UNIVERSITAT ROVIRA I VIRGILI
ESCOLA TÈCNICA SUPERIOR D'ENGINYERIA

Predicting Acute Kidney Injury in Septic Patients Using Logistic Regression with MIMIC-III Data

Master in Health Data Science — MHEDAS

ASSESSMENT III

Subject: Electronic Health Records

Authors: Antonio García Tierno,
Daniel Girbes Sardaña,
Ravneet-Rahul Sandhu Singh

Date: January 23, 2025

UNIVERSITAT ROVIRA i VIRGILI
BIOINFORMATICS BARCELONA
UAB Universitat Autònoma de Barcelona
UNIVERSITAT DE BARCELONA
Universitat de Lleida
UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH
UNIVERSITAT DE VIC UNIVERSITAT CENTRAL DE CATALUNYA
Universitat de Girona
UGA Université Grenoble Alpes
eit Health
Co-funded by the European Union

ABSTRACT

Acute Kidney Injury (AKI) is a severe complication in critically ill patients, particularly those with sepsis. Using the MIMIC-III database, we developed and evaluated a logistic regression model to predict AKI. ICU patients with sepsis were selected based on criteria such as diagnosis codes, age, and hospital stay duration. Data preprocessing addressed challenges such as missing values and class imbalance. Key predictors, including estimated Glomerular Filtration Rate (eGFR), creatinine, and Blood Urea Nitrogen (BUN), were identified as the most significant contributors. The model achieved an Area Under the Curve (AUC) of 0.810 (95% CI: 0.788–0.832), demonstrating strong performance with the following metrics: accuracy (74.060%), precision (76.970%), recall (67.340%), and F1-score (71.830%).

Non-Scientific Abstract

Acute Kidney Injury (AKI) is a serious condition where the kidneys suddenly stop working properly, which can happen to people who are already very sick, especially with infections like sepsis. To help doctors predict when this might happen, we used a large database of hospital records to build a logistic model. We looked at patients in intensive care and filtered them by factors like kidney test results and age to train the model. Some parameters, like estimated Glomerular Filtration Rate (eGFR), creatinine, and Blood Urea Nitrogen (BUN), were especially helpful in predicting AKI. The model isn't perfect but works well enough to be useful, correctly identifying about 74% of cases.

Table of Contents

1. Introduction

The kidneys play a vital role in maintaining the body's physiological balance, making their proper function critical for health. Recent advancements in Machine Learning (ML) have shown promising results in predicting medical conditions, relying heavily on key clinical features with significant influence on model performance. Many of these features are derived from Electronic Health Records (EHRs), which have become indispensable for driving advancements in healthcare research and outcomes.

1.1 Acute Kidney Injury

The kidneys maintain the body's balance by filtering blood to remove waste and excess fluids as urine, regulating electrolytes, acid-base levels, and blood pressure. They also produce erythropoietin for red blood cell production and activate vitamin D for bone health.

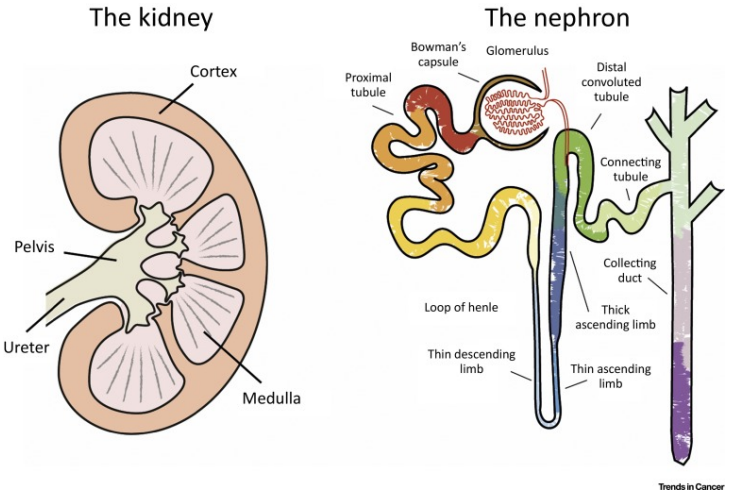

Figure 1.1: Kidney and nephron anatomy. Source: [1].

Each kidney consists of two main structural regions: an outer cortex and an inner medulla, housing functional units called nephrons, as shown in **Figure 1.1**.

The kidneys can lose their ability to function, a condition called Acute Kidney Injury (AKI), marked by a sudden decline in kidney performance, waste accumulation, and in severe cases, life-threatening complications. AKI affects 14% of hospitalized patients globally, with higher rates in intensive care [2]. Its primary causes include reduced kidney blood flow (prerenal), tissue damage (intrinsic), and urine flow obstruction (postrenal).

1.2 State of the Art

In recent years, various ML techniques have been employed to predict AKI, showcasing diverse methodologies and outcomes. Support Vector Machines (SVMs) have been widely explored for AKI prediction due to their ability to handle high-dimensional data. A systematic review by Vagliano et al. [3] assessed multiple ML models, including SVMs, for

predicting AKI. The review reported that SVM performance metrics varied significantly across studies, with area under the receiver operating characteristic curve (AUC) values ranging from 0.552 to 0.890. The authors emphasized that despite its potential, SVM performance is heavily dependent on careful tuning and proper data preprocessing.

Another study by Wei et al. [4] focused on the application of the K-Nearest Neighbors (KNN) algorithm to predict 90-day renal recovery in patients with severe AKI undergoing prolonged intermittent renal replacement therapy. The KNN model demonstrated superior performance compared to other algorithms, showing promise in long-term prognosis (AUC = 0.621). The authors highlighted KNN's ability to capture patient similarities effectively, which is useful in identifying patterns in datasets.

In a separate approach, Yue et al. [5] applied the Extreme Gradient Boosting (XGBoost) algorithm to predict AKI in critically ill septic patients. The XGBoost model achieved an AUC of 0.817, outperforming other models such as artificial neural networks. The authors attributed the superior performance of XGBoost to its capability to handle complex, high-dimensional data, making it especially suitable for critical care settings.

The success of these ML models in predicting AKI relies heavily on the selection of relevant clinical and laboratory features. Among the most commonly utilized features are serum creatinine and estimated Glomerular Filtration Rate (eGFR), which provide critical insights into kidney filtration efficiency. Blood urea nitrogen (BUN), another key biomarker, reflects the levels of nitrogenous waste in the blood and is often used alongside creatinine to assess renal function. Urine output, a direct indicator of kidney activity, is frequently incorporated to monitor acute changes. Additionally, demographic factors such as age and pre-existing conditions, including diabetes and hypertension, are crucial for understanding a patient's risk profile.

1.3 MIMIC-III Clinical Database

The MIMIC-III Clinical Database is a publicly available dataset containing data from over 40,000 ICU patients at Beth Israel Deaconess Medical Center in Boston, spanning more than 58,000 hospital admissions between 2001 and 2012. Developed through collaboration with the Massachusetts Institute of Technology, it includes detailed information on physiological signals, medications, lab results, diagnostics, and medical notes. The data is fully de-identified in compliance with HIPAA, using unique patient codes to protect privacy [6].

2. Methodology

This section outlines the process used to prepare the data, identify key features, and build predictive models for AKI. The workflow began with cohort selection, followed by data preprocessing. Feature selection techniques were applied to determine the most relevant predictors, and a linear regression model was developed and evaluated to predict AKI.

2.1 Cohort Selection

ICU patient data from the MIMIC-III Clinical Database was extracted, applying the following restrictions:

- **Diagnosis Restriction:** Sepsis diagnosed patients.
- **Age Restriction:** Patients aged 18–89 years old.
- **Length of Stay:** Hospital stays of at least 48h.
- **Feature Data:** Data from the first 24h.

The selected variables are summarized in **Table 2.1**. Demographics provide baseline variability, comorbidities indicate underlying risks, laboratory features capture critical biochemical markers, chartevent features reflect real-time clinical conditions, and interventions highlight treatment intensity. For laboratory and chartevent features, maximum and minimum values were extracted and specific ranges filters were applied to remove outliers.

Table 2.1: Selected features for analysis (33 features in total).

Category	Features
Demographics	Age, Ethnicity, Gender
Comorbidities	Acute Kidney Injury (AKI), Chronic Kidney Disease (CKD), Coronary Artery Disease (CAD), Hypertension (HYP), Type 2 Diabetes Mellitus (DM2)
Laboratory Features	Albumin, Anion Gap, Bilirubin, Blood Urea Nitrogen (BUN), Chloride, Creatinine, estimated Glomerular Filtration Rate (eGFR), Glucose, Hematocrit, Hemoglobin, International Normalized Ratio (INR), Lactate, Partial Thromboplastin Time (PTT), Platelet Count, Potassium, Sodium
Chartevent Features	Diastolic Blood Pressure (DBP), Heart Rate (HR), Height, Oxygen Saturation (SpO2), Systolic Blood Pressure (SBP), Temperature, Weight
Interventions	Mechanical Ventilation, Vasopressor Use

2.2 Data Preprocessing

This process involved several key steps aimed at preparing the raw data for modeling. Column mapping was performed to standardize and organize data fields. The eGFR was calculated according to the steps described at MIMIC's GitHub. Missing data (NaN values) were addressed systematically to minimize biases, and class imbalances were handled using Synthetic Minority Oversampling Technique (SMOTE).

2.2.1 Column Mapping

Gender was encoded as a binary variable, where `Male` was mapped to 1 and `Female` to 0. For ethnicities, the most prevalent category was assigned a value of 1, the second most prevalent category was assigned a value of 2, and all remaining categories were mapped to 3. To ensure consistency in data handling, the data type for all columns was set to `float`, except for comorbidities, gender, and ethnicities which were retained as `int` due to their categorical nature.

2.2.2 eGFR Calculation

The eGFR is a crucial marker for kidney function, providing an estimate of the filtration efficiency of the kidneys. For this analysis, eGFR was calculated using the following equation [7]:

$$\begin{aligned} \text{eGFR} = 175 \times (\text{Scr})^{-1.154} \times (\text{Age})^{-0.203} \\ \times (0.742 \text{ if Female}) \\ \times (1.212 \text{ if African American}). \end{aligned} \tag{2.1}$$

In this formula, `Scr` represents the patient's serum creatinine level in mg/dL, and `Age` is measured in years. The gender correction factor is 0.74 if the patient is female and 1 otherwise, while the ethnicity correction factor is 1.21 for African American patients and 1 otherwise.

2.2.3 NaN Removal

First, features corresponding to *Mechanical Ventilation* and *Vasopressor Use* were eliminated because no entries in the dataset for the selected subjects contained values for the associated `ITEMID`. Next, vertical elimination (i.e. feature elimination) was performed by excluding columns with more than 15% missing data. This step ensured that features with excessive missing values, which could introduce noise or bias into the analysis, were removed. Following this, horizontal elimination (i.e. subject elimination) was applied to rows with more than 20% missing data, further reducing the dataset by excluding subjects with insufficient information.

To address missing values, we applied an iterative imputation method based on the Multivariate Imputation by Chained Equations (MICE) framework [8]. This method estimates missing values by iteratively predicting them using regression models built from the observed data. Each column with missing values is treated as a dependent variable in turn, while other columns act as predictors. The imputation process continues iteratively until convergence, ensuring that the relationships and patterns within the dataset are preserved. In our case, we used a maximum of 10 iterations for the imputation process.

2.2.4 Class imbalance

To address the class imbalance in the dataset, we applied SMOTE. This method generates synthetic samples for the minority class by interpolating between existing minority class samples [9]. As shown in Figure, this process uses the KNN algorithm to identify similar data points in the minority class and creates new samples along the line segments connecting these neighbors (in our case, 5 neighbors). The effects of SMOTE and the rest of the steps are summarized in **Table 2.2**.

Table 2.2: Summary of cohort adjustments. Counts represent the number of subjects with at least one NaN value across their columns, categorized by AKI and Non-AKI groups.

	AKI		Non-AKI	
Step	**NaN**	**Non-NaN**	**NaN**	**Non-NaN**
Interventions Features Out	2211	159	1230	58
Column and row filtering	223	2135	110	1173
Imputer	0	2358	0	1283
SMOTE	0	2358	0	2358

2.3 Feature Selection

Before applying feature selection techniques, the data was normalized to ensure all features were on the same scale. Normalization was performed using Min-Max scaling, which transforms each feature to a range of [0, 1]. The Min-Max is defined as:

$$X_{\text{scaled}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}}, \tag{2.2}$$

where X represents the original feature values, X_{min} is the minimum value, and X_{max} is the maximum value of the feature. By normalizing the data, we ensured that all features contributed equally to the feature selection process.

2.3.1 Variance Inflation Factor (VIF)

VIF is a statistical measure used to quantify multicollinearity among features in a dataset. Multicollinearity arises when two or more features are highly correlated, which can destabilize regression coefficients and reduce the interpretability of the model. The VIF for a given feature X_i is calculated as:

$$\text{VIF}_i = \frac{1}{1 - R_i^2}, \tag{2.3}$$

where R_i^2 is the coefficient of determination obtained by regressing the feature X_i against all other features. A high R_i^2 value indicates that X_i can be predicted well by the other features, leading to a high VIF value. Generally, a VIF threshold of 10 is used, with values above this indicating multicollinearity.

In this study, an iterative approach was employed to remove features with high VIF values. First, a constant term was added to the dataset to account for the intercept in regression calculations. The VIF was computed for each feature by regressing it against the other features and calculating the R_i^2 values. Features with VIF values exceeding 10 were flagged as highly collinear. Among these, the feature with the highest VIF value was removed from the dataset. The process was repeated, recalculating the VIF after each removal, until all remaining features had VIF values below the threshold [10].

2.3.2 Significant Features

After identifying non-collinear features using VIF, we performed statistical tests to determine the most significant predictors of AKI. Specifically, quantitative features were analyzed using Student's t-test (t), while qualitative features were evaluated using a Chi-squared test (χ^2). Features with a p-value ≤ 0.05 were retained as statistically significant. The most significant features identified through these tests are presented in **Table 2.3**.

The t-test is a parametric statistical test that compares the means of two groups—in our case, AKI and Non-AKI cohorts. This test assumes that the data follows a normal distribution and calculates whether the observed differences in means are likely due to random chance. For instance, if a quantitative feature such as creatinine levels significantly differs between the two groups, it suggests a meaningful association with AKI. On the other hand, the χ^2-test is a non-parametric test that examines the association between categorical variables. It compares the observed frequencies of categories in AKI and Non-AKI groups against the expected frequencies under the null hypothesis of no association. For example, if certain ethnicities are disproportionately represented in the AKI group, the test will highlight this relationship [11].

2.4 Model Development

To generate the logistic model, we split the dataset into training and testing sets. The dataset was divided using a 70/30 split, where 70% of the data was allocated for training and 30% for testing.

Logistic regression is a statistical method used for binary classification, which predicts the probability of an outcome belonging to one of two classes. The model uses the logistic function, defined as:

$$P(y = 1|x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_n x_n)}}, \tag{2.4}$$

Table 2.3: Comparison of features between AKI and Non-AKI patients. Quantitative features are summarized by mean and standard deviation (SD), while categorical features are presented as counts for each group. A feature is considered statistically significant if the p-value is ≤ 0.05.

Feature (Unit)	AKI Mean (SD) / Count	Non-AKI Mean (SD) / Count	p-value
Age (years)	65.457 (14.986)	62.623 (15.045)	1.002×10^{-10}
Minimum Creatinine (mg/dL)	1.930 (1.513)	1.387 (1.753)	1.441×10^{-29}
Maximum Anion gap (mmol/L)	16.749 (4.605)	14.624 (3.604)	1.476×10^{-67}
Minimum Anion gap (mmol/L)	14.186 (3.753)	12.650 (3.070)	4.015×10^{-52}
Maximum Chloride (mmol/L)	108.420 (7.428)	107.488 (6.011)	2.234×10^{-6}
Minimum Chloride (mmol/L)	105.122 (7.557)	104.669 (5.856)	2.144×10^{-2}
Maximum Glucose (mg/dL)	174.921 (92.680)	155.957 (67.823)	1.341×10^{-15}
Maximum Platelet Count ($10^3/\mu L$)	225.550 (137.671)	237.127 (142.619)	4.588×10^{-3}
Maximum Potassium (mmol/L)	4.504 (0.822)	4.243 (0.683)	4.955×10^{-32}
Minimum Potassium (mmol/L)	3.862 (0.636)	3.707 (0.527)	1.777×10^{-19}
Maximum Sodium (mmol/L)	140.210 (6.018)	139.444 (4.198)	4.048×10^{-7}
Minimum Sodium (mmol/L)	137.386 (5.812)	137.058 (4.498)	3.030×10^{-2}
Minimum Hematocrit (%)	29.201 (5.483)	29.885 (4.839)	5.719×10^{-6}
Maximum Hemoglobin (g/dL)	10.647 (1.869)	10.867 (1.668)	2.086×10^{-5}
Minimum eGFR (mL/min/1.730 m²)	54.405 (43.681)	97.813 (80.100)	5.736×10^{-112}
Minimum BUN (mg/dL)	39.019 (26.021)	20.502 (14.097)	2.487×10^{-185}
Minimum SBP (mmHg)	84.855 (14.873)	86.624 (13.538)	1.976×10^{-5}
Minimum DBP (mmHg)	41.174 (10.049)	42.601 (9.722)	7.452×10^{-7}
Maximum Temperature (°C)	37.611 (0.981)	37.858 (0.925)	7.728×10^{-19}
Minimum Temperature (°C)	36.077 (0.856)	36.221 (0.781)	1.383×10^{-9}
Gender	{0: 962, 1: 1396}	{0: 1172, 1: 1186}	9.689×10^{-10}
Ethnicity	{1: 1693, 2: 260, 3: 405}	{1: 1651, 2: 349, 3: 358}	2.707×10^{-4}
DM2	{0: 1382, 1: 976}	{0: 1597, 1: 761}	1.044×10^{-10}
CAD	{0: 1691, 1: 667}	{0: 1853, 1: 505}	5.794×10^{-8}
CKD	{0: 1593, 1: 765}	{0: 1972, 1: 386}	1.355×10^{-37}
HYP	{0: 1209, 1: 1149}	{0: 1451, 1: 907}	1.473×10^{-12}

where $P(y = 1|x)$ is the probability of the positive class, β_0 is the intercept, $\beta_1, \beta_2, \ldots, \beta_n$ are the coefficients, and $x_1, x_2, \ldots, x_n$ are the features. The logistic regression model was trained using an L_2 penalty to control regularization and prevent overfitting. A regularization strength (C) of 100 was chosen, which reduces the regularization effect and allows the model to fit the data more closely. Additionally, a maximum of 200 iterations was set to ensure the optimization algorithm had sufficient time to converge to an optimal solution [12].

Once trained, the model was evaluated across the testing dataset. Predictions were generated for each set, and the model's performance was assessed using key metrics, including accuracy, precision, recall, F1-score, and the AUC. The AUC quantifies the model's ability to discriminate between the two classes (AKI and Non-AKI).

In addition, a bootstrapping procedure was applied to compute a 95% confidence interval for the AUC [13]. This process involved resampling the data with replacement to generate multiple subsets . For each subset, the Receiver Operating Characteristic (ROC) curve and AUC were computed. The bootstrapped AUC scores were used to calculate the confidence interval, providing a range within which the true AUC value is likely to fall.

3. Results

This section presents the findings of the study, focusing on the performance metrics of the logistic regression model, the contribution of individual features, and a comparison with existing studies.

3.1 Model Performance Metrics

The performance metrics in **Table 3.1** demonstrate consistent results between the training and testing datasets. The test set shows slightly higher precision, suggesting the model effectively reduces false positives, while recall is higher in the training set, reflecting a trade-off between sensitivity and precision. The F1-score aligns with these trends.

Table 3.1: Performance metrics for the training and testing datasets.

Metric	Train	Test
Accuracy	0.737	0.741
Precision	0.751	0.770
Recall	0.716	0.673
F1-score	0.733	0.718

Regarding the AUC, both the training and testing datasets achieve a strong balance between sensitivity (True Positive Rate) and specificity (False Positive Rate). As shown in **Figure 3.1**, for the training set, the AUC is 0.815 with a 95% confidence interval of (0.800, 0.830). Similarly, the test set achieves an AUC of 0.810 with a 95% confidence interval of (0.788, 0.832). The overlap between the confidence intervals of the training and testing sets suggests that the model generalizes well, showing no evidence of significant overfitting or underfitting.

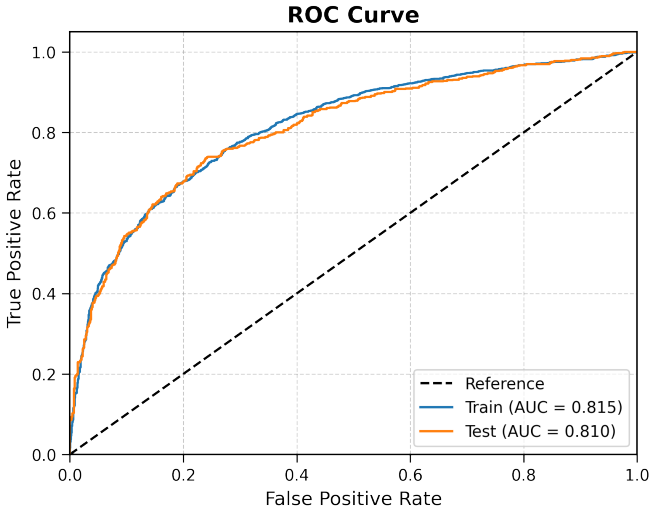

Figure 3.1: ROC curves for the training and testing datasets.

3.2 Feature Contribution Analysis

The bar plot in **Figure 3.2** highlights the importance of various features in the logistic regression model, with eGFR emerging as the most influential predictor. eGFR is a key marker of kidney function, estimating the rate at which the kidneys filter waste from the blood. A higher eGFR value, reflected by the positive coefficient, is strongly associated with improved outcomes. Clinically, low eGFR values (<60 mL/min/1.730 m^2) indicate CKD or renal dysfunction, which are linked to adverse outcomes, including cardiovascular disease and increased mortality [14, 15].

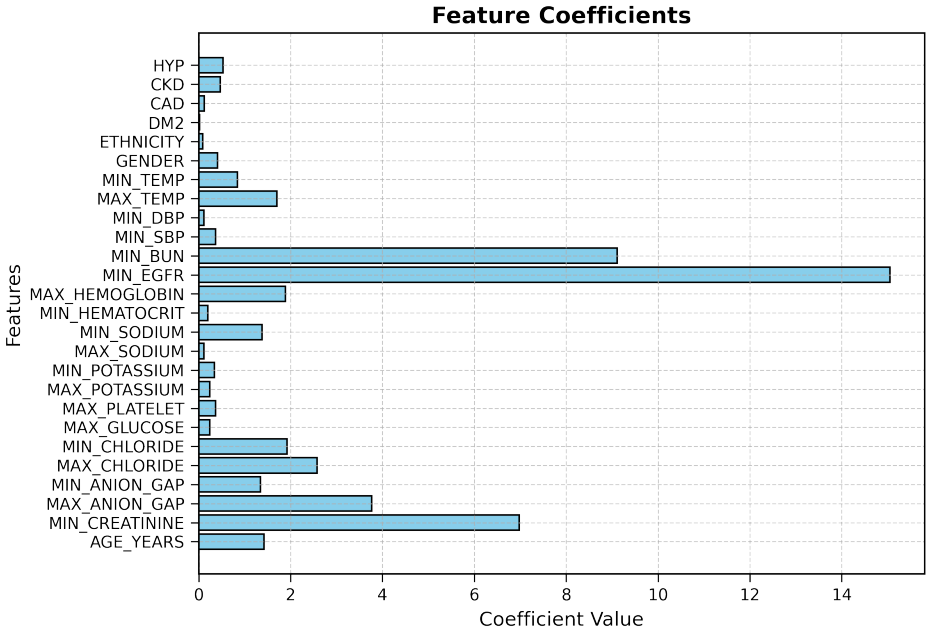

Figure 3.2: Feature importance in the logistic regression model.

Creatinine and BUN are also highly influential features, underscoring the model's reliance on markers of kidney health. Creatinine, a byproduct of muscle metabolism, is a core component of eGFR calculations, with elevated levels indicating reduced kidney function [16]. Similarly, BUN, which measures the amount of nitrogen derived from protein metabolism, is another strong indicator of renal health. Elevated BUN and creatinine levels often point to kidney dysfunction, dehydration, or acute injury [17]. Additionally, chloride, an essential electrolyte involved in acid-base balance and cellular homeostasis, demonstrates moderate importance. Abnormal chloride levels are associated with metabolic disturbances, often linked to kidney or respiratory dysfunction [18, 19]. Other features, such as demographic factors (age, gender, and ethnicity) and vital signs (minimum systolic and diastolic blood pressure), contribute less significantly but remain important for stratifying classes. For example, age is a key factor in disease progression, particularly for chronic conditions like CKD and cardiovascular disease [20]. Electrolytes, including sodium and potassium, along with hemoglobin, hematocrit, and glucose, provide additional insight into metabolic and systemic health [21, 22].

3.3 Benchmarking Against Existing Studies

To evaluate the performance of our model, we compared it to other studies that also used logistic regression under similar conditions. Roknaldin et al. [23] and Jiang et al. [24] both utilized the MIMIC-III database, imposing similar restrictions regarding age and the inclusion of ICU-admitted patients. Roknaldin et al. specifically targeted septic patients, employing 23 features such as urine output, bilirubin levels, and eGFR. In addition to these features, they included variables such as mechanical ventilation and vasopressor use, which are critical for distinguishing AKI and non-AKI cases, a key difference from our approach. Malhotra et al. [25] used a broader, multicenter dataset, incorporating data from the University of California, San Diego (UCSD), and the Mayo Clinic to predict AKI in ICU patients. Their logistic regression model included 25 binary and 4 continuous predictors, spanning chronic comorbidities like chronic kidney disease and hypertension, as well as acute risk factors such as hypotension and mechanical ventilation.

The comparison of AUC values and their corresponding 95% confidence intervals (CI) for the models is summarized in **Figure 3.3**. Roknaldin et al.'s model achieved the highest performance with an AUC of 0.887 (95% CI: 0.861–0.915). Malhotra et al.'s model demonstrated solid performance with an AUC of 0.810 (95% CI: 0.780–0.830), reflecting its effectiveness across a more heterogeneous ICU population. Jiang et al.'s model reported an AUC of 0.760 (95% CI: 0.700–0.820), indicating a modest performance. Our proposed method achieved a strong AUC of 0.810 (95% CI: 0.788–0.832), comparable to Malhotra et al.'s results and competitive with those reported in the literature.

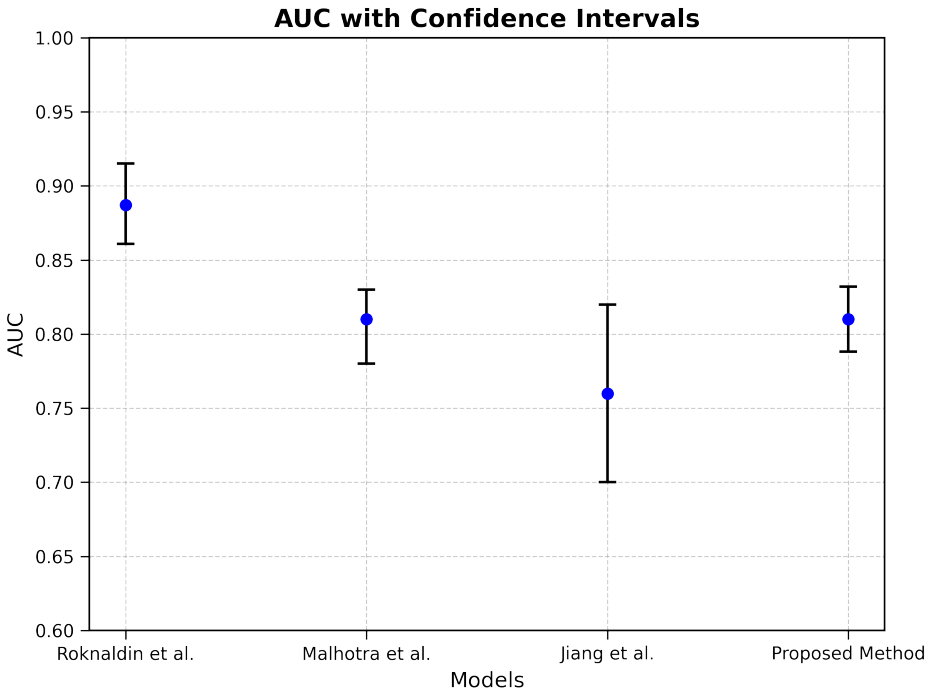

Figure 3.3: Comparison of AUC values across studies for AKI prediction.

4. Conclusions

This study developed and evaluated a predictive model for Acute Kidney Injury (AKI) using ICU patient data from the MIMIC-III database. The logistic regression model demonstrated strong performance, achieving an Area Under the Curve (AUC) of 0.810 (95% CI: 0.788–0.832). On the test set, the model achieved an accuracy of 74.060%, precision of 76.970%, recall of 67.340%, and an F1-score of 71.830%. Key predictors, such as estimated Glomerular Filtration Rate (eGFR), creatinine, and Blood Urea Nitrogen (BUN), were identified as the most significant contributors to AKI prediction, while secondary features like age and blood pressure provided additional insights.

Despite its good performance, there are areas for improvement. Including additional features such as ICU interventions could enhance predictive accuracy and provide a more comprehensive understanding of AKI risk factors. The model's reliance on data from a single source limits its generalizability, and validating the model on external datasets would help assess its applicability across different clinical settings. Furthermore, the simplicity of logistic regression may not fully capture complex, like non-linear relationships which might be present in the data, suggesting the potential benefit of exploring advanced machine learning techniques. Finally, the imputation of missing values introduces some degree of uncertainty, which could be mitigated by incorporating methods to handle missing data more robustly or building richer datasets with fewer missing values.

References

[1] H. C. Looker, M. Mauer, and R. G. Nelson, "Role of kidney biopsies for biomarker discovery in diabetic kidney disease," *Advances in Chronic Kidney Disease*, vol. 25, no. 2, pp. 192–201, 2018. Diabetic Kidney Disease (c. 2018).

[2] E. Adiyeke, Y. Ren, S. Fogel, P. Rashidi, M. Segal, E. A. Shenkman, A. Bihorac, and T. Ozrazgat-Baslanti, "Epidemiology, trajectories and outcomes of acute kidney injury among hospitalized patients: A retrospective multicenter large cohort study," *arXiv preprint arXiv:2403.08020*, Mar 2024.

[3] G. Vagliano *et al.*, "A systematic review of machine learning models for predicting acute kidney injury," *Clinical Kidney Journal*, vol. 15, no. 12, pp. 2266–2277, 2022.

[4] J. Wei *et al.*, "K-nearest neighbors algorithm for predicting 90-day renal recovery in severe aki patients," *BMC Medical Informatics and Decision Making*, 2023.

[5] F. Yue *et al.*, "Application of xgboost algorithm for predicting aki in critically ill septic patients," *BMC Translational Medicine*, 2022.

[6] A. Johnson, T. Pollard, and R. Mark, "Mimic-iii clinical database (version 1.4)," *PhysioNet*, 2016.

[7] M. L. for Computational Physiology, "Mimic code issue #608." GitHub. Accessed: 2025-01-23.

[8] S. Van Buuren and K. Groothuis-Oudshoorn, "mice: Multivariate imputation by chained equations in r," *Journal of Statistical Software*, vol. 45, no. 3, pp. 1–67, 2011.

[9] G. Kovács, "Smote-variants: A python implementation of 85 minority oversampling techniques," *Neurocomputing*, vol. 366, pp. 352–354, 2019.

[10] U. Community, "When using variance inflation factor (vif), should i do the removal recursively?." Cross Validated (Stack Exchange). Accessed: 2025-01-23.

[11] B. In, "T-test vs. chi-square: What's the difference?." Built In. Accessed: 2025-01-23.

[12] S. learn Developers, "Logisticregression — scikit-learn 1.5 documentation." Scikit-learn Documentation. Accessed: 2025-01-23.

[13] U. Community, "Confidence interval auc with the bootstrap method." Stack Overflow. Accessed: 2025-01-23.

[14] A. S. Levey and J. Coresh, "Chronic kidney disease," *The Lancet*, vol. 379, no. 9811, pp. 165–180, 2012.

[15] A. S. Go, G. M. Chertow, D. Fan, C. E. McCulloch, and C.-Y. Hsu, "Chronic kidney disease and the risks of death, cardiovascular events, and hospitalization," *New England Journal of Medicine*, vol. 351, no. 13, pp. 1296–1305, 2004.

[16] P. Delanaye, E. Cavalier, and H. Pottel, "Creatinine: not so simple!," *Nephron Clinical Practice*, vol. 120, no. 3, pp. c194–c197, 2012.

[17] H. Carroll, "Blood urea nitrogen," *Journal of Clinical Pathology*, vol. 50, no. 5, pp. 473–473, 1997.

[18] J. W. Ter Avest, J. P. van Dijk, L. D. Dikkeschei, and T. S. van der Werf, "Chloride: the queen of electrolytes?," *Clinical Chemistry*, vol. 46, no. 9, pp. 1435–1438, 2000.

[19] K. J. Gunnerson and J. A. Kellum, "Clinical review: the meaning of acid–base abnormalities in the intensive care unit–effects on patient outcome," *Critical Care*, vol. 9, no. 5, p. 508, 2005.

[20] M. Tonelli, N. Wiebe, B. Culleton, A. House, C. Rabbat, M. Fok, F. McAlister, and A. X. Garg, "Chronic kidney disease and mortality risk: a systematic review," *Journal of the American Society of Nephrology*, vol. 17, no. 7, pp. 2034–2047, 2006.

[21] N. E. Madias, "Hyponatremia," *Kidney International*, vol. 52, pp. S1–S8, 1997.

[22] J. H. Galla, "Disorders of potassium homeostasis: hypokalemia and hyperkalemia," *Critical Care Clinics*, vol. 16, no. 2, pp. 311–328, 2000.

[23] A. Roknaldin, Z. Zhang, J. Xu, K. Alaei, and M. Pishgar, "Utilizing machine learning models to predict acute kidney injury in septic patients from mimic-iii database," *arXiv preprint arXiv:2412.03737*, 2023.

[24] Z. Jiang, X. An, Y. Li, C. Xu, H. Meng, and Y. Qu, "Construction and validation of a risk assessment model for acute kidney injury in patients with acute pancreatitis in the intensive care unit," *BMC Nephrology*, vol. 24, p. 123, 2023.

[25] R. Malhotra, K. B. Kashani, E. Macedo, J. Kim, J. Bouchard, S. Wynn, G. Li, L. Ohno-Machado, and R. Mehta, "A risk prediction score for acute kidney injury in the intensive care unit," *Nephrology Dialysis Transplantation*, vol. 32, no. 5, pp. 814–822, 2017.