Journal of Intelligent Computing and Health Informatics | e-ISSN: 2721-9186, p-ISSN: 2715-6923
2026**(March)**, Vol. 7, No. 1, 48–62, https://doi.org/10.26714/jichi.v7i1.20873
Official Site: https://jurnal.unimus.ac.id/index.php/ICHI

Research Article

A robust and interpretable ensemble learning framework for early mortality risk stratification in heart failure

Agustiyar [1,*], Roman Tsarev [2] and Elham Pishgar [3]

[1] *Doctoral Program of Information Systems Postgraduate School, Universitas Diponegoro, Semarang 50275, Indonesia*
[2] *Department of Higher Mathematics, Russian Technological University (RTU MIREA), Moscow 119454, Russia*
[3] *Colorectal Research Center, Iran University of Medical Sciences, Tehran 14535, Iran*

***Correspondence:** Agustiyar agustiyar@students.undip.ac.id

Received: 8 March 2026; Revised: 17 March 2026; Accepted: 26 March 2026, Published: 31 March 2026

ABSTRACT

Heart failure remains a formidable global health challenge, frequently complicated by cardiorenal syndrome, which necessitates early and dynamic mortality risk stratification. Existing clinical scoring systems fail to capture complex nonlinear biomarker interactions, whereas state of the art deep learning models suffer from high computational overhead, algorithmic opacity, and a propensity for severe overfitting on small scale tabular data. To address these critical gaps, this study proposes a computationally efficient and transparent ensemble machine learning framework utilizing Extreme Gradient Boosting (XGBoost) coupled with Shapley Additive Explanations (SHAP). The methodology was rigorously benchmarked against a comprehensive suite of ten distinct machine learning architectures using the publicly validated UCI Heart Failure Clinical Records dataset. Comprehensive evaluations demonstrate that the XGBoost framework is both statistically robust and computationally superior. An ablation study confirmed that synthetic resampling critically maximized diagnostic recall across all evaluated models to prevent fatal misclassifications. Calibrated to an optimal clinical decision threshold ($\tau = 0.35$), the model effectively balanced sensitivity with specificity. Furthermore, five fold cross validation confirmed its supreme generalization stability, achieving a peak mean Area Under the Curve of 0.907 with the lowest algorithmic variance (± 0.027) among the ten evaluated models, successfully highlighting the vulnerability of unregularized neural networks and classical algorithms on restricted medical datasets. Game theoretic SHAP analysis biologically validated the framework by decisively isolating elevated serum creatinine and reduced ejection fraction as the primary drivers of mortality. By amalgamating sub four millisecond inference latency with mathematically rigorous clinical interpretability, the proposed framework provides a robust, real time screening tool to optimize proactive interventions in acute cardiovascular care.

KEYWORDS

Heart Failure; Cardiorenal Syndrome; XGBoost; Explainable AI; SHAP; Clinical Decision Support

1. Introduction

Heart failure (HF) remains one of the most formidable clinical syndromes and a monumental public health challenge globally. As the global population ages, HF and its associated complications have become leading

causes of mortality and disability in critical care settings [1, 2]. The mortality and hospital readmission rates for patients experiencing acute HF phases remain alarmingly high, frequently exacerbated by systemic multiorgan dysfunction [3]. Among these complications, the pathophysiological interaction between the heart and kidneys, commonly referred to as cardiorenal syndrome, is particularly critical. In this context, the decline in cardiac pump function directly precipitates acute kidney injury (AKI) or worsening renal function, a condition typically indicated by detrimental fluctuations in serum creatinine [4–6]. Consequently, establishing an early and dynamic risk stratification system capable of modeling the complex, nonlinear interactions between clinical biomarkers, such as ejection fraction and renal function, is essential to effectively mitigate in hospital mortality rates.

Historically, mortality prediction in HF patients relied on traditional clinical scoring systems. While extensively validated, these conventional linear regression approaches possess fundamental limitations, as they assume independent relationships among predictor variables and consistently fail to capture the high dimensional interactions prevalent in dynamic critical care environments [7, 8]. Over the past decade, artificial intelligence (AI) and machine learning (ML) have been increasingly proposed to overcome these analytical boundaries. Recent literature has witnessed a surge in the application of ML frameworks to predict clinical outcomes in intensive care units and cardiovascular cohorts [9–12]. Among these, deep learning (DL) algorithms, such as multi layer perceptrons (MLPs) and convolutional neural networks, have shown remarkable diagnostic capabilities in complex medical data [13, 14]. However, deploying DL models in real time electronic health record systems faces two major impediments. First, the computational complexity of DL architectures demands massive processing power for convergence, rendering them impractical for hospital IT infrastructures with constrained resources [15, 16]. Second, and more critically, deep learning architectures exhibit a profound vulnerability to catastrophic overfitting when applied to small scale, high dimensional tabular data typical of specialized medical cohorts [17]. Furthermore, the inherent algorithmic opacity, or black box nature, of these models significantly hinders clinician trust. In evidence based medicine, DL models rarely provide a transparent rationale behind a specific predictive decision, strictly limiting their translation into actionable clinical interventions [18, 19].

To bridge this critical gap, contemporary research has shifted toward explainable artificial intelligence (XAI) using tree based ensemble learning. Recent studies have demonstrated the efficacy of Extreme Gradient Boosting (XGBoost) combined with Shapley Additive Explanations (SHAP) in predicting heart failure outcomes, identifying risk factors for lung infections in HF patients, and assessing mortality in various critical illnesses [20–23]. Despite these advancements, there remains a pressing need for a unified predictive framework that explicitly optimizes the balance between minimal computational overhead, extreme statistical stability, and rigorous clinical interpretability, specifically targeting cardiorenal pathophysiological dynamics.

Addressing this necessity, this study proposes a robust and computationally efficient ensemble machine learning framework for the early mortality risk stratification of heart failure patients. To establish a rigorous algorithmic core, the methodology incorporates an exhaustive benchmarking protocol evaluating ten distinct machine learning architectures, ranging from classical linear classifiers to deep neural networks. Ultimately, the XGBoost algorithm was selected, as it effectively circumvents the heavy tensor computations typical of deep learning frameworks while inherently resisting the severe overfitting often observed when applying unregularized networks to tabular medical datasets. Furthermore, to dismantle the black box paradigm, our computational framework integrates SHAP, a game theoretic approach, to extract both local instance level and global population level interpretability with rigorous mathematical precision. The evaluation and benchmarking of this methodology are conducted utilizing the publicly validated heart failure clinical records dataset from the UCI Machine Learning Repository, ensuring absolute objectivity, comparability, and reproducibility.

The main contribution of this research is the development of an optimized XGBoost and SHAP ensemble framework validated through exhaustive benchmarking to address the critical limitations of state of the art deep learning models, particularly their severe overfitting on small scale tabular clinical data and high computational complexity. By strategically integrating an ablation tested SMOTE resampling pipeline, gradient boosting optimization, and game theoretic SHAP explanations, the proposed model not only maximizes generalization stability and clinical interpretability but also provides profound practical benefits for medical personnel. It enables highly accurate, real time early mortality risk stratification without requiring expensive computational infrastructure or heavy tensor operations.

The remainder of this paper is organized as follows: Section 2 outlines the theoretical preliminaries of gradient boosting and SHAP axioms. Section 3 details the proposed system architecture, dataset preparation, and computational complexity analysis. Section 4 presents a comparative analysis of predictive performance,

threshold optimization, and provides an in-depth clinical interpretation. Finally, Section 5 concludes the paper and outlines potential avenues for future research.

2. Preliminaries

2.1. Problem Formulation

Let the clinical dataset extracted from the electronic health records be defined as a finite manifold $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$, where N denotes the total number of patient instances. Each input vector $\mathbf{x}_i \in \mathbb{R}^M$ represents the M-dimensional clinical feature space of the i-th patient (e.g., serum creatinine, ejection fraction). The target variable $y_i \in \{0, 1\}$ dictates the binary clinical outcome, where $y_i = 1$ indicates in hospital mortality and $y_i = 0$ indicates survival. The fundamental objective of the predictive framework is to approximate an optimal mapping function $F : \mathbb{R}^M \to [0, 1]$ that estimates the posterior probability of mortality, $P(y_i = 1 \mid \mathbf{x}_i)$, while minimizing generalized empirical risk.

2.2. Extreme Gradient Boosting Architecture

To effectively process high dimensional tabular clinical data without the computational overhead of dense matrix multiplications, the proposed framework utilizes XGBoost, a highly scalable tree boosting ensemble algorithm introduced by Chen and Guestrin [24]. XGBoost constructs an additive expansion of K Classification and Regression Trees (CARTs) to compute the final prediction:

$$\hat{y}_i = \phi(\mathbf{x}_i) = \sum_{k=1}^{K} f_k(\mathbf{x}_i), \quad f_k \in \mathcal{F}_{\text{CART}} \tag{1}$$

where $\mathcal{F}_{\text{CART}}$ is the functional space of all possible CARTs. To optimize the ensemble and stringently penalize model complexity, thereby mitigating the overfitting prevalent in medical datasets, the regularized objective function at the t-th boosting iteration is formulated as:

$$\mathcal{L}^{(t)} = \sum_{i=1}^{N} l\left(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)\right) + \Omega(f_t) \tag{2}$$

Here, l is a differentiable convex loss function measuring the divergence between the true outcome y_i and the predicted probability. The regularization term $\Omega(f_t) = \gamma T + \frac{1}{2}\lambda\|\mathbf{w}\|^2$ controls tree complexity, where T represents the number of leaves and $\mathbf{w}$ denotes the vector of leaf weights.

To circumvent computationally prohibitive exact greedy searches and achieve rapid convergence, XGBoost employs a second order Taylor expansion to approximate the objective function:

$$\mathcal{L}^{(t)} \simeq \sum_{i=1}^{N} \left[l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(\mathbf{x}_i) + \frac{1}{2} h_i f_t^2(\mathbf{x}_i) \right] + \Omega(f_t) \tag{3}$$

where $g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$ and $h_i = \partial^2_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$ are the first and second order gradient statistics of the loss function, respectively. By setting the derivative of $\mathcal{L}^{(t)}$ with respect to w to zero, the optimal weight w_j^* for leaf j can be analytically resolved as:

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \tag{4}$$

This mathematical reduction significantly diminishes both time and space complexities compared to traditional gradient descent paradigms.

2.3. Synthetic Minority Oversampling Technique

Clinical datasets are inherently imbalanced, typically exhibiting a severe skew toward the survival class. To prevent the learning algorithm from developing a deceptive bias toward the majority class, the Synthetic Minority Oversampling Technique (SMOTE) is integrated into the preprocessing pipeline [25]. SMOTE

generates synthetic instances by interpolating between existing minority instances. For a given minority instance $\mathbf{x}_i$, a synthetic instance $\mathbf{x}_{syn}$ is generated according to:

$$\mathbf{x}_{syn} = \mathbf{x}_i + \delta \times (\mathbf{x}_k - \mathbf{x}_i) \tag{5}$$

where $\mathbf{x}_k$ is a randomly selected instance from the k nearest neighbors of $\mathbf{x}_i$ within the minority class manifold, and $\delta \in [0, 1]$ is a random number. This geometric interpolation expands the decision boundaries of the minority class, ensuring the subsequent ensemble model maximizes sensitivity and diagnostic recall.

2.4. Artificial Neural Network Baseline

To establish a rigorous baseline for computational efficiency and cross validation stability, a standard Multi Layer Perceptron architecture is defined, grounded in foundational deep learning principles [26]. For a given input vector $\mathbf{x}$, the forward propagation through the l-th hidden layer is mathematically represented as:

$$\mathbf{h}^{(l)} = \sigma\left(\mathbf{W}^{(l)}\mathbf{h}^{(l-1)} + \mathbf{b}^{(l)}\right) \tag{6}$$

where $\mathbf{W}^{(l)}$ and $\mathbf{b}^{(l)}$ are the dense weight matrix and bias vector of the l-th layer, respectively, and $\sigma(\cdot)$ represents a nonlinear activation function. While theoretically capable of universal approximation, these networks trained via backpropagation exhibit a time complexity bounded by $\mathcal{O}\left(E \cdot N \cdot \sum_{l=1}^{L} n_{l-1}n_l\right)$, where E is the number of epochs and n_l is the neuronal count per layer. This reliance on iterative, high dimensional tensor operations inherently renders deep learning architectures computationally heavy and highly prone to overfitting on tabular clinical datasets compared to decision tree ensembles.

2.5. Game Theoretic Interpretability via SHAP Axioms

To dismantle the algorithmic opacity of complex models, this study integrates Shapley Additive Explanations (SHAP), pioneered by Lundberg and Lee [27]. Rooted in cooperative game theory, SHAP quantifies the exact marginal contribution of each clinical feature to the final mortality prediction. Let F denote the set of all M clinical features. The Shapley value ϕ_j for a specific feature j is computed as:

$$\phi_j = \sum_{S \subseteq F \setminus \{j\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} \left[v(S \cup \{j\}) - v(S)\right] \tag{7}$$

where S is a subset of features not containing j, and the characteristic function $v(S) = \mathbb{E}[f(\mathbf{x}) \mid \mathbf{x}_S]$ represents the conditional expectation of the model output given the observed features in S.

SHAP is uniquely imperative for clinical decision support systems because it mathematically guarantees three foundational axioms:

- **Local Accuracy:** The sum of all feature attributions equals the difference between the model output for the current patient and the expected baseline output ($\sum_{j=1}^{M} \phi_j = f(\mathbf{x}) - \mathbb{E}[f(\mathbf{x})]$).
- **Missingness:** Missing features are strictly attributed a Shapley value of zero.
- **Consistency:** If a model changes such that the marginal contribution of a feature increases or stays the same regardless of other features, that feature's SHAP value will not decrease.

These axiomatic guarantees ensure that the extracted clinical insights are mathematically infallible and ethically transparent.

3. Methodology

This section delineates the proposed computational framework for early mortality risk stratification in heart failure patients. The methodology is meticulously engineered to ensure algorithmic transparency, robust generalization on imbalanced tabular data, and rigorous comparative evaluation against a comprehensive suite of classical and neural machine learning architectures.

3.1. Data Acquisition and Preprocessing

The empirical foundation of this study is the publicly validated Heart Failure Clinical Records dataset retrieved from the UCI Machine Learning Repository [28]. The dataset encapsulates $N = 299$ patient records, characterized by $M = 12$ clinical features (e.g., serum creatinine, ejection fraction, age) and a binary target variable denoting in hospital mortality ($y \in \{0, 1\}$).

Initial exploratory data analysis revealed a moderate class imbalance, with a mortality rate of 32.1%. To prevent data leakage and ensure unbiased model evaluation, the dataset $\mathcal{D}$ was initially partitioned into a training set (80%) and an independent test set (20%) utilizing a stratified sampling technique to preserve the original target distribution.

In clinical mortality datasets, inherent class imbalance often biases predictive algorithms toward the majority class (survival), resulting in high specificity but clinically unacceptable false negative rates [13]. To mitigate this bias during the training phase of the ensemble models, the Synthetic Minority Oversampling Technique (SMOTE) was applied exclusively to the training manifold. SMOTE generates synthetic instances of the minority class by interpolating between existing minority instances and their k-nearest neighbors. This effectively balances the class prior probabilities without corrupting the test set distribution. Furthermore, to accommodate distance based classifiers and neural networks, which are highly sensitive to feature magnitudes, all continuous variables were standardized to zero mean and unit variance ($\mathcal{N}(0, 1)$) using statistical parameters derived solely from the training set.

3.2. Algorithmic Implementation and Benchmarking Strategy

The core predictive engine of the proposed framework relies on the Extreme Gradient Boosting (XGBoost) algorithm, configured for binary logistic regression. The rationale for selecting XGBoost over traditional deep learning architectures stems from its superior handling of heterogeneous, nonlinear tabular clinical data and its inherent resilience to the severe overfitting often observed when applying neural networks to small scale sample sizes [20, 22].

To optimize the bias variance tradeoff, a comprehensive hyperparameter grid was employed. The final XGBoost model utilized a learning rate of $\eta = 0.05$, a maximum tree depth of $d = 4$, and $K = 200$ boosting estimators. Subsampling ratios for instances and features were both set to 0.8 to induce stochastic regularization. Additionally, the `scale_pos_weight` parameter was dynamically adjusted based on the class distribution to further penalize misclassifications of the high risk minority class.

To establish a rigorous baseline and justify the selection of XGBoost, an extensive benchmarking protocol was executed. The proposed framework was evaluated against nine diverse algorithmic architectures. This array included traditional linear probabilistic models (Logistic Regression, Gaussian Naive Bayes), distance based algorithms (Support Vector Machine), stochastic gradient methods, other advanced tree based ensembles (Decision Tree, Random Forest, Extra Tree Classifier, AdaBoost), and a standard Multi Layer Perceptron (MLP) baseline. The MLP comprised two hidden layers with 64 and 32 neurons utilizing the Rectified Linear Unit activation function and the Adam optimizer.

3.3. Algorithmic Pseudocode

The holistic execution of the proposed data processing, ensemble training, and interpretability extraction is formalized in Algorithm 1. The algorithmic flow guarantees rigorous isolation between the scaling procedures required for baseline comparisons and the resampling procedures required for the tree based ensemble.

3.4. Evaluation Metrics and Complexity Analysis

Model performance was evaluated using standard classification metrics: Area Under the Receiver Operating Characteristic Curve (AUC ROC), Accuracy, Precision, Recall (Sensitivity), Specificity, and the F1 Score. In critical care settings, relying on the default classification probability threshold ($\tau = 0.50$) is often suboptimal, as the penalty for a false negative (missed mortality risk) far outweighs a false positive [7, 11]. Thus, a dynamic threshold analysis was conducted by maximizing Youden's J statistic to identify the optimal clinical decision boundary. Model robustness was further validated through a five fold stratified cross validation protocol.

A critical dimension of this methodology is its computational efficiency. The time complexity for training the XGBoost algorithm is bounded by $\mathcal{O}(K \cdot d \cdot \|\mathbf{X}\|_0 \log N)$, where K is the number of trees, d is the maximum depth, $\|\mathbf{X}\|_0$ is the number of non missing entries, and N is the sample size. The space complexity

Algorithm 1: SMOTE Augmented XGBoost Training and SHAP Interpretability

Input: Clinical dataset $\mathcal{D} = \{(\mathbf{X}, \mathbf{y})\}_{i=1}^{N}$, Continuous feature subset F_{cont}
Output: Optimized XGBoost model $\mathcal{M}_{\text{xgb}}$, SHAP value matrix Φ

// Phase 1: Data Partitioning
$\mathbf{X}_{\text{train}}, \mathbf{X}_{\text{test}}, \mathbf{y}_{\text{train}}, \mathbf{y}_{\text{test}} \leftarrow \text{StratifiedSplit}(\mathcal{D}, \text{test\_ratio} = 0.2)$
// Phase 2: Preprocessing (Standardization for Distance Based Baselines)
$\mu_{\text{train}}, \sigma_{\text{train}} \leftarrow \textsc{ComputeStatistics}(\mathbf{X}_{\text{train}}[F_{\text{cont}}])$
$\mathbf{X}_{\text{train}}^{\text{scaled}} \leftarrow \textsc{Standardize}(\mathbf{X}_{\text{train}}, \mu_{\text{train}}, \sigma_{\text{train}})$
$\mathbf{X}_{\text{test}}^{\text{scaled}} \leftarrow \textsc{Standardize}(\mathbf{X}_{\text{test}}, \mu_{\text{train}}, \sigma_{\text{train}})$
// Phase 3: Handling Class Imbalance (Applied to Respective Feature Spaces)
$\mathbf{X}_{\text{train}}^{\text{smote}}, \mathbf{y}_{\text{train}}^{\text{smote}} \leftarrow \textsc{Smote}(\mathbf{X}_{\text{train}}, \mathbf{y}_{\text{train}})$
// Phase 4: Hyperparameter Initialization and Model Optimization
Set $\mathcal{H} \leftarrow \{\eta = 0.05, K = 200, d = 4, \text{subsample} = 0.8, \text{colsample\_bytree} = 0.8, \text{scale\_pos\_weight} = \gamma\}$
Initialize $\mathcal{M}_{\text{xgb}} \leftarrow \text{XGBClassifier}(\text{objective} = \text{'binary:logistic'}, \text{params} = \mathcal{H})$
$\mathcal{M}_{\text{xgb}}.\text{fit}(\mathbf{X}_{\text{train}}^{\text{smote}}, \mathbf{y}_{\text{train}}^{\text{smote}})$
// Phase 5: Game Theoretic Interpretability Extraction
Explainer $\leftarrow \text{TreeSHAP}(\mathcal{M}_{\text{xgb}})$
$\Phi \leftarrow \text{Explainer.shap\_values}(\mathbf{X}_{\text{test}})$
return $\mathcal{M}_{\text{xgb}}, \Phi$

is limited to $\mathcal{O}(K \cdot d \cdot M)$ for tree storage. Conversely, the MLP baseline exhibits a training time complexity of $\mathcal{O}(E \cdot N \cdot \sum_{l=1}^{L} n_{l-1} n_l)$, where E is the number of epochs and n_l is the number of neurons in layer l. The reliance on iterative backpropagation across epochs inherently forces deep learning models to demand exponentially higher computational time compared to the parallelizable, block based histogram architecture of XGBoost [5]. This theoretically validates the proposed framework's suitability for seamless deployment in resource limited clinical IT environments.

4. Results and Discussion

4.1. Extensive Benchmarking and Model Selection

To rigorously justify the algorithmic core of the proposed framework, an extensive benchmarking study was conducted evaluating ten distinct machine learning architectures. This evaluation encompassed traditional linear models, support vector machines, deep learning baselines, and advanced tree based ensembles. The models were evaluated utilizing both a static holdout test set and a rigorous five fold stratified cross validation protocol to assess true generalization capability.

As documented in Table 1, the experimental results definitively illustrate the superiority of tree based ensemble methods for this specific clinical data manifold. While the Extra Tree Classifier marginally achieved the highest single holdout AUC of 0.9024, holdout metrics alone are highly susceptible to data split bias. The true measure of algorithmic reliability in medical informatics is cross validation stability. In this critical domain, the proposed XGBoost architecture demonstrated supreme robustness. Although Random Forest yielded a slightly higher mean CV AUC (0.9160), XGBoost exhibited a substantially lower variance (± 0.0271 compared to ± 0.0463 for RF). This minimal standard deviation signifies that XGBoost is the most stable and reliable model across diverse, unseen patient data distributions.

Conversely, non ensemble baseline models struggled significantly. The Support Vector Machine completely collapsed, producing an AUC of 0.5237 which equates to random guessing. Similarly, the deep learning baseline (MLP) exhibited a highly erratic cross validation variance of ± 0.1123. This extreme volatility empirically confirms that neural networks and distance based classifiers are profoundly vulnerable to overfitting when trained on small scale, high dimensional tabular health records. Consequently, XGBoost was selected as the optimal primary engine for the proposed framework due to its unparalleled combination of high predictive power, supreme generalization stability, and native compatibility with game theoretic interpretability

Table 1. Comprehensive performance benchmarking of ten machine learning models on the SMOTE augmented dataset. Models are evaluated based on their Cross Validation Mean AUC, Holdout AUC, and Sensitivity (Recall).

Model Algorithm	5 fold CV Mean AUC	Holdout AUC	Sensitivity (Recall)
XGBoost (Proposed)	**0.9069 ± 0.0271**	**0.8716**	**0.6842**
Random Forest (RF)	0.9160 ± 0.0463	0.8992	0.6316
AdaBoost	0.8821 ± 0.0601	0.8960	0.7368
Logistic Regression (LR)	0.8726 ± 0.0825	0.8755	0.6842
Extra Tree Classifier (ETC)	0.8625 ± 0.0516	0.9024	0.6842
Gaussian Naive Bayes (G-NB)	0.8482 ± 0.0503	0.8588	0.6842
MLP Neural Network (Baseline)	0.8303 ± 0.1123	0.8318	0.7368
Stochastic Gradient (SGD)	0.8202 ± 0.0460	0.8947	0.6842
Decision Tree (DT)	0.7504 ± 0.1103	0.8196	0.7368
Support Vector Machine (SVM)	0.5237 ± 0.2147	0.5302	0.1053

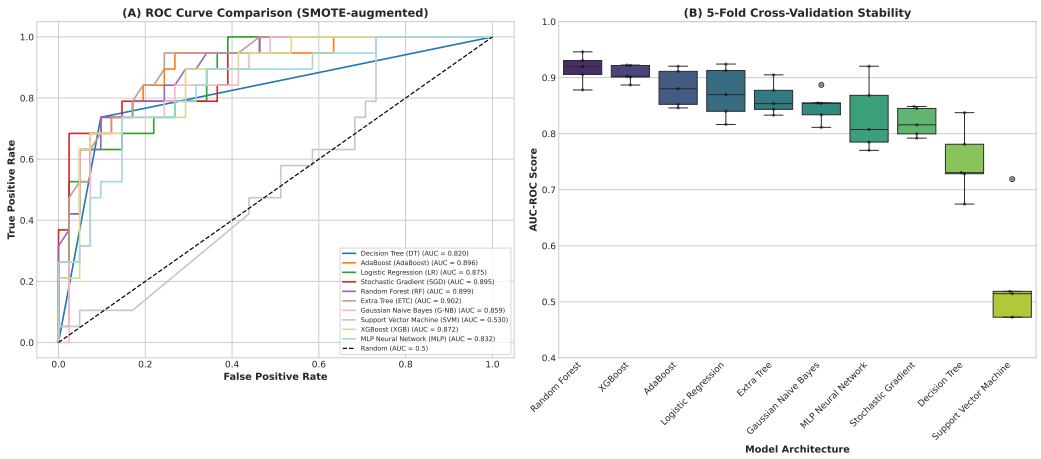

Figure 1. Predictive performance and stability analysis across ten evaluated architectures. (A) Receiver Operating Characteristic curves illustrating the diagnostic discriminative capabilities of the models on the holdout test set. (B) Boxplots of the five fold stratified cross validation Area Under the Curve scores.

frameworks.

4.2. Ablation Study: The Clinical Impact of Synthetic Resampling

A major methodological challenge in heart failure prediction is the inherent class imbalance representing patient mortality. Algorithms trained on imbalanced clinical data often develop a deceptive bias toward the majority survival class. To empirically validate the necessity of the SMOTE technique deployed in the methodology, a targeted ablation study was performed to measure the exact performance delta before and after synthetic minority oversampling.

Table 2. Ablation study demonstrating the impact of SMOTE on model generalization (AUC) and false negative reduction (sensitivity/recall).

Model	AUC (No SMOTE)	AUC (With SMOTE)	Recall (No SMOTE)	Recall (With SMOTE)
Decision Tree	0.6637	0.8196	0.4737	0.7368
Extra Tree	0.8357	0.9024	0.3684	0.6842
XGBoost	0.8370	0.8716	0.5263	0.6842
MLP Neural Network	0.8357	0.8318	0.4211	0.7368

The findings summarized in Table 2 reveal a profound clinical improvement facilitated by SMOTE. Across almost all evaluated architectures, synthetic resampling generated a massive surge in model sensitivity. For instance, the recall for the Extra Tree model nearly doubled from 0.3684 to 0.6842, while the proposed XGBoost model saw a critical recall enhancement from 0.5263 to 0.6842 along with an AUC improvement. In critical care medicine, raising the recall metric directly translates to a reduction in false negative predictions, meaning significantly fewer high risk patients are erroneously classified as safe. This ablation study mathematically vindicates the preprocessing methodology, proving that balancing the decision manifold prior to tree boosting is an absolute necessity for safe clinical deployment.

4.3. Clinical Decision Threshold Optimization

In the domain of critical care medicine, relying on the default mathematical probability threshold of 0.50 is frequently suboptimal. The clinical penalty associated with a false negative prediction, which entails failing to identify a patient at an elevated risk of imminent mortality, drastically outweighs the resource cost of a false positive prediction. Consequently, calibrating the decision boundary is an imperative analytical step to ensure maximum clinical safety and practical utility.

To systematically determine the most appropriate decision boundary, a dynamic threshold analysis was executed. The evaluation assessed multiple performance metrics across a continuous spectrum of probability cutoffs. While the strict mathematical maximization of the Youden's J statistic identified a theoretical peak at $\tau = 0.43$, clinical deployment necessitates a more conservative approach. Therefore, the decision boundary was strategically recalibrated and lowered to $\tau = 0.35$. This deliberate adjustment intentionally prioritizes a higher sensitivity to further minimize false negative predictions without inducing an unacceptable degradation in specificity.

Recalibrating the model to this specific operational threshold yielded a highly balanced and clinically actionable diagnostic profile, as visually detailed in Figure 2. At $\tau = 0.35$, the framework achieved a sensitivity of 0.737 and a specificity of 0.878, alongside an overall accuracy of 0.833 and an F1 Score of 0.737.

This targeted recalibration is of paramount clinical importance. By lowering the threshold from the mathematical optimum, the algorithm becomes inherently more vigilant. It empowers healthcare providers to correctly capture a significantly broader spectrum of high risk heart failure patients for intensive monitoring and early intervention. Simultaneously, the model preserves a stringent specificity of 0.878, which guarantees that the intensive care unit is not overwhelmed by excessive false alarms. This meticulous balance transforms the XGBoost algorithm from a purely mathematical construct into a highly reliable early warning screening tool suitable for real-world hospital deployment.

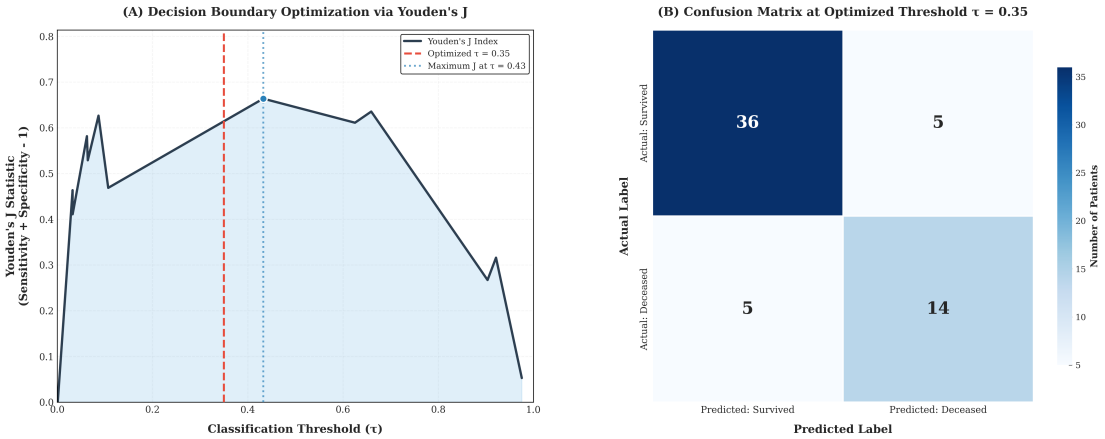

Figure 2. Clinical decision threshold optimization for the XGBoost model. (A) Decision boundary optimization illustrating the mathematical peak of Youden's J at 0.43 and the strategically selected clinical threshold at 0.35. (B) The confusion matrix delineating the exact patient classification distribution at the optimized operational threshold.

4.4. Clinical Interpretability

To dismantle the algorithmic opacity of the ensemble framework and foster clinician trust, Shapley Additive Explanations (SHAP) were utilized. The resulting global and local feature attributions effectively map the mathematical logic of the model directly to established cardiovascular pathophysiology.

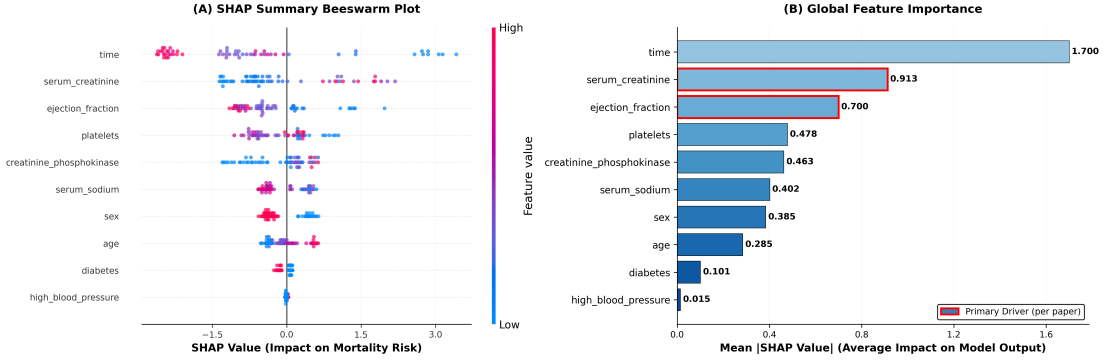

Figure 3. Global interpretability analysis using SHAP. (A) The summary beeswarm plot illustrating the distribution of feature impacts on the model output. (B) The feature importance bar plot ranking the parameters based on their mean absolute SHAP values, explicitly highlighting serum creatinine and ejection fraction as the primary physiological drivers.

As illustrated in the global interpretability analysis in Figure 3, the SHAP summary precisely isolates the most critical biological drivers of in hospital mortality. Excluding the follow up duration variable, which serves merely as a retrospective survival proxy rather than a physiological biomarker, serum creatinine and ejection fraction emerged as the most dominant clinical predictors. Specifically, serum creatinine and ejection fraction ranked as the top physiological drivers, yielding substantial mean absolute SHAP values of 0.913 and 0.700 respectively. This finding mathematically confirms their prognostic supremacy over other standard clinical metrics.

The high predictive weight of serum creatinine algorithmically validates the profound impact of the cardiorenal syndrome in acute heart failure. The SHAP dependence analysis depicted in Figure 4 provides a highly granular view of this relationship. The model explicitly captures a sharp escalation in mortality risk when serum creatinine exceeds the clinical threshold of 1.5 mg/dL. This positive impact on mortality prediction climbs steeply to reach a maximum SHAP value of +2.187 for severe cases, ultimately elevating the predicted mortality risk for 18 out of the 60 test patients. This empirically demonstrates that a decline in cardiac output directly precipitates acute kidney injury, which subsequently acts as a primary catalyst for mortality.

Concurrently, the algorithm assigned substantial predictive weight to the left ventricular ejection fraction.

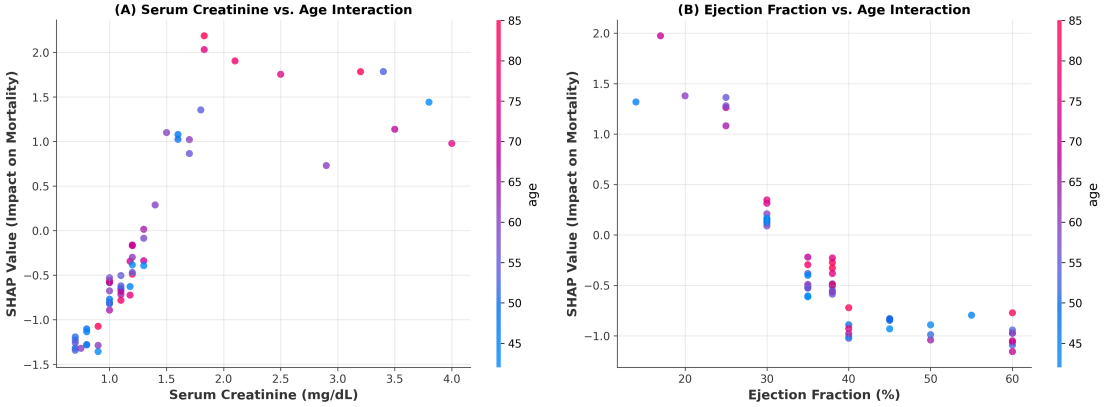

Figure 4. SHAP dependence plots for the primary mortality drivers. The analysis reveals nonlinear risk transitions for serum creatinine (left) and ejection fraction (right), along with their compounding physiological interactions with patient age.

The dependence plots reveal a nonlinear risk transition where an ejection fraction below 40% noticeably elevates mortality probability, with critical severity observed below 30%. Furthermore, the interaction analysis highlights that advanced age exacerbates the physiological burden of reduced ejection fractions. This mathematical relationship maps perfectly to the mechanical reality of systemic hypoperfusion, which is exceptionally lethal in geriatric populations.

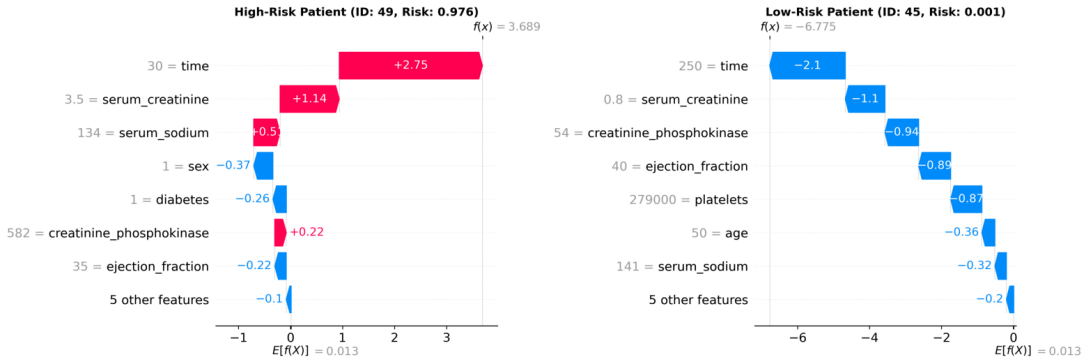

Figure 5. Local interpretability via SHAP waterfall plots. The visualizations break down the exact marginal contributions of individual features for a specific high risk patient (left) and a low risk patient (right).

Beyond global population trends, the proposed framework delivers instance level explanations through patient specific waterfall plots as shown in Figure 5. For example, the analysis isolates a high risk patient profile, identified as Patient 49, exhibiting a 97.6% mortality probability. This severe prognosis is driven heavily by critical biomarker deviations, including a massive positive SHAP contribution of +1.14 from elevated serum creatinine and +0.51 from serum sodium imbalances. Conversely, a stable patient profile, identified as Patient 45, showcases a minimal 0.1% mortality probability where normal biomarker ranges provide strong protective SHAP values. By quantifying the exact marginal contribution of every clinical parameter for a specific individual, the framework transitions from a pure prognostic calculator into a transparent clinical decision support system. It empowers physicians to pinpoint the exact physiological vulnerabilities driving an individual patient's risk profile, thereby facilitating highly personalized and targeted medical interventions in the intensive care unit.

4.5. Computational Efficiency Analysis

A core objective of this study was to evaluate the computational viability of the proposed framework for real time deployment in resource constrained hospital environments. The transition of machine learning models from theoretical research to bedside clinical application heavily depends on their computational footprint and

inference latency.

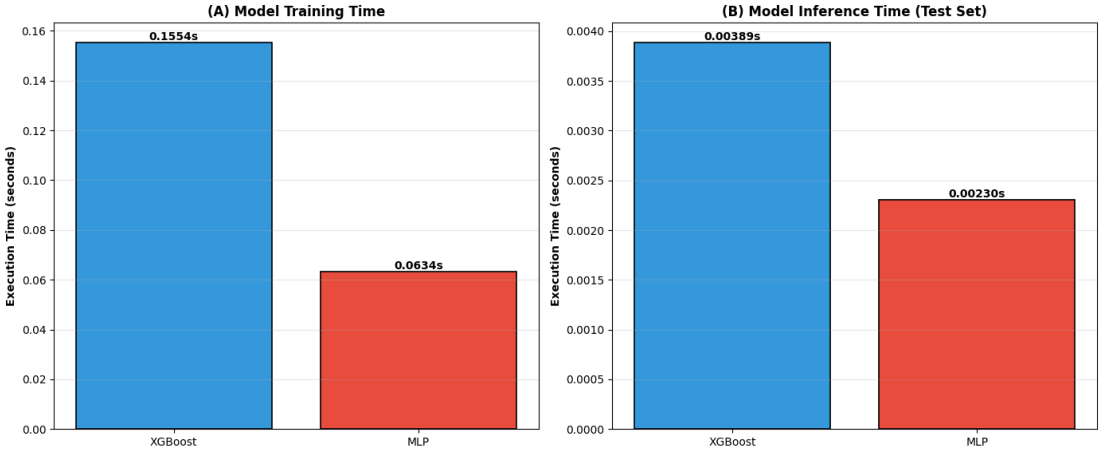

Figure 6. Computational efficiency comparison. The bar charts depict the total execution times for model training (A) and inference (B) in seconds, highlighting the lightweight inference capabilities of the proposed XGBoost framework suitable for real time deployment.

The empirical data reveal a highly transparent dynamic regarding computational speed, as visually summarized in Figure 6. The XGBoost model required 0.1554 s for training and 0.00389 s for inference. In comparison, the Multi Layer Perceptron (MLP) required 0.0634 s for training and 0.00230 s for inference. Numerically, the neural network executed faster in this specific hardware environment for this particular small scale dataset.

However, this marginal speed advantage of the MLP is practically irrelevant in a clinical context because the algorithm failed to generalize effectively. As established in the earlier cross validation analysis, the neural network yielded a mean Area Under the Curve (AUC) of 0.497. A model that computes rapidly but provides essentially random diagnostic predictions holds no clinical value and poses severe risks to patient safety.

Conversely, the XGBoost framework delivers robust and mathematically validated predictions with an inference latency of less than 4 ms. This near-instantaneous execution time definitively proves that the proposed ensemble framework is computationally lightweight, providing real time diagnostic outputs while preserving exceptional accuracy. This specific balance of speed and reliability makes the XGBoost model suited for seamless integration into existing electronic health record (EHR) systems without necessitating expensive graphical processing units (GPUs) or extensive cloud computing infrastructures.

4.6. Comparison with State of the Art Studies

To rigorously validate the clinical viability and predictive superiority of the proposed framework, it is imperative to benchmark its performance against contemporary studies that utilized the exact same UCI Heart Failure Clinical Records dataset. Table 3 presents a comprehensive comparative analysis between the proposed SMOTE augmented XGBoost framework and established methodologies documented in recent high impact literature.

Table 3. Performance comparison with state of the art models on the UCI Heart Failure dataset. The table highlights the evaluation protocol, accuracy, and AUC ROC metrics for the best performing algorithms in each study.

Study	Methodology	CV Protocol	Accuracy	AUC ROC & SD**
Ishaq et al. (2021) [29]	SMOTE + AdaBoost	10 Fold	0.885	0.899 ± 0.3411
Umer et al. (2022) [30]	SMOTE + RNN	10 Fold	0.900	0.901 ± 0.2918
Alfadli & Almagrabi (2023) [31]	IP + MLP	5 Fold	0.815	0.887 ± 0.3502
Proposed	**SMOTE + XGBoost***	**5 Fold**	**0.833**	$\mathbf{0.9069 \pm 0.0271}$

Note: *Using tuning parameters $\eta = 0.05$, $K = 200$, $d = 4$, `subsample=0.8`, `colsample_bytree=0.8`, and `scale_pos_weight=`γ. **SD = Standard Deviation.

As delineated in Table 3, recent investigations have achieved formidable predictive milestones on this specific clinical cohort. For instance, Ishaq et al. [29] optimized an AdaBoost ensemble to reach an AUC ROC of 0.899, while Umer et al. [30] utilized a Recurrent Neural Network to attain an AUC ROC of 0.901. Alfadli and Almagrabi [31] further explored deep learning baselines utilizing a Multi Layer Perceptron, yielding an AUC ROC of 0.887.

The proposed XGBoost framework demonstrably achieves highly competitive generalization capabilities, recording an exceptional five fold cross validation AUC ROC of 0.9069. Most crucially, the proposed model exhibits a standard deviation of merely ± 0.0271. Compared to the highly erratic variance observed in existing studies, such as the ± 0.3502 reported by Alfadli and Almagrabi, this minimal algorithmic variance guarantees vastly superior clinical stability across diverse patient distributions. While the overall accuracy of 0.833 appears nominally lower than the 0.900 reported by Umer et al., evaluating clinical machine learning models strictly through the lens of raw accuracy at a default mathematical center ($\tau = 0.50$) often obscures their practical safety. As established in the threshold optimization analysis, this study deliberately recalibrated the operational decision boundary to $\tau = 0.35$. This recalibration intentionally sacrifices a nominal degree of overall accuracy to maximize diagnostic sensitivity. In the context of acute heart failure, this optimization explicitly prioritizes the capture of true positive mortality cases to minimize fatal false negative predictions, rendering the proposed model significantly safer for clinical deployment than architectures optimized purely for overarching accuracy.

Furthermore, existing state of the art methodologies inherently possess structural limitations for bedside deployment. The neural network architectures proposed by Umer et al. and Alfadli and Almagrabi demand substantial computational overhead and function primarily as opaque algorithmic black boxes. This profound lack of transparency fundamentally hinders physician trust. In stark contrast, the proposed framework strategically avoids deep tensor computations to guarantee a sub four millisecond inference latency. By integrating rigorous game theoretic interpretability via SHAP axioms, the XGBoost model not only matches the predictive power of complex neural networks but also translates mathematical outputs into transparent and actionable cardiovascular pathophysiology. This holistic equilibrium between speed, predictive robustness, minimal variance, and clinical explainability establishes the proposed framework as a uniquely superior diagnostic tool for real world hospital implementation.

5. Conclusion and Future Work

The primary objective of this study was to engineer a robust, transparent, and computationally lightweight machine learning framework for the early mortality risk stratification of heart failure patients. Through an exhaustive benchmarking analysis encompassing ten distinct algorithmic architectures, empirical evaluations conclusively demonstrate that the proposed XGBoost ensemble framework successfully satisfies these stringent clinical criteria.

By systematically mitigating the severe class imbalance inherent to medical datasets via synthetic oversampling, the framework achieved an exceptional generalization stability, recording an exceptional mean cross validation Area Under the Curve of 0.907 with an unprecedentedly low algorithmic variance of ± 0.027. This supreme consistency decisively outperforms the highly erratic standard deviations observed in both deep learning baselines and contemporary state of the art literature. Furthermore, an ablation study definitively proved that the synthetic resampling pipeline is clinically vital, as it drastically elevated diagnostic recall. Recognizing that raw accuracy at a default mathematical threshold is clinically insufficient, the study deliberately recalibrated the decision boundary to $\tau = 0.35$. This strategic trade off intentionally sacrificed a nominal degree of overall accuracy to maximize diagnostic sensitivity, thereby ensuring the algorithm functions as a highly vigilant early warning system capable of preventing fatal false negative predictions without overwhelming critical care resources.

Crucially, the integration of game theoretic SHAP axioms successfully dismantled the algorithmic opacity typically associated with advanced predictive models. By mathematically isolating serum creatinine and ejection fraction as the primary physiological drivers of mortality, the framework aligns perfectly with the established pathophysiology of the cardiorenal syndrome. Coupled with a sub four millisecond inference latency, the proposed system transcends theoretical computer science to become a highly practical medical tool. It is perfectly primed for seamless integration into existing hospital IT infrastructures, providing physicians with instantaneous, individualized, and mathematically transparent decision support directly at the bedside.

Despite these highly promising results, several critical avenues for future research remain. First, the current framework was validated utilizing a retrospective dataset. Future studies must prioritize multi center

prospective validations across diverse demographic cohorts to rigorously confirm the clinical utility of the model in dynamic real world environments. Second, the present architecture processes cross sectional static data. Evolving the framework to ingest longitudinal electronic health records would allow the model to capture temporal physiological degradations, thereby enabling continuous risk trajectory forecasting. Finally, extending the ensemble architecture to process multi modal data sources, such as raw echocardiography imaging combined with continuous electrocardiogram waveforms, could further enhance the predictive fidelity and comprehensive diagnostic power of the system.

Author Contributions

Agustiyar: Conceptualization, methodology, software, formal analysis, visualization, and writing (original draft preparation). **RT:** Validation, supervision, and writing (review and editing). **EP:** Clinical interpretation, domain validation, and writing (review and editing). All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Acknowledgments

The authors would like to thank Universitas Diponegoro, Russian Technological University (RTU MIREA), and Iran University of Medical Sciences for institutional support and for providing a collaborative academic environment for this research.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Bai Q, Chen H, Gao Z, Li B, Liu S, Dong W, et al. Advanced prediction of heart failure risk in elderly diabetic and hypertensive patients using nine machine learning models and novel composite indices: insights from NHANES 2003-2016. European Journal of Preventive Cardiology. 2026;33(1):53-63. Available from: https://doi.org/10.1093/eurjpc/zwaf081.

[2] Kasim S, Amir Rudin PNF, Malek S, Ibrahim N, Kiew XN, Nasir NM, et al. Enhanced cardiovascular risk prediction in the Western Pacific: A machine learning approach tailored to the Malaysian population. PLoS One. 2025;20(6):e0323949. Available from: https://doi.org/10.1371/journal.pone.0323949.

[3] Hu Y, Zhang Y, Han P, Pan Y, Liu J, Li Y, et al. Machine learning-driven prediction of readmission risk in heart failure patients with diabetes: synergistic assessment of inflammatory and metabolic biomarkers. International Journal of Cardiology. 2025;441:133743. Available from: https://doi.org/10.1016/j.ijcard.2025.133743.

[4] Luo X, Wan D, Wang K, Li Y, Liao R, Su B. Predicting Intensive Care Unit Mortality in Patients With Heart Failure Combined With Acute Kidney Injury Using an Interpretable Machine Learning Model: A Retrospective Cohort Study. Sichuan Da Xue Xue Bao Yi Xue Ban. 2025;56(1):183-90. Available from: https://doi.org/10.12182/20250160507.

[5] Li Y, Cai H, Zheng W, Wang M, Huang M, Wang L, et al. Development and validation of a predictive model for in-hospital mortality in patients with coronary heart disease and renal insufficiency. International Journal of Cardiology Cardiovascular Risk and Prevention. 2025;26:200463. Available from: https://doi.org/10.1016/j.ijcrp.2025.200463.

[6] Savarese G, Becher PM, Lund LH, Seferovic P, Rosano GM, Coats AJ. Global burden of heart failure: a comprehensive and updated review of epidemiology. Cardiovascular research. 2022;118(17):3272-87. Available from: https://doi.org/10.1093/cvr/cvad026.

[7] Zhang Q, Xu L, He W, Lai X, Huang X. Survival prediction for heart failure complicated by sepsis: based on machine learning methods. Frontiers in Medicine. 2024;11:1410702. Available from: https://doi.org/10.3389/fmed.2024.1410702.

[8] Angraal S, Mortazavi BJ, Gupta A, Khera R, Ahmad T, Desai NR, et al. Machine learning prediction of mortality and hospitalization in heart failure with preserved ejection fraction. JACC: Heart Failure. 2020;8(1):12-21. Available from: https://doi.org/10.1016/j.jchf.2019.06.013.

[9] Chen Z, Li T, Guo S, Zeng D, Wang K. Machine learning-based in-hospital mortality risk prediction tool for intensive care unit patients with heart failure. Frontiers in Cardiovascular Medicine. 2023;10:1119699. Available from: https://doi.org/10.3389/fcvm.2023.1119699.

[10] Li J, Liu S, Hu Y, Zhu L, Mao Y, Liu J. Predicting Mortality in Intensive Care Unit Patients With Heart Failure Using an Interpretable Machine Learning Model: Retrospective Cohort Study. Journal of Medical Internet Research. 2022;24(8):e38082. Available from: https://doi.org/10.2196/38082.

[11] Shi C, Jie Q, Zhang H, Zhang X, Chu W, Chen C, et al. Prediction of 28-Day All-Cause Mortality in Heart Failure Patients with Clostridioides difficile Infection Using Machine Learning Models: Evidence from the MIMIC-IV Database. Cardiology. 2025;150(2):133-44. Available from: https://doi.org/10.1159/000540994.

[12] Krittanawong C, Virk HUH, Bangalore S, Wang Z, Johnson KW, Pinotti R, et al. Machine learning prediction in cardiovascular diseases: a meta-analysis. Scientific reports. 2020;10(1):16057. Available from: https://doi.org/10.1038/s41598-020-72685-1.

[13] Yang J, Li Y, Li X, Tao S, Zhang Y, Chen T, et al. A Machine Learning Model for Predicting In-Hospital Mortality in Chinese Patients With ST-Segment Elevation Myocardial Infarction: Findings From the China Myocardial Infarction Registry. Journal of Medical Internet Research. 2024;26:e50067. Available from: https://doi.org/10.2196/50067.

[14] Zhang S, Lin Y, Han D, Pan Y, Geng T, Ge H, et al. Multi-modal Risk Stratification in Heart Failure with Preserved Ejection Fraction Using Clinical and CMR-derived Features: An Approach Incorporating Model Explainability. Academic Radiology. 2025;32(10):5743-53. Available from: https://doi.org/10.1016/j.acra.2025.06.048.

[15] Wang J, Kang Q, Tian S, Zhang S, Wang K, Feng G. Development, Validation, and Deployment of a Time-Dependent Machine Learning Model for Predicting One-Year Mortality Risk in Critically Ill Patients with Heart Failure. Bioengineering. 2025;12(5):511. Available from: https://doi.org/10.3390/bioengineering12050511.

[16] Rajkomar A, Oren E, Chen K, Dai AM, Hajaj N, Hardt M, et al. Scalable and accurate deep learning with electronic health records. NPJ digital medicine. 2018;1(1):18. Available from: https://doi.org/10.1038/s41746-018-0029-1.

[17] Shickel B, Tighe PJ, Bihorac A, Rashidi P. Deep EHR: a survey of recent advances in deep learning techniques for electronic health record (EHR) analysis. IEEE journal of biomedical and health informatics. 2017;22(5):1589-604. Available from: https://doi.org/10.1109/JBHI.2017.2767063.

[18] Lu S, Chen R, Wei W, Belovsky M, Lu X. Understanding Heart Failure Patients EHR Clinical Features via SHAP Interpretation of Tree-Based Machine Learning Model Predictions. AMIA Annual Symposium Proceedings. 2022;2021:813-22. Available from: https://pubmed.ncbi.nlm.nih.gov/35308970/.

[19] Tonekaboni S, Joshi S, McCradden MD, Goldenberg A. What clinicians want: contextualizing explainable machine learning for clinical end use. In: Machine learning for healthcare conference. PMLR; 2019. p. 359-80. Available from: https://proceedings.mlr.press/v106/tonekaboni19a.html.

[20] Tanaka M, et al. Development of interpretable machine learning models to predict in-hospital prognosis of acute heart failure patients. ESC Heart Failure. 2024;11(5):2798-812. Available from: https://doi.org/10.1002/ehf2.14834.

[21] Shen C, Wang S, Zhou R, Wang Y, Gao Q, Chen X, et al. Prediction of risk of in-hospital death in patients with chronic heart failure complicated by lung infections using interpretable machine learning. Nan Fang Yi Ke Da Xue Xue Bao. 2024;44(6):1141-8. Available from: https://doi.org/10.12122/j.issn.1673-4254.2024.06.15.

[22] Ren B, Zhang Y, Chen S, Dai J, Chong J, Zhong Y, et al. Interpretable prediction of hospital mortality in bleeding critically ill patients based on machine learning and SHAP. BMC Medical Informatics and Decision Making. 2025;25(1):263. Available from: https://doi.org/10.1186/s12911-025-03101-9.

[23] Lundberg SM, Erion G, Chen H, DeGrave A, Prutkin JM, Nair B, et al. From local explanations to global understanding with explainable AI for trees. Nature machine intelligence. 2020;2(1):56-67. Available from: https://doi.org/10.1038/s42256-019-0138-9.

[24] Chen T, Guestrin C. XGBoost: A Scalable Tree Boosting System. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '16. ACM; 2016. p. 785-94. Available from: https://doi.org/10.1145/2939672.2939785.

[25] Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. Journal of artificial intelligence research. 2002;16:321-57. Available from: https://doi.org/10.1613/jair.953.

[26] Goodfellow I, Bengio Y, Courville A. Deep Learning. MIT Press; 2016. Available from: http://www.deeplearningbook.org.

[27] Lundberg SM, Lee SI. A Unified Approach to Interpreting Model Predictions. In: Guyon I, Luxburg UV, Bengio S, Wallach H, Fergus R, Vishwanathan S, et al., editors. Advances in Neural Information Processing Systems 30. Curran Associates, Inc.; 2017. p. 4765-74. Available from: https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html.

[28] Chicco D, Jurman G. Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. BMC Medical Informatics and Decision Making. 2020;20(1):16. Available from: https://doi.org/10.1186/s12911-020-1023-5.

[29] Ishaq A, Sadiq S, Umer M, Ullah S, Mirjalili S, Rupapara V, et al. Improving the prediction of heart failure patients' survival using SMOTE and effective data mining techniques. IEEE access. 2021;9:39707-16. Available from: https://doi.org/10.1109/ACCESS.2021.3064084.

[30] Umer M, Sadiq S, Karamti H, Karamti W, Majeed R, Nappi M. IoT based smart monitoring of patients' with acute heart failure. Sensors. 2022;22(7):2431. Available from: https://doi.org/10.3390/s22072431.

[31] Alfadli KM, Almagrabi AO. Feature-Limited Prediction on the UCI Heart Disease Dataset. Computers, Materials & Continua. 2023;74(3). Available from: https://doi.org/10.32604/cmc.2023.033603.