

< Heart Disease Prediction Project >

A Project Work Synopsis

Submitted in the partial fulfilment 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. Akanksha Moral >



**CHANDIGARH
UNIVERSITY**

Discover. Learn. Empower.

**CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413,
PUNJAB
April, 2023**

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. Research Objective

4. Methodologies

5. Conclusion

6. References

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:-

Our 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 (CG reports in some cases). The goal of our project is to improve the accuracy and efficiency of the diagnosis process, as well as to provide early detection and prevention for individuals at risk of heart disease. The project will involve collecting and analyzing various datasets 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, feature extraction , and model optimization to achieve considerably high accuracy in predicting heart disease.

2.3 Literature Review Summary

S. No.	Author	Title	Purpose	Algorithms used and accuracy
1	Devansh Shah · Samir Patel · Santosh Kumar Bharti	Heart Disease Prediction using Machine Learning Techniques	Aim of this research is to predict whether or not a patient will develop heart disease. This research was done on supervised machine learning classification techniques using Naïve Bayes, decision tree, random forest, and K-nearest neighbor on UCI repository.	Naïve Bayes 74% Naïve Bayes 95.556% Decision tree SVM (Highest accuracy obtained by SVM)

2	Divya Krishnani, Anjali Kumari, Akash Dewangan, Aditya Singh, Nenavath Srinivas Naik	Prediction of Coronary Heart Disease using Supervised Machine Learning Algorithms	We propose a preprocessing extensive work where Random Forest is the most compatible contender for the prediction model and gives the highest performance measure among K-Nearest Neighbour and Decision Tree.	The accuracy, recall, precision, specificity and F1 score of RF on the proposed work are 96.71%, 98.74%, 94.4%, 99%, 96.61% respectively
3	LIAQAT ALI , ATIQUR RAHMAN, AURANGZEB KHAN, MINGYI ZHOU, ASHIR JAVEED, AND JAVED ALI KHAN	An Automated Diagnostic System for Heart Disease Prediction Based on 2 Statistical Model and Optimally Configured Deep Neural Network	In this paper, we have developed an automated diagnostic system for the diagnosis of heart disease. The proposed diagnostic system used 2 statistical model for features refinement and DNN for classification.	We achieved classification accuracy of 93.33% on 11 features using neural network with two hidden layers.
4	MOHAMMAD AYOUB KHAN	An IoT Framework for Heart Disease Prediction Based on MDCNN Classifier	Wearable technologies can be utilized effectively in healthcare industry, particularly in chronic heart disease. The monitoring and prediction systems can help to save many lives by instant intervention specially when patient is located at remote place where medical facilities are not present.	Accuracy of MDCNN - 93.3%
5	LIAQAT ALI , AWAIS NIAMAT, JAVED ALI KHAN, NOORBAKHS AMIRI GOLILARZ, XIONG XINGZHONG, ADEEB NOOR, REDHWAN NOUR, AND SYED AHMAD CHAN BUKHARI	An Optimized Stacked Support Vector Machines Based Expert System for the Effective Prediction of Heart Failure	In this paper, an expert system based on stacked SVMs was proposed to facilitate the diagnosis of heart failure. The first SVM model was used to eliminate irrelevant features while the second model was used as predictive model. Both the models were optimized using a hybrid grid search algorithm.	L1 Linear SVM +L2 Linear & RBF SVM - Accuracy 92.22%
6	SENTHILKUMAR MOHAN , CHANDRASEGAR THIRUMALAI, AND GAUTAM SRIVASTAVA	Effective Heart Disease Prediction Using Hybrid Machine Learning Techniques	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.	multiple models are used, with accuracy varying from 75% to 88,4%
7	JIAN PING LI , AMIN UL HAQ , SALAH UD DIN , JALALUDDIN KHAN , ASIF KHAN , AND ABDUS SABOOR	Heart Disease Identification Method Using Machine Learning Classification in E-Healthcare	Machine learning classifiers include LR, K-NN, ANN, SVM, NB, and DT are used in the designing of the system. Four standard feature selection algorithms including Relief, MRMR, LASSO, LLBFS, and proposed a novel feature selection algorithm FCMIM used to solve feature selection problem. LOSO cross-validation method is used in the system for the best hyperparameters selection.	Proposed model FCMIM - SVM , with accuracy - 92.37%

8	Dr. M. Kavitha, G. Gnaneswar , R. Dinesh, Y. Rohith Sai, R. Sai Suraj	Heart Disease Prediction using Hybrid machine Learning Model	There are many diagnosis processes available in the medical industry. However, in terms of accuracy, machine learning is considered the best choice. The proposed work uses a TkInter Python designed application for the heart disease prediction. The proposed system using combinations of Decision Tree and Random forest for heart disease prediction as a hybrid model. Cleveland database is used for this study.	Decision Tree 79% Random Forest 81% Hybrid (Decision Tree+ Random Forest) 88%
---	---	--	--	---

This table provides a summary of eight different research papers that propose machine learning models for heart disease prediction. The algorithms used and their respective accuracy are also listed.

1. Devansh Shah et al. proposed a heart disease prediction model using Naïve Bayes, decision tree, random forest, and K-nearest neighbor algorithms with the highest accuracy obtained by SVM, which was 95.556% using Naïve Bayes and 74% using other algorithms.
2. Divya Krishnani et al. proposed a heart disease prediction model using Random Forest, K-Nearest Neighbor, and Decision Tree algorithms, with Random Forest giving the highest performance measure among the three models. The accuracy, recall, precision, specificity, and F1 score of the Random Forest model were 96.71%, 98.74%, 94.4%, 99%, and 96.61%, respectively.
3. LIAQAT ALI et al. proposed an automated diagnostic system for heart disease using 2 statistical model for features refinement and DNN for classification, which achieved a classification accuracy of 93.33% on 11 features using neural network with two hidden layers.

4. MOHAMMAD AYOUB KHAN proposed an IoT framework for heart disease prediction based on MDCNN Classifier with an accuracy of 93.3%.
5. LIAQAT ALI et al. proposed an expert system based on stacked SVMs for the effective prediction of heart failure, which used L1 Linear SVM +L2 Linear & RBF SVM and achieved an accuracy of 92.22%.
6. Senthilkumar Mohan et al. proposed a heart disease prediction model using a hybrid approach combining Random Forest (RF) and Linear Method (LM), which proved to be quite accurate with multiple models used, and accuracy varying from 75% to 88.4%.
7. JIAN PING LI et al. proposed a heart disease identification method using machine learning classification in e-healthcare, which used LR, K-NN, ANN, SVM, NB, and DT algorithms, along with four standard feature selection algorithms and a proposed novel feature selection algorithm FCMIM, achieving an accuracy of 92.37% with the proposed model FCMIM - SVM.
8. Dr. M. Kavitha et al. proposed a heart disease prediction model using a hybrid model of Decision Tree and Random forest algorithms, with an accuracy of 88% for the hybrid model, 81% for Random Forest, and 79% for Decision Tree.

Overall, the research papers show that machine learning models are effective in predicting heart disease, with accuracies ranging from 74% to 97.6%. The choice of algorithm and feature selection technique plays a critical role in the accuracy of the prediction models.

3. OBJECTIVE:-

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.

4. 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 metric

5.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.

6.REFERENCES:-

1. Shah, Devansh, Samir Patel, and Santosh Kumar Bharti. "Heart disease prediction using machine learning techniques." *SN Computer Science* 1 (2020): 1-6.

2. Krishnani, Divya, et al. "Prediction of coronary heart disease using supervised machine learning algorithms." *TENCON 2019-2019 IEEE Region 10 Conference (TENCON)*. IEEE, 2019.
3. Ali, Liaqat, et al. "An automated diagnostic system for heart disease prediction based on χ^2 statistical model and optimally configured deep neural network." *Ieee Access* 7 (2019): 34938-34945.
4. Khan, Mohammad Ayoub. "An IoT framework for heart disease prediction based on MDCNN classifier." *IEEE Access* 8 (2020): 34717-34727. [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.
5. Ali, Liaqat, et al. "An optimized stacked support vector machines based expert system for the effective prediction of heart failure." *IEEE Access* 7 (2019): 54007-54014. [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**.
6. Mohan, Senthilkumar, Chandrasegar Thirumalai, and Gautam Srivastava. "Effective heart disease prediction using hybrid machine learning techniques." *IEEE access* 7 (2019): 81542-81554. [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.

7. Li, Jian Ping, et al. "Heart disease identification method using machine learning classification in e-healthcare." *IEEE Access* 8 (2020): 107562-107582. [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).
8. Kavitha, M., et al. "Heart disease prediction using hybrid machine learning model." *2021 6th international conference on inventive computation technologies (ICICT)*. IEEE, 2021. [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.