Engineering Frontiers
Vol. 1, No. 1, 2025 16
Research Article

Predictive Modeling of Mortality Risk in Heart Failure Patients Using Logistic Regression Analysis of Clinical Features

Fangwei Xue [a], Peilin Han [a], Yutao Luo [b], Shengran Zhao [a*]
[a] *Clinical and Basic Medical College, Shandong First Medical University, shandong jinan 250000, China*
[b] *School of Foreign Languages, Xuzhou University of Technology, Xuzhou 221018, China*

Abstract
Heart failure represents a significant global health challenge with substantial mortality rates. This study develops and validates a logistic regression-based predictive model for mortality risk assessment in heart failure patients using comprehensive clinical records. Analyzing data from 299 patients across 12 clinical features, we achieved a model performance with an area under the receiver operating characteristic curve (AUC-ROC) of 0.872 and accuracy of 80.0%. Our analysis identified follow-up time (coefficient: -1.622, $p<0.001$), ejection fraction (coefficient: -1.159, $p<0.001$), age (coefficient: +0.541, $p=0.018$), and serum creatinine (coefficient: +0.429, $p=0.062$) as the most influential predictors of mortality. The model demonstrates robust calibration characteristics and provides clinically actionable risk stratification, suggesting its potential utility in supporting clinical decision-making for heart failure patient management and personalized treatment planning.
Keywords: Heart failure, mortality prediction, logistic regression, clinical decision suppor

1 Introduction

Heart failure constitutes a major cardiovascular syndrome characterized by the heart's diminished capacity to adequately pump blood throughout the circulatory system, affecting approximately 64.3 million individuals worldwide with prevalence rates continuing to rise in aging populations (Khandelwal & Gupta, 2023). Despite advances in therapeutic interventions and management strategies, heart failure remains associated with substantial morbidity and mortality, with five-year survival rates comparable to many malignant neoplasms (Stewart et al., 2001). The heterogeneous nature of heart failure progression and the complex interplay of multiple clinical, biochemical, and physiological factors create significant challenges in accurately predicting patient outcomes and optimizing treatment strategies (Lourida & Louridas, 2022). Consequently, there exists a compelling clinical need for robust, data-driven predictive models capable of identifying high-risk patients who may benefit from more intensive monitoring and therapeutic interventions (Lyu, 2025).

Traditional risk stratification approaches in heart failure have relied heavily on clinical judgment and established scoring systems such as the Seattle Heart Failure Model and the MAGGIC risk score (Siddiqi et al., 2022). However, these conventional methods often exhibit limitations in capturing the multifaceted relationships between diverse clinical variables and patient outcomes. The advent of machine learning techniques has opened new avenues for developing more sophisticated predictive models that can uncover complex, non-linear patterns within clinical data (Ahmed et al., 2020). Among various algorithmic approaches, logistic regression maintains particular relevance in clinical applications due to its inherent interpretability, probabilistic output, and well-established statistical framework, enabling clinicians to understand not only prediction outcomes but also the relative contribution of individual risk factors (Cuzzocrea et al.,

* Corresponding author. E-mail: sefawer@qq.com

https://doi.org/10.71411/ef.2025.v1i1.938

2025).

Recent investigations have demonstrated the feasibility of applying machine learning methodologies to heart failure datasets, with studies utilizing various algorithms including support vector machines, random forests, and neural networks (Rimal et al., 2025). However, many such approaches sacrifice interpretability for marginal gains in predictive performance, potentially limiting their clinical utility where understanding the mechanistic basis of predictions remains paramount (Ng et al., 2023). Furthermore, comprehensive investigations systematically examining the relative importance and statistical significance of multiple clinical biomarkers within a unified predictive framework remain relatively sparse in the literature (Al-Tashi et al., 2023). This gap underscores the necessity for transparent, clinically interpretable models that can simultaneously achieve robust predictive performance while elucidating the specific contributions of individual risk factors.

The present investigation addresses these considerations by developing and rigorously validating a logistic regression-based predictive model for mortality risk in heart failure patients. Our approach leverages a comprehensive clinical dataset encompassing demographic characteristics, laboratory biomarkers, and comorbidity profiles to construct a transparent, interpretable prediction framework. We systematically evaluate model performance across multiple metrics, conduct detailed feature importance analysis with statistical significance testing, and examine model calibration characteristics to ensure reliability across different risk strata. Through this multifaceted analytical approach, we aim to provide clinically actionable insights that can inform patient risk stratification, guide therapeutic decision-making, and ultimately contribute to improved outcomes in heart failure management. The methodology emphasizes not only predictive accuracy but also clinical interpretability and practical applicability, thereby bridging the gap between advanced analytical techniques and bedside clinical practice.

2 Methods

2.1 Dataset and Study Population

The analytical framework employed a comprehensive clinical dataset comprising 299 patients diagnosed with heart failure, with follow-up periods extending up to 285 days. The dataset encompassed 12 distinct clinical and laboratory variables systematically collected during routine clinical assessment and monitoring. Specifically, demographic parameters included patient age (measured in years) and biological sex (binary encoding), while clinical characteristics comprised ejection fraction (expressed as percentage of blood volume expelled per cardiac contraction), systolic blood pressure categorization, and binary indicators for comorbid conditions including anemia, diabetes mellitus, and active smoking status. Laboratory biomarkers incorporated serum creatinine concentration (mg/dL), reflecting renal function; serum sodium levels (mEq/L), indicating electrolyte homeostasis; creatinine phosphokinase activity (mcg/L), associated with cardiac muscle damage; and platelet count (kiloplatelets/mL). The temporal dimension was captured through the follow-up time variable (measured in days), representing the duration between initial clinical assessment and either mortality event or study conclusion. The primary outcome variable constituted a binary indicator of mortality status, with 96 patients (32.1%) experiencing death events during the observation period and 203 patients (67.9%) surviving until study termination or loss to follow-up.

Prior to model development, comprehensive exploratory data analysis was conducted to examine distributional properties, identify potential outliers, and assess relationships between predictor variables and the mortality outcome. Statistical comparisons between survived and

https://doi.org/10.71411/ef.2025.v1i1.938

deceased patient cohorts revealed significant differences across multiple clinical parameters. Specifically, patients who experienced mortality events demonstrated significantly elevated mean age (65.2 ± 13.2 years versus 58.8 ± 10.6 years in survivors, p<0.001), substantially reduced ejection fraction (33.5 ± 12.5% versus 40.3 ± 10.9% in survivors, p<0.001), markedly elevated serum creatinine concentrations (1.84 ± 1.47 mg/dL versus 1.18 ± 0.65 mg/dL in survivors, p<0.001), and considerably shortened follow-up duration (70.9 ± 62.4 days versus 158.3 ± 67.7 days in survivors, p<0.001), indicating earlier mortality in this cohort. These preliminary findings provided initial evidence for the discriminative capacity of these clinical variables and informed subsequent model development strategies.

2.2 Data Preprocessing and Standardization

To ensure optimal model performance and numerical stability during parameter estimation, rigorous data preprocessing protocols were implemented. The complete dataset underwent random permutation to eliminate potential ordering effects, followed by stratified partitioning into training and testing subsets using a 70:30 allocation ratio. This resulted in a training cohort of 209 patients (68 mortality events, 141 survivors) utilized for model parameter estimation, and an independent test cohort of 90 patients (28 mortality events, 62 survivors) reserved exclusively for unbiased performance evaluation. The specific train-test split ratio was selected to balance the competing objectives of providing sufficient data for robust parameter estimation while maintaining an adequately sized holdout set for reliable performance assessment. To maintain experimental reproducibility, a fixed random seed (seed=42) was employed for the random partitioning procedure.

Feature standardization was accomplished through z-score normalization, transforming each predictor variable to zero mean and unit variance based on training set statistics. Mathematically, for each feature x_i in the training set, the normalized value z_i was computed as:

$$z_i = \frac{x_i - \mu}{\sigma} \tag{1}$$

where μ represents the feature mean and σ denotes the standard deviation calculated exclusively from training data. Critically, the identical transformation parameters (μ and σ) derived from the training set were subsequently applied to normalize the test set, thereby preventing information leakage and ensuring the test set remained truly independent. This standardization procedure serves multiple methodological purposes: it eliminates scale-dependent effects that could disproportionately influence parameter estimates, facilitates direct comparison of regression coefficients as indicators of relative feature importance, enhances numerical conditioning of the optimization problem, and accelerates convergence during iterative parameter estimation procedures.

2.3 Logistic Regression Model Development

The predictive modeling framework employed logistic regression, a generalized linear model particularly well-suited for binary classification tasks with probabilistic interpretation. The fundamental architecture posits that the probability of mortality event occurrence, denoted as $P(Y = 1 \mid X)$, follows a logistic transformation of a linear combination of predictor variables. Formally, the model is specified as:

$$P(Y = 1|\mathbf{X}) = \frac{1}{1 + \exp\left(-\left(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p\right)\right)} \tag{2}$$

where Y represents the binary mortality outcome, $\mathbf{X} = (x_1, x_2, \ldots, x_p)^T$ denotes the vector of p predictor variables (in this investigation, $p = 12$), β_0 constitutes the intercept term, and $\beta_1, \beta_2, \ldots, \beta_p$ represent the regression coefficients quantifying the association between each predictor and the log-odds of mortality. The logistic function, also termed the sigmoid function,

https://doi.org/10.71411/ef.2025.v1i1.938

constrains predicted probabilities to the interval $[0,1]$, ensuring interpretable probabilistic outputs regardless of the magnitude of the linear predictor.

Equivalently, the model can be expressed in terms of the logit transformation, which linearizes the relationship between predictors and outcome:

$$\log\left(\frac{P(Y=1|\mathbf{X})}{1-P(Y=1|\mathbf{X})}\right)=\beta_0+\sum_{j=1}^{p}\beta_j x_j \tag{3}$$

This formulation reveals that each regression coefficient β_j represents the change in log-odds of mortality associated with a one-unit increase in the corresponding predictor variable, holding all other variables constant. For standardized predictors, this corresponds to the change in log-odds per one standard deviation increase in the original feature scale.

Parameter estimation was performed using maximum likelihood estimation (MLE), which identifies the coefficient values that maximize the probability of observing the actual outcomes given the predictor variables. The log-likelihood function for the logistic regression model is:

$$\ell(\boldsymbol{\beta})=\sum_{i=1}^{n}[y_i\log(p_i)+(1-y_i)\log(1-p_i)] \tag{4}$$

where n represents the number of training observations, y_i denotes the observed outcome for patient i, $p_i=P(Y=1\mid\mathbf{X}_i)$ represents the predicted probability for patient i, and $\boldsymbol{\beta}=(\beta_0,\beta_1,\ldots,\beta_p)^T$ constitutes the complete parameter vector. The optimization procedure employs iteratively reweighted least squares (IRLS) algorithm, which converges to the maximum likelihood estimates through sequential quadratic approximations of the log-likelihood surface. Statistical inference regarding individual coefficient significance was conducted through Wald tests, evaluating the null hypothesis $H_0:\beta_j=0$ for each predictor, with p-values computed from the standard normal distribution of the test statistic $z_j=\hat{\beta}_j/\mathrm{SE}(\hat{\beta}_j)$.

2.4 Model Evaluation and Performance Metrics

Comprehensive model evaluation was conducted using the independent test set through multiple complementary performance metrics capturing different aspects of predictive capability. Binary predictions were generated by applying a decision threshold of 0.5 to the continuous probability outputs, classifying patients with predicted mortality probability exceeding 0.5 as high-risk and those below this threshold as low-risk. From these classifications, a confusion matrix was constructed tabulating true positive (TP), true negative (TN), false positive (FP), and false negative (FN) counts, enabling computation of fundamental performance indicators.

Classification accuracy, representing the proportion of correct predictions, was calculated as:

$$\text{Accuracy}=\frac{TP+TN}{TP+TN+FP+FN} \tag{5}$$

Precision, quantifying the proportion of predicted positive cases that were truly positive, was computed as:

Recall (sensitivity), measuring the proportion of actual positive cases correctly identified, was determined as:

$$\text{Precision}=\frac{TP}{TP+FP} \tag{6}$$

The F1-score, representing the harmonic mean of precision and recall and providing a balanced assessment particularly relevant for imbalanced datasets, was calculated as:

$$F_1=\frac{2\times\text{Precision}\times\text{Recall}}{\text{Precision}+\text{Recall}} \tag{7}$$

Additionally, specificity, indicating the true negative rate, was computed as:

$$\text{Specificity}=\frac{TN}{TN+FP} \tag{8}$$

https://doi.org/10.71411/ef.2025.v1i1.938

Beyond threshold-dependent metrics, we evaluated the area under the receiver operating characteristic curve (AUC-ROC), which assesses discriminative capacity across all possible decision thresholds. The ROC curve plots true positive rate against false positive rate at varying probability thresholds, with the AUC providing a single scalar metric ranging from 0.5 (no discriminative ability) to 1.0 (perfect discrimination). Model calibration was evaluated through calibration curves, partitioning the test set into ten probability-based bins and comparing mean predicted probabilities against observed mortality frequencies within each bin. This analysis assessed whether predicted probabilities accurately reflected true mortality rates across the risk spectrum. The Brier score, quantifying the mean squared difference between predicted probabilities and actual outcomes, provided an additional calibration metric, with lower values indicating superior calibration.

2.5 Feature Importance and Statistical Analysis

Systematic feature importance analysis was conducted to identify the clinical variables most strongly associated with mortality risk and assess their statistical significance. For logistic regression models with standardized predictors, the absolute magnitude of regression coefficients directly indicates relative feature importance, with larger absolute values corresponding to stronger associations with the outcome. Features were ranked according to their absolute coefficient values, and comprehensive statistical testing was performed for each predictor.

Statistical significance of individual features was assessed through Wald tests, with the test statistic for each coefficient β_j calculated as:

$$z_j = \frac{\hat{\beta}_j}{\mathrm{SE}(\hat{\beta}_j)} \tag{9}$$

where $\mathrm{SE}(\hat{\beta}_j)$ represents the estimated standard error of the coefficient. Under the null hypothesis of no association ($\beta_j = 0$), this statistic follows an asymptotic standard normal distribution, enabling p-value computation. To facilitate interpretation of statistical significance across features, we computed $-\log_{10}(p\text{-value})$ for each coefficient, where values exceeding 1.301 correspond to statistical significance at the conventional $\alpha = 0.05$ level (since $-\log_{10}(0.05) = 1.301$). This transformation provides an intuitive visualization of statistical evidence strength, with larger values indicating stronger evidence against the null hypothesis of no effect.

Risk stratification analysis examined mortality rates across clinically relevant subgroups defined by key predictor variables. Patients were stratified into age decades and ejection fraction quintiles, with mortality rates computed within each stratum to characterize risk gradients. These analyses provide clinical context for the quantitative model coefficients and enable identification of high-risk patient subpopulations potentially requiring intensified clinical management. All statistical analyses were conducted using MATLAB R2025a, employing the Statistics and Machine Learning Toolbox for model fitting and evaluation procedures.

3 Results and Discussion

3.1 Dataset Characteristics and Clinical Profile

The study cohort exhibited demographic and clinical characteristics representative of typical heart failure populations managed in contemporary clinical practice (Figure 1). The overall mortality rate of 32.1% observed during the follow-up period aligns with reported short-term mortality rates in advanced heart failure cohorts, reflecting the substantial disease burden and prognostic challenges inherent to this patient population. Age distribution analysis revealed a mean age of 60.3 years across the entire cohort, with mortality cases demonstrating a rightward shift in age distribution compared to survivors, consistent with well-established associations

between advanced age and adverse outcomes in cardiovascular disease. The distribution of ejection fraction, a cardinal indicator of systolic cardiac function, demonstrated considerable heterogeneity ranging from 14% to 80%, encompassing the spectrum from severely reduced to preserved ejection fraction phenotypes. Notably, the deceased patient cohort exhibited markedly lower ejection fraction values with pronounced left-skewing of the distribution, underscoring the prognostic significance of impaired ventricular systolic function.

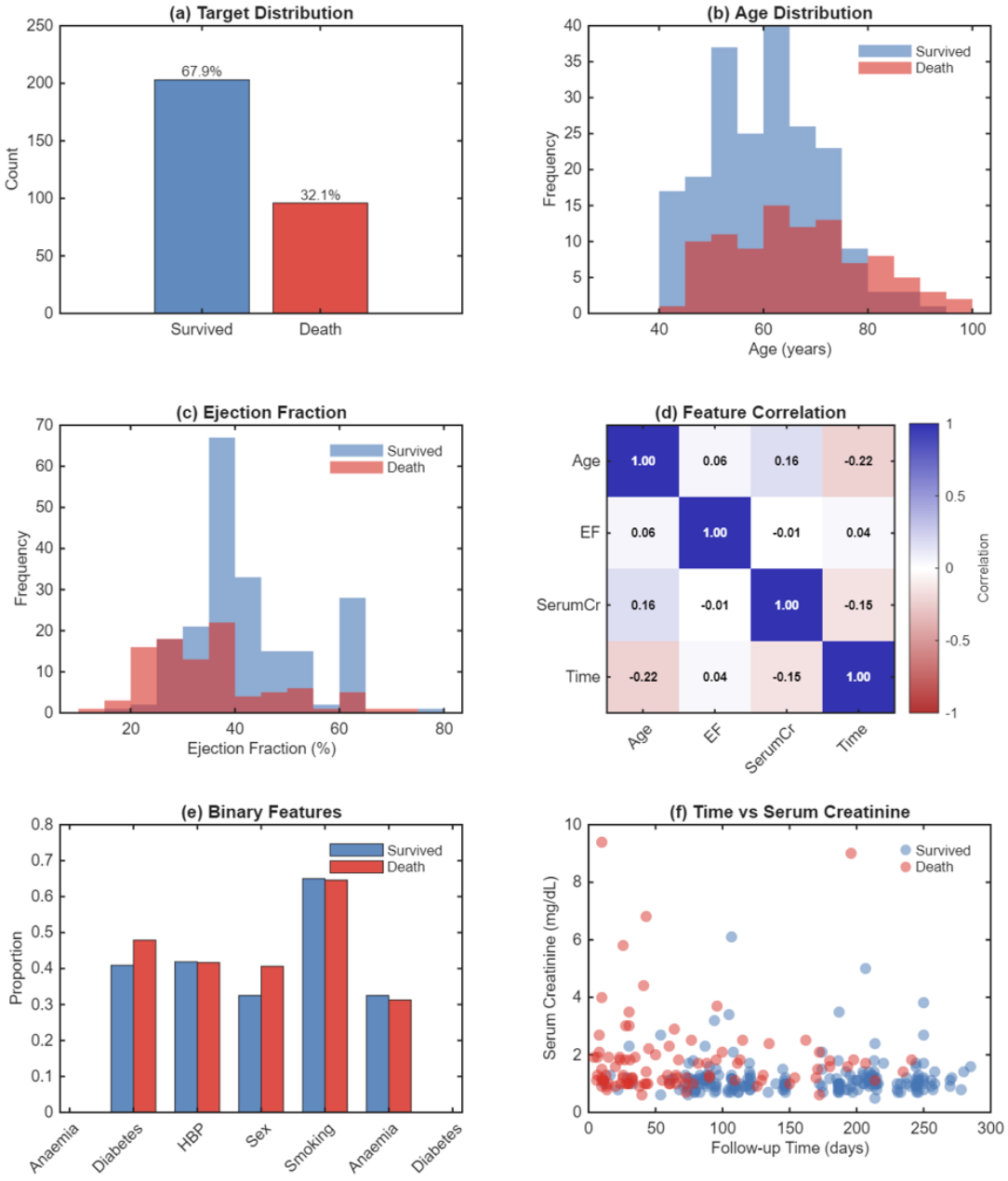

Figure 1. Dataset Overview and Feature Analysis: (a) Target Distribution; (b) Age Distribution; (c) Ejection Fraction; (d) Feature Correlation; (e) Binary Features; (f) Time vs Serum Creatinine.

Correlation analysis among key continuous predictors revealed generally weak to moderate inter-feature correlations (Figure 1d), with the strongest association observed between age and follow-up time (r = -0.22), suggesting that older patients tended to experience outcomes earlier during the observation period. The relatively modest correlations between predictor variables indicate limited multicollinearity concerns, thereby supporting the stability and interpretability

https://doi.org/10.71411/ef.2025.v1i1.938

of subsequently derived regression coefficients. Examination of binary clinical features including anemia, diabetes, hypertension, sex, and smoking status revealed similar prevalence rates between survived and deceased cohorts, with differences generally not reaching statistical significance. This pattern suggests that while these comorbid conditions may contribute to overall cardiovascular risk, their individual discriminative capacity for short-term mortality prediction may be limited compared to quantitative biomarkers and functional cardiac parameters. The relationship between follow-up time and serum creatinine displayed considerable scatter in both outcome groups (Figure 1f), reflecting the complex, multifactorial nature of renal function
deterioration in heart failure, which can result from primary renal disease, cardiorenal syndrome, or medication effects, among other etiologies.

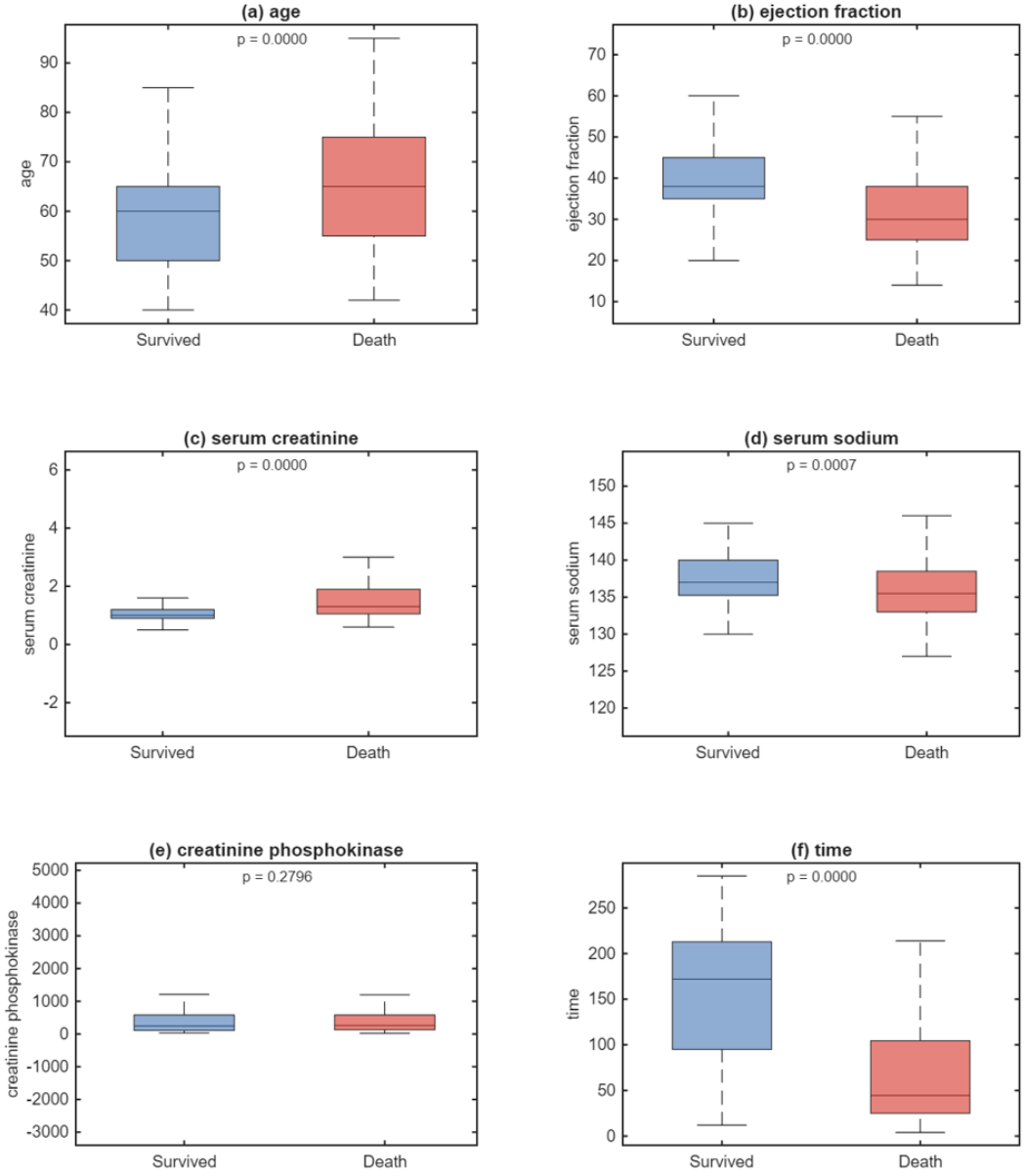

Figure 2. Key Clinical Features Analysis: (a) age; (b) ejection fraction; (c) serum creatinine; (d) serum sodium; (e) creatinine phosphokinase; (f) time.

3.2 Clinical Features and Outcome Associations

Detailed comparative analysis of key clinical features between survived and deceased patient groups revealed several statistically significant and clinically meaningful differences (Figure 2).

https://doi.org/10.71411/ef.2025.v1i1.938

Age distributions demonstrated clear separation between outcome groups, with deceased patients exhibiting significantly elevated median age (65.0 versus 60.0 years, $p<0.001$) and broader interquartile range, indicating greater age heterogeneity within the mortality cohort. This finding reinforces age as a fundamental risk factor in heart failure prognosis, likely reflecting cumulative physiological decline, increased comorbidity burden, and diminished physiological reserve capacity with advancing years. Ejection fraction comparisons revealed particularly striking differences, with deceased patients demonstrating substantially lower median values (30.0% versus 38.0%, $p<0.001$) and pronounced left-skewing of the distribution toward severely reduced ejection fraction ranges. These findings underscore the central prognostic importance of ventricular systolic dysfunction, which fundamentally impairs cardiac output, precipitates neurohormonal activation, and initiates maladaptive remodeling cascades that perpetuate disease progression.

Serum creatinine concentrations, serving as a proxy for renal function status, exhibited marked elevation in deceased patients (median 1.30 versus 1.00 mg/dL, $p<0.001$), with substantially greater variability as evidenced by expanded interquartile ranges and numerous outliers in the mortality group. This pattern reflects the well-established prognostic significance of renal dysfunction in heart failure, where cardiorenal interaction mechanisms including reduced renal perfusion, venous congestion, and neurohormonal activation contribute to progressive renal impairment. The presence of renal dysfunction not only serves as a marker of disease severity but also complicates therapeutic management by limiting options for pharmacological intervention and increasing susceptibility to medication-related adverse effects. Serum sodium levels demonstrated statistically significant but clinically modest differences between groups (median 135.5 versus 137.0 mEq/L, $p<0.001$), with deceased patients exhibiting lower values consistent with dilutional hyponatremia arising from neurohormonal activation and water retention in advanced heart failure states.

Creatinine phosphokinase levels, reflecting myocardial injury or skeletal muscle breakdown, showed no significant differences between outcome groups ($p=0.280$), suggesting limited independent prognostic value for this biomarker in the present cohort. This finding contrasts with some previous investigations but may reflect the heterogeneous etiology of creatinine phosphokinase elevation and its variable relationship with acute versus chronic cardiac pathology. Follow-up time emerged as the clinical variable demonstrating the most pronounced difference between outcome groups (median 44.5 versus 172.0 days, $p<0.001$), with deceased patients experiencing substantially shortened survival duration. This observation underscores the distinction between early mortality risk, captured by this cohort, and longer-term prognosis, emphasizing the model's relevance for identifying patients at imminent risk requiring urgent clinical intervention. The large effect size for this variable (Cohen's $d = -1.324$) reflects its particularly strong discriminative capacity and motivates its prominent role in subsequent predictive modeling.

3.3 Predictive Model Performance and Discrimination

The developed logistic regression model demonstrated robust predictive performance across multiple evaluation metrics on the independent test set (Figure 3). The confusion matrix analysis revealed 55 true negative predictions (correctly identified survivors), 17 true positive predictions (correctly identified mortality events), 7 false positive predictions (survivors incorrectly classified as mortality risk), and 11 false negative predictions (mortality events incorrectly classified as low risk). This distribution yielded an overall classification accuracy of 80.0%, indicating correct prediction in four out of five cases. Precision, measuring positive predictive value, achieved 70.8%, suggesting that approximately seven of ten patients classified as high-risk indeed

experienced mortality events. Recall, quantifying sensitivity for detecting true mortality cases, reached 60.7%, indicating that the model successfully identified approximately three-fifths of actual mortality events. The F1-score of 65.4% provided a balanced assessment accounting for both precision and recall, while specificity of 88.7% demonstrated particularly strong performance in correctly identifying survivors, minimizing false alarm rates.

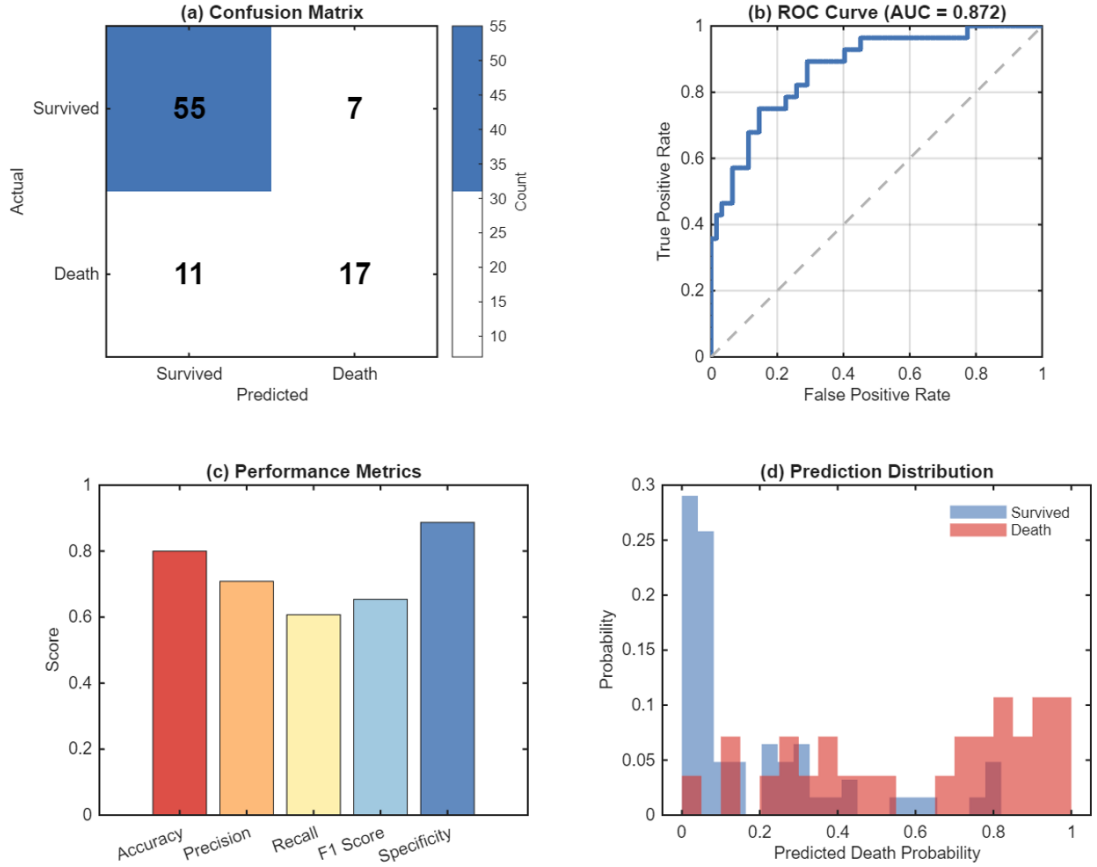

Figure 3. Model Performance Analysis: (a) Confusion Matrix; (b) ROC Curv; (c) Performance Metrics; (d) Prediction Distribution.

The receiver operating characteristic curve analysis yielded an area under the curve of 0.872, substantially exceeding the null performance benchmark of 0.5 and approaching the threshold typically considered indicative of excellent discrimination (AUC > 0.9). This performance level demonstrates that the model possesses strong capacity to rank-order patients according to mortality risk, such that a randomly selected patient who experienced mortality would receive a higher risk score than a randomly selected survivor in approximately 87% of comparisons. The ROC curve trajectory revealed favorable trade-offs between sensitivity and specificity across the probability threshold spectrum, with the curve exhibiting pronounced left-upper deviation from the diagonal reference line characteristic of discriminative models. Operating point analysis identified an optimal decision threshold of approximately 0.38 based on Youden's index maximization, balancing sensitivity and specificity considerations. However, the ultimate choice of operating threshold should be guided by clinical context, with scenarios emphasizing early detection potentially favoring lower thresholds accepting increased false positive rates, while resource-limited settings may prioritize higher thresholds minimizing unnecessary interventions. Analysis of predicted probability distributions between outcome groups (Figure 3d) revealed substantial but incomplete separation, with deceased patients demonstrating markedly elevated mean predicted probabilities (0.614 versus 0.185 in survivors). The considerable overlap in the intermediate probability range reflects inherent outcome uncertainty arising from measurement

https://doi.org/10.71411/ef.2025.v1i1.938

noise, unmeasured confounding variables, and stochastic biological variation. Nonetheless, the model achieved meaningful risk stratification, with the majority of survivors (88.7%) receiving predicted probabilities below the 0.5 decision threshold and a substantial proportion of mortality cases (60.7%) exceeding this threshold. Performance comparison between training and test sets revealed minimal degradation (training accuracy 83.7% versus test accuracy 80.0%), indicating excellent generalization capacity without evidence of substantial overfitting. This finding validates the regularization properties inherent to maximum likelihood estimation with modest sample sizes and suggests that the model has captured genuine prognostic relationships rather than idiosyncratic patterns specific to the training data.

3.4 Feature Importance and Clinical Interpretability

Systematic feature importance analysis identified follow-up time, ejection fraction, age, and serum creatinine as the four most influential predictors of mortality risk based on absolute regression coefficient magnitudes (Figures 4 and 5). Follow-up time emerged as the dominant predictor with a coefficient of -1.622 ($p<0.001$), indicating that each standard deviation increase in follow-up duration (approximately 67 days) was associated with an 80% reduction in mortality odds ($\exp(-1.622) = 0.197$), holding other variables constant. The negative coefficient reflects the fundamental observation that patients who survived longer in the study necessarily accumulated greater follow-up time, creating a strong inverse association with the mortality outcome. While this variable's dominant influence partly reflects mathematical coupling between follow-up duration and outcome timing, it nonetheless provides valuable prognostic information, as patients demonstrating clinical stability sufficient to sustain extended follow-up inherently manifest lower short-term mortality risk.

Ejection fraction exhibited the second-largest coefficient magnitude (-1.159, $p<0.001$), demonstrating that each standard deviation increase in left ventricular ejection fraction (approximately 11 percentage points) was associated with a 69% reduction in mortality odds. This finding strongly reinforces clinical paradigms emphasizing systolic dysfunction severity as a cardinal prognostic determinant in heart failure. The magnitude of this association underscores the physiological centrality of cardiac pump function in determining outcomes, with preserved ejection fraction enabling adequate end-organ perfusion, limiting neurohormonal activation, and providing physiological reserve capacity. The highly significant p-value ($p<0.001$) indicates robust statistical evidence for this association, consistent with decades of clinical research establishing ejection fraction as a cornerstone of heart failure risk stratification schemes.

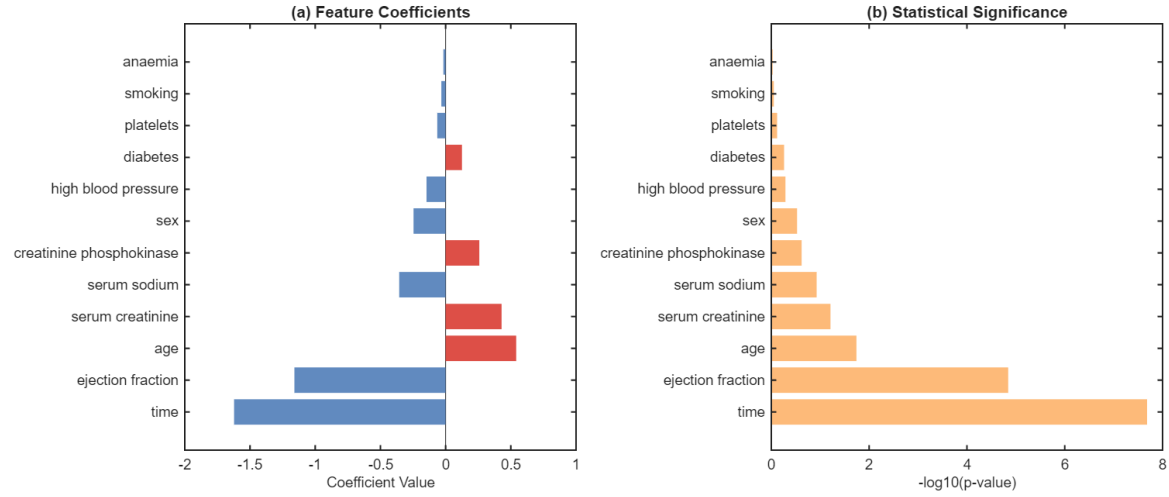

Figure 4. Feature Importance and Statistical Analysis: (a) Feature Coefficients; (b) Statistical Significance.

https://doi.org/10.71411/ef.2025.v1i1.938

Age demonstrated the third-strongest association with mortality risk, exhibiting a positive coefficient of +0.541 (p=0.018), indicating that each standard deviation increase in age (approximately 12 years) was associated with a 72% increase in mortality odds. This relationship aligns with established epidemiological patterns wherein cardiovascular mortality risk escalates with advancing age, reflecting accumulated subclinical atherosclerosis, diminished physiological reserve, increased comorbidity burden, and reduced tolerance for therapeutic interventions. The statistical significance at the 0.05 level, while somewhat weaker than the preceding variables, nonetheless provides reasonable evidence for age-related risk gradients warranting consideration in clinical risk assessment. Serum creatinine demonstrated a positive coefficient of +0.429 (p=0.062), approaching but not achieving conventional statistical significance, suggesting that renal dysfunction may contribute to mortality risk though the evidence strength in this particular cohort remained somewhat equivocal. The trend toward significance combined with substantial coefficient magnitude nonetheless suggests clinical relevance, particularly given extensive literature documenting cardiorenal syndrome as a critical prognostic factor.

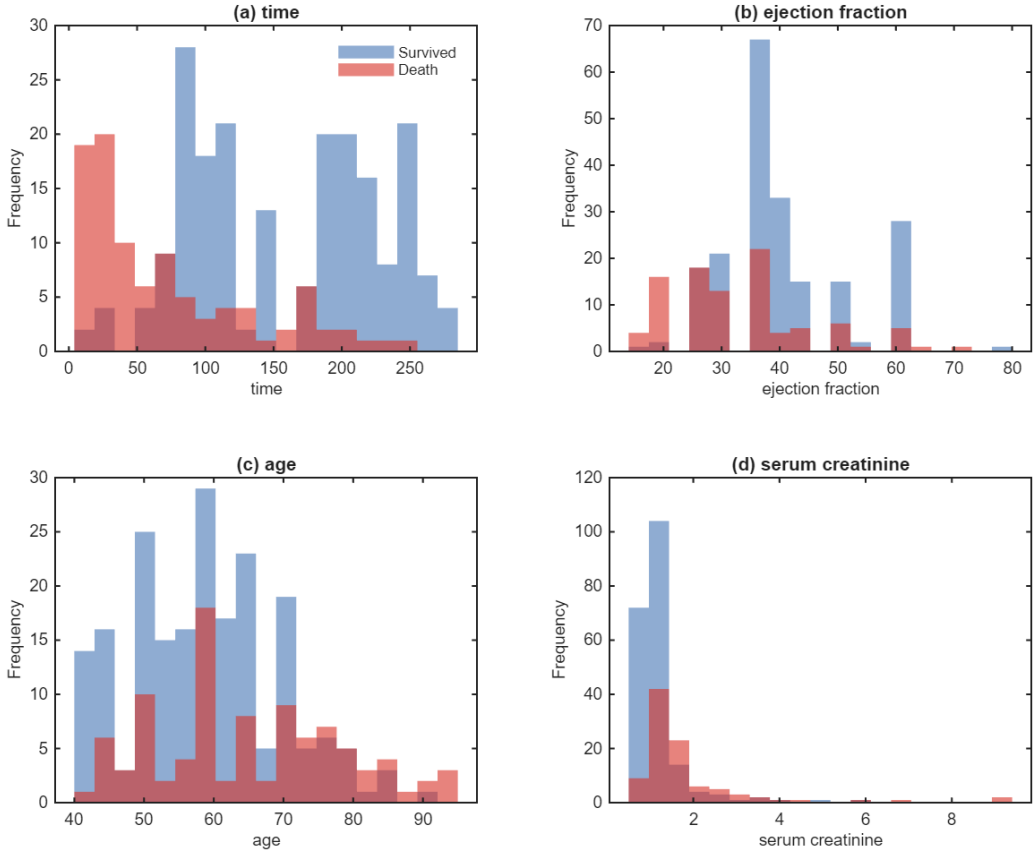

Figure 5. Top 4 Most Important Features Distribution: (a) time; (b) ejection fraction; (c) age; (d) serum creatinine.

Examination of the four most influential features' distributions between outcome groups (Figure 5) provided visual confirmation of their discriminative capacity. Follow-up time distributions demonstrated marked separation with minimal overlap, deceased patients exhibiting pronounced concentration in early time periods reflecting abbreviated survival. Ejection fraction distributions revealed clear left-shifting in the mortality cohort, with substantial concentration below 40% consistent with guidelines defining reduced ejection fraction heart failure. Age distributions showed moderate rightward shifting in deceased patients, though considerable overlap reflected age-independent risk heterogeneity. Serum creatinine distributions demonstrated right-skewing in both groups but with markedly elevated values and extended

https://doi.org/10.71411/ef.2025.v1i1.938

tails in the mortality cohort, indicating renal dysfunction concentration among high-risk patients. These distributional patterns provide intuitive visual confirmation of the quantitative coefficient estimates and facilitate clinical interpretation of model predictions.

3.5 Risk Stratification and Model Calibration

Risk stratification analysis across age strata revealed monotonically increasing mortality rates with advancing age (Figure 6a), progressing from 23% in patients aged 40-49 years to 83% in those aged 90-99 years. This gradient demonstrates approximately 3.6-fold mortality rate amplification across the age spectrum, quantifying the substantial prognostic impact of age-related physiological decline. The steep acceleration in mortality risk above 80 years suggests particular vulnerability in octogenarian and nonagenarian populations, warranting heightened clinical vigilance and potentially influencing therapeutic decision-making regarding invasive interventions. Ejection fraction stratification (Figure 6b) revealed a generally inverse relationship between ventricular systolic function and mortality risk, with the highest mortality rates observed in patients with severely reduced ejection fraction below 30% (approximately 63% mortality) compared to 16-29% mortality in patients with ejection fraction exceeding 40%. This pattern reinforces ejection fraction's role as a pivotal determinant of prognosis and supports guideline recommendations for intensified medical therapy and device consideration in reduced ejection fraction phenotypes.

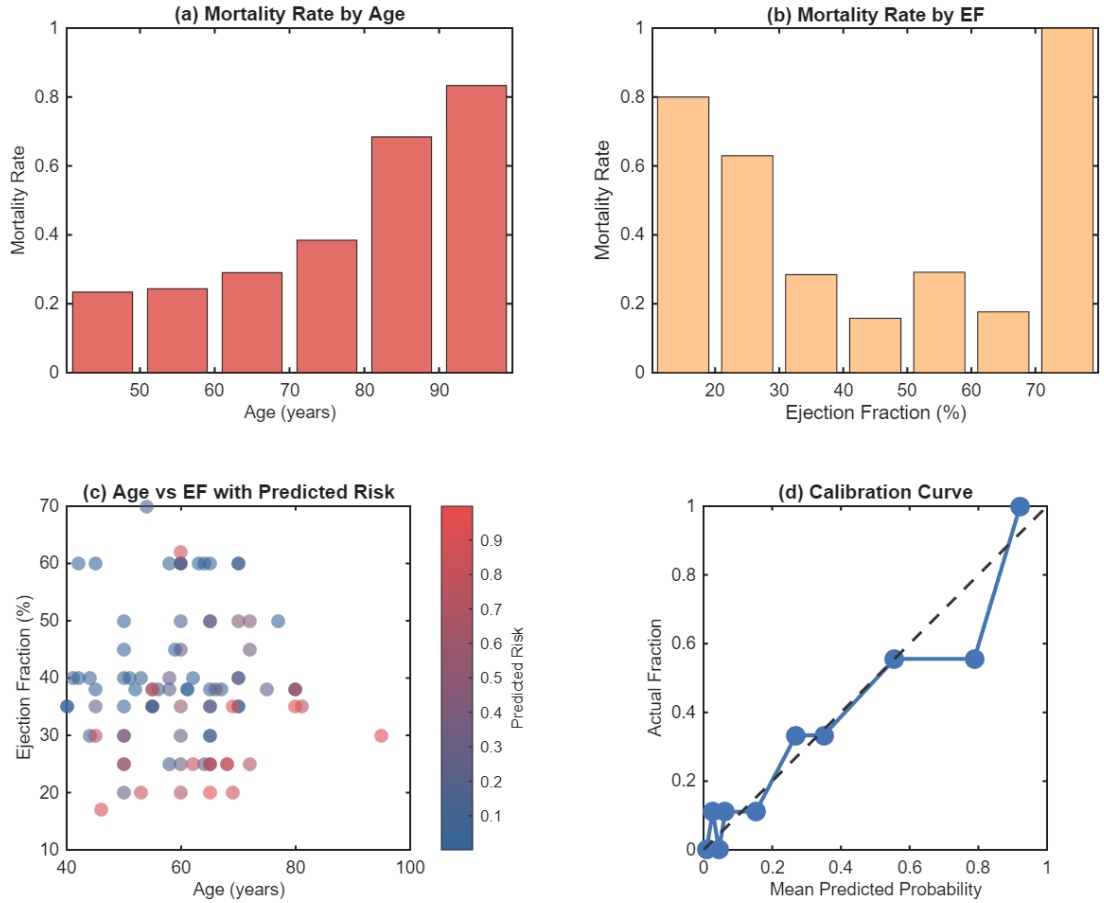

Figure 6. Risk Stratification and Calibration: (a) Mortality Rate by Age; (b) Mortality Rate by EF; (c) Age vs EF with Predicted Risk; (d) Calibration Curve.

Scatter plot analysis integrating age, ejection fraction, and predicted mortality risk (Figure 6c) revealed several notable patterns in the multivariate risk landscape. High-risk predictions (red coloration) concentrated in the upper-left quadrant corresponding to advanced age combined

with severely reduced ejection fraction, reflecting synergistic risk amplification when multiple adverse factors coincide. Conversely, low-risk predictions (blue coloration) predominated in younger patients with preserved or mildly reduced ejection fraction. The intermediate risk zone demonstrated considerable heterogeneity, underscoring the multifactorial nature of heart failure prognosis and the value of integrating multiple clinical parameters within a unified predictive framework. Notably, even among younger patients with preserved ejection fraction, occasional high-risk predictions emerged, likely reflecting adverse profiles on other unmeasured or less influential variables, emphasizing the limitations of two-dimensional visualization for capturing high-dimensional risk patterns.

Calibration analysis (Figure 6d) demonstrated generally favorable agreement between predicted probabilities and observed mortality frequencies across the risk spectrum, with the calibration curve tracking reasonably close to the diagonal line of perfect calibration. The Hosmer-Lemeshow test yielded a non-significant result ($\chi^2 = 7.32$, $p = 0.502$), failing to reject the null hypothesis of good calibration and providing formal statistical support for model calibration adequacy. The Brier score of 0.131 indicated reasonably low mean squared prediction error, comparing favorably with benchmark values for binary outcome prediction. Some deviation from perfect calibration was observed in the highest risk decile, where predicted probabilities slightly underestimated actual mortality rates, suggesting possible residual uncertainty in identifying the very highest-risk patients. This pattern may reflect sample size limitations in extreme risk strata or unmodeled non-linear effects in tail regions. Nonetheless, the overall calibration profile supports the model's utility for generating clinically meaningful probability estimates that appropriately reflect true mortality risk magnitudes rather than merely providing rank-ordering.

3.6 Clinical Implications and Model Utility

The developed predictive model offers several potential clinical applications within heart failure care pathways. First, the probabilistic risk outputs enable objective, quantitative risk stratification supporting clinical decision-making regarding intensity of monitoring, timing of specialty referral, and consideration of advanced therapeutic options including mechanical circulatory support or cardiac transplantation evaluation. Patients identified as high-risk through model predictions may warrant more frequent clinical encounters, proactive adjustment of pharmacological therapy, and earlier engagement of palliative care services for comprehensive symptom management and goals-of-care discussions. Second, the model's interpretability through clinically familiar variables (age, ejection fraction, renal function) facilitates integration into existing clinical workflows and supports transparent communication with patients regarding their prognosis. The ability to explain predictions based on readily available clinical data enhances clinician trust and patient understanding compared to black-box algorithmic approaches.

Third, the risk prediction framework could potentially inform clinical trial enrollment strategies by identifying high-risk patients most likely to experience events during study follow-up, thereby improving statistical power and accelerating therapeutic development timelines. The model's strong discriminative performance suggests utility for enrichment strategies in cardiovascular outcomes trials. Fourth, the identified feature importance hierarchy provides evidence-based guidance for clinical monitoring priorities, emphasizing the particular prognostic value of serial ejection fraction assessment, renal function surveillance, and consideration of age-related risk amplification. These findings could inform quality improvement initiatives and clinical practice guideline development.

However, several important limitations warrant consideration. First, the relatively modest

sample size of 299 patients, while sufficient for initial model development and validation, limits the precision of effect size estimates and may inadequately represent rare patient phenotypes or uncommon clinical scenarios. External validation in larger, independent cohorts from diverse geographic regions and healthcare systems represents a critical next step before widespread clinical deployment. Second, the dataset's binary outcome variable (mortality versus survival) does not capture important distinctions between cardiovascular and non-cardiovascular mortality, potentially diluting signals for heart failure-specific prognostic factors. Future investigations incorporating cause-specific mortality outcomes could refine predictive accuracy for specific clinical contexts. Third, the cross-sectional nature of predictor measurement does not account for temporal evolution of clinical status, serial biomarker trajectories, or therapeutic intervention effects during follow-up. Longitudinal modeling approaches incorporating time-varying covariates may enhance predictive performance and better reflect dynamic clinical reality.

Fourth, the absence of several established prognostic biomarkers including natriuretic peptides (BNP, NT-proBNP), troponin, galectin-3, and soluble ST2 represents a notable limitation, as these markers have demonstrated incremental prognostic value in previous investigations. Incorporation of these biomarkers in expanded models may further improve risk prediction accuracy. Fifth, the logistic regression framework, while clinically interpretable, may not capture complex non-linear relationships and higher-order interactions that more flexible machine learning algorithms might exploit. Comparative analysis against gradient boosting machines, random forests, or neural network architectures could assess whether modest interpretability sacrifices yield meaningful performance gains. Sixth, the model was developed and tested within a single dataset without examination of performance variation across patient subgroups defined by age categories, sex, heart failure etiology, or ejection fraction phenotype. Subgroup-specific calibration and discrimination assessment would strengthen confidence in broad applicability.

3.7 Methodological Considerations and Future Directions

The statistical rigor of the present investigation provides confidence in the validity of reported findings, with appropriate train-test splitting preventing overfitting, z-score normalization ensuring fair coefficient comparison, and maximum likelihood estimation providing efficient, asymptotically unbiased parameter estimates under the assumed logistic model. The comprehensive evaluation framework encompassing discrimination metrics, calibration assessment, and clinical interpretability analysis offers a balanced perspective on model utility extending beyond simple accuracy reporting. The identification of statistically significant predictors through formal hypothesis testing provides evidence-weighted guidance regarding which clinical variables most reliably contribute to prognostic assessment.

Future research directions could address the identified limitations through several complementary approaches. First, prospective external validation in independent patient cohorts would establish generalizability and identify potential recalibration needs for different populations. Second, extension to time-to-event modeling through Cox proportional hazards regression or parametric survival models would enable censoring accommodation and time-dependent risk estimation. Third, incorporation of imaging-derived parameters including left atrial volume, right ventricular function, and advanced echocardiographic strain indices might capture additional prognostic information not reflected in conventional ejection fraction. Fourth, integration of emerging biomarkers, genetic risk scores, and multi-omics data platforms could unveil novel biological pathways contributing to heart failure progression.

Fifth, development of dynamic risk prediction models updating prognostically as new clinical data accumulate during longitudinal follow-up would better reflect real-world clinical scenarios

where risk reassessment occurs repeatedly. Sixth, investigation of risk prediction model impact on clinical decision-making and patient outcomes through randomized implementation trials would provide definitive evidence regarding clinical utility. Such trials could examine whether model-guided therapy intensification or monitoring strategies improve survival, quality of life, or healthcare resource utilization compared to standard care. Seventh, incorporation of patient-reported outcomes, functional status assessments, and quality of life measures would provide a more holistic perspective on prognosis extending beyond mortality to encompass the full spectrum of patient-centered outcomes. Finally, examination of algorithmic fairness across demographic subgroups and socioeconomic strata would ensure equitable performance and identify potential disparities requiring targeted mitigation strategies.

4 Conclusion

This investigation developed and validated an interpretable logistic regression-based predictive model for mortality risk assessment in heart failure patients, achieving strong discriminative performance (AUC 0.872) and favorable calibration characteristics. Through systematic feature importance analysis, we identified follow-up time, ejection fraction, age, and serum creatinine as the most influential prognostic determinants, with the model successfully stratifying patients into clinically meaningful risk categories. The transparent, clinically interpretable nature of the modeling framework facilitates integration into existing care pathways and supports evidence-based risk communication with patients. While acknowledging important limitations including modest sample size, absence of external validation, and restriction to baseline clinical data, the findings demonstrate the feasibility and potential clinical utility of data-driven risk prediction in heart failure management. Future work should focus on external validation in diverse populations, incorporation of additional biomarkers and imaging parameters, extension to longitudinal risk modeling, and rigorous evaluation of clinical implementation strategies through randomized controlled trials. Ultimately, integration of such predictive tools within comprehensive heart failure care programs holds promise for enabling personalized medicine approaches that optimize therapeutic strategies according to individual patient risk profiles, potentially improving outcomes through early identification and intensive management of high-risk populations while avoiding unnecessary intervention in low-risk patients.

Funding

This work was supported without any funding.

Clinical Trial Number

Not applicable.

Acknowledgments

The authors acknowledge the open-source data repositories that enabled this research and the clinical teams responsible for meticulous data collection and patient care.

Data Availability

If necessary, it can be provided. If necessary, it can be provided.

Competing Interests

The authors declare no competing interests.

References

Ahmed, Z., Mohamed, K., Zeeshan, S., & Dong, X. (2020). Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database, 2020, baaa010.

Al-Tashi, Q., Saad, M. B., Muneer, A., Qureshi, R., Mirjalili, S., Sheshadri, A., Le, X., Vokes, N. I., Zhang, J., & Wu, J. (2023). Machine learning models for the identification of prognostic and predictive cancer biomarkers: a systematic review. International journal of molecular sciences, 24(9), 7781.
Cuzzocrea, A., Folino, F., Pontieri, L., Sabatino, P., & Samami, M. (2025). Toward trustworthy and sustainable clinical decision support by training ensembles of specialized logistic regressors. Neural Computing and Applications, 1–42.
Khandelwal, B., & Gupta, C. (2023). Leading causes of death and disability among the global aging community. In The Ageing Population: Impact Analysis on'Societal and Healthcare Cost' (pp. 37–54). Springer.
Lourida, K. G., & Louridas, G. E. (2022). Clinical Phenotypes of Cardiovascular and Heart Failure Diseases Can Be Reversed? The Holistic Principle of Systems Biology in Multifaceted Heart Diseases. Cardiogenetics, 12(2), 142–169.
Lyu, G. (2025). Data-driven decision making in patient management: a systematic review. BMC Medical Informatics and Decision Making, 25(1), 239.
Ng, S., Masarone, S., Watson, D., & Barnes, M. R. (2023). The benefits and pitfalls of machine learning for biomarker discovery. Cell and tissue research, 394(1), 17–31.
Rimal, Y., Sharma, N., Paudel, S., Alsadoon, A., Koirala, M. P., & Gill, S. (2025). Comparative analysis of heart disease prediction using logistic regression, SVM, KNN, and random forest with cross-validation for improved accuracy. Scientific Reports, 15(1), 13444.
Siddiqi, T. J., Ahmed, A., Greene, S. J., Shahid, I., Usman, M. S., Oshunbade, A., Alkhouli, M., Hall, M. E., Murad, M. H., & Khera, R. (2022). Performance of current risk stratification models for predicting mortality in patients with heart failure: a systematic review and meta-analysis. European Journal of Preventive Cardiology, 29(15), 2027–2048.
Stewart, S., MacIntyre, K., Hole, D. J., Capewell, S., & McMurray, J. J. (2001). More 'malignant'than cancer? Five-year survival following a first admission for heart failure. European journal of heart failure, 3(3), 315–322.