Machine Learning for Predicting Postoperative Atrial Fibrillation After Cardiac Surgery: A Scoping Review of Current Literature

Adham H. El-Sherbini[a], Aryan Shah, BSc[b], Richard Cheng, BHSc[b], Abdelrahman Elsebaie, BHSc[a], Ahmed A. Harby, MSc[c], Damian Redfearn, MD, PhD[d], and Mohammad El-Diasty, MD, PhD[e,*]

Postoperative atrial fibrillation (POAF) occurs in up to 20% to 55% of patients who underwent cardiac surgery. Machine learning (ML) has been increasingly employed in monitoring, screening, and identifying different cardiovascular clinical conditions. It was proposed that ML may be a useful tool for predicting POAF after cardiac surgery. An electronic database search was conducted on Medline, EMBASE, Cochrane, Google Scholar, and *ClinicalTrials.gov* to identify primary studies that investigated the role of ML in predicting POAF after cardiac surgery. A total of 5,955 citations were subjected to title and abstract screening, and ultimately 5 studies were included. The reported incidence of POAF ranged from 21.5% to 37.1%. The studied ML models included: deep learning, decision trees, logistic regression, support vector machines, gradient boosting decision tree, gradient-boosted machine, K-nearest neighbors, neural network, and random forest models. The sensitivity of the reported ML models ranged from 0.22 to 0.91, the specificity from 0.64 to 0.84, and the area under the receiver operating characteristic curve from 0.67 to 0.94. Age, gender, left atrial diameter, glomerular filtration rate, and duration of mechanical ventilation were significant clinical risk factors for POAF. Limited evidence suggest that machine learning models may play a role in predicting atrial fibrillation after cardiac surgery because of their ability to detect different patterns of correlations and the incorporation of several demographic and clinical variables. However, the heterogeneity of the included studies and the lack of external validation are the most important limitations against the routine incorporation of these models in routine practice. Artificial intelligence, cardiac surgery, decision tree, deep learning, gradient-boosted machine, gradient boosting decision tree, k-nearest neighbors, logistic regression, machine learning, neural network, postoperative atrial fibrillation, postoperative complications, random forest, risk scores, scoping review, support vector machine. **(Am J Cardiol 2023;209:66−75)**

Keywords: machine learning, postoperative atrial fibrillation, cardiac surgery, artificial intelligence, postoperative complications, risk scores, deep learning, decision tree, logistic regression, support vector machine, gradient boosting decision tree, gradient-boosted machine, k-nearest neighbors, neural network, random forest, scoping review

Postoperative atrial fibrillation (POAF) is the most common arrhythmic complication after cardiac surgery and can be generally attributed to a combination of inflammatory response, oxidative stress, autonomic imbalance, and structural-functional remodeling of the atrial muscle.[1] Its peak incidence occurs on the second postoperative day with a reported average incidence rate of 35% among all cardiac surgical procedures.[2,3]

In general, risk factors for developing POAF include advanced age, obesity, the presence of co-morbidities (such as chronic obstructive pulmonary disease, diabetes mellitus, and arterial hypertension),[4] the need for intraoperative blood transfusion, surgery for valvular heart disease,[5,6] and the development of postoperative complications (such as stroke, and infections).[6]

It was also reported that POAF may an independent predictor of several adverse outcomes in post-cardiac surgery patients, including renal insufficiency, stroke, cognitive dysfunction, and both short and long-term mortality.[2,7] In addition, POAF can be associated with prolonged hospital length of stay and increased resource utilization.[2] Despite advancements in surgical and anesthetic techniques, the incidence of POAF has not decreased significantly in the last decades, and it is expected to increase given the increasingly aging and poly-morbid patient population.[8] Therefore, identifying

[a]Faculty of Health Sciences, Queen's University, Kingston, Ontario, Canada; [b]School of Medicine, Queen's University, Kingston, Ontario, Canada; [c]The School of Computing, Queen's University, Kingston, Ontario, Canada; [d]Division of Cardiology, Department of Medicine, Queen's University, Kingston, Ontario, Canada; and [e]Division of Cardiac Surgery, University Hospitals Cleveland Medical Center, Cleveland, Ohio, USA. Manuscript received August 11, 2023; revised manuscript received and accepted September 21, 2023.

Funding: none.

See page 74 for Declaration of Competing Interest.

*Corresponding author.

E-mail address: cardiac.science.lab@gmail.com (M. El-Diasty).

0002-9149/© 2023 Elsevier Inc. All rights reserved.
https://doi.org/10.1016/j.amjcard.2023.09.079

patients with a high probability of developing POAF and initiating preventive treatments during the perioperative period is critical for effective management.[8,9]

The mechanisms for developing new-onset AF after cardiac surgery are very complex and multi-factorial.[10] Nevertheless, multiple risk stratification models have been introduced to predict POAF, including the FHS (Framingham Heart Study), the ARIC (Atherosclerosis Risk in Communities Study), and the Cohorts for Heart and Ageing Research in Genomic Epidemiology- Atrial fibrillation (CHARGE-AF) scores.[11−13] These models employ easily obtainable variables, such as age, ethnicity, height, weight, blood pressure, smoking status, antihypertensive medication use, history of diabetes, and heart failure.[14] Moreover, structural cardiac abnormalities, such as atrial fibrosis and atrial enlargement have also been used to predict the risk of developing POAF.[14]

Several healthcare industries have been increasingly applying machine learning (ML) for diagnosis, image interpretation, treatment strategy, and outcome prediction.[7] ML algorithms, a subdiscipline of artificial intelligence, can process complex inputs and identify subtle relations that traditional statistical methods may miss.[15] It is generally divided into 3 main types: supervised, unsupervised and reinforcement learning.[15] During training, the former type of learning requires labels, such as whether a POAF event has occurred or not and as a result, the algorithm is given both the input and output labels. Conversely, unsupervised learning aims to find connections between the data without the aid of labels, and as such, several techniques, including clustering, have been described for this type of learning. The concept of reward maximization is used in the latter form of learning, where the ML algorithm takes on the role of an agent that gets either positive or negative reinforcement to aid it in the decision-making process.[15]

The aim of this scoping review is to consolidate the available literature on the role of different ML models in the prediction of POAF in patients who underwent cardiac surgical procedures.

Methods

Search strategy and study selection

The Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) were followed when conducting this scoping review.[16] Requests for access to the extracted data or the data extraction template may be provided upon contact with the corresponding author. The following databases were electronically searched for primary studies evaluating the use of ML in predicting POAF in patients who underwent cardiac surgical procedures: Medline, EMBASE, Cochrane, *ClinicalTrials.gov*, and Google Scholar (December 30, 2022). A combination of keywords, “machine learning,” “artificial intelligence,” “cardiac surgery,” and “postoperative atrial fibrillation,” were used in the literature search. For consideration, articles had to be written in English, be primary studies including adults at least 18 years old, and present results for an ML model used to predict POAF after cardiac surgery.

Study outcomes

Outcomes of interest included specificity, sensitivity, area under the receiver operating characteristic curve (AUROC), and incidence of POAF.

Screening

Literature search results were uploaded to Covidence review software (Covidence Systematic Review Software, Veritas Health Innovation, Melbourne, Australia. http://www.covidence.org).

Data extraction

The author, publication date, study design, prevalence of POAF, initial study population demographics, and pertinent results were among the data that were extracted. A table of study characteristics was created to extract and compile all the articles. Two independent blinded reviewers assessed all of the studies’ quality using the Newcastle Ottawa Quality Assessment.

Results

Through the literature search, a total of 5,955 citations underwent title and abstract screening by 2 blinded independent reviewers (AES and AS), of which all conflicts were resolved by a third reviewer (RC). Fourteen studies were eligible for a full-text evaluation by 2 blinded independent reviewers (AES and RC), and conflicts were resolved by a third reviewer (AS). From that, 7 studies were ultimately included for data extraction in this scoping review (Figure 1).[7,9,17−19] Five of the studies were retrospective, and 2 were prospective in design, compiling data from a total of 26,703 patients who underwent cardiac surgery.[7,9,17−19] The mean age of the patients across all the studies ranged from 50.4 to 65.8 years, and the proportion of males across all studies ranged from 51.6% to 75.2%.[7,9,17−19] The incidence of POAF across all the studies ranged from 21.5% to 37.1% across the studies.[7,9,17−19]

The studies within this scoping review varied in their inclusion criteria and cardiac surgical procedures (Table 1[7,9,17−21], Figure 2). The most common cardiac procedure included was coronary artery bypass grafting (CABG). Other studies included patients who underwent single valve surgery,[7,9,17,18] multiple valve surgery,[18] aortic surgery,[7,9,17−20] minimally invasive surgery,[17] or a combination of these procedures. The diagnostic modalities used to identify POAF across the studies were also diverse, with 3 studies relying on electrocardiogram (ECG) data,[17,18,21] while the remaining 4 studies utilized clinical documentation, administrative data, or Holter monitoring.[7,9,19,20]

The ML algorithms utilized in the studies included deep learning (DL),[21] decision trees (DT),[9] logistic regression (LR) models,[7,9,19] support vector machines (SVMs),[7,9,18,20] gradient boosting DT (GBDT),[7,17] gradient-boosted machine (GBM),[9] K-nearest neighbors (KNNs),[9] neural network (NN),[20] and random forest (RF) models.[9,20] These ML algorithms used a range of patient characteristics such as age, sex, body mass index, history of hypertension, diabetes, congestive heart failure, glomerular filtration rate,

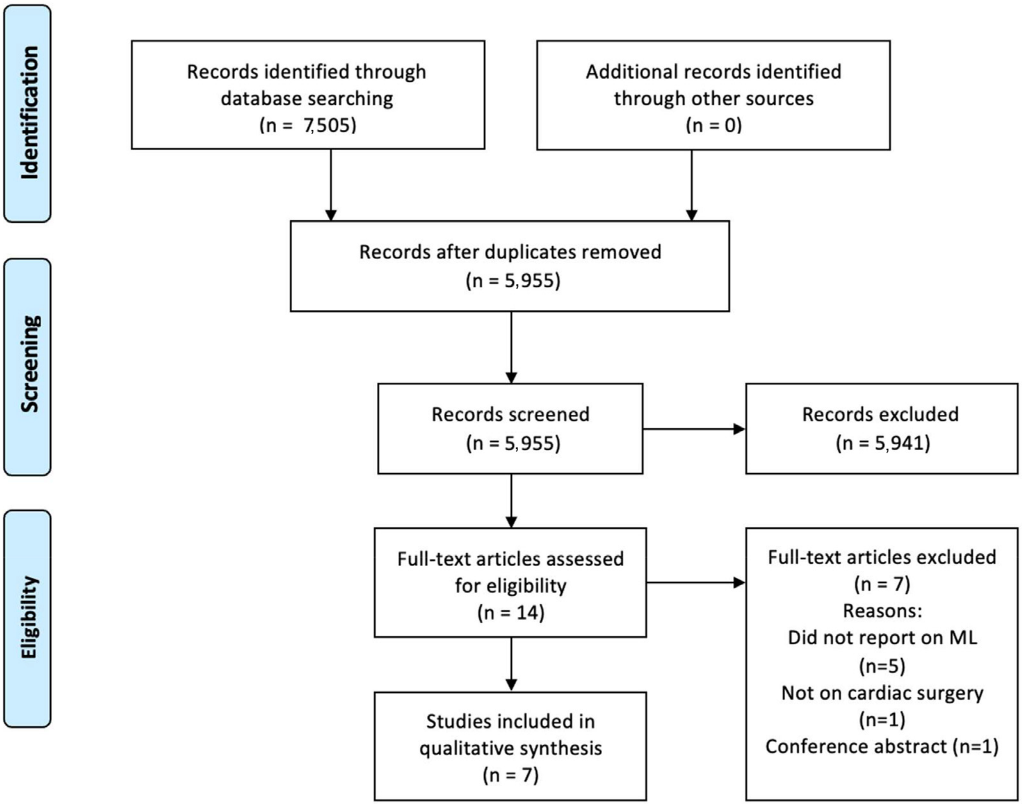

Figure 1. PRISMA diagram.

left atrial diameter, previous myocardial infarction, and chronic obstructive pulmonary disease to predict the probability of a POAF occurrence.[7,9,17−19] Four studies also incorporated intraoperative variables such as cardiopulmonary bypass time and mechanical ventilation time as risk predictors.[7,18−20] Reported performance metrics, such as sensitivity and specificity, varied across the studies and ranged from 0.22 to 0.91 and 0.64 to 0.84, respectively; Table 2.[7,9,17−21] The area under the receiver operating characteristic curve (AUC-ROC) values ranged from 0.67 to 0.94, and the overall accuracy of the ML algorithms ranged from 67% to 72%.[7,9,17−19] Study outcomes are summarized in Table 3.[7,9,17−21]

The model that was produced by Hiraoka et al,[17] was based on using a GBDT ML model to detect POAF after a variety of cardiac surgical procedures using pulse rate data output from an Apple watch with built-in photoplethysmography. In this study, pulse rate was collected from 79 patients who underwent cardiac surgery for 24 hours during hospitalization and continuously after discharge for up to 14 days.[17] Features of pulse data were computed every minute up to 10 min of POAF diagnosis and treated respectively as a single record for training.[17] The median of mean heart rate and SD up to the time of diagnosis were compared with baseline.[17] The GBDT model was trained on 59 patients and tested on 20 patients.[17] Specificity was measured at 0.838, sensitivity was measured at 0.909, and AUROC was measured at 0.942.[17] The authors found the main predictors contributing to AF diagnosis (with GBDT) to be age and baseline changes in HR.[17]

Another model, produced by He et al,[18] used a SVM ML model to predict POAF after cardiac surgery. Long-term single-lead ECG was collected on 94 patients more than 24-H before surgery and more than 7 days after surgery.[18] The model was trained in total on 38 patients and tested on 56 patients with optimal hyperparameters determined by fivefold cross-validation and grid search.[18] Two schemes were adopted such that one included all patient data (scheme A) while another included random patient data sets to ensure data balance as far as possible (scheme B).[18] In scheme A, an accuracy of 0.66, a specificity of 0.74 and a sensitivity of 0.22 were reported[18] while in scheme B, an accuracy of 0.67, a specificity of 0.78 and a sensitivity of 0.56 were reported.[18] Two multivariate prediction models were then adopted: one based solely on clinical patient characteristics: age, sex, left atrial diameter, glomerular filtration rate, and duration of mechanical ventilation (model 1), and another with these characteristics in addition to P-wave ECG data ($P_{max.}$, Pstd, and PWd) (model 2).[18] Models 1 and 2 reported an AUROC of 0.86 and 0.89, respectively, suggesting that the model combining P-wave parameters and clinical data performed better in predicting POAF.[18]

In their study, Tohyama et al[21] reported the use of DL on 12-lead ECG data collected for 30 days before surgery to predict POAF. A total of 27,563 patients were included in this study and the DL model was trained, tuned, and internally validated in a ratio of 7:1:2, respectively.[21] At 7 days postoperatively, the model was found to achieve a sensitivity of 79.9%, specificity of 73.5%, positive predictive value of 10.2%, and negative predictive value of 99.0%.[21]

Table 1
Study Characteristics

Author	Country	Study design	Machine learning model(s)	Incidence of POAF	Type of cardiac surgery	Cross-validation	Diagnostic modality	Training sample size	Validation sample size	Test sample size
He et al. 2022[18]	China	Prospective	support vector machine (Two models: clinical model or clinical + ECG model)	31.00%	single valve surgery (n=41), multiple valve surgery (n=32), CABG (n=7), AVR (n=21)	5-fold	Single-led ECG and clinical data	38	NA	NA
Hiraoka et al. 2022[17]	Japan	Prospective	gradient boosting decision tree	34.20%	off-pump CABG (n=18), valve surgery (n=57), other surgery (n=4), minimally invasive surgery (n=7)	Bayesian Optimization with cross-validation	Wearable device (Apple Watch Series 4)	59	NA	20
Karri et al. 2021[9]	Australia	Retrospective	Random forest classifier, decision tree classifier, logistic regression, K neighbours' classifier, support vector machine, and gradient boosted machine	21.50%	CABG, valvular operation, revision procedures, and indicators of cardiac surgery such as cardioplegia or cardiopulmonary bypass	5-fold	Multivariate Characteristics	80% of 6040 = 4832	NA	20% of 6040 = 1208
Lu et al. 2023[7]	China	Retrospective	Logistic Regression, Gradient Boosting Decision Tree, Support Vector Machine	37.10%	Valve and/or Coronary Artery Bypass Grafting Surgery under Cardiopulmonary Bypass (CPB)	10-fold	Multivariate Characteristics	70% of 1400 = 980	NA	30% of 1400 = 420
Magee et al[19]	United States	Retrospective	Logistic Regression	21.50%	CABG	Hosmer-Lemeshow $\chi 2$	Multivariate Characteristics	NA	NA	NA
Parise et al[20]	Netherlands	Retrospective	multivariate adaptive regression spline, neural network, random forest, support vector machine	10.66%	CABG	10-fold	Multivariate Characteristics	296 patients (75%)	NA	96 patients (25%)
Tohyama et al[21]	Japan	Retrospective	deep learning	3.60%	NA	NA	ECG	30786 ECG (70%)	8796 ECG (20%)	NA

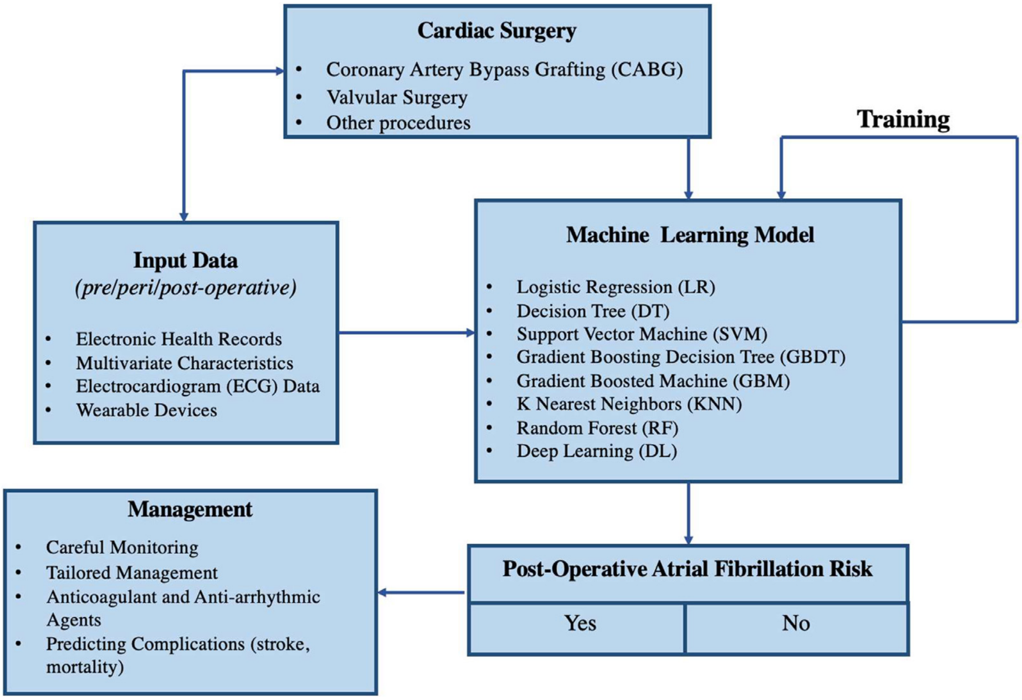

Figure 2. Machine learning process of predicting postoperative atrial fibrillation.

The C-statistic was reported to be 0.83, which can be compared with a reference of 0.63 derived from a single network (without ML) of clinical variables that included age and sex only.[21] The authors noted a high negative predictive value of the model to underscore the potential for preoperative ECG data analyzed with DL to be effective in screening and identifying high-risk patients with POAF who will require appropriate ECG monitoring during the postoperative period.[21]

The remaining 4 studies developed ML models based exclusively on clinical data of multivariate patient characteristics, either preoperatively, postoperatively, or a combination of such factors. For example, Karri et al,[9] generated multiple ML models using a RF classifier, DT classifier, LR, KNN, SVM, and GBM to predict POAF after cardiac surgery. These ML models were trained on 4,832 participants and tested on 1,208 participants, with fivefold cross-validation.[9] Performance was then compared with the established gold standard scoring tool known as the POAF score, which was reported in this study, and consistently in literature, to have a specificity of 0.65, a sensitivity of 0.60 and an AUROC of 0.63.[9] The best model in this study was the GBM, which performed better than the POAF score (specificity 0.64, sensitivity 0.73, AUROC 0.74).[9] Notably, all ML models performed superior (AUROC range from 0.67 to 0.74) to the POAF score except for the DT classifier, which substantially underperformed (AUROC of 0.59).[9] The authors found age, other cardiac procedures, congestive heart failure, and valvular disease to be among the largest patient characteristics predictive for POAF.[9]

In another study, Lu et al,[7] generated an LR, GBDT, and SVM model that was trained on 980 participants and tested on 420 participants with a 10-fold cross-validation. All models performed similarly with a specificity range of 0.72 to 0.79, a sensitivity range of 0.60 to 0.68, and an AUROC of 0.77 to 0.78.[7] As the relative importance of predictors showed by GBDT algorithm, the 6 most influential predictive variables were left atrial diameter postoperative white blood cell count count, advanced age, preoperative platelet count, arrhythmia, and type of surgery.[7]

In their publication, researchers Magee et al[19] produced an LR model from perioperative risk factors for POAF after CABG across 19,083 patients and compared the prediction

Table 2
Study findings

Study	ML Models	Specificity	Sensitivity	AUROC
He et al.[18]	SVM (clinical +/- ECG)	0.74-0.78	0.22-0.56	0.86 − 0.89 (+ ECG)
Hiraoka et al.[17]	GBDT	0.838	0.909	0.942
Karri et al.[9]	RF, DT, LR, KNN, SVM, GBM	0.64-0.83	0.36-0.73	0.67 − 0.74 (GBM)
Lu et al.[7]	LR, GBDT, SVM	0.716-0.795	0.600-0.684	0.77− 0.78
Magee et al.[19]	LR	NA	NA	0.72
Parise et al.[20]	MARS, NN, RF, SVM, DL	0.74-1	0.67-1	0.78− 0.95 (SVM)
Tohyama et al.[21]	DL	0.735	0.799	0.83

Table 3
General study conclusions

Study	General Conclusion	Risk factors for POAF
Tohyama et al. (2023)[21]	DL model is effective in screening for POAF in the post-operative period.	Preoperative ECG changes
Parise et al. (2023)[20]	RF outperformed all other models in clinical prediction of POAF following CABG.	Age, preoperative creatinine values, time of aortic cross-clamping, body surface area (BSA), and Logistic Euro-Score
Lu et al. (2022)[7]	SVM was the best predictor and may be an effective tool for predicting POAF.	Increased left atrium diameter (LAD), postoperative white blood cell count (WBC) and age
He at al. (2022)[18]	Clinical + ECG model and the ML model based on P-wave parameters could predict POAF.	P wave parameters
Hiraoka et al (2022)[17]	Apple Watch could potentially detect AF with a ML classifier during the recovery period after heart surgery.	Age and baseline standard deviation (SD) of heart rate
Magee et al. (2022)[19]	The model demonstrates acceptable accuracy and concordance, good selectivity.	Age, the need for prolonged ventilation (24 hours or more), the use of cardiopulmonary bypass and preoperative arrhythmias.
Karri et al. (2021)[9]	ML can outperform clinical scoring tools.	Age, coagulopathy, valvular disease, valve operations, renal failure, liver disease, other neurological disorders and congestive heart failure

model to the true outcomes of the disease. Prolonged ventilator use, pump status (on/off), and race were factors found to be of significant consideration within the study (odds ratio >1.50).[19] To that end, a weighted variable algorithm consisting of 14 readily obtainable clinical indicators was created and demonstrated to be of strong prediction capability for POAF after CABG (AUROC 0.72).[19]

Finally, Parise et al[20] assessed the performance of ML models based on multivariate characteristics in 394 patients who underwent CABG surgery. Researchers trained and tested 4 separate models (multivariate adaptive regression spline [MARS], NN with 3 hidden layers, RF, and SVM) in a 75:25 participant ratio.[20] When analyzed within a confusion matrix calculated at the threshold value of 0.50, the authors reported the RF model to outperform all others in the clinical prediction of POAF after CABG with observed specificity, sensitivity, and accuracy of 0.81, 0.60, and 0.79, respectively.[20] All other models ranged from (0.60 to 0.70), (0.57 to 0.70), and (0.58 to 0.69) within the same respective categories.[20] Furthermore, the MARS, NN, RF, and SVM, reported a maximal AUROC of 0.87, 0.94, 0.78, and 0.95, respectively.[20] As noted, no single model achieved maximal ROC, sensitivity, and specificity together. Among all multivariate characteristics, the RF model demonstrated that age, preoperative creatinine values, time of aortic cross-clamping, and body surface area, to be among the greatest predictive factors (normalized contribution to model greater than 40%).[20]

Discussion

Based on the available literature, ML models can potentially predict POAF after cardiac surgery with promising specificity, sensitivity, and AUROC scores. The mechanics of ML require several crucial elements to operate fluidly together. To effectively train an ML model, it is imperative to provide sufficient and reliable data. Various forms of monitoring, such as ECG, Holter monitoring, and wearable devices, can also provide a source of very valuable data that can be input into these models.[22] In order for ML algorithms to function appropriately, it is imperative that the data be preprocessed before it is utilized in the algorithms.[23] This entails determining and eliminating outliers, dealing with missing data, and scaling the data to guarantee that all variables are considered equally.[23] Numerous algorithms are available, including supervised learning algorithms (SVM, DT, or LR) and unsupervised learning techniques such as clustering and dimensionality reduction to be trained on the chosen diagnostic modality. The basic principles, advantages, and limitations of each of these models are summarized in (Table 4[24–35]).[36]

The algorithm depends on the diagnostic modalities' respective data and the purpose of the model.[37] While SVM works by identifying the best hyperplane that separates different classes of data points,[38] other models such as GBDT, RF, DT, KNN, and GBM are ensemble algorithms that combine multiple weak models to create a stronger prediction model.[39] Also, MARS, NN, and DL are nonparametric regression algorithms that can capture nonlinear relations between predictor variables and outcomes.[39] The algorithm receives input data and the related output during the training.[37] After learning from the input-output pairings, the algorithm modifies its internal parameters to minimize the discrepancy between the predicted and actual output.[37] The algorithm is refined this way until it achieves acceptable accuracy when it is prepared for testing. The test sets findings are utilized to fine-tune the model and provide improvements where they are required.[40]

In the study by Hiraoka et al,[17] the AUROC curve shows that their algorithm had a diagnostic accuracy of 0.9416 (Sensitivity 0.909 and Specificity 0.838 at the point closest to the top left). Other publications showed that wearable technology has good diagnostic accuracy for non-surgical AF with an area under the ROC curve of ≥0.9.[41–43] These, however, are predicated on data sampling with a preselected group of patients who have a history of AF and in a constrained ideal setting (e.g., at rest or when data collection duration is only a few hours).

Table 4
Benefits and Setbacks of ML models

ML Algorithm Model	Description of Model	Advantages	Limitations
Logistic regression (LR)[24,25]	• Uses flow properties, such as continuous, discrete, or hybrid, prior to linearly combining the inputs by passing them through a logistic function. • Most widely used supervised ML algorithm as it produces more reliable results in large datasets.	• Associated with low variance. • Offers probability for output. • Simple to use, and training takes little time.	• May be modified to multiclass classification using a variety of applications but cannot naturally categorize multi-class data. • Does not function well in the presence of interconnected attributes • In situations where there are too few observations compared to features, logistic regression may result in overfitting. • Regression models need relevant and biologically important independent predictor variables to be valid. • Collinearity can cause errors or uncertainty in estimating effects. • Variables must have consistent association magnitudes, and interactions between predictors must be considered for valid estimates.
K-nearest neighbor (KNN)[26]	• Simplest of the supervised ML algorithms • To apply the algorithm, attribute vectors must be constructed. • Next number of neighbourhoods is specified by the k parameter, which also determines the nearest neighbor that each data input should be assigned to.	• Simple to implement and comprehend since it is free of assumptions. • Heuristic in nature. • Swiftly adapts to input changes while being used in real-time. • Could be easily utilized to solve multiclass classification problems.	• The pace of the algorithm slows down significantly as the amount of data increases. • As the number of variables rises, it gets more challenging to obtain the desired output. • It is impacted by outliers and is unable to handle missing numbers. • The variable characteristics must all be stated in the same scale for the system to function effectively.
Naïve Bayes (NB)[27,28]	• Implements the conditional independence rule, which stipulates that all properties are independent variables and thus changes in each variable would do not impact others. • Useful supervised ML algorithm when classifying larger data sets.	• Very reliable findings are produced using this relatively quick and adaptable model. • Highly suited for larger data sets. • Training doesn't require much of your time. • Reduces unnecessary specifications to improve grading performance.	• To attain favourable results, large data records are required. • Compares poorly to the other classifiers in terms of performance depending on the type of problem. • Conditional Independence assumption may not always work due to feature dependence, causing issues like zero-probability and suboptimal binning with Multinomial Naive Bayes. • Imbalanced data may not be handled well with Complement Naive Bayes.
Support vector machine (SVM)[29,30]	• A supervised ML algorithm that is based on statistical learning theory. • Used prominently for the classification of binary, multiclass or non-linear data. • The foundation of SVM is the prediction of the decision function that can discriminate between classes.	• Produces reliable findings despite the lack of sufficient data. • Particularly effective with unstructured data. • Uses a handy kernel solution function to solve difficult problems. • Is comparatively excellent at high dimensional data scaling.	• Choosing the right kernel solution function is often challenging. • When working with extensive data sets, training takes a while. • The model could be challenging to perceive and comprehend due to issues brought on by individual circumstances and varied weights. • The contribution of each variable to the outcome varies because the variable weights are not constant. • SVM struggles with complex classification problems.
Random forest (RF)[31,32]	• Simple supervised ML algorithm utilized for classification of data and generation of decision trees since it is not vulnerable to overfitting.	• Because of the relationship between training and testing data, the likelihood of encountering a classifier that does not perform well is decreased.	• Visually challenging to comprehend and interpret. • Significantly more time-consuming and difficult to build than decision trees. • Computation-intensive and the algorithm itself is less heuristic.

(continued)

Table 4 (***Continued***)

ML Algorithm Model	Description of Model	Advantages	Limitations
	• When training this ML algorithm, it produces many decision trees from the subset of the problem and estimates each tree. • The most voted estimations are used for classification.	• Extremely flexible with very high accuracy that is present even when a large proportion of the data are missing. • Versatility in solving regression and classification problems. • Handling both categorical and continuous variables • Automatic addressing of missing values, no feature scaling requirement • Efficient handling of non-linear parameters • Robust to outliers and noise • Stable in handling new data without significant impact on accuracy.	
Extreme gradient boosting (XGBoost)[33]	• A scalable form of gradient boosting which combines outputs from trees to generate predictions. • By building more trees, this supervised ML algorithm reduces the errors of the prior trees and as such the model increases in reliability as trees are added.	• When the data is clean, it can avoid overfitting. • Can deal with missing values. • Enables for cross-validation at each iteration of the process, maximizing the number of iterations.	• More challenging to comprehend compared to other linear algorithms. • Data that is noisy could overfit.
Light gradient boosting machine (LightGBM)[34,35]	• Another type of supervised, decision tree-based ML algorithm that effectively implements aspects of the gradient boosting framework. • Main difference is that it grows trees vertically instead of horizontally as other ML models would.	• Utilizes an optimization technique called histogram-based split finding, which accelerates the training process by dividing continuous feature values into discrete bins. • Reduces memory use by switching continuous data to discrete bins. • By using a leaf-wise tree growth strategy instead of a level-wise split approach (the primary element in getting greater accuracy), it creates far more complicated trees. • With a considerable reduction in training time, it can perform just as well with larger datasets as other gradient boosting models	• Since it employs leaf-wise tree development to build deeper, more intricate trees, it can be more challenging to analyze and comprehend. • Prone to overfitting, especially in small data sets

Several artificial intelligence models are routinely used in cardiac surgery to enhance patient outcomes and optimize surgical techniques. Convolutional Neural Networks (CNNs) are oftentimes used to evaluate diagnostic data to detect patients at the highest risk of cardiovascular disease or related risk factors.[44] Similarly, recurrent NNs are frequently used to assess physiological data, such as ECGs, to detect aberrant cardiac rhythms.[45] DL, a subset of ML, can also to forecast results and identify patients vulnerable to postoperative complications.[46,47]

For POAF detection, both conventional and DL ML techniques have been employed.[48] In studies comparing ML and traditional risk scores for predicting POAF after cardiac surgery, ML has been demonstrated to have several advantages. First, conventional risk measures, such as CHA2DS2-VASc score, are based on a few clinical criteria and do not account for all the critical risk factors for POAF.[49] ML algorithms, on the other hand, can examine a wide range of factors, such as patient characteristics, medical history, and laboratory data, to find patterns and associations that might be predictive of POAF.[20] Second, traditional risk assessments need more flexibility, as they may be unrevised after the release of new information.[49] Third, ML algorithms may be able to handle missing data more effectively because they do not rely on data distribution assumptions and can perform more sophisticated computations. In addition to assisting in processing imaging or electrocardiographic data, ML algorithms may also incorporate and understand vast amounts of clinical data and isolate novel clinical patterns and concepts.[14,50] Finally, it is also important to understand that clinical risk assessment models for POAF assume that each of the risk factors identified by LR has a linear relation with the dependent variable, POAF in this case.[9] However, because of the sensitivity of multicollinear independent variables in the model, LR is unsuitable for analyzing various variables, especially if the correlations between them are not linear in nature.[7]

Future Directions of Research

There are several domains that future studies should focus to improve the performance of ML models in predicting POAF. First, data from electronic health records should be integrated in different ML algorithms to collect pertinent clinical data for the prediction of POAF and for creating quicker and more accurate management strategies. Second, real-time monitoring of patients and the processing of continuous ECG data, collected using wearable or implantable devices, can be effective in improving the diagnostic capability of ML models. Third, in addition to identifying individuals who are more likely to develop POAF, future ML models may be implemented to personalize therapeutic options, such as the use of antiarrhythmic and anticoagulant therapies, and the detection or forecasting of POAF-related complications such as stroke or mortality.

Limitations

The current review has several limitations. First, several studies had a relatively small training and testing sample size with lack of external validation. Second, there were several methodological inconsistencies in some of the included studies such as the limited-quality ECG signals or the lack of continuous ECG monitoring which could potentially have resulted in underestimation of the reported incidence of POAF.[7,18] Third, there was a degree of heterogeneity in terms of defining POAF and the reported period of patient monitoring. Finally, the multi-factorial nature of POAF and its underlying complex mechanism may make it challenging to create a "perfect" predictive ML model with reliable performance that can be incorporated in routine practice.

Conclusions

Our findings suggest that ML may play a role in predicting the development of atrial fibrillation after cardiac surgery. ML modes may offer an advantage over conventional risk scores because of their ability to analyze different correlations and their potential for incorporating several demographic and clinical variables. Future ML models may also be useful to predict complications related to POAF and tailor management strategies to each patient's clinical data and risk profile. However, the methodologic heterogeneity among the included studies and the lack of external validation mandate the need for future studies with adequate sample size that can be robust to overfitting.

Declaration of Competing Interest

The authors have no competing interests to declare.

Data Availability

All data is available upon request to the corresponding author.

1. Rezaei Y, Peighambari MM, Naghshbandi S, Samiei N, Ghavidel AA, Dehghani MR, Haghjoo M, Hosseini S. Postoperative atrial fibrillation following cardiac surgery: from pathogenesis to potential therapies. *Am J Cardiovasc Drugs* 2020;20:19–49.
2. Greenberg JW, Lancaster TS, Schuessler RB, Melby SJ. Postoperative atrial fibrillation following cardiac surgery: a persistent complication. *Eur J Cardiothorac Surg* 2017;52:665–672.
3. Yang H, Yuan C, Yang J, Xiang H, Lan W, Tang Y. A novel predictive model for new-onset atrial fibrillation in patients after isolated cardiac valve surgery. *Front Cardiovasc Med* 2022;9:949259.
4. Seo EJ, Hong J, Lee HJ, Son YJ. Perioperative risk factors for new-onset postoperative atrial fibrillation after coronary artery bypass grafting: a systematic review. *BMC Cardiovasc Disord* 2021;21:418.
5. Ishibashi H, Wakejima R, Asakawa A, Baba S, Nakashima Y, Seto K, Kobayashi M, Okubo K. Postoperative atrial fibrillation in lung cancer lobectomy-analysis of risk factors and prognosis. *World J Surg* 2020;44:3952–3959.
6. Auer J, Weber T, Berent R, Ng CK, Lamm G, Eber B. Risk factors of postoperative atrial fibrillation after cardiac surgery. *J Card Surg* 2005;20:425–431.
7. Lu Y, Chen Q, Zhang H, Huang M, Yao Y, Ming Y, Yan M, Yu Y, Yu L. Machine learning models of postoperative atrial fibrillation prediction after cardiac surgery. *J Cardiothorac Vasc Anesth* 2023;37:360–366.
8. Filardo G, Damiano RJ, Ailawadi G, Thourani VH, Pollock BD, Sass DM, Phan TK, Nguyen H, da Graca B. Epidemiology of new-onset atrial fibrillation following coronary artery bypass graft surgery. *Heart (Br Card Soc)* 2018;104:985–992.
9. Karri R, Kawai A, Thong YJ, Ramson DM, Perry LA, Segal R, Smith JA, Penny-Dimri JC. 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;30:1929–1937.
10. Gaudino M, Di Franco A, Rong LQ, Piccini J, Mack M. Postoperative atrial fibrillation: from mechanisms to treatment. *Eur Heart J* 2023;44:1020–1039.
11. Schnabel RB, Sullivan LM, Levy D, Pencina MJ, Massaro JM, D'Agostino RB, Newton-Cheh C, Yamamoto JF, Magnani JW, Tadros TM, Kannel WB, Wang TJ, Ellinor PT, Wolf PA, Vasan RS, Benjamin EJ. Development of a risk score for atrial fibrillation (Framingham Heart Study): a community-based cohort study. *Lancet* 2009;373:739–745.
12. Chamberlain AM, Agarwal SK, Folsom AR, Soliman EZ, Chambless LE, Crow R, Ambrose M, Alonso A. A clinical risk score for atrial fibrillation in a biracial prospective cohort (from the Atherosclerosis Risk in Communities [ARIC] study). *Am J Cardiol* 2011;107:85–91.
13. Alonso A, Krijthe BP, Aspelund T, Stepas KA, Pencina MJ, Moser CB, Sinner MF, Sotoodehnia N, Fontes JD, Janssens AC, Kronmal RA, Magnani JW, Witteman JC, Chamberlain AM, Lubitz SA, Schnabel RB, Agarwal SK, McManus DD, Ellinor PT, Larson MG, Burke GL, Launer LJ, Hofman A, Levy D, Gottdiener JS, Kääb S, Couper D, Harris TB, Soliman EZ, Stricker BH, Gudnason V, Heckbert SR, Benjamin EJ. Simple risk model predicts incidence of atrial fibrillation in a racially and geographically diverse population: the CHARGE-AF consortium. *J Am Heart Assoc* 2013;2:e000102.
14. Tseng AS, Noseworthy PA. Prediction of atrial fibrillation using machine learning: a review. *Front Physiol* 2021;12:752317.
15. Jovel J, Greiner R. An introduction to machine learning approaches for biomedical research. *Front Med (Lausanne)* 2021;8:771607.
16. Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, Shamseer L, Tetzlaff JM, Akl EA, Brennan SE, Chou R, Glanville J, Grimshaw JM, Hróbjartsson A, Lalu MM, Li T, Loder EW, Mayo-Wilson E, McDonald S, McGuinness LA, Stewart LA, Thomas J, Tricco AC, Welch VA, Whiting P, Moher D. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ* 2021;372:n71.
17. Hiraoka D, Inui T, Kawakami E, Oya M, Tsuji A, Honma K, Kawasaki Y, Ozawa Y, Shiko Y, Ueda H, Kohno H, Matsuura K, Watanabe M, Yakita Y, Matsumiya G. Diagnosis of atrial fibrillation using machine learning with wearable devices after cardiac surgery: algorithm development study. *JMIR Form Res* 2022;6:e35396.
18. He K, Liang W, Liu S, Bian L, Xu Y, Luo C, Li Y, Yue H, Yang C, Wu Z. Long-term single-lead electrocardiogram monitoring to detect new-onset postoperative atrial fibrillation in patients after cardiac surgery. *Front Cardiovasc Med* 2022;9:1001883.

19. Magee MJ, Herbert MA, Dewey TM, Edgerton JR, Ryan WH, Prince S, Mack MJ. Atrial fibrillation after coronary artery bypass grafting surgery: development of a predictive risk algorithm. *Ann Thorac Surg* 2007;83(5):1707–1712:discussion 1712.
20. Parise O, Parise G, Vaidyanathan A, Occhipinti M, Gharaviri A, Tetta C, Bidar E, Maesen B, Maessen JG, La Meir M, Gelsomino S. Machine learning to identify patients at risk of developing new-onset atrial fibrillation after coronary artery bypass. *J Cardiovasc Dev Dis* 2023;10:82.
21. Tohyama T, Ide T, Ikeda M, Nagata T, Tagawa K, Hirose M, Funakoshi K, Sakamoto K, Kishimoto J, Todaka K, Nakashima N, Tsutsui H. Deep learning of ECG for the prediction of postoperative atrial fibrillation. *Circ Arrhythm Electrophysiol* 2023;16:e011579.
22. Bayoumy K, Gaber M, Elshafeey A, Mhaimeed O, Dineen EH, Marvel FA, Martin SS, Muse ED, Turakhia MP, Tarakji KG, Elshazly MB. Smart wearable devices in cardiovascular care: where we are and how to move forward. *Nat Rev Cardiol* 2021;18:581–599.
23. Wei J, Chu X, Sun X, Xu K, Deng H, Chen J, Wei Z, Lei M. Machine learning in materials science. *InfoMat* 2019;1:338–358.
24. Tolles J, Meurer WJ. Logistic regression: relating patient characteristics to outcomes. *JAMA* 2016;316:533–534.
25. Tu JV. Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes. *J Clin Epidemiol* 1996;49:1225–1231.
26. Cunningham P, Delany SJ. k-Nearest Neighbour Classifiers - A Tutorial. *ACM Comput Surv* 2022;54:1–25.
27. Wickramasinghe I, Kalutarage H. Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation. *Soft Comput* 2021;25:2277–2293.
28. Chen H, Hu S, Hua R, Zhao X. Improved naive Bayes classification algorithm for traffic risk management. *EURASIP J Adv Signal Process* 2021;2021:30.
29. Auria L, Moro RA. Support vector machines (SVM) as a technique for solvency analysis. SSRN. Available at: http://www.ssrn.com/abstract=1424949. Accessed on May 19, 2023.
30. Pisner DA, Schnyer DM. Support vector machine. In: Mechelli A, Vieira S, eds. Machine Learning: Mathods and Applications to Brain Disorders. Amsterdam: Elsevier; 2020:101–121.
31. Statnikov A, Wang L, Aliferis CF. A comprehensive comparison of random forests and support vector machines for microarray-based cancer classification. *BMC Bioinformatics* 2008;9:319.
32. Fawagreh K, Gaber MM, Elyan E. Random forests: from early developments to recent advancements. *Syst Sci Control Eng* 2014;2:602–609.
33. Nielsen D. Tree boosting with XGBoost - why does XGBoost win “every” machine learning competition? Available at: https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2433761. Accessed on May 19, 2023.
34. Rufo DD, Debelee TG, Ibenthal A, Negera WG. Diagnosis of diabetes mellitus using gradient boosting machine (LightGBM). *Diagnostics (Basel)* 2021;11:1714.
35. Fan J, Ma X, Wu L, Zhang F, Yu X, Zeng W. Light Gradient Boosting Machine: an efficient soft computing model for estimating daily reference evapotranspiration with local and external meteorological data. *Agric Water Manag* 2019;225:105758.
36. Bkassiny M, Li Y, Jayaweera SK. A survey on machine-learning techniques in cognitive radios. *IEEE Commun Surv Tutorials* 2013;15:1136–1159.
37. El Naqa I, Murphy MJ. What is machine learning? In: El Naqa I, Li R, Murphy MJ, eds. Machine Learning in Radiation Oncology. Cham: Springer International Publishing; 2015:3–11.
38. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol* 2019;19:64.
39. Che D, Liu Q, Rasheed K, Tao X. Decision tree and ensemble learning algorithms with their applications in bioinformatics. In: Arabnia HR, Tran QN, eds. Software Tools and Algorithms for Biological Systems. New York: Springer; 2011:191–199.
40. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. *Science* 2015;349:255–260.
41. Tison GH, Sanchez JM, Ballinger B, Singh A, Olgin JE, Pletcher MJ, Vittinghoff E, Lee ES, Fan SM, Gladstone RA, Mikell C, Sohoni N, Hsieh J, Marcus GM. Passive detection of atrial fibrillation using a commercially available Smartwatch. *JAMA Cardiol* 2018;3:409–416.
42. Aschbacher K, Yilmaz D, Kerem Y, Crawford S, Benaron D, Liu J, Eaton M, Tison GH, Olgin JE, Li Y, Marcus GM. Atrial fibrillation detection from raw photoplethysmography waveforms: A deep learning application. *Heart Rhythm* 2020;1:3–9.
43. Harju J, Tarniceriu A, Parak J, Vehkaoja A, Yli-Hankala A, Korhonen I. Monitoring of heart rate and inter-beat intervals with wrist plethysmography in patients with atrial fibrillation. *Physiol Meas* 2018;39:065007.
44. Shirakawa T, Koyama Y, Shibata R, Fukui S, Tatsuoka M, Yoshitatsu M, Toda K, Fukuda I, Sawa Y. Automated heart segmentation using a convolutional neural network accelerates 3D model creation for cardiac surgery. *Eur Heart J Cardiovasc Imaging* 2021;22(suppl 1).jeaa356.353.
45. Chenga L, Tavakolia M. Neural-network-based heart motion prediction for ultrasound-guided beating-heart surgery. IEEE. Available at: https://ieeexplore.ieee.org/document/8843136/. Accessed on March 27, 2023.
46. Raghu VK, Moonsamy P, Sundt TM, Ong CS, Singh S, Cheng A, Hou M, Denning L, Gleason TG, Aguirre AD, Lu MT. Deep learning to predict mortality after cardiothoracic surgery using preoperative chest radiographs. *Ann Thorac Surg* 2023;115:257–264.
47. Allou N, Allyn J, Provenchere S, Delmas B, Braunberger E, Oliver M, De Brux JL, Ferdynus C, EpiCard investigators. Clinical utility of a deep-learning mortality prediction model for cardiac surgery decision making [published online February 2, 2023]. *J Thorac Cardiovasc Surg* doi:10.1016/j.jtcvs.2023.01.022.
48. Denysyuk HV, Pinto RJ, Silva PM, Duarte RP, Marinho FA, Pimenta L, Gouveia AJ, Gonçalves NJ, Coelho PJ, Zdravevski E, Lameski P, Leithardt V, Garcia NM, Pires IM. Algorithms for automated diagnosis of cardiovascular diseases based on ECG data: a comprehensive systematic review. *Heliyon* 2023;9:e13601.
49. Chua SK, Shyu KG, Lu MJ, Lien LM, Lin CH, Chao HH, Lo HM. Clinical utility of CHADS2 and CHA2DS2-VASc scoring systems for predicting postoperative atrial fibrillation after cardiac surgery. *J Thorac Cardiovasc Surg* 2013;146:919–926.e1.
50. Yu Y, Peng C, Zhang Z, Shen K, Zhang Y, Xiao J, Xi W, Wang P, Rao J, Jin Z, Wang Z. Machine learning methods for predicting long-term mortality in patients after cardiac surgery. *Front Cardiovasc Med* 2022;9:831390.