Nomogram Development for Predicting In-Hospital Mortality in Sepsis-Associated Elderly Patients

N. Abirami, T. Su, R.Tanwar, S.Rajasekaran, V. Warke

Abstract

Background: Sepsis-Associated Acute Kidney Injury (SA-AKI) is a disorder in which sepsis, a severe and sometimes fatal response to infection, causes the kidneys to function worse. Acute kidney damage can result from the kidneys' general critical function of filtering waste materials and extra fluid from the blood. As a result, physicians use additional instruments, including nomograms, to forecast incipient old patients.

Methods: Using improved nomograms, we demonstrated a machine learning method for predicting death in critically sick patients with sepsis-associated patients. We have incorporated in A rise in serum creatinine of > 0.3 mg/dL within 48 hours and an increase in serum creatinine of ≥ 1.5 times baseline, which is known or assumed to have happened within the previous seven days, were the criteria used to identify which individuals had AKI. In order to compare outcomes and improve nomogram prediction, a number of machine learning (ML) models were employed as baseline models. .

Results: Our proposed model produced an area under the receiver operating characteristics (AUROC) of 0.8776, which is an improvement over the results of the baseline model and existing literature. We used the Medical Information Mart for Intensive Care IV (MIMIC-IV) database of 10,761 ICU patients with SA-AKI for the selections discussed in this research paper.

Conclusions: The proposed approach, with our updated feature selection and implemented nomogram, 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, nomogram

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, 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 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 [3].

Our study is different from earlier research in several significant respects. Unlike previous studies that mostly relied on traditional statistical techniques or a single machine learning model, we first construct and evaluate a number of advanced algorithms, including Random Forest, CatBoost, and Neural Networks, in order to discover the optimal way for mortality prediction. Second, our work encompasses a greater variety of clinical parameters and uses advanced feature selection techniques based on variance thresholds rather than conventional LASSO approaches.

Our method is unique since it generates nomograms that are more clinically interpretable and have higher prediction accuracy. Our models are more relevant to actual clinical decision-making because they provide information on the relative importance of multiple clinical parameters, in contrast to earlier research that just examined prediction accuracy. Our study combines the predictive capability of advanced machine learning with interpretable visualization tools, bridging the gap between clinical utility and statistical sophistication. Furthermore, our study overcomes the limitations of earlier research by utilizing a larger feature set and more reliable preprocessing techniques. Our findings have practical ramifications that could result in better clinical decision-making in intensive care units and early identification of high-risk patients to improve patient outcomes through more targeted medicines..

Methods

Database Introduction

Our analysis includes deidentified electronic health data for individuals admitted to the Beth Israel Deaconess Medical Center in Boston, Massachusetts, between 2008 and 2019. This database is known as the Medical Information Mart for Intensive Care IV (MIMIC-IV). Comprehensive, de-identified, and integrated is the MIMIC-IV clinical dataset.

Data Source and Inclusion Criteria

Individuals with sepsis who experienced AKI within 48 hours after being hospitalized to the intensive care unit are included in this research. For this investigation, the ICD9 codes for sepsis are 995-91 and 995-92 [4]. The criteria used to identify individuals with AKI were an increase in

serum creatinine of ≥ 0.3 mg/dL and a rise in serum creatinine of > 1.5 times baseline, which is known or assumed to have happened within the previous 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.

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	

Table 1 Features used in this study.

Participants

In this study, the cohort comprised 10,761 and 1,024 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.

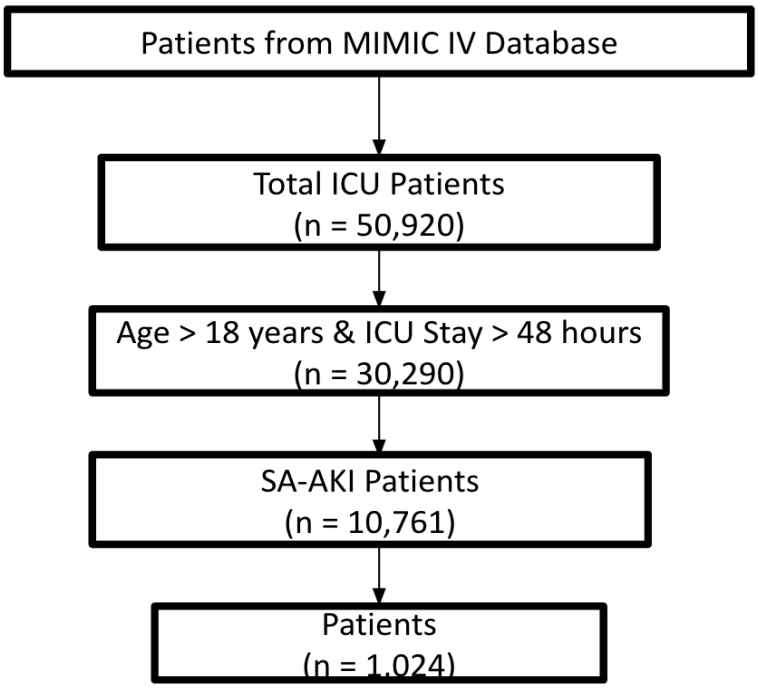

Fig. 1 Overview of the dataset. This figure illustrates an overview of the dataset. The number of total ICU patients, patients greater than 18 years old and ICU stay greater than 48 hours, and SA-AKI patients were retrieved from data engineering transformations.

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 and 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. While maintaining crucial information for model training, this thorough preparation method guaranteed data quality. Histograms and boxplots were used to illustrate the distribution of characteristics before and after preprocessing in order to confirm the efficacy of our methodology.

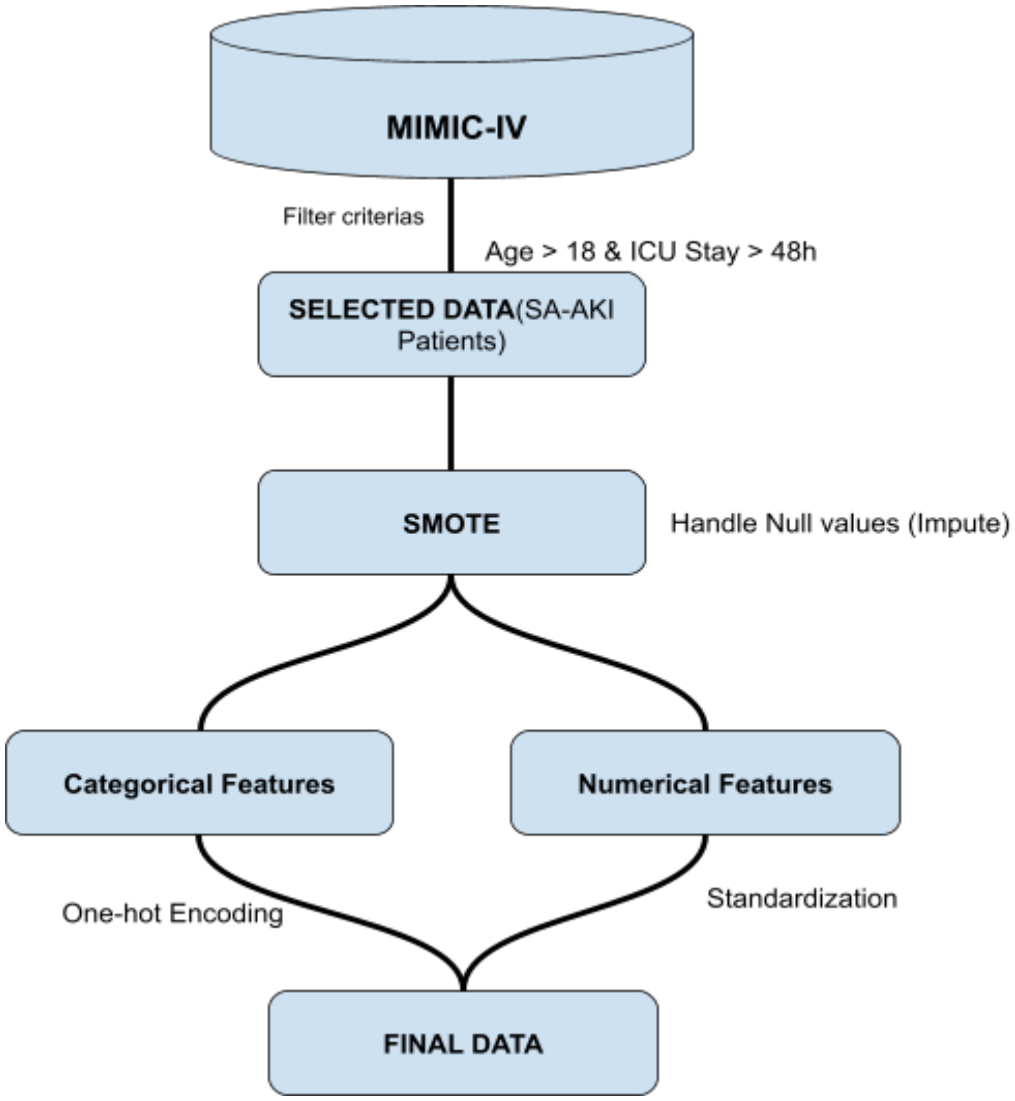

Figure 2 illustrates the data workflow and preprocessing pipeline used; this preprocessing flow

sets the place for applying various models for analysis. By calculating the IQR for each numerical feature and defining a threshold usually 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 the Figure 3.

Lastly, Figure 3 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].

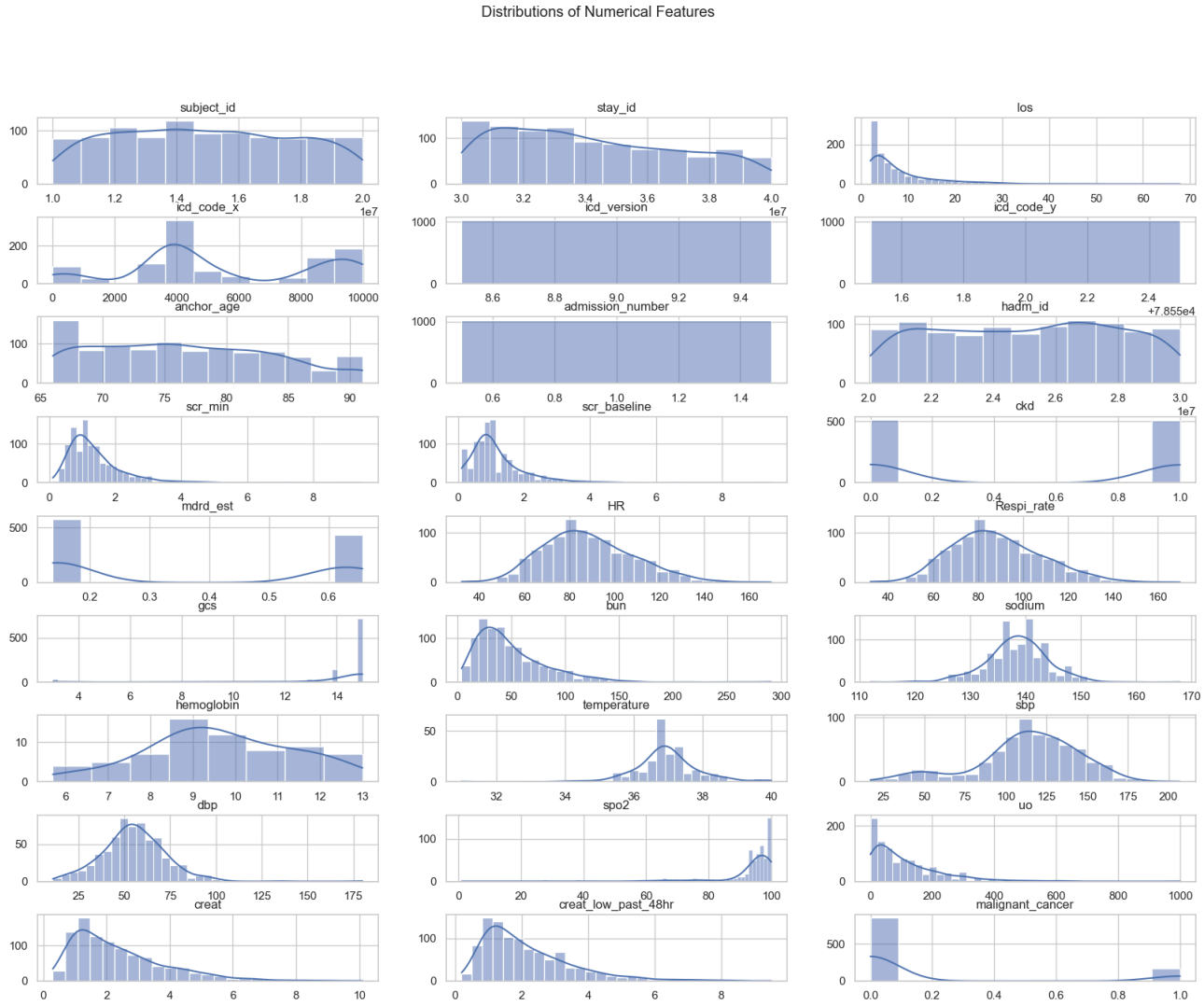

Fig. 3 Data Distribution after Imputation.

The data distribution for the numerical variables about the imputation process is depicted in this figure. Missing information was handled quickly and easily. The data's wide distribution was maintained while providing resilience against outliers by imputed missing values for numerical variables using median values. Mode imputation was applied to categorical data, which uses the most common category to fill in missing values. This approach, which was less complicated than

MICE, preserved data integrity and performed well for our study while avoiding the computational cost of more sophisticated imputation techniques.

.

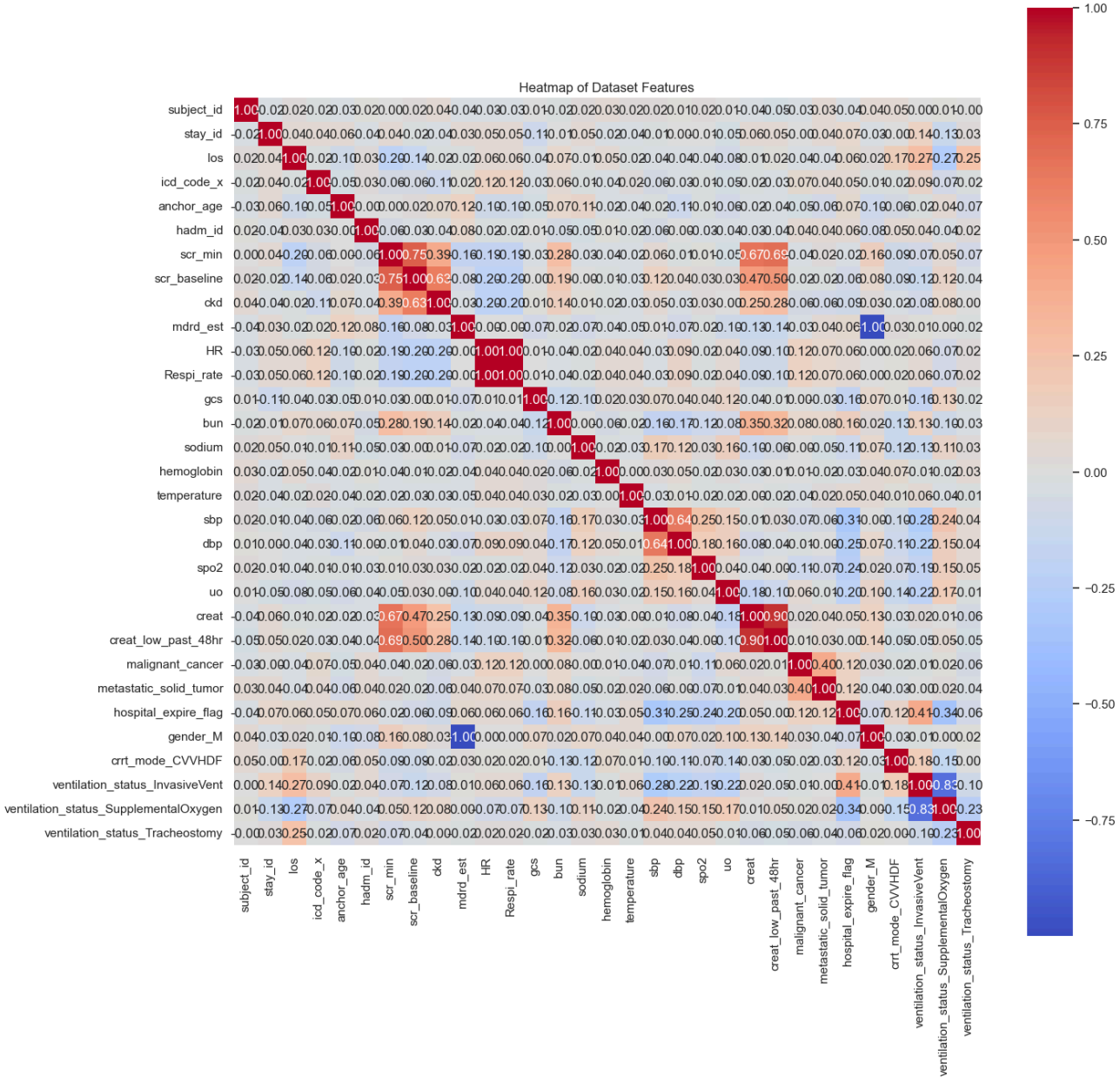

Fig. 4 Correlation Heatmap. Variables that are substantially connected, favorably or negatively correlated, or show little to no association at all can be rapidly identified using the correlation heatmap. There was no possibility of multicollinearity problems because there were no significant relationships.

SpO2 Desat Limit and Sodium have a high positive connection (1.00), suggesting that these two variables often fluctuate together. Additionally, there is a clear negative link between temperature in degrees Celsius and ventilation status, indicating that these two variables have the opposite relationship with oxygen saturation. Significantly, the fact that most variables have little

multicollinearity suggests that they can contribute to predictive models on their own without skewing the outcomes because of strong correlations.

Prediction Approach - Proposed Model

We present a machine learning model that uses the logistic regression framework to predict mortality in critically unwell SA-AKI patients. Decision tree algorithms are used by LightGBM techniques to classify and predict. LightGBM is more effective at classifying trees because it builds trees based on the leaves that have the greatest potential to reduce loss rather than building trees flat. The default settings, including a maximum depth of five and a boosting type of "dart" (Dropout Additive Regression Trees), were used to train the suggested logistic regression model. In order to avoid overfitting the training cohort, the "dart" boosting technique randomly eliminates trees during training. This enables the model to make predictions on the testing cohort with greater generality. The maximum depth parameter places a cap on the length of the decision tree's longest path from the root node to a leaf. In order to help avoid overfitting on the training cohort, controlling this parameter can help reduce the decision tree's complexity. In order to improve the model's performance, grid search with cross-validation was used as a hypertuning strategy to nurture the ideal set of parameters. To identify the ideal set of parameters, a number of values for various parameters, including learning rate, number of estimators, maximum depth, and gamma, were assessed. The goal was to find the combination of parameters that would produce the highest AUROC rating. These selected parameters are then used to train the model, 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 ROC score.

Ablation study

Knowing how each feature affects the model's performance is essential for predictive modeling. By methodically eliminating each characteristic and monitoring the difference in model performance, an ablation research sheds light on how important each one is. This method aids in identifying aspects that are essential for forecasts as well as those that may be unnecessary or even harmful. A Random Forest classifier—which is well-known for its effectiveness in managing intricate datasets with numerous feature interactions—was used for our study. Assessing each feature's effect on the model's capacity to forecast in-hospital mortality, as indicated by the Receiver Operating Characteristics (ROC) curve's Area Under the Curve (AUC), was the goal. A Random Forest model was used for the ablation investigation, and it had the the following steps:

Data Preparation: The dataset was divided into subgroups for testing (20%) and training (80%). Model Training: The entire dataset was used to train a Random Forest classifier. Baseline Performance: To create a point of comparison, the fully-featured model's AUC was computed.Model Evaluation and Feature Removal: The model was retrained, the AUC was

recalculated, and each feature was extracted from the dataset separately. Every feature in the dataset underwent this iteration.

Feature	AUC with Feature	AUC without Feature	Difference
subject_id	0.875	0.871	0.004
stay_id	0.875	0.876	-0.001
los	0.875	0.866	0.009
icd_code_x	0.875	0.867	0.008
anchor_age	0.875	0.867	0.008
hadm_id	0.875	0.860	0.015
scr_min	0.875	0.869	0.006
scr_baseline	0.875	0.872	0.003
ckd	0.875	0.882	-0.007
mdrd_est	0.875	0.881	-0.006
HR	0.875	0.877	-0.002
Respi_rate	0.875	0.877	-0.002
gcs	0.875	0.882	-0.007
bun	0.875	0.874	0.001
sodium	0.875	0.863	0.012
hemoglobin	0.875	0.881	-0.006
temperature	0.875	0.878	-0.003
sbp	0.875	0.889	-0.014
dbp	0.875	0.863	0.012
spo2	0.875	0.875	0.000

uo	0.875	0.862	0.013
creat	0.875	0.888	-0.013
creat_low_past_48hr	0.875	0.873	0.002
malignant_cancer	0.875	0.884	-0.009
metastatic_solid_tumor	0.875	0.878	-0.003
gender	0.875	0.870	0.005
crrt_mode_CVVHDF	0.875	0.875	0.000
ventilation_status_InvasiveVent	0.875	0.874	0.001
ventilation_status_Supplemental Oxygen	0.875	0.883	-0.008
ventilation_status_Tracheostomy	0.875	0.874	0.001

` Table 2 shows the Ablation study results

A number of important characteristics that have a big influence on model performance were brought to light by the ablation investigation. When eliminated, features including "los," "icd_code_x," and "anchor_age" demonstrated a discernible drop in AUC, highlighting their significance in patient outcome prediction. However, when eliminated, characteristics such as "stay_id" and "ventilation_status_SupplementalOxygen" did not significantly alter or even slightly enhance AUC, indicating that they may be redundant or not predictive of the result in this model configuration.

Baseline Model Development

To compare the outcomes of the process mining and ML approach utilizing ML algorithms, a baseline model was employed.In order to test several machine learning techniques for prediction, the baseline model was trained using the MIMIC-IV data. The same variables were fed into the suggested model as the initial stage in developing the model. In order to categorize critically ill patients with acute kidney injury linked to sepsis, several machine learning algorithms were tested. Among these algorithms were Random Forest, K-nearest neighbors, Support Vector Machine, Logistic Regression, and Decision Trees. These models were trained using gridsearch to find the best model parameters [8]. Finding the best model was the goal of this search method, which was based on the validation cohort's AUROC.

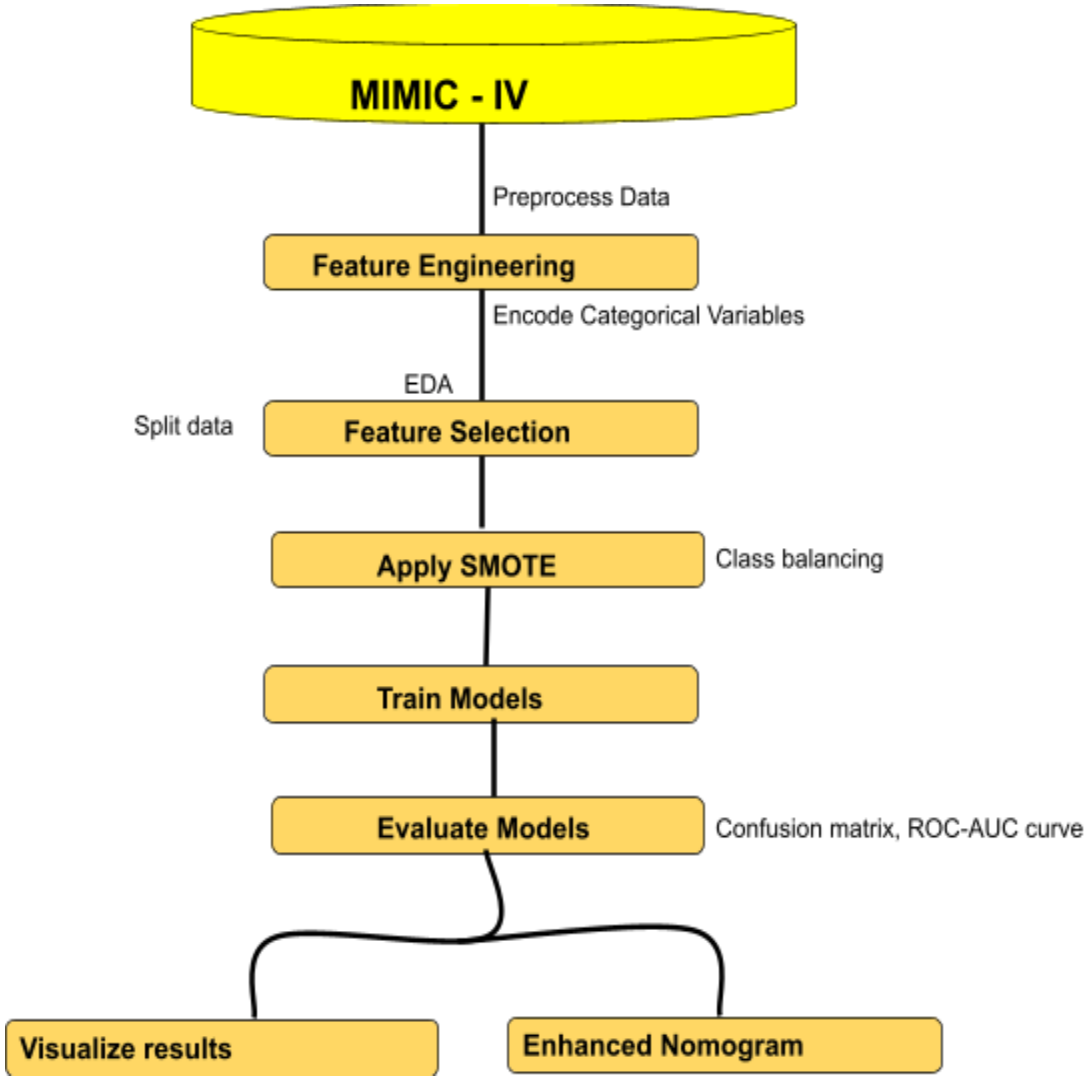

Fig. 5: Process Flow Overview. An outline of the rigorous data process flow is shown in this picture. Data engineering transformation was carried out utilizing the MIMIC IV Data, which comprises thorough demographics, vitals, lab results, and clinical data. After that, in order to prepare the data for the models being developed, exploratory data analysis (EDA) was carried out to find trends and identify any anomalies. Following their construction, the ML and DL models were assessed appropriately to guarantee their precision and dependability in a clinical context, and the outcomes were displayed.

Statistical Analysis Between Cohorts

Continuous variables were shown to have a normal distribution using the interquartile range (IQR) and median, In contrast, categorical variables were expressed in terms of percentages and numbers. Chi-Square and t-tests were used for continuous variables, respectively. Based on $P < 0.05$, the significant level was established. Python 3.6 was utilized to perform statistical analysis, model creation, and descriptive statistics.

Results

Cohort Characteristics Model Completion

Data from the MIMIC-IV database, which originally contained information on 50,920 patients in critical care units, was utilized in the study. Patients required to be at least 18 years old and have spent more than 48 hours in the intensive care unit (ICU) in order to create a focused research population and meet stringent inclusion requirements. Following the initial filtering process, 30,290 patients made up the population. The data that met certain diagnostic criteria for damage associated with sepsis were included in the filtered sample. The ICD-9 codes 995-91, 995-92, and 785-52 were used to identify sepsis, severe sepsis, and septic shock, respectively.

10,761 SA-AKI cases were found using this identification method. 1,024 people made up the final research group, for whom comprehensive clinical data was collected. In addition to accurate clinical measurements like vital signs (heart rate, blood pressure, and respiration rate), Glasgow Coma Scale scores, and significant laboratory values like BUN, creatinine, and hemoglobin levels, this data also included demographic variables like age and gender. Additionally, treatment data was gathered, with a focus on the usage of Continuous Renal Replacement Therapy (CRRT) and the status of mechanical ventilation. Important comorbidities, such as cancer and chronic renal illness, were also disclosed for each patient.

In order to give a complete picture of older persons with acute renal impairment associated with sepsis, this carefully chosen cohort was designed to take into account the complexity of their clinical presentations and the available treatments. The large amount of data was utilized to create robust prediction models that primarily focused on variables that could be used to forecast in-hospital mortality. For this high-risk patient population, the cohort offers a strong basis for assessing outcomes and creating prediction models because of strict selection standards and meticulous data gathering methods.

Characteristics	**Train cohort** (N = 8,608)	**Validation cohort** (N = 2,152)	**P value**
Mortality	4173 (48.4)	1072 (49.8)	0.03
Age mean (std)	65.8 (14.8)	66.4 (14.5)	< 0.001
Female (%)	41.3	41.6	< 0.001
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

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
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

Table 3 Comparison of the variables between train and validation cohorts.

Evaluation Metrics

Based on Table 4, 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.

Model	**Our AUC**	**Study AUC**
Logistic Regression	0.805	0.730
SVM	0.809	0.680
KNN	0.795	0.601
Decision Tree	0.637	0.585
Random Forest	0.877	0.778
XGBoost	0.80	0.794
AdaBoostClassifier	0.82	-
GradientBoostingClassifier	0.83	-

Table 4 Comparison of our model AUC scores with the baseline study AUC scores.

From Figure 6, 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.

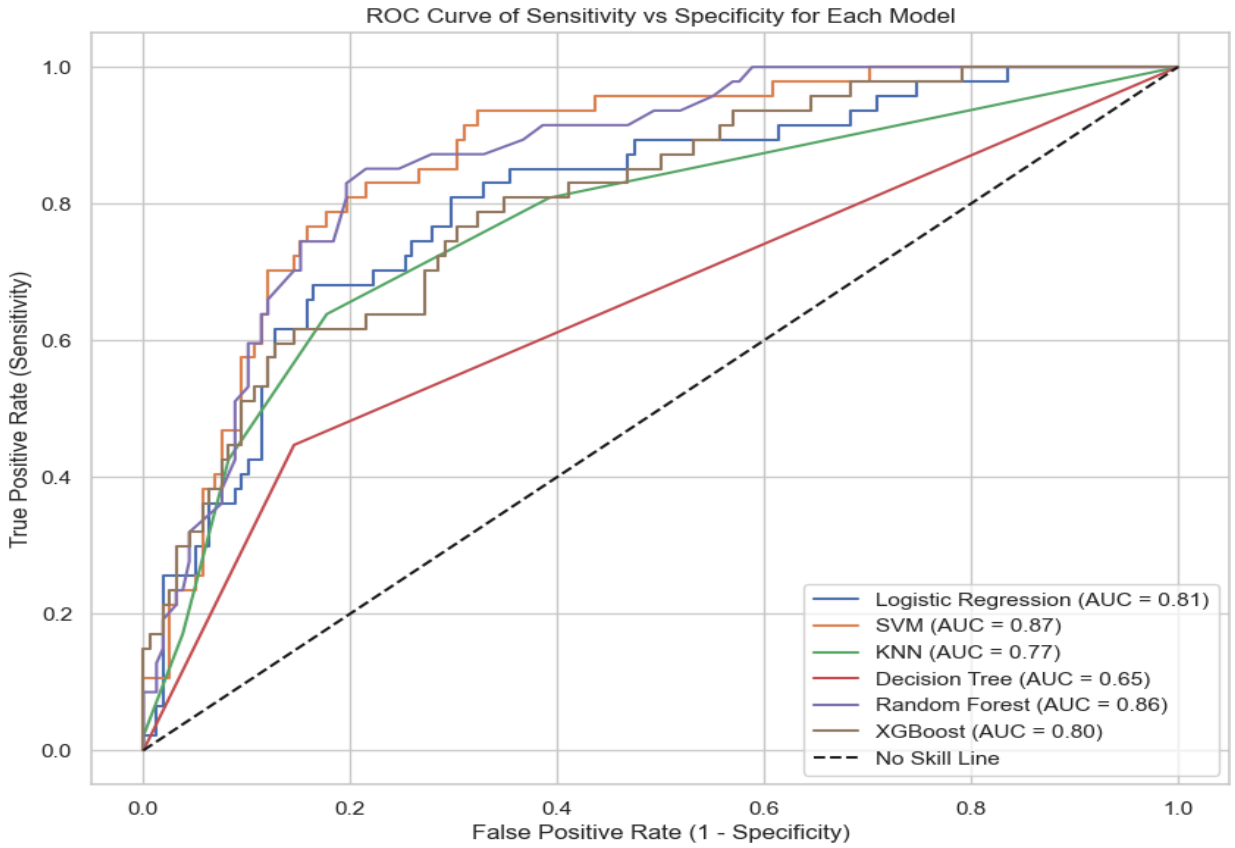

Fig. 6 Overview of the Model ROC Curve. This figure illustrates an overview of each ML model's ROC Curve and Sensitivity .

Figure 7 illustrates the ROC 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.

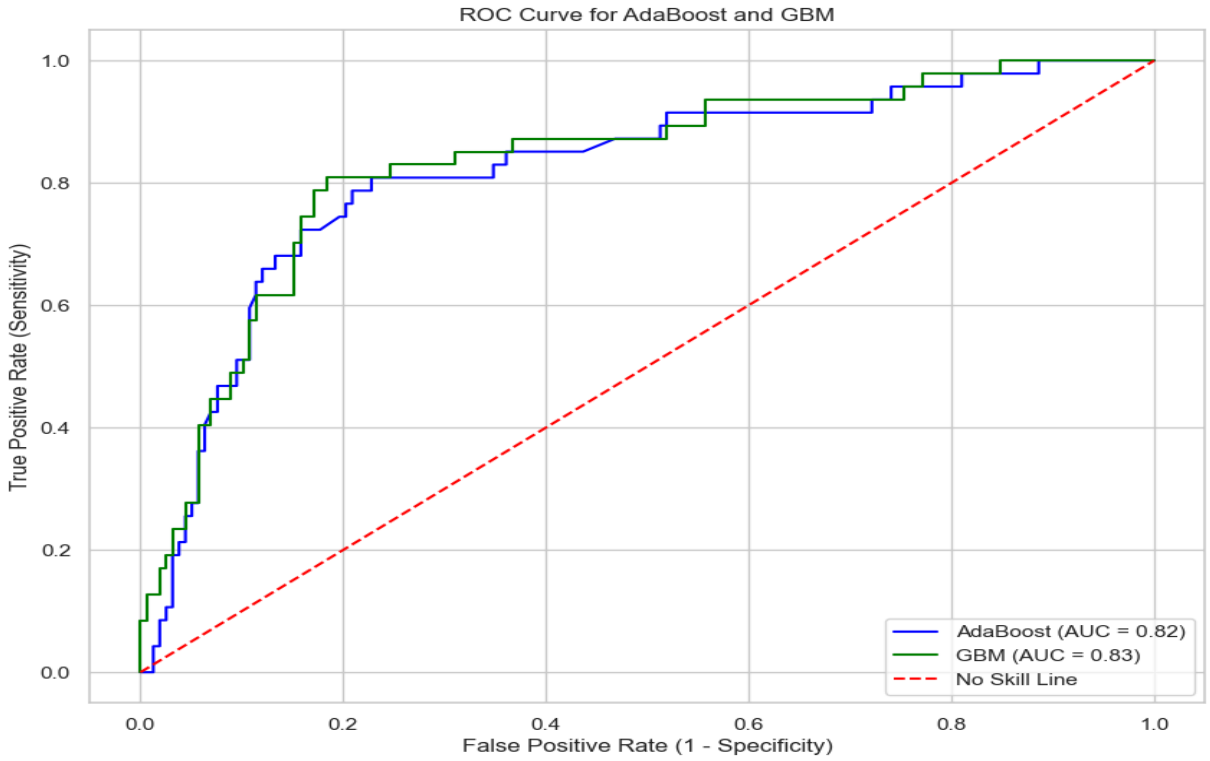

Fig. 7 LightGBM AUROC curve result.

This figure illustrates the ROC curve for the ML model algorithms, AdaBoost and GBM. CatBoost and NN curve result depicts the ROC curve for the ML model algorithms, we can see that the CatBoost and the Neural Network clearly outperforms the baseline model by getting the AUROC curve as 0.84 and 0.83 respectively. These show that both models can predict mortality outcomes with Catboost doing slightly better with balancing sensitivity and specificity across the threshold values.

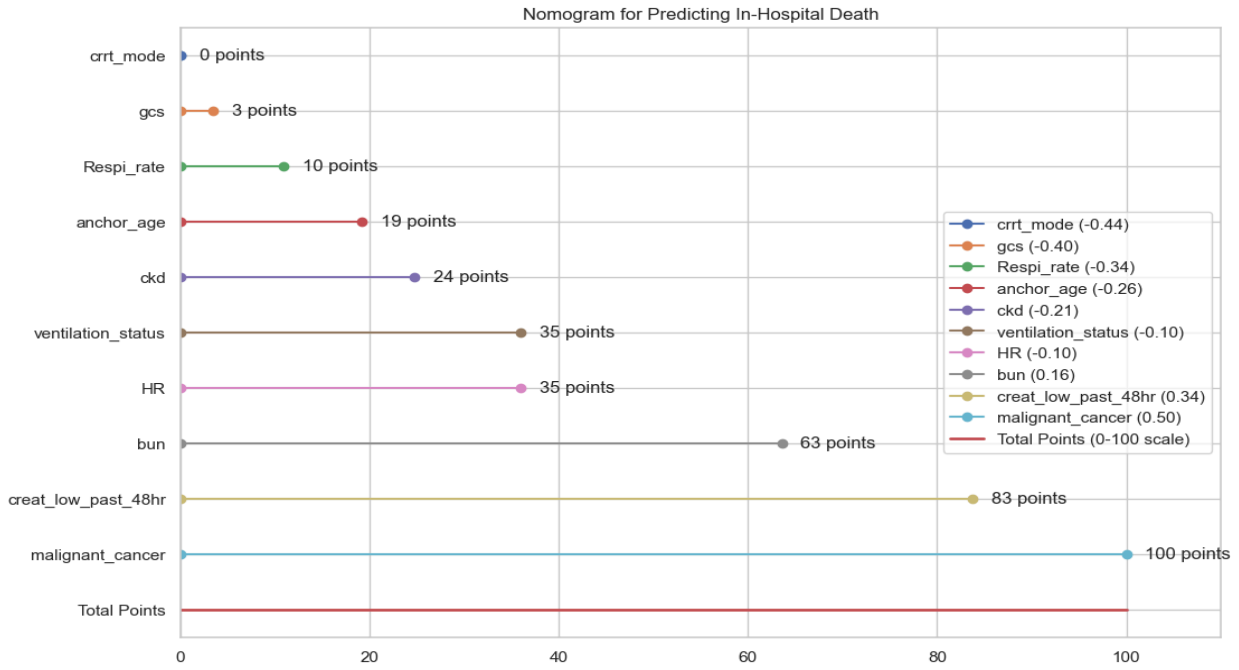

Figure 8 illustrates Enhanced Nomogram for Random Forest . The relative significance of many characteristics in predicting in-hospital mortality is demonstrated by the Random Forest nomogram. This figure shows that the best indicators of death for SA-AKI patients are renal function markers and cancer status.

The Figure 9 shows both positive and negative correlations with mortality outcomes are highlighted by the logistic regression nomogram, which shows normalized coefficient contributions across all features.

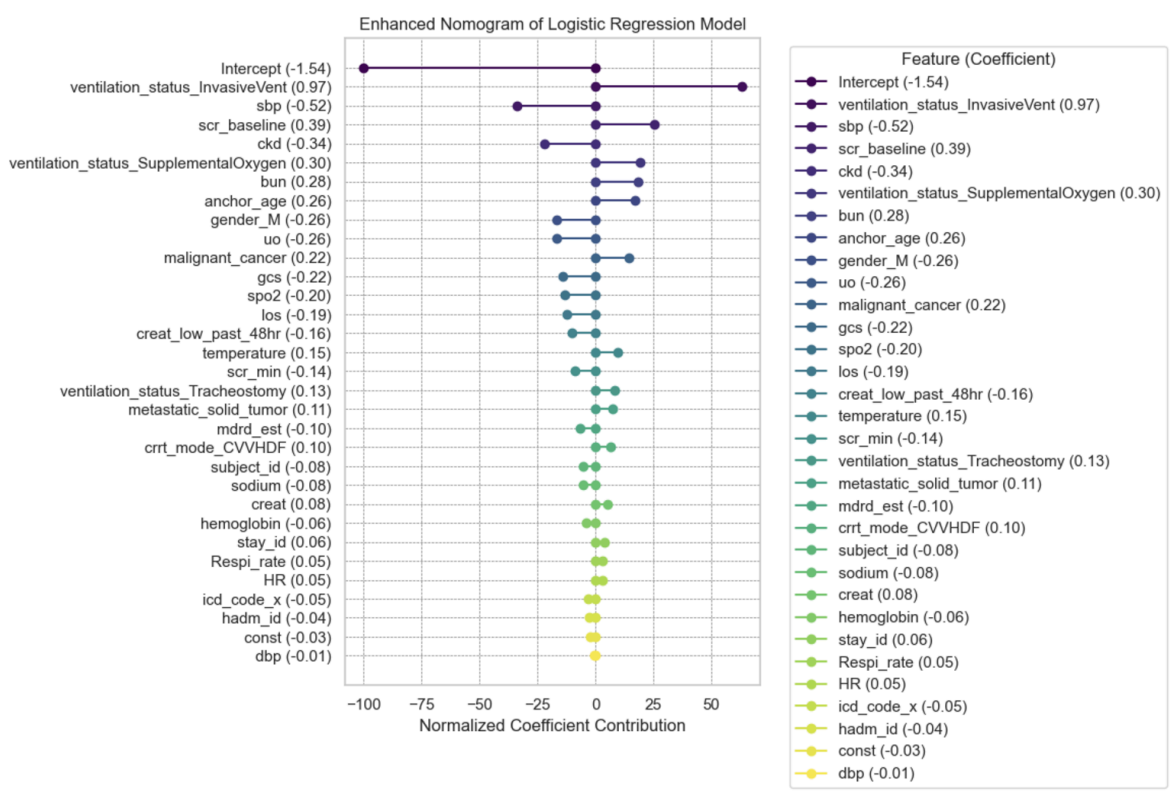

Figure 10 illustrates Enhanced Nomogram for the Logistic Regression model

These illustrations show how our models were able to preserve interpretability while effectively capturing intricate relationships in the data. The ROC curves confirm the predictive efficacy of the models, and the nomograms offer therapeutically meaningful insights into risk factors.

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, Random forest achieved the best AUC score while in the baseline model of the study, XGBoost performed the 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. We selected a threshold of 0.01 for feature selection, which efficiently eliminated features with low variance, in contrast to prior research that used LASSO. In order to ascertain which factors had the most effect on mortality prediction, the feature importance was assessed using correlation analysis and Random Forest's built-in important metrics..

Six ML approaches were created in this study: logistic regression, decision trees, random forests (RF), support vector machines (SVM), k-nearest neighbors (KNN), extreme gradient boosting (XGBoost), and random forests (RF). The mortality prediction models were developed and evaluated using them. We used the median and interquartile range (IQR) to represent the normal distribution of continuous data. On the other hand, numbers and percentages were used to describe categorical variables. Because the retrospective modeling study was conducted at a single site using the MIMIC IV database, the baseline model's ability to infer cause and effect was further limited.

The literature review plotted calibration curves and conducted decision curve analysis (DCA) following the comparison of the AUC score. The expected behavior of the model was fitted using the Local Interpretable Model-Agnostic Explanations (LIME) approaches. Finally, a sensitivity analysis of the findings was carried out. The best model for the dataset, according to the AUC score, is Random Forest. In conclusion, integrating additional machine learning models, there were no appreciable variations in the baseline attributes between the training and validation sets that are more important. The current literature has limitations, such as depending only on internal model validation and using data from a limited sample size, which means the findings do not accurately reflect the general population. Because the current literature used the imputation approach to calculate specific missing data, which may have resulted in a departure from the real value, the new method performed better than the existing literature. Furthermore, in addition to including the same metrics from the primary literature review, our suggested model also developed more sophisticated machine learning models, including CatBoost, LightGBM, Gradient Boosting, and AdaBoost.

Limitations

The study's limitations include feature selection, dimensionality, data imbalance, and data quality. ML models depend on the completeness and quality of the input data for data quality. The model's performance and generalization may be impacted in this particular instance by medical data that contains errors, missing fields, or inconsistencies due to the failure to consider multiple center validation. Class imbalance affects mortality predictions, which can result in biased models for the dominant class and subpar forecast results for the minority class. With so many clinical criteria and laboratory findings available, feature selection and dimensionality occur from choosing particular attributes from a large number of variables. Furthermore, the correlations among these attributes could result in overfitting and heightened computational complexity. Moreover, automated feature selection algorithms and feature engineering approaches may be used in future developments. Every enhancement has the potential to uncover pertinent attributes from extensive healthcare databases. With the help of the aforementioned techniques, our model aims to increase precision by making the data easier to understand and addressing overfitting issues.

Conclusion

Using the MIMIC-IV database, we effectively designed and validated a nomogram to predict in-hospital mortality among older patients with sepsis-associated acute kidney damage (SA-AKI). Important clinical factors such as BUN, GCS score, AKI stage, and mechanical ventilation status were found through the use of expert machine learning approaches such as Random Forest and Logistic Regression models. The Random Forest model with the best AUC of 0.8776 's accuracy and dependability were proved by our findings, which showed outstanding performance.

By using sophisticated preprocessing techniques including imputation and outlier treatment, as well as expert clinical opinions in feature selection, the model's interpretability and practical usefulness were guaranteed. The developed nomogram gives medical professionals a simple way to evaluate patient risk, enabling them to treat high-risk patients quickly. In addition to outperforming previous research in terms of projected accuracy, our method showed how adding fewer but clinically meaningful characteristics improved model stability and interpretability.

Clinically speaking, this study demonstrates how early diagnosis and tailored treatment regimens made possible by data-driven solutions may enhance patient outcomes in critical care settings. The future research should concentrate on extending findings using external validation datasets and investigating advanced feature engineering approaches in order to improve model performance. If such prediction algorithms are incorporated into standard clinical practices, treatment decisions for critically unwell SA-AKI patients may be altered.

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
SAPS: Simplified Acute Physiology Score
MICE: Multiple Imputation by Chained Equations
MIMIC-IV: Medical Information Mart for Intensive Care IV
DCA: Decision Curve Analysis
GBM: Gradient Boosting Machine
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 Electrical and Computer Engineering, University of Southern California, Hughes Aircraft Electrical Engineering Center, 3740 McClintock Ave Suite 100, Los Angeles, CA 90089
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
[15] 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/
[16] 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
[17] Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, et al. LightGBM: A highly efficient gradient-boosting decision tree. Adv Neural Inf Process Syst. 2017;30:3146–54.
[18] Natekin A, Knoll A. Gradient boosting machines, a tutorial. Front Neurorobot. 2013;7:21.
[19] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; San Francisco: 2016. p. 785-94.
[20] MIMIC-IV (Medical Information Mart for Intensive Care, Version 4.0). Laboratory for Computational Physiology, Massachusetts Institute of Technology; 2020. https://mimic.mit.edu/. Accessed 22 Feb 2024.
[21] Goh KH, Wang L, Yeow AY, Poh H, Li K, Yeow JJ, Tan GY. Artificial intelligence in sepsis early prediction and diagnosis using unstructured data in healthcare. Nat Commun. 2021;12(1):711.
[22] Zhu R, Lu D, Xu Y, E W, Cao J, Zuo Y, et al. Deep learning-based prediction of in-hospital mortality for sepsis. Sci Rep. 2020;10:12345
[23]Pishgar M, Karim F, Majumdar S, Darabi H. Pathological Voice Classification Using Mel-Cepstrum Vectors and Support Vector Machines. In: 2018 IEEE International Conference on Big Data (Big Data). 2018. pp. 5267–71. https://doi.org/10.1109/BigData.2018.8622208.
[24]National Institute of General Medical Sciences. Sepsis [Internet]. U.S. Department of Health https://www.nigms.nih.gov/education/fact-sheets/Pages/sepsis.aspx.
[25]Evans T. Diagnosis and management of sepsis. Clin Med. 2018;18(2):146.

[26]Jarczak D, Kluge S, Nierhaus A. Sepsis-pathophysiology and therapeutic concepts. Front Med (Lausanne). 2021;8:640675.
[27]Lever A, Mackenzie I. Sepsis: definition, epidemiology, and diagnosis. BMJ. 2007;335(7625):879–83.
[28]Gyawali B, Ramakrishna K, Dhamoon AS. Sepsis: The evolution in definition, pathophysiology, and management. SAGE Open Med. 2019;7:2050312119835043.
[29]Knoop V, Süveges D, Sveen U, Johnsen L, Vikse B, Rizzi M. Survival prediction of patients with sepsis from age, sex, and septic episode number alone. Sci Rep. 2020;10(1):16725.
[30]Hotchkiss R, Moldawer L, Opal S, et al. Sepsis and septic shock. Nat Rev Dis Primers. 2016;2:16045.