Predicting Risk for Trauma Patients Using Static and Dynamic Information from the MIMIC III Database

ABSTRACT

Risk quantification algorithms in the ICU serve a dual purpose: (1) providing clinicians with early alerts for patients at extreme risk, and (2) optimizing the allocation of limited resources, including enabling remote management. Leveraging large datasets from electronic health records, predictive models can quantify patient risk effectively. In this study, we developed and evaluated machine learning models to classify trauma patients as high-risk or low-risk, where high-risk patients were defined as those who expired within the next 10 hours or within the last 10% of their ICU stay. Using the MIMIC-III database, we filtered 5,558 trauma patient records, of which 1,387 were non-survivors. Each record included 5 static variables (e.g., age, gender) and 28 dynamic variables (e.g., vital signs, metabolic panel).

Dynamic variables were processed using a 3-hour moving time window, treating each window as a unique patient-time fragment. From these fragments, we extracted statistical features, including the mean and median, as inputs for training. Gradient Boosting, Logistic Regression, and Random Forest classifiers were trained and evaluated, achieving a best AUROC score of 84.33%, as measured by the area under the receiver operating characteristic curve. This work demonstrates the potential of dynamic patient-time fragment modeling in reducing the complexity of ICU data, providing clinicians with actionable insights for high-risk trauma patients.

INTRODUCTION

Trauma remains the leading cause of death in the United States for individuals under 46 and is responsible for the highest number of overall expected years of life lost. Intensive care units (ICUs), with their continuous monitoring of patients, generate vast amounts of data that can be leveraged to predict patient mortality, assess time-dependent risk, and identify potential opportunities for data science and machine learning applications. Given the significant variability between patients, data-driven approaches are particularly well-suited for predicting patient outcomes, as developing patient-specific mechanistic models based on first principles is complex and often injury-specific. Additionally, time-series data presents unique opportunities for capturing patient trajectories, further enhancing predictive accuracy.

A growing body of work has explored the potential of existing databases in predicting ICU outcomes. For instance, Alistair et al. achieved an impressive AUROC of 92.4% by predicting mortality using features extracted from the first 24 hours of a patient's ICU stay. Harutyunyan et al. similarly predicted mortality within 24 hours, achieving an AUROC of 91.1%, while also forecasting average length of stay—an essential metric for ICU efficacy. Our group's earlier study predicted mortality within the National Trauma Data Bank (NTDB) with an AUROC of 91.8%. On the other hand, physics-based models have also been developed to simulate trauma outcomes, such as the system of ordinary differential equations by Ursino et al. to model the circulatory system under traumatic bleeding conditions. Despite the effectiveness of these models in simulating blood loss patterns, they often struggle to correlate specific injuries with outcomes. Furthermore, Hirshberg et al. explored the impact of blood transfusions on coagulopathy, demonstrating that resuscitation with more than five units of red blood cells could lead to dilutional coagulopathy, highlighting the complexity of trauma management and the need for better predictive models.

In this paper, we use the MIMIC-III ICU database to predict risk of death in trauma patients and whether a patient's health will begin to rapidly decline (analogously, a rapid rise in risk). We pose this problem as a time-series classification problem where the input is a fixed-length window of patient properties (both dynamic and static) with a 1-hour step size. Each patient 3-hour window is regarded as an individual patient-time fragment that is used for training or evaluating a model. The goal of this work is to develop a model that can continuously assess and predict patient mortality probability (a metric for quantifying patient risk) as data becomes available in real time.

METHODOLOGY

I. Patient Data Filtering

The data extraction process involved a meticulous filtering approach to ensure the relevance and quality of the dataset for predicting trauma patient risk in the ICU. First, trauma patients admitted to the ICU were identified using ICD-9 codes, ensuring the inclusion of only those whose conditions were relevant to the study's scope. Second, neonatal patients and those under 16 years of age were excluded to focus on adult and adolescent trauma cases, aligning with the study's objectives and mitigating physiological variability in younger populations. Third, patients with a "do not resuscitate" (DNR) status were excluded if they died within four hours of ICU admission. This step was crucial to avoid confounding factors in the dataset, as such cases often reflect pre-determined clinical decisions rather than the ICU's impact on patient outcomes. These comprehensive filtering criteria resulted in a refined dataset that allowed for more robust and accurate modeling of trauma patient risk.

II. Data Preprocessing

In this project, extensive data preprocessing was conducted to prepare the dataset for model training. First, outliers were identified and addressed to ensure the robustness of the models. Categorical variables, such as sex, were transformed into numerical representations using label encoding to make them compatible with machine learning algorithms. Missing values in static variables, including demographic details, were handled using mean imputation to replace missing values with the average, preserving the overall distribution. For critical static variables like height and weight, regression imputation was employed to predict their values based on other correlated variables in the dataset. This method was chosen as it leverages relationships between features to provide more accurate estimations compared to simpler imputation methods, which is crucial for these variables due to their impact on clinical predictions.

For dynamic variables, which included time-series data such as vital signs and lab results, linear interpolation was applied to fill in missing values. This method maintains the temporal continuity of the data, ensuring that the dynamic variables reflect realistic trends over time. Additionally, patient-time fragments were created using a 3-hour moving window, allowing for the extraction of summary statistics like the mean and median, which capture essential trends and patterns in the data. These preprocessing steps ensured that the data was clean, complete, and ready for input into machine learning models, ultimately improving the performance and reliability of the predictions.

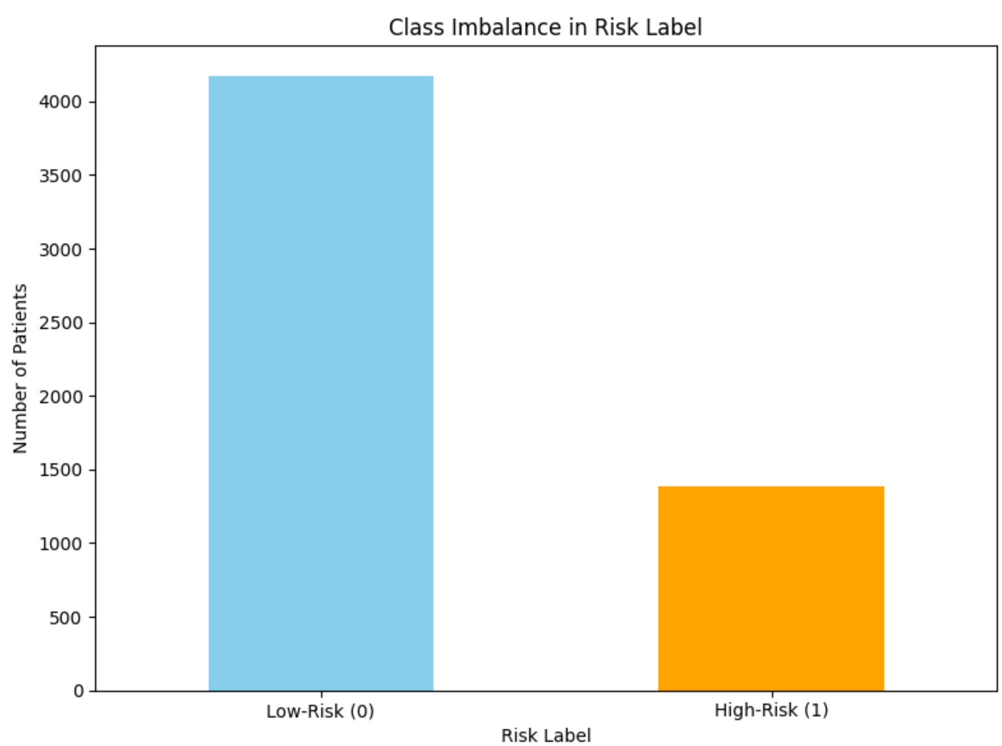
III. Data Formatting

In the data formatting process, we structured the dataset using a 3-hour moving time window approach to convert the raw time-series data into a format suitable for predictive modeling. Data from the three most recent time points, spaced at 1-hour intervals, was used as input to predict mortality risk over the subsequent 10 hours. This setup enabled the creation of our target variable, 'risk_label,' which categorized patients into two groups: high-risk for those who expired within the next 10 hours and low-risk for those who survived beyond that period. Data formatting was crucial for transforming the complex, continuous, and high-dimensional time-series data into manageable fragments, making it easier for machine learning algorithms to process and learn patterns. We hypothesized that recent patient history would provide more predictive power than long-term historical data, enhancing the model's ability to assess real-time risk. Within each 3-hour time fragment, we computed the mean and median of all features to capture essential statistical summaries, ensuring that the model effectively represented dynamic changes in patient conditions while maintaining computational efficiency.

IV. Class Imbalances

In our dataset, we observed a significant class imbalance, with 1,387 high-risk patients and 4,169 low-risk patients. This imbalance indicates that high-risk cases constitute a much smaller portion of the dataset compared to low-risk cases. Class imbalance poses a challenge in predictive modeling because models trained on such data may become biased toward the majority class (low-risk) and fail to accurately predict the minority class (high-risk).

Addressing class imbalance is crucial in healthcare applications, where accurately identifying high-risk patients can have life-saving implications. To mitigate this issue, we applied class imbalance handling techniques, such as oversampling the minority class, under sampling the majority class, or adjusting the class weights in our models. These methods help ensure that the model pays equal attention to both classes, improving its ability to distinguish between high-risk and low-risk patients. By addressing this imbalance, we aimed to enhance the model’s performance and reliability, particularly in correctly predicting high-risk patients.

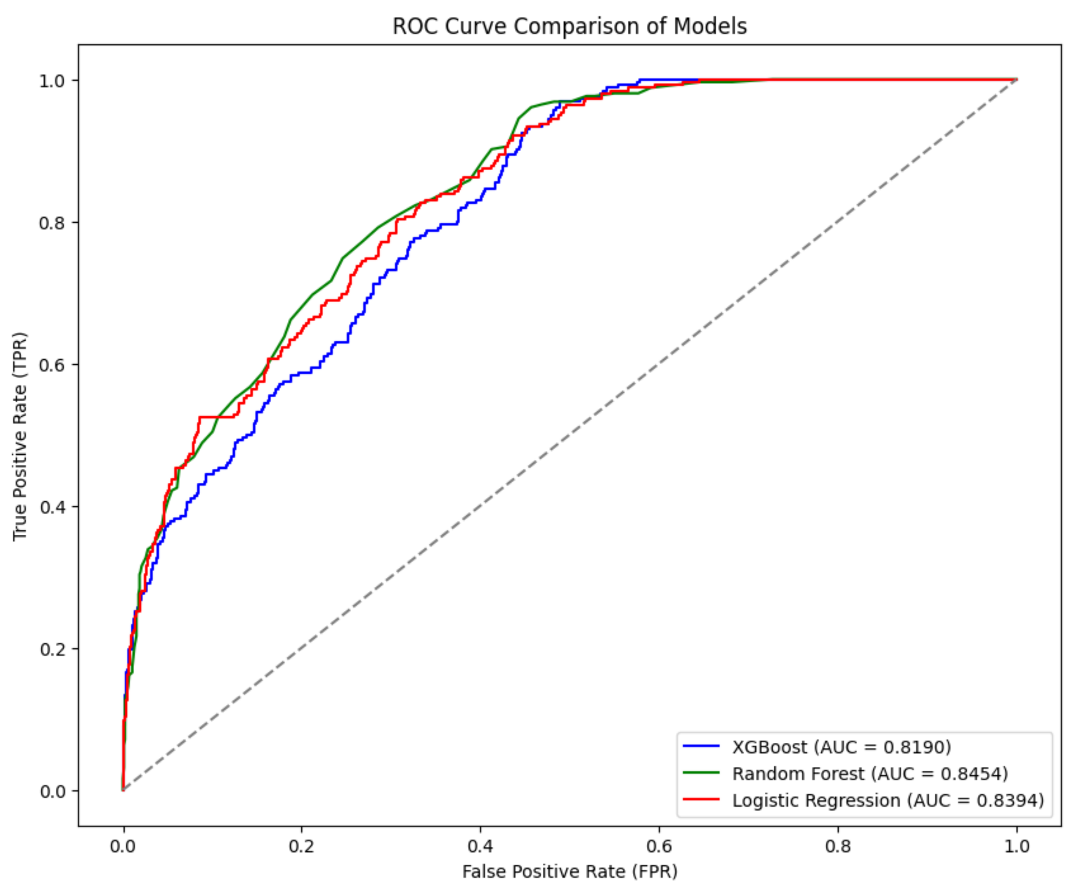


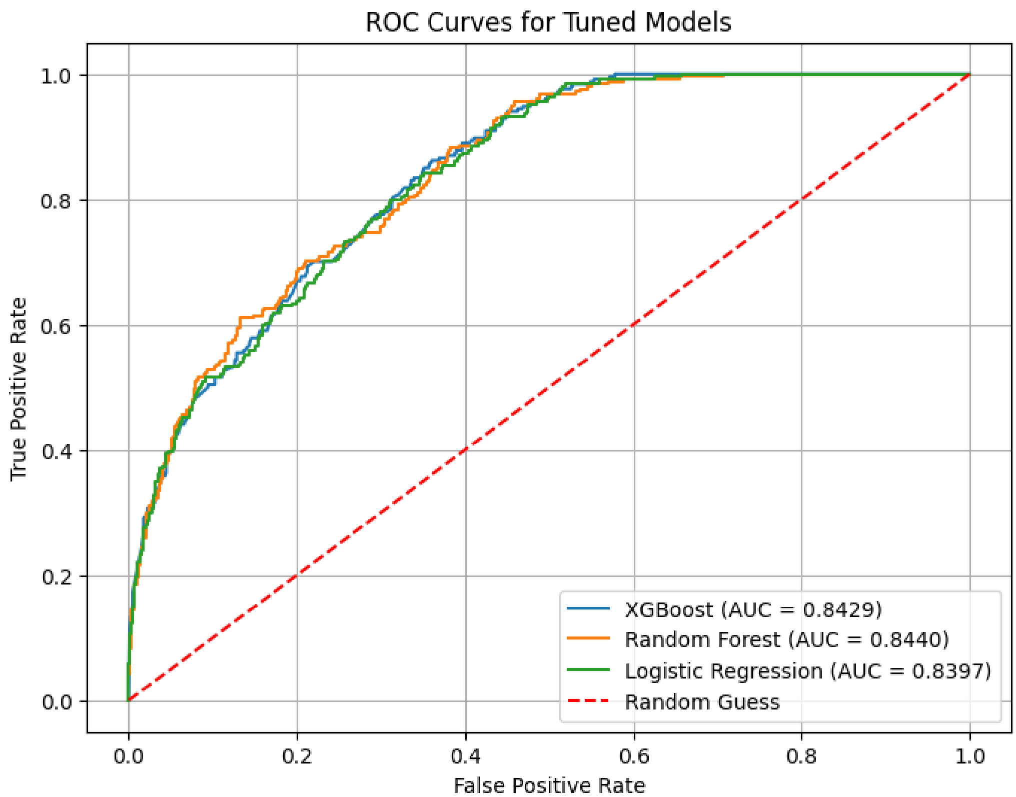
V. Data Modelling

We followed a systematic approach to model training and evaluation for predicting high-risk and low-risk trauma patients in the ICU using machine learning. First, we trained three different models: Gradient Boosting, Random Forest, and Logistic Regression. The models were evaluated on several metrics, including accuracy, precision, recall, F1 score, and ROC AUC. The results showed that all models performed well, with Logistic Regression and Random Forest both achieving an accuracy of 82.19%. The best performance in terms of AUC was from Logistic Regression, with an AUC of 0.6989, slightly outperforming the others.

Next, we performed hyperparameter tuning to optimize the performance of the models using grid search and cross-validation. For XGBoost, the best parameters were found to be a learning rate of 0.01, a maximum depth of 3, 100 estimators, and a subsample rate of 0.8, yielding an AUC of 0.8432. For Random Forest, the optimal hyperparameters included a max depth of 10, a minimum samples leaf of 2, and 300 estimators, leading to an AUC of 0.8422. Logistic Regression's best parameters were found to be a regularization strength of 1 and L2 penalty, with an AUC of 0.8433, the highest among the models after tuning.

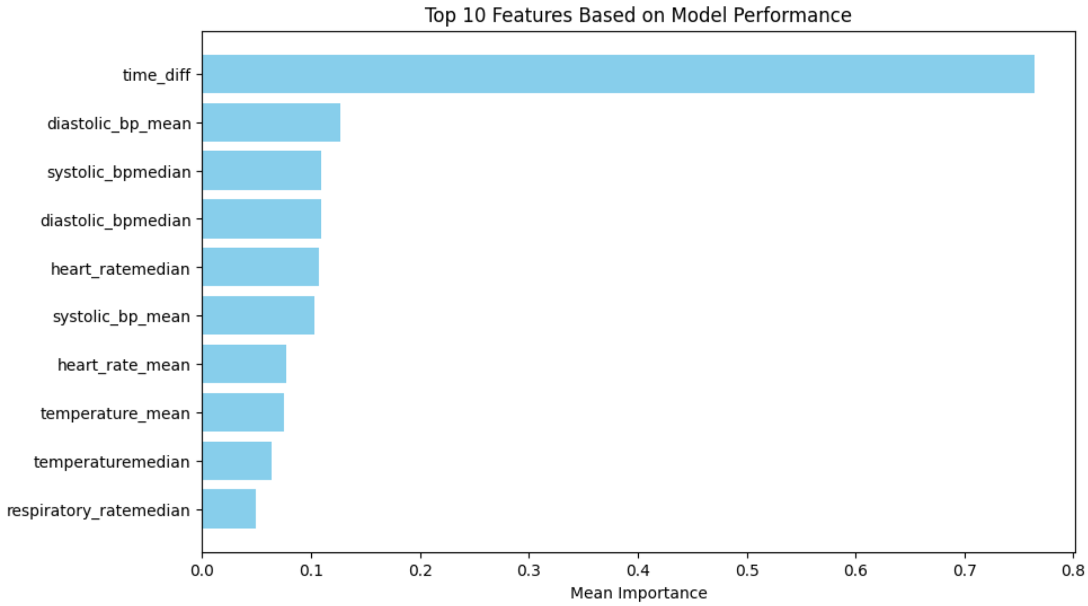
After hyperparameter tuning, we tested the models on a test set, and the final ROC-AUC scores indicated that Logistic Regression performed best with a score of 0.8433.





Feature Importance

we conducted a feature importance analysis for the Logistic Regression model, which emerged as the best-performing model for predicting high-risk trauma patients in the ICU. Feature importance helps us understand which variables contribute the most to the model's decision-making process. Among the top 10 features, the most influential was time_diff, which had the highest mean importance score of 0.7639, indicating its significant role in predicting patient risk. Other vital features included diastolic_bp_mean (0.1270) and systolic_bpmedian (0.1094), highlighting the importance of blood pressure measurements in assessing patient condition. Variables such as heart_rate_mean (0.0767) and temperature_mean (0.0752) were also crucial, suggesting that vital signs play a critical role in determining patient risk. This analysis not only underscores the relevance of physiological data in predicting patient outcomes but also helps clinicians prioritize the most impactful variables for decision-making in critical care settings.

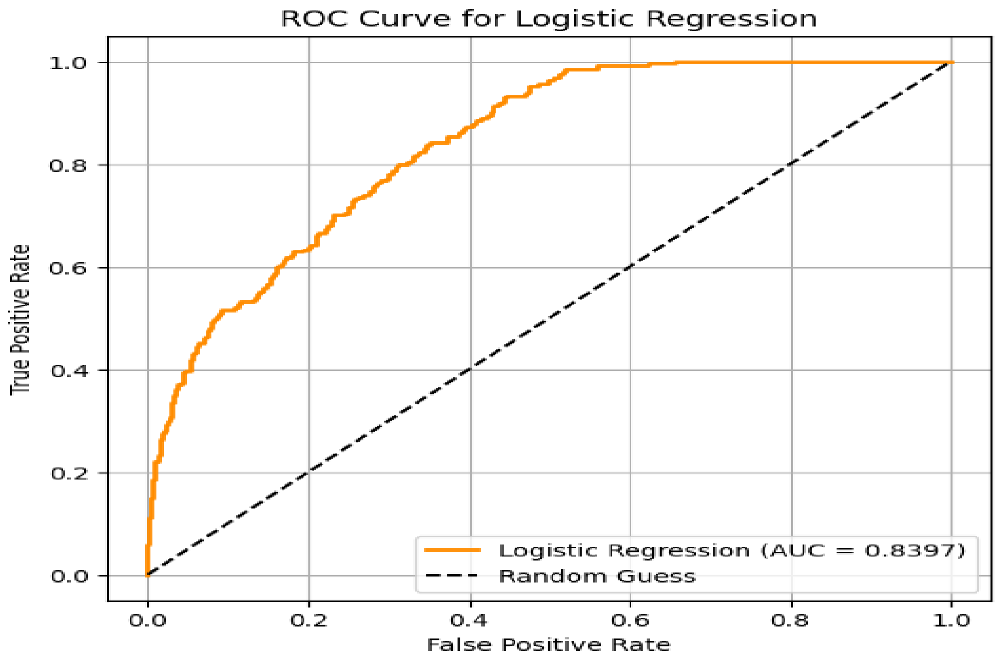


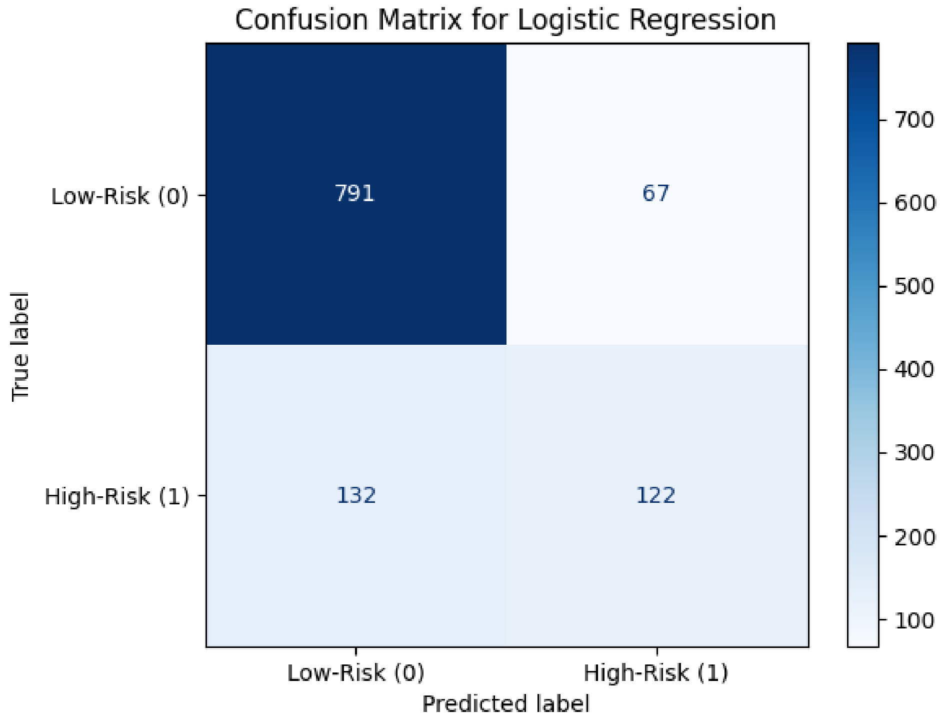
Results

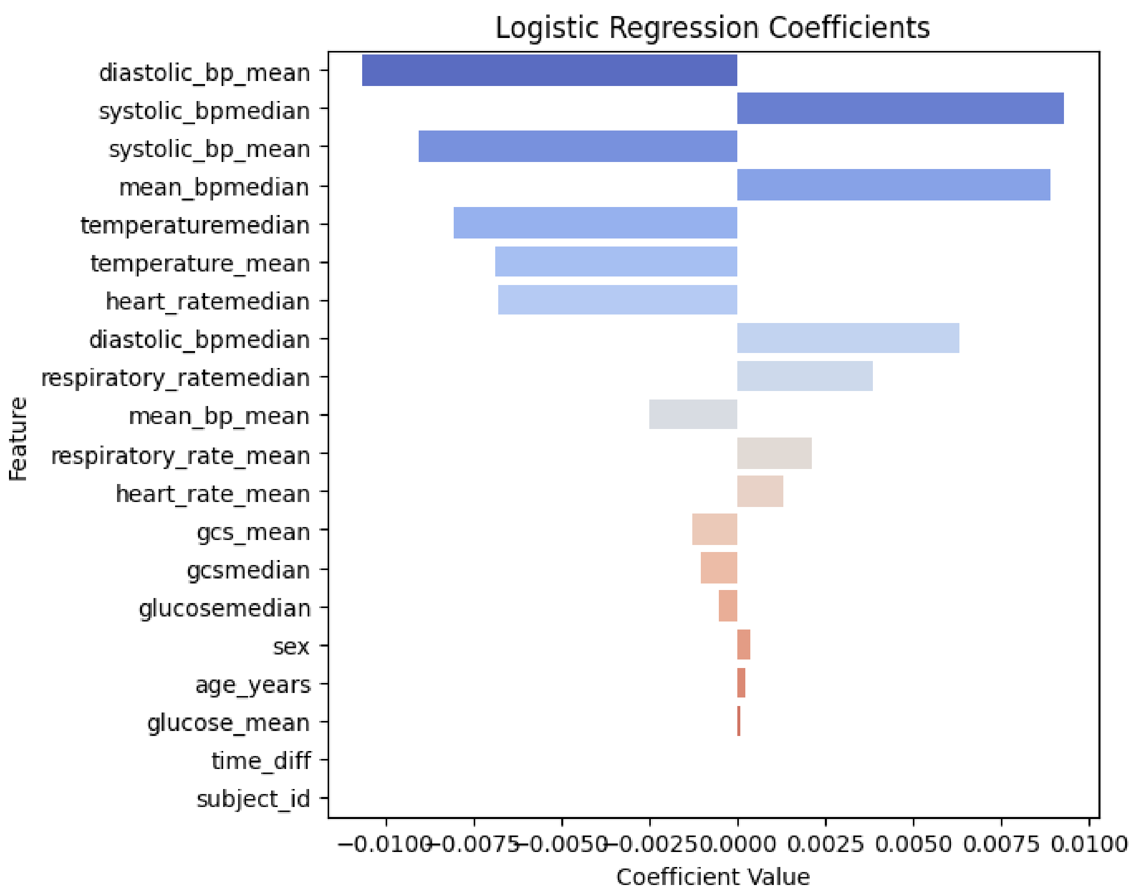
The evaluation of the three models—XGBoost, Random Forest, and Logistic Regression—demonstrated varied performance across different metrics. In terms of accuracy, both Random Forest and Logistic Regression achieved 82.19%, outperforming XGBoost with an accuracy of 79.77%. However, when considering precision, Random Forest performed the best (71.54%), followed by Logistic Regression (65.22%) and XGBoost (57.36%). Despite its higher precision, Random Forest showed a lower recall (36.61%) compared to Logistic Regression (47.24%) and XGBoost (44.49%), indicating that Random Forest was less sensitive to identifying high-risk patients. The F1 score, which balances precision and recall, revealed a similar trend, with Logistic Regression leading (0.5479), followed by XGBoost (0.5011) and Random Forest (0.4844).
In terms of ROC AUC, which is critical for assessing model performance in imbalanced datasets, Logistic Regression again emerged as the best model with an AUC of 0.6989, followed closely by XGBoost with an AUC of 0.6735 and Random Forest with an AUC of 0.6615. Hyperparameter tuning improved the models' performance, with Logistic Regression achieving the highest AUC after tuning (0.8433). XGBoost and Random Forest also benefitted from tuning, with their best AUC scores being 0.8432 and 0.8422, respectively. After testing the final model, Logistic Regression's ROC AUC score of 0.8397 confirmed its superior ability to distinguish between high-risk and low-risk patients.

Further analysis of Logistic Regression's coefficients revealed the influence of key features on the model's predictions. Features such as diastolic_bp_mean (-0.0106) and heart_ratemedian (-0.0068) had negative coefficients, indicating that higher diastolic blood pressure and lower heart rates were associated with lower risks. Conversely, systolic_bpmedian (0.0093) and mean_bpmedian (0.0089) exhibited positive coefficients, suggesting that these vital signs were linked to higher mortality risk. Additionally, features like temperature_mean (-0.0069) and respiratory_rate_mean (0.0021) had more moderate effects, showing how temperature and

respiratory rate influence the risk classification. The small coefficients for factors like age_years (0.0002) and *sex* (0.0004) suggested minimal impact on the outcome, emphasizing the importance of physiological features in predicting high-risk trauma patients.
Overall, while Random Forest showed strong precision, Logistic Regression's balance between recall, AUC, and its interpretability through coefficients made it the most effective model for this task.




Confusion Matrix for Logistic Regression
Low-Risk (0)
High-Risk (1)
True label
791
67
132
122
Low-Risk (0)
High-Risk (1)
Predicted label
700
600
500
400
300
200
100


Logistic Regression Coefficients
diastolic_bp_mean
systolic_bpmedian
systolic_bp_mean
mean_bpmedian
temperaturemedian
temperature_mean
heart_ratemedian
diastolic_bpmedian
respiratory_ratemedian
mean_bp_mean
respiratory_rate_mean
heart_rate_mean
gcs_mean
gcsmedian
glucosemedian
sex
age_years
glucose_mean
time_diff
subject_id
Feature
−0.0100
−0.0075
−0.0050
−0.0025
0.0000
0.0025
0.0050
0.0075
0.0100
Coefficient Value

Discussions

The models developed in this project, particularly Logistic Regression, demonstrated the ability to classify patients into high-risk and low-risk categories with a substantial ROC AUC score of 0.8397. This could have profound healthcare implications, as early identification of high-risk patients can allow clinicians to prioritize care, allocate resources more efficiently, and potentially improve patient outcomes by taking timely interventions. It also emphasizes the importance of using predictive models for decision support in critical care settings, where resource management is often a challenge.

However, there were several challenges encountered during the project. One of the most notable was handling class imbalance, where the low-risk class vastly outnumbered the high-risk class. This imbalance could have led to model bias, where the model may have been inclined to predict the majority class (low-risk) more often. Techniques like class balancing and careful evaluation using metrics like ROC AUC helped mitigate this issue, but it also highlighted the importance of choosing the right performance metrics when dealing with imbalanced datasets. Additionally, data preprocessing posed its own set of challenges, such as dealing with missing values using imputation techniques like regression for continuous features and transforming categorical variables. Proper feature engineering, including the creation of time windows for dynamic features, was essential to capture patient trajectories and ensure that the model had access to relevant information.

Findings:

- The process taught us that even slight adjustments in data handling (e.g., imputation methods) and model choice can substantially impact results. For instance, Logistic Regression, despite being a simpler model, outperformed more complex ones in this case, underscoring the importance of model selection based on the task at hand.
- Evaluating the model using multiple metrics like precision, recall, F1-score, and AUC is necessary to understand performance comprehensively, especially in healthcare applications where false negatives (misclassifying high-risk patients as low-risk) can have dire consequences.
- Data preprocessing and feature engineering are key to the success of any machine learning project, especially in healthcare where patient data can be complex and sparse.
- Class imbalance is a significant challenge in predictive healthcare models, and addressing it early through techniques like resampling or careful evaluation is essential.
- Hyperparameter tuning is critical to enhance model performance, but different models may require different approaches.

Recommendations:

- Given the promising results with Logistic Regression, future studies should explore ensemble methods or deep learning to see if they can further improve accuracy and handle complex interactions between variables.
- Real-time integration of the predictive models into hospital management systems could lead to more actionable insights for clinicians.
- Additional research could focus on incorporating other types of data, such as patient history, genetic data, or real-time monitoring, which could potentially refine risk predictions.

Conclusion

This project demonstrates the transformative potential of machine learning in enhancing healthcare delivery, particularly in the ICU environment. By leveraging dynamic patient data and addressing challenges such as class imbalance, we have developed a robust predictive model capable of identifying high-risk trauma patients with greater accuracy. Our findings underscore the importance of continuous model evaluation and refinement, particularly in high-stakes settings like critical care. The successful application of these techniques not only contributes to improved patient outcomes but also offers significant potential for optimizing resource allocation and early intervention strategies in ICU management. Moving forward, we aim to refine our models further and explore their integration into real-time clinical decision support systems, ultimately advancing the quality and efficiency of patient care in critical environments.

References

https://journals.plos.org/plosone/article/figures?id=10.1371/journal.pone.0262523

https://physionet.org/content/mimiciii/1.4/

https://www2.gov.bc.ca/assets/gov/health/practitioner-pro/medical-services-plan/diag-codes_injury.pdf

DataPreprocessingHealthcare.pdf
https://canvas.csun.edu/courses/158104/files/26565792?module_item_id=7401254