**< Heart Disease Prediction Project >**

**A Project Work Synopsis**

*Submitted in the partial fulfillment for the award of the degree of*

**BACHELOR OF ENGINEERING**

**IN**

**COMPUTER SCIENCE WITH SPECIALIZATION IN**

**ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING**

**Submitted by:**

21BCS6325- Nikhil Sharma

21BCS3713-Kartik Kaushik

21BCS11016- Satyam Kumar Singh

21BCS11060- Aryan Chaudhary

**Under the Supervision of:**

**<** Ms. Akansha Moral **>**



**CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413,**

**PUNJAB**

**September, 2022**

# Table of Contents

|  |  |
| --- | --- |
| Title Page | i |
| Abstract | ii |
| 1.    Introduction |  |
| 1.1           Problem Definition |  |
| 1.2 Project Overview |  |
| 1.3 Hardware Specification |  |
| 1.4 Software Specification |  |
| 2.    Literature Survey |  |
| 2.1 Existing System |  |
| 2.2 Proposed System |  |
| 2.3 Literature Review Summary |  |
| 3.    Problem Formulation |  |
| 4.    Research Objective |  |
| 5.    Methodologies |  |
| 6.    Experimental Setup |  |
| 7.    Conclusion |  |
| 8.    Tentative Chapter Plan for the proposed work |  |
| 9.    Reference |  |

# Abstract

Heart disease is the leading cause of death for people worldwide. To treat heart disease, many advanced techniques are used. In medical centers, the most common problem is that many medical staff do not have the same knowledge and expertise to treat their patients, so they make their own decisions that turn out to be poor and sometimes lead to death. To overcome these problems, predicting heart diseases using machine learning algorithms and data mining techniques becomes easy for automatic diagnosis in hospitals as they play an important role in this. Heart diseases can be predicted by analyzing various health parameters of patients. There are different algorithms to predict heart diseases such as Naive Bayes, k nearest neighbors (KNN), decision trees, and artificial neural networks (ANN). We use different parameters to predict heart disease. The parameters are age, sex, cerebral palsy (CP), Gender, Cerebpalsylsey (CP), Blood Pressure (bp), Fasting blood sugar test (FBS), etc. Experiments show that Naïve Bayes has the highest accuracy of almost 90%.

Keywords:Naive Bayes, k Nearest Neighbor (KNN), Decision tree, Artificial Neural Network (ANN), Random Forest, Heart Disease

# 1. INTRODUCTION

## 1.1 Problem Definition

Heart disease prediction project aims to develop a machine-learning model that can accurately predict the presence or absence of heart disease in patients based on their medical history and diagnostic test results. The goal of the project is to improve the accuracy and efficiency of the diagnosis process, as well as to provide early detection and intervention for individuals at risk of heart disease. The project will involve collecting and analyzing a dataset of patient records and diagnostic test results, which will be used to train and test the machine learning model. The model will be developed using various techniques such as feature engineering, feature selection, and model optimization to achieve high accuracy in predicting heart disease. The ultimate objective of the project is to provide healthcare practitioners with a reliable and efficient tool for diagnosing and managing heart disease in patients. This project **could** **have** a significant public health **impact** by improving **diagnostic** accuracy and early detection of heart disease, a leading cause of death worldwide.

## 1.2 Problem Overview

## Heart disease prediction is a crucial problem in healthcare that aims to accurately predict the likelihood of a person developing heart disease. Heart disease, also known as cardiovascular disease, is a leading cause of death globally and is responsible for numerous health complications.

## 1.3 Hardware Specification

Processor: A modern, multi-core processor such as an Intel Core i5 or i7, or an AMD Ryzen processor.

RAM: At least 8GB of RAM, although 16GB or more is recommended for more demanding use cases.

Storage: Sufficient storage space for the software application, as well as any necessary datasets and models. A solid-state drive (SSD) is recommended for faster read/write speeds

Internet connection: A stable and reliable internet connection is necessary for accessing online datasets, as well as for downloading updates and models.

## 1.4 Software Specification

Operating System - Any OS with clients to access the internet

Network - Wi-Fi Internet or cellular Network

Visio Studio - Create and design Data Flow and Context Diagram Github - Versioning Control

Google Chrome - Medium to find references to do system testing, displaying and run shiny app

# 2. LITERATURE SURVEY

## 2.1 Existing System

There are many existing systems for heart disease prediction using machine learning. Here are a few examples:

1. Framingham Risk Score: The Framingham Heart Study developed a risk score based on demographic, medical, and lifestyle factors to predict the risk of developing coronary heart disease.
2. Support Vector Machine (SVM) Model: SVM is a popular machine learning algorithm used for classification and regression analysis. It has been used to predict the risk of heart disease based on patient data.
3. Random Forest Model: Random forest is a machine learning algorithm that builds multiple decision trees and combines their outputs to make predictions. It has been used to predict heart disease risk based on patient data.
4. Logistic Regression Model: Logistic regression is a statistical method used for binary classification problems. It has been used to predict the likelihood of heart disease based on patient data.
5. Deep Learning Model: Deep learning is a subset of machine learning that uses neural networks to learn patterns in data. It has been used to predict heart disease risk based on patient data.

## 2.2 Proposed System

## 2.3 Literature Review Summary (Minimum 7 articles should refer)

.

|  |  |  |  |
| --- | --- | --- | --- |
| Author | Title | Purpose | Algorithms used and accuracy |
| 1. Mr. Santhana.J,   Dr. Geetha.S | Prediction of Heart Disease Using Machine Learning Algorithms Year-2018 | In this system, a heart disease dataset is used. The main aim of this system is to predict the possibilities of occurring heart disease of the patient in terms of percentage.This is performed through data mining classification techniques. | Decision tree 91% Native Bayes 87% |
| 1. M.Marimuthu, 2. S. Deivarani, Gayathri.R | Analysis of Heart Disease Prediction using various Machine Learning Techniques Year 2018 | To achieve better accuracy and to make the system more efficient so that It can predict the chances of heart attack | KNN 83.60%, NB 80.66%, Decision Tress 75.5%, SVM 65.56% |
| 1. Sanchayita Dhar, Pritha Datta, Ankur Biswas, Tanusree Dey, Krishna Roy | A hybrid Machine Learning Approach for prediction of Heart Diseases.  Year 2018 | In this paper, to develop a prediction system that be capable to envisage heart diseases based on measurements , are extracted from the The ERIC laboratpry consisting of 209 test cases. | NB, Decision tree, random forest |
| 1. Rajesh N, T Mannesha, Shaik Hafeez, Hari Krishna | Predction of Heart Dsease Using Machine Learning Algorithms Year 2018 | In this paper, processing patient’s dataset and a data of patients to whom we need to predict the chance of occurrence of a heart disease. | NB, Decision Tree(ID3 ALgorithm) |

# 4. OBJECTIVES

The purpose of this project is to determine if a patient can be diagnosed with the cardiovascular disease based on gender, age, and medical characteristics such as chest pain and fasting blood glucose. Data sets are selected from the UCI repository along with medical records. patient's attributes. We use this data set to predict whether a patient has heart disease. To predict this, it uses a patient's 14 medical characteristics and classifies whether the patient has a heart condition. These medical properties are trained on 4 algorithms: NB, logistic regression, KNN, and random forest classifier. The most efficient of these algorithms is KNN, which gives an accuracy of 88.52%. Finally, we classify patients at risk of cardiovascular disease and this method is completely cost-effective.

# 5. METHODOLOGY

The methodology for a heart disease prediction project may vary depending on the data available and the specific goals of the project. However, a general approach to building a heart disease prediction model can be broken down into the following steps:

1. Data collection and cleaning: Collecting data from various sources and cleaning the data to remove missing or inconsistent values, and to ensure that the data is in a format suitable for analysis.
2. Data exploration and visualization: Exploring and visualizing the data to gain insights and identify patterns or relationships that may be relevant to the prediction of heart disease.
3. Feature selection: Selecting the most relevant features (variables) that are likely to have an impact on the prediction of heart disease. This may involve using statistical techniques such as correlation analysis or principal component analysis (PCA).
4. Model selection: Selecting an appropriate machine learning model based on the type of problem, available data, and the level of accuracy required. Some common models for heart disease prediction include logistic regression, decision trees, and neural networks.
5. Model training and evaluation: Splitting the data into training and testing sets, training the model on the training data, and evaluating its performance on the testing data. The evaluation may involve metrics such as accuracy, precision, recall, and F1 score.
6. Hyperparameter tuning: Fine-tuning the model parameters to optimize its performance on the testing data.
7. Deployment and monitoring: Deploying the model in a production environment and monitoring its performance over time to ensure that it continues to produce accurate predictions.

The analysis of various machine learning algorithms, the algorithms that are used in this paper are K nearest neighbors (KNN), Logistic Regression, and Random Forest Classifiers which can be helpful for practitioners or medical analysts to accurately diagnose Heart Disease. The methodology gives a framework for the proposed model**.** The methodology is a process that includes steps that transform given data into recognized data patterns for the knowledge of the users. The first step is referred to as the collection of the data in the second stage extracts significant values and the 3rd is the preprocessing stage where we explore the data. Data preprocessing deals with missing values, cleaning of data, and normalization depending on the algorithms used**.** After pre-processing of data, the classifier is used to classify the pre-processed data the classifier used in the proposed model is KNN, Logistic Regression, and Random Forest Classifier. Finally, the proposed model is undertaken, where we evaluated our model on the basis of accuracy and performance using various performance metrics.

# 6.EXPERIMENTAL SETUP

# 7.CONCLUSION

# Identifying the processing of raw healthcare data of heart information will help in the long-term saving of human lives and early detection of abnormalities in heart conditions. Machine learning techniques were used in this work to process raw data and provide a new and novel discernment towards heart disease. Heart disease prediction is challenging and very important in the medical field. However, the mortality rate can be drastically controlled if the disease is detected at the early stages and preventative measures are adopted as soon as possible. Further extension of this study is highly desirable to direct the investigations to real-world datasets instead of just theoretical approaches and simulations. The proposed hybrid HRFLM approach is used combining the characteristics of Random Forest (RF) and Linear Method (LM). HRFLM proved to be quite accurate in the prediction of heart disease. The future course of this research can be performed with diverse mixtures of machine learning techniques to better prediction techniques. Furthermore, new feature selection methods can be developed to get a broader perception of the significant features to increase the performance of heart disease prediction.

# 8. TENTATIVE CHAPTER PLAN FOR THE PROPOSED WORK

**CHAPTER 1: INTRODUCTION**

**CHAPTER 2: LITERATURE REVIEW**

**CHAPTER 3: OBJECTIVE**

**CHAPTER 4: METHODOLOGIES**

**CHAPTER 5: EXPERIMENTAL SETUP**

**CHAPTER 6: CONCLUSION AND FUTURE SCOPE**

## REFERENCES

[1] Soni J, Ansari U, Sharma D & Soni S (2011). *Predictive data mining for medical diagnosis: an overview of heart disease prediction.* International Journal of Computer Applications, ***17*(8), 43-8**

[2] Dangare C S & Apte S S (2012). *Improved study of heart disease prediction system using data mining classification techniques*. *International Journal of Computer Applications*, ***47*(10), 44-8**.

[3] Ordonez C (2006). *Association rule discovery with the train and test approach for heart disease prediction*. *IEEE Transactions on Information Technology in Biomedicine*, *10*(2), **334-43**.

[4] Shinde R, Arjun S, Patil P & Waghmare J (2015). *An intelligent heart disease prediction system using k-means clustering and Naïve Bayes algorithm*. *International Journal of Computer Science and Information Technologies*, ***6*(1)**, 637-9.

[5] Jee S H, Jang Y, Oh D J, Oh B H, Lee S H, Park S W & Yun Y D (2014). *A coronary heart disease prediction model: the Korean Heart Study*. *BMJ open*, ***4*(5)**, e005025.

[6] Ganna A, Magnusson P K, Pedersen N L, de Faire U, Reilly M, Ärnlöv J & Ingelsson E (2013). *Multilocus genetic risk scores for coronary heart disease prediction*. *Arteriosclerosis, thrombosis, and vascular biology*, ***33*(9)**, 2267-72.

[7] Jabbar M A, Deekshatulu B L & Chandra P (2013, March). *Heart disease prediction using lazy associative classification*. In *2013 International Mutli-Conference on Automation, Computing,Communication, Control and Compressed Sensing (iMac4s)* (pp. 40- 6). **IEEE**.

[8] Dangare Chaitrali S and Sulabha S Apte. *"Improved study of heart disease prediction system using data mining classification techniques." International Journal of Computer Applications* 47.10 (2012): 44-8.

[9] Soni Jyoti. "Predictive data mining for medical diagnosis: *An overview of heart disease prediction*." *International Journal of Computer Applications* 17.8 (2011): 43-8.

[10] Chen A H, Huang S Y, Hong P S, Cheng C H & Lin E J (2011, September). *HDPS: Heart disease prediction system*. In *2011 Computing in Cardiology* (pp. 557-60). **IEEE**.

[11] Parthiban, Latha and R Subramanian. *"Intelligent heart disease prediction system using CANFIS and genetic algorithm." International Journal of Biological, Biomedical and Medical Sciences* 3.3 (2008).

[12] Wolgast G, Ehrenborg C, Israelsson A, Helander J, Johansson E & Manefjord H (2016). *Wireless body area network for heart attack detection [Education Corner]*. *IEEE antennas and propagation magazine*, ***58*(5)**, 84-92.

[13] Patel S & Chauhan Y (2014). *Heart attack detection and medical attention using motion sensing device -kinect*. *International Journal of Scientific and Research Publications*, ***4*(1)**, 1-4.

[14] Raihan M, Mondal S, More A, Sagor M O F, Sikder G, Majumder M A & Ghosh K (2016, December). *Smartphone based ischemic heart disease (heart attack) risk prediction using clinical data and data mining approaches, a prototype design*. In *2016 19th*

*International Conference on Computer and Information Technology (ICCIT)* (pp. 299-303). **IEEE**.

[15] Buechler K F & McPherson P H (1999). *U.S. Patent No. 5,947,124*. Washington, DC: U.S. Patent and Trademark Office.

[16] Takci H (2018). *Improvement of heart attack prediction by the feature selection methods*. *Turkish Journal of Electrical Engineering & Computer Sciences*, ***26*(1)**, 1-10.

[17] Worthen W J, Evans S M, Winter S C & Balding D (2002). *U.S. Patent No. 6,432, 124*. Washington, DC: U.S. Patent and Trademark Office.

[18] Brown N, Young T, Gray D, Skene A M & Hampton J R (1997). *Inpatient deaths from acute myocardial infarction*, 1982-92: analysis of data in the Nottingham heart attack register. *BMJ*, ***315*(7101)**, 159-64.

[19] Folsom A R, Prineas R J, Kaye S A & Soler J T (1989). *Body fat distribution and self-reported prevalence of hypertension, heart attack, and other heart disease in older women*. *International journal of epidemiology*, ***18*(2)**, 361-7.

[20] Kiyasu J Y (1982). *U.S. Patent No. 4,338,396*. Washington, DC: U.S. Patent and Trademark Office.