BMC Medical Informatics and Decision Making

RESEARCH

Open Access

Predicting in-hospital mortality in patients with heart failure combined with atrial fibrillation using stacking ensemble model: an analysis of the medical information mart for intensive care IV (MIMIC-IV)

Panpan Chen[1], Junhua Sun[1], Yingjie Chu[2*] and Yujie Zhao[1*]

Abstract

Background Heart failure (HF) and atrial fibrillation (AF) usually coexist and are associated with a poorer prognosis. This study aimed to develop a model to predict in-hospital mortality in patients with HF combined with AF.

Methods Patients with HF and AF were obtained from the Medical Information Mart for Intensive Care IV (MIMIC-IV) database from 2008 to 2019. Feature selection was based on the Mann-Whitney U test and the least absolute shrinkage and selection operator (LASSO) regression model. Random Forest, eXtreme Gradient Boosting (XGBoost), Light Gradient Boosting Machine (LGBM), K-Nearest Neighbor (KNN) models, and their stacked model (the stacking ensemble model) were established. The area under of the curve (AUC) with 95% confidence interval (CI), sensitivity, specificity, as well as accuracy were applied to assess the performance of the predictive models.

Results A total of 5,998 patients with HF combined with AF were included, of which 4,198 patients were assigned to the training set and 1,800 to the testing set (7:3). Among these 4,198 patients, 624 (14.86%) died in-hospital and 3,574 (85.14%) survived. Twenty-two features were used to construct the predictive model. Among these four single models, the AUC was 0.747 (95%CI: 0.717–0.777) for the Random Forest model, 0.755 (95%CI: 0.725–0.785) for the XGBoost model, 0.754 (95%CI: 0.724–0.784) for the LGBM model, and 0.746 (95%CI: 0.716–0.776) for the KNN model in the testing set. The stacking ensemble model had the highest AUC compared to the four single models, with AUCs of 0.837 (95%CI: 0.821–0.852) and 0.768 (95%CI: 0.740–0.796) for the training set and testing set, respectively.

Conclusion The stacking ensemble model showed a good predictive effect in predicting in-hospital mortality in patients with HF combined with AF and may provide clinicians with a reference tool for early identification of mortality risk.

Keywords Heart failure, Atrial fibrillation, Mortality, Prediction, Stacking ensemble model

*Correspondence:
Yingjie Chu
yingjiechu_cardio@outlook.com
Yujie Zhao
yujiezhao_cardio@outlook.com

[1]Department of Cardiovascular Medicine, Zheng Zhou Cardiovascular Hospital, The 7th People's Hospital of Zheng Zhou, No. 17, Jingnan Fifth Road, Huizhuang Development Zone, Zhengzhou, Henan 450000, China
[2]Department of Cardiovascular Medicine, Henan Provincial People's Hospital, No. 7, Weiwu Road, Jinshui District, Zhengzhou, Henan 450000, China

Background

Heart failure (HF) is a cardiovascular disease caused by abnormalities in the structure or function of the heart that can lead to increased intracardiac pressure or decreased cardiac output [1]. Atrial fibrillation (AF) is the most common arrhythmia and frequently coexists in patients with HF [2]. AF occurs in approximately one-third of patients with HF and these patients have higher morbidity and mortality than either HF or AF alone [3]. Therefore, early identification of the risk of mortality in patients with HF combined with AF is important for disease management and burden reduction.

Many tools for risk or prognosis prediction in patients with HF have been reported, including biomarkers, risk scores, and their combined metrics, but most of them have limited predictive validity [4–7]. A meta-analysis demonstrated that the predictive ability of the existing model was mediocre (c-index < 0.71) and was not applicable to the general population (e.g., only to those who were able to calculate a risk score) [8]. Adler et al. indicated that machine learning algorithms can be used to capture features associated with mortality in patients with HF to construct models that can improve the predictive effectiveness of existing models [9]. Different machine learning algorithms have been applied in the diagnosis and prognosis prediction of both HF and AF, but model effectiveness varies depending on the modeling approach [10–12]. Since each machine learning algorithm may excel or have drawbacks in different situations, models integrating multiple machine learning methods are applied. Stacking is a powerful integration technique that utilizes the predictions of multiple base learners as features to train new meta-learners, often exhibiting better performance than any single model [13]. Recently, Chiu et al. constructed a stacking ensemble model for predicting mortality in HF patients based on six base classifiers, and their model demonstrated good prediction results [14]. However, the effectiveness of the stacking ensemble models for predicting in-hospital mortality in patients with HF combined with AF is unclear. Thus, the purpose of this study was to construct a stacking ensemble model for predicting the risk of mortality in patients with HF combined with AF for use in assisting the clinical management of patients.

Methods

Study design and data source

This study utilized a retrospective cohort study design to develop models for predicting mortality in patients with HF combined with AF. Data were obtained from the Medical Information Mart for Intensive Care IV (MIMIC-IV) database between 2008 and 2019 period. MIMIC-IV is a large, single-center database containing real hospitalization data for patients admitted to the ICU at Beth Israel Deaconess Medical Center between 2008 and 2019 (https://mimic.mit.edu/docs/iv/). MIMIC-IV contains comprehensive information for each patient, including demographics, vital signs, laboratory measurements, medications, clinical measurements, and medical history [15]. Patients diagnosed with HF combined with AF at first admission to the ICU were included. Patients younger than 18 years of age, admitted to the ICU for less than 24 h, or missing survival data were excluded. HF and AF were identified according to the International Classification of Diseases, ninth/ten revision (ICD-9, ICD-10) codes. HF includes acute HF (ICD-9: 42821, 42823, 42831, 42833, 42841, 42843; ICD-10: I5021, I5023, I5031, I5033, I5041, I5043, I50811, I50813) and chronic HF (ICD-9: 42822, 42832, 42842; ICD-10: I5022, I5032, I5042, I50812). The ICD codes for AF are 42,731 for ICD-9 and I480-, I481-, I482-, I4891- for ICD-10. MIMIC-IV was approved by the Institutional Review Boards of Beth Israel Deaconess Medical Center and the Massachusetts Institute of Technology. Informed consent was not required because all protected health in the database was de-identified and did not influence clinical care [16]. All methods were performed in accordance with the relevant guidelines and regulations.

Outcome and data collection

The outcome was the in-hospital mortality in patients with HF combined with AF. Follow-up was conducted from admission to hospital until discharge or death. For patients with multiple admissions records, only data from the patient's first admission were used. Data collection included age, gender (female, male), race (Black, White, others, unknown), ICU type [cardiac care unit (CCU), medical ICU (MICU), surgical ICU (SICU), others], HF type (acute, chronic, unspecified), weight, heart rate, systolic blood pressure (SBP), diastolic blood pressure (DBP), respiratory rate, temperature, saturation of peripheral oxygen (SPO_2), Charlson comorbidity index, Simplified Acute Physiology Score (SAPS) II, ventilation (no, yes), vasopressor (no, yes), antiarrhythmic (no, yes), antiplatelet (no, yes), anticoagulation (no, yes), Beta 1 receptor agonist (no, yes), coronary artery bypass grafting (CABG) (no, yes), catheter ablation (no, yes), white blood cell (WBC), platelet, hemoglobin, red blood cell distribution width (RDW), creatinine, blood urea nitrogen (BUN), glucose, sodium, potassium, chloride, bicarbonate, estimated glomerular filtration rate (eGFR), anion gap, and in-hospital follow-up time.

Predictive model construction and evaluation

All data were randomly divided into the training set and the testing set in a ratio of 7:3. The data from the training set was used for the construction of the model (feature selection, model training), and the data from the testing

set was used for the internal validation of the model. Due to the imbalance in the incidence of outcomes in the dataset, the synthetic minority oversampling technique (SMOTE) was used to address the data imbalance during model training (training set). The SMOTE method balances the data set by oversampling to increase the number of minority class samples [17].

Feature selection was first performed using the Mann-Whitney U test, which was utilized to compare differences in characteristics between survivors and non-survivors. Among the initial 33 features, 8 features were excluded with a $P>0.05$. Then, the remaining 25 features were screened using the least absolute shrinkage and selection operator (LASSO) regression model. LASSO is a compression estimation that compresses the regression coefficients of some features by constructing a penalty function. To ensure the stability and efficacy of the features, the feature set with a higher value is selected from the 10-fold cross-validation results, that is, the features whose coefficient is not 0 are retained. Finally, 22 of these 25 features were retained and included in the predictive model.

Four single models including the Random Forest model, eXtreme Gradient Boosting (XGBoost) model, Light Gradient Boosting Machine (LGBM), and K-Nearest Neighbor (KNN) model were constructed. In addition, a stacking ensemble model consisting of these four single models was established. Random Forest is an extension of Bagging integrated learning that uses decision trees as the basic classifiers. Random Forest generates many classifiers and combines their results by majority voting. XGBoost is an efficient gradient boosting decision tree algorithm, which integrates multiple weak learners into a strong learner by certain methods, i.e., the results of all the classifiers are accumulated to get the result. LGBM is a machine learning algorithm based on gradient boosting decision trees, which progressively improves the performance of a model by iteratively training multiple decision trees. KNN is one of the most basic and simple algorithms in the machine learning algorithm model, which can be used for both classification and regression by measuring the distance between different feature values. The stacking ensemble model of this study was performed using the categorical boosting technique. Stacking is an integration method that connects several different types of classification models through meta-classifiers, by combining several weak learners to obtain a model with stronger generalization ability. In the stacking ensemble model of this study, Random Forest, XGBoost, LGBM, and KNN were used as the base classifiers in the first stage, and the outputs obtained from each single model in the first stage were fed into the meta-classifiers in the second stage. Then, the meta-classifiers were fitted to the output meta-features of each classification model by the categorical boosting integration technique.

The performances of the four single models and the stacking ensemble model were assessed by the area under of the receiver operating characteristic curve (AUC) with 95% confidence interval (CI), sensitivity, specificity, as well as accuracy. The value of AUC is greater than 0.75 indicating that the model has good predictive ability. The modeling process of this study was shown in Fig. 1. The optimal parameters for the different models were presented in Supplementary Table 1.

Statistical analysis

Continuous data were presented as mean±standard deviation (SD) or median and quartiles [M (Q1, Q3)], and categorical data were presented as numbers and percentages [n (%)]. Differences in continuous data were compared using the t-test or Wilcoxon signed-rank test, and differences in categorical data were compared by the Chi-square test or Fisher's exact probability test.

For missing values, variables with ≥20% missing data (e.g., lactate) were excluded, and variables with <20% missing data (e.g., respiratory rate) were interpolated using the random forest imputation. Sensitivity analysis was conducted before and after missing data processing. Descriptive statistical analyses were completed using R version 4.3.1 software (Institute for Statistics and Mathematics, Vienna, Austria). The construction and visualization of the model were performed using Python version 3.9.12 software (Python Software Foundation, Delaware, USA). A P-value<0.05 was considered statistically significant.

Results

Characteristics of patients

Between 2008 and 2019, MIMIC-IV documented a total of 7,097 patients diagnosed with HF combined with AF. After excluding 1,099 patients who were admitted to ICU for less than 24 h, 5,998 patients were included in the analysis (Supplement Fig. 1). Of these 5,998 patients, 4,198 were assigned to the training set and 1,800 to the testing set. Table 1 presents the baseline characteristics of patients in the training set. Among these 4,198 patients, 624 (14.86%) died in-hospital and 3,574 (85.14%) survived. The mean age was 74.35±11.48 years, 2,397 (57.1%) were males, and 3,112 (74.13%) were White. There were 1,040 (24.77%) patients from CCU, 1,428 (34.02%) patients from MICU, and 693 (16.51%) patients from SICU. For the type of HF, 2,205 (52.53%) patients had acute HF, 1,257 (29.94%) patients had chronic HF, and 736 (17.53%) patients had unspecified HF. The mean SAPS II score was 41.23±12.69. The median length of in-hospital follow-up time and ICU stay was 7.89 (5.17, 12.91) days and 2.98 (1.83, 5.28) days, respectively.

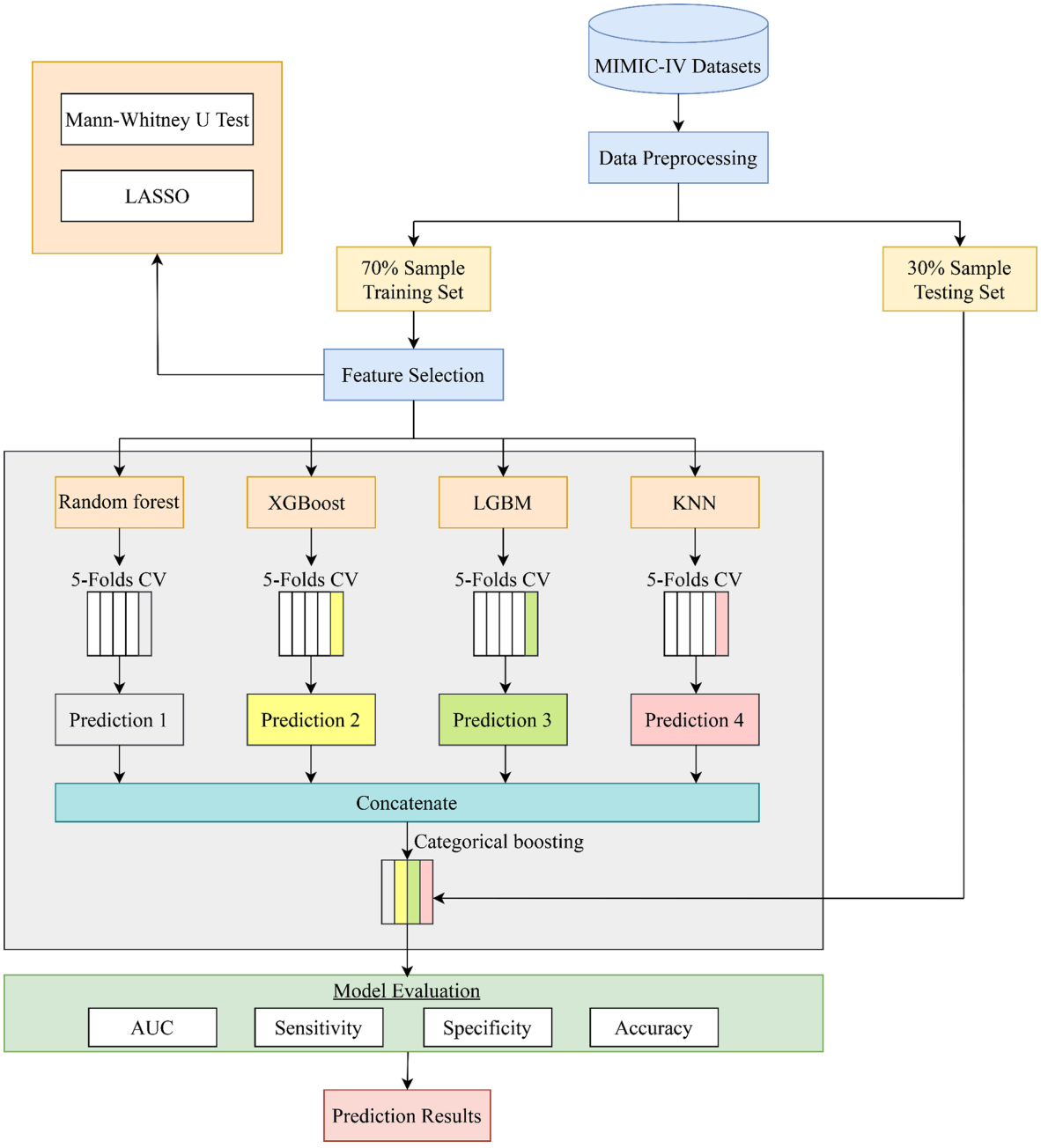

Fig. 1 The modeling process of this study. MIMIC-IV, the Medical Information Mart for Intensive Care IV database; LASSO, the least absolute shrinkage and selection operator; XGBoost, eXtreme Gradient Boosting; LGBM, Light Gradient Boosting Machine; KNN, K-Nearest Neighbor; CV, cross-validation; AUC, the area under of the curve

The results of LASSO regression on feature screening were shown in Fig. 2. The 10-fold cross-validation was used for LASSO regression, a λ-value (λ=0.00187) was determined when the mean squared error (MSE) value was the smallest (Fig. 2A) and 22 features were selected based on the λ-value (Fig. 2B). A total of 22 features were used to construct the predictive model, including age, weight, heart rate, SBP, respiratory rate, SPO_2, Charlson comorbidity index, SAPS II, RDW, BUN, glucose, eGFR, anion gap, race (White), ICU type (MICU, SICU, others), HF type (chronic), ventilation (yes), vasopressor (yes), anticoagulation (yes), and beta-1 receptor agonist (yes). The correlation heat map for these 22 features was presented in Supplement Fig. 2.

Table 1 Baseline characteristics of patients with heart failure (HF) combined with atrial fibrillation (AF) in the training set

Variables	Total (*N* = 4198)	Survivors (*N* = 3574)	Non-survivors (*N* = 624)	*P*
Age, years, Mean (± SD)	74.35 (± 11.48)	73.92 (± 11.53)	76.81 (± 10.87)	< 0.001
Gender, n (%)				0.141
Female	1801 (42.9)	1516 (42.42)	285 (45.67)	
Male	2397 (57.10)	2058 (57.58)	339 (54.33)	
Race, n (%)				< 0.001
Black	290 (6.91)	248 (6.94)	42 (6.73)	
Others	383 (9.12)	321 (8.98)	62 (9.94)	
Unknown	413 (9.84)	315 (8.81)	98 (15.71)	
White	3112 (74.13)	2690 (75.27)	422 (67.63)	
ICU type, n (%)				< 0.001
CCU	1040 (24.77)	885 (24.76)	155 (24.84)	
MICU	1428 (34.02)	1170 (32.74)	258 (41.35)	
Others	1037 (24.70)	975 (27.28)	62 (9.94)	
SICU	693 (16.51)	544 (15.22)	149 (23.88)	
HF type, n (%)				< 0.001
Acute	2205 (52.53)	1867 (52.24)	338 (54.17)	
Chronic	1257 (29.94)	1110 (31.06)	147 (23.56)	
Unspecified	736 (17.53)	597 (16.7)	139 (22.28)	
Weight, kg, Mean (± SD)	83.17 (± 23.72)	83.53 (± 23.60)	81.07 (± 24.33)	0.017
Heart rate, bpm, Mean (± SD)	88.86 (± 21.63)	88.12 (± 21.35)	93.07 (± 22.75)	< 0.001
Systolic, mmHg, Mean (± SD)	119.73 (± 23.81)	120.26 (± 23.65)	116.71 (± 24.49)	0.001
Diastolic, mmHg, Mean (± SD)	65.66 (± 18.18)	65.73 (± 17.92)	65.28 (± 19.62)	0.599
Respiratory rate, bpm, Mean (± SD)	19.59 (± 5.99)	19.32 (± 5.90)	21.12 (± 6.27)	< 0.001
Temperature, ℃, Mean (± SD)	36.56 (± 0.73)	36.57 (± 0.69)	36.48 (± 0.92)	0.027
SPO_2, %, M (Q_1, Q_3)	98 (95–100)	98 (95–100)	97 (94–100)	< 0.001
Charlson comorbidity index, score, Mean (± SD)	4.06 (± 2.14)	3.94 (± 2.09)	4.77 (± 2.30)	< 0.001
SAPSII, score, Mean (± SD)	41.23 (± 12.69)	39.56 (± 11.75)	50.82 (± 13.59)	< 0.001
Ventilation, n (%)				< 0.001
No	389 (9.27)	360 (10.07)	29 (4.65)	
Yes	3809 (90.73)	3214 (89.93)	595 (95.35)	
Vasopressor, n (%)				< 0.001
No	2100 (50.02)	1903 (53.25)	197 (31.57)	
Yes	2098 (49.98)	1671 (46.75)	427 (68.43)	
Antiarrhythmic, n (%)				< 0.001
No	1895 (45.14)	1664 (46.56)	231 (37.02)	
Yes	2303 (54.86)	1910 (53.44)	393 (62.98)	
Antiplatelet, n (%)				0.603
No	4191 (99.83)	3567 (99.80)	624 (100.00)	
Yes	7 (0.17)	7 (0.20)	0 (0)	
Anticoagulation, n (%)				< 0.001
No	898 (21.39)	805 (22.52)	93 (14.90)	
Yes	3300 (78.61)	2769 (77.48)	531 (85.10)	
Beta 1 receptor agonist, n (%)				< 0.001
No	3739 (89.07)	3272 (91.55)	467 (74.84)	
Yes	459 (10.93)	302 (8.45)	157 (25.16)	
CABG, n (%)				0.155
No	4134 (98.48)	3515 (98.35)	619 (99.20)	
Yes	64 (1.52)	59 (1.65)	5 (0.80)	
Catheter ablation, n (%)				0.842
No	4105 (97.78)	3496 (97.82)	609 (97.60)	
Yes	93 (2.22)	78 (2.18)	15 (2.40)	
WBC, K/uL, M (Q1, Q3)	10.80 (7.80–14.90)	10.60 (7.80–14.50)	12.30 (8.60-17.22)	< 0.001
Platelet, K/uL, Mean (± SD)	200.51 (± 97.29)	199.43 (± 95.14)	206.68 (± 108.67)	0.118

Table 1 (continued)

Variables	Total (*N* = 4198)	Survivors (*N* = 3574)	Non-survivors (*N* = 624)	*P*
Hemoglobin, g/dL, Mean (± SD)	10.30 (± 2.18)	10.33 (± 2.19)	10.15 (± 2.15)	0.051
RDW, %, Mean (± SD)	15.69 (± 2.28)	15.54 (± 2.19)	16.58 (± 2.59)	< 0.001
Creatinine, mg/dL, M (Q1, Q3)	1.20 (0.90–1.90)	1.20 (0.90–1.80)	1.50 (1.10–2.60)	< 0.001
BUN, mg/dL, Mean (± SD)	35.42 (± 24.93)	33.65 (± 23.75)	45.57 (± 28.80)	< 0.001
Glucose, mg/dL, M (Q1, Q3)	132 (108-168.03)	131 (107–166)	139.5 (111–180)	< 0.001
Sodium, mEq/L, Mean (± SD)	137.45 (± 5.41)	137.42 (± 5.30)	137.63 (± 6.00)	0.423
Potassium, mEq/L, Mean (± SD)	4.35 (± 0.80)	4.34 (± 0.79)	4.43 (± 0.86)	0.009
Chloride, mEq/L, Mean (± SD)	102.38 (± 6.78)	102.54 (± 6.67)	101.41 (± 7.34)	< 0.001
Bicarbonate, mEq/L, Mean (± SD)	23.74 (± 5.02)	23.94 (± 4.86)	22.55 (± 5.67)	< 0.001
eGFR, mL/min/1.73m^2, Mean (± SD)	58.14 (± 27.25)	59.94 (± 26.99)	47.84 (± 26.41)	< 0.001
Anion gap, mEq/L, Mean (± SD)	15.69 (± 5.04)	15.27 (± 4.85)	18.10 (± 5.38)	< 0.001
In-hospital follow time, days, M (Q1, Q3)	7.89 (5.17–12.91)	8.03 (5.35–12.97)	6.64 (3.21–12.31)	< 0.001
Length of ICU stay, days, M (Q1, Q3)	2.98 (1.83–5.28)	2.88 (1.76–4.97)	4.08 (2.24–8.38)	< 0.001

Note ICU, intensive care unit; CCU, cardiac care unit; MICU, medical ICU; SICU, surgical ICU; SPO_2, saturation of peripheral oxygen; SAPS II, Simplified Acute Physiology Score II; CABG, coronary artery bypass grafting; WBC, white blood cell; RDW, red blood cell distribution width; BUN, blood urea nitrogen; eGFR, estimated glomerular filtration rate

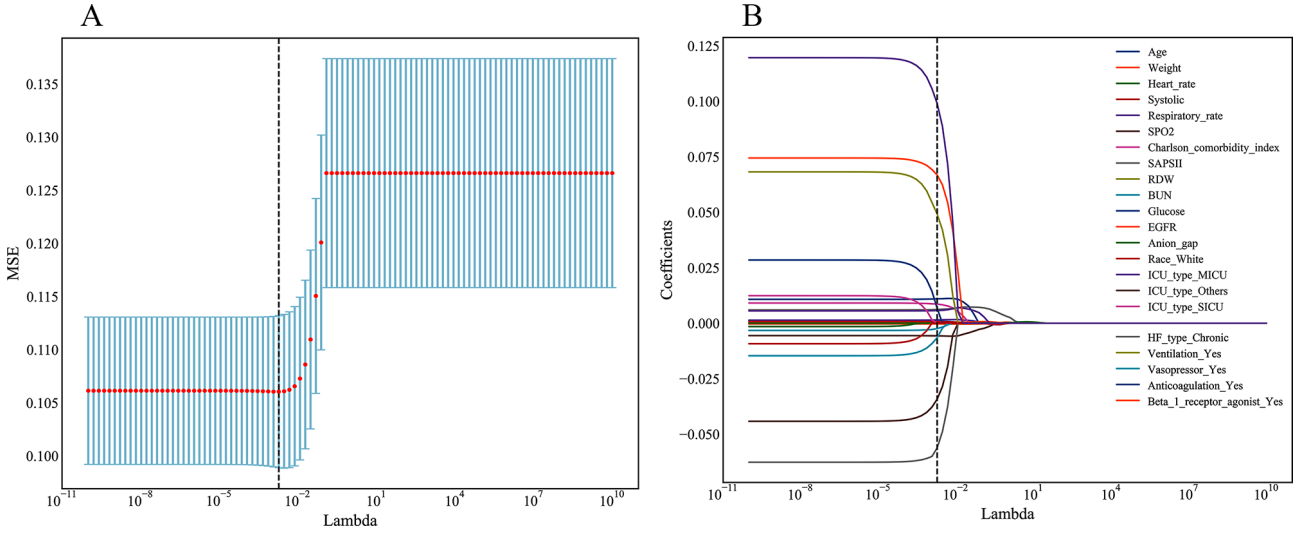

Fig. 2 Feature selection using the least absolute shrinkage and selection operator (LASSO) regression. (**A**) changes in mean squared error (MSE) during LASSO regression screening; (**B**) Changes in the coefficient profiles during LASSO regression screening

Model performance for predicting in-hospital mortality in patients with HF combined with AF

Table 2 shows the performance of the Random Forest model, XGBoost model, LGBM model, KNN model, and stacking ensemble model in predicting in-hospital mortality in patients with HF combined with AF. The single models of Random Forest, XGBoost, LGBM, and KNN performed better in predicting the in-hospital mortality, with model AUCs of 0.818 (95%CI: 0.801–0.835), 0.827 (95%CI: 0.811–0.843), 0.811 (95%CI: 0.794–0.829), and 0.824 (95%CI: 0.808–0.840) in the training set, respectively. In the testing set, the AUC was 0.747 (95%CI: 0.717–0.777) for the Random Forest model, 0.755 (95%CI: 0.725–0.785) for the XGBoost model, 0.754 (95%CI: 0.724–0.784) for the LGBM model, and 0.746 (95%CI: 0.716–0.776) for the KNN model. Moreover, the stacking ensemble model had the highest AUC compared to the four single models, with AUCs of 0.837 (95%CI: 0.821–0.852) and 0.768 (95%CI: 0.740–0.796) for the training set and testing set, respectively. The receiver operating characteristic (ROC) curves of these models were shown in Fig. 3.

Comparisons of AUC between the stacking ensemble model and the four single models were presented in Table 3. In predicting in-hospital mortality in patients with HF combined with AF, the AUC of the stacking ensemble model was superior to that of the four single models on both the training set and the testing set ($P < 0.05$).

In addition, the predictive performance of these models was analyzed in populations with different HF types (chronic, acute) (Supplement Table 2). The stacking

Table 2 Model performance in predicting in-hospital mortality in patients with HF combined with AF

Models	Dataset	Sensitivity (95%CI)	Specificity (95%CI)	AUC (95%CI)	Accuracy (95%CI)
Random forest	Training set	0.740 (0.706–0.775)	0.740 (0.725–0.754)	0.818 (0.801–0.835)	0.740 (0.727–0.753)
	Testing set	0.629 (0.572–0.685)	0.715 (0.692–0.738)	0.747 (0.717–0.777)	0.702 (0.681–0.723)
XGBoost	Training set	0.803 (0.772–0.834)	0.685 (0.670–0.700)	0.827 (0.811–0.843)	0.702 (0.689–0.716)
	Testing set	0.704 (0.650–0.757)	0.670 (0.647–0.694)	0.755 (0.725–0.785)	0.676 (0.654–0.697)
LGBM	Training set	0.713 (0.678–0.749)	0.764 (0.750–0.778)	0.811 (0.794–0.829)	0.756 (0.743–0.769)
	Testing set	0.632 (0.576–0.689)	0.757 (0.735–0.778)	0.754 (0.724–0.784)	0.737 (0.717–0.758)
KNN	Training set	0.838 (0.809–0.867)	0.648 (0.633–0.664)	0.824 (0.808–0.840)	0.677 (0.662–0.691)
	Testing set	0.739 (0.688–0.791)	0.637 (0.613–0.661)	0.746 (0.716–0.776)	0.653 (0.631–0.675)
Stacking ensemble	Training set	0.812 (0.782–0.843)	0.705 (0.690–0.720)	0.837 (0.821–0.852)	0.721 (0.708–0.735)
	Testing set	0.682 (0.628–0.737)	0.693 (0.670–0.716)	0.768 (0.740–0.796)	0.691 (0.670–0.712)

Note HF, heart failure; AF, atrial fibrillation; XGBoost, eXtreme Gradient Boosting; LGBM, Light Gradient Boosting Machine; KNN, K-Nearest Neighbor; AUC, the area under of the curve; CI, confidence interval

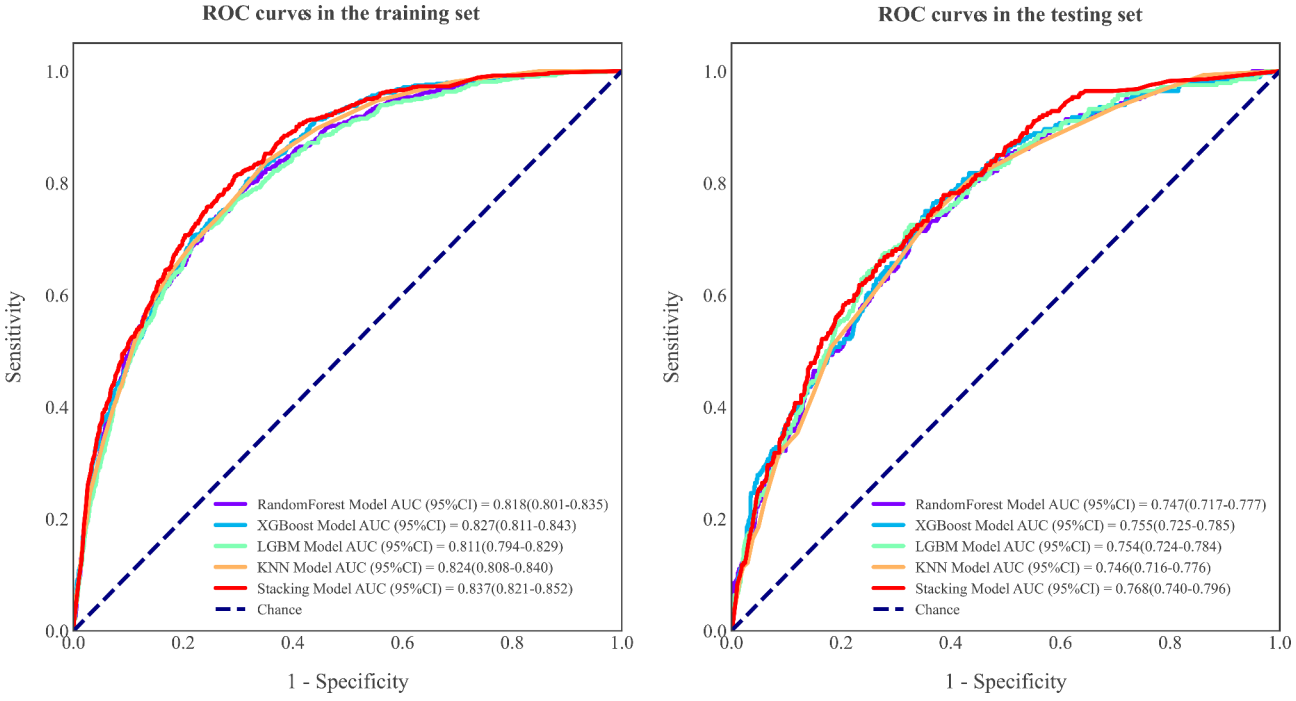

Fig. 3 The receiver operating characteristic (ROC) curves of the models to predict in-hospital mortality in patients with heart failure (HF) combined with atrial fibrillation (AF). XGBoost, eXtreme Gradient Boosting; LGBM, Light Gradient Boosting Machine; KNN, K-Nearest Neighbor; AUC, the area under of the curve; CI, confidence interval

ensemble model still showed good ability to predict in-hospital mortality in both patients with chronic HF combined with AF [Training set: (AUC=0.907, 95%CI: 0.886–0.928); Testing set: (AUC=0.800, 95%CI: 0.746–0.853)] and patients with acute HF combined with AF [Training set: (AUC=0.828, 95%CI: 0.806–0.851); Testing set: (AUC=0.743, 95%CI: 0.699–0.786)].

Discussion

Patients with HF combined with AF tend to have a poorer prognosis. This study constructed a model to predict in-hospital mortality in patients with HF combined with AF using four single models and the stacking ensemble model, respectively. Among the four single models, the LGBM model and the XGBoost model had good predictive ability for in-hospital mortality, with model AUCs of 0.754 and 0.755 in the testing set, respectively. The AUC of the stacking ensemble model was superior to that of the four single models, with AUCs of 0.837 and 0.768 for the training set and testing set, respectively.

Previous studies have reported models for predicting mortality in patients with HF [12, 18]. Li et al. used machine learning methods to build a model for predicting mortality in patients with HF, in which the XGBoost model had the best prediction, with an AUC of 0.824 on the training set [12]. Chen et al. demonstrated that the XGBoost model used in the prediction of in-hospital mortality in patients with HF outperformed conventional

Table 3 Delong test for comparison of AUC of different models

Dataset	Models	AUC (95%CI)	Statistics	*P*
Training set	Random forest	0.818 (0.801–0.835)	6.71	< 0.001
	XGBoost	0.827 (0.811–0.843)	3.58	< 0.001
	LGBM	0.811 (0.794–0.829)	9.43	< 0.001
	KNN	0.824 (0.808–0.840)	2.99	0.003
	Stacking ensemble	0.837(0.821–0.852)	Ref	
Testing set	Random Forest	0.747 (0.717–0.777)	4.08	< 0.001
	XGBoost	0.755 (0.725–0.785)	2.52	0.012
	LGBM	0.754 (0.724–0.784)	2.86	0.004
	KNN	0.746 (0.716–0.776)	2.61	0.009
	Stacking ensemble	0.768(0.740–0.796)	Ref	

Note AUC, the area under of the curve; CI, confidence interval; XGBoost, eXtreme Gradient Boosting; LGBM, Light Gradient Boosting Machine; KNN, K-Nearest Neighbor

risk prediction methods, with an AUC of 0.771 for the model in the external validation set [18]. Segar et al. showed that machine learning models predicted HF mortality better than the traditional Get With The Guidelines-Heart Failure (GWTG-HF) model (C-statistic: 0.82 vs. 0.69) [19]. Since each machine learning model may have strengths and weaknesses in different situations, the stacking ensemble model achieves better model performance by integrating multiple machine learning models. In addition, HF and AF frequently coexist and are associated with a worse prognosis. However, models for predicting the risk of mortality in patients with HF combined with AF have not been reported. Our study used a machine learning approach to compare the performance of different models in predicting mortality in patients with HF combined with AF. All four single models (Random Forest, XGBoost, LGBM, KNN) showed good predictive ability of in-hospital mortality, and the AUC of the models in the training set exceeded 0.81. However, the predictive ability of these models on the testing set (AUC > 0.746) is slightly weaker than on the training set. This may be related to the fact that the data distributions of the training and testing sets are inconsistent (the SMOTE method was used in the training set to deal with the data imbalance problem) leading to difficulties in generalizing the models to the testing set. The stacking ensemble model consisting of these four single models showed better predictions than any of the single models. The AUC of the stacking ensemble model in the training set and testing set were 0.837 and 0.768, respectively.

In this study, the sensitivity and specificity of the model represent the model's recognition of patients at risk of mortality and patients not at risk of mortality, respectively, while the accuracy represents the overall recognition performance of the model for patients at risk of mortality and patients not at risk of mortality. Our stacking ensemble model had a sensitivity of 0.812, a specificity of 0.705, and an accuracy of 0.721. Although the specificity of the model was not high, the model had a high sensitivity of 0.812. For models with mortality as an outcome, models with high sensitivity for the prediction of patient mortality may be more clinically valuable. However, the performance of the stacking ensemble model in the testing set was similarly weaker than in the training set. Moreover, chronic HF and acute HF were combined for analysis in this study. To test whether this was reasonable, we examined the performance of the model in the acute HF and chronic HF populations separately. The results demonstrated that the stacking ensemble model showed a good ability to predict in-hospital mortality in both patients with chronic HF combined with AF (AUC = 0.907) and patients with acute HF combined with AF (AUC = 0.828). This suggests that it is feasible to combine chronic HF and acute HF for predictive models with mortality as the outcome. The stacking ensemble model may provide a reference for a real mortality risk assessment tool for clinical practice. HF and AF interact, with common risk factors (e.g., age, hypertension, obesity) and comorbidities (e.g., valvular and ischemic and cardiac disease), neurohormonal and electrophysiologic changes, as well as alterations to cardiac myocytes combining to create an environment in which the heart is susceptible to HF and AF [20, 21]. Increased ventricular rate and arrhythmias caused by AF can shorten the left ventricular filling time, resulting in decreased cardiac output, which increases left atrial pressure. The increased cardiac filling pressures in HF can lead to atrial stretching, cardiac fibrosis, dysregulation of intracellular calcium regulation, and autonomic and neuroendocrine dysfunction, all of which may cause AF [22]. Weak atrial contraction impairs ventricular filling and worsens diastolic function. Ventricular remodeling with ventricular dilatation is a response to chronically elevated blood pressure, and this remodeling can lead to worsening of AF and HF [20, 23, 24].

In our predictive model, 22 characteristics (e.g., age, RDW, BUN, blood glucose, eGFR, anion gap) were used to construct the model. The relationship between these characteristics and HF or AF has also been reported. Age was an independent predictor of all-cause mortality in patients with HF, and age was found to significantly influence the effect of body mass index on patient mortality [25]. RDW reflects the variability of circulating red blood cell size, and a high RDW was associated with morbidity and mortality in patients with HF [26, 27]. The association between RDW and HF may be related to nutritional deficiencies, renal insufficiency, hepatic congestion, and inflammatory stress [26, 27]. BUN is a marker of kidney function that measures protein metabolism in the blood. Previous studies have shown that BUN is a key predictor

of mortality in patients with HF [8, 9, 28]. Diabetes is a common risk factor for HF, and elevated blood glucose levels are an independent predictor of 30-day mortality in patients with HF [29, 30]. eGFR is an assessment of renal function, and impaired renal function is a prognostic indicator of acute and chronic HF [31]. Serum anion gap is used in the differential diagnosis of acid-base imbalance and metabolic acidosis, and high serum anion gap levels were linked to an increased risk of mortality in patients with HF [32]. However, the order of importance of these 22 features for the stacking ensemble model cannot be known. Since the results of the stacking ensemble model are based on the outputs of four single models, there are differences in the importance of these 22 features for each single model.

Our study constructed a stacking ensemble model for predicting in-hospital mortality in patients with HF combined with AF. The stacking ensemble model combines the strengths of multiple machine learning models and shows better predictive performance than a single model. As no predictive model for in-hospital mortality in HF combined with AF has been reported, our stacking ensemble model may provide a reference for a true mortality risk assessment tool in clinical practice.

Limitations

Some limitations of this study should be noted. First, this study mainly included ICU patients, and the model's prediction of mortality risk in the general population needs to be further tested. Second, some biomarkers such as Troponin-T and N-terminal pro-B-type natriuretic peptides were not considered due to too many missing values (more than 60%), which may affect the prediction effect of the model. Third, the model lacks external validation, which is necessary before the model can be applied in clinical practice. Fourth, this study was unable to analyze HF into three phenotypes, HF with preserved ejection fraction (HFpEF), HF with reduced ejection fraction (HFrEF), and HF with midrange ejection fraction (HFmEF), and thus patients with HFrEF, HFpEF, and HFmrEF phenotypes could not be analyzed separately. Fifth, this study could not determine whether some patients had isolated or combined pulmonary hypertension associated with right ventricular dysfunction, making it impossible to perform a classification analysis based on isolated or combined pulmonary hypertension. Sixth, the cause of the patient's heart failure is unknown.

Conclusions

This study constructed models for predicting in-hospital mortality in patients with HF combined with AF. The stacking ensemble model consisting of Random Forest, XGBoost, LGBM, and KNN has better AUC than any of single model. The stacking ensemble model may provide a reference for a true mortality risk assessment tool in clinical practice among patients with HF combined with AF. Moreover, external validation is necessary before the model can be applied in clinical practice.

Abbreviations

HF	Heart failure
AF	Atrial fibrillation
MIMIC-IV	Medical Information Mart for Intensive Care IV
ICD-9, ICD-10	International Classification of Diseases, ninth/ten revision
SBP	Systolic blood pressure
DBP	Diastolic blood pressure
SPO_2	Saturation of peripheral oxygen
SAPS	Simplified Acute Physiology Score
CABG	Coronary artery bypass grafting
WBC	White blood cell
RDW	Red blood cell distribution width
BUN	Blood urea nitrogen
eGFR	Estimated glomerular filtration rate
CCU	Cardiac care unit
MICU	Medical ICU
SICU	Surgical ICU
LASSO	Least absolute shrinkage and selection operator
XGBoost	eXtreme Gradient Boosting
LGBM	Light Gradient Boosting Machine
KNN	K-Nearest Neighbor
AUC	Area under of the receiver operating characteristic curve
CI	Confidence interval
SD	Standard deviation

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12911-024-02829-0.

Supplementary Material 1

Acknowledgements

Not applicable.

Author contributions

PC, YC, and YZ designed the study. PC wrote the manuscript. JS and YC collected, analyzed, and interpreted the data. YC and YZ critically reviewed, edited, and approved the manuscript. All authors read and approved the final manuscript.

Funding

Not applicable.

Data availability

The datasets generated during and/or analyzed during the current study are available in the MIMIC-IV database, https://mimic.mit.edu/docs/iv/.

Declarations

Ethics approval and consent to participate

MIMIC-IV was approved by the Institutional Review Boards of Beth Israel Deaconess Medical Center and the Massachusetts Institute of Technology. Informed consent was not required because all protected health in the database was de-identified and did not influence clinical care. All methods were performed in accordance with the relevant guidelines and regulations.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 1 February 2024 / Accepted: 17 December 2024
Published online: 23 December 2024

References

1. Metra M, Teerlink JR. Heart failure. Lancet (London England). 2017;390:1981–95.
2. Sartipy U, Dahlström U, Fu M, Lund LH. Atrial fibrillation in heart failure with Preserved, mid-range, and reduced ejection fraction. JACC Heart Fail. 2017;5:565–74.
3. Brown LAE, Boos CJ. Atrial fibrillation and heart failure: factors influencing the choice of oral anticoagulant. Int J Cardiol. 2017;227:863–8.
4. McKie PM, Cataliotti A, Lahr BD, Martin FL, Redfield MM, Bailey KR, et al. The prognostic value of N-terminal pro-B-type natriuretic peptide for death and cardiovascular events in healthy normal and stage A/B heart failure subjects. J Am Coll Cardiol. 2010;55:2140–7.
5. Sartipy U, Dahlström U, Edner M, Lund LH. Predicting survival in heart failure: validation of the MAGGIC heart failure risk score in 51,043 patients from the Swedish heart failure registry. Eur J Heart Fail. 2014;16:173–9.
6. Sawano M, Shiraishi Y, Kohsaka S, Nagai T, Goda A, Mizuno A, et al. Performance of the MAGGIC heart failure risk score and its modification with the addition of discharge natriuretic peptides. ESC Heart Fail. 2018;5:610–9.
7. Lanfear DE, Levy WC, Stehlik J, Estep JD, Rogers JG, Shah KB et al. Accuracy of Seattle Heart failure model and HeartMate II risk score in Non-inotrope-dependent Advanced Heart failure patients: insights from the ROADMAP Study (Risk Assessment and comparative effectiveness of left ventricular assist device and Medical Management in Ambulatory Heart failure patients). Circulation Heart Fail. 2017; 10.
8. Ouwerkerk W, Voors AA, Zwinderman AH. Factors influencing the predictive power of models for predicting mortality and/or heart failure hospitalization in patients with heart failure. JACC Heart Fail. 2014;2:429–36.
9. 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. 2020;22:139–47.
10. Fuadah YN, Lim KM. Optimal classification of Atrial Fibrillation and Congestive Heart failure using machine learning. Front Physiol. 2021;12:761013.
11. Falsetti L, Rucco M, Proietti M, Viticchi G, Zaccone V, Scarponi M, et al. Risk prediction of clinical adverse outcomes with machine learning in a cohort of critically ill patients with atrial fibrillation. Sci Rep. 2021;11:18925.
12. Li J, Liu S, Hu Y, Zhu L, Mao Y, Liu J. Predicting Mortality in Intensive Care Unit patients with heart failure using an interpretable machine learning model: Retrospective Cohort Study. J Med Internet Res. 2022;24:e38082.
13. Naimi AI, Balzer LB. Stacked generalization: an introduction to super learning. Eur J Epidemiol. 2018;33:459–64.
14. Chiu CC, Wu CM, Chien TN, Kao LJ, Li C, Jiang HL. Applying an Improved Stacking Ensemble Model to predict the mortality of ICU patients with heart failure. J Clin Med. 2022; 11.
15. Johnson A, Bulgarelli L, Pollard T, Horng S, Celi LA, Mark R. (2021). MIMIC-IV (version 1.0). PhysioNet. 2021. https://doi.org/10.13026/s6n6-xd98. Accessed October 22, 2023.
16. Johnson AE, Pollard TJ, Shen L, Lehman LW, Feng M, Ghassemi M, et al. MIMIC-III, a freely accessible critical care database. Sci data. 2016;3:160035.
17. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer W. P. SMOTE: synthetic minority over-sampling technique. J Artif Intell Res. 2002;16:321–57.
18. Chen Z, Li T, Guo S, Zeng D, Wang K. Machine learning-based in-hospital mortality risk prediction tool for intensive care unit patients with heart failure. Front Cardiovasc Med. 2023;10:1119699.
19. Segar MW, Hall JL, Jhund PS, Powell-Wiley TM, Morris AA, Kao D, et al. Machine Learning-Based Models Incorporating Social Determinants of Health vs Traditional models for Predicting In-Hospital mortality in patients with heart failure. JAMA Cardiol. 2022;7:844–54.
20. Gopinathannair R, Chen LY, Chung MK, Cornwell WK, Furie KL, Lakkireddy DR, et al. Managing Atrial Fibrillation in patients with heart failure and reduced ejection Fraction: A Scientific Statement from the American Heart Association. Circulation Arrhythmia Electrophysiol. 2021;14:Hae0000000000000078.
21. Anter E, Jessup M, Callans DJ. Atrial fibrillation and heart failure: treatment considerations for a dual epidemic. Circulation. 2009;119:2516–25.
22. Kloosterman M, Santema BT, Roselli C, Nelson CP, Koekemoer A, Romaine SPR, et al. Genetic risk and atrial fibrillation in patients with heart failure. Eur J Heart Fail. 2020;22:519–27.
23. Batul SA, Gopinathannair R. Atrial Fibrillation in Heart failure: a therapeutic challenge of our Times. Korean Circulation J. 2017;47:644–62.
24. Ling LH, Kistler PM, Kalman JM, Schilling RJ, Hunter RJ. Comorbidity of atrial fibrillation and heart failure. Nat Reviews Cardiol. 2016;13:131–47.
25. Regan JA, Kitzman DW, Leifer ES, Kraus WE, Fleg JL, Forman DE, et al. Impact of Age on comorbidities and outcomes in Heart failure with reduced ejection fraction. JACC Heart Fail. 2019;7:1056–65.
26. Felker GM, Allen LA, Pocock SJ, Shaw LK, McMurray JJ, Pfeffer MA, et al. Red cell distribution width as a novel prognostic marker in heart failure: data from the CHARM program and the Duke Databank. J Am Coll Cardiol. 2007;50:40–7.
27. Sotiropoulos K, Yerly P, Monney P, Garnier A, Regamey J, Hugli O, et al. Red cell distribution width and mortality in acute heart failure patients with preserved and reduced ejection fraction. ESC Heart Fail. 2016;3:198–204.
28. 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. 2020;8:12–21.
29. Khan H, Kunutsor SK, Kauhanen J, Kurl S, Gorodeski EZ, Adler AI, et al. Fasting plasma glucose and incident heart failure risk: a population-based cohort study and new meta-analysis. J Card Fail. 2014;20:584–92.
30. Mebazaa A, Gayat E, Lassus J, Meas T, Mueller C, Maggioni A, et al. Association between elevated blood glucose and outcome in acute heart failure: results from an international observational cohort. J Am Coll Cardiol. 2013;61:820–9.
31. Biegus J, Zymliński R, Testani J, Marciniak D, Zdanowicz A, Jankowska EA, et al. Renal profiling based on estimated glomerular filtration rate and spot urine sodium identifies high-risk acute heart failure patients. Eur J Heart Fail. 2021;23:729–39.
32. Xu H, Xia J, Wang A, Zong L, An X, Sun X. Serum anion gap is associated with mortality in intensive care unit patients with diastolic heart failure. Sci Rep. 2023;13:16670.

Publisher's note