Available Online: 30 April 2026 DOI: 10.54254/3049-5458/2026.33245

Application of left atrial reservoir strain combined with SII and SIRI in stroke risk stratification in patients with atrial fibrillation: a machine learning study

Cuicui Wu[1], Boyan Jia[1], Yeting Wang[1], Jingqiu Zhang[1], Lizhen Li[2*]

[1]Youjiang Medical University for Nationalities, Baise, China.
[2]Department of Ultrasound, Affiliated Hospital of Youjiang Medical University for Nationalities, Baise, China.

*Corresponding Author. Email: 15607761729@163.com

Abstract. Non-Valvular Atrial Fibrillation (NVAF) is closely associated with ischemic stroke, and accurate risk stratification is crucial for individualized anticoagulation. This study aimed to develop a machine learning model for ischemic stroke risk stratification in NVAF patients by integrating Left Atrial Strain Parameters and Systemic Inflammatory Indices (SII, SIRI), and to evaluate its clinical utility. A retrospective case-control study enrolled 109 NVAF patients (46 with stroke, 63 without); Left Atrial Reservoir Strain (LASr), Conduit Strain (LAScd), Contractile Strain (LASc) were measured via two-dimensional speckle-tracking echocardiography, and SII, SIRI were calculated. Independent risk factors identified by multivariate logistic regression were used to construct 11 machine learning models, whose performance was assessed by AUC, accuracy, and F1-score; the optimal model was interpreted by SHAP and its clinical net benefit evaluated by decision curve analysis. SIRI, LASr, and SII were independently associated with stroke (all $p < 0.05$); the Random Forest model performed best, with a testing set AUC of 0.940 (95% CI: 0.852-1.000), and SHAP analysis and decision curve analysis confirmed its reliability and clinical value. The multimodal machine learning model integrating left atrial reservoir function and inflammatory status effectively stratifies stroke risk in NVAF patients, and may serve as a complementary tool to the CHA_2DS_2-VASc score for individualized anticoagulation decisions.

Keywords: left atrial strain, stroke risk, speckle-tracking echocardiography, atrial dysfunction, systemic immune inflammation index, system inflammation response index

1. Introduction

Atrial Fibrillation (AF) is the most common sustained cardiac arrhythmia in clinical practice. Ischemic stroke secondary to AF is associated with high rates of disability and mortality, imposing a heavy burden on the public health system [1]. Current international guidelines recommend using the CHA_2DS_2-VASc score for stroke risk stratification in patients with Non-Valvular Atrial Fibrillation (NVAF) to guide anticoagulation therapy [2]. However, as a simplified tool designed to aid clinical decision-making, the CHA_2DS_2-VASc score exhibits only moderate predictive performance for identifying "high-risk" patients, with a c-index of

approximately 0.60–0.65 [3]. In clinical practice, it is not uncommon for patients classified as low-to-intermediate risk (e.g., score 1-2) to experience stroke,while anticoagulation decisions for some high-risk patients are complicated by factors such as advanced age and high bleeding risk [4, 5]. Therefore,moving beyond traditional clinical scores to develop a multimodal assessment tool that can more accurately identify high-risk individuals and provide evidence for individualized anticoagulation strategies has become an urgent need in the field of AF-related stroke prevention.

In terms of imaging assessment, conventional echocardiographic parameters, such as Left Atrial (LA) diameter and volume index,can reflect overt structural atrial remodeling but lack sensitivity for detecting subclinical LA mechanical dysfunction [6]. Two-Dimensional Speckle-Tracking Echocardiography (2D-STE), owing to its angle independence and excellent reproducibility, has emerged as the preferred tool for quantitative assessment of LA function [7, 8]. This technique enables precise analysis of the three phases of LA mechanical function: Longitudinal Reservoir Strain (LASr), reflecting LA compliance and reservoir function; Conduit Strain (LAScd), reflecting passive emptying function; and Contractile Strain (LASc), reflecting active booster pump function [9]. These parameters are sensitive to atrial myocardial fibrosis and serve as ideal imaging biomarkers for evaluating the severity of atrial cardiomyopathy [10].

In recent years, numerous studies have highlighted the clinical utility of integrated inflammatory indices derived from complete blood cell counts—namely, the Systemic Immune-inflammation Index (SII) and the Systemic Inflammation Response Index (SIRI) [11]. These indices have garnered widespread attention due to their convenience, low cost,and ability to comprehensively reflect the body's immune, inflammatory, and coagulation activation status. SII and SIRI integrate neutrophil, lymphocyte, monocyte, and platelet counts, representing a composite measure of systemic inflammation and prothrombotic state [12]. The recent Ziqi Jin study also revealed associations between SII, SIRI, and the risk of cardiovascular disease and all-cause mortality [13]. Notably, the systemic inflammatory state reflected by SII and SIRI and the LA dysfunction resulting from atrial cardiomyopathy may exert synergistic effects in thrombogenesis: inflammation can promote atrial fibrosis, exacerbating LA functional impairment, while atrial blood stasis creates a local milieu conducive to inflammation-mediated thrombosis. However, up to now, no research has combined 2D-STE parameters that allow accurate assessment of regional atrial mechanical function with SII and SIRI, which represent systemic inflammatory status, to explore their combined role in predicting stroke risk among patients with atrial fibrillation.

Currently, there is a lack of research integrating precisely quantified regional LA mechanical function parameters derived from 2D-STE (LASr, LAScd, LASc) with hematological indicators of the Systemic Prothrombotic Inflammatory environment (SII, SIRI) to construct interpretable risk stratification models using machine learning algorithms. Therefore, we hypothesized that a machine learning-based risk stratification model integrating these multimodal indicators might outperform the traditional CHA_2DS_2-VASc score,enabling more accurate stroke risk stratification in patients with Non-Valvular Atrial Fibrillation (NVAF). To test this hypothesis, we conducted a retrospective case-control study aimed at investigating the value of combining left atrial strain parameters derived from Two-Dimensional Speckle-Tracking Echocardiography (2D-STE) with systemic inflammatory indices for stroke risk stratification in NVAF patients. By constructing and comparing multiple machine learning models, employing the SHAP method for interpretability analysis, and utilizing decision curve analysis to evaluate model discrimination and clinical net benefit, we ultimately aim to provide a multimodal, interpretable, and precise assessment tool that integrates imaging indicators and inflammatory burden for stroke risk stratification in patients with Atrial Fibrillation (AF).

2. Methods

2.1. Study population

This study employed a retrospective case-control design. A total of 109 consecutive patients with Non-Valvular Atrial Fibrillation (NVAF) who met the inclusion criteria were enrolled from the Department of Cardiology, Affiliated Hospital of Youjiang Medical University for Nationalities, between October 2023 and October 2025. Patients were divided into two groups based on the presence or absence of ischemic stroke: the case group (NVAF with stroke, n = 46)and the control group (NVAF without stroke, n = 63).

Inclusion criteria: (1) Diagnosis of AF confirmed by Electrocardiography (ECG) or Holter monitoring, consistent with the guidelines outlined in the 2023 Chinese Guidelines for the Diagnosis and Management of Atrial Fibrillation; (2) Diagnosis of ischemic stroke confirmed by cranial Computed Tomography (CT) or Magnetic Resonance Imaging (MRI), according to the 2023 Chinese Guidelines for the Diagnosis and Treatment of Acute Ischemic Stroke; (3) Complete clinical data, including Two-dimensional Speckle-Tracking Echocardiography (2D-STE) data and inflammatory markers; (4) Informed consent obtained from the patient or their legal representative.

Exclusion criteria: (1) Concomitant severe hepatic or renal dysfunction, malignancy, or autoimmune disease; (2) Contraindications to echocardiography or poor image quality precluding speckle-tracking analysis; (3) History of acute infection, surgery, or trauma within the preceding month; (4) Missing key clinical data.

For patients in the stroke group, all echocardiographic examinations and blood sample collections were performed within 24 hours of admission. Therefore, the models constructed in this study reflect atrial function and inflammatory status characteristics following a stroke event,aiming to investigate the discriminative ability of these markers for stroke patients.

2.2. Data collection and measurement

2.2.1. Clinical baseline data collection

Baseline clinical information, including sex, age, comorbidities (hypertension, diabetes mellitus, coronary heart disease, heart failure, etc.), and type of AF (paroxysmal, persistent, permanent), were extracted from the hospital's electronic medical record system. These baseline data were not incorporated into model development but were used solely for descriptive characterization of the clinical cohort.

2.2.2. Two-dimensional speckle-tracking echocardiography

All patients underwent standard transthoracic echocardiography using a Philips EPIQ 7C ultrasound system equipped with an X5-1 transducer (1-5 MHz). Patients were examined in the left lateral decubitus position during quiet breathing. Dynamic two-dimensional images from the apical four-chamber, two-chamber, and three-chamber views were acquired and stored over three consecutive cardiac cycles, with a frame rate > 50 frames per second [14].

Stored images were transferred to dedicated speckle-tracking analysis software (TomTec Arena 2.0) for offline analysis. The Left Atrial (LA) endocardial border was manually traced, and the software automatically tracked speckle motion to compute the LA global longitudinal strain. Left Atrial Reservoir Strain (LASr), Conduit Strain (LAScd), and Contractile Strain (LASc) were obtained, and the average value from three cardiac cycles was used as the final result for each parameter. All measurements were performed independently by two experienced echocardiographers who were blinded to the patients'clinical grouping. If the difference between the two measurements exceeded 10%, a third senior echocardiographer reviewed the images to determine the final value.

2.2.3. Inflammatory marker measurement

Fasting venous blood samples were collected from all patients within 24 hours of admission and prior to the administration of anti-infective therapy. Neutrophil Count (NEUT), Lymphocyte Count (LYM), Monocyte Count (MONO), and Platelet Count (PLT) were measured using an automated hematology analyzer. Based on these measurements, the following inflammatory indices were calculated: Systemic Immune-Inflammation Index (SII) = (Platelet count × Neutrophil count)/Lymphocyte count. Systemic Inflammation Response Index (SIRI) = (Monocyte count × Neutrophil count)/Lymphocyte count [15].

2.3. Statistical analysis and model development

2.3.1. Baseline comparison and feature selection

Statistical analyses were performed using SPSS version 22.0 and R software (version 4.2.2). Normally distributed continuous variables were expressed as mean±standard deviation and compared between groups using the independent samples t-test. Non-normally distributed continuous variables were presented as median (interquartile range) and analyzed using the Mann-Whitney U test. Categorical variables were presented as frequencies (percentages) and compared using the chi-square test or Fisher's exact test. A two-sided p-value < 0.05 was considered statistically significant.

Feature Selection Strategy: Initially, univariate analysis was conducted to compare clinical, echocardiographic, and laboratory parameters between the two groups. Variables with $p < 0.10$ were entered into multivariate logistic regression analysis (using the backward elimination method) to identify factors independently associated with stroke. Based on the results of the multivariate analysis, the identified independent correlates (SIRI, LASr, and SII) were selected as input features for subsequent machine learning models.

2.3.2. Machine learning model development and evaluation

Model development and evaluation were performed using Python 3.8 with machine learning libraries including Scikit-learn (version 1.0.2), XGBoost (version 1.5.0), and LightGBM (version 3.3.2).

Dataset Splitting: The entire cohort of 109 patients was randomly divided into a training set (n = 76) and an independent testing set (n = 33) using a 7:3 ratio. The training set was used for model training and hyperparameter tuning, while the testing set was strictly reserved for evaluating the final models'generalization performance and remained completely isolated throughout the training process [16].

Model Construction: Based on the three selected features (SIRI, LASr, SII), the following 11 machine learning models were constructed using the training set: (1) Logistic Regression, (2) K-Nearest Neighbors, (3) Gaussian Naïve Bayes, (4) Linear Discriminant Analysis, (5) Support Vector Machine (SVM), (6) Random Forest, (7) Gradient Boosting Machine (implemented using the XGBoost algorithm), (8) LightGBM, (9) Multi-Layer Perceptron, (10) AdaBoost, and (11) Partial Least Squares (PLS). This collection of models encompasses major algorithm categories, including linear models, tree-based models,ensemble methods, and neural networks.

Hyperparameter Tuning and Validation: To optimize model performance and mitigate the risk of overfitting, hyperparameter tuning was performed within the training set using 5-fold cross-validation combined with grid search. The optimal hyperparameter combination for each model was selected based on the average cross-validated AUC.

Model Performance Evaluation: The 11 tuned models were used to make predictions on the independent testing set. Classification performance was comprehensively evaluated using accuracy, sensitivity, specificity, and F1-score. Discriminative ability was quantified by the area under the receiver operating characteristic

curve (AUC) along with its 95% Confidence Interval (CI). The DeLong test was employed to compare the statistical significance of differences in AUC between the best-performing model and other models [17].

Model Interpretability and Clinical Utility: SHAP (SHapley Additive exPlanations) analysis was applied to provide global and local interpretations of the final optimal model, quantifying the contribution and direction of influence for each feature variable (SIRI, LASr, SII). Furthermore, Decision Curve Analysis (DCA) was conducted to evaluate the clinical net benefit of the optimal model across different risk thresholds, comparing it against the strategies of "treat all" and "treat none".

3. Results

3.1. Baseline characteristics of the study population and selection of core associated factors

A total of 109 patients with Non-Valvular Atrial Fibrillation (NVAF) were enrolled in this study, including 46 patients in the stroke group and 63 in the non-stroke group. Statistical analysis demonstrated comparability of clinical characteristics between the training and testing sets ($p > 0.05$). Table 1 presents the baseline clinical characteristics of the study participants. No statistically significant differences were observed between the training and testing sets in terms of age, sex, hypertension, coronary heart disease, or heart failure (all $p > 0.05$). However, the distribution of diabetes mellitus differed between the two groups ($p = 0.036$), with a lower proportion of diabetes in the training set (3.95%) compared to the testing set (18.18%).

Univariate analysis revealed no statistically significant differences between the stroke and non-stroke groups in age, sex, or comorbidities (hypertension, diabetes mellitus, coronary heart disease, heart failure) (all $p > 0.05$). In contrast, statistically significant differences were observed between the two groups for Left Atrial Reservoir Strain (LASr), Systemic Immune-inflammation Index (SII), and Systemic Inflammation Response Index (SIRI) (all $p < 0.05$). No significant differences were found for Left Atrial Conduit Strain (LAScd) or Contractile Strain (LASc) between the two groups ($p > 0.05$).

Variables with $p < 0.10$ in the univariate analysis were entered into multivariate logistic regression analysis. The results identified SIRI (OR = 12.08, 95% CI: 2.64–55.22, $p = 0.001$), LASr (OR = 0.84, 95% CI: 0.72–0.99, $p = 0.032$), and SII (OR = 1.01, 95% CI: 1.01–1.01, $p = 0.003$) as factors independently associated with the occurrence of stroke in patients with atrial fibrillation. Based on these findings, these three indicators were selected as input features for subsequent machine learning models.

Table 1. Baseline characteristics of patients

Variables	Total (n = 109)	test (n = 33)	train (n = 76)	Statistic	*p*
Age, Mean ± SD	63.88 ± 11.88	62.12 ± 11.25	64.64 ± 12.13	t = -1.02	0.310
Left atrial anterior-posterior diameter, Mean ± SD	38.75 ± 8.97	36.97 ± 8.46	39.53 ± 9.13	t = -1.37	0.173
Peak E, Mean ± SD	89.74 ± 31.27	82.36 ± 29.72	92.95 ± 31.57	t = -1.64	0.104
Log Nt Probnp, Mean ± SD	6.09 ± 1.63	5.82 ± 1.54	6.21 ± 1.66	t = -1.14	0.255
LASr, Mean ± SD	13.72 ± 8.59	15.85 ± 9.97	12.80 ± 7.81	t = 1.72	0.089
LAScd, Mean ± SD	8.57 ± 3.82	8.43 ± 3.92	8.63 ± 3.80	t = -0.24	0.807
LASc, Mean ± SD	7.86 ± 5.64	8.92 ± 7.80	7.41 ± 4.38	t = 1.29	0.201
SII, Mean ± SD	495.93 ± 225.52	477.54 ± 251.43	503.91 ± 214.60	t = -0.56	0.577
SIRI, Mean ± SD	1.14 ± 0.76	1.07 ± 0.64	1.17 ± 0.80	t = -0.61	0.543

Table 1. Continued

Gender, n(%)				$\chi^2 = 1.99$ 0.158
Male	65 (59.63)	23 (69.70)	42 (55.26)	
Female	44 (40.37)	10 (30.30)	34 (44.74)	
Grouping, n (%)				$\chi^2 = 0.15$ 0.696
0	63 (57.80)	20 (60.61)	43 (56.58)	
1	46 (42.20)	13 (39.39)	33 (43.42)	
Heart failure, n (%)				$\chi^2 = 0.05$ 0.815
0	51 (46.79)	16 (48.48)	35 (46.05)	
1	58 (53.21)	17 (51.52)	41 (53.95)	
hypertension, n (%)				$\chi^2 = 0.00$ 0.954
0	59 (54.13)	18 (54.55)	41 (53.95)	
1	50 (45.87)	15 (45.45)	35 (46.05)	
diabetes, n (%)				$\chi^2 = 4.42$ 0.036
0	100 (91.74)	27 (81.82)	73 (96.05)	
1	9 (8.26)	6 (18.18)	3 (3.95)	
Coronary heart disease, n (%)				$\chi^2 = 0.01$ 0.915
0	57 (52.29)	17 (51.52)	40 (52.63)	

3.2. Comparison of discriminative performance among different machine learning models

Based on the three selected features (SIRI, LASr, SII), risk stratification models were constructed using 11 machine learning algorithms,and their discriminative performance was evaluated in both the training set and the independent testing set.

Model Performance Overview: The detailed performance metrics (accuracy, sensitivity, specificity, PPV, NPV, F1-score,Youden index) for the 11 models in the training and testing sets are presented below. Overall, all models demonstrated a high level of discriminative performance, albeit with some variability.

Training Set Performance: In the training set, the Boosting Method (XGBoost) model exhibited the best comprehensive discriminative performance, with an accuracy of 90.79%, sensitivity of 90.91%, specificity of 90.70%, F1-score of 89.55%, and Youden index of 0.816. The Lasso model and Bayes Method model also showed excellent discriminative ability. The Neighbor Method model demonstrated the weakest performance, with an accuracy of only 68.42% and a Youden index of 0.357, significantly inferior to the other models.

Training Set AUC Comparison: The discriminative ability of each model was further assessed using the area under the receiver operating characteristic curve (AUC). The Bayes Method achieved the highest AUC in the training set, at 0.969 (95% CI: 0.939–0.999), followed by Lasso (0.949, 95% CI: 0.903–0.994), Boosting Method (0.941, 95% CI: 0.887–0.995), Logistic Model (0.939, 95% CI: 0.889–0.988), and Discriminant Model (0.936, 95% CI: 0.883–0.989). Random Forest, Gradient Boosting, SVM_Kernel, Neural Net, and PLS Model all achieved AUCs above 0.91, indicating good discriminative performance. Consistent with the previous findings, the Neighbor Method had the lowest AUC (0.814, 95% CI: 0.722–0.907).

Testing Set Performance: The performance metrics for the 11 models in the independent testing set are shown in Table 2 In the testing set, the Logistic Model and Lasso models achieved the highest classification accuracy (both 87.88%), with a sensitivity of 76.92%, specificity of 95.00%, PPV of 90.91%, NPV of 86.36%, F1-score of 0.8333, and Youden index of 0.719, demonstrating the most outstanding performance among all

models. The Random Forest, Boosting Method, and Neural Net models showed identical performance metrics (accuracy 84.85%, sensitivity 69.23%, specificity 95.00%, F1-score 0.7826), indicating highly similar stratification capabilities. The Gradient Boosting, SVM_Kernel, and Neighbor Method models also exhibited identical performance metrics (accuracy 84.85%, sensitivity 76.92%, specificity 90.00%, F1-score 0.8000). The PLS Model, Bayes Method, and Discriminant Model each achieved an accuracy of 81.82%;among these, Bayes Method had relatively higher sensitivity (76.92%) but lower specificity (85.00%), while PLS Model had the highest specificity (95.00%) but the lowest sensitivity (61.54%).

Table 2. Performance indicators of 11 models in the test set

Models	Sensitivity	Specificity	Accuracy	PPV	NPV	F1	Youden's index
Random Forest	0.6923	0.9500	0.8485	0.9000	0.8261	0.7826	0.6423
Gradient Boosting	0.7692	0.9000	0.8485	0.8333	0.8571	0.8000	0.6692
SVM_Kernel	0.7692	0.9000	0.8485	0.8333	0.8571	0.8000	0.6692
Logistic Model	0.7692	0.9500	0.8788	0.9091	0.8636	0.8333	0.7192
Neighbor Method	0.7692	0.9000	0.8485	0.8333	0.8571	0.8000	0.6692
PLS Model	0.6154	0.9500	0.8182	0.8889	0.7917	0.7273	0.5654
Boosting Method	0.6923	0.9500	0.8485	0.9000	0.8261	0.7826	0.6423
Neural Net	0.6923	0.9500	0.8485	0.9000	0.8261	0.7826	0.6423
Bayes Method	0.7692	0.8500	0.8182	0.7692	0.8500	0.7692	0.6192
Discriminant Model	0.6923	0.9000	0.8182	0.8182	0.8182	0.7500	0.5923
Lasso	0.7692	0.9500	0.8788	0.9091	0.8636	0.8333	0.7192

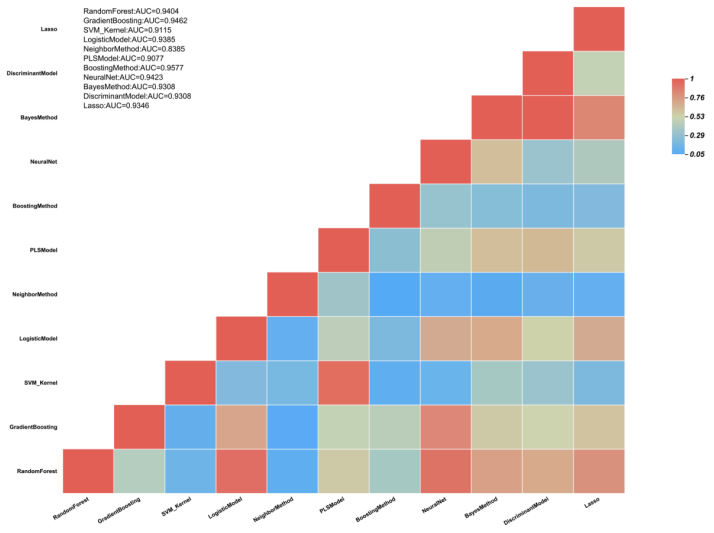

Figure 1. Comparison of ROC curves for the test set

Testing Set AUC Comparison: A comparison of the ROC curves for the testing set is shown in Figure 1, and the AUC values with 95% confidence intervals for each model are presented in Figure 2 (forest plot). The Boosting Method attained the highest AUC in the testing set, at 0.958 (95% CI: 0.895–1.000), followed by Gradient Boosting (0.946, 95% CI: 0.865–1.000), Neural Net (0.942,95% CI: 0.863–1.000), Random Forest

(0.940, 95% CI: 0.852–1.000), and Logistic Model (0.938, 95% CI: 0.861–1.000). The Lasso (0.935), Bayes Method (0.931), and Discriminant Model (0.931) also achieved AUCs exceeding 0.93. Although the performance of most models in the testing set was slightly lower than in the training set, the magnitude of fluctuation was small, suggesting good generalization ability without significant overfitting.

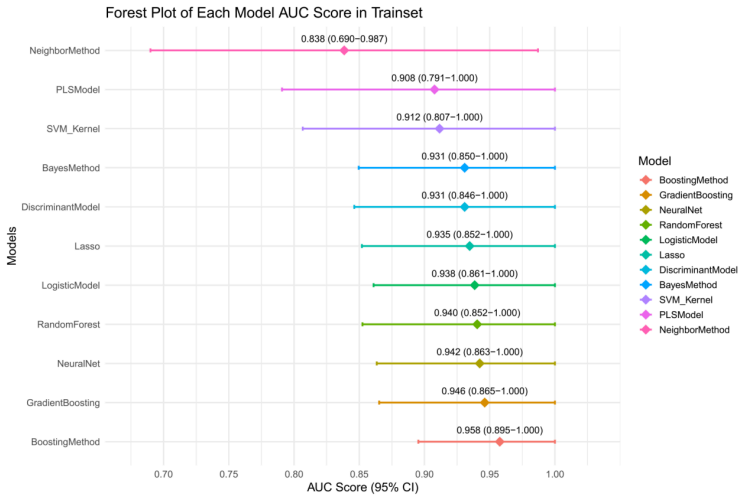

Figure 2. Forest plot of AUC for the test set

3.3. Model calibration assessment

Calibration curves were employed to assess the agreement between the predicted probabilities from each model and the observed incidence of stroke. In the training set, the calibration curves for most models (e.g., Logistic Model, Lasso, Bayes Method) closely aligned with the ideal diagonal line, indicating good calibration of their predicted probabilities. Ensemble models (e.g., Boosting Method, Random Forest) showed slight deviations in the low-probability region, but overall remained within an acceptable range. The curve for the Neighbor Method deviated markedly from the diagonal, consistent with its poor discriminative performance.

In the testing set (Figure 3), Logistic Model and Lasso continued to exhibit the best calibration, with curves nearly coinciding with the diagonal, demonstrating that their predicted probabilities remained reliable on unseen data. Boosting Method and Random Forest showed minor fluctuations in calibration, but were generally acceptable. The calibration curve for the Neighbor Method deviated severely, further confirming its insufficient generalization ability.

Taking Random Forest, which demonstrated the best overall performance, as an example, Figures 4 display its calibration curves in the training and testing sets, respectively. The model showed good calibration in the training set, and although there was a slight deviation in the testing set, it remained within an acceptable range overall.

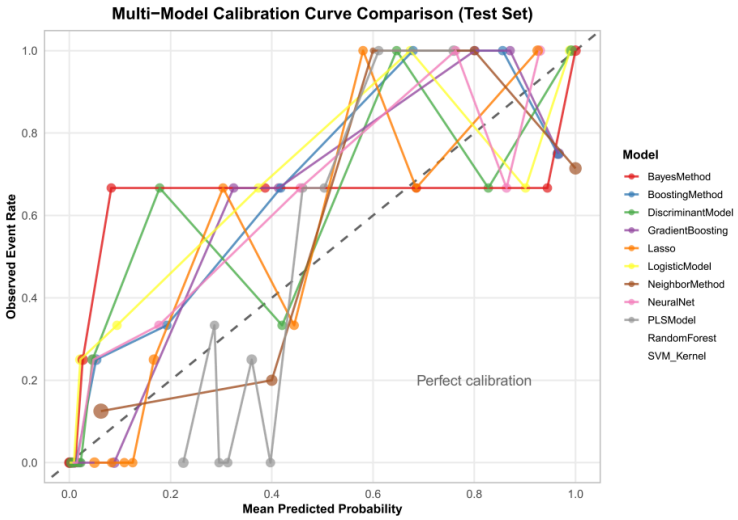

Figure 3. Calibration curves of multiple models for the test set

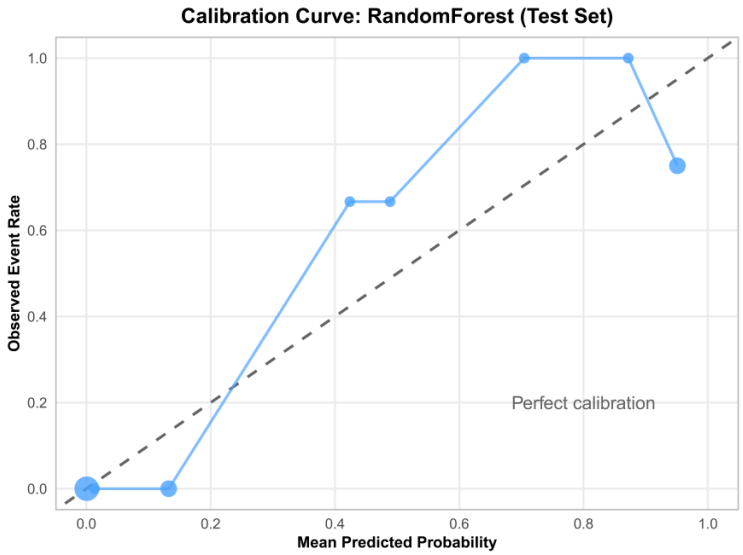

Figure 4. Calibration curve of the best model test set

3.4. Confusion matrix validation of the optimal model

Confusion matrices were employed to visually validate the classification performance of one of the optimal models (Random Forest). In the training set, the model correctly classified 37 non-stroke patients (true negatives) and 32 stroke patients (true positives), while misclassifying 6 stroke patients as negative (false negatives) and 1 non-stroke patient as positive (false positive), yielding a training set accuracy of 90.8%. In the testing set (Figure 5), the model correctly classified 19 non-stroke patients and 10 stroke patients, with 1 stroke patient misclassified as negative and 3 non-stroke patients misclassified as positive, resulting in a testing set accuracy of 87.9%. These results demonstrate that the optimal model possesses good discriminative ability for identifying stroke and non-stroke patients, with a low misclassification rate.

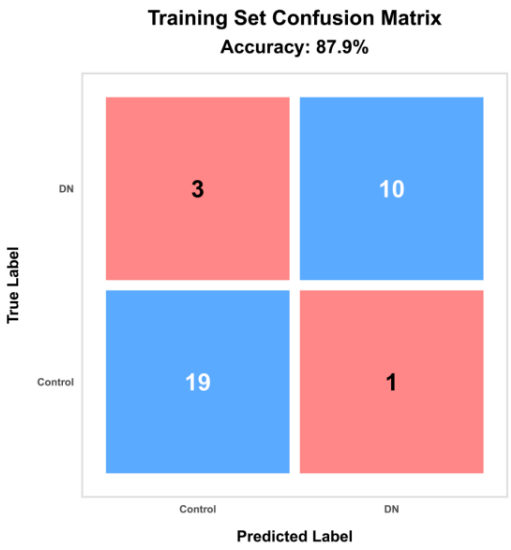

Figure 5. Confusion matrix of test set

3.5. Model interpretability analysis:feature contribution and local explanation

The SHAP algorithm was employed to perform interpretability analysis on the optimal model, quantifying the contribution and direction of influence of each input feature on the model's output.

Global Feature Importance: As shown in Figure 6, among the three input features, SIRI contributed the most to the model output, with a mean absolute SHAP value of 0.160, followed by LASr (0.143) and SII (0.133).

Direction of Feature Influence: The SHAP beeswarm plot (Figure 7) intuitively illustrates the relationship between feature values and the model output. High SIRI values (red) corresponded to positive SHAP values, indicating that elevated SIRI is associated with increased stroke risk. Low LASr values (blue) corresponded to positive SHAP values, suggesting that reduced LASr (impaired left atrial function) is associated with increased stroke risk. High SII values also positively drove the model output. These findings were highly consistent with clinical expectations.

Local Explanation Example: Figure 8 presents a SHAP force plot shown. For this patient, the low LASr value (8.7%) made the largest positive contribution to the model output (+0.124), followed by a high SII value (518) contributing + 0.058, and a high SIRI value (1.03) contributing + 0.015. The combined effect of these features increased the model's predicted risk probability from a baseline value of 0.437 to 0.635. This example intuitively demonstrates the driving effect of core features on individual risk prediction.

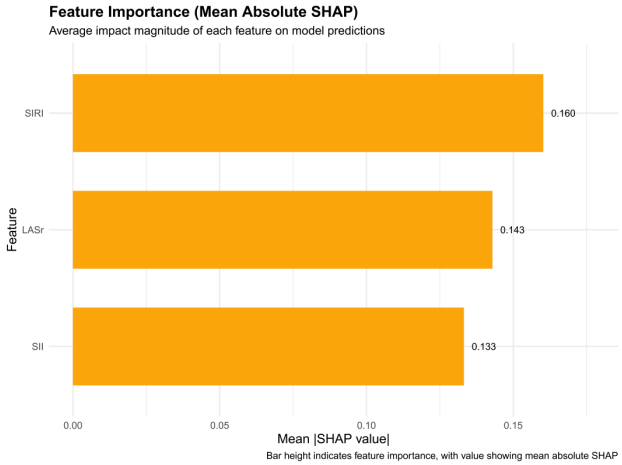

Figure 6. Bar chart of SHAP feature importance

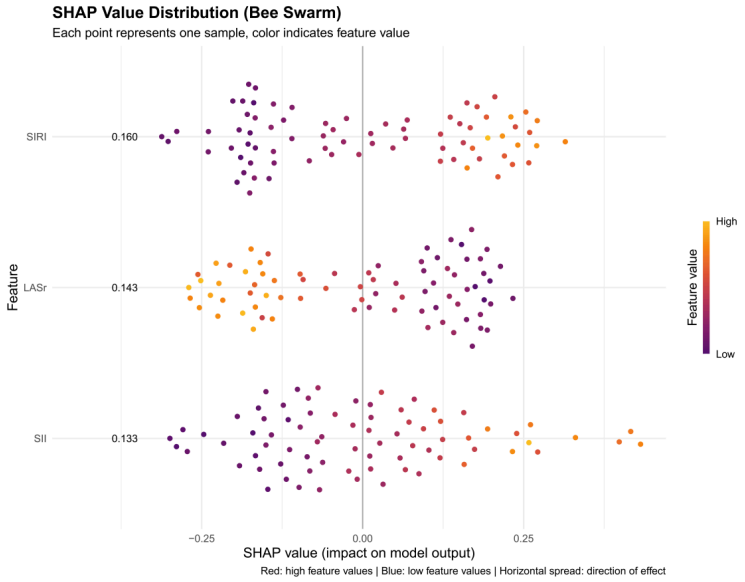

Figure 7. SHAP swarm chart

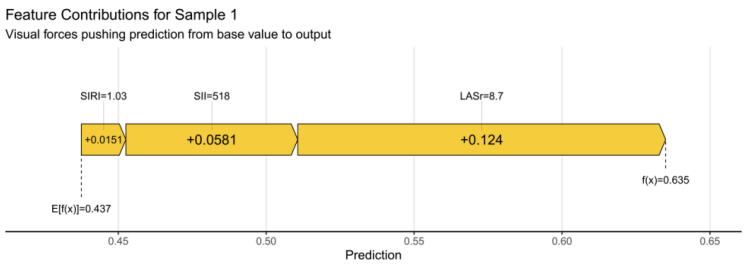

Figure 8. SHAP force diagram of a typical patient in the stroke group

3.6. Clinical utility assessment:decision curve analysis

Decision Curve Analysis: Decision Curve Analysis (DCA) was performed to evaluate the clinical net benefit of each model. In the training set, within a clinically reasonable range of risk thresholds (0.1–0.6), the net benefit curves for most models lay above the strategies of "treat all" and "treat none", indicating that these models possess good clinical utility. In the testing set (Figure 9), models including Logistic Model, Lasso, and Boosting Method consistently maintained a high net benefit across the threshold range of 0.2–0.6.

Clinical Impact Curve: To further evaluate the models'value in real-world clinical decision-making, clinical impact curves were plotted. In the training set, the optimal model (Random Forest) demonstrated that, across a wide range of risk thresholds (0.2–0.6), the proportion of actual stroke patients among those predicted to be at high risk was significantly higher than that expected by a random strategy, indicating that the model can effectively enrich for high-risk individuals. In the testing set (Figure 10), the model also maintained good clinical net benefit, with particularly good agreement between the number of predicted positive cases and actual positive cases within the threshold range of 0.3–0.5. This suggests that applying the model within this threshold range may help avoid overtreatment or undertreatment.

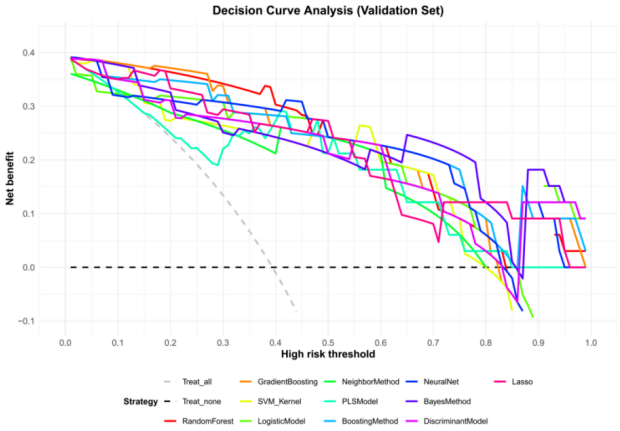

Figure 9. Decision curve of the test set

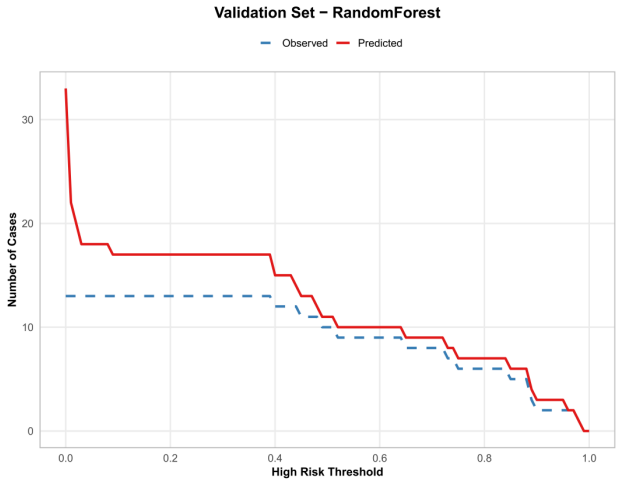

Figure 10. Clinical impact curve of the test set for the best model(random forest)

4. Discussion

The occurrence of stroke in patients with Atrial Fibrillation (AF) results from the interplay between hemodynamic abnormalities and inflammatory responses. Two-Dimensional Speckle-Tracking Echocardiography (2D-STE), as a non-invasive and precise ultrasound technique, can sensitively reflect myocardial microstructural changes and functional status. Among the parameters derived from this technique, Left Atrial Reservoir Strain (LASr) is a core indicator for assessing Left Atrial (LA) reservoir function. A reduction in LASr suggests LA myocardial fibrosis, decreased compliance,and impaired reservoir function, which may lead to blood stasis within the LA, promoting thrombus formation and thereby increasing the risk of stroke [18]. In the present study, LASr was significantly lower in the stroke group compared to the non-stroke group, consistent with previous studies [19].

Mannina et al. [20] conducted a 10.9-year follow-up study in 806 community-dwelling individuals without a history of atrial fibrillation or stroke and found that those with reduced peak Left Atrial Longitudinal Strain (LASr) (lowest quintile) had a 3.12-fold increased risk of new-onset ischemic stroke. Notably, this study population excluded patients with AF, implying that even in the absence of AF—a "classic source of thrombosis"—isolated LA functional impairment (reduced LASr) is sufficient to significantly increase stroke risk [20]. By extension, in patients with AF, where atrial activity and function are further disrupted, reduced LASr may signify more severe atrial myopathy and blood stasis, potentially conferring an even higher stroke risk. Vera et al. [18], in a prospective follow-up of 92 patients with Cryptogenic Stroke (CS), reported that patients with LASr < 23% had significantly higher rates of stroke recurrence or mortality during a 28-month follow-up period (log-rank $p = 0.009$), and LASr was an independent predictor of outcomes after adjusting for clinical risk factors (HR = 0.90). Bhat et al. [21] similarly confirmed the association between LASr and recurrent thrombotic events in a CS cohort. In patients with AF, varying degrees of left atrial dysfunction are frequently encountered.Furthermore, the prognostic relevance of LASr in CS populations is believed to arise from underlying subclinical AF or atrial cardiomyopathy. Therefore, utilizing LASr for risk stratification in AF patients is well grounded in pathophysiology [21].

SII and SIRI, as novel composite inflammatory indices, integrate counts of neutrophils, lymphocytes, and platelets, providing a more comprehensive reflection of systemic inflammatory activation compared to single inflammatory markers [22]. In this study, SIRI demonstrated the highest contribution in the SHAP analysis

(0.160), with high values positively driving stroke risk, indicating that inflammatory burden plays a critical role in the development of AF-related stroke. The biological plausibility of this finding is supported by previous research: neutrophils play a key role in the inflammatory response of atherosclerosis, secreting numerous inflammatory mediators, chemokines, and oxygen radicals that can lead to endothelial cell damage and subsequent tissue ischemia [23, 24]; monocyte activation and their transformation into lipid-laden macrophages are essential processes in atherosclerotic plaque formation [25]; while lymphocytes exert inhibitory effects in inflammatory regulation, potentially conferring protective effects against atherosclerosis [26]. Thus, the association between SIRI and stroke risk in AF patients has a clear biological foundation [27].

These findings are highly consistent with recent studies on inflammatory indices predicting cardiovascular events [28]. Xia et al. [29] published a large-scale cohort study based on the NHANES database, enrolling 42,875 US adults with up to 20 years of follow-up, systematically evaluating the associations of SII and SIRI with mortality risk. The results showed that individuals with SII > 655.56 had a 29% increased risk of all-cause mortality (HR 1.29) and a 33% increased risk of cardiovascular mortality (HR 1.33); those with SIRI > 1.43 had a 55% increased risk of cardiovascular mortality (HR 1.55) [29]. The strengths of that study lie in its large sample size, long follow-up period, and inclusion of a healthy population free of baseline cardiovascular disease, enabling a more reliable assessment of the impact of inflammatory burden itself on long-term prognosis.

Integrating the findings of Xia et al. [29] with the results of the present study forms a coherent evidence chain:in healthy populations without cardiovascular disease, systemic inflammation (elevated SII, SIRI) predicts an increased long-term risk of cardiovascular death; whereas in populations already diagnosed with AF, inflammation not only contributes to the maintenance and progression of AF but is also closely linked to the occurrence of thrombotic events such as stroke. The underlying mechanism involves inflammation creating a "prothrombotic microenvironment" through platelet activation,vascular endothelial injury, and upregulation of tissue factor expression. In AF patients with pre-existing atrial blood stasis, superimposing this prothrombotic environment dramatically increases the risk of thrombus formation. Given their simplicity and cost-effectiveness, SII and SIRI show considerable potential as composite inflammatory biomarkers for assessing thrombotic risk and guiding anticoagulation decisions in AF patients. Subsequent studies should explore whether adding these indices to the CHA_2DS_2-VASc score could improve its ability to predict thrombotic events [30].

This study represents the first attempt to integrate LA strain parameters with inflammatory indices, employing 11 machine learning algorithms to construct stroke risk stratification models for patients with AF, and systematically comparing the discriminative performance of these models. The results revealed that in the training set, the Random Forest model exhibited optimal performance; however, in the independent testing set, the Logistic Model and Lasso models demonstrated the best overall discriminative performance, with an accuracy of 87.88%and leading AUC and F1-score values. This finding carries important methodological implications: Logistic Model, as a traditional statistical model, features a simple structure and strong interpretability, making it less prone to overfitting in small-sample clinical data; the Lasso model, through L1 regularization, automatically performs feature selection, effectively controlling model complexity. The superior performance of these two models in the testing set suggests that for clinical studies with limited sample sizes, linear models or regularized models may possess better generalization ability compared to complex ensemble learning models. This observation aligns with the conclusions of recent studies comparing machine learning with traditional regression, which noted that when the sample size is < 200, logistic regression tends to be more robust than complex models like Random Forest. Therefore, for clinical application, we recommend prioritizing Logistic Model or Lasso models as risk stratification tools.

SHAP interpretability analysis revealed a gradient in the contribution of SIRI, LASr, and SII to the model output (SIRI > LASr > SII), with directions of influence highly consistent with clinical expectations: high SIRI reflecting increased systemic inflammatory burden, low LASr indicating impaired LA function, and high SII reflecting activation of the immune-inflammation-coagulation axis—all positively driving stroke risk. From a pathophysiological perspective, elevated SIRI signifies activation of innate immunity mediated by neutrophils and monocytes, which can release inflammatory cytokines such as IL-6 and TNF-α, damaging vascular endothelium and promoting atherosclerosis and thrombosis [31]; SII further integrates platelet information, directly reflecting the interaction between platelet activation and inflammation, with high SII suggesting a hypercoagulable state [32]; while reduced LASr represents LA myocardial fibrosis and loss of reservoir function, leading to intra-atrial blood stasis. Through SHAP analysis, this study provides data-driven evidence for the synergistic role of these three dimensions in AF-related stroke, offering a multidimensional theoretical basis for clinical risk stratification.

The Random Forest model developed in this study can be directly applied in clinical practice. Its input features require only LASr (obtainable from routine echocardiography), SII, and SIRI (derivable from complete blood count), ensuring high clinical accessibility. According to decision curve analysis, across a broad range of risk thresholds (0.2–0.6), the clinical net benefit of the model was significantly higher than the strategies of "treat all" or "treat none". This implies that physicians can select an appropriate intervention threshold within the 0.2–0.6 range based on the model's output risk probability, combined with individual patient factors (e.g., bleeding risk,patient preference). For example, if a threshold of 0.4 is chosen, patients with a model-output risk > 0.4 would be recommended for anticoagulation, while those with risk < 0.4 could defer anticoagulation with close follow-up, thereby enabling precise stratification and individualized decision-making. Furthermore, the excellent calibration of the Random Forest model ensures the reliability of these probability estimates, allowing physicians to directly use the absolute values output by the model for risk communication. Future development of a web-based calculator or mobile application based on this model could further facilitate its clinical translation.

In this study, model performance was generally superior in the training set compared to the testing set, consistent with the objective laws of machine learning model training. During the training phase, models can iteratively optimize to adequately fit the characteristics of the training set, potentially adapting to some noise; hence, the Boosting Method achieved an accuracy of 90.79% in the training set. In contrast,testing set data are completely unseen during training, and models can only rely on general patterns learned from the training set for prediction, making a slight decrease in performance a normal phenomenon. Notably, the magnitude of the performance difference between the training and testing sets in this study (an accuracy difference of approximately 3 percentage points) was within a reasonable range, and Random Forest maintained excellent discrimination (AUC 0.94) and calibration in the testing set, indicating that the feature associations learned by the model have practical clinical significance and that no severe overfitting occurred. Another possible reason for the slightly lower testing set performance is the relatively small sample size (n = 33), which may lead to chance variations in feature distribution; future studies with larger sample sizes are needed to further validate model stability.

This study has several limitations. First, its single-center retrospective design and relatively limited sample size (total 109 patients, testing set 33) may introduce selection bias, and the wide confidence intervals for the testing set AUC suggest that the findings require further validation in multicenter prospective cohorts. Second, the model input features were limited to three core indicators (LASr, SII, SIRI), without integrating commonly used clinical parameters such as the CHA_2DS_2-VASc score, LA volume index, or coagulation function indices; future research could construct comprehensive models incorporating more variables to explore optimal feature

combinations. Third, inherent measurement errors in echocardiography exist; although reproducibility was ensured through independent measurements by two observers and adjudication, image quality and operator experience may still influence LASr measurements. Fourth, the retrospective design precludes establishing causality; the ultrasound and blood parameters in this study were collected after the stroke event, reflecting post-event associations rather than pre-event predictions. Prospective cohort studies are needed to validate the true predictive value of this model. Fifth, there was a statistically significant difference in diabetes prevalence between the training and testing sets ($p = 0.036$), with a higher proportion of diabetes in the testing set. This difference may result from chance variation due to the small sample size during random splitting. Nevertheless, the model maintained excellent discriminative performance (AUC 0.94) and calibration in the testing set, suggesting that the predictive ability of the core features (LASr, SII, SIRI) for stroke risk is robust and not substantially affected by differences in baseline characteristic distributions. Future studies with larger samples and multicenter external validation are warranted to confirm the model's stability.

5. Conclusion

Among the 11 machine learning models evaluated, the Random Forest model demonstrated the best overall performance. It achieved an accuracy of 84.21%, sensitivity of 78.79%, and specificity of 88.37% in the training set, and an AUC of 0.940 (95% CI: 0.852–1.000) in the independent testing set, placing it within the top tier of ensemble models alongside Boosting Method (0.958) and Gradient Boosting (0.946). Critically, the Random Forest model exhibited excellent calibration—its calibration curve closely aligned with the ideal diagonal line,indicating a high degree of agreement between predicted probabilities and observed stroke incidence, which was markedly superior to other ensemble models.

SIRI, LASr, and SII were identified as the core features driving the model's risk stratification output. Their directions of influence (high SIRI, low LASr, and high SII positively driving risk) are highly consistent with clinical pathophysiological mechanisms, establishing them as key biomarkers for stroke risk stratification in patients with atrial fibrillation. This integrated risk stratification model provides imaging-inflammatory multidimensional evidence to guide individualized anticoagulation decisions for AF patients in clinical practice, offering significant potential for clinical translation.

Future research should focus on conducting multicenter, large-sample prospective studies to further validate the model's external validity, explore the inclusion of additional clinical variables (e.g., CHA_2DS_2-VASc score, left atrial volume index) to optimize model performance, and develop web-based or mobile-accessible calculation tools to facilitate its widespread adoption in clinical practice.

References

[1] Paludan-Müller, C., Vad, O. B., Stampe, N. K., Diederichsen, S. Z., Andreasen, L., Monfort, L. M., Fosbøl, E. L., Køber, L., Torp-Pedersen, C., Svendsen, J. H., Olesen, M. S. (2024). Atrial fibrillation: Age at diagnosis, incident cardiovascular events, and mortality. *European Heart Journal, 45*(24). 2119–2129

[2] Chan, J. (2024). *CHA_2DS_2-VASc score for atrial fibrillation stroke risk* [updated]. Evidence to Action: Official Journal of MDCalc. 042024(01)

[3] Lip, G. Y. H., Teppo, K., Nielsen, P. B. (2024). CHA2DS2-VASc or a non-sex score (CHA2DS2-VA) for stroke risk prediction in atrial fibrillation: Contemporary insights and clinical implications. *European Heart Journal, 45*(36). 3718–3720

[4] Serna, M. J., Rivera-Caravaca, J. M., López-Gálvez, R., Soler-Espejo, E., Lip, G. Y. H., Marín, F., Roldán, V. (2024). Dynamic assessment of CHA2DS2-VASc and HAS-BLED scores for predicting ischemic stroke and

major bleeding in atrial fibrillation patients. *Revista Española de Cardiología (English Edition), 77*(10). 835–842

[5] Szymanski, F. M., Lip, G. Y. H., Filipiak, K. J., Platek, A. E., Hrynkiewicz-Szymanska, A., Opolski, G. (2015). Stroke risk factors beyond the CHA2DS2-VASc score: Can we improve our identification of "high stroke risk" patients with atrial fibrillation? *The American Journal of Cardiology, 116*(11). 1781–1788

[6] Anagnostopoulos, I., Kousta, M., Kossyvakis, C., Paraskevaidis, N. T., Schizas, N., Vrachatis, D., Deftereos, S., Giannopoulos, G. (2023). Atrial strain and occult atrial fibrillation in cryptogenic stroke patients: A systematic review and meta-analysis. *Clinical Research in Cardiology, 112*(11). 1600–1609

[7] Obokata, M., Negishi, K., Kurosawa, K., Tateno, R., Tange, S., Arai, M., Amano, M., Kurabayashi, M. (2014). Left atrial strain provides incremental value for embolism risk stratification over CHA2DS2-VASc score and indicates prognostic impact in patients with atrial fibrillation. *Journal of the American Society of Echocardiography, 27*(7). 709-716.e4

[8] Park, J., Hwang, I., Park, J. J., Park, J., Cho, G. (2021). Left atrial strain to predict stroke in patients with acute heart failure and sinus rhythm. *Journal of the American Heart Association, 10*(13). e020414

[9] Sonaglioni, A., Nicolosi, G. L. (2025). Left atrial reservoir strain in cardiovascular and systemic disease: Advances and clinical applications from physiology to practice. *Reviews in Cardiovascular Medicine, 26*(12). 46198

[10] Abduch, M. C. D., Alencar, A. M., Mathias Jr., W., Vieira, M. L. D. C. (2014). *Cardiac mechanics evaluated by speckle tracking echocardiography*. Arquivos Brasileiros de Cardiologia

[11] Li, N., Zhou, H., Tang, Q. (2017). Red blood cell distribution width: A novel predictive indicator for cardiovascular and cerebrovascular diseases. *Disease Markers*. 2017. 1–23

[12] Lyngbakken, M. N., Myhre, P. L., Røsjø, H., Omland, T. (2019). Novel biomarkers of cardiovascular disease: Applications in clinical practice. *Critical Reviews in Clinical Laboratory Sciences, 56*(1). 33–60

[13] Jin, Z., Wu, Q., Chen, S., Gao, J., Li, X., Zhang, X., Zhou, Y., He, D., Cheng, Z., Zhu, Y., Wu, S. (2021). The associations of two novel inflammation indexes, SII and SIRI with the risks for cardiovascular diseases and all-cause mortality: A ten-year follow-up study in 85, 154 individuals. *Journal of Inflammation Research. Volume 14.* 131–140

[14] Gala, D., Kim, E. J., Radfar, N., Makaryus, A. N. (2025). The role of left atrial strain in patients with low CHA2DS2-VASc scores: Refining stroke risk assessment. *Journal of the American Society of Echocardiography, 38*(11). 1090–1098

[15] Jiang, P., Chen, J., Li, J. (2025). Association of the systemic immune-inflammatory index and systemic inflammatory response index with all-cause and cardiovascular mortality in individuals with metabolic inflammatory syndrome. *European Journal of Medical Research, 30*(1). 444

[16] Brattain, L. J., Telfer, B. A., Dhyani, M., Grajo, J. R., Samir, A. E. (2018). Machine learning for medical ultrasound: Status, methods, and future opportunities. *Abdominal radiology (New York), 43*(4). 786–799

[17] Demler, O. V., Pencina, M. J., D'Agostino, R. B. (2012). Misuse of DeLong test to compare AUCs for nested models. *Statistics in Medicine, 31*(23). 2577–2587

[18] Vera, A., Cecconi, A., Ximénez-Carrillo, Á., Ramos, C., Martínez-Vives, P., Lopez-Melgar, B., Sanz-García, A., Ortega, G., Aguirre, C., Montes, Á., Vivancos, J., Jiménez-Borreguero, L. J., Alfonso, F. (2024). Left atrial strain predicts stroke recurrence and death in patients with cryptogenic stroke. *The American Journal of Cardiology, 210*. 51–57

[19] Bufano, G., Radico, F., D'Angelo, C., Pierfelice, F., De Angelis, M. V., Faustino, M., Pierdomenico, S. D., Gallina, S., Renda, G. (2022). Predictive value of left atrial and ventricular strain for the detection of atrial fibrillation in patients with cryptogenic stroke. *Frontiers in Cardiovascular Medicine. 9*. 869076

[20] Mannina, C., Ito, K., Jin, Z., Yoshida, Y., Matsumoto, K., Shames, S., Russo, C., Elkind, M. S. V., Rundek, T., Yoshita, M., DeCarli, C., Wright, C. B., Homma, S., Sacco, R. L., Di Tullio, M. R. (2023). Association of left atrial strain with ischemic stroke risk in older adults. *JAMA Cardiology, 8*(4). 317

[21] Bhat, A., Chen, H. H. L., Khanna, S., Mahajan, V., Gupta, A., Burdusel, C., Wolfe, N., Lee, L., Gan, G. C. H., Dobbins, T., MacIntyre, C. R., Tan, T. C. (2022). Diagnostic and prognostic value of left atrial function in identification of cardioembolism and prediction of outcomes in patients with cryptogenic stroke. *Journal of the American Society of Echocardiography, 35*(10). 1064–1076

[22] Kurtul, A., Ornek, E. (2019). Platelet to lymphocyte ratio in cardiovascular diseases: A systematic review. *Angiology, 70*(9). 802–818

[23] Abete, I., Lu, Y., Lassale, C., Verschuren, M., Van Der Schouw, Y., Bueno-de-Mesquita, B. (2019). White cell counts in relation to mortality in a general population of cohort study in the netherlands: A mediating effect or not? *BMJ Open. 9*(10). e030949

[24] Gill, D., Monori, G., Georgakis, M. K., Tzoulaki, I., Laffan, M. (2018). Genetically determined platelet count and risk of cardiovascular disease: Mendelian randomization study. *Arteriosclerosis, Thrombosis, and Vascular Biology, 38*(12). 2862–2869

[25] Madjid, M., Fatemi, O. (2013). Components of the complete blood count as risk predictors for coronary heart disease. *Texas Heart Institute Journal, 40*(1). 17–29

[26] Wettersten, N., Horiuchi, Y., Maisel, A. (2021). Advancements in biomarkers for cardiovascular disease: Diagnosis, prognosis, and therapy. *Faculty Reviews. 10*

[27] Shah, A. D., Denaxas, S., Nicholas, O., Hingorani, A. D., Hemingway, H. (2016). Low eosinophil and low lymphocyte counts and the incidence of 12 cardiovascular diseases: A CALIBER cohort study. *Open Heart, 3*(2). e000477

[28] Kim, J. H., Lim, S., Park, K. S., Jang, H. C., Choi, S. H. (2017). Total and differential WBC counts are related with coronary artery atherosclerosis and increase the risk for cardiovascular disease in koreans. *PLOS ONE, 12*(7). e0180332

[29] Xia, Y., Xia, C., Wu, L., Li, Z., Li, H., Zhang, J. (2023). Systemic immune inflammation index (SII), system inflammation response index (SIRI) and risk of all-cause mortality and cardiovascular mortality: A 20-year follow-up cohort study of 42, 875 US adults. *Journal of Clinical Medicine, 12*(3). 1128

[30] Watson, T., Shantsila, E., Lip, G. Y. (2009). Mechanisms of thrombogenesis in atrial fibrillation: Virchow's triad revisited. *The Lancet, 373*(9658). 155–166

[31] Cimato, T. R. (2016). Persistent stem cell-driven inflammation in patients with prior MI and stroke. *European Heart Journal.* ehw323

[32] Mancinelli, M., Moscucci, F., Cofini, V., De Nino, A. L., Bocale, R., Savoia, C., Baratta, F., Desideri, G. (2025). Metabolic determinants of systemic inflammation dynamics during hemodialysis: Insights from the systemic immune–inflammation index in a single-center observational study. *Metabolites, 15*(10). 651