BMC Cardiovascular Disorders

RESEARCH

Open Access

Machine-learning model for predicting left atrial thrombus in patients with paroxysmal atrial fibrillation

Wanli Xiong[1,2,3†], Qiqi Cao[1,2,3†], Lu Jia[1,2,3], Min Chen[4], Tao Liu[1,2,3], Qingyan Zhao[1,2,3], Yanhong Tang[1,2,3], Bo Yang[1,2,3], Li Li[5], Shaobo Shi[1,2,3*], He Huang[1,2,3] and Congxin Huang[1,2,3] on behalf of China Atrial Fibrillation Center Project Team

Abstract

Objective Left atrial thrombus (LAT) poses a significant risk for stroke and other thromboembolic complication in patients with atrial fibrillation (AF). This study aimed to evaluate the incidence and predictors of LAT in patients with paroxysmal AF, utilizing machine learning techniques based on data from the Chinese Atrial Fibrillation study.

Methods A large-scale multi-center retrospective study was conducted involving patients diagnosed with non-valvular paroxysmal AF. LAT incidence was assessed, and potential risk factors were analyzed. Machine learning algorithms, including decision tree, random forest, AdaBoost, k-Nearest Neighbor, and logistic regression, were employed to develop a predictive model for LAT.

Results Of the 49,515 patients with paroxysmal AF, 1,058 patients (2.1%, 95% CI 2.0%-2.3%) were identified with LAT. Sixty-one variables were initially included to train machine learning models, with the random forest algorithm demonstrating the best predictive performance (AUC 0.833, 95%CI 0.730–0.924). The final model, refined to include nine essential features, achieved an AUC of 0.787 (95%CI 0.670–0.883). Calibration analysis indicated no significant difference between predicted and observed values ($p = 0.181$). The median predicted probabilities of LAT across quintiles were 2.3%, 7.0%, 11.8%, 16.6%, and 21.5%.

Conclusion This simplified prediction model effectively identifies the risk of LAT in patients with paroxysmal AF, providing a valuable tool for clinical decision-making. Further studies are needed to explore AF management and risk stratification in other AF subtypes.

Keywords Paroxysmal atrial fibrillation, Left atrial thrombus, Transesophageal echocardiography, Machine-learning model

†Wanli Xiong and Qiqi Cao contributed equally to this work.

*Correspondence:
Shaobo Shi
shiyige@whu.edu.cn
Full list of author information is available at the end of the article

Introduction

Atrial fibrillation (AF) is the most prevalent cardiac arrhythmia and is associated with substantial morbidity and mortality. Globally, it affects approximately 33 million individuals, with nearly 5 million new cases diagnosed annually [1]. It significantly increases the risk of thromboembolic events, particularly ischemic stroke, which can be life-threatening [2]. Alarmingly, nearly one-third of patients with AF are asymptomatic [3], making early detection and intervention crucial for mitigating complications. This underscores the necessity for effective AF screening within public health strategies aimed at enhancing patient outcomes.

The identification of thrombus formation in the left atrium is essential for the prevention of stroke and other thromboembolic complications, with transesophageal echocardiography (TEE) remaining the gold standard for detecting left atrial thrombus (LAT) [4, 5]. TEE exhibits a sensitivity of 100% and a specificity of 99% for diagnosing LAT, rendering it a reliable diagnostic modality for clinical practice [6]. Paroxysmal AF is characterized by episodes lasting less than 7 days, often resolving spontaneously. While it is widely recognized that non-paroxysmal AF poses a higher risk of thrombus formation and subsequent embolism, which thought to be associated with more severe hemodynamic abnormalities [7]. Epidemiological data regarding thrombus formation in paroxysmal AF are limited and inconsistency, with a lack of comprehensive research into the risk factors associated with LAT in this patient population. In response to this knowledge gap, the present study aims to identify key risk factors for LAT in patients with paroxysmal AF using machine learning (ML) algorithms. By leveraging data from a large multicenter cohort, this study seeks to develop a clinically applicable predictive model to enhance systematic management of paroxysmal AF patients, optimize risk stratification, and guide clinical decision-making.

Materials and methods

Study design and population

This study was part of the real-world study of Chinese atrial fibrillation (RWS-CAF; registration number: ChiCTR1900021250,registered on February 3, 2019) [8], and was conducted in accordance with the Declaration of Helsinki. It received approval from the Ethical Review Committee of Renmin Hospital of Wuhan University. Between September 2018 and September 2023, consecutive patients were retrieved from the China Atrial Fibrillation Center database using the keywords "paroxysmal atrial fibrillation" and "transesophageal echocardiography" (https://www.china-afc.org/). The final study cohort consisted of patients with non-valvular paroxysmal AF who underwent transesophageal echocardiography. The occurrence of LAT as a clinical event was analyzed using ML algorithms to establish a predictive model based on clinical characteristics. Given that this research is retrospective, basic information regarding all patients were not disclosed throughout the study, thus the requirement for informed consent was waived by the Ethics Commission. We adhered to the STROBE statement for reporting observational studies.

Data collection and clinical variables

The diagnosis of AF was based on the 2020 European Society of Cardiology Guidelines for the diagnosis and management of atrial fibrillation [9], generally involving tracking with a 12-lead electrocardiogram for at least 30 s. Paroxysmal AF is defined as AF that terminates spontaneously or through intervention within 7 days of onset. In this study, all references to AF pertain to non-valvular AF, specifically excluding cases with moderate or severe mitral stenosis or those following mechanical valve replacement surgery. The patient selection flowchart is depicted in Fig. 1. Patients with incomplete demographic characteristics, medical history, laboratory tests, echocardiograms, and treatment data were excluded from the analysis. Medical history such as hypertension, diabetes mellitus, coronary heart disease (CHD), stroke/transient ischemic attack (TIA), and chronic heart failure (CHF) were diagnosed according to relevant guidelines. The CHA_2DS_2-VASc score was utilized to evaluate stroke risk in patients with non-valvular AF, assigning points as follows: heart failure, hypertension, age 65–74, diabetes mellitus, vascular disease, and female-point 1 each, age $\geq$ 75 and stroke/transient ischemic attack (TIA) -point 2 each [9].

Laboratory test indicators include international normalized ratio (INR), creatinine clearance rate (CrCl), brain natriuretic peptide (BNP) and N-terminal pro-brain natriuretic peptide (NT-proBNP). Echocardiography data include left atrial diameter (LAD), right atrial diameter (RAD), left ventricular end-diastolic diameter (LVEDD), and left ventricular ejection fraction (LVEF). LAT refers to both left atrial thrombus and left atrial sludge. LA sludge is a characterized as a dynamic, viscous, layered echogenic area that is denser than spontaneous echography (SEC) but less dense than thrombus [10]. It is considered an intermediary stat between SEC and thrombus, and is often challenging to differentiate from true organized thrombosis via TEE. Consequently, this study does not strictly distinguish between thrombosis and sludge.

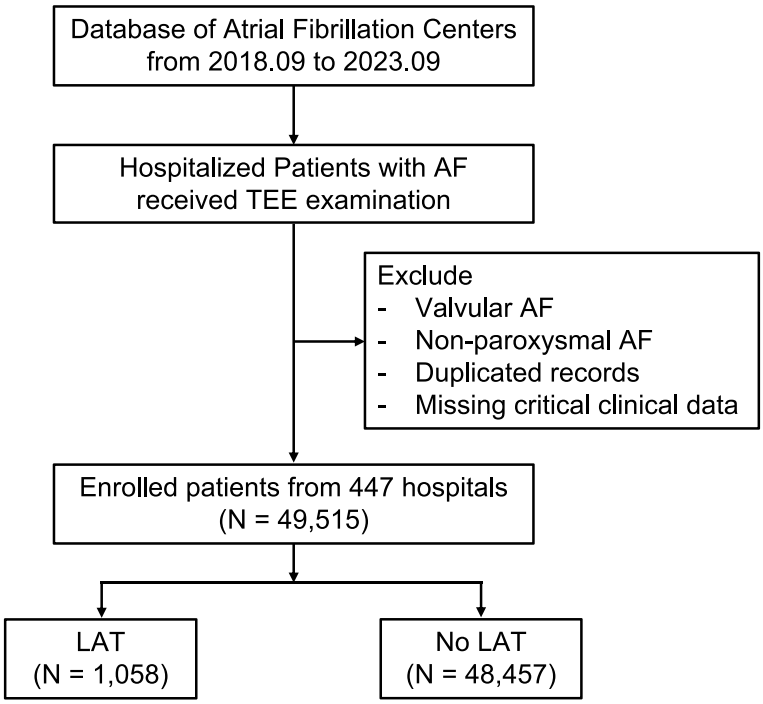


Fig. 1 Flowchart of participants selection. AF, atrial fibrillation; LAT, left atrial thrombus; TEE, transesophageal echocardiography

Machine learning models

Five commonly utilized ML algorithms were employed in this study to select critical features and develop the risk prediction model: decision tree, random forest, AdaBoost, k-Nearest Neighbor (kNN), and logistic regression. Patients were divided into training and internal validation cohorts in an 80:20 ratio. These ML models were trained using a tenfold cross-validation approach on the training set. Each ML model plotted the Receiver Operating Characteristic (ROC) curve to evaluate predictive performance, quantified by the Area Under the Curve (AUC), and identified the major features relevant for predicting LAT risk.

Statistical analysis

Continuous variables are presented as mean ± standard deviation (SD) or median (IQR), with T-tests applied for those conforming to normal distribution. Categorical variables are reported as counts and percentages (%), with comparisons conducted using the chi-square test or Fisher's exact test. Comparisons were made between demographic characteristics, medical history, laboratory tests, echocardiograms, and treatment data between participants in the LAT group and the non-LAT group. The kNN algorithm, with k = 5, effectively compensates for missing data through estimation. The difference in LAT detection rates between various subgroups was calculated and compared using the odds ratio (OR) values and 95% confidence intervals (CI). Statistical analysis was performed using SPSS 23.0, with significance defined as a $p < 0.05$. Five ML models were implemented using Python (Version 3.8.13).

Results

Characteristics of included patients

Among the 49,515 patients included in the final analysis, the mean age was 64.8 years, with 43.1% being female. A total of 1,058 patients (2.1%) were identified with LAT. Baseline characteristics of all included patients are presented in Table 1. Compared with patients without LAT, those with LAT exhibited a lower proportion of females and cardiology referrals. These patients demonstrated a higher prevalence of a CHA_2DS_2-VASc score ≥ 2, an elevated average heart rate, and a greater incidence of comorbidities including hypertension, CHD, chronic heart failure, stroke/TIA, arterial disease, cardiomyopathy, and thrombotic disorders. Additionally, patients with LAT had reduced LVEF, increased LAD and RAD. Furthermore, the usage rates of anticoagulants, diuretics, and angiotensin-converting enzyme inhibitors were higher among patients with LAT, whereas the usage rate of antiarrhythmic medications was lower (Table 1). The detection rate of LAT varies among different patient groups (Table 2). Patients from non-cardiology departments, those with CHA_2DS_2-VASc score greater than 2,

Table 1 Clinical characteristics of included patients

Characteristics	Total	LAT		
	(*n* = 49,515)	Yes (*n* = 1,058)	No (*n* = 48,457)	*p* value
Gender (Female)	21,346 (43.1)	416 (39.3)	20,930 (43.2)	0.012
Age (years)	64.8 ± 11.5	65.2 ± 11.2	64.8 ± 11.5	0.279
Cardiology department	45,768 (92.4)	932 (88.1)	44,836 (92.5)	< 0.001
CHA_2DS_2-VASc score				
0	8,352 (16.8)	155 (14.7)	8,197 (16.9)	0.005
1	12,091 (24.4)	230 (21.7)	11,861 (24.5)	
≥ 2	29,072 (58.7)	673 (63.6)	28,399 (58.6)	
Asymptomatic	1,355 (2.7)	32 (3.0)	1,323 (2.7)	0.562
Smoking	5,504 (11.1)	116 (11.0)	5,388 (11.1)	0.874
Alcohol	2,423 (4.8)	57 (5.4)	2,366 (4.9)	0.451
BMI (Kg/m^2)	24.3 ± 3.4	24.3 ± 3.7	24.3 ± 3.4	0.746
Systolic BP (mmHg)	130 ± 18	130 ± 19	130 ± 18	0.946
Heart rate (beat per min)	79 ± 17	81 ± 19	79 ± 17	0.002
Medical history				
Hypertension	19,316 (39.0)	453 (42.8)	18,863 (38.9)	0.01
CHD	6,653 (13.4)	166 (15.7)	6,487 (13.4)	0.03
CHF	6,053 (12.1)	173 (16.4)	5,862 (12.1)	< 0.001
Diabetes	5,440 (11.0)	129 (12.2)	5,311 (11.0)	0.205
Stroke/TIA	4,652 (9.4)	129 (12.2)	4,523 (9.3)	0.002
Arterial disease	2,929 (5.9)	79 (7.5)	2,850 (5.9)	0.031
Thyroid disease	1,270 (2.5)	25 (2.4)	1,245 (2.6)	0.674
COPD	999 (2.0)	25 (2.4)	974 (2.0)	0.419
Valvular heart disease	746 (1.5)	21 (2.0)	725 (1.5)	0.208
Cardiomyopathy	725(1.5)	24 (2.3)	701 (1.4)	0.028
Implantable cardiac devices	454 (0.9)	11 (1.0)	443 (0.9)	0.672
Thrombotic disorders	315 (0.6)	32 (3.0)	283 (0.6)	< 0.001
Cardiac echocardiography				
LVEF (%)	60.6 ± 16.1	60.0 ± 7.1	60.6 ± 6.1	0.004
LAD (mm)	38.2 ± 6.7	39.2 ± 5.9	38.2 ± 6.7	< 0.001
LVEDD (mm)	46.8 ± 4.5	47.0 ± 5.0	46.8 ± 4.5	0.065
RAD (mm)	36.8 ± 4.8	37.9 ± 6.4	36.8 ± 4.7	< 0.001
Laboratory results				
INR	1.10 (1.02–1.07)	1.17 (1.03–1.08)	1.10 (1.02–1.07)	0.001
BNP (pg/ml)	133 (99–179)	134 (104–179)	133 (99–179)	< 0.001
NT-proBNP (pg/ml)	405 (204–649)	462 (236–744)	405 (204–638)	0.004
CrCl (ml/min)	75 (62–92)	73 (62–91)	75 (62–92)	0.179
Therapeutic Drugs				
Antiarrhythmics	26,795 (54.1)	531 (50.2)	26,264 (54.2)	0.010
Statins	17,898 (36.1)	406 (38.4)	17,492 (36.1)	0.127
Anticoagulant	20,037 (40.4)	478 (45.2)	19,559 (40.4)	0.002
Digoxin	7,038 (14.2)	166 (15.7)	6,872 (14.2)	0.165
Antiplatelet	6,897 (13.9)	152 (14.4)	6,745 (13.9)	0.381
CCB	6,457 (13.0)	133 (12.6)	6,324 (13.1)	0.674
Diuretic	6,387 (12.8)	198 (18.7)	6,189 (12.8)	< 0.001
ACEIs/ARBs	3,251 (6.6)	90 (8.5)	3,161 (6.5)	0.010

Data presented as mean ± SD, median (IQR), or n (%)

ACEIs angiotensin-converting enzyme inhibitors, *AF* atrial fibrillation, *ARBs* angiotensin receptor antagonists, *BMI* body mass index, *BNP* brain natriuretic peptide, *BP* blood pressure, *CCB* calcium-channel blockers, *CHD* coronary heart disease, *CHF* chronic heart failure, *COPD* chronic obstructive pulmonary disease, *CrCl* creatinine clearance, *DCM* dilated cardiomyopathy, *HCM* hypertrophic cardiomyopathy, *INR* international standardized ratio, *LAD* left atrial diameter, *LVEF* left ventricular ejection fraction, *LVEDD* left ventricular end-diastolic diameter, *NT-proBNP* N-terminal pro-B-type natriuretic peptide, *RAD* right atrial diameter, *RVEDD* right ventricular end-diastolic diameter, *TIA* transient ischemic attack

Table 2 Detection rate and odds ratio of LAT in subgroups

Subgroups	LAT	OR	*p* value	*p* value for interaction
Gender				
Female	1.9 (1.8–2.1)	0.85 (0.75–0.97)	0.012	0.277
Male	2.3 (2.1–2.5)			
Department				
Cardiology	2.0 (1.9–2.2)	0.60 (0.50–0.73)	< 0.001	< 0.001
Non-cardiology	3.4 (2.8–3.9)			
Symptom				
Yes	2.1 (2.0–2.3)	0.90 (0.64–1.3)	0.562	0.002
No	2.4 (1.6–3.2)			
BMI (Kg/m^2)				
< 28	2.1 (2.0–2.2)	0.88 (0.74–1.06)	0.174	0.938
≥ 28	2.4 (2.0–2.8)			
CHA_2DS_2-VASc score				
0	1.9 (1.6–2.1)			0.206
1	1.9 (1.7–2.1)	1.03 (0.84–1.25)	0.810	
≥ 2	2.3 (2.1–2.5)	1.22 (1.05–1.48)	0.012	
Hypertension				
Yes	2.3 (2.1–2.6)	1.17 (1.04–1.32)	0.010	0.307
No	1.9 (1.8–2.2)			
Diabetes				
Yes	2.4 (2.0–2.8)	1.13 (0.94–1.35)	0.205	0.003
No	2.0 (1.9–2.2)			
CVD				
Yes	2.8 (2.3–3.2)	1.35 (1.12–1.61)	0.002	0.93
No	2.1 (1.9–2.2)			

Data presented as % (95% CI)

BMI body mass index, *CVD* cardiovascular disease, *LAT* left atrial thrombus, *OR* odds ratio

and those with hypertension, diabetes, and cardiovascular disease exhibited higher LAT detection rates. An interaction effect of age on LAT detection was observed in relation to department and diabetes status.

Establishment of LAT prediction model

The random forest model demonstrated the best predictive performance (AUC 0.833, 95%CI 0.730–0.924), followed by Xgboost, decision tree, logistic regression, and kNN models. To further validate the necessity of our research, we used logistic algorithm to calculate the predictive performance of CHA2DS2-VASc scores, with an AUC of 0.532 (95%CI 0.514–0.549). The predictive ability of LAT is significantly inferior to that of random forests (Fig. 2). Consequently, the random forest model was selected for further refinement. The feature weights in this model are illustrated in Fig. 3A. To enhance clinical applicability, the number of features was systematically reducing, revealing that the minimal number of features required to achieved 99% of the predictive performance was nine (Fig. 3B), with an AUC of 0.787 (95%CI 0.670–0.883) in finial random forest model (Fig. 3C).

Calibration and application of LAT prediction model

A significant discrepancy was observed between predicted and actual values in the final model ($p < 0.001$, Fig. 4A). After calibration, this discrepancy was no longer statistically significant ($p = 0.181$, Fig. 4B). The calibrated model was subsequently employed to stratify participants based on the quintile range of the estimated LAT detection rate, yielding LAT probabilities of 2.3%, 7.0%, 11.8%, 16.6%, and 21.5% (Fig. 5A). The distribution of the LAT detection rate as predicted by the final model is illustrated in Fig. 5B.

Discussion

In this large-scale, multi-center retrospective study, we report an incidence of LAT in paroxysmal AF of 2.1%. This finding aligns closely with a previous study which including data from 14,653 patients, demonstrated a LAT incidence of 2.7% [11]. Notably, this prior study found that the incidence of LAT was approximately four times higher in patients with non-paroxysmal AF compared to those with paroxysmal AF. The discrepancy in LAT incidence between our study and previous research may be attributed to several factors, including potentially inadequate screening practices for AF in China [12]. Such gaps in screening could result in missed diagnoses of AF, which may contribute to an elevated incidence of thromboembolic events. Furthermore, our findings indicate that both the LAT group and the non-LAT group had anticoagulant medication histories of less than 50%. A previous study demonstrated that after four weeks of anticoagulation therapy, 16 out of 18 patients with non-valvular AF had resolution of their thrombus, underscoring the significance of even short-term anticoagulation in preventing LAT formation [13]. This may further explain the higher incidence of thrombosis observed in our cohort.

Age has consistently been identified as a risk factor for AF and is positively correlated with the presence of LAT in affected patients [14, 15]. Studies have indicated that patients with non-paroxysmal AF are typically older, with higher CHA_2DS_2-VASc scores and increased prevalence of comorbidities such as hypertension and diabetes. While our study excluded age as a risk factor for LAT in paroxysmal AF, we confirmed associations with gender, body mass index (BMI), CHA_2DS_2-VASc score, hypertension, and CVD. Additionally, we noted significant interactions between asymptomatic AF, non-cardiology department presentations, and age,

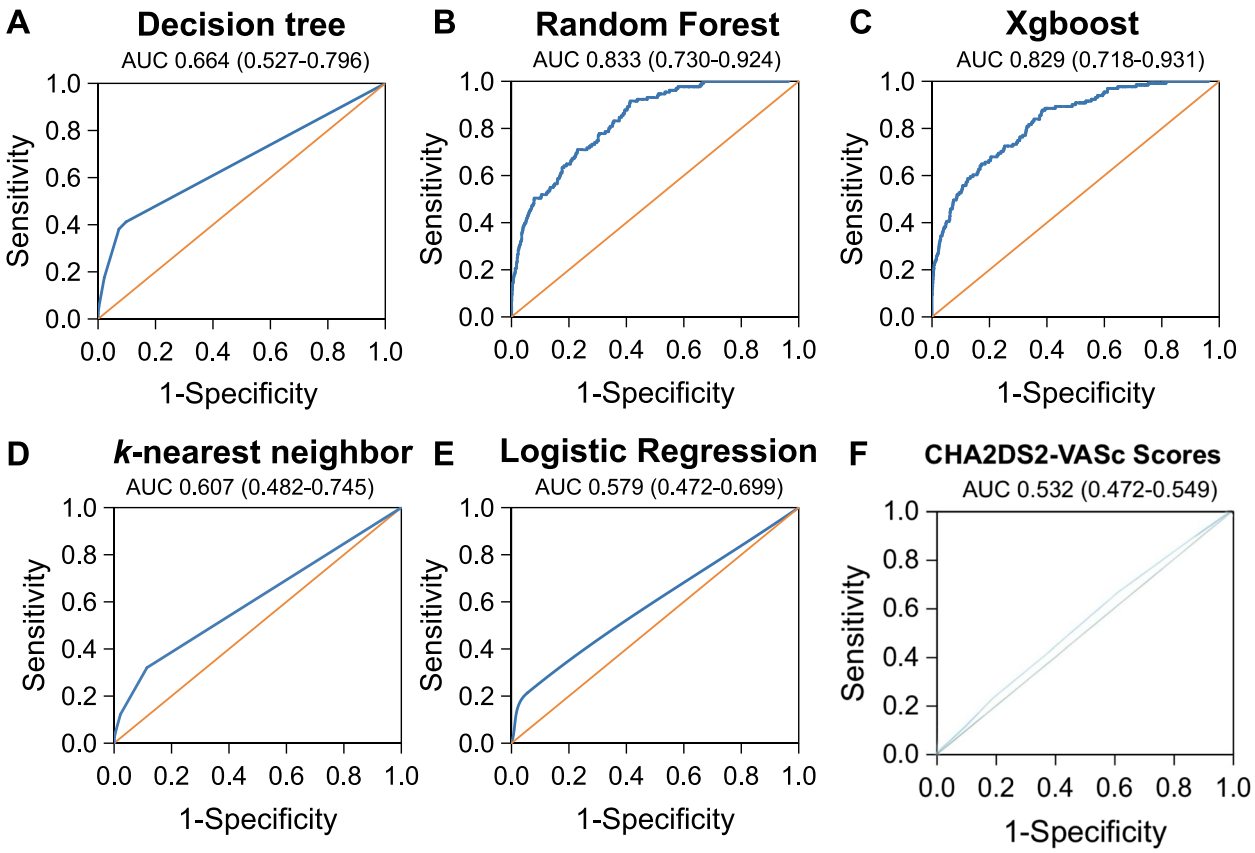


Fig. 2 Receiver operating characteristic curves for comparative analysis of machine learning algorithms and CHA2DS2-VASc scores in LAT. AUC, area under the curve; CI, confidence interval; LAT, left atrial thrombus

suggesting that asymptomatic patients may be overlooked during AF screening efforts.

In addressing the challenges of predicting LAT, we utilized various machine learning algorithms, including logistic regression and decision trees, to optimize classification and predictive modeling capabilities in the context of extensive and diverse datasets. Our study's strengths include a large sample size and substantial event rates for LAT, along with the application of machine learning-based statistical tools that facilitate variable selection and risk modeling. These tools adeptly manage large multidimensional datasets without the constraints imposed by traditional statistical assumptions, and we validated our model independently using an external cohort of AF patients.

This study introduces a predictive model for LAT in AF patients, leveraging machine learning techniques to achieve robust accuracy via a random forest algorithm. Compared to earlier studies employing machine learning for thrombus prediction [16, 17], particularly in AF populations, our model represents a significant advancement in both methodological rigor and clinical applicability. Previous research explored diverse machine learning methodologies, yet often encountered challenges related to model interpretability and the relevance of selected features. Our findings demonstrate that a refined random forest model, which integrates nine essential features, maintains a high predictive performance with an area under the curve (AUC) of 0.787, suggesting a model that is both interpretable and practical for clinical deployment.

Unlike prior studies that often-employed generalized datasets with broader feature sets [18, 19], our model was specifically trained on a large cohort focused on LAT risk factors. This targeted approach enabled the model to accurately capture critical variables such as the CHA_2DS_2-VASc score, LVEF, and atrial dimensions, all of which have been associated with thrombus formation in AF patients. While previous studies have identified these factors individually, they frequently lacked the comprehensive integration afforded by a robust machine learning model capable of managing complex variable interactions [19, 20]. Importantly, our model's reliance on established risk factors aligns with findings from earlier logistic regression studies, which have demonstrated effectiveness but may not adapt to the complexities of

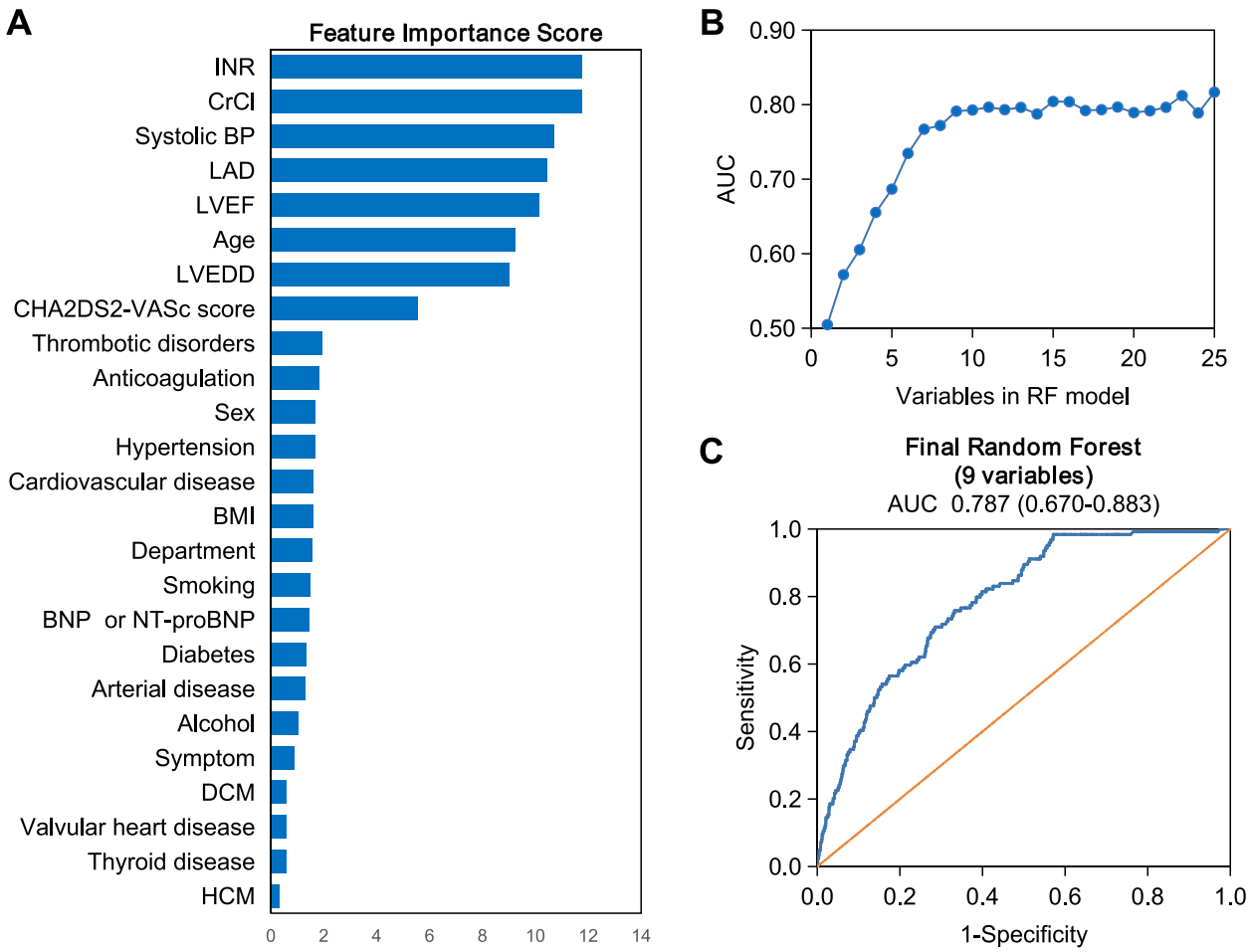


Fig. 3 Variable importance and performance analysis of LAT predictive modeling using random forest algorithm. AF, atrial fibrillation; BMI, body mass index; BNP, brain natriuretic peptide; BP, blood pressure; CrCl, creatinine clearance; DCM, dilated cardiomyopathy; HCM, hypertrophic cardiomyopathy; INR, international standardized ratio; LAD, left atrial diameter; LVEF, left ventricular ejection fraction; LVEDD, left ventricular end-diastolic diameter; NT-proBNP, N-terminal pro-B-type natriuretic peptide; TIA, transient ischemic attack. ROC, receiver operating characteristic scurves; AUC, area under the curve; CI, confidence interval

individual patients as effectively as our machine learning approach.

The clinical significance of this study is emphasized by its potential applications across various clinical settings. The stratified LAT risk probabilities facilitate patient categorization into quintiles, providing valuable insights for resource allocation and treatment prioritization. In comparison to existing risk scores, our machine learning model offers enhanced granularity in risk stratification, guiding clinicians in selecting appropriate preventive interventions, such as anticoagulant therapy or intensified monitoring for high-risk patients. Given the high risk of stroke associated with LAT, the implementation of this model in clinical practice could substantially improve patient outcomes through earlier detection and intervention [21].

Moreover, our calibrated model addresses the initial prediction-to-reality discrepancies and adjusts them to a non-significant level, thereby enhancing its applicability in real-world settings. This calibration is particularly pertinent for integration into electronic health record systems, where an interpretable model can automate the identification of high-risk patients based on readily available clinical data. Such integration has the potential to alleviate clinician burden by flagging individuals who would most benefit from further cardiovascular evaluation or anticoagulation therapy, ultimately promoting more targeted and effective care for AF populations.

The low anticoagulation rate (< 50% in both groups) likely contributed to the observed LAT incidence. Prior studies demonstrate that even short-term anticoagulation significantly reduces thrombus resolution [22–24]. Our findings underscore the need for improved adherence

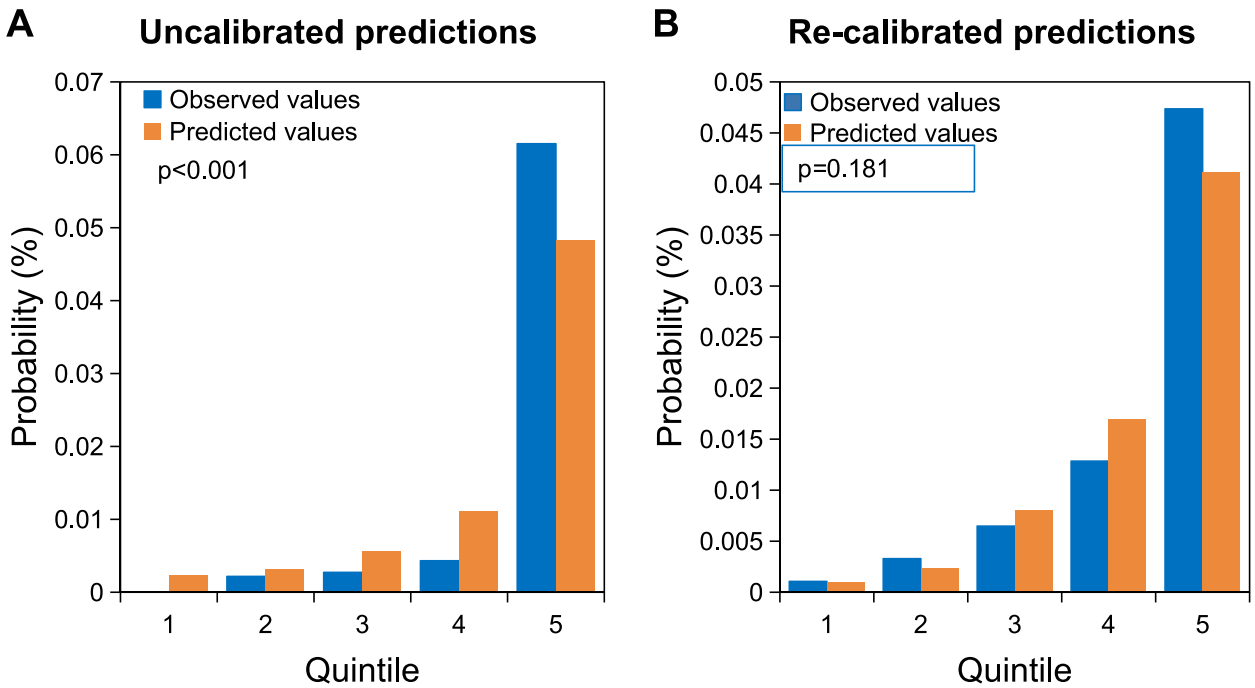
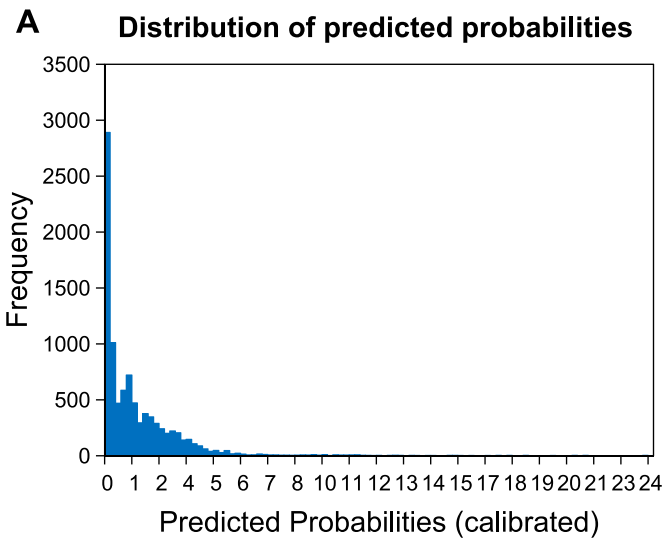


Fig. 4 Verification of model performance using testing cohort and post-calibration analysis. Model performance was evaluated using testing cohort data before (left) and after calibration (right). Blue bars represent predicted values and orange bars indicate actual observed values. Predictive impact was evaluated by stratifying detection rates into quintile. There was a statistically significant disparity between predicted and actual values before calibration ($p < 0.001$), and the disparity was no longer statistically significanct after calibration ($p = 0.284$)



B

Quantile	Median	Minimum	Maximum
1st	0.023	0.000	0.045
2nd	0.070	0.048	0.093
3rd	0.118	0.095	0.141
4th	0.166	0.143	0.188
5th	0.215	0.191	0.238

Fig. 5 Predicted distribution of LAT detection rates in the final model

to anticoagulation guidelines in paroxysmal AF management, particularly in high-risk subgroups.

All participating centers were located in China, which may limit generalizability to other populations due to genetic, socioeconomic, and healthcare system differences. However, key predictors (e.g., CHA2DS2-VASc score, LVEF) are globally recognized, suggesting broader applicability pending external validation.

While TEE cannot definitively distinguish thrombus from sludge, both entities confer thromboembolic risk. Future studies incorporating cardiac MRI may

improve differentiation, but our combined endpoint remains clinically relevant for guiding anticoagulation decisions. Although isoproterenol [25] may be able to distinguish thrombus and sluge, this is not a routine clinical approach.

A recent study [26] showed that the AUC of left atrial thrombus in patients with NVAF predicted by scores alone was 0.593 (95% CI: 0,495–0,690), much lower than our model(AUC 0.787). Although the study population was all patients with non-valvular atrial fibrillation, and we studied paroxysmal atrial fibrillation, it is still useful for reference.

Although a study [27] is similar with our approach, Their single-center cohort ($n = 1,222$) reported an AUC of 0.850, outperforming our model's AUC of 0.787. However, their study focused on a combined endpoint of LAT/SEC and included NVAF patients undergoing catheter ablation, whereas our multicenter cohort ($n = 49,515$) specifically targeted paroxysmal AF patients. While their model prioritized interpretability using SHAP values, our approach emphasized generalizability through multicenter data and refined feature selection. Notably, both studies identified left atrial diameter (LAD) as a critical predictor, underscoring its universal importance in thrombus risk stratification. Despite methodological differences, these findings collectively highlight the complementary roles of machine learning in addressing distinct clinical scenarios—Huang et al. for personalized risk visualization and our model for broader population-level screening.

In conclusion, this study advances the accuracy and clinical usability of LAT prediction by combining machine learning techniques with clinically relevant features. Future research should focus on validating this model across different healthcare settings to assess its generalizability, as well as exploring real-time applications within electronic health record systems to streamline management of AF patients. These advancements could support a proactive, precision medicine approach to LAT risk management, with the potential to mitigate the significant cardiovascular burden associated with thrombus-related complications in AF patients.

Limitations

Although the data and results are large sample, there were several limitations in our study. First, the retrospective design may introduce selection and information biases, as the data relied on existing medical records, which could lead to inaccuracies or omissions in clinical information. Second, while this is a multicenter study, all participating centers are located in China, potentially limiting the generalizability of the findings to broader populations or different geographic regions. Third, the low rate of anticoagulant usage (less than 50%) in both the LAT and non-LAT groups may affect the incidence rates of LAT, complicating the assessment of anticoagulant efficacy. Additionally, the lack of long-term follow-up data restricts the ability to observe changes in LAT incidence over time or the sustained impact of anticoagulation therapy. Furthermore, the heterogeneity of the patient sample regarding clinical characteristics and comorbidities may affect the applicability of the findings across diverse patient populations [28]. Lastly, although machine learning models offer robust predictive capabilities, their interpretability may be limited, making it challenging for clinicians to fully understand the clinical implications of the model outputs. Additionally, incomplete medical records could lead to information bias, particularly regarding anticoagulation adherence. Moreover, ESC guidelines do not explicitly recommend routine TEE for all asymptomatic atrial fibrillation patients, but all of our patients have undergone LAT examination through TEE, which may cause bias. Future prospective studies are warranted to validate these findings. These limitations underscore the need for further research to validate these findings and enhance their clinical relevance.And we haven't paid attention to the prevention of atrial fibrillation, which can still lead to pulmonary arterial hypertension even after radiofrequency ablation [29].

Conclusion

This large-scale multi-center retrospective study reveals a 2.1% incidence of LAT in patients with paroxysmal AF, highlighting the need for enhanced screening and anticoagulation practices. Our findings identify key clinical factors, including gender, BMI, CHA_2DS_2-VASc score, hypertension, and cardiovascular disease, as significant predictors of LAT risk. Furthermore, the development of a machine learning-based predictive model demonstrates strong accuracy in stratifying LAT risk, offering practical implications for clinical decision-making. This study contributes to the advancement of precision medicine in AF management, and future research should focus on validating this model across diverse populations and integrating it into electronic health records to improve patient outcomes.

Abbreviations

LAT	Left Atrial Thrombus
AF	Atrial Fibrillation
TEE	Transesophageal Echocardiography
ML	Machine Learning
CHD	Coronary Heart Disease
TIA	Stroke/Transient Ischemic Attack

CHF	Chronic Heart Failure
INR	International Normalized Ratio
CrCl	Creatinine Clearance Rate
BNP	Brain Natriuretic Peptide
NT-proBNP	N-terminal pro-brain Natriuretic Peptide
LAD	Left Atrial Diameter
RAD	Right Atrial Diameter
LVEDD	Left Ventricular End-diastolic Diameter
LVEF	Left Ventricular Ejection Fraction
SEC	Spontaneous Echography
kNN	K-Nearest Neighbor
ROC	Receiver Operating Characteristic
AUC	Area Under the Curve
OR	Odds Ratio
CI	Confidence Intervals
BMI	Body Mass Index

Acknowledgements
The authors would like to thank Chinese atrial fibrillation center and those researchers who contribute to the BMC Cardiovascular Disorders.

Authors' contributions
Wanli, Xiong and Qiqi Cao (Co-First authors):Conceptualization, Data Curation, Formal Analysis, Formal Analysis, Methodology, Validation, Visualization and Writing—Original Draft. Lu Jia: Formal Analysis, Investigation, Writing—Original Draft. Min Chen: Formal Analysis, Investigation, Software, Visualization. Liu Tao, Qingyan Zhao, Yanhong Tang and Bo Yang: Project Administration, Resources, Supervision, Writing—Review & Editing. Li Li: Formal Analysis, Methodology, Software, Supervision, Validation, Visualization. Shaobo Shi(Corresponding Author): Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Project Administration, Validation, Visualization, Writing—Review & Editing. He Huang and Congxin Huang: Conceptualization, Funding Acquisition, Project Administration, Resources, Supervision, Writing—Review & Editing. All authors reviewed the manuscript.

Funding
The study was supported by The Interdisciplinary Innovative Talents Foundation from Renmin Hospital of Wuhan University (JCRCZN-2022–003) and the Fundamental Research Funds for the Central Universities (2042024YXB003).

Data availability
The data that support the study will be available in Chinese atrial fibrillation database.

Declarations

Ethics approval and consent to participate
This study was part of the real-world study of Chinese atrial fibrillation RWS-CAF; registration number: ChiCTR1900021250,registered on February 3, 2019**,** and was conducted in accordance with the Declaration of Helsinki. It received approval from the Ethical Review Committee of Renmin Hospital of Wuhan University. As a retrospective study, it does not involve sensitive patient information, has been approved by the ethics committee, and is exempt from patient informed consent.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Cardiology, Renmin Hospital of Wuhan University, Hubei Province, 238 Jiefang Road, Wuchang District, Wuhan 430060, China. [2]Cardiovascular Research Institute, Wuhan University, Wuhan 430060, China. [3]Key Laboratory of Cardiovascular Disease, Wuhan 430060, China. [4]Wuhan Shinall Technology Co, Ltd, Wuhan 430070, China. [5]School of Electronic Information, Wuhan University, Wuhan 430060, China.

Received: 10 January 2025 Accepted: 12 May 2025
Published online: 02 June 2025

References

1. Ladwig KH, Goette A, Atasoy S, Johar H. Psychological aspects of atrial fibrillation: A systematic narrative review: Impact on incidence, cognition, prognosis, and symptom perception. Curr Cardiol Rep. 2020;22(11):137.
2. Brundel BJJM, Ai X, Hills MT, Kuipers MF, Lip GYH, de Groot NMS. Atrial fibrillation. Nat Rev Dis Primers. 2022;8(1):21.
3. Glotzer TV, Ziegler PD. Cryptogenic stroke: Is silent atrial fibrillation the culprit? Heart Rhythm. 2015;12(1):234–41.
4. Akhmerov A, Ramzy D. Left Atrial Thrombus. N Engl J Med. 2022;387(5):e8.
5. Romero J, Cao JJ, Garcia MJ, Taub CC. Cardiac imaging for assessment of left atrial appendage stasis and thrombosis. Nat Rev Cardiol. 2014;11(8):470–80.
6. Manning WJ, Weintraub RM, Waksmonski CA, Haering JM, Rooney PS, Maslow AD, Johnson RG, Douglas PS. Accuracy of transesophageal echocardiography for identifying left atrial thrombi. A prospective, intraoperative study. Ann Intern Med. 1995;123(11):817–22.
7. Kim YG, Shim J, Boo KY, Kim DY, Oh SK, Lee KN, Choi JI, Kim YH. Different influence of cardiac hemodynamics on thromboembolic events in patients with paroxysmal and non-paroxysmal atrial fibrillation. PLoS ONE. 2019;14(3):e0214743.
8. Zhao QY, Shi SB, Huang H, Jiang H, Yang B, Wu G, Bao MW, Liu Y, Tang YH, Wang X, Zhang S, Huang DJ, Huo Y, Ge JB, Huang CX. Contemporary characteristics, management, and outcomes of patients hospitalized for atrial fibrillation in China: results from the real-world study of Chinese atrial fibrillation registry. Chin Med J (Engl). 2020;133(23):2883–4.
9. Hindricks G, Potpara T, Dagres N, Arbelo E, Bax JJ, Blomström-Lundqvist C, Boriani G, Castella M, Dan GA, Dilaveris PE, Fauchier L, Filippatos G, Kalman JM, La Meir M, Lane DA, Lebeau JP, Lettino M, Lip GYH, Pinto FJ, Thomas GN, Valgimigli M, Van Gelder IC, Van Putte BP, Watkins CL, ESC Scientific Document Group. 2020 ESC Guidelines for the diagnosis and management of atrial fibrillation developed in collaboration with the European Association for Cardio-Thoracic Surgery (EACTS): The Task Force for the diagnosis and management of atrial fibrillation of the European Society of Cardiology (ESC) Developed with the special contribution of the European Heart Rhythm Association (EHRA) of the ESC. Eur Heart J. 2021;42(5):373–498.
10. Lowe BS, Kusunose K, Motoki H, Varr B, Shrestha K, Whitman C, Tang WH, Thomas JD, Klein AL. Prognostic significance of left atrial appendage "sludge" in patients with atrial fibrillation: a new transesophageal echocardiographic thromboembolic risk factor. J Am Soc Echocardiogr. 2014;27(11):1176–83.
11. Lurie A, Wang J, Hinnegan KJ, McIntyre WF, Belley-Côté EP, Amit G, Healey JS, Connolly SJ, Wong JA. Prevalence of Left Atrial Thrombus in Anticoagulated Patients With Atrial Fibrillation. J Am Coll Cardiol. 2021;77(23):2875–86.
12. Du X, Guo L, Xia S, Du J, Anderson C, Arima H, Huffman M, Yuan Y, Zheng Y, Wu S, Guang X, Zhou X, Lin H, Cheng X, Dong J, Ma C. Atrial fibrillation prevalence, awareness and management in a nationwide survey of adults in China. Heart. 2021;107(7):535–41.
13. Collins LJ, Silverman DI, Douglas PS, Manning WJ. Cardioversion of nonrheumatic atrial fibrillation. Reduced thromboembolic complications with 4 weeks of precardioversion anticoagulation are related to atrial thrombus resolution. Circulation. 1995;92(2):160–3.
14. Frost L, Johnsen SP. Age and Risk of Stroke in Asians With Atrial Fibrillation. Stroke. 2018;49(8):1809–10.
15. Wang N, Yu Y, Sun Y, Zhang H, Wang Y, Chen C, Tan X, Wang B, Lu Y. Acquired risk factors and incident atrial fibrillation according to age and genetic predisposition. Eur Heart J. 2023;44(47):4982–93.
16. Bernardini A, Bindini L, Antonucci E, Berteotti M, Giusti B, Testa S, Palareti G, Poli D, Frasconi P, Marcucci R. Machine learning approach for prediction of outcomes in anticoagulated patients with atrial fibrillation. Int J Cardiol. 2024;15(407):132088.

17. Antúnez-Muiños P, Vicente-Palacios V, Pérez-Sánchez P, Sampedro-Gómez J, Sánchez-Puente A, Dorado-Díaz PI, Nombela-Franco L, Salinas P, Gutiérrez-García H, Amat-Santos I, Peral V, Morcuende A, Asmarats L, Freixa X, Regueiro A, Caneiro-Queija B, Estevez-Loureiro R, Rodés-Cabau J, Sánchez PL, Cruz-González I. Predictive Power for Thrombus Detection after Atrial Appendage Closure: Machine Learning vs. Classical Methods J Pers Med. 2022;12(9):1413.
18. Lip GYH, Tran G, Genaidy A, Marroquin P, Estes C, Landsheft J. Improving dynamic stroke risk prediction in non-anticoagulated patients with and without atrial fibrillation: comparing common clinical risk scores and machine learning algorithms. Eur Heart J Qual Care Clin Outcomes. 2022;8(5):548–56.
19. Zhao Y, Cao LY, Zhao YX, Wang F, Xie LL, Xing HY, Wang Q. Medical record data-enabled machine learning can enhance prediction of left atrial appendage thrombosis in nonvalvular atrial fibrillation. Thromb Res. 2023;223:174–83.
20. Ruzieh M, Bai C, Meisel E, Kramer EF, Frechette RR, Nassereddin AT, Smoot M, Edwards ES, Kurup V, Naccarelli GV, Naik D, Kimmel SE, Mardini MT. Predictors of left atrial appendage thrombus in atrial fibrillation patients undergoing cardioversion. J Interv Card Electrophysiol. 2024;67(9):2059–66. https://doi.org/10.1007/s10840-024-01868-x. Epub 2024 Jul 12. PMID: 38995603.
21. Melduni R, Nkomo VT, Wysokinski W, Gersh BJ, Deshmukh A, Padang R, Greene EL, Oh JK, Lee HC. Risk of left atrial appendage thrombus and stroke in patients with atrial fibrillation and mitral regurgitation. Heart. 2022;108(1):29–36.
22. Zhou Q, Liu X, Gu ZC, Yang X, Huang XH, Wu YZ, Tao YY, Wei M. Short-term antiplatelet versus anticoagulant therapy after left atrial appendage closure: a systematic review and meta-analysis. J Thromb Thrombolysis. 2024;57(2):194–203. https://doi.org/10.1007/s11239-023-02919-2. (Epub 2024 Jan 5).
23. Yadlapati A, Groh C, Passman R. Safety of short-term use of dabigatran or rivaroxaban for direct-current cardioversion in patients with atrial fibrillation and atrial flutter. Am J Cardiol. 2014;113(8):1362–3. https://doi.org/10.1016/j.amjcard.2013.12.044. (Epub 2014 Jan 31).
24. Lorente-Ros Á, Alonso-Salinas GL, Monteagudo Ruiz JM, Abellás-Sequeiros M, Vieítez-Florez JM, Sánchez Vega D, Álvarez-Garcia J, Sanmartín-Fernández M, Lorente-Ros M, Del Prado DS, Fernández Golfín C, Zamorano Gómez JL. Effect of Duration of Anticoagulation in the Incidence of Stroke in Patients With Left-Ventricular Thrombus. Am J Cardiol. 2022;15(185):115–21. https://doi.org/10.1016/j.amjcard.2022.09.005. (Epub 2022 Oct 12).
25. Goyal SK, Hyder S, Liu S, Vannan MA. Isoproterenol-Assisted Differentiation Between Sludge and Organized Thrombus to Guide Left Atrial Appendage Occlusion. JACC Clin Electrophysiol. 2023;9(1):111–6. https://doi.org/10.1016/j.jacep.2022.10.026. Epub 2022 Nov 30.
26. Zhang Y, Yuan YQ. Value of left atrial diameter with CHA2DS2-VASc score in predicting left atrial/left atrial appendage thrombosis in non-valvular atrial fibrillation. Arq Bras Cardiol. 2021;116(2):325–31.
27. Huang C, Shu S, Zhou M, Sun Z, Li S. Development and validation of an interpretable machine learning model for predicting left atrial thrombus or spontaneous echo contrast in non-valvular atrial fibrillation patients. PLoS ONE. 2025;20(1):e0313562. https://doi.org/10.1371/journal.pone.0313562.
28. Mohanty S, Torlapati PG, La Fazia VM, Kurt M, Gianni C, MacDonald B, Mayedo A, Allison J, Bassiouny M, Gallinghouse GJ, Burkhardt JD, Horton R, Di Biase L, Al-Ahmad A, Natale A. Best anticoagulation strategy with and without appendage occlusion for stroke-prophylaxis in postablation atrial fibrillation patients with cardiac amyloidosis. J Cardiovasc Electrophysiol. 2024;35(7):1422–8. https://doi.org/10.1111/jce.16308. (Epub 2024 May 15).
29. Mohanty S, Della Rocca DG, Torlapati PG, Chierchia GB, Dello Russo A, Casella M, Gianni C, MacDonald B, Mayedo A, La Fazia VM, Bassiouny M, Gallinghouse GJ, Burkhardt JD, Horton R, Al-Ahmad A, Di Biase L, Pannone L, de Asmundis C, Natale A. Pulsed-Field Ablation Does Not Worsen Baseline Pulmonary Hypertension Following Prior Radiofrequency Ablations. JACC Clin Electrophysiol. 2024;10(3):477–86. https://doi.org/10.1016/j.jacep.2023.11.005. (Epub 2023 Nov 22).

Publisher's Note