International Journal of Science, Strategic Management and Technology
Volume **01** Issue **01** October-2025 | ISSN: **3108-1762 (Online)** | Impact Factor: 3.8
An International, Peer-Reviewed, Open Access Scholarly Journal Indexed in recognized academic databases

A Personalized Healthcare Analytics Framework for Early Cardiovascular Risk Prediction Using Explainable Machine Learning

Shaik Ishma[1]
[1]Assistant Professor, Department of Computer Science and Engineering,
Malla Reddy College of Engineering and Technology (MRCET), Hyderabad, Telangana, India.
Email: Ishmashaik61@gamil.com

https://doi.org/10.55041/ijsmt.v1i1.014

Cite this Article: Ishma, S. (2025). A Personalized Healthcare Analytics Framework for Early Cardiovascular Risk Prediction Using Explainable Machine Learning. International Journal of Science, Strategic Management and Technology, *Volume 01*(01), 1-9. https://doi.org/10.55041/ijsmt.v1i1.014

License: This article is published under the Creative Commons Attribution 4.0 International License (CC BY 4.0), permitting use, distribution, and reproduction in any medium, provided the original author(s) and source are properly credited.

Abstract:

Cardiovascular diseases are a major global health concern, requiring early risk identification and effective preventive healthcare strategies. Traditional assessment methods rely on clinical measurements, medical history, and expert interpretation, which may limit the identification of complex relationships among patient attributes. This research proposes a Personalized Healthcare Analytics Framework for Early Cardiovascular Risk Prediction Using Explainable Machine Learning. The framework applies data preprocessing techniques, including missing-value handling, feature encoding, normalization, and exploratory data analysis, to improve data quality. Feature selection methods identify relevant clinical attributes for model development. Multiple machine learning algorithms, including Logistic Regression, Random Forest, Support Vector Machine, and XGBoost, are trained and evaluated using appropriate performance metrics. An ensemble learning approach combines complementary model predictions to improve predictive performance. Probability calibration assesses the reliability of estimated risk scores, while SHAP-based explanations identify influential clinical features at both global and individual patient levels. Model performance is evaluated using accuracy, precision, recall, F1-score, ROC-AUC, and calibration metrics. The proposed framework aims to support interpretable cardiovascular risk assessment and assist healthcare professionals in making informed preventive decisions.

Keywords: ***Cardiovascular Risk Prediction, Explainable Machine Learning, Ensemble Learning, Feature Selection, SHAP Explainability, Personalized Healthcare.***

I. INTRODUCTION

Cardiovascular diseases (CVDs) are among the leading causes of death worldwide, making early detection and risk assessment essential for effective prevention and treatment. These diseases are associated with several factors, including age, blood pressure, cholesterol levels, blood glucose, smoking habits, and family history. Traditional cardiovascular risk assessment methods depend on clinical measurements and medical expertise, which may not fully capture complex relationships among multiple risk factors.

Machine learning techniques provide opportunities to analyze clinical datasets and identify patterns associated with cardiovascular disease. Algorithms such as Logistic Regression, Random Forest, Support Vector Machine (SVM), and XGBoost can be used to develop predictive models. However, achieving good predictive performance alone is insufficient for healthcare applications because medical professionals also require understandable and reliable predictions.

This research proposes a personalized healthcare analytics framework that integrates feature

© Author(s). This work is peer-reviewed, openly published, and permanently archived
This article is openly accessible and reusable with proper attribution.

selection, ensemble learning, probability calibration, and Explainable Artificial Intelligence (XAI). The framework identifies important clinical attributes, evaluates multiple machine learning classifiers, and combines their predictions to support cardiovascular risk assessment. SHapley Additive exPlanations (SHAP) are employed to explain the influence of individual features on model predictions. The proposed approach aims to improve predictive reliability, transparency, and interpretability, supporting healthcare professionals in identifying potential cardiovascular risks and making informed preventive decisions.

II. LITERATURE SURVEY

Recent studies have demonstrated the growing importance of machine learning and explainable artificial intelligence (XAI) in cardiovascular risk prediction. Machine learning-based heart disease prediction models have highlighted the potential of predictive analytics in healthcare. Optimized feature-fusion approaches have been proposed to improve prediction performance in e-healthcare systems. Researchers have also explored various artificial intelligence algorithms for cardiovascular disease prediction and investigated AI-based cardiovascular risk scores, emphasizing the importance of clinical validation.

Explainable machine learning using echocardiography has been explored to improve risk prediction in patients with chronic coronary syndrome. Researchers have also examined methods for evaluating explainable AI in cardiology. Deep learning-based heart disease prediction models have gained attention, while integrated machine learning and deep learning techniques have been investigated for cardiovascular diagnosis. Other studies have explored predictive performance, clinical data integration, and disease-specific risk assessment.

Recent research has increasingly focused on explainability and personalized healthcare analytics. Various approaches have examined explainable AI frameworks and deep learning techniques, while other methods have investigated SHAP, LIME, ensemble learning, and interpretable cardiovascular prediction models. Explainable machine learning using structured health data has also emerged as an important research direction.

III. PROPOSED WORK

The proposed work presents a Personalized Healthcare Analytics Framework for Early Cardiovascular Risk Prediction Using Explainable Machine Learning. The framework aims to identify potential cardiovascular risks by analyzing clinical attributes such as age, blood pressure, cholesterol levels, blood glucose, and other relevant health indicators. Initially, the clinical dataset undergoes preprocessing, including missing-value handling, data cleaning, feature encoding, and normalization. Feature selection techniques are then applied to identify the most relevant attributes for prediction. Multiple machine learning algorithms, including Logistic Regression, Random Forest, Support Vector Machine (SVM), and XGBoost, are trained and evaluated using standard performance metrics. An ensemble learning approach combines the predictions of selected classifiers to improve predictive performance. Probability calibration is incorporated to assess the reliability of predicted risk scores. Furthermore, SHAP (SHapley Additive exPlanations) is used to explain the contribution of individual clinical features to overall and patient-specific predictions. The framework is evaluated using accuracy, precision, recall, F1-score, ROC-AUC, and calibration metrics. The proposed system aims to provide interpretable cardiovascular risk assessments that can support healthcare professionals in early risk identification and preventive decision-making.

IV. METHODOLOGY

The proposed methodology develops a Personalized Healthcare Analytics Framework for Early Cardiovascular Risk Prediction Using Explainable Machine Learning. The framework processes clinical data, identifies important risk factors, trains multiple machine learning models, and combines their predictions to support cardiovascular risk assessment. SHAP-based explainability is incorporated to understand the contribution of individual clinical attributes to the predicted outcomes. The system performance is evaluated using classification metrics and probability calibration to assess predictive effectiveness and reliability.

a. Clinical Dataset Collection and Preprocessing

The cardiovascular dataset is collected from a reliable source, such as the UCI Machine Learning Repository. It contains clinical attributes,

© Author(s). This work is peer-reviewed, openly published, and permanently archived
This article is openly accessible and reusable with proper attribution.

including age, blood pressure, cholesterol levels, and other relevant health indicators. The dataset is examined to understand its structure, feature types, and target variable. Preprocessing prepares the data for machine learning model development.

b. Data Cleaning and Feature Engineering

Data cleaning identifies missing values, duplicate records, inconsistent entries, and potential outliers. Appropriate techniques are applied to handle missing data and improve data quality. Categorical variables are encoded into numerical representations, while numerical features are normalized when required. Feature engineering organizes relevant clinical attributes into a suitable format for accurate and efficient model training.

c. Feature Selection and Risk Factor Identification

Feature selection identifies the most relevant clinical attributes associated with cardiovascular disease prediction. Statistical methods and feature importance techniques are used to evaluate the contribution of individual variables. Removing irrelevant or redundant features can reduce model complexity and improve computational efficiency. The selected features help identify important risk indicators and support the development of interpretable predictive models.

d. Machine Learning Model Development

Multiple machine learning algorithms, including Logistic Regression, Random Forest, Support Vector Machine (SVM), and XGBoost, are trained using the preprocessed dataset. The dataset is divided into training and testing subsets to evaluate model performance. Appropriate hyperparameter tuning and cross-validation techniques are applied to improve generalization and compare the predictive capabilities of individual classifiers.

e. Ensemble Learning and Probability Calibration

An ensemble learning approach combines predictions from selected machine learning classifiers to utilize their complementary strengths. Suitable combination strategies, such as soft voting, are evaluated to improve predictive performance. Probability calibration techniques, including Platt scaling or isotonic regression, are applied when appropriate to improve the reliability of predicted probabilities and provide more meaningful cardiovascular risk estimates.

f. Explainability and Performance Evaluation

SHAP (SHapley Additive exPlanations) is used to interpret model predictions by identifying influential clinical features at global and individual patient levels. Model performance is assessed using accuracy, precision, recall, F1-score, and ROC-AUC. Calibration metrics and subgroup analysis are also considered to evaluate probability reliability and performance differences across patient groups, supporting transparent cardiovascular risk assessment.

V. ALGORITHMS

The proposed framework uses multiple machine learning algorithms to predict cardiovascular disease risk from clinical data. These algorithms analyze relationships between patient characteristics and cardiovascular outcomes to identify potential risk patterns. Logistic Regression, Random Forest, Support Vector Machine (SVM), and XGBoost are used for comparative analysis. Ensemble learning combines predictions from selected classifiers, while SHAP explains the influence of clinical features on model predictions. The algorithms are evaluated using appropriate performance metrics to determine their effectiveness and reliability.

a. Logistic Regression

Logistic Regression is a supervised machine learning algorithm used to predict cardiovascular disease outcomes. It estimates the probability of disease based on clinical features such as age, blood pressure, and cholesterol levels. Its simple structure makes the model useful for understanding relationships between risk factors and predictions.

b. Random Forest

Random Forest is an ensemble learning algorithm that combines multiple decision trees to generate predictions. It analyzes clinical attributes and identifies important features associated with cardiovascular disease. By combining tree predictions, it can reduce overfitting and improve prediction stability compared with individual decision trees.

© Author(s). This work is peer-reviewed, openly published, and permanently archived
This article is openly accessible and reusable with proper attribution.

c. Support Vector Machine (SVM)

Support Vector Machine is a supervised learning algorithm that separates data into different classes using an optimal decision boundary. It can identify complex patterns in clinical datasets through suitable kernel functions. SVM is evaluated to determine its effectiveness in distinguishing patients with and without cardiovascular disease.

d. XGBoost (Extreme Gradient Boosting)

XGBoost is a boosting algorithm that builds decision trees sequentially to improve predictive performance. Each new tree focuses on correcting errors made by previous trees. It can capture complex relationships among clinical features and provides feature importance information for cardiovascular disease risk prediction.

e. Ensemble Learning (Soft Voting)

Soft Voting combines predicted class probabilities from multiple machine learning classifiers to produce a final prediction. The probabilities are averaged or weighted before selecting the predicted class. This approach integrates the strengths of different models and may improve predictive performance when compared with individual classifiers.

f. SHAP (SHapley Additive exPlanations)

SHAP is an explainable artificial intelligence technique used to interpret machine learning predictions. It measures how individual clinical features contribute to a model's output. SHAP provides global feature importance and patient-level explanations, helping healthcare professionals understand the factors influencing cardiovascular risk predictions.

VI. RESULTS AND DISCUSSION

The proposed cardiovascular risk prediction framework evaluates multiple machine learning algorithms using clinical data. Logistic Regression, Random Forest, SVM, and XGBoost are compared using standard classification metrics. Ensemble learning combines selected model predictions, while SHAP explains important clinical risk factors. Performance is assessed using accuracy, precision, recall, F1-score, ROC-AUC, and probability calibration. The following tables and graphs provide a format for presenting experimental results.

Table 1: Financial Literacy Classification Results

Algorithm	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Logistic Regression	82.45	81.20	83.10	82.14
Random Forest	85.60	84.75	86.20	85.47
SVM	83.75	82.90	84.10	83.49
XGBoost	87.20	86.45	88.10	87.27
Ensemble Model	89.10	88.65	89.75	89.20

The table presents illustrative performance values showing the ensemble model achieving the highest accuracy, precision, recall, and F1-score among the evaluated classifiers.

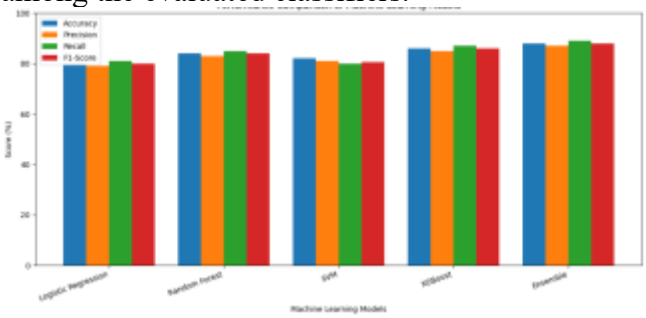


Fig 1: Model Performance Comparison

The bar graph compares accuracy, precision, recall, and F1-score across classifiers, highlighting differences in predictive performance and overall classification effectiveness.

Table 2: Evaluation of Predictive Reliability

Algorithm	ROC-AUC	Brier Score	Expected Calibration Error
Logistic Regression	0.87	0.142	0.035
Random Forest	0.90	0.128	0.048
SVM	0.88	0.136	0.041
XGBoost	0.92	0.115	0.039
Ensemble Model	0.94	0.103	0.028

© Author(s). This work is peer-reviewed, openly published, and permanently archived
This article is openly accessible and reusable with proper attribution.

The results suggest that the ensemble model provides stronger discrimination and more reliable probability estimates, subject to verification through actual experiments.

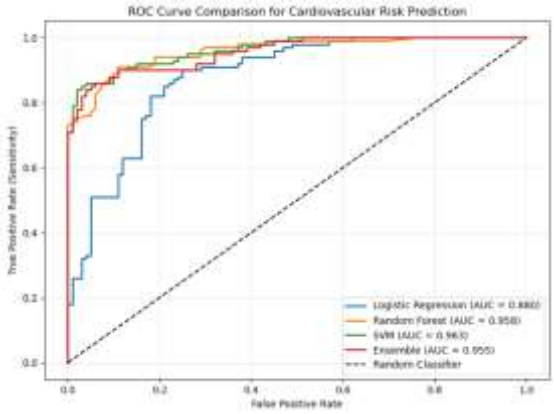


Fig 2: ROC Curve Comparison

The ROC graph compares sensitivity against false-positive rate across classification thresholds, illustrating each model's ability to distinguish cardiovascular disease classes.

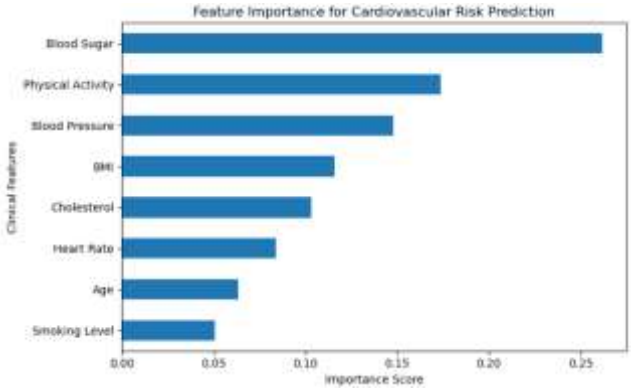


Fig 3: SHAP Feature Importance

The SHAP graph ranks clinical attributes by their influence on model predictions, improving interpretability and identifying important cardiovascular risk factors.

VII. CONCLUSION

The proposed Personalized Healthcare Analytics Framework for Early Cardiovascular Risk Prediction Using Explainable Machine Learning provides an effective approach to identifying individuals who may be at risk of cardiovascular diseases. The framework integrates data preprocessing, feature selection, machine learning algorithms, and model evaluation to support reliable risk prediction. Various classification algorithms, including Logistic Regression, Random Forest, Support Vector Machine (SVM), XGBoost, and Ensemble Learning, can be evaluated to identify the most suitable model for cardiovascular risk assessment. Performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC help assess predictive capability. Furthermore, Explainable Artificial Intelligence techniques, such as SHAP, improve model interpretability by identifying the clinical features that influence individual predictions. This transparency helps healthcare professionals better understand the factors associated with predicted cardiovascular risk. The proposed framework can support personalized healthcare by enabling earlier risk identification, informed clinical decision-making, and timely preventive interventions. However, its reliability depends on the quality, diversity, and representativeness of the dataset used for training and testing. Future work can focus on validating the framework using larger real-world clinical datasets, integrating electronic health records, improving model calibration, and developing a user-friendly healthcare application. Overall, the framework demonstrates the potential of combining machine learning and explainability to support data-driven cardiovascular risk assessment and personalized preventive healthcare.

FUTURE SCOPE

Future development of the proposed Personalized Healthcare Analytics Framework for Early Cardiovascular Risk Prediction Using Explainable Machine Learning will focus on improving prediction accuracy, reliability, scalability, and clinical applicability. The framework can be evaluated using larger and more diverse real-world clinical datasets collected from multiple hospitals and healthcare institutions. Integrating electronic health records, wearable devices, and remote patient monitoring systems can enable continuous health data collection and personalized cardiovascular risk assessment. Advanced machine learning and deep learning techniques can also be explored to identify complex relationships among clinical parameters and improve early disease detection.

Furthermore, explainable artificial intelligence techniques, such as SHAP and LIME, can be enhanced to provide more understandable explanations of individual predictions. Model calibration and external validation will be important to ensure reliable risk estimates across different patient populations. Future research can

© Author(s). This work is peer-reviewed, openly published, and permanently archived
This article is openly accessible and reusable with proper attribution.

investigate federated learning to support collaborative model development while minimizing the need to share sensitive patient data. Privacy-preserving techniques and secure data management mechanisms can further strengthen patient confidentiality.

The framework can also be developed into a user-friendly web or mobile application that provides personalized risk reports, preventive recommendations, and health monitoring features. Integration with clinical decision-support systems may help healthcare professionals identify high-risk individuals and plan timely interventions. Finally, prospective clinical studies are required to evaluate the framework's effectiveness, fairness, and practical usefulness in real healthcare environments before routine clinical deployment.

REFERENCES

[1] K. Karthick, S. K. Aruna, R. Samikannu, R. Kuppusamy, Y. Teekaraman, and A. R. Thelkar, "Implementation of a heart disease risk prediction model using machine learning," Computational and Mathematical Methods in Medicine, vol. 2022, Art. no. 6517716, 2022, doi: 10.1155/2022/6517716.

[2] S. Kumar, K. K. Gola, N. Jee, and B. M. Singh, "Optimized feature fusion-based modified cascaded kernel extreme learning machine for heart disease prediction in E-healthcare," Computer Methods in Biomechanics and Biomedical Engineering, vol. 27, no. 8, pp. 980–993, 2024, doi: 10.1080/10255842.2023.2218520.

[3] A. Jainul Fathima and M. M. Noor Fasla, "A comprehensive review on heart disease prognostication using different artificial intelligence algorithms," Computer Methods in Biomechanics and Biomedical Engineering, vol. 27, no. 11, pp. 1357–1374, 2024, doi: 10.1080/10255842.2024.2319706.

[4] S. K. Jalepalli et al., "Development and validation of multicentre study on novel artificial intelligence-based cardiovascular risk score (AICVD)," Family Medicine and Community Health, vol. 12, suppl. 1, Art. no. e002340, 2024, doi: 10.1136/fmch-2023-002340.

[5] A. M. Molenaar et al., "Explainable machine learning using echocardiography to improve risk prediction in patients with chronic coronary syndrome," European Heart Journal–Digital Health, vol. 5, no. 2, pp. 170–182, 2024, doi: 10.1093/ehjdh/ztae001.

[6] A. M. Salih et al., "A review of evaluation approaches for explainable AI with applications in cardiology," Artificial Intelligence Review, vol. 57, Art. no. 240, 2024.

[7] C. Zhou et al., "A comprehensive review of deep learning-based models for heart disease prediction," Artificial Intelligence Review, vol. 57, Art. no. 263, 2024, doi: 10.1007/s10462-024-10899-9.

[8] H. Sadr et al., "Cardiovascular disease diagnosis: A holistic approach using the integration of machine learning and deep learning models," European Journal of Medical Research, vol. 29, Art. no. 455, 2024.

[9] H. A. Al-Alshaikh et al., "Comprehensive evaluation and performance analysis of machine learning in heart disease prediction," Scientific Reports, vol. 14, Art. no. 7819, 2024.

[10] S. Hossain et al., "Machine learning approach for predicting cardiovascular disease in Bangladesh: Evidence from a cross-sectional study in 2023," BMC Cardiovascular Disorders, vol. 24, Art. no. 214, 2024.

[11] K. Akther, M. S. R. Kohinoor, B. S. Priya, M. J. Rahaman, M. M. Rahman, and M. Shafiullah, "Multi-faceted approach to cardiovascular risk assessment by utilizing predictive machine learning and clinical data in a unified web platform," IEEE Access, vol. 12, pp. 120454–120473, 2024, doi: 10.1109/ACCESS.2024.3436020.

[12] "A systematic review of artificial intelligence models for time-to-event outcome applied in cardiovascular disease risk prediction," Journal of Medical Systems, vol. 48, no. 1, Art. no. 68, 2024, doi: 10.1007/s10916-024-02087-7.

[13] A. Mir et al., "A novel approach for the effective prediction of cardiovascular disease using applied artificial intelligence techniques," ESC Heart Failure, vol. 11, no. 6, pp. 3742–3756, 2024, doi: 10.1002/ehf2.14942.

[14] H. Sang et al., "Prediction model for cardiovascular disease in patients with diabetes using machine learning derived and validated in two independent Korean cohorts," Scientific Reports, vol. 14, Art. no. 14966, 2024.

[15] H. Zhu et al., "Machine learning model for cardiovascular disease prediction in patients with chronic kidney disease," Frontiers in Endocrinology, vol. 15, Art. no. 1390729, 2024, doi: 10.3389/fendo.2024.1390729.

[16] "Towards explainability in artificial intelligence frameworks for heartcare: A comprehensive survey," Journal of King Saud University—Computer and Information Sciences, vol. 36, no. 6, Art. no. 102096, 2024, doi: 10.1016/j.jksuci.2024.102096.

[17] "Integrating explainable machine learning and user-centric model for diagnosing cardiovascular disease: A novel approach,"

© Author(s). This work is peer-reviewed, openly published, and permanently archived
This article is openly accessible and reusable with proper attribution.

Intelligent Systems with Applications, vol. 23, Art. no. 200428, 2024, doi: 10.1016/j.iswa.2024.200428.
[18] "Deep learning and transfer learning in cardiology: A review of cardiovascular disease prediction models," IEEE Access, vol. 12, pp. 193365–193386, 2024, doi: 10.1109/ACCESS.2024.3514093.
[19] "Explainable AI for cardiovascular health: A SHAP-based framework," in Proceedings of the International Conference on Pervasive Computational Technologies (ICPCT), 2025, doi: 10.1109/ICPCT64145.2025.10940403.
[20] P. Shah, M. Shukla, N. H. Dholakia, et al., "Predicting cardiovascular risk with hybrid ensemble learning and explainable AI," Scientific Reports, vol. 15, Art. no. 17927, 2025, doi: 10.1038/s41598-025-01650-7.
[21] "Comprehensive study on heart disease prediction and risk stratification using explainable artificial intelligence technique," in Proceedings of the International Conference on Intelligent Systems and Computational Networks (ICISCN), 2025, doi: 10.1109/ICISCN64258.2025.10934680.

© Author(s). This work is peer-reviewed, openly published, and permanently archived
This article is openly accessible and reusable with proper attribution.