Prediction of Death Risk in Heart Failure Patients Based on Explainable Machine Learning Models

Yue Yan, Hao Zhu*, Luyao Zhou, Qi Yang, Canyan Liao, Wei He, Baolin Zhou
College of Physics and Information Engineering, Zhaotong University, Zhaotong, China
**Corresponding Author*

Abstract: This study was designed to construct an interpretable machine learning model for predicting mortality risk in intensive care unit (ICU) patients with heart failure. Data were obtained from the MIMIC-IV and eICU databases, comprising 30,411 heart failure patients, and were subsequently split into training and testing sets. The study employed four machine learning algorithms—eXtreme Gradient Boosting (XGBoost), random forest (RF), Classification and Regression Trees (CART), and logistic regression (LR)—to build predictive models for adverse outcomes. The SHapley Additive exPlanations (SHAP) method was applied to interpret the model and identify key prognostic factors. The optimal model was selected according to its predictive accuracy and the area under the receiver operating characteristic curve (AUC). Among the four models, the XGBoost model demonstrated the highest predictive performance, achieving an AUC value of 0.88. Finally, the optimal model is explained using the Shapley Additive Explanation value. The SHAP value shows that the average value of APSIII is the most important predictive variable. An interpretable machine learning model not only performs well in predicting mortality rates in heart failure patients, but is also crucial for clinicians to develop personalized prevention and treatment plans.

Keywords: Heart Failure; Machine Learning; Extreme Gradient Boosting; Shapley

1. Introduction

1.1 Research Background and Significance

Heart failure (HF) refers to the syndrome of cardiac circulatory disorders caused by dysfunction of the systolic or diastolic function of the heart, which prevents sufficient venous blood flow back to the heart, resulting in blood stasis in the venous system and inability to inject blood into the arterial system normally. The disease has the characteristics of high incidence rate, high medical cost and poor prognosis, and has become a major public health problem worldwide [1,2]. Heart failure is not an independent disease, but rather the late stage of the development of various heart diseases. It is a complex clinical comprehensive disease, such as impaired ventricular filling, decreased ejection function, ventricular dysfunction, insufficient cardiac output, pulmonary or systemic circulation congestion, insufficient organ or tissue blood perfusion, etc.

On a global scale, the prevalence of heart failure ranges from 1% to 7%, and the hospital mortality rate ranges from 5% to 20%. In the first year after discharge, about half of heart failure patients need to be readmitted for treatment, at which point the mortality rate increases to 20% -30% [3]. The prognosis of patients with acute heart failure is very poor, and the combined mortality rate and readmission rate within 60 days have reached 35.2% [4,5]. Within one year after diagnosis, the mortality rate is 20%, within five years after diagnosis, the mortality rate is 50%, and the 10-year mortality rate is as high as 90%. Heart failure affects over 5 million people, with approximately 2% of adults and 10% of the elderly suffering from this disease. In the United States, there are over 400,000 new cases of heart failure each year, with approximately 1 million hospitalized patients, of which more than 80% are over 65 years old. 10% of the annual healthcare budget in the United States is spent on the management of cardiovascular disease and heart failure, and 75% is spent on hospital care. The mortality rate of heart failure in different age groups is higher than other cardiovascular diseases, reaching 59% [6].

Heart failure is an increasingly serious public health problem, and the incidence increases

with age. Therefore, it is necessary to conduct research on the risk stratification and mortality assessment of heart failure patients, which can provide basic strategies for clinical decision-making and practical information for health policies and insurance services. In order to predict the mortality rate of heart failure patients more accurately, artificial intelligence, especially machine learning, is a useful tool. Unlike conventional predictive approaches that rely on a predefined set of variables, machine learning can leverage computational power to integrate a large number of variables, thereby enhancing prediction accuracy. Moreover, it employs various feature selection techniques to refine the model, improving both its precision and effectiveness. Therefore, this article establishes a predictive model based on interpretable machine learning that can be widely applied in practice. This model can effectively utilize clinical data of heart failure patients, establish a high-precision heart failure prognosis evaluation model, and predict the occurrence of adverse events as soon as possible and assist doctors in treatment, which has important clinical significance.

1.2 Current Research Status at Home and Abroad

In recent years, machine learning models have accurately classified the complex pathology and intervention outcomes of heart failure. Aaron et al. [7] established a multivariate proportional risk survival model using 80 clinical feature data from 268 outpatient patients with advanced heart failure. Gao et al. [8] conducted research on heart failure risk assessment methods based on electronic health records and deep learning, exploring new approaches for data-driven prognostic evaluation. Zeng et al. [9] constructed a machine learning-based risk prediction model for death or readmission in acute heart failure patients during the vulnerable phase, validating its effectiveness in a clinical cohort. Xu et al. [10] systematically reviewed the research progress of machine learning-based prediction models for heart failure risk, summarizing methodological advances and application challenges. Xie et al. [11] evaluated a heart failure discrimination model for patients with NT-proBNP grey zone values based on machine learning algorithms, providing a solution for a specific clinical subgroup. Research has confirmed that in many cases, machine learning algorithms establish models with better performance than traditional methods.

In order to construct a more accurate assessment of the mortality risk of heart failure patients, this article uses interpretable machine learning algorithms to screen adverse factor variables of heart failure patients and establish a mortality prediction model for heart failure patients, in order to correct and improve the prognosis of patients as soon as possible, and to construct efficient and feasible cardiac arrest warning mechanisms and prevention strategies.

2. Materials and Methods

2.1 Data Source and Introduction

MIMIC-IV is a freely accessible, single-center database comprising clinical data for more than 40,000 intensive care unit (ICU) patients admitted between 2008 and 2019. Its dataset includes patient demographic information, laboratory test values, medication treatment records, and recorded vital signs. Another widely utilized resource is the eICU Collaborative Research Database (eICU-CRD), which is derived from multiple longitudinal, multicenter retrospective cohort studies involving 335 ICUs across the United States from 2014 to 2015. This database contains demographic information, physiological parameters obtained from bedside monitoring, diagnosis records coded according to the International Classification of Diseases, Ninth Revision (ICD-9), as well as additional laboratory data.

2.2 Study Population

The exclusionary criteria comprised patients aged below 18 years or those with missing outcome data. For individuals with multiple ICU admissions, only data from the initial admission were included. Following the implementation of these criteria, the study encompassed 15,983 and 14,428 CHF patients in the training and testing sets, respectively. The patient selection flowchart is presented in Figure 1.

2.3 Data Extraction

Considering clinical significance and data accessibility, the following variables were extracted: demographic characteristics (age, gender, race, weight, height, and admission

severity scores (SOFA, APSIII)); vital signs (heart rate (HR), systolic blood pressure (SBP), diastolic blood pressure (DBP), mean arterial pressure, respiratory rate, body temperature, pulse oxygen saturation (SpO_2), and first 24-hour urine output); Complications (hypertension, atrial fibrillation, ischemic heart disease, diabetes, depression, iron deficiency anemia, hyperlipidemia, chronic kidney disease (CKD), and chronic obstructive pulmonary disease (COPD)); and laboratory variables (hematocrit, red blood cell count, mean corpuscular hemoglobin, mean red blood cell hemoglobin concentration, mean corpuscular volume (MCV), red blood cell distribution width (RDW), platelet count, white blood cells, neutrophils, basophils, lymphocytes, prothrombin time (PT), international normalized ratio (INR), NT-proBNP, creatine kinase, creatinine, blood urea nitrogen (BUN), glucose, potassium, sodium, calcium, chloride, magnesium, anion gap, bicarbonate, lactate, and hydrogen ion concentration (pH), along with vasopressor records including adrenaline, antidiuretic hormone, dopamine, and phenylephrine). Demographic characteristics and vital signs were recorded within the first 24 hours after ICU admission, whereas laboratory variables were assayed throughout the entire ICU hospitalization. Comorbidities were identified using ICD-9 codes, and for variables with repeated measurements, the mean value was calculated for subsequent data analysis.

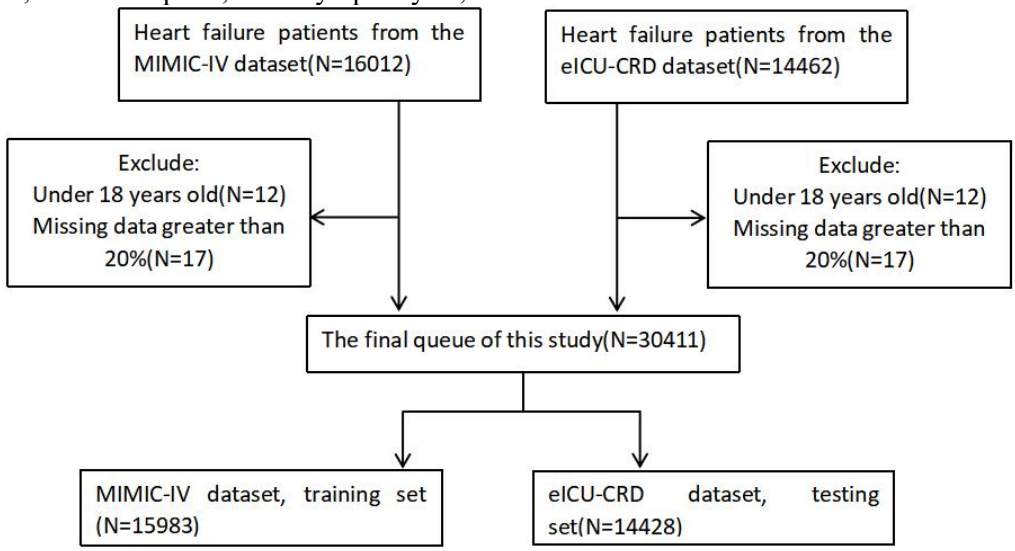


Figure 1. Patient Selection Flowchart. eICU-CRD, eICU Collaborative Research Database; MIMIC-IV, Intensive Care Medical Information Market.

2.4 Machine Learning and Explainable Method

The identification of robust and clinically relevant features is critical for accurate mortality risk prediction. To this end, this research applied both feature mining and visualization techniques to examine feature-related mortality risk. The study cohort was randomly stratified into a training set (90%) and a validation set (10%). Four machine learning classifiers—eXtreme Gradient Boosting (XGBoost), Random Forest (RF), Classification and Regression Trees (CART), and Logistic Regression (LR)—were developed and validated. The receiver operating characteristic (ROC) curve was utilized to select and evaluate the optimal model. Subsequently, the best-performing model employed the SHapley Additive exPlanations (SHAP) method to visualize the key features influencing mortality risk. Based on game theory, SHAP interprets machine learning model outputs by calculating the contribution of each feature. This interpretable framework assists clinicians in comprehending risk factors at both the variable and patient levels. The resulting explanations provide human-readable insights, which, when effectively utilized, can contribute to improved clinical decision-making and patient outcomes.

2.5 Statistic analysis

The normality of continuous variables was assessed using the Kolmogorov-Smirnov test. Normally distributed variables are presented as mean ± standard deviation (SD), while non-normally distributed variables are summarized as median and interquartile range (IQR). Categorical variables are expressed as numbers

and percentages. Intergroup comparisons for continuous variables were performed using the Student's t-test or the Mann-Whitney U test, as appropriate. For categorical variables, the Pearson chi-square test or Fisher's exact test was applied for group comparisons based on the data characteristics.

To mitigate potential bias from missing data, variables with missing values greater than 20% were excluded from the analysis. Other variables with smaller missing values, multiple imputation techniques were employed for analysis.

A p-value below 0.05 was regarded as statistically significant. All statistical analyses in this study were performed using Yi Yan software and R software.

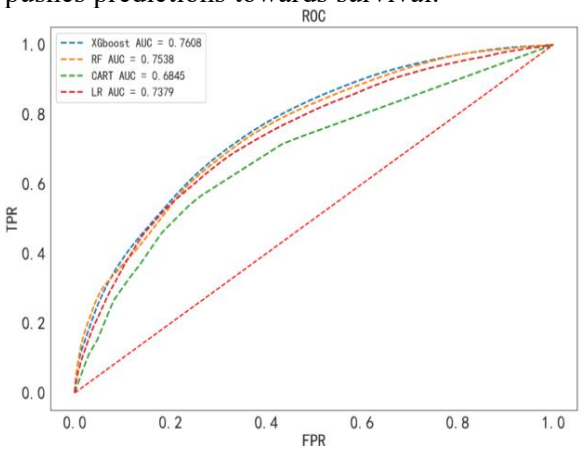
3. Results

3.1 Multiple Model Comparison

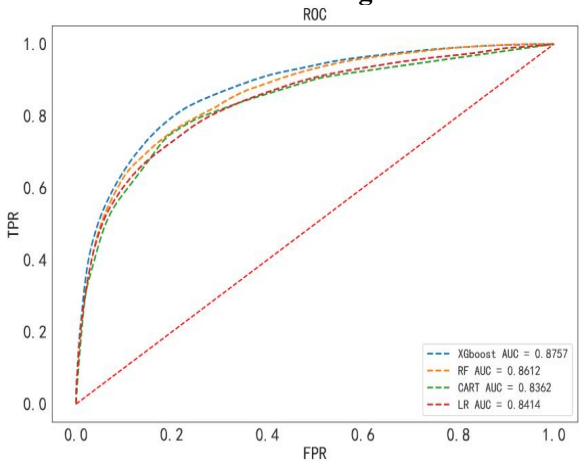
XGBoost, LR, RF, and CART models were developed using the training dataset. The predictive performance of these four models was evaluated by comparing their receiver operating characteristic (ROC) curves. The area under the ROC curve (AUC) quantifies the model's discriminative ability, with values between 0.5 and 1. A higher AUC closer to 1 indicates stronger predictive power, while an AUC of 0.5 suggests no discriminative capacity. In the training set, the AUC values for the XGBoost, LR, RF, and CART models were 0.7608, 0.7379, 0.7538, and 0.6845, respectively (Figure 2A). Corresponding AUCs in the test set were 0.8757, 0.8414, 0.8612, and 0.8362 (Figure 2B). Comparative analysis revealed that the XGBoost model achieved the highest predictive performance across both datasets. Consequently, the XGBoost-based model was selected as the final model for subsequent analysis.

3.2 Interpreting XGBoost Model with SHAP Method

The SHAP algorithm was employed to assess the contribution of each predictive variable to the outcomes of the XGBoost model. The relative importance of these variables is displayed in descending order in the feature importance plot, as presented in Figure 3. The top 12 were: Acute Physiological Score (APSIII), Age, Norepinephrine, Red Blood Cell Distribution Width (RDW), Respiratory Rate (RR), Sequential Organ Failure Score (SOFA), Vasopressin, Anion Gap (AG), Intubated, Systolic Blood Pressure (SBP), Dopamine, Phenylephrine. Additionally, to analyze the directional associations between predictive factors and the target outcome, SHAP values were applied to identify risk factors associated with mortality, as illustrated in Figure 4. The horizontal position of each point indicates the direction of its influence on the prediction, while the color represents whether the observed value of that variable is relatively high (red) or low (blue). A wider spread of points along the horizontal axis corresponds to a greater magnitude of influence for that feature. Therefore, this graph is generally wide at the top and narrow at the bottom (the one with the greatest influence is placed above). We can see that the increase in APSIII has a positive impact and pushes predictions towards death, while the increase in RDW has a negative impact and pushes predictions towards survival.



A. The Training Set



B. The Testing Set

Figure 2. The Working Characteristic Curves of Subjects in Four Models of Heart Failure Patients

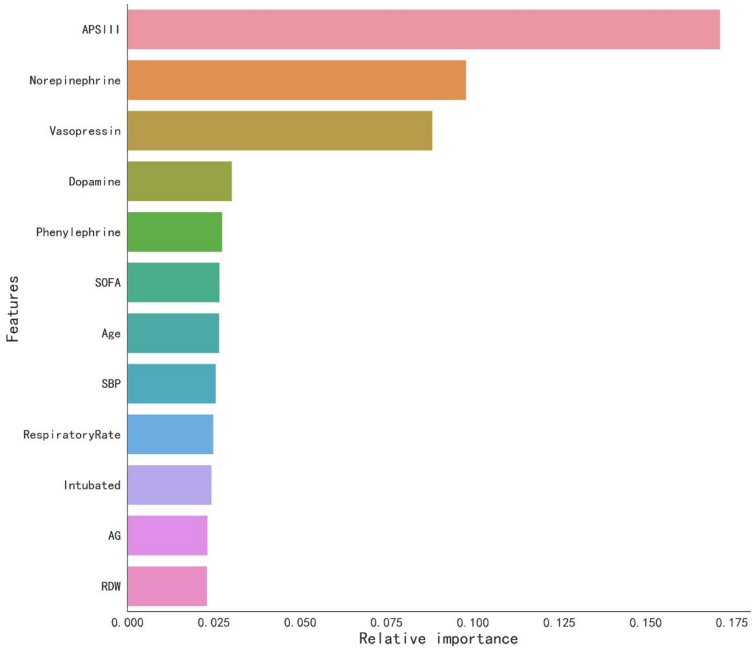


Figure 3. The Ranking of Features Importance

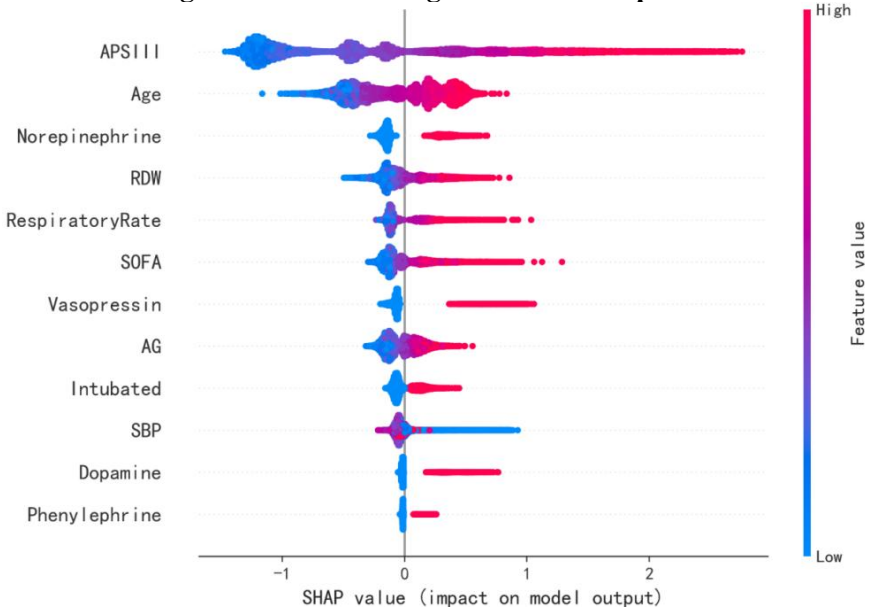


Figure 4. The Distribution of the Impacts of Features

4. Conclusion

This research constructs an interpretable XGBoost model, which shows outstanding performance in predicting mortality risk among heart failure patients. The adopted interpretable machine learning techniques serve to precisely pinpoint key risk factors and bolster clinicians' trust in the predictive outputs. This assists clinicians in rapidly identifying high-risk heart failure patients, allowing for timely and suitable clinical interventions. Consequently, the framework enables more personalized and proactive patient management. The model's transparent design further promotes its adoption into standard clinical workflows, thereby enhancing decision-making support.

Acknowledgements

The research was supported by: Yunnan Provincial Department of Education Science Research Fund Project (2023J1209); Science and Technology planning Project of Yunnan Province (202001AP070046).

References

[1] Alexander R L. Heart failure resulting from cancer treatment: still serious but an opportunity for prevention. systematic reviews, 2018, 7(1):313-323.

[2] Mahoto K. The Concept of Heart Failure: Chronic Diseases Accompanied by an Attack of Acute Exacerbation. Tokyo: Therapeutic Strategies for Heart Failure. 2018:1-15.

[3] Peng Peichi, Liu Wenxian, Zhao Han, et al Analysis of prognostic factors for heart failure caused by hypertension Chinese Journal of Modern Medicine, 2018 (4): 9-13.

[4] Chen L M, Levine D A, Hayward R, et al. Relationship between Hospital 30-Day Mortality Rates for Heart Failure and Patterns of Early Inpatient Comfort Care. Journal of Hospital Medicine, 2018, 13(3):170-176.

[5] Rohan K, Kumar D, Harlan M. et al. Rising Mortality in Patients With Heart Failure in the United States. Jacc Heart Failure, 2018, 6(7):610-612.

[6] Ma Liyuan, Wu Yazhe, Wang Wen, et al Interpretation of Key Points in the 2017 China Cardiovascular Disease Report Chinese Journal of Cardiology, 2018, 23 (1): 3-6.

[7] Aaronson K D, Schwartz J S, Chen T M, et al. Development and prospective validation of a clinical index to predict survival in ambulatory patients referred for cardiac transplant evaluation. Circulation, 1997, 95(12):2660-2667.

[8] Gao, Z. Y. Research on Deep Learning Methods for Heart Failure Risk Assessment Based on Electronic Health Records. University of Science and Technology Beijing, 2025. DOI: 10.26945/d.cnki.gbjku.2025.000095.

[9] ZENG Jing, HE Xiaolong, HU Huajuan, et al., Construction of risk prediction model for predicting death or readmission in acute heart failure patients during vulnerable phase based on machine learning, Journal of Army Medical University, 2024, 46(7): 738-745, http://dx.doi.org/10.16016/j.2097-0927.202312147.

[10] Xu Qian, Xu Cuirong, Cai Xue, et al. Research Progress of Machine Learning-Based Prediction Models for Heart Failure Risk. Journal of Nursing Science, 2024, 52(5): 807-815.

[11] Xie Qiuhua, Lu Zuohua, Deng Shengqiong, et al. Evaluation of Heart Failure Discrimination Model for NT-proBNP Grey Value Patients Based on Machine Learning Algorithm. Journal of Tongji University: Medicine, 2021, 42 (3) : 6. DOI: 10.12289 / j.i SSN. 1008-0392.20433.

