

**+HEART DISEASE PREDICTION USING DATA MINING APPROACH**

**BY**

**Mahimul Islam Nadim**

**ID: 161-15-847**

**Jane Alam**

**ID: 161-15-920**

**Humaira Khanam**

**ID: 161-15-1002**

This Report Presented in Partial Fulfillment of the Requirements for the Degree of  
Bachelor of Science in Computer Science and Engineering

Supervised By

**Mr. Sheikh Abujar**

Senior Lecturer

Department of CSE

Daffodil International University

Co-Supervised By

**Mr. Dewan Mamun Raza**

Lecturer

Department of CSE

Daffodil International University



**DAFFODIL INTERNATIONAL UNIVERSITY**

**DHAKA, BANGLADESH**

**NOVEMBER 2019**

## **APPROVAL**

This Project titled “**Heart Disease Prediction using Data Mining Approach**”, submitted by Mahimul Islam Nadim, Jane Alam and Humaira Khanam to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on Date: 11-25-2019

## **BOARD OF EXAMINERS**

---

**Dr. Syed Akhter Hossain**

**Professor and Head**

Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

**Chairman**

---

**Dr. S M Aminul Haque**

**Associate Professor & Associate Head**

Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

**Internal Examiner**

---

**Saif Mahmud Parvez**

**Lecturer**

Department of Computer Science and Engineering  
Faculty of Science & Information Technology  
Daffodil International University

**Internal Examiner**

---

**Dr. Mohammad Shorif Uddin**

**Professor**

Department of Computer Science and Engineering  
Jahangirnagar University

**External Examiner**

## DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Mr. Sheikh Abujar, Senior Lecturer, Department of CSE**, Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

### Supervised by:

---

**Mr. Sheikh Abujar**  
Senior Lecturer  
Department of CSE  
Daffodil International University

### Co-Supervised by:

---

**Mr. Dewan Mamun Raza**  
Lecturer  
Department of CSE  
Daffodil International University

### Submitted by:

---

**Mahimul Islam Nadim**  
ID: 161-15-847  
Department of CSE  
Daffodil International University

---

**Jane Alam**  
ID: 161-15-920  
Department of CSE  
Daffodil International University

---

**Humaira Khanam**  
ID: 161-15-1002  
Department of CSE  
Daffodil International University

## ACKNOWLEDGEMENT

First we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Mr. Sheikh Abujar, Senior Lecturer, Department of CSE**, Daffodil International University, Dhaka. Deep Knowledge & keen interest of our supervisor in the field of “Data Mining” to carry out this project. His endless patience ,scholarly guidance ,continual encouragement , constant and energetic supervision, constructive criticism , valuable advice ,reading many inferior draft and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to supervision, co-supervision and Head, Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

## **ABSTRACT**

Most of the data in today's world is computerized, these are usually scattered and not properly utilized. Data mining works like a blind stick to utilize these data. As elsewhere, Bangladesh also contains heart disease which puts people at a significant risk of death. According to the World Health Organization (WHO), an estimated 17.9 million people died from cardiovascular diseases (CVD) [10]. Although there is a lot of data, we have little knowledge of decision making. We have identified the major sources of heart disease by reviewing these data and using data mining techniques. In this paper, the data mining system uses for medical sections such as Smoke, blood pressure, diabetes, systolic blood pressure, diastolic blood pressure like 15 attributes to find predict heart disease. This prediction predicts by some data mining algorithm namely Decision tree, Artificial Neural Network (ANN), SVM. The accuracy of these algorithms is (76.94%), (79.40%) and (84.12%). This work is carried out to track the performance of certain data mining techniques on certain selected attributes, as explained later.

**Index Term:** Data Mining, Heart Disease, ANN, Decision Tree, SVM

## TABLE OF CONTENTS

<b>CONTENTS</b>	<b>PAGE</b>
Board of examiners	ii
Declaration	iii
Acknowledgements	iv
Abstract	v
<b>CHAPTER</b>	
<b>CHAPTER 1: Introduction.....</b>	<b>1-7</b>
1.1 Data Mining.....	1
1.2 Data mining techniques.....	2
1.2.1 SVM (Support Vector Machine).....	2
1.2.2 ANN (Artificial Neural Network).....	3
1.2.3 Decision Tree (DT).....	4
1.3 Motivation.....	6
1.4 Research objective.....	7
<b>CHAPTER 2: Background.....</b>	<b>8-9</b>
2.1 Introduction.....	8
2.2 Related Works.....	8
2.3 Challenges.....	8-9

<b>CHAPTER 3: Overview of Disease.....</b>	<b>10-17</b>
3.1 Heart disease.....	10
3.2 Facts of Heart Disease.....	11
3.3 Causes of Heart Disease.....	11
3.3.1 Risk Factors of heart disease.....	12
3.3.2 Discuss about some of Risk Factors.....	12
3.4 Symptoms of heart disease.....	14
3.5 Remedies for heart disease.....	17
<b>CHAPTER 4: Research Methodology.....</b>	<b>18-19</b>
4.1 Introduction .....	18
4.2 Research object and instrumentation .....	18
4.2.1 Filtering statistically significant features.....	18
4.2.2 Applying Algorithm.....	19
<b>CHAPTER 5:.....</b>	<b>20-25</b>
5.1 SVM Performance.....	20
5.2 Applying Decision Tree.....	21
5.3 Applying ANN.....	22
5.4 Dataset Type.....	23
5.5 Comparative performance of algorithms.....	25

<b>CHAPTER 6: Conclusion, limitations and future research...</b>	<b>26-26</b>
6.1 Conclusion.....	26
6.2 limitations.....	26
6.3 future research.....	26
<b>CHAPTER 7: REFERENCES .....</b>	<b>27-30</b>



## LIST OF FIGURES

FIGURES	PAGE NO
Figure 1.1: Data Mining	1
Figure 1.2: General SVM classification for linear hyper plane [7]	2
Figure 1.3: SVM classification for curve hyper plane [7]	3
Figure 1.4: Multi-Layer Perceptron Neural Network (MLPNN)	4
Figure 1.5: Basic Decision tree and its pruned version [5]	5
Figure 1.6: Annual Mortality Rate Graph of Bangladesh with the World. [5]	6
Figure 3.1: Human Heart	10
Figure 3.2: Norman Heart VS Damage Heart	13
Figure 3.3: Cholesterol Level High, Low, Good & Bad	13
Figure 3.4: Heart Attack	14
Figure 3.5: Swelling feet	15
Figure 3.6: Changes in the structure of the hands	16
Figure 4.1: Methodology	19
Figure 5.1: Data points of SVM (Graph in 3D)	20
Figure 5.2: Data point of Decision Tree algorithm (3D view)	21
Figure 5.3: Working process of ANN	22

Figure 5.4: Data point of ANN algorithm (3D view)

23

## LIST OF TABLE

<b>TABLE</b>	<b>PAGE NO</b>
Table 5.1: Dataset	24
Table 5.2: Comparison ANN, Decision Tree and SVM.	25

# CHAPTER 1

## Introduction

This chapter highlights some of the basic criteria to clarify our work overview. It describes briefly the different techniques of data mining that we used for comparison.

### 1.1 Data Mining

In the area of computer sciences, data mining is a rapidly growing field. Data mining is a method for collecting useful information from so many databases. Due to non-trivial information in large volumes, it is most effective in exploratory research. For converting raw data into useful data, the method of data mining is used. It can be used in manufacturing control, customer retention, sports, astrology and Web Surf-Aid areas [1]. It's a multi-stage process as shown in Figure 1.1

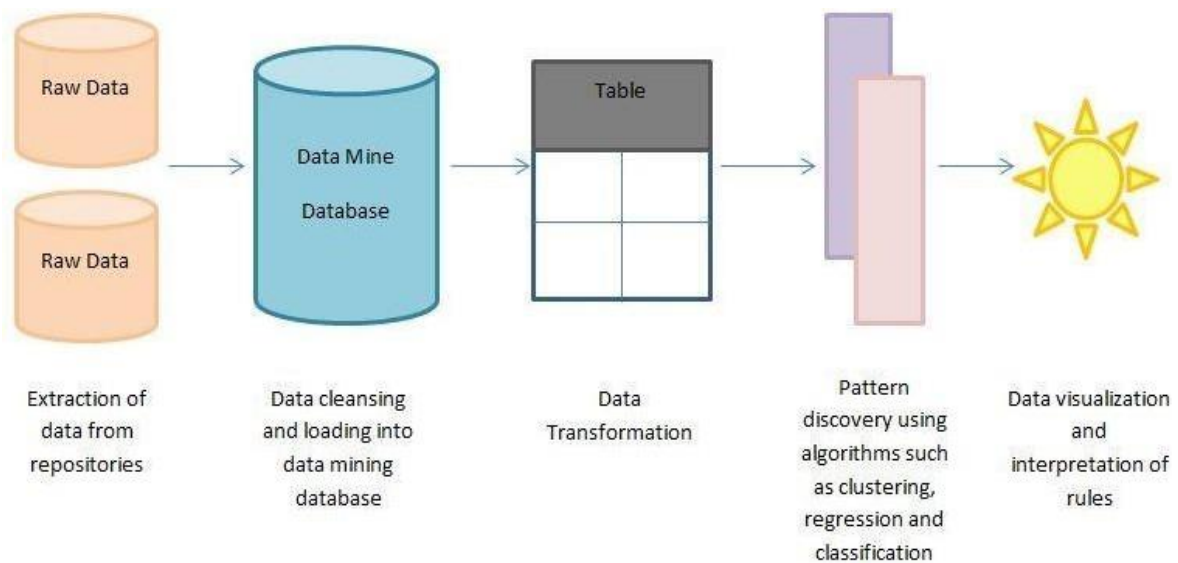


Figure 1.1: Data mining

In short, data mining is a process by which data is analyzed and obtained from different perspectives. Knowledge that has been discovered could be used in a wide range of industries, such as the health sector. The health sector stores large amounts of health information, which is unfortunately not "mined" for discovering hidden knowledge in order to make effective decisions. Data mining techniques can help to improve this situation, especially in the prediction of heart

disease. Heart disease Prediction is a highly sensitive and risky factor. If done properly, the health-

administrative services could use it. Researchers have explored various ways to implement data mining in healthcare over the last decade to achieve accurate coronary heart disease predictions.

## 1.2 Data mining techniques

Data mining brings a new dimension to disease prediction in the medical sector. There are many algorithms that can be used to study data. The few we are using are as follows:

### 1.2.1 SVM (Support Vector Machine)

Support Vector Machine is a concept of defining decision boundaries. A decision planes separate different class items. In the below figure the plane divides the example points to either red or green section. Any new example (data) that will fall right to the plane will be labeled as green otherwise it will be labeled as red [7].

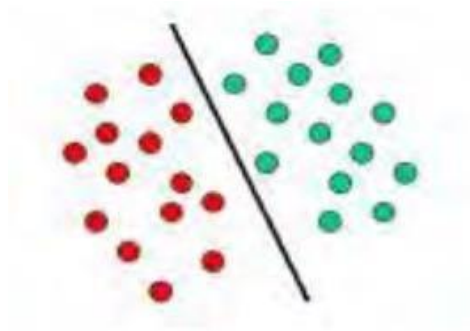


Figure 1.2: General SVM classification for linear hyper plane [7]

The above example was a linear type classification. But most of the classification examples are not simple and needs complex techniques to reach an optimal separation. For the below example, we can guess that there needs a curve plane to distinguish between objects. The drawing of these planes for classification task is called hyper plane classifiers.

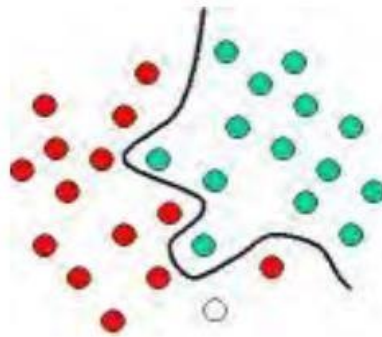


Figure 1.3: SVM classification for curve hyper plane [7]

### 1.2.2 ANN (Artificial Neural Network)

Artificial neural network (ANN) is a computational or mathematical model based on the structure and functions of biological neural network. One of Artificial Neural Network's most important models is Multilayer Perceptron (MLP), which maps input data to a suitable output array.

Multilayer Perceptron Neural Network (MLPNN) is the type of architecture used to implement the system. It uses three or more neural layers (nodes) with non-linear activation where two layers are respectively input and output and one or even more hidden layers are used as shown in Figure 1. 3.

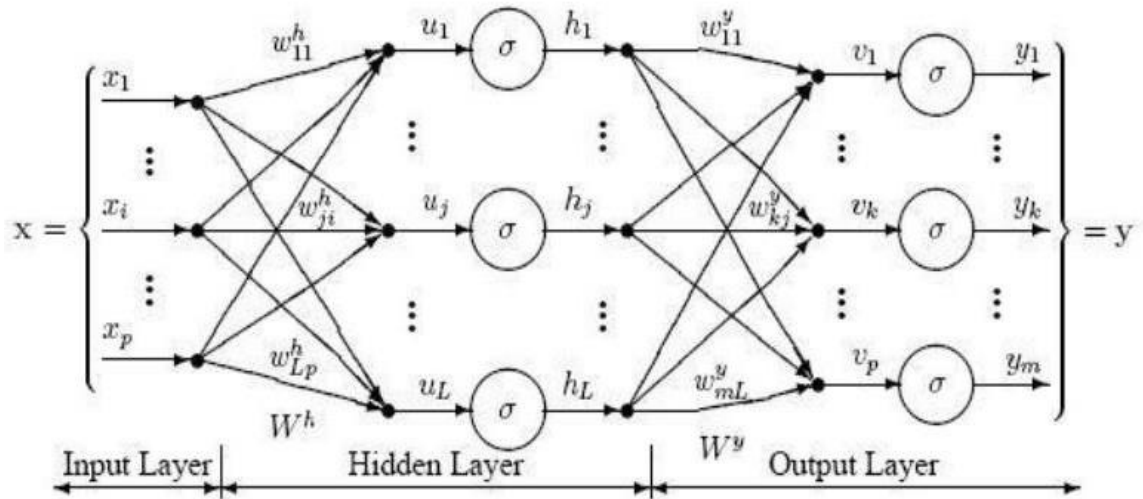


Figure 1.4: Multi-Layer Perceptron Neural Network (MLPNN)

The input layer receives signals from outside nodes. The output of the input layer is given by weighted connection links to the hidden layer [10].

### 1.2.3 Decision Tree (DT)

For classification problems, the decision tree approach is more efficient. A decision tree consists of a root node which is the parent node, branches representing test results and leaf nodes representing the class label or class label distribution. The DT is mainly used to achieve the ultimate goal as effectively as possible. Moreover, it is also used as a powerful machine learning tool. Several common algorithms are in the decision tree, including CART, ID3, C4.5, CHAID and J48. The DT contains a flowchart type structure where every single node has a corresponding test element and Decision tree implementation starts with its construction. For the root node, a feature test is chosen to help partition the training data in a manner that causes the class label associated with the data to be maximally disambiguated. This feature test will cause maximum reduction in class entropy while going through the training data. Then we drop from the root node to a set of child node, one for each partition of the training data created by the feature test at the root node. Here, the number of child to be able to get depends on how symbolic the features are.



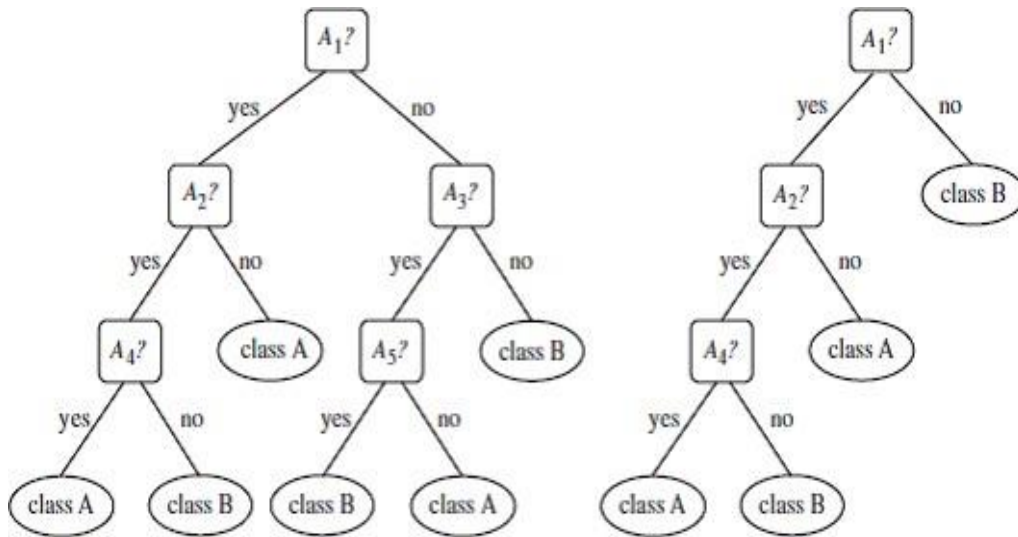


Figure 1.5: Basic Decision tree and its pruned version [5]

If our features are numerical, we find decision thresholds that divide the data and we drop to child nodes. Then we pose the same question for each child node that we asked for the root node that would disambiguate the class labels associated with that child node's training data to the maximum. Selecting the best test attribute is the challenging part of decision trees. Measurement of information gain is used to select the test attribute for each tree node. There is a term called entropy, which is a measure of how pure an arbitrary instance set is [9].

Let,  $S$  be a set which consists of data samples  $s$ . Class label attribute has  $m$  distinct values where  $m$  defines numerous categories,  $C_k$ .  $S_i$  be the number of samples consisting in  $S$  in  $C_k$  class.

Classification is done by the given equation:

$$I(S_1, S_2, \dots, S_m) = \log_2 P_k \dots\dots$$

$P_k$  is the probability that an arbitrary sample belongs to class  $C_k$  and is estimated by  $S_k / s$ . Let,  $A$  have  $v$  individual values,  $\{a_1, a_2, \dots, a_v\}$ . This attribute  $A$  is used to divide  $S$  into  $v$  subsets where  $S_j$  contains samples in  $S$  that have value  $a_j$  of  $A$ .

### 1.3 Motivation

A heart attack is a serious risk factor, where life and death comes very close. Habits or lifestyles are often the cause of heart disease prevalence. Most of us don't have any idea that we could have heart diseases or heart problems. If you have ideas in advance, most of them can be cured. Reduced physical activity as well as lack of exercise and less activity, these factors have, unfortunately, increased the rate of heart disease to a high percentage. The rate of heart disease has the same impact in a developing country like ours. According to the WHO country profile for 2018, cardiovascular disease alone kills 2.56 lakh people in Bangladesh accounting for 30 per cent of deaths caused by Non-Communicable Diseases where in 2014, it was 1.5 lakh or 17 per cent of the deaths caused by the Non-Communicable Diseases [3].

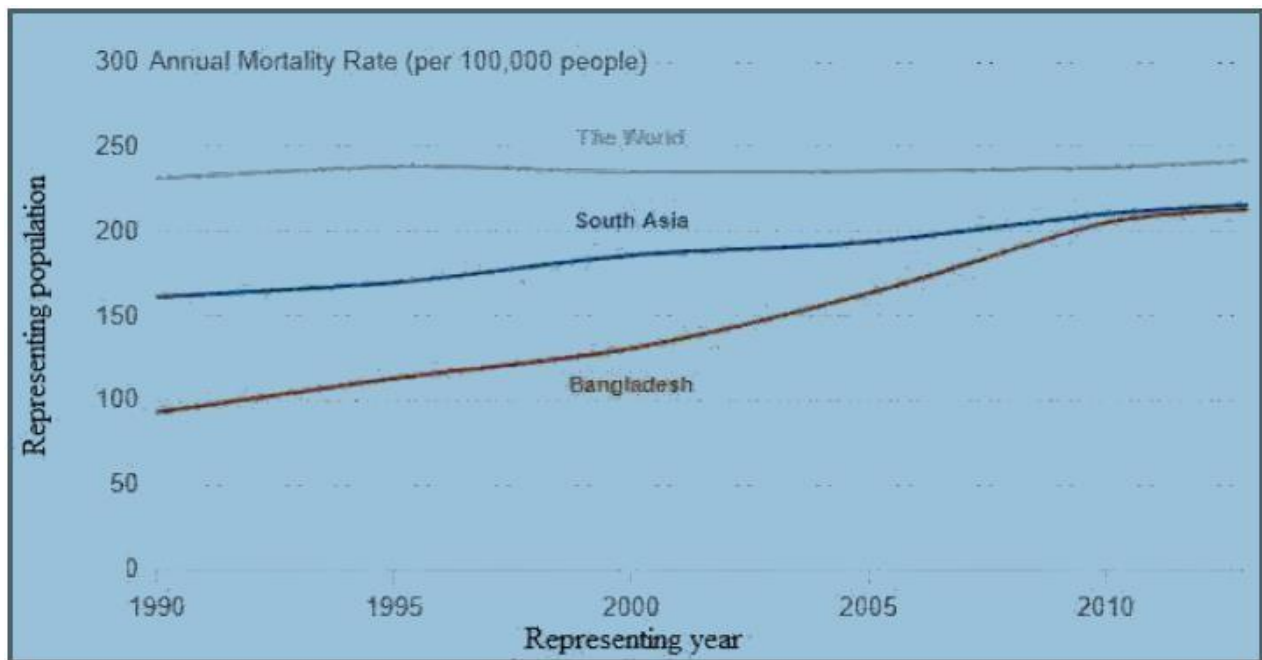


Figure 1.6: Annual Mortality Rate Graph of Bangladesh with the World. [5]

Predicting heart disease is a challenging and risky task because it depends directly on the health of people, so accuracy is a major factor. Poor predictions can lead to disastrous consequences. Therefore, this research focuses on the comparison of different data mining techniques to predict this. It shows the comparative analysis of the different methods.

#### **1.4 Research objective**

- Disease detection using symptoms
- Initiative of primary treatment
- Assist to preliminary tests
- Reduce medical diagnosis cost

## **CHAPTER 2**

### **Background**

#### **2.1 Introduction**

This sector contains the mechanisms used in existing works. Many researchers have contributed in this sector in using different techniques that helped to predict heart diseases in efficient ways. However, this field is still new to our type of research since more contributions can be made using different techniques, finding new patterns, improving accuracy etc.

#### **2.2 Related Works**

[13] In this paper a web based user friendly system was developed by using data mining techniques, namely Naïve Bayes and WAC (weighted associative classifier). They developed it to use as a training tool to medical students and train nurses to diagnose patients with heart disease and also to provide decision support to assist doctors to make better clinical decisions.

[15] A study by Chaitrali S. Dangare and Sulabha S. Apte uses Decision trees, Naive Bayes & Neural Networks algorithm to predict Heart disease by using 15 attributes. They got 100% accuracy in by applying Neural Network algorithm.

[16] To develop the multi-parametric feature by assessing Heart Rate Variability from ECG and heart disease pattern, a classification method was used by Kiyong Noh. The dataset distributed into two groups, namely patients and normal people with heart disease consisting of 670 peoples.

#### **2.3 Challenges**

The data in most of the hospitals are confidential in Bangladesh. We tried to collect data for research purposes but were denied. The data allowed to collect mostly contained name, age and sex which are not enough for prediction. Heart diseases depends on a lot of variables like hypertension,

diabetes, exercise schedules, chest pain etc. Collection of these data was against the hospital rules. Although the data was finally collected, it has a lot of missing data. The data was filled by using median method. The data set was also sparse which made prediction difficult if the scale factor was set to get a refine and more accurate prediction.

## CHAPTER 3

### Overview of Disease

#### 3.1 Heart disease

Heart disease usually refers to age-related structural changes in the heart, blood vessels, and veins, and kidney disease. High blood pressure is the main cause of the disease. Heart disease can occur at all ages. But older people are more at risk for the disease. Cholesterol levels generally increase with age and 80% of people with heart disease over the age of 65 die of heart disease. Again, the probability of having a stroke increases by twice the age of 50-55 years. As age increases, the elasticity of the arteries begins to decline, resulting in coronary artery disease. Men have more heart disease than women. It is mainly caused by smoking, cholesterol, hypertension, diabetes and many more.

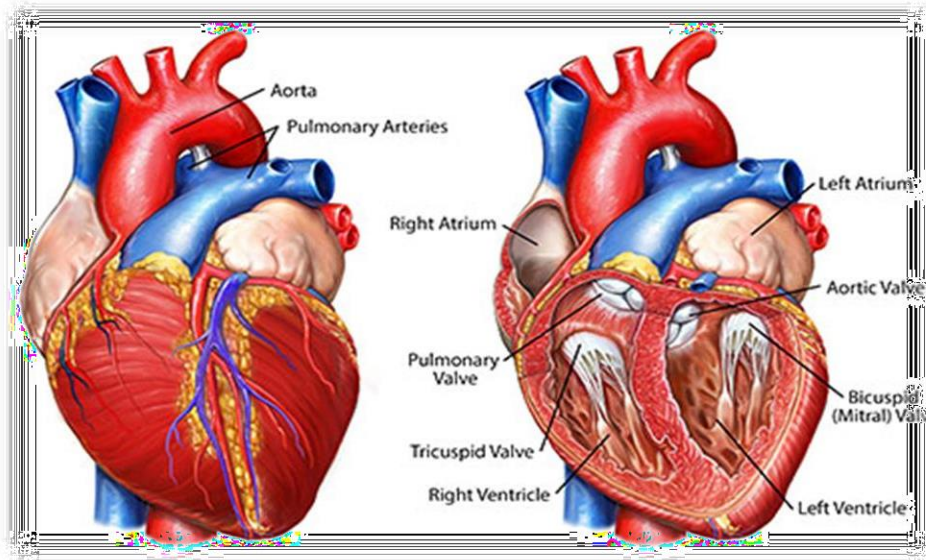


Figure 3.1: Human Heart

Heart disease can be many kinds. Such as hypertensive heart disease, congenital heart disease, coronary heart disease, heart failure, cardio-myopathy,, core pulmonary (right side of the heart and dysfunction of the respiratory tract), cerebrovascular disease (blood vessel dysfunction in the brain,

e.g. - stroke), marginal artery disease, rheumatic heart disease (heart and valve damage due to arthritis), cardiac dysrhythmias e Yadi.

To avoid this, the elimination of nutritious food, the elimination of smoking and tobacco products, maintaining the normal healthy weight, Hyper tension can also be control, regular medicine care, diabetes controlling ,hard work have to be eliminated and must be changing life style.

### **3.2 Facts of Heart Disease**

- Although heart disease is the cause of death for both men and women, the number of heart diseases and deaths is higher among men than women.
- Most people die of heart disease every year in the United States of america,1 people out of every 4 people.
- In the United States, the number of patients with heart disease is so high that every 43 second there is an attack and 1 people die every minute.
- Asia has become the heart of the continent after cancer. That has left people in a panic.
- About 17 million deaths every year worldwide are caused by CVD.

### **3.3 Causes of Heart Disease**

The following are some of the factors that can cause heart disease:-

- Age
- Sex/ Gender
- High blood pressure
- Smoking
- Diabetes (increase in blood sugar)
- Alcohol
- Hyper Tension
- Consuming tobacco products
- Increased Cholesterol In The Blood (Bad Cholesterol)
- Obesity (extra weight gain)

- Genetic continuity (if you have parents, grandparents at above the age of 50up, you are more than 50% more likely to be comparison an ordinary person)
- Lack of physical exercise/ poor exercise/ physical inactivity
- Unconsciousness in eating and drinking
- High lipid

It also increases the potential risk of a life-changing heart disease.

### 3.3.1 Risk Factors of heart disease

Following are some of the “**Risk Factors**” that can cause heart disease

- Smoking
- High blood pressure
- Increased Cholesterol In The Blood
- Genetic continuity
- Hyper Tension
- Obesity
- Age

Since, they are causes of very high risk for heart disease. To avoid this, you need to lose weight, regular exercise, by changing life style, change your eating habits, be free to smoke, have regular check-ups, and be aware of your family's age.

### 3.3.2 Discuss about some of Risk Factors

- **Smoking:** Smoking is a tobacco product made through a special process. It is usually made into smoke and burned into the body through the smoke inhaled. This is extremely harmful. Excessive smoking causes the tar to accumulate in the lungs, blocking blood vessels that can result in a major disease such as a stroke or a heart attack. **(A) Normal Heart VS (B) Damage Heart**



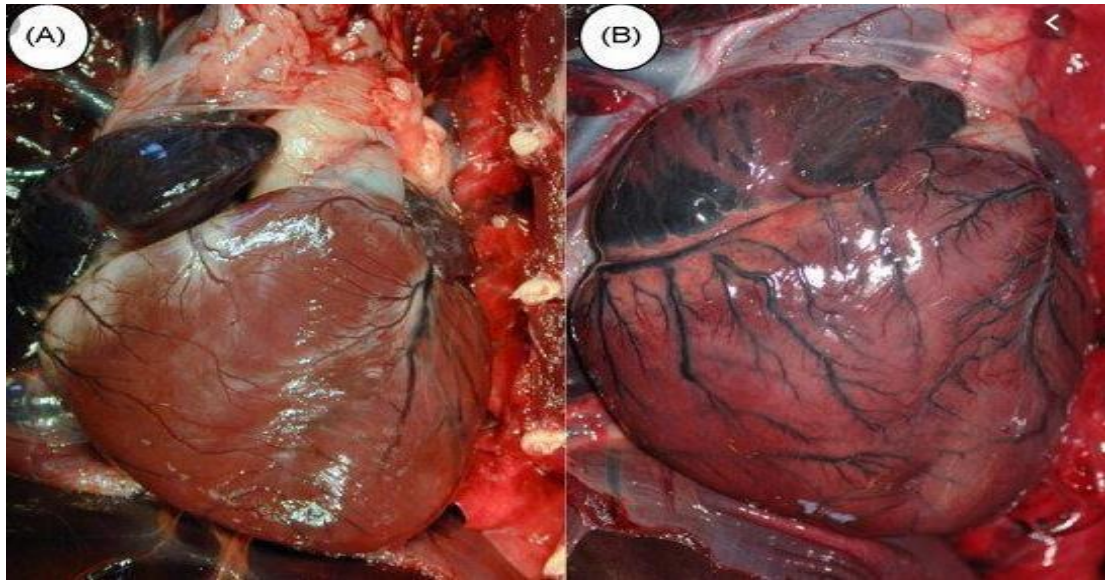


Figure 3 2: Norman Heart VS Damage Heart

- **Cholesterol:** Cholesterol is a type of fat cell. As it grows in the body, a variety of diseases are created. When cholesterol increases in the blood vessels, the artery becomes blocked. As a result, blood circulation stops and various parts of the body swell. Cholesterol accumulates in the heart, thus blocking blood circulation, which increases the risk of heart attack. Not only that, the brain can also be infected. Cholesterol increases, mainly for our eating, not exercising regularly, smoking, drinking, etc. To stay healthy, you have to give up all these bad habits.

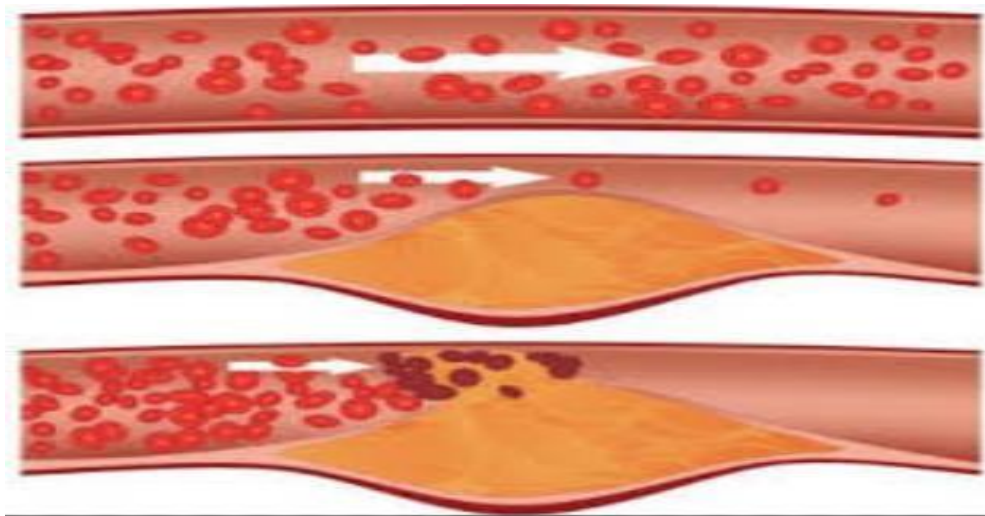


Figure 3.3: Cholesterol Level High, Low, Good & Bad

### 3.4 Symptoms of heart disease

The heart is an important organ in every organ of the body. The death of someone whose action is stopped. The heart is the main body of the human that is very important to stay active. So you need to know the symptoms of heart disease and follow up with good health.

**To know if you have a heart disease, we need to carefully consider the following symptoms:**

- **Chest pain:** Chest pain in the heart is mainly in the middle of the chest or spread around. Sometimes it can happen again on the left side of the chest. The pain can be severe and varied. The characteristic of this pain is usually acute stress. Many times the pain can be limited just to the middle of the chest or to the heart along the chest. Again, the pain can spread to the left shoulder, left arm, left hand, left shoulder finger. Sometimes this pain can occur in the right shoulder. Again, this pain can come from the throat and jaw. Someone again feels the pain on the stomach.



Figure 3.4: Heart Attack

- **Swelling of hands and feet:** If you feel the hands and feet are swollen at times, then it is a symptom of a heart problem. Normal blood vessels are stopped due to heart problems



which cause blood and body to swell in the arteries.

Figure 3.5: Swelling feet

- **Breathing:** Another important symptom of heart problems is breathing which is often avoided by people. When the heart is not functioning properly, there is a lack of oxygen in the body which results in shortness of breath. Therefore, it is advisable to consult a doctor quickly at this time.
- **Digestive Problems:** If you think you have been having digestive problems for some time and have not been able to cure any medication then you will find that many problems are present in the heart which causes digestive problems, preventing normal blood flow.
- **Changes in the structure of the hands and feet:** Often, the fingers or nails of the elderly are changed. For example, the hand or toes are bent and it is when the nails grow higher than the skin. It is called clubbing. This symptom usually occurs when the tissue of the fingers becomes thick. These changes are seen only when enough blood from the heart cannot reach the fingers.



Figure 3.6: Change in the structure of the hands

- **Excessive sweating:** If you feel excessive sweating, even if sitting under the wings at normal temperature. If water is flowing from the body then you should understand that a problem is being created in the heart. Excessive sweating cause's body fatigue and body dehydrate.
- **Sleepiness:** Even after normal sleep, if you do not get sleep rash or if you are feeling sleepy all the time, then you have to understand that heart problems are being created.

The symptoms of heart disease are not always manifest through chest pain, shortness of breath, weakness. Gender differences can also lead to different symptoms. In men and women, children and old people may have different symptoms. So be careful.

### 3.5 Remedies for heart disease

We have to know if we want to live safely. So it's important to know about heart disease remedy, now know how to easily cure heart disease and how to survive:

- Smoking, drugs and tobacco products are the main enemies of heart disease, so stay away from them. Say, no to smoking and no to drugs.
- No need to worry. Try to be completely free of thought. Only then can it be good.
- Occasionally, you should consult a doctor and try to higher to lower your body's blood pressure and try to maintain a normal blood pressure.

- Control diabetes / keep normal. Exercise regularly to keep your body healthy.

- Put plenty of vegetables on the diet list and delete eating healthy and cholesterol foods.  
Stay healthy Keep healthy Build a beautiful life.

All things considered, heart disease is a complex disease. There may be a loss of life for some negligence, but it can also be very good if you take good care. The chances of being perfectly good though are very low. So be aware at all times and seek the advice of a doctor from time to time. Food lists should change with age. It is possible to live a normal life by following the rules.

## CHAPTER 4

### Research Methodology

#### 4.1 Introduction

In this research we are analyzing various kind of attribute which are related with heart disease. All of syndromes to find we used data mining technique. Almost every attribute are visible in heart disease patients.

#### 4.2 Research object and instrumentation

The purpose of this research is to effectively analyze the possible result of the patient's dataset. To apply a prediction method, some features were developed in a model to identify the disease based on the characteristics of diseases. The steps in the process are as follows:

##### 4.2.1 Filtering statistically significant features

To build an optimized machine learning model it is always vital to determine which independent attributes or features has more statistical significance. Approach that has taken to create the learning model is as follows:

a) Feature Selection:

To find out highly significant features there are six methods available and they are All in, Raw data collect, Attribute analysis form the dataset and find the best reason which are actual courses in the heart disease patients, Data analysis, to find the accuracy using the data mining approaches, outcome analysis, check accuracy from final output respectively.

Step 01: Collect the raw dataset.

Step 02: Fit the data with all possible attributes.

Step 03: Analyze raw data from the dataset

Step 04: Filtering data using data mining technique (algorithm).

© Daffodil International University



Step 05: Result analysis from the output of dataset.

Step 06: Check the accuracy from the result.

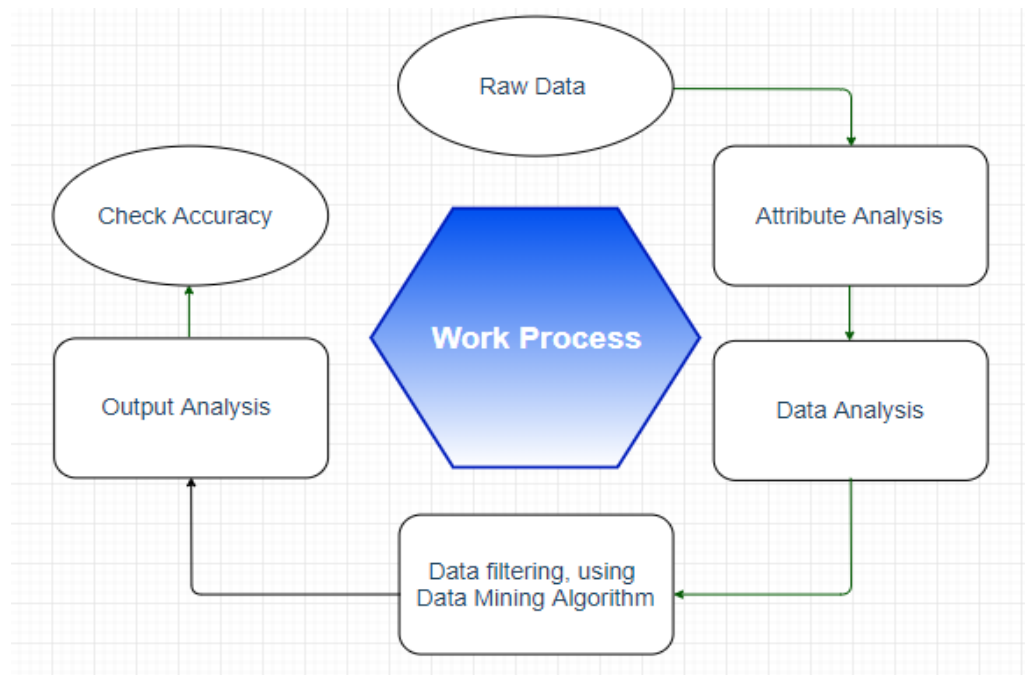


Figure 4.1: Methodology

### 4.2.2 Applying Algorithm

When we have our dataset ready with highly significant features we have applied ANN (Artificial Neural Network), Decision Tree, SVM (Support Vector Machine) to the dataset to predict the outcome of each point based on every independent feature that we have selected.

After finding the result of these algorithms, we have generated an accuracy report to evaluate the performance of the model and we have taken 'Accuracy score' and 'f1-score' into consideration as a numerical performance factors. The reason we took 'f1-score' over precision and recall is that the data we have used in this study is imbalanced and what 'f1-score' does is it combines precision and recall as harmonic means which punishes extreme values and gives a better intuition about the performance of the operation.

## CHAPTER 5

### Experimental Results and Discussion

In this study, we used authenticated survey data for analysis and research purposes. The data were analyzed by the similarity measure (Calculated by “Euclidean distance”) between data points and their output. We have plotted data points before and after analyzing to show an abstract of operation. Data was optimized into three-dimension using G-point (Graph point) to plot in a 3D graph. Every bar is represented by different colors and legends are used to identify them. A classification report is presented to give an insight on performance such as accuracy score, precession, recall, f1-score, etc.

#### 5.1 SVM Performance

The Heart disease dataset is loaded by the python ‘pandas’ library and plotted by ‘Jupyter Notebook’ data visualization python.

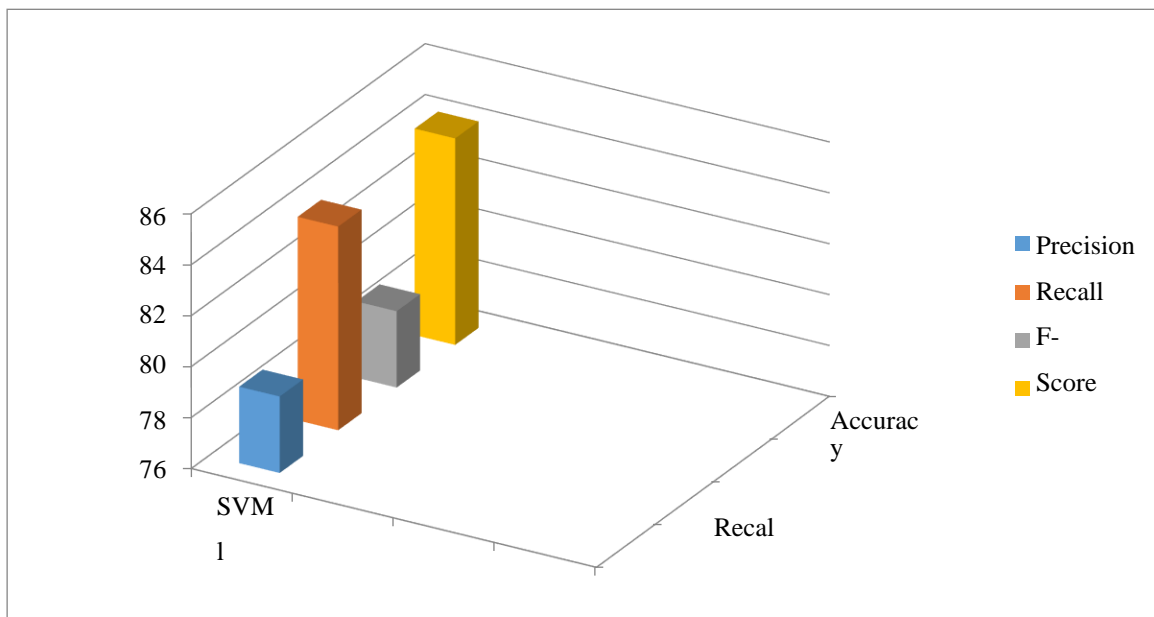


Figure 5.1: Data points of SVM (Graph in 3D)

## 5.2 Applying Decision Tree

Decision-making algorithm and show the result generated by a tree. It is easy for a tree-base decision-making algorithm. To help make and take any type of adjudication just a moment. The Decision tree is one of the most algorithms to show an accurate result. Parent attribute will be root. Sequentially, child and sibling are visible in this method. The performance is shown by the bar chart with a 3-D graphical view. Fig.5.2

The data points are shown using the color bar below:

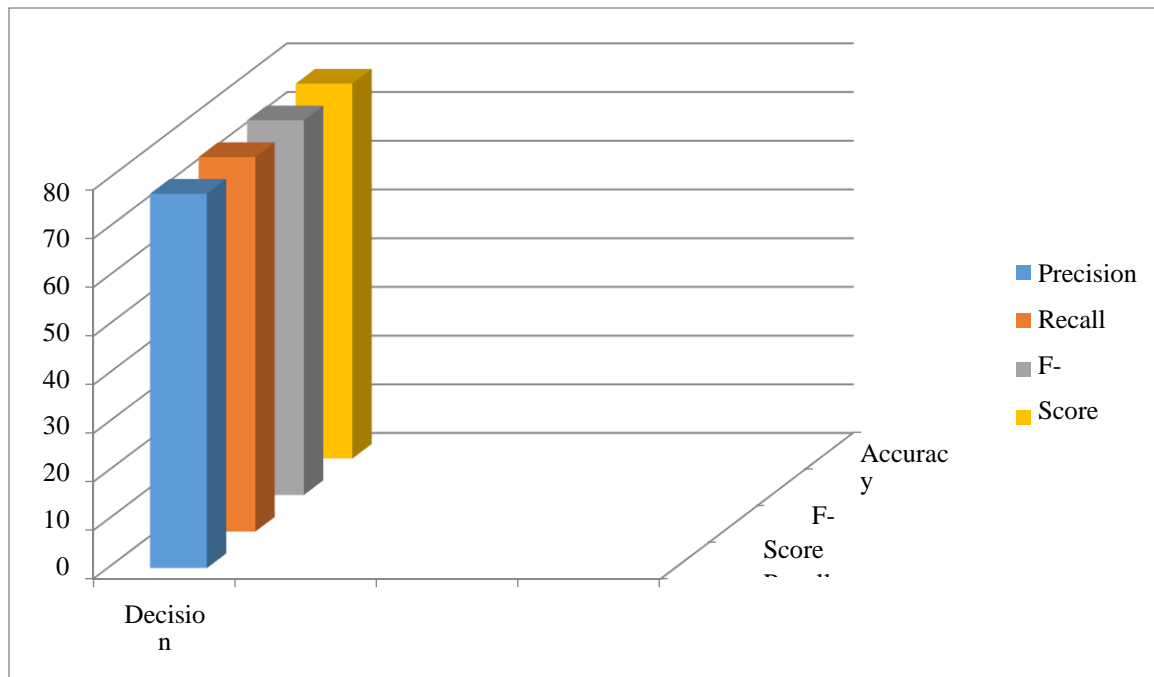


Figure 5.2: Decision Tree algorithm (3D view)

### 5.3 Applying ANN

The process of An Artificial Neural Network (ANN) is an interconnection one node to another node. This algorithm is the shape of the human brain. The circle is denoted by an artificial neuron and arrow represents one connection to another connection which is dependence. [11]

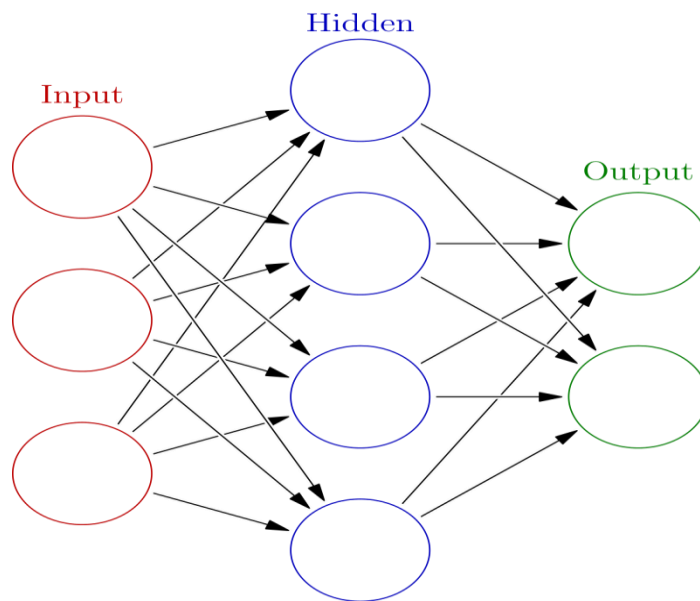


Figure.5.3: Process of ANNs

The classified algorithm Artificial Neural Network (ANN) result is shown Fig.5.4 where

has given the precision, recall, f-score and finally are total accuracy of these algorithm.  
Accuracy is 79.40%.

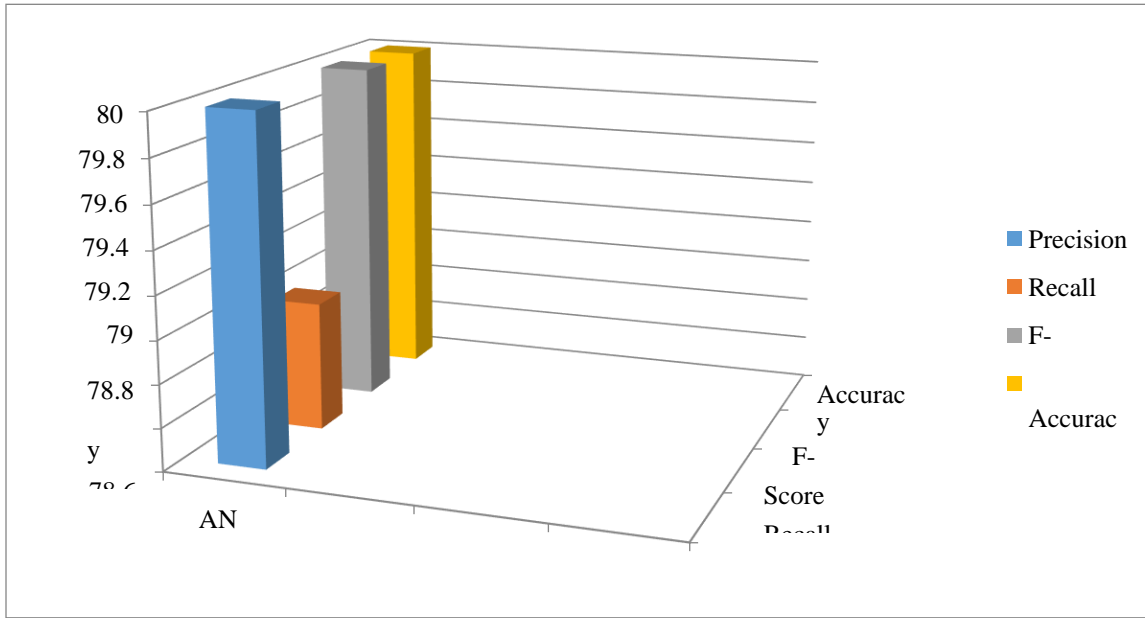


Figure.5.4: ANN algorithm with color bar

## 5.4 Dataset Type

The dataset consists of 14 more attributes and many objects. The age is between 36 to 75. Gender is male and female {1: male, 0: female}. Current Smoke define as Boolean that means 1 if they continue and 0 if there give up or no interested. Another related attribute is cigarette per day (clgs per day) are added as a numerical number. Blood Pressure Medicine (BD Meds), Prevalent Stoke, Prevalent Hypertension (Prevalent Hyp), Diabetes are added in this dataset.

Four most important issue gets in this set one in Systolic Blood Pressers (Sys BP), Diastolic Blood Pressers (Dia BP), Body Mass Index (BMI) and Heart Rate. The class attribute has a value of 0 for normal and 1 for attending the patient's heart disease. All these attribute has been taken to following some process [Table: 5.1].

gender	age	education	currentSmoker	cigsPerDay	BPMeds	prevalentStroke	prevalentHyp	diabetes	totChol	sysBP	diaBP	BMI	heartRate	glucose	T
1	39	4	0	0	0	0	0	0	195	106.0	70.0	26.97	80	77	
0	46	2	0	0	0	0	0	0	250	121.0	81.0	28.73	95	76	
1	48	1	1	20	0	0	0	0	245	127.5	80.0	25.34	75	70	
0	61	3	1	30	0	0	1	0	225	150.0	95.0	28.58	65	103	
0	46	3	1	23	0	0	0	0	285	130.0	84.0	23.10	85	85	
0	43	2	0	0	0	0	1	0	228	180.0	110.0	30.30	77	99	
0	63	1	0	0	0	0	0	0	205	138.0	71.0	33.11	60	85	
0	45	2	1	20	0	0	0	0	313	100.0	71.0	21.68	79	78	
1	52	1	0	0	0	0	1	0	260	141.5	89.0	26.36	76	79	
1	43	1	1	30	0	0	1	0	225	162.0	107.0	23.61	93	88	
0	50	1	0	0	0	0	0	0	254	133.0	76.0	22.91	75	76	
0	43	2	0	0	0	0	0	0	247	131.0	88.0	27.64	72	61	
1	46	1	1	15	0	0	1	0	294	142.0	94.0	26.31	98	64	
0	41	3	0	0	1	0	1	0	332	124.0	88.0	31.31	65	84	
0	39	2	1	9	0	0	0	0	226	114.0	64.0	22.35	85	76	
0	38	2	1	20	0	0	1	0	221	140.0	90.0	21.35	95	70	
1	48	3	1	10	0	0	1	0	232	138.0	90.0	22.37	64	72	
0	46	2	1	20	0	0	0	0	291	112.0	78.0	23.38	80	89	
0	38	2	1	5	0	0	0	0	195	122.0	84.5	23.24	75	78	

Table 5.1: Dataset



## 5.5 Comparative performance of algorithms

Algorithm's Name	Precision	Recall	F1-Score	Accuracy (In %)
Artificial Neural Network (ANN)	0.80	0.79	0.80	79.40
Decision Tree	0.77	0.77	0.77	76.94
Support Vector Machine (SVM)	0.79	0.84	0.79	84.12

Table 5.2: Comparison ANN, Decision Tree and SVM.

It can be inferred from Table 5.1 that Support Vector Machine (SVM) does a very good job in compare with Decision Tree and Artificial Neural Network almost 84% accuracy. From this table, the Artificial Neural Network's outcome accuracy is 79.40%. On the other head, Decision Tree's performance is 77% which is almost good but not well according to the other algorithm.

## CHAPTER 6

### Conclusion, limitations and future research

#### 6.1 Conclusion

Finally, as outcome result from Table: 5.1, the performance of Support Vector Machine (SVM) algorithm gives the overall classification accuracy percentage of 84.12%, which compared to the other algorithm like as Artificial Neural Network (ANN) and Decision Tree in classifying the stage of heart disease patients. The problem of constraining and summarizing different kinds of data mining technique is used in the field of medical prediction are discussed. We used 14 more attributes a patient for the most accuracy, what are the main reasons and find the specific motive of heart problem.

#### 6.2 Limitations

- ❖ In primary test cases, we skip some diagnosis factors such as ‘Chest Pain’, ‘Alcohol Intake’, ‘previous medical records’ so on.
- ❖ This approach mostly relies on the working dataset. The more accurate and larger amount of data, more accuracy in the result.

#### 6.3 Future scope

- ❖ The proposed model can be compared with other data mining techniques specially supervised learning techniques to check reliability.
- ❖ Another classifier like Naive Bayes, confusion matrix, etc. can be used to reduce the complexity and increase the accuracy.

## CHAPTER 7

### References

- [1] Soni, J., Ansari, U., Sharma, D. and Soni, S., 2011. Predictive data mining for medical diagnosis: An overview of heart disease prediction. *International Journal of Computer Applications*, 17(8), pp.43-48.
- [2] Masethe, H.D. and Masethe, M.A., 2014, October. Prediction of heart disease using classification algorithms. In *Proceedings of the world Congress on Engineering and computer Science (Vol. 2, pp. 22-24)*.
- [3] Pattekari, S.A. and Parveen, A., 2012. Prediction system for heart disease using Naïve Bayes. *International Journal of Advanced Computer and Mathematical Sciences*, 3(3), pp.290-294.
- [4] Thenmozhi, K. and Deepika, P., 2014. Heart disease prediction using classification with different decision tree techniques. *International Journal of Engineering Research and General Science*, 2(6), pp.6-11.
- [5] Subbalakshmi, G., Ramesh, K. and Rao, M.C., 2011. Decision support in heart disease prediction system using naive bayes. *Indian Journal of Computer Science and Engineering (IJCSE)*, 2(2), pp.170-176.
- [6] Imah, E.M. and Rahmawati, E., 2019. Classification of Emotional State Based on EEG Signal using AMGLVQ. *Procedia Computer Science*, 157, pp.552-559.
- [7] Latha, C.B.C. and Jeeva, S.C., 2019. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. *Informatics in Medicine Unlocked*, 16, p.100203.
- [8] Amin, M.S., Chiam, Y.K. and Varathan, K.D., 2019. Identification of significant features and data mining techniques in predicting heart disease. *Telematics and Informatics*, 36, pp.82-93.
- [9] Ilayaraja, M. and Meyyappan, T., 2015. Efficient data mining method to predict the risk of heart diseases through frequent itemsets. *Procedia Computer Science*, 70, pp.586-592.
- [10] [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)\(11-5-2019](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)(11-5-2019) at 10:57 pm)
- [11] [https://en.wikipedia.org/wiki/Artificial\\_neural\\_network](https://en.wikipedia.org/wiki/Artificial_neural_network) (23-11-2019 at 3:47 pm )
- [12] Xia Huosong, Du Kuanqi, Cai Shuqin, "Enterprise knowledge tree model and factors of KMS based on E-C", *Journal of Knowledge Management*, Vol.7-No.1, pp. 96-106, March 2003.
- [13] Chauhan R, Kaur H, Alam M A, "Data Clustering Method for Discovering Clusters in Spatial Cancer Databases", *International Journal of Computer Applications* , (0975 – 8887) Vol.10-No.6, November 2010.
- [14] NEWAGE Bangladesh, available at << <http://www.newagebd.net/article/51904/cardiac-patients-rising-in-bangladesh-25-lakh-die-annually> >>last accessed on 23-05-2019 at 5:00 AM.
- [15] Amandeep Kaur Mann, Navneet Kaur, "Survey Paper on Clustering Techniques", Volume 2, Issue 4, April 2013, ISSN: 2278 – 7798.
- [16] HealthGrove, "Cardiovascular Diseases in Bangladesh," by Graphiq. [Online]. Available: <http://global-disease-burden.healthgrove.com/l/41188/Cardiovascular-Diseases-in-Bangladesh#Overview&s=3D2kvJ>. [Accessed: 03-Aug-2017].

[17] Miss. Chaitrali S. Dangare, Dr. Mrs. Sulabha S. Apte, "A DATA MINING APPROACH FOR PREDICTION OF HEART DISEASE USING NEURAL NETWORKS", Volume 3, Issue 3, October -

December (2012), pp. 30-40

- [18] StatSoft, Inc. (2013). Electronic Statistics Textbook. Tulsa, OK: StatSoft. WEB: <http://www.statsoft.com/textbook/>.
- [19] M. A. Nishara Banu ; B. Gomathy, “Disease Forecasting System Using Data Mining Methods”, International Conference on Intelligent Computing Applications, March 2014.
- [20] G. Karthiga, C. Preethi, and R. D. H. Devi, “Heart Disease Analysis System Using Data,” vol.3, no. 3, pp. 3101–3105, 2014.
- [21] J. Han, J. Pei, and M. Kamber, Data Mining: Concepts and Techniques, 3rd ed. Elsevier Inc., 2011.
- [22] Vaishali, G. and Kalaivani, V., 2016, January. Big data analysis for heart disease detection system using map reduce technique. In 2016 International Conference on Computing Technologies and Intelligent Data Engineering (ICCTIDE'16) (pp. 1-6). IEEE.
- [23] Krishnaiah, V., Narsimha, G. and Chandra Subhash, N., 2016. Heart disease prediction system using data mining techniques and intelligent fuzzy approach: a review. International Journal of Computer Applications, 136(2), pp.43-51.
- [24] Latha, C.B.C. and Jeeva, S.C., 2019. Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques. Informatics in Medicine Unlocked, 16, p.100203.
- [25] Buchan, K., Filannino, M. and Uzuner, Ö., 2017. Automatic prediction of coronary artery disease from clinical narratives. Journal of biomedical informatics, 72, pp.23-32.
- [26] Amin, M.S., Chiam, Y.K. and Varathan, K.D., 2019. Identification of significant features and data mining techniques in predicting heart disease. Telematics and Informatics, 36, pp.82-93.
- [27] Ezekowitz, J.A., O'Meara, E., McDonald, M.A., Abrams, H., Chan, M., Ducharme, A., Giannetti, N., Grzeslo, A., Hamilton, P.G., Heckman, G.A. and Howlett, J.G., 2017. 2017 Comprehensive update of the Canadian Cardiovascular Society guidelines for the management of heart failure. Canadian Journal of Cardiology, 33(11), pp.1342-1433.
- [28] M.Umamaheswari, Dr. P. Isakki @Devi, “Myocardial Infarction Prediction using K-means Clustering Algorithm”, International Journal of Innovative Research in Computer and Communication, Vol. 5, Special Issue 1, March 2017.
- [29] Muhammad Zulfadhilah, Imam Riadi, Yudi Prayudi, “Log Classification using K-means Clustering for Identify Internet User Behaviours”, International Journal of Compiler Applications, (0975-8887), Vol.154No.3,November 2016.
- [30] Oyelade, O. J, Oladipupo, O. O and Obagbuwa, I. C, “Application of k-Means Clustering algorithm for prediction of Students’ Academic Performance”, International Journal of Computer Science and Information Security, Vol. 7, o. 1, 2010.
- [31] K.Rajalakshmi,Dr.S.S.Dhenakaran,N.Roobin“Comparative Analysis of K-Means Algorithm in Disease Prediction”, International Journal of Science, Engineering and Technology Research (IJSETR), Vol. 4, Issue 7, July 2015.

[32] A. K. Pandey, P. Pandey, K. L. Jaiswal, and A. K. Sen, "Data Mining Clustering Techniques in the

- Prediction of Heart Disease using Attribute Selection Method,” International Journal of Science, Engineering and Technology Research (IJSETR), ISSN: 2277798, Vol 2, Issue10, October 2013.
- [33] R. Das, I. Turkoglu, and A. Sengur, “Diagnosis of valvular heart disease through neural networks ensembles,” Elsevier, 2009.
- [34] M. Karaolis, J. A. Moutiris, and C. S. Pattichis, “Association rule analysis for the assessment of the risk of coronary heart events,” Proceedings of the 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2009.
- [35] S. Shilna and E. Navya, “Heart disease forecasting system using kmeans clustering algorithm with PSO and other data mining method,” International Journal On Engineering Technology and Sciences (IJETS™), ISSN(P): 2349-3968, ISSN (O): 2349-3976, Vol. 3, Issue 4, April 2016.
- [36] K.R. Lakshmi, M. V. Krishna and S. P. Kumar, “Performance Comparison of Data Mining Techniques for Predicting of Heart Disease Survivability,” International Journal of Scientific and Research Publications, ISSN 2250-3153, Vol.3, Issue.6, June 2013.
- [37] K. Solanki, P. Berwal and S. Dalal, “Analysis of application of data mining techniques in healthcare,” International Journal of Computer Applications, Vol. 148, No.2, August 2016.
- [38] M. Verma, M. Srivastava, N. Chack, A. K. Diswar and Nidhi Gupta,” A Comparative Study of Various Clustering Algorithms in Data Mining, ”International Journal of Engineering Research and Applications (IJERA), Vol. 2, Issue 3, pp. 1379-1384, 2012.
- [39] S. Revathi and T. NalinI, “Performance Comparison of Various Clustering Algorithm,” Vol. 3, Issue 2, ISSN: 2277 128X, February 2013.
- [40] R. Chauhan, H. Kaur and A. Alam, “Data Clustering Method for Discovering Clusters in Spatial Cancer
- [41] AmandeepKaurMann ,NavneetKaur ,”Survey Paper on Clustering Techniques “Volume 2, Issue 4, April 2013 ISSN: 2278 – 7798.
- [42] Jain A.K., Murty M.N., and Flynn P.J., “Data Clustering: A Review”, ACM Computing Surveys, 31 (3). pp. 264-323, 1999.