Performance of an AI prediction tool for new-onset atrial fibrillation after coronary artery bypass grafting

Hualong Ma,[a,h] Dalong Chen,[b,h] Weitao Lv,[c,h] Qiuying Liao,[d] Jingyi Li,[e] Qinai Zhu,[f] Ying Zhang,[d] Lizhen Deng,[d] Xiaoge Liu,[a] Qinyang Wu,[a] Xianliang Liu,[g,**] and Qiaohong Yang[a,*]

[a]Jinan University School of Nursing, Guangzhou, Guangdong, China
[b]Yunfu People's Hospital, Yunfu, Guangdong, China
[c]Intensive Care Unit, The First Affiliated Hospital of Jinan University, Guangzhou, Guangdong, China
[d]Cardiovascular Surgery Department, The First Affiliated Hospital of Jinan University, Guangzhou, Guangdong, China
[e]Qingyuan People's Hospital, Qingyuan, Guangdong, China
[f]Laboratory Department, The First Affiliated Hospital of Jinan University, Guangzhou, Guangdong, China
[g]School of Nursing and Health Studies, Hong Kong Metropolitan University, Hong Kong SAR, China

eClinicalMedicine
2025;81: 103131
Published Online xxx
https://doi.org/10.1016/j.eclinm.2025.103131

Summary

Background There is lack of tools to predict new-onset postoperative atrial fibrillation (NOAF) after coronary artery bypass grafting (CABG). We aimed to develop and validate a novel AI-based bedside tool that accurately predicts predict NOAF after CABG.

Methods Data from 2994 patients who underwent CABG between March 2015 and July 2024 at two tertiary hospitals in China were retrospectively analyzed. 2486 patients from one hospital formed the derivation cohort, split 7:3 into training and test sets, while the 508 patients from a separate hospital formed the external validation cohort. A stacking model integrating 11 base learners was developed and evaluated using Accuracy, Precision, Recall, F1 score, and Area Under Curve (AUC). SHapley Additive exPlanations (SHAP) values were calculated and plotted to interpret the contributions of individual characteristics to the model's predictions.

Findings Seventy-seven predictive characteristics were analyzed. The stacking model achieved superior performance with AUCs 0·931 and F1 scores 0·797 in the independent external validation, outperforming CHA2DS2-VASc, HATCH, and POAF scores (AUC 0·931 vs. 0·713, 0·708, and 0·667; $p < 0·05$). SHAP value indicate that the importance of predictive features for NOAF, in descending order, include: Brain natriuretic peptide, Left ventricular end-diastolic diameter, Ejection fraction, BMI, β-receptor blockers, Duration of surgery, Age, Neutrophil percentage-to-albumin ratio, Myocardial infarction, Left atrial diameter, Hypertension, and smoking status. Subsequently, we constructed an easy-to-use bedside clinical tool for NOAF risk assessment leveraging these characteristics.

Interpretation The AI-based tool offers superior prediction of NOAF, outperforming three existing predictive tools. Future studies should further explore how various patient characteristics influence the timing of NOAF onset, whether early or late.

Funding This work was funded by Lingnan Nightingale Nursing Research Institute of Guangdong Province, and Guangdong Nursing Society (GDHLYJYZ202401).

Keywords: Artificial intelligence; Machine learning; Web tool; Coronary artery bypass grafting; Postoperative atrial fibrillation; Prediction model

Introduction

Coronary artery bypass grafting (CABG) is a widely utilized surgery for treating coronary artery disease by bypassing severely stenotic or obstructed coronary arteries that cannot be effectively treated with angioplasty or stent placement. New-onset postoperative atrial

*Corresponding author. School of Nursing, Jinan University, Room 528, Jinan University School of Nursing, Guangzhou, Guangdong, China.
**Corresponding author. Hong Kong Metropolitan University, School of Nursing and Health Studies, Hong Kong Metropolitan University, Hong Kong SAR, China.
E-mail addresses: yqiaohong@163.com (Q. Yang), dxliu@hkmu.edu.hk (X. Liu).
[h]Contributed equally to this work.

Research in context

Evidence before this study

We conducted a comprehensive review of the literature to identify studies on predicting new-onset atrial fibrillation (NOAF) after coronary artery bypass grafting (CABG). Our search spanned multiple databases, including PubMed, Embase, and Cochrane Library, China National Knowledge Infrastructure (China), Wanfang Database (China), Weipu Database (China) covering studies published from dataset development, to July 31, 2024. We used search terms such as "atrial fibrillation," "postoperative atrial fibrillation," "coronary artery bypass grafting," "risk prediction," and "artificial intelligence" et al. Studies were included if they addressed risk factors for NOAF, validated risk prediction models, or involved the application of machine learning, ensemble learning, deep learning in postoperative risk assessment. Non-English or Non-Chinese language studies were considered if translated versions were available.

Our review revealed that commonly used prediction tools, such as the CHA2DS2-VASc, HATCH, and POAF scores, have limited predictive power (area under the curve [AUC]: 0.64–0.76). While these tools are simple to use, they often fail to capture the complexity of NOAF's multifactorial etiology. We also noted that most studies focused on linear regression models, small single-center datasets, or postoperative predictors, which limits their practical applicability for preoperative decision-making. AI applications have shown promise in enhancing other diseases predictive accuracy but lacked validation in large, diverse populations, especially for NOAF after CABG.

Added value of this study

This study represents the most comprehensive effort to develop and validate an AI-based tool for predicting NOAF in CABG patients. By leveraging a large multicenter dataset of 2994 patients and using ensemble learning, we achieved significantly higher predictive accuracy compared to three existing tools. Unlike traditional models, our web-based bedside tool is preoperatively applicable, relying on easily accessible clinical parameters and omitting postoperative metrics, thus enabling proactive interventions. Furthermore, we evaluated the bedside tool's decision-support capabilities in real-world settings, demonstrating its ability to enhance the diagnostic consistency of healthcare providers, irrespective of their experience levels.

Implications of all the available evidence

The findings suggest that AI-driven tools can address key gaps in predicting and managing NOAF, especially in resource-limited or training-intensive settings. Incorporating this bedside tool into routine clinical practice can help reduce unnecessary interventions, optimize preoperative planning, and standardize decision-making across diverse clinical environments. Future research could explore how the predictive characteristics identified in existing studies affect the timing of NOAF occurrence.

fibrillation (NOAF) is the most common complication following CABG, with an incidence ranging from 20% to 40%.[1] NOAF typically develops within 2–4 days post-surgery[2] and is associated with prolonged hospital stays,[3] increased readmission rates,[4] cerebrovascular events (10-year hazard ratio: 1.53),[5] and higher long-term mortality.[1] Despite established guidelines for its management, the incidence of NOAF has not significantly decreased in recent years.[6] Most patients undergoing CABG maintain sinus rhythm postoperatively, meaning the routine use of prophylactic medications could expose many to unnecessary side effects and costs.[7] Thus, tailoring treatment strategies based on patient-specific risk factors is crucial to minimizing unnecessary interventions while effectively preventing NOAF and its associated complications.

Although age is a well-known risk factor for NOAF, other definitive factors remain unclear, indicating a need for further research on a larger scale.[8] Current prediction tools, such as the CHA2DS2-VASc,[9] HATCH,[10] POAF score,[11] demonstrate suboptimal predictive performance (area under the curve [AUC] 0·64–0·76), and primarily focusing on simple and linear factors. The POAF score, for example, incorporates only age, preoperative conditions, and left ventricular function, etc. Recent advancements in AI have shown potential for enhancing prediction models in healthcare. These technologies are capable of identifying complex, nonlinear relationships within large datasets, which could lead to more accurate and potentially more generalized predictions.[12] However, while there is growing interest in their application, the clinical benefits of AI in this context remain debated. The real-world effectiveness of AI-driven models has yet to be fully established, and much of the current research is limited by small, single-center datasets or simplified linear models. Therefore, further large-scale, multi-center studies are necessary to assess the true potential of AI for improving NOAF risk prediction. This study aims to develop an AI-driven NOAF prediction tool for CABG patients (Fig. 1 Part A), and create a user-friendly bedside web tool (Fig. 1 Part B).

Methods

Study design and setting

This prognostic, multicenter study was conducted in two phases, as outlined in Fig. 1. The first phase involved developing a predictive model for NOAF in patients undergoing CABG, which including training

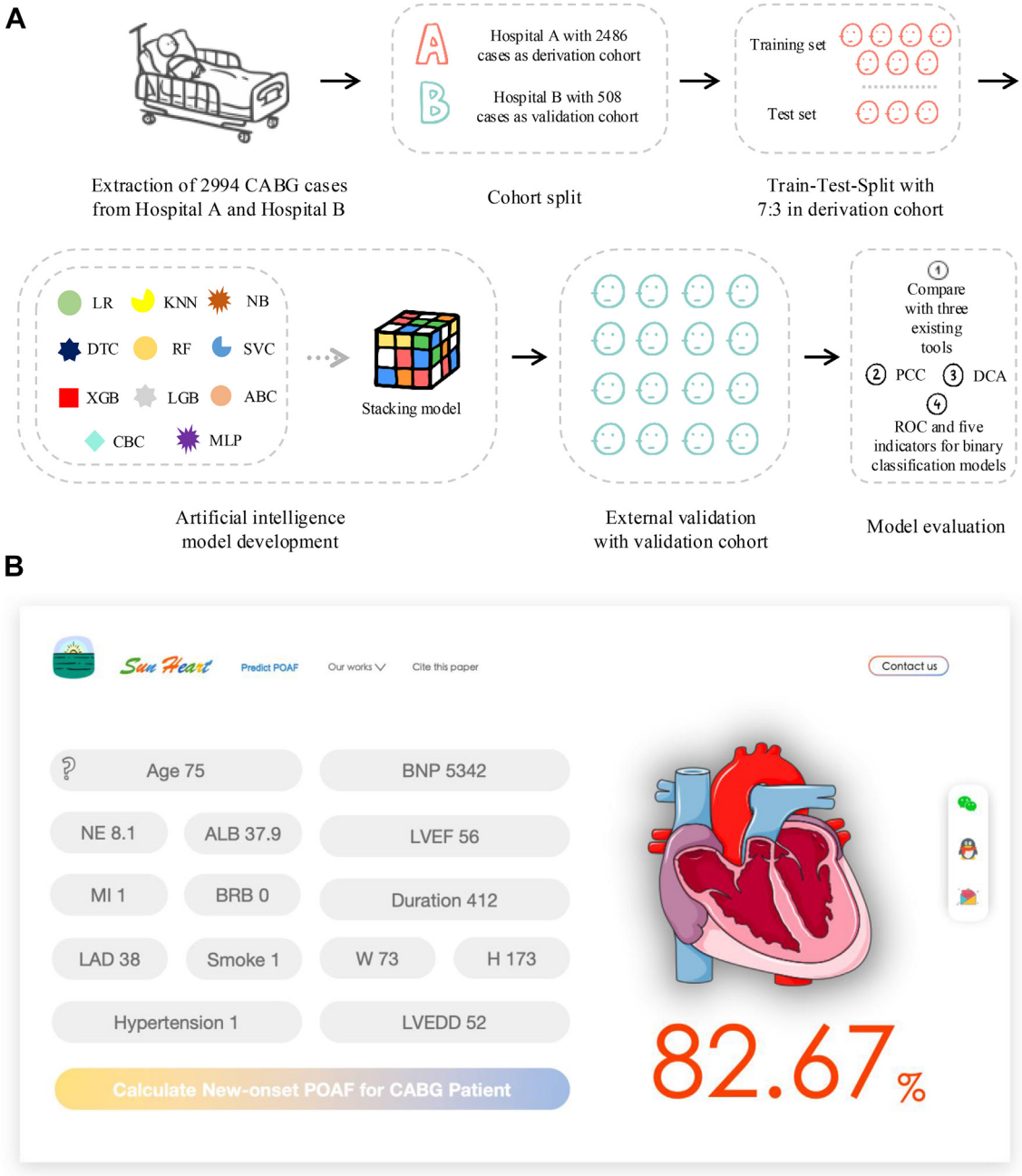

Fig. 1: Note: ABC, Adaptive Boosting Classifier; ALB, albumin; AUC, area under curve; BNP, brain natriuretic peptide; BRB, β-receptor blockers; CABG, coronary artery bypass grafting; CBC, CatBoost Classifier; DCA, decision curve analysis; DTC, Decision Tree Classifier; H, height; KNN, K-Nearest Neighbors; LAD, left atrial diameter; LGB, Light Gradient Boosting Machine; LR, Logistic regression; LVEF, left ventricular ejection fraction; LVEDD, left ventricular end diastolic diameter; MI, myocardial infarction; MLP, Multilayer Perceptron; NB, Naive Bayes; NE, neutrophil; PCC, probability calibration curve; RF, Random Forest; ROC, receiver operating characteristic curve; SVC, Support Vector Classifier; W, weight; XGB, eXtreme Gradient Boosting. Part A: This is a prognostic study, the first phase of which involved developing a predictive model for NOAF in patients undergoing CABG using 11 base learners stacked via a stacking technique, with internal and external validation. The corresponding results are presented in Fig. 3. Part B: The second phase focused on creating a user-friendly bedside tool. The tool features a user-friendly interface. A prompt, explaining the abbreviations and units for all data inputs (e.g., BNP: brain natriuretic peptide (pg/ml), etc.), will appear automatically every time the user opens the tool and cannot be closed for 5 s. Medical professionals can simply input the relevant parameters and click "Calculate New-onset POAF for CABG Patient" to determine the probability of NOAF occurrence. If the probability of NOAF exceeds 95% or falls below 5%, our tool provides a prompt saying, "The prediction probability is too [HIGH/LOW], please check the numerical inputs carefully, especially the units.".

the model on a large dataset, and validating the tool with an independent external dataset, to ensure the generalization. The second phase focused on creating a user-friendly bedside tool to ensure the clinical implications. The study adheres to the TRIPOD checklist to report results. The glossary is presented in the Supplementary File part 1.

Participants

Data were retrospectively collected from 2994 patients who underwent CABG between March 2015 and July 2024 at two tertiary hospitals in China (Hospital A: 2486 cases; Hospital B: 508 cases). This geographical and institutional diversity between the two hospitals ensures that the cohorts reflect different clinical settings and demographics, making the tool highly generalizable. NOAF was diagnosed based on the following criteria: absence of prior atrial fibrillation, and postoperative electrocardiographic evidence of p-wave disappearance replaced by f-waves, absolute irregularity in the QRS wave rhythm, and variable RR intervals, with the episode lasting for more than 30 s.[13] Furthermore, NOAF occurring after discharge was not defined as NOAF within this study.[14] Patients in both source hospitals underwent at least 7 days of continuous ECG monitoring postoperatively. Upon detection of NOAF, confirmation by both the attending cardiologist and nurse was required.

Inclusion criteria were as follows: (1) No prior cardiac surgery and first-time CABG; (2) Preoperative electrocardiogram showing sinus rhythm; (3) Age ≥18 years. Exclusion criteria included: (1) Coagulation disorders, thyroid dysfunction, heart valve disease, or other congenital heart malformations; (2) History of pacemaker implantation or previous atrial fibrillation; (3) Prophylactic use of amiodarone during the perioperative period.

Predictors and outcomes

A total of 77 predictive characteristics were collected (see Supplementary File part 2), categorized into eight groups: demographic data, cardiac ultrasound results, surgery-related information, preoperative nursing records, preoperative laboratory tests, medical history, preoperative drug usage, and postoperative nursing records.

Smoking status was categorized as current smoker if the patient had quit for less than one year; otherwise, they were considered a non-smoker.[15] Current alcohol consumption was defined as drinking one or more alcoholic beverages per week; individuals who consumed less were categorized as non-drinkers.[16] Nutritional risk was assessed using the NRS2002, where a score of ≥3 indicated a risk of malnutrition. Independence in daily living was evaluated using the Barthel Activities of Daily Living (ADL) Index, which ranges from 0 to 100, with higher scores reflecting greater independence. For this study, POAF, CHA2DS2-VASc, and HATCH scores were independently evaluated by four senior professionals. The formal medical records of the patients served as the standard for all scoring items. The POAF score was calculated as follows: 1 point for age 60–69 years, 2 points for 70–79 years, and 3 points for 80 years and older; 1 point for comorbid chronic obstructive pulmonary disease, estimated glomerular filtration rate (eGFR) <15 ml/min/1·73 m^2, emergency surgery, intra-aortic balloon counterpulsation, left ventricular ejection fraction (LVEF) <30%, or valve surgery.[11] The CHA2DS2-VASc score was calculated as: 2 points for age ≥75 years, 1 point for age 65–74 years, female sex, heart failure, hypertension, diabetes, or peripheral artery disease, and 2 points for a history of stroke or transient ischemic attack.[9] The HATCH score was determined by assigning 1 point for hypertension, age ≥75 years or chronic obstructive pulmonary disease, 2 points for stroke or transient ischemic attack or heart failure.[10] Given that the Neutrophil Percentage-to-Albumin Ratio (NPAR) has emerged as a novel inflammation marker in recent years, and that elevated NPAR values are associated with increased 30-day all-cause mortality in patients with severe coronary artery disease, we constructed and explored NPAR as a potential predictor of NOAF.[17]

Statistical methods

Following the 10-events-per-variable rule, a minimum of 770 cases (77 variables × 10 events per variable) was required.[18] Data analysis was conducted using Python (v3·12·5). The pattern of missing data in the original dataset was analyzed using Little's MCAR test, which revealed a χ^2 value of 5888·25 ($p < 0·001$), indicating that the dataset was not missing completely at random (MCAR). Consequently, simply removing data with a missing rate >10% may introduce bias. To address this concern, we constructed models separately with and without patients with more than 10% missing data. A sensitivity analysis was performed using the DeLong test to compare the model performance (Supplementary File part 4.1). The results demonstrated the robustness of the model. Therefore, in this study, patients with more than 10% missing data were excluded to ensure that the data were as representative of real-world conditions as possible. A visualization of the missing data proportions for the remaining variables is presented in the Supplementary File part 3. We then employed multiple imputation using chained equations (MICE) to address the missing data, generating 10 imputed datasets for subsequent analysis. Sensitivity analysis using DeLong test for different missing data imputation methods were shown in the Supplementary File part 4.2.

Kaplan–Meier analysis was performed to assess the cumulative incidence of NOAF across postoperative days, with the analysis conducted for each day to illustrate the rate of NOAF development. Quantile–quantile (Q–Q) plots were used to evaluate the normality of

data. Normally distributed variables were presented as mean ± standard deviation, and non-normally distributed data as median and interquartile range. Categorical variables were summarized as counts and percentages. For continuous variables, the independent t-test was used for normally distributed data, and the t'-test for non-normal distributions (Fig. 2). Categorical data were analyzed using the chi-square or Fisher's exact test. Type I error rate was set at 0·05. Statistically significant features were further evaluated using Least Absolute Shrinkage and Selection Operator (LASSO) regression, with optimal penalty determined through 5-fold cross-validation. Non-zero coefficient features were then entered into logistic regression to identify significant predictors. We used Synthetic Minority Oversampling Technique (SMOTE) to address the class imbalance of NOAF positive samples, aiming to make the model more sensitive towards the under-represented NOAF-positive cases. The benefits of SMOTE were shown in Supplementary File part 5.

Optimal hyperparameters for base learners were identified using grid search and 5-fold cross-validation. A total of 11 base learners were constructed, including Logistic Regression (LR), K-Nearest Neighbors (KNN), Naive Bayes (NB), Decision Tree Classifier (DTC), Random Forest (RF), Support Vector Classifier (SVC), eXtreme Gradient Boosting (XGB), Light Gradient Boosting Machine (LGB), Adaptive Boosting Classifier (ABC), CatBoost Classifier (CBC), and Multilayer Perceptron (MLP). These learners were integrated using a Stacking technique, with Logistic Regression serving as the meta-learner. Similarly, grid search and 5-fold cross-validation were also used in ensemble learning to prevent overfitting in Stacking Model. The rationale behind selecting these 11 base learners was to introduce a diverse set of models, each capable of capturing different patterns and relationships in the data. Models such as RF, XGB, and LGB are ensemble techniques known for their ability to capture complex, non-linear relationships and handle high-dimensional data, while LR and NB offer simpler, interpretable linear models. Additionally, KNN offers distance-based learning and MLP incorporates deep learning architecture. This varied selection of models allows the stacking model to leverage the strengths of each approach and mitigate their individual weaknesses, leading to improved model generalization and predictive performance. To prevent data leakage, all preprocessing steps (such as scaling) were performed separately for the training and validation cohorts. Additionally, five-fold cross-validation was applied within the training set to evaluate model performance, ensuring no information from the validation cohort influenced the training process.

Model performance was compared using Accuracy, Precision, Recall, F1 Score, Area Under the Receiver Operating Characteristic Curve (AUC), probability calibration, and the Hosmer–Lemeshow test. To define the "best" model, we prioritized F1 Score if AUC exceeded 0·9, as this indicated strong discriminative ability and a balanced trade-off between precision and recall. Otherwise, we selected the model based primarily on AUC, which provides a comprehensive measure of overall performance, particularly in imbalanced datasets. The best-performing model was then used to develop a user-friendly web-based bedside tool for clinical application. SHAP (Shapley Additive Explanations) values were calculated and plotted to interpret the contributions of individual features to the model's predictions, providing insight into the relative importance of each predictor in assessing NOAF risk.

Ethics statement

We collected de-identified data and adhered strictly to the ethical principles outlined in the Declaration of Helsinki for medical research. Ethical approval was granted by the IRB of the First Affiliated Hospital of Jinan University (Approval No. JNUKY-2021-001). As this was a retrospective study, the IRB waived the requirement for informed consent from patients.

Role of funding source

The funding sources had no role in the design of the study, data collection, analysis, or interpretation, nor in the writing of the manuscript or the decision to submit it for publication. The authors were not paid to write this article by any pharmaceutical company or other agency.

Results

Basic patient characteristics

Initially, data were collected from 4003 patients; after applying inclusion and exclusion criteria, a total of 2994 cases were included in the study: 2486 from Hospital A and 508 from Hospital B (Fig. 1A). The mean age of the derivation cohort, used for model development, was 63·98 years (SD = 6·89), with 75·62% of patients being male. The overall incidence of NOAF was 26·69%. On postoperative third day, the sinus rhythm maintenance rate was 90·64% (95% CI: 89·54%–91·63%). Kaplan–Meier analysis (Fig. 3A) revealed that the highest incidence of NOAF occurred between postoperative days 1 and 5. Detailed demographic and clinical characteristics of the derivation cohort are presented in the Supplementary File part 6.

As illustrated in Fig. 2, univariate analysis identified 21 characteristics with statistically significant differences between the NOAF and non-NOAF groups, including Age, Education level, and Current smoking et al. (all $p < 0·05$). These characteristics were subjected to LASSO regression, which identified 12 variables with non-zero coefficients under the optimal penalty: Age, Left Ventricular Ejection Fraction (LVEF), Left Atrial Diameter (LAD), Left Ventricular End-Diastolic Diameter (LVEDD), Body Mass Index (BMI), Duration of surgery, Brain Natriuretic Peptide (BNP), Neutrophil

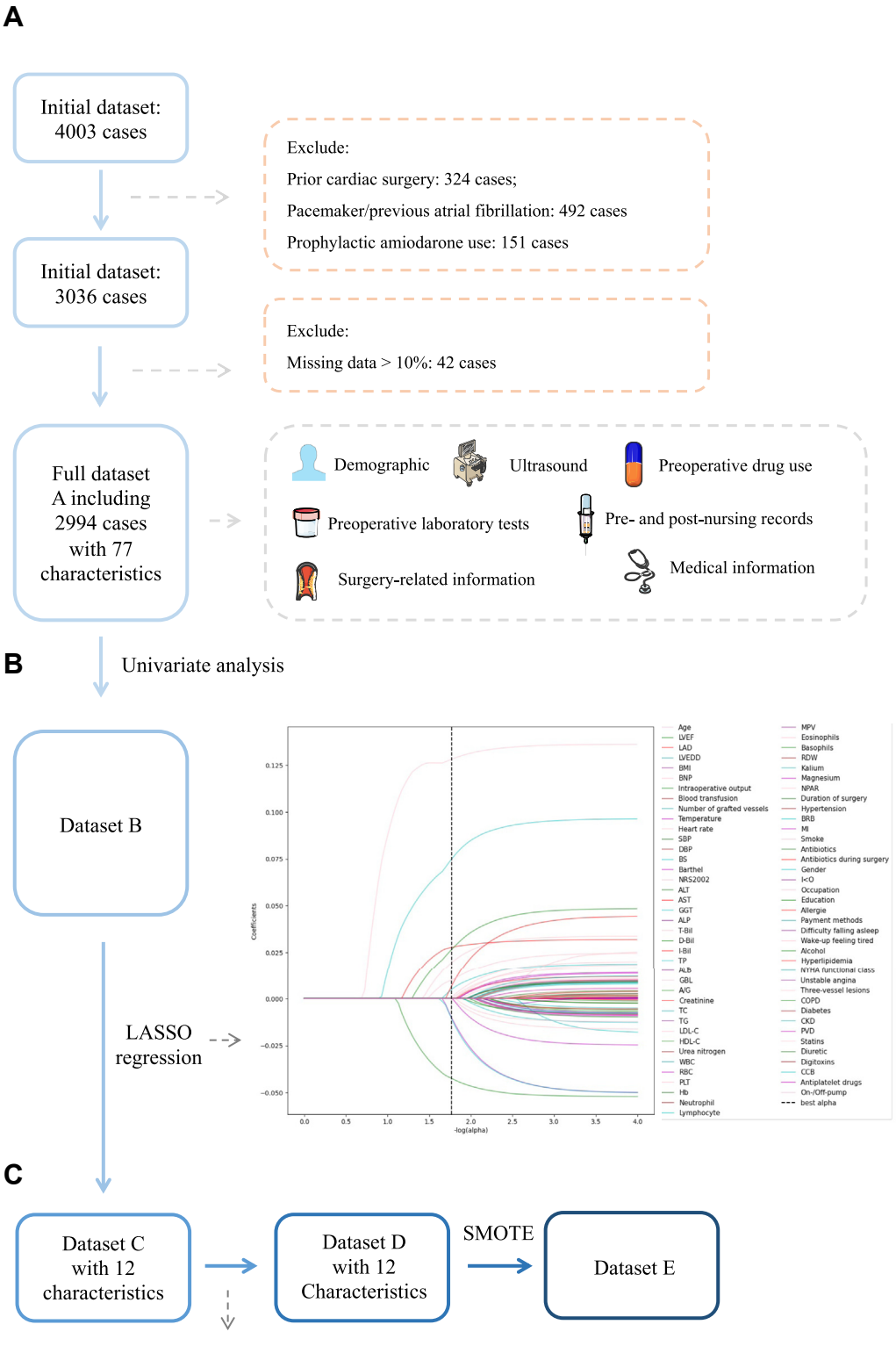

Fig. 2: Note: A/G, ALB/GLB ratio; ALB, albumin; ALP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; ADS, antibiotics during surgery; BMI, body mass index; BNP, brain natriuretic peptide; BRB, β-receptor blockers; BS, blood sugar; CCB, calcium

Percentage-to-Albumin Ratio (NPAR), Hypertension, β-Blocker use (BRB), Myocardial Infarction (MI), and Current smoking. These 12 variables were then incorporated into binary logistic regression (Supplementary File part 7), with the model log-likelihood ratio yielding a p-value < 0·001, indicating statistical significance and a valid model. All 12 variables were found to be significant predictors in the logistic regression analysis.

Stacking model development

As shown in Figs. 2C and 1A, 11, base learners were developed using 12 variables from dataset E as inputs, with NOAF as the predictive label. These base learners were integrated using the Stacking strategy, and both internal and external validation of the model were performed. Fig. 3C and D depict the ROC and decision curve analysis (DCA) results for internal and external validation, respectively. Table 1 shows the performance metrics of each model, with the Stacking model achieving the highest AUC and F1 scores in both the derivation and validation cohorts. This indicates that the Stacking model provides the best overall predictive performance, as confirmed by the Hosmer–Lemeshow test (p = 0·217), and the probability calibration curves are presented in Fig. 3E. Among the base learners, AdaBoost exhibited the highest AUC in the derivation cohort, demonstrating its strong performance. Consequently, SHAP values were plotted based on the AdaBoost model (Fig. 3B), revealing the most important features predicting NOAF, ranked in descending order: BNP, LVEDD, LVEF, BMI, BRB, Duration of surgery, Age, NPAR, MI, LAD, Hypertension, and smoking status. Furthermore, as illustrated in Fig. 3D, the AUC value of Stacking model was higher than those of existing POAF prediction tools in the independent external validation ([POAF score: 0·667, HATCH score: 0·708, CHA2DS2-VASc score: 0·713] vs. 0·931, all p < 0·05), demonstrating superior predictive performance.

The bedside tool

Using the Stacking model, we developed a bedside tool (Fig. 1B) that allows for the automatic calculation of NOAF risk based on simple numeric inputs, providing clinicians with real-time predictive insights.

Discussion

In this study, we successfully developed an AI-based predictive model for NOAF following CABG, followed by the development of a web-based bedside tool based on this model. To the best of our knowledge, this is the most comprehensive study to date in terms of the diversity and number of characteristics used for predicting NOAF in this context. A key strength of our tool lies in the rigorous exclusion of postoperative variables (with only surgical duration retained in the final model), while maintaining satisfactory predictive performance. Consequently, upon completion of the CABG, clinicians possess the full complement of patient characteristics necessary for prediction, facilitating timely postoperative intervention in patients at high risk of NOAF.

Previous studies have demonstrated that postoperative NT-proBNP levels are associated with the development of atrial fibrillation.[19] In our study, we found that elevated preoperative BNP levels also predict NOAF, with BNP being the most important predictive marker. BNP is a key biomarker of myocardial injury; higher levels of BNP correlate with greater myocardial damage, which leads to compensatory tachycardia and, consequently, a higher risk of NOAF.[20] Increased LVEDD and decreased LVEF, associated with left ventricular dysfunction, are also significant predictors of NOAF, indicating notable myocardial fibrosis. These structural changes promote abnormal electrophysiological activity, thereby increasing the likelihood of POAF.[21] Elevated BMI is another risk factor for NOAF, likely due to the enlargement of the left atrium and ventricle, which predisposes obese patients to arrhythmias.[22] The administration of preoperative β-blockers has been shown to prevent POAF by inhibiting sympathetic activity and reducing heart rate, which also contributes to improved myocardial electrical stability.[23] Prolonged surgery duration can further increase the risk of NOAF, potentially due to a combination of sympathetic hyperactivity, myocardial stretching, and ischemia-reperfusion injury during the procedure. These factors may delay sinus node and intra-atrial conduction,

channel blocker; CI, confidence interval; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; D-Bil, direct bilirubin; DBP, diastolic blood pressure; GGT, gamma-glutamyl transferase; GLB, globulin; Hb, hemoglobin; HDL-C, high-density lipoprotein cholesterol; I < O, input is less than output; I-Bil, indirect bilirubin; LAD, left atrial diameter; LASSO, least absolute shrinkage and selection operator; LDL-C, low-density lipoprotein cholesterol; LVEF, left ventricular ejection fraction; LVEDD, left ventricular end diastolic diameter; MI, myocardial infarction; MPV, mean platelet volume; NPAR, neutrophil percentage-to-albumin ratio; NYHA, New York Heart Association; OR, odds ratio; PLT, platelet; PVD, peripheral vascular disease; RBC, red blood cell; RDW, red cell volume distribution width; SBP, systolic blood pressure; SMOTE, synthetic minority oversampling technique; T-Bil, total bilirubin; TC, total cholesterol; TG, triglyceride; TP, total protein; WBC, white blood cell. Part A: Participant flow chart. Finally, a total of 77 predictive characteristics were collected, categorized into eight groups. Part B: Twenty-one features were included in the LASSO regression. The dashed line in the figure indicates the optimal alpha value, with 12 features to the left having non-zero coefficients, proceeding to the next selection step. Part C: Twelve features were subjected to further selection using logistic regression, with results available in Supplementary File part 9. The statistically significant twelve features underwent SMOTE processing to address the imbalance in NOAF positive samples, resulting in the formation of the final dataset E.

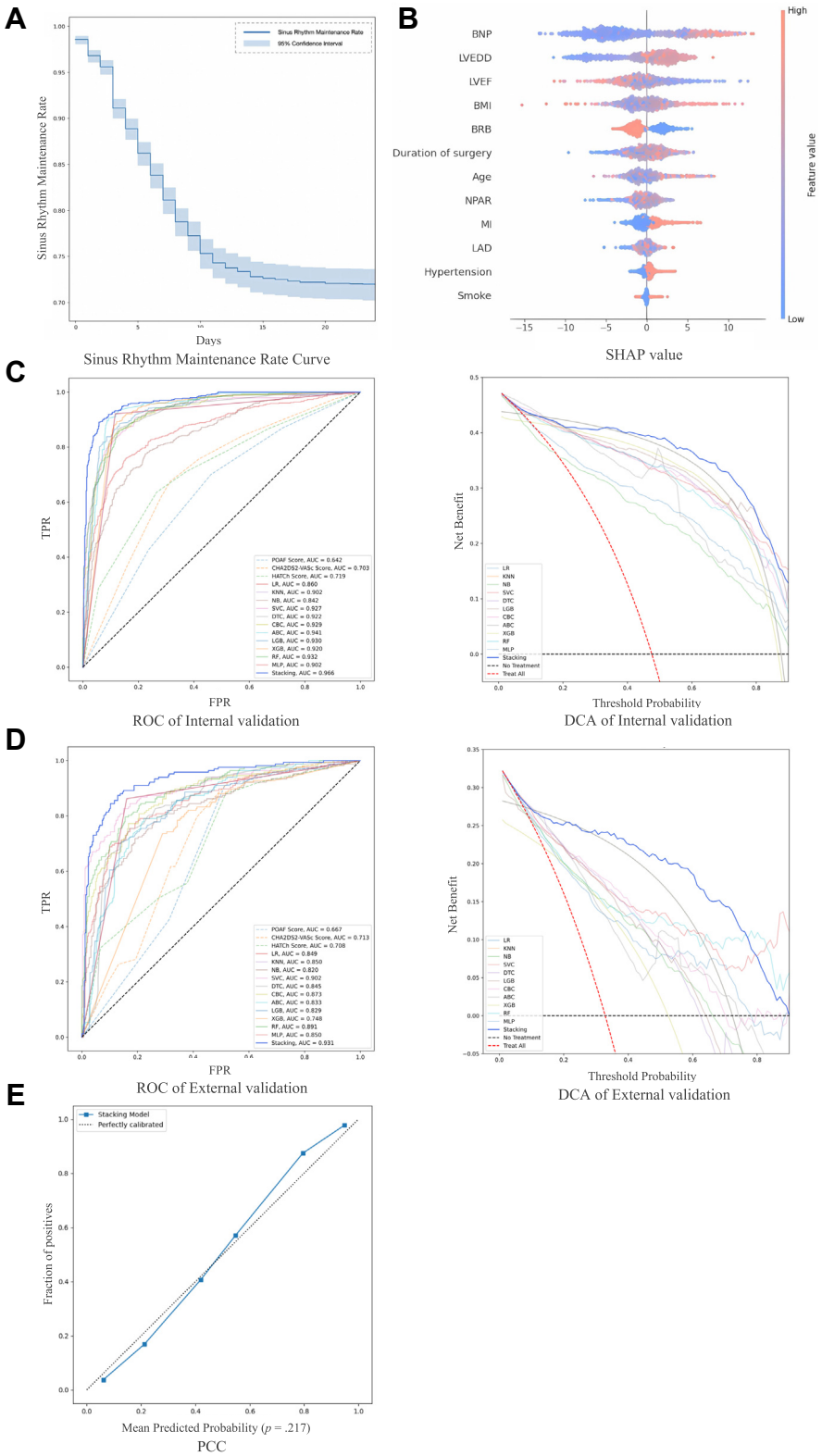

Fig. 3: Note: AUC, area under curve; BNP, brain natriuretic peptide; BRB, β-receptor blockers; CBC, CatBoost Classifier; DCA, decision curve analysis; DTC, Decision Tree Classifier; FPR, false positive rate; KNN, K-Nearest Neighbors; LAD, left atrial diameter; LGB, Light Gradient Boosting

Models	Accuracy	Precision	Recall	F1 score	AUC
Derivation cohort					
LR	0·802	0·810	0·764	0·786	0·860
KNN	0·901	0·878	0·921	0·899	0·902
NB	0·767	0·804	0·675	0·734	0·842
DTC	0·868	0·857	0·868	0·862	0·922
RF	0·864	0·856	0·855	0·858	0·933
SVC	0·857	0·848	0·853	0·850	0·927
XGB	0·882	0·865	0·892	0·878	0·920
LGB	0·873	0·862	0·860	0·868	0·930
ABC	0·899	0·907	0·879	0·893	0·941
CBC	0·855	0·834	0·870	0·851	0·929
MLP	0·878	0·860	0·889	0·874	0·902
Stacking	0·912	0·920	0·893	0·906	0·966
Validation cohort					
LR	0·748	0·584	0·808	0·678	0·849
KNN	0·847	0·724	0·862	0·787	0·851
NB	0·754	0·597	0·773	0·674	0·820
DTC	0·766	0·609	0·802	0·693	0·845
RF	0·799	0·650	0·844	0·734	0·891
SVC	0·815	0·670	0·862	0·754	0·902
XGB	0·695	0·524	0·773	0·625	0·748
LGB	0·766	0·615	0·767	0·683	0·829
ABC	0·784	0·667	0·683	0·675	0·834
CBC	0·772	0·611	0·838	0·707	0·873
MLP	0·770	0·606	0·856	0·710	0·851
Stacking	0·876	0·861	0·743	0·797	0·931

Note: AUC, Area Under Curve; LR, Logistic regression; KNN, K-Nearest Neighbors; NB, Naive Bayes; DTC, Decision Tree Classifier; RF, Random Forest; SVC, Support Vector Classifier; XGB, eXtreme Gradient Boosting; LGB, Light Gradient Boosting Machine; ABC, Adaptive Boosting Classifier; CBC, CatBoost Classifier; MLP, Multilayer Perceptron. The closer the Accuracy, Precision, Recall, F1 score, and AUC values are to 1, the better the model's performance.

Table 1: **Model performance on derivation and validation cohort.**

ultimately triggering POAF.[24] Advancing age is another predictor of NOAF, as cardiac function progressively declines with age. This decline is accompanied by ventricular remodeling, cardiomyocyte loss, and increased myocardial fibrosis, all of which contribute to the development of POAF.[25] In addition, our study identified that elevated NPAR were associated with NOAF, though its predictive significance was less than that of the other factors mentioned above. The relationship between inflammation and AF is well-documented[2]; elevated serum inflammatory biomarkers in AF patients are linked to increased expression of inflammatory markers and cytokines in cardiac tissue. This inflammatory response in the atrial wall may result in oxidative stress, myocardial apoptosis, fibrosis, and disruptions in intracellular calcium handling, all of which contribute to the development of arrhythmias, including AF.[26] NPAR reflects both neutrophil count and serum albumin levels. Neutrophils, as key mediators of inflammation, and albumin, a marker of nutritional status that is affected by inflammation, both play a role in the pathogenesis of NOAF.[17] Low serum albumin levels have also been strongly correlated with poor clinical outcomes in patients with coronary artery disease,[17,27] explaining why NPAR can serve as a predictor of POAF.

Our findings demonstrate that the Stacking model outperforms three established tools for predicting POAF, namely the POAF, CHA2DS2-VASc, and HATCH score. To facilitate its use in clinical settings, we developed a user-friendly bedside tool that automatically calculates the risk of NOAF based on the input of a few key values. Notably, the AI-driven tool surpassed even senior healthcare professionals in identifying NOAF manually. This suggests that the tool has the potential to improve both diagnostic accuracy and consistency, and could be incorporated into medical and nursing decision-making. Manual assessment typically depends on clinical experience and subjective interpretation, which can lead to variability in diagnostic outcomes.[28] Different healthcare providers may draw on varying experiences or interpret the same clinical data differently when assessing NOAF risk. In contrast, our tool offers objective and standardized assessments, ensuring that the evaluation of NOAF risk remains

Machine; LR, Logistic regression; LVEF, left ventricular ejection fraction; LVEDD, left ventricular end diastolic diameter; MI, myocardial infarction; MLP, Multilayer Perceptron; NB, Naive Bayes; PCC, probability calibration curve; RF, Random Forest; ROC, receiver operating characteristic curve; SVC, Support Vector Classifier; TPR, true positive rate; XGB, eXtreme Gradient Boosting; NPAR, neutrophil percentage-to-albumin ratio. Part A: The figure illustrates the maintenance rate of sinus rhythm after coronary artery bypass grafting, indicating that new-onset atrial fibrillation (NOAF) is prevalent in the first 1–5 days postoperatively. Part B: This is a Shapley Additive Explanations (SHAP) diagram. The twelve features presented effectively predict NOAF. The features' importance for predicting NOAF increases as they are ranked higher in the diagram, with preoperative BNP levels being the most significant and smoking history the least significant. Part C: The left side presents the ROC curves of the internal validation set, with each curve representing a different model. The closer a curve is to the top-left corner of the box, the better the model's performance. It is evident that our Stacking model demonstrates the best performance. Furthermore, all our models outperform the existing three tools. The right side displays the clinical decision curves for the internal validation set, where each curve also signifies a model. Again, curves closer to the top-right corner indicate superior model performance, confirming that our Stacking model is the best. Part D: Similar to Part C, this figure illustrates the performance of all models on the external validation set, indicating that our Stacking model performs the best. Part E: This figure illustrates the probability calibration curve, with the blue line representing the performance of the Stacking model; the closer it is to the dashed line, the better the model's performance. The *p*-value from the Hosmer–Lemeshow test is 0·217, which exceeds 0·05, indicating no significant evidence of systematic bias between the model's predicted probabilities and the observed actual outcomes, thus reflecting good model performance.

highly accurate and consistent across different users. However, while the model demonstrates strong overall performance, the Recall could be improved, particularly in patients with mild NOAF. Additionally, patients with complex multimorbidity, where multiple comorbid conditions interact in intricate ways, were more likely to be misclassified. Further refinement of the model to better handle these cases is needed.

This tool was developed and validated to enhance predictive performance on larger datasets. Rather than prompting immediate treatment, such as beta-blockers, it supports clinicians in making more informed decisions regarding patient management, particularly for those at high NOAF risk. This early postoperative NOAF identification allows clinicians to implement preventive strategies promptly, potentially improving outcomes by optimizing management and preventing complications.

This study has several limitations: (1) Although our study presented sinus rhythm maintenance curves for the overall population, we did not investigate the impact of specific patient characteristics on the timing of NOAF. Future research could explore these temporal patterns more thoroughly by employing stratified survival analysis or time-dependent covariate models to assess how various factors (e.g., age, comorbidities, surgery duration) influence the early vs. late onset of NOAF; (2) The retrospective nature of this study meant that the POAF, HATCH, and CHA2DS2-VASc scores were assigned based on available clinical data. While multiple senior clinician assessments were conducted, the retrospective approach may still have introduced selection bias. Future studies should consider using prospective cohort designs to gather real-time data for more accurate score assignment; (3) Another potential limitation is the risk of overfitting, despite efforts to mitigate it through 5-fold cross-validation, grid search optimization of hyperparameters, and independent external validation. The relatively small sample size may still introduce some bias, and the model's performance requires further validation in larger, independent cohorts; (4) While our study demonstrates improved predictive performance of the AI tool over three current scores, it does not yet provide direct evidence of a positive impact on clinical outcomes, and its retrospective design may introduce selection bias. Randomized controlled trials are needed to further validate the predictive tool and assess its real-world clinical impact.

In conclusion, we developed an AI-based Stacking model, along with a bedside tool featuring a user-friendly interface. The tool was both internally and externally validated, demonstrating strong performance in predicting NOAF following CABG. Our tool also outperformed three existing POAF predictive tools, namely POAF scores, CHA2DS2-VASc, and HATCH. Future studies should further explore how various patient characteristics influence the timing of POAF onset, whether early or late.

Contributors

Hualong Ma: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Validation, Software, Visualization, Writing—original draft, Writing—review & editing; Dalong Chen: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Validation, Writing—review & editing; Weitao Lv: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Writing—review & editing; Qiuying Liao: Conceptualization, Investigation, Methodology, Resources, Writing—review & editing; Jingyi Li: Conceptualization, Investigation, Methodology, Writing—review & editing; Qinai Zhu: Conceptualization, Data curation, Resources, Writing—review & editing; Ying Zhang: Conceptualization, Data curation, Resources, Writing—review & editing; Lizhen Deng: Conceptualization, Data curation, Resources, Writing—review & editing; Xiaoge Liu: Conceptualization, Investigation, Resources, Writing—review & editing; Qinyang Wu: Conceptualization, Resources, Writing—review & editing; Xianliang Liu: Conceptualization, Project administration, Supervision, Validation, Writing—review & editing; Qiaohong Yang: Conceptualization, Funding acquisition, Project administration, Supervision, Validation, Writing—review & editing.

Hualong Ma, Dalong Chen, Weitao Lv contributed equally to this work.

Qiaohong Yang, Xianliang Liu contributed equally to corresponding authors.

Hualong Ma, Dalong Chen, Weitao Lv, Qiaohong Yang have directly accessed and verified the data. All authors were responsible for the decision to submit the manuscript.

Data sharing statement

Deidentified participant data collected for this study, along with the data dictionary defining each field in the dataset, will be made available to others upon reasonable request. These data will be available beginning with publication and will remain accessible indefinitely. Requests for data can be directed to the corresponding author, Qiaohong Yang (email: yqiaohong@163.com), or the first author, Hualong Ma (email: Ma980383402@outlook.com). Access to the data will be granted subject to approval of a research proposal and the signing of a data access agreement, in compliance with legal and ethical considerations.

Declaration of interests

None.

Acknowledgements

This work was funded by Lingnan Nightingale Nursing Research Institute of Guangdong Province, and Guangdong Nursing Society (GDHLYJYZ202401).

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.eclinm.2025.103131.

References

1 Eikelboom R, Sanjanwala R, Le M-L, Yamashita MH, Arora RC. Postoperative atrial fibrillation after cardiac surgery: a systematic review and meta-analysis. *Ann Thorac Surg*. 2021;111:544–554.
2 Zhang H, Qiao H, Yang B, et al. Development and validation of a diagnostic model based on left atrial diameter to predict postoperative atrial fibrillation after off-pump coronary artery bypass grafting. *J Thorac Dis*. 2023;15:3708–3725.
3 Engin M, Aydın C. Investigation of the effect of HATCH score and coronary artery disease complexity on atrial fibrillation after on-pump coronary artery bypass graft surgery. *Med Princ Pract*. 2020;30:45.
4 Lotter K, Yadav S, Saxena P, Vangaveti V, John B. Predictors of atrial fibrillation post coronary artery bypass graft surgery: new scoring system. *Open Heart*. 2023;10:e002284.
5 Benedetto U, Gaudino MF, Dimagli A, et al. Postoperative atrial fibrillation and long-term risk of stroke after isolated coronary artery bypass graft surgery. *Circulation*. 2020;142:1320–1329.
6 Shen J, Lall S, Zheng V, Buckley P, Damiano RJ, Schuessler RB. The persistent problem of new-onset postoperative atrial

fibrillation: a single-institution experience over two decades. *J Thorac Cardiovasc Surg*. 2011;141:559–570.
7 Fan K, Chen L, Liu F, et al. Predicting new-onset postoperative atrial fibrillation following isolated coronary artery bypass grafting: development and validation of a novel nomogram. *Int J Gen Med*. 2022;15:937–948.
8 Arslan AK, Erdil N, Guldogan E, Colak C, Akca B, Colak MC. Prediction of postcoronary artery bypass grafting atrial fibrillation: POAFRiskScore tool. *Thorac Cardiovasc Surg*. 2023;71:282–290.
9 Chen Y-L, Zeng M, Liu Y, et al. CHA2DS2-VASc score for identifying patients at high risk of postoperative atrial fibrillation after cardiac surgery: a meta-analysis. *Ann Thorac Surg*. 2020;109:1210–1216.
10 Emren V, Aldemir M, Duygu H, et al. Usefulness of HATCH score as a predictor of atrial fibrillation after coronary artery bypass graft. *Kardiol Pol*. 2016;74:749–753.
11 Mariscalco G, Biancari F, Zanobini M, et al. Bedside tool for predicting the risk of postoperative atrial fibrillation after cardiac surgery: the POAF score. *J Am Heart Assoc*. 2014;3:e000752.
12 Shin TH, Ashley SW, Tsai TC. Defining the role of machine learning in optimizing surgical outcomes. *JAMA Surg*. 2024;159(12):1432. https://doi.org/10.1001/jamasurg.2024.4297.
13 Chinese Society of Cardiology, Chinese Medical Association, Heart Rhythm Committee of Chinese Society of Biomedical Engineering. Chinese guidelines on diagnosis and management of atrial fibrillation. *Chinese J Cardiovasc Dis*. 2023;51:572–618.
14 Swinkels BM, de Mol BA, Kelder JC, Vermeulen FE, Ten Berg JM. New-onset postoperative atrial fibrillation after aortic valve replacement: effect on long-term survival. *J Thorac Cardiovasc Surg*. 2017;154:492–498.
15 A C, Vm C, Yh O, et al. Association of intracranial aneurysm rupture with smoking duration, intensity, and cessation. *Neurology*. 2017;89. https://doi.org/10.1212/WNL.0000000000004419.
16 Cochrane J, Chen H, Conigrave KM, Hao W. Alcohol use in China. *Alcohol Alcohol*. 2003;38:537–542.
17 Sun T, Shen H, Guo Q, et al. Association between neutrophil percentage-to-albumin ratio and all-cause mortality in critically Ill patients with coronary artery disease. *Biomed Res Int*. 2020;2020:8137576.
18 Riley RD, Ensor J, Snell KIE, et al. Calculating the sample size required for developing a clinical prediction model. *BMJ*. 2020;368:m441.
19 Chokengarmwong N, Yeh DD, Chang Y, et al. Elevated admission N-terminal pro-brain natriuretic peptide level predicts the development of atrial fibrillation in general surgical intensive care unit patients. *J Trauma Acute Care Surg*. 2017;83:485–490.
20 Akazawa T, Nishihara H, Iwata H, Warabi K, Ohshima M, Inada E. Preoperative plasma brain natriuretic peptide level is an independent predictor of postoperative atrial fibrillation following off-pump coronary artery bypass surgery. *J Anesth*. 2008;22:347–353.
21 Kochetkov AI, Iyu Orlova, Ostroumova OD, et al. Connection between cardiac fibrosis biomarkers and echocardiography parameters in patients with various forms of atrial fibrillation. *Medicinskij Sovet*. 2024;6:170–179.
22 Zhang J, Zhu C, Nie C, et al. Impact of body mass index on postoperative atrial fibrillation in patients with hypertrophic cardiomyopathy undergoing septal Myectomy. *J Am Heart Assoc*. 2022;11:e023152.
23 Francisco AC, Awata WMC, Lima TS, et al. Three generation β-blockers for atrial fibrillation treatment. *Curr Hypertens Rev*. 2023;19:34–41.
24 Dave S. Incidence and risk factors for development of atrial fibrillation after cardiac surgery under cardiopulmonary bypass. *Indian J Anaesth*. 2019;63:512.
25 Akintoye E, Sellke F, Marchioli R, Tavazzi L, Mozaffarian D. Factors associated with postoperative atrial fibrillation and other adverse events after cardiac surgery. *J Thorac Cardiovasc Surg*. 2018;155:242–251.e10.
26 Mukai Y. Inflammation and atrial fibrillation. *J Arrythm*. 2024;40:26–27.
27 Plakht Y, Gilutz H, Shiyovich A. Decreased admission serum albumin level is an independent predictor of long-term mortality in hospital survivors of acute myocardial infarction. Soroka Acute Myocardial Infarction II (SAMI-II) project. *Int J Cardiol*. 2016;219:20–24.
28 Creavin ST, Noel-Storr AH, Langdon RJ, et al. Clinical judgement by primary care physicians for the diagnosis of all-cause dementia or cognitive impairment in symptomatic people. *Cochrane Database Syst Rev*. 2022;6:CD012558.