International Journal of Engineering Applied Science and Management
ISSN (Online): 2582-6948
Vol. 5 Issue 8, August 2024

# Heart Disease Prediction with Machine Learning

Pooja Ballodia[1], Mr. Ashish Suryavanshi[2]
Research Scholar, Department of Electronics and Computer Science Engineering ,School of Engineering & Technology , Vikram University Ujjain[1]
Assistant Professor, Department of Electronics and Computer Science Engineering, School of Engineering & Technology, Vikram University Ujjain[2]

***Abstract**: This paper presents a cardiovascular disease detection model developed using three machine learning classification techniques: Logistic Regression, Random Forest Classifier, and K-Nearest Neighbors (KNN). The project predicts individuals with cardiovascular disease by extracting medical history data from a dataset, including attributes such as chest pain, sugar level, and blood pressure. The Heart Disease Detection System assists patients based on their clinical information and history of heart disease. The model achieves an accuracy of 87.5%, illustrating that incorporating more training data can enhance the model's predictive accuracy. Utilizing computer-aided techniques allows for quicker and more cost-effective patient predictions, surpassing traditional methods and benefiting both patients and doctors. Our project improves heart disease prediction by cleaning the dataset and applying Logistic Regression and KNN, achieving an average accuracy of 87.5%, which is higher than previous models with an accuracy of 85%. Among the three algorithms, KNN exhibited the highest accuracy at 88.52%. Analysis of the dataset revealed that 44% of individuals are suffering from heart disease. This project demonstrates the potential of machine learning techniques to improve the accuracy and efficiency of heart disease diagnosis.*

***Keywords:** Cardiovascular disease detection, Machine learning, Logistic Regression, K-Nearest Neighbors (KNN), Random Forest Classifier*

## 1. Introduction

Machine learning enables the manipulation and extraction of implicit, previously unknown, and potentially useful information from data. It is a vast and rapidly growing field, incorporating various classifiers from supervised, unsupervised, and ensemble learning. These classifiers are used to predict outcomes and assess the accuracy of datasets. In the context of our Heart Disease Prediction System (HDPS), this knowledge is invaluable. Cardiovascular diseases, affecting millions worldwide, are the leading cause of death in adults. The World Health Organization estimates 17.9 million global deaths annually from cardiovascular diseases. Our project aims to predict individuals at risk of heart disease based on their medical history, helping to diagnose and treat patients more effectively with fewer medical tests.

This project employs three data mining techniques: Logistic Regression, K-Nearest Neighbors (KNN), and Random Forest Classifier. Our system achieved an accuracy of 87.5%, outperforming previous systems that used only one technique. Logistic regression, a supervised learning method, deals with discrete values, while KNN and Random Forest Classifier enhance prediction accuracy. The objective is to determine if a patient is likely to develop cardiovascular disease based on attributes like gender, age, chest pain, and fasting blood sugar levels. Using a dataset from the UCI repository containing patients' medical histories, we employ 14 attributes to predict heart disease risk. Training these attributes under the three algorithms, we found KNN to be the most efficient, with an accuracy of 88.52%. Ultimately, this

cost-effective method classifies patients at risk, aiding in early diagnosis and treatment.

## 2. Related Work

Significant work related to the diagnosis of cardiovascular heart disease using machine learning algorithms has motivated this study. This paper includes a brief literature survey and presents an efficient cardiovascular disease prediction system using various algorithms, including Logistic Regression, K-Nearest Neighbors (KNN), and Random Forest Classifier. The results demonstrate that each algorithm has its strengths in meeting defined objectives. The model incorporating the Intelligent Heart Disease Prediction System (IHDPS) could calculate decision boundaries using both traditional and contemporary machine learning and deep learning models. It emphasized crucial factors such as family history of heart disease. However, the accuracy of the IHDPS model was lower compared to newer models, such as those detecting coronary heart disease using artificial neural networks and advanced machine and deep learning algorithms.

McPherson et al. identified risk factors for coronary heart disease or atherosclerosis using neural network techniques, accurately predicting whether a test patient had the disease. Additionally, R. Subramanian et al. introduced a method for diagnosing and predicting heart disease and blood pressure using neural networks. They built a deep neural network incorporating disease-related attributes, with an output perceptron and nearly 120 hidden layers, ensuring accurate results for test datasets. The supervised network is recommended for heart disease diagnosis. During model testing with unfamiliar data, the trained model used previously learned data to predict results, thereby calculating the model's accuracy.

## 3. Data Source

An organized dataset of individuals, considering their history of heart problems and other medical conditions, was selected for this study. Heart diseases encompass various conditions that affect the heart. According to the World Health Organization (WHO), cardiovascular diseases are the leading cause of death among middle-aged people. The dataset used includes medical histories of 304 patients of different age groups, providing crucial medical attributes such as age, resting blood pressure, and fasting sugar levels. These attributes help determine if a patient is diagnosed with heart disease.

This dataset, obtained from the UCI repository, contains 13 medical attributes for 304 patients, aiding in the classification of patients at risk of heart disease. The dataset is divided into training and testing sets, containing 303 rows and 14 columns, with each row representing a single record. The patterns leading to heart disease detection are extracted from these records, facilitating accurate predictions and classifications of patients at risk. All attributes are listed in 'Table 1'.

**Table 1 Various Attributes used are listed**

| S. No | Observation | Description | Values |
|---|---|---|---|
| 1 | Age | Age in Years | Continuous |
| 2 | Sex | Sex of Subject | Male/Female |
| 3 | CP | Chest Pain | Four Types |
| 4 | Trestbps | Resting Blood Pressure | Continuous |
| 5 | Chol | Serum Cholesterol | Continuous |
| 6 | FBS | Fasting Blood Sugar <,or> 120 mg/dl | Yes/No |
| 7 | Restecg | Resting Electrocardiograph | |
| 8 | Thalach | Maximum Heart Rate Achieved | Continuous |
| 9 | Exang | Exercise Induced Angina | Yes/No |
| 10 | Oldpeak | ST Depression when Workout compared to the Amount of Rest Taken | Continuous |
| 11 | Slope | Slope of Peak Exercise ST segment | up/Flat/Down |
| 12 | Ca | Gives the number of Major Vessels Coloured by Fluoroscopy | 0-3 |
| 13 | Thal | Defect Type | Reversible/Fixed/Normal |
| 14 | Num | Heart Disease (Not Present/Present in the Four Major types) | Five Values |

## 4. Methodology

This paper analyzes various machine learning algorithms, including K-Nearest Neighbors (KNN), Logistic Regression, and Random Forest Classifiers, to aid practitioners and medical analysts in accurately diagnosing heart disease. The research involves reviewing recent journals, published papers, and data on cardiovascular diseases. The methodology provides a framework for the proposed model, outlining steps that transform raw data into recognized patterns for user insights. The proposed

International Journal of Engineering Applied Science and Management
ISSN (Online): 2582-6948
Vol. 5 Issue 8, August 2024

methodology (Figure 1) begins with data collection, followed by extracting significant values. The third stage, data preprocessing, involves handling missing values, data cleaning, and normalization based on the algorithms used. After preprocessing, classifiers (KNN, Logistic Regression, and Random Forest) are employed to categorize the data. The model is then evaluated for accuracy and performance using various metrics.

An effective Heart Disease Prediction System (EHDPS) has been developed using these classifiers, utilizing 13 medical parameters such as chest pain, fasting blood sugar, blood pressure, cholesterol, age, and sex for prediction.

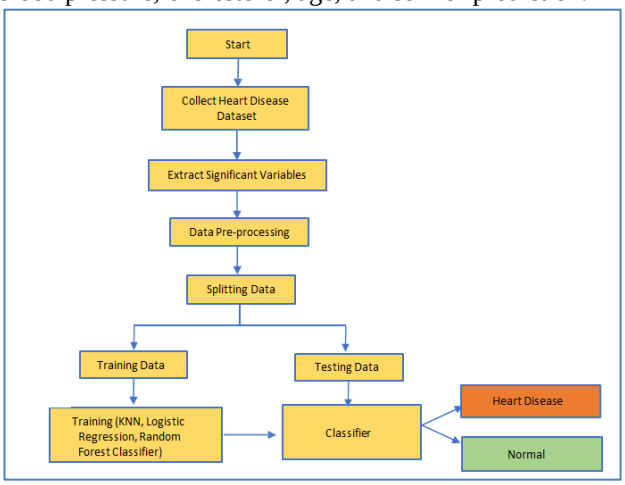


Figure 1 Proposed Model

## 5. Results & Discussions

The results indicate that while various researchers use different algorithms such as SVC and Decision Tree for detecting heart disease, our use of K-Nearest Neighbors (KNN), Random Forest Classifier, and Logistic Regression yields better results. These algorithms are more accurate, cost-efficient, and faster compared to those used in previous studies. Notably, KNN and Logistic Regression achieve a maximum accuracy of 88.5%, which is equal to or greater than the accuracies reported in earlier research. The improvement in accuracy can be attributed to the increased number of medical attributes used from the dataset. Our findings also demonstrate that Logistic Regression and KNN outperform the Random Forest Classifier in predicting heart disease. This highlights the superiority of KNN and Logistic Regression for heart disease diagnosis.

Figures 2, 3, 4, and 5 illustrate the classification of patients based on age group, resting blood pressure, sex, and chest pain, as predicted by the classifiers. These visualizations underscore the effectiveness of our chosen algorithms in segregating and predicting patients diagnosed with heart disease.

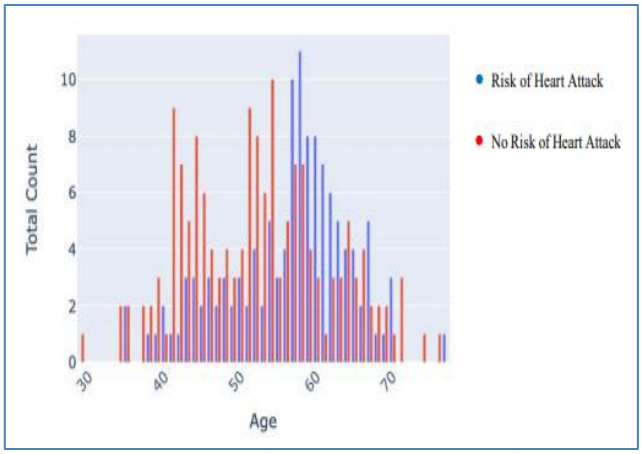


Figure 2 Shows the Risk of Heart Attack on the basis of their age

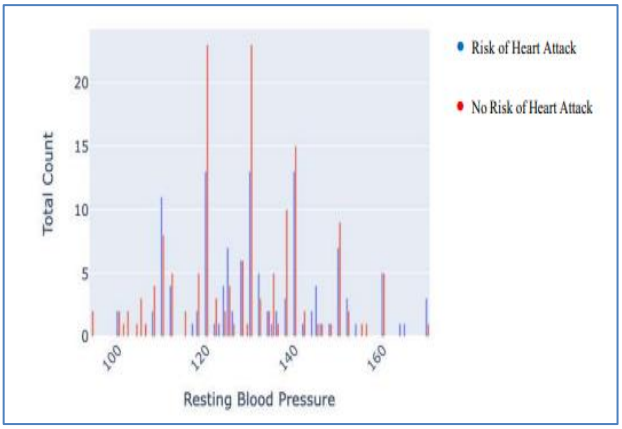


Figure 3 Shows the Risk of Heart Attack on the basis of their Resting Blood Pressure

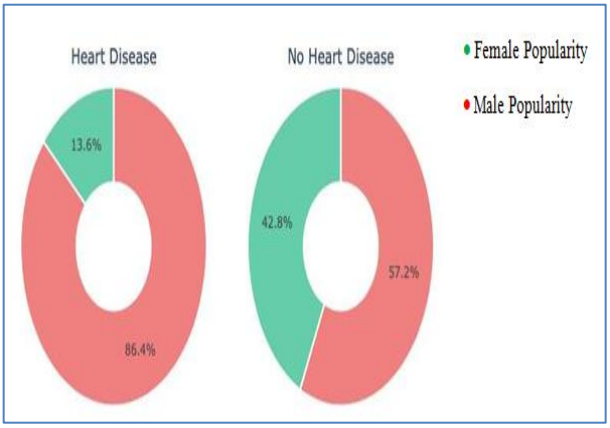


Figure 4 Shows the patients having or not having Heart Disease on the basis of Sex

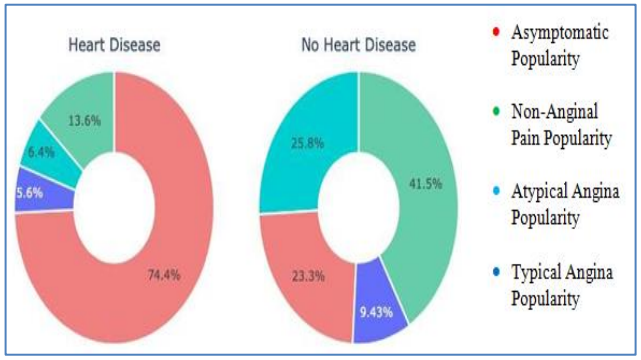


Figure 5 Shows the patients having or not having Heart Disease on the basis of type of Chest Pain

## 6. Conclusion

In conclusion, a cardiovascular disease detection model has been developed using three machine learning classification techniques: Logistic Regression, Random Forest Classifier, and K-Nearest Neighbors (KNN). This project predicts individuals with cardiovascular disease by extracting medical history data from a dataset, including attributes such as chest pain, sugar level, and blood pressure. The Heart Disease Detection System assists patients based on their clinical information and history of heart disease.

The model achieves an accuracy of 87.5%, demonstrating that using more training data can improve the model's ability to accurately predict heart disease. The use of computer-aided techniques allows for faster and more cost-effective patient predictions, outperforming traditional methods and aiding both patients and doctors. Our project enhances heart disease prediction by cleaning the dataset and applying Logistic Regression and KNN, achieving an average accuracy of 87.5%, which is higher than previous models with an accuracy of 85%. Among the three algorithms used, KNN exhibited the highest accuracy at 88.52%. Additionally, analysis of the dataset revealed that 44% of individuals are suffering from heart disease. This project demonstrates the potential of machine learning techniques to improve the accuracy and efficiency of heart disease diagnosis.

## References

[1] V. Ramalingam, A. Dandapath, and M. K. Raja, "Heart disease prediction using machine learning techniques: a survey," International Journal of Engineering Technology, vol. 7, no. 2.8, pp. 684–687, 2018.

[2] Rajdhan, A. Agarwal, M. Sai, D. Ravi, and P. Ghuli, "Heart disease prediction using machine learning," International Journal of Research and Technology, vol. 9, no. 04, pp. 659–662, 2020.

[3] J. Patel, D. TejalUpadhyay, and S. Patel, "Heart disease prediction using machine learning and data mining technique," Heart Disease, vol. 7, no. 1, pp. 129–137, 2015.

[4] D. Shah, S. Patel, and S. K. Bharti, "Heart disease prediction using machine learning techniques," SN Computer Science, vol. 1, no. 6, pp. 1–6, 2020.

[5] Y. Khourdifi and M. Bahaj, "Heart disease prediction and classification using machine learning algorithms optimized by particle swarm optimization and ant colony optimization," International Journal of Intelligent Engineering and Systems, vol. 12, no. 1, pp. 242–252, 2019.

[6] U. Haq, J. P. Li, M. H. Memon, S. Nazir, and R. Sun, "A hybrid intelligent system framework for the prediction of heart disease using machine learning algorithms," Mobile Information Systems, vol. 2018, 2018.

[7] R. Katarya and S. K. Meena, "Machine learning techniques for heart disease prediction: a comparative study and analysis," Health and Technology, vol. 11, no. 1, pp. 87–97, 2021.

[8] K. Budholiya, S. K. Shrivastava, and V. Sharma, "An optimized xgboost based diagnostic system for effective prediction of heart disease," Journal of King Saud University-Computer and Information Sciences, 2020.

[9] Naiyer, Vaseem, Jitendra Sheetlani, and Harsh Pratap Singh. "Software Quality Prediction Using Machine Learning Application." Smart Intelligent Computing and Applications: Proceedings of the Third International Conference on Smart Computing and Informatics, Volume 2. Springer Singapore, 2020.

[10] Pasha, Shaik Imran, and Harsh Pratap Singh. "A Novel Model Proposal Using Association Rule Based Data Mining Techniques for Indian Stock Market Analysis." Annals of the Romanian Society for Cell Biology (2021): 9394-9399.

[11] Md, Abdul Rasool, Harsh Pratap Singh, and K. Nagi Reddy. "Data Mining Approaches to Identify Spontaneous Homeopathic Syndrome Treatment." Annals of the Romanian Society for Cell Biology (2021): 3275-3286.

[12] Vijay Vasanth, A., et al. "Context-aware spectrum sharing and allocation for multiuser-based 5G cellular networks." Wireless Communications and Mobile Computing 2022 (2022).

[13] Singh, Harsh Pratap, and Rashmi Singh. "Exposure and Avoidance Mechanism Of Black Hole And Jamming Attack In Mobile Ad Hoc Network." International Journal of Computer Science, Engineering and Information Technology 7.1 (2017): 14-22.

[14] Singh, Harsh Pratap, et al. "Design and Implementation of an Algorithm for Mitigating the Congestion in

Mobile Ad Hoc Network." International Journal on Emerging Technologies 10.3 (2019): 472-479.
[15] Singh, Harsh Pratap, et al. "Congestion Control in Mobile Ad Hoc Network: A Literature Survey."
[16] Rashmi et al.. "Exposure and Avoidance Mechanism Of Black Hole And Jamming Attack In Mobile Ad Hoc Network." International Journal of Computer Science, Engineering and Information Technology 7.1 (2017): 14-22.
[17] Sharma et al., "Guard against cooperative black hole attack in Mobile Ad-Hoc Network." Harsh Pratap Singh et al./International Journal of Engineering Science and Technology (IJEST) (2011).
[18] Singh, et al., "A mechanism for discovery and prevention of coopeartive black hole attack in mobile ad hoc network using AODV protocol." 2014 International Conference on Electronics and Communication Systems (ICECS). IEEE, 2014.
[19] Harsh et al., "Design and Implementation of an Algorithm for Mitigating the Congestion in Mobile Ad Hoc Network." International Journal on Emerging Technologies 10.3 (2019): 472-479.
[20] M. Shouman, T. Turner, and R. Stocker, "Using decision tree for diagnosing heart disease patients," in Proceedings of the Ninth Australasian Data Mining Conference-Volume 121, 2011, pp. 23–30.
[21] S. A. Pattekari and A. Parveen, "Prediction system for heart disease using naïve bayes," International Journal of Advanced Computer and Mathematical Sciences, vol. 3, no. 3, pp. 290–294, 2012.
[22] M. A. Jabbar, B. L. Deekshatulu, and P. Chandra, "Prediction of heart disease using random forest and feature subset selection," in Innovations in bio-inspired computing and applications. Springer, 2016, pp. 187–196.
[23] S. Asadi, S. Roshan, and M. W. Kattan, "Random forestswarm optimization-based for heart diseases diagnosis," Journal of Biomedical Informatics, vol. 115, p. 103690, 2021.