NOVEL IMPLEMENTATION OF CARDIO-VASCULAR(CVD) USING MACHINE LEARNING TECHNIQUES

A PROJECT REPORT

Submitted by

R. RAMKUMAR (421118104070)

in partial fulfillment for the award of the degree
of

BACHELOR OF ENGINEERING

In

COMPUTER SCIENCE AND ENGINEERING

IFET COLLEGE OF ENGINEERING, VILLUPURAM – 605 108
(AN AUTONOMOUS INSTITUTION)

ANNA UNIVERSITY: CHENNAI 600 025

APRIL2021

ANNA UNIVERSITY: CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this project report **"NOVEL IMPLEMENTATION OF CARDIO-VASCULAR DISEASE (CVD) USING MACHINE LEARNING TECHNIQUES"** is the bonafide work of "**R. RAMKUMAR**(421118104070)**"** who carried out the project work under my supervision.

SIGNATURE	**SIGNATURE**
Dr. P.KANIMOZHI (Ph.D),	**Mr.M.O.RAMKUMAR ,**
HEAD OF THE DEPARTMENT,	**SUPERVISOR,**
ASSOCIATE PROFESSOR,	ASSOCIATE PROFESSOR,
Department of Computer Science and Engineering,	Department of Computer Science and Engineering,
IFET College of Engineering,	IFET College of Engineering,
Villupuram – 605 108.	Villupuram – 605 108.

Submitted for the project viva voice examination held on……………………

INTERNAL EXAMINER **EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

I thank the almighty, for the blessings that have been showered upon me to bring forth the success of the project. I would like to express my sincere gratitude to our Chairman **Mr.K.V.Raja**, and our Secretary **Mr.K.Shivram Alva** for providing us with an excellent infrastructure and necessary resources to carry out this project and I extend my gratitude to our principal **Dr.G.Mahendran**, for his constant support of my work.

I also take this opportunity to express my sincere thanks to our Vice-Principal and Dean of Academics **Dr.S.Matilda** and our Head Placement and Corporate Affairs **Prof.S.Viswanathan** has provided all the needful help in executing the project successfully.

I wish to express my thanks to our Head of the Department **Dr.P.Kanimozhi**, Associate Professor for her persistent encouragement and support to complete this project. I express my heartfelt gratitude to my guide **Mr.M.O.RAMKUMAR,** Assistant Professor, Department of Computer Science and Engineering for her priceless guidance and motivation which helped me to bring this project to perfect shape. And I thank our Faculty Advisor **Mr.R.Parthiban,** Associate Professor, Department of Computer Science and Engineering who encouraged me in every step of this project to complete it successfully.

I express my deep sense of thanks to all faculty members and lab technicians in my department for their cooperation and interest shown at every stage of our endeavor in making a project work successful.

Last but not the least, I wholeheartedly thanks to our lovely parents and friends for their moral support in tough times and their constructive criticism which made me succeed in my work.

ABSTRACT

The medical field is growing at a rapid pace with new diseases cropping up daily with the need for the invention of an appropriate course of treatment. The heart is a clenched human fist-sized muscular organ, which is responsible for blood circulation. Though heart/cardiac disease is the name given to diseases affecting the heart in general, many diseases come under this name including coronary artery diseases (CAD), cardiomyopathy, Cardio Vascular Disease (CVD), and so on depending on the circulation of blood throughout the body. To support clinicians in the diagnosis of heart disease, heart disease data prediction has been so designed to analyze medical data with clinical expertise. Through improvement in these predicting systems, there can be an enhancement in the quality of medical diagnostic decisions for heart disease. Data mining plays a crucial part in the prediction of cardiac disease. In this work, the Naive Bayes (NB) classifier, C4.5 classifier, and Artificial Neural Network (RNN)-Back Propagation (BP) methods are used. These traditional methods are utilized for predicting heart disease. When the conditionality of the input is huge, the NB classifier method derived from the Bayesian theorem is used. Despite being simple, it performs better than other protocols. The C4.5 protocol builds decision trees from a training data-set by utilizing data entropy perception. It is a well-known and used protocol and is also called the statistical classifier. For solving several decision modeling problems, RNN has been utilized as a tool in typical cases. The use of RNNs is evidenced in areas of modeling, pattern recognition, data processing, and sequence recognition systems.

TABLE OF CONTENTS

CHAPTER NO.	TITLE	Page No.
	ABSTRACT	**iv**
	LIST OF TABLE	**v**
1	**INTRODUCTION**	**1**
	1.1 GENERAL OVERVIEW	**1**
2	**LITERATURE SURVEY**	**2**
	2.1 LITERATURE SURVEY	2
	2.2 MOTIVATION	4
	2.3 PROBLEM STATEMENT	4
	2.4 RESEARCH OBJECTIVES	5
	2.5 DATA MINING REVIEW	5
	2.6 METHODOLOGY	7
	2.6.1 DATA SOURCE	7
	2.6.2 MINING MODELS	9
	2.6.3 VALIDATING MODEL EFFECTIVENESS	10
	2.6.4 EVALUATION OF MINING GOALS	13
	2.7 BENEFITS AND LIMITATIONS	17
3	**SYSTEM ANALYSIS**	**18**
	3.1 EXISTING SYSTEM	18
	3.2 DISADVANTAGE OF EXISTING SYSTEM	18

4	**PROPOSED SYSTEM**	19
	4. PROPOSED SYSTEM	19
	4.1 ADVANTAGE OF PROPOSED SYSTEM	20
5	**SYSTEM SPECIFICATION**	21
	5.1 SYSTEM REQUIREMENTS	21
6	**SYSTEM DESIGN**	22
	6.1 INTERFAEC DESIGN	22
	6.2 FRONT END DESIGN	22
	6.3 BACK END DESIGN	23
	6.4 SYSTEM ARCHITECTURE	24
7	**FUNCTIONAL ENVIRONMENT**	25
	7.1 FEASIBILITY CONSIDERATION	25
	7.2 BEHAVIORAL FEASIBILITY	26
8	**SYSTEM TESTING**	27
	8.1 TYPES OF TESTS	27
	8.1.1 UNIT TESTING	27
	8.1.2 INTEGRATION TESTING	29
	8.1.3 FUNCTIONAL TEST	29
	8.1.4 SYSTEM TEST	30
	8.1.5 ACCEPTANCE TESTING	30
9	**CONCLUSION**	31

10	**APPENDIX**	32
	APPENDIX 1 - CODE	32
	APPENDIX 2 - LOGIN CODE	41
	APPENDIX 3 - SCREEN SHOTS	44
11	**REFERENCE**	48

ABBREVATION

DB	Database
RDBMS	Relational Database Management System
SQL	Structured Query Language
OCR	Oracle Cluster Registry
SID	System Identifier
GUI	Graphical User Interface
OCI	Oracle Call Interface
DBA	Database Administrator
DBCA	Database Configuration Assistant
GSD	Global Single Database
ONS	Oracle Notification Server
ACID	Atomicity, Consistency, Isolation, Durability
ADT	Abstract Data type
BLOB	Binary Large Object
CLOB	Character Large Object
DBMS	Database Management System
DDL	Data Definition Language
DML	Data Manipulation Language
DTP	Distributed Transaction Processing
ISQL	Interactive SQL
LOB	Large Object
MIS	Management Information Services
NCLOB	National Character Large Object
ODBMS	Object Database Management System
ODL	Object Query Language
OODBMS	Object-Oriented Database Management System
OQL	Object Query Language
ORDBMS	Object-Relational Database Management System
OSQL	Object SQL

OWS	Oracle Web Server
PL/SQL	Procedural Language/SQL
SAG	SQL Access Group
WAN	Wide Area Network
TPS	Transactions per Second
OMF	Oracle Management File

CHAPTER - 1

1. INTRODUCTION

1.1 GENERAL OVERVIEW

A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering effective treatments. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and/or decision support systems. Most hospitals today employ some sort of hospital information system to manage their healthcare or patient data. These systems are designed to support patient billing, inventory management, and the generation of simple statistics. Some hospitals use decision support systems, but they are largely limited. Clinical decisions are often made based on doctors' intuition and experience rather than on the knowledge-rich data hidden in the database. This practice leads to unwanted biases, errors, and excessive medical costs which affects the quality of service provided to patients.

CHAPTER–2

LITERATURESURVEY

1. Survival prediction of heart failure patients using machine learning techniques

Authors: Asif Newaz * , Nadim Ahmed, Farhan Shahriyar Haq

The goal of this research is to develop a reliable decision-support system for the survival prediction of heart failure patients by utilizing their clinical records and laboratory test results. Forecasting heart-failure related events in clinical practice tend to be quite inaccurate and highly variable. Identifying the key drivers of heart failure is also clinically very important. In this regard, we develop a model to accurately identify the patients who are at risk utilizing machine learning techniques. This can help clinicians make informed decisions regarding the intensity of treatment required for a patient. For this study, we have utilized a heart failure data-set originally collected from the Faisalabad Institute of Cardiology and the Allied Hospital in Faisalabad, Pakistan. Sampling strategy is incorporated into the ensemble learning framework to develop a more robust Random Forest Classifier that can effectively deal with the imbalanced nature of the data and provide a more generalization result with higher accuracy. Two different feature selection techniques - Chi-square test and Recursive Feature Elimination are utilized to identify the features that are most significant in terms of survival prediction of heart failure patients. Using our proposed approach, a maximum G-mean score of 76.83% with a sensitivity score of 80.21% was achieved.

2. **Prediction of Cardiac Disease using Supervised Machine Learning Algorithms**

Author: R.Jane Preetha Princy Karunya Institute of Technology and Sciences, Coimbatore

The healthcare industry is dealing with billions of patients all over the world and producing massive data. The machine learning-based models are dissecting the multi dimensional medical data-sets and generating better insights. In this study, a cardiovascular data-set is classified by using several state-of-the-art Supervise d Machine Learning algorithms that are precisely used for disease prediction. The results indicate that the Decision Tree classification model predicted the cardiovascular disease s better than Naive Bayes, Logistic Regression, Random Forest, SVM and KNN based approaches. The Decision Tree bequeathed the best result with the accuracy of 73%. This approach could be helpful for doctors to predict the occurrence of heart disease s in advance and provide appropriate treatment.

3. **Heart Disease Prediction Using Machine Learning Algorithms**

Author: Archana Singh, Rakesh Kumar

Machine learning is the sub branch of artificial intelligence and it is making computers to learn from data without being explicitly programmed Heart disease prediction is used to determine the root cause of getting heart attack and the probability of getting a heart attack, group the people into different clusters based on getting heart attack or not There are five levels in heart attack from level 0 to level 4. There are 14 important attributes to be considered in analysis of heart attack namely age, BP, CHOL, gender, CP, CA, THAL.

2.2. Motivation

A major challenge facing healthcare organizations (hospitals, medical centers) is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering effective treatments. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information and/or decision support systems.

Most hospitals today employ some sort of hospital information system to manage their healthcare or patient data. These systems typically generate huge amounts of data which take the form of numbers, text, charts, and images. Unfortunately, these data are rarely used to support clinical decision-making. There is a wealth of hidden information in these data that is largely untapped. This raises an important question: "How can we turn data into useful information that can enable healthcare practitioners to make intelligent clinical decisions?" This is the main motivation for this research.

2.3. Problem statement

Many hospital information systems are designed to support patient billing, inventory management, and the generation of simple statistics. Some hospitals use decision support systems, but they are largely limited. They can answer simple queries like "What is the average age of patients who have heart disease?", "How many surgeries had resulted in hospital stays longer than 10 days?" "Identify the female patients who are single, above 30 years old, and who have been treated for cancer." However, they cannot answer complex queries like "Identify the important preoperative predictors that increase the length of hospital stay", "Given patient records on cancer, should treatment include chemotherapy alone, radiation alone, or both chemotherapy and radiation?", and "Given patient records, predict the probability of patients getting a heart disease".

Clinical decisions are often made based on doctor's intuition and experience rather than on the knowledge-rich data hidden in the database. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. Wu, et al proposed that integration of clinical decision support with computer-based patient records could reduce medical errors, enhance patient safety, decrease unwanted practice variation, and improve patient outcome. This suggestion is promising as data modeling and analysis tools, e.g., data mining, have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions.

2.4. Research objectives

The main objective of this research is to develop a prototype Intelligent Heart Disease Prediction System (IHDPS) using three data mining modeling techniques, namely, Decision Trees, Naive Bayes, and Neural networks.

IHDPS can discover and extract hidden knowledge (patterns and relationships) associated with heart disease from a historical heart disease database. It can answer complex queries for diagnosing heart disease and thus assist healthcare practitioners to make intelligent clinical decisions that traditional decision support systems cannot. Providing effective treatments also helps to reduce treatment costs. To enhance visualization and ease of interpretation, it displays the results both in tabular and graphical forms.

2.5. Data mining review

Although data mining has been around for more than two decades, its potential is only being realized now. Data mining combines statistical analysis, machine learning, and database technology to extract hidden patterns and relationships from large databases . Fayyad defines data mining as "a process of nontrivial extraction of implicit, previously unknown and potentially useful information from the data stored in a database”. Giudici defines it as “a process of selection, exploration, and modeling of large quantities of data to discover regularities or relations that are at first unknown to obtain clear and useful results for the owner of the database”. Data mining uses two strategies: supervised and unsupervised learning. In supervised learning, a training set is used to learn model parameters whereas in unsupervised learning no training set is used (e.g., k-means clustering is unsupervised). Each data mining technique serves an afferent purpose depending on the modeling objective. The two most common modeling objectives are classification and prediction. Classification models predict categorical labels (discrete, unordered) while prediction models predict continuous-valued functions. Decision Trees and Neural Networks use classification algorithms while Regression, Association Rules, and Clustering use prediction algorithms .

Decision Tree algorithms include CART (Classification and Regression Tree), ID3 (Iterative Dichotomized 3), and C4.5. These algorithms differ in the selection of splits when to stop a node from splitting, and assignment of class to a non-split node. CART uses the Gini index to measure the impurity of a partition or set of training tuples. It can handle high-dimensional categorical data. Decision Trees can also handle continuous data (as in regression) but they must be converted to categorical data.

Naive Bayes or Bayes’ Rule is the basis for many machine-learning and data mining methods. The rule (algorithm) is used to create models with predictive capabilities. It provides new ways of exploring and understanding data. It learns from the “evidence” by calculating the correlation between the target (i.e., dependent) and other (i.e., independent) variables.

Neural Networks consist of three layers: input, hidden and output units (variables). The connection between input units and hidden and output units is based on the relevance of the assigned value (weight) of that particular input unit. The higher the weight the more important it is. Neural Network algorithms use Linear and Sigmoid transfer functions. Neural Networks are suitable for training large amounts of data with few inputs. It is used when other techniques are unsatisfactory.

2.6. Methodology

IHDPS uses the CRISP-DM methodology to build the mining models. It consists of six major phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The business understanding phase focuses on understanding the objectives and requirements from a business perspective, converting this knowledge into a data mining problem definition, and designing a preliminary plan to achieve the objectives. The data understanding phase uses the raw data and proceeds to understand the data, identify its quality, gain preliminary insights, and detect interesting subsets to form hypotheses for hidden information. The data preparation phase constructs the final data-set that will be fed into the modeling tools. This includes table, record, and attribute selection as well as data cleaning and transformation. The modeling phase selects and applies various techniques, and calibrates their parameters to optimal values. The evaluation phase evaluates the model to ensure that it achieves the business objectives. The deployment phase specifies the tasks that are needed to use the models. Data Mining Extension (DMX), a SQL-style query language for data mining, is used for building and accessing the models' contents. Tabular and graphical visualizations are incorporated to enhance the analysis and interpretation of results.

2.6.1. Data source

A total of 909 records with 15 medical attributes (factors) were obtained from the Cleveland Heart Disease Database. Figure 1 lists the attributes. The records were split equally into two data-sets: the training data-set (455 records) and the testing data-set (454 records). To avoid bias, the records for each set were selected randomly.

For the sake of consistency, only categorical attributes were used for all three models. All the non-categorical medical attributes were transformed into categorical data. The attribute “Diagnosis” was identified as the predictable attribute with a value of “1” for patients with heart disease and a value of “0” for patients with no heart disease.

The attribute “Patient ID” was used as the key; the rest are input attributes. It is assumed that problems such as missing data, inconsistent data, and duplicate data have all been resolved

Predictable attribute

1. Diagnosis (value 0: < 50% diameter narrowing (no heart disease); value 1: > 50% diameter narrowing (has heart disease))

Key attribute

1. PatientID – Patient’s identification number

Input attributes

1. Sex (value 1: Male; value 0: Female)

2. Chest Pain Type (value 1: typical type 1 angina, value 2: typical type angina, value 3: non-angina pain; value 4: asymptomatic)

3. Fasting Blood Sugar (value 1: > 120 mg/dl; value 0: < 120 mg/dl)

4. Restecg – resting electrographic results (value 0: normal; value 1: 1 having ST-T wave abnormality; value 2: showing probable or definite left ventricular hypertrophy)

5. Exam – exercise-induced angina (value 1: yes; value 0: no)

6. Slope – the slope of the peak exercise ST segment (value 1: unsloping; value 2: flat; value 3: down sloping)

7. CA – number of major vessels colored by fluoroscope (value 0 – 3)

8. Thal (value 3: normal; value 6: fixed defect; value 7: reversible defect)

9. Trest Blood Pressure (mm Hg on admission to the hospital)

10. Serum Cholesterol (mg/dl)

11. Thalach – maximum heart rate achieved

12. Oldpeak – ST depression induced by exercise relative to rest

13. Age in Year

2.6.2. Mining models

Data Mining Extension (DMX) query language was used for model creation, model training, model prediction, and model content access. All parameters were set to the default setting except for parameters "Minimum Support =1" for Decision Tree and "Minimum Dependency Probability = 0.005" for Naive Bayes [10]. The trained models were evaluated against the test data-sets for accuracy and effectiveness before they were deployed in IHDPS. The models were validated using Lift Chart and Classification Matrix.

Counts for Neural Network on Diagnosis Group

Predicted	0 (Actual)	1 (Actual)
0	211	30
1	35	178

Counts for Neural Network on Diagnosis Group

Predicted	0 (Actual)	1 (Actual)
0	211	30
1	35	178

2.6.3. Validating model effectiveness

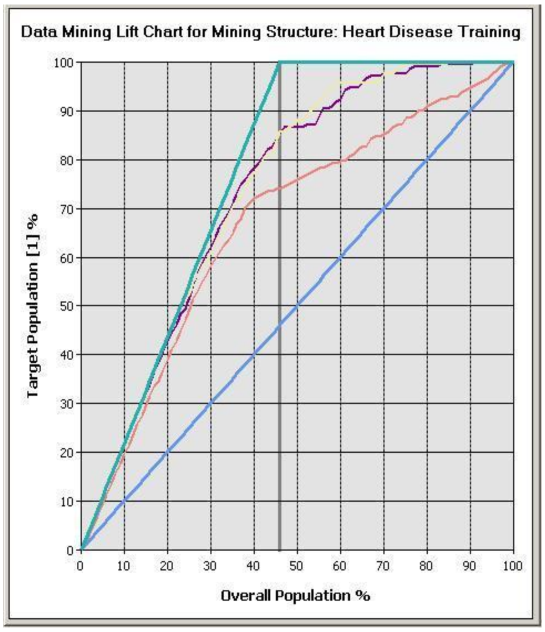
The effectiveness of models was tested using two methods: Lift Chart and Classification Matrix. The purpose was to determine which model gave the highest percentage of correct predictions for diagnosing patients with a heart disease.

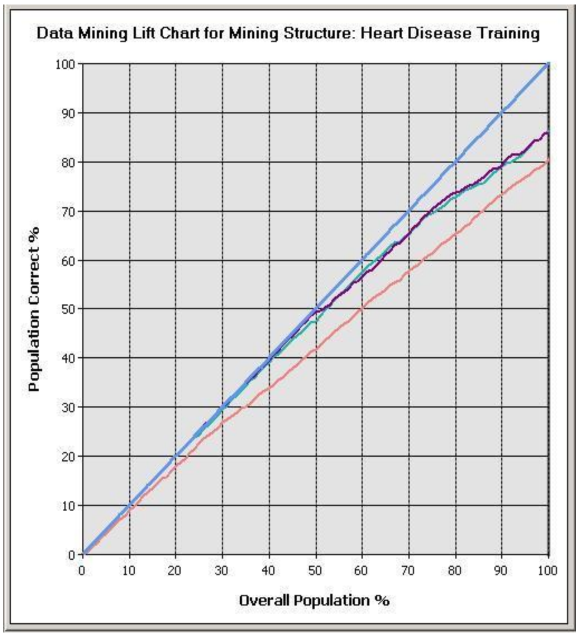
Lift Chart with predictable value. To determine if there was sufficient information to learn patterns in response to the predictable attribute, columns in the trained model were mapped to columns in the test data-set. The model, predictable column to chart against, and the state of the column to predict patients with heart disease (predict value = 1) were also selected. Figure 2 shows the Lift Chart output. The X-axis shows the percentage of the test data-set used to compare predictions while the Y-axis shows the percentage of values predicted to the specified state. The blue and green lines show the results for the random-guess and ideal model respectively. The purple, yellow and red lines show the results of the Neural Network, Naive Bayes, and Decision Tree models respectively. The top green line shows the ideal model; it captured 100% of the target population for patients with heart disease using 46% of the test data-set. The bottom blue line shows the random line which is always a 45-degree line across the chart. It shows that if we randomly guess the result for each case, 50% of the target population would be captured using 50% of the test data-set. All three model lines (purple, yellow, and red) fall between the random-guess and ideal model lines, showing that all three have sufficient information to learn patterns in response to the predictable state.

Counts for Neural Network on Diagnosis Group

Predicted	0 (Actual)	1 (Actual)
0	211	30
1	35	178

Lift Chart with no predictable value. The steps for producing Lift Chart are similar to the above except that the state of the predictable column is left blank. It does not include a line for the random-guess model. It tells how well each model fared at predicting the correct number of the predictable attribute. Figure 3 shows the Lift Chart output. The X-axis shows the percentage of test data-set used to compare predictions while the Y-axis shows the percentage of predictions that are correct. The





blue, purple,green and red lines show the ideal, Neural Network, Naive Bayes and Decision Trees models respectively. The chart shows the performance of the models across all possible states. The model ideal line (blue) is at 45-degree angle, showing that if 50% of the test data-set is processed, 50% of test data-set is predicted correctly.

Counts for Decision Tree on Diagnosis Group

Predicted	0 (Actual)	1 (Actual)
0	219	62
1	27	146

The chart shows that if 50% of the population is processed, Neural Network gives the highest percentage of correct predictions (49.34%) followed by Naive Bayes (47.58%) and Decision Trees (41.85%). If the entire population is processed, Naive Bayes model appears to perform better than the other two as it gives the highest

number of correct predictions (86.12%) followed by Neural Network (85.68%) and Decision Trees (80.4%).

Processing less than 50% of the population causes the Lift lines for Neural Network and Naive Bayes to be always higher than that for Decision Trees, indicating that Neural Network and Naive Bayes are better at making high percentage of correct predictions than Decision Trees. Along the X-axis the Lift lines for Neural Network and Naive Bayes overlap, indicating that both models are equally good for predicting correctly. When more than 50% of population is processed, Neural Network and Naive Bayes appear to perform better as they give high percentage of correct predictions than Decision Trees. This is because the Lift line for Decision Trees is always below that of Neural Network and Naive Bayes. For some population range, Neural Network appears to fare better than Naive Bayes and vice-versa.

Classification Matrix. Classification Matrix displays the frequency of correct and incorrect predictions. It compares the actual values in the test data-set with the predicted values in the trained model. In this example, the test data-set contained 208 patients with heart disease and 246 patients without heart disease. Figure 4 shows the results of the Classification Matrix for all the three models. The rows represent predicted values while the columns represent actual values (1 for patients with heart disease,'0' for patients with no heart disease). The left-most columns show values predicted by the models. The diagonal values show correct predictions

Counts for Naive Bayes on Diagnosis Group

Predicted	0 (Actual)	1 (Actual)
0	211	28
1	35	180

Figure 5 summarizes the results of all three models. Naive-Bayes appears to be most effective as it has the highest percentage of correct predictions (86.53%) for patients with heart disease, followed by Neural Network (with a difference of less than 1%) and Decision Trees. Decision Trees, however, appears to be most effective for predicting patients with no heart disease (89%) compared to the other two models

2.6.4. Evaluation of Mining Goals

Five mining goals were defined based on exploration of the heart disease data-set and objectives of this research. They were evaluated against the trained models. Results show that all three models had achieved the stated goals, suggesting that they could be used to provide decision support to doctors for diagnosing patients and discovering medical factors associated with heart disease. The goals are as follows:

Goal 1:

Given patients' medical profiles, predict those who are likely to be diagnosed with heart disease. All three models were able to answer this question using singleton query and batch or prediction join query. Both queries could predict on single input cases and multiple input cases respectively. IHDPS supports prediction using "what if" scenarios. Users enter values of medical attributes to diagnose patients with heart disease. For example, entering values Age = 70, CA = 2, Chest Pain Type = 4, Sex = M, Slope = 2 and Thal = 3 into the models, would produce the output in Figure 6. All three models showed that this patient has a heart disease. Naive Bayes gives the highest probability (95%) with 432 supporting cases, followed closely by Decision Tree (94.93%) with 106 supporting cases and Neural Network (93.54%) with 298 supporting cases. As these values Are high, doctors could recommend that the patient should undergo further heart examination. Thus performing "what if" scenarios can help prevent a potential heart attack.

Goal 2:

Identify the significant influences and relationships in the medical inputs associated with the predictable state – heart disease. The Dependency viewer in Decision Trees and Naive Bayes models shows the results from the most significant to the least significant (weakest) medical predictors. The viewer is especially useful when there are many predictable attributes. Figures 7 and 8 show that in both models, the most significant factor influencing heart disease is "Chest Pain Type". Other significant factors include Thal, CA and Exang. Decision Trees model shows 'Trest Blood Pressure" as the weakest factor while Naive Bayes model shows 'Fasting Blood Sugar' as the weakest factor. Naive Bayes appears to fare better than Decision Trees as it shows the significance of all input attributes. Doctors can use this information to further analyze the strengths and weaknesses of the medical attributes associated with heart disease.

Goal 3:
Identify the impact and relationship between the medical attributes in relation to the predictable state – heart disease. Identifying the impact and relationship between the medical attributes in relation to heart disease is only found in Decision Trees viewer (Figure 9). It gives a high probability (99.61%) that patients with heart disease are found in the relationship between the attributes (nodes): "Chest Pain Type = 4 and CA = 0 and Exang = 0 and Trest Blood Pressure >= 146.362 and < 158.036." Doctors can use this information to perform medical screening on these four attributes instead of on all attributes on patients who are likely to be diagnosed with heart disease. This will reduce medical expenses, administrative costs, and diagnosis time. Information on least impact (5.88%) is found in the relationship between the attributes: "Chest Pain Type not = 4 and Sex = F". Also given is the relationship between attributes for patients with no heart disease. Results show that the relationship between the attributes: "Chest Pain Type not = 4 and Sex = F" has the highest impact (92.58%). The least impact (0.2%) is found in the attributes: "Chest Pain Type = 4 and CA = 0 and Exang = 0 and Trest Blood Pressure >= 146.362 and < 158.036". Additional information such as identifying

patients' medical profiles based selected nodes can also be obtained by using the drill through function.

Doctors can use the Decision Tree viewer to perform

Goal 4: Identify characteristics of patients with heart disease. Only Naive Bayes model identifies the characteristics of patients with heart disease. It shows the probability of each input attribute for the predictable state. Figure 10 shows that 80% of the heart disease patients are males (Sex = 1) of which 43% are between ages 56 and 63. Other significant characteristics are: high probability in fasting blood sugar with less than 120 mg/dl reading, chest pain type is asymptomatic, slope of peak exercise is flat, etc. Figure 11 shows the characteristics of patients with no heart disease with high

Attributes	Values	Probability %
FastingBloodSugar	FastingBloodSugar = 0	86.179
Exang	Exang = 0	83.74
CA	ca = 0	80.488
Thal	thal = 3	79.268
Oldpeak	Oldpeak < 0.63	67.073
Slope	slope = 1	65.854
Restecg	Restecg = 0	57.724
Sex	Sex = 1	56.911
Sex	Sex = 0	43.089
Restecg	Restecg = 2	41.463
Chest	ChestPainType = 3	41.057
ThalachMaxHeartRate	ThalachMaxHeartRate >= 167.58	38.211
	1 2 3 4	

probability in fasting blood sugar with less than 120 mg/dl reading, no exercise induced, number of major vessels is zero, etc. These results can be further analyzed

Goal 5:

Determine the attribute values that differentiate nodes favoring and disfavoring the predictable states: (1) patients with heart disease (2) patients with no heart disease. This query can be answered by analyzing the results of attribute discrimination viewer of Naive Bayes and Neural Network models. The viewer provides information on the impact of all attribute values that relate to the predictable state.Naive Bayes model (Figure 12) shows the most important attribute favoring patients with heart disease: “Chest Pain Type = 4” with 158 cases and 56 patients with no heart disease. The input attributes “Thal = 7” with 123 (75.00%) patients, “Exang = 1” with 112 (73.68%) patients,” Slope =2” with 138 (66.34%) patients, etc. also favor predictable state. In contrast, the attributes “Thal = 3” with 195 (73.86%) patients, “CA = 0” with 198 (73.06%) patients, “Exang = 0” with 206 (67.98%), etc. favor predictable state for patients with no heart disease.

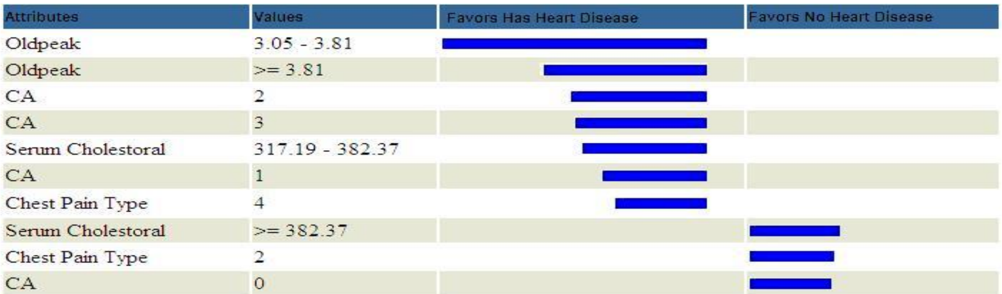


Figure 10. A Tornado Chart for Attribute Discrimination Viewer in descending order for Naive Bayes

Neural Network model (Figure 13) shows that the most important attribute value that favors patients with heart disease is “Old peak = 3.05 – 3.81” (98%). Other attributes that favor heart disease include “Old peak >=3.81”, “CA=2”, “CA=3”, etc. Attributes like “Serum Cholesterol >= 382.37”, “Chest Pain Type = 2”, “CA =0”, etc. also favor the predictable state for patients with no heart disease

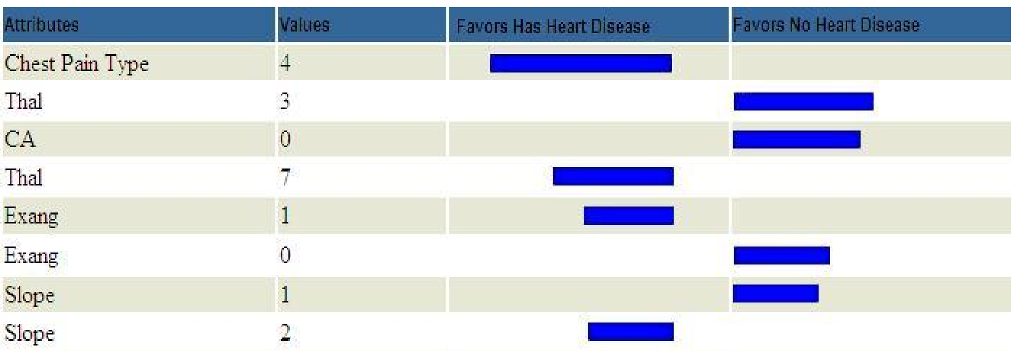


Figure 11. Attribute Discrimination Viewer in descending order for Neural Network

2.7. Benefits and limitations

IHDPS can serve a training tool to train nurses and medical students to diagnose patients with heart disease. It can also provide decision support to assist doctors to make better clinical decisions or at least provide a "second opinion." The current version of IHDPS is based on the 15 attributes listed in Figure 1. This list may need to be expanded to provide a more comprehensive diagnosis system. Another limitation is that it only uses categorical data. For some diagnosis, the use of continuous data may be necessary. Another limitation is that it only uses three data mining techniques. Additional data mining techniques can be incorporated to provide better diagnosis. The size of the data-set used in this research is still quite small. A large data-set would definitely give better results. It is also necessary to test the system extensively with input from doctors, especially cardiologists, before it can be deployed in hospitals. [Access to the system is currently restricted to stakeholders.

CHAPTER – 3

3. SYSTEM ANALYSIS

3.1 EXISTING SYSTEM

Clinical decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in the database. **Medical Misdiagnoses** are a serious risk to our healthcare profession. If they continue, then people will fear going to the hospital for treatment. We can put an end to medical misdiagnosis by informing the public and filing claims and suits against the medical practitioners at fault. There are many ways that a medical misdiagnosis can present itself. Whether a doctor is at fault, or hospital staff, a misdiagnosis of a serious illness can have very extreme and harmful effects. This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients. The National Patient Safety Foundation cites that **42% of medical patients feel they have had experienced a medical error or missed diagnosis**. Patient safety is sometimes negligently given the back seat for other concerns, such as the cost of medical tests, drugs, and operations.

3.2 DISADVANTAGES OF EXISTING SYSTEM

- There are many ways that a medical misdiagnosis can present itself. Whether a doctor is at fault, or hospital staff, a misdiagnosis of a serious illness can have very extreme and harmful effects.
- This practice leads to unwanted biases, errors and excessive medical costs which affects the quality of service provided to patients.
- The National Patient Safety Foundation cites that 42% of medical patients feel they have had experienced a medical error or missed diagnosis.
- Patient safety is sometimes negligently given the back seat for other concerns, such as the cost of medical tests, drugs, and operations

CHAPTER – 4

4. PROPOSED SYSTEM

This proposed system acts as a decision support system and will prove to be an aid for the physicians with diagnosis , the algorithm, Fuzzy c means uses clustering and makes use of clusters and data points to predict the relativity of an attribute, each point is associated with multiple clusters depending upon the membership degrees, the training data is trained by using proposed machine learning algorithm yolov5 classification clustering and adaboost algorithm is explained in detail. CNN-based object detectors are primarily applicable for recommendation systems. YOLO (You only Look Once) models are used for object detection with high performance. YOLO divides an image into a grid system and each grid detects objects within itself. They can be used for real-time object detection based on the data streams. They require very few computational resources

The main objective of this research is to develop a Intelligent Heart Disease Prediction System using three data mining modeling technique, namely, Naive Bayes. It is implemented as web based questionnaire application Based on the user answers, it can discover and extract hidden knowledge (patterns and relationships) associated with heart disease from a historical heart disease database. It can answer complex queries for diagnosing heart disease and thus assist healthcare practitioners to make intelligent clinical decisions which traditional decision support systems cannot. By providing effective treatments, it also helps to reduce treatment costs.

4.1 ADVANTAGES OF PROPOSED SYSTEM

- This suggestion is promising as data modeling and analysis tools, e.g., data mining, have the potential to generate a knowledge-rich environment which can help to significantly improve the quality of clinical decisions.

- The main objective of this research is to develop a prototype Intelligent Heart Disease Prediction System (IHDPS) using three data mining modeling techniques, namely, Decision Trees, Naive Bayes and Neural Network.

- So, it providing effective treatments, it also helps to reduce treatment costs. To enhance visualization and ease of interpretation.

- High performance and accuracy rate

- Yolo classifications very flexible and is widely in various domains with high rate of success

CHAPTER - 5

SYSTEM SPECIFICATION

5.1 HARDWARE REQUIREMENTS

Server Side

Processor	: Intel
HDD	: Minimum 20 MB Disk Space
RAM	: Minimum 64 MB
Database	: SQL Server 2000

Client Side

Processor	**:** AMD, Intel
HDD	**:** Minimum 30MB free disk space
RAM	**:** Minimum 32MB
OS	**:** Windows 98 or above

5.2 SOFTWARE REQUIREMENTS

Operating System	: Windows XP
Front-End	: ASP.NET with C#
Back-End	: SQL Server 7.0
Web Server	: IIS

CHAPTER - 6

6. SYSTEM DESIGN

6.1 INTERFACE DESIGN

Database and web server act as interface Login, username, password and other role are stored and checked in Database. Employee details are viewed use of database. All the information about the concern is viewed use of database

6.2 FRONT END DESIGN

The front end designed using ASP.Net Using c# Language and Standard controls, toolbox. Microsoft .Net is a set of Microsoft Software technologies for connecting your world of information, people, systems, and devices. It enables an unprecedented level of software integration through the use of XML, Web services, small, discrete, building block application that connect to each other.

- JIT -> Just in time
- MSIL ->Microsoft Intermediate Language
- CLR ->Common Language Runtime
- BCL ->Base Class Library

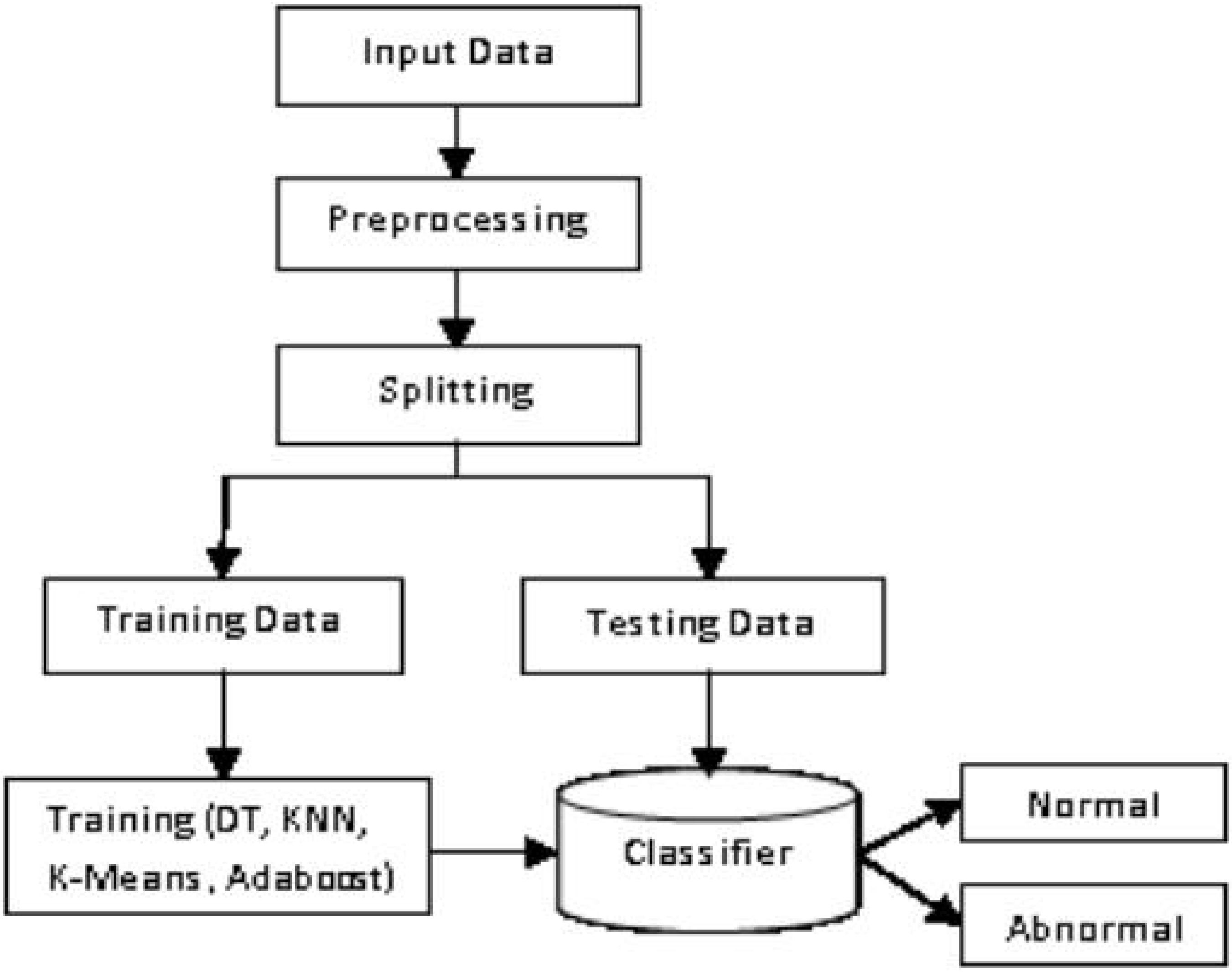
Core Concepts

- Web development platform.
- New programming model.
- Separate layout and business logic.
- Use services provided by the .NET Framework.
- Code is compiled the first time a page is requested
- State management

6.3 BACK END DESIGN

Back end is mainly used to store the data and also retrieve the data. SQL SERVER is one of the most powerful Server. It is used to maintaining the data.

6.4 SYSTEM ARCHITECTURE



CHAPTER - 7

FUNCTIONAL ENVIRONMENT

7.1 FEASIBILITY CONSIDERATION

Three key considerations are involved in feasibility analysis: economic, technical and behavioral. Let's briefly review each consideration and its relation to systems effort.

TECHNICAL FEASIBILITY

Technical feasibility centers on the existing computer system (hardware, software, etc,) and to what extent it can support the proposed addition. For example, if the current computer is operating at 80 percent capacity an arbitrary ceiling. then running another application could overload the system or require additional hardware. This involves financial considerations to accommodate technical enhancements.

ECONOMICAL AND SOCIAL FEASIBILITY

Economic analysis is the most frequently used method for evaluating the effectiveness of a candidate system. More commonly known as cost/benefit analysis, the procedure is to be determining the benefits and savings that are expected from a candidate system and compare them with costs.

Otherwise, further justification or alterations in the proposed system will have to be made if it is to have a chance of being approved. This is ongoing effort that improves accuracy, at each phase of the system life.

7.2 BEHAVIORAL FEASIBILITY

People are inherently resistant to change and computers have been known to facilitate change. An estimate should be made of how strong a reaction the user staff is likely to have towards the development of a computerized system. It is the common knowledge that computer installations have something to do with turnover, transfer, retraining and changes in employee job status. Therefore, it is understandable.

STEPS IN FEASIBILITY STUDY

- Form a project team and appoint a project leader.
- Prepare system flowcharts.
- Enumerate potential candidate system.
- Describe and identify characteristics of candidate systems.
- Form a project team and appointing a project leader.
- Prepare system flowcharts.
- Enumerate potential candidate system.
- Describe and identify characteristics of candidate systems.
- Determine and evaluate performance and cost effectiveness of each candidate system.
- Weight system performance and cost data.
- Select the best candidate system.
- Prepare and report final project directive to management.
- Form a project team and appointing a project leader.
- Prepare system flowcharts.
- Enumerate potential candidate system.
- Describe and identity characteristics of candidate systems.
- Determine and evaluate performance and cost effectiveness.
- Weight system performance and cost data.
- Select the best candidate system.
- Prepare and report the final project directive to management

CHAPTER – 8

8. SYSTEM TESTING

Software testing is an important element of software quality assurance and represents the ultimate review of specification, design and coding. The increasing visibility of s/w as a system element and the costs associated with an s/w failure are motivating forces for well planned through testing.

8.1 TYPES OF TESTS

8.1.1 UNIT TESTING

Unit Test comprises the set of tests performed by an individual programmer prior to integration of the unit into a large system It is illustrated as,

Code and Debugging→Unit Testing→Integration

A program unit is usually small enough that the programmer who developed it can test it in great detail and certainly this will be possible when the unit test integrated into an evolving software product.

There are 4 categories of test that a programmer will typically perform on a program unit.

a) Functional Test

Functional Test cases involve exercising the code with the nominal input value for which the expected results are known.

b) Performance Test

Performance Test determines the amount of execution time spend in various parts of unit, program throughput and response time and device utilization by the program unit.

c) Stress Test

Stress Tests are those designed to intentionally break the unit. A great deal can be learned about the strength and limitations of a program by examining the manner in which a program unit breaks.

d) Structure Test

Structure test are concentrated with exercising the internal logic of a program and traversing particular execution path. Program error can be classified as the missing path errors, computational error and domain error.

8.1.2 INTEGRATION TESTING

The first step in the testing is the top-down approach was the integration is carried from low-level module to the top. In the bottom up approach the integration is carried out from the top-level module to the bottom. The modules are generally tested using the bottom-up approach by introducing steps from the top-level function.

8.1.3 FUNCTIONAL TEST

Functional tests provide systematic demonstrations that functions tested are available as specified by the business and technical requirements, system documentation, and user manuals.

Functional testing is centred on the following items:

Valid Input: identified classes of valid input must be accepted. Invalid Input: identified classes of invalid input must be rejected. Functions: identified functions must be exercised.

Output: identified classes of application outputs must be exercised. Systems/Procedures: interfacing systems or procedures must be invoked. Organization and preparation of functional tests is focused on requirements, key functions, or special test cases. In addition, systematic coverage pertaining to identify Business process flows; data fields, predefined processes, and successive processes must be considered for testing. Before functional testing is complete, additional tests are identified and the effective value of current tests is determined.

8.1.4 SYSTEM TEST

System testing ensures that the entire integrated software system meets requirements. It tests a configuration to ensure known and predictable results. An example of system testing is the configuration-oriented system integration test. System testing is based on process descriptions and flows, emphasizing pre-driven process links and integration point

TEST OBJECTIVES

There are several rules that can serve objectives. They are

- Testing is process of executing a program with the intent of finding an error.
- A good test case is one that has a high probability of finding an undiscovered error.
- A successful test is one that uncovers an undiscovered error.

8.1.5 ACCEPTANCE TESTING

User Acceptance Testing is a critical phase of any project and requires significant participation by the end user. It also ensures that the system meets the functional requirements.

CHAPTER – 9

CONCLUSION

A prototype heart disease prediction system is developed using three data mining classification modeling techniques. The system extracts hidden knowledge from a historical heart disease database. DMX query language and functions are used to build and access the models. The models are trained and validated against a test data-set. Lift Chart and Classification Matrix methods are used to evaluate the effectiveness of the models. All three models are able to extract patterns in response to the predictable state. The most effective model to predict patients with heart disease appears to be Naive Bayes followed by Neural Network and Decision Trees.

Five mining goals are defined based on business intelligence and data exploration. The goals are evaluated against the trained models. All three models could answer complex queries, each with its own strength with respect to ease of model interpretation, access to detailed information and accuracy. Naive Bayes could answer four out of the five goals; Decision Trees, three; and Neural Network, two. Although not the most effective model, Decision Trees results are easier to read and interpret. The drill through feature to access detailed patients' profiles is only available in Decision Trees. Naive Bayes fared better than Decision Trees as it could identify all the significant medical predictors. The relationship between attributes produced by Neural Network is more difficult to understand. IHDPS can be further enhanced and expanded. For example, it can incorporate other medical attributes besides the 15 listed in Figure 1. It can also incorporate other data mining techniques, e.g., Time Series, Clustering and Association Rules. Continuous data can also be used instead of just categorical data. Another area is to use Text Mining to mine the vast amount of unstructured data available in healthcare databases. Another challenge would be to integrate data mining and text mining.

CHAPTER 10 - APPENDIX

APPENDIX 1

SOURCE

CODE

```
using System;
using System.Data;
using System.Configuration;
using System.Collections;
using System.Web;
using System.Web.Security;
using System.Web.UI;
using System.Web.UI.WebControls;
using System.Web.UI.WebControls.WebParts;
using System.Web.UI.HtmlControls;
using System.Data.SqlClient;
using System.Web.Configuration;
using System.Net;
using System.Globalization;
using Microsoft.AnalysisServices.AdomdServer;
using Microsoft.AnalysisServices.AdomdClient;
public partial class Default2 : System.Web.UI.Page
{
    AdomdConnection con = new
    AdomdConnection(ConfigurationManager.ConnectionStrings["Mine"].ConnectionSt
    ring);
    protected void Page_Load(object sender, EventArgs e)
    {
        if (Session["Username"] == null || Session["Username"] == "")
        {
```

```
            Response.Redirect("Login.aspx");
        }
        Label2.Text = "Welcome" + "  " + Session["Username"].ToString();
    }
    protected void Button1_Click(object sender, EventArgs e)
    {
```

IMPLEMENTATION OF NAVIE BAYES METHOD

```
Microsoft.AnalysisServices.AdomdClient.AdomdDataReader dr;

Microsoft.AnalysisServices.AdomdClient.AdomdCommand cmd3 = new
Microsoft.AnalysisServices.AdomdClient.AdomdCommand();

        cmd3.CommandText =
"SELECT[naviebaisemethod].[Output],PredictProbability([Output])As
pro ,PredictSupport([Output])As supp From  [naviebaisemethod] NATURAL
PREDICTION JOIN (SELECT " + txtAge.Text + " AS [Age]," +
ddlSex.SelectedValue + "AS [Sex]," + ddlChestpain.SelectedValue + " As [Chest Pain
Type]," + txtRestingbloodPressure.Text + "AS [Trest Blood Pressure], " +
txtserumcholestral.Text + "AS [Serum Cholestorral]," +
ddlFastingBloodSugar.SelectedValue + "AS [Fasting Blood Sugar], " +
ddlrestecng.SelectedValue + " AS [Resting Electrocardiographic Results], " +
txtThalach.Text + "AS [Maximum Heart Rate]," + ddlExang.SelectedValue + " AS
[Exercise Induced Angina]," + txtOldPeak.Text + " AS [ST Depression Induced By
Exercise Relative To Rest]," + ddlSlope.SelectedValue + "AS [The Slope Of The Peak
Exercise ST Segment]," + ddlCA.SelectedValue + " AS [Number Of Major Vessels
Colored By Flourosopy]," + ddlThal.SelectedValue + "  AS [Thal])AS t";

        cmd3.Connection = con;
```

```
con.Open();

string prob;

    dr = cmd3.ExecuteReader();

    if (dr.Read())

{

    txtnbpredic.Text =dr.GetValue(0).ToString();

    prob = dr.GetValue(1).ToString();

    decimal  per = Convert.ToDecimal(prob);

    decimal per1 = System.Math.Round((per * 100),2);

   txtnbprob.Text = Convert.ToString(per1);

  decimal supp=Convert.ToDecimal(dr.GetValue(2).ToString());

   decimal supp1=System.Math.Round(supp);

   txtnbsupport.Text = Convert.ToString(supp1);



}

dr.Close();

con.Close();
```

IMPLEMENTATION OF DECISION TREE METHOD

```
Microsoft.AnalysisServices.AdomdClient.AdomdDataReader dr1 ;

    Microsoft.AnalysisServices.AdomdClient.AdomdCommand cmd4 = new
Microsoft.AnalysisServices.AdomdClient.AdomdCommand();

    cmd4.CommandText =
"SELECT[Descision_tree_method].[Output],PredictProbability([Output])As
pro ,PredictSupport([Output])As supp From  [Descision_tree_method] NATURAL
PREDICTION JOIN (SELECT " + txtAge.Text + " AS [Age]," +
ddlSex.SelectedValue + "AS [Sex]," + ddlChestpain.SelectedValue + " As [Chest Pain
Type]," + txtRestingbloodPressure.Text + "AS [Trest Blood Pressure], " +
txtserumcholestral.Text + "AS [Serum Cholestorral]," +
ddlFastingBloodSugar.SelectedValue + "AS [Fasting Blood Sugar], " +
ddlrestecng.SelectedValue + " AS [Resting Electrocardiographic Results], " +
txtThalach.Text + "AS [Maximum Heart Rate]," + ddlExang.SelectedValue + " AS
[Exercise Induced Angina]," + txtOldPeak.Text + " AS [ST Depression Induced By
Exercise Relative To Rest]," + ddlSlope.SelectedValue + "AS [The Slope Of The Peak
Exercise ST Segment]," + ddlCA.SelectedValue + " AS [Number Of Major Vessels
Colored By Flourosopy]," + ddlThal.SelectedValue + "  AS [Thal])AS t";

    cmd4.Connection = con;

    con.Open();

    dr1 = cmd4.ExecuteReader();

      if (dr1.Read())

    {
```

```
        txtdtpredic.Text =dr1.GetValue(0).ToString();

        prob = dr1.GetValue(1).ToString();

        decimal per = Convert.ToDecimal(prob);

        decimal per1 = System.Math.Round((per * 100), 2);

        txtdtprob.Text =Convert.ToString(per1);

        decimal supp = Convert.ToDecimal(dr1.GetValue(2).ToString());

        decimal supp1 = System.Math.Round(supp);

        txtdtsupport.Text = Convert.ToString(supp1);

    }

    dr1.Close();

    con.Close();
```

IMPLEMENTATION OF NEURAL NETWORK METHOD

```
    Microsoft.AnalysisServices.AdomdClient.AdomdDataReader dr2;

    Microsoft.AnalysisServices.AdomdClient.AdomdCommand cmd5 = new
Microsoft.AnalysisServices.AdomdClient.AdomdCommand();
```

```
cmd5.CommandText =
"SELECT[neural_network_method].[Output],PredictProbability([Output])As pro ,PredictSupport([Output])As supp From  [neural_network_method] NATURAL PREDICTION JOIN (SELECT " + txtAge.Text + " AS [Age]," + ddlSex.SelectedValue + "AS [Sex]," + ddlChestpain.SelectedValue + " As [Chest Pain Type]," + txtRestingbloodPressure.Text + "AS [Trest Blood Pressure], " + txtserumcholestral.Text + "AS [Serum Cholestorral]," + ddlFastingBloodSugar.SelectedValue + "AS [Fasting Blood Sugar], " + ddlrestecng.SelectedValue + " AS [Resting Electrocardiographic Results], " + txtThalach.Text + "AS [Maximum Heart Rate]," + ddlExang.SelectedValue + " AS [Exercise Induced Angina]," + txtOldPeak.Text + " AS [ST Depression Induced By Exercise Relative To Rest]," + ddlSlope.SelectedValue + "AS [The Slope Of The Peak Exercise ST Segment]," + ddlCA.SelectedValue + " AS [Number Of Major Vessels Colored By Flourosopy]," + ddlThal.SelectedValue + "  AS [Thal])AS t";

cmd5.Connection = con;

con.Open();

   dr2 = cmd5.ExecuteReader();

    if (dr2.Read())

{

     txtnnpredic.Text  =dr2.GetValue(0).ToString();

     prob = dr2.GetValue(1).ToString();

    decimal per = Convert.ToDecimal(prob);

     decimal per1 = System.Math.Round((per * 100), 2);
```

```
            txtnnprob.Text = Convert.ToString(per1);

            decimal supp = Convert.ToDecimal(dr2.GetValue(2).ToString());

            decimal supp1 = System.Math.Round(supp);

            txtnnsupport.Text = Convert.ToString(supp1);

        }

        dr2.Close();

        con.Close();

    }
```

IMPLEMENTATION OF INPUT ATTRIBUTTES

```
protected void btnClear_Click(object sender, EventArgs e)
  {
     txtAge.Text ="";
     ddlSex.SelectedIndex=0;
     ddlChestpain.SelectedIndex=0;
     txtRestingbloodPressure.Text = "";
     txtserumcholestral.Text = "";
     ddlFastingBloodSugar.SelectedIndex=0;
     ddlrestecng.SelectedIndex=0;
     txtThalach.Text = "";
     ddlExang.SelectedIndex=0;
     txtOldPeak.Text="";
     ddlSlope.SelectedIndex=0;
     ddlCA.SelectedIndex=0;
     ddlThal.SelectedIndex = 0;
     txtnnpredic.Text="";
```

```
        txtnnprob.Text="";

       txtnnsupport.Text = "";

        txtdtpredic.Text = "";

        txtdtprob.Text = "";

        txtdtsupport.Text = "";

        txtnbpredic.Text = "";

        txtnbprob.Text = "";

        txtnbsupport.Text = "";

    }

    protected void lnkbtnhome_Click(object sender, EventArgs e)

    {

        Response.Redirect("Default3.aspx");

    }

}
```

APPENDIX 2

LOGIN CODE

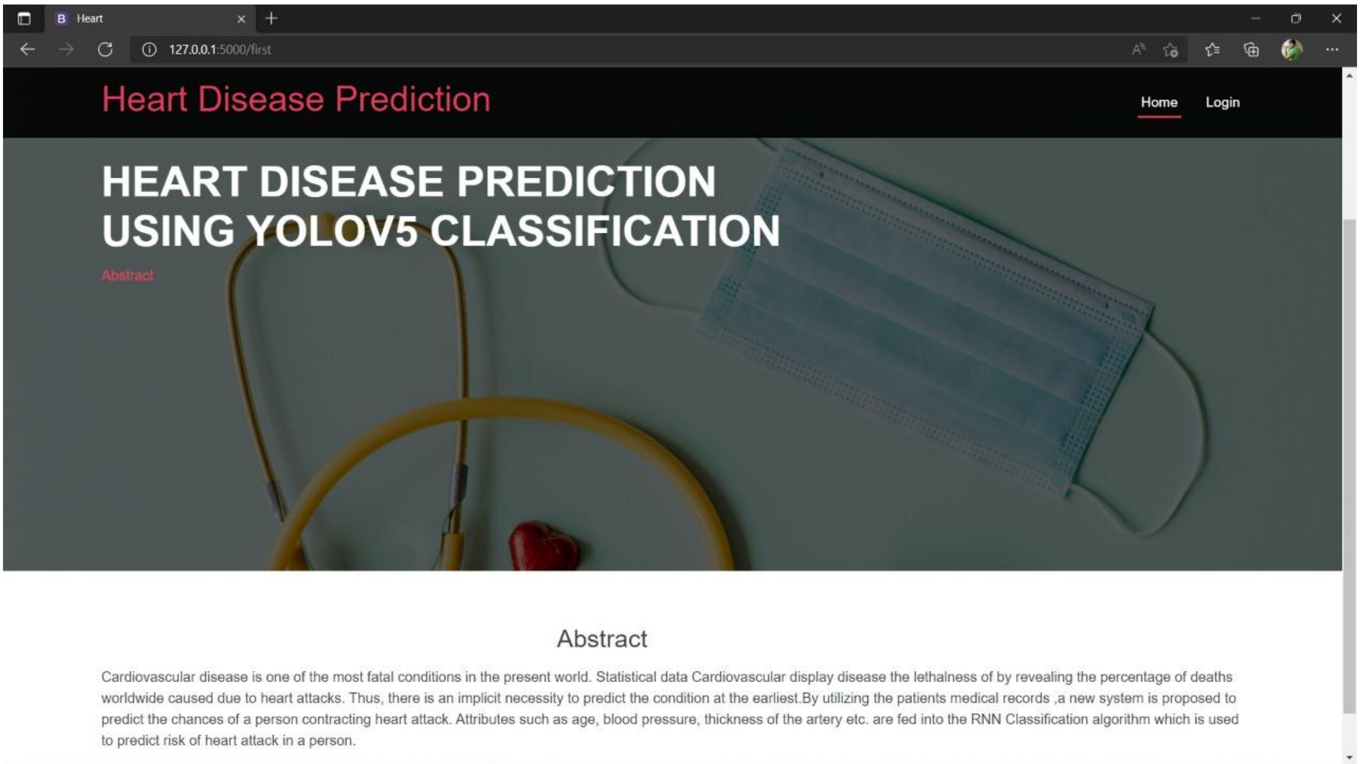
```
using System;

using System.Data;

using System.Configuration;

using System.Collections;

using System.Web;

using System.Web.Security;

using System.Web.UI;

using System.Web.UI.WebControls;

using System.Web.UI.WebControls.WebParts;

using System.Web.UI.HtmlControls;

using System.Data.SqlClient;

using System.Web.Configuration;

public partial class Login : System.Web.UI.Page

{

    SqlConnection con = new
    SqlConnection(WebConfigurationManager.AppSettings["db"]);
```

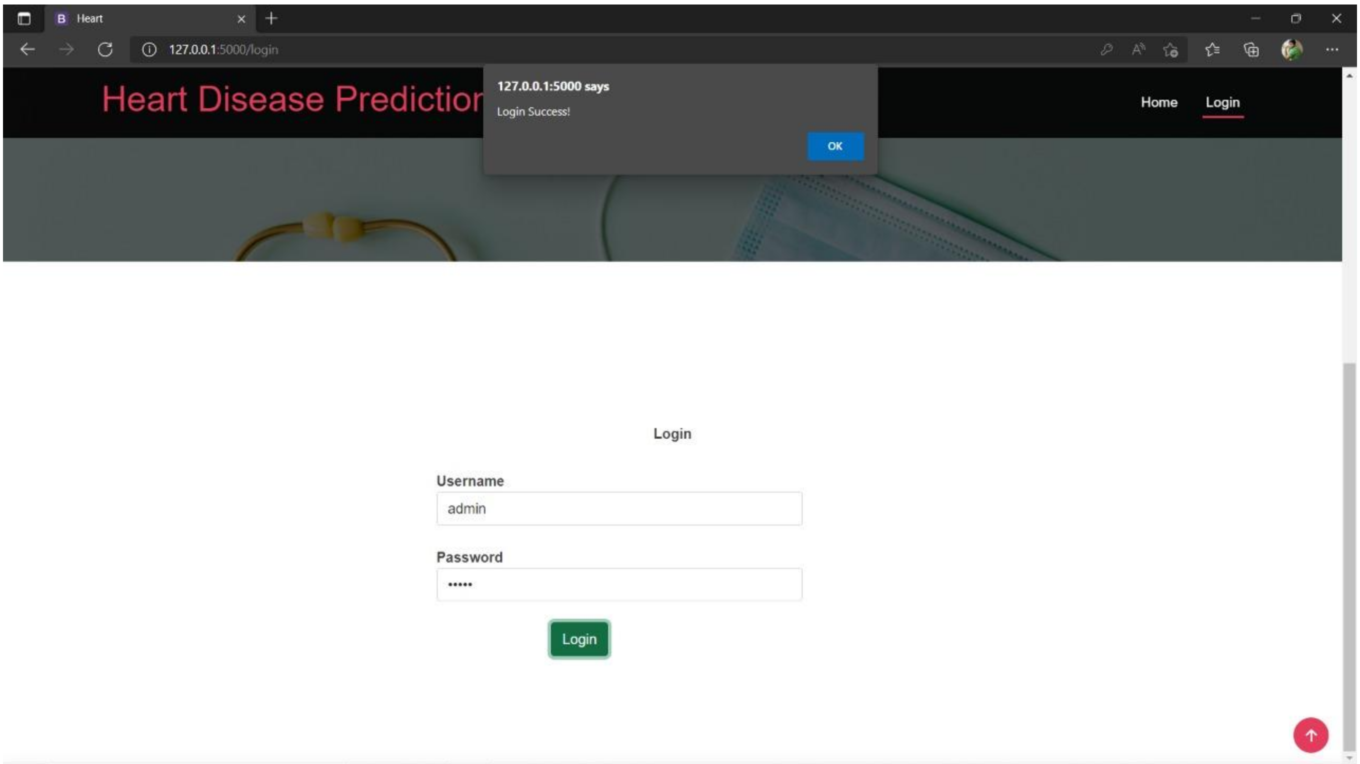
```
SqlCommand cmd;

SqlDataReader dr;

string pwd;

protected void Page_Load(object sender, EventArgs e)

{

}

protected void Button1_Click(object sender, EventArgs e)

{

    con.Open();

    cmd = new SqlCommand("Select Password from UserInfo where UserId='" +
TextBox1.Text + "'", con);

    dr = cmd.ExecuteReader();

    while(dr.Read())

    {

        pwd=dr.GetValue(0).ToString();

    }

    if(pwd==TextBox2.Text.ToString())
```

```
        {
        Session["Username"] = TextBox1.Text.ToString();
        Response.Redirect("Default2.aspx");
        }
        else
        {
           Response.Write("<script>alert('Invalid ! Please Try Again')</script>");
        }

    }
}
```

APPENDIX 3

SCREEN SHOTS





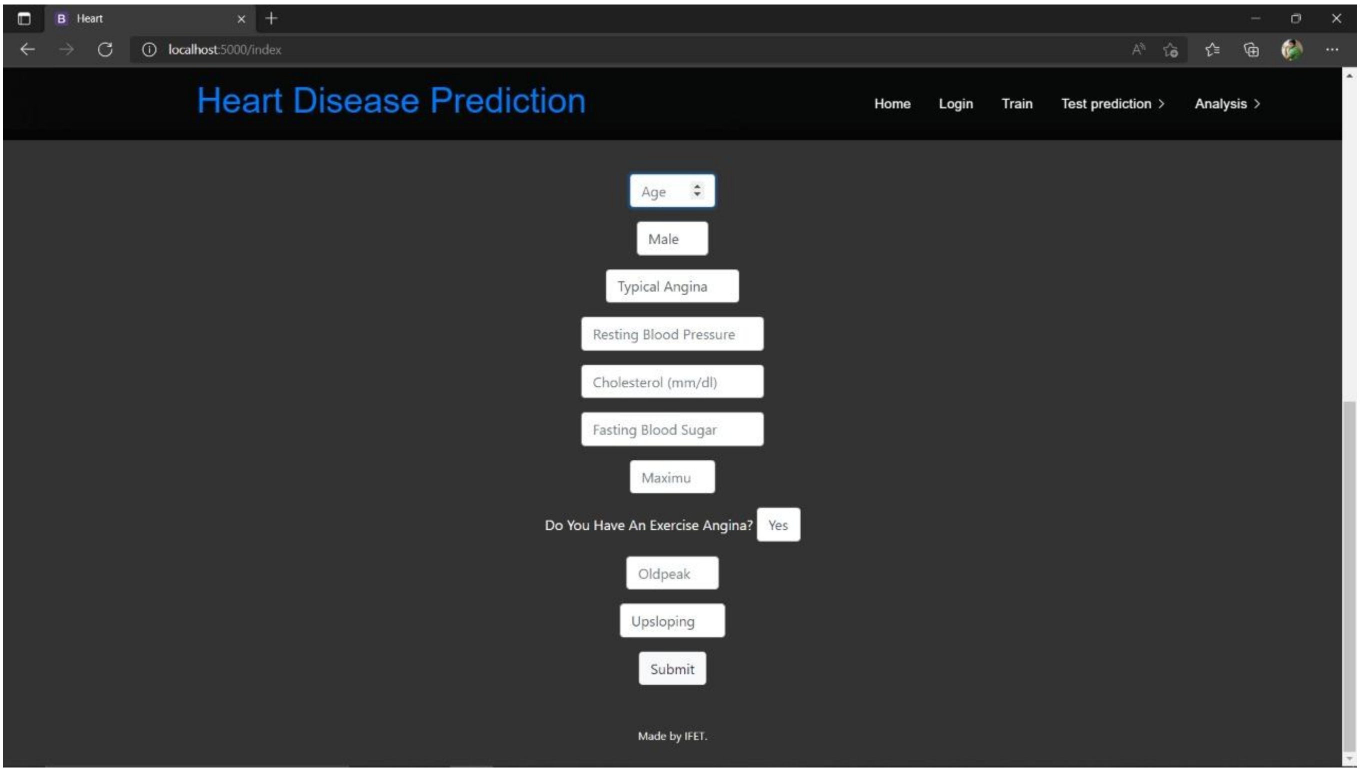


Figure of Input Field

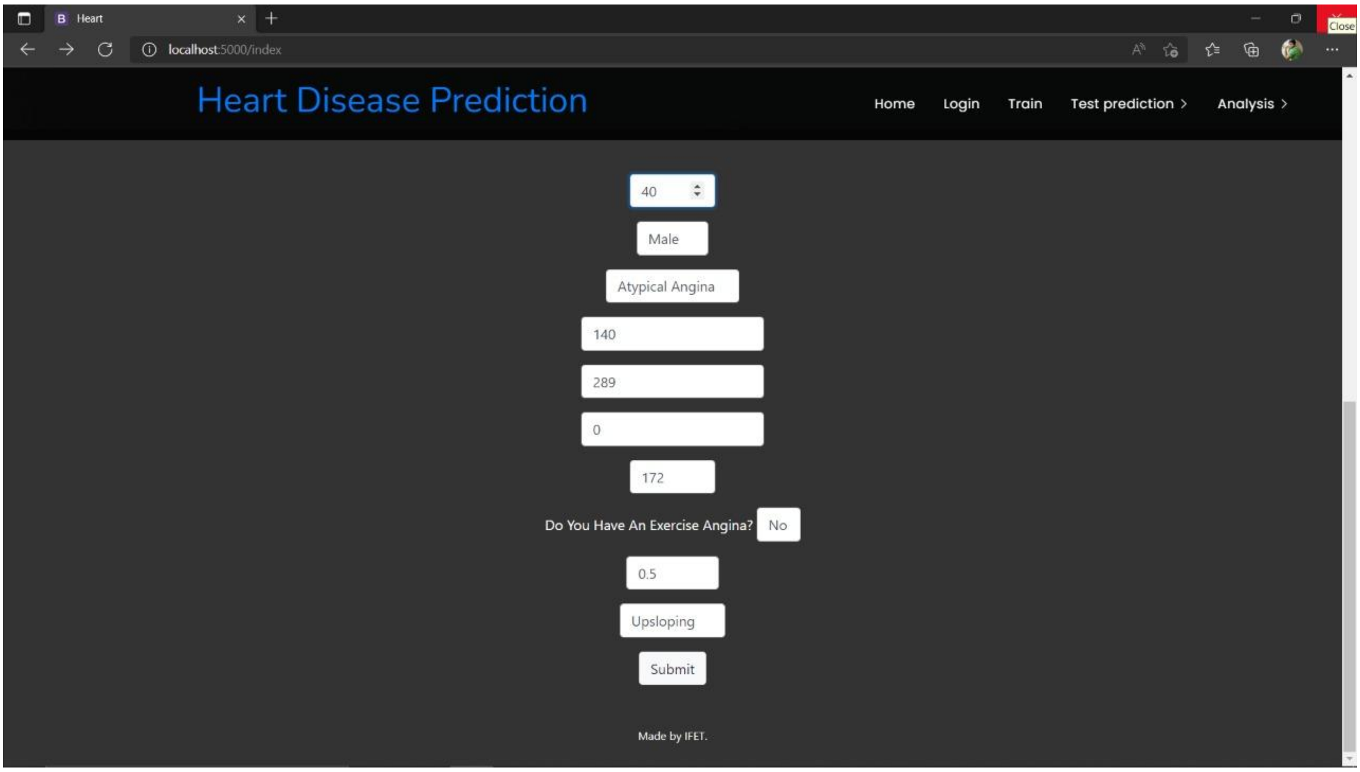


Figure of Input data

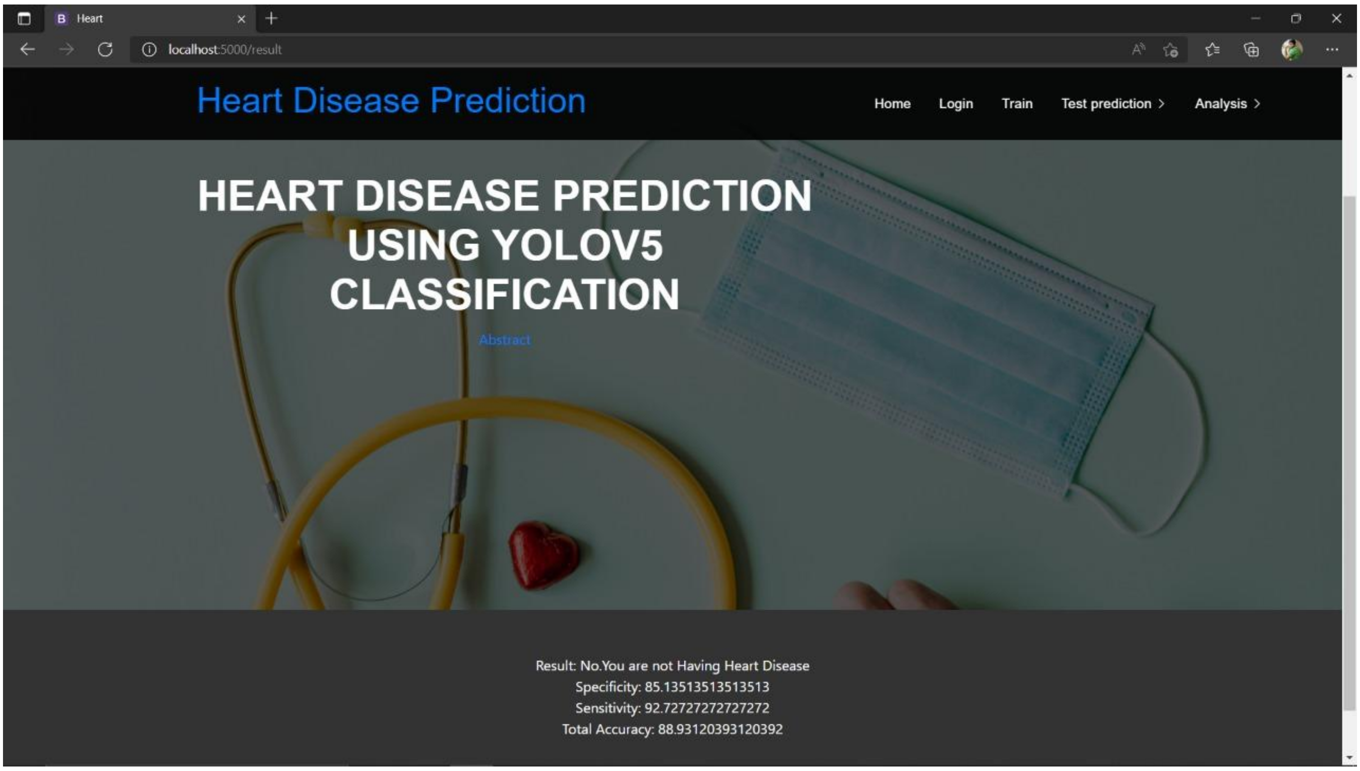


Figure of Output

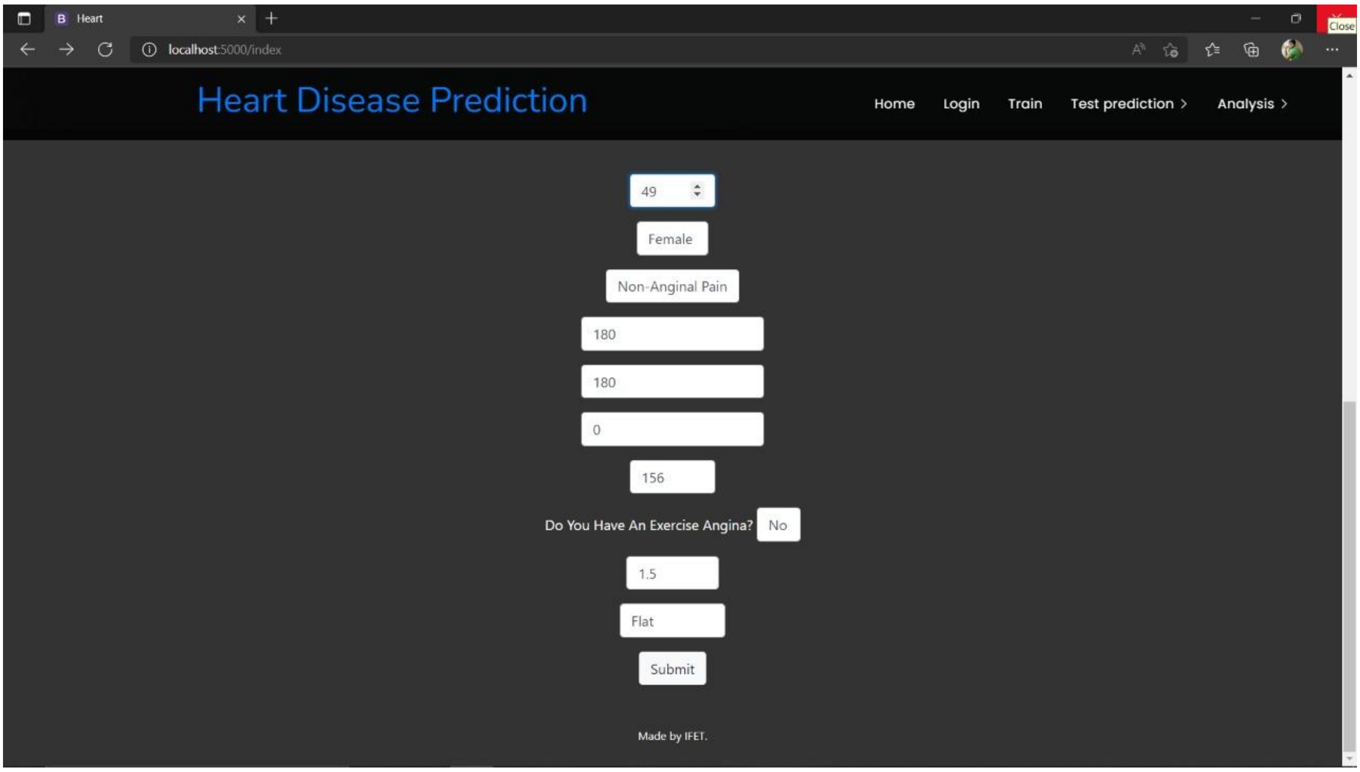


Figure of Input data

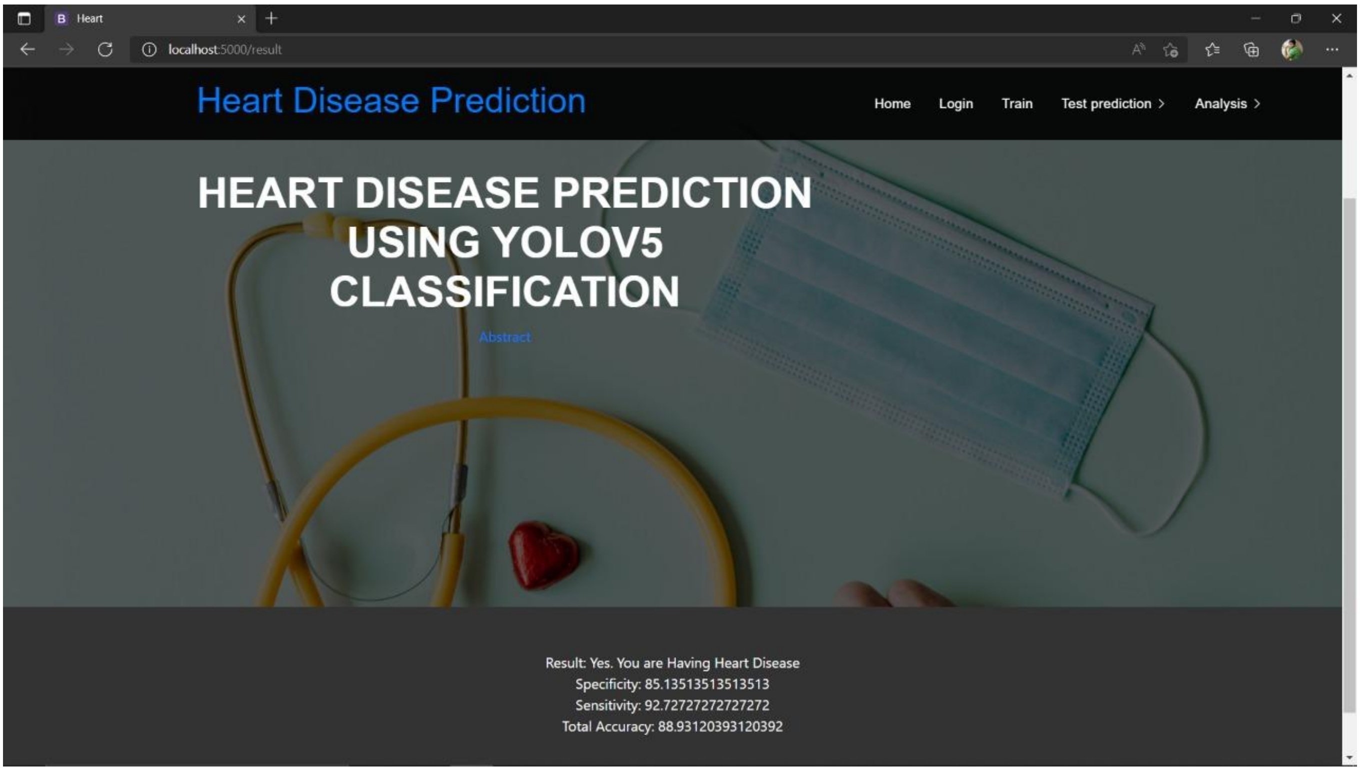


Figure of Output

11. REFERENCES

[1] Blake, C.L., Mertz, C.J.: “UCI Machine Learning Databases”, http://mlearn.ics.uci.edu/databases/heart-disease/, 2004.

[2] Chapman, P., Clinton, J., Kerber, R. Khabeza, T., Reinartz, T., Shearer, C., Wirth, R.: “CRISP-DM 1.0: Step by step data mining guide”, SPSS, 1-78, 2000.

[3] Charly, K.: “Data Mining for the Enterprise”, 31st Annual Hawaii Int. Conf. on System Sciences, IEEE Computer, 7, 295-304, 1998.

[4] Fayyad, U: “Data Mining and Knowledge Discovery in Databases: Implications fro scientific databases”, Proc. of the 9th Int. Conf. on Scientific and Statistical Database Management, Olympia, Washington, USA, 2-11, 1997.

[5] Giudici, P.: “Applied Data Mining: Statistical Methods for Business and Industry”, New York: John Wiley, 2003.

[6] Han, J., Kamber, M.: “Data Mining Concepts and Techniques”, Morgan Kaufmann Publishers, 2006.

[7] Ho, T. J.: “Data Mining and Data Warehousing”, Prentice Hall, 2005.

[8] Kaur, H., Wasan, S. K.: “Empirical Study on Applications of Data Mining Techniques in Healthcare”, Journal of Computer Science 2(2), 194-200, 2006.

[9] Mehmed, K.: “Data mining: Concepts, Models, Methods and Algorithms”, New Jersey: John Wiley, 2003.

[10] Mohd, H., Mohamed, S. H. S.: “Acceptance Model of Electronic Medical

Record", Journal of Advancing Information and Management Studies. 2(1), 75-92, 2005.

[11] Microsoft Developer Network (MSDN). http://msdn2.microsoft.com/en-us/virtuallabs/aa740409.aspx, 2007.

[12] Obenshain, M.K: "Application of Data Mining Techniques to Healthcare Data", Infection Control and Hospital Epidemiology, 25(8), 690–695, 2004.

[13] Sellappan, P., Chua, S.L.: "Model-based Healthcare Decision Support System", Proc. Of Int. Conf. on Information Technology in Asia CITA'05, 45-50, Kuching, Sarawak, Malaysia, 2005

[14] Tang, Z. H., MacLennan, J.: "Data Mining with SQL Server 2005", Indianapolis: Wiley, 2005.

[15] Thuraisingham, B.: "A Primer for Understanding and Applying Data Mining", IT Professional, 28-31, 2000.

[16] Weiguo, F., Wallace, L., Rich, S., Zhongju, Z.: "Tapping the Power of Text Mining", Communication of the ACM. 49(9), 77-82, 2006.

[17] Wu, R., Peters, W., Morgan, M.W.: "The Next Generation Clinical Decision Support: Linking Evidence to Best Practice", Journal Healthcare Information Management. 16(4), 50-55, 2002.