TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING

PULCHOWK CAMPUS

THESIS NO: 2072/MSCS-658

NEURAL NETWORK BASED CORONARY ARTERY DISEASE PREDICTION USING FEATURE SELECTION BY RANDOM FOREST AND DOMAIN EXPERT

BY:

NEERAV ADHIKARI

A THESIS

SUBMITTED TO THE DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE IN COMPUTER SYSTEMS AND KNOWLEDGE ENGINEERING

DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING

November, 2019

NEURAL NETWORK BASED CORONARY ARTERY DISEASE PREDICTION USING FEATURE SELECTION BY RANDOM FOREST AND DOMAIN EXPERT

by:

Neerav Adhikari

2072-MSCS-658

Thesis Supervisor:

Dr. Aman Shakya

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Computer Systems and Knowledge Engineering

Department of Electronics and Computer Engineering
Institute of Engineering, Pulchowk Campus
Tribhuwan University
Lalitpur, Nepal

November, 2019

Copyright ©

The author has agreed that the library, Department of Electronics and Computer Engineering, Institute of Engineering, Pulchowk Campus, may make this thesis freely available for inspection. Moreover the author has agreed that the permission for extensive copying of this thesis work for scholarly purpose may be granted by the professor(s), who supervised the thesis work recorded herein or, in their absence, by the Head of the department, wherein this thesis was done. It is understood that the recognition will be given to the author of this thesis and to the Department of Electronics and Computer Engineering, Pulchowk Campus in any use of the material of this thesis. Copying of publication or other use of this thesis for financial gain without approval of the Department of Electronics and Computer Engineering, Institute of Engineering, Pulchowk Campus and author's written permission is prohibited.

Request for permission to copy or to make any use of the material in this thesis in whole or part should be addressed to:

Head
Department of Electronics and Computer Engineering
Institute of Engineering, Pulchowk Campus
Pulchowk, Lalitpur, Nepal

TRIBHUVAN UNIVERSITY

INSTITUTE OF ENGINEERING PULCHOWK CAMPUS

DEPARTMENT OF ELECTRONICS AND COMPUTER ENGINEERING

Recommendation

The undersigned certify that they have read and recommended to the Department of Electronics and Computer Engineering for acceptance, a thesis entitled "**Neural Network Based Coronary Artery Disease Prediction Using Feature Selection by Random Forest and domain Expert**", submitted by **Neerav Adhikari** in partial fulfillment of the requirement for the award of the degree of "**Master of Science in Computer Systems and Knowledge Engineering**".

....

Supervisor: Dr. Aman Shakya

Department of Electronics and Computer Engineering,

Institute of Engineering, Pulchowk Campus

....

External Examiner

Mr. Bikash Bahadur Shrestha

....

Committee Chairperson, Dr. Aman Shakya

Program Coordinator

Department of Electronics and Computer Engineering

Date …………………………………….

Departmental Acceptance

The thesis entitled "**Neural Network Based Coronary Artery Disease Prediction Using Feature Selection by Random Forest and domain Expert**", submitted by **Mr. Neerav Adhikari** in partial fulfillment of the requirement for the award of the degree of "Master of Science in Computer System and Knowledge Engineering" has been accepted as a bona-fide record of work independently carried out by him in the department.

Dr. Surendra Shrestha

Head of the Department

Department of Electronics and Computer Engineering,

Pulchowk Campus, Institute of Engineering,

Tribhuvan University, Nepal.

Abstract

Coronary Artery Disease (CAD) has become very common nowadays and is the leading cause of death across the world. According to WHO data published in May 2017 death caused by CAD in Nepal reached 30,559 which is 18.72 % of total death. Angiography is often regarded as best and more accurate method for CAD diagnosis in hospitals however it is more costly and has many side effects.

An attempt is being made here for developing a better modal for the classification of CAD by using the clinical, ECG and laboratorial features of patients. Optimum features were selected with the help of domain expert and random forest (mean decreasing accuracy) method. Neural Network based classification model was developed with the selected features for the classification of Left Anterior Descending artery (LAD), Left Circumflex artery (LCX) and Right Coronary Artery (RCA) in human heart. To get the optimum result from the neural network model with 10- fold cross validation method was adopted.

The optimized model reached its maximum accuracy of 91.397 % for LAD, 88.09 % for LCX and 90.36 % for RCA classification. Similar new model with common 22 features as input was developed with the data collected from TUTH, Manmohan Cardiovascular and Transplant Centre IOM. Maximum accuracy of 65 % for LAD, 66.67 % for LCX and 60 % for RCA classification was achieved.

Key Words: Coronary Artery Disease (CAD), Left Anterior Descending artery (LAD), Left Circumflex artery (LCX), Right Coronary Artery (RCA)

Acknowledgement

I would like to express my sincere thanks to my respected Supervisor also the Program Coordinator of MSCS, Department of Electronics and Computer Engineering, Pulchowk Campus, **Dr. Aman Shakya**, for granting his precious time to provide feedback, comments, valuable suggestion, encouragement and continuous support during my research work.

I would further take the opportunity to convey my gratitude to Head of Department **Dr. Surendra Shrestha**. Furthermore, I would like to thank **Prof. Dr. Shashidhar Ram Joshi, Prof. Dr. Subarna Shakya**, **Dr. Dibakar Raj Panta** and **Dr. Sanjeeb Pd. Panday, Dr. Basanta Joshi** for their valuable comments and guidance during my study.

I would also like to express my sincere gratitude to **Dr. Ravi Sahi** (Cardiology TUTH) and **Dr. Amrit Bogati** (Cardiologist, Shahid Gangalal National Heart Centre) for providing valuable insights and continuous support to during my research for the selection of important features and providing data sample for my research work.

I would like to thank all my classmates who shared their views, ideas and valuable suggestion during my work. I would also like to thank Department of Electronics and Computer Engineering, Pulchowk Campus, for providing necessary resources and information during my research.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ANN	Artificial Neural Network
BMI	Body Mass Index
CAD	Coronary Artery Disease
CRF	Chronic Renal Failure
CVA	Cerebrovascular Accident
DM	Diabetes Mellitus
FBS	Fasting Blood Sugar
HBP	High Blood Pressure
HDL	High Density Lipoprotein
LAD	Left Anterior Descending artery
LCX	Left Circumflex artery
LVH	Left Ventricular Hypertrophy
MDA	Mean Decreasing Accuracy
RCA	Right Coronary Artery
RPROP	Resilient Back Propagation
SVM	Support Vector Machine
TG	Triglyceride
WHO	World Health Organization

CHAPTER I: INTRODUCTION

1.1 Background and Motivation

Heart disease describes a range of conditions that affect our heart. Diseases under the heart disease umbrella include blood vessel diseases, such as coronary artery disease; heart rhythm problems (arrhythmias); and heart defects someone born with (congenital heart defects), among others. The term "heart disease" is often used interchangeably with the term "Cardiovascular Disease (CAD)". Cardiovascular disease generally refers to conditions that involve narrowed or blocked blood vessels that can lead to a heart attack, chest pain (angina) or stroke. Other heart conditions, such as those that affect your heart's muscle, valves or rhythm, also are considered forms of heart disease. Cardiovascular Disease is the number one killer in the world responsible for 17.7 million death per year representing to 31% of all global death. [1]

Once CAD was thought to be the problem of developed countries however today it is a global health problem adding extra burden in developing countries like Nepal. Hypertension, Diabetes, physical inactivity are conventional risk factors of CAD. Different studies have shown high prevalence of these risk factors. The prevalence of hypertension in four different geographic regions of Nepal showed highest rate in urban communities of Kathmandu (9.98%), followed by rural Terai (8.11%), and the mountain region, Jumla (5.3%) [2]. According to the latest WHO data published in May 2017 Coronary Artery Disease death in Nepal reached 30,559 which is 18.72 % of total death. The age adjusted death rate is 158.35 per 100,000 of population, which is ranked at 41 in the world. [3]

Healthcare industry (Hospital/Labs) collects large amounts of data which are not mined to discover hidden information for making decision. With the increasing access to huge datasets and corresponding demands to analyze these data has led to the development of new algorithms for performing machine learning on data streams. Different data mining technique with enormous investigations has been done in this filed till date. Since these healthcare date are unstructured and heterogeneous with lots of attributes in it so it should be analyzed to predict to provide information for making diagnosis with better decision support system in the health care industry. As there is uncertainty in manual prediction system, an attempt is being made here by introducing a random forest method and domain experts(doctors) suggesting for important feature selection and the using resilient back propagation method for predicting the presence coronary artery disease in patients. The main aim of this thesis work is to suggest an automated diagnosis of coronary artery diseases i.e. blockage of arteries Left Anterior Descending artery (LAD), Left Circumflex artery (LCX) or Right Coronary Artery (RCA) by taking into account earlier information and data. Healthcare organizations can reduce costs by accomplishment of computer based data and/or decision support systems by using this model.

1.2 Problem Statement

Often the success of any machine learning project depends on the features used. Much human effort has been spend on designing good features selection method which are usually knowledge-based and engineered by domain experts over years of trial and error, as data might contain noise and outliers, which greatly affect the performance of the model. When choosing algorithms, one may need to verify which algorithm or parameters work best on the dataset.

There are many algorithm for selecting the important features from the data set. However while only considering the machine learning or domain experts may leads to poor classification. Following are the challenges found for better classification of coronary artery disease:

- Researches works have been done using different machine learning approach for predicting of Coronary Artery Disease (CAD), but they didn’t include the consultation with heart cardiologist or specialist for better understanding of features and their association with CAD along with Random Forest based feature selection.

- Most of the research work has been done in prediction of coronary artery disease but very few work has been done for the blockage for each artery (LAD, LCX & RCA) independently.

1.3 Objective:

The objective of this study are

1. To develop the classification model using neural network for better prediction of coronary artery disease considering the feature selection based on domain expert and random forest.
2. To develop neural network model for the prediction of blockage in three arteries (LAD, LCX and RCA) of human heart.

1.4 Scope of Work

This system is concerned on developing better classification model for forecasting the coronary artery disease in human body using random forest and domain expert based feature selecting technique. This Neural Network based classification model is expected to have the potential to generate a knowledge-rich environment which can help to reduce medical errors, enhance patient safety and significantly improve the quality of clinical decisions system for patients with coronary artery disease.

The system will also classify the individual blockage in each arteries i.e. Left Anterior Descending artery (LAD), Left Circumflex artery (LCX) or Right Coronary Artery (RCA) of human heart. The proposed model can also be used as a reference for medical practitioner and health care provider from the same field.

CHAPTER II: LITERATURE REVIEW

This section explains the human circulatory system and a gentle introduction to coronary arteries in human body. It also review of works related to Coronary Artery Disease (CAD) using different methods for different periods of time where each review is divided into different sections according the proposed method that they used for forecasting:

2.1 Human Circulatory System and Anomalies

The human heart is an amazing organ that continuously pumps oxygen and nutrient-rich blood throughout your body to sustain life. Pumped blood circulates throughout the body via the circulatory system, supplying oxygen and nutrients to the tissues and removing carbon dioxide and other wastes. This fist-sized powerhouse beats (expands and contracts) 100,000 times per day, pumping five or six quarts of blood each minute, or about 2,000 gallons per day. [4]

There are three main types of blood vessels:

- **Arteries**: They begin with the aorta, the large artery leaving the heart. Arteries carry oxygen-rich blood away from the heart to all of the body's tissues. They branch several times, becoming smaller and smaller as they carry blood farther from the heart.
- **Capillaries:** These are small, thin blood vessels that connect the arteries and the veins. Their thin walls allow oxygen, nutrients, carbon dioxide, and other waste products to pass to and from our organ's cells.
- **Veins:** These are blood vessels that take blood back to the heart; this blood lacks oxygen (oxygen-poor) and is rich in waste products that are to be excreted or removed from the body. Veins become larger and larger as they get closer to the heart. The superior vena cava is the large vein that brings blood from the head and arms to the heart, and the inferior vena cava brings blood from the abdomen and legs into the heart.

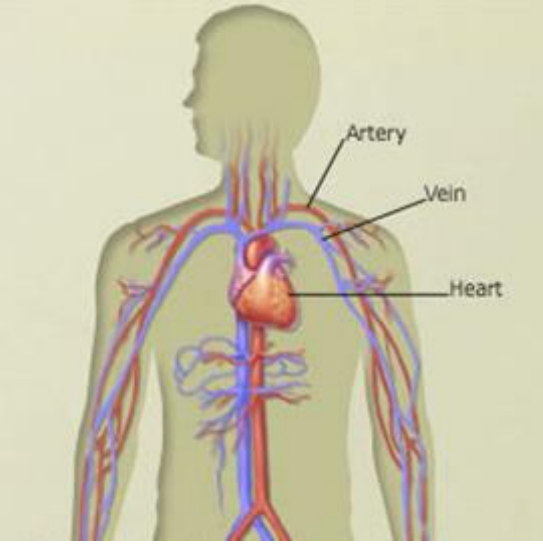


Figure 1 : Human Circulatory System.

https://www.webmd.com/heart-disease/high-cholesterol-healthy-heart#1

© 2005 - 2019 WebMD LLC

2.1.1 Human heart anatomy

In humans, the heart is roughly the size of a large fist and weighs between about 10 to 12 ounces (280 to 340 grams) in men and 8 to 10 ounces (230 to 280 grams) in women, according to Henry Gray's "Anatomy of the Human Body." The human heart has four chambers: two upper chambers (the atria) and two lower ones (the ventricles). The right atrium and right ventricle together make up the "right heart," and the left atrium and left ventricle make up the "left heart." A wall of muscle called the septum separates the two sides of the heart. [5]

2.1.2 Human heart function

The heart circulates blood through two pathways: the **Pulmonary Circuit** and the **Systemic Circuit**.

In the **pulmonary circuit,** deoxygenated blood leaves the right ventricle of the heart via the pulmonary artery and travels to the lungs, then returns as oxygenated blood to the left atrium of the heart via the pulmonary vein.

In the **systemic circuit**, oxygenated blood leaves the body via the left ventricle to the aorta, and from there enters the arteries and capillaries where it supplies the body's tissues with oxygen. Deoxygenated blood returns via veins to the venae cave, re-entering the heart's right atrium.

2.1.3 Coronary Arteries

Heart is made of tissue that requires a supply of oxygen and nutrients. Although its chambers are full of blood, the heart receives no nourishment from this blood. The heart receives its own supply of blood from a network of arteries, called the **coronary arteries**. Two major coronary arteries branch off from the aorta near the point where the aorta and the left ventricle meet. These arteries and their branches supply all parts of the heart muscle with blood

These two major branched coronary arteries are:

- **Right Coronary Artery(RCA)**
  - The main portion of the right coronary artery provides blood to the right side of the heart.
  - Run across the surface of the heart's underside and provide blood circulation to the bottom portion of both ventricles and back of the septum.
  - The right side of the heart is smaller because it pumps blood only to the lungs.

- **Left main Coronary Artery**: The left side of the heart is larger and more muscular because it pumps blood to the rest of the body. It branches into the **Circumflex Artery (LCX)** and the **Left Anterior Descending Artery (LAD).**

  - **Left Circumflex Artery(LCX) -** supplies blood to the left atrium, side and back of the left ventricle
  - **Left Anterior Descending artery (LAD) -** supplies the front and bottom of the left ventricle and the front of the septum.

The heart muscle, like every other organ or tissue in your body, needs oxygen-rich blood to survive. Blood is supplied to the heart by its own vascular system, called **Coronary Circulation**.[6]

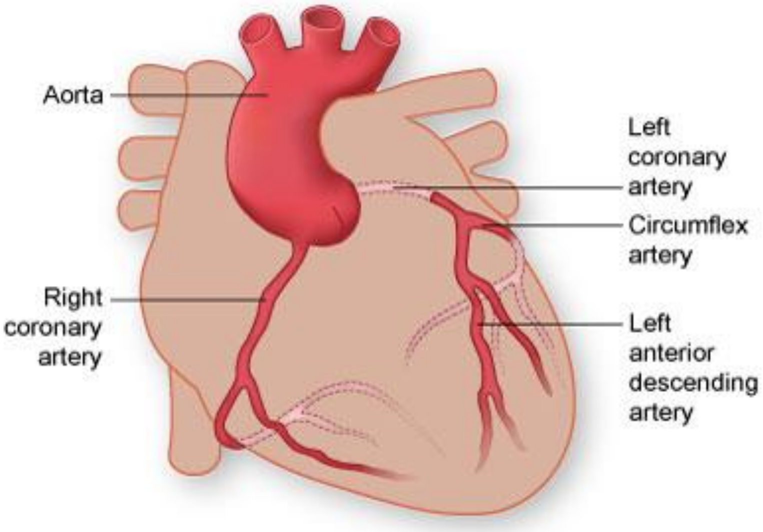


Figure 2 : The Human Heart. https://www.texasheart.org/wpcontent/uploads/2017/12/thi-coronary-arteries.jpg © Texas Heart Institute

2.1.4 Coronary Artery Disease

Coronary artery disease (CAD) or coronary heart disease is atherosclerosis (plaque in artery walls) of the inner lining of the blood vessels that supply blood to the heart. Coronary artery disease is a common form of heart disease and is a major cause of illness and death. Coronary artery disease begins when hard cholesterol substances (plaques) are deposited within a coronary artery. The coronary arteries arise from the aorta, which is adjacent to the heart. The plaques narrow the internal diameter of the arteries which may cause a tiny clot to form which can obstruct the flow of blood to the heart muscle. [7]

Too much plaque buildup and narrowed artery walls can make it harder for blood to flow through your body. When heart muscle doesn't get enough blood, one may have chest pain or discomfort, called angina. Over time, CAD can weaken the heart muscle. This may lead to heart failure, a serious condition where the heart can't pump blood the way that it should. An irregular heartbeat, or arrhythmia, also can develop. [8]

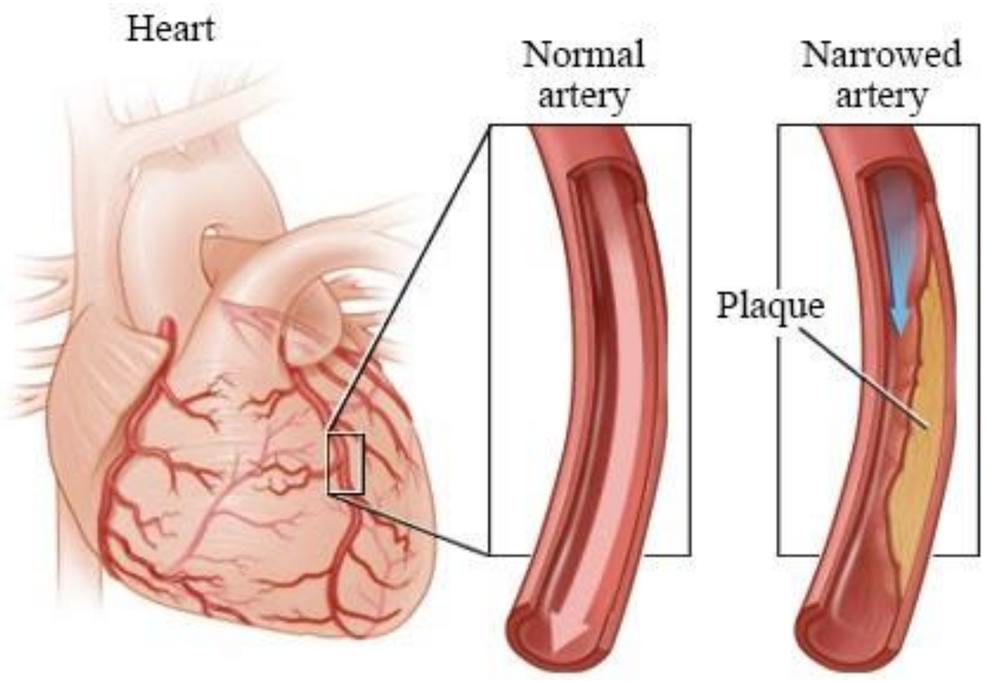


Figure 3 : Narrowed or blocked coronary artery https://metrohealth.net/healthwise/coronary-angioplasty/ © 1995-2018 Healthwise, Incorporated

Collateral Circulation is a network of tiny blood vessels, and, under normal conditions, not open. When the coronary arteries narrow to the point that blood flow to the heart muscle is limited (coronary artery disease), collateral vessels may enlarge and become active. This allows blood to flow around the blocked artery to another artery nearby or to the same artery past the blockage, protecting the heart tissue from injury. [9]

2.2 Random forest based prediction

Min Liu, Xiaowei Xu and Ye Tao and Xiadong Wang (2017) proposed "An Improved Random Forest Method Base on RELIEFF for Medical Diagnosis". Paper presents a hybrid approach that combines the Relief-F algorithm (RA) and Random Forest for modeling and analysis of complex disease, with improved speed and stability. [10] The dataset consists of patients' health information from 2010 to 2014 at a hospital. It includes 17,337 samples, 14 features and 1 class label for 11 disease. The method was tested with eleven datasets from real medical records. To increase the reliability of the result Relief-F, features selection algorithm was run 20 time. The result was taken as average weight of the each features. Experimental results prove that the proposed method improves the accuracy of medical data classification for illness diagnosis. In their model they conducted two experiments, in 1st experiment Relief-F algorithm (RA) was used for feature selection and then a model was constructed with Random Forest. Building a single model for all eleven kind of disease show very poor result with maximum accuracy of 33.57 % only when experiment was run for 100 times. In their 2nd experiment individual model was made for each disease and the accuracy was increased distinctly. Maximum accuracy achieved by this method for coronary artery disease was at around 90.28 % with only 5 features (Age, Fasting Blood Glucose, Postprandial Blood, Glucose (PBG), and Low-density Lipoprotein (LDL).

Burak Kolukisa, Hilal Hacilar, Gokhan Goy and Burcu Bakir-Gungor (2018) proposed a "Evaluation of Classification Algorithms, Linear Discriminant Analysis and a New Hybrid Feature Selection Methodology for the Diagnosis of Coronary Artery Disease".[11] The data set used were from Cleveland, South Africa and Z-Alizadehsani. In order to develop robust model different features selection methods were used like Information Gain (IG) and Gain Ratio (GR), Relief-F (RF) and Chi-Squared (CS) test. In their study they found that feature selection method improves the performance and accuracy of the classifier. There were also the case when feature selection method didn't work example is for South African data set. Maximum accuracy of 87.12 % with Sensitivity of 92.6 and Specificity 98 % was achieved using features like Typical Chest Pain, Exertional CP, Q Wave, Region with RWMA, Age, Sex, Weight, BMI, Obesity, DM, FBS, HTN, BP, Current Smoker, Ex-Smoker, FH, LDL, HDL with Random Forest.

2.3 Naïve Bayes based prediction

Roohallah Alizadehsani, Jafar Habibi, Mohammad Javad Hosseini, and Z-Alizadeh Sani (2016) proposed “A Data Mining Approach for Diagnosis of Coronary Artery Disease” the data set used in this paper is gathered from 303 random visitors to Tehran’s Shaheed Rajaei Cardiovascular, Medical and Research Center where each set has 54 feature.[12] For feature selection method Information Gain along with feature creating algorithm were applied to increase the accuracy of the proposed model. Maximum accuracy archived by Naïve Bays method was around 47.87% with Sensitivity of 28.70 and Specificity 95.40 %. The features selected were Typical Chest Pain Region RWMA, Age, EF, HTN, DM, T-inversion , ESR Q wave, ST elevation ,PR ,BMI ,Lymph ,BP ,Dyspnea, HDL, CR, WBC, Weight ,Function Class ,Airway disease ,HB, TG, BBB, Na, Sex, LVH, Hb , FH.

Gokhan Goy, Mustafa Kus, and Burcu Bakir-Gungor (2018) proposed a “Evaluation of Classification Algorithms, Linear Discriminant Analysis and a New Hybrid Feature Selection Methodology for the Diagnosis of Coronary Artery Disease”.[11] The data set used were from Cleveland, South Africa and Z-Alizadehsani. In order to develop robust model different features selection methods were used like Information Gain (IG) and Gain Ratio (GR), Relief-F (RF). Although different features selection methods were there maximum accuracy for Naïve Bayes was archived using features selected by Fremingham Heart Study with following feature: Typical Chest Pain, Exertional CP, Q Wave, Region with RWMA, Age, Sex, Weight, BMI, Obesity, DM, FBS, HTN, BP, Current Smoker, Ex-Smoker, FH, LDL, HDL. Accuracy archived by Naïve Bays method was around 83.49% with Sensitivity of 86.7 and Specificity 83.3 %.

2.4 Bagging based prediction

Zahra Alizadeh Sani, Roohallah Alizadehsani , Jafar Habibi proposed a comparative study “Diagnosing Coronary Artery Disease via Data Mining Algorithms by Considering Laboratory and Echocardiography Features”.[13] Different features were measured and collected from potential patients in order to construct a dataset, which was later utilized for model extraction. The data were gathered from 303 random visitors to Rajaie Cardiovascular Medical and Research center. Among the total attributes feature selection method (Gini Index and Information Gain) was used for feature selection from the collected data and then the C4.5 and Bagging Algorithm were used to forecast the LAD (Left Anterior descending) LCX (Left Circumflex) and RCA (Right Coronary Artery). Use of Feature selection based on Information gain enhance the accuracy for LAD and LCX but it has opposite effect on LCX. For LAD 79.54% of accuracy with feature selection based on information gain and Bagging was found and for RCA 68.95 % of accuracy with feature selection based on Gini Index and bagging was found, whereas for LCX maximum accuracy of 65.09 with no feature selection and bagging method was found.

2.5 Support Vector Machine based prediction

Haleh Ayatollahi and Leila Gholamhosseini (2019) proposed a comparative study “Predicting coronary artery disease: a comparison between two data mining algorithms” They used same data set and for ANN and SVM Classified for the prediction of disease.[14] Data collected were from medical records of the patients with coronary artery disease who were hospitalized in three hospitals affiliated to AJA University of Medical Sciences between March 2016 and March 2017 (n=1324). Totally, 25 variables affecting CAD were selected and related data were extracted. Checklist for dataset was designed based on the variables used in the guideline of the Cleveland heart disease dataset policy in UCI (University of California) repository. The attributes used were gender, age, weight, marital status, occupation, address, family history, smoking, comorbidity, diabetes, pulse rate, T.S.T waves, high blood pressure (HBP), cholesterol, triglyceride (TG), hemoglobin (Hgb), blood glucose level, creatinine, systolic blood pressure, diastolic blood pressure, chest pain, low density lipoprotein (LDL), high density lipoprotein (HDL), CAD diagnosis, and the length of hospitalization. They used simple neural network with 25 neurons in input layer multiple hidden layer and 2 neuron in output layer. SVM algorithm presented higher accuracy and better performance than the ANN model and was characterized with higher power and sensitivity. Overall, SVM provided a better classification for the prediction of CAD.

2.6 Neural Network based prediction

František Babič and Jaroslav Olejár (2017) proposed a "Predictive and Descriptive Analysis for Heart Disease Diagnosis". [15] The data set taken for their study was from Z-Alizadeh Sani, Cleveland and South African Heart disease data set. Different algorithms were like Naïve Bayes, Support Vector Machine and Network were developed for the classification of the model. For each algorithm 10 times repeated experiments were performed with maximum accuracy of 86.32% in Neural Network model for Z-Alizadeh Sani data set. Features used were Age, Diabetes Mellitus, Fasting Blood Suger(mg/dl), Pulse Rate, ST- Elevation, ST-Depression, Low-Density Lpoprotein(mg/dl), White Blood Cell, Obesity, Creatine(mg/dl), Ex- Smoker, Hemoglobin(g/dL),, Non-angina CP, and Heart Disease (CAD).

Roohallah Alizadehsani, Jafar Habibi, Mohammad Javad Hosseini, and Z-Alizadeh Sani (2016) proposed "A Data Mining Approach for Diagnosis of Coronary Artery Disease" the data set used in this paper is gathered from 303 random visitors to Tehran's Shaheed Rajaei Cardiovascular, Medical and Research Center where each set has 54 feature.[12] Different Feature selection method like Information Gain, Gini Index along with feature creating algorithm were applied to increase the accuracy. Differed classification method SVM, Sequential Minimal Optimization (SMO), ANN were applied but accuracy of 87.13% with Sensitivity of 90.28 % and Specificity 79.31 % was achieved by Neural Network along with feature selection method with following features, Typical Chest Pain Region RWMA, Age, EF, HTN, DM, T-inversion , ESR Q wave, ST elevation ,PR ,BMI ,Lymph ,BP ,Dyspnea, HDL, CR, WBC, Weight ,Function Class ,Airway disease ,HB, TG, BBB, Na, Sex, LVH, Hb , FH .

Zeinab Arabasadi, Roohallah Alizadehsani and Mohamad Roshanzamir (2017) proposed a "Computer aided decision making method for heart disease detection using hybrid neural network". [16] The data set was collected from Z-Alizadeh Sani dataset containing 303 patients data set with 54 feature for each patient. Different feature selection algorithm (Gini Index, Weight by SVM, Information gain and Principal Component Analysis) were applied in pre-processing of data. Initial weight of Neural Network (NN) was identified by Genetic Algorithm and then NN was trained. They made use of error back propagation algorithm with sigmoid exponential function to train their Neural Network. Proposed method obtained accuracy of 88 % with Sensitivity of 91 and Specificity 89.4 % for the detection of CAD.

Oluwarotimi Williams Samuela, Grace Mojisola Asogbona, Arun Kumar Sangaiahc (2016) proposed a “An Integrated Decision Support System Based on ANN and Fuzzy_AHP for Heart Failure Risk Prediction”.[17] The data set used were from clevland contained 303 data samples of patients with some missing values and 13 attributes. In their study they used 13 attributes and their contribution were determined by an experienced cardiac clinician. A Fuzzy analytic hierarchy process (Fuzzy_AHP) technique was used to compute the global weights for the attributes based on their individual contribution. Then the global weights that represent the contributions of the attributes were applied to train an ANN classifier for the prediction of HF risks in patients. The result shows promising result with the accuracy of 81.10 %.

Table 1 : Summary of Literature review

S.No	Author	Dataset	Feature Selecting method	Algorithm	Accuracy
1	Alizadehsani R, Habibi J [13] (2016)	Z-Alizadeh Sani Data set for LAD, RCA and LCX	Gini Index(GI), Information Gain(IG)	C.45 Bagging	LAD - 79.54% (IG and Bagging) RCA -68.95 % (GI and Bagging) LCX – 65.09 % (No Feature extraction and Bagging)
2.	Zeinab Arabasadi, Roohallah Alizadehsani (2017) [16]	Switzerland Cleveland & Z-Alizadeh Sani dataset for CAD	Gini Index (GI), Weight by SVM, Information Gain (IG)	Neural Network	Switzerland – 71.5 % Cleveland – 89.4 % Z-Alizadeh Sani – 88 .85 %
3.	R.Alizadehsani, J. Habibi (2016) [12]	Z-Alizadeh Sani dataset for CAD	Information Gain (IG) and Gini Index (GI),	Bagging, Neural Network & Naïve Bays	Bagging – 89.44% NN – 87.13 % Naïve Bays – 47.84 %

4.	Min Liu, Xiaowei Xu and Ye Tao (2017) [10]	Data set CAD from China Hospita, China	ReliefF	Random Forest	CAD - 90.28 %
5	Burak Kolukisa, Hilal Hacilar, Gokhan Goy (2018) [11]	Cleveland, and Z-Alizadehsani CAD Data	Information Gain (IG), Gain Ratio (GR), Relief-F (RF)	SVM Random Forest(RF) Naïve Bayes (NB), Bagging	SVM (Z-Alizadehsani) – 86.07% RF (Clevland) – 83.16 % RF (Z-Alizadehsani) – 87.12 % NB (Clevland) – 83.49 % NB (Z-Alizadehsani) – 80.57 % Bagging (Clevland) – 83.16% Bagging(Z-Alizadehsani) – 87.7 %

6.	František Babič and Jaroslav Olejár (2017) [15]	Cleveland, Z-Alizadeh Sani Hungarian		Support Vector Machine(SVM) and Neural Network(NN)	NN (Clevland) - 86.32 % SVM (Z-Alizadeh Sani) - 86.67 % SVM (Hungarian) - 73.7 %
7.	Oluwarotimi Williams Samuela and Grace Mojisola Asogbona (2016) [17]	Cleveland data set for CAD	Fuzzy_AHF	Neural Network	81.10%
8.	Gokhan Goy, Mustafa Kus, and Burcu Bakir-Gungor (2018) [11]	Cleveland, South Africa and Z-Alizadeh sani.	Information Gain (IG) and Gain Ratio (GR), Relief-F (RF)	Naïve Bays	83.49 %

Performance of any classification model greatly depend on the features section technique used in it. Most of the work done by author uses feature selection technique like Gini Index, Information gain, Relief-F, Gain Ratio etc. Considering only machine learning technique to feature selection might lead to poor result. So feature selection by Random Forest with mean decreasing accuracy and Domain expert is used here in this work. The classification is done by Neural Network.

Also most of the work done so far in this field were concerned only for the classification of coronary artery disease, very few work has been done for detail classification of LAD, LCX and RCA. Only one work has been done for detail classification LAD, LCX and RCA with accuracy of 79.54% for LAD, 68.95 % RCA and 69.09% for LCX as done by Alizadehsani R, Habibi J.

Comparison of Baseline model based with working modal is given in table 1 and 2:

Baseline model (Roohallah Alizadehsani, Jafar Habibi, Mohammad Javad Hosseini, and Z-Alizadeh Sani 2016)

Table 2 : Summary of Baseline modal

Classification model	For Coronary Artery Disease (CAD)
Feature Selection method	Applied (Gini Index)
Doctors Recommendation on features selection	No
Data Source	Z-Alizadeh Sani dataset
No of Features	28
Type of Modal	ANN Back-propagation (28-10-5-1)
Validation with another data source	NO

Working Neural Network Model

Table 3 : Summary of working Modal

Classification model	For LAD, LCX and RCA for Coronary Artery Disease
Feature Selection method	Applied (Random Forest)
Doctors Recommendation on features selection	Yes
Data Source	Z-Alizadeh Sani dataset
No of Features	35 for LAD and 33 for LCX and 34 for RCA
Type of modal	ANN Resilient Back propagation
Verification with another data source	Yes (Data provided by Dr. Ravi Sahi, Cardiology TUTH at from TUTH)

Chapter III: METHODOLOGY

3.1. Overview of the working model

Data set collected form the UCS repository contains 54 features, domain expert were asked to mark important features, only 27 of the features were marked important by them, detail of the features can be found in section 3.4 another features selection method random forest (Mean Decreasing Accuracy) was adopted to get better classification results, detail of the method can be found in section 3.3 Then total features selected form both the method was adopted to Neural Network classification model, 10-fold cross validation was used as small sample were available.

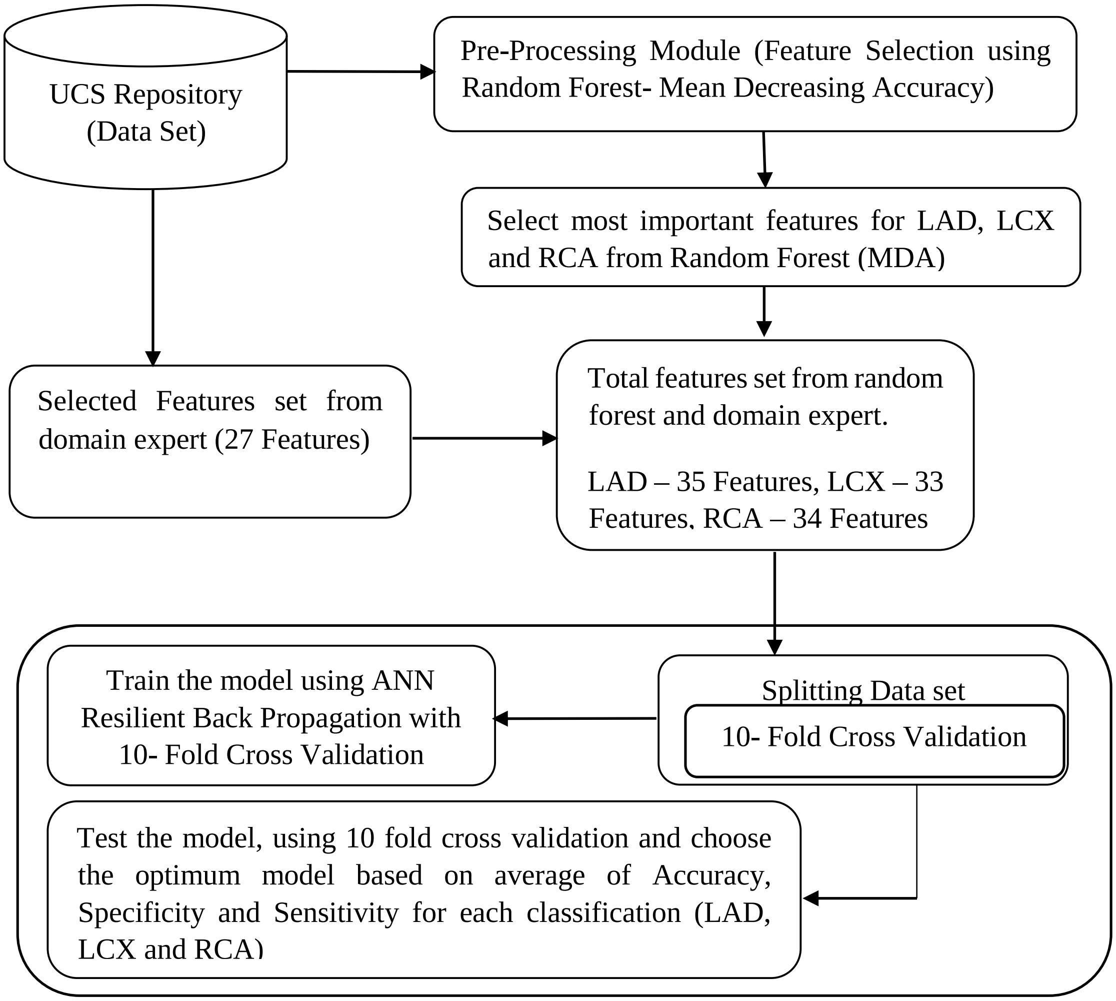

UCS Repository
(Data Set)
Pre-Processing Module (Feature Selection using Random Forest- Mean Decreasing Accuracy)
Select most important features for LAD, LCX and RCA from Random Forest (MDA)
Selected Features set from domain expert (27 Features)
Total features set from random forest and domain expert.
LAD – 35 Features, LCX – 33 Features, RCA – 34 Features
Train the model using ANN Resilient Back Propagation with 10- Fold Cross Validation
Splitting Data set
10- Fold Cross Validation
Test the model, using 10 fold cross validation and choose the optimum model based on average of Accuracy, Specificity and Sensitivity for each classification (LAD, LCX and RCA)

Figure 4 : Block diagram of working model

3.2. Preprocessing and Features Selection

Data preprocessing is an important step in the data mining process. Real-world data is often incomplete, inconsistent, and/or lacking in certain behaviors or trends, and is likely to contain many errors. Data preprocessing is a proven method of resolving such issues. Among many pre-processing techniques data classification, data normalization and data cleaning are used in for this project

Feature selection is the process of selecting a subset of relevant features (predictors) to be used in the model. This techniques is used because it simplification of models to make them easier to interpret, it makes model faster by reducing training times, avoid the curse of dimensionality. In order to get the benefits of the feature section technique good feature selection method must be adopted. In this research work, Doctor's Recommendation along with random forest (mean decreasing accuracy) was used to identify the most relevant features from dataset which were then used in the classification of coronary artery disease.

3.2.1 Description of Data Set

The data sample for this research was collected by Z-Alizadeh Sani at Tehran's Shaheed Rajaei Cardiovascular, Medical and Research Centre, Iran. This dataset contains 303 records with 54 features with two target classes: Suffered by Coronary Artery Disease or Normal. A patient is categorized as Stenotic, if his/her diameter narrowing is greater than or equal to 50%, and otherwise as Normal. The dataset can be used for detail classification of coronary artery disease, i.e. stenosis diagnosis of each LAD, LCX and RCA arteries in human heart. They divided the original set of variables into four groups: Demographic, Symptoms and Examination, ECG, Laboratory and Echo. The data set is available in UCI machine learning repository.

Table 4 : Total features of the data set and their distribution

S.No	Feature Type	Feature Description	Value
1	Demographic	Age	30-86
2		Weight	48-120
3		Length	140-188
4		Sex	M, F
5		BMI (Body Mass Index Kg/m)	18-41
6		DM (Diabetes Mellitus)	Yes, No
7		HNT (Hyper tension)	Yes , No
8		Current Smoker	Yes , No
9		Ex- Smoker	Yes , No
10		FH(Family History)	Yes , No
11		Obesity (MBI >25)	Yes , No
12		CRF (Chronic Renal Failure)	Yes , No
13		CVA (Cerebrovascular Accident)	Yes , No
14		Airway Disease	Yes , No
15		Thyroid Disease	Yes , No
16		CHF (Congestive Heart Failure)	Yes , No
17		DLP (Dyslipidemia)	Yes , No
18	Symptom and Examination	BP (Blood Pressure in mmHg)	90-190
19		PR (Pulse Rate ppm)	50-110
20		Edema	Yes, No
21		Weak Peripheral Pulse	Yes, No
22		Lung Rales	Yes, No
23		Systolic Murmur	Yes, No
24		Diastolic Murmur	Yes, No
25		Typical Chest Pain	Yes, No
26		Dyspnea	Yes, No

27		Function Class	1,2,3,4
28		Atypical	Yes, No
29		Nonanginal CP	Yes, No
30		Exertional CP (Exertional Chest Pain)	Yes, No
31		Low Th Ang (Low Thershold Angina)	Yes, No
32	ECG	Q Wave	Yes, No
33		ST Elevation	Yes, No
34		ST Depression	Yes, No
35		T Inversion	Yes, No
36		LVH (Left Ventricular Hyertrophy)	Yes, No
37		Poor R wave Progression	Yes, No
38		BBB	Yes, No
39	Laboratory and echo	FBS (Fasting Blood Sugar in mg/dl)	62-400
40		Cr (Creatine in mg/dl)	0.5 - 2.2
41		TG (Triglyceride in mg/dl)	37-1050
42		LDL (Low Density Lipoprotein in mg/dl)	18-232
43		HDL (High Density Lipoprotein in mg/dl)	15-111
44		BUN (Blood Urea Nitrogen in mg/dl)	6-52
45		ESR (Erythrocyte Sedimentation rate in mm/h)	1-90
46		HB (Hemoglobin in g/dl)	8.9 - 17.6
47		K (Potassium in mEq/lit)	3 - 6.6
48		Na (Sodium in mEq/lit)	128- 156
49		WBC (White Blood Cell in Cells / Ml)	37000-18,000
50		Lymph (Lymphocyte in %)	7-60
51		Neut (Neutrophil in %)	32-89
52		PLT (Platelet in 1000/ml)	25-742
53		EF-TTE (Ejection Fraction in %)	15-60
54		Region RWMA (Regional Wall Motion Abnormality)	0.1,2,3,4
55	**Classification Variable**	**LAD - Diagnosis with LAD Blockage**	**Stenotic/ Normal**
56	**Classification Variable**	**LCX - Diagnosis with LCX Blockage**	**Stenotic/ Normal**
57	**Classification Variable**	**RCA - Diagnosis with RCA Blockage**	**Stenotic/ Normal**

Table 4 shows the total data set used in this study. There are total 57 variable with 55th (LAD), 56th LCX and 57th RCA being the detection variable. These Data set were updated at 2017-11-17.

3.2.2 Data Preprocessing

Data preprocessing is an integral step in Machine Learning as the quality of data and the useful information that can be derived from it directly affects the ability of our model to learn. The real world data may contain many errors, null values or might contain irrelevant features and noise thus making it incomplete. Data preprocessing is a proven method for resolving such issues. Therefore, so it is extremely important that we preprocess our data before feeding it into our model.

The outcome of data pre-processing is the ultimate data set with reduced attributes. The major weakness with clinical data set is the presence of redundant records. The occurrence of redundant instances causes the learning algorithm to be biased towards frequent records and unbiased towards infrequent records. These redundant records were removed in order to improve the detection accuracy.

Steps for data pre-processing used in the study were:

a. **Data Classification**

Among all the features of the original data set, the input variables like Sex, Obesity, CRF, CVA , Airway disease ,Thyroid Disease ,CHF ,DLP ,BP ,Weak Peripheral Pulse ,Lung rales , Systolic Murmur, Diastolic Murmur, Dyspnea, Atypical, Nonanginal , Exertional CP, Low TH Ang, LVH, Poor R Progression ,BBB have string values. All these features have two Male/ Female for feature named sex and Yes/ No for rest.

Binary encoding method was applied for these feature which have string values as:
1 for Male and -1 for Female and
1 for Yes and -1 for No

Also the output class LAD, LCX and RCA contains string value Normal and Stenotic these values are also encoded as:
Normal: 0
Stenotic: 1

b. Data Cleaning

The collected data set contains some features whose standard deviation is zero, means it has the same values in all observation set. Such features doesn't carry any information for the decision class so such features were removed. Among the total observation set two such features "Chronic Renal Failure" and "External Chet Pain" were removed. Histogram plot for each features set was done to know the data distribution among the feature set. The histogram plot for each feature set is attached in ANNEX I.

c. Data Normalization

Normalization is a technique applied as part of data preparation for machine learning. It improves the performance and training stability of the model. The goal of normalization is to change the values of numeric columns in the dataset to a common scale, without distorting differences in the ranges of values.

For machine learning, every dataset does not require normalization. It is required only when features have different ranges. In our case the feature vector WBC has large value in input data set then rest features which will intrinsically influence the result more due to its larger value but this doesn't necessarily mean it is more important for a predicting class.

Among two popular normalization Mix-max normalization and Z-score normalization, Z-Score normalization is chosen as features were not uniformly distributed across a fixed range and we see that the feature distribution does not contain extreme outliers.

Z-Score Normalization = $(X_i - \mu) / \sigma$ ……. Equation 3.1

Where

X_i - Feature value in data set

μ - Mean value of the feature

σ – Standard deviation of the feature

3.3 Random Forest based Feature Selection

Data set may contains many features which are irrelevant to the decision variable, and we often want a way to create a model that only includes the most important features. Features selection is a means to reduce the burden of dimensionality. It has befits like the model is more simple to interpret or Better visualization and data understanding, variance in the model is reduced, also the over-fitting and finally reduction in the computational cost (and time) of training a model. The process of identifying only the most relevant features is called feature selection.

Random forest algorithm is a supervised algorithm which provides relatively good accuracy, robustness and ease of use. As the name suggest, this algorithm creates the forest with a number of trees. Each tree in the ensemble is built from a sample drawn with replacement (i.e. a bootstrap sample) from the training set. In addition, instead of using all the features, a random subset of features is selected, further randomizing the tree. As a result, the bias of the forest increases slightly, but due to the averaging of less correlated trees, its variance decreases, resulting in an overall better model.

General rule of thumb is

$$Mtry = \sqrt[2]{Number\ of\ features(54)} \quad \ldots \text{ Equation 3.2}$$

So M_{try} is choose as 7 for our model.

Random Forests is often used for classification model however it also provide two straightforward methods for feature selection based on: Mean Decrease Impurity and Mean Decrease Accuracy.

In in research work importance features were selected from Mean Decreasing Accuracy (MDA) from a random forest model generated.

MDA is also called permutation importance. This is because when a decision tree is created based on a set of learning datasets divided through subsampling, the intuition behind permutation has an importance that is not a useful feature for predicting an outcome. OOB (Out-Of-Bag) is one of the subsampling techniques to calculate prediction error of each of the training samples utilizing bootstrap aggregation. MDA is the method that calculates variable importance by permutation and the method uses OOB to divide its sample data. In other words, OOB estimates more accurate prediction value by computing OOB accuracy before and after the permutation of variable X_j and compute the difference.

Since , t ϵ {1,2,3,4… n_{tree}}, the variable importance of X_j in tree t is the averaged value of the difference between predicted class before permuting X_j, which is $y_i = f(X_i)$, and after permuting variable X_j , which is $y_i = f(x^j_i)$, in certain observation i.

The Formula of mean decreasing Accuracy is:

$$VI(\mathrm{Xj}) = \frac{1}{ntree}\sum_{t=1}^{ntree}\frac{\sum_{i\epsilon OBB} I\big(yi=f(xi)\big)-\sum_{i\epsilon OBB} I\big(yi=f(xij)\big)}{|OBB|} \quad \text{… Equation 3.3}$$

Where,
VI is the Variable Importance

To increase the reliability of the features being selected for this work, random forest based feature selection algorithm was run and only 50 percent variables with high score in MDA were selected. Again the model was run using the new variables set and return the error rate. Implementing these procedure iteratively until the error rate does not decrease. [22] Following are the key factors used in experiment for random forest:

Method: Random Forest Classification model
Factor considered for feature selection: Mean Decreasing Accuracy (MDA)
No. of variables tried at each split (M_{try}): 7
Number of trees grown in each experiment: 500
Experiment run: 10 times and Average of MDA was take
Lowest OBB Error achieved: 16.47 (LAD), 17.34 (LCX) and 17.94 (RCA)

Table 5 contains the list of important features selected from mean decreasing accuracy of random forest algorithm.

Table 5 : List of important features selected from Random Forest (MDA)

Features for LAD	**MDA**	**Features for LCX**	**MDA**	**Features for RCA**	**MDA**
Typical_Chest_Pain	15.270	Age	12.176	Typical_Chest_Pain	8.767
Age	13.602	Typical_Chest_Pain	9.967	DM	8.63722
Region_RWMA	10.730	CR	5.807	Age	6.4029
EF_TTE	10.037	PLT	4.136	Neut	5.94175
Nonanginal	8.770	TG	3.531	Poor_R_Progression	5.70714
Attypical	7.331	BP	3.340	Attypical	5.151
CR	6.295	HTN	3.253	ESR	4.98655
Dyspnea	4.299	Weight	3.220	Lymph	4.68516
DLP	3.385	Sex	3.190	FBS	3.50314
St_Elevation	3.239	K	3.002	Q_Wave	3.26672
Lymph	3.141	FBS	2.997	Nonanginal	2.83763
Current_Smoker	3.136	Attypical	2.987	PR	2.55514
Diastolic_Murmur	3.048	Length	2.864	Current_Smoker	2.43402
DM	2.953	Current_Smoker	2.289	HTN	2.2322
Q_Wave	2.855	DM	2.268	Function_Class	2.22704
Poor_R_Progression	2.746	DLP	2.220	Diastolic_Murmur	2.15592
Neut	2.575	BUN	2.214	Region_RWMA	2.13948
BP	2.435	EF_TTE	2.073	Na	2.00046
FBS	2.381	HDL	1.980	TG	1.78406
Function_Class	2.222	Q_Wave	1.944	WBC	1.40648
Tinversion	2.148	ESR	1.904	BMI	1.23156
ESR	1.833	EX_Smoker	1.860		
Weight	1.597	Tinversion	1.622		
Lung_rales	1.449	Region_RWMA	1.559		
Sex	1.383				
HTN	1.303				

Variable importance plot for these selected features is found in ANNEX II

3.4 Feature selection based on medical recommendation

Expert knowledge includes facts of related domain and requires the use of data and information. Machine learning seeks to represent generalizations, that is, not to represent each individual situation, but to group the situations that share important properties. There are many algorithm for selecting the important of features for any project. However while only considering the machine learning or domain experts may leads to poor classification. One of the major advantage of this project work is that the feature selection during the pre-processing from the total data set was done consulting the domain expert as well. Approach based on expert knowledge searches the most relevant features from perspective of application domain and hence greatly improved gap for creating the misclassification error.

An arrangement was made to meet Dr. Amrit Bogati, Cardiologist, Shahid Gangalal National Heart Centre and Dr. Ravi Sahi, MBBS, MD, DM Cardiology TUTH, Manmohan Cardiovascular and Transplant Centre IOM Maharajgunj who helped me to list down the important features in total data set. They said they follow American Heart Association recommendation for the prediction of CAD along with Framingham Heart Study. Among the total listed features in the data set table 6 shows features were marked most important from the doctors:

Table 6 : List of important features selected from domain expert

S. No.	**Selected features by doctor for coronary artery disease**
1	Age
2	Weight
3	Length
4	Sex
5	BMI
6	DM
7	HTN
8	Current Smoker
9	Ex-Smoker
10	FH
11	Obesity
12	DLP
13	Typical Chest Pain
14	Dyspnea
15	Function Class
16	Atypical
17	Nonanginal CP
18	Q-Wave
19	St Elevation
20	St Depression
21	Tinversion
22	FBS
23	CR
24	TG
25	LDL
26	HDL
27	EF-TTE

3.5 Input Data set

The input data set for each class LAD, LCX and RCA is made combining both methods mentioned above. List in the data set contains the features that are used as input is shown in table 7.

Table 7: Features selected from domain expert and random forest (MDA)

S.No	LAD and Dr. Recommendation	LCX and Dr. Recommendation	RCA and Dr. Recommendation
1	Age	Age	Age
2	Attypical	Attypical	Attypical
3	BMI	BMI	BMI
4	BP	BP	CR
5	CR	BUN	Current_Smoker
6	Current_Smoker	CR	Diastolic_Murmur
7	Diastolic_Murmur	Current_Smoker	DLP
8	DLP	DLP	DM
9	DM	DM	EF_TTE
10	Dyspnea	Dyspnea	ESR
11	EF_TTE	EF_TTE	EX_Smoker
12	ESR	ESR	FBS
13	EX_Smoker	EX_Smoker	FH
14	FBS	FBS	Function_Class
15	FH	FH	HDL
16	Function_Class	Function_Class	HTN
17	HDL	HDL	LDL
18	HTN	HTN	Length
19	LDL	K	Lymph
20	Length	LDL	Lymph
21	Lung_rales	Length	Neut
22	Lymph	Nonanginal	Nonanginal
23	Neut	Obesity	Obesity
24	Nonanginal	PLT	Poor_R_Progression
25	Obesity	Q_Wave	PR
26	Poor_R_Progression	Region_RWMA	Q_Wave
27	Q_Wave	Sex	Region_RWMA
28	Region_RWMA	St_Depression	St_Depression
29	Sex	St_Elevation	St_Elevation
30	St_Depression	TG	TG
31	St_Elevation	Tinversion	Tinversion
32	TG	Typical_Chest_Pain	Typical_Chest_Pain
33	Tinversion	Weight	WBC
34	Typical_Chest_Pain		Weight
35	Weight		

3.6 Resilient Back Propagation

Training a neural network is the process of finding values for the weights and biases so that, for a set of training data with known input and output values, the computed outputs of the network closely match the known outputs. The most common technique used to train neural networks is the back-propagation algorithm. Back propagation requires a value for a parameter called the learning rate. The effectiveness of back propagation is highly sensitive to the value of the learning rate. [18]

Resilient back propagation is an algorithm that can be used to train a neural network, is similar to the more common (regular) back-propagation. But it has two main advantages over back propagation:

- Resilient Back Propagation is often faster than back propagation.
- It doesn't require you to specify any free parameter values, as opposed to back propagation which needs values for the learning rate.

3.6.1 Working Principle Resilient Back Propagation

Many machine learning algorithms, including Resilient Back Propagation Algorithm, is based on a mathematical concept called the gradient. A gradient is made up of several partial derivatives for weight and bias i.e. A gradient is just a collection of the all-partial derivatives for all the weights and biases. A partial derivative for a weight can be thought of as the slope of the tangent line (the slope) to the error function for some value of the weight.

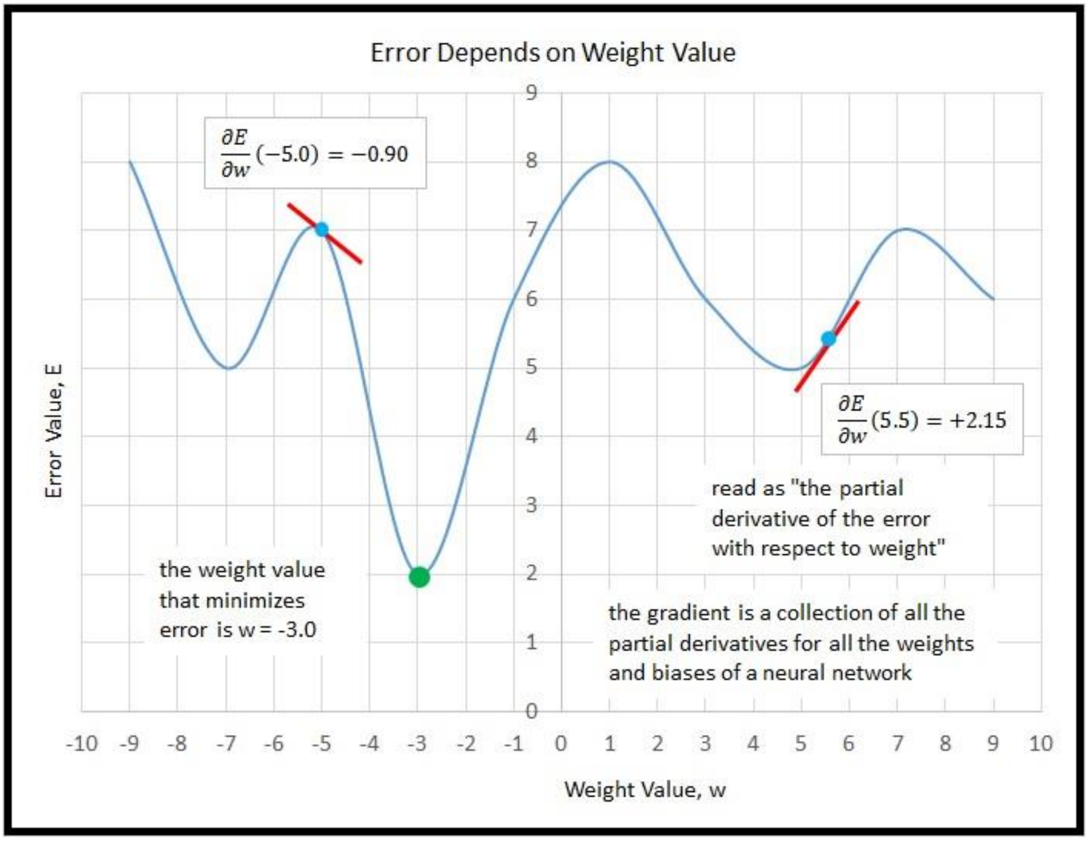


Figure 5 : Partial Derivatives and the Gradient

All natural network have some measure of error (Sum of Square Error, Root Mean Square, Mean Square etc), and that the value of error will change as the value of one weight and bias changes. During training, regular back propagation uses the magnitudes of the partial derivatives to determine how much weight is to be adjusted. This seems very reasonable, but some time if the learning rate and maximum gradient are not fixed in regular back propagation method then, new weight could swing wildly back and overshoot again in the other direction. This oscillation could continue and the weight for the minimum error would never be found.

With regular back propagation normally make use a small learning rate, which along with the magnitude of the gradient, to determines the weight delta in a training iteration. This means we likely won't overshoot an optimal answer, but it means training will be very slow as you creep closer and closer to a weight that gives minimum error.

Algorithm of Resilient Back Propagation (RPROP) Neural Network

Resilient propagation, in short, RPROP is one of the fastest training algorithms available. The algorithm just refers to the direction of the gradient. It is a supervised learning method. It works similarly to back propagation, except that the weight updates is done in a different manner.

The Resilient Back Propagation Algorithm makes two significant changes to the back-propagation algorithm.

- It doesn't use the magnitude of the gradient to determine a weight delta; instead, it uses only the sign of the gradient.

$$\Delta_{ij}^{(t)} = \begin{cases} \eta^{+} \times \Delta_{ij}^{(t-1)} & , \ if \quad \dfrac{\partial E}{\partial w_{ij}}^{(t-1)} \times \dfrac{\partial E^{(t)}}{\partial w_{ij}} > 0 \\ \eta^{-} \times \Delta_{ij}^{(t-1)} & , \ if \quad \dfrac{\partial E}{\partial w_{ij}}^{(t-1)} \times \dfrac{\partial E^{(t)}}{\partial w_{ij}} < 0 \\ \Delta_{ij}^{(t-1)} & , \ else \end{cases}$$

$$where \quad 0 < \eta^{-} < 1 < \eta^{+}$$

… Equation 3.5

The update-value Δ_{ij} evolves during the learning process based on the sign of the error gradient of the previous iteration $\frac{\partial\, E(t-1)}{\partial Wij}$ and the error gradient of the current iteration $\frac{\partial\, E(t)}{\partial Wij}$. Every time the partial derivative (error gradient) of the corresponding weight W_{ij} changes its sign, which indicates that the last update was too big and the algorithm has jumped over a local minimum, the update-value Δ_{ij} is decreased by the factor η- , which is a constant with a value of 0.5. If the derivative retains its sign, the update value is slightly increased by the factor η+ in order to accelerate convergence in shallow regions. η+, is a constant with a value of 1.2. If the derivative is 0 then we do not change the update-value.

- Instead of using a single learning rate for all weights and biases, it maintains separate weight deltas for each weight and bias, and adapts these deltas during training.

$$\Delta w_{ij}^{(t)} = \begin{cases} -\Delta_{ij}^{(t)} \quad , & if \quad \dfrac{\partial E^{(t)}}{\partial w_{ij}} > 0 \\ +\Delta_{ij}^{(t)} \quad , & if \quad \dfrac{\partial E^{(t)}}{\partial w_{ij}} < 0 \\ 0 \quad , & else \end{cases}$$

$$w_{ij}^{(t+1)} = w_{ij}^{(t)} + \Delta w_{ij}^{(t)}$$

… Equation 3.6

Once the update-value is calculated for each weight, the weight-update is then calculated. There are two rules to follow to calculate the weight-update. The first rule is that if the current derivative and the previous derivative retain their signs then above equation is used to calculate the weight-update.

If the current derivative is a positive value meaning the previous value was also a positive value (increasing error), then the weight is decreased by the update value. If the current derivative is negative value meaning the previous value was also a negative value (decreasing error) then the weight is increased by the update value.

The second rule is that if the current derivative and the previous derivative have changed their signs i.e. there was a big step taken then chances are that a minimum was missed. To avoid such big jumps, the weights need to be reverted to the previous state as.

$$\Delta w_{ij}^{(t)} = -\Delta w_{ij}^{(t-1)}, \quad if \quad \frac{\partial E}{\partial w_{ij}}^{(t-1)} * \frac{\partial E}{\partial w_{ij}}^{(t)} < 0$$

… Equation 3.7

If the weight was reverted then the previous derivative needs to also be changed, otherwise when the weight is updated again then it will reapply the same changes, repeating this scenario. Therefore, the previous derivative $\frac{\partial\, E(t-1)}{\partial Wij}$ is set to 0. [19]

Parameters used in resilient back propagation neural network
Δ_0 is the initial value of the delta
Δ_{max} - maximum value a delta update can have
Δ_{min} - minimum value for delta update
η^+ and η^- Reducing factor

Activation function

It is a function that is use to get the output of node in a neural network. It maps the resulting values in between 0 to 1 or -1 to 1 etc. (depending upon the function types). Neural networks are to support nonlinearity and more complexity, the activation function to be used has to be robust enough to maps inputs to outputs and should be able to get access to a much richer hypothesis space that would benefit from deep representations. It should be differential, shouldn't cause gradients to vanish and should be simple and fast in processing. It should be able to learn and represent almost anything. [20]

Resilient Back Propagation Neural Network use sigmoid as activation function.

The sigmoid function is a mathematical function that produces a sigmoidal curve; a characteristic curve for its S shape. This squashes the input to any value between 0 and 1, and makes the model logistic in nature.

This function refers to a special case of logistic function defined by the following formula:

$$F(x) = \frac{1}{1+\ e-x} \qquad \text{…equation 3.8}$$

Weight Initialization

Weight of the neural network was initialized randomly in the model and the model was trained.

Loss Function / Error Function

Loss function is a function that tells us, how good our neural network for a certain task. The intuitive way to do it is, take each training example, pass through the network to get the number, subtract it from the actual number we wanted to get and square it (because negative numbers are just as bad as positives).

The loss function used in our project is “sum of squared error” and it can be expressed as:

$$E_{see}(y, y^\wedge) = \sum_{k=1}^{i} (y - y^\wedge)^2 \qquad \text{…equation 3.9}$$

Reducing Factor and Min-Max value of delta

Since the RPROP uses the reducing factor (η) and min-max (Δ_{max} / Δ_{min}) value of delta for train the model. So neural network model was trained with various value of η , Δ_{max} and Δmin to get the optimum result. Δ_{max} used in this modal is 1.2 and Δ_{min} as 0.5 .

3.6.2 K-Fold Cross-Validation

Cross-validation is a resampling procedure used to evaluate machine learning models on a limited data sample. The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into. As such, the procedure is often called k-fold cross-validation. When a specific value for k is chosen, it may be used in place of k in the reference to the model, such as k=10 becoming 10-fold cross-validation. [21]

Cross-validation is primarily used in applied machine learning to estimate the skill of a machine learning model on unseen data. That is, to use a limited sample in order to estimate how the model is expected to perform in general when used to make predictions on data not used during the training of the model. 10-Fold Cross Validation method has been used to train the proposed model.

Algorithm of K- Fold Cross Validation

Step 1. Split the dataset into k groups

Step 2. For each unique group:

a. Take the group as a test data set & the remaining groups as a training data set

b. Fit a model on the training set and evaluate it on the test set

c. Retain the evaluation score and discard the model

Step 3. Repeat step 2 until every K-fold serve as the test set. Then take the average of the recorded scores to get the performance metric for the model.

3.6.3 Optimized Resilient Back Propagation Neural Network Model

Three Neural Network model were created using the Resilient Back Propagation algorithm as there were different set for input for each model of natural network as selected by pre-processed data. For the classification of LAD (Left Anterior Descending artery) following 35 features were selected from pre-processed data Age, Attypical, BMI, BP, CR, Current_Smoker, Diastolic_Murmur, DLP, DM, Dyspnea, EF_TTE, ESR, EX_Smoker, FBS, FH, Function_Class, HDL, HTN, LDL, Length, Lung_rales, Lymph, Neut, Nonanginal, Obesity, Poor_R_Progression, Q_Wave, Region_RWMA, Sex, St_Depression, St_Elevation, TG, Tinversion, Typical_Chest_Pain, Weight. One hidden layers waw used with 18 neurons. The output of neural network consist of two class either Normal (0) or Stenotic (1).

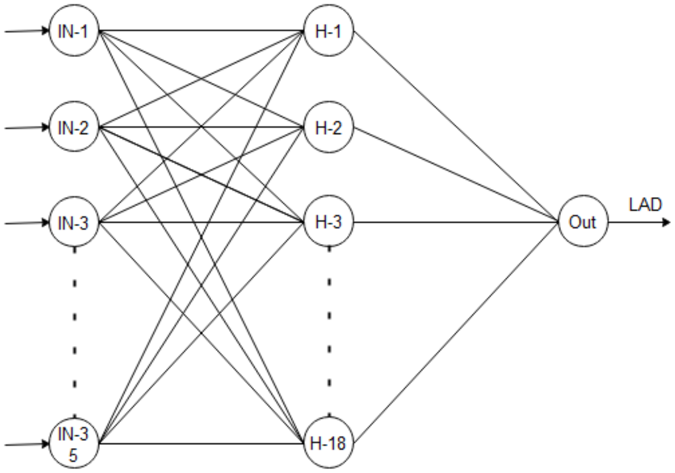


Figure 6 : Optimized Neural Network for LAD classification

For the classification of LCX (Left Circumflex artery) following 33 features were selected from pre-processed data Age, Attypical, BMI, BP, BUN, CR, Current_Smoker, DLP, DM, Dyspnea, EF_TTE, ESR, EX_Smoker, FBS, FH"Function_Class, HDL, HTN, K, LDL, Length, Nonanginal, Obesity, PLT, Q_Wave, Region_RWMA, Sex, St_Depression, St_Elevation, TG, Tinversion, Typical_Chest_Pain, Weight. Two hidden layers were used with 25-15 neurons in each layer. The output of neural network consist of two class either Normal (0) or Stenotic (1).

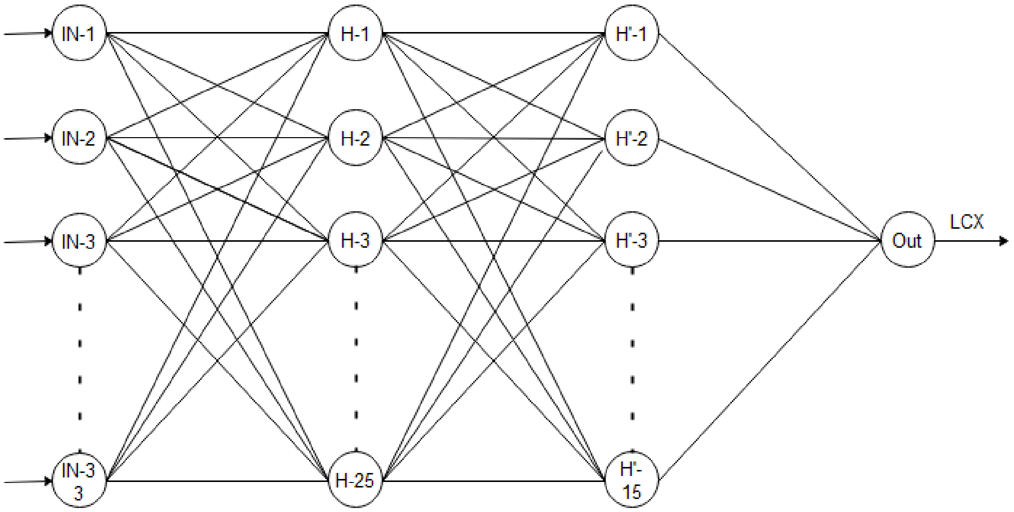


Figure 7 : Optimized Neural Network modal for LCX classification

For the classification of RCA (Right Coronary Artery) following 34 features were selected from pre-processed data Age, Attypical, BMI, CR, Current_Smoker, Diastolic_Murmur, DLP, DM, Dyspnea, EF_TTE, ESR, EX_Smoker, FBS, FH, Function_Class, HDL, HTN, LDL, Length, Lymph, Na, Neut, Nonanginal, Obesity, Poor_R_Progression, PR, Q_Wave, Region_RWMA, St_Depression, St_Elevation, TG, Tinversion, Typical_Chest_Pain, WBC, Weight. Two hidden layers were used with 15-10 neurons in each layer. The output of neural network consist of two class either Normal (0) or Stenotic (1).

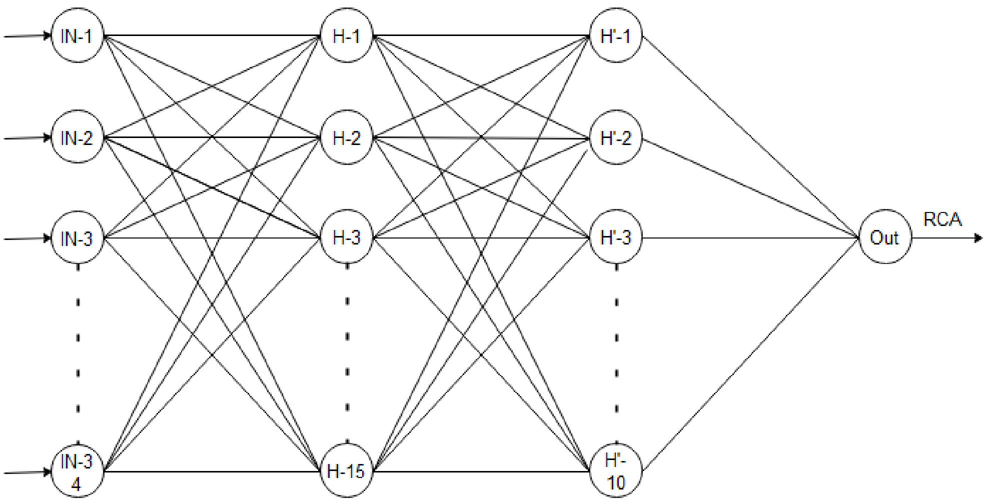


Figure 8 : Optimized Neural Network modal for RCA classification

To get the optimized neural network model from each class, model was run by variation in size of neural network, and learning rate for multiple time with 10-fold cross validation technique and different performance merits were considered like sensitivity, specificity and accuracy. Above all optimized neural network was achieved by using multiple trial of experiments in the data set selected from pre-processing.

3.6.4 Performance Metric

Evaluating a model involves checking if the predicted value is equal to the actual value during the testing phase. The metrics that we choose to evaluate the model is very important. Choice of metrics influences how the performance of machine learning algorithms is measured and compared. There are various metrics available to check the model. Following performance metrics were used for optimizing the proposed neural network model:

a. Confusion Matrix:

The Confusion matrix is one of the most intuitive and easiest metrics used for finding the correctness and accuracy of the model. When the values of the classification are plotted in a NxN matrix (2x2 in case of binary classification), the matrix is called the confusion matrix. All the evaluation metrics can be derived from the confusion matrix itself:

Table 8 : Confusion Matrix

	Predicted Value	**Predicted Value**
Actual Value	Positive	Negative
Positive	True Positive (TP)	False Positive (FP)
Negative	False Negative (FN)	True Negative (TN)

- **True Positives (TP):** True positives are the cases when the actual class of the data point was True and the predicted is also True

- **True Negatives (TN):** True negatives are the cases when the actual class of the data point was False and the predicted is also false.

- **False Positives (FP):** False positives are the cases when the actual class of the data point was False and the predicted is true. False is because the model has predicted incorrectly and positive because the class predicted was a positive one.

- **False Negatives (FN):** False negatives are the cases when the actual class of the data point was true and the predicted is False. False is because the model has predicted incorrectly and negative because the class predicted was a negative one.

b. Accuracy

Accuracy in classification problems is the number of correct predictions made by the model over all kinds predictions made. Accuracy is a good measure when the target variable class in the data are nearly balanced. Accuracy should never be used as a measure when the target variable classes in the data are a majority of one class.

$$Accuracy\ = \frac{\mathrm{TP+TN}}{TP+FP+\ FN+TN} \qquad \text{…equation 3.10}$$

c. Recall / Sensitivity

Recall is the number of True Positives divided by the number of True Positives and the number of False Negatives. Put another way it is the number of positive predictions divided by the number of positive class values in the test data. It is also called Sensitivity or the True Positive Rate. Sensitivity is also known as the ability to distinguish the sick from the true ill.

$$Recall = \frac{\mathrm{TP}}{TP+\ FN} \qquad \text{…equation 3.11}$$

d. Specificity

Specificity is the number of True Negative divided by the number of True Negative and the number of False Positives. It is opposite of recall. Specificity is the ability to distinguish the healthy from the true healthy.

$$Specificity = \frac{\mathrm{TN}}{TN+\ FN} \qquad \text{…equation 3.12}$$

e. Receiver Operating Characteristic curve

A Receiver Operating Characteristic (ROC) curve is a graphical visual that illustrates the predictive ability of a binary classifier system. The ROC curve is created by plotting a graph of the True Positive Rate (TPR) against the False Positive Rate (FPR). This gives us Sensitivity versus (1 - Specificity). The larger the area under ROC curve, the higher the performance of the algorithm is. A ROC curve typically looks like as shown in figure 9.

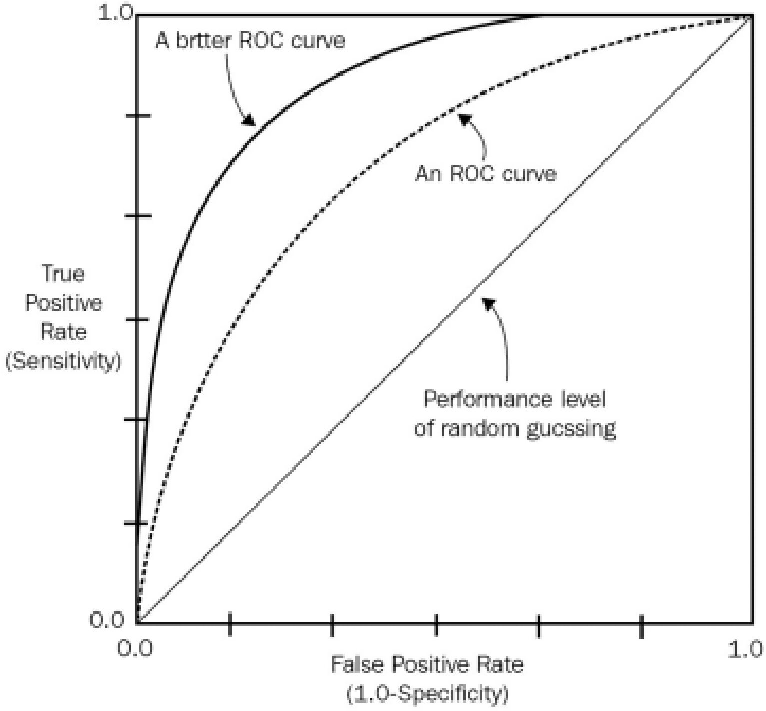


Figure 9 : Typical ROC curve

$FPR = \frac{\mathrm{FP}}{TN + FP}$ …equation 3.13 $TPR = \frac{\mathrm{TP}}{TP + FN}$ …equation 3.14

Receiver Operating Characteristics (ROC) analysis is an established method of measuring diagnostic performance for the analysis of medical test performance. The ROC curve is a good measure when the performance of different classifiers needs to be compared. ROC analysis is a standard approach used to determine the sensitivity and specificity of the diagnosis.

3.6.5 Baseline model comparison with proposed Neural Network Model

To compare baseline model proposed by Roohallah Alizadehsani, Jafar Habibi, Mohammad Javad Hosseini, and Z-Alizadeh Sani (2016) with current proposed neural network model, same architecture was recreated with 28 input features as shown in figure 10:

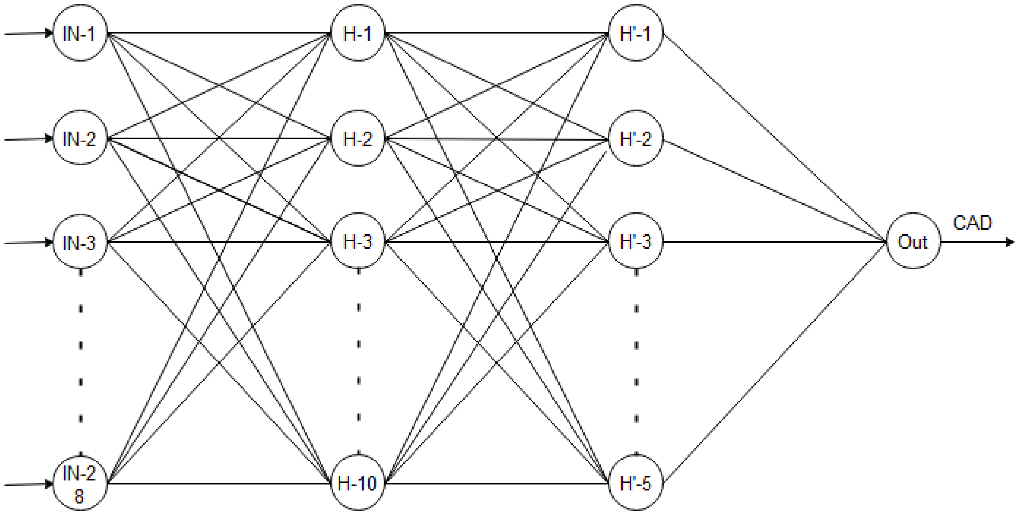


Figure 10 : Base line NN model with 28 input features and two hidden layer

The base line model consist of 28 input variables selected from feature selecting method (gain index) Typical Chest Pain, Region RWMA, Age, EF, HTN, DM, T-inversion , ESR, Q wave, ST, elevation ,PR ,BMI ,Lymph ,BP ,Dyspnea, HDL, CR, WBC, Weight ,Function Class ,Airway disease ,HB, TG, BBB, Na, Sex, LVH , FH. The model was trained with back propagation neural network with learning rate 0.01. Activation function used for recreating the base line model is sigmoid with error function sse (some of square error) to produce classification result for CAD.
Then same base line model with same set of 28 features were used, same learning rate was used for the classification of LAD, LCX and RCA independently. The Accuracy received was around 85% for LAD, 83% of accuracy was obtained for RCA and 85% for LCX classification.

3.7 Software and Tools used

Table 9: Tools used in the study

Item	**Tool**
Operating System	**Windows 10 64-bit**
Processor	**Intel® Core(TM) i7- 3820QM CPU @ 2.70 GHz**
Random Access Memory	**12 GB**
Graphics Processing Unit	**GeForce GTX 1050 ti 4 GB**
Programming Language	**R Studio**

CHAPTER IV: RESULTS

4.1. Results of optimized Resilient Back Propagation Neural Network model

The optimized Resilient Back Propagation Neural Network model was trained and was tested for LAD, LCX and RCA Classification using 10- fold cross validation method. Single model was made for the classification of LAD, LCX, and RCA. Maximum accuracy of 91.397% with 90.32 % specificity and 95 % sensitivity for LAD classification, Maximum accuracy of 88.09 % with 80 % specificity and 91.52 % sensitivity for LCX classification and Maximum accuracy of 90.36 % with 72.72 % specificity and 95.71 % sensitivity for RCA classification was achieved by the proposed model. To find the optimum result the, multiple experiments were carried out with different combination. Tabulated result for finding the optimized model is attached in Annex III.

4.2. Validation of Resilient Back Propagation Neural Network model

To validate the model different performance metrics were considers like Specificity, Sensitivity, ROC curve and Accuracy.

4.2.1 Calculation of Accuracy Metrics

The optimized proposed model for LAD was obtained with 35 input features selected from pre-processed data, 1 hidden layers 18 neuron in hidden layer. To get the optimized model 10- fold cross validation with repeated experimental trials was performed. Performance merits for the optimized model is show below.

Table 10: Confusion matrix for optimized model of LAD Classification

	Predicted Value	**Predicted Value**
Actual Value	Normal (0)	Stenotic (1)
Normal (0)	28	5
Stenotic (1)	3	57

Table 11: Specificity, Sensitivity and Accuracy for LAD Classification

No of Input features	No of hidden layers	No of neurons in each hidden layer	Performance Merits		
			Specificity	Sensitivity	Accuracy
35	1	18	95 %	96.32 %	91.397 %

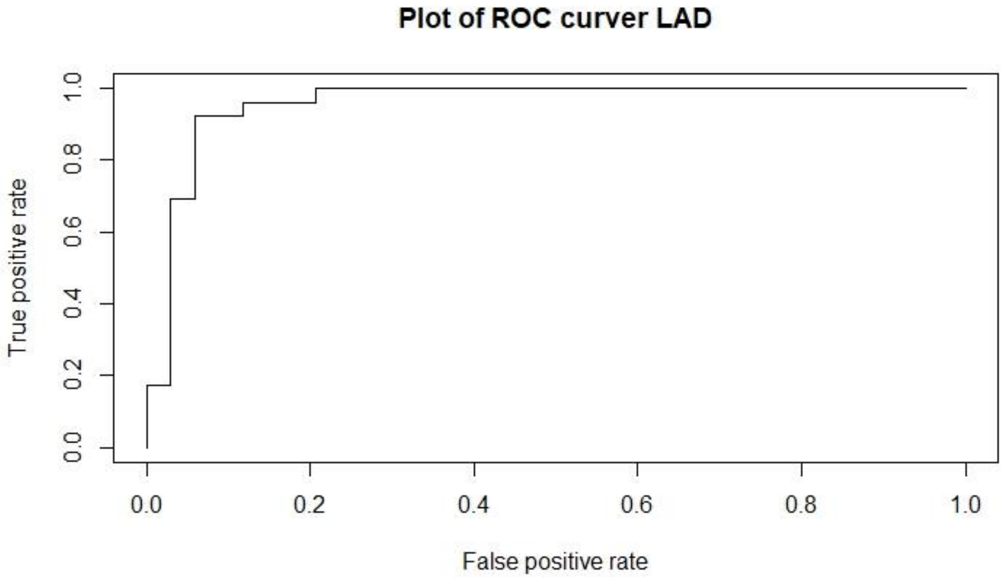


Figure 11 : ROC curve for LAD Classification model

The proposed model for LCX was obtained with 33 input features selected from pre-processed data, 2 hidden layers 25 and 15 neuron in each hidden layer. To get the optimized model 10- fold cross validation with repeated experimental trials was performed. Performance merits for the optimized model is show in table 13.

Table 12: Confusion matrix for optimized model of LCX Classification

	Predicted Value	**Predicted Value**
Actual Value	Normal (0)	Stenotic (1)
Normal (0)	54	5
Stenotic (1)	5	20

Table 13: Specificity, Sensitivity and Accuracy for LCX Classification

No of Input features	No of hidden layers	No of neurons in each hidden layer	Performance Merits		
			Specificity	Sensitivity	Accuracy
33	2	25-15	80 %	91.52 %	88.09 %

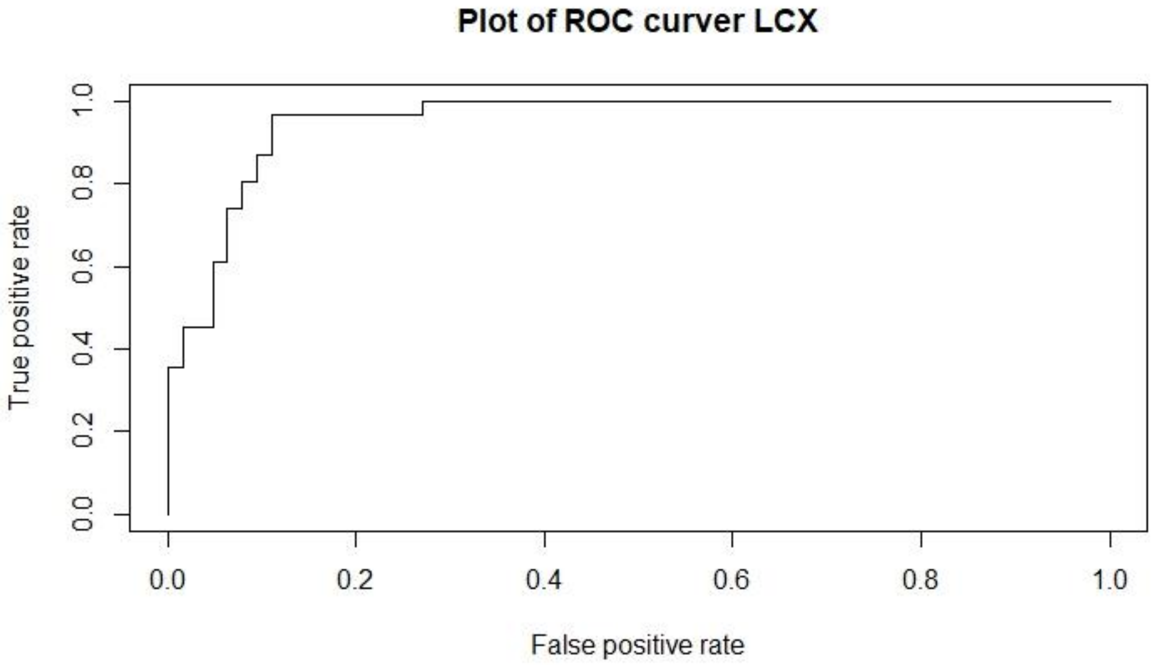


Figure 12: ROC curve for LCX Classification model

The proposed model for RCA was obtained with 34 input features selected from pre-processed data, 2 hidden layers 15 and 10 neuron in each hidden layer. To get the optimized model 10- fold cross validation with repeated experimental trials was performed. Performance merits for the optimized model is show in table 15.

Table 14 : Confusion matrix for optimized model of RCA Classification

	Predicted Value	**Predicted Value**
Actual Value	Normal (0)	Stenotic (1)
Normal (0)	67	5
Stenotic (1)	3	8

Table 15: Specificity, Sensitivity and Accuracy for RCA Classification

No of Input features	No of hidden layers	No of neurons in each hidden layer	Performance Merits		
			Specificity	Sensitivity	Accuracy
34	2	15-10	71.72 %	95.71 %	90.36 %

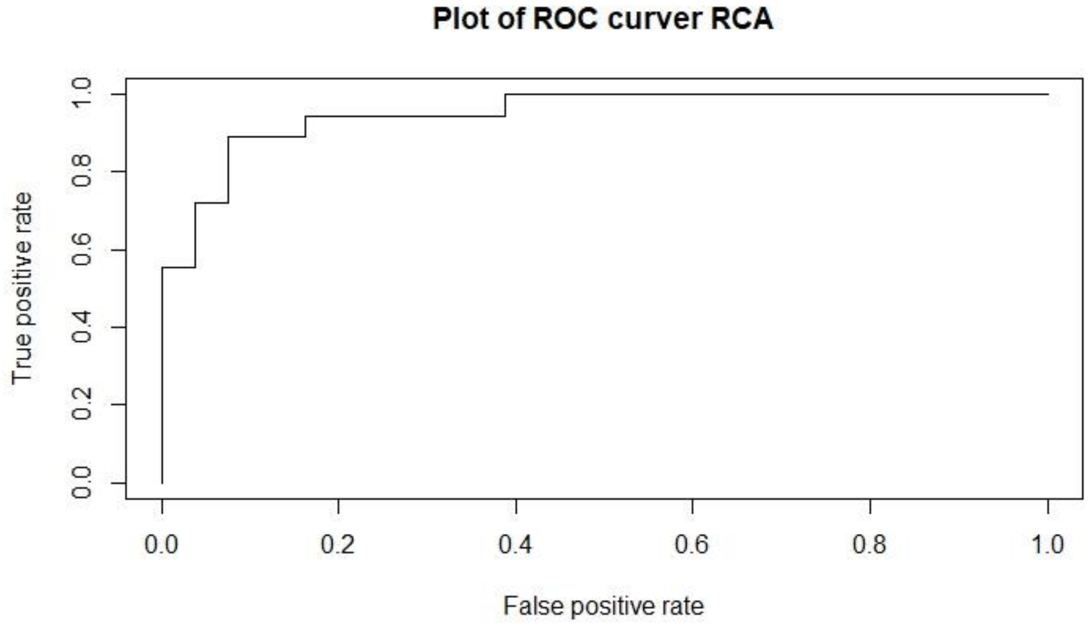


Figure 13 : ROC curve for RCA Classification model

4.2.2 Evaluating the model with TUTH data set

To evaluate the model with the data sample collected from TUTH, as provided by Dr. Ravhi Sahi was used. The data contains 60 features but only 22 features were found common with Z-Alidesani data set. New model was created and trained with Z-Alizadeh Sani data set with common 22 features as present in TUTH. The features used were Age, Weight, Length, BMI, LDL, HDL, FBS, TG, Sex, Smoking, DM, HTN, FH, Obesity, DLP, Typical_Chest_Pain, Dyspnea, Q_Wave, St_Elevation, St_Depression, Tinversion. Model was trained multiple times with different combination of hidden layer and neuron to get the best result. Performance merits for the optimized model for TUTH data is show in table 17.

Table 16 : Confusion matrix for optimized model for TUTH data set

	Predicted Value	**Predicted Value**	
Actual Value	Normal (0)	Stenotic (1)	
Normal (0)	11	9	LAD
Stenotic (1)	12	28	
Normal (0)	29	14	LCX
Stenotic (1)	6	11	
Normal (0)	24	21	RCA
Stenotic (1)	3	12	

Table 17 : Specificity, Sensitivity and Accuracy for TUTH Data Set

No of Input features	No of hidden layers	No of neurons in each hidden layer	**Performance Merits**			Remarks
			Specificity	Sensitivity	Accuracy	
22	2	12-8	70	47.82	65	LAD
22	2	15-10	64.7	82.85	66.67	LCX
22	2	12-8	69	78.89	60	RCA

Result in table 16 and table 17 show performance metric for real data (data collected from TUTH) greatly differs from the result of proposed model above, same is the case for the test set being selected from   Z-Alizadeh Sani data set. This might be because some very important features might be missing among the set of 22 features, as only 22 common features were found between TUTH data set and Z-Alizadeh Sani data. Result for multiple experimental data is attached in Annex IV.

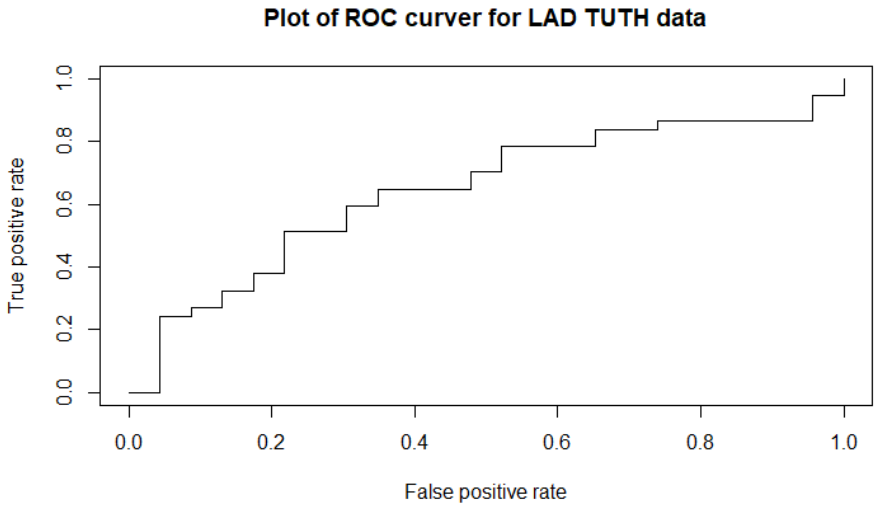


Figure 14 : ROC curve for LAD TUTH Data

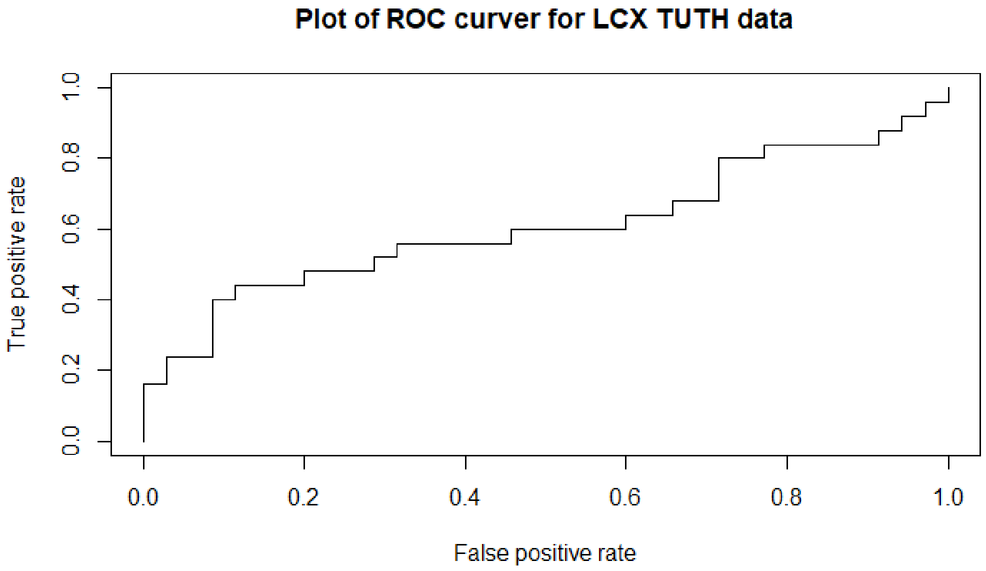


Figure 15 : ROC curve for LCX TUTH Data

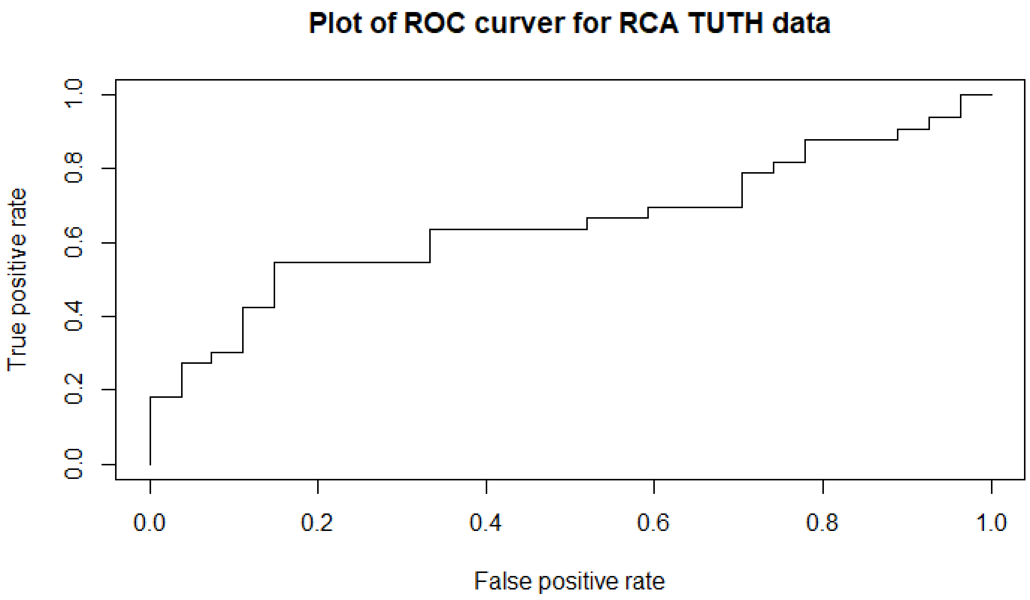


Figure 16 : ROC curve for RCA TUTH Data

4.3 Comparing the proposed model with others work

Among the works done by many of authors, following work are found most similar with the proposed model:

František Babič and Jaroslav Olejár uses Z-Alizadeh Sani data set with features set of Age, Diabetes Mellitus, Fasting Blood Suger(mg/dl), Pulse Rate, ST- Elevation, ST-Depression, Low-Density Lpoprotein(mg/dl), White Blood Cell, Obesity, Creatine(mg/dl), Ex- Smoker, Hemoglobin(g/dL), Non-angina CP. 10-Fold cross validation technique was used to get maximum accuracy of 86.32% for CAD classification using Neural Network model .

Roohallah Alizadehsani, Jafar Habibi uses Z-Alizadeh Sani data set and different feature selection method like Information Gain, Gini Index to select relevant features from total set of 54 features. Typical Chest Pain Region RWMA, Age, EF, HTN, DM, T-inversion , ESR Q wave, ST elevation ,PR ,BMI ,Lymph ,BP ,Dyspnea, HDL, CR, WBC, Weight ,Function Class ,Airway disease ,HB, TG, BBB, Na, Sex, LVH, Hb , FH were used for Neural Network modal to get accuracy of 87.13% with Sensitivity of 90.28 % and Specificity 79.31 % for CAD classification.

Only one model has been found for detail classification of LAD, LCX and RCA classification in human heart. Zahra Alizadeh Sani, Roohallah Alizadehsani , Jafar Habibi proposed modal for classification of LAD, LCX and RCA from the data collected form 303 random visitors to Rajaie Cardiovascular Medical and Research center. Gini Index and Information Gain was used for features selecting form the total data set. Maximum accuracy of 79.54 % for LAD with feature selection based on information gain was found and for RCA 68.95 % of accuracy with feature selection based on Gini Index, whereas for LCX maximum accuracy of 65.09 with no feature selection was found. This low accuracy in each classification might be because of selection of unimportant features.

CHAPTER V: CONCLUSION AND RECOMMENDATION

5.1. Conclusion

The study work for this project used 10-fold cross validation with resilient back propagation neural network for the detail classification of coronary artery disease. Original data set contains 54 features so features selection method was adopted to enhance the performance of classifier. Feature selection based on domain's expert and an algorithm Random Forest (Mean decreasing accuracy) combined was used to enhance the performance of the classifier. After feature selection 35 different features were selected for LAD and 33 for LCX and 34 features were selected for RCA classification. While constructing the model initially features set with doctor recommendation and random forest algorithm were individually tested but better results were not achieved. After combining the features selected from both method, improvement was seen in the performance of the classifier with classification accuracy of 91.397 % for LAD classification, 88.09 % for LCX and 90.36% for RCA was achieved.

To evaluate the model with data collected from TUTH as provided by Dr. Ravi new model with 22 common features set Age, Weight, Length, BMI, LDL, HDL, FBS, TG, Sex, Smoking, DM, HTN, FH, Obesity, DLP, Typical_Chest_Pain, Dyspnea, Q_Wave, St_Elevation, St_Depression, Tinversion was created. The model was trained and tested with Z-Alizadeh Sani data. Multiple experiments were carried out to get the best result and finally the selected optimum model was used to validate the data collected from TUTH. The Performance merits were not as high as in the test data. This might be because the data used is from another source, un-uniform distribution of data set, important features like might be missing among the selected feature set. Maximum accuracy achieved during evaluation is 65% for LAD 66.67 % For LCX and 60 % RCA. Set of features that were not found in TUTH data set were Attypical, BP, BUN, CR, Diastolic_Murmur, EF_TTE, ESR, EX_Smoker, Function_Class, K, Lymph, Neut, Nonanginal, PLT, Poor_R_Progression, PR, Region_RWMA, WBC. Also as per doctor Ravi Sahi low accuracy might be because of missing of important features like Ex-Smoking, Attypical, Age, Function Class and finally the as test data was collected form young age group (28-47 years).

5.2. Recommendations

Feature selection is more important work for any classification model. As it helps in selecting most relevant features which will help in generating better classification model. Domain expert along with Random Forest, mean decreasing accuracy was used for feature selection in this research work. Among many available algorithms Neural Network was used here for detail classification of coronary artery disease.

Despite of good results obtained by the proposed model, my future work is built robust classification model by taking large number of data sample from all possible categories of patients and geographical region along with use of different feather selection technique like Information gain, Recursive Feature Elimination, Ant Colony, Artificial Bee Colony (ABC), Along with guidance of domain expert for listing the important features.

REFERENCES

1. WHO "Cardiovascular diseases Key facts", 17th May 2017 [Online]. Available: https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)

2. Abhinav Vaidya, Ramjee Prasad Pathak, and Mrigendra Raj Pandey, Eds., "Prevalence of hypertension in Nepalese community triples in 25 years: a repeat cross-sectional study in rural Kathmandu," [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3861166/

3. Coronary Heart Disease in Nepal, WHO, 2017. [Online]. Available: https://www.worldlifeexpectancy.com/nepal-coronary-heart-disease

4. Anatomy and Circulation of the Heart, WebMD, 4th Feb 2019. [Online]. Available: https://www.webmd.com/heart-disease/high-cholesterol-healthy-heart#1

5. Human Heart Anatomy, Live Science, 22nd March, 2016. [Online]. Available: https://www.livescience.com/34655-human-heart.html

6. Coronary Arteries, Texas Heart Institute, 14th Aug 2018. [Online]. Available: https://www.texasheart.org/heart-health/heart-information-center/topics/the-coronary-arteries/

7. Coronary Artery Disease, American Heart Association, 31st June 2015. [Online]. Available: https://www.heart.org/en/health-topics/consumer-healthcare/what-is-cardiovascular-disease/coronary-artery-disease

8. Coronary artery disease, Mayo Clinic, 23rd Nov 2018. [Online]. Available: https://www.mayoclinic.org/diseases-conditions/coronary-artery-disease/symptoms-causes/syc-20350613

9. Coronary Arteries, Cleveland Clinic, 5th Jan 2019. [Online]. Available: https://my.clevelandclinic.org/health/articles/17063-coronary-arteries

10. Min Liu, Xiaowei Xu, Ye Tao and Xiadong Wang," An Improved Random Forest Method Base on RELIEFF for Medical Diagnosis", in Conf. 2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC)

11. Burak Kolukisa, Hilal Hacilar, Gokhan Goy and Burcu Bakir-Gungor, "Evaluation of Classification Algorithms, Linear Discriminant Analysis and a New Hybrid Feature Selection Methodology for the Diagnosis of Coronary Artery Disease", 2018 IEEE International Conference on Big Data (Big Data)

12. Roohallah Alizadehsani, Jafar Habibi, Mohammad Javad Hosseini, and Z-Alizadeh Sani, "A Data Mining Approach for Diagnosis of Coronary Artery Disease", 2016 Computer Methods and Programs in Biomedicine

13. Zahra Alizadeh Sani, Roohallah Alizadehsani and Jafar Habibi, "Diagnosing Coronary Artery Disease via Data Mining Algorithms by Considering Laboratory and Echocardiography Features", 2013 Research Cardiovascular Medicine

14. Haleh Ayatollahi and Leila Gholamhosseini, "Predicting coronary artery disease: a comparison between two data mining algorithms", 2019 Ayatollahi et al. BMC Public Health

15. František Babič and Jaroslav Olejár, "Predictive and Descriptive Analysis for Heart Disease Diagnosis", 2017 IEEE Computer Science and Information System

16. Zeinab Arabasadi, Roohallah Alizadehsani and Mohamad Roshanzamir, "Computer aided decision making method for heart disease detection using hybrid neural network", 2017 Elsevier Computer Methods and Programs in Biomedicine

17. Oluwarotimi Williams Samuela, Grace Mojisola Asogbona, Arun Kumar Sangaiahc “An Integrated Decision Support System Based on ANN and Fuzzy_AHP for Heart Failure Risk Prediction”, 2016 Expert Systems With Applications

18. How to Use Resilient Back Propagation to Train Neural Networks, Visual Studio Magazine,3rdSept2015.[Online].Available:https://visualstudiomagazine.com/articles/2015/03/01/resilient-back-propagation.aspx

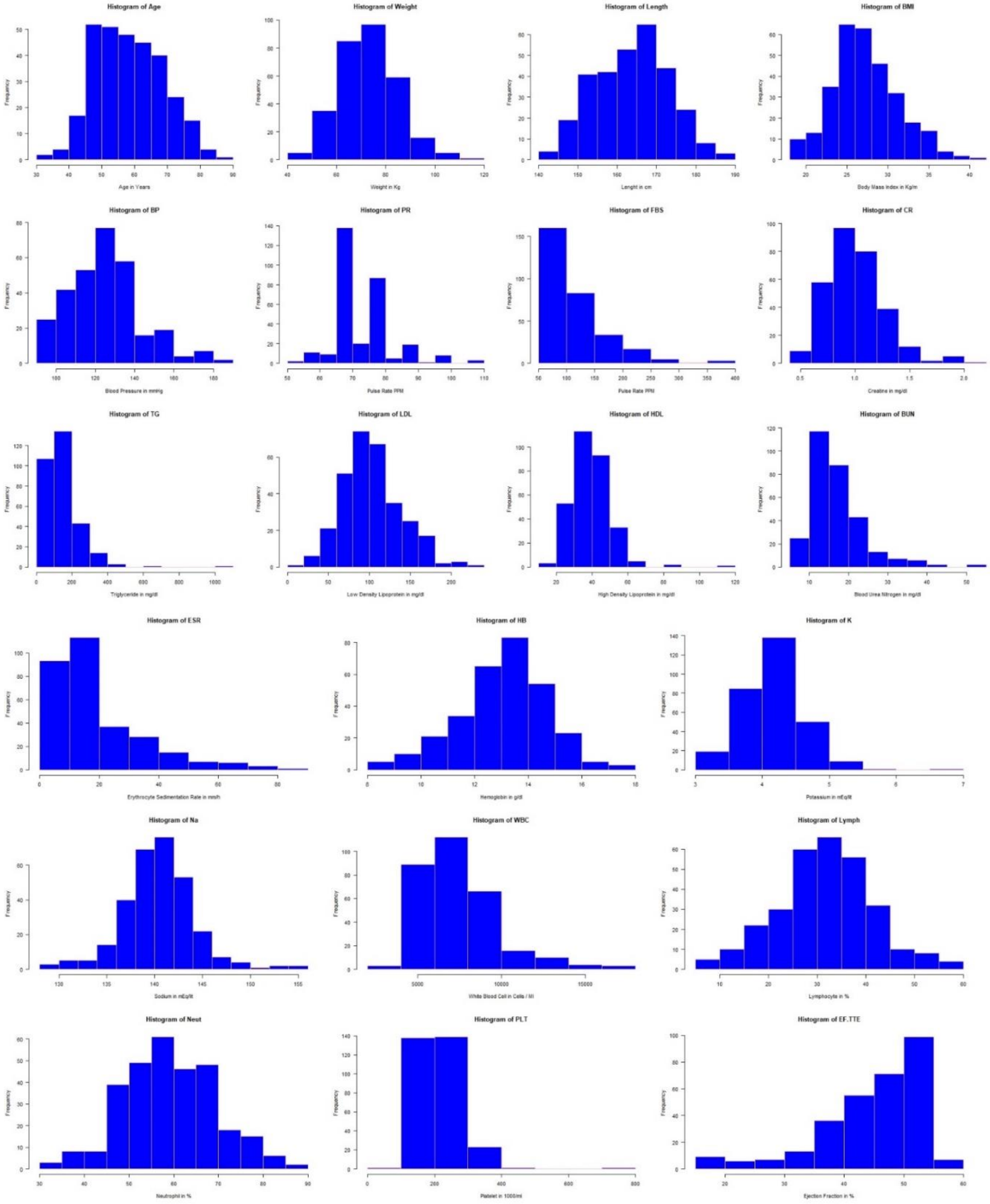
19. Navneel Prasad, Rajeshni Singh and Sunil Pranit Lal,”Comparison of Back Propagation and Resilient Propagation Algorithm for Spam Classification”, [Online]. Available: https://core.ac.uk/download/pdf/77223336.pdf

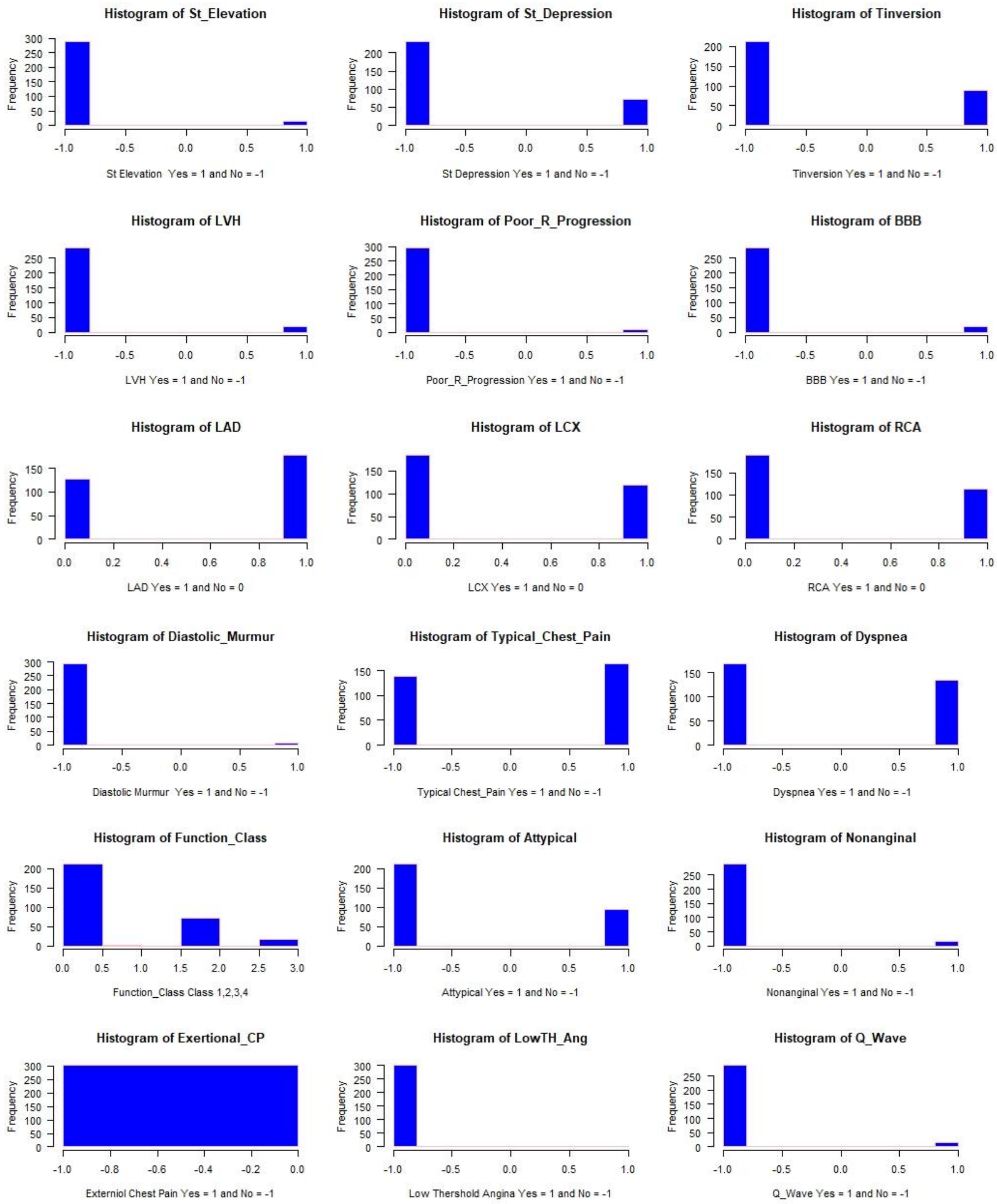
20. Balaji Venkateswaran, Giuseppe Ciaburro,”Activaiton Funciton”, Neural Network with R, Packt Publishing Ltd, Mumbai, 2017, pp. 42-52

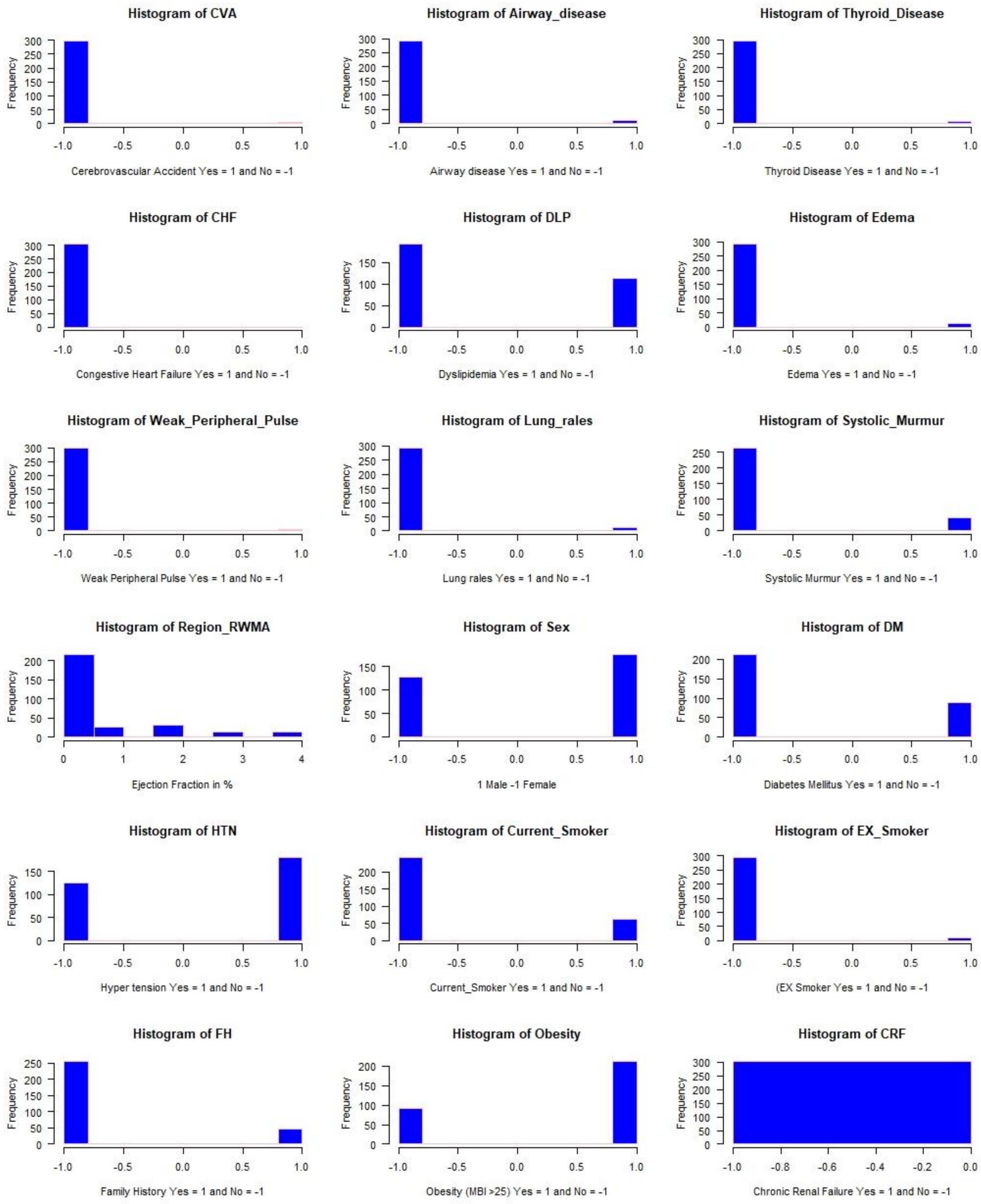
21. A Gentle Introduction to k-fold Cross-Validation, Machine Learning Mastery, 8th Aug 2019 [Online]. Available: https://machinelearningmastery.com/k-fold-cross-validation/

22. Hong Han , Xiaoling Guo and Hua Yu, “Variable Selection Using Mean Decrease Accuracy and Mean Decrease Gini Based on Random Forest” 2016 IEEE International Conference on Computer Methods

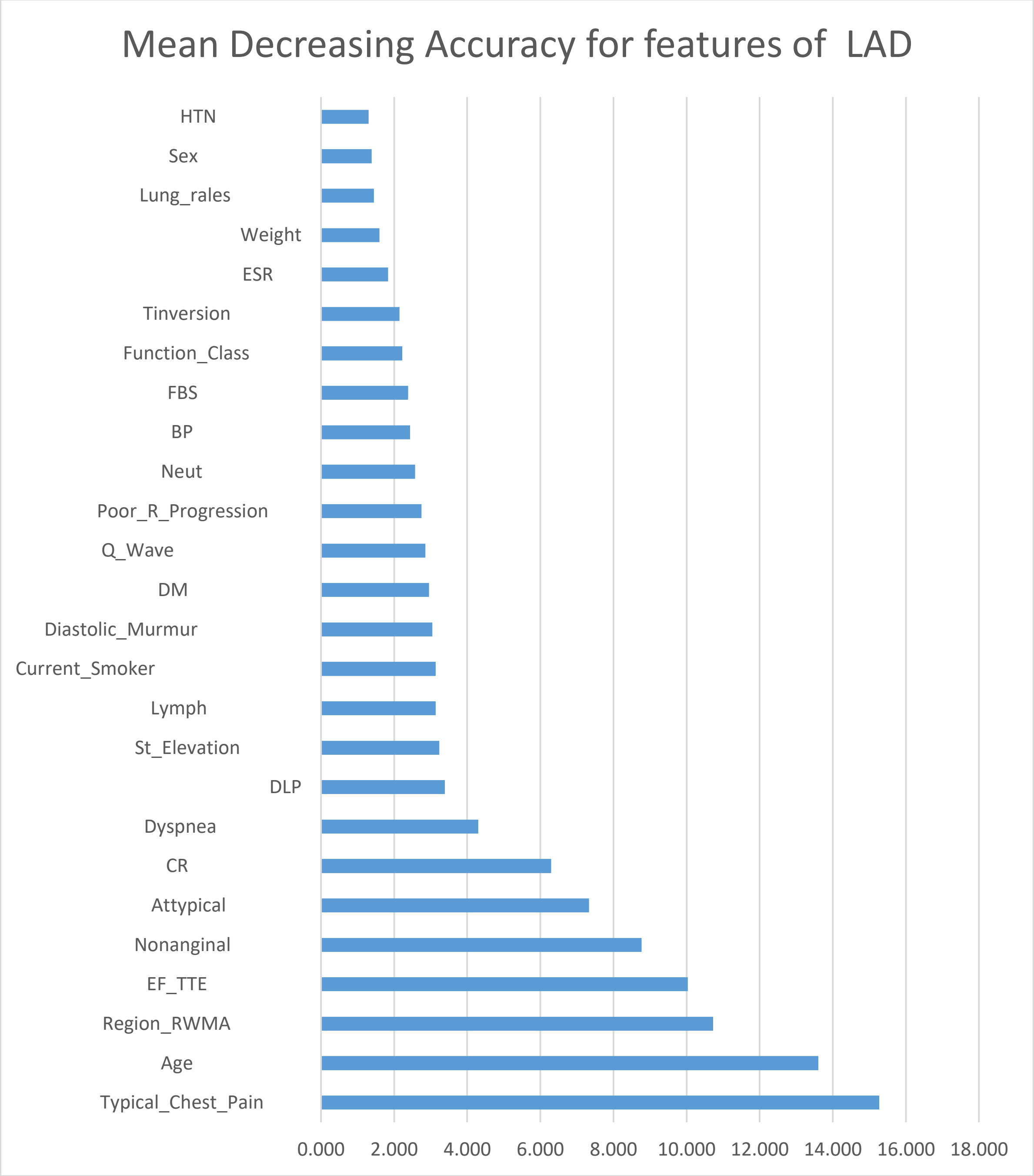
ANNEX –I: Histogram Plot for all features

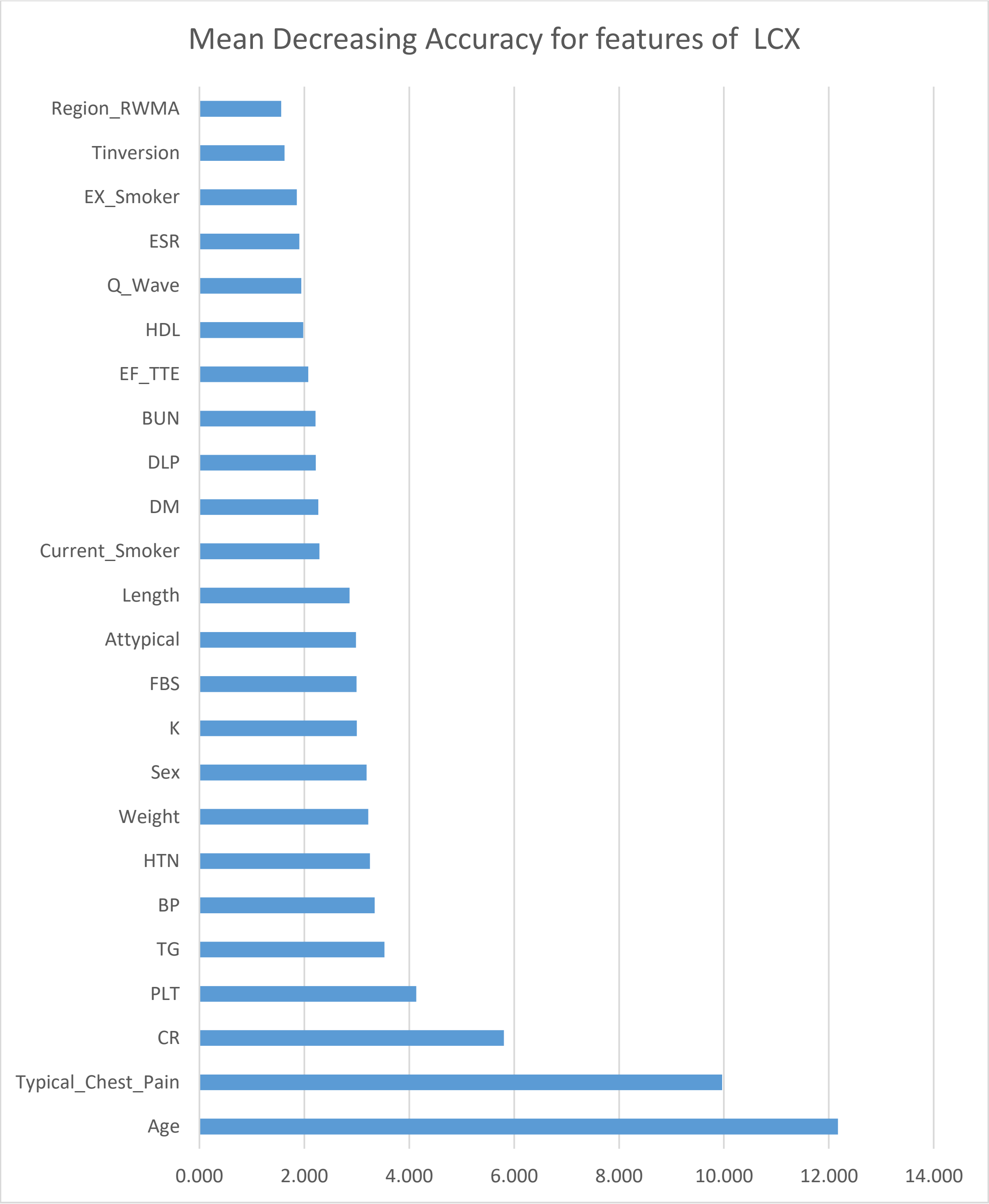


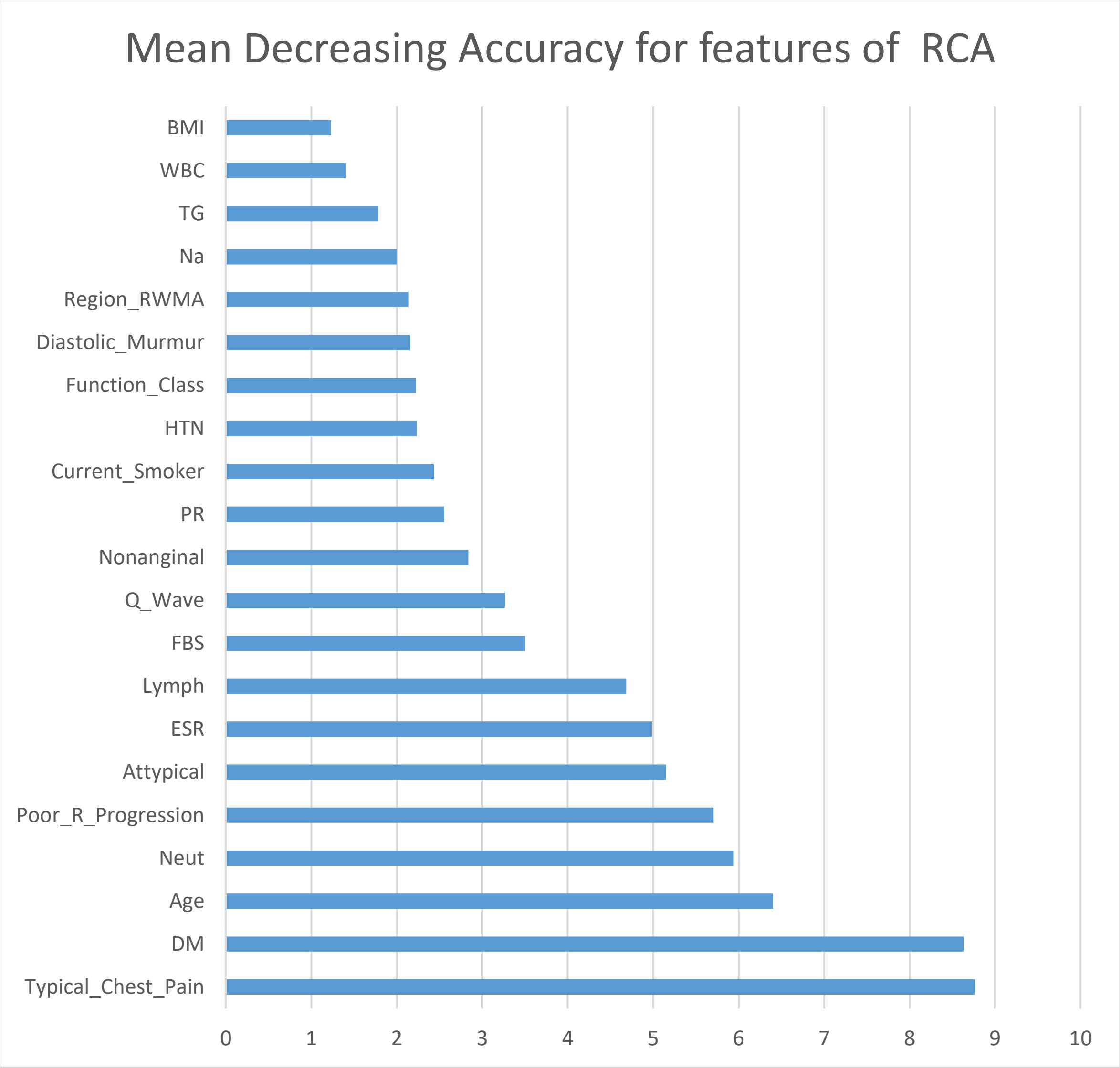

Histogram of St_Elevation
Frequency
300
250
200
150
100
50
0
-1.0
-0.5
0.0
0.5
1.0
St Elevation Yes = 1 and No = -1
Histogram of St_Depression
Frequency
200
150
100
50
0
-1.0
-0.5
0.0
0.5
1.0
St Depression Yes = 1 and No = -1
Histogram of Tinversion
Frequency
200
150
100
50
0
-1.0
-0.5
0.0
0.5
1.0
Tinversion Yes = 1 and No = -1
Histogram of LVH
Frequency
250
200
150
100
50
0
-1.0
-0.5
0.0
0.5
1.0
LVH Yes = 1 and No = -1
Histogram of Poor_R_Progression
Frequency
300
250
200
150
100
50
0
-1.0
-0.5
0.0
0.5
1.0
Poor_R_Progression Yes = 1 and No = -1
Histogram of BBB
Frequency
250
200
150
100
50
0
-1.0
-0.5
0.0
0.5
1.0
BBB Yes = 1 and No = -1
Histogram of LAD
Frequency
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
LAD Yes = 1 and No = 0
Histogram of LCX
Frequency
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
LCX Yes = 1 and No = 0
Histogram of RCA
Frequency
150
100
50
0
0.0
0.2
0.4
0.6
0.8
1.0
RCA Yes = 1 and No = 0
Histogram of Diastolic_Murmur
Frequency
300
250
200
150
100
50
0
-1.0
-0.5
0.0
0.5
1.0
Diastolic Murmur Yes = 1 and No = -1
Histogram of Typical_Chest_Pain
Frequency
150
100
50
0
-1.0
-0.5
0.0
0.5
1.0
Typical Chest_Pain Yes = 1 and No = -1
Histogram of Dyspnea
Frequency
150
100
50
0
-1.0
-0.5
0.0
0.5
1.0
Dyspnea Yes = 1 and No = -1
Histogram of Function_Class
Frequency
200
150
100
50
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
Function_Class Class 1,2,3,4
Histogram of Attypical
Frequency
200
150
100
50
0
-1.0
-0.5
0.0
0.5
1.0
Attypical Yes = 1 and No = -1
Histogram of Nonanginal
Frequency
250
200
150
100
50
0
-1.0
-0.5
0.0
0.5
1.0
Nonanginal Yes = 1 and No = -1
Histogram of Exertional_CP
Frequency
300
250
200
150
100
50
0
-1.0
-0.8
-0.6
-0.4
-0.2
0.0
Externiol Chest Pain Yes = 1 and No = -1
Histogram of LowTH_Ang
Frequency
300
250
200
150
100
50
0
-1.0
-0.5
0.0
0.5
1.0
Low Thershold Angina Yes = 1 and No = -1
Histogram of Q_Wave
Frequency
250
200
150
100
50
0
-1.0
-0.5
0.0
0.5
1.0
Q_Wave Yes = 1 and No = -1


Histogram of CVA
Frequency
300
250
200
150
100
50
0
-1.0
-0.5
0.0
0.5
1.0
Cerebrovascular Accident Yes = 1 and No = -1
Histogram of Airway_disease
Airway disease Yes = 1 and No = -1
Histogram of Thyroid_Disease
Thyroid Disease Yes = 1 and No = -1
Histogram of CHF
Congestive Heart Failure Yes = 1 and No = -1
Histogram of DLP
Dyslipidemia Yes = 1 and No = -1
Histogram of Edema
Edema Yes = 1 and No = -1
Histogram of Weak_Peripheral_Pulse
Weak Peripheral Pulse Yes = 1 and No = -1
Histogram of Lung_rales
Lung rales Yes = 1 and No = -1
Histogram of Systolic_Murmur
Systolic Murmur Yes = 1 and No = -1
Histogram of Region_RWMA
0
1
2
3
4
Ejection Fraction in %
Histogram of Sex
1 Male -1 Female
Histogram of DM
Diabetes Mellitus Yes = 1 and No = -1
Histogram of HTN
Hyper tension Yes = 1 and No = -1
Histogram of Current_Smoker
Current_Smoker Yes = 1 and No = -1
Histogram of EX_Smoker
(EX Smoker Yes = 1 and No = -1
Histogram of FH
Family History Yes = 1 and No = -1
Histogram of Obesity
Obesity (MBI >25) Yes = 1 and No = -1
Histogram of CRF
-0.8
-0.6
-0.4
-0.2
Chronic Renal Failure Yes = 1 and No = -1

ANNEX –II: Mean Decreasing Accuracy for features for LAD, LCX and RCA




Mean Decreasing Accuracy for features of LCX
Region_RWMA
Tinversion
EX_Smoker
ESR
Q_Wave
HDL
EF_TTE
BUN
DLP
DM
Current_Smoker
Length
Attypical
FBS
K
Sex
Weight
HTN
BP
TG
PLT
CR
Typical_Chest_Pain
Age
0.000
2.000
4.000
6.000
8.000
10.000
12.000
14.000


Mean Decreasing Accuracy for features of RCA
BMI
WBC
TG
Na
Region_RWMA
Diastolic_Murmur
Function_Class
HTN
Current_Smoker
PR
Nonanginal
Q_Wave
FBS
Lymph
ESR
Attypical
Poor_R_Progression
Neut
Age
DM
Typical_Chest_Pain
0
1
2
3
4
5
6
7
8
9
10

ANNEX- III: Experimental Trial for LAD, LCX and RCA Classification with 10-fold cross validation

Experimental Trial for LAD Classification with 10-fold cross validation

S. No.	No of Hidden Layer	No of Neurons in each hidden layer	Accuracy	Sensitivity	Specificity
1	2	25- 10	0.884409	0.8382563	0.914566873
2	2	25- 12	0.887957	0.8335037	0.905883496
3	2	25- 15	0.894301	0.8235037	0.902580277
4	2	25- 18	0.907849	0.8548039	0.918525104
5	2	25- 20	0.910409	0.8700023	0.921659433
6	2	25- 8	0.901183	0.8825023	0.930497752
7	2	25- 5	0.912086	0.8936134	0.930130105
8	2	30- 10	0.887849	0.8227801	0.910953848
9	2	30- 15	0.891183	0.8254388	0.904181709
10	2	30- 20	0.904301	0.8178198	0.906052043
11	2	30- 5	0.904516	0.874972	0.92439691
12	2	30- 12	0.894409	0.8263609	0.903395565
13	2	15- 10	0.897849	0.8406466	0.911834542
14	2	15- 15	0.894516	0.8544561	0.92152417
15	2	15- 12	0.901183	0.8927007	0.927458514
16	2	15- 8	0.904301	0.869218	0.927368774
17	2	15- 5	0.891183	0.8409944	0.90814992
18	2	15- 20	0.897849	0.8503595	0.917645786
19	2	15- 18	0.877742	0.7935037	0.890960526
20	2	15- 25	0.884731	0.8763609	0.922437888
21	2	15- 30	0.881398	0.8129388	0.897946675
22	2	15- 17	0.887742	0.8281466	0.908366943
23	2	15- 19	0.91109	0.8638609	0.921327243
24	2	12-10	0.89129	0.8744468	0.925343752
25	2	12-8	0.89086	0.8596849	0.915182096
26	2	12-5	0.867849	0.8103688	0.896044872
27	2	10-10	0.877957	0.8301914	0.901793946
28	2	10-15	0.9166	0.881356	0.92783656
29	2	10-5	0.877634	0.8257563	0.901168868
30	1	8	0.90086	0.8832828	0.92072235
31	1	5	0.890538	0.845846	0.89834423
32	1	3	0.85129	0.8578662	0.890749127

33	1	16	0.897097	0.8460606	0.902969545
34	1	10	0.887419	0.8855739	0.921606099
35	1	15	0.904086	0.8504411	0.910367432
36	1	20	0.894409	0.8801633	0.924928415
37	1	25	0.920753	0.8961355	0.929547828
38	1	30	0.890968	0.8593239	0.912264278
39	1	12	0.884301	0.8906334	0.926402526
40	**1**	**18**	**0.91397**	**0.963213**	**0.95**
41	3	25-15-3	0.857777	0.9557231	0.889462366
42	3	30-5-8	0.908405	0.8921505	0.873048866
43	4	30-15-10-5	0.802151	0.9250538	0.867949546

Experimental Trial for LCX Classification with 10-fold cross validation

S. No.	No of Hidden Layer	No of Neurons in each hidden layer	Accuracy	Sensitivity	Specificity
1	2	30- 10	0.858172	0.896542	0.815665
2	2	30- 15	0.854946	0.875585	0.806402
3	2	30- 20	0.874946	0.897113	0.835501
4	2	30- 5	0.868065	0.885292	0.832718
5	2	30- 12	0.844731	0.884057	0.807859
6	2	15- 10	0.864731	0.879488	0.813622
7	2	15- 15	0.851183	0.875431	0.799654
8	2	15- 12	0.870441	0.914042	0.857346
9	2	15- 8	0.864409	0.882251	0.816732
10	2	15- 5	0.854624	0.867098	0.797057
11	2	15- 20	0.848387	0.866542	0.801028
12	2	15- 18	0.871505	0.900292	0.828501
13	2	15- 25	0.847957	0.862515	0.791936
14	2	15- 30	0.868172	0.881849	0.816587
15	2	15- 17	0.867634	0.897376	0.840573
16	2	15- 19	0.864516	0.89432	0.828473
17	2	12-10	0.874624	0.89307	0.829018
18	2	12-8	0.87129	0.894737	0.831486

19	2	12-5	0.861075	0.912515	0.857374
20	2	10-10	0.838065	0.861418	0.774343
21	2	10-15	0.874409	0.898487	0.833144
22	2	10-5	0.878065	0.881126	0.824647
23	2	25- 10	0.874731	0.885515	0.816613
24	2	25- 12	0.861505	0.868125	0.793876
25	**2**	**25- 15**	**0.8809**	**0.91522**	**0.80**
26	2	25- 18	0.854731	0.893157	0.817494
27	2	25- 20	0.874731	0.891282	0.817887
28	2	25- 8	0.868065	0.909165	0.834645
29	2	25- 5	0.871613	0.907235	0.83372
30	1	10	0.871075	0.891659	0.826421
31	1	15	0.874624	0.89716	0.826757
32	1	20	0.865054	0.904368	0.840834
33	1	12	0.877849	0.899497	0.843882
34	1	18	0.877742	0.890697	0.831297
35	1	8	0.875054	0.885943	0.820772
36	1	5	0.855054	0.891458	0.811453
37	1	3	0.84172	0.902971	0.833155
38	1	22	0.848495	0.883815	0.797172
39	1	28	0.877849	0.91936	0.874463
40	1	24	0.848495	0.870491	0.788721
41	1	16	0.884516	0.91336	0.876421
42	1	25	0.874839	0.889102	0.825366
43	3	30-10-8	0.861937	0.882301	0.874106
44	4	25-10-10-8	0.86712	0.897124	0.881748

Experimental Trial for RCA Classification with 10-fold cross validation

S. No.	No of Hidden Layer	No of Neurons in each hidden layer	Accuracy	Sensitivity	Specificity
1	2	25- 10	0.880968	0.925234	0.745238
2	2	25- 12	0.897634	0.945403	0.756349
3	2	25- 15	0.89086	0.931518	0.702857

4	2	25- 18	0.897849	0.938125	0.744484
5	2	25- 20	0.877527	0.938435	0.709127
6	2	25- 8	0.887849	0.933069	0.728095
7	2	25- 5	0.894301	0.925212	0.720952
8	**2**	**15- 10**	**0.903616**	**0.907146**	**0.727272**
9	2	15- 15	0.900894	0.933442	0.868946
10	2	15- 12	0.897742	0.946294	0.780992
11	2	15- 8	0.884624	0.941828	0.726071
12	2	15- 5	0.897634	0.94939	0.789762
13	2	12-10	0.887957	0.927842	0.738045
14	2	12-8	0.900968	0.944641	0.774444
15	2	12-5	0.881398	0.93072	0.715754
16	2	10-10	0.874624	0.932563	0.725714
17	2	10-15	0.874516	0.93383	0.734167
18	2	10-5	0.854839	0.91072	0.654087
19	2	30- 10	0.88412	0.96103	0.84015
20	2	30- 15	0.901301	0.965396	0.856429
21	2	30- 20	0.878065	0.935341	0.749167
22	2	30- 5	0.864409	0.915102	0.688651
23	2	30- 12	0.867957	0.920505	0.6725
24	1	10	0.874731	0.917011	0.734167
25	1	15	0.877957	0.926844	0.747937
26	1	20	0.877957	0.916709	0.705476
27	1	25	0.874409	0.936515	0.747698
28	1	30	0.884409	0.933873	0.739921
29	1	12	0.887957	0.955143	0.817778
30	1	18	0.897634	0.939864	0.805556
31	1	8	0.894516	0.957638	0.839762
32	1	5	0.874731	0.920566	0.70254
33	1	3	0.88129	0.938163	0.750278
34	1	22	0.861505	0.892156	0.667817
35	1	28	0.888065	0.942868	0.782698
36	1	24	0.894301	0.941347	0.777857
37	1	16	0.880968	0.926737	0.730238
38	3	30-15-10	0.875673	0.781769	0.858803
39	3	30-15-12	0.881856	0.657778	0.822224

ANNEX- IV: Experimental Trial for evaluation of model with TUTH and Z-Alizadeh Sani Data Set with 22 features set

S. No.	No of Hidden Layer	No of Neurons in each hidden layer	TUTH Data Set			Z-Alizadeh Sani Data Set			For
			Accuracy	Sensitivity	Specificity	Accuracy	Sensitivity	Specificity	
1	2	20-15	64.33	56.52	70.68	79.01	85.12	91.11	LAD
2	2	20-10	63.33	52.17	70.27	84.26	74.28	84.84	
3	2	20-8	59.86	54167	68.67	84.67	81.32	86.72	
4	2	12-10	61.66	60.86	71.87	78.78	75.58	84.37	
5	**2**	**12-8**	**65**	**47.82**	**70**	82.5	81.48	89.79	
6	2	12-5	62.82	50.33	62.76	84.03	81.81	86.84	
7	2	15-10	61.61	56.52	70.58	79.56	78.78	87.27	
8	2	15-8	61.8	59.42	70.43	81.61	79.33	85.53	
9	2	15-12	58.44	60.6842	63.31	83.67	81.9	87.32	
10	1	10	64.28	56.52	77.22	81.3	81.08	89.06	
11	1	8	64.66	56.52	72.79	78.57	77.41	85.71	
12	1	12	63.33	56.52	71.42	81.91	77.56	81.25	
13	1	15	51.66	43.47	61.76	80.55	70	83.09	
14	1	18	56.66	43.47	64.86	89.04	92.87	95.12	
15	2	20-15	63.367	76.52	60.42	82.33	87.45	83.08	LCX
16	2	20-10	61.28	73.67	63.22	77.78	87.63	77.56	
17	2	20-8	65.28	70.37	56.53	78.12	83.45	73.65	
18	2	12-10	66.67	82.85	64.7	79.47	83.58	71.05	
19	2	12-8	60	71.42	53.38	81.67	87.06	79.42	
20	2	12-5	64.67	69.22	53.9	79.54	87.71	76.86	
21	**2**	**15-10**	**66.67**	**82.85**	**64.7**	**73.46**	**79.66**	**67.56**	
22	2	15-8	63.33	74.28	57.14	68	68.75	64.54	
23	2	15-12	58.44	60.6842	63.31	83.67	86.33	78.92	
24	1	10	58.33	74.28	50	80.61	90.16	80	
25	1	8	60	74.28	52.63	78.04	85.71	75.86	
26	1	12	60	74.28	52.63	78.12	80.7	71.5	
27	1	15	66.67	88.57	69.23	78.94	88.88	73.07	
28	1	18	58.33	62.85	50	73.86	81.13	68.75	
29	2	20-15	53.33	66.67	60.87	77.42	79.87	73.11	
30	2	18-10	55	62.96	61.53	69.32	78.78	72.89	

31	2	20-8	58.33	74.07	68.18	79.47	88.38	72.17	
32	2	12-10	56.67	74.07	66.67	72.04	84.28	65.05	
33	**2**	**12-8**	**60**	**78.88**	**69**	78.16	89.55	73.33	
34	2	12-5	58.33	88.89	78.57	79.36	87.33	71.51	
35	2	15-10	58.33	77.78	70	81.72	93.42	64.42	
36	2	15-8	53.33	70.37	61.9	78.75	91.22	70.47	
37	2	15-12	58.32	81.89	80	78.67	85.32	65.78	
38	1	10	58.33	74.28	50	77	83.11	66	
39	1	8	51.66	77.78	62.5	79.79	87.5	64	
40	1	12	53.33	59.25	63.33	77.14	85.18	67.67	
41	1	15	56.67	70.37	65.52	72.82	87.3	60	
42	1	18	55	77.77	65.52	75	81.08	72.52	