International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 6, pp 23-28, June-2025

https://doi.org/10.47001/IRJIET/2025.906004

Early Heart Disease Prediction Using Machine Learning Algorithm

[1]Arpita Gangadhar Awate, [2]Shweta Rajendra Tirpude, [3]Mangla Ganpat Bhoyar, [4]Asst. Prof. Suraj S. Bankar

[1,2,3]Student, Computer Science and Engineering, Shri Sai College of Engineering and Technology Bhadrawati, Chandrapur, India

[4]Assistant Professor, Computer Science and Engineering, Shri Sai College of Engineering and Technology Bhadrawati, Chandrapur, India

***Abstract* - One of the most common health problems in the world, cardiovascular disease accounts for around 32% of all fatalities yearly. Effective treatment and illness management of cardiac disorders depend on early detection and diagnosis. In spite of medical professionals efforts, Misdiagnosis and misunderstanding of test results by cardiologists and cardiovascular surgeons may occur daily. According to the World Health Organization (WHO), cardiovascular diseases (CVDs) cause 32% of all deaths around the world, which makes them a significant global health concern. As Artificial Intelligence (AI) techniques like as Machine Learning (ML) and Deep Learning (DL) have advanced, they have become essential tools for detecting and predicting CVDs. By carefully comparing a number of strong existing machine learning algorithms, this study seeks to create an ML system for the early prediction of cardiovascular illnesses. This study analyzes and validates the system's performance using statlog cardiac datasets from global platforms. A variety of machine learning techniques, such as decision trees and random forests are trained using the Cleveland dataset. To determine the best hypermetric variables that illustrate the optimal performance of the algorithms used, various evaluation methods have been applied. As a result, hyperparameter tuning methods have been utilized. cross-validation featuring a confidence interval of 95%. The results of the study focus on ML's advantage for enhancing early prediction and diagnosis of cardiovascular diseases. This study helps in the progress of ML methods within medicine by examining and comparing how well the algorithms used performed on the Cleveland and Statlog heart datasets. The ML system created provides healthcare professionals with a valuable resource for the early prediction and diagnosis of cardiovascular diseases, and it may have implications for the prediction and diagnosis of other diseases as well.**

Keywords: Cardiovascular diseases, Artificial intelligence, Machine learning, Deep learning, Prediction.

I. INTRODUCTION

According to the World Health Organization (WHO), Cardiovascular Diseases (CVDs) are now a major global health problem, responsible for 32% of all deaths worldwide and causing to 17.9 million death occurrences annually. 1In this context, techniques such as Machine Learning (ML) and Deep Learning (DL), which are subsets of Artificial Intelligence (AI), have become as significantly more valuable resources for scientists and medical experts in their work to predict and identify CVDs. Artificial Intelligence includes a wide range of concepts and definitions. The importance of it has been developing and may vary based on the area in which it is used. To put it differently, AI can be briefly described as the utilization of machines with learning capabilities similarly to those of human cognitive functions. In reality, ML is a crucial part of the AI domain. In the context of supervised learning, it involves to the utilization of trained algorithms that enable machines to independently learn, carry out tasks, and resolve equations based on known inputs and outputs from prior instances. When it comes to unsupervised learning, the outputs are not known. ML and DL are applicable in various domains, such as data science, image analysis, voice and noise processing, city traffic management, digital marketing, self-driving driving, fraud detection, handwritten recognition, among others. In the field of applied medicine, earlier studies have proved the practicality of ML and DL methods for predicting a variety of diseases. Traditional diagnostic approaches typically depend on physical examinations, patient medical histories, and a range of biological tests. Certainly, these approaches can be time-consuming and costly, and they do not guarantee an accurate diagnosis in every instance. ML has become an efficient instrument in medicine, offering a unique method for predicting and diagnosing cardiovascular diseases. ML systems can identify patterns and correlations that may not be immediately visible to healthcare professionals through visual inspection by adding massive medical data and employing advanced algorithms. This can aid in the early identification and prediction of cardiovascular diseases, resulting in enhanced patient outcomes and a decrease in the worldwide burden of these conditions.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 6, pp 23-28, June-2025

https://doi.org/10.47001/IRJIET/2025.906004

This study aims to further the development of ML techniques within medicine and to offer healthcare professionals a valuable resource for early prediction and diagnosis of cardiovascular diseases. As well, this research will reveal perspective on the possible uses of ML for predicting and diagnosing other illnesses. The main points and ideas of the present study can therefore be summarized as follows.

- To avoid fatalities among patients, doctors are very interested in accurate early predictions of heart disease risk.
- It is proposed to establish a diagnostic decision support system that addresses the problem of cardiologists making misdiagnoses and helps prevent possible misunderstandings of test results.
- The dataset of heart disease in Cleveland is examined through feature engineering methods and utilized to train and test the suggested ML models.
- A managed selection of ML algorithms is trained using the Cleveland heart disease dataset, followed by hyperparameter tuning.
- The prediction system that was developed attained a peak accuracy of 92%, overcoming similar research, and achieved an accuracy of 91.18% with the Statlog heart dataset, with performance verification via 10-fold cross-validation within a 95% confidence interval.

II. METHODOLOGY

Our suggested approach structures around creating a heart disease prediction system through a verification of ML algorithms' performance. In our study, we utilized the Heart Disease Cleveland dataset, which is commonly used in heart disease prediction research. The dataset includes multiple health-related factors that are used to predict the existence of heart disease. We started our inquiry with a thorough examining and evaluation of the Cleveland database, a well-known dataset that has been widely employed in research related to heart disease prediction. The database includes a wide collection of health-related features recognized to affect the prediction of cardiovascular disease. The main aim of our study was to identified the essential properties among these variables and categorize them as the main risk factors determining the accurate prediction of heart disease.

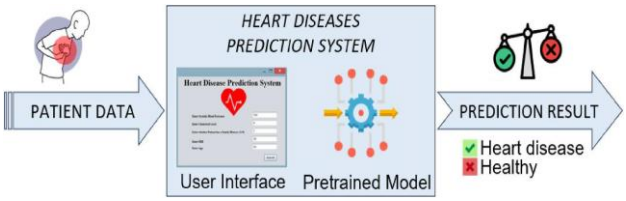

Figure 1: Layout of Proposed System

Data Processing: At first, we started our preprocessing phase by identifying duplicates and missing values, resulting in the deletion of 16 samples from the dataset of 303. The next step is to check for and address outliers, as they significantly affect the results of statistical and ML models. This process targets to identify and address data points that show significant deviation from the majority of the dataset. Outlier detection can be performed using different techniques. The suitable method varies based on the characteristics of the data, making it necessary to select the correct approach for detecting outliers. Once outliers have been managed with, checking the data types of the attributes is necessary. To avoid errors during analysis and modeling, it is essential to convert attributes to the correct data type when necessary, as incorrect data types can lead to such errors.

To make sure the dataset is free of missing or null values, all preprocessing stages are implemented. This matters, as problems can occur and results can be incorrect when building an ML model if there are missing or null values. One can verify the absence of missing or null values in the dataset by employing suitable programming tools and techniques. It is essential to carry out outlier detection after the dataset has been cleaned of any missing or null values. Outliers are data points that are considerably separate from other values in the dataset. They can greatly influence the outcomes of an ML model. If not managed properly, they can lead to overfitting or underfitting of the model with respect to the data and can introduce bias into the results. A common method for identifying outliers is to create box plots for each feature in the dataset. Box plots visually describe the distribution of a dataset and clearly indicate the presence of outliers.

Table 1

Sr. No.	Attributes	Definitions
1	Age	This feature provides information on how old a patient is, which is an important factor to consider in heart disease prediction as the risk of heart disease increases with age.
2	GD	Refers to the biological sex of a person, which can be either "Woman" or "Man".
3	CP	Chest pain is classified into four types: Typical angina, Atypical angina, Non-angina pain, and Asymptomatic.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 6, pp 23-28, June-2025

https://doi.org/10.47001/IRJIET/2025.906004

4	Restbps	The "resting blood pressure" attribute refers to the patient's blood pressure when they are in a relaxed state (mm Hg)
5	Chol	Refers to the levels of cholesterol in a patient's blood. (mg/dl)
6	Fbs	Diabetes is a binary variable that indicates whether a patient has been diagnosed with diabetes.
7	Restecg	The electrocardiographic results attribute is a categorical variable that describes the results of an electrocardiogram (ECG) test performed on a patient.
8	Thalach	The Heart rate attribute refers to the number of times the heart beats per minute.
9	Exang	The Angina attribute is a categorical variable that indicates whether a patient has angina, a type of chest pain that occurs when there is not enough blood flow to the heart.
10	Oldpeak	Refers to the ST depression induced by exercise relative to rest.
11	Slope	Refers to the slope of the peak exercise ST segment, which is a measure of the electrical activity of the heart.
12	Number of Major Vesseels (0-3)	Number of major vessels (0–3) attribute refers to the number of major blood vessels (0–3) that are visible by fluoroscopy (a type of X-ray that uses a continuous X-ray beam to produce real-time images).
13	Thal	The thallium heart scan attribute refers to the results of a thallium heart scan, which is a type of nuclear imaging test used to evaluate blood flow to the heart muscle.

Credit authorship contribution statement:

Arpita Gangadhar Awate: Writing – original draft, website, Methodology, Data collection, Formal analysis, Imagination.
Shweta Rajendra Tirpude: Writing – review & editing, Website, Resources, Data collection, Formal analysis, Imagination.
Mangla Ganpant Bhoyar: Writing – review & editing, Methodology, Data collection, Formal analysis, Imagination.
Asst. Suraj S. Bankar: Writing- review and editing, Data Collection, Methodology, Formal Analysis.

Data availability statement

The data used in this paper is publicly available and referenced in the article.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

III. CONCLUSION

As heart disease continues to be one of the top causes of death globally, it is essential to predict it accurately and early for immediate action and effective treatment. This study examined the use of machine learning methods, particularly the Random Forest algorithm, to predict heart disease based on a structured health-related dataset. The results from our experiments indicate that the Random Forest algorithm performs well at identifying patterns and relationships among various risk factors related to heart disease. Random Forest, as an ensemble learning technique, utilizes the strengths of multiple decision trees and merges their outputs to generate predictions that are more accurate and stable. It manages non-linear data effectively, lowers the risk of overfitting compared to single decision trees, and offers valuable insights into feature significance. These properties are especially valuable in medicine, where data is frequently loud, diverse, and imbalanced. With its high accuracy, precision, recall, and F1-score, our model proved reliable for predicting heart disease outcomes. In addition to its predictive performance, the Random Forest model's interpretability enables doctors to gain insight into how various features—like age, blood pressure, cholesterol levels, and chest pain type—contribute to the final prediction. This supports in diagnosis and improves the decision-making process by drawing attention to important.

Our research corroborates the effectiveness of Random Forest for predicting heart disease; however, there remain opportunities for improvement and investigation. Future studies might concentrate on the integration of more varied and larger datasets from different healthcare facilities to enhance flexibility. Model accuracy and personalization could be further improved by including real-time patient monitoring data, electronic health records (EHR), and genetic information.

ISSN (online): 2581-3048
Volume 9, Issue 6, pp 23-28, June-2025
https://doi.org/10.47001/IRJIET/2025.906004

ACKNOWLEDGMENT

Our honest gratitude goes to all who provided support and guidance throughout this research project on predicting heart disease with the Random Forest algorithm. Firstly, we owe a great thankful to our project guide Prof. Suraj Bankar, for their constant support, valuable advice, and constant motivation. We would also like to express our gratitude to all the faculty members and staff of Computer science and Engineering, Shri Sai College of Engineering and technology Bhadrawati, for offering us the appropriate environment and necessary resources for our research. We were able to successfully complete this work thanks to their direct and indirect support. We were able to successfully complete this project. We would like to extend our gratitude to the developers and administrators of the datasets we utilized, particularly the UCI Machine Learning Repository. Our ability to conduct experiments and effectively test our model was greatly helped by the presence of secure data. We would also like to express our gratitude for the support and feedback from our classmates and friends, who frequently shared valuable ideas and motivated us throughout this journey. Through discussions with them, we were able to examine the problem from various angles and develop improved solutions. This research would not have been possible without the support and backing of all these amazing persons. We truly appreciate all of you.

REFERENCES

[1] Lowlesh Yadav and Asha Ambhaikar, “IOHT based Tele-Healthcare Support System for Feasibility and performance analysis,” Journal of Electrical Systems, vol. 20, no. 3s, pp. 844–850, Apr. 2024, doi: 10.52783/jes.1382.

[2] L. Yadav and A. Ambhaikar, "Feasibility and Deployment Challenges of Data Analysis in Tele-Healthcare System," 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI), Raipur, India, 2023, pp. 1-5, doi: 10.1109/ICAIIHI57871.2023.10489389.

[3] L. Yadav and A. Ambhaikar, "Approach Towards Development of Portable Multi-Model Tele-Healthcare System," 2023 International Conference on Artificial Intelligence for Innovations in Healthcare Industries (ICAIIHI), Raipur, India, 2023, pp. 1-6, doi: 10.1109/ICAIIHI57871.2023.10489468.

[4] Lowlesh Yadav and Asha Ambhaikar, Exploring Portable Multi-Modal Telehealth Solutions: A Development Approach. International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC), vol. 11, no. 10, pp. 873–879, Mar. 2024.11(10), 873–879.

[5] Lowlesh Yadav, Predictive Acknowledgement using TRE System to reduce cost and Bandwidth, March 2019. International Journal of Research in Electronics and Computer Engineering (IJRECE), VOL. 7 ISSUE 1 (JANUARY- MARCH 2019) ISSN: 2393-9028 (PRINT) | ISSN: 2348-2281 (ONLINE).

[6] Sandhya.S. Bachar, Neehal.B. Jiwane, Ashish.B. Deharkar “Sentiment analysis of social media” DOI: 10.17148/IJARCCE.2022.111234 International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Impact Factor 7.918 Vol. 11, Issue 12, December 2022.

[7] Akshay A. Zade, Lowlesh N. Yadav, Neehal B. Jiwane. “A Review on Voice Browser” DOI: 10.17148/IJARCCE.2022.111238 International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Impact Factor 7.918 Vol. 11, Issue 12, December 2022.

[8] Omkar K. Khadke, Lowlesh N. Yadav, Neehal B. Jiwane. “Review on Challenges and Issues in Data Mining” DOI: 10.17148/IJARCCE.2022.111149 International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Impact Factor 7.918 Vol. 11, Issue 11, November 2022.

[9] Arpita Awate: Miss. Vaishali Vaidya, Mr. Vijay Rakhade, Mr. Neehal B. Jiwane. “VOICE CONTROLLED ROBOTIC CAR BY USING ARDUINO KIT” DOI: 10.17148/IJARCCE.2022.111232 International Journal of Advanced Research in Computer and Communication Engineering ISO 3297:2007 Certified Impact Factor 7.918 Vol. 11, Issue 12, December 2022.

[10] Atharv Arun Yenurkar, Asst Prof. Neehal B. Jiwane, Asst. Prof. Ashish B. Deharkar. “Effective Validation for Pervasive Computing and Mobile Computing Using MAC Algorithm”. International Journal of Research Publication and Reviews, Vol 3, no 12, pp 470-473 December 2022.

[11] Pooja Raju Katore, Asst. Prof. Ashish B. Deharkar, Asst. Prof. Neehal B. Jiwane. “Cloud Computing and Cloud Computing Technologies: A-Review”. International Journal of Research Publication and Reviews, Vol 3, no 12, pp 538-540 December 2022.

[12] Combining Vedic & Traditional Mathematic Practices for Enhancing Computational Speed in Day-To-Day Scenarios, Speed in Day-To-Day Scenarios, Conference: Industrial Engineering Journal ISSN: 0970-2555 Website: www.ivyscientific.org, At: Industrial Engineering Journal ISSN: 0970-2555 , Website: www.ivyscientific.org. (UGC JOURNAL).

ISSN (online): 2581-3048
Volume 9, Issue 6, pp 23-28, June-2025
https://doi.org/10.47001/IRJIET/2025.906004

[13] python.net, December 2022, DOI:10.17148/IJARCCE.2022.111237, Conference: International Journal of Advanced Research in Computer and Communication Engineering.

[14] A Survey for Credit Card Fraud Detection Using Machine Learning, December 2022, DOI:10.17148/IJARCCE.2022.111221, Conference: International Journal of Advanced Research in Computer and Communication Engineering. GRB 210217A: a short or a long GRB? December 2022, DOI: 10.1007/s12036-022-09822, Journal of Astrophysics and Astronomy, Published by Online ISSN: 0973-7758, Print ISSN: 0250-6335.

[15] Pronunciation Problems of English Language Learners in India, November 2022, DOI: 10.17148/IJARCCE.2022.111151, Conference: International Journal of Advanced Research in Computer and Communication Engineering.

[16] Photometric and spectroscopic analysis of the Type II SN 2020jfo with a short plateau, November 2022, DOI:10.48550/arXiv.2211.02823.

[17] Artificial Neural Network, May 2022, DOI: 10.17148/IJARCCE.2022.115196, Conference: International Journal of Advanced Research in Computer and Communication Engineering.

[18] Cloud Storage Security Based on Dynamic key Generation Technique, May 2022 DOI: 10.17148/IJARCCE.2022.115189, Conference: International Journal of Advanced Research in Computer and Communication Engineering.

[19] Research on Techniques for Resolving Big Data Issues, May 2022, DOI: 10.17148/IJARCCE.2022.115192, Conference: International Journal of Advanced Research in Computer and Communication Engineering.

[20] STUDY on INTERNET of THINGS BASED APPLICATION, May 2022, DOI: 10.17148/IJARCCE.2022.115179, Conference: International Journal of Advanced Research in Computer and Communication Engineering.

[21] Research on Data Mining, May 2022, DOI: 10.17148/IJARCCE.2022.115176, Conference: International Journal of Advanced Research in Computer and Communication Engineering.

[22] Security Solution of the Atm and Banking System, May 2022, DOI: 10.17148/IJARCCE.2022.115165, Conference: International Journal of Advanced Research in Computer and Communication Engineering.

[23] Study on Positive and Negative Effects of Social Media on Society, May 2022, DOI: 10.17148/IJARCCE.2022.115161, Conference: International Journal of Advanced Research in Computer and Communication Engineering.

[24] Research on Association Rule Mining Algorithms, May 2022, DOI: 10.17148/IJARCCE.2022.115152, Conference: International Journal of Advanced Research in Computer and Communication Engineering.

[25] Block chain Technology, May 2022, DOI: 10.17148/IJARCCE.2022.115154, Conference: International Journal of Advanced Research in Computer and Communication Engineering.

[26] INTERNET of THINGS RESEARCH CHALLANGES and FUTURE SCOPE, May 2022 DOI: 10.17148/IJARCCE.2022.115150, Conference: International Journal of Advanced Research in Computer and Communication Engineering.

[27] Data Collection and Analysis in a Smart Home Automation System, May 2022 DOI: 10.17148/IJARCCE.2022.115148, Conference: International Journal of Advanced Research in Computer and Communication Engineering.

[28] Using Encryption Algorithms in Cloud Computing for Data Security and Privacy, May 2022, DOI:10.17148/IJARCCE.2022.115149, Conference: International Journal of Advanced Research in Computer and Communication Engineering.

[29] An Efficient Way to Detect the Duplicate Data in Cloud by using TRE Mechanism, May 2022, DOI:10.17148/IJARCCE.2022.115139, Conference: International Journal of Advanced Research in Computer and Communication Engineering , Volume: 11.

International Research Journal of Innovations in Engineering and Technology (IRJIET)

ISSN (online): 2581-3048

Volume 9, Issue 6, pp 23-28, June-2025

https://doi.org/10.47001/IRJIET/2025.906004

Citation of this Article:

Arpita Gangadhar Awate, Shweta Rajendra Tirpude, Mangla Ganpat Bhoyar, & Asst. Prof. Suraj S. Bankar. (2025). Early Heart Disease Prediction Using Machine Learning Algorithm. *International Research Journal of Innovations in Engineering and Technology - IRJIET*, 9(6), 23-28. Article DOI https://doi.org/10.47001/IRJIET/2025.906004

****

