Predicting presence of heart disease in patients using neural network

MAJOR PROJECT II REPORT
SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY

IN

INFORMATION SYSTEMS

Submitted By:

Mohammad Fariz Hussain

(2K18/ISY/05)

Under the supervision of
Ritu Agarwal

DEPARTMENT OF INFORMATION TECHNOLOGY
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

JUNE, 2020

CANDIDATE'S DECLARATION

I, **Mohammad Fariz Hussain Roll No 2K18/ISY/05** student of M.Tech Information Systems, hereby declare that the Major Project II titled "**Predicting the presence of heart disease in patient using neural network**" which is submitted by me to the Department of Information Technology, Delhi Technological University, Bawana Road, Delhi-110042 in partial fulfilment of the requirement for the award of the degree of Master of Technology, is original and not copied from any source without proper citation. This work has not previously formed the basis for the award of any degree, Diploma Associateship, Fellowship or other similar title or recognition.

Place: Delhi NAME : **Mohammad Fariz Hussain**

Date:

CERTIFICATE

I hereby certify that the Project Dissertation titled "Predicting the presence of heart disease in patient using neural network" which is submitted by **Mohammad Fariz Hussain, Roll No 2K18/ISY/05 Information Technology,** Delhi Technological University, Bawana Road, Delhi-110042 in partial fulfilment of the requirement for the award of the degree of Master of Technology, is a record of the project work carried out by the student under my supervision. To the best of my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere.

Place: Delhi **Ritu Agarwal**

Date: **Assistant Professor**

ACKNOWLEDGEMENT

I express my gratitude to my major project guide **Ritu Agarwal**, lT Dept., Delhi Technological University, Bawana Road, Delhi-110042 for the valuable support and guidance she provided in making this major project. It is my pleasure to record my sincere thanks to my respected guide for her constructive criticism and insight without which the project would not have shaped as it has.

I humbly extend my words of gratitude to other faculty members of this department for providing their valuable help and time whenever it was required**.**

Mohammad Fariz Hussain

Roll No. 2K18/ISY/05

M.Tech (Information Systems)

E-mail: -Fariz.hussain@gmail.com

ABSTRACT

.

However, training neural networks require large datasets, otherwise, it cannot give accurate classification. Inspite all the data availability, there are some issues which lack enough data. Medical images, rare animals species to name a few examples with relatively less number of information. In our experiment, we have taken the data from the heart deprtment, which tells the condition of the heart. We will be comparing the cross entropy loss function with out own loss function.

CONTENTS

List of Figures

List of Tables

List of Symbols, Abbreviations and Nomenclature

ResNet – Residual Network

DenseNet – Dense Neural Network

FCN – Fully Connected Network

CNN – Convolutional Neural Network

ReLu - Rectified Linear Unit

RCNN – Region Convolutional Neural Network

YOLO – You Only Look Once

2D – Two Dimensional

3D – Three Dimensional

SVM – Support Vector Machine

DCNN – Deep Convolutional Neural Network

RGB – Red Green Blue

PCA – Principal Component Analysis

ICA – Independent Component Analysis

μ – Mean

σ – Variance

Σ – Summation

Ω - Subset

Chapter 1

Introduction

From the beginning of mankind, humans have faced many problems to survive on planet earth. We have managed to solve our problems one by one over time. We created things for our comfort. Little could we imagine what we were created in parallel. Diseases are the creation of humankind on this planet.

We created materialistic things but neglected the circumstances of the creation. In this thesis, I will be discussing the evolution of one of the diseases over the years. I.e. heart disease.

Heart diseases are nowadays very common in society. Heart disease results from a variety of situations that cause distress to the heart. There can be numerous causes of ailments related to the heart. Some causes include coronary artery disease, blood vessel disease, heart rhythm problem i.e. arrhythmias. In one severe chronic condition, you can be born with a defect in your heart i.e. congenital heart defects.

It is relevant to point out that the terminology used for heart disease by the doctor is cardiovascular disease. You can use this word interchangeably with heart disease. [1]

There are a variety of ailments that can severely distress heart health. Some of them are mentioned below.

Table 1: Diseases caused by the organ and their symptoms.

Symptoms caused by	Signs
Blood-vessels (atherosclerotic disease)[2] (tightening of blood vessels in different body parts)	→ Chest-pain. The patient can also experience tightness, pressure, or discomfort around the area of the chest. → Breathing problems are also commonly reported. One example is shortness in breath flow. → Pain in the areas of the neck, jaw, or throat. A patient can also experience pain in the upper abdomen and back. → Pain accompanied by numbness and weakness in the legs or arms. A tingling cold sensation is also reported.
Abnormal-heartbeats (heart arrhythmias)[2]	→ A throbbing sensation in the chest. → Slow heartbeat.

	→ Shortness of breath. → Light-headed-ness. → Dizzy-ness. → Fainting or near fainting. → Racing heartbeat. → Chest pain or discomfort.
Heart-defects[2]	→ Pale grey or blue skin color. → In an infant, short-ness of breath during feeding, leading to poor weight gain. → Bump in the legs, abdomen, or areas around the eyes.
Weak-heart-muscle(dilated cardiomyopathy)[2]	→ Breathless-ness with exertion or at rest. → Irregular heartbeats that feel fast, pounding, or fluttering. → Dizziness, lightheaded-ness, and fainting. → Swelling of the legs, ankles, and feet.
Heart-infections[2]	→ Shortness of breath. → Swelling in your legs or abdomen. → Dry or persistent cough. → Skin rashes or rare sports. → Fever. → Weak-ness or fatigue. → Changes in your heart rhythm.
Valvular heart-disease[2]	→ Short-ness of breath. → Swollen feet or ankles. → Chest-pain. → Fainting (syncope). → Fatigue. → Irregular heartbeat.

The following risk factors can also aggravate the causes of danger to heart health. [2]

- → Age: Old age increases the likelihood of weakness in the heart muscles.
- → Sex: Males are at high risk in comparison to women.
- → Family profile or genetics: If your close relatives have had previous medical history related to heart health. It increases the chances of you developing similar conditions.
- → Smoking: Nicotine and carbon monoxide are very dangerous for the heart. Smokers are highly vulnerable to develop heart ailments as compared to non-smokers.

- → Poor diet: High amounts of sugar, fat, salt, and cholesterol in your diet can increase the likelihood of poor heart health.
- → High blood pressure: In patients with High blood pressure, the arteries get hardened and thick. It results in narrowing the blood vessels through which the blood flows.
- → High level of blood cholesterol: As we all know, cholesterol is not good for health.
- → Diabetes: We have to control it. Or it may result in heart disease and other diseases too.
- → Obesity: Being overweight worsens other risk factors.
- → Physical inactivity: We have to move and keep our bodies active regularly to prevent us from heart diseases.
- → Stress: Stress is inevitable in today's lifestyles. It is our invisible enemy which can cause damage to our hearts. High degrees of stress increases the level of a hormone known as cortisol. Stress also harms our testosterone levels. Therefore, we need to take care of our stress levels and find ways to release them regularly. Exercise, Meditation, following a creative hobby are good ways to beat stress levels and keep your heart immune to stress and anxiety.
- → Poor hygiene: Hygiene is an essential part of our life. It is directly correlated with our immunity. If we don't maintain good hygiene, it may cause viral and bacterial infection and can lead to a heart infection.

Poor heart health can lead to the following conditions

- → Heart-failure.
- → Stroke
- → Peripheral artery disease.
- → Unexpected cardiac arrest.
- → Heart-attack.
- → Aneurysm.

In today's times, we are equipped with the right knowledge of heart diseases through advances in the areas of medical science and technology over some time. We can empower ourselves to use computer technology and data analytics to avert the fatalities caused by poor heart conditions.

We have be using the data gathered by the medical devices relevant to heart diseases to expect the chances of heart disease in human beings. We can classify the risk into two categories based on our existing knowledge. One category has chances to develop heart disease and the other is not having a heart disease. So it becomes a classification problem which we have solved using our developments in neural networks and machine learning techniques.

From the data provided by the patients, we have trained our model. We have be training the model for the two outcomes. The first category includes those having heart disease

and the other category covers those not having a heart problem. We have find out the extent of the accuracy we obtain. It will help us discover a model that can be used in the future for the betterment of human life.

Chapter 2

Literature Review

We have been working on the data collected regarding heart diseases based on the medical inputs. As stated earlier, we have classified our data into two categories. The first category of data set includes patients with heart ailments and the second category of patients are characterized by those not having any heart disease. We have tried this with different techniques. The techniques are as follows.

2.1 Logistic regression:-

This technique of data analysis is used to predict categorical data. I.e. binary data. It is a arithmetic bass technique that habits a logistic function to shape a technique of the binary dependent variable. While numerous more difficult expressions are out there. It is used to predict the probability of an event or a class. Example pass/fail, healthy/sick, win/lose, alive/dead, Weak-heart-muscle (dilated cardiomyopathy) [2]. For example, if a student passes in the exam. That prediction can be done using the previous data. The output will be only two those are a person will pass or fail. Only two outputs to predict. [3]

$$l = log_b \frac{p}{1-p} = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$$\frac{p}{1-p} = b^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}$$

$$p = \frac{b^{\beta_0 + \beta_1 x_1 + \beta_2 x_2}}{b^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + 1}} = \frac{1}{1 + b^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2)}}$$

Where *x_1 and x_2 are predictors*
l is log-odd
b is the base of a logarithm
β_i are parameters of the model

2.2 K nearest neighbour:-

In this basic technique, the original data is classified based on a predefined parameter by using the existing data. E.g. distance. You will find the usage of this technique for pattern recognition and statistical estimation.

The ballot of the neighbour is used as a means of classification whenever an original item arises in the data set. Further, the case is assigned to the maximum mutual k nearest neighbour corresponding to the distance function. [4]

Distance-functions:-

→ Euclidean distance function [5][6]

$$\sqrt{\sum_{i=1}^{k} (x_i - y_i)^2}$$

→ Manhattan [7][6]

$$\sum_{i=1}^{k} |x_i - y_i|$$

→ Minkowski [8]

$$[\sum_{i=1}^{k} |x_i - y_i|^q]^{\frac{1}{q}}$$

The above-mentioned equations are used for the constant variables. When it comes to definite variables, you are supposed to use Hamming distance in your equation.

Hamming distance:-

$$D_H = \sum_{i=1}^{k} |x_i - y_i|$$

$$x = y \Rightarrow D = 0$$
$$x \neq y \Rightarrow D = 1$$

The optimal value k is found after the inspection of the data. In general, the value of k is taken larger as it reduces the value of overall noise. [9]

2.3 Support-Vector-Machine

It is a way developed founded on supervised machine intelligence. You can use this technique to perform regression analysis as well as to meet the challenges posed by classification. The prominent usage of support vector machines is seen in grouping problems. In such a machine we plot the available data on the n-dimension space where n is the number of structures we have. Individually feature has its particular location on the n dimension plane. And each feature provides a particular coordinate. After that, we have to find a hyperplane that differentiates the classes from each other. [10]

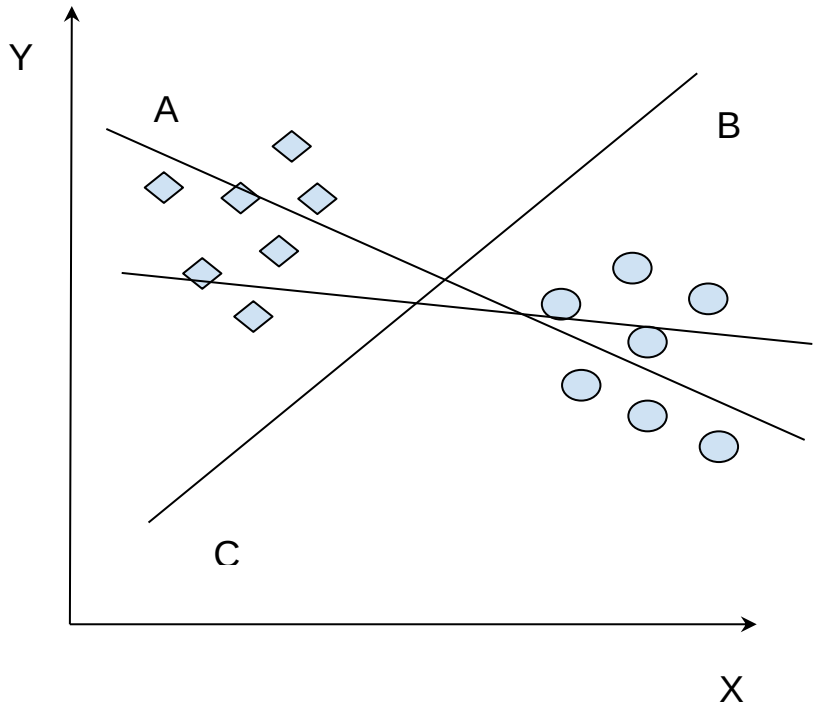

Figure 1: Graphical image of SVM to choose best fit.

We have to select the plane which separates the two classes of objects perfectly. The only rule is which plane separates the class better. In this given figure the plane B performs the best.

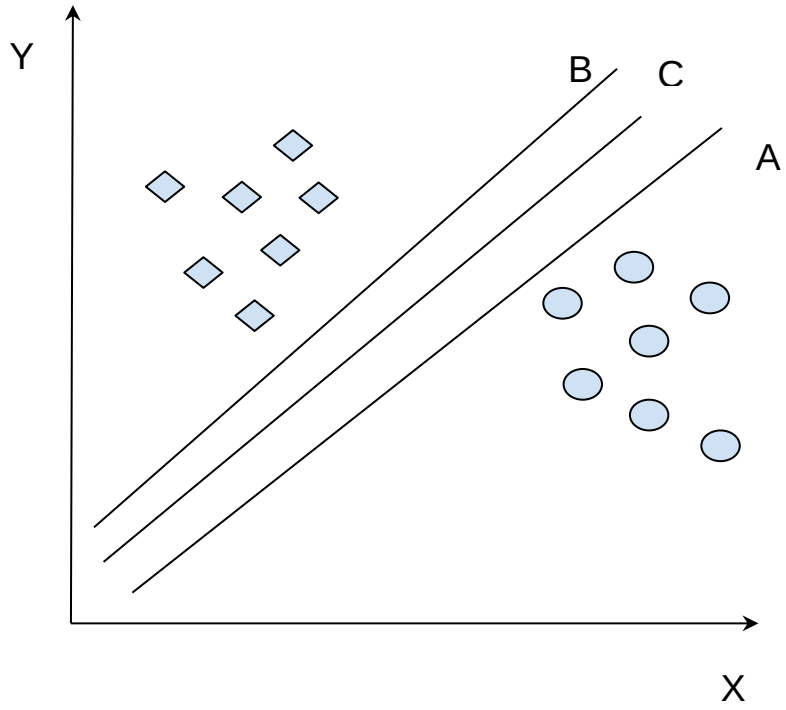

Figure 2: Graphical representation of SVM when we have more than one best fit.

In the above case, all three perform well in separating the classes but we have to select only one. So we calculate the separation between the plane and the adjacent data point. Which has the maximum distance will be the right hyperplane.

2.4 Naive Bayes Algorithm

This Algorithm is a technique for grouping. This method is built on the Bayes Theorem. The assumption is that predictors are independent of each other. In simple terms, a particular feature is independent of the feature present in the same class. It is mainly used for very large datasets. [11]

$$P(c \mid x) = \frac{P(x \mid c)P(c)}{P(x)}$$

$$P(c \mid x) = P(x_1|c)*P(x_2|c)*.....*P(x_n|c)*PI$$

Where
P (c | x) is the later probability of *class* (c, *target*) given *predictor* (x, *attributes*)
*P*I is the previous probability of *class*
P (*x* | *c*) is the probability which is the probability of *predictor* given *class*
P(*x*) is the prior likelihood of *predictor*

2.5 Decision-Tree Algorithm

A decision tree is a powerful tool in the machine learning library. The decision tree algorithm, unlike other supervised learning methods, can also be recycled to answer prediction problem and grouping techniques the end objective of using this tool is to generate a working out model that can be recycled by studying the rules of judgment based on previous data (test models) to guess the class or price of the goal variables.

The level of interpretation of the algorithm for the Decision Trees is too simple likened with other grouping algorithms. The decision-tree algorithm tries to crack the difficulty by using tree illustration. The inner node of each tree matches an attribute, and separately leaf node matches a class tag. [12]

Decision Tree Procedure: - Keep the best attribute to the dataset at the tree root. Divide the curriculum collapsed into subsets. Subsets can be rendered in such a way that any subset holds data with the similar value for an attribute.

Repeat the above procedure and step 2 on each subset until leaf nodes are found in all branches of the tree. Indecision trees we continue to guess a class mark for a record from the root of the list. They equate the ideals of the root attribute with those of the record attribute.

Grounded on contrast, we follow the branch matching to that value and move to the next node. We remain to equate our record’s attribute values with other internal tree nodes until we get to a leaf node with expected class value. So we learn how to forecast the target class or value using the decision tree based on it. Now let’s understand how we can create a pattern for the tree of decisions.

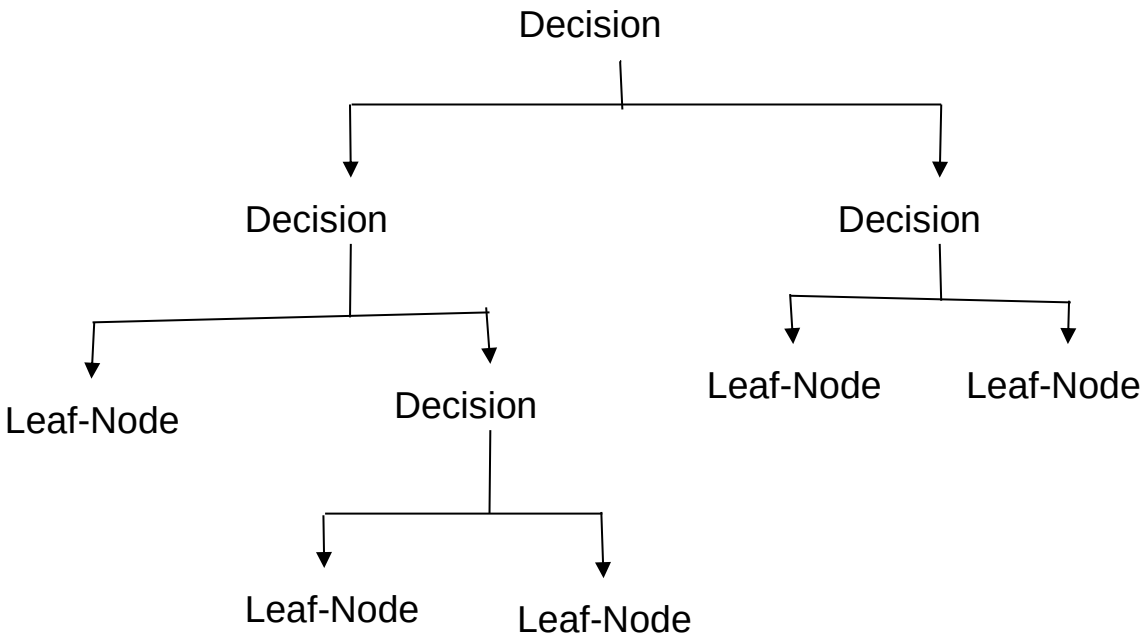

Figure 3 Decision tree example.

2.6 Random-Forest Classifier

It is an education algorithm based on a tree-based ensemble. The Random-Forest Classifier is an array of decision trees from a casually selected training set subset. It totals the votes for the final class of the test object from, unlike decision trees.

Ensemble algorithms are those that incorporate more than one same or different type of algorithm for classifying objects. For e.g., running predictions over Naive Bayes, SVM, and Decision Tree and then voting for the class for test object final consideration. [13]

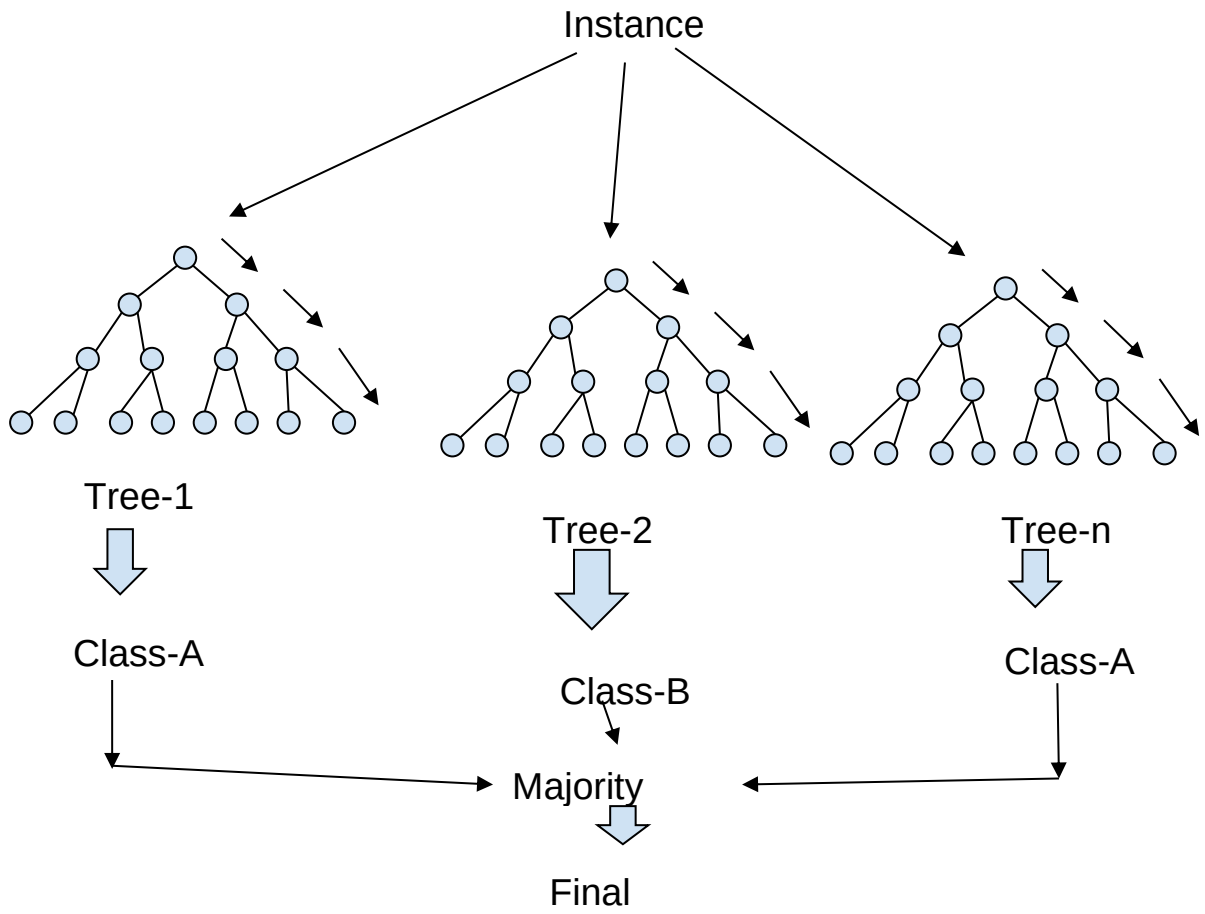

Figure 4: Graphical representation of Random forest classifier.

Features and Benefits of Random Forest: -

- → It produces a highly accurate classificatory on behalf of many records sets. It successfully functions on large datasets.
- → With no variable deletion, it facilitates you to group more than 1000 input variables
- → It provides you with approximations according to the significance of the grouping variables.
- → You get an inner, approximation regarding the mistake of simplification without bias.
- → You also get an actual estimation system for lost data. It helps you to ensure correctness even when a huge volume of data is lost.

Limitations of Random-Forest: -

Random forests can prove to be inefficient for data sets involving noisy classification and data sets involving complex regression analysis.

This algorithm gives ineffective and inconsistent variable position results for certain kinds of data set which have extra layers of data in its sample size. E.g. business is

observed while reporting in datasets with varying degrees of complex categorical variables.

Chapter 3

Proposed work

In this model, we have used the neural network for the estimate of heart disease in the patient. We have introduced a new cost function in this model and find the updated results. And check whether this new function can be proved helpful to us or not in improving the time or the accuracy of the model.

3.1 Neural network

We have to recognise the meaning of the neural network to know the whole procedure. The term Neural is taken from the root word neurons. It denotes the base unit for referencing the nervous system. This power of neurons can be used to create a powerful network and can be used for different purposes in day to day life. Hence the artificial neural network is introduced. And it is used widely in our life to solve many issues of our life. [14]

3.2 Artificial neural networks

Artificial neural networks can be simply described as biologically motivated models that are simulated on the computer to perform a well-defined series of tasks such as clustering, sorting, pattern acknowledgment, etc.

Typically, Artificial Neural Networks is a genetically determined neural network (of an artificial nature) needed to work a complex set of tasks.[15]

3.2.1 Working of artificial neural networks

You can understand such complex networks better if you understand the concept of weighted directed graphs. In this case, the neurons represent the nodes on one hand. On the other hand, the nexus between neural outputs and inputs is reflected by the directional endpoints by using weights. It takes signals from the environment as form vectors. Such incoming vectors can include patterns and images within their ambit. Further, the inputs are given algebraic notations in the form of x (n) to denote the increase in the number of inputs. Each input is further assigned a parallel weightage. [16]

(These weights are a crucial measure to solve specific problems under this model. You want to showcase the degree of connectivity or robustness of the neural interconnection

inside the artificial network primarily. Assigning weight is the first step in this direction. All weighted inputs are systematically summed up through internal computations.

At this point, it is crucial to briefly discuss the inner workings of the system. A bias is added when the weighted sum becomes zero. It ensures that the system gives a non-zero output. It is also useful when you are attempting to scale up in response to the system.

Bias is correlated to weight and represents the input value to 1. It should be noted that the number of weighted inputs can range from 0 to infinity.

So benchmarking is employed to manage the result within the range of desired value. A predefined threshold is capped for this purpose.

After that, the activation function is implemented to check through the sum of weighted inputs. The transfer functions that are grouped to generate the targeted outcomes is known as the activation function. The activation function comes in the form of varied linear and nonlinear sets.

The most preferred activation functions in terms of usage are as follows:

- → The Binary activation function
- → Sigmoidal activation function (linear)
- → Tan hyperbolic sigmoidal activation function (nonlinear)

These activation paragraphs are discussed in depth in the forthcoming paragraphs.

3.2.2 Architecture of Artificial Neural Networks

We need to consider what a traditional neural network comprises to understand the architecture of an artificial neural network. A typical neural network consists of numerous artificial neurons (yes, of course, that's why it's called an artificial neural network) these are called units grouped in a sequence of strata. Let's look at the various kinds of layers integral to a neural artificial network:

Input-layer: It takes an input of the data. The data will enter from the input layer first and then it will pass on to the other layers.

Output-layer: We will get the result from the output layers. The number of nodes will represent the kind of output we are expecting. The output layer is the final layer in the network.

Hidden-layer: The hidden layer is in between the input and the output layers. It is mainly hidden. It helps the network to convert the input data to some information and pass that to the output.

The neural network has all of the three layers. They are connected. Nodes can be connected in any manner. It can be connected to all of the nodes in the other layers or to some selected ones.

3.3 Work of the Activation Function in a Neural-Network Model

The input data which is in the numeric form are given to the input layers in a neural network. Every neuron has a weight attached to it. Which is then multiplied by the input to that neuron. And the final value is passed on to the next layer to some neuron. The activation function is a mathematical operation which acts as a gate between the input neuron and the output neuron. It acts as a threshold that allows the data to transfer if it satisfies the conditions otherwise not. [17]

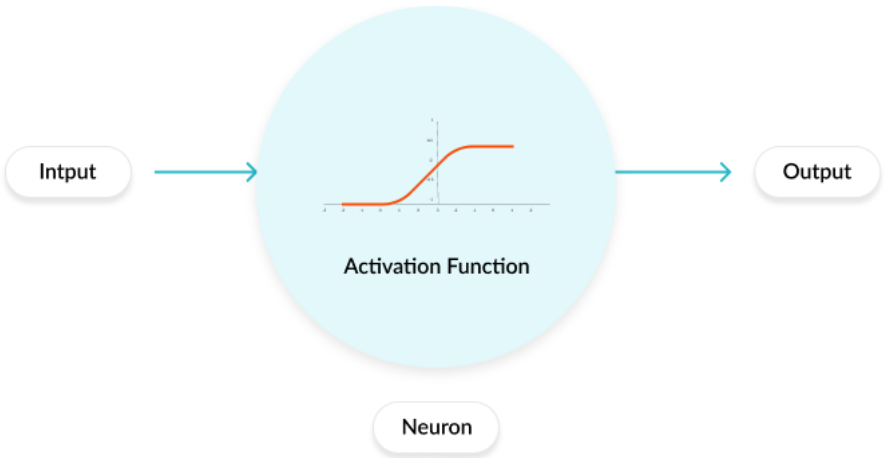

Figure 5: Internal working of neuron. [18]

Neural networks are extremely using nonlinear activation functions that can help the network study complex data, calculate and learn almost any function that poses a question, and give precise predictions.

3.4 Activation Functions Types

3.4.1 Binary-Step Function

It prevails as a function for activation grounded on a threshold. If the input value is overhead or underneath a specific threshold, it triggers the neuron and sends the same signal to the next layer.

The difficulty by a phase method is that it does not accept multi-value outputs — it cannot endorse, for example, classifying the inputs into one of many groups. [19][17]

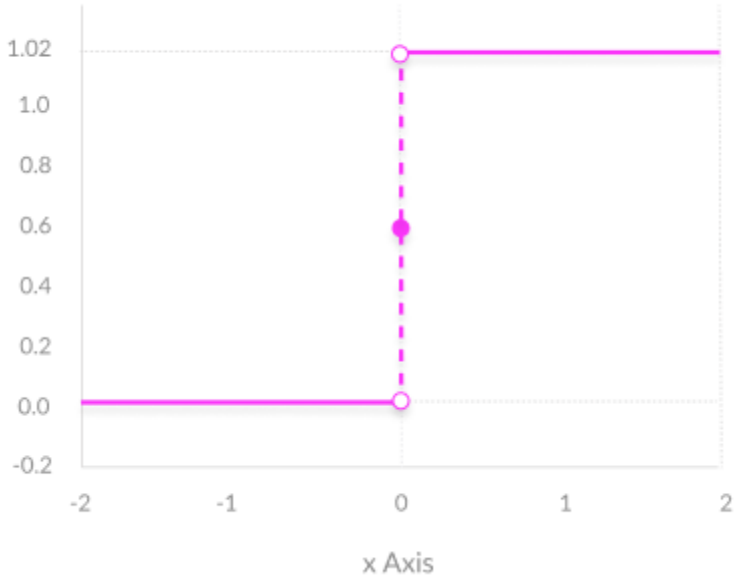

Figure 6: Binary Step Function Graph [18]

3.4.2 Linear-Activation Function

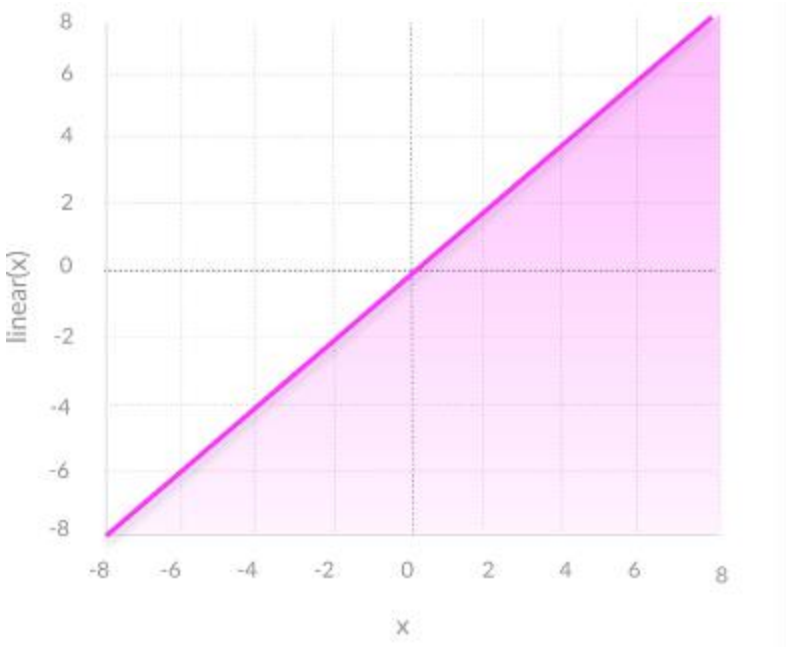

Figure 7: Linear-Activation Function Graph [18]

It takes the input and multiplies the input by the weights of the neuron. The output signal is proportional to the input. The liner is a good function compared to the step function. That is because the step function outputs only two values that are true or false. But the liner will output more than the two results.

There are some limitations with the linear activation function:

Backpropagation (gradient descent) cannot be used to train the model — the function's derivative is a constant. It is unrelated to the data. , X. It renders it implausible to drive back and interpret the weights that enable better prediction in the input neurons.

All neural network layers collapse into one — by linear activation functions. It ensures that there is no hassle in the number of layers for establishing a neural connection. The last layer becomes a linear first layer function. Accordingly, the neural network crystallizes into a single layer owing to the linear characteristics. It further makes the architecture nothing but a linear regression model. It has partial power and skill to handle input data parameters of varying complexity.

3.4.3 Non-Linear Activation Functions

Up-to-date models of neural networks utilize nonlinear activation functions. They permit the model to make compound mappings among the inputs and outputs of the network. This functionality makes learning possible. It also enables the demonstration of multi-dimensional and dynamic data such as non-linear or high dimensional audio, images, video, and data sets.

In a neural network, virtually any imaginable process can be represented as a functional computation, provided a nonlinear activation function.
Nonlinear functions answer linear activation problem:

- → Backpropagation is allowed because they have an input-related derivative function.
- → They require many coatings of neurons to “stack” to make a profound neural network. To study complex data sets with high precision stages, numerous hidden layers of neurons are wanted.

3.5 Seven Common Nonlinear Activation Functions

3.5.1 Sigmoid or Logistic

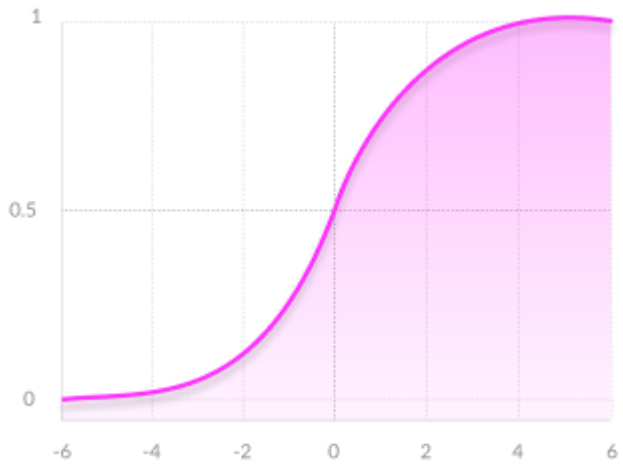

Figure 8: Sigmoid or Logistic Activation Function Graph [18]

Advantages

The performance values are in the form of a smooth curve without any jumps in it.
0 and 1 are the only outputs assured modifying each neuron output.
Strong predictions – it brings the value of Y to the edge very near to 1 or 0 near to the curve. That helps in good predictions.

Disadvantages

Vanishing gradient – for very high values of X that can be small or greater, there is nearly no adjustment to the forecast, causing vanishing gradient issues. It will affect the network to learn more or take more time to conclude.
Outputs not based on zero.
Costs on computation

3.5.2 TanH / Hyperbolic Tangent

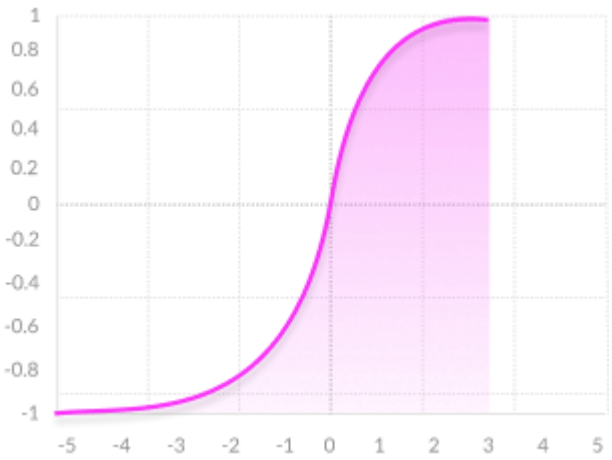

Figure 9: TanH/Hyperbolic Tangent Graph [18]

Advantages

Zero cantered — making modelling inputs with muscularly negative, neutral, and muscularly positive values easier.
As with the Sigmoid-function, otherwise.

Disadvantages

Similar functioning Sigmoid

3.5.3 ReLU-(Rectified Linear Unit)

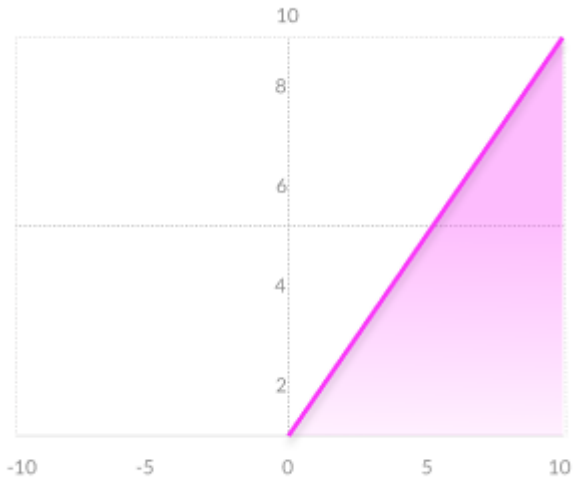

Figure 10: ReLU Activarion Function Graph [18]

Advantages

Computationally effective — permits the network to unite rapidly

Non-linear-while it appears similar to a linear function, Because of the derivation function, it is easy to back propagate with ReLU.

Disadvantages

The Dying ReLU issue — when inputs reach zero, or it can be negative, the function gradient becomes zero, the network is unable to achieve backpropagation and can’t acquire.

3.5.4 Leaky ReLU

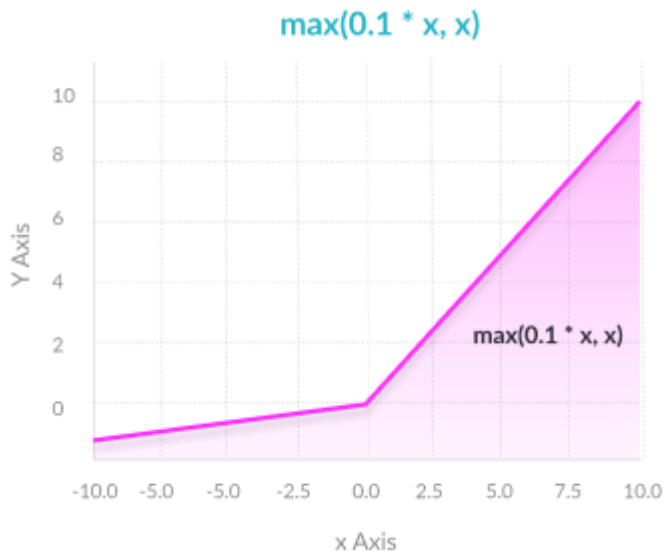

Figure 11: Leaky ReLU Activarion Function and Graph [18]

Advantages
Stops dying ReLU problem — this difference of ReLU has a small helpful slope in the bad area, allowing backpropagation even for negative inputs
As with ReLU, otherwise

Disadvantages
Not reliable results —for the negative input values leaky ReLU does not provide a reliable prediction.

3.5.7 Parametric-ReLU

$$f(x)=max(a\partial,\partial)$$

Figure 12: Parametric-ReLU Function [18]

Advantages
Permits the negative slope to be educated — distinct leaky ReLU, this function offers the slope of the negative part of the function as an argument. And backpropagation can be performed and the most suitable value of α can be learned.
As with ReLU, otherwise

Disadvantages
Can do different performances for different problems.

3.5.6 Softmax

$$\sigma(z)j = \frac{ezj}{\sum_{k=1}^{k} ezk} \quad for\ j = 1....k.$$

Figure 13: Softmax Function [18]

Capable of managing several classes in other activation functions only one class — standardizes the outputs for each session between 0 and 1, and splits by their sum, charitable the probability that the input value is of a particular class.

Valuable for output neurons — softmax is typically castoff for output layer only, for neural networks which are essential to categorize inputs into manifold groups.

3.5.7 Swish

Swish is a fresh, self-contained activation feature that was exposed by Google scholars. It does well than ReLU, with a comparable degree of numerical effectiveness, according to their article. ImageNet accuracy increases to 0.6 to 0.9 with the same model using ReLU and Swish with the new function.

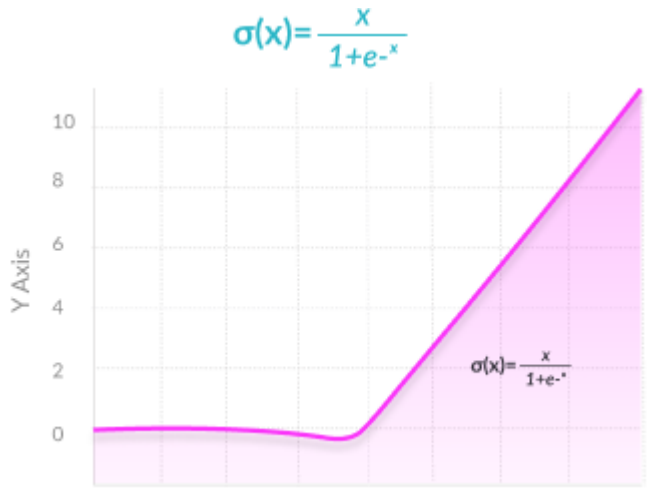

Figure 14: Swish Activation Function and Graph [18]

3.6 What does the machine learn?

It can change form model to model but mainly it learns a function f(x) corresponds to y. It means that words the model to get the value of x and with the help of the x and with that function to forecast the value of y. The model learns two parameters that form the equations y = mx+b. where x is the variable m is the slope and b is the bias.
Bias is equivalent to y when the x is 0. Both are scalar parameters. By multiple inputs of which we have the output we train the model and update the parameters. After sufficient amounts of inputs the model gets trained and we can use this model to predict the y for the new x inputs. [20]

3.7 Learning parameters: Cost functions

The performance of the model is calculated by the cost function for the given data. It calculates the error based on the expected value and the predicted value form the model. After that, it represents it in the form of a single real number. The cost function can be formed in different ways depending upon the problem. There are mainly two purposes of the cost function.
Minimized – Cost, error, or loss are the names given to the returned value. Its work is to find the values of model indicators on which the function gives the smallest value possible.
Maximized – Reward is the name given to the value it yields. Its work is to find the values of model parameters where the function gives the largest value possible. [21]

Chapter 4

Experiment Works and Results

4.1 Data Exploration

We have data that classifies whether or not patients have heart disease based on the attributes in it.
Our emphasis is on constructing a model that attempts to give a predictive analysis of whether a patient suffers from a heart ailment or not.

	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
0	63	1	3	145	233	1	0	150	0	2.3	0	0	1	1
1	37	1	2	130	250	0	1	187	0	3.5	0	0	2	1
2	41	0	1	130	204	0	0	172	0	1.4	2	0	2	1
3	56	1	1	120	236	0	1	178	0	0.8	2	0	2	1
4	57	0	0	120	354	0	1	163	1	0.6	2	0	2	1

Figure 16: Example of data [22]

4.2 Data contains

- → age – age of the patient in years
- → sex – 1 is for male and 0 is for female
- → cp – the types of chest pain
- → trestbps – blood pressure of resting patient (in mm Hg)
- → chol – serum cholesterol in mg/dl
- → fbs – (blood sugar of fasting patient > 120 mg/dl) 1 is for true and 0 is for false
- → restecg – electrocardiographic results in resting patient
- → thalach – heart rate inpatient which is maximum
- → exang – 1 is for yes and 0 is for no (exercise-induced angina)
- → oldpeak – ST depression induced by exercise relative to rest
- → slope – the slope of the peak exercise ST segment
- → ca – number of major vessels (0-3) colored by fluoroscopy
- → thal – 3 is for normal; 6 is for the fixed defect; 7 is for reversible defect
- → target – diseased or not 1 is for yes and 0 is for no.

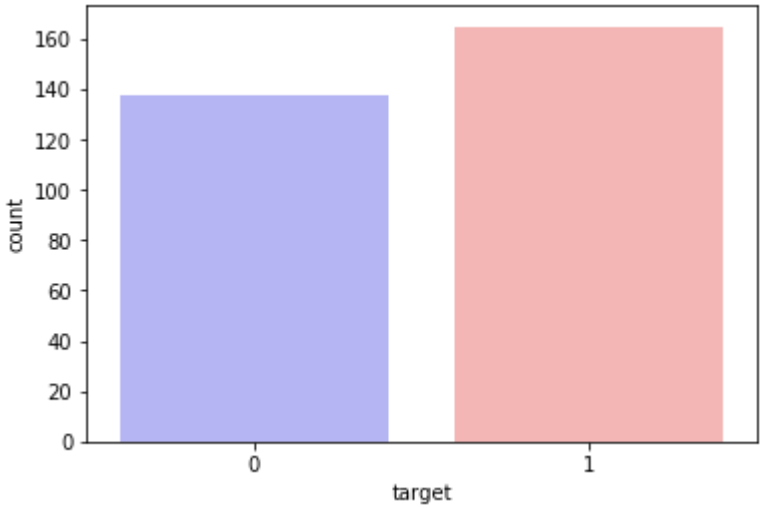

Figure 17: Graphical picture of the data according to the target [23]

In this bar graph, the data is shown for the patient having or not having the heart disease those are 54.46% is diseased and 45.54% is not diseased.

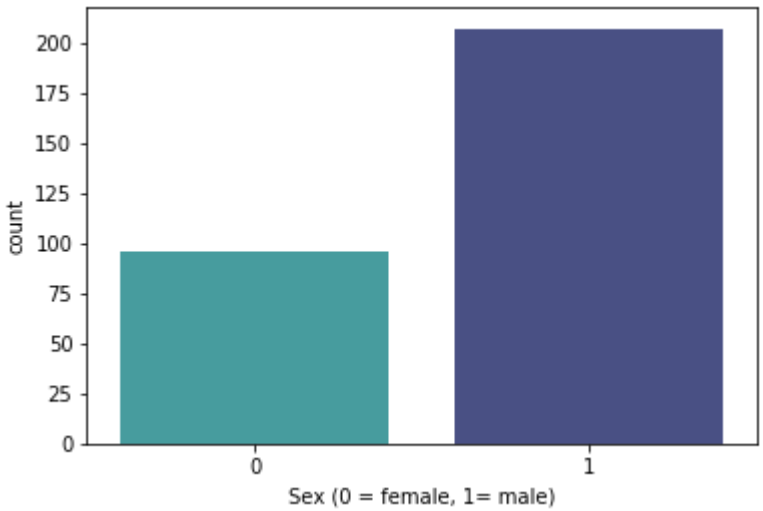

Figure 18: Graphical picture of the data according to the male and female in data. [23]

In this bar graph, the data is shown for the patient base on the sex those are 31.68% is female and 68.32% is male.

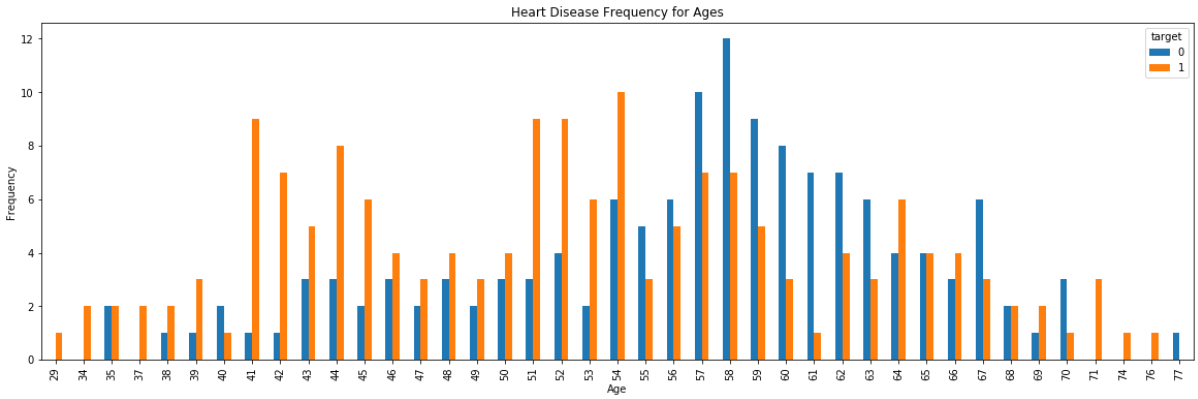

Figure 19: Graphical picture of the data according to the age [23]

In this bar graph, the data is shown for the patient base on age and sex.

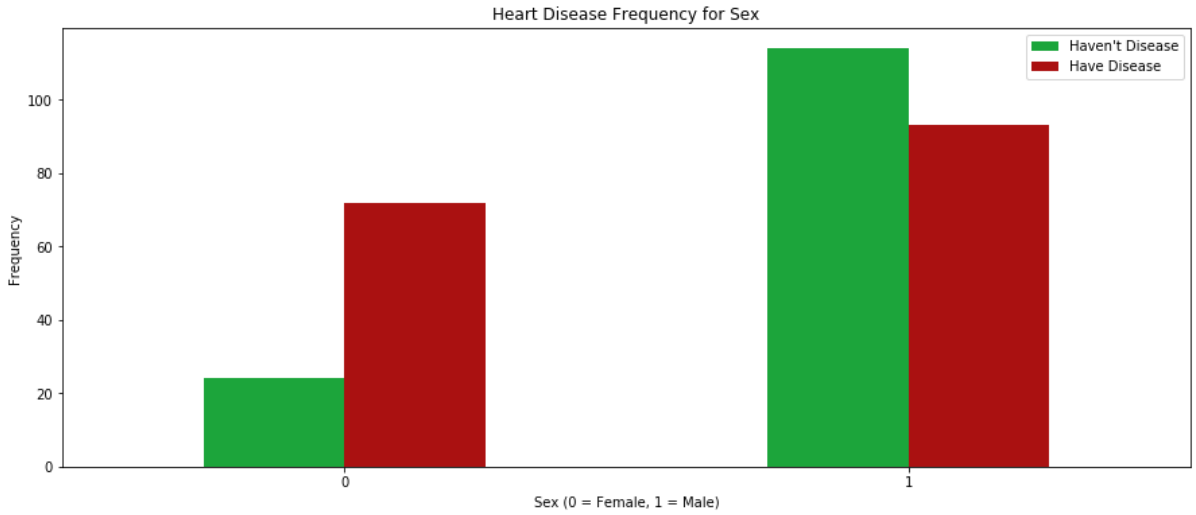

Figure 20: Graphical picture of the data according to the sex. [23]

In this bar graph, the data is shown for the patient base on the disease and sex.

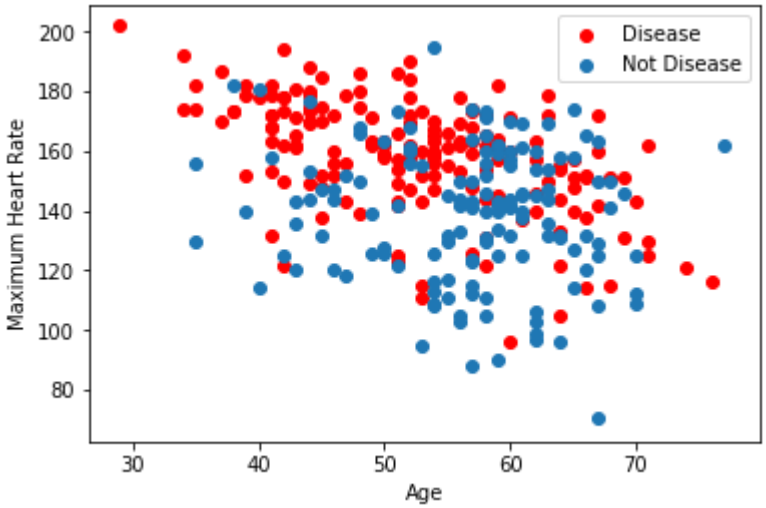

Figure 21: Graphical picture of the data according to the heart rate. [23]

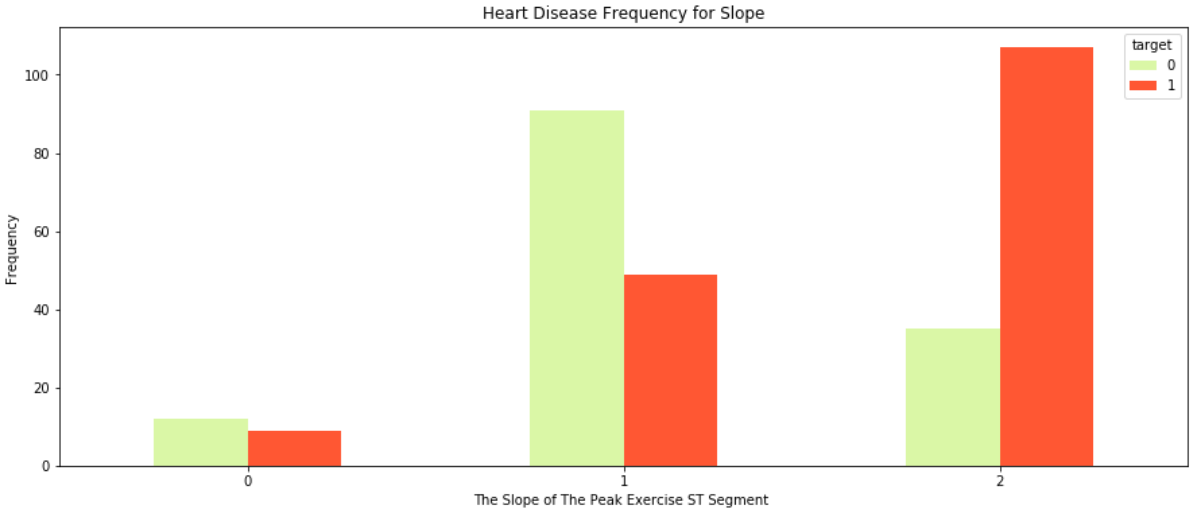

Figure 22: Graphical picture of the data according to the ST segment. [23]

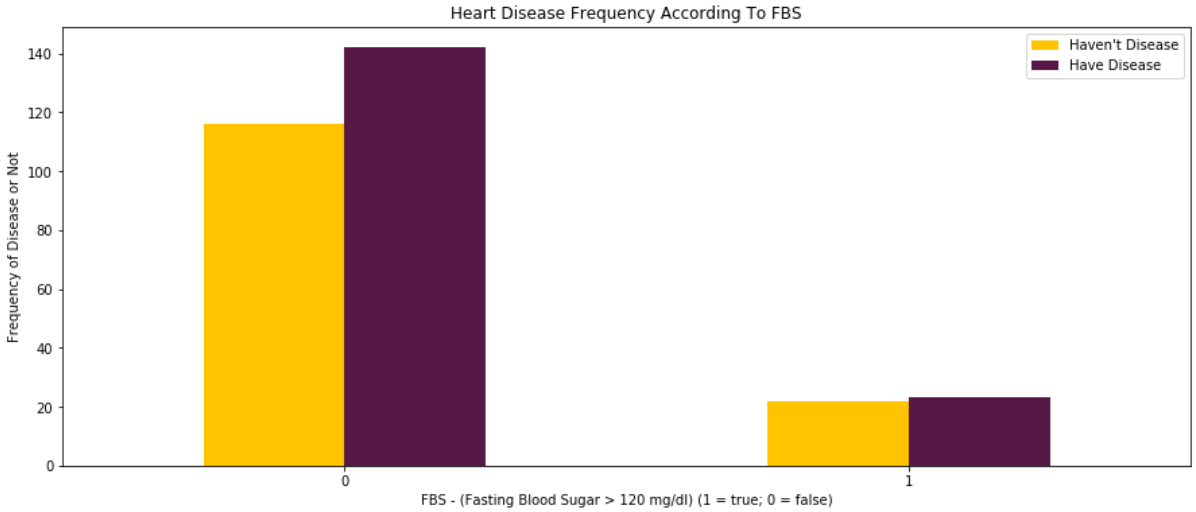

Figure 23: Graphical picture of the data according to the FBS(Fasting Blood Sugar) [23]

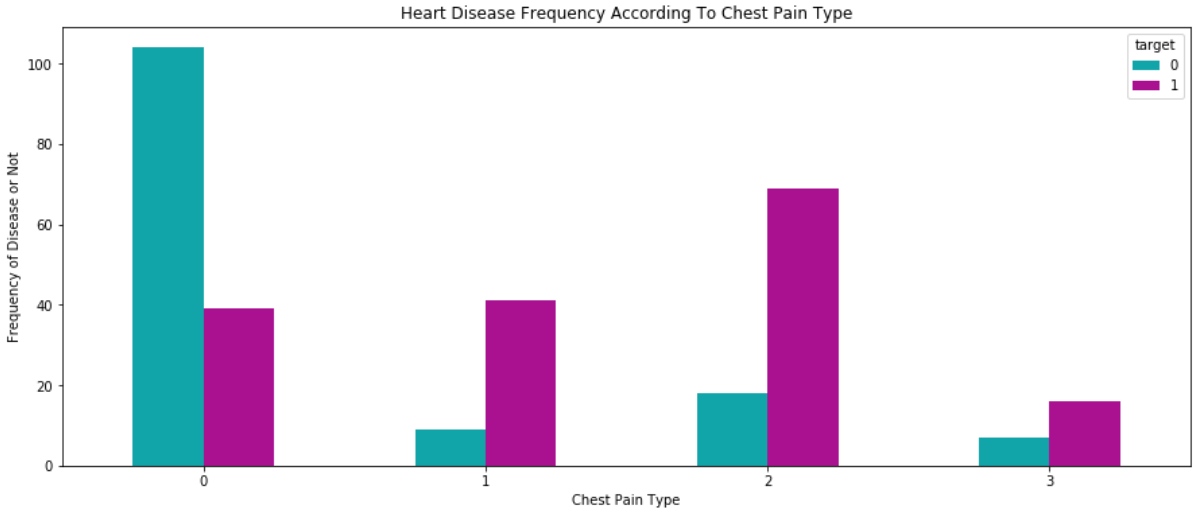

Figure 24: Graphical picture of the data according to the chest pain type. [23]

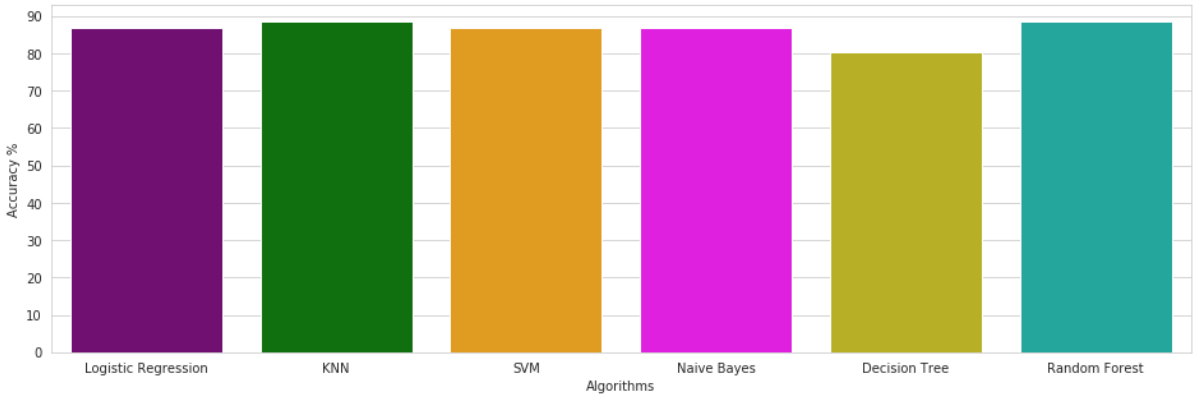

Figure 25: Graphical picture of the accuracy achieved using different models. [23]

Previously this model uses the cross-entropy loss function and the graph we get after the execution of the program is shown below. The cost we get is 0.33720228563306326 on 10000 iterations.

$$- (y \log(p) + (1-y) \log(1-p))$$

Where y is a binary indicator
p is the prediction

```
iteration: 10000
cost: 0.33720228563306326
```

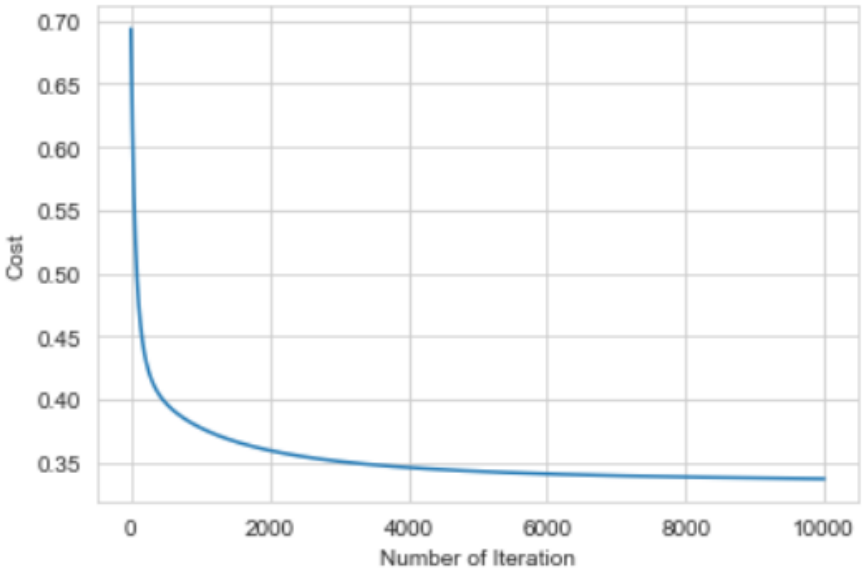

Figure 26: Result of old cost function and graph [23]

While using the updated cost function the loss becomes less and drops to -1.7540951990199793 on 10000 iterations using the below-cost function.

$$- \quad (1 – \log(y – p))$$

Where y is a binary indicator
p is the prediction

```
iteration: 10000
cost: -1.7540951990199793
```

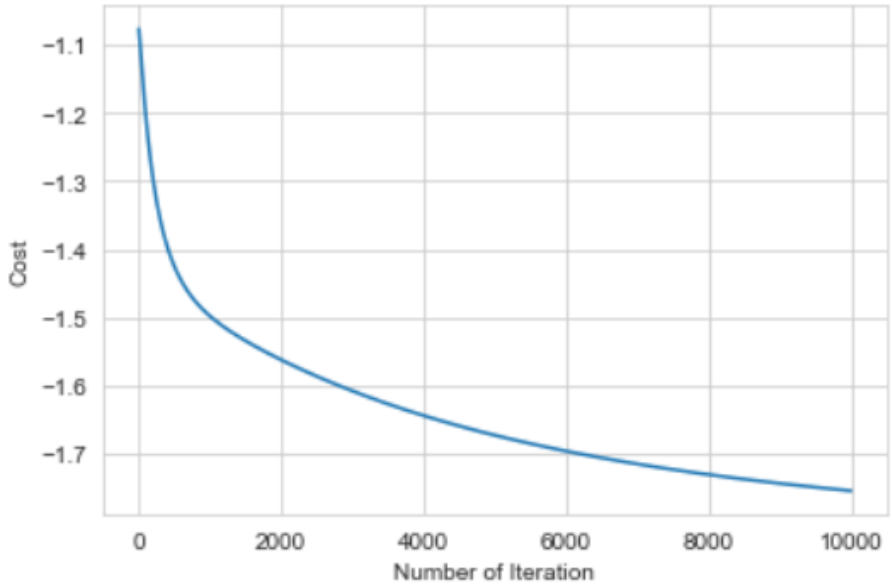

Figure 27: Result from the new cost function and its graph

Table 2: Comparison between the cost of cross-entropy loss and new loss function.

Cost using a cross-entropy loss function	0.33720228563306326
Cost using a new loss function	-1.7540951990199793

Chapter 5

Conclusion and Future Work

Conclusion

In this mode, we have been working on the updating of the cost function. We have seen that we are using our updated model in which we using our new loss function. We observed that the loss become less in comparison to the previous loss function. The previous loss function that is used is a cross-entropy loss function.

We have not been able to improve the accuracy using this loss function of ours. But it decreases the loss to a greater extend. So it can be a useful function for other models too. Maybe it will help decrease the loss and improve the accuracy in some other models.

Future Work

The new cost function used in this model can be used in other models too. We can have the possibility to improve the accuracy and the time to calculate the result. We have seen the cost function reduces the cost of the model which can be a great help to the result calculation in the other models.
We have not seen any improvement in the present model but it may help to improve the other model. So we have to try the cost function in other models. It may lead to some great results.

References

[1] D. J. P. Barker, “Fetal origins ofcoronary heart disease.”

[2] “Heart disease - Symptoms and causes - Mayo Clinic.” [Online]. Available: https://www.mayoclinic.org/diseases-conditions/heart-disease/symptoms-causes/syc-20353118. [Accessed: 27-Jun-2020].

[3] “Sample Data for Gender and Recommendation for Remedial Reading Instruction.”

[4] L. Baoli, Y. Shiwen, and L. Qin, “An Improved k-Nearest Neighbor Algorithm for Text Categorization 1.”

[5] I. Dokmani´c, R. Parhizkar, J. Ranieri, and M. Vetterli, “Euclidean Distance Matrices Essential Theory, Algorithms and Applications.”

[6] D. Sinwar, R. Kaushik, and M. Tech Scholar, “Study of Euclidean and Manhattan Distance Metrics using Simple K-Means Clustering,” *www.ijraset.com*, vol. 2, 2014.

[7] S. Craw, “Manhattan Distance,” in *Encyclopedia of Machine Learning and Data Mining*, Springer US, 2017, pp. 790–791.

[8] H. B. Çolakoğlu, “A generalization of the Minkowski distance and a new definition of the ellipse,” Mar. 2019.

[9] M. Norouzi, D. J. Fleet, and R. Salakhutdinov, “Hamming Distance Metric Learning.”

[10] G. Cauwenberghs and T. Poggio, “Incremental and Decremental Support Vector Machine Learning.”

[11] K. Vembandasamyp, R. R. Sasipriyap, and E. Deepap, “Heart Diseases Detection Using Naive Bayes Algorithm,” 2015.

[12] Y. Ben-Haim and Y. I. Com, “A Streaming Parallel Decision Tree Algorithm Elad Tom-Tov,” 2010.

[13] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, “An assessment of the effectiveness of a random forest classifier for land-cover classification,” 2012.

[14] R. Hecht-Nielsen, “Theory of the Backpropagation Neural Network.”

[15] T. L. Fine, “Fundamentals of Artificial Neural Networks [Book Reviews],” *IEEE Trans. Inf. Theory*, vol. 42, no. 4, p. 1322, 2005.

[16] K. M. M. Anil K. Jain, Jianchang Mao, “Artificial Neural Networks - A Tutorial-Institute of Electrical & Electronics Enginee (1996).” p. 14, 1996.

[17] P. Sibi, S. A. Jones, and P. Siddarth, “ANALYSIS OF DIFFERENT ACTIVATION FUNCTIONS USING BACK PROPAGATION NEURAL NETWORKS,” *J. Theor. Appl. Inf. Technol.*, vol. 31, no. 3, 2013.

[18] “7 Types of Activation Functions in Neural Networks: How to Choose?” [Online]. Available: https://missinglink.ai/guides/neural-network-concepts/7-types-neural-network-activation-functions-right/. [Accessed: 30-Jun-2020].

[19] A. Wanto, A. P. Windarto, D. Hartama, and I. Parlina, “Use of Binary Sigmoid Function And Linear Identity In Artificial Neural Networks For Forecasting Population Density,” *IJISTECH (International J. Inf. Syst. Technol.*, vol. 1, no. 1, p. 43, Nov. 2017.

[20] J. Burrell, “How the machine ‘thinks’: Understanding opacity in machine learning algorithms,” *Big Data Soc.*, vol. 3, no. 1, p. 205395171562251, Jan. 2016.

[21] C. W. J. Granger, “Prediction with a Generalized Cost of Error Function,” *J.*

Oper. Res. Soc., vol. 20, no. 2, pp. 199–207, Jun. 1969.
[22] "Heart Disease - Classifications (Machine Learning) | Kaggle." [Online]. Available: https://www.kaggle.com/cdabakoglu/heart-disease-classifications-machine-learning. [Accessed: 30-Jun-2020].
[23] "Heart Disease - Classifications (Machine Learning) | Kaggle." [Online]. Available: https://www.kaggle.com/cdabakoglu/heart-disease-classifications-machine-learning. [Accessed: 26-Jun-2020].

LIST OF PUBLICATION BY CANDIDATE

[1] Accepted: Advantages of Capsule network over the convolutional neural network: A review