**Predicting Mortality in Critically Ill Patients with Sepsis-Associated Acute Kidney Injury**

P. Hardikar, U. Latnekar, A. Patel, D. Liau, T. Su

**Abstract**

**Background:** Sepsis-Associated Acute Kidney Injury (SA-AKI) is a condition where the kidney function deteriorates as a result of sepsis, which is a severe and potentially life-threatening reaction to infection. Since the kidneys play a crucial role in filtering waste products and excess fluids from the blood, when compromised, it can lead to acute kidney injury.

**Methods and Results:** We presented a deep learning approach for the prediction of mortality in critically ill patients with sepsis-associated acute kidney injury. To determine which patients had AKI, the conditions used were an increase in serum creatinine by ≥ 0.3 mg/dL within 48 hours and an increase in serum creatinine to ≥ 1.5 times baseline, which is known or presumed to have occurred within the prior seven days. Several machine meaning (ML) models were used as the baseline models to compare results.

**Results:**  For the selections mentioned in this research paper, we used the Medical Information Mart for Intensive Care IV (MIMIC-IV) database of 10,761 ICU patients with SA-AKI, our proposed model yielded an area under the receiver operating characteristics (AUROC) of 0.821, which is an improvement from the results of the baseline model and existing literature.

**Conclusions:** The proposed approach, with our updated feature selection, allowed for more accurate predictions to be made about the mortality among patients based on a patient's medical history from previous hospital visits.

**Keywords:** Machine learning, Acute kidney injury, Hospital readmission, Sepsis-associated

**Background**

Sepsis is a complex medical illness brought on by an infection that sets off a systemic inflammatory response [1]. In severely ill individuals, it is one of the most frequent and deadly causes of disease and death. Acute kidney damage (AKI) is commonly known as a consequence of sepsis. About 40% of patients with severe sepsis develop AKI, which makes treatment more challenging, expensive, and likely to result in death [1]. The complex syndrome, sepsis-associated acute kidney injury (SA-AKI), has several factors that are linked to a worse prognosis, an extended hospital stay, and a higher number of comorbidities than in sepsis patients without AKI [2]. Due to critical conditions, it is important to precisely forecast the prognosis for patients with SA-AKI in the ICU.

The prognosis of SA-AKI patients is an ongoing topic in critical care medicine. There are numerous scoring systems used to forecast patients with SA-AKI outcomes. However, they are limited to specificity and sensitivity, which have resulted in unsatisfactory performance. The Acute Physiology and Chronic Health Evaluation II, the Simplified Acute Physiology Score II (SAPS II), and the Sequential Organ Failure Assessment (SOFA) score are examples of the scoring systems used in practice. Furthermore, a few multivariate prediction models have been created to forecast the cycle of patients with SA-AKI. These models are based on widely used statistical methods such as the Cox proportional risk model and logistic regression, which constructed a death prediction model for 2066 patients with SA-AKI, demonstrating a better forecast performance [3].

The Cox proportional risk model was created to handle linear correlations between dependent and independent variables, potentially oversimplifying more complex nonlinear interactions. Nevertheless, the links between variables are intricate, including both linear and non-linear relationships. Multicollinearity between variables can also influence the Cox proportional risk model,and impair its effectiveness. Therefore, it is imperative to look for more accurate and successful prediction methods for the treatment of SA-AKI patients [3].

Medical practitioners are interested in and accepting of machine learning (ML), in addition to the advancements in statistical theory and computer technology. Modern machine learning techniques have benefited predictive models for a wide range of diseases, allowing them to surpass their more traditional logistic and Cox regression-based counterparts. Machine learning has been applied in a variety of clinical domains, with applications ranging from diagnosis to prediction. ML techniques have also been applied to predict critically ill patients' prognosis, yielding better results than the more traditional approaches of logistic regression and Cox regression analysis. However, the benefit of ML algorithms in predicting death within patients with SA-AKI has not yet been proven. This study aims to develop and validate machine learning models for early prediction of mortality for critically-ill patients with SA-AKI.

The present study aims to introduce a new process mining/deep learning approach that incorporates a patient's past medical history from previous hospital visits and the time information related to the variables in order to predict mortality in critically ill patients with sepsis-associated acute kidney injury. The Transparent Reporting of a Multivariable Predictive Model for Individual Prognosis Or Diagnosis (TRIPOD) project principles were followed in the development of a prediction model.

**Methods**

**Database Introduction**

We used the Medical Information Mart for Intensive Care IV (MIMIC-IV) database, which consists of deidentified electronic health records for patients admitted to the Beth Israel Deaconess Medical Center in Boston, Massachusetts, from 2008 to 2019. MIMIC-IV is an integrated, de-identified, and comprehensive clinical dataset.

**Data Source and Inclusion Criteria**

This study includes adults with sepsis that developed AKI within 48 hours of being admitted in the ICU. The sepsis ICD9 Codes used for this study are 995-91 and 995-92 [4]. To determine which patients had AKI, the conditions used were an increase in serum creatinine by ≥ 0.3 mg/dL and an increase in serum creatinine to ≥ 1.5 times baseline, which is known or presumed to have occurred within the prior seven days.

**Variable Selection**

We gathered data across different aspects and from various sources regarding the patient’s demographic features, comorbidities, lab features, and clinical/vital features, totalling to 24 features.

The demographic features obtained were gender and age. The comorbidities include diabetes, hypertension, and chronic kidney disease. Lab features include Albumin (low levels can indicate nephrotic syndrome or malnutrition, both of which can be risk factors for AKI), Creatinine/Urine (a decrease in urine output is a diagnostic criterion for AKI), Hemoglobin (anemia can be a risk factor for AKI), HrApacheII Score (Apache II Scoring system related to heart rate), INR (coagulation abnormalities can be associated with AKI, especially in conditions like disseminated intravascular coagulation), PT (coagulation abnormalities can be associated with AKI, especially in conditions like disseminated intravascular coagulation), Sodium (electrolytes, especially potassium and sodium, as imbalances can be both a cause and a result of AKI), and Urea Nitrogen (BUN (Blood Urea Nitrogen) elevated levels can indicate reduced kidney function; high levels can indicate reduced kidney function). Clinical and Vital Features are Arterial Blood Pressure diastolic (hypotension is a significant risk factor for AKI, especially in the ICU setting), Arterial Blood Pressure systolic (hypotension is a significant risk factor for AKI, especially in the ICU setting), Heart Rate (tachycardia or bradycardia may indicate conditions leading to AKI), Heart rate Alarm - High (alarm for high heart rate), Respiratory Rate (respiratory complications can result in hypoxia, a potential cause of AKI), Temperature Celsius (fever may suggest infections, which can progress to sepsis-associated AKI), and SpO2 Desat Limit (low oxygen saturation can be a sign of hypoxia, a risk factor for AKI). A summary of these features is shown in Table 1 below.

Table 1 Features used in this study.

| **Included Features** | | | | |
| --- | --- | --- | --- | --- |
| **Scores** | **Demographics** | **Comorbidities** | **Lab Features** | **Clinical Features** |
| SOFA Score | Gender | Diabetes | Albumin | Arterial Blood Pressure diastolic |
| SAPS Score | Age | Hypertension | Creatinine | Arterial Blood Pressure systolic |
|  |  | Chronic Kidney Disease | Hemoglobin | Heart Rate |
|  |  |  | INR (PT) | Heart Rate Alarm - High |
|  |  |  | PT | Respiratory Rate |
|  |  |  | PTT | SpO2 Desat Limit |
|  |  |  | Platelet Count | Temperature (C) |
|  |  |  | Sodium |  |
|  |  |  | Urea Nitrogen |  |
|  |  |  | HrApacheII Score |  |

**Participants**

In this study, the cohort comprised 10,761 individuals diagnosed with sepsis-associated acute kidney injury (SA-AKI), as illustrated in Figure 1. The overall population included 50,920 patients admitted to the intensive care unit (ICU). Among these ICU patients, 30,290 were adults aged over 18 years who had an ICU stay exceeding 48 hours. This subset was crucial and specifically selected to ensure a comparable and relevant sample for assessing the impacts and outcomes associated with SA-AKI under extended critical care conditions. Figure 1 below shows the overview of our dataset.

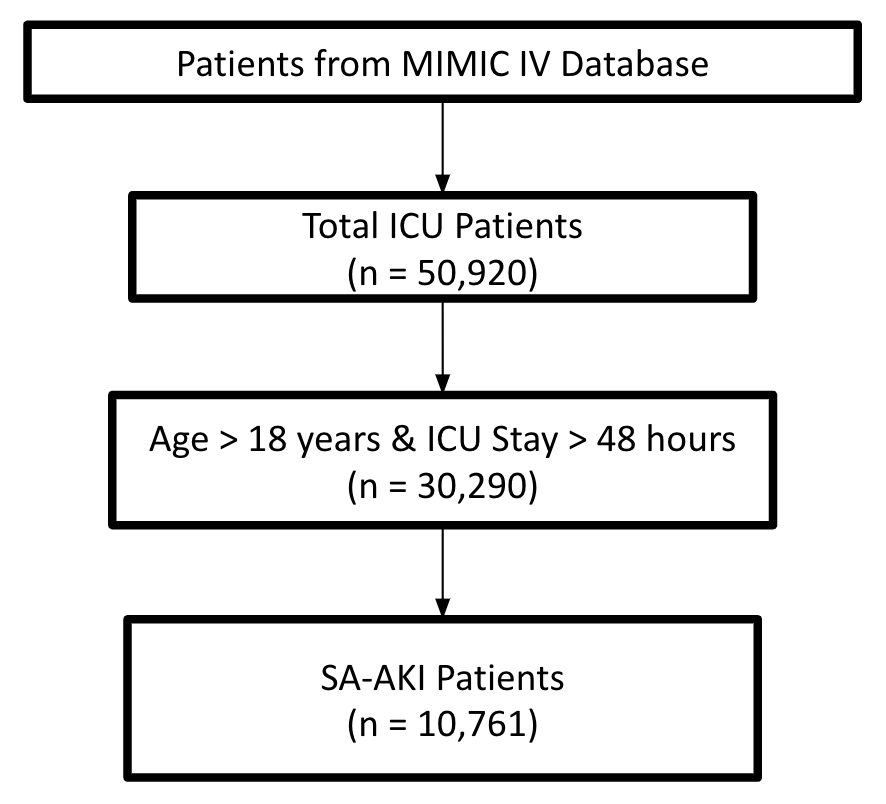


Fig. 1 Overview of the dataset. This figure illustrates an overview of the dataset. The number of total ICU patients, patients greater than 18 years old and ICU stay greater than 48 hours, and SA-AKI patients were retrieved from data engineering transformation in BigQuery.

**Data Preprocessing**

In this study, we have addressed several key data categorical variables that were encoded to convert medical and health data into a format suitable for later analysis, enabling the application of machine learning algorithms. Missing values were handled using the mean and outliers that caused imbalanced data were also handled. Feature selection was performed by analyzing univariate and bivariate statistical tests [5].

Initially, categorical variables were identified within the dataset, which were subsequently transformed into numerical format. This conversion involved employing either One-hot encoding or label encoding techniques. Outlier detection and removal were conducted using the Interquartile Range (IQR) method. Figure 2 below shows the data before outlier detection and handling.

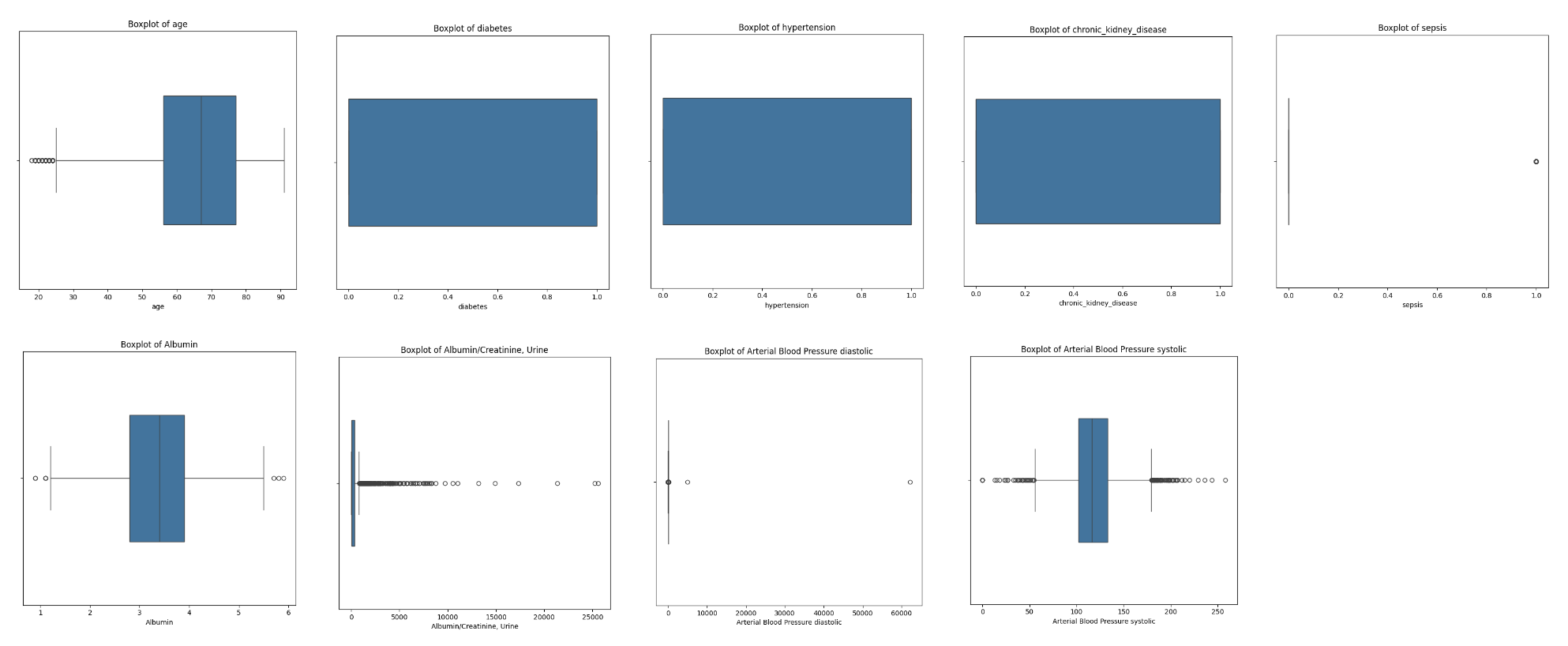


Fig. 2 Before Outlier Detection. This figure illustrates box plots for numeric variables before performing outlier detection.

By calculating the IQR for each numerical feature and defining a threshold typically set at 1.5 times the IQR, outliers were identified and subsequently eliminated from the dataset. This step ensured that the data remained representative and free from potentially skewed observations, as shown in Figure 3.

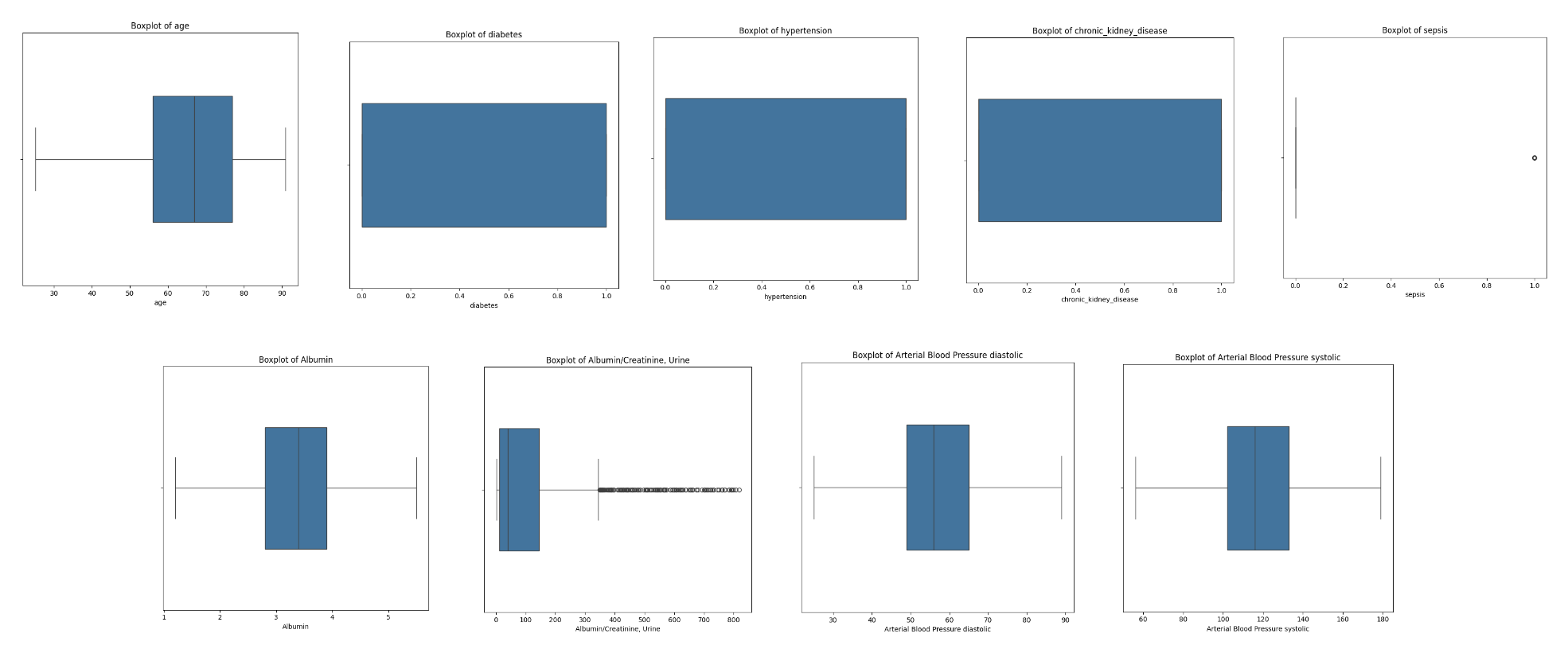


Fig. 3 After Outlier Detection with IQR. This figure illustrates the box plots for numeric variables after performing outlier detection with IQR.

Lastly, Figure 4 below shows that feature scaling using the Standard Scaler technique was applied to normalize the numerical features. This process standardized the features to have a mean of 0 and a standard deviation of 1, facilitating fair comparison and enhancing the performance of machine learning models [6, 15].

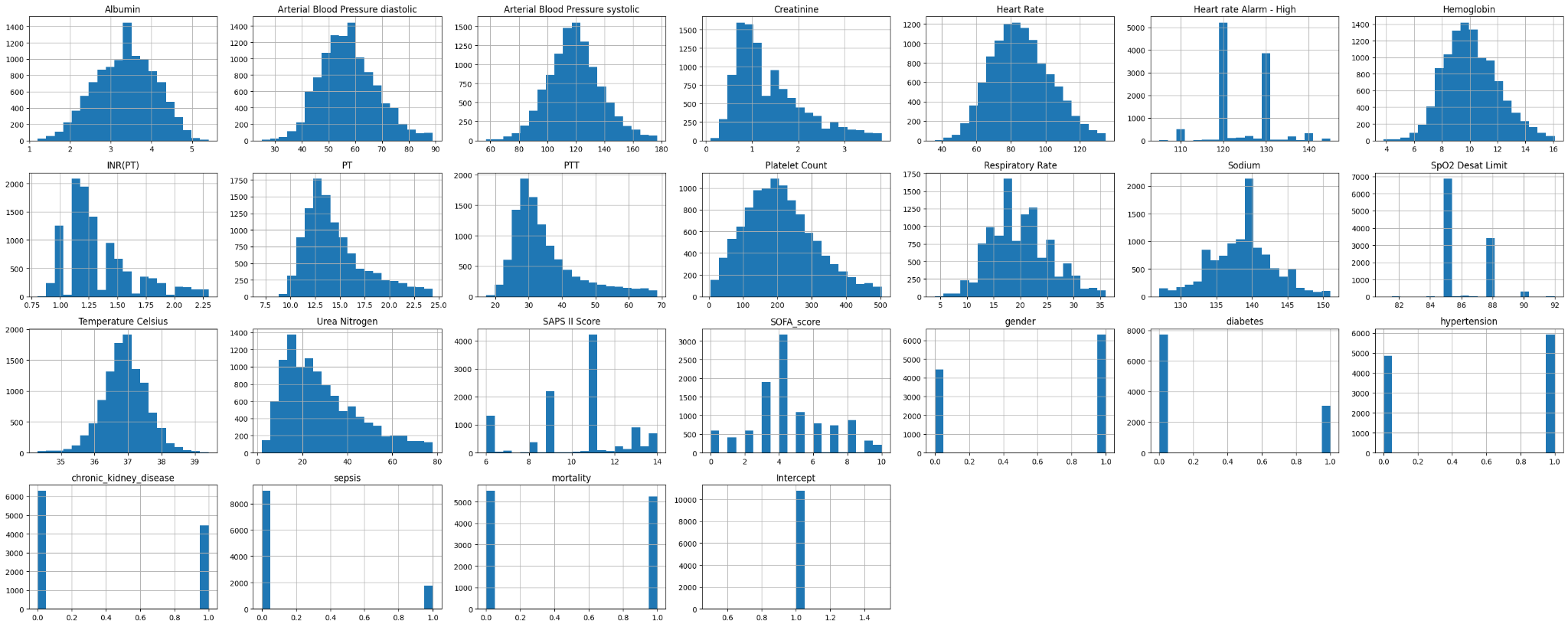


Fig. 4 Data Distribution after Imputation. This figure illustrates the data distribution for numeric variables about imputation was performed.

Subsequently, we handled missing data by using the algorithm. This iterative approach, known as Multiple Imputation by Chained Equations (MICE), effectively fills in missing information by creating imputed datasets [7]. By using this method, we improved the datasets integrity and precision allowing for thorough analyses. Additionally, incorporating up to date knowledge and treatment approaches into the datasets not only boosts the significance but also enhances the predictive accuracy and usefulness of models for clinical decision making.

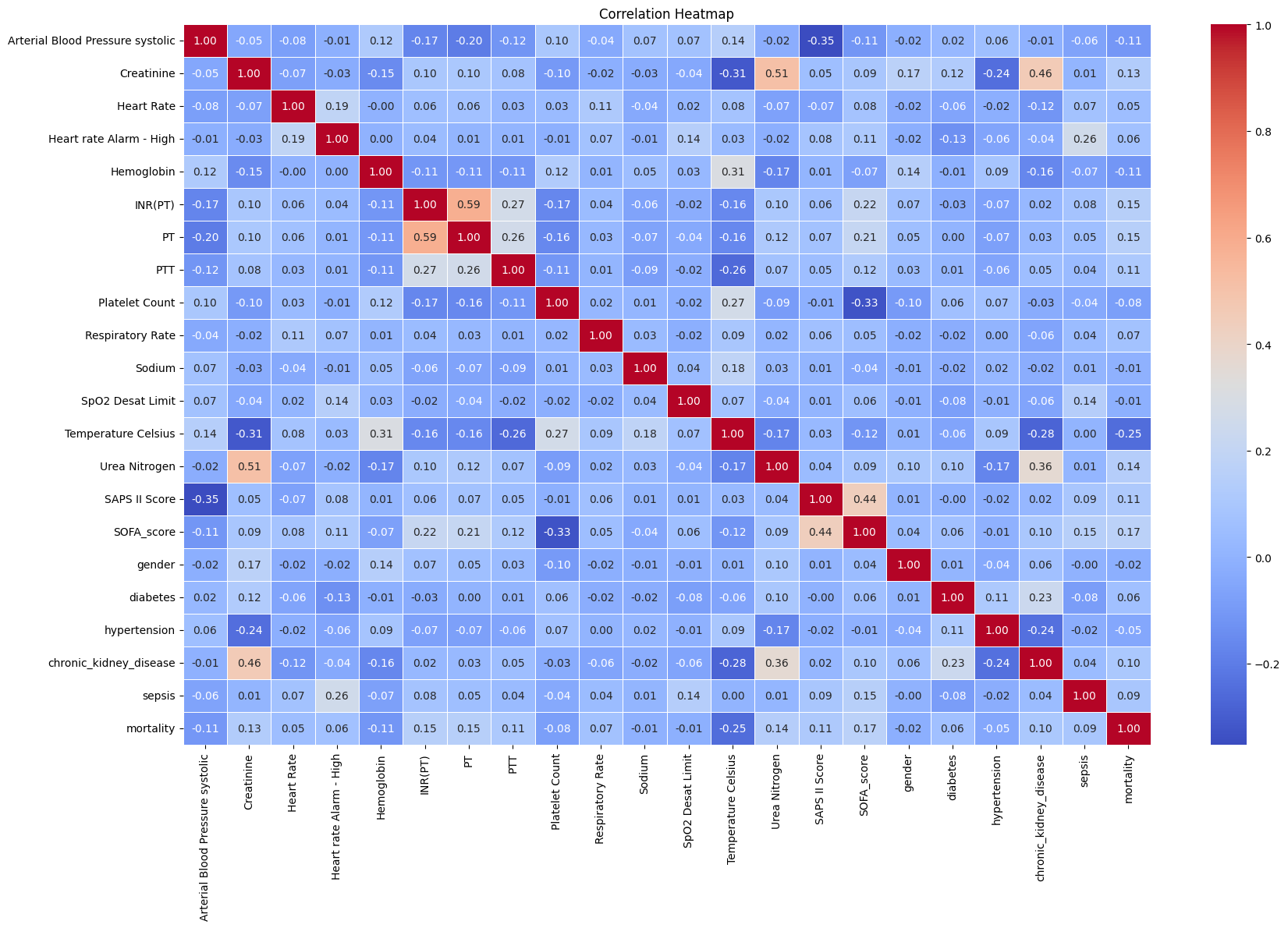


Fig. 5 Correlation Heatmap. The correlation heatmap quickly discerns which variables are strongly correlated, positively or negatively, or exhibit little to no correlation at all. There were no strong correlations; therefore, there was no potential for multicollinearity issues.

A strong positive relationship (1.00) between Hemoglobin and INR (PT) as well as between Sodium and SpO2 Desat Limit indicating that these pairs of variables tend to change together. Moreover, Temperature Celsius has a noticeable negative correlation with Urea Nitrogen, suggesting an opposite connection between these two factors. Importantly, the limited presence of multicollinearity among most variables implies that they can independently play a role in predictive models without causing bias in the results due to high correlations.

**Prediction Approach - Proposed Model**

To predict mortality in critically-ill patients with SA-AKI, we propose a machine learning model using the LightGBM gradient-boosting framework. LightGBM methods use decision tree algorithms to make predictions and classifications. Rather than constructing trees on a level basis, LightGBM aims to build trees based on leafs that have the most potential to minimize loss, making them more efficient at making such classifications.

The proposed LightGBM model was initially trained using the default parameters such as a boosting type of ‘dart’ (Dropout Additive Regression Trees) and a maximum depth of 5. The ‘dart’ boosting algorithm drops trees randomly during training, which can help prevent overfitting of the training cohort, allowing the model to generalize more when making predictions on the testing cohort. The maximum depth parameter sets a limit for how long the longest path from the root node of the decision tree to a leaf. Regulating this measure can help curb the complexity of the decision tree in order to aid in preventing overfitting on the training cohort. To enhance the performance of the model, the optimal set of parameters chosen for the model were cultivated using grid search with cross-validation as a hypertuning technique. Multiple values for different parameters such as learning rate, number of estimators, maximum depth, and gamma were evaluated in order to determine which combination of parameters would result in the highest AUROC value, leading to the optimal set of parameters. The model is then trained based on these chosen parameters, and can then be used to make predictions on an unseen set of data (the testing/validation cohort). The model’s performance at predicting mortality for this dataset is then evaluated and quantified using the AUROC score.

**Baseline Model Development**

Multiple baseline models were utilized in order to compare the results of the process mining and ML/DL approach using ML algorithms.The baseline models were trained using the MIMIC-IV data to explore other ML algorithms for prediction. To develop these models, the first step was to feed the same variables to the proposed model. Various ML algorithms were evaluated to classify critically ill patients with sepsis-associated acute kidney injury. The algorithms included were Logistic Regression, Support Vector Machine, K-nearest neighbors, Decision Trees, Random Forest, and XGBoost. The training process of these models included gridsearch to determine the most optimal model parameters [8]. This search process aimed to find the best model which was determined based on the AUROC of the validation cohort.

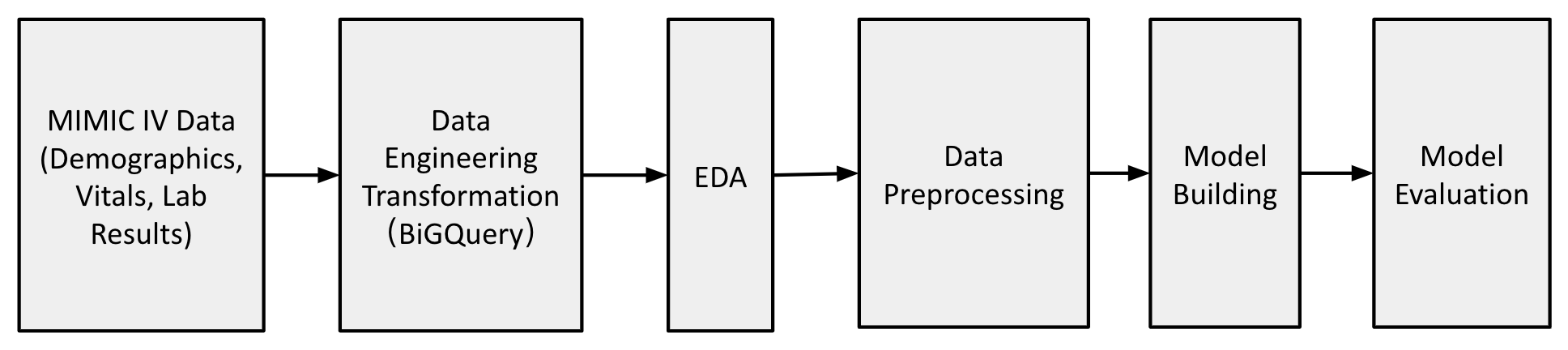


Fig. 6 Overview of the Process Flow. This figure illustrates an overview of the rigorous data process flow. From using the MIMIC IV Data, which includes comprehensive demographics, vitals, lab results, and clinical data, BigQuery was used in order to perform data engineering transformation. From there, exploratory data analysis (EDA) was conducted to uncover patterns and spot any anomalies, followed by data preprocessing to prepare the data for the models being built. Once the ML models and DL models were built, they were evaluated accordingly to ensure model’s accuracy and reliability in a clinical setting.

**Statistical Analysis Between Cohorts**

The interquartile range (IQR) and median were applied to showcase the normal distribution of continuous variables, and numbers and percentages were employed to describe categorical variables. Chi-Square tests were performed, and for continuous variables, t-tests were implied. The significant level was determined based on P < 0.05. Descriptive statistics, model development, and statistical analysis were conducted using Python, version 3.6.

**Machine Learning**

We have developed several machine learning models to evaluate their performance in our analysis. The models we built were the following Logistic Regression, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Decision Tree, Random Forest, XGBoost, LightGBM, GBM, and AdaBoost to assess their predictive accuracy and generalizable results across the study.

**Results**

**Cohort Characteristics Model Completion**

In the paper, out of 8,129 eligible patients, the median age among patients is 68.7 years old, and 42.1% of patients are female. Congestive heart failure was 34.8%, diabetes was 31.6%, and chronic pulmonary illness (2358/8129, 29.0%) were the three most prevalent comorbidities. A summary of the characteristics of the dataset is presented in Table 2 below.

Table 2 Comparison of the variables between train and validation cohorts.

| **Characteristics** | **Train cohort**  (N = 8,608) | **Validation cohort**  (N = 2,152) | **P value** |
| --- | --- | --- | --- |
| *Outcome variable N, (%)* |  |  |  |
| Mortality | 4173 (48.4) | 1072 (49.8) |  |
| *Demographics* |  |  |  |
| Age mean (std) | 65.8 (14.8) | 66.4 (14.5) | < 0.001 |
| Female (%) | 41.3 | 41.6 | < 0.001 |
| *Comorbidities N, (%)* |  |  |  |
| Diabetes | 2455 (28.5) | 594 (27.6) | < 0.001 |
| Hypertension | 4706 (54.7) | 1208 (56.1) | < 0.001 |
| Chronic Kidney Disease | 3590 (41.7) | 869 (40.4) | < 0.001 |
| *Laboratory findings mean (std)* |  |  |  |
| Albumin | 3.4 (0.7) | 3.33 (0.7) | < 0.001 |
| Creatinine, Urine | 1.4 (0.7) | 1.4 (0.7) | < 0.001 |
| Hemoglobin | 10.1 (1.9) | 10.1 (2.0) | < 0.001 |
| *Clinical/Vital mean (std)* |  |  |  |
| Heart Rate | 86.2 (17.3) | 85.9 (17.2) | < 0.001 |
| Respiratory Rate | 19.9 (5.6) | 19.8 (5.4) | < 0.001 |
| SpO2 Desat Limit | 86.1 (1.5) | 86.2 (1.6) | 0.08 |
| Temperature | 36.9 (0.7) | 36.9 (0.7) | < 0.001 |

**Evaluation Metrics**

Based on Table 3 below, the AUC score measures the area under the Receiver Operating Characteristic (ROC) curve, indicating how effectively the model can distinguish classes. Higher scores would suggest a better performance [9, 10]. The ROC curve illustrates the trade off, between sensitivity, which indicates the positive rate and specificity 1 false positive rate at various threshold levels.

Table 3 Comparison of our model AUC scores with the baseline study AUC scores.

| **Model** | **Our AUC** | **Study AUC** |
| --- | --- | --- |
| Logistic Regression | 0.741 | 0.730 |
| SVM | 0.740 | 0.680 |
| KNN | 0.627 | 0.601 |
| Decision Tree | 0.633 | 0.585 |
| Random Forest | 0.742 | 0.778 |
| XGBoost | 0.762 | 0.794 |
| Cat Boost | 0.765 | - |
| LightGBM | 0.821 | - |
| Gradient Boosting | 0.761 | - |
| Ada Boost | 0.711 | - |

From Figure 7, our study revealed that the LightGBM model exhibited higher accuracy in terms of Area Under the Curve (AUC). The LightGBM indicates its efficacy in identifying binary outcomes in comparison to other models. On the other hand, the XGBoost model performed better than all other models in the study. This difference is due to model settings, dataset features or implementation details.

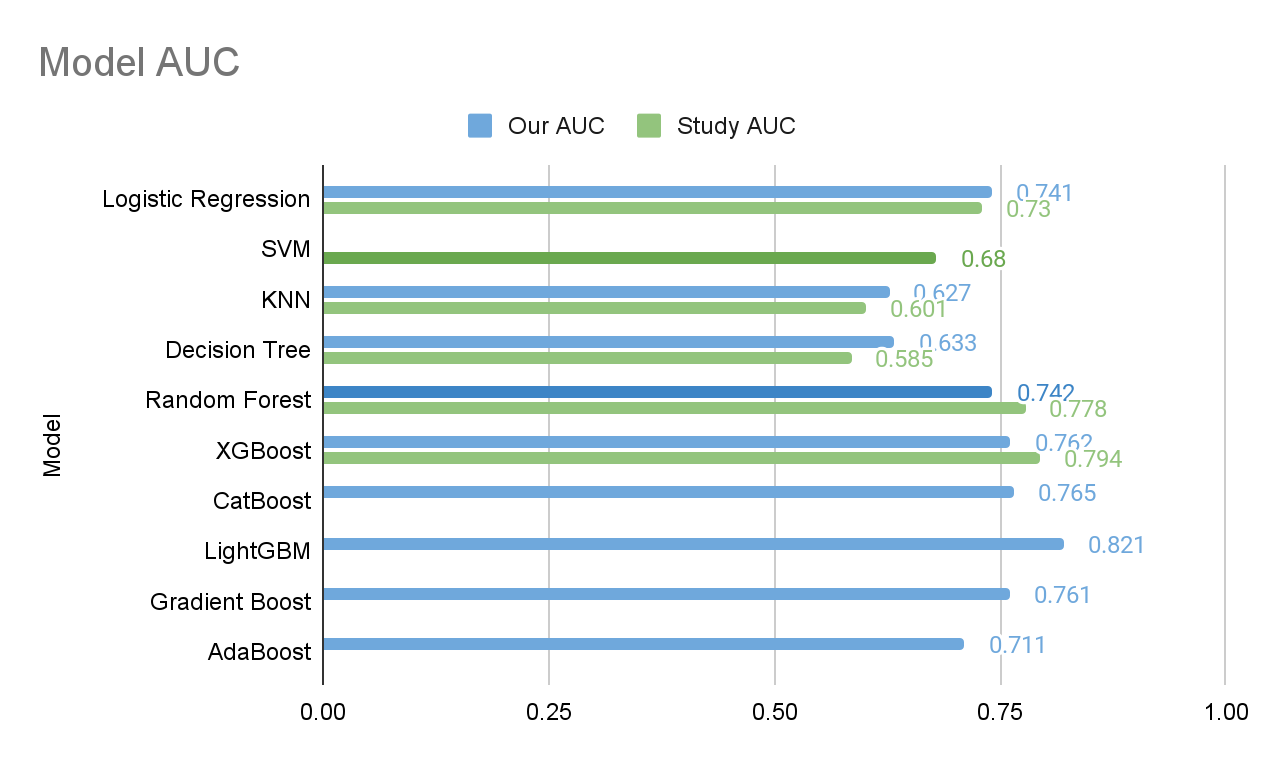


Fig. 7 Overview of the Model AUC. This figure illustrates an overview of each ML model’s AUC.

Figure 8 (below) illustrates the AUROC curve for LightGBM. The curve is positioned towards the corner of the graph, reflecting a high true positive rate and a relatively low false positive rate.

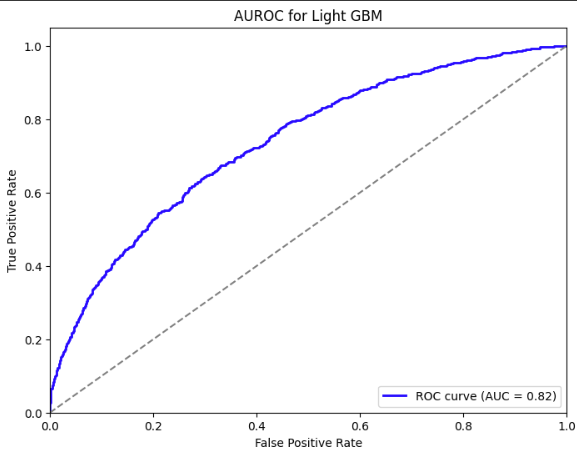
****

Fig. 8 LightGBM AUROC curve result. This figure illustrated the AUROC curve for the ML model, LightGBM.

**Discussion**

**Summary of Existing Model Compilation**

The impacts of SA-AKI can lead to long term consequences for kidney health, potentially leading to chronic kidney disease or irreversible kidney damage. Managing SA-AKI involves treating the underlying infection that supports kidney function and often requires care to monitor and manage the various complications that arise. Detection and intervention are vital for improving outcomes in individuals facing this challenging condition. Therefore, in order to determine the effectiveness of the model, we have conducted several machine learning (ML) and deep learning (DL) models.

In our study, all models were evaluated using metrics of accuracy, AUC score, and ROC curve. Accuracy measures the overall correctness of a classification model. The AUC score measures the area under the Receiver Operating Characteristic (ROC) curve indicating how effectively the model can distinguish classes. Higher scores suggest performance [9, 10]. The ROC curve illustrates the trade off between sensitivity (positive rate) and specificity (1 false positive rate) at various threshold levels.

We introduced four new machine learning models: CatBoost, LightGBM, Gradient Boost and AdaBoost. Compared to the original model, LightGBM achieved the best AUC score while in the baseline model of the study, XGBoost performed best.

**Comparison with Literature Results**

Compared to the other authors, data was gathered on the patient’s demographic features, chronic disease history, vital signs, laboratory results, treatments, illness severity scores, and outcomes. The LASSO technique was utilized to reduce the dimension of features, and SHAP was used to rank the importance of the data. The existing literature had a total of 24 features out of 44 clinical characteristics and results showed that age was a significant risk factor in critically-ill patients with SA-AKI.

The six machine learning (ML) methods the existing study built were logistic regression, support vector machine (SVM), k-nearest neighbor (KNN), decision tree, random forest (RF), and extreme gradient boosting (XGBoost). They were implemented in order to create and test each model for predicting mortality. The median and interquartile range (IQR) were applied to describe the normal distribution of continuous variables. On the other hand, numbers and percentages were used to describe categorical variables. The baseline model was also limited in its ability to conclude cause and effect due to the retrospective modeling study being conducted at a single location, utilizing the MIMIC IV database.

After comparing the AUC score, the literature study performed decision curve analysis (DCA) and plotted calibration curves. The Local Interpretable Model-Agnostic Explanations (LIME) techniques were used to fit the model’s expected behavior. Lastly, the study conducted a sensitivity analysis of the results. XGBoost has the best AUC score, which was stated as the best model for the dataset. From the SHAP values, sequential organ failure assessment score, respiration, simplified acute physiology score II, and age were the variables with the most influence in the XGBoost model. To summarize, there were no significant differences in the baseline features between the training and validation sets.

Table 4 Summary of the literature review.

| **Study** | **Method** | **Results** |
| --- | --- | --- |
| Li el al. (2023) [3] | Logistic Regression, SVM, KNN, Decision Tree, Random Forest,XGBoost, LASSO Regression | XGBoost (AUC 0.794) |
| Jiang et al.(2023) [11] | Logistic Regression | Logistic Regression (95% CI [0.45%–0.88%]) - Validation cohorts (0.780 (95% CI: 0.75–0.82) and 0.80 (95% CI: 0.75–0.85)) |
| Wang et al. (2022) [12] | Logistic Regression | Logistic Regression (95% CI [46.8%–70.3%]) - AKI was 58.6% (95% CI [46.8%–70.3%]) in MIMIC IV and 44.6% (95% CI [12.7%–76.4%]) |
| Zhou et al.(2023) [13] | CatBoost, Accuracy (ACC), Youden Index, Sensitivity, Specificity, F1 Score, Positive Predictive Value (PPV), Negative Predictive Value (NPV) | CatBoost (AUC 0.827) - ACC: 75%, Youden Index: 50%, Sensitivity: 75%, Specificity: 75%, F1 score: 0.56, PPV: 44%, NPV: 92%] |
| Xiao-Qin et al. (2022) [14] | XGBoost | XGBoost (AUC 0.848 - 0.804) |

From our method, we performed variable selection differently to determine which features would have more significance, as well as including other machine learning models. Drawbacks of the existing literature include data from a smaller sample size, leading to the results not representing the broader population, and relying solely on conducted internal model validation. The proposed method outperformed the existing literature since the existing literature calculated specific missing data using the imputation method, which could have led to deviation from the actual value. In addition, our proposed model not only adopted the same metrics from the main literature review, but also further built more advanced ML models, such as CatBoost, LightGBM, Gradient Boosting, and AdaBoost.

**Limitations**

The limitations of the study are data quality, data imbalance, and feature selection and dimensionality. For data quality, ML models rely on the input data’s quality and completeness. Specifically in this case, medical data may consist of incomplete fields, errors, or have inconsistencies from not considering multiple center validation, affecting the performance and generalization of the model. Data imbalance comes from the prediction for mortality is affected by class imbalance, which can lead to biased models on the majority class and poor prediction performance on the minority class. Feature selection and dimensionality arises from selecting specific characteristics from a wide range of variables, given the numerous clinical factors and laboratory results available. In addition, the relationships between these characteristics may lead to overfitting and increased computing complexities. Moreover, potential improvements could involve utilizing feature engineering techniques and automated feature selection algorithms. All of the improvements could identify relevant qualities from large healthcare datasets. The above strategies for our model seek to improve the comprehensibility of the data, and further address problems related to overfitting and ultimately enhance precision.

**Conclusion**

As sepsis-associated acute kidney injury (SA-AKI) is a health issue where the kidneys struggle to properly filter and remove waste materials and extra fluids as a result of sepsis. We developed and validated machine learning models and deep learning models that successfully predict early mortality risk in patients with SA-AKI. The study represents a significant step forward in the application of machine learning. Out of all the algorithms, the LightGBM model is the most efficient. In addition, the SHAP values approach revealed that age, SAPS II, respiratory rate, and SOFA score were significant predictors of death in patients with SA-AKI. Clinical prediction would benefit from these findings.

To ensure that this approach in machine learning is broadly applicable and transferable across scenarios and connected to improved clinical judgment and results. It is necessary to conduct studies involving multiple centers. Additionally, the MIMIC IV dataset contains information such as detailed clinical narratives and examples that can serve as input data for the models. This further opens up more research opportunities. As a result, it is critical to incorporate advanced automated feature selection tools and incorporate sophisticated feature engineering strategies.

**Abbreviations**

ML: Machine Learning

DL: Deep Learning

SA-AKI: Sepsis-Associated Acute Kidney Injury

SVM: Support Vector Machine

KNN: K-Nearest Neighbors

AUC: Area Under the Curve

ICU: Intensive Care Unit

IQR: InterQuartile Range

EDA: Exploratory Data Analysis

SOFA: The Sequential Organ Failure Assessment

SHAP: Simplified Acute Physiology Score

MICE: Multiple Imputation by Chained Equations

MIMIC-IV: Medical Information Mart for Intensive Care IV

DCA: Decision Curve Analysis

NN: Neural Network

**Acknowledgements**

The authors would like to express our special appreciation and thanks to our Professor Maryam Pishgar for guidance and would also like to thank the developers of MIMIC-IV for providing a detailed and comprehensive public EHR dataset.

**Declarations**

**Ethics Approval and Consent to Participate**

The dataset supporting the conclusions of this article is available in the Medical Information Mart for Intensive Care version IV (MIMIC-IV). This database is a public de-identified database thus informed consent and approval of the Institutional Review Board was waived. All methods were performed in accordance with the relevant guidelines and regulations.

**Consent for Publication**

Not applicable.

**Author Details**

Department of Industrial and Systems Engineering, University of Southern California, 3470 Trousdale Parkway, Los Angeles, CA 90089, USA.

**References**

[1] Camacho-Collados, J., Marín, C., Anido-Rifón, L. E., & Lanza-Gutiérrez, J. M. (2019). PM4Py: A Python Library for Process Mining. Journal of Healthcare Informatics Research, 3, 1-17. <https://doi.org/10.1007/s41666-019-00052-1>

[2] Nasir, A., et al. (2024). Prediction of mortality in critically ill patients with sepsis-associated acute kidney injury: an ensemble machine learning approach. Critical Care, 28(1), 77. https://doi.org/10.1186/s13054-024-04877-4

[3] Li, X., Wu, R., Zhao, W. et al. Machine learning algorithm to predict mortality in critically ill patients with sepsis-associated acute kidney injury. Sci Rep 13, 5223 (2023).<https://doi.org/10.1038/s41598-023-32160-z>

[4] Makam, Anil N., et al. "Use and Outcomes of Venovenous Extracorporeal Membrane Oxygenation for Acute Respiratory Failure in Adults: A National Database Analysis." Annals of the American Thoracic Society, vol. 11, no. 6, 2014, pp. 891-897. https://pubmed.ncbi.nlm.nih.gov/24871149/

[5] Paudel, S., & Khanal, S. (2021). Applications of Machine Learning and Deep Learning in Healthcare Systems: A Systematic Review. SN Operations Research Forum. https://doi.org/10.1007/s40747-021-00637-x

[6] Ongolo-Zogo, P., Lavis, J. N., Tomson, G., Sewankambo, N. K., & others. (2015). How to identify, assess and utilize expert knowledge as evidence for policy-making in the public health domain: A guide for policymakers and researchers. BMC Medical Research Methodology, 15(1), 1-11. https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-015-0022-1

[7] Hernán, Miguel A., et al. "Estimating the Effect of Acutely Harmful Exposures on Outcomes: The Limited Utility of Including all Acute Disease Events." American Journal of Epidemiology, vol. 179, no. 6, 2014, pp. 764-771. https://academic.oup.com/aje/article/179/6/764/107562

[8] Ntalampiras, S., & Manolopoulos, Y. (2016). An overview of recent advances in event detection in multimedia content. Neurocomputing, 180, 140-153. https://doi.org/10.1016/j.neucom.2015.08.108

[9] Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A., van Ginneken, B., & Sánchez, C. I. (2017). Medical image segmentation using deep learning: A review. Physics in Medicine & Biology, 61(21), R41-R89. https://doi.org/10.1088/1361-6560/aab4b1

[10] Banerjee, S., & Kulkarni, R. (2021). Deep learning for COVID-19 detection and prognostication using chest radiographs and CT scans: A systematic review and meta-analysis. PLoS One, 16(12), e0261053. https://doi.org/10.1371/journal.pone.0261053

[11] Wang Z, Weng J, Yang J, Zhou X, Xu Z, Hou R, Zhou Z, Wang L, Chen C, Jin S. Acute kidney injury-attributable mortality in critically ill patients with sepsis. PeerJ. 2022;10:e13184. Published online 2022 Mar 25. doi: 10.7717/peerj.13184.

[12] Zhou H, Liu L, Zhao Q, Jin X, Peng Z, Wang W, Huang L, Xie Y, Xu H, Tao L, Xiao X, Nie W, Liu F, Li L, Yuan Q. Machine learning for the prediction of all-cause mortality in patients with sepsis-associated acute kidney injury during hospitalization. Front Immunol. 2023 Apr 3;14:1140755. doi: 10.3389/fimmu.2023.1140755. PMID: 37077912; PMCID: PMC10106833.

[13] Xiao-Qin Luo, Ping Yan, Shao-Bin Duan, Yi-Xin Kang, Ying-Hao Deng, Qian Liu, Ting Wu, and Xi Wu. "Development and Validation of Machine Learning Models for Real-Time Mortality Prediction in Critically Ill Patients With Sepsis-Associated Acute Kidney Injury." Frontiers in Medicine, Section: Intensive Care Medicine and Anesthesiology, Volume 9, 2022, Article ID 853102. https://doi.org/10.3389/fmed.2022.853102.

[14] Jiang W., Zhang, C., Yu, J., Shao, J., & Zheng, R. (2023). Development and validation of a nomogram for predicting in-hospital mortality of elderly patients with persistent sepsis-associated acute kidney injury in intensive care units: A retrospective cohort study using the MIMIC-IV database. BMJ Open, 13(3), 1. doi:https://doi.org/10.1136/bmjopen-2022-069824

[4] Makam, Anil N., et al. "Use and Outcomes of Venovenous Extracorporeal Membrane Oxygenation for Acute Respiratory Failure in Adults: A National Database Analysis." Annals of the American Thoracic Society, vol. 11, no. 6, 2014, pp. 891-897. https://pubmed.ncbi.nlm.nih.gov/24871149/

[15] Singh, A., Srivastava, P. R., & Sharma, M. (2021). A comprehensive review on outlier detection techniques: Critical aspects, applications, and challenges. Journal of Network and Computer Applications, 180, 103077. https://www.sciencedirect.com/science/article/abs/pii/S1574013720304068