



OPEN ACCESS

EDITED BY

Surjeet Dalal,
Amity University Gurgaon, India

REVIEWED BY

Samuel Huang,
Virginia Commonwealth University,
United States
Nancy Arya,
GD Goenka University, India

*CORRESPONDENCE

Yimin Yao

✉ 20173266@zcmu.edu.cn

Xiaochun Wang

✉ 20173107@zcmu.edu.cn

RECEIVED 06 December 2024

ACCEPTED 05 March 2025

PUBLISHED 17 March 2025

CITATION

Pu J, Yao Y and Wang X (2025) Development and validation of a machine learning model for online predicting the risk of in heart failure: based on the routine blood test and their derived parameters.
Front. Cardiovasc. Med. 12:1539966.
doi: 10.3389/fcvm.2025.1539966

COPYRIGHT

© 2025 Pu, Yao and Wang. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Development and validation of a machine learning model for online predicting the risk of in heart failure: based on the routine blood test and their derived parameters

Jianchen Pu¹, Yimin Yao^{2*} and Xiaochun Wang^{2*}

¹Medical Laboratory, Nanxun Hospital of Traditional Chinese Medicine, Huzhou, China, ²Medical Laboratory, The First Affiliated Hospital of Zhejiang Chinese Medical University (Zhejiang Provincial Hospital of Chinese Medicine), Hangzhou, China

Background: Heart failure (HF), a core component of cardiovascular diseases, is characterized by high morbidity and mortality worldwide. By collecting and analyzing routine blood data, machine learning models were built to identify the patterns of changes in blood indicators related to HF.

Methods: We conducted a statistical analysis of routine blood data from 226 patients who visited Zhejiang Provincial Hospital of Traditional Chinese Medicine (Hubin) between May 1, 2024, and June 30, 2024. The patients were divided into an experimental group (HF patients) and a normal control group. Additionally, 211 patients from the Qiantang and Xixi centers formed an independent external validation cohort. This study used both univariate and multivariate analyses to identify the risk factors associated with HF. Variables associated with HF were selected using LASSO regression analysis. In addition, eight different machine learning algorithms were applied for prediction, and the prediction performances of these algorithms were comprehensively evaluated using the receiver operating characteristic curve, area under the curve (AUC), calibration curve analysis, and decision curve analysis and confusion matrix.

Conclusions: Using LASSO regression analysis, leukocyte, neutrophil, red blood cell, hemoglobin, platelet, and monocyte-to-lymphocyte ratios were identified as risk factors for HF. Among the evaluated models, the random forest model exhibited the best performance. In the validation cohort, the area under the curve (AUC) of the model was 0.948, while that of the test cohort was 1.000. The calibration curve revealed good agreement between the actual and predicted probabilities, whereas the decision curve showed the significant clinical application of the model. Additionally, the AUC of the model in the external independent test cohort was 0.945.

Discussion: We used an online predictive tool to develop a predictive machine-learning model. The main purpose of this model was to predict the probability of developing HF in the future. This prediction can provide strong support and references for clinicians when making decisions. This online forecasting tool not only processes a large amount of data but also continuously optimizes and adjusts the accuracy of the model according to the latest medical research and clinical data. We hope to identify high-risk patients for early intervention to reduce the incidence of HF and improve their quality of life.

KEYWORDS

routine blood test, heart failure, machine learning, clinical research, Random Forest

1 Introduction

With an increasing aging population, the incidence of HF is increasing annually, placing great pressure on the global medical system (1). Recently, there has been increasing interest in identifying biomarkers that contribute to the early detection, prognosis, and monitoring of cardiovascular diseases. One such biomarker is brain diuretic natriuretic peptide (Pro-BNP) (2–4). Pro-BNP is a precursor peptide synthesized and released from the ventricular myocardium in response to cardiac stress and strain. High pro-BNP levels are associated with various cardiac conditions such as HF, myocardial infarction, and atrial fibrillation. Measurement of pro-BNP has become an important tool for the diagnosis and prognosis of patients with HF in clinical application (5–8). However, the implementation of pro-BNP testing poses certain difficulties for primary healthcare institutions. Therefore, exploring early diagnostic markers and effective treatments for HF is crucial for improving patient prognosis, enhancing the quality of life, and reducing healthcare costs.

The CBC is a routine clinical test that reflects the status of the inflammatory response in the body. Recent research has found that certain indicators in complete blood cell count, such as white blood cell count, neutrophil count, and lymphocyte count, are closely related to the occurrence and development of HF (9). These indicators are not only simple, rapid, and economical but can also reflect the response status and degree of patients with HF, which has important reference value for early diagnosis.

Machine learning (ML) is a branch of artificial intelligence that focuses on how computers simulate or implement human learning processes to master new knowledge or skills and reorganize existing knowledge structures to continuously improve their performance (10). In recent years, machine learning has been increasingly used in healthcare, particularly for predicting HF (11). As simple and economical tests, routine blood tests show great potential for predicting HF (9). By collecting and analyzing large amounts of routine blood data, machine learning models can be constructed to identify patterns of change in blood indicators associated with HF. These models can be used to assess the risk of HF and provide strong support for clinical decision-making.

HF exhibits features of high morbidity and high mortality (12). By collecting and analyzing routine blood data, machine learning models were built to identify the patterns of changes in blood indicators related to HF. The CBC (complete blood cell count) (CBC) is a common blood test that provides information about the cellular components of the blood, including red blood cells, white blood cells, and platelets. Changes in CBC parameters have been observed in various cardiovascular diseases, and these changes reflect underlying pathophysiological processes (13–16). Therefore, monitoring changes in these indicators can provide an important reference for the early diagnosis and treatment of HF.

2 Materials and methods

This study was approved by the Ethics Committee of First Affiliated Hospital of Zhejiang Chinese Medical University

(reference number 2024-KL-551-01). The subjects of the study were 226 patients (including 127 patients with HF) who received treatment at the Hubin Campus of Zhejiang Provincial Hospital of Traditional Chinese Medicine between May 1, 2024, and June 30, 2024, and 211 patients from the other two hospital districts were retrospectively analyzed. Patient data, including electronic medical records and laboratory test indicators, were obtained from a hospital information system. The inclusion criteria were as follows: (1) Availability of complete clinical data; (2) A definite diagnosis of HF; (3) An average age of 65–75 years. The exclusion criteria were as follows: (1) Other infectious diseases; (2) Concurrent malignancy or severe blood/immune system disease; and (3) Blood transfusion therapy in the last month.

Detailed patient clinical information and data including age, sex, complications, and routine peripheral blood status were collected. Routine blood examination data (white blood cell count, neutrophil count, lymphocyte count, monocyte count, red blood cell count, hemoglobin level, and platelet count) were also recorded. The following conventional blood-derived indicators were calculated: neutrophil-to-lymphocyte ratio (NLR), dNLR (derived neutrophil ratio (dNLR)), monocyte-to-lymphocyte ratio (MLR), NMLR (sum of the neutrophil-to-monocyte-to-lymphocyte ratio (NMLR)), SIRI (systemic inflammatory response index (SIRI)), and SII (systemic immune inflammation index). The calculation is as follows: $NLR = \frac{\text{Neutrophil counts } (10^9/L)}{\text{lymphocyte count } (10^9/L)}$, $dNLR = \frac{\text{neutrophil count } (10^9/L)}{\text{white blood cell count-lymphocyte count } (10^9/L)}$, $MLR = \frac{\text{monocyte count } (10^9/L)}{\text{lymphocyte count } (10^9/L)}$, $NMLR = \frac{(\text{monocyte count} + \text{neutrophil count}) (10^9/L)}{\text{lymphocyte count } (10^9/L)}$, $SIRI = \frac{\text{neutrophil count } (10^9/L)}{\text{Monocyte count } (10^9/L) / \text{lymphocyte count } (10^9/L)}$, $SII = \frac{\text{platelet count } (10^9/L)}{\text{Neutrophil count } (10^9/L) / \text{lymphocyte count } (10^9/L)}$. Based on the examination results of the patient, the group of 226 patients were divided into two groups: a HF group and a normal control group.

The Beckman Coulter DxAI platform (<https://www.xsmartanalysis.com/login/>) was used for statistical analysis. Minimum Absolute contraction and selection operator (LASSO) regression analyses were used to identify the factors associated with HF. XGBoost, Logistic Regression (LR), LightGBM (LGBM), Random Forest (RF), AdaBoost, Decision Tree (DT), Gradient Boosting Decision Tree (GBDT), and Gaussian Naive Bayes (GNB), were the eight candidate models which were evaluated based on calibration plots and assessed for their predictive performance in terms of sensitivity, specificity, accuracy, predictive value, and area under the curve (AUC) in both test and validation queues. Random draw of 15.00% of the data in the overall sample as the test queues and the remaining samples as the training queues for 2-fold cross-validation (One served as training queues and one as validation queues) and tested in $AUC = 0.9310 \pm 0.0009$ in the validation queues. The final model had an $AUC = 0.8889$ and an accuracy = 0.7647 in the test queues. This identified the optimal machine learning model. The filtered models were subsequently validated in an external independent test cohort.

SPSS Modeler 16.0 and the R software version 4.2.3 were used in this study. For measurement data, the *t*-test and Wilcoxon

signed-rank test were used for analysis, and for count data, the chi-square test was used for comparison. The *t*-test was used for numerical variables with a normal distribution and homogeneity of variance, and the Wilcoxon signed-rank test was used for numerical variables with a normal distribution but uneven variance. LASSO regression analysis was used to identify factors associated with the development of HF and to assess the performance of these factors using receiver operating characteristic (ROC) curves. The statistical significance level was set at $P < 0.05$.

3 Results

3.1 Baseline characteristics

Table 1 presents the baseline characteristics of 226 patients. Among these patients, 127 (56.19%) were diagnosed with HF, of which 52 (40.95%) were female patients and 75 (59.05%) were male patients. The remaining 99 (43.81%) patients were healthy of which 49 (49.5%) were female and 50 (50.50%) were male. In this study, sex did not show a significant difference ($P = 0.200$). All other factors (Table 1) were significantly different between the groups.

3.2 Feature selection related to HF

LASSO regression analysis was performed to identify factors associated with the risk of HF (Table 1). The analysis showed that white blood cell count (WBC), neutrophil count (NE), lymphocyte count (LY), monocyte count (MO), red blood cell count (RBC), hemoglobin (HGB), platelet count (PLT), monocyte-to-lymphocyte ratio (MLR), and systemic inflammation index (SII) were risk factors for HF (see Figure 1). In addition, we assessed the AUC values of these factors by ROC analysis (Figure 2). The AUC values for WBC, NE, LY, MO,

RBC, HGB, PLT, MLR, and SII, were 0.617, 0.718, 0.835, 0.584, 0.872, 0.857, 0.648, 0.854, and 0.730.

3.3 ML algorithm feature recognition

The REFCV, SVMREFCV, and Boruta algorithms were used to identify markers (Figures 3A–C), and a Venn diagram was drawn using the R language (Figure 3D). Considering the intersections of the three algorithms, six overlapping markers were identified: WBC, NE, RBC, HGB, PLT, and MLR.

3.4 Optimal model identification

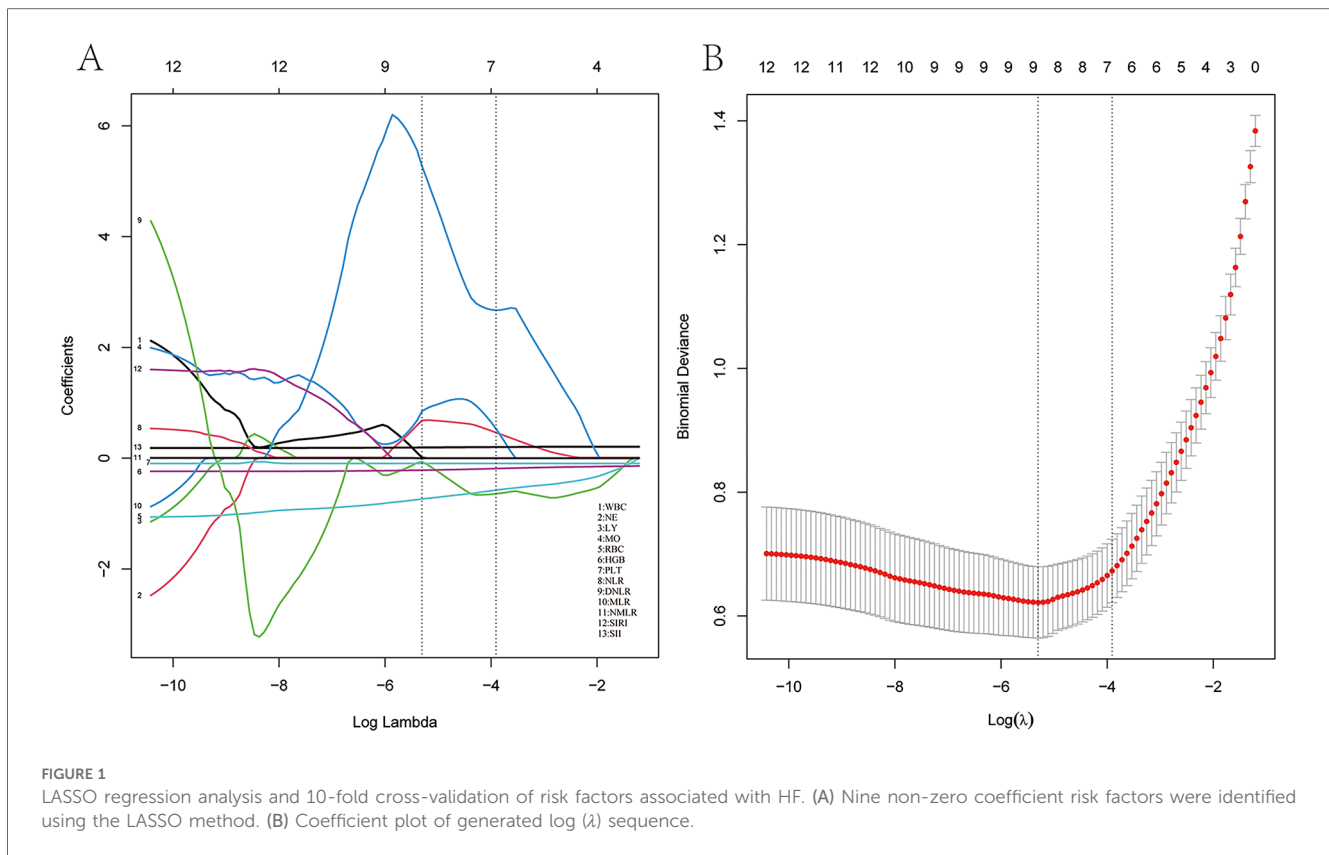
The performances of the eight machine learning models in the training and validation stages are presented in Table 2 and Figure 4. The random forest (RF) model was the most prominent in its prediction accuracy, with AUC values of 0.948 and 1.000 in the validation and testing stages, respectively. In addition, both the calibration and decision curves confirmed the excellent performance of the random forest (RF) model and its value in clinical applications.

3.5 Analysis of the random forest (RF) model

As shown in Table 3 and Figure 5A–C, the AUC values of the test cohort were similar to the AUC values of the validation cohort, and the validation cohort did not surpass the training cohort in Figure 5D, indicating a strong fit of the random forest (RF) model. Table 3 shows that accuracy, sensitivity, and specificity exceeded the 70% threshold for the test sequence. Furthermore, the calibration curve demonstrated a significant correlation between the actual and predicted probabilities (Figure 5E), while the decision curve demonstrated the significant clinical utility of the model (Figure 5F), thus confirming the superior performance

TABLE 1 Baseline characteristics of the two groups of patients.

Characteristics		Normal control ($n = 99$)	Heart failure group ($n = 127$)	<i>P</i> -value
Sex, <i>n</i> (%)	Female	49 (49.495%)	52 (40.945%)	0.200
	Male	50 (50.505%)	75 (59.055%)	
Variable category				
WBC, median [IQR]		5.700 [5.000, 6.600]	6.500 [5.000, 8.200]	0.003
NE, median [IQR]		3.100 [2.500, 3.800]	4.300 [3.200, 6.300]	<0.001
LY, mean (\pm SD)		1.940 \pm 0.444	1.256 \pm 0.525	<0.001
MO, median [IQR]		0.400 [0.400, 0.500]	0.500 [0.400, 0.600]	0.027
RBC, median [IQR]		4.550 [4.300, 4.930]	3.670 [3.130, 4.090]	<0.001
HGB, mean (\pm SD)		138.980 \pm 13.151	109.291 \pm 23.368	<0.001
PLT, median [IQR]		232.000 [184.000, 260.000]	179.000 [135.000, 239.000]	<0.001
NLR, median [IQR]		1.607 [1.294, 1.957]	3.625 [2.400, 6.250]	<0.001
DNLR, median [IQR]		0.844 [0.806, 0.868]	0.878 [0.833, 0.908]	<0.001
MLR, median [IQR]		0.222 [0.182, 0.267]	0.400 [0.308, 0.556]	<0.001
NMLR, median [IQR]		1.867 [1.500, 2.176]	4.143 [2.667, 6.889]	<0.001
SIRI, median [IQR]		0.720 [0.500, 0.978]	1.733 [0.921, 3.300]	<0.001
SII, median [IQR]		375.158 [270.667, 480.000]	696.000 [373.333, 1,195.000]	<0.001



of the Random Forest (RF) model. The results of the confusion matrix show the performance of the model differences on different datasets. In the training set (Figure 5G), the true positive rate (sensitivity) was 93.6% and the true negative rate (specificity) was 89.2%. In the test set (Figure 5H), the true positive rate was 77.8% and the true negative rate was 81.3%. Figure 6 shows the overall SHAP interpretation for all model covariates applicable to predict the probability of occurrence of heart failure.

3.6 External validation of the random forest (Rf) model

An external independent test cohort of 211 patients from the other two districts. The area under the curve (AUC) is 0.945 (Figure 7A), and the decision curve revealed a significant clinical benefit (Figure 7B).

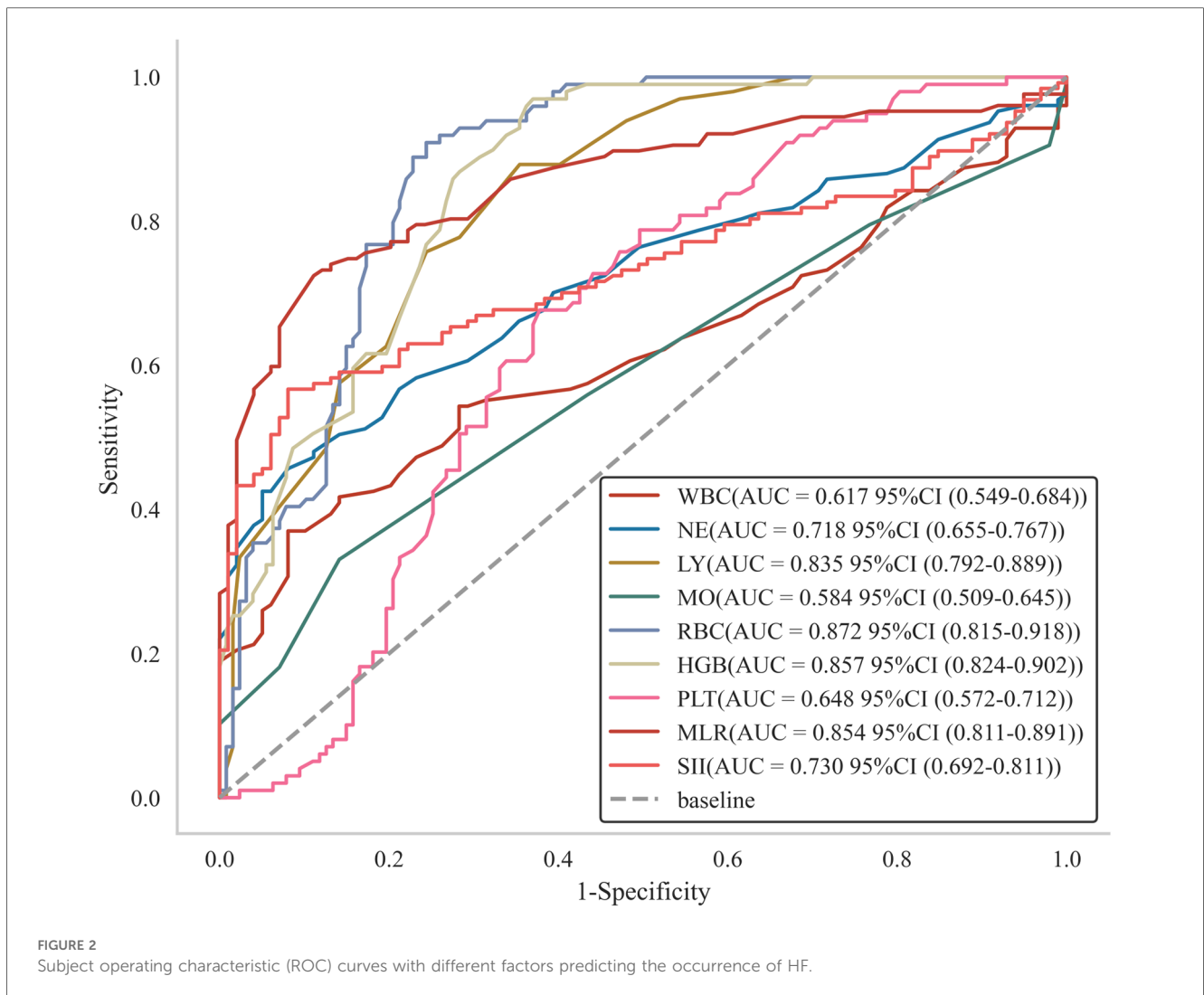
3.7 Online forecast site

Following the above analysis, we developed an online prediction tool designed to help primary clinicians assess the risk of HF occurrence in patients with suspected HF. This tool allows users to input blood indicators (WBC, NE, RBC, HGB, PLT, and MLR) to predict the likelihood of disease occurrence (Figure 8). If the results indicate a high risk of death, clinicians should be

vigilant and prepare for treatment in advance. (<http://www.xsmartanalysis.com/model/list/predict/model/html?mid=18904&symbol=71aW7309uF6WP3170Zn1>).

4 Discussion

In HF patients, a decline in the heart's ability to pump blood leads to an increased cardiac workload, which, in turn, prompts cardiac muscle cells to secrete more pro-BNP. Therefore, pro-BNP levels can serve as a key indicator for assessing cardiac function (17). However, despite the important clinical value of the pro-BNP test in the diagnosis of HF, it can help physicians quickly identify and evaluate the severity of HF, thus providing timely and effective treatment. However, in primary medical institutions, the popularization and application of pro-BNP testing still faces certain challenges owing to equipment and technical limitations. Many primary hospitals lack advanced testing equipment to perform accurate pro-BNP testing and the professional level of technicians not standardized, further limiting the wide application of pro-BNP testing in primary medical institutions. Therefore, identifying new early diagnostic markers is of great importance for improving the prognosis of patients with HF, improving their quality of life, and reducing the burden on the medical system. Moreover, by reducing the burden on the medical system, the allocation of medical resources can be optimized such that more patients can obtain timely and effective treatment, thereby improving the overall health of society.



With the widespread adoption of electronic health records, hospital information systems have accumulated abundant patient visit data worldwide, creating an ideal environment for machine learning (ML) applications. Traditional regression analysis methods have limitations in handling complex high-dimensional interaction information in large datasets, technically limiting the ability of models to make effective predictions regarding complex relationships. Machine learning can effectively overcome these challenges (18). When processing complex data, machine learning (ML) does not need to preset the nature of the data distribution or the linear or nonlinear connections between features. Machine learning can assist in identifying potential predictor variables and models using computationally intensive iterative algorithms rather than relying on manually selected features, thus improving the prediction accuracy of the model (19). In recent years, machine-learning technology has been widely used in cardiovascular medicine, particularly in the management of HF patients. For example, machine learning techniques have been used in the diagnosis of HF, death risk prediction, and evaluation of rehospitalization rates, and have demonstrated excellent efficacy (20, 21). The use of random

forest models can accurately predict the risk of heart failure in middle-aged and elderly individuals in pre-diabetes or diabetic states (22). The random forest algorithm utilizes the a reliable method for a combined model of individual features to improve the accuracy of high-frequency predictions (23). Studies have demonstrated the ability of the random forest model to identify risk factors in patients with HF (24). To enhance clinical physicians' recognition and trust in machine learning models, this study drew on the research findings of Alexander A. Huang et al. (25, 26) and introduced the SHAP value analysis method. This method, by enhancing the transparency of the model, further strengthens the model's credibility and the practicality of its clinical application.

Routine blood testing contains a lot of data on the disease, and is more economical and acceptable to patients than pro-BNP. By collecting and analyzing routine blood data, machine learning models were designed to identify the patterns of changes in blood indicators related to HF. This method adopts advanced data analysis technology, which can screen valuable information from a large number of blood indicators and provide a scientific basis for the early diagnosis and treatment of HF. Using this

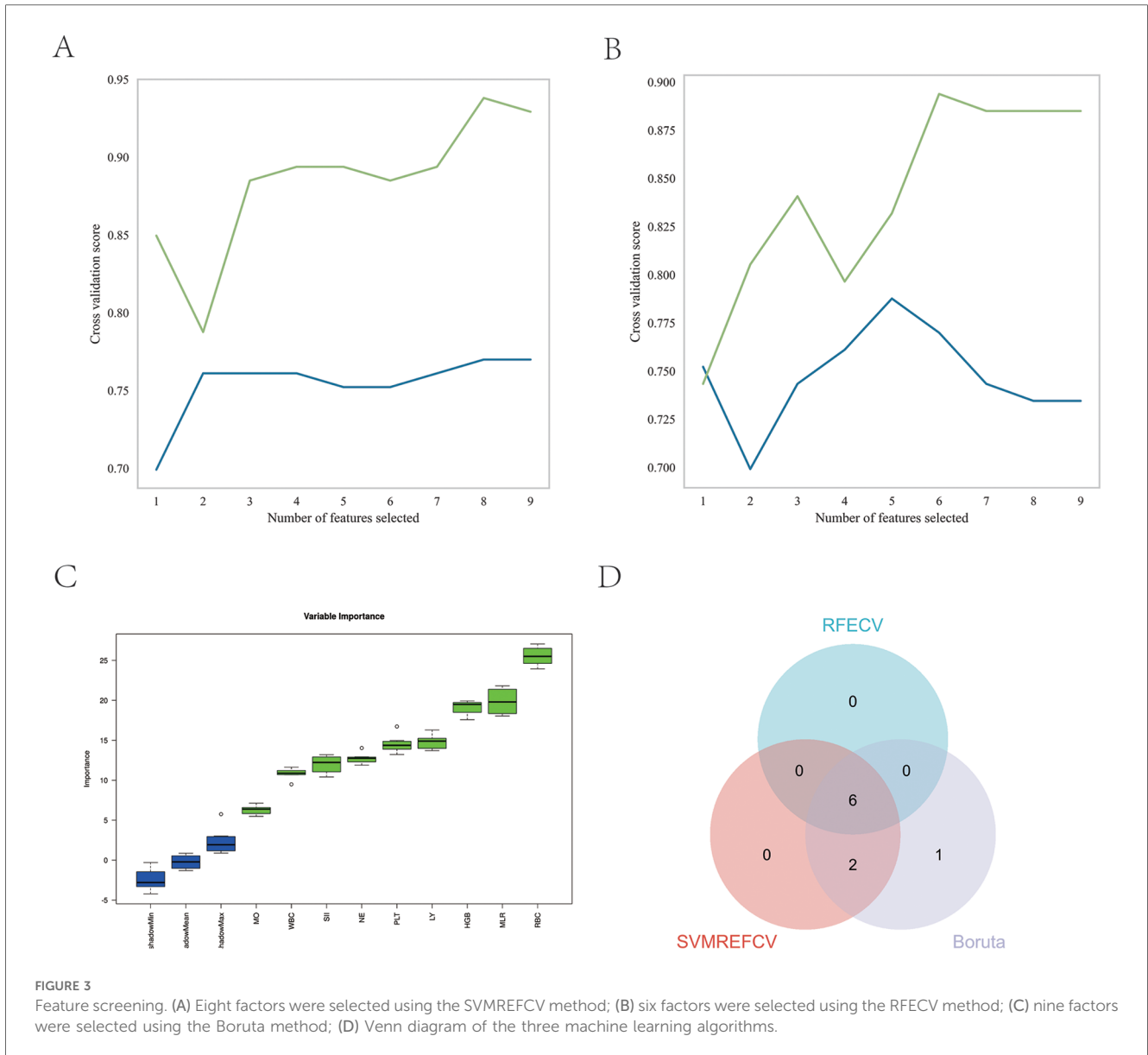


TABLE 2 The diagnostic effect of the eight classification models in the training and validation cohorts.

Classifier	Cohorts	AUC	Cutoff	Accuracy	Sensitivity	Specificity	Positive predictive value	Negative predictive value	F1
Logistic	Training	0.945	0.671	0.858	0.769	0.974	0.975	0.766	0.860
	Validation	0.927	0.671	0.783	0.681	0.923	0.916	0.699	0.775
XGBoost	Training	1.000	0.771	0.994	0.990	1.000	1.000	0.987	0.995
	Validation	0.934	0.771	0.837	0.780	0.907	0.908	0.776	0.839
LightGBM	Training	1.000	0.757	0.994	0.990	1.000	1.000	0.988	0.995
	Validation	0.899	0.757	0.783	0.730	0.859	0.886	0.688	0.800
RandomForest	Training	1.000	0.525	0.992	0.990	0.994	0.995	0.987	0.993
	Validation	0.948	0.525	0.837	0.822	0.851	0.886	0.801	0.848
AdaBoost	Training	1.000	0.501	0.994	0.990	1.000	1.000	0.988	0.995
	Validation	0.920	0.501	0.859	0.826	0.930	0.963	0.708	0.889
DecisionTree	Training	1.000	1.000	0.450	0.000	1.000	NaN	0.450	NaN
	Validation	0.803	1.000	0.391	0.000	1.000	NaN	0.391	NaN
GBDT	Training	1.000	0.866	0.994	0.990	1.000	1.000	0.987	0.995
	Validation	0.908	0.866	0.783	0.683	0.920	0.889	0.695	0.771
GNB	Training	0.941	0.155	0.864	0.831	0.906	0.918	0.809	0.872
	Validation	0.928	0.155	0.793	0.686	0.948	0.947	0.693	0.790

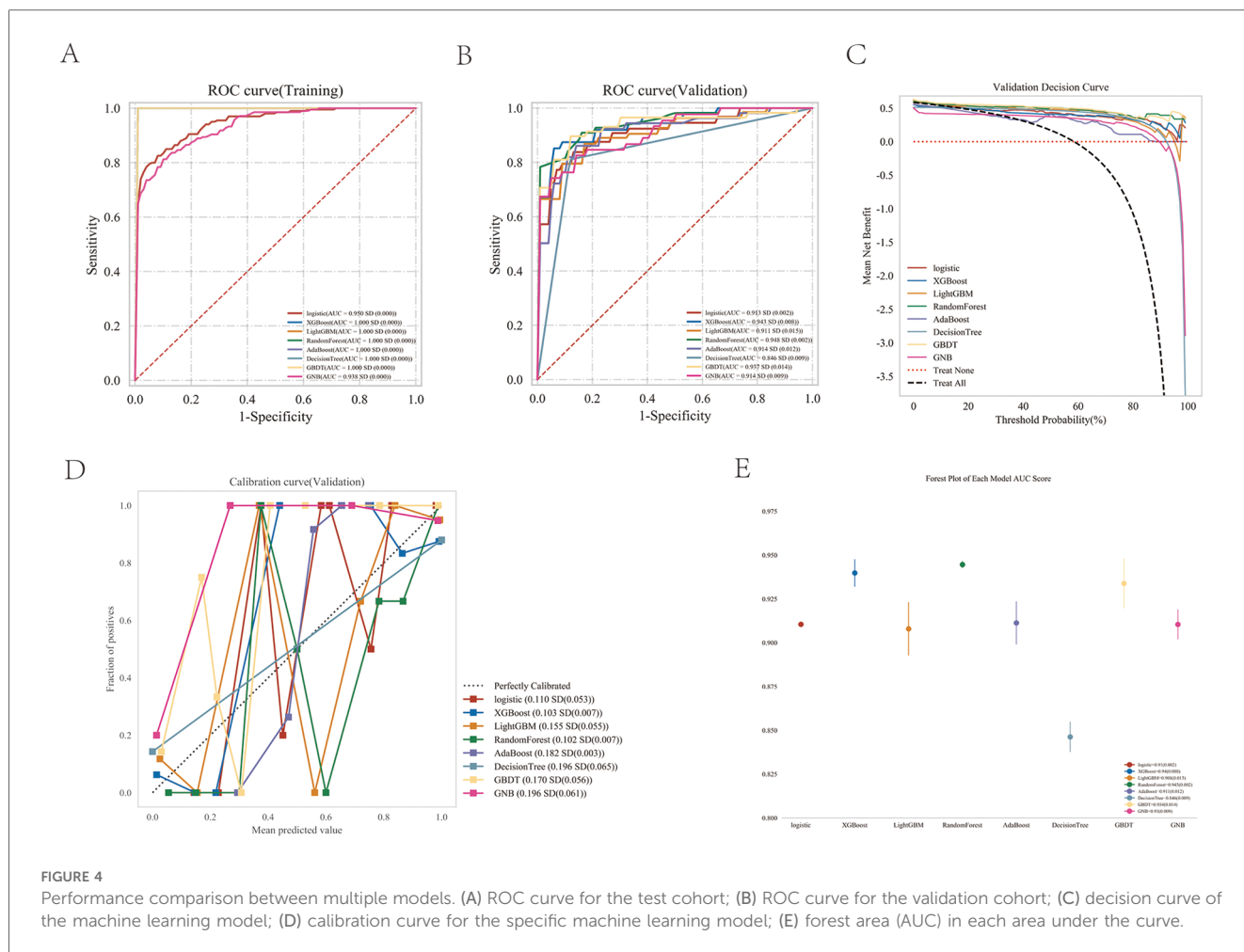


TABLE 3 Diagnosis effect of random forest (RF) models in the test and validation cohort.

Cohorts	AUC	Cutoff	Accuracy	Sensitivity	Specificity	Positive predictive value	Negative predictive value	F1
Training	0.999	0.600	0.974	0.963	0.988	0.991	0.953	0.977
Validation	0.931	0.600	0.844	0.788	0.915	0.928	0.776	0.849
Testing	0.889	0.550	0.754	0.722	0.813	0.813	0.722	0.765

method, doctors can more accurately assess a patient’s condition and develop targeted treatment plans to improve the treatment effect and reduce the morbidity and mortality. Therefore, after evaluating eight machine learning models, we developed a prediction model that used laboratory data to predict the risk factors associated with the occurrence of HF.

This study demonstrated significant associations between six blood markers (white blood cell count, neutrophil count, red blood cell count, hemoglobin level, platelet count, and monocyte-lymphocyte ratio) and the occurrence of HF. Among the eight machine learning algorithms evaluated, the Random Forest (RF) model showed the highest prediction accuracy and achieved high AUC values in both the validation and test sets. This result is consistent with Chicco D et al. (24).

Leukocytes constitute a key part of the immune system, and their main role is to resist infection and participate in

inflammatory responses. HF manifests as a decline in heart function, making the heart unable to efficiently pump blood to meet its needs (27). The association between white blood count and HF is mainly reflected in the inflammatory response. Studies have found that damage due to HF can lead to the release of inflammatory mediators, such as interleukins (ILs) and tumor necrosis factor (TNF), which stimulate the bone marrow to release more white blood cells into circulation, leading to an increase in the white blood cell count. Furthermore, according to relevant research findings, there is a strong association between a higher neutrophil count and HF, even within the normal range. Specifically, the comparison between neutrophil counts of $6-7 \times 10^9/L$ vs. $2-3 \times 10^9/L$ shows a strong correlation with HF (HR: 2.04; 95% CI: 1.82–2.06). This suggests that the white blood cell count, especially the level of neutrophils, may be associated with an increased risk of HF (28) therefore, the analysis of the

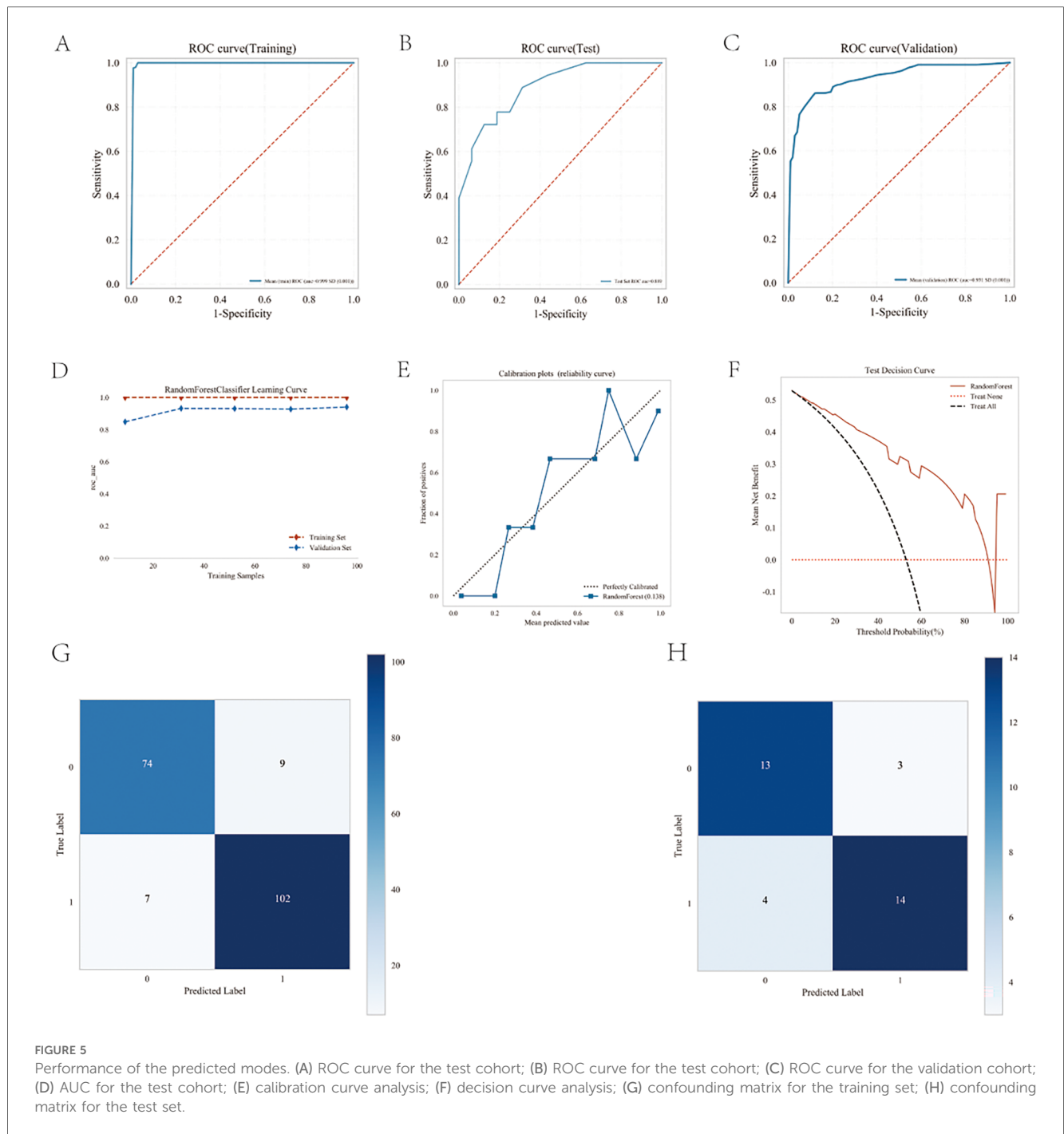
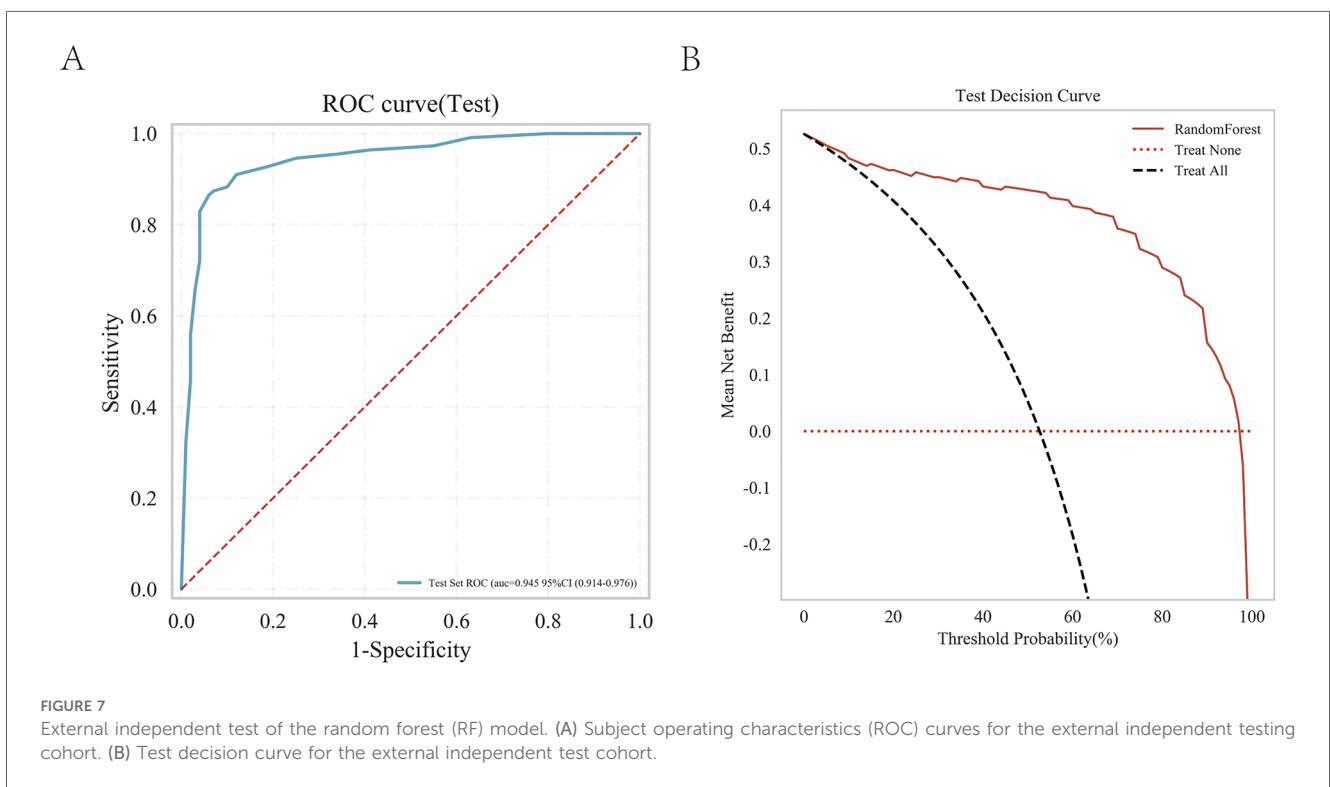
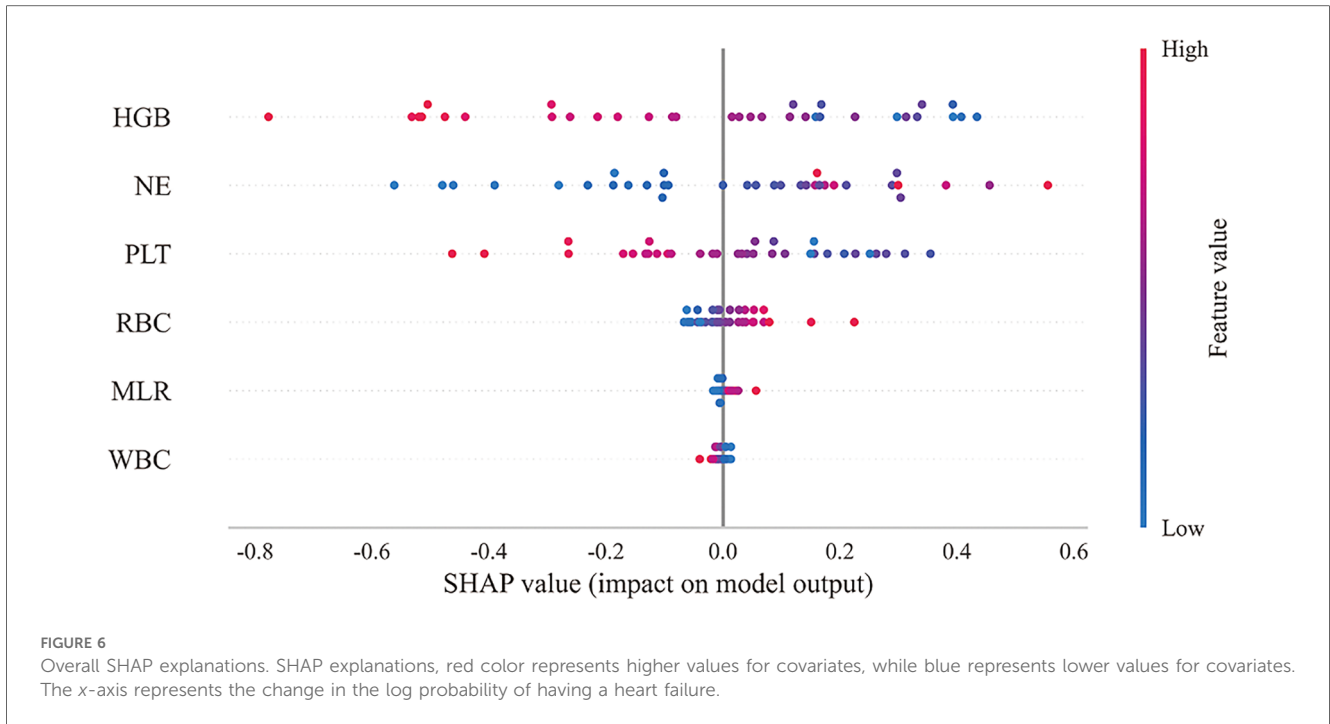


FIGURE 5 Performance of the predicted modes. (A) ROC curve for the test cohort; (B) ROC curve for the test cohort; (C) ROC curve for the validation cohort; (D) AUC for the test cohort; (E) calibration curve analysis; (F) decision curve analysis; (G) confounding matrix for the training set; (H) confounding matrix for the test set.

white blood cell count and subtypes may be very valuable in assessing the inflammatory status and prognosis of patients with HF. In this study, white blood cell counts were generally higher in patients with HF than in healthy controls, and the results were consistent with the findings of previous studies. An increase in the white blood cell count can be used as an indicator of inflammation in HF to assess disease severity and prognosis.

As primary oxygen-carrying cells, a reduction in the number of red blood cells leads to a decrease in the ability of blood to transport oxygen, which can exacerbate myocardial ischemia and hypoxia. To compensate for this deficiency, the heart attempts to

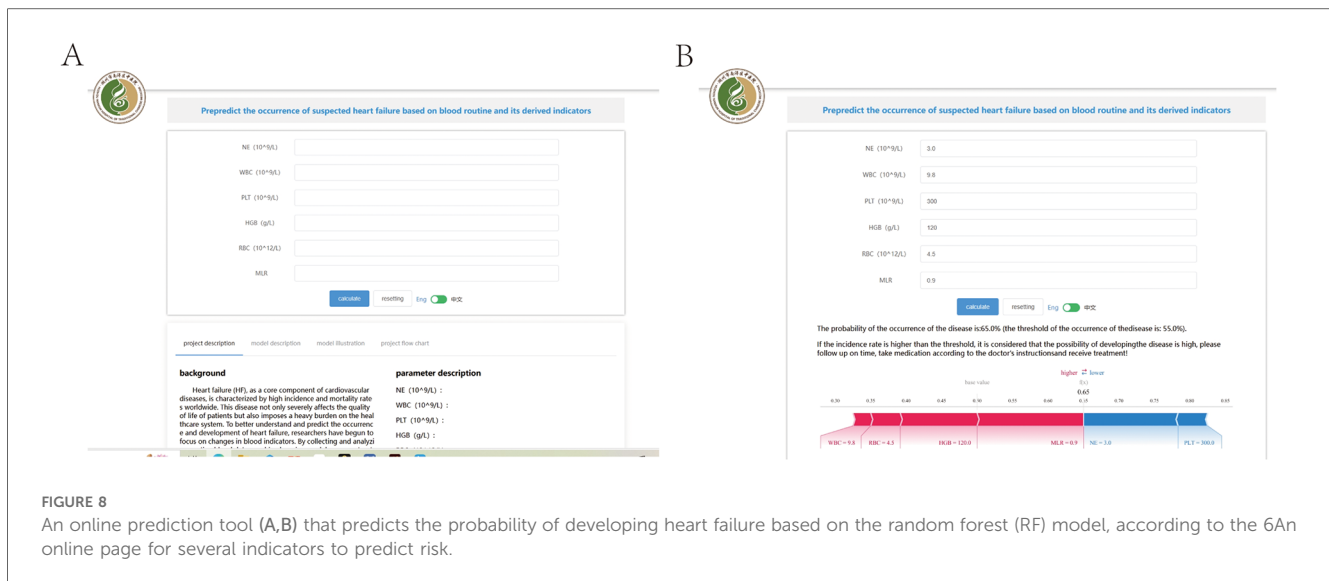
increase blood supply by accelerating the heart rate and enhancing myocardial contractility. However, this also increases the workload on the heart. Long-term reduction in red blood cells, along with myocardial ischemia and hypoxia, can lead to gradual deterioration of heart function. To meet the body's oxygen demand, the heart must exert more force to pump blood. This persistent increase in cardiac workload may lead to a gradual enlargement of the heart, weakened myocardial contractility, and ultimately result in the occurrence of HF (29). This is also highly consistent with the results of this study, further validating our findings and conclusions.



Platelets are also involved in blood flow. When the number of thrombocytes is reduced, the heart requires more power to maintain blood circulation, thereby increasing the burden on the heart. Long-term burden may lead to gradual impairment of cardiac function, leading HF. Studies have found that thrombocytopenia is a common platelet abnormality in patients with HF (30). In this study, the platelet count in the

experimental group was lower than that in the control group, which may have been because the majority of the patients were taking anticoagulant drugs, resulting in reduced platelet levels.

Our study has certain limitations. First, the sample size was relatively small. Second, the data were derived from a single center and were retrospective. Moreover, the population we selected is the elderly population, Therefore, this model has an



accurate predictive power for the probability of developing heart failure in the elderly population. In the future, we will gradually expand the sample size and include more covariates, as well as broaden the scope of the model, based on the progress of our research and the availability of resources and we will consider combining this model with additional clinical and imaging data to enhance the predictive power of the model.

5 Conclusions

This study identified key factors associated with the occurrence of HF using the LASSO regression method, including white blood cell count (WBC), neutrophil count (NE), red blood cell count (RBC), hemoglobin level (HGB), platelet count (PLT), and monocyte to the lymphocyte ratio (MLR). After evaluating eight different machine learning models, the Random Forest (RF) model was found to perform the best, demonstrating excellent predictive accuracy and clinical utility. A significant positive correlation was observed between the probability of actual occurrence and the probability predicted by the model. In addition, the online prediction tool can effectively assist doctors in primary medical institutions to predict a patient's condition more accurately.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Ethics statement

This study was approved by the Zhejiang Provincial Hospital of Traditional Chinese Medicine (2024-KL-551-01). All studies were performed in accordance with the ethical standards of the 1964 Declaration of Helsinki and its later amendments or comparable

ethical standards. Owing to the retrospective nature of the analysis and according to national legislation and institutional requirements, this study did not require written informed consent.

Author contributions

JP: Data curation, Formal analysis, Software, Writing – original draft. YY: Funding acquisition, Methodology, Writing – review & editing. XW: Project administration, Resources, Writing – review & editing.

Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

Acknowledgments

We gratefully acknowledge the JiZhi Analytics Platform for providing algorithm support and data analysis tools for this research.

Conflict of interest

The authors declare that this study was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Pandey A, Keshvani N, Pina IL. Prevalence and incidence of Pre-HF among Hispanics/Latinos. *JACC Heart Fail.* (2023) 11(8 Pt 1):958–60. doi: 10.1016/j.jchf.2023.05.021
- Isnard R. NT-BNP/BNP for screening left ventricular hypertrophy in hypertension: what else? *Arch Cardiovasc Dis.* (2008) 101(5):295–7. doi: 10.1016/j.acvd.2008.05.004
- Jensen J, Ma LP, Bjurman C, Hammarsten O, Fu ML. Prognostic values of NTproBNP/BNP ratio in comparison with NTproBNP or BNP alone in elderly patients with chronic heart failure in a 2-year follow up. *Int J Cardiol.* (2012) 155(1):1–5. doi: 10.1016/j.ijcard.2011.01.083
- Ghantous CM, Kamareddine L, Farhat R, Zoueini FA, Mondello S, Kobeissy F, et al. Advances in cardiovascular biomarker discovery. *Biomedicines.* (2020) 8(12):552. doi: 10.3390/biomedicines8120552
- Cao Z, Jia Y, Zhu B. BNP and NT-proBNP as diagnostic biomarkers for cardiac dysfunction in both clinical and forensic medicine. *Int J Mol Sci.* (2019) 20(8):1820. doi: 10.3390/ijms20081820
- Panagopoulou V, Deftereos S, Kossyvakis C, Raisakis K, Giannopoulos G, Bouras G, et al. NTproBNP: an important biomarker in cardiac diseases. *Curr Top Med Chem.* (2013) 13(2):82–94. doi: 10.2174/1568026611313020002
- Mair J, Gerda F, Renate H, Ulmer H, Andrea G, Pachinger O, et al. Head-to-head comparison of B-type natriuretic peptide (BNP) and NT-proBNP in daily clinical practice. *Int J Cardiol.* (2008) 124(2):244–6. doi: 10.1016/j.ijcard.2006.11.230
- Ceriello A, Lalic N, Montanya E, Valensi P, Khunti K, Hummel M, et al. NT-proBNP point-of-care measurement as a screening tool for heart failure and CVD risk in type 2 diabetes with hypertension. *J Diabetes Complications.* (2023) 37(3):108410. doi: 10.1016/j.jdiacomp.2023.108410
- Torfi E, Bahreiny SS, Saki N, Khademi R, Sarbazjoda E, Nezhad IA, et al. Evaluation of pro-BNP biomarker in heart failure patients and its relationship with complete blood count parameters: a case-control study. *Health Sci Rep.* (2024) 7(9):e70083. doi: 10.1002/hsr2.70083
- Heinrichs B, Eickhoff SB. Your evidence? Machine learning algorithms for medical diagnosis and prediction. *Hum Brain Mapp.* (2020) 41(6):1435–44. doi: 10.1002/hbm.24886
- Guo C, Wu M, Cheng H. The comprehensive machine learning analytics for heart failure. *Int J Environ Res Public Health.* (2021) 18(9):4943. doi: 10.3390/ijerph18094943
- Rogers JG, Patel CB, Mentz RJ, Granger BB, Steinhilber KE, Fuzat M, et al. Palliative care in heart failure: the PAL-HF randomized, controlled clinical trial. *J Am Coll Cardiol.* (2017) 70(3):331–41. doi: 10.1016/j.jacc.2017.05.030
- Melanson SE, Lewandrowski EL. Laboratory testing for B-type natriuretic peptides (BNP and NT-proBNP): clinical usefulness, utilization, and impact on hospital operations. *Am J Clin Pathol.* (2005) 124(Suppl):S122–8. doi: 10.1309/EUOG395MYPWNTND7
- Lassale C, Curtis A, Abete I, van der Schouw YT, Verschuren WMM, Lu Y, et al. Elements of the complete blood count associated with cardiovascular disease incidence: findings from the EPIC-NL cohort study. *Sci Rep.* (2018) 8(1):3290. doi: 10.1038/s41598-018-21661-x
- Haybar H, Pezeshki S, Saki N. Evaluation of complete blood count parameters in cardiovascular diseases: an early indicator of prognosis? *Exp Mol Pathol.* (2019) 110:104267. doi: 10.1016/j.yexmp.2019.104267
- Madjid M, Fatemi O. Components of the complete blood count as risk predictors for coronary heart disease. *Tex Heart Inst J.* (2013) 40(1):17–29.
- Oremek GM, Passek K, Holzgreve F, von der Eltz V, Droge J. The biomarkers BNP and NT-proBNP. *Zentralbl Arbeitsmed Arbeitsschutz Ergon.* (2023) 73(2):89–95. doi: 10.1007/s40664-022-00491-9
- Li X, Shang C, Xu C, Wang YT, Xu JC, Zhou Q, et al. Development and comparison of machine learning-based models for predicting heart failure after acute myocardial infarction. *BMC Med Inform Decis Mak.* (2023) 23(1):165. doi: 10.1186/s12911-023-02240-1
- Wang Q, Li B, Chen K, Yu F, Su H, Hu K, et al. Machine learning-based risk prediction of malignant arrhythmia in hospitalized patients with heart failure. *ESC Heart Fail.* (2021) 8(6):5363–71. doi: 10.1002/ehf2.13627
- Tabassian M, Sunderji I, Erdei T, Sanchez-Martinez S, Degiovanni A, Marino P, et al. Diagnosis of heart failure with preserved ejection fraction: machine learning of spatiotemporal variations in left ventricular deformation. *J Am Soc Echocardiogr.* (2018) 31(12):1272–84. doi: 10.1016/j.echo.2018.07.013
- 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(1):12–21. doi: 10.1016/j.jchf.2019.06.013
- Wang Y, Hou R, Ni B, Jiang Y, Zhang Y. Development and validation of a prediction model based on machine learning algorithms for predicting the risk of heart failure in middle-aged and older US people with prediabetes or diabetes. *Clin Cardiol.* (2023) 46(10):1234–43. doi: 10.1002/clc.24104
- Yuan H, Fan X, Jin Y, He JX, Gui Y, Song LY, et al. Development of heart failure risk prediction models based on a multi-marker approach using random forest algorithms. *Chin Med J.* (2019) 132(7):819–26. doi: 10.1097/CM9.0000000000000149
- Chicco D, Jurman G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC Med Inform Decis Mak.* (2020) 20(1):16. doi: 10.1186/s12911-020-1023-5
- Huang AA, Huang SY. Dendrogram of transparent feature importance machine learning statistics to classify associations for heart failure: a reanalysis of a retrospective cohort study of the medical information mart for intensive care III (MIMIC-III) database. *PLoS One.* (2023) 18(7):e288819. doi: 10.1371/journal.pone.0288819
- Huang AA, Huang SY. Shapely additive values can effectively visualize pertinent covariates in machine learning when predicting hypertension. *J Clin Hypertens.* (2023) 25(12):1135–44. doi: 10.1111/jch.14745
- Zhang H, Pan B, Wu P, Parajuli N, Rekhter MD, Goldberg AL, et al. PDE1 inhibition facilitates proteasomal degradation of misfolded proteins and protects against cardiac proteinopathy. *Sci Adv.* (2019) 5(5):w5870. doi: 10.1126/sciadv.aaw5870
- Shah AD, Denaxas S, Nicholas O, Hingorani AD, Hemingway H. Neutrophil counts and initial presentation of 12 cardiovascular diseases: a CALIBER cohort study. *J Am Coll Cardiol.* (2017) 69(9):1160–9. doi: 10.1016/j.jacc.2016.12.022
- Kadia TM, Reville PK, Wang X, Rausch C R, Borthakur G, Pemmaraju N, et al. Phase II study of venetoclax added to cladribine plus low-dose cytarabine alternating with 5-azacitidine in older patients with newly diagnosed acute myeloid leukemia. *J Clin Oncol.* (2022) 40(33):3848–57. doi: 10.1200/JCO.21.02823
- Getawa S, Bayleyegn B. Platelet, neutrophil and lymphocyte quantitative abnormalities in patients with heart failure: a retrospective study. *Vasc Health Risk Manag.* (2023) 19:69–78. doi: 10.2147/VHRM.S394765