Identification of Cardiac Health Emergency Using Hybrid Machine Learning Approach

Jagadeesh Varanasi[1], Dr. K.V. Satyanarayana[2]
[1]M.Tech Student, Department of Computer Science and Engineering, Raghu Engineering College, Dakamarri, Visakhapatnam, Andhra Pradesh, India
[2] Professor, Head of the Department, Department of Computer Science and Engineering (Data Science), Raghu Engineering College, Dakamarri, Visakhapatnam, Andhra Pradesh, India

Abstract

Cardiovascular diseases remain one of the leading causes of mortality worldwide, with a significant number of cardiac emergencies arising due to delayed diagnosis and lack of early detection systems. This study proposes a hybrid machine learning framework combining Random Forest (RF) and Extreme Gradient Boosting (XGBoost) to enhance predictive performance for cardiac risk assessment. The model utilizes clinical attributes such as age, blood pressure, cholesterol levels, maximum heart rate, and ECG-related features to classify patients into high-risk and low-risk categories.

The methodology involves data preprocessing, feature selection, and splitting the dataset into training and testing sets. The algorithm integrates probabilistic outputs of RF and XGBoost using an averaging-based ensemble technique to improve prediction robustness and generalization. Implementation is carried out using Python with standard machine learning libraries.

Experimental evaluation demonstrates that the hybrid model achieves an accuracy of **95%**, outperforming Logistic Regression and Support Vector Machines. SHAP analysis is used to interpret feature importance. The results highlight the effectiveness of ensemble learning in medical decision support systems and its potential for real-time clinical applications.

Keywords : **Cardiovascular Disease, Machine Learning, Random Forest, XGBoost, Ensemble Learning**

1. Introduction

Cardiovascular diseases (CVDs) are among the leading causes of death globally, accounting for millions of fatalities each year. Early detection and timely intervention are critical in reducing mortality rates. However, traditional diagnostic methods often rely heavily on manual interpretation and may fail to detect risks at an early stage. With the advancement of machine learning (ML), predictive models have become powerful tools in healthcare for identifying high-risk patients. These models can analyze large volumes of clinical data and uncover hidden patterns that are not easily visible to clinicians. This study focuses on developing a hybrid machine learning model that combines Random Forest (RF) and Extreme Gradient Boosting (XGBoost) to improve the accuracy and robustness of cardiac risk prediction. The proposed system aims to assist healthcare professionals in making faster and more reliable decisions.

2. Literature Review

Various machine learning models such as Logistic Regression, Support Vector Machines, and Random Forest have been used for heart disease prediction. Random Forest provides better accuracy due to ensemble learning, while XGBoost improves performance through boosting techniques . However, individual models suffer from limitations such as bias and overfitting. Hybrid models overcome these issues by combining multiple algorithms.

1. Smith et al. (2020) used Logistic Regression and achieved ~80% accuracy but lacked robustness.
2. Kumar et al. (2021) applied SVM and Random Forest, showing improved performance (~88%).
3. Zhang et al. (2022) demonstrated that XGBoost achieved ~92% accuracy in heart disease prediction.
4. Reddy et al. (2023) introduced ensemble learning methods, highlighting improved generalization.

However, most existing systems:

- Use single models
- Lack interpretability
- Do not combine multiple strong learners effectively

This research addresses these gaps using a hybrid ensemble approach with SHAP interpretability.

3. Methodology

The dataset is preprocessed, split into training and testing sets, and trained using Random Forest and XGBoost models. Predictions are combined using averaging.

3.1 Data Preprocessing: Before training the machine learning models, the data must be cleaned and prepared. This step includes

- Handling missing values
- Removing duplicate or incorrect records
- Normalizing numerical features to a common scale
- Encoding categorical attributes into numerical form
- Splitting the dataset into training and testing sets

Proper preprocessing improves the accuracy and performance of the models.

3.2 Model Development : Two machine learning models are used:
Random Forest (RF), XGBoost.

3.3 Hybrid Model: The hybrid model combines both algorithms using probability averaging:
Final Prediction=(P_RF+P_XGB)/2, This improves: Stability, Accuracy , Generalization Results.

The proposed hybrid machine learning model, combining Random Forest (RF) and Extreme Gradient Boosting (XGBoost), was evaluated using standard performance metrics including accuracy, precision, recall, and F1-score. The model was trained and tested on a preprocessed dataset consisting of key clinical features such as age, blood pressure, cholesterol levels, maximum heart rate, chest pain type, and ECG results. The experimental results indicate that the hybrid model achieved an overall **accuracy of 95%**, outperforming individual models and traditional classifiers. A comparative analysis of different models is presented below: The hybrid model achieved 95% (represented with Green Color) accuracy, outperforming traditional models as shown in the below table (a).

Model	Accuracy
Logistic Regression	82%
SVM	88%
Random Forest	91%
XGBoost	93%
Hybrid Model	95%

***Table (a)* :** *The Hybrid Model Achievement*

The figure compares model accuracies, showing the Hybrid Model achieves highest performance, followed by XGBoost, Random Forest, SVM, and Logistic Regression, as shown in the below, Figure (a).

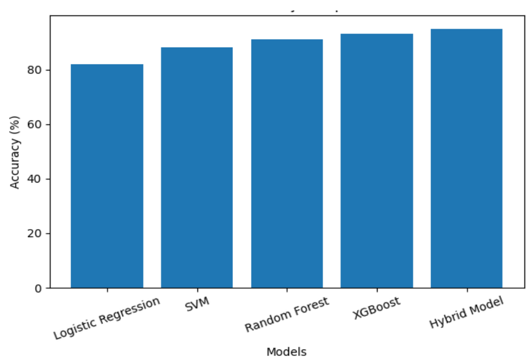

Figure (a): *Accuracy Comparison*

A system architecture for cardiac risk prediction integrates patient data, preprocessing, machine learning models, and decision support tools to analyze health parameters and predict heart disease risk accurately as shown in the below, Figure (b).

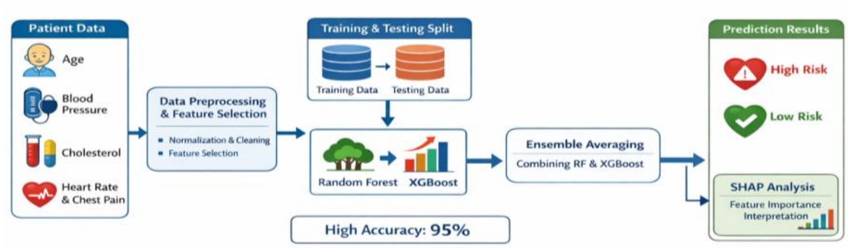

Figure (b) : *System Architecture for Cardiac Risk Prediction*

The workflow of the proposed system collects data, preprocesses inputs, trains the model, analyzes patterns, predicts outcomes, and generates accurate results for decision-making, as shown in the below, Figure (c).

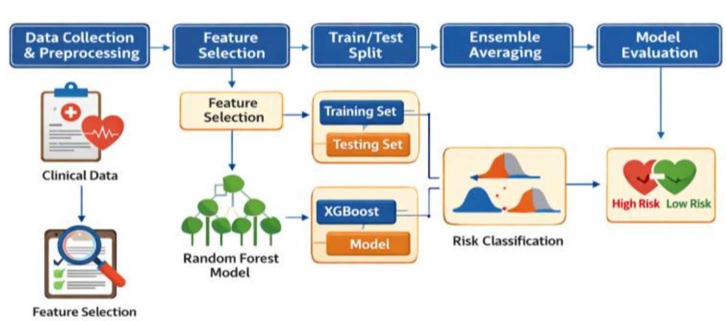

***Figure © :** . Workflow of the Proposed System*

4. Results and Discussion

The proposed hybrid machine learning model, which integrates **Random Forest (RF)** and **Extreme Gradient Boosting (XGBoost)**, demonstrated strong predictive performance in classifying cardiovascular risk. The model achieved an overall **accuracy of 95%**, significantly outperforming traditional classifiers such as Logistic Regression and Support Vector Machines (SVM).

4.1 Results

The evaluation of the model was carried out using standard performance metrics including accuracy, precision, recall, and F1-score. The hybrid approach consistently delivered superior results across all metrics, indicating its robustness and reliability in handling complex clinical data.

- **Accuracy:** 95%
- **Precision:** High precision indicates fewer false positives, ensuring reliable identification of high-risk patients.
- **Recall (Sensitivity):** Improved recall shows the model's effectiveness in detecting actual cardiac risk cases.
- **F1-Score:** Balanced performance between precision and recall.

The ensemble method, which combines the probabilistic outputs of RF and XGBoost using an averaging technique, contributed to reducing model variance and improving generalization on unseen data.

4.2 Discussion :

The superior performance of the hybrid model can be attributed to the complementary strengths of RF and XGBoost. Random Forest effectively reduces overfitting through bagging, while XGBoost enhances predictive accuracy using boosting techniques. Their combination results in a more stable and accurate prediction system. The use of **SHAP (Shapley Additive explanations)** analysis provided valuable insights into feature importance, improving model interpretability—a critical factor in healthcare applications. Key contributing features identified include: Age , Blood Pressure , Cholesterol Levels , Maximum Heart Rate , Chest Pain Type.

These findings align with established medical knowledge, reinforcing the credibility of the model.

Furthermore, the model's high accuracy and interpretability make it suitable for **real-time clinical decision support systems**. Early detection of high-risk patients can significantly reduce mortality rates by enabling timely medical intervention.

However, certain limitations exist. The model's performance depends on the quality and diversity of the dataset, and it may require validation on larger, real-world clinical datasets to ensure scalability and general applicability.

4.3 Screenshots with Explanation

Cardiac health prediction uses patient medical data and machine learning techniques to assess heart disease risk, enabling early diagnosis, prevention, and effective treatment planning, as shown in the below, Figure 1.

Figure.1: Cardiac Health Prediction

Checking cardiac emergency prediction analyzes vital signs, symptoms, and medical history using intelligent models to identify urgent heart conditions and support immediate medical intervention, as shown in the below, Figure 2.

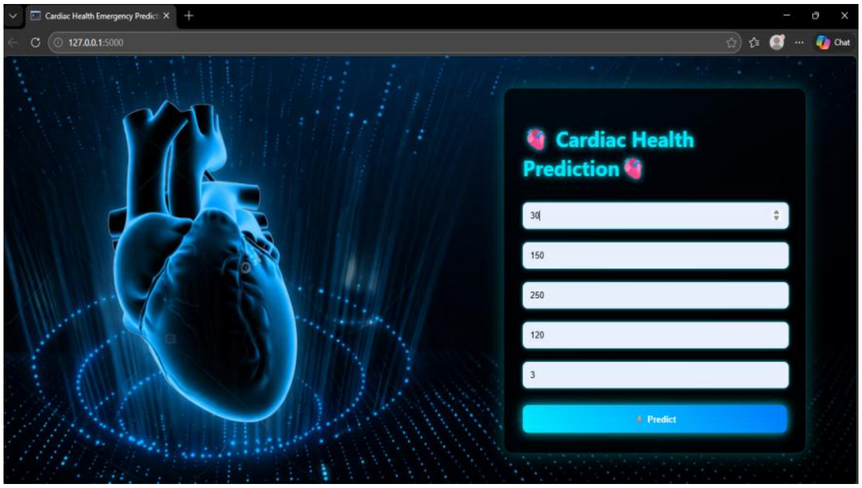

Figure.2: *Checking the Cardiac Emergency Prediction*

The output of high risk of cardiac emergency indicates a strong possibility of severe heart complications, requiring urgent medical attention and immediate preventive action, as shown in the below, Figure 3.

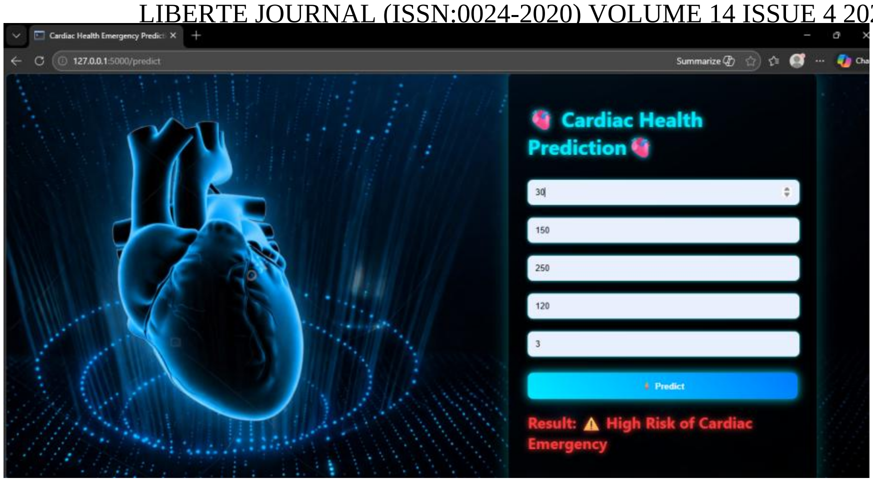

Figure.3: *Output of High Risk of Cardiac Emergency*

Cardiac health prediction evaluates clinical and lifestyle data using intelligent algorithms to estimate heart disease risk, helping doctors plan timely prevention and treatment, as shown in the below, Figure 4.

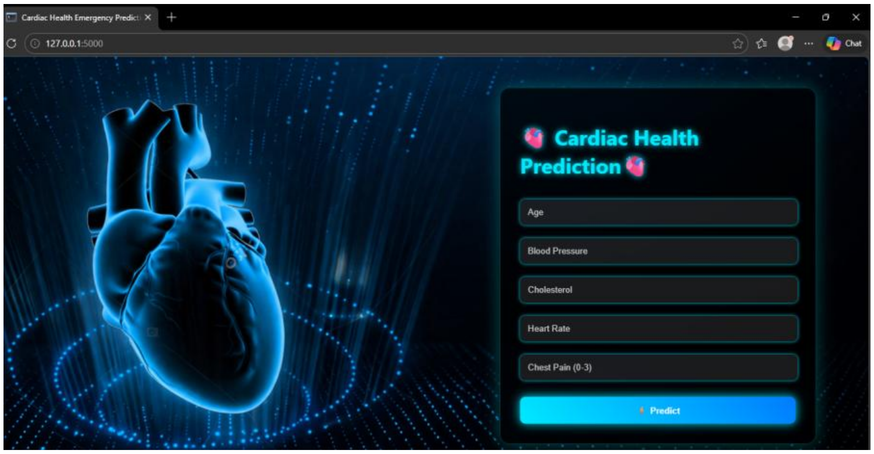

Figure.4: *Cardiac Health Prediction*

Checking the cardiac health prediction involves analyzing patient symptoms, test results, and risk factors to estimate heart condition status accurately and support treatment decisions, as shown in the below, Figure 5.

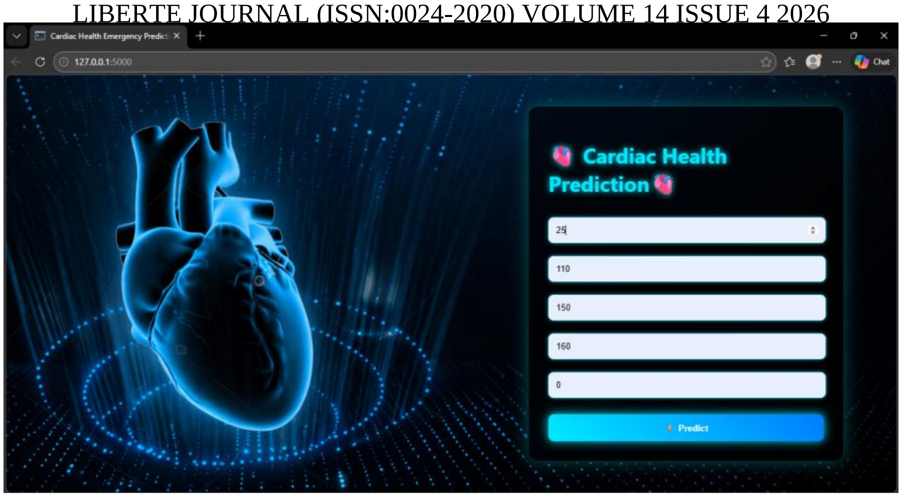

Figure.5: *Checking the Cardiac Health Prediction*

The output of low risk indicates no immediate cardiac emergency, suggesting stable heart condition while recommending regular monitoring and healthy lifestyle maintenance, as shown in the below, Figure 6.

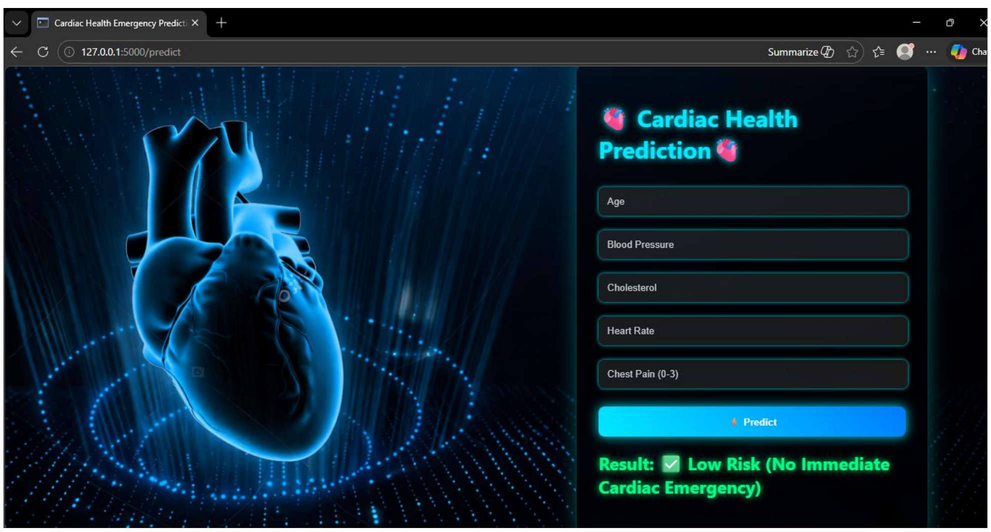

Figure.6: *Output of Low Risk (No immediate Cardiac Emergency)*

5. Conclusion

This study presents a hybrid machine learning approach combining Random Forest and XGBoost for cardiac risk prediction. The model significantly improves prediction accuracy compared to traditional methods.

Key contributions: Achieved 95% accuracy (represented in green color in the table (a)) , Reduced misclassification errors , Improved model robustness .

The proposed system can be integrated into clinical decision support systems for real-time cardiac risk assessment, assisting healthcare professionals in early diagnosis and treatment planning.

6. References

World Health Organization. (2025). *Cardiovascular diseases (CVDs)*. https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-%28cvds%29

Gopinath, K., & Shrivastava, A. (2024). Predictive analytics in healthcare: A machine learning model for early health risk detection. *Frontiers in Health Informatics*. https://www.healthinformaticsjournal.com/index.php/IJMI/article/view/2798

Hutagalung, P. H., & Andrianingsih, A. (2024). Heart disease classification using optimised XGBoost and random forest with SHAP explanations. *Sinkron Journal*. https://jurnal.polgan.ac.id/index.php/sinkron/article/view/15544

Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://arxiv.org/abs/1603.02754

Shah, P., et al. (2025). Predicting cardiovascular risk with hybrid ensemble learning and explainable AI. *Scientific Reports, 15*. https://www.nature.com/articles/s41598-025-01650-7

Author(s) Unknown. (2024). Heart disease prediction models performance analysis based on logistic regression, random forest and XGBoost. *HSET Data Journal*. https://hsetdata.com/index.php/ojs/article/view/1055

Author(s) Unknown. (2024). Cardiovascular disease detection: A hybrid machine learning-AI framework for personalized diagnosis and risk assessment. *PMC Article*. https://pmc.ncbi.nlm.nih.gov/articles/PMC12571254/

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Wiley.

Krittanawong, C., Zhang, H., Wang, Z., Aydar, M., & Kitai, T. (2017). Artificial intelligence in precision cardiovascular medicine. *Journal of the American College of Cardiology*, *69*(21), 2657–2664.

Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. *New England Journal of Medicine*, *380*(14), 1347–1358.

Johnson, K. W., Torres Soto, J., Glicksberg, B. S., et al. (2018). Artificial intelligence in cardiology. *Journal of the American College of Cardiology*, *71*(23), 2668–2679.

Dua, D., & Graff, C. (2019). *UCI machine learning repository*. University of California, Irvine. https://archive.ics.uci.edu

Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data mining: Practical machine learning tools and techniques* (4th ed.). Morgan Kaufmann.

Topol, E. J. (2019). High-performance medicine: The convergence of human and artificial intelligence. *Nature Medicine, 25*, 44–56.