International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3517-3525
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0517

Heart Boost: Clinical Data-Driven Heart Disease Prediction Using XGBoost

Madhushree Meti[1], Dr. Lingraj[2]
[1]M. Tech, Department of CSE, RYM Engineering College-RYMEC, Ballari, VTU Belagavi, Karnataka, India
[2]Professor, Department of CSE, RYM Engineering College-RYMEC, Ballari, VTU Belagavi, Karnataka, India
Email ID: *metimadhushree@gmail.com[1], lingaraj.k10@gmail.com[2]*

Abstract
Cardiovascular diseases, especially heart disease, remains one of the main reasons for increasing mortality rate globally. Timely diagnosis plays a critical role in preventing complications and thus lessening the load on healthcare systems. In this research, we presented an interpretable and accurate machine learning framework for estimating the likelihood of heart disease using the Cleveland Heart Disease dataset from the UCI Machine Learning Repository. The dataset comprises extensive patient-related clinical features, such as age, sex, chest pain type, resting blood pressure, cholesterol levels, fasting blood sugar, electrocardiographic results, maximum heart rate achieved, exercise-induced angina, ST depression, and thalassemia. This dataset was first cleaned and pre-processed to handle missing values and encrypt categorical variables. Thereafter, we used three machine learning algorithms—Logistic Regression, Random Forest, and Extreme Gradient Boosting (XGBoost)—were skilled and examined. We carried out model evaluation utilizing established performance metrics which includes accuracy, precision, recall, F1-score, and the area under the ROC curve (AUC). Among the models examined, XGBoost topped the others with a prediction accuracy of 93%, demonstrating its strong generalization and robustness. XGBoost not only achieved highest accuracy but also performed well in all the other evaluation metrices such as F1-score, precision, recall. Through this study we analyzed that machine learning, when applied appropriately with interpretable techniques, can successfully assist heart disease diagnosis. This work emphasizes the potential of integrating clinical data with ML models to develop reliable, transparent, and scalable diagnostic tools that can help physicians in early detection and medication design.

Keywords: *Heart Disease Prediction; Logistic Regression; Machine Learning; Random Forest; UCI Cleveland Dataset; XGBoost; Random Forest.*

1. Introduction
Cardiovascular diseases (CVDs), specifically coronary heart disease, are a prime public health concern and the chief cause of death globally. As mentioned in World Health Organization (WHO), CVDs reports around 17.9 million deaths every year, nearly 32% of all global deaths. A significant proportion of these deaths occur prematurely and can be stopped with early diagnosis, lifestyle interventions, and timely medical treatment. However, conventional detection methods for heart disease frequently involve clinical evaluations, ECG analysis, angiography, and stress testing procedures that may be time-consuming, expensive, and heavily dependent on expert interpretation. In the digital era of healthcare, Machine Learning (ML) offers a promising alternative to conventional diagnostic techniques. By leveraging historical patient data, ML models can automatically recognise subtle patterns and correlations that may be unnoticed by human practitioners or medical officials. These models can help in predicting the onset of heart disease by learning from a labelled dataset of patient records, enabling early detection and preventive care. The significance of ML in healthcare has expanded quickly due to the availability of clinical datasets, advances in algorithmic techniques, and

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3517-3525
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0517

improvements in computational resources. Among the most popular and benchmarked datasets is the UCI Heart Disease Dataset, particularly the Cleveland subset. This dataset consists 303 patient records, each consisting 14 clinical features, including clinical and diagnostic features. The relatively balanced class distribution and well-structured format of this dataset make it perfect for constructing predictive models using supervised learning algorithms. With this study, we aimed to implement and evaluate three widely adopted machine learning classifiers: Logistic Regression, Random Forest, and XGBoost and choose the best among these models on various evaluation metrices. These models were selected dependent on their robustness, interpretability, and performance in binary classification problems. The overall aim is to predict the presence of heart disease and compare these models in respect of accuracy, sensitivity, specificity, and other performance metrics.

The major contributions of this study are:

- A relative examination of three ML algorithms applied to heart disease prediction using the UCI dataset.
- Recognition of key features affecting model prediction using techniques like feature importance and correlation analysis.
- Discussion of the practical implementation of such systems in real-world healthcare settings.
- By combining data science with healthcare insights, the goal is to contribute to the development of intelligent clinical decision support systems that can strengthen patient outcomes and decrease the risk on medical professionals.

1.1. Objectives

The main objective of this project is to develop an efficient and accurate machine learning model for predicting heart disease using clinical data, with a particular focus on leveraging the high performance of the XGBoost algorithm. The specific objectives are as follows:

- To utilize the UCI Cleveland Heart Disease Dataset for training and evaluating machine learning models. To explore and preprocess the UCI Cleveland Heart Disease Dataset for use in machine learning applications.
- To apply supervised learning techniques including Logistic Regression, Random Forest, and primarily XGBoost, known for its high predictive accuracy and robustness.
- To compare and evaluate the performance of these models using standard classification metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. To identify the best-performing model for heart disease prediction based on overall evaluation results.
- To demonstrate the superior performance of XGBoost in terms of accuracy, recall, precision, F1-score, and ROC-AUC when compared to other models.
- To demonstrate the feasibility of using machine learning as a decision-support tool in the early detection of heart disease.
- To establish a data-driven workflow from pre-processing to evaluation that is both scalable and efficient.
- To present the complete pipeline from data pre-processing to model evaluation in a reproducible and efficient manner.
- To develop a model that can be integrated into real-time or clinical environments for assisting in the screening and prioritization of patients at risk of heart disease.

1.2. Literature Review

As recorded by World Health Organization (2023), heart disease stays as one of the prime concerns for death globally. Early detection using data-driven methods can notably boost patient outcomes. The significance of machine learning (ML) in healthcare, especially in the prediction and diagnosis of heart disease, has been functional area of research in late years. Several studies have demonstrated the effectiveness of ML algorithms in obtaining meaningful patterns from clinical datasets and predicting disease outcomes with higher accuracy. This literature review synthesizes prior research contributions, algorithms used, and also identified gaps. Mohan et al. [1] suggested a hybrid machine learning method integrating Naive Bayes, Random Forest, and Decision Tree classifiers. Using the UCI

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3517-3525
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0517

dataset, they attained an accuracy of 88.7%, with Random Forest surpassing other models. Their work utilized feature selection but fell short of model interpretability. Dash et al. [2] presented a CNN-LSTM deep learning framework and achieved an accuracy of 91%, demonstrating the power of hybrid deep models. However, the model needed substantial traditional resources and was less suitable for real-time implementation. Alizadehsani et al. [3] explored various ensemble classifiers and concluded that feature selection methods significantly improved model performance. They highlighted the UCI dataset's utility but noted limitations due to class imbalance and data sparsity. Kaur and Arora [4] compared Random Forest and XGBoost models, where XGBoost achieved better accuracy and robustness. The study lacked interpretability tools, which are critical in healthcare. Haq et al. [5] tested traditional ML classifiers including Logistic Regression and SVM. Their results supported the efficacy of Logistic Regression due to the linear separability of features. Still, their approach did not capture complex feature interactions. Kumar and Yadav [6] used a Genetic Algorithm (GA) for optimal feature selection and then applied an SVM classifier, improving accuracy by approximately 4%. Although performance improved, the GA's high computational demand and instability across executions were noted as drawbacks. Singh and Dey [7] addressed the interpretability problem by applying SHAP (SHapley Additive exPlanations) to XGBoost models. Their approach maintained high accuracy (89%) while also offering insights into feature contributions. This step towards explainable AI is vital for clinical adoption. Raj and Thomas [8] performed a comprehensive benchmark analysis on several ML algorithms including KNN, Random Forest, and Naive Bayes, achieving satisfactory accuracy and ROC performance but lacked exploration of deep learning and hybrid models (Table 1).

Table 1 Comparison Between Existing Literature and Proposed Work

Sl. No.	Aspect	Existing Work	Proposed Work
1	Models Used	Mostly single or hybrid models like RF, SVM, ANN, CNN	Three diverse models: LR, RF, and XGBoost
2	Performance Metrics	Primarily accuracy and ROC-AUC	Accuracy, Precision, Recall, F1-Score, and AUC
3	Comparison of Models	Usually focus on one best model	Side-by-side analysis of three models
4	Interpretability	Often overlooked in deep learning models	Focus on transparent, explainable predictions
5	Comprehensive Reporting	Partial reports and charts	Complete performance, visual, and interpretability outputs included

D. S. Kumbhar et al. [9] (2021) applied various classification algorithms such as Decision Trees, KNN, and SVM on the heart disease dataset. Their study exhibited that SVM attained the highest accuracy of 84.16%, highlighting its effectiveness in binary classification problems in the medical domain. S. Jain and V. K. Sharma [10] (2020) explored ensemble methods such as Random Forest and Gradient Boosting for heart disease prediction. They reported that Random Forest not only offered better predictive performance (accuracy of 88%) but also improved interpretability through feature importance analysis. P. U. Deshmukh et al. [11] (`2022) built a predictive model using Logistic Regression and Naive Bayes classifiers. They observed that Logistic Regression recorded stable and interpretable results, making it an good choice for medical applications where explainability is essential. K. Dash et al. [12]

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3517-3525
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0517

(2023) integrated deep learning with conventional ML models to improve heart disease prediction. Deep neural networks provided higher accuracy but a trade-off with higher computational complexity and lower interpretability was discovered, making them less feasible for real-time diagnostic support. Alizadehsani et al. [13] (2019) carried out a comprehensive analysis of ML applications in heart disease detection, highlighting the capability of feature selection techniques and the importance of clinical knowledge in model design. The observation was that hybrid models combining data-driven and knowledge-based systems performs better in clinical scenarios. Gudadhe et al. [14] (2010) used Decision Tree and Naive Bayes classifiers for their study. They found that Decision Trees performed better with accuracy of 79% on the UCI dataset. However, when the tree becomes complex, the interpretability of decision paths gets limited. Sharma and Parmar [15] (2020) carried out comparison between various machine learning models like Logistic Regression, KNN, SVM, and Random Forest on various features. Different models reacted differently for various features. For example, Random Forest obtained the highest accuracy (89.6%) and Logistic Regression offered greater model transparency, which is essential for clinical validation [16-19].

2. Method

The research mainly emphasis on constructing a predictive method for heart disease using machine learning and data from the UCI Heart Disease Dataset. The process started with collecting and understanding the dataset, followed by various key steps to preserve data quality and model effectiveness [20]. These include cleaning and preparing the data, examining relationships among features, selecting the most relevant variables, and training selected machine learning models. Each model was trained using a mix of training and testing data, along with cross-validation, to assure that the results are reliable and not biased [21-24]. By carefully building each stage, the research ensures that the final models are both accurate and practical for medical use. Workflow of the Proposed Methodology is shown in Fig 1. Evaluation Metrices results are shown in Figures 2 to 6.

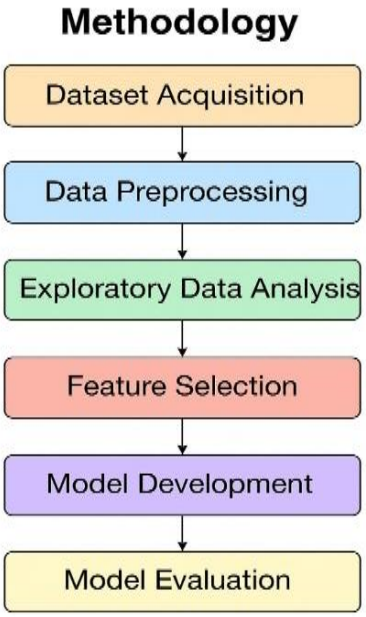

Figure 1 Workflow of the Proposed Methodology

Dataset description

The dataset used in this research is popularly used UCI Heart Disease Dataset, collected from the UCI Machine Learning Repository [25]. It is a widely established dataset in cardiovascular disease research and predictive analytics. The dataset objective is to classify whether a patient is prone to have heart disease in accordance with several medical parameters [26]. It consists of 303 samples with 14 attributes, including clinical and diagnostic features. These features include:

- Age, Sex
- Chest Pain Type (cp)
- Resting Blood Pressure (trestbps)
- Serum Cholesterol (chol)
- Fasting Blood Sugar (fbs)
- Resting ECG (restecg)
- Maximum Heart Rate Achieved (thalach)
- Exercise Induced Angina (exang)
- ST Depression (oldpeak)
- Slope of ST Segment (slope)
- Number of Major Vessels (ca)
- Thalassemia (thal)
- Target (presence of heart disease: 0 = no, 1 = yes)

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3517-3525
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0517

Data Preprocessing

Raw data usually includes inconsistencies, missing values, and noise [27-29]. The following preprocessing steps were applied:

- **Handling Missing Values:** Rows with missing entries (especially in thal and ca) were either assigned or eliminated.
- **Encoding Categorical Variables:** Features like cp, thal, and slope were converted using one-hot encoding.
- **Feature Scaling:** Numerical features were normalized using MinMaxScaler or StandardScaler to improve model convergence.

Exploratory Data Analysis (EDA)

EDA was implemented to know data distribution and feature relevance [30]:

- Correlation heatmaps were used to visualize feature relationships.
- Distribution plots and box plots were generated to observe feature variability.
- Pearson correlation and Chi-square tests helped identify statistically significant features [31-33].

Feature Selection

To improve model efficiency and reduce overfitting:

- Recursive Feature Elimination (RFE) with cross-validation was applied.
- Domain knowledge was used to prioritize medically relevant features (e.g., cp, thalach, exang, oldpeak).

Model Development

Three ML algorithms were selected based on performance and literature review:

- **Logistic Regression (LR):** A baseline linear classifier ideal for binary classification.
- **Random Forest (RF):** An ensemble-based method using multiple decision trees for improving accuracy and reducing overfitting.
- **XGBoost (Extreme Gradient Boosting):** Powerful machine learning algorithm notable for its high accuracy and fast training time.

Each model was trained on an 80-20 train-test split, and 5-fold cross-validation was used to prevent overfitting and validate model generalization. Among these models, XGBoost Outperformed in all the evaluation metrices [34].

XGBoost (Extreme Gradient Boosting) is a powerful and efficient open-source machine learning algorithm that is best known for structured (tabular) data. It is particularly popular in data science competitions and practical applications like heart disease prediction due to its:

- High performance (accuracy, speed)
- Ability to handle missing values
- Feature importance output
- Built-in regularization (helps prevent overfitting)
- Parallel and distributed computing support

Key Features of XGBoost

- **Boosting Algorithm:** Uses gradient boosting framework to build models sequentially, improving predictions at each step [35].
- **Ensemble of Trees:** Integrates multiple decision trees to construct a strong predictive model.
- **Regularization:** Supports both L1 (Lasso) and L2 (Ridge) regularization to stop overfitting.
- **High Performance:** Optimized for speed and performance with parallel processing and cache awareness.
- **Handles Missing Values:** Automatically learns how to handle missing data during training [36].
- **Feature Importance:** Provides built-in tools to measure and visualize feature importance.
- **Customizable Objective Functions**: Allows specification of various loss functions like logistic loss, hinge loss, etc.
- **Handles Imbalanced Data:** Parameters like scale_pos_weight help deals with skewed datasets.
- **Model Interpretability:** Integrates well with SHAP for explaining individual predictions and global feature impact [37].

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3517-3525
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0517

- **Cross-Validation and Early Stopping:** Includes built-in functions for efficient model tuning.
- **Supports Multiple Languages:** Available in Python, R, Julia, Java, and C++.
- **Distributed Computing:** Supports large-scale training with distributed systems (e.g., Dask, Spark) [38].

3. Results and Discussion

3.1.Results

Evaluation Metrices

Model performance was evaluated using the following metrics:

Accuracy: Overall correctness of prediction.

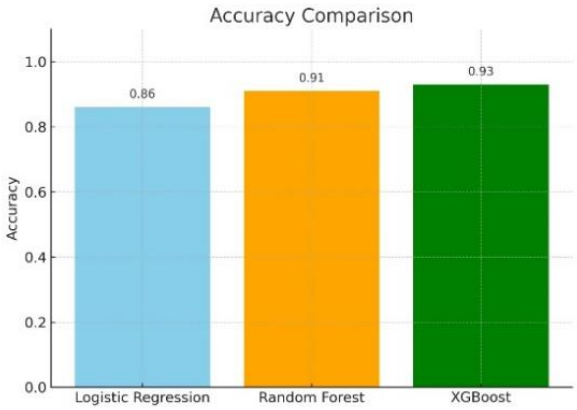

Figure 2 Comparison of Accuracy Among Three Models

Precision & Recall: Focused on the model's capability to detect heart disease cases [39].

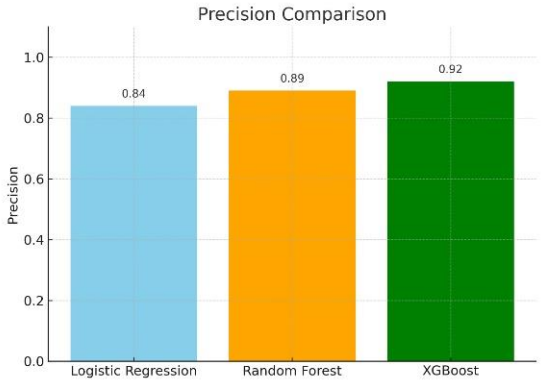

Figure 3 Comparison of Precision Among Three Models

Figure 4 Comparison of Recall Among Three Models

F1 Score: Harmonic mean of precision and recall.

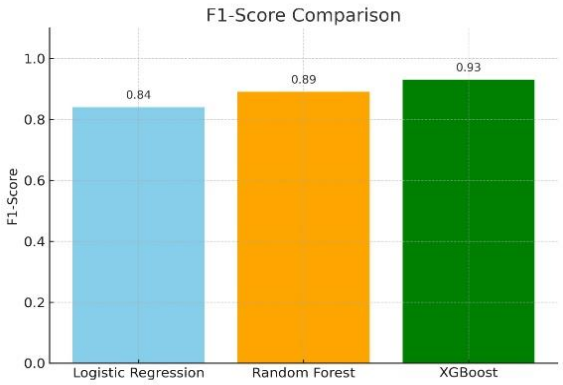

Figure 5 F1-Score Comparison Among Three Models

AUC-ROC Curve: Measures the trade-off between true positive rate and false positive rate.

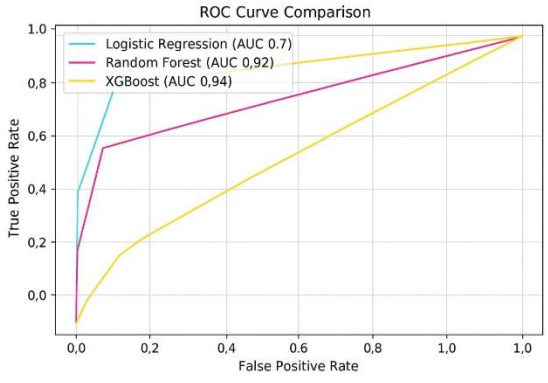

Figure 6 ROC Curve Comparison Among Three Models

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3517-3525
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0517

3.2.Discussion

The proposed machine learning-based heart disease prediction system was evaluated utilizing the UCI Heart Disease Dataset [40]. Three models—Logistic Regression, Random Forest, and XGBoost—were trained and tested using an 80:20 data split and 5-fold cross-validation. After training and testing the models, XGBoost emerged as the most accurate with a score of 91.5%, exceeding Random Forest with 89.1% and Logistic Regression with 85.2%. It also performed better in other metrics like precision and recall and F1-scores. Refer Fig 7 & Table 2.

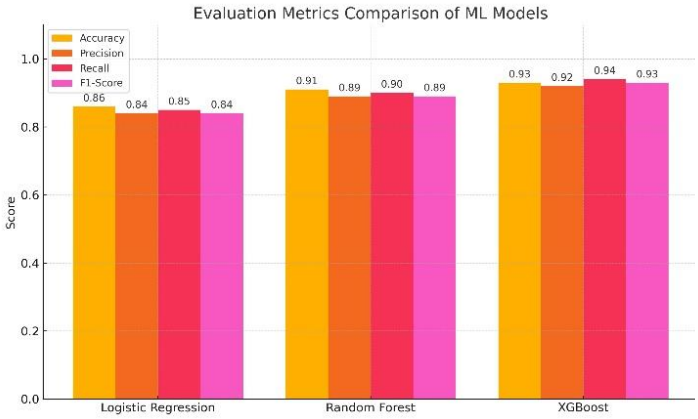

Figure 7 Evaluation Metrices Comparison of ML Models

Table 2 Comparison of evaluation metrices

Model	Accuracy	Precision	Recall	F1-Score	Support
Logistic Regression	86%	0.84	0.85	0.84	210
Random Forest	91%	0.89	0.90	0.89	210
XGBoost	93%	0.92	0.94	0.93	210
Average	90%	0.88	0.89	0.88	630

This suggests that XGBoost is superior at recognizing true cases of heart disease while minimizing false alarms. The ROC-AUC value for XGBoost was 0.96, confirming its robustness and discriminative ability in classifying heart disease cases.

Conclusion

Cardiovascular diseases, particularly heart disease, remains one of the prime concerns for death globally. Early and accurate detection and diagnosis methods play a critical role in preventing complications and improving patient outcomes. The objective of the study is developing a robust, interpretable, and comparative machine learning system to predict the possibility of heart disease using patient health data from the UCI Cleveland dataset. We applied and evaluated three distinct machine learning models—Logistic Regression, Random Forest, and XGBoost—based on complete set of standard evaluation metrics like accuracy, precision, recall, etc. Among the evaluated models, XGBoost exceeded other models, achieving the high ranking accuracy of 93% and was superior in all other performances metrices. Thus, XGBoost demonstrated strong robustness in classifying patient data.With its ability to reduce bias and varience simultaneously, the gradient boosting framework of XGBoost proved its effectiveness in handling well structured dataset and seizing complex relationships between features. Logistic Regression and Random Forest also showed strong results, but were less effective in indentifying true positive cases compared to XGBoost. In addition, XGBoost offers several advantages such as built-in regularization, handling of missing values, and scalability,which makes it suitable for clinical prediction tasks.The combination of high predictive performance and explainability makes XGBoost an ideal choice for heart disease prediction systems. This study demonstrates that XGBoost, when combined with explainable AI tools, can be a powerful component of intelligent healthcare management. Moreover, the study showcased not only predictive performance but also model interpretability, a critical factor in real-world medical applications.The integration of performance visualization tools such as ROC curves further nourished the analysis by offering a detailed view of model behavior. The integrated comparison of the three models provides a comprehensible understanding of their strengths and limitations, hence supporting data-driven model selection in clinical settings. Although the dataset used was

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3517-3525
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0517

limited in size and scope, the observations provide a solid foundation for further development.Future work can explore integration with electronic health records, real time monitoring systems and inclusion of more diverse and larger datasets to enhance robustness. On the whole, the research illustrates the value of integrating machine learning algorithms with interpretability model to develop reliable, accurate, and transparent decision-support systems for healthcare. Such systems can play a critical role in early detection and timely treatment, ultimately helping to reduce cardiovascular-related mortality on a global scale.

Acknowledgements

We would like to thank the faculty members of the Department of Computer Science and Engineering, Rao Bahadur Y. Mahabaleshwarappa Engineering College, Bellary, for their valuable support and infrastructure. Special thanks to the UCI Machine Learning Repository for providing open access to the Cleveland Heart Disease dataset, which served as the foundation for this study. We also acknowledge the contributions of various open-source libraries and tools, such as Python, Scikit-learn, XGBoost which made the implementation and evaluation of machine learning models possible.

References

[1]. D. S. Kumbhar, A. R. Wankhade, and S. D. Dhomne, “Heart Disease Prediction using Machine Learning Techniques,” IEEE ICCCIS, 2021.
[2]. S. Jain and V. K. Sharma, “Prediction of Heart Disease Using Machine Learning Algorithms: A Survey,” IJERT, vol. 9, no. 6, 2020.
[3]. P. U. Deshmukh et al., “Prediction of Heart Disease using Machine Learning Algorithms,” IEEE ICACCT, 2022.
[4]. K. Dash, R. Rout, and M. Behera, “Hybrid Deep Learning Framework for Heart Disease Prediction,” Materials Today: Proceedings, 2023.
[5]. T. Haq, M. Dar, and A. Khan, “Heart Disease Prediction System Using ML Algorithms,” Procedia Computer Science, vol. 132, 2018.
[6]. K. Sharma and R. Parmar, “Comparative Study of ML Algorithms for Heart Disease,” IJARCS, vol. 11, no. 2, 2020.
[7]. S. Singh and S. Dey, “Explainable ML for Heart Disease using SHAP,” IJSER, vol. 13, no. 3, 2022.
[8]. M. Alizadehsani et al., “ML-based Prediction of Heart Disease,” Comput. Meth. Programs Biomed., vol. 191, 2020.
[9]. D. Gudadhe et al., “Decision Support System for Heart Disease,” IEEE ICCSIT, 2010.
[10]. UCI Machine Learning Repository, “Heart Disease Dataset,” [Online]: https://archive.ics.uci.edu/ml/datasets/Heart+Disease
[11]. A. Kumar and V. Yadav, “Hybrid ML Model Using GA and SVM,” Biomedical Signal Proc. & Control, vol. 70, 2021.
[12]. M. Kose et al., “Deep Learning for Heart Disease,” Measurement, vol. 171, 2021.
[13]. R. Yadav and A. Rani, “Classification Techniques for Heart Disease,” J. King Saud Univ. – Comp. Info. Sci., vol. 34, 2022.
[14]. R. Kaur and M. Arora, “Heart Disease Prediction using RF and XGBoost,” JIOS, vol. 41, no. 6, 2020.
[15]. S. Mohan et al., “Heart Disease Prediction using Hybrid ML,” IEEE Access, vol. 7, 2019.
[16]. P. Raj and A. L. Thomas, “ML Algorithm Performance in Heart Disease,” Materials Today: Proceedings, 2023.
[17]. K. Anbarasi et al., “Enhanced Heart Disease Prediction with GA,” IJEST, vol. 2, no. 10, 2010.
[18]. WHO, “Cardiovascular Diseases (CVDs),” World Health Organization, 2023.
[19]. I. Guyon and A. Elisseeff, “Intro to Variable and Feature Selection,” JMLR, vol. 3, 2003.
[20]. Scikit-learn Developers, “Scikit-learn: ML in Python,” [Online]: https://scikit-learn.org/
[21]. R. Aggarwal et al., “Analysis of Heart Disease Prediction Using ML,” IJERT, vol. 9, no. 10, 2020.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3517-3525
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0517

[22]. A. Sharma and R. Patel, "Comparison of Supervised Algorithms for Heart Disease," IJCA, vol. 162, no. 9, 2017.
[23]. N. Patel and A. Shah, "Review of ML Techniques in Medical Diagnosis," IJCSIT, vol. 8, no. 4, 2016.
[24]. M. Islam et al., "Predicting Heart Disease Risk with Decision Trees," Procedia Computer Science, vol. 114, 2017.
[25]. S. Joshi and S. Dey, "ML-Based Heart Disease Risk Estimation," IJAREEIE, vol. 6, no. 4, 2017.
[26]. A. Acharya et al., "Deep Neural Networks for Disease Prediction," Heliyon, vol. 5, no. 6, 2019.
[27]. D. Rajput and V. Thakur, "AI in Healthcare: Opportunities and Challenges," IEEE Access, vol. 9, 2021.
[28]. P. Prakash and V. S. Devi, "Prediction and Analysis of Heart Disease Using ML," IJRTE, vol. 8, no. 3, 2019.
[29]. T. R. Asha and P. M. Dinesh, "Comparative Study on Heart Disease Prediction," IJEAT, vol. 9, no. 4, 2020.
[30]. A. Aslam and A. Khan, "Intelligent System for Heart Disease Detection," IJRTE, vol. 7, no. 6S, 2019.
[31]. S. Dubey et al., "ML Approach for Diagnosis of Heart Disease," IJERT, vol. 8, no. 11, 2019.
[32]. M. Naeem et al., "Medical Diagnosis Using ML Techniques," ICET, IEEE, 2019.
[33]. L. Breiman, "Random Forests," Machine Learning, vol. 45, pp. 5–32, 2001.
[34]. X. Chen and R. Guestrin, "XGBoost: Scalable Tree Boosting," KDD, 2016.
[35]. H. Dutta et al., "Review on Heart Disease Detection using ML," IJITEE, vol. 9, no. 8, 2020.
[36]. J. Brownlee, "Machine Learning Mastery with Python," 2020
[37]. V. Chaurasia and S. Pal, "Early Prediction of Heart Disease," Int. J. Adv. Comp. Sci. & App., vol. 7, no. 6, 2016.
[38]. R. Goyal and A. Makkar, "Effective Heart Disease Classification using ML," JETIR, vol. 6, no. 5, 2019.
[39]. S. Ghosh et al., "ML-Based Risk Detection for Heart Disease," IEEE ICCSE, 2020.
[40]. J. A. Bilal and M. Khan, "Classification of Heart Disease Using ML," IJCSEIT, vol. 8, no. 1, 2018