## -1.Introduction

### 1.1 Background and Motivation

Intensive Care Units (ICUs) play a critical role in healthcare, handling patients with life-threatening conditions that demand close monitoring and advanced medical interventions. ICU mortality is a key indicator of healthcare quality, reflecting the effectiveness of care strategies, resource utilization, and clinical expertise. Factors influencing ICU mortality include patient demographics, comorbidities, and the availability of healthcare resources. Identifying and predicting factors associated with ICU mortality can allow assess of severity, guide interventions, prioritize care, and allocate resources effectively to improve patient survival [1, 2]. Studies leveraging machine learning (ML) and traditional statistical models have demonstrated significant potential in predicting ICU outcomes using physiological, demographic, and clinical variables collected within the first day of admission [3, 4]. For example, conditional medical generative adversarial networks (cMedGAN) and XGBoost-based models have been applied to datasets like MIMIC-III to predict mortality with high accuracy [5-7]. However, the reliability of these models depends on data completeness, model choice, and parameter tuning. Importantly, the first 24 hours serve as a snapshot of the patient’s condition, providing a foundation for predicting outcomes while leaving room for continuous assessment and updates.

Among these factors, the "weekend effect" has emerged as a significant area of investigation. The term refers to the observed differences in patient outcomes based on the timing of hospital admission, particularly during weekends. Studies have consistently shown increased mortality rates for weekend admissions across a variety of conditions, including emergency surgery [8] pulmonary embolism [9] and ischemic stroke [10]. Reduced staff availability, limited access to specialists, and constrained resources during weekends are among the hypothesized contributors to this phenomenon.

The "weekend effect" in ICU settings warrants deeper exploration, given the critical nature of care provided. Prior research highlights disparities in outcomes for stroke patients and ICU patients with complex conditions [9, 10]. Understanding these disparities is vital for implementing system-wide changes aimed at ensuring equitable and effective care throughout the week.

### 1.2 Role of the MIMIC-III Dataset

The Medical Information Mart for Intensive Care (MIMIC-III) database provides an extensive repository of de-identified clinical data from over 60,000 ICU admissions at the Beth Israel Deaconess Medical Center between 2001 and 2012. Its rich dataset includes demographics, physiological parameters, medication records, and outcomes, making it a valuable resource for studying ICU mortality and other critical care outcomes.

This study leverages the MIMIC-III dataset to investigate the weekend effect in ICU mortality, focusing on differences in mortality risk between weekend and weekday admissions. Through advanced statistical modelling and machine learning approaches, this study aims to contribute to the body of evidence on the weekend effect and inform interventions to improve patient outcomes in critical care settings.

## 2. Methods

### 2.1 Data Understanding and Preprocessing

The analysis was conducted using the MIMIC-III (Medical Information Mart for Intensive Care) dataset, a comprehensive and publicly available database containing detailed clinical data from ICU admissions. The dataset includes demographics, ICU admission details, clinical measurements, interventions, and outcomes for over 40,000 ICU patients. This rich dataset enables robust investigations into ICU mortality and related factors.

The initial dataset, ‘master\_data’, was constructed by systematically refining and merging key tables from the MIMIC-III database. To ensure data quality and relevance, several inclusion and exclusion criteria were applied. Patients aged 18–89 years were included to comply with HIPAA regulations while focusing on adult ICU populations. Only the first ICU admission lasting at least 24 hours was considered to ensure adequate observation data and independence between samples. Duplicates based on ICU stay identifiers (icustay\_id) were removed, and records missing critical variables such as expire\_flag or intime were excluded. Records with missing critical variables, such as demographic information or admission times, were excluded from the analysis.

Key tables were merged to enrich the dataset, a range of variables extracted from the `patients`, `admissions`, `icustays`, and `chartevents` tables. Demographic variables included age, gender, and ethnicity, which were crucial for understanding population characteristics. ICU admission details, such as length of stay, the first and last care units, admission time, and ICU stay time, were included to capture the context and duration of care. Patient demographics, including gender and date of birth, were integrated from the `patients` table. Clinical measurements encompassed vital signs, such as heart rate, respiratory rate, and blood pressure, as well as laboratory results like creatinine, lactate, and blood urea nitrogen, which are essential indicators of patient condition. Additional clinical insights were included from weight measurements and microbiology data, such as the presence of multi-drug resistants organisms (MDROs). A binary indicator, `is\_weekend\_admission`, was derived from admission timestamps to facilitate the weekend effect analysis. Outcome variables, including in-hospital mortality and discharge details, were used as primary endpoints for analysis.

The data preparation process involved merging key tables using unique patient (`subject\_id`) and admission (`hadm\_id`) identifiers, ensuring a comprehensive dataset for each ICU stay. Patients with ICU stays lasting at least 24 hours were selected using the length of stay column (`los`) from the `icustays` table. Demographic data were obtained from the `patients` table, while the `admissions` table provided additional details on admission and discharge times, insurance status, and marital status. ICU-specific information, including unit type and duration, was extracted from the `icustays` table, and time-series data on vital signs and laboratory results were derived from the `chartevents` table. Aggregated statistical summaries, such as mean, maximum, minimum, and trends, were calculated from the first 24 hours of time-series data to capture the severity of patient conditions.

Missing data were addressed using imputation strategies. Numeric columns were imputed with mean values, while categorical variables were imputed using the mode.

### 2.2 Feature Engineering

Feature engineering focused on deriving meaningful predictors from the available data. Aggregated features were created by calculating statistical summaries, such as mean, maximum, and trends, for vital signs and lab results over the first 24 hours of ICU stay. These summaries provided a concise representation of the patients’ physiological status during this critical period.

Categorical indicators, such as ICU type and weekend admission status, were also included. Handling missing data was approached differently for two analysis styles. For Style A, variables with more than 25% missingness were excluded, ensuring a high-quality dataset. For Style B, a more lenient approach was adopted, retaining rows with sufficient data coverage (≥75% of variables), and imputing missing values using statistical methods. These approaches provided distinct datasets for comparative analysis, facilitating robust model evaluations.

The primary outcome variable was mortality, represented by the expire\_flag. This binary outcome served as the dependent variable in predictive modeling and weekend effect analyses.

### 2.3 Predictive Modeling

Several machine learning models were employed to predict ICU mortality based on data from the first 24 hours of ICU stay. Logistic regression was used as baseline predictive model for its interpretability, while Random Forest and XGBoost were included to explore non-linear relationships and interaction effects among variables and compare interpretability and predictive performance.

Hyperparameter tuning was conducted using GridSearchCV and RandomSearchCV to optimize model performance. Models were evaluated using a range of metrics, including ROC-AUC, F1 score, precision, and recall, ensuring a comprehensive assessment of predictive capabilities. Data were split into training (80%) and validation (20%) sets, with stratification to maintain class balance. The training process included sensitivity analyses to evaluate robustness across different feature sets and data imputation strategies.

### 2.4 Statistical Analysis for weekend effect

To investigate the weekend effect on ICU mortality, logistic regression models were used to evaluate the association between weekend admission and mortality. These models adjusted for covariates, such as patient demographics and clinical measurements. Propensity score matching (PSM) was employed to balance baseline covariates between weekend and weekday admissions, reducing potential confounding. The propensity scores were estimated using logistic regression and matched using nearest neighbor matching. After matching, McNemar’s test was conducted to assess paired differences in mortality rates, providing a robust comparison of outcomes between the two groups.

The data preprocessing, exploratory data analysis, and feature engineering steps were primarily performed in R (documented in the Rmd file), while advanced modeling and hyperparameter tuning were conducted in Python (Jupyter Notebook). This dual-software approach allowed leveraging the strengths of each platform for different stages of the analysis.

## 3. Results and Discussion

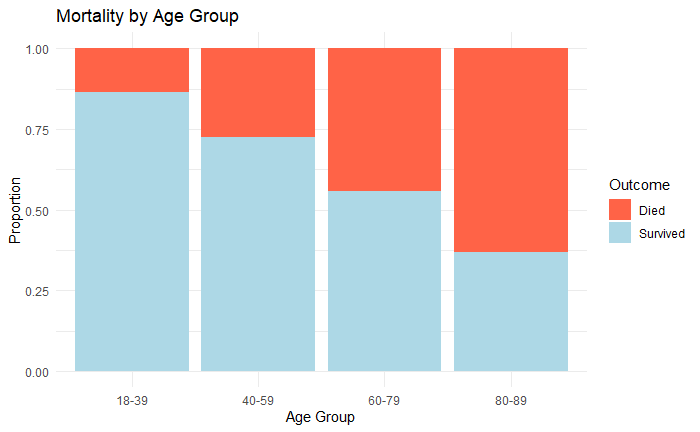
### 3.1 Exploratory Data Analysis (EDA):

#### Summary of dataset structure and missingness

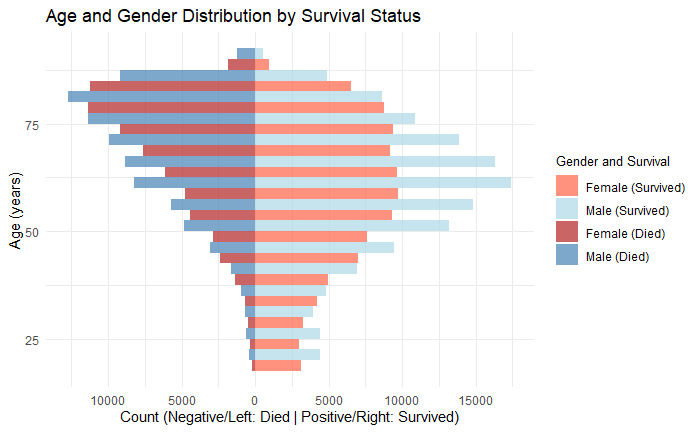
The dataset extracted from MIMIC-III incorporates a broad range of demographic, clinical, and ICU admission-related variables. The primary dataset the `master\_data` used for time-series analysis had 375,552 rows and 83 columns.

#### Demographic distribution and ICU characteristics

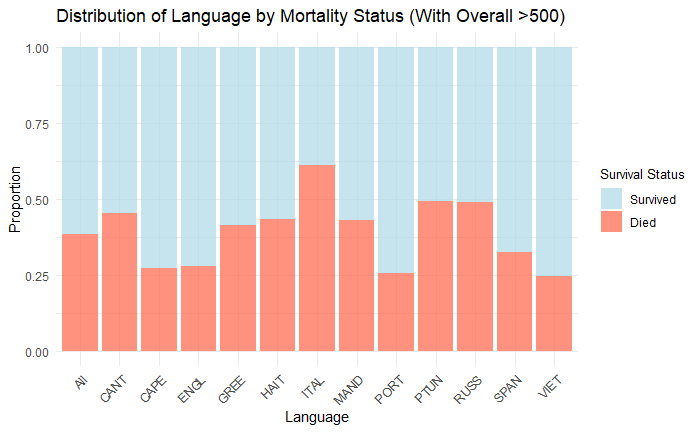
Age significantly influenced mortality risk. Patients aged 18–39 exhibited minimal mortality rates, while those aged 80–89 had the highest mortality proportion, as shown in the "Mortality by Age Group" chart. The age and gender distribution of ICU patients reveals notable patterns in mortality. Mortality is significantly higher



among older age groups for both genders, with non-survivor bars dominating in the higher age categories. Males tend to have a slightly higher proportion of survivors in the younger age groups compared to females, while the overlap in the middle age range suggests similar mortality rates between genders for these age groups.



Statistical tests further validate these observations. A Welch Two-Sample t-test indicates a statistically significant difference in the mean age of survivors (57.95 years) compared to non-survivors (68.57 years), with a p-value < 2.2e-16. This finding underscores the critical role of age in predicting mortality outcomes. Additionally, males exhibit slightly higher mortality rates, comprising 59.02% of non-survivors compared to 40.98% for females. A Chi-squared test confirms a significant association between gender and mortality (p-value < 2.2e-16). While the association is statistically significant, further analysis is required to determine the clinical relevance of gender as a predictor of ICU mortality. These findings highlight age and gender as important factors to consider in ICU outcome prediction models.



Among languages with over 500 occurrences, notable variability in mortality rates exists. For instance, Italian-speaking patients exhibit the highest mortality proportion, with nearly 50% of the cohort labeled as "Died." Other languages, such as English and Mandarin, show relatively lower mortality proportions, indicating better survival outcomes. Language differences may reflect underlying disparities, such as access to healthcare resources, communication barriers, or cultural factors influencing care-seeking behavior.

Ethnic groups such as Black/African American and Hispanic/Latino show slightly higher mortality proportions compared to White and Asian groups. Disparities in mortality among ethnic groups may stem from structural inequities, differences in comorbidity burdens, or access to high-quality care. The "Unknown" ethnicity group also displays a relatively higher mortality rate, potentially highlighting challenges in data completeness or underrepresentation in healthcare settings. Addressing these disparities may require incorporating ethnicity as a covariate in predictive modeling and performing subgroup-specific analyses.

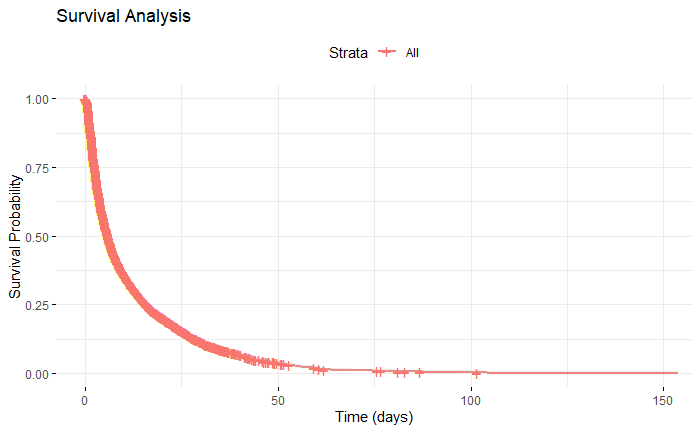
A graph of ethnicity status

Description automatically generated

A graph of a number of bars

Description automatically generated with medium confidence

The distribution of ICU types reveals distinct patterns in mortality rates, with significant variation across different units. The Coronary Care Unit (CCU) exhibits a moderate mortality rate, reflecting its specialized focus on cardiac conditions. The Cardiac Surgery Recovery Unit (CSRU) shows the lowest mortality rate, likely due to its controlled environment for post-surgical recovery. In contrast, the Medical Intensive Care Unit (MICU) has the highest mortality rate, consistent with its role in treating patients with severe and complex medical conditions. The Surgical Intensive Care Unit (SICU) has a moderate mortality rate, likely influenced by the challenges of post-surgical care, while the Trauma/Surgical Intensive Care Unit (TSICU) reports a lower mortality rate, highlighting its effectiveness in managing trauma and surgical emergencies. A chi-squared test confirms a strong association between ICU type and mortality (p-value < 2.2e-16), emphasizing the critical role ICU specialization plays in patient outcomes. Further analysis of patient characteristics within each ICU type, such as age and comorbidities, could provide deeper insights into these observed differences.



The survival probability plot indicated a steep decline in survival during the initial days of ICU stay, followed by a plateau. This suggests the critical nature of the first 24–48 hours for patient outcomes. Patients who survived beyond the initial critical period were more likely to recover or stabilize, as reflected in the leveling off of survival probability. These findings underscore the importance of early interventions and the inclusion of time-sensitive variables in predictive models.

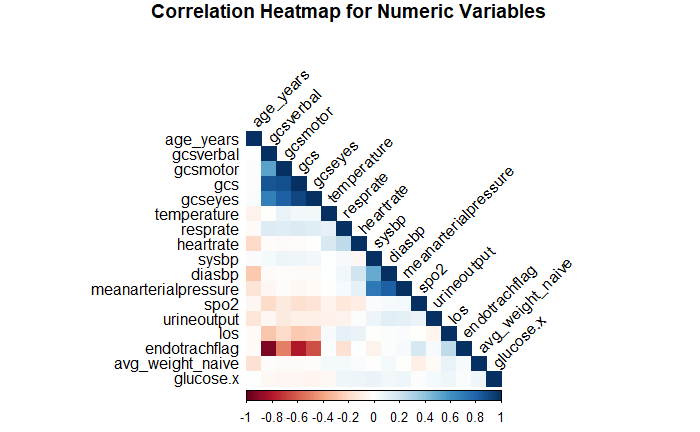
#### Key variable distributions and missing percentage

The dataset comprises a wide range of variables extracted from MIMIC-III tables, representing demographic data, ICU-specific information, and clinical measurements. The variables have been categorized based on their source tables and missing data percentages, enabling a comprehensive understanding of the dataset quality. Variables with ≤ 25% missingness were directly used in modelling and analysis. For variables with moderate missingness (25% to ≤ 75%), imputation strategies included mean substitution for numeric fields and mode substitution for categorical ones. Variables with > 75% missingness were excluded from most analyses to avoid introducing significant biases.



Style A and Style B data processing strategies are applied. Style A approach focuses on simplifying the dataset by removing variables with significant missingness (>25%) and imputing the remaining missing values in numeric columns. But important columns with >25% missingness may excluded. Style B approach prioritizes retaining variables, even those with high missingness, by filtering rows that meet a completeness threshold (e.g., 75% of columns have non-missing values). However, due to row filtering results in a smaller dataset size. In addition, potential bias introduced by selective row retention, as the retained rows may not represent the overall population. Both approaches use `SimpleImputer` for handling missing values: Numeric columns are imputed with the mean. Categorical columns are imputed with the most frequent value (mode). Boolean columns are converted to integers for imputation compatibility and reverted back to boolean format post-imputation.

#### Correlation Analysis



The correlation analysis highlights relationships among numeric variables in the dataset, as visualized in the heatmap. Notable correlations include the expected strong associations between the Glasgow Coma Scale (GCS) and its subcomponents (`gcseyes`, `gcsmotor`, `gcsverbal`), as well as among blood pressure metrics (`sysbp`, `diasbp`, `meanarterialpressure`). These align with clinical expectations, given their interdependent physiological measurements. The analysis also reveals multicollinearity in variables such as `gcs` and its subcomponents, which may necessitate removing or consolidating highly correlated variables during regression or model training to improve model stability and interpretability. Additionally, the negative correlation between `gcsverbal` and `endotrachflag` aligns with clinical insights, as intubation (indicated by `endotrachflag`) often impairs a patient’s verbal response (`gcsverbal`). These insights guide feature selection and engineering in subsequent predictive modeling steps.

### 3.2 Predictive Modeling Results

#### Performance of Baseline and Tuned Models



The performance of baseline and tuned models was evaluated across Style A and Style B datasets using key metrics such as ROC-AUC, precision, recall, F1-score, and accuracy. For Style A, Random Forest emerged as the top-performing model with a nearly perfect ROC-AUC score of 0.998, demonstrating exceptional predictive capabilities. Tuned XGBoost closely followed with a ROC-AUC of 0.988, offering a robust balance between precision and recall. Artificial Neural Networks (ANN) achieved moderate performance with a ROC-AUC of 0.781, while Logistic Regression, despite being interpretable, lagged significantly behind with the lowest ROC-AUC of 0.753.

For Style B, XGBoost demonstrated the best performance with a ROC-AUC of 0.905, followed by ANN with 0.810. Random Forest showed competitive results with a ROC-AUC of 0.853, outperforming Logistic Regression, which achieved a ROC-AUC of 0.785. While Random Forest exhibited strong predictive power for Style A's larger dataset, XGBoost's performance in Style B highlights its robustness in handling datasets with reduced sample sizes and higher clinical feature inclusion. Logistic Regression's simplicity and interpretability remain its primary strengths, albeit at the cost of predictive accuracy.

#### Performance of Baseline and Tuned

The ROC curves for each model provided a visual comparison of their performance across datasets. For Style A, the steep curve of Random Forest underscored its superior predictive capability, followed closely by XGBoost. ANN exhibited a moderate curve, while Logistic Regression showed the least favorable performance. In Style B, XGBoost's ROC curve led the models, emphasizing its adaptability to the dataset's complexities. Confusion matrices further highlighted the balance between sensitivity and specificity for each model, particularly for the tuned Random Forest and XGBoost.

A graph of a curve

Description automatically generated

A graph of different colored lines

Description automatically generated

#### Feature Importance Analysis

Feature importance analysis revealed key drivers of predictions for Random Forest and XGBoost. In Style A, the most critical features identified by XGBoost included `age\_years`, `last\_wardid`, and `sysbp\_normal`, emphasizing their clinical relevance in mortality prediction. Similarly, Random Forest confirmed `age\_years` as the top predictor, reflecting the strong association between age and ICU mortality. In Style B, XGBoost identified additional clinically relevant variables such as `spo2\_normal`, `creatinine`, and `bloodureanitrogen`, showcasing the importance of lab results and vital signs in predicting patient outcomes. These findings highlight the models' ability to capture and interpret meaningful clinical patterns, contributing to improved decision-making in critical care settings.

A graph with blue bars

Description automatically generated

A graph of a number of blue bars

Description automatically generated with medium confidence

### 3.3 Weekend Effect Analysis:

#### Mortality Rates Across Weekend and Weekday Admissions

The weekend effect analysis aimed to investigate differences in ICU mortality rates between weekend and weekday admissions. Mortality rates were calculated for both weekend and weekday admissions in Style A and Style B datasets. In Style A, the weekend mortality rate was 41.17%, higher than the weekday mortality rate of 37.72%. Similarly, in Style B, the weekend mortality rate was 43.52%, compared to 39.32% on weekdays. These observations consistently suggest an elevated mortality risk associated with weekend admissions, indicative of a potential weekend effect.

#### Statistical Validation with Chi-Square Tests

To statistically validate these findings, Chi-Square tests were performed. In Style A, the test yielded a highly significant result (Chi-Square Statistic is 310.65, P-value is 1.57e-69), while for Style B, the result was also significant Chi-Square Statistic is 25.93, P-value is 3.57e-7). These results confirm a significant association between admission timing and mortality, further reinforcing the evidence of a weekend effect.

#### Propensity Score Matching for Confounding Control

Propensity score matching (PSM) was employed to address potential confounding variables and ensure a more balanced comparison between weekend and weekday admissions. Logistic regression models were used to calculate propensity scores based on selected covariates, and nearest neighbor matching was performed. Post-matching analysis showed that mortality remained higher for weekend admissions (Style A: 41.17% vs. 37.72%; Style B: 43.55% vs. 39.18%). This consistent trend highlights the robustness of the weekend effect observation.

#### Covariate Balance Assessment

The balance of covariates between matched groups was evaluated using standardized mean differences (SMD). For Style A, most covariates, such as `avg\_weight\_naive` and `diasbp`, were well-balanced, with SMD values below the 0.1 threshold. However, variables like `last\_wardid` and `age\_years` showed slight imbalances. Similarly, in Style B, covariates such as `bloodureanitrogen` and `hemoglobin` were balanced, but `age\_years` and `hematocrit` exhibited residual imbalances. These results underline the effectiveness of PSM in reducing confounding, while also emphasizing the need for careful interpretation of results given residual imbalances in some covariates.

A graph of blue rectangular bars

Description automatically generated with medium confidence

A graph of blue rectangular bars

Description automatically generated with medium confidence

#### Conclusions on the Weekend Effect

Overall, the findings support the existence of a weekend effect in ICU admissions, with higher mortality rates observed during weekends compared to weekdays. This effect persists even after addressing potential confounders, suggesting systemic differences in care delivery or patient characteristics between weekend and weekday admissions. Further investigations are warranted to explore the underlying causes of this phenomenon and develop targeted interventions to mitigate it.

## 4. Conclusion

This study utilized the MIMIC-III dataset to investigate ICU mortality and explore the presence of the "weekend effect" through a combination of advanced statistical analyses and machine learning models. The results revealed several critical findings. First, age and ICU admission characteristics significantly influence mortality, with older patients and those admitted to specific ICU types (e.g., MICU) facing higher risks. Second, Style A and Style B data imputation approaches yielded distinct insights into the trade-offs between dataset size and variable inclusion, with Style A excelling in predictive power due to its larger dataset and Style B capturing nuanced clinical features. Third, Random Forest and XGBoost consistently emerged as the top-performing models, with XGBoost demonstrating superior robustness in datasets with higher missingness.

The weekend effect analysis provided strong evidence of increased mortality for weekend ICU admissions, even after controlling for confounders using propensity score matching. This underscores systemic disparities in resource availability, care quality, or patient acuity during weekends. Residual imbalances in covariates such as age and hematocrit highlight the challenges of achieving complete control over confounding variables.

Future research should focus on validating these findings in other datasets and incorporating additional variables, such as socioeconomic and institutional factors, to better understand and address disparities. Prospective studies could further investigate mechanisms underlying the weekend effect, such as staffing patterns, care protocols, and patient severity differences. Lastly, integrating real-time predictive models into clinical workflows may aid in resource allocation and early intervention, potentially mitigating adverse outcomes associated with ICU admissions during weekends.

## Reference

1. Vincent, J.L., et al., *Mean arterial pressure and mortality in patients with distributive shock a retrospective analysis of the MIMIC-III database.* Annals of intensive care, 2018. **8**(1).

2. Yang, W., et al., *Mortality prediction among ICU inpatients based on MIMIC-III database results from the conditional medical generative adversarial network.* Heliyon, 2023. **9**(2): p. 10.

3. Qiu, Y., et al., *Predictive nomogram for 28-day mortality risk in mitral valve disorder patients in the intensive care unit: A comprehensive assessment from the MIMIC-III database.* International journal of cardiolog, 2024. **407**: p. 10.

4. Li, F., et al., *Prediction model of in-hospital mortality in intensive care unit patients with heart failure machine learning-based, retrospective analysis of the MIMIC-III database.* BMJ open, 2021. **11**(7): p. 17.

5. Zhu, Y., et al., *Machine Learning Prediction Models for Mechanically Ventilated Patients: Analyses of the MIMIC-III Database.* Frontiers in medicine, 2021. **8**: p. 9.

6. Hou, N., et al., *Predicting 30‑days mortality for MIMIC‑III patients with sepsis‑3: a machine learning approach using XGboost.* Journal of translational medicine, 2020. **18**(1): p. 14.

7. Yang, S., et al., *A Retrospective Cohort Study: Predicting 90-Day Mortality for ICU Trauma Patients with a Machine Learning Algorithm Using XGBoost Using MIMIC-III Database.* Journal of multidisciplinary healthcare,, 2023. **16**: p. 16.

8. Faust, L., K. Feldman, and N.V. Chawla, *Examining the weekend effect across ICU performance metrics.* Critical care (London, England), 2019. **23**(1): p. 9.

9. Aujesky, D., et al., *Weekend Versus Weekday Admission and Mortality After Acute Pulmonary Embolism.* Circulation 2009. **119**(7): p. 7.

10. Saposnik, G., et al., *Weekends: A dangerous time for having a stroke?* Stroke (1970), 2007. **38**(4).