Performance Evaluation of Machine Learning Algorithms for Coronary Artery Disease Features

Md. Shah Jalal
Student Id: 012 162 017

A Thesis

in

The Department

of

Computer Science and Engineering

Presented in Partial Fulfillment of the Requirements

For the Degree of Master of Science in Computer Science and Engineering

United International University

Dhaka, Bangladesh

April 2019

Approval Certificate

This thesis titled "**Performance Evaluation of Machine Learning Algorithms for Coronary Artery Disease Features**” submitted by **Md. Shah Jalal**, Student ID: **012 162 017**, has been accepted as Satisfactory in fulfillment of the requirement for the degree of Master of Science in Computer Science and Engineering, United International University (UIU), Dhaka, Bangladesh on April, 2019.

Board of Examiners

1.
________________________________ Supervisor

Dr. A.K.M. Muzahidul Islam
Professor, Department of CSE, UIU.
Dhaka-1212, Bangladesh.

2.
________________________________ Co-Supervisor

Dr. Mohammad Nurul Huda
Professor & Director-MSCSE Program, UIU.
Dhaka-1212, Bangladesh.

3.
________________________________ Head Examiner

Dr. Khondaker Abdullah Al Mamun
Professor, Department of CSE, UIU.
Dhaka-1212, Bangladesh.

4.
________________________________ Examiner-I

Mr. Suman Ahmmed
Assistant Professor, Department of CSE, UIU.
Dhaka-1212, Bangladesh.

5.
________________________________ Examiner-II

Rubaiya Rahtin Khan
Assistant Professor, Department of CSE, UIU.
Dhaka-1212, Bangladesh.

6.
________________________________ Ex-Officio

Dr. Md. Abul Kashem Mia
Professor & Dean, Department of CSE, UIU.
Dhaka-1212, Bangladesh.

Declaration

This is to certify that the work entitled "**Performance Evaluation of Machine Learning Algorithms for Coronary Artery Disease Features**" is the outcome of the research carried out by me under the supervision of Dr. A.K.M. Muzahidul Islam, Professor and Dr. Mohammad Nurul Huda, Professor & Director-MSCSE program, Department of CSE, UIU.

____

Md. Shah Jalal
Id: 012 162 017
Department of Computer Science & Engineering (CSE)
United International University (UIU)
Dhaka-1212, Bangladesh.

In my capacity as supervisor/co-supervisor of the candidate's thesis, I certify that the above statements are true to the best of my knowledge.

____

Dr. A.K.M. Muzahidul Islam
Professor
Department of Computer Science & Engineering (CSE)
United International University (UIU)
Dhaka-1212, Bangladesh.

____

Dr. Mohammad Nurul Huda
Professor & Director-MSCSE Program,
Department of Computer Science & Engineering (CSE)
United International University (UIU)
Dhaka-1212, Bangladesh.

Abstract

Cardiovascular disease is the leading cause of mortality in the world. Bangladesh probably has the highest rates of cardiovascular disease among all South Asian countries and yet is the least studied. A proper prediction mechanism system can significantly reduce this death toll. In this work, we propose an intelligent system that can make an effectively prediction of a possible cardiac attack using only twelve (12) features. We also apply six (06) well known supervised machine learning algorithms on two different datasets (e.g. NICVD patient's data and UCI dataset) to analyze the prediction accuracy. The overall process can be categorized into four phases. Phase 1: we have provided a comprehensive literature review where we summarize various related machine learning algorithms. Phase 2: we have collectedheart disease patients' data through survey questionnaires from NICVD, Bangladesh to create a dataset. Phase 3: we have reduced feature vector dimensionality with heart disease prediction model. Finally, feed the data to appropriate machine learning algorithms to determine if the predictive model is accurate. It is observed that using NICVD patient's data for 12 features the classification accuracy of Artificial Neural Network (ANN) is 92.80% and it performs better than others classification algorithms such as Decision Tree (82.50%), Naïve Bayes (85%), Support Vector Machine (SVM) (75%), Logistic Regression (77.50%), Random Forest (75%).Whereas using UCI dataset with 12 attributes the classification accuracy of ANN is 91.7% and it also performs better than the other classification algorithms, such as Decision Tree (76.90%), Naïve Bayes (86.50%), SVM (76.33%), Random Forest (67.33%), and Logistic Regression (81.52%).In Bangladesh most of the people don't go for regular medical checkup. The main reasons behind this ignorance are lack of money, as the whole medical process is too costly so most people can't afford this, and lack of awareness.Among all diseases,heart attack is most common and hazardous is not only Bangladesh but also universal. Heart attack occurs when stream of blood cannot go to the brain for any blockage. The angiographic heart disease status is costly and also it is not available in rural area. This intelligent system can be predict heart disease status by measuring some simple information that will be helpful for everyone as it will not take much time and will be budget friendly. It can be assist to medical practitioners or non-cardiologists, especially in rural area.

Acknowledgement

I would like to start by expressing my deepest gratitude to the Almighty Allah for giving me the ability and the strength to finish the task successfully within the scheduled time.

This thesis titled "**Performance Evaluation of Machine Learning Algorithms for Coronary Artery Disease Features**" has been prepared to fulfill the requirement of MSCSE degree. I am very much fortunate that I have received sincere guidance, supervision, co-supervision and co-operation from various persons.

Firstly, I would like to express my heartiest gratitude to my supervisor, Dr. A.K.M. Muzahidul Islam, Professor, Department of Computer Science & Engineering (CSE) and co-supervisor Dr. Mohammad Nurul Huda, Professor & Director of MSCSE Program, Department of Computer Science & Engineering (CSE), United International University, for their continuous guidance, encouragement, and patience, and for giving me the opportunity to do this work. Their valuable suggestions and strict guidance made it possible to prepare a well-organized thesis report.

Secondly, I would like to thank Professor Dr. MD Afzalur Rahman, Director of National Institute of Cardiovascular Disease (NICVD), Associate Professor Dr. Abdul Momen and Dr. Farzana Sultana, Registrar of NICVD, Bangladesh for their assistance and support in this research.

Finally, my deepest gratitude and love to my parents for their support, encouragement, and endless love.

Table of Contents

LIST OF TABLES....vii

LIST OF FIGURES.... viii

1. Introduction 1

1.1 Motivation....2

1.2 Objectives....2

1.3 Contribution....3

1.4 Organization of the chapter....3

2. Background and Literature Review....4

2.1 Coronary heart disease 4

2.2 Overview of machine learning 5

2.2.1 Data preprocessing....6

2.2.2 Dimension reduction techniques....7

2.2.2.1 Feature extraction....7

2.2.2.2 Feature selection....8

2.2.3 Classification algorithms....8

2.2.3.1 Decision Tree....9

2.2.3.2 Random Forest....10

2.2.3.3 Support Vector Machine (SVM)....11

2.2.3.4 Naïve Bayes....12

2.2.3.5 Artificial Neural Network (ANN)....12

2.3 Related work....14

2.3.1 Using fuzzy logic....14

2.3.2 Using data mining techniques....15

2.3.3 Using genetic algorithm..15

2.3.4 Different approaches of J48 Decision Tree..16

2.3.5 Different dataset and Decision Tree (C4.5, Fast)..16

2.3.6 Prototype based IHDPS..17

2.3.7 Literature findings..17

3. Methodology.. 19

3.1 Proposed model..25

3.2 Data Set..26

3.3 Experimental decision..27

4. Results..28

4.1 Experimental results..28

5. Discussion & Conclusion..34

5.1 Summary..34

5.2 Discussion..34

5.3 Limitation..35

5.4 Future work..35

5.5 Conclusion..36

6. References..37

7. Appendix A..40

LIST OF TABLES

Table 1: Various supervised and unsupervised learning algorithms 9

Table 2: Summarizes the results of decision tree using different approaches..................16

Table 3: Comparison of accuracy for different dataset..17

Table 4: Literature analysis of various machine learning algorithms...............................18

Table 5: Previous most used clinical features collected from UCI repository..................20

Table 6: Clinical features name and description..21

Table 7: Heart disease data collected from NICVD, Bangladesh.....................................27

Table 8: Properties of various machine learning classifiers...29

Table 9: Performance analysis of various ML methods using NICVD dataset................30

Table 10: Performance analysis of various supervised methods using UCI dataset with 14 features..31

Table 11: Performance analysis of various supervised methods using UCI dataset with 12 features..31

Table 12:Comparison table of various ML methods performance between 14 and 12 features using UCI dataset...32

Table 13: Comparison table of various ML methods performance between NICVD and UCI dataset...33

Table 14: CA of ML methods using leave-one-out method for NICVD dataset.............33

LIST OF FIGURES

Figure 1: Coronary artery disease with stages of plaque form..4

Figure 2: Overview of machine learning...6

Figure 3: Structure diagram of decision tree 10

Figure 4: Support vector machine diagram..11

Figure 5: Artificial neural network diagram..13

Figure 6: Representation of normal ECG & various types of ST segment.........................24
.
Figure 7: Overview /model of this research...25

Figure 8: Classification accuracy graph of various machine learning algorithms using NICVD dataset...30

Figure 9: Comparison graph of classification accuracy between 14 and 12 attributes using UCI dataset...32

Figure 10: Comparison graph of classification accuracy between NICVD dataset and UCI dataset 33

Figure 11: Model of IoT enabled system..35

Chapter 1

Introduction

Every year more than 17 million people die globally, where 30% of them die due to the cardiac attack and 80% of them happen in the under developing countries [1]. However, the exact number of incidence of the disease in Bangladesh is not known. In most cases the identification of heart disease depends on costly and composite data of pathological and clinical [2].

Machine learning (ML) is among the main approaches that may be used to detect the status of cardiac disease based on the feature of the clinical data [3, 4]. This method is utilized to determine areas of earlier unidentified knowledge. Several techniques such as ANN, SVM, Naïve Bayes, Decision Tree, Logistic regression, and Random Forest may be used in the identification of heart disease. However, it is important to determine the appropriate approach that can provide accurate prediction of a malfunction cardiac status. This research focuses on coronary artery disease features, comprehensive literature review, and various supervised machine learning algorithms for effective coronary heart disease prediction.

Firstly, we provide a comprehensive literature review [6-22] where we summarize various related ML algorithms. We reviewed most of all the related work that has been done till now and learn about the machine learning algorithms they used most.Therefore, we implement these mostly usable machine learning algorithms to train and test our model and find the best one for our system that fit the system best.

Secondly, we construct an actual valid raw survey dataset and a survey questionnaires for obtain data from National Institute of Cardiovascular Disease (NICVD), Bangladesh. Where we have collected 40 exercise tolerance test positive (ETT+) patient's data out of 200 sable angina patients' sample with 12 (twelve) important coronary artery disease features.

Thirdly, we propose an intelligent system to predict effective heart disease prediction using minimized twelve (12) features, namely age, sex, chest pain, blood pressure, cholesterol, blood sugar, ECG results, heart rate, exercise induced angina, old peak, slope, and history of stenosis on coronary artery angiogram (CAG).**Fourthly**, we apply six (06) different ML algorithms (e.g., Artificial Neural Networks, Decision Tree, Random Forest, Naïve Bayes, Support Vector Machine, and Logistic Regression) on two different datasets(e.g., NICVD patient's data and UCI: dataset). Finally, we analyze their performances.

1.1 Motivation

Cardiovascular disease means a class of malady that involves to the heart or blood veins. Among all diseases, heart disease is most common and hazardous in not only Bangladesh but also universal. Heart attack occurs when stream of blood cannot go to the brain for any blockage. The blockage is most often a buildup fat, cholesterol or any other substances. The diagnosis of heart disease is a step by step process, angiogram is one of them. The angiographic coronary heart disease status is costly and time consuming. In 2005, among all causes of deaths, cardiovascular disease (CVD) was accounted in favor of 30% [2]. Deaths caused by cardiovascular diseases have increased by 17% between 2006 and 2015, whereas deaths from other diseases decline by 3% [2]. Thus, this problem has become a big threat. If a system can developed which can predict heart attack by measuring some simple information that will be helpful for everyone as it will not take much time and will be budget friendly.The goal of this research is, developing a system or structure for predicting heart disease status easily by some simple heart disease symptoms. For this, we have used machine learning (ML) algorithms. ML is an approach which has been developed to teach machine how to recognize distinctive features in datasets and can predict disease status.

1.2 Objectives

- To collect or construct actual valid raw data from NICVD, Bangladesh.

- To construct a survey questionnaire for collect heart disease data.
- To develop a system/model to predict coronary heart disease using minimum features with low error.
- To analyze the cardiac data using various machine learning methods/data mining techniques for better prediction and accuracy.
- To analyze the performances of different dataset with different machine learning algorithms.

1.3 Contribution

- We have provided a comprehensive literature review [6-20], where we summarize various related ML algorithms.
- We have constructed a survey questionnaire for collect clinical heart disease data of stable angina patients.
- We have constructed a dataset of heart disease with ETT+(Exercise Tolerance test positive) patients.
- We have reduced feature vector dimensionality 14 to 12.

1.4 Organization of the chapter

This book of research is organized as follows. Chapter 2 describes the background and literature review with various literatures findings. Chapter 3 describes the methodology of this research with existing features & proposed features, proposed model, describes the dataset, and describes the features. Depict the experimental results in Chapter 4. Chapter5 concludes the research with limitation and shows some future work directions.

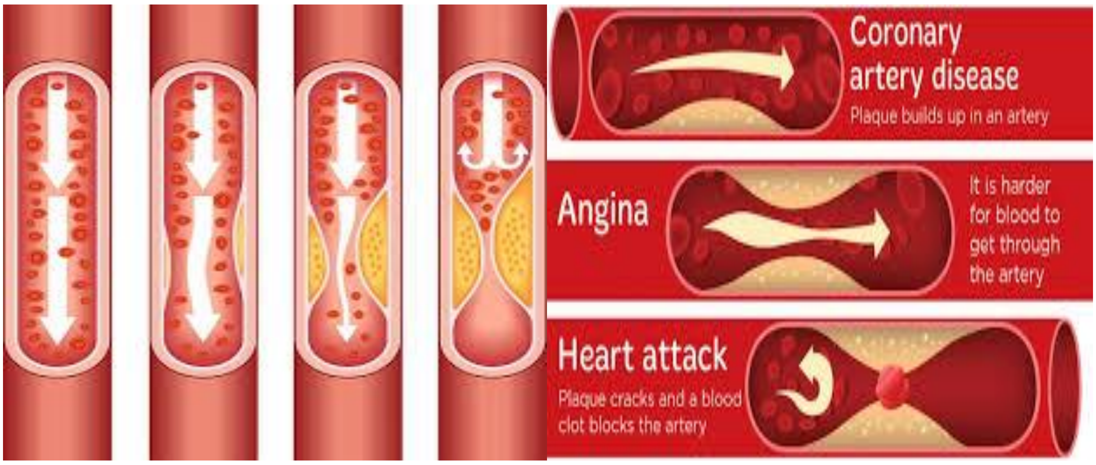
Chapter 2

Background and Literature Review

2.1 Coronary heart disease

The disease of the heart and coronary arteries is known by the coronary heart disease (CHD) caused by the increased fatty materials in the blood vessels. This is the root of cheat pain, heart attack. A heart attack happened when the brain is interrupted due to the blood vessels cannot supply blood with oxygen to the heart.Coronary heart disease generally appears when the blood vessel becomes blocked.

Coronary artery disease (CAD) generally ensures when cholesterol gathers on the artery walls; creating plaques (see the Figure 1(a)). The heart’s network of blood vessels is coronary artery (CA). The narrow arteries that decrease blood flow to the heart Coronary arteries are the heart's net of blood vessels. The supply of oxygen/ blood to the heart may become decrease, while the coronary arteries narrow shown in Figure1(b), especially during physical movement.



Stage-1 **Stage-2** **Stage-3** **Stage-4**

Figure 1(a): Coronary artery disease with stages of palque forms

In above this Figure 1 we have observed, at **stage-1:** normal blood flow (heart ok), **stage-2:** plaque forms in lining of artery (this reduction in blood flow may not produce any symptoms), **stage-3:** plaque grows; as fatty deposits, or plaques, build up in the coronary arteries (lining of artery is damaged), **stage-4:**plaque cracks (Heart Attack).

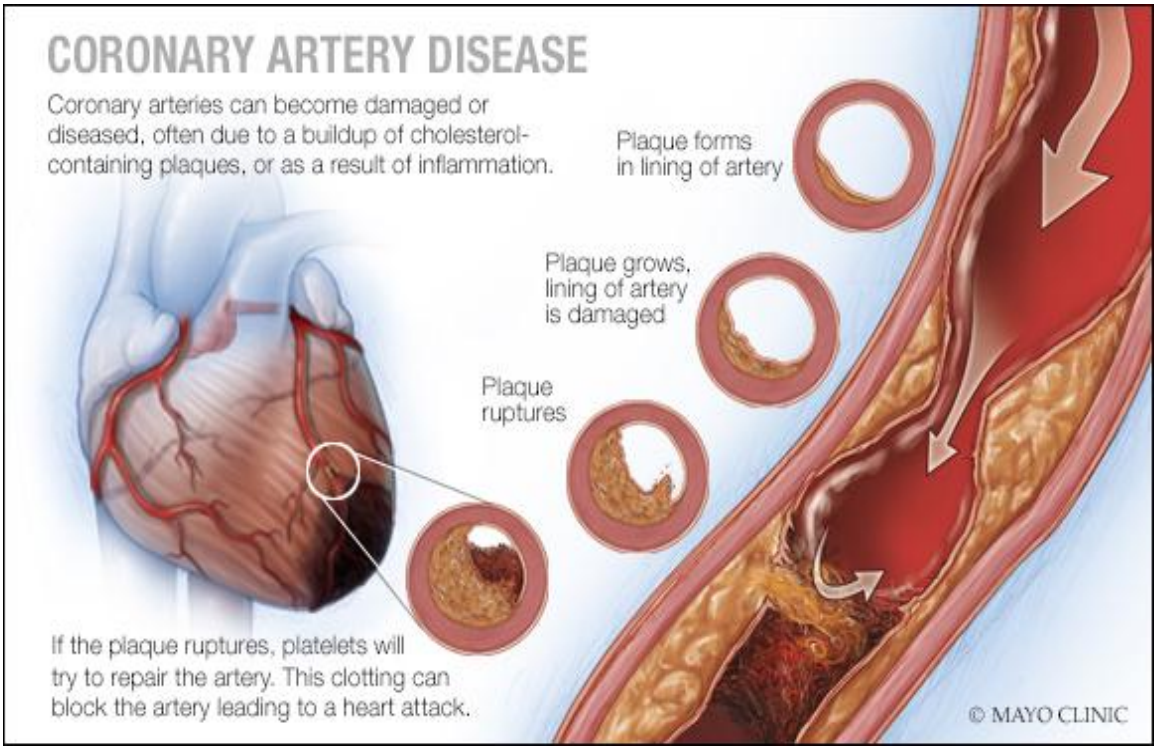


Figure1(b): Coronary artery disease [collected through onloine].

Nowadays, ML is rapidly growing attention to the research community. This is because ML method can produce reliable results and can learn from earlier computation. ML can be main approaches to predict heart attack based on clinical data, where analyzing data or building predictive models manually could be impossible. Data that are used in ML are basically of two types: unlabeled and labeled data. In labeled data, features are given and have some class of mark or sense attached to the data, therefore used in supervised learning (SL). Labeled feature is either numerical or class based (categorical). Numerical data are used for predicting the value and class based data are used in classification. Whereas, in unlabeled data, have only input data and according to input data no class exists to assist. It is used in unsupervised learning (USL) to categorize the patterns or any construction present in the dataset. Thus, the labeled data and unlabeled data are used with SL and USL, correspondingly [5].

2.2 Overview of machine learning

Machine Learning(ML) is the knowledge and practice of algorithms, mathematical optimization and statistical methods that is concerned with computer programs. In computer science ML is part of Artificial Intelligence (AI), which is used for data analysis, prediction, forecasting, knowledge discovering and improve their performance progressively.

ML algorithms construct a mathematical model by various sample data. According to sample data specification machine learning can make a decision with better performance. The sample data is divided into two parts, namely training data and test data. These are used in ML methodsare delivered from mathematical optimizationand is closely related to statistical method. In ML, training detailsused for learning knowledge discovering,whereas the test data is usedto evaluate accuracy and performance of a specific task.The Overview/Model of Machine Learning is shown in Figure 1.

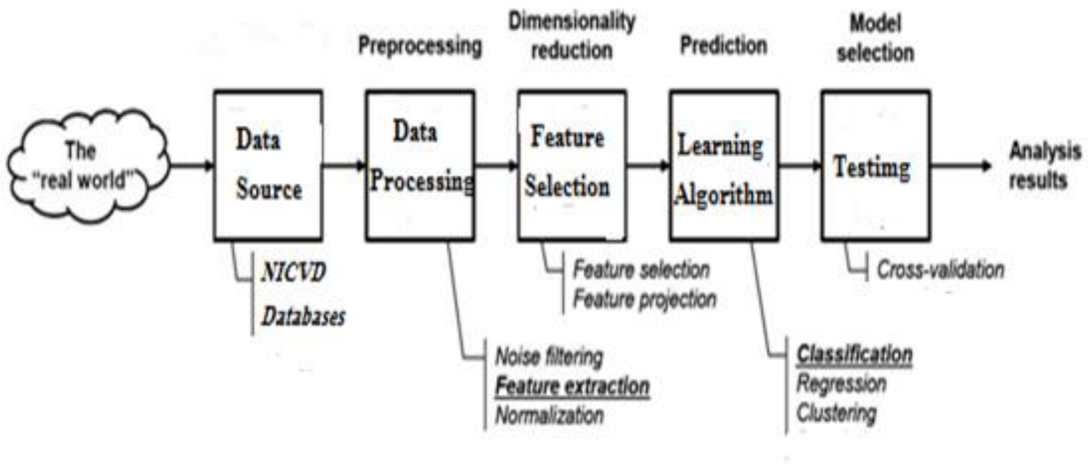


Figure 2: Overview of machine learning

2.2.1 Data preprocessing

In the present advanced world, there is tremendous amount of data gathered from various sources like machine: internet, sensor; people: surveys; and organizations: databases etc. In this research, data is collected from National Institute of Cardiovascular Disease(NICVD) through surveys questioners with clinical information and UCI

(University of California)database. Since data is gathered from various sources it requires filtering (reduce noise&distortions), cleaning (reduce missing & garbage value), processing and storing into database or data warehouse for analysis.

Data preprocessing isthe preliminary step in the machine learning overview.Regarding data learning (i.e. data training and data testing) anddata analysis it requires a suitable data format. Sometimes data preprocessing is also known as data wrangling (e.g. data in CSV format, noise reduction, normalization, etc.), where data wrangling refers to be a set of procedures which transforms raw data into suitable environmental data format. In this research, we have applied CSV (Comma Separated Value) data format. To obtainbetter accuracy we applied data normalization and data standardization operation on dataset. Data normalization operation is done when the features or attributes contain different assessing value.

2.2.2 Dimension reduction techniques

A dimensionality retrenchment or reduction technique is themethod or cognitive operation to eliminate a large number of variables or dimensions without losing information.A large number of dimensions, duplicate variables, and duplicate or large dataset increase the inconsistency and complexity. To reduce inconsistency, time complexity, computational complexity with better performance including accurate analysis dimensionality retrenchment techniques is required[19, 109-110].Dimensionality reduction techniques have two types: Feature Selection andFeature Extraction.Feature extraction and selection is described below.

2.2.2.1 Feature extraction

Feature extraction is a dimensionality reduction method, where an early set of variables is reduced to others convenient groups (features) for analysis & processing. While the dataset is describe correctly and entirely. According to input data, the method of transforming to the set of features is called feature extraction. Feature extraction is the techniques of extract features from the input data by transforming method, where features are represent the input data

There are some benefits of feature extraction process are:

- Decrease the cost of feature dimensionality.
- Build up classifier effectiveness.
- Increase the classification accuracy.

2.2.2.2 Feature selection

In subset or feature selection, *d* dimensions or feature are selected out of *D* dimensions with valuable information and eliminate the (*D-d*) features. The best feature or subset provides the least number of dimensions with low error function and better accuracy [19,109-110].There are some benefits of feature selection technique are:

- Reduce Over fitting: Less over fitting or duplicate features, redundant data can make decision quickly based on noise.

- Improves Performance: Less misleading features provides the better accuracy.

- Reduce Computational Time: Least dimensions or features mean that learning algorithm train faster.

2.2.3 Classification algorithms

Machine Learning method or classification algorithm can produce reliable results and can learn from earlier computation where analyzing data or building predictive models manually could be impossible. Data that are used for machine learning are basically of two types: unlabeled data use for unsupervised learning and labeled data use for supervise learning. The various supervised learning and unsupervised learning of Machine learning algorithms are given in Table 1.Supervised Learning provides a predictive learning model by applying input variables a set of X and output variables Y [5]. Unsupervised learning provides a learning model by applying input variables a set of X with unknown output. By re-present particular input patterns of data the system can learn and perform complex processing task.

Table 1: Various supervised and unsupervised learning algorithms

Machine Learning Algorithms	
Supervised Learning	***Unsupervised Learning***
Decision Tree	K Mean Clustering
Naïve Bayes	Self-organizing Map
Artificial Neural Network (ANN)	Fuzzy C-Mean Clustering
Support Vector Machine(SVM)	Hierarchical Clustering
Hidden Markov Model (HMM)	
K-Nearest Neighbor (KNN)	
Random Forest	
Bagging	
Logistic Regression	

In this research,we have applied well known most popular supervised learning algorithms. Some most popular supervised learning algorithms described below.

2.2.3.1 Decision Tree

Decision tree is a supervised learning classifier and most powerful tools which can be used in discrete/continuous datafor prediction or classification.In machine learning decision tree represents tree structure form and classifies the data or instance by starting at the root of the tree has no incoming edges& moving through it until a leaf node has exactly one incoming edge. Decision tree contains decision nodes, leaf nodes, edges and path. Using this contents decision tree can make a decision by input objects or a set of attributes. There are some algorithms used in decision trees like ID3(Iternative Dichotomizer), C4.5, CART, C5.0, and J48 in WEKA tools etc.This research we have applied ID3 algorithm in MATLAB and J48 in WEKA. Structure diagram of decision tree shown in Figure 2.

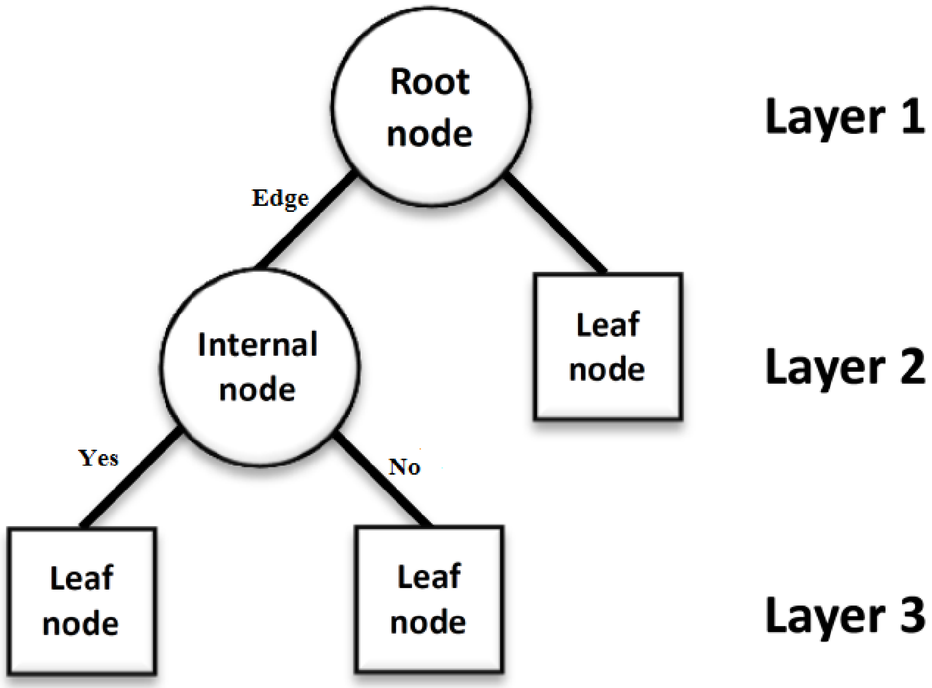


Figure 3: Structure diagram of Decision Tree

2.2.3.2 Random Forest

Random forest algorithm is a supervised classifier that builds various decision trees and mixes them collectively to acquire better performance. Random forest creates a forest randomly, where each decision tree (DT) is formed by applying this method on the dataset. Random forest provides the decision or prediction by a majority vote over the entity trees' prediction [19].

The Random Forest method works in following steps:

1. Select the point's data of k randomly from the input training dataset.
2. Constructs a tree regarding these data points.
3. Select N tree subset from the decision trees and executes the step 1 & step 2.
4. On basis of majority's vote determine the result.

2.2.3.3 Support Vector Machine (SVM)

A Support Vector Machine (SVM) is a supervised machine learning algorithm or discriminative classifierdefined by a separating hyper-plane. This hyper-plane is a line dividing a plane in two parts in two dimensional spaces where in each class lay in either sides and distinct the various classes of data.Support vector machine constructedwith the training data and it outputsthe hyper-plane in the test data [20]. It tries to find the space in the matrix of data where different classes can be widely separated and draws a hyper-plane.

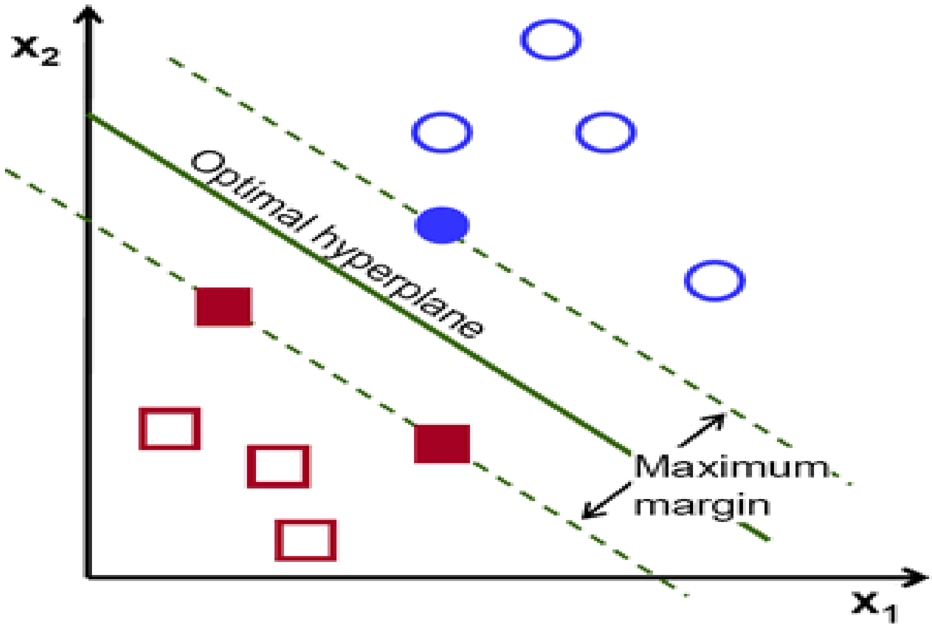


Figure 4: Support Vector Machine.Collected from introduction to SVM [21]

In Figure 3, the classes of training data points defined by red and blue color where data is labeled. In this space have more than one way to draw a hyper-plane for classify the data linearly. An optimal hyper-plane is taken which maximizes the margin between the classes. Support Vector Machine (SVM) can also work as a non-linear classifier using kernel-trick function.

2.2.3.4 Naïve Bayes

Naïve Bayes is a well-known supervised machine learning algorithm or classifier. To classify the data Naïve Bayes applies the Bayes theorem to classify the data and it assumes that the probability of certain attribute X is totally independent of another attribute Y [21,397]. Bayes theorem provides a theory to compute the probability with prior knowledge of the hypothesis.

$\text{Posterior} = \frac{\text{Likelihood} * \text{Prior}}{\text{Evidence}}$,where Posterior -is the posterior probability of class (target) given predictor (attribute), Likelihood -is the probability of predictor given class, Prior -is the prior probability of class, Evidence -is the prior probability of predictor.

There are three types of Naïve Bayes: Gaussian Naïve Bayes use for classification problems, Multinomial Naïve Bayes is applied in multinomial distributed data problems and Bernoulli Naïve Bayes is applied in data with multivariate Bernoulli distribution problem. In this research we have applied Gaussian Naïve Bayes theorem.

2.2.3.5 Artificial Neural Network (ANN)

A neural network comprises of a band of neurons (working element). Each neuron associated to other neuron by direct connection, each connection has some weight committed with them. The weight depicts knowledge used by network to solve a problem. Neurons are reassembled in layers: an input layer, an output layer and one or more hidden layers. Data disseminates from one layer to other layer by neurons which have some weight and input vectors connected with them. For each neuron, the weighted input vectors are tallied and a threshold valueθjis added. This summed up input Ij is then passed through an activation function (a non-linear function) f(Ij) to develop the output of the neuron yj. The output of one neuron renders the input to the neurons in the next layer. Mathematically it is represented as:

$$I_j = \sum W_{ji} X_i + \theta_j \ldots\ldots\ldots\ldots\ldots\ldots \text{(1)}$$

$$y_j = f(I_j) \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(2)}$$

There are varieties of neural network models and learning procedures. In this analyze, Multilayer perceptron (MLP) neural network has been used. MLP is the most important neural network. It employs linear combination function in input layer to compute single output from multiple inputs and then enforce nonlinear activation function on yielded output. The universal structure of MLP neural network is shown in Figure 4.

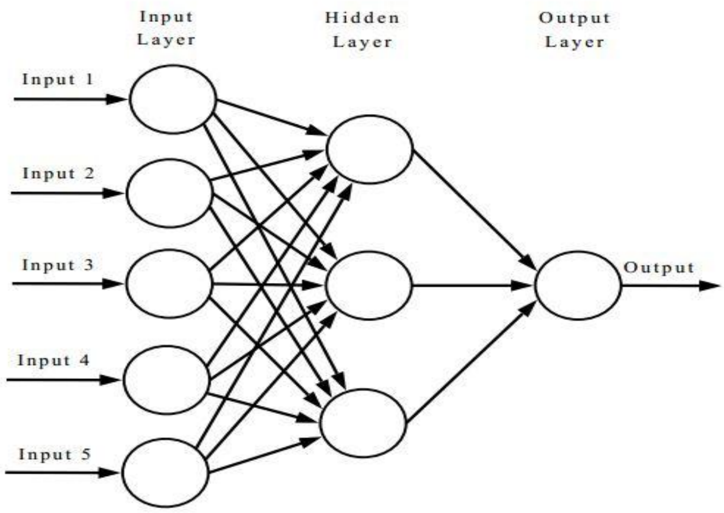


Figure 5: Artificial Neural Network

In the first layer, we have input layer which have the input neuron. Each input neuron represents an input data. Second layer is called hidden layer. MLP can consist of more than one hidden layer. In the last layer, it is the output layer which has output neurons. Output neurons comprise the anticipated value. Mathematically, this is represented as:

$$Y = \varphi\left(\sum_{i=0}^{n}(W_i X_i + b)\right) \ldots\ldots\ldots\ldots\ldots\ldots.(3)$$

Where $W_i$announces the vector weights, X_i is the vector of inputs (i=1, 2…...n), b is the bias, Y is the output and φ is the activation function [22]. MLP is schooled using back propagation algorithm. There are tons of ways, where back propagation algorithm can be enforced. At the origin of the training the network weights are initialized with random value.

This algorithm operates iteratively by advancing a network's interlinking weights such that the overall fault is reduced. MLP can be employed where we have limited knowledge of relationship between inputs and targets.

The generalization computational model of central nervous system can be represented as an artificial neural network. Neural network has potentiality to study from experience to evoke performance of model. Artificial Neural Network (ANN) has firm capacity of pattern recognition, classification and forecasting from their learning experience. Artificial Neural Network (ANN) applies the conception of nonlinear mapping which is practicable when data given is incomplete and noisy where principles cannot be determined.

2.3 Related work

In this work we have reviewed some related works. Where we tried to read out or find out the gap of these work. The most recent papers we have reviewed. Most of the authors of these related works have presented UCI dataset with 14 features. And also they have implemented some different ML methods. In details of related work described below.

2.3.1 Using fuzzy logic

Fuzzylogic and set theory can be suitable for developing knowledge basedsystems for diagnosis of diseases in healthcare. In [6], the authors have proposed a method that combines the genetic algorithms for feature selection and fuzzy expert system (Mamdani Model) for effective classification. Here experiments are taken by using fuzzy tool in MATLAB. The dataset is used from UCI machine learning repository and six (06) attributes are used in the experiment. In this system, the input is the set of all selected features, whereas the output is either a value 0 or 1 (0-absence, 1-presence) of heart disease in a patient. Here, genetic algorithms and fuzzy logic are used. It is to be noted that fuzzy logic is a mathematical tool and is a subfield of AI. The problem with using only fuzzy logic in AI is that the system can only implement the rules and cannot learn as it goes along.

2.3.2 Using data mining techniques

In [7], the authors have presented a comparison of four (04) different data mining algorithms that predict the risk of heart diseases, namely C5.0, Neural Network, Support Vector Machine (SVM), and K-Nearest Neighborhood (KNN). It is observed that C5.0 Decision tree has greater accuracy of 93.02% than KNN (88.37%), SVM (86.05%), and Neural Network (80.23%). The results produced using decision tree is interpretable, applicable and easily understandable by different clinical practitioner. Here, 270 records with thirteen (13) features are used from the UCI and four (04) classifiers including C5.0, SVM, KNN and Neural Network are used. Data divided into training set and test set (70% and 30%, respectively). The training set is used to build the classifier and test set used to validate it. It is observed that decision tree outperforms others with 93.02% accuracy and the accuracy result of ANN is very low.

In [8], the authors have presented data classification based on various ML algorithms, namely KNN, Naïve Bayes, and Decision List. A data mining tool, known as TANAGRA, is used to classify the data where the data is evaluated using 10-fold cross validation. Here the training data set of 3000 instances with fourteen (14) different attributes is used in the experiments. Depending upon the attributes, the dataset is classified into two parts: 70% of the data is used for training and 30% is used for testing. It is observed that the Naïve Bayes algorithm (52.33% accuracy with 609ms) performs better than the other two algorithms i.e., Decision List with 52% accuracy and KNN with 45.67% accuracy and the accuracy of these algorithms are very low.

2.3.3 Using genetic algorithm

In [9], in order to reduce the number of attributes, from thirteen (13) attributes to six (06) attributes, the authors have proposed Genetic Algorithm which is applied on a data set of 909records. Different classifiers corresponding to these attributes are used for cardiac status prediction. It is observed that the Decision Tree has outperformed others with highest accuracy and least mean absolute error. Here, the authors used a limited number of reduced attributes such as chest pain, exercise induced angina, number of major vessel colored, resting blood pressure, maximum heart rate achieved and old peak. However, for an intelligent and effective heart disease prediction important attributes such as age, sex, resting ECG, fasting blood sugar, cholesterol, the slope of the peak exercise ST segment is totally ignored.

2.3.4 Different approaches of J48 Decision Tree

In [17], the authors have presented a prediction model of heart disease using 14 attributes including heart disease results collected from UCI database. They have applied decision tree J48 algorithm using different approaches, such as pruned, un-pruned, and reduced error pruning approach for prediction heart disease based on these attributes. Where they observed J48 reduced error pruning approach is (almost 76%) better than J48 un-pruned (almost 73%) and J48 pruned (almost 74%) approach. Also they have observed that fasting blood sugar is most important feature which gives better classification result than other features. However, the problem is fasting blood sugar is not provided good accuracy. The summary of the experimental results of decision tree J48 algorithm using different approaches are given in Table 2.

Table 2: Results of decision tree using the different approaches

Performance	**Pruned approach**	**Un-pruned approach**	**Reduced Error Pruning approach**
Accuracy	73.79 %	72.82 %	**75.73 %**
Precision	72%	68%	74%
F -Measure	77%	78%	79%
No.Correct	76	75	78
No. of Rules	33	14	11
Size of Tree	56	24	17

2.3.5 Different datasets and Decision Tree

This study has presented the datasets conducting with the heart disease problem, gives various results according to implementing by different decision tree (C4.5 and Fast Decision Tree) of machine learning algorithm on different datasets.

This paper has been viewed the result of accuracy of different decision trees. They have applied different dataset collected from various sources, especially UCI repository and have applied 10 fold cross validation on environment of WEKA tool. The dataset have contained 14 features. The Classification Accuracy (CA) of different decision tree compared shown in Table3.

Table 3: Comparison of accuracy for different dataset

Dataset	**CA of C4.5**	**CAof FDT**	**Records**
Cleveland	78.54%	77.55%	303
Hungarian	78.57%	78.23%	294
V.A	71.50%	69.50%	200
Statlog	76.60%	76.60%	270

2.3.6 Prototype based IHDPS

In [10], the authors have proposed a prototype (web-based system implemented on the .NET platform) known as Intelligent Heart Disease Prediction System (IHDPS), which is based on data mining techniques e.g., Neural Networks, Naïve Bayes and Decision Trees. A total number of 15 attributes were considered in IHDPS and 909 records collected from Cleveland Heart Disease Database were used in the system. It is observed that the prediction of Naïve Bayes achieves better results than Neural Network and Decision Trees.

2.3.7 Literature findings

We have summarized various related ML algorithms [7-19]. Previous most have used different machine learning methods and performances (accuracy, precision, recall) with have applied heart disease dataset (dataset name and no. feature). Table 4 summarizes the findings.

Table 4: Literature analysis of various machine learning algorithms

Techniques/Methods		Reference NO.	Accuracy (%)	Precision (%)	Recall (%)	Number of features	Dataset
Supervised Learning (SL)	***Unsupervised Learning (USL)***						
Decision Tree/List	-	[7]	93.02	90.90	-	14	UCI
		[8]	52.00	48.55	48.97	14	Sensing Data
		[9]	99.20	99.78	99.78	15	Sellapanetal
		[10]	89.00	-	-	15	Cleveland,UCI
		[11]	89.00	-	-	14	APRHI Data
		[12]	89.60	-	-	14	Survival Data
		[13]	72.93	82.60	82.20	14	Cleveland,UCI
		[18]	92.50	92.50	92.50	14	UCI
Naïve Bayes	-	[8]	52.33	-	-	14	Sensing Data
		[9]	96.50	96.60	96.80	15	Sellapanetal
		[10]	86.53	-	-	15	Cleveland,UCI
		[11]	83.53	-		14	APRHI Data
		[16]	82.31	93.10	85.70	14	UCI
		[18]	91.20	92.30	91.20	14	UCI
Random Forest	-	[18]	88.70	-	-	14	UCI
SVM	-	[7]	86.05	89.47		14	UCI
		[12]	92.10	-	-	14	Survival Data
		[18]	68.80	72.70	68.80	14	UCI
ANN	-	[7]	80.23	83.78		14	UCI
		[10]	85.53	-	-	15	Cleveland,UCI
		[11]	83.00	-	-	14	APRHI Data
		[12]	91.00	-	-	14	Survival Data
KNN	-	[7]	88.37	88.09	-		UCI
		[8]	45.67	54.79	45.52	14	Sensing Data
-	Clustering	[9]	88.30	83.26	95.20	15	Sellapanetal
-	k-mean based	[14]	74.00	78.00	67.00	-	-
Decision Tree J48	-	[16]	84.35	86.20	97.20	14	UCI
		[17]	75.73	-	-	14	UCI
Bagging		[16]	85.03	86.10	98.40	14	UCI

In the summary of this chapter**,** we have provided a strong background of CHD with the overview of machine learning. We reviewed mostof the related work that has been done till now and learn about the machine learning algorithms they have used most. Most of researchers conducted UCI dataset with 14 or 15 attributes. They have applied different data mining techniques Therefore, we implement these mostlyusable machine learning algorithms to train and test our model and find the best one for our system that fit the system best.

Chapter 3

Methodology

Coronary angiography is the important with standard tests for the identifying /detection the CAD. As with any invasive approach, there are some difficulties of procedure and patient related hasslesthat are innate to the test. Coronary angiogram is more complex, time consuming and costly to predict heart disease; especially in Bangladesh. We propose a heart disease prediction model to assist medical practitioners, non-cardiologists, professionals in predicting heart disease status based on the patient's clinical datafrom cardiologist report. To avoid coronary angiogramhassle and to predict CAD easilywe have used 12 most important clinical features in this model. This features are most responsible for occur heart disease.

We havecollected the data of 40 ETT+ (Exercise Tolerance Testpositive) patientsthrough survey questioners out of 200 stable angina patients in 6 (six) months from NICVD, Bangladesh. Where 23 patients has heart disease and 17patients has no heart disease.According to ETT positive patients, we have collected 16 clinical features of coronary artery disease. In this research we have used 12 important clinical features of coronary artery disease.

Also we have used UCI machine learning dataset collectedfrom the University of California, Irvine (UCI) for more analysis in this research.It has 303 patients record with 75 attributes, where 164 instances or patientshasno heart disease and 139 patients or instances has heart disease. Earlier most of researchers conducted this dataset for 14 or 15 features (see the Table5.) including angiographic disease status or data (heart disease=yes, heart disease=no) and applied different data mining techniques [6-25]. But in this research we have applied most important 12 clinical features [proposed] and previous most have used 14 features or attributes. Also we have implemented different supervised learning methods on these datasets for comparison.

Table 5: Previous most used clinical features collected from UCI repository [10]

SL No.	Features Name	Description
1	Age	Instance age in years (33-72)
2	Sex	Instance gender (1 = male; 0 = female)
3	CP	Chest pain type (1= typical angina, 2=atypical angina, 3=non-angina pain,4=asymptomatic)
4	RBP	Resting blood pressure (Systolic in mmHg on admission to the hospital), normal 120/80 mm HG
5	Cholesterol	Total serum cholesterol in mg/dl (<200 mg/dl)
6	FBS	Fasting blood sugar > 120 mg/dl (1= T; 0 = F)
7	ECG	Resting electrocardiographic results, results (0=normal, 1=abnormal)
8	HR	Maximum heart rate achieved
9	EIangina	Exercise induced angina (1 = yes, 0 = no)
10	Old peak	ST depression induced by exercise relative to rest
11	Slope	The slope of the peak exercise ST segment (value1= up sloping, 2=flat, 3=down sloping)
12	Stenosis	1 = normal; 2 = defect; 3 = reversible defect
13	Num	Number of major vessels (0-3) colored by fluoroscopy (Angiographic disease status)
14	CA	Diagnosis of heart disease, Value 0 =< 50% diameter narrowing, 1 = > 50% diameter narrowing (Angiographic disease status)

To avoid coronary angiogram hassle or complex procedure of angiographic status and to reduce angiogram test cost, we have used 12 most important clinical features in this research model. Where we have discarded 2 angiographic features (Num& CA) from UCI heart disease dataset, and have updated 1 feature (stenosis). These12 features are most responsible for occur heart disease and can play as good predictor to predict heart disease status without angiographic features. There are some coronary artery diseases features, which are presented in this research. The proposed features are described in Table 6 and below.

Table 6: Clinical features name and description

SL No.	Features Name	Description
1	Age	Instance age in years (30-74)
2	Sex	Instance gender (1 = male; 0 = female)
3	CPT	Chest pain type (CPT) (1= typical angina, 2=atypical angina, 3=non-angina pain, 4=asymptomatic)
4	RBP	Resting blood pressure (Systolic in mm Hg on admission to the hospital), normal 120/80 mm HG
5	Cholesterol	Total cholesterol in mg/dl (<200 mg/dl)
6	FBS	Fasting blood sugar (FBS)>120 mg/dl or 6.5mmol/L (1= T;0 =F)
7	ECG	Resting electrocardiographic (ECG) results, results (0=normal, 1=abnormal)
8	HR	Maximum heart rate (MHR)
9	EIangina	Exercise induced angina(EIangina) (1 = yes, 0 = no)
10	Old peak	ST depression induced by exercise relative to rest
11	Slope	The slope of the peak exercise ST segment (value 1= up sloping, 2=flat, 3=down sloping)
12	History of Stenosis	1 = normal, 2 = fixed defect, 3 = reversible defect;(=>70% stenosis on CAG of 2mm or more sized vessel,=>50% LM)

Age: Aging increases risk of damaged and narrowed arteries and weakened heart muscle. Men age 45 or older and women age 55 or older are more likely to have a heart attack than are younger men and women. In this survey we have collected different patient of age; 30 to 75 years.

Sex: Men are generally at greater risk of heart disease. However, women's risk increases after menopause. We have collected 9 woman and 31 men’s data where the woman age 40 or older. We have considered instance age in years 1= male, 0= female.

RBP: High blood pressure can damage arteries that feed our heart. Uncontrolled high blood pressure that occurs with other conditions, such as obesity, high cholesterol or diabetes, narrowing the vessels through which blood flows increases risk. Blood pressure is measured by two numbers. The first number, called systolic blood pressure, measures the pressure in our blood vessels when our heart beats. The second number, called diastolic blood pressure, measures the pressure in our blood vessels when our heart rests between beats. If the measurement reads 120 systolic and 80 diastolic, we would write 120/80 mmHg. A blood pressure less than 120/80 mmHg is normal. A blood pressure of 140/90 mmHg or more is too high. People with levels in between 120/80 and 140/90 have a condition called pre-hypertension, which means they are at high risk for high blood pressure. We have collected **Systolic** blood pressure in mmHg all over the patients' data on admission to the hospital.

CPT: Chest pain as a squeezing, pressure, heaviness, tightening, burning, or aching across the chest. It usually starts behind the breastbone. The pain often spreads to the neck, jaw, arms, shoulders, throat, back, or even the teeth. In this survey we have classified chest pain into four classes. We have considered value1= typical angina;Typical angina is define as the presence of sub-sternal chest pain or discomfort that was provoked by exertion or emotional stress and was relieved by rest. The history of the patient is classical and chance of having coronary artery blockages is high, 2=atypical angina;atypical angina is used to describe a form of anginal chest pain that does not fit the typical presentation. Typical angina symptoms occur during times of stress or activity due to decreased blood supply to the heart, 3=non-angina pain; Patients has no idea about pain, 4=asymptomatic; it means that do not show any symptoms of the pain or disease.

Cholesterol: LDL cholesterol (bad cholesterol) alone is a relatively poor predictor of risk. In biomedical report or blood test report we have found total cholesterol, HDL cholesterol, and LDL cholesterol. An optimal level of LDL cholesterol is <130 milligrams per deciliter (mg/dL), or 3.37 millimoles per liter (mmol/L), Total cholesterol is <200 mg/dL (5.2 mmol/L), and HDL is >40 mg/dL (1.3 mmol/L). In this survey we have considered total cholesterol.

FBS: Diabetes increases risk of heart disease. A fasting blood sugar level less than 100 mg/dL(5.6 mmol/L) is normal. A fasting blood sugar level from 100 to 125 mg/dL(5.6 to 6.9 mmol/L) is considered pre-diabetes. If it's 126 mg/dL (7.0mmol/L) or higher on two separate tests, it considered diabetes. In this research we have considered fasting blood sugar>120 mg/dl, where 1=True, 0=false.

ECG: An electrocardiogram (ECG) is a test which measures the electrical activity of heart to show whether or not it is working normally.Representation of normal ECG wave has shown in Figure 6(a). From ECG report we have considered Value 0: normal, Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV). We have extracted the value from ECG report.

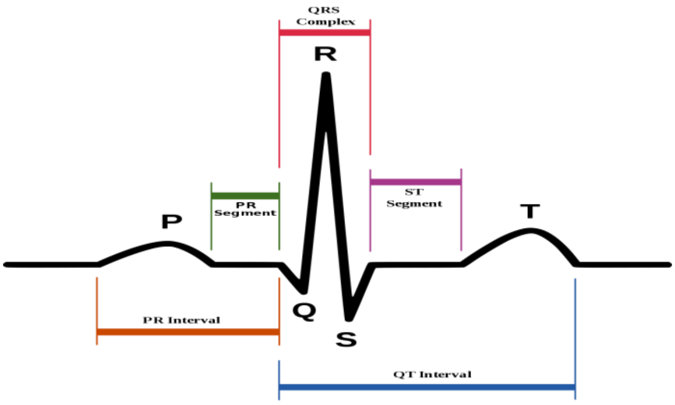


Figure 6(a): Representation of normal ECG with PQRST waves

HR: Maximum heart rate extracted from Exercise Treadmill Test report (ETT). The age related number of beats per minute of the heart when working at its maximum that is usually estimated as 220 minus one's age. The calculation is simple with this method: 226 – age for women, 220 – age for men. Let, a men age is 47 years. His maximum heart rate will be (220-47)= 173 beats per minute (bpm). In order to remain in the endurance zone, between 60 and 75% of his MHR, the number of beats of his heart during the session will be between 103.8 or104bpm and 129.75 or 130bpm (173x60%=103.8, 173x75%=129.75). If target heart rate is not achieved with poor exercise capacity causes will be ETT positive. In this research we have considered HR in bpm.

EIangina: Exercise-induced angina is a common complaint of cardiac patients,patient acute chest pain during exercise. The patient has request stopping, because of acute chest pain. We have considered exercise induced 1= yes means acute chest pain,0= no.

Oldpeak: ST depression induced by exercise relative to rest. ST segment depression may be determined by measuring the vertical distance between the patient's trace and the isoelectric line or base line at a location 2-3 millimeters from the QRS complex. It is significant if it is more than 1 mm in V5-V6, or 1.5 mm in AVF or III. In this research we have considered ST depression under the base line (see the figure 6(b)).

Slope: The ST segment depression has three types shown in Figure 6 (b). The normal ST segment has a slight upward concavity. Flat, down-sloping, or depressed ST segments may indicate coronary heart attack. We have considered value 1= up sloping, 2=flat, 3=down sloping.

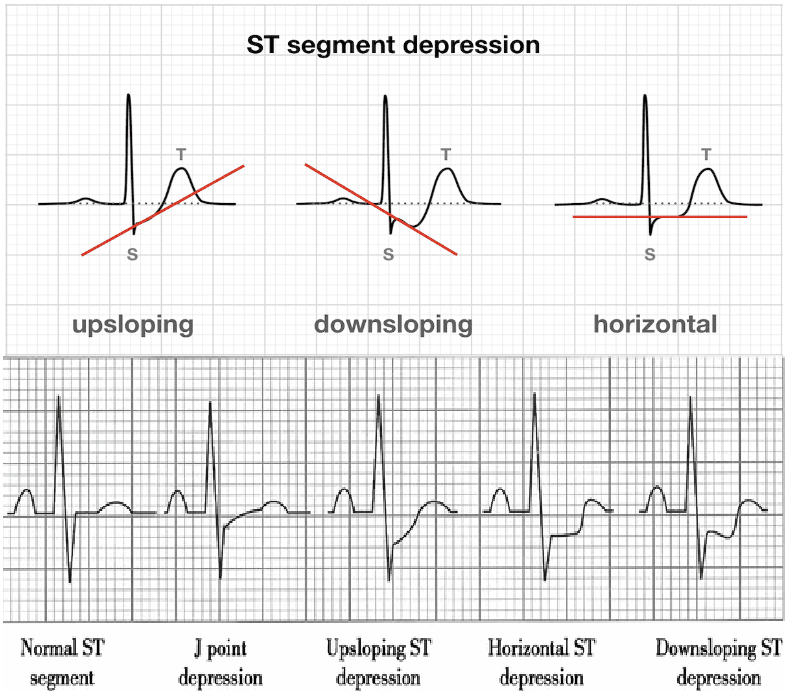


Figure 6(b): Representation of ST segment depression with various types.

Stenosis: There are three arteries that run over the surface of the heart and supply it with blood, where is one artery on the right side, and two arteries on the left side. The one on the right is known as the right coronary. On the left side, which is the main side, typically we call heart blockage less than 40% mild. We have considered 70% or more stenosis on CAG of 2mm or more sized vessel on the right side, and 50% or more thanon the Left middle is very significant for heart attack.

3.1 Proposed model

The proposed model (see the Figure 7(a)) of this work described below:

- To collect actual valid raw survey data from NICVD, Bangladesh.
- To develop a system to predict heart disease from cardiologist report using minimum features with low error.
- To analyze the cardiac data using various machine learning methods/data mining techniques for better prediction and accuracy.
- To analyze the performance of different datasets with different machine learning algorithms.
- To validate the heart disease prediction result by medical professionals.

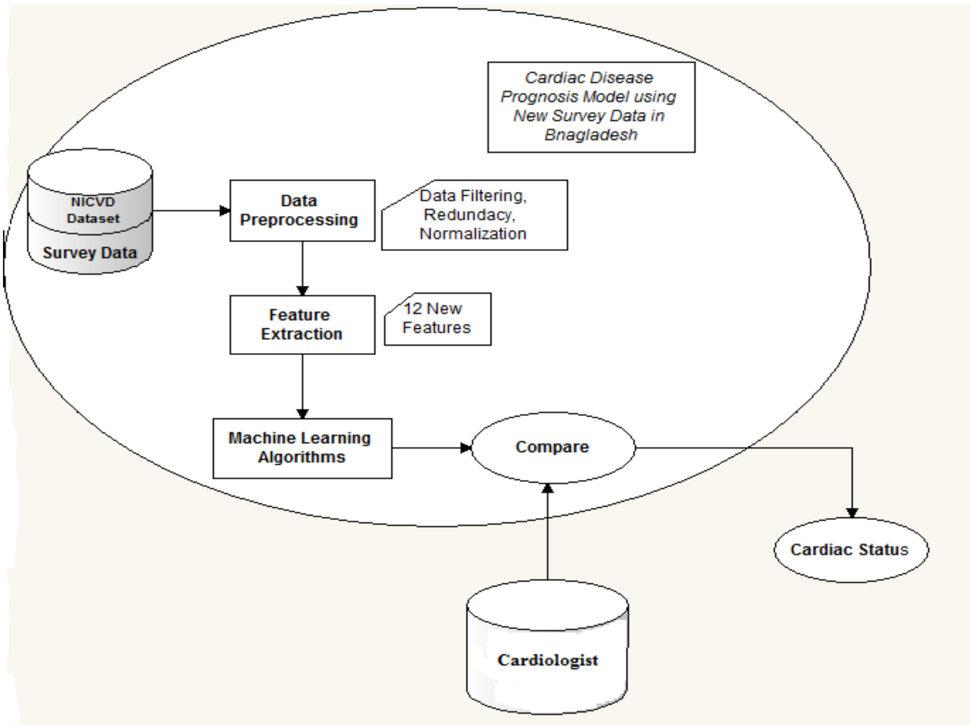


Figure 7(a): Overview /model of this research

We select 12 important attributes or clinical features, i.e., age, sex, chest pain type (CPT), blood pressure (RBP), cholesterol (CholesT), blood sugar (FBS), ECG results, heart rate (HR), exercise induced angina (EIangina), old peak, slope, and stenosis. We then choose six (06) well known supervised machine learning algorithms namely, Decision Tree (J48), Naïve Bayes, Random Forest, SVM, ANN, and Logisticfor classifying and predicting heart disease based on these clinical features vector (see Figure 7(b)). The results are then observed and analyzed carefully that shows that ANN outperforms other classifiers.

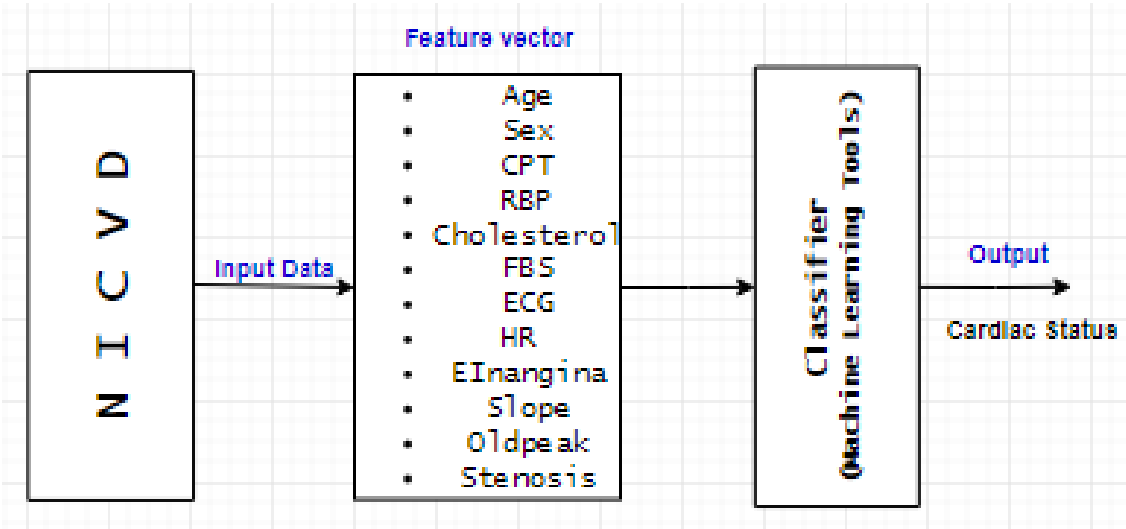


Figure 7(b): Overview /model of this research

3.2 Dataset

Data is collected from National Institute of Cardiovascular Disease(NICVD), Bangladesh through survey questionnaires. We have collected40 patient's data with 12 important clinicalfeatures or attributes out of 200 stable angina patients. Some amount of heart disease clinical data collected from NICVD, Bangladesh shown in Table 7.

Table 7: Heart disease data collected from NICVD, Bangladesh

Age	Sex	CPT	RBP	CholesT	FBS	ECG	HR	EIangina	Old Peak	Slope	Stenosis
50	0	3	140	220	0	0	155	0	1.4	2	0
59	1	1	130	222	0	0	110	1	0.5	3	3
47	1	4	120	231	1	2	127	1	1.8	3	1
60	1	1	110	212	1	2	160	1	0	1	2
31	1	1	140	198	2	2	127	0	1.4	3	0
50	1	4	190	170	1	0	136	0	2.6	3	1
63	1	3	130	204	0	0	140	1	0	2	0
62	1	4	110	226	0	2	127	1	1.6	3	1
43	0	3	110	153	0	0	156	1	1.8	3	0
55	0	4	140	232	1	0	139	0	1	3	1
64	1	3	130	330	0	0	155	0	0	1	0
63	0	4	145	220	0	2	117	0	0	2	0
39	1	3	130	245	0	0	182	0	3.2	3	0
54	1	4	135	223	1	2	161	1	3.1	3	1
57	1	3	130	250	1	2	137	1	0.8	2	1
49	1	2	120	230	0	0	152	1	1	3	2
59	0	4	140	320	0	2	162	0	0.4	1	1
61	1	4	140	330	0	2	115	1	0.6	3	3
58	1	4	120	221	1	0	158	1	1.6	1	1
44	0	2	110	202	0	0	171	0	1	2	0
73	1	2	150	277	1	0	169	0	0.5	1	0
58	1	4	145	288	0	0	118	1	0.8	3	3
51	1	1	120	159	0	2	145	0	0.6	1	0

3.3 Experimental decision

In this experiment we have observed that using NICVD: 40 patient’s data with 12 attributes, the classification accuracy of Artificial Neural Network (ANN) is almost 94% and it performs better than others classification algorithms such as Decision tree (82.50%), Naïve Bayes (85%), SVM (75%), Simple logistic (87.50%), Random forest (75%), and Bagging (62.50%). Whereas using UCI dataset with 12 attributes, the classification accuracy of Artificial Neural Network (ANN) is almost 92% and it performs better than others classification algorithms, such as Decision tree (76.90%), Naïve Bayes (86.50%), SVM (76.33%), Simple logistic (81.52%), Random forest (67.33%), and Bagging (55.12%).It is also observed that using UCI dataset the performance of 14 attributes and 12 attributes are almost equal.

Chapter 4

Results

4.1 Experimental results

In this study we used NICVD dataset with proposed 12 important features. Where we have validated patient's clinical data & features by cardiologist with coronary artery disease angiogram status (i.e., Cardiac disease: 1= positive, 0=negative).In the results of 40 patients, 17 patients instance has no heart disease and 23 patients or instances has heart disease. Also we have used UCI dataset with 12 important attributes and 14 (fourteen) attributes with coronary artery disease angiogram (i.e., Cardiac disease: 1= positive, 0=negative) results of 303 patients where 164 patients or instances has no heart disease and 139 patients or instances has heart disease. Regarding heart disease angiogram results we used binary classifier (i.e., Heart disease: 1= positive, 0=negative) where data is labeled. Since we have label data, we have used supervised learning (SL) method/algorithm(s) on the datasets. We have performed data preprocessing with data normalization operation on the datasetsfor finding better accuracy by classifier.

Thereafter, we applied different data mining or supervised machine learning algorithms in the environment of MATLAB and WEKA 3.8 set to compare between different classifiers and different data sets.Also we have applied10-fold cross validation (70% of the data is used for training and 30% is used for testing) on the datasets.The classifiers name with various properties of classifiers described in Table 8. We have four Investigations for ML classifiers:

i) Train (212) + Test(91) on UCI instances (14 features) : 10-fold cross validation
ii) Train (212) + Test(91) on UCI instances (12 features) : 10-fold cross validation
iii) Train (28) +Test(12) on NICVD instances (12 features): 10-fold cross validation
iv)Train (39) + Test(1) on NICVD instances (12 features): Leave one out validation

The experiment results (i.e., Various ML methods performance table, comparison graph) are displayed in Table 9, 10, 11, 12, 13& Figure 8, 9, and 10.

Table 8: Properties of various machine learning classifiers

SL No.	Classifiers Name	Properties of Classifiers
1	ANN	Input unit= 12, Hidden layer=2, Output layer=1, Activation Function: Sigmoid. Learning rate, η=0.01, Momentum coefficient, μ=0.9 Weight update, $\Delta W_{kj}= \eta * e_k(n) * x_j(n)$,where error, $e_k(n)$= desired output $(d_k) - (y_k)$ output signal of neuron K at time n.
2	Decision Tree	max_depth=32,min_samples_leaf=0.1, min_samples_split=0.1 max_features=12
3	Naive Bayes	Normal (Gaussian) distribution, Posterior, $\boldsymbol{P}(\boldsymbol{c}\|\boldsymbol{x}) = \frac{\text{Likelihood}\boldsymbol{P}(\boldsymbol{x}\|\boldsymbol{c}) * \text{Prior}\boldsymbol{P}(\boldsymbol{c})}{\text{Evidence}\boldsymbol{P}(\boldsymbol{x})}$;
4	Random Forest	Total no. of trees=10
5	SVM	Kernel function: Linear, c=1.0;$k(x,y)= x^T y + c$, where c is a trade off parameter between error and margin
6	Logistic Regression	Model: Linear with sigmoid function; logistic function,p(x)= $\frac{1}{1+e^{-(1+x_1+\cdots+x_{12})}}$

It is observed that using NICVD: 40 patient's data with 12 attributes, the classification accuracy of Artificial Neural Network (ANN) is almost 94% and it performs better than others classification algorithms such as Decision tree (82.50%), Naïve Bayes (85%), SVM (75%), Simple logistic (87.50%), Random forest (75%), and Bagging (62.50%). Whereas using UCI dataset with 12 attributes, the classification accuracy of Artificial Neural Network (ANN) is almost 92% and it performs better than others classification algorithms, such as Decision tree (76.90%), Naïve Bayes (86.50%), SVM (76.33%), logistic regression (81.52%), Random forest (67.33%)It is also observed that using UCI dataset the performance of 14 attributes and 12 attributes are almost equal.

$$\textbf{Classification Accuracy (CA)} = \frac{\text{Correct Classification}}{\text{Total Classification}} \times 100\%$$

$$\textbf{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\textbf{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Where,TP - True Positive, FP - False Positive, FN -False Negative.
Here, Positive means Heart Attack has been happened and Negative means No Heart Attack.

F1: This is a function of Precision and Recall. The F1 score is the harmonic mean of Precision and Recall.

Table 9: Performance analysis of various supervised methods using NICVD dataset

Techniques/Methods	CA	Precision	Recall	F1
Decision Tree (J48)	82.50%	85.20%	82.50%	82.60%
Naïve Bayes	85.00%	86.10%	85.00%	85.10%
Random Forest	75.00%	78.60%	75.00%	73.00%
SVM	75.00%	75.60%	75.00%	75.10%
ANN	**92.80%**	**92.91%**	**92.80%**	**92.79%**
Logistic	77.50%	78.60%	77.50%	77.60%

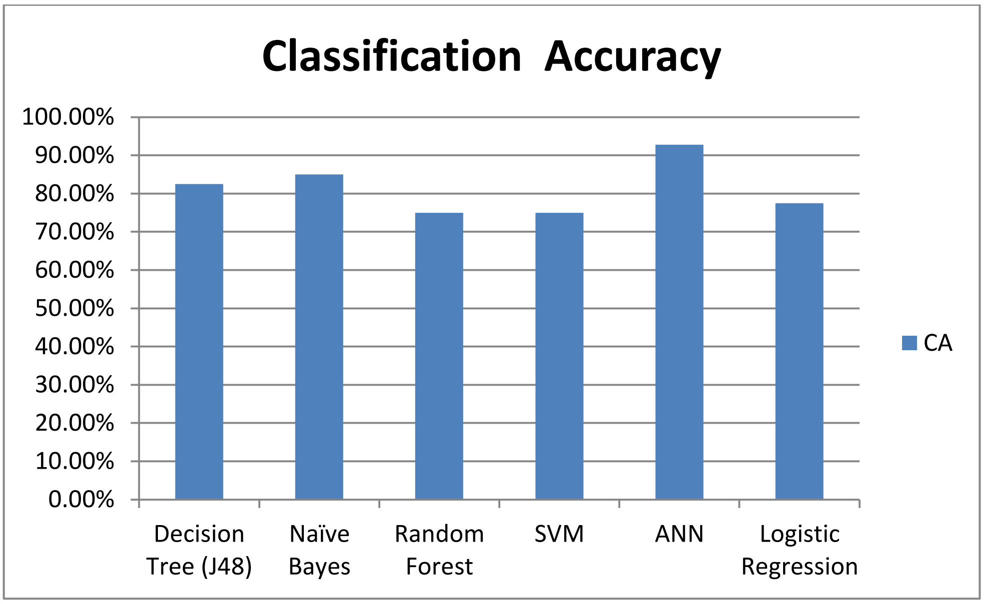


Figure 8: Classification accuracy graph of various ML algorithms using NICVD dataset

Table 10: Performance analysis of ML methods using UCI dataset with 14 attributes

Techniques/Methods (14 features)	**CA**	**Precision**	**Recall**	**F1**
Decision Tree (J48)	76.57%	76.8%	76.6%	76.6%
Naïve Bayes	86.51%	86.6%	82.65%	86.4%
Random Forest	69.64%	69.9%	69.6%	69.1%
SVM	76.89%	77.4%	76.9%	76.9%
ANN	**93.5%**	**93.8%**	**93.68%**	**93.56%**
Logistic	67.70%	67.6%	67.7%	67.6%

Table 11: Performance analysis of various ML methods using UCI dataset with 12 features

Techniques/Methods (12 features)	**CA**	**Precision**	**Recall**	**F1**
Decision Tree (J48)	76.90%	77.3%	76.9%	76.9%
Naïve Bayes	86.50%	86.3%	86.2%	86.1%
Random Forest	67.33%	67.3%	67.3%	67.0%
SVM	76.33%	76.3%	76.3%	76.3%
ANN	**91.7%**	**91.62%**	**91.67%**	**91.7%**
Logistic	67.01%	67.1%	67.01%	67.01%

Table 12: Comparison table of various ML methods performance between 14 features and 12 features using UCI dataset

Method Name	**CA (%)**		**Precision (%)**		**Recall (%)**	
	14 features	**12 features**	**14 features**	**12 features**	**14 features**	**12 features**
Decision Tree (J48)	76.7	76.9	76.8	77.3	76.6	76.9
Naïve Bayes	86.5	86.5	86.6	86.3	82.6	86.2
Random Forest	69.6	67.3	69.9	67.3	69.6	67.3
SVM	76.8	76.3	77.4	76.3	76.9	76.3
ANN	93.5	91.7	93.8	91.6	93.7	91.7
Logistic Regression	67.7	67.0	67.6	67.1	67.7	67.0

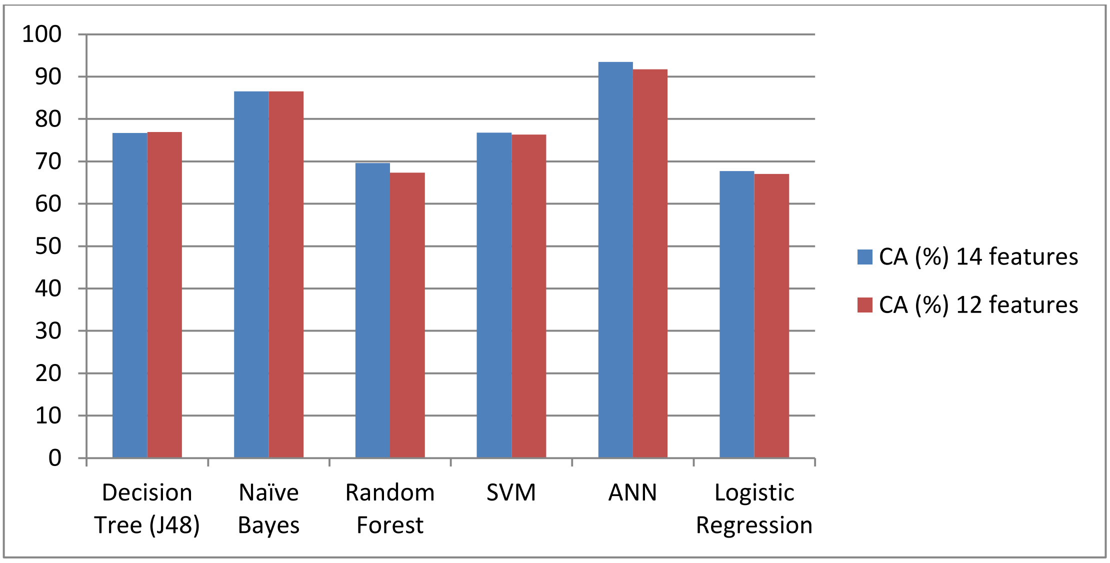


Figure 9: Comparison graph of classification accuracy between 14 and 12 features using UCI dataset.

Table 13: Comparison table of various machine learning methods performance between NICVD dataset and UCI dataset using 12 features

Methods Name	CA (%)		Precision (%)		Recall (%)	
	NICVD Dataset	UCI Dataset	NICVD Dataset	UCI Dataset	NICVD Dataset	UCI Dataset
Decision Tree (J48)	82.5	76.9	85.2	77.3	82.5	76.9
Naïve Bayes	85.0	86.5	86.1	86.3	85.0	86.2
Random Forest	75.0	67.3	78.6	67.3	75.0	67.3
SVM	75.0	76.3	75.6	76.3	75.0	76.3
ANN	92.8	91.7	92.9	91.6	92.8	91.7
Logistic Regression	77.5	67.0	78.6	67.1	77.5	67.0

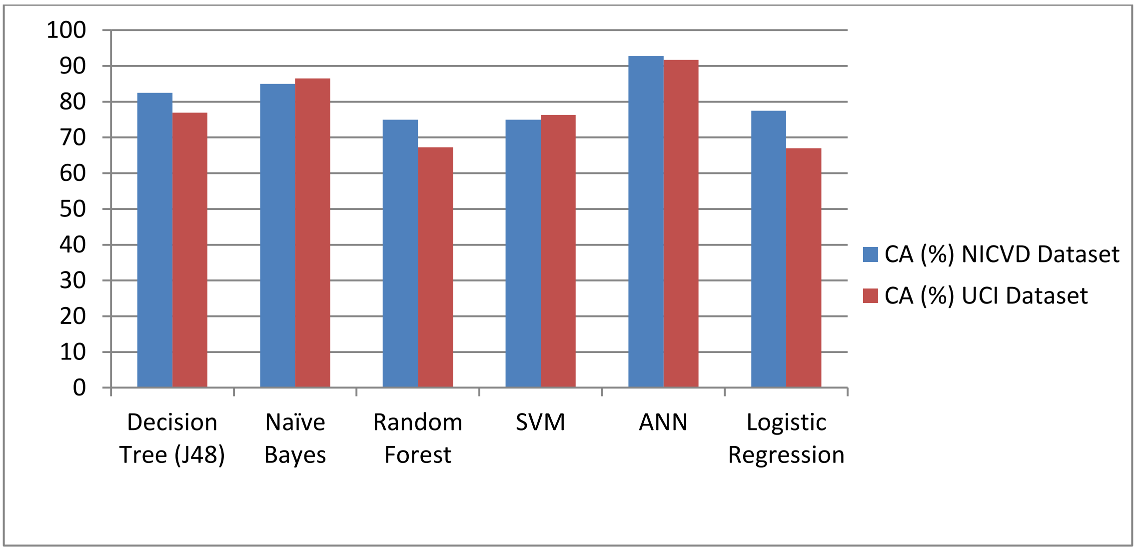


Figure 10: Comparison graph of classification accuracy betweenNICVD and UCIdataset

Table 14: CA of ML methods using leave-one out method for NICVD dataset

Techniques/Methods	CA
Decision Tree (J48)	71.00%
Naïve Bayes	75.10%
Random Forest	71.42%
SVM	75.00%
ANN	**84.80%**
Logistic	75.33%

Chapter 5

Discussion & Conclusion

5.1 Summary

In this work we have collected data from NICVD, Bangladesh and developed a cardiac disease prediction model to assist medical practitioners in predicting heart disease status based on the patients' clinical data. We have selected twelve (12) vital clinical features of coronary artery disease and developed a prediction model using six (06) supervised machine learning algorithms. It is observed that ANN is the best predictor with 92.8% accuracy.

5.2 Discussion

In this research we have applied different machine learning algorithms for 12 features. Previous most of the authors or researchers have implemented various machine learning algorithms using UCI dataset for 14 features (with angiographic status).We have four investigations for ML classifiers. One is NICVD: 40 patient's data with 12 attributes, the classification accuracy of ANN is (92.80%), Decision Tree (82.50%), Naïve Bayes (85%), SVM (75%), Logistic Regression (77.50%), and Random forest (75%). This dataset contains small amount of data with numeric value of 12 features. For the noisy and numeric data ANN gives good result. When the data will be more, the accuracy result may be change. For the small amount of training data in the classifiers, the data can be over fitted. To reduce over fitting problem we have implemented 10-fold cross validation on the dataset. Also we have applied leave one out cross validation on the NICVD dataset. The accuracy of ANN is (84.80%), Decision Tree (71%), Naïve Bayes (75.10%), SVM (75%), Logistic Regression (75.33%), and Random forest (71.42%) by leave one method. Where we have found the results of classifiers is nearby to more accurate.

Third one is UCI dataset with 12 attributes, the classification accuracy of Artificial Neural Network (ANN) is (91.7%) and it performs better than others classification algorithms, such as Decision tree (76.90%), Naïve Bayes (86.50%), SVM (76.33%), Logistic regression (67%), and Random forest (67.33%). Fourth one is UCI dataset with 14features, the classification accuracy of ANN is (93.5%), Decision tree (76.70%), Naïve Bayes (86.50%), SVM (76.80%), Logistic regression (67.70%), and Random Forest (69.6%).It is also observed that using UCI dataset the performance of 14 attributes and 12 attributes are almost equal. For the large amount of data result will be change.

5.3 Limitation

In this research we have found some limitation and have not considered some things are:

- We have collected small scale corpus for NICVD (only 40).
- We have not collected more features for NICVD.
- We have not considered smoking and genetic (family history of CAD) features.
- We have not considered Troponin I test for detect heart disease.

5.4 Future work

In future we would like to concentrate on three things:

- Prepare a repository database based on Bangladeshi patients.
- Prepare an IoT enabled system to collect data from machines (see Figure 11).
- Developed a mobile based cardiac status application through internet.

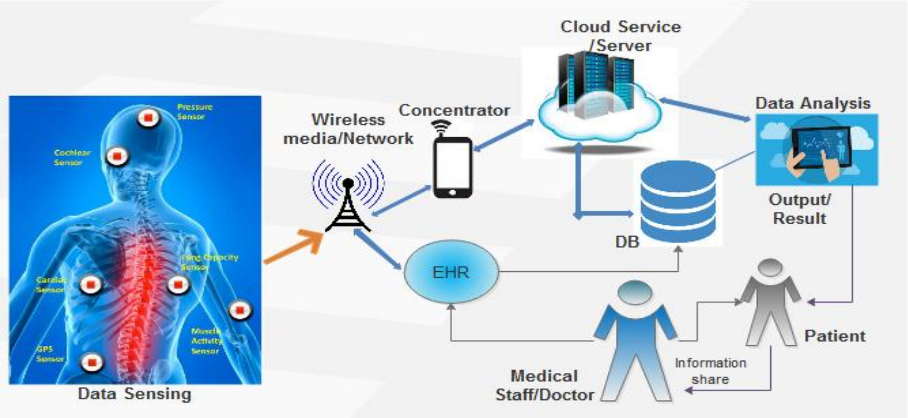


Figure 13: Model of IoT enabled system

In this system, wearable sensors are used to measure various physiological parameters e.g. blood pressure, body temperature, ECG, ETT, Cholesterol level, etc. After sensing data, the sensors transmit the gathered information to a gateway server through a wireless connection/communication.

5.5 Conclusion

On the whole aim of our work is to predict heart disease status more precisely and build a dataset of our own which will be first coronary heart disease data repository in Bangladesh. Therefore, we have collected heart disease patient’s data through our own constructed survey questionnaire. We also finalize minimum number of features dimensionality for our system to reduce angiogram hassles. However, we excluded smoking feature as it does not have much impact behind occurring heart disease in our country. Six data mining classification techniques such as Naïve Bayes, Decision Tree, Logistic Regression, Random Forest, Support Vector Machine and simple Neural Network have been applied. It has been seen that Neural Network works better than other classifiers with 92.80% accuracy.

References

[1] Wong, N. D., “Epidemiological studies of CHD and the evolution preventive cardiology,”Nature Reviews cardiology,Vol. 11, pp. 276–289,California, Irvine, CA 92697, USA, 2014.

[2] Wu R, Peters W, Morgan MW, “The next generation clinical decision support: linking evidence to best practice,” Journal of Healthcare Information Management, Vol. 16, No. 4 , USA, 2002.

[3] Anbarasi M, Anupriya E, and Iyengar NCHSN, “Enhanced prediction of heart disease with feature subset selection using genetic algorithm,” International Journal of Engineering Science and Technology, Vol. 2,No.10, 2:5370-76, 2010.

[4] Palaniappan S, Awang R., “Intelligent heart disease prediction system using data mining techniques,” International Journal of Computer Science and Network Security, Vol.8, No.8,8:343-50,August 2008.

[5] Cord, Matthieu, Cunningham, Padraig (Eds.), “Machine Learning Techniques for Multimedia Case Studies on Organization and Retrieval,” Springer,XVI, 289 p, hardcover, 2008.

[6] E.P.Ephzibah, Dr. V. Sundarapandian, “Framing Fuzzy Rules using Support Sets for Effective Heart Disease Diagnosis,” International Journal of Fuzzy Logic Systems (IJFLS), Vol.2, No.1, February 2012.

[7] Moloud Abdar, Sharareh R. Niakan Kalhori, Tole Sutikno, Imam Much Ibnu Subroto, Goli Arji, “Comparing Performance of Data Mining Algorithms in Prediction Heart Diseases,” IJECE, Vol. 5, No. 6, pp. 1569 – 576,ISSN: 2088-8708, December 2015.

[8] Asha Rajkumar, G. Sophia Reena, “Diagnosis of Heart Disease Using Data mining Algorithm,” Global Journal of Computer Science and Technology, Page | 38, Vol. 10 No. 10, Ver. 1.0 ,September 2010.

[9] M. Anbarasi, E. Anupriya, and N.Ch.S.N.Iyengar, “Enhanced Prediction of Heart Disease with Feature Subset Selection using Genetic Algorithm,” International Journal of Engineering Science and Technology, Vol. 2, No. 10, 2010.

[10] Sellappan Palaniappan, Rafiah Awang, “Intelligent Heart Disease Prediction System Using Data Mining Techniques,” International Journal of Computer Science and Network Security (IJCSNS), Vol. 8, No. 8, August 2008.

[11] k.Srinivas, Dr.G.Raghavendra Rao, and Dr. A.Govardhan, “Analysis of coronary heart disease and prediction of heart attack in coal mining regions using data mining techniques,” International Conference on Computer Science and Education (ICCSE), pp. 1344 - 1349, 2010.

[12] Yanwei Xing, “Combination Data Mining Methods with New Medical Data to Predicting Outcome of Coronary Heart Disease,” IEEE Transactions on Convergence Information Technology, pp. 868 – 872, November 2007.

[13] Vikas Chaurasia and Saurabh Pal, “Early Prediction of Heart Diseases using Data Mining Techniques,” Caribbean Journal of Science and Technology, ISSN 0799-3757, Vol. 1, pp. 208-21,2013.

[14] Hnin Wint Khaing, “Data Mining based Fragmentation and Prediction of Medical Data”, International Computer Research and Dvelopment (ICCRD),2011.

[15] Max Bramer, “Principles of Data Mining,” 2nd ed. Springer, ISSN1863-7310, 2013.

[16] V. Chauraisa and S. Pal, “Data Mining Approach to Detect Heart Diseases,” International Journal of Advanced Computer Science and Information Technology (IJACSIT),Vol. 2, No. 4, pp 56-66, 2013.

[17] Atul Kumar Pandey,Prabhat Pandey,K.L. Jaiswal, and Ashish Kumar Sen, “A Heart Disease Prediction Model using Decision Tree,”Journal of Computer Engineering (IOSR-JCE), e-ISSN: 2278-0661, p- ISSN 2278-8727,Vol. 12, No. 6, pp. 83-86, August 2013.

[18] Thendral Puyalnithi and V. Madhu Viswanatham, “Preliminary Cardiac Disease Risk Prediction Based on Medical and Behavioural Data Set Using Supervised Machine Learning Techniques,”Indian Journal of Science and Technology, Vol. 9, No. 31,August 2016.

[19] Shai Shalev-Shwartz, Shai Ben-David, “Understanding Machine Learning,” Cambridge University Press, New York,NY 10013-2473, 2014.

[20] Introduction to Support Vector Machines [online]. Open-CV Documentation; 7 March 2017.

URL:http://docs.opencv.org/2.4/doc/tutorials/ml/introduction_to_svm/introduction_to_svm.html

[21] Ethan Alapydin,“Introduction to Machine Learning,” 2nd ed. Cambridge Massachusetts, MIT Press, London, 2010.

[22] N. Q. Hung, M. S. Babel, S. Weesakul, and N. K. Tripathi, 2008: "An artificial neural network model for rainfall forecasting" in Bangkok, Thailand, 2008.

[23] Randa El-Bialy , Mostafa A. Salamay, Omar H. Karam, and M.Essam Khalifa, "Feature Analysis of Coronary Artery Heart Disease Data Sets," International Conference on Communication, Management and Information Technology (ICCMIT), 2015.

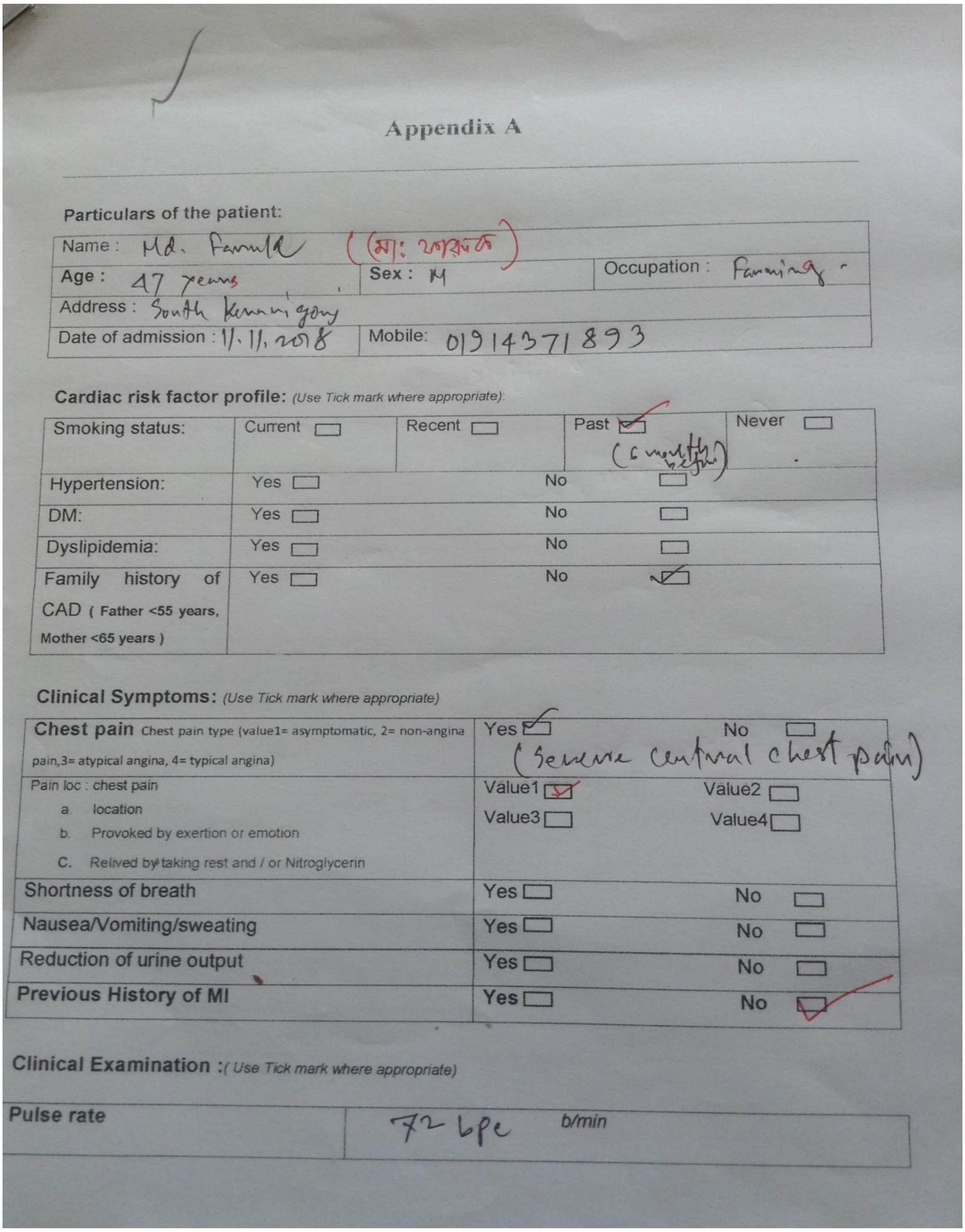
[24] AKM Monwarul Islam, AKM Mohibullah, and Timir Paul "Cardiovascular Disease in Bangladesh: A Review," Bangladesh Heart Journal, Vol. 31, No. 2,July 2016.

[25] Mai Shouman, Tim Turner, and Rob Stocker "Applying k-Nearest Neighbor in Diagnosing Heart Disease Patients," International Conference on Knowledge Discover (ICKD), Singapore, 2012.

[26] Markos G. Tsipouras, Dimitrios I. Fotiadis, Katerina K. Naka, and Lampros K. Michalis "Automated Diagnosis of Coronary Artery Disease Based on Data Mining and Fuzzy Modeling,"

Appendix A

Data Collection Form



Appendix A

Particulars of the patient:

Name : Md. Farukh ((মা: [illegible]))	
Age : 47 years	Sex : M / Occupation : Farming
Address : South Keranigonj	
Date of admission : 1.11.2018	Mobile: 01914371893

Cardiac risk factor profile: *(Use Tick mark where appropriate)*:

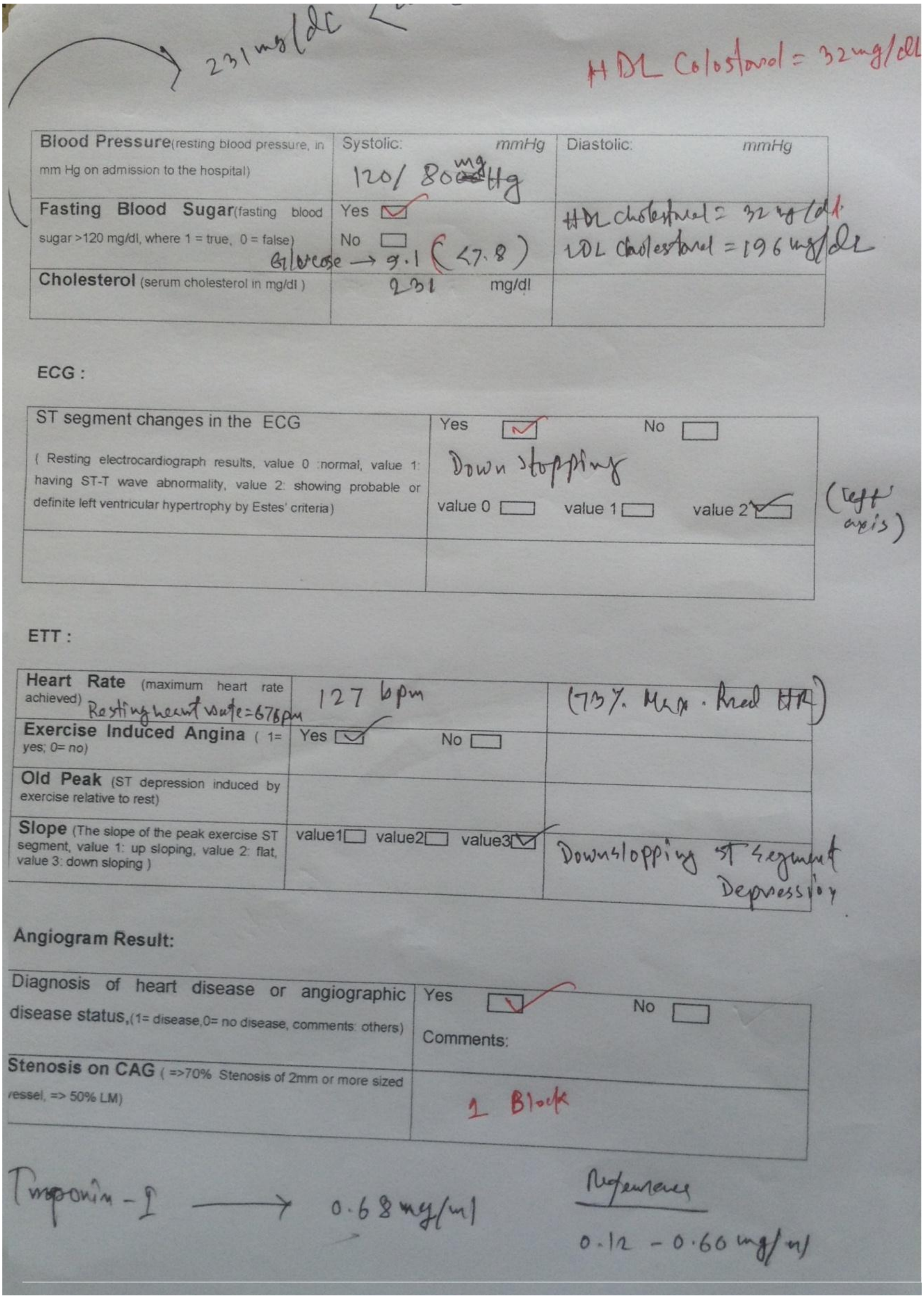
Smoking status:	Current ☐	Recent ☐	Past ☑ (6 month before)	Never ☐
Hypertension:	Yes ☐		No ☐	
DM:	Yes ☐		No ☐	
Dyslipidemia:	Yes ☐		No ☐	
Family history of CAD (Father <55 years, Mother <65 years)	Yes ☐		No ☑	

Clinical Symptoms: *(Use Tick mark where appropriate)*

Chest pain Chest pain type (value1= asymptomatic, 2= non-angina pain,3= atypical angina, 4= typical angina)	Yes ☑ (Severe central chest pain)	No ☐
Pain loc : chest pain a. location b. Provoked by exertion or emotion C. Relived by taking rest and / or Nitroglycerin	Value1 ☑ Value3 ☐	Value2 ☐ Value4 ☐
Shortness of breath	Yes ☐	No ☐
Nausea/Vomiting/sweating	Yes ☐	No ☐
Reduction of urine output	Yes ☐	No ☐
Previous History of MI	Yes ☐	No ☑

Clinical Examination :*(Use Tick mark where appropriate)*

Pulse rate	72 bpe b/min



231 mg/dL

HDL Colostorol = 32 mg/dL

Blood Pressure (resting blood pressure, in mm Hg on admission to the hospital)	Systolic: mmHg 120/ 80 mmHg	Diastolic: mmHg
Fasting Blood Sugar (fasting blood sugar >120 mg/dl, where 1 = true, 0 = false)	Yes ☑ No ☐ Glucose → 9.1 (<7.8)	HDL cholesterol = 32 mg/dl LDL Cholesterol = 196 mg/dL
Cholesterol (serum cholesterol in mg/dl)	231 mg/dl	

ECG :

ST segment changes in the ECG (Resting electrocardiograph results, value 0 :normal, value 1: having ST-T wave abnormality, value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria)	Yes ☑ No ☐ Down sloping value 0 ☐ value 1 ☐ value 2 ☑

(left axis)

ETT :

Heart Rate (maximum heart rate achieved) Resting heart rate = 67 bpm	127 bpm	(73% Max. Pred HR)
Exercise Induced Angina (1= yes; 0= no)	Yes ☑ No ☐	
Old Peak (ST depression induced by exercise relative to rest)		
Slope (The slope of the peak exercise ST segment, value 1: up sloping, value 2: flat, value 3: down sloping)	value1 ☐ value2 ☐ value3 ☑	Downslopping ST segment Depression

Angiogram Result:

Diagnosis of heart disease or angiographic disease status,(1= disease,0= no disease, comments: others)	Yes ☑ No ☐ Comments:
Stenosis on CAG (=>70% Stenosis of 2mm or more sized vessel, => 50% LM)	1 Block

Troponin - I → 0.68 ng/ml

Reference

0.12 - 0.60 ng/ml

Application for data collection from NICVD, Bangladesh:

ULAB
UNIVERSITY OF LIBERAL ARTS
BANGLADESH

REF: ULAB/CSE/FAC/02

Date: July 01, 2018

To
The Director
National Institute of Cardiovascular Diseases (NICVD)
Sher-e-Bangla Nagar
Dhaka-1207, Bangladesh

Ref: Data collection for Evaluation Probability of Coronary Artery Disease in Patient with Stable Angina Evaluated by Coronary Angiogram by PRE-CAG Risk Model.

Esteemed Sir,

With due respect and honour I, Dr. A.K.M. Muzahidul Islam, Associate Professor and Head of CSE department, ULAB, Dhanmondi, Dhaka would like to mention that currently I and Prof. Dr. Mohammad Nurul Huda of UIU are jointly supervising a Masters student Mr. Shah Jalal who is pursuing research on Coronary Artery Disease.

We would like to conduct part of the research work at NICVD under supervision of the chief-supervisor Prof. Dr. Md. Afzalur Rahman sir and the supervisor Assoc. Prof. Dr. Abdul Momen sir, of NICVD.

In this research, we need to obtain some research data which may be collected from your institute. We have studied this matter jointly with respected Assoc. Dr. Momen and have prepared a survey questionnairy form which is attached with this email.

Thus you would be highly appreciated if you kindly could allow us to conduct a survey at your institute from July 01 to July 31, 2018.

Sincerely,

Dr. A.K.M. Muzahidul Islam
Associate Professor and Head
Department of Computer Science and Engineering (CSE)
University of Liberal Arts Bangladesh (ULAB)
Email : muzahidul.islam@ulab.edu.bd

Authorization letter from NICVD, Bangladesh:

গণপ্রজাতন্ত্রী বাংলাদেশ সরকার
পরিচালক ও অধ্যাপকের কার্য্যালয়
জাতীয় হৃদরোগ ইনস্টিটিউট ও হাসপাতাল
শেরে বাংলা নগর, ঢাকা- ১২০৭।

স্মারক নং- এনআইসিভিডি/একাডেমিক/ষ্ট্যাডি/২০১৮/

তারিখঃ- .০৯.২০১৮ ইং।

অফিস আদেশ

অত্র প্রতিষ্ঠানের স্মারক নং-এনআইসিভিডি/একাডেমিক/নোটিশ-২০১৪/১৫৫৯ তারিখঃ ০২.০৪.২০১৮ মোতাবেক ক্লিনিক্যাল ও প্রশাসনিক বিষয়ে তথ্য সংগ্রহের নিমিত্তে আগ্রহী ছাত্র/ছাত্রীদের আবেদন যাচাই বাছাই এর জন্য গঠিত কমিটির সভা অনুষ্ঠিত হয়।

উক্ত সভায় নিম্নলিখিত ০২ জন ছাত্রকে অত্র প্রতিষ্ঠানের কমিটি কর্তৃক টাইটেল অনুযায়ী তথ্য সংগ্রহের সুপারিশের আলোকে নিম্নোক্ত তত্ত্বাবধায়কের অধিনে থেকে প্রার্থীদের থিসিস/ডিজারটেশন/ষ্ট্যাডি অনুমতি প্রদাণ করা হলো।

ক্রমিক নং	প্রার্থীর নাম	টাইটেল	তত্ত্বাবধায়ক
1.	Dr. Shahriar Faruque MD part-III Department of Psychiatry BSMMU Reg.5113	**Premature ejaculation among post coronary artery stenting in cardiology OPD in tertiary care hospital.**	Dr. Pijous Biswas
2.	Md. Shah Jalal (Jamil) MSCSE Program, United International University (UIU)	**Evaluation Probalility of Coronary Artery Disease in Patient with Stable Angina Evaluated by Coronary Angiogram by PRE-CAG Risk Model.**	Dr. Abdul Momen

শর্তাবলী ঃ

১। হাসপাতালের নিয়ম-কানুন যথাযথভাবে মেনে চলা ;
২। তথ্য সংগ্রহের ক্ষেত্রে তত্ত্বাবধায়ককে অবহিত করন ;
৩। তথ্য সংগ্রহ শেষে সভাপতি স্ট্যাডি কমিটির নিকট অগ্রায়নসহ চূড়ান্ত প্রতিবেদন দাখিল করতে হবে।

(অধ্যাপক ডাঃ মোঃ আফজালুর রহমান)
পরিচালক ও অধ্যাপক

স্মারক নং- এনআইসিভিডি/একাডেমিক/ষ্ট্যাডি/২০১৮/ ৪৭০৭

তারিখঃ- ১৮.০৯.২০১৮ ইং।

অনুলিপি সদয় অবগতি ও প্রয়োজনীয় ব্যবস্থা গ্রহণের জন্য প্রেরিত হলোঃ

১। অধ্যাপক/সহযোগী অধ্যাপক..।
২। ডাঃ আব্দুল মোমেন, সহযোগী অধ্যাপক, কার্ডিওলজী,এনআইসিভিডি, ঢাকা।
৩। ডাঃ পীযুষ বিশ্বাস,আবাসিক চিকিৎসক, এনআইসিভিডি, ঢাকা।
৪। ডাঃ..,এনআইসিভিডি, ঢাকা।
৫। মোঃ শাহ জালাল, এম এস সি এস ই, ইউ আই ইউ, ঢাকা।
৬। মেট্রন, এনআইসিভিডি, ঢাকা।
৭। পরিচালক ও অধ্যাপক মহোদয়ের ব্যক্তিগত সহকারী, এনআইসিভিডি, ঢাকা।
৮। .... ওয়ার্ড সিষ্টার ইনচার্জ, এনআইসিভিডি, ঢাকা।
৯। অফিস নথি।

১৮/৯/১৮
পরিচালক ও অধ্যাপক

Clinical patient’s ETT Report:

পপুলার ডায়াগনস্টিক সেন্টার লিঃ

POPULAR DIAGNOSTIC CENTRE LTD.

2 No. English Road, Dhaka-1100, Bangladesh. Tel: 9574917-21, Mob: 01938-821943, 01954-249757, 01764-486997, 01764-486838, Fax: 88 02 7115872

Head Office : House # 16, Road # 02, Dhanmondi, Dhaka. Tel : 9669480, Fax : 9666804, E-mail : popular@popularbd.com

POPULAR

EXERCISE TREADMILL REPORT

ID No : **E134315**

Patient's Name : Md. Faruque

Refd. By : Dr. Md. Jashim Uddin, MBBS,FPGCS,D-Card

Receive Date: **18/10/2018** Delivery Date: **18/10/2018**

Age: **47 Year(s)** Sex: **Male**

CLINICAL INFORMATION

- Height : 160 cm
- Weight : 60kg
- Resting ECG : within normal limits.

PURPOSE : Screening of IHD.

Test Summary:

Phase	Stage	Time (Min)	Speed (MPH)	Elevation (% Grade)	Heart rate (bpm)	BP (mmHg)
PRE- EX	Standing	01.04	0.00	00	69	120/80
Exercise (Bruce)	Stage 1	02.31	1.70	10	123	140/80
	Stage 2					
	Stage 3					
	Stage 4					
Recovery	2 min	02.00	0.00	00	76	150/70
	6 min	02.00	0.00	00	105	120/80

- Reason for Termination : **Patient requests stopping because of severe central chest pain.**
- Exercise Time : **2:31 min**
- Max. Speed : **1.70 MPH**
- Max Grade. : **10 %**
- Estimated Max. Workload : **4.60 METs**
- Peak HR achieved : **127 bpm (73 % Max. Pred HR)**
- Significant ST segment changes : **Down slopping ST segment depression in recovery** phase **Present in lead V_1-V_6.**

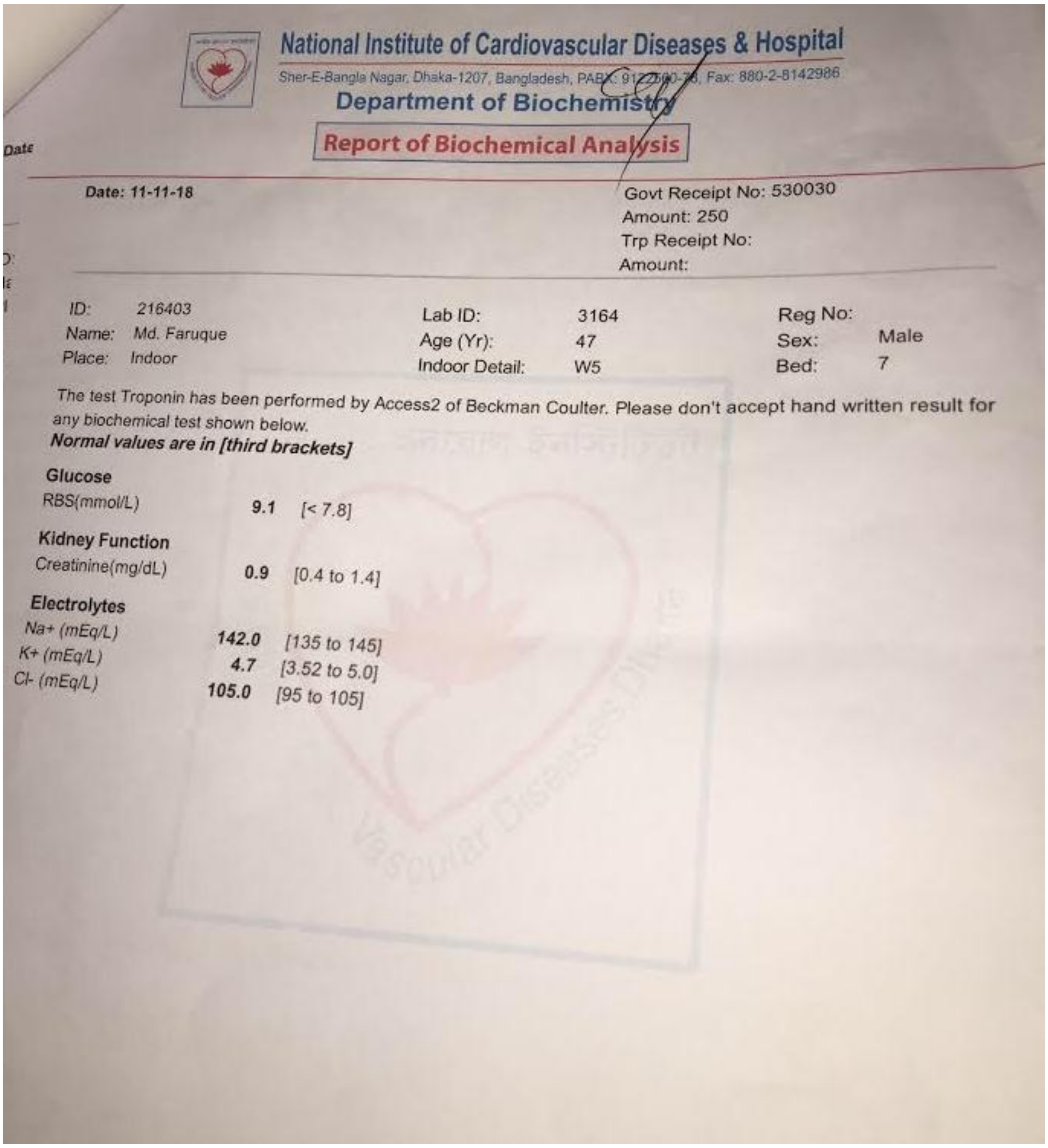
Impressions :

- **Poor exercise capacity.**
- **Target heart rate is not achieved.**
- Normal haemodynamic response.
- No arrhythmia observed during the procedure.
- NYHA functional class – I.
- ETT is **Strongly Positive** for ECG evidence of provo~~cable~~ myocardial ischaemia.

18.10.18

DR. MD. MUSADDEQUL ALAM (DOLON)
MBBS.(DHAKA) D-CARD (BSMMU)
FCPS (CADIOLOGY- thesis)
ASSTANT PROFESSOR CARDIOLOGY DEPT
DHAKA NATIOLAL MEDICAL COLLEGE

Patients’Biomedical Report:



National Institute of Cardiovascular Diseases & Hospital

Sher-E-Bangla Nagar, Dhaka-1207, Bangladesh, PABX: 912[illegible], Fax: 880-2-8142986

Department of Biochemistry

Report of Biochemical Analysis

Date: 11-11-18

Govt Receipt No: 530030

Amount: 250

Trp Receipt No:

Amount:

ID:	216403	Lab ID:	3164	Reg No:	
Name:	Md. Faruque	Age (Yr):	47	Sex:	Male
Place:	Indoor	Indoor Detail:	W5	Bed:	7

The test Troponin has been performed by Access2 of Beckman Coulter. Please don't accept hand written result for any biochemical test shown below.

Normal values are in [third brackets]

Glucose		
RBS(mmol/L)	**9.1**	[< 7.8]
Kidney Function		
Creatinine(mg/dL)	**0.9**	[0.4 to 1.4]
Electrolytes		
Na+ (mEq/L)	***142.0***	*[135 to 145]*
K+ (mEq/L)	***4.7***	*[3.52 to 5.0]*
Cl- (mEq/L)	***105.0***	*[95 to 105]*

Patients’ Biomedical Report:

NEW DHAKA MODERN CLINIC
(Hospital & Computerized Diagnostic Complex)

ID	: N26301	Received Date: 9/6/2018	Delivered Date: 9/6/2018
Patient Name	: MR. FARUK		
Referred By	: PROF. SYED AZIZUL HAQUE. MBBS.FCPS.MD.FACC. FRCP		Age : 47 Year(s) Sex : Male
Specimen	: Blood		

COMPUTERIZED BIOCHEMICAL REPORT

Estimations are Carried out by automated Chemistry Analyzer. Cobas-c 311

Test	Result	Reference Value
Plasma Glucose Random	7.25 mmol/L	Adult: 4.1-7.8 mmol/L >=11.1 mmol/L (Diabetic) Neonate: 1.67-4.42 mmol/L
Corr. Urine Sugar	Nil	
S. ALT (SGPT)	58 U/L	Upto : 40 U/L
Cholesterol (Total)	231 mg/dL	<200 mg/dL
S. HDL Cholesterol	32 mg/dL	>40 mg/dL
S. LDL Cholesterol	196 mg/dL	<130 mg/dL
S. Triglyceride	184 mg/dL	<150 mg/dL
T. Cholesterol/HDL Ratio	7.19	Male: <5.1 Female: <4.4
LDL/HDL Ratio	6.13	Male: <3.3 Female: <2.9

Checked by 1 Checked by 2 Printed by - Belal

Sultan M. Atikur Rahman
B.Sc (Hon's), M.Sc (R.U.)
Dept. Biochemistry & Molecular Biology
Biochemist.
New Dhaka Modern Clinic.

Clubbing ...

পপুলার ডায়াগনস্টিক সেন্টার লিঃ

POPULAR DIAGNOSTIC CENTRE LTD.

2 No. English Road, Dhaka-1100, Bangladesh. Tel: 9574917-21, Mob: 01930-021943, 01954-249757, 01764-406997, 01764-406038, Fax: 88 02 7115072
Head Office : House # 16, Road # 02, Dhanmondi, Dhaka. Tel : 9669480, Fax : 9666804, E-mail : popular@popularbd.com

POPULAR

EXERCISE TREADMILL REPORT

Receive Date: **18/10/2018** Delivery Date: **18/10/2018**
Age: **47 Year(s)** Sex: **Male**

ID No. : **E134315**
Patient's Name : Md. Faruque
Refd. By : Dr. Md. Jashim Uddin, MBBS.FPGCS.D-Card

CLINICAL INFORMATION: * Height : 160 cm * Weight : 60kg
* Resting ECG : within normal limits.

PURPOSE : Screening of IHD.

Test Summary:

Phase	Stage	Time (Min)	Speed (MPH)	Elevation (% Grade)	Heart rate (bpm)	BP (mmHg)
PRE- EX	Standing	01.04	0.00	00	69	120/80
Exercise (Bruce)	Stage 1	02.31	1.70	10	123	140/80
	Stage 2					
	Stage 3					
	Stage 4					
Recovery	2 min	02.00	0.00	00	76	150/70
	6 min	02.00	0.00	00	105	120/80

- Reason for Termination : **Patient requests stopping because of severe central chest pain.**
- Exercise Time : **2:31 min**
- Max, Speed : **1.70 MPH**
- Max Grade. : **10 %**
- Estimated Max, Workload : **4.60 METs**
- Peak HR achieved : **127 bpm (73 % Max. Pred HR)**
- Significant ST segment changes : **Down slopping ST segment depression in recovery phase Present in lead V_1-V_6.**

Impressions :

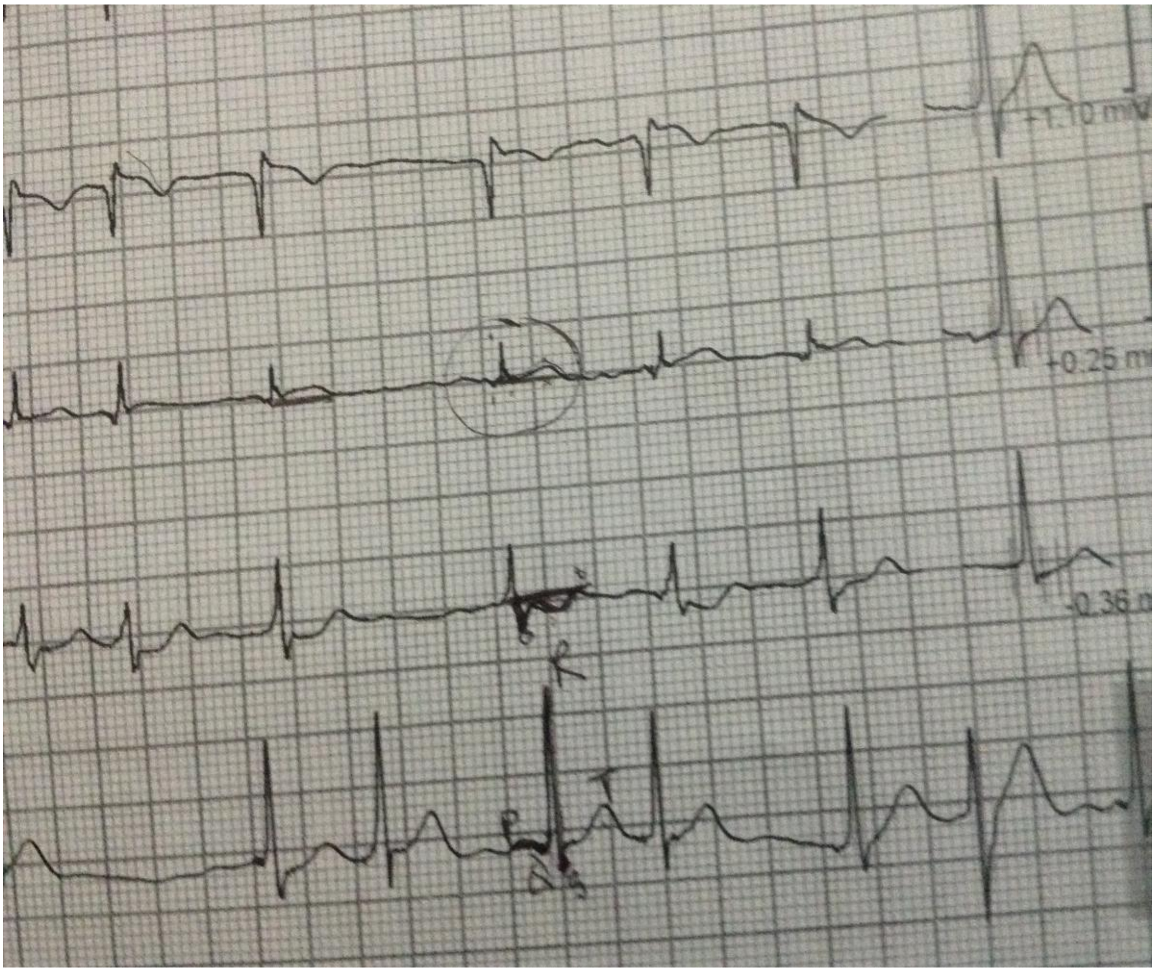
- **Poor exercise capacity.**
- **Target heart rate is not achieved.**
- Normal haemodynamic response.
- No arrhythmia observed during the procedure.
- NYHA functional class – I.
- ETT is **Strongly Positive** for ECG evidence of provo~~cable~~ myocardial ischaemia.

18.10.18

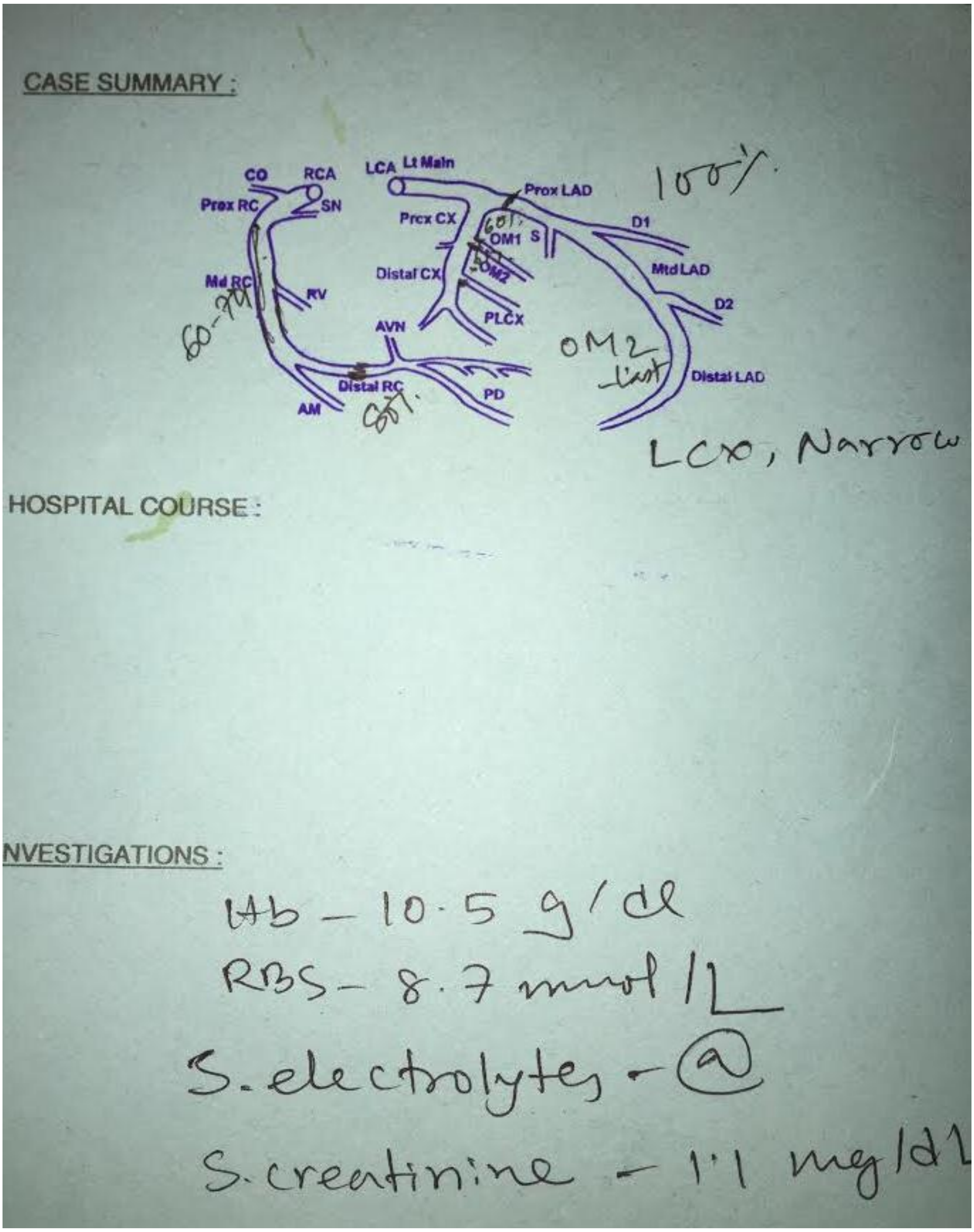
DR. MD. MUSADDEQUL ALAM (DOLON)
MBBS.(DHAKA) D-CARD (BSMMU)
FCPS (CADIOLOGY- thesis)
ASSTANT PROFESSOR CARDIOLOGY DEPT
DHAKA NATIOLAL MEDICAL COLLEGE

Prepared by________

Ptients' ETT report:



Patients' Angiogram Report:



National Institute of Cardiovascular Diseases

Date: 12-11-2018 **Department of Cardiology** Cath Lab: 0

Sr. N0	Name	Age	Ward/Bed	Diagnosis	Procedure	Category	Comment
1	Kiamuddin → 530782	60	05/14	AMI(ANT)	CAG+		
2	Shahid Al Mamun	55	C-2	UA+DM	CAG+(RADIL)		
3	Abul Kalam Azad PCI	59	05/11	CSA(ETT+VE)+HTN	PCI TO LAD+RCA		
4	Setab Ali →	41	05/28	AMI(EXT-ANT)	CAG+		
~~5~~	~~Joynal Abedin~~ PCI	50	05/40	AMI(A/S)POST PCI TO LAD	PCI TO RCA/LCX		
6	Borhan Uddin	56	05/48	AMI(INF) SK+	CAG+		
7	ATM Nasir Uddin PCI	63	C-09	CSAPOST PCI 2012	PCI TO LM+ LCX		
8	Md. Sirajul Haque	57	C-19	AMI(INF)	CAG+		
	Ferdoushi Hossain →	50	08/17	CSA+ ETT+ve	CAG+	(R) D-II	
10	Ali Chowdhury	62	C-20	CSA(ETT+VE)DM	CAG+		
11	Obaidul Haque →	45	05/04	AMI(INF)RVI+HTN	CAG+		
12 →	Md. Mahmud Sheekh ? 530955	50	05/*16	UA+HTN+DM	CAG+		
13	A.Rob PCI	70	05/34	UaOMI(INF)	PCI TO LAD+LCX	D-III *2	
14	Md.Faruk	47	05/7	CSA + ETT+ve	CAG+	D-III	Dr Jasim