

# Artificial Intelligence for the prediction of Acute Kidney Injury during the perioperative period: Systematic Review and Meta-Analysis of Diagnostic Test Accuracy

**Hanfei Zhang**

University of Electronic Science and Technology of China

**Amanda Y Wang**

George Institute for Global Health, UNSW

**Shukun Wu**

University of Electronic Science and Technology of China

**Johnathan Ngo**

University of Sydney

**Yunlin Feng**

University of Electronic Science and Technology of China

**Xin He**

Affiliated Hospital of Southwest Medical University

**Yingfeng Zhang**

University of Electronic Science and Technology of China

**Xingwei Wu**

University of Electronic Science and Technology of China

**Daqing Hong** (✉ [hongdaqing11@126.com](mailto:hongdaqing11@126.com))

University of Electronic Science and Technology of China

---

## Research Article

**Keywords:** artificial intelligence, machine learning, acute kidney injury, acute kidney failure, perioperative period

**Posted Date:** February 15th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1209770/v2>

**License:** © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

---

## Abstract

## Background

Acute kidney injury (AKI) is independently associated with morbidity and mortality in a wide range of surgical settings. Nowadays, with the increasing use of electronic health records (EHR), advances in patient information retrieval, and cost reduction in clinical informatics, artificial intelligence is increasingly being used to improve early recognition and management for perioperative AKI. However, there is no quantitative synthesis of the performance of these methods.

## Objective

To estimate the sensitivity and specificity of artificial intelligence for the prediction of acute kidney injury during the perioperative period.

## Methods

Pubmed, Embase, and Cochrane Library were searched to 2nd October 2021. Studies presenting diagnostic performance of artificial intelligence in the early detection of perioperative acute kidney injury were included. Two independent evaluators extracted data. The risk of bias of eligible studies was assessed using the PROBAST tool.

## Results

Nineteen studies involving 304,076 patients were included. Quantitative random-effects meta-analysis using the Rutter and Gatsonis hierarchical summary receiver operating characteristics (HSROC) model revealed pooled sensitivity, specificity, and diagnostic odds ratio of 0.77 (95% CI: 0.73 to 0.81), 0.75 (95% CI: 0.71 to 0.80), and 10.7 (95% CI 8.5 to 13.5), respectively. Threshold effect was found to be the only source of heterogeneity, and there was no evidence of publication bias.

## Conclusions

Our review demonstrates the promising performance of artificial intelligence for early prediction of perioperative AKI. Further studies should focus on the improvement of existing models, novel biomarkers, and clinical effectiveness.

## Introduction

Acute Kidney Injury (AKI) is a clinical syndrome characterised by a sudden decrease in glomerular filtration rate, defined by a rapid increase in serum creatinine, decrease in urine output, or both <sup>1</sup>. Noteworthy, AKI in the perioperative period is one of the most serious yet under-recognised complications, associated with increased risk of morbidity and mortality, chronic kidney disease, long-term adverse events, and increased cost and resource utilisation <sup>2-4</sup>. Nephrologists should recognise the huge medical burden.

Despite remarkable improvements in the identification of high-risk patients <sup>5</sup>, assessment of AKI is still based on two relatively non-specific markers that may lack utility in discriminating patients with incipient AKI: serum creatinine (SCr) and urine output (UO) <sup>6</sup>. Oliguria is neither sensitive nor specific and more likely to be associated with haemodynamic response to both hypovolemia and intracellular dehydration <sup>7</sup>. Moreover, SCr detected may vary in critically ill patients (e.g., severe hepatic disease) or by diet (e.g., food rich in proteins). In addition, sarcopenia and sepsis lead to reduced creatine release and decreased creatinine production <sup>6</sup>. The lack of an early marker contributes to delays in recognition and may attenuate the opportunity for early successful intervention.

The advances in clinical informatics and the widespread adoption of electronic health records (EHR) have allowed for the development of artificial intelligence (AI) models which can automatically trigger an electronic alert to physicians <sup>8</sup>. Since Thottakkara's study on a machine learning prediction model of postoperative acute kidney injury in 2016, AIs have evolved quickly and demonstrated improved accuracy in identifying patients at risk of developing AKI, as well as early recognition of subclinical AKI, compared with traditional multivariate regression models <sup>9</sup>. However, there is no quantitative synthesis of the diagnostic accuracy of these methods. Researchers have tried different ways, including but not limited to expanding sample sizes, use of real-time predictive analytics, finding novel biomarkers, and optimising algorithms, in an attempt to raise diagnostic accuracy but have received conflicting results <sup>10,11</sup>.

We conducted a systematic review and meta-analysis to quantitatively analyse the diagnostic accuracy of the AIs in detecting acute Kidney Injury during the perioperative period and investigated the factors that affected diagnostic accuracy.

## Methods

### Data sources and searches

Two independent evaluators searched PubMed, Embase, and the Cochrane Library using combined free texts and MeSH terms relating to the perioperative period, acute kidney injury, and AI (prior to October 2021). The abstracts of all identified studies were reviewed to exclude irrelevant articles. Full-text reviews were conducted to determine whether the inclusion criteria were satisfied in all the studies. We also manually checked the reference lists of relevant

publications including reviews and commentaries to include eligible studies. Disagreements were resolved by a discussion between two evaluators. Appendix 1 shows the detailed search strategy.

### **Selection Criteria**

Studies were eligible if they met the following inclusion criteria: (1) AKI was defined using consensus criteria such as RIFLE, AKIN, and KDIGO, or studies with clear AKI definitions, (2) the main outcome was the onset of AKI during the immediate pre-operative period until the time of discharge, (3) application of the AI algorithm for the prediction of perioperative acute kidney injury, (4) inclusion of diagnostic performance indices of the AI algorithm, including specificity, sensitivity, positive likelihood ratio (PLR), negative likelihood ratio (NLR), positive predictive value (PPV), negative predictive value (NPV), or the figure of the area under the receiver operating characteristic curve, which enables the construction of a 2x2 diagnostic table, and (5) human adult subjects.

The exclusion criteria were the studies that were not original studies such as letters, comments, editorials, protocols or reviews

### **Data extraction and quality assessment**

The data that was extracted independently by two investigators included study characteristics (authors and year of publication), characteristics of the sample set (sample size, age, sex, and type of surgery), characteristics of the index test (external validation, number of predictors, and type of AIs), characteristics of reference standard, and accuracy data (number of true positives, true negatives, false positives, and false negatives). If different types of models were compared in the same study, we only included the model which had the highest diagnostic accuracy. When original studies reported the sensitivity and specificity under multiple thresholds, we extracted the accuracy data under the threshold with the largest Youden's index, defined as the sum of sensitivity and specificity minus one. If both the internal validation and external validation were performed, the two-by-two data of the latter was extracted, because of better generalisability.

We assessed the methodological quality in 20 signalling questions in 4 key domains: participants, predictors, outcome, and analysis of each study using the Prediction model Risk Of Bias Assessment Tool (PROBAST), which is a risk of bias assessment tool designed for systematic reviews of diagnostic or prognostic prediction models<sup>12,13</sup>. According to the signal problem and the author's judgment, each of the domains was divided into "high", "low" and "unclear". Overall risk of bias is graded as low risk when all domains are considered low risk, and overall risk of bias is considered high risk when at least one of the domains is considered high risk.

### **Data synthesis and analysis**

Extracted two-by-two data were first graphically shown in the forest plot with the point estimate of sensitivity and specificity and their 95% confidence intervals (Cis). To remove the effect of a possible heterogeneous threshold, we conducted a quantitative random-effects meta-analysis using Rutter and Gatsonis hierarchical summary receiver operating characteristics (HSROC) model to combine summary receiver operating characteristic curves (SROC) curve which was the standard method for meta-analysing diagnostic studies reporting pairs of sensitivity and specificity<sup>14</sup>. This method comprehensively considers the effect of diagnostic tests under different diagnostic thresholds and converts the diagnostic odds ratio (DOR) by the sensitivity and specificity of each pair as the only metric of diagnostic analysis<sup>15</sup>.

Subgroup analysis and meta-regression were used to explore the potential heterogeneity. The following pre-specified subgroup analyses were performed based on AI type, surgery type, number of patients, external validation, diagnostic criteria, and methodological quality of included studies. We regarded the factor as a source of heterogeneity if the coefficient of the covariate was statistically significant ( $P < 0.05$ ). Because the Metandi and Midas package of STATA required a minimum of four studies to conduct the diagnostic test accuracy meta-analysis (reference), if less than four studies were enrolled in the subgroup analysis, Meta-DiSc 1.4 using the 'Moses-Shapiro-Littenberg method' was used (reference).

We performed sensitivity analysis to evaluate the robustness of our main outcomes by exploring the effect of excluding one study at a time and used Deek's funnel plot<sup>16</sup> to assess the presence of publication bias. All the data analysis were conducted in STATA (version 16.0) with the two-tailed probability of type I error of 0.05 ( $\alpha = 0.05$ ).

## **Results**

### **Identification of Relevant Studies**

A total of 540 articles were identified by searching three electronic databases. Among them, 105 were duplicate studies, and 384 were excluded during the initial screening by reviewing titles and abstracts. The full texts of the remaining 53 articles were thoroughly reviewed. Among these, 34 studies were excluded from the final analysis due to the following reasons: abstract ( $n = 15$ ), review ( $n = 11$ ), clinical score ( $n = 2$ ), study with incomplete data ( $n = 2$ ), failed to get the original text ( $n = 3$ ) and did not pertain to topic ( $n = 1$ , the topic of this article was automated identification of the electronic medical record). The remaining 19 studies were included in the final analysis, which was shown in Figure 1.

### **Characteristics of eligible studies**

The total number of subjects tested in the included studies was 304,076, with the sample size ranged from 109 to 96,653<sup>17-35</sup>.

Seventeen studies described the demographic characteristics of their study population, of whom the mean age was 37 to 71 years old and the percentage of males was 16% to 88%<sup>17,19-31,33-35</sup>.

The included studies were categorized based on the type of the surgery participants received, including cardiothoracic surgery, any inpatient operative procedure, liver transplantation, total knee arthroplasty<sup>17-35</sup>.

Enrolled studies presented the performance of the AI algorithms with test dataset (internal validation), and there were only four studies<sup>21,26,27,34</sup> that presented the performance of external validation. Nine studies<sup>21-25,28,32-34</sup> established the AI algorithm based on the gradient boosting machine (GBM), three studies<sup>17,19,35</sup> established random forest (RF)-based algorithms, three studies<sup>20,27,29</sup> established two types of artificial neural network (ANN)-based algorithms, one study<sup>26</sup> established Bayesian network (BN)-based algorithm, one study<sup>31</sup> established decision-tree (DT)-based algorithm, one study<sup>30</sup> established an ensemble algorithm, and another study even conducted a novel machine learning risk algorithm<sup>18</sup> called: MySurgeryRisk.

Fifteen studies applied the Kidney Disease Improving Global Outcomes (KDIGO) definition for AKI<sup>17-19,21,22,24-27,29-34</sup>. Among these, some used serum creatinine changes only to define AKI while urine output criteria were not adopted<sup>21,23,25,29,34</sup>. Two studies applied the Acute Kidney Injury Network (AKIN) criteria<sup>20,23</sup>.

These characteristics (modifiers) were evaluated as potential sources of heterogeneity through subgroup analysis and meta-regression. (Table 1) shows the detailed characteristics of the studies.

### Methodological Quality of the Studies (Figure 2)

Among the 19 studies<sup>17-35</sup> in the final analysis, 4 studies<sup>18,25,32,33</sup> showed low risk of bias, 2 studies<sup>26,29</sup> showed unclear risk of bias, and 13 studies<sup>17,19-24,26-28,30-33</sup> showed high risk of bias.

Regarding the participants domain, the risk of bias was low in 18 studies<sup>17-25,27-35</sup> and unclear in one due to insufficient information describing the sampling method in external validation<sup>26</sup>.

Concerning the predictors domain, we considered the risk of bias unclear in one study<sup>31</sup> because the details of the predictors were not reported.

In terms of the outcomes, 15 studies<sup>17-19,21,22,24-27,29-34</sup> applied the Kidney Disease Improving Global Outcomes (KDIGO) definition for AKI, but we considered the risk of bias unclear in five studies<sup>21,22,24,29,34</sup> because they utilised creatinine changes only. The risk of bias was high in one study<sup>27</sup> because only patients with severe AKI were enrolled. In addition, two studies<sup>28,35</sup> which used their own criteria for AKI were also considered to have high risk of bias.

The most concerning issue seen in the analysis was the high risk of bias in majority of the included studies (13/19). The risk of bias in 12 studies<sup>17,19-23,27,28,30,31,34,35</sup> was considered high.

Overall, studies<sup>17,19-24,26-28,30-33</sup> with high risk in at least one of the four domains were rated as low methodological quality in the diagnostic test accuracy of artificial intelligence for the prediction of acute kidney injury during the perioperative period (Figure 2)

### Diagnostic Test Accuracy of Artificial Intelligence for the prediction of Acute Kidney Injury during perioperative period

The Figure 3 showed the paired forest plot for sensitivity and specificity with the corresponding 95% CIs for each study. The SROC curve, with a 95% confidence region, was illustrated in Figure 4. The following summarised estimates using the HSROC model were also calculated: sensitivity 0.77 (95% CI: 0.73 to 0.81), specificity 0.75 (95% CI: 0.71 to 0.80), positive likelihood ratio 3.2 (95% CI: 2.7 to 3.7), negative likelihood ratio 0.30 (95% CI: 0.26 to 0.35), and diagnostic odds ratio 10.7 (95% CI 8.5 to 13.5). To investigate the clinical utility of AI, a Fagan nomogram was generated. Assuming a 50% prevalence of AKI during the perioperative period, the Fagan nomogram shows that the posterior probability of AKI was 76% if the test was positive, and the posterior probability of the absence of AKI was 23% if the test was negative (Figure 5).

### Exploring Heterogeneity with Meta-Regression and Subgroup Analysis

The shape of the SROC curve was symmetric (Figure 4). However, we observed a medium positive correlation after logit transformed TPR and FPR (Spearman correlation coefficient=0.48), and an asymmetric parameter,  $\beta$ , with a significant P-value (P=0.036) indicating threshold heterogeneity among the studies.

The heterogeneity was not found among the included studies in the joint model of meta-regression (type of AI [P=0.58], number of included patients [P=0.22], type of surgery [P=0.17], methodological quality [P=0.93], external validation [P=0.69], the definition of AKI [p=.14] Figure 6)

(Table 2) shows the detailed results of subgroup analysis exploring the potential source of between-study heterogeneity.

### Sensitivity analysis

After excluding one study at a time, the results (figure 7) showed that every result is 95% within the confidence interval, combined DOR was 10.66 (95% CI: 8.47 to 13.40), which meant the outcomes of meta-analysis was robust.

### Publication Bias

Publication bias were assessing using Deek's funnel plot for the prediction of AKI during the perioperative period (Figure 8). The plot was grossly symmetrical with respect to the regression line. The Deek's funnel plot asymmetry test showed no evidence of publication bias (P=0.62).

## Discussion

To the best of our knowledge, this is the first systematic review and meta-analysis assessing the predictive utility of artificial intelligences (AIs) in AKI during the perioperative period. Due to heterogeneous thresholds, the current optimal way to merge data is using the hierarchical summary receiver operating characteristics (HSROC) model<sup>14</sup>. Our study showed that the AIs can correctly detect 77% (95% CI: 0.73 to 0.81) of the patients with perioperative AKI and exclude 75% (95% CI: 0.71 to 0.80) of patients without perioperative AKI. These results presented better performance compared to the clinical scoring tools physicians used<sup>18,28,34</sup> and implied application prospects of artificial intelligences in perioperative AKI.

In a lot of cases, perioperative AKI are managed by non-nephrologists who may have reduced awareness of AKI and have a paucity of effective interventions<sup>36</sup>. In the developed countries, 30~45% of patients experienced drug-related adverse events in the non-nephrology departments<sup>37,38</sup>. The delayed recognition of nephrotoxins in other departments was associated with higher mortality compared to those in the nephrology or urology department<sup>36</sup>. A widespread application of AI could send electronic alerts, provide a second opinion, and offer opportunities for identifying patients at risk within a time window that enables renal referral<sup>39,40</sup>. Currently, how physicians would react to the early prediction made by AIs is not clear. Therefore, a prospective study based on the application of AI in clinical practice is needed.

Another important finding of this study is the robustness of the predictive performance of the AI algorithm, irrespective of the modifiers detected during the systematic review process such as the type of AIs, the type of surgery, or the criteria used in diagnosis.

The gradient boosted machine was attention-worthy as it showed the best performance in both liver transplantation and cardiac surgery<sup>19-21,23</sup>. However, after comparing the performance of seven artificial intelligence algorithms using meta-regression, no significant difference among them were found. In subgroup analysis, RF (random forest) even was superior to GBM (gradient boosting machine) with pooled sensitivity and specificity of 0.82 and 0.74 compared with 0.77 and 0.69, respectively, indicating that other algorithms might also have great potential in clinical application with predictive accuracy as good as gradient boosted machine.

The occurrence of acute kidney injury in patients receiving cardiac and vascular surgery has been widely reported, but less information was available regarding non-cardiac surgery<sup>41</sup>, probably due to its overall lower incidence which is approximately 1% of general surgery cases<sup>42</sup>. Therefore, more research is required before we draw a conclusion regarding the influence of surgery type.

Our study showed that none of pre-specified subgroups showed an impact on the predictive accuracy. It suggested that the development of artificial intelligence might have hit a plateau and it might be difficult to further optimise predictive accuracy through existing methods without technological innovation. Previous studies have also shown that although physicians' practice effectively improved, e-alerts alone could not reduce the mortality and the rate of severe AKI<sup>43-46</sup>. Currently, AKI diagnosis depends on changes in serum creatinine. However, novel biomarkers such as neutrophil gelatinase-associated lipocalin (NGAL), kidney injury molecule-1 (KIM-1), Cystatin C, IGFBP7, and osteopontin, as reliable measurement tools for detecting AKI have shown promising results<sup>47-50</sup>. NGAL or KIM-1, reportedly directly released from kidney injury might further provide methods to promptly predict an AKI event and patient prognosis in the early phase<sup>51</sup>. Cystatin C, a molecule with a short half-life in the serum (two hours), is completely filtered at the glomerulus of healthy kidneys, so it might be an ideal surrogate for glomerular filtration rate and tubular cell integrity<sup>52,53</sup>. Due to insufficient data about novel biomarkers on AKI risk prediction models in current studies, the real value of novel biomarkers applied in AI could not be evaluated. Further studies using novel biomarkers as input variables are essential.

Nowadays, e-alerts based on AI were widely used in conjunction with AKI care bundles to construct integrated clinical decision support system (CDS). Is the system truly rational at its current stage? Perhaps not, as the evidence base around clinical decision support system is growing but conflicting<sup>54,55</sup>, but if it can be tied to novel biological markers or even molecular imaging of kidney diseases, it might be.

### Limitations

Despite the promising results, important limitations have to be considered. Firstly, many arguably exaggerated claims exist about AIs equivalence with (or superiority over) clinicians. It is not enough to show good predictive performance on the training set only because most show optimistic results, external validation studies are scarce, and when performed, tend to show reduced accuracy of the studied model<sup>56</sup>. In fact, few AI models have described any clinical effects of their use. Thus, we do not know whether it will improve (or worsen) clinical decisions<sup>57</sup>. Secondly, if a user strongly trusts in the e-alerts of the automatic system, they might present an indolent attitude and wait for AKI alert trigger from the model before taking action. The model requires these actions to dynamically adjust parameters and trigger the alert. This may lead to missed opportunities to mitigate or prevent AKI<sup>58</sup>. Thirdly, most of the enrolled studies were conducted at a single centre, which limits the reproducibility and the generalisability of the results. Fourth, AI entering the field of nephrology must adapt to legal and ethical concerns. The inability to clarify the features used because of a black-box nature conflicts with general data protection requirements<sup>59</sup>. Additionally, used by and serving the interests of private finance, corporations, and start-ups, AI can lead to widening social inequalities, which violates the 'right to health legislation'<sup>60,61</sup>.

## Conclusion

Our review demonstrates the promising performance of artificial intelligence for early prediction of perioperative AKI. Further studies should focus on the improvement of existing models, novel biomarkers, and clinical effectiveness.

## Declarations

## Ethics approval and consent to participate

Not applicable

## Consent for publication

Not applicable

## Availability of data and materials

All data generated or analysed during this study are included in this article

## Competing interests

The authors declare that they have no competing interests.

## Funding

Dr Amanda Y Wang is supported by RACP Jacquot Research Establishment Award, Australia.

Daqing Hong is supported by Sichuan Hemodialysis Quality Control Platform (Project of Sichuan Provincial Department of Science and Technology 2019JDPT0007)

## Authors' Contributions:

The authors' responsibilities were as follows – research idea and study design: H.F.Z. and X.H., data acquisition: H.F.Z. and Y.L.F., statistical analysis/interpretation: H.F.Z., S.K.W., manuscript writing: H.F.Z., Y.F.Z., A.Y.W., and J.N., supervision or mentorship: D.Q.H., A.Y.W., X.W.W., and J.N.. Each author contributed important intellectual content during manuscript drafting or revision and agrees to be personally accountable for the individual's own contributions and to ensure that questions pertaining to the accuracy or integrity of any portion of the work, even one in which the author was not directly involved, are appropriately investigated and resolved, including with documentation in the literature if appropriate.

## Acknowledgements

Not applicable

## References

1. Ronco C, Bellomo R, Kellum JA. Acute kidney injury. *The Lancet*. 2019;394(10212):1949–1964.
2. Bhosale SJ, Kulkarni AP. Preventing Perioperative Acute Kidney Injury. *Indian J Crit Care Med*. 2020;24(Suppl 3):S126-S128.
3. Hobson C, Ruchi R, Bihorac A. Perioperative Acute Kidney Injury: Risk Factors and Predictive Strategies. *Crit Care Clin*. 2017;33(2):379–396.
4. Zarbock A, Koyner JL, Hoste EAJ, Kellum JA. Update on Perioperative Acute Kidney Injury. *Anesth Analg*. 2018;127(5):1236–1245.
5. Bihorac A, Yavas S, Subbiah S, et al. Long-term risk of mortality and acute kidney injury during hospitalization after major surgery. *Annals of surgery*. 2009;249(5):851–858.
6. Group KDIGOAW. KDIGO clinical practice guideline for anemia in chronic kidney disease. *Kidney Int Suppl*. 2012;2(4):279–335.
7. Prowle JR, Liu YI, Fau - Licari E, Licari E, Fau - Bagshaw SM, et al. Oliguria as predictive biomarker of acute kidney injury in critically ill patients. (1466-609X (Electronic)).
8. Hodgson LE, Selby N, Huang TM, Forni LG. The Role of Risk Prediction Models in Prevention and Management of AKI. *Semin Nephrol*. 2019;39(5):421–430.
9. Thottakkara P, Ozrazgat-Baslanti T, Hupf BB, et al. Application of Machine Learning Techniques to High-Dimensional Clinical Data to Forecast Postoperative Complications. *PloS one*. 2016;11(5):e0155705.
10. Chan L, Vaid A, Nadkarni GN. Applications of machine learning methods in kidney disease: hope or hype? *Current opinion in nephrology and hypertension*. 2020;29(3):319–326.
11. Gameiro J, Branco T, Lopes JA. Artificial Intelligence in Acute Kidney Injury Risk Prediction. *J Clin Med*. 2020;9(3).
12. Moons KGM, Wolff RF, Riley RD, et al. PROBAST: A Tool to Assess Risk of Bias and Applicability of Prediction Model Studies: Explanation and Elaboration. *Annals of internal medicine*. 2019;170(1):W1-W33.
13. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: A Tool to Assess the Risk of Bias and Applicability of Prediction Model Studies. *Annals of internal medicine*. 2019;170(1):51–58.
14. Reitsma JB, Glas AS, Rutjes AW, Scholten RJ, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *J Clin Epidemiol*. 2005;58(10):982–990.
15. Deeks JJ, Higgins JP, Altman DG, Group CSM. Analysing data and undertaking meta-analyses. *Cochrane handbook for systematic reviews of interventions*. 2019:241–284.

16. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol.* 2005;58(9):882–893.
17. Adhikari L, Ozrazgat-Baslanti T, Ruppert M, et al. Improved predictive models for acute kidney injury with IDEA: Intraoperative Data Embedded Analytics. *PloS one.* 2019;14(4):e0214904.
18. Bihorac A, Ozrazgat-Baslanti T, Ebadi A, et al. MySurgeryRisk: Development and Validation of a Machine-learning Risk Algorithm for Major Complications and Death After Surgery. *Annals of surgery.* 2019;269(4):652–662.
19. Filiberto AC, Ozrazgat-Baslanti T, Loftus TJ, et al. Optimizing predictive strategies for acute kidney injury after major vascular surgery. *Surgery.* 2021;170(1):298–303.
20. Hofer IS, Lee C, Gabel E, Baldi P, Cannesson M. Development and validation of a deep neural network model to predict postoperative mortality, acute kidney injury, and reintubation using a single feature set. *npj Digital Medicine.* 2020;3(1).
21. Ko S, Jo C, Chang CB, et al. A web-based machine-learning algorithm predicting postoperative acute kidney injury after total knee arthroplasty. *Knee surgery, sports traumatology, arthroscopy: official journal of the ESSKA.* 2020.
22. Lee HC, Yoon HK, Nam K, et al. Derivation and Validation of Machine Learning Approaches to Predict Acute Kidney Injury after Cardiac Surgery. *J Clin Med.* 2018;7(10).
23. Lee HC, Yoon SB, Yang SM, et al. Prediction of Acute Kidney Injury after Liver Transplantation: Machine Learning Approaches vs. Logistic Regression Model. *J Clin Med.* 2018;7(11).
24. 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):3321–3328.
25. Lei VJ, Luong T, Shan E, et al. Risk Stratification for Postoperative Acute Kidney Injury in Major Noncardiac Surgery Using Preoperative and Intraoperative Data. *JAMA network open.* 2019;2(12):e1916921.
26. Li Y, Xu J, Wang Y, et al. A novel machine learning algorithm, Bayesian networks model, to predict the high-risk patients with cardiac surgery-associated acute kidney injury. *Clinical cardiology.* 2020;43(7):752–761.
27. Meyer A, Zverinski D, Pfahringer B, et al. Machine learning for real-time prediction of complications in critical care: a retrospective study. *The Lancet Respiratory Medicine.* 2018;6(12):905–914.
28. Penny-Dimri JC, Bergmeir C, Reid CM, Williams-Spence J, Cochrane AD, Smith JA. Machine Learning Algorithms for Predicting and Risk Profiling of Cardiac Surgery-Associated Acute Kidney Injury. *Seminars in thoracic and cardiovascular surgery.* 2021;33(3):735–745.
29. Rank N, Pfahringer B, Kempfert J, et al. Deep-learning-based real-time prediction of acute kidney injury outperforms human predictive performance. *npj Digital Medicine.* 2020;3(1).
30. Tseng PY, Chen YT, Wang CH, et al. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Critical Care.* 2020;24(1).
31. Xin W, Yi W, Liu H, et al. Early prediction of acute kidney injury after liver transplantation by scoring system and decision tree. *Renal failure.* 2021;43(1):1137–1145.
32. Xue B, Li D, Lu C, et al. Use of Machine Learning to Develop and Evaluate Models Using Preoperative and Intraoperative Data to Identify Risks of Postoperative Complications. *JAMA network open.* 2021;4(3):e212240.
33. Yayac M, Aman ZS, Rondon AJ, Tan TL, Courtney PM, Purtill JJ. Risk Factors and Effect of Acute Kidney Injury on Outcomes Following Total Hip and Knee Arthroplasty. *The Journal of arthroplasty.* 2021;36(1):331–338.
34. Zhang Y, Yang D, Liu Z, et al. An explainable supervised machine learning predictor of acute kidney injury after adult deceased donor liver transplantation. *Journal of translational medicine.* 2021;19(1).
35. Zhou C, Wang R, Jiang W, et al. Machine learning for the prediction of acute kidney injury and paraplegia after thoracoabdominal aortic aneurysm repair. *Journal of cardiac surgery.* 2020;35(1):89–99.
36. Yang L, Xing G, Wang L, et al. Acute kidney injury in China: a cross-sectional survey. *The Lancet.* 2015;386(10002):1465–1471.
37. Cox ZL, McCoy AB, Matheny ME, et al. Adverse drug events during AKI and its recovery. *Clinical journal of the American Society of Nephrology: CJASN.* 2013;8(7):1070–1078.
38. Herrera-Gutierrez ME, Seller-Perez G, Sanchez-Izquierdo-Riera JA, Maynar-Moliner J, group Ci. Prevalence of acute kidney injury in intensive care units: the "Corte de prevalencia de disFuncion RenAl y DEpuracion en criticos" point-prevalence multicenter study. *J Crit Care.* 2013;28(5):687–694.
39. Kellum JA, Bihorac A. Artificial intelligence to predict AKI: is it a breakthrough? *Nature reviews Nephrology.* 2019;15(11):663–664.
40. Tomasev N, Glorot X, Rae JW, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature.* 2019;572(7767):116–119.
41. Van Biesen W, Vanmassenhove J, Decruyenaere J. Prediction of acute kidney injury using artificial intelligence: are we there yet? *Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association.* 2020;35(2):204–205.
42. Gumbert SD, Kork F, Jackson ML, et al. Perioperative Acute Kidney Injury. *Anesthesiology.* 2020;132(1):180–204.
43. Lachance P, Villeneuve PM, Rewa OG, et al. Association between e-alert implementation for detection of acute kidney injury and outcomes: a systematic review. *Nephrology, dialysis, transplantation: official publication of the European Dialysis and Transplant Association - European Renal Association.* 2017;32(2):265–272.

44. Rind DM, Safran C, Fau - Phillips RS, Phillips Rs Fau - Wang Q, et al. Effect of computer-based alerts on the treatment and outcomes of hospitalized patients. (0003-9926 (Print)).
45. Wilson FP, Shashaty M, Testani J, et al. Automated, electronic alerts for acute kidney injury: a single-blind, parallel-group, randomised controlled trial. *The Lancet*. 2015;385(9981):1966–1974.
46. Colpaert K, Hoste EA, Steurbaut K, et al. Impact of real-time electronic alerting of acute kidney injury on therapeutic intervention and progression of RIFLE class. *Critical care medicine*. 2012;40(4):1164–1170.
47. Flechet M, Guiza F, Schetz M, et al. AKIpredictor, an online prognostic calculator for acute kidney injury in adult critically ill patients: development, validation and comparison to serum neutrophil gelatinase-associated lipocalin. *Intensive Care Med*. 2017;43(6):764–773.
48. Grieshaber P, Möller S, Arneith B, et al. Predicting Cardiac Surgery-Associated Acute Kidney Injury Using a Combination of Clinical Risk Scores and Urinary Biomarkers. *The Thoracic and cardiovascular surgeon*. 2020;68(5):389–400.
49. Ibrahim NE, McCarthy CP, Shrestha S, et al. A clinical, proteomics, and artificial intelligence-driven model to predict acute kidney injury in patients undergoing coronary angiography. *Clinical cardiology*. 2019;42(2):292–298.
50. Wang JJ, Chi NH, Huang TM, et al. Urinary biomarkers predict advanced acute kidney injury after cardiovascular surgery. *Critical care (London, England)*. 2018;22(1):108.
51. Park S, Lee H. Acute kidney injury prediction models: current concepts and future strategies. *Current opinion in nephrology and hypertension*. 2019;28(6):552–559.
52. Mazul-Sunko B, Zarkovic N, Vrkic N, et al. Proatrial natriuretic peptide (1-98), but not cystatin C, is predictive for occurrence of acute renal insufficiency in critically ill septic patients. *Nephron Clin Pract*. 2004;97(3):c103-107.
53. Villa P, Jimenez M, Soriano MC, Manzanares J, Casasnovas P. Serum cystatin C concentration as a marker of acute renal dysfunction in critically ill patients. *Critical care (London, England)*. 2005;9(2):R139-143.
54. Kashani KB. Automated acute kidney injury alerts. *Kidney international*. 2018;94(3):484–490.
55. Zhao Y, Zheng X, Wang J, et al. Effect of clinical decision support systems on clinical outcome for acute kidney injury: a systematic review and meta-analysis. *BMC Nephrol*. 2021;22(1):271.
56. Reilly BM, Evans AT. Translating clinical research into clinical practice: impact of using prediction rules to make decisions. *Annals of internal medicine*. 2006;144(3):201–209.
57. Laupacis A, Sekar N, Stiell IG. Clinical prediction rules. A review and suggested modifications of methodological standards. *Jama*. 1997;277(6):488–494.
58. Bastin AJ, Ostermann M, Slack AJ, Diller GP, Finney SJ, Evans TW. Acute kidney injury after cardiac surgery according to Risk/Injury/Failure/Loss/End-stage, Acute Kidney Injury Network, and Kidney Disease: Improving Global Outcomes classifications. *J Crit Care*. 2013;28(4):389–396.
59. Wilson FP. Machine Learning to Predict Acute Kidney Injury. *Am J Kidney Dis*. 2020;75(6):965–967.
60. Fukuda-Parr S, Gibbons E. Emerging Consensus on 'Ethical AI': Human Rights Critique of Stakeholder Guidelines. *Global Policy*. 2021;12(S6):32–44.
61. Human Rights BDaTPH. <https://www.hrbdt.ac.uk/health/>. Published Identifying opportunities and threats to the right to health in a new data-driven economy. Accessed.

## Tables

**Table 1. Clinical characteristics of the included studies.**

Author, year	Number of patients	External validation	Type of surgery	AKI definition	Age (y), mean±SD/median(range)	Male (%)	Model type	Predictors	TP
Adhikari,2019	2,911	No	Any type of inpatient operative procedure	KDIGO	60(49, 69)	60	RF	69	942
Bihorac,2019	51,457	No	Any type of inpatient operative procedure	KDIGO	NR	NR	MySurgeryRisk	16	16020
Filiberto,2021	1,631	No	Cardiovascular surgery	KDIGO	68(59, 75)	66	RF	367	503
Ko,2020	455	Yes	Total joint arthroplasty	KDIGO	71±6	16	GBM	6	13
Lee(1),2018	363	No	Liver transplantation	AKIN	53(48,60)	68	GBM	20	91
Lee(2),2018	1,005	No	Cardiovascular surgery	KDIGO	64(55,71)	73	GBM	72	251
Lei,2019	8,494	No	Noncardiac surgery	KDIGO stage I	58±16	54	GBM	339	696
Lei,2020	270	No	Cardiovascular surgery	KDIGO	48±10	74	GBM	20	125
Li,2020	1,894	Yes	Cardiovascular surgery	KDIGO	56±13	58	BN	12	469
Meyer,2018	5,898	Yes	Cardiovascular surgery	KDIGO stage III	68(59, 76)	69	ANN	52	117
Penny-Dimri,2020	96,653	No	Cardiovascular surgery	Self-defined+RRT	NR	73	GBM	56	3357
Rank,2020	350	No	Cardiovascular surgery	KDIGO stage I or II	69±14	67	ANN	96	149
Tseng,2020	202	No	Cardiovascular surgery	KDIGO	63(53,71)	65	Ensemble	94	44
Xin,2021	109	No	Liver transplantation	KDIGO	54±9	83	DT	NR	45
Xue,2021	106870	No	Any type of inpatient operative procedure	KDIGO	NR	NR	GBM	711	5222
Yayac,2021	20800	No	Total joint arthroplasty	KDIGO	66±11	55	GBM	41	606
Zhang,2021	195	Yes	Liver transplantation	KDIGO	47±10	88	GBM	111	80
Zhou,2020	212	No	Cardiovascular surgery	RRT	37±10	70	RF	7	20

ANN = artificial neural network; GBM = gradient boosting machine; RF = random forest; BN = Bayesian network; DT = decision tree; RRT= renal replacement therapy; FP = false positive; FN = false negative; TN = true negative; NR = not reported.

**Table 2. Summary of diagnostic test accuracy and subgroup analysis of the included studies**

Subgroup	Number of included studies	Sensitivity (95% CI)	Specificity (95% CI)	PLR	NRL	DOR
<b>Type of AI</b>						
GBM	9	0.77 (0.76-0.78)	0.69 (0.69-0.69)	2.7 (2.4-3.0)	0.34 (0.29-0.41)	7.8 (6.1-10)
RF	3	0.82 (0.80-0.84)	0.74 (0.72-0.76)	3.5 (1.9-6.4)	0.25 (0.22-0.27)	13 (6.5-26)
ANN	3	0.62 (0.59-0.64)	0.87 (0.86-0.87)	4.9 (4.0-6.0)	0.29 (0.14-0.60)	16 (7.8-34)
<b>Number of patients</b>						
<1000	8	0.79 (0.76-0.82)	0.77 (0.75-0.79)	3.4 (2.6-4.3)	0.25 (0.17-0.36)	14 (9.0-21)
≥1000	11	0.78 (0.78-0.79)	0.71 (0.71-0.71)	3.1 (2.7-3.7)	0.33 (0.28-0.39)	9.6 (7.3-13)
<b>Type of surgery</b>						
Cardiovascular surgery	9	0.73 (0.72-0.74)	0.71 (0.71-0.71)	3.4 (2.7-4.4)	0.33 (0.28-0.38)	11 (8.0-15)
Any type of inpatient operative procedure	4	0.79 (0.78-0.80)	0.73 (0.73-0.73)	3.7 (2.8-5.0)	0.31 (0.23-0.41)	12 (9.0-17)
Liver transplantation	3	0.82 (0.77-0.87)	0.73 (0.69-0.78)	2.7 (1.6-4.6)	0.26 (0.20-0.34)	11 (4.9-23)
Total joint arthroplasty	2	0.75 (0.72-0.78)	0.60 (0.60-0.61)	2.8 (1.2-6.3)	0.27 (0.07-1.01)	11 (1.2-110)
<b>Methodological quality</b>						
Low quality	13	0.73 (0.72-0.74)	0.72 (0.72-0.72)	3.4 (2.7-4.2)	0.32 (0.26-0.38)	11 (8.2-15)
Unclear quality	2	0.72 (0.70-0.75)	0.82 (0.80-0.84)	3.9 (3.5-4.4)	0.27 (0.14-0.54)	15 (6.8-32)
High quality	4	0.80 (0.80-0.80)	0.71 (0.71-0.71)	2.6 (2.0-3.5)	0.30 (0.26-0.35)	8.6 (5.6-13)
<b>External validation</b>						
No	15	0.78 (0.78-0.79)	0.71 (0.71-0.71)	3.1 (2.7-3.6)	0.31 (0.26-0.36)	10 (8.0-13)
Yes	4	0.72 (0.69-0.75)	0.85 (0.84-0.85)	3.7 (2.5-5.6)	0.30 (0.22-0.42)	13 (7.0-24)
<b>AKI definition</b>						
KDIGO	14	0.80 (0.79-0.80)	0.71 (0.71-0.71)	2.9 (2.5-3.5)	0.30 (0.27-0.34)	10 (7.8-13)
Self-defined	3	0.73 (0.72-0.74)	0.71 (0.71-0.71)	4.1 (2.1-8.1)	0.32 (0.22-0.45)	13 (4.7-37)
AKIN	2	0.60 (0.55-0.61)	0.88 (0.87-0.89)	4.6 (4.1-5.1)	0.34 (0.15-0.80)	13 (5.8-29)

ANN = artificial neural network; GBM = gradient boosting machine; RF = random forest; KDIGO = Kidney Disease: Improving Global Outcomes; AKIN = Acute Kidney Injury Network

## Figures

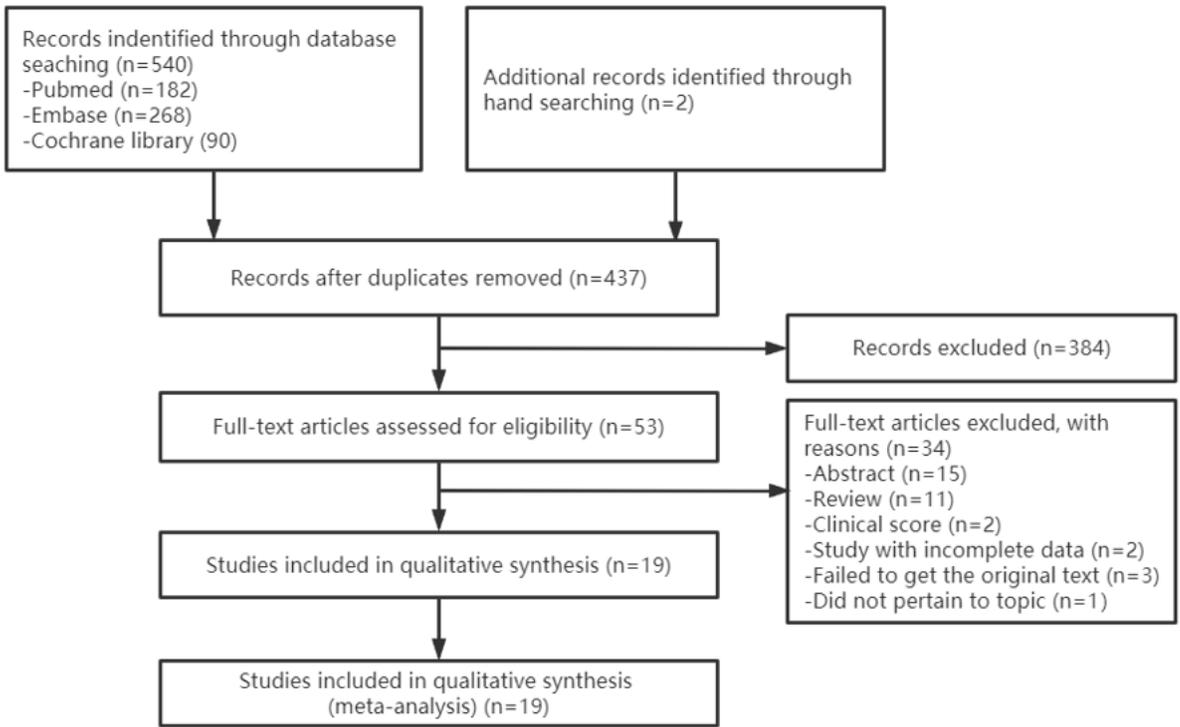


Figure 1

Flow diagram of the identification of relevant studies

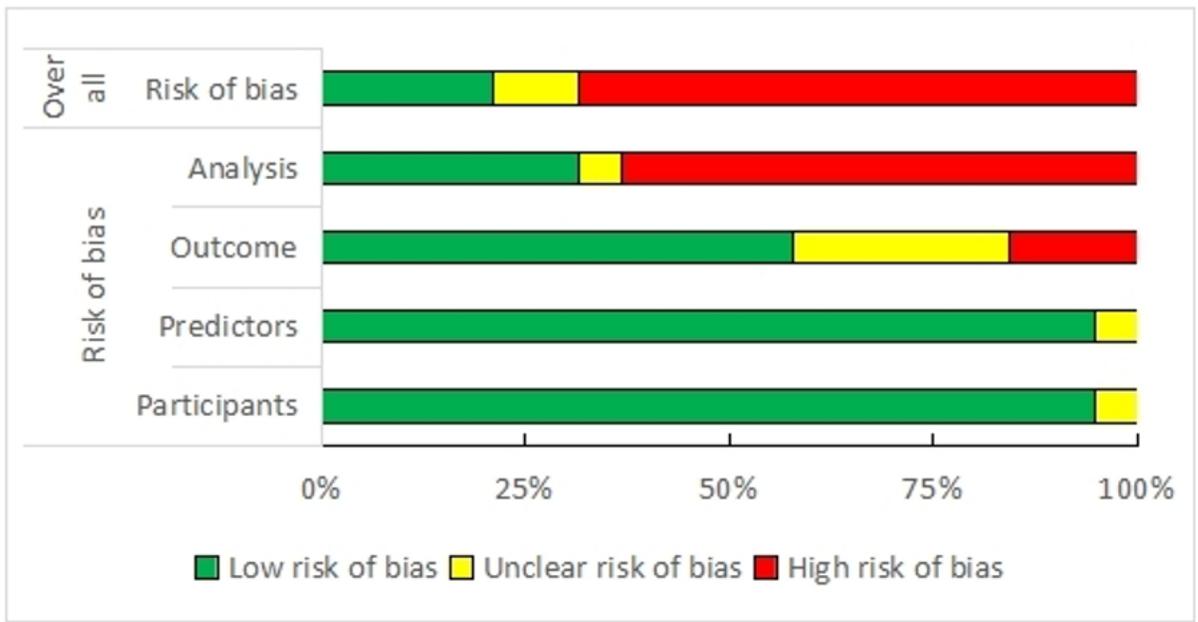


Figure 2

Risk of bias assessment (using PROBAST) based on four domains

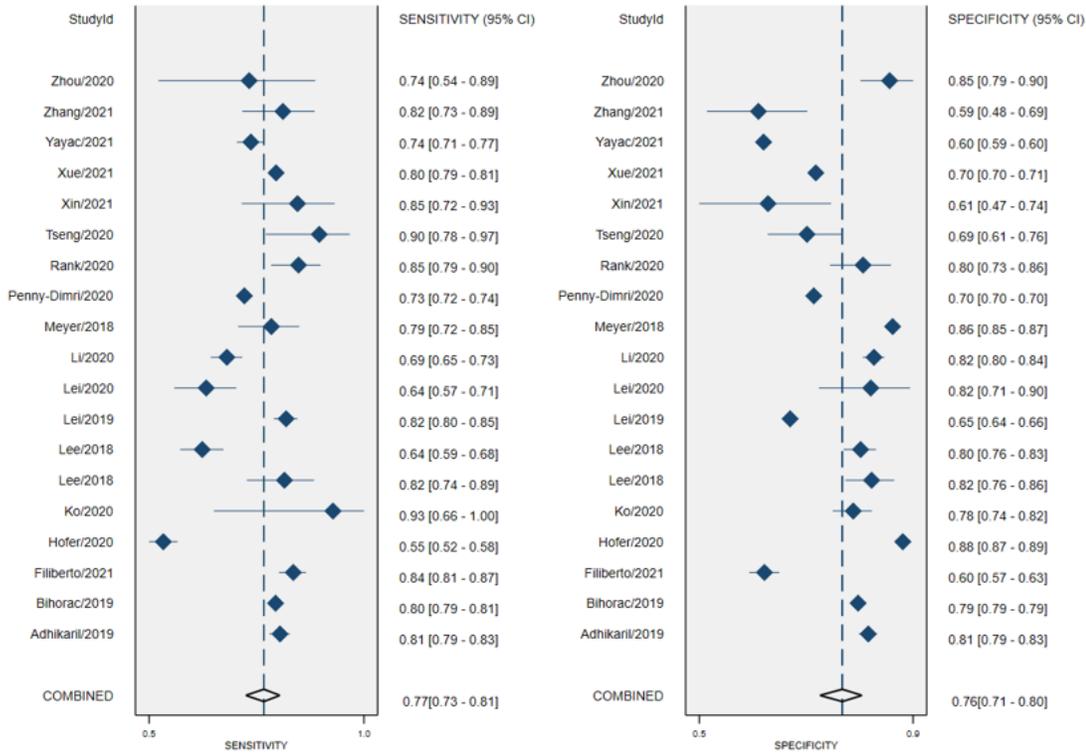


Figure 3

Forest plots of sensitivity and specificity of artificial intelligence algorithm for the prediction of Acute Kidney Injury during the perioperative period.

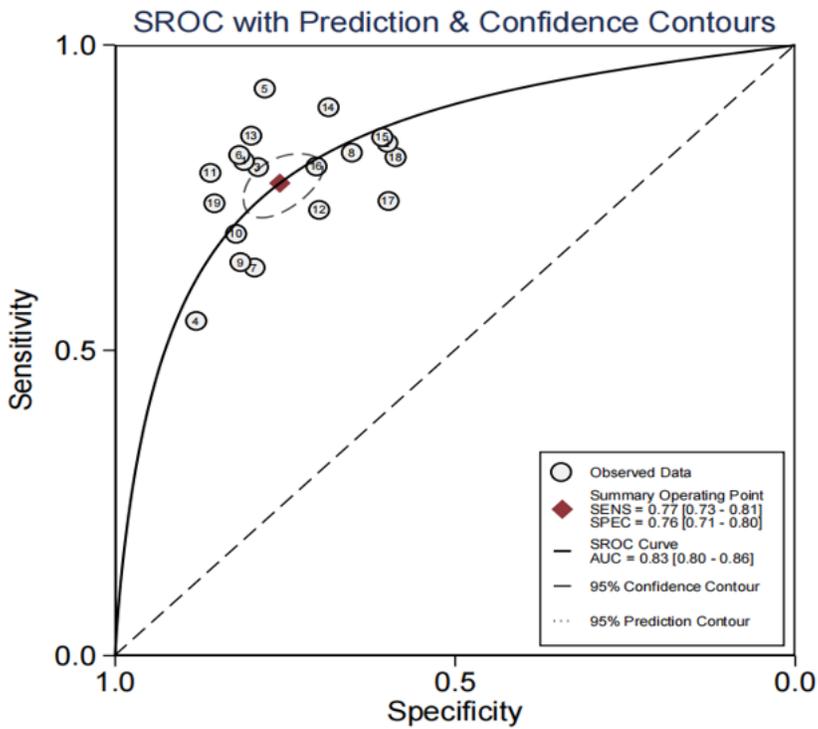


Figure 4

Summary receiver operating characteristic curve with 95% confidence region for the prediction of AKI during the perioperative period.

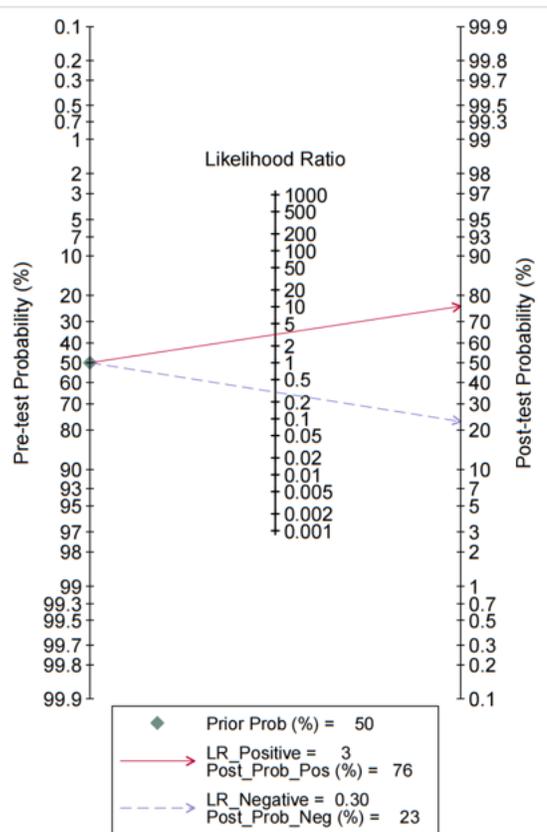


Figure 5

Fagan nomogram for the prediction of AKI during the perioperative period.

Figure 6

Meta-regression for the reason of heterogeneity in the diagnostic test accuracy meta-analysis. Nopt<sup>2</sup> number of patients.

Figure 7

Sensitivity analysis for the prediction of AKI during the perioperative period.

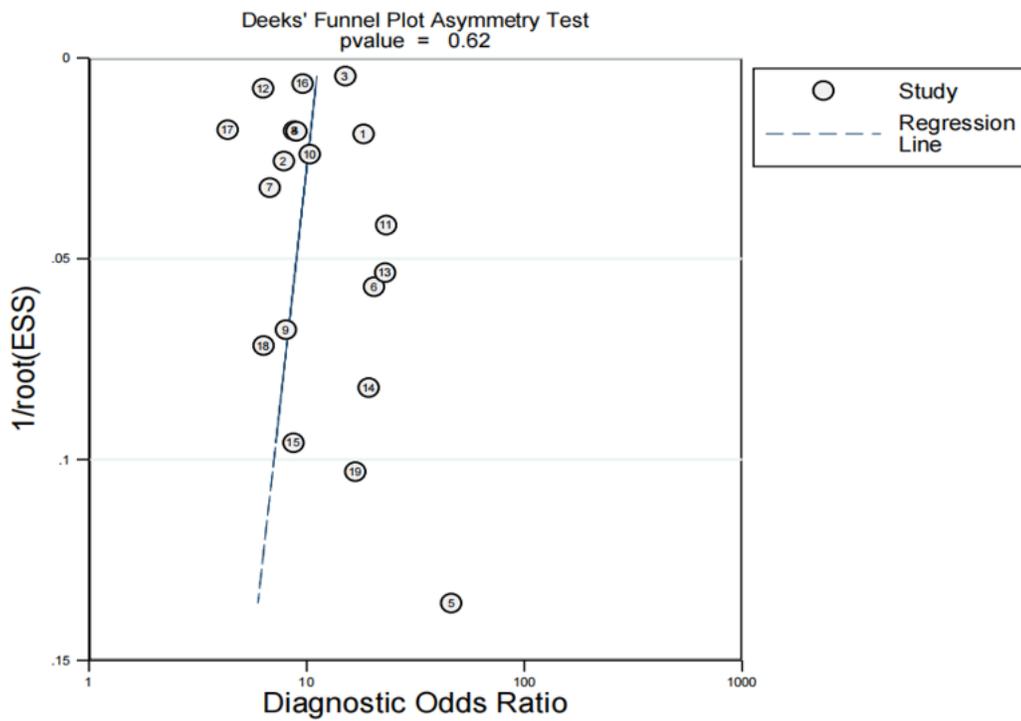


Figure 8

Deek's funnel plot for the prediction of AKI during the perioperative period.

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Appendix1.docx](#)