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

ISSN (online): 2581-3048

Volume 9, Issue 6, pp 281-286, June-2025

https://doi.org/10.47001/IRJIET/2025.906037

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.**

Cardiovascular diseases (CVDs) such as hypertension, heart failure, stroke, and coronary artery disease are now the major causes of early death worldwide, particularly in low and middle-income countries. Early detection of these disorders could lower the number of people who die prematurely. Researchers have proposed many techniques for CVD prediction, such as data mining, machine learning (ML), and the Internet of Things (IoT), for the early detection and monitoring of cardiac patients. Although these techniques are suggested and sometimes used, there is still much worry regarding their efficacy in situations where the error rate is high and accuracy is doubtful. As a result, it is necessary to select a prediction technique that can deliver more accuracy and fewer errors. This paper proposes an effective ensemble method based on the Random Forest (RF) algorithm for improving accuracy by combining multiple feature selection technique.

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 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 an 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 281-286, June-2025

https://doi.org/10.47001/IRJIET/2025.906037

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 variety of disease.

II. HEART DISEASE

The Heart is the most important organ of human body. If it does not function properly then it affects other organ of the body. According to a report 7,000,000 die from heart attacks each year. According to WHO report around 17.9 million people die due to CVDS in 2016. 31% of the death of people is due to Heart disease around the globe in every year. The pumping of blood to the human body is the vital function of heart which supplies oxygen and nutrients to the human body and also removes other metabolic waste from the body. If there is deficiency of blood in human body then heart doesn't function properly and it stop working which causes the death of human being. Angina occurs when there is temporary loss of blood to the heart causing chest pain. Cardiovascular disease is of two types.

(1) Heart Attack-It occurs when the heart blood vessels are suddenly blocked.

(2) Heart Failure-It results from coronary heart disease, hypertension, cardiomyopathy. Heart failure is basically when the heart is unable to maintain a strong blood flow and these results in chronic tiredness resist physical activities and shortness of breath. Heart failure can be divided into three types 1. Right side heart failure 2. Left side heart failure 3. congestive heart failure.

Right sided heart failure usually causes left sided heart failure. In right sided heart failure blood backs up into other tissues such as liver and in the abdomen causing congestion in these areas. As a result of right sided heart failure, we can have Hepatomegaly and Anciles.

In left sided heart failure oxygenated blood cannot be pumped out from heart to the rest of the body. So, blood can back flow. Blood can accumulate in lung veins causing fluid accumulation in lungs causes shortness of breath and oedema.

Table 1: Major cause of heart disease [10]

Disease Type
Smoking
High Blood Pressure
High Cholesterol
Diabetes and Prediabetes
Being overweight
Physical inactivity

Risk factor that cannot control for heart disease:

1) Family history
2) 55 years or older
3) History of preeclampsia

Symptoms of Heart attack

a) Nausea
b) Dizziness
c) Jaw pain
d) Abdominal pain

Living a healthy life style can reduce the effect of heart disease. Drinking plenty of water, eating green vegetables, fat free food, doing exercises, regular check-up of heart, consulting with the doctor if there any family history of heart disease can reduce the effect of heart disease Metabolicsyndrome.

III. 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

International Research Journal of Innovations in Engineering and Technology (IRJIET)
ISSN (online): 2581-3048
Volume 9, Issue 6, pp 281-286, June-2025
https://doi.org/10.47001/IRJIET/2025.906037

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.

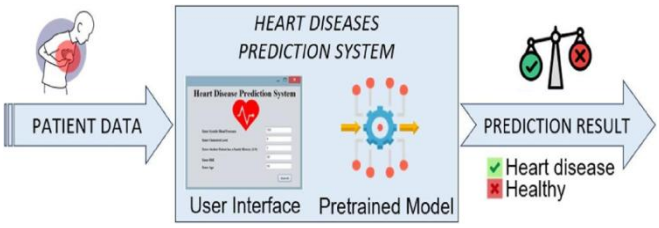

Figure 1: Layout of Proposed System

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 2: Features for data prediction

Atrribute	Meaning
Age1	Age is continuous
Gender1	1=male 0=female
Cp1	Chest pain
Trestbps	Resting blood pressure results during hospitalised: continuous(mmHg)
Chol	Cholesterol level in mg/dl
Fbs1	Fasting blood sugar 0:<=120mg/dl, 1:>120mg/dl
Restecg	Electrocardiographic results during resting 1=true 0=false
Thalach	Maximum heart rate achieved: continuous
exang	Exercise induced angina
oldpeak	ST depression
slope	ST segment slope
Ca	Number of major vessels coloured by fluoroscopy: discrete (0,1,2,3)
thal	3:normal 6:fixed defect 7:reversible defect

ISSN (online): 2581-3048
Volume 9, Issue 6, pp 281-286, June-2025
https://doi.org/10.47001/IRJIET/2025.906037

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, Methhodology, 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.

IV. 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 diseases. Each of the above-mentioned algorithms has performed extremely well in some cases but poorly in some other cases.

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 would like to express special thanks of gratitude to our guide Prof. Suraj Bankar, who gave us the golden opportunity to do this wonderful project on the topic Heart Disease Prediction Using Machine Learning With the Random Forest Algorithm, which also helped us in doing a lot of research and we came to know about so many new things. We would also like to thank our H.O.D. Jiwne Sir for providing us the opportunity to implement our project. We are thankful to both of them. Finally, we would also like to thank our department staff members and our parents & friends who helped us a lot in finalizing this project within the limited time frame.

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.

ISSN (online): 2581-3048
Volume 9, Issue 6, pp 281-286, June-2025
https://doi.org/10.47001/IRJIET/2025.906037

[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] 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).

[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.*

[15] 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.

[16] 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.*

[17] Photometric and spectroscopic analysis of the Type II SN 2020jfo with a short plateau, November 2022, DOI:10.48550/arXiv.2211.02823, License CC BY 4.0.

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

[19] 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.*

[20] 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.*

[21] STUDY on INTERNET of THINGS BASED APPLICATION, May 2022, DOI: 10.17148/IJARCCE.2022.115179, *Conference:*

ISSN (online): 2581-3048
Volume 9, Issue 6, pp 281-286, June-2025
https://doi.org/10.47001/IRJIET/2025.906037

International Journal of Advanced Research in Computer and Communication Engineering.

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

[23] 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.*

[24] 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.*

[25] 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.*

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

[27] 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.*

[28] 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.*

[29] 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.*

[30] 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.

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), 281-286. Article DOI https://doi.org/10.47001/IRJIET/2025.906037

****

