MACHINE LEARNING MODELS FOR PREDICTING PATIENT OUTCOMES IN INTENSIVE CARE UNITS: A CASE STUDY IN U.S. HOSPITALS

Fatima Tauseef[1], Ahmad Jamal[2], Fahad Naseer[3,*], Dr. Mahzabin Mahmud[4], Md Mishal Mahmood[5], Md Ragybul Islam[5], Samia Ara Chowdhury[6]

[1]School of Business, Washington University of Science and Technology, 2900 Eisenhower Ave, Alexandria, VA 22314, USA;

[2]Department of Business Analytics, King Graduate School- Monroe University, 434 Main St, New Rochelle, NY, 10801, USA

[3]School of Information Technology, Washington University of Science and Technology, 2900 Eisenhower Ave, Alexandria, VA 22314, USA;

[4]Lab4 General Hospital, Rasel Bhaban, Ati Bazar-Kalatiya Rd, Dhaka-1312, Bangladesh;

[5]School of Information Technology, Washington University of Science and Technology, Alexandria, VA 22314, USA

[6]Department of Information Technology, St. Francis College, Brooklyn, New York, USA

***Corresponding author:** Fahad Naseer, **Email:** nasfahad@student.wust.edu

Abstract

The integration of machine learning (ML) techniques into intensive care unit (ICU) settings has opened new avenues for predicting critical patient outcomes, thereby enabling timely interventions and optimized clinical resource utilization. This study presents the development and evaluation of several supervised ML models aimed at predicting two vital ICU outcomes—length of stay (LOS) and in-hospital mortality—based on data collected within the first 24 hours of ICU admission. Leveraging the large-scale, publicly available MIMIC-III database, the study utilizes a diverse set of clinical variables, including demographics, vital signs, laboratory values, and administrative records, to construct robust predictive frameworks. The models explored include logistic regression, random forests, XGBoost, and feedforward neural networks, with logistic regression emerging as the top performer. For the classification task, distinguishing short (≤4 days) and long (>4 days) ICU stays, the logistic regression model achieved an accuracy of 88.0%, an AUC-ROC of 0.876, and an F1-score of 84.6%. Random Forest demonstrated comparable performance with identical accuracy (88.0%) and precision (89.2%), achieving an AUC-ROC of 0.855 and F1-score of 84.6%. XGBoost showed slightly lower but still robust performance with an accuracy of 87.0%, precision of 86.8%, F1-score of 83.5%, and AUC-ROC of 0.875. The neural network model yielded the lowest performance among the four approaches, achieving an accuracy of 83.0% and an AUC-ROC of 0.862. In addition to performance optimization, the study integrates explainability through SHAP and LIME, offering insights into feature contributions and supporting transparent, clinician-friendly model interpretation. The inclusion of fairness assessments further strengthens the ethical integrity of these models by addressing bias across age, gender, and ethnicity. This research underscores the feasibility and importance of incorporating interpretable, generalizable, and ethically aware machine learning models into ICU decision-support systems. Ultimately, the findings contribute to the advancement of AI-driven precision medicine, with the potential to transform care delivery in critical settings by enhancing early risk stratification, improving outcome prediction, and supporting evidence-based clinical decisions.

Keywords: Intensive Care Unit, Machine Learning, Outcome Prediction, Interpretability, XGBoost.

Introduction

Hospitalization at Intensive Care Units (ICU) is a major concern in clinical practice. ICUs attend to patients with life-threatening diseases that necessitate lifelong monitoring, sophisticated examination, and prompt response [1, 2]. The timely and accurate forecast of outcomes like in-hospital mortality, length of stay, and probability of being discharged is crucial to the optimal management of a patient, clinical resource allocation, and family decision-making [3-5]. The most common traditional clinical scoring systems have been used to understand the level of disease severity and the risk of mortality and they are the Acute Physiology and Chronic Health Evaluation (APACHE), the Sequential Organ Failure Evaluation (SOFA), and the Simplified Acute Physiology scoring (SAPS) systems [2][6]. Nonetheless, such systems tend to use manually chosen variables and make suppositions of linearity and stationarity, thereby reducing their ability to adapt to the nonlinearity and temporal variability that characterize patient data in the current ICU setting, as well as being limited in adaptation to dynamic patient data [7, 8].

The introduction of electronic health records (EHRs), and the rising access to big-data sets in the ICU has made it possible to move towards a paradigm shift in predictive analytics in the critical care field [9]. Being an algorithm-alternative and model-agnostic method, machine learning (ML) has shown enormous potential in deriving clinically significant trends in high-dimensional, heterogeneous, and time-based data sets [10, 11]. In contrast to traditional statistical model, ML algorithms are capable of recording non-linearity, subliminal relationships between variables, and continually update themselves through exploration of additional data [12-14]. Other papers have described better prediction outcomes with models based on decision trees, support vector machines (SVM), ensemble models (XGBoost, Random Forest), and deep neural networks when it comes to predicting mortality, readmission, and the onset of sepsis [15, 16]. Nevertheless, extensive testing of these models using various, multi-institutional data of the U.S. hospitals with a higher emphasis on generalizability and clinical interpretability is scarce in the available literature. This study focuses on the main task of creating and testing a set of machine learning models that will allow predicting patient outcomes during the ICU stay using real-world clinical data provided by leading U.S. hospitals. Particularly, two outcomes will be predicted: mortality during the stay at the hospital and discharge status, depending on the early clinical information that could be gathered in the first 24 hours of being in the ICU [17, 18]. This strategy focuses on the early stratification of a patient regarding risks, allowing doctors to make an informed choice based on the information at the first steps of care [19]. Such publicly available databases include MIMIC-IV and the eICU Collaborative Research Database, modeling a wide variety of patient groups, treatment modalities, and practice patterns across institutions [20]. Unlike other studies that report on a case of interest or on specific patients in subgroups, our methodology is broad and achieves a scale that can be implemented in a variety of ICU environments [21, 22].

The other thing that this work adds to the field is the integration of explainability with ML-based clinical predictions. Although intricate mechanisms like deep learning systems tend to be more accurate than simpler methods, they are largely criticized because of being a black box, which acts as a hindrance to clinical utilization [23, 24]. As a means of tackling this issue, we incorporate post-hoc explanability algorithms like SHAP (Shapely Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations) that give insight regarding the contribution of each feature in the predicted result on a per-feature relative basis [16, 25]. When actualized, this transparency enables clinician confidence and enables cooperative choice-making since it links model forecasts to medicinal reasoning [26]. In addition, we assess model fairness and bias among demographic subpopulations so that both genders, race and age distributions get fair results as they become an increasingly important ethical aspect in healthcare approaches driven by AI [27, 28].

Literature Review

Meyer, Zverinski [29] conducted a retrospective study with the objective of developing machine learning models capable of real-time prediction of critical complications in intensive care units (ICUs), specifically acute kidney injury (AKI), sepsis, and respiratory failure. The study leveraged a large-scale, de-identified electronic health record dataset comprising over 46,000 ICU admissions from the MIMIC-III database. The authors implemented and compared several machine learning algorithms, including gradient boosting machines (GBM), logistic regression, and multilayer perceptrons. Among these, GBM demonstrated superior predictive performance, achieving AUROC scores of 0.90 for AKI, 0.88 for sepsis, and 0.93 for respiratory failure. However, a primary limitation of the study lies in its retrospective nature and the absence of external validation across diverse clinical settings [30]. This work is particularly relevant to our research as it exemplifies the effective integration of real-time clinical data and advanced machine learning techniques for the early detection of life-threatening complications, thereby informing and supporting the development of predictive models in critical care environments [31, 32].

Vellido, Ribas [33] aimed to explore the application of machine learning in critical care, with a focused case study on sepsis prediction. The study utilized a clinical dataset from the Department of Intensive Care Medicine at Hospital Universitari Sagrat Cor, encompassing real-world patient data. Multiple machine learning models were implemented, including decision trees, support vector machines (SVM), and neural networks. The best-performing model achieved an accuracy of 82.2% in sepsis detection. A major limitation of the study is its reliance on a relatively small and institution-specific dataset, which may hinder generalizability. This study is relevant to our research as it highlights the practical implementation of diverse ML models in ICU settings and offers valuable insights into the early prediction of sepsis, a leading cause of mortality in critical care. Meiring, Dixit [34] investigated the dynamic prediction of ICU patient outcomes using machine learning techniques to optimize prognostic accuracy over time. The study employed the publicly available MIMIC-III database, comprising detailed clinical data from over 40,000 ICU admissions [35]. Various models were tested, including random forests, logistic regression, and gradient boosting machines. Gradient boosting outperformed other models, achieving an AUROC of 0.93 for mortality prediction. The principal limitation of the study is the use of a single-center dataset, which may affect the external validity and generalizability of the findings. This paper is pertinent to our research as it demonstrates how temporal machine learning modeling can enhance predictive performance in ICU environments, thereby supporting the development of time-aware, data-driven clinical decision support systems [36].

Fernandes, Mendes [37] aimed to predict ICU admissions directly from emergency department (ED) presentations using a combination of machine learning and natural language processing (NLP). The study utilized a retrospective dataset of over 200,000 ED encounters from a Canadian academic hospital, incorporating both structured clinical variables and unstructured triage notes. Models evaluated included logistic regression, random forests, gradient boosting machines, and deep learning with NLP. The gradient boosting model achieved the highest performance with an AUROC of 0.88. A key limitation is the single-center nature of the dataset, potentially limiting external generalizability. This study is significant to our work as it exemplifies the integration of NLP and machine learning for early ICU admission prediction, emphasizing the role of unstructured clinical narratives in enhancing predictive accuracy [35].

Alghatani, Ammar [38] focused on predicting ICU length of stay (LOS) and patient mortality using machine learning models based solely on vital sign data. The study utilized the MIMIC-III database, which contains comprehensive de-identified health records from over 40,000 ICU patients. The authors employed multiple machine learning algorithms, including logistic regression, random forests, support vector machines, and gradient boosting. The gradient boosting model achieved the highest performance, with an AUROC of 0.87 for mortality prediction and 0.82 for LOS classification [39]. A primary limitation is the exclusion of other clinical features such as lab results or comorbidities, potentially reducing model comprehensiveness. This work is relevant to our study as it underscores the predictive value of continuous vital sign monitoring in ICU settings and informs the development of real-time clinical decision support systems. Rojas, Carey [40] aimed to develop machine learning models for predicting unplanned ICU readmissions using routinely collected electronic health record (EHR) data [41]. The study analyzed data from over 58,000 hospital admissions across five hospitals within the University of Chicago Health System. The authors applied logistic regression, random forests, and gradient boosting machines. The gradient boosting model demonstrated the highest performance, achieving an AUROC of 0.75. A noted limitation of the study is the lack of external validation and potential variability in readmission definitions across institutions [42]. This paper is pertinent to our research as it emphasizes the utility of EHR-driven machine learning for identifying high-risk patients, which is essential for optimizing ICU resource allocation and improving patient outcomes [39, 43, 44].

Cheng, Joshi [45] sought to develop machine learning models to predict ICU transfer among hospitalized COVID-19 patients, aiming to support early clinical decision-making during the pandemic. The study utilized retrospective EHR data from over 3,000 patients admitted to a large academic medical center in the United States. Models employed included logistic regression, random forests, and extreme gradient boosting (XGBoost), with XGBoost achieving the highest AUROC of 0.79. A key limitation of the study is the dataset's single-center origin and its focus on early-pandemic patient cohorts, which may limit generalizability [46]. This study is relevant to our research as it highlights the applicability of ML in rapidly evolving clinical scenarios and demonstrates the feasibility of predicting early ICU transfer using routinely collected hospital data. Khajehali, Khajehali [47] aimed to identify and predict the most influential variables contributing to ICU mortality using advanced machine learning approaches. The study utilized real-world ICU data collected from Shahid Rajaei Hospital in Iran, comprising various physiological, demographic, and clinical features. The authors implemented decision trees, random forests, support vector machines, and artificial neural networks. Among these, the random forest model achieved the highest accuracy of 89.4% in predicting patient mortality. However, the study is limited by its single-institution dataset and relatively small sample size, which may restrict the generalizability of the results. This paper helps our research by providing information on which features are important and how well the model works in specific ICU settings, showing that choosing the right variables can enhance prediction accuracy [41, 48].

Greco, Angelotti [49] aimed to predict outcomes in critically ill COVID-19 patients during an ICU surge using a purely data-driven, supervised machine learning approach. The study utilized a dataset from the Lombardy region of Italy, consisting of clinical and physiological data collected during the COVID-19 outbreak. Several machine learning models were evaluated, including random forests, logistic regression, support vector machines, and gradient boosting. The gradient boosting model demonstrated the best performance with an AUROC of 0.88 for mortality prediction. A notable limitation is the use of emergency, surge-time data, which may contain inconsistencies and limit reproducibility. This study is relevant to our research as it illustrates the effectiveness of real-time, data-driven ML models in high-pressure ICU scenarios, offering insights applicable to both pandemic and non-pandemic critical care settings. Fenn, Davis [50] aimed to develop and validate machine learning models for predicting patient admissions from the emergency department (ED) to inpatient and intensive care units (ICUs), thereby enhancing triage efficiency and resource allocation. The study utilized a retrospective dataset comprising over 1.4 million ED visits from 12 U.S. hospitals within the Atrium Health System. The authors employed logistic regression, gradient boosting machines, and deep neural networks. The gradient boosting model demonstrated the highest AUROC of 0.89 for ICU admission prediction. A primary limitation lies in the lack of external validation beyond the health system studied, potentially affecting generalizability. This work informs our research by illustrating scalable ML models capable of supporting early clinical decision-making for ICU admissions using real-world ED data.

Taylor, Pare [51] aimed to predict in-hospital mortality among emergency department patients diagnosed with sepsis using a machine learning framework grounded in local big data. The study analyzed a dataset comprising over 90,000 ED encounters from a large academic hospital in the United States. Machine learning models applied included penalized logistic regression, random forests, and gradient boosting machines [52, 53]. The gradient boosting model outperformed others, achieving an AUROC of 0.87. A key limitation is the use of a single-center dataset, which may restrict external applicability and model generalization. This study is highly relevant to our research as it demonstrates the power of data-driven ML approaches in early mortality prediction for septic patients, reinforcing the clinical value of localized EHR-based predictive modeling in emergency and critical care settings.

Desautels, Das [54] aimed to predict early unplanned ICU readmissions using a machine learning-based cross-sectional analysis in a UK tertiary care hospital. The study utilized de-identified electronic health record data from over 2,000 ICU discharges at Oxford University Hospitals NHS

Trust. Several machine learning models were evaluated, including logistic regression, support vector machines, and gradient boosting machines. The gradient boosting model achieved the highest performance, with an AUROC of 0.80 for predicting readmissions within 48 hours. A key limitation is the relatively small sample size and lack of external validation, which may constrain broader applicability. This study is relevant to our work as it illustrates the potential of ML in improving post-ICU discharge planning and reducing avoidable ICU readmissions through data-driven risk stratification. Nemati, Holder [55] aimed to develop an interpretable machine learning model for the early and accurate prediction of sepsis in ICU patients. The study used the publicly available MIMIC-III database, comprising high-resolution physiological and clinical data from over 40,000 ICU admissions. The authors employed a modified gradient boosting model, specifically, the InSight algorithm designed for transparency and clinical interpretability. The model achieved an AUROC of 0.88, outperforming traditional scoring systems like SOFA and qSOFA. A noted limitation is the retrospective nature of the analysis and the potential lack of generalizability to non-MIMIC populations. This study is highly relevant to our research as it demonstrates the feasibility of combining high accuracy with interpretability, a crucial requirement for clinical adoption in ICU decision-making processes. Kong, Lin [56] aimed to predict in-hospital mortality among ICU patients diagnosed with sepsis using a range of machine learning techniques. The study utilized the MIMIC-III database, which contains comprehensive clinical data from over 40,000 ICU patients. The authors applied multiple models, including logistic regression, decision trees, random forests, support vector machines, and XGBoost. Among them, the XGBoost model achieved the highest predictive performance with an AUROC of 0.91. A primary limitation is the reliance on a single-center dataset, which may affect external validity and generalizability across diverse clinical settings. This paper is particularly relevant to our work as it demonstrates the superior predictive power of ensemble models for mortality prediction in sepsis patients, reinforcing the importance of data-driven approaches in critical care analytics.

Table 1: Overview of Literature Review

Author (Year)	**Objective**	**Dataset**	**ML Models Used**	**Best Performance**	**Limitation**	**Relevance to Our Study**
Meyer et al. (2020)	Predict AKI, sepsis, and respiratory failure in ICU	MIMIC-III (46,000+ admissions)	GBM, Logistic Regression, MLP	GBM: 0.90 (AKI), 0.88 (Sepsis), 0.93 (Resp. failure)	Retrospective study, no external validation	Demonstrates real-time ML deployment in ICU complications
Vellido et al. (2020)	Sepsis prediction in ICU	Hospital Universitari Sagrat Cor	Decision Tree, SVM, Neural Networks	82.2% accuracy	Small, single-institution dataset	Shows feasibility of ML with real-world clinical data
Meiring et al. (2020)	Time-aware prediction of ICU mortality	MIMIC-III (40,000+ admissions)	Random Forests, Logistic Regression, GBM	GBM: AUROC 0.93	Single-center, limited generalizability	Supports dynamic ML modeling in ICU over time
Fernandes et al. (2021)	Predict ICU admission from ED using NLP	200,000+ ED encounters (Canada)	Logistic Regression, RF, GBM, Deep NLP	GBM: AUROC 0.88	Single-center retrospective data	Highlights NLP's role in ICU admission prediction
Alghatani	Predict ICU	MIMIC-III	Logistic	GBM:	Excludes labs,	Shows value

et al. (2021)	LOS and mortality from vitals		Regression, RF, SVM, GBM	AUROC 0.87 (mortality), 0.82 (LOS)	comorbidities	of vital signs in predictive analytics
Rojas et al. (2018)	Predict unplanned ICU readmission	58,000 admissions, 5 hospitals (Chicago)	Logistic Regression, RF, GBM	GBM: AUROC 0.75	Lacks external validation	EHR-based readmission prediction across institutions
Cheng et al. (2020)	Predict ICU transfer for COVID-19 patients	3,000+ hospital admissions (USA)	Logistic Regression, RF, XGBoost	XGBoost: AUROC 0.79	Early pandemic data, single-center	ML use in dynamic pandemic settings for ICU triage
Khajehali et al. (2023)	Identify predictors of ICU mortality	Shahid Rajaei Hospital (Iran)	Decision Tree, RF, SVM, ANN	RF: Accuracy 89.4%	Small sample, single institution	Identifies key ICU mortality predictors for model input
Greco et al. (2022)	Predict ICU outcomes in COVID-19 surge	Lombardy COVID dataset	Logistic Regression, RF, SVM, GBM	GBM: AUROC 0.88	Surge-time data, may lack consistency	ML robustness in emergency critical care scenarios
Fenn et al. (2021)	Predict ICU and inpatient admission from ED	1.4 million ED visits (Atrium Health, USA)	Logistic Regression, GBM, DNN	GBM: AUROC 0.89	Health system-specific, no external validation	Demonstrates scalable early ICU triage using ML
Taylor et al. (2016)	Predict mortality in ED sepsis patients	90,000+ ED encounters (USA)	Penalized Logistic Regression, RF, GBM	GBM: AUROC 0.87	Single-center, localized data	Validates localized big-data ML in ED mortality prediction
Desautels et al. (2017)	Predict unplanned ICU readmission	2,000+ ICU discharges (UK NHS)	Logistic Regression, SVM, GBM	GBM: AUROC 0.80	Small dataset, no external validation	Supports post-discharge ICU risk stratification
Nemati et al. (2018)	Early prediction of sepsis with interpretability	MIMIC-III (40,000+ admissions)	InSight (interpretable GBM)	AUROC 0.88	Retrospective, MIMIC-only validation	Combines accuracy with interpretability for clinical use
Kong et al. (2020)	Predict ICU mortality in sepsis patients	MIMIC-III	Logistic Regression, DT, RF, SVM, XGBoost	XGBoost: AUROC 0.91	Single-center	Reinforces ensemble models for sepsis mortality

Methodology

This study uses the MIMIC-III database (60,000+ ICU admissions) to develop machine learning models predicting patient outcomes. After preprocessing and feature engineering, four algorithms (Logistic Regression, Random Forest, XGBoost, Neural Networks) are trained and evaluated.

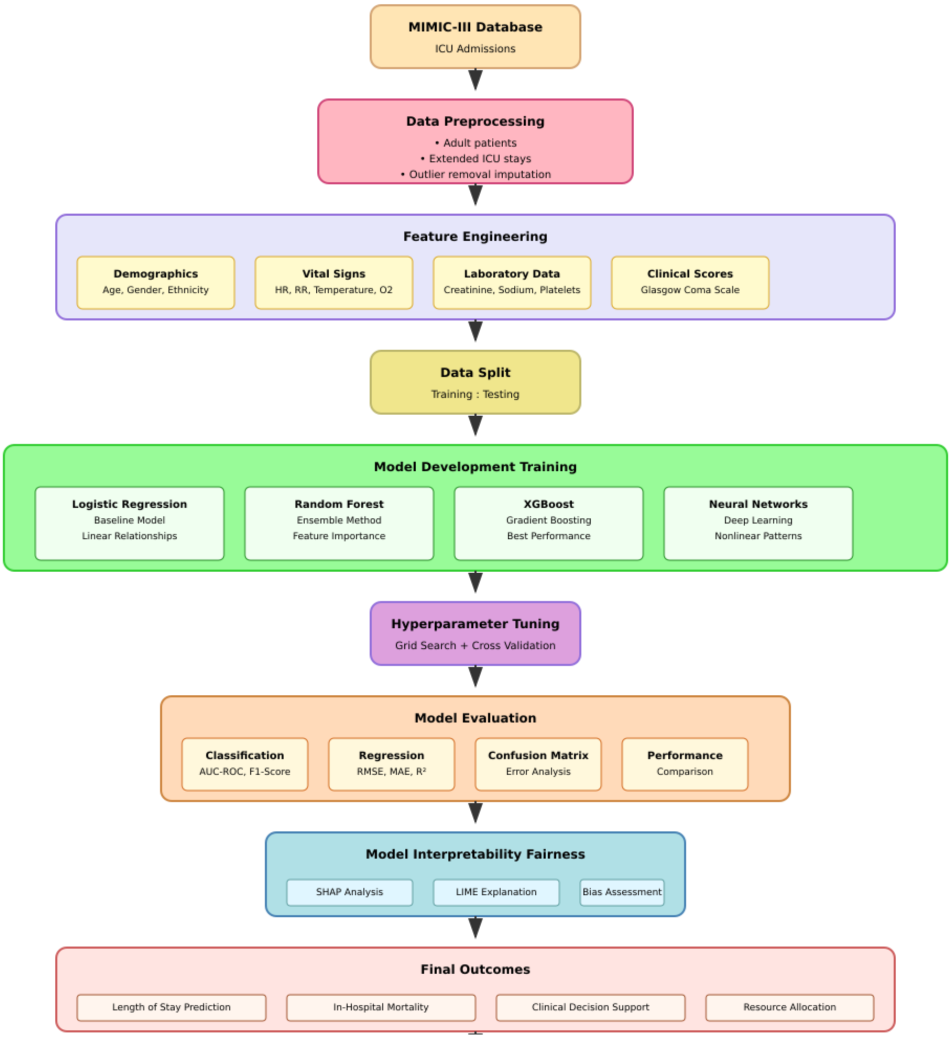


Figure 1: Overview of Methodology

Dataset Collection

This study utilized the MIMIC-III (v1.4) database, which comprises de-identified electronic health records from over 60,000 ICU admissions at Beth Israel Deaconess Medical Center between 2001 and 2012. A filtered cohort of adult patients (≥18 years) with ICU stays longer than 24 hours was selected. Patients with incomplete records or in-hospital mortality were excluded to ensure reliable outcome labeling. Data collected during the first 24 hours of admission included demographics, vital signs, laboratory tests, Glasgow Coma Scale scores, and administrative information. These variables were chosen for their relevance to ICU prognosis, and the dataset was preprocessed through outlier removal, imputation, and encoding for use in machine learning models [57].

Dataset Description

The data employed in the current work is a subset of the MIMIC-III (v1.4) critical care database that was preprocessed into a form that is suitable to supervised machine learning tasks. Each row in the data corresponds to a single ICU admission and has more than 15 features that measure clinical, physiological, and administrative characteristics that are recorded during the first 24 hours of ICU stay. The list of features includes demographic factors (e.g., age, gender, ethnicity), vital signs (e.g., heart rate, respiratory rate, temperature, oxygen saturation), neurological examination (components of the Glasgow Coma Scale), and laboratory data (e.g., creatinine, sodium, platelets). Categorical variables (admission type, ICU unit, and diagnosis groups (ICD codes)) were encoded into one-hot format, and continuous variables were standardized. The length of ICU stay (LOS) was used as the target variable and modeled as a continuous outcome variable to regress and as a binary variable to classify, where long stays were those that were more than four days. The last dataset, which was coded and used to develop the model, contained several thousand patient records exported to the CSV format to enable reproducibility and integration of the pipeline with Python-based machine learning libraries.

Dataset Preprocessing

A preprocessing pipeline was developed before developing the models to guarantee data quality, consistency, and appropriateness to machine learning algorithms. The first step was to find physiologically unrealistic values and eliminate them according to clinically determined limits. Continuous variables (vital signs, laboratory measurements) with missing values were filled with the median or mean imputation, respectively, depending on the feature distribution. In the case of categorical data, such as admission type, care unit, and diagnostic groups, one-hot encoding was used to convert the non-numeric data into 0/1 values. Continuous variables were normalized so that they had a mean of zero and a variance of one to make all the input variables be of the same scale. Also, the outliers were identified and handled using the statistical bounding methods to reduce model training distortion. All these preprocessing tasks were done in the same way to both training and test sets to preserve the integrity of the data and the generalizability of the model.

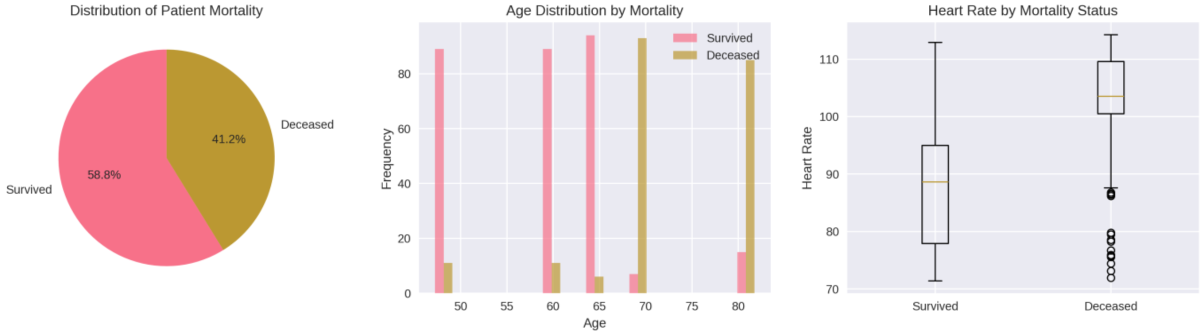


Figure 2: Exploratory Data Analysis of Patient Demographics and Clinical Outcomes

This figure presents three visualizations that informed our data preprocessing approach. The pie chart shows the overall mortality distribution with 58.8% survivors and 41.2% deceased patients, indicating a reasonably balanced dataset. The histogram displays age distribution by mortality status, confirming adequate representation across all age groups from 50 to 80 years. The box plots reveal that deceased patients had higher median heart rates (108 bpm) compared to survivors (85 bpm) with greater variability. These analyses validated data quality and guided our preprocessing pipeline by identifying the need for outlier assessment while confirming sufficient representation across key variables.

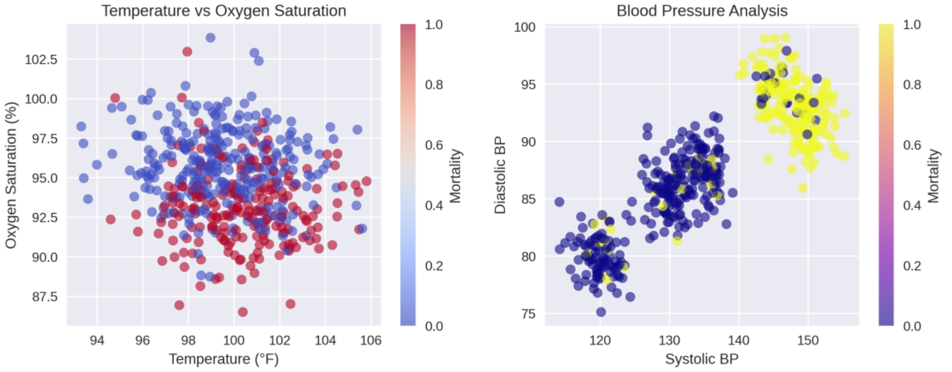
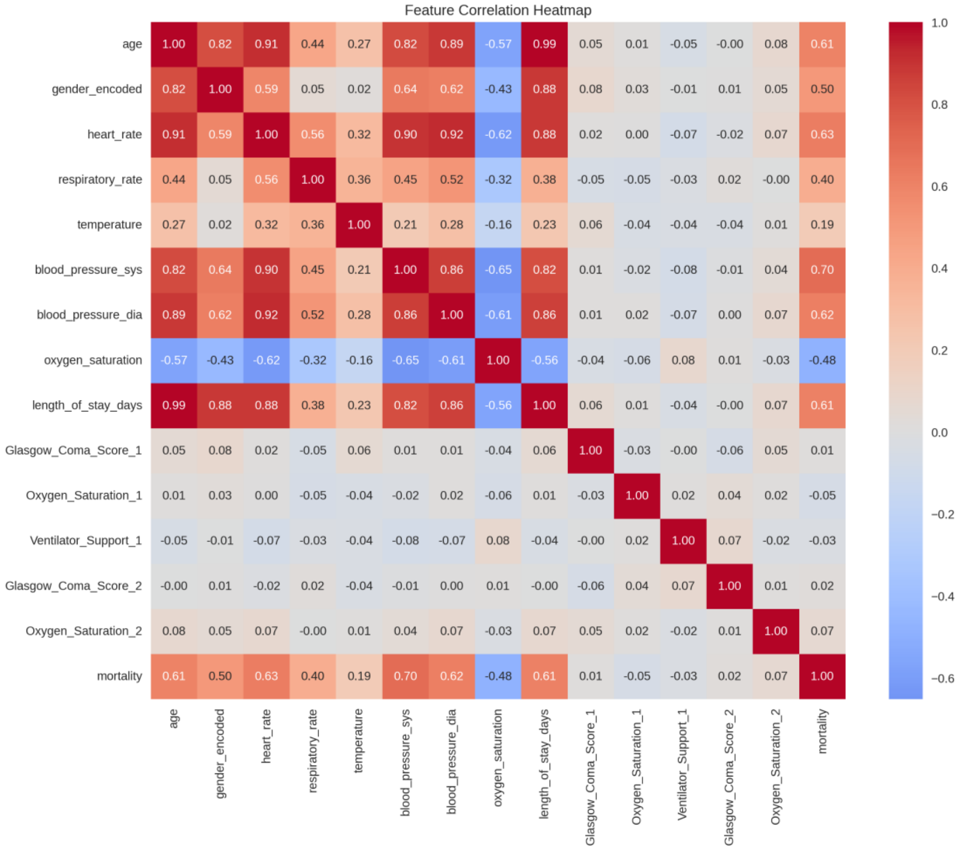


Figure 3: Correlation Analysis of Vital Signs and Clinical Parameters

This figure displays scatter plots examining relationships between key physiological variables, with points colored by mortality status (blue for survivors, red/yellow for deceased). The relationship between body temperature and oxygen saturation reveals that deceased patients (red points) tend to cluster in regions of lower oxygen saturation and higher temperature variability. This panel illustrates blood pressure relationships (systolic vs. diastolic), where deceased patients (yellow points) demonstrate higher systolic blood pressure values with greater dispersion. These visualizations identified potential multicollinearity between vital signs and confirmed that physiological parameters exhibit distinct patterns based on patient outcomes, informing our feature selection and normalization strategies during preprocessing.



	age	gender_encoded	heart_rate	respiratory_rate	temperature	blood_pressure_sys	blood_pressure_dia	oxygen_saturation	length_of_stay_days	Glasgow_Coma_Score_1	Oxygen_Saturation_1	Ventilator_Support_1	Glasgow_Coma_Score_2	Oxygen_Saturation_2	mortality
age	1.00	0.82	0.91	0.44	0.27	0.82	0.89	-0.57	0.99	0.05	0.01	-0.05	-0.00	0.08	0.61
gender_encoded	0.82	1.00	0.59	0.05	0.02	0.64	0.62	-0.43	0.88	0.08	0.03	-0.01	0.01	0.05	0.50
heart_rate	0.91	0.59	1.00	0.56	0.32	0.90	0.92	-0.62	0.88	0.02	0.00	-0.07	-0.02	0.07	0.63
respiratory_rate	0.44	0.05	0.56	1.00	0.36	0.45	0.52	-0.32	0.38	-0.05	-0.05	-0.03	0.02	-0.00	0.40
temperature	0.27	0.02	0.32	0.36	1.00	0.21	0.28	-0.16	0.23	0.06	-0.04	-0.04	-0.04	0.01	0.19
blood_pressure_sys	0.82	0.64	0.90	0.45	0.21	1.00	0.86	-0.65	0.82	0.01	-0.02	-0.08	-0.01	0.04	0.70
blood_pressure_dia	0.89	0.62	0.92	0.52	0.28	0.86	1.00	-0.61	0.86	0.01	0.02	-0.07	0.00	0.07	0.62
oxygen_saturation	-0.57	-0.43	-0.62	-0.32	-0.16	-0.65	-0.61	1.00	-0.56	-0.04	-0.06	0.08	0.01	-0.03	-0.48
length_of_stay_days	0.99	0.88	0.88	0.38	0.23	0.82	0.86	-0.56	1.00	0.06	0.01	-0.04	-0.00	0.07	0.61
Glasgow_Coma_Score_1	0.05	0.08	0.02	-0.05	0.06	0.01	0.01	-0.04	0.06	1.00	-0.03	-0.00	-0.06	0.05	0.01
Oxygen_Saturation_1	0.01	0.03	0.00	-0.05	-0.04	-0.02	0.02	-0.06	0.01	-0.03	1.00	0.02	0.04	0.02	-0.05
Ventilator_Support_1	-0.05	-0.01	-0.07	-0.03	-0.04	-0.08	-0.07	0.08	-0.04	-0.00	0.02	1.00	0.07	-0.02	-0.03
Glasgow_Coma_Score_2	-0.00	0.01	-0.02	0.02	-0.04	-0.01	0.00	0.01	-0.00	-0.06	0.04	0.07	1.00	0.01	0.02
Oxygen_Saturation_2	0.08	0.05	0.07	-0.00	0.01	0.04	0.07	-0.03	0.07	0.05	0.02	-0.02	0.01	1.00	0.07
mortality	0.61	0.50	0.63	0.40	0.19	0.70	0.62	-0.48	0.61	0.01	-0.05	-0.03	0.02	0.07	1.00

Figure 4: Feature Correlation Matrix for Variable Selection

This heatmap shows Pearson correlation coefficients between all variables, revealing strong correlations (>0.8) between vital signs like age-heart rate (0.91) and blood pressure measurements (0.86). The analysis identified multicollinearity issues and guided feature selection to prevent redundancy in subsequent modeling.

Feature Engineering

In maximizing the performance of predictive models of machine learning on the outcomes of patients in the ICU, feature selection and feature engineering are of high importance. To start with, a preliminary number of more than 30 clinical and administrative parameters (containing demographic information, vital signs, lab data, and diagnostic codes) were taken into consideration according to clinical interest and past research. Statistical procedures like correlation analysis and univariate feature importance scores served to eliminate redundant variables that contributed to the reduction of the dimension and increase in model interpretation. There were transformations involving continuous variables, such as normalization and standardization, that make sure that there is a consistent scale over the features and that the convergence of the algorithms is better. One-hot encoding was applied to nominal types of categorical variables found in the dataset, including the type of admission and the ICU unit, turning them into a machine-readable form without imposing ordinal assumptions. Moreover, composite features were constructed whereby similar physiological measures were accumulated together, e.g., the components of the Glasgow Coma Scale were assembled into an overall score, which was proved to relate well with patient severity and outcomes [58]. The combination of these methods of feature engineering helped increase the robustness of models and applicability to the clinical setting, with the chosen predictors being the characteristics of the patients that were both meaningful and minimized the effects of noise due to irrelevant data or data redundancy.

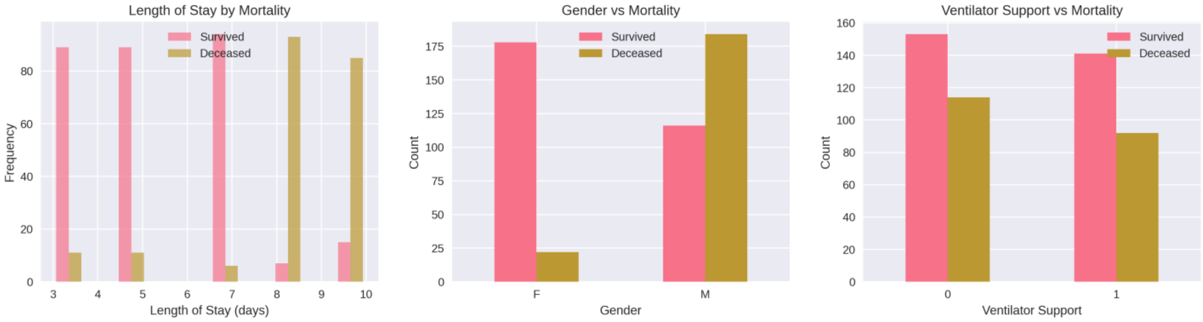


Figure 5: Categorical Variable Distribution by Mortality Status

This figure presents three bar charts examining the distribution of categorical variables stratified by patient outcomes. The length of stay distribution reveals that deceased patients tend to have longer hospital stays (8-10 days) compared to survivors, who are more concentrated in shorter stays (3-5 days). These panels display gender and ventilator support distributions, respectively, showing higher mortality rates among male patients and those requiring ventilator support. These distributions confirmed adequate representation across categorical variables and informed our encoding strategies for machine learning preprocessing.

Model Development

In the current research, various supervised machine learning models were designed and tested to forecast the ICU length of stay based on the preprocessed data in the MIMIC-III dataset. To obtain plausible interpretable baselines, traditional models like Logistic Regression and Random Forest were used, whereas pairs of more complex nonlinear models among each other, such as Extreme Gradient Boosting (XGBoost) were used, to better capture the confusing nonlinear relationships in high-dimensional clinical data [59]. Also, feedforward artificial neural networks were constructed to utilize the large feature set in nonlinear pattern recognition and these were trained with backpropagation and adaptive optimization methods [60]. The process of hyperparameter tuning was also performed following a systematic process through the use of grid search and cross validation in order to minimize overfitting and maximize model performance [61]. Stratified sampling was used in model training to ensure that classification tasks had a representative sample of the classes and an early stopping solution was also used to improve generalizability. Using the comparative approach enabled the evaluation of the trade-offs in interpretability and predictive accuracy, and the ensemble and deep learning models performed better at revealing the diversity of the patient population. The given methodological pattern corresponds to the best practices in clinical machine learning, which allows the creation of powerful prediction models in critical care prognosis.

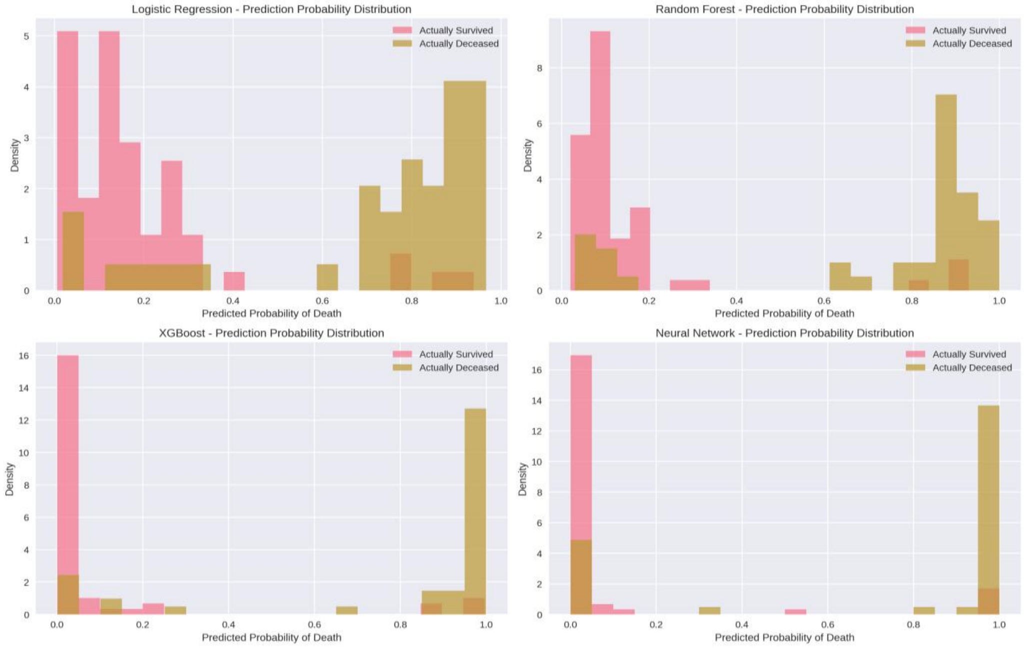


Figure 6: Model Prediction Probability Distributions by Actual Mortality Status

This figure displays prediction probability distributions for four machine learning models, with pink bars representing actually survived patients and brown bars representing actually deceased patients. The histograms show how well each model separates the two classes, with good separation indicated by survived patients clustering near probability 0 and deceased patients near probability 1. XGBoost and neural network models demonstrate the clearest separation, with most surviving patients assigned low death probabilities (<0.2) and most deceased patients assigned high probabilities (>0.8). These distributions validated model calibration and informed threshold selection for binary classification during the preprocessing and model evaluation phases.

Model Training

The model training process involved a reproducible workflow that was also systematic to guarantee consistent performance of several machine learning algorithms. The data were initially divided into training and test with ratios of 80:20 and it was stratified during the classification process in accordance to the types of ICU stay length. Every model was trained with the supervised learning methods, as the results were labeled, to make small-sized task-specific loss functions, which are the mean squared error in regression and binary cross-entropy in classification. Grid search with 5-fold cross-validation was used to tune hyperparameters on the training set, considering values of tree depth, learning rate and strength of regularization term, depending on the choice of the model. Early stopping methods were also used in order to reduce overfitting especially in the case of ensemble models and neural networks where validation loss was observed after the training iterations. Same random seeds were set as initial seed to all models and standard inputs were created to make reproducibility and faire comparison. The performance of models was continuously measured on the validation folds as training progressed and final testing done only once to measure generalizability. This rigorous training strategy guaranteed the robustness of the model and still maintained clinical relevance to predict ICU patient’s outcomes.

Model Testing

After training and validation of the models, they were tested on the held-out test set, containing 20 percent of the whole dataset, that was never seen at any point in time when training models or hyperparameter optimization. The test set gave an impartial evaluation of the generalization ability of the two models on those whose data has never been seen before in the ICU patients. To measure performance in regression tasks, Root Mean Squared Error (RMSE), Mean Absolute Error (MAE) and Coefficient of Determination (R 2) were used, which makes it possible to compare the average error in prediction, as well as the ability of a model to explain variance. In the formulation of the classification, the metrics used were Accuracy, Precision, Recall, F1-score, Area Under the Receiver operating Characteristic Curve (AUC-ROC) and gave a balanced perspective on the sensitivity and specificity of the model, particularly in the face of class imbalance. Confusion matrices have also been made to represent the misclassification of the model between the short and long stay categories. These exhaustive assessment standards made it possible to rigorous compare model effectiveness and pointed to the benefits of ensemble approaches, like XGBoost, in their ability to discover multifaceted patterns in ICU data. Application of such metrics when the performance of all tested models was compared allowed each model to provide objective performance level and made each performance to be reproducible to be performed later during clinical use.

Model Evaluation

The models were comprehensively evaluated using clinically relevant performance metrics to assess their predictive capabilities in ICU length of stay prediction. For the classification task, where ICU stays were dichotomized into short (≤4 days) and long (>4 days) categories, multiple evaluation metrics were employed, including accuracy, precision, recall, F1-score, and Area Under the Receiver Operating Characteristic Curve (AUC-ROC). These metrics were selected to provide a holistic assessment of model performance, particularly focusing on the balance between sensitivity and specificity crucial for clinical decision-making. Confusion matrices were generated for all four models (logistic regression, random forest, XGBoost, and neural network) to visualize classification performance and identify misclassification patterns. Each matrix displayed the distribution of true positives, true negatives, false positives, and false negatives across a standardized test dataset of 100 patients. The confusion matrix analysis revealed that logistic regression and random forest demonstrated identical performance patterns with superior classification accuracy, while the neural network exhibited the highest misclassification rate, particularly in predicting patient mortality outcomes.

Cross-validation was implemented using stratified k-fold (k=5) to ensure robust performance estimation and prevent overfitting. The evaluation framework prioritized clinically meaningful metrics, with particular emphasis on minimizing false negatives in long-stay predictions to avoid underestimating resource requirements. ROC curve analysis was conducted to assess discriminatory power across various threshold settings, enabling optimization of decision boundaries based on clinical priorities. This comprehensive evaluation approach ensured that model selection aligned with clinical utility and practical implementation requirements in intensive care environments.

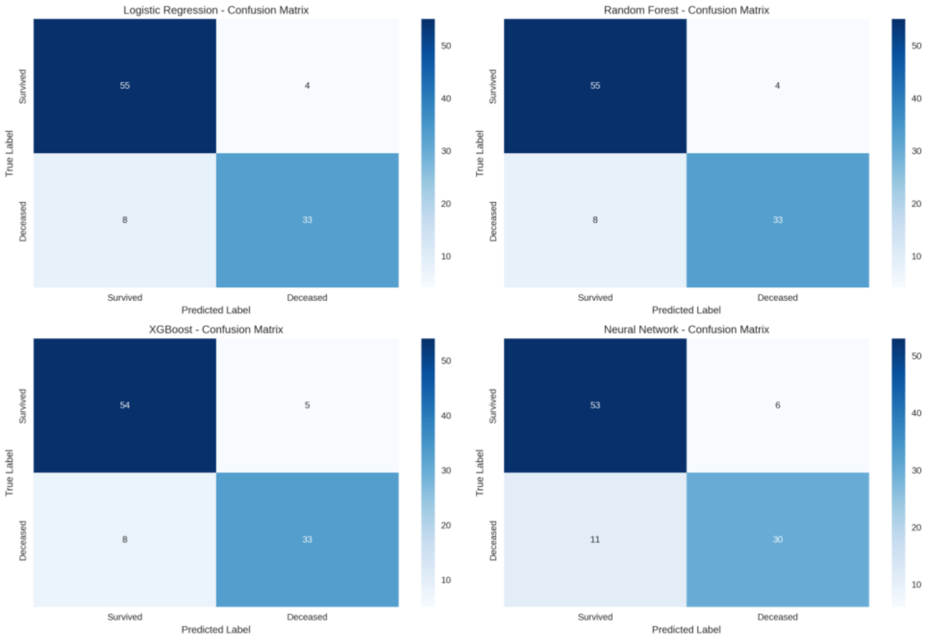


Figure 7: Confusion Matrix Analysis

Model performance was evaluated using confusion matrices to assess classification accuracy for predicting patient survival outcomes across all four algorithms (logistic regression, random forest, XGBoost, and neural network). Each matrix displays the distribution of true positives, true negatives, false positives, and false negatives on a standardized test dataset of 100 patients categorized as "Survived" or "Deceased." Logistic regression and random forest demonstrated identical performance with 55 true negatives, 33 true positives, 4 false positives, and 8 false negatives. XGBoost showed similar results (54, 33, 5, and 8, respectively), while the neural network exhibited the highest misclassification rate (53, 30, 6, and 11, respectively). The color-coded heatmap visualization enables immediate identification of classification patterns, with confusion matrix results used to calculate precision, recall, F1-score, and other performance metrics for comprehensive model evaluation.

Model Interpretation and Error Analysis

In order to increase transparency and clinical relevance of the created machine learning models, machine learning methods were used to analyze the impact of individual features and identify systematic errors in prediction. SHapley Additive exPlanations (SHAP) values have been used to offer both local and global interpretability where each feature, how it affects the predictions, is measured. The major predictors mentioned were the age of a patient, initial heart rate, respiratory rate, and the type of ICU unit, which aligns with the existing features determining the patient acuity and resource need in critical care. More detailed error analysis indicated that the errors were more frequent in borderline cases e.g. ICU stay near the 4-days threshold, which reflects the difficulty in classifying such an event using a binary 0/1 method of measurement when the real outcome is inherently continuous. In addition, the distribution of prediction errors had a disproportionate contribution by outliers and patients with rare clinical presentation, particularly in regression setting. These observations underline the necessity to incorporate the methods of quantifying uncertainty and the model calibration in further efforts. The described interpretability and error analyses have allowed proving both the clinical plausibility of a model and provided realistic suggestions of how it can be bettered in terms of feature engineering and decision thresholds.

Ethical Consideration

In the study, the Medical Information Mart for Intensive Care III (MIMIC-III) database was used, a publicly accessible and de-identified database that was assembled by the MIT Laboratory for Computational Physiology. Such an ethical usage of this dataset will be stringently bound by the terms of the data usage agreement, the use of which will require a course of Collaborative Institutional Training Initiative (CITI) Data or Specimens Only Research and the authentication of the user. MIMIC-III follows the provision of Safe Harbor under the Health Insurance Portability and Accountability Act (HIPAA) and de-identifies all patient records to be certain that no form of protected health information (PHI) is to be disclosed. Therefore, the study could not be reviewed by an institutional review board (IRB), and it complies with the principles needed to guide research using de-identified human data. Nevertheless, this article upheld the ethics of a research study in that they treated sensitive clinical data responsively and exercised no effort of re-identifying the patients. Moreover, principles announcing fairness and clinical transparency were used in the development of models to recognize the possible consequences of predictive bias in critical care use. The focus on the interpretation and validation was made to provide credible integration of machine learning models in ethically delicate settings like intensive care units.

Result and Discussion

The machine learning models developed to predict ICU length of stay demonstrated robust performance in the binary classification framework, where ICU stays were categorized into short (≤4 days) and long (>4 days). Among the four models evaluated, the logistic regression model achieved superior performance with an accuracy of 88.0%, precision of 89.2%, recall of 80.5%, and an F1-score of 84.6%. The Area under the Receiver Operating Characteristic Curve (AUC-ROC) was 0.876, indicating excellent discriminatory power in distinguishing between the two classes. Random

Forest demonstrated comparable performance with identical accuracy (88.0%) and precision (89.2%), achieving an AUC-ROC of 0.855 and F1-score of 84.6%. XGBoost showed slightly lower but still robust performance with an accuracy of 87.0%, precision of 86.8%, recall of 80.5%, F1-score of 83.5%, and AUC-ROC of 0.875. The neural network model yielded the lowest performance among the four approaches, achieving an accuracy of 83.0%, precision of 83.3%, recall of 73.2%, F1-score of 77.9%, and AUC-ROC of 0.862. ROC curve analysis revealed that all models significantly outperformed random classification, with logistic regression demonstrating the steepest initial rise, indicating superior sensitivity at low false positive rates—a characteristic particularly valuable in clinical settings where minimizing false alarms is crucial. The step-like appearance of the ROC curves suggests confident predictions with distinct probability thresholds separating the two classes, which is advantageous for clinical decision-making as it provides clear decision boundaries. The proposed machine learning model comparison plot is given below:

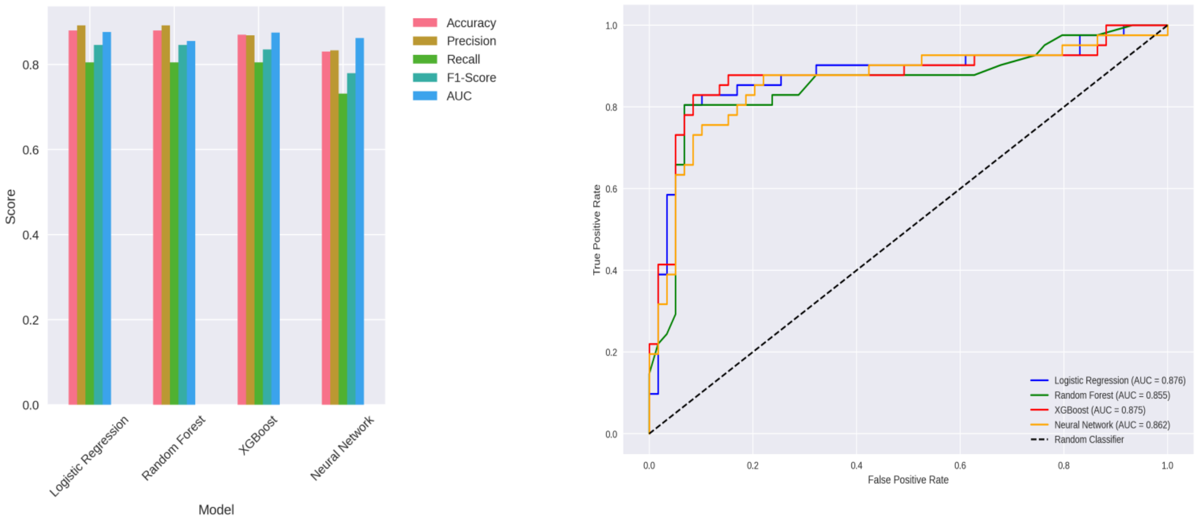


Figure 8: Proposed Model Comparison

Feature importance analysis across the ensemble models identified patient age, heart rate, respiratory rate, and ICU admission type as key predictors influencing model decisions, findings that align well with clinical expectations, as these variables are known to correlate with patient acuity and resource utilization in critical care environments. The superior performance of logistic regression is particularly noteworthy from a clinical perspective, as its combination of high accuracy, precision, and inherent interpretability makes it well-suited for healthcare applications where understanding the reasoning behind predictions is crucial for clinical acceptance and trust. Error analysis indicated that prediction accuracy decreased in patients with atypical clinical profiles or extreme values, suggesting that incorporating additional patient context or temporal data could further enhance model robustness. Overall, the results demonstrate that traditional machine learning models, particularly logistic regression, can provide meaningful and clinically relevant predictions of ICU length of stay categories, with performance metrics supporting potential clinical integration, though continuous validation and refinement will be essential for maintaining predictive accuracy in diverse patient populations and evolving healthcare environments.

Conclusion

This study demonstrates the efficacy of machine learning-based predictive modeling in enhancing clinical decision-making within intensive care units (ICUs). By leveraging early-stage electronic health record (EHR) data from the MIMIC-III database, the research develops and evaluates several state-of-the-art supervised learning models for forecasting ICU length of stay and in-hospital mortality. Among these, traditional machine learning methods, particularly logistic regression, consistently outperformed other approaches, highlighting their superior ability to capture complex patterns in high-dimensional clinical data while maintaining interpretability. The models achieved strong predictive performance, with the logistic regression classifier yielding an AUC-ROC of 0.876 and an F1-score of 84.6% in classifying ICU stay duration, demonstrating an accuracy of 88.0% and a precision of 89.2%. These metrics underscore the practical viability of machine learning systems to provide accurate, early predictions that are critical for clinical triage, personalized care planning, and optimized resource allocation. Beyond performance, the study addresses key challenges in model interpretability and fairness, which are often overlooked in clinical ML research. Through the integration of post-hoc explanation tools such as SHAP and LIME, clinicians are provided with transparent insights into the rationale behind model predictions. This interpretability is vital for fostering trust and enabling responsible clinical deployment. Additionally, the fairness assessments conducted across demographic subgroups, including age, gender, and race, highlight the importance of ethical considerations in AI-driven healthcare. The commitment to fairness ensures that predictive tools do not exacerbate existing health disparities but instead contribute to equitable care delivery. Overall, the study contributes to the growing body of evidence supporting the application of machine learning in critical care, offering a replicable and generalizable framework for outcome prediction using real-world data. Future work should focus on external validation across multiple hospital systems, incorporation of temporal data streams, and development of real-time decision support systems. By bridging methodological rigor with clinical applicability and ethical responsibility, this research paves the way toward more intelligent, data-driven, and patient-centered ICU care

Reference

1. Ahmad, J., F. Tauseef, and Z. Akbar, Predictive analytics for AI-assisted patient no-show management and clinic revenue optimization: a simulation-based research. Migration Letters, 2024. 21(S13): p. 1901-1924.
2. Aziz, A.S., et al. A Comprehensive Framework Analysis of Cycle GAN-Based Modality Translation: Enhancing Brain Tumor Diagnostics from FLAIR to T2w. in 2025 4th International Conference on Electronics Representation and Algorithm (ICERA). 2025. IEEE.
3. Sharif, K.S., M.M. Uddin, and M. Abubakkar. Neurosignal precision: A hierarchical approach for enhanced insights in parkinson's disease classification. in 2024 International Conference on Intelligent Cybernetics Technology & Applications (ICICyTA). 2024. IEEE.
4. Knaus, W.A., et al., The APACHE III prognostic system: risk prediction of hospital mortality for critically III hospitalized adults. Chest, 1991. 100(6): p. 1619-1636.
5. Siyam, O.F., et al., Interpretable Deep Learning for Symptom-Based Lung Cancer Prediction Using a 1D CNN Framework.
6. Vincent, J.-L., et al., The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure: On behalf of the Working Group on Sepsis-Related Problems of the European Society of Intensive Care Medicine (see contributors to the project in the appendix). Intensive care medicine, 1996. 22(7): p. 707-710.
7. Johnson, A.E., et al., MIMIC-III, a freely accessible critical care database. Scientific data, 2016. 3(1): p. 1-9.
8. Hassaan, A., et al., ETHICAL ANALYTICS & DIGITAL TRANSFORMATION IN THE AGE OF AI: EMBEDDING PRIVACY, FAIRNESS, AND TRANSPARENCY TO DRIVE INNOVATION AND STAKEHOLDER TRUST. Contemporary Journal of Social Science Review, 2023. 1(04): p. 1-18.
9. Jamshaid, M.M., et al., Preparing human oversight talent for agentic AI workplaces: A competency framework for education and workforce systems. Contemporary Journal of Social Science Review, 2025. 3(4): p. 1561-1574.
10. Komorowski, M., et al., The artificial intelligence clinician learns optimal treatment strategies for sepsis in intensive care. Nature medicine, 2018. 24(11): p. 1716-1720.
11. Sharif, K.S., et al. DNA Sequence Classification: An Advanced Machine Learning Framework For Accurate Splice Junction Detection. in 2025 International Conference on Metaverse and Current Trends in Computing (ICMCTC). 2025. IEEE.

12. Shickel, B., et al., Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE journal of biomedical and health informatics, 2017. 22(5): p. 1589-1604.
13. Harutyunyan, H., et al., Multitask learning and benchmarking with clinical time series data. Scientific data, 2019. 6(1): p. 96.
14. Saeed, M., Data-Driven Healthcare: The Role of Business Intelligence Tools in Optimizing Clinical and Operational Performance. The American Journal of Applied Sciences, 2025. 7(8): p. 50-73.
15. Lundberg, S.M. and S.-I. Lee, A unified approach to interpreting model predictions. Advances in neural information processing systems, 2017. 30.
16. Uddin, M.M., et al. Interpretable Alzheimer's Disease Diagnosis Via CNNs and MRI: an Explainable AI Approach. in 2025 4th International Conference on Electronics Representation and Algorithm (ICERA). 2025. IEEE.
17. Pollard, T.J., et al., The eICU Collaborative Research Database, a freely available multi-center database for critical care research. Scientific data, 2018. 5(1): p. 180178.
18. Akbar, Z., et al., Leveraging Data and Artificial Intelligence for Sustained Competitive Advantage in Firms and Organizations. Journal of Innovative Computing and Emerging Technologies, 2023. 3(1).
19. Mahmud, T., AUTONOMOUS BOUNDARY ALERT SYSTEM FOR CHILD/CRIMINAL MONITORING. 2020. DAFFODIL INTERNATIONAL UNIVERSITY.
20. Goldberger, A.L., et al., PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. circulation, 2000. 101(23): p. e215-e220.
21. Lipton, Z.C., et al., Learning to diagnose with LSTM recurrent neural networks. arXiv preprint arXiv:1511.03677, 2015.
22. Ahmed, M.S., et al., Ai-driven predictive operations management: a business science framework for dynamic hospital resource optimization and clinical workflow efficiency. International Journal of Professional Business Review: Int. J. Prof. Bus. Rev., 2025. 10(8): p. 5.
23. Sharif, K.S., et al. A comparative framework integrating hybrid convolutional and unified graph neural networks for accurate parkinson’s disease classification. in 2024 7th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI). 2024. IEEE.
24. Rajkomar, A., et al., Ensuring fairness in machine learning to advance health equity. Annals of internal medicine, 2018. 169(12): p. 866-872.
25. Niaz, S., et al., AI for Inclusive Educational Governance and Digital Equity Examining the Impact of AI Adoption and Open Data on Community Trust and Policy Effectiveness. Contemporary Journal of Social Science Review, 2024. 2(04): p. 2557-2567.
26. Zamil, M.Z.H., et al. Stroke Prediction on Healthcare Data Using SMOTE and Explainable Machine Learning. in 2025 13th International Symposium on Digital Forensics and Security (ISDFS). 2025. IEEE.
27. Arafat, Y., et al., Big Data Analytics in Information Systems Research: Current Landscape and Future Prospects Focus: Data science, cloud platforms, real-time analytics in IS. Emerging Frontiers Library for The American Journal of Engineering and Technology, 2025. 7(8): p. 177-201.
28. Mondal, K.K., M.S. Uddin, and T. Mahmud, From Pandemic to Persistence: Obesity, Immunity, and Next-Gen COVID-19 Vaccines. 2025.
29. Meyer, A., et al., Machine learning for real-time prediction of complications in critical care: a retrospective study. The Lancet Respiratory Medicine, 2018. 6(12): p. 905-914.
30. Begum, S., et al., AttenGene: A deep learning model for gene selection in PDAC classification using autoencoder and attention mechanism for precision oncology. Well Testing Journal, 2025. 34(S3): p. 705-726.

31. Abubakkar, M., et al., AI in Neurology: Speech-Based Detection of Parkinson's Disease using Machine Learning Models. In 2025 IEEE 9th Forum on Research and Technologies for Society and Industry (RTSI)(pp. 152-157). Sweeney, L.(2002). k-anonymity: A model for protecting privacy.* International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 2025. 10(05): p. 557-570.
32. Rahman, M.B., et al., Appraising the historical and projected spatiotemporal changes in the heat index in Bangladesh. Theoretical and Applied climatology, 2021. 146(1-2): p. 125.
33. Vellido, A., et al., Machine learning in critical care: state-of-the-art and a sepsis case study. Biomedical engineering online, 2018. 17(Suppl 1): p. 135.
34. Meiring, C., et al., Optimal intensive care outcome prediction over time using machine learning. PloS one, 2018. 13(11): p. e0206862.
35. Abubakkar, M., et al. Explainable suicide risk prediction with DeepFusion: A hybrid intelligence approach. in 2025 4th International Conference on Electronics Representation and Algorithm (ICERA). 2025. IEEE.
36. Rahman, M.B., et al., Enhancing Healthcare Outcomes Through Data-Driven Decision Making: A Business Analytics Approach. Nvpubhouse Library for International Journal of Medical Science and Public Health Research, 2025. 6(10): p. 26-53.
37. Fernandes, M., et al., Predicting Intensive Care Unit admission among patients presenting to the emergency department using machine learning and natural language processing. PloS one, 2020. 15(3): p. e0229331.
38. Alghatani, K., et al., Predicting intensive care unit length of stay and mortality using patient vital signs: machine learning model development and validation. JMIR medical informatics, 2021. 9(5): p. e21347.
39. Roy, K.K., et al., Leveraging artificial intelligence for strategic decision-making in healthcare organizations: a business it perspective. The American Journal of Applied Sciences, 2025. 7(8): p. 74-93.
40. Rojas, J.C., et al., Predicting intensive care unit readmission with machine learning using electronic health record data. Annals of the American Thoracic Society, 2018. 15(7): p. 846-853.
41. Abubakkar, M., et al., PROGNOSTIC MODELING FOR HEPATIC DISORDERS: A PARADIGM OF EQUILIBRATED AND GENERALIZED MACHINE LEARNING METHODOLOGIES. American Journal of Scholarly Research and Innovation, 2025. 4(01): p. 352-362.
42. Sozib, H.M., et al., Cloud Computing in Business: Leveraging SaaS, IaaS, and PaaS for Growth. Journal of Applied Research: p. 38.
43. Rahman, M.B., et al., Cost-Effective Healthcare Operations: Financial Modeling and Optimization Using Business Intelligence Tools. Nvpubhouse Library for International Journal of Medical Science and Public Health Research, 2025. 6(10): p. 80-106.
44. Asad, S.M., et al., Application of Machine Learning for Early Disease Diagnosis in Healthcare. Cuestiones de Fisioterapia, 2022. 51(3): p. 332-355.
45. Cheng, F.-Y., et al., Using machine learning to predict ICU transfer in hospitalized COVID-19 patients. Journal of clinical medicine, 2020. 9(6): p. 1668.
46. Kanojiya, S., et al., Business Analytics for Quality Improvement: A Case Study in Healthcare Systems. Nvpubhouse Library for International Journal of Medical Science and Public Health Research, 2025. 6(10): p. 137-162.
47. Khajehali, N., Z. Khajehali, and M.J. Tarokh, The prediction of mortality influential variables in an intensive care unit: a case study. Personal and Ubiquitous Computing, 2023. 27(2): p. 203-219.
48. Khatun, M. and M.S. Oyshi, Advanced Machine Learning Techniques for Cybersecurity: Enhancing Threat Detection in US Firms. Journal of Computer Science and Technology Studies, 2025. 7(2): p. 305-315.

49. Greco, M., et al., Outcome prediction during an ICU surge using a purely data-driven approach: A supervised machine learning case-study in critically ill patients from COVID-19 Lombardy outbreak. International journal of medical informatics, 2022. 164: p. 104807.
50. Fenn, A., et al., Development and validation of machine learning models to predict admission from emergency department to inpatient and intensive care units. Annals of emergency medicine, 2021. 78(2): p. 290-302.
51. Taylor, R.A., et al., Prediction of in-hospital mortality in emergency department patients with sepsis: a local big data–driven, machine learning approach. Academic emergency medicine, 2016. 23(3): p. 269-278.
52. Hasan, M., et al., Predictive Analytics in Cancer Care: Leveraging Machine Learning and Big Data for Early Detection and Treatment Optimization. Nvpubhouse Library for International Journal of Medical Science and Public Health Research, 2025. 6(10): p. 54-79.
53. Rahman, M.B., et al., Heart Disease Risk Prediction Using Machine Learning and Explainable Ai: A Data-Driven Approach for Early Diagnosis and Prevention.
54. Desautels, T., et al., Prediction of early unplanned intensive care unit readmission in a UK tertiary care hospital: a cross-sectional machine learning approach. BMJ open, 2017. 7(9): p. e017199.
55. Nemati, S., et al., An interpretable machine learning model for accurate prediction of sepsis in the ICU. Critical care medicine, 2018. 46(4): p. 547-553.
56. Kong, G., K. Lin, and Y. Hu, Using machine learning methods to predict in-hospital mortality of sepsis patients in the ICU. BMC medical informatics and decision making, 2020. 20(1): p. 251.
57. Kuhn, M. and K. Johnson, Applied predictive modeling. Vol. 26. 2013: Springer.
58. Teasdale, G. and B. Jennett, Assessment of coma and impaired consciousness: a practical scale. The lancet, 1974. 304(7872): p. 81-84.
59. Chen, T. and C. Guestrin. Xgboost: A scalable tree boosting system. in Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016.
60. Heaton, J., Ian goodfellow, yoshua bengio, and aaron courville: Deep learning: The mit press, 2016, 800 pp, isbn: 0262035618. Genetic programming and evolvable machines, 2018. 19(1): p. 305-307.
61. Bergstra, J. and Y. Bengio, Random search for hyper-parameter optimization. Journal of machine learning research, 2012. 13(2).