


SYSTEMATIC REVIEW

Open Access



# Machine learning for the prediction of acute kidney injury post cardiac surgery: a systematic review and meta-analysis

Raja Ahsan Aftab<sup>1,2\*</sup> , Zirwa Asim Butt<sup>1</sup>, Baharudin Ibrahim<sup>1</sup> and Lim Soo Kun<sup>3</sup>

## Abstract

**Background and objectives** Cardiac surgery associated acute kidney injury can lead to increased morbidity, mortality, and hospitalization. The available risk assessment tools have limited predictive ability. Machine learning has been increasingly utilized to predict acute kidney injury in cardiac surgery patients in recent times due to its ability to handle complex clinical data. However, its predictive value remains uncertain. This study evaluates the predictive performance of machine learning models for acute kidney injury post-cardiac surgery.

**Methods** A systematic review and meta-analysis was conducted by searching Web of Science, PubMed, Science Direct, Google Scholar, Scopus, and Cochrane Library up to 31st December 2025. PRISMA guidelines were followed. Included studies were assessed for machine learning model performance and acute kidney injury predictors, with effect measures including area under the receiver operator characteristic curve (AUC), sensitivity, and specificity. Pooled estimates were calculated using a random-effects model with 95% confidence intervals. Risk of bias was assessed using PROBAST. The meta-analysis in our study was performed using R version 4.5.0.

**Results** The systematic search yielded 45 studies that met our inclusion criteria, encompassing 13 distinct model types, which include 81 models for training and 162 for validation. The overall pooled AUC was 0.83 (95% CI: 0.79–0.85) in the training and 0.76 (95% CI: 0.75–0.78) in the validation cohorts. Pooled sensitivity and specificity in the training dataset were 0.75 (95% CI: 0.71–0.79) and 0.81 (95% CI: 0.72–0.87), respectively. In the validation dataset, pooled sensitivity was 0.61 (95% CI: 0.53–0.69), while specificity was 0.82 (95% CI: 0.77–0.86). Analysis showed an overall 44.4% high risk of bias, particularly due to the analysis domain of PROBAST.

**Conclusion** This study suggests that machine learning based models could potentially serve as a viable framework for predicting the risk of post-cardiac surgery AKI, however, highlighting the need for model optimization and validation in a diverse population before clinical implementation.

**Trial registration** The study was registered with PROSPERO (CRD42024576556).

**Keywords** Acute kidney injury, Cardiac surgery, Machine learning, Systematic review, Meta analysis, Prediction

\*Correspondence:

Raja Ahsan Aftab  
ahsan1025@yahoo.com

<sup>1</sup>Department of Clinical Pharmacy and Pharmacy Practice, Faculty of Pharmacy, Universiti Malaya, Kuala Lumpur 50603, Malaysia

<sup>2</sup>School of Health, Science and Society, University of Suffolk, Ipswich, United Kingdom

<sup>3</sup>Department of Medicine, Faculty of Medicine, Universiti Malaya, Kuala Lumpur 50603, Malaysia



© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Introduction

Acute Kidney Injury (AKI) following cardiac surgery presents as a serious complication with poor prognosis and increased risk of mortality [1]. It is associated with a multifaceted array of exposures like cardiopulmonary bypass (CPB), tissue injury, cardiac insufficiency, and hemolysis, etc. The incidence rate of AKI associated with cardiac surgery ranges between 7% and 40% [2]. The pathophysiology of cardiac surgery-associated AKI is multifactorial and remains incompletely understood. Hypoperfusion, ischemia-reperfusion injury, neurohormonal activation, inflammation, nephrotoxin exposure, and CPB related non-pulsatile perfusion are known to be contributing factors [3]. Early prediction and interventions can be pivotal in mitigating adverse outcomes.

The clinical management of post-cardiac surgery AKI remains reactive rather than preventive. Standard Kidney Disease Improving Global Outcomes (KDIGO) criteria rely on serum creatinine, which is a delayed kidney marker, often rising only when significant loss of glomerular filtration occurs. While electronic alert systems have been implemented to improve recognition, a recent meta-analysis suggests that they have failed to consistently improve patient survival [4]. This highlights the critical need for proactive risk stratification of AKI. Recent literature indicates that early interventions using AKI care bundles can enhance kidney outcomes, but their effectiveness relies on identifying high-risk patients during a critical 'golden window' prior to clinical injury [5]. Hence, automated risk stratification along with targeted biomarker assessment can lead to reduced morbidity and need for continuous renal replacement therapy [6].

Machine learning (ML) offers a promising solution for the early risk stratification of AKI by using multidimensional data, surpassing conventional methods. Recent advancements in ML have equipped clinical settings with effective tools for improving AKI risk prediction. These ML algorithms enhance traditional predictive models by offering greater accuracy and the ability to handle complex, multi-dimensional datasets [7]. These predictive models can encompass a range of factors, both preoperative, intraoperative, and postoperative, to anticipate the likelihood of AKI with greater accuracy [8].

Several ML models, including random forest (RF), extreme gradient boosting (XGBoost), logistic regression (LR), artificial neural networks (ANN), and support vector machine (SVM), have been utilized to predict cardiac surgery-associated AKI and have demonstrated better accuracy in risk prediction in comparison to the traditional statistical measures, making them particularly suitable for clinical applications [9]. Understanding various ML algorithms for the prediction of AKI and their comparative efficacy is essential for risk stratification,

informed decision-making, and successful implementation of interventions to prevent AKI after cardiac surgeries.

Despite the development of numerous ML models for predicting AKI, their predictive accuracy remains inconsistent. Therefore, this review aims to comprehensively evaluate the predictive capabilities of a wide range of models, particularly in the context of AKI following cardiac surgery. Understanding these variations is crucial for improving model reliability and clinical application.

## Methods

### Overview

All the articles related to the prediction of cardiac surgery-associated AKI were systematically identified, which were published in online databases as scientific literature. This systematic review and meta-analysis strictly adhered Preferred Reporting Items for Systematic Review and Meta-analysis guidelines (PRISMA). Moreover, the checklist adhering to PRISMA guidelines is provided in the *Supplementary file*. The protocol for this review has also been registered with PROSPERO [CRD42024576556].

### Search strategy

PubMed, Google Scholar, Scopus, Web of Science, Science direct and Cochrane database were searched from inception to 31st Dec, 2025. The search utilized a combination of medical subject headings (MeSH) and text/keywords related to *Acute Kidney Injury, Cardiac Surgery, and Machine learning*. Search queries with full synonym search terms are given in the *Supplementary file Table S1*.

### Study eligibility criteria

Only studies published in English were included, specifically quantitative studies such as randomized controlled trials, case-control studies, and cohort studies. The focus was on research involving ML models designed to predict AKI after cardiac surgery. Studies reporting the performance metrics of ML models were also included. Reviews, abstracts, letters, scientific correspondence, posters, animal studies, case reports, advertisements, theses, opinions, and editorials were excluded due to insufficient methodological data and performance metrics required for robust risk of bias assessment and quantitative meta-analysis. A detailed inclusion and exclusion criteria is given in the *Supplementary file*.

### Literature search

The search strategy was developed following the study protocol, targeting articles that met the predefined inclusion criteria. After retrieving the articles, duplicate entries were removed using EndNote 21. A preliminary

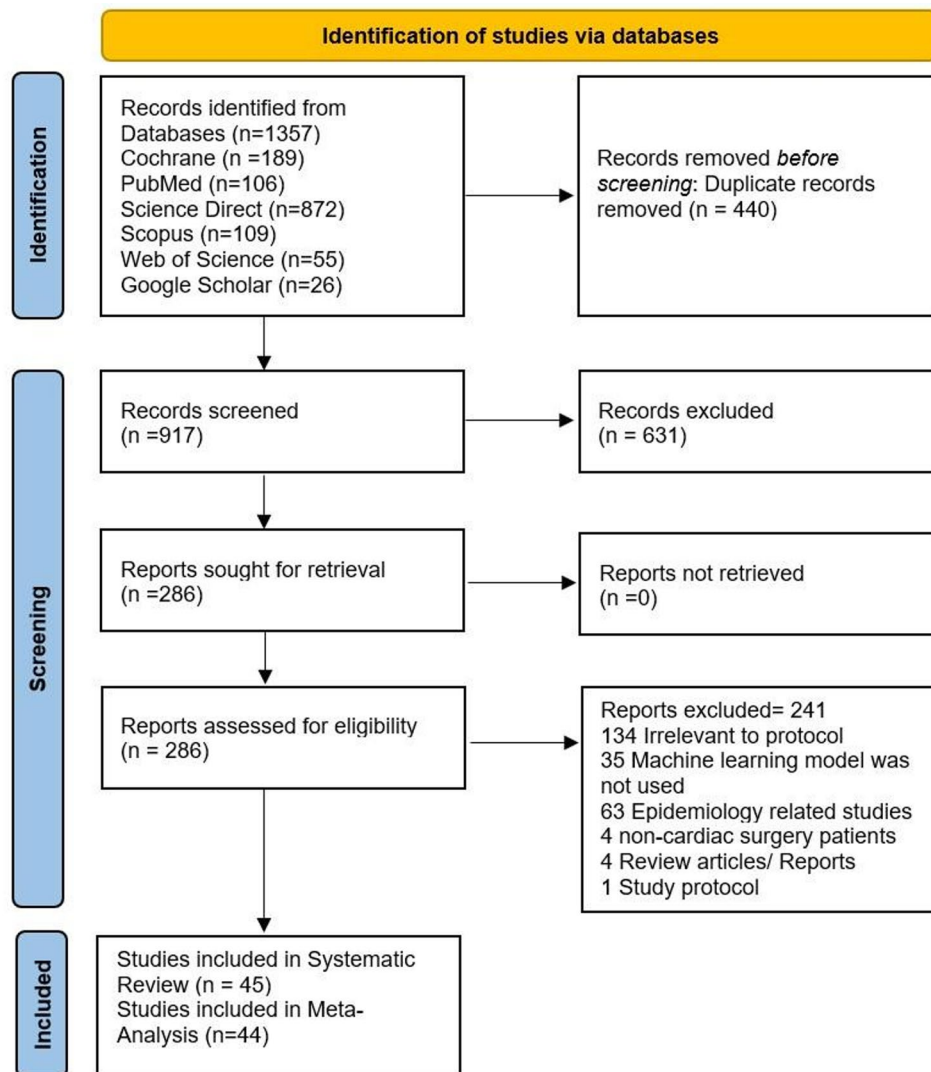
screening of titles and abstracts was conducted to filter studies related to the protocol of our study. Full-text articles of potentially relevant studies were retrieved and carefully reviewed to determine eligibility by two authors independently. The articles for inclusion and exclusion, with any discrepancies, were resolved through discussion among other authors. Articles lacking critical information necessary for the analysis were excluded from the review. The process of article selection, including the number of articles at each stage of the review, is illustrated in the PRISMA flow chart shown in Fig. 1.

### Data extraction

Data extraction focused on gathering key information from each included study based on a predefined data extraction form (Supplementary file Table S2). Two authors independently extracted data from the included

studies. This process involved collecting details such as the first author, publication year, study design, sources of patient data, AKI diagnostic criteria, total number of AKI cases, and overall sample size. Additionally, we recorded the specific ML models used, the number of AKI cases, the sample size in both the training and validation datasets, methods applied to prevent overfitting, strategies for handling missing data, and approaches for variable selection and feature screening.

Performance metrics such as the Area Under the Curve (AUC), sensitivity, and specificity were also documented. The definition of AKI followed the criteria used in each study, such as KDIGO [10] or AKIN [11]. When studies did not report necessary data, we calculated it manually using available information. Studies lacking essential data were excluded from the meta-analysis but were included in the systematic review. Studies which were considered



**Fig. 1** PRISMA flow diagram of the study selection process. Flow diagram illustrating the identification, screening, eligibility assessment, and inclusion of studies in the systematic review and meta-analysis in accordance with the PRISMA guidelines

for inclusion but were excluded for protocol deviations are listed in the Supplementary file Table S3. This inclusion and exclusion of data was finalized upon mutual agreement between authors.

### **Risk of bias assessment**

We assessed the potential for bias in the studies included in our analysis using the PROBAST tool for risk of bias assessment (Supplementary file Table S4). It evaluates four main domains: participants, predictors, outcomes, and analysis methods. Each domain has a specific number of questions: two, three, six, and nine, respectively, that collectively determine the study's risk of bias and overall reliability. Each question provides three possible responses: "yes/probably yes," "no/probably no," and "no information." If any question within a domain is answered with "no" or "probably no," the domain is classified as high risk. Conversely, if all questions are answered with "yes" or "probably yes," the domain is rated as low risk.

### **Quality of evidence**

The Grading of Recommendations Assessment, Development and Evaluation (GRADE) methodology was used to assess the quality of evidence of all the outcomes of our study. Outcomes included pooled training and validation AUC, sensitivity, and specificity of ML models predicting AKI after cardiac surgery. Due to the observational nature of all the included studies, evidence started at low and was downgraded based on the risk of bias (PROBAST), inconsistency (high heterogeneity), indirectness (training cohorts vs. independent validation), and imprecision (wide confidence intervals). Final certainty ratings were classified as high, moderate, low, or very low.

### **Data analysis**

We conducted a detailed statistical analysis to evaluate the predictive performance of the machine learning models included in this study. For models that were validated in more than two independent datasets, excluding the model development set, a random-effects meta-analysis was carried out to estimate their overall performance and accuracy. This analysis also included models that were internally validated through methods such as bootstrapping or cross-validation and externally validated in at least two independent datasets.

The discriminative ability of the models was assessed by extracting reported AUC. In instances where the AUC did not have a reported 95% confidence interval (CI) or standard error, we estimated the standard error using a formula based on the number of events and participants, following the guidelines proposed by Debray et al. [12]. Due to expected clinical, methodological, and algorithmic differences from variations in predictor selection, outcome definitions, and machine learning techniques,

we derived pooled estimates of AUC using a random effect model with logit-transformation and restricted maximum likelihood (REML) estimation. Additionally, sensitivity and specificity were extracted from included studies, and their meta-analysis was performed using a bivariate random effect model to account for the inherent correlation between sensitivity and specificity. This model uses logit-transformed values to include both within-study and between-study variability. When primary studies did not report  $2 \times 2$  contingency tables, true positives (TP), false positives (FP), false negatives (FN), and true negatives (TN) were generated using reported sensitivity, specificity, total sample size, and number of outcome events. These reconstructed counts were then used for bivariate random effects meta-analysis following the Reitsma model. Results were stratified by cohort (training and validation) and by ML model architecture type.

Heterogeneity was assessed using Cochran's Q statistic and  $I^2$  statistic. Heterogeneity statistics were calculated overall and within subgroups defined by ML model type. The significance of heterogeneity was evaluated using the Q test. For the Q test, a p-value less than 0.10 indicated a significant presence of heterogeneity. The  $I^2$  statistics quantified the proportion of total variation across studies that was due to heterogeneity instead of sampling error. Values of 25%, 50%, and 75% represented low, moderate, and high heterogeneity, respectively.

Moreover, publication bias and small study effects were assessed among validated prediction models using the funnel plot and Egger's test. For this, the reported primary model per study was included to ensure the statistical independence, using the AUC and its standard error derived from the 95% confidence interval. Funnel plot for visualization of asymmetry and Egger's regression test was applied to formally evaluate small study effects.

All statistical analyses and visualizations were performed using R (version 4.5.0). The "meta" package was used to perform meta-analysis of AUC, while, "mada" package was employed to execute bivariate random effect meta-analysis for sensitivity and specificity. Data preparation, subgroup filtering, and high-resolution forest plots were generated using "readxl", "dplyr", "forestplot", "magick", and "grid" packages.

## **Results**

### **Study selection**

The databases search yielded initially 1357 articles, including Cochrane (189), PubMed (106), Science Direct (872), Scopus (109), Web of Science (55), and Google Scholar (26). After duplicate removal of 440 articles, the remaining were screened for titles and abstracts. After a thorough screening process, ultimately 45 research

articles were included in the review. The PRISMA flow diagram for the study selection process is given in Fig. 1.

### Characteristics of included studies

The studies included in this systematic review and meta-analysis examined 13 distinct machine learning models applied to 401,579 cardiac surgery patients. These models included Logistic Regression (LR), Random Forest (RF), Support Vector Machine (SVM), eXtreme Gradient Boosting (XGBoost), LightGBM, k-Nearest Neighbors (k-NN), Naive Bayes (NB), Artificial Neural Networks (ANN), Ensemble methods, Decision Tree (DT), Adaptive boosting (AdaBoost), Bayesian Model Network, SoftMax, and other Gradient Boosting Decision Tree (GBDT) models.

The studies comprised 4 prospective cohort studies, 38 retrospective cohort studies, 1 retrospective case control, 1 post-hoc RCT analysis and 1 retrospective/prospective multi-center cohort study. These were retrieved from six major databases. The countries represented in the included literature were predominantly China (35 studies), followed by the United States (3 studies), Iran, Korea, Australia, Germany, France, Canada, and Taiwan, each contributing one study (Fig. 2). The detailed description of the characteristics of the included studies is given in Table 1.

### Risk of bias assessment

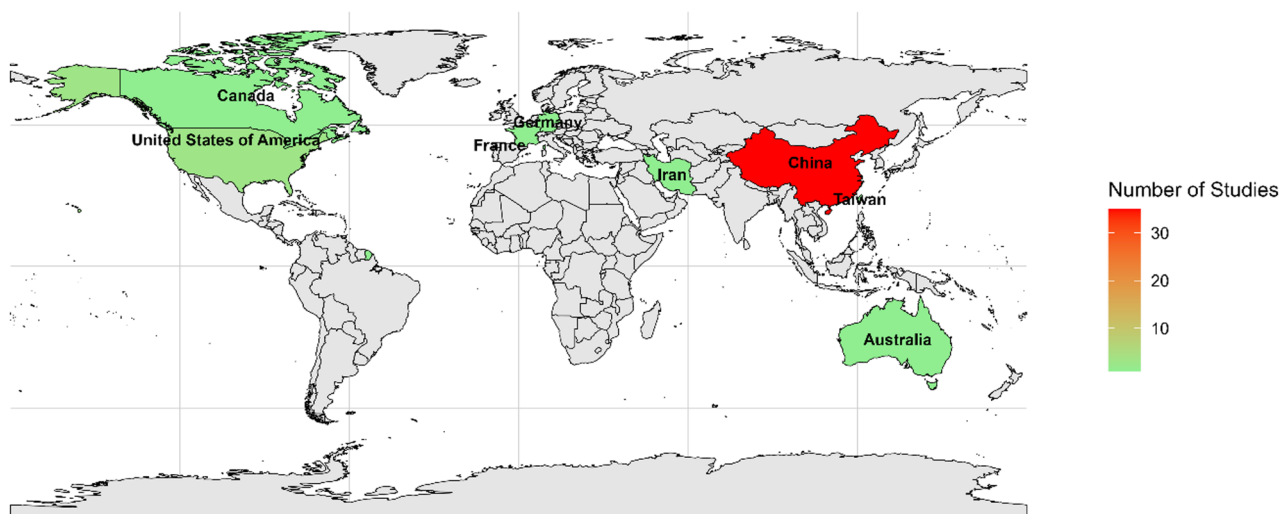
The quality assessment of the included studies was evaluated using the PROBAST tool. The domain-specific assessment is provided in Fig. 3, while detailed individual study ratings are available in the Supplementary file Table S5. Approximately 26.7% of the included studies were classified as having a low overall risk of bias, 44.4% were rated as high risk, and the remaining 28.9% were

categorized as unclear. The participant domain showed a relatively small proportion of high risk, with 86.7% of studies rated as low risk and only 11.1% rated as high risk. Regarding the predictors domain, low risk of bias was observed in 71.1% of the studies, while 22.2% were rated as unclear. In the outcome domain, assessment of AKI outcome was rated as low risk in 86.7% of the studies, with the remaining 13.3% classified as unclear. However, the analysis domain raised more concern, with 44.4% of studies being classified as having a high risk of bias, whereas 33.3% of studies were rated as low risk, and 22.2% were unclear (Fig. 3).

### AUC

Figure 4 illustrates that a total of 81 models were included in the meta-analysis of AUC in the training cohort for the prediction of AKI after cardiac surgery, while 162 models were included in the validation cohort derived from 44 primary studies included in the meta-analysis. Using random effect meta-analysis, the pooled AUC was estimated to be 0.83 (95% CI: 0.79–0.85) in the training cohort and 0.76 (95% CI: 0.75–0.78) in the validation cohort. Subgroup analysis stratified by ML model architecture suggested significant differences in discriminative performance across model types in both cohorts. In the training cohort, RF 0.90 (95% CI: 0.77–0.96), GBDT 0.87 (95% CI: 0.65–0.96), and XGBoost 0.85 (95% CI: 0.79–0.90) exhibited comparatively higher pooled AUC estimates. In the validation cohort, XGBoost 0.79 (95% CI: 0.76–0.82), RF 0.79 (95% CI: 0.76–0.81), and Ensemble methods 0.79 (95% CI: 0.77–0.81) showed relatively higher discriminative performance compared with other models.

Substantial heterogeneity was observed across most ML model types in both training and validation cohorts. In the training cohort, the random effect meta-analysis



**Fig. 2** Geographic distribution of studies included in the systematic review and meta-analysis

**Table 1** Characteristics of included studies

Authors	Country	Study type	Cardiac surgery types	Participants	Training data set	Validation data set	Diagnostic criteria for AKI	AKI cases	Generation way of verification set	External validation	Missing data handling	Machine learning models used
Mathieu Legrand (2013) [13]	France	Retrospective cohort	CPB for endocarditis	202	202	202	AKIN	120	V-fold CV	No	Imputation	Ensemble
Zhongli Chen (2020) [14]	China	Prospective cohort	ATAAD, TBAAD	402	204	198	AKIN	103, severe AKI 87	Two separate cohorts for screening and validation	No	Complete case analysis	LR
Yang Li (2020) [15]	China	Prospective cohort	CABG, Valvular, Aortic, and combination of these	5,533	3,639	1,894	KDIGO	2075	Temporal split, 10-fold CV	Yes	Complete case analysis	BN, LR
Guiyu Lei (2020) [16]	China	Retrospective cohort	Aortic arch	897	627	270	KDIGO	652	Random split (70:30), 10-fold CV	No	Complete case analysis	LR, RF, SVM, LightGBM
Po-Yu Tseng (2020) [17]	Taiwan	Retrospective cohort	CABG, Valvular, Aortic, and others using CPB	671	469	202	KDIGO	163	Random split (70:30)	No	Mean or median imputation	LR, DT, SVM, RF, XGBoost, Ensemble
Tim Coulson (2021) [18]	Australia	Retrospective cohort	CABG, Valve, Aortic, and others	22,731	17,048	5,683	KDIGO	5829	Random split (75:25)	No	Complete case analysis	LR
Penghua Hu (2021) [19]	China	Retrospective cohort	Valve, CABG with CPB	848	597	251	KDIGO	524	Random split, bootstrapping (n = 1000)	No	Multiple imputation	LR
Hao Cui (2021) [20]	China	Prospective cohort	CABG, Valve, Aorta, and others	214	95	119	AKIN	122	Clinical cohort split (CABG+VR vs. AA)	Yes	Complete case analysis	RF, SVM, LR
Penghua Hu (2021) [21]	China	Retrospective cohort	CABG, Valve, Aortic, CHD, and others	22,348	15,701	6,647	KDIGO	5,576	10-fold cross-validation, Random split	No	Multiple imputation	LR
Xin Xue (2022) [22]	China	Prospective cohort	Valve replacement, CABG, large vessel surgery, combination of these, congenital heart disease corrections	135	108	27	KDIGO	44	Random split (80:20), 5-fold CV	No	NR	LR, RF, XGBoost, SVM
Li Xinsai (2022) [23]	China	Retrospective cohort	TAAAD, TBAAD	456	319	137	KDIGO	201	Random split (70:30), 5-fold CV, bootstrapping (n = 1000)	No	Complete case analysis	LR, DT, XGBoost, LightGBM, RF
Hang Zhang (2022) [24]	China	Retrospective cohort	CABG, valve, and combination	1,457	1,170	287	KDIGO	353	Bootstrapping (n = 1000)	No	Multiple imputation	RF, Deep Forest, XGBoost, LR
Jurij Matijak (2022) [25]	Germany	Retrospective cohort	Cardiac surgery with CPB	7,507	6,756	751	KDIGO	1,699	Random split (90:10), 10-fold CV	No	Complete case analysis	LR, XGBoost
Arman Kilic (2022) [26]	United States	Retrospective cohort	Aortic valve replacement	243,142	194,513	48,629	KDIGO	6,911	Random split (80:20), bootstrapping (n = 1000)	No	Complete case analysis	XGBoost

**Table 1** (continued)

Authors	Country	Study type	Cardiac surgery types	Participants	Training data set	Validation data set	Diagnostic criteria for AKI	AKI cases	Generation way of verification set	External validation	Missing data handling	Machine learning models used
Jizhang Li (2022) [27]	China	Retrospective cohort	Acute aortic syndrome	1,637	1,318	319	KDIGO	NR	10-fold CV	Yes (319)	k-NN, imputation	XGBoost, LR
Charat Thongprayoon (2022) [28]	USA	Retrospective cohort	CABG, Valve, Heart transplant, Pericardiectomy	13,158	9,244	3,914	KDIGO	4,745	Hold out validation, random split (70:15:15)	No	Multiple imputation using RF	Ensemble, RF, DT, XGBoost, ANN, LR
Hang Zhang (2022) [29]	China	Retrospective cohort	CABG, Valve, and combination of these	11,740	8,197	3,543	KDIGO	3,237	Random split (70:30), 10-fold CV	No	Multiple imputation	LR
Azar Eijmalian (2022) [30]	Iran	Retrospective/Prospective cohort	CABG, Valvular, Transplant, Aortic	1,435	1,148	287	KDIGO	672	Random split (80:20), 10-fold CV	No	Multiple/Iterative imputation, k-NN, and Exclusion	LR, RF, DT, XGBoost, SVM, MLP, AdaBoost
Yelena Petrosyan (2022) [31]	Canada	Retrospective cohort	CABG, Single valve, and combination of these	6,522	4,566	1,956	KDIGO	1,760	Random split (70:30)	No	NR	Ensemble, LR
Tianchen Jia (2023) [32]	China	Retrospective cohort	CABG	2,780	2,224	556	KDIGO	510	Random split (80:20)	Yes (2051)	NR	LightGBM, SVM, RF, Softmax
Yun Yan (2023) [33]	China	Retrospective cohort	Valvular surgery	3,392	2,374	1,018	KDIGO	1,713	Bootstrapping (n=1000)	No	Imputation	LR, RF, XGBoost
Jiakang Shao (2023) [34]	China	Retrospective cohort	Cardiac surgery	2,108	1,686	422	KDIGO	637	5-fold CV, Cohort based split	Yes (422)	Complete case analysis	RF, ANN, LR, SVM, DT, k-NN
Yefeng Tong (2023) [35]	China	Retrospective case control	On-pump CABG	1,264	884	380	KDIGO	372	Random split (70:30), 10-fold CV	Yes	Handled by ML algorithms	XGBoost, LR, LightGBM, RF, k-NN, AdaBoost
Sai Zheng (2023) [36]	China	Retrospective cohort	Off-pump CABG	477	334	143	KDIGO	88	Random split (70:30)	No	Mean imputation	SVM, RF, GBDT, XGBoost, DT
Qian Li (2023) [37]	China	Post-hoc RCT analysis*	Cardiac surgery with CPB	2,416	1,933	483	KDIGO	1,819	Random split (80:20), 5-fold CV	Yes (562, 3517)	Mean and mode imputation	LR, SVM, k-NN, DT, RF, NB, GBDT, XGBoost, LightGBM, CatBoost, AdaBoost, Extra Trees
Jicheng Jiang (2023) [38]	China	Retrospective cohort	CABG, Valve, and others	2,310	1,848	462	KDIGO	1,020	Random split (70:30)	No	Multiple imputation	LR, DT, RF, GBDT, NB, MLP

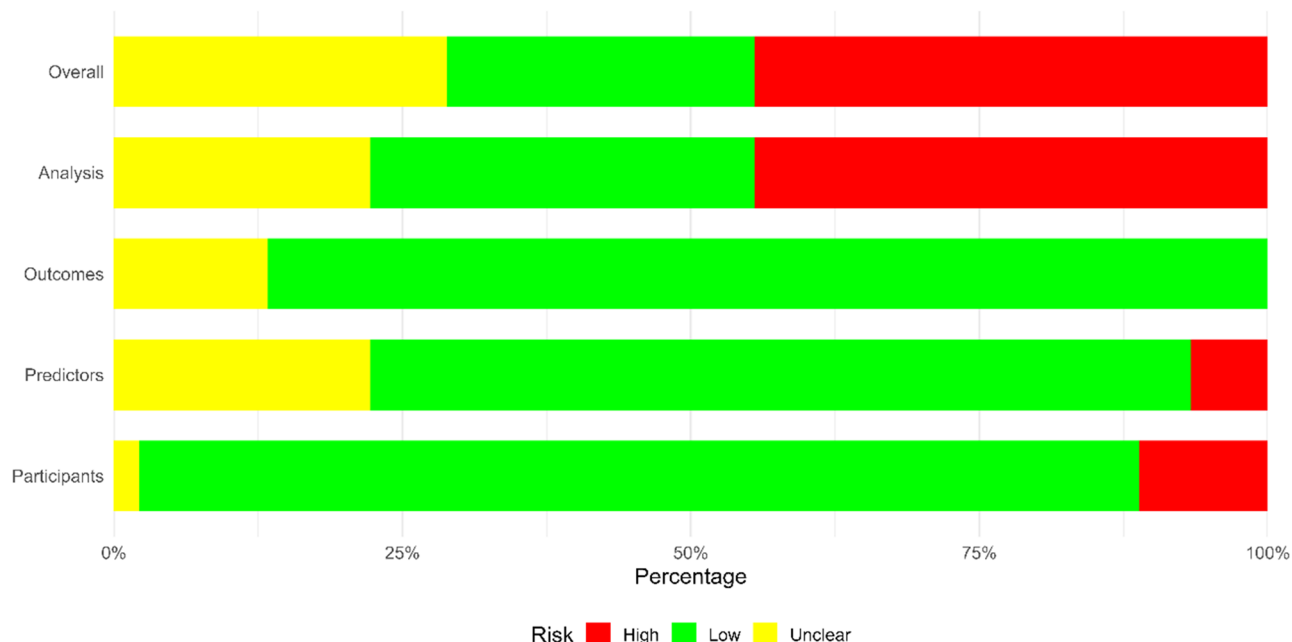
**Table 1** (continued)

Authors	Country	Study type	Cardiac surgery types	Participants	Training data set	Validation data set	Diagnostic criteria for AKI	AKI cases	Generation way of verification set	External validation	Missing data handling	Machine learning models used
Rui Fan (2023) [39]	China	Retrospective cohort	CABG, Valve, and combination of these	778	452	326	KDIGO	210	10-fold CV	No	Multiple imputation	DF, RF, XGBoost
Anran Dai (2023) [40]	China	Retrospective cohort	Acute Aortic dissection	265	212	53	KDIGO	191	Random split (80:20), 10-fold CV	No	k-NN imputation	LR, XGBoost, RF, SVM
Yuchen Gao (2023) [41]	China	Retrospective cohort	CABG, Valve, Aortic, CHD, Cardiomyopathy, Cardiac tumour	15,880	11,117	4763	KDIGO	4845	Random split (70:30), 5-fold CV	No	Multiple imputation	XGBoost, LR
Xuejian Hou (2024) [42]	China	Retrospective cohort	CABG	1,767	1,414	353	KDIGO	417	Random split (80:20), 5-fold repetition	No	NR	RF, LightGBM, XGBoost, LR, NB
Zhihe Zeng (2024) [43]	China	Retrospective cohort	Off-pump CABG	1,110	777	333	KDIGO	642	Random split (70:30), 5-fold CV	No	Deep learning-based imputation	LR, RF, DT, XGBoost
Xiaolong Liu (2024) [44]	China	Retrospective cohort	Total Aortic arch replacement	572	394	178	KDIGO	131	Random split (70:30), 10-fold CV, bootstrapping (n=100)	No	NR	ANN, LR
Yuanhan Chen (2024) [45]	China	Retrospective cohort	On-pump, Valve, Aortic, CHD, combination of these and others	5,368	3,768	1,600	KDIGO	2681	Random split (70:30)	No	Multiple imputation	LR
Changho HAN (2024) [46]	Korea	Retrospective cohort	Cardiac surgery with CPB	2,003	1,402	601	KDIGO	322	Random split (70:30)	No	NR	XGBoost
Yuezi Song (2024) [47]	China	Retrospective cohort	Off-pump CABG	701	561	140	KDIGO	73	Random split (80:20)	No	Complete case analysis (Excluded)	LR, GBDT, XGBoost, AdaBoost, RF, SVM, k-NN, DT
XinPei Liu (2024) [48]	China	Retrospective cohort	Active infective endocarditis cardiac surgery	527	368	159	KDIGO	261	Random split (70:30)	No	Mean or median imputation	LR, Linear SVM, Radial SVM, XGBoost, DT, RF
Yang Zhang (2025) [49]	China	Retrospective cohort	CABG	520	364	156	KDIGO	135	Random split (70:30), Grid search CV	No	k-NN imputation	RF, XGBoost, LR, LightGBM, Softmax, SVM
Kuroush Nezaferi (2025) [50]	USA	Retrospective cohort	CABG, Valvular, Aortic, Septal myectomy, and combination using CPB	602	481	121	KDIGO	110	Random split (80:20), 5-fold Grid search CV	No	Complete case analysis excluded	RF, LSTM
Ling Chen (2025) [51]	China	Retrospective cohort	ATAAD	1,010	758	252	KDIGO	466	Random split (75:25)	Yes (201)	Complete case analysis excluded	LR

**Table 1** (continued)

Authors	Country	Study type	Cardiac surgery types	Participants	Training data set	Validation data set	Diagnostic criteria for AKI	AKI cases	Generation way of verification set	External validation	Missing data handling	Machine learning models used
Haiming Li (2025) [52]	China	Retrospective cohort	CABG	2,155	1,509	646	KDIGO	365	Random split (70:30)	No	Mean imputation (< 20% missing), otherwise excluded	LR, DT, RF, SVM, XGBoost, AdaBoost, GBDT, LightGBM, k-NN
Biao Hou (2025) [53]	China	Retrospective cohort	CABG	3,043	2,130	913	KDIGO	465	Random split (70:30)	Yes (878)	Complete case analysis excluded	SVM, DT, RF, AdaBoost, XGBoost
Zishan Li (2025) [54]	China	Retrospective cohort	Cardiac surgery	4,565	3,196	1,369	KDIGO	3,046	Random split (70:30), 10-fold CV	No	Multiple imputation	RF, LASSO regression
Yang Xu (2025) [55]	China	Retrospective cohort	CABG, Valve, and combination of these	1,304	916	388	KDIGO	1,028	Random split (70:30), 10-fold CV	No	Multiple imputation (< 20% missing)	LR, LASSO regression, RF
Zheyuan Chen (2025) [56]	China	Retrospective cohort	ATAAD	1,350	1,148	202	KDIGO	586	Random split (85:15), 10-fold CV	No	Multiple imputation	GBDT, LightGBM, RF, k-NN, ANN, NB, LR
Qin Sun (2025) [57]	China	Retrospective cohort	CABG, Valve or combination of these	2,277	1,593	684	KDIGO	514	Random split (70:30)	Yes (1174)	missForest imputation	LR, RF, GBDT, XGBoost, SVM, Ensemble

Abbreviations: AKIN, Acute Kidney Injury Network; KDIGO, Kidney Disease: Improving Global Outcomes; ATAAD, Acute Type A Aortic Dissection; CABG, Coronary Artery Bypass Grafting; CHD, Congenital Heart Disease; CPB, Cardiopulmonary Bypass; RCT, Randomized Controlled Trial; TAAAD, Total Acute Aortic Arch Dissection; TBAAD, Type B Acute Aortic Dissection; AdaBoost, Adaptive Boosting; ANN, Artificial Neural Network; BMA, Bayesian Model Averaging; BN, Bayesian Network; CatBoost, Categorical Boosting; DF, Deep Forest; DT, Decision Tree; GBDT, Gradient Boosting Decision Tree; k-NN, k-Nearest Neighbors; LASSO, Least Absolute Shrinkage and Selection Operator; LightGBM, Light Gradient Boosting Machine; LR, Logistic Regression; LSTM, Long Short-Term Memory; MLP, Multilayer Perceptron; NB, Naive Bayes; RF, Random Forest; SVM, Support Vector Machine; XGBoost, Extreme Gradient Boosting



**Fig. 3** Risk of bias assessment results using the PROBAST tool

revealed considerable heterogeneity ( $I^2 = 98.2\%$ ,  $p < 0.001$ ). Similarly, the validation cohort exhibited very high heterogeneity ( $I^2 = 98.7\%$ ,  $p < 0.001$ ), indicating marked variability in predictive performance across included ML approaches. Furthermore, model specific subgroup analysis revealed persistently high within subgroup heterogeneity for most algorithms (Supplementary file Table S6 and S7). Furthermore, tests for subgroup differences indicated significant between model heterogeneity in both cohorts (training  $Q = 26.7$ ,  $p = 0.0034$ ; validation  $Q = 51.37$ ,  $p < 0.001$ ).

#### Sensitivity and specificity

Figure 5 illustrates a total of 17 models were evaluated in the training cohort, reporting sensitivity and specificity, which represented 7,997 patients undergoing cardiac surgery. The pooled sensitivity across these ML models was 0.75 (95% CI: 0.71–0.79), while the pooled specificity was 0.81 (95% CI: 0.72–0.87). In the validation cohorts, 72 models reported sensitivity and specificity, encompassing 10,063 patients. In this setting, the pooled sensitivity declined to 0.61 (95% CI: 0.53–0.69), whereas the pooled specificity remained consistently high at 0.82 (95% CI: 0.77–0.86).

Bivariate random effect meta-analysis of diagnostic performance of included ML models showed moderate to substantial between-model heterogeneity, with variability evident in both training ( $\tau = 0.42$ –1.04) and validation ( $\tau = 1.27$ –1.44) cohorts. Detailed heterogeneity estimates are presented in the Supplementary file Table S8.

#### Quality of evidence

Using the GRADE framework, the certainty of evidence for pooled estimates of AUC, sensitivity, and specificity was rated as very low across both training and validation cohorts. Detailed GRADE assessments for each outcome are presented in the supplementary file Table S9.

Moreover, the assessment of small study effects using Egger's regression test did not indicate evidence of asymmetry ( $p = 0.41$ ) (Supplementary file Figure S1).

#### Predictors of acute kidney injury

Across the included studies, predictors contributing to model performance were extracted based on reported feature importance rankings. The top five features utilized by ML models reported in each study were identified and summarized in Fig. 6.

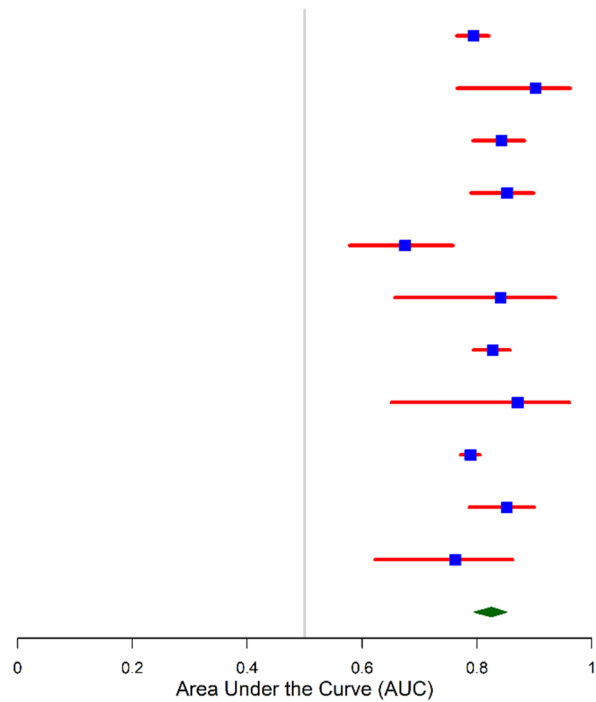
Among the most frequently reported predictors were preoperative serum creatinine, eGFR, age, and hypertension. Other predictors that appeared prominently were preoperative hemoglobin, intraoperative urine output, CPB time, and blood transfusion. A detailed list of predictors to develop ML models utilized by studies included in this review is given in the Supplementary file Table S10.

#### Discussion

The use of ML for predicting AKI after cardiac surgery has surged in recent years. Despite considerable progress, systematic reviews and meta-analyses that assess the predictive performance of these models remain limited. This systematic review and meta-analysis evaluated multiple

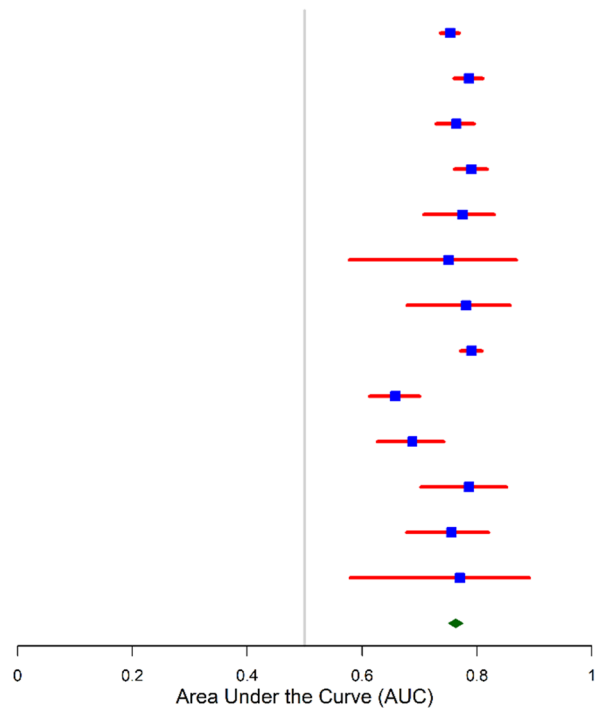
### (A) Training Cohort

Model	No. of Studies (N)	AUC (95% CI)
LR	22	0.79 (0.76–0.82)
RF	12	0.90 (0.77–0.96)
SVM	9	0.84 (0.79–0.88)
XGBoost	12	0.85 (0.79–0.90)
DT	7	0.67 (0.58–0.76)
ANN	3	0.84 (0.66–0.94)
AdaBoost	4	0.83 (0.79–0.86)
GBDT	3	0.87 (0.65–0.96)
NB	2	0.79 (0.77–0.81)
k-NN	2	0.85 (0.79–0.90)
LightGBM	5	0.76 (0.62–0.86)
<b>Summary</b>	<b>81</b>	<b>0.83 (0.79–0.85)</b>

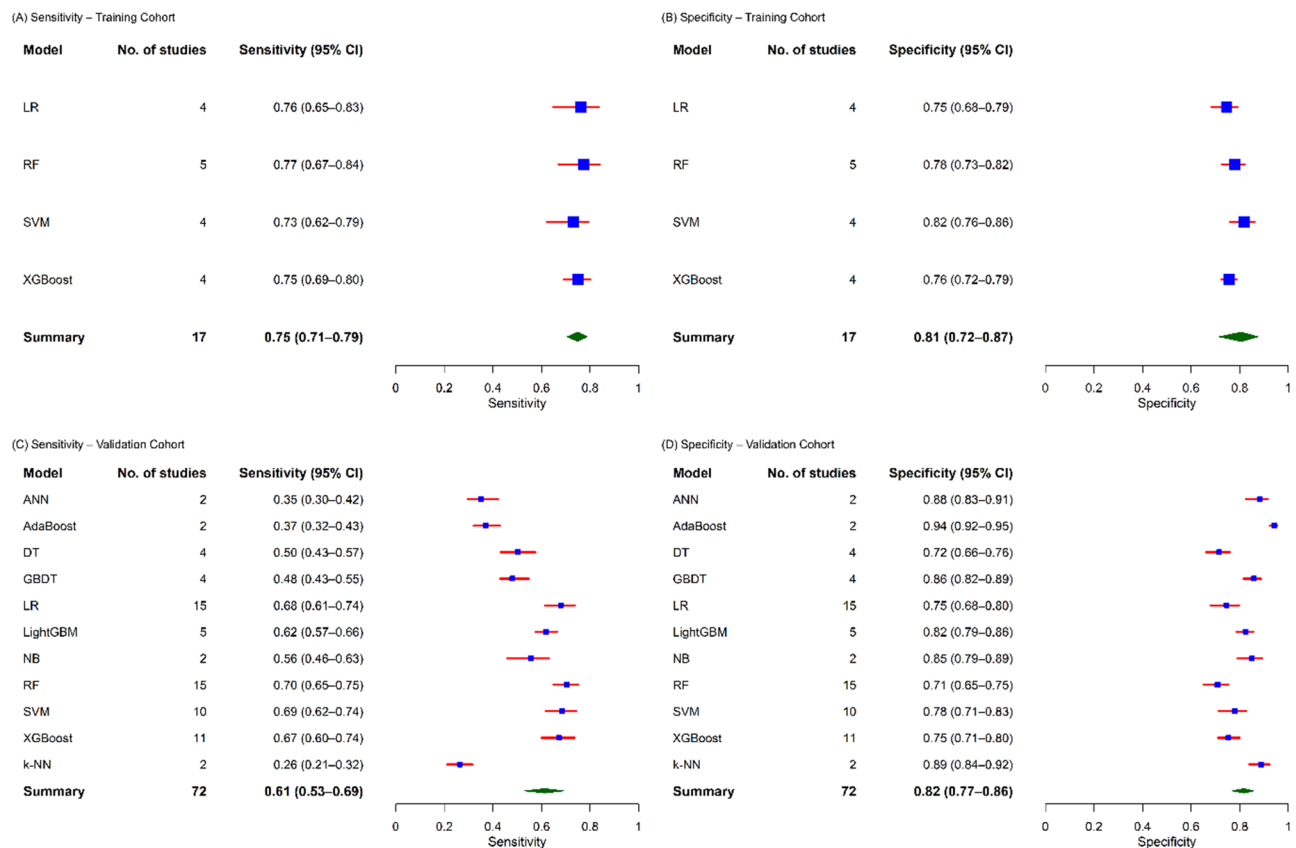


### (B) Validation Cohort

Model	No. of Studies (N)	AUC (95% CI)
LR	36	0.75 (0.74–0.77)
RF	31	0.79 (0.76–0.81)
SVM	19	0.76 (0.73–0.80)
XGBoost	23	0.79 (0.76–0.82)
LightGBM	10	0.77 (0.71–0.83)
NB	3	0.75 (0.58–0.87)
Softmax	3	0.78 (0.68–0.86)
Ensemble	3	0.79 (0.77–0.81)
k-NN	7	0.66 (0.61–0.70)
DT	11	0.69 (0.63–0.74)
GBDT	6	0.79 (0.70–0.85)
AdaBoost	6	0.76 (0.68–0.82)
ANN	4	0.77 (0.58–0.89)
<b>Summary</b>	<b>162</b>	<b>0.76 (0.75–0.78)</b>



**Fig. 4** Forest plots of pooled AUCs for machine learning models predicting AKI: training (A), validation (B)



**Fig. 5** Forest plots of sensitivity and specificity for machine learning models: training (A–B) and validation (C–D) cohorts

ML models and highlighted their predictive performance using AUC, sensitivity, and specificity in training and validation cohorts, offering valuable insights into their practical application in clinical settings. Across the 13 distinct models evaluated, ranging from traditional algorithms like LR to more advanced techniques such as gradient boosting (XGBoost, LightGBM) and ensemble methods, predictive outcomes were observed to have moderate to high performance. Our findings indicate that ML algorithms demonstrate good discriminative ability for AKI, with a pooled AUC of 0.83 (95% CI: 0.80–0.85) in the training and 0.76 (95% CI: 0.75–0.78) in the validation cohorts.

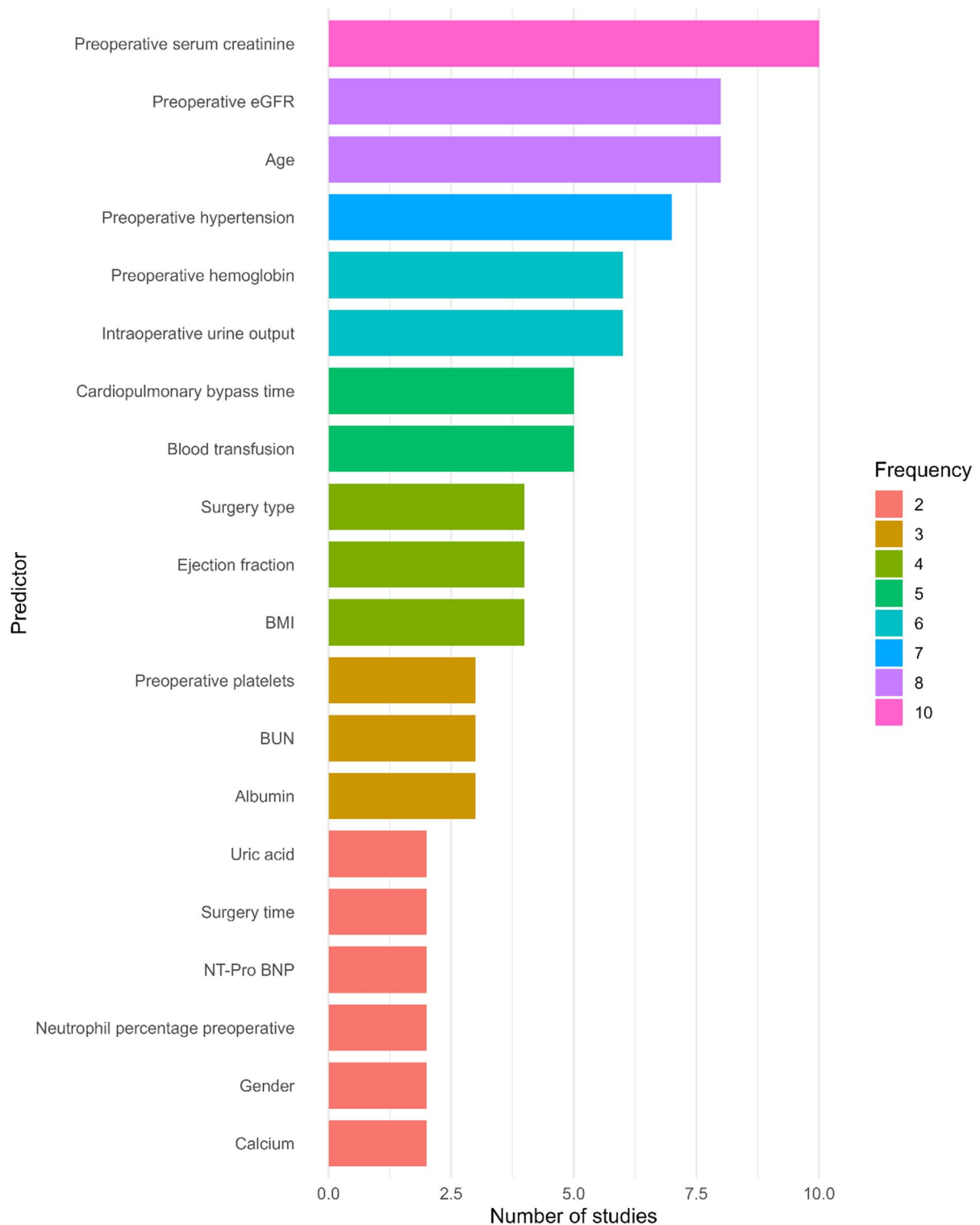
Risk stratification of AKI in cardiac surgery patients is usually performed with traditional risk analysis systems like regression models and risk scores. However, they have significant limitations in terms of handling complex, high-dimensional data, adapting to new information, and providing personalized, real-time predictions [22]. These shortcomings make them less suited for modern, data-rich healthcare environments, where ML models offer improved flexibility, accuracy, and predictive power [58].

The more advanced models, like ensemble techniques, RE, gradient boosting achieved the higher pooled AUCs in this study, whereas simpler approaches

like DT showed lower discrimination. Ensemble-based approaches benefit from their ability to combine predictions from multiple weak learners, effectively reducing variance and overfitting [59]. Gradient boosting methods iteratively correct misclassification from their previous trees, enabling them to capture complex nonlinear relationships among predictors that are common in clinical settings [60]. However, simpler models like DT suffer, as they are more prone to instability and overfitting, particularly when applied to heterogeneous clinical datasets with noisy or correlated predictors [1].

Notably, LR exhibited stable performance in both cohorts, suggesting that LR can be robust when applied to well-established clinical predictors with approximate linear associations to AKI. Conventional statistical approaches can remain clinically relevant, particularly when interpretability is crucial [61]. Moreover, the overall decrease in predictive performance of several ML models across training to validation cohorts signals potential overfitting in the training datasets, emphasizing the necessity of evaluation on independent datasets for robust prediction.

Bivariate random effect meta-analysis of sensitivity and specificity accounted for the trade-off between identifying patients at risk and avoiding false positives. In the



**Fig. 6** Frequently reported predictors of ML models for the prediction of AKI after cardiac surgery

validation cohort, LR and RF achieved moderate sensitivity (0.68–0.79) and specificity (0.71–0.78), whereas algorithms such as k-NN and ANN showed lower sensitivity. Ensemble and boosting methods generally

achieved higher specificity, reflecting greater accuracy in correctly identifying patients without AKI. These trends highlight the importance of tailoring model selection to the intended clinical objective, such as prioritizing higher

sensitivity to minimize missed AKI cases or specificity to reduce unnecessary interventions.

A substantial heterogeneity was observed across ML models in predicting AKI after cardiac surgery, which is common in the meta-analysis of diagnostic accuracy and ML-based prediction models. Unlike conventional intervention studies, predictive modelling research inherently varies in terms of model architecture, feature selection, dataset characteristics, and validation strategies. This between-model variance is consistent with the prior studies, which have shown that differences in patient populations, input variables, and outcome definitions can influence the predictive performance of ML models. While we did not explore the additional sources of heterogeneity due to inconsistent reporting of covariates, these findings highlight that model performance may vary depending on the study design and data characteristics, emphasizing the need for external validation before clinical implementation.

The clinical implementation of ML-based prediction faces several practical considerations. High-performing models such as Ensemble, RF, and XGBoost may offer superior accuracy but are less interpretable, potentially limiting clinician acceptance. The limitation of complex models, often referred to as “black box” models, necessitates the use of Explainable AI frameworks, such as SHapley Additive Explanation (SHAP), for a better understanding of the underlying features involved in predictions [62]. This understanding offers clinical insights regarding the specific physiological drivers behind an AKI alert. In contrast, LR and other transparent models provide interpretable outputs that can facilitate decision-making at the bedside.

Beyond algorithmic performance, the predictive accuracy of the ML models is substantially influenced by the selection of optimal modelling variables and their quality. Across the included studies, the input variables were a mix of demographic, preoperative, intraoperative, and early postoperative biomarkers. The most frequently reported features for the prediction of AKI after cardiac surgery were preoperative serum creatinine, eGFR, age, and hypertension, reflecting their well-established roles in AKI risk. Preoperative serum creatinine levels and eGFR are critical in assessing baseline renal function, and their levels often indicate pre-existing kidney damage or reduced renal reserve, placing patients at higher risk of AKI following surgery [63]. These variables also provide essential information on renal reserve and vulnerability to ischemic or inflammatory insults during cardiac surgery. Intraoperative urine output is another indicator of renal function during surgery, and its decrease is one of the earliest signs of impending AKI [64]. The prominence of renal function markers across studies suggests that they should remain a core component of any predictive

model for AKI. Similarly, age is a known risk factor for AKI due to the decline in kidney function with advancing age and the increased likelihood of comorbidities in older patients [65]. Hypertension and baseline hemoglobin also emerged as important predictors of postoperative AKI across the included studies, highlighting their clinical relevance in AKI risk stratification models.

A notable finding from this review is geographical bias, with 77.7% (35/45) of studies originating from China. This concentration introduces a potential geographical bias that could limit the global applicability of the results. Healthcare systems in Southeast Asia, i.e., China, may differ significantly from the West or other regions in terms of surgical protocols, perioperative care practices, and genetic predispositions to AKI that are specific to different ethnic groups. Furthermore, the prevalence of comorbidities like diabetes and the timing of surgical procedures often vary by region. Therefore, although the ML models analyzed in this review demonstrated acceptable performance within their study populations, their generalizability remains uncertain. Broader evaluation in diverse, multi-ethnic, and international populations is required, as only a limited proportion of included studies performed external validation using large, publicly available datasets such as the US-based MIMIC database [55].

Despite the strengths of this meta-analysis, several limitations must be acknowledged. First, substantial heterogeneity was observed across the included models, likely reflecting differences in modeling strategies and validation approaches. Although a random-effect model was used to account for between-model variability, the pooled performance metrics must be considered as average estimates across heterogeneous settings rather than universally applicable benchmarks. Second, most included studies were retrospective and relied on single-center datasets, with a limited proportion of studies that conducted external validations, raising concerns about the global applicability. Third, a lack of standardization in ML techniques across studies, such as differences in hyperparameter tuning, model optimization, and handling of missing data, also complicates the interpretation and replicability of results. Finally, the effect of publication bias cannot be fully excluded, as high-performance studies have more chances to get published. Although funnel plot asymmetry and statistical tests were explored, these methods have limited power in meta-analysis of predictive modeling studies.

## Conclusion

This meta-analysis offers a comprehensive evaluation of the effectiveness of various ML models in predicting AKI after cardiac surgery. Overall, the ML models exhibited moderate to high predictive ability, with a pooled AUC indicating reliable discrimination across the studies.

However, significant variability was noted, likely due to differences in modeling approaches, predictor selection, and validation strategies. Despite this heterogeneity and the associated risk of bias, our findings support the potential of ML-based prediction to assist in the risk stratification of AKI, which could help guide preventive strategies in clinical practice. Future research should prioritize external validation, standardization of reporting, and the integration of ML models into clinical workflows to evaluate their real-world effectiveness.

#### Abbreviations

AKI	Acute kidney injury
AUC	Area under the curve
SVM	Support vector machine
ML	Machine learning
PRISMA	Preferred reporting items for systematic reviews and meta-analyses
SHAP	SHapley additive explanation
KDIGO	Kidney disease improving global outcomes
AKIN	Acute kidney injury network
RF	Random forest
LR	Logistic regression
k-NN	k-Nearest neighbours
ANN	Artificial neural networks
XGBoost	eXtreme gradient boosting
DT	Decision tree
GBDT	Gradient boosting decision tree
AdaBoost	Adaptive boosting
NB	Naive bayes

#### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12911-026-03358-8>.

Supplementary Material 1

#### Acknowledgements

Not applicable.

#### Author contributions

RAA contributed to study conception, design, analysis of data, reviewing, editing of the manuscript, and supervision of ZAB. ZAB contributed to the study conception, design, data collection, analysis of data, interpretation and draft writing. BI and LSK contributed to the study design and critical review of the manuscript. RAA, BI, LSK and ZAB independently accessed and verified all extracted data. All authors had full access to all the data in the study and had final responsibility for the decision to submit to publication.

#### Funding

This study was funded by Universiti Malaya Research Excellence Grant UMREG020-2023. The funder played no role in study design, data collection, analysis and.

#### Data availability

The dataset supporting the conclusions of this article is included within the article and its additional files.

#### Declarations

##### Ethics approval and consent to participate

This study is a systematic review and meta-analysis of previously published studies. No new data were collected from human or animal participants, and therefore ethical approval and consent to participate were not required.

##### Consent for publication

Not applicable.

##### Competing interests

The authors declare no competing interests.

Received: 8 November 2025 / Accepted: 23 January 2026

Published online: 03 March 2026

#### References

1. Tseng PY, Chen YT, Wang CH, Chiu KM, Peng YS, Hsu SP, Chen KL, Yang CY, Lee OKS. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Crit Care*. 2020;24:1–13. <https://doi.org/https://pubmed.ncbi.nlm.nih.gov/32736589/>.
2. Thongprayoon C, Hansrivijit P, Bathini T, Vallabhajosyula S, Mekraksakit P, Kaewput W, Cheungpasitporn W. Predicting acute kidney injury after cardiac surgery by machine learning approaches. *J Clin Med*. 2020;9(6):1767. <https://doi.org/10.3390/jcm9061767>.
3. Yu X, Ji Y, Huang M, Feng Z. Machine learning for acute kidney injury: changing the traditional disease prediction mode. *Front Med*. 2023;10:1050255. <https://doi.org/10.3389/fmed.2023.1050255>.
4. Chen JJ, Lee TH, Chan MJ, Tsai TY, Fan PC, Lee CC, Wu VC, Tu YK, Chang CH. Electronic alert systems for patients with acute kidney injury: A systematic review and Meta-Analysis. *JAMA Netw Open*. 2024;7(8):e2430401. <https://doi.org/10.1001/jamanetworkopen.2024.30401>.
5. See CY, Pan HC, Chen JY, Wu CY, Liao HW, Huang YT, Liu JH, Wu VC, Ostermann M. Improvement of composite kidney outcomes by AKI care bundles: a systematic review and meta-analysis. *Crit Care*. 2023;27(1):390. <https://doi.org/10.1186/s13054-023-04641-0>.
6. Goldstein SL, Krallman KA, Roy JP, Collins M, Chima RS, Basu RK, Chawla L, Fei L. Real-Time acute kidney injury risk Stratification-Biomarker directed fluid management improves outcomes in critically ill children and young adults. *Kidney Int Rep*. 2023;8(12):2690–700. <https://doi.org/10.1016/j.ekir.2023.09.019>.
7. Zhang Q, Zheng P, Hong Z, Li L, Liu N, Bian Z, Chen X, Wu H, Zhao S. Machine learning in risk prediction of continuous renal replacement therapy after coronary artery bypass grafting surgery in patients. *Clin Exp Nephrol*. 2024;1–11. <https://doi.org/10.1007/s10157-024-02472-z>
8. Lee H-C, Yoon H-K, Nam K, Cho YJ, Kim TK, Kim WH, Bahk J-H. Derivation and validation of machine learning approaches to predict acute kidney injury after cardiac surgery. *J Clin Med*. 2018;7(10):322. <https://doi.org/10.3390/jcm7100322>.
9. Song Z, Yang Z, Hou M, Shi X. Machine learning in predicting cardiac surgery-associated acute kidney injury: A systemic review and meta-analysis. *Front Cardiovasc Med*. 2022;9:951881. <https://doi.org/10.3389/fcvm.2022.951881>.
10. Khwaja A. KDIGO clinical practice guidelines for acute kidney injury. *Nephron Clin Pract*. 2012;120(4):c179–84. <https://doi.org/10.1159/000339789>.
11. Lopes JA, Jorge S. The RIFLE and AKIN classifications for acute kidney injury: a critical and comprehensive review. *Clin Kidney J*. 2012;6(1):8–14. <https://doi.org/10.1093/ckj/sfs160>.
12. Debray TPA, Damen JAAG, Riley RD, Snell K, Reitsma JB, Hooft L, Collins GS, Moons KGM. A framework for meta-analysis of prediction model studies with binary and time-to-event outcomes. *Stat Methods Med Res*. 2018;28(9):2768–86. <https://doi.org/10.1177/0962280218785504>.
13. Section 5. Dialysis interventions for treatment of AKI. *Kidney Int Supplements*. 2012;2(1):89–115. <https://doi.org/10.1038/kisup.2011.35>.
14. Chen Z, Chen L, Yao G, Yang W, Yang K, Xiong C. Novel blood cytokine-based model for predicting severe acute kidney injury and poor outcomes after cardiac surgery. *J Am Heart Association*. 2020;9(22). <https://doi.org/10.1161/JAHA.120.018004>.
15. Li Y, Xu J, Wang Y, Zhang Y, Jiang W, Shen B, Ding X. A novel machine learning algorithm, bayesian networks model, to predict the high-risk patients with cardiac surgery-associated acute kidney injury. *Clin Cardiol*. 2020;43(7):752–61. <https://doi.org/10.1002/clc.23377>.
16. Lei G, Wang G, Zhang C, Chen Y, Yang X. Using machine learning to predict acute kidney injury after aortic arch surgery. *J Cardiothorac Vasc Anesth*. 2020;34(12):3321–8. <https://doi.org/10.1053/j.jvca.2020.06.007>.
17. Tseng P-Y, Chen Y-T, Wang C-H, Chiu K-M, Peng Y-S, Hsu S-P, Chen K-L, Yang C-Y, Lee OK-S. Prediction of the development of acute kidney injury following

- cardiac surgery by machine learning. *Crit Care*. 2020;24(1):478. <https://doi.org/10.1186/s13054-020-03179-9>.
18. Coulson T, Bailey M, Pilcher D, Reid CM, Seevanayagam S, Williams-Spence J, Bellomo R. Predicting acute kidney injury after cardiac surgery using a simpler model. *J Cardiothorac Vasc Anesth*. 2021;35(3):866–73. <https://doi.org/10.1053/j.jvca.2020.06.072>.
  19. Hu P, Chen Y, Wu Y, Song L, Zhang L, Li Z, Fu L, Liu S, Ye Z, Shi W, et al. Development and validation of a model for predicting acute kidney injury after cardiac surgery in patients of advanced age. *J Card Surg*. 2021;36(3):806–14. <https://doi.org/10.1111/jocs.15249>.
  20. Cui H, Shu S, Li Y, Yan X, Chen X, Chen X, Hu Y, Chang Y, Hu Z, Wang X, Song J. Plasma Metabolites-Based prediction in cardiac Surgery-Associated acute kidney injury. *J Am Heart Assoc*. 2021;10(22):e021825. <https://doi.org/10.1161/jaha.121.021825>.
  21. Hu P, Mo Z, Chen Y, Wu Y, Song L, Zhang L, Li Z, Fu L, Liang H, Tao Y, et al. Derivation and validation of a model to predict acute kidney injury following cardiac surgery in patients with normal renal function. *Ren Fail*. 2021;43(1):1205–13. <https://doi.org/10.1080/0886022X.2021.1960563>.
  22. Xue X, Liu Z, Xue T, Chen W, Chen X. Machine learning for the prediction of acute kidney injury in patients after cardiac surgery. *Front Surg*. 2022;9. <https://doi.org/10.3389/fsurg.2022.946610>.
  23. Xinsai L, Zhengye W, Xuan H, Xueqian C, Kai P, Sisi C, Xuyan J, Suhua L. Prediction model of acute kidney injury after different types of acute aortic dissection based on machine learning. *Front Cardiovasc Med*. 2022;9. <https://doi.org/10.3389/fcvm.2022.984772>.
  24. Zhang H, Wang Z, Tang Y, Chen X, You D, Wu Y, Yu M, Chen W, Zhao Y, Chen X. Prediction of acute kidney injury after cardiac surgery: model development using a Chinese electronic health record dataset. *J Transl Med*. 2022;20(1):166. <https://doi.org/10.1186/s12967-022-03351-5>.
  25. Kalisnik JM, Bauer A, Vogt FA, Stickl FJ, Zibert J, Fittkau M, Bertsch T, Kounev S, Fischlein T. Artificial intelligence-based early detection of acute kidney injury after cardiac surgery. *Eur J Cardiothorac Surg*. 2022;62(5). <https://doi.org/10.1093/ejcts/ezac289>.
  26. Kilic A, Goyal A, Miller JK, Gleason TG, Dubrawski A. Performance of a machine learning algorithm in predicting outcomes of aortic valve replacement. *Ann Thorac Surg*. 2021;111(2):503–10. <https://doi.org/10.1016/j.athoracsur.2020.05.107>.
  27. Li J, Gong M, Joshi Y, Sun L, Huang L, Fan R, Gu T, Zhang Z, Zou C, Zhang G, Qian X, Qiao C, Chen Y, Jiang W, Zhang H. Machine learning prediction model for acute renal failure after acute aortic syndrome surgery. *Front Med (Lausanne)*. 2021;8:728521. <https://doi.org/10.3389/fmed.2021.728521>.
  28. Thongprayoon C, Pattharantitima P, Kattah AG, Mao MA, Keddiss MT, Dillon JJ, Kaewput W, Tangpanithandee S, Krisanapan P, Qureshi F, Cheungpasitporn W. Explainable preoperative automated machine learning prediction model for cardiac surgery-associated acute kidney injury. *J Clin Med*. 2022;11(21). <https://doi.org/10.3390/jcm11216264>.
  29. Zhang H, Yu M, Wang R, Fan R, Zhang K, Chen W, Chen X. Derivation and validation a risk model for acute kidney injury and subsequent adverse events after cardiac surgery: a multicenter cohort study. *Int J Gen Med*. 2022;15:7751–60. <https://doi.org/10.2147/IJGM.S354821>.
  30. Ejmalian A, Aghaei A, Nabavi S, Abedzadeh Darabad M, Tajbakhsh A, Abin AA, Ebrahimi Moghaddam M, Dabbagh A, Jahangirifard A, Memary E, Sayyadi S. Prediction of acute kidney injury after cardiac surgery using interpretable machine learning. *Anesth Pain Med*. 2022;12(4):e127140. <https://doi.org/10.5812/aapm-127140>.
  31. Petrosyan Y, Mesana TG, Sun LY. Prediction of acute kidney injury risk after cardiac surgery: using a hybrid machine learning algorithm. *BMC Med Inf Decis Mak*. 2022;22(1). <https://doi.org/10.1186/s12911-022-01859-w>.
  32. Jia T, Xu K, Bai Y, Lv M, Shan L, Li W, Zhang X, Li Z, Wang Z, Zhao X, Li M, Zhang Y. Machine-learning predictions for acute kidney injuries after coronary artery bypass grafting: a real-life multicenter retrospective cohort study. *BMC Med Inf Decis Mak*. 2023;23(1). <https://doi.org/10.1186/s12911-023-02376-0>.
  33. Yan Y, Gong H, Hu J, Wu D, Zheng Z, Wang L, Lei C. Perioperative parameters-based prediction model for acute kidney injury in Chinese population following valvular surgery. *Front Cardiovasc Med*. 2023;10. <https://doi.org/10.3389/fcvm.2023.1094997>.
  34. Shao J, Liu F, Ji S, Song C, Ma Y, Shen M, Sun Y, Zhu S, Guo Y, Liu B, Wu Y, Qin H, Lai S, Fan Y. Development, External Validation, and visualization of machine learning models for predicting occurrence of acute kidney injury after cardiac surgery. *Rev Cardiovasc Med*. 2023;24(8). <https://doi.org/10.31083/j.rcm240829>.
  35. Tong Y, Niu X, Liu F. Prediction of acute kidney injury after extracorporeal cardiac surgery (CSA-AKI) by machine learning algorithms. *Heart Surg Forum*. 2023;26(5):E537–51. <https://doi.org/10.59958/hfsf.5673>.
  36. Zheng S, Li Y, Luo C, Chen F, Ling G, Zheng B. Machine learning for predicting the development of postoperative acute kidney injury after coronary artery bypass grafting without extracorporeal circulation. *Cardiovasc Innovations Appl*. 2023;7. <https://doi.org/10.15212/CVIA.2023.0006>.
  37. Li Q, Lv H, Chen Y, Shen J, Shi J, Zhou C. Development and validation of a machine learning predictive model for cardiac surgery-associated acute kidney injury. *J Clin Med*. 2023;12(3). <https://doi.org/10.3390/jcm12031166>.
  38. Jiang J, Liu X, Cheng Z, Liu Q, Xing W. Interpretable machine learning models for early prediction of acute kidney injury after cardiac surgery. *BMC Nephrol*. 2023;24(1). <https://doi.org/10.1186/s12882-023-03324-w>.
  39. Fan R, Qin W, Zhang H, Guan L, Wang W, Li J, Chen W, Huang F, Zhang H, Chen X. Machine learning in the prediction of cardiac surgery associated acute kidney injury with early postoperative biomarkers. *Front Surg*. 2023;10. <https://doi.org/10.3389/fsurg.2023.1048431>.
  40. Dai A, Zhou Z, Jiang F, Guo Y, Asante DO, Feng Y, Huang K, Chen C, Shi H, Si Y, Zou J. Incorporating intraoperative blood pressure time-series variables to assist in prediction of acute kidney injury after type a acute aortic dissection repair: an interpretable machine learning model. *Ann Med*. 2023;55(2). <https://doi.org/10.1080/07853890.2023.2266458>.
  41. Gao Y, Wang C, Dong W, Li B, Wang J, Li J, Tian Y, Liu J, Wang Y. An explainable machine learning model to predict acute kidney injury after cardiac surgery: A retrospective cohort study. *Clin Epidemiol*. 2023;15:1145–57. <https://doi.org/10.2147/CLEPS404580>.
  42. Hou X, Zhang K, Liu T, Xu S, Zheng J, Li Y, Dong R. Prediction of acute kidney injury following isolated coronary artery bypass grafting in heart failure patients with preserved ejection fraction using machine learning with a novel nomogram. *Rev Cardiovasc Med*. 2024;25(2). <https://doi.org/10.31083/j.rcm2502043>.
  43. Zeng Z, Tian X, Li L, Diao Y, Zhang T. An interpretable machine learning model to predict off-pump coronary artery bypass grafting-associated acute kidney injury. *Adv Clin Experimental Med*. 2024;33(5):473–81. <https://doi.org/10.17219/acem/169609>.
  44. Liu X, Fang M, Wang K, Zhu J, Chen Z, He L, Liang S, Deng Y, Chen C. Machine learning-based model to predict severe acute kidney injury after total aortic arch replacement for acute type A aortic dissection. *Heliyon*. 2024;10(13):e34171. <https://doi.org/10.1016/j.heliyon.2024.e34171>.
  45. Chen Y, Mo Z, Chu H, Hu P, Fan W, Wu Y, Song L, Zhang L, Li Z, Liu S, et al. A model for predicting postoperative persistent acute kidney injury (AKI) in AKI after cardiac surgery patients with normal baseline renal function. *Clin Cardiol*. 2024;47(1):e24168. <https://doi.org/10.1002/clc.24168>.
  46. Han C, Soh S, Kim HI, Song JW, Yoon D. Machine learning with clinical and intraoperative biosignal data for predicting cardiac Surgery-Associated acute kidney injury. *Stud Health Technol Inf*. 2024;316:286–90. <https://doi.org/10.3233/shti240400>.
  47. Song Y, Zhai W, Ma S, Wu Y, Ren M, Van den Eynde J, Nardi P, Pang PYK, Ali JM, Han J, Guo Z. Machine learning-based prediction of off-pump coronary artery bypass grafting-associated acute kidney injury. *J Thorac Dis*. 2024;16(7):4535–42. <https://doi.org/10.21037/jtd-24-711>.
  48. Liu X, Ai S, Yu R, Zhang C, Miao Q. Development and evaluation of a machine learning model for post-surgical acute kidney injury in active infective endocarditis. *Front Cardiovasc Med*. 2024;11:1425275. <https://doi.org/10.3389/fcvm.2024.1425275>.
  49. Zhang Y, Cai D, Deng Y, Wang Z, Zhang Z, Zhang H, Wang Q, Feng S, Sun L, Wei J. Machine learning based prediction of postoperative acute kidney injury risk in coronary artery bypass grafting patients. *Clin Interv Aging*. 2025;20:2033–48. <https://doi.org/10.2147/cia.S543303>.
  50. Nezafati K, Cheruku S, Wanyan T, Yang DM, Zhou Q, Jessen ME, Neyra JA, Zhan X, Xie Y, Fox A. Predicting In-Hospital acute kidney injury after cardiac surgery using machine learning. *J Cardiothorac Vasc Anesth*. 2025. <https://doi.org/10.1053/j.jvca.2025.09.045>.
  51. Chen L, Pan Y, Zhang H, Xue H, Qiu Z, Chen L. Early postoperative acute kidney injury prediction in patients with acute type A aortic Dissection. *iScience*. 2025;114222. <https://doi.org/10.1016/j.isci.2025.114222>.
  52. Li HM, Hu H, Li JX, Peng WX. Machine learning-based model for the prediction of acute kidney injury following coronary artery bypass graft surgery in elderly Chinese patients. *J Thorac Disease*. 2025;17(4):2519–27. <https://doi.org/10.21037/jtd-2025-264>.
  53. Hou B, Liu T, Yan P, Wang Y, Hou X, Li L, Yang H, Chen L, Liu T, Zhang K, Xu S, Li Y, Dong R. Deep learning-based prediction model of acute kidney

- injury following coronary artery bypass grafting in coronary heart disease patients: a multicenter clinical study from China. *Front Cardiovasc Med.* 2025;12:1600012. <https://doi.org/10.3389/fcvm.2025.1600012>.
54. Li Z, Wang L, Zhang X, Wu A, Liu T. Integration of machine learning and large Language models for screening and identifying key risk factors of acute kidney injury after cardiac surgery. *Front Med (Lausanne).* 2025;12:1618222. <https://doi.org/10.3389/fmed.2025.1618222>.
  55. Xu Y, Song C, Wei W, Miao R. Development and validation of a cardiac surgery-associated acute kidney injury prediction model using the MIMIC-IV database. *PLoS ONE.* 2025;20(6):e0325151. <https://doi.org/10.1371/journal.pone.0325151>.
  56. Chen Z, Lu X, Liu M, Yu Y, Yu L, Cheng S, Zhu Z, Lai Y, Liu N. Artificial Intelligence-Driven prediction of acute kidney injury following acute type A aortic dissection surgery in a Chinese population. *J Cardiothorac Vasc Anesth.* 2025;39(10):2729–38. <https://doi.org/10.1053/j.jvca.2025.05.047>.
  57. Sun Q, Ma MQ, Du X, Wan X, Cao CC, Liang XB. Explainable machine learning models for early prediction of acute kidney injury after cardiac surgery. *Int Urol Nephrol.* 2025. <https://doi.org/10.1007/s11255-025-04807-5>.
  58. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. *BMC Med Res Methodol.* 2019;19(1):64. <https://doi.org/10.1186/s12874-019-0681-4>.
  59. Sagi O, Rokach L. Ensemble learning: A survey. *WIREs Data Min Knowl Discov.* 2018;8(4):e1249. <https://doi.org/10.1002/widm.1249>.
  60. Bao W, Zhao Y, Canes A, Steinberg D, Lyashevskaya O. Predictive analytics with gradient boosting in clinical medicine. *Annals Translational Med.* 2019;7:7:152. <https://doi.org/10.21037/atm.2019.03.29>.
  61. Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol.* 2019;110:12–22. <https://doi.org/10.1016/j.jclinepi.2019.02.004>.
  62. Hettikankanamage N, Shafiqabady N, Chatteur F, Wu RMX, Ud Din F, Zhou J. eXplainable artificial intelligence (XAI): a systematic review for unveiling the black box models and their relevance to biomedical imaging and sensing. *Sens (Basel).* 2025;25(21). <https://doi.org/10.3390/s25216649>.
  63. Parolari A, Pesce LL, Pacini D, Mazzanti V, Salis S, Sciacovelli C, Rossi F, Alamanni F. Risk factors for perioperative acute kidney injury after adult cardiac surgery: role of perioperative management. *Ann Thorac Surg.* 2012;93(2):584–91. <https://doi.org/10.1016/j.athoracsur.2011.09.073>.
  64. Moreira R, Jacinto T, Neves P, Vouga LA, Baeta C. Predictors of acute kidney injury associated with cardiopulmonary bypass. *Rev Port Cir Cardiorac Vasc.* 2019;26(2):109–15 <https://doi.org/10.48729/pjctvs.123>.
  65. Khademi S, Mehr LS, Janati M, Jouybar R, Dehghanpisheh L. Association of urine output during cardiopulmonary bypass and postoperative acute kidney injury in patients undergoing coronary artery bypass graft. *Perfusion.* 2022;38(3):567–73. <https://doi.org/10.1177/02676591211068973>.

### Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.