall flowchart of this study. MIMIC-IV, Medical Information Mart for Intensive Care IV; CKD, chronic kidney disease; ICU, intensive care unit; DKA, v; AKI, acute kidney injury; HR, heart rate; RR, respiratory rate; WBC, white blood cell count; PLT, Platelet count; Hb, hemoglobin; CHF, congestive heart failure; XGBoost, eXtreme Gradient Boosting; GNB, Gaussian Naïve Bayes; CNB, Complement Naive Bayes; MLP, multi-layer perceptron neural network; SVM, support vector machine; LASSO CV, least absolute shrinkage and selection operator cross-validation.

本研究的总体流程图。MIMIC-IV，重症监护医学信息市场 IV；CKD，慢性肾脏疾病；ICU，重症监护室；DKA，v；阿基，急性肾损伤；HR，心率；RR，呼吸频率；WBC，白色血细胞计数；PLT，血小板计数；Hb，血红蛋白；CHF，充血性心力衰竭；XGBoost，极限梯度增强；GNB，高斯朴素贝叶斯；CNB，补体朴素贝叶斯；MLP，多层感知器神经网络；SVM，支持向量机；LASSO CV，最小绝对收缩和选择算子交叉验证。

## 2.3. Data extraction and pre-processing 2.3.数据提取和预处理

The following patient data were retrieved from the MIMIC-IV database: demographics, vital signs, comorbidities, laboratory tests, interventions, prognosis, and scoring systems, as shown in Supplementary Table S1. Features missing greater than 20% were excluded in the follow-up study, and other variables were duplicated using nearest neighbor imputation algorithms (excluded missing variables were shown in Supplementary Table S2). All items were recorded within 24 h of admission to the ICU. And due to repeat sampling, we only retained the results of the first test.

从 MIMIC-IV 数据库中检索以下患者数据：人口统计学、生命体征、合并症、实验室检查、干预、预后和评分系统，如 Supplementary Table S1 所示。在随访研究中排除缺失大于 20% 的特征，并使用最近邻插补算法重复其他变量（排除的缺失变量见 Supplementary Table S2）。所有项目均在入住 ICU 后 24 小时内记录。由于重复采样，我们只保留了第一次测试的结果。

## 2.4. Outcome 2.4.结果

Our study's primary outcome was the occurrence of AKI of DKA patients receiving ICU care within a week. Utilizing KDIGO criteria, a diagnosis of AKI was determined (5).

我们研究的主要结果是接受 ICU 护理的 DKA 患者在一周内发生阿基。利用 KDIGO 标准，确定阿基的诊断（5）。

## 2.5. Model development and validation 2.5.模型开发和验证

Supervised ML algorithms have been playing an important role in various clinical prediction models. In the present study, supervised ML algorithms were used to construct later predictions of the risk of AKI in DKA patients. The least absolute shrinkage and selection operator cross-validation (LASSO CV) method was used for model selection in order to reduce model complexity and the risk of overfitting and to optimize model training speed. Machine learning has contributed massively applications in medical diagnosis, treatment, and prediction (16). In our present study, the overall dataset was divided into 2 groups randomly, 85% in training cohort and 15% in validation cohort, respectively. To minimize overfitting and identify the optimization hyperparameters, 10-fold cross-validation (CV) was also performed. And then, 8 ML model algorithms, including eXtreme Gradient Boosting (XGBoost); logistic regression, Light BGM, Ada Boost, Gaussian Naïve Bayes (GNB), multi-layer perceptron neural network (MLP), Complement Naive Bayes (CNB), support vector machine (SVM). Before establishing the models, the Lasso CV features selection method was applied to better select feature parameters. Furthermore, we reported several parameters related to model performance in this study, including area under the receiver operating characteristic curve (AUC), sensitivity, specificity, and accuracy. The final model was determined according to the highest AUC value after comparing the performance of the 8 different ML models. Finally, feature importance was calculated for the final models to evaluate the contribution of the candidate predictors. The programming written using Python (package version 3.8) was used to exert data analysis.

监督 ML 算法在各种临床预测模型中发挥着重要作用。在本研究中，监督 ML 算法用于构建 DKA 患者阿基风险的后期预测。采用最小绝对收缩和选择算子交叉验证（LASSO CV）方法进行模型选择，以降低模型复杂度和过拟合风险，优化模型训练速度。机器学习在医疗诊断、治疗和预测方面有着广泛的应用（16）。在本研究中，将整个数据集随机分为 2 组，分别为训练队列的 85% 和验证队列的 15%。为了最小化过拟合并识别优化超参数，还进行了 10 倍交叉验证（CV）。然后，8 种 ML 模型算法，包括极限梯度提升（XGBoost）、逻辑回归、Light BGM、Ada Boost、高斯朴素贝叶斯（GNB）、多层感知器神经网络（MLP）、互补朴素贝叶斯（CNB）、支持向量机（SVM）。在建立模型之前，应用 Lasso CV 特征选择方法来更好地选择特征参数。此外，我们报告了本研究中与模型性能相关的几个参数，包括受试者工作特征曲线下面积（AUC）、灵敏度、特异性和准确性。在比较 8 种不同 ML 模型的性能后，根据最高 AUC 值确定最终模型。最后，计算最终模型的特征重要性，以评估候选预测因子的贡献。使用 Python（软件包版本 3.8）编写的程序用于进行数据分析。

## 2.6. Statistical analysis 2.6.统计分析

Variables that followed a normal distribution were expressed as the mean  $\pm$  standard deviation (SD); if they did not follow a normal distribution the quartiles were generally used. Student's t-test or Mann-Whitney U-test was used to analyze continuous variables. In the case of categorical variables, the  $\chi^2$  test or Fisher's exact test was generally used. All Statistical analyses were performed using R version 3.6.3 and python version 3.7. The core

code for model construction was provided by the Extreme Smart Analysis platform<sup>1</sup> and uploaded to the supplemental material.

遵循正态分布的变量表示为平均值±标准差（SD）；如果不遵循正态分布，则通常使用四分位数。使用 Student t 检验或 Mann-Whitney U 检验分析连续变量。对于分类变量，通常使用χ<sup>2</sup> 检验或 Fisher 精确检验。所有统计分析均使用 R 版本 3.6.3 和 python 版本 3.7 进行。模型构建的核心代码由极智分析平台 1 提供，并上传到补充材料中。

### 3. Results 3.结果

#### 3.1. Patients’ characteristics 3.1.患者特征

1,322 patients diagnosed with DKA in total were eventually enrolled in our study. Among the total cohort, 689 (52.1%) were female, and the median age was 50 [IQR 35, 62] years. Within a week of admission to the ICU, 497 (37.6%) progressed to DKA-AKI. The incidence of DKA-AKI was 36.4% (181/497) in stage 1, 34.4% (171/497) in stage 2, and 29.2% (145/497) in stage 3. CRRT was applied to 17.4% (73/497) of AKI patients. The AKI group had higher ventilator utilization, hospital length of stay (HLOS), and in-hospital mortality. Except for gender, systolic blood pressure (SBP), liver disease, history of hypertension, platelet count (PLT), calcium, BG, and infusion volume, the statistical analysis of the candidate predictors revealed significant differences between the AKI and non-AKI groups (Table 1).

最终共有 1, 322 例诊断为 DKA 的患者入组我们的研究。在整个队列中，689 例（52.1%）为女性，中位年龄为 50 岁[IQR 35, 62]。在入住 ICU 的一周内，497 例（37.6%）进展为 DKA-AKI。DKA-AKI 的发生率在第 1 阶段为 36.4%（181/497），在第 2 阶段为 34.4%（171/497），在第 3 阶段为 29.2%（145/497）。

17.4%（73/497）的阿基患者接受了 CRRT 治疗。阿基组的呼吸机使用率、住院时间（HLOS）和住院死亡率较高。除性别、收缩压（SBP）、肝脏疾病、高血压病史、血小板计数（PLT）、钙、BG 和输注量外，候选预测因子的统计分析显示阿基组和非 AKI 组之间存在显著差异（Table 1）。

Table 1.  
Characteristic at baseline between AKI and non-AKI group.

Variable	变量	Total (n = 1,322)	Non-AKI (n = 825)	AKI (n = 497)	p value
Age, years	年龄，岁	50 [35, 62]	43 [30, 57]	58 [46, 68]	<0.001
Gender (Female)	性别（女性）	689 (52.1)	425 (51.5)	264 (53.1)	0.572
Weight, Kg	体重，千克	73.2 [62.0, 87.1]	70.0 [60.3, 83.6]	78.5 [66.0, 94.0]	<0.001
Ethnicity	种族				0.028
White	白人	742 (56.1)	474 (57.4)	268 (53.9)	

Variable	变量	Total (n = 1,322)	Non-AKI (n = 825)	AKI (n = 497)	p value
African-American	非洲裔美国人	364 (27.5)	223 (27.0)	141 (28.3)	
Hispanic-American	西班牙裔美国人	76 (5.7)	51 (6.2)	25 (5.0)	
Asian	亚洲人	28 (2.1)	21 (2.5)	7 (1.4)	
Other	其他	112 (8.5)	56 (6.8)	56 (11.3)	
DM type	糖尿病类型				<0.001
T1DM	1型糖尿病	821 (62.1)	546 (66.2)	275 (55.3)	
T2DM	2型糖尿病	373 (28.2)	209 (25.3)	164 (33.1)	
Other	其他	128 (9.7)	70 (8.5)	58 (11.7)	
HR, beats/min	心率, 次/分钟	100 [88, 111.000]	101 [89, 113]	98 [85, 109]	<0.001
RR, breaths/min	呼吸频率, 次/分钟	19 [16, 23]	19 [16, 23]	20 [17, 24]	0.002
SBP, mmHg	收缩压, 毫米汞柱	129 [114, 145]	128 [115, 144]	130 [111, 147]	0.907
DBP, mmHg	舒张压, 毫米汞柱	71 [60, 83]	72 [62, 83]	69 [56, 82]	0.002
Microangiopathy (Yes)	微血管病变 (是)	691 (52.3)	411 (49.8)	280 (56.3)	0.022
Macroangiopathy (Yes)	大血管病变 (是)	319 (24.1)	260 (31.5)	59 (11.9)	<0.001
Preexisting CKD	既往慢性肾脏病				<0.001
Non-CKD	无慢性肾脏病	1,006 (76.1)	704 (85.3)	302 (60.8)	
Stage1-3	1-3期	251 (18.9)	115 (13.9)	136 (27.4)	
Stage3-4	3-4期	65 (4.9)	6 (0.7)	59 (11.9)	
UTI (Yes)	尿路感染 (是)	147 (11.120)	71 (8.606)	76 (15.292)	<0.001
Pneumonia (Yes)	肺炎 (是)	58 (4.387)	18 (2.182)	40 (8.048)	<0.001
Liver disease (Yes)	肝病 (是)	124 (9.4)	73 (8.8)	51 (10.3)	0.393



Variable	变量	Total (n = 1,322)	Non-AKI (n = 825)	AKI (n = 497)	p value
History of hypertension (Yes)	高血压病史 (是)	513 (38.8)	305 (37.0)	208 (41.9)	0.078
History of CHF (Yes)	心力衰竭病史 (是)	208 (15.7)	69 (8.4)	139 (27.9)	<0.001
History of AMI (Yes)	急性心肌梗死病史 (是)	213 (16.1)	83 (10.1)	130 (26.2)	<0.001
History of ACI (Yes)	急性脑梗死病史 (是)	91 (6.9)	28 (3.4)	63 (12.7)	<0.001
Malignant Cancer (Yes)	恶性肿瘤 (是)	59 (4.463)	33 (4.000)	26 (5.231)	0.294
Bicarbonate, mEq/L	碳酸氢根, 毫当量/升	21.0 [18.0, 25.0]	21.0 [17.0, 25.0]	22.0 [18.0, 25.0]	0.005
WBC, K/ $\mu$ L	白细胞计数, 千/微升	8.2 [5.9, 11.9]	7.8 [5.8, 11.0]	9.0 [6.4, 13.5]	<0.001
PLT, K/ $\mu$ L	血小板计数, 千/微升	228.0 [175.0, 289.0]	227.0 [177.0, 283.0]	229.0 [167.0, 298.0]	0.711
Hb, g/dl	血红蛋白, 克/分升	10.9 [9.3, 12.3]	11.4 [10.1, 12.6]	9.8 [8.5, 11.3]	<0.001
Sodium, mEq/L	钠, 毫当量/升	138.0 [135.0, 140.0]	137.0 [135.0, 140.0]	138.0 [135.0, 141.0]	0.006
Chloride, mEq/L	氯, 毫当量/升	104.0 [100.0, 108.0]	104.0 [101.0, 108.0]	103.000 [99.0, 109.0]	0.004
Calcium, mEq/L	钙, 毫当量/升	8.400 [8.0, 8.8]	8.400 [8.0, 8.8]	8.400 [7.9, 8.8]	0.339
Phosphate, mEq/L	磷, 毫当量/升	2.7 [1.900, 3.700]	2.400 [1.7, 3.2]	3.300 [2.3, 4.5]	<0.001
AG	阴离子间隙	14.0 [12.0, 17.0]	14.0 [12.0, 17.0]	15.000 [12.0, 18.0]	0.015

Variable	变量	Total (n = 1,322)	Non-AKI (n = 825)	AKI (n = 497)	p value
Total osmotic pressure	总渗透压	494.600 [440.0, 560.2]	492.0 [437.6, 555.2]	498.400 [447.4, 570.6]	0.024
BUN, mg/dl	血尿素氮, 毫克/分升	16.0 [9.0, 31.0]	13.0 [8.0, 20.0]	28.0 [14.0, 46.0]	<0.001
Scr, mg/dl	血清肌酐, 毫克/分升	0.9 [0.7, 1.5]	0.8 [0.6, 1.1]	1.4 [0.9, 2.5]	<0.001
Potassium, mEq/L	钾, 毫当量/升	4.1 [3.7, 4.5]	4.0 [3.7, 4.4]	4.2 [3.8, 4.7]	<0.001
Blood glucose, mg/dl	血糖, 毫克/分升	188.0 [135.0, 257.000]	191.0 [139.0, 258.0]	183.000 [130.0, 251.0]	0.174
Infusion volume, mL	输液量, 毫升	12020.0 [7700.0, 17320.000]	11890.0 [7700.0, 16600.0]	12400.000 [7805.0, 18010.0]	0.061
Urine output, mL	尿量, 毫升	1950.0 [1085.0, 3000.0]	2300.0 [1470.0, 3440.0]	1300.000 [600.0, 2250.0]	<0.001
eGFR	估算肾小球滤过率	0.992 [0.854, 1.093]	1.000 [0.882, 1.114]	0.930 [0.829, 1.067]	<0.001
Use of NaHCO3 (Yes)	使用碳酸氢钠 (是)	123 (9.304)	41 (4.970)	82 (16.499)	<0.001
Mechanical ventilation (Yes)	机械通气 (是)	152 (11.498)	31 (3.758)	121 (24.346)	<0.001
GCS	格拉斯哥昏迷评分	15.000 [15.000, 15.000]	15.000 [15.000, 15.000]	15.000 [15.000, 15.000]	<0.001
CRRT (Yes)	连续肾脏替代治疗 (是)	77 (5.825)	4 (0.485)	73 (14.688)	<0.001
OASIS	牛津急性疾病严重程度评分	25.000 [21.000, 31.000]	23.000 [20.000, 27.000]	30.000 [25.000, 39.000]	<0.001

Variable	变量	Total (n = 1,322)	Non-AKI (n = 825)	AKI (n = 497)	p value
SOFA	顺序器官衰竭评分	3.000 [1.000, 5.000]	2.000 [1.000, 3.000]	5.000 [3.000, 8.000]	<0.001
SAPS-II	简化急性生理评分II	26.000 [19.000, 36.000]	22.000 [16.000, 29.000]	36.000 [26.000, 45.000]	<0.001
HLOS, days	住院时间, 天	4.838 [2.969, 8.487]	3.890 [2.647, 6.051]	7.654 [4.158, 13.670]	<0.001
Hospital mortality (Yes)	医院死亡 (是)	61 (4.614)	11 (1.333)	50 (10.060)	<0.001

AKI, acute kidney injury; DM, diabetic mellitus; T1DM, type 1 DM; T2DM, type 2 DM; HR, heart rate; RR, respiratory rate; SBP, systolic blood pressure, DBP, diastolic blood pressure, CKD, chronic kidney diseases, UTI, urinary tract infection; CHF, congestive heart failure; AMI, acute myocardial infarction; ACI, acute cerebral infarction; WBC, white blood cell; PLT, platelets count; Hb, hemoglobin; AG, anion gap; BUN, blood urea nitrogen; Scr, serum creatinine; eGFR, estimated glomerular filtration rate; GCS, Glasgow coma scale; OASIS, oxford acute severity of illness score; SOFA, sequential organ failure assessment; SAPS-II, simplified acute physiology score II; HLOS, hospital length of stay.

AKI, 急性肾损伤; DM, 糖尿病; T1DM, 1 型 DM; T2DM, 2 型 DM; HR, 心率; RR, 呼吸频率; SBP, 收缩压; DBP, 舒张压; CKD, 慢性肾脏疾病; UTI, 尿路感染; CHF, 充血性心力衰竭; AMI, 急性心肌梗死; ACI, 急性脑梗死; WBC, 白色细胞; PLT, 血小板计数; Hb, 血红蛋白; AG, 阴离子间隙; BUN, 血尿素氮; Scr, 血清肌酐; eGFR, 估计肾小球滤过率; GCS, 格拉斯哥昏迷量表; OASIS, 牛津急性疾病严重程度评分; SOFA, 序贯器官衰竭评估; SAPS-II, 简化急性生理学评分 II; HLOS, 住院时间。

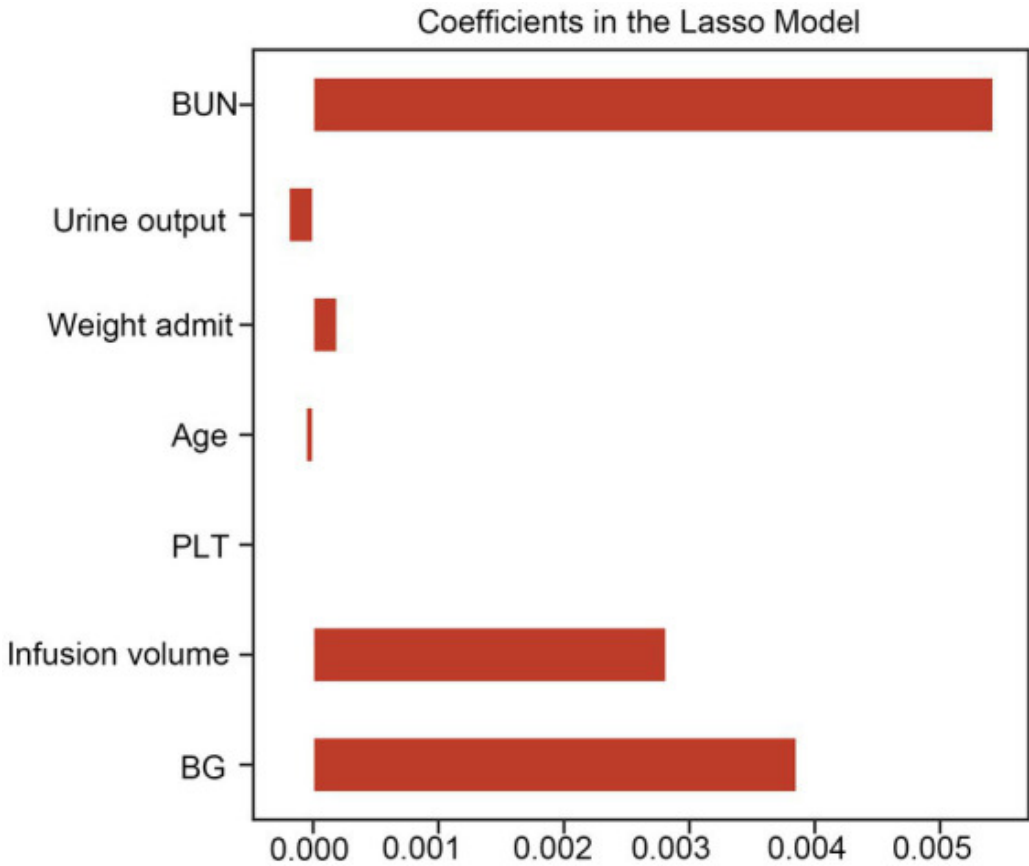
### 3.2. Predictive model performance 3.2.预测模型性能

The results of LASSO CV screened seven variables as input variables for the model (Figure 2). Eight MI-based models have been developed and validated, and the XGBoost model performed the strongest in predicting DKA-AKI with the highest AUC value (0.835) (Figure 3A). In addition, the XGBoost model had a sensitivity of 0.716, a specificity of 0.823, and an accuracy of 0.782 in the training set (Table 2). When internal validation was performed, the XGBoost model still showed optimal performance with an AUC of 0.800 (Figure 3B), and its sensitivity, specificity, and accuracy were 0.718, 0.790, and 0.749, respectively (Table 3). Therefore, the XGBoost model was selected as the final mode. Furthermore, decision curve analysis (DCA) (Figure 4A) and calibration plots (Figure 4B) were also performed to further demonstrate the performance of the XGBoost model.

LASSO CV 的结果筛选了 7 个变量作为模型的输入变量（Figure 2）。已经开发并验证了八个基于 MI 的模型，并且 XGBoost 模型在预测 DKA-AKI 方面表现最强，具有最高 AUC 值（0.835）（Figure 3A）。此外，

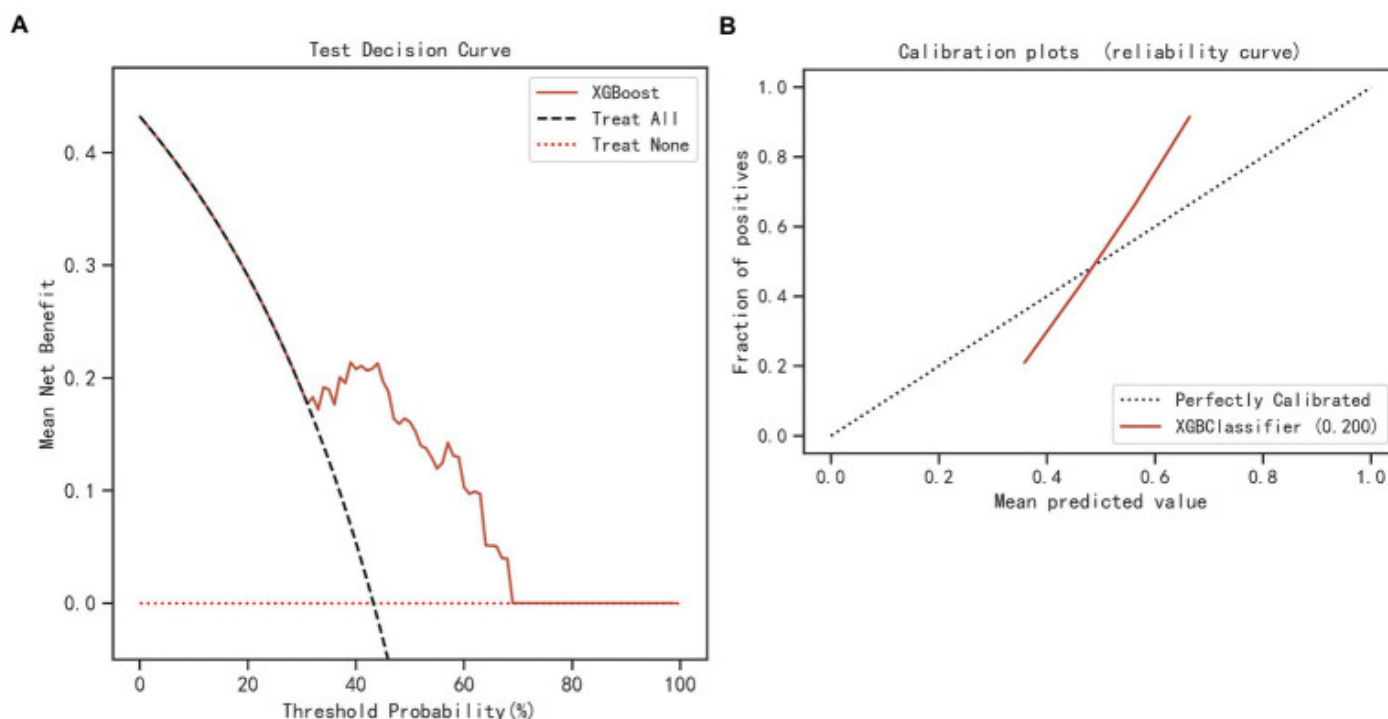
XGBoost 模型在训练集中的灵敏度为 0.716，特异性为 0.823，准确度为 0.782（Table 2）。当进行内部验证时，XGBoost 模型仍显示出最佳性能，AUC 为 0.800（Figure 3B），其灵敏度、特异性和准确性分别为 0.718、0.790 和 0.749（Table 3）。因此，选择 XGBoost 模型作为最终模式。此外，还进行了决策曲线分析（DCA）（Figure 4A）和校准图（Figure 4B），以进一步证明 XGBoost 模型的性能。

Figure 2.



Lasso CV method was used to conduct feature selection. LASSO CV, least absolute shrinkage and selection operator cross-validation, BUN, blood urea nitrogen; BG, blood glucose; PLT, platelet count. 采用 Lasso CV 方法进行特征选择。LASSO CV，最小绝对收缩和选择操作员交叉验证，BUN，血尿素氮；BG，血糖；PLT，血小板计数。

Figure 3.



Comparing the different ML models' AUC in the training (A) and validation (B) sets. ML, machine learning; AUC, area under the receiver operating characteristic curve; XGBoost, eXtreme Gradient Boosting; GNB, Gaussian Naïve Bayes; CNB, Complement Naive Bayes; MLP, multi-layer perceptron neural network; SVM, support vector machine.

比较训练 (A) 和验证 (B) 集中不同 ML 模型的 AUC。ML, 机器学习; AUC, 受试者工作特征曲线下面积; XGBoost, 极限梯度提升; GNB, 高斯朴素贝叶斯; CNB, 互补朴素贝叶斯; MLP, 多层感知器神经网络; SVM, 支持向量机。

Table 2.

Model parameters in training set.

Model	AUC	Cutoff	Accuracy	Sensitivity	Specificity	PPV	NPV	F1-Score
XG Boost	0.835 (0.004)	0.463 (0.013)	0.782 (0.007)	0.716 (0.025)	0.823 (0.025)	0.714 (0.022)	0.824 (0.009)	0.714 (0.007)
Logistic	0.773 (0.004)	0.468 (0.033)	0.744 (0.012)	0.627 (0.040)	0.817 (0.042)	0.679 (0.036)	0.782 (0.012)	0.649 (0.007)
Light GBM	0.547 (0.079)	0.600 (0.800)	0.634 (0.031)	0.343 (0.322)	0.777 (0.292)	NA	0.666 (0.052)	NA
AdaBoost	0.820 (0.005)	0.467 (0.002)	0.743 (0.008)	0.770 (0.013)	0.729 (0.019)	0.634 (0.016)	0.837 (0.007)	0.695 (0.006)
GNB	0.795 (0.005)	0.351 (0.032)	0.742 (0.010)	0.720 (0.031)	0.757 (0.031)	0.642 (0.020)	0.817 (0.011)	0.678 (0.010)

Model	AUC	Cutoff	Accuracy	Sensitivity	Specificity	PPV	NPV	F1-Score
CNB	0.678 (0.005)	0.993 (0.022)	0.664 (0.013)	0.660 (0.010)	0.674 (0.007)	NA	0.754 (0.042)	NA
MLP	0.607 (0.074)	0.397 (0.029)	0.612 (0.097)	0.614 (0.282)	0.588 (0.330)	NA	0.710 (0.063)	NA
SVM	0.518 (0.061)	0.449 (0.057)	0.573 (0.101)	0.409 (0.322)	0.673 (0.347)	0.491 (0.084)	0.667 (0.037)	0.363 (0.146)

Data are shown as means  $\pm$  standard deviations (SD).

数据显示为平均值 $\pm$ 标准差（SD）。

AUC, area under the receiver operating characteristic curve; PPV, positive prediction value; NPV, negative prediction value; GNB, Gaussian Naïve Bayes; CNB, Complement Naive Bayes; GNB, Gaussian Naïve Bayes; MLP, multi-layer perceptron neural network; SVM, support vector machine; NA, not applicable.  
AUC，受试者工作特征曲线下面积; PPV，阳性预测值; NPV，阴性预测值; GNB，高斯朴素贝叶斯; CNB，互补朴素贝叶斯; GNB，高斯朴素贝叶斯; MLP，多层感知器神经网络; SVM，支持向量机; NA，不适用。

Table 3.  
Model parameters in validation set.

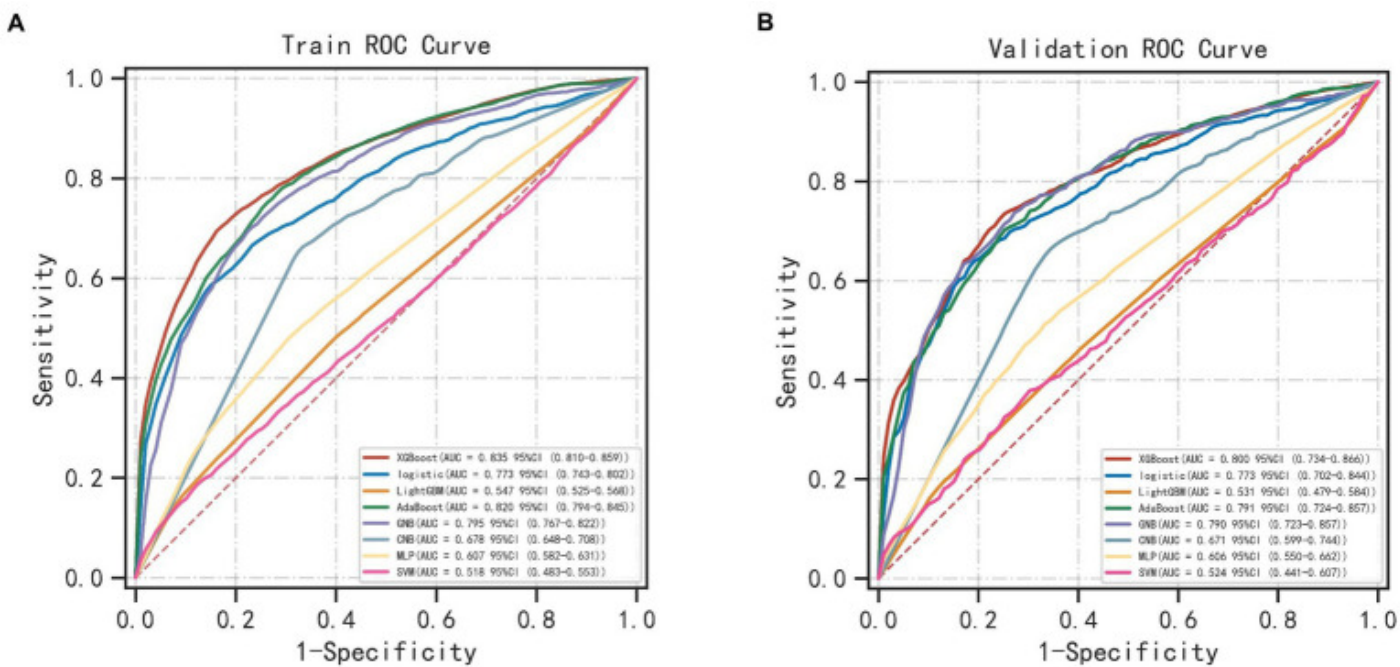
Model	AUC	Cutoff	Accuracy	Sensitivity	Specificity	PPV	NPV (SD)	F1-Score
XGBoost	0.800 (0.019)	0.463 (0.013)	0.749 (0.017)	0.718 (0.060)	0.790 (0.036)	0.643 (0.035)	0.812 (0.026)	0.677 (0.035)
Logistic	0.773 (0.025)	0.468 (0.033)	0.738 (0.028)	0.650 (0.065)	0.821 (0.064)	0.649 (0.051)	0.788 (0.045)	0.648 (0.050)
Light GBM	0.531 (0.081)	0.600 (0.800)	0.626 (0.042)	0.350 (0.308)	0.751 (0.292)	NA	0.651 (0.038)	NA
AdaBoost	0.791 (0.028)	0.467 (0.002)	0.714 (0.031)	0.695 (0.065)	0.782 (0.053)	0.584 (0.056)	0.820 (0.036)	0.633 (0.052)
GNB	0.790 (0.031)	0.351 (0.032)	0.738 (0.037)	0.694 (0.065)	0.807 (0.073)	0.639 (0.066)	0.809 (0.045)	0.664 (0.057)
CNB	0.671 (0.029)	0.993 (0.022)	0.653 (0.031)	0.677 (0.042)	0.661 (0.028)	NA	0.739 (0.066)	NA
MLP	0.606 (0.077)	0.397 (0.029)	0.617 (0.107)	0.548 (0.305)	0.664 (0.289)	NA	NA	NA

Model	AUC	Cutoff	Accuracy	Sensitivity	Specificity	PPV	NPV (SD)	F1-Score
SVM	0.524 (0.093)	0.449 (0.057)	0.569 (0.100)	0.554 (0.292)	0.576 (0.316)	0.467 (0.126)	0.717 (0.146)	0.456 (0.155)

Data are shown as means  $\pm$  standard deviations (SD).  
数据显示为平均值 $\pm$ 标准差（SD）。

AUC, area under the receiver operating characteristic curve; PPV, positive prediction value; NPV, negative prediction value; GNB, Gaussian Naïve Bayes; CNB, Complement Naive Bayes; GNB, Gaussian Naïve Bayes; MLP, multi-layer perceptron neural network; SVM, support vector machine; NA, not applicable.  
AUC, 受试者工作特征曲线下面积; PPV, 阳性预测值; NPV, 阴性预测值; GNB, 高斯朴素贝叶斯; CNB, 互补朴素贝叶斯; GNB, 高斯朴素贝叶斯; MLP, 多层感知器神经网络; SVM, 支持向量机; NA, 不适用。

Figure 4.



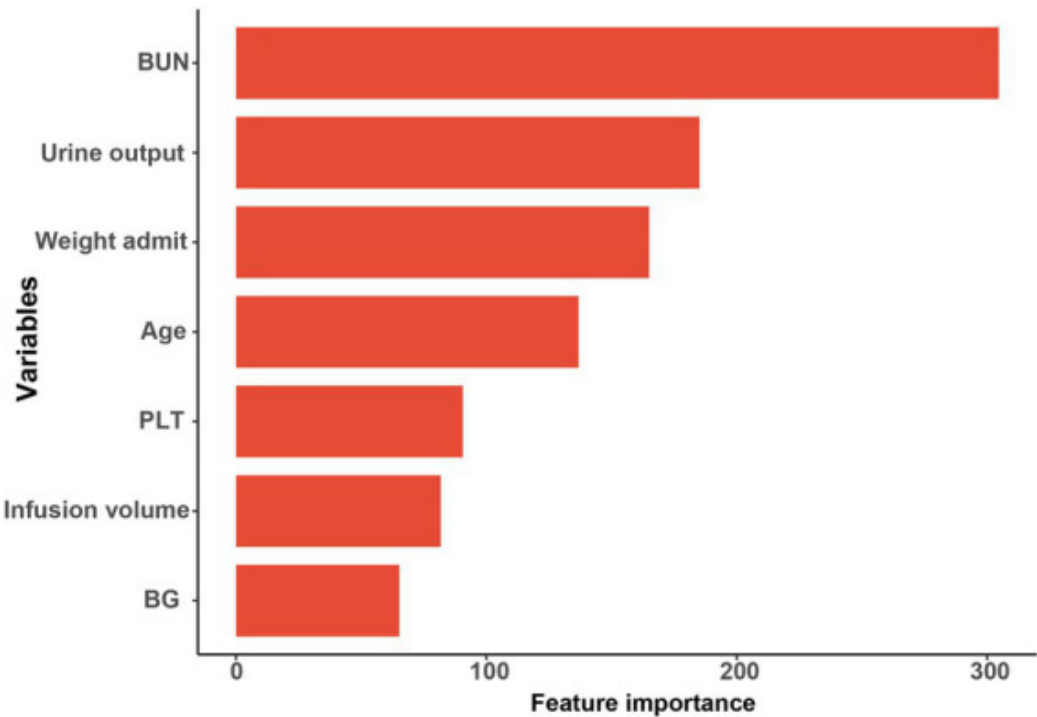
DCA (A) and calibration curve (B) of the XGBoost and simplified model. DCA, decision curve analysis; XGBoost, eXtreme Gradient Boosting.  
XGBoost 和简化模型的 DCA (A) 和校准曲线 (B)。DCA, 决策曲线分析; XGBoost, 极端梯度提升。

### 3.3. Relative importance of variables 3.3.变量的相对重要性

Based on the XGBoost model with superior performance, 7 characteristics of interest were finally obtained in this study. Feature importance analysis was conducted to interpret the importance of the variables, which revealed that the top 5 contributing variables were blood urea nitrogen (BUN), urine output, weight, age, and platelet count (PLT), in that order (Figure 5 and Supplementary Table S3).

基于性能上级的 XGBoost 模型，本研究最终获得了 7 个感兴趣的特征。进行特征重要性分析以解释变量的重要性，其显示前 5 个贡献变量依次为血尿素氮（BUN）、尿量、体重、年龄和血小板计数（PLT）（ Figure 5 和 Supplementary Table S3 ）。

Figure 5.



BUN, blood urea nitrogen; PLT, platelet count, BG, blood glucose.  
BUN，血尿素氮; PLT，血小板计数; BG，血糖。

## 4. Discussion 4.讨论

Although the incidence of DKA-AKI was lower than in previous studies by Junzhe Chen (3) and Jean-Christophe Orban (8), it was comparable to the findings of our previous study in the MIMIC-III database (9). Thirty-eight candidate predictors were applied to train and validate 8 MI models to predict the risk of DKA-AKI. The XGBoost model outperforms all other ML method in terms of discrimination and accuracy, with an AUC value of 0.835 in the training set and 0.800 in the validation set. BUN, urine output, weight, age, PLT, fluid volume, and glucose were the 7 variables that contributed to the XGBoost model orderly. The mortality rate among patients in the AKI group was found to be 10 times higher than that of the non-AKI group, with rates of 10.1 and 1.3%, respectively. Additionally, patients who suffered from AKI were more likely to require mechanical ventilation and spend longer periods of time in the ICU, leading to increased medical expenses. Accurately predicting the incidence of DKA-AKI can assist clinicians in identifying high-risk DKA-AKI patients in the ICU, and timely treatment and management can significantly improve the prognosis of these patients.

尽管 DKA-AKI 的发生率低于 Junzhe Chen（3）和 Jean-Christophe Orban（8）之前的研究，但与我们之前在 MIMIC-III 数据库（9）中的研究结果相当。应用 38 个候选预测因子来训练和验证 8 个 MI 模型以预测 DKA-AKI 的风险。XGBoost 模型在区分度和准确性方面优于所有其他 ML 方法，训练集的 AUC 值为 0.835，验证集为



0.800。BUN、尿量、体重、年龄、PLT、液体体积和葡萄糖是对 XGBoost 模型有贡献的 7 个变量。阿基组患者的死亡率比非 AKI 组高 10 倍，分别为 10.1%和 1.3%。此外，患有阿基的患者更可能需要机械通气并在 ICU 中花费更长时间，导致医疗费用增加。准确预测 DKA-AKI 的发生率可以帮助临床医生识别 ICU 中的高危 DKA-AKI 患者，及时的治疗和管理可以显著改善这些患者的预后。

With the evolution of electronic medical and advent of the Big Data era, ML has already achieved remarkable achievements in the diagnosis and prognosis of diseases (17). For instance, ML methods have been maturely applied to develop predictive models for AKI in patients with sepsis, as well as in those undergoing cardiac and liver surgeries (13, 18–21).

随着电子医疗的发展和大数据时代的到来，ML 在疾病诊断和预后方面已经取得了显著的成就（17）。例如，ML 方法已被广泛应用于开发脓毒症患者以及接受心脏和肝脏手术的患者阿基预测模型（13，18 - 21）。

BUN and urine output at baseline were included in our final model and were ranked higher in importance, so more attention should be given to patients with DKA who already had abnormal BUN and urine volume at ICU admission. Additionally, body weight was found to be a predictor for DKA-AKI in our study, with a median weight of 70.0 [IQR 60.3, 83.6] (Kg) in the AKI group compared to 78.5 [IQR 66.0, 94.0] (Kg) in the non-AKI group. Shi et al. (22) reported that overweight and obese patients were at a significantly higher risk of cardiac surgery-related AKI. However, since height is largely absent of ICU patients from the database, we were unable to calculate body mass index (BMI). Notably, increased intra-abdominal pressure in critically ill obese patients causes venous obstruction and insufficient blood flow to arterial organs, which may account for why obese patients are more likely to experience AKI (22–24). Age was also found to be significantly associated with AKI in our study, consistent with previous research (3, 8). The median age was higher in the AKI group (58 years [IQR 46, 68]) compared to the non-AKI group (43 years [IQR 30, 57]). Age-related structural and functional changes in the kidney include glomerulosclerosis, a decline in estimated glomerular filtration rate (eGFR), and an increase in glomerular capillary pressure. The kidney becomes more vulnerable to acute injury as it ages because its capacity to self-regulate declines (25). Long-term poor glycemic control, which is common in older patients with diabetes, can also lead to persistent kidney damage through inflammation, oxidative stress, and glycosylation (26). Both fluid therapy and low-dosage insulin are crucial treatments for DKA. Our study found that patients with AKI received more fluid infusions and experienced less urination compared to patients without AKI, leading to a higher cumulative fluid balance. Similar findings have been reported in earlier studies (27–29). Furthermore, Inkinen et al. (30) demonstrated that this phenomenon is linked to a lack of recovery from AKI. Excessive fluid intake may result in interstitial renal edema, which can raise renal perfusion pressure and impair kidney function. Ischemia–reperfusion (I/R) injury represents one of the crucial mechanisms of AKI. Following this, coagulation and inflammation are activated, and platelets are crucial to this process. In an animal model of AKI induced by I/R, Jansen et al. demonstrated that a significant proportion of activated platelets were present in the necrotic zone. Furthermore, the application of clopidogrel prior to modeling reduced tubular necrosis and preserved some renal function in mice (31). Also, an observational cohort study discovered that a significant connection between preoperative aspirin use and a reduced risk of AKI linked to cardiac surgery (32). Further research is necessary to determine whether antiplatelet medications can be utilized to prevent and treat other causes of AKI. BG was still included in the final XGBoost model and was negatively correlated with the outcome even though there was no statistically significant difference between the AKI (191.0 mg/dl [IQR 139.0, 258.0]) and the

non-AKI groups (183 mg/dl [IQR 130.0, 251.0]) in terms of BG level. This could be due to the fact that individuals with kidney damage are more likely to experience hypoglycemia when taking insulin (33).

在我们的最终模型中纳入了基线时的 BUN 和尿量，并且重要性排名较高，因此应更多关注在 ICU 入院时已存在 BUN 和尿量异常的 DKA 患者。此外，在我们的研究中发现体重是 DKA-AKI 的预测因子，阿基组的中位体重为 70.0 [IQR 60.3, 83.6] (Kg)，而非 AKI 组为 78.5 [IQR 66.0, 94.0] (Kg)。Shi 等人 (22) 报告称，超重和肥胖患者发生心脏手术相关阿基的风险显著较高。然而，由于 ICU 患者的身高在数据库中基本缺失，因此我们无法计算体重指数 (BMI)。值得注意的是，重症肥胖患者的腹内压升高会导致静脉阻塞和动脉器官血流不足，这可能是肥胖患者更容易发生阿基的原因 (22 - 24)。在我们的研究中还发现年龄与阿基显著相关，与之前的研究一致 (3, 8)。阿基组的中位年龄 (58 岁 [IQR 46, 68]) 高于非 AKI 组 (43 岁 [IQR 30, 57])。肾脏中与肾小球疾病相关的结构和功能变化包括肾小球硬化、估计肾小球滤过率 (eGFR) 下降和肾小球毛细血管压升高。随着年龄的增长，肾脏变得更容易受到急性损伤，因为它的自我调节能力下降 (25)。在老年糖尿病患者中常见的长期血糖控制不良也可通过炎症、氧化应激和糖基化导致持续性肾损伤 (26)。液体治疗和低剂量胰岛素都是 DKA 的重要治疗方法。我们的研究发现，与无阿基的患者相比，AKI 患者接受了更多的液体输注，排尿较少，导致累积液体平衡较高。在早期的研究中也报告了类似的结果 (27 - 29)。此外，Inkinen 等人 (30) 证明，这种现象与阿基恢复不足有关。过量的液体摄入可能导致间质性肾水肿，这可能升高肾灌注压并损害肾功能。缺血再灌注 (I/R) 损伤是阿基的重要机制之一。在此之后，凝血和炎症被激活，血小板对这一过程至关重要。在 I/R 诱导的阿基动物模型中，Jansen 等人证明，坏死区中存在显著比例的活化血小板。此外，在建模前应用氯吡格雷减少了肾小管坏死并保留了小鼠的一些肾功能 (31)。此外，一项观察性队列研究发现，术前使用阿司匹林与心脏手术相关的阿基风险降低之间存在显著联系 (32)。需要进一步研究以确定抗血小板药物是否可用于预防和治疗阿基的其他原因。BG 仍被纳入最终 XGBoost 模型中，并且与结局呈负相关，尽管阿基组 (191.0 mg/dl [IQR 139.0, 258.0]) 和非 AKI 组 (183 mg/dl [IQR 130.0, 251.0]) 之间的 BG 水平无统计学显著性差异。这可能是由于肾损伤的个体在服用胰岛素时更容易发生低血糖 (33)。

In contrast to previous clinical studies, we have successfully constructed a tool for assessing the risk of DAK-AKI based on several supervised ML algorithms. This model is constructed using a large real-world database and plays a crucial role in risk assessment and stratification of AKI, particularly in critically ill patients. To our knowledge, our study is the first to develop and validate a ML-based model of DKA-AKI, assisting clinicians in identifying high-risk individuals' early and addressing associated risk factors promptly. Furthermore, our sample size of DKA patients is the largest compared to similar studies, ensuring more robust results. Encouragingly, the prediction models constructed in our study exhibit promising performance in both the training and validation sets.

与以前的临床研究相比，我们已经成功地构建了一种基于几种监督 ML 算法的 DAK-AKI 风险评估工具。该模型是使用大型真实世界数据库构建的，在阿基的风险评估和分层中发挥着至关重要的作用，特别是在重症患者中。据我们所知，我们的研究是第一个开发和验证基于 ML 的 DKA-AKI 模型的研究，有助于临床医生早期识别高危人群并及时解决相关风险因素。此外，与类似研究相比，我们的 DKA 患者样本量最大，确保了更稳健的结果。令人鼓舞的是，在我们的研究中构建的预测模型在训练集和验证集中都表现出了良好的性能。

Limitations of our present study should be noted. Firstly, our research was conducted as a single-center study on ICU patients, and we obtained data from the MIMIC-IV database. Further research will require data collection on patients from various nations or regions as well as general medical wards to enhance the generalizability of the findings.

我们目前研究的局限性应该被注意到。首先，我们的研究是以 ICU 患者为对象的单中心研究，我们从 MIMIC-IV 数据库中获得数据。进一步的研究将需要收集来自不同国家或地区以及普通医疗病房的患者数据，以提高研究结果的普遍性。

Second, missing data were present in this study, and some variables with missing data of >20% were excluded. However, we made every effort to address the limitations of missing data by applying the KNN algorithm to the dataset for interpolation. Third, this is a retrospective study based on a public database the results have some limitations. Finally, in the absence of prior monitoring data in the MIMIC-IV database, it becomes challenging to assess if a patient with DKA has developed AKI. Hence, we cannot exclude this group of patients from the study, which may result in biased results.

其次，本研究中存在缺失数据，并排除了一些缺失数据>20%的变量。然而，我们尽一切努力通过将 KNN 算法应用于数据集进行插值来解决缺失数据的局限性。第三，这是一项基于公共数据库的回顾性研究，结果有一定的局限性。最后，在 MIMIC-IV 数据库中缺乏既往监测数据的情况下，评估 DKA 患者是否发生阿基变得具有挑战性。因此，我们不能从研究中排除这组患者，这可能导致偏倚结果。

## 5. Conclusion 5.结论

Collectively, we extracted the data from the MIMIC-IV database to build ML-based model to predict the DKA-AKI. The results showed that BUN, urine output, weight, age, PLT, infusion volume, and glucose, in order of importance, were predictors of the occurrence of DKA-AKI. This model can identify high-risk patients at an early stage, assist clinical decision-making, and may improve the prognosis of DKA patients to some extent.

总的来说，我们从 MIMIC-IV 数据库中提取数据，建立基于 ML 的模型来预测 DKA-AKI。结果显示，BUN、尿量、体重、年龄、PLT、输注量和葡萄糖（按重要性顺序）是 DKA-AKI 发生的预测因子。该模型可早期识别高危患者，辅助临床决策，并可能在一定程度上改善 DKA 患者的预后。

## Data availability statement 数据可用性声明

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

支持本文结论的原始数据将由作者提供，无不当保留。