

基于XGBoost和SHAP的急性肾损伤可解释预测模型

罗妍^① 王枏^② 叶文玲*^③

^①(北京邮电大学计算机学院(国家示范性软件学院) 北京 100876)

^②(北京邮电大学可信分布式计算与服务教育部重点实验室 北京 100876)

^③(中国医学科学院中国协和医科大学北京协和医院肾内科 北京 100730)

摘要:重症监护病房(ICU)住院期间发生的急性肾损伤(AKI)与患者发病率和死亡率的增加有关。该研究的目的是提出一个基于机器学习的框架,用于危重病患者的可解释AKI预测,该框架能够同时实现良好的预测和解释能力。该文从重症监护医学信息数据库III(MIMIC-III)中提取的数据包括患者的年龄、性别、生命体征和ICU入院第1天及随后的化验值。在该研究中,通过将XGBoost模型与其他4种机器学习模型进行比较,证明了XGBoost模型的预测性能。此外,SHAP(SHapley Additive exPlanation)模型可解释器用于提供个性化评估和解释,以实现个性化的临床决策支持。结果表明,XGBoost能较好地预测AKI,与以往的预测模型相比,此模型更为简单有效,仅用21个特征变量即得到了更稳定的预测结果,预测精度高,模型准确率和受试者工作特征曲线下面积(AUC)分别为0.824和0.840,均高于既往研究结果。此外,该文对两组指标进行了特征依赖分析,发现24h尿量减少和血尿素氮升高可增加AKI风险。综上所述,该可解释预测模型可能有助于临床医生更准确快速地识别重症监护中存在AKI风险的患者,为患者提供更好的治疗。此外,可解释性框架的使用增加了模型透明度,便于临床医生分析预测模型的可靠性。

关键词:急性肾损伤;重症监护;模型解释;临床决策支持;预测模型

中图分类号: TP391

文献标识码: A

文章编号: 1009-5896(2022)01-0027-12

DOI: 10.11999/JEIT210931

An Interpretable Prediction Model for Acute Kidney Injury Based on XGBoost and SHAP

LUO Yan^① WANG Cong^② YE Wenling^③

^①(School of Computer Science (National Pilot Software Engineering School), Beijing University of Posts and Telecommunications, Beijing 100876, China)

^②(Key Laboratory of Trustworthy Distributed Computing and Service, Beijing University of Posts and Telecommunications, Ministry of Education, Beijing 100876, China)

^③(Department of Nephrology, Peking Union Medical College Hospital, Chinese Academy of Medical Sciences & Peking Union Medical College, Beijing 100730, China)

Abstract: The development of Acute Kidney Injury (AKI) during admission to the Intensive Care Unit (ICU) is associated with increased morbidity and mortality. The objective of this study is to develop a machine learning-based framework for interpretable AKI prediction in critical care that can achieve both good prediction and interpretation capability. Data extracted from the Medical Information Mart for Intensive Care III (MIMIC-III) include patient age, gender, vital signs and lab values during the first day of ICU admission and subsequent hospitalization. In this study, the prediction performance of the XGBoost model is demonstrated by comparing it to four other machine learning models. In addition, the SHapley Additive exPlanation (SHAP) framework is used to provide individualized evaluation and explanations to enable personalized clinical decision support. The results show that XGBoost can predict AKI robustly with an Accuracy and the area Under the receiver operating Characteristic curve (AUC) of 0.824 and 0.840, respectively, which are higher than previous prediction models. Furthermore, a feature dependency analysis is conducted for two pairs of features and found decrease in urine volume and elevation of blood urea nitrogen indicates an increase of AKI risk. To sum up, this

interpretable predictive model may help clinicians more accurately identify patients at risk of AKI in intensive care and provide better treatment for patients. In addition, the use of this interpretability framework increases model transparency and facilitates clinicians to analyze the reliability of predictive models.

Key words: Acute kidney injury; Intensive Care Unit (ICU); Model interpretation; Clinical decision support; Prediction model

1 引言

急性肾损伤(Acute Kidney Injury, AKI)的特点是在48h内肾功能迅速下降^[1],患者主要表现为血清肌酐(Serum Creatinine, Scr)升高或尿量减少^[2]。根据2012年改善全球肾脏病预后组织(Kidney Disease: Improving Global Outcomes, KDIGO)提出的诊断标准,AKI被定义为48小时内Scr增加至少 $26.5\mu\text{mol}\cdot\text{L}^{-1}$ ($0.3\text{ mg}\cdot\text{dL}^{-1}$)或达到基线值的1.5倍及以上,且明确或经推断上述情况发生在7d之内,尿量减少小于 $0.5\text{ mL}\cdot(\text{kg}\cdot\text{h})^{-1}$,且时间持续6h以上^[3]。AKI在各种临床环境中都会经常遇到,占重症监护病房(Intensive Care Unit, ICU)患者的20%~50%^[4-6],住院患者的5%~7%^[7,8]。此外,AKI还与高死亡风险相关,与无AKI的患者相比,患有AKI患者的死亡风险增加了7倍^[9]。有研究称在ICU,AKI患者的死亡率高达10.3%~26.5%^[10-14]。美国的住院数据显示,AKI造成每年估计47~240亿美元的费用^[15]。

有研究表明,AKI分期越高,需要肾脏替代治疗的患者越多,病死率越高^[16-18]。也有越来越多的证据表明,即使AKI在症状上治愈,其长期慢性肾病、心血管疾病和死亡的风险也会增加^[19-21]。因此,早期识别AKI风险患者,并对AKI高风险患者进行早期干预,对于减少AKI的发病率、有效缩短疾病进展并降低死亡率上具有重大意义。

在AKI概念出现之前,常用的概念是急性肾功能衰竭(Acute Renal Failure, ARF),为了强调早期肾损害在诊断标准中的重要性才被AKI取代^[22]。而早期诊断和干预可延缓AKI进展为严重阶段和出院后慢性肾病的发展^[23]。考虑到AKI的高死亡率和早期发现的重要性,临床医生需要新的方法在显著肾损害发生之前早期预测AKI。

近年来,机器学习(Machine Learning, ML)算法在疾病预测模型中应用广泛,有研究发现可以利用Scr水平和现有的临床指标来预测AKI的发生。Flechet等人^[24]对252名危重患者进行了前瞻性研究,并将医生对AKI预测的结果与基于随机森林算法的AKI预测模型进行了比较,发现AKI预测模型能筛选出高危患者和减少假阳性率,并能比医生更早提供预测。

但目前流行的AKI风险预测模型主要关注提高模型的准确性,而对模型解释的关注较少,这阻碍了机器学习模型在临床实践中的应用^[25,26]。因此,为了支持临床医生的决策,有必要在保持准确预测的同时理解和解释这些关系。然而,同时实现良好的可预测性和可解释性是具有挑战性的,因为许多预测模型通常是“黑箱”的,而可解释的模型通常具有较差的预测性能^[27]。为此,本文的目标是开发一个可进行AKI早期预测的可解释模型,重点是模型解释,并期望该模型能广泛应用于临床。

本文的结构如下:第2节介绍本研究的相关工作。第3节将深入讨论XGBoost(eXtreme Gradient Boosting)和SHAP(SHapley Additive exPlanation)。第4节和第5节对AKI风险预测模型的结果和性能进行分析和讨论,并通过SHAP提供综合特征解释。最后,对研究结果、优势和局限性进行详细的讨论。

2 相关工作

疾病的早期预测可以在支持医疗卫生专业人员方面发挥重要作用,据统计,11%的医院死亡是由于未能及时识别和治疗病情恶化所致^[28]。近5年来,ML方法在准确、及时预测AKI高危患者方面发挥了重要作用。例如,Flechet等人^[29]开发并验证了随机森林预测模型,成功预测了成年ICU患者的AKI,受试者工作特征曲线下面积(Area Under the receiver operating characteristics Curve, AUC)达到0.84。

Mohamadlou等人^[30]使用梯度提升算法以生命体征和Scr为特征预测重症AKI,AUC值在第48小时和72小时分别为0.76和0.73。次年,Lei等人^[31]也使用同样的机器学习模型来识别术后出现AKI的患者,最终模型的AUC为0.82。此外,Koyner等人^[32,33]开发并在最近对其梯度提升算法模型进行了外部验证,该预测模型对48小时内1期AKI的预测AUC值为0.67。

Simonov等人^[34]利用来自同一卫生保健系统中3家医院的169859名成人住院患者的回顾性数据开发了一个简单的逻辑回归模型,并证明基于现有实验室数据的简单模型可以准确预测AKI,AUC为0.74。

Xu等人^[35]用ML模型(逻辑回归、随机森林和梯度提升算法)对58976例重症监护病房AKI患者的死亡风险进行预测,发现梯度提升算法的AUC为0.75,在预测死亡率方面优于其他模型。

Zimmerman等人^[36]对23950例成年危重病患者进行了研究,建立了逻辑回归预测模型,用于入ICU 72 h后AKI的早期预测,AUC为0.78。

Rashidi等人^[37]开发并内部验证了ML模型,比较了50名烧伤和51名创伤患者的AKI早期识别。他们的递归神经网络模型能够提前62h准确预测AKI,AUC为0.92。

如前所述,大多数模型的AUC表现一般^[30,32-35],而一些研究集中于特定的患者群体,如心脏手术患者^[31],或研究样本量小^[37],从而限制了这些模型的使用。此外,从表1可以看出,目前大多数研究中使用的特征向量非常繁琐,收集和计算也很复杂。这些研究大多缺乏模型可解释性,只有有限的解释提供简单的特征重要性结果。相比之下,本文在前期研究的基础上,使用公开可访问的包含超过46000名患者的去识别健康数据的重症监护医学信息数据库MIMIC(Medical Information Mart for Intensive Care)III进行数据分析和模型开发^[38],最后基于XGBoost算法构建了重症监护病房患者的AKI早期预测模型,并比较了XGBoost与其他4种流行的机器学习技术的性能。模型中仅使用常见的生命体征和实验室检测指标,通过有效的数据预处理和XGBoost模型参数调整,取得了良好的AKI早期风险预测性能。然后,利用SHAP估计的Shapley值从全局和局部两个角度对预测模型进行解释。解释结果不依赖所使用的预测模型,这保证了结果的可靠性并为解决临床问题提供更多的证据支撑。这些成为这项工作的主要贡献。

3 方法

3.1 数据源

本研究中属于回顾性队列研究,数据来源于一个公开的、大规模的重症医学数据库——MIMIC-III。该数据库由麻省理工学院计算生理学实验室开发,整合了2001年至2012年贝斯以色列女执事医疗中心46520例成年ICU患者所有连续的匿名综合临床数据,为重症医学科医生开展临床研究提供了极大的便利^[38]。

本研究的目的是预测ICU患者AKI的发生。我们使用KDIGO诊断标准对AKI进行定义^[3],在48 h内Scr水平达到基线值的1.5倍及以上或增加0.3 mg/dL。考虑到AKI分期的特异性较差^[39],以及我们的数据集集中缺乏足够符合要求的尿量数据,因此本研究排除了尿量标准。基线值定义为入院后24 h内首次Scr测量值。为明确早期、亚临床的患者生理基线恶化是预测AKI的重要原因,使患者得到及时的治疗,本研究主要使用第1天的生理变量检测值来预测未来48 h内ICU住院患者发生AKI的风险。在19001例ICU记录中,共提取出4532例满足以上条件的AKI病例。

3.2 特征变量提取

本研究的主要结果是入院后3 d内AKI的发生率。其时间窗示意图如图1所示,Pt(Prediction time)表示预测时间,Et(Event time)表示AKI发生时间,即Pt发生后48 h。

其中Et的预测结果被编码为一个二元分类的因变量,分为两类:0(患者在Et期间未出现AKI)和1(患者在Et期间出现AKI)。为了预测AKI的发生,本文根据ICU常用的疾病风险预测评分系统(如Apache, SAPS等)、专家意见和数据集可用变量情况,从MIMIC III数据库中提取了患者的24 h临床

表1 既往研究结果表明ML技术可以有效地用于AKI预测

研究	数据集	记录	ML 算法	病人	AUC
Mohamadlou 等人 ^[30]	MIMIC-III; Stanford University data set	48582 19737	Gradient Boosting	ICU patients/ inpatients	0.73-0.80
Lei等人 ^[31]	University of Pennsylvania Health System	42615	Gradient Boosting	noncardiac surgery	0.82
Koyner等人 ^[32]	University of Chicago (an urban tertiary referral hospital)	121158	Gradient Boosting	adult admissions	0.73
Churpek等人 ^[33]	3 health systems	495971	Gradient Boosting	adult admissions	0.67-0.72
Simonov等人 ^[34]	3 hospitals	169859	logistic regression	adult admissions	0.74
Xu等人 ^[35]	MIMIC-III	8181	Gradient Boosting Decision Tree	ICU patients	0.73
Zimmerman等人 ^[36]	MIMIC-III	23950	Logistic Regression	adult critical care patients	0.78
Rashidi等人 ^[37]	UC Davis clinical laboratory	101	Recursive Neural Network	burn and trauma patients	0.92

资料记录,包括患者的年龄、性别、Scr、生命体征和实验室化验结果,以及每小时尿量、24 h尿量等。为了更好地理解数据,表2列出了AKI疾病组和正常对照组两组变量的分布情况(百分比)。该分析是在预处理任务之前进行的。

3.3 数据预处理

由于AKI事件的发生与时间相关,这给数据提取和使用带来了障碍。在MIMIC-III数据库中,一些与时间相关的特征测量结果缺失,缺失值的百分比高达60%以上。此外,表2的数据表明,AKI患

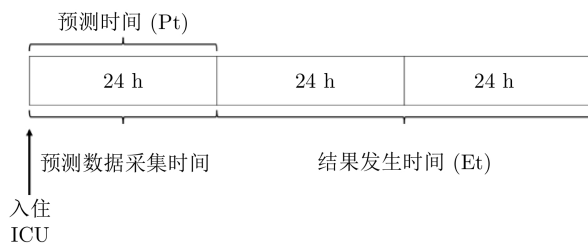


图1 数据提取时间间隔示意图(AKI的确诊时间窗为ICU入院24 h后)

者和非AKI患者的类别比例极不平衡。这两个问题使得AKI早期预测模型的准确预测和解释变得更加困难。因此,处理缺失的值并解决类别的不平衡十分重要。

与大多数临床和实验室数据集一样,本文的基础数据集包含缺失值,例如,与平均动脉压、心率、体温、肺泡-动脉氧梯度、红细胞比容、白蛋白和胆红素相关的时间序列有55%~90%的数据缺失。考虑到这些变量的重要性,我们仍然保留着它们。对于这些缺失值,我们首先分别用均值和中位数来填充,但两者的效果都很差。在严重缺失的情况下,中位数甚至更不可靠。使用平均值填充的优点是不会造成严重的数据失真,并且可以保证填充后的平均值不会发生太大的变化。但在严重缺失数据的情况下,这种填充会导致整个样本信息量大幅减少,破坏数据的随机性。因此,当严重缺失的特征很重要时,将缺失的特征作为一个类别来处理是十分有效的,这样可以既保留特征又保证数据的无偏估计。本文没有对缺失的数值进行任何插补,因

表2 各变量在基础数据集中的分布情况(%)

变量	正常对照组	疾病组
Age, years	63.41±16.25	66.54±14.77
Gender	Female	1,846(40.73%)
	Male	2,686(59.27%)
	Median, Interquartile Range (IQR; 25%–75%)	Median, Interquartile Range (IQR; 25%–75%)
GCS	14.00 (8.00–15.00)	10.00 (3.00–15.00)
Scr_baseline, mg/dL	0.90 (0.70–1.10)	1.40 (0.90–2.40)
MAP, mmHg	77.29 (70.39–84.90)	74.54 (68.21–81.96)
HR, beats/min	84.39 (74.06–95.84)	85.38 (76.44–97.98)
T, °C	37.10 (36.75–37.46)	37.07 (36.65–37.41)
R, /min	18.35 (16.09–21.16)	18.62 (16.02–21.93)
PO ₂ , mmHg	124.67 (93.00–159.00)	124.17 (97.50–152.18)
A-aDO ₂ , mmHg	439.00 (351.00–532.00)	475.50 (381.00–556.75)
HCT, %	30.00 (27.00–34.00)	28.50 (25.15–32.00)
WBC, ×10 ⁹ /L	10.60 (7.80–14.00)	11.62 (8.22–15.90)
BUN, mg/dL	17.00 (12.00–26.00)	27.50 (17.67–44.67)
Na, mmol/L	138.67 (136.00–141.00)	138.00 (135.88–140.50)
K, mmol/L	4.03 (3.75–4.37)	4.27 (3.93–4.65)
ALB, mg/dL	3.10 (2.60–3.60)	2.80 (2.40–3.20)
BIL, mg/dL	0.70 (0.40–1.40)	1.00 (0.50–3.11)
pH	7.39 (7.35–7.43)	7.36 (7.32–7.40)
Glu, mg/dL	127.00 (107.58–153.00)	132.21 (114.00–161.00)
FU, ml/h	105.88 (70.42–156.76)	67.00 (34.49–108.39)
TVU, ml/24 h	1,810.00 (1,194.00–2,663.00)	1,148.00 (549.75–1,995.00)

缩略词: GCS, 格拉斯哥昏迷评分; R, 呼吸率; PO₂, 动脉血氧分压; WBC, 白细胞; BUN, 血尿素氮; Na, 血钠; K, 血钾; pH, 酸碱度; Glu, 血糖; TVU, 24 h总尿量

为缺失值的估算并不能都为以电子健康记录为基础的预测模型提供一致的改进效果^[40]。

如表2所示, 正常组纳入14469条病例, AKI病例入选仅4532条。非AKI比AKI的比率接近3:1, 即该数据集的分布是高度不平衡的。目前一般采用过采样和欠采样技术来处理不平衡数据。并且由于欠采样技术往往会导致大量数据的丢失, 从而导致模型精度的下降^[41]。所以, 过采样方法通常是首选。SMOTE(Synthetic Minority Oversampling Technique)是一种常用的过采样方法, 它通过线性插值的方法在两个少数类样本间合成新的样本, 从而有效缓解由随机过采样引起的过拟合问题。本研究使用了Chawla等人^[42]提出的SMOTE方法来平衡数据。

最后, 由于本文使用的数据集中原始数据的取值范围已经确定, 为了保持原始数值之间的关系, 我们选择了Min-max归一化方法将我们的数据特征值归一化到[0,1]的范围。

3.4 预测模型

本研究使用XGBoost算法对ICU患者发生AKI的风险进行建模。XGBoost最初由Chen和Guestrin在2016年提出^[43]。XGBoost是一个集成分类器, 它通过将多个决策树模型进行集成来增强分类能力。也就是说, 需要将多棵树的得分相加得到最终的预测得分(每次迭代, 都在现有的树上添加一个新的函数, 拟合上一轮预测结果与真实值之间的残差)。为了尽可能大地降低目标函数, XGBoost使用了2阶泰勒展开来自定义损失函数, 可以在不选定损失函数具体形式的情况下, 仅通过输入数据值就能进行叶子分裂优化计算, 即将损失函数的选取和模型算法优化分开了。这种去耦合增加了XGBoost的适用性, 可以使其更为精准地逼近真实的损失函数。XGBoost在许多领域都有出色的表现^[44,45], 并显示出更高的预测精度和更快的处理时间, 同时计算成本更低, 复杂性更低^[43,46]。这些特性正好适合在预测AKI时应对稀疏和高维临床数据的需要。XGBoost在应用机器学习方面的优势在于其极高的精度和速度。在本研究中, 我们还将XGBoost与其他4种常用的ML算法, 包括逻辑回归(Logistic Regression, LR)、支持向量机(Support Vector Machine, SVM)、随机森林分类器(Random Forest, RF)和人工神经网络(Artificial Neural Network, ANN)的性能进行了比较, 以验证XGBoost是否能够提供最佳性能。

3.5 参数调整

根据Chen等人^[43]的研究, XGBoost模型的性能可以通过调整参数来最大化。参数调优对于XG-

Boost防止过拟合很重要, 由于本研究中使用的方法是监督学习, 而网格搜索是监督学习中调整参数和提高模型泛化性能的有效方法^[47], 所以我们选择了网格搜索来调整参数以获得最优性能。

网格搜索是执行超参数调优以确定给定模型最优值的过程。手动执行此操作可能会花费相当多的时间和资源, 因此我们使用GridSearchCV来自动化超参数的调优, 其中“CV”代表“交叉验证”。GridSearchCV可对估计器的指定参数值穷举搜索, 使某些估计器更加有效。在本文中, 交叉验证后的最佳XGBoost超参数值是: learning_rate: 0.05, max_depth: 6, n_estimators: 700, min_child_weight: 1, gamma: 0.6, colsample_bytree: 0.7, subsample: 0.9, reg_alpha: 0.1, reg_lambda: 3。

本研究中所有的预测模型和相关分析都是使用Python3.7中的开源库(scikit-learn, XGBoost和SHAP)在Jupyter notebook平台上进行的。

3.6 模型评估

本文将数据集随机分为训练集(70%的队列)和测试集(30%的队列), 预测AKI风险的模型仅使用来自训练集的数据建立, 针对训练集进行不重复抽样随机分为5份。每次都使用其中4份来训练模型, 余下的1份用来验证4份训练出来的模型的准确率, 重复该步骤5次, 直到每个子集都有一次机会作为验证集, 其余机会作为训练集。计算5组测试结果的平均值作为模型精度的估计, 并作为当前5折交叉验证下模型的性能指标。交叉验证可以在一定程度上减小过拟合, 并从有限的数据中获取尽可能多的有效信息。在模型开发过程中, 为模型性能的测量提供了一种更加稳定可靠的方法^[48]。最后, 我们利用剩余的30%测试集进行内部验证, 这有利于我们评估训练好的模型在新数据上的表现。

KDIGO定义的1期AKI预测结果是本文的主要结果, 本文使用指标受试者工作特征曲线(Receiver operating characteristic curves)下面积AUC评估模型性能。除此之外, 我们还引入了以下4个常用指标: 准确性(ACC)、敏感度(Sen, 真正阳性率, 即召回率)、精确率(Pre, 阳性预测值)和F1值^[49]。其计算方法为

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

$$Sen = \frac{TP}{TP + FN} \quad (2)$$

$$Pre = \frac{TP}{TP + FP} \quad (3)$$

$$F1 = \frac{2 \text{精确率} \times \text{召回率}}{\text{精确率} + \text{召回率}} \quad (4)$$

TP, FP, TN, FN分别表示真阳性、假阳性、真阴性、假阴性。AUC为受试者操作特征曲线下面积, 而该曲线图是反映敏感性与特异性之间关系的曲线。横坐标X轴为1-特异性, 也称为假阳性率(误报率), X轴越接近零准确率越高; 纵坐标Y轴称为敏感度, 也称为真阳性率(敏感度), Y轴越大代表准确率越好^[50]。值得注意的是, 考虑到模型的AUC值对预测建模和决策有显著意义, 主要表现为高特异性较少引起干预, 而高敏感性更容易触发干预^[51], 因此我们在模型选择和最终报告中使用的主要指标是AUC。此外, 为了了解模型性能的不确定性, 我们计算了95%置信区间。

3.7 模型解释

尽管机器学习在许多领域取得了巨大的成功, 但因缺乏可解释性严重限制了其在现实任务尤其是安全敏感任务中的应用。可解释性在医学领域中是非常重要的。医疗辅助诊断系统必须是可理解的、可解释的, 理想情况下, 它应该能够向所有相关方解释提供对应决策的完整逻辑, 才能获得医生的信任。由于医学数据自有的特点, 构建用于医学结构化数据分析的可解释深度学习模型与其他领域中的应用是不同的。本文重点关注SHAP可解释方法在AKI早期辅助诊断模型中的应用。SHAP基于博弈论^[52]和局部解释^[53], 属于经典的事后解释框架, 可以提供Shapley值来估计每个特性的贡献。Shapley值是一种描述模型在对特定数据点进行预测时特定特征的“权重”或“重要性”的方法, 是SHAP的核心所在。与传统的特征重要性方法(如XGBoost的特征重要性)相比, SHAP具有更好的一致性, 可以呈现各预测因子相对于目标变量的正/负关系, 用于局部和全局解释^[54,55]。对于局部可解释性, 每种特征都有自己的一组Shapley值。因此, 它可以用来解释每个样本每个特征对预测的贡献, 增加了透明度, 便于临床医生分析预测模型的可靠性。将所有样本中对应该变量的Shapley值取平均值作为该特征的重要性值, 即可得到全局解释。在本文中, 通过计算各ICU临床指标的Shapley值来分析各特征与AKI风险的相关性, 并结合相关临床研究结果和临床实际表现情况来进行模型解释。

因为Shapley值是解释特征重要性的唯一方法, 满足局部准确性和一致性的数学特性, 而Tree Explainer 是专门解释树模型的解释器, 通过树集成和加法方法激活Shapley值用于特征归因。因此, 本文采用 $\text{mean}|\text{Tree SHAP}|$, 一种基于个性化

的启发式SHAP平均的全局属性方法, 对特征进行排序, 以明确模型中的重要预测因素。

4 结果

本研究使用MIMIC III进行数据提取和模型验证。表1为研究样本的统计结果, 需要注意的是, 在MIMIC III数据集中, 我们使用icustay ID作为唯一标识。阴性样本数量通常远远多于阳性样本数量, 这是临床数据集普遍存在的问题。基础数据集中包含19001例患者, 其中发生AKI的患者4532例, 约占23.85%。在这个数据集中, 他们的平均年龄(标准差[SD])为64.15(15.96)岁, 其中女性为8110例(42.68%)。与非AKI患者相比, AKI患者年龄平均大3岁, 女性较少(40.73% vs 43.29%)(表1)。从表1中我们发现AKI患者基础肾功能差, 格拉斯哥昏迷评分(GCS)加重, 更常伴发糖尿病、肝细胞功能障碍。

4.1 XGBoost 模型

在本研究中, XGBoost是在13001例(70%)随机选择的数据上进行训练的, 其余的6000例(30%)用于模型测试。图2为XGBoost模型五折交叉验证的AUC结果。我们可以看到AUC在0.828和0.844之间变化。总的来说, 这些指标明显偏高, 说明XGBoost可以用于AKI风险检测的建模。XGBoost最终在30%的测试集上测试, 最终得到的性能指标为: 准确性为0.816, AUC为0.839, 再次表明该模型性能稳定, 表现显著。XGBoost稳定的高AUC(高灵敏度和高特异性)在其他研究中得到了进一步的检验和验证^[56]: 如Zimmerman等人^[36]认为大多数预测模型的AUC在0.75左右。此外, 与Gong等人^[27], Zimmerman等人^[36]和Li等人^[57]的工作

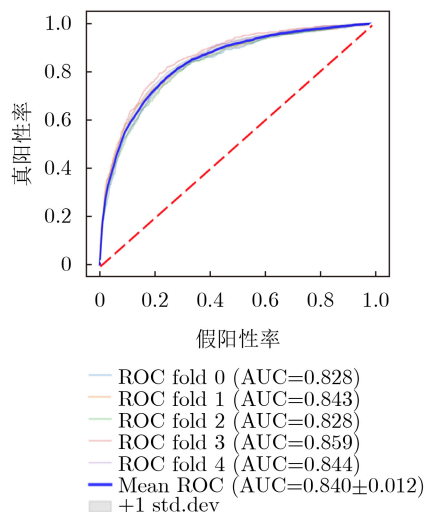


图2 XGBoost模型5折交叉验证的AUC值结果

相比，本文的模型获得了更好的AUC结果(Gong等人：0.781；Zimmerman等人：0.783；Li等人：0.779)。并且其中两项研究没有进一步研究各特征向量间的详细关系并提供对应的模型解释。

此外，本文还比较了XGBoost与其他4种流行的ML算法的性能，结果见表3(各模型五倍交叉验证的平均值；括号中的值为95%置信区间)和图3。从测试的5种算法中，XGBoost表现最好。实验结果也表明，其余4种模型各有优缺点。总体而言，集成学习算法的性能优于基分类器，特别是在灵敏度方面。这说明基础算法不适合解决这类复杂的分类问题。

4.2 特征重要性分析

除对AKI预测模型的预测能力进行评估外，本文还进一步分析了训练模型提供的关于特征重要性的信息。根据ICU常用的Apache, SAPS等风险预测评分系统，结合临床专家的意见，从MIMIC III数据库中提取了与AKI发生、发展高度相关的特征，并在每个患者入院后最初24小时内取平均值。最终共筛选出21个变量组成基本数据集，包括年龄、性别、生命体征和临床实验室结果。

图4显示了可引起AKI的前20个结构化临床指标。根据结果，入院当天基线Scr在AKI风险预测中发挥了关键作用。总尿量(Total Volume Urine in

24 hours, TVU)是影响肾功能最重要的变量之一，其次是动脉血氧分压(partial Pressure of Oxygen, PO₂)、体温和血尿素氮(Blood Urea Nitrogen, BUN)。预测AKI的关键信息集中在权重排名前5位的预测因子中。这一结果符合AKI病理生理学发展^[58]，即Scr、尿量和BUN与AKI发生高度相关。PO₂和体温也是AKI的危险因素，表示这两个指标的异常可能增加ICU患者罹患AKI的概率。我们希望这些发现的关键变量能够引起重视，进而为协助临床医生更快地判断ICU患者发生AKI的风险，缩小临床检查和检测的范围。

4.3 解释模型结果

为了进一步明确各指标相对于目标变量的正/负关系，本文使用SHAP的特殊变体，TreeSHAP用于模型解释。图5为AKI风险预测模型特征重要性排序的SHAP摘要图。该摘要图结合了特征重要度和特征的影响。摘要图上的每个点都是一个特征和一个实例的Shapley值，y轴上的位置由特征决定，x轴上的位置由Shapley值决定，颜色代表特征值从小到大，重叠点在y轴方向上浮，因此我们可以了解每个特征的Shapley值的分布。

我们可以看到入ICU 24h内首次SCr测量值(Scr_baseline)对模型的影响最大。此外，该特征值越高，Shapley值越高，发生AKI的概率就越

表 3 5种AKI风险预测模型的性能比较

模型	准确率[95% CI]	精确率[95% CI]	敏感度[95% CI]	F1[95% CI]
SVM	0.793[0.789, 0.798]	0.781[0.772, 0.789]	0.793[0.789, 0.798]	0.747[0.741, 0.752]
LR	0.788[0.785, 0.793]	0.765[0.759, 0.771]	0.789[0.785, 0.793]	0.762[0.758, 0.766]
RF	0.824[0.813, 0.835]	0.809[0.803, 0.816]	0.822[0.817, 0.827]	0.807[0.796, 0.818]
ANN	0.806[0.799, 0.815]	0.808[0.797, 0.819]	0.821[0.812, 0.829]	0.806[0.797, 0.814]
XGBoost	0.824[0.815, 0.832]	0.812[0.802, 0.822]	0.824[0.815, 0.832]	0.813[0.804, 0.822]

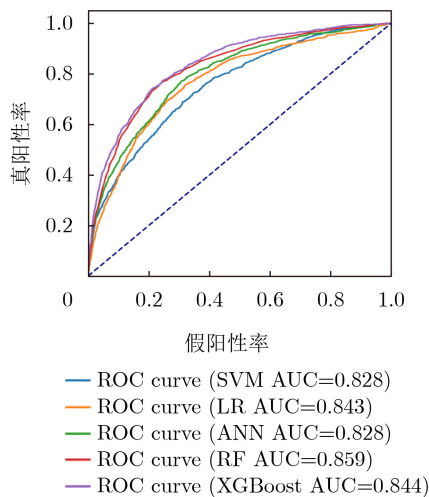


图 3 由测试集计算得到的各模型AUC结果

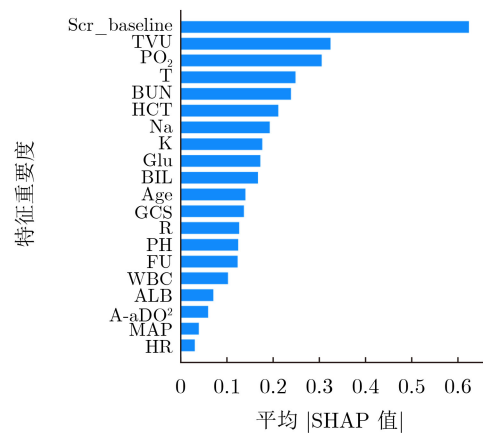


图 4 预测模型中前20个变量的排名

大。总尿量是第2个最重要的特征，该特征值越低，发生AKI的概率就越高，与GCS相同。在生命体征特征方面，图5也显示较高的体温和呼吸频率的患者发生AKI的概率增高。此外，尽管胆红素(Bilirubin, BIL)的重要性明显低于上述特征，但它对部分患者而言影响最大，且对AKI的发生有积极的影响。此外，在实验室检查方面，红细胞比容(Hematocrit, HCT)和白蛋白(Albumin, ALB)越高越容易发生AKI，但它们的重要性也明显较低。上述实验结果与临床及相关研究结果一致^[59]，进一步证实了GCS, BIL, HCT和相关生命体征对AKI风险的影响。尽管图5的摘要图可以简要说明特征对AKI风险的正面或负面影响，但信息仍然有限。为了帮助医生更好地了解各变量之间的交互效应，有待进一步研究各预测指标之间的关系。图6给出了两个SHAP特征依赖图的例子。

如图6(a)所示，我们选择TVU作为特征来确定Scr从0.5增加到3.5时的影响。红色的点代表Scr的高值，蓝色的点代表低值。当TVU较低时，Scr的SHAP值会增加，说明增加TVU会降低AKI发生的可能性。即降低TVU，增加Scr会导致AKI的概率增加。而Scr较低时的SHAP值小于零，说明增加TVU而Scr较低时，AKI发生的概率较低。

图6(b)显示了BUN和Scr对预测AKI带来的影响。忽略图形的颜色，关注BUN变化对模型输出的影响。尽管存在噪声，但在BUN值为25之前，SHAP值大多为负值。在25~75范围内，Shapley值逐渐增大并为正，当BUN约为150时，SHAP值再次逐渐减小并为负。BUN的高度非线性影响证明了只看一个参数(例如，它是正的还是负的)往往是不够的。从本质上看，SHAP值表明BUN的主要影响为正；即当BUN升高超过25时，AKI风险增加。在BUN较高时，红点出现越多；说明Scr和BUN呈正相关。从图6可以看出，TVU降低，BUN升高，说明AKI风险增加。一些研究报道BUN和尿

量是AKI、终末期肾脏疾病和死亡率的重要指标^[60,61]。本研究给出的结果也验证了这一结论。

值得注意的是，这些依赖图说明的是相关性，而不是因果关系。基本上，它们提供了特定范围下的信息，可以提醒特定环境下可能出现的结果变化趋势。因此有必要将这些信息与医生的经验和患者的病情相结合，以确定该特征是否可以作为干预的选择。

最后，为了评估本文中的XGBoost模型，我们检索了利用MIMIC III数据集进行AKI预测的最新研究，并进行了对比。如表4所示，本研究获得了优越的性能，这说明我们提出的模型非常有效。

5 讨论

在本研究中，使用从MIMIC III数据库中提取的一组数据，训练XGBoost模型对AKI的早期预测进行建模。共纳入13001个病例进行模型训练，准确率为0.816，AUC为0.839。此外，模型训练中AUC值(即五倍交叉验证)也证实了该算法在预测AKI方面的高性能。该结果优于文献中报道的基线模型^[36]。本研究还构建了基于RF, ANN, SVM和LR算法的AKI风险预测模型。将XGBoost模型与

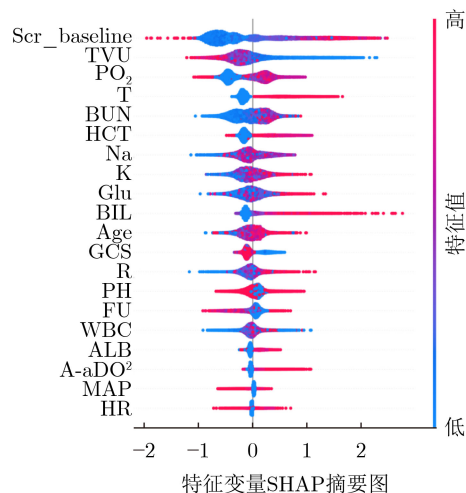
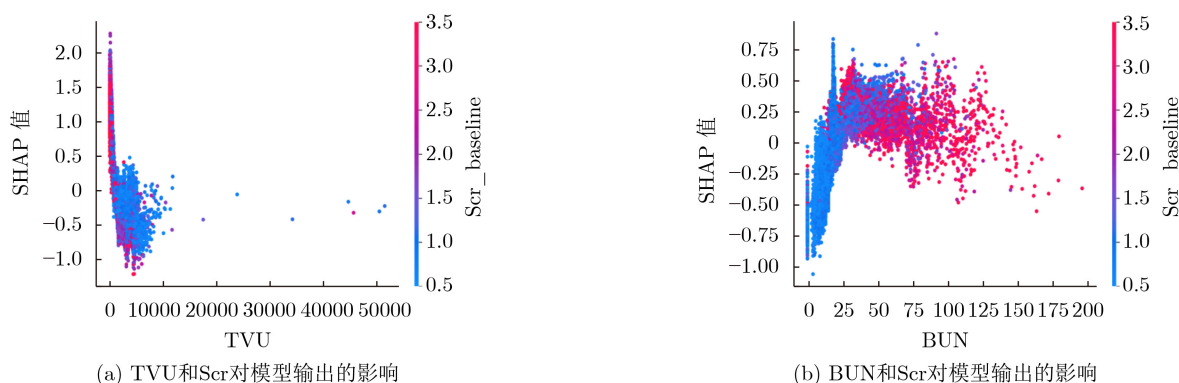


图5 平均排名前20的特征变量SHAP摘要图



(a) TVU和Scr对模型输出的影响

(b) BUN和Scr对模型输出的影响

图6 重要指标的SHAP特征依赖图示例

表4 MIMIC-III数据集上的最新研究（仅用于AKI预测）不同模型性能的比较

研究	记录	ML 算法	条件	AUC
Mohamadlou等人 ^[30]	48582	Gradient Boosting	ICU patients	0.728-0.761
Zimmerman等人 ^[36]	23950	Logistic Regression	ICU patients	0.78
Li等人 ^[57]	16560	Knowledge-guided CNN	ICU patients	0.779
Shawwa 等人 ^[62]	18529	Gradient Boosting	ICU patients	0.656
Sun 等人 ^[63]	16558	Mixed-feature CNN	ICU patients	0.83
本文方法	19001	XGBoost	ICU patients	0.839

上述模型的性能进行比较,结果表明,在所有评价指标中,XGBoost模型的预测性能最好。相比之下,支持向量机在所有5种模型中表现出了最弱的性能。

由于支持向量机模型等基础算法背后的数学逻辑对临床医生来说很难理解,除了预测结果之外,它不能给临床医生提供更多的信息。然而,作为一种基于决策树的集成机器学习算法,XGBoost的另一个好处是它不受多重共线性的影响。因此,即使两个变量在模型中的作用相同,我们也可以保留两者,因为我们可以通过SHAP进行特征重要性分析。因此,我们认为XGBoost模型具有良好的模型性能和临床可解释性,在AKI风险预测方面具有很大的潜力。

利用SHAP对最终的XGBoost模型进行特征重要性分析,发现肾功能相关特征(尤其是Scr)对模型中AKI发生的概率有实质性影响。相比之下,心率是除性别外最不重要的特征,我们注意到数据集中缺失了大量的心率数据(83.4%),这可能是导致该结果的原因之一。该模型的结果也可以为临床决策提供信息。以温度为例,温度对AKI的发生概率有积极的影响,可能对高温患者采取更多的降温措施可以缓解AKI的发生。血钾水平升高可能是肾脏损伤早期电解质紊乱的表现^[64]。而Glu的特征表明糖尿病共病和程序风险可能导致AKI发病,与文献一致^[65]。特征变量重要性排序结果证实,临床记录中有临床意义的关键词可以用于预测AKI,我们所建立的模型确实捕捉到了这些特征变量。进一步进行SHAP依赖分析,获取并描述两对特征对模型的影响(图6),该方法不仅能够评价特征对模型输出影响的重要性的方向,还可以提取特征对模型输出的复杂非线性影响^[66]。值得注意的是,在这项工作中,TVU和BUN的交互影响提供了有意义的信息,这是大多数其他技术无法捕捉到的。

我们的研究还有以下几个优势。首先,本研究使用的数据库是一个公开的、大规模的ICU数据库,样本量充足。其次,我们采用XGBoost算法构

建预测模型,该模型能够容纳大量特征变量的数据,并对影响AKI预测的变量进行重要性排序。第三,本文采用了五倍交叉验证,以获得更性能更稳定的模型。第四,我们所使用的预测变量可以在临床实践中得到,这就保证了该模型的临床可用性。最后,本模型具有临床可解释性,预测模型的可靠性,且可以帮助医生更好地了解各变量与目标变量之间及两变量相互之间的交互效应。

这项回顾性研究仍有一些局限性。首先,本文使用的数据是从单中心数据库中提取的。因此,本研究获得的结论仍需进一步的外部验证,以进行泛化和前瞻性试验,评估其临床效用。其次,有许多因素可导致或促成AKI,包括有效循环血量不足、肾毒性药物和败血症^[5,67]。然而,本研究使用的数据集仅包含了部分在MIMIC III数据库中定义的结构化信息,而其他急性肾损伤危险因素,如脓毒症、烧伤、创伤、心脏手术、肾毒性药物等^[3],也可能有助于提升我们模型的性能。第三,由于我们没有获得ICU入院前的SCr数据,因此只能以ICU入院后第1次SCr检测值作为基线。最后,本研究中发现的重要特征和拐点可作为AKI风险的早期体征,但在其推荐范围是否可作为对照的参考还需进一步验证。

6 结论

由于AKI与不良临床结果和额外的医疗资源消耗相关,本文开发了一个高性能的XGBoost预测模型,通过使用入ICU后的前24 h数据,早期预测所有危重成人患者患AKI的风险。虽然基于XGBoost算法的AKI风险预测模型的临床适用性还需要在实际的临床实践中进行检验,但我们认为XGBoost模型由于其性能和临床可解释性,在未来ICU的临床工作中具有很大应用价值,有助于ICU临床医师避免延误高危AKI患者的治疗,这对改善AKI患者的预后至关重要。

参考文献

- [1] LUNDBERG S M and LEE S I. A unified approach to

- interpreting model predictions[C]. Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, USA, 2017: 4765–4774.
- [2] ZHOU Jiaojiao, LIU Yun, TANG Yi, *et al.* A comparison of RIFLE, AKIN, KDIGO, and Cys-C criteria for the definition of acute kidney injury in critically ill patients[J]. *International Urology and Nephrology*, 2016, 48(1): 125–132. doi: [10.1007/s11255-015-1150-6](https://doi.org/10.1007/s11255-015-1150-6).
- [3] KELLUM J A, 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 International Supplements*, 2012, 2(1): 1–138. doi: [10.1038/kisup.2012.1](https://doi.org/10.1038/kisup.2012.1).
- [4] COCA S G, YUSUF B, SHLIPAK M G, *et al.* Long-term risk of mortality and other adverse outcomes after acute kidney injury: A systematic review and meta-analysis[J]. *American Journal of Kidney Diseases*, 2009, 53(6): 961–973. doi: [10.1053/j.ajkd.2008.11.034](https://doi.org/10.1053/j.ajkd.2008.11.034).
- [5] CHERTOW G M, BURDICK E, HONOUR M, *et al.* Acute kidney injury, mortality, length of stay, and costs in hospitalized patients[J]. *Journal of the American Society of Nephrology*, 2005, 16(11): 3365–3370. doi: [10.1681/ASN.2004090740](https://doi.org/10.1681/ASN.2004090740).
- [6] CHAWLA L S, BELLOMO R, BIHORAC A, *et al.* Acute kidney disease and renal recovery: Consensus report of the Acute Disease Quality Initiative (ADQI) 16 Workgroup[J]. *Nature Reviews Nephrology*, 2017, 13(4): 241–257. doi: [10.1038/nrneph.2017.2](https://doi.org/10.1038/nrneph.2017.2).
- [7] KELLUM J A and PROWLE J R. Paradigms of acute kidney injury in the intensive care setting[J]. *Nature Reviews Nephrology*, 2018, 14(4): 217–230. doi: [10.1038/nrneph.2017.184](https://doi.org/10.1038/nrneph.2017.184).
- [8] HOSTE E A J and SCHURGERS M. Epidemiology of acute kidney injury: How big is the problem?[J]. *Critical Care Medicine*, 2008, 36(4): S146–S151.
- [9] ZENG Xiaoxi, MCMAHON G M, BRUNELLI S M, *et al.* Incidence, outcomes, and comparisons across definitions of AKI in hospitalized individuals[J]. *Clinical Journal of the American Society of Nephrology*, 2014, 9(1): 12–20. doi: [10.2215/CJN.02730313](https://doi.org/10.2215/CJN.02730313).
- [10] YANG Li, XING Guolan, WANG Li, *et al.* Acute kidney injury in China: A cross-sectional survey[J]. *The Lancet*, 2015, 386(10002): 1465–1471. doi: [10.1016/S0140-6736\(15\)00344-X](https://doi.org/10.1016/S0140-6736(15)00344-X).
- [11] LIBÓRIO A B, LEITE T T, DE OLIVEIRA NEVES F M, *et al.* AKI complications in critically ill patients: Association with mortality rates and RRT[J]. *Clinical Journal of the American Society of Nephrology*, 2015, 10(1): 21–28. doi: [10.2215/CJN.04750514](https://doi.org/10.2215/CJN.04750514).
- [12] DE OLIVEIRA MARQUES F, OLIVEIRA S A, DE LIMA E SOUZA P F, *et al.* Kinetic estimated glomerular filtration rate in critically ill patients: Beyond the acute kidney injury severity classification system[J]. *Critical Care*, 2017, 21(1): 280. doi: [10.1186/s13054-017-1873-0](https://doi.org/10.1186/s13054-017-1873-0).
- [13] SUSANTITAPHONG P, CRUZ D N, CERDA J, *et al.* World incidence of AKI: A meta-analysis[J]. *Clinical Journal of the American Society of Nephrology*, 2013, 8(9): 1482–1493. doi: [10.2215/CJN.00710113](https://doi.org/10.2215/CJN.00710113).
- [14] SRISAWAT N, SILEANU F E, MURUGAN R, *et al.* Variation in risk and mortality of acute kidney injury in critically ill patients: A multicenter study[J]. *American Journal of Nephrology*, 2015, 41(1): 81–88. doi: [10.1159/000371748](https://doi.org/10.1159/000371748).
- [15] DASTA J F and KANE-GILL S. Review of the literature on the costs associated with acute kidney injury[J]. *Journal of Pharmacy Practice*, 2019, 32(3): 292–302. doi: [10.1177/0897190019852556](https://doi.org/10.1177/0897190019852556).
- [16] OSTERMANN M, CHANG R, and The Riyadh ICU Program Users Group. Correlation between the AKI classification and outcome[J]. *Critical Care*, 2008, 12(6): R144. doi: [10.1186/cc7123](https://doi.org/10.1186/cc7123).
- [17] THAKAR C V, CHRISTIANSON A, FREYBERG R, *et al.* Incidence and outcomes of acute kidney injury in intensive care units: A veterans administration study[J]. *Critical Care Medicine*, 2009, 37(9): 2552–2558. doi: [10.1097/CCM.0b013e3181a5906f](https://doi.org/10.1097/CCM.0b013e3181a5906f).
- [18] SAWHNEY S, MARKS A, FLUCK N, *et al.* Intermediate and long-term outcomes of survivors of acute kidney injury episodes: A large population-based cohort study[J]. *American Journal of Kidney Diseases*, 2017, 69(1): 18–28. doi: [10.1053/j.ajkd.2016.05.018](https://doi.org/10.1053/j.ajkd.2016.05.018).
- [19] WANG J H, JOSLIN J, JENKINS R, *et al.* Outcomes of elderly patients with acute kidney injury on a renal high dependency unit[J]. *International Journal of Clinical Practice*, 2015, 69(10): 1209–1210. doi: [10.1111/ijcp.12683](https://doi.org/10.1111/ijcp.12683).
- [20] WU V C, WU C H, HUANG T M, *et al.* Long-term risk of coronary events after AKI[J]. *Journal of the American Society of Nephrology*, 2014, 25(3): 595–605. doi: [10.1681/ASN.2013060610](https://doi.org/10.1681/ASN.2013060610).
- [21] VILLENEUVE P M, CLARK E G, SIKORA L, *et al.* Health-related quality-of-life among survivors of acute kidney injury in the intensive care unit: A systematic review[J]. *Intensive Care Medicine*, 2016, 42(2): 137–146. doi: [10.1007/s00134-015-4151-0](https://doi.org/10.1007/s00134-015-4151-0).
- [22] LIN Ke, HU Yonghua, and KONG Guilan. Predicting in-hospital mortality of patients with acute kidney injury in the ICU using random forest model[J]. *International Journal of Medical Informatics*, 2019, 125: 55–61. doi: [10.1016/j.ijmedinf.2019.02.002](https://doi.org/10.1016/j.ijmedinf.2019.02.002).
- [23] PONCE D, DE PIETRO FRANCO ZORZENON C, DOS SANTOS N Y, *et al.* Early nephrology consultation can have an impact on outcome of acute kidney injury patients[J]. *Nephrology Dialysis Transplantation*, 2011, 26(10): 3202–3206. doi: [10.1093/ndt/gfr359](https://doi.org/10.1093/ndt/gfr359).

- [24] FLECHET M, FALINI S, BONETTI C, *et al.* Machine learning versus physicians' prediction of acute kidney injury in critically ill adults: A prospective evaluation of the AKIpredictor[J]. *Critical Care*, 2019, 23(1): 282. doi: [10.1186/s13054-019-2563-x](https://doi.org/10.1186/s13054-019-2563-x).
- [25] ANDREU-PEREZ J, POON C C Y, MERRIFIELD R D, *et al.* Big data for health[J]. *IEEE Journal of Biomedical and Health Informatics*, 2015, 19(4): 1193–1208. doi: [10.1109/JBHI.2015.2450362](https://doi.org/10.1109/JBHI.2015.2450362).
- [26] XIAO Cao, CHOI E, and SUN Jimeng. Opportunities and challenges in developing deep learning models using electronic health records data: A systematic review[J]. *Journal of the American Medical Informatics Association*, 2018, 25(10): 1419–1428. doi: [10.1093/jamia/ocy068](https://doi.org/10.1093/jamia/ocy068).
- [27] GONG Kaidi, LEE H K, YU Kaiye, *et al.* A prediction and interpretation framework of acute kidney injury in critical care[J]. *Journal of Biomedical Informatics*, 2021, 113: 103653. doi: [10.1016/j.jbi.2020.103653](https://doi.org/10.1016/j.jbi.2020.103653).
- [28] TOMAŠEV N, GLOROT X, RAE J W, *et al.* A clinically applicable approach to continuous prediction of future acute kidney injury[J]. *Nature*, 2019, 572(7767): 116–119. doi: [10.1038/s41586-019-1390-1](https://doi.org/10.1038/s41586-019-1390-1).
- [29] FLECHET M, GÜIZA 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 Medicine*, 2017, 43(6): 764–773. doi: [10.1007/s00134-017-4678-3](https://doi.org/10.1007/s00134-017-4678-3).
- [30] 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]. *Canadian Journal of Kidney Health and Disease*, 2018, 5: 1–9.
- [31] LEI V J, LUONG T, SHAN E, *et al.* Risk stratification for postoperative acute kidney injury in major noncardiac surgery using preoperative and intraoperative data[J]. *JAMA Network Open*, 2019, 2(12): e1916921. doi: [10.1001/jamanetworkopen.2019.16921](https://doi.org/10.1001/jamanetworkopen.2019.16921).
- [32] KOYNER J L, CAREY K A, EDELSON D P, *et al.* The development of a machine learning inpatient acute kidney injury prediction model[J]. *Critical Care Medicine*, 2018, 46(7): 1070–1077. doi: [10.1097/CCM.0000000000003123](https://doi.org/10.1097/CCM.0000000000003123).
- [33] CHURPEK M M, CAREY K A, EDELSON D P, *et al.* Internal and external validation of a machine learning risk score for acute kidney injury[J]. *JAMA Network Open*, 2020, 3(8): e2012892. doi: [10.1001/jamanetworkopen.2020.12892](https://doi.org/10.1001/jamanetworkopen.2020.12892).
- [34] SIMONOV M, UGWUOWO U, MOREIRA E, *et al.* A simple real-time model for predicting acute kidney injury in hospitalized patients in the US: A descriptive modeling study[J]. *PLoS Medicine*, 2019, 16(7): e1002861. doi: [10.1371/journal.pmed.1002861](https://doi.org/10.1371/journal.pmed.1002861).
- [35] XU Zhenxing, LUO Yuan, ADEKKANATTU P, *et al.* Stratified mortality prediction of patients with acute kidney injury in critical care[J]. *Studies in Health Technology and Informatics*, 2019, 264: 462–466.
- [36] ZIMMERMAN L P, REYFMAN P A, SMITH A D R, *et al.* Early prediction of acute kidney injury following ICU admission using a multivariate panel of physiological measurements[J]. *BMC Medical Informatics and Decision Making*, 2019, 19(S1): 16. doi: [10.1186/s12911-019-0733-z](https://doi.org/10.1186/s12911-019-0733-z).
- [37] RASHIDI H H, SEN S, PALMIERI T L, *et al.* Early recognition of burn- and trauma-related acute kidney injury: A pilot comparison of machine learning techniques[J]. *Scientific Reports*, 2020, 10(1): 205. doi: [10.1038/s41598-019-57083-6](https://doi.org/10.1038/s41598-019-57083-6).
- [38] JOHNSON A E W, POLLARD T J, SHEN Lu, *et al.* MIMIC-III, a freely accessible critical care database[J]. *Scientific Data*, 2016, 3: 160035. doi: [10.1038/sdata.2016.35](https://doi.org/10.1038/sdata.2016.35).
- [39] RALIB A, PICKERING J W, SHAW G M, *et al.* The urine output definition of acute kidney injury is too liberal[J]. *Critical Care*, 2013, 17(3): R112. doi: [10.1186/cc12784](https://doi.org/10.1186/cc12784).
- [40] RAZAVIAN N and SONTAG D. Temporal convolutional neural networks for diagnosis from lab tests[J]. arXiv preprint arXiv: 1511.07938, 2015.
- [41] HAN Hui, WANG Wenyuan, and MAO Binghuan. Borderline-SMOTE: A new over-sampling method in imbalanced data sets learning[C]. 2005 International Conference on Intelligent Computing, Hefei, China, 2005: 878–887.
- [42] CHAWLA N V, BOWYER K W, HALL L O, *et al.* SMOTE: Synthetic minority over-sampling technique[J]. *Journal of Artificial Intelligence Research*, 2002, 16: 321–357. doi: [10.1613/jair.953](https://doi.org/10.1613/jair.953).
- [43] CHEN Tianqi and GUESTRIN C. Xgboost: A scalable tree boosting system[C]. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, USA, 2016: 785–794.
- [44] NGUYEN H, BUI X N, BUI H B, *et al.* Developing an XGBoost model to predict blast-induced peak particle velocity in an open-pit mine: A case study[J]. *Acta Geophysica*, 2019, 67(2): 477–490. doi: [10.1007/s11600-019-00268-4](https://doi.org/10.1007/s11600-019-00268-4).
- [45] MA Xiaojun, SHA Jinglan, WANG Dehua, *et al.* Study on a prediction of P2P network loan default based on the machine learning LightGBM and XGboost algorithms according to different high dimensional data cleaning[J]. *Electronic Commerce Research and Applications*, 2018, 31: 24–39. doi: [10.1016/j.elerap.2018.08.002](https://doi.org/10.1016/j.elerap.2018.08.002).
- [46] MOUSA S R, BAKHIT P R, and ISHAK S. An extreme gradient boosting method for identifying the factors contributing to crash/near-crash events: A naturalistic driving study[J]. *Canadian Journal of Civil Engineering*, 2019, 46(8): 712–721. doi: [10.1139/cjce-2018-0117](https://doi.org/10.1139/cjce-2018-0117).
- [47] SYARIF I, PRUGEL-BENNETT A, WILLS G. SVM

- parameter optimization using grid search and genetic algorithm to improve classification performance[J]. *Telkomnika*, 2016, 14(4): 1502–1509. doi: [10.12928/telkomnika.v14i4.3956](https://doi.org/10.12928/telkomnika.v14i4.3956).
- [48] ARLOT S and LERASLE M. Choice of V for V-fold cross-validation in least-squares density estimation[J]. *The Journal of Machine Learning Research*, 2016, 17(1): 7256–7305.
- [49] HANLEY J A and MCNEIL B J. The meaning and use of the area under a receiver operating characteristic (ROC) curve[J]. *Radiology*, 1982, 143(1): 29–36. doi: [10.1148/radiology.143.1.7063747](https://doi.org/10.1148/radiology.143.1.7063747).
- [50] FLACH P, HERNÁNDEZ-ORALLO J, and FERRI C. A coherent interpretation of AUC as a measure of aggregated classification performance[C]. Proceedings of the 28th International Conference on International Conference on Machine Learning, Bellevue, USA, 2011.
- [51] WILSON F P. Machine learning to predict acute kidney injury[J]. *American Journal of Kidney Diseases*, 2020, 75(6): 965–967. doi: [10.1053/j.ajkd.2019.08.010](https://doi.org/10.1053/j.ajkd.2019.08.010).
- [52] ŠTRUMBELJ E and KONONENKO I. Explaining prediction models and individual predictions with feature contributions[J]. *Knowledge and Information Systems*, 2014, 41(3): 647–665. doi: [10.1007/s10115-013-0679-x](https://doi.org/10.1007/s10115-013-0679-x).
- [53] RIBEIRO M T, SINGH S, and GUESTRIN C. "Why should I trust you?": Explaining the predictions of any classifier[C]. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, USA, 2016: 1135–1144.
- [54] LUNDBERG S M, NAIR B, VAVILALA M S, *et al.* Explainable machine-learning predictions for the prevention of hypoxaemia during surgery[J]. *Nature Biomedical Engineering*, 2018, 2(10): 749–760. doi: [10.1038/s41551-018-0304-0](https://doi.org/10.1038/s41551-018-0304-0).
- [55] LUNDBERG S M, ERION G, CHEN H, *et al.* From local explanations to global understanding with explainable AI for trees[J]. *Nature Machine Intelligence*, 2020, 2(1): 56–67. doi: [10.1038/s42256-019-0138-9](https://doi.org/10.1038/s42256-019-0138-9).
- [56] PARSA A B, MOVAHEDI A, TAGHIPOUR H, *et al.* Toward safer highways, application of XGBoost and SHAP for real-time accident detection and feature analysis[J]. *Accident Analysis & Prevention*, 2020, 136: 105405.
- [57] LI Yikuan, YAO Liang, MAO Chengsheng, *et al.* Early prediction of acute kidney injury in critical care setting using clinical notes[C]. Proceedings of 2018 IEEE International Conference on Bioinformatics and Biomedicine, Madrid, Spain, 2018: 683–686.
- [58] BASILE D P, ANDERSON M D, and SUTTON T A. Pathophysiology of acute kidney injury[J]. *Comprehensive Physiology*, 2012, 2(2): 1303–1353.
- [59] SOVIK S, ISACHSEN M S, NORDHUUS K M, *et al.* Acute kidney injury in trauma patients admitted to the ICU: A systematic review and meta-analysis[J]. *Intensive Care Medicine*, 2019, 45(4): 407–419. doi: [10.1007/s00134-019-05535-y](https://doi.org/10.1007/s00134-019-05535-y).
- [60] OH H J, SHIN D H, LEE M J, *et al.* Urine output is associated with prognosis in patients with acute kidney injury requiring continuous renal replacement therapy[J]. *Journal of Critical Care*, 2013, 28(4): 379–388. doi: [10.1016/j.jcrc.2012.11.019](https://doi.org/10.1016/j.jcrc.2012.11.019).
- [61] HUSSEIN R H, CALICE-SILVA V, EVANS R, *et al.* Diagnosis of acute kidney injury in children hospitalized in a sub-saharan african unit by saliva urea nitrogen dipstick test[J]. *Blood Purification*, 2020, 49(1/2): 185–196.
- [62] SHAWWA K, GHOSH E, LANIUS S, *et al.* Predicting acute kidney injury in critically ill patients using comorbid conditions utilizing machine learning[J]. *Clinical Kidney Journal*, 2021, 14(5): 1428–1435. doi: [10.1093/ckj/sfaa145](https://doi.org/10.1093/ckj/sfaa145).
- [63] SUN Mengxin, BARON J, DIGHE A, *et al.* Early prediction of acute kidney injury in critical care setting using clinical notes and structured multivariate physiological measurements[J]. *Studies in Health Technology and Informatics*, 2019, 264: 368–372.
- [64] CHEN Daonan, DU Jiang, XIE Yun, *et al.* Relationship between early serum sodium and potassium levels and AKI severity and prognosis in oliguric AKI patients[J]. *International Urology and Nephrology*, 2021, 53(6): 1171–1187. doi: [10.1007/s11255-020-02724-3](https://doi.org/10.1007/s11255-020-02724-3).
- [65] HURSH B E, RONSLEY R, ISLAM N, *et al.* Acute kidney injury in children with type 1 diabetes hospitalized for diabetic ketoacidosis[J]. *JAMA Pediatrics*, 2017, 171(5): e170020. doi: [10.1001/jamapediatrics.2017.0020](https://doi.org/10.1001/jamapediatrics.2017.0020).
- [66] KUMAR I E, VENKATASUBRAMANIAN S, SCHEIDEGGER C, *et al.* Problems with shapley-value-based explanations as feature importance measures[C]. Proceedings of the 37th International Conference on Machine Learning, 2020: 5491–5500.
- [67] BELLOMO R, KELLUM J A, and RONCO C. Acute kidney injury[J]. *The Lancet*, 2012, 380(9843): 756–766. doi: [10.1016/S0140-6736\(11\)61454-2](https://doi.org/10.1016/S0140-6736(11)61454-2).
- 罗 妍: 女, 1988年生, 博士生, 研究方向为智能信息处理、医疗数据挖掘与分析。
- 王 枫: 女, 1958年生, 教授, 研究方向为智能信息处理、医疗数据挖掘与分析。
- 叶文玲: 女, 1969年生, 主任医师, 研究方向为肾脏疾病的诊断和治疗、医疗数据处理。