

# Detect And Predict Heart Disease Using Machine Learning

S. SHIRELY<sup>1</sup>, M. SHOUBIYA<sup>2</sup>

<sup>1</sup>MCA, (Assistant Professor, Master of Computer Applications)

<sup>2</sup>MCA, Christ College of Engineering and Technology  
Moolakulam, Oulgaret Municipality, Puducherry-605010.

**Abstract**—Heart diseases are one of the most severe challenges in the current medical sector and are a major cause of deaths worldwide. Prompt diagnosis and correct risk estimation have been given prime importance in preventing death and increasing the survival rate of the patient. The current traditional diagnosis process is a time-consuming and expensive process with a strong reliance on the expertise of the doctor, resulting in fluctuating diagnosis results. Keeping in mind these issues, a Machine Learning system for the Prediction of Heart Diseases was proposed in this paper.

The system evaluates several variables including age, gender, blood pressure, levels of cholesterol, heart rate, and other variables of clinical significance for classifying patients as being at a higher or a lesser risk. Several machine learning models have been trained and tested on the popular UCI heart diseases datasets, which ensures reliable testing of the proposed work. The system is tested for its performance using accuracy and other measures for assessing the efficiency of the proposed system.

Experimental results also show that the system yields high predictive accuracy, which promises its ability to be an efficient, reliable, and scalable solution for heart disease risk assessment. The proposed model has great potential for supporting healthcare professionals by enabling quicker diagnosis and improvement of preventive care, thereby reducing the overall burden on healthcare systems. This study implicates the importance of machine learning techniques in modern healthcare analytics and their potential role in enhancing early disease detection and patient outcomes.

**Index Terms**—Prediction of Heart Diseases, Machine Learning, Learning Algorithms in Supervised Learning, Analysis in Healthcare, Medical Data Mining, Predictive Modeling in Healthcare, Classification of Risk, Clinical Support Systems in Medicine, Early Detection of Diseases, Monitoring of Patient's Healthcare, Analysis of Cardiovascular Diseases in Medicine, Supporting Systems in Medicine for Diagnosis, Health Informatics in UCIheart disease dataset – Analysis of Classifications in Medicine –

**Metrics for Performance Evaluation – Preventive Healthcare.**

## I. INTRODUCTION

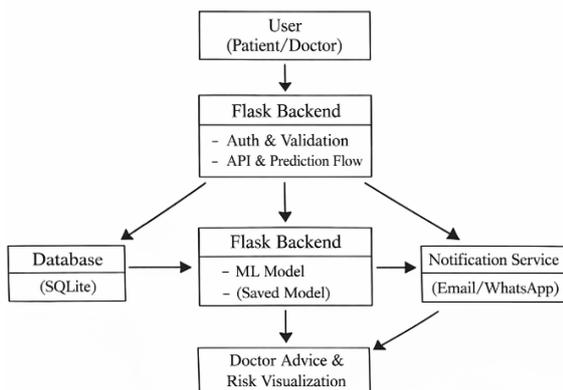
Cardiovascular diseases (CVDs) are known to be responsible for approximately millions of deaths worldwide every year, making them one of the top-leading causes for mortality globally. Given its intensity levels, complications, and ever-increasing numbers, CVDs can be termed one of the most important challenges for public healthcare today [1], [2]. There are certain prominent factors related to lifestyle, stresses, smoking, being overweight, being physically inactive, or improper food habits that tend to substantially raise figures concerning those affected by cardiovascular diseases.

Identification and diagnosis of patients at risk of having cardiovascular diseases remain important in preventing complications and improving treatment outcomes among patients. However, conventional methods for diagnosing patients at risk of having cardiovascular diseases often take lengthy procedures, are costly, and require specialized skills. With the ever-increasing number of cases of medical and health-related data, machine learning approaches have become efficient for the analysis of a variety of complicated datasets in the field of medicine and the recognition of patterns or trends which are not easily identified through conventional statistical analysis methods [6], [17]. Keeping the above application areas in perspective, the project undertaken for this thesis revolves around the design of an online heart disease prediction system that utilizes machine learning techniques along with alerting and advising functionalities to support healthcare practitioners and patients. The proposed system has been designed to facilitate early predictions, support decision-making, and encourage preventive healthcare through timely alerts and personal advising.

## II. LITERATURE SURVEY

There have been some studies exploring the use of machine learning techniques for predicting heart diseases. Smith et al. in [3] designed and developed a logistic regression approach that showed promising results. Kumar and Devi in [4] experimentally showed that ensemble approaches result in better predictions using decision trees and random forest. Shinde & Kulkarni [7] used supervised learning classifiers like Naïve Bayes & K-NN for classifying patients with heart disease. Polat et al. [8] employed the use of neural networks for improving the process of diagnosing heart disease. Recent research has recognized the efficiency of standardized data sets like UCI Heart Disease for effective classification [5], [16]. Nonetheless, most of the current technologies do not offer patient alert services or personal recommendations, which this paper purposed to solve.

## III. SYSTEM ARCHITECTURE



The user interacts with the web interface to input the health information of the patients. This information is then handled by the Flask backend and transmitted to the developed machine learning model. The result is stored in the database. In the case of a high-risk patient, warning functions like sending an email or WhatsApp message are enabled. According to the result, health-related suggestions are given to the user.

## VI. PROPOSED METHODOLOGY

### 4.1 Data Collection

The data used for the analysis are collected from the UCI Machine Learning Repository [5]. The data comprises the following 13 medical features: age,

chest pain type, blood pressure, cholesterol, fasting blood sugar, and maximum heart rate.

### 4.2 Data Preprocessing

Preprocessing techniques include handling Missing Values, scaling of features by normalization methods, and division of datasets into Training and Test datasets [18].

### 4.3 Model Training

A supervised machine learning classification technique is used for developing the model for prediction of the heart disease. Logistic Regression and Random Forest classifiers have been used in this task.

Logistic Regression is employed as a baseline model to predict the probability of the occurrence of heart disease given the input clinical variables. It attempts to establish the relation between the independent variables and the binary output or target (heart disease or no heart disease) using the sigmoid function, which makes the model appropriate for medical risk prediction tasks [6].

The performance of the prediction can be enhanced, as well as the interaction of the features, by using a Random Forest Classifier. The Random Forest Classifier is an ensemble learning algorithm, which trains a forest of multiple decision trees to make a prediction.

This method helps in preventing overfitting and improves classification accuracy by combining multiple tree outputs [20].

The dataset is preprocessed before the training phase because the features in the dataset are normalized through feature scaling methods in order to make all the medical attributes in the dataset make an equal contribution to the training process. The dataset is split into two sets: the training set and the test set. During the training phase, the models develop important correlations and patterns in the attributes like age, chest pain type, cholesterol, and maximum heart rate.

The models produce a decision boundary that identifies high-risk and low-risk patients and predicts heart disease effectively.

The trained model creates a decision boundary that differentiates high-risk patients from low-risk patients with optimal results.

#### 4.4 Predictions and Classification

They include:

The trained model classifies patients into:

- High Risk- It shows there is more likelihood of having heart disease
- Low Risk - Low probability of the presence of heart disease

$$\text{Result} = \begin{cases} \text{High Risk, if } P \geq 0.5 \\ \text{Low Risk, if } P < 0.5 \end{cases}$$

#### 4.5 Alert and Advisory Module

The alert and advisory component is initiated depending on the result of the prediction outcome. In the case where a patient is identified as a High Risk individual, the automatic alert notifications are sent to both the patient as well as the hospital via email and WhatsApp services [12].

In both high and low risk cases, patient-specific doctor recommendations are provided by incorporating predefined medical guidelines and prediction output by the system itself. The advisory component aims at providing advice on lifestyle, food, and activities that result in improved preventive health awareness among patients [16], [25].

### V. RESULTS AND DISCUSSION

As a result, experimental evaluation shows that the proposed heart disease prediction system achieves an accuracy of roughly 85%, which is comparable to or even better than several existing systems reported in the literature [3], [9], [24]. The model's usability and effectiveness are further enhanced by visualization techniques in the form of bar charts, clearly representing the probability distributions associated with the outcomes of predictions, making them more comprehensible to users.

These results support the fact that machine learning-based prediction systems can serve as valuable clinical decision-support tools by facilitating early diagnosis, timely intervention, and evidence-based medical decision-making. The findings demonstrate the potential of machine learning approaches in improving diagnostic accuracy and efficiency in modern healthcare environments.

### VI. CONCLUSION

The paper proposed a system for predicting heart disease based on clinical parameters utilizing machine learning algorithms. The system is capable of classifying patient risks accurately into high and low categories. The proposed system, which uses a combination of web-based technology and machine learning algorithms, eliminates the reliance on manual diagnosis [3], [4].

The addition of real-time alert systems like Email and WhatsApp notification systems improves patient safety by ensuring that patients with a higher risk condition seek medical attention promptly [2], [12]. The provision of personalized advice from doctors also educates patients and helps them adopt healthy practices. The experimental results show that this prediction method provides a reliable prediction outcome and can be applied as a decision-support tool [7], [16].

### VII. FUTURE ENHANCEMENTS

- Integration with wearable health monitoring tools can support real-time collection of health information such as heart rate and activity level, which can increase accuracy of prediction models [2], [18].
- More sophisticated models of deep learning techniques like CNN and RNN can be integrated to address the issue of handling medical data [19], [20].
- The model can be extended to predict multiple chronic conditions like diabetes and hypertension on the same platform [11], [21].
- Hosting the cloud on-premise may ensure the protection of patient information, but security can be a challenge when cloud services are accessed remotely [20].
- A mobile application development can increase engagement levels among users as it will enable them to receive instant notifications, reports, as well as doctor's advice at any given time and at any given location [12], [25].

### REFERENCES

- [1] World Health Organization. (2022). Cardiovascular Diseases (CVDs). Geneva
- [2] American Heart Association, Heart Disease and Stroke Statistics—2021 Update, USA, 2021.

- [3] J. Smith et al., "Heart disease prediction using machine learning techniques," IEEE J. Biomed.Health Inform., 2021.
- [4] R. Kumar and S. Devi, "Comparative analysis of ML algorithms for heart disease Prediction
- [5] D. Dua and C. Graff, "UCI Machine Learning Repository," Univ. of California, Irvine,
- [6] T.M. Mitchell, Machine Learning
- [7] S. Shinde and P. Kulkarni, "Prediction of heart disease using ML," IJCSIT,
- [8] K. Polat et al., "Heart disease diagnosis using neural networks," Expert Systems with Applications, vol. 34, Applications,
- [9] A. Sen et al., "Heart disease prediction using decision tree," IJACR, 2013.
- [10] M. Jabbar, et al., "Heart Disease Classification Using KNN," Procedia Technology, vol.
- [11] Y. Jain et al., "Minimization of heart disease risk using data mining," IJARCS, 2016.
- [12] S. Rajesh et al., "Early prediction of heart disease," IJERT, 2020
- [13] N. Al-Ajam and E. Al-Shammari, "ANN-based heart disease prediction,"
- [14] H. Masethe & M. Masethe, "Heart disease prediction using classification,"
- [15] P. Thakkar and S. Mehta, "Survey on heart disease prediction," IJSR,
- [16] S. Palaniappan and R. Awang, "Intelligent heart disease prediction," IJ
- [17] I. Witten et al., Data Mining, Morgan Kaufmann, 2011
- [18] A. Uysal and S. Gunal, "Impact of preprocessing on classification," Computers & Security, vol. Electrical Engineering, 2018
- [19] F. Chollet, Deep Learning with Python, Manning,
- [20] J. Han et al., Data Mining Concepts and Techniques, Morgan Kaufmann, 2012.
- [21] M. Mythili and A. Shanavas, "Heart disease prediction analysis," IJCA
- [22] A. Gupta et al., "ML approach for heart disease prediction," IJARCET, 2017.
- [23] S. Shinde, and P. Kulkarni, "Heart disease prediction using ML," IJERT
- [24] M. Nilashi et al., "Analytical method for heart disease prediction," Telematics and Informatics
- [25] K. Rao et al., "ML approach for heart disease prediction," Journal of Medical Systems, 2020.