

Multilevel predictors categorization for post-CABG atrial fibrillation prediction

Karina I. Shakhgeldyan^{1,2}, Vladislav Y. Rublev^{2,*}, Nikita S. Kuksin², Boris I. Geltser², Regina L. Pak²

¹Center of Artificial Intelligence, Vladivostok State University, Vladivostok 690014, Russia

²Laboratory of Big Data Analysis in Medicine, School of Medicine and Life Sciences, Far Eastern Federal University, Vladivostok 690922, Russia

*Corresponding author. Laboratory of Big Data Analysis in Medicine, School of Medicine and Life Sciences, Far Eastern Federal University, Ajax Village, 10, building 25, Vladivostok 690922, Russia. E-mail: dr.rublev.v@gmail.com

Abstract

Postoperative atrial fibrillation (PoAF) is a common complication after coronary artery bypass grafting (CABG). Despite its association with increased risk of ischemic stroke, bleeding, acute renal failure and mortality there is still no ideal predictive tool with proper clinical interpretability. A retrospective single-center cohort study enrolled 1305 electronic medical records of patients with elective isolated CABG. PoAF was identified in 280 (21.5%) patients. Prognostic models with continuous variables were developed utilizing multivariate logistic regression (MLR), random forest and eXtreme gradient boosting methods. Predictors were dichotomized via grid search for optimal cut-off points, centroid calculation, and Shapley additive explanation (SHAP). For multilevel categorization, we proposed to use threshold values combination identified during dichotomization, as well as ranking cut-off thresholds by MLR weighting coefficients (multimetric categorization method). Based on multistage selection, nine PoAF predictors were identified and validated. After categorization, prognostic models with continuous and multilevel categorical variables were developed. The best XGB model employing continuous predictors demonstrated an AUC = 0.76. Models in which predictors were derived utilizing the multimetric categorization approach showed comparable predictive performance (AUC = 0.758). The main advantage of models with multilevel predictors categorization was their superior explainability and clinical interpretability in predicting POAF. Multilevel predictors categorization represents a promising tool for improving the explainability of POAF predictive development estimates. Using the developed prognostic models, it was demonstrated that the categorization procedures proposed by the authors ensure both high predictive accuracy and transparency of the generated clinical conclusions.

Keywords: prognostic models; multilevel categorization; dichotomization; postoperative atrial fibrillation; stochastic gradient boosting; SHapley Additive exPlanations (SHAP) method

Highlights

1. Validation of new 1st diagnosed atrial fibrillation predictors were performed in patients with coronary heart disease after coronary artery bypass grafting with subsequent development of predictive models utilizing machine learning methods.
2. A new multilevel categorization method was tested, allowing to identify threshold values of predictors with the greatest predictive value, which were classified as risk factors for postoperative atrial fibrillation.
3. The best quality metrics (AUC - 0.802) were demonstrated by a stochastic gradient boosting prognostic model based on predictors identified by the multilevel categorization method.

Introduction

Postoperative atrial fibrillation (PoAF) affects 20%–40% of patients after coronary artery bypass grafting (CABG) [1], with stable rates despite preventive strategies [2, 3] while some authors have even shown a potential trend to increase [4]. PoAF increases the risk of stroke, bleeding, and renal failure approximately fourfold, and doubles mortality at 30 days and 6 months [5]. The lack of a unified pathophysiological model has driven the creation of forecasting tools to personalize risk assessment [6–9].

Among PoAF prediction studies, the PoAF score [7] developed using MLR methods achieved an accuracy by an area under the ROC

curve (AUC) of 0.63–0.65, with 0.6 sensitivity and 0.65 specificity values [8, 10]. Such limited accuracy prompted the use of new machine learning (ML) methods, allowing to improve model quality measured by AUC up to 0.7–0.75 [8, 11]. These models employed continuous and dichotomous predictors, with binary variables used to assess concomitant diseases. However, previous works lacked clinical justification for threshold values used in PoAF risk prediction. Multilevel categorization was only applied to age in some studies, with cut-off points set arbitrarily [7, 11].

This study aimed to develop new prognostic models of PoAF in patients with coronary artery disease after isolated CABG based

Received: 14 September 2025; Revised: 29 November 2025; Accepted: 05 December 2025

© The Author(s) 2025. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

on preoperative predictors set and their multilevel categorization efficiency evaluation to improve prognosis quality and its clinical interpretation.

Material and methods

Data

Single-center cohort retrospective study results are presented, during which electronic health records data of patients with coronary artery disease admitted for planned isolated CABG to the Vladivostok “Primorye Regional Clinical Hospital No. 1” cardiac surgery department from 2008 to 2023 were analyzed. Exclusion criteria included the presence of atrial fibrillation (of any form) in anamnesis, as well as combination of CABG with any other surgery. Thus, the final dataset was represented by 1305 patients (992 men and 313 women) aged 35 to 83 years. The study protocol met local institutional requirements and received full approval; patient consent was not required. Far eastern federal university review board “Ethics approval: IRB protocol number: №1/3,” were approved on 19/03/2023. As the study involved a retrospective review of medical records, the requirement for patient consent was waived. Data were accessed from 19/03/23 to 12/08/23 for research purposes. During this period authors had access to information that could identify individual participants during data collection (DOB and medical record number). To validate the results obtained in the present study, data from the University Clinic of the Far Eastern Federal University (Vladivostok) for the period 2022–2024 were used. The external validation dataset was constructed from electronic medical records in accordance with the inclusion and exclusion criteria of the primary dataset. The test dataset comprised medical records of 200 patients with coronary artery disease who underwent elective coronary artery bypass grafting (CABG).

First diagnosed PoAF episode was considered as the endpoint. AF episodes lasting more than 30 seconds, verified by the results of continuous electrocardiogram monitoring for at least 96 hours after CABG, were considered as PoAF development evidence. The PoAF presence was coded “1,” the absence – “0.” Thus, two patient groups were identified among the examined cohort. The first included 280 (21.5%) patients with AF paroxysms recorded during postoperative period in the hospital, the second—1025 (78.5%) patients without cardiac arrhythmias.

The preoperative clinical and functional status of patients was assessed on the first day of hospital treatment by 130 factors, the main ones of which are presented in [Supplementary Appendix A](#). In addition to demographic, anthropometric, anamnestic data and physical examination results, clinical blood test indicators were analyzed. The diameters of the left (LAD) and right (RAD) atria, longitudinal dimensions of the left (LAL) and right (RAL) atrium, end-systolic (ESD) and diastolic (EDD) dimensions of the left ventricular (LV), ejection fraction (LVEF), and mean pulmonary artery pressure (MPAP) were determined. The ECG results were also analyzed: duration of P wave and QRS complex, PQ, QT intervals and RR.

Statistical methods

Continuous characteristics distribution according to the Kolmogorov-Smirnov test differed from normal, so consequently nonparametric mathematical statistics methods were used for them. The indicators were presented as median (Me) and interquartile ranges (Q1; Q3), the Mann-Whitney test was used for

continuous variables intergroup comparisons, and χ^2 for categorical ones. For binary variables, odds ratios (OR) and their 95% confidence interval (CI) were calculated by Fisher’s exact test. Differences were considered statistically significant at P -value < .05.

Machine learning

PoAF predictive models were developed using MLR, random forest (RF) and eXtreme gradient boosting (XGB) methods. Their quality was assessed by 6 metrics: AUC, sensitivity, specificity, F1-score, positive predictive value (PPV) and negative predictive value (NPV). For optimal hyperparameters selection, the Grid Search Cross-Validation (GridSearchCV) optimization method from sklearn Python library was used.

The dataset was split into 2 samples: for training and cross-validation (80%) and for final testing (20%). The training and cross-validation procedure was performed by stratified k-Fold technique on 10 folds. The average AUC quality metric was used for best model selection, predictors picking and validation, and optimal hyperparameters selection by searching through a grid of acceptable values (GridSearchCV). For final testing, the best MLR, RF and XGB models with optimal parameters and hyperparameters were trained on 80% of the dataset, and tested on the final testing sample (20%). For quality metrics confidence, the assessment procedure was repeated 500 times, followed by metrics averaging, performing the initial division randomly using the bootstrapping method ([Fig. 1](#)). Models were developed by utilizing open-source libraries Python version 3.9.16 (scikit-learn version 0.24.2, xgboost version 1.5.1).

Variables categorization

This study utilized a multilevel categorization method that was previously reported by the authors [[12](#)].

To dichotomize potential predictors, we used grid-step optimization methods $\Delta = (\max - \min) / 100$: P -value minimization—Min(P -value), AUC maximization—Max(AUC), quartile method [[13](#)], centroid method and SHapley Additive exPlanation (SHAP) [[14](#)]. The Shapley method allowed us to identify thresholds at which the predictor influence function on the endpoint demonstrated singularity, which can be observed several times during the range of changes in the continuous attribute values [[12](#)]. To carry out multi-level categorization, we combined all threshold values identified with indicators dichotomization utilizing various methods, including the SHAP method. In this case, close threshold values were combined into one by averaging. The centroid method assumed usage of the analyzed characteristics median in the comparison groups (with and without PoAF) and values equidistant from them (centroids), with the help of which 4 categories were identified for each indicator [[14](#)]. The quartile method involves identifying 4 categories for each variable based on the results of assessing their medians, 2nd and 3rd quartiles [[15](#)].

For indicators endpoint influence degree assessment, the SHapley Additive exPlanation method was used.

Study design

The study design included 6 stages ([Fig. 1](#)). At the very first of them utilizing intergroup comparisons tests, a potential PoAF predictors pool was formed. At the second stage of the study, PoAF prognostic models with predictors in a continuous form were developed by ML methods. The prognostic significance of

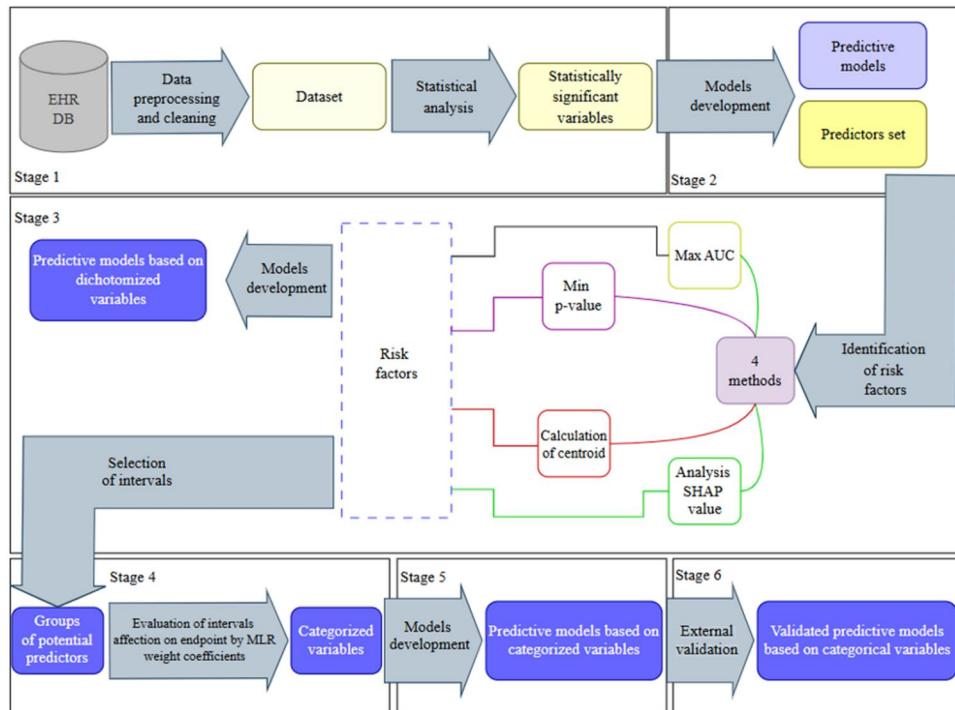


Figure 1 Study design

the predictor was confirmed by AUC value increase after its inclusion in the model. During models development, all variables were considered, regardless of statistically significant differences in comparison groups, and hyperparameters were adjusted at the same stage. Models development and cross-validation was carried out on 80% of the dataset (derivation cohort), and the final testing was carried out on 20% (validation cohort). For further steps, the predictors and hyperparameters obtained at this stage were utilized. At the third stage, using various threshold values identification methods, binarization of continuous variables was carried out using a derivation cohort, and on their basis, PoAF prognostic models were developed, which were validated on the validation cohort. At the fourth stage of the study, multi-level categorization of variables was carried out using 4 approaches. In the first of them, only thresholds identified by the SHAP method were taken into account; in the second, the set of threshold values obtained by other dichotomization methods was expanded. In addition, thresholds obtained by the centroid method were considered, taking into account the medians of the groups with and without PoAF, as well as using quartiles Q1, Q2 and Q3. For risk factors endpoint influence degree assessment, MLR models were developed, whose weights were used to code multilevel categorical predictors. Risk factors with negative or close to 0 weight coefficients in the MLR model were excluded from consideration. At the fifth stage of the study, 4 new PoAF prognostic models were developed by XGB method, the predictors of which were obtained by different methods of multilevel categorization. To assess statistically significant differences between the performance metrics obtained using bootstrapping ($n=100$), 95% CI and pairwise comparisons using the Mann–Whitney U test were applied. At the final stage, external validation of the developed models was performed. The 95% confidence intervals for the selected performance metrics were estimated using the bootstrap method by randomly generating 2000 resampled datasets, each

containing $N - 10$ observations, where N denotes the number of records in the external dataset.

Results

Subject characteristics

Intergroup analysis of clinical, demographic and laboratory parameters demonstrated that patients with PoAF were distinguished by older age, an increased prevalence of tricuspid regurgitation (TR) among them, lower levels of platelets, total protein and triglycerides in the blood. Individuals in this group had higher values of LV ESD, LAD, RAD and RAL, Ao/LV systolic pressure gradient and an increased duration of the QT and PQ intervals (Table 1 and Supplementary Appendix A).

Machine learning models

During the second stage, PoAF prognostic models were developed, validated and tested utilizing RF, XGB and MLR methods. For all models, the best AUC metric results were obtained by usage of ECG indicators (duration of QRS, QT, PQ, RR, and P wave intervals), age, RAD, ESD, and TR as predictors. Developed models predictive value comparison showed that the XGB and RF methods provide higher forecast accuracy compared with MLR (AUC—0.76 and 0.752 vs 0.697) (Table 2). To assess potential overfitting, model performance was additionally evaluated on the training dataset (Supplementary Appendix B). The absolute difference between performance metrics obtained in the training and testing sets did not exceed 0.05, indicating adequate model generalizability and the absence of clinically relevant overfitting. Supplementary Appendix B shows MLR model weight coefficients.

Categorization

During third stage, PoAF predictors in a continuous form were dichotomized utilizing searching for the optimal cutoff threshold

Table 1. Clinical and functional characteristics of patients with coronary artery disease.

Predictor	Group 1 (n = 280)	Group 2 (n = 1025)	OR (95% CI)	P-value
Age, years	66 (61; 71)	63 (58; 69)	–	<.000001
Female, abs. (%)	76 (27.14%)	237 (23.12%)	1.24 [0.917; 1.673]	.188
Height, cm	170 (165; 176)	170 (165; 176)	–	.802
Weight, kg	80 (73; 90)	80 (73; 90)	–	.86
BMI, kg/m ²	27.9 (25.4; 31)	28.1 (25.8; 31)	–	.27
LVEF, %	60 (54; 64)	60 (51; 65)	–	.89
RLVMI, c. u.	1.01 (0.86; 1.18)	0.98 (0.84; 1.16)	–	.263
RTI, c. u.	0.408 (0.363; 0.455)	0.417 (0.37; 0.458)	–	.442
LV ESD, cm	3.3 (3.2; 3.7)	3.35 (3; 3.8)	–	.185
LV EDD, cm	5.1 (4.8; 5.425)	5.1 (4.7; 5.4)	–	.284
Systolic pressure gradient Ao/LV, mm Hg	7 (5; 9)	6 (5; 8)	–	.051
MPAP, mm Hg	25 (23; 30)	25 (23; 30)	–	.221
LAL, cm	3.9 (3.6; 4.2)	3.9 (3.6; 4.3)	–	.339
LAD, cm	4.6 (4; 5.1)	4.3 (3.8; 4.9)	–	.0002
Indexed LA volume, ml/m ²	33.3 (24.5; 43.1)	30.2 (23.7; 39.3)	–	.125
RAL, cm	3.8 (3.5; 4.1)	3.7 (3.4; 4)	–	.0023
RAD, cm	4.5 (4.1; 4.9)	4.3 (3.8; 4.8)	–	.0000033
P, ms	100 (100; 100)	100 (100; 100)	–	.69
PQ, ms	160 (140; 180)	150 (140; 180)	–	.025
QRS, ms	80 (80; 100)	80 (80; 100)	–	.00036
RR, ms	950 (895; 1090)	900 (800; 1080)	–	.129
QT, ms	400 (360; 420)	380 (360; 420)	–	.00143
Creatinine, μmol/l	92.71 (79; 110)	97 (83; 110)	–	.036
GFR, ml/min	158.8 (135.1; 188.7)	162.8 (135.1; 190.9)	–	.43
CHF III-IV FC, abs. (%)	43 (15.36%)	124 (12.1%)	1.56 [0.76; 1.47]	.169
Extracardiac arteriopathy	96 (34.3%)	343 (33.46%)	1.04 [0.79; 1.38]	.85
AH, abs. (%)	225 (80.35%)	813 (79.31%)	1.7 [0.81; 3.7]	.81
Aortic stenosis, abs. (%)	6 (2.1%)	20 (1.95%)	1.24 [0.49; 3.15]	.74
TR, abs. (%)	52 (18.57%)	112 (10.93%)	1.8 [1.26; 2.6]	.0009
MR, abs. (%)	89 (31.78%)	296 (28.88%)	1.13 [0.85; 1.51]	.38
AR, abs. (%)	20 (7.15%)	67 (6.5%)	1.12 [0.66; 1.88]	.82
CKD, abs. (%)	75 (26.78%)	262 (25.56%)	1.07 [0.79; 1.43]	.74
COPD, abs. (%)	45 (16.07%)	137 (13.36%)	1.2 [0.72; 1.95]	.29
DM, abs. (%)	68 (24.28%)	257 (25.07%)	0.95 [0.7; 1.3]	.82
Previous stroke, abs. (%)	27 (9.64%)	89 (8.68%)	1.2 [0.71; 1.76]	.70
Hemoglobin, g/l	142 (131; 152)	142 (131; 153)	–	.79
Red blood cells, 10 ¹² /l	4.69 (4.31; 5.03)	4.69 (4.30; 5.06)	–	.97
Leukocytes, 10 ⁹ /l	6.8 (5.7; 8)	6.9 (5.8; 8.37)	–	.16
Lymphocytes, 10 ⁹ /l	1.87 (1.44; 2.42)	2 (1.53; 2.53)	–	.09
Platelets, 10 ⁹ /l	228 (182; 266)	232 (192; 273)	–	.0336
Total cholesterol, mmol/l	4.35 (3.66; 5.4)	4.45 (3.7; 5.43)	–	.15
Glucose, mmol/l	5.63 (5.11; 6.21)	5.67 (5.13; 6.52)	–	.144
Total protein, g/l	70.9 (66.3; 73.7)	71.4 (67.9; 75.2)	–	.004
Total bilirubin, μmol/l	16.96 (12.1; 23.813)	16.3 (11.7; 22.95)	–	.338
Triglycerides, mmol/l	1.48 (1.16; 1.88)	1.63 (1.2; 2.2)	–	.003
Urea, mmol/l	6 (5; 7.13)	6 (4.97; 7.35)	–	.954
Thrombin time, s	19.6 (16.6; 21.5)	19.9 (17.1; 21.4)	–	.67
PTI, %	93.3 (86; 99.6)	94 (86.88; 102)	–	.148
INR	1.05 (1; 1.13)	1.03 (0.98; 1.1)	–	.073
SBP, mm Hg	130 (130; 150)	130 (125; 140)	–	.0229
DBP, mm Hg	80 (75; 80)	80 (70; 80)	–	.00775
Heart rate, beats/min	68 (62; 75)	68 (62; 72)	–	.178

CI = confidence interval; BMI = body mass index; LV = left ventricle; LVEF = LV ejection fraction; RLVMI = relative left ventricular myocardial mass index; RTI = relative thickness index of left ventricle posterior wall; LV ESD = end systolic dimension; LV EDD = end diastolic dimension; MPAP = mean pulmonary artery pressure; LAL = left atrium medial-lateral size; LAD = left atrium anterior-posterior size; RAL = right atrium longitudinal size; RAD = right atrium transverse size; GFR = glomerular filtration rate; CHF = congestive heart failure; FC = functional class; AH = arterial hypertension; TR = tricuspid regurgitation; MR = mitral regurgitation; AR = aortic regurgitation; CKD = chronic kidney disease; COPD = chronic obstructive pulmonary disease; DM = diabetes mellitus; PTI = prothrombin time index; INR = international normalized ratio; SBP = systolic blood pressure; DBP = diastolic blood pressure.

on the grid methods [Min(P-value) and Max(AUC)], along with the SHAP and centroid calculation method (Table 3). Threshold values usage, deviation from which is associated with PoAF likelihood increase, allows us to consider binarized data as risk factors for adverse events. The risk factor is coded “1” if the predictor value exceeds the threshold with the postfix “+” or does not reach it—with the postfix “–,” in other cases—“0.”

Study results showed substantial variation in threshold values across binarization methods. For example, the cutoff point for

QRS according to SHAP was 80 ms, while when maximizing AUC, the cutoff point was fixed at 81 ms, and for the P wave, values above 130ms were risk factors, while for Max(AUC)—above 100 ms (Table 3). The first three dichotomization methods considered isolated indicators and did not take into account predictive models. The SHAP method was applied to a multifactorial XGB model and the threshold value were defined as the point where the shap-value exceeded the level of 0.2 arbitrary units. Thus, due to dichotomization, the following PoAF risk factors

Table 2. Assessment of the accuracy of prognostic models for PoAF using predictors in continuous form.

Metrics	Cross-validation			Final testing		
	MLR	XGB	RF	MLR	XGB	RF
AUC	0.696 [0.693; 0.699]	0.762 [0.759; 0.765]	0.755 [0.752; 0.758]	0.697 [0.688; 0.706]	0.76 [0.752; 0.768]	0.752 [0.745; 0.76]
Sen	0.655 [0.65; 0.66]	0.693 [0.689; 0.698]	0.684 [0.679; 0.69]	0.655 [0.638; 0.673]	0.693 [0.679; 0.708]	0.683 [0.668; 0.698]
Spec	0.64 [0.637; 0.642]	0.691 [0.688; 0.694]	0.701 [0.698; 0.704]	0.643 [0.635; 0.651]	0.686 [0.677; 0.694]	0.697 [0.689; 0.706]
PPV	0.306 [0.304; 0.309]	0.354 [0.351; 0.357]	0.359 [0.355; 0.362]	0.308 [0.302; 0.316]	0.35 [0.344; 0.357]	0.356 [0.349; 0.363]
NPV	0.886 [0.885; 0.888]	0.905 [0.903; 0.906]	0.903 [0.901; 0.905]	0.885 [0.88; 0.89]	0.464 [0.456; 0.472]	0.901 [0.897; 0.905]
F1-score	0.415 [0.412; 0.418]	0.466 [0.463; 0.469]	0.468 [0.464; 0.472]	0.419 [0.41; 0.428]	0.464 [0.456; 0.472]	0.467 [0.458; 0.475]

Table 3. PoAF continuous predictors dichotomization using different methods.

Predictors	Min (P-value)	Max (AUC)	Centroid	SHAP
Age, years	60.0+	60.0+	64.0+	61+
LV ESD, cm	3.0+	3.0+	3.33+	[3.1; 4.1]
RAD, cm	4.17+	4.17+	4.4+	[4.2; 5.3]
QRS, ms	81-	81-	80-	80-
QT, ms	420+	382+	390+	390+
PQ, ms	163+	163+	155+	[170; 210]
RR, ms	882+	882+	925+	[700; 750], [880; 1000], 1100
P, ms	120+	100+	100+	130+

LV = left ventricle; LV ESD = end systolic dimension; RAD = right atrium transverse size.

were identified: age over 61 years, RAD more than 4.2 cm, ESD—3.1 cm, QRS duration less than 80 ms, QT more than 390 ms, P—130 ms, PQ—170 ms, RR—700 ms (Fig. 2).

Annotation: The blue and red dotted lines indicate the cutoff thresholds. Abbreviations: LV—left ventricle, ESD—LV end systolic dimension, RAD—right atrium transverse size.

Using the QRS diagram as an example (Fig. 2), it can be seen that the probability of PoAF developing in the range from 40 to 80 ms remains consistently high, but sharply decreases at its values ≥ 90 ms. Exceeding the QT parameter value more than 390 ms increases the risk of arrhythmia, but its maximum probability is fixed when the QT value is above 450 ms. Assessing the dynamics of changes in shap-value allows us to explain the relationship between various predictor values and study endpoint, which was the basis for utilizing this method in multilevel categorization procedures. For example, the distribution of the PQ interval demonstrates that the highest risk of POAF is observed among patients whose PQ values fall within the range of 170–210 ms.

On the basis of MLR POAF prognostic models with dichotomous predictors were developed (Table 4). It was found that the model constructed using the centroid method was significantly inferior to the models developed using the Max(AUC) and SHAP-based approaches, which demonstrated acceptable predictive performance (AUC: 0.739–0.744).

At the fourth stage of the study, utilizing various multilevel categorization methods, 4 groups of PoAF risk factors were formed. The first pool of risk factors was obtained from the shap-value analysis results (Fig. 2). The second pool expanded the first due to threshold values obtained at the third stage of the study by several dichotomization methods. The third group of risk factors was represented by the predictors medians in the comparison groups and their centroids, and the fourth group used threshold values corresponding to predictors quartiles. To

encode multilevel categorical predictors values, we used the weight coefficients (WC) of the MLR models developed for each risk factors group (Table 5). We call the approach that ensures the formation of a second pool of risk factors and their WC—multimetric categorization.

Models based on multilevel categorical predictors

At the fifth stage of the study, based on multilevel predictors obtained by various methods, 4 PoAF prognostic models were developed by XGB (Table 6). The best predictive properties were demonstrated by the model with predictors identified by the multilevel categorization method. The latter had comparable accuracy with the models including continuous variables or risk factors (AUC 0.758 and 0.76). In addition, this model enabled explanation of POAF risk based on the assessment of threshold values and the relative contributions of its predictors (Table 5). Taking these findings into account, the highest probability of POAF in patients with CAD after CABG was associated with a P-wave duration ≥ 130 ms (WC = 1.67), RAD of 4.2–5.3 cm (WC = 1.161), RR interval of 700–750 ms (WC = 1.159), PQ interval of 170–210 ms (WC = 0.987), and ESD ≥ 5 cm (WC = 0.956). An association with POAF was also identified for ESD in the range of 3.1–4.1 cm, QT interval ≥ 390 ms, QRS duration ≤ 80 ms, RR intervals of 880–1100 ms and ≥ 1100 ms, P-wave duration of 100–130 ms, age > 60 years, and the presence of CHF and TR.

A comparative accuracy assessment between prognostic models with predictors identified by dichotomization and by multimetric multilevel categorization methods demonstrated the advantages of the latter, which was confirmed by statistically significant AUC metric differences (P-value $< .001$).

The assessment of individual PoAF predictors' influence on its development was performed utilizing the SHAP and XGB methods (Fig. 3). The strongest influence is demonstrated by the QT indicator (shap-value 0.94) after 450 ms, TR presence, the duration of the RR intervals (700–1100 ms) and RAD (4.2–5.3 cm). A low PoAF development probability is associated with the younger age (patients under 60 years), ESD below 3 cm and RAD up to 4.1 cm, PQ interval less than 150 ms and QRS above 100 ms.

At the final stage of the study, the predictive performance of POAF models with continuous predictors and those with predictors derived using the multimetric categorization approach was evaluated on an external cohort of patients treated at the University Clinic of the Far Eastern Federal University (Table 7). The validation results demonstrated comparable predictive accuracy of the MLR, SGB, and SL models with continuous predictors (Table 2) and the SGB models with categorical predictors (Table 6). The similar predictive performance observed during both model development and external validation may be explained by the use of contemporary modeling approaches as well as by the set of selected predictors (patient age, ECG and echocardiographic parameters), which are largely independent of

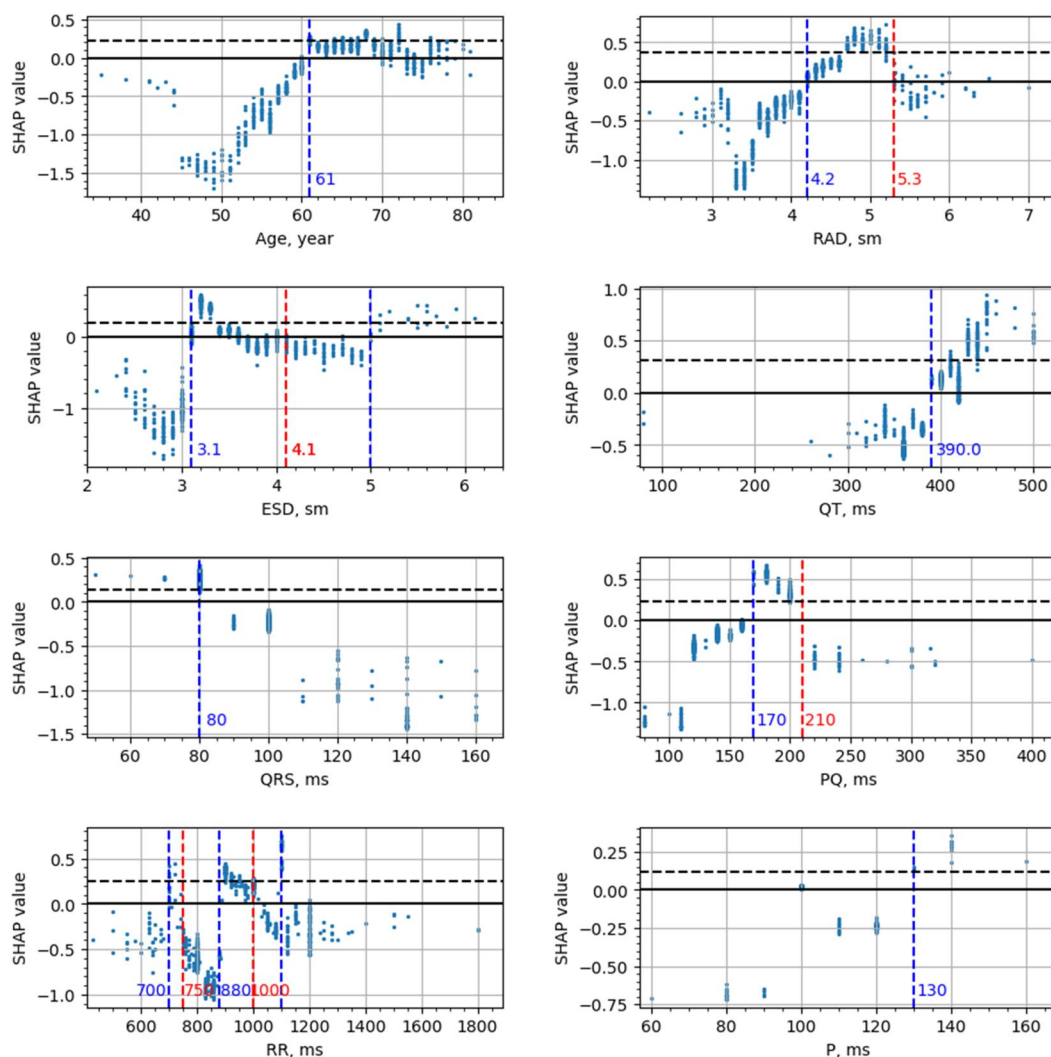


Figure 2 Continuous indicators and their threshold values endpoint impact assessment by the SHAP method. The blue and red dotted lines indicate the cutoff thresholds. ESD = LV end systolic dimension, LV = left ventricle, RAD = right atrium transverse size

Table 4. Performance evaluation of POAF prognostic models based on dichotomous predictors.

Metrics	Min (P-value)	Max (AUC)	Centroid	SHAP
AUC	0.722 [0.715; 0.731]	0.739 [0.732; 0.747]	0.628 [0.618; 0.637]	0.744 [0.736; 0.752]
Sen	0.653 [0.637; 0.669]	0.683 [0.671; 0.695]	0.601 [0.583; 0.618]	0.698 [0.681; 0.714]
Spec	0.661 [0.653; 0.669]	0.696 [0.689; 0.703]	0.6 [0.587; 0.613]	0.644 [0.634; 0.653]
PPV	0.319 [0.313; 0.326]	0.354 [0.348; 0.362]	0.269 [0.262; 0.276]	0.323 [0.317; 0.329]
NPV	0.887 [0.882; 0.892]	0.901 [0.897; 0.904]	0.861 [0.856; 0.866]	0.898 [0.894; 0.903]
F-score	0.428 [0.419; 0.437]	0.466 [0.458; 0.473]	0.369 [0.361; 0.379]	0.441 [0.433; 0.448]

the professional, organizational, and resource-related characteristics of healthcare institutions.

Conclusion

In recent years, prognostic models utilizing ML methods are being developed, wide usage of which in clinical practice is limited by the complexity of prognostic results interpretation. Promising tools for this problem solving are explainable artificial intelligence (XAI) algorithms, the elements of which includes the predictors threshold values determination and their ranking according to their influence intensity on the endpoint. Predictors

threshold values determination is carried out using their categorization, which allows to detail the relationship between indicators of the clinical and functional patients status with the resulting variable. According to the literature, the most accessible method of multilevel categorization is descriptive statistics with the medians, quartiles or quantile calculation [16]. However, most of the categorization criticisms are associated precisely with this approach, which is primarily due to the dependence of such threshold values on a specific sample, lack of relationship with the clinical context, ignoring possible non-linear relationships, etc [16]. An alternative method that takes into account the clinical context is searching for optimal

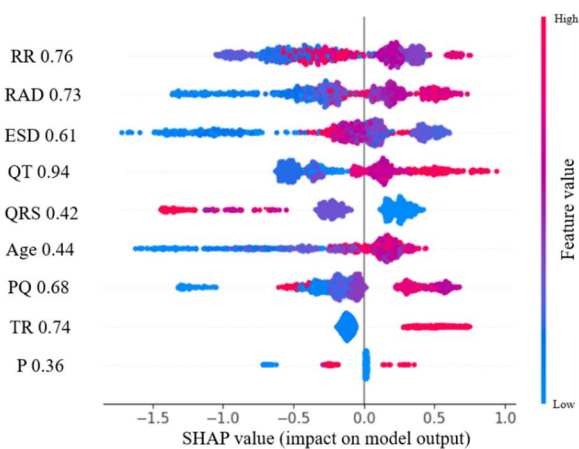
Table 5. Predictors weight coefficients and thresholds obtained by multilevel categorization methods.

Predictors	SHAP		Multimetric categorization		Group medians and centroids		Quartiles	
	Thresholds	WC	Thresholds	WC	Thresholds	WC	Thresholds	WC
Age, years	61+	0.683	60+	0.729	[63; 66]		[58; 64]	
					66+	0.397	[64; 69]	0.72
						0.317	69+	0.874
ESD, cm	[3.1; 4.1]	0.958	[3.1; 4.1]	0.915	[3.3; 3.33]		[3.1; 3.3]	
	5+	1.154	5+	0.956	3.33+	0.674	[3.3; 3.8]	1.422
						0.18	3.8+	1.061
							0.886	
RAD, cm	[4.2; 5.3]	0.917	[4.2; 5.3]	1.161	[4.4; 4.5]	0.361	[3.8; 4.3]	0.19
							[4.3; 4.8]	0.307
QRS, ms	80-	0.255	80-	0.253	80-	0.381	[80; 100]	0.662
QT, ms	390+	0.542	390+	0.549	[380; 400]		[395; 420]	
					400+	0.173	420+	0.839
						0.59		0.661
RR, ms	[700; 750]	1.145	[700; 750]	1.159	950+	0.219	[920; 1080]	
	[880; 1100]	0.911	[880; 1100]	0.926			1080+	0.331
	1100	0.740	1100	0.738				0.091
PQ, ms	[170; 210]	0.98	[170; 210]	0.987	155+	0.596	[160; 180]	
							180+	0.393
								0.573
P, ms	130+	1.335	[100; 130]	0.625	100+	1.28	100+	1.41
			130+	1.67				
TR	1	0.734	1	0.726	1	0.68	1	0.76

WC = weight coefficient; LV = left ventricular; ESD = LV end systolic dimension; RAD = right atrium transverse size; TR = tricuspid regurgitation.

Table 6. Accuracy assessment of PoAF prognostic models based on predictors with multilevel categorization.

Metrics	Multilevel SHAP	Multimetric categorization	Group medians and centroids	Quartiles
AUC	0.754 [0.747; 0.761]	0.758 [0.751; 0.765]	0.663 [0.654; 0.672]	0.681 [0.672; 0.69]
Sen	0.657 [0.642; 0.6717]	0.665 [0.65; 0.679]	0.612 [0.596; 0.627]	0.635 [0.621; 0.65]
Spec	0.688 [0.679; 0.697]	0.691 [0.683; 0.699]	0.6223 [0.615; 0.631]	0.641 [0.634; 0.649]
PPV	0.339 [0.333; 0.346]	0.344 [0.338; 0.351]	0.386 [0.378; 0.394]	0.301 [0.295; 0.307]
NPV	0.893 [0.889; 0.897]	0.895 [0.891; 0.899]	0.283 [0.277; 0.289]	0.879 [0.875; 0.883]
F1-score	0.447 [0.439; 0.455]	0.453 [0.445; 0.461]	0.869 [0.864; 0.874]	0.408 [0.4; 0.416]

**Figure 3** Assessment of predictor impact on PoAF using the SHAP method

threshold values based on minimization or maximization of objective functions, such as Min(P-value) or Max(AUC).

Recent literature about PoAF prediction problem analysis showed that utilizing categorization methods, PoAF risk factors were identified, which included the age of patients over 60 years

[7, 17], 66 years [5, 18] or 70 years [19], increased LA size [5, 20], including with LAD > 4.5 cm [18] or > 3.9 cm [11] and reduced LVEF < 30% [7], increased P wave duration according to standard (>116 ms) [21, 22]. A number of anamnestic data are also validated as risk factors for PoAF: male gender [5, 23], the presence of arterial hypertension, chronic heart failure, chronic obstructive pulmonary disease, chronic kidney disease, diabetes mellitus [5], rheumatic heart disease [18, 23], mitral valve disease [3], previous cardiac surgery, metabolic syndrome and obesity [5, 6]. In our study, anamnestic features did not demonstrate predictive potential. The TR and RAD indicators were firstly verified as PoAF predictors. Our study confirmed the predictive value of age and ECG in relation to PoAF, but did not reveal a relationship between laboratory data and LVEF with the PoAF development.

Utilizing the patients with coronary artery disease after the CABG database example, we analyzed the effectiveness of various predictors threshold values searching methods, deviations from which increased their predictive potential and allowed to attribute them as PoAF risk factors. It was identified that the SHAP method, which considered as one of the promising XAI technologies, is a useful categorization tool due to the effective determination of cut-off thresholds, in particular for multilevel categorization and predictors relationship analysis both in continuous and categorical forms with the study endpoint. At the

Table 7. External validation of POAF prognostic models using predictors in continuous and categorical form.

Metrics	Continuous form of predictors			Categorical form of predictors		
	MLR	XGB	RF	MLR	XGB	RF
AUC	0.673 [0.664; 0.682]	0.787 [0.779, 0.795]	0.766 [0.758, 0.774]	0.773 [0.764; 0.781]	0.775 [0.767, 0.783]	0.757 [0.749, 0.765]
Sen	0.56 [0.545; 0.575]	0.698 [0.684; 0.712]	0.765 [0.752; 0.778]	0.709 [0.695; 0.723]	0.715 [0.701; 0.729]	0.715 [0.701, 0.729]
Spec	0.601 [0.593; 0.608]	0.689 [0.681; 0.696]	0.644 [0.637; 0.651]	0.693 [0.686; 0.7]	0.717 [0.71, 0.724]	0.704 [0.697; 0.711]
PPV	0.264 [0.256; 0.272]	0.385 [0.375; 0.394]	0.37 [0.362; 0.378]	0.391 [0.381; 0.4]	0.414 [0.404; 0.424]	0.4 [0.391; 0.41]
NPV	0.85 [0.845, 0.855]	0.908 [0.903; 0.912]	0.923 [0.919; 0.928]	0.911 [0.906; 0.915]	0.915 [0.911; 0.919]	0.914 [0.91; 0.918]
F-score	0.348 [0.339; 0.358]	0.689 [0.681; 0.696]	0.484 [0.475; 0.493]	0.485 [0.475; 0.495]	0.504 [0.494; 0.515]	0.495 [0.485; 0.504]

same time, multilevel categorical predictors obtained by combining SHAP data with other dichotomization methods results have been shown to provide higher predictive accuracy. Potential risks of information loss during new categorization methods usage were overcome by detailing knowledge about the interconnection of individual risk factors with the study endpoint. This was confirmed by predictive models quality criteria comparison for predictors both in continuous and multilevel categorical forms. Thus, for the best model with continuous predictors, AUC was 0.76, while utilizing multimetric categorization, it was 0.758. Categorization methods based on descriptive statistics did not compensate for the loss of predictive accuracy through the introduction of additional information.

It should also be noted that our models demonstrated higher predictive accuracy than those reported in a previously published study that relied exclusively on preoperative variables and employed MLR, RF, and XGB methods (best AUC in the present study: 0.76 vs. 0.74 in [8]). XGB-based models yielded the best overall performance metrics. Marked differences were observed between predictor importance estimates obtained using SHAP and those based on regression coefficients in the MLR models. Specifically, in our study, the greatest importance for achieving the endpoint according to MLR coefficients was associated with a P-wave duration exceeding 130 ms, whereas the highest risk according to the Shapley method was linked to QT > 450 ms. These findings indicate the need for further research aimed at quantifying the magnitude of predictor effects on the clinical endpoint, which is of critical importance for the clinical interpretation of predictive outcomes.

In the present study, machine learning methods (MLR, XGB, and RF) and two approaches to the multilevel categorization of POAF predictors—multimetric categorization and the SHAP-based method—were evaluated using a database of patients with coronary artery disease after isolated coronary artery bypass grafting. Using the developed prognostic POAF models, it was demonstrated that the proposed categorization procedures provide both high predictive accuracy and enhanced transparency of risk prediction results.

Clinical significance

The application of prognostic models developed using preoperative predictors and modern machine learning methods enables accurate estimation of the risk of postoperative atrial fibrillation in patients with coronary artery disease undergoing coronary artery bypass grafting. Multilevel categorization of predictors enhances the transparency of risk prediction and serves as a valuable tool for justifying the selection of individualized preventive strategies.

Dataset limitations

The limitations of this study are primarily related to its retrospective design and the data collection mainly from a single medical center. External validation was performed on smaller sample and only from one healthcare institute. A second limitation is that most of the analyzed variables (with the exception of sex and age) contained missing values, with the proportion of missing data ranging from 0.08% to 25%. The missingness was assumed to be random, making a systematic bias unlikely. Finally, several clinical and functional characteristics that have previously been validated as predictors of POAF in other studies were not included in the present analysis. We plan to digitize and incorporate these variables in future research. The full source code for the software implementation of the method is

Limitations imposed by model overfitting

Increasing the complexity of machine-learning models by incorporating numerous predictors or by deepening tree structures in ensemble algorithms is well known to increase the risk of overfitting [24]. In the present study, unexpectedly high predictive performance was initially observed when only ECG-derived variables (PQ, QRS, RR, QT intervals and P-wave duration) were used in ensemble learning methods (XGB and RF), whereas the MLR model based on the same predictors demonstrated poor discriminative ability (AUC = 0.58).

Further analysis revealed pronounced overfitting of the SGB model despite the limited number of predictors: the AUC reached 0.999 on the training dataset but dropped to 0.78 on the test dataset. Moreover, any modification of the predictor structure, including categorical transformation, resulted in a substantial deterioration of model performance (AUC = 0.60), confirming model instability under these conditions.

To mitigate overfitting, additional clinically relevant predictors (RAD, age and ESD) were incorporated. This adjustment ensured both high predictive performance on the test dataset (AUC = 0.76) and stable generalization, as evidenced by near-identical AUC values for the training and test sets (0.762 and 0.76, respectively). Importantly, predictor dichotomization did not lead to a significant loss of discriminative ability, and the predictive accuracy of the MLR model increased markedly compared with the ECG-only model (AUC 0.697 vs. 0.58).

The full source code for the software implementation of the method is available at the following link: <https://github.com/NikitaKuksin/MultilevelPredictorsCategorizationForAtrialFibrillationPredictionInPatientsWithCoronaryHeartDisease.git>

Author contributions

Karina I. Shakhgeldyan (Writing—original draft, Visualization, Validation, Software, Methodology, Funding acquisition, Data curation), Vladislav Y. Rublev (Writing—original draft, Resources,

Supervision, Formal analysis, Data curation, Conceptualization), Nikita S. Kuksin (Writing—original draft, Investigation, Funding acquisition), Boris I. Geltser (Writing—original draft, Writing—review & editing, Supervision, Conceptualization, Resources, Methodology), and Regina L. Pak (Writing—original draft, Conceptualization, Resources, Data curation)

Supplementary data

Supplementary data are available at *Biology Methods and Protocols* online.

Conflict of interest statement. The authors declare that there are no obvious or potential conflicts of interest related to the publication of this article, they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Funding

This work was supported by the Ministry of Science and Higher Education of the Russian Federation (the project FZNS-2023-0010 of the State Assignment of the Far Eastern Federal University—FEFU).

Code availability

The full source code for the software implementation of the method is available at the following link: <https://github.com/NikitaKuksin/MultilevelPredictorsCategorizationForAtrialFibrillationPredictionInPatientsWithCoronaryHeartDisease.git>

References

- Gaudino M, Di Franco A, Rong LQ et al. Postoperative atrial fibrillation: from mechanisms to treatment. *Eur Heart J* 2023; **44**:1020–39.
- Xu Z, Qian L, Zhang L et al. Predictive value of NT-proBNP, procalcitonin and CVP in patients with new-onset postoperative atrial fibrillation after cardiac surgery. *Am J Transl Res* 2022; **14**:3481–7.
- Shen J, Lall S, Zheng V et al. The persistent problem of new-onset postoperative atrial fibrillation: a single-institution experience over two decades. *J Thorac Cardiovasc Surg* 2011; **141**:559–70.
- Filardo G, Damiano RJ, Ailawadi G et al. Epidemiology of new-onset atrial fibrillation following coronary artery bypass graft surgery. *Heart* 2018; **104**:985–92.
- Greenberg JW, Lancaster TS, Schuessler RB et al. Postoperative atrial fibrillation following cardiac surgery: a persistent complication. *Eur J Cardiothorac Surg* 2017; **52**:665–72. <https://doi.org/10.1093/ejcts/ezx039>
- Lu Y, Chen Q, Zhang H et al. Machine learning models of postoperative atrial fibrillation prediction after cardiac surgery. *J Cardiothorac Vasc Anesth* 2023; **37**:360–6.
- 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.
- Karri R, Kawai A, Thong YJ et al. Machine learning outperforms existing clinical scoring tools in the prediction of postoperative atrial fibrillation during intensive care unit admission after cardiac surgery. *Heart Lung Circ* 2021; **30**:1929–37.
- Parise O, Parise G, Vaidyanathan A et al. Machine learning to identify patients at risk of developing new-onset atrial fibrillation after coronary artery bypass. *J Cardiovasc Dev Dis* 2023; **10**:82.
- El-Sherbini AH, Shah A, Cheng R et al. Machine learning for predicting postoperative atrial fibrillation after cardiac surgery: a scoping review of current literature. *Am J Cardiol* 2023; **209**:66–75. <https://doi.org/10.1016/j.amjcard.2023.09.079>
- 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**:3708–25.
- Shakhgelyan KI, Kuksin NS, Domzhalov IG et al. Interpretable machine learning for in-hospital mortality risk prediction in patients with ST-elevation myocardial infarction after percutaneous coronary interventions. *Comput Biol Med* 2024; **170**:107953.
- Valente F, Henriques J, Paredes S et al. A new approach for interpretability and reliability in clinical risk prediction: acute coronary syndrome scenario. *Artif Intell Med* 2021; **117**:102113.
- Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: *Advances in Neural Information Processing Systems. Proceedings of the 31st Annual Conference on Neural Information Processing Systems*. Long Beach, USA, 2017.
- Evenson KR, Wen F, Herring AH. Associations of accelerometry-assessed and self-reported physical activity and sedentary behavior with all-cause and cardiovascular mortality among US adults [published correction appears in *Am J Epidemiol*. 2017 Jul 1; 186(1):129]. *Am J Epidemiol* 2016; **184**:621–32.
- Mabikwa OV, Greenwood DC, Baxter PD, Fleming SJ. Assessing the reporting of categorised quantitative variables in observational epidemiological studies. *BMC Health Serv Res* 2017; **17**:307.
- Lin SZ, Crawford TC, Suarez-Pierre A et al. A novel risk score to predict new onset atrial fibrillation in patients undergoing isolated coronary artery bypass grafting. *Heart Surg Forum* 2018; **21**:E489–96.
- Mathew JP, Parks R, Savino JS et al.; MultiCenter Study of Perioperative Ischemia Research Group. Atrial fibrillation following coronary artery bypass graft surgery: predictors, outcomes, and resource utilization. *JAMA* 1996; **276**:300–6.
- Mathew JP, Fontes ML, Tudor IC et al.; Multicenter Study of Perioperative Ischemia Research Group. A multicenter risk index for atrial fibrillation after cardiac surgery. *JAMA* 2004; **291**:1720–9.
- Manning WJ, Gelfand EV. Left atrial size and postoperative atrial fibrillation: the volume of evidence suggests it is time to break an old habit. *J Am Coll Cardiol* 2006; **48**:787–9.
- Zaman AG, Archbold RA, Helft G et al. Atrial fibrillation after coronary artery bypass surgery: a model for preoperative risk stratification. *Circulation* 2000; **101**:1403–8.
- Steinberg JS, Zelenkofske S, Wong SC et al. Value of the P-wave signal-averaged ECG for predicting atrial fibrillation after cardiac surgery. *Circulation* 1993; **88**:2618–22.
- Creswell LL, Schuessler RB, Rosenbloom M, Cox JL. Hazards of postoperative atrial arrhythmias. *Ann Thorac Surg* 1993; **56**:539–49.
- Ying X. An overview of overfitting and its solutions. *J Phys Conf Ser* 2019; **1168**:022022. <https://doi.org/10.1088/1742-6596/1168/2/022022>

© The Author(s) 2025. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Biology Methods and Protocols, 2025, 11, 1–9

<https://doi.org/10.1093/biomethods/bpaf092>

Methods Article