Original Paper

Predicting Mortality in Intensive Care Unit Patients With Heart Failure Using an Interpretable Machine Learning Model: Retrospective Cohort Study

Jili Li[1*], BS; Siru Liu[2*], PhD; Yundi Hu[3], BDS; Lingfeng Zhu[4], BCS; Yujia Mao[1], BS; Jialin Liu[5], MD

[1]West China School of Medicine, Sichuan University, Chengdu, China

[2]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, TN, United States

[3]School of Data Science, Fudan University, Shanghai, China

[4]Department of Computer Science, Sichuan University, Chengdu, China

[5]Department of Medical Informatics, West China Hospital, Sichuan University, Chengdu, China

[*]these authors contributed equally

Corresponding Author:
Jialin Liu, MD
Department of Medical Informatics
West China Hospital
Sichuan University
No 37 Guoxue Road
Chengdu, 610041
China
Phone: 86 28 85422306
Fax: 86 28 85582944
Email: dljl8@163.com

Abstract

Background: Heart failure (HF) is a common disease and a major public health problem. HF mortality prediction is critical for developing individualized prevention and treatment plans. However, due to their lack of interpretability, most HF mortality prediction models have not yet reached clinical practice.

Objective: We aimed to develop an interpretable model to predict the mortality risk for patients with HF in intensive care units (ICUs) and used the SHapley Additive exPlanation (SHAP) method to explain the extreme gradient boosting (XGBoost) model and explore prognostic factors for HF.

Methods: In this retrospective cohort study, we achieved model development and performance comparison on the eICU Collaborative Research Database (eICU-CRD). We extracted data during the first 24 hours of each ICU admission, and the data set was randomly divided, with 70% used for model training and 30% used for model validation. The prediction performance of the XGBoost model was compared with three other machine learning models by the area under the curve. We used the SHAP method to explain the XGBoost model.

Results: A total of 2798 eligible patients with HF were included in the final cohort for this study. The observed in-hospital mortality of patients with HF was 9.97%. Comparatively, the XGBoost model had the highest predictive performance among four models with an area under the curve (AUC) of 0.824 (95% CI 0.7766-0.8708), whereas support vector machine had the poorest generalization ability (AUC=0.701, 95% CI 0.6433-0.7582). The decision curve showed that the net benefit of the XGBoost model surpassed those of other machine learning models at 10%~28% threshold probabilities. The SHAP method reveals the top 20 predictors of HF according to the importance ranking, and the average of the blood urea nitrogen was recognized as the most important predictor variable.

Conclusions: The interpretable predictive model helps physicians more accurately predict the mortality risk in ICU patients with HF, and therefore, provides better treatment plans and optimal resource allocation for their patients. In addition, the interpretable framework can increase the transparency of the model and facilitate understanding the reliability of the predictive model for the physicians.

Introduction

Heart failure (HF), the terminal phase of many cardiovascular disorders, is a major health care issue with an approximate prevalence of 26 million worldwide and more than 1 million hospital admissions annually in both the United States and Europe [1]. Projections show that by 2030 over 8 million Americans will have HF, leading to an increase of 46% from 2012 [2]. Each year, HF costs an estimated US $108 billion, constituting 2% of the health care budget globally, and it is predicted to continue to rise, yet half of it is possibly avoidable [3]. As COVID-19 continues to spread across the world, HF, a severe complication, is associated with poor outcome and death from COVID-19 [4,5].

The critically ill patients in intensive care units (ICUs) demand intensive care services and highly qualified multidisciplinary assistance [6]. Although ICU plays an integral role in maintaining patients' life, this also implies the workforce shortage, limited medical resources, and heavy economic burden [7]. Therefore, early hospital mortality detection for patients is necessary and may assist in delivering proper care and providing clinical decision support [8].

In recent years, artificial intelligence has been widely used to explore the early warning predictors of many diseases. Given the inherent powerful feature of capturing the nonlinear relationships with machine learning algorithms, more researchers advocate the use of new prediction models based on machine learning to support appropriate treatment for patients rather than traditional illness severity classification systems such as SOFA, APACHE II, or SAPS II [9-11]. Although a large number of predictive models have shown promising performance in research, the evidence for their application in clinical setting and interpretable risk prediction models to aid disease prognosis are still limited [12-15].

The purpose of our study is to develop an interpretable model to predict the risk mortality for patients with HF in the ICU, using the free and open critical care database—the eICU Collaborative Research Database (eICU-CRD). In addition, the SHapley Additive exPlanations (SHAP) method is used to explain the extreme gradient boosting (ie, XGBoost) model and explore prognostic factors for HF.

Methods

Data Source

The eICU-CRD (version 2.0) is a publicly available multicenter database [16], containing deidentified data associated with over 200,000 admissions to ICUs at 208 hospitals of the United States between 2014-2015. It records all patients, demographics, vital sign measurements, diagnosis information, and treatment information in detail [17].

Ethical Considerations

Ethical approval and individual patient consent was not necessary because all the protected health information was anonymized.

Study Population

All patients in the eICU-CRD database were considered. The inclusion criteria were as follows: (1) patients were diagnosed with HF according to the International Classification of Diseases, ninth and tenth Revision codes (Multimedia Appendix 1); (2) the diagnosis priority label was "primary" when admitted to the ICU in 24 hours; (3) the ICU stay was more than 1 day; and (4) patients were aged 18 years or older. Patients who had more than 30% missing values were excluded [18].

Predictor Variables

The prediction outcome of the study is the probability of in-hospital mortality, defined as patient's condition upon leaving the hospital. Based on previous studies [19-22] and experts' opinion (a total of 6 independent medical professionals and cardiologists in West China Hospital of Sichuan University), demographics, comorbidities, vital signs, and laboratory findings (Multimedia Appendix 2) were extracted from the eICU-CRD, using Structured Query Language (MySQL) queries (version 5.7.33; Oracle Corporation). The following tables from eICU-CRD were used: "diagnosis," "intakeoutput," "lab," "patient," and "nursecharting." Except for the demographic characteristics, other variables were collected during the first 24 hours of each ICU admission. Furthermore, to avoid overfitting, Least Absolute Shrinkage and Selection Operator (LASSO) is used to select and filter the variables [23,24].

Missing Data Handling

Variables with missing data are a common occurrence in eICU-CRD. However, analyses that ignore missing data have the potential to produce biased results. Therefore, we used multiple imputation for missing data [25]. All selected variables contained <30% missing values. Data were assumed missing at random and were imputed using fully conditional specification with the "mice" package (version 3.13.0) for R (version 4.1.0; R Core Team).

Machine Learning Explainable Tool

The interpretation of the prediction model is performed by SHAP, which is a unified approach to calculate the contribution and influence of each feature toward the final predictions precisely [26]. The SHAP values can show how much each predictor contributes, either positively or negatively, to the target variable. Besides, each observation in the data set could be interpreted by the particular set of SHAP values.

Statistical Analysis

All statistical analysis and calculations were performed using R software and Python (version 3.8.0; Python Software Foundation). The categorical variables are expressed as total

numbers and percentages, and the χ^2 test or Fisher exact test (expected frequency <10) is used to compare the differences between groups. The continuous variables are expressed as median and IQR, and the Wilcoxon rank sum test is used when comparing the two groups.

Four machine learning models—XGBoost, logistic regression (LR), random forest (RF), and support vector machine (SVM)— were used to develop the predictive models. The prediction performance of each model was evaluated by the area under the receiver operating characteristic curve. Moreover, we calculated the accuracy, sensitivity, positive predictive values, negative predictive values, and F_1 score when the prediction specificity was fixed at 85%. Additionally, to assess the utility of models for decision-making by quantifying the net benefit at different threshold probabilities, decision curve analysis (DCA) was conducted [27].

Results

Patient Characteristics

Among 17,029 patients with HF in eICU-CRD, a total of 2798 adult patients diagnosed with primary HF were included in the final cohort for this study. The patient screening process is shown in Figure 1. The data set was randomly divided into 2 parts: 70% (n=1958) of the data were used for model training, and 30% (n=840) of the data were used for model validation. The LASSO regularization process resulted in 24 potential predictors on the basis of 1958 patients in the training data set, which were used for model developing. Patients in the nonsurvivor group were older than the ones in the survivor group (*P*<.001). The hospital mortality rate was 9.96% (195/1958) in the training data set and 10% (84/840) in the testing data set (Multimedia Appendix 3). Table 1 shows the comparisons of predictor variables between survivors and nonsurvivors during hospitalization.

Figure 1. Flowchart of patient selection. ICD: International Classification of Diseases; ICU: intensive care unit.
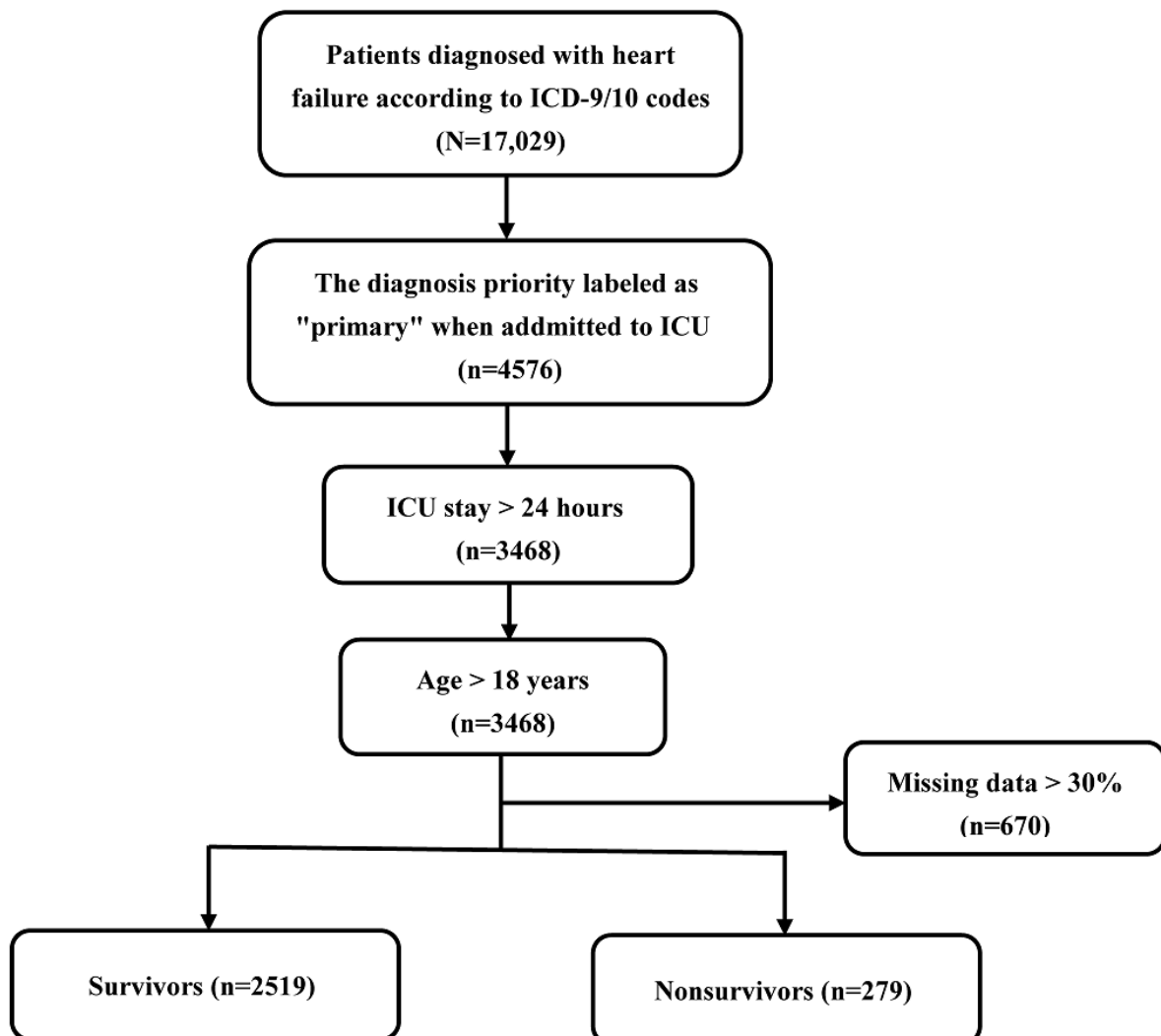
Table 1. All predictor variables for patients with heart failure (N=2798).

	Survivors (n=2519)	Nonsurvivors (n=279)	*P* value
Age (years), median (IQR)	71 (60-80)	76 (66-82)	<.001
Gender (male), n (%)	1338 (53.1)	170 (60.9)	.02
Comorbidities, n (%)			
Hypertension	654 (26)	46 (16.5)	<.001
Acute renal failure	441 (17.5)	78 (28.0)	<.001
Vital signs, median (IQR)			
Heartrate_min[a]	70 (61-80)	74 (62-86)	<.001
Respiratory rate_avg[b]	20.1 (17.8-23.0)	21.8 (19.0-26.0)	<.001
Respiratory rate_max[c]	27 (24-32)	32 (26-38)	<.001
Nibp[d]_systolic_avg	120.0 (107.1-134.8)	109.0 (100.1-121.4)	<.001
Nibp_systolic_min	95 (83-110)	84 (72-97)	<.001
Nibp_diastolic_min	49 (41-57)	45 (35-52.5)	<.001
Temperature_max	37 (37-37)	37 (37-38)	.03
Temperature_min	36 (36-37)	36 (36-37)	.007
Laboratory variables, median (IQR)			
Urineoutput	1550 (599-2750)	875 (140-1900)	<.001
SpO_2[e]_min	92 (88-95)	90 (84.5-94)	<.001
SpO_2_avg	96.6 (95.1-98.0)	96.5 (94.5-97.9)	.04
Anion_gap_max	11.0 (9.0-14.0)	12.0 (10.0-15.0)	<.001
Creatinine_min	1.45 (1.01-2.30)	1.70 (1.19-2.50)	.001
Blood_urea_nitrogen_avg	30.0 (21.0-47.6)	42.0 (28.0-58.5)	<.001
Calcium_min	8.6 (8.1-9.0)	8.5 (7.9-8.9)	.005
Chloride_min	101 (97-104)	99 (95-104)	.01
Platelets×1000_min	193 (149-249)	180 (140-235.5)	.008
White_blood_cell×1000_min	9.1 (6.8-12.1)	10.9 (7.6-15.7)	<.001
RDW[f]_min	15.7 (14.4-17.3)	16.4 (15.0-18.2)	<.001
Hemoglobin_max	10.6 (9.2-12.3)	10.4 (8.95-12.0)	.059

[a]Min: minimum.

[b]Avg: average.

[c]Max: maximum.

[d]Nibp: noninvasive blood pressure.

[e]SpO_2: O_2 saturation.

[f]RDW: red blood cell distribution width.

Model Building and Evaluation

Within the training data set, the XGBoost, LR, RF, and SVM models were established, and the testing data set obtained AUCs of 0.824, 0.800, 0.779, and 0.701, respectively (Table 2 and Figure 2). Comparatively, XGBoost had the highest predictive performance among the four models (AUC=0.824, 95% CI 0.7766-0.8708), whereas SVM had the poorest generalization ability (AUC=0.701, 95% CI 0.6433-0.7582). DCA was performed for four machine learning models in the testing data set to compare the net benefit of the best model and alternative

approaches for clinical decision-making. Clinical net benefit is defined as the minimum probability of disease, when further intervention would be warranted [28]. The plot measures the net benefit at different threshold probabilities. The orange line in Figure 3 represents the assumption that all patients received intervention, whereas the yellow line represents that none of the patients received intervention. Due to the heterogeneous profile of the study population, treatment strategies informed by any of the four machine learning–based models are superior to the default strategies of treating all or no patient. The net benefit of the XGBoost model surpassed those of the other

XSL•FO

RenderX

machine learning models at 10%~28% threshold probabilities (Figure 3).

Table 2. Performance of each model for prediction.

Model	AUC[a] (%)	Sensitivity (%)	F_1 score	Accuracy (%)	PPV[b]	NPV[c]
XGBoost	0.824	0.595	0.407	0.826	0.307	0.950
LR[d]	0.800	0.607	0.413	0.827	0.311	0.951
RF[e]	0.779	0.571	0.392	0.823	0.298	0.947
SVM[f]	0.701	0.345	0.258	0.801	0.204	0.921

[a]AUC: area under the curve.

[b]PPV: positive predictive value.

[c]NPV: negative predictive value.

[d]LR: logistic regression.

[e]RF: random forest.

[f]SVM: support vector machine.

Figure 2. The receiver operating characteristic curve among the four models for patients with heart failure. SVM: support vector machine.
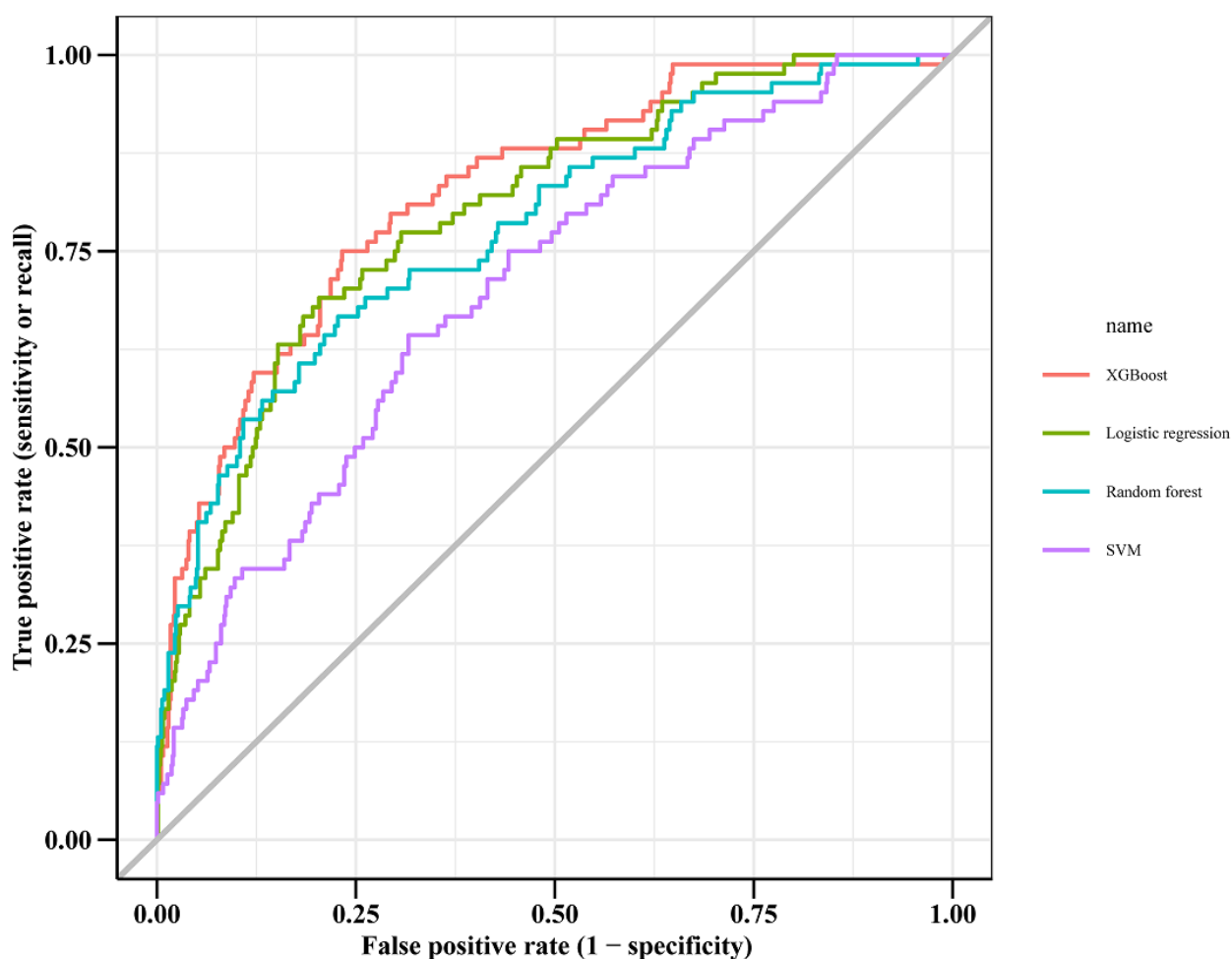
Figure 3. Decision curve analysis of four models plotting the net benefit at different threshold probabilities. SVM: support vector machine.
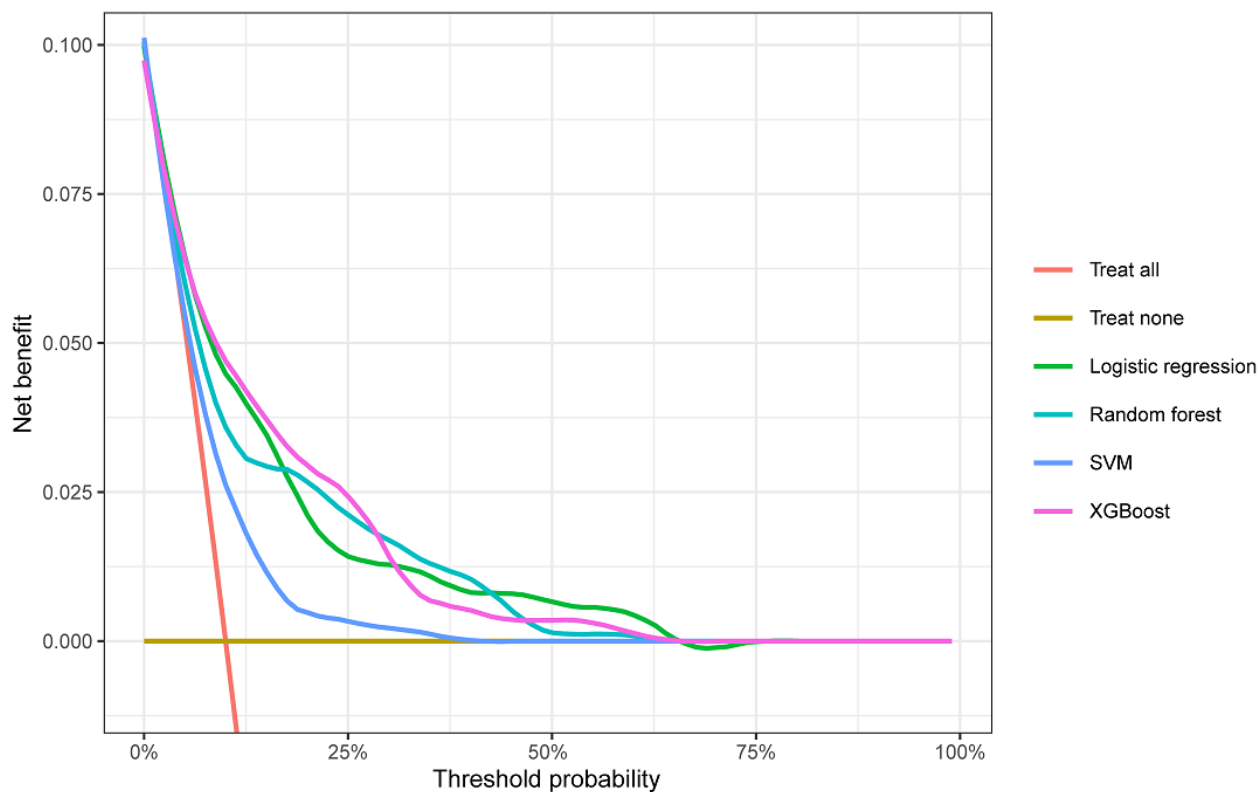
Explanation of XGBoost Model With the SHAP Method

The SHAP algorithm was used to obtain the importance of each predictor variable to the outcome predicted by the XGBoost model. The variable importance plot lists the most significant variables in a descending order (Figure 4). The average of blood urea nitrogen (BUN) had the strongest predictive value for all prediction horizons, followed quite closely by the age factor, the average of noninvasive systolic blood pressure, urine output, and the maximum of respiratory rate. Furthermore, to detect the positive and negative relationships of the predictors with the target result, SHAP values were applied to uncover the mortality risk factors. As presented in Figure 5, the horizontal location shows whether the effect of that value is associated with a higher or lower prediction and the color shows whether that variable is high (in red) or low (in blue) for that observation; we can see that increases in the average BUN has a positive impact and push the prediction toward mortality, whereas increases in urine output has a negative impact and push the prediction toward survival.

Figure 4. The weights of variables importance. Avg: average; BUN: blood urea nitrogen; max: maximum; min: minimum; NIBP: noninvasive blood pressure; RDW: red blood cell distribution width; SHAP: SHapley Additive exPlanation; SpO$_2$: O$_2$ saturation; WBC: white blood cell.
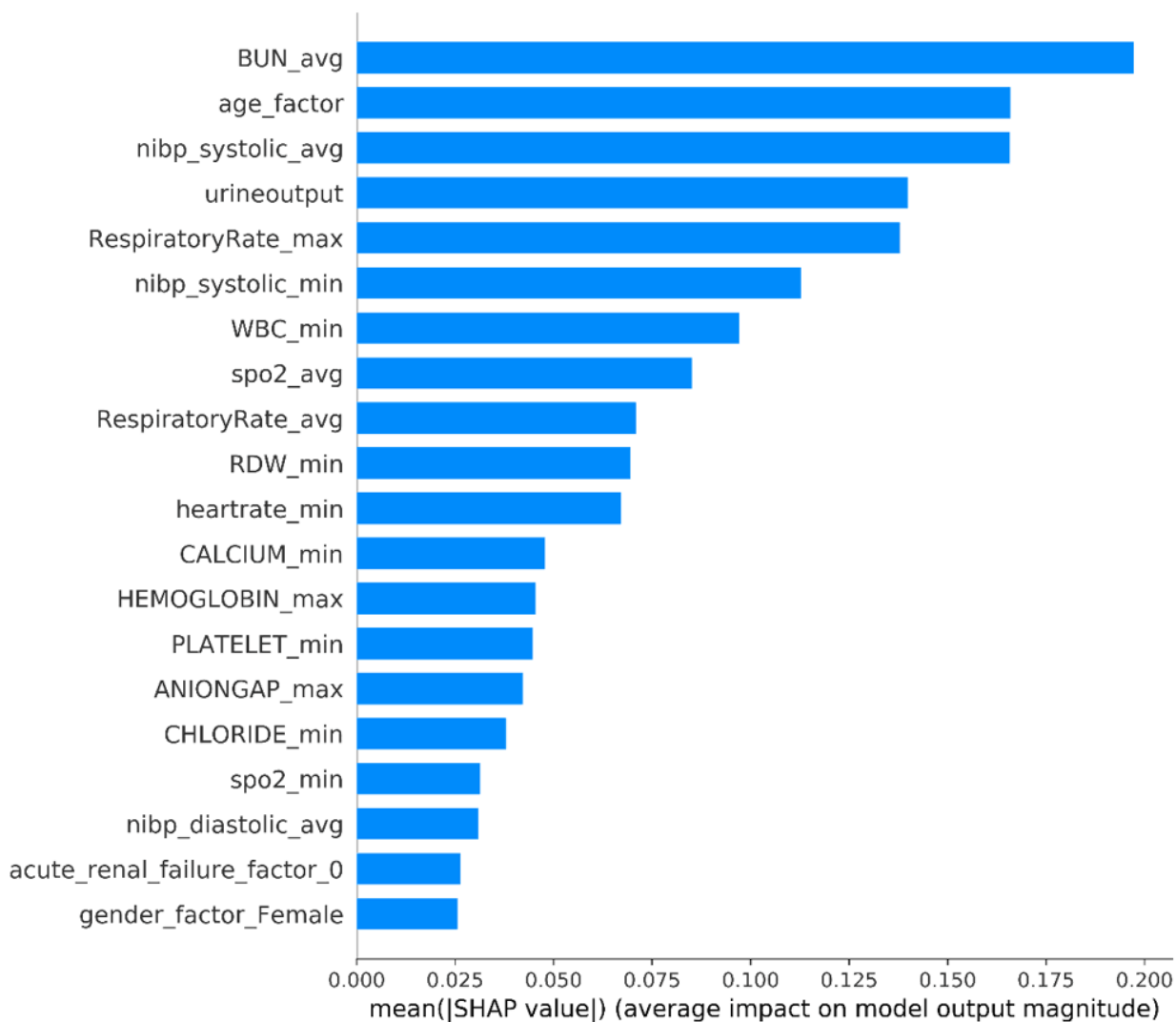
Figure 5. The SHapley Additive exPlanation (SHAP) values. Avg: average; BUN: blood urea nitrogen; max: maximum; min: minimum; NIBP: noninvasive blood pressure; RDW: red blood cell distribution width; SpO$_2$: O$_2$ saturation; WBC: white blood cell.
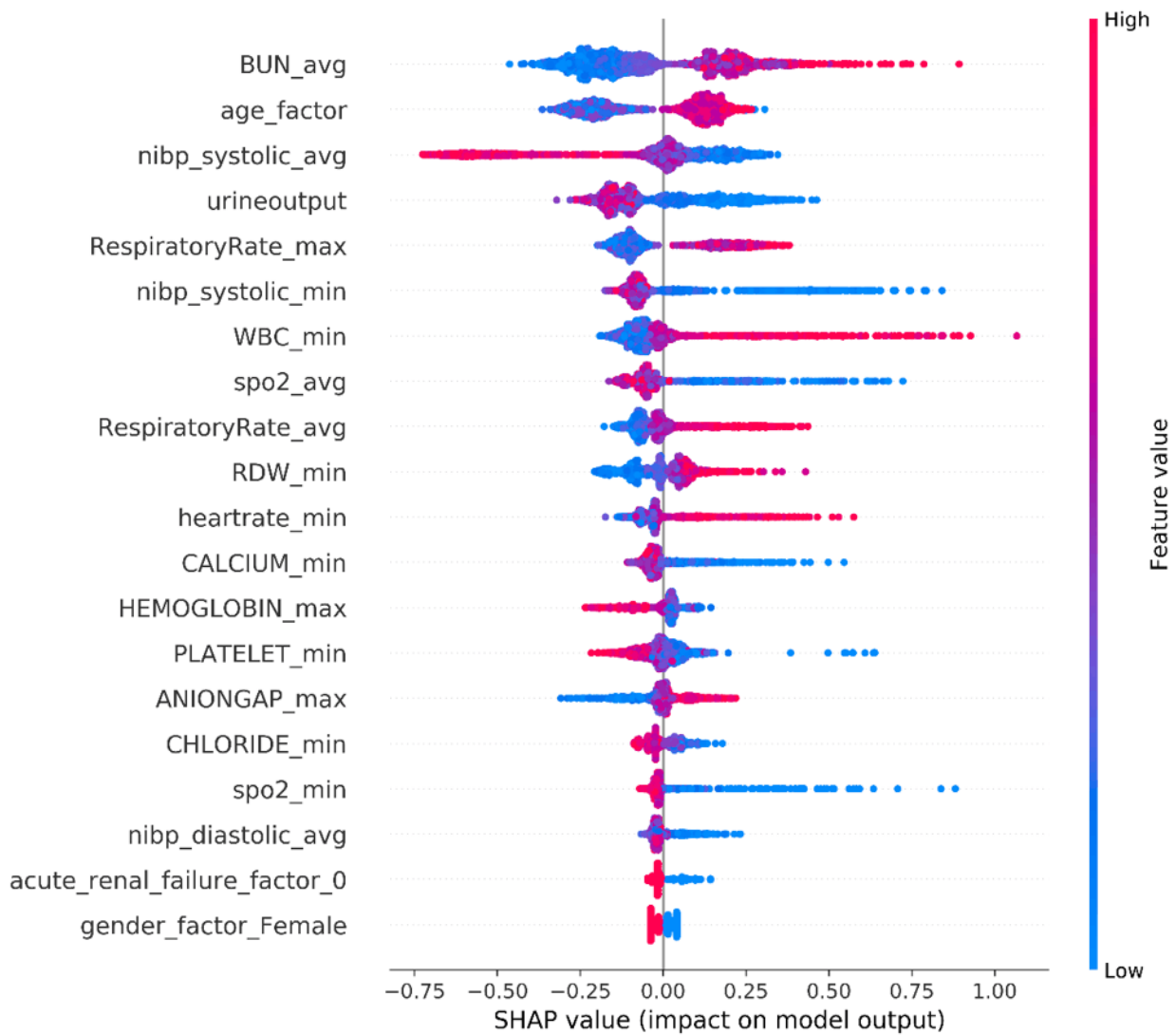
SHAP Individual Force Plots

Figure 6 shows the individual force plots for patients who (A) did not survive and (B) survived. The SHAP values indicate the prediction-related feature of individual patients and the contribution of each feature to the mortality prediction. The bold-faced numbers are the probabilistic predicted values (f(x)), whereas the base values are the values predicted without giving input to the model. The f(x) is the log odds ratio of each observation. The red features (on the left) indicate features that increase the mortality risk, and the blue features indicate features that decrease the mortality risk. The length of the arrows helps visualize the magnitude of the effect on the prediction. The longer the arrow, the larger the effect.

Figure 6. SHapley Additive exPlanation (SHAP) force plot for two selected patients.
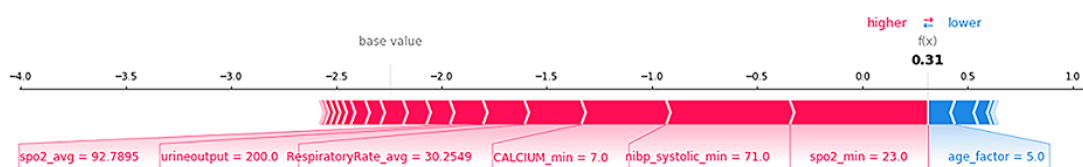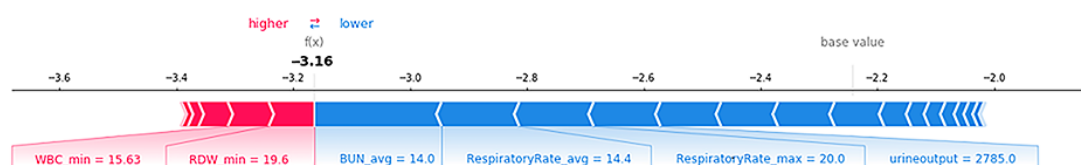
Discussion

Principal Findings

In this retrospective cohort study of a large-scale public ICU database, we developed and validated four machine learning algorithms to predict the mortality of patients with HF. The XGBoost model outperforms the performance of LR, RF, and SVM. The SHAP method is used to explain the XGBoost model, which ensures the model performance and clinical interpretability. This will help physicians better understand the decision-making process of the model and facilitates the use of prediction results. Besides, to avoid ineffective clinical interventions, the relevant threshold probability range of DCA that we focused on was between 15% and 25%, during which XGBoost performed better. In critical care research, XGBoost has been widely used to predict the in-hospital mortality of patients and may assist clinicians in decision-making [29-31]. However, the mortality of patients with HF included in the final cohort is just 9.97%. Although DCA shows that the XGBoost model is better than the two default strategies, the positive predictive value is just 0.307 when the prediction specificity is fixed at 85%. Therefore, the XGBoost model may not be fully acceptable to provide decision-making support for clinicians. Evaluation of the benefits of earlier prediction of mortality and its additional cost is necessary in clinical practice.

Using SHAP to explain the XGBoost model, we identified some important variables associated with in-hospital mortality of patients with HF. In this study, the average BUN was recognized as the most important predictor variable. As a renal function marker to measure the amount of nitrogen in blood that comes from protein metabolism, previous studies also showed that BUN was the key predictor of HF mortality prediction with machine learning algorithms [32,33]. Kazory [34] concludes that BUN may be a biomarker of neurohormonal activation in patients with HF. From the perspective of pathophysiology, the activity of sympathetic nervous systems and the renin-angiotensin-aldosterone system increases with the aggravation of HF, which causes the vasoconstriction of the afferent arterioles. A reduction in renal perfusion further leads to water and sodium retention and promotes urea reabsorption, ultimately resulting in an increased BUN. However, further research is needed to explore the applicability of this SHAP method, due to the lack of an external validation cohort.

Limitations

This study had some limitations. First, our data were extracted from a publicly available database, and some variables were missing. For example, we intended to include more predictor variables that may affect in-hospital mortality such as prothrombin time as well as brain natriuretic peptide and lactate; however, the missing values were over 70%. Second, all data were derived from the ICU patients from the United States, so the applicability of our model remained unclear in other populations. Third, our mortality prediction models were based on data available within 24 hours of each ICU admission, which may neglect the subsequent events that change the prognosis and cause confounders to some extent. Fourth, due to lack of an external validation cohort, the applicability of the developed XGBoost model may not be very efficient in clinical practice. Currently, we are trying to collect data of patients with HF in ICUs from West China Hospital of Sichuan University. Although we have obtained some preliminary data, it is not feasible for the external prospective validation because of the limited sample size.

Conclusions

We developed the interpretable XGBoost prediction model that has the best performance in estimating the mortality risk in patients with HF. In addition, the interpretable machine learning approach can be applied to accurately explore the risk factors of patients with HF and enhance physicians' trust in prediction

models. This will help physicians identify patients with HF who have a high mortality risk so as to timely apply appropriate treatments for them.

Authors' Contributions

J Liu, J Li, and SL conceived the study. SL, J Liu, J Li, YH, LZ, and YM performed the analysis, interpreted the results, and drafted the manuscript. All authors revised the manuscript. All authors read and approved the final manuscript.

Conflicts of Interest

None declared.

Multimedia Appendix 1

Supplementary material 1: the ICD 9/10 codes.
[XLS File (Microsoft Excel File), 25 KB-Multimedia Appendix 1]

Multimedia Appendix 2

Supplementary material 2: the selected predictor variables.
[DOCX File , 17 KB-Multimedia Appendix 2]

Multimedia Appendix 3

Supplementary material 3: all predictor variables for patients in the training and testing data set.
[DOCX File , 21 KB-Multimedia Appendix 3]

References

1. Ambrosy AP, Fonarow GC, Butler J, Chioncel O, Greene SJ, Vaduganathan M, et al. The global health and economic burden of hospitalizations for heart failure: lessons learned from hospitalized heart failure registries. J Am Coll Cardiol. Apr 1, 2014;63(12):1123-1133. [FREE Full text] [doi: 10.1016/j.jacc.2013.11.053] [Medline: 24491689]
2. Heidenreich PA, Albert NM, Allen LA, Bluemke DA, Butler J, Fonarow GC, et al. Forecasting the impact of heart failure in the United States: a policy statement from the American Heart Association. Circ Heart Fail. May 2013;6(3):606-619. [FREE Full text] [doi: 10.1161/HHF.0b013e318291329a] [Medline: 23616602]
3. Liao L, Allen LA, Whellan DJ. Economic burden of heart failure in the elderly. Pharmacoeconomics. 2008;26(6):447-462. [Medline: 18489197]
4. Kim HJ, Park M, Shin JI, Park J, Kim D, Jeon J, et al. Associations of heart failure with susceptibility and severe complications of COVID-19: a nationwide cohort study. J Med Virol. Nov 05, 2021:1138-1145. [FREE Full text] [doi: 10.1002/jmv.27435] [Medline: 34738248]
5. Yonas E, Alwi I, Pranata R, Huang I, Lim MA, Gutierrez EJ, et al. Effect of heart failure on the outcome of COVID-19 - a meta analysis and systematic review. Am J Emerg Med. Aug 2021;46:204-211. [FREE Full text] [doi: 10.1016/j.ajem.2020.07.009] [Medline: 33071085]
6. Monteiro F, Meloni F, Baranauskas JA, Macedo AA. Prediction of mortality in intensive care units: a multivariate feature selection. J Biomed Inform. Jul 2020;107:103456. [FREE Full text] [doi: 10.1016/j.jbi.2020.103456] [Medline: 32454242]
7. Kramer AA, Dasta JF, Kane-Gill SL. The impact of mortality on total costs within the ICU. Crit Care Med. Sep 2017;45(9):1457-1463. [doi: 10.1097/CCM.0000000000002563] [Medline: 28658024]
8. Khurrum M, Asmar S, Joseph B. Telemedicine in the ICU: innovation in the critical care process. J Intensive Care Med. Dec 2021;36(12):1377-1384. [doi: 10.1177/0885066620968518] [Medline: 33111599]
9. Kang MW, Kim J, Kim DK, Oh K, Joo KW, Kim YS, et al. Machine learning algorithm to predict mortality in patients undergoing continuous renal replacement therapy. Crit Care. Feb 06, 2020;24(1):42. [FREE Full text] [doi: 10.1186/s13054-020-2752-7] [Medline: 32028984]
10. Liu C, Liu X, Mao Z, Hu P, Li X, Hu J, et al. Interpretable machine learning model for early prediction of mortality in ICU patients with rhabdomyolysis. Med Sci Sports Exerc. Sep 01, 2021;53(9):1826-1834. [doi: 10.1249/MSS.0000000000002674] [Medline: 33787533]
11. Pirracchio R, Petersen ML, Carone M, Rigon MR, Chevret S, van DLMJ. Mortality prediction in intensive care units with the Super ICU Learner Algorithm (SICULA): a population-based study. Lancet Respir Med. Jan 2015;3(1):42-52. [FREE Full text] [doi: 10.1016/S2213-2600(14)70239-5] [Medline: 25466337]
12. Pan P, Li Y, Xiao Y, Han B, Su L, Su M, et al. Prognostic Assessment of COVID-19 in the intensive care unit by machine learning methods: model development and validation. J Med Internet Res. Nov 11, 2020;22(11):e23128. [FREE Full text] [doi: 10.2196/23128] [Medline: 33035175]

13. Stenwig E, Salvi G, Rossi PS, Skjærvold NK. Comparative analysis of explainable machine learning prediction models for hospital mortality. BMC Med Res Methodol. Feb 27, 2022;22(1):53. [FREE Full text] [doi: 10.1186/s12874-022-01540-w] [Medline: 35220950]

14. Wang K, Tian J, Zheng C, Yang H, Ren J, Liu Y, 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. Oct 2021;137:104813. [FREE Full text] [doi: 10.1016/j.compbiomed.2021.104813] [Medline: 34481185]

15. Zihni E, Madai VI, Livne M, Galinovic I, Khalil AA, Fiebach JB, et al. Opening the black box of artificial intelligence for clinical decision support: a study predicting stroke outcome. PLoS One. 2020;15(4):e0231166. [FREE Full text] [doi: 10.1371/journal.pone.0231166] [Medline: 32251471]

16. eICU Collaborative Research Database. URL: https://eicu-crd.mit.edu/ [accessed 2022-07-20]

17. Pollard TJ, Johnson AEW, Raffa JD, Celi LA, Mark RG, Badawi O. The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Sci Data. Sep 11, 2018;5:180178. [FREE Full text] [doi: 10.1038/sdata.2018.178] [Medline: 30204154]

18. Li K, Shi Q, Liu S, Xie Y, Liu J. Predicting in-hospital mortality in ICU patients with sepsis using gradient boosting decision tree. Medicine (Baltimore). May 14, 2021;100(19):e25813. [FREE Full text] [doi: 10.1097/MD.0000000000025813] [Medline: 34106618]

19. Abraham WT, Fonarow GC, Albert NM, Stough WG, Gheorghiade M, Greenberg BH, et al. OPTIMIZE-HF InvestigatorsCoordinators. Predictors of in-hospital mortality in patients hospitalized for heart failure: insights from the Organized Program to Initiate Lifesaving Treatment in Hospitalized Patients with Heart Failure (OPTIMIZE-HF). J Am Coll Cardiol. Jul 29, 2008;52(5):347-356. [FREE Full text] [doi: 10.1016/j.jacc.2008.04.028] [Medline: 18652942]

20. Fonarow GC, Adams KF, Abraham WT, Yancy CW, Boscardin WJ, ADHERE Scientific Advisory Committee, Study Group,Investigators. Risk stratification for in-hospital mortality in acutely decompensated heart failure: classification and regression tree analysis. JAMA. Feb 02, 2005;293(5):572-580. [doi: 10.1001/jama.293.5.572] [Medline: 15687312]

21. Knaus WA, Draper EA, Wagner DP, Zimmerman JE. APACHE II: a severity of disease classification system. Crit Care Med. Oct 1985;13(10):818-829. [Medline: 3928249]

22. Peterson PN, Rumsfeld JS, Liang L, Albert NM, Hernandez AF, Peterson ED, et al. American Heart Association Get With the Guidelines-Heart Failure Program. A validated risk score for in-hospital mortality in patients with heart failure from the American Heart Association get with the guidelines program. Circ Cardiovasc Qual Outcomes. Jan 2010;3(1):25-32. [doi: 10.1161/CIRCOUTCOMES.109.854877] [Medline: 20123668]

23. Tibshirani R. Regression shrinkage and selection via the Lasso. J R Stat Soc Series B Stat Methodol. 1996;58(1):267-288. [FREE Full text] [doi: 10.1111/j.2517-6161.1996.tb02080.x]

24. Muthukrishnan R, Rohini R. LASSO: A feature selection technique in predictive modeling for machine learning. 2016. Presented at: IEEE international conference on advances in computer applications (ICACA); Oct 24; Coimbatore, India. [doi: 10.1109/icaca.2016.7887916]

25. Blazek K, van Zwieten A, Saglimbene V, Teixeira-Pinto A. A practical guide to multiple imputation of missing data in nephrology. Kidney Int. Jan 2021;99(1):68-74. [doi: 10.1016/j.kint.2020.07.035] [Medline: 32822702]

26. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. 2017. Presented at: NIPS'17; Dec 4-9:4768-4777; Long Beach, CA.

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. Dec 2018;74(6):796-804. [FREE Full text] [doi: 10.1016/j.eururo.2018.08.038] [Medline: 30241973]

28. Lee C, Light A, Alaa A, Thurtle D, van der Schaar M, Gnanapragasam VJ. Application of a novel machine learning framework for predicting non-metastatic prostate cancer-specific mortality in men using the Surveillance, Epidemiology, and End Results (SEER) database. Lancet Digit Health. Mar 2021;3(3):e158-e165. [FREE Full text] [doi: 10.1016/S2589-7500(20)30314-9] [Medline: 33549512]

29. 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. Dec 07, 2020;18(1):462. [FREE Full text] [doi: 10.1186/s12967-020-02620-5] [Medline: 33287854]

30. Liu J, Wu J, Liu S, Li M, Hu K, Li K. Predicting mortality of patients with acute kidney injury in the ICU using XGBoost model. PLoS One. 2021;16(2):e0246306. [FREE Full text] [doi: 10.1371/journal.pone.0246306] [Medline: 33539390]

31. Zhang Z, Ho KM, Hong Y. Machine learning for the prediction of volume responsiveness in patients with oliguric acute kidney injury in critical care. Crit Care. Apr 08, 2019;23(1):112. [FREE Full text] [doi: 10.1186/s13054-019-2411-z] [Medline: 30961662]

32. Adler ED, Voors AA, Klein L, Macheret F, Braun OO, Urey MA, et al. Improving risk prediction in heart failure using machine learning. Eur J Heart Fail. Jan 2020;22(1):139-147. [FREE Full text] [doi: 10.1002/ejhf.1628] [Medline: 31721391]

33. Angraal S, Mortazavi BJ, Gupta A, Khera R, Ahmad T, Desai NR, et al. Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction. JACC Heart Fail. Jan 2020;8(1):12-21. [FREE Full text] [doi: 10.1016/j.jchf.2019.06.013] [Medline: 31606361]

34.   Kazory A. Emergence of blood urea nitrogen as a biomarker of neurohormonal activation in heart failure. Am J Cardiol. Sep 01, 2010;106(5):694-700. [doi: 10.1016/j.amjcard.2010.04.024] [Medline: 20723648]

Abbreviations

AUC: area under the curve
BUN: blood urea nitrogen
DCA: decision curve analysis
eICU-CRD: eICU Collaborative Research Database
HF: heart failure
ICU: intensive care unit
LASSO: Least Absolute Shrinkage and Selection Operator
LR: logistic regression
RF: random forest
SHAP: SHapley Additive exPlanation
SVM: support vector machine