Kidney Medicine

Acute Kidney Injury Prognosis Prediction Using Machine Learning Methods: A Systematic Review

Yu Lin, Tongyue Shi, and Guilan Kong

Rationale & Objective: Accurate estimation of in-hospital outcomes for patients with acute kidney injury (AKI) is crucial for aiding physicians in making optimal clinical decisions. We aimed to review prediction models constructed by machine learning methods for predicting AKI prognosis using administrative databases.

Study Design: A systematic review following PRISMA guidelines.

Setting & Study Populations: Adult patients diagnosed with AKI who are admitted to either hospitals or intensive care units.

Search Strategy & Sources: We searched PubMed, Embase, Web of Science, Scopus, and Cumulative Index to Nursing and Allied Health for studies published between January 1, 2014 and February 29, 2024. Eligible studies employed machine learning models to predict in-hospital outcomes of AKI based on administrative databases.

Data Extraction: Extracted data included prediction outcomes and population, prediction models with performance, feature selection methods, and predictive features.

Analytical Approach: The included studies were qualitatively synthesized with assessments of quality and bias. We calculated the pooled model discrimination of different AKI prognoses using random-effects models.

Results: Of 3,029 studies, 27 studies were eligible for qualitative review. In-hospital outcomes for patients with AKI included acute kidney disease, chronic kidney disease, renal function recovery or kidney failure, and mortality. Compared with models predicting the mortality of patients with AKI during hospitalization, the prediction performance of models on kidney function recovery was less accurate. Meta-analysis showed that machine learning methods outperformed traditional approaches in mortality prediction (area under the receiver operating characteristic curve, 0.831; 95% CI, 0.799-0.859 vs 0.772; 95% CI, 0.744-0.797). The overlapping predictive features for in-hospital mortality identified from ≥6 studies were age, serum creatinine level, serum urea nitrogen level, anion gap, and white blood cell count. Similarly, age, serum creatinine level, AKI stage, estimated glomerular filtration rate, and comorbid conditions were the common predictive features for kidney function recovery.

Limitations: Many studies developed prediction models within specific hospital settings without broad validation, restricting their generalizability and clinical application.

Conclusions: Machine learning models outperformed traditional approaches in predicting mortality for patients with AKI, although they are less accurate in predicting kidney function recovery. Overall, these models demonstrate significant potential to help physicians improve clinical decision making and patient outcomes.

Registration: CRD42024535965.

Complete author and article information provided before references.

Correspondence to G. Kong (guilan.kong@hsc.pku.edu.cn)

Kidney Med. 7(1):100936. Published online November 15, 2024.

doi: 10.1016/j.xkme.2024.100936

INTRODUCTION

Acute kidney injury (AKI) is a critical condition among hospitalized patients, especially in those admitted to the intensive care unit (ICU), with a mortality rate of 26.9%, almost 4 times higher than patients without AKI during an ICU stay.[1] In addition to high mortality rates, patients with AKI may have a high possibility of experiencing long-term complications such as acute kidney disease, kidney failure, or cardiovascular events.[2-5] Despite decades of research, few effective therapies exist to improve kidney function recovery in patients with AKI.[6] Early recognition of the onset and development of AKI and timely interventions are essential for the effective management and the prevention of long-term kidney damage in patients with AKI.

In recent years, numerous reviews have been published to provide an overview of AKI prediction model development, with some of them focused on specific populations, such as patients with hospital- or ICU-acquired AKI,[7,8] those undergoing cardiac or noncardiac surgery,[9-11] liver transplant surgery,[12] or other conditions.[13-18] Many studies in the past decade have employed traditional regression models, mostly logistic or Cox regression, to predict new-onset AKI or its prognosis.[19,20] However, these methods may be limited when there is collinearity between variables. Advancements in machine learning techniques have provided better solutions to disentangle the nonlinear relationship between predictors and outcomes and effectively use longitudinal data, offering new opportunities for early intervention.[21-23]

In clinical practice, the therapeutic strategy for stage 3 AKI, characterized by volume overload or electrolyte abnormalities, is using kidney replacement therapy (KRT) to

correct metabolic disorders and alleviate fluid overload.[24] However, early KRT initiation does not improve the overall survival of patients with AKI, based on existing evidence.[25,26] Several clinical trials[27,28] highlighted the lack of effective tools to identify which patients with AKI would benefit from KRT at an early stage. Developing prediction models that accurately predict patients with AKI with a high risk of adverse outcomes can significantly improve the management and treatment of those patients. Several models have been developed for AKI prognosis prediction using machine learning methods, and their performances were compared with those of traditional statistical approaches.[29-31] Machine learning models have demonstrated impressive performance in predicting mortality compared with traditional approaches, providing clinicians with useful information regarding further adverse events or long-term prognosis. Given the intricate nature of time dependency of clinical data, some deep learning approaches offer real-time probability prediction of clinical outcomes,[32] which potentially enable the integration of the algorithms with e-alert systems for the timely detection of adverse events.

In the subsequent sections, we will explore in depth the methodologies, datasets, and performance metrics used in AKI prognosis prediction studies, as well as the challenges of real-world implementation and future directions for machine learning in this field.

METHODS

Literature Search

We preregistered the protocol for this review in the International Database of Prospectively Registered Systematic Reviews (registration ID: CRD42024535965) and followed the PRISMA guidelines[33] as well as a previous systematic review guideline on prediction model performance[34] to assess machine learning models for clinical outcomes in patients with AKI. We searched PubMed, Embase, Web of Science, Scopus, and Cumulative Index to Nursing and Allied Health databases for literature published in the last 10 years, from January 1, 2014 to February 29, 2024, to explore the application of machine learning models in AKI prognosis-related studies. The keywords for literature search are detailed in Item S1: Supplementary Text and the search strategies are displayed in Fig S6. We also extensively searched the reference list of a published systematic review focused on AKI prognosis.[35]

Research ethics committee approval: As a systematic review based on published literature, this study did not involve new data collection from human participants, thus exempting it from formal ethics approval.

Informed consent: Because this study used only deidentified data from published sources, informed consent was waived.

Inclusion and Exclusion Criteria

We included studies that were as follows: (1) focused on adult patients admitted into hospitals or ICUs; (2) used routinely collected data from administrative databases such as electronic health records or electronic medical records; (3) developed and validated prediction models for the clinical outcomes in patients with AKI; and (4) were published within the last 10 years from January 1, 2014 to February 29, 2024. Additionally, studies that used single-center or multicenter research databases primarily consisting of regularly collected electronic health record or electronic medical record data were also included in this review. The exclusion criteria were as follows: (1) studies that did not use administrative databases such as registry-based studies; (2) validation studies that did not develop the prediction models; (3) conference abstracts; (4) unpublished literature such as preprint articles; and (5) articles not written in English.

Study Selection and Data Extraction

After applying the search strategies, we initially retrieved studies with titles or abstracts related to our topic. Two researchers independently screened the literature using the inclusion and exclusion criteria, extracted data, and cross-checked the findings. During the screening process, the researchers first excluded irrelevant literature by title and accessed the abstracts with full texts to determine the articles for final inclusion. Any discrepancies were resolved by discussion within groups. For the data extraction, 2 researchers exacted the information of each study, including the prediction outcomes and population (Table 1),[29-31,36-59] prediction models with performance (Table 2),[29-31,36-59] feature selection methods, and predictive features (Table 3).[29-31,36-59] We also completed a quality assessment and evaluated the risk of bias (Tables S1 and S2) of the included studies.

Data Synthesis and Quality Assessment

First, we qualitatively reviewed the included studies for the model characteristics and predictive features for outcomes. The model discrimination was recorded for models validated internally or externally. The best C-statistic or area under the receiver operating characteristic curve (AUROC) was reported for both machine learning and traditional models when studies used multiple prediction algorithms. If the study did not provide the average model performance for the prediction at several time points, we reported the best C-statistic or AUROC for the prediction model at different time points. We further conducted meta-analyses for studies that developed models using both machine learning and traditional methods. Detailed methods for the meta-analyses are shown in Item S1: Supplementary Text. The quality of the prediction model in each study was evaluated using the following criteria: (1) handling of missing data, (2) validation method used, (3) inclusion of external validation, (4) calibration of the model, (5) scope of outcome assessed, and (6) model availability. Incompleteness was not meant to indicate low quality but may implicate the potential limitations of each publication. We employed the standard tool Prediction model Risk of Bias Assessment Tool to evaluate the risk of bias and the applicability of the included studies.[60,61]

Table 1. Characteristics of Patients With AKI and Prediction Outcomes in the Selected Literature

Study	Prediction Outcome	Data Source	Population	Sample Size in Included Population/ Training Set	No. of Participants With Outcome	Incidence of Outcome
He et al[36] (2021)	Acute kidney disease	Development: Beijing Friendship Hospital Validation: MIMIC III	Sepsis-associated AKI	209	116	55.5%
Liu et al[37] (2022)	Mortality	University of Kentucky Hospital (UK Albert B. Chandler Hospital)	AKI with dialysis	570	237	41.6%
Liu et al[38] (2021)	Mortality	Development: University of Kentucky Hospital (UK Albert B. Chandler Hospital) Validation: MIMIC III	AKI with dialysis	608	247	40.6%
Yang et al[39] (2023)	Mortality	MIMIC IV	Sepsis-associated AKI	9,158	1,940	21.2%
Chang et al[40] (2022)	Mortality	Pooled dataset of MIMIC III and eICU-CRD	patients receiving KRT for AKI	11,558	3,412	29.5%
Liu et al[41] (2021)	Mortality	eICU-CRD	ICU patients with AKI	7,548	1,234	16.4%
Luo et al[42] (2022)	Mortality	Development: MIMIC IV Validation: eICU-CRD	Sepsis-associated AKI	6,066	1,127	18.6%
Huang et al[43] (2021)	Mortality	Development: MIMIC III Validation: eICU-CRD and the Second People's Hospital of Shenzhen	ICU patients with AKI	3,411	838	24.6%
Cunha et al[44] (2016)	Mortality	MIMIC III	ICU patients with AKI	2,362	1,148	48.6%
Hung et al[31] (2022)	Mortality	Changhua Christian Hospital (CCH) Clinical Research Database (CCHRD)	AKI patients who received CKRT	2,932	2,024 (in-hospital) 1,733 (28-d) 1,984 (90-d)	69.0% (in-hospital) 59.1% (28-d) 67.7% (90-d)
Huang et al[45] (2021)	Mortality	MIMIC III	Hospitalized patients with AKI	2,247	824	36.7%
Tang et al[46] (2024)	Mortality	MIMIC IV	Sepsis-associated AKI	8,426	1,813	21.5%
Li et al[47] (2023)	Mortality	MIMIC IV	Sepsis-associated AKI	8,129	1,629	20.0%
Lin et al[48] (2019)	Mortality	MIMIC III	ICU patients with AKI	19,044	2,586	13.6%
Fan et al[29] (2023)	Mortality	Development: MIMIC IV Validation: Hangzhou First People's Hospital Affiliated to Zhejiang University School of Medicine	Sepsis-associated AKI	2,599	712	27.4%
Zhou et al[49] (2023)	Mortality	MIMIC IV, Xiangya Hospital and Third Xiangya Hospital of Central South University, Changsha, China	Sepsis-associated AKI	16,154	3,318	20.5%

(Continued)

Table 1 (Cont'd). Characteristics of Patients With AKI and Prediction Outcomes in the Selected Literature

Study	Prediction Outcome	Data Source	Population	Sample Size in Included Population/ Training Set	No. of Participants With Outcome	Incidence of Outcome
Neyra et al[50] (2023)	Multiple outcomes: Mortality Major adverse kidney events (death, KRT or long-term decrease in eGFR)	Development: The University of Kentucky Hospital Validation: The University of Texas Southwestern (UTSW) Medical Center	Critically ill patients with AKI	7,354	2,383	32.4%
Nateghi et al[51] (2023)	Multiple outcomes: Mortality CKD	AZ Groeninge Hospital in Kortrijk	ICU patients diagnosed with AKI stage 3	101	43 (mortality) 47 (CKD)	42.6% (mortality) 46.5% (CKD)
Pike et al[52] (2015)	Multiple outcomes: Mortality Kidney function recovery	27 US Veterans Affairs- and university-affiliated centers	Critically ill patients with AKI	817	298 (recovery) 415 (mortality)	36.5% (recovery) 50.8% (mortality)
Wei et al[53] (2023)	Multiple outcomes: Mortality Kidney function recovery	Shanghai General Hospital, Shanghai Jiao Tong University School of Medicine	AKI patients who received PIKRT	493	256 (mortality) 150 (recovery at 30 d) 163 (recovery at 90 d)	51.9% (mortality) 30.4% (recovery at 30 d) 33.1% (recovery at 90 d)
Wu et al[54] (2023)	Multiple outcomes: Mortality KRT	Chinese Renal Disease Data System	Hospitalized patients with AKI	137,084	1,864	1.4%
Liu et al[55] (2021)	Non-recovery AKI	Chang Gung Memorial Hospitals in Taiwan	Patients with AKI at admission (index_AKI) and/ or during the hospitalization	8,600	3,871	45.0%
Lee et al[56] (2019)	Kidney function recovery	21 Kaiser Permanente-owned hospitals	Hospitalized patients with AKI	2,214	905	40.9%
Huang et al[30] (2023)	Kidney function recovery	Multicenter EPaNIC database	Critically ill patients with ICU-acquired AKI stage 3	229	86	37.6%
Zhao et al[57] (2022)	Kidney function recovery	MIMIC IV	Hospitalized patients with AKI	12,321	8,364	67.9%
Low et al[58] (2019)	KRT	A tertiary institution in Singapore	Hospitalized patients with AKI	3,333	174	5.2%
Pattharanitima et al[59] (2021)	KRT-free survival	MIMIC III	Critically ill patients with AKI requiring CKRT	684	205	30.0%

Abbreviations: AKI, acute kidney injury; CKD, chronic kidney disease; CKRT, continuous kidney replacement therapy; eGFR, estimated glomerular filtration rate; eICU-CRD, eICU Collaborative Research Database; EPaNIC, Early Parenteral Nutrition Completing Enteral Nutrition in Adult Critically Ill Patients; ICU, intensive care unit; KRT, kidney replacement therapy; MIMIC, Medical Information Mart for Intensive Care; PIKRT, prolonged intermittent kidney replacement therapy.

Table 2. Characteristics of Models for Predicting the Clinical Outcomes of Patients With AKI

Study	Prediction Outcome	Time for Outcome Prediction	Algorithm	Model Discrimination (Traditional Methods)	Model Discrimination (Machine Learning)	Other Metrics to Measure Performance	Time Series Features Included
He et al[36] (2021)	Acute kidney disease	7 d after AKI-initiating event	Recurrent neural network-LSTM, decision tree, logistic regression	AUROC: 0.728	AUROC: 1.000	NA	No
Liu et al[37] (2022)	Mortality	Mortality during hospitalization	Knowledge-gulded Time-aware (KIT)-LSTM, LSTM, T-LSTM, Phased-LSTM, RE-TAIN, ATTAIN, Transformer, XGBoost, SVM	NA	AUROC: 0.75 ± 0.10	Recall: 0.62 ± 0.21, Precision: 0.70 ± 0.10, F3 score: 0.62 ± 0.20	Yes
Liu et al[38] (2021)	Mortality	Rolling mortality prediction during hospitalization (24 h, 48 h, 72 h)	Knowledge graph guided double attention LSTM model (KGDAL), LSTM, XGBoost, random forest	NA	AUROC: 0.76	Accuracy: 0.71, Precision: 0.66, Recall: 0.87, F1 score: 0.75	Yes
Yang et al[39] (2023)	Mortality	28-dmortality after ICU admission	XGBoost, random forest, GBM, logistic regression, SAPS II score	AUROC: 0.850 (95% CI, 0.836-0.864)	AUROC: 0.873 (95% CI, 0.860-0.886)	NA	No
Chang et al[40] (2022)	Mortality	30-d mortality	XGBoost, random forest, MLP, logistic regression	AUROC: 0.819 (95% CI, 0.787-0.851)	AUROC: 0.823 (95% CI, 0.791-0.854)	Accuracy: 0.758, Sensitivity: 0.635, Specificity: 0.832, PPV: 0.697, NPV: 0.790	No
Liu et al[41] (2021)	Mortality	Mortality during hospitalization	XGBoost, SVM, random forest, logistic regression	AUROC: 0.662	AUROC: 0.796	Accuracy: 0.860, Precision: 0.860, Recall: 0.994, F1 score: 0.922	No
Luo et al[42] (2022)	Mortality	Mortality in 48, 72, and 120 h and in the first 28 d after ICU admission	XGBoost, SOFA score, SAPS II score	AUROC: 0.763 (95% CI, 0.751-0.775)	AUROC: 0.848 (95% CI, 0.838-0.858)	Accuracy: 0.795, Sensitivity: 0.674, Specificity: 0.803	Yes
Huang et al[43] (2021)	Mortality	Mortality during hospitalization and at 28 and 90 d	Random forest, SOFA score, SAPS II score, APACHE IV score	AUROC: 0.74	AUROC: 0.82	Accuracy: 0.75, Sensitivity: 0.70, Specificity: 0.77, F1 score: 0.67	No
Cunha et al[44] (2016)	Mortality	24-h, 1-y mortality	Fuzzy c-means, Gustafson-Kessel algorithm	NA	AUROC: 0.76 ± 0.04	Accuracy: 0.69 ± 0.05, Sensitivity: 0.69 ± 0.04, Specificity: 0.69 ± 0.05	No

(Continued)

Table 2 (Cont'd). Characteristics of Models for Predicting the Clinical Outcomes of Patients With AKI

Study	Prediction Outcome	Time for Outcome Prediction	Algorithm	Model Discrimination (Traditional Methods)	Model Discrimination (Machine Learning)	Other Metrics to Measure Performance	Time Series Features Included
Hung et al[31] (2022)	Mortality	Mortality in 28 d, 90 d, and during hospitalization	SVM, random forest, GBM, XGBoost	NA	AUROC: 0.823 (95% CI, 0.788-0.858)	Sensitivity: 0.742, Specificity: 0.787, PPV: 0.881, NPV: 0.590, F1 score: 0.806, Accuracy: 0.756	No
Huang et al[45] (2021)	Mortality	1-year mortality	XGBoost, artificial neural network	NA	AUROC: 0.83	Accuracy: 0.81, Sensitivity: 0.81, Specificity: 0.94, PPV: 0.69, NPV: 0.83	No
Tang et al[46] (2024)	Mortality	Mortality during hospitalization	SVM, GBM, AdaBoost, XGBoost, CatBoost, Naïve Bayesian, neural network, MLP, KNN, random forest, logistic regression	AUROC: 0.771	AUROC: 0.804	Youden index: 0.470, F1 score: 0.402, Accuracy: 0.810 (95% CI, 0.790-0.830), Sensitivity: 0.290 (95% CI, 0.250-0.330), Specificity: 0.950 (95% CI, 0.940-0.960), PPV: 0.640 (95% CI, 0.580-0.700), NPV: 0.830 (95% CI, 0.810-0.840)	No
Li et al[47] (2023)	Mortality	Mortality during hospitalization	XGBoost, random forest, SVM, KNN, decision tree, logistic regression, SAPS II score	AUROC: 0.730 (95% CI, 0.694-0.765)	AUROC: 0.794 (95% CI, 0.762-0.827)	Accuracy: 0.832, Sensitivity: 0.793, Specificity: 0.752, Average precision: 0.660	No
Lin et al[48] (2019)	Mortality	Mortality during hospitalization	Random forest, SVM, artificial neural network, customized SAPS II model based on logistic regression	AUROC: 0.795 (95% CI, 0.781-0.809)	AUROC: 0.866 (95% CI, 0.862-0.870)	Brier score: 0.085 (95% CI, 0.084-0.086), Accuracy: 0.728 (95% CI, 0.715-0.741), F1 score: 0.459 (95% CI, 0.449-0.470)	No
Fan et al[29] (2023)	Mortality	Mortality in 7, 14, or 28 d	XGBoost, random forest, MLP, support vector classifier, logistic regression	AUROC: 0.75 (95% CI, 0.73-0.77)	AUROC: 0.91 (95% CI, 0.90-0.92)	Accuracy: 0.85, Precision: 0.85, F-score: 0.85, Recall: 0.85	No

(Continued)

Table 2 (Cont'd). Characteristics of Models for Predicting the Clinical Outcomes of Patients With AKI

Study	Prediction Outcome	Time for Outcome Prediction	Algorithm	Model Discrimination (Traditional Methods)	Model Discrimination (Machine Learning)	Other Metrics to Measure Performance	Time Series Features Included
Zhou et al[49] (2023)	Mortality	Mortality during hospitalization	CatBoost, GBM, LightGBM, AdaBoost, XGBoost, KNN, MLP, SVM, logistic regression	AUROC: 0.788	AUROC: 0.827	Accuracy: 0.750, Youden index: 0.500, Sensitivity: 0.750, Specificity: 0.750, F1 score: 0.560, PPV: 0.440, NPV: 0.920	No
Neyra et al[50] (2023)	Multiple outcomes: Mortality MAKE: death, KRT, or long-term decrease in eGFR	Mortality during hospitalization Mortality or KRT or 50% reduction of eGFR within or at 120 d after discharge KRT within the last 48 h before discharge	Random forest, SVM, XGBoost, logistic regression	Mortality AUROC: 0.71 (95% CI, 0.71-0.71) MAKE AUROC: 0.67 (95% CI, 0.67-0.67)	Mortality AUROC: 0.74 (95% CI, 0.73-0.74) MAKE AUROC: 0.73 (95% CI, 0.72-0.74)	Mortality – Accuracy: 0.65 (95% CI, 0.64-0.66), Precision: 0.18 (95% CI, 0.17-0.18), Sensitivity: 0.69 (95% CI, 0.67-0.71), Specificity: 0.64 (95% CI, 0.63-0.65), F1 score: 0.28 (95% CI, 0.27-0.29), PPV: 0.18 (95% CI, 0.17-0.18), NPV: 0.95 (95% CI, 0.95-0.95), Calibration slope: 1.17 (95% CI, 1.12-1.21) MAKE – Accuracy: 0.67 (95% CI, 0.66-0.69), Precision: 0.42 (95% CI, 0.40-0.43), Sensitivity: 0.67 (95% CI, 0.64-0.69), Specificity: 0.68 (95% CI, 0.64-0.71), F1 score: 0.51 (95% CI, 0.50-0.52), PPV: 0.42 (95% CI, 0.40-0.43), NPV: 0.85 (95% CI, 0.85-0.86), Calibration slope: 1.34 (95% CI, 1.27-1.41)	No
Nateghi et al[51] (2023)	Multiple outcomes: Mortality CKD	3 and 6 mo after experiencing stage 3 AKI	Random survival forests, survival XGBoost, logistic regression	Kidney recovery AUROC: 0.717 Mortality C-statistic: 0.776 ± 0.081	Kidney recovery AUROC: 0.846 Mortality C-statistic: 0.825 ± 0.057	Kidney recovery PR-AUC: 0.895	No

(Continued)

Table 2 (Cont'd). Characteristics of Models for Predicting the Clinical Outcomes of Patients With AKI

Study	Prediction Outcome	Time for Outcome Prediction	Algorithm	Model Discrimination (Traditional Methods)	Model Discrimination (Machine Learning)	Other Metrics to Measure Performance	Time Series Features Included
Pike et al[52] (2015)	Multiple outcomes: Mortality Kidney function recovery	Kidney recovery or mortality at day 60	LASSO, stepwise model	Renal recovery AUROC: 0.73 (95% CI, 0.68-0.78) Mortality AUROC: 0.74 (95% CI, 0.69-0.78)	Kidney recovery AUROC: 0.76 (95% CI, 0.71-0.81) Mortality AUROC: 0.78 (95% CI, 0.73-0.82)	Net Reclassification Index: Kidney recovery 0.27 (95% CI, 0.08-0.47) Mortality 0.41 (95% CI, 0.22-0.59)	No
Wei et al[53] (2023)	Multiple outcomes: Mortality Kidney function recovery	30-d mortality, 30- and 90-d renal function recovery	SVM, KNN, Naïve Bayes, perceptron, stochastic gradient descent, decision tree, random forest, logistic regression	Mortality AUROC: 0.665 Renal recovery AUROC: 0.534	Mortality AUROC: 0.679 Renal recovery AUROC: 0.665	Mortality – Accuracy: 0.669, Precision: 0.755, Recall: 0.545, F1 score: 0.633 Renal recovery – Accuracy: 0.709, Precision: 0.637, Recall: 0.493, F1 score: 0.556	No
Wu et al[54] (2023)	Multiple outcomes: Mortality KRT	24 h, 28 h, 72 h, and 7d	Bidirectional LSTM model	NA	Mortality AUROC: 0.924 KRT AUROC: 0.776	Mortality – Accuracy: 0.885, Precision: 0.673, Recall: 0.930, F1 score: 0.755 KRT – Accuracy:0.885, Precision: 0.673, Recall: 0.934, F1 score: 0.755	Yes
Liu et al[55] (2021)	Non-recovery AKI	AKI non-recovery during hospitalization	LASSO, random forest, XGBoost, LightGBM, stepwise logistic regression	AUROC: 0.790 ± 0.014	AUROC: 0.808 ± 0.015	Sensitivity: 0.661 ± 0.037, Specificity: 0.796 ± 0.050, Precision: 0.800 ± 0.031, F1 score: 0.723 ± 0.016	No
Lee et al[56] (2019)	Kidney function recovery	Kidney function recovery without KRT dependence within 90 d after KRT initiation and survival for ≥4 wk after KRT discontinuation	logistic regression, CART	C-statistic: 0.94	C-statistic: 0.64	NA	No
Huang et al[30] (2023)	Kidney function recovery	Kidney function recovery during hospitalization	LASSO	Biomarker NGAL_AKI: 0.54	AUROC: 0.71	NA	No

(Continued)

Table 2 (Cont'd). Characteristics of Models for Predicting the Clinical Outcomes of Patients With AKI

Study	Prediction Outcome	Time for Outcome Prediction	Algorithm	Model Discrimination (Traditional Methods)	Model Discrimination (Machine Learning)	Other Metrics to Measure Performance	Time Series Features Included
Zhao et al[57] (2022)	Kidney function recovery	72 h after AKI initiated event	XGBoost, Bayesian networks, random forest, SVM, logistic regression	AUROC: 0.807 ± 0.010	AUROC: 0.836 ± 0.020	Accuracy: 0.730 ± 0.010, Sensitivity: 0.502 ± 0.030, Specificity: 0.958 ± 0.010	No
Low et al[58] (2019)	KRT	Diagnosed with AKI >1 y	Decision tree and logistic regression	AUROC: 0.930	Accuracy: 0.704	PPV: 0.970; NPV: 0.780	No
Pattharanitima et al[59] (2021)	KRT-free survival	≥7 d before hospital discharge without the requirement of KRT	MLP + LSTM, MLP, XGBoost, random forest, SVM, AdaBoost, logistic regression	AUROC: 0.57 (95% CI, 0.52-0.62)	AUROC: 0.70 (95% CI, 0.67-0.73)	Sensitivity: 0.71, Specificity: 0.59, PPV: 0.42, NPV: 0.81	Yes

Abbreviations: AdaBoost, adaptive boosting; AKI, acute kidney injury; APACHE, Acute Physiology And Chronic Health Evaluation; AUROC, area under the receiver operating characteristic curve; CART, Classification and Regression Trees; CatBoost, category boosting; CI, confidence interval; CKD, chronic kidney disease; eGFR, estimated glomerular filtration rate; GBM, gradient boosting machine; ICU, intensive care unit; KNN, K-nearest neighbor; KRT, kidney replacement therapy; LASSO, least absolute shrinkage and selection operator; LSTM, long short-term memory; MAKE, major adverse kidney events; MLP, multilayer perceptron; NA, not assessed; NGAL, neutrophil gelatinase-associated lipocalin; NPV, negative predictive value; PPV, positive predictive value; PR-AUC, precision-recall area under the curve; SAPS II, Simplified Acute Physiology Score II; SOFA, Sequential Organ Failure Assessment; SVM, support vector machine; XGBoost, extreme gradient boosting.

RESULTS

After applying the search strategies, we retrieved 3,029 articles without duplication. Among these manuscripts, 2,823 were excluded based on their titles and abstracts, and 156 were selected for complete review (Fig 1). In accordance with our inclusion and exclusion criteria, we eventually included 27 studies for systematic review.[29-31,36-59] For the clinical outcomes of patients with AKI, the post-AKI conditions reported in the selected studies include acute kidney disease (n = 1), chronic kidney disease (n = 1), kidney function recovery or kidney failure requiring KRT (n = 9), and mortality (n = 20). Some studies focused on general hospitalized patients with AKI (n = 6), whereas most studies predicted the in-hospital outcomes for patients with AKI with specific conditions (n = 21), such as sepsis (n = 7), admission to the ICU (n = 8), and requirement for dialysis (n = 6). Separate datasets were used in 13 studies for model development and validation. The sample size for model training ranged from 101-137,084. Twenty-five studies evaluated the model prediction capability through retrospective validation, whereas Liu et al[55] and Nateghi et al[51] assessed the model prediction performance through prospective validation.

Prediction Models

The prediction model characteristics were extracted and are displayed in Table 2. The timing of predicting the prognosis outcomes of patients with AKI varied between different studies. Most studies predicted the mortality or kidney function recovery in patients with AKI during hospitalization (n = 10), whereas some studies predicted AKI prognoses at specific time points, such as 72 hours (n = 4), 7 days (n = 5), or 28 days (n = 5). The timing of prognosis prediction in patients with AKI may also affect the predictive capability of models, because the disease status may change over time but not necessarily at a constant rate.

Twenty-five studies examined the performance of >1 model in outcome predictions. Among 4 studies employing deep learning methods, 3 studies[37,38,59] constructed a deep learning model with superior performance to other machine learning methods such as extreme gradient boosting (XGBoost), support vector machine, or random forest for mortality prediction or KRT-free survival prediction. Additionally, Wu et al[54] proposed a deep learning model with better prediction accuracy for short-term mortality (24 hours: AUROC = 0.934) and KRT (24 hours: AUROC = 0.883) than longer prediction times (48 hours, 72 hours, and 7 days). Other commonly used machine learning models for AKI prognosis prediction included XGBoost (n = 17), random forest (n = 15), support vector machine (n = 11), decision tree (n = 5), multilayer perceptron (n = 5), gradient boosting machine (n = 4), and K-nearest neighbor (n = 4). A brief introduction to these machine learning methods is provided in Item S1: Supplementary Text and Figs S1-S5.

Table 3. Characteristics of Feature Selection Methods and Top Predictive Features for AKI Prognosis Prediction

Study	Time Points of Predictor Availability	Feature Selection Methods	Top Predictive Features
He et al[36](2021)	Within 3 d after ICU admission	RNN-LSTM model	Delta non-kidney SOFA, creatinine on day 3, hypertension, diuretics, delta creatinine, emergency department, non-kidney SOFA at day 3, baseline creatinine, kidney toxic drug, AKI stage, CCI
Liu et al[37] (2022)	Temporal features were used before the outcome	Not reported	Not reported
Liu et al[38] (2021)	Features were selected before outcome	Not reported	Not reported
Yang et al[39] (2023)	Not reported	Univariate regression, correlation analysis, and Boruta were combined for feature selection	Age, BMI, albumin, respiratory rate, pH, urine output, SpO_2, temperature, glucose, SUN, sodium, anion gap, WBC, $PaCO_2$, bicarbonate, heart rate, creatinine, Hgb, lactate, PaO_2, neutrophils, DBP, SBP, PaO_2/FiO_2, cerebrovascular disease, metastatic solid tumor, severe liver disease, mild liver disease, dementia
Chang et al[40] (2022)	Within 24 h before KRT initiation	Not reported	Creatinine, platelet, FiO_2, anion gap, GCS, age, mean arterial pressure, vasopressor, breaths per minute, bicarbonate
Liu et al[41] (2021)	Within 24 h of ICU admission	LASSO	Age, BMI, SCr, sodium, platelets, bicarbonate, chloride, BP, SUN, RBC, heart rate, respiratory rate, potassium
Luo et al[42] (2022)	predictors were selected within each time window (48 h, 72 h, 120 h, and 28 d)	Not reported	GCS, urine output, ICU length of stay, age, SUN, lactate, mechanical ventilation, vasopressors, heart rate, respiratory rate, SpO_2, BP, WBC, potassium, sodium, temperature, PTT, platelets, serum total bilirubin, Hgb, INR, pH, chloride, creatinine, KRT, ethnicity, PaO_2, $PaCO_2$, bicarbonate, albumin, sex, loop diuretics
Huang et al[43] (2021)	Within 24 h of ICU admission	Univariate analysis using random forest and logistic regression	Comfort measures only, do not resuscitate, SAPS II, urine output, SOFA, anion gap, SUN, lactate, BP, total CO_2, bicarbonate, platelets, respiration rate, age, INR, prothrombin time; WBC, length of stay, PTT, sodium
Cunha et al[44] (2016)	Not reported	Not reported	Not reported
Hung et al[31] (2022)	Within 24 h after CKRT initiation	Recursive feature elimination	APACHE II, albumin level, timing of CKRT initiation, age, potassium levels, SpO_2, mean arterial pressure, INR, creatinine levels, vasopressor use
Huang et al[45] (2021)	Not reported	Not reported	Age, cancer, kidney failure, hypotension, shock, anemia, hemorrhage, acidosis, hypertension, creatinine level, urine output, prothrombin time, hematocrit, Hgb, WBC, SUN, serum bicarbonate, serum sodium, platelet count, BP, sex
Tang et al[46] (2024)	Within 24 h of admission	Recursive feature elimination	AKI stage, PaO_2, lactate, urine output, norepinephrine injection rate, SUN, invasive mechanical ventilation, base excess, and anion gap
Li et al[47] (2023)	During the first 24 h after ICU admission	LASSO	SOFA, respiratory rate, SAPS II, age, cerebrovascular disease, body temperature, urine output, serum sodium, serum chloride, weight, INR, Hgb, heart rate, cancer, SCr, liver disease, PTT, platelets, anion gap, SUN
Lin et al[48] (2019)	Measured before the outcome	Not reported	Urine output, SBP, age, serum bicarbonate level, heart rate
Fan et al[29] (2023)	In the first 24 h of ICU stay	Not reported	SOFA, urine output, lactate, SpO_2, respiratory rate, INR, mean BP, anion gap, bicarbonate, heart rate, age, weight, temperature, chloride, glucose, sodium, SUN, potassium, hematocrit, WBC, creatinine, platelets

(Continued)

Table 3 (Cont'd). Characteristics of Feature Selection Methods and Top Predictive Features for AKI Prognosis Prediction

Study	Time Points of Predictor Availability	Feature Selection Methods	Top Predictive Features
Zhou et al[49] (2023)	Within 24 h of admission	Recursive feature elimination	Urine output, SUN, norepinephrine injection rate, anion gap, creatinine, RBC volume distribution width, INR, heart rate, temperature, respiratory rate, FiO_2, creatinine, GCS, diabetes, stroke
Neyra et al[50] (2023)	During the first 3 d of ICU admission	Features were ranked according to importance in logistic regression, Random Forest, SVM, and XGBoost, and the overlapping features were used as final predictive features	Mortality: Last KDIGO, urine output, FiO_2, pressor, SUN, platelets, age, bilirubin, Hgb, pH, heart rate, ICU admission SCr, serum sodium, maximum KDIGO MAKE: pressor, urine output, FiO_2, platelets, SUN, lowest temperature, hours in ICU, bilirubin, last KDIGO, age, pH, heart rate, highest temperature, ECMO, maximum KDIGO
Nateghi et al[51] (2023)	CysC and creatinine were measured at ICU admission and at the time of developing AKI stage 3; other variables were measured in the first day of ICU admission	Not reported	Mortality: SOFA, SAPS II, fluid balance, clinically COPD, suspected infection, oncological history, eGFR, ventilation, arterial hypertension, diabetes CKD: GFR, SCr, CysC, age, SOFA, SAPS II, length of stay in-hospital, length of stay in ICU, fluid balance, arterial hypertension, BMI
Pike et al[52] (2015)	Not reported	Pearson's correlation, partial eta squared and χ^2 test	Mortality: sodium, total protein, LDH, phosphorus, thrombin time Kidney function recovery: hematocrit, chlorine, total protein, uric acid, blood phosphorus, prothrombin time, CKD stage, diabetes
Wei et al[53] (2023)	Not reported	Multivariate logistic regression	Age, mean arterial pressure, mechanical ventilation, bilirubin, and plasma IL-8
Wu et al[54] (2023)	Within 24 h of hospital admission	Not reported	Not reported
Liu et al[55] (2021)	Before and during the index hospitalization for AKI	Recursive feature elimination	Age at index date, HA-AKI, index AKI stage at index admission, CCI, CKD, cancer, index SCr, baseline SCr, SUN, potassium, LDL cholesterol, SUA, calcium, CRP, albumin, erythrocyte sedimentation rate, WBC, lymphocyte count, neutrophil count, use of health service
Lee et al[56] (2019)	Demographics and inpatient laboratory values were recorded on the RRT initiation date; outpatient laboratory values and vital signs were measured 7 to 365 d before admission	Stepwise logistic regression applied to bootstrapped samples	eGFR, preadmission Hgb level, chronic liver disease, age
Huang et al[30] (2023)	The first day of AKI stage 3 in the ICU	Correlation-based feature selection method (CfsSubsetEval)	Age, sepsis on admission, surgery/trauma diagnostic group, KRT on the first AKI stage 3 day in ICU, cardiac surgery diagnostic group
Zhao et al[57] (2022)	Within the first 24 h after ICU admission and the diagnosis of AKI	Not reported	SCr, antibiotic duration, sodium, anion gap, temperature, chloride, SAPS II, Hgb, urine output volume to weight ratio, sepsis, heart rate, prothrombin time, SpO_2, AKI stage
Low et al[58] (2019)	Not reported	Multivariate logistic regression	AKI onset in ICU, hematological malignancy, higher delta SCr (SCr increase from AKI detection until peak), higher serum potassium, baseline eGFR
Pattharanitima et al[59] (2021)	From hospital admission and up to CKRT initiation	Not reported	Not reported

Abbreviations: AKI, acute kidney injury; APACHE, Acute Physiology And Chronic Health Evaluation; AST, aspartate aminotransferase; BMI, body mass index; BP, blood pressure; CCI, Charleson comorbidity index; CKD, chronic kidney disease; COPD, chronic obstructive pulmonary disease; CRP, C-reactive protein; CRRT, continuous renal replacement therapy; CysC, cystatin C; DBP, diastolic blood pressure; ECMO, extracorporeal membrane oxygenation; eGFR, estimated glomerular filtration rate; FiO_2, fraction of inspired oxygen; GCS, Glasgow Coma Scale; HA-AKI, hospital acquired acute kidney injury; Hgb, hemoglobin; ICU, intensive care unit; IL, interleukin; INR, international normalized ratio; KDIGO, Kidney Disease: Improving Global Outcomes; KRT, kidney replacement therapy; LASSO, least absolute shrinkage and selection operator; LDL, low-density lipoprotein; MAKE, major adverse kidney events; $PaCO_2$, partial pressure of carbon dioxide; PaO_2, partial pressure of oxygen; PTT, partial thromboplastin time; RBC, red blood cell; RNN-LSTM, recurrent neural network long short-term memory; RRT, renal replacement therapy; SAPS II, Simplified Acute Physiology Score II; SBP, systolic blood pressure; SCr, serum creatinine; SOFA, Sequential Organ Failure Assessment; SpO_2, oxygen saturation; SUA, serum uric acid; SUN, serum urea nitrogen; SVM, support vector machine; WBC, white blood cell; XGBoost, extreme gradient boosting.

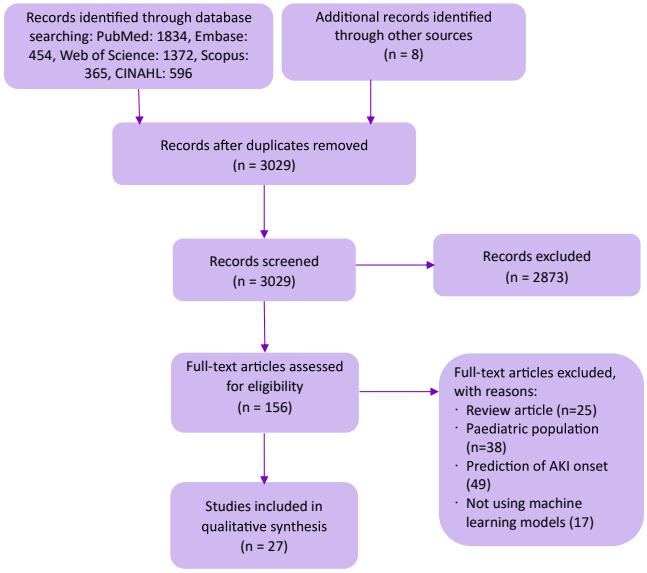

Figure 1. Flow diagram illustrating systematic literature search to identify eligible studies. AKI, acute kidney injury.

Of the selected 27 studies, 21 compared traditional methods (such as severity scoring systems [Simplified Acute Physiology Score II, Acute Physiology And Chronic Health Evaluation III] or logistic regression) with machine learning models. Except for the studies conducted by Low et al[58] and Lee et al,[56] 19 of 21 studies found that machine learning models had a higher AUROC or C-statistic than traditional statistical models or severity assessment scores. For mortality prediction (n = 20), the AUROC of the machine learning models in different studies ranged from 0.679-0.924, with 18 of them exceeding an AUROC >0.750. We meta-analyzed the model discrimination for mortality prediction (Fig S7), and machine learning approaches significantly outperformed traditional methods (AUROC, 0.831; 95% confidence interval [CI], 0.799-0.859 vs 0.772; 95% CI, 0.744-0.797).

The prediction models for kidney function recovery or kidney failure were less accurate than the mortality prediction models, with an AUROC ranging from 0.640-0.846. Only 4 studies provided sufficient data for meta-analysis of kidney functional recovery. Although machine learning showed a trend toward higher model discrimination, the difference was not statistically significant (AUROC, 0.781; 95% CI, 0.674-0.861 vs. 0.734; 95% CI, 0.547-0.863) for different models, which might be because of the limited number of studies that predicted kidney functional recovery (Fig S8). The Egger's test indicated no significant publication bias for all meta-analyses (Fig S9). Across the meta-analyses, there were high study heterogeneities with significant I^2 values exceeding 90% ($P < 0.001$), which is common in prediction model meta-analyses because of non-overlapping CIs among studies.[62,63] In addition to AUROC, a few studies reported the model performance using other model evaluation metrics. For example, sensitivity (recall) (n = 19), accuracy (n = 17), specificity (n = 13), F1 score (n = 11), precision (n = 10), positive predictive value (n = 8), and negative predictive value (n = 8) were commonly used in these studies.

Predictors for the Clinical Outcomes of Patients With AKI

We summarized each study's feature selection method and top predictive features (Table 3). Five studies applied traditional regression-based or correlation-based approaches for the feature selection. Six studies applied feature selection methods belonging to wrapped methods that can maximize the model prediction ability using a combined subset of features, such as recursive feature elimination or CfsSubsetEval from Weka software.

Except for 4 studies, 23 studies reported the most predictive features in their models. Among them, 18 studies reported important features for mortality, including age, serum creatinine level, serum urea nitrogen level, anion gap, and white blood cell count, which were overlapping predictive features for mortality that appeared in >6 studies. A multicenter study in China investigated risk factors for death in elderly AKI patients across 15 hospitals and found that age and comorbid conditions such as cardiovascular disease, cancer, and mechanical ventilation were significantly associated with mortality risk.[64] The increase in serum urea nitrogen level was linked to long-term mortality in patients with various diseases such as heart failure[65] or cardiovascular disease.[66] Additionally, serum anion gap has been reported to be robustly associated with mortality risk in critically ill patients with chronic obstructive pulmonary disease.[67] White blood cells, such as lymphocytes, play an important role in AKI development by producing proinflammatory cytokines.[68] For the outcome of kidney function recovery or kidney failure, age, serum creatinine level, estimated glomerular filtration rate, AKI stage, and comorbid conditions were frequently used predictive features. A retrospective cohort study showed that patients with a decreased estimated glomerular filtration rate during hospitalization would have an 18% increased mortality in the following year and a 267% increased risk of kidney failure within 10 years.[69] The various stages of AKI, particularly AKI requiring dialysis, which is the most severe form of AKI, were associated with increased mortality and kidney non-recovery in patients after surgery.[70] In the case of acute kidney disease, factors such as serum creatinine level, comorbid conditions, and medications played significant roles in determining acute kidney disease persistence.

Quality Assessment of Included Studies

The quality of the included studies was assessed from 6 perspectives, including handling of missing data, validation methods, validation type, calibration of the model, scope of the population included in model development, and model availability (Table S1). Of the 27 studies, 21

reported the time points of feature measurements used in model construction (Table 3). Nine studies used variables measured within 24 hours of ICU admission, whereas 3 used variables measured at AKI initiation to predict clinical outcomes. Eight studies were externally validated on different datasets, indicating that the variables used are widely available, and these models could be deployed across different hospitals if needed. Nineteen studies employed various methods to handle missing values, suggesting that some predictive variables may be of low quality and require further processing before risk estimation. Of the 21 studies comparing traditional methods with machine learning models, 14 reported model calibration, with 7 providing detailed calibration metrics. Four studies reported the slope of the calibration curve or the correlation between predicted and observed outcomes, ranging from 0.840-1.349 (a slope of 1 indicating perfect calibration). Three studies reported Hosmer-Lemeshow test P values (all > 0.05), indicating good model fit. Finally, 14 studies used the study population from a single hospital to predict the clinical outcomes of AKI patients. It was challenging to evaluate the generalizability of these prediction models trained and internally validated within a single hospital dataset without external validation. Specifically, 2 prediction models from Fan et al[29] and Neyra et al[50] are publicly available for nephrologists and can be used to predict mortality based on clinical variables routinely measured in patients. These models offer a foundation for incorporating machine learning into clinical practice, intending to reduce patient risk.

We further assessed the risk of bias for the included studies and the applicability of the prediction models using the Prediction model Risk Of Bias Assessment tool (Table S2). Machine learning models that were only developed and validated in a single dataset from a single hospital were considered as having low applicability ($n = 14$). Those constructing models from relatively small sample sizes, without feature selection, or without calibration were identified as having a high risk of bias for methodology. We identified only 4 studies with relatively low bias risk and minimal concerns regarding the applicability of their prediction models for patients with AKI.

DISCUSSION

In this comprehensive systematic review of 27 studies on AKI prognosis, we found that machine learning models provide better discrimination in predicting mortality than kidney function recovery, and they outperform traditional approaches in mortality prediction among patients with AKI. We identified a key set of predictive features for in-hospital mortality including age, serum creatinine level, serum urea nitrogen level, anion gap, and white blood cell count, whereas age, serum creatinine level, AKI stage, estimated glomerular filtration rate, and comorbid conditions played an important role in kidney function recovery or kidney failure prediction. The deep learning models that can identify the risk of adverse events dynamically with time-dependent features incorporated have superior model performance over static models. With the capability of identifying high-risk adverse events in patients with AKI, machine learning models might aid physicians in making timely and optimal interventions and ultimately improve the clinical outcomes of high-risk patients.

Given its potential effect on patient survival, the prognosis of AKI patients is a highly concerning topic in post-AKI management. The timing of KRT initiation remains controversial in AKI patients without heavy fluid overload or major metabolic disorders, such as acidosis, hyperkalemia, and uremia.[71] Meta-analysis showed no difference in mortality rate and dialysis dependence between early or late initiation of KRT.[26] Previous studies[27,28] highlighted an absence of risk stratification tools in patients with AKI without severe complications to guide KRT initiation. Machine learning models that accurately predict clinical outcomes of patients with AKI are capable of (1) helping clinicians to identify patients at a high risk of death or nonrecovery early and obtain the opportunity to provide timely interventions; (2) facilitating clinical trials for AKI patients at a high risk of adverse outcomes, potentially reducing mortality or loss of kidney function; and (3) assisting medical counseling and informing patients about the risks of deteriorating conditions. In addition to traditional machine learning methods, the rising trend of using deep learning models to identify AKI risk is a popular strategy for the early detection of patients with AKI. Deep learning models can capture useful information from time series data, demonstrating better prediction performance than traditional statistical approaches, machine learning models, and even medical experts.[72,73] A study systematically reviewed published studies on AKI prognosis with a wide range of prediction models and concluded that some clinical models were internally validated but unavailable for easy application in clinical settings. Our study extends the current understanding of the AKI prognosis prediction models by incorporating a broader range of machine learning approaches, providing valuable insights into the utilization of machine learning models in patients with AKI.

In this review, we summarized several overlapping predictive features from various machine learning models. Several biomarkers were used in AKI prediction, but these biomarkers have limited effects in predicting the prognosis of AKI.[74–77] Baseline estimated glomerular filtration rate, proteinuria, age, diabetes, AKI stage, and comorbid conditions have been shown to influence the probability of kidney function recovery.[78–81] A meta-analysis reviewed 63 studies to identify the biomarkers for KRT in AKI and found that neutrophil gelatinase-associated lipocalin, blood creatinine, and cystatin C were highly predictive for receiving KRT in patients with AKI.[82] Our study indicated that age and serum creatinine

levels have been repeatedly reported as significant risk factors for mortality in patients with AKI. In addition to serum creatinine, age is an essential indicator of kidney function.[83] A previous study highlighted the increased susceptibility of aging individuals to AKI and the high incidence of progression to kidney failure.[84] Delahunt et al[85] investigated patients with progeria and found that a patient showed focal glomerulosclerosis and tubular atrophy, a symptom of kidney aging. During the aging process, the deterioration of significant artery stiffness with the increased systolic blood pressure and diastolic blood pressure contributed to kidney microvascular damage and increased the possibility of trauma, further leading to kidney failure.[84,86,87] Multimorbidities in patients with AKI, such as diabetes, hypertension, and cardiovascular disease, are frequently reported to be associated with nonrecovery AKI.[74,88-90] Aging and chronic diseases accompanied by diminished glomerular reserve may lead to advanced-stage AKI.[91] Collective evidence has demonstrated that the severity of disease conditions is highly possible to be associated with hemodynamic instability and decreased nephron mass, eventually resulting in kidney nonrecovery.

This study has its limitations. Most of these studies developed prediction models within specific hospital settings and did not validate them at a large scale in hospitals at a national level or across countries, limiting broad application of these models. Furthermore, most included studies did not provide further discussion about the clinical use of these machine learning models, restricting their utilization in clinical practice and continuous improvement of these models. In addition to comparing machine learning models with traditional methods, future research should prioritize developing models based on commonly used clinical variables and validate these models across different hospitals to ensure their robustness and generalizability. Making these models publicly available and integrating them into electronic medical record systems would greatly assist nephrologists in making clinical decisions and help reduce patient risk. Collaboration between model developers, health care institutions, and electronic health record providers will be crucial to achieving this integration.

CONCLUSIONS

In conclusion, we found that machine learning models outperform traditional methods in predicting mortality, with better model discrimination for mortality than kidney function recovery. The application of deep learning models in AKI prognosis suggests they can effectively predict adverse outcomes with time changes, showing the potential to assist clinicians in identifying high-risk patients. Notably, these models need to be validated across diverse datasets and made publicly accessible to maximize their benefit for patients with AKI.

SUPPLEMENTARY MATERIAL

Supplementary File (PDF)

Item S1: Supplementary Text

Figure S1: The overview of decision tree

Figure S2: The overview of random forest

Figure S3: The overview of gradient boosting machine

Figure S4: The overview of support vector machine

Figure S5: The overview of Multilayer perceptron

Figure S6: Search strategies

Figure S7: Meta-analyses of machine learning models and traditional methods in predicting mortality in AKI patients

Figure S8: Meta-analyses of machine learning models and traditional methods in predicting kidney function recovery in AKI patients

Figure S9: The Funnel plots and Egger's test of meta-analyses

Table S1: Quality Assessment of Included Studies From Different Aspects.

Table S2: Risk of Bias and Application of Prediction Models Assessment According to the PROBAST Tool.

ARTICLE INFORMATION

Authors' Full Names and Academic Degrees: Yu Lin, PhD, Tongyue Shi, MS, and Guilan Kong, PhD

Authors' Affiliations: National Institute of Health Data Science, Peking University, Beijing, China (YL, TS, GK); Advanced Institute of Information Technology, Peking University, Hangzhou, China (YL, TS, GK); Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing, China (YL); and Institute for Artificial Intelligence, Peking University, Beijing, China (TS, GK).

Address for Correspondence: Guilan Kong, PhD, National Institute of Health Data Science, Peking University, 38 Xueyuan Road, Haidian District, Beijing 100191, China. Email: guilan.kong@hsc.pku.edu.cn

Authors' Contributions: Research area and study design: YL, GK; data acquisition: YL, TS; data analysis and interpretation: YL; supervision or mentorship: GK. Each author contributed important intellectual content during manuscript drafting or revision and accepts accountability for the overall work by ensuring that questions pertaining to the accuracy or integrity of any portion of the work are appropriately investigated and resolved.

Support: This work was supported by the National Natural Science Foundation of China under Grant No. 82372095; Zhejiang Provincial Natural Science Foundation of China under Grant No. LZ22F020014; Beijing Natural Science Foundation under Grant No. 7212201; and Humanities and Social Science Project of the Ministry of Education of China under Grant No. 22YJA630036.

Financial Disclosure: The authors declare that they have no relevant financial interests.

Acknowledgments: We would like to express our sincere gratitude to Dr Xin Zhang from the Renal Division, Department of Medicine, Peking University First Hospital, for his invaluable contribution to the interpretation of the clinical implications of machine learning models, which enriched the clinical relevance of our work.

Data Sharing: Data supporting the findings are publicly available in the manuscript of each study (see references 29-31 and 36-59).

Peer Review: Received June 10, 2024 as a submission to the expedited consideration track with 2 external peer reviews. Direct editorial input from the Editor-in-Chief. Accepted in revised form October 9, 2024.

REFERENCES

1. Hoste EAJ, Bagshaw SM, Bellomo R, et al. Epidemiology of acute kidney injury in critically ill patients: the multinational AKI-EPI study. *Intensive Care Med*. 2015;41(8):1411-1423.
2. Srisawat N, Sileanu FE, Murugan R, et al. Variation in risk and mortality of acute kidney injury in critically ill patients: a multicenter study. *Am J Nephrol*. 2015;41(1):81-88.
3. Shirakabe A, Hata N, Kobayashi N, et al. Long-term prognostic impact after acute kidney injury in patients with acute heart failure. *Int Heart J*. 2012;53(5):313-319.
4. Wu VC, Wu CH, Huang TM, et al. Long-term risk of coronary events after AKI. *J Am Soc Nephrol*. 2014;25(3):595-605.
5. James MT, Ghali WA, Knudtson ML, et al. Associations between acute kidney injury and cardiovascular and renal outcomes after coronary angiography. *Circulation*. 2011;123(4):409-416.
6. Ostermann M, Bellomo R, Burdmann EA, et al. Controversies in acute kidney injury: conclusions from a Kidney Disease: Improving Global Outcomes (KDIGO) Conference. *Kidney Int*. 2020;98(2):294-309.
7. Hodgson LE, Sarnowski A, Roderick PJ, Dimitrov BD, Venn RM, Forni LG. Systematic review of prognostic prediction models for acute kidney injury (AKI) in general hospital populations. *BMJ Open*. 2017;7(9):e016591.
8. Huang CY, Grandas FG, Flechet M, Meyfroidt G. Clinical prediction models for acute kidney injury. *Rev Bras Ter Intensiva*. 2020;32(1):123-132.
9. Huen SC, Parikh CR. Predicting acute kidney injury after cardiac surgery: a systematic review. *Ann Thorac Surg*. 2012;93(1):337-347.
10. Wilson T, Quan S, Cheema K, et al. Risk prediction models for acute kidney injury following major noncardiac surgery: systematic review. *Nephrol Dial Transplant*. 2016;31(2):231-240.
11. Allen DW, Ma B, Leung KC, et al. Risk prediction models for contrast-induced acute kidney injury accompanying cardiac catheterization: systematic review and meta-analysis. *Can J Cardiol*. 2017;33(6):724-736.
12. Caragata R, Wyssusek KH, Kruger P. Acute kidney injury following liver transplantation: a systematic review of published predictive models. *Anaesth Intensive Care*. 2016;44(2):251-261.
13. Silver SA, Shah PM, Chertow GM, Harel S, Wald R, Harel Z. Risk prediction models for contrast induced nephropathy: systematic review. *BMJ*. 2015;351:h4395.
14. Safari S, Yousefifard M, Hashemi B, et al. The role of scoring systems and urine dipstick in prediction of rhabdomyolysis-induced acute kidney injury: a systematic review. *Iran J Kidney Dis*. 2016;10(3):101-106.
15. Bajaj T, Koyner JL. Artificial intelligence in acute kidney injury prediction. *Adv Chronic Kidney Dis*. 2022;29(5):450-460.
16. Park S, Lee H. Acute kidney injury prediction models: current concepts and future strategies. *Curr Opin Nephrol Hypertens*. 2019;28(6):552-559.
17. Song X, Liu X, Liu F, Wang C. Comparison of machine learning and logistic regression models in predicting acute kidney injury: a systematic review and meta-analysis. *Int J Med Inform*. 2021;151:104484.
18. Xiao Z, Huang Q, Yang Y, et al. Emerging early diagnostic methods for acute kidney injury. *Theranostics*. 2022;12(6):2963-2986.
19. Feng Y, Wang AY, Jun M, et al. Characterization of risk prediction models for acute kidney injury: a systematic review and meta-analysis. *JAMA Netw Open*. 2023;6(5):e2313359.
20. Van Acker P, Van Biesen W, Nagler EV, Koobasi M, Veys N, Vanmassenhove J. Risk prediction models for acute kidney injury in adults: An overview of systematic reviews. *PLoS One*. 2021;16(4):e0248899.
21. Tomašev 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.
22. Lim B, van der Schaar M. Disease-atlas: Navigating disease trajectories using deep learning. Paper presented at: Machine Learning for Healthcare Conference; August 16-18, 2018; Stanford, CA.
23. Kellum JA, Bihorac A. Artificial intelligence to predict AKI: is it a breakthrough? *Nat Rev Nephrol*. 2019;15(11):663-664.
24. Khwaja A. KDIGO clinical practice guidelines for acute kidney injury. *Nephron Clin Pract*. 2012;120(4):c179-c184.
25. Gaudry S, Hajage D, Benichou N, et al. Delayed versus early initiation of renal replacement therapy for severe acute kidney injury: a systematic review and individual patient data meta-analysis of randomised clinical trials. *Lancet*. 2020;395(10235):1506-1515.
26. Li X, Liu C, Mao Z, Li Q, Zhou F. Timing of renal replacement therapy initiation for acute kidney injury in critically ill patients: a systematic review of randomized clinical trials with meta-analysis and trial sequential analysis. *Crit Care*. 2021;25(1):15.
27. STARRT-AKI Investigators, Canadian Critical Care Trials Group, Australian and New Zealand Intensive Care Society Clinical Trials Group, et al. Timing of initiation of renal-replacement therapy in acute kidney injury. *N Engl J Med*. 2020;383(3):240-251.
28. Gaudry S, Hajage D, Schortgen F, et al. Initiation strategies for renal-replacement therapy in the intensive care unit. *N Engl J Med*. 2016;375(2):122-133.
29. Fan Z, Jiang J, Xiao C, et al. Construction and validation of prognostic models in critically Ill patients with sepsis-associated acute kidney injury: interpretable machine learning approach. *J Transl Med*. 2023;21(1):406.
30. Huang CY, Güiza F, De Vlieger G, et al. Development and validation of clinical prediction models for acute kidney injury recovery at hospital discharge in critically ill adults. *J Clin Monit Comput*. 2023;37(1):113-125.
31. Hung PS, Lin PR, Hsu HH, Huang YC, Wu SH, Kor CT. Explainable machine learning-based risk prediction model for in-hospital mortality after continuous renal replacement therapy initiation. *Diagnostics (Basel)*. 2022;12(6):1496.
32. King Z, Farrington J, Utley M, et al. Machine learning for real-time aggregated prediction of hospital admission for emergency patients. *NPJ Digit Med*. 2022;5(1):104.
33. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*. 2021;372:n71.
34. Debray TPA, Damen JAAG, Snell KIE, et al. A guide to systematic review and meta-analysis of prediction model performance. *BMJ*. 2017;356:i6460.
35. Haredasht FN, Vanhoutte L, Vens C, Pottel H, Viaene L, De Corte W. Validated risk prediction models for outcomes of acute kidney injury: a systematic review. *BMC Nephrol*. 2023;24(1):133.
36. He J, Lin J, Duan M. Application of machine learning to predict acute kidney disease in patients with sepsis associated acute kidney injury. *Front Med (Lausanne)*. 2021;8:792974.
37. Liu LJ, Ortiz-Soriano V, Neyra JA, Chen J. KIT-LSTM: knowledge-guided time-aware LSTM for continuous clinical risk prediction. *Proceedings (IEEE Int Conf Bioinformatics Biomed)*. 2022;2022:1086-1091.

38. Liu LJ, Ortiz-Soriano V, Neyra JA, Chen J. KGDAL: knowledge graph guided double attention LSTM for rolling mortality prediction for AKI-D patients. *ACM BCB*. 2021;2021:53.
39. Yang J, Peng H, Luo Y, Zhu T, Xie L. Explainable ensemble machine learning model for prediction of 28-day mortality risk in patients with sepsis-associated acute kidney injury. *Front Med (Lausanne)*. 2023;10:1165129.
40. Chang HH, Chiang JH, Wang CS, et al. Predicting mortality using machine learning algorithms in patients who require renal replacement therapy in the critical care unit. *J Clin Med*. 2022;11(18):5289.
41. Liu J, Wu J, Liu S, Li M, Hu K, Li K. Predicting mortality of patients with acute kidney injury in the ICU using XGBoost model. *PLoS One*. 2021;16(2):e0246306.
42. Luo XQ, Yan P, Duan SB, et al. Development and validation of machine learning models for real-time mortality prediction in critically ill patients with sepsis-associated acute kidney injury. *Front Med (Lausanne)*. 2022;9:853102.
43. Huang H, Liu Y, Wu M, Gao Y, Yu X. Development and validation of a risk stratification model for predicting the mortality of acute kidney injury in critical care patients. *Ann Transl Med*. 2021;9(4):323.
44. Cunha VS, Salgado CM, Vieira SM, Sousa JMC. Fuzzy modeling to predict short and long-term mortality among patients with acute kidney injury. Presented at: 2016 IEEE International Conference On Fuzzy Systems (FUZZ-IEEE); July 24-29, 2016; Vancouver, BC, Canada.
45. Huang CJ, Wu B, Li X, et al. Predicting 1-year mortality of acute kidney injury: A risk model using electronic health records. Paper presented at: International Forum on Medical Imaging in Asia 2021; January 24-27, 2021; Taipei, Taiwan.
46. Tang J, Huang J, He X, et al. The prediction of in-hospital mortality in elderly patients with sepsis-associated acute kidney injury utilizing machine learning models. *Heliyon*. 2024;10(4):e26570.
47. Li X, Wu R, Zhao W, et al. Machine learning algorithm to predict mortality in critically ill patients with sepsis-associated acute kidney injury. *Sci Rep*. 2023;13(1):5223.
48. Lin K, Hu Y, Kong G. Predicting in-hospital mortality of patients with acute kidney injury in the ICU using random forest model. *Int J Med Inform*. 2019;125:55-61.
49. Zhou H, Liu L, Zhao Q, et al. Machine learning for the prediction of all-cause mortality in patients with sepsis-associated acute kidney injury during hospitalization. *Front Immunol*. 2023;14:1140755.
50. Neyra JA, Ortiz-Soriano V, Liu LJ, et al. Prediction of mortality and major adverse kidney events in critically ill patients with acute kidney injury. *Am J Kidney Dis*. 2023;81(1):36-47.
51. Nateghi Haredasht F, Viaene L, Pottel H, De Corte W, Vens C. Predicting outcomes of acute kidney injury in critically ill patients using machine learning. *Sci Rep*. 2023;13(1):9864.
52. Pike F, Murugan R, Keener C, et al. Biomarker enhanced risk prediction for adverse outcomes in critically ill patients receiving RRT. *Clin J Am Soc Nephrol*. 2015;10(8):1332-1339.
53. Wei W, Cai Z, Chen L, Yuan W, Fan Y, Rong S. Short-term prognostic models for severe acute kidney injury patients receiving prolonged intermittent renal replacement therapy based on machine learning. *BMC Med Inform Decis Mak*. 2023;23(1):133.
54. Wu C, Zhang Y, Nie S, et al. Predicting in-hospital outcomes of patients with acute kidney injury. *Nat Commun*. 2023;14(1):3739.
55. Liu CL, Tain YL, Lin YC, Hsu CN. Prediction and clinically important factors of acute kidney injury non-recovery. *Front Med (Lausanne)*. 2021;8:789874.
56. Lee BJ, Hsu CY, Parikh R, et al. Predicting renal recovery after dialysis-requiring acute kidney injury. *Kidney Int Rep*. 2019;4(4):571-581.
57. Zhao X, Lu Y, Li S, et al. Predicting renal function recovery and short-term reversibility among acute kidney injury patients in the ICU: comparison of machine learning methods and conventional regression. *Ren Fail*. 2022;44(1):1326-1337.
58. Low S, Vathsala A, Murali TM, et al. Electronic health records accurately predict renal replacement therapy in acute kidney injury. *BMC Nephrol*. 2019;20(1):32.
59. Pattharanitima P, Vaid A, Jaladanki SK, et al. Comparison of approaches for prediction of renal replacement therapy-free survival in patients with acute kidney injury. *Blood Purif*. 2021;50(4-5):621-627.
60. Fernandez-Felix BM, López-Alcalde J, Roqué M, Muriel A, Zamora J. CHARMS and PROBAST at your fingertips: a template for data extraction and risk of bias assessment in systematic reviews of predictive models. *BMC Med Res Methodol*. 2023;23(1):44.
61. Wolff RF, Moons KGM, Riley RD, et al. PROBAST: a tool to assess the risk of bias and applicability of prediction model studies. *Ann Intern Med*. 2019;170(1):51-58.
62. Shen A, Wei X, Zhu F, et al. Risk prediction models for breast cancer-related lymphedema: a systematic review and meta-analysis. *Eur J Oncol Nurs*. 2023;64:102326.
63. Zhang Z, Yang L, Han W, et al. Machine learning prediction models for gestational diabetes mellitus: meta-analysis. *J Med Internet Res*. 2022;24(3):e26634.
64. Liu JQ, Cai GY, Liang S, et al. Characteristics of and risk factors for death in elderly patients with acute kidney injury: a multicentre retrospective study in China. *Postgrad Med J*. 2018;94(1111):249-253.
65. Duan S, Li Y, Yang P. Predictive value of blood urea nitrogen in heart failure: a systematic review and meta-analysis. *Front Cardiovasc Med*. 2023;10:1189884.
66. Hong C, Zhu H, Zhou X, et al. Association of blood urea nitrogen with cardiovascular diseases and all-cause mortality in USA adults: results from NHANES 1999-2006. *Nutrients*. 2023;15(2):461.
67. Chen X, Yang Q, Gao L, et al. Association between serum anion gap and mortality in critically ill patients with COPD in ICU: data from the MIMIC IV database. *Int J Chron Obstruct Pulmon Dis*. 2024;19:579-587.
68. Gameiro J, Lopes JA. Complete blood count in acute kidney injury prediction: a narrative review. *Ann Intensive Care*. 2019;9(1):87.
69. Efros O, Beckerman P, Basson AA, et al. Fluctuations in serum creatinine levels during hospitalization and long-term end-stage kidney disease and mortality. *JAMA Netw Open*. 2023;6(8):e2326996.
70. Wang TJ, Lin CH, Wei HJ, Wu MJ. Long-term outcomes and risk factors of renal failure requiring dialysis after heart transplantation: a nationwide cohort study. *J Clin Med*. 2020;9(8):2455.
71. Bouchard J, Mehta RL. Timing of kidney support therapy in acute kidney injury: what are we waiting for? *Am J Kidney Dis*. 2022;79(3):417-426.
72. Ozrazgat-Baslanti T, Loftus TJ, Ren Y, Ruppert MM, Bihorac A. Advances in artificial intelligence and deep learning systems in ICU-related acute kidney injury. *Curr Opin Crit Care*. 2021;27(6):560-572.
73. Vagliano I, Chesnaye NC, Leopold JH, Jager KJ, Abu-Hanna A, Schut MC. Machine learning models for predicting acute kidney injury: a systematic review and critical appraisal. *Clin Kidney J*. 2022;15(12):2266-2280.

74. Srisawat N, Wen X, Lee M, et al. Urinary biomarkers and renal recovery in critically ill patients with renal support. *Clin J Am Soc Nephrol*. 2011;6(8):1815-1823.
75. Srisawat N, Murugan R, Lee M, et al. Plasma neutrophil gelatinase-associated lipocalin predicts recovery from acute kidney injury following community-acquired pneumonia. *Kidney Int*. 2011;80(5):545-552.
76. Fiorentino M, Tohme FA, Murugan R, Kellum JA. Plasma biomarkers in predicting renal recovery from acute kidney injury in critically ill patients. *Blood Purif*. 2019;48(3):253-261.
77. Dewitte A, Joannès-Boyau O, Sidobre C, et al. Kinetic eGFR and novel AKI biomarkers to predict renal recovery. *Clin J Am Soc Nephrol*. 2015;10(11):1900-1910.
78. Hsu CY, Chertow GM, McCulloch CE, Fan D, Ordoñez JD, Go AS. Nonrecovery of kidney function and death after acute on chronic renal failure. *Clin J Am Soc Nephrol*. 2009;4(5):891-898.
79. Cerdá J, Liu KD, Cruz DN, et al. Promoting kidney function recovery in patients with AKI requiring RRT. *Clin J Am Soc Nephrol*. 2015;10(10):1859-1867.
80. Hickson LJ, Chaudhary S, Williams AW, et al. Predictors of outpatient kidney function recovery among patients who initiate hemodialysis in the hospital. *Am J Kidney Dis*. 2015;65(4):592-602.
81. Zhao X, Li C, Lu Y, et al. Characteristics and risk factors for renal recovery after acute kidney injury in critically ill patients in cohorts of elderly and non-elderly: a multicenter retrospective cohort study. *Ren Fail*. 2023;45(1):2166531.
82. Klein SJ, Brandtner AK, Lehner GF, et al. Biomarkers for prediction of renal replacement therapy in acute kidney injury: a systematic review and meta-analysis. *Intensive Care Med*. 2018;44(3):323-336.
83. Zhou XJ, Rakheja D, Yu X, Saxena R, Vaziri ND, Silva FG. The aging kidney. *Kidney Int*. 2008;74(6):710-720.
84. Chang-Panesso M. Acute kidney injury and aging. *Pediatr Nephrol*. 2021;36(10):2997-3006.
85. Delahunt B, Stehbens WE, Gilbert-Barness E, Shozawa T, Rüger BM. Progeria kidney has abnormal mesangial collagen distribution. *Pediatr Nephrol*. 2000;15(3-4):279-285.
86. Franklin SS, Gustin W 4th, Wong ND, et al. Hemodynamic patterns of age-related changes in blood pressure. The Framingham Heart Study. *Circulation*. 1997;96(1):308-315.
87. Chirinos JA, Segers P, Hughes T, Townsend R. Large-artery stiffness in health and disease: JACC state-of-the-art review. *J Am Coll Cardiol*. 2019;74(9):1237-1263.
88. Kellum JA, Sileanu FE, Bihorac A, Hoste EAJ, Chawla LS. Recovery after acute kidney injury. *Am J Respir Crit Care Med*. 2017;195(6):784-791.
89. Bagshaw SM. Epidemiology of renal recovery after acute renal failure. *Curr Opin Crit Care*. 2006;12(6):544-550.
90. Gameiro J, Marques F, Lopes JA. Long-term consequences of acute kidney injury: a narrative review. *Clin Kidney J*. 2021;14(3):789-804.
91. Forni LG, Darmon M, Ostermann M, et al. Renal recovery after acute kidney injury. *Intensive Care Med*. 2017;43(6):855-866.