# Predicting Long-Term Mortality and Assessing Weekend Admission Effects in ICU Patients: A Multi-Model Approach Using Early Admission Data

Liyang Li

November 28, 2024

## 1 Introduction

Predicting mortality in intensive care unit (ICU) patients has become a focal point in medical research, aiming to assist clinicians in identifying high-risk patients, optimizing resource allocation, and improving management strategies[1]. Machine learning models, such as Random Forest and XGBoost, have demonstrated significant potential in leveraging early physiological data to predict high-risk patients. For example, Johnson et al. highlighted the importance of variables such as blood glucose levels, positive blood culture results, and ventilator use in predicting short-term mortality. However, most studies primarily focus on in-hospital or short-term mortality, while the factors influencing long-term all-cause mortality remain underexplored and warrant further investigation.

This study aims to fill this gap by developing a machine learning model based on the first 24 hours of physiological and therapeutic data of ICU admissions from the MIMIC-III database. Our goal is to improve the accuracy of long-term mortality predictions and provide a practical decision-support tool for clinical practice.

Additionally, we explore the "weekend effect," a phenomenon where ICU patients admitted over the weekend may experience higher in-hospital mortality rates compared to those admitted on weekdays. Specifically, Bell and Redelmeier (2001) found that patients admitted during weekends exhibited significantly higher in-hospital mortality, potentially due to limited medical resources or treatment delays[2]. However, the impact of the weekend effect on long-term all-cause mortality remains unclear. This study will analyze whether weekend admissions independently increase the risk of long-term mortality, offering new insights into this phenomenon.

# 2 Methods

## 2.1 Study Population

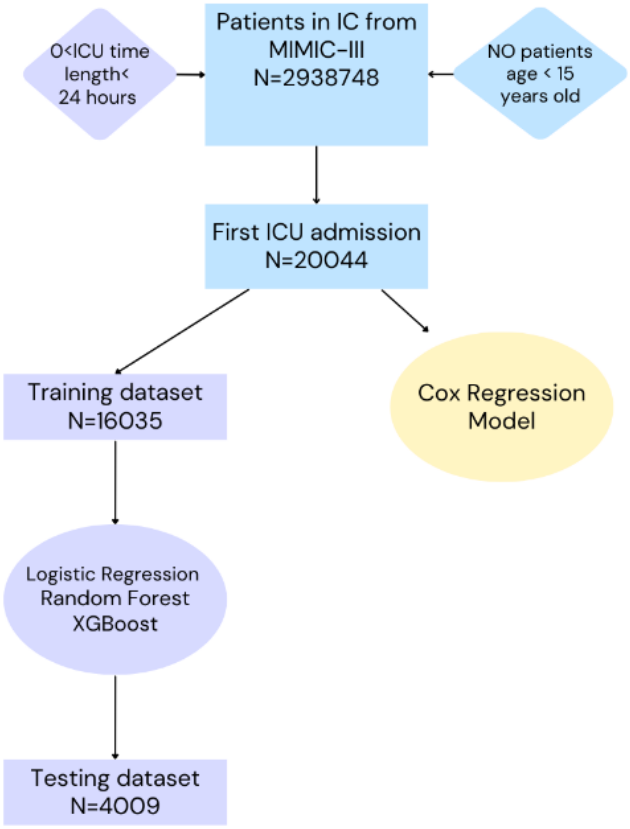


Figure 1: Study flow chart. According to the inclusion and exclusion criteria of the study, the populations of the training set and test set were obtained. Exclusion of patients with ICU stay less than 24 hours applies only to Research Question 1. No patients under the age of 15 were included. The variable selection process and the number of variables are not shown in this figure.

This study conducted a retrospective cohort analysis based on the MIMIC-III database, a publicly available and de-identified ICU database encompassing over 40,000 ICU admissions from 2001 to 2012 [3]. The database integrates data from the Philips CareVue and MDsoft MetaVision ICU systems, providing comprehensive patient information, including demographics, laboratory results, blood cultures, vital signs, Glasgow Coma Scale (GCS) scores, ICU admission records, and survival data. The study primarily utilized tables containing the most relevant variables. Since all data are de-identified and the study involves no clinical interventions, no ethical committee approval or informed consent was required. Inclusion criteria were patients listed in the "selected" table, excluding those under 15 years of age or non-first ICU admissions. For research question 1, patients with ICU stays of less than 24 hours were further excluded.

## 2.2 Data Preprocessing

The dataset underwent preprocessing, including variable selection, missing value handling, and data integration. Variables were categorized into categorical variables (e.g., *expire_flag*, *admission_type*) and continuous variables (e.g., *age_years*, *glucose*, GCS scores). Missing values for continuous variables were imputed using the median to maintain distributional consistency, while missing GCS scores were filled with 0 to reflect unrecorded statuses. Categorical variables were imputed with "Unknown" to represent missing categories. Additionally, to mitigate the impact of outliers on the model, continuous variables were trimmed at the 1% and 99% percentiles.

For continuous variables with multiple daily measurements (e.g., vital signs and laboratory test results), the model utilized the maximum or most critical values of the day (e.g., lowest blood pressure, highest lactate level) based on recommendations from previous literature [1], reflecting the most severe condition of the patient.

## 2.3 Variable Selection

This study identified key variables related to ICU patient mortality based on two main criteria: (1) commonly used variables in the literature, including physiological parameters (e.g., *systolic_bp*, *heart_rate*), laboratory markers (e.g., *lactate*, *whitebloodcell*), and Glasgow Coma Scale (GCS) scores; (2) variables deemed clinically significant for ICU management and patient outcomes based on expert recommendations and practical experience (e.g., *spo2* and *admission_type*). A candidate variable list was compiled through a review of the literature and clinical knowledge, and the presence of these variables in the dataset was systematically verified to ensure their availability for analysis.

For the second research question, Lasso regression [10] was further employed to perform feature selection, introducing regularization constraints to effectively reduce redundant variables and narrow down the feature set. Additionally, important confounding factors identified in the first research question were included to ensure that these critical variables were appropriately controlled in subsequent analyses. The selected variables were further evaluated using univariate Cox regression [11] to assess their significance in predicting survival outcomes, laying the foundation for constructing the multivariate Cox regression model.

## 2.4 ML Model

This study utilized Python 3.12 and the scikit-learn 0.22.2 library [7] to construct and train machine learning models. The dataset was randomly split into 80% for training and 20% for testing. To optimize model performance, hyperparameters were tuned using grid search combined with five-fold cross-validation. Logistic regression (baseline), random forest, and XGBoost models were trained and compared.

The primary evaluation metric was AUROC[4] along with its 95% confidence interval.

Secondary evaluation metrics included the F1[5] score , accuracy, and AUPRC[6] , to comprehensively assess model performance in predicting in-hospital mortality. Calibration curves and the Brier score were also used to evaluate the reliability of predicted probabilities.

Finally, feature importance in the XGBoost model was visualized using feature weights, mean gain, and coverage. SHAP[9] was employed to interpret the contribution of variables to the prediction outcomes, revealing key features influencing ICU patient mortality.

### 2.5 Survival Model

All significant variables identified through the selection process were included in the multivariate Cox regression model [11]. Stepwise regression (bidirectional selection) was a pplied to further refine the variable selection. To ensure model optimization and stabil-i ty, ANOVA and Likelihood Ratio Tests (LRT) were conducted to evaluate and compare the three models, determining the retained variables and the overall reasonableness of the final model.

### 2.6 Statistical Analysis

The statistical analysis in this study included the following components: categorical variables were described using frequencies and percentages (e.g., gender, admission source), while continuous variables were summarized using medians and interquartile ranges (IQR) to describe central tendencies and distribution characteristics (e.g., age, laboratory data). For group comparisons, Fisher's exact test was applied to categorical variables, and the Mann-Whitney U test was used for continuous variables to analyze differences between weekend and weekday admissions.

Additionally, patients admitted on Saturdays and Sundays were categorized as "weekend admissions," and the proportions and distribution differences between weekend and weekday admissions were calculated to investigate the "weekend effect." Statistical significance was assessed using two-sided tests, with a significance threshold of $P < 0.05$. All analyses were conducted using Python 3.12 (Ubuntu 22.04) and R version 4.1 (R Studio, Inc.).

## 3 Results

### 3.1 Baseline Characteristics

The median age of patients in this study was 64 years (IQR: 51–77 years), and the median hospital length of stay was 4.56 days. Most patients were admitted through the emergency department (89.04%, data from Table 1), with a slightly higher proportion

on weekends (90.74% vs. weekdays 90.19%). The proportion of elective admissions was significantly lower on weekends (3.32% vs. weekdays 5.58%, data from Table 1).

## 3.2 Mortality Rates

The overall mortality rate was 66.06%, with weekend admissions having a significantly lower mortality rate (60.71%) compared to weekdays (68.10%, $p < 0.001$, data from Table 2). Additionally, the median age of patients admitted on weekends (65 years) was significantly higher than that on weekdays (63 years, $p < 0.001$, data from Table 1).

## 3.3 Impact of Age on Mortality

Age was identified as a crucial factor influencing mortality, with mortality rates increasing significantly across age groups ($p < 0.001$, data from Table 2). Specifically, the mortality rate was 29.27% for patients under 40 years of age, 51.14% for those aged 40–60 years, 80.05% for those aged 60–80 years, and 92.18% for patients over 80 years. The high mortality among elderly patients may be associated with heavier comorbid burdens and reduced physiological reserve.

## 3.4 Admission Patterns and Survival Probabilities

The proportion of admissions through the emergency department was significantly higher on weekends (55.05% vs. 43.09%, data from Table 1), while the proportion of transfer admissions was slightly higher on weekends (24.63% vs. weekdays 23.69%, data from Table 1). Kaplan-Meier survival curve analysis revealed that factors such as insurance status (e.g., Medicare), admission type (emergency department or physician referral), and transfer mechanisms significantly influenced patient survival probabilities. Patients with Medicare insurance or those admitted through physician referral or the emergency department exhibited higher survival rates, possibly due to more timely medical interventions or resource access.

## 3.5 ICU Admission and Disposition

Regarding ICU admissions, the majority of patients were first admitted to the MICU (Medical ICU), accounting for the highest proportion (44.15%, data from Table 1), while the CSRU (Cardiac Surgery Recovery Unit) accounted for the lowest proportion (8.52%, data from Table 1). The final disposition of patients was generally consistent between weekday and weekend admissions, indicating that the allocation of critical care resources was not significantly affected by admission timing.

Table 1: Number of hospital admissions according to sociodemographic, clinical and health resource characteristics broken down by weekday or weekend

| Category | Total | | Weekday | | Weekend | |
|---|---|---|---|---|---|---|
| | N | % | N | % | N | % |
| **Total** | 2,938,748 | 100.00 | 2,139,505 | 72.80 | 799,243 | 27.20 |
| **Insurance Type** | | | | | | |
| Medicare | 1,588,673 | 54.07 | 1,162,221 | 54.32 | 425,452 | 53.23 |
| Private | 923,544 | 31.42 | 676,634 | 31.63 | 246,910 | 30.89 |
| Medicaid | 284,673 | 9.69 | 205,280 | 9.59 | 79,393 | 9.93 |
| Self Pay | 45,588 | 1.55 | 25,452 | 1.19 | 20,136 | 2.52 |
| Government | 97,270 | 3.31 | 69,918 | 3.27 | 27,352 | 3.42 |
| **Admission Type** | | | | | | |
| Emergency | 2,616,792 | 89.05 | 1,896,104 | 88.62 | 720,688 | 90.17 |
| Elective | 235,804 | 8.02 | 199,293 | 9.31 | 36,511 | 4.57 |
| Urgent | 86,152 | 2.93 | 44,108 | 2.06 | 42,044 | 5.26 |
| **Admission Location** | | | | | | |
| Emergency Room Admit | 1,330,784 | 45.29 | 922,160 | 43.10 | 408,624 | 51.13 |
| Transfer from Hosp/Extram | 776,542 | 26.43 | 582,112 | 27.21 | 194,430 | 24.33 |
| Clinic Referral/Premature | 501,217 | 17.05 | 370,230 | 17.30 | 130,987 | 16.39 |
| Phys Referral/Normal Deli | 303,629 | 10.34 | 248,473 | 11.61 | 55,156 | 6.90 |
| Transfer from Skilled Nur | 24,068 | 0.82 | 14,946 | 0.70 | 9,122 | 1.14 |
| Transfer from Other Health | 2,229 | 0.08 | 1,305 | 0.06 | 924 | 0.12 |
| ** Info Not Available ** | 267 | 0.01 | 267 | 0.01 | 0 | 0.00 |
| Trsf Within This Facility | 12 | 0.00 | 12 | 0.00 | 0 | 0.00 |
| **Age Group** | | | | | | |
| 60–79 | 1,193,325 | 40.63 | 896,043 | 41.88 | 297,282 | 37.20 |
| 40–59 | 872,544 | 29.69 | 633,128 | 29.59 | 239,416 | 29.96 |
| >80 | 548,282 | 18.65 | 382,680 | 17.89 | 165,602 | 20.72 |
| 20–39 | 298,301 | 10.15 | 210,578 | 9.84 | 87,723 | 10.98 |
| 0–19 | 26,296 | 0.89 | 17,076 | 0.80 | 9,220 | 1.15 |
| **First Careunit** | | | | | | |
| MICU | 1,297,350 | 44.15 | 942,055 | 44.03 | 355,295 | 44.45 |
| TSICU | 445,823 | 15.18 | 310,068 | 14.49 | 135,755 | 16.99 |
| CCU | 334,982 | 11.40 | 242,386 | 11.33 | 92,596 | 11.59 |
| SICU | 610,015 | 20.76 | 416,167 | 19.45 | 193,848 | 24.25 |
| CSRU | 250,578 | 8.51 | 228,829 | 10.70 | 21,749 | 2.72 |
| **Last Careunit** | | | | | | |
| MICU | 1,294,284 | 44.03 | 941,465 | 44.00 | 352,819 | 44.14 |
| TSICU | 435,892 | 14.83 | 301,890 | 14.11 | 134,002 | 16.77 |
| CCU | 297,683 | 10.13 | 218,600 | 10.22 | 79,083 | 9.89 |
| SICU | 659,551 | 22.45 | 449,708 | 21.02 | 209,843 | 26.26 |
| CSRU | 251,338 | 8.56 | 227,842 | 10.65 | 23,496 | 2.94 |

Table 2: Hospital Admissions and Mortality Rates by Category (Including 95% Confidence Intervals)

| Category | Weekday | | Weekend | |
|---|---|---|---|---|
| | **N** | **% (95% CI)** | **N** | **% (95% CI)** |
| **Total** | 2,139,505 | 51.64 (51.58–51.71) | 799,243 | 54.26 (54.16–54.37) |
| **Insurance Type** | | | | |
| Medicare | 1,162,221 | 64.36 (64.27–64.45) | 425,452 | 70.45 (70.32–70.59) |
| Private | 676,634 | 35.99 (35.88–36.11) | 246,910 | 34.34 (34.16–34.53) |
| Medicaid | 205,280 | 41.91 (41.70–42.12) | 79,393 | 41.27 (40.93–41.61) |
| Self Pay | 25,452 | 28.05 (27.50–28.61) | 20,136 | 61.20 (60.52–61.87) |
| Government | 69,918 | 28.93 (28.59–29.27) | 27,352 | 14.87 (14.45–15.29) |
| **Admission Type** | | | | |
| Emergency | 1,896,104 | 53.24 (53.17–53.31) | 720,688 | 54.16 (54.05–54.28) |
| Elective | 199,293 | 32.63 (32.42–32.83) | 36,511 | 58.69 (58.19–59.20) |
| Urgent | 44,108 | 68.87 (68.44–69.30) | 42,044 | 52.16 (51.68–52.64) |
| **Admission Location** | | | | |
| Emergency Room Admit | 922,160 | 59.37 (59.27–59.47) | 408,624 | 55.90 (55.75–56.05) |
| Transfer from Hosp/Extram | 582,112 | 53.56 (53.43–53.69) | 194,430 | 56.05 (55.83–56.28) |
| Clinic Referral | 370,230 | 38.28 (38.12–38.44) | 130,987 | 41.25 (40.99–41.52) |
| Phys Referral | 248,473 | 36.51 (36.32–36.70) | 55,156 | 60.60 (60.19–61.01) |
| Transfer from Skilled Nur | 14,946 | 81.06 (80.42–81.68) | 9,122 | 87.02 (86.32–87.69) |
| Transfer from Other Health | 1,305 | 61.69 (59.02–64.29) | 924 | 98.05 (96.94–98.76) |
| ** Info Not Available ** | 267 | 100.00 (98.58–100.00) | 0 | 0.00 (0.00–0.00) |
| **Age Group** | | | | |
| 60–79 | 896,043 | 57.46 (57.36–57.57) | 297,282 | 63.91 (63.74–64.08) |
| 40–59 | 633,128 | 41.56 (41.44–41.68) | 239,416 | 39.47 (39.28–39.67) |
| >80 | 382,680 | 73.52 (73.38–73.66) | 165,602 | 79.94 (79.75–80.13) |
| 20–39 | 210,578 | 20.83 (20.66–21.00) | 87,723 | 17.57 (17.32–17.83) |
| 0–19 | 17,076 | 10.01 (9.57–10.47) | 9,220 | 15.26 (14.54–16.01) |
| **First Care Unit** | | | | |
| MICU | 942,055 | 59.62 (59.52–59.72) | 355,295 | 59.01 (58.85–59.17) |
| TSICU | 310,068 | 40.67 (40.50–40.85) | 135,755 | 48.21 (47.95–48.48) |
| CCU | 242,386 | 59.68 (59.48–59.87) | 92,596 | 49.63 (49.30–49.95) |
| SICU | 416,167 | 47.81 (47.66–47.97) | 193,848 | 51.69 (51.47–51.91) |
| CSRU | 228,829 | 32.14 (31.95–32.33) | 21,749 | 57.24 (56.59–57.90) |
| **Last Care Unit** | | | | |
| MICU | 941,465 | 60.20 (60.10–60.30) | 352,819 | 58.75 (58.59–58.91) |
| TSICU | 301,890 | 40.33 (40.15–40.50) | 134,002 | 44.65 (44.38–44.91) |
| CCU | 218,600 | 61.57 (61.36–61.77) | 79,083 | 54.04 (53.69–54.38) |
| SICU | 449,708 | 47.53 (47.39–47.68) | 209,843 | 53.66 (53.45–53.87) |
| CSRU | 227,842 | 29.89 (29.70–30.07) | 23,496 | 47.95 (47.31–48.59) |

## 3.6 ML Model

This study collected 82 clinical features from the first 24 hours of ICU admission and, through a combination of literature review and clinical expertise, identified 18 features significantly associated with long-term mortality. Logistic regression (baseline), random forest, and XGBoost models were employed for training and performance comparison.

| ML | Cohort | Accuracy | AUC (95% CI) | Sensitivity | Specificity |
|---|---|---|---|---|---|
| LR | Training cohort | 0.6366 | 0.7523 (0.7521, 0.7526) | 0.8302 | 0.5065 |
| | Testing cohort | 0.6348 | 0.7487 (0.7482, 0.7491) | 0.8238 | 0.5077 |
| RF | Training cohort | 0.7241 | 0.8469 (0.8467, 0.8471) | 0.8713 | 0.6252 |
| | Testing cohort | 0.6687 | 0.7607 (0.7602, 0.7612) | 0.7934 | 0.5849 |
| XGB | Training cohort | 0.7345 | 0.8009 (0.8007, 0.8011) | 0.5829 | 0.8364 |
| | Testing cohort | 0.7131 | 0.7730 (0.7726, 0.7735) | 0.5428 | 0.8277 |

Table 3: Performance Comparison of Machine Learning Models: Logistic Re- gression (LR), Random Forest (RF), and XGBoost (XGB)

The primary evaluation metric was AUROC along with its 95% confidence interval. Secondary evaluation metrics included the F1 score, accuracy, and AUPRC. On the test set, XGBoost achieved an AUROC of **0.773 (95% CI: 0.7726, 0.7735)**, significantly outperforming random forest (**AUROC = 0.761, 95% CI: 0.7602, 0.7612**) and logistic regression (**AUROC = 0.749, 95% CI: 0.7482, 0.7491**). These results demonstrate that XGBoost has superior discriminatory power in differentiating positive and negative samples. Additionally, XGBoost achieved an AUPRC of **0.691**, surpassing random forest (**0.665**) and logistic regression (**0.654**), highlighting its significant advantage in handling imbalanced data.

For secondary metrics, the F1 score, which balances precision and recall, was **0.6034** for XGBoost, slightly lower than random forest (**0.6583**) and logistic regression (**0.6447**), indicating a trade-off between precision and recall. However, XGBoost demonstrated superior overall classification performance with an accuracy of **71.31%**, compared to logistic regression (**63.48%**) and random forest (**66.87%**). These results highlight the robust classification capability of XGBoost.

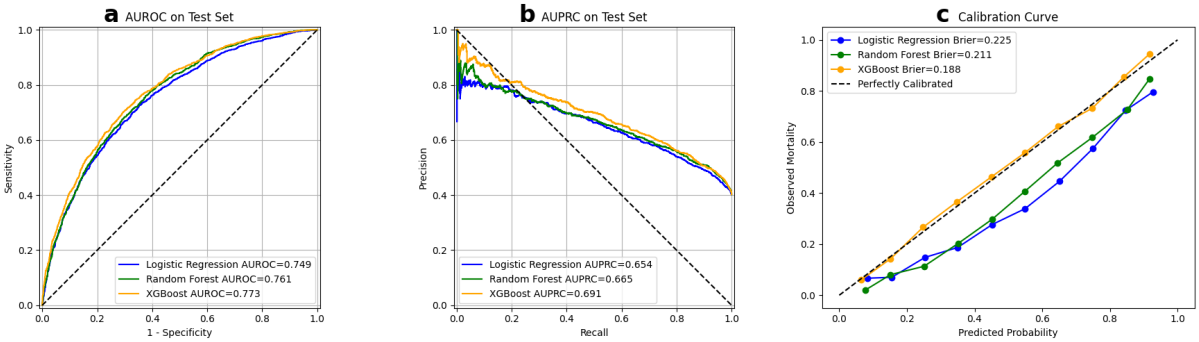


Figure 2: (a) The Area Under the Receiver Operating Characteristic (AUROC) curve shows the sensitivity vs. 1-specificity, comparing Logistic Regression, Random Forest, and XGBoost models on the test set. (b) The Area Under the Precision-Recall Curve (AUPRC) evaluates the trade-off between precision and recall for the same models. (c) Calibration plots illustrate the agreement between predicted probabilities and observed mortality, with the black 45-degree line representing perfect calibration where predicted and observed probabilities align.

Figure 2 shows that XGBoost outperformed other models in both discriminatory power and calibration ability, demonstrating superior performance in distinguishing between positive and negative samples. Moreover, the predictive probabilities of XGBoost aligned more closely with actual observations, indicating excellent calibration.

In comparison, random forest and logistic regression were slightly less effective in both calibration and discrimination. Further analysis of calibration ability revealed that XGBoost achieved a Brier score of 0.188, which was significantly lower than that of random forest (0.211) and logistic regression (0.225). This indicates that XGBoost had the smallest error between predicted probabilities and actual outcomes, achieving the best calibration.

Additionally, calibration curves demonstrated that across different predicted probability ranges, XGBoost's predictions were more consistent with actual outcomes compared to other models. These results highlight the high reliability of XGBoost in predicting long-term mortality probabilities. Therefore, XGBoost is the optimal model for predicting long-term mortality in ICU patients.

From Figure 3, the SHAP summary plot of the XGBoost model highlights significant differences in the contribution of various features to predicting ICU patient mortality. *age_years* (age) emerged as the most influential predictor, with SHAP values showing that older age significantly increases the predicted probability of mortality, aligning with the known health risks associated with aging. *admission_location_EMERGENCY ROOM ADMIT* (emergency room admission location) and *bloodureanitrogen* (blood urea nitrogen) were also identified as key predictors, suggesting elevated risks for patients admitted through the emergency room and those with higher blood urea nitrogen levels.

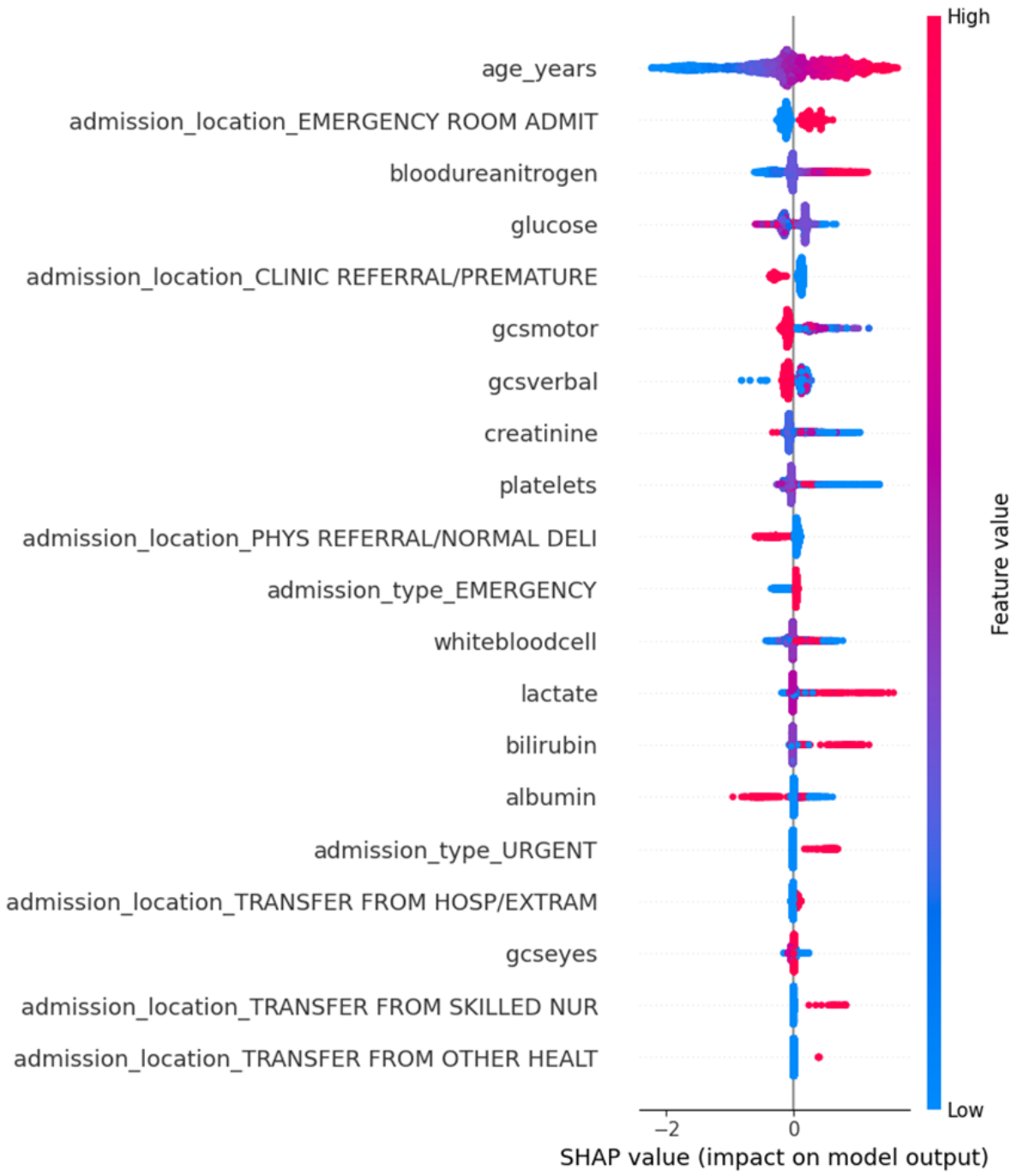


Figure 3: Feature importance ranking for predicting long-term mortality risk in ICU patients. This figure is based on the SHAP analysis of an XGBoost model, providing stability and interpretability ranking for 18 features. Each point represents the SHAP value for an individual sample, with colors transitioning from blue (low feature value) to red (high feature value). The greater the spread of points, the greater the impact of the feature on the model's predictions. Age (age_years), blood urea nitrogen (bloodureanitrogen), and lactate are the most influential predictors, contributing significantly to mortality risk.

## 3.7 Survival Model

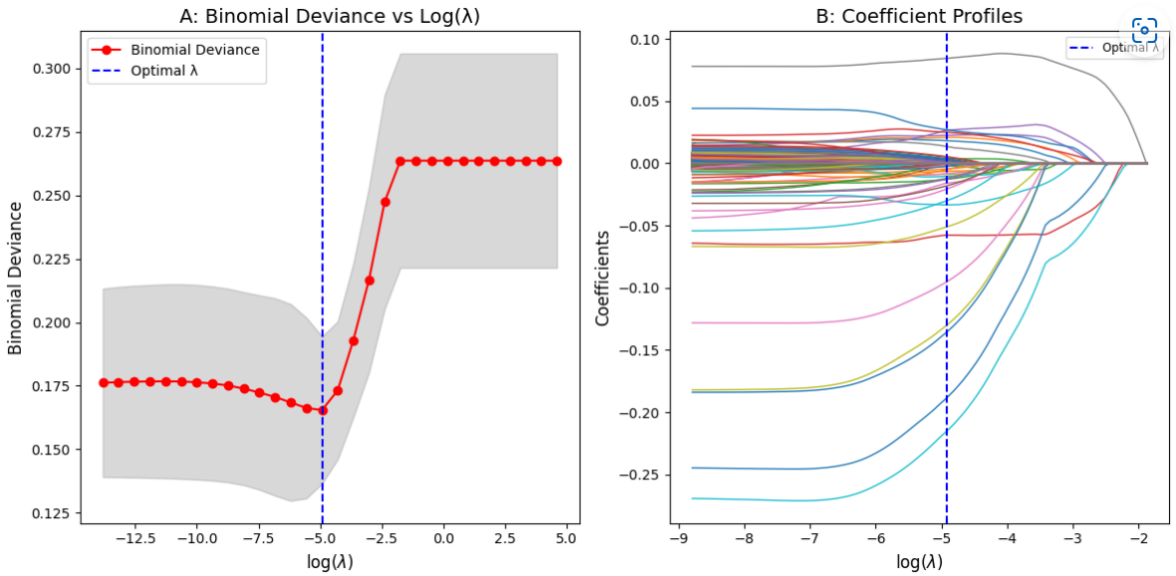


Figure 4: Feature selection using logistic regression with Least Absolute Shrinkage and Selection Operator (LASSO). (A) Regularization parameter ($\lambda$) selection using 5-fold cross-validation. The figure shows the binomial deviance curve as a function of $\log(\lambda)$, with the optimal $\lambda$ value determined at the minimum deviance point. The shaded area represents the standard error. The dashed line indicates the selected optimal $\lambda$ value, which was used for subsequent feature selection. (B) Coefficient path plot of the LASSO model. As $\lambda$ increases, non-essential features are shrunk to zero, retaining only the most predictive variables. The dashed line shows the optimal $\lambda$ value selected through cross-validation, ultimately resulting in 21 features with non-zero coefficients, which were used to generate the final model.

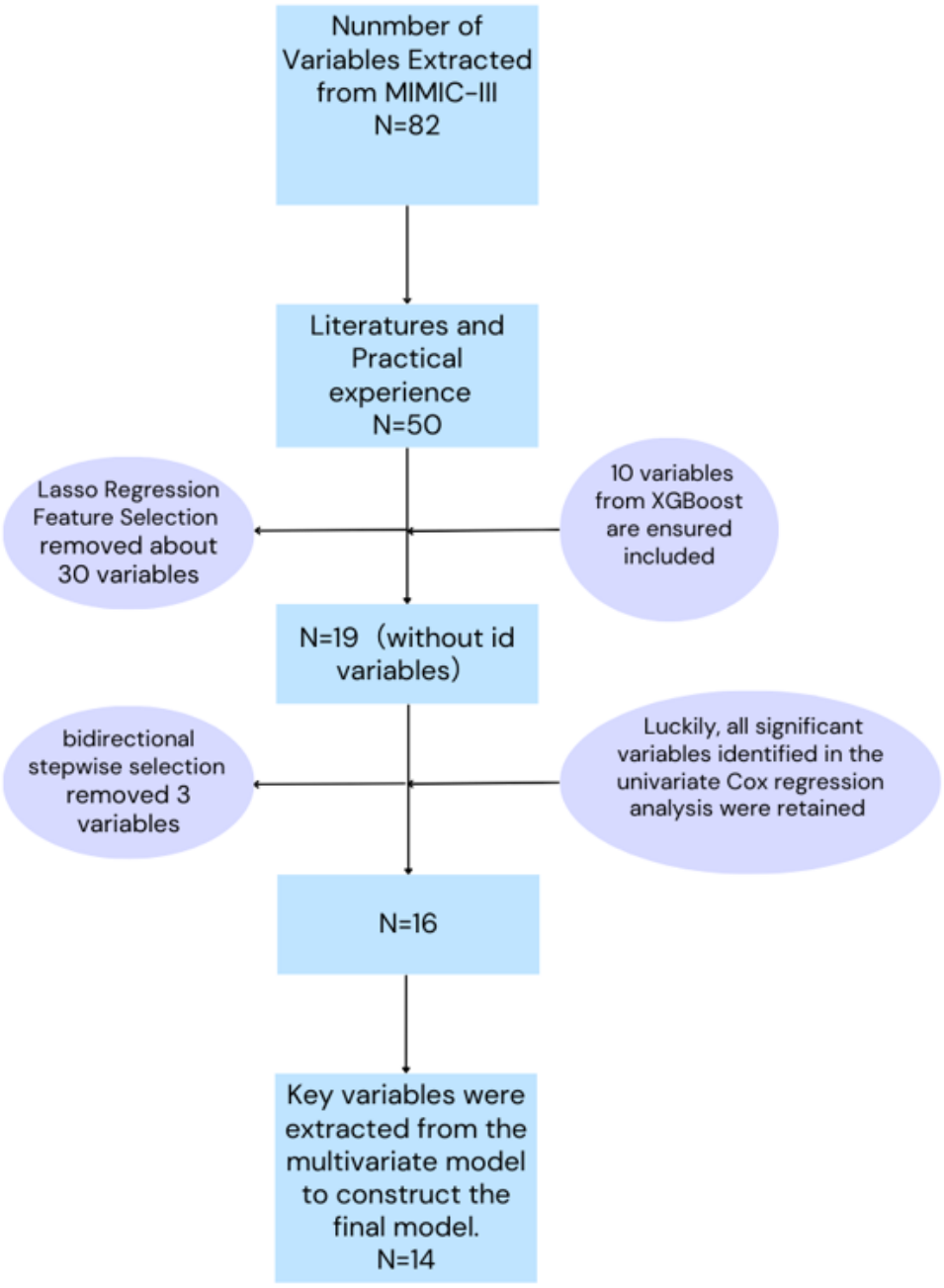


Figure 5: Stepwise Variable Selection Process for Constructing the Final Explanatory Model

After Lasso regression [10] (Figure 4), we ensured that the top ten most important candidate variables from the first research question (Figure 6) were included in the model as confounders. Subsequently, using bidirectional stepwise selection in the multivariate model, all significant variables identified in the univariate Cox regression analysis were retained to construct the Full Model. Finally, significant variables were selected from the multivariate model to form the Reduced Model, achieving a balance between interpretability and model complexity.

Table 4: Comparison of Three Cox Models (HR, 95% Confidence Intervals)

| **Variable** | **Model 1** | **Model 2** | **Model 3** |
|---|---|---|---|
| Intercept | Baseline Hazard | Baseline Hazard | Baseline Hazard |
| Weekend Admission | 0.99 | 0.97 | 0.97 |
| | (0.94, 1.04) | (0.92, 1.02) | (0.92, 1.02) |
| Positive Culture | – | 1.06* | 1.06* |
| | | (1.00, 1.12) | (1.00, 1.11) |
| GCS Score | – | 1.05** | 1.04** |
| | | (1.01, 1.08) | (1.01, 1.08) |
| GCS Motor Response | – | 0.92*** | 0.92*** |
| | | (0.88, 0.97) | (0.88, 0.97) |
| Lactate | – | 1.02* | 1.02* |
| | | (1.00, 1.04) | (1.00, 1.04) |
| Blood Urea Nitrogen | – | 1.00*** | 1.00*** |
| | | (1.00, 1.00) | (1.00, 1.00) |
| Hemoglobin | – | – | 0.98 |
| | | | (0.95, 1.01) |
| INR | – | 1.02* | 1.02* |
| | | (1.00, 1.04) | (1.00, 1.04) |
| Chloride | – | 0.99*** | 0.99*** |
| | | (0.98, 0.99) | (0.98, 0.99) |
| Hematocrit | – | 0.98*** | 0.99* |
| | | (0.98, 0.99) | (0.98, 1.00) |
| Age | – | 1.02*** | 1.02*** |
| | | (1.02, 1.03) | (1.02, 1.03) |
| Insurance (Medicare) | – | 1.17*** | 1.16*** |
| | | (1.10, 1.24) | (1.09, 1.24) |
| **Admission Location** | | | |
| Emergency Room Admit | – | – | 0.82 |
| | | | (0.66, 1.02) |
| Physician Referral | – | 0.61*** | 0.50*** |
| | | (0.56, 0.66) | (0.40, 0.63) |
| Clinic Referral | – | 0.75*** | 0.62*** |
| | | (0.71, 0.80) | (0.50, 0.77) |
| Transfer from Hospital | – | 0.72*** | 0.59*** |
| | | (0.68, 0.76) | (0.48, 0.74) |
| **Model Fit** | | | |
| AIC | 140,726 | 138,296 | 138,295 |
| BIC | 140,733 | 138,401 | 138,414 |
| **Tests** | **Model Comparison** | **Chisq (df)** | **p-value** |
| Likelihood Ratio Test | Model 2 vs Model 1 | 2,462.4 (16) | < 2.2e-16 |
| | Model 3 vs Model 2 | 5.1 (2) | 0.08 |

***Notes:*** *Values are Hazard Ratios (HR); *** $p < 0.001$ (highly significant); ** $p < 0.01$ (moderately significant); * $p < 0.05$ (significant at 5% level).*
*95% confidence intervals indicate the range of plausible parameter estimates.*
*Baseline hazard is part of the hazard function in Cox models and is not directly estimated.*

The reduced model was selected because it achieved an effective balance between performance and complexity. ANOVA and the Likelihood Ratio Test (LRT) showed that the log-likelihood values of the reduced model and the full model were very close (-69,133 vs. -69,131; $\chi^2 = 5.1$, degrees of freedom = 2, $p = 0.08$), indicating that the reduced model's fit was almost identical to the full model despite excluding non-significant variables.

Compared to the unadjusted model, the reduced model significantly improved model fit ($\chi^2 = 2,457$, degrees of freedom = 14, $p < 0.001$), with the log-likelihood increasing from -70,362 to -69,133. This highlights the importance of adjusting for confounding variables.

Model fit metrics further supported the selection of the reduced model. Its AIC was 138,296, nearly identical to the full model (138,295) but substantially lower than the unadjusted model (140,726). The BIC for the reduced model was 138,401, which was lower than both the full model (138,414) and the unadjusted model (140,733). Since BIC imposes a stronger penalty for model complexity, this indicates that the reduced model achieved a better balance between fit and simplicity.

By retaining significant predictors (e.g., positive culture, GCS score, and lactate) and the primary exposure variable (weekend admission status) while excluding variables with minimal contributions to model fit, the reduced model enhanced interpretability and statistical significance. Therefore, it represents an efficient and practical choice, avoiding the excessive complexity of the full model.

In the final Cox regression model, the hazard ratio (HR) for the primary variable *weekend admission* was 0.97 (95% CI: 0.92–1.02, $p > 0.05$), indicating no significant difference in long-term mortality risk between weekend and weekday admissions. However, the model did not fully satisfy the proportional hazards (PH) assumption (Table 5). Key variables like **GCS** and **lactate** violated this assumption[11], potentially affecting the model's accuracy and interpretability. The global test also indicated deviation from the PH assumption.

Therefore, it is recommended to perform stratified analyses for these variables or use a time-dependent Cox regression model to better capture their dynamic effects over time. When refining the model, it's important to balance complexity and interpretability to ensure it accurately reflects variable dynamics while remaining practical and usable.

# 4 Discussion and Conclusions

## 4.1 Study Highlights

This study integrated early clinical features of ICU patients and compared the performance of different models in mortality prediction, providing valuable reference for clinical decision support and optimization of medical resource allocation. The main advantages of the study include: (1) a comprehensive comparison of three models—logistic regression, random forest, and XGBoost—highlighting the superiority of XGBoost; (2) the proposal of a Reduced Model that balances performance and complexity, suitable for practical clin-

ical applications; and (3) an exploration of the impact of weekend admission on long-term mortality, offering data support for the allocation of medical resources.

## 4.2 Study Limitations

However, the study also has some limitations. First, the data was sourced solely from the MIMIC database, which may limit the generalizability of the results to other populations. Future studies should validate the findings using multicenter datasets. In addition, since the study only utilized data from the first 24 hours after ICU admission, it may be more suitable for predicting short-term mortality (e.g., 7-day, 30-day, or 90-day mortality) rather than long-term outcomes. Third, the Cox model did not fully satisfy the proportional hazards assumption, with some key variables violating this assumption. Future research could use stratified analysis or time-dependent models to improve prediction accuracy and interpretability. Finally, the study mainly focused on comparing weekend and weekday admissions without further refining the potential effects of specific admission times (e.g., morning, afternoon, or night), which may be closely related to medical resource allocation and patient outcomes.

## 4.3 Conclusions

This study demonstrates that XGBoost performs best among machine learning models, while the Reduced Model achieves more efficient interpretability and balance by adjusting for confounders. The analysis results indicate that there is no significant difference in long-term mortality between weekend and weekday admissions, providing supportive evidence for rational medical resource allocation. This study offers effective analytical tools for predicting long-term mortality and emphasizes the importance of integrating early ICU admission features for assessing mortality risk, thereby providing theoretical support and practical basis for optimizing ICU patient management and resource allocation.

# Appendix A: Graphs and Tables

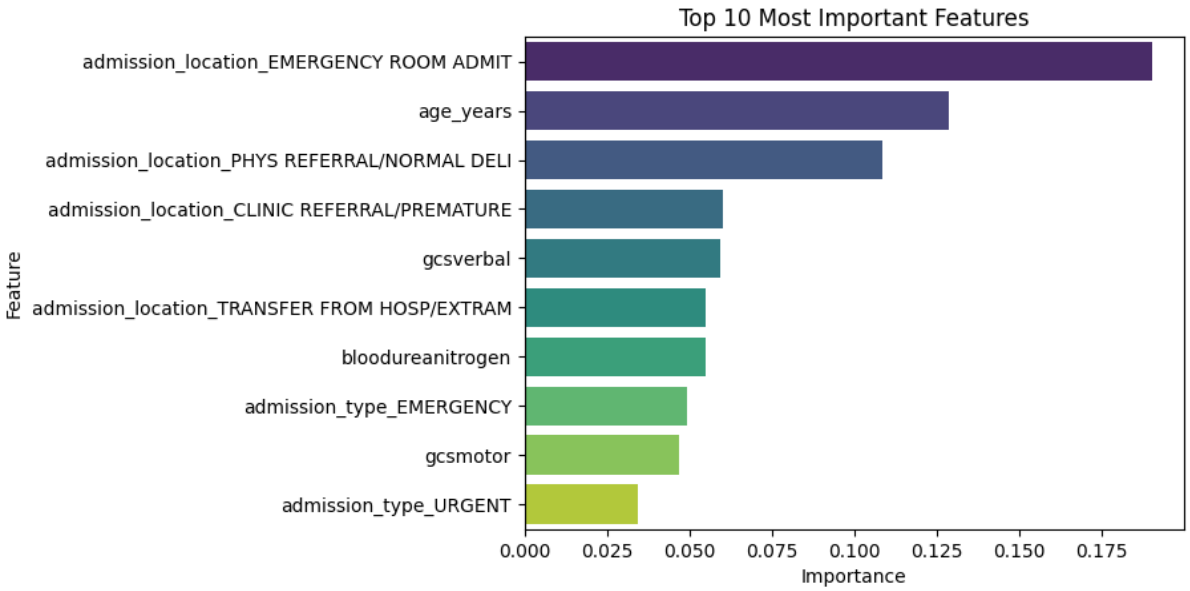


Figure 6: Top 10 Most Influential Features Identified by the XGBoost Model for ICU Mortality Prediction

| **Variable** | **Chi-sq** | **df** | **p** |
|---|---|---|---|
| positiveculture | 3.3905 | 1 | 0.0656 |
| gcs | 62.4989 | 1 | 2.7e-15 |
| gcsmotor | 8.7610 | 1 | 0.0031 |
| lactate | 10.6842 | 1 | 0.0011 |
| bloodureanitrogen | 2.8542 | 1 | 0.0911 |
| intnormalisedratio | 0.0146 | 1 | 0.9038 |
| chloride | 1.9497 | 1 | 0.1626 |
| hematocrit | 0.0493 | 1 | 0.8243 |
| age_years | 0.1723 | 1 | 0.6781 |
| insurance_Medicare | 0.0557 | 1 | 0.8134 |
| gcsverbal | 110.4847 | 1 | 2e-16 |
| admission_location_PHYS.REFERRAL.NORMAL.DELI | 4.8556 | 1 | 0.0276 |
| admission_location_CLINIC.REFERRAL.PREMATURE | 0.0869 | 1 | 0.7682 |
| admission_location_TRANSFER.FROM.HOSP.EXTRAM | 15.8305 | 1 | 6.9e-05 |
| is_weekend_admission | 4.4516 | 1 | 0.0349 |
| GLOBAL | 171.5329 | 15 | 2e-16 |

Table 5: Cox.zph Proportionality Assumption Test Results For The Final Model

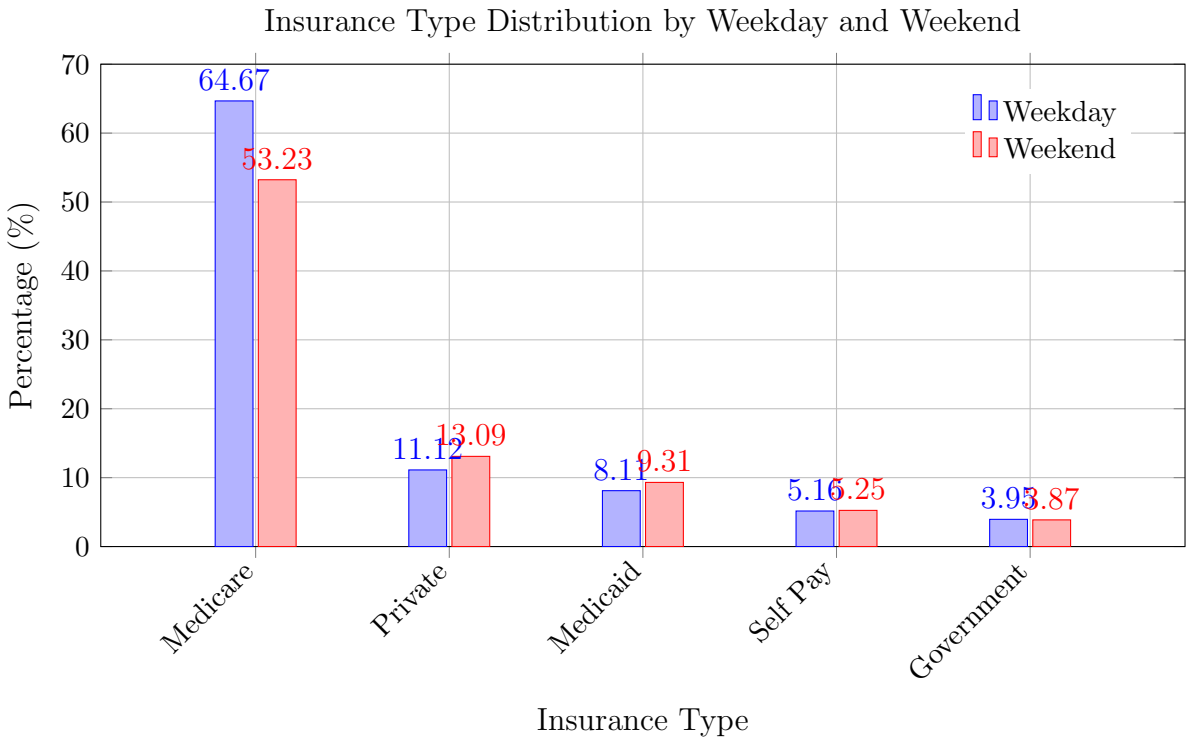


Figure 7: Insurance Type Distribution for Weekday and Weekend Admissions.

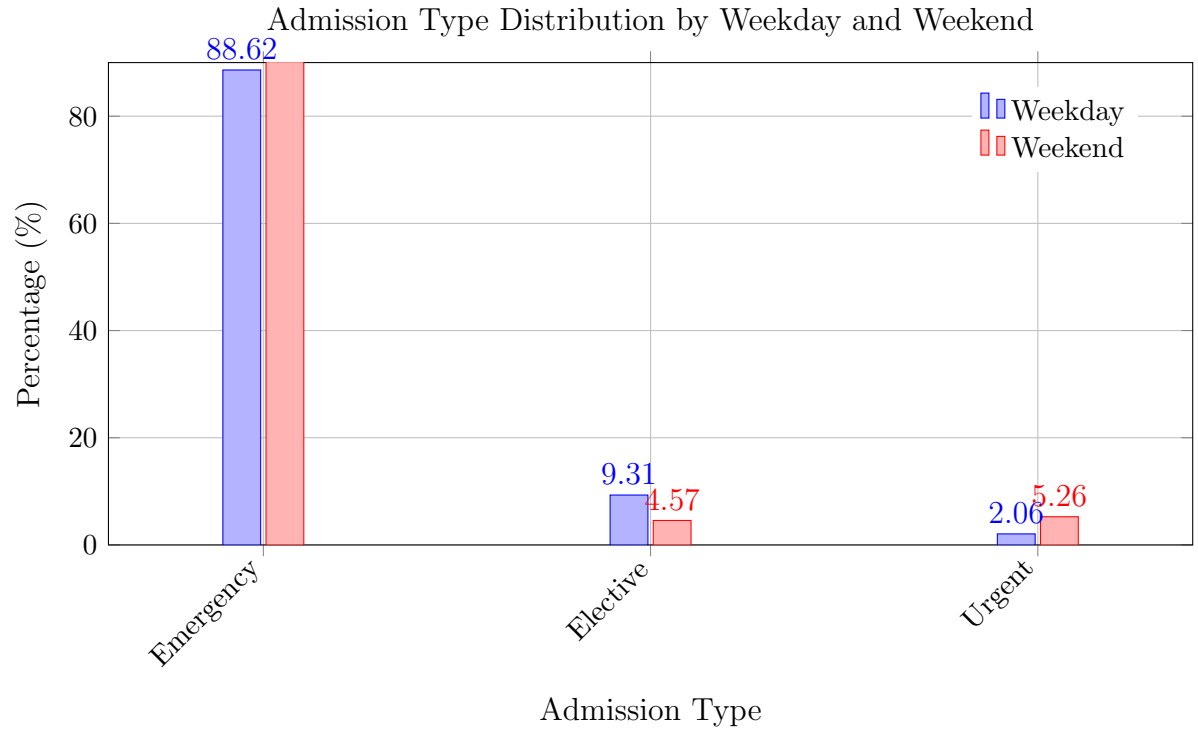


Figure 8: Admission Type Distribution for Weekday and Weekend Admissions.

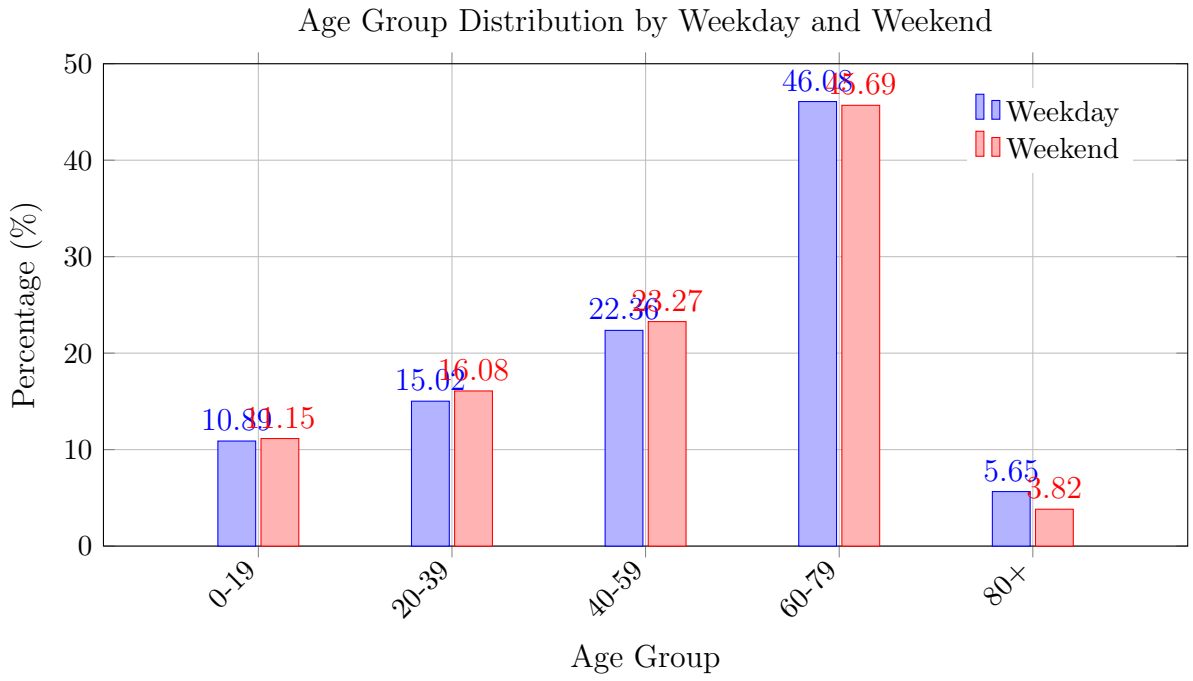


Figure 9: Age Group Distribution for Weekday and Weekend Admissions.

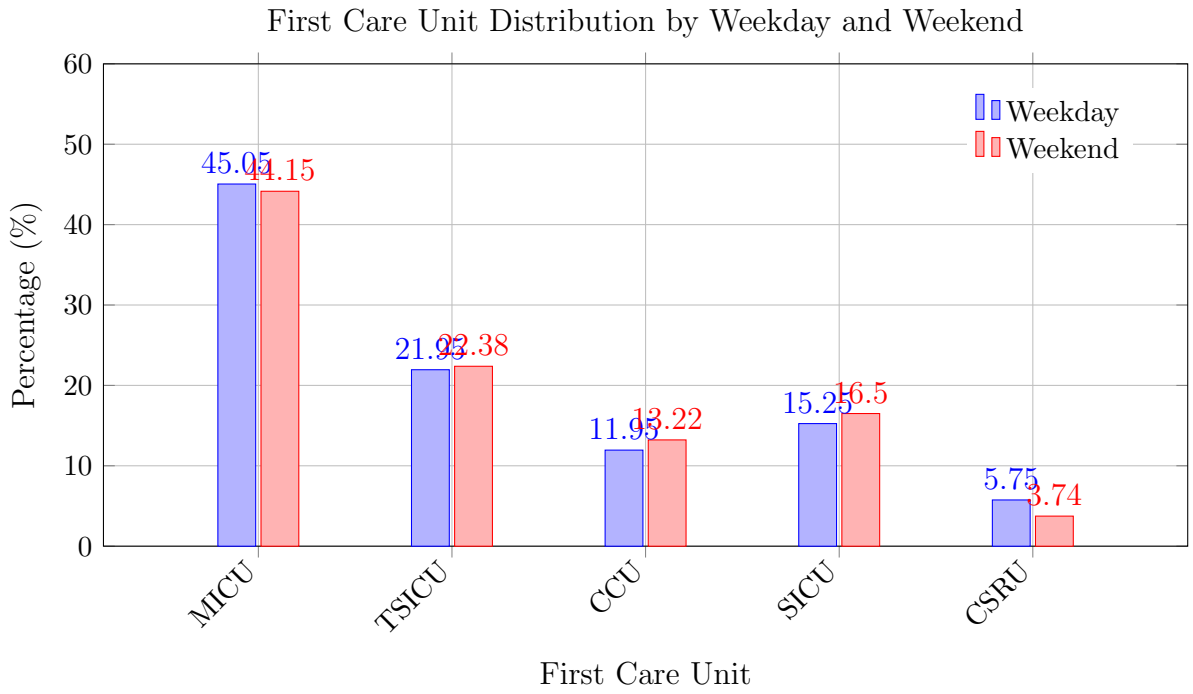


Figure 10: First Care Unit Distribution for Weekday and Weekend Admissions.

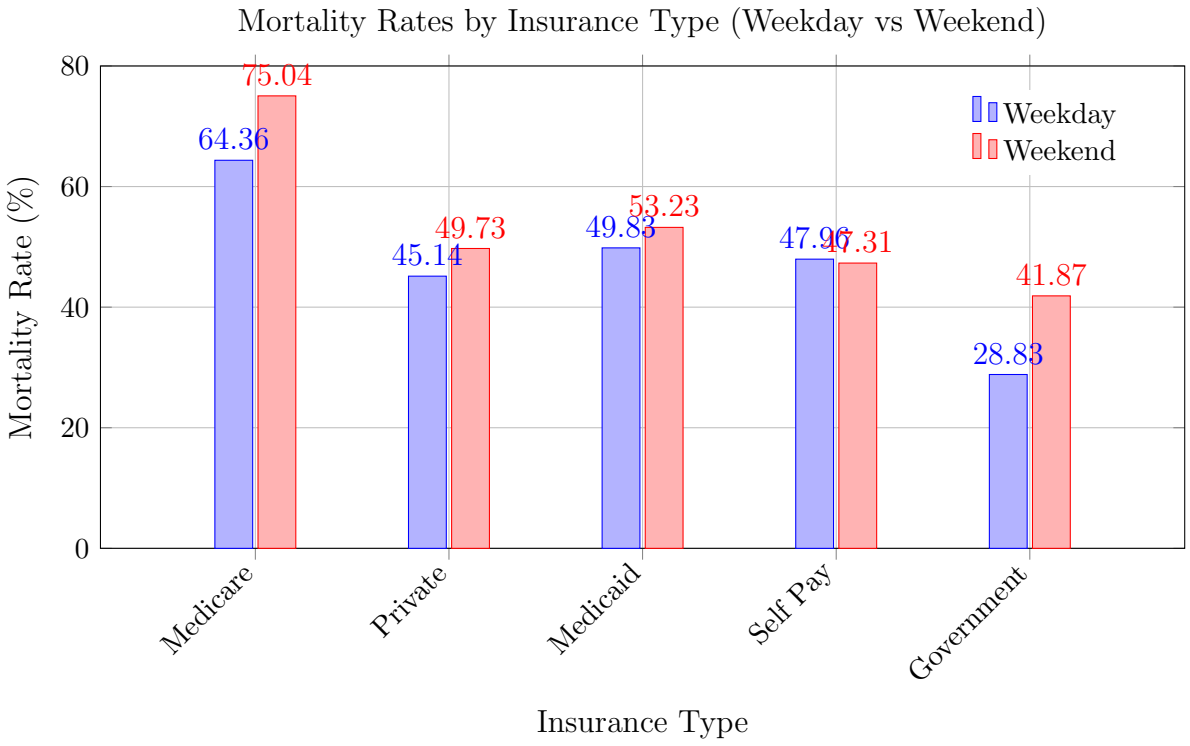


Figure 11: Mortality Rates by Insurance Type for Weekday and Weekend Admissions.

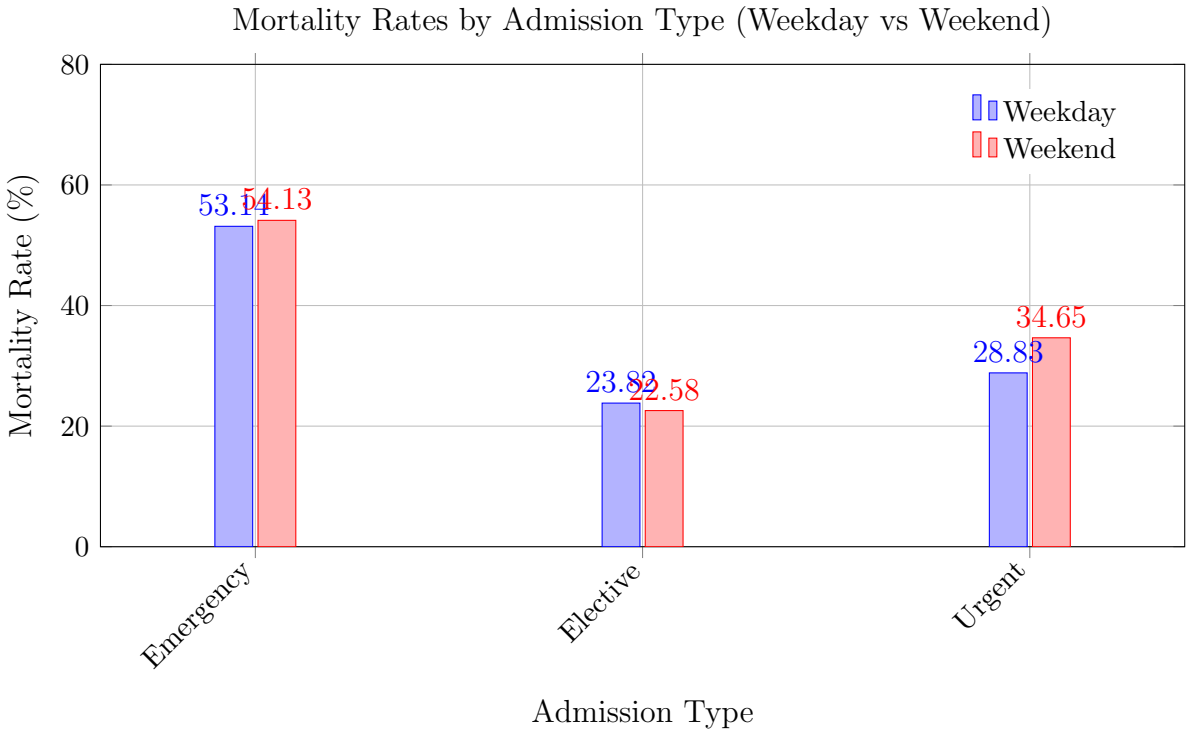


Figure 12: Mortality Rates by Admission Type for Weekday and Weekend Admissions.

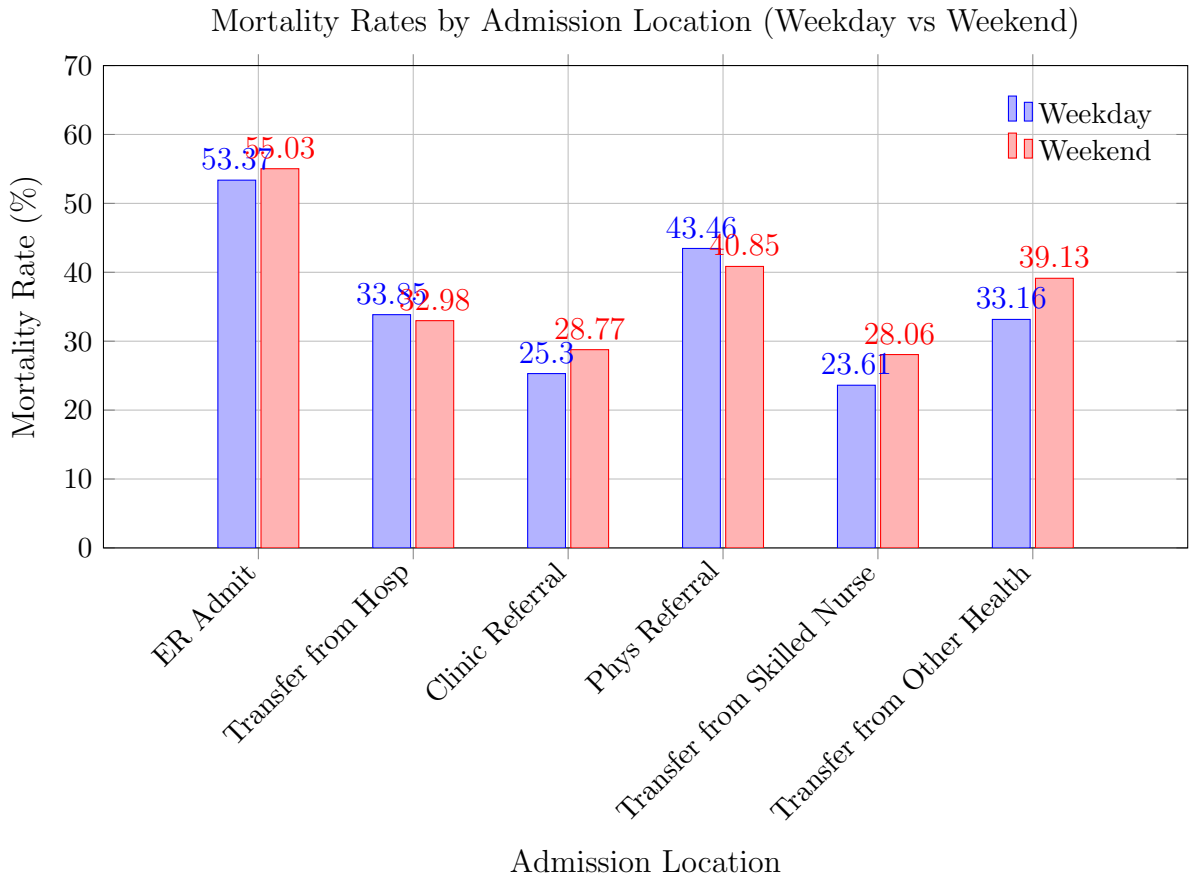


Figure 13: Mortality Rates by Admission Location for Weekday and Weekend Admissions. Data shows varied mortality rates by admission location, with weekend rates generally higher.

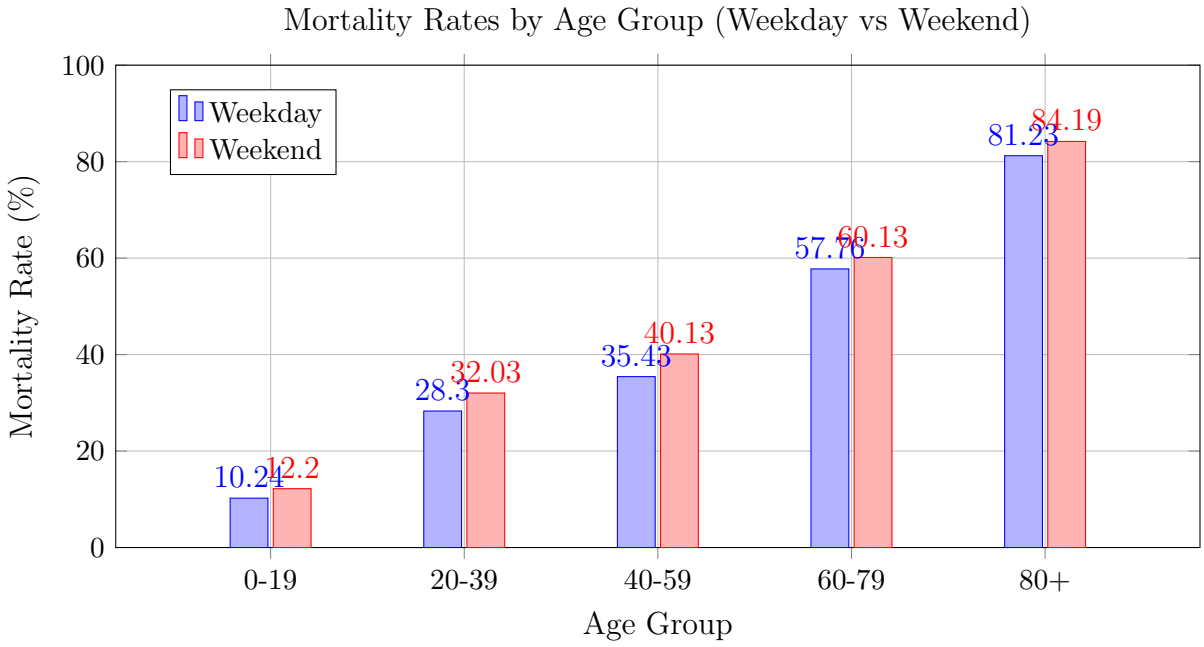


Figure 14: Mortality Rates by Age Group for Weekday and Weekend Admissions. Mortality rates increase significantly with age, with higher rates for older patients on weekends.

# Appendix B: Survival Models

## Model 1 (Baseline Model)

**Specification:**

$$\text{Surv}(\text{los\_x}, \text{expire\_flag}) \sim \text{is\_weekend\_admission}$$

**Interpretation:**

This univariate model includes only *is_weekend_admission* as the explanatory variable. It serves as a baseline to measure the impact of admission time (weekend vs. weekday) on survival time. Without adjusting for other covariates, it provides a limited understanding of the survival outcome.

Model 1 considers the sole effect of weekend admissions (*is_weekend_admission*) on survival time. The model's poor fit indicates that a single variable is insufficient to comprehensively explain survival time variability.

## Model 2

**Specification:**

$$\begin{aligned}\text{Surv}(\text{los\_x}, \text{expire\_flag}) \sim\ & \text{positiveculture} + \text{gcs} + \text{gcsmotor} + \text{lactate} \\ & + \text{bloodureanitrogen} + \text{intnormalisedratio} + \text{chloride} \\ & + \text{hematocrit} + \text{age\_years} + \text{insurance\_Medicare} \\ & + \text{gcsverbal} + \text{admission\_location\_PHYS.REFERRAL.NORMAL.DELI} \\ & + \text{admission\_location\_CLINIC.REFERRAL.PREMATURE} \\ & + \text{admission\_location\_TRANSFER.FROM.HOSP.EXTRAM}\end{aligned}$$

**Interpretation:**

This simplified multivariable model retains only the significant variables from Model 3. By removing non-significant variables, such as *is_weekend_admission* and *hemoglobin*, it reduces complexity while maintaining high explanatory power.

Model 2 simplifies Model 3 by focusing only on statistically significant predictors. This reduces the risk of overfitting and enhances interpretability while maintaining strong predictive performance.

## Model 3

**Specification:**

$$\begin{aligned}\text{Surv}(\text{los\_x}, \text{expire\_flag}) \sim\ & \text{is\_weekend\_admission} + \text{positiveculture} + \text{gcs} + \text{gcsmotor} \\ & + \text{lactate} + \text{bloodureanitrogen} + \text{hemoglobin} \\ & + \text{intnormalisedratio} + \text{chloride} + \text{hematocrit} \\ & + \text{age\_years} + \text{insurance\_Medicare} + \text{gcsverbal} \\ & + \text{admission\_location\_EMERGENCY.ROOM.ADMIT} \\ & + \text{admission\_location\_PHYS.REFERRAL.NORMAL.DELI} \\ & + \text{admission\_location\_CLINIC.REFERRAL.PREMATURE} \\ & + \text{admission\_location\_TRANSFER.FROM.HOSP.EXTRAM}\end{aligned}$$

**Interpretation:**

Building upon Model 1, this model incorporates additional clinical and demographic variables, such as *is_weekend_admission* and *hemoglobin*. By including a broader range of factors, Model 3 captures more patient-specific differences and enhances the model's predictive ability.

Model 3 improves upon Model 1 by significantly expanding its scope, incorporating variables like *bloodureanitrogen* and *admission location*, which are crucial for understanding survival outcomes. This results in better model fit and explanatory power.

# References


[1] Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L.-W. H., Feng, M., Ghassemi, M., *et al.* (2016). MIMIC-III, a freely accessible critical care database. *Scientific Data*, *3*, Article 160035. https://doi.org/10.1038/sdata.2016.35

[2] Kuo, Y., Barbaro, R. P., Wallace, D. J., Bartlett, R. H., Davis, M. M., Kahn, J. M., & Iwashyna, T. J. (2019). Examining the weekend effect across ICU performance metrics. *Critical Care*, *23*(1), 1–9. https://doi.org/10.1186/s13054-019-2479-5

[3] Johnson, A. E. W., Pollard, T. J., Shen, L., Lehman, L. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., & Mark, R. G. (2016). MIMIC-III, a freely accessible critical care database. *Scientific Data*, *3*, 160035. https://doi.org/10.1038/sdata.2016.35

[4] Hanley, J. A., & McNeil, B. J. (1982). The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology*, *143*(1), 29–36. https://doi.org/10.1148/radiology.143.1.7063747

[5] Powers, D. M. W. (2011). Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*, *2*(1), 37–63. https://arxiv.org/abs/2010.16061

[6] Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, *10*(3), e0118432. https://doi.org/10.1371/journal.pone.0118432

[7] Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830. http://jmlr.org/papers/v12/pedregosa11a.html

[8] Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). ACM. https://doi.org/10.1145/2939672.2939785

[9] Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems* (Vol. 30). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

[10] Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288. https://www.jstor.org/stable/2346178

[11] Cox, D. R. (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, *34*(2), 187–220. https://www.jstor.org/stable/2985181