Journal of Cardiothoracic and Vascular Anesthesia 39 (2025) 666–674

Contents lists available at ScienceDirect

Journal of Cardiothoracic and Vascular Anesthesia

journal homepage: www.jcvaonline.com

Original Article

Machine Learning for In-hospital Mortality Prediction in Critically Ill Patients With Acute Heart Failure: A Retrospective Analysis Based on the MIMIC-IV Database

Jun Li, BSc*,#, Yiwu Sun, MD*,#, Jie Ren, MD†, Yifan Wu, MD‡, Zhaoyi He, MD§,#,[1]

*Department of Anesthesiology, Dazhou Central Hospital, Dazhou, Sichuan, China
†Department of Anesthesiology, Guizhou Provincial People's Hospital, Guiyang, Guizhou, China
‡Department of Anesthesiology, Shanghai Sixth People's Hospital, Shanghai, China
§Surgical Anesthesia Center, The First People's Hospital of Longquanyi District, Chengdu, China

Background: The incidence, mortality, and readmission rates for acute heart failure (AHF) are high, and the in-hospital mortality for AHF patients in the intensive care unit (ICU) is higher. However, there is currently no method to accurately predict the mortality of AHF patients.

Methods: The Medical Information Mart for Intensive Care Ⅳ (MIMIC-Ⅳ) database was used to perform a retrospective study. Patients meeting the inclusion criteria were identified from the MIMIC-Ⅳ database and randomly divided into a training set (n = 3,580, 70%) and a validation set (n = 1,534, 30%). The variates collected include demographic data, vital signs, comorbidities, laboratory test results, and treatment information within 24 hours of ICU admission. By using the least absolute shrinkage and selection operator (LASSO) regression model in the training set, variates that affect the in-hospital mortality of AHF patients were screened. Subsequently, in the training set, five common machine learning (ML) algorithms were applied to construct models using variates selected by LASSO to predict the in-hospital mortality of AHF patients. The predictive ability of the models was evaluated for sensitivity, specificity, accuracy, the area under the curve of receiver operating characteristics, and clinical net benefit in the validation set. To obtain a model with the best predictive ability, the predictive ability of common scoring systems was compared with the best ML model.

Results: Among the 5,114 patients, in-hospital mortality was 12.5%. Comparing the area under the curve, the XGBoost model had the best predictive ability among all ML models, and the XGBoost model was chosen as the final model for its higher net benefit. Its predictive ability was superior to common scoring systems.

Conclusions: The XGBoost model can effectively predict the in-hospital mortality of AHF patients admitted to the ICU, which may assist clinicians in precise management and early intervention for patients with AHF to reduce mortality.

Key Words: prediction model; machine learning; acute heart failure; intensive care unit; in-hospital mortality; MIMIC-IV database

ACUTE HEART FAILURE (AHF) is a clinical syndrome characterized by dyspnea or exertional limitation due to impairment of ventricular filling or ejection of blood or both.[1] In addition, heart failure (HF) has high incidence, mortality, and readmission rates, affecting more than 64 million people worldwide.[2,3] Research showed that between 2009 to 2012 and 2013 to 2016, the prevalence of HF among American adults increased from 5.7 million to 6.2 million, and the annual incidence rate of HF among adults over 55 years old in the

[1]*Address correspondence to* Zhaoyi He, Surgical Anesthesia Center, The First People's Hospital of Longquanyi District, No. 43, Yidudonglu, Chengdu, China.

E-mail address: 18215605376@126.com (Z. He).

[#]These authors contributed equally to this work.

https://doi.org/10.1053/j.jvca.2024.12.016

United States increased from 870,000 from 2005 to 2011 to 1 million in 2014.[4,5] It is expected that by 2030, the estimated prevalence of HF will increase by 24%, reaching approximately 8.5 million.[6,7] In the United States, approximately 10% to 51% of HF inpatients are admitted to the intensive care unit (ICU), and compared with patients in the general medical ward, patients admitted to the ICU have a significantly higher adjusted in-hospital mortality rate.[8,9,10] It was reported that the in-hospital mortality rate for patients treated in the ICU was 10.6% compared with 4.0% for all HF patients.[11]

Machine learning (ML) belongs to the category of artificial intelligence, and the strength of ML algorithms lies in their ability to automatically learn patterns from large datasets for predictive analysis, allowing users to collect knowledge from past data and apply it to future predictions. Compared with traditional prediction models based on linear parameter models, ML is not only applicable in highly complex and nonlinear environments, but also suitable for processing big data.[12] ML has already been applied in detection and diagnosis, treatment, and outcome prediction and prediction evaluation,[13] for example, in the diagnosis of diabetic retinopathy from retinal fundus photographs and breast cancer.[14,15]

Although previous studies have established predictive models for mortality in AHF patients, the area under the curve (AUC) of receiver operating characteristics (ROC) of the best models constructed by these studies were 0.764 and 0.720, respectively.[16,17] Applying these models for prediction cannot obtain accurate results, and the models still need to be optimized. This study was designed to develop and validate multiple ML models based on clinical features, to find the model with the best predictive performance for predicting in-hospital mortality in AHF patients admitted to the ICU.

Materials and Methods

Study Design and Data Source

A retrospective analysis was conducted using all relevant data extracted from the Medical Information Mart for Intensive Care Ⅳ (MIMIC-IV) database,[18] which consists of comprehensive and high-quality data on patients admitted to the ICU at the Beth Israel Deaconess Medical Center between 2008 and 2019. To access the database, one author (S.Y.W.) passed the National Institutes of Health Protecting Human Research Participants web-based training course and obtained approval to extract data from the MIMIC-IV for research purposes (Certification Number: 50778029).

Study Patients

When patients were diagnosed with AHF using the ICD-9 codes 42821, 42831, 42841, 42823, 42833, and 42843, or ICD-10 codes I5021, I5031, I5041, I50811, I5023, I5033, I5041, I5043 and I50813, patient eligibility was considered. Patients who were ≥18 years old at the time of ICU admission were included in the study; Only the first ICU admission was included for patients with multiple ICU admissions. Patients without an ICU record were excluded from the study. The flow chart in Figure 1 showed the selection of patients. Finally, 5,114 adult patients were included in this study.

Data Extraction and Processing

In the MIMIC-IV cohort, based on previous research,[16,19] clinical relevance, and general availability, clinical and laboratory variates were extracted within the first day of ICU admission and included: demographic characteristics, vital signs, comorbidities and laboratory variates, treatment information, and scoring systems. A total of 143 variates were extracted, and to avoid bias caused by a large number of missing values, variates with more than 30% missing value were directly excluded[20] (Missing value details in Supplementary Table 1.) For variates with missing values equaling less than 30%, multiple imputation was used to fill data imputation by R software ("mice" package). Ultimately, 54 variates were obtained for further analysis. In-hospital mortality of the AHF patients was also extracted as an outcome.

Missing Data

Variates with missing data totaling >30% were directly excluded. Variates with missing data totaling <30%were processed by multiple imputation using the "mice" package in R software. Multiple imputation is a two-stage approach. Missing values were input a number of times using a statistical model based on the available data, producing several datasets from which parameters could be estimated. These parameters were combined to provide an effective estimate of the parameters.[21,22]

ML Model Building and Assessment

The dataset was divided into a training set and a validation set, and models were constructed in the training set. Considering there were still many variates in the training set, to effectively prevent overfitting, variates were further selected by using the least absolute shrinkage selection operator (LASSO) regression. LASSO is a regression-based methodology, which was conducted via a continuous shrinking operation and minimizing regression coefficients, to reduce the likelihood of overfitting. LASSO has the unique feature of penalizing the absolute value of a regression coefficient. The greater the penalization, the greater the shrinkage of coefficients, with some reaching 0, thus automatically removing unnecessary and/or noninfluential covariates.[21,23] Finally, the variates selected through LASSO were used to construct the models in the training set.

In this study, five common algorithms, including logistic regression (LR), k-nearest neighbor (KNN), eXtreme gradient boosting (XGBoost), decision tree (DT), and random forest (RF), were applied to construct models for predicting in-hospital mortality of AHF patients in the ICU. The training set was used for model establishment and adjustment (by a grid

search), while the validation set was used for model evaluation. The model with the best predictive performance was determined by comparing the AUC values of these models in the validation set and the clinical net benefit of these models through decision curve analysis (DCA).

Statistical Analysis

According to the normality of the distribution, continuous variates are described as mean ± standard deviation (SD) or median and interquartile range (IQR), which were compared by Student's *t*-test or the Mann–Whitney U test. The categorical variates are expressed as a percentage of the total, and were compared by Pearson's chi-squared test or Fisher's exact test.

The LASSO penalization method was used to further screen the predictive variates. By conducting tenfold cross-validation on the established LASSO model, a key parameter (lambda) was determined, and variates with predictive significance were screened out.

After that, the patients were randomly divided into two groups, of which 70% were used as the training set and the remaining 30% as the validation set. Five common ML methods were applied to develop the models in the training set and validate them in the validation set. The performance of the models was assessed by comparing the AUC and DCA in the validation set. Finally, the model with the best predictive ability was obtained.

Sequential Organ Failure Assessment, Logistic Organ Dysfunction System, and Simplified Acute Physiology Score II scores are commonly used tools to evaluate the severity and prognosis of diseases in ICU patients, and their predictive ability was compared with the best ML model.

All analyses were performed using the statistical software packages R version 4.2.2 (http://www.R-project.org, The R Foundation). p-Values < 0.05 (two-sided test) were considered statistically significant.

Results

Baseline Characteristics

Figure 1 shows a flow chart describing the procedure for subject selection. As shown in Figure 1, 5,114 AHF patients were included. Among the included patients, 637 died (12.5%) and 4,477 survived during in-hospital. Table 1 summarized the comparison of baseline characteristics, vital signs, and laboratory parameters between deaths and survivors during in-hospital. There were significant differences between the death group and the survival group, with the death group having older age, lower urine output, higher creatinine and blood urea nitrogen (BUN) values, faster heart rate (HR), lower blood pressure, faster respiratory rate (RR), lower blood oxygen saturation, and a higher proportion of vasopressor drug use. At the same time, all patients were also randomly divided into a training set (70%, 3,580) and a validation set (30%, 1,534) for subsequent model construction and validation. Between the two datasets, except for the sbp_max, mbp_max, and the proportion of patient with chronic obstructive pulmonary disease, there were no significant differences in other indicators.

Features Selected in Models

The LASSO regularization process screened 18 potential predictive factors based on 3,580 patients in the training set

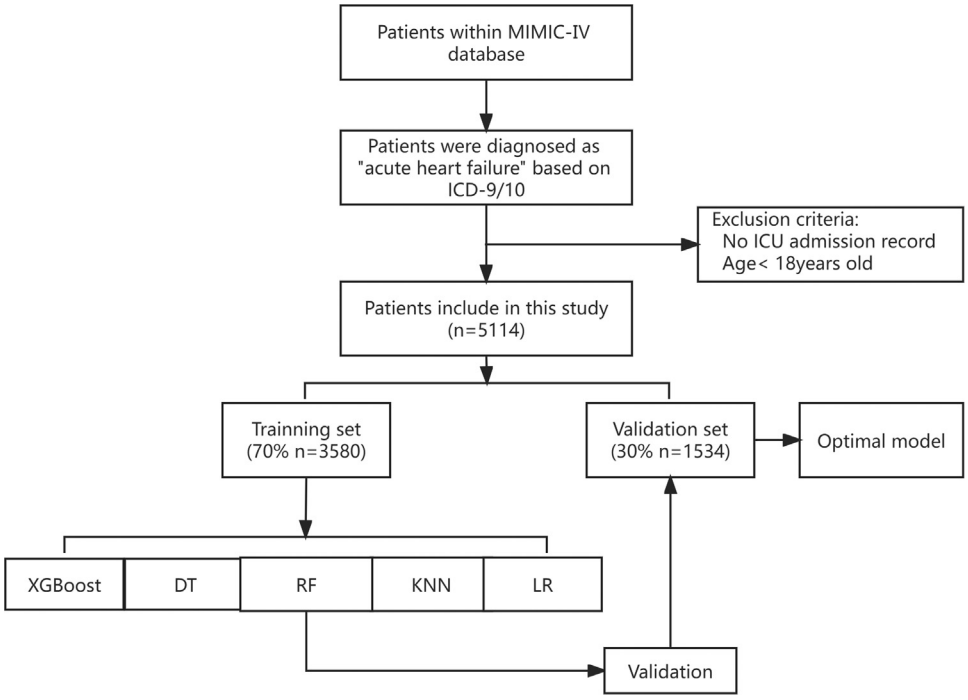

Fig 1. Flowchart of patient selection.

DT, decision tree; ICD-9/10, 9/10th revision of the International Classification of Diseases; ICU, intensive care unit; KNN, k-nearest neighbor; LR, logistic regression; MIMIC-Ⅳ, Medical Information Mart for Intensive Care Ⅳ; RF, random forest; XGBoost, eXtreme gradient boosting.

Table 1
Characteristics of Participants

	Total (n = 5,114)	Alive (n = 4,477)	Died (n = 637)	p-Value	Training set (n = 3,580)	Validation set (n = 1,534)	p-Value
Age, mean (SD)	73.17 (13.82)	72.59 (13.98)	77.24 (11.89)	<0.001	73.04 (13.90)	73.45 (13.64)	0.34
Male, mean (SD)	0.54 (0.50)	0.54 (0.50)	0.56 (0.50)	0.38	0.54 (0.50)	0.54 (0.50)	0.695
Platelets_min, mean (SD)	197.64 (98.70)	198.91 (96.58)	188.69 (112.19)	0.014	198.39 (99.74)	195.88 (96.26)	0.405
Platelets_max, mean (SD)	233.85 (109.55)	234.52 (106.75)	229.15 (127.49)	0.247	234.27 (110.78)	232.88 (106.67)	0.678
Wbc_min, mean (SD)	10.87 (6.62)	10.56 (5.99)	13.11 (9.70)	<0.001	10.84 (6.44)	10.94 (7.03)	0.622
Wbc_max, mean (SD)	14.57 (9.56)	14.20 (8.87)	17.19 (13.18)	<0.001	14.51 (9.09)	14.71 (10.58)	0.48
Aniongap_min, mean (SD)	13.63 (3.48)	13.40 (3.25)	15.26 (4.43)	<0.001	13.66 (3.46)	13.55 (3.50)	0.272
Aniongap_max, mean (SD)	17.06 (4.58)	16.71 (4.25)	19.45 (5.90)	<0.001	17.12 (4.68)	16.91 (4.34)	0.146
BUN_min, mean (SD)	31.46 (21.90)	29.96 (20.57)	42.00 (27.37)	<0.001	31.59 (22.27)	31.16 (21.02)	0.521
BUN_max, mean (SD)	36.44 (24.20)	34.75 (22.87)	48.33 (29.41)	<0.001	36.54 (24.31)	36.20 (23.95)	0.644
Creatinine_min, mean (SD)	1.52 (1.28)	1.48 (1.27)	1.83 (1.33)	<0.001	1.53 (1.34)	1.50 (1.13)	0.368
Creatinine_max, mean (SD)	1.80 (1.52)	1.74 (1.50)	2.20 (1.58)	<0.001	1.81 (1.57)	1.78 (1.39)	0.562
INR_min, mean (SD)	1.50 (0.78)	1.46 (0.76)	1.74 (0.92)	<0.001	1.50 (0.80)	1.48 (0.75)	0.291
INR_max, mean (SD)	1.79 (1.39)	1.73 (1.31)	2.27 (1.83)	<0.001	1.80 (1.40)	1.78 (1.39)	0.611
PT_min, mean (SD)	16.29 (7.91)	15.96 (7.67)	18.65 (9.09)	<0.001	16.38 (8.06)	16.10 (7.57)	0.243
PT_max, mean (SD)	19.35 (13.81)	18.67 (12.84)	24.15 (18.59)	<0.001	19.42 (13.92)	19.17 (13.55)	0.548
PTT_min, mean (SD)	34.67 (15.77)	34.07 (15.25)	38.92 (18.54)	<0.001	34.79 (15.86)	34.38 (15.57)	0.391
PTT_max, mean (SD)	54.49 (38.22)	53.07 (37.24)	64.50 (43.26)	<0.001	55.16 (38.85)	52.94 (36.67)	0.057
Urine output, mean (SD)	1,900.86 (1,399.54)	1,991.61 (1,393.81)	1,263.05 (1,268.50)	<0.001	1,908.21 (1,414.12)	1,883.70 (1,365.21)	0.566
Heart rate_min, mean (SD)	71.91 (15.98)	71.73 (15.47)	73.19 (19.20)	0.032	72.00 (15.95)	71.71 (16.08)	0.552
Heart rate_max, mean (SD)	104.70 (22.36)	104.06 (22.10)	109.14 (23.65)	<0.001	104.86 (22.34)	104.33 (22.39)	0.439
Heart rate_mean, mean (SD)	86.26 (16.92)	85.74 (16.60)	89.90 (18.67)	<0.001	86.34 (16.90)	86.07 (16.97)	0.597
Sbp_min, mean, SD)	89.48 (16.94)	90.57 (16.41)	81.81 (18.58)	<0.001	89.26 (16.95)	89.99 (16.93)	0.157
Sbp_max, mean (SD)	144.68 (24.16)	145.36 (23.82)	139.90 (25.96)	<0.001	144.23 (23.54)	145.73 (25.52)	0.042
Sbp_mean, mean (SD)	115.43 (16.44)	116.36 (16.27)	108.87 (16.12)	<0.001	115.25 (16.40)	115.86 (16.53)	0.22
Dbp_min, mean (SD)	44.63 (11.46)	45.22 (11.26)	40.50 (12.02)	<0.001	44.65 (11.54)	44.58 (11.29)	0.84
Dbp_max, mean (SD)	88.17 (21.05)	88.32 (20.90)	87.09 (22.07)	0.167	87.86 (20.24)	88.88 (22.84)	0.112
Dbp_mean, mean (SD)	61.85 (11.45)	62.21 (11.48)	59.36 (10.94)	<0.001	61.90 (11.41)	61.74 (11.53)	0.639
Mbp_min, mean (SD)	57.00 (13.26)	57.76 (12.83)	51.63 (14.94)	<0.001	56.88 (13.47)	57.28 (12.76)	0.315
Mbp_max, mean (SD)	104.77 (27.21)	104.75 (26.77)	104.90 (30.15)	0.895	104.25 (26.29)	105.99 (29.23)	0.036
Mbp_mean, mean (SD)	76.61 (10.66)	77.07 (10.60)	73.40 (10.52)	<0.001	76.59 (10.60)	76.64 (10.79)	0.88
RR_min, mean (SD)	13.30 (3.70)	13.25 (3.62)	13.63 (4.23)	0.017	13.30 (3.73)	13.28 (3.64)	0.867
RR_max, mean (SD)	29.49 (6.45)	29.30 (6.35)	30.77 (6.95)	<0.001	29.50 (6.44)	29.46 (6.47)	0.856
Respiratory rate_mean, mean (SD)	20.53 (3.85)	20.38 (3.74)	21.58 (4.38)	<0.001	20.55 (3.84)	20.50 (3.87)	0.693
Temperature_ mean, mean (SD)	36.77 (0.51)	36.78 (0.49)	36.73 (0.66)	0.044	36.78 (0.52)	36.75 (0.50)	0.125
SpO_2_min, mean (SD)	90.41 (6.37)	90.87 (5.32)	87.21 (10.71)	<0.001	90.44 (6.41)	90.36 (6.28)	0.686
SpO_2_max, mean (SD)	99.33 (1.36)	99.34 (1.17)	99.25 (2.28)	0.109	99.35 (1.31)	99.30 (1.45)	0.321
SpO_2_mean, mean (SD)	96.33 (2.32)	96.41 (2.03)	95.76 (3.73)	<0.001	96.34 (2.31)	96.29 (2.34)	0.475
Glucose_mean, mean (SD)	360.92 (8153.17)	296.26 (7042.84)	815.39 (13604.70)	0.133	358.55 (7997.25)	366.47 (8508.61)	0.975
Prior HF (%)	1931 (37.8)	1707 (38.1)	224 (35.2)	0.162	1337 (37.3)	594 (38.7)	0.369
COPD (%)	279 (5.5)	251 (5.6)	28 (4.4)	0.244	180 (5.0)	99 (6.5)	0.047
CHD (%)	1312 (25.7)	1178 (26.3)	134 (21.0)	0.005	916 (25.6)	396 (25.8)	0.892
DM (%)	650 (12.7)	581 (13.0)	69 (10.8)	0.145	446 (12.5)	204 (13.3)	0.435
HBP (%)	1546 (30.2)	1380 (30.8)	166 (26.1)	0.016	1076 (30.1)	470 (30.6)	0.702
MI (%)	1563 (30.6)	1332 (29.8)	231 (36.3)	0.001	1117 (31.2)	446 (29.1)	0.139
Dobutamine (%)	315 (6.2)	203 (4.5)	112 (17.6)	<0.001	222 (6.2)	93 (6.1)	0.9
Dopamine (%)	341 (6.7)	238 (5.3)	103 (16.2)	<0.001	242 (6.8)	99 (6.5)	0.733
Epinephrine (%)	528 (10.3)	428 (9.6)	100 (15.7)	<0.001	371 (10.4)	157 (10.2)	0.93
Norepinephrine (%)	1397 (27.3)	1014 (22.6)	383 (60.1)	<0.001	973 (27.2)	424 (27.6)	0.76
Phenylephrine (%)	997 (19.5)	812 (18.1)	185 (29.0)	<0.001	694 (19.4)	303 (19.8)	0.791
Lods, mean (SD)	5.54 (3.29)	5.09 (2.96)	8.73 (3.70)	<0.001	5.54 (3.29)	5.55 (3.31)	0.887
SAPSII, mean (SD)	39.62 (13.14)	37.95 (11.99)	51.33 (14.79)	<0.001	39.61 (13.14)	39.65 (13.15)	0.918
SOFA, mean (SD)	5.09 (3.32)	4.76 (3.11)	7.35 (3.82)	<0.001	5.07 (3.32)	5.14 (3.31)	0.462
In_hospital mortality (%)	637 (12.5)	0 (0.0)	637 (100.0)	<0.001	446 (12.5)	191 (12.5)	1

Data are presented as %, mean ± standard deviation, or median (interquartile range).
Abbreviations: BUN, blood urea nitrogen; CHD, coronary heart disease; COPD, chronic obstructive pulmonary disease; Dbp, diastolic blood pressure, DM, diabetes mellitus; HBP, hypertension; HF, heart failure; INR, international normalized ratio; LODS, Logistic Organ Dysfunction System; Mbp, mean blood pressure; MI, myocardial infarction; PT, prothrombin time; PTT, partial thromboplastin time; RR, respiratory rate; SAPSII, Simplified Acute Physiology Score II; Sbp, systolic blood pressure; SOFA, Sequential Organ Failure Assessment; SpO_2, saturation pulse oxygen; Wbc, white blood cell.

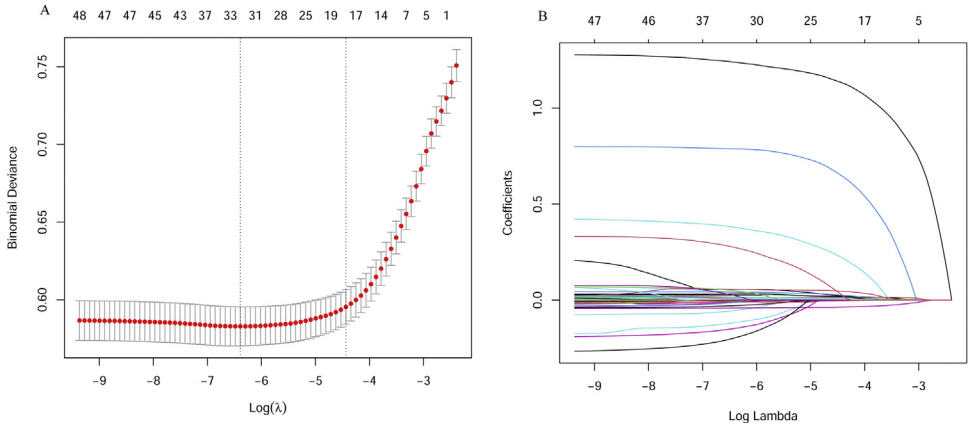

Fig 2. Feature selection using the least absolute shrinkage and selection operator (LASSO) binary logistic regression model. (A) Tuning parameter (λ) selection in the LASSO model used tenfold cross-validation via minimum criteria. The partial likelihood deviance (binomial deviance) curve was plotted versus log(λ). Dotted vertical lines were drawn at the optimal values by using the minimum criteria and the one SE of the minimum criteria (the 1-SE criteria). λ value of 0.011845 was chosen (1-SE criteria) according to tenfold cross-validation. (B) LASSO coefficient profiles of the 54 features. A coefficient profile plot was produced against the log(λ) sequence. The vertical line was drawn at the value selected using tenfold cross-validation, where optimal resulted in 18 features with nonzero coefficients.

(Fig 2A, B). Figure 2B shows the LASSO-selected predictors (shrinkage parameter λ = 0.011845).

Development and Comparison of ML Models

Within the first 24 hours of admission to the ICU, after excluding variates with missing values >30%, a total of 54 clinical features were collected. Multiple imputation was used to fill in missing data. These results are shown in Table 1. LASSO regression was used to further screen for characteristic variates of in-hospital mortality in AHF patients. Finally, 18 variates were selected for model construction, including: age, WBC_min, Aniongap_min, Aniongap_max, BUN_min, INR_min, INR_max, PTT_min, Urine output, HR_mean, SBP_mean, DBP_min, RR_mean, SpO_2_min, dobutamine, dopamine, norepinephrine, phenylephrine. We selected five ML methods (including XGBoost, KNN, RF, DT, and LR) based on the training set to construct five ML models to predict the in-hospital mortality risk of AHF patients (Fig 3A).

The performance of the models was evaluated using the validation set. Figure 3B and Table 2 describe the performance of these prediction models, indicating that, compared with other ML models, the XGBoost model is relatively better than the other four models (AUC: XGBoost, 0.82; DT, 0.61; RF, 0.81; LR, 0.81; KNN, 0.76). The DCA curve (Fig 5), also indicated that the XGBoost model performed better. Therefore, in this study, the XGBoost model was chosen as the final model. Meanwhile, the XGBoost model was also compared with common scoring systems for disease severity and prognosis in the ICU, such as the SOFA, LODS, and SAPSII. The results are shown in Figure 4A, B, and Table 3. The results indicated that XGBoost had better predictive ability than common scoring systems (AUC: XGBoost, 0.82; LODS, 0.78; SAPSⅡ, 0.75; SOFA, 0.70). Calibration curves were also plotted for each model in the validation set. Calibration curves depict the calibration of the model in terms of the agreement between the predicted risk of in-hospital mortality and observed in-hospital mortality. The closer the dashed line is to the solid line, the

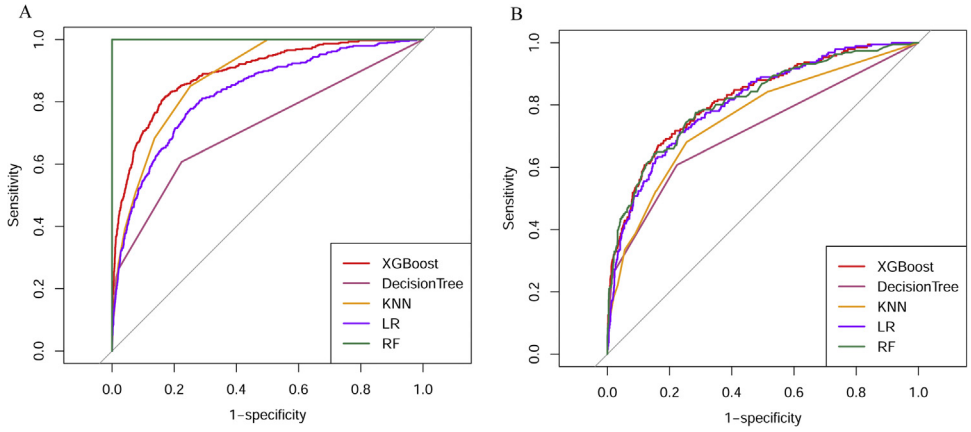

Fig 3. Receiver operator characteristic (ROC) curves for the machine learning (ML) models predict in-hospital mortality (training set and validation set). ROC curves for five ML models predicting in-hospital mortality in the training set (A) and validation set (B), respectively. DT, decision tree; KNN, k-nearest neighbor; LR, logistic regression; RF, random forest; XGBoost, eXtreme gradient boosting.

Table 2
Predictive Performances of the Five Machine Learning Models for Predicting In-hospital Mortality

Model	AUC	95% CI	Accuracy	Sensitivity	Precision	F1 Score	Specificity
XGB	0.82	0.78-0.85	0.90	0.99	0.90	0.94	0.25
DT	0.61	0.57-0.64	0.88	0.98	0.90	0.94	0.20
RF	0.81	0.78-0.85	0.89	1.00	0.89	0.94	0.15
LR	0.81	0.78-0.84	0.88	0.98	0.90	0.94	0.23
KNN	0.76	0.72-0.80	0.88	0.99	0.89	0.94	0.12

Abbreviations: AUC, area under the curve; DT, decision tree; KNN, k-nearest neighbor; LR, logistic regression; RF, random forest; XGBoost, eXtreme gradient boosting.

better the accuracy of the model. The XGBoost model had the best accuracy (Supplementary Figure 1).

Discussion

Given the critical condition of AHF patients, the risk of admission to the ICU is relatively high. If physicians can accurately predict patient prognosis, they can better choose treatment strategies and allocate medical resources. This study used data from the MIMIC-IV database to develop and validate five prediction models based on the ML algorithm for predicting the in-hospital mortality rate of AHF patients in the ICU. To avoid overfitting, LASSO regression analysis was used to screen for independent risk factors for in-hospital mortality before constructing a ML model using the collected data.

The results indicated that the XGBoost model performed best in AHF prediction, which was consistent with some previous studies.[24,25] XGBoost is a ML technique with the remarkable features of processing missing data efficiently and flexibly and assembling weak prediction models to build an accurate one.[26] The XGBoost model had better predictive ability than the LR model in predicting in-hospital mortality in AHF patients. The LR algorithm cannot detect complex nonlinear relationships and interactions between independent and dependent variates, and is sensitive to multicollinearity of independent variates, which may lead to inaccurate models. The XGBoost model can effectively address the shortcomings of LR and develop more accurate prediction models. Therefore, the application of the XGBoost algorithm is increasing.[27]

Finally, the XGBoost model was selected as the best predictive model and its predictive ability was compared with traditional ICU disease severity scores. The results showed that the XGBoost model had better predictive ability. The importance of model variates based on the XGBoost algorithm were also ranked, and the top ten variates were: norepinephrine, urine output, age, white blood cell (Wbc)_min, saturation pulse oxygen (SpO_2)_min, RR_mean, BUN_min, systolic blood pressure (Sbp)_mean, partial thromboplastin time (PTT)_min, and HR_mean (Fig 6).

Norepinephrine is a highly effective and reliable vasopressor with many advantages, including: (1) raising blood pressure effectively, (2) increasing the cardiac index without increasing HR and without excessively increasing myocardial oxygen consumption, and (3) not acting on β-2 receptors without increasing lactate levels. Research has shown that the level of plasma norepinephrine is an important marker for predicting mortality in HF patients.[28,29] In the current study, the use of norepinephrine was also an important predictor of mortality in AHF patients. This may be due to the poor cardiac function of the patients, and the insufficient enhancement of sympathetic nervous activity to improve cardiac function, requiring exogenous supplementation of norepinephrine.

A large proportion of HF patients were admitted to the hospital due to worsening fluid overload.[5,30] Urination can reduce

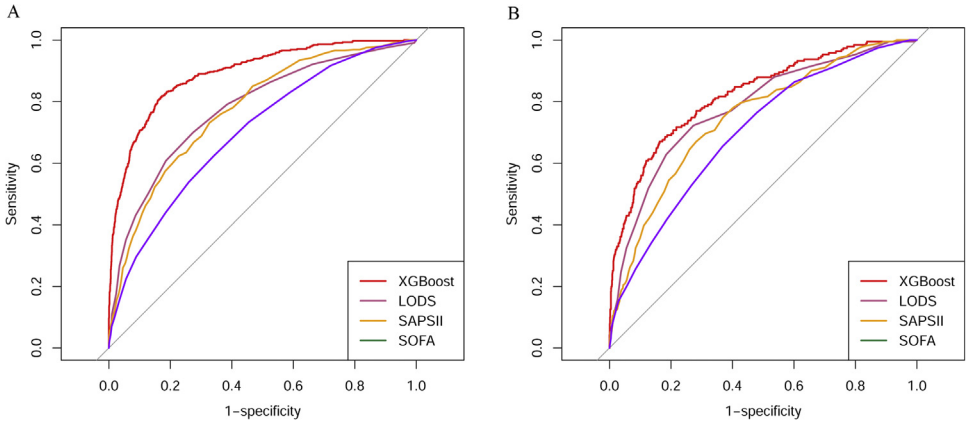

Fig 4. Receiver operator characteristic (ROC) curves for the XGBoost model and ROC curves for common scoring systems predict in-hospital mortality (training set and validation set). ROC curves for four models predicting in-hospital mortality in the training set (A) and validation set (B), respectively. DT, decision tree; KNN, k-nearest neighbor; LR, logistic regression; RF, random forest; XGBoost, eXtreme gradient boosting.

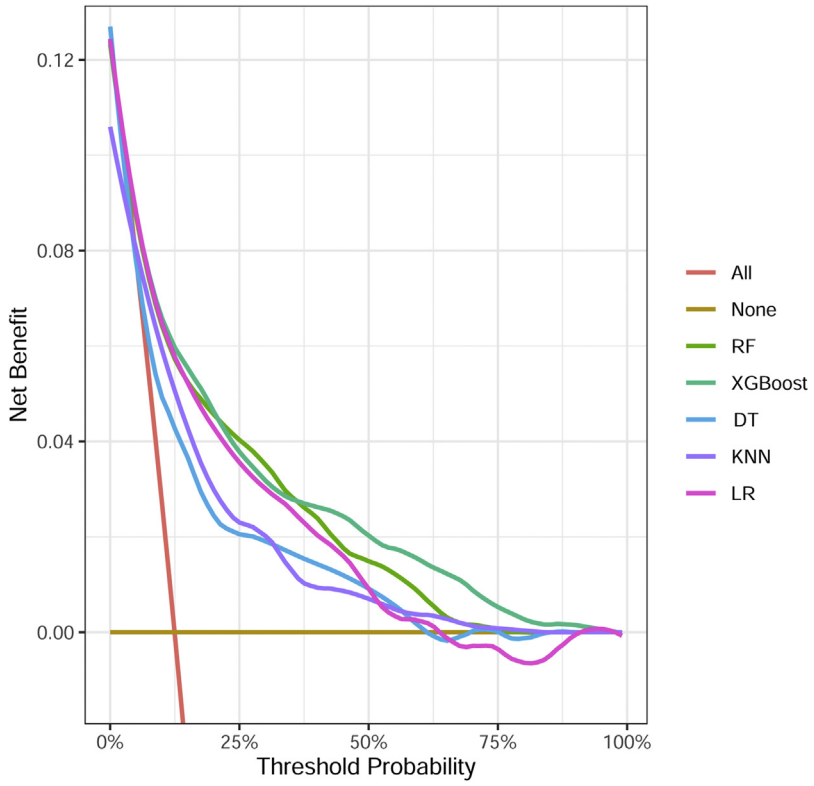

Fig 5. Decision curve analysis for machine learning (ML) models. The y-axis measures the net benefit. The "All" line represents the assumption that all patients die in the hospital. The "None" line represents the assumption that no patients die in the hospital. DT, decision tree; KNN, k-nearest neighbor; LR, logistic regressio; RF, random forest; XGBoost, eXtreme gradient boosting.

volume load. What's more, urine output can reflect the patient's organ perfusion. There are also many studies that have shown that urine output is an important variate in predicting the risk of death in patients.[31,32] This is consistent with the current research findings. However, in the data extracted, the proportion of patients with renal function impairment or renal failure was very small. Although these impairments were also associated with in-hospital mortality, they were not considered independent variates to avoid bias in the results. Further analysis should be conducted in subsequent studies.

HF is a common condition in older adults that results from the complex interplay of age-related diseases and age-associated physiologic changes.[33] First, age is associated with reduced responsiveness to beta-adrenergic stimulation, which causes maximum HR declines [34]; Second, aging alters left ventricular diastolic filling, which effects cardiac filling[35]; the change in age also directly leads to the development or deterioration of HF.[33] In this study, age was also an important variate affecting the in-hospital mortality of HF patients.

The number of Wbc is a cellular marker of systemic inflammation. Compared with the level of specific inflammatory markers that only focus on certain immune systems, Wbc counts reflected the overall activity of the immune system.[36] Studies have shown that Wbc count is associated with HF risk.[37] Moreover, Wbc count is an indicator reflecting overall immune system activity and has been reported to be associated with the risk of short- and long-term mortality of HF patients.[38,39] In this study, Wbc was also shown to be an important variate in predicting in-hospital mortality in AHF patients.

SpO_2, RR, BUN, systolic blood pressure, HR, and partial thromboplastin time have also been proven to be significant predictors of mortality in HF patients in other studies.[16,19,40] This is also consistent with the current research findings.

Table 3
Predictive Performances of the XGBoost Model and Common Scoring Systems for Predicting In-hospital Mortality

Model	AUC	95% CI	Accuracy	Sensitivity	Precision	F1 Score	Specificity
XGB	0.82	0.78-0.85	0.90	0.99	0.90	0.94	0.25
LODS	0.78	0.74-0.82	0.88	0.99	0.89	0.93	0.11
SAPSII	0.75	0.71-0.78	0.88	0.99	0.89	0.93	0.12
SOFA	0.70	0.66-0.74	0.88	1.00	0.89	0.93	0.03

AUC, area under the curve; LODS, Logistic Organ Dysfunction System; SAPSII, Simplified Acute Physiology Score II; SOFA, Sequential Organ Failure Assessment; XGBoost, eXtreme gradient boosting.

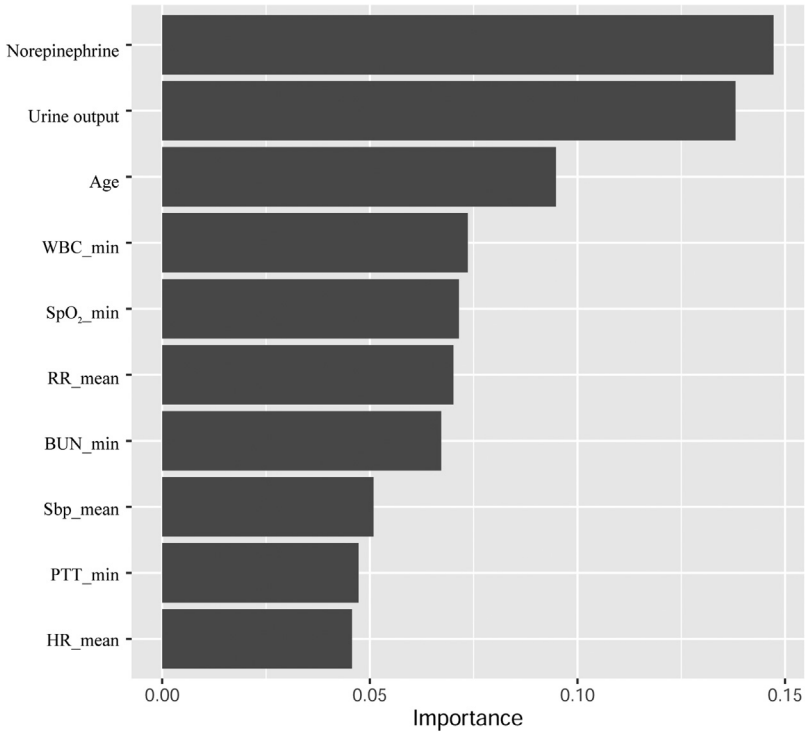

Fig 6. Importance ranking of variates. BUN_min, blood urea nitrogen minimum; HR_mean, heart rate mean; PTT_min, partial thromboplastin time minimum; RR_mean, respiratory rate mean; Sbp_mean, systolic blood pressure mean; SpO_2_min, saturation pulse oxygen minimum; WBC_min, white blood cell minimum.

Some limitations of this study must be considered: First, this was a retrospective study, which may lead to inevitable selection bias. Second, as the data only came from the MIMIC-IV database, this may affect the extension of the prediction model to other populations. Therefore, it is necessary to conduct large-scale, multicenter research for external validation. Third, multiple imputation methods were used to fill in missing data, which may lead to deviations from the true values. Fourth, although a XGBoost model with good predictive ability was constructed, the model lacks interpretability due to the "black box" of ML. However, the authors believe that the constructed model is helpful for clinicians to effectively treat AHF patients.

Conclusion

In conclusion, this study demonstrated that ML based on the XGboost algorithm is indeed superior to traditional LR and common scoring systems, which may assist clinicians in precise management and early intervention of patients with AHF to reduce mortality.

Declaration of competing interest

The authors declare that they have no conflicts of interest.

CRediT authorship contribution statement

Jun Li: Formal analysis, Data curation, Conceptualization. **Yiwu Sun:** Writing – original draft, Formal analysis, Data curation, Conceptualization. **Jie Ren:** Conceptualization. **Yifan Wu:** Writing – original draft. **Zhaoyi He:** Methodology, Formal analysis, Data curation, Conceptualization.

Ethics Approval and Consent to Participate

The Institutional Review Board at the Beth Israel Deaconess Medical Center granted a waiver of informed consent and approved the data-sharing initiative. After passing the "Protecting Human Research Participants" exam, we were granted access to the MIMIC-IV database (Certification Number: 50778029). The study was designed and conducted in accordance with relevant guidelines and regulations (Declaration of Helsinki).

Consent for Publication

Not applicable.

Availability of Data and Material

The data that support this study are available from the MIMIC-IV database (https://physionet.org/content/mimiciv/), but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. However, data are available from the corresponding author upon reasonable request.

Funding

None.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1053/j.jvca.2024.12.016.

References

1 Murphy SP, Ibrahim NE, Januzzi JL Jr.. Heart failure with reduced ejection fraction: A review. JAMA 2020;324:488–504.
2 Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: A systematic analysis for the Global Burden of Disease Study 2017. Lancet 2018;392:1789–858.
3 Conrad N, Judge A, Canoy D, et al. Temporal trends and patterns in mortality after incident heart failure: A longitudinal analysis of 86 000 individuals. JAMA Cardiol 2019;4:1102–11.
4 Virani SS, Alonso A, Benjamin EJ, et al. Heart disease and stroke statistics-2020 update: A report from the American Heart Association. Circulation 2020;141:e139–596.
5 Mozaffarian D, Benjamin EJ, Go AS, et al. Heart disease and stroke statistics—2015 update: A report from the American Heart Association. Circulation 2015;131:e29–322.
6 Heidenreich PA, Albert NM, Allen LA, et al. Forecasting the impact of heart failure in the United States: A policy statement from the American Heart Association. Circ Heart Fail 2013;6:606–19.
7 Benjamin EJ, Virani SS, Callaway CW, et al. Heart disease and stroke statistics-2018 Update: A report from the American Heart Association. Circulation 2018;137:e67–492.
8 Conrad N, Judge A, Tran J, et al. Temporal trends and patterns in heart failure incidence: A population-based study of 4 million individuals. Lancet 2018;391:572–80.
9 Safavi KC, Dharmarajan K, Kim N, et al. Variation exists in rates of admission to intensive care units for heart failure patients across hospitals in the United States. Circulation 2013;127:923–9.
10 van Diepen S, Bakal JA, Lin M, et al. Variation in critical care unit admission rates and outcomes for patients with acute coronary syndromes or heart failure among high- and low-volume cardiac hospitals. J Am Heart Assoc 2015;4:e001708.
11 Adams KF Jr., Fonarow GC, Emerman CL, et al. Characteristics and outcomes of patients hospitalized for heart failure in the United States: Rationale, design, and preliminary observations from the first 100,000 cases in the Acute Decompensated Heart Failure National Registry (ADHERE). Am Heart J 2005;149:209–16.
12 Olsen CR, Mentz RJ, Anstrom KJ, et al. Clinical applications of machine learning in the diagnosis, classification, and prediction of heart failure. Am Heart J 2020;229:1–17.
13 Rajkomar A, Dean J, Kohane I. Machine learning in medicine. New Engl J Med 2019;380:1347–58.
14 Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. JAMA 2016;316:2402–10.
15 Samala RK, Chan HP, Hadjiiski L, et al. Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography. Med Phys 2016;43:6654.
16 Peng S, Huang J, Liu X, et al. Interpretable machine learning for 28-day all-cause in-hospital mortality prediction in critically ill patients with heart failure combined with hypertension: A retrospective cohort study based on medical information mart for intensive care database-IV and eICU database. Front Cardiovasc Med 2022;9:994359.
17 Angraal S, Mortazavi BJ, Gupta A, et al. Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction. JACC Heart Fail 2020;8:12–21.
18 Goldberger AL, Amaral LA, Glass L, et al. PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals. Circulation 2000;101:E215–20.
19 Smith JG, Newton-Cheh C, Almgren P, et al. Assessment of conventional cardiovascular risk factors and multiple biomarkers for the prediction of incident heart failure and atrial fibrillation. J Am Coll Cardiol 2010;56:1712–9.
20 Hu C, Li L, Huang W, et al. Interpretable machine learning for early prediction of prognosis in sepsis: A discovery and validation study. Infect Dis Ther 2022;11:1117–32.
21 Sun Y, He Z, Ren J, et al. Prediction model of in-hospital mortality in intensive care unit patients with cardiac arrest: A retrospective analysis of MIMIC -IV database based on machine learning. BMC Anesthesiol 2023;23:178.
22 Lee KJ, Simpson JA. Introduction to multiple imputation for dealing with missing data. Respirology 2014;19:162–7.
23 McEligot AJ, Poynor V, Sharma R, et al. Logistic LASSO regression for dietary intakes and breast cancer. Nutrients 2020;12:2652.
24 Wang K, Tian J, Zheng C, et al. Interpretable prediction of 3-year all-cause mortality in patients with heart failure caused by coronary heart disease based on machine learning and SHAP. Comput Biol Med 2021;137:104813.
25 Li J, Liu S, Hu Y, et al. Predicting mortality in intensive care unit patients with heart failure using an interpretable machine learning model: Retrospective cohort study. J Med Internet Res 2022;24:e38082.
26 Hou N, Li M, He L, et al. Predicting 30-days mortality for MIMIC-III patients with sepsis-3: A machine learning approach using XGboost. J Transl Med 2020;18:462.
27 Yue S, Li S, Huang X, et al. Machine learning for the prediction of acute kidney injury in patients with sepsis. J Transl Med 2022;20:215.
28 Anand IS, Fisher LD, Chiang YT, et al. Changes in brain natriuretic peptide and norepinephrine over time and mortality and morbidity in the Valsartan Heart Failure Trial (Val-HeFT). Circulation 2003;107:1278–83.
29 Minatoguchi S. Heart failure and its treatment from the perspective of sympathetic nerve activity. J Cardiol 2022;79:691–7.
30 Costanzo MR, Jessup M. Treatment of congestion in heart failure with diuretics and extracorporeal therapies: Effects on symptoms, renal function, and prognosis. Heart Fail Rev 2012;17:313–24.
31 Heffernan AJ, Judge S, Petrie SM, et al. Association between urine output and mortality in critically ill patients: A machine learning approach. Crit Care Med 2022;50:e263–71.
32 Zhang L, Huang T, Xu F, et al. Prediction of prognosis in elderly patients with sepsis based on machine learning (random survival forest). BMC Emerg Med 2022;22:26.
33 Dharmarajan K, Rich MW. Epidemiology, pathophysiology, and prognosis of heart failure in older adults. Heart Fail Clinics 2017;13:417–26.
34 Fleg JL, Strait J. Age-associated changes in cardiovascular structure and function: A fertile milieu for future disease. Heart Fail Rev 2012;17:545–54.
35 Loffredo FS, Nikolova AP, Pancoast JR, et al. Heart failure with preserved ejection fraction: Molecular pathways of the aging myocardium. Circ Res 2014;115:97–107.
36 Horne BD, Anderson JL, John JM, et al. Which white blood cell subtypes predict increased cardiovascular risk? J Am Coll Cardiol 2005;45:1638–43.
37 Pfister R, Sharp SJ, Luben R, et al. Differential white blood cell count and incident heart failure in men and women in the EPIC-Norfolk study. Eur Heart J 2012;33:523–30.
38 Novack V, Pencina M, Zahger D, et al. Routine laboratory results and thirty day and one-year mortality risk following hospitalization with acute decompensated heart failure. PLoS ONE 2010;5:e12184.
39 Wu Q, Liu JH, Ma QH, et al. White blood cell count as a mediator of the relationship between depressive symptoms and all-cause mortality: A community-based cohort study. Arch Gerontol Geriatr 2021;94:104343.
40 deFilippi CR, de Lemos JA, Christenson RH, et al. Association of serial measures of cardiac troponin T using a sensitive assay with incident heart failure and cardiovascular mortality in older adults. JAMA 2010;304:2494–502.