Journal of the American Heart Association

ORIGINAL RESEARCH

Machine Learning for Mortality Prediction in Patients With Heart Failure With Mildly Reduced Ejection Fraction

Pengchao Tian, MB; Lin Liang, MB; Xuemei Zhao, MD; Boping Huang, MB; Jiayu Feng, MB; Liyan Huang, MB; Yan Huang, MD; Mei Zhai, MD; Qiong Zhou, MS; Jian Zhang, MD, PhD; Yuhui Zhang, MD, PhD

BACKGROUND: Machine-learning-based prediction models (MLBPMs) have shown satisfactory performance in predicting clinical outcomes in patients with heart failure with reduced and preserved ejection fraction. However, their usefulness has yet to be fully elucidated in patients with heart failure with mildly reduced ejection fraction. This pilot study aims to evaluate the prediction performance of MLBPMs in a heart failure with mildly reduced ejection fraction cohort with long-term follow-up data.

METHODS AND RESULTS: A total of 424 patients with heart failure with mildly reduced ejection fraction were enrolled in our study. The primary outcome was all-cause mortality. Two feature selection strategies were introduced for MLBPM development. The "All-in" (67 features) strategy was based on feature correlation, multicollinearity, and clinical significance. The other strategy was the CoxBoost algorithm with 10-fold cross-validation (17 features), which was based on the selection result of the "All-in" strategy. Six MLBPMs with 5-fold cross-validation based on the "All-in" and the CoxBoost algorithm with 10-fold cross-validation strategy were developed by the eXtreme Gradient Boosting, random forest, and support vector machine algorithms. The logistic regression model with 14 benchmark predictors was used as a reference model. During a median follow-up of 1008 (750, 1937) days, 121 patients met the primary outcome. Overall, MLBPMs outperformed the logistic model. The "All-in" eXtreme Gradient Boosting model had the best performance, with an accuracy of 85.4% and a precision of 70.3%. The area under the receiver-operating characteristic curve was 0.916 (95% CI, 0.887–0.945). The Brier score was 0.12.

CONCLUSIONS: The MLBPMs could significantly improve outcome prediction in patients with heart failure with mildly reduced ejection fraction, which would further optimize the management of these patients.

Key Words: HFmrEF ■ machine learning ■ mortality ■ prediction

Heart failure (HF) has become a major concern in public health worldwide. Accurate risk stratification is one of the key components in managing patients with HF. Several risk prediction models have been developed, such as the Seattle Heart Failure Model,[1] the Barcelona Bio-Heart Failure risk calculator,[2] and the Meta-Analysis Global Group in Chronic Heart Failure risk score.[3] However, Pau et al. reported that these 3 contemporary prediction models were less accurate and either underestimated or overestimated the risk of patient mortality.[4]

With the developments in the field of computer science, machine learning (ML) and artificial intelligence (AI) have demonstrated their unique strength in the prevention, diagnosis, and outcome prediction of cardiovascular diseases.[5] In HF, ML has demonstrated its

Correspondence to: Jian Zhang, MD, PhD, and Yuhui Zhang, MD, PhD, Heart Failure Center, State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center for Cardiovascular Disease, Chinese Academy of Medical Science, Peking Union Medical College, Key Laboratory of Clinical Research for Cardiovascular Medications, National Health Committee, No. 167 Fuwai Hospital, Beilishi Rd. Xicheng District, 10037 Beijing, China.
Email: fwzhangjian62@126.com, yuhuizhangjoy@163.com

This manuscript was sent to Julie K. Freed, MD, PhD, Associate Editor, for review by expert referees, editorial decision, and final disposition.

Supplemental Material is available at https://www.ahajournals.org/doi/suppl/10.1161/JAHA.122.029124

For Sources of Funding and Disclosures, see page 14.

JAHA is available at: www.ahajournals.org/journal/jaha

CLINICAL PERSPECTIVE

What Is New?

- Heart failure with mildly reduced ejection fraction, accounting for nearly 25% of patients with heart failure, has become the new focus.
- However, evidence on the effectiveness of machine learning in heart failure with mildly reduced ejection fraction outcome prediction is still being determined.
- Based on the electronic clinical data, the machine-learning–based prediction model showed good predictive power and could further refine risk stratification in heart failure with mildly reduced ejection fraction.

What Are the Clinical Implications?

- By integrating the optimization of feature selection strategy and machine learning algorithms, a better balance of the prediction performance and clinical utility could be reached, further refining the management of patients with heart failure with mildly reduced ejection fraction and improving their quality of life.
- In addition, newer phenotype and prognostic markers can also be found and used in clinical practice.

Nonstandard Abbreviations and Acronyms

ADHERE	Acute Decompensated Heart Failure National Registry
ATTEND	Acute Decompensated heart failure syndromes
AUPRC	area under the precision-recall curve
CBCV	CoxBoost with 10-fold cross-validation
EHFS II	Euro Heart Failure Survey II
HFmrEF	heart failure with mildly reduced ejection fraction
KorAHF	Korean Acute Heart Failure
ML	machine learning
MLBPM	machine-learning-based prediction model
RF	random forest
SVM	support vector machine
XGBoost	eXtreme Gradient Boosting

superiority in mortality prediction in comparison to the benchmark models.[6–9] Nevertheless, most contemporary studies focused on the whole heart failure population or mainly on the reduced ejection or preserved ejection subgroups.[7,10–15] In an explanatory analysis of the Machine learning Assessment of Risk and EaRly mortality in Heart Failure (MARKER-HF) study,[16] an ML-based score comprising 8 variables demonstrated its value in predicting clinical outcomes in a subgroup of patients with HF with mildly reduced ejection fraction (HFmrEF).[17] However, the MARKER-HF score was not mainly derived from patients with HFmrEF. Critical characteristics such as NT-proBNP (N-terminal pro-brain natriuretic peptide), serum sodium, and systolic blood pressure (SBP) were not included. Therefore, it may not fully reflect the risk characteristics of patients with HFmrEF. More importantly, recent studies mainly reported the performance of ML models in predicting short-term outcomes, such as during hospitalization or 30 days after discharge. Its predictive power in long-term outcomes remains elusive. To fill that gap, our study aimed to evaluate the predictive value and clinical usefulness of MLBPMs in patients with HFmrEF during a long-term follow-up.

METHODS

Data, Materials, and Code Disclosure Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request. The data are not publicly available due to privacy or ethical restrictions.

Patients and Outcomes

In this pilot study, 467 patients who met the diagnostic criteria of HFmrEF were extracted from a consecutively enrolled prospective HF cohort of our Heart Failure Care Unit, Fuwai Hospital, Chinese Academy of Medical Science and Peking Union Medical College (CAMS&PUMC), from December 2006 to October 2017. The diagnosis of HFmrEF was made, confirmed, or revised by 2 cardiologists according to the 2016 European Society of Cardiology (ESC) heart failure guideline.[18] A generally used definition of HFmrEF was used in this study, which was defined as patients with HF with a left ventricular ejection fraction (LVEF) of 40% to 49%.[19] Although the LVEF was not fully concomitant with what the 2021 ESC heart failure guideline had defined,[20] a particular subanalysis on patients with HFmrEF with LVEF 41% to 49% was independently conducted.

Patients were excluded if they were concurrent with 1 or more of the following situations: (1) malignancy; (2) severe infection or sepsis; (3) acute pulmonary embolism; (4) acute myocarditis; (5) cardiac amyloidosis; (6) autoimmune diseases; (7) cor pulmonale; or (8) aortic dissection.

A total of 43 patients met the exclusion criteria. The final analytical cohort included 424 patients with HFmrEF (Figure S1). This study was approved by the ethics committee of Fuwai Hospital and was conducted

under the Declaration of Helsinki. All patients signed consent forms after they were admitted.

The primary outcome was all-cause mortality. Follow-up was conducted by telephone or clinical visits every 3 to 6 months after discharge.

Data Collection and Candidate Features

Demographic and clinical data were all extracted from the well-structured electronic database of our center, which was recorded by trained physicians and nurses from the hospital information system of Fuwai Hospital CAMS&PUMC. In total, 107 baseline features within the first 24 hours after Heart Failure Care Unit admission were enrolled in this study, including vital signs, such as blood pressure and heart rate, and laboratory results, such as complete blood cell count, blood biochemistry, and NT-proBNP.

Data Processing and Feature Selection

Twenty-five features with >50% missing values were removed. The remaining 82 features were collected for further processing (Tables S1 and S2). For features with missing values of <10%, median/mean imputation was used. Those with 10% to 20% missing values were imputed by the MissForest package (version 1.5) in R. The largest feature missing rate is 15.09%. The results of MissForest imputation are listed in Table S3. The comparison of features before and after data imputation is shown in Table S4. No significant discriminations were observed. Two different selection strategies were used after imputation for ML-based model development. The first was called "All-in," mainly based on feature correlation, multicollinearity, and clinical significance. The criteria for collinearity diagnostics are the coefficient between 2 variables >0.7 or the variance inflation factor >10. Features that met these criteria or without explicit clinical significance were removed directly or converted into a ratio (Figure S2 and Table S5). Sixty-seven features were retained. The second was called "CBCV," which was selected by the CoxBoost algorithm (CoxBoost package, version 1.5) with 10-fold cross-validation. It is an effective strategy to fit a Cox proportional hazards model by component-wise likelihood-based boosting and is particularly suitable for models with a large number of predictors.[21] Seventeen features were ultimately determined. Moreover, 14 benchmark risk predictors were chosen for the logistic regression model. Features involved in this study are summarized in Table S1.

Model Derivation, Validation, and Performance

Three machine-learning-based prediction models (MLBPMs) with the "All-in" feature selection strategy, 3 MLBPMs with the "CBCV" feature selection strategy, and 1 logistic regression model were developed to assess the mortality risk of patients with HFmrEF. The MLBPMs were constructed by the eXtreme Gradient Boosting (XGBoost) algorithm, the random forest (RF) algorithm, and the support vector machine (SVM) algorithm from the tidymodels package (version 1.0.0). Grid research with 5-fold cross-validation was used to search the best hyperparameters for the final MLBPMs. The principle and the best hyperparameters of the MLBPMs in this study are summarized in Table S6. A 5-fold cross-validation strategy was applied to develop the derivation and validation cohorts. All study patients were randomly divided into 5 subsets with similar event rates. Four of them (80%) were combined and treated as the training set, and the other 1 (20%) was treated as the validation set. This process was repeated 5 times, and each subset would be treated as the validation set, thereby comprehensively considering the variability of patients and estimating the mortality risk (Figure 1). The logistic regression model with the enter method, containing 14 predictors, was set as the reference model.

Multiple measurements and plots were used to evaluate the model performance. Model discrimination was assessed by receiver operating characteristic curves and the corresponding area under the curve (AUC). As the event rate was relatively low in our study population, the discrimination was also assessed by precision-recall (PR) curves and the area under the PR curve.[22] The Brier score was used to assess model calibration. Moreover, model accuracy, precision, recall, and F1-score were also introduced to measure the performance of ML-based models. The prediction distribution from each model was plotted in the order of the patient's risk. The net clinical fit was evaluated by decision curve analysis. The Shapley additive explanations (SHAP) analysis was introduced to better understand the impact of features on the prognostic performance of XGBoost models.[23] The best-performing model with the highest accuracy, AUC, area under the PR curve, and the smallest Brier score was chosen for feature importance, SHAP, and subgroup analysis. We used the TRIPOD checklist when writing our report.[24]

Statistical Analysis

Continuous variables are expressed as the mean ± SD or the median (interquartile range). Categorical variables are sorted as frequencies (percentages). The Student *t* test, the Mann–Whitney *U* test, or the χ^2 test were applied for baseline comparison, as appropriate. Correlations between each feature were assessed by the Spearman ρ coefficients and visualized by the heatmap. All analyses were conducted by SPSS 25.0 (IBM, Chicago, IL) and R version 4.2.0. A 2-sided *P* value <0.05 was considered statistically significant.

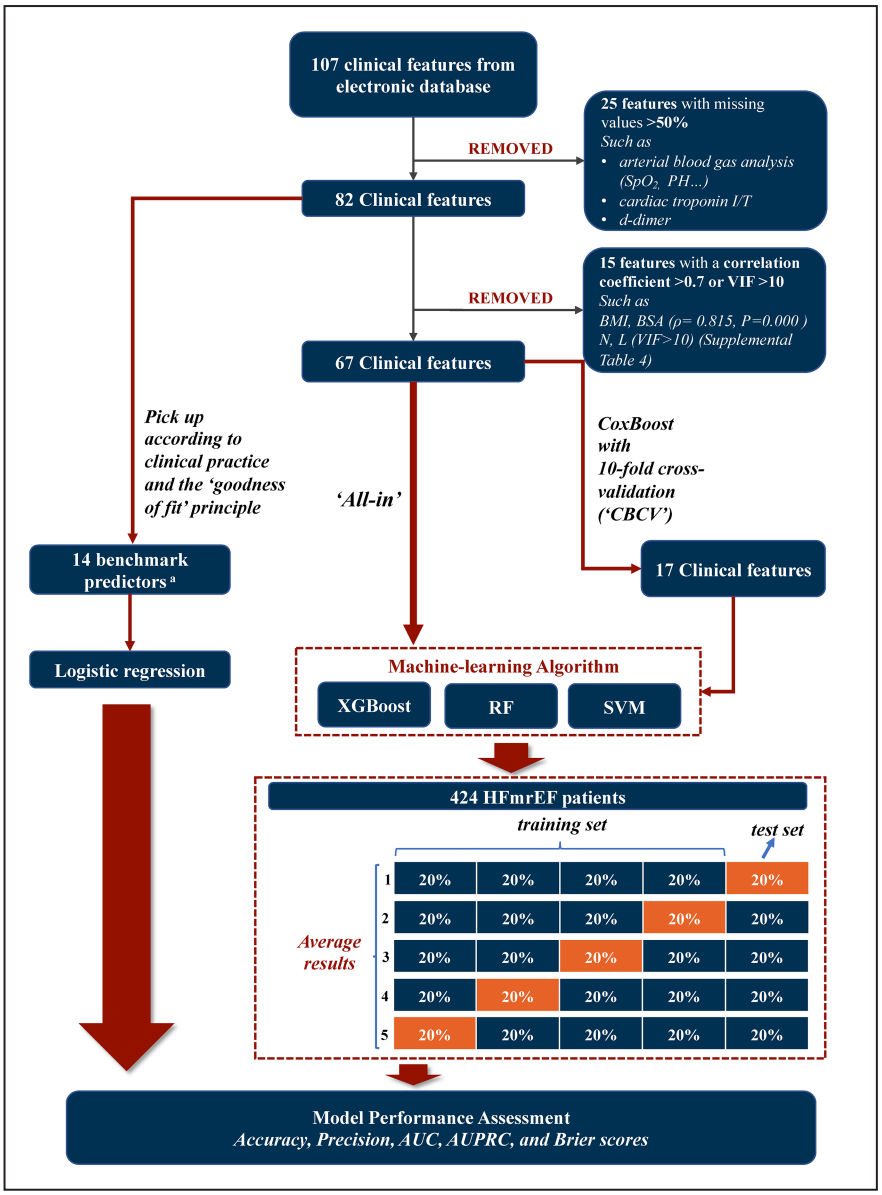


Figure 1. Flowchart for feature selection and model development.
[a]14 benchmark predictors: age, sex, SBP, Na, NYHA III/IV, eGFR*, ischemic heart failure, diabetes, anemia, NT-proBNP, ACEIs/ARBs treatment, β-blockers treatment, LVEF, and hemoglobin. ACEI indicates angiotensin-converting enzyme inhibitor; ARB, angiotensin receptor blocker; AUC, area under the receiver operating characteristic curve; AUPRC, area under the precision-recall curve; BMI, body mass index; BSA, body surface area; CBCV, Cox Boost with 10-fold cross-validation; eGFR, estimated glomerular filtration rate; HFmrEF, heart failure with mildly reduced ejection fraction; L, lymphocyte; LVEF, left ventricular ejection fraction; N, neutrophil; NT-proBNP, N-terminal probrain natriuretic peptide; NYHA, New York Heart Association; RF, random forest; SBP, systolic blood pressure; SVM, support vector machine; and XGBoost, eXtreme Gradient Boosting. *eGFR was calculated from the creatinine by the chronic kidney disease epidemiology collaboration (CKD-EPI) equation.

RESULTS

Baseline Characteristics

The baseline characteristics of the study population are summarized in Table 1. Among the included patients, 71.2% were men, and the mean age was 59.34±14.25 years old. Nearly 70% of patients were in New York Heart Association III or IV. Ischemic heart disease was the leading cause of heart failure (43.39%). The mean LVEF of these patients was 43.0%. The use rates of angiotensin-converting enzyme inhibitors or angiotensin receptor blockers, β-blockers, and

Table 1. Baseline Characteristics of Patients With Heart Failure With Mildly Reduced Ejection Fraction

Characteristics	All patients (n=424)	Survivor (n=303)	Nonsurvivor (n=121)	*P* value
Demographic features				
Male, n (%)	302 (71.2)	226 (74.6)	76 (62.8)	0.021
Age, y	59.34±14.25	57.10±14.39	64.97±12.27	<0.001
Vital signs				
Heart rate, bpm	76.00 (66.00, 88.00)	77.00 (66.00, 89.00)	75.00 (66.00, 84.00)	0.338
Body mass index, kg/m^2	24.87±4.33	25.23±4.33	23.96±4.21	0.006
Systolic pressure, mm Hg	123.41±20.73	123.69±20.32	122.69±21.78	0.652
Diastolic pressure, mm Hg	73.53±13.40	74.30±13.41	71.58±13.24	0.059
NYHA III/IV, n (%)	293 (69.1)	198 (65.3)	95 (78.5)	0.011
Laboratory data				
NT-proBNP, pg/mL	1712.50 (816.45, 3523.39)	1486.30 (669.67, 2886.15)	2596.00 (1188.30, 5039.60)	<0.001
Neutrophil-to-lymphocyte ratio	2.64 (1.92, 3.67)	2.49 (1.82, 3.37)	3.08 (2.23, 4.61)	<0.001
Platelet, ×10^9/L	203.85±76.93	207.63±76.15	194.38±78.36	0.109
Hemoglobin, g/L	136.54±22.38	139.98±21.55	127.93±22.18	<0.001
Red blood cell distribution, %	13.40 (12.70, 14.50)	13.30 (12.70, 14.10)	14.00 (13.00, 15.20)	<0.001
Total protein, g/L	68.18±6.78	68.41±6.68	67.60±7.03	0.264
Albumin, g/L	39.90±4.75	40.32±4.72	38.86±4.68	0.004
AST/ALT ratio	1.07 (0.75, 1.56)	1.00 (0.72, 1.40)	1.40 (0.93, 1.87)	<0.001
Alkaline phosphatase, IU/L	68.00 (55.75, 86.25)	68.00 (55.00, 85.00)	70.00 (57.00, 91.00)	0.223
γ-Glutamyl transpeptidase, IU/L	39.00 (26.00, 68.25)	39.00 (26.00, 66.50)	40.00 (28.00, 69.00)	0.501
Total bilirubin, μmol/L	18.20 (13.40, 25.85)	17.80 (12.90, 25.30)	19.60 (14.00, 27.80)	0.062
K, mmol/L	4.05±0.52	4.02±0.51	4.11±0.55	0.143
Na, mmol/L	137.70±3.93	138.15±3.68	136.59±4.31	<0.001
Cl, mmol/L	102.60±4.57	102.98±4.52	101.65±4.58	0.007
Ca, mmol/L	2.28 (2.19, 2.37)	2.29 (2.20, 2.38)	2.24 (2.14, 2.35)	0.007
P, mmol/L	1.28±0.25	1.28±0.26	1.26±0.22	0.448
Glucose, mmol/L	5.30 (4.72, 6.51)	5.26 (4.70, 6.24)	5.47 (4.96, 6.90)	0.007
Cr/BUN ratio	12.30 (10.21, 15.09)	12.49 (10.47, 15.11)	11.66 (9.76, 14.29)	0.036
eGFR*, mL/min per 1.73 m^2	70.50±25.42	74.18±24.04	61.26±26.42	<0.001
Uric acid, μmol/L	424.51 (344.38, 515.36)	421.82 (338.36, 510.55)	429.00 (358.99, 543.70)	0.232
Creatine kinase, IU/L	62.00 (41.00, 91.25)	62.00 (42.00, 89.00)	62.00 (40.00, 99.00)	0.824
Lactate dehydrogenase, IU/L	187.00 (159.00, 233.00)	184.00 (153.50, 224.00)	208.00 (167.00, 262.00)	<0.001
Apolipoprotein, mg/L	174.10 (83.33, 365.06)	174.10 (80.84, 377.09)	174.10 (86.14, 312.43)	0.699
High-sensitivity C-reactive protein, mg/L	3.33 (1.46, 8.66)	2.87 (1.17, 7.21)	4.74 (2.55, 10.68)	<0.001
Triglyceride, mmol/L	1.44 (1.04, 1.94)	1.49 (1.10, 1.98)	1.27 (0.89, 1.74)	0.002
Total cholesterol, mmol/L	4.11 (3.41, 4.74)	4.15 (3.49, 4.81)	3.87 (3.21, 4.55)	0.003
High-density lipoprotein, mmol/L	0.97 (0.82, 1.20)	0.98 (0.84, 1.21)	0.94 (0.77, 1.18)	0.038
Erythrocyte sedimentation rate, mm/h	8.00 (3.75, 19.00)	7.00 (3.00, 17.00)	12.00 (5.00, 27.00)	<0.001
Prothrombin time, s	13.90 (13.30, 14.90)	13.70 (13.20, 14.60)	14.20 (13.60, 15.50)	<0.001
Activated partial thromboplastin time, s	37.70 (34.77, 41.50)	37.40 (35.00, 40.95)	38.60 (34.40, 42.90)	0.172
Thrombin time, s	16.30 (15.60, 17.20)	16.30 (15.70, 17.20)	16.30 (15.30, 17.30)	0.355
Fibrinogen, g/L	3.54 (2.99, 4.23)	3.51 (2.91, 4.11)	3.68 (3.19, 4.60)	0.017
Free triiodothyronine, pg/mL	2.68 (2.38, 2.94)	2.72 (2.47, 2.99)	2.48 (2.22, 2.74)	<0.001
Free thyroxine, ng/mL	1.19 (1.08, 1.37)	1.19 (1.07, 1.35)	1.23 (1.11, 1.43)	0.034
Total triiodothyronine, ng/mL	0.92 (0.77, 1.13)	0.94 (0.81, 1.13)	0.86 (0.66, 1.04)	<0.001
Total thyroxine, μg/mL	7.93 (6.66, 9.12)	7.93 (6.69, 9.15)	7.90 (6.65, 9.10)	0.65
Thyroid-stimulating hormone, μIU/mL	1.91 (1.14, 2.94)	1.91 (1.20, 2.91)	1.87 (1.08, 3.16)	0.613

(*Continued*)

Table 1. Continued

Characteristics	All patients (n=424)	Survivor (n=303)	Nonsurvivor (n=121)	*P* value
Echocardiography parameters				
Left atrial diastolic diameter, mm	44.00 (40.00, 49.00)	44.00 (40.00, 49.00)	44.00 (40.00, 50.00)	0.578
Left ventricular end-diastolic diameter, mm	60.00 (55.00, 65.00)	61.00 (56.00, 66.00)	59.00 (53.00, 64.00)	0.047
Left ventricular posterior wall, mm	9.70 (9.00, 10.00)	9.70 (9.00, 10.00)	9.70 (8.00, 10.00)	0.254
Interventricular septum, mm	10.00 (9.00, 11.00)	10.00 (9.00, 11.00)	10.00 (9.00, 10.00)	0.72
Left ventricular ejection fraction, %	43.00 (40.00, 45.00)	43.00 (40.00, 45.00)	43.00 (40.00, 46.00)	0.959
Right ventricular end-diastolic diameter, mm	23.00 (21.00, 25.00)	22.73 (20.99, 25.00)	23.00 (21.00, 26.00)	0.106
Left ventricular aneurysm, n (%)	24 (5.7)	18 (5.9)	6 (5.0)	0.871
Ventricular thrombus, n (%)	8 (1.9)	6 (2.0)	2 (1.7)	1
Ventricular wall movement abnormalities, n (%)	406 (95.8)	290 (95.7)	116 (95.9)	1
Comorbidities, n (%)				
Mitral valve regurgitation	134 (31.6)	91 (30.0)	43 (35.5)	0.325
Tricuspid valve regurgitation	93 (21.9)	60 (19.8)	33 (27.3)	0.121
Aortic valve regurgitation	52 (12.3)	37 (12.2)	15 (12.4)	1
Mitral valve stenosis	15 (3.5)	10 (3.3)	5 (4.1)	0.898
Aortic valve stenosis	25 (5.9)	15 (5.0)	10 (8.3)	0.28
Atrial fibrillation	142 (33.5)	98 (32.3)	44 (36.4)	0.498
Anemia	70 (16.5)	40 (13.2)	30 (24.8)	0.006
Hypertension	232 (54.7)	174 (57.4)	58 (47.9)	0.096
Diabetes	134 (31.6)	79 (26.1)	55 (45.5)	<0.001
Ischemic heart failure	184 (43.4)	116 (38.3)	68 (56.2)	0.001
Guide-directed medicine treatments, n (%)				
ACEIs or ARBs	280 (66.0)	217 (71.6)	63 (52.1)	<0.001
β-blockers	377 (88.9)	275 (90.8)	102 (84.3)	0.081
Mineralocorticoid receptor antagonists	318 (75.0)	233 (76.9)	85 (70.2)	0.192
Diuretics	366 (86.3)	260 (85.8)	106 (87.6)	0.742

ACEI indicates angiotensin-converting enzyme inhibitor; ALT, alanine transaminase; ARB, angiotensin receptor blocker; AST, aspartate aminotransferase; BUN, blood urea nitrogen; Cr, creatinine; eGFR, estimated glomerular filtration rate; NT-proBNP, N-terminal pro-B-type natriuretic peptide; and NYHA, New York Heart Association.

*eGFR was calculated from the creatinine by the Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) equation.

mineralocorticoid receptor antagonists were 66%, 88.9%, and 86.3%, respectively.

During a median follow-up of 1008 (750, 1937) days, 121 patients met the primary end point. The 1-year, 3-year, and 5-year mortality rates were 10.6%, 20%, and 24.29%, respectively. In comparison with survivors, the deceased was older with worse cardiac and renal function and higher comorbidity of ischemic heart disease (IHD), diabetes, and anemia (Table 1).

Model Performance of the "All-in" MLBPMs

The performance of the "All-in" MLBPMs and the reference logistic model are summarized in Table 2. The predictive performance of the XGBoost model was superior to that of the other MLBPMs and the reference logistic model (accuracy, 85.4%; precision, 70.3%; recall, 84.3%; F1-score, 0.767; AUC, 0.916; 95% CI, 0.887–0.945; area under the precision-recall curve [AUPRC], 0.838; Brier score, 0.12). This was followed by the RF model, with an accuracy and precision of 82.1% and 64.3%, respectively. It also had good discrimination and calibration, with an AUC of 0.900 (0.870–0.930) and a Brier score of 0.13. The performance of the SVM model was inferior to that of the RF model but was better than that of the reference logistic model. The receiver operating characteristic curves and PR curves for the MLBPMs and reference logistic model are plotted in Figures 2A and 2B. Decision curve analyses showed that the MLBPMs had a higher clinical net benefit than the reference logistic model within the threshold risk of 5% to 76%. Among the MLBPMs, the XGBoost model had the highest clinical net benefit within the risk threshold of 22% to 76%. The RF model had better clinical performance than the SVM and

Table 2. Results of Different Models for Mortality Prediction in Patients With Heart Failure With Mildly Reduced Ejection Fraction

Model performance	Accuracy	Precision	Recall	F1-Score	AUC (95% CI)	AUPRC	Sensitivity	Specificity	Brier score
"All-in" MLBPMs									
XGBoost	0.854	0.703	0.843	0.767	0.916 (0.887–0.945)	0.838	0.843	0.858	0.12
Random forest	0.821	0.643	0.835	0.727	0.900 (0.870–0.930)	0.793	0.835	0.815	0.13
SVM	0.781	0.588	0.777	0.669	0.807 (0.757–0.857)	0.700	0.777	0.782	0.15
"CBCV" MLBPMs									
XGBoost	0.804	0.630	0.760	0.689	0.852 (0.811–0.893)	0.728	0.760	0.822	0.14
Random forest	0.774	0.575	0.793	0.667	0.858 (0.821–0.896)	0.734	0.793	0.766	0.14
SVM	0.771	0.58	0.719	0.642	0.780 (0.727–0.832)	0.640	0.719	0.792	0.16
Reference model									
Logistic*	0.747	0.551	0.620	0.584	0.755 (0.702–0.807)	0.577	0.620	0.798	0.17

AUC indicates area under the receiver operating curve; AUPRC, area under the precision-recall curve; CBVC, CoxBoost with 10-fold cross-validation; MLBPM, machine-learning-based prediction model; SVM, support vector machine; and XGBoost, extreme Gradient Boosting.

*Adjusted for 14 benchmarker predictors: age, sex, systolic blood pressure, hemoglobin, N-terminal pro-B-type natriuretic peptide, sodium, left ventricular ejection fraction, estimated glomerular filtration rate, New York Heart Association III/IV, ischemic heart failure, diabetes, anemia, treatment with angiotensin-converting enzyme inhibitors/angiotensin receptor blockers, treatment with β-blockers.

the reference logistic model within a risk threshold of 22% to 56% and 5% to 59%, respectively (Figure 2C). Moreover, prediction distribution plots demonstrated that the XGBoost model, the RF model, and the SVM model could be more precise in stratifying patients at risk of all-cause mortality than the reference logistic model by positively clustering patients who died during long-term follow-up in the order of risk (Figure 3A through 3D).

Model Performance of the MLBPMs With "CBCV" Selection Strategy

The performance of the "CBCV" MLBPMs are also summarized in Table 2. Still, the XGBoost model had the best predictive accuracy, precision, and F1 score, which were 80.4%, 60.3%, and 0.689%, respectively. The RF model seemed to have a better discrimination performance than the XGBoost model, with an AUC and AUPRC of 0.858 and 0.734, respectively. However, the De Long test showed that the discrimination performance between these 2 models was equivalent (P=0.673). Overall, the XGBoost model and the RF model outperformed the SVM model and the reference logistic model in mortality prediction in patients with HFmrEF (Table 2 and Figure 4A and 4B). Moreover, model performance declined in "CBCV" MLBPMs compared with the corresponding "All-in" MLBPMs, with lower accuracy, precision, F1-score, recall, AUC (P<0.05 for all comparisons), and AUPRC, and higher Brier scores (Table 2 and Figure S3A through S3F). Decision curve analyses showed that the 3 "CBCV" MLBPMs had greater clinical net benefits than the reference logistic model within a risk threshold of 22% to 70%. The XGBoost model and RF model had better clinical performance than the SVM model. The RF model outperformed the XGBoost model when the risk threshold was >58%. The SVM model outperformed the RF model when the threshold was >70% (Figure 4C). Moreover, when combined with the "All-in" MLBPMs, the clinical performance of the "CBCV" XGBoost model was inferior to that of the "All-in" model, and the "CBCV" RF model was superior to that of the "All-in" models (Figure S3G and S3H). The clinical net benefits were similar in the "All-in" and the "CBCV" SVM model (Figure S3I). The prediction distribution plots demonstrated that the MLBPMs could also be more precise in stratifying patients at higher risk of all-cause mortality than the reference logistic model, which was similar to the "All-in" MLBPMs (Figure S4A through S4D).

Feature Importance

As the "All-in" XGBoost model performed best in mortality prediction, its feature importance was further analyzed. The feature importance plot lists the enrolled variables in descending order (Figure 5A). Age was the strongest predictive feature, followed

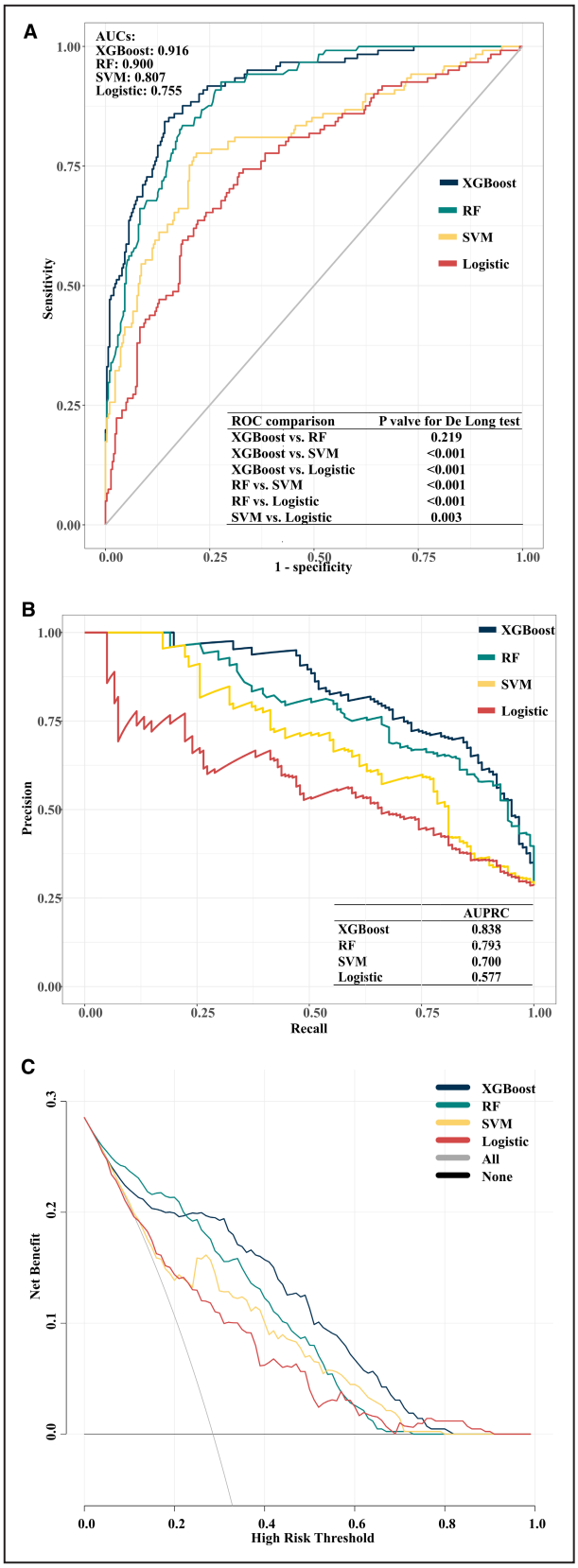


Figure 2. Model prediction performance and clinical usefulness of the "All-in" MLBPMs.
"All-in" MLBPMs showed good discrimination performance by ROC (**A**) and PR curves (**B**) and higher clinical net benefit **(C)**. AUC indicates area under the receiver operating characteristic curve; AUPRC, area under the precision-recall curve; MLBPM, machine-learning-based prediction model; PR, precision-recall; RF, random forest; ROC, receiver operating characteristic; SVM, support vector machine; and XGBoost, eXtreme Gradient Boosting.

by the ratio of aspartate aminotransferase to alanine transaminase, serum sodium level, serum potassium level, NT-proBNP, etc. SHAP analysis showed that older age and higher aspartate aminotransferase/alanine transaminase and NT-proBNP were associated with a higher risk of all-cause mortality. A higher hemoglobin level was associated with a lower risk of all-cause mortality. The results of other features in the SHAP analysis are demonstrated in Figure 5B. The gain of each feature to the model is listed in Table S7.

Subgroup Analysis

We conducted a series of subgroup analyses based on the "All-in" XGBoost model, and the results are summarized in Table 3. The predictive power of the "All-in" XGBoost model was good in patients with HFmrEF with LVEF 41% to 49% (accuracy, 86.5%; precision, 71.8%; AUC, 0.910 [95% CI, 0.869–0.951]; Brier score: 0.12) and was consistent with that of the whole study population (P=0.808 for AUC comparison). Among different HFmrEF subgroups, such as between patients ≤65 and >65 years old, among different body mass index, between different SBP or diastolic blood pressure levels, and with or without renal dysfunction, that power remained strong and was consistent (P>0.05 for all AUC comparison; Table 3). Moreover, the "all-in" XGBoost model also showed satisfying predictive performance when the observation time of outcome was limited to 1, 3, and 5 years. The clustered prediction distribution plots also showed good risk stratification among these subgroup patients and at different observation times (Figure S5A through S5Y). However, model precision, recall, and AUPRC were remarkably lower at the 1-year follow-up than at the other follow-ups, which may be related to the lower event rate. The recall and AUPRC of patients with estimated glomerular filtration rate ≤60 mL/min per 1.73 m^2 or ≥65 years old were significantly higher than those with estimated glomerular filtration rate>60 mL/min per 1.73 m^2 or <65 years old, which may be attributed to the impact of renal dysfunction or old age on mortality. This was also observed in patients with or without diabetes, which the adverse effect of diabetes on nonevent survival could well explain. The predictive specificity for the underweight subgroup was only 45.5%, which may be mainly due to the limited sample size. Only 22 patients were underweight (body mass index <18.5 kg/m^2) in this study.

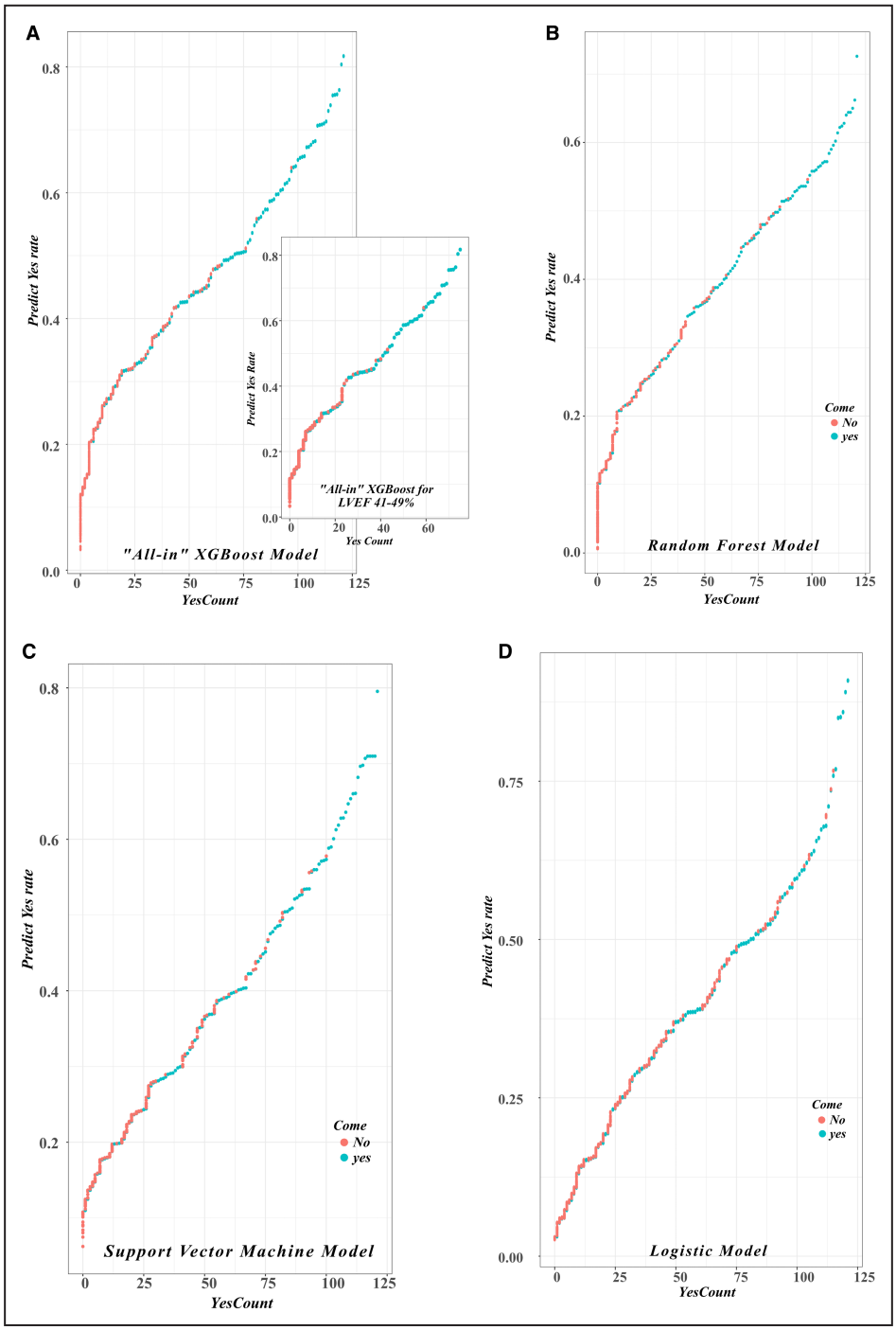


Figure 3. Prediction distribution of "All-in" MLBPMs in patients with HFmrEF.
A, The XGBoost model. **B**, The RF model. **C**, The SVM model. **D**, The reference logistic model. The machine learning model could be more precise in stratifying patients at risk of all-cause mortality by positively clustering patients, in order of risk, who died during long-term follow-up. The prediction distributions were consistent between the whole HFmrEF population and those with LVEF 41% to 49%. HFmrEF indicates heart failure with mildly reduced ejection fraction; LVEF, left ventricular ejection fraction; MLBPM, machine-learning-based prediction model; RF, random forest; SVM, support vector machine; and XGBoost, eXtreme Gradient Boosting.

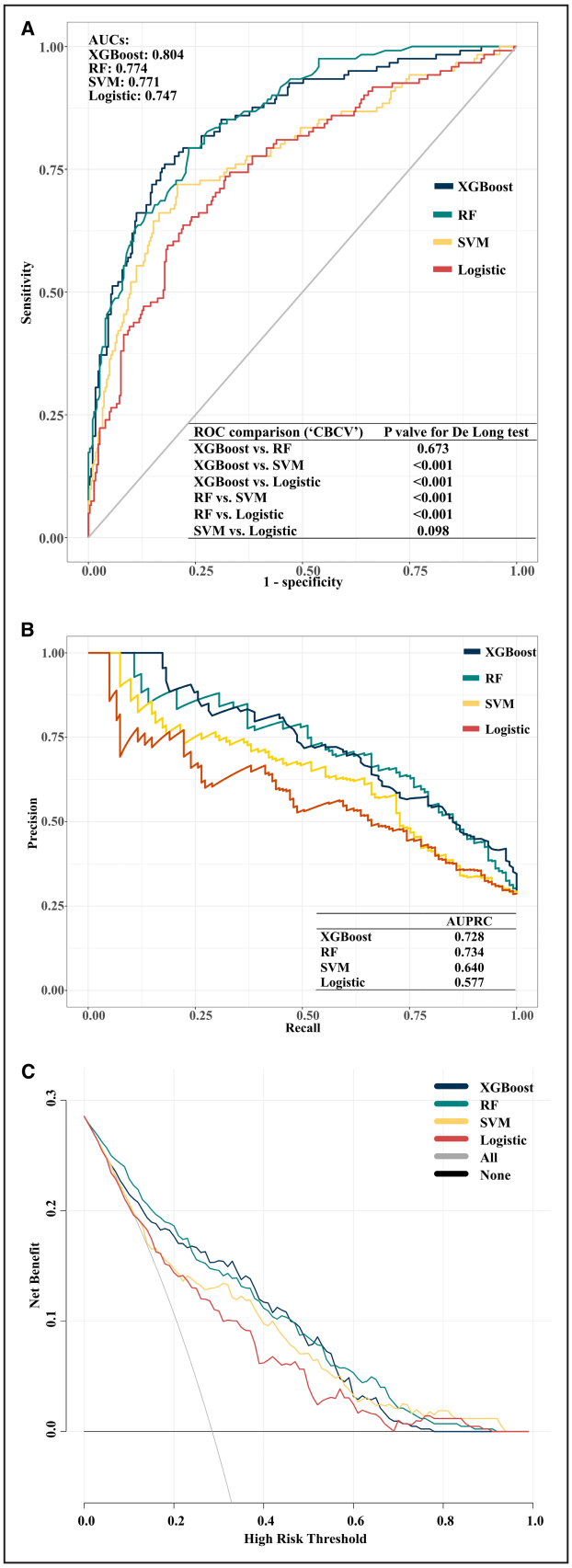


Figure 4. Model prediction performance and clinical usefulness of the "CBCV" MLBPMs.
"CBCV" MLBPMs showed good discrimination performance by ROC (**A**) and PR curves (**B**) and higher clinical net benefit (**C**). AUC indicates area under the receiver operating characteristic curve; AUPRC, area under the precision-recall curve; CBCV, CoxBoost with 10-fold cross-validation; MLBPMs, machine-learning-based prediction models; ROC, receiver operating characteristic; PR, precision-recall; RF, random forest; SVM, support vector machine; and XGBoost, eXtreme Gradient Boosting.

DISCUSSION

Few ML-based model studies in HF have focused on patients with HFmrEF and their long-term outcomes. In this pilot study, we evaluate the predictive value of MLBPMs in long-term all-cause mortality and their preponderance in aiding decision-making in clinical practice in patients with HFmrEF. Our main findings were (1) the overall predictive performance of the MLBPMs was superior to that of the reference logistic model and had greater clinical net benefit. (2) The "All-in" MLBPMs outperformed the "CBCV" MLBPMs in mortality prediction, highlighting the importance of comprehensive evaluation in risk prediction while indicating the complex relationship between different variables. (3) Compared with other ML algorithms, the XGBoost-derived prediction models had the best predictive performance. (4) The predictive power of the "All-in" XGBoost model remained strong in different subgroups of patients and at different outcome observation times.

First proposed in 2016, HFmrEF has become a new hot topic in heart failure. HFmrEF accounts for 21.8% of patients with heart failure in China[25] and ≈25% worldwide.[19] Accurate risk stratification is one of the key components in managing these patients. The Seattle Heart Failure Model score system, the BCN Bio-HF Calculator, and the Meta-Analysis Global Group in Chronic Heart Failure score were all used to predict the mortality of patients with heart failure with reduced ejection. The AUCs or c-indexes were all lower than 0.8.[1–3,26] The BIOSTAT-HF score, used to assess the risk for heart failure with preserved ejection fraction and patients with heart failure with reduced ejection, reached similar results with c-indexes of only 0.74 and 0.72, respectively.[27] Therefore, the existing risk score systems neither specifically targeted the HFmrEF population nor showed an excellent outcome prediction.

Moreover, numerous studies have proven the feasibility and effectiveness of MLBPMs in heart failure short-term mortality prediction,[6,7,9,10,16,28] but few have focused on their long-term power in patients with HFmrEF. Our study filled that gap in this area. We assessed the long-term performance of 3 MLBPMs (XGBoost, RF, and SVM), including 67 clinical features

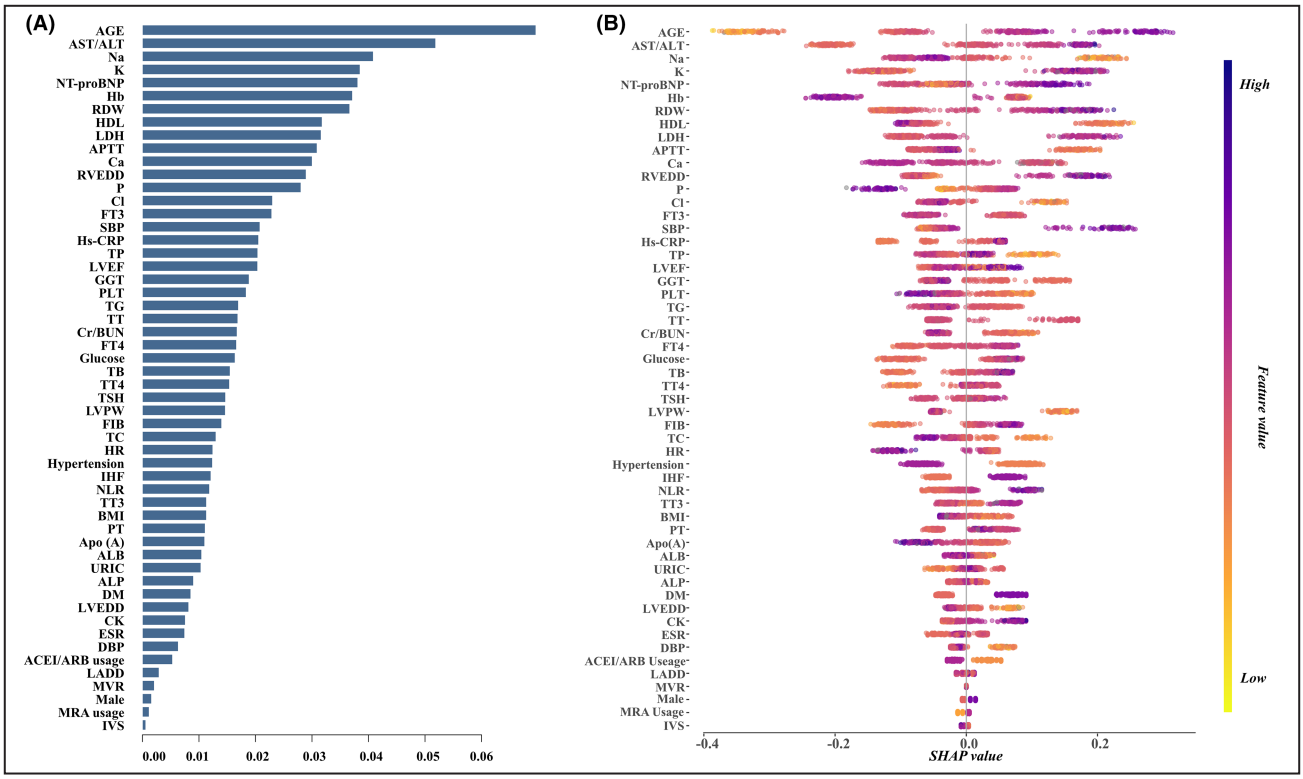


Figure 5. Feature importance and SHAP analysis of the "All-in" XGBoost model.
Feature importance plot of the "all-in" XGBoost (**A**) and the explanation by SHAP analysis (**B**). SHAP analysis was helpful in understanding the direction of a specific feature to a higher or lower prediction, and the color indicates whether that feature is high (in purple) or low (in yellow) for that observation. ACEI indicates angiotensin-converting enzyme inhibitor; ALB, albumin; ALT, alanine transaminase; ALP, alkaline phosphatase; Apo (A), apolipoprotein; APTT, activated partial thromboplastin time; ARB, angiotensin receptor blocker; AST, aspartate aminotransferase; BMI, body mass index; BUN, blood urea nitrogen; CK, creatine kinase; Cr, creatinine; DBP, diastolic blood pressure; DM, diabetes mellitus; ESR, erythrocyte sedimentation rate; FIB,fibrinogen; FT3, free triiodothyronine; FT4, free thyroxine; GGT, γ-glutamyl transpeptidase; Hb, hemoglobin; HDL, high-density lipoprotein; HR, heart rate; Hs-CRP, high-sensitivity C-reactive protein; IHF, ischemic heart failure; IVS, interventricular septum; LADD, left atrial diastolic diameter; LDH, lactate dehydrogenase; LVEDD, left ventricular end-diastolic diameter; LVEF, left ventricular ejection fraction; LVPW, left ventricular posterior wall; MRA, mineralocorticoid receptor antagonist; MVR, mitral valve regurgitation; NLR, neutrophil to lymphocyte ratio; NT-proBNP, N-terminal probrain natriuretic peptide; NYHA, New York Heart Association; PLT, platelet; PT, prothrombin time; RDW, red blood cell distribution; RVEDD, right ventricular end-diastolic diameter; SBP, systolic blood pressure; SHAP, Shapley additive explanations; TB, total bilirubin; TG, triglyceride; TC, total cholesterol; TSH, thyroid-stimulating hormone; TT, thrombin time; TT3, total triiodothyronine; TT4, total thyroxine; URIC, uric acid; and XGBoost, eXtreme Gradient Boosting.

(the "All-in" model), in predicting mortality and showed a remarkable increase in predictive accuracy, precision, discrimination, and calibration to the reference logistic model, especially the XGBoost model (accuracy, 85.4%; precision, 70.3%; AUC, 0.916 [95% CI, 0.887–0.945]; Brier score: 0.12). Compared with other algorithms, the XGBoost model had higher accuracy and precision, and better discrimination and calibration (Table 2). This result is consistent with the existing short-term outcome studies that have been reported in patients with acute heart failure (AHF). Polo Friz et al showed that the XGBoost model best predicted 30-day readmission in patients with AHF, with an AUC of 0.803.[29] Li et al reported an accuracy and AUC of 82.6% and 0.824 of the XGBoost model in predicting in-hospital mortality of patients with AHF, which was higher than the RF model and SVM model, with an accuracy of 82.3% and 80.1% and an AUC of 0.779 and 0.701, respectively.[30]

Although the comparison between studies should be made with caution, our "All-in" XGBoost model performance outperformed the CoxBoost model proposed by Park et al[28] at the 3-year follow-up (0.900 [95% CI, 0.863–0.935] versus 0.761 [95% CI, 0.754–0.767]). Furthermore, to the best of our knowledge, this is the first study that reported the performance of the XGBoost model in predicting 5-year mortality (accuracy, 83.5%; precision, 61.5%; AUC, 0.910 [95% CI, 0.878–0.941]; Brier score: 0.11) in this area. However, the prediction performance was less satisfactory when the observation time was limited to 1 year. The model accuracy and precision were only 75.5% and 29.4%,

Table 3. Performance of "All-in" eXtreme Gradient Boosting Model in Different Subgroups and Observation Times

Model performance	Accuracy	Precision	Recall	F1-score	AUC (95% CI)	*P* value for AUC comparison	AUPRC	Sensitivity	Specificity	Brier score
LVEF 41%–49%*	0.865	0.718	0.813	0.762	0.910 (0.869–0.951)	0.808‡	0.836	0.813	0.884	0.12
Nonischemic	0.883	0.705	0.811	0.754	0.918 (0.875–0.961)	0.723	0.804	0.811	0.904	0.11
Ischemic	0.815	0.702	0.868	0.776	0.907 (0.865–0.949)		0.868	0.868	0.784	0.15
Age >65 y	0.813	0.705	0.902	0.791	0.914 (0.868–0.961)	0.984	0.901	0.902	0.755	0.15
Age ≤65 y	0.877	0.701	0.783	0.740	0.915 (0.878–0.952)		0.765	0.783	0.904	0.11
Male	0.874	0.732	0.789	0.759	0.914 (0.878–0.950)	0.777	0.818	0.789	0.903	0.12
Female	0.803	0.667	0.933	0.778	0.922 (0.878–0.967)		0.875	0.933	0.727	0.14
SBP <100 mm Hg	0.851	0.722	0.867	0.788	0.925 (0.854–0.996)	0.811	0.848	0.867	0.844	0.13
SBP ≥100 mm Hg	0.854	0.701	0.840	0.764	0.915 (0.884–0.947)		0.838	0.840	0.860	0.12
DBP <60 mm Hg	0.75	0.621	0.947	0.75	0.882 (0.788–0.977)	0.459	0.851	0.947	0.621	0.15
DBP ≥60 mm Hg	0.867	0.724	0.824	0.771	0.920 (0.889–0.951)		0.844	0.824	0.883	0.12
eGFR ≤60 mL/min per 1.73 m^2	0.821	0.714	0.948	0.815	0.929 (0.889–0.970)	0.357	0.916	0.948	0.732	0.14
eGFR>60 mL/min per 1.73 m^2	0.870	0.691	0.746	0.718	0.902 (0.860–0.944)		0.752	0.746	0.905	0.11
NYHA III/IV	0.826	0.683	0.863	0.763	0.905 (0.869–0.941)	0.342	0.848	0.863	0.808	0.14
NYHA I/II	0.916	0.8	0.769	0.874	0.937 (0.882–0.992)		0.828	0.769	0.952	0.10
Underweight†	0.727	0.647	1	0.786	0.818 (0.631–1)	*P*>0.05 for all	0.802	1	0.455	0.19
Normal weight†	0.856	0.755	0.851	0.8	0.934 (0.891–0.977)		0.899	0.851	0.859	0.13
Overweight†	0.832	0.661	0.822	0.733	0.897 (0.845–0.950)		0.802	0.822	0.836	0.13
Obesity†	0.907	0.722	0.765	0.743	0.924 (0.858–0.990)		0.796	0.765	0.938	0.09
With diabetes	0.851	0.761	0.927	0.761	0.930 (0.890–0970)	0.276	0.902	0.927	0.797	0.14
Without diabetes	0.855	0.654	0.773	0.708	0.897 (0.854–0.94)		0.770	0.773	0.879	0.11
With Af	0.859	0.722	0.886	0.796	0.933 (0.894–0.972)	0.364	0.863	0.886	0.847	0.13
Without Af	0.851	0.692	0.818	0.75	0.908 (0.869–0.946)		0.828	0.818	0.863	0.12
1-y Observation time	0.755	0.294	0.933	0.447	0.912 (0.862–0.962)	*P*>0.05 for all	0.628	0.933	0.734	0.09
3-y Observation time	0.821	0.531	0.894	0.667	0.900 (0.863–0.935)		0.696	0.894	0.802	0.11
5-y Observation time	0.835	0.615	0.854	0.715	0.910 (0.878–0.941)		0.796	0.854	0.829	0.11

Af indicates atrial fibrillation; AUC, area under receiver operating curve; AUPRC, area under the precision-recall curve; DBP, diastolic blood pressure; eGFR, estimated glomerular filtration rate; LVEF, left ventricular ejection fraction; NYHA, New York Heart Association; and SBP, systolic blood pressure.

*According to the definition of heart failure with mildly reduced ejection fraction in 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure.

†The body mass index (BMI) cut-off values for underweight, normal weight, overweight, and obesity were <18.5, 18.5≤BMI<23.5, 23.5≤BMI<27.5, and ≥27.5, according to the recommendation of the World Health Organization for Asia population.

‡Compared with the whole heart failure with mildly reduced ejection fraction population in this study.

respectively, which might be attributed to the lower event rate (10.6%) and the limited sample size. Two similar studies based on the MIMIC-III database and e-ICU-CRD for in-hospital mortality prediction showed similar results,[30,31] with an accuracy of 76% and 82.6%, and a precision of 43.2% and 30.7%, respectively. The event rates of these 2 studies were 13.52% and 9.97%.

Moreover, our study assessed the effectiveness of ML feature selection and the performance of 3 MLBPMs (the "CBCV" model) in predicting mortality. Our results showed that the "CBCV" MLBPMs had better prediction performance and higher clinical net benefit than the reference logistic model. However, the model performance of these MLBPMs was still inferior to that of the "All-in" models (Table 2 and Figure S3A through S3F), which could be explained by feature insufficiency. Nevertheless, the improvement in clinical benefit was different among the "CBCV" MLBPMs. The "CBCV" RF model yielded a higher clinical net benefit than that of the "All-in" model, while the benefit declined in the "CBCV" XGBoost model and was equivalent between the 2 SVM models. This result implied the importance of optimizing the strategy for both feature selection and model construction for patients with HFmrEF in the future to better balance the prediction performance and clinical utility.

Model interpretation has long been a major obstacle to applying ML-based models in clinical practice. They have been called the "black box." Using the SHAP method, our study showed that age was the most important feature in mortality prediction. This was consistent with the results of the HFmrEF subgroup analysis reported in a previous study.[28] Subgroup analyses on ages ≤65 and >65 years old showed that the overall performance of our XGBoost model was consistent. Limited cohort size and old age may explain the model's remarkably higher recall observed in patients older than 65 years. Our study also found that the aspartate aminotransferase/alanine transaminase ratio was important in outcome prediction. A higher ratio was associated with a worse outcome, possibly due to severe myocardial pathology.[32] Its clinical significance needs to be further elucidated in HFmrEF. Moreover, it is interesting to observe that the right ventricular end-diastolic diameter ranked higher than the left ventricular end-diastolic diameter (LVEDD) in mortality prediction, with a larger diameter indicating a worse outcome. This finding supplements what Alberto et al. have reported.[33] They found that right ventricular end-diastolic diameter predicts poor outcomes better than left ventricular end-diastolic diameter in heart failure with preserved ejection fraction and heart failure with reduced ejection. The narrowed LVEF defining the HFmrEF restrains the variation in left ventricular end-diastolic diameter, which may also explain this finding. In addition, only 18 patients had a SBP <90 mm Hg at admission, which could reasonably explain the contradictory results that patients with lower SBP were predicted to have a better prognosis. In summary, the SHAP results of the present study need further validation in an external HFmrEF cohort.

Compared with patients of other ethnicities, patients with heart failure in China have some unique clinical characteristics. The China-HF registry reported that patients with heart failure in China were ≈5 years younger (64.8±15.2 in China-HF versus 72±14 in ADHERE [Acute Decompensated Heart Failure National Registry] and 70±13 in EHFS II), with 8 to 16 mm Hg lower SBP (128±26 mm Hg in China-HF versus 136±31 mm Hg in KorAHF [Korean Acute Heart Failure] and 144±33 mm Hg in ADHERE), lower body mass index (23.7±4.3 kg/m^2 in China-HF versus 26.8 in EHFS II) and lower comorbidity of coronary heart disease (49.6% in China-HF versus 57% in ADHERE and 53.6% in EHFS II), atrial fibrillation (24.4% in China-HF versus 31% in ADHERE and 39 in EHFS II), and diabetes (21% in China-HF versus 44% in ADHERE, 33.8 in ATTEND [Acute Decompensated heart failure syndromes], 36% in KorAHF, and 32.8% in EHFSII).[34] The contemporary China-HF stage II registry study showed similar results compared with the GWTG-HF study in the United States.[25] However, few studies have reported the comparison mentioned above in a solely HFmrEF cohort, which is worth investigating in the future. Nevertheless, our study was the first to evaluate the effectiveness of MLBPM in mortality prediction in patients with HFmrEF in China and has demonstrated its good predictive power, although partially affected by limited sample size, in patients with or without ischemic heart disease, diabetes, and across different body mass indexes and ages (Table 3). Further nationwide registry studies are needed to clarify the prevalence, management, and outcomes of patients with HFmrEF in China and to develop a more precise MLBPM in the future.

Limitations

The present study has several limitations. First, our study was a single-center study. However, as the largest national treatment center for heart failure in China, the phenotype of our patients with HFmrEF is typical and highly representative. Second, the sample size was limited. To avoid overfitting, we used 5-fold cross-validation to develop all MLBPMs and the grid search for hyperparameters. Third, this study did not include other parameters of interest in mortality prediction, such as arterial blood gas analysis, cardiac troponin I/T, and D-dimer. Each of those variables had >50% missing values in this analysis. Fourth, the treatment effect of angiotensin receptor-neurolysin inhibitor and sodium-glucose cotransporter 2 inhibitors could not

be evaluated because they were not available in China until the spring of 2018 and 2020, respectively. Fifth, we did not compare the performance of other ML models and feature selection strategies due to the limited sample size. We will further explore these aspects in subsequent studies. Last, due to the lack of an external validation cohort, the generality and applicability of our results may be compromised.

CONCLUSIONS

In this pilot study, we presented 6 machine-learning-based models based on the XGBoost, RF, and SVM algorithm with good accuracy and favorable performance in mortality prediction during a long-term follow-up in patients with HFmrEF. The machine-learning model can be more precise in risk stratification and significantly improve the clinical benefit, further optimizing patient management. Feature selection based on machine learning also showed good effectiveness and thus is worth further investigation in the future.

ARTICLE INFORMATION

Received December 7, 2022; accepted May 10, 2023.

Affiliations

Heart Failure Center, State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center for Cardiovascular Disease, Chinese Academy of Medical Science, Peking Union Medical College, Beijing, China (P.T., L.L., X.Z., B.H., J.F., L.H., Y.H., M.Z., Q.Z., J.Z., Y.Z.); and Key Laboratory of Clinical Research for Cardiovascular Medications, National Health Committee, Beijing, China (J.Z.).

Acknowledgments

We want to thank all enrolled patients, their relatives, and all staff of the heart failure center of Fuwai Hospital, CAMSA&PUMC.

Sources of Funding

This research was supported by the Key Projects in the National Science and Technology Pillar Program of the 12th Five-Year Plan Period (grant number 2011BAI11B08), Beijing, People's Republic of China; Key Projects in the National Science and Technology 6 Pillar Program of the 13th Five-Year Plan Period (grant number 2017YFC1308300), Beijing, People's Republic of China; Key Projects in the National Science and Technology 6 Pillar Program of the 13th Five-Year Plan Period (grant number 2017YFC1308305), Beijing, People's Republic of China; and Chinese Academy of Medical Science Innovation Fund for Medical Science (grant number 2021-CXGC08).

Disclosures

None.

Supplemental Material

Tables S1–S7
Figures S1–S5

REFERENCES

1. Levy WC, Mozaffarian D, Linker DT, Sutradhar SC, Anker SD, Cropp AB, Anand I, Maggioni A, Burton P, Sullivan MD, et al. The Seattle Heart Failure Model: prediction of survival in heart failure. *Circulation.* 2006;113:1424–1433. doi: 10.1161/CIRCULATIONAHA.105.584102
2. Lupón J, de Antonio M, Vila J, Peñafiel J, Galán A, Zamora E, Urrutia A, Bayes-Genis A. Development of a novel heart failure risk tool: the barcelona bio-heart failure risk calculator (BCN bio-HF calculator). *PLoS One.* 2014;9:e85466. doi: 10.1371/journal.pone.0085466
3. Pocock SJ, Ariti CA, McMurray JJ, Maggioni A, Kober L, Squire IB, Swedberg K, Dobson J, Poppe KK, Whalley GA, et al. Predicting survival in heart failure: a risk score based on 39 372 patients from 30 studies. *Eur Heart J.* 2013;34:1404–1413. doi: 10.1093/eurheartj/ehs337
4. Codina P, Lupon J, Borrellas A, Spitaleri G, Cediel G, Domingo M, Simpson J, Levy WC, Santiago-Vacas E, Zamora E, et al. Head-to-head comparison of contemporary heart failure risk scores. *Eur J Heart Fail.* 2021;23:2035–2044. doi: 10.1002/ejhf.2352
5. Johnson KW, Torres Soto J, Glicksberg BS, Shameer K, Miotto R, Ali M, Ashley E, Dudley JT. Artificial intelligence in cardiology. *J Am Coll Cardiol.* 2018;71:2668–2679. doi: 10.1016/j.jacc.2018.03.521
6. Shin S, Austin PC, Ross HJ, Abdel-Qadir H, Freitas C, Tomlinson G, Chicco D, Mahendiran M, Lawler PR, Billia F, et al. Machine learning vs. conventional statistical models for predicting heart failure readmission and mortality. *ESC Heart Fail.* 2021;8:106–115. doi: 10.1002/ehf2.13073
7. Konig S, Pellissier V, Hohenstein S, Bernal A, Ueberham L, Meier-Hellmann A, Kuhlen R, Hindricks G, Bollmann A. Machine learning algorithms for claims data-based prediction of in-hospital mortality in patients with heart failure. *ESC Heart Fail.* 2021;8:3026–3036. doi: 10.1002/ehf2.13398
8. Negassa A, Ahmed S, Zolty R, Patel SR. Prediction model using machine learning for mortality in patients with heart failure. *Am J Cardiol.* 2021;153:86–93. doi: 10.1016/j.amjcard.2021.05.044
9. Jing L, Ulloa Cerna AE, Good CW, Sauers NM, Schneider G, Hartzel DN, Leader JB, Kirchner HL, Hu Y, Riviello DM, et al. A machine learning approach to management of heart failure populations. *JACC Heart Fail.* 2020;8:578–587. doi: 10.1016/j.jchf.2020.01.012
10. Angraal S, Mortazavi BJ, Gupta A, Khera R, Ahmad T, Desai NR, Jacoby DL, Masoudi FA, Spertus JA, Krumholz HM. Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction. *JACC Heart Fail.* 2020;8:12–21. doi: 10.1016/j.jchf.2019.06.013
11. Luo C, Zhu Y, Zhu Z, Li R, Chen G, Wang Z. A machine learning-based risk stratification tool for in-hospital mortality of intensive care unit patients with heart failure. *J Transl Med.* 2022;20:136. doi: 10.1186/s12967-022-03340-8
12. Lv H, Yang X, Wang B, Wang S, Du X, Tan Q, Hao Z, Liu Y, Yan J, Xia Y. Machine learning-driven models to predict prognostic outcomes in patients hospitalized with heart failure using electronic health records: retrospective study. *J Med Internet Res.* 2021;23:e24996. doi: 10.2196/24996
13. Woolley RJ, Ceelen D, Ouwerkerk W, Tromp J, Figarska SM, Anker SD, Dickstein K, Filippatos G, Zannad F, Metra M, et al. Machine learning based on biomarker profiles identifies distinct subgroups of heart failure with preserved ejection fraction. *Eur J Heart Fail.* 2021;23:983–991. doi: 10.1002/ejhf.2144
14. Segar MW, Jaeger BC, Patel KV, Nambi V, Ndumele CE, Correa A, Butler J, Chandra A, Ayers C, Rao S, et al. Development and validation of machine learning-based race-specific models to predict 10-year risk of heart failure: a multicohort analysis. *Circulation.* 2021;143:2370–2383. doi: 10.1161/CIRCULATIONAHA.120.053134
15. Sanchez-Martinez S, Duchateau N, Erdei T, Kunszt G, Aakhus S, Degiovanni A, Marino P, Carluccio E, Piella G, Fraser AG, et al. Machine learning analysis of left ventricular function to characterize heart failure with preserved ejection fraction. *Circ Cardiovasc Imaging.* 2018;11:e007138. doi: 10.1161/CIRCIMAGING.117.007138
16. Adler ED, Voors AA, Klein L, Macheret F, Braun OO, Urey MA, Zhu W, Sama I, Tadel M, Campagnari C, et al. Improving risk prediction in heart failure using machine learning. *Eur J Heart Fail.* 2020;22:139–147. doi: 10.1002/ejhf.1628
17. Greenberg B, Adler E, Campagnari C, Yagil A. A machine learning risk score predicts mortality across the spectrum of left ventricular ejection fraction. *Eur J Heart Fail.* 2021;23:995–999. doi: 10.1002/ejhf.2155
18. Ponikowski P, Voors AA, Anker SD, Bueno H, Cleland JGF, Coats AJS, Falk V, González-Juanatey JR, Harjola VP, Jankowska EA, et al. 2016 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure: the Task Force for the diagnosis and treatment of acute and chronic heart failure of the European Society of Cardiology (ESC). Developed with the special contribution of the Heart Failure Association (HFA) of the ESC. *Eur Heart J.* 2016;37:2129–2200. doi: 10.1093/eurheartj/ehw128

19. Savarese G, Stolfo D, Sinagra G, Lund LH. Heart failure with mid-range or mildly reduced ejection fraction. *Nat Rev Cardiol.* 2022;19:100–116. doi: 10.1038/s41569-021-00605-5
20. McDonagh TA, Metra M, Adamo M, Gardner RS, Baumbach A, Böhm M, Burri H, Butler J, Čelutkienė J, Chioncel O, et al. 2021 ESC Guidelines for the diagnosis and treatment of acute and chronic heart failure. *Eur Heart J.* 2021;42:3599–3726. doi: 10.1093/eurheartj/ehab368
21. Binder H, Schumacher M. Allowing for mandatory covariates in boosting estimation of sparse high-dimensional survival models. *BMC Bioinformatics.* 2008;9:14. doi: 10.1186/1471-2105-9-14
22. Davis J, Goadrich M. The relationship between precision-recall and ROC curves. 2006.
23. Lundberg S, Lee S-I. A Unified Approach to Interpreting Model Predictions. 2017.
24. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *BMJ.* 2015;350:g7594. doi: 10.1136/bmj.g7594
25. Zhang Y, Gao C, Greene SJ, Greenberg BH, Butler J, Yu J, Zheng Z, Ma G, Wang L, Yang P, et al. Clinical performance and quality measures for heart failure management in China: the China-Heart Failure registry study. *ESC Heart Fail.* 2023;10:342–352. doi: 10.1002/ehf2.14184
26. Khanam SS, Choi E, Son JW, Lee JW, Youn YJ, Yoon J, Lee SH, Kim JY, Ahn SG, Ahn MS, et al. Validation of the MAGGIC (Meta-Analysis Global Group in Chronic Heart Failure) heart failure risk score and the effect of adding natriuretic peptide for predicting mortality after discharge in hospitalized patients with heart failure. *PLoS One.* 2018;13:e0206380. doi: 10.1371/journal.pone.0206380
27. Voors AA, Ouwerkerk W, Zannad F, van Veldhuisen DJ, Samani NJ, Ponikowski P, Ng LL, Metra M, Ter Maaten JM, Lang CC, et al. Development and validation of multivariable models to predict mortality and hospitalization in patients with heart failure. *Eur J Heart Fail.* 2017;19:627–634. doi: 10.1002/ejhf.785
28. Park J, Hwang IC, Yoon YE, Park JB, Park JH, Cho GY. Predicting long-term mortality in patients with acute heart failure by using machine learning. *J Card Fail.* 2022;28:1078–1087. doi: 10.1016/j.cardfail.2022.02.012
29. Polo Friz H, Esposito V, Marano G, Primitz L, Bovio A, Delgrossi G, Bombelli M, Grignaffini G, Monza G, Boracchi P, et al. Machine learning and LACE index for predicting 30-day readmissions after heart failure hospitalization in elderly patients. *Intern Emerg Med.* 2022;17:1727–1737. doi: 10.1007/s11739-022-02996-w
30. 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. doi: 10.2196/38082
31. Li F, Xin H, Zhang J, Fu M, Zhou J, Lian Z. Prediction model of in-hospital mortality in intensive care unit patients with heart failure: machine learning-based, retrospective analysis of the MIMIC-III database. *BMJ Open.* 2021;11:e044779. doi: 10.1136/bmjopen-2020-044779
32. Glinghammar B, Rafter I, Lindström AK, Hedberg JJ, Andersson HB, Lindblom P, Berg AL, Cotgreave I. Detection of the mitochondrial and catalytically active alanine aminotransferase in human tissues and plasma. *Int J Mol Med.* 2009;2:621–631. doi: 10.3892/ijmm_00000173
33. Palazzuoli A, Ruocco G, Evangelista I, De Vivo O, Nuti R, Ghio S. Prognostic significance of an early echocardiographic evaluation of right ventricular dimension and function in acute heart failure. *J Card Fail.* 2020;26:813–820. doi: 10.1016/j.cardfail.2020.01.002
34. Zhang Y, Zhang J, Butler J, Yang X, Xie P, Guo D, Wei T, Yu J, Wu Z, Gao Y, et al. Contemporary epidemiology, management, and outcomes of patients hospitalized for heart failure in China: results from the China Heart Failure (China-HF) Registry. *J Card Fail.* 2017;23:868–875. doi: 10.1016/j.cardfail.2017.09.014

Supplemental material for Machine learning for mortality prediction in patients with heart failure with mildly reduced ejection fraction

Table S1. Enrolled features in this study

	Feature name
Features with missing value ⩽50% (enrolled features for analysis, 82 features)	Age, gender, BMI, BSA, HR, SBP, DBP, NYHA class, NT-proBNP, PT, APTT, TT, FIB, RBC, hemoglobin, HCT, RDW, RDW-SD, WBC, N, L, PLT, TP, ALB, ALT, AST, ALP, GGT, TB, DB, potassium, sodium, chlorine, calcium, phosphorus, glucose, Cr, BUN, eGFR, UA, CK, LDH, apolipoprotein, hs-CRP, TG, TC, HDL, LDL-T, LDL-C, ESR, CRP, FT3, FT4, TT3, TT4, TSH, LAD, LVEDD, RVEDD, LVPW, IVS, LVEF, ventricular aneurysm, ventricular thrombus, ventricular wall abnormal movement, hypertension, diabetes mellitus, IHF, anemia, secondary pulmonary hypertension, AF, MVR, MVS, AVR, AVS, TVR, ACEI/ARB treatment, ACEI treatment, ARB treatment, β-blocker treatment, MRA treatment, diuretic treatment
'All-in' strategy (67 features)	Age, gender, HR, BMI, SBP, DBP, NYHA class, NT-proBNP, NLR, PLT, hemoglobin, RDW, TP, ALB, AST/ALT, ALP, GGT, TB, potassium, sodium, chlorine, calcium, phosphorus, glucose, Cr/BUN, UA, CK, LDH, Apolipoprotein, hs-CRP, TG, TC, HDL, ESR, PT, APTT, TT, FIB, FT3, FT4, TT3, TT4, TSH, LAD, LVEDD, LVPW, IVS, LVEF, RVEDD, left ventricular aneurysm, ventricular thrombus, ventricular wall abnormal movement, MVR, TVR, AVR, MVS, AVS, AF, anemia, hypertension, diabetes mellitus, IHF, ACEI/ARB treatments, β-blocker treatment, MRA treatment, diuretic treatment
CoxBoost strategy without 10-fold cross-validation (46 features)	Age, SBP, PT, NLR, RDW, ALB, ALP, TB, UA, CK, LDH, hs-CRP, TT4, NT-proBNP, IVS, LVEF, ventricular wall abnormal movement, MVR, TVR, diabetes mellitus, IHF, Diuretic treatment, HR, NYHA class, TT, PLT, TP, AST/ALT, GGT, sodium, chlorine, Glu, Cr/BUN, apolipoprotein, TG, TC, HDL, FT3, TT3, LAD, LVPW, ventricular thrombus, AVR, hypertension, ACEI/ARB treatment, β-blocker treatment
CoxBoost strategy after 10-fold cross-validation ('CBCV' strategy,17 features)	Age, NLR, hemoglobin, RDW, AST/ALT, Na, UA, LDH, hs-CRP, TC, FT3, NT-proBNP, TVR, hypertension, DM, IHD, ACEI/ARB treatment
Metric Logistic model (14 features)	Age, gender, SBP, sodium, eGFR, hemoglobin, NYHA class, anemia, diabetes mellitus, NT-proBNP, IHF, LVEF, ACEI/ARB treatment, β-blocker treatment

Abbreviations: HR = heart rate; BMI = body mass index; BSA = body surface area; SBP = systolic blood pressure; DBP = diastolic blood pressure; NYHA = New York Heart Association; PT = prothrombin time; APTT = activated partial thromboplastin time; TT = thrombin time; FIB = fibrinogen; WBC = white blood cell; N = neutrophil; L = lymphocyte; RBC = red blood cell; RDW = red blood cell distribution; RDW-SD = red blood cell distribution standard division; PLT = platelet; Hb = hemoglobin; HCT = hematocrit; TP = total protein; ALB = albumin; AST = aspartate aminotransferase; ALT = alanine transaminase; ALP = alkaline phosphatase; GGT= gamma-glutamyl transpeptidase; TB = total bilirubin; DB = direct bilirubin; eGFR = estimated glomerular filtration rate; Cr = creatinine; BUN= blood urea nitrogen; CK = creatine kinase; LDH = lactate dehydrogenase; Apo A = apolipoprotein; HS-CRP = high-sensitivity C-reaction protein; TG = triglyceride; TC = total cholesterol; HDL = high-density lipoprotein; LDL-T = low-density lipoprotein transporter; LDL-C = low-density lipoprotein cholesterol; ESR = erythrocyte sedimentation rate; CRP = C-reaction protein; FT3 = free triiodothyronine; FT4 = free thyroxine; TT3 = total triiodothyronine; TT4 = total thyroxine; TSH = thyroid stimulating hormone; LADD = left atrial diastolic diameter; LVEDD = left ventricular end-diastolic diameter; LVPW = left ventricular posterior wall; IVS = interventricular septum; LVEF = left ventricular ejection fraction; RVEDD = right ventricular end-diastolic diameter; VAN = left ventricular anerysum; THRO = left ventricular thrombus; LVAM = left ventricular abnormal movement; MVR = mitral valve regurgitation; TVR = tricuspid

valve regurgitation; AVR = aortic valve regurgitation; MSS = mitral valve stenosis; ASS = aotric valve stenosis; PAH = pulmonary artery hypertrnsion; Af = atrial fibrillation; HBP = hypertension; DM = diabetes mellitus; IHF = ischemic heart failure; ACEI = angiotensin-converting enzyme inhibitor; ARB = angiotensin receptor blocker; MRA = mineralocorticoid receptor antagonist; DM = Diabetes Mellius; MVR = Mitral valve regurgitation; HR = heart rate

Table S2 Characteristics of enrolled features with missing values ≤50%

Baseline Characteristics	HFmrEF Patients (N=424)
Demographic characteristics	
Age, years	59.34±14.25
Males, n (%)	302 (71.2%)
Vital signs	
Body mass index, kg/m^2	24.87±4.47
Body surface area, mm/m^2	1.78±0.23
Heart rate, bpm	76 [66,88]
Systolic pressure (mmHg)	123.41±20.73
Diastolic pressure (mmHg)	73.53±13.40
NYHA function class, (%)	
I	10 (2.4%)
II	121 (28.5%)
III	201 (47.4%)
IV	92 (21.7%)
Laboratory results	
NT-proBNP, ng/ml	1701.8 [695.6,3779.2]
Prothrombin time, s	13.90 [13.30, 14.90]
Activated partial thromboplastin time, s	37.70 [34.77, 41.50]
Thrombin time, s	16.30 [15.60, 17.20]
Fibrinogen, g/L	3.54 [2.99, 4.23]
Red blood cell, x10^{12}/L	4.53±0.72
Hematocrit, %	40.93±6.15
Hemoglobin, g/L	136.55±22.49
Red blood cells wide distribution, %	13.4 (12.7,14.53)
Red blood cells wide distribution standard division, fl	44 (41.4,47.58)
White blood cell, x10^9/L	6.67 (5.53,8.21)
Neutrophil, x10^9/L	4.30 (3.37,5.52)
Lymphocyte, x10^9/L	1.63 (1.29,2.13)
Platelet, x10^9/L	203.85±77.11
Total Protein, g/L	68.18±6.79
Albumin, g/L	39.90±4.76
Alanine transaminase, IU/L	21 (14,33)
Aspartate aminotransferase, IU/L	22 (17,30)
Alkaline phosphatase, IU/L	68 (55,88)
Gamma-glutamyl transpeptidase, IU/L	39 (26,70)
Total bilirubin, μmol/L	18.2 (13.35,26.3)
Direct bilirubin, μmol/L	3.6 (2.4,5.55)
K, mmol/L	4.05±0.52
Na, mmol/L	137.70±3.94
Cl, mmol/L	102.60±4.58
Ca, mmol/L	2.28 (2.18,2.37)
P, mmol/L	1.28±0.26
Glucose, mmol/L	5.30 (4.70,6.54)
Creatine, μmol/L	94.5 (77.71,116.31)

Baseline Characteristics	HFmrEF Patients (N=424)
Blood urea nitrogen, mmol/L	7.70 (5.78,10.28)
eGFR, ml/min·1.73m^2	70.50±25.42
URIC, μmol/L	445.24±143.24
Creatine kinase, IU/L	62 (41,92)
Lactate dehydrogenase, IU/L	187 (159,233.5)
Apolipoprotein, mg/L	174.1 (80.02,371.79)
High-sensitivity C-reaction protein, mg/L	3.26 (1.43,8.95)
Triglyceride, mmol/L	1.44 (1.03,1.97)
Total cholesterol, mmol/L	4.11 (3.39,4.75)
High-density lipoprotein, mmol/L	0.97 (0.82,1.23)
Low-density lipoprotein T, mmol/L	2.31 (1.73,2.87)
Low-density lipoprotein cholesterol, mmol/L	2.42 (1.87,2.98)
Erythrocyte sedimentation rate, mm/h	8 (3,19)
C-reactive protein, mg/L	4.27 (2.21,8.75)
Free triiodothyronine, pg/ml	2.37 (2.68,2.97)
Free thyroxine, ng/ml	1.19 (1.08,1.37)
Total triiodothyronine, ng/ml	0.92 (0.75,1.13)
Total thyroxine, μg/ml	7.94 (6.60, 9.20)
Thyroid Stimulating Hormone, μIU/ml	1.91 (1.14, 2.94)
Echocardiography	
Left atrial diastolic diameter, mm	44 (40,49)
Left ventricular end-diastolic diameter, mm	60.36±8.69
Right ventricular end-diastolic diameter, mm	23 (20,25)
Left ventricular posterior wall, mm	8.15 (9.7,10)
Interventricular septum, mm	10 (9,11)
Left ventricular ejection fraction, n (%)	43.21±2.99
Ventricular Aneurysm, n (%)	24 (5.7%)
Ventricular Thrombus, n (%)	8 (1.9%)
Ventricular Wall Abnormal Movement, n (%)	344 (81.1%)
Comorbidities, n (%)	
Hypertension	232 (54.7%)
Diabetes mellitus	134 (31.6%)
Ischemic Heart Failure	184 (43.4%)
Anemia	70 (16.5%)
Secondary Pulmonary Hypertension	64 (15.1%)
Atrial Fibrillation	142 (33.5%)
Mitral Valve Regurgitation	134 (31.6%)
Mitral Valve Stenosis	15 (3.5%)
Aortic Valve Regurgitation	52 (12.3%)
Aortic Valve Stenosis	25 (5.9%)
Tricuspid Valve Regurgitation	93 (21.9%)
Guide directed medicine treatments, n (%)	
ACEIs or ARBs	280 (66.0%)
Angiotensin-converting enzyme inhibitors	223 (52.6%)
Angiotensin receptor blockers	57 (13.4%)
β-blockers	377 (88.9%)
MRAs	318 (75.0%)
Diuretics	366 (86.3%)

ACEI = angiotensin-converting enzyme inhibitor; ARB = angiotensin receptor blocker; MRA = Mineralocorticoid receptor antagonists

Table S3. Results of MissForest data imputation

Missforest Imputation results (OBB error)	NRMSE	PFC
	0.021	0.036

OBB error =out of box error NRMSE = normalized root-mean-square error PFC = roportion of falsely classified

Table S4. Results of data before and after imputation

Characteristics	Missing rate (%)	Number of Patients	Before Imputation	After Imputation	P value
Body mass index, kg/m^2	6.1%	26	24.87±4.47	24.87±4.33	0.287
Body surface area, mm/m^2	6.1%	26	1.78±0.23	1.78±0.22	0.271
Fibrinogen*, g/L	14.4%	61	3.53 (2.91,4.29)	3.54 (2.99,4.23)	0.755
TT*, s	15.3%	65	16.1 (15.4,17.0)	16.3 (15.60,17.20)	0.06
ALP, IU/L	4%	17	68 (55,88)	68 (55.25,86.75)	0.996
GGT, IU/L	4.2%	18	39 (26,70)	39 (26,68.75)	0.998
Ca, mmol/L	4%	17	2.28 (2.17,2.37)	2.28 (2.19,2.37)	1.000
P, mmol/L	4.7%	20	1.28±0.26	1.28±0.25	0.396
Glucose, mmol/L	2.8%	12	5.30 (4.70,6.54)	5.30 (4.72,6.52)	1.000
Apo A, mg/L	4%	17	174.1 (80.02,371.79)	174.1 (82.51,365.61)	1.000
HSCRP, mg/L	3.8%	16	3.26 (1.43,8.95)	3.33 (1.45,8.67)	0.945
TG, mmol/L	2.6%	11	1.44 (1.03,1.97)	1.44(1.04,1.94)	0.996
TC, mmol/L	2.6%	11	4.11 (3.39,4.75)	4.11 (3.40,4.74)	0.999
HDL, mmol/L	2.6%	11	1.03±0.31	1.03±0.31	0.715
LDLT, mmol/L	5.2%	22	2.31 (1.73,2.87)	2.31 (1.75,2.84)	0.997
LDLC, mmol/L	4%	17	2.42 (1.87,2.98)	2.42 (1.91,2.94)	1.000
FT3, pg/ml	4.2%	18	2.68 (2.37,2.97)	2.68 (2.37,2.94)	1.000
FT4, ng/ml	4.2%	18	1.19 (1.08,1.37)	1.19 (1.08,1.37)	0.981
TT3, pg/ml	4.2%	18	0.92 (0.74,1.13)	0.92 (0.77,1.13)	0.988
TT4, ng/ml	4.2%	18	7.94 (6.60,9.20)	7.94 (6.66,9.18)	1.000
TSH, ng/ml	4.2%	18	1.91 (1.11,3.16)	1.91 (1.14,2.95)	0.998
NT-proBNP*, pg/ml	14.4%	61	1701.8 (695.6,3779.2)	1712.5 (815.4,3530.80)	0.770
LVPW, mm	9.2%	39	9.7 (8.15,10)	9.7 (9,10)	0.995
IVS, mm	7.8%	33	10 (9,11)	10 (9,11)	0.768
RVEDD*, mm	15.09%	64	23 (20,25)	23 (21,25)	0.931
Left Ventricular Aneurysm*, n (%)	14.6%	62	24 (6.6%)	24 (5.7%)	0.572
Ventricular Thrombus*, n (%)	14.6%	62	8 (2.2%)	8 (1.9%)	0.749
Ventricular Wall Movement Abnormalities*, n (%)	14.6%	62	344 (95.0%)	406 (95.8%)	0.627
Mitral Valve Regurgitation*, n (%)	14.6%	62	134 (34.6%)	134 (31.6%)	0.361
Tricuspid Valve Regurgitation*, n (%)	14.6%	62	93 (24.0%)	93 (21.9%)	0.478
Aortic Valve Regurgitation*, n (%)	14.6%	62	52 (13.4%)	52 (12.3%)	0.618

Characteristics	Missing rate (%)	Number of Patients	Before Imputation	After Imputation	P value
Mitral Valve Stenosis*, n (%)	14.6%	62	15 (3.9%)	15 (3.5%)	0.799
Aortic Valve Stenosis*, n (%)	14.6%	62	25 (6.5%)	5.9 (1.8%)	0.739

*Imputated with MissForest
Footnote as Table S1.

Table S5. Results of feature multicollinearity

Features	Tolerance	VIF
Gender	.344	2.910
AGE	.272	3.675
HR	.558	1.793
BMI	.176	5.677
BSA	.150	6.688
SBP	.398	2.514
DBP	.386	2.590
NYHA	.713	1.402
PT	.580	1.725
APTT	.471	2.125
TT	.663	1.509
FIB	.358	2.792
WBC	.006	155.018
N	.009	117.575
L	.051	19.683
PLT	.510	1.962
RBC	.024	41.581
Hb	.037	26.745
HCT	.028	36.340
RDW_SD	.064	15.672
RDW	.054	18.378
TP	.374	2.676
ALB	.357	2.801
ALT	.194	5.144
AST	.119	8.434
ALP	.493	2.029
GGT	.468	2.138
TB	.159	6.272
DB	.136	7.370
K	.616	1.622

Features	Tolerance	VIF
Na	.548	1.826
Cl	.591	1.691
Ca	.715	1.400
P	.665	1.504
Glucose	.447	2.237
CR	.117	8.526
eGFR	.127	7.873
BUN	.232	4.316
URIC	.472	2.121
CK	.581	1.721
LDH	.243	4.107
Apo (A)	.732	1.366
Hs-CRP	.371	2.694
TG	.031	32.301
TC	.009	110.873
HDL	.103	9.671
LDLT	.015	68.845
LDLC	.046	21.945
ESR	.226	4.418
CRP	.391	2.557
FT3	.117	8.561
FT4	.163	6.142
TT3	.195	5.122
TT4	.303	3.300
TSH	.807	1.239
NT-proBNP	.470	2.128
LADD	.518	1.931
LVEDD	.506	1.978
LVPW	.457	2.189
IVS	.510	1.959
LVEF	.757	1.322
RVEDD	.569	1.757
VAN	.785	1.273
THRO	.788	1.269
LVAM	.825	1.213
MVR	.513	1.949
TVR	.451	2.216
AVR	.573	1.746
MSS	.690	1.449

Features	Tolerance	VIF
ASS	.656	1.524
PAH	.542	1.845
Af	.547	1.828
Anemia	.418	2.392
HBP	.584	1.713
DM	.519	1.925
IHF	.459	2.178
ARB	.754	1.326
ACEI/ARB	.551	1.815
β-Blockers	.696	1.436
MRA	.637	1.570
Diuretics	.827	1.209

*ACEI was excluded during calculation.
Footnote as Table S1.

Table S6. Principle of the machine-learning algorithm and the hyperparameter

Machine-learning algorithm	Principle	Best hyperparameter	
		'All-in'	'CBCV'
XGBoost	XGBoost is a scalable end-to-end tree-boosting system that retrofits the tree-boosting algorithm in handling sparse data by weight quantile sketch for approximate learning and introduces the column block for parallel learning(1).	mtry=3 min_n=9 tree_depth=2 learn_rate=0.0463 loss_reduction=0.746	mtry=3 min_n=9 tree_depth=2 learn_rate=0.0463 loss_reduction=0.746
RF	RF is a special kind of bagging method integrating multiple decision tree results. It can effectively avoid the underlying over-fitting caused by the decision tree prediction and increase prediction accuracy(2).	mtry=6 n_tree=500 min_n=70	mtry=6 n_tree=250 min_n=70
SVM	SVM is a supervised learning algorithm used for classification and regression analysis. Its main idea is to create a hyperplane in a high-dimensional space that separates the data into two or more classes(3).	cost=9.88 sigma=0.000584	cost=2.67 sigma=0.00244

*All hyperparameters were determined according to the best accuracy by 5-fold cross-validation.

Abbr: XGBoost= eXtreme Gradient Boosting; RF= random forest; SVM = support vector machine

1. Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System2016. 785-94 p.

2. Chen C, Breiman L. Using Random Forest to Learn Imbalanced Data. University of California, Berkeley. 2004.

3. Suthaharan S. Support Vector Machine. In: Suthaharan S, editor. Machine Learning Models and Algorithms for Big Data Classification: Thinking with Examples for Effective Learning. Boston, MA: Springer US; 2016. p. 207-35.

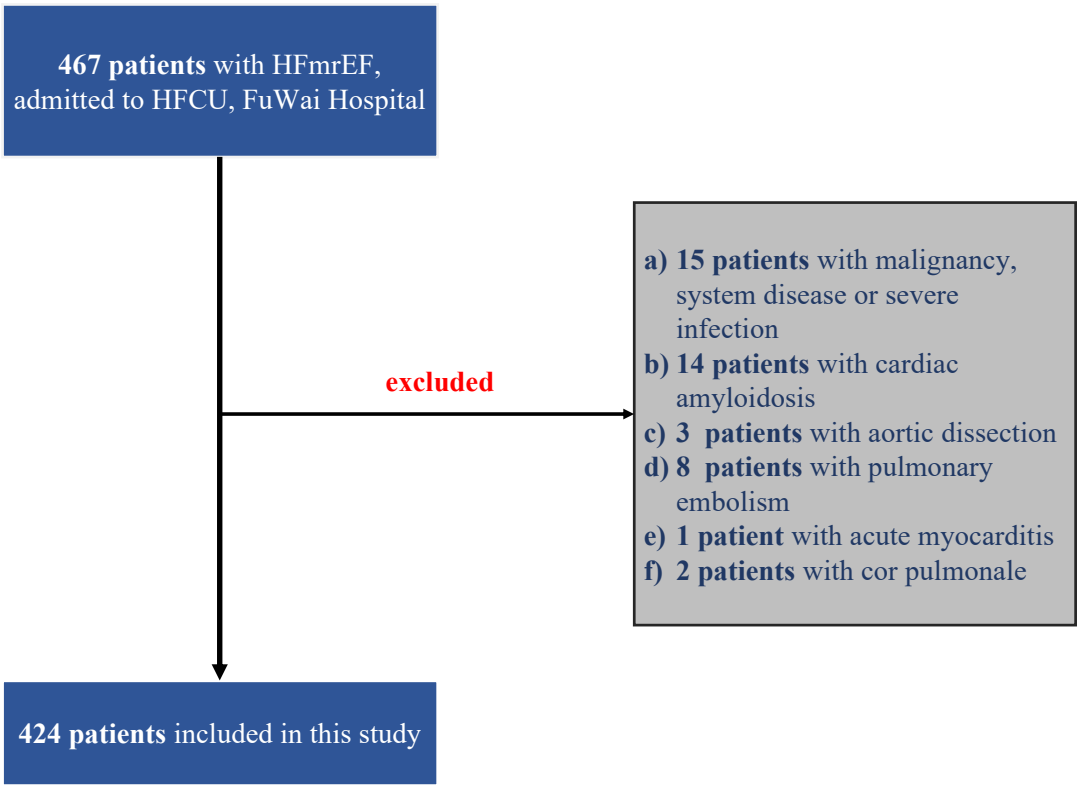
Table S7. Feature importance of the 'all-in' XGBoost model

Feature	Gain	Cover	Frequency	Importance
AGE	0.069522	0.052148	0.048359	0.069522
AST/ALT	0.051782	0.036465	0.037997	0.051782
NA	0.040745	0.026595	0.025907	0.040745
K	0.038405	0.030378	0.029361	0.038405
NT-proBNP	0.038001	0.035648	0.036269	0.038001
Hb	0.037056	0.027296	0.02418	0.037056
RDW	0.036577	0.026126	0.02418	0.036577
HDL	0.031691	0.031899	0.031088	0.031691
LDH	0.031509	0.025529	0.02418	0.031509
APTT	0.030796	0.033928	0.032815	0.030796
Ca	0.029934	0.029225	0.027634	0.029934
RVEDD	0.028859	0.028591	0.025907	0.028859
P	0.027944	0.032431	0.031088	0.027944
Cl	0.02291	0.017559	0.017271	0.02291
FT3	0.022775	0.017754	0.017271	0.022775
SBP	0.020688	0.021645	0.020725	0.020688
HSCRP	0.020446	0.017471	0.017271	0.020446
TP	0.020304	0.025726	0.025907	0.020304
LVEF	0.020277	0.021644	0.02418	0.020277
GGT	0.018771	0.020573	0.020725	0.018771
PLT	0.018249	0.015681	0.017271	0.018249
TG	0.016874	0.017021	0.018998	0.016874
TT	0.016774	0.020722	0.020725	0.016774
CB	0.01664	0.015688	0.013817	0.01664
FT4	0.016537	0.014398	0.017271	0.016537
Glucose	0.016286	0.016612	0.017271	0.016286
TB	0.015409	0.016572	0.017271	0.015409
TT4	0.015294	0.015647	0.017271	0.015294
TSH	0.0146	0.018139	0.020725	0.0146
LVPW	0.014561	0.021723	0.020725	0.014561
FIB	0.013907	0.01992	0.020725	0.013907
TC	0.0129	0.015251	0.013817	0.0129
HR	0.012343	0.013734	0.013817	0.012343
Hypertension	0.012281	0.017121	0.015544	0.012281
Ischemic heart failure	0.012012	0.014918	0.017271	0.012012
NLR	0.01176	0.017612	0.017271	0.01176
TT3	0.011226	0.017757	0.017271	0.011226
BMI	0.011222	0.009908	0.01209	0.011222

Feature	Gain	Cover	Frequency	Importance
PT	0.010994	0.015575	0.015544	0.010994
Apo (A)	0.010907	0.01517	0.015544	0.010907
ALB	0.010371	0.011897	0.013817	0.010371
URIC	0.010237	0.014561	0.017271	0.010237
ALP	0.008924	0.00991	0.01209	0.008924
DM	0.008449	0.012385	0.010363	0.008449
LVEDD	0.008077	0.010562	0.01209	0.008077
CK	0.007475	0.01105	0.01209	0.007475
ESR	0.007364	0.010761	0.010363	0.007364
DBP	0.006237	0.011795	0.010363	0.006237
ACEI/ARB Treatment	0.005219	0.006542	0.005181	0.005219
LADD	0.002831	0.002806	0.003454	0.002831
MVR	0.001996	0.003158	0.003454	0.001996
Gender	0.001497	0.001625	0.001727	0.001497
MRA	0.001074	0.003374	0.003454	0.001074
IVS	0.000483	0.001771	0.001727	0.000483

Footnote as Table S1.

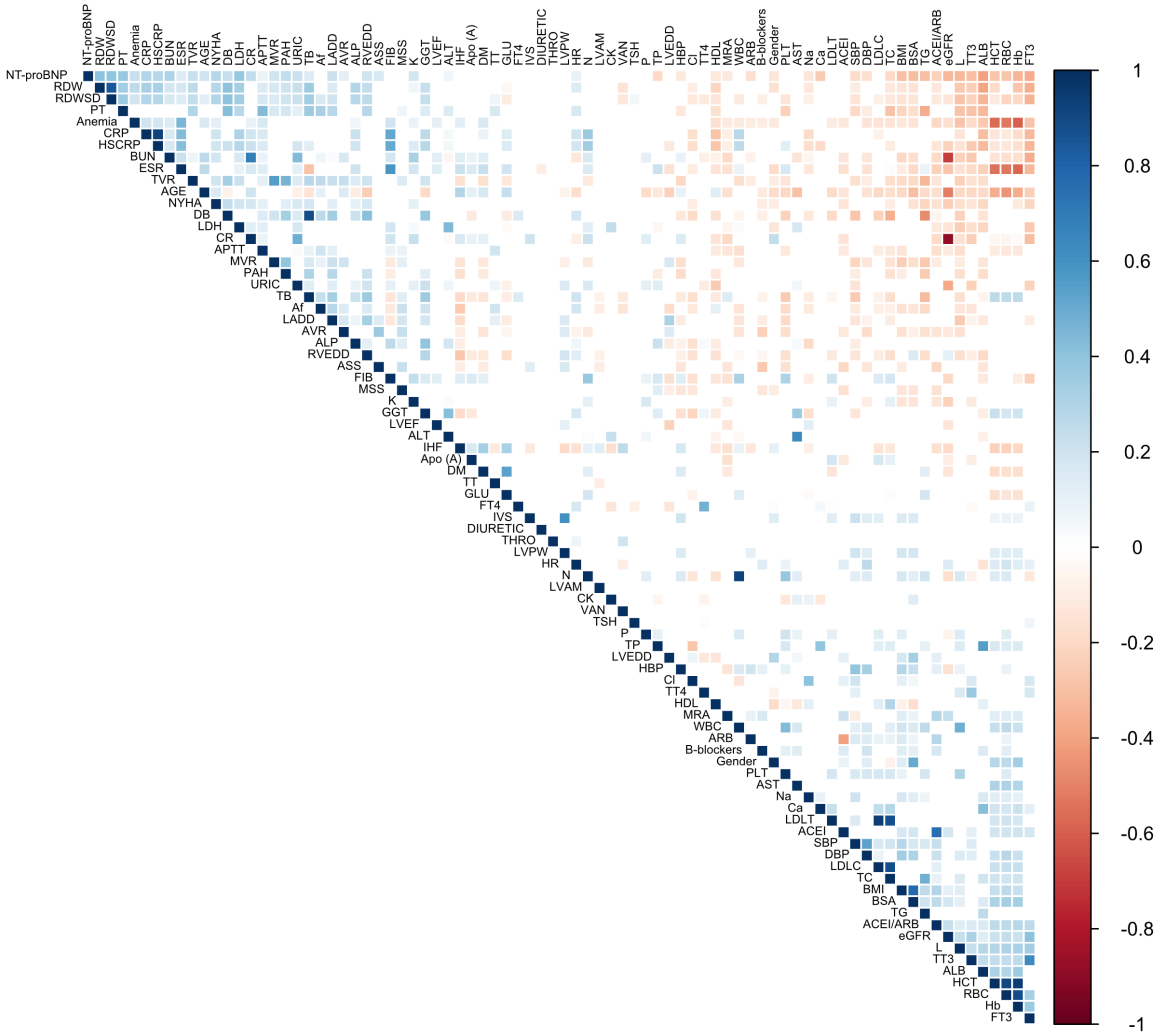
Figure S1 Flow chart of the study



HFmrEF = heart failure with mildly reduced ejection fraction; HFCU = heart failure care unit

A total of 467 patients were enrolled in this study, with 43 patients met the exclusion criteria, and a total of 424 patients finally enrolled.

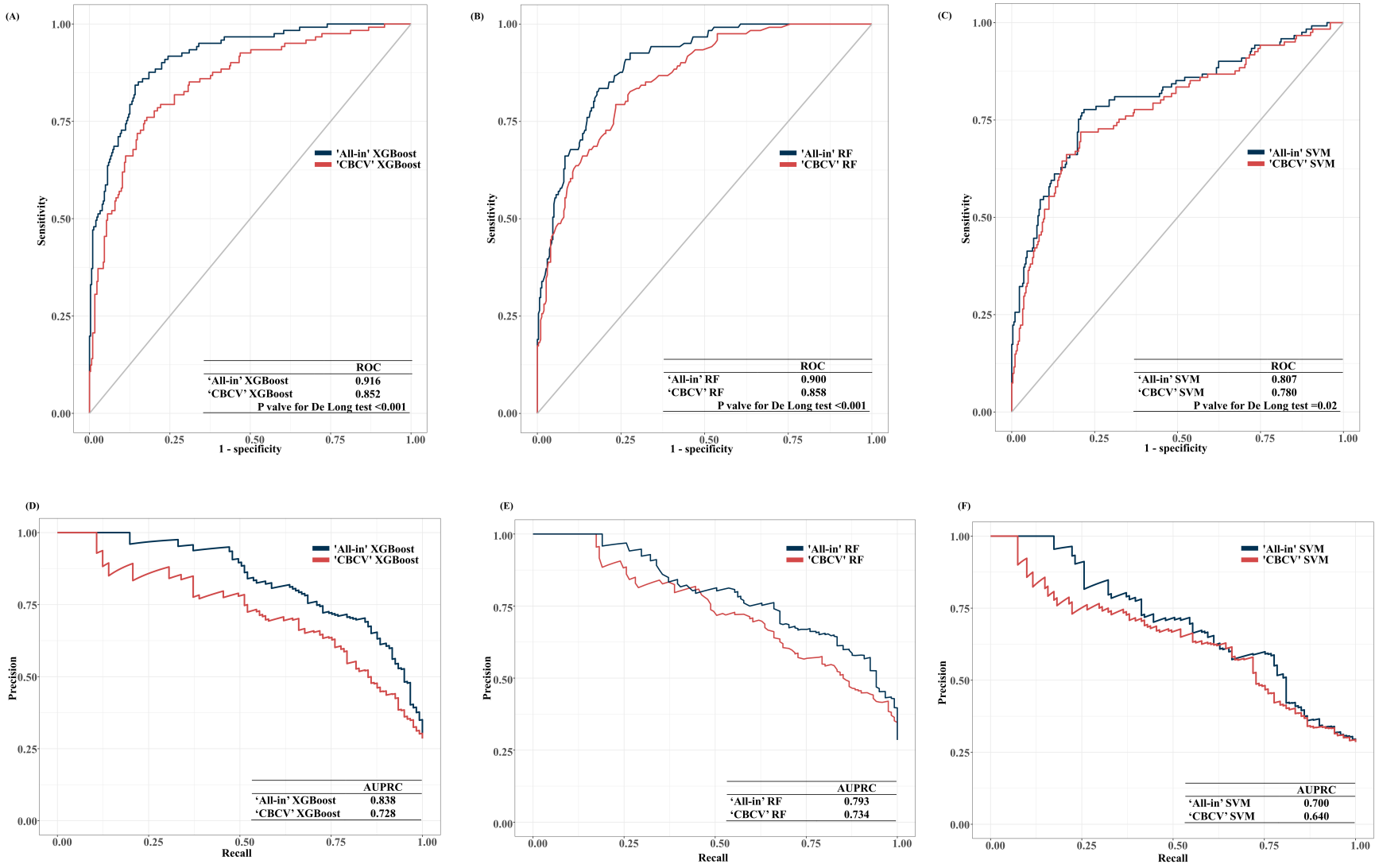
Figure S2. Correlation heatmap between eighty-two recruited features

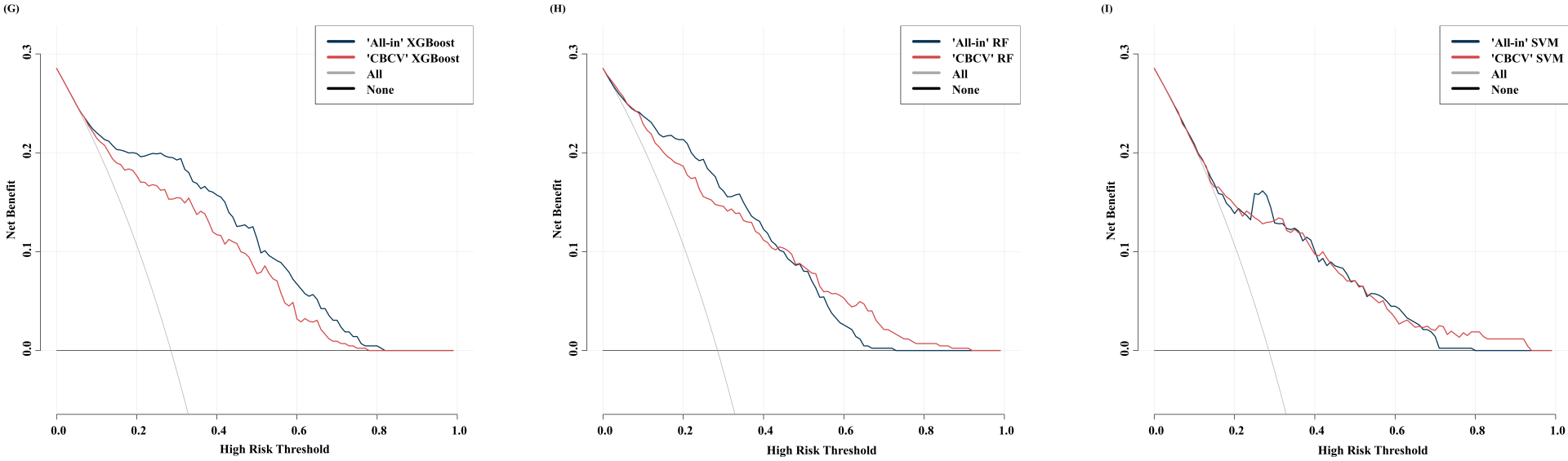


As shown in the figure, Features with strong correlations were in dark blue (positive correlation) or dark red (negative correlation). Such as, the BMI and BSA had a strong positive correlation (Spearman coefficient ρ= 0.815, P=0.000), the Cr and eGFR had a strong negative correlation (Spearman coefficient ρ=-0.885, P=0.000)

HR = heart rate; BMI = body mass index; BSA = body surface area; SBP = systolic blood pressure; DBP = diastolic blood pressure; NYHA = New York Heart Association; PT = prothrombin time; APTT = activated partial thromboplastin time; TT = thrombin time; FIB = fibrinogen; WBC = white blood cell; N = neutrophil; L = lymphocyte; RBC = red blood cell; RDW = red blood cell distribution; RDW-SD = red blood cell distribution standard division; PLT = platelet; Hb = hemoglobin; HCT = hematocrit; TP = total protein; ALB = albumin; AST = aspartate aminotransferase; ALT = alanine transaminase; ALP = alkaline phosphatase; GGT= gamma-glutamyl transpeptidase; TB = total bilirubin; DB = direct bilirubin; eGFR = estimated glomerular filtration rate; Cr = creatinine; BUN= blood urea nitrogen; CK = creatine kinase; LDH = lactate dehydrogenase; Apo A = apolipoprotein; HS-CRP = high-sensitivity C-reaction protein; TG = triglyceride; TC = total cholesterol; HDL = high-density lipoprotein; LDL-T = low-density lipoprotein transporter; LDL-C = low-density lipoprotein cholesterol; ESR = erythrocyte sedimentation rate; CRP = C-reaction protein; FT3 = free triiodothyronine; FT4 = free thyroxine; TT3 = total triiodothyronine; TT4 = total thyroxine; TSH = thyroid stimulating hormone; LADD = left atrial diastolic diameter; LVEDD = left ventricular end-diastolic diameter; LVPW = left ventricular posterior wall; IVS = interventricular septum; LVEF = left ventricular ejection fraction; RVEDD = right ventricular end-diastolic diameter; VAN = left ventricular anerysum; THRO = left ventricular thrombus; LVAM = left ventricular abnormal movement; MVR = mitral valve regurgitation; TVR = tricuspid valve regurgitation; AVR = aortic valve regurgitation; MSS = mitral valve stenosis; ASS = aotric valve stenosis; PAH = pulmonary artery hypertrnsion; Af = atrial fibrillation; HBP = hypertension; DM = diabetes mellitus; IHF = ischemic heart failure; ACEI = angiotensin-converting enzyme inhibitor; ARB = angiotensin receptor blocker; MRA = mineralocorticoid receptor antagonist; DM = Diabetes Mellius; MVR = Mitral valve regurgitation

Figure S3. Performance Comparison between ‘All-in’ MLBPMs and ‘CBCV’ MLBPMs

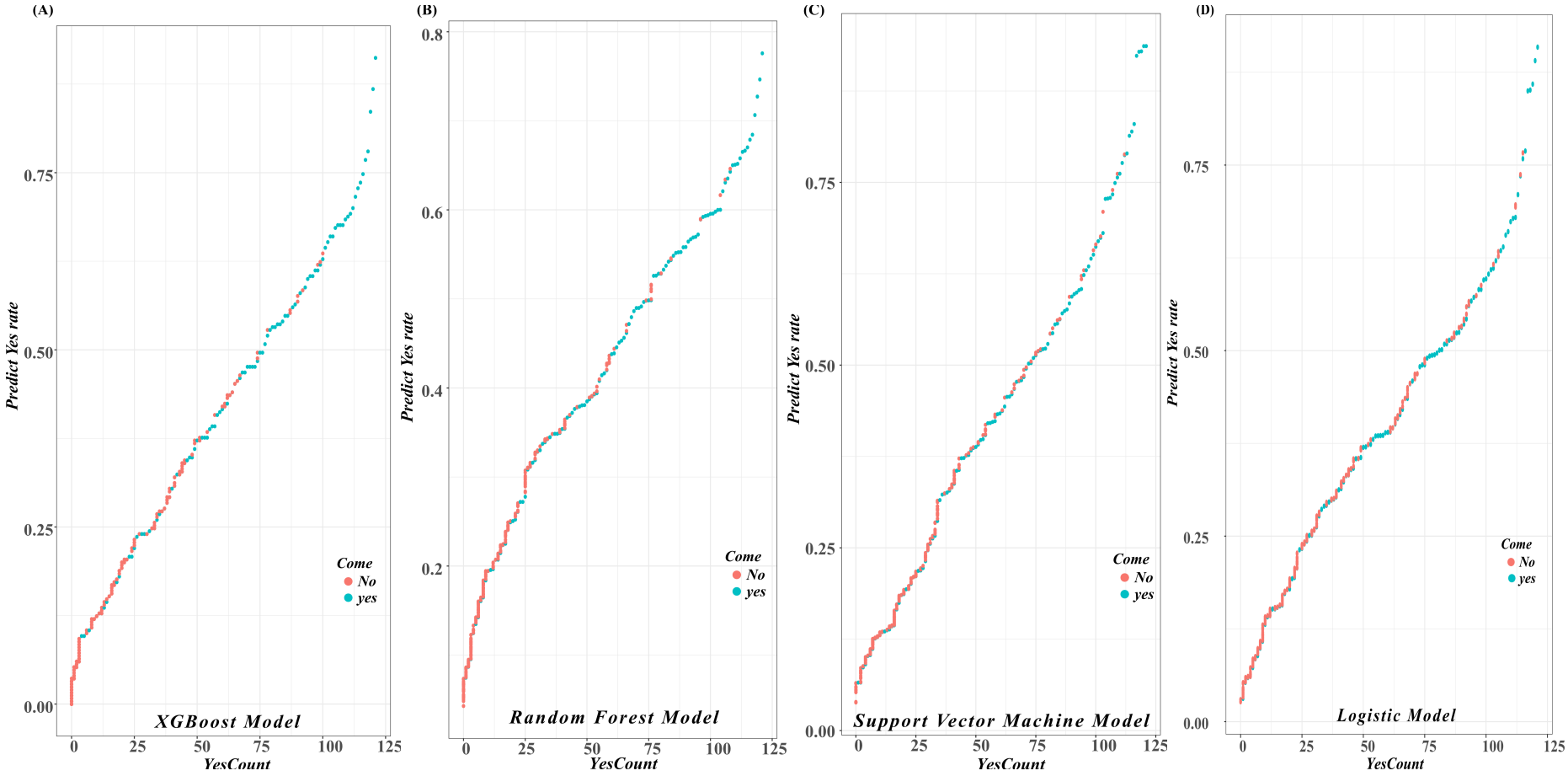




As shown in this figure, model performance declined in 'CBCV' MLBPMs compared with the corresponding 'All-in' MLBPMs, with lower ROC-AUC (A-C) and AUPRC (D-F). Decision curve analyses showed that the clinical performance of the 'CBCV' XGBoost model (G) was inferior to that of the 'All-in' model, and the 'CBCV' RF model were superior to the 'All-in' model (H). The clinical net benefits were similar in the 'All-in' and the 'CBCV' SVM model (I).

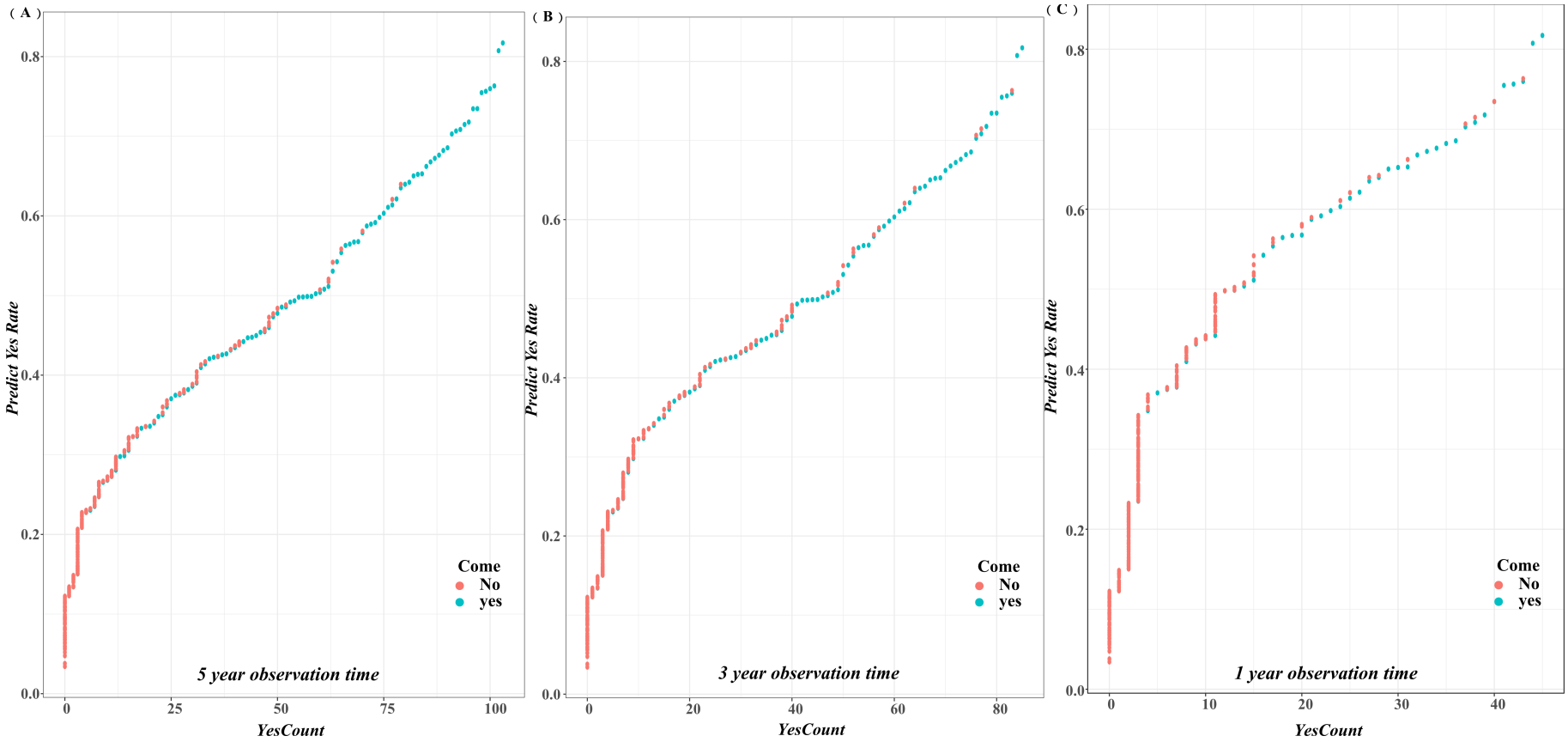
Abbreviations: RF = random forest; SVM = support vector machine

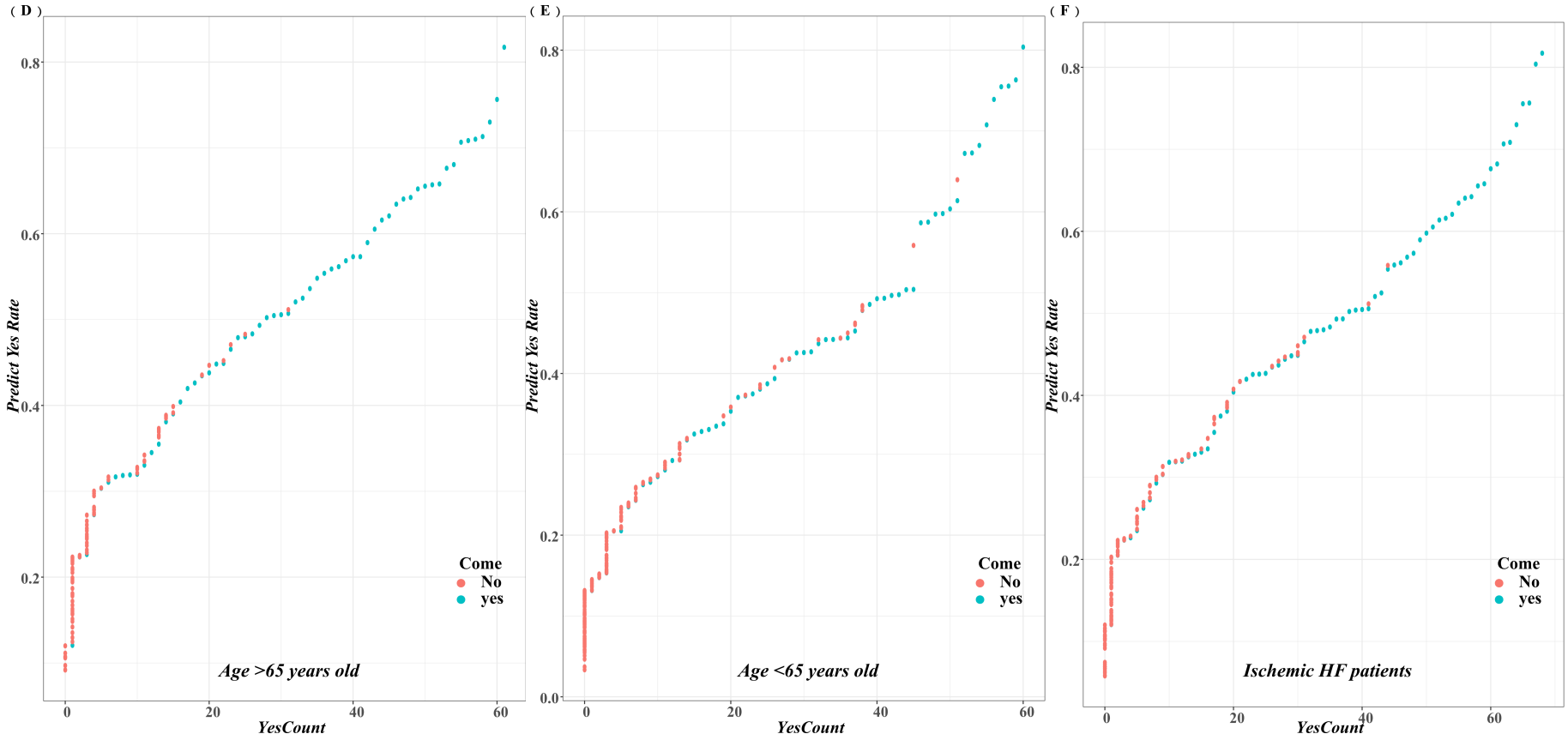
Figure S4. Prediction distribution plots of the 'CBCV' MLBPMs and reference logistic model

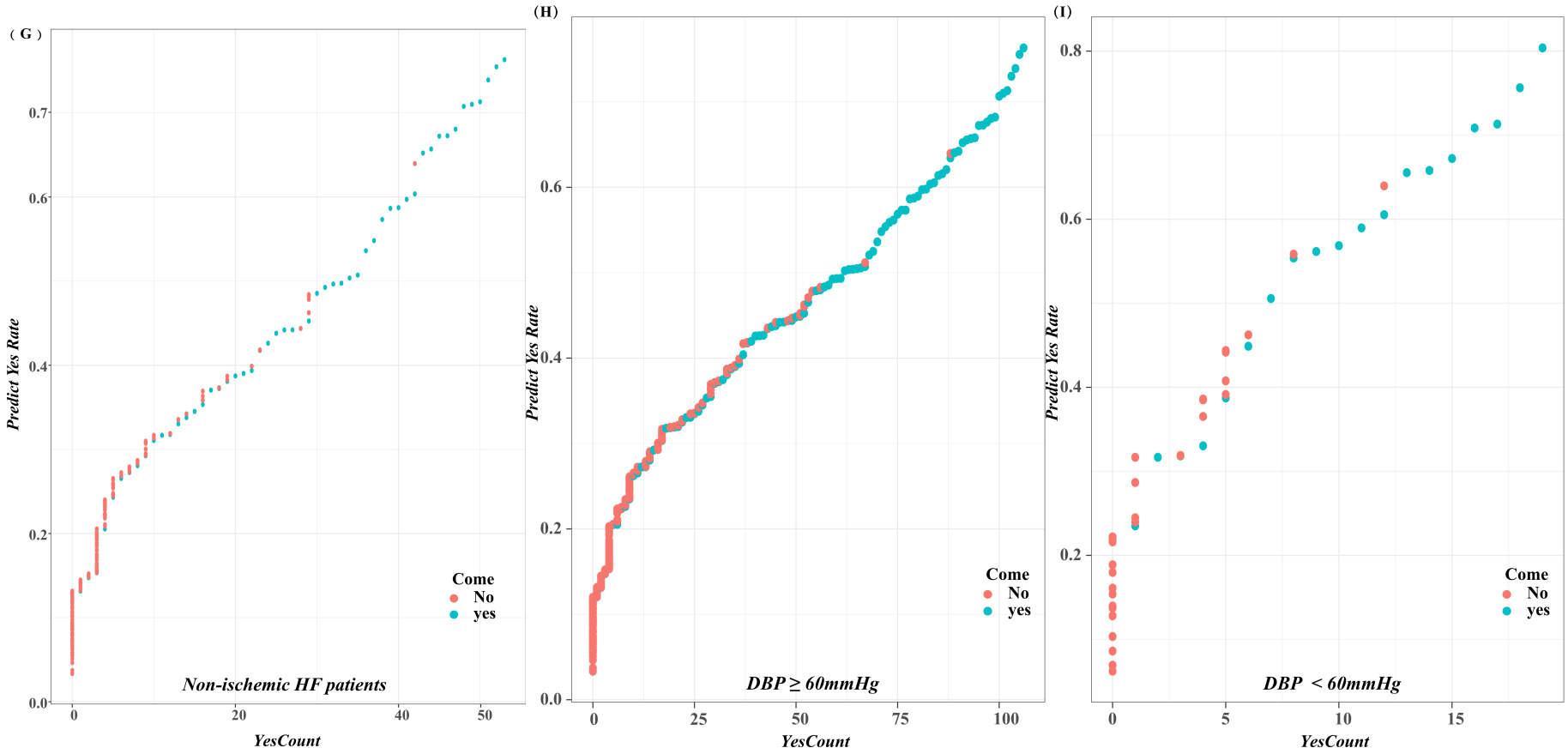


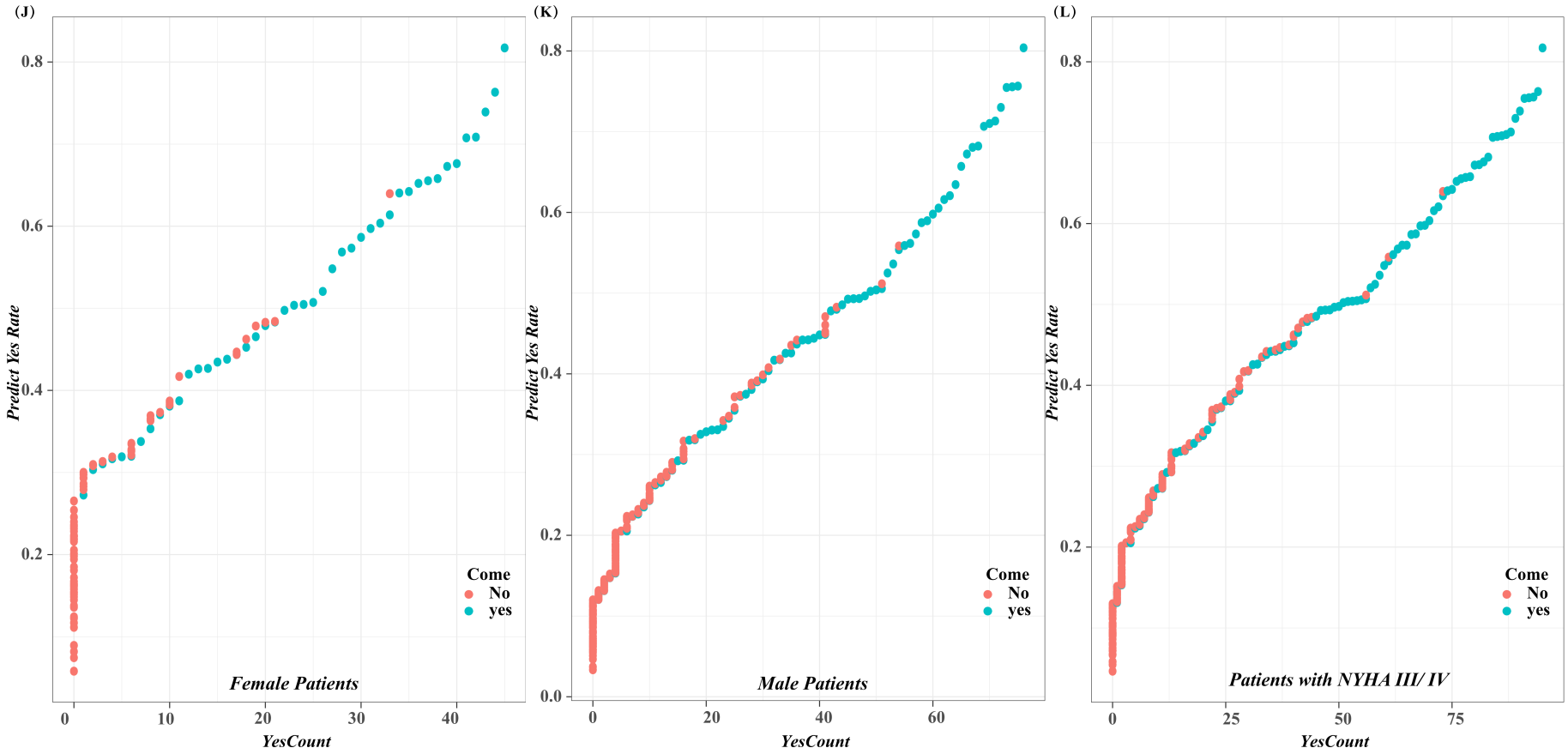
As shown in this figure, all MLBPMs could be more precise in stratifying patients at risk of all-cause mortality than the reference logistic model by positively clustering patients who died during long-term follow-up in the order of risk.

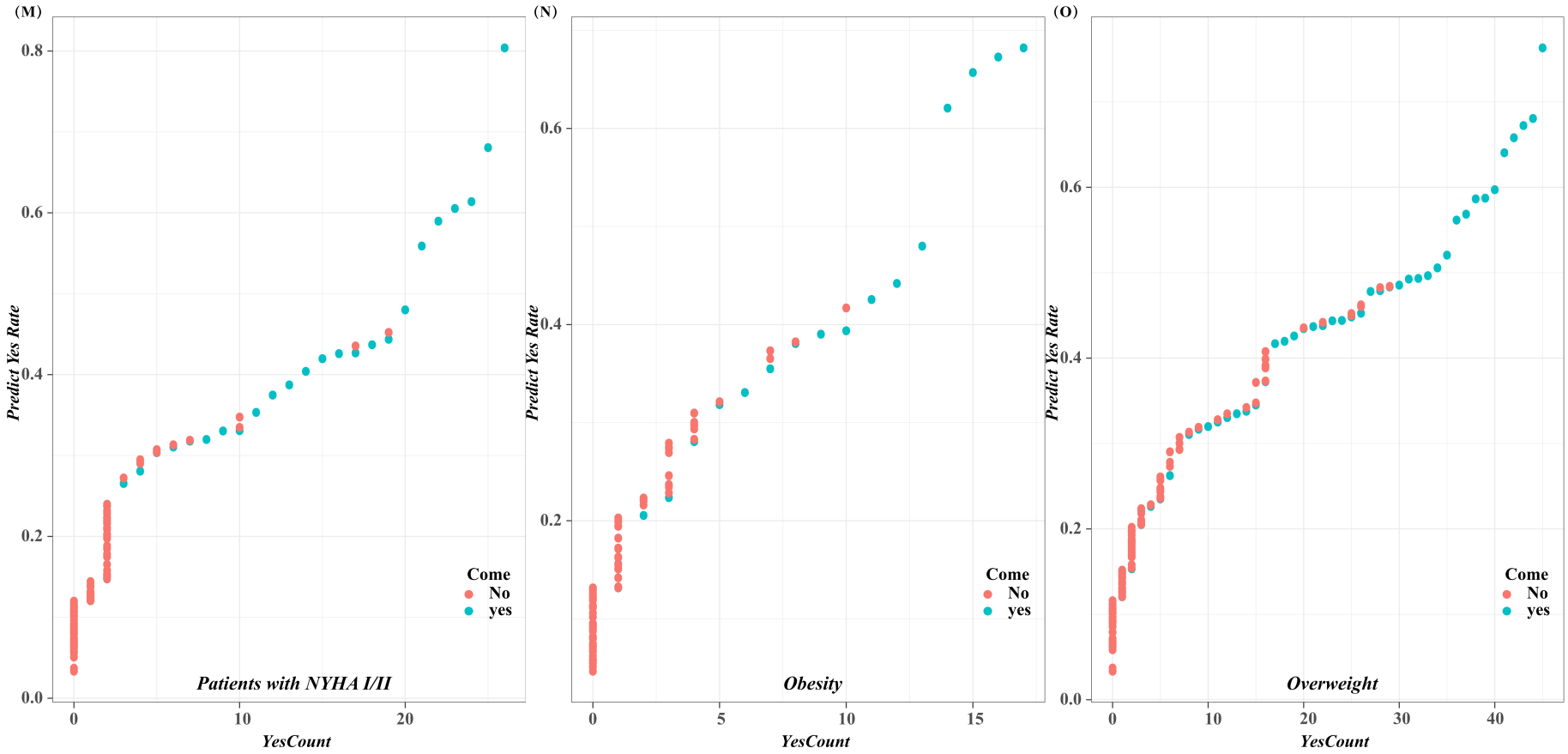
Figure S5. Prediction distribution plots of the 'All-in' XGBoost Model among different subgroup patients and observation times

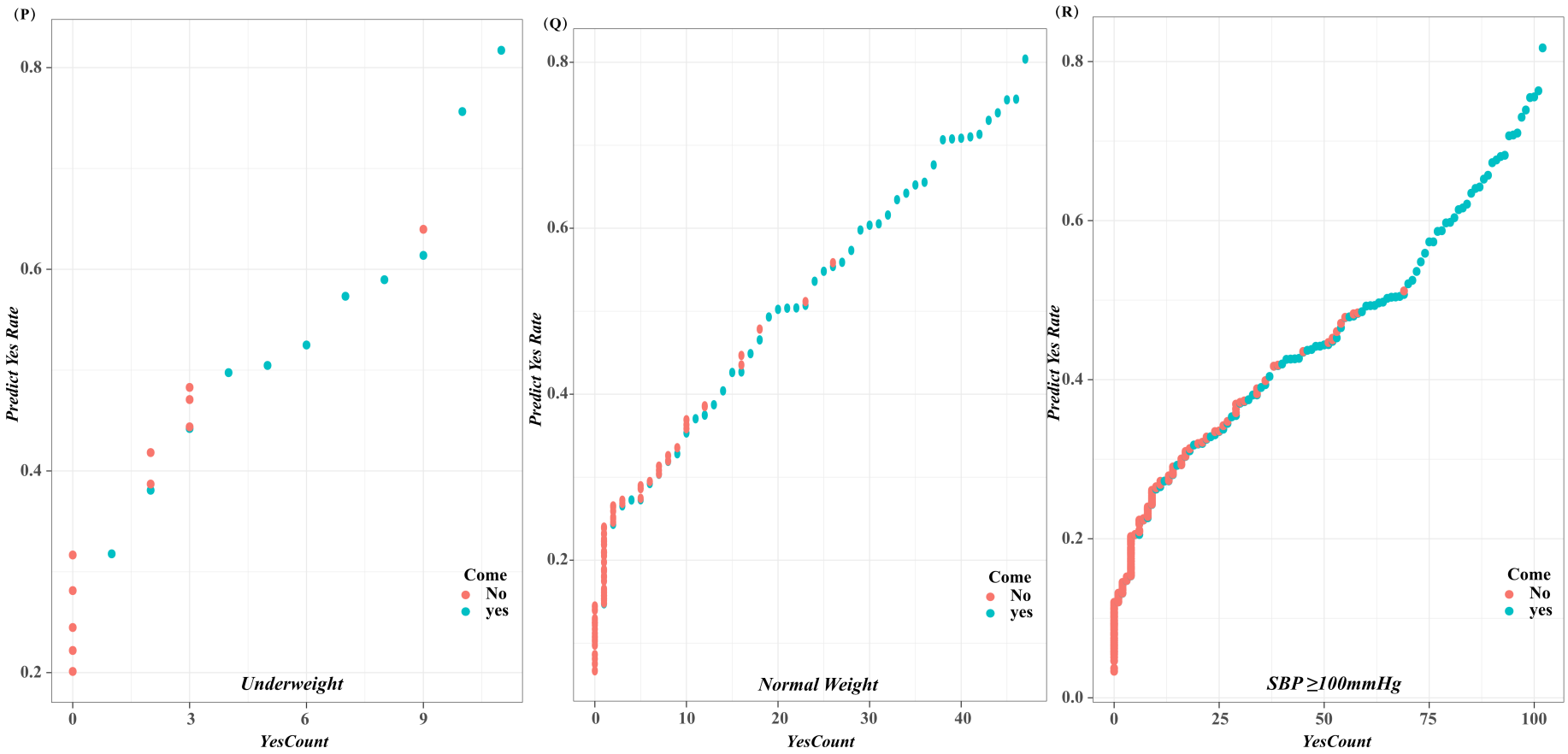


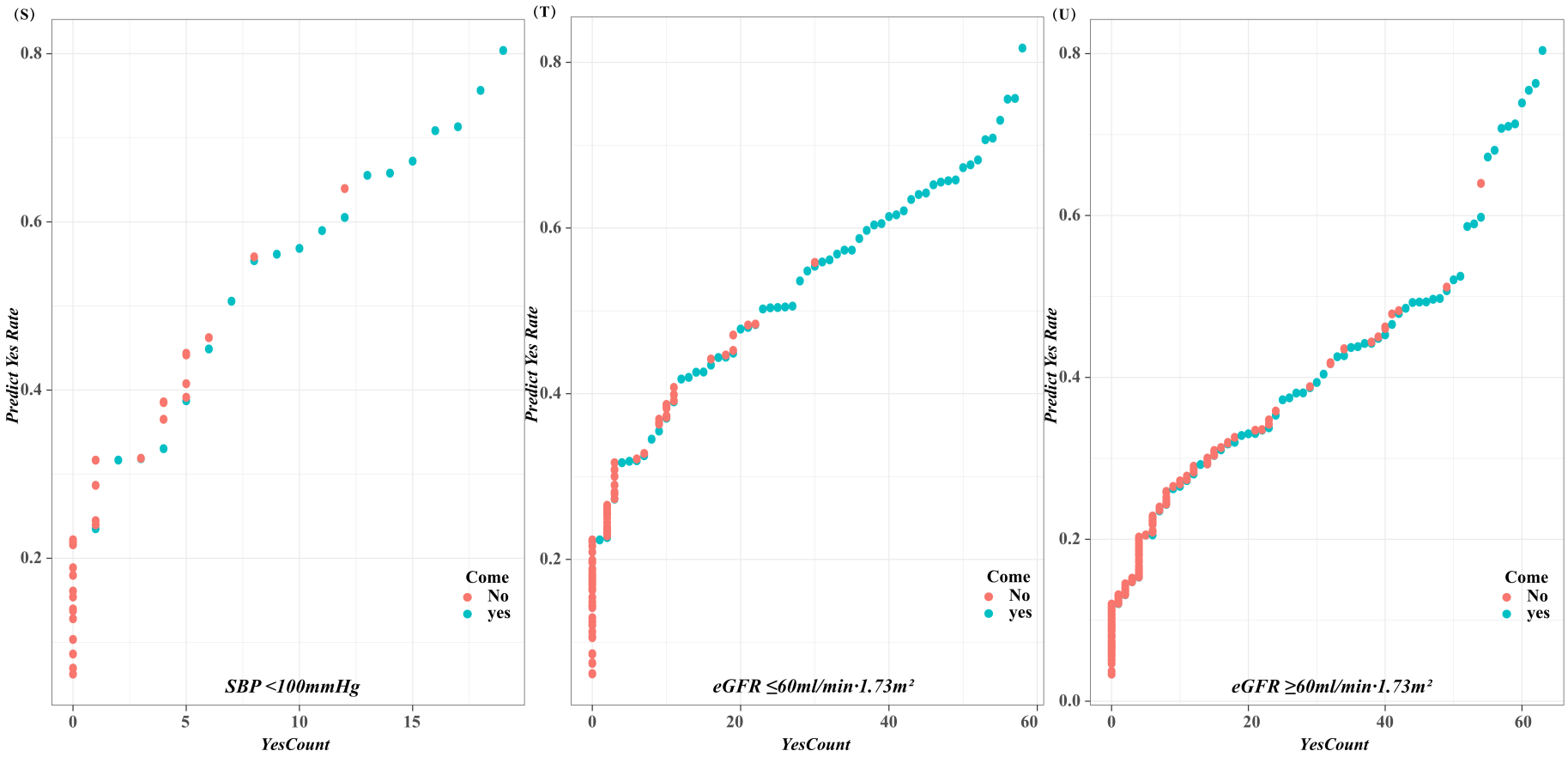

(D)
Predict Yes Rate
0.8
0.6
0.4
0.2
0
20
40
60
YesCount
Come
No
yes
Age >65 years old
(E)
Predict Yes Rate
0.8
0.6
0.4
0.2
0.0
0
20
40
60
YesCount
Come
No
yes
Age <65 years old
(F)
Predict Yes Rate
0.8
0.6
0.4
0.2
0
20
40
60
YesCount
Come
No
yes
Ischemic HF patients

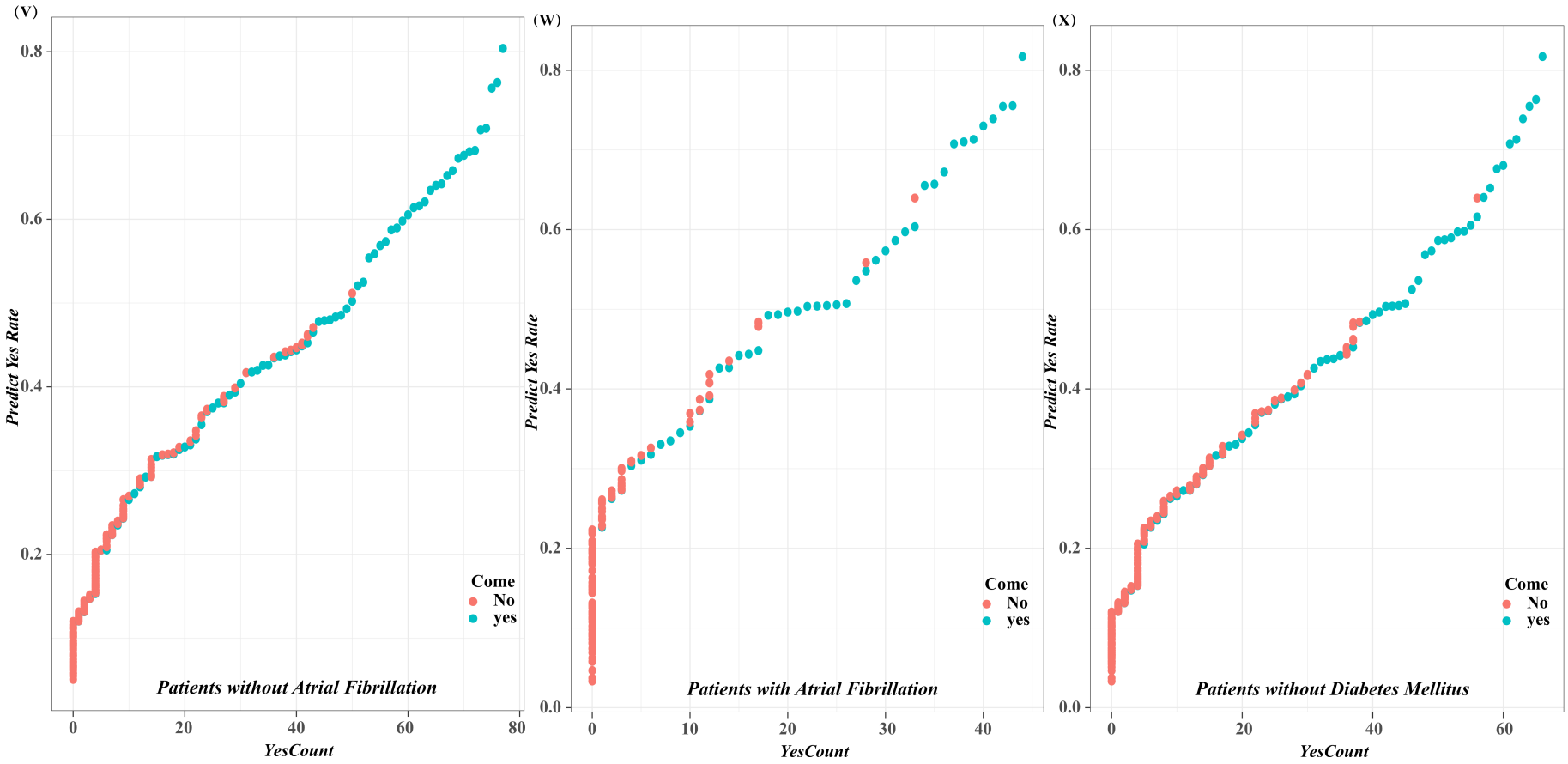

(G)
Predict Yes Rate
0.0
0.2
0.4
0.6
0.7
0
20
40
50
YesCount
Come
No
yes
Non-ischemic HF patients
(H)
Predict Yes Rate
0.0
0.2
0.4
0.6
0
25
50
75
100
YesCount
Come
No
yes
DBP ≥ 60mmHg
(I)
Predict Yes Rate
0.2
0.4
0.6
0.8
0
5
10
15
YesCount
Come
No
yes
DBP < 60mmHg

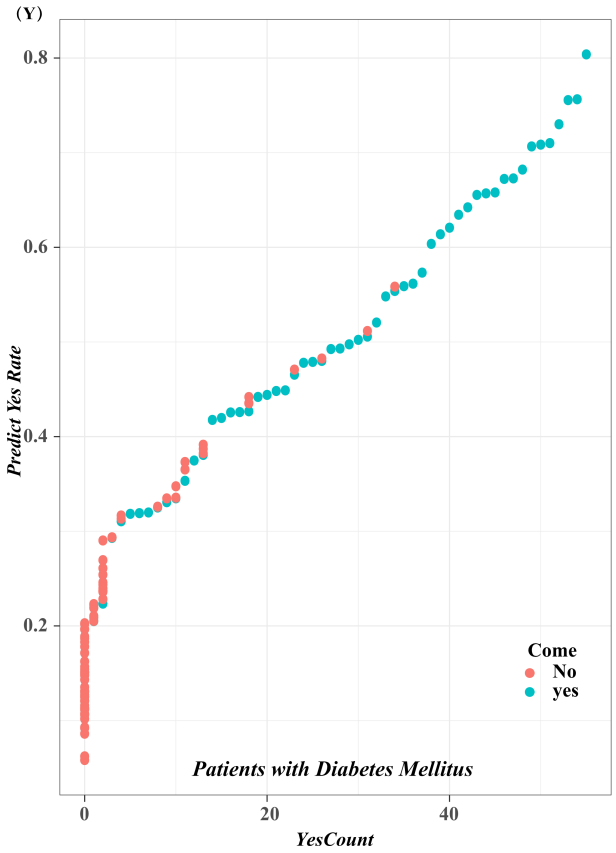

(J)
Predict Yes Rate
0.8
0.6
0.4
0.2
0
10
20
30
40
YesCount
Come
No
yes
Female Patients
(K)
Predict Yes Rate
0.8
0.6
0.4
0.2
0.0
0
20
40
60
YesCount
Come
No
yes
Male Patients
(L)
Predict Yes Rate
0.8
0.6
0.4
0.2
0
25
50
75
YesCount
Come
No
yes
Patients with NYHA III/ IV


(M)
Predict Yes Rate
0.0
0.2
0.4
0.6
0.8
0
10
20
YesCount
Come
No
yes
Patients with NYHA I/II
(N)
Predict Yes Rate
0.2
0.4
0.6
0
5
10
15
YesCount
Come
No
yes
Obesity
(O)
Predict Yes Rate
0.0
0.2
0.4
0.6
0
10
20
30
40
YesCount
Come
No
yes
Overweight


(P)
Predict Yes Rate
0.8
0.6
0.4
0.2
0
3
6
9
YesCount
Come
No
yes
Underweight
(Q)
Predict Yes Rate
0.8
0.6
0.4
0.2
0
10
20
30
40
YesCount
Come
No
yes
Normal Weight
(R)
Predict Yes Rate
0.8
0.6
0.4
0.2
0.0
0
25
50
75
100
YesCount
Come
No
yes
SBP ≥100mmHg


(S)
Predict Yes Rate
0.8
0.6
0.4
0.2
0
5
10
15
YesCount
Come
No
yes
SBP <100mmHg
(T)
Predict Yes Rate
0.8
0.6
0.4
0.2
0
20
40
60
YesCount
Come
No
yes
eGFR ≤60ml/min·1.73m²
(U)
Predict Yes Rate
0.8
0.6
0.4
0.2
0.0
0
20
40
60
YesCount
Come
No
yes
eGFR ≥60ml/min·1.73m²


(V)
Predict Yes Rate
0.8
0.6
0.4
0.2
0
20
40
60
80
YesCount
Come
No
yes
Patients without Atrial Fibrillation
(W)
Predict Yes Rate
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
YesCount
Come
No
yes
Patients with Atrial Fibrillation
(X)
Predict Yes Rate
0.8
0.6
0.4
0.2
0.0
0
20
40
60
YesCount
Come
No
yes
Patients without Diabetes Mellitus



The ‘all-in’ XGBoost model showed good performance in patients’ risk stratifying at observation times of 5(A), 3 (B), and 1 (C) years, and in different subgroup patients (D) age<65 years (E) age ⩾65 years (F) ischemic (G) non-ischemic (H) DBP⩾60mmHg (I) DBP <60mmHg (J) female patients (K) male patients (L) NYHA III/IV (M) NYHA I/II (N) obesity (O) overweight (P) underweight (Q) normal weight (R) SBP⩾100mmHg (S) SBP <100mmHg (T) eGFR ⩽60ml/min·1.73m^2 (U) eGFR >60ml/min·1.73m^2 (V) without atrial fibrillation (W) with atrial fibrillation (X) without diabetes mellitus (Y) without diabetes mellitus

Abbreviations: NYHA = New York Heart Association, eGFR = estimated filtration rate