

# HEART DISEASE PREDICTION

Anderson Victor.A<sup>1</sup>, Bharat Natrayn.G<sup>2</sup>, Dhivahar.M<sup>3</sup>, Sugania.S. S<sup>4</sup>

<sup>1,2,3</sup>Final Year Student, Dept of Computer Science and Engineering, Jeppiaar SRR Engineering College, Chennai.

<sup>4</sup>Assistant Professor, Dept. of Computer Science and Engineering, Jeppiaar SRR Engineering College, Chennai.

\*\*\*

**Abstract** - The proposed project consists of a web-based application that would allow individuals to obtain the predicted potential of attaining such CVDs during their lifetime through a machine-learning model. The proposal consists of 5 models which individually test the input given by the user to identify any possible likelihood that they could attain such conditions. The models along with the accuracy of themselves are displayed to the user after they provide the necessary details that are required by the models to calculate the probability.

The machine learning models use a dataset of about 350 individual records of real-world patients to calculate this probability. The models select 14 attributes such as age, sex, chest pain, resting blood pressure, resting blood sugar, serum cholesterol, etc., and compare that data to whether that individual has attained any form CVD at any point during their lifetime. Different models provide different accuracies based on how they are trained. However, as the number of users rise and each user adds their individual data to the model, the accuracy of the proposed model will ultimately increase and provide users with dependable observations and allow cardiologists to adapt their techniques to better suit the condition of the person being diagnosed.

**Key Words:** Cardio-Vascular Diseases, Machine Learning Algorithms, Logistic Regression, Random-forest classifier, Sci-kit Learn, pre-processing, HTML & CSS

## 1. INTRODUCTION

In the modern society, the rise of industrialization and capitalism has changed the lifestyle of an individual to that in which severe medical conditions such as obesity, high and low blood pressure, tachycardia, high cholesterol and diabetes have normalized to such an extent where each and every person past the age of 40 have been diagnosed with at least one of the aforementioned conditions. If necessary-initiatives are not taken within the next few years, the world would be faced with the risk of doubling the number of cases diagnosed with the same each year.

The proposed project consists of a web-based application that would allow individuals to obtain the predicted potential of attaining such CVDs during their lifetime through a machine-learning model. Machine Learning involves building programs using certain tools to build a module that could help the computer to discover how to perform tasks without being explicitly programmed to do so.

Compared to simpler tasks which, require miniscule effort by a human to create an algorithm for, it would be more efficient to help the machine develop its own algorithm. The proposal consists of 5 models which individually test the input given by the user to identify any possible likelihood that they could attain such conditions. The models along with the accuracy of themselves are displayed to the user after they provide the necessary details that are required by the models to calculate the probability.

The machine learning models use a dataset of about 350 individual records of real-world patients to calculate this probability. The models select 14 attributes such as age, sex, chest pain, resting blood pressure, resting blood sugar, serum cholesterol, etc., and compare that data to whether that individual has attained any form CVD at any point during their lifetime. Different models provide different accuracies based on how they are trained. However, as the number of users rise and each user adds their individual data to the model, the accuracy of the proposed model will ultimately increase and provide users with dependable observations and allow cardiologists to adapt their techniques to better suit the condition of the person being diagnosed.

## 2. RELATED WORKS

Sneha Grampurohit and Chetan Sagarnal have developed a classifier system using machine learning algorithms is to immensely help to solve the health-related issues. A Sample dataset of about 4920 patients' records diagnosed with 41 diseases was selected for analysis and then developed using machine leaning algorithms such as Decision Tree Classifier, Random forest Classifier and Naïve Bayes classifier. Diseases and health related problems like malaria, dengue, Impetigo, Diabetes, Migraine, Jaundice, Chicken Pox etc., cause significant effect on one's health and sometimes might also lead to death if ignored. The healthcare industry can make an effective decision making by 'mining' the huge database they possess i.e., by extracting the hidden patterns and relationships in the database. Data mining algorithms like Decision Tree, Random forest and Naïve Bayes algorithms can give a remedy to this situation. Thus an automated system that can discover and extract hidden knowledge associated with the diseases from a historical (diseases-symptoms) databse according to the rule set of the respective algorithms has been developed.

Sayantan Saha and Argha Roy Chowdhuri developed a web-based Disease Detection System which aims at building up a website where user can know about the disease by submitting the symptoms that he is experiencing. The system was developed using clean technology of ID3 (Iterative Dichotomiser 3) algorithm. The algorithm analyzes the training set and based on it teaches the machine to identify disease based on symptoms. Once the disease is detected, search for corresponding remedy for the disease. The user is advised to try the remedy for about a week or two and if the symptoms persist consult a doctor.

Shadab Adam developed an Intelligent System using a specific data mining technique, namely, Naive Bayes. It is implemented as a web-based application in which the user answers the predefined questions. It retrieves the data from the stored database and compares the user values with trained data set. It is able to answer complex queries for diagnosing heart disease. Hence, it can assist healthcare practitioners to make intelligent clinical decisions which traditional decision support systems cannot. By providing effective solutions, it can also help to reduce treatment costs.

### 3. PROPOSED SYSTEM

Cardio Vascular Diseases (CVDs) are the most common cause of death in the modern society constituting to about 35% of total deaths. People underestimate the value of health in the modern society. This proposal aims in the production of a successfully built web application that would allow people to obtain a reliable source to predict whether they are at any risk to any CVDs during their lifetime. The browser provides the user with predictions from a total of 5 different machine learning models which are logistic regression classifier, naïve bayes classifier, decision tree classifier, SVC classifier and random forest classifier.

#### 3.1 Advantages of Proposed System

1. The existing system focuses on identification of diseases whereas the proposed system focuses on the prediction of CVDs in patients.
2. The ability of the existing system ceases with the determining of the disease present in the diagnosed patient at that particular time.
3. There exists no application where a random end-user could identify the risk of themselves being diagnosed with a Cardio-Vascular disease (CVDs).

### 4. SYSTEM ARCHITECTURE

The proposed system consists of a developer who is responsible for developing the machine learning model with the help of the dataset obtained from the hospital. The end user is able to access the built model from the User Interface which was made by the developer.

The dataset which is collected from the Hospital is stored in the database and is used for building the machine learning models. The model is then trained and then deployed in the web application as the user interface for the end users to interact with.

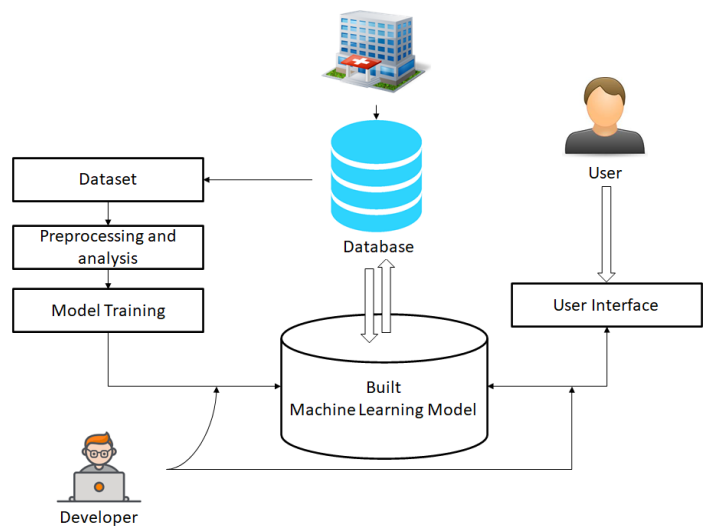


Fig-1: System Architecture

### 5. REQUIREMENTS SPECIFICATION

#### 5.1 Hardware Requirements

**Processor:** Intel ® Pentium ®

**RAM:** 4 GB and above

**Hard Disk:** 120 GB

**Speed:** 1.6 GHz and above

**Monitor:** 15" LED SVGA

**Input Devices:** Keyboard, Mouse

#### 5.2 Software Requirements

**Operating System:** Windows 8/8.1/10

**Coding Language:** Python, HTML & CSS

**Platform:** Chrome v50 and above

**Special Tools:** Visual Studio Code, Jupyter Notebooks

### 6. MODULES AND DESCRIPTION

#### 6.1 Recognizing Libraries and Algorithms

The application includes several libraries such as sci-kit learn and numpy which are python libraries that are involved in data processing. The sci-kit learn (sklearn) library consists of

several machine learning algorithms such as logistic regression classifier, KNN classifier, Random Forest Classifier, etc. Numpy is a library for python, adding support for large, multi-dimensional arrays and matrices, along with a collection of high-level mathematical functions to operate on these arrays.

## 6.2 Data Pre-Processing

Initially, before the data can be used to build a machine learning model, it has to be pre-processed to obtain a dataset which does not contain any null and missing values. The model which will be built is capable of producing an accurate prediction using the machine learning algorithms such as Random Forest classifier, Decision Tree classifier, Naïve Bayes classifier, Logistic Regression Classifier, etc. The data is processed to obtain histograms, correlation graphs, bar graphs, etc to produce detailed analysis of the dataset that we possess.

## 6.3 Model Implementation and Prediction

The implementation model consists of a collection of components as well as the implementation subsystems that house them. Both deliverable components, such as executables, and components from which deliverables are made, such as source code files, are considered components.

The following are the phases in the modelling process:

1. Examine the problem. We must first thoroughly investigate the situation in order to pinpoint the problem and fully comprehend its central issues

2. Construct a model

3. Complete the model

4. Verify and interpret the solution of the model

5. Prepare a report on the model

6. Keep the model going

## 6.4 Front End Deployment in HTML and CSS

Front end development is the process of creating a web application using HTML and CSS languages which would be favorable for a user to interact with it and understand the contents of the webpage. HTML stands for the Hyper Text Markup Language which is used to convert text into images, tables, links and other representations. CSS stands for Cascading Style Sheets which is a programming language that governs the appearance of the webpage.

## 6.5 Integration of Frontend and the model

The integration of the web application created using HTML needs to be integrated with the model generated previously

to obtain an application which could successfully predict the presence of heart disease in the individual. Using frameworks such as Flask and Django, it is successfully integrated to produce an application which would change from page to page as the user navigates through it. The frameworks are written in python allowing an easier user-friendly approach towards integration. They are highly compatible with modern technologies comparatively.

## 6.6 Model Deployment

Deployment is the process of integrating a machine learning model into an established production system in order to make data-driven business decisions. It's one of the last steps in the machine learning process, and it's also one of the most time-consuming. Although the application could be run on the home server, it can also be deployed using modern Cloud Service Providers such as Amazon Web Services (AWS), Google Cloud Platform and Microsoft Azure.

## 7. RESULTS

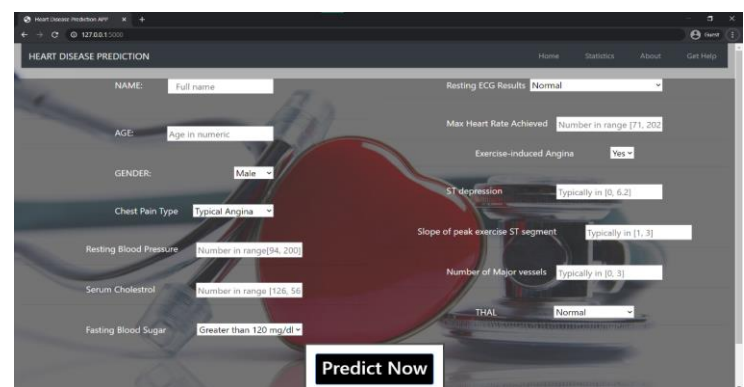


Fig 2- Deployed Home Page

The dataset obtained from the hospital is pre-processed and analyzed which is later developed into a machine learning model using python libraries such as sci-kit learn which has inbuilt functions for machine learning classifiers. The model is then deployed using Flask framework for deployment of python-based applications.

## 8. CONCLUSION

Records show that about 3 million deaths occur annually in India due to Cardio-Vascular Diseases. According to the WHO, heart diseases are the major cause of death in Indians accounting for about 35%. Several millions of dollars have been spent cumulatively by various leading companies to solve this issue. Our goal is to provide every citizen with the accessibility to identify risks, if there are any, at an earlier stage so that they may be well prepared.

**REFERENCES**

- [1] Sneha Grampurohit and Chetan Sagarnal, "Disease Prediction using Machine Learning Algorithms", 2020 International Conference for Emerging Technology (INCET), DOI: 10.1109/INCET49848.2020.9154130.
- [2] Sayantan Saha, Argha Roy Chowdhuri et,al "Web Based Disease Detection System",IJERT, ISSN:22780181,Vol.2 Issue 4, April-2013
- [3] Shadab Adam et.al "Prediction system for Heart Disease using Naïve Bayes", International Journal of advanced Computer and Mathematical Sciences, ISSN 2230- 9624, Vol 3,Issue 3,2012,pp 290- 294[Accepted- 12/06/2012].
- [4] Min Chen, Yixue Hao et.al "Disease Prediction by Machine Learning over big data from Healthcare Communities", IEEE[Access 2017]
- [5] Mr Chintan Shah,Dr. Anjali Jivani, "Comparison Of Data Mining Classification Algorithms for Breast Cancer Prediction", IEEE-31661
- [6] Qulan, J.R. 1986. "Induction of Decision Trees". Mach.Learn. 1,1 (Mar. 1986),81-10
- [7] Palli Suryachandra, Prof.Venkata Subba Reddy, "Comparison of Machine Learning algorithms For Breast Cancer", IEEE.
- [8] Andrew Alikberov, Stephan Broadly et.al "The Learning Machine", Accessed on: March 26,2020. [Online].