

Predictive Performance and Clinical Utility of Artificial Intelligence Models for Hospital-Acquired Acute Kidney Injury: A Systematic Review and Meta-Analysis

Roshan Rajesh Menon¹, Shobhit Sharma², Urja Nikhil Parikh³, Siddhanth Santhosh⁴, Bridget Achiaa Quansah⁵, Cecilia Agada⁶, Waqas Alauddin^{7*}

¹Junior Doctor, Dept of Orthopedics, Dr Sangtani's Ortho Relief Hospital & Research Centre, Nagpur, Maharashtra, India

²Medical Officer, Department of Medicine, Railway Health Unit, Ludhiana, Punjab, India

³General Physician, Grodno State Medical University, Grodno, Belarus

⁴Doctor, Kempegowda Institute of Medical Sciences, Bangalore, Karnataka, India

⁵Student (Masters in Public Health), University of New South Wales, Sydney, New South Wales, Australia

⁶MS4 (Fourth-Yr Medical Student) Windsor University School of Medicine, St. Kitts & Windsor Institute of Research & Education, Chicago, Illinois, United States

⁷Associate Professor, Department of Physiology, Venkateshwara Institute of Medical Sciences, Amroha, India

***Address for Correspondence:** Dr. Waqas Alauddin, Associate Professor, Department of Physiology, Venkateshwara Institute of Medical Sciences, Amroha, India

E-mail: dr.waqas7@gmail.com

Received: 29 Jun 2026/ Revised: 17 Jul 2026/ Accepted: 20 Aug 2026

ABSTRACT

Background: Hospital-acquired acute kidney injury (HA-AKI) is associated with substantial morbidity and mortality. Artificial intelligence (AI) and machine learning (ML) may enable earlier prediction of HA-AKI, but reported performance varies across studies.

Methods: A systematic review and meta-analysis was conducted according to PRISMA 2020 guidelines. PubMed, Scopus, Web of Science, EMBASE, CINAHL, CENTRAL, EBSCOhost, and Google Scholar were searched from January 2000 to August 2026. Studies evaluating AI/ML models for predicting HA-AKI in hospitalized adults were eligible. The primary outcome was discriminatory performance measured by AUROC/AUC. Random-effects meta-analysis was performed, with heterogeneity assessed using I^2 and Cochran's Q. Subgroup analysis compared deep-learning with tree-based/classical ML models.

Results: Twelve studies were included. The pooled AUROC was 0.847 (95% CI: 0.82–0.87), indicating good discrimination. Heterogeneity was substantial ($I^2 = 96.9\%$; $p < 0.001$), with a 95% prediction interval of 0.74–0.92. Deep-learning models had a pooled AUROC of 0.862 (95% CI: 0.82–0.89), compared with 0.837 (95% CI: 0.79–0.87) for tree-based/classical ML models; the difference was not significant ($p = 0.34$).

Conclusion: AI/ML models demonstrate good predictive performance for HA-AKI, but substantial heterogeneity indicates variable performance across clinical settings. External and prospective validation, calibration, and clinical utility assessment are needed before widespread implementation.

Key-words: Hospital-acquired acute kidney injury; Artificial intelligence; Machine learning; predictive models; AUROC; Deep learning; Clinical prediction

How to cite this article

Menon RR, Sharma S, Parikh UN, Santhosh S, Quansah BA. Predictive Performance and Clinical Utility of Artificial Intelligence Models for Hospital-Acquired Acute Kidney Injury: A Systematic Review and Meta-Analysis. SSR Inst Int J Life Sci., 2026; 12(5): 10910-10919.



Access this article online

<https://ijls.com/>

INTRODUCTION

Acute kidney injury (AKI) is a common and clinically important complication among hospitalized patients and is associated with increased mortality, prolonged hospitalization, healthcare utilization, and subsequent chronic kidney disease. Approximately one in five hospitalized adults develops AKI, emphasizing the importance of early identification and prevention. ^[1,2] HA-AKI is particularly important because it develops after



admission and may provide an opportunity for early risk identification and intervention. ^[1,2]

AKI is primarily diagnosed using changes in serum creatinine and urine output according to established criteria such as Kidney Disease: Improving Global Outcomes (KDIGO). ^[3] However, serum creatinine is a delayed marker of kidney injury. Early prediction of patients at risk of HA-AKI could therefore facilitate timely monitoring, avoidance of nephrotoxic exposures, and optimization of hemodynamic status. ^[2,3] Conventional AKI prediction models use demographic, clinical, laboratory, and physiological variables. However, the multifactorial nature of AKI and the increasing availability of longitudinal electronic health record (EHR) data may limit their ability to capture complex relationships. AI and ML can analyze large, heterogeneous datasets and may enable more dynamic and individualized risk prediction. ^[4,5]

Several studies have demonstrated promising performance of AI/ML models for AKI prediction. Koyner *et al.* reported AUCs of 0.90 and 0.87 for predicting stage 2 AKI at 24 and 48 hours, respectively. ^[6] Tomašev *et al.* demonstrated that a continuous deep-learning model could predict future AKI using longitudinal clinical data, including dialysis-requiring episodes. ^[7] Other studies have evaluated recurrent neural networks, gradient boosting, random forests, convolutional neural networks, and ensemble approaches, with variable performance across populations and healthcare systems. ^[8-11] Despite these advances, the generalizability and clinical applicability of AI-based AKI prediction models remain uncertain. A systematic review of 46 ML-based AKI studies reported AUCs ranging from 0.49 to 0.99 and identified concerns regarding risk of bias, limited external validation, and inadequate calibration assessment. ^[4] Similarly, evidence from externally validated models has highlighted limited population diversity and challenges in transportability across healthcare settings. ^[5] Recent evidence also indicates differences between internal and external validation performance, emphasizing the importance of independent validation. ^[12]

Although previous reviews have evaluated AI/ML approaches for AKI prediction, substantial variation remains in populations, AKI definitions, prediction horizons, algorithms, predictors, and validation strategies. Quantitative evidence specifically addressing

HA-AKI and comparing predictive performance across model families remains limited. Therefore, this systematic review and meta-analysis aimed to evaluate the predictive performance and clinical utility of AI/ML models for HA-AKI in adults. The primary outcome was pooled AUROC/AUC, with sensitivity, specificity, accuracy, predictive values, calibration, and clinical utility assessed where available. We also evaluated between-study heterogeneity and compared deep-learning with tree-based/classical ML approaches.

MATERIALS AND METHODS

This systematic review and meta-analysis was conducted in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 guidelines and designed following PRISMA-P recommendations. ^[13] The primary objective was to assess, synthesize, and compare the predictive performance and clinical utility of (AI)-based models for HA-AKI. The research question used the Population, Index test, Comparator, and Outcome (PICO) framework. The Population comprised adults (≥ 18 years) hospitalized in ICU or other inpatient settings who were at risk of or evaluated for HA-AKI. The Index test included AI/machine learning-based models developed or validated for HA-AKI prediction or early detection using routinely available clinical, demographic, laboratory, physiological, or electronic health record data. Comparators included conventional risk assessments, established prediction models, baseline statistical models, or usual care, where available. Outcomes included model discrimination, calibration, and clinical utility, assessed using AUROC, sensitivity, specificity, accuracy, predictive values, calibration metrics, decision-curve analysis, and net benefit. We also considered variations in patient populations, clinical settings, AKI definitions, prediction horizons, model characteristics, predictors, validation methods, and study methodology.

Literature Search- A comprehensive literature search was conducted in PubMed/MEDLINE, Scopus, Web of Science, EBSCOhost, CENTRAL, CINAHL, and EMBASE, with Google Scholar searched additionally for supplementary studies. Studies published from January 2000 to August 2026 were considered. The search targeted studies evaluating AI/machine learning (ML)-based models for predicting or detecting HA-AKI in

adults. Keywords, MeSH terms, controlled vocabulary, and Boolean operators (“AND” and “OR”) were used to combine terms related to hospitalized patients, AI/ML predictive models, HA-AKI, prediction, and clinical utility. Search strategies were adapted for each database, with complete strategies provided in the Supplementary Material. Additional sources included ClinicalTrials.gov, WHO ICTRP, ISRCTN, OSF Registries, relevant grey literature, and PROSPERO. Backward and forward citation screening of relevant studies and reviews was also performed to identify additional eligible studies. [14-16]

Inclusion and Exclusion Criteria- Studies were eligible if they were original research articles published from January 2000 to August 2026 evaluating AI/ML models for prediction or early detection of HA-AKI in hospitalized adults (≥ 18 years), including ICU, medical-surgical, emergency, and other inpatient settings. Models using clinical, demographic, physiological, laboratory, medication-related, or electronic health record data were included. Eligible algorithms included logistic regression within predictive/ML frameworks, decision trees, random forests, support vector machines, neural networks, deep learning, gradient boosting, XGBoost, ensemble methods, and other supervised ML approaches.

Studies were required to define HA-AKI using established or clearly specified criteria (e.g., KDIGO, AKIN, or RIFLE) and a defined prediction horizon. Outcomes included predictive discrimination, calibration, and clinical utility, assessed using AUROC/AUC, sensitivity, specificity, accuracy, predictive values, calibration measures, Brier score, decision-curve analysis, or net benefit. Only English-language studies were included.

We excluded reviews, meta-analyses, scoping/narrative reviews, editorials, conference abstracts without full text, case reports, protocols, letters, dissertations, and other non-original publications. Pediatric studies; studies without HA-AKI prediction or AI/ML modeling; studies focused exclusively on CKD, community-acquired AKI, dialysis-dependent kidney failure, or non-AKI outcomes; risk-factor studies without predictive modeling; non-human or simulated datasets; and methodological studies without clinical evaluation were also excluded. Duplicate publications were removed, and overlapping cohorts were managed to avoid double-counting.

Data Extraction- Two reviewers independently extracted data using a standardized, prespecified form. Extracted data included study characteristics, participant demographics, clinical setting, data source, HA-AKI definition, prediction timing and horizon, and AI/ML model characteristics, including algorithms, predictors, preprocessing, and validation methods. Performance measures included AUROC/AUC, sensitivity, specificity, accuracy, predictive values, calibration, Brier score, decision-curve analysis, net benefit, and corresponding 95% confidence intervals or other uncertainty measures. Methodological features, including model development, validation, overfitting, data leakage, and class imbalance, were also recorded. Potential sources of heterogeneity were documented, including population, setting, AKI definition, prediction horizon, data source, algorithm, predictors, and validation strategy. Multiple models were extracted separately while avoiding double-counting, with preference given to external or independent validation estimates most closely matching the prespecified outcome and setting. Discrepancies were resolved by consensus or a third reviewer, and authors were contacted when essential information was unclear or unavailable.

Data Synthesis- A quantitative meta-analysis was performed to evaluate the predictive performance of AI/ML models for HA-AKI. The primary outcome was AUROC/AUC, with 95% confidence intervals (CIs), standard errors, or other measures of precision extracted where available. Secondary measures included sensitivity, specificity, accuracy, predictive values, F1 score, calibration, Brier score, and clinical utility. Pooled estimates with 95% CIs were presented using forest plots. Random-effects models were used to account for anticipated clinical, methodological, and statistical heterogeneity. Heterogeneity was assessed using Cochran’s Q and I^2 statistics. Subgroup analyses were performed, where feasible, according to clinical setting, population, AKI definition, prediction horizon, algorithm, and validation strategy. When multiple models were reported, they were extracted separately while avoiding double-counting, with preference given to primary or externally/independently validated models aligned with the prespecified outcome and setting. Where quantitative pooling was inappropriate, findings were synthesized narratively.

Small-study effects and publication bias were assessed when sufficient studies were available. PROBABST+AI assessments informed interpretation and sensitivity analyses, including exclusion of studies with high risk of bias, applicability concerns, unclear HA-AKI definitions, or lack of independent/external validation, where feasible.

Ethical Considerations- As this review used previously published data, ethical approval and informed consent were not required. The review followed PRISMA 2020 guidelines to promote methodological rigor and transparency. The review protocol was registered prospectively with PROSPERO (CRD420261507962).

RESULTS

A total of 1,086 records were identified, including 1,048 from electronic databases and 38 from other sources. Before screening, 800 duplicates and 59 records removed for other reasons were excluded, leaving 189 records for title and abstract screening. Of these, 146 were excluded, and 43 reports were sought for retrieval. Sixteen could not be retrieved, leaving 27 reports for full-

text assessment. Fifteen were excluded because they were unrelated to HA-AKI ($n = 6$), did not use AI/ML models for AKI prediction ($n = 5$), or were not prediction/prognostic studies ($n = 4$). The 38 records identified through other sources included OpenGrey.eu ($n = 4$), WorldCat.org ($n = 7$), ClinicalTrials.gov ($n = 2$), WHO ICTRP ($n = 9$), PROSPERO ($n = 5$), OSF Registries ($n = 3$), and ISRCTN Registry ($n = 8$). Of these, 28 could not be retrieved, while 10 underwent eligibility assessment. All 10 were excluded because the population did not meet eligibility criteria ($n = 6$), insufficient data were available to assess predictive performance ($n = 2$), or the records were conference abstracts, protocols, reviews, or editorials ($n = 2$). Ultimately, 12 studies (12 reports) were included in the systematic review and meta-analysis, as shown in Fig. 1.

The diagram was generated using the PRISMA 2020 Flow Diagram web application developed by Haddaway NR, McGuinness LA, and colleagues (United Kingdom). No automated tools were used in the screening or selection of studies.

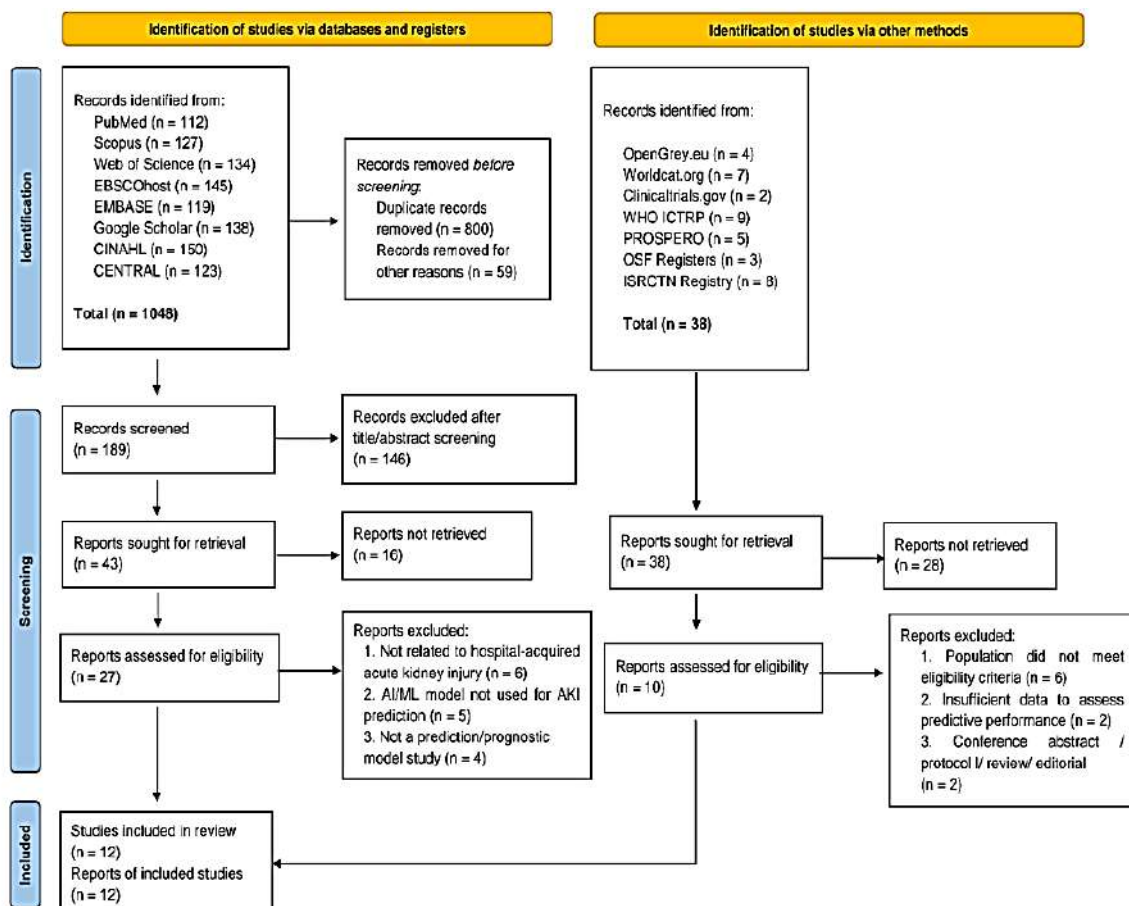


Fig. 1: PRISMA flow diagram of study selection process.

The PRISMA-guided selection process identified 12 studies for qualitative and quantitative synthesis (Fig. 1). All were observational prediction-model studies, mainly retrospective or prospective cohorts, evaluating AI/ML models for HA-AKI in hospitalized adults across medical-surgical wards, ICUs, emergency/critical care, and specialized settings.

HA-AKI was defined using KDIGO, AKIN, RIFLE, or other specified criteria, with varying prediction horizons and clinical, laboratory, physiological, medication-related, or EHR predictors. Models included supervised ML, ensemble methods, neural networks, deep learning, gradient boosting, and random forests. Performance was primarily assessed using AUROC/AUC, with additional measures including sensitivity, specificity, accuracy, predictive values, F1 score, calibration, Brier score, and clinical utility where reported. Substantial variation was observed in populations, settings, sample sizes, AKI

definitions, predictors, algorithms, prediction horizons, and validation strategies.

The 12 studies were published between 2018 and 2025, mainly in the United States, with additional studies from South Korea, Singapore, and China. Most used retrospective EHR-based cohorts, with sample sizes ranging from approximately 4,104 to 703,782 patients or admissions. AI/ML approaches included gradient boosting, XGBoost, random forests, support vector machines, neural networks, and ensemble models is summarized in Table 1. Prediction horizons generally ranged from 24 to 72 hours before AKI onset. AUROC/AUC was the primary performance measure, with some studies also reporting calibration, external validation, interpretability, decision-curve analysis, alert burden, and clinical utility. Overall, models demonstrated moderate-to-excellent discrimination, with performance varying across populations and healthcare settings.

Table 1: Characteristics of Included Studies

Author (yr)	Country	Study design	Sample size	Primary focus	Key findings
Koyner <i>et al.</i> ^[6]	USA	Retrospective ML development/validation	121,158	Gradient boosting	AUC 0.90 (24 h) and 0.87 (48 h).
He <i>et al.</i> ^[17]	USA	Retrospective cohort; comparative ML	76,957	Five ML models	AUC 0.720–0.764.
Tomašev <i>et al.</i> ^[7]	USA/UK	Retrospective EHR deep-learning study	703,782	Deep learning	Predicted 55.8% of AKI and 90.2% of dialysis-requiring AKI.
Churpek <i>et al.</i> ^[8]	USA	Multicenter retrospective validation	495,971 admissions	Gradient boosting	AUC 0.85–0.86 for stage 2 AKI.
Song <i>et al.</i> ^[18]	USA	Multicenter retrospective development/external validation	153,821 encounters	AI model	Performance varied across sites; SHAP improved interpretability.
Kim <i>et al.</i> ^[9]	South Korea	Retrospective multicenter development/external validation	149,108	RNN	Externally validated real-time AKI prediction.
Chua <i>et al.</i> ^[11]	Singapore	Retrospective longitudinal ML study	20,732 admissions	ML/EHR	Predicted HA-AKI ≥48 h before onset.
Le <i>et al.</i> ^[10]	USA	Retrospective cohort; ML development/validation	12,347	CNN	AUROC 0.86; outperformed XGBoost and SOFA.
Gong <i>et al.</i> ^[19]	USA	Retrospective ML development/interpretability study	ICU cohort	Ensemble ML	Good discrimination; interpretable predictors.
Lee <i>et al.</i> ^[20]	South Korea	Retrospective comparative ML study	4,104	SVM/RF/XGBoost/LightGBM	LightGBM AUC 0.810; greater net benefit than logistic regression.
Griffin ^[21]	USA	Retrospective cohort; RNN/GRU	14,480 patients;	GRU/RNN	Precision 0.60; reduced false alerts.

		development/testing	18,180 admissions		
Yu <i>et al.</i> [22]	China	Retrospective ML development/external validation	59,936	XGBoost/SVM/RF	XGBoost AUC 0.930–0.936 internally; 0.812–0.833 externally.

In Fig. 2, the pooled AUROC was 0.847 (95% CI: 0.82–0.87) across 12 studies, indicating good discrimination. Heterogeneity was substantial ($I^2 = 96.9\%$, $\tau^2 = 0.110$, $Q = 352.5$, $df = 11$, $p < 0.001$). The 95% prediction interval

(0.74–0.92) indicates considerable variability in performance across new settings, suggesting that the pooled estimate may overstate expected performance in individual hospitals.

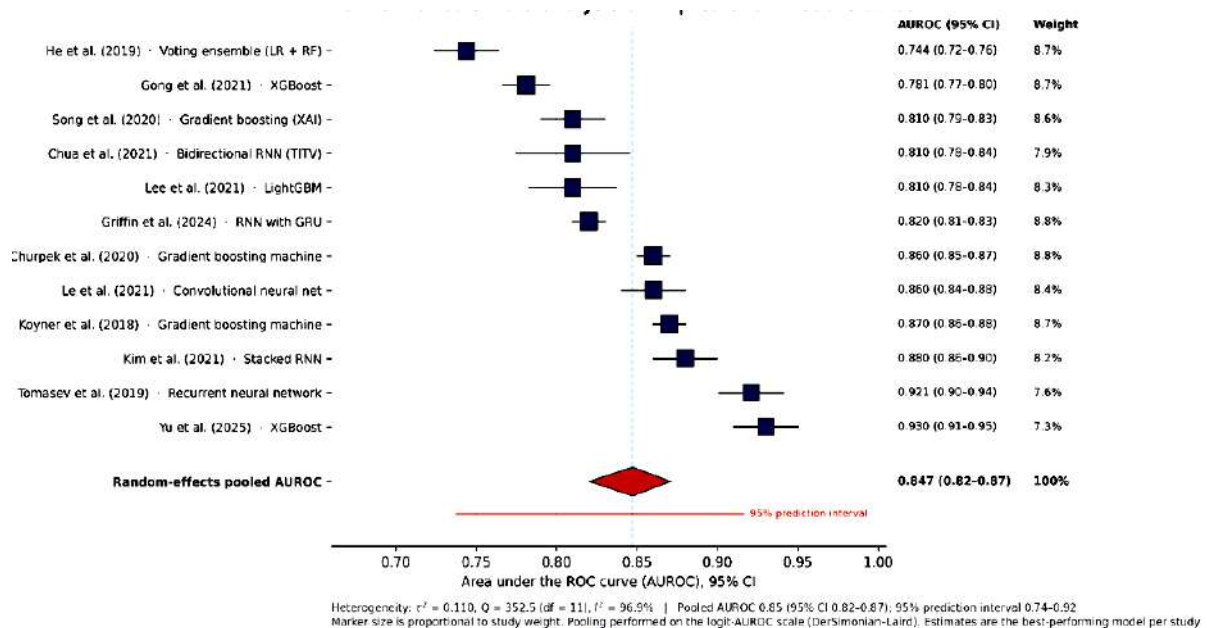


Fig. 2: Overall Forest Plot

In Fig. 3, Deep-learning models had a pooled AUROC of 0.862 (95% CI: 0.82–0.89), compared with 0.837 (95% CI: 0.79–0.87) for tree-based/classical ML models. The difference was not significant ($z = 0.96$, $p = 0.34$).

Heterogeneity remained high in both groups ($I^2 > 94\%$), suggesting that model family contributed little to the observed variability.

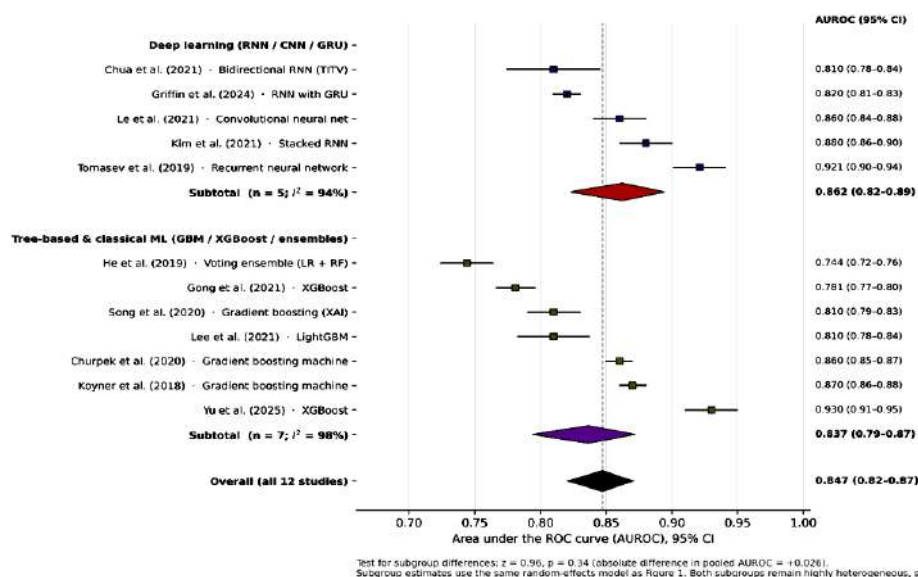


Fig. 3: Subgroup Analysis by Model Family

DISCUSSION

This systematic review and meta-analysis evaluated the predictive performance and clinical utility of AI/ML models for HA-AKI. Across 12 studies, the pooled AUROC was 0.847 (95% CI: 0.82–0.87), indicating good discrimination. However, heterogeneity was substantial ($I^2 = 96.9\%$), with a 95% prediction interval of 0.74–0.92, indicating considerable variability in performance across populations and healthcare settings. Thus, the pooled estimate should be interpreted as an overall summary rather than the expected performance of an individual model.

The observed performance is consistent with previous studies demonstrating the potential of AI/ML for early AKI prediction. Koyner *et al.* reported AUCs of 0.90 and 0.87 for stage 2 AKI prediction at 24 and 48 hours, respectively. ^[5] Churpek *et al.* reported AUCs of 0.85–0.86 for stage 2 or higher AKI across multiple health systems, supporting the potential for clinically useful discrimination following external validation. ^[8] Tomašev *et al.* further demonstrated that continuous deep learning using longitudinal EHR data could predict future AKI, including dialysis-requiring episodes. ^[7]

The high heterogeneity observed in our analysis is clinically important. Previous evidence has demonstrated substantial variation in AKI prediction studies regarding patient populations, clinical settings, predictors, prediction horizons, AKI definitions, algorithms, and validation strategies. ^[4] Vagliano *et al.* reported AUCs ranging from 0.49 to 0.99 across 46 ML-based AKI studies and identified frequent concerns regarding risk of bias, limited external validation, and inadequate calibration assessment. ^[4] These differences likely contribute to the variability observed in the present analysis.

Subgroup analysis showed no significant difference between deep-learning and tree-based/classical ML models. Deep-learning models had a pooled AUROC of 0.862 (95% CI: 0.82–0.89), compared with 0.837 (95% CI: 0.79–0.87) for tree-based/classical ML models ($z = 0.96$, $p = 0.34$). Heterogeneity remained high in both groups ($I^2 > 94\%$), suggesting that model architecture alone does not explain the observed variability. Patient characteristics, AKI stage, predictor availability, prediction horizon, data quality, and validation strategy may have greater influence on performance. ^[4]

The prediction interval of 0.74–0.92 further emphasizes the importance of model transportability. Although the

pooled AUROC indicates good discrimination, performance may differ substantially when models are applied to new hospitals or populations. Churpek *et al.* demonstrated relatively consistent performance across multiple health systems, but external validation remains limited among AKI prediction models. ^[4,8] Independent and prospective validation is therefore important before widespread clinical implementation.

Importantly, discrimination alone does not establish clinical utility. Models with high AUROC may have limited practical value if calibration is poor, false-positive alerts are excessive, or predictions do not lead to effective clinical interventions. Studies such as those by Lee *et al.* and Griffin *et al.* highlight the potential importance of decision-curve analysis and reducing alert burden when evaluating clinical implementation. ^[20,21] Future studies should therefore report calibration, decision-curve analysis, alert burden, interpretability, and patient-centered outcomes alongside discrimination measures.

The methodological quality and transparency of AI prediction studies also warrant attention. TRIPOD+AI provides guidance for transparent reporting of AI-based prediction models, while PROBAST+AI provides a framework for assessing risk of bias and applicability. ^[23,24] Adoption of these frameworks may improve reproducibility and facilitate comparison across models. The clinical implications of these findings are that AI/ML models show promise as early-warning tools for HA-AKI, but their performance appears highly context dependent. Future research should prioritize multicenter external and prospective validation, standardized AKI definitions and prediction horizons, calibration assessment, interpretability, and evaluation of clinical outcomes. Assessment of alert burden and workflow integration will also be important to determine whether improved prediction translates into meaningful improvements in patient care. ^[4,23,24]

STRENGTHS

This systematic review has several strengths, including a comprehensive literature search, predefined eligibility criteria, quantitative synthesis of predictive performance, assessment of heterogeneity, and subgroup analysis by model family. The prediction interval additionally provides clinically relevant information regarding the expected variability of model performance across future settings.

LIMITATIONS

Several limitations should be acknowledged. Only 12 studies were included in the quantitative synthesis, limiting the statistical power of subgroup analyses. Considerable clinical and methodological heterogeneity existed in patient populations, AKI definitions, prediction horizons, predictors, algorithms, and validation strategies. Most studies were observational and predominantly retrospective, EHR-based, which may limit generalizability. Furthermore, AUROC primarily reflects discrimination and does not fully capture calibration or real-world clinical effectiveness.

CONCLUSIONS

AI and machine learning models demonstrate good overall discrimination for predicting hospital-acquired acute kidney injury, with a pooled AUROC of 0.847 (95% CI: 0.82–0.87). However, the substantial heterogeneity ($I^2 = 96.9\%$) and wide prediction interval (0.74–0.92) indicate considerable variability in performance across clinical settings. Deep-learning and tree-based/classical ML models showed comparable predictive performance, suggesting that model architecture alone does not explain this variability. These findings support the potential of AI/ML as early-warning tools for HA-AKI, while emphasizing the need for robust external and prospective validation, calibration assessment, and evaluation of clinical utility before widespread clinical implementation.

CONTRIBUTION OF AUTHORS

Research concept- Roshan Rajesh Menon, Shobhit Sharma, Urja Nikhil Parikh, Siddhanth Santhosh, Bridget Achiaa Quansah, Cecilia Agada, Waqas Alauddin

Research design- Roshan Rajesh Menon, Shobhit Sharma, Urja Nikhil Parikh, Siddhanth Santhosh, Bridget Achiaa Quansah, Cecilia Agada, Waqas Alauddin

Supervision- Roshan Rajesh Menon, Shobhit Sharma, Urja Nikhil Parikh, Siddhanth Santhosh, Bridget Achiaa Quansah, Cecilia Agada, Waqas Alauddin

Materials- Roshan Rajesh Menon, Shobhit Sharma, Urja Nikhil Parikh, Siddhanth Santhosh, Bridget Achiaa Quansah, Cecilia Agada, Waqas Alauddin

Data collection interpretation - Roshan Rajesh Menon, Shobhit Sharma, Urja Nikhil Parikh, Siddhanth Santhosh, Bridget Achiaa Quansah, Cecilia Agada, Waqas Alauddin

Literature search- Roshan Rajesh Menon, Shobhit

Sharma, Urja Nikhil Parikh, Siddhanth Santhosh, Bridget Achiaa Quansah, Cecilia Agada, Waqas Alauddin

Writing article- Roshan Rajesh Menon, Shobhit Sharma, Urja Nikhil Parikh, Siddhanth Santhosh, Bridget Achiaa Quansah, Cecilia Agada, Waqas Alauddin

Critical review- Roshan Rajesh Menon, Shobhit Sharma, Urja Nikhil Parikh, Siddhanth Santhosh, Bridget Achiaa Quansah, Cecilia Agada, Waqas Alauddin

Article editing- Roshan Rajesh Menon, Shobhit Sharma, Urja Nikhil Parikh, Siddhanth Santhosh, Bridget Achiaa Quansah, Cecilia Agada, Waqas Alauddin

Final approval- Roshan Rajesh Menon, Shobhit Sharma, Urja Nikhil Parikh, Siddhanth Santhosh, Bridget Achiaa Quansah, Cecilia Agada, Waqas Alauddin

REFERENCES

- [1] Susantitaphong P, Cruz DN, Cerda J, Abulfaraj M, Alqahtani F, et al. Acute Kidney Injury Advisory Group of the American Society of Nephrology. World incidence of AKI: a meta-analysis. *Clin J Am Soc Nephrol.*, 2013; 8(9): 1482-93. doi: 10.2215/CJN.00710113. Epub 2013 Jun 6. Erratum in: *Clin J Am Soc Nephrol.*, 2014; 9(6): 1148.
- [2] Makris K, Spanou L. Acute Kidney Injury: Definition, Pathophysiology and Clinical Phenotypes. *Clin Biochem Rev.*, 2016; 37(2): 85-98.
- [3] Khwaja A. KDIGO clinical practice guidelines for acute kidney injury. *Nephron Clin Pract.* 2012; 120(4):c179-84. doi: 10.1159/000339789.
- [4] Vagliano I, Chesnaye NC, Leopold JH, Jager KJ, Abu-Hanna A, et al. Machine learning models for predicting acute kidney injury: a systematic review and critical appraisal. *Clin Kidney J.*, 2022; 15(12): 2266-80. doi: 10.1093/ckj/sfac181.
- [5] Wainstein M, Flanagan E, Johnson DW, Shrapnel S. Systematic review of externally validated machine learning models for predicting acute kidney injury in general hospital patients. *Front Nephrol.*, 2023; 3:1220214. doi: 10.3389/fneph.2023.1220214.
- [6] Koyner JL, Carey KA, Edelson DP, Churpek MM. The Development of a Machine Learning Inpatient Acute Kidney Injury Prediction Model. *Crit Care Med.*, 2018; 46(7): 1070-77. doi: 10.1097/CCM.0000000000003123.



- [7] Tomašev N, Glorot X, Rae JW, Zielinski M, Askham H, et al. A clinically applicable approach to continuous prediction of future acute kidney injury. *Nature*, 2019; 572(7767): 116-19. doi: 10.1038/s41586-019-1390-1.
- [8] Churpek MM, Carey KA, Edelson DP, Singh T, Astor BC, et al. Internal and External Validation of a Machine Learning Risk Score for Acute Kidney Injury. *JAMA Netw Open*, 2020; 3(8): e2012892. doi: 10.1001/jamanetworkopen.2020.12892.
- [9] Kim K, Yang H, Yi J, Son HE, Ryu JY, et al. Real-Time Clinical Decision Support Based on Recurrent Neural Networks for In-Hospital Acute Kidney Injury: External Validation and Model Interpretation. *J Med Internet Res.*, 2021; 23(4): e24120. doi: 10.2196/24120.
- [10] Le S, Allen A, Calvert J, Palevsky PM, Braden G, et al. Convolutional Neural Network Model for Intensive Care Unit Acute Kidney Injury Prediction. *Kidney Int Rep.*, 2021; 6(5): 1289-98. doi: 10.1016/j.ekir.2021.02.031.
- [11] Chua HR, Zheng K, Vathsala A, Ngiam KY, Yap HK, et al. Health Care Analytics With Time-Invariant and Time-Variant Feature Importance to Predict Hospital-Acquired Acute Kidney Injury: Observational Longitudinal Study. *J Med Internet Res.*, 2021; 23(12): e30805. doi: 10.2196/30805.
- [12] Cama-Olivares A, Braun C, Takeuchi T, O'Hagan EC, Kaiser KA, et al. Systematic Review and Meta-Analysis of Machine Learning Models for Acute Kidney Injury Risk Classification. *J Am Soc Nephrol.*, 2025; 36(10): 1969-83. doi: 10.1681/ASN.0000000702.
- [13] Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ*, 2016; 354: i4086. doi: 10.1136/bmj.i4086.
- [14] Siddaway AP, Wood AM, Hedges LV. How to do a systematic review: a best practice guide for conducting and reporting narrative reviews, meta-analyses, and meta-syntheses. *Annu Rev Psychol.*, 2019; 70: 747-70. doi: 10.1146/annurev-psych-010418-102803.
- [15] Bethel AC, Shaw N, Abbot R, et al. Insights from search summary tables for evidence and gap maps: a case study on peer support interventions. *J Med Libr Assoc.*, 2025; 113: 177-83. doi: 10.5195/jmla.2025.1831.
- [16] James J, Costello JT, Drahota AK: A benchmark for the quality of reflexology intervention reporting using the template for intervention description and replication (TIDieR) checklist: a systematic review. *Eur J Integ Med.*, 2024; 1: 102391. doi: 10.1016/j.eujim.2024.102391.
- [17] He J, Hu Y, Zhang X, Wu L, Waitman LR, Liu M. Multi-perspective predictive modeling for acute kidney injury in general hospital populations using electronic medical records. *JAMIA Open*, 2019; 2(1): 115-22. doi: 10.1093/jamiaopen/ooy043.
- [18] Song X, Yu ASL, Kellum JA, Waitman LR, Matheny ME, et al. Cross-site transportability of an explainable artificial intelligence model for acute kidney injury prediction. *Nat Commun.*, 2020; 11(1): 5668. doi: 10.1038/s41467-020-19551-w.
- [19] Gong K, Lee HK, Yu K, Xie X, Li J. A prediction and interpretation framework of acute kidney injury in critical care. *J Biomed Inform.*, 2021; 113: 103653. doi: 10.1016/j.jbi.2020.103653.
- [20] Lee Y, Ryu J, Kang MW, Seo KH, Kim J, et al. Machine learning-based prediction of acute kidney injury after nephrectomy in patients with renal cell carcinoma. *Sci Rep.*, 2021; 11(1): 15704. doi: 10.1038/s41598-021-95019-1.
- [21] Griffin BR, Mudireddy A, Horne BD, Chonchol M, Goldstein SL, et al. Predicting Nephrotoxic Acute Kidney Injury in Hospitalized Adults: A Machine Learning Algorithm. *Kidney Med.*, 2024; 6(12): 100918. doi: 10.1016/j.xkme.2024.100918.
- [22] Yu X, Wang W, Wu R, Gong X, Ji Y, Feng Z. Construction of a machine learning-based interpretable prediction model for acute kidney injury in hospitalized patients. *Sci Rep.*, 2025; 15(1): 9313. doi: 10.1038/s41598-025-90459-5.
- [23] Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ*, 2024; 385: e078378. doi: 10.1136/bmj-2023-078378. Erratum in: *BMJ*, 2024; 385: q902. doi: 10.1136/bmj.q902.
- [24] Moons KGM, Damen JAA, Kaul T, Hooft L, Andaur Navarro C, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for



prediction models using regression or artificial intelligence methods. *BMJ*, 2025; 388: e082505. doi: 10.1136/bmj-2024-082505.

Open Access Policy:

Authors/Contributors are responsible for originality, contents, correct references, and ethical issues. SSR-IIJLS publishes all articles under Creative Commons Attribution- Non-Commercial 4.0 International License (CC BY-NC). <https://creativecommons.org/licenses/by-nc/4.0/legalcode>

