

RESEARCH

Open Access



Elucidating predictors of preoperative acute heart failure in older people with hip fractures through machine learning and SHAP analysis: a retrospective cohort study

Qili Yu^{1,2}, Mingming Fu⁴, Zhiyong Hou^{3*} and Zhiqian Wang^{1*}

Abstract

Background Acute heart failure (AHF) has become a significant challenge in older people with hip fractures. Timely identification and assessment of preoperative AHF have become key factors in reducing surgical risks and improving outcomes.

Objective This study aims to precisely predict the risk of AHF in older people with hip fractures before surgery through machine learning techniques and SHapley Additive exPlanations (SHAP), providing a scientific basis for clinicians to optimize patient management strategies and reduce adverse events.

Methods A retrospective study design was employed, selecting patients admitted for hip surgery in the Department of Geriatric Orthopedics at the Third Hospital of Hebei Medical University from January 2018 to December 2022 as research subjects. Data were analyzed using logistic regression, random forests, support vector machines, AdaBoost, XGBoost, and GBM machine learning methods combined with SHAP analysis to interpret relevant factors and assess the risk of AHF.

Results A total of 2,631 patients were included in the final cohort, with an average age of 79.3 ± 7.7 . 33.7% of patients experienced AHF before surgery. A predictive model for preoperative AHF in older people hip fracture patients was established through multivariate logistics regression: $\text{Logit}(P) = -2.262 - 0.315 \times \text{Sex} + 0.673 \times \text{Age} + 0.556 \times \text{Coronary heart disease} + 0.908 \times \text{Pulmonary infection} + 0.839 \times \text{Ventricular arrhythmia} + 2.058 \times \text{Acute myocardial infarction} + 0.442 \times \text{Anemia} + 0.496 \times \text{Hypokalemia} + 0.588 \times \text{Hypoalbuminemia}$, with a model nomogram established and an AUC of 0.767 (0.723–0.799). Predictive models were also established using five machine learning methods, with GBM performing optimally, achieving an AUC of 0.757 (0.721–0.792). SHAP analysis revealed the importance of all variables, identifying acute myocardial infarction as the most critical predictor and further explaining the interactions between significant variables.

Qili Yu and Mingming Fu are co-first authors.

*Correspondence:

Zhiyong Hou
drzyhou@hebmu.edu.cn
Zhiqian Wang
37800709@hebmu.edu.cn

Full list of author information is available at the end of the article



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

Conclusion This study successfully developed a predictive model based on machine learning that accurately predicts the risk of AHF in older people with hip fractures before surgery. The application of SHAP enhanced the model's interpretability, providing a powerful tool for clinicians to identify high-risk patients and take appropriate preventive and therapeutic measures in preoperative management.

Keywords Heart failure, Hip fracture, Preoperative, Machine learning, SHAP, Prediction model

Introduction

With the acceleration of global population aging, hip fractures in the older people have become a significant public health challenge. Since records began in 1990, over 1.6 million people worldwide have suffered from hip fractures. It is predicted that, over time, especially among the older people, the incidence of hip fractures will show a gradually increasing trend. By 2050, the number of individuals affected is expected to rise to at least 4.5 million [1, 2]. These fractures not only increase the mortality and disability rates of patients, severely affecting the quality of life, but also impose a significant burden on the healthcare system, including the costs of surgical treatment, long-term rehabilitation, and the subsequent socio-economic burdens. In particular, the risk of AHF before surgery in older people with hip fractures has significantly increased, becoming a key complication closely associated with high mortality rates and prolonged hospital stays, further exacerbating the risk of postoperative complications, including infections and re-fractures [3]. Therefore, developing effective prediction and prevention strategies is crucial for improving the treatment outcomes of this patient group.

In clinical practice, we have observed that most surgeons tend to overlook the assessment of heart failure biomarkers such as Brain Natriuretic Peptide (BNP) or N-Terminal pro-B-Type Natriuretic Peptide (NT-proBNP) in the preoperative evaluation of older people with hip fractures. This oversight could miss patients who have developed AHF, thus failing to intervene timely in this potentially high-risk state [4]. Machine learning offers a new perspective and approach by analyzing vast amounts of patient data to predict complications that may arise after a hip fracture, such as preoperative AHF, thereby providing a scientific basis for clinical decision-making, optimizing patient management strategies, and reducing the incidence of adverse events [5, 6]. This study utilizes machine learning methods and SHAP values aimed at precisely predicting the risk of AHF before surgery in older people with hip fractures. By analyzing clinical data to reveal the complex associations between patient characteristics, laboratory test results, and preoperative complications, this research offers a new perspective and method. It not only enhances the accuracy of predictions but also provides actionable data support for doctors, optimizing patient management strategies, and reducing the occurrence of adverse events.

Therefore, the establishment and use of this study's model can alert physicians to conduct a more comprehensive preoperative assessment, including the measurement of BNP or NT-proBNP, thus identifying those high-risk patients. Such an integrated preoperative approach can not only reduce surgical risks and postoperative complications but also shorten hospital stays and potentially lower mortality rates. It provides a safer and more effective treatment plan for older people with hip fractures, significantly improving their prognosis and ultimately achieving the goal of improving the clinical outcomes of older people with hip fractures.

Materials and methods

Study design and patients

This retrospective study selected inpatients who underwent hip surgery at the Department of Geriatric Orthopedics, Hebei Medical University Third Hospital, from January 2018 to December 2022, as the study subjects. The inclusion criteria were: (1) aged 65 years and older; (2) hip fractures confirmed by radiographic examinations such as X-rays; (3) patients with complete medical records, laboratory test results, and other necessary medical documents. Exclusion criteria were lack of complete medical records, laboratory test results, or other necessary medical documents, and patients who did not meet the diagnostic criteria for hip fractures.

Ethical statement

This study, based on the retrospective analysis of existing case data, ensured that all patient data collection and analysis were conducted anonymously to protect patient privacy. Furthermore, the study was in compliance with the Declaration of Helsinki and had been approved and supported by the Institutional Review Board of Hebei Medical University Third Hospital (Approval No.: 2021-087-1).

Disease definition

AHF is a condition where there is a sudden decrease in the heart's ability to pump blood, leading to the body's circulatory volume being insufficient to meet metabolic demands. According to the European Society of Cardiology, AHF is caused by acute changes in the structure or function of the heart, accompanied by increased filling pressures and/or a significant reduction in ejection fraction. Common symptoms include shortness of breath,

pulmonary congestion, and inadequate organ perfusion. A key biochemical marker for diagnosing AHF is a significant increase in serum BNP or NT-proBNP levels [7]. Different thresholds of BNP and NT-proBNP are used for diagnosing AHF to accommodate patients of varying ages. Specifically, the diagnostic threshold for BNP is ≥ 300 pg/mL for patients of all ages. For NT-proBNP, the thresholds are age-stratified: >450 pg/mL for patients under 55 years of age; >900 pg/mL for those between 55 and 75 years; and >1800 pg/mL for patients 75 years and older. These elevated markers are instrumental in confirming the diagnosis of AHF [8]. In clinical practice, differentiating between an acute exacerbation of heart failure and chronic heart failure is key, focusing on changes in clinical symptoms and acute variations in BNP or NT-proBNP levels. Through this approach, physicians can more accurately diagnose AHF, thereby providing appropriate treatment for patients.

Data collection

This retrospective study is based on data collected from patients who underwent hip surgery in the Department of Geriatric Orthopedics at Hebei Medical University Third Hospital, from January 2018 to December 2022. Data on patients prior to surgery were gathered through the medical record system, including details on heart failure status, sex, age, admission time, and comorbidities such as hypertension, Old cerebral infarction, coronary artery disease, diabetes, chronic obstructive pulmonary disease (COPD), cancer, arrhythmias, pulmonary infections, ventricular arrhythmias, acute myocardial infarction, acute cerebrovascular disease, stress hyperglycemia, stress ulcers, urinary tract infections, anemia, hypokalemia, hyponatremia, hypoalbuminemia, and lower limb venous thrombosis. By analyzing this data, we aim to gain a deeper understanding of the risk factors for AHF in older people with hip fractures before surgery, thereby providing precise intervention suggestions for clinical practice.

Model establishment

Dataset configuration and variable selection

In this study, we utilized the logistic regression method to predict the occurrence of AHF in older people with hip fractures before surgery. Initially, the collected data were divided into training and validation sets at a ratio of 7:3 to ensure the adequacy of the training process and the independence of the evaluation process. To identify risk factors significantly associated with AHF and avoid overfitting, we used Backward Elimination for variable selection. In this approach, we started with a full model that included all candidate variables and progressively removed those with low statistical significance to optimize the model [9]. Furthermore, we utilized a

multivariate logistic regression model to assess the relationship between these factors and AHF, ensuring that only statistically significant predictors were included in the final model. By constructing a nomogram of the model, we made the prediction outcomes and the contributions of various variables both intuitive and easy to understand.

Multi-model development and validation process

Beyond the basic logistic regression model, we delved into five additional machine learning algorithms, including Random Forest (RF), Support Vector Machine (SVM), Adaptive Boosting (AdaBoost), Extreme Gradient Boosting (XGBoost) and gradient boosting machine (GBM). To ensure methodological consistency and allow for direct performance comparison across different models, the same set of variables selected via the Backward Elimination was used in developing each of these predictive models. The models were trained using a 5-fold cross-validation method, which helped us to more accurately evaluate their performance on unseen data [10]. Given the imbalance between positive and negative samples in our dataset, we incorporated the Synthetic Minority Over-sampling Technique (SMOTE) focusing on the minority class samples at the boundary to improve the sample distribution and optimize model performance.

Model interpretability and comprehensive evaluation

During the model evaluation phase, we not only focus on the model's discriminative ability, assessed by calculating the Area Under the Curve (AUC) value, but also on the model's calibration through the Hosmer-Lemeshow test. Additionally, Decision Curve Analysis (DCA) is applied to compute the net benefit at different thresholds, comprehensively evaluating the model's practical value in clinical decision-making [11]. The Clinical Impact Curve (CIC) is utilized to visualize the benefit values brought by different thresholds. To enhance the model's interpretability, we employed SHapley Additive exPlanations (SHAP) analysis, which demonstrates the contribution of different variables at the individual level. Furthermore, every observation in the dataset can be explicated with designated SHAP values.

Through the aforementioned strategy, our goal is to construct a model that is both precise and highly interpretable, providing a powerful tool for clinicians. This will enable them to identify the risk of AHF before surgery more promptly when treating older people with hip fractures.

Statistical analysis

In this study, our aim was to reveal the relationship between AHF and various risk factors in older people with hip fractures. Initially, we analyzed the baseline

information of participants through descriptive statistics. To streamline our analysis and enhance the robustness of our findings, we employed nonparametric methods for all continuous variables. This approach eliminates the need to differentiate between normally and non-normally distributed data, ensuring uniformity in our statistical treatment and enhancing the interpretability of our results. The distribution characteristics of categorical variables were presented in frequencies and percentages. Variance Inflation Factor (VIF) and tolerance were calculated to assess potential collinearity among parameters, with a VIF below 5 and tolerance above 0.1 considered as standards indicating no significant collinearity. All statistical analyses were conducted using R language. The level of statistical significance was set at $P < 0.05$.

Results

Patient baseline characteristics

Between January 2018 and December 2022, a total of 4,170 older people with hip fractures were included in our study. After screening, 1,539 patients were excluded, leaving 2,631 patients in the final analysis. The excluded patients comprised 1,077 with non-hip fractures, 328 non-surgical patients, and 134 with incomplete data (Fig. 1).

Table 1 presents the baseline clinical characteristics of the overall sample and compares those between the AHF group and the non-AHF group among older people with hip fractures, analyzed using R software. Overall, the mean age of the patients was 79.3 ± 7.7 , with 766 males (29.1%) and 1,865 females (70.9%). Among them,

888 patients (33.7%) experienced AHF before surgery. There were statistically significant differences in gender distribution, age, and age groups (< 75 years and ≥ 75 years) between the two groups ($p < 0.05$), as determined by R software. Regarding comorbidities, the prevalence of coronary artery disease, COPD, and arrhythmias was significantly higher in the AHF group compared to the non-AHF group ($p < 0.05$), with analysis conducted in R software. Additionally, preoperative complications such as pulmonary infection, ventricular arrhythmias, acute myocardial infarction, acute cerebrovascular disease, and urinary tract infections also showed a higher incidence in the AHF group, with significant statistical differences ($p < 0.05$) identified through R software.

Univariate analysis of laboratory data and ultrasound examinations

Table 2 displays the preoperative laboratory and lower limb venous ultrasound characteristics of older people with hip fractures. The incidences of anemia, hypokalemia, hyponatremia, and hypoalbuminemia were significantly higher in the AHF group compared to the non-AHF group, showing significant statistical differences ($p < 0.05$). However, there was no significant difference in the incidence of lower limb venous thrombosis between the two groups.

Development and validation of nomograms

Using R, patients were randomly divided into a training set and a test set in a 7:3 ratio, with 1,843 patients in the training set and 788 in the test set. Initial analysis

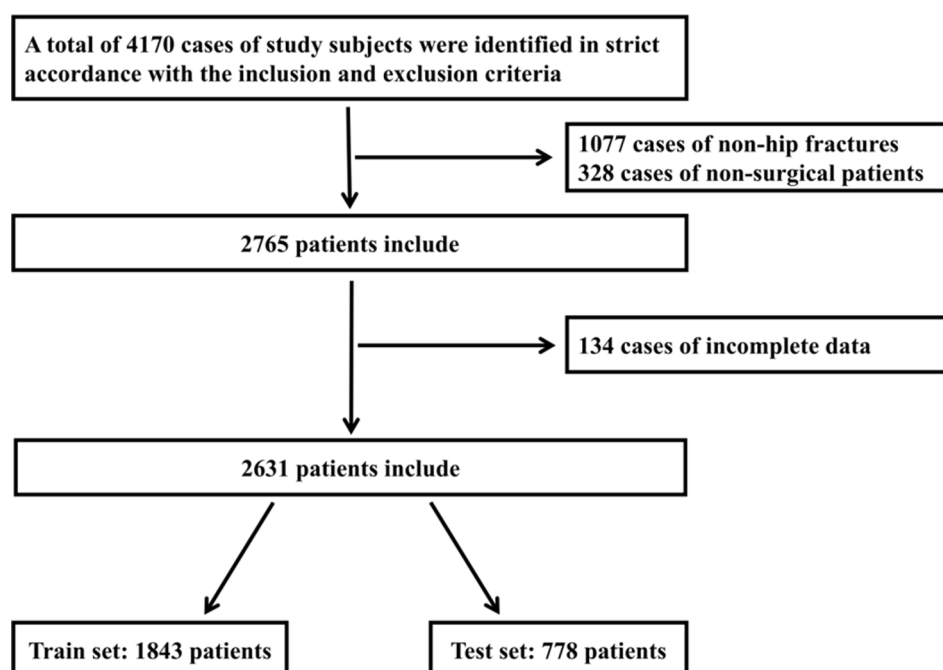


Fig. 1 The patient flow chart in our study

Table 1 Baseline clinical characteristics of hip fracture patients classified by acute heart failure

Variables	Total(N = 2631)	Non-acute heart failure(N = 1743)	Acute heart failure(N = 888)	p-value
Gender, N (%)				0.016
Male	766(29.1%)	534(30.6%)	232(26.1%)	
Female	1865(70.9%)	1209(69.4%)	656(73.9%)	
Age, mean $\pm$ SD (years)	79.3 $\pm$ 7.7	78.2 $\pm$ 7.8	81.6 $\pm$ 7.0	<0.001
Age group, N (%)				<0.001
<75years	746(28.4%)	588(33.7%)	158(17.8%)	
$\geq$ 75 years	1885(71.6%)	1155(66.3%)	730(82.2%)	
Admission time				0.331
<48 h	1813(68.9%)	1212(69.5%)	601(67.7%)	
$\geq$ 48 h	818(31.1%)	531(30.5%)	287(32.3%)	
Comorbidity N (%)				
Hypertension				0.530
Yes	1370(52.1%)	900(51.6%)	470(52.9%)	
No	1261(47.9%)	843(48.4%)	418(47.1%)	
Old cerebral infarction				0.145
Yes	1207(45.9%)	782(44.9%)	425(47.9%)	
No	1424(54.1%)	961(55.1%)	463(52.1%)	
Coronary heart disease				<0.001
Yes	705(26.8%)	426(24.4%)	279(31.4%)	
No	1926(73.2%)	1317(55.1%)	609(68.6%)	
Diabetes				0.967
Yes	662(25.2%)	439(25.2%)	223(25.1%)	
No	1969(74.8%)	1304(74.8%)	665(74.9%)	
COPD				0.001
Yes	300(11.4%)	173(9.9%)	127(4.3%)	
No	2331(88.6%)	1570(90.1%)	761(85.7%)	
Cancer				0.547
Yes	107(4.1%)	68(3.9%)	39(4.4%)	
No	2524(95.9%)	1675(96.1%)	849(95.6%)	
Arrhythmia				0.021
Yes	175(6.7%)	102(5.9%)	73(8.2%)	
No	2456(93.3%)	1641(94.1%)	815(91.8%)	
Complications				
Pulmonary infection				<0.001
Yes	376(14.3%)	198(11.4%)	178(20.0%)	
No	2255(85.7%)	1545(88.6%)	710(80.0%)	
Ventricular arrhythmia				<0.001
Yes	424(16.1%)	212(12.2%)	212(23.9%)	
NO	2207(83.9%)	1531(87.8%)	676(76.1%)	
Acute myocardial infarction				<0.001
Yes	172(6.5%)	53(3.0%)	119(13.4%)	
No	2459(93.5%)	1690(97.0%)	769(86.6%)	
Acute cerebrovascular disease				0.029
Yes	250(9.5%)	150(8.6%)	100(11.3%)	
No	2381(90.5%)	1593(91.4%)	778(88.7%)	
Stress hyperglycemia				0.084
Yes	38(1.4%)	20(1.1%)	18(2.0%)	
No	2593(98.6%)	1723(98.9%)	870(98.0%)	
Stress ulcer				0.838
Yes	27(1.0%)	19(1.1%)	8(0.9%)	
No	2604(99.0%)	1724(98.9%)	880(99.1%)	
Urinary tract infection				0.027

Variables	Total(N= 2631)	Non-acute heart failure(N= 1743)	Acute heart failure(N=888)	p-value
Yes	115(4.4%)	65(3.7%)	50(5.6%)	
No	2516(95.6%)	1678(96.3%)	838(94.4%)	

Values are presented as mean ± standard deviation, median (interquartile range), or number (percentage) as appropriate, SD Standard deviation, COPD Chronic Obstructive Pulmonary Disease

Table 2 The results of univariate analysis of laboratory data and ultrasound examination

Variables	Total(N= 2631)	Non-acute heart failure(N= 1743)	Acute heart failure(N=888)	p-value
Anemia				<0.001
Yes	802(30.5%)	476(27.3%)	326(36.7%)	
No	1829(69.5%)	1267(72.7%)	562(63.3%)	
Hypokalemia				<0.001
Yes	588(22.3%)	341(19.6%)	247(27.8%)	
No	2043(77.7%)	1402(80.4%)	641(72.2%)	
Hyponatremia				<0.001
Yes	739(28.1%)	445(25.5%)	294(33.1%)	
No	1892(71.9%)	1298(74.5%)	594(66.9%)	
Hypoalbuminemia				0.001
Yes	863(32.8%)	534(30.6%)	329(37.0%)	
No	1768(67.2%)	1209(69.4%)	559(63.0%)	
Lower extremity venous thrombosis				0.234
Yes	850(32.3%)	577(33.1%)	273(30.7%)	
No	1781(67.7%)	1166(66.9%)	615(69.3%)	

Values are presented as median (interquartile range), or number (percentage) as appropriate

Table 3 Prediction factors of preoperative acute heart failure in geriatric patients with hip fracture

	B	Odds ratio	P value	95%CI
Sex(Male)	-0.315	0.730	0.014	0.565–0.938
Age(≥75)	0.673	1.960	<0.001	1.509–2.563
Coronary heart disease	0.556	1.745	<0.001	1.374–2.216
Pulmonary infection	0.908	2.480	<0.001	1.865–3.302
Ventricular arrhythmia	0.839	2.313	<0.001	1.758–3.304
Acute myocardial infarction	2.058	7.836	<0.001	4.787–13.044
Anemia	0.442	1.556	<0.001	1.223–1.977
Hypokalemia	0.496	1.642	<0.001	1.271–2.120
Hypoalbuminemia	0.588	1.800	<0.001	1.391–2.333
Constant	-2.262	0.104	<0.001	

with Backward Elimination on the training set selected 15 variables out of 22. Subsequent multivariable logistic regression analysis identified gender, age, coronary heart disease, pulmonary infection, ventricular arrhythmia, acute myocardial infarction, anemia, hypokalemia, and hypoalbuminemia as independent risk factors for the occurrence of AHF before surgery in older people with hip fractures (Table 3; Fig. 2). Based on these independent risk factors, we developed a nomogram model to predict the probability of pre-surgical AHF in older people with hip fractures (Fig. 3). The predictive model is given by $\text{Logit}(P) = -2.262 - 0.315 \times \text{Sex} + 0.673 \times \text{Age} + 0.556 \times \text{Coronary heart disease} + 0.908 \times \text{Pulmonary infection} + 0.839 \times \text{Ventricular arrhythmia} + 2.058 \times \text{Acute myocardial infarction} + 0.442 \times \text{Anemia} + 0.496 \times$

$\text{Hypokalemia} + 0.588 \times \text{Hypoalbuminemia}$. The variance inflation factor (VIF) was calculated for each variable in the model, indicating all predictor variables had VIF values well below the threshold of 5, specifically: Sex 1.01, Age 1.01, Coronary heart disease 1.01, Pulmonary infection 1.01, Ventricular arrhythmia 1.02, Acute myocardial infarction 1.01, Anemia 1.11, Hypokalemia 1.02, Hypoalbuminemia 1.12.

The nomogram was evaluated through 1,000 bootstrap resampling, and the results showed that the calibration curve deviated only slightly from the perfect prediction line, indicating good agreement between the model's predictions and the actual observations (Fig. 4). The Area Under the Curve (AUC) of the Receiver Operating Characteristic (ROC) for the validation dataset was 0.767 (95% CI: 0.723–0.799), indicating robust predictive performance of the model (Fig. 5). Additionally, the cross-validation process was implemented across the entire dataset, yielding an average AUC of 0.760. These cross-validation metrics further substantiate the model's performance and underscore its reliability in diverse clinical settings. Moreover, the nomogram model's corrected C-statistic obtained through bootstrap resampling was 0.776, demonstrating good performance in internal validation. This means that the model has strong discriminative ability and can accurately predict the risk of AHF in patients. DCA indicates significant clinical decision-making value with a probability range of 8–90% in the training set (Fig. 6A) and 9–86% in the validation set

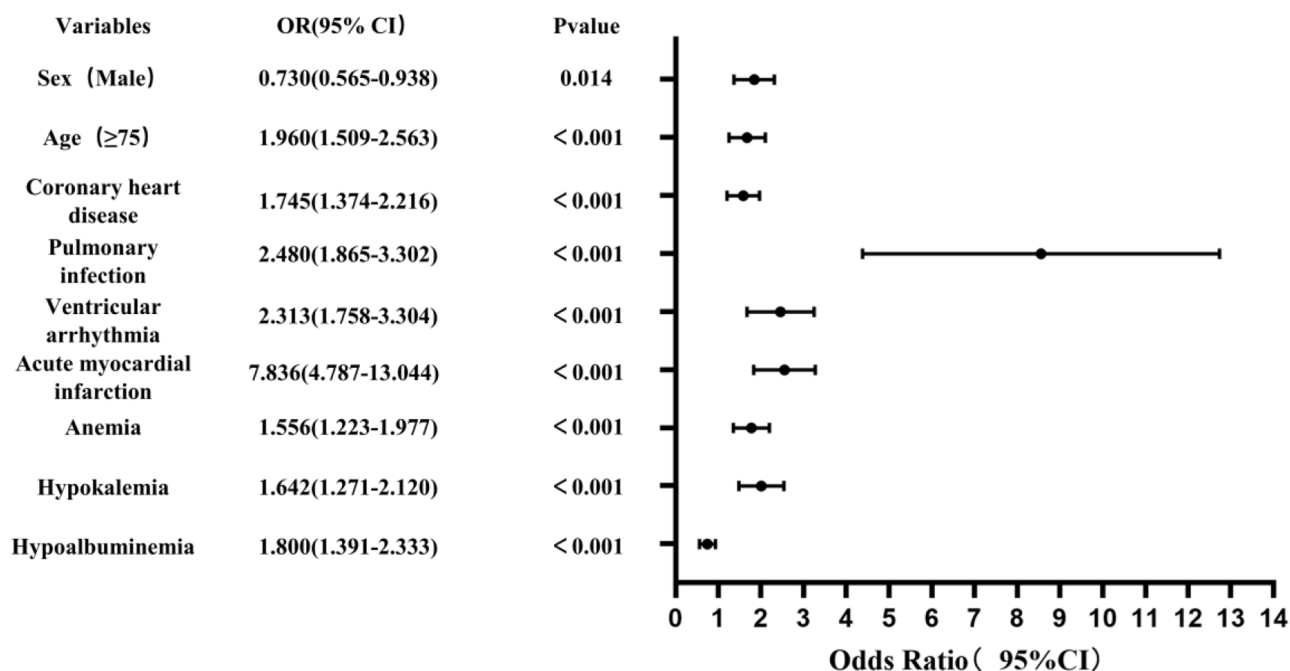


Fig. 2 Forest plot showing the relationship between risk factors and the occurrence of preoperative acute heart failure in older people with hip fracture

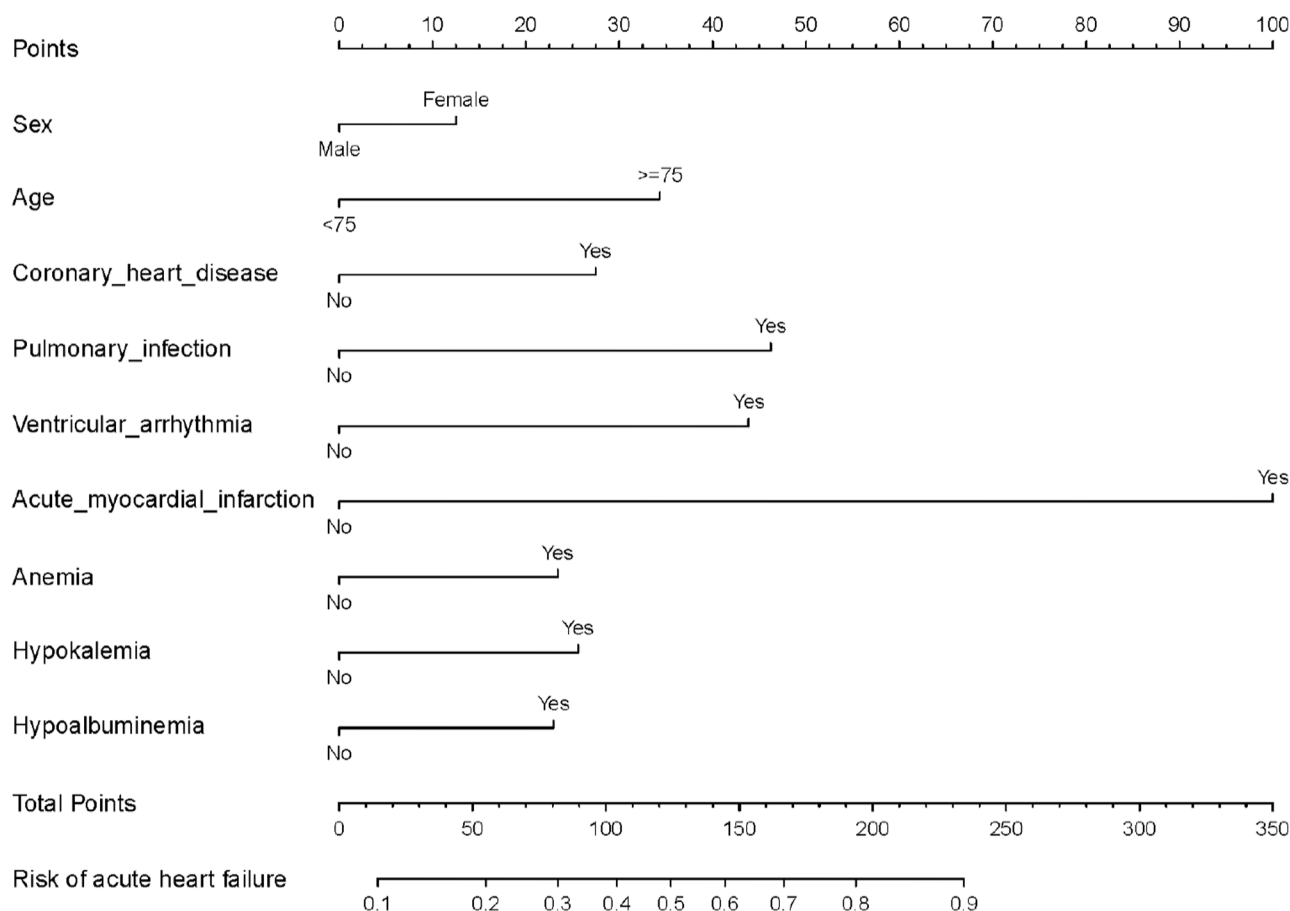


Fig. 3 A nomogram model for predicting the occurrence of preoperative acute heart failure in older people with hip fractures

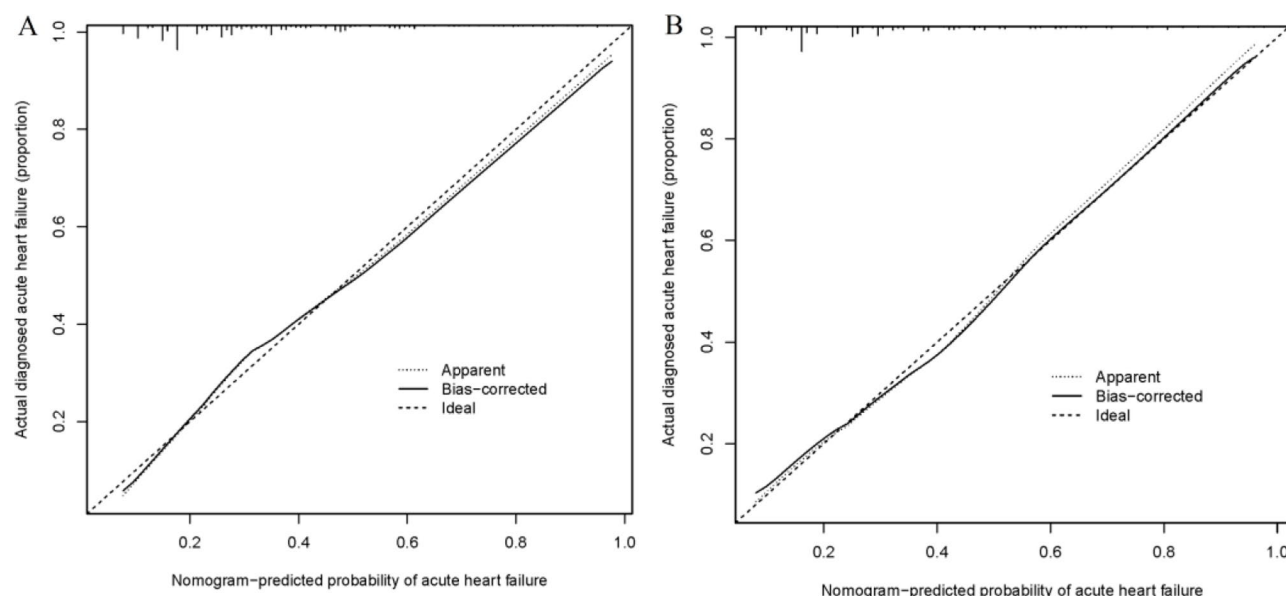


Fig. 4 Calibration curves of the acute heart failure nomogram prediction in the cohort. Panel **A** shows the calibration curve for the training dataset, and Panel **B** shows the curve for the test dataset. The x-axis represents the predicted acute heart failure risk. The y-axis represents the actual diagnosed acute heart failure. The diagonal dotted line represents a perfect prediction by an ideal model. The solid line represents the performance of the nomogram, of which a closer fit to the diagonal dotted line represents a better prediction

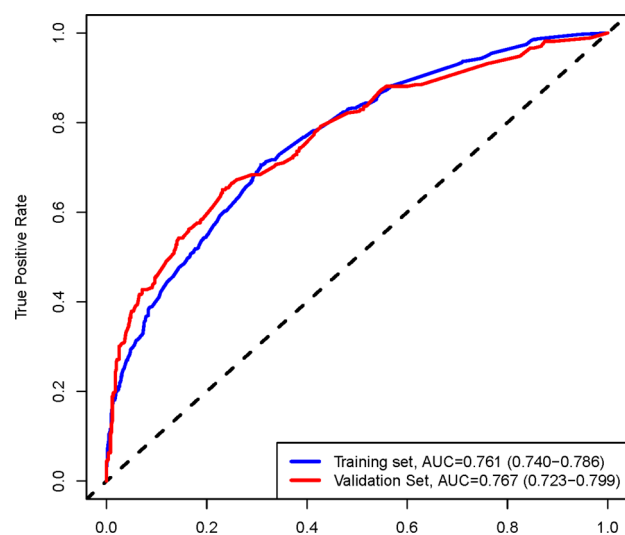


Fig. 5 Analysis of the ROC curve for the predictive values of preoperative acute heart conditions. The blue curve represents the ROC for the training set, with an area under the curve (AUC) of 0.761 (95% CI: 0.740–0.786), illustrating the model's performance on the dataset used for model development. The red curve represents the ROC for the validation set, with an AUC of 0.767 (95% CI: 0.723–0.799), indicating the model's performance on a separate dataset used to test the model. The dashed diagonal line represents the line of no discrimination, which a purely random classifier would achieve. The closer the ROC curve is to the top left corner, the higher the test's overall accuracy

(Fig. 6B). Additionally, the Clinical Impact Curve (CIC) demonstrates the effect of different threshold settings on the number of patients predicted by the model (Fig. 6C and D). This further suggests the model has substantial

application potential, especially in predicting the risk of AHF in older people after hip fractures. The model provides a powerful tool to more precisely predict the likelihood of AHF, thereby guiding clinicians towards more appropriate preventative and therapeutic measures. Implementing clinical interventions based on this model's predictions can effectively optimize patient management, likely leading to positive impacts on patient health outcomes.

Development of predictive models using machine learning methods

All raw data was preprocessed prior to being input into the machine learning model, including cleaning and transformation steps, to ensure data integrity and high quality for accurate handling and analysis by the machine learning algorithms. The features with the highest importance scores in standardization were Acute Myocardial Infarction, Ventricular Arrhythmia, Pulmonary Infection, and Anemia (Fig. 7A; Table 4). Correlations between variables were also calculated and are displayed in Figure (Fig. 7B), which presents a detailed correlation matrix of all input variables. The heatmap uses color gradients to represent the strength and direction of linear correlations, with red indicating positive correlations, blue indicating negative correlations, and white representing no correlation. This visualization enables a clearer understanding of the relationships between input variables, highlighting potential collinearity issues.

Subsequently, the predictive models including Random Forest (RF), Support Vector Machine (SVM), Adaptive

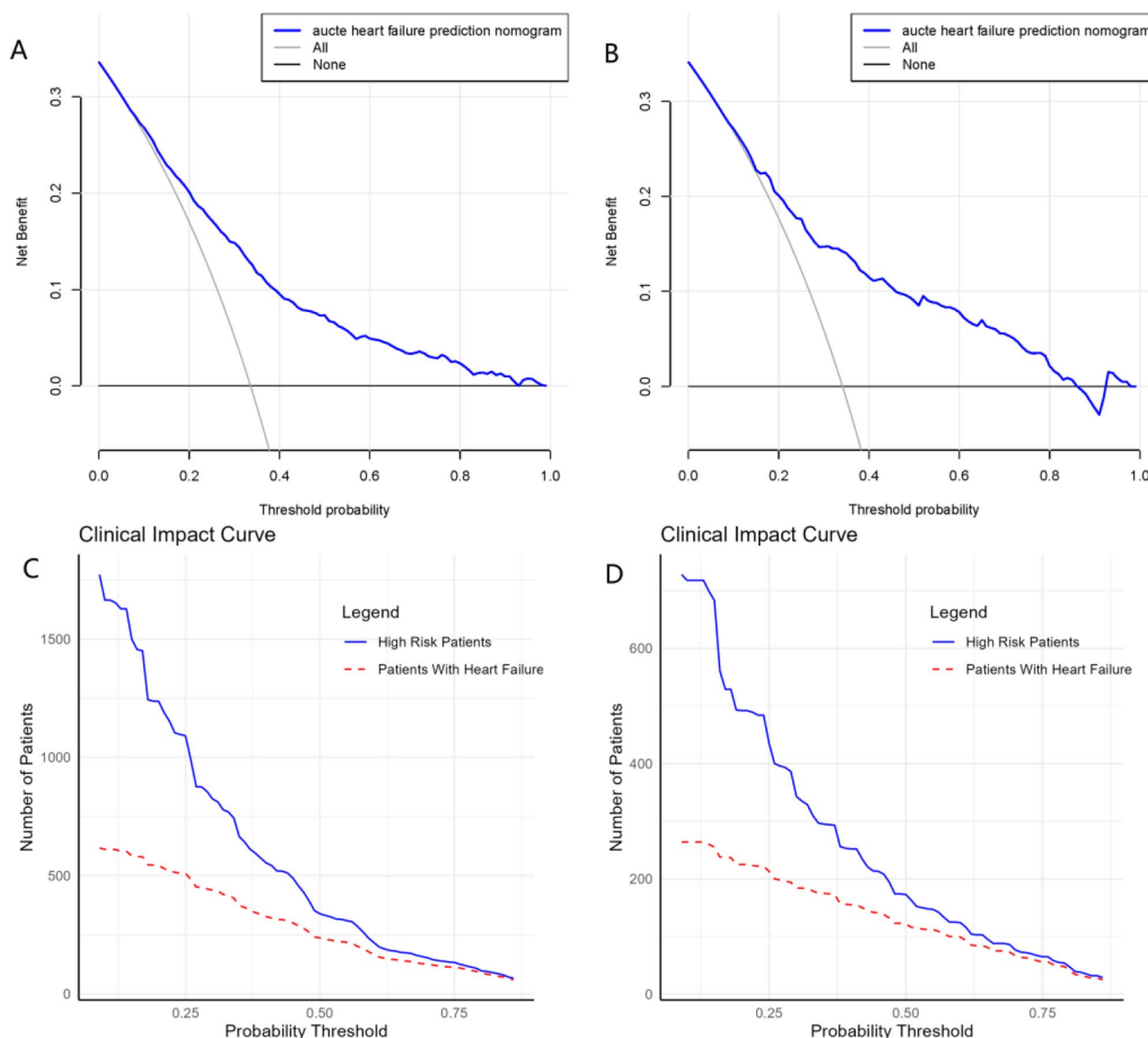


Fig. 6 Decision curve analysis (DCA) and Clinical Impact Curves (CIC) for the acute heart failure nomogram. **A** and **B** depict the DCA for the training and test datasets respectively, with the y-axis measuring net benefit. The blue line in each represents the performance of the acute heart failure risk nomogram. The grey solid line assumes all patients have acute heart failure, and the grey dashed line assumes no patients have the condition. **C** and **D** show the CIC for the training and test datasets respectively, with the y-axis indicating the number of patients. In **C** and **D**, the solid blue line represents high-risk patients as identified by the nomogram, and the dashed red line indicates the actual patients with heart failure. These graphs suggest that the nomogram provides a positive net benefit for clinical decision-making within a probability threshold range

Boosting (AdaBoost), Extreme Gradient Boosting (XGBoost), and Gradient Boosting Machine (GBM) were developed using the set of 15 variables selected through the Backward Elimination process. This consistent variable set across all models allows for a robust comparison of their effectiveness in predicting preoperative AHE. The models were evaluated on their performance metrics, with the Area Under the Curve (AUC) values obtained as follows: RF 0.746 (0.710–0.782), SVM 0.714 (0.676–0.752), AdaBoost 0.735 (0.699–0.772), XGBoost 0.747

(0.711–0.783), GBM 0.757 (0.721–0.792), with GBM showing the best AUC among the models (Fig. 8).

Accuracy, sensitivity, precision, and F1 scores were assessed for each predictive model. The results revealed that in terms of accuracy, GBM achieved the highest at 73.1%, closely followed by XGBoost with 73.0%. AdaBoost also displayed strong performance with an accuracy of 72.5%. When evaluating sensitivity, GBM outperformed all other models, reaching an impressive 95.6%, significantly higher than the rest, with SVM also performing well at 90.6%. In terms of precision, XGBoost

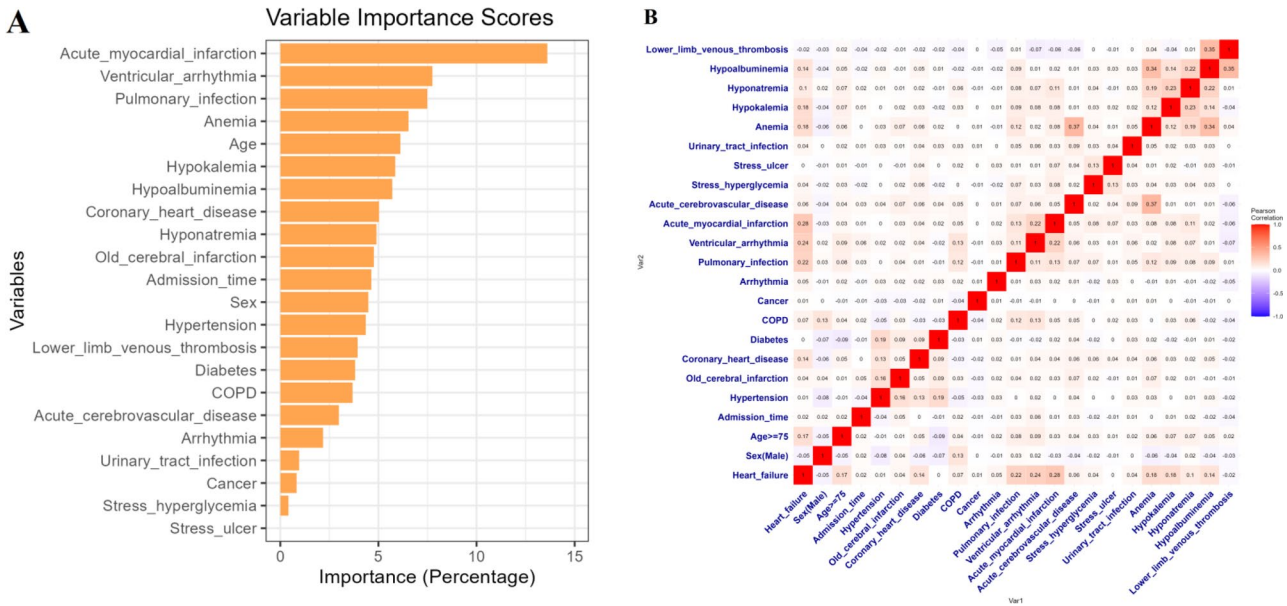


Fig. 7 Variable Importance and Correlation Matrix from Preprocessed Data in Machine Learning Model Analysis. **A** displays the variable importance scores, with the most significant features for the model's standardization being Acute Myocardial Infarction, Ventricular Arrhythmia, Pulmonary Infection, and Anemia. **B** shows the correlation matrix of the variables, with red indicating a strong positive correlation, blue a strong negative correlation, and white indicating no correlation. These visualizations provide the relationships between different clinical variables

Table 4 The exact data of importance of all the variables

Variables	Importance	Normalized importance
Acute myocardial infarction	100	0.13605442
Ventricular arrhythmia	57	0.07755102
Pulmonary infection	55	0.07482993
Anemia	48	0.06530612
Age	45	0.06122449
Hypokalemia	43	0.0585034
Hypoalbuminemia	42	0.05714286
Coronary heart disease	37	0.05034014
Hypertension	36	0.04897959
Old cerebral infarction	35	0.04761905
Hyponatremia	34	0.0462585
Admission time	33	0.04489796
Sex	32	0.04353741
Lower limb venous thrombosis	29	0.03945578
Diabetes	28	0.03809524
COPD	27	0.03673469
Acute cerebrovascular disease	22	0.02993197
Arrhythmia	16	0.02176871
Urinary tract infection	7	0.00952381
Cancer	6	0.00816327
Stress hyperglycemia	3	0.00408163
Stress ulcer	0	0

led the group with 75.9%, followed closely by RF at 74.9%. GBM's precision was lower at 72.7%. For the F1 score, which reflects a balance between precision and sensitivity, GBM showed the best result at 82.6%, indicating its effectiveness in providing a harmonious balance of recall

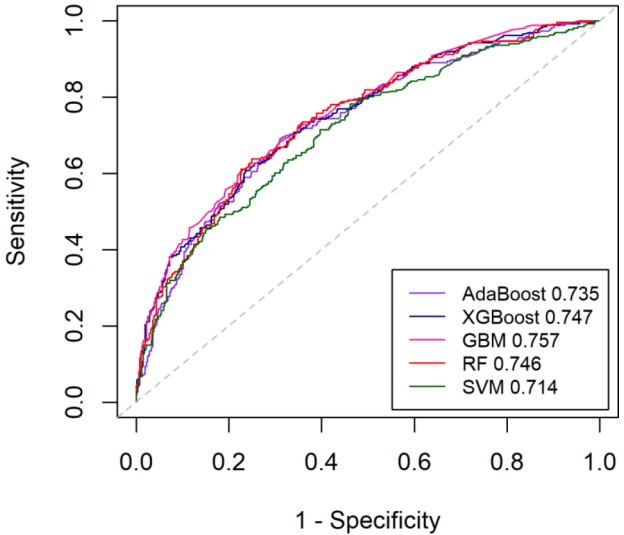


Fig. 8 Receiver Operating Characteristic (ROC) curves for various machine learning models in the evaluation of the dataset. The curves compare the sensitivity (true positive rate) and 1 - specificity (false positive rate) across different thresholds for Random Forest (RF), Support Vector Machine (SVM), AdaBoost, Extreme Gradient Boosting (XGBoost), and Gradient Boosting Machine (GBM). Area Under the Curve (AUC) values are displayed in the legend, with GBM showing the highest AUC of 0.757

and precision. Additionally, Youden's Index ($YI = \text{Sensitivity} + \text{Specificity} - 1$) was calculated and incorporated as an additional performance metric to evaluate the balance between sensitivity and specificity for each model. Among the predictive models, GBM demonstrated the highest YI (0.687), highlighting its superior ability to

Table 5 Comparing parameters of preoperative acute heart failure models in older people with hip fractures

	AUC	Accuracy	Sensitivity	Precision	F1	YI
RF	0.746	72.0%	87.5%	74.9%	80.7%	0.595
SVM	0.714	71.7%	90.6%	73.1%	80.9%	0.623
AdaBoost	0.735	72.6%	87.2%	44.0%	80.8%	0.578
XGBoost	0.747	73.0%	87.9%	75.9%	81.4%	0.609
GBM	0.757	73.1%	95.6%	72.7%	82.6%	0.687
LR	0.761	72.5%	79.5%	24.4%	37.3%	0.520

AUC area under the curve of ROC, RF random Forest, SVM support vector machine, AdaBoost adaptive boosting, XGBoost extreme gradient boosting, GBM gradient boosting machine, LR logistic regression, YI Youden's index

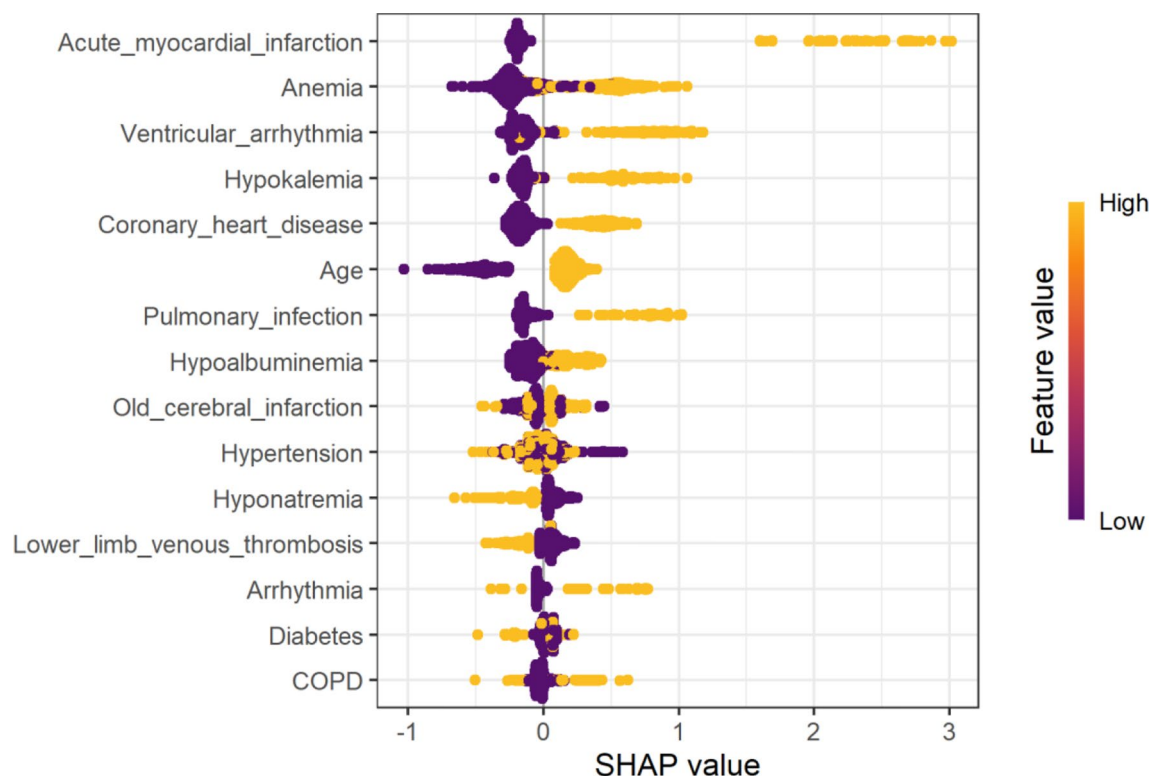


Fig. 9 SHAP value analysis for predictive modeling of acute heart Failure in older people with hip fractures. This Feature Importance Plot visualizes the impact of individual features on the prediction of acute heart failure risk. Each dot represents an observation, plotted against its SHAP value on the x-axis. The direction and magnitude of these SHAP values indicate whether the feature increases (positive value) or decreases (negative value) the risk of acute heart failure according to the model. The color gradient signifies the value of the feature, ranging from low (purple) to high (yellow)

achieve an optimal trade-off between sensitivity and specificity. SVM ranked second with a YI of 0.623, followed by XGBoost with a YI of 0.609 (Table 5).

SHAP analysis was conducted to understand the impact of multiple features on the predictive model for AHF in older people with hip fractures before surgery (Fig. 9). The Feature Importance Plot shows each observation as a dot, with the SHAP value on the x-axis indicating the impact of the feature on the model's output. Positive values indicate contributions that increase risk, while negative values indicate contributions that decrease risk. The color gradient from purple to yellow represents feature values from low to high. It is observed that the SHAP values for Acute Myocardial Infarction are distributed in the positive region, with several higher positive

points indicating that the presence of acute myocardial infarction significantly increases the risk of AHF. Conversely, Hyponatremia shows both positive and negative SHAP values, concentrated near zero, suggesting a relatively small or individual-dependent impact on the prediction. However, the SHAP values for COPD are mainly in the negative region, possibly indicating a lower risk of AHF in patients with COPD in this model. Through individual-level predictive behavior analysis using the SHAP algorithm, the model revealed key variables influencing the risk of AHF for four patients, showing the contribution of each factor to the prediction and identifying Acute Myocardial Infarction as the main variable affecting all patients. Its SHAP value was significantly higher than other features, and we also found that Anemia,

Ventricular Arrhythmia, and Pulmonary Infection play important roles in increasing the risk of heart failure (Figs. 10A-D). The SHAP values of these variables provide positive contributions, reinforcing their importance in risk assessment, consistent with the overall trends in the Feature Importance Plot.

By constructing multivariate dependence plots (Fig. 11), we suggest interactions between variable features, such as between Acute Myocardial Infarction and Anemia, where scatter plots reveal their association in predicting the risk of AHF. With an increase in the feature value of acute myocardial infarction, a significant rise in SHAP values is observed, especially at higher

feature values of acute myocardial infarction, where we see a cluster of yellow dots in the upper right corner of the graph. These yellow dots represent higher values of anemia, implying that in the context of high values of acute myocardial infarction, anemia's predictive contribution to the risk of AHF increases. Conversely, when the feature values of acute myocardial infarction are lower, the dots, mostly shown in purple and concentrated in the lower left corner of the graph, represent a smaller predictive contribution to heart failure risk. In addition to the demonstrated relationship between Acute Myocardial Infarction (AMI) and Anemia, Fig. 11 also presents multivariate dependence plots that reveal important

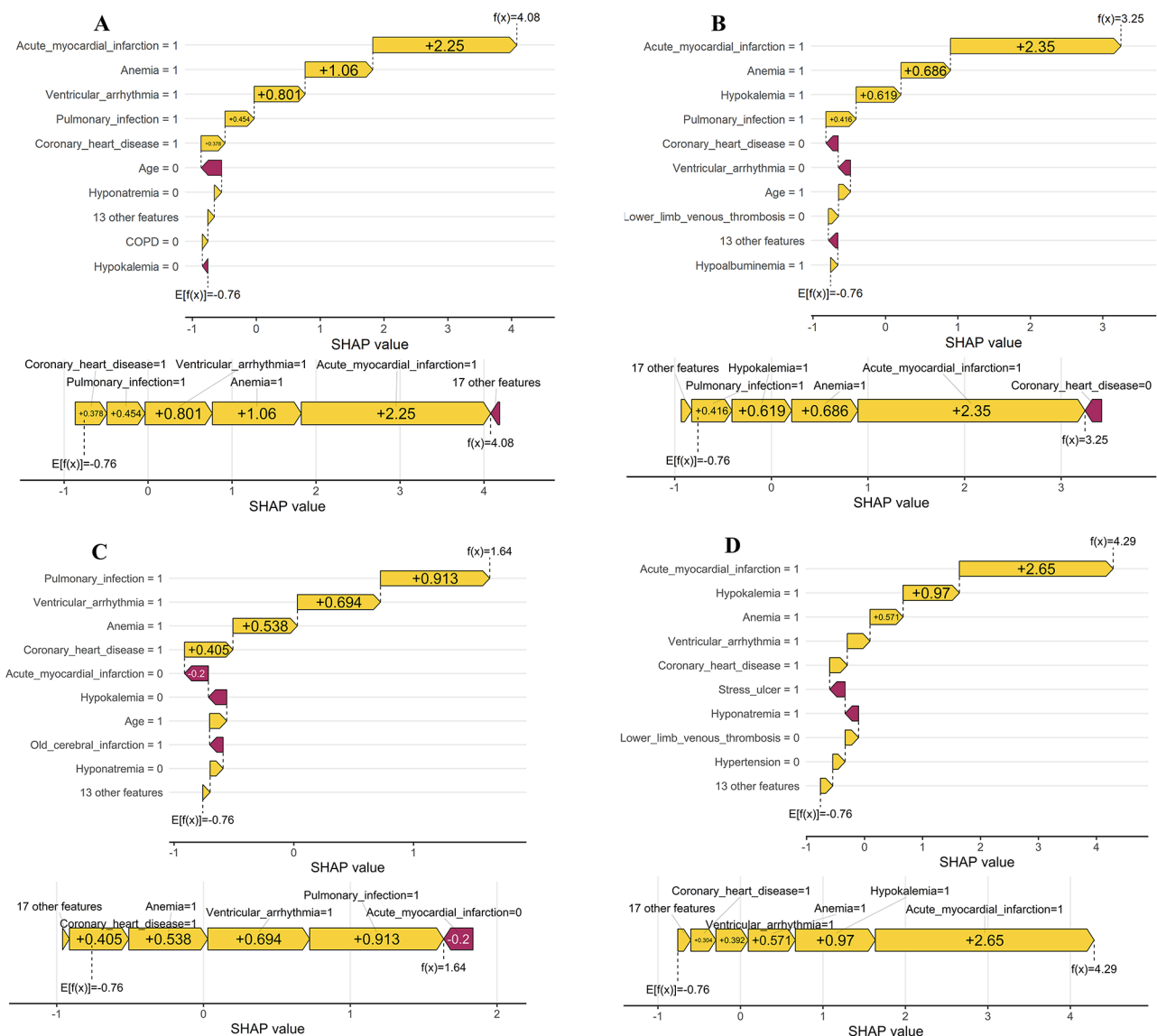


Fig. 10 SHAP value distributions for individual predictive analysis across four patients (A-D). These plots display the influence of various clinical features on the model's prediction of acute heart failure risk for each patient. In each subfigure, the x-axis represents the SHAP value, indicating the impact level of each feature. Features with higher SHAP values contribute more significantly to the prediction. Across all patients, Acute Myocardial Infarction is consistently the most influential variable with the highest SHAP values, indicating a strong association with increased heart failure risk

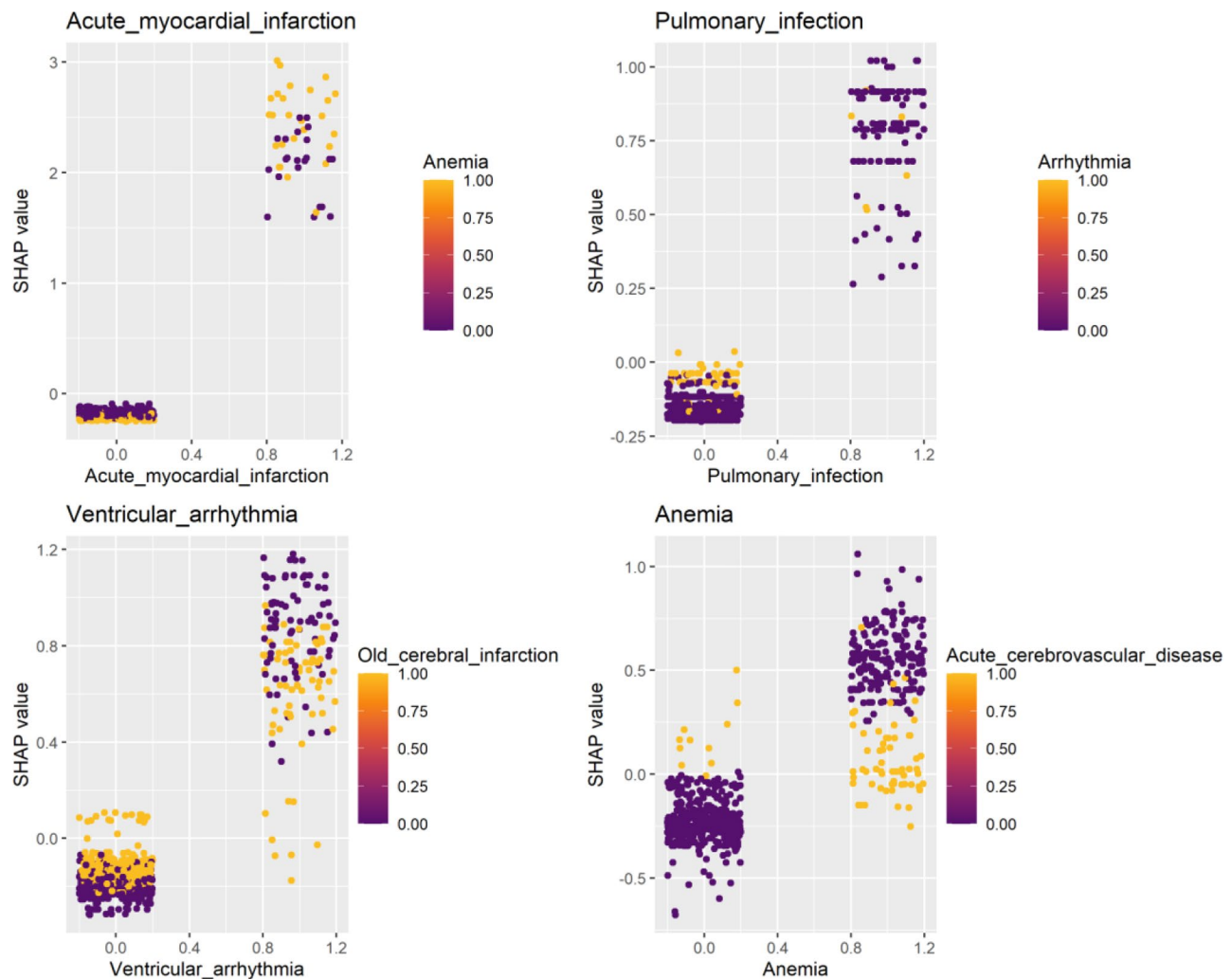


Fig. 11 Multivariate dependence plots demonstrating feature interactions in acute heart failure risk prediction. Each plot illustrates the relationship between a specific feature and SHAP values, which quantify the impact on the model's output. The color gradient, from purple to yellow, shows the value of one feature relative to another, with yellow indicating higher values. The plots reveal non-linear interactions between features, indicating complex relationships that are crucial for understanding the model's predictions

interactions between other clinical features. **Pulmonary Infection and Anemia:** The scatter plot shows an increasing trend of SHAP values for anemia as the severity of pulmonary infection intensifies. The dots transition from lower to higher SHAP values with increased pulmonary infection severity, suggesting a strengthening influence of anemia on heart failure risk predictions in the context of worsening pulmonary conditions. **Ventricular Arrhythmia and Old Cerebral Infarction:** This plot indicates a correlation between the presence of ventricular arrhythmias and higher SHAP values when old cerebral infarction is noted. More intense yellow dots appear as the feature values of ventricular arrhythmia increase alongside old cerebral infarction, highlighting a potential interactive effect on heart failure risk. **Anemia and Acute Cerebrovascular Disease:** Here, the distribution of dots illustrates how elevated anemia levels, combined

with acute cerebrovascular disease, lead to higher SHAP values. This pattern suggests an amplified predictive significance of anemia under the burden of acute cerebrovascular conditions.

Discussion

Hip fractures are a common type of fracture among the older people population, significantly impacting the quality of life of patients. Therefore, timely surgical intervention is crucial for restoring normal life functions and independence in patients [12]. However, the occurrence of AHF (AHF) preoperatively in older people with hip fractures is a common and serious complication. This complication not only increases surgical risks but may also prolong hospital stays, elevate medical costs, and make postoperative recovery more challenging, even leading to patient mortality [13, 14]. In older people with

hip fractures, predicting the risk of preoperative AHF is key to improving patient outcomes and reducing medical costs [15]. Recent studies have shown that predictive models constructed using multivariate logistic regression models and machine learning methods can identify high-risk patients. It has been found that advanced age (≥ 70 years), hypertension, anemia, hypoalbuminemia, and surgical duration exceeding 120 min are risk factors for heart failure in older people with hip fractures. Understanding these risk factors provides important references for the perioperative management of older people with hip fractures [16]. However, we have developed a logistic regression model and five machine learning models through retrospective studies to predict the likelihood of AHF preoperatively in older people with hip fractures and employed SHAP to offer an explanation of feature vector importance and the interactions among vectors in machine learning models, enhancing model transparency and interpretability. This provides clinicians with a quantitative tool to assess the risk of AHF preoperatively in older people with hip fractures, allowing for more targeted preventive and therapeutic measures in preoperative management, thereby improving patient outcomes.

With the advent of the big data era, machine learning models have gained increasing attention due to their ability to handle large datasets, identify complex nonlinear relationships, and interactive effects. This technology has shown immense potential in medical fields such as heart failure, where analyzing big data from electronic health records can not only identify subtypes of heart failure but also improve risk prediction, offering possibilities for personalized medicine [17]. A recent article published in *The Lancet* highlights the advantages of machine learning methods. Amitava Banerjee's team utilized machine learning to classify and predict outcomes of heart failure by analyzing large electronic health record datasets, clarifying classifications of heart failure patients, utilizing polygenic risk scores to measure their relevancy, and explaining potential biological mechanisms between different heart failure subtypes [18]. For the perioperative assessment of patients with hip fractures, a research team developed a predictive model to evaluate the risk of AHF perioperatively. This model is based on multivariate logistic regression analysis, covering factors such as respiratory diseases, history of heart disease, and ASA scores [19].

However, it's noteworthy that our research differs from previous studies as we focus more on predicting AHF preoperatively in older people with hip fractures. Our study found that, in terms of AUC value, the LR model (0.761) marginally outperforms the GBM model (0.757), suggesting a slightly better overall capability in distinguishing patients with or without AHF. AUC was used as the primary metric to compare model performance

because it is a threshold-independent measure that reflects the model's overall discriminative ability across all possible classification thresholds. At the same time, the nomogram developed from the logistic regression model provides a significant advantage by offering a visual and user-friendly tool for risk assessment of preoperative AHF in older people with hip fractures. This nomogram translates complex clinical data into a straightforward points system, where each predictor variable is assigned a score. Clinicians can quickly calculate a patient's total risk score by summing these individual scores, which directly corresponds to a probability of outcome on a visual scale. This functionality not only simplifies the decision-making process but also enhances the understanding of patient-specific risk factors, facilitating tailored intervention strategies. Such a tool is invaluable in clinical settings, supporting rapid and informed decisions that are critical in managing the acute care of these patients.

Although there is little difference in accuracy between the two models, GBM exhibits higher performance in sensitivity, precision, and F1 score. These threshold-dependent metrics provide complementary information to AUC by offering additional insights into the practical classification capabilities of the model under a specific threshold (e.g., 0.5). Specifically, in clinical settings such as preoperative AHF prediction, high sensitivity ensures that most true positive cases are correctly identified, which is critical for minimizing the risk of missed diagnoses (false negatives). This is particularly important given the severe consequences of untreated acute heart failure. While a lower specificity indicates a higher false positive rate, this trade-off can be acceptable in scenarios where the clinical priority is to identify as many high-risk patients as possible for further evaluation and intervention.

Additionally, Youden's Index ($YI = \text{Sensitivity} + \text{Specificity} - 1$) was calculated as a supplemental performance metric to better evaluate the models' abilities to balance sensitivity and specificity. Among the predictive models, GBM achieved the highest YI (0.687), reinforcing its superior capability to strike an optimal balance between true positive and true negative predictions. These results provide a more comprehensive understanding of model performance and further validate the robustness of GBM in clinical applications. Specifically, the YI values align with the previously reported advantages of GBM in sensitivity and precision, highlighting its reliability in minimizing misclassification and improving diagnostic accuracy. GBM's superior performance in these areas suggests it is more reliable for identifying patients genuinely at risk, potentially reducing the likelihood of misdiagnosis and improving positive predictive accuracy. Given these considerations, GBM might often be more

suitable as a diagnostic tool for preoperative AHF in older people with hip fractures due to its enhanced ability to correctly classify affected individuals. However, the choice of the model in clinical research should still be dictated by comprehensive consideration of the actual application scenarios and needs. Factors such as the clinical team's familiarity with the model, the availability of computational resources, and the necessity for transparent, interpretable results all play critical roles in this decision. In environments where quick, clear decisions are paramount, and the stakes of misclassification are high, the enhanced sensitivity and precision of GBM could be particularly valuable. Nonetheless, the accessibility and straightforwardness of logistic regression might favor its use in contexts where simplicity and speed are prioritized over maximal predictive accuracy. This nuanced approach to model selection, emphasizing a balance between statistical performance and practical applicability, is essential for effectively implementing predictive models in real-world clinical settings.

In our study, the use of the acute heart failure nomogram in clinical decision-making is critically evaluated through Decision Curve Analysis (DCA) and Clinical Impact Curves (CIC), presented in Fig. 6. These analyses are pivotal in understanding the balance between benefit and harm as the threshold probabilities for predicting acute heart failure are adjusted. As observed in panels A and B of Fig. 6, when the threshold probability is set lower than 0.25, there is a noticeable decrease in the net benefit. This is indicative of a higher rate of false positives, where the model predicts heart failure in more patients than those who actually have the condition. Such a lower threshold might be employed in clinical settings where the cost of missing a true case of heart failure is deemed higher than the risk of unnecessary treatment for patients without the condition. However, this comes at the cost of increased interventions based on false positives, which could lead to unwarranted patient anxiety, unnecessary tests, and treatments. The Clinical Impact Curves in panels C and D further illustrate the effect of these thresholds on clinical practice. At lower thresholds, more patients are identified as at risk, potentially ensuring that no actual cases are missed. However, this also means treating many patients who do not need treatment, which can strain healthcare resources and affect the overall quality of care. Through detailed threshold analysis, our study underscores the importance of tailored threshold settings based on specific clinical environments and patient populations. This tailored approach ensures that predictive models, while powerful, are applied thoughtfully to enhance patient outcomes and resource utilization effectively.

In recent research, SHAP values have played a crucial role in interpreting complex machine learning models in

the field of heart failure, helping to identify key predictive factors that could impact patient outcomes [20]. For instance, studies have utilized SHAP values to highlight the importance of different clinical variables in predicting the 3-year all-cause mortality rate among patients with chronic heart failure, providing clinicians with valuable model interpretations [21]. In assessing the risk of AHF preoperatively in older people with hip fractures, our machine learning model combined with SHAP values offers more objective and effective support for clinical decision-making. This approach allows us to quantify the contribution of each clinical feature variable to the prediction model, which is particularly valuable in handling multivariate and complex medical data.

In our study, through the analysis combining machine learning models with SHAP, we found that acute myocardial infarction, ventricular arrhythmia, pulmonary infection, and anemia are the four most important feature variables affecting the model's predictions. These factors are closely related to the occurrence of AHF. Acute myocardial infarction, a significant manifestation of cardiovascular disease, directly relates to a sharp decline in cardiac function, which is particularly important in the older people population as it may exacerbate existing cardiac burdens [22]. Ventricular arrhythmias could be an early warning of insufficient cardiac pump function, and in high-risk populations, it may precede heart failure [23]. Pulmonary infections can increase cardiac load, especially in older people with hip fractures requiring high cardiac output, potentially exacerbating existing cardiac conditions [24]. Anemia, by reducing oxygen-carrying capacity, can affect the cardiac oxygenation status, thereby increasing the cardiac workload [25].

Through multiple variable partial dependence plot analyses of feature variables in our study, we observed that although each variable contributes uniquely to the risk prediction of heart failure, they are not isolated. There may be interactions among them, meaning the presence of certain variable combinations could increase or decrease the risk of AHF. For example, an anemic condition could exacerbate the risk of heart failure caused by arrhythmias. This finding is consistent with other studies, such as Richard J. and colleagues, who found that patients with chronic kidney disease (CKD) and end-stage renal disease (ESRD) often have anemia and electrolyte imbalances, which may promote electrical instability, induce reentrant arrhythmias, and ultimately lead to congestive heart failure or even induce sudden cardiac death [26]. By identifying these key predictive factors, we can better understand and interpret the results of model predictions. These insights remind clinicians to promptly identify and focus on these key risk factors when assessing the surgical risk of older people with hip fractures. More targeted strengthening of cardiac protection and

monitoring, optimizing preoperative management strategies, improving overall treatment effectiveness, improving long-term prognosis, and preventing adverse events are advised.

Limitations

Although this study has constructed a risk prediction model for preoperative AHF in older people with hip fractures using machine learning methods, it still faces several limitations. First, the selection and scope of samples are restricted, as the study is based on the data of older people femoral fracture patients from a specific hospital. This selection may limit the general applicability of the study results, especially under different regional and medical conditions. Second, this retrospective study may have missed some patients with heart failure not included in the sample, introducing a certain bias in sample selection. Third, the interpretability of machine learning models remains a concern. Despite the increased interpretability through SHAP analysis, machine learning models are often considered “black box” models, which may limit their application in clinical decision-making. Fourth, there may be important predictive variables not included in the model that could significantly affect the risk of heart failure. Fifth, although the study was divided into training and validation sets and cross-validation was performed, it still belongs to internal validation without external validation, which is a limitation.

Conclusion

In summary, we have constructed a prediction model for preoperative AHF in older people with hip fractures using LR and five machine learning methods, among which GBM exhibited the best performance in terms of AUC, accuracy, sensitivity, and F1 score. Additionally, the application of SHAP analysis has enhanced the interpretability of the model, providing clinicians with an effective assessment method, significantly improving the scientific accuracy and precision of preoperative evaluation and decision-making by clinicians. Our research not only offers a new methodological perspective but also brings new thoughts and exploration directions to the fields of heart failure and orthopedic research, demonstrating the significant role of the big data era in advancing medical science development.

Acknowledgements

We are grateful to all those who took part in or assisted with this study project.

Author contributions

QLY conceived of the study and drafted the manuscript. MMF gathered and processed the data. ZQW and ZYH supervision, and revised the manuscript. All authors contributed to the article and approved the submitted version.

Funding

None.

Data availability

The datasets utilized in the present study are contained within the internal network of the Third Hospital of Hebei Medical University. Due to existing data privacy policies, these datasets are not publicly accessible. However, they can be made available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

The ethical review board of the Third Hospital of Hebei Medical University evaluated and sanctioned this research protocol, ensuring adherence to the Helsinki Declaration. The approval was granted under the reference number 2021–087–1. Due to the retrospective nature of data gathering in this study, the board also provided a waiver for informed consent. Prior to analysis, all patient data were anonymized to protect privacy.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Department of Geriatric Orthopedics, Third Hospital of Hebei Medical University, Shijiazhuang, Hebei 050051, China

²Department of Cardiology, The First Hospital of Qinhuangdao, Qinhuangdao, Hebei 066000, China

³Department of Orthopaedic Surgery, Third Hospital of Hebei Medical University, Shijiazhuang, Hebei 050051, China

⁴Third Hospital of Hebei Medical University, Shijiazhuang, Hebei 050051, China

Received: 16 April 2024 / Accepted: 9 April 2025

Published online: 23 April 2025

References

1. Gullberg B, Johnell O, Kanis JA. World-wide projections for hip fracture. *Osteoporos INT*. 1997;7(5):407–13.
2. Tewari P, Sweeney JBF, Lemos JL, Shapiro L, Gardner MJ, Morris AM, Baker LC, Harris AS, Kamal RN. Evaluation of systemwide improvement programs to optimize time to surgery for patients with hip fractures: A systematic review. *Jama Netw Open*. 2022;5(9):e2231911.
3. Boddaert J, Raux M, Khiami F, Riou B. Perioperative management of elderly patients with hip fracture. *Anesthesiology*. 2014;121(6):1336–41.
4. Duceppe E, Patel A, Chan M, Berwanger O, Ackland G, Kavsak PA, Rodseth R, Biccard B, Chow CK, Borges FK, et al. Preoperative N-Terminal Pro-B-Type natriuretic peptide and cardiovascular events after noncardiac surgery: A cohort study. *Ann Intern Med*. 2020;172(2):96–104.
5. Shameer K, Johnson KW, Glicksberg BS, Dudley JT, Sengupta PP. Machine learning in cardiovascular medicine: are we there yet? *Heart*. 2018;104(14):1156–1164.
6. Taleb I, Kyriakopoulos CP, Fong R, Ijaz N, Demertzis Z, Sideris K, Wever-Pinzon O, Koliopoulou AG, Bonios MJ, Shad R, et al. Machine learning multicenter risk model to predict right ventricular failure after mechanical circulatory support: the Stop-RVF score. *Jama Cardiol*. 2024;9(3):272–82.
7. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Bohm M, Burri H, Butler J, Celutkiene J, Chioncel O, et al. 2021 ESC guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J*. 2021;42(36):3599–726.
8. Ibrahim NE, Januzzi JL. Established and emerging roles of biomarkers in heart failure. *CIRC Res*. 2018;123(5):614–29.
9. Haring M, Grotli EI, Riemer-Sorensen S, Seel K, Hanssen KG. A Levenberg-Marquardt algorithm for sparse identification of dynamical systems. *IEEE T Neur Net Lear*. 2023;34(11):9323–36.
10. Mahesh TR, Dhillip Kumar V, Vinoth Kumar V, Asghar J, Geman O, Arulkumaran G, Arun N, Muhammad A, Ahmad M. AdaBoost Ensemble Methods Using K-Fold Cross Validation for Survivability with the Early Detection of Heart Disease. *Comput Intel Neurosc*. 2022;2022:9005211–9005278.

11. Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, Roobol MJ, Steyerberg EW. Reporting and interpreting decision curve analysis: A guide for investigators. *EUR UROL*. 2018;74(6):796–804.
12. Alexiou KI, Roushias A, Varitimidis SE, Malizos KN. Quality of life and psychological consequences in elderly patients after a hip fracture: a review. *Clin Interv Aging*. 2018;13:143–50.
13. Carbone L, Buzkova P, Fink HA, Lee JS, Chen Z, Ahmed A, Parashar S, Robbins JR. Hip fractures and heart failure: findings from the cardiovascular health study. *Eur Heart J*. 2010;31(1):77–84.
14. Kamijikkoku S, Y Yoshimura 2023 Concurrent negative impact of undernutrition and heart failure on functional and cognitive recovery in hip fracture patients. *Nutrients* 15 22 4800.
15. Fu M, Zhang Y, Guo J, Zhao Y, Hou Z, Wang Z, Zhang Y. Application of integrated management bundle incorporating with multidisciplinary measures improved in-hospital outcomes and early survival in geriatric hip fracture patients with perioperative heart failure: a retrospective cohort study. *Aging Clin Exp Res*. 2022;34(5):1149–58.
16. You F, Ma C, Sun F, Liu L, Zhong X. The risk factors of heart failure in elderly patients with hip fracture: what should we care. *Bmc Musculoskel Dis*. 2021;22(1):832.
17. Mohammad MA. Advancing heart failure research using machine learning. *Lancet Digit Health*. 2023;5(6):e331–2.
18. Banerjee A, Dashtban A, Chen S, Pasea L, Thygesen JH, Fatemifar G, Tyl B, Dyszynski T, Asselbergs FW, Lund LH, et al. Identifying subtypes of heart failure from three electronic health record sources with machine learning: an external, prognostic, and genetic validation study. *Lancet Digit Health*. 2023;5(6):e370–9.
19. Tian M, Li W, Wang Y, Tian Y, Zhang K, Li X, Zhu Y. Risk factors for perioperative acute heart failure in older hip fracture patients and establishment of a nomogram predictive model. *J Orthop Surg Res*. 2023;18(1).
20. Wang K, Tian J, Zheng C, Yang H, Ren J, Liu Y, Han Q, Zhang Y. 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.
21. Sun Z, Dong W, Shi H, Ma H, Cheng L, Huang Z. Comparing machine learning models and statistical models for predicting heart failure events: A systematic review and Meta-Analysis. *Front Cardiovasc Med*. 2022;9:812276.
22. Kochar A, Doll JA, Liang L, Curran J, Peterson ED. Temporal trends in post myocardial infarction heart failure and outcomes among older adults. *J Card Fail*. 2022;28(4):531–9.
23. McMurray J. Beta-blockers, ventricular arrhythmias, and sudden death in heart failure: not as simple as it seems. *Eur Heart J*. 2000;21(15):1214–5.
24. Drozd M, Garland E, Walker AMN, Slater TA, Koshy A, Straw S, Gierula J, Paton M, Lowry J, Sapsford R et al. Infection-Related hospitalization in heart failure with reduced ejection fraction. *Circulation: Heart Fail*. 2020;13(5).
25. Anand IS. Anemia and chronic heart failure. *J Am Coll Cardiol*. 2008;52(7):501–11.
26. Glasscock RJ, Pecoits-Filho R, Barberato SH. Left ventricular mass in chronic kidney disease and ESRD. *Clin J Am Soc Nephro*. 2009;4(Supplement 1):S79–91.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.