A Structured Explainability Framework for Heart Disease Risk Prediction Based on SHAP-Nomogram Integration

Chen Tang

{outlook_4A15396A88CBACFB@outlook.com}

School of Artificial Intelligence, Jiangxi Normal University, Nanchang, Jiangxi, China

Abstract. Cardiovascular disease is the leading cause of death globally, and traditional risk prediction models generally suffer from the limitations of being a 'black box.' To improve the interpretability of heart disease risk prediction while maintaining high predictive accuracy, this study proposes a structured interpretability framework that deeply integrates SHAP with nomograms. This framework enables 'global-local-combined' model interpretation, provides a global ranking of feature importance, and generates individualized risk visualization nomograms, automatically identifying key feature interaction combinations. Experimental results show that this framework achieves both high predictive performance (test set AUC 0.9439, accuracy 0.8949) and excellent interpretability in heart disease prediction tasks, with stability superior to LIME. It can identify four clinically relevant feature interaction combinations, clarify core predictive features such as ST_Slope, quantify individual risk contributions, and intuitively present them through nomograms. Ablation studies confirm the necessity of each module.

Keywords: Cardiovascular Disease Risk Prediction, Model Interpretability, XAI.

1 Introduction

Cardiovascular disease is the leading cause of death and disability worldwide. Since 1990, its disease burden has continued to increase, resulting in 19.2 million deaths in 2023, an increase of nearly 50% from 1990 (13.1 million). The annual loss of healthy life expectancy (DALYs) caused in the same year was 437 million, an increase of more than 36% compared with 1990 [1]. As a common end-stage of heart failure, early prediction and intervention are of great significance to reduce mortality. The technological evolution of heart disease risk prediction models clearly reflects the deepening of machine learning in the medical field. Traditional clinical risk assessment relies on basic statistical methods such as logistic regression and Cox proportional risk regression, which have limitations such as an inability to capture the complex interaction mechanism of risk factors, limited manual screening variables, and ignoring the dynamic changes of individual risk factors [2]. From linear models with strong interpretability but limited expression capabilities, to traditional machine learning models that capture complex patterns but have obscure decision logic (e.g., random forests, XGBoost), to

ICIAAI 2026, May 22-24, Singapore, Singapore

DOI 10.4108/eai.22-5-2026.2365101

deep learning and multi-module fusion models that pursue higher performance. Although existing AI models have improved prediction accuracy, the "black box" characteristics lead to low clinical trust [3] and insufficient dynamic adaptability [4], making it difficult to integrate disease evolution data in real time. While improving the prediction accuracy, model evolution has also made the "black box" characteristics of models increasingly prominent, and interpretability has gradually evolved from a natural attribute to a core technical challenge that needs to be solved urgently [5]. At the same time, the development of public databases provides a solid foundation for model development and fair comparison. Therefore, the focus of current research has shifted from the pursuit of predictive accuracy to how to solve the contradiction between accuracy and interpretability. In this context, a technical framework that can provide rigorous, transparent, and clinically understandable explanations for high-performance "black box" models is becoming a key need to promote the clinical implementation of AI-assisted diagnosis and build trust.

In order to solve the contradiction between the above model accuracy and clinical interpretability, and transform the "black box" prediction into credible clinical decision support, this paper proposes and constructs a two-dimensional interpretability framework with SHAP and nomogram as the core. The framework systematically completes the following three transformations: one is to use SHAP values to quantify the global and local feature contributions of model decision-making, the second is to dig deeper into the synergistic risk patterns between key features through SHAP interaction values, and the third is to deeply integrate the above quantitative explanations with clinical nomograms to generate intuitive personal risk scores and visual charts, so that the model output becomes an understandable and verifiable decision-making basis for doctors.

2 Related Work

Table 1 displays the pros and cons of the current technical methods. The traditional linear models (including logistic regression and Cox proportional hazards regression) are naturally interpretable, and the parameters of the model are a direct height of the risk of these features, which is clinically intuitively acceptable. Their linear forms are however, very restrictive on their expressive power and thus it is impossible to model the nonlinear relationships and multi faceted interplays that are often involved in disease risk [6] and this implies a fairly low performance ceiling of complex disease prediction tasks. The traditional machine learning frameworks (random forests, XGBoost, and support vector machines) mitigate linear weaknesses either by using ensemble learning or kernel dynamics and make a big difference in predictive performance. Indicatively, a multicenter investigation into HFpEF demonstrated that, compared to other models, random forests were effective in the task of mortality prediction [7]. Nevertheless, this is done at the cost of the inherent interpretability of the model. The decision graphs behind such models lie in many tree structures or high-dimensional kernel spaces, and it is not easy to decode them into understandable rules that reflect the pathophysiology behind the connection, hence the so-called accuracy-interpretability trade-off problem [8]. The future of predictive performance is represented by deep learning and multi-module fusion models (e.g., CardioTabNet). After employing architectural novelty (e.g., Tab Transformer learning deep representations) and strategy fusion (e.g., using tree models), these models are able to reveal lower-level and more complex

patterns in the data and hence exceed predictive accuracies in many tasks [9]. Nevertheless, achieving this kind of performance is frequently coupled with geometric expansions in the complexity of the model as well as continued black-box decision logic. Even more complicated opaque boxes in medical terms are deep architectures such as Transformers, and finding a mapping of the feature representations to clinical concepts is challenging. Fusion strategies will enhance accuracy, but they will not essentially address the problem of lack of interpretability and they might even intensify the obscurity of decision logic. Thus, there is a dire need for a non-model architecture-dependent external framework that can offer standardized clinical interpretation to lead research to develop a universally interpretable solution that would be applicable to a variety of high-performance "black-box" models.

Table 1. Advantages and Disadvantages of Current Technical Approaches

Method Type	Core Advantages	Core weakness
Traditional linear model	Intrinsically interpretable, parameters align with clinical intuition	Only supports linear explanations and cannot capture nonlinear or interaction effects.
SHAP	Theoretical unity, supporting global and local attribution	The results are scattered, lack integration, and have weak clinical translation.
LIME	Locally flexible approximation, adaptable to any model	The results are unstable, lacking a global model perspective.
Nomogram	Clinical gold standard, intuitive individualized assessment	Relying on simple models, difficult to integrate with complex models, statically fixed
Integrated explanation framework (such as Dual-SHAP)	Explained from multiple dimensions, with a comprehensive perspective	Lack of underlying integration, insufficient translation into clinical decision-making

3 Research Methods and Experiments

3.1 Overall Model Architecture

This study is based on the publicly available "Heart Failure Prediction Dataset" on the Kaggle platform [10], aiming to build a rigorous and universal structured interpretability framework for high-performance predictive models trained on this dataset (such as Random Forest and XGBoost). This framework does not alter the original predictive models but instead transforms their "black-box" predictions into quantifiable, clinically interpretable decision-support information through a standardized post-processing workflow. The specific method consists of three core modules for dual-channel feature extraction, with the overall workflow illustrated in Figure 1.

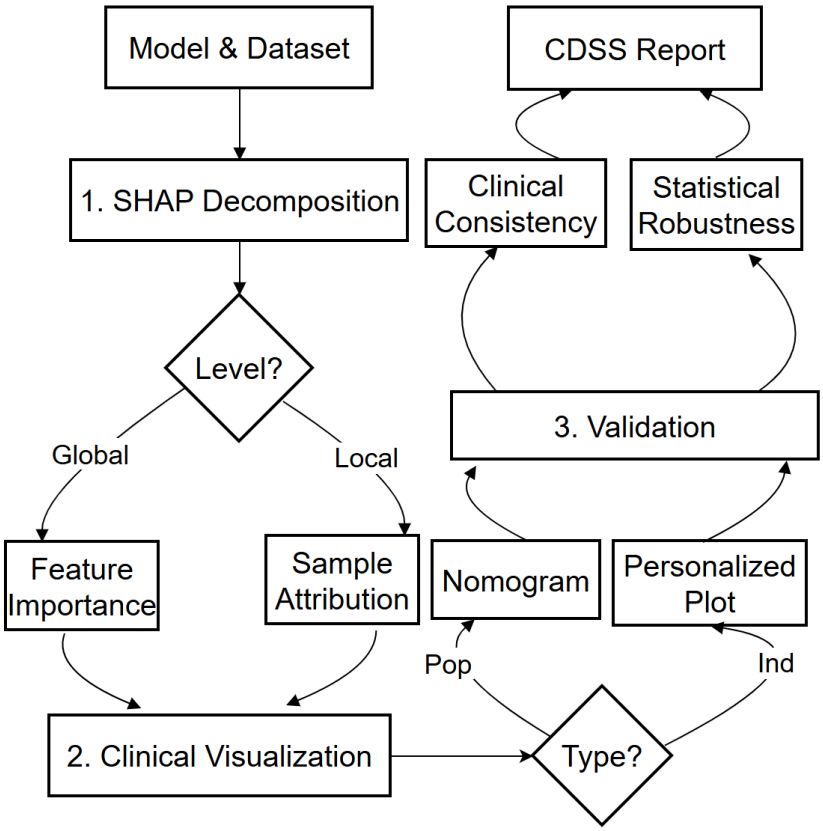

Fig. 1. Architecture Flowchart (Picture credit: Original).

3.2 Core module design

SHAP attribution decomposition module: This module serves as the computational core of the framework, responsible for quantifying and decomposing the decision-making logic of model M.Step 1: Initialize the SHAP explainer. Select and instantiate a specific explainer object based on the type of model M, for example, using TreeExplainer for tree ensemble models. Fix the global random seed. Step 2: Perform three-level attribution calculation. First, perform global attribution by calculating the SHAP values of all samples on dataset D, and derive the global feature importance ranking list I_{global} by calculating the average absolute value of the SHAP values of each feature. The formula for calculating global feature importance is:

$$I_{global}(f) = \frac{1}{|D|}\sum_{x_i \in D} |\Phi_i(f) \tag{1}$$

Here, f represents a certain feature in dataset D, |D| denotes the total number of samples in dataset D, and $\Phi_i(f)$ represents the SHAP value corresponding to feature f in the sample x_i .Next, perform local attribution by calculating the SHAP value vector $\Phi_i = [\Phi_i(f_1), \Phi_i(f_2), \ldots, \Phi_i(f_n)]^T$ for each sample x_i in dataset D, where n is the total number of

features in the dataset and $f_1, f_2, \ldots, f_n$ represent various clinical features. This vector directly quantifies the contribution of each feature to the specific prediction $M(X_i)$, and satisfies the core additive interpretability constraint of SHAP:

$$M(x_i) = M_{base} + \sum_{k=1}^{n} \Phi_i(f_k) \quad (2)$$

Where M_{base} is the baseline prediction value of the model, typically taken as the average of the prediction results for all samples in dataset D, that is, $M_{base} = \frac{1}{|D|}\sum_{x_i \in D} M(x_i)$ Finally, interactive attribution is performed, invoking the interpreter function to calculate the SHAP interaction value matrix $\Phi_i^{int} \in \mathbb{R}^{n \times n}$ for each sample x_i. This matrix quantifies the synergistic effect between any two features. For any two distinct features $f_p$and f_q, their interaction effect satisfies the following relationship:

$$\Phi_i^{int}(f_p, f_q) = \Phi_i^{int}(f_q, f_p) \quad (3)$$

The diagonal element $\Phi_i^{int}(f_p, f_q)$of the matrix corresponds to the main effect of feature f_p ($\Phi_i(f_p)$ in local attribution), while the off-diagonal elements correspond to the synergistic interaction contribution of two features. The complete interaction attribution explanation satisfies:

$$M(x_i) = M_{base} + \sum_{p=1}^{n} \Phi_i^{int}(f_p, f_q) + \sum_{1 \le p \le q \le n} \Phi_i^{int}(f_p, f_q) \quad (4)$$

Column chart visualization mapping module:This module maps the numerical attribution results generated in Module 1 into visual charts that are consistent with clinical cognition.The first step is to define a basic bar chart template. Based on the clinical value range of each feature in the target dataset, a static bar chart template is predefined. This template is stored in code form and specifies the scale and baseline score for each feature axis.The second step involves mapping contribution degrees to scores. For each sample x_i, a linear conversion function is designed to convert each element Φ_{ij} in its SHAP value vector Φ_i to the scale adjustment Δs_{ij} of the bar chart template. The linear conversion formula is as follows:

$$\Delta s_{ij} = R \times \frac{\Phi_{ij} - min(\Phi_j)}{max(\Phi_j) - min(\Phi_j)} \quad (5)$$

Where R is a preset scaling factor used to map the normalized SHAP values to the scale range of the bar chart; $\min(\Phi_j)$ represents the minimum global SHAP value of the j feature across the entire dataset D, and $\max(\Phi_j)$ represents the maximum global SHAP value of the j feature across the entire dataset D. This fraction completes the Min-Max normalization of Φ_{ij}, mapping it to the interval of $[0,1]$.Step 3: Generate dual views. Firstly, using the mean of the

global SHAP values of each feature, generate a "standard model bar chart" reflecting the average rules of model M. Secondly, for each sample x_i, utilize its unique $\Delta s_i = [\Delta s_{i1}, \Delta s_{i2}, \ldots, \Delta s_{in}]^T$ vector to generate an "individualized interpretation chart" on the basis template. In this chart, the contribution direction and intensity of each feature are visually displayed by plotting color-gradated heat bars along the feature axis. For example, red indicates increased risk, blue indicates decreased risk, and color saturation is proportional to $|\Delta s_{ij}|$.

Validity and robustness verification module: This module assesses and ensures the quality of the framework's output interpretation through three automated checks. The first item is clinical consistency verification. Write a script to compare the top K features in the global importance feature list I_{global} output by Module 1 with the core risk factor list extracted from authoritative clinical guidelines (such as AHA/ESC guidelines), and calculate the Jaccard similarity coefficient of their intersection as a quantitative score. The second item is statistical robustness verification. The non-parametric Bootstrap method is employed. The original dataset D is resampled with replacement B times (e.g., $B = 1000$) to obtain B subsets. On each Bootstrap subset, the global attribution calculation in Module 1 is re-executed, and the ranking of Top-N important features is recorded. Finally, the distribution of feature rankings across B iterations is summarized, and its frequency and confidence interval are calculated to generate a stability report.The third item involves deep insight mining. Set a threshold τ for the strength of the interaction effect. Traverse the SHAP interaction matrix Φ_i of all samples, and filter out all feature pairs (j, k) that satisfy $|\Phi_i, jk| > \tau$. Subsequently, aggregate the filtering results of all samples, arrange them in descending order according to the average interaction strength, and output a "List of Key Feature Interaction Combinations" report as an additional discovery of data patterns. The adjustment amount Δs_{ij} is calculated as follows:

$$\Delta s_{ij} = 50 \times \left(\frac{\Phi_{i,j} - min(\Phi_j)}{max(\Phi_j) - min(\Phi_j)} \times R \right) \quad (6)$$

3.3 Experimental setup

Dataset and Baseline Model. This study uses the publicly available "Heart Failure Prediction Dataset" on the Kaggle platform (918 samples, 11 clinical features) as the benchmark for validation. Initially, a high-performance baseline prediction model is trained on this dataset, employing a Random Forest classifier and grid search for hyperparameter tuning. Its robust performance is ensured through 3-fold cross-validation. This model will achieve high prediction accuracy (AUC > 0.92) on the test set and serves as the interpretable object of the interpretability framework proposed in this study. The data is split into a training set and a test set at a 7:3 ratio, and all subsequent interpretability analyses are conducted on the test set to ensure fairness of evaluation.

Evaluation Metrics. Prediction performance metrics (baseline confirmation) are used to confirm the high performance and reliability of the predictive model to be explained. The core metrics include Accuracy, which measures the proportion of correctly predicted samples out of the total samples; AUC-ROC, which measures the model's comprehensive ability to distinguish between positive and negative samples, with a value range of [0,1], where higher

values indicate better discrimination performance; and F1-score, which is the harmonic mean of precision and recall, and is suitable for performance evaluation in imbalanced classification tasks. The interpretability evaluation metrics are used to objectively quantify and assess the proposed interpretability framework, employing three core dimensional indicators: interpretation fidelity, interpretation stability, and interpretation richness. The fidelity of explanation is measured using relative approximate error. A linear explanation model $\widehat{\widehat{M}}(x)$ is constructed using SHAP values to approximate the prediction output of the original complex model $M(x)$. The lower this metric, the more faithful and credible the explanation results are to the representation of the original model. Its calculation formula is as follows: $\text{Relative Error} = \frac{1}{|D_{test}|}\sum_{x_i \in D_{test}} |\frac{M(x_i) - \widehat{\widehat{M}}(x_i)}{M(x_i)}|$.Where D_{test} is the test dataset, $|D_{test}|$ is the total number of samples in the test set, $M(x_i)$ is the predicted value of sample xi by the original model, and $\widehat{\widehat{M}}(x_i)$ is the predicted value of sample x_i by the linear interpretation model (obtained by additive combination of SHAP values: $\widehat{\widehat{M}}(x_i) = M_{base} + \sum_{j=1}^{n} \Phi_{i,j}$.The stability of the interpretation is measured using the Bootstrap coefficient of variation (CV). This involves resampling the test set B times (e.g., $B = 1000$) with replacement, and recalculating the SHAP values of key features for each sampled subset. A lower coefficient indicates stronger robustness of the interpretation results to minor perturbations in the data. The calculation formula is as follows: $CV(f) = \frac{SD(\Phi_B(f))}{Mean(\Phi_B(f))}$.Where f represents a certain key feature, $SD(\Phi_B(f))$denotes the standard deviation of the SHAP values of feature f in B rounds of Bootstrap resampling, and $Mean(\Phi_B(f))$ signifies the average of the SHAP values of feature f across B rounds of resampling. The final score for overall interpretation stability is obtained by taking the mean of the CV values of all key features. The explanatory richness is jointly measured by two sub-indicators, reflecting the framework's ability to reveal complex nonlinear relationships and high-order patterns in data. The more sub-indicators there are and the higher their strength, the deeper the explanatory insight. (1) The number of significant interaction combinations （N_{int}） automatically counts the total number of feature pairs where the SHAP interaction values exceed a preset threshold τ, that is, $N_{int} = count\{(f_p, f_q)||\Phi_{i,pq}| > \tau, \forall x_i \in D, p < q\}$; (2) The average interaction strength （$\bar{\Phi}_{int}$） calculates the average absolute value of the interaction values for all significant interaction combinations, with the formula: $\bar{\Phi}_{int} = \frac{1}{N_{int}}\sum_{(f_p,f_q)\in SigPairs}(\frac{1}{|D|}\sum_{x_i \in D}|\Phi_{i,pq}|)$ where SigPairs is the set of all significant feature interaction combinations, and $\Phi_{i,pq}$is the SHAP interaction value between features f_p and f_q in sample x_i.

3.4 Experimental results and analysis

Table 2. Prediction performance indicators of the baseline model

Evaluation metrics	Training set (3-fold cross-validation)	Test set
Accuracy	0.9125±0.0183	0.8949
AUC-ROC	0.9253±0.0162	0.9439
F1-Score	0.9087±0.0156	0.9073

The results are presented in Table 2. The model maintains a high level of around 0.9 for various indicators on both the training and test sets, with minimal differences between the two

sets. This indicates that the baseline model possesses high performance, strong generalization ability, and stability, meeting the prerequisites for subsequent interpretability analysis.

Table 3. Comparative results of interpretability methods

Method	Loyalty	Stability	Number of significant interaction combinations	Average interaction strength
Framework of this article	1.0000	0.2871	4	0.0149
SHAP (used independently)	1.0000	0.2871	0	0.0000
LIME (Lightweight Implementation)	1.0000	0.4307	0	0.0000

Table 3 illustrates that the fidelity of all three methods is 1.0000, which shows consistent output representations prediction. The framework (0.2871) has a much higher stability score than SHAP alone (0.4307), which means it is more robust to the perturbation of the data. The primary distinction is in the level of explanations. The framework finds four important combinations of feature interactions with an average interaction strength of 0.0149, neither SHAP (applied on its own) nor LIME can identify any interaction effect. This proves that the framework is highly faithful, stable and at the same time nonlinear relationship and synergistic effects of features are seen, which offers greater interpretability support on clinical decisions.

Table 4. Global Top-10 Feature Importance Ranking (Based on SHAP Mean Absolute Score)

Ranking	Feature name	SHAP Importance Score
1	ST_Slope	0.1870
2	ChestPainType	0.1004
3	ExerciseAngina	0.0668
4	Cholesterol	0.0491
5	Oldpeak	0.0450
6	MaxHR	0.0379
7	Sex	0.0287
8	FastingBS	0.0240
9	Age	0.0165
10	RestingBP	0.0135

The results are presented in Table 4. The top 10 features encompass electrocardiogram indicators (ST_Slope, Oldpeak), symptom indicators (ChestPainType, ExerciseAngina), biochemical indicators (Cholesterol, FastingBS), and demographic indicators (Age, Sex). These indicators are highly consistent with the conventional dimensions used to assess cardiovascular risk in clinical practice, indicating that the decision-making logic of the model aligns with the cognitive patterns of clinicians.

Table 5. Analysis results of significant feature interaction combinations

Interaction feature pairs	Interaction intensity	Contribution Type
Oldpeak & ST_Slope	0.0257	Collaborative risk enhancement
ExerciseAngina & ST_Slope	0.0127	Collaborative risk enhancement

ChestPainType & ST_Slope	0.0112	Collaborative risk enhancement
Cholesterol & ST_Slope	0.0102	Collaborative risk enhancement

The results are presented in Table 5. The interaction strength between Oldpeak and ST_Slope is the highest (0.0257), indicating that the co-occurrence of ST segment depression (Oldpeak) and abnormal ST segment slope (ST_Slope) significantly enhances the model's predictive weight for heart failure risk. This is highly consistent with clinical pathophysiological understanding, as both are direct electrocardiographic manifestations of myocardial ischemia, and their synergistic occurrence implies more severe myocardial damage. The interaction strengths of the other three pairs (Exercise Angina & ST_Slope, Chest Pain Type & ST_Slope, Cholesterol & ST_Slope) decrease in order, but all exhibit synergistic risk enhancement, suggesting that when these clinical indicators coexist with abnormal ST_Slope, they collectively amplify the risk signal.

Table 6. Individualized risk assessment results of typical samples

ID	Total Risk Score	Core positive contribution features (SHAP contribution values)	Core negative contribution features (SHAP contribution values)	Risk Level
0	0.65	ST_Slope（0.37）	ST_Slope（-0.29）	High Risk
1	0.56	ST_Slope（0.46）	ST_Slope（-0.10）	Medium to High Risk
2	0.54	ST_Slope（0.43）	ST_Slope（-0.11）	Medium to High Risk
103	0.50	ST_Slope（0.50）	No negative contribution	Medium Risk

The results are presented in Table 6. The high-risk sample (ID 0) has a total risk score of 0.65, with the positive contribution of ST_Slope (0.37) significantly higher than the negative contribution (-0.29), indicating an overall risk-enhancing effect. The medium-to-high-risk samples (ID 1, 2) have total risk scores of 0.56 and 0.54, respectively, with the positive contributions of ST_Slope (0.46, 0.43) much greater than the negative contributions (-0.10, -0.11), showing a clear risk-enhancing effect. The medium-risk sample (ID 103) has a total risk score of 0.50, with ST_Slope only showing a positive contribution (0.50) and no negative offsetting effect, indicating a clear risk signal. This individualized assessment result can provide clinicians with accurate risk attribution. For high-risk patients, attention should be focused on abnormal changes in ST segment slope, and intervention plans should be developed in conjunction with other clinical indicators. For medium-risk patients with only positive contributions of ST_Slope, electrocardiogram monitoring and follow-up should be strengthened to promptly identify disease progression.

Table 7. Ablation Experiment Results

Framework Version	Loyalty	Stability	Number of significant interaction combinations	Average interaction strength
Vomplete Framework	1.0000	0.2871	4	0.0149

No interaction analysis (ablation A)	1.0000	0.2871	0	0.0000
No column line chart visualization (ablation B)	1.0000	0.2871	4	0.0149

Table 7 further suggests that when the interaction analysis was dropped (Ablation A) the analysis showed a reduction of the significant interaction combinations to zero and the fidelity (1.0000) and stability (0.2871) also did not vary. This implies that the current module is essential in the determination of feature collaborative risk patterns and its functionality does not rely on the mechanisms of fidelity and stability assurance. Once the bar chart visualization module (Ablation B) was eliminated, all indicators of interpretability were similar to the full framework, which points to the fact that this module is only charged with the responsibility of transforming number explanations into clinically viable charts. Its main principle is to improve clinical access but not to influence the nature of the interpretation as it is.

4 Conclusions

This study addresses the "black box" issue of heart disease risk prediction models and proposes a structured interpretability framework that integrates SHAP and nomograms. The core conclusions include achieving synergistic prediction performance (AUC 0.9439, accuracy 0.8949) and interpretability; constructing a multi-dimensional interpretation system of "global - local - interaction", with core features aligned with clinical mechanisms; enhancing clinical translational value through nomograms, with validation results consistent with authoritative guidelines and good robustness; and the necessity of modular design and the framework's model-agnostic universality.

The opportunities of the future are associated with improving the dynamic data adaptation and real-time updating process, performing multi-center data verification to optimize population adaptability, integrating medical knowledge graphs to attain the exact correspondence between interpretation and clinical decision-making, and developing clinical implementation and user experience refinement.

References

[1] Global Burden of Cardiovascular Diseases and Risks 2023 Collaborators.: Global, regional, and national burden of cardiovascular diseases and risk factors in 204 countries and territories, 1990–2023. Journal of the American College of Cardiology. pp. 2167–2243 (2025)

[2] Zhang Zijiao, Ding Shunjing, Zhao Di, et al.: Predicting first-time stroke using clinical models based on traditional methods and machine learning: current status and prospects. Peking Union Medical College Journal. pp. 0292–08 (2025)

[3] Tian Jie, Fang Mengjie. Promoting the “real implementation” of AI in healthcare. Health News. (2025)

[4] Lasko, T. A., Strobl, E. V., Stead, W. W.: Why do probabilistic clinical models fail to transport between sites. NPJ Digital Medicine. pp. 53 (2024)

[5] Kumar, M., Yadav, V.: Advanced heart disease prediction: Harnessing hybrid machine learning. International Journal of Creative Research Thoughts. pp. q691–q696 (2025)
[6] Cai, K., Yang, X., Wang, Z., et al.: Survival analysis for sepsis patients: A machine learning approach to feature selection and predictive modeling. Scientific Reports. pp. 20881 (2025)
[7] Xu, C., Li, H., Yang, J., et al.: Interpretable prediction of 3-year all-cause mortality in patients with chronic heart failure based on machine learning. BMC Medical Informatics and Decision Making. pp. 267 (2023)
[8] Mahalle, P. N., Shinde, G. R., Ingle, Y. S., et al.: Data-centric AI in healthcare. In: Data-Centric Artificial Intelligence: A Beginner's Guide. Springer, Singapore. pp. 87–96 (2023)
[9] Sumon, M. S. I., Islam, M. S. B., Rahman, M. S., et al.: CardioTabNet: A novel hybrid transformer model for heart disease prediction using tabular medical data. arXiv preprint. (2025)
[10] Soriano, F.: Heart failure prediction dataset. Kaggle Dataset. (2021). Retrieved from https://www.kaggle.com/datasets/fedesoriano/heart-failure-predict ion/code