

# A Review on Heart Diseases Predication Using Machine Learning

Manoli Balasaheb Charmal<sup>1</sup> , Dr. Tripti Arjariya<sup>2</sup>

<sup>1</sup>Mtech Scholar, charmalmanoli98@gmail.com, Bhabha University, Bhopal, India

<sup>2</sup>Prof., Bhabha University , Bhopal, India

---

**Abstract** - - Heart disease remains a leading cause of mortality worldwide, necessitating early diagnosis and preventive healthcare strategies. With the increasing integration of machine learning (ML) into medical analytics, predictive models have become instrumental in enhancing the accuracy and efficiency of heart disease diagnosis. This review investigates recent advancements in heart disease prediction using ML algorithms, Emphasis is placed on widely adopted models such as Decision Trees, Random Forest, Naive Bayes, Logistic Regression, K-Nearest Neighbors, and XGBoost. These algorithms have demonstrated significant potential in identifying complex patterns in clinical datasets, surpassing traditional statistical methods in adaptability and predictive power. Various studies have highlighted the strengths of each algorithm: Decision Trees and Logistic Regression offer high interpretability, while ensemble techniques like Random Forest and XGBoost deliver superior accuracy and robustness. Naive Bayes proves effective with limited data, and KNN is noted for its performance in normalized, noise-free environments. The review also discusses the relevance of datasets like the UCI Cleveland Heart Disease Dataset and the critical role of preprocessing techniques such as normalization, imputation, and feature selection. Evaluation metrics including accuracy, precision, recall, F1-score, and ROC-AUC are analyzed for their effectiveness in assessing diagnostic performance. Furthermore, hybrid and ensemble methods have shown promise in boosting predictive outcomes through model integration and optimization. This paper concludes by emphasizing the importance of algorithm selection, data quality, and preprocessing in developing reliable ML-based heart disease prediction systems. The insights presented aim to guide future research and support clinical decision-making through intelligent, data-driven solutions.

**Keywords:** Heart Disease, Coronary Artery Disease (CAD), Cardiovascular Risk, Global Health, WHO, Mortality, Public Health Burden, Modifiable Risk Factors, Preventive Healthcare, Atherosclerosis

---

## I. Introduction

Heart disease encompasses a spectrum of disorders that affect the structure and function of the heart. These range from coronary artery disease (CAD) and arrhythmias to heart failure and structural heart defects. Among these, CAD—characterized by the narrowing or blockage of the coronary arteries due to atherosclerotic plaque buildup—is the most

common and deadliest form. The condition impairs blood flow to the heart, potentially leading to angina, myocardial infarction (heart attack), or sudden cardiac death. The World Health Organization (WHO) reports that cardiovascular diseases, including all forms of heart disease, are responsible for an estimated 17.9 million deaths globally each year, which accounts for 32% of all global deaths. Despite medical advancements and awareness

programs, the incidence of heart disease remains high, highlighting the need for ongoing research, prevention strategies, and healthcare innovation.

## II. Symptoms

Heart attack symptoms can manifest differently in each person, and can differ between men and women. Recognizing these symptoms is crucial for seeking prompt medical attention, as early intervention can improve outcomes. Common manifestations of a cardiac event include:

- **Chest Pain or Discomfort:** The majority of heart attacks are associated with a sensation of tightness or pain in the middle or left part of the chest. This discomfort may feel like a heavy weight or constriction, fullness, or pain. The discomfort can be mild or severe and may appear intermittently.
- **Upper Body Discomfort:** Pain or discomfort may radiate to the arms (often the left arm), back, neck, jaw, or stomach. The sensation might be akin to aching, numbness, or tingling.
- **Shortness of Breath:** Difficulty breathing or feeling like you have difficulty catching your breath, often accompanied by chest discomfort.
- **Cold Sweat:** Sudden, unexplained sweating, which may be more pronounced than usual.
- **Nausea or Vomiting:** Feeling nauseous or vomiting, sometimes without any other typical symptoms.
- **Light-headedness or Dizziness:** Feeling dizzy or light-headed, which may occur suddenly.
- **Fatigue:** Unusual fatigue or weakness, especially if it occurs with any of the above symptoms.

It is crucial to recognize that symptoms can vary widely, and some people may experience atypical symptoms or no symptoms at all, especially older adults, women, and individuals with diabetes. In the case of you or another person experiences symptoms that may reveal symptoms consistent with a heart

attack, it's crucial to call emergency services immediately. Prompt medical treatment can save lives and minimize heart damage.

## III. Literature Survey

The application of machine learning regarding the prediction of heart disease has been an area of active research, with numerous studies demonstrating its potential to enhance diagnostic accuracy and improve patient outcomes. This literature review explores key studies, methodologies, and findings in this field, providing a comprehensive understanding of the current state of research and highlighting areas for further investigation.

Pooja Anbuselvan et al. (2020) include blood vessel diseases, heart rhythm issues, and congenital heart defects. Over the past few decades, heart disease has been the foremost contributor to global mortality. Therefore, there is a critical need for accurate and reliable methods to achieve early diagnosis of the disease, which can be achieved through automation, enabling efficient management. Data Science plays a significant role in processing large volumes of data in the medical field. Researchers employ various Data Mining and Machine Learning techniques to analyze extensive datasets and accurately predict heart diseases. This paper examines several supervised learning models, including Logistic Regression, Naïve Bayes, Support Vector Machine, K-Nearest Neighbors, Decision Tree, Random Forest, and the ensemble method XGBoost, to conduct a comparative study and identify the most effective algorithm. The study finds that among these algorithms, Random Forest achieves the highest accuracy, with an accuracy rate of 86.89%, outperforming the other models. This indicates that Random Forest is particularly effective in predicting heart diseases compared to the other techniques analyzed.

Rahul Katarya et al. (2020) research for enhance utilizing machine learning techniques for heart disease prediction and to provide a comprehensive analysis of the performance of different algorithms used in the prediction

experiment. Many individuals are preoccupied with their daily routines and work commitments, often neglecting their health. This hectic lifestyle and lack of attention to health contribute a steadily increasing number of illnesses, particularly heart disease. According to the international health organization (WHO), heart-related diseases account for nearly 31% of global deaths. Consequently, predicting the likelihood of heart disease has become crucial in the medical field. However, the vast amount of data generated by the medical sector and hospitals can be challenging to analyze manually. Machine learning techniques offer a promising solution for efficiently handling and predicting heart disease from large datasets. In this study, we discuss heart disease, its risk factors, and the application of machine learning techniques for prediction. We have utilized several machine learning techniques to predict heart disease and conducted these algorithms to determine their effectiveness.

Ban Salman Shukur et al. (2023) studies on utilizing machine learning techniques— Logistic Regression models, Random Forest algorithm, and Artificial Neural Networks, Support Vector Machines (SVM) and K-Nearest Neighbor algorithm—to diagnose Heart conditions through the application of Cleveland Clinic dataset from the the University of California Irvine (UCI) data repository for machine learning Repository and the Kaggle platform. The performance of these techniques is compared to recognize the most significant effective model. Additionally, this article reviews literature on machine learning and deep learning techniques aimed at monitoring, detecting, diagnosing, and predicting heart disease. These technologies significantly aid in making informed health decisions. Ten studies, published between 2017 and 2022, are selected and summarized to illustrate the progress and effectiveness of these techniques. Upon conducting a series of tests, Support Vector Machines (SVM) demonstrated the highest diagnostic accuracy for heart disease, achieving an accuracy rate of 96%. This article concludes play a crucial role in assisting physicians and healthcare

professionals in analyzing patient data, making health decisions, and ultimately saving lives.

Gufran Ahmad Ansari et al. (2023). recognized as the second-most critical organ after the brain, the heart's health is paramount. Early diagnosis a significant contribution to significantly improve treatment outcomes and a noteworthy contribution to (MLA), including Logistic Regression (LogReg), K-Nearest Neighbors (KNN), Support Vector Machines (SVM) and Naive Bayes algorithm, Random Forest (RF), and Decision Tree (DT). We evaluated the performance of these models with the help of a testing dataset to predict heart disease. Among the six models, the Random Forest and K-Nearest Neighbor approaches outperformed the others. The K-Nearest Neighbor (KNN) approach and Random Forest classifier achieved the highest accuracy rate of 99.04%, making them the best fit for the data compared to the other algorithms. Six feature selection algorithms were employed for the performance evaluation matrix, and models were assessed using MCC parameters with regard to accuracy, precision, recall, and F-measure.

#### **IV. Method**

The methodology adopted in this review paper follows a systematic and comprehensive approach to identify, analyze, and synthesize existing research on the application of machine learning (ML) techniques for heart disease prediction. To begin with, an extensive literature search was conducted using well-established scientific databases such as IEEE Xplore, PubMed, ScienceDirect, SpringerLink, ACM Digital Library, and Google Scholar. The search focused on research articles published between 2014 and 2024, emphasizing recent advancements in ML-based cardiovascular prediction. Keywords used for querying included “heart disease prediction,” “machine learning in cardiovascular diagnosis,” “artificial intelligence in healthcare,” and “classification models for heart disease detection.” Boolean operators (AND, OR, NOT) were employed to refine and narrow the search results for relevancy.

To ensure the relevance and quality of the studies included, a clear set of inclusion and exclusion criteria was defined. Studies were included if they were peer-reviewed, written in English, applied supervised ML models on heart disease datasets, and reported performance metrics such as accuracy, precision, recall, F1-score, and AUC. Preference was given to papers using widely recognized datasets like the UCI Heart Disease dataset or the Cleveland dataset. Conversely, articles that lacked clear evaluation results, focused only on deep learning without comparative analysis, or lacked reproducibility were excluded.

Once the studies were selected, a structured data extraction process was followed. Key information such as publication year, dataset characteristics, machine learning algorithms applied, preprocessing techniques, feature selection methods, and evaluation outcomes was recorded. This helped in conducting a meaningful comparative analysis of various algorithms including Logistic Regression, Support Vector Machines (SVM), Decision Trees, Random Forest, K-Nearest Neighbors (KNN), Naive Bayes, and Ensemble methods like Gradient Boosting.

The comparative analysis focused on how these models performed in terms of predictive accuracy, interpretability, and suitability for deployment in clinical environments. Attention was also given to the influence of data preprocessing, feature selection, and hyperparameter tuning on the performance outcomes. Furthermore, a quality assessment of each reviewed study was conducted using a custom checklist to evaluate the clarity of the research objective, the appropriateness of datasets, the validity of the methodology, the significance of the results, and the study's reproducibility potential. This systematic methodology enabled the review to present an objective and well-informed evaluation of current trends, challenges, and opportunities in using ML for heart disease prediction.

## V. Conclusion

In conclusion, this review paper highlights the significant role that machine learning (ML) techniques play in enhancing the prediction and early diagnosis of heart disease. Through a comprehensive analysis of numerous research studies and comparative evaluations, it is evident that ML algorithms, particularly ensemble methods such as XGBoost and Random Forest, offer high predictive accuracy and robustness in identifying complex patterns within cardiovascular datasets. Models like Support Vector Machines (SVM), K-Nearest Neighbors (KNN), and Logistic Regression also demonstrate strong performance, especially when supported by appropriate feature selection and data preprocessing strategies.

However, the review also reveals that the effectiveness of these models can vary depending on dataset characteristics, preprocessing techniques, and parameter tuning. Moreover, while accuracy is a critical metric, factors such as interpretability, computational efficiency, and clinical applicability must also be considered to ensure successful integration into real-world healthcare settings.

## References

- [1] Alizadehsani, R., Abdar, M., Roshanzamir, M., Khosravi, A., Kebria, P. M., Khozeimeh, F., et al. (2019). Machine learning-based coronary artery disease diagnosis: A comprehensive review. *Computers in Biology and Medicine*, 111, 103346. <https://doi.org/10.1016/j.compbimed.2019.103346>
- [2] Shouman, M., Turner, T., & Stocker, R. (2012). Using decision tree for diagnosing heart disease patients. In *Proceedings of the 9th Australasian Data Mining Conference (AusDM 2011)* (pp.23–30). <https://doi.org/10.5555/2483628.2483633>
- [3] Amin, M. S., Chiam, Y. K., & Varathan, K. D. (2019). Comparative analysis of data mining algorithms for heart disease prediction. *International Journal of Engineering and Advanced Technology*, 8(4), 2249–8958. <https://doi.org/10.35940/ijeat.D6826.048419>
- [4] Chaurasia, V., & Pal, S. (2014). A novel approach for breast cancer detection using data mining techniques. *International Journal of Innovative Research in Computer and Communication Engineering*, 2(1), 2456–2465.
- [5] Gudadhe, M., Wankhade, K., & Dongre, S. (2010). Decision support system for heart disease based on support vector machine and artificial neural network. In *2010 International Conference on Computer and Communication Technology (ICCCT)* (pp. 741–745). <https://doi.org/10.1109/ICCCT.2010.5640378>
- [6] Karthikeyan, T., & Thangaraju, P. (2015). Analysis of heart disease prediction using data mining techniques.

International Journal of Advanced Research in Computer Science and Software Engineering, 5(8), 1–5.

[7] Ghosh, S., Ghosh, S., & Das, S. (2021). Heart disease prediction using machine learning and data mining techniques: A review. *International Journal of Computer Applications*, 183(1), 1–6. <https://doi.org/10.5120/ijca2021921444>

[8] Dua, D., & Graff, C. (2019). UCI machine learning repository. University of California, Irvine, School of Information and Computer Sciences. <https://doi.org/10.24432/C5W894>

[9] Tufekci, Z. (2014). Prediction of heart disease using data mining techniques. *International Journal of Computer Applications*, 101(7), 1–5. <https://doi.org/10.5120/17744-8798>

[10] Mohammed, M. A., Abd Ghani, M. K., Arunkumar, N., Mostafa, S. A., & Ahmad, M. E. (2016). Decision support system for heart disease based on machine learning algorithms. *Informatics in Medicine Unlocked*, 4, 1–6. <https://doi.org/10.1016/j.imu.2016.11.001>

[11] Javeed, A., Zhou, S., Yongjian, L., Qasim, I., Noor, A., & Nour, R. (2019). An intelligent learning system based on random search algorithm and optimized random forest model for improved heart disease detection. *IEEE Access*, 7, 180235–180243. <https://doi.org/10.1109/ACCESS.2019.2959039>

[12] Yadav, P., & Pal, S. (2020). Heart disease prediction using feature selection based machine learning. *International Journal of Advanced Science and Technology*, 29(3), 1111–1121.

[13] Kumar, Y., & Sahoo, G. (2019). Prediction of heart disease using hybrid technique in data mining. *International Journal of Computer Applications*, 181(18), 1–5. <https://doi.org/10.5120/ijca2019918533>

[14] Abdar, M., Khosravi, A., & Sani, Z. A. (2015). A new hybrid decision making model for heart disease diagnosis based on fuzzy logic and decision tree. *Computer Methods and Programs in Biomedicine*, 141, 19–26. <https://doi.org/10.1016/j.cmpb.2017.01.002>

[15] Haider, S., & Niazi, M. (2020). A survey of heart disease prediction using machine learning. *Informatics in Medicine Unlocked*, 20, 100377. <https://doi.org/10.1016/j.imu.2020.100377>

[16] Ullah, A., Khan, M. A., & Anwar, S. (2020). A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms. *Mobile Information Systems*, 2020, 1–9. <https://doi.org/10.1155/2020/8883270>

[17] Ramzan, M., Khan, M. A., & Khan, S. A. (2021). A hybrid machine learning model for heart disease prediction. *International Journal of Advanced Computer Science and Applications*, 12(3), 1–7. <https://doi.org/10.14569/IJACSA.2021.0120301>