CHAPTER 1

INTRODUCTION

The heart is one of the main organs of the human body. It pumps blood through the blood vessels of the circulatory system. The circulatory system is extremely important because it transports blood, oxygen and other materials to the different organs of the body. Heart plays the most crucial role in circulatory system. If the heart does not function properly then it will lead to serious health conditions including death. Heart disease has created a lot of serious concerned among researches; one of the major challenges in heart disease is correct detection and finding presence of it inside a human. Early techniques have not been so much efficient in finding it even medical professor are not so much efficient enough in predicating the heart disease. There are various medical instruments available in the market for predicting heart disease there are two major problems in them, the first one is that they are very much expensive and second one is that they are not efficiently able to calculate the chance of heart disease in human. Heart disease refers to issues and deformities in the structure or in the functioning of the heart or of the blood vessels supplying the heart, that damages its usual operations. It is a stern condition with elevated commonness, about 3-5% of the people who are being admitted to the hospital are associated with heart failure occurrences. Up to 2% of the total health expenses in several countries is being spent for treating the patients associated with heart diseases. The standard of services in medical field infers diagnosing patients correctly and administering treatments which are successful. Faulty infirmary conclusions may lead to unpleasant mishaps which are therefore intolerable. Establishing a probable illness administration structure needs inspection of huge volume of figures, advance prognosis of the illness, examining the gravity and advance forecast of untimely incidents. Large bulk of data in the area of medical science is being saved each day in various digital forms such as Electronic Health Records(EHRs) and registers. The quantity of information in the field of medicine is huge and stored in computers as well. These figures are not deployed or brought into effective use. This data collected if scrutinized and studied could be utilized in predicting of diseases or preventing them.

1.1 Motivation

A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and/or decision support systems. Most hospitals today employ some sort of hospital information systems to manage their healthcare or patient data. These systems typically generate huge amounts of data which take the form of numbers, text, charts and images. Unfortunately, these data are rarely used to support clinical decision making. There is a wealth of hidden information in these data that is largely untapped. This raises an important question: "How can we turn data into useful information that can enable healthcare practitioners to make intelligent clinical decisions?" This is the main motivation for this research [9].

1.2 Objective of our Project

The main objective of this research is to develop a prototype Intelligent Heart Disease Prediction System (IHDPS) using three data mining modeling techniques, namely, Decision Trees, Naïve Bayes and Neural Network. IHDPS can discover and extract hidden knowledge (patterns and relationships) associated with heart disease from a historical heart disease database. It can answer complex queries for diagnosing heart disease and thus assist healthcare practitioners to make intelligent clinical decisions which traditional decision support systems cannot. By providing effective treatments, it also helps to reduce treatment costs. To enhance visualization and ease of interpretation, it displays the results both in tabular and graphical forms.

1.3 Overview of Problem

Prediction and diagnosing of heart disease has become a challenging factor faced by doctors and hospitals both in India and abroad. In order to reduce the large scale of deaths from heart diseases, a quick and efficient detection technique is to be discovered. Development of software to be used for practical applications rather than only in research field. Doctors need not only to know the main symptoms in a patient but also the other factors in life and surrounding that lead to disease in the patient.

1.4 Need for the proposed system

Currently Heart disease prediction system predicts the likelihood of patients getting heart disease. There is a need for more accurate predictions and also to classify the type of heart disease. This helps doctors to easily know the type of heart disease. Thus can provide required medical help to the patients.

1.5 Advantages of Proposed System

- Early prediction of heart diseases can be done.
- The cost of medication will be minimized.
- Accuracy rate will be high.
- Reduce the large scale of death caused from heart diseases
- Helps to take accurate decision for doctors to diagnose the patient

CHAPTER 2

LITERATURE SURVEY

"Review on Heart Disease Prediction using Ann and Classifier" - In the proposed system we use the Artificial Intelligence. The Artificial Neural Network i.e. ANN is use to predict the heart disease by decision condition. In the existing system the main disadvantage is that it gives low accuracy in the results. Hence for increasing the accuracy we use the ANN. For normalizing the data set Data mining classifier is used. From the analysis it is concluded that artificial neural network algorithm is best for classification of knowledge data from large amount of medical data [1].

"Classification of Heart Disease using Artificial Neural Network and Feature Subset Selection" - In this paper we have proposed a new feature selection method for heart disease classification using ANN and various feature selection methods for Andhra Pradesh Population. We applied different feature selection methods to rank the attributes which contribute more towards classification of heart disease, which indirectly reduces the no. of diagnosis tests to be taken by a patient. Our experimental results indicate that on an average with ANN and feature subset selection provides on the average better classification accuracy and dimensionality reduction [2].

"A Review on Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques" - This paper summarises some of the recent works done in data mining related to cardiovascular diseases. Careful selection of the combination of mining techniques and accurate implementation of those techniques on the data set yields a fast and effective implementation of a system for heart disease management. This paper summarizes some of the current research on predicting heart diseases using data mining techniques, analyse the various combinations of mining algorithms used and conclude which technique(s) are effective and efficient. Study on various data cleaning and pruning techniques that

prepare and make a dataset suitable for mining [3].

"Prediction Of Heart Disease Using Artificial Neural Network" In this paper, we enlighten the number of techniques in Artificial Neural Network (ANN). The accuracy is calculated and visualized such as ANN gives 94.7% but with Principle Component Analysis (PCA) accuracy rate improve to 97.7%.The model of the KDD which involves the collection of data sets, its preprocessing, building patterning matching classifier model with ANN. After training them we will ensemble them to get the better decision we use Weka tool for the implementation of the methodology. Accuracy can be further increased by changing the setting and making them more optimized according to each algorithm and nature of data [4].

"A survey on Prediction of Heart Disease using Machine Learning Algorithms" - This survey paper will discusses about various machine learning algorithm for heart disease and their relative comparison on the various parameter. It also shows future prospectus of machine learning algorithm in heart disease. This paper also does a deep analysis on utilization of deep learning in field of predicting heart disease. Algorithms such as Naive Bayes, Decision Tree, KNN, Neural Network, are used. An analytical comparison has been done for finding out best available algorithm for medical dataset. To carry forward the work of temporal medical dataset, where dataset varies with time and retraining of dataset is required [5].

"Intelligent Heart Disease Prediction System Using Data Mining Techniques" - In this paper a prototype heart disease prediction system is developed using three data mining classification modeling techniques. The system extracts hidden knowledge from a historical heart disease database. DMX query language and functions are used to build and access the models. The models are trained and validated against a test dataset. Lift Chart and Classification Matrix methods are used to evaluate the effectiveness of the models. The most effective model to predict patients with heart disease appears to be Naïve Bayes followed by Neural Network and Decision Trees. It can also incorporate other data mining techniques, e.g., Time Series, Clustering and Association Rules. Continuous data can also be used instead of just categorical data [6].

"A Hybrid Approach for Heart Disease Prediction Using Artificial Neural Network and K-means" - The data is classified according to their properties. This classification is

implemented by developing a model using Artificial Neural Network and K-means algorithm. The focus of our model is to classify the data according to cardiovascular diseases better so as to have more reliable diagnosis. The hybrid approach gives higher accuracy rate of about 93.52% than earlier proposed method [7].

"Review of Heart Disease Prediction System Using Data Mining and Hybrid Intelligent Techniques" - In this paper Heart attack prediction system methodology is categorized in three types. At first type data mining technique (mainly classification technique) are analyzed. The second type intelligent techniques used for heart disease prediction are analyzed. The final type the role of feature subset in the heart disease prediction is discussed. From the analysis it is concluded that, data mining plays a major role in heart disease classification. The classification accuracy can be improved by reduction in features [8].

"Decision Support System for Heart Disease Based on Support Vector Machine and Artificial Neural Network" presented a decision support system for heart disease classification. Support vector machine (SVM) and artificial neural network (ANN) were the two main methods used in this system. A multilayer perceptron neural network (MLPNN) with three layers was employed to develop a decision support system for the diagnosis of heart disease. This multilayer perceptron neural network was trained by back-propagation algorithm which is computationally an efficient method. Results showed that a MLPNN with back-propagation technique can be successfully used for diagnosing heart disease[9]

CHAPTER 3

SYSTEM REQUIREMENT SPECIFICATION

Requirement Specification (RS) is a description of a software and hardware system to be developed, laying out functional and non-functional requirements, and may include a set of use cases that describe interactions the users will have with the software.

3.1 Software Requirement and Specification

Software Requirements Specification (SRS) provides an overview of the entire SRS with purpose, scope, definitions, acronyms, abbreviations, references and overview of the SRS. A software requirements specification (SRS) is a comprehensive description of the intended purpose and environment for software under development. The SRS fully describes what the software will do and how it will be expected to perform the various gestures and determining its accuracy. The SRS is a requirements specification for a software system, is a description of the behavior of a system to be developed and may include a set of use cases that describe interactions the users will have with the software. In addition it also contains non-functional requirements. Nonfunctional requirements impose constraints on the design or implementation. Software requirements specification establishes the basis for agreement between customers and contractors or suppliers on what the software product is to do as well as what it is not expected to do. Software requirements specification permits a rigorous assessment of requirements before design can begin and reduces later redesign. It should

also provide a realistic basis for estimating product costs, risks, and schedules. An SRS minimizes the time and effort required by developers to achieve desired goals and also minimizes the development cost. A good SRS defines how an application will interact with system hardware, other programs and human users in a wide variety of real-world situations. Parameters such as operating speed, response time, availability, portability, maintainability, footprint, security and speed of recovery from adverse events are evaluate.

3.2 FUNCTIONAL REQUIREMENTS

A functional requirement defines a function of a software system or its components. A function is described as a set of inputs, the behavior, and outputs. Functional requirements may be calculations, technical details, data manipulation and processing and other specific functionality that define what a system is supposed to accomplish. Behavioral requirements describing all the cases where the system uses the functional requirements are captured in use cases. Functional requirements are supported by non-functional requirements (also known as quality requirements), which impose constraints on the design or implementation (such as performance requirements, security, or reliability).

3.3 NON FUNCTIONAL REQUIREMENTS

In systems engineering and requirements engineering, a non-functional requirement is a requirement that specifies criteria that can be used to judge the operation of a system, rather than specific behaviors. They are contrasted with functional requirements that define specific behavior or functions. Non-functional requirements define how a system is supposed to be. Non-functional requirements are in the form of "system shall be <requirement>", an overall property of the system as a whole or of a particular aspect and not a specific function. The system's overall properties commonly mark the difference between whether the development project has succeeded or failed. Non-functional requirements are often called "quality attributes" of a system.

3.4 HARDWARE REQUIREMENTS:

- PROCESSOR: Pentium 4 and higher (1.2GHZ and higher)
- RAM: 4GB and higher
- HARD DISK: 20GB or higher.
- Monitor: 15”CRT or LCD monitor
- Keyboard: normal
- Mouse: compatible mouse

3.5 SOFTWARE REQUIREMENTS:

- Operating system: Windows 8(and higher).
- Coding Language: Python, JavaScript, HTML,CSS
- Python Libraries: pandas, numpy, matplotlib, seaborn, itertools, plotly, pprint

CHAPTER 4

SYSTEM ANALYSIS

Large systems are always decomposed into modules or subsystem that provides some related set of services. The initial design process for identifying these modules and establishing framework for modules or subsystem control and communication is called architectural design and output of this design process is a description of the software architecture.

4.1 Existing System

In the existing system there is less accuracy in predicting the heart disease and there is no classification of its type. So there is a need for effective predictive system that also classifies the type of heart disease with more accuracy.

4.2 Proposed System

In this section we have mentioned about the system architecture and its modules are explained below. The Fig.4.1 represents the overview of systems architecture with the core modules that are shown in the figure.

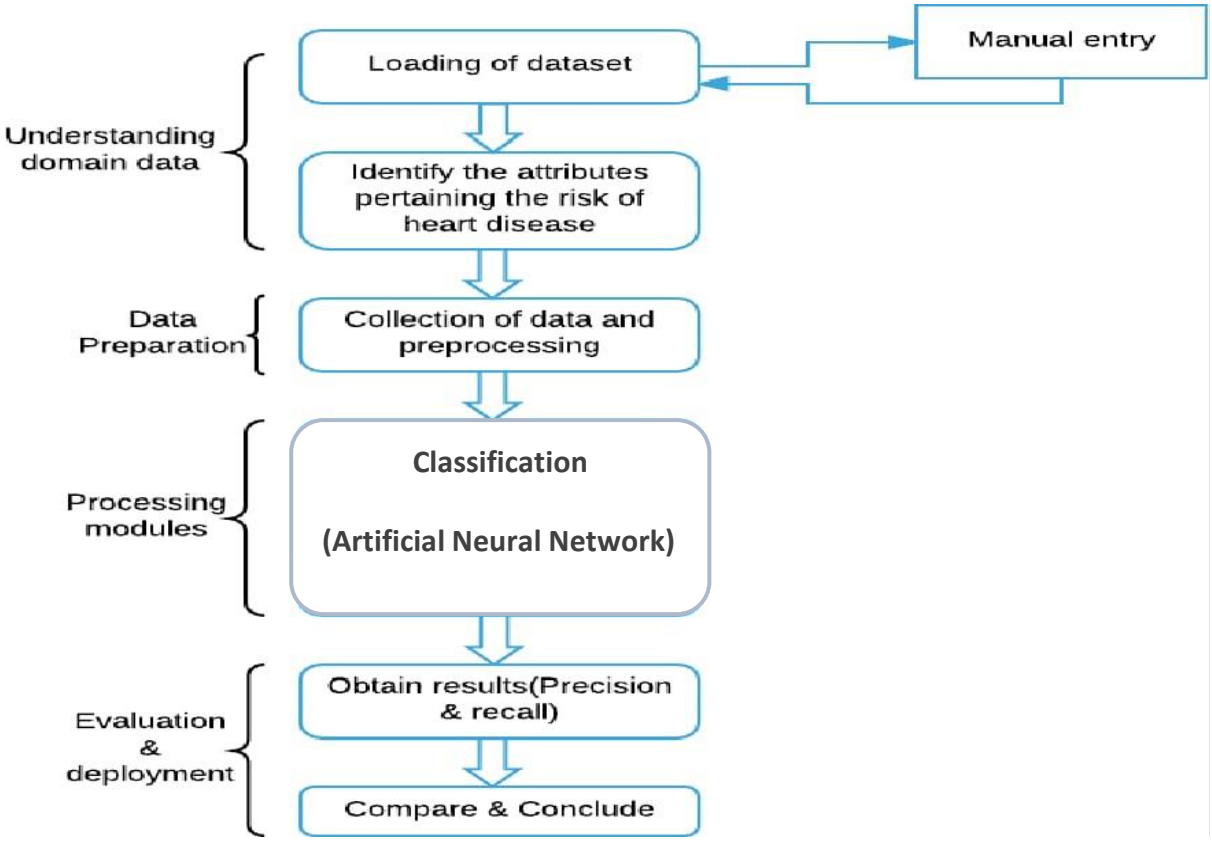


Fig.4.1. Heart Disease Prediction System Architecture

The core modules of the proposed system consist of:

- Understanding Domain Dataset: Cleveland dataset is provided as input to system, detail about the Cleveland dataset is discussed in section V. We also provided feature of loading excel dataset or manual entries into the system.
- Data Preparation: When data is manipulated in a way that it can be suitable for future examination is called data preparation.
- Module: It talks about the algorithmic approach applied over the system in order to obtain high accuracy result. As an algorithmic approach we use ANN classifier and K-means clustering data mining techniques
- Evaluation and deployment: Final Analysing modules states information related to

generated output. We obtain a confusion matrix, our system compares and conclude about measurable resultant are facts like sensitivity, accuracy, decision support etc.

4.3 Classifier

Classification means the procedure of arranging the data into different groups or classes according to their similarities. In our approach we use ANN classifier on various heart diseases attribute on the basis of common characteristics which can either be descriptive such as age, sex, cholesterol, fps etc. Whenever using statistics and the areas where huge amounts of data are collected and analyzed, it's important to sort the info points into sub-groups often. This can be a very hard task for human beings, who often arent able to recognize which class a data point belongs to due to large amounts of data present in each data point. Instead, a digital classier is used.

4.3.1 ANN Algorithm:

Neural networks resembles biological network of brain neurons. Similar to central nervous systems, an interconnected group of neurons forms neural network. Each node gets input from other nodes and weights between nodes adjust so that the whole network learns to compute the output. There are various types of neural networks structures with each having its own learning algorithm. The neural network is a novel computer architecture compared to traditional computers. An artificial network of neurons connected with each other to give a specified output on applying input is called neural network.

Types of ANNs

- Single Layer Perceptron
- Multilayer Perceptron (MLPs)
- Radial-Basis Function Networks (RBFs)
- Hopfield Network
- Boltzmann Machine
- Self-Organization Map (SOM)
- Modular Networks (Committee Machines)

In our approach we used Multilayer Perceptron. It is used to map input into outputs with the help of input layer, various hidden layers and output layer. In order to get optimal solution, it makes use of trial and error method. The design of neural network is shown in Fig 4.2 below.

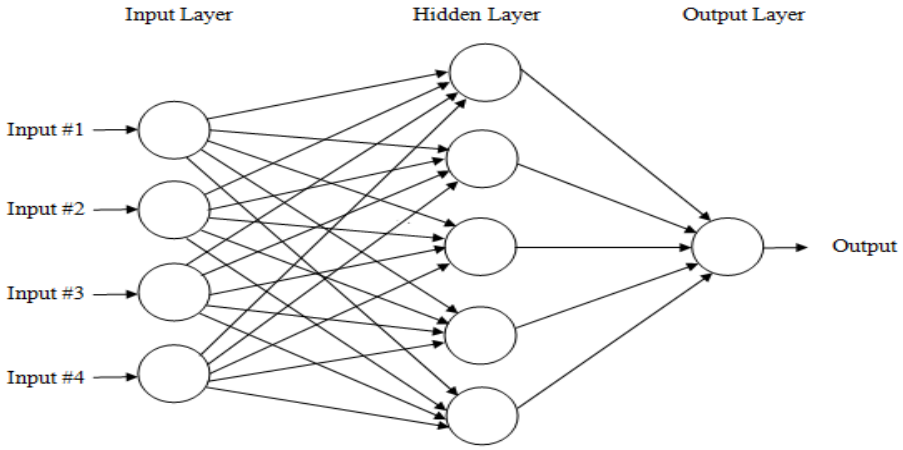


Fig.4.2. Layers of ANN

Pseudo code

STEP 1: Input nodes i, given input x_i:

For each input node i

$output_i = x_i$

STEP 2: Hidden layer nodes j

For each hidden neuron j

$output_j = \sum_i phi(w_{ji}.output_i)$

STEP 3: Output layer neurons k

For each output neuron k

$output_k = \sum_k phi(w_{ij}.output_j)$

phi – activation function

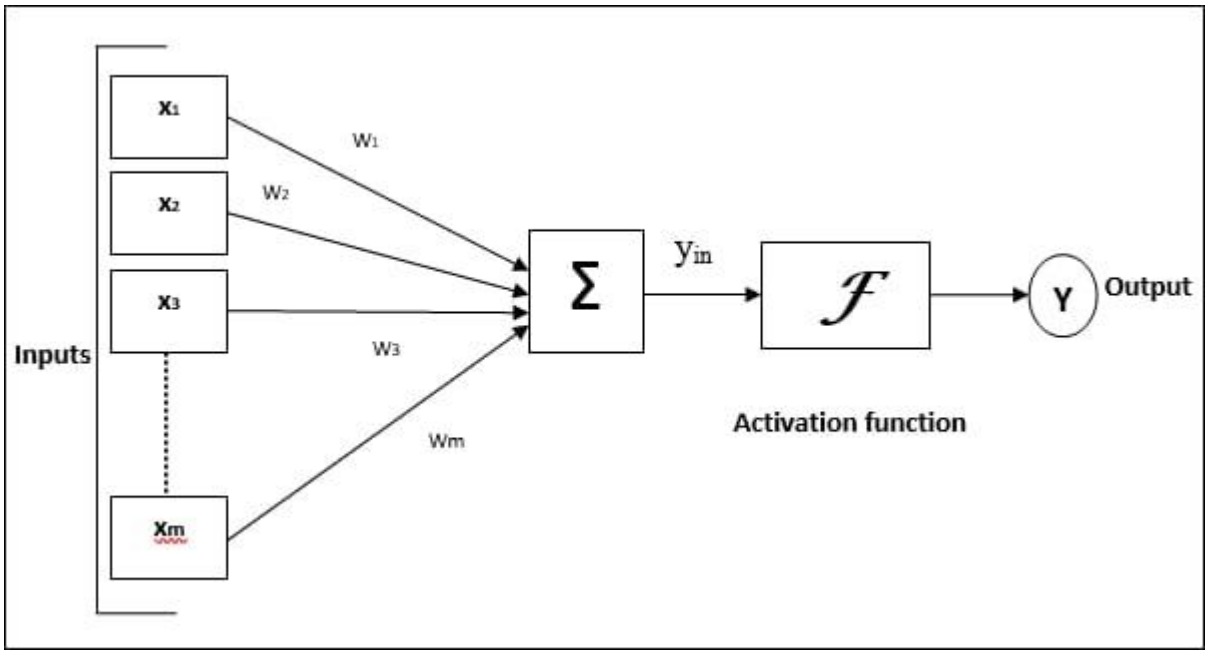


Fig.4.3. Graphical Representation Of Neuron In ANN.

Where, The output is calculated by below function.

$$y_j = \sum_{f=1}^{k} w_{ij} x_i$$

Y_j represents output neuron.

X_i is input neuron

W_{ij} is the weight connecting X_i and Y_j

$\sum$ is sigmoidal function

Input: A vector of real values (attribute, value) / obtained clusters from K-means.

Output: Classified classed with associated weight Activation function: sigmoidal function

Complexity: depends on the structure and number of layers in a neural network.

In our case in order to obtain optimal solution: O(2^n) to converge to optimal solution. Whereas below equation gives a Complexity solution to calculate approximate solution.

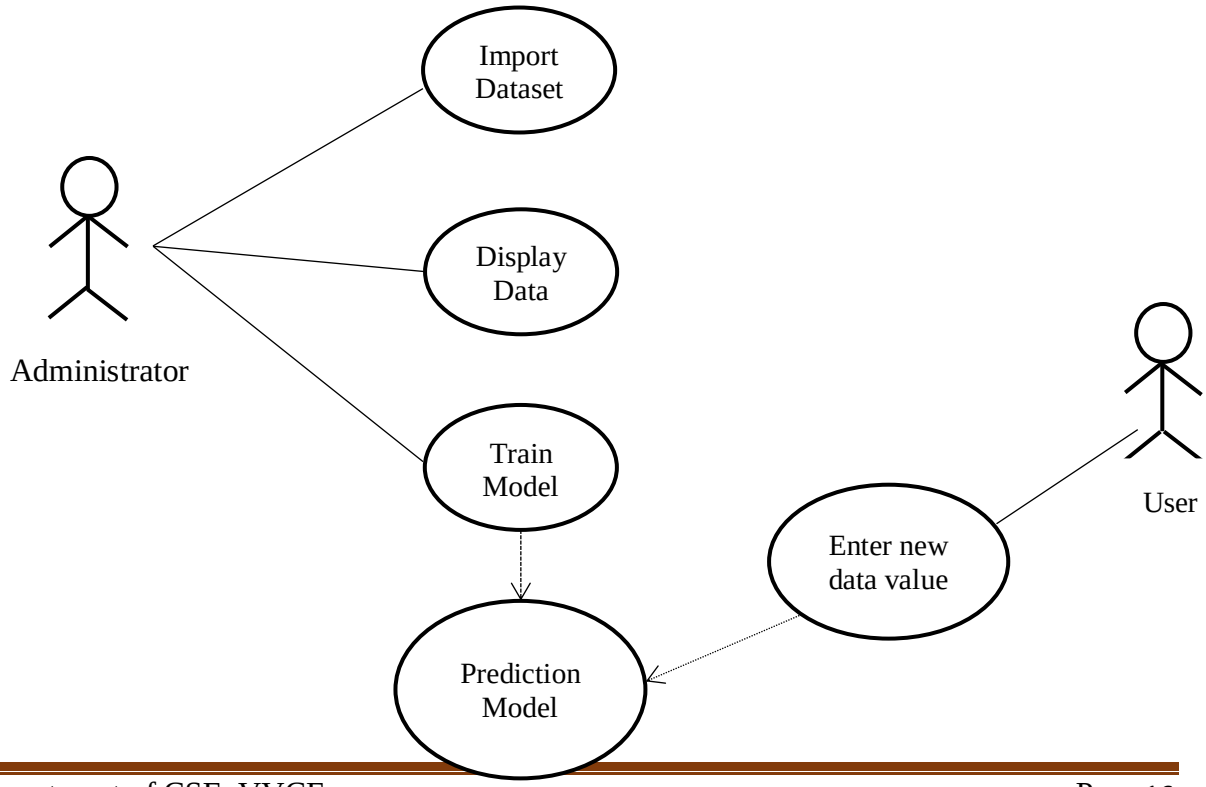
O (#epochs * # examples * #features * #neurons) => approximate solution.

CHAPTER 5

SYSTEM DESIGN

Systems design is the process of defining the architecture, components, modules, interfaces, and data for a system to satisfy specified requirements. Systems design could be seen as the application of systems theory to product development.

5.1 Use Case Diagram



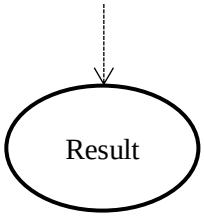


Fig.5.1. Use Case Diagram

A use case diagram is a dynamic or behavior diagram in UML. Use case diagrams model the functionality of a system using actors and use cases. Use cases are a set of actions, services, and functions that the system needs to perform. In this context, a "system" is something being developed or operated, such as a web site. The "actors" are people or entities operating under defined roles within the system. Use case diagrams are valuable for visualizing the functional requirements of a system that will translate into design choices and development priorities. They provide a good high level analysis from outside the system. Use case diagrams specify how the system interacts with actors without worrying about the details of how that functionality is implemented.

5.2 Data Flow Diagram

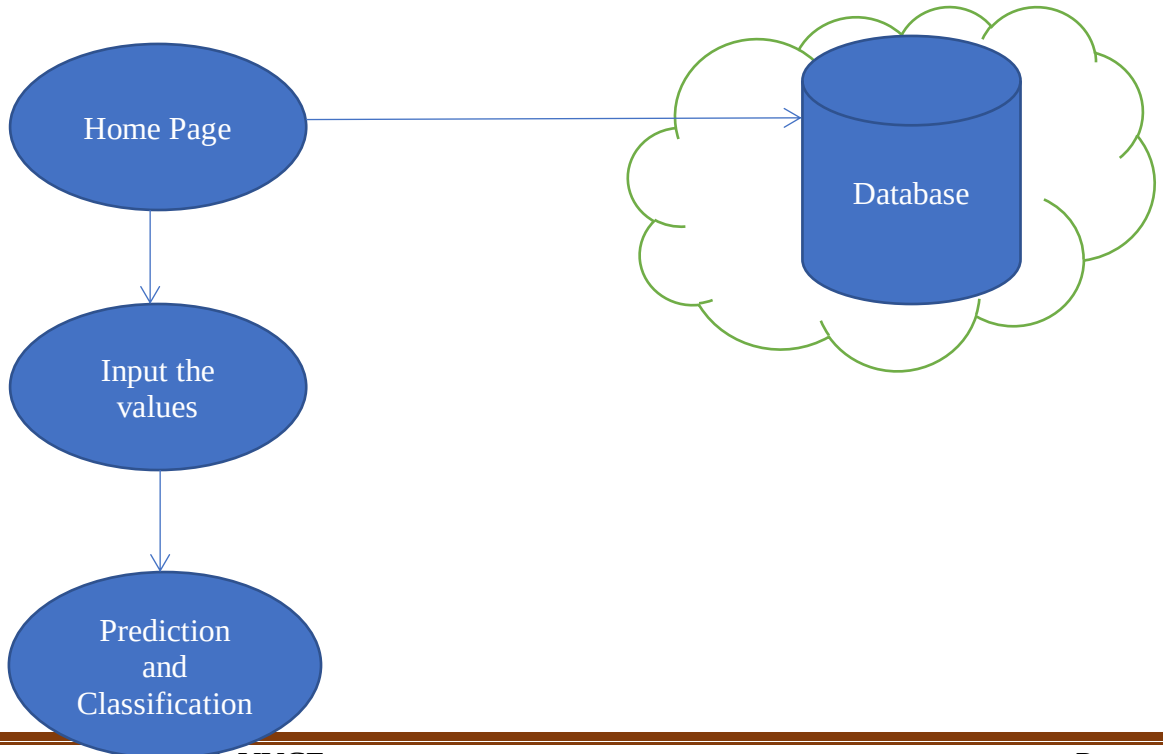


Fig.5.2. Data Flow Diagram

A data flow diagram (DFD) maps out the flow of information for any process or system. It uses defined symbols like rectangles, circles and arrows, plus short text labels, to show data inputs, outputs, storage points and the routes between each destination. Data flowcharts can range from simple, even hand-drawn process overviews, to in-depth, multi-level DFDs that dig progressively deeper into how the data is handled. They can be used to analyze an existing system or model a new one.

5.3 Sequence Diagram

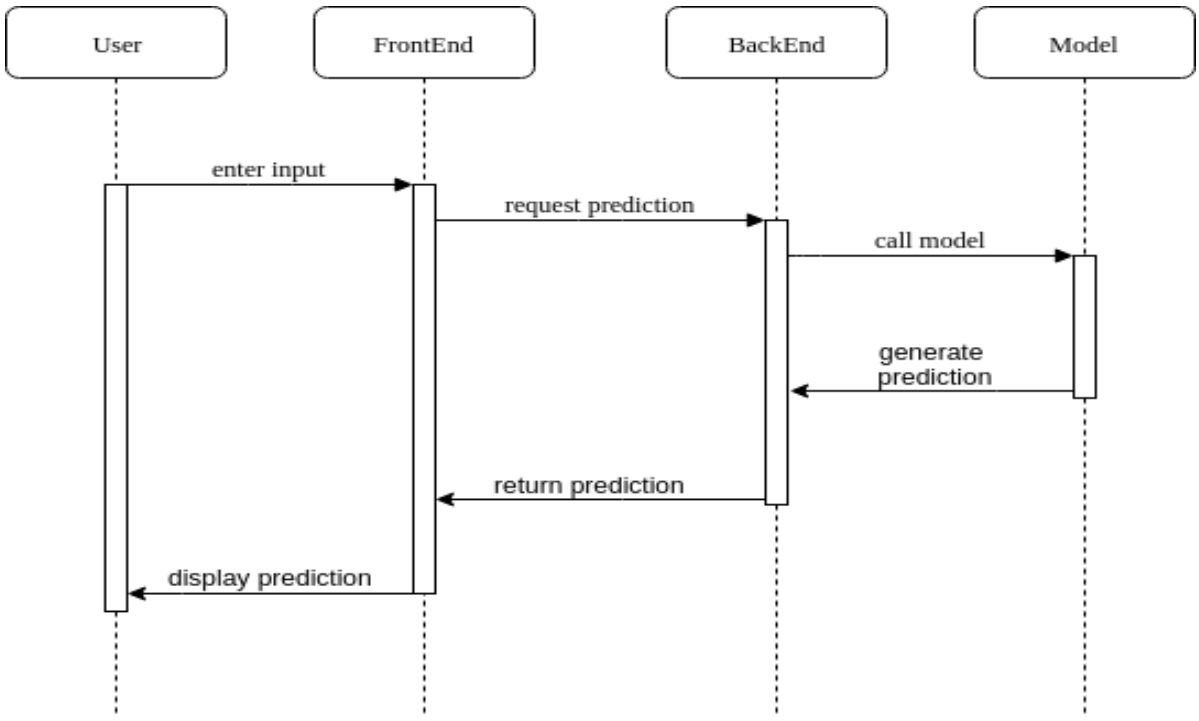


Fig.5.3. Sequence Diagram

The above diagram shows object interactions in chronological order. It depicts the sequence of messages exchanged between objects which are needed to carry out the functionality in the shown scenario. For the use case of generate a prediction, the above diagram shows the sequence of steps and interaction between the different components of the system.

5.4 Activity Diagram

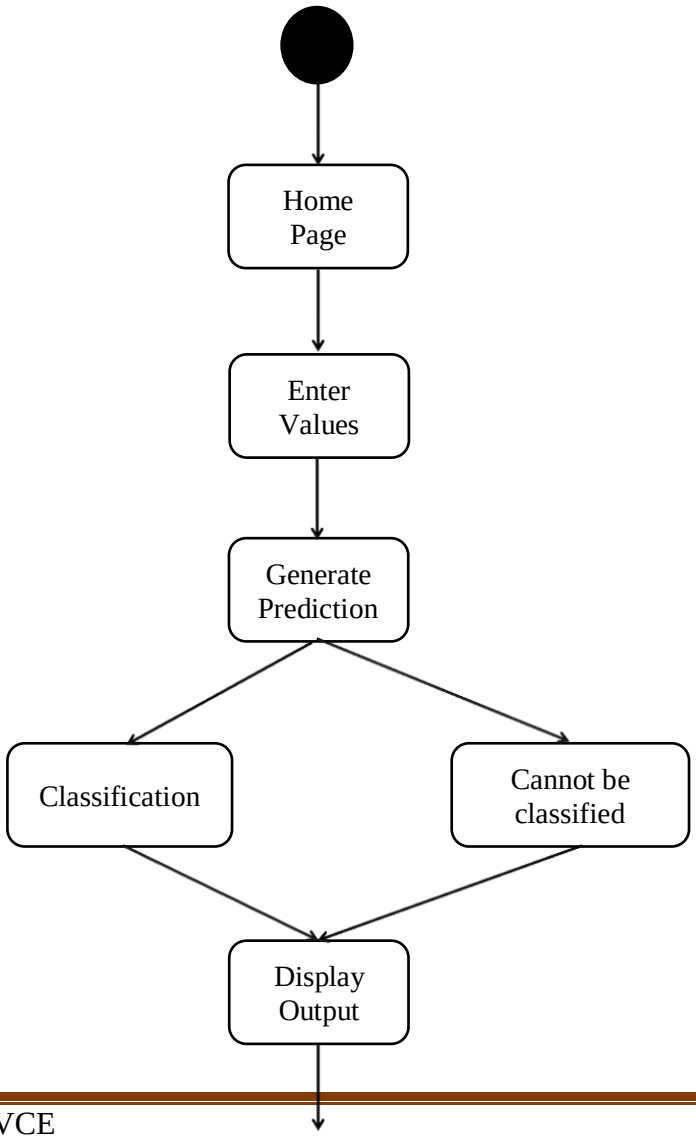


Fig.5.4. Activity Diagram

The above figure represents the activity diagram depicted using the Unified Modeling Language (UML). This diagram exhibits the dynamic aspects of the system. It depicts the flow from one activity to another.

A user can chose to enter values and check for heart disease and its type.

5.5 High Level Design

High-level design (HLD) explains the architecture that would be used for developing a software product. The architecture diagram provides an overview of an entire system, identifying the main components that would be developed for the product and their interfaces. The HLD uses possibly nontechnical to mildly technical terms that should be understandable to the administrators of the system.

5.5.1 System Architecture

A system architecture or systems architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system.

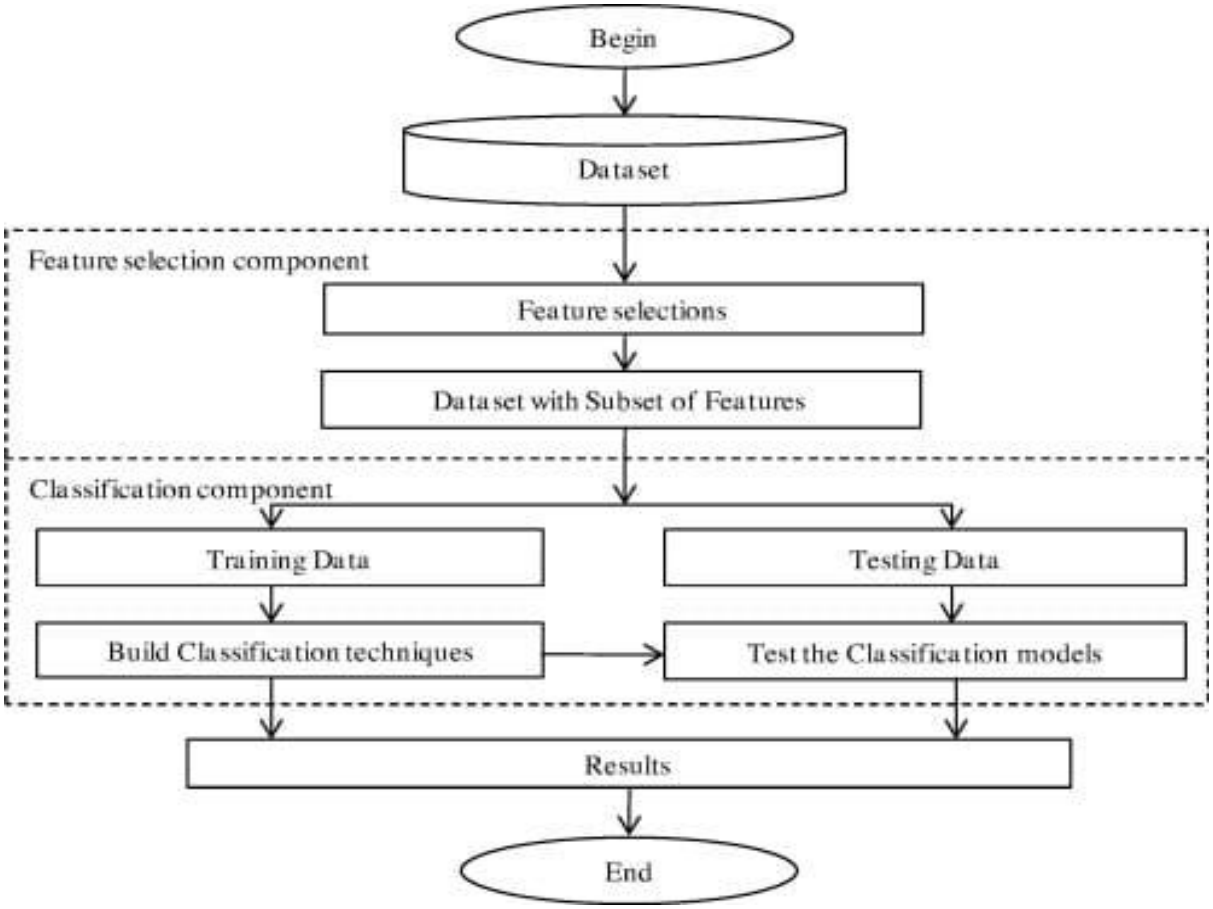


Fig.5.5. System Architecture

The system contains the following blocks:

A. As we are following the model which involves the collection of data sets, its preprocessing, building patterning matching classifier model with ANN. After training them we will ensemble them to get the better decision for the implementation of the methodology as shown in the . Figure 5.1 In first step data is preprocessed for the training using algorithm then the processed data is input in the ANN.

B. Datasets: As we are following Heart disease diagnosis medical reports data used "cleaveland heart data" is taken from UCI data repository 13 attributes are select for the research such as age (years), sex (1 = male; 0 = female), cp, tretbps (resting blood pressure in mm Hg), chol (cholesterol in mg/dl), fbs(fasting blood sugar), restecg (resting electrocardiographic results), oldpeak (ST depression induced by

exercise relative to rest), slope(the slope of the peak exercise ST segment), ca(number of major vessels), thal(reversible defect) and Num(Class attribute 1=heart disease; 0 for no heart disease)

C. Feature Subset Selection Feature subset selection is a preprocessing step commonly used in machine learning. It is effective in reducing dimensionality and removes irrelevant data thus increases learning accuracy. It refers to the problem of identifying those features that are useful in predicting class. Features can be discrete, continuous or nominal. Generally features are of three types. 1) Relevant, 2) Irrelevant, 3) Redundant. Feature selection methods wrapper and embedded models. Filter model rely on analyzing the general characteristics of data and evaluating features and will not involve any learning algorithm, where as wrapper model uses après determined learning algorithm and use learning algorithms performance on the provided features in the evaluation step to identify relevant feature. Embedded models incorporate feature selection as a part of the model training process.

D. Training Data: is labeled data used to train your machine learning algorithms and increase accuracy.

E. Testing Data: Every machine learning model needs to be tested in the real world to measure how robust its predictions are. This is data that it has never seen before. Just as a student comes across fresh problems while in an exam, models too, need to be similarly challenged so as to evaluate their performance.

F. Building classification technique using ANN: We use ANN technique for classification.

G. Test the classification Model: Here the model is tested by giving the input values and based on the results the model is verified.

H. Results: Result contains whether the person is having heart disease or not. If heart disease is present then with the help of (PHDTC) Predicting Heart Disease and Type Classification we classify the type of heart disease.

CHAPTER 6
IMPLEMENTATION

An implementation is a realization of a technical specification or algorithm as a program, software component, or other computer system through computer programming and deployment. Many implementations may exist for a given specification or standard. For example, web browsers contain implementations of World Wide Web Consortium-recommended specifications, and software development tools contain implementations of programming languages.

6.1 Data collection

Data collection is meant to be used by the patients and the probability generation to be used by doctors or patients. In Data collection the main functionality of the system is carried out. The system would direct the patients to answer a few questions in order to get necessary diagnosis. The information collected from the patient includes personal details, associated complaints and reactions experienced. The dataset consists of databases related to diagnosis of heart disease. Numeric values are assigned to attributes. The data was collected from the Cleveland Clinic Foundation (cleveland.data).

age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
47	1	2	130	253	0	1	179	0	0	2	0	2	1
35	1	1	122	192	0	1	174	0	0	2	0	2	1
58	1	1	125	220	0	1	144	0	0.4	1	4	3	1
56	1	1	130	221	0	0	163	0	0	2	0	3	1
56	1	1	120	240	0	1	169	0	0	0	0	2	1
55	0	1	132	342	0	1	166	0	1.2	2	0	2	1
41	1	1	120	157	0	1	182	0	0	2	0	2	1
38	1	2	138	175	0	1	173	0	0	2	4	2	1
38	1	2	138	175	0	1	173	0	0	2	4	2	1
67	1	0	160	286	0	0	108	1	1.5	1	3	2	0
67	1	0	120	229	0	0	129	1	2.6	1	2	3	0
62	0	0	140	268	0	0	160	0	3.6	0	2	2	0
63	1	0	130	254	0	0	147	0	1.4	1	1	3	0
53	1	0	140	203	1	0	155	1	3.1	0	0	3	0
56	1	2	130	256	1	0	142	1	0.6	1	1	1	0

Table 6.1. Dataset Structure for Prediction

A	B	C	D	E	F	G	H	I	J	K	L	M	N
age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	dpc
56	0	1	140	294	0	0	153	0	1.3	1	0	2	0
54	1	0	140	239	0	1	160	0	1.2	2	0	2	0
58	0	3	150	283	1	0	162	0	1	2	0	2	0
43	1	0	150	247	0	1	171	0	1.5	2	0	2	0
69	0	3	140	239	0	1	151	0	1.8	2	2	2	0
61	1	2	150	243	1	1	137	1	1	1	0	2	0
71	0	1	160	302	0	1	162	0	0.4	2	2	2	0
65	0	2	140	417	1	0	157	0	0.8	2	1	2	0
54	1	2	150	232	0	0	165	0	1.6	2	0	3	0
65	0	2	155	269	0	1	148	0	0.8	2	0	2	0
65	0	2	160	360	0	0	151	0	0.8	2	0	2	0
51	0	2	140	308	0	0	142	0	1.5	2	1	2	0
39	1	2	140	321	0	0	182	0	0	2	0	2	0
44	1	2	140	235	0	0	180	0	0	2	0	2	0

Table 6.2. Dataset Structure for Classification

Following are the attributes with their descriptions:

Serial No.	Attribute	Description
1	Age	In Year
2	Sex	value 1: Male value 0: Female
3	Chest pain type	value 1: typical type 1 angina value 2: typical type angina value 3: non-angina pain value 4: asymptomatic
4	Trest Blood Pressure	mmHg on admission to Hospital
5	Cholestrol	(mg/dl)

6	Fasting Blood Sugar	value 1: > 120 mg/dl value 0:< 120 mg/d
7	RestECG	resting electrographic results value 0:normal value 1: 1 having ST-T wave abnormality value 2: showing probable/definite left ventricular hypertroph
8	Slope	the slope of the peak exercise ST segment value 1: unsloping value 2: flat value 3: downsloping
9	Thalach	maximum heart rate achieved
10	Exang	exercise induced angina value 1: yes value 0: no
11	Oldpeak	ST depression induced by exercise
12	Thal	value 3: normal value 6: fixed defect value 7: reversible defect
13	Ca	number of major vessels colored by flouropsy (value 0 –3)
14	Value	Diagnosis of heart disease (angiographic disease status) Value 0: < 50% diameter narrowing Value 1: > 50% diameter narrowing

Table 6.3 List of Attributes

- Training Data

Training Data is categorized data that is mainly used to instruct machine learning algorithms and enhance precision.

- Test Data

All models of machine learning are tested to check how strong is their predictions. That is the output data which is never seen before. It is just like challenging students with new problems in exams, the models also must be challenged to measure its performance. Here 80 percent of dataset is for training and 20 percent is for testing[8].

6.2 Probability Generation

In the below Fig 4.5, the patient will enter his records and the algorithm will generate the probability depending upon the following steps.

- The values in data set will be filtered in order to get discrete values.
- Then probability of each attribute will be calculated and then the combined probability will be shown

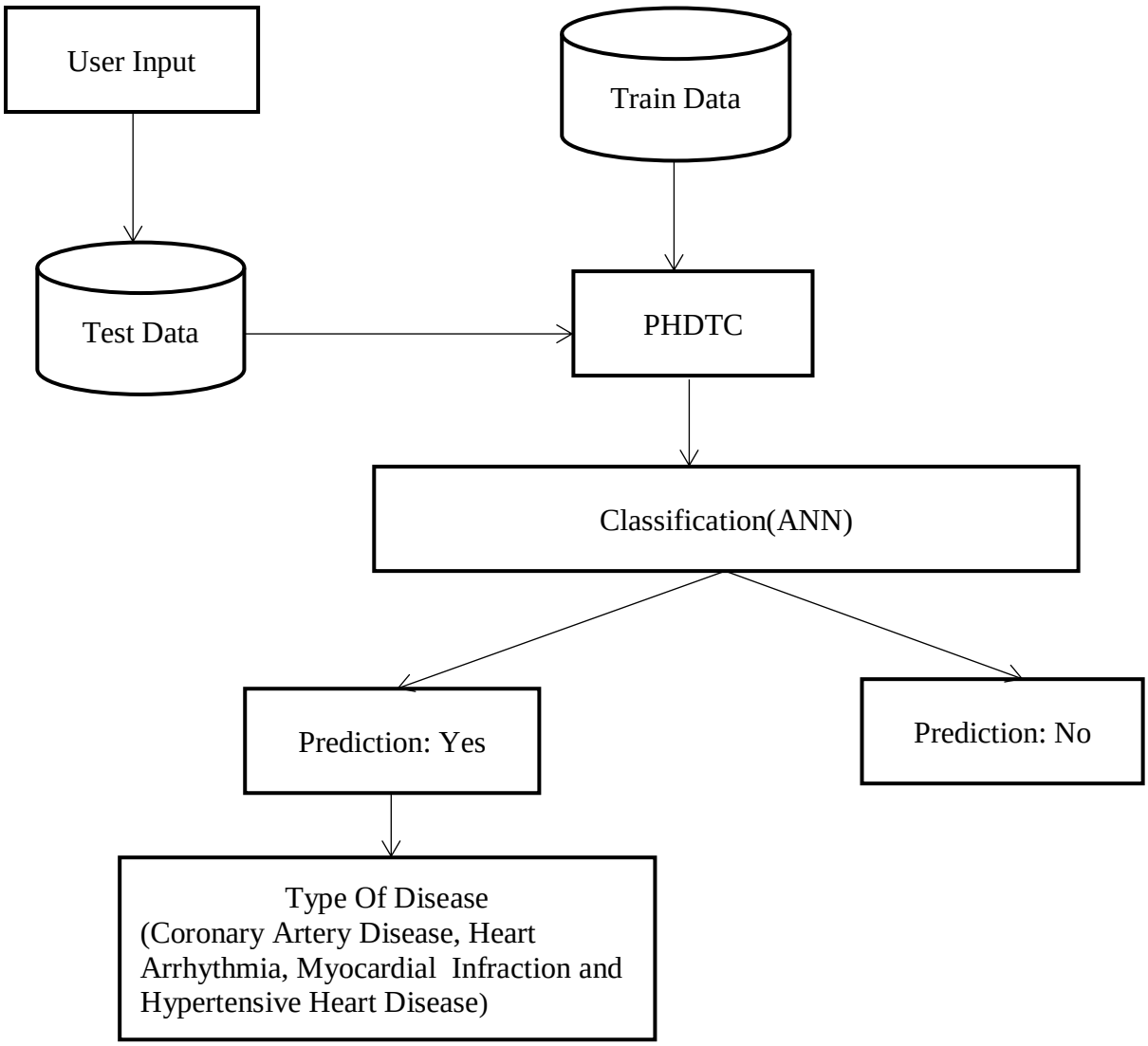


Fig 6.1. Probability Generation

After the collection of the data entered by the patient, the proposed system will generate the probability whether the patient is in risk of a heart disease or not. If present then

it classifies the type of heart disease.

The types of heart diseases are:

Coronary Artery Disease (CAD):

Coronary artery disease (CAD), also known as ischemic heart disease (IHD), involves the reduction of blood flow to the heart muscle due to build-up of plaque in the arteries of the heart. It is the most common of the cardiovascular diseases. Types include stable angina, unstable angina, myocardial infarction, and sudden cardiac death. A common symptom is chest pain or discomfort which may travel into the shoulder, arm, back, neck, or jaw. Occasionally it may feel like heartburn. Usually symptoms occur with exercise or emotional stress, last less than a few minutes, and improve with rest. Shortness of breath may also occur and sometimes no symptoms are present. In many cases, the first sign is a heart attack. Other complications include heart failure or an abnormal heartbeat.

Myocardial infarction (MI):

Myocardial infarction (MI), commonly known as a heart attack, occurs when blood flow decreases or stops to a part of the heart, causing damage to the heart muscle. The most common symptom is chest pain or discomfort which may travel into the shoulder, arm, back, neck, or jaw. Often it occurs in the center or left side of the chest and lasts for more than a few minutes. The discomfort may occasionally feel like heartburn. Other symptoms may include shortness of breath, nausea, feeling faint, a cold sweat, or feeling tired. About 30% of people have atypical symptoms. Women more often present without chest pain and instead have neck pain, arm pain, or feel tired. Among those over 75 years old, about 5% have had an MI with little or no history of symptoms. An MI may cause heart failure, an irregular heartbeat, cardiogenic shock, or cardiac arrest.

Hypertensive heart disease :

Hypertensive heart disease includes a number of complications of high blood pressure that affect the heart. While there are several definitions of hypertensive heart disease in the

medical literature, the term is most widely used in the context of the International

Classification of Diseases (ICD) coding categories. The definition includes heart failure and other cardiac complications of hypertension when a causal relationship between the heart disease and hypertension is stated or implied on the death certificate. In 2013 hypertensive heart disease resulted in 1.07 million deaths as compared with 630,000 deaths in 1990.
According to ICD-10, hypertensive heart disease, and its subcategories: hypertensive heart disease with heart failure and hypertensive heart disease without heart failure are distinguished from chronic rheumatic heart diseases other forms of heart disease (I30-I52) and ischemic heart diseases. However, since high blood pressure is a risk factor for atherosclerosis and ischemic heart disease, death rates from hypertensive heart disease provide an incomplete measure of the burden of disease due to high blood pressure.

Heart arrhythmia :

Heart arrhythmia (also known as arrhythmia, dysrhythmia or irregular heartbeat) is a group of conditions in which the heartbeat is irregular, too fast or too slow. A heart ratethat is too fast – above 100 beats per minute in adults – is called tachycardia, and a heart rate that is too slow – below 60 beats per minute – is called bradycardia. Many types of arrhythmia have no symptoms. When symptoms are present, these may include palpitations or feeling a pause between heartbeats.[1] In more serious cases, there may be lightheadedness, passing out, shortness of breath or chest pain.[1] While most types of arrhythmia are not serious, some predispose a person to complications such as stroke or heart failure. Others may result in cardiac arrest.

Cannot be predicted:

When two diseases have the same related factors, type of disease cannot be predicted.

6.3 Generating the Results

Depending upon the probability of the heart disease of the patient, the further

questions will appear. The symptoms will be entered by the patients and depending upon the probability, the results will be generated and will be display the type of disease.

6.4 About Tools

The tools which were used for implementation of the project are as follows

- Bootstrap
- Keras
- CSS
- Tensorflow Js
- Google Colab

6.4.1 Bootstrap

Bootstrap is a free and open-source CSS framework directed at responsive, mobile-first front-end web development. It contains CSS- and (optionally) JavaScript-based design templates for typography, forms, buttons, navigation and other interface components.
Bootstrap is a web framework that focuses on simplifying the development of informative web pages (as opposed to web apps). The primary purpose of adding it to a web project is to apply Bootstrap's choices of color, size, font and layout to that project. As such, the primary factor is whether the developers in charge find those choices to their liking. Once added to a project, Bootstrap provides basic style definitions for all HTML elements. The end result is a uniform appearance for prose, tables and form elements across web browsers. In addition, developers can take advantage of CSS classes defined in Bootstrap to further customize the appearance of their contents. For example, Bootstrap has provisioned for light- and dark-colored tables, page headings, more prominent pull quotes, and text with a highlight.

Bootstrap also comes with several JavaScript components in the form of jQuery plugins. They provide additional user interface elements such as dialog boxes, tooltips, and carousels.

Each Bootstrap component consists of an HTML structure, CSS declarations, and in some cases accompanying JavaScript code. They also extend the functionality of some existing interface elements, including for example an auto-complete function for input fields.

6.4.2 KERAS

Keras is an open-source neural-network library written in Python. It is capable of running on top of TensorFlow, Microsoft Cognitive Toolkit, Theano, or PlaidML.Designed to enable fast experimentation with deep neural networks, it focuses on being user-friendly, modular, and extensible. It was developed as part of the research effort of project ONEIROS (Open-ended Neuro-Electronic Intelligent Robot Operating System) and its primary author and maintainer is François Chollet, a Google engineer. Chollet also is the author of the XCeption deep neural network model.

Keras contains numerous implementations of commonly used neural-network building blocks such as layers, objectives, activation functions, optimizers, and a host of tools to make working with image and text data easier. The code is hosted on GitHub, and community support forums include the GitHub issues page, and a Slack channel.

In addition to standard neural networks, Keras has support for convolutional and recurrent neural networks. It supports other common utility layers like dropout, batch normalization, and pooling.

Keras allows users to productize deep models on smartphones (iOS and Android), on the web, or on the Java Virtual Machine. It also allows use of distributed training of deep-learning models on clusters of Graphics Processing Units (GPU) and Tensor processing units (TPU).

6.4.3 CSS

Cascading Style Sheets (CSS) is a style sheet language used for describing the presentation of a document written in a markup language like HTML. CSS is a cornerstone technology of the World Wide Web, alongside HTML and JavaScript.

CSS is designed to enable the separation of presentation and content, including layout, colors,

and fonts. This separation can improve content accessibility, provide more flexibility and

control in the specification of presentation characteristics, enable multiple web pages to share formatting by specifying the relevant CSS in a separate .css file, and reduce complexity and repetition in the structural content.

Separation of formatting and content also makes it feasible to present the same markup page in different styles for different rendering methods, such as on-screen, in print, by voice (via speech-based browser or screen reader), and on Braille-based tactile devices. CSS also has rules for alternate formatting if the content is accessed on a mobile device.

The name cascading comes from the specified priority scheme to determine which style rule applies if more than one rule matches a particular element. This cascading priority scheme is predictable.

The CSS specifications are maintained by the World Wide Web Consortium (W3C). Internet media type (MIME type) text/css is registered for use with CSS by RFC 2318 (March 1998). The W3C operates a free CSS validation service for CSS documents.

In addition to HTML, other markup languages support the use of CSS including XHTML, plain XML, SVG, and XUL.

6.4.4 GOOGLE COLAB

Google Colab is a free cloud service and now it supports free GPU. You can improve your Python programming language coding skills. We can develop deep learning applications using popular libraries such as Keras, TensorFlow, PyTorch, and OpenCV.
The most important feature that distinguishes Colab from other free cloud services is: Colab provides GPU and is totally free. Detailed information about the service can be found on the faq page. Colaboratory is a Google research project created to help disseminate machine learning education and research. It's a Jupyter notebook environment that requires no setup to

use and runs entirely in the cloud. This means that as long as you have a Google account, you can freely train your models on a K80 GPU.

6.4.5 Tensorflow. js

TensorFlow.js is largely modeled after TensorFlow, with a few exceptions that are specific to the JS environment. Like TensorFlow, the core data structure is the Tensor. The TensorFlow.js API provides methods to create tensors from JS arrays, as well as mathematical functions that operate on tensors. Figure 1 shows a high level schematic view of the architecture. TensorFlow.js consists of two sets of APIs: the Ops API which provides lower-level linear algebra operations (e.g. matrix multiplication, tensor addition, etc.), and the Layers API, which provides higher-level model building blocks and best practices with emphasis on neural networks. The Layers API is modeled after the tf.keras namespace in TensorFlow Python, which is based on the widely adopted TensorFlow.js is designed to run in-browser and server-side. When running inside the browser, it utilizes the GPU of the device via WebGL to enable fast parallelized floating point computation. In Node.js, TensorFlow.js binds to the TensorFlow C library, enabling full access to TensorFlow. TensorFlow.js also provides a slower CPU implementation as a fallback (omitted in the figure for simplicity), implemented in plain JS. This fallback can run in any execution environment and is automatically used when the environment has no access to WebGL or the TensorFlow binary.

TensorFlow.js was designed to bring ML to the JS ecosystem, empowering a diverse group of JS developers with limited or no ML experience. At the same time, we wanted to enable experienced ML users and teaching enthusiasts to easily migrate their work to JS, which necessitated wide functionality and an API that spans multiple levels of abstraction. These two goals are often in conflict, requiring a fine balance between ease-of-use and functionality. Lastly, as a new library with a growing user base,
missing functionality was prioritized over performance.

CHAPTER 7
RESULTS AND ANALYSIS

7.1 TEST ING

Testing is the process of evaluating a system or its components with the intent to find whether it satisfies the specified requirements or not. This activity results in the actual, expected and difference between their results. In simple words testing is executing a system in order to identify any gaps, errors or missing requirements in contrary to the actual desire or requirements. Testing is the practice of making objecting judgments regarding the extent to which the system (device) meets, exceeds or fails to meet stated objectives.

Test cases	Function	Input	Expected Output	Obtained Output
1	Disease Present/ Not	<63,1,3,145,233,1,0,150,0,2.3,0,0,1,1>	Disease Present	Disease Present
2	Disease Present/ Not	<58,1,1,120,284,0,0,160,0,1.8,1,0,2,0>	Disease Not Present	Disease Not Present
3	Disease Type	<44,1,2,140,235,0,0,180,0,0,2,0,2,0>	Coronary Heart Disease	Coronary Heart Disease
4	Disease Type	<57,1,2,150,168,0,1,174,0,1.6,2,0,2,1>	Myocardial Infraction	Myocardial Infraction
5	Disease Type	<42,0,0,102,265,0,0,122,0,0.6,1,0,2,2>	Hypertensive heart disease	Hypertensive heart disease
6	Disease Type	<56,0,1,140,294,0,0,153,0,1.3,1,0,2,3>	Heart Arrhythmia	Heart Arrhythmia
7	Disease Type	<44,1,2,140,235,0,0,180,0,0,2,0,2,4>	Sorry cannot classify the type	Sorry cannot classify the type

Table 7.1 Test cases

The results obtained in the above table are all the possible outcomes obtained from our project. Here function represents the purpose of testing. Expected outputs of a project are the deliverables from the research work. The obtained outputs are the outputs generated after execution of the input values given by the user. From this system the type classification is restricted only to a few types. If there is a confusion as to which type of disease the input values belong to then the system fails to classify the type of heart disease.

7.2 SNAPSHOTS

Home page of Predicting Heart Disease and Type Classification System.

Fig 7.2(a) Home page

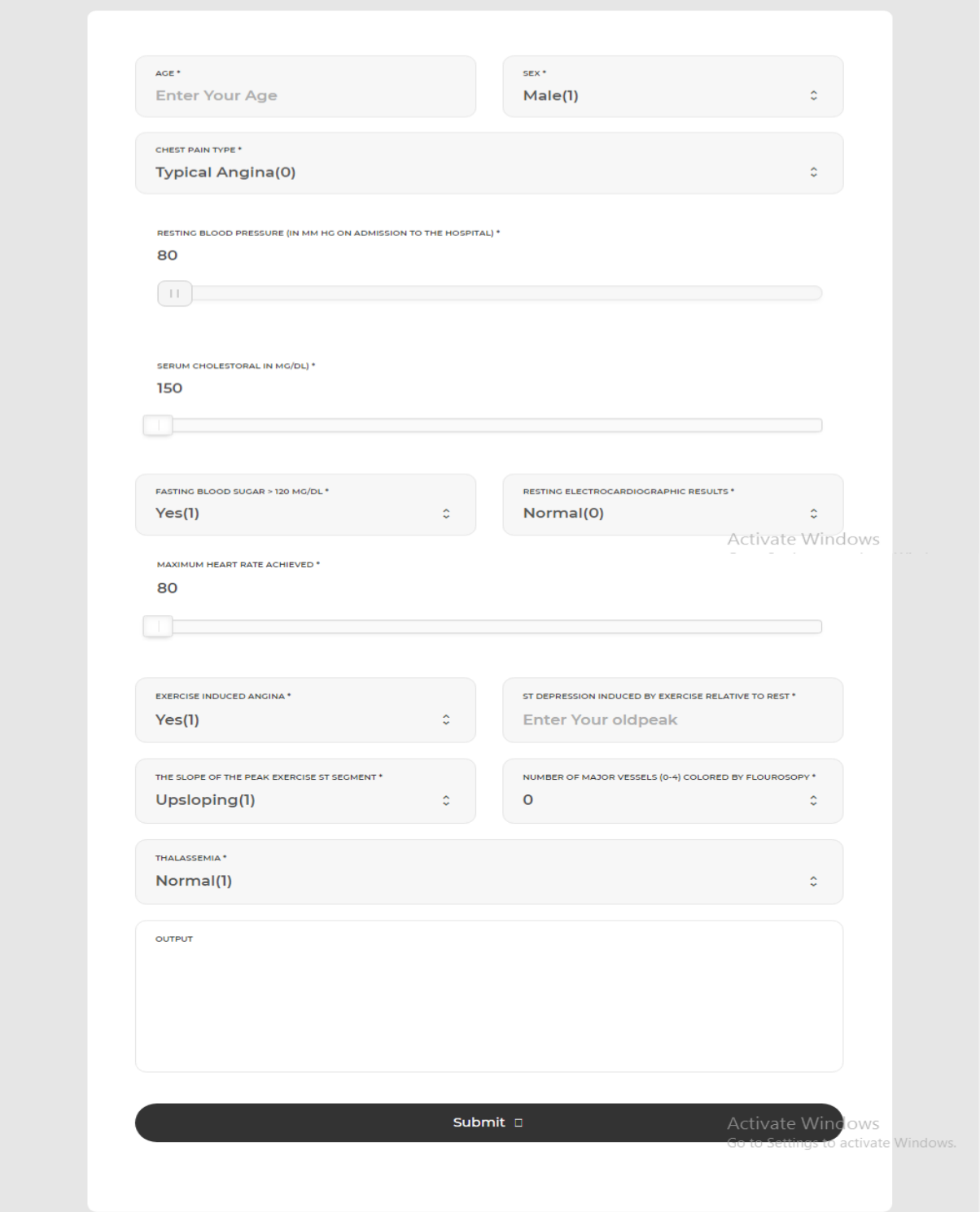


Fig 7.2(b) Attributes form

Predicting whether the disease is present or not.

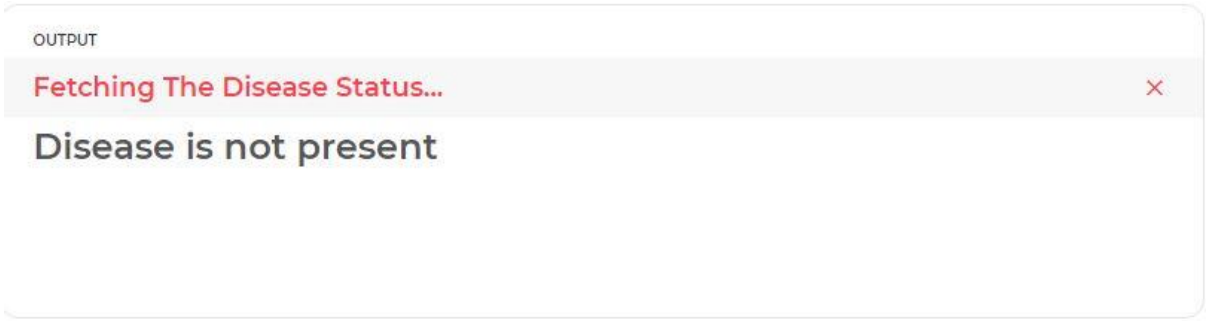


Fig 7.2(c) Disease Prediction

Disease classified as Coronary heart disease.

Fig 7.2(d) Coronary heart disease type Classification

Disease classified as Heart arrhythmia.

Fig 7.2(e) Heart Arrhythmia type Classification

Disease classified as Hypertensive Heart Disease.

Fig 7.2(f) Hypertensive Heart disease type Classification

Disease classified as Myocardial Infraction.

Fig 7.2(g) Myocardial Infraction type Classification

Disease is present but type cannot be classified.

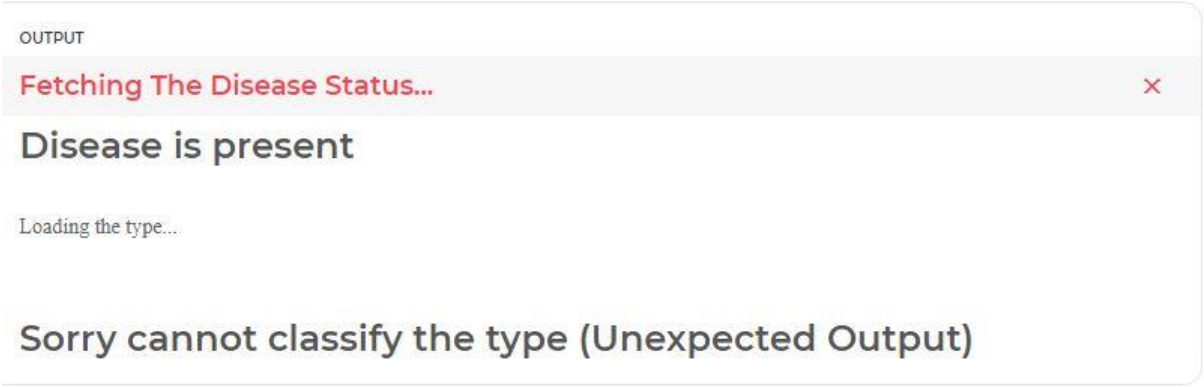


Fig 7.2(h) Disease cannot be classified

CONCLUSION

Heart disease must be predicted early because it is one of the prime causes of death in the world. An automated heart disease prediction system can be used as a device by doctors to determine heart disease. This survey paper presents a systematic heart disease prediction model which uses machine learning algorithm with comparatively more accuracy.

Here we are classifying certain types of heart diseases based on the dataset of patients who is having heart disease. The Machine learning algorithm, Artificial Neural Network is used for this purpose. As this paper suggests a new approach to classify heart disease accuracy is around 70 to 80 percent. In future approach more weightage must be given to accuracy.

FUTURE ENHANCEMENT

Heart disease is a leading cause of death. Predicting and classifying of heart disease is important to reduce death rate. There are many types of heart disease. We have classified only four of them. In future enhancement other types of heart disease to be classified with more accuracy rate. For this ANN can be used. More attributes can also be taken for accurate classification. Large amount of data to be trained to get more accuracy.

REFERENCES

[1] Ankita R. Mokashi et al. "Review on Heart Disease Prediction using Ann and Classifier", International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 IJERTV4IS090814 www.ijert.org Vol. 4 Issue 09, September-2015

[2] M. Akhil Jabbar, B.L Deekshatulu & Priti Chandra "Classification of Heart Disease using Artificial Neural Network and Feature Subset Selection" Global Journal of Computer Science and Technology Neural & Artificial Intelligence Volume 13 Issue 3 Version 1.0 Year 2013

[3] Animesh Hazra, Subrata Kumar Mandal, Amit Gupta, Arkomita Mukherjee and Asmita, Mukherjee "A Review on Heart Disease Diagnosis and Prediction Using Machine Learning and Data Mining Techniques" Advances in Computational Sciences and Technology ISSN 0973-6107,

[4] Shahid Mehmood Awan ,Muhammad Usama Riaz , Abd Ghaffar Khan, "Prediction Of Heart Disease Using Artificial Neural Network" Vfast Transactions on Software Engineering ISSN(e): 2309-6519; ISSN(p): 2411-6327 Volume 13, Number 3, September-December, 2018

[5] Himanshu Sharma, M A Rizvi, "A survey on Prediction of Heart Disease using Machine Learning Algorithms", International Journal on Recent and Innovation Trends in Computing and Communication ISSN: 2321-8169.

[6] Sellappan Palaniappan , Rafiah Awang, "Intelligent Heart Disease Prediction System Using Data Mining Techniques", IJCSNS International Journal of Computer Science and Network Security.

[7] Amita Malav, Kalyani Kadam "A Hybrid Approach for Heart Disease Prediction Using Artificial Neural Network and K-means" International Journal of Pure and

Applied Mathematics Volume 118 No. 8 2018, 103-110 ISSN: 1311-8080 (printed version); ISSN: 1314-3395

[8] R. Chitra and V. Seenivasagam "Review of Heart Disease Prediction System Using Data Mining and Hybrid Intelligent Techniques", Ictact Journal On Soft Computing, July 2013, VOLUME: 03, ISSUE: 04

[9] Mrudula Gudadhe, Kapil Wankhade, and Snehlata Dongre, Sept 2010, "Decision Support System for Heart Disease Based on Support Vector Machine and Artificial Neural Network", International Conference on Computer and Communication Technology (ICCCT), DOI:10.1109/ICCCT.2010.5640377, 17-19.