A systematic review of artificial intelligence algorithms for predicting acute kidney injury

M.R. BACCI1,2, C.V.B. MINCZUK1, F.L.A. FONSECA2

1Nephrology Department, Centro Universitário Faculdade de Medicina do ABC, Santo Andre, Brazil
2Department of Clinical Analysis Laboratory, Centro Universitário Faculdade de Medicina do ABC, Santo Andre, Brazil

Abstract. – OBJECTIVE: Acute kidney injury (AKI) increases mortality and costs in hospitalized patients. New methods for early AKI identification have been developed with targeted biomarkers and electronic health records data analysis. Machine learning (ML) use in diagnostics and health data analysis has recently increased. We performed a systematic review to analyze the use of ML for AKI prediction in hospitalized adults.

MATERIALS AND METHODS: Tubmed, EMBASE, Cochrane, and Web of Science databases were searched until 31st March of 2023. English-language studies using ML in adults for AKI prediction were included using predetermined eligibility search terms such as acute kidney injury, machine learning, and artificial intelligence. Two reviewers evaluated the publications' titles, abstracts, and full texts separately and obtained appropriate data. The main outcome was an area under the curve (AUC) result of at least 0.70.

RESULTS: Ten studies in 102 articles were included involving 242,251 patients. Deep learning (AUC 0.907 in critical care AKI; AUC 0.797 in hospitalized patients AKI) was similar to Logistic regression (AUC 0.877 in critical care AKI; AUC 0.789 in hospitalized patients). Decision tree constructions had similar AUC.

CONCLUSIONS: In this review, most ML models analyzed fulfilled the main outcome. AKI is multifactorial; however, ML performed well with different etiologies, such as cardiac-related AKI, drug-related AKI, and critical care patients. Overfitting data and constructing black box models are limitations that might jeopardize the generalization and comprehension of the results. Most studies were single-center, and three manuscripts used the same database with a predominantly Caucasian population, resulting in a lack of diversity and reducing external generalization. In conclusion, ML could effectively predict AKI in hospitalized adults. Future directions rely on including a more diverse population and completing prospective and controlled trials.

Key Words: Acute kidney injury, Machine learning, Artificial intelligence, Logistic regression, Biomarkers.

Introduction

Acute kidney injury (AKI) is a serious illness that worsens the prognosis for hospitalized patients1. AKI lengthens the patient’s hospital stay and increases mortality risk2. To offer an accurate and consistent diagnosis and stage for AKI, many criteria have been proposed, although there is a significant discrepancy among them3. The accepted approach to stage kidney impairment is the estimated glomerular function (eGFR)4. Nonetheless, equations provide a clear picture of stable patients, in contrast to the critical care AKI scenario, which includes unstable patients. The biomarker for diagnosing, staging AKI, and estimating GFR is serum creatinine (SCr); however, its levels are influenced by sex, and the individual’s muscle degree, and a rise does not occur in AKI’s early stages4. Cystatin C is a better biomarker than SCr but is more expensive and rarely used, which makes it difficult for a comparison using electronic datasets. AKI early detection influences patients’ outcomes, the length of their treatment, and the dosage of their medications5. Personalized patient stratification using targeted biomarkers is feasible for the early detection of AKI and for creating prediction scores based on AKI’s etiology with early-stage focused treatment4-7.

An area of artificial intelligence (AI) called machine learning (ML) is devoted to understanding and creating learning techniques that use data to improve task performance. ML algorithms build a model based on sample data, also called classification models, to provide predictions or choices without being explicitly instructed7. ML use is promising to study subjects in outcome prediction, diagnosis6, and image interpretation in the health sciences, which have also been used for AKI prediction8. To the matter of concern using ML to predict AKI, there is huge variability in the studies’ methods shown so far, which use various training variables and different electronic databases. No comprehensive review of AKI predic-
tion with all published information has yet been done. We conducted a thorough analysis of the development of AKI with ML prediction models in adult hospitalized patients.

**Materials and Methods**

**Electronic Search Databases**

Using combined MeSH keywords, two independent researchers examined four online databases: EMBASE, MEDLINE, Web of Science, and Cochrane Library from the beginning of the registries through March 31, 2023. When a disagreement occurred, a third independent researcher selected the manuscript. The report was made through the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) standards.

**Researching Selection and Methodology**

The title and the abstract were evaluated using the keywords “acute kidney injury” AND “machine learning” AND “artificial intelligence”. The AND conjoiner was chosen to narrow the search to these terms seen together only. The studies that were included in this review were selected based on the PICOS strategy: (1) Participants: patients with AKI of any cause; (2) Intervention: use of AI or ML to diagnose AKI; (3) Comparison: Kidney Disease Improving Global Outcomes (KDIGO) criteria served as the standard for AKI diagnostic, and the AI models developed were compared by their performance; (4) Outcomes: the primary outcome was the results of the Area Under the Receiver Operating Curves (ROC-AUC) above 0.700 to successfully diagnose AKI; (5) Design of studies: trials and randomized controlled trials. The 2012 KDIGO criteria classified AKI as a binary occurrence based on the SCr level measured at hospital admission. A 1.5-fold increase in SCr during a seven-day interval or a 0.3 mg/dL (26.5 umol/L) increase within 48 hours. AKI was not diagnosed based on urine output because of the absence of consistent data in the studies. The studies chosen for review did not include dialysis patients.

The search filters used consisted of selecting trials or clinical studies with adult patients, published in English, without animals in the sample, and excluding preprints. These filters resulted in the primary selection of 102 manuscripts.

**Extraction and Evaluation of Data**

The two reviewers retrieved the data separately after reading the abstract and article. The criteria used for selection involved evaluation of the methodology for diagnosing AKI based on KDIGO criteria, the nature of the AI model utilized, the type of Electronic Health Records (EHR) used, and the need to match AI models with AUC findings. All the ML models reported in the manuscripts were considered for the final discussion. Still, only those with the best AUC performance entered the discussion section. Logistic regression (LR), random forest (RF), eXtreme gradient boosting (XGBoost), decision tree (DT), support vector machine (SVC), naïve Bayes (NB), gradient boost machine (GBM), and deep learning (DL) were selected as the machine learning models. This work interpreted LR as an ML construction model without human selection of variables. The articles were eliminated if they used solely cluster analysis of AKI patients to compare ML models and had a percentage of missing data over 10%.

We evaluated the manuscript’s methodological quality using the first edition of the Risk Of Bias In Non-randomized Studies of Exposure tool (ROBINS-E) created by the Cochrane Collaboration. Specific information about the ROBINS-E is found in the supplemental methods section. Based on the risk of bias, manuscripts were categorized as high risk, moderate concerns, or low risk. Manuscripts having a higher likelihood of bias were omitted from the final decision. To properly analyze the weight of each variable in the final validation model, we picked studies that classified variables and used techniques to determine the presence of an imbalance in data distribution, such as the Synthetic Minority Oversampling Technique (SMOTE). The SHapley Additive exPlanations (SHAP) value evaluated internal observation of the black box of findings for each research, a tool for interpreting the influence of each variable and grading its usage in the supplied external outcome. This research required no submission to an ethics committee since all of the analyses were included in previously published articles. The methodology of this systematic review was previously registered with the number INPLASY20230025 on the International Platform of Registered Systematic Reviews and Meta-analysis Protocols (INPLASY).

**Results**

**Identification of Studies**

A total of 102 studies were identified by searching EMBASE, MEDLINE, Web of Science, and the Cochrane Library database. Among them, seventy-three studies were duplicates removed using the EndNote 20 duplicate removal tool.
Fifteen articles were removed after the review of the title and abstract sections. The final step had fourteen full-text manuscripts thoroughly reviewed, and four were removed due to the use of clustering without employing ML models and an excessive proportion of missing data. The remaining ten articles were included in the final analysis and shown in the diagram (Figure 1). The final sample size involved 242,251 AKI patients evaluated across the studies.

The ROBINS-E tool is shown in Figure 2 with the risk of bias for observational studies plots. All the studies were evaluated for the seven domains, and none had the final judgment of high concern that would have ended in exclusion.

Characteristics of Eligible Studies and ML Models Used

The data for this review was separated according to the number of patients evaluated in each study and the associated cause of AKI; if the study used data from a single institution or multicenter, the database source used for the model building, the ML model used for the comparison and the related AUC result from the calculation with the given outcome.

The principal AKI etiologies studied in this review involved cardiac-related AKI (cardiac surgery and acute myocardial infarction), AKI in hospitalized patients, critical care AKI, AKI related to the postoperative period, AKI in can-
A systematic review of artificial intelligence algorithms for predicting acute kidney injury

cancer patients treated with immune checkpoint inhibitors and AKI related to deceased liver donor transplantation (Table I). Cardiac-related AKI was the most frequent etiology found in this selection.

The ML models utilized in the research varied considerably, making their comparisons unfeasible due to the different AKI etiologies and variables used for training and validation. However, the GBM model in the multi-center evaluation had a solid AUC performance in detecting AKI stage 2 (AUC 0.810) and AKI stage 3 (AUC 0.870) after 48 hours of hospital admission in a general population14. Logistic regression was used in half of the studies evaluated in this review and achieved a good AUC result equal to 0.70 in all analyses. Deep learning was assessed by Alfieri et al15, Chua et al16, Thongprayoon et al17, and Yu et al18 using artificial neural networks (ANN) with the best AUC of 0.907 in the evaluation of hospitalized non-oliguric AKI stages 2 and 3 patients15. Support vector machines (SVM) are supervised models that analyze data from classification and regression. SVM had an AUC of 0.718 in AKI related to acute myocardial infarction19 and 0.824 in AKI related to immune checkpoint inhibitors in cancer patients18. In all cardiac-related AKI, the RF model, a decision tree derivative, had the highest AUC result20.

Electronic Health Records and AKI Incidence

Most of the studies used single-center data analysis. Four14,15,19,26 of the ten studies compared data from different health institutions. The Medical Information Mart for Intensive Care (MIMIC) III and MIMIC-IV electronic databases were the most used in three of the ten analyses. These databases have a predominantly white Caucasian population. The National Patient-Centered Research Network (PCORnet) was the most significant EHR for gathering information from twelve independent health systems from nine USA states14 and with a higher diversity compared to the others with the inclusion of Hispanics, Native Americans, and African Americans. Four databases represented Asian individuals, with evaluations conducted in Singapore, Taipei, and China.

Figure 2. ROBINS-E plot regarding overall quality of studies selected.
<table>
<thead>
<tr>
<th>Author, Year</th>
<th>Patients</th>
<th>Single/ Multi-center</th>
<th>AKI</th>
<th>Database</th>
<th>AKI Incidence (%)</th>
<th>ML model</th>
<th>AUC</th>
<th>Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alfieri et al 2022</td>
<td>10,596</td>
<td>Multi-center</td>
<td>ICU patients</td>
<td>MIMIC-III eICU</td>
<td>4.17</td>
<td>LR DL (ANN)</td>
<td>0.877 0.907</td>
<td>Development of oliguric stage 2/3 AKI</td>
</tr>
<tr>
<td>Bishara et al 2022</td>
<td>8,858</td>
<td>Single center</td>
<td>Post-operative patients</td>
<td>The University of California San Francisco</td>
<td>4.3</td>
<td>LR GBM</td>
<td>0.730 0.850</td>
<td>Development of AKI in the post-operative period</td>
</tr>
<tr>
<td>Cai et al 2022</td>
<td>3,882</td>
<td>Multi-center</td>
<td>AMI patients</td>
<td>MIMIC-III MIMIC-IV</td>
<td>29.4</td>
<td>SVM DT RF NB XGBoost</td>
<td>0.718 0.663 0.696 0.724 0.666</td>
<td>AKI mortality during hospitalization</td>
</tr>
<tr>
<td>Chua et al 2021</td>
<td>16,288</td>
<td>Single center</td>
<td>Hospitalized patients</td>
<td>National University Hospital, Singapore</td>
<td>4.4</td>
<td>LR DL (RNN)</td>
<td>0.789 0.797</td>
<td>Development of AKI after 48 h of admission</td>
</tr>
<tr>
<td>Cruz et al 2021</td>
<td>32,581</td>
<td>Multi-center</td>
<td>Cardiac-surgery related AKI</td>
<td>MIMIC-III Mount Sinai Health System</td>
<td>4.5</td>
<td>LR DT RF GBM</td>
<td>0.796 0.733 0.842 0.821</td>
<td>Development of Cardiac-surgery related AKI during hospitalization.</td>
</tr>
<tr>
<td>Song et al 2020</td>
<td>153,821</td>
<td>Multi-center</td>
<td>Hospitalized patients</td>
<td>PCORNet</td>
<td>14.2</td>
<td>GBM any AKI GBM AKI stage 2 GBM AKI stage 3</td>
<td>0.760 0.810 0.870</td>
<td>Development of AKI 48 h post admission</td>
</tr>
<tr>
<td>Thongprayoon et al 2022</td>
<td>13,158</td>
<td>Single center</td>
<td>Cardiac-surgery related AKI</td>
<td>Mayo Clinic</td>
<td>36</td>
<td>DT RF DL (ANN) XGBoost</td>
<td>0.640 0.780 0.750 0.770</td>
<td>Development of post-operative AKI</td>
</tr>
<tr>
<td>Tseng et al 2020</td>
<td>671</td>
<td>Single center</td>
<td>Cardiac-surgery related AKI</td>
<td>Far Eastern Memorial Hospital, New Taipei City</td>
<td>24.3</td>
<td>LR SVM XGBoost DT RF</td>
<td>0.806 0.825 0.837 0.781 0.839</td>
<td>Development of post-operative AKI</td>
</tr>
<tr>
<td>Zhang et al 2021</td>
<td>780</td>
<td>Single center</td>
<td>AKI after Liver Transplantation</td>
<td>University Lingnan Hospital, Lingnan</td>
<td>55.1</td>
<td>LR SVM RF GBM</td>
<td>0.730 0.750 0.750 0.760</td>
<td>Development of AKI pos deceased donor Liver Transplantation</td>
</tr>
<tr>
<td>Yu et al 2022</td>
<td>1,616</td>
<td>Single center</td>
<td>Immune Checkpoint Inhibitors related</td>
<td>PLA General Hospital, Beijing</td>
<td>6.9</td>
<td>SVM DT RF LR NB XGBoost DL (ANN)</td>
<td>0.824 0.776 0.816 0.720 0.813 0.786 0.786</td>
<td>Cancer patients with ICI AKI related to AKI</td>
</tr>
</tbody>
</table>

The incidence of AKI varied across the etiologies observed. The highest incidence occurred in liver deceased donor transplantation, with 55.1%. Cardiac-related AKI, when surgery was made, had an incidence of 36% and after acute myocardial infarction of 29.4%; however, in critical care patients with KDIGO stage 2 to 3 oliguric patients, the incidence was only 4.17%.

Discussion

In this systematic review evaluating AI models’ use in AKI prediction, we found AUC results above 0.70 in most studies analyzed using different ML tools. The best performance occurred with the ANN model, which was expected through the continuous learning process that this model experiences. Most studies16-18,25,27,28 used a single-center evaluation, and the main AKI etiology was the cardiac-associated AKI, involving post-operative cardiac surgery patients and acute myocardial infarction.

AKI is a serious condition that increases mortality by 77% and contributes to longer lengths of stay (LOS) in hospitalized patients1,21. The existence of risk factors such as male gender, heart disease, higher age, and black ethnicity was responsible for a worse outcome in AKI patients1,21. In this review, the studies did not evaluate mortality or LOS; however, AKI early identification remains an essential key to hospitalized patient-centered care. AKI definition evolved in the past decades from the Risk, Injury, Failure, Loss and End-stage renal disease (RIFLE) 2004 classification, the Acute Kidney Injury Network (AKIN) 2007 classification to the 2012 KDIGO AKI definition and classification22,23. All the studies used the same AKI definition2. Since the timing of AKI recognition is essential and serum creatinine rise occurs late, new identification methods have been developed in the search for early biomarkers, such as the ratio between the tissue inhibitor of metalloproteinases-2 (TIMP-2) and insulin-like growth factor-binding-protein 7 (IGFBP7)6 and ML to analyze a list of variables to AKI development prediction7.

Most trials20,24 used decision tree structures where each internal node or “leaf” represents an input feature. Many models are derived from this classification tree, such as DT, RF20, GBM, and XGBoost24. There is a prone to data overfitting in decision tree models such as DT, GBM, and XGBoost GBMs and XGBoost, particularly when utilizing a large number of boosting rounds or complex base learners. On the other hand, DL models such as ANN are computer systems inspired by the brain’s organic neural networks. The notion depends on inputs that yield a result, with self-adjustment based on a learning rule that uses these outputs as a learning curve. Once enough data is inserted, the model can choose the best information for the exposure and outcome chosen. The studies conducted by Alfieri15 (AUC 0.907), Chua16 (AUC 0.797), Thongprayoon17 (AUC 0.750), and Yu18 (AUC 0.786) used DL models with solid AKI prediction results through different etiologies. While these results sound promising, DL and RF models are considered “black box” models. This lack of interpretability could limit their use in healthcare settings, where understanding is critical. Moreover, they are computationally expensive, requiring powerful hardware and potentially longer training times. LR is a traditional and well-known method to analyze multiple variables in each outcome. In our analysis, the studies that used LR as one of the comparators had AUC results higher than 0.73015,16,18,25-28. Nonetheless, there is a perception that AI models are better than conventional statistical approaches that analyze numerous variables, such as LR. In this systematic review, LR was as good as ML models regarding the AUC results reported for AKI prediction29,30. Comparing models from research with diverse etiologies of AKI is another challenge that must be answered. As SCr is a late AKI biomarker and urine debt requires a minimum of six hours of monitoring, the accurate determination of the time of AKI onset is crucial when known. This explains why so much research exists on cardiac-related AKI and post-operative AKI. Following the previous statement, we had four cardiac-related AKI studies27,19,26,27 and two with post-operative AKI after non-cardiac surgeries28,30.

The strength of this review relies on gathering ten studies with 242,251 patients from three continents (North America, Europe, and Asia) with the inclusion of different AKI etiologies. Moreover, we correlated ML models and DL neural network development with a consistent final result. Limitations were not few and revealed the use of the same electronic database in some of the analyses, such as the MIMIC III31 and MIMIC-IV32, resulting in the inclusion of a majority white and Caucasian population, reducing the diversity of population and generalization of the findings in three studies35,36,37. Using different variables led to
inconsistent AUC findings using the same model across trials. The heterogeneity in these findings was anticipated because the trials were examined across various AKI etiologies, and the factors may not have been the same\textsuperscript{19}. LR, conversely, benefits from human intervention and knowledge of the choice of the variables\textsuperscript{30}. There is always the risk of human bias selection, indeed. Still, as experts, researchers tend to be more accurate in choosing variables when using the knowledge of pathophysiology and specific exposure and outcome. Moreover, few studies\textsuperscript{14} performed a multi-center evaluation in the same country (USA); the rest studied the single-center hospital database.

**Conclusions**

In summary, the findings of this systematic review, which included ten studies, show that machine learning models are highly effective at predicting cardiac-related-AKI, post-operative-related-AKI, AKI in hospitalized patients, and medication-associated-AKI using the 2012 KDIGO criteria. All the studies conducted a retrospective analysis of the data, and the development of prospective and controlled trials with ML to AKI prediction is needed. Further enhancement of electronic hospital databases, the definition of standard variables in the analysis, and the inclusion of a more diverse population are critical future directions to reduce the burden of AKI.

**Funding**

No funding was used for this work.

**Authors’ Contributions**

MRB and CVBM performed the choice of the keywords and the search of the manuscripts; MRB performed the writing of the manuscript; FLAF reviewed the final choice and the final draft of the manuscript. All the authors performed the final proofreading and revision of the final version of the manuscript.

**ORCID ID**

MRB: 0000-0001-8578-8404
CVBM: 0000-0001-8419-3740
FLAF: 0000-0003-1223-1589

**Conflict of Interest**

The authors declare there are no conflicts of interest related to this manuscript.

**Data Availability**

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

**Ethics Approval**

This study did not require the approval of the Institutional Review Board.

**Informed Consent**

Not applicable.

**References**


