Archives of Academic Emergency Medicine. 2025; 13(1): e45

OPEN ACCESS

ORIGINAL RESEARCH

An Ensemble Machine Learning Model for Early Prediction of Vancomycin-Induced Acute Kidney Injury in ICU Patients

Faezeh Aghamirzaei[1], Ahmad Ali Abin[1]*, Farzaneh Futuhi[2]

1. Faculty of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran.
2. Department of Nephrology, Shahid Beheshti University of Medical Science, Tehran, Iran.

Received: February 2025; Accepted: March 2025; Published online: 15 April 2025

Abstract: **Introduction:** Acute Kidney Injury (AKI) is a severe complication of vancomycin treatment due to its nephrotoxic effects. However, research on predicting AKI in this high-risk group remains limited. This study presents a stacking ensemble machine learning model designed to predict the onset of AKI in this patient population. **Methods:** Leveraging data from 314 ICU patients, the model incorporates SHapley Additive exPlanations (SHAP) for enhanced interpretability, identifying key predictors such as serum creatinine levels, glucose variability, and patient age. The model achieved an Area Under the Curve (AUC) of 0.94, outperforming existing predictive approaches. By utilizing readily available clinical data and determining an optimal temporal prediction window, this model facilitates proactive clinical decision-making, aiming to reduce the risk of AKI and improve patient outcomes. **Results:** The stacking ensemble model achieved 92% accuracy, 93% precision, 92% sensitivity, and 0.94 AUC in 314 ICU patients, pinpointing creatinine, glucose variability, and age as critical AKI predictors. **Conclusion:** The findings suggest that integrating advanced machine learning techniques with interpretable artificial intelligence (AI) can provide a scalable and cost-effective solution for early AKI detection in diverse healthcare settings.

Keywords: Acute Kidney Injury; Vancomycin; Machine Learning; Prognosis, Intensive Care Unit

Cite this article as: Aghamirzaei F, Abin AA, Futuhi F. An Ensemble Machine Learning Model for Early Prediction of Vancomycin-Induced Acute Kidney Injury in ICU Patients. Arch Acad Emerg Med. 2025; 13(1): e45. https://doi.org/10.22037/aaemj.v13i1.2560.

1. Introduction

Acute Kidney Injury (AKI) is a sudden loss of renal function linked to numerous acute and chronic comorbidities. Even mild AKI episodes may lead to chronic kidney disease, while severe or recurrent cases can progress to end-stage renal disease, significantly increasing mortality and reducing quality of life (1-3). In intensive care units (ICUs), AKI is highly prevalent, affecting nearly 50% of patients (4). AKI is characterized by a sudden rise in serum creatinine levels and a decrease in urine volume (5). The duration of AKI strongly correlates with survival rates, highlighting the need for timely prediction and intervention.

At present, Vancomycin is primarily used in the treatment of infections caused by methicillin-resistant Staphylococcus aureus (MRSA). However, its use is associated with significant AKI occurrences (6, 7); resulting in high mortality, chronic kidney disease, and extended hospital stays (8). Vancomycin's nephrotoxicity is influenced by multiple factors, including dosing regimen and concurrent use of other nephrotoxic agents. The pathophysiology includes direct tubular toxicity and interstitial nephritis, with elevated trough levels and prolonged therapy as key risk factors for renal damage (9). There is considerable variation in the reported incidence of Vancomycin-associated AKI (VA-AKI), which ranges from 5 to 43% (10). Despite the significant risk of AKI in vancomycin-treated patients, research specifically targeting this subgroup remains limited.

An effective AKI predictor must be minimally invasive, widely accessible, cost-effective, and easily integrated into routine clinical workflows. However, accurately predicting AKI in vancomycin-treated patients remains challenging. Traditional clinical markers of kidney injury, such as serum creatinine, often only exhibit substantial changes after considerable kidney damage has occurred, thereby limiting their predictive value. While novel biomarkers and risk scores show promise, their limited validation and lack of widespread implementation reduce their clinical utility (11).

Artificial intelligence (AI) refers to systems that simulate human intelligence, often through statistical methods or machine learning (12). Machine learning, a subset of AI, enables systems to learn from data and improve performance without explicit programming. In healthcare, machine learning has shown great promise in enhancing predictive accuracy and improving patient outcomes, particularly in mobile health monitoring platforms (13). Additionally, consensus clustering classification methods have been used to identify

* **Corresponding Author:** Ahmad Ali Abin; Faculty of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran. Phone number: (+98) 912 725 3488. Email: a_abin@sbu.ac.ir, ORCID: https://orcid.org/0000-0002-0916-0348.

This open-access article distributed under the terms of the Creative Commons Attribution NonCommercial 3.0 License (CC BY-NC 3.0).
Downloaded from: https://journals.sbmu.ac.ir/aaem/index.php/AAEM/index

disease subclasses and predict outcomes, further strengthening the robustness of healthcare predictive models (14). Recent advances in ML highlight its transformative impact across specialties, from anesthesiology to pain management optimization (15-18).

For example, Wu et al. developed an ensemble machine learning model to predict AKI in neuro intensive care unit (NICU) patients following brain surgery. The model achieved strong early predictive performance for postoperative AKI risk (19). Similarly, Xunliang Li et al. developed an Extreme Gradient Boosting (XGBoost) model to predict in-hospital mortality among patients with sepsis-associated AKI (SA-AKI). The model achieved an AUC of 0.794, demonstrating its effectiveness in mortality risk prediction (20).

Despite advancements in AKI prediction, most machine-learning models are designed for broad ICU populations and overlook the complexities of vancomycin-induced nephrotoxicity (10). Vancomycin-induced AKI is primarily driven by oxidative stress and tubular damage, mechanisms that necessitate a targeted predictive approach (9). Only a limited number of studies have concentrated on developing predictive models tailored to vancomycin-induced AKI. For example, a model employing XGBoost achieved moderate predictive performance, with AUC values ranging from 0.85 to 0.88, highlighting key risk factors including serum creatinine levels and vancomycin trough concentrations (21). Another study used causal modeling to explore the link between vancomycin exposure and AKI onset, achieving an AUC of 0.83. While these studies mark substantial progress, they often lack the level of predictive accuracy and interpretability required for widespread clinical integration. Moreover, simpler models such as risk-scoring systems and decision trees have been developed to provide practical tools for clinicians; however, these approaches typically yield AUC values below 0.85, reflecting limitations in both precision and adaptability across diverse patient populations (22, 23). There is a critical need for a robust and interpretable predictive model that enhances accuracy while addressing the unique dynamics of vancomycin-induced nephrotoxicity. Such a model would support personalized and timely clinical decision-making. This study addresses the gap by developing a machine-learning model tailored for vancomycin-treated patients in an ICU setting. The model incorporates key clinical features, including 'Max Creatinine,' 'Max Phosphate,' and 'Age,' which are critical indicators of kidney function and AKI risk, particularly for nephrotoxic drug exposure (11).

2. Methods

2.1. Study design and setting

Initially, data were collected from ICU patients who received vancomycin treatment at Loghman Hakim Hospital in Tehran, Iran, between November 2021 and December 2022. The study aimed to develop a machine-learning model to predict AKI in patients treated with vancomycin. This study strictly adhered to ethical guidelines. The experimental protocols and methodologies employed were rigorously reviewed and approved by the Iranian Research Institute for Information Science and Technology (IranDoc), under approval number 2952802.

2.2. Participants

The initial dataset included 430 ICU patients treated with vancomycin. Inclusion and exclusion criteria were applied to refine the cohort. Adult patients aged 18 to 90 years who received vancomycin for at least three days were included. patients with pre-existing chronic kidney disease (CKD), a glomerular filtration rate (GFR) below 60 mL/min/1.73m², or evidence of proteinuria were excluded to focus on patients at risk for new-onset AKI. Additionally, patients outside the specified age range or treated with vancomycin for fewer than three days, as well as records with more than 20% missing data in critical variables, were excluded to ensure robust analysis.

Following these criteria, the final cohort consisted of 314 patients, providing a focused representation of ICU patients at risk for vancomycin-induced AKI. This refined dataset offered a solid foundation for developing and validating the machine learning models aimed at predicting AKI, leveraging both demographic and clinical features gathered during patients' ICU stay.

2.3. Data gathering

Data collection involved recording static and dynamic clinical features from vancomycin-treated ICU patients to predict AKI. Static features included age, gender, body mass index (BMI), and comorbidities, which remained constant during the ICU stay. Dynamic features consisted of daily clinical and laboratory measurements, including creatinine levels, anion gap, phosphate, red cell distribution width (RDW), and partial thromboplastin time (PTT). Data were collected starting from the first day of vancomycin treatment and continued up to 24 hours before the onset of AKI. This temporal window was selected to focus on predictive features that precede AKI onset, excluding data from the final 24 hours to avoid potential overlap with diagnostic markers of AKI.

2.4. Data Preprocessing

Data preprocessing was conducted to ensure the dataset's quality and suitability for machine learning model development. This process included managing missing data, aggregating dynamic features into static metrics, and selecting the most relevant features for training the models. One-Hot Encoding:

Categorical variables, such as gender and urine color, were converted into numerical formats using one-hot encoding. This method preserved the categorical nature of the data while ensuring compatibility with machine learning models.

Handling Missing Data:

To maintain dataset quality and robustness, records with

This open-access article distributed under the terms of the Creative Commons Attribution NonCommercial 3.0 License (CC BY-NC 3.0).
Downloaded from: https://journals.sbmu.ac.ir/aaem/index.php/AAEM/index

more than 20% missing values in critical variables were excluded. For the remaining records, missing values were imputed using the K-nearest neighbors (KNN) algorithm, which estimated missing values based on patient similarities while preserving underlying data patterns and clinical relevance.
Feature Aggregation and Temporal Feature Extraction:
Temporal feature aggregation was used to transform dynamic time-series data into fixed-length feature vectors suitable for machine learning models. Each dynamic feature was summarized into static metrics by calculating several statistical descriptors:
• Mean: Represented the average feature value over the observation period, reflecting overall levels.
• Maximum and Minimum: Captured the range of variability, highlighting peaks and troughs that could signal clinical risk thresholds.
• Skewness and Kurtosis: Described the shape of the feature distribution. Skewness reflected data asymmetry, while kurtosis identified extreme values and outliers.
• Slope: Indicated feature trends over time, such as a rising slope in creatinine levels, which might suggest worsening kidney function as an early marker of AKI.
This process preserved temporal patterns and trends from dynamic features in a format compatible with machine learning models requiring fixed-length inputs. Figure 5 illustrates this approach with an example of a patient treated with vancomycin for seven days, showing AKI onset on the seventh day. By converting dynamic measurements into static vectors, the model could better identify relevant patterns for AKI prediction.
Feature Selection:
The initial dataset included 193 features, which were refined through a rigorous feature selection process to enhance model efficiency and interpretability. Methods such as Pearson correlation, chi-square tests, Recursive Feature Elimination (RFE), and wrapper-based approaches (e.g., Lasso, LightGBM, Random Forest) were applied. Features selected by at least three methods were retained, resulting in 62 key features. This multi-method approach reduced dimensionality and minimized the risk of overfitting. Table 1 provides a summary of the selected features and the methods used for their selection.

2.5. Data Labeling

The labeling process categorized patients based on AKI development using the Kidney Disease Improving Global Outcomes (KDIGO) criteria (5). According to KDIGO, AKI is defined as an increase in serum creatinine by 0.3 mg/dL within 48 hours or a rise to 1.5 times the baseline within the prior seven days. These criteria were systematically applied to determine whether each patient experienced AKI during their ICU stay. Patients meeting any of these conditions were labeled as AKI-positive, while those who did not were classified as AKI-negative.

2.6. Statistical Analysis

Numeric features were presented as medians with interquartile ranges, while categorical features were reported as counts with percentages. The final column shows p-values from the Mann-Whitney U test, highlighting statistically significant differences in feature distributions between the two groups. A low p-value indicates a significant variation in the feature between AKI and Non-AKI patients

2.7. Model Development and Evaluation

To develop predictive models for AKI risk, six machine learning algorithms were evaluated:
• Random Forest (RF): An ensemble learning algorithm that builds multiple decision trees during training. Each tree contributes to the final prediction through majority voting, which helps mitigate overfitting and captures intricate feature interactions (24). RF is particularly well-suited for identifying key predictors in healthcare applications due to its robustness and interpretability (25). Figure 1 illustrates the structure of the Random Forest model used in this study.
• Neural Networks (NN): These models mimic the structure of the human brain through layers of interconnected "neurons" capable of learning complex, non-linear patterns. By processing data through these layers, NNs effectively capture relationships that are difficult for traditional models to identify (26, 27). Figure 2 shows the architecture of the Neural Network model.
• XGBoost: Known for its efficiency and high performance, this gradient-boosting algorithm constructs sequential trees, each correcting the errors of its predecessors. XGBoost excels in handling diverse data types, missing values, and outliers, making it a preferred choice for clinical prediction tasks (28). XGBoost has demonstrated superior performance in kidney disease prediction studies, consistently outperforming other algorithms due to its robustness and high classification precision (29). Figure 3 provides a schematic representation of the XGBoost workflow. • CatBoost: Optimized for datasets with significant categorical variables, CatBoost employs advanced encoding methods to convert non-numeric data into numerical representations without loss of information. This model is particularly effective in handling medical datasets where categorical features such as "Urine Color" and "Urine Appearance" are crucial (30).
• LightGBM: A gradient boosting algorithm that adopts a unique leaf-wise growth strategy, which improves computational efficiency and enhances predictive performance on high-dimensional data (31). Its scalability and capacity to handle diverse data types make it a valuable tool for analyzing complex patient data (32). As shown in Figure 4, this leaf-wise growth approach selectively grows leaves with the highest loss reduction, contrasting with the traditional level-wise growth strategy.
• Extra Trees (ET): This ensemble method creates multiple decision trees with added randomness in selecting split points, improving generalization and reducing variance. ET

This open-access article distributed under the terms of the Creative Commons Attribution NonCommercial 3.0 License (CC BY-NC 3.0).
Downloaded from: https://journals.sbmu.ac.ir/aaem/index.php/AAEM/index

is effective in identifying important clinical features influencing AKI risk (33-35).

Following initial training, models demonstrating strong predictive performance were further refined through Random Search. This hyperparameter optimization method efficiently samples combinations of parameters in high-dimensional spaces, offering a computationally effective alternative to exhaustive Grid Search (36).

Hyperparameters optimized for each model included:

• Extra Trees Classifier: Parameters such as maximum features, minimum samples split, and maximum depth.

• LightGBM, XGBoost, and CatBoost: Learning rate, maximum depth, boosting rounds, and regularization factors.

• Random Forest: Parameters including maximum features, maximum depth, and minimum samples split.

• Neural Networks: Adjustments to learning rate, number of hidden layers, neurons per layer, activation functions, and dropout rates. To ensure robust performance evaluation, a 10-fold cross-validation technique was applied. This approach partitioned the dataset into ten subsets, where each subset served as the test set once, and the remaining subsets were used for training. This methodology provided a comprehensive assessment of model performance and reduced risks of overfitting and data imbalance, challenges prevalent in ICU datasets. Recent research in critical care has underscored the importance of such validation strategies. For example, Wu et al. demonstrated the efficacy of 10-fold cross-validation in achieving reliable predictions for AKI in neuro-intensive care patients (19).

2.8. Screening performance characteristics of the model

In the model development and evaluation process, several key metrics were utilized to assess the performance of the machine learning models. These metrics include accuracy, precision, sensitivity, specificity, F1 score, and the Area Under the Curve (AUC) (37).

• Accuracy was defined as the proportion of true results (both true positives and true negatives) among the total number of cases examined.

• Precision, or the positive predictive value, is the proportion of true positives among the total predicted positives.

• Sensitivity, also known as recall or true positive rate, is the proportion of actual positives correctly identified.

• The F1 score is the harmonic mean of precision and sensitivity, providing a balanced measure between these two metrics. The formula for F1 is mathematically represented as the equation below.

F1 Score = 2 × (Precision × sensitivity) / (Precision + sensitivity)

• Specificity is the proportion of actual negative cases that are correctly identified as negative by the model. It indicates the model's effectiveness in avoiding false positives.

• The Area Under the Curve (AUC) refers to the area under the ROC (Receiver Operating Characteristic) curve. This metric is employed to evaluate the performance of a binary classification system, with values ranging from 0.5 (no better than random guessing) to 1.0 (perfect classifier).

3. Results

314 patients with a median age of 65.0 (IQR: 52.0-76.0) years were studied (38.25% male). 159 (50.63%) cases developed AKI. Table 2 summarizes the baseline characteristics of the study population, detailing differences between patients who developed AKI and those who did not.

The Extra Trees classifier emerged as the best-performing model, achieving an accuracy of 90%, a precision of 91%, and an Area Under the Curve (AUC) of 91%. It demonstrated a sensitivity of 89% and a specificity of 88%, highlighting its balanced performance in accurately detecting AKI cases while effectively minimizing false positives. Additionally, its F1 score of 90% underscores the model's overall reliability and robustness. The optimal hyperparameters for this model were found to be a max feature value of 0.96, a minimum sample split of 36, and a max depth of 20. Table 3 provides a summary of model performance metrics, including 95% confidence intervals for each metric.

Despite the standout performance of the Extra Trees model, other models also achieved notable results. This encouraged further exploration using a stacking ensemble approach, combining the strengths of these individual models to potentially enhance predictive performance. Ensemble learning techniques, like stacking, have been shown to improve prediction results by leveraging multiple model strengths, particularly in healthcare settings where data complexity and variability can impact individual model performance (38).

Optimal Hyperparameters for Each Model:

• Extra Trees: max feature = 0.96, min sample split = 36, max depth = 20

• LightGBM: learning rate = 0.1, num leaves = 1598, min data in leaf = 36

• XGBoost: max depth = 9, min child weight = 3, subsample = 0.49

• CatBoost: learning rate = 0.2, depth = 9, l2 leaf reg = 3.13, rsm = 0.17

• Random Forest: max features = 0.21, max depth = 18, min samples split = 13

• Neural Network: learning rate = 0.01, dense layer size = 64

To further explore potential improvements, a stacking ensemble model was developed, where individual models served as first-level learners, while LightGBM acted as the second-level meta-model. This stacking ensemble approach demonstrated superior performance, delivering an accuracy of 92%, precision of 93%, sensitivity of 92%, specificity of 90%, an F1 score of 92%, and an AUC of 94%. This balance of sensitivity and specificity underscores the model's robustness in detecting AKI cases while minimizing false positives. The improvement in metrics highlights the effectiveness of combining models to leverage the individual strengths of each, thereby boosting overall predictive power (39). The en-

This open-access article distributed under the terms of the Creative Commons Attribution NonCommercial 3.0 License (CC BY-NC 3.0).
Downloaded from: https://journals.sbmu.ac.ir/aaem/index.php/AAEM/index

semble model's AUC score, with a 95% confidence interval of 0.92 to 0.96, further supports its reliability. Such validation is critical for clinical application, ensuring that the model generalizes well beyond the training data and maintains consistent performance across diverse patient populations.

The ensemble model's Receiver Operating Characteristic (ROC) curve, depicted in Figure 6, demonstrates an AUC of 0.94. This result reflects the model's strong ability to distinguish between AKI-positive and AKI-negative cases across various thresholds, reinforcing its potential for reliable early risk stratification in clinical settings.

4. Discussion

The findings of our study demonstrate the potential of machine learning, particularly a stacking ensemble approach, in accurately predicting AKI in patients treated with vancomycin.

In our in-depth analysis of the model's interpretability, we utilized SHAP (SHapley Additive exPlanations) to better understand the features contributing to AKI prediction in patients undergoing Vancomycin therapy (40). The SHAP results provided a clear breakdown of how different clinical and laboratory features influence the model's predictions. Among the most influential were maximum creatinine levels, glucose variability (measured by the kurtosis of glucose), and patient age. These results align well with clinical expectations, as high creatinine levels directly indicate impaired kidney function, while glucose variability can signal metabolic stress that impacts renal health. Age also stands out as a significant risk factor, since older patients are more prone to reduced renal reserve and impaired kidney perfusion. Incorporating these features into our model not only improves predictive accuracy but also offers clinicians practical insights that can help tailor patient care. Figure 7 shows a SHAP summary plot highlighting the relative importance of these features.

The plot visually conveys the importance of these features in predicting AKI. The plot shows that "Max Creatinine" was the most significant predictor, followed by other features such as "Max Phosphate" and "Age." This distribution of feature importance underscores the multifactorial nature of AKI, suggesting that the model considers a complex interplay of physiological factors when assessing a patient's risk. The integration of SHAP values enhances the transparency of the model, allowing clinicians to understand not only the likelihood of AKI occurrence but also the underlying factors contributing to each prediction. This interpretability is essential for fostering trust in the model's decisions and for supporting clinical decision-making.

The prediction of vancomycin-induced AKI using machine learning has been relatively underexplored, with only a few studies addressing this specific subgroup. For instance, Kim et al. developed a risk-scoring system utilizing multivariate logistic regression, elastic net, random forest, and support vector machine models, achieving an AUC of approximately 0.735 (22). While foundational, their approach relied heavily on traditional statistical methods and simpler machine learning models that lacked the capacity to capture the complex interactions inherent in ICU data. Similarly, Mu et al. employed the XGBoost algorithm to predict vancomycin-induced AKI across patients with varying underlying conditions, reporting AUC values between 0.85 and 0.88 (21). Despite these moderate successes, their model's reliance on specialized biomarkers, such as vancomycin trough levels, limits its feasibility in resource-constrained settings.

In contrast, our study adopted a stacking ensemble approach that combines multiple machine learning models to leverage their individual strengths and overcome the limitations of single-model methods. This strategy yielded a significantly higher AUC of 94%, demonstrating superior predictive performance compared to previously reported models. The effectiveness of ensemble learning in handling complex clinical prediction challenges, where individual models often falter, is evident in these results.

Additionally, unlike prior studies that required expensive and specialized tests, such as vancomycin trough concentration measurements, our model was built using routinely collected laboratory data. These data are readily available in most healthcare facilities, making the model more practical and applicable across diverse healthcare settings. This design choice ensures greater scalability, cost-effectiveness, and broader clinical impact, particularly in public hospitals where budget constraints limit access to costly biomarkers.

Beyond enhanced accuracy and broader applicability, our study uniquely addresses the need for model interpretability. Machine learning models often face skepticism from clinicians due to their "black-box" nature, which can limit trust in predictions. By incorporating SHapley Additive exPlanations (SHAP), our model offers clear insights into the contribution of individual features to predictions. Key features such as "Max Creatinine," "Max Phosphate," and "Age" were identified as significant predictors of AKI risk, aligning with established clinical knowledge. This transparency fosters clinician trust and supports the integration of the model into clinical workflows, where understanding the rationale behind predictions is essential for informed decision-making.

We embarked on a series of experiments to test the impact of the duration between data observation and the onset of AKI on model performance. Our primary motivation was to investigate how manipulating the temporal window of data observation impacts the predictive performance of our models. AKI is a dynamic condition, and its onset and progression could be influenced by various factors over time. The temporal window for observation, therefore, plays a critical role in determining the accuracy and reliability of our models.

Initially, we leveraged a broad temporal window—from the first day of vancomycin treatment until 24 hours before the onset of AKI (see Figure 8-a). This approach was based on the assumption that the entirety of the patient's medical history would provide a comprehensive view, enabling the ac-

This open-access article distributed under the terms of the Creative Commons Attribution NonCommercial 3.0 License (CC BY-NC 3.0).
Downloaded from: https://journals.sbmu.ac.ir/aaem/index.php/AAEM/index

curate prediction of AKI. To ensure consistent comparisons across all experiments, we used the CatBoost model, which achieved an accuracy of 85%, precision of 88%, recall of 83%, an F1 score of 85%, and an AUC of 89%.

We then extended the time window of prediction to 48 hours before the onset of AKI and observed a slight decrease in model performance (see Figure 8-b). The CatBoost model achieved an accuracy of 85%, precision of 88%, recall of 83%, an F1 score of 85%, and an AUC of 89%. Extending the prediction window further to 72 hours before AKI onset resulted in a more significant performance drop (see Figure 8-c). This time, the CatBoost model's accuracy fell to 76%, precision to 79%, recall to 75%, F1 score to 77%, and AUC to 78%. This substantial decrease in model performance indicated that essential predictive information might be missed when extending the data collection window too far from the onset of AKI.

These findings prompted an exploration of the last few days before AKI onset. We restricted our data observation window to between 72 and 24 hours before AKI onset (see Figure 8-d), hypothesizing that crucial data might be contained within this timeframe. Indeed, the results supported our hypothesis, as the CatBoost model achieved an accuracy of 84%, precision of 85%, recall of 85%, an F1 score of 85%, and an AUC of 88%. This notable improvement in performance compared to the 72-hour prediction window underscored the importance of this immediate pre-onset period, supporting our belief that crucial physiological changes indicative of impending AKI may become most apparent within this time.

To further evaluate the model's performance in a different temporal context, we conducted an additional experiment by restricting the observation window to the first three days of vancomycin treatment. This approach was aimed at determining whether early lab values alone could provide sufficient predictive power for AKI while enhancing the model's practical utility in clinical settings. Using the CatBoost model, we observed an accuracy of 78%, precision of 80%, recall of 76%, an F1 score of 78%, and an AUC of 83%. While these results demonstrated moderate predictive performance, they were lower compared to models trained on longer temporal windows, particularly those closer to AKI onset. This decline in performance reflects the importance of capturing dynamic changes, such as creatinine trends, that occur over time and are critical for accurate AKI prediction.

These experiments suggest that the best predictive performance was achieved with data collected from the first day of vancomycin treatment until 24 hours before AKI onset. Notably, data collected in the 72-24 hours preceding AKI onset was found to be of critical importance to the predictive ability of the models. This underlines the significance of monitoring these patients closely in the days leading up to AKI onset while also highlighting the potential for early prediction using limited initial data.

This study presents a robust stacking ensemble machine-learning model specifically designed to predict AKI in patients treated with vancomycin. Our model achieved high predictive accuracy, interpretability, and cost-effectiveness, making it well-suited for diverse healthcare settings, including those with limited resources. Unlike previous studies that relied heavily on specialized biomarkers, our approach utilizes routinely collected clinical data, enhancing the model's accessibility and scalability.

The relatively underexplored nature of vancomycin-induced AKI prediction highlights the significance of our work in filling a critical gap in the current literature. By focusing on widely available data and identifying an optimal temporal window for intervention, our model supports early and proactive clinical decision-making, ultimately helping to reduce the incidence of AKI and improve patient outcomes. Additionally, the integration of SHAP values ensures transparency and facilitates clinician trust, which is vital for the successful adoption of machine learning tools in healthcare. Future work should validate this model across multiple healthcare settings and extend its application to other nephrotoxic agents, potentially broadening its clinical impact. The advancements presented in this study mark an important step towards more personalized, data-driven healthcare in critical care environments, offering a practical solution for the early detection and prevention of AKI in high-risk patients.

5. Limitations

Despite promising results, our study has limitations. The dataset was derived from a single institution, which may limit the generalizability of our findings. Additionally, the relatively small sample size of 314 patients may limit the power of the conclusions and the generalizability of the findings. Larger-scale validation across multiple institutions is needed to enhance the robustness of the model.

6. Conclusions

The findings suggest that integrating advanced machine learning techniques with interpretable AI can provide a scalable and cost-effective solution for early AKI detection in diverse healthcare settings.

7. Declarations

7.1. Acknowledgments

None.

7.2. Ethical Approval

This study strictly adhered to ethical guidelines. The experimental protocols and methodologies employed were rigorously reviewed and approved by the Iranian Research Institute for Information Science and Technology (IranDoc), under approval number 2952802.

This open-access article distributed under the terms of the Creative Commons Attribution NonCommercial 3.0 License (CC BY-NC 3.0).
Downloaded from: https://journals.sbmu.ac.ir/aaem/index.php/AAEM/index

7.3. Consent to Participate

The authors confirm that informed consent was obtained from all participants. Patients admitted to the ICU primarily post-neurosurgery were thoroughly informed about the scope and intent of this study. Explicit consent was procured from each patient allowing the research team to access and utilize pertinent medical information which encompasses their demographic details, type of ailment, associated comorbidities, and diagnostic test results. If patients' post-surgery and during their ICU stay met the inclusion criteria, their medical data was meticulously documented for the purposes of this research.

7.4. Consent to Publish

Consent to publish was obtained from all participants prior to their inclusion in the study. Patients were informed that their anonymized data would be used for publication purposes and that no identifying information would be disclosed.

7.5. Funding

The authors did not receive support from any organization for the submitted work.

7.6. Competing Interests

The authors declare no competing interests.

7.7. Author contributions statement

Faeze Aghamirzaei was the principal author who conducted the experiments, gathered data, and prepared the original draft. Ahmad Ali Abin conceptualized the problem, supervised the research project, and reviewed and edited the manuscript. Farzaneh Futuhi conceived the original idea, defined the problem, validated the results, and reviewed the manuscript. All authors read and approved the final manuscript.

7.8. Availability of data and materials

The datasets generated during the current study are available from the corresponding author upon reasonable request.

7.9. Using artificial intelligence chatbots

This study did not use artificial intelligence chatbots.

References

1. Chen Y-W, Wu M-Y, Mao C-H, Yeh Y-T, Chen T-T, Liao C-T, et al. Severe acute kidney disease is associated with worse kidney outcome among acute kidney injury patients. Scientific Reports. 2022;12(1).
2. Kellum JA, Romagnani P, Ashuntantang G, Ronco C, Zarbock A, Anders H-J. Acute kidney injury. Nature Reviews Disease Primers. 2021;7(1):52.
3. Pickkers P, Darmon M, Hoste E, Joannidis M, Legrand M, Ostermann M, et al. Acute kidney injury in the critically ill: an updated review on pathophysiology and management. Intensive Care Medicine. 2021;47(8):835-50.
4. Ronco C, Bellomo R, Kellum JA. Acute kidney injury. The Lancet. 2019;394(10212):1949-64.
5. Khwaja A. KDIGO Clinical Practice Guidelines for Acute Kidney Injury. Nephron Clinical Practice. 2012;120(4):c179-c84.
6. Liu C, Bayer A, Cosgrove SE, Daum RS, Fridkin SK, Gorwitz RJ, et al. Clinical Practice Guidelines by the Infectious Diseases Society of America for the Treatment of Methicillin-Resistant Staphylococcus aureus Infections in Adults and Children. Clinical Infectious Diseases. 2011;52(3):e18-e55.
7. Kwak S, Kim JY, Cho H. Vancomycin-induced nephrotoxicity in non-intensive care unit pediatric patients. Scientific Reports. 2021;11(1).
8. Yang Z, Wang C, Wang H, Wang S, Liu R, Wang X, et al. Cross-sectional survey on adult acute kidney injury in Chinese ICU: the study protocol (CARE-AKI). BMJ Open. 2018;8(6):e020766.
9. Filippone E, Kraft W, Farber J. The Nephrotoxicity of Vancomycin. Clinical Pharmacology amp; Therapeutics. 2017;102(3):459-69.
10. Van Hal SJ, Paterson DL, Lodise TP. Systematic Review and Meta-Analysis of Vancomycin-Induced Nephrotoxicity Associated with Dosing Schedules That Maintain Troughs between 15 and 20 Milligrams per Liter. Antimicrobial Agents and Chemotherapy. 2013;57(2):734-44.
11. Notice. Kidney International Supplements. 2012;2(1):1.
12. Bindra S, Jain R. Artificial intelligence in medical science: a review. Irish journal of medical science. 2023;193.
13. Boursalie O, Samavi R, Doyle TE. Machine Learning and Mobile Health Monitoring Platforms: A Case Study on Research and Implementation Challenges. Journal of Healthcare Informatics Research. 2018;2(1):179-203.
14. Alyousef AA, Nihtyanova S, Denton C, Bosoni P, Bellazzi R, Tucker A. Nearest Consensus Clustering Classification to Identify Subclasses and Predict Disease. Journal of Healthcare Informatics Research. 2018;2(4):402-22.
15. Hashemi S, Yousefzadeh Z, Abin AA, Ejmalian A, Nabavi S, Dabbagh A. Machine Learning-Guided Anesthesiology: A Review of Recent Advances and Clinical Applications. Journal of Cellular & Molecular Anesthesia. 2024;9(1).
16. Nikkinen O, Kolehmainen T, Aaltonen T, Jämsä E, Alahuhta S, Vakkala M. Developing a supervised machine learning model for predicting perioperative acute kidney injury in arthroplasty patients. Computers in Biology and Medicine. 2022;144:105351.
17. Gao W, Wang J, Zhou L, Luo Q, Lao Y, Lyu H, et al. Prediction of acute kidney injury in ICU with gradient boosting decision tree algorithms. Computers in Biology and Medicine. 2022;140:105097.
18. Ejmalian A, Aghaei A, Nabavi S, Abedzadeh Darabad M, Tajbakhsh A, Abin AA, et al. Prediction of Acute Kidney

This open-access article distributed under the terms of the Creative Commons Attribution NonCommercial 3.0 License (CC BY-NC 3.0).
Downloaded from: https://journals.sbmu.ac.ir/aaem/index.php/AAEM/index

Injury After Cardiac Surgery Using Interpretable Machine Learning. Anesth Pain Med. 2022;12(4):e127140.
19. Wu M, Jiang X, Du K, Xu Y, Zhang W. Ensemble machine learning algorithm for predicting acute kidney injury in patients admitted to the neurointensive care unit following brain surgery. Scientific Reports. 2023;13(1).
20. Li X, Wu R, Zhao W, Shi R, Zhu Y, Wang Z, et al. Machine learning algorithm to predict mortality in critically ill patients with sepsis-associated acute kidney injury. Scientific Reports. 2023;13(1).
21. Mu F, Cui C, Tang M, Guo G, Zhang H, Ge J, et al. Analysis of a machine learning–based risk stratification scheme for acute kidney injury in vancomycin. Frontiers in Pharmacology. 2022;13.
22. Kim JY, Kim KY, Yee J, Gwak HS. Risk Scoring System for Vancomycin-Associated Acute Kidney Injury. Frontiers in Pharmacology. 2022;13.
23. Miyai T, Imai S, Kashiwagi H, Sato Y, Kadomura S, Yoshida K, et al. A Risk Prediction Flowchart of Vancomycin-Induced Acute Kidney Injury to Use When Starting Vancomycin Administration: A Multicenter Retrospective Study. Antibiotics. 2020;9(12):920.
24. Biau G. Analysis of a random forests model. J Mach Learn Res. 2012;13(null):1063–95.
25. Yadav D. Prediction of Heart Disease Using Feature Selection and Random Forest Ensemble Method. International Journal for Pharmaceutical Research Scholars. 2020;12:56-66.
26. Yu Z, Wang K, Wan Z, Xie S, Lv Z. Popular deep learning algorithms for disease prediction: a review. Cluster Computing. 2023;26(2):1231-51.
27. Lisboa PJ, Taktak AFG. The use of artificial neural networks in decision support in cancer: A systematic review. Neural Networks. 2006;19(4):408-15.
28. Ogunleye A, Wang QG. XGBoost Model for Chronic Kidney Disease Diagnosis. IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2020;17(6):2131-40.
29. Raihan MJ, Khan MA-M, Kee S-H, Nahid A-A. Detection of the chronic kidney disease using XGBoost classifier and explaining the influence of the attributes on the model using SHAP. Scientific Reports. 2023;13(1).
30. Bentéjac C, Csörgő A, Martínez-Muñoz G. A comparative analysis of gradient boosting algorithms. Artificial Intelligence Review. 2021;54(3):1937-67.
31. Zhang D, Gong Y. The Comparison of LightGBM and XGBoost Coupling Factor Analysis and Prediagnosis of Acute Liver Failure. IEEE Access. 2020;8:220990-1003.
32. Rufo DD, Debelee TG, Ibenthal A, Negera WG. Diagnosis of Diabetes Mellitus Using Gradient Boosting Machine (LightGBM). Diagnostics. 2021;11(9):1714.
33. Sharma D, Gujral R, Jain A. Breast cancer prediction based on neural networks and extra tree classifier using feature ensemble learning. Measurement: Sensors. 2022;24:100560.
34. Shafique R, Mehmood A, Ullah DS, Choi GS. Cardiovascular Disease Prediction System Using Extra Trees Classifier2019.
35. Hameed M, Khalid M, Khaleel F, Al-Ansari N. An Extra Tree Regression Model for Discharge Coefficient Prediction: Novel, Practical Applications in the Hydraulic Sector and Future Research Directions. Mathematical Problems in Engineering. 2021;2021:1-19.
36. Bergstra J, Bengio Y. Random search for hyper-parameter optimization. J Mach Learn Res. 2012;13(null):281–305.
37. Vujovic Z. Classification Model Evaluation Metrics. International Journal of Advanced Computer Science and Applications. 2021;Volume 12:599-606.
38. Yu K, Xie X. Predicting Hospital Readmission: A Joint Ensemble-Learning Model. IEEE Journal of Biomedical and Health Informatics. 2020;24(2):447-56.
39. Mahajan P, Uddin S, Hajati F, Moni MA. Ensemble Learning for Disease Prediction: A Review. Healthcare. 2023;11(12):1808.
40. Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Proceedings of the 31st International Conference on Neural Information Processing Systems; Long Beach, California, USA: Curran Associates Inc.; 2017. p. 4768–77.

This open-access article distributed under the terms of the Creative Commons Attribution NonCommercial 3.0 License (CC BY-NC 3.0).
Downloaded from: https://journals.sbmu.ac.ir/aaem/index.php/AAEM/index

Table 1: Overview of feature selection results

Number	Feature	Pearson Correlation	Chi-square	RFE	Lasso	Random Forest	LightGBM	Occurrence Count
1	Slope - Creatinine	×	×	×	×	×	×	6
2	Max - Phosphate	×	×	×	×	×	×	6
3	Age	×	×	×	×	×	×	6
4	Skew - Platelet Count	×	×	×	×	×	×	6
5	Max - RDW	×	×	×	×	×	×	6
6	Mean - Creatinine	×	×	×	×	×		5
7	Mean - RDW	×	×	×		×	×	5
8	Slope - Bicarbonate	×	×	×	×	×		5
9	Mean - Anion Gap	×	×	×		×	×	5
10	Max - Urea Nitrogen	×	×	×		×	×	5
11	Min - PTT	×	×	×	×	×		5
12	Max - Anion Gap	×	×	×	×	×		5
13	Slope - Urea Nitrogen	×	×	×		×	×	5
14	Slope - Anion Gap	×	×	×		×		4
15	Skew - Urea Nitrogen	×		×		×	×	4
16	Mean - Platelet Count	×		×	×	×		4
17	Mean - Phosphate	×	×	×	×			4
18	Max - Protein	×	×			×	×	4
19	Max - Urine Appearance	×	×	×	×			4
20	Mean - Urobilinogen	×	×			×	×	4
21	Min - RDW	×	×	×		×		4
22	Mean - Magnesium	×	×	×		×		4
23	Mean - Calcium	×		×		×	×	4
24	Slope - pCO2	×	×	×	×			4
25	Max - pCO2	×	×	×	×			4

The table illustrates the result of each feature selection method applied in the study. Each row corresponds to one of the top 25 features, and each column corresponds to a specific feature selection method. The '×' symbol indicates that the corresponding feature was selected by the respective method. The "Occurrence Count" column represents the total number of methods by which each feature was selected and the more the better.

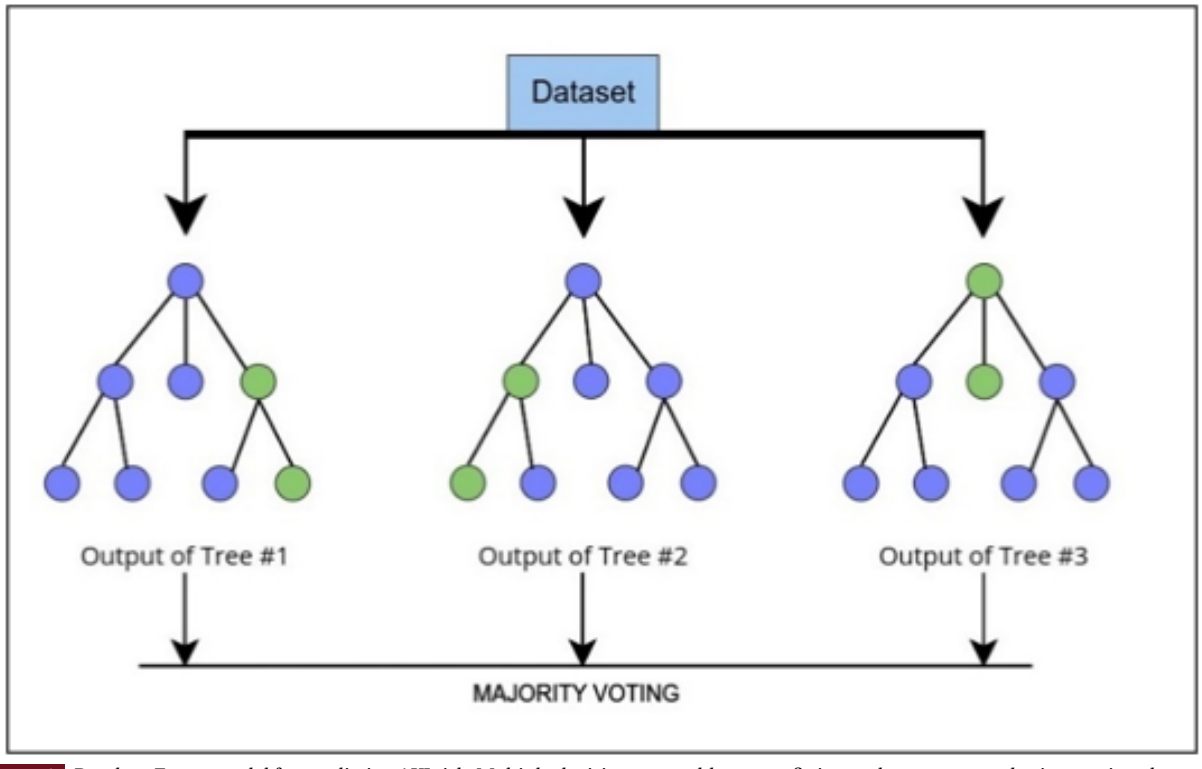

Figure 1: Random Forest model for predicting AKI risk. Multiple decision trees address overfitting and capture complex interactions between features such as creatinine levels and patient age.

This open-access article distributed under the terms of the Creative Commons Attribution NonCommercial 3.0 License (CC BY-NC 3.0).
Downloaded from: https://journals.sbmu.ac.ir/aaem/index.php/AAEM/index

Table 2: Baseline characteristics of patients by acute kidney injury status

Features	Total (n = 314)	Acute kidney injury		P
		No (n =155)	Yes (n = 159)	
Age (year)				
Median (IQR)	65.0 (52.0-76.0)	65.0 (52.0-77.0)	66.0 (54.75-74.25)	0.708
Gender				
Male	120 (38.25)	61 (39.35)	59 (37.12)	0.7
Laboratory features				
Slope Creatinine	0.01 (-0.03-0.1)	-0.02 (-0.05-0.01)	0.08 (0.01-0.18)	<0.001
Slope Bicarbonate	0.21 (-0.35-0.73)	0.37 (-0.17-0.98)	-0.02 (-0.46-0.61)	0.003
Skew Platelet Count	0.28 (-0.11-0.66)	0.17 (-0.17-0.52)	0.36 (0.01-0.71)	0.008
Mean RDW	15.33 (14.26-17.18)	14.93 (13.95-16.18)	15.92 (14.65-18.33)	<0.001
Mean Creatinine	1.07 (0.74-1.59)	0.83 (0.56-1.15)	1.36 (0.98-2.06)	<0.001
Mean Anion Gap	13.14 (11.52-14.75)	12.4 (11.1-13.74)	13.74 (12.06-15.62)	<0.001
Max RDW	15.9 (14.8-17.9)	15.5 (14.4-16.7)	16.35 (15.1-19.35)	<0.001
Max Creatinine	1.3 (1.0-2.2)	1.0 (0.7-1.35)	1.9 (1.3-2.8)	<0.001
Max Anion Gap	16.0 (14.0-18.0)	15.0 (14.0-17.0)	17.0 (15.0-19.0)	<0.001
Slope Urea Nitrogen	0.27 (-0.79-2.52)	-0.18 (-1.19-0.88)	1.3 (-0.44-4.53)	<0.001
Slope Anion Gap	-0.04 (-0.4-0.29)	-0.12 (-0.48-0.12)	0.04 (-0.3-0.4)	0.001
Skew pH	-0.7 (-1.5-0.0)	-0.71 (-1.65-0.12)	-0.7 (-1.27-0.0)	0.037
Min PTT	28.7 (25.55-32.8)	28.0 (24.85-30.35)	29.35 (26.08-34.92)	0.001
Mean Phosphate	3.4 (2.93-3.99)	3.2 (2.82-3.56)	3.71 (3.14-4.33)	<0.001
Mean PTT	34.1 (28.36-42.9)	32.6 (27.32-37.41)	35.81 (29.87-48.13)	<0.001
Mean Leukocytes	0.0 (0.0-0.5)	0.0 (0.0-0.21)	0.0 (0.0-0.5)	0.001
Max Urea Nitrogen	34.0 (22.0-54.5)	27.0 (19.5-41.0)	45.5 (29.0-69.5)	<0.001
Max Phosphate	4.4 (3.7-5.25)	4.0 (3.5-4.5)	4.85 (4.0-5.9)	<0.001
Max Leukocytes	0.0 (0.0-0.5)	0.0 (0.0-0.5)	0.0 (0.0-1.0)	<0.001
Kurt MCHC	-0.92 (-1.28-0.49)	-1.03 (-1.32-0.62)	-0.83 (-1.25-0.3)	0.01
Slope Phosphate	0.05 (-0.06-0.2)	0.05 (-0.07-0.1)	0.07 (-0.03-0.27)	0.005
Slope MCH	-0.02 (-0.1-0.07)	-0.04 (-0.13-0.05)	0.02 (-0.08-0.1)	0.011
Min Urobilinogen	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.0 (0.0-0.0)	0.168
Min RDW	14.8 (13.9-16.35)	14.6 (13.65-15.6)	15.3 (14.2-16.92)	0.001
Min Phosphate	2.5 (2.0-3.0)	2.4 (1.9-2.8)	2.6 (2.2-3.1)	0.002
Mean Urobilinogen	0.0 (0.0-0.2)	0.0 (0.0-0.2)	0.0 (0.0-0.12)	0.629
Mean Platelet Count	226.86(143.18-344.02)	278.17 (176.29-365.48)	204.33 (118.75-308.42)	<0.001
Mean Magnesium	2.04 (1.91-2.18)	2.0 (1.89-2.12)	2.08 (1.95-2.27)	0.001
Mean Calcium, Total	8.16 (7.78-8.54)	8.14 (7.74-8.48)	8.16 (7.81-8.63)	0.191
Max Urine Appearance	0.0 (0.0-1.0)	0.0 (0.0-1.0)	0.0 (0.0-1.0)	0.001
Max Platelet Count	312.0 (191.5-443.0)	357.0 (217.0-478.0)	266.5 (167.0-404.25)	0.002
Max CO2	29.0 (26.0-32.0)	30.0 (27.5-33.0)	29.0 (24.0-31.0)	0.008
Kurt White Blood Cells	-0.86 (-1.34-0.35)	-0.93 (-1.36-0.49)	-0.76 (-1.22-0.12)	0.072
Kurt Urea Nitrogen	-1.01 (-1.41-0.5)	-1.06 (-1.45-0.77)	-0.92 (-1.37-0.35)	0.045
Kurt Creatinine	-0.95 (-1.37-0.36)	-0.84 (-1.36-0.22)	-1.04 (-1.45-0.44)	0.129
Slope pCO2	0.02 (-0.69-0.6)	0.02 (-0.6-0.6)	0.02 (-0.86-0.56)	0.708
Slope White Blood Cells	-0.04 (-0.57-0.6)	-0.1 (-0.72-0.31)	0.1 (-0.48-0.95)	0.011
Slope RDW	0.05 (-0.01-0.16)	0.04 (-0.01-0.11)	0.08 (-0.02-0.18)	0.066
Slope Magnesium	0.01 (-0.01-0.05)	0.01 (-0.01-0.04)	0.02 (-0.01-0.07)	0.05
Skew Urea Nitrogen	0.18 (-0.17-0.67)	0.04 (-0.22-0.55)	0.35 (-0.08-0.75)	0.013
Skew Phosphate	0.11 (-0.29-0.57)	0.05 (-0.36-0.52)	0.18 (-0.28-0.64)	0.158
Skew INR(PT)	0.22 (-0.17-0.71)	0.22 (-0.31-0.67)	0.22 (-0.0-0.71)	0.288
Min pCO2	35.0 (31.5-39.0)	35.0 (32.0-39.0)	35.5 (31.0-40.0)	0.372
Min Urea Nitrogen	18.0 (10.0-28.0)	14.0 (9.0-20.0)	21.0 (12.75-34.0)	<0.001
Min Platelet Count	161.0 (91.5-248.0)	182.0 (116.0-262.0)	135.5 (75.0-222.5)	0.004
Min Creatinine	0.8 (0.5-1.2)	0.7 (0.5-0.9)	0.9 (0.6-1.42)	<0.001
Min Anion Gap	11.0 (9.0-12.0)	10.0 (9.0-11.0)	11.0 (9.0-13.0)	0.001
Mean pH	7.2 (6.83-7.37)	7.21 (6.93-7.41)	7.17 (6.65-7.36)	0.017
Mean Urea Nitrogen	24.78 (15.39-38.0)	19.62 (14.5-29.29)	31.84 (21.94-48.0)	<0.001
Max PTT	38.5 (31.55-58.85)	35.8 (29.95-46.05)	44.95 (33.2-74.95)	0.001
Max Magnesium	2.3 (2.2-2.5)	2.3 (2.1-2.4)	2.4 (2.2-2.6)	0.001
Kurt RDW	-0.98 (-1.37-0.53)	-0.99 (-1.38-0.6)	-0.97 (-1.32-0.5)	0.938
Kurt Chloride	-0.93 (-1.34-0.39)	-0.93 (-1.32-0.37)	-0.93 (-1.34-0.47)	0.641
Kurt Anion Gap	-0.85 (-1.23-0.25)	-0.89 (-1.24-0.26)	-0.79 (-1.22-0.25)	0.74
Slope Red Blood Cells	-0.03 (-0.07-0.03)	-0.03 (-0.07-0.02)	-0.03 (-0.07-0.03)	0.831

This open-access article distributed under the terms of the Creative Commons Attribution NonCommercial 3.0 License (CC BY-NC 3.0).
Downloaded from: https://journals.sbmu.ac.ir/aaem/index.php/AAEM/index

Table 2: Baseline characteristics of patients by acute kidney injury status

Features	Total (n = 314)	Acute kidney injury		P
		No (n =155)	Yes (n = 159)	
Slope Hemoglobin	-0.07 (-0.22-0.09)	-0.07 (-0.24-0.09)	-0.07 (-0.18-0.09)	0.517
Slope Chloride	-0.2 (-0.83-0.44)	-0.29 (-0.83-0.36)	-0.18 (-0.82-0.6)	0.384
Skew Sodium	0.0 (-0.38-0.45)	-0.03 (-0.37-0.43)	0.0 (-0.41-0.47)	0.526

Data are presented as frequency (%) or median (IQR). IQR: interquartile range.

Table 3: Performance metrics for the evaluated models, with their respective 95% confidence intervals

Model	Accuracy	Sensitivity	Specificity	Precision	F1 Score	AUC
Extra Trees	0.90(0.87–0.93)	0.89(0.86–0.92)	0.88(0.84–0.92)	0.91(0.88–0.94)	0.90(0.87–0.93)	0.91(0.88–0.94)
LightGBM	0.88 (0.86–0.91)	0.88 (0.86–0.91)	0.87 (0.84–0.90)	0.88 (0.85–0.91)	0.88(0.85–0.91)	0.89(0.87–0.91)
XGBoost	0.86 (0.83–0.89)	0.77 (0.74–0.80)	0.84 (0.81–0.87)	0.96 (0.94–0.98)	0.85(0.82–0.88)	0.92(0.90–0.94)
CatBoost	0.85 (0.83–0.87)	0.87 (0.84–0.90)	0.86 (0.84–0.88)	0.85 (0.82–0.88)	0.86 (0.83–0.89)	0.88(0.86–0.90)
RF	0.86 (0.83–0.89)	0.87 (0.84–0.90)	0.86 (0.84–0.88)	0.87 (0.84–0.90)	0.87 (0.84–0.90)	0.87(0.84–0.90)
NN	0.76 (0.73–0.79)	0.75 (0.72–0.78)	0.78 (0.75–0.81)	0.79 (0.76–0.82)	0.77 (0.74–0.80)	0.82(0.79–0.85)

Data are presented with 95% confidence interval. RF: Random Forest; NN: Neural Network.

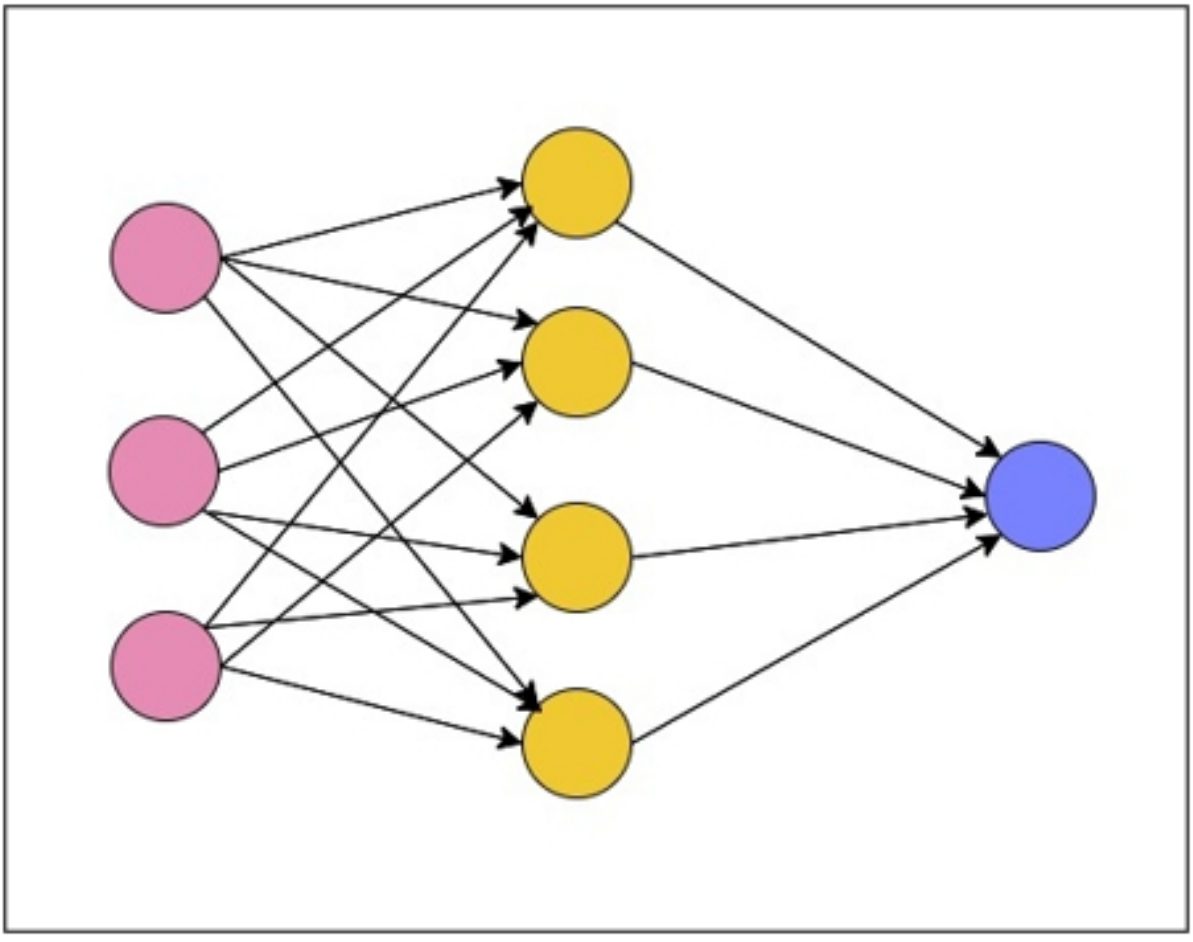

Figure 2: Neural network with input, hidden, and output layers for AKI prediction. Captures complex, non-linear relationships in ICU patient data for enhanced prediction accuracy.

This open-access article distributed under the terms of the Creative Commons Attribution NonCommercial 3.0 License (CC BY-NC 3.0).
Downloaded from: https://journals.sbmu.ac.ir/aaem/index.php/AAEM/index

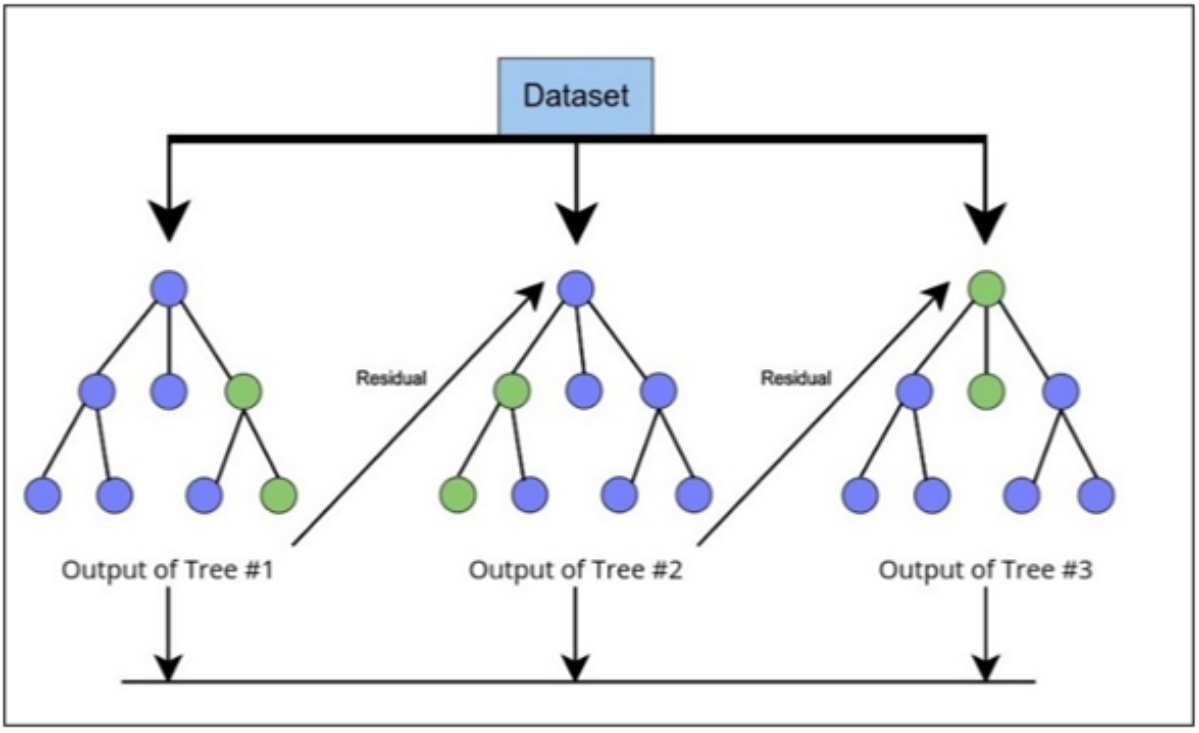

Figure 3: XGBoost model workflow showing sequential tree building to correct prediction errors. Features like creatinine levels and anion gap are key contributors.

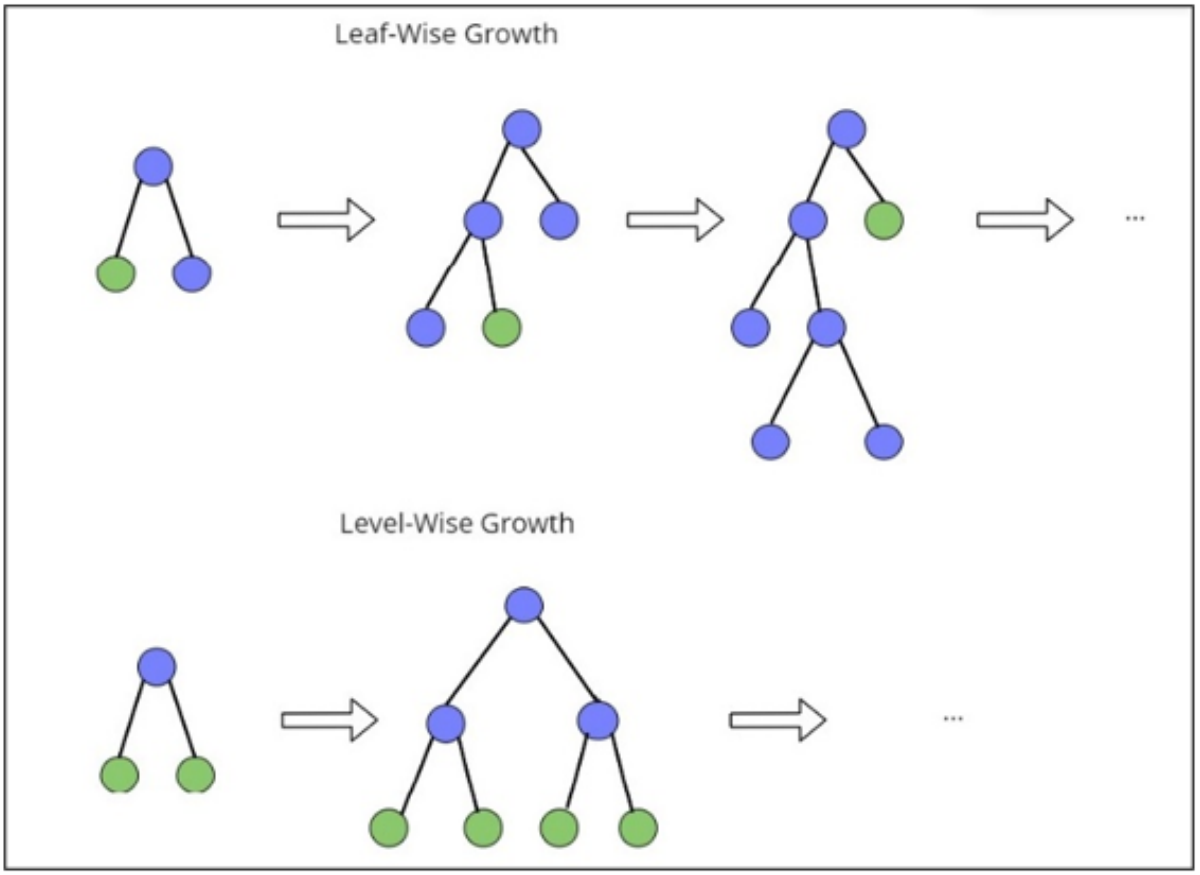

Figure 4: Comparison of LightGBM's leaf-wise growth versus level-wise growth. The figure illustrates how the leaf-wise approach enhances efficiency and prediction accuracy.

This open-access article distributed under the terms of the Creative Commons Attribution NonCommercial 3.0 License (CC BY-NC 3.0).
Downloaded from: https://journals.sbmu.ac.ir/aaem/index.php/AAEM/index

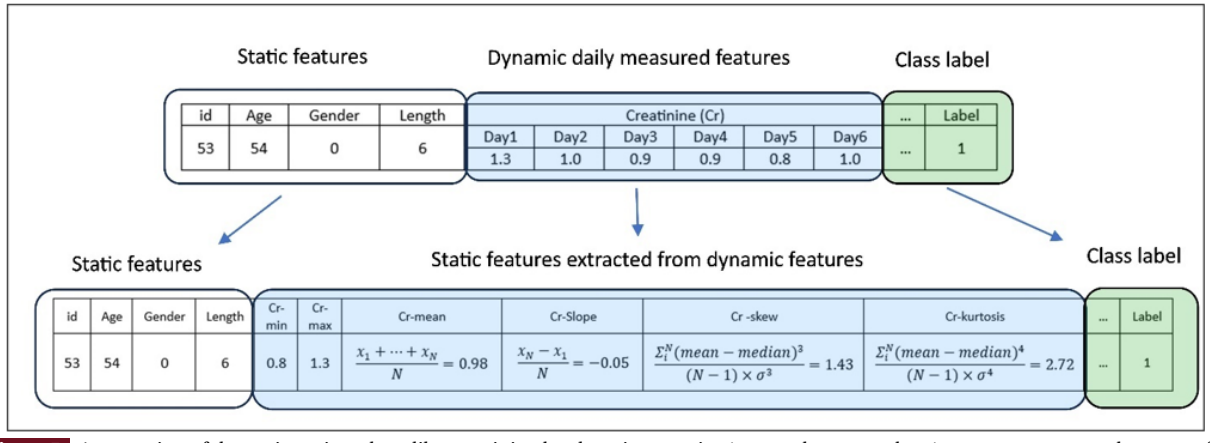

Figure 5: Aggregation of dynamic patient data, like creatinine levels, using metrics (mean, skewness, slope) to capture temporal patterns for AKI prediction.

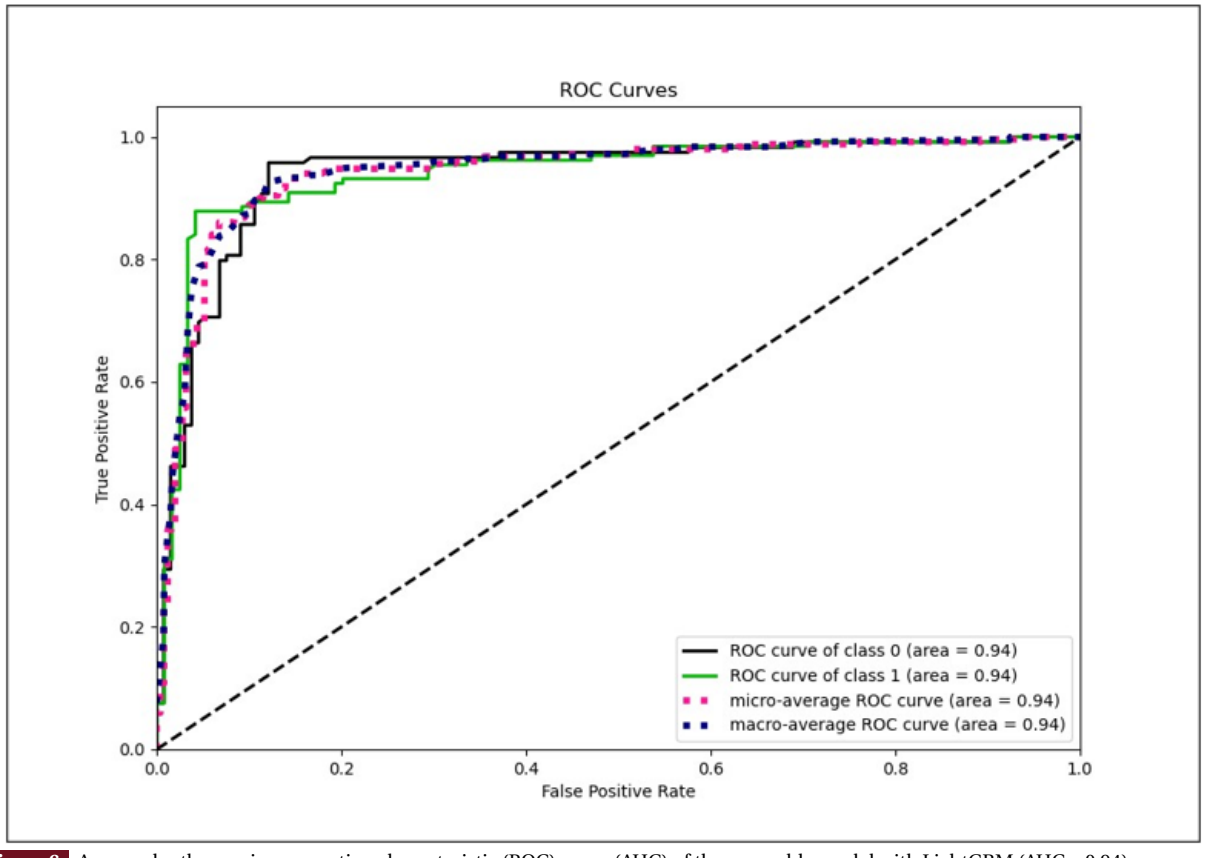

Figure 6: Area under the receiver operating characteristic (ROC) curve (AUC) of the ensemble model with LightGBM (AUC = 0.94).

This open-access article distributed under the terms of the Creative Commons Attribution NonCommercial 3.0 License (CC BY-NC 3.0).
Downloaded from: https://journals.sbmu.ac.ir/aaem/index.php/AAEM/index

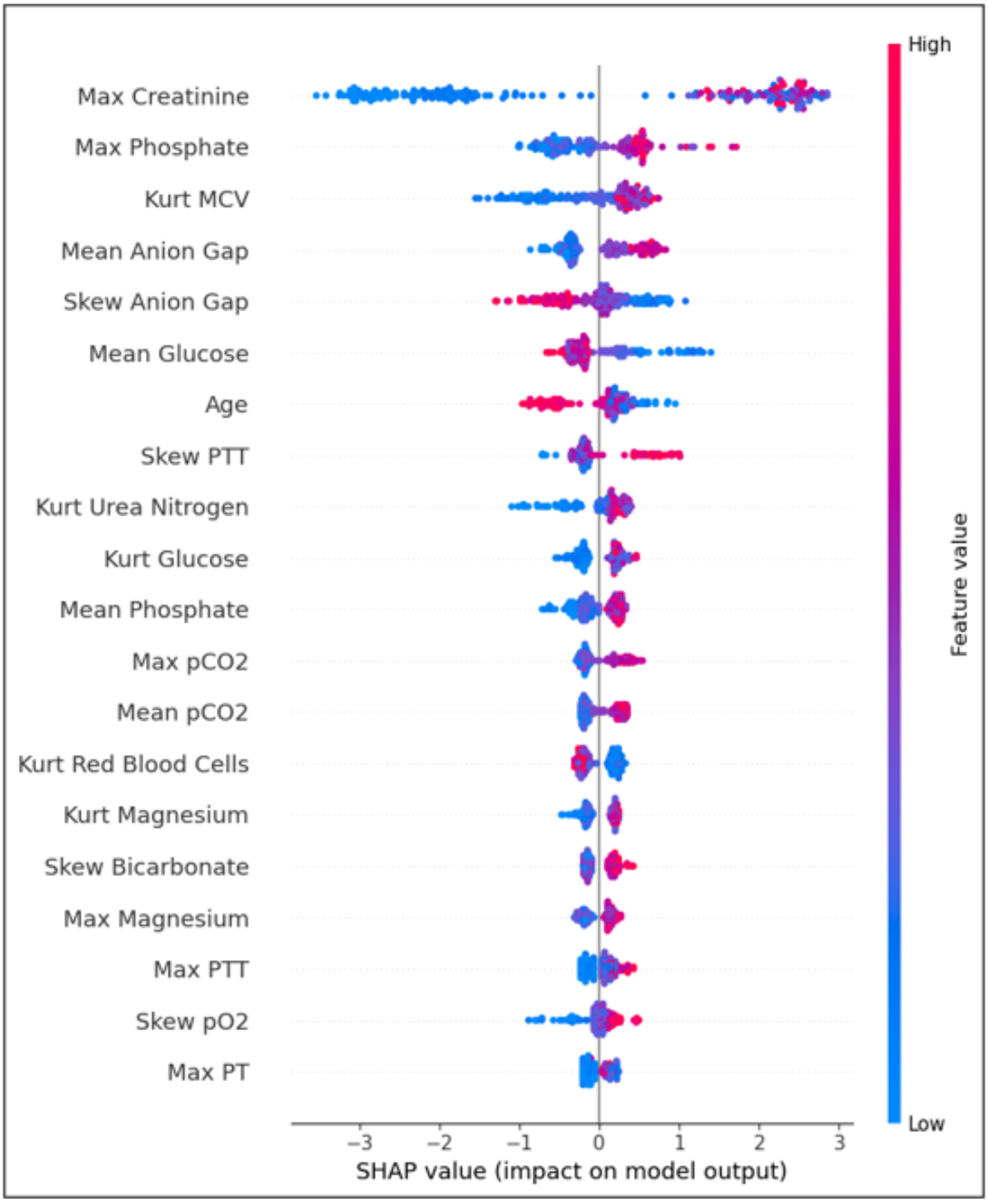

Figure 7: SHAP summary plot highlighting key features like 'Max Creatinine' and 'Age' that contribute most significantly to predicting acute kidney injury (AKI) risk.

This open-access article distributed under the terms of the Creative Commons Attribution NonCommercial 3.0 License (CC BY-NC 3.0).
Downloaded from: https://journals.sbmu.ac.ir/aaem/index.php/AAEM/index

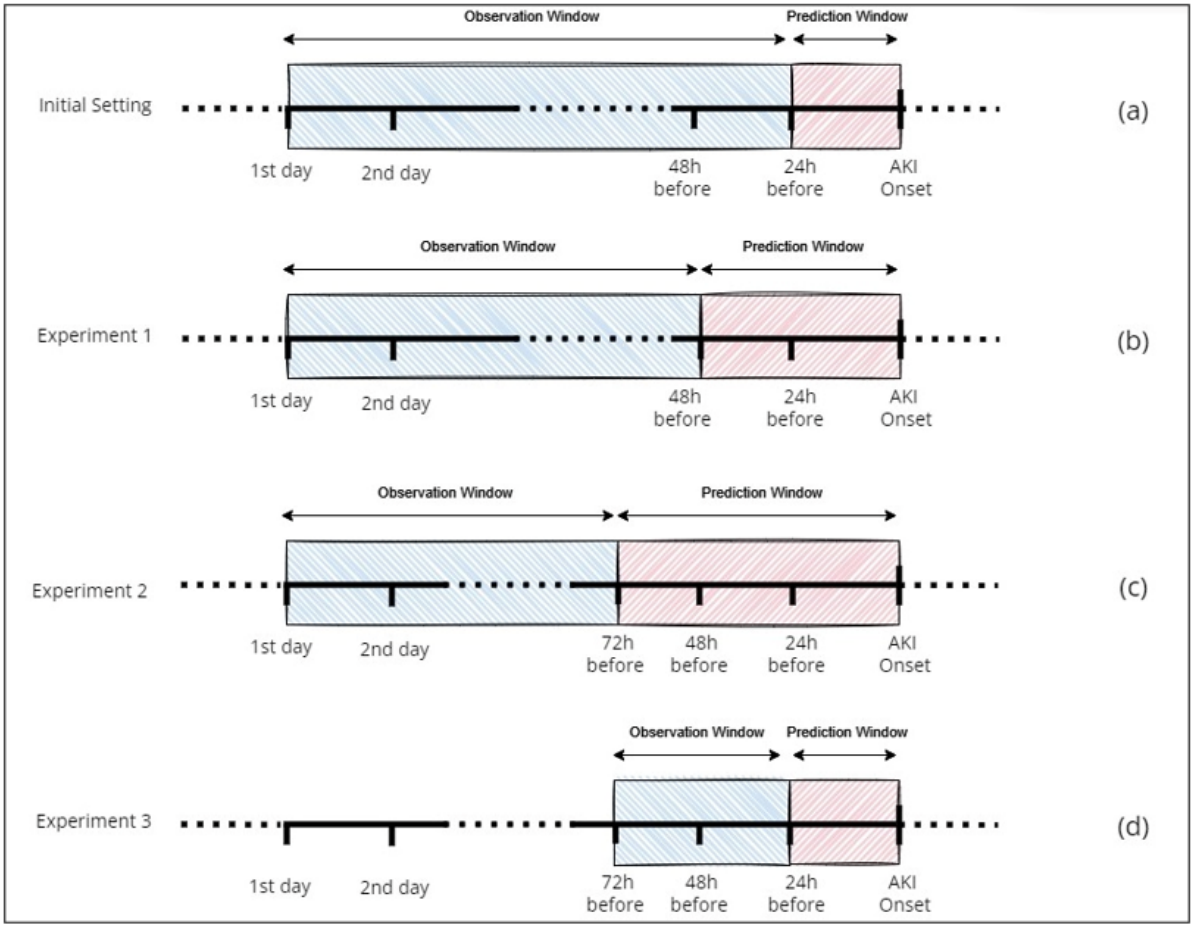

Figure 8: Schematic of experimental settings analyzing the impact of different temporal windows (24, 48, 72 hours) on acute kidney injury (AKI) prediction model performance.

This open-access article distributed under the terms of the Creative Commons Attribution NonCommercial 3.0 License (CC BY-NC 3.0).
Downloaded from: https://journals.sbmu.ac.ir/aaem/index.php/AAEM/index