Received: 1 October 2025 | Revised: 10 November 2025 | Accepted: 30 November 2025 | Published online: 12 December 2025

RESEARCH ARTICLE

Medinformatics
2026, Vol. 3(3) 275–285
DOI: 10.47852/bonviewMEDIN52027812

BON VIEW PUBLISHING

Advancing Interpretable AI for Cardiovascular Risk Assessment: A Stacking Regression Approach in Clinical Data from Bangladesh

Suhana Tasnim[1], Mohammad Mamun[2,3], Safiul Haque Chowdhury[2,*], Mohammed Ibrahim Hussain[3] and Muhammad Minoar Hossain[3]

[1] *Department of Life Sciences, Independent University Bangladesh, Bangladesh*

[2] *Department of Computer Science and Engineering, Jahangirnagar University, Bangladesh*

[3] *Department of Computer Science and Engineering, Bangladesh University, Bangladesh*

Abstract: Cardiovascular diseases (CVDs) are complex conditions affecting a large portion of the global population, and their early, accurate, and timely prediction remains a significant challenge. Conventional CVD risk assessment often relies on limited parameters and fails to capture the complex interactions among genetic, lifestyle, and environmental factors. Recent machine learning studies have improved predictive performance; however, they often rely on small or retrospective datasets, lack real-time or external validation, and offer limited interpretability for clinical use. This study introduces a novel stacking ensemble framework that integrates ridge regression (RR), Theil-Sen regressor, and gradient boosting regressor. To our knowledge, this is the first application of a regression-based stacking approach for CVD risk prediction that embeds explainable artificial intelligence as a core component, a combination rarely explored in low-resource healthcare contexts. Using a real-world dataset of 1,529 patients from Jamalpur Medical College Hospital, Bangladesh, the proposed model achieved 96% predictive accuracy, outperforming most existing methods. The dataset itself represents a rare contribution, as most prior studies rely on UCI, Framingham, or other benchmark repositories rather than contemporary hospital data from underrepresented populations. Through SHapley Additive exPlanations analysis, our model identifies BMI, diabetes, and blood pressure as the most influential factors, aligning with established medical knowledge and providing clinically actionable insights. Unlike prior black-box models, our framework not only improves prediction accuracy but also delivers transparent explanations that enhance trust and support public health decision-making. This integration of accuracy, explainability, and context-specific clinical insight underscores the novelty and practical relevance of our approach for advancing interpretable AI in CVD prediction, particularly in resource-limited healthcare settings.

Keywords: cardiovascular diseases, machine learning, ensemble model, ridge regressor, Theil-Sen regressor, gradient boosting regressor, explainable AI

1. Introduction

Cardiovascular diseases (CVDs) are the leading cause of mortality worldwide, accounting for an enormous health and economic burden. Among these, heart attacks and strokes represent approximately 85% of CVD-related deaths, making them a priority concern for global health systems [1]. The significant risk factors associated with CVD include smoking, excessive alcohol consumption, unhealthy diets, sedentary lifestyles, and hypertension, which often occur together and increase the likelihood of adverse outcomes [2]. Traditional treatment methods, such as medications, lifestyle modifications, and surgical interventions, have been widely employed to manage symptoms and mitigate risk [3]. In critical cases, invasive procedures like angioplasty or bypass surgery are performed to restore blood circulation and reduce arterial blockages [4]. Additionally, growing attention has been given to herbal remedies and plant-derived compounds for their potential cardioprotective benefits [5]. Despite these efforts, the global statistics remain alarming, with nearly 20 million CVD-related deaths reported in 2022 alone, representing almost one-third of all deaths worldwide [6]. In low- and middle-income countries (LMICs) such as Bangladesh, the burden is increasing rapidly due to demographic aging, urbanization, and lifestyle transitions. Recent epidemiological data indicate that hypertension, diabetes, and obesity are rising across both rural and urban populations, yet early risk detection and preventive interventions remain inadequate. The limited availability of structured, high-quality clinical data in regional hospitals further constrains the ability to develop reliable prediction tools tailored to local populations [7].

Machine learning (ML) methods have shown significant promise in improving cardiovascular risk prediction through data-driven modeling. However, several persistent challenges limit their

***Corresponding author:** Safiul Haque Chowdhury, Department of Computer Science and Engineering, Jahangirnagar University, Bangladesh. Email: 4169safiulhaque@juniv.edu

clinical translation. Most prior studies have relied on benchmark or Western datasets (e.g., UCI, Framingham), which do not reflect the demographic and environmental diversity of LMICs. Moreover, while many models achieve high accuracy, they often function as "black boxes," providing little insight into how predictions are made, an essential aspect for clinical trust and adoption. There is also limited work exploring hybrid or ensemble-based regression models that integrate interpretability tools to ensure transparency in risk attribution [8, 9].

To address these limitations, the present study develops an explainable stacking ensemble framework that integrates ridge regression (RR), Theil-Sen regressor (TSR), and gradient boosting regressor. Using a real-world dataset from Jamalpur Medical College Hospital, Bangladesh, this study aims to improve prediction accuracy while providing transparent explanations of influential risk factors through SHAP (SHapley Additive exPlanations) analysis. This dual focus on performance and interpretability makes the approach both technically robust and clinically meaningful. Furthermore, by leveraging contemporary hospital data from an underrepresented population, this work contributes novel regional evidence to the global literature on AI-driven cardiovascular risk prediction.

The main objectives of this study are therefore threefold:

1) To construct a comprehensive and representative dataset for CVD risk prediction in a low-resource clinical setting
2) To develop and validate a regression-based stacking ensemble framework that enhances predictive performance
3) To integrate explainable AI techniques to interpret model outputs and identify the most influential risk factors
4) Together, these contributions aim to support the development of interpretable, context-specific AI tools for preventive cardiovascular healthcare.

2. Literature Review

To identify the research gap, we conducted a comprehensive review of existing methods relevant to this study. This section highlights the contributions, outcomes, limitations, and future directions of prior works. Dritsas and Trigka [10] employed logistic regression (LR), achieving an accuracy of 87.8% in predicting CVD risk. Khan et al. [11] utilized random forest (RF), which achieved the highest accuracy of 85.01% with the lowest misclassification error. However, a key limitation was the absence of real-time clinical validation; the authors recommended incorporating deep learning models to enhance accuracy. RF has been widely used due to its robustness, but it often struggles with imbalanced CVD datasets, where minority cases (e.g., positive diagnoses) are underrepresented. Its reliance on bagging and majority-vote decision trees can bias predictions toward the dominant class, leading to higher false negatives and lower sensitivity in clinical contexts. Bhatt et al. [12] proposed an ML framework for CVD prediction, reporting that the multilayer perceptron (MLP) achieved the highest accuracy of 87.28% with cross-validation. Nevertheless, the study did not consider temporal trends or genetic factors, which are crucial for a comprehensive assessment of CVD risk. The authors suggested that future work incorporate deep learning, such as convolutional neural networks (CNNs), to improve predictive performance. Chandrasekhar and Peddakrishna [13] reviewed the use of ML in enhancing heart disease risk prediction. They used six ML classifiers and an ensemble soft voting method (SVE). The ensemble model achieved an accuracy of 93.44%; however, the study used only 302 cases, which may not be sufficient for generalizing to diverse populations. Kanagarathinam et al. [14] employed a CatBoost ML classifier and achieved a mean accuracy of 94.34% which was validated via 10-fold cross-validation. CatBoost outperformed other models due to its gradient boosting (GB) framework, which is tailored to handle categorical and heterogeneous data. Unlike RF, CatBoost uses ordered boosting and efficient encoding of categorical features, enabling it to capture complex, nonlinear interactions among risk factors (e.g., age, diabetes, hypertension) while minimizing overfitting on small datasets. Islam et al. [15] developed an integrated system combining the Internet of Things (IoT) with LR to predict CVD risk levels, achieving an F1-score of 91% for binary classification and 80.4% for ternary classification. While the prototype demonstrated high accuracy and usability, future efforts should address hardware limitations, data heterogeneity, and broader clinical integration to maximize impact. Stonier et al. [16] developed a model to predict CVD risk using RF, achieving an accuracy of 88.52%. However, the study's small sample size of 301 patients limits its generalizability, indicating a need for larger datasets. Huang et al. [17] applied five ML models to predict cardiovascular risk in middle-aged and elderly Chinese populations. Among these, the LightGBM (LGB) model achieved the highest accuracy of 81.7%. However, its F1-score was relatively low at 0.509, indicating a high false-negative rate. Future studies should prioritize improving model sensitivity and the F1-score. Zaidi et al. [18] developed HeartEnsembleNet, a hybrid ensemble learning approach for predicting CVD risk. Their hybrid model HeartEnsembleNet with Hybrid Random Forest Linear Models (HRFLM), which combined RF and k-Nearest Neighbors, achieved the highest accuracy of 92.25%. Despite this strong performance, the model lacks interpretability. Incorporating explainable AI techniques such as SHAP or Local Interpretable Model-agnostic Explanations (LIME) would be essential to enhance clinical trust. Cheng et al. [19] evaluated multiple ML models for predicting CVD risk in Taiwanese adults. Among these, GB achieved the highest accuracy of 76.2% and an F1-score of 56.7%. However, the study's reliance on self-reported outcomes and absence of external validation limit its immediate clinical applicability. Future work should emphasize validation, interpretability, and integration of diverse data sources.

Table 1 highlights significant progress in CVD risk prediction using a wide range of ML models, including LR, RF, GB, LGB, CatBoost, deep learning architectures such as MLP, and hybrid

Table 1. Summary of the existing methods

Author	ML Model	CVD Risk Accuracy	Dataset Size	Limitation	Future Direction
Dritsas and Trigka [10]	LR	87.8%	Not specified	–	–
Khan et al. [11]	RF	85.01%	Not specified	No real-time clinical validation	Explore deep learning models

(*Continued*)

Table 1. (*Continued*)

Author	ML Model	CVD Risk Accuracy	Dataset Size	Limitation	Future Direction
Bhatt et al. [12]	MLP	87.28%	Not specified	Ignored temporal/genetic factors	Apply CNN/deep learning
Chandrasekhar and Peddakrishna [13]	SVE	93.44%	Small (302)	Limited generalizability	Use larger datasets
Kanagarathinam et al. [14]	CatBoost	94.34%	Not specified	–	–
Islam et al. [15]	IoT + LR	F1: 91% (binary), 80.4% (ternary)	Prototype, heterogeneous IoT data	Hardware/data heterogeneity	Improve scalability and integration
Stonier et al. [16]	RF	88.52%	Small (301)	Limited generalizability	Larger datasets
Huang et al. [17]	LGB	81.7%	Not specified	Low F1-score (0.509)	Improve sensitivity
Zaidi et al. [18]	HRFLM	92.25%	Not specified	Lack of explainability	Apply SHAP/LIME
Cheng et al. [19]	GB	76.2%	Not specified	Reliance on self-reported data	Emphasize validation

frameworks like HRFLM. Reported accuracies vary from 76.2% to 94.34%, with ensemble and hybrid models (e.g., CatBoost, SVE, HRFLM) generally outperforming traditional approaches. While these results are promising, several studies reporting accuracies above 90% have common challenges that remain. Many works suffer from small sample sizes (e.g., < 310 patients), reliance on retrospective or self-reported datasets, lack of external or real-time clinical validation, and limited consideration of diverse features such as genetic, temporal, or longitudinal data. Furthermore, some models struggle with poor sensitivity (e.g., LGB, with an F1-score of 0.509) or lack transparency (e.g., HRFLM), which limits their clinical trustworthiness. Hardware and data heterogeneity issues have also been noted as barriers to scalability. Overall, despite progress, most models have yet to demonstrate robust generalizability across broader populations or real-world clinical deployment.

3. Materials and Methodology

The primary objective of this research is to predict the level of risk associated with CVD based on patients' information using an ML-based approach. To achieve the aim of this research, several regression analysis models were applied, and the best models were utilized in a Stacking method as the primary technique. Figure 1 provides an overview of the core architecture of this research, and Sections 3.1 to 3.7 describe the steps in detail.

The methodological design of this study was guided by two primary objectives: (1) to construct a robust and generalizable predictive framework for CVD risk using real-world hospital data from Bangladesh and (2) to ensure that the resulting model is transparent and clinically interpretable. To achieve these aims, we developed an innovative stacking ensemble that integrates RR, TSR, and GBR, three models with complementary strengths. This combination allows for the capture of both linear and nonlinear relationships while maintaining resistance to outliers and reducing overfitting.

The methodological innovation of this approach lies in the integration of regression-based stacking with explainable artificial intelligence (XAI) through SHAP analysis. Unlike prior works that treat CVD prediction as a classification problem using opaque "black-box" algorithms, our framework employs continuous risk estimation through regression, offering finer granularity and clinical relevance. The inclusion of SHAP interpretability further enhances transparency by quantifying the contribution of each risk factor, bridging the gap between data science and medical decision-making.

3.1. Dataset

In this study, the dataset was obtained from a comprehensive CVD risk assessment available on Kaggle [20]. The dataset comprises information on 1,529 patients collected from Jamalpur Medical College Hospital in Jamalpur, Bangladesh, between January 20, 2024, and January 1, 2025. The dataset consists of 22 columns, of which 21 are predictive variables and 1 is the decision variable. Both numeric and nominal data were used in this research. The dataset is described in more detail in Table 2.

3.2. Data refining and preprocessing

In the initial preprocessing stage, missing values in the dataset were addressed through data imputation. Data imputation is a method used to replace missing or incomplete entries with estimated values based on existing data, often through statistical or computational approaches [21]. A total of 948 missing values were identified, all within numerical features, and these were imputed using the mean of the respective variables. The original blood pressure (mmHg) column, which contained paired systolic and diastolic readings, was removed to avoid redundancy and multicollinearity. Instead, these values were transformed into a single categorical variable representing hypertension stages (e.g., normal, stage 1, stage 2, and stage 3). This approach preserved the clinical relevance of blood pressure while simplifying the dataset and improving model interpretability.

Following this, nominal variables were converted into numerical form to make them suitable for analysis. For example, in the sex column, female entries were encoded as 1 and male entries as

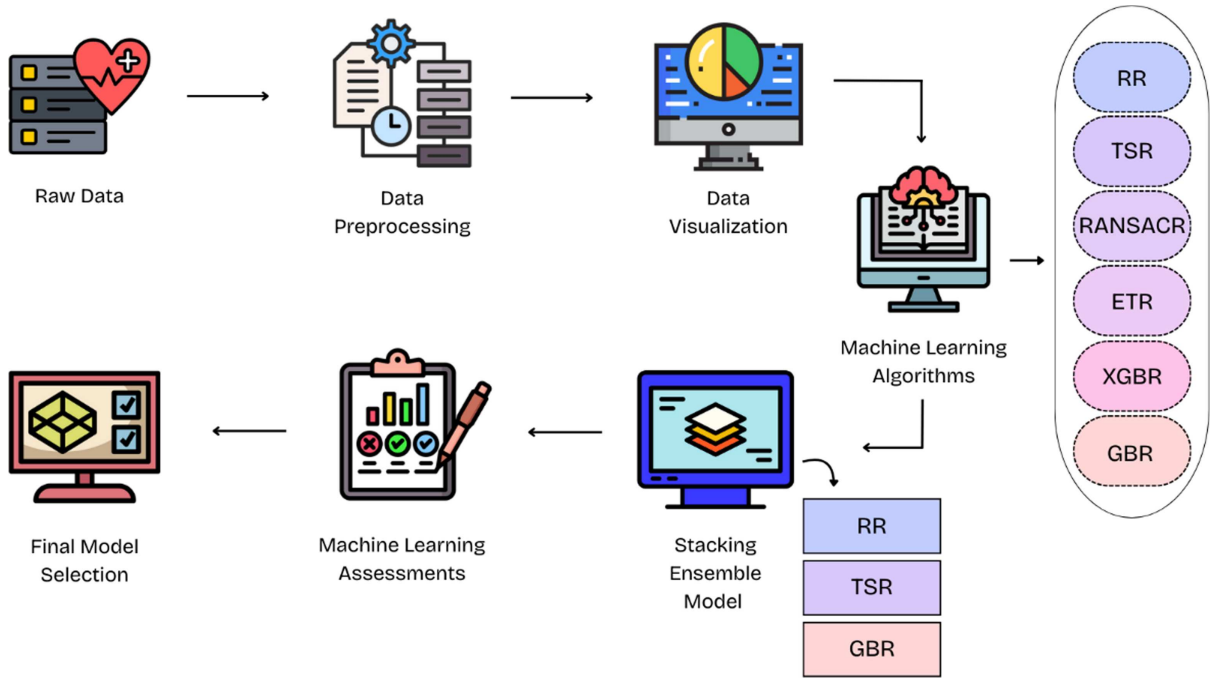

Figure 1. The core framework of the applied approach

Table 2. An in-depth description of the dataset used

Parameter	Description	Type
Sex	Biological sex of the individual (male/female)	Nominal
Age	Age in years	Numerical
Weight (kg)	Body weight measured in kilograms	Numerical
Height (m)	Height measured in meters	Numerical
BMI	Body mass index, calculated as weight/height^2	Numerical
Abdominal circumference (cm)	Waist circumference in centimeters	Numerical
Blood pressure (mmHg)	Recorded blood pressure (systolic/diastolic)	Nominal
Total cholesterol (mg/dL)	Total cholesterol concentration in the blood	Numerical
HDL (mg/dL)	High-density lipoprotein cholesterol level	Numerical
Fasting blood sugar (mg/dL)	Blood glucose level after fasting	Numerical
Smoking status	Whether the individual smokes	Nominal
Diabetes status	Whether the individual has diabetes	Nominal
Physical activity level	Level of physical activity (low/medium/high)	Nominal
Family history of CVD	Presence of cardiovascular disease in family history	Nominal
CVD risk level	Categorical risk level (low/intermediary/high)	Nominal

0. Similarly, for smoking status, diabetes status, and family history of CVD, a "Yes" response was encoded as 1 and a "No" as 0. For ordinal variables describing severity levels, numerical values from 1 to 4 were assigned. For instance, patients with normal blood pressure were assigned a value of 1, while those with hypertension stage 2 were assigned a value of 4. The same encoding approach was applied to physical activity level and CVD risk level.

3.3. Data visualization

This research examines various visualizations of the dataset to understand the trends and patterns in the data. For this research, histograms and density plots were merged and used for the data visualization.

3.3.1. Histogram and density plot

A histogram is a statistical chart that visualizes the frequency distribution of numeric data by dividing it into contiguous intervals (bins) and representing the count of data points within each bin as vertical bar [22]. When augmented with a density plot, a smoothed representation of the distribution estimated using kernel density estimation, the resulting visualization provides a comprehensive view of both the absolute frequencies and the underlying probability density function [23]. In the context of CVD risk modeling, these combined charts for features such as age, BMI, total cholesterol, fasting blood sugar, LDL, and smoking status

offer critical insight into data characteristics, including skewness, modality, and dispersion. Furthermore, they allow for the visual comparison of distributional characteristics across subgroups, such as sex-based differences, highlighting shifts in central tendency and variability. Figure 2 showcases these distributions, guiding essential preprocessing decisions, informing feature engineering, and supporting the development of predictive models that account for demographic heterogeneity to improve predictive accuracy and generalizability in CVD risk assessment.

3.3.2. Heatmap

Figure 3 presents a correlation matrix heatmap for the analysis of CVD risk. A heatmap provides a visual representation of correlation coefficients between variables, using a color gradient to indicate the strength and direction of relationships [24]. Warmer colors denote stronger positive correlations, while cooler colors represent stronger negative correlations. This visualization enables the identification of linear dependencies among variables such as BMI, total cholesterol, blood pressure, and LDL, which can reveal potential multicollinearity issues in predictive modeling. Recognizing these interrelationships supports more informed feature selection and model design, ultimately improving the accuracy and interpretability of CVD risk prediction models.

3.4. Machine learning algorithms

In this research, we applied a range of ML algorithms to predict CVD risk and evaluated their performance using multiple metrics. The models that achieved the highest accuracy were RR [25], TSR [26], RANSAC regressor (RANSACR) [27], extra trees regressor (ETR) [28], XGBoost regressor (XGBR) [29], and GBR [30]. To ensure the reliability of these results, we also performed 5-fold and 10-fold cross-validation, which helped identify the models that generalized best to unseen data. Based on these findings, we constructed an ensemble model using the stacking method. In this approach, selected high-performing models were used as base learners, and a final meta-model combined their predictions to produce the overall output. This strategy enabled us to leverage the complementary strengths of different algorithms, thereby improving prediction accuracy compared to individual models.

3.5. Stacking ensemble model

Ensemble stacking is an ML technique that combines the predictions of multiple base models through a secondary model, known as a meta-learner. The base models are trained on the same dataset, and their outputs are used as inputs to the meta-learner, which learns how to integrate them optimally. This approach leverages the complementary strengths of different algorithms, often achieving higher predictive accuracy and robustness than any individual model alone [31]. In this study, the stacking ensemble combined three models: RR, TSR, and GBR. Each model was first trained separately on the same training data, and their predictions were then combined into a final Ridge Regression model, which learned how to optimize its outputs to make the final prediction. To further enhance performance, the original input features were also provided to the final model alongside the base model predictions, giving it additional information to improve accuracy.

3.6. Machine learning assessments

The performance of regression models is typically evaluated using a variety of metrics, including mean squared error (MSE), mean absolute error (MAE), coefficient of determination (R^2), peak signal-to-noise ratio and signal-to-noise ratio (SNR) [32, 33]. Table 3 summarizes the results obtained across these evaluation measures. Collectively, these metrics provide a comprehensive view of a model's predictive accuracy and its ability to

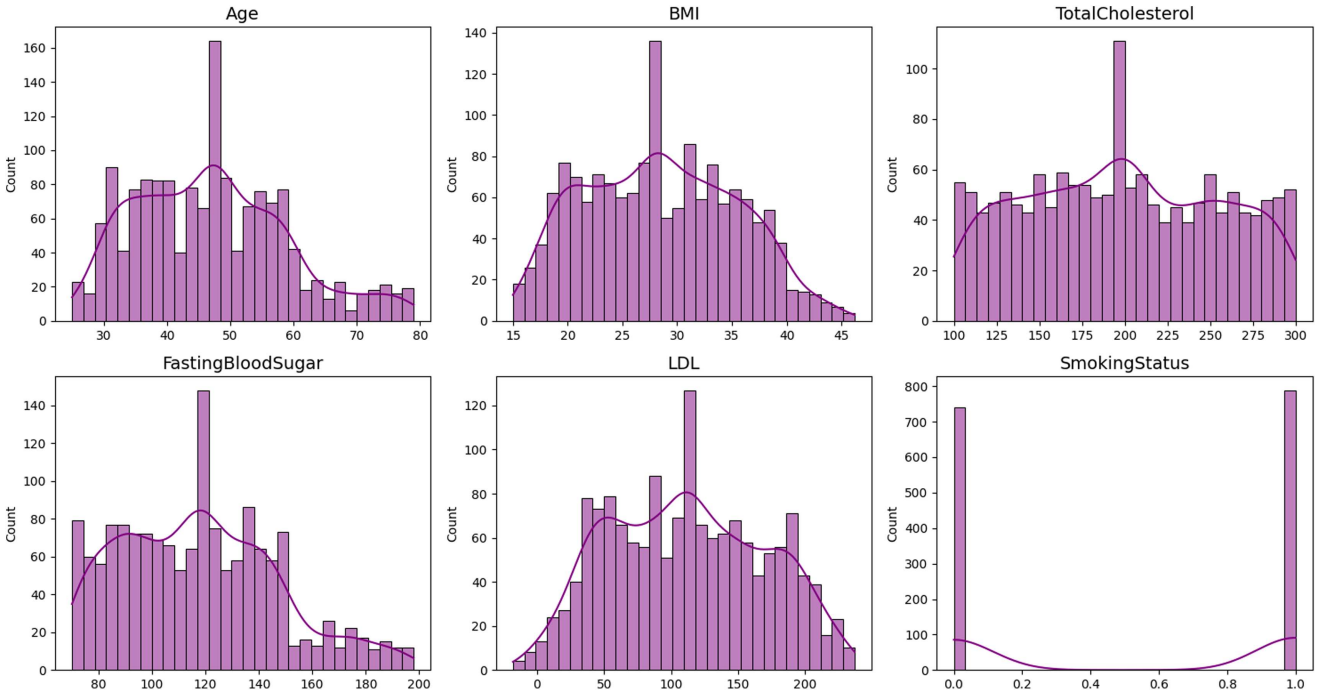

Figure 2. A merge of histograms and density plots showing distribution in the following variables: (a) age, (b) BMI, (c) total cholesterol, (d) fasting blood sugar, (e) LDL, and (f) smoking status

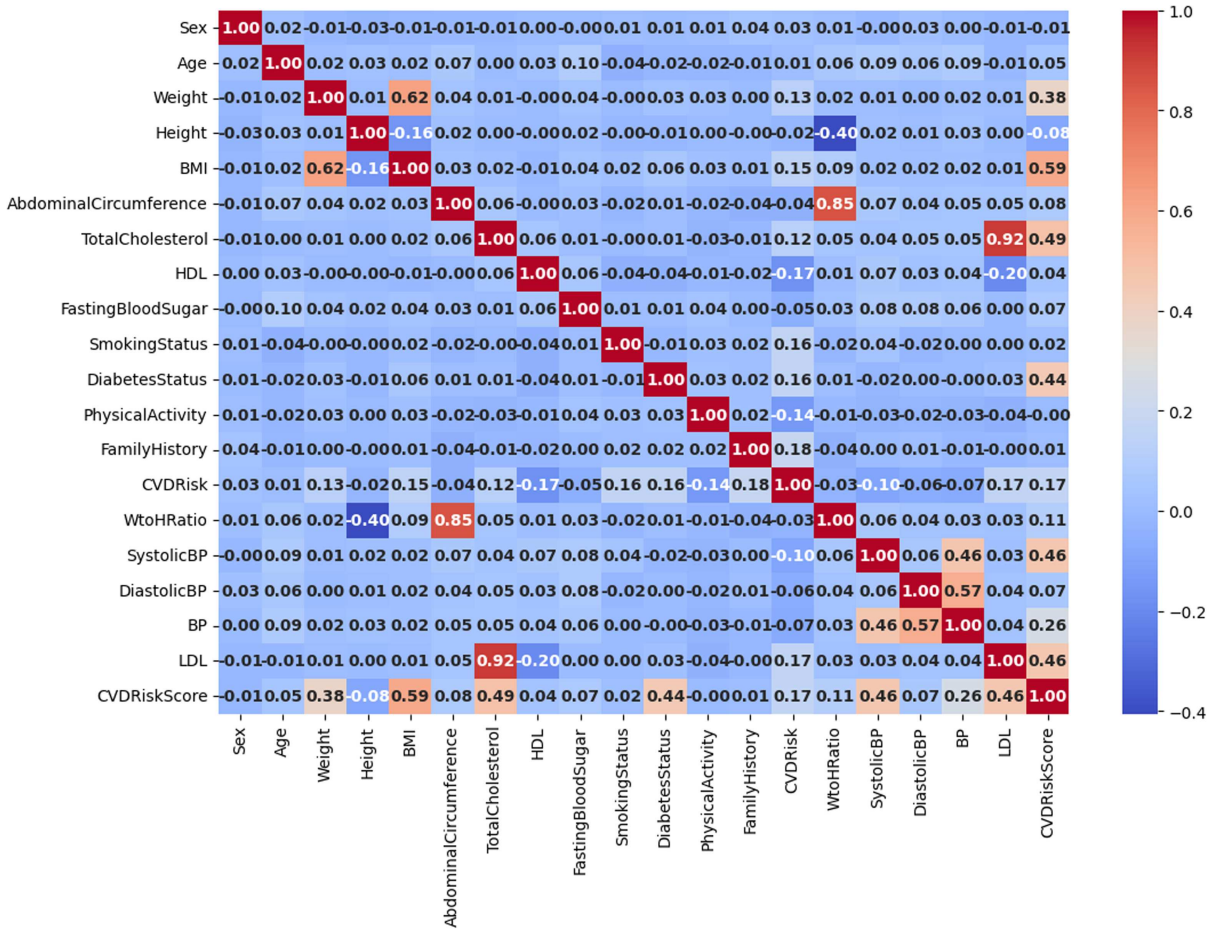

Figure 3. Heatmap illustrating the correlation of CVD with health and lifestyle factors

Table 3. Complete forms of the assessing metrics along with their definitions and formulas

Metric	Definition	Formula
MAE	Average absolute difference between predictions and actual values.	$\frac{1}{n}\sum_{i=1}^{n}\lvert y_i-\hat{y}_i\rvert$
MSE	Average of squared differences between predictions and actual values.	$\frac{1}{n}\sum_{i-1}^{n}(y_i-\hat{y}_i)^2$
R^2	Proportion of variance in data explained by the model.	$1-\frac{\sum_{i=1}^{n}(\hat{y}_i-y_i)^2}{\sum_{i=1}^{n}(y_i-\bar{y}_i)^2}$
PSNR	Log ratio of peak signal power to reconstruction error power.	$10\log_{10}\frac{MAX_I^2}{MSE}$
SNR	Log ratio of signal power to noise power.	$10\log_{10}\left(\frac{P_{signal}}{P_{noise}}\right)$

generalize beyond the training data. In general, lower MSE and MAE values indicate minor prediction errors, while higher R^2, PSNR, and SNR values reflect stronger explanatory power and better signal fidelity. A balanced consideration of these metrics is therefore essential, as relying on a single measure may provide a biased or incomplete assessment of model performance. By jointly analyzing these indicators, one can more reliably determine the robustness and effectiveness of the regression model in practical applications.

3.7. Final model selection

Based on the evaluation of various models under the specified parameters, RR emerged as the most effective individual model, consistently demonstrating superior performance. Nevertheless, to further improve predictive accuracy and harness the complementary strengths of different algorithms, an ensemble stacking strategy was adopted. In this framework, RR, TSR, and GBR were combined, with RR serving as the meta-model. This design

enabled the ensemble to maintain the stability and robustness of RR while also benefiting from the nonlinear modeling capacity of TSR and the variance reduction provided by GBR. As a result, the stacked model achieved more reliable, accurate, and generalizable predictions compared to any single model alone.

To complement the final model selection, XAI techniques were applied to ensure transparency and interpretability of the predictions. XAI refers to a set of methods and tools designed to make ML models understandable to humans, bridging the gap between complex "black-box" algorithms and clinical decision-making. Among various XAI methods, SHAP was employed in this study, as it provides a unified and theoretically grounded approach to quantify the contribution of each feature to the model's output [34]. By assigning Shapley values derived from cooperative game theory, SHAP highlights how individual factors, such as BMI, diabetes status, and blood pressure, influence CVD prediction both globally (across the entire dataset) and locally (for individual patients). This integration of XAI into the final stacked model not only improved trust and interpretability but also aligned the results with established clinical knowledge, making the framework more suitable for real-world healthcare applications [35].

Overall, the methodological workflow, encompassing data cleaning, feature transformation, model selection, and ensemble integration, was directly aligned with the study's objectives. By combining robust regression models with explainable ensemble learning, the proposed method advances CVD risk prediction beyond accuracy toward interpretability and contextual relevance. This design ensures that the model not only performs well statistically but also yields insights that are clinically actionable and aligned with the broader research goals outlined in the introduction.

4. Results and Discussion

The performance of each regression model was evaluated using multiple statistical indicators, including MAE, MSE, and the R^2. Lower MAE and MSE values represent smaller prediction errors, while higher R^2 values indicate stronger model explanatory power. Table 4 summarizes the training results, showing that several models such as TSR, RANSACR, and GBR achieved near-perfect scores (MSE ≈ 0, R^2 ≈ 1). However, these results likely reflect *overfitting* to the training data, meaning that the models captured noise instead of generalizable patterns.

To obtain a more realistic evaluation of model performance, 5-fold and 10-fold cross-validation were performed. Ridge Regression produced stable and consistent results across folds, with an average MAE of 0.29, MSE of 0.38, and R^2 of 0.93 (Table 5). These metrics indicate that RR generalized well to unseen data, balancing accuracy with robustness. Accuracy values across folds ranged from approximately 89% to 96%, confirming its reliability.

Based on these findings, we developed a stacking ensemble that combined RR, TSR, and GBR. This ensemble leveraged the complementary strengths of the three models: RR's stability against multicollinearity, TSR's robustness to outliers, and GBR's ability to capture nonlinear interactions. Under 10-fold cross-validation, the ensemble achieved an MAE of 0.30, MSE of 0.29, and R^2 of 0.96, corresponding to a predictive accuracy of approximately 96%. These results demonstrate that the stacking framework improved generalization while maintaining strong predictive power, outperforming individual models.

Table 5. The results of the 10-fold cross-validation performed for RR

MAE	MSE	RMSE	R^2
0.28	0.255	0.5	0.96
0.3	0.337	0.58	0.96
0.24	0.26	0.51	0.95
0.26	0.353	0.59	0.94
0.29	0.306	0.55	0.94
0.33	0.577	0.76	0.9
0.28	0.399	0.63	0.93
0.34	0.551	0.74	0.89
0.33	0.485	0.7	0.9
0.24	0.235	0.48	0.96

To understand which factors most strongly influenced CVD risk predictions, SHAP analysis was applied to the final ensemble model (Figure 4). The SHAP plot is a visualization tool based on cooperative game theory that explains the contribution of each feature to the predictions made by an ML model. It quantifies how individual input variables affect the predicted risk, thereby enhancing model transparency and interpretability in clinical decision-making [36].

The results highlight BMI as the most important predictor, followed by diabetes status and the categorical blood pressure stage derived from systolic and diastolic measurements. Elevated LDL and total cholesterol were also associated with higher CVD risk, whereas higher HDL appeared protective. This indicates that although the original systolic and diastolic readings were excluded during preprocessing, their clinical significance was retained through the encoded blood pressure categories used in the model. The prominence of BMI over smoking in our model's predictions may reflect specific epidemiological patterns in Bangladesh. While smoking is a well-established global risk factor for CVD, its relative contribution can vary depending on population-level exposure and competing health risks. In Bangladesh, high rates of overweight and obesity have emerged in both urban and rural settings, driven by dietary transitions toward calorie-dense foods, reduced physical activity, and socioeconomic shifts. These changes have

Table 4. Performance comparison of regression models for CVD risk prediction

Models	MAE	MSE	R^2	PSNR	SNR
Ridge	0.5	0.17	0.99	39.16	37.60
TSR	0.00	0.00	1.00	255.99	254.43
RANSACR	0.00	0.00	1.00	267.93	266.37
GBR	0.00	0.00	1.00	97.30	95.74
XGBR	0.00	0.00	1.00	92.68	91.11

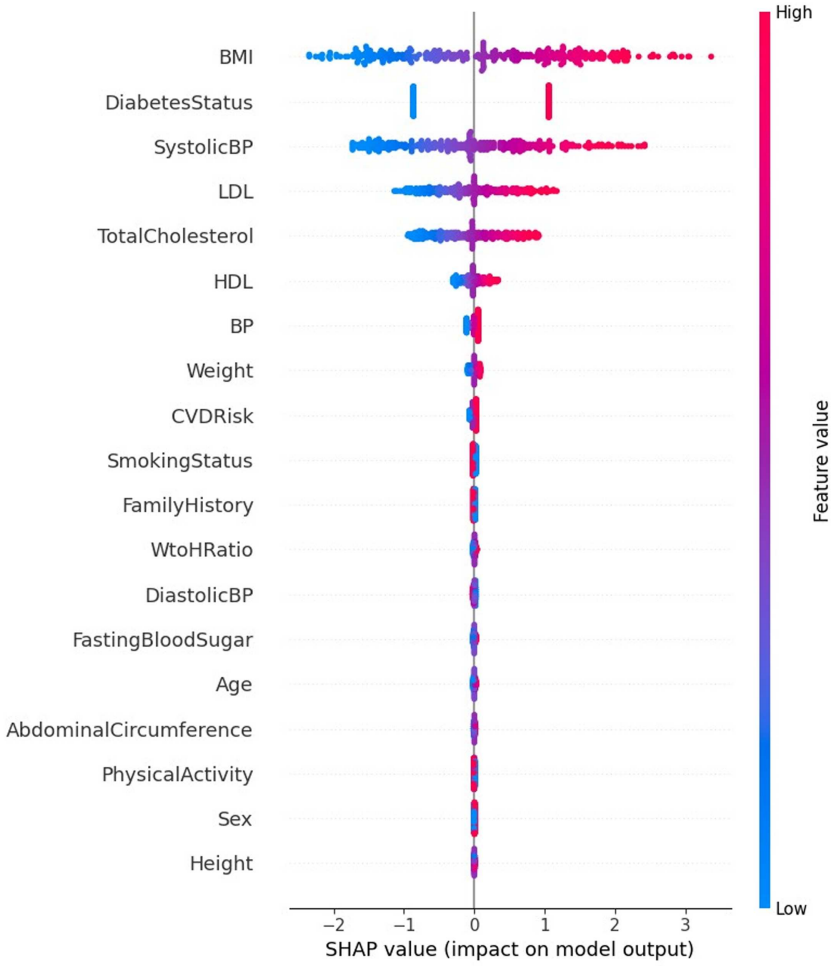

Figure 4. SHAP plot for the stacked ensemble model

produced a rising burden of metabolic syndrome, type 2 diabetes, and hypertension, all strongly mediated through elevated BMI. In contrast, although smoking remains prevalent, its cardiovascular effects may be overshadowed in this cohort by the more substantial, more immediate metabolic impacts of excess body weight. Clinically, this suggests that interventions targeting obesity, dietary modification, and physical inactivity could yield more significant reductions in CVD risk than focusing on smoking cessation alone, particularly in regions where obesity-related comorbidities are accelerating.

Table 6 compares existing studies on CVD risk prediction with the proposed framework using standardized criteria. Previous studies report accuracies ranging from 76% to 94%, but most lack 10-fold validation and XAI integration. Only one prior work (CatBoost) explicitly applied 10-fold cross-validation, and none incorporated explainability.

The proposed framework uniquely combines RR, TSR, and GBR into a stacking ensemble, which is validated using 10-fold cross-validation. It achieves 96% accuracy, outperforming earlier studies, and integrates explainable AI (SHAP) to identify clinically meaningful predictors such as BMI, diabetes, and blood pressure. This dual emphasis on accuracy and interpretability positions it as more robust and clinically actionable than previous black-box approaches.

This study demonstrates the potential of ML for accurate prediction of CVD risk using clinical and lifestyle features. Ridge regression achieved stable performance with ~93% accuracy, while the stacking ensemble of RR, TSR, and GBR further improved accuracy to 96%. SHAP analysis revealed BMI, diabetes status, and blood pressure as the most influential predictors, in agreement with established risk factors, while cholesterol levels also played a significant role. These findings reinforce the clinical relevance of the proposed model. Compared with previous research, our results are competitive and often superior. Prior works employing LR, RF, and MLP achieved accuracies between 85% and 88%, while CatBoost and SVE reached ~94%. By contrast, our stacking ensemble achieved 96% accuracy and offered greater interpretability through SHAP analysis. This addresses a key limitation of many earlier studies, which prioritized accuracy at the expense of transparency, a crucial factor for clinical adoption. Beyond simply achieving higher accuracy, our stacking ensemble outperforms models such as CatBoost and HRFLM because it integrates complementary learning strategies: Ridge Regression ensures stability and resilience against multicollinearity, Theil-Sen provides robustness to outliers, and GB captures complex, nonlinear relationships. This synergy yields not only better predictive power but also more consistent generalization across varying data distributions. Importantly, the embedded SHAP analysis delivers transparent risk attribution, which CatBoost offers only in a limited form and HRFLM largely lacks. As a result, our framework strikes a better balance between predictive strength and clinical interpretability. A key limitation, however, is that our dataset originates from a single institution,

Table 6. Comparison of accuracy between existing studies and the present study

Existing Study	ML Model Used	Accuracy	10-Fold	XAI
Kabir et al. [7]	LR	87.8%	Not specified	No
Alowais et al. [8]	RF	85.01%	No	No
Mathur et al. [9]	MLP	87.28%	Cross-validation (not always 10-fold)	No
Dritsas and Trigka [10]	SVE	93.44%	No	No
Khan et al. [11]	CatBoost	94.34%	10-fold	No
Bhatt et al. [12]	IoT + LR	F1: 91% (binary)	No	No
Chandrasekhar and Peddakrishna [13]	RF	88.52%	No	No
Kanagarathinam et al. [14]	LGB	81.7% (F1 = 0.509)	No	No
Islam et al. [15]	HRFLM	92.25%	No	No
Stonier et al. [16]	GB	76.2% (F1 = 56.7%)	No	No
Proposed framework (this study)	Stacking ensemble (RR + TSR + GBR) + SHAP	96% (R^2 = 0.96)	Yes (10-fold CV)	Yes (SHAP analysis)

which constrains generalizability to other populations. While the model performed well on internal validation, future work should incorporate transfer learning or federated learning frameworks to adapt the model across multicenter datasets while preserving patient privacy. Such efforts will enhance scalability and ensure that the model remains both accurate and clinically relevant in diverse healthcare contexts. The results highlight the promise of ML models as decision-support tools for early CVD risk stratification. By identifying modifiable predictors such as BMI, blood pressure, and cholesterol, the model can inform targeted interventions and preventive strategies. Despite current limitations, this study underscores the value of ensemble ML methods in advancing precise and interpretable CVD risk prediction.

5. Conclusion

This study developed an ML framework for predicting CVD risk using clinical and lifestyle data, achieving up to 96% accuracy with a stacking ensemble of RR, TSR, and GBR. Unlike prior approaches, our method uniquely combines regression-based ensemble learning with explainable AI, delivering not only predictive strength but also transparent, clinically relevant insights. The integration of SHAP analysis highlighted modifiable factors such as BMI, diabetes status, and blood pressure, offering actionable guidance for prevention strategies and enhancing trust in model outputs.

Beyond its technical performance, the study contributes conceptually to the advancement of interpretable AI in healthcare. By emphasizing explainability, this work bridges the gap between data-driven prediction and clinical applicability, showing that ML can support, not replace, expert judgment. The results demonstrate that integrating transparency within high-performing models can foster greater clinician trust and more responsible AI adoption in medical decision-making. Furthermore, by using a real-world hospital dataset from Bangladesh, this study provides rare and valuable evidence from a low-resource healthcare setting, expanding the global scope of AI research that has been dominated by Western datasets.

The scientific value of this work lies not only in its technical performance but also in its contribution to the advancement of interpretable AI in clinical practice. Unlike many prior studies that rely on opaque "black-box" algorithms or benchmark datasets, this research demonstrates how explainable ensemble regression can deliver clinically transparent insights using real-world, low-resource hospital data. This enhances the validity and practical relevance of AI-driven models by aligning predictive outcomes with medically recognized risk factors.

The findings have important implications for public health policy and clinical practice. In particular, they suggest that targeted interventions addressing obesity, diabetes management, and hypertension control could substantially reduce CVD risk in similar populations. The explainable ensemble framework proposed here could also be adapted for hospital-based screening systems or telehealth platforms to assist clinicians in early risk assessment.

While promising, the current work has limitations related to its single-center dataset and retrospective design. Future research should aim to validate the model using multicenter data and explore federated or transfer learning to ensure generalizability across different demographic and clinical contexts. Additionally, incorporating temporal and genetic data may further improve predictive accuracy. Nevertheless, further validation using larger, multicenter datasets is essential to confirm generalizability across populations. Future studies should also explore federated learning and longitudinal data integration to strengthen predictive validity and clinical applicability. Overall, this study contributes a scientifically grounded, interpretable, and contextually relevant framework for advancing precision cardiovascular risk assessment through XAI.

Ethical Statement

This study utilized data obtained from a publicly available dataset. The authors confirm that the data were originally collected under the ethical approvals and reporting standards of the source study. No new data involving human or animal subjects were collected for this research, and as such, independent ethical approval was not required.

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

Data Availability Statement

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

Author Contribution Statement

Suhana Tasnim: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Project administration. **Mohammad Mamun:** Supervision. **Safiul Haque Chowdhury:** Writing – review & editing, Supervision. **Mohammed Ibrahim Hussain:** Supervision. **Muhammad Minoar Hossain:** Supervision.

References

[1] World Health Organization. (2025). *Cardiovascular diseases (CVDs)*. https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

[2] Ciumărnean, L., Milaciu, M. V., Negrean, V., Orășan, O. H., Vesa, S. C., Sălăgean, O., . . . , & Vlaicu, S. I. (2022). Cardiovascular risk factors and physical activity for the prevention of cardiovascular diseases in the elderly. *International Journal of Environmental Research and Public Health*, *19*(1), 207. https://doi.org/10.3390/ijerph19010207

[3] Kaminsky, L. A., German, C., Imboden, M., Ozemek, C., Peterman, J. E., & Brubaker, P. H. (2022). The importance of healthy lifestyle behaviors in the prevention of cardiovascular disease. *Progress in Cardiovascular Diseases*, *70*, 8–15. https://doi.org/10.1016/j.pcad.2021.12.001

[4] Karvandi, M. (2021). Review of laser therapy in cardiovascular diseases. *Journal of Lasers in Medical Sciences*, *12*, e52. https://doi.org/10.34172/jlms.2021.52

[5] Netala, V. R., Teertam, S. K., Li, H., & Zhang, Z. (2024). A comprehensive review of cardiovascular disease management: Cardiac biomarkers, imaging modalities, pharmacotherapy, surgical interventions, and herbal remedies. *Cells*, *13*(17), 1471. https://doi.org/10.3390/cells13171471

[6] Joseph, P., Lanas, F., Roth, G., Lopez-Jaramillo, P., Lonn, E., Miller, V., . . . , & Yusuf, S. (2025). Cardiovascular disease in the Americas: The epidemiology of cardiovascular disease and its risk factors. *The Lancet Regional Health - Americas*, *42*, 100960. https://doi.org/10.1016/j.lana.2024.100960

[7] Kabir, A., Karim, M. N., & Billah, B. (2022). Health system challenges and opportunities in organizing non-communicable diseases services delivery at primary healthcare level in Bangladesh: A qualitative study. *Frontiers in Public Health*, *10*, 1015245. https://doi.org/10.3389/fpubh.2022.1015245

[8] Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., . . . , & Albekairy, A. M. (2023). Revolutionizing healthcare: The role of artificial intelligence in clinical practice. *BMC Medical Education*, *23*(1), 689. https://doi.org/10.1186/s12909-023-04698-z

[9] Mathur, P., Srivastava, S., Xu, X., & Mehta, J. L. (2020). Artificial intelligence, machine learning, and cardiovascular disease. *Clinical Medicine Insights: Cardiology*, *14*, 117954682092740. https://doi.org/10.1177/1179546820927404

[10] Dritsas, E., & Trigka, M. (2023). Efficient data-driven machine learning models for cardiovascular diseases risk prediction. *Sensors*, *23*(3), 1161. https://doi.org/10.3390/s23031161

[11] Khan, A., Qureshi, M., Daniyal, M., & Tawiah, K. (2023). A novel study on machine learning algorithm-based cardiovascular disease prediction. *Health & Social Care in the Community*, *2023*(1), 1406060. https://doi.org/10.1155/2023/1406060

[12] Bhatt, C. M., Patel, P., Ghetia, T., & Mazzeo, P. L. (2023). Effective heart disease prediction using machine learning techniques. *Algorithms*, *16*(2), 88. https://doi.org/10.3390/a16020088

[13] Chandrasekhar, N., & Peddakrishna, S. (2023). Enhancing heart disease prediction accuracy through machine learning techniques and optimization. *Processes*, *11*(4), 1210. https://doi.org/10.3390/pr11041210

[14] Kanagarathinam, K., Sankaran, D., & Manikandan, R. (2022). Machine learning-based risk prediction model for cardiovascular disease using a hybrid dataset. *Data & Knowledge Engineering*, *140*, 102042. https://doi.org/10.1016/j.datak.2022.102042

[15] Islam, M. N., Raiyan, K. R., Mitra, S., Mannan, M. M. R., Tasnim, T., Putul, A. O., & Mandol, A. B. (2023). Predictis: An IoT and machine learning-based system to predict risk level of cardio-vascular diseases. *BMC Health Services Research*, *23*(1), 171. https://doi.org/10.1186/s12913-023-09104-4

[16] Stonier, A. A., Gorantla, R. K., & Manoj, K. (2024). Cardiac disease risk prediction using machine learning algorithms. *Healthcare Technology Letters*, *11*(4), 213–217. https://doi.org/10.1049/htl2.12053

[17] Huang, Q., Jiang, Z., Shi, B., Meng, J., Shu, L., Hu, F., & Mi, J. (2025). Characterisation of cardiovascular disease (CVD) incidence and machine learning risk prediction in middle-aged and elderly populations: Data from the China health and retirement longitudinal study (CHARLS). *BMC Public Health*, *25*(1), 518. https://doi.org/10.1186/s12889-025-21609-7

[18] Zaidi, S. A. J., Ghafoor, A., Kim, J., Abbas, Z., & Lee, S. W. (2025). HeartEnsembleNet: An innovative hybrid ensemble learning approach for cardiovascular risk prediction. *Healthcare*, *13*(5), 507. https://doi.org/10.3390/healthcare13050507

[19] Cheng, C.-H., Lee, B.-J., Nfor, O. N., Hsiao, C.-H., Huang, Y.-C., & Liaw, Y.-P. (2024). Using machine learning-based algorithms to construct cardiovascular risk prediction models for Taiwanese adults based on traditional and novel risk factors. *BMC Medical Informatics and Decision Making*, *24*(1), 199. https://doi.org/10.1186/s12911-024-02603-2

[20] Dumlao, J. (2025). *CAIR-CVD-2025: Cardiovascular risk from Bangladesh* [Data set]. Kaggle. https://www.kaggle.com/datasets/jocelyndumlao/cair-cvd-2025-cardiovascular-risk-from-bangladesh

[21] Lee, D.-H., & Kim, H.-J. (2023). A self-attention-based imputation technique for enhancing tabular data quality. *Data*, *8*(6), 102. https://doi.org/10.3390/data8060102

[22] Unnikrishnan, S., Donovan, J., Macpherson, R., & Tormey, D. (2020). An integrated histogram-based vision and machine-learning classification model for industrial emulsion processing. *IEEE Transactions on Industrial Informatics*, *16*(9), 5948–5955. https://doi.org/10.1109/tii.2019.2959021

[23] Lotfi, M., Javadi, M., Osório, G. J., Monteiro, C., & Catalão, J. P. S. (2020). A novel ensemble algorithm for solar power forecasting based on kernel density estimation. *Energies*, *13*(1), 216. https://doi.org/10.3390/en13010216

[24] Gu, Z. (2022). Complex heatmap visualization. *iMeta*, *1*(3), e43. https://doi.org/10.1002/imt2.43

[25] Arashi, M., Roozbeh, M., Hamzah, N. A., & Gasparini, M. (2021). Ridge regression and its applications in genetic studies.

PLoS ONE, *16*(4), e0245376. https://doi.org/10.1371/journal.pone.0245376

[26] Bal, A. (2024). Improving the robustness of the Theil-Sen estimator using a simple heuristic-based modification. *Symmetry*, *16*(6), 698. https://doi.org/10.3390/sym16060698

[27] Martínez-Otzeta, J. M., Rodríguez-Moreno, I., Mendialdua, I., & Sierra, B. (2023). RANSAC for robotic applications: A survey. *Sensors*, *23*(1), 327. https://doi.org/10.3390/s23010327

[28] Sudhamathi, T., & Perumal, K. (2024). Ensemble regression based Extra Tree Regressor for hybrid crop yield prediction system. *Measurement: Sensors*, *35*, 101277. https://doi.org/10.1016/j.measen.2024.101277

[29] Avanijaa, J., Sunitha, G., Madhavic, K. R., Korad, P., & Sai Vittal, R. H. (2021). Prediction of house price using XGBoost regression algorithm. *Turkish Journal of Computer and Mathematics Education*, *12*(2), 2151–2155. https://doi.org/10.17762/turcomat.v12i2.1870

[30] Konstantinov, A. V., & Utkin, L. V. (2021). Interpretable machine learning with an ensemble of gradient boosting machines. *Knowledge-Based Systems*, *222*, 106993. https://doi.org/10.1016/j.knosys.2021.106993

[31] Shahhosseini, M., Hu, G., & Pham, H. (2022). Optimizing ensemble weights and hyperparameters of machine learning models for regression problems. *Machine Learning with Applications*, *7*, 100251. https://doi.org/10.1016/j.mlwa.2022.100251

[32] Ateeq, M., Afzal, M. K., Naeem, M., Shafiq, M., & Choi, J-G. (2020). Deep learning-based multiparametric predictions for IoT. *Sustainability*, *12*(18), 7752. https://doi.org/10.3390/su12187752

[33] Bilali, A. E., & Taleb, A. (2020). Prediction of irrigation water quality parameters using machine learning models in a semi-arid environment. *Journal of the Saudi Society of Agricultural Sciences*, *19*(7), 439–451. https://doi.org/10.1016/j.jssas.2020.08.001

[34] Wang, Y., Lang, J., Zuo, J. Z., Dong, Y., Hu, Z., Xu, X., . . . , & Li, H. (2022). The radiomic-clinical model using the SHAP method for assessing the treatment response of whole-brain radiotherapy: A multicentric study. *European Radiology*, *32*(12), 8737–8747. https://doi.org/10.1007/s00330-022-08887-0

[35] Chowdhury, S. H., Mamun, M., Shaikat, T. A., Hussain, M. I., Iqbal, S., & Hossain, M. M. (2025). An ensemble approach for artificial neural network-based liver disease identification from optimal features through hybrid modeling integrated with advanced explainable AI. *Medinformatics*, *2*(2), 107–119. https://doi.org/10.47852/bonviewMEDIN52024744

[36] Chowdhury, S. H., Mamun, M., Shaikat, T. A., Hussain, M. I., & Hossain, M. M. (2025). Improving network classification accuracy through feature clustering and ensemble machine learning with explainable AI. In *2025 International Conference on Electrical, Computer and Communication Engineering*, 1–6. https://doi.org/10.1109/ECCE64574.2025.11013433

How to Cite: Tasnim, S., Mamun, M., Chowdhury, S. H., Hussain, M. I., & Hossain, M. M. (2026). Advancing Interpretable AI for Cardiovascular Risk Assessment: A Stacking Regression Approach in Clinical Data from Bangladesh. *Medinformatics*, *3*(3), 275–285. https://doi.org/10.47852/bonviewMEDIN52027812