# Detection of Cardiac Abnormalities Using Machine Learning

Ms. Aishwarya R. Dhote[1], Mr. S. S. Mhaske[2], Dr C. M. Jadhao[3]
[1]*PG Student, Department of Electronics and Telecommunication Engineering, MGI-COET, Shegaon*
[2]*HOD, Department of Electronics and Telecommunication Engineering, MGI-COET, Shegaon*
[3]*Principal & Professor, Mauli Group of Institutions, College of Engineering and Technology, Shegaon, Maharashtra, India*

**Abstract: Cardiovascular diseases have become the leading cause of death worldwide over the last few decades, affecting both developed and developing countries. Early detection of cardiac diseases, along with continuous supervision by clinicians, can significantly reduce the mortality rate. However, accurate detection of heart diseases in all cases and 24-hour patient consultations by doctors are often not feasible due to the demands of time and expertise. The medical field is witnessing notable advancements through machine learning techniques, which have improved the accuracy of cardiac disease predictions. This methodology actively aims to highlight significant factors associated with heart disease. This study employs the Random Forest (RF) classification algorithm to train the model.**

***Keywords: Cardiac disease, Machine learning, Random Forest***

## 1. INTRODUCTION

The heart is a muscular organ that pumps blood throughout the body and serves as the central component of the cardiovascular system, which also includes the lungs. The cardiovascular system consists of a vast network of blood vessels, such as veins, arteries, and capillaries, responsible for delivering blood to all parts of the body. Abnormalities in normal blood flow from the heart lead to various types of heart diseases, collectively known as cardiovascular diseases (CVD). These diseases are a leading cause of death globally. According to the World Health Organization (WHO), approximately 17.5 million deaths occur each year due to heart attacks and strokes, with more than 75% of these deaths occurring primarily in middle-income and low-income countries. Furthermore, 80% of deaths attributed to CVDs result from strokes and heart attacks. Therefore, early detection of cardiac abnormalities and the development of predictive tools for heart diseases are crucial for saving lives and enabling doctors to design effective treatment plans, ultimately reducing mortality rates associated with cardiovascular diseases.

Advancements in healthcare systems have resulted in the availability of substantial patient data (i.e., Big Data in Electronic Health Record Systems), which can be utilized to create predictive models for cardiovascular diseases. Python, a programming language known for its high-level object-oriented abstraction, offers a dynamic array of building options and rapid development cycles. It is widely recognized as one of the safest programming languages, with numerous applications in the medical field. Additionally, it enjoys popularity and acceptance for its applications in AI-based software development and various other web applications.

## 2. LITERATURE SURVEY

The paper by Sree et al. (2023) [1] presents a comprehensive analysis of risk prediction for heart disease using four machine learning algorithms. The authors trained the model using four classification methods: Decision Tree (DT), K-Nearest Neighbors (K-NN), Random Forest (RF), and Support Vector Machine (SVM). They evaluated the performance of these algorithms based on several metrics, including recall, accuracy, precision, and specificity. The results demonstrated that SVM provided the best performance among the methods evaluated.

In the research conducted by Alkurdi (2023) [2], the Heart Disease UCI dataset was utilized to train machine learning models for heart disease diagnosis.

The study applied various preprocessing techniques, such as mean normalization, SMOTE, and correlation analysis. After preprocessing the data, it was fed into decision tree, random forest, SVM, and K-NN classification algorithms. The methodology yielded encouraging results, highlighting the importance of effective data preparation, as indicated by the accuracy, precision, recall, F1 score, and ROC AUC scores obtained.

Swain et al. (2023) [3] explored the use of IoT to assess risk factors associated with poorly managed cardiac conditions. They noted a significant technological gap between researchers in healthcare, patients, doctors, and machine learning architects. Their review emphasizes the potential to leverage advanced technologies in the field of machine learning.

Another study by Swain et al. (2023) [4] investigated how machine learning models could predict an individual's risk of developing coronary heart disease (CHD) over the following ten years, based on medical data and personal habits. Their goal was to identify the most accurate model for heart disease prediction using various machine learning classification techniques, including random forests (RF), logistic regression (LR), artificial neural networks (ANNs), and decision trees (DT). The findings indicated that the random forest model achieved the highest accuracy.

Handan Tanyildizi-Kökkülünk (2023) [5] focused on predicting heart disease using machine learning and data mining techniques. The categorization process was carried out using multiple linear regressions (MLR) in R Studio. By employing the Akaike information criterion to identify key characteristics significantly contributing to predictions, the study achieved an MLR model with an accuracy of 88%, precision of 93%, sensitivity of 86%, and specificity of 91%.

Ruqiya (2023) [6] provided a detailed review of the Cleveland Heart Disease Dataset through machine learning. From this review, they recommended increasing the sample size of the dataset for better evaluation and suggested implementing deep learning approaches, along with considering additional features for classification.

Zabeeulla et al. (2023) [7] introduced a machine learning algorithm for diagnosing cardiac issues, achieving an impressive accuracy of 98.8%. They utilized a publicly available heart disease dataset from Kaggle, which comprised thirteen features. The primary aim of their study was to determine whether patients had a 10-year risk of developing coronary heart disease (CHD).

Rindhe et al. (2021) [8] performed simulations using a dataset from the UCI Machine Learning repository, which included 303 samples with fourteen input features and one output feature. They applied supervised machine learning algorithms, including random forest, support vector machine, and artificial neural networks. The accuracies achieved were 84.0% for Support Vector Classifier, 83.5% for Neural Network, and 80.0% for Random Forest Classifier.

Anusuya & Gomathi (2021) [9] proposed a methodology that involved preprocessing, feature selection, eigenvector extraction, and classification. Additionally, they employed multi-objective-based Ant Colony Optimization (MOACO) to select the most relevant features from the extracted set, enhancing classification and clustering. Their proposed approach was compared with existing methodologies in terms of accuracy, recall, precision, normalized mutual information (NMI), and execution time, concluding that their solution was more effective for datasets of various sizes.

## 3. METHODOLOGY

The methodology for heart disease prediction involves a systematic approach, as illustrated in Figure 1. This process consists of several steps: data preprocessing of the input dataset, feature extraction, feature selection, splitting the data into training and testing datasets, classification, and ultimately, the output of the classifier, which predicts heart abnormalities.

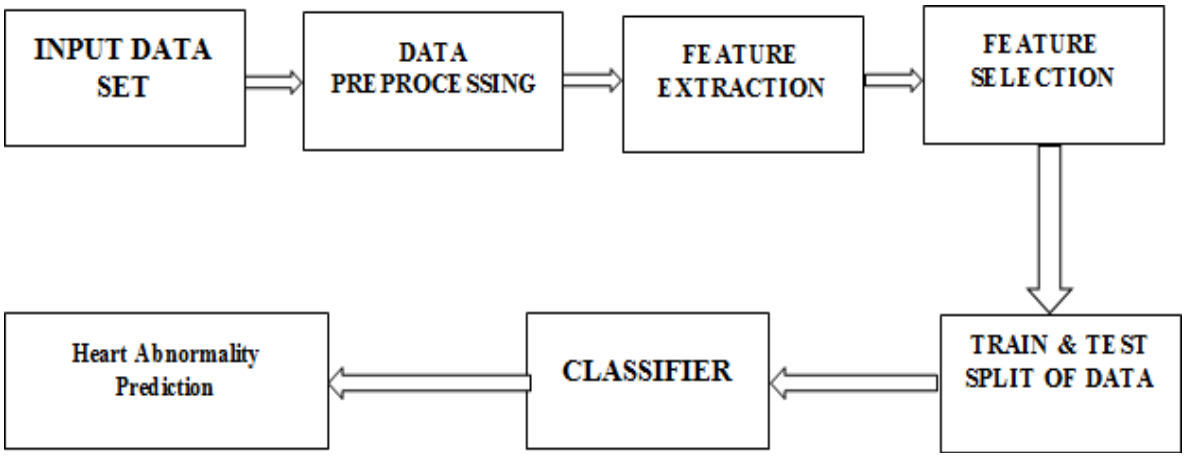


Fig. 1: System flow for prediction of heart disease

This study utilizes a dataset sourced from an online repository containing 1,025 samples for training a Random Forest classifier. The dataset includes critical parameters such as age, sex, chest pain type (CP), resting blood pressure (trestbps), cholesterol levels, fasting blood sugar (fbs), resting electrocardiographic measurements (restecg), maximum heart rate (thalach), exercise-induced angina (exang), ST depression (oldpeak), and the slope of ST.

Initially, the data undergoes preprocessing to ensure its quality and readiness for analysis. This step involves handling missing values, removing duplicates, and standardizing numerical features. After preprocessing, feature extraction techniques are applied to gather relevant information from the dataset, followed by feature selection to identify the most informative attributes for predicting heart disease.

The dataset is then divided into training and testing subsets to enable model training and evaluation. The Random Forest classifier is chosen for its effectiveness in handling classification tasks and is trained using the training dataset. Once the model is trained, it is deployed within a Graphical User Interface (GUI) developed using Python.

In the GUI, users can input various parameters related to heart health, including age, sex, chest pain type, blood pressure, and cholesterol levels. Based on these inputs, the Random Forest classifier generates predictions regarding the likelihood of heart disease, providing outputs that indicate either the possibility of heart disease or the absence of it. This information is valuable for healthcare professionals.

This methodology represents a comprehensive and accurate approach to predicting heart disease, leveraging machine learning techniques and GUI development to enhance accessibility and usability in healthcare settings.

## 4. CONCLUSION

In this study, an efficient machine learning-based diagnostic system has been developed for diagnosing heart disease. The application of machine learning in healthcare to predict health issues is rapidly growing and is essential for saving lives. Using the Python programming language, a GUI-based model employing the Random Forest method was created, achieving 100% accuracy. The output response time of this GUI-based model is minimal, making it user-friendly and easy to understand.

## REFERENCE

[1] Sree, P. K., Prasad, M., PBV, R. R., Ramana, C. V., Murty, P. T. S., Mallesh, A. S., & Raju, P. J. R. S. (2023, October 10). A Comprehensive Analysis on Risk Prediction of Heart Disease using Machine Learning Models. International Journal on Recent and Innovation Trends in Computing and Communication, 11(11s), 605–610. https://doi.org/10.17762/ijritcc.v11i11s.8295

[2] Alkurdi, A. A. H. (2023, January 1). Enhancing Heart Disease Diagnosis Using Machine Learning Classifiers. https://doi.org/10.54216/fpa.130101

[3] Swain, S. K., Behera, N., Swain, A. K., Jayasingh, S. K., Patra, K. J., Pattanayak, B. K., Mohanty, M. N., Naik, K. D., & Rout, S. (2023, October 7). Application of IoT Framework for Prediction of Heart Disease using Machine Learning. International Journal on Recent and Innovation

Trends in Computing and Communication. https://doi.org/10.17762/ijritcc.v11i10s.7616

[4] Swain, S., Chakravarty, S., Paikaray, B., & Bhoyar, H. (2023). Heart Disease Detection Using Machine Learning Techniques. Lecture Notes in Electrical Engineering, 273–284. https://doi.org/10.1007/978-981- 19-9090-8_24

[5] H, T. K. (2023, January 5). Prediction of Heart Disease Using Machine Learning with Data Mining. Physical Science & Biophysics Journal, 7(1), 1–6. https://doi.org/10.23880/psbj-16000228

[6] Ruqiya. (2023, June 30). Review on Cleveland Heart Disease Dataset using Machine Learning. Volume 21, Issue 1, 21(1), 87–98. https://doi.org/10.52584/qrj.2101.11

[7] Zabeeulla, M., Sharma, C., & Anand, A. (2023, March 1). Early Detection of Heart Disease Using Machine Learning Approach. CARDIOMETRY, 26, 342–347. https://doi.org/10.18137/cardiometry.2023.26.342347

[8] Rindhe, B. U., Ahire, N., Patil, R., Gagare, S., & Darade, M. (2021, May 12). Heart Disease Prediction Using Machine Learning. International Journal of Advanced Research in Science, Communication and Technology, 267–276. https://doi.org/10.48175/ijarsct-1131

[9] Anusuya, V., & Gomathi, V. (2021, March 25). An Efficient Technique for Disease Prediction by Using Enhanced Machine Learning Algorithms for Categorical Medical Dataset. Information Technology and Control, 50(1), 102–122. https://doi.org/10.5755/j01.itc.50.1.25349

[10] Shadman Nashif, Md. Rakib Raihan, Md. Rasedul Islam, Mohammad Hasan Imam. Heart Disease Detection by Using Machine Learning Algorithms and a Real-Time Cardiovascular Health Monitoring System. World Journal of Engineering and Technology, 2018, 6, 854-873. http://www.scirp.org/journal/wjet

[11] D. (n.d.). Datasets/DS/heart.csv at main DataThinkers/Datasets. GitHub. https://github.com/DataThinkers/Datasets/blob/main/DS/heart.csv

[12] Online, E. (2023, September 29). World Heart Day: 27% of deaths in India are caused by cardiovascular diseases. The Economic Times. https://economictimes.indiatimes.com/magazines/panache/world-heart-day27-per-cent-of-deaths-in-india-are-caused-by-cardiovascular-diseases/articleshow/104031929.cms