

Research Question: What are the predictors of in-hospital death of patients diagnosed with heart failure on admission?

Group name: Group A3 Q

Members: Evangelia Paraskevadaki, and Julius Agyei Okanta

Abstract

Background

Heart failure (HF) is a leading cause of hospitalizations, particularly among older adults, with many requiring admissions to intensive care units (ICUs). ICU admission is associated with significantly higher in-hospital mortality rates compared to general ward admissions. Accurate predictive models for mortality can assist in optimizing clinical decision-making and resource allocation.

Objective

To develop a predictive model for in-hospital mortality among HF patients admitted to the ICU using the Medical Information Mart for Intensive Care (MIMIC-III) database.

Methods

This retrospective cohort study analysed 1,152 adult HF patients. Data, including demographics, vital signs, comorbidities, and laboratory results, were collected during the first 24 hours of admission. Key predictors were identified using logistic regression and least absolute shrinkage and selection operator (LASSO) techniques. The model was trained on 70% of the data and validated on the remaining 30%.

Results

Age, renal function (blood urea nitrogen and creatinine), metabolic parameters (anion gap), blood pressure, and ICU length of stay were significant predictors of mortality. The model achieved robust predictive performance, with Nagelkerke R^2 values of 0.336 (training) and 0.300 (validation).

Conclusions:

The study highlights the utility of predictive models in identifying high-risk HF patients, supporting more targeted interventions. External validation and additional predictors are recommended for enhancing clinical applicability.

Introduction

Heart failure (HF) is a clinical syndrome caused by defects in the structure and function of the myocardium, resulting in the impairment of the ejection of blood or ventricular ejection¹. Heart failure accounts for one of the highest causes of hospital admissions in patients aged ≥ 65 years². In the USA, about 10% - 15% of patients with heart failure are admitted to the intensive care unit¹. According to Li et. al (2021), patients admitted to the ICU have significantly higher in-hospital mortality compared to those admitted to hospital wards³. In-hospital mortality rates for all patients with heart failure has been reported to be 4.0% compared to an in-hospital mortality rate of 10.6% for patients who received treatment in the ICU³.

Previously, studies on risks factors for heart failure outcomes have largely considered patient's clinical and demographic characteristics. Risk factors such as age, medical history, and comorbidities are reported to be related to a higher risk of mortality⁴. Due to variation in mortality risks across patient populations, a mortality prediction model that estimates a patient's risk can aid clinical decision making and the reduction of disease burden². Patient's risk for adverse outcomes is often not considered by physicians when calibrating HF therapies, therefore failing to administer effective therapies to patients with the highest risks, for whom will have the greatest benefits of the therapy⁵.

Several predictive models have been developed for the prognosis of heart failure; however, these models have not been used widely in clinical practice^{1,5}. For example, some existing studies had populations from clinical trials or single centers, which may not be representative of the general HF population⁵. In a study conducted by Peterson et. al (2009), the authors developed the Get With The Guidelines - Heart Failure (GWTG - HF) risk score for using clinical variables to predict in-hospital mortality using clinical variables. It also provided clinicians with a validated tool for risk stratification^{1,5}. Using a multivariable logistic regression, they found that age, systolic blood pressure, blood urea nitrogen, heart rate, sodium, chronic obstructive pulmonary disease, and nonblack race were predictive of in-hospital mortality⁵. The tool was proposed for a non-contemporary cohort, hence limited its applicability to current practice¹. The lack of reproducibility of results has become of great concern in scientific literature. In contrast, there is an increase in the utilisation of electronic-based systems in hospitals globally, which has led to vast amount of data being collected in hospitals⁶. Large clinical databases offer opportunities to address the challenges associated with the lack of evidence in medical practice⁶.

The objective of this study is to develop a predictive model for in-hospital mortality among HF patients admitted to the ICU based on data in a large clinical database, the Medical Information Mart for Intensive Care (MIMIC-III) database.

Materials and methods

Data source

The Medical Information Mart for Intensive Care (MIMIC) is a comprehensive, publicly available database containing deidentified health data from patients admitted to the critical care units at Beth Israel Deaconess Medical Center. MIMIC-III, which includes data from 2001 to 2012, was collected using Metavision and CareVue bedside monitoring systems. The database encompasses a wide range of information, including demographics, bedside vital sign measurements, laboratory results, procedures, medications, caregiver notes, imaging reports, and mortality outcomes (both in-hospital and post-discharge).

Study patients

A total of 1152 patients with a first diagnosis of HF were identified by relevant ICD-9 codes. Only patients who were ≥ 18 years old at the time of hospital admission were included in the study.

Data extraction and variable selection

With Structured Query Language queries (MariaDB on DBeaver) we extracted demographic characteristics, vital signs and laboratory data from the following tables in MIMIC-III: ADMISSIONS, PATIENTS, ICUSTAYS, D_ICD DIAGNOSIS, DIAGNOSIS_ICD, LABEVENTS, CHARTEVENTS. More specifically, based on previous studies (Han et al., 2022; Li et al., 2021) we included the following data as possible predictors of in-hospital mortality:

- Demographic characteristics: age, race, sex
- Comorbidities: diabetes type II (DM2), coronary artery disease (CAD), chronic kidney disease (CKD), hypertension, chronic obstructive pulmonary disease (COPD), acute myocardial infarction (AMI), hepatic failure (HepF)
- Vital signs averaged values (heart rate, respiratory rate, systolic blood pressure (SBP)).
- Laboratory test indexes (minimum and maximum values taken in the first 24 hours of a patient's stay) include blood urea nitrogen (BUN), glucose, anion gap (AG), international normalised ratio (INR), prothrombin time (PT), chloride, creatinine, potassium, sodium, platelet, haemoglobin, white blood cell count etc.

Demographic characteristics and vital signs extracted were recorded during the first 24 hours of each admission and laboratory variables were measured during the 1st day of hospital stay with a 1-hour time interval. For laboratory measurements, 1st day minimum and maximum values were included in the analysis (Johnson et al., 2017). Excluding the laboratory measurement, for variable data with multiple measurements, the calculated mean value was included for analysis. Comorbidities were identified using ICD-9 codes. The primary outcome of the study was in-hospital mortality, defined as the vital status at the time of hospital discharge in survivors and non-survivors.

Variables with more than 10% of missing values were excluded from the analysis. Patients with incomplete data were also excluded from the analysis.

Statistical analysis

We present baseline patient characteristics in both samples using a percentage of the total for categorical variables and mean \pm SD for continuous variables. For categorical variables, we used a Pearson's χ^2 test

to assess differences in proportions between the two groups. For all continuous variables, we used independent samples t-tests to compare the mean and standard deviation values.

A total of 47 demographic, clinical and biochemical variables were considered as candidate predictors (Table I). The data was split into training (70%) and validation (30%) datasets. Two methods were used to select the most important predictors for the in-hospital mortality prediction model from the training dataset. First, we used correlation analysis, selecting the predictors which present a bivariate correlation coefficient with mortality, of over 0.15. Then, multicollinearity between candidate predictors was assessed and variable selection was performed to retain variables that do not correlate to other variables with coefficients over 0.80.

Secondly, we applied the least absolute shrinkage and selection operator (LASSO) technique, a regression-based method, to simultaneously perform variable selection and regularization. This approach improves both the prediction accuracy and interpretability of the statistical model, making it particularly effective for reducing high-dimensional data. Variables with non-zero coefficients in the LASSO regression were chosen for further analysis.

Results

Descriptive statistics

The study included 1,152 patients diagnosed with heart failure at admission. Key demographic, clinical, and laboratory characteristics are summarized in Table I. The average age of patients was 69.83 years (SD: 13.28), with a range of 19.18 to 88.96 years. Most of the patients were White (68.1%), followed by Black or African American (13.6%). Females constituted 42% of the cohort. Hypertension (44%) and coronary artery disease (57%) were the most prevalent comorbidities, while diabetes mellitus (50%) and chronic kidney disease (41.9%) were also significant contributors. Chronic obstructive pulmonary disease (27.9%) and acute myocardial infarction (21.2%) were less frequent but notable comorbidities. The mean heart rate at admission was 83.71 bpm (range: 46.94–129.14), while systolic blood pressure averaged 115.88 mmHg. Diastolic blood pressure had a median of 58.08 mmHg. The average length of stay was 10.38 days, with a significant number requiring ICU stays averaging 4.56 days. Mortality during hospitalization was 20.7%.

Univariate analysis

In-hospital mortality was not significantly related to gender ($\chi^2(1) = 0.027$, $p = 0.869$) or ethnicity ($\chi^2(25) = 35.35$, $p = 0.080$). Both the first and last recorded ages (AGE_FIRST and AGE_LAST) were significantly higher in non-survivors compared to survivors ($p < 0.001$). While the average length of stay (AVG_LOS) was slightly higher for non-survivors (10.99 ± 7.53 vs. 10.22 ± 7.38), the difference was not statistically significant ($p = 0.160$). However, ICU length of stay (AVG_ICULOS_days) was significantly longer in non-survivors (5.73 ± 5.74 vs. 4.25 ± 5.21 , $p < 0.001$), suggesting that patients with worse outcomes required more intensive care.

In-hospital mortality was not significantly related to diabetes mellitus ($\chi^2(1) = 0.61$, $p = 0.435$), CAD ($\chi^2(1) = 0.46$, $p = 0.460$), COPD ($\chi^2(1) = 0.045$, $p = 0.833$), CKD ($\chi^2(1) = 3.27$, $p = 0.070$) and AMI ($\chi^2(1) = 3.09$, $p = 0.078$). On the other hand, mortality was associated with HepF ($\chi^2(1) = 21.93$, $p < 0.001$), indicating that patients with HF and hepatic failure are more likely to die during their hospital stay. Non-survivors had higher average heart rates (86.59 ± 12.09 vs. 82.95 ± 12.08 , $p < 0.001$) and significantly lower average systolic blood pressure (SBP) and diastolic blood pressure (DBP).

Both minimum and maximum creatinine and BUN levels were significantly higher in non-survivors ($p < 0.001$), reflecting worse renal function, which is a known risk factor for mortality in heart failure. Non-survivors had significantly higher minimum potassium levels and anion gap levels, indicating more pronounced metabolic dysregulation ($p < 0.001$). There were also significant differences in prothrombin time (PT), partial thromboplastin time (PTT), and international normalized ratio (INR) between groups.

Predictors' selection based on univariate analysis.

Based on the univariate analysis presented in the previous section, non-survivors consistently exhibited worse metabolic, renal, and hemodynamic parameters. Variables to represent these parameters were selected after producing the correlation matrix and removing the variables that presented correlations over 0.8 (Figure I). The variables that were included in the first logistic regression model were `aniongap_max`, `inr_max`, `bun_min`, `AVG_ICU_LOS_days`, `HEARTRATE_avg`, `SBP_avg` and `DBP_avg` (log transformed values).

The results in Table II summarize the logistic regression analysis based on predictors selected through univariate analysis. Nagelkerke R^2 value of 0.255 indicates that approximately 25.5% of the variance in in-hospital mortality is explained by the selected predictors. McFadden's R^2 value of 0.174 suggests a modest but meaningful fit for the logistic regression model. The Variance Inflation Factors (VIFs) for all predictors are below 2, indicating no multicollinearity issues. The minimum BUN value was a significant predictor of in-hospital mortality (OR = 2.73, 95% CI [1.83, 4.06], $p < 0.001$), indicating that a one-unit increase in the log-transformed minimum blood urea nitrogen (BUN) is associated with a 2.73 times increase in the odds of in-hospital mortality. Similarly, a one-unit increase in the log-transformed maximum anion gap is associated with 3.49 times increase in the odds of mortality (OR = 3.49, 95% CI [1.21, 10.04], $p < 0.05$). Each one-unit increase in the log-transformed average heart rate is associated with an exponential increase in odds (OR = 58.89, 95% CI [12.01, 288.75], $p < 0.001$). Lastly, a one-unit increase in the log-transformed average diastolic blood pressure is associated with a significant reduction in mortality risk (OR = 0.04, 95% CI [0.01, 0.19], $p < 0.001$). The findings reinforce the critical importance of monitoring renal function (BUN), heart rate, acid-base status (anion gap), and blood pressure (DBP) in predicting outcomes in hospitalized patients with heart failure.

The validation model is based on the same predictors as the model in Table II, but it is applied to the test dataset (30% of the total sample). Regarding model fit, the Nagelkerke R^2 is 0.205, indicating that 20.5% of the variance in in-hospital mortality is explained by the predictors. This is lower than the 25.5% reported for the training model in Table II. Also, McFadden's R^2 is 0.139, suggesting a slightly weaker fit compared to the training model's value of 0.174. Therefore, the lower R^2 values suggest that the model explains less variance in the validation data compared to the training data. The odds ratio of BUN is smaller in the validation model compared to the training model (OR=1.89, $p = 0.0462$), but still statistically significant, confirming the importance of renal function as a predictor. The OR of anion gap is larger in the validation model (OR =7.04, $p = 0.0150$), suggesting greater importance in the validation cohort. Heart rate is not a significant predictor in the validation, despite the increased effect size in the training model. The effect of diastolic blood pressure is consistent across models, confirming its importance in predicting mortality. The effect of systolic blood pressure becomes more pronounced in the validation model. Overall, the validation model exhibits weaker overall fit (lower R^2 values). Renal function (log `bun_min`), acid-base balance (log `aniongap_max`), and blood pressure (log `SBP_avg` and log `DBP_avg`) consistently predict mortality, while predictors like heart rate (log `HEARTRATE_avg`) and ICU stay duration (log `AVG_ICULOS_days`) show variability, highlighting the need for further investigation into their context-specific impacts.

Predictors' selection based on LASSO regression.

The LASSO regularization extracted 20 potential predictors based on 742 patients in the training dataset (Figures II, III). The results in Table III provide insights into the predictive factors for in-hospital mortality among patients, based on a logistic regression model with multiple predictors. Older age is strongly associated with higher mortality risk, with an OR of 11.73 (95% CI [2.67, 51.50], $p = 0.0014$). Each unit increase in the log-transformed average length of stay reduces the odds of mortality (OR = 0.35, 95% CI [0.21, 0.58], $p < 0.001$). This suggests that longer hospital stays are protective, possibly due to improved monitoring or care overtime. Yet, increased ICU length of stay is a strong predictor of mortality, with an OR of 2.86 (95% CI [1.77, 4.63], $p < 0.001$). This may be since ICU stays reflect the severity of illness and correlate with worse outcomes. Higher average heart rates are associated with a substantial increase in mortality risk (OR = 98.30, 95% CI [16.73, 577.73], $p < 0.001$). Lower systolic blood pressure (OR = 0.14, 95% CI [0.02, 0.94], $p = 0.0443$) and diastolic blood pressure (OR = 0.09, 95% CI [0.01, 0.45], $p = 0.0049$) are strongly associated with increased mortality risk. Higher minimum creatinine levels (OR = 3.40, 95% CI [1.46, 7.94], $p = 0.0041$) and minimum BUN levels (OR = 1.93, 95% CI [1.10, 3.40], $p = 0.0217$) are also associated with increased mortality risk.

The validation model is based on the same predictors as the model in Table III, but it is applied to the testing dataset (30% of the total sample). Regarding model fit, the Nagelkerke R^2 is 0.300, indicating that 30% of the variance in in-hospital mortality is explained by the predictors. This is lower than the 33.6% reported for the training model in Table III. Also, McFadden's R^2 is 0.211, suggesting a slightly weaker fit compared to the training model's value of 0.236. Therefore, the validation model shows slightly lower R^2 values compared to the training data. Age (log_AGE_LAST) and systolic blood pressure (log_SBP_avg) consistently predict in-hospital mortality across both models. Hepatic failure (HepF) becomes more prominent in the validation model, indicating its potential as a key mortality predictor. Predictors like heart rate (log_HEARTRATE_avg) and BUN (log_bun_min) show reduced significance in the validation model, suggesting possible overfitting in the training dataset or cohort-specific effects. The validation model shows a slightly weaker fit and fewer significant predictors compared to the training model. This highlights the importance of external validation and the potential for overfitting in the training data.

Discussion and conclusion

The logistic model with LASSO predictors is preferable because it retains its predictive power in the validation dataset, as shown by a Nagelkerke R^2 of 0.300 in the validation phase, which is slightly lower than its performance on the training data ($R^2 = 0.336$). Key predictors like age, and systolic blood pressure remain significant in both training and validation datasets, indicating robustness across datasets. Also, it includes a broader set of clinically relevant predictors, such as HepF (Hepatic Failure), creatinine levels for renal function (log_creatinine_min). While the proposed model includes more predictors, all have reasonable Variance Inflation Factors (VIFs), indicating no significant multicollinearity.

Data of the MIMIC-III database come from one center (Beth Israel Deaconess Medical Center). Further limitations include the use of log transformation to account for outliers and the exclusion of variables with missing values (> 10%). Future research may take a different approach in outlier and missing value handling to compare results. Also, other algorithms for predictor selection may be applied to examine the robustness of the model. Lastly, additional predictors that are available in the MIMIC-III database may also be examined.

References

1. Lagu, T. et al. (2016) 'Validation and comparison of seven mortality prediction models for hospitalized patients with acute decompensated heart failure', *Circulation: Heart Failure*, 9(8). doi:10.1161/circheartfailure.115.002912.
2. Han, D. et al. (2022) 'Early prediction of in-hospital mortality in patients with congestive heart failure in intensive care unit: a retrospective observational cohort study', *BMJ Open*, 12(7). doi:10.1136/bmjopen-2021-059761.
3. Li, F. et al. (2021) '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*, 11(7). doi:10.1136/bmjopen-2020-044779. 2
4. Pongiglione, B. et al. (2021) 'Patient, hospital and country-level risk factors of all-cause mortality among patients with chronic heart failure: Prospective International Cohort Study', *PLOS ONE*, 16(5). doi:10.1371/journal.pone.0250931.
5. Peterson, P.N. et al. (2010) 'A validated risk score for in-hospital mortality in patients with heart failure from the American Heart Association Get With The guidelines program', *Circulation: Cardiovascular Quality and Outcomes*, 3(1), pp. 25–32. doi:10.1161/circoutcomes.109.854877. 51
6. Johnson, A.E. et al. (2017) 'The mimic code repository: Enabling reproducibility in Critical Care Research', *Journal of the American Medical Informatics Association*, 25(1), pp. 32–39. doi:10.1093/jamia/ocx084.

Appendix

Table I: Demographic characteristics, clinical and laboratory data by in-hospital mortality groups

Variable	Total sample (N=1152)	In-hospital survivals n = 913, 79.3% (mean \pm SD)	In-hospital deaths n = 239, 20.8% (mean \pm SD)	P-value
AGE_FIRST	69.83 \pm 13.28	68.76 \pm 13.59	73.91 \pm 11.13	<0.001
AGE_LAST	71.03 \pm 12.95	69.89 \pm 13.29	75.37 \pm 10.51	<0.001
AVG_LOS	10.38 \pm 7.42	10.22 \pm 7.38	10.99 \pm 7.53	0.160
AVG_ICULOS_days	4.56 \pm 5.35	4.25 \pm 5.21	5.73 \pm 5.74	<0.001
HEARTRATE_avg	83.71 \pm 12.17	82.95 \pm 12.08	86.59 \pm 12.09	<0.001
SBP_avg	115.88 \pm 16.59	117.50 \pm 16.23	109.71 \pm 16.55	<0.001
DBP_avg	58.08 \pm 11.92	59.16 \pm 12.49	53.98 \pm 8.26	<0.001
aniongap_min	14.04 \pm 2.99	13.73 \pm 2.87	15.18 \pm 3.16	<0.001
aniongap_max	16.77 \pm 3.90	16.36 \pm 3.69	18.35 \pm 4.29	<0.001
bicarbonate_min	24.48 \pm 5.04	24.69 \pm 4.87	23.68 \pm 5.56	0.011
bicarbonate_max	27.17 \pm 4.78	27.30 \pm 4.66	26.67 \pm 5.22	0.089
creatinine_min	1.95 \pm 1.48	1.84 \pm 1.44	2.38 \pm 1.56	<0.001
creatinine_max	2.24 \pm 1.73	2.12 \pm 1.70	2.70 \pm 1.75	<0.001
chloride_min	99.29 \pm 5.47	99.48 \pm 5.33	98.59 \pm 5.93	0.036
chloride_max	102.79 \pm 5.73	103.03 \pm 5.61	101.87 \pm 6.10	0.009
glucose_min	114.76 \pm 35.05	113.88 \pm 33.81	118.10 \pm 39.32	0.130
glucose_max	184.75 \pm 76.53	183.81 \pm 76.68	188.35 \pm 76.00	0.412
hematocrit_min	30.68 \pm 5.71	30.77 \pm 5.87	30.36 \pm 5.07	0.284
hematocrit_max	35.32 \pm 5.16	35.48 \pm 5.19	34.70 \pm 5.00	0.035
hemoglobin_min	10.17 \pm 1.91	10.21 \pm 1.95	10.01 \pm 1.75	0.113
hemoglobin_max	11.57 \pm 1.79	11.64 \pm 1.80	11.29 \pm 1.74	0.006
platelet_min	210.53 \pm 93.37	212.45 \pm 94.38	203.16 \pm 89.18	0.158
platelet_max	244.66 \pm 106.23	246.57 \pm 106.91	237.34 \pm 103.50	0.225
potassium_min	3.92 \pm 0.48	3.89 \pm 0.45	4.05 \pm 0.53	<0.001
potassium_max	4.88 \pm 0.77	4.86 \pm 0.76	4.97 \pm 0.82	0.055
ptt_min	34.74 \pm 13.22	33.98 \pm 12.18	37.70 \pm 16.35	0.002
ptt_max	50.40 \pm 29.43	49.09 \pm 28.83	55.48 \pm 31.23	0.006
inr_min	1.63 \pm 0.81	1.58 \pm 0.80	1.80 \pm 0.82	<0.001
inr_max	2.01 \pm 1.71	1.89 \pm 1.58	2.43 \pm 2.10	<0.001
pt_min	16.90 \pm 6.20	16.64 \pm 6.27	17.88 \pm 5.83	0.005
pt_max	19.19 \pm 9.98	18.65 \pm 9.44	21.24 \pm 11.61	0.002
sodium_min	136.13 \pm 4.42	136.20 \pm 4.30	135.86 \pm 4.83	0.327
sodium_max	139.18 \pm 4.15	139.27 \pm 3.92	138.85 \pm 4.93	0.222
bun_min	38.87 \pm 23.10	35.90 \pm 21.71	50.19 \pm 24.74	<0.001
bun_max	43.82 \pm 24.67	40.61 \pm 23.07	56.10 \pm 26.71	<0.001
wbc_min	9.76 \pm 4.80	9.63 \pm 4.74	10.27 \pm 5.02	0.077

Notes. P-values were calculated with independent samples t-test.

Figure I: Correlation matrix of predictors derived by correlation analysis.

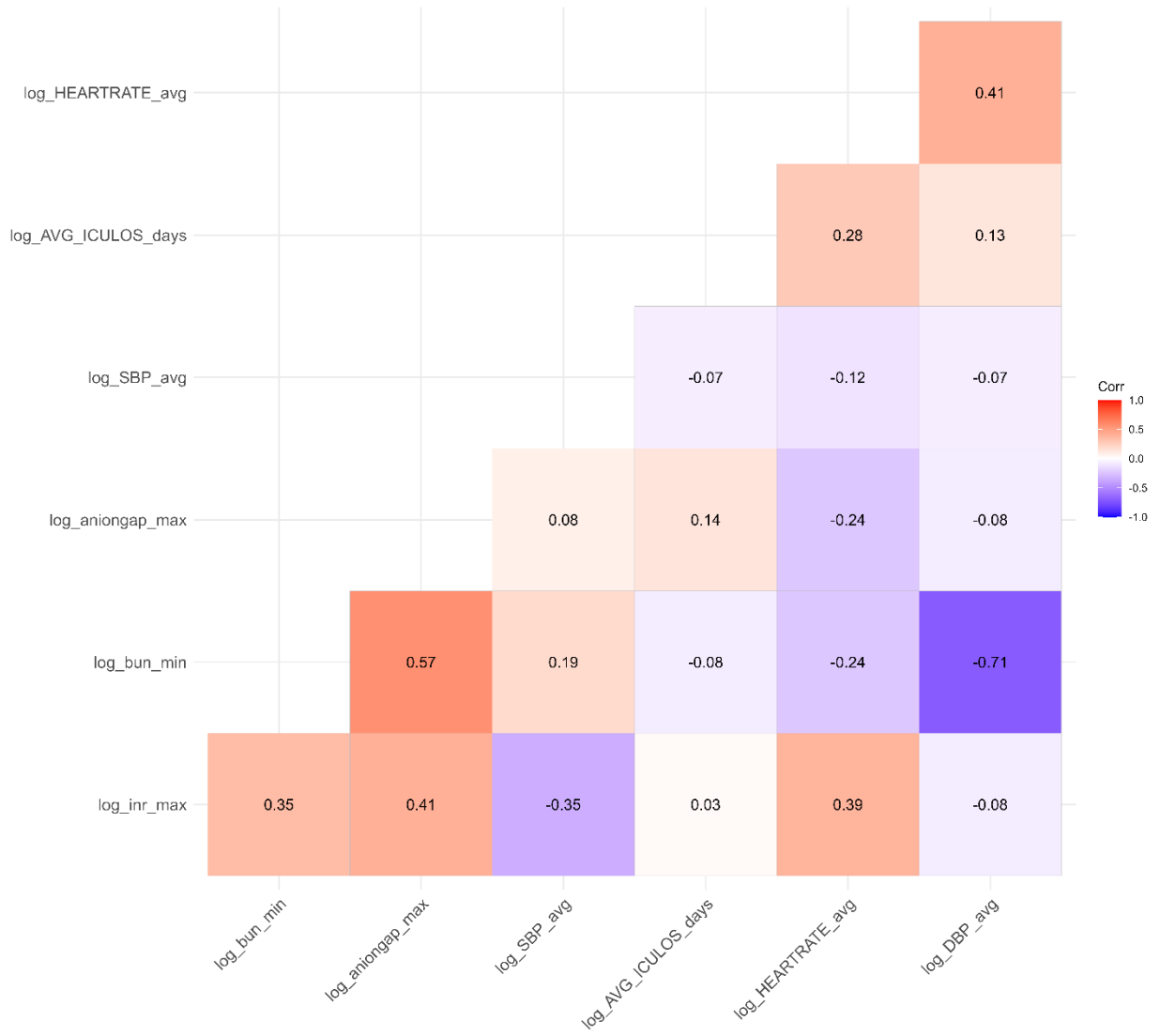


Table II: Logistic Regression Results for Predicting In-Hospital Mortality (train data, predictors selected based on univariate analysis)

Predictor	Estimate	Std. Error	z-value	p-value	OR	95% CI (OR)	VIF
(Intercept)	-9.9988	5.8718	-1.703	0.0886	-	-	-
log_bun_min	1.0044	0.2115	4.749	2.04e-06	2.73	[1.83, 4.06]	1.268
log_aniongap_max	1.2484	0.5399	2.312	0.0208	3.49	[1.21, 10.04]	1.206
log_AVG_ICULOS_days	0.2748	0.1636	1.680	0.0930	1.32	[0.93, 1.87]	1.034
log_HEARTRATE_avg	4.0757	0.8304	4.908	9.18e-07	58.89	[12.01, 288.75]	1.384
log_inr_max	0.3877	0.2796	1.387	0.1655	1.47	[0.86, 2.50]	1.078
log_SBP_avg	-0.9450	0.8243	-1.146	0.2517	0.39	[0.08, 1.96]	1.216
log_DBP_avg	-3.2318	0.7596	-4.255	2.09e-05	0.04	[0.01, 0.19]	1.277

Note. Odds Ratio (OR) is calculated as $\exp(\text{Estimate})$. 95% CI of OR is calculated as $\exp(\text{Estimate} \pm 1.96 * \text{Std. Error})$. Significance levels: *** $p < .001$, ** $p < .01$, * $p < .05$, $p < .10$. Model Statistics: Nagelkerke R^2 : 0.255, McFadden's R^2 : 0.174, Null Deviance: 766.30, Residual Deviance: 633.12, AIC: 649.12.

Figure II: Partial likelihood deviance (binomial deviance) curve versus tuning parameter (λ) in the log scale. The left dotted vertical line represents the value of λ that minimizes the binomial deviance, which corresponds to 20 demographic and clinical variables.

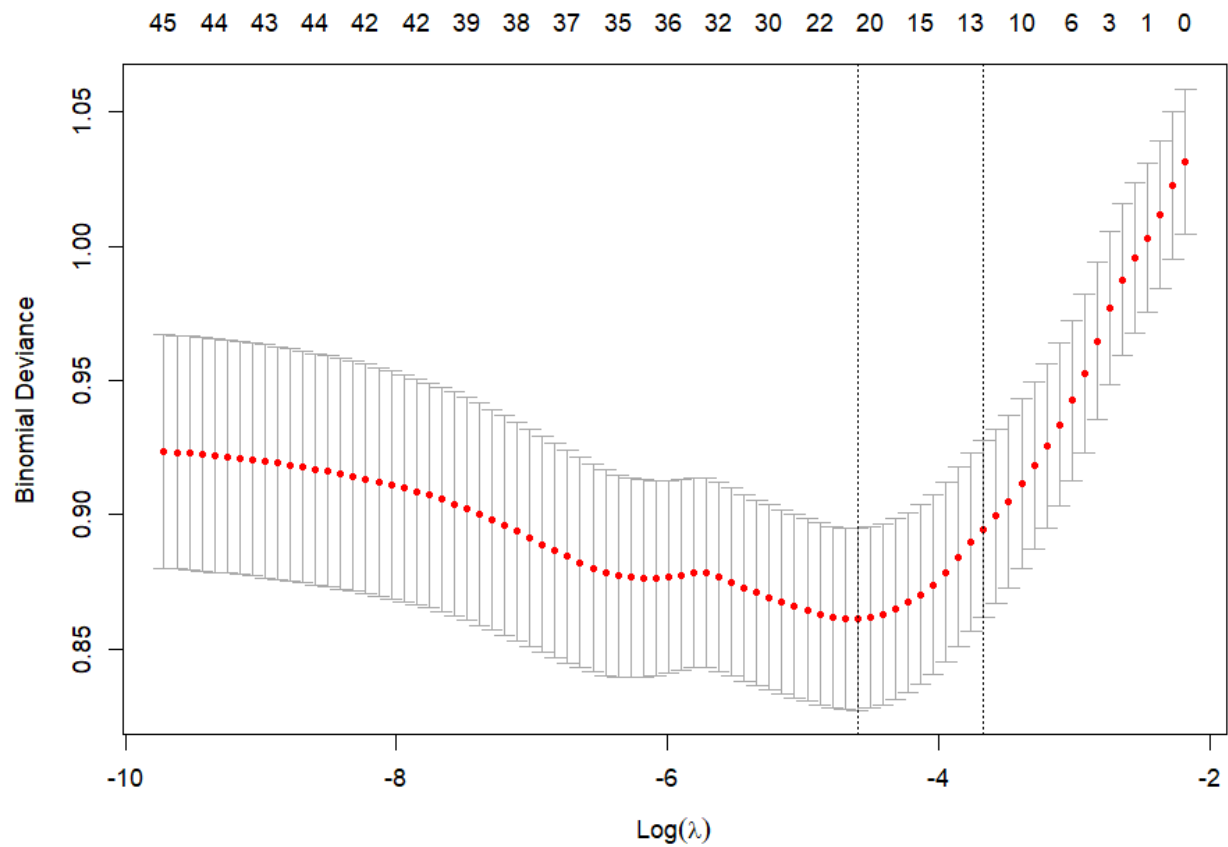


Figure III Bivariate correlation coefficient between the variables selected via LASSO regression from the training dataset

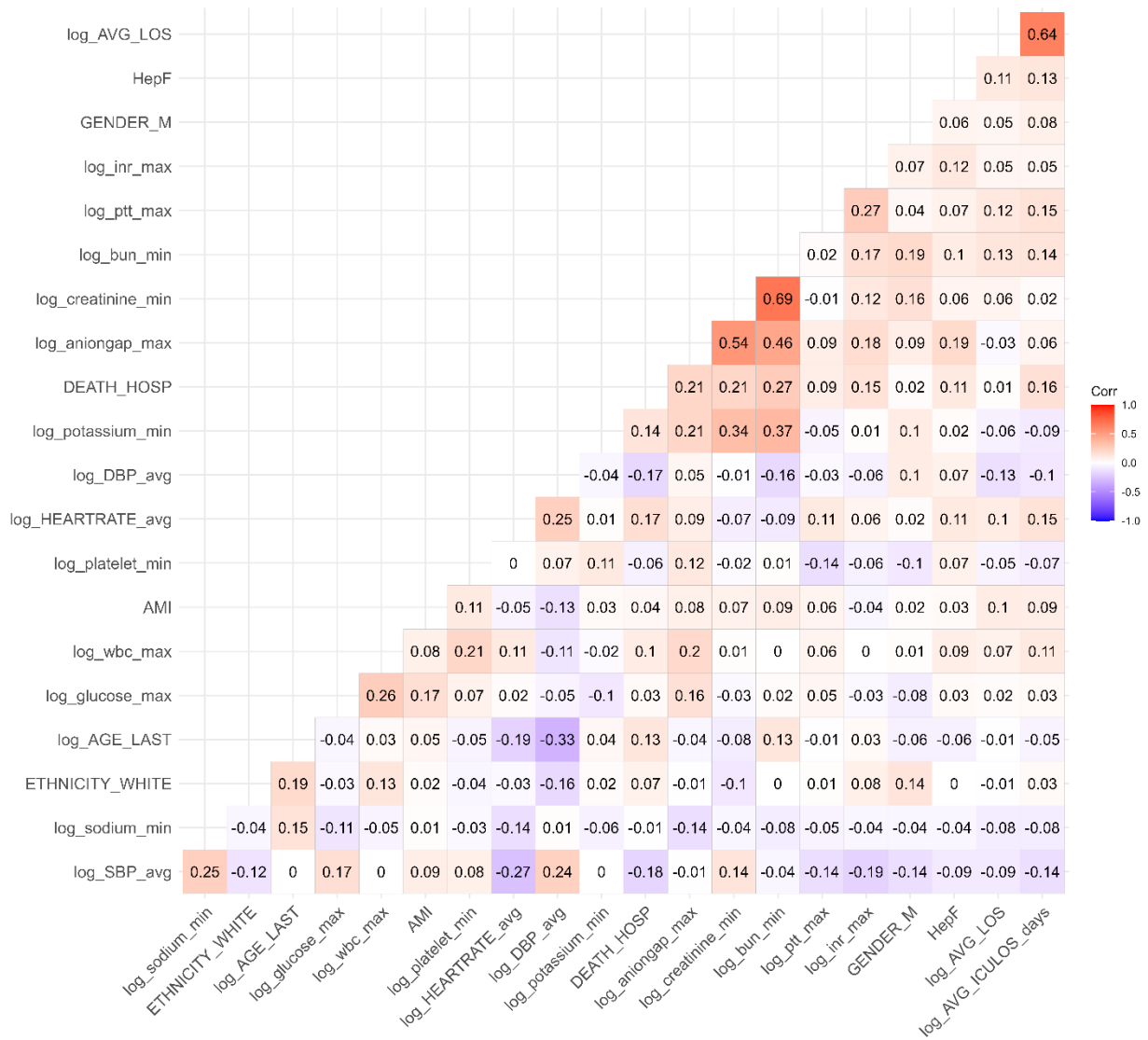


Table III: Logistic Regression Results for Predicting In-Hospital Mortality

Predictor	Estimate	Std. Error	z-value	p-value	OR	95% CI (OR)	VIF
(Intercept)	-43.9760	17.5261	-2.509	0.0121			
AMI	0.2159	0.2533	0.852	0.3940	1.24	[0.74, 2.07]	1.078
HepF	0.7539	0.4507	1.673	0.0944	2.13	[0.87, 5.21]	1.132
ETHNICITY_WHITE	0.1811	0.2465	0.735	0.4625	1.20	[0.74, 1.95]	1.114
GENDER_M	-0.3176	0.2311	-1.374	0.1694	0.73	[0.46, 1.16]	1.190
log_AGE_LAST	2.4633	0.7721	3.190	0.0014	11.73	[2.67, 51.50]	1.334
log_AVG_LOS	-1.0524	0.2614	-4.025	5.69e-05	0.35	[0.21, 0.58]	2.176
log_AVG_ICULOS_days	1.0516	0.2449	4.293	1.76e-05	2.86	[1.77, 4.63]	2.116
log_HEARTRATE_avg	4.5889	0.9012	5.092	3.54e-07	98.30	[16.73, 577.73]	1.447
log_SBP_avg	-1.9392	0.9642	-2.011	0.0443	0.14	[0.02, 0.94]	1.562
log_DBP_avg	-2.4159	0.8593	-2.812	0.0049	0.09	[0.01, 0.45]	1.550
log_aniongap_max	0.1791	0.6506	0.275	0.7832	1.20	[0.33, 4.37]	1.705
log_creatinine_min	1.2219	0.4254	2.873	0.0041	3.40	[1.46, 7.94]	2.578
log_glucose_max	0.4474	0.3423	1.307	0.1912	1.56	[0.85, 2.86]	1.307
log_platelet_min	-0.3719	0.2662	-1.397	0.1624	0.69	[0.41, 1.14]	1.168
log_potassium_min	1.6645	1.2384	1.344	0.1789	5.28	[0.54, 51.40]	1.246
log_ptt_max	0.1031	0.2404	0.429	0.6682	1.11	[0.69, 1.78]	1.145
log_inr_max	0.4425	0.2969	1.491	0.1361	1.56	[0.87, 2.79]	1.163
log_sodium_min	4.5964	3.3630	1.367	0.1717	99.26	[0.35, 282.24]	1.196
log_bun_min	0.6562	0.2858	2.296	0.0217	1.93	[1.10, 3.40]	1.968
log_wbc_max	0.2724	0.2986	0.912	0.3616	1.31	[0.73, 2.37]	1.246

Note. Odds Ratio (OR) is calculated as $\exp(\text{Estimate})$. 95% CI of OR is calculated as $\exp(\text{Estimate} \pm 1.96 * \text{Std. Error})$. Significance levels: *** $p < .001$, ** $p < .01$, * $p < .05$, $p < .10$. Model Statistics: Nagelkerke R^2 : 0.336, McFadden's R^2 : 0.236, Null Deviance: 766.30, Residual Deviance: 585.53, AIC: 627.53