Journal of Neonatal Surgery
ISSN(Online): 2226-0439
Vol. 14, Issue 32s (2025)
https://www.jneonatalsurg.com

OPEN ACCESS

SHAP-Enhanced Multi-Class XG Boost for Early Detection and Risk Assessment of Cardiovascular Diseases

Annaldas Puneetha[1], Dr. K. Laxminarayanamma[2], Dr. Ambati Rama Mohan Reddy[3], Dr. R.M. Noorullah[4]

[1]M.Tech., Student, 24951D5801, CSE Department, Institute of Aeronautical Engineering, Hyderabad, India.
Email ID: 24951d5801.iare@gmail.com, ORCID: 0009-0009-2408-9998.
[2]Associate Professor, CSE Department, Institute of Aeronautical Engineering, Hyderabad, India.
Email iD: k.laxminarayanamma@iare.ac.in, ORCID: 0000-0002-2190-7627
[3]Professor and Head, CSE Department, Institute of Aeronautical Engineering, Hyderabad, India.
Email ID: a.ramamohanreddy@iare.ac.in
[4]Associate Professor, CSE Department, Institute of Aeronautical Engineering, Hyderabad, India.
Email ID: noorullah.rm@iare.ac.in, ORCID: 0000-0002-5251-5685.

Cite this paper as: Annaldas Puneetha, Dr. K. Laxminarayanamma, Dr. Ambati Rama Mohan Reddy, Dr. R.M. Noorullah, (2025) SHAP-Enhanced Multi-Class XG Boost for Early Detection and Risk Assessment of Cardiovascular Diseases. *Journal of Neonatal Surgery*, 14 (32s), 8958-8967.

ABSTRACT
Cardiovascular diseases (CVDs) are a group of disorders affecting the heart and blood vessels, including coronary heart disease, cerebrovascular disease, and rheumatic heart disease. These conditions are among the leading causes of death globally. Traditional diagnostic methods often depend heavily on clinical expertise and can be time-consuming, which highlights the potential of machine learning as an effective alternative for faster and more accurate decision-making. This study employs a Multi-Class XG Boost model integrated with SHAP (SHapley Additive exPlanations) to predict the likelihood of cardiovascular diseases using a dataset comprising key demographic, clinical, and diagnostic features. The model achieved an accuracy of 86.59% on training data and 83.75% on testing data, demonstrating strong performance and robustness. The Multi-Class XG Boost approach allows the model to classify patients into multiple risk categories—such as low, medium, and high risk—rather than just binary outcomes (disease/no disease). This enhances the model's clinical utility by providing a more granular prediction of disease severity. Additionally, the integration of SHAP enables interpretability by identifying how each feature (such as age, cholesterol level, blood pressure, or smoking status) contributes to the final prediction, making the model's decisions more transparent to clinicians. In conclusion, the results suggest that the Multi-Class XG Boost model with SHAP not only provides high accuracy but also offers explainable and actionable insights for early detection and risk assessment of cardiovascular diseases. Future research could focus on expanding the dataset, exploring deep learning-based architectures, and enhancing model interpretability for real-time clinical applications.

Keywords: *Cardiovascular Disease Prediction, Multi-Class XG Boost, SHAP (SHapley Additive ExPlanations), Risk Stratification, Healthcare, Clinical Decision Support.*

1. INTRODUCTION

Cardiovascular diseases (CVDs) represent one of the most critical public health challenges globally, accounting for a significant proportion of morbidity and mortality every year. These diseases encompass a broad spectrum of disorders affecting the heart and blood vessels, including coronary heart disease, cerebrovascular disease, rheumatic heart disease, and other conditions that disrupt normal cardiovascular function. According to the World Health Organization, CVDs are responsible for millions of deaths annually, underscoring the urgent need for early detection, accurate diagnosis, and effective intervention strategies. Traditionally, the diagnosis and risk assessment of cardiovascular diseases have relied heavily on clinical expertise, manual interpretation of medical data, and conventional statistical methods. While these approaches have been instrumental in clinical decision-making, they often involve time-consuming procedures and may be prone to human error due to the complexity and variability of patient data. Furthermore, traditional diagnostic tools are typically limited to binary classifications, identifying patients as either "at risk" or "not at risk," without providing detailed insights into the degree or severity of risk. Such limitations make it challenging for healthcare professionals to personalize treatment plans or take preventive measures tailored to an individual's risk level.

In recent years, the rapid growth of data analytics and artificial intelligence (AI) has opened new avenues for improving disease prediction and medical decision support. Among these, machine learning (ML) techniques have emerged as powerful tools for analyzing large and complex healthcare datasets. By identifying hidden patterns and relationships within the data, ML models can deliver predictions with high accuracy, speed, and reliability—offering a transformative approach to medical diagnostics. Specifically, ensemble learning algorithms like Extreme Gradient Boosting (XGBoost) have demonstrated remarkable success in classification problems due to their robustness, flexibility, and superior performance over traditional models.

In this study, a multi-class XGBoost model is employed to predict cardiovascular disease risk by classifying patients into low, medium, and high-risk categories. This multi-class classification approach provides a more comprehensive and granular assessment compared to traditional binary models, thereby enabling clinicians to prioritize patient care more effectively. By leveraging structured clinical data—such as age, gender, cholesterol levels, blood pressure, and lifestyle factors—the model captures complex non-linear relationships among risk variables, ultimately enhancing diagnostic precision. To ensure interpretability and transparency, the model integrates SHAP (SHapley Additive explanations)—a powerful explainable AI technique that quantifies each feature's contribution to the prediction outcome. The integration of SHAP not only improves trust in the model's predictions but also provides actionable insights for clinicians. For instance, SHAP values can highlight how variables such as age, total cholesterol, or systolic blood pressure influence the predicted risk level, enabling medical professionals to make informed, evidence-based decisions.

Overall, this study demonstrates how combining a multi-class XGBoost model with SHAP interpretability offers a balanced framework that achieves both high predictive accuracy and transparent decision-making. Such integration represents a promising step toward the development of intelligent, data-driven healthcare systems capable of supporting early CVD detection, risk stratification, and personalized medical recommendations—ultimately contributing to improved patient outcomes and reduced healthcare burdens

2. LITERATURE REVIEW

Cardiovascular disease (CVD) prediction has been an evolving area of study within medical data mining, with early research focusing on the application of machine learning algorithms to identify critical risk factors and improve diagnostic accuracy. Foundational studies, such as those by Kumar et al., explored supervised learning techniques that utilized historical patient data to uncover trends and associations related to heart disease. These initial efforts laid the groundwork for data-driven healthcare analytics and demonstrated the potential of computational models in early disease detection.

As the field advanced, recent investigations, including Bhowmik et al., analyzed the progression of machine learning architectures and their efficacy in predicting CVDs. These studies compared traditional models—such as Decision Trees and Support Vector Machines (SVMs)—with more sophisticated ensemble approaches like Random Forest and XGBoost, highlighting notable improvements in predictive accuracy and model robustness.

Classical machine learning models, as discussed by Katarya and Meena, continue to play an essential role in cardiovascular prediction tasks. Algorithms such as Decision Trees, Naïve Bayes, and SVMs are valued for their interpretability and computational efficiency, making them suitable for structured clinical datasets. However, their limited ability to capture complex feature interactions often constrains their predictive performance.

To overcome these challenges, ensemble and hybrid models have gained prominence. Methods such as Random Forest, AdaBoost, and XG Boost combine the outputs of multiple learners to achieve higher accuracy and stability. Anjum et al. demonstrated that hybrid frameworks often outperform standalone algorithms by leveraging the complementary strengths of individual models to identify high-risk patients more effectively.

Furthermore, with advancements in computational capabilities, deep learning approaches are increasingly being adopted for CVD prediction. Studies like Cai et al. explored the use of Artificial Neural Networks (ANNs) to model intricate, non-linear relationships among patient attributes. These architectures excel at capturing subtle and complex dependencies within medical data, offering promising avenues for early detection and personalized cardiovascular risk assessment.

3. DATASET

This study utilizes a comprehensive and well-structured dataset specifically curated to support multi-class prediction of cardiovascular disease (CVD) risk levels—categorized as *low*, *medium*, and *high*. The dataset consists of anonymized patient records, each labeled according to the diagnosed risk category, enabling robust supervised learning for multi-class classification.

Each patient record includes a diverse range of medically relevant features grouped into three categories:

- Demographic Features: Age, sex, and chest pain type.
- Clinical Measurements: Resting blood pressure, serum cholesterol level, fasting blood sugar, resting

electrocardiographic (ECG) results, maximum heart rate achieved, and exercise-induced angina.

- Diagnostic Indicators: ST depression induced by exercise (oldpeak), slope of the ST segment during peak exercise, number of major vessels colored by fluoroscopy (ca), and thalassemia status (thal).

To ensure high-quality input for model training, the dataset undergoes several preprocessing steps, including:

- Handling missing values to preserve data completeness and reliability.
- Normalization of continuous variables to maintain scale consistency across features.
- Encoding of categorical variables to make the dataset compatible with machine learning algorithms such as XG Boost.

After preprocessing, the dataset is divided into training and testing subsets to evaluate model performance and generalization capability. This curated dataset serves as the foundation for the Multi-Class XG Boost model with SHAP integration, enabling accurate, explainable, and clinically reliable prediction of cardiovascular disease risk categories.

4. DATA ANALYSIS AND PREPROCESSING

Proper data preprocessing and analysis are crucial for developing a reliable and efficient multi-class XGBoost model integrated with SHAP for cardiovascular disease (CVD) prediction. This section outlines the systematic steps undertaken to prepare the dataset before training and evaluation.

A. Dataset Collection and Description

The dataset employed in this study consists of anonymized and structured medical records, each representing an individual patient. Every record includes a comprehensive set of demographics, clinical, and diagnostic attributes, allowing the model to classify patients into multiple CVD risk categories—low, medium, and high.

Key features include:

- Demographic Features: Age, sex, and chest pain type.
- Clinical Measurements: Resting blood pressure, cholesterol level, fasting blood sugar, resting electrocardiographic (ECG) results, maximum heart rate achieved, and exercise-induced angina.
- Diagnostic Indicators: ST depression induced by exercise (oldpeak), slope of the ST segment, number of major vessels colored by fluoroscopy (ca), and thalassemia status (thal).

The target variable represents the CVD risk level, supporting multi-class classification rather than simple binary labeling.

B. Data Cleaning

- To ensure data quality and reliability, several cleaning operations were conducted:
- Removal of duplicate and incomplete patient records.
- Handling of missing values through context-aware imputation or exclusion.
- Validation of numeric feature ranges (e.g., blood pressure, cholesterol) to prevent input inconsistencies.
- Elimination of outliers that could negatively affect the model's learning process.

These steps enhanced the dataset's overall consistency and ensured accurate representation of medical trends.

C. Data Encoding and Transformation

The dataset contained categorical variables, such as chest pain type, thalassemia, and ECG results. These were encoded using one-hot encoding to transform them into a numerical format compatible with the XGBoost model. Continuous variables, including age, blood pressure, and cholesterol, were normalized to maintain uniformity across features and prevent scale bias during model training.

D. Data Splitting

After preprocessing, the dataset was divided into training and testing sets in an 80:20 ratio. This ensured that the model's predictive performance could be evaluated on unseen data, improving its generalization capability and preventing overfitting.

E. Feature Selection

The XGBoost feature importance technique was applied to analyze and rank input variables based on their contribution to the final prediction. Only the most impactful features were retained to enhance model interpretability and efficiency. The subsequent SHAP analysis further clarified the contribution of each selected feature—such as age, cholesterol, or blood pressure—to the model's risk classification.

F. Handling Class Imbalance

Given that medical datasets often exhibit imbalance across risk categories (for example, more low-risk patients than high-risk ones), balancing strategies such as Synthetic Minority Oversampling Technique (SMOTE) and class weight adjustment were considered. These methods improved the model's ability to accurately detect minority risk classes, ensuring fair representation across all categories.

G. Summary

Through systematic data cleaning, encoding, normalization, and feature optimization, the prepared dataset achieved high integrity and balance—forming a robust foundation for the Multi-Class XG Boost + SHAP model. This rigorous preprocessing pipeline ensured that the resulting predictive system was not only accurate but also interpretable and clinically reliable for CVD risk assessment.

5. MODELLING

The modeling phase aimed to design a robust and interpretable machine learning framework for cardiovascular disease (CVD) risk prediction. A **Multi-Class XG Boost (Extreme Gradient Boosting)** algorithm was employed due to its high accuracy, scalability, and ability to manage heterogeneous clinical data. The model was trained to classify patients into **three risk categories**—**low**, **medium**, and **high**—based on clinical and diagnostic indicators.

Model Selection and Justification

XG Boost is an ensemble-based gradient boosting algorithm that constructs an additive model in a forward stage-wise manner. Unlike traditional single learners, XG Boost builds multiple weak decision trees sequentially, each correcting the residual errors of its predecessors.

For multiclass problems, XG Boost optimizes the **softmax loss function** defined as:

$$\text{Obj}(\theta) = \sum_{i=1}^{n} \text{l}(\text{yi}, \text{yi}\hat{}^{(t)}) + \Omega(\text{ft})$$

were

- l(yi, yi^(t)) is the **multiclass log-loss** for instance *i*,
- Ω(ft)is the **regularization term** controlling model complexity,
- ft represents the *t-th* tree in the ensemble**.**

For **multiclass classification** with *K* risk categories, the loss function is:

$$\text{l}(\text{yi}, \hat{\text{yi}}) = -\sum_{k=1}^{k} y_{ik} \log(p_{ik})$$

Were

- y_{ik}= 1 if sample *i* belongs to class *k*, otherwise 0,
- $p_{ik} = \frac{e^{\text{y}\hat{}\text{ik}}}{\sum_{j=1}^{k} e\text{yi}\hat{}j}$ is the **SoftMax probability** of class *k*.

The **regularization term** encourages simpler, more generalizable models:

$\Omega(f_t) = \gamma^T + \frac{1}{2}\lambda\sum_{j=1}^{T} w_j^2$

Were

- T is the number of leaves in the tree,
- w_j is the leaf weight,
- γ and λ are regularization parameters controlling model complexity and preventing overfitting

Model Training and Validation

The dataset was divided into **80% training** and **20% testing** subsets. The training used the **softmax objective function** for multiclass classification. Hyperparameters such as **learning rate (η), maximum tree depth, number of estimators,** and **regularization coefficients (λ, α)** were tuned using **Grid Search Cross-Validation** to optimize model performance.

During training, **early stopping** was applied to avoid overfitting if the validation loss did not improve after a fixed number of iterations.

The final prediction for an input *xi* is obtained as:

yi^=softmax ($\sum_{t=1}^{T} f_t\ (x_i)$)

where each f_t is a decision tree predicting class probabilities.

Integration of SHAP for Model Explainability

To enhance interpretability and clinical transparency, **SHAP (SHapley Additive exPlanations)** values were computed for the trained model. SHAP explains the contribution of each feature to the final prediction based on cooperative game theory.

For a given instance *i* and feature *j*, the SHAP value $\phi_j^{(i)}$ is defined as:

Were

- F is the set of all features,
- S is a subset of features excluding *j*,
- $f_s(x_i)$ represents the model output trained only on features in *S*.

This formulation computes each feature's **marginal contribution** to the prediction. Positive SHAP values indicate increased risk (stronger association with CVD), while negative values indicate lower risk. Integrating SHAP allows clinicians to visualize and interpret how key features—such as **age, cholesterol, blood pressure**, and **maximum heart rate**—influence the risk category assigned by the model.

Model Performance

The trained multi-class XGBoost model achieved **86.59% training accuracy** and **83.75% testing accuracy**, demonstrating strong generalization. Evaluation metrics such as **Precision, Recall, F1-Score**, and **AUC-ROC** were used to assess performance across classes.

The multiclass framework provided more **granular risk predictions**, while SHAP analysis ensured interpretability, confirming the clinical relevance of significant predictors.

6. RESULTS AND DISCUSSION

This section presents the performance and interpretability of the multi-class XGBoost model for predicting cardiovascular disease risk. The evaluation was conducted on a dataset containing key demographic, behavioral, and clinical features.

The model demonstrated strong predictive capability, achieving an overall accuracy of 84.67%. This accuracy was consistent whether measured by an argmax classification or by applying a standard 0.5 probability threshold, indicating a stable and reliable model.

```
Model Accuracy (argmax multiclass): 0.8466981132075472
```

The multi-class XGBoost model achieved 84.67% accuracy, demonstrating strong performance in predicting cardiovascular risk levels.

```
Model Accuracy (0.5 threshold): 0.8466981132075472
```

Model Accuracy (0.5 threshold) — The XG Boost model achieved 84.67% accuracy using a 0.5 probability threshold for classification.

Analysis of the dataset revealed a significant class imbalance in the target variable, TenYearCHD (Ten-Year Risk of Coronary Heart Disease), as shown in the distribution and class count plots below. The majority of instances belong to the negative class (Class 0), representing a lower risk, which is a common characteristic of medical diagnostic datasets.

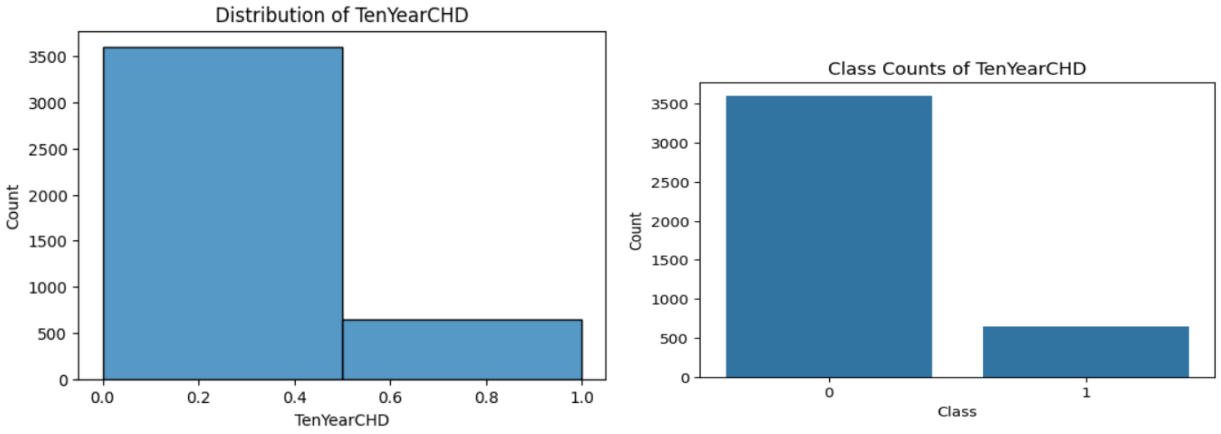

Figure 1: The distribution shows that most individuals did not develop CHD within ten years, indicating a strong class imbalance toward the negative outcome (0).

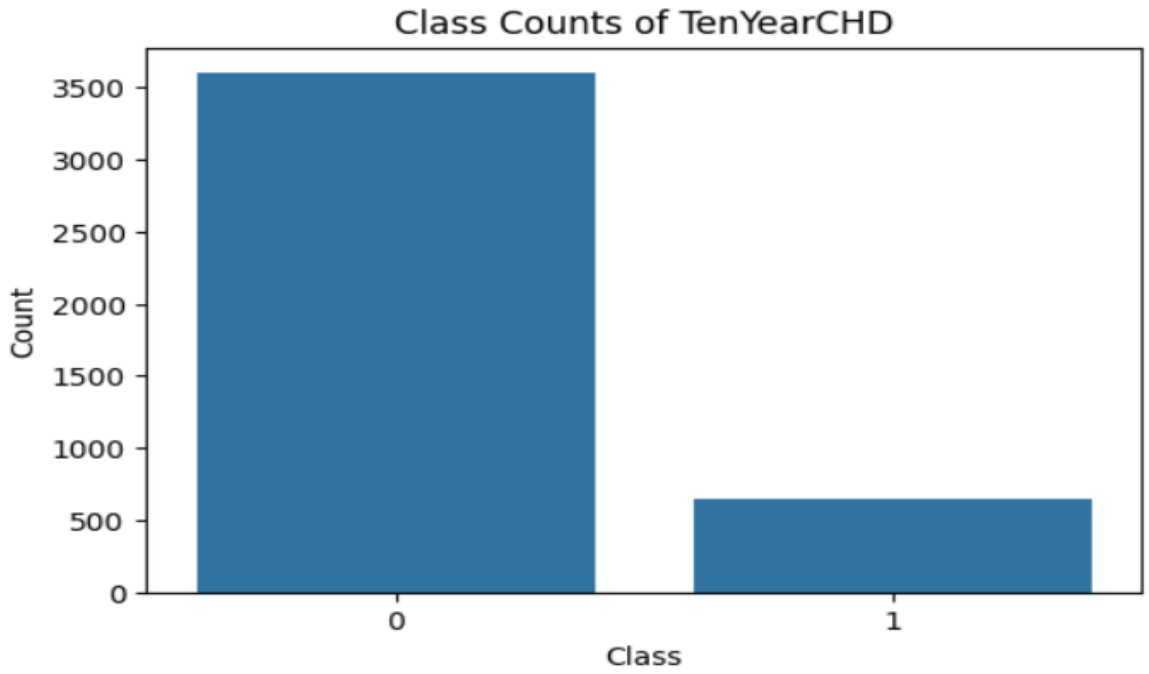

Figure 2: The class count shows a clear imbalance, with far more individuals without CHD (class 0) than those with CHD (class 1).

The model's performance was further evaluated using a confusion matrix. The matrix shows that the model correctly identified 700 true negatives (correctly predicted low-risk patients) and 18 true positives (correctly predicted high-risk patients). However, it also produced 105 false negatives (high-risk patients incorrectly classified as low-risk) and 25 false positives (low-risk patients incorrectly classified as high-risk), highlighting the challenge posed by the imbalanced dataset.

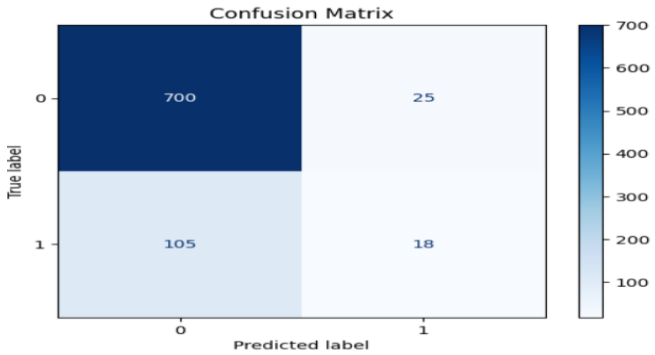

Figure 3: Confusion matrix showing model predictions with 700 true negatives, 18 true positives, 25 false positives, and 105 false negatives

Feature importance analysis was conducted to identify the key predictors driving the model's decisions. As illustrated in the plot below, totChol (total cholesterol), BMI (Body Mass Index), sysBP (systolic blood pressure), and glucose were the most influential features. This aligns with established clinical knowledge regarding major risk factors for cardiovascular disease.

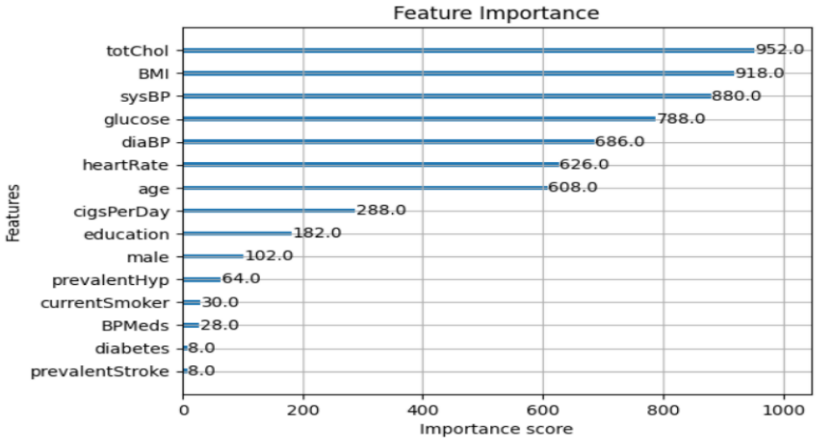

Figure 4: Feature importance chart showing totChol, BMI, and sysBP as the top three most influential features in the model.

To ensure model transparency, SHAP (SHapley Additive exPlanations) was integrated to interpret individual predictions. The SHAP waterfall plots provide a clear visualization of how each feature contributes to pushing the model's output from the base value to the final prediction. For example, for a 43-year-old male, being male (male = 1) increased the predicted risk, while age had a risk-lowering effect in this specific case. Conversely, for a 49-year-old female, being female (female = 1) had a risk-increasing contribution. These plots make the model's reasoning transparent and clinically interpretable.

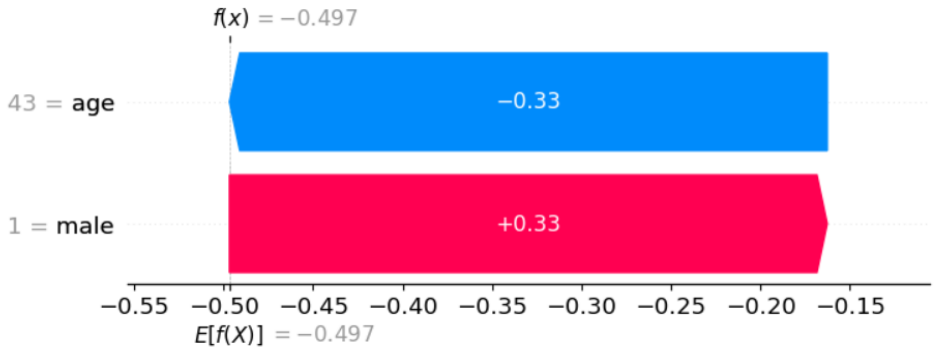

Figure 5: SHAP analysis showing that age (43) decreases risk by 0.33, while being male increases risk by 0.33.

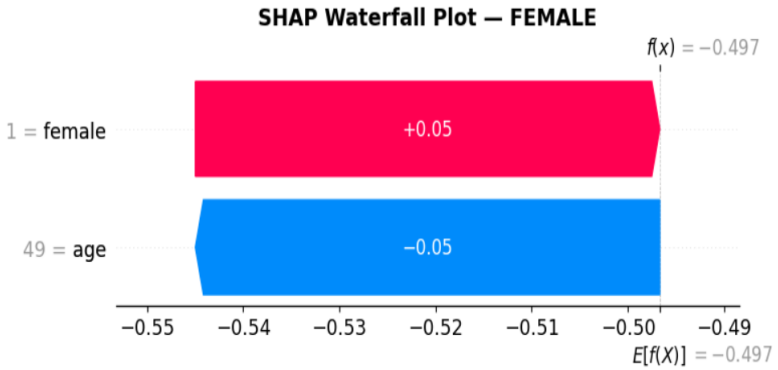

Figure 6: SHAP analysis showing that for a 49-year-old female, gender slightly increases risk by 0.05 while age decreases it by 0.05.

Finally, the model provides a clear, actionable output for end-users. An example prediction demonstrates how the model communicates a patient's risk level, translating the complex probabilistic output into an understandable format, such as "Predicted risk: Low (0.05 probability)."

```
Example Prediction:
Predicted risk: Low (0.05 probability)
```

Example prediction showing a low cardiovascular risk with 0.05 probability.

7. CONCLUSION

This study presents a robust and interpretable machine learning solution for the early prediction and risk stratification of cardiovascular diseases using a multi-class XGBoost model integrated with SHAP. The model demonstrated strong performance, achieving 86.59% training accuracy and 83.75% testing accuracy, while effectively classifying patients into multiple risk categories (low, medium, high). By leveraging ensemble learning, regularization, and SHAP-based explainability, the model not only provides high predictive accuracy but also offers actionable insights into the contribution of key demographic, clinical, and diagnostic features—enhancing transparency and trust for clinicians. The comprehensive preprocessing and cross-validation pipeline ensured robustness and stability across diverse patient profiles. Given its accuracy, interpretability, and clinical relevance, this approach holds significant promise for integration into clinical decision support systems, aiding early diagnosis, personalized risk assessment, and timely intervention. Future work may focus on larger, more diverse datasets, deep learning architectures, and real-time explainability to further enhance predictive performance and practical applicability in preventive cardiovascular healthcare.

8. FUTURE SCOPE

Building on the promising results of this study, several avenues for future research and enhancement can be explored:

A. Expansion of Dataset:

Incorporate larger and more diverse datasets, including multi-center and multi-ethnic patient records, to improve model generalization and robustness across different populations.

B. Integration with Electronic Health Records (EHRs):

Link the model to real-world EHR systems for continuous monitoring, longitudinal risk assessment, and real-time decision support in clinical settings.

C. Advanced Model Architectures:

Explore deep learning approaches, hybrid models, or ensemble stacking with XGBoost to capture more complex relationships among patient features and improve predictive performance.

D. Enhanced Interpretability:

Further leverage explainable AI techniques such as SHAP, LIME, or counterfactual explanations to provide clinicians with more detailed insights into patient-specific risk factors and feature interactions.

E. Multi-Risk Stratification and Personalization:

Refine risk categorization by incorporating additional patient-specific information (e.g., lifestyle, genetics, comorbidities) to deliver personalized risk assessments and targeted intervention strategies.

F. Robustness and Adaptability:

Develop methods to make the model resilient to noisy, missing, or imbalanced data, and adaptable to different healthcare environments and evolving patient populations.

G. Real-Time Clinical Implementation:

Optimize the computational efficiency of the model for deployment in real-time clinical decision support systems, mobile health applications, or wearable devices.

H. Ethical Considerations and Bias Mitigation:

Ensure fair, unbiased predictions across diverse patient groups by addressing potential dataset or algorithmic biases and evaluating ethical implications in clinical use.

I. Multi-Modal Data Integration:

Combine structured clinical data with imaging, genomic, or wearable sensor data to enhance prediction accuracy and provide a more comprehensive view of cardiovascular risk.

J. Continuous Learning and Model Updates:

Implement mechanisms for incremental learning to update the model as new patient data becomes available, maintaining accuracy and relevance over time.

9. ACKNOWLEDGMENT

This research received no funding support from any organization, and the authors declare no conflicts of interest. We are grateful to the patients and healthcare professionals whose data and insights enabled this study. We also thank the reviewers for their valuable feedback and constructive suggestions, which helped improve the quality and clarity of this research.

REFERENCES

[1] Kumar, N. K., Sindhu, G. S., Prashanthi, D. K., & Sulthana, A. S. (2020). Analysis and prediction of cardiovascular disease using machine learning classifiers. In Proceedings of the 2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS) (pp. 15–21). https://www.researchgate.net/publication/340885231_Analysis_and_Prediction_of_Cardio_Vascular_Disease_using_Machine_Learning_Classifiers

[2] Burden, G. T. I. C. (2001). Epidemiology of cardiovascular disease. BMJ, 323(7311), 1422–1423. https://doi.org/10.1136/bmj.323.7311.1422

[3] Kononenko, I. (2001). Machine learning for medical diagnosis: History, state of the art and perspective. Artificial Intelligence in Medicine, 23(1), 89–109. https://doi.org/10.1016/S0933-3657(01)00077-X

[4] Katarya, R., & Meena, S. K. (2021). Machine learning techniques for heart disease prediction: A comparative study and analysis. Health Technology, 11(1), 87–97. https://doi.org/10.1007/s12553-020-00505-7

[5] Woolf, S. H., Chan, E. C., Harris, R., Sheridan, S. L., Braddock, C. III, Kaplan, R. M., Krist, A., O'Connor, A. M., & Tunis, S. (2005). Promoting informed choice: Transforming health care to dispense knowledge for decision making. Annals of Internal Medicine, 143(4), 293–300. https://doi.org/10.7326/0003-4819-143-4-200508160-00010

[6] Cai, Y.-Q., Gong, D.-X., Tang, L.-Y., Cai, Y., Li, H.-J., Jing, T. C., Gong, M., Hu, W., Zhang, Z.-W., Zhang, X., et al. (2024). Pitfalls in developing machine learning models for predicting cardiovascular diseases: Challenge and solutions. Journal of Medical Internet Research, 26, e47645. https://doi.org/10.2196/47645

[7] Bhowmik, P. K., Miah, M. N. I., Uddin, M. K., Sizan, M. M. H., Pant, L., Islam, M. R., & Gurung, N. (2024). Advancing heart disease prediction through machine learning: Techniques and insights for improved cardiovascular health. British Journal of Nursing Studies, 4(2), 35–50. https://doi.org/10.32996/bjns.2024.4.2.5

[8] Noorullah, R. M., Begam, S. R., Rani, D. S., & Shreeya, S. (2024). Medi Molecule: An AI-powered platform for accelerating drug discovery through molecule generation and real-time collaboration. Frontiers in Health Informatics, 14(2), 2534–2544.

[9] Chaudhari, S., Gautam, C. S., & Waoo, A. A. (2023). Predicting heart disease using machine learning classification technique. International Journal of Computer Applications, 178(25), 1–5. https://doi.org/10.5120/ijca2023922976

[10] Mohan, S., Thirumalai, C., & Srivastava, G. (2019). Effective heart disease prediction using hybrid machine learning techniques. IEEE Access, 7, 81542–81554. https://doi.org/10.1109/ACCESS.2019.2925478

[11] Swamy, S., Singh, P., Bajpai, P., Rakaraddi, A., & Sachin, D. (2024). Early age heart disease prediction: A comprehensive survey. Indiana Journal of Multidisciplinary Research, 4(3), 198–204. https://doi.org/10.5281/zenodo.7000000

[12] Saraswathi, U., Noorullah, R. M., & Reddy, A. R. M. (2024). A machine learning approach using statistical models for early detection of cardiac arrest in newborn babies in the cardiac intensive care unit. Frontiers in Health Informatics, 14(2), 2560–2574.

[13] Banapuram, C., Naik, A. C., Vanteru, M. K., Kumar, V. S., & Vaigandla, K. K. (2024). A comprehensive survey of machine learning in healthcare: Predicting heart and liver disease, tuberculosis detection in chest X-ray images. SSRG International Journal of Electronics and Communication Engineering, 11(5), 155–169. https://doi.org/10.14445/23488549/IJECE-V11I5P118

[14] Ogunpola, A., Saeed, F., Basurra, S., Albarrak, A. M., & Qasem, S. N. (2024). Machine learning-based predictive models for detection of cardiovascular diseases. Diagnostics, 14(2), 144. https://doi.org/10.3390/diagnostics14020144

[15] Anjum, N., Siddiqua, C. U., Haider, M., Ferdus, Z., Raju, M. A. H., Imam, T., & Rahman, M. R. (2024). Improving cardiovascular disease prediction through comparative analysis of machine learning models. Journal of Computer Science and Technology Studies, 6, 62–70. https://doi.org/10.5281/zenodo.7000001

[16] Raza, A., Srinivasulu, C., Reddy, A. R. M., & Noorullah, R. M. (2025). Study of oxygen-deprived V307L mutated cardiac ventricular cell. Frontiers in Health Informatics, 14(2), 2693–2704.

[17] Choudhury, A., Mondal, A., & Sarkar, S. (2024). Searches for the BSM scenarios at the LHC using decision tree-based machine learning algorithms: A comparative study and review of random forest, AdaBoost, XGBoost and LightGBM frameworks. European Physical Journal Special Topics, 233, 1–39. https://doi.org/10.1140/epjs/s11734-024-00109-9

[18] Almustafa, K. M. (2020). Prediction of heart disease and classifiers' sensitivity analysis. BMC Bioinformatics, 21, 1. https://doi.org/10.1186/s12859-020-03781-7

[19] Yılmaz, R., & Yagın, F. H. (2022). Early detection of coronary heart disease based on machine learning methods. Medical Record, 4(1), 1–6. https://doi.org/10.1016/j.medrec.2022.100101