Scientific Reports

https://doi.org/10.1038/s41598-026-37828-w

Article in Press

Application of XGBoost and logistic regression in predicting 90 days mortality for elderly severe acute renal failure patients

Received: 25 November 2025

Accepted: 27 January 2026

Published online: 03 February 2026

Cite this article as: Zeng J., Zhu Y., Ye F. *et al.* Application of XGBoost and logistic regression in predicting 90 days mortality for elderly severe acute renal failure patients. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-37828-w

Jinping Zeng, Yiying Zhu, Feng Ye, Qin Song, Haiyan Ma, Jun Yang & Yinyin Wu

Application of XGBoost and logistic regression in predicting 90 days mortality for elderly severe acute renal failure patients

Jinping Zeng[1], MPH; Yiying Zhu[1], MPH; Feng Ye[1], MPH; Qin Song[2], PhD; Haiyan Ma[1], PhD; Jun Yang[3*], PhD; and Yinyin Wu[1*], PhD

1 Department of Epidemiology and Health Statistics, School of Public Health and Nursing, Hangzhou Normal University, Hangzhou, China
2 Department of Occupational and Environmental Health, School of Public Health and Nursing, Hangzhou Normal University, Hangzhou, China
3 Department of Nutrition and Toxicology, School of Public Health and Nursing, Hangzhou Normal University, Hangzhou, China
*Correspondence: gastate@zju.edu.cn (Jun Yang) and 20110106@hznu.edu.cn (Yinyin Wu)

Abstract

Background: Acute renal failure (ARF) is one of the most common conditions encountered in the intensive care unit (ICU). ARF has a complex pathogenesis and due to the progressive weakening of the structure and function of the kidney, the incidence of ARF increases significantly in the aging group. Therefore, the development of reliable predictive model is of great importance to identify those patients in high risk for ARF, in order to provide timely and effective interventions to improve their prognosis.

Objective: Extreme gradient boosting (XGBoost) is an efficient integrated learning algorithm with advantages over traditional logistic regression (LR) methods. The purpose of this study was to compare the performance of the two models in predicting 90-day mortality in elderly patients with ARF.

Methods: Data of elderly patients (>60years) with ARF in ICU

were extracted from MIMIC IV with 90-day mortality as end-point. The performance of the two predictive models was tested and compared by receiver operating characteristic curve and decision curve analysis (DCA). Cumulative residual distribution plot and residual box-plot were then used to determine the fit of the model. Finally, the model with better overall diagnostic value was selected and a breakdown plot was drawn.

Results: Data of 7,500 elderly ARF patients were analyzed, of whom 1,150 died within 90 days. Both models showed good discriminatory ability, but the XGBoost model had a larger area under the curve value. DCA results revealed that the net benefit of the XGBoost model had a greater range than the LR model. Moreover, the XGBoost model had the smallest sample residuals and root-mean-square residuals, indicating a better fitting of the XGBoost algorithm. Finally, a breakdown plot based on the XGBoost model was created as an individualized tool for prognosis prediction in elderly patients with ARF.

Conclusions: Our study find that the XGBoost algorithm model was a better model for predicting 90-day mortality in elderly ICU patients with ARF compared to the LR model. The model may have clinical applications for elderly patients with ARF and may help healthcare professionals to develop detailed treatment plans as well as provide accurate care.

Keywords: Acute renal failure; 90-day mortality; XGBoost; Logistic regression; The elderly

Introduction

Acute renal failure (ARF) is a common syndrome characterized

by high morbidity, high mortality and poor prognosis [1]. It is mainly caused by impaired renal function due to renal ischemia or obstruction, as well as the consequence of certain types of nephrotoxic drugs [2]. ARF was defined using the KDIGO criteria, requiring meeting any of the following within 48 hours: (1) a serum creatinine increase ≥0.3 mg/dl, (2) a 1.5-fold elevation from baseline, or (3) urine output <0.5 ml/kg/h for 6 or more hours. ARF can lead to rapid decline of renal function, causing symptoms such as azotemia, water-electrolyte disturbances, and acid-base imbalance [3]. In the elderly, ARF is more likely to progress to multi-organ failure with a mortality rate as high as 70% due to the decline of body functions and the combination of multiple organ pathologies [4]. Even worse, ARF patients can only slow down the progression of the disease and improve the quality of life through long-term medication or dialysis treatment [5].

ARF is also one of the most common disorders with a prevalence of 10-15% in the intensive care unit (ICU) [6], and as high as 50-60% in critically ill patients [7], thus it is a serious global health problem that we are facing [8]. On the other hand, the incidence of ARF has increased significantly in the aging population, probably due to the progressive weakening of the structure and function of the human kidney [9]. A survey showed that 52.2% of all in-hospital ARF patients were over 60 years of age [10]. In recent years, despite improvements in the diagnosis and treatment of the disease, the mortality rate of ARF has not been significantly reduced [11], and the burden of disease caused by ARF remains high, especially in ICU [12]. Consequently, early recognition and diagnosis of ARF is extremely important. Therefore, the development of reliable predictive models is particularly urgent to identify those patients at risk and provide

them with timely and effective interventions to improve their prognosis.

In recent years, the predictive performance of machine learning (ML) technology has been greatly improved due to the fast development of computer technology including artificial intelligence and the establishment of many databases. Extreme gradient boosting (XGBoost) is an ensemble learning algorithm that it eratively builds multiple decision trees, with each tree correcting the residuals of the previous one. The greedy algorithm is adopted to select the best splitting point, supporting parallel computing and missing value processing. It is applicable to nonlinear problems and has strong generalization ability, but it is prone to over-fitting for small sample data and needs to suppress through parameter adjustment [13,14]. It has the distinctive features of efficiently and flexibly handling of missing data and assembling weak predictive models to build accurate models [15]. XGBoost algorithm has been widely used in the medical field, including disease diagnosis, rational and safe use of medication and drug development, which helps to improve the efficiency and quality of decision making [16,17]. Logistic regression (LR) is a linear model , which fits the data by adjusting the weights and the bias, and the decision boundary is linear . It belongs to a binary classification problem. The model is simple and highly interpret-able, but has limited ability to handle nonlinear problems .It has also been applied in medical research, such as disease diagnosis, patient prognosis assessment, and drug response prediction [18]. Interestingly, some studies have compared the performance of the XGBoost algorithm model with the LR model [19], and while in some cases XGBoost was more accurate than LR [20], the opposite was shown in other cases [21]. This study systematically compared these two models to

explore their applicability in predicting the risk of death in elderly patients with ARF.

In recent years, with the continuous development of deep learning and uncertain artificial intelligence, new paths have been opened up for long-term stable prediction in the field of medical diagnosis [22]. For example, the deep self-supervised framework proposed by Zhang *et al*. [23], combined with feature elimination and selection, has demonstrated outstanding performance in the multi-dimensional health risk classification of blood tests. The dimensionality reduction method proposed by Garcia *et al*. [24] has demonstrated significant robustness in clinical practice for randomly missing data in the diagnosis of thyroid cancer. Compared with previous studies, this research focuses on core demands such as clinical interpret-ability, real-time evaluability, and data robustness. This plan aims to strike a balance between algorithm performance and clinical practicality. In the future, we plan to utilize advanced feature elimination techniques to further enhance the interpret-ability and robustness of the model in high-dimensional and noisy clinical data.

Methods

Database

Medical information mart for intensive care-IV (MIMIC-IV) is a publicly available critical care database that contains information about patients hospitalized at the Higher Medical Center in Boston from 2008 to 2019. The researchers had completed all the training course and were certified to access the database. This study focuses on data from four modules: hosp, core, icu, and ed [25]. For a description of the main tables in these four modules, please see

Table 1 for details. All data, including demographic data, medication data, comorbidities, laboratory test data, vital signs and disease severity score data, were extracted from the official Physionet website (http://mimic.physionet.org/). The relevant code for data extraction is available on the official GitHub website (https://github.com/mit-lcp/mimic-iv).

Table 1. Main tables in the MIMIC-IV database

Module	Table name	Relevant introduction
hosp	diagnoses_icd	Patient diagnostic information data
hosp	prescriptions	Patient's prescription records
hosp	labevents	Records of the patient's laboratory tests
hosp	pharmacy	Patient pharmacy data records
core	admission	Patient admission information
core	patient	General information about the patient
core	transfers	Patient turnover bed records
icu	icu_stays	ICU admission time log
icu	chartevents	Patient chart data
ed	vitalsign	Patient vital signs data

Research population

This study included patients with a clinical diagnosis of ARF, among which 9,768 patients were over 60 years of age. For this study, the inclusion criteria were: over 60 years old, and admission

to the ICU longer than 24 h. Exclusion criteria were: patients who died within 24 h of ICU admission and patients with incomplete data. For patients who were admitted to ICU multiple times, only data from their first admission were used. The multiple interpolation method is adopted to handle missing values, avoiding the introduction of noise .Then, eliminate the variables with a missing ratio of 30%, as a large missing ratio will affect the accuracy of the prediction model. Ultimately, a total of 7,500 patients were enrolled in this study.

Data extraction

PostgreSQL (v13.0) and Navicat Premium (v15.0) software were used to extract the data related to elderly patients with ARF. Then, the data were processed using R software. The main process of data processing is shown in Figure 1. General information included age at admission, body weight, length of stay in ICU, etc. Treatment measures included: vasopressor use, nor-epinephrine use, the use of continuous renal replacement therapy, etc. Related comorbidities included the following: cerebrovascular disease, mild liver disease, severe liver disease, metastatic solid tumor, etc. Disease severity scores included: sequential organ failure assessment (SOFA), acute physiology score-III (APSIII), logistic organ dysfunction system (LODS), Oxford acute severity of illness score (OASIS), simplified acute physiology score-II (SPASII), and systemic inflammatory response syndrome (SIRS). The first laboratory test results after admission to ICU included: white blood cells (WBC), prothrombin time (PT), partial thromboplastin time (PTT), anion gap (AG), and urine output, etc. Vital signs included the following: systolic blood pressure (SBP), diastolic blood pressure (DBP), pulse oxygen saturation (SpO_2), etc. Because of the high sampling frequency, the maximum, the minimum, and the

average values were used to represent vital signs and laboratory test results.

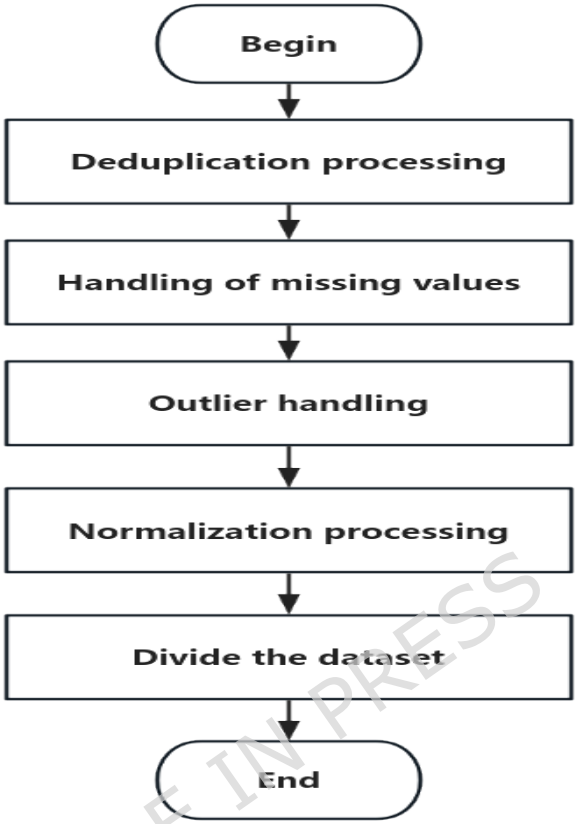

Figure 1. Data processing flow

Statistical analysis

Elderly patients with ARF were divided into two groups based on 90-day mortality. Continuous variables were represented using the median and quartile, and were compared by Mann-Whitney U test. Categorical variables were expressed in terms of frequency or percentage, and compared using the Chi-square tests or Fisher's exact tests. Elderly patients with ARF were randomly assigned to an 80% training set and a 20% validation set. This stratified segregation provides an equitable initial data distribution for model training and initial validation. The model training phase incorporated a nested cross-validation framework. This entailed an

outer 5-fold random split, with models trained on 80% of the data and their performance evaluated on the remaining 20%. Within each outer fold, an inner 5-fold cross-validation combined with grid search was employed for hyper-parameter optimization. The final reported model performance is the average across all 5 outer validation folds, ensuring reliability and reducing the risk of over-fitting.

During the model construction phase, the LR model employed AIC-based backward stepwise regression ($P < 0.05$) for feature selection, while the XGBoost model relied on its internal feature importance mechanism, reflecting the fundamental distinction and complementary nature between statistical significance testing and algorithm-based gain calculation [26]. The XGBoost objective function consists of a loss term and a regularization term, whereas the LR model uses the log-loss function to measure performance. After feature selection, XGBoost optimized its hyperparameters (such as eta and max_depth) through grid search, ultimately constructing both models. In the model comparison, ROC and DCA analyses showed that XGBoost achieved a slightly higher AUC (0.851) than LR (0.838, $P < 0.05$), but DCA validation indicated that the difference in net benefit between the two models was minimal at clinically relevant thresholds, suggesting that their impact on clinical decision-making may be limited [27]. The model fit was validated as acceptable through residual distribution and boxplot analyses [28]. In summary, we selected the XGBoost model with superior overall performance as the predictive tool and generated a breakDown plot to visually illustrate the contribution of each variable to the outcome prediction.

Results

Baseline characteristics

A total of 7,500 elderly patients with ARF were included in this study, among which 1,150 patients died within 90 days and 6,350 patients survived. Comparisons between groups showed that there were significant differences in the age at admission, body weight, vasopressor use, severe liver disease, metastatic solid tumor, urine output, PTT_max, etc. There were no statistically significant differences between the variables including myocardial infarction, glucose_min, SpO2_max, etc. Other baseline characteristics were shown in Tables 2-4.

Table 2. General information and treatment of the patients

	Death within 90 days	***Survival within 90 days***	***P***
General information			
Number (sample size)	1150	6350	
Age, year	80.00(73.28,86.81)	78.31(71.55,85.03)	**<0.001**
Gender (%)			**<0.001**
Female	552(48.0)	2738(43.1)	
Male	598(52.0)	3612(56.9)	
Body weight, kg	74.80(63.05,87.38)	77.73(66.20,91.09)	**<0.001**
Ethnicity (%)			**<0.001**
White	827(71.9)	4500(70.9)	
Black	93(8.1)	703(11.1)	
Yellow	37(3.2)	175(2.7)	
Other	193(16.8)	972(15.3)	
Length of stay in the ICU, day	4.56(2.30,8.91)	2.96(1.85,5.29)	**<0.001**
First care unit (%)			**<0.001**
CCU	177(15.4)	1140(18.0)	
SICU	134(11.7)	595(9.4)	
MICU	301(26.2)	1607(25.3)	
CVICU	63(5.5)	898(14.1)	
Other	475(41.3)	2110(33.2)	
The treatment			
Antibiotic (%)			**<0.001**
No	65(5.7)	944(14.9)	
Yes	1085(94.3)	5406(85.1)	
Dobutamine (%)			**<0.001**

No	1065(92.6)	6208(97.7)	
Yes	85(7.4)	142 (2.3)	
Dopamine (%)			**< 0.001**
No	1033(89.8)	6047(95.2)	
Yes	117(10.2)	303(4.8)	
Nerve blockers (%)			**< 0.001**
No	1082(94.1)	6290(99.1)	
Yes	68(5.9)	60(0.9)	

Table 2 (Continued)

	Death within 90 days	***Survival within 90 days***	***P***
Epinephrine (%)			**< 0.001**
No	1038(90.3)	6036(95.1)	
Yes	112(9.7)	314(4.9)	
Norepinephrine(%)			**< 0.001**
No	539(46.9)	4736(74.6)	
Yes	611(53.1)	1614(25.4)	
Phenylephrine (%)			**< 0.001**
No	798(69.4)	5292(83.3)	
Yes	352(30.6)	1058(16.7)	
Vasopressor (%)			**< 0.001**
No	825(71.7)	5968(94.0)	
Yes	325(28.3)	382(6.0)	
CRRT (%)			**< 0.001**
No	1104(96.0)	6287(99.0)	

Yes	46(4.0)	63(0.1)	

ICU intensive care unit, *CCU* coronary care unit, *SICU* surgical intensive care unit, *MICU* medical intensive care unit, *CVICU* cardiac vascular intensive care unit, *CRRT* continuous renal replacement therapy, P value less than 0.05 are shown in bold text.

Table 3. Comorbidity and score system of the patients

	Death within 90 days	***Survival within 90 days***	***P***
Comorbidity			
Myocardial_infarct (%)			0.180
No	812(70.6)	4604(72.5)	
Yes	338(29.4)	1746(27.5)	
Congestive_heart_failure (%)			0.090
No	530(46.1)	3095(48.7)	
Yes	620(53.9)	3255(51.3)	
Peripheral_vascular_disease (%)			**0.020**
No	933(81.1)	5335(84.0)	
Yes	217(18.9)	1015(16.0)	

Cerebrovascular_disease (%)			**< 0.001**
No	956(83.1)	5530(87.1)	
Yes	194(16.9)	820(12.9)	
Dementia (%)			0.680
No	1055(91.7)	5848(92.1)	
Yes	95(8.3)	502(7.9)	
Chronic_pulmonary_disease (%)			**0.020**
No	747(65.0)	4350(68.5)	
Yes	403(35.0)	2000(31.5)	
Rheumatic_disease (%)			0.590
No	1093(95.0)	6058(95.4)	
Yes	57(5.0)	292(4.6)	
Peptic_ulcer_disease (%)			0.280
No	1099(95.6)	6111(96.2)	
Yes	51(4.4)	239(3.8)	
Mild_liver_disease (%)			**< 0.001**
No	943(82.0)	5797(91.3)	
Yes	207(18.0)	553(8.7)	
Diabetes uncomplicated (%)			0.310
No	813(70.7)	4392(69.2)	
Yes	337(29.3)	1958(30.8)	
Diabetes complicated (%)			**0.020**
No	991(86.2)	5298(83.4)	
Yes	159(13.8)	1052(16.6)	
Paraplegia (%)			**< 0.001**
No	1094(95.1)	6160(97.0)	
Yes	56(4.9)	190(3.0)	

Table 3 (Continued)

	Death within 90 days	***Survival within 90 days***	***P***
Renal_disease (%)			**0.030**
No	674(58.6)	3497(55.1)	
Yes	476(41.4)	2853(44.9)	
Malignant_cancer (%)			**< 0.001**
No	849(73.8)	5413(85.2)	
Yes	301(26.2)	937(14.8)	

Severe_liver_disease (%)			**< 0.001**
No	1049(91.2)	6138(96.7)	
Yes	101(8.8)	212(3.3)	
Metastatic_solid_tumor (%)			**< 0.001**
No	963(83.7)	5980(94.2)	
Yes	187(16.3)	370(5.8)	
Aids (%)			0.680
No	1148(99.8)	6341(99.8)	
Yes	2(0.2)	9(0.2)	
Score system			
SOFA	8.00(6.00,12.00)	5.00(3.00,8.00)	**< 0.001**
APSIII	75.00(59.00,97.00)	50.00(41.00,63.00)	**< 0.001**
LODS	9.00(6.00,11.00)	5.00(4.00,7.00)	**< 0.001**
OASIS	40.00(34.00,47.00)	33.00(27.00,39.00)	**< 0.001**
SAPSII	52.00(43.00,63.00)	41.00(35.00,49.00)	**< 0.001**
SIRS	3.00(2.00,3.00)	3.00(2.00,3.00)	**< 0.001**

SOFA sequential organ failure assessment, *APSIII* acute physiology and chronic health score III, *LODS* logistic organ dysfunction system, *OASIS* Oxford acute severity of illness score, *SAPSII* simplified acute physiology score II, *SIRS* systemic inflammatory response syndrome, P value less than 0.05 are shown in bold text.

Table 4. Laboratory tests and vital signs of the patients

	Death within 90 days	***Survival within 90 days***	***P***
Laboratory tests			
Hematocrit_min (%)	28.00(23.90,32.50)	28.70(24.80,33.20)	**<0.001**
Hematocrit_max (%)	32.10(28.40,36.48)	32.70(29.30,37.10)	**<0.001**
Hemoglobin_min (g/dL)	9.00(7.70,10.60)	9.40(8.10,10.90)	**<0.001**
Hemoglobin_max (g/dL)	10.30(9.00,11.70)	10.70(9.40,12.20)	**<0.001**
Platelets_min (k/uL)	164.00(100.00,225.00)	167.00(122.00,223.75)	**<0.001**
Platelets_max (k/uL)	199.00(132.00,264.00)	201.00(154.00,267.00)	**<0.001**
WBC_min (k/uL)	9.90(7.10,13.50)	9.50(6.93,12.40)	**<0.001**
WBC_max (k/uL)	13.00(10.00,17.98)	12.70(9.20,17.00)	**<0.001**
AG_min (mEq/L)	14.00(12.00,17.00)	13.00(11.00,15.00)	**<0.001**
AG_max (mEq/L)	17.00(16.00,21.00)	17.00(14.00,19.00)	**<0.001**
Bicarbonate_min (mEq/L)	20.00(16.00,23.00)	21.00(18.00,23.75)	**<0.001**
Bicarbonate_max (mEq/L)	23.00(20.00,25.00)	23.00(21.00,26.00)	**<0.001**
BUN_min (mg/dL)	34.00(25.00,49.75)	30.00(21.00,42.00)	**<0.001**

BUN_max (mg/dL)	39.00(29.00,58.00)	36.00(25.00,49.00)	**<0.001**
Calcium Total_min (EU/dL)	8.00(7.50,8.50)	8.10(7.70,8.60)	**<0.001**
Calcium Total_max (EU/dL)	8.60(8.00,9.00)	8.60(8.10,9.00)	**0.040**
Chloride_min(mEq/L)	102.00(97.00,105.00)	102.00(98.00,106.00)	**<0.001**
Chloride_max (mEq/L)	105.00(101.00,109.75)	106.00(102.00,110.00)	**<0.001**
Creatinine_min (g/dL)	1.30(1.10,1.90)	1.30(1.10,1.70)	**<0.001**
Creatinine_max (g/dL)	1.60(1.30,2.30)	1.60(1.30,2.10)	**<0.001**
Sodium_min (mEq/L)	137.00(134.00,140.00)	137.00(134.00,140.00)	0.190
Sodium_max (mEq/L)	140.00(137.00,143.00)	140.00(137.00,142.00)	0.410
Potassium_min (mEq/L)	4.00(3.60,4.50)	4.00(3.60,4.40)	0.080
Potassium_max (mEq/L)	4.60(4.20,5.20)	4.60(4.20,5.00)	**0.020**
PT_min (s)	13.90(13.10,16.40)	13.60(12.30,14.98)	**<0.001**
PTT_min (s)	29.50(27.70,34.98)	29.30(26.40,32.70)	**<0.001**
PTT_max (s)	39.95(31.13,66.50)	34.19(29.00,49.70)	**<0.001**
Glucose_min (mg/dL)	110.50(93.00,138.00)	110.00(94.00,130.00)	0.110
Glucose_max (mg/dL)	155.00(127.00,200.00)	150.00(123.00,188.00)	**<0.001**

Urine output (ml)	812.0(397.75,1430.75)	1315.0(880.0,1990.0)	**< 0.001**
Vital Signs			
Heart rate_min (min-1)	73.00(63.00,86.00)	68.00(60.00,78.00)	**< 0.001**
Heart rate_max (min-1)	107.00(94.05,124.00)	99.00(87.00,114.00)	**< 0.001**
Heart rate_mean (min-1)	89.24(77.50,101.57)	81.62(72.38,92.32)	**< 0.001**

Table 4 (Continued)

	Death within 90 days	***Survival within 90 days***	***P***
SBP_min (mmHg)	84.00(75.00,91.00)	88.00(80.00,98.00)	**< 0.001**
SBP_max (mmHg)	141.00(126.00,156.00)	144.00(132.00,160.0)	**< 0.001**
SBP_mean (mmHg)	108.15(100.49,117.5)	113.75(105.62,125.26)	**< 0.001**
DBP_min (mmHg)	41.00(35.00,47.00)	42.00(37.00,48.00)	**< 0.001**
DBP_max (mmHg)	84.00(72.00,97.00)	84.00(72.00,97.00)	0.690
DBP_mean (mmHg)	57.81(51.97,63.91)	58.12(52.44,64.87)	**0.030**
Respiratory_rate_min (min-1)	13.00(11.00,16.00)	13.00(11.00,16.00)	**< 0.001**
Respiratory_rate_max(min-1)	30.00(26.00,34.00)	28.00(24.00,31.00)	**< 0.001**
Respiratory rate_mean(min-1)	20.45(18.15,23.54)	19.24(17.13,21.72)	**< 0.001**
Temperature_min (℃)	36.39(36.06,36.56)	36.39(36.17,36.61)	**< 0.001**

Temperature_max (℃)	37.11(36.83,37.50)	37.11(36.89,37.44)	**0.040**
Temperature_mean (℃)	36.73(36.51,36.92)	36.75(36.56,36.95)	**< 0.001**
SpO2_min (%)	91.00(87.00,94.00)	92.00(90.00,94.00)	**< 0.001**
SpO2_max (%)	100.00(99.85,100.00)	100.00(99.00,100.00)	0.070
SpO2_mean (%)	96.70(94.96,98.30)	96.88(95.50,98.20)	**< 0.001**

WBC white blood cells, *AG* anion gap, *BUN* blood urea nitrogen, *PT* prothrombin time, *PTT* partial thromboplastin time, *SBP* systolic blood pressure, *DBP* diastolic blood pressure, *SpO2* pulse oxygen saturation, *Max* maximum, *Min* minimum , P value less than 0.05 are shown in bold text.

Model construction

(1) XGBoost model

Using the method of backward stepwise regression analysis, the variables with p-values less than 0.05 were screened out, and then the XGBoost model was constructed. The analysis of different variables according to the XGBoost model showed that APSIII, LODS, length of stay in ICU, urine output, age at admission, vasopressor, metastatic solid tumor, SpO2_mean, platelets_max, respiratory rate_mean, PTT_max, SBP_mean, heart rate_min, hemoglobin_max, and body weight were the 15 most important features, which were all strongly correlated with the 90-day mortality rate as shown in Figure 2.

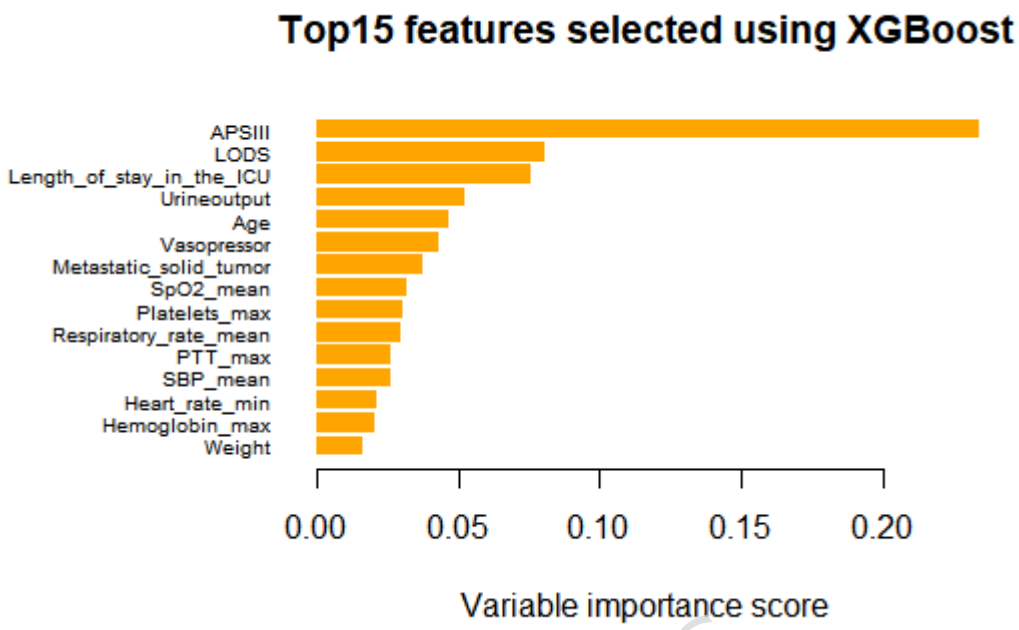

Figure 2. Top 15 features selected using XGBoost.

(2) LR model

Using the method of backward stepwise regression analysis, the variables with p-values less than 0.05 were screened out, and then the LR model was constructed. LR analysis showed that age, vasopressor, severe liver disease, and metastatic solid tumor were significant risk factors that increased the risk of death in elderly ARF patients. Urine output, serum creatinine, body temperature_min, and SpO2_mean were protective factors for elderly ARF patients. Detailed information of LR analysis was shown in Table 5.

Table 5. Features selected in the logistic regression

Variable	OR(95%CI)	*P*
(Intercept)	0.006(0.004-0.01)	< 0.001
Age	1.349(1.227-1.485)	< 0.001
Body weight	0.827(0.753-0.907)	< 0.001
Dobutamine	1.874(1.247-2.803)	< 0.001
Dopamine	1.576(1.125-2.192)	< 0.001
Norepinephrine	1.324(1.068-1.642)	0.010
Vasopressor	2.540(1.955-3.300)	< 0.001
Myocardial_infarct	1.364(1.122-1.655)	< 0.001
Peripheral_vascular_disease	1.299(1.035-1.624)	0.020
Cerebrovascular_disease	1.391(1.075-1.791)	0.010
Chronic_pulmonary_disease	1.336(1.112-1.605)	< 0.001
Mild_liver_disease	1.812(1.356-2.408)	< 0.001
Paraplegia	2.399(1.512-3.743)	< 0.001
Malignant_cancer	1.601(1.266-2.017)	< 0.001
Severe_liver_disease	1.892(1.257-2.840)	< 0.001
Metastatic_solid_tumor	3.144(2.342-4.215)	< 0.001
Hemoglobin_max	0.796(0.727-0.871)	< 0.001
Platelets_min	1.396(1.179-1.655)	< 0.001
Platelets_max	0.656(0.553-0.775)	< 0.001

WBC_min	1.119(1.030-1.217)	< 0.001
Chloride_min	0.905(0.829-0.986)	0.020
Creatinine_max	0.823(0.753-0.898)	< 0.001
PTT_max	1.175(1.084-1.272)	< 0.001
Urine output	0.744(0.668-0.827)	< 0.001
Heart_rate_min	1.268(1.159-1.388)	< 0.001
SBP_min	1.273(1.126-1.442)	< 0.001
SBP_mean	0.875(0.770-0.993)	0.040
Respiratory_rate_mean	1.253(1.146-1.370)	< 0.001
Temperature_min	0.874(0.802-0.953)	< 0.001
SpO2_mean	0.811(0.743-0.883)	< 0.001
APSIII	1.029(1.023-1.036)	< 0.001
LODS	1.159(1.107-1.214)	< 0.001
OASIS	0.981(0.966-0.997)	0.020

OR odds ratio, *WBC* white blood cells, *PTT* partial thromboplastin time, *SBP* systolic blood pressure, *SpO2* pulse oxygen saturation, *APSIII* acute physiology and chronic health score III, *LODS* logistic organ dysfunction system, *OASIS* Oxford acute severity of illness score.

Model comparison

In the model validation phase, both the XGBoost algorithm model and the LR model showed good discrimination ability. The area under the curve (AUC) values of the two models were 0.838 (LR) and 0.851 (XGBoost), respectively, whereas the AUC value of the XGBoost model was larger (Figure 3). There were significant differences in AUC among different models (P=0.013). The DCA mainly directly assesses the practicability of the model for clinical

decision-making by quantifying the "net benefit" under different intervention thresholds . As can be seen from the figure 4, the curves of the two models are higher than those of the "all treatment" or "all no treatment" strategies in the vast majority of threshold ranges, indicating that these two models have a clinical net benefit. By further observing the DCA of the two prediction models, we found that the net benefit of the XGBoost model had a larger range than the LR model, indicating that the XGBoost model had a higher clinical utility (Figure 4). Furthermore, the XGBoost model had smaller sample residuals and root-mean-square residuals, indicating that the XGBoost algorithm fitted better and the predicted values of the model were closer to the actual values (Figure 5). Taken together, comparison of the two models showed that the XGBoost algorithm model was a better model for predicting 90-day mortality in elderly patients with ARF.

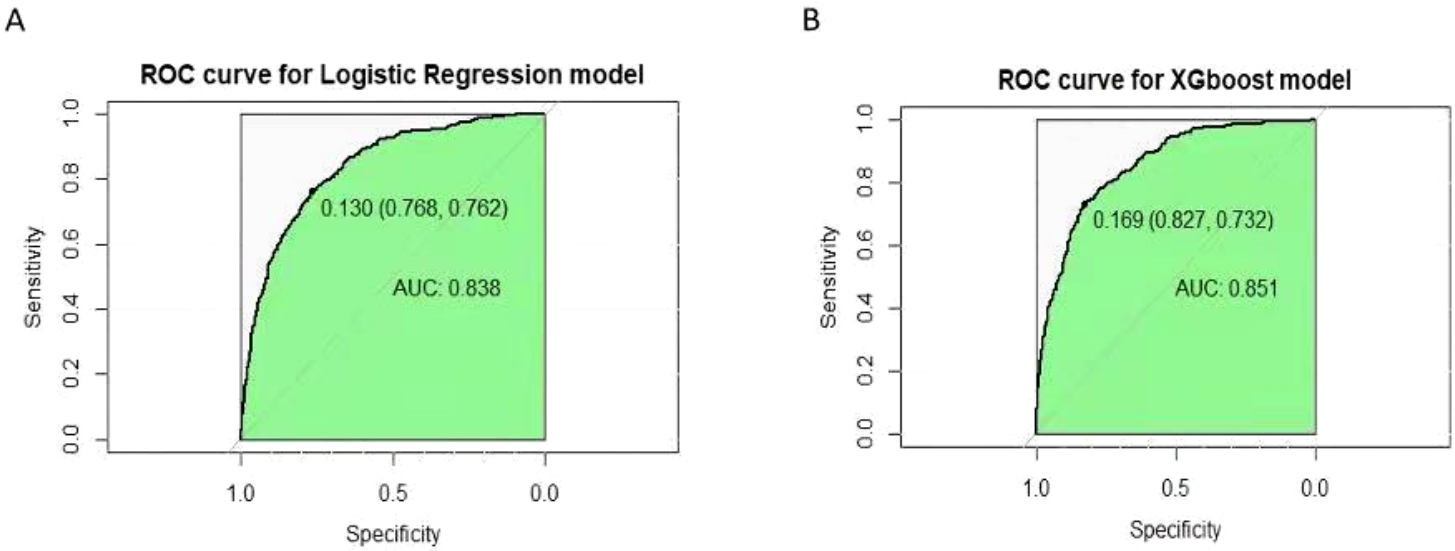

Figure 3. The receiver operating characteristic curves. A Logistic regression model, area under curves (AUC) is 0.838; B XGBoost model, AUC is 0.851.

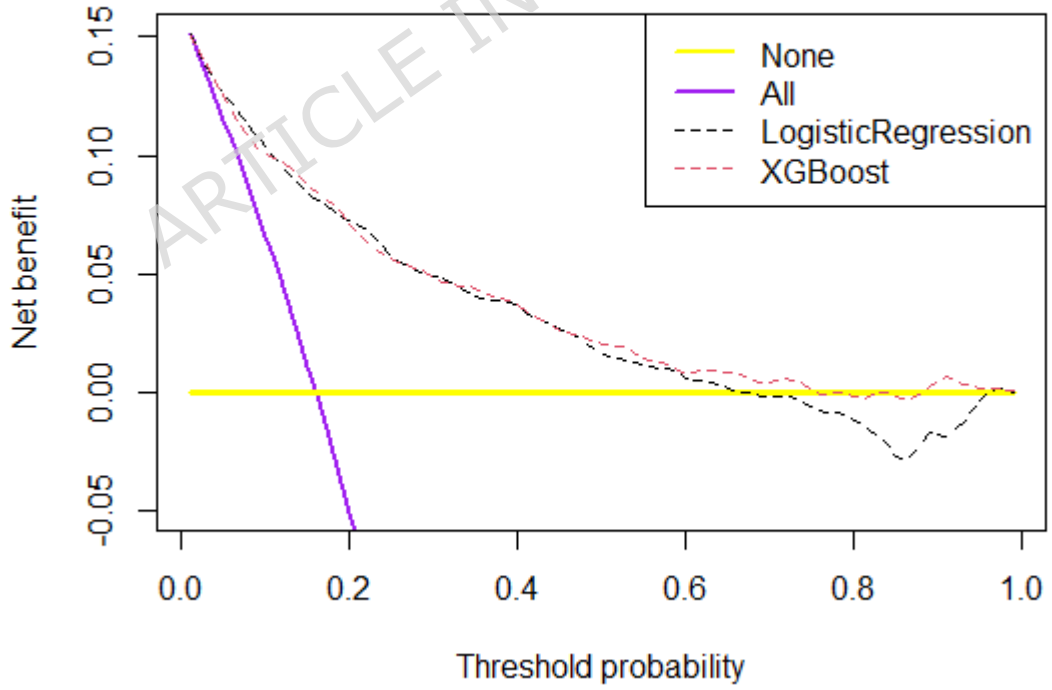

Figure 4. Decision curve analysis of the two prediction models.

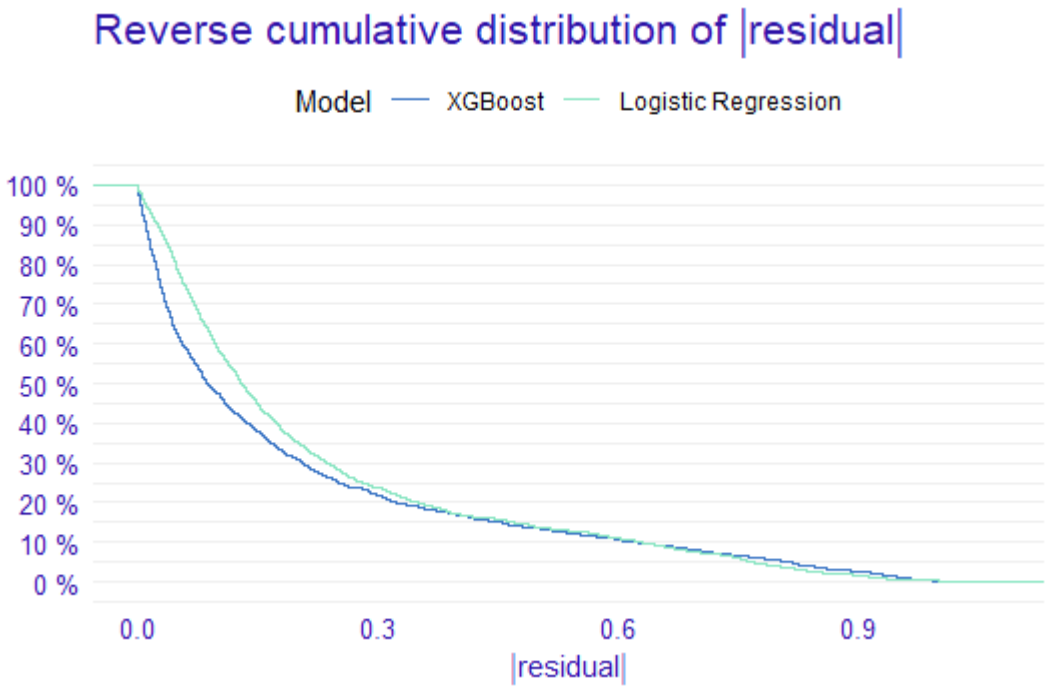

Figure 5. Cumulative residual distribution plot for two models. X-axis indicates the absolute residual value and Y-axis indicates the cumulative percentage of residuals. Solid blue line=XGBoost, Solid sky-blue line=Logistic regression. The preferred model is the XGBoost model.

Optimal model analysis

Given that the XGBoost model was superior to the LR model in terms of discrimination, clinical validity, and degree of fit, breakdown plots were constructed on the basis of the XGBoost model to predict the risk of death at the individual level in elderly patients with ARF. Breakdown plots were constructed using significant variables such as age, vasopressor, urine output, metastatic solid tumor, etc. Green color indicated a positive effect on outcome indicators and red color indicated a negative effect. The size of the absolute value indicated the degree of risk, with larger values indicating higher risk. As shown in Figure 6, the extent to which each indicator contributed to the outcome variable could be clearly seen. This study identified through the breakdown

plot that critical illness score, urine output, age, weight, hemoglobin, heart rate, partial thromboplastin time, blood oxygen saturation, vasopressin and metastatic solid tumor were the ten key predictive factors affecting the mortality rate of this patient. These factors together explained the total predictive power of the model. Finally, the risk of death at 90 days for the patient was obtained.

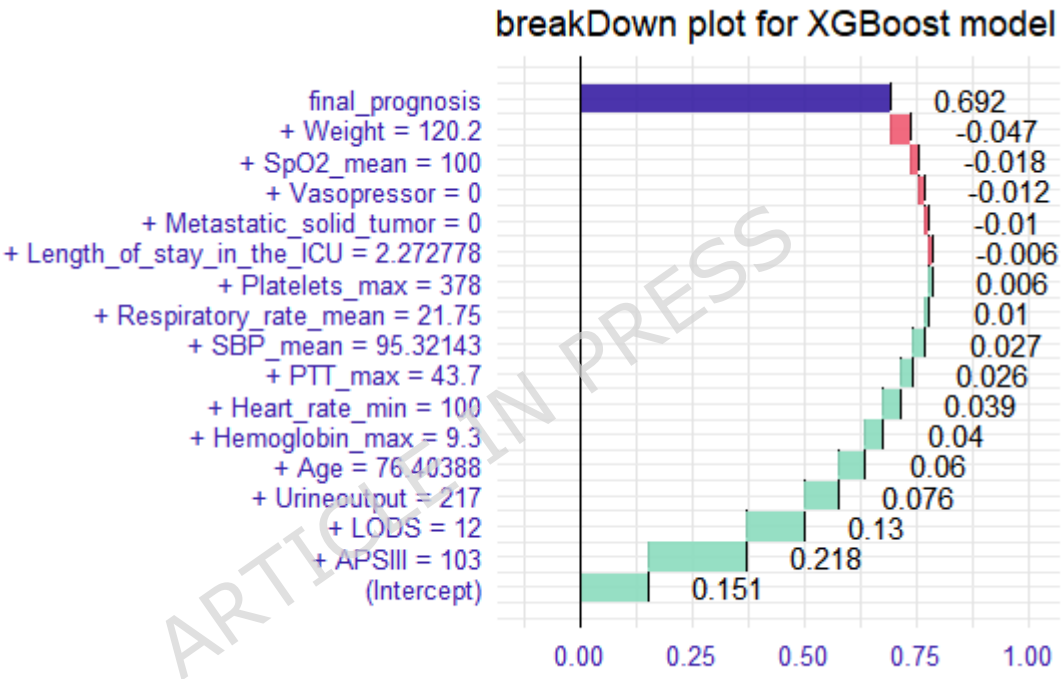

Figure 6. Breakdown plot for XGBoost model.

Discussion

Compared with previous studies using the MIMIC database to predict ARF in ICU [29], there are several advantages in this study [30]. Firstly, the group with the highest ARF mortality rate was used in this study. Elderly ARF patients in ICU were preferentially selected for the study of short-term mortality in ARF. This is because age-related changes in organ structure and function

render elderly patients in ICU more susceptible to ARF [31]. Moreover, relevant studies have shown that the short-term mortality rate of ARF is more than 60% [32]. Secondly, the metric of goodness of fit was added to determine the performance of the model in this study, and the comparison of goodness of fit was mainly demonstrated by cumulative residual distribution plot and residual box-plot. Finally, this study utilized breakdown plots to predict the impact of different variables from the XGBoost model on the outcome metrics.

The results of this study showed that the XGBoost model has better performance than the LR model for predicting ARF mortality in the elderly by AUC and DCA. The advantage of XGBoost lies in its powerful nonlinear modeling capability, which can capture the interaction relationships and threshold effects in complex clinical data. However, its "black box" nature may affect the trust of clinicians. For example, Mohamadlou *et al.* [33] utilized the XGBoost algorithm to predict the risk of ARF in critically ill patients. Lei *et al*. [34] built the XGBoost algorithm to identify the risk of ARF after surgery. Koyner *et al*. [35] also utilized the XGBoost algorithm model to make a prediction of the risk of ARF occurring within 48 h of admission, thus identifying influencing factors that would increase the risk of ARF development. In contrast, the advantage of LR lies in its simplicity and interpretability, but it is difficult to handle nonlinear relationships and multicollinearity problems. For example, Kristovic *et al*. [36] applied LR method to construct a predictive model for ARF in postoperative patients. An *et al*. [37] utilized LR modeling to study the common risk factors of ARF patients in neurosurgical ICU. However, several studies have found that LR methods have lower AUC values compared to some of the newer techniques [38],

suggesting that the measured performance metrics are relatively low, or with higher prediction errors [39]. Our result is in consistent with some previous studies in which the XGBoost model was shown to be superior. For instance, Yue *et al*. [40] studied the risk of ARF in patients with sepsis and found that the XGBoost model was the most effective model among all the prediction models. Furthermore, a meta-analysis study showed that XGBoost is more effective than LR and other ML algorithms in predicting ARF [43]. It is noteworthy that while XGBoost demonstrated statistically significant advantages in key clinical metrics (e.g., clinical net benefit in critical thresholds), the overall predictive performance of traditional logistic regression remained comparable. This aligns with a recent comparative study in the Iranian ED setting, which concluded that LR can perform as well as advanced ensemble models in predicting in-hospital mortality [44]. While both XGBoost and LR models showed good predictive performance, clinical implementation requires interpretable decision support. As demonstrated in recent studies, nomograms offer key advantages by consolidating risk factors into a visual format, enabling rapid risk estimation via point-scoring, and clearly marking clinical thresholds (e.g., ICU admission at 85% probability) [45].

Although the XGBoost model only slightly outperforms the logistic regression model in terms of the area under the ROC curve (AUC), this difference holds significant clinical and statistical importance. First, the AUC difference demonstrates statistical significance ($p < 0.05$), indicating that the XGBoost model consistently outperforms traditional methods in distinguishing between high-risk and low-risk patients. Second, clinical decision-making is not solely based on a single metric. Decision curve

analysis (DCA) confirms that the XGBoost model provides higher clinical net benefits across a broader range of decision thresholds. This means that when using the XGBoost model, clinicians can identify more true high-risk patients at the same false-positive rate, or reduce unnecessary interventions at the same true-positive rate. Additionally, the model fit evaluation (residual analysis) shows that the XGBoost model has a lower root mean square error compared to the LR model, indicating that its predictions are not only more accurate but also more stable. This subtle difference is particularly crucial in resource-intensive ICU settings, as it may enable more precise timing for critical interventions. Therefore, we recommend that future research further validate the added value of these two models in real-world clinical decision-making through prospective, multicenter cohorts, and explore their potential applications in personalized treatment planning.

Then, for clinical scenarios with abundant data and complex interactions among variables, XGBoost is a better choice[46]. However, in clinical Settings where resources are limited and rapid explanations are required, LR still holds significant value. Therefore, for large medical institutions, it is recommended to use XGBoost to build predictive models and combine them with interpret-able tools such as breakdown plots to assist clinical decision-making. For primary medical institutions, the LR model can be adopted because of its relatively low demand for computing resources. Both models can be used in clinical decision support systems, but they need to be optimized in combination with feedback from clinicians.

In the XGBoost algorithm model, APSIII, urine output, age at admission, vasopressor, metastatic solid tumor, SpO2_mean, SBP_mean, and body weight were all strongly correlated with the

mortality rate of elderly ARF patients. APSIII score had the greatest weight among these characteristics. APSIII score is commonly used for determining disease severity and predicting mortality, and performs well in the timely identification of high-risk patients and the development of intervention strategies [47]. Urine output has long been recognized as the most common influencing factor for ARF [48]. Oliguria is often presented as the first clinical sign of ARF and is one of the criteria for the diagnosis of ARF by KDIGO (Kidney Disease: Improving Global Outcomes) [49]. Also, decreased urine output can cause hypovolemia, which can promote the development of ARF [50]. Prompt rehydration therapy restores circulating blood volume and improves impaired renal perfusion [51,52]. Furthermore, vasoactive substances may influence ARF progression; for instance, vasopressors can increase glomerular perfusion pressure and urine output, potentially elevating ARF risk [53]. Our model demonstrated superior predictive performance for 90-day mortality in elderly ARF patients, with vasopressin identified as a key predictor. However, limitations exist: (1) CRRT-exclusive enrollment due to ICU data completeness may limit generalizability to non-ICU populations; (2) the 90-day observation period may miss long-term outcomes; (3) inter-institutional dialysis criteria variability could introduce heterogeneity. Future multicenter studies with extended follow-up and comprehensive dialysis tracking are recommended. In addition, metastatic solid tumor is a common comorbidity in elderly patients with ARF, and Rosner and Perazella suggested that the production of inflammatory cytokines resulted in an increased mortality rate of patients with ARF [54]. Because these factors are easy to assess at the time of patient admission, they can be used as predictors in elderly patients with ARF.

Since the XGBoost model outperforms the LR model, we constructed breakdown plot for interpreting the XGBoost model. The breakdown plot helps doctors quickly identify key risk factors by quantifying the contribution rate of each clinical variable to the predicted outcome (such as age, laboratory indicators, etc.), thereby enabling targeted adjustment of treatment plans. Meanwhile, it is also a visualization tool used to discover how the specific value of each variable affects the prognosis of the model. Finally, by comparing the consistency between the contribution values and clinical experience, the rationality of the model logic is verified to provide data-driven optimization basis for decision-making [55]. It can help doctors provide the best medical plans for patients and offer reliable conclusions for research. Therefore, this study suggests that the predictive results of the XGBoost model should be regarded as an auxiliary, higher-precision risk stratification tool, rather than an isolated decision-making basis. Clinically, it is recommended to integrate the predictions of such models with dynamic renal function monitoring (e.g., daily urine output changes, electrolyte levels) and the comprehensive judgment of physicians, thereby achieving more precise and personalized patient management.

This study also has some limitations: first, it is a single-center study and lacks external validation. Second, the MIMIC-IV database does not provide patient history and long-term follow-up events, and some key impact variables (contrast agent exposure and nephrotoxic drug exposure) might be overlooked. Thirdly, when the number of deaths is scarce, the model tends to favor the majority category, leading to missed diagnoses. In subsequent research, it is planned to optimize the model by adjusting parameters or through oversampling techniques. Finally, this is a retrospective study in

which most of the patients were white, which may impact on the results. In future studies, it needs to be validated in conjunction with further prospective multi-center studies.

Conclusions

The study shows that for predicting 90-day mortality in elderly ARF patients in the ICU, the XGBoost algorithm model is significantly better than the traditional LR model. APSIII, urine output, vasopressor medications, and metastatic solid tumor were all found to be strongly associated with ARF mortality in the elderly.

Abbreviations

ARF: acute renal failure; ICU: intensive care unit; XGBoost: extreme gradient boosting; LR: logistic regression; DCA: decision curve analysis; ML: machine learning; MIMIC-IV: medical information mart for intensive care-IV; SOFA: sequential organ failure assessment; APSIII: acute physiology score-III; LODS: logistic organ dysfunction system; OASIS: Oxford acute severity of illness score; SPASII: simplified acute physiology score-II; SIRS: systemic inflammatory response syndrome; WBC: white blood cells; PT: prothrombin time; PTT: partial thromboplastin time; AG: anion gap; SBP: systolic blood pressure; DBP: diastolic blood pressure; SpO2: pulse oxygen saturation; ROC: receiver operating characteristic curve; AUC: area under the curve; CCU: coronary care unit; SICU: surgical intensive care unit; MICU: medical intensive care unit; CVICU: cardiac vascular intensive care unit; CRRT: continuous renal replacement therapy; BUN: blood urea nitrogen; Max: maximum; Min: minimum; OR: odds ratio.

Declarations

Ethics approval and consent to participate

This study was conducted in accordance with the Declaration of Helsinki. Institutional review board approval and informed consent were not required in current study because MIMIC-IV data is publicly available and all patient data are de-identified. Informed consent of all subjects and/or their legal guardians was obtained when MIMIC-IV was established.

Consent to publish

Not applicable.

Availability of data and materials

The data that support the findings of this study are openly available on the MIMIC-IV website at https://physionet.org/content/mimiciv/1.0/. The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was supported in part by grants from the National Natural Science Foundation of China (No. 32270186) and the Huadong Medicine Joint Funds of the Zhejiang Provincial Natural Science Foundation (No. LHDMZ24H040001).

Acknowledgement

Not applicable.

Authors' contributions

JZ created the study protocol, performed the statistical analyses and wrote the first manuscript draft. YZ and FY assisted with data collection. QS assisted with data interpretation. HM and JY assisted with manuscript revision. YW conceived the study and critically revised the manuscript. All authors read and approved the final manuscript.

References

1. Levey AS, James MT. Acute Kidney Injury. Ann Intern Med. 2017; 167(9): ITC66-ITC80.

2. Li Q, Zhao M, Du J, Wang X. Outcomes of renal function in elderly patients with acute kidney injury. Clin Interv Aging. 2017; 12: 153-160.

3. Koza Y. Acute kidney injury: current concepts and new insights. J Inj Violence Res. 2016; 8(1): 58-62.

4. Chang-Panesso M. Acute kidney injury and aging. Pediatr Nephrol. 2021; 36(10): 2997-3006.

5. Weidner DA, Yoo MJ. A Review of Acute Kidney Injury. J Spec Oper Med. 2022; 22(3): 70-74.

6. Al-Jaghbeer M, Dealmeida D, Bilderback A, Ambrosino R, Kellum JA. Clinical decision support for in-Hospital AKI. J Am Soc Nephrol. 2018; 29(2): 654-660.

7. Hoste EA, Bagshaw SM, Bellomo R, Cely CM, Colman R, Cruz DN, et al. Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. Intensive Care Med. 2015; 41(8): 1411-1423.

8. Holmes J, Roberts G, Meran S, Williams JD, Phillips AO; Welsh AKI Steering Group. Understanding Electronic AKI Alerts: Characterization by Definitional Rules. Kidney Int Rep. 2016; 2(3): 342-349.

9. Wang JL, Yan YC. Research progress of acute kidney injury in elderly patients. Chinese Journal of Integrated Traditional and Western Nephropathy. 2012; 13(3): 272-275.

10. Lu RH, Fang Y, Gao JY, Cai H, Zhu ML, Zhang MF, et al. Analysis of incidence and risk factors of acute kidney injury in inpatients. Chinese Journal of Critical Care and Emergency Medicine. 2011; 23(7): 413-417.
11. Yang YJ, Cai GY. Diagnosis and treatment of acute kidney injury in elderly patients. Chin J Clinical Healthcare. 2020; 23(1): 15-19.
12. Hoste EAJ, Kellum JA, Selby NM, Zarbock A, Palevsky PM, Bagshaw SM, et al. Global epidemiology and outcomes of acute kidney injury. Nat Rev Nephrol. 2018; 14(10): 607-625.
13. Lee HC, Yoon SB, Yang SM, Kim WH, Ryu HG, Jung CW, et al. Prediction of Acute Kidney Injury after Liver Transplantation: Machine Learning Approaches vs Logistic Regression Model. J Clin Med. 2018; 7(11): 428.
14. Hsu CN, Liu CL, Tain YL, Kuo CY, Lin YC. Machine Learning Model for Risk Prediction of Community-Acquired Acute Kidney Injury Hospitalization From Electronic Health Records: Development and Validation Study. J Med Internet Res. 2020; 22(8): e16903.
15. Yuan KC, Tsai LW, Lee KH, Cheng YW, Hsu SC, Lo YS, et al. The development an artificial intelligence algorithm for early sepsis diagnosis in the intensive care unit. Int J Med Inform. 2020; 141: 104176.
16. Qi QN, Liu Y, Chen JH, Liu XZ, Yang R, Zhang JY, et al. Research progress of machine learning XGBoost algorithm in medical field. Journal of Molecular Imaging. 2019; 44(5): 856-862.
17. Huang YC, Li SJ, Chen M, Lee TS, Chien YN. Machine-Learning Techniques for Feature Selection and Prediction of Mortality in Elderly CABG Patients. Healthcare (Basel). 2021; 9(5):547.
18. Schober P, Vetter TR. Logistic Regression in Medical Research. Anesth Analg. 2021; 132(2): 365-366.
19. Yan Z, Wang J, Dong Q, Zhu L, Lin W, Jiang X. XGBoost algorithm and logistic regression to predict the postoperative 5-year outcome in patients with glioma. Ann Transl Med. 2022; 10(16): 860.
20. Hou N, Li M, He L, Xie B, Wang L, Zhang R, et al. Predicting 30-days

mortality for MIMIC-III patients with sepsis-3: a machine learning approach using XGboost. J Transl Med. 2020; 18(1): 462.

21. Xu Y, Han D, Huang T, Zhang X, Lu H, Shen S, et al. Predicting ICU Mortality in Rheumatic Heart Disease: Comparison of XGBoost and Logistic Regression. Front Cardiovasc Med. 2022; 9: 847206.

22. Foersch S, Glasner C, Woerl A. Multistain deep learning for prediction of prognosis and therapy response in colorectal cancer. Nature Medicine. 2023; 30(3): 430-439.

23. Zhang L, Wang H, Chen Y. Deep self-supervised machine learning algorithms with a novel feature elimination and selection approaches for blood test-based multi-dimensional health risks classification. BMC Bioinformatics. 2023; 24(1).

24. Garcia M, Smith J, Lee K, Patel R. Reprogramming of myeloid cells and their progenitors in patients with non-medullary thyroid carcinoma. Nature Communications. 2022; 13(1).

25. Ulrich H, Behrend P, Wiedekopf J, Drenkhahn C, Kock-Schoppenhauer AK, Ingenerf J. Hands on the Medical Informatics Initiative Core Data Set-Lessons Learned from Converting the MIMIC-IV. Stud Health Technol Inform. 2021;283:119-126.

26. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent Reporting of a multi-variable prediction model for Individual Prognosis or Diagnosis (TRIPOD): the TRIPOD statement. Ann Intern Med. 2015; 162(1): 55-63.

27. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, et al. Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. Eur Urol. 2018; 74(6): 796-804.

28. Pan X, Jin X, Wang J, Hu Q, Dai B. Placenta inflammation is closely associated with gestational diabetes mellitus. Am J Transl Res. 2021;13(5): 4068-4079.

29. Wei C, Zhang L, Feng Y, Ma A, Kang Y. Machine learning model for predicting acute kidney injury progression in critically ill patients. BMC Med

Inform Decis Mak. 2022; 22(1): 17.

30. Zhao X, Lu Y, Li S, eGuo F, Xue H, Jiang L, et al. Predicting renal function recovery and short-term reversibility among acute kidney injury patients in the ICU: comparison of machine learning methods and conventional regression. Ren Fail. 2022; 44(1): 1326-1337.

31. Liu F, Chen J, Li Z, Meng X. Recent Advances in Epigenetics of Age-Related Kidney Diseases. Genes (Basel). 2022; 13(5): 796.

32. Doyle JF, Forni LG. Acute kidney injury: short-term and long-term effects. Crit Care. 2016; 20(1): 188.

33. Mohamadlou H, Lynn-Palevsky A, Barton C, Chettipally U, Shieh L, Calvert J, et al. Prediction of acute kidney injury with a machine learning algorithm using electronic health record data. Canadian Journal of Kidney Health and Disease. 2018; 5: 1-9.

34. Lei VJ, Luong T, Shan E, Chen X, Neuman MD, Eneanya ND, et al. Risk Stratification for Postoperative Acute Kidney Injury in Major Noncardiac Surgery Using Preoperative and Intraoperative Data. JAMA Netw Open. 2019; 2(12): e1916921.

35. Koyner JL, Carey KA, Edelson DP, Churpek MM. The Development of a Machine Learning Inpatient Acute Kidney Injury Prediction Model. Crit Care Med. 2018; 46(7): 1070-1077.

36. Kristovic D, Horvatic I, Husedzinovic I, Sutlic Z, Rudez I, Baric D, et al. Cardiac surgery-associated acute kidney injury: risk factors analysis and comparison of prediction models. Interact Cardiovasc Thorac Surg. 2015; 21(3): 366-373.

37. An S, Luo H, Wang J, Gong Z, Tian Y, Liu X, et al. An acute kidney injury prediction nomogram based on neurosurgical intensive care unit profiles. Ann Transl Med. 2020; 8(5): 194.

38. Xiao J, Ding R, Xu X, Guan H, Feng X, Sun T, et al. Comparison and development of machine learning tools in the prediction of chronic kidney disease progression. J Transl Med. 2019; 17(1): 119.

39. Li YM, Li ZL, Chen F, Liu Q, Peng Y, Chen M. A LASSO-derived risk model for long-term mortality in Chinese patients with acute coronary syndrome. J Transl Med. 2020; 18(1): 157.

40. Yue S, Li S, Huang X, Liu J, Hou X, Zhao Y, et al. Machine learning for the prediction of acute kidney injury in patients with sepsis. J Transl Med. 2022; 20(1): 215.

41. Song X, Liu X, Liu F, Wang C. Comparison of machine learning and logistic regression models in predicting acute kidney injury: a systematic review and meta-analysis. Int J Med Inform.2021; 151: 104484.

42. Rahmatinejad Z, Dehghani T, Hoseini B, Rahmatinejad F, Lotfata A, Reihani H, et al. A comparative study of explainable ensemble learning and logistic regression for predicting in-hospital mortality in the emergency department. Sci Rep. 2024;14(1):3406.

43. Rahmatinejad Z, Peiravi S, Hoseini B, Rahmatinejad F, Eslami S, Abu-Hanna A, et al. Comparing In-Hospital Mortality Prediction by Senior Emergency Resident's Judgment and Prognostic Models in the Emergency Department. Biomed Res Int. 2023; 2023: 6042762.

44. Schmidt JB. Conformal Prediction and Venn-ABERS Calibration for Reliable Machine Learning-Based Prediction of Bacterial Infection Focus. bioRxiv. 2025.

45. Mercado MG, Smith DK, Guard EL. Acute Kidney Injury: Diagnosis and Management. Am Fam Physician. 2019; 100(11): 687-694.

46. Rahman M, Shad F, Smith MC. Acute kidney injury: a guide to diagnosis and management. Am Fam Physician. 2012; 86(7): 631-639.

47. Kellum JA, Lameire N, KDIGO AKI Guideline Work Group. Diagnosis, evaluation and management of acute kidney injury: a KDIGO summary (Part1). Crit Care.2013; 17(1): 204.

48. Xu LB, Zhu LG, Liu GP, Yu L. A single-center retrospective study on the epidemiology of acute kidney injury in elderly patients in Inner Mongolia. Chinese Journal of Integrative Nephropathy. 2020; 21(10): 871-874.

49. Lin DX, Du B. Fluid management of acute kidney injury. J Cardiopulmonary and Vascular Diseases. 2019; 38(11): 1188-1191.

50. Grangeon-Chapon C, Dodoi M, Esnault VL,Favre G. Osmotic stress and mortality in elderly patients with kidney failure: a retrospective study. Clin Interv Aging. 2019; 14: 225-229.

51. Sato R, Luthe SK, Nasu M. Blood pressure and acute kidney injury. Crit Care. 2017; 21(1): 28.

52. Rosner MH, Perazella MA. Acute kidney injury in the patient with cancer. Kidney Res Clin Pract. 2019; 38(3): 295-308.

53. Collins GS, Reitsma JB, Altman DG, Moons KG. A step-by-step guide to developing and validating clinical prediction models. Journal of Clinical Epidemiology. 2024; 150: 105-120.